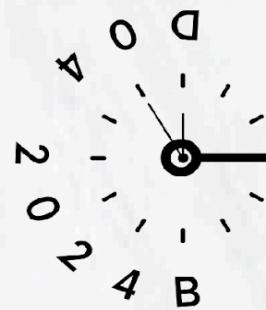


# PROYECTO FINAL

- CASTAÑEDA HERNÁNDEZ GEMMA  
ISELA 304190485
- MONTES ROSALES BRAULIO  
FERMIN 214799184
- JOSÉ CARLOS TAPIA COBIAN  
218632179



## **PROYECTO FINAL**

**¿Qué nivel de felicidad tiene la población mexicana según sus actividades?**

### **1. Entendimiento del negocio**

#### **a. Objetivos del negocio**

##### **i. Antecedentes**

El Instituto Nacional de Estadística y Geografía (INEGI) es un organismo público autónomo de México, responsable de normar y coordinar el Sistema Nacional de Información Estadística y Geográfica. Su función principal es captar y difundir información sobre el territorio, los recursos, la población y la economía del país, con el objetivo de proporcionar datos que faciliten la toma de decisiones y el desarrollo nacional.

La Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019 tiene como propósito generar datos estadísticos que permitan medir las diversas formas de trabajo que realizan las personas, tanto remunerado como no remunerado. Además, busca destacar la relevancia de la producción doméstica y su impacto en la economía, analizar cómo distribuyen su tiempo hombres y mujeres, y evaluar su percepción de bienestar. Todo esto se enfoca en la población de 12 años o más en contextos urbanos y rurales.

##### **ii. Objetivo comercial o de investigación del proyecto**

Analizar y comparar la distribución del tiempo en horas dedicadas a sus actividades para identificar patrones que respondan a la pregunta: "¿Qué nivel de felicidad tiene la población mexicana según sus actividades?". Este análisis busca visibilizar las contribuciones económicas y sociales de diferentes sectores de la población, así como generar información relevante en temas orientados a mejorar la calidad de vida y el bienestar general. Además, se pretende poner de manifiesto la relación del tiempo dedicado a actividades no remuneradas (muchas veces ignoradas) con el nivel de felicidad expresado por la población.

### **iii. Criterios de rendimiento**

- **Análisis Diferenciado:** Identificar y documentar las diferencias en cantidad de horas en el tiempo destinado a actividades esenciales y no esenciales con un enfoque en la relación con el sentimiento de su felicidad percibida.
- **Visualización Clara:** Crear visualizaciones efectivas que permitan interpretar fácilmente los datos, mediante gráficos y tablas de manera comprensible.
- **Relevancia y Aplicabilidad:** Generar conclusiones que sirvan como antecedente para quienes busquen información sobre la población mexicana, su uso del tiempo y su percepción sobre cómo se perciben respecto a su felicidad. Por ejemplo para una persona que esté elaborando una tesis en población mexicana en el área de psicología social. Para evaluar si las conclusiones derivadas del análisis generan insumos para hallazgos que les resultan útiles o aplicables en proyectos de investigación los resultados de este proyecto junto con el dataset, se subirán a Kaggle y contabilizar la cantidad de descargas y citas realizadas en un año, consideramos que de 50 a 100 descargas como rango podría considerarse un buen inicio.
- **Impacto Social:** Generar conocimiento que permita identificar hechos actuales tanto sobre la desigualdad en el uso del tiempo para actividades diversas, como la importancia de las actividades no remuneradas, que puedan utilizarse para futuros proyectos enfocados a grupos vulnerables de México.
- **Rigor Metodológico:** Utilizar técnicas de minería de datos que garanticen la validez y confiabilidad de los resultados obtenidos.

### **b. Valoración de la situación**

#### **i. Inventario de recursos**

- Datos, conocimiento y herramientas

Tipo de datos e información que almacenan:

#### Datos Demográficos:

- **Entidad Federativa:** Clave numérica que identifica la ubicación geográfica de la persona dentro de las 32 entidades federativas de México.
- **Sexo:** Identidad de género de la persona, clasificada como "Hombre" o "Mujer".
- **Edad:** Número de años cumplidos por la persona.

#### Trabajo:

- **Trabajó al menos una hora:** Indica si la persona ha realizado alguna actividad laboral remunerada durante al menos una hora.
- **Horas de trabajo entre semana:** Cantidad de tiempo, en horas, que la persona dedica a actividades laborales en un día promedio de lunes a viernes.
- **Horas de trabajo fin de semana:** Cantidad de tiempo, en horas, dedicado a actividades laborales durante un día promedio de fin de semana.
- **Horas de traslado entre semana:** Tiempo promedio, en horas, invertido en el desplazamiento hacia el lugar de trabajo en un día laboral.

- **Horas de traslado fin de semana:** Tiempo promedio, en horas, invertido en el desplazamiento hacia el lugar de trabajo durante un día de fin de semana.
- **Salario Mensual:** Remuneración económica mensual que recibe la persona por su trabajo.
- **Permisos laborales:** Disponibilidad de permisos en el trabajo por motivos de salud, maternidad o accidentes.
- **Derecho a vacaciones:** Indica si la persona tiene derecho a un periodo vacacional en su empleo.
- **Derecho a jubilación:** Indica si la persona puede acceder a una pensión al finalizar su vida laboral.
- **Derecho a servicios médicos:** Indica si la persona tiene acceso a atención médica a través de su empleo.

#### **Uso del Tiempo:**

- **Horas de sueño entre semana:** Tiempo promedio, en horas, que la persona duerme diariamente entre semana.
- **Horas de sueño fin de semana:** Tiempo promedio, en horas, que la persona duerme diariamente durante el fin de semana.
- **Horas para comer entre semana:** Tiempo promedio diario, en horas, dedicado a la alimentación entre semana.

- **Horas para comer el fin de semana:** Tiempo promedio diario, en horas, dedicado a la alimentación durante el fin de semana.
- **Horas de aseo personal entre semana:** Tiempo promedio diario, en horas, dedicado al aseo personal entre semana.
- **Horas de aseo personal fin de semana:** Tiempo promedio diario, en horas, dedicado al aseo personal durante el fin de semana.
- **Horas de estudio entre semana:** Tiempo promedio diario, en horas, dedicado a actividades académicas entre semana.
- **Horas de estudio fin de semana:** Tiempo promedio diario, en horas, dedicado a actividades académicas durante el fin de semana.
- **Horas trasladando estudiantes entre semana:** Tiempo promedio diario, en horas, dedicado a transportar estudiantes entre semana.
- **Horas trasladando estudiantes fin de semana:** Tiempo promedio diario, en horas, dedicado a transportar estudiantes durante el fin de semana.
- **Horas de quehaceres domésticos entre semana:** Tiempo promedio diario, en horas, dedicado a tareas del

hogar entre semana.

- **Horas de quehaceres domésticos fin de semana:** Tiempo promedio diario, en horas, dedicado a tareas del hogar durante el fin de semana.
- **Horas de ganadería entre semana:** Tiempo promedio diario, en horas, dedicado a la cría de animales para autoconsumo entre semana.
- **Horas de ganadería fin de semana:** Tiempo promedio diario, en horas, dedicado a la cría de animales para autoconsumo durante el fin de semana.
- **Horas de recolección entre semana:** Tiempo promedio diario, en horas, dedicado a la recolección de productos (agrícolas, etc.) para autoconsumo entre semana.
- **Horas de recolección fin de semana:** Tiempo promedio diario, en horas, dedicado a la recolección de productos para autoconsumo durante el fin de semana.
- **Horas de agricultura entre semana:** Tiempo promedio diario, en horas, dedicado a labores agrícolas para autoconsumo entre semana.
- **Horas de agricultura fin de semana:** Tiempo promedio diario, en horas, dedicado a labores agrícolas para autoconsumo durante el fin de semana.

- **Horas de confección casera entre semana:** Tiempo promedio diario, en horas, dedicado a la elaboración de productos textiles o artesanales en el hogar entre semana.
- **Horas de confección casera fin de semana:** Tiempo promedio diario, en horas, dedicado a la elaboración de productos textiles o artesanales en el hogar durante el fin de semana.
- **Horas de mantenimiento del hogar entre semana:** Tiempo promedio diario, en horas, dedicado a la reparación y mantenimiento del hogar entre semana.
- **Horas de mantenimiento del hogar fin de semana:** Tiempo promedio diario, en horas, dedicado a la reparación y mantenimiento del hogar durante el fin de semana.
- **Horas de compras del hogar entre semana:** Tiempo promedio diario, en horas, dedicado a la adquisición de bienes para el hogar entre semana.
- **Horas de compras del hogar fin de semana:** Tiempo promedio diario, en horas, dedicado a la adquisición de bienes para el hogar durante el fin de semana.
- **Horas de trámites y pagos entre semana:** Tiempo promedio diario, en horas, dedicado a la gestión de trámites y pagos entre semana.

- **Horas de trámites y pagos fin de semana:** Tiempo promedio diario, en horas, dedicado a la gestión de trámites y pagos durante el fin de semana.
- **Horas de organización del hogar entre semana:** Tiempo promedio diario, en horas, dedicado a la organización del hogar entre semana.
- **Horas de organización del hogar fin de semana:** Tiempo promedio diario, en horas, dedicado a la organización del hogar durante el fin de semana.
- **Horas de mantenimiento del hogar por terceros entre semana:** Tiempo promedio diario, en horas, que terceros dedican al mantenimiento del hogar de la persona entre semana.
- **Horas de mantenimiento del hogar por terceros fin de semana:** Tiempo promedio diario, en horas, que terceros dedican al mantenimiento del hogar de la persona durante el fin de semana.
- **Horas de cuidados especiales entre semana:** Tiempo promedio diario, en horas, dedicado al cuidado de personas con necesidades especiales entre semana.
- **Horas de cuidados especiales fin de semana:** Tiempo promedio diario, en horas, dedicado al cuidado de personas con necesidades especiales durante el fin de semana.

- **Horas de crianza de menores de 0 a 5 años entre semana:** Tiempo promedio diario, en horas, dedicado al cuidado de niños entre 0 y 5 años entre semana.
- **Horas de crianza de menores de 0 a 5 años fin de semana:** Tiempo promedio diario, en horas, dedicado al cuidado de niños entre 0 y 5 años durante el fin de semana.
- **Horas de crianza de menores de 0 a 14 años entre semana:** Tiempo promedio diario, en horas, dedicado al cuidado de niños y adolescentes entre 0 y 14 años entre semana.
- **Horas de crianza de menores de 0 a 14 años fin de semana:** Tiempo promedio diario, en horas, dedicado al cuidado de niños y adolescentes entre 0 y 14 años durante el fin de semana.
- **Horas de acompañamiento familiar (15 a 59 años) entre semana:** Tiempo promedio diario, en horas, dedicado a compartir con familiares entre 15 y 59 años entre semana.
- **Horas de acompañamiento familiar (15 a 59 años) fin de semana:** Tiempo promedio diario, en horas, dedicado a compartir con familiares entre 15 y 59 años durante el fin de semana.
- **Horas de acompañamiento familiar (60 años o más) entre semana:** Tiempo promedio diario, en horas,

dedicado a compartir con familiares de 60 años o más entre semana.

- **Horas de acompañamiento familiar (60 años o más) fin de semana:** Tiempo promedio diario, en horas, dedicado a compartir con familiares de 60 años o más durante el fin de semana.
- **Horas de apoyo a otro hogar entre semana:** Tiempo promedio diario, en horas, dedicado a brindar apoyo a otro hogar entre semana.
- **Horas de apoyo a otro hogar fin de semana:** Tiempo promedio diario, en horas, dedicado a brindar apoyo a otro hogar durante el fin de semana.
- **Horas de voluntariado entre semana:** Tiempo promedio diario, en horas, dedicado a realizar actividades de voluntariado entre semana.
- **Horas de voluntariado fin de semana:** Tiempo promedio diario, en horas, dedicado a realizar actividades de voluntariado durante el fin de semana.
- **Horas de actividades recreativas entre semana:** Tiempo promedio diario, en horas, dedicado al ocio y la recreación entre semana.
- **Horas de actividades recreativas fin de semana:** Tiempo promedio diario, en horas, dedicado al ocio y la

recreación durante el fin de semana.

- **Horas de atención a la salud entre semana:** Tiempo promedio diario, en horas, dedicado al cuidado de la salud física y mental entre semana.
- **Horas de atención a la salud fin de semana:** Tiempo promedio diario, en horas, dedicado al cuidado de la salud física y mental durante el fin de semana.

#### **Sentimientos:**

- **Sentimientos sobre los quehaceres del hogar:** Percepción subjetiva de la persona sobre el tiempo dedicado a las tareas del hogar.
- **Sentimientos sobre el estudio:** Percepción subjetiva de la persona sobre el tiempo dedicado al estudio.
- **Sentimientos sobre el trabajo:** Percepción subjetiva de la persona sobre el tiempo dedicado al trabajo.
- **Sentimientos sobre el cuidado y apoyo a familiares:** Percepción subjetiva de la persona sobre el tiempo dedicado al cuidado y apoyo de familiares.
- **Sentimientos sobre el ocio y la convivencia:** Percepción subjetiva de la persona sobre el tiempo dedicado al ocio y la convivencia.

- **Sentimientos sobre el traslado:** Percepción subjetiva de la persona sobre el tiempo dedicado al traslado.
- **Sentimientos sobre los trámites y pagos:** Percepción subjetiva de la persona sobre el tiempo dedicado a trámites y pagos.
- **Satisfacción con la vida en general:** Grado de satisfacción general de la persona con su vida.
- **Satisfacción con la vida familiar:** Grado de satisfacción de la persona con su vida familiar.
- **Satisfacción con la vida afectiva:** Grado de satisfacción de la persona con su vida afectiva y relaciones.
- **Satisfacción con la vida social:** Grado de satisfacción de la persona con sus relaciones sociales.
- **Satisfacción con la situación económica:** Grado de satisfacción de la persona con su situación económica.
- **Satisfacción con la vivienda:** Grado de satisfacción de la persona con su vivienda.
- **Nivel de felicidad general:** Nivel de felicidad general experimentado por la persona.

### *Herramientas disponibles y conocimiento relevante*

Lenguaje de programación:

- Python.

Herramientas para data mining:

- Orange.

Módulos de Python enfocados a machine learning, data mining y estadística:

- Scikit-learn
- Scipy
- Pandas
- Orange

Algoritmos de agrupamiento y clasificación.

- Naive Bayes
- K-Means

## **ii. Requisitos, supuestos y restricciones**

- **Requisitos:**

*Obtención de datos:*

1. Descargar el conjunto de datos completo de la "Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019" proporcionado por el INEGI.
2. Asegurar que el dataset incluye variables relacionadas con:
  - a. Sexo (hombres y mujeres).
  - b. Actividades remuneradas (trabajo, negocios, etc.).
  - c. Actividades no remuneradas (cuidado del hogar, hijos, familiares, etc.).
  - d. Tiempo dedicado a cada actividad.

*Preprocesamiento de datos:*

- Limpiar datos incompletos, duplicados o con valores inconsistentes.

- Estandarizar formatos (fechas, valores categóricos y numéricos).
- Crear nuevas variables si es necesario, como "total de horas dedicadas por actividad."

*Análisis exploratorio:*

- Realizar estadísticas descriptivas para entender la distribución del tiempo entre sexo.
- Generar visualizaciones para comparar el tiempo invertido en actividades remuneradas y no remuneradas.

*Técnicas de minería de datos:*

- Utilizar el algoritmo de Naive Bayes para clasificar las actividades como remuneradas o no remuneradas en función de variables sociodemográficas como sexo y edad.
- Implementar el algoritmo K-means para realizar una clusterización que identifique patrones en el tiempo dedicado a actividades remuneradas y no remuneradas.
- Analizar si los clusters reflejan diferencias claras entre hombres y mujeres o si se agrupan por otras características como edad o tipo de actividad.
- Realizar clasificación para predecir el tipo de actividad en función de variables según el sexo.

*Generación de resultados:*

- Interpretar hallazgos relevantes sobre diferencias entre hombres y mujeres.
- Visualizar los resultados en gráficos claros.
- Preparar un informe final detallando hallazgos y conclusiones.

- **Supuestos**

Se asume que los datos proporcionados por INEGI son precisos, consistentes y representativos.

Se cuenta con acceso a las herramientas necesarias (Python y Orange)

La clasificación de actividades en "remuneradas" y "no remuneradas" es clara y válida.

El modelo analítico asumirá que las categorías de actividades y las variables sociodemográficas (edad, sexo, etc.) son independientes.

- **Restricciones**

- La principal limitación identificada radica en las variables disponibles en el dataset, dado que la problemática abarca múltiples aspectos de la población, sería ideal contar con variables adicionales que proporcionen más detalles de los grupos estudiados como su nivel socioeconómico.
- En cuanto al desarrollo del modelo, se presenta la limitante de que este es el primer acercamiento a un problema relacionado con minería de datos, lo que implica que carecemos de experiencia previa en el tema.
- En la decisión de los modelos a utilizar, nos vemos limitados en el hecho de que no disponemos un amplio conocimiento sobre todas las técnicas de minería de datos.
- La realización de este proyecto está restringida por el tiempo disponible, ya que se tiene previsto completarlo en un plazo de mes y medio. Sin embargo, el tiempo efectivo

de desarrollo es aún más limitado debido a la cercanía del cierre del semestre.

### iii. Riesgos y contingencias

- El muestreo seleccionado no sea lo suficientemente representativo al dataset completo. Para ello podemos generar múltiples muestras y evaluar si los resultados obtenidos generalizan bien al dataset completo.
- El equipo se enfoca en detalles secundarios o análisis excesivamente profundos sin haber abordado la problemática principal. Para ello debemos definir claramente el objetivo principal desde el inicio y establecer prioridades.
- Debido a la falta de experiencia, es posible derivar a errores durante el desarrollo de los algoritmos o la interpretación de los resultados. En función a evitar este problema, podemos practicar la implementación de los algoritmos en datasets más sencillos para familiarizarnos con su funcionamiento antes de aplicarlos al proyecto.

### iv. Terminología

- **Actividades de cuidado:** En el contexto de nuestro proyecto, se refiere a la identificación y análisis de actividades destinadas al cuidado de las personas, como servicios médicos o familiares.
- **Algoritmo:** Procedimientos o métodos que se usan para identificar patrones y tendencias en un conjunto de datos.
- **Bienestar:** Es un estado en el que se contempla subjetivamente una vida positiva, que promueve mejoras en la persona y se pretende en la mayoría de los individuos llegar a ese estado.
- **Calidad de Vida:** Se refiere al bienestar general de las personas, incluyendo aspectos físicos, mentales y sociales.

- **CLI/ Interfaz de línea de comandos:** Es una herramienta de bajo impacto que puede utilizarse de forma independiente o junto con la consola para llevar a cabo tareas en Oracle Cloud Infrastructure. La CLI ofrece las mismas funcionalidades principales que la consola, además de incluir comandos adicionales. Algunos de estos comandos, como la posibilidad de ejecutar scripts, añaden capacidades que amplían las funciones de la consola.
- **Conjunto:** colección de datos o atributos utilizados para análisis.
- **Correlación:** Es una medida estadística que indica el grado en que dos variables están relacionadas. En minería de datos, la correlación ayuda a identificar relaciones y patrones entre variables.
- **Data mining / Minería de datos:** La minería de datos es un método para identificar información útil dentro de grandes volúmenes de datos. Emplea técnicas matemáticas para descubrir patrones y tendencias presentes en los datos que, debido a su complejidad o a la gran cantidad de información, no pueden ser detectados mediante métodos tradicionales de análisis.
- **ENUT :** Encuesta Nacional sobre Uso del Tiempo realizada por el Instituto Nacional de Estadística y Geografía.
- **Felicidad:** es un estado general de satisfacción con la vida que perdura a lo largo de períodos prolongados. Se considera una de las principales maneras en que las personas evalúan qué tan bien están viviendo; es decir, cuán contentas se sienten consigo mismas y su entorno.
- **Fusión:** Para nuestro proyecto, fusión significa unir 2 o más atributos de un dataset en un solo atributo.

- **Índice Silhouette/Score Silhouette:** Métrica usada para evaluar la calidad de un clustering. Mide qué tan bien se agrupan los puntos en su cluster en comparación con clusters vecinos.
- **Kaggle:** es una comunidad en línea de científicos de datos e ingenieros de aprendizaje automático. Kaggle permite a los usuarios encontrar conjuntos de datos que desean utilizar para crear modelos de IA, publicar conjuntos de datos, trabajar con otros científicos de datos e ingenieros de aprendizaje automático y participar en competencias para resolver desafíos de la ciencia de datos.
- **K-Means:** Algoritmo de agrupamiento para identificar patrones ocultos y formar grupos de personas con características o comportamientos similares sin categorizar de antemano.
- **Machine learning / Aprendizaje automático :** Es un campo de la inteligencia artificial (IA) y la informática que se enfoca en utilizar datos y algoritmos para que la IA reproduzca el proceso de aprendizaje humano, perfeccionando su exactitud de manera progresiva.
- **Modelo:** en minería de datos se genera al aplicar un algoritmo a un conjunto de datos. Sin embargo, no se limita a ser solo un algoritmo o un conjunto de metadatos; se trata de una combinación de datos, estadísticas y patrones que pueden utilizarse para analizar nuevos datos, realizar predicciones y descubrir relaciones.
- **Muestra:** Subconjunto representativo de los datos originales que se usa para analizar o entrenar modelos.
- **Naive Bayes:** Algoritmo de clasificación para catalogar o predecir cómo diferentes factores (como sexo) influyen en la felicidad general de las personas.
- **Numpy:** Es una biblioteca de Python que proporciona un objeto de matriz multidimensional, varios objetos derivados (como

matrices y matrices enmascaradas) y una variedad de rutinas para operaciones rápidas en matrices, incluidas operaciones matemáticas, lógicas, manipulación de formas, ordenamiento, selección, E/S, transformadas de Fourier discretas, álgebra lineal básica, operaciones estadísticas básicas, simulación aleatoria y mucho más.

- **Orange:** Herramienta de código abierto para minería de datos y aprendizaje automático que ofrece una interfaz gráfica intuitiva, facilitando el análisis y visualización de datos sin necesidad de programación.
- **Pandas:** Biblioteca de Python que permite manejar y analizar datos de manera fácil y eficiente. Con ella, es posible cargar datos desde archivos (como Excel, CSV, bases de datos, etc.), organizarlos, filtrarlos y analizarlos.
- **Python:** Es un lenguaje de programación versátil y sencillo de aprender. Ofrece estructuras de datos avanzadas y eficientes, complementadas por un sistema de programación orientado a objetos que es simple pero muy funcional. Su sintaxis clara y elegante, junto con el uso de tipado dinámico y su carácter interpretado, lo hacen perfecto para la creación de scripts y el desarrollo ágil de aplicaciones en una amplia variedad de áreas y plataformas.
- **Salud:** según la OMS, y para este proyecto “es un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades”.
- **Sexo:** Se refiere a los rasgos biológicos y fisiológicos que distinguen a hombres y mujeres. Según esta definición, la OMS reconoce "hombre" y "mujer" como categorías de sexo.
- **Tiempo por actividad:** número de horas empleadas en determinada actividad.

- **Variables categóricas:** variables que representan datos cualitativos que describen categorías o grupos.
- **Variables numéricas:** variables que representan datos cuantitativos que se expresan en números.

## v. Análisis de costo/beneficio

El tiempo estimado para la realización de este proyecto será alrededor de 1 mes y 2 semanas del actual año.

Como principal beneficio se encuentra a nivel académico aprender a trabajar con datos reales y aplicar técnicas de minería de datos como Naive Bayes y K-Means utilizando dataset reales. Así como fortalecer habilidades en el manejo de herramientas analíticas y de programación.

Otro beneficio que encontramos es el de la reflexión individual acerca de la carga desigual de actividades remuneradas y no remuneradas entre hombres y mujeres en México.

## c. Objetivos de la minería de datos

### i. Metas de la minería de datos

- Aplicar el algoritmo de K-Means para identificar patrones ocultos y formar grupos de personas con características similares sin categorizar de manera premeditada.
- Utilizar el algoritmo Naive Bayes para evaluar el desempeño general en la clasificación de la felicidad general, centrándonos en métricas globales como la precisión y la estabilidad del modelo, sin analizar la influencia individual de cada atributo.

### ii. Criterio de rendimiento

- Para evaluar el rendimiento de los modelos se utilizarán los siguientes criterios:

- Para K-Means: **Índice Silhouette**, medir que tan bien se separan los clústeres entre sí, con valores cercanos a 1 indicando una buena separación y **visualización de los clústeres**, interpretando los resultados gráficamente para validar si los grupos formados tienen sentido y coherencia.
- Para Naive Bayes: Con base a la **precisión obtenida** del modelo y haciendo uso de un **conjunto de prueba** se evalúa mediante probabilidad posterior el porcentaje de predicción del algoritmo.

#### **d. Plan de proyecto**

##### **i. Plan de proyecto**

- **Fase 1: Recolección y Familiarización con los Datos (Semana 1)**
  - Descargar el conjunto de datos de la ENUT 2019 desde la fuente oficial (INEGI).
  - Revisar la documentación disponible sobre la encuesta (diccionario de datos, metodología, etc.).
  - Explorar los datos iniciales para entender su estructura, variables y posibles desafíos.
- **Fase 2: Limpieza y Preprocesamiento de Datos (Semana 2)**
  - Identificar y manejar valores faltantes.
  - Estandarizar formatos de variables.
  - Crear nuevas variables relevantes para el análisis (por ejemplo, total de horas dedicadas a actividades remuneradas y no remuneradas).
  - Detectar y tratar valores atípicos (outliers) que puedan afectar el análisis.
- **Fase 3: Análisis Exploratorio de Datos (Semana 3)**
  - Realizar estadísticas descriptivas para comprender la

distribución de las variables clave.

- Generar visualizaciones iniciales para identificar tendencias y patrones en la distribución del tiempo entre hombres y mujeres.

- **Fase 4: Modelado y Evaluación (Semana 4)**

- Naive Bayes:
  - i. Implementar el algoritmo para clasificar actividades y analizar la precisión general en la clasificación de la clase "Felicidad General".
  - ii. Evaluar la estabilidad de las métricas obtenidas mediante el cálculo de la desviación estándar entre iteraciones.
  - iii. Interpretar los resultados para garantizar que el modelo no depende significativamente de particiones específicas de los datos.
- K-Means:
  - i. Aplicar el algoritmo para identificar grupos con patrones similares en el uso del tiempo.
  - ii. Evaluar la calidad de los clusters generados mediante métricas como el Silhouette Score.
  - iii. Utilizar herramientas de visualización (gráficos de dispersión, mapas de calor, etc.) para interpretar y comunicar los resultados.

- **Fase 5: Interpretación y Conclusiones (Semana 5)**

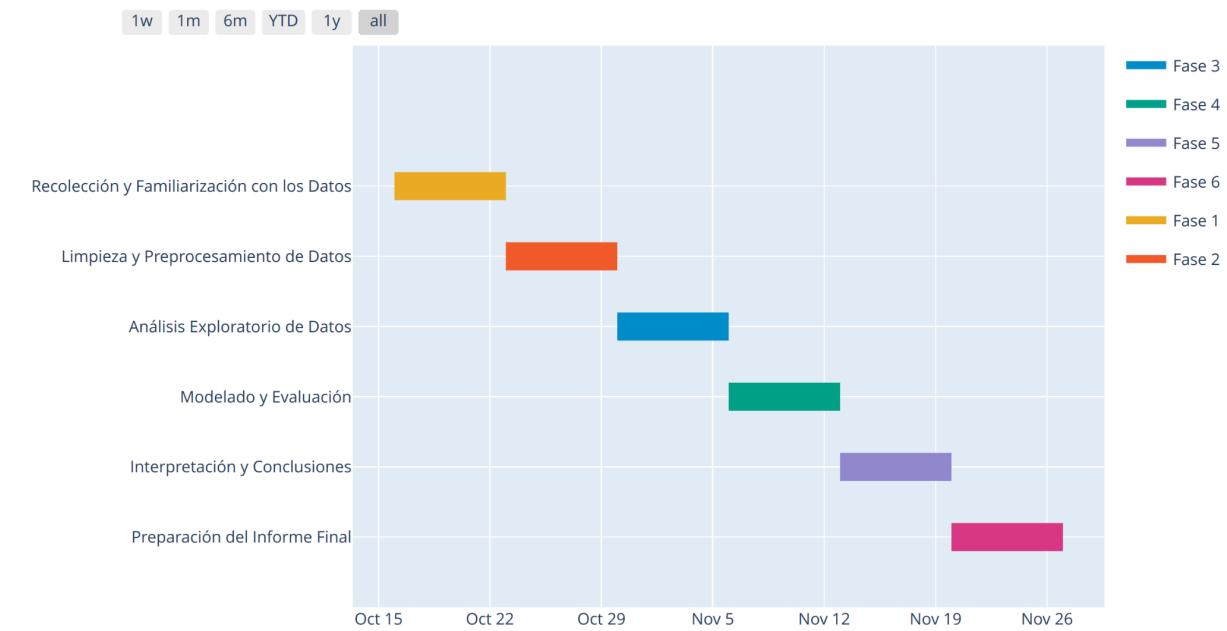
- Analizar los resultados de los modelos en términos del negocio y los objetivos de investigación.
- Extraer conclusiones sobre las diferencias de género en el uso del tiempo.
- Formular recomendaciones o propuestas basadas en los hallazgos.

- **Fase 6: Preparación del Informe Final (Semana 6)**

- Documentar todo el proceso y los resultados en un informe claro y conciso.
- Preparar visualizaciones efectivas para comunicar los hallazgos clave.

## Diagrama de gantt

Plan de Proyecto: Análisis del Uso del Tiempo ENUT 2019



## ii. Valoración de herramientas técnica

### Python

Ventajas:

1. Versatilidad y facilidad de uso:
  - Es un lenguaje fácil de aprender y usar.
  - Permite trabajar con diferentes bibliotecas diseñadas para análisis y modelado de datos.
2. Integración:
  - Compatible con muchas herramientas de análisis de datos y machine learning.
3. Automatización y repetibilidad:

- Facilita la creación de scripts reutilizables para el análisis de datos.

Desventajas:

1. Velocidad de ejecución:
  - Es más lento en comparación con lenguajes como C++ o Java para tareas muy complejas.
2. Gestión de memoria:
  - Puede consumir más memoria al trabajar con datasets grandes.
3. Curva de aprendizaje inicial:
  - Aunque es fácil de aprender, requiere tiempo para dominar sus bibliotecas avanzadas.

## **Pandas**

Ventajas:

1. Eficiencia:
  - Facilita la manipulación y análisis de datos estructurados con métodos intuitivos.
2. Capacidades avanzadas:
  - Ofrece funciones potentes para limpieza de datos, filtrado y agrupación.
3. Flexibilidad:
  - Compatible con otros frameworks de Python.
4. Documentación extensa:
  - Amplios recursos y tutoriales disponibles.

Desventajas:

1. Limitaciones con datos muy grandes:
  - No está optimizado para datasets que no caben en

memoria.

## 2. Dependencia de sintaxis:

- Puede ser confuso para usuarios nuevos debido a su sintaxis compleja en algunos casos.

## NumPy

Ventajas:

- Eficiencia:
  - Los arrays de NumPy permiten un almacenamiento y procesamiento de datos numéricos mucho más eficiente que las listas de Python.
  - Las operaciones matemáticas se optimizan gracias a la homogeneidad de los arrays (elementos del mismo tipo).
- Funcionalidad:
  - Ofrece una amplia gama de funciones matemáticas para álgebra lineal, transformadas de Fourier, estadísticas y más.
  - Broadcasting facilita operaciones entre arrays de diferentes formas.
- Base para otras herramientas:
  - NumPy es la base de muchas librerías científicas en Python como SciPy, Pandas y Scikit-learn.
- Rendimiento:
  - Muchas de sus funciones están implementadas en C, lo que resulta en un rendimiento superior.

Desventajas:

- Curva de aprendizaje:
  - Aunque intuitivo en general, dominar conceptos

como el broadcasting o la indexación avanzada puede requerir tiempo.

- Limitaciones con datos no numéricos:
  - NumPy está optimizado para datos numéricos. Si necesitas trabajar con datos de texto o estructuras más complejas, Pandas puede ser una mejor opción.
- Manejo de datos faltantes:
  - NumPy tiene un soporte limitado para datos faltantes en comparación con Pandas.

## **Orange**

Ventajas:

1. Interfaz amigable:
  - No requiere programación, lo que facilita el uso para usuarios no técnicos.
2. Visualización intuitiva:
  - Ofrece herramientas gráficas avanzadas para analizar y presentar resultados.
3. Integración con Python:
  - Permite utilizar scripts en Python dentro del flujo de trabajo para personalización.
4. Ideal para análisis rápido:
  - Útil para explorar datos y probar rápidamente diferentes enfoques.

Desventajas:

1. Capacidades limitadas:
  - Menos flexible que el uso directo de bibliotecas de Python.

2. Rendimiento:
  - Puede ser más lento con datasets grandes o complejos.
3. Dependencia de componentes:
  - Algunos widgets requieren configuraciones adicionales o conocimientos específicos.

## **2. Entendimiento de los datos**

### **a. Recolección de datos iniciales**

#### **i. Requerimientos de los datos**

Los datos provienen del portal del Instituto Nacional de Estadística y Geografía (INEGI), específicamente en el apartado de encuestas regulares, este conjunto de datos fue recopilado en el año 2019 y publicado en diciembre del mismo año, tiene como objetivo proporcionar datos estadísticos de las formas de trabajo de los individuos, tanto remunerado como no remunerado, así como, identificar la importancia de la producción doméstica y su contribución a la economía, como usan su tiempo los hombres y las mujeres y como consideran su bienestar. Abarca población de 12 años en adelante de áreas urbanas, rurales e indígenas.

#### **1. Orígenes de datos y su disponibilidad**

Los datos se obtienen de un comprimido con múltiples carpetas con los diccionarios, conjunto de datos, y catálogos cada uno de estos disponible como archivo plano (.csv) delimitado por comas.

#### **ii. Criterios de selección**

##### **1. Criterios de selección de los datos**

Para este análisis, se utilizarán las siguientes variables extraídas, fusionadas y renombradas de la ENUT 2019

- **Variables demográficas:**
  - **Entidad:** Permite analizar diferencias regionales en la distribución del tiempo y el nivel de felicidad.

- **Sexo:** Fundamental para examinar las disparidades de género en la asignación de tiempo a distintas actividades y su relación con el bienestar.
- **Edad:** Permite analizar cómo la distribución del tiempo y la percepción de felicidad varían a lo largo del ciclo de vida.
- **Variables de actividad:**
  - **TrabajoAlMenos1Hora,** **TiempoTrabajoLV,** **TiempoTrabajoSD:** Miden la participación en actividades remuneradas y el tiempo dedicado a ellas, lo que es crucial para entender la carga laboral.
  - **TiempoTrasladoLV,** **TiempoTrasladoSD:** Consideran el tiempo dedicado a desplazamientos, un factor que puede afectar el bienestar y la disponibilidad de tiempo para otras actividades.
  - **Salario:** Permite analizar la relación entre ingresos y bienestar, así como explorar posibles desigualdades en la remuneración.
  - **TiempoDedicadoADormirLV,** **TiempoDedicadoADormirSD,** **TiempoDedicadoAComerLV,** **TiempoDedicadoAComerSD,** **TiempoDedicadoAAseoPersonalLV,** **TiempoDedicadoAAseoPersonalSD:** Incluyen actividades esenciales para el bienestar físico y mental.
  - **TiempoDedicadoAActividadesEstudioLV,** **TiempoDedicadoAActividadesEstudioSD,** **TiempoDedicadoATrasladarEstudianteLV,**

**TiempoDedicadoATrasladarEstudianteSD:**

Capturan la participación en actividades educativas y el tiempo invertido en ellas, lo cual es relevante para analizar el desarrollo personal y las oportunidades futuras.

● **Variables de trabajo no remunerado:**

- QuehaceresDomésticosLV,
- QuehaceresDomésticosSD,
- TiempoGanaderíaAutoconsumoLV,
- TiempoGanaderíaAutoconsumoSD,
- TiempoActividadesRecolecciónLV,
- TiempoActividadesRecolecciónSD,
- TiempoAgriculturaAutoconsumoLV,
- TiempoAgriculturaAutoconsumoSD,
- TiempoConfecciónCaseraLV,
- TiempoConfecciónCaseraSD,
- TiempoMantenimientoHogarLV,
- TiempoMantenimientoHogarSD,
- TiempoComprasDelHogarLV,
- TiempoComprasDelHogarSD,
- TiempoTrámitesPagosLV,
- TiempoTrámitesPagosSD,
- TiempoOrganizaciónHogarLV,
- TiempoOrganizaciónHogarSD,
- TiempoMantenimientoHogarTercerosLV,
- TiempoMantenimientoHogarTercerosSD,
- TiempoCuidadosEspecialesOtraPersonaLV,
- TiempoCuidadosEspecialesOtraPersonaSD,
- TiempoCrianzaMenores0a5LV,
- TiempoCrianzaMenores0a5SD,
- TiempoCrianzaMenores0a14LV,

**TiempoCrianzaMenores0a14SD,**  
**TiempoAcompañamientoFamiliar15a59LV,**  
**TiempoAcompañamientoFamiliar15a59SD,**  
**TiempoAcompañamientoFamiliar60LV,**  
**TiempoAcompañamientoFamiliar60SD,**  
**TiempoApoyoOtroHogarLV,**  
**TiempoApoyoOtroHogarSD,**  
**TiempoVoluntariadoLV, TiempoVoluntariadoSD:**  
Estas variables permiten analizar la distribución del trabajo no remunerado, que a menudo recae desproporcionadamente sobre las mujeres.

- **Variables de bienestar:**

- **SentimientoQuehaceresHogar,**  
**SentimientoEstudio, SentimientoTrabajo,**  
**SentimientoCuidadoApoyoFamiliares,**  
**SentimientoOcioConvivencia,**  
**SentimientoTraslado,**  
**SentimientoTramitesPagos,**  
**SentimientoVidaGeneral,**  
**SentimientoVidaFamiliar,**  
**SentimientoVidaAfactiva,**  
**SentimientoVidaSocial,**  
**SentimientoSituaciónEconómica,**  
**SentimientoVivienda,**  
**SentimientoFelicidadGeneral21:** Estas variables subjetivas proporcionan información sobre la percepción individual de bienestar en diferentes ámbitos de la vida.

**Justificación:**

La selección de estas variables se basa en la necesidad de comprender cómo la distribución del tiempo entre actividades remuneradas, no remuneradas y de cuidado se relaciona con el bienestar subjetivo de la población mexicana, considerando las diferencias de género, edad y ubicación geográfica. El análisis de estas variables permitirá identificar patrones de desigualdad y explorar su impacto en la felicidad y calidad de vida de las personas.

**2. En el caso de BD, listar las tablas de interés y sus diccionarios de datos**

**iii. Inserción de datos**

**1. Si fueron diferentes orígenes de datos, indicar los mecanismos de adquisición.**

No aplica, ya que todos los datos fueron adquiridos del mismo conjunto de datos proporcionados por la Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019 del Instituto Nacional de Estadística y Geografía (INEGI).

**b. Descripción de los datos**

**i. Análisis volumétrico**

**1. Criterios de Exclusión**

PCA

**1. Número de instancias:** 1000

**2. Número de atributos:** 75

**3. Número de instancias por clase:**

**ii. Definición del dominio del atributo (nombre, tipos de datos y formato)**

Variable	Tipo	de	Descripción	Rango/Categorías
----------	------	----	-------------	------------------

	<b>Variable</b>		
Entidad	Numérica (entera)	Código numérico que representa una entidad federativa.	1: Aguascalientes 2: Baja California 3: Baja California Sur 4: Campeche 5: Coahuila de Zaragoza 6: Colima 7: Chiapas 8: Chihuahua 9: Distrito Federal 10: Durango 11: Guanajuato 12: Guerrero 13: Hidalgo 14: Jalisco 15: México 16: Michoacán de Ocampo 17: Morelos 18: Nayarit 19: Nuevo León 20: Oaxaca 21: Puebla 22: Querétaro 23: Quintana Roo 24: San Luis Potosí 25: Sinaloa 26: Sonora 27: Tabasco 28: Tamaulipas 29: Tlaxcala 30: Veracruz de Ignacio de la Llave 31: Yucatán 32: Zacatecas

Sexo	Categórica	Sexo de la persona.	['Mujer' 'Hombre']
Edad	Numérica (entera)	Edad de la persona en años.	12 - 95
TrabajoAlMenos1Hora	Categórica	Indica si la persona trabajó al menos una hora.	['No' 'Si']
TiempoTrabajoLV	Numérica (continua)	Tiempo dedicado al trabajo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 96.0
TiempoTrabajoSD	Numérica (continua)	Tiempo dedicado al trabajo en un día de fin de semana promedio en sábado y domingo (en horas).	0.0 - 48.0
TiempoTrasladoLV	Numérica (continua)	Tiempo dedicado al traslado al trabajo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 40.0
TiempoTrasladoSD	Numérica (continua)	Tiempo dedicado al traslado al trabajo en un día de fin de semana promedio (en horas).	0.0 - 40.0
Salario	Numérica (continua)	Salario mensual de la persona.	0.0 - 780000.0
PermisoEnfermedadMaternidadAccidente	Categórica	Indica si la persona tiene derecho a permiso por enfermedad, maternidad o accidente.	['No especificado' 'Si' 'No']
PermisoVacaciones	Categórica	Indica si la persona tiene derecho a vacaciones.	['No especificado' 'Si' 'No']

DerechoJubilación	Categórica	Indica si la persona tiene derecho a jubilación.	['No especificado' 'Si' 'No']
DerechoServiciosMedicos	Categórica	Indica si la persona tiene derecho a servicios médicos.	['No especificado' 'Si' 'No']
TiempoDedicadoADormirLV	Numérica (continua)	Tiempo dedicado a dormir en un día laboral promedio de lunes a viernes (en horas).	2.0 - 76.0
TiempoDedicadoADormirSD	Numérica (continua)	Tiempo dedicado a dormir en un día de fin de semana promedio (en horas).	0.0 - 35.0
TiempoDedicadoAComerLV	Numérica (continua)	Tiempo dedicado a comer en un día laboral promedio de lunes a viernes (en horas).	0.0 - 77.5
TiempoDedicadoAComerSD	Numérica (continua)	Tiempo dedicado a comer en un día de fin de semana promedio de sábado y domingo(en horas).	0.0 - 45.0
TiempoDedicadoAAseoPersonalLV	Numérica (continua)	Tiempo dedicado al aseo personal en un día laboral promedio de lunes a viernes(en horas).	0.0833333333333333 - 21.0
TiempoDedicadoAAseoPersonalSD	Numérica (continua)	Tiempo dedicado al aseo personal en un día de fin de semana promedio en sábado y domingo (en horas).	0.0 - 20.0

TiempoDedicadoAAActividadesEstudioLV	Numérica (continua)	Tiempo dedicado a actividades de estudio en un día laboral promedio de lunes a viernes (en horas).	0.0 - 80.0
TiempoDedicadoAAActividadesEstudioSD	Numérica (continua)	Tiempo dedicado a actividades de estudio en un día de fin de semana promedio (en horas).	0.0 - 20.0
TiempoDedicadoATrasladarEstudianteLV	Numérica (continua)	Tiempo dedicado a trasladar a un estudiante en un día laboral promedio de lunes a viernes (en horas).	0.0 - 15.0
TiempoDedicadoATrasladarEstudianteSD	Numérica (continua)	Tiempo dedicado a trasladar a un estudiante en un día de fin de semana promedio (en horas).	0.0 - 3.0
QuehaceresDomésticosLV	Numérica (continua)	Tiempo dedicado a quehaceres domésticos en un día laboral promedio de lunes a viernes (en horas).	0.0 - 84.5833333333334
QuehaceresDomésticosSD	Numérica (continua)	Tiempo dedicado a quehaceres domésticos en un día de fin de semana promedio (en horas).	0.0 - 46.35

TiempoGanaderíaAutoconsumoLV	Numérica (continua)	Tiempo dedicado a la ganadería para autoconsumo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 35.0
TiempoGanaderíaAutoconsumoSD	Numérica (continua)	Tiempo dedicado a la ganadería para autoconsumo en un día de fin de semana promedio (en horas).	0.0 - 18.0
TiempoActividadesRecolecciónLV	Numérica (continua)	Tiempo dedicado a actividades de recolección en un día laboral promedio de lunes a viernes (en horas).	0.0 - 20.0
TiempoActividadesRecolecciónSD	Numérica (continua)	Tiempo dedicado a actividades de recolección en un día de fin de semana promedio (en horas).	0.0 - 8.0
TiempoAgriculturaAutoconsumoLV	Numérica (continua)	Tiempo dedicado a la agricultura para autoconsumo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 18.0
TiempoAgriculturaAutoconsumoSD	Numérica (continua)	Tiempo dedicado a la agricultura para autoconsumo en un día de fin de semana promedio (en horas).	0.0 - 5.0

TiempoConfecciónCaseraLV	Numérica (continua)	Tiempo dedicado a la confección casera en un día laboral promedio de lunes a viernes (en horas).	0.0 - 25.0
TiempoConfecciónCaseraSD	Numérica (continua)	Tiempo dedicado a la confección casera en un día de fin de semana promedio (en horas).	0.0 - 8.0
TiempoMantenimientoHogarLV	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar en un día laboral promedio de lunes a viernes (en horas).	0.0 - 28.0
TiempoMantenimientoHogarSD	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar en un día de fin de semana promedio (en horas).	0.0 - 15.0
TiempoComprasDelHogarLV	Numérica (continua)	Tiempo dedicado a las compras del hogar en un día laboral promedio de lunes a viernes (en horas).	0.0 - 13.0
TiempoComprasDelHogarSD	Numérica (continua)	Tiempo dedicado a las compras del hogar en un día de fin de semana promedio (en horas).	0.0 - 10.0
TiempoTramitesPagosLV	Numérica (continua)	Tiempo dedicado a trámites y pagos en un día laboral promedio de lunes a viernes (en horas).	0.0 - 6.0

TiempoTramite sPagosSD	Numérica (continua)	Tiempo dedicado a trámites y pagos en un día de fin de semana promedio (en horas).	0.0 - 2.0
TiempoOrganiz aciónHogarLV	Numérica (continua)	Tiempo dedicado a la organización del hogar en un día laboral promedio de lunes a viernes (en horas).	0.0 - 7.0
TiempoOrganiz aciónHogarSD	Numérica (continua)	Tiempo dedicado a la organización del hogar en un día de fin de semana promedio (en horas).	0.0 - 2.0
TiempoManten imientoHogarTe rcerosLV	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar por terceros en un día laboral promedio de lunes a viernes (en horas).	0.0 - 50.0
TiempoManten imientoHogarTe rcerosSD	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar por terceros en un día de fin de semana promedio (en horas).	0.0 - 30.0
TiempoCuidado sEspecialesOtr aPersonaLV	Numérica (continua)	Tiempo dedicado a cuidados especiales de otra persona en un día laboral promedio de lunes a viernes (en horas).	0.0 - 81.83333333333333

TiempoCuidadosEspecialesOtraPersonaSD	Numérica (continua)	Tiempo dedicado a cuidados especiales de otra persona en un día de fin de semana promedio (en horas).	0.0 - 46.0
TiempoCrianzaMenores0a5LV	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 5 años en un día laboral promedio (en horas).	0.0 - 85.0
TiempoCrianzaMenores0a5SD	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 5 años en un día de fin de semana promedio (en horas).	0.0 - 33.0
TiempoCrianzaMenores0a14LV	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 14 años en un día laboral promedio (en horas).	0.0 - 103.0
TiempoCrianzaMenores0a14SD	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 14 años en un día de fin de semana promedio (en horas).	0.0 - 39.0
TiempoAcompañamientoFamiliar15a59LV	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 15 a 59 años en un día laboral promedio (en horas).	0.0 - 60.0

TiempoAcompañamientoFamiliar15a59SD	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 15 a 59 años en un día de fin de semana promedio (en horas).	0.0 - 24.0
TiempoAcompañamientoFamiliar60LV	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 60 años o más en un día laboral promedio (en horas).	0.0 - 62.0
TiempoAcompañamientoFamiliar60SD	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 60 años o más en un día de fin de semana promedio (en horas).	0.0 - 48.0
TiempoApoyoOtroHogarLV	Numérica (continua)	Tiempo dedicado a apoyar a otro hogar en un día laboral promedio de lunes a viernes (en horas).	0.0 - 40.0
TiempoApoyoOtroHogarSD	Numérica (continua)	Tiempo dedicado a apoyar a otro hogar en un día de fin de semana promedio (en horas).	0.0 - 32.0
TiempoVoluntariadoLV	Numérica (continua)	Tiempo dedicado al voluntariado en un día laboral promedio de lunes a viernes (en horas).	0.0 - 10.0

TiempoVoluntariadoSD	Numérica (continua)	Tiempo dedicado al voluntariado en un día de fin de semana promedio (en horas).	0.0 - 26.0
TiempoActividadesRecreativasLV	Numérica (continua)	Tiempo dedicado a actividades recreativas en un día laboral promedio de lunes a viernes (en horas).	0.0 - 116.0
TiempoActividadesRecreativasSD	Numérica (continua)	Tiempo dedicado a actividades recreativas en un día de fin de semana promedio (en horas).	0.0 - 71.5
TiempoAtenciónSaludLV	Numérica (continua)	Tiempo dedicado a la atención de la salud en un día laboral promedio de lunes a viernes (en horas).	0.0 - 99.0
TiempoAtenciónSaludSD	Numérica (continua)	Tiempo dedicado a la atención de la salud en un día de fin de semana promedio (en horas).	0.0 - 48.0
SentimientoQuehaceresHogar	Categórica	Sentimiento general sobre los quehaceres del hogar.	[Quisiera dedicarle mas tiempo' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle menos tiempo' 'No aplica (no hizo la actividad)' 'No especificado']
SentimientoEstudio	Categórica	Sentimiento general sobre el estudio.	[Quisiera dedicarle mas tiempo' 'No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico'

			'Quisiera dedicarle menos tiempo']
SentimientoTrabajo	Categórica	Sentimiento general sobre el trabajo.	['No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle menos tiempo' 'Quisiera dedicarle mas tiempo']
SentimientoCuidadoApoyoFamiliares	Categórica	Sentimiento general sobre el cuidado y apoyo a familiares.	['Quisiera dedicarle mas tiempo' 'No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle menos tiempo' 'No especificado']
SentimientoOcioConvivencia	Categórica	Sentimiento general sobre el ocio y la convivencia.	['Esta bien el tiempo que le dedico' 'No aplica (no hizo la actividad)' 'Quisiera dedicarle mas tiempo' 'Quisiera dedicarle menos tiempo' 'No especificado']
SentimientoTraslado	Categórica	Sentimiento general sobre el traslado.	['Esta bien el tiempo que le dedico' 'No aplica (no hizo la actividad)' 'Quisiera dedicarle menos tiempo' 'Quisiera dedicarle mas tiempo']
SentimientoTramitesPagos	Categórica	Sentimiento general sobre los trámites y pagos.	['No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle mas tiempo' 'Quisiera dedicarle menos tiempo' 'No especificado']
SentimientoVidaGeneral	Categórica	Sentimiento general sobre la vida en general.	['Satisfecho(a)' 'Poco satisfecho(a)' 'Mas o menos satisfecho(a)' 'Muy']

			satisfecho(a)' 'Nada satisfecho(a)']
SentimientoVidaFamiliar	Categórica	Sentimiento general sobre la vida familiar.	['Muy satisfecho(a)' 'Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Poco satisfecho(a)' 'Nada satisfecho(a)']
SentimientoVidaAfectiva	Categórica	Sentimiento general sobre la vida afectiva.	['Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Poco satisfecho(a)' 'Muy satisfecho(a)' 'Nada satisfecho(a)']
SentimientoVidaSocial	Categórica	Sentimiento general sobre la vida social.	['Mas o menos satisfecho(a)' 'Satisfecho(a)' 'Muy satisfecho(a)' 'Nada satisfecho(a)' 'Poco satisfecho(a)']
SentimientoSituaciónEconómica	Categórica	Sentimiento general sobre la situación económica.	['Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Poco satisfecho(a)' 'Muy satisfecho(a)' 'Nada satisfecho(a)']
SentimientoVivienda	Categórica	Sentimiento general sobre la vivienda.	['Muy satisfecho(a)' 'Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Nada satisfecho(a)' 'Poco satisfecho(a)']
SentimientoFelicidadGeneral	Categórica	Sentimiento general de felicidad.	['Muy Feliz' 'Feliz' 'Poco feliz' 'Mas o menos feliz' 'Nada']

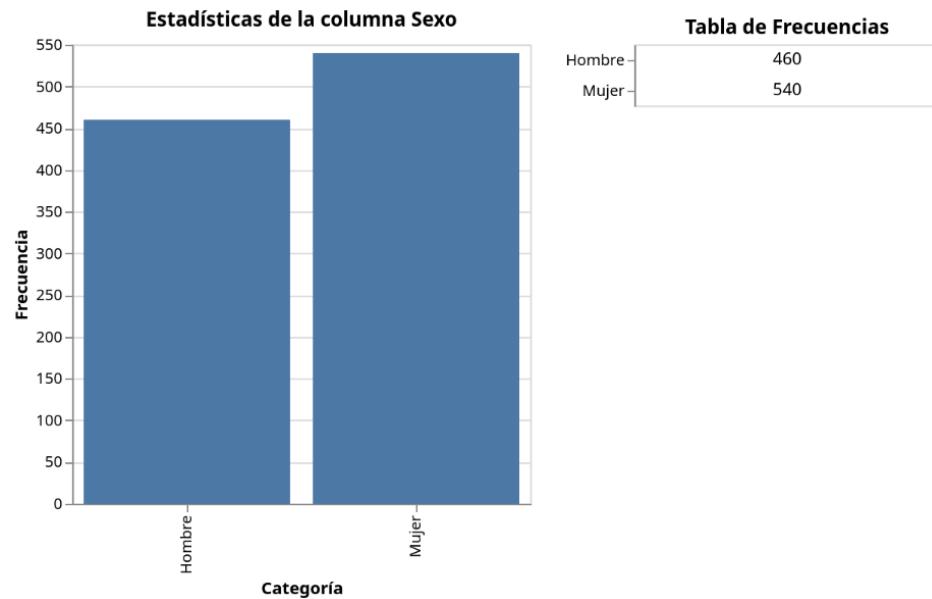
## 1. Diccionario de datos

### c. Exploración de los datos (Por atributo)

**i. Resumen estadístico, en el caso de los numéricos un Box plot con la descripción de los datos. En el caso de valores categóricos la moda, el histograma.**

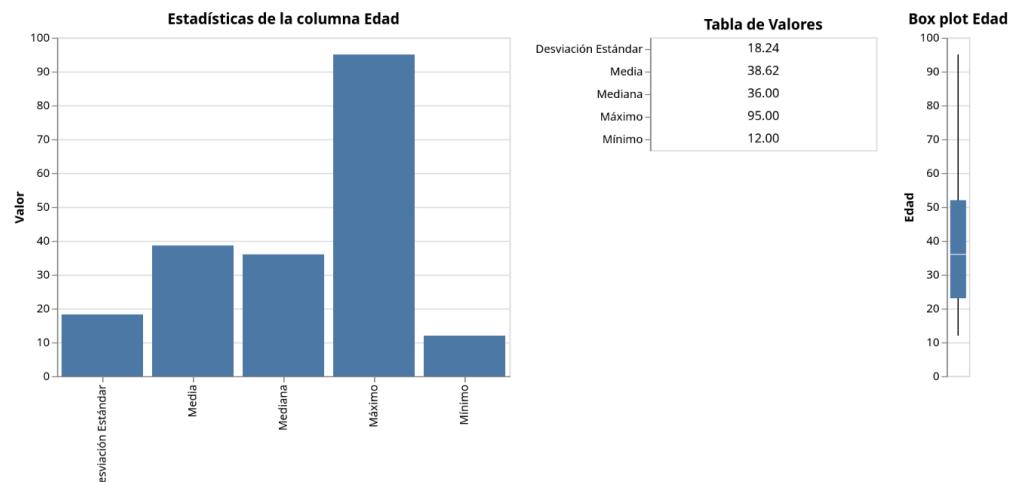
A continuación se muestran los Box Plot y gráficas de barras de nuestros datos:

### Sexo

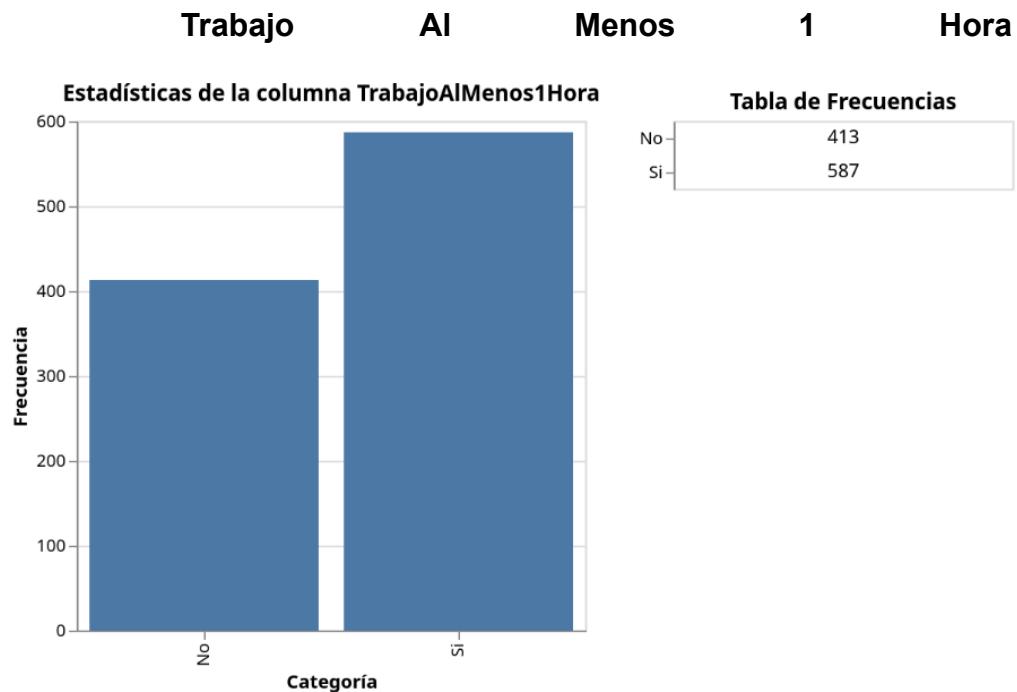


En la muestra que tomamos de nuestro dataset predominaron las mujeres, esto no implica que el dataset tiene una muestra mayor de mujeres, tiene que ver con la muestra que obtuvimos.

### Edad

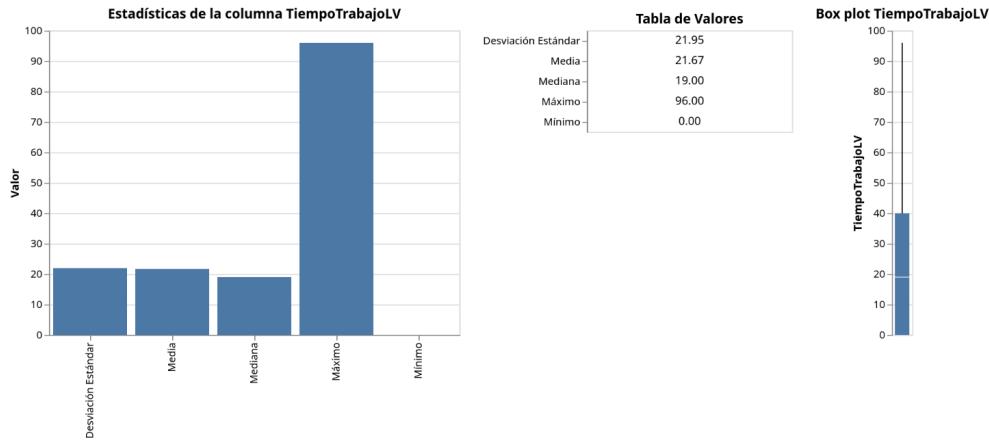


La distribución de la edad en nuestra muestra muestra una media de 38 años y una mediana de 36 años. Esto sugiere que la mayoría de los individuos en nuestro estudio se encuentran en la edad adulta, lo que es relevante para nuestro análisis ya que la percepción de felicidad y la distribución del tiempo pueden variar según la etapa de la vida. Los valores atípicos en esta gráfica resaltan individuos cuya edad es muy diferente del resto de la muestra, casos poco comunes pero que son propias de la comunidad y que son características válidas pero extremas de la población estudiada.



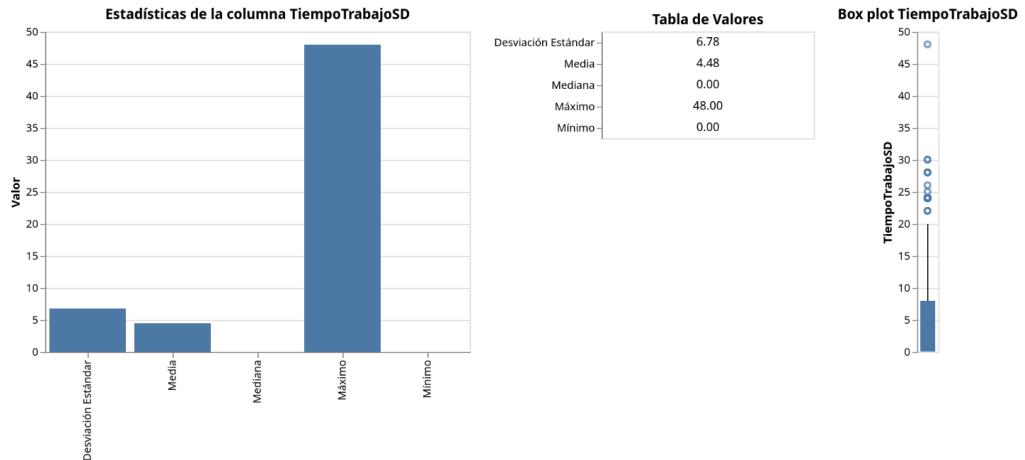
Tal como se supuso en el atributo de la edad, la mayoría de nuestras muestras trabajan.

### Tiempo Trabajo LV



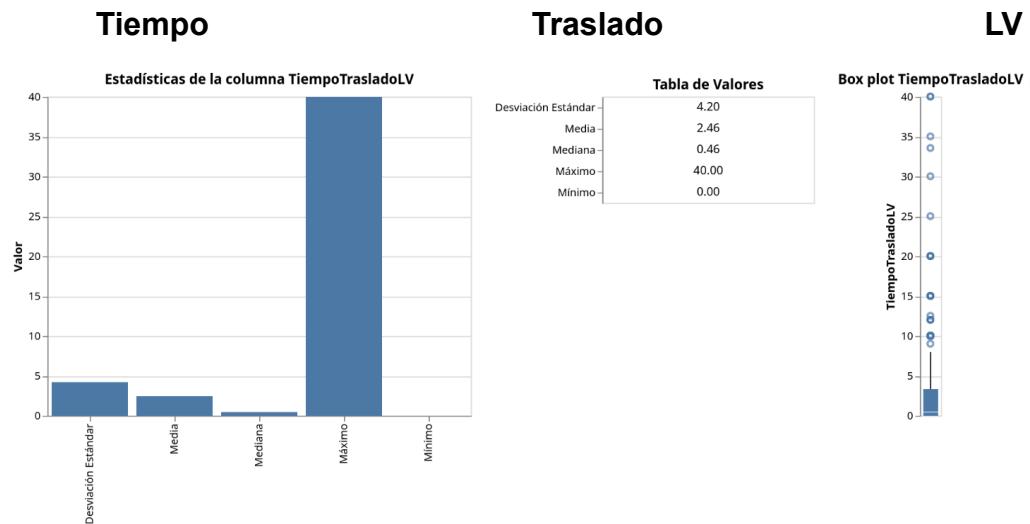
La media de trabajo semanal fue de 21 hrs, esto deriva de que poco menos de la mitad del dataset no trabaja y esto reduce en gran medida la media, lo notamos en el número gigantesco que tiene la desviación estándar. Los valores atípicos de esta gráfica destacan casos donde las horas de trabajo entre semana son muy bajas o muy altas, lo que podría reflejar diferencias en la situación laboral o particularidades en los estilos de vida y empleos de los individuos.

## Tiempo Trabajo SD



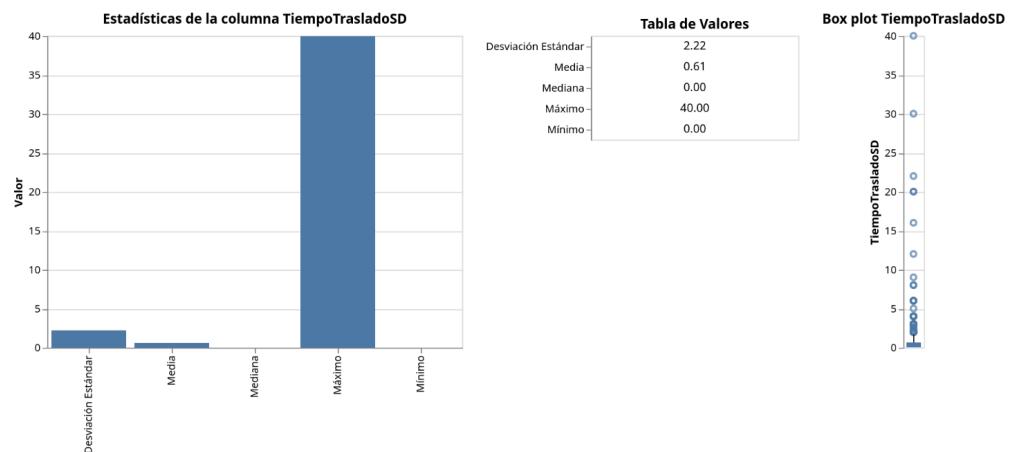
Como podemos observar la media es de 4 hrs y media, sin embargo las instancias que trabajan rondan las 25 hrs esta diferencia probablemente tiene que ver con que la mayoría no trabaja en estos días. Los datos extremos o atípicos podrían deberse a sectores o personas con necesidades específicas de trabajar los fines de semana

o personas con alta carga laboral temporalmente o actividades extraordinarias.



Al parecer la media se ve bastante afectada por las personas que no trabajan, podemos notar una concentración de puntos en 20 hrs, por lo cual podríamos intuir que es la mediana de los trabajadores de lunes a viernes.

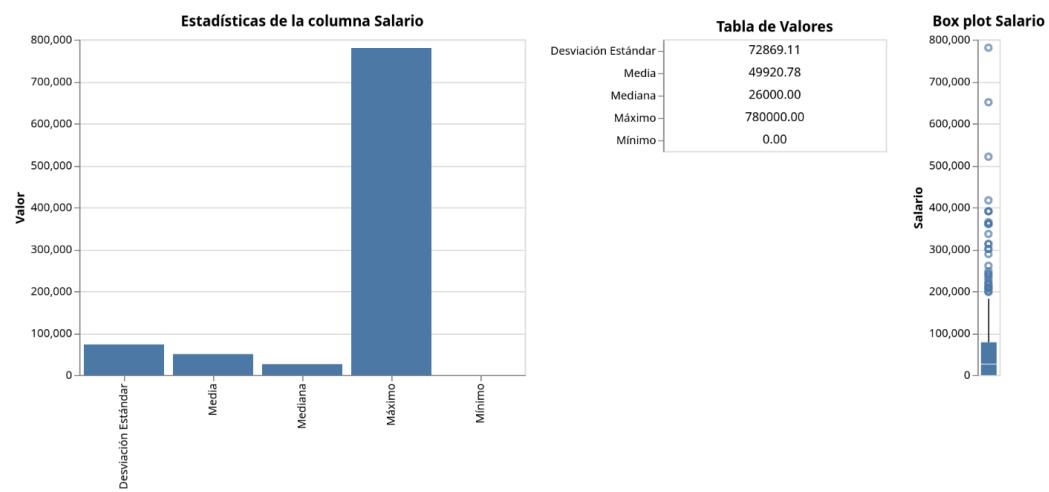
## Tiempo Traslado SD



En la barplot podemos observar que el tiempo de traslado de sábado a domingo máximo es el que destaca con un valor de 40; así mismo en el box plot se reconoce que la media está muy cerca del 0, pero hay una desviación estándar de 2.2 y algunos datos fuera de este rango con

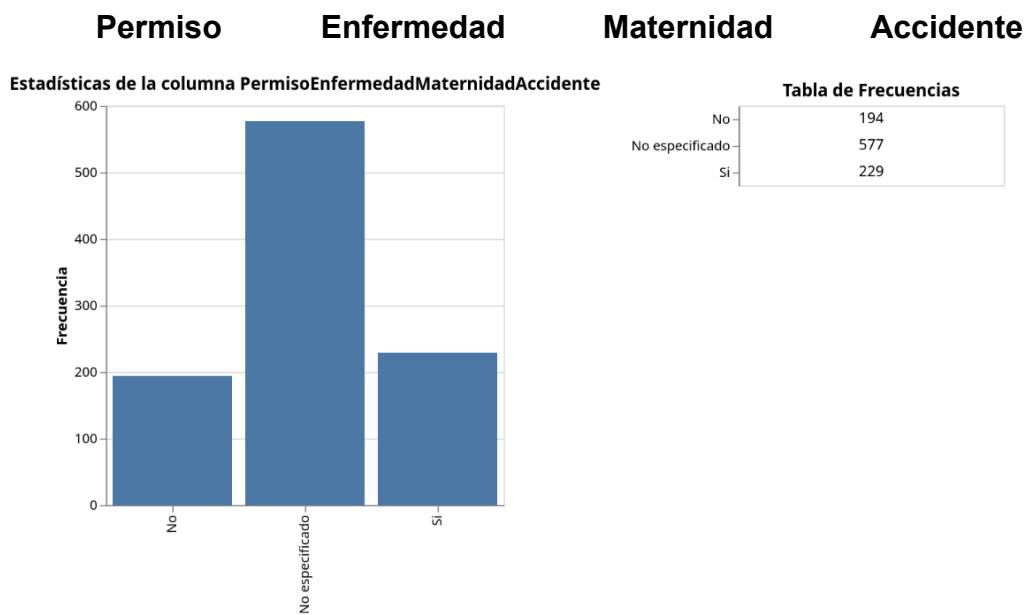
mayor valor, aunque no son significativos comparados con la mayor parte de la muestra. Los datos atípicos en este caso se pueden deber a que hay personas que por diversos motivos como laburar en fines de semana o tener actividades de recreación necesitan trasladarse más a lo habitual que el promedio.

## Salario anual

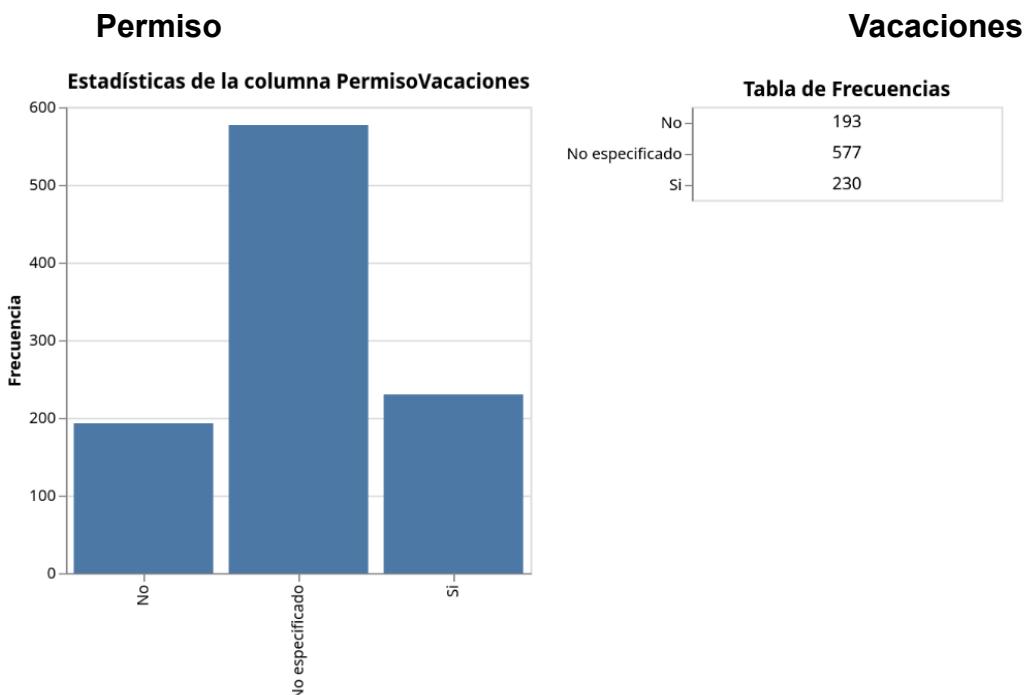


Los salarios en este dataset tienen una distribución altamente desigual, con unos pocos valores extremadamente altos y otros que directamente son 0 que distorsionan el promedio (media).

La mayoría de los salarios se encuentran cerca del rango bajo (alrededor de la mediana de 26,000) y contamos con una media del salario es de 49920, sin embargo, al ver el boxplot notamos que las personas que si trabajan tienen un salario que se concentra en alrededor de 200 000.

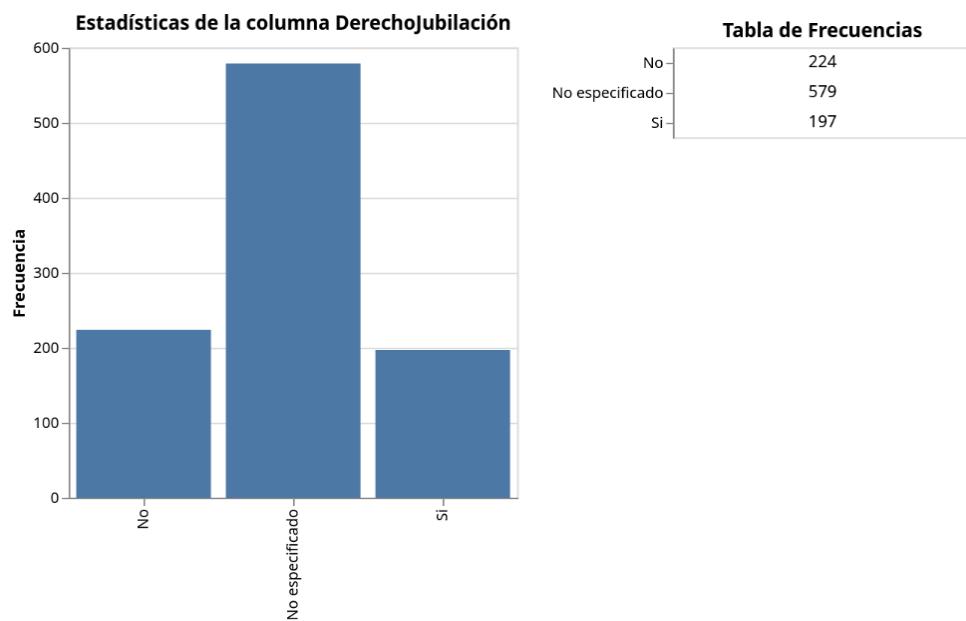


De los registros que especificaron su respuesta, más personas tienen permisos ("Sí") que no ("No"), lo que podría indicar una tendencia favorable en el acceso a este tipo de permisos. Los datos no especificados es posible que se deban a que no tienen empleo y por ello no requieren en sí un permiso, o que estén en un trabajo no formal en el que no pueden ejercer su derecho a permiso laboral.



Esta gráfica y tabla de frecuencias muestran información sobre la columna **PermisoVacaciones**, que indica si las personas tienen acceso a permisos para vacaciones. Aunque la mayoría de los datos en esta columna no están especificados, quienes indicaron tener permisos de vacaciones son más numerosos que quienes no los tienen. Los datos no especificados es posible que se deban a que no tienen empleo y por ello no requieren en sí un permiso, o que estén en un trabajo no formal en el que no pueden ejercer su derecho a permiso en el área laboral.

### Derecho Jubilación



La gráfica y la tabla de frecuencias revelan información sobre la variable **DerechoJubilación**, que indica si las personas tienen derecho a jubilación, más de la mitad del dataset tiene valores faltantes o no especificados para esta columna; además, de los datos que sí se especifican, más personas no tienen derecho a jubilación (224) que las que sí tienen (197). Los datos no especificados es posible que se deban a que no tienen empleo y por ello no cuentan con jubilación, o que estén en un trabajo no formal en el que no pueden ejercer ese derecho.

### Derecho

### Servicios

### Médicos

Estadísticas de la columna DerechoServiciosMedicos

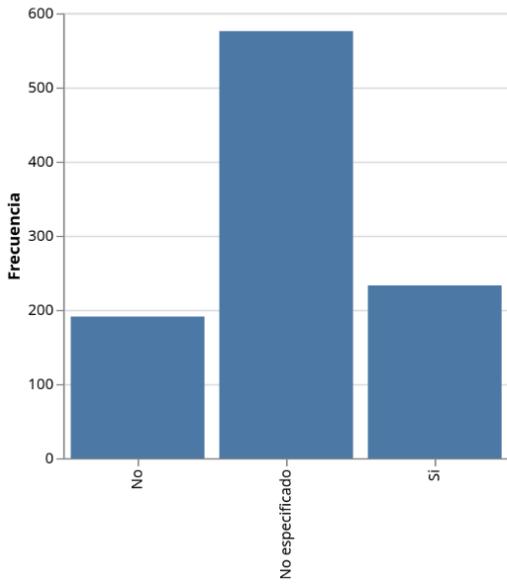
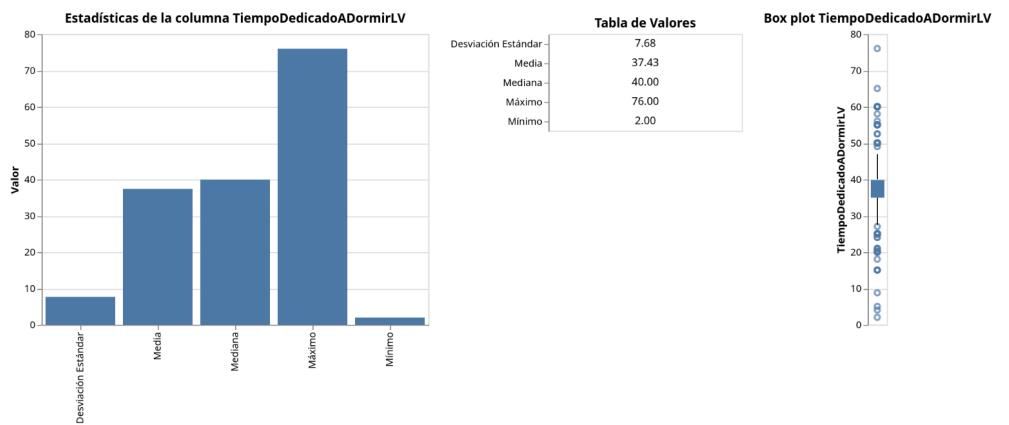


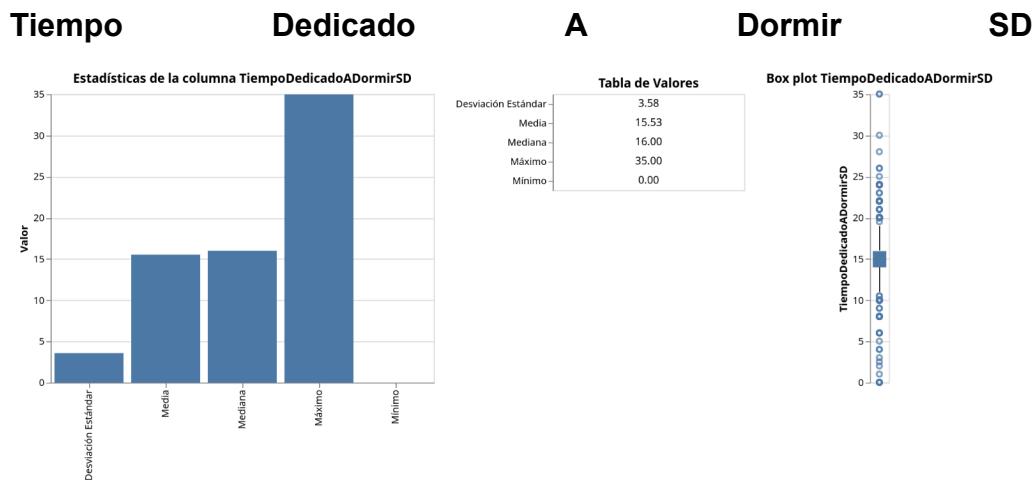
Tabla de Frecuencias

No	191
No especificado	576
Sí	233

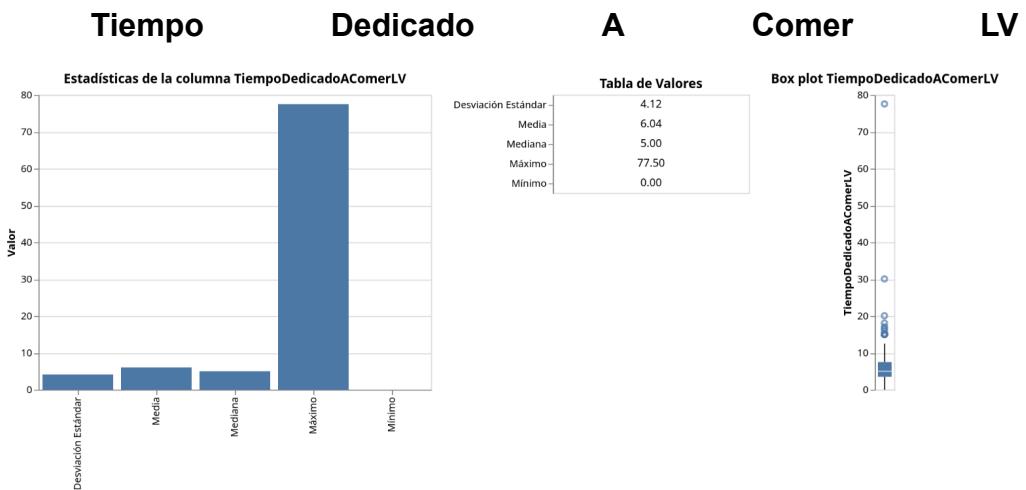
La gráfica y la tabla de frecuencias muestran información sobre la columna **DerechoServiciosMedicos**, entre las personas que especificaron su respuesta, más personas indicaron que tienen acceso a servicios médicos ("Sí") que las que no tienen ("No"). Los datos no especificados es posible que se deban a que no tienen empleo y por ello no cuentan con derecho a servicios médicos, o que estén en un trabajo no formal en el que no pueden ejercer ese derecho.

**Tiempo****Dedicado****A****Dormir****LV**

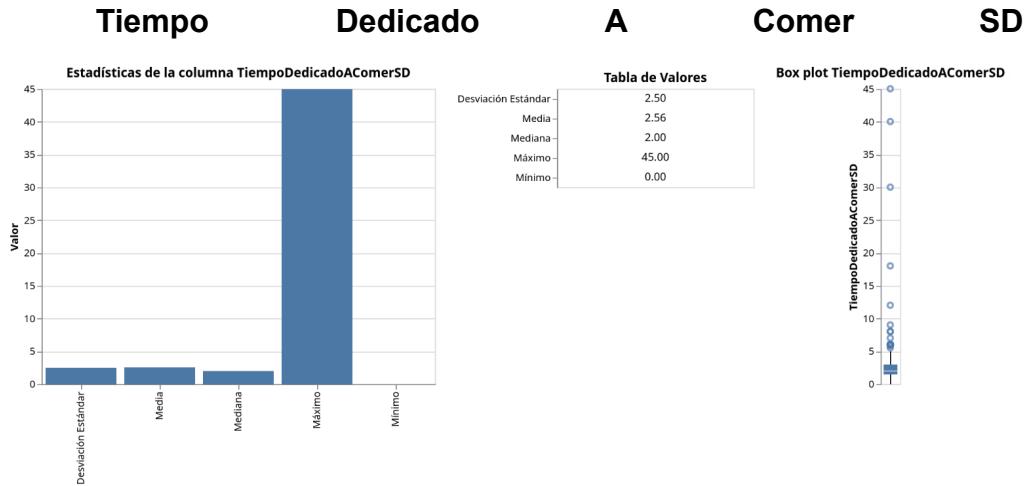
Según podemos observar la media es de 37 hrs de lunes a viernes esto da alrededor de 7 hrs y media, dado que todos duermen coincide con el boxplot, algo interesante sería ver si las personas que no trabajan o estudian duermen más en promedio. Los valores atípicos representan individuos que duermen muy poco durante los días entre semana (cerca de 2 horas en este caso) cuyas posibles causas son personas con alta carga laboral o escolar que reduce significativamente su tiempo para dormir, insomnio u otros trastornos del sueño, personas que priorizan otras actividades (trabajo nocturno, ocio, etc.). Además hay casos extremadamente altos que representan personas que duermen mucho más que el promedio (hasta 76 horas en este caso), lo cual es posible que se deba a que son personas en situación de desempleo, sin actividades demandantes o con horarios flexibles, problemas de salud que requieren más tiempo de descanso.



Al parecer la media de horas de sueño aumenta un poco los fines de semana, nuevamente notamos que coincide con el box plot, es interesante ver que hay gente que duerme 17 horas al día. Los valores atípicos representan personas que duermen muy poco durante el fin de semana, lo que puede deberse a una alta dedicación a actividades sociales, familiares o laborales incluso durante el fin de semana, problemas de insomnio o condiciones de salud que afectan el sueño, personas con horarios desbalanceados que sacrifican el tiempo de descanso. Aunado a esto hay individuos que dedican gran parte del fin de semana a dormir (alrededor del máximo de 35 horas), lo cual puede deberse a que acumulan sueño atrasado de los días entre semana, otra posibilidad es que son personas con pocas actividades demandantes durante el fin de semana o que necesitan dormir por recuperación por condiciones de salud o agotamiento extremo.

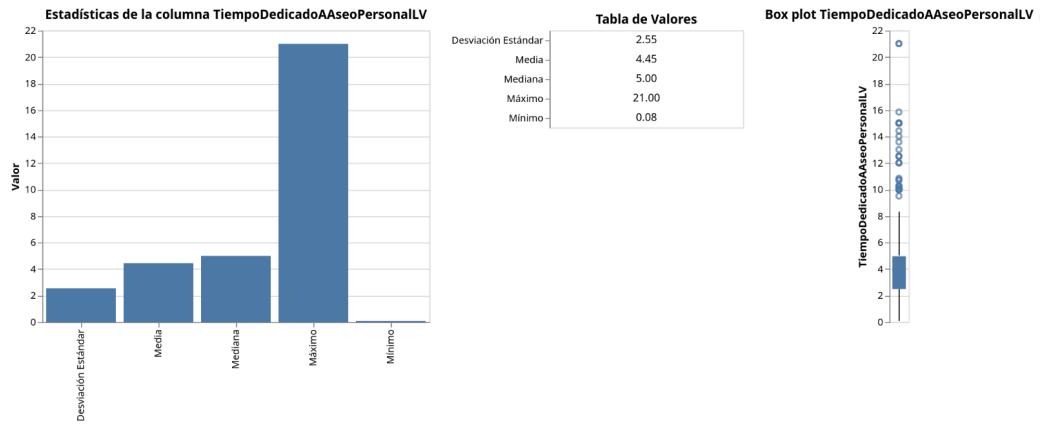


La media es de 6 horas comiendo, esto da 25 minutos por comida al día, además con ayuda del boxplot notamos que hay otra gran acumulación de instancias en alrededor de 15 hrs, con unas alrededor de una hora por comida. Los valores atípicos proporcionan información valiosa sobre hábitos alimenticios que se desvían significativamente del comportamiento promedio, en este caso los valores atípicos bajos (cerca de 0) representan personas que dedican muy poco tiempo a comer entre semana, es posible que sea debido a que tienen estilos de vida acelerados que priorizan otras actividades sobre las comidas, tienen hábitos como el consumo de alimentos rápidos o la omisión de comidas y/o que presentan condiciones que limitan el tiempo disponible para alimentarse, como trabajo o estudios. En el caso de los valores atípicos altos (alrededor de 77.50 horas) indica que hay individuos que dedican un tiempo extremadamente alto a comer entre semana, que disfrutan de largas pausas para comer como parte de su rutina, profesiones relacionadas con la preparación o consumo de alimentos, como chefs o críticos gastronómicos.



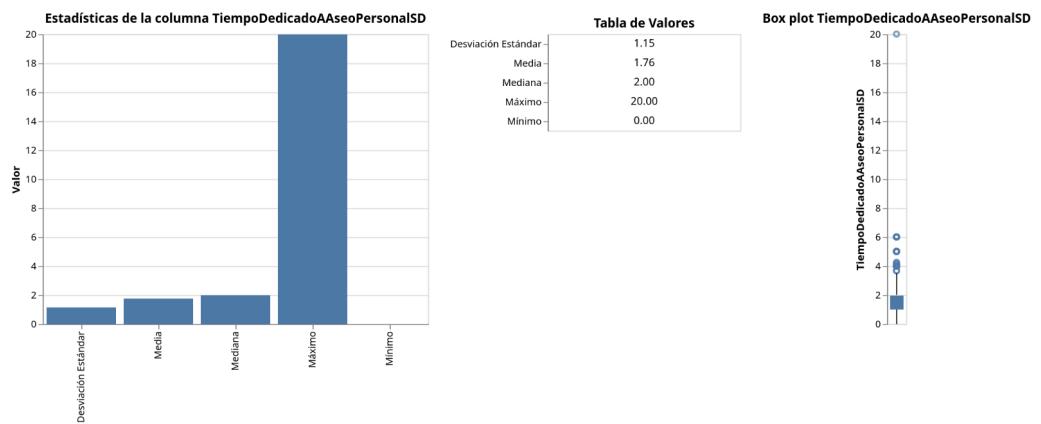
Los sábados y domingos notamos un tiempo de comida mucho más considerable de 25 minutos, el box plot sugiere lo mismo. Los valores atípicos representan extremos en los hábitos alimenticios durante este periodo, Valores atípicos bajos (cercaos a 0) representan individuos que casi no dedican tiempo a comer durante el fin de semana; posiblemente por realizar otras actividades en lugar de dedicar tiempo a las comidas, consumo rápido de alimentos, como comidas rápidas o snacks, saltarse comidas por razones personales, sociales o económicas. Los valores atípicos altos (cercaos al máximo de 45 horas) reflejan individuos que invierten mucho tiempo en comer durante el fin de semana, como posibles causas están que el tiempo dedicado a reuniones familiares o sociales alrededor de comidas largas, actividades relacionadas con la preparación y disfrute prolongado de alimentos, posibles errores en el registro de datos.

## Tiempo dedicado A Aseo Personal LV



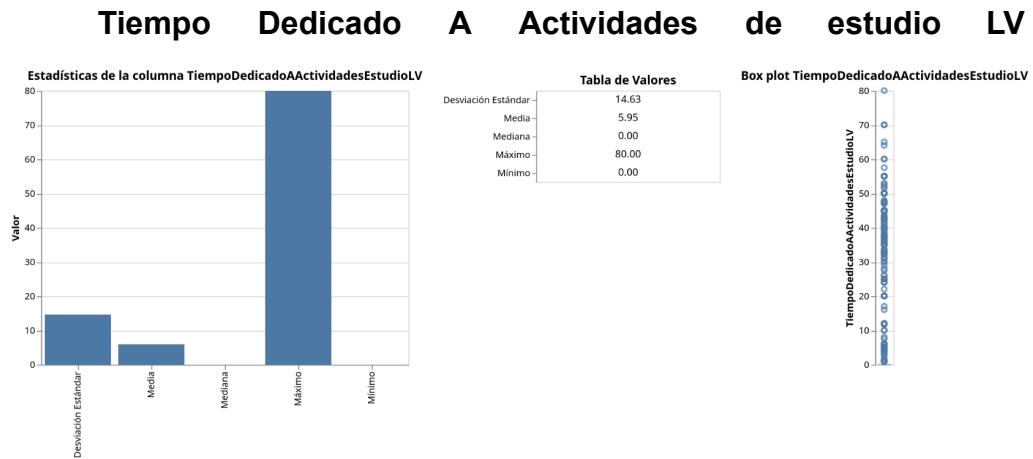
La gráfica, la tabla y box plot exponen que tenemos una media de tiempo de aseo de 4.45 de lunes a viernes lo cual se traduce en 54 minutos al día, en el box plot se puede notar otra aglomeración considerable en alrededor de 10 hrs, es probable que al separar por sexo se note la diferencia tan considerable de valores. Los valores atípicos en esta gráfica destacan diferencias individuales significativas en el tiempo dedicado al aseo personal durante los días entre semana. Estos valores pueden estar influenciados por estilos de vida, preferencias personales o factores externos.

## Tiempo Dedicado A Aseo Personal SD



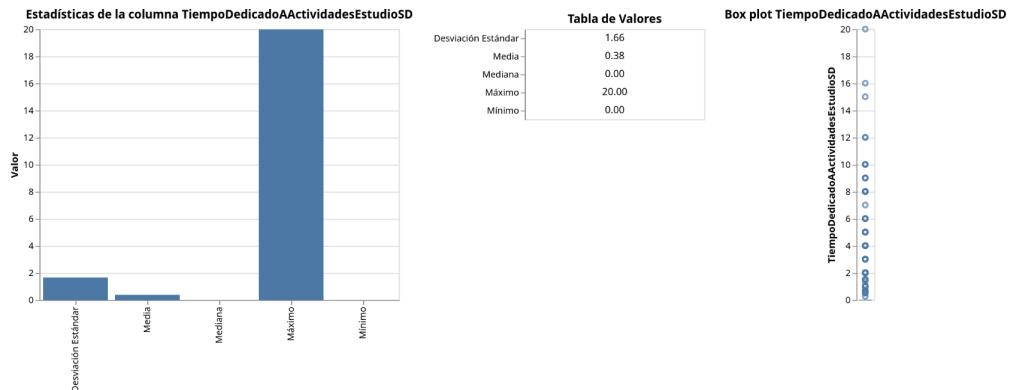
Contamos con una media de 1.76 un poco superior a la media entre semana, de la misma forma notamos en el boxplot que hay una acumulación de instancias en alrededor de 4 hrs que es

correspondiente al incremento entre semana. Los valores atípicos reflejan personas que dedican poco o nada de tiempo al aseo personal durante los fines de semana, esto puede deberse a un estilo de vida relajado durante el fin de semana, priorizando otras actividades, preferencia por concentrar actividades de aseo personal en días previos, diferencias en el concepto de "aseo personal" entre los encuestados. Los valores atípicos altos (cercaos al máximo de 20 horas) representan a individuos que dedican una cantidad considerable de tiempo a actividades relacionadas con el aseo personal, como posibles causas están las actividades específicas como sesiones de spa, cuidado estético intensivo o preparación para eventos sociales, interpretación amplia de la actividad, rutinas o prácticas más detalladas durante el fin de semana.

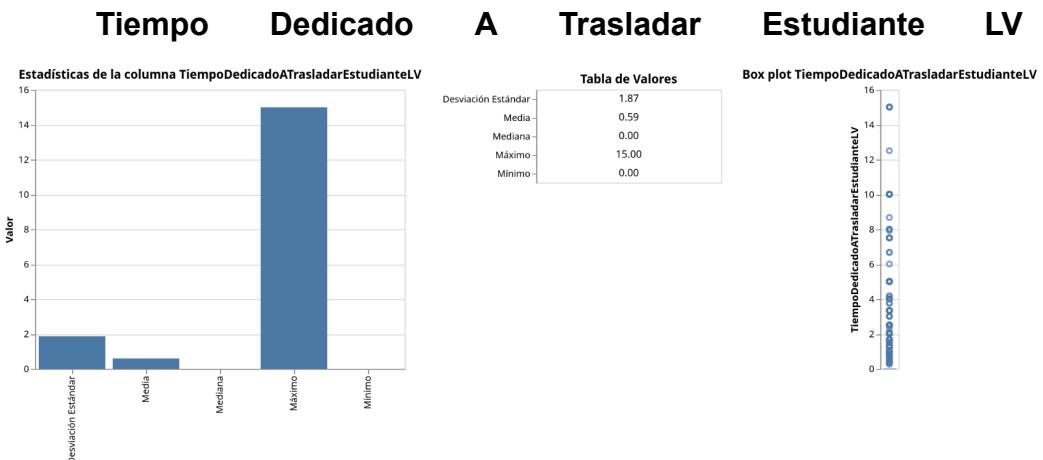


La gráfica, la tabla y box plot ofrecen información sobre la columna **TiempoDedicadoAAActividadesEstudioLV**, que refleja el tiempo dedicado a actividades de estudio durante los días laborables (LV). La mayoría de las personas no dedica tiempo a actividades de estudio durante los días laborales, aunque existe un grupo pequeño que reporta una dedicación significativa. Esto sugiere una distribución desigual en los hábitos de estudio.

### Tiempo Dedicado A Actividades de estudio SD



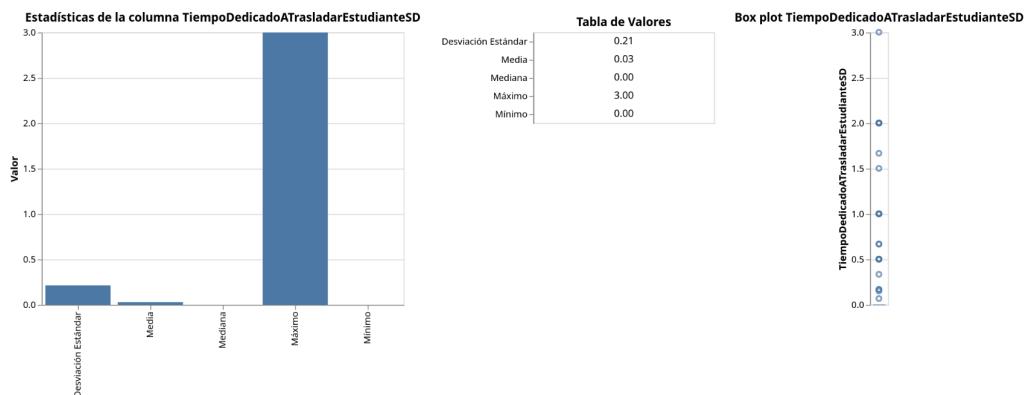
La gráfica, la tabla y box plot indican información sobre la columna **TiempoDedicadoAAActividadesEstudioSD**, que refleja el tiempo dedicado a actividades de estudio durante los días de fin de semana (SD). La mayoría de las personas no dedica tiempo al estudio durante los fines de semana, lo que podría deberse a otras responsabilidades o preferencias. Sin embargo, existen algunos casos que muestran una alta dedicación, lo cual es interesante analizar más a fondo.



La gráfica, la tabla y box plot evidencian información sobre la columna **TiempoDedicadoATrasladarEstudianteLV**, que refleja el tiempo dedicado a trasladar a estudiantes durante los días laborables (LV). La mayoría de las personas no dedican tiempo al traslado de estudiantes en días laborales, pero existe un pequeño grupo que sí lo hace, con una alta dedicación en algunos casos particulares. Esto podría reflejar

diferentes niveles de responsabilidad o necesidades específicas en este ámbito.

### Tiempo Dedicado A Trasladar SD

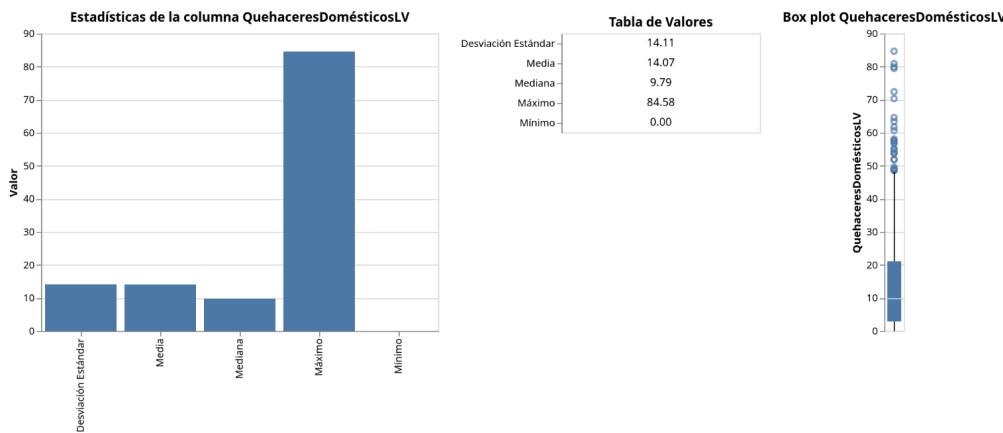


La gráfica, la tabla y box plot revelan información sobre la columna **TiempoDedicadoATrasladarEstudianteSD**, que refleja el tiempo dedicado a trasladar a estudiantes durante los días de fin de semana (SD). La mayor parte de los valores se concentran en 0 horas, Un caso muestra 3 horas dedicadas, lo cual es el tiempo más alto reportado. reflejando la baja participación en esta actividad durante los fines de semana. La mayoría de las personas no dedican tiempo al traslado de estudiantes durante los fines de semana, con solo algunos casos específicos reportando tiempo dedicado, en rangos bajos como máximo de 3 horas.

## Quehaceres

## Domésticos

LV

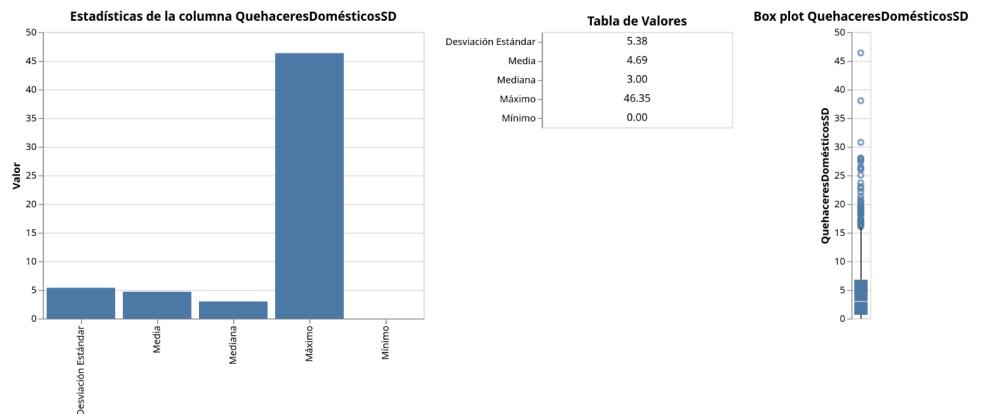


La gráfica, la tabla y box plot presentan información sobre la columna **QuehaceresDomésticosLV**, que refleja el tiempo dedicado a quehaceres domésticos durante los días laborales (LV). La mayoría de las personas dedica menos de 20 horas semanales a quehaceres domésticos durante los días laborales, aunque existe un grupo que reporta tiempos mucho más altos, lo que sugiere un compromiso doméstico intensivo en ciertos casos.

## Quehaceres

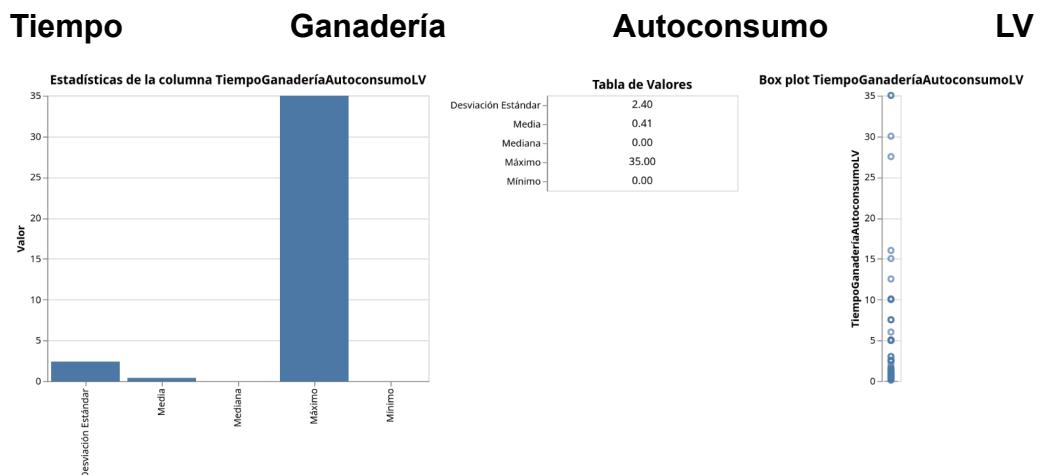
## Domésticos

SD

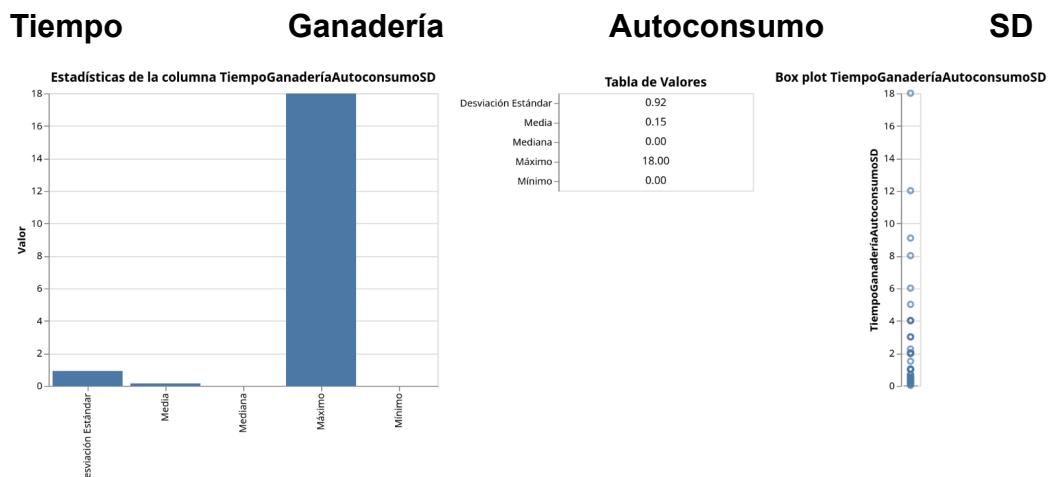


La gráfica, la tabla y box plot exponen información sobre la columna **TiempoGanaderíaAutoconsumoLV**, que refleja el tiempo dedicado a actividades de ganadería para autoconsumo durante los días de lunes a viernes (LV). La mayoría de las personas no realiza actividades de ganadería para autoconsumo durante los días laborales. Sin embargo,

existen casos aislados con dedicación considerable que podrían corresponder a personas en contextos rurales o con estilos de vida particulares.



La gráfica, la tabla y box plot ofrecen nuevamente la información sobre la columna **TiempoGanaderíaAutoconsumoLV**, que mide el tiempo dedicado a actividades de ganadería para autoconsumo de lunes a viernes (LV). La actividad de ganadería para autoconsumo es poco común durante los días laborales, con la mayoría de las personas reportando 0 horas. Solo unos pocos casos dedican tiempo significativo, probablemente en contextos rurales o específicos.

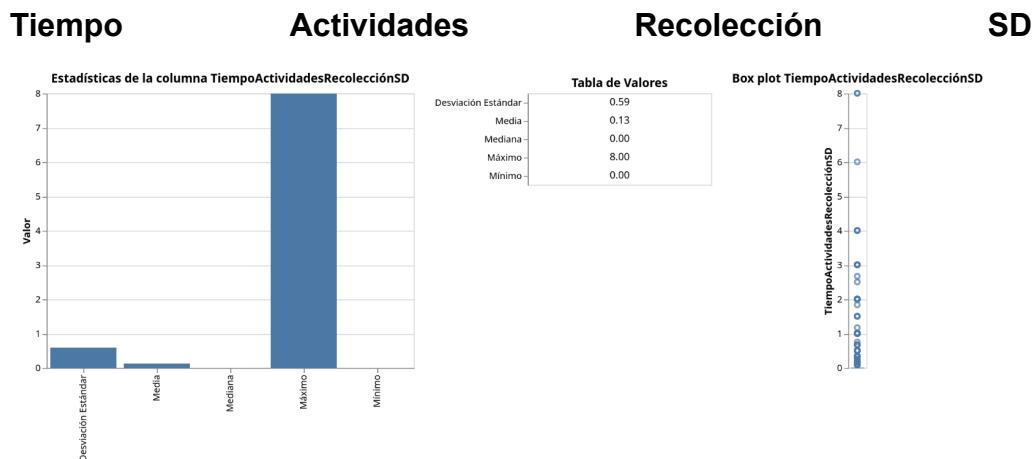


La gráfica, la tabla y box plot manifiestan información sobre la columna

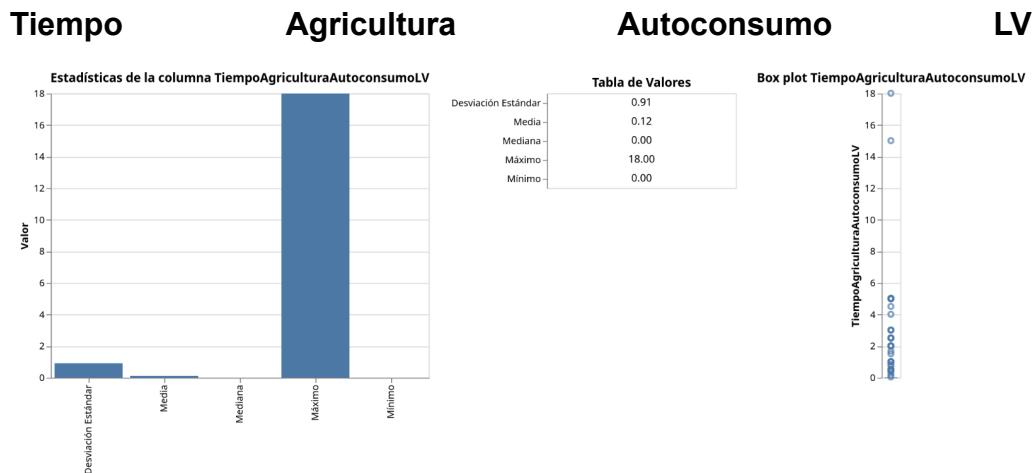
**TiempoGanaderíaAutoconsumoSD**, que mide el tiempo dedicado a actividades de ganadería para autoconsumo durante los días de fin de semana (SD). La actividad de ganadería para autoconsumo durante los fines de semana es rara, con solo unos pocos casos que reportan tiempo significativo. La mayoría de los valores son nulos, lo que refleja que esta práctica no es generalizada en la población analizada.



La gráfica y la tabla muestran información sobre la columna **TiempoActividadesRecolecciónLV**, que refleja el tiempo dedicado a actividades de recolección de lunes a viernes (LV). La actividad de recolección de lunes a viernes es mínima o inexistente para la mayoría de las personas en el dataset. Sin embargo, los pocos casos que dedican tiempo significativo reflejan una actividad intensiva, probablemente vinculada a ocupaciones específicas o contextos rurales.

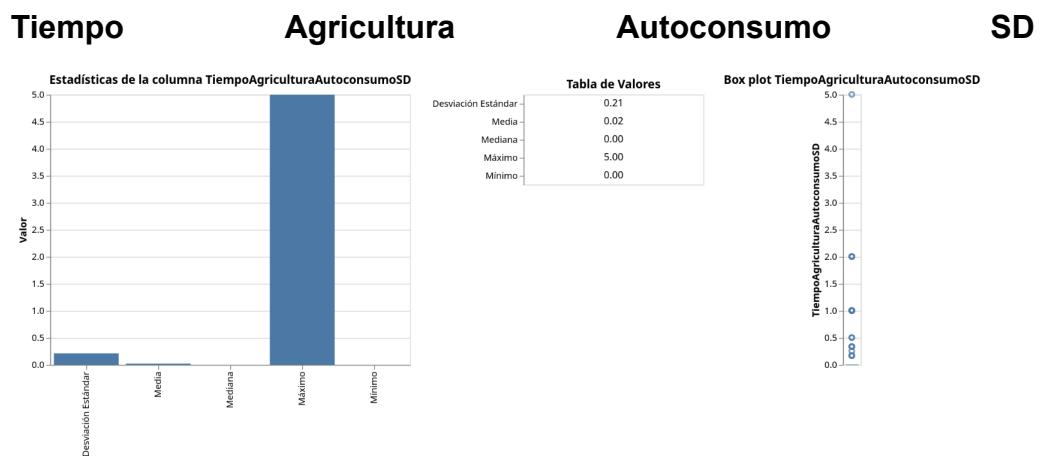


La gráfica, la tabla y box plot presentan información sobre la columna **TiempoActividadesRecolecciónSD**, que refleja el tiempo dedicado a actividades de recolección durante los días de fin de semana (SD). La actividad de recolección durante los fines de semana es mínima para la mayoría de las personas, con solo unos pocos casos reportando tiempos significativos. Esto sugiere que la recolección no es una actividad común para la mayoría de los encuestados. Algunas personas que reportan datos atípicos podrían deberse a que tienen como principal actividad la recolección o estilos de vida específicos.

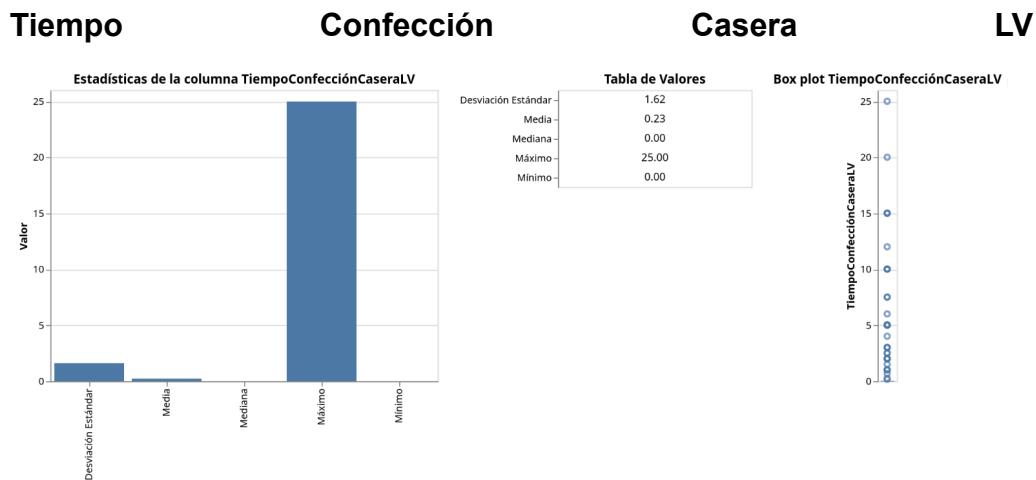


La gráfica, la tabla y box plot revelan información sobre la columna **TiempoAgriculturaAutoconsumoLV**, que refleja el tiempo dedicado a actividades agrícolas para autoconsumo entre semana (LV). La actividad agrícola para autoconsumo entre semana es mínima para la

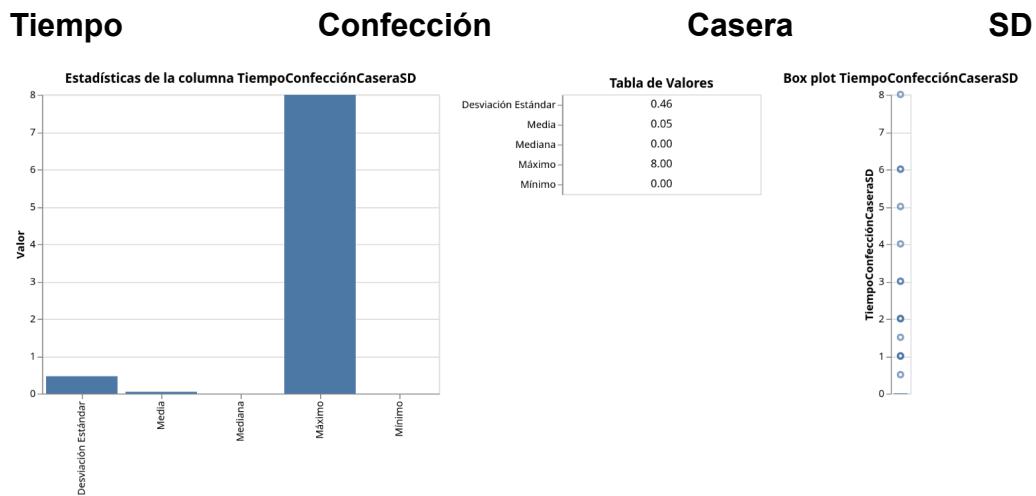
mayoría de las personas en el dataset. Solo unos pocos casos reportan tiempos significativos, lo que indica que esta práctica es relevante sólo para un pequeño segmento de la población. Los datos atípicos altos representan personas que posiblemente tienen estilos de vida en los que por preferencia deciden sembrar su propio alimento o que se encuentran en un lugar en el que no les es posible comprar y necesitan proveerse por sí mismos(as) su alimento.



La gráfica, la tabla y box plot evidencian información sobre la columna **TiempoAgriculturaAutoconsumoSD**, que mide el tiempo dedicado a actividades agrícolas para autoconsumo durante los días de fin de semana (SD). La actividad agrícola para autoconsumo durante los fines de semana es mínima para la mayoría de la población en el dataset, con solo unos pocos casos reportando tiempo significativo. Esto sugiere que es una práctica poco común y probablemente específica de ciertos grupos o contextos.

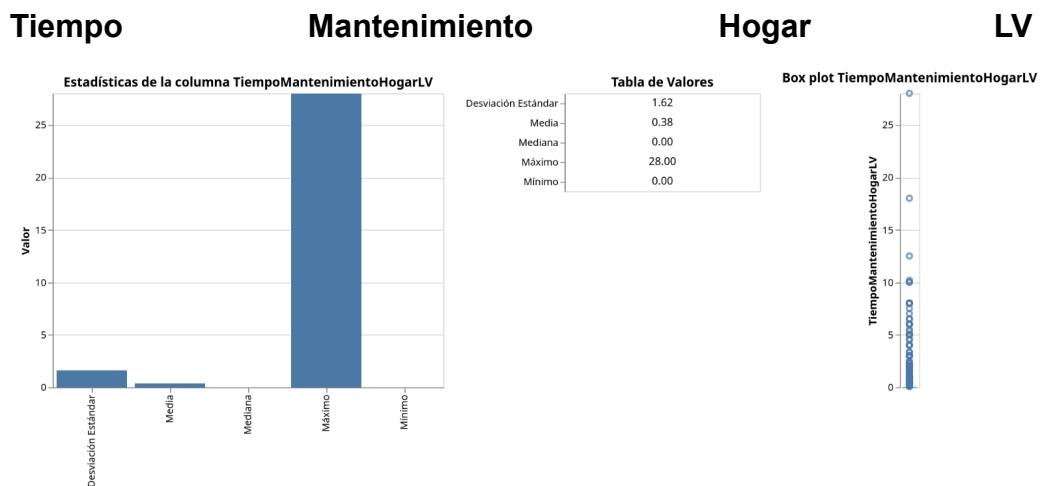


La gráfica, la tabla y box plot evidencian información sobre la columna **TiempoConfecciónCaseraLV**, que mide el tiempo dedicado a actividades de confección casera de lunes a viernes (LV). La confección casera entre semana es una actividad poco común, con la mayoría de las personas dedicando 0 horas a esta tarea. Sin embargo, los casos que reportan tiempos significativos indican que puede ser una actividad importante para ciertos grupos específicos.



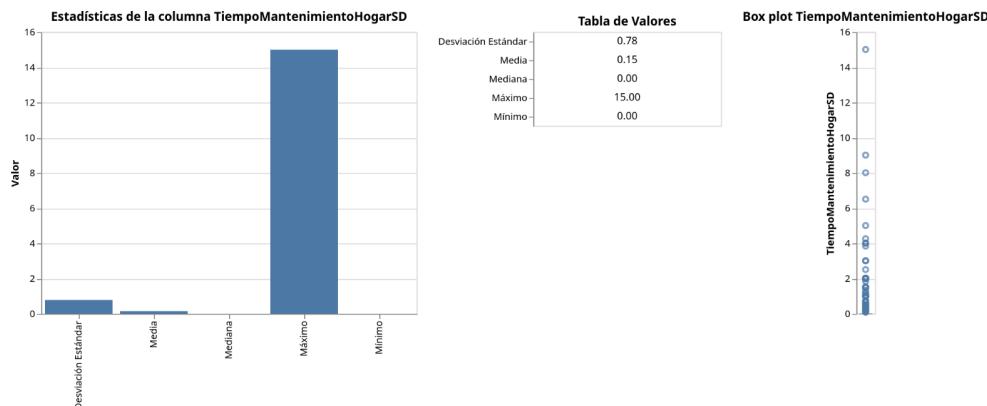
La gráfica, la tabla y box plot presentan información sobre la columna **TiempoConfecciónCaseraSD**, que refleja el tiempo dedicado a actividades de confección casera durante los fines de semana. La actividad de confección casera durante el fin de semana es muy poco común, con solo unos pocos casos que reportan tiempos significativos.

Esto sugiere que es una práctica ocasional o específica para ciertos grupos.



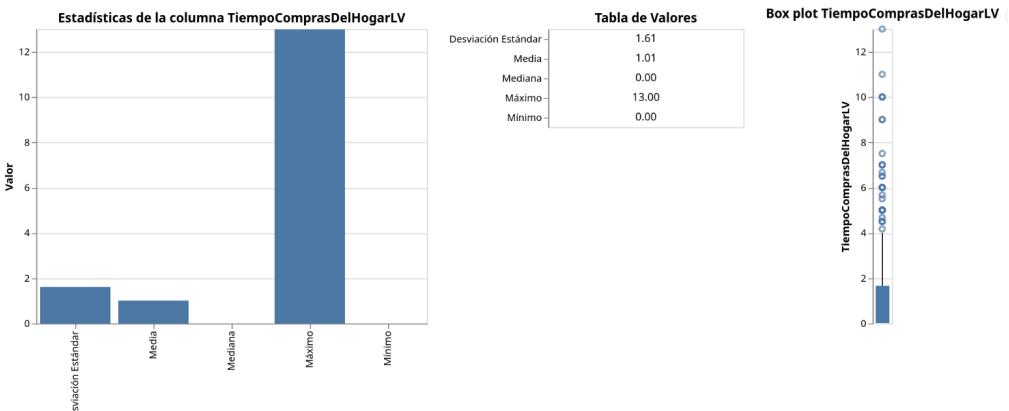
La gráfica, la tabla y box plot revelan información sobre la columna **TiempoMantenimientoHogarLV**, que mide el tiempo dedicado a actividades de mantenimiento del hogar entre semana (de lunes a viernes). El mantenimiento del hogar entre semana es una actividad poco común para la mayoría de las personas, con solo unos pocos casos reportando tiempos significativos. Esto sugiere que, en general, esta actividad podría ser relegada a los fines de semana o no considerarse relevante en su contexto diario; en el caso de quienes reportan alta cantidad de horas para mantenimiento del hogar, es posible que estén en un hogar con condiciones especiales que necesiten alta dedicación o mejoras.

## Tiempo Mantenimiento Hogar SD

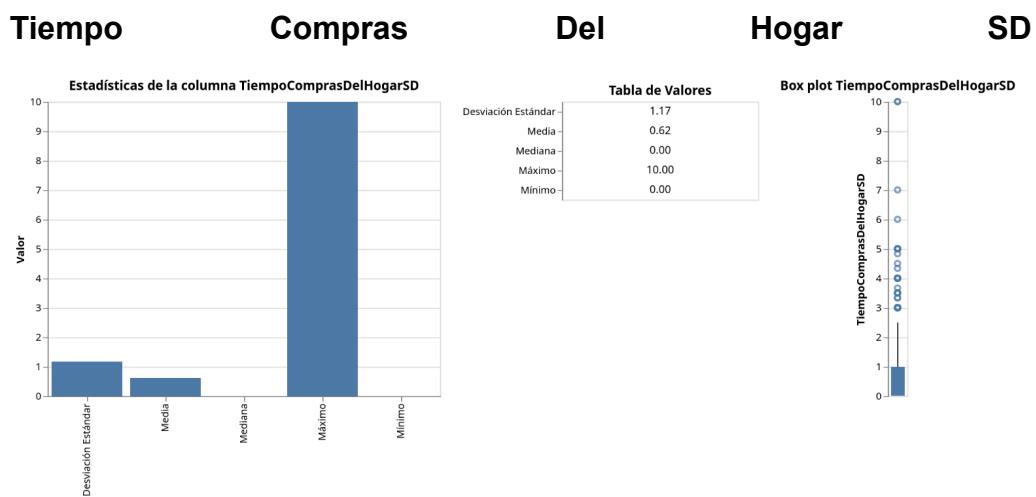


La gráfica, la tabla y box plot ofrecen información sobre la columna **TiempoMantenimientoHogarSD**, que refleja el tiempo dedicado a actividades de mantenimiento del hogar durante los fines de semana. El mantenimiento del hogar durante los fines de semana es una actividad poco común para la mayoría, con solo unos pocos casos que reportan tiempos significativos. Esto sugiere que, en general, esta actividad no es una prioridad principal durante este periodo. Para el caso de los altos niveles de horas reportadas para esta actividad se infiere que existen casos particulares de hogares que requieren mucho mantenimiento o personas que desean tener de manera impecable su hogar y saben cómo mejorarlo.

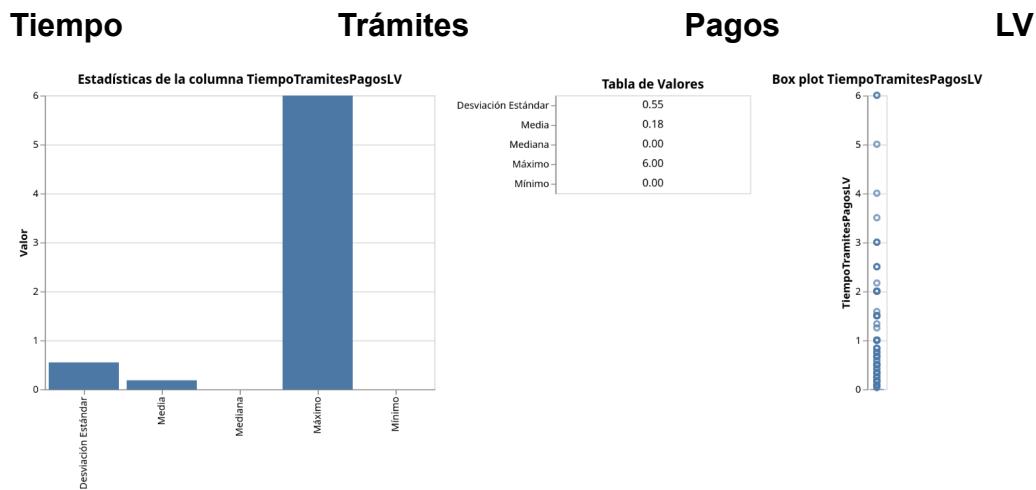
## Tiempo Compras Del Hogar LV



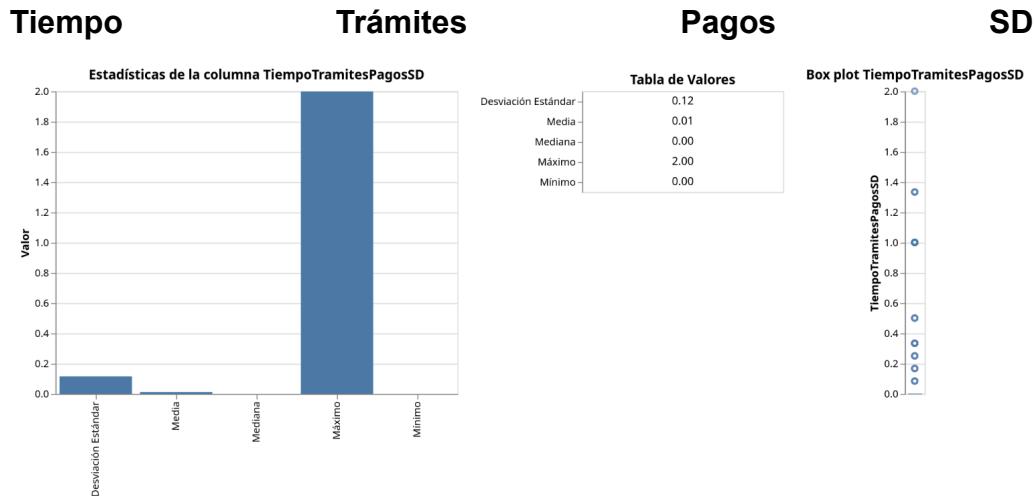
La gráfica, la tabla y box plot destapa información sobre la columna **TiempoComprasDelHogarLV**, que mide el tiempo dedicado a realizar compras del hogar entre semana (de lunes a viernes). Realizar compras del hogar entre semana es una actividad poco común para la mayoría de las personas, aunque hay algunos casos destacados que reportan tiempos significativos. Esto sugiere que la frecuencia de esta actividad puede depender del rol de la persona en el hogar o de su disponibilidad de tiempo.



La gráfica, la tabla y el box plot exponen información sobre la columna **TiempoComprasDelHogarSD**, que mide el tiempo dedicado a realizar compras del hogar durante los fines de semana. Realizar compras del hogar durante los fines de semana es una actividad poco común para la mayoría de las personas, pero algunos casos reportan tiempos significativos, posiblemente relacionados con compras planificadas o tareas intensivas de abastecimiento.



La gráfica, la tabla y el box plot manifiestan información sobre la columna **TiempoTrámitesPagosLV**, que mide el tiempo dedicado a realizar trámites y pagos entre semana (de lunes a viernes). Realizar trámites y pagos entre semana es una actividad poco común para la mayoría de las personas, pero algunos casos reportan tiempos significativos. Esto sugiere que esta actividad podría depender de factores como la disponibilidad de tiempo o necesidades específicas.

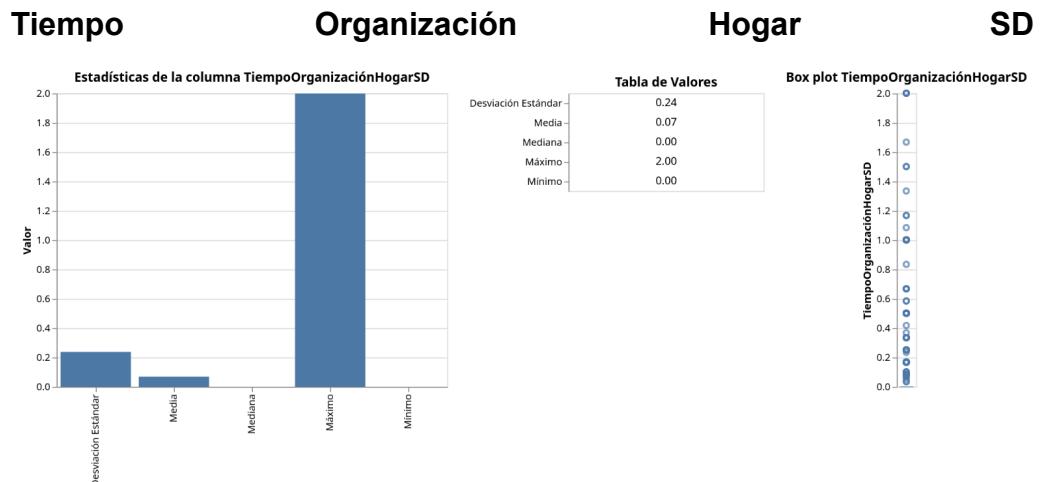


La gráfica, la tabla y el box plot revelan información sobre la columna **TiempoTrámitesPagosSD**, que mide el tiempo dedicado a realizar trámites y pagos durante los fines de semana. La actividad de realizar trámites y pagos durante los fines de semana es poco común, con solo

unos pocos casos que dedican tiempo significativo. Esto podría deberse a restricciones de horarios o a la planificación de estos trámites en otros días.

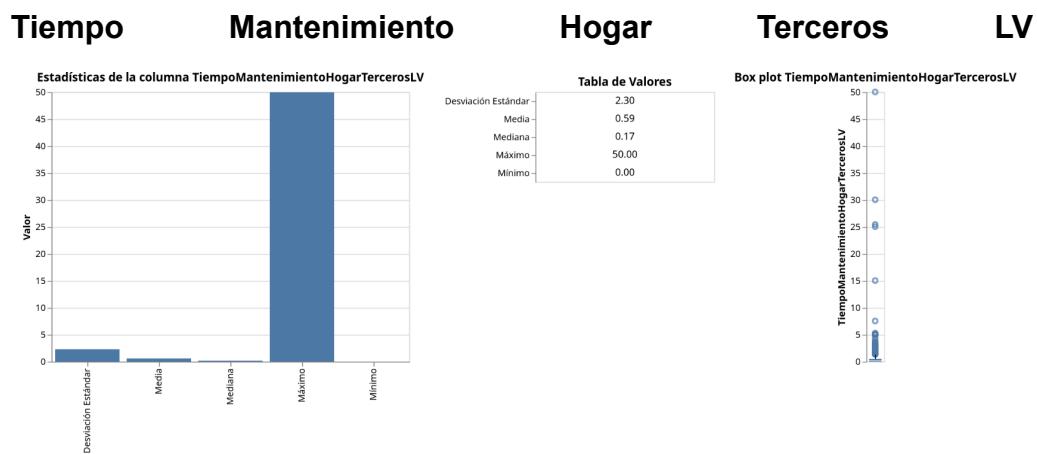


La gráfica, la tabla y el box plot ofrecen información sobre la columna **TiempoOrganizaciónHogarLV**, que mide el tiempo dedicado a organizar el hogar entre semana (de lunes a viernes). Organizar el hogar entre semana es una actividad poco frecuente para la mayoría de las personas, pero algunos casos reportan tiempos significativos. Esto puede deberse a roles específicos en el hogar o disponibilidad de tiempo.

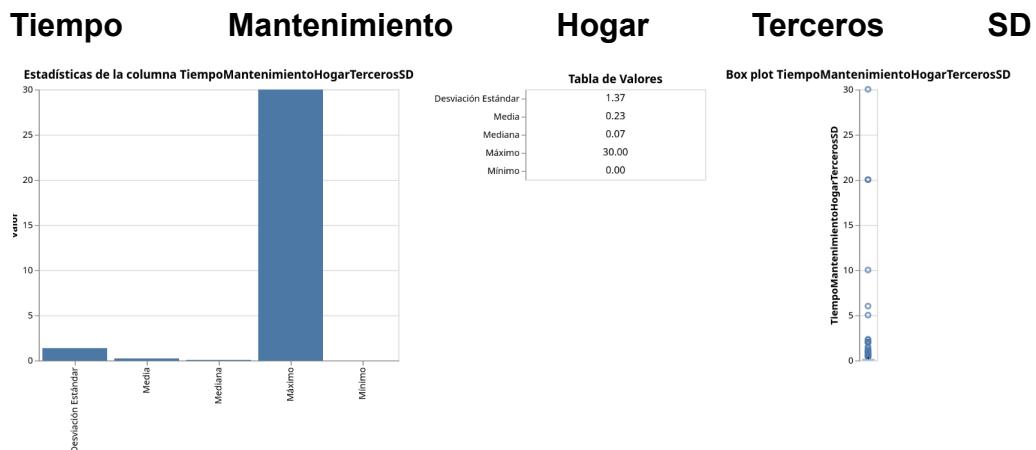


La gráfica, la tabla y el box plot evidencian información sobre la

columna **TiempoOrganizaciónHogarSD**, que mide el tiempo dedicado a organizar el hogar durante los fines de semana. Organizar el hogar durante los fines de semana es una actividad poco frecuente para la mayoría de las personas. Sin embargo, algunos casos aislados muestran una dedicación significativa, probablemente por razones específicas, como particularidades de los gustos por la persona a ser muy organizada.

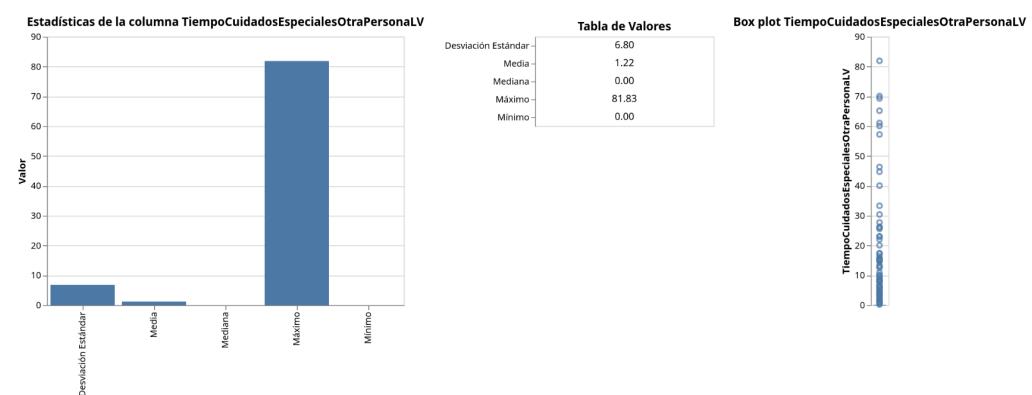


La gráfica, la tabla y el box plot presentan información sobre la columna **TiempoMantenimientoHogarTercerosLV**, que mide el tiempo dedicado entre semana (de lunes a viernes) al mantenimiento del hogar de terceros. El mantenimiento de hogares de terceros entre semana es una actividad poco común, pero algunos casos excepcionales reflejan dedicación significativa, probablemente vinculada a ocupaciones específicas o situaciones en las que se apoya a algún familiar que está pasando por malos momentos.



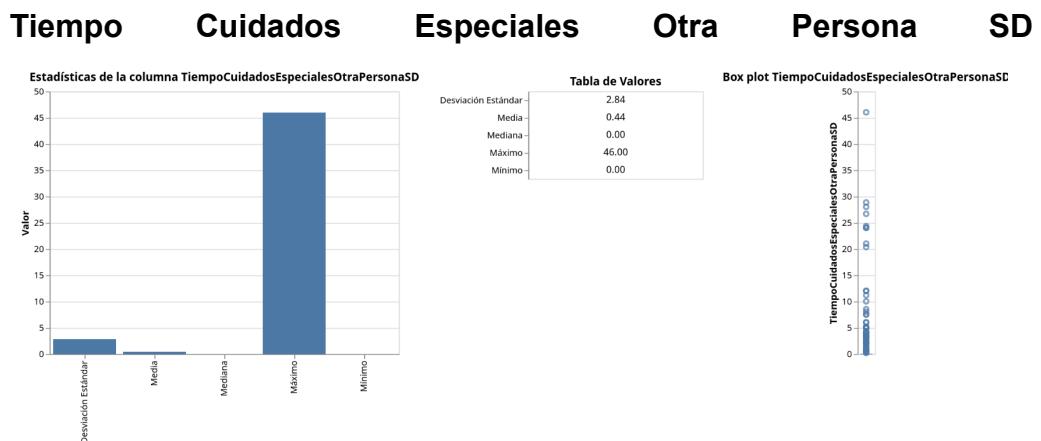
La gráfica, la tabla y el box plot revelan información sobre la columna **TiempoMantenimientoHogarTercerosSD**, que mide el tiempo dedicado durante los fines de semana al mantenimiento de hogares de terceros. Brindar cuidados especiales entre semana es una actividad minoritaria para la mayoría de las personas, pero algunos casos extremos reflejan dedicación significativa, posiblemente vinculados a contextos laborales o familiares específicos.

### Tiempo Cuidados Especiales Otra Persona LV

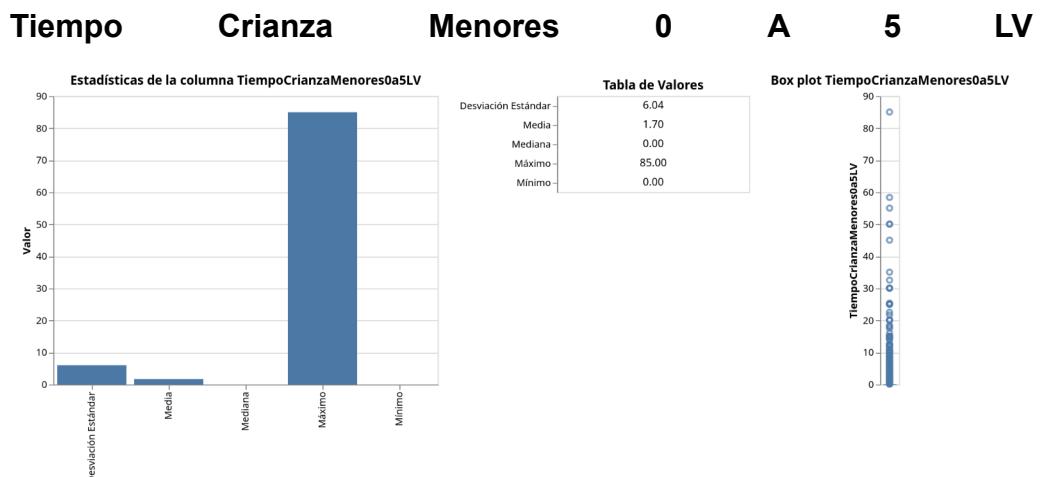


La gráfica, la tabla y el box plot presentan información sobre la columna **TiempoCuidadosEspecialesOtraPersonaLV**, que mide el tiempo dedicado entre semana (de lunes a viernes) a brindar cuidados especiales a otra persona. Brindar cuidados especiales entre semana

es una actividad minoritaria para la mayoría de las personas, pero algunos casos extremos reflejan dedicación significativa, posiblemente vinculados a contextos laborales o familiares específicos.



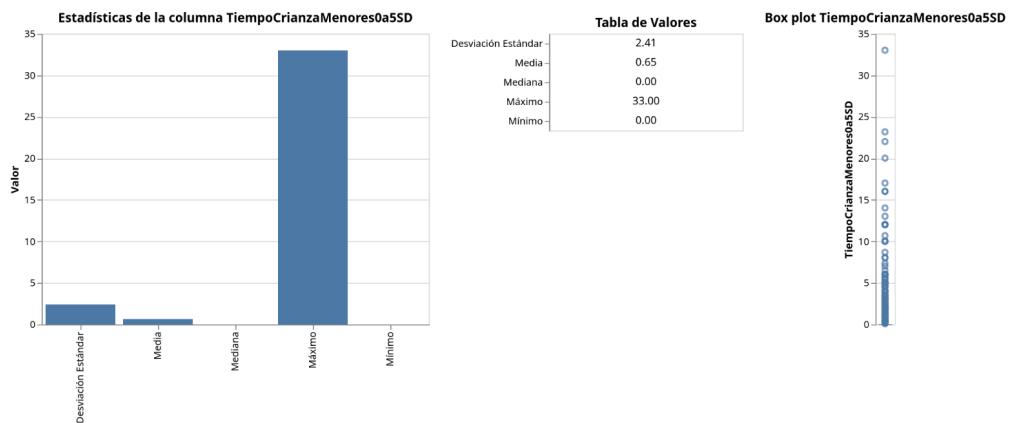
La gráfica, la tabla y el box plot ofrecen información sobre la columna **TiempoCuidadosEspecialesOtraPersonaSD**, que mide el tiempo dedicado a brindar cuidados especiales a otra persona durante el fin de semana. Los cuidados especiales a otra persona durante el fin de semana son una actividad que no realiza la mayoría, pero hay casos particulares donde se dedican muchas horas, posiblemente por motivos laborales o familiares.



La gráfica, la tabla y el box plot exhiben información sobre la columna **TiempoCrianzaMenores0a5LV**, que mide el tiempo dedicado al

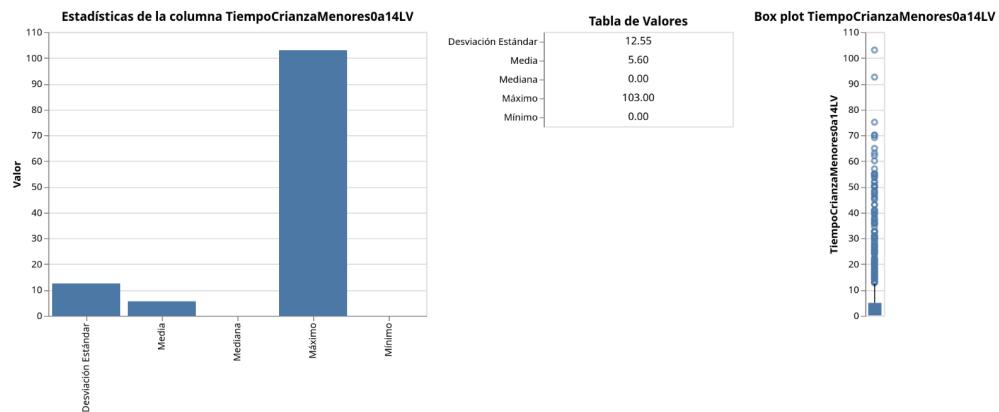
cuidado de niños de 0 a 5 años entre semana. El cuidado de niños menores de 5 años es una actividad que no realiza la mayoría entre semana, pero hay casos donde consume una gran cantidad de tiempo, reflejando posibles necesidades familiares o laborales específicas.

### Tiempo Crianza Menores 0 A 5 SD



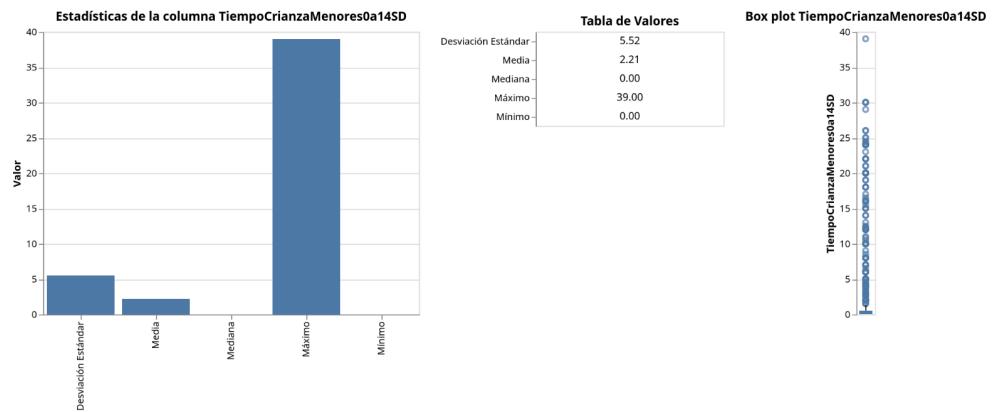
La gráfica, la tabla y el box plot revelan los datos de la columna **TiempoCrianzaMenores0a5SD**, que mide el tiempo dedicado al cuidado de niños de 0 a 5 años durante el fin de semana. El tiempo dedicado al cuidado de niños pequeños durante el fin de semana es generalmente bajo para la mayoría, pero existen casos excepcionales con valores altos. Esto sugiere dinámicas familiares diversas como por ejemplo, que una sola persona se dedique a dicha actividad o tenga muchos hijos en este rango de edad.

## Tiempo Crianza Menores 0 A 14 LV



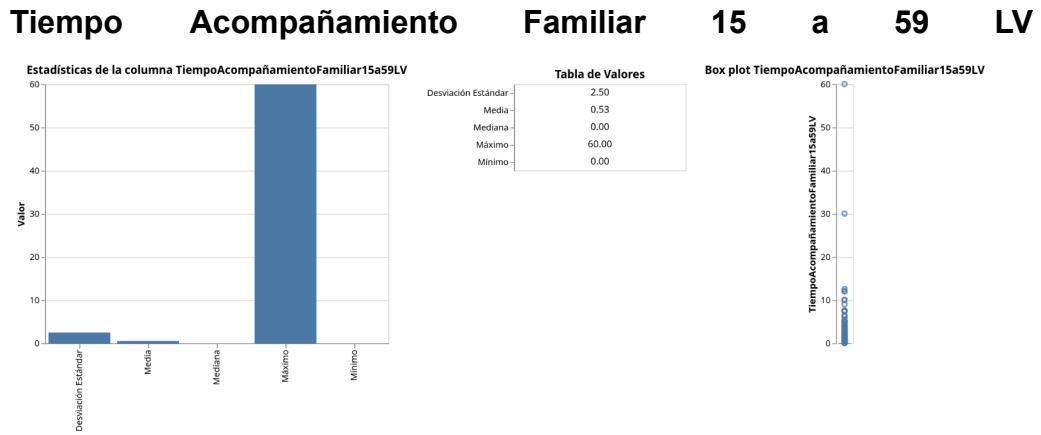
La gráfica, la tabla y el box plot exponen los datos de la columna **TiempoCrianzaMenores0a14LV**, que mide el tiempo dedicado al cuidado de niños de 0 a 14 años entre semana. Aunque muchas personas no dedican tiempo al cuidado de niños de 0 a 14 años entre semana, los datos muestran que hay una minoría significativa que invierte una cantidad considerable de tiempo. Esto refleja diferentes realidades familiares y sociales, como la distribución de responsabilidades en el hogar o el acceso a guarderías o escuelas.

## Tiempo Crianza Menores 5 A 14 SD

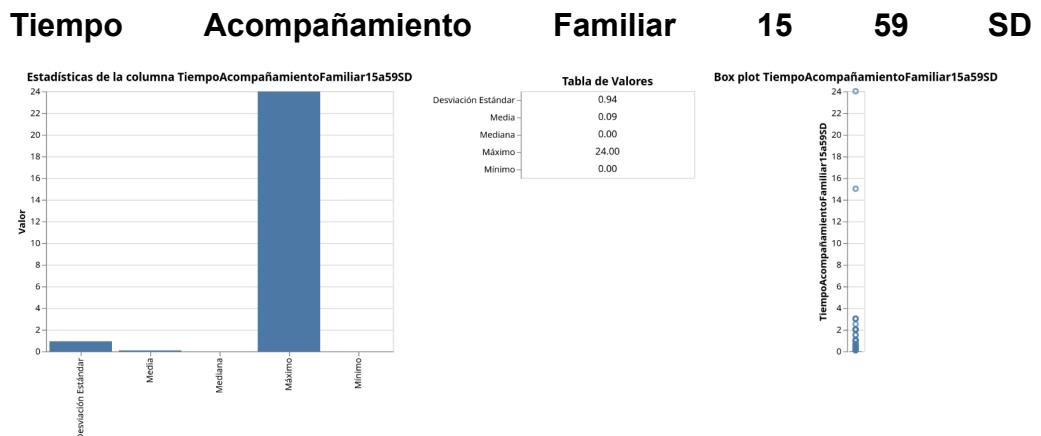


La gráfica, la tabla y el box plot presentan los datos de la columna **TiempoCrianzaMenores0a14SD**, que mide el tiempo dedicado al cuidado de niños de 0 a 14 años los fines de semana. Aunque una gran parte de las personas no dedica tiempo al cuidado de niños de 0 a 14

años los fines de semana, los datos muestran que hay una minoría significativa que invierte mucho tiempo en esta actividad. Esto refleja dinámicas familiares específicas y posibles desigualdades en la distribución de responsabilidades.

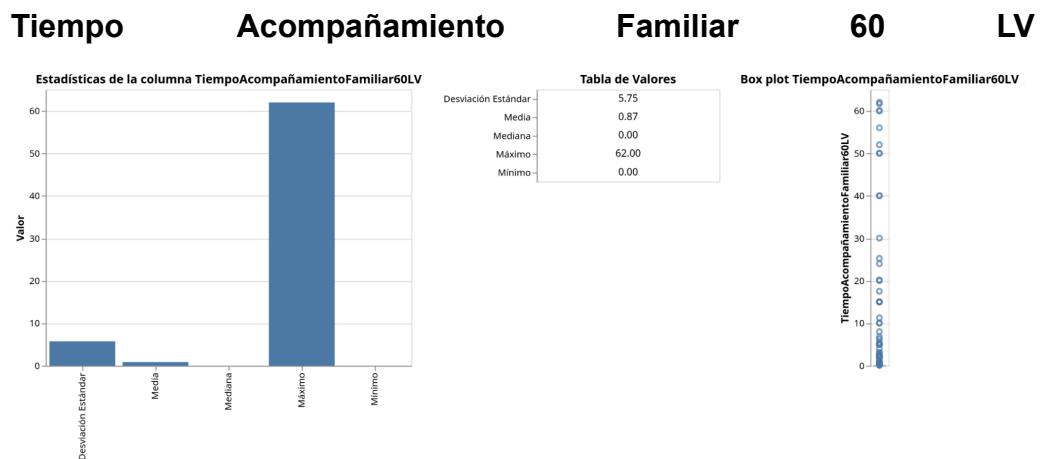


La gráfica, la tabla y el box plot evidencian la actividad de acompañamiento familiar entre semana que muestra un patrón polarizado: una mayoría que no participa en esta actividad y una minoría que dedica mucho tiempo, probablemente debido a circunstancias particulares, lo que puede deberse a la desigualdad en la distribución de actividades parentales. Este análisis puede ayudar a identificar necesidades y políticas de apoyo para quienes tienen responsabilidades familiares significativas.



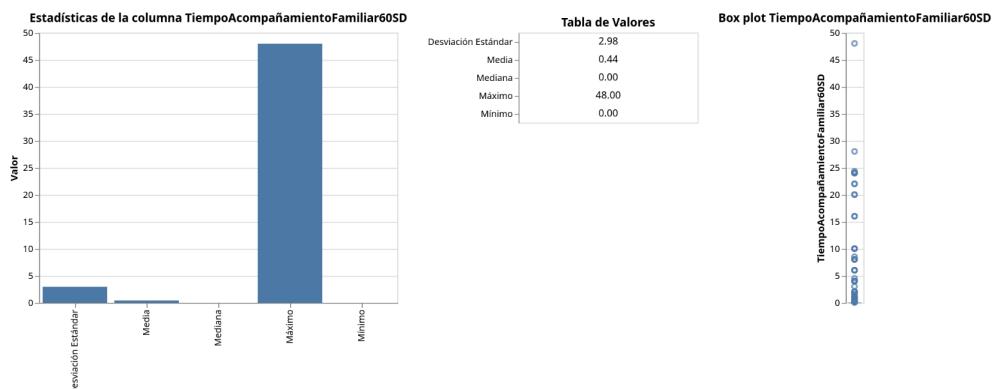
La gráfica, la tabla y el box plot corresponden a la columna

**TiempoAcompañamientoFamiliar15a59SD**, que mide el tiempo dedicado al acompañamiento de familiares (de 15 a 59 años) durante el fin de semana. El tiempo dedicado al acompañamiento de familiares de 15 a 59 años durante el fin de semana es mínimo para la mayoría de los encuestados, con algunas excepciones significativas, lo que puede deberse a la desigualdad en la distribución de actividades parentales. Este análisis puede ser útil para entender dinámicas familiares y planificar apoyos específicos para personas con mayores responsabilidades de cuidado.



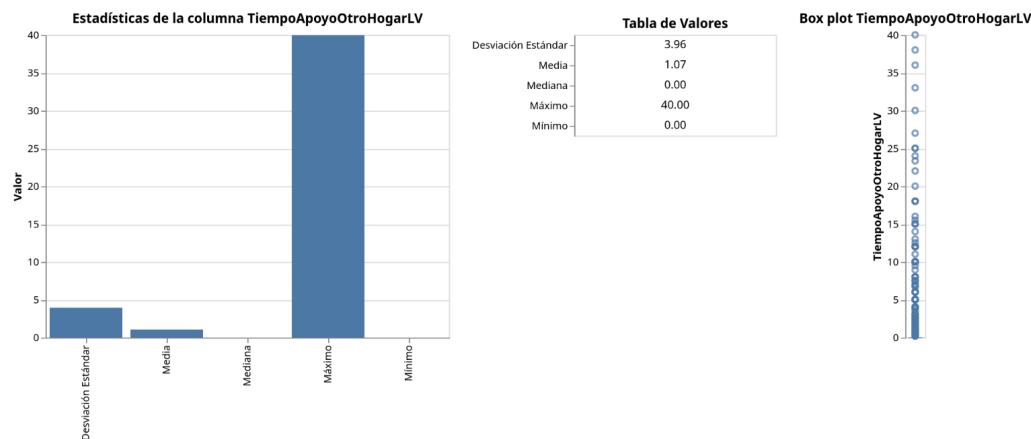
La gráfica, la tabla y el box plot corresponden a la columna **TiempoAcompañamientoFamiliar60LV**, que mide el tiempo dedicado al acompañamiento de familiares mayores de 60 años entre semana. La actividad de acompañamiento a familiares mayores de 60 años entre semana es mínima para la mayoría, con algunos casos que dedican tiempo significativo, probablemente por responsabilidades familiares específicas. Este análisis es útil para diseñar políticas de apoyo o intervenciones dirigidas a cuidadores.

## Tiempo Acompañamiento Familiar 60 SD



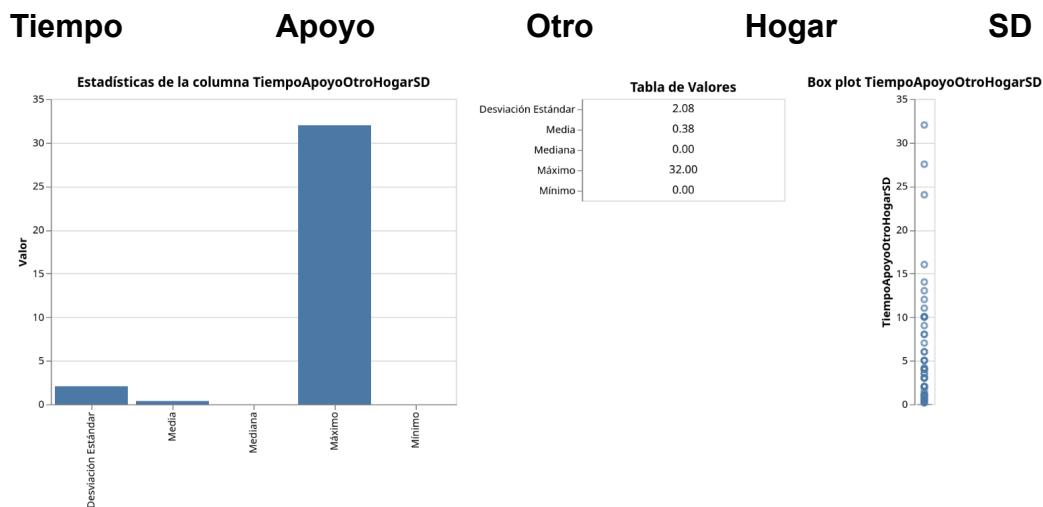
La gráfica, la tabla y el box plot corresponden a la columna **TiempoAcompañamientoFamiliar60SD**, que mide el tiempo dedicado al acompañamiento de familiares mayores de 60 años durante los fines de semana. Aunque la dedicación al acompañamiento de familiares mayores de 60 años durante los fines de semana es mayor que entre semana, sigue siendo baja para la mayoría. Sin embargo, los valores extremos sugieren la existencia de subgrupos que podrían necesitar atención especial.

## Tiempo Apoyo Otro Hogar LV

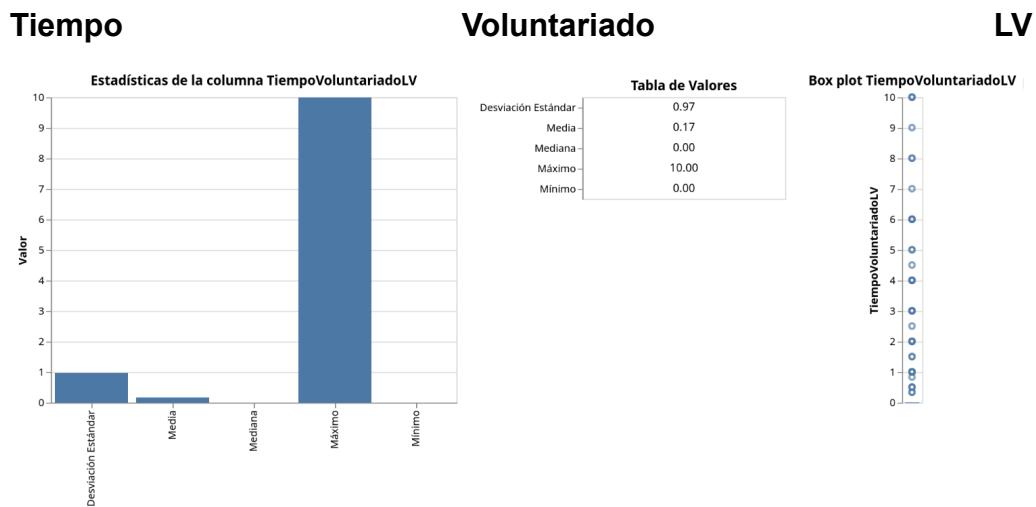


La gráfica, la tabla y el box plot corresponden a la columna **TiempoApoyoOtroHogarLV** (Entre semana). El tiempo dedicado a actividades de apoyo en otro hogar entre semana es limitado para la mayoría, pero hay casos extremos que reflejan compromisos

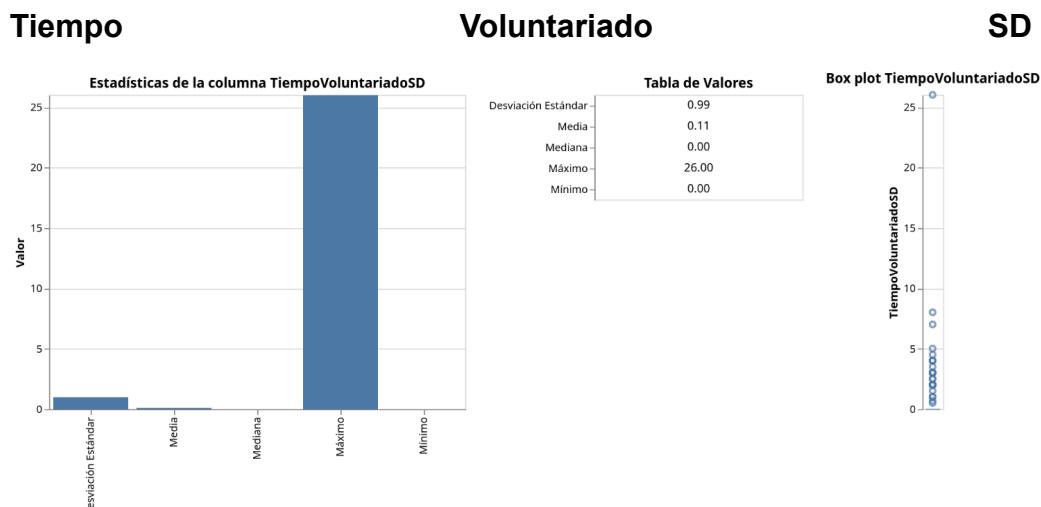
significativos, lo que refleja que hay sectores de la población que podrían requerir atención específica y que hay personas en la familia que los apoya pero que el tiempo entre otros miembros de la familia no es compartido equitativamente, lo que refleja que hay situaciones que podrían mejorarse entre la familia, la sociedad y el gobierno para atender las necesidades de estas personas.



La gráfica, la tabla y el box plot corresponden a la columna **TiempoApoyoOtroHogarSD** (Fin de semana). El tiempo dedicado a actividades de apoyo en otro hogar durante los fines de semana es generalmente bajo, con casos aislados que reportan valores elevados. Este comportamiento puede estar asociado a circunstancias específicas, lo que sugiere que hay familias que por sí solas no pueden ganarse la vida y necesitan de apoyo por parte de los familiares y que este apoyo no es brindado equitativamente por todos los miembros y solo una o pocas personas apoyan a este hogar.

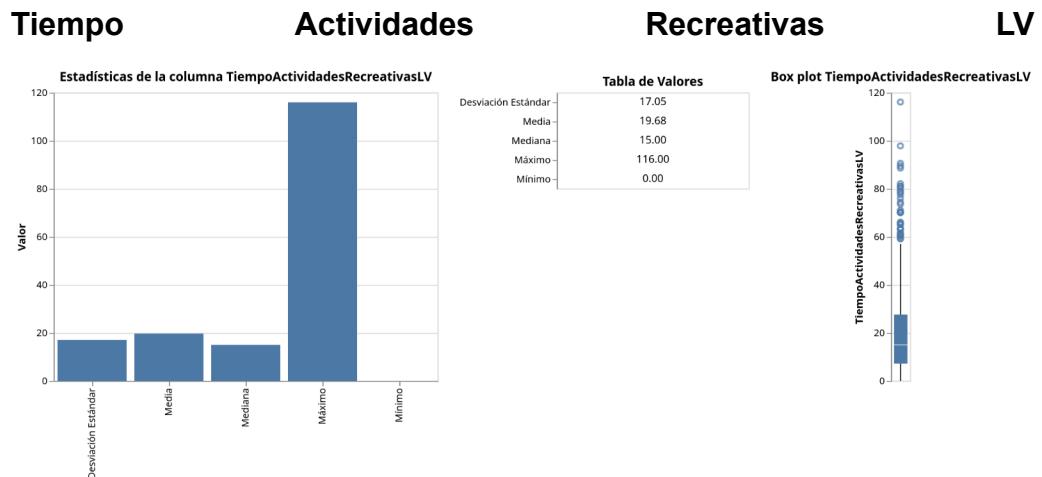


La gráfica, la tabla y el box plot muestran a la columna **TiempoVoluntariadoLV** (entre semana) que corresponde a la dedicación al voluntariado entre semana es mínima en promedio, aunque algunos casos muestran un fuerte compromiso. Este comportamiento podría reflejar limitaciones de tiempo por las obligaciones laborales o personales o falta de interés, pero también presenta una oportunidad para fomentar actividades de voluntariado mediante programas adaptados. En el caso opuesto (alta dedicación) refleja que hay personas que les gusta implicarse en estas actividades.

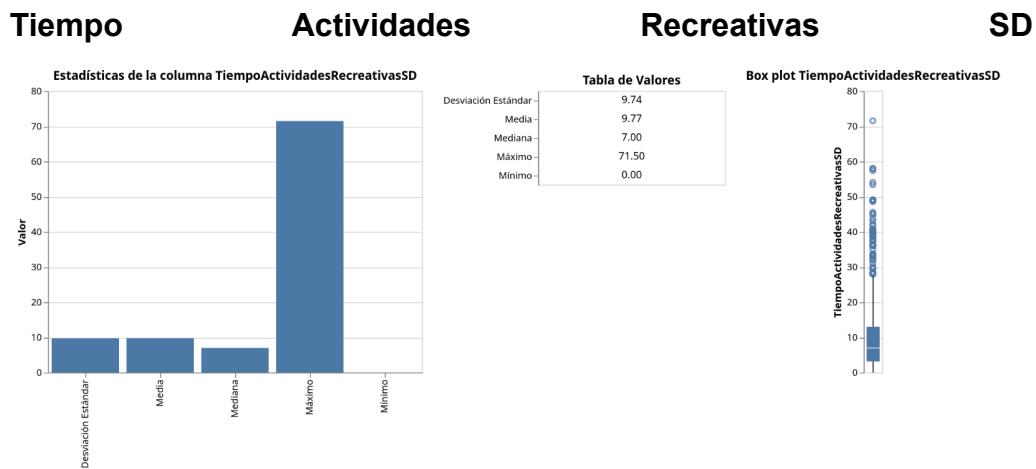


La gráfica, la tabla y el box plot presentan la columna **TiempoActividadesRecreativasLV** (entre semana). El tiempo dedicado a

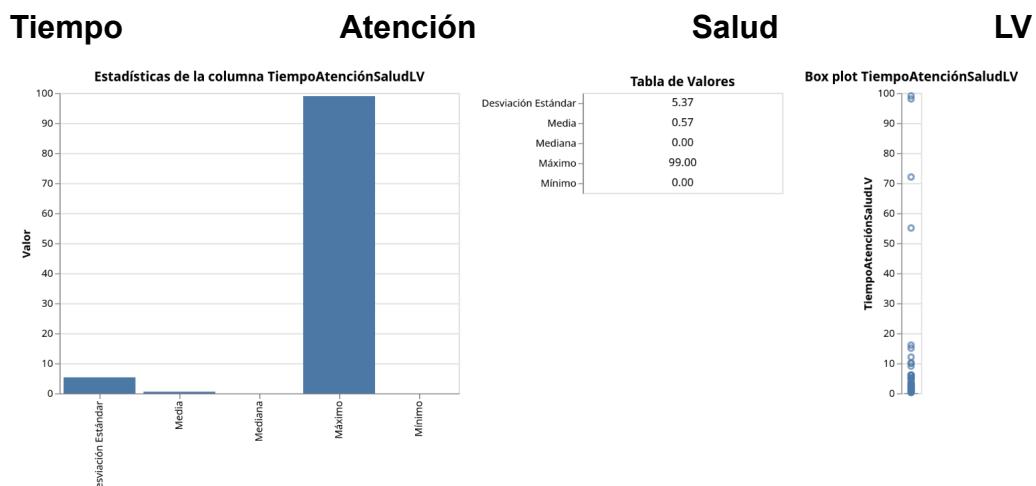
actividades recreativas entre semana es significativo para muchas personas, con una mediana de 15 horas semanales. Sin embargo, la amplia variabilidad refleja diferencias en prioridades y estilos de vida. Promover el equilibrio entre trabajo y recreación podría beneficiar a quienes no reportan tiempo en estas actividades.



La gráfica, la tabla y el box plot presentan la columna **TiempoActividadesRecreativasLV** (entre semana), en la cual se puede visualizar que existe una gran diversidad en el tiempo que las personas dedican a actividades recreativas entre semana. Aunque la mayoría se encuentra en un rango moderado de tiempo (alrededor de la media de 19.68 horas), hay casos particulares que dedican mucho más tiempo, posiblemente porque tienen una baja carga laboral o tienen mucho tiempo libre. En el caso en el que las actividades recreativas son mínimas o nulas es posible que se deba a que tienen uno o múltiples trabajos altamente demandantes, o que realizan diferentes actividades que determinan como prioridad antes que dedicar tiempo a la recreación.

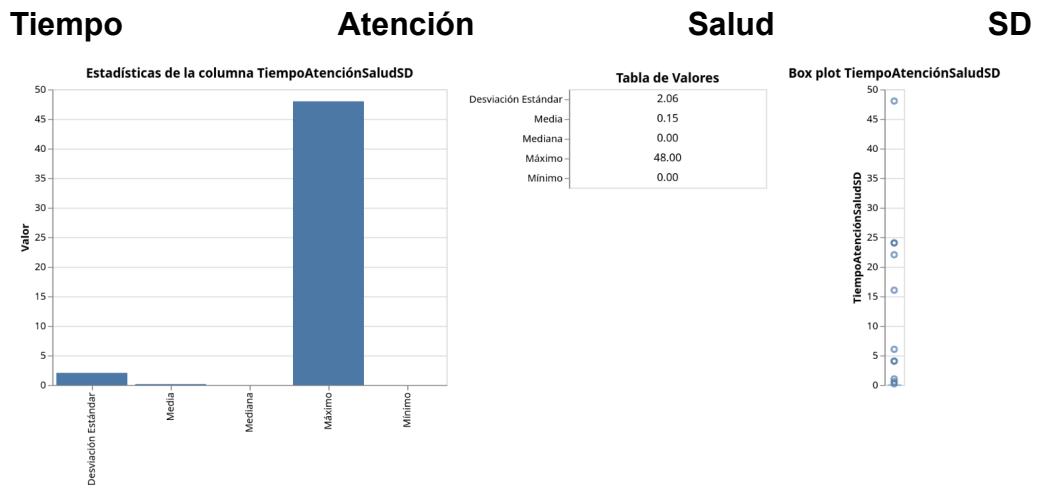


La gráfica, la tabla y el box plot se exponen los datos de la columna **TiempoActividadesRecreativasSD** (fin de semana) en la que se identifica que el tiempo dedicado a actividades recreativas durante el fin de semana presentan una considerable diversidad en el tiempo dedicado a actividades recreativas. Aunque el promedio está cerca de 10 horas, hay personas que no realizan estas actividades (0 horas) y otras que las consideran una prioridad significativa (71.50 horas), posiblemente a que hay personas que tienen mucho tiempo libre o en el caso opuesto que realizan muchas actividades y no tienen tiempo de recreación.



La gráfica, la tabla y el box plot se exponen los datos de la columna **TiempoAtenciónSalud** entre semana, en el cual se puede observar

que la mayoría de las personas no destinan tiempo entre semana a la atención de su salud (mediana de 0 horas), posiblemente por falta de necesidad, tiempo disponible, o acceso a servicios de salud.



La gráfica, la tabla y el box plot se exponen los datos de la columna **TiempoAtenciónSaludSD** (fin de semana). La mayoría de las personas no dedican tiempo a la atención de la salud durante los fines de semana, lo que podría ser debido a actividades diferentes de salud en estos días o falta de necesidad. Los valores altos pueden corresponder a casos particulares de atención médica prolongada, que no son representativos del comportamiento general.

## Sentimiento

## Quehaceres

## Hogar

Estadísticas de la columna **SentimientoQuehaceresHogar**

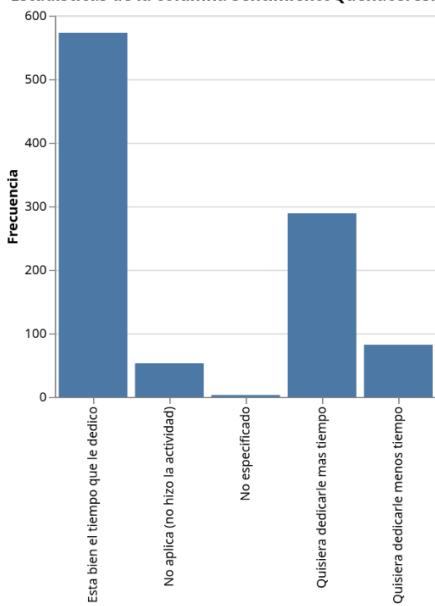
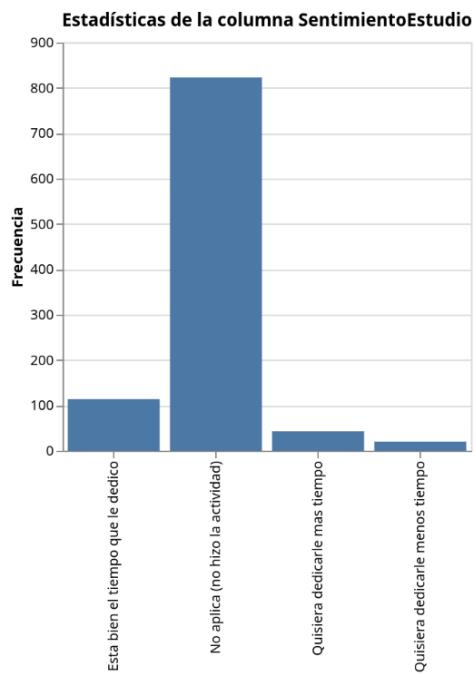


Tabla de Frecuencias

Esta bien el tiempo que le dedico	573
No aplica (no hizo la actividad)	53
No especificado	3
Quisiera dedicarle mas tiempo	289
Quisiera dedicarle menos tiempo	82

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoQuehaceresHogar**, que captura cómo se sienten las personas respecto al tiempo que dedican a los quehaceres del hogar. La mayoría de las personas están satisfechas, pero también hay un número significativo (289) que desearía dedicar más tiempo, mientras que un menor grupo (82) desearía dedicar menos, lo anterior se puede deber a que puede existir desigualdad en el reparto de las actividades del hogar o que haya personas que se dedican exclusivamente a ello y desean que la carga de trabajo disminuya en el caso quienes desean dedicar menos tiempo. En la situación opuesta en la que se desea dedicar más tiempo, posiblemente sea porque no hay alguna persona que apoye en ese sentido al hogar y desean mejorar la limpieza u otras tareas requeridas propias de los quehaceres.

## Sentimiento



## Estudio

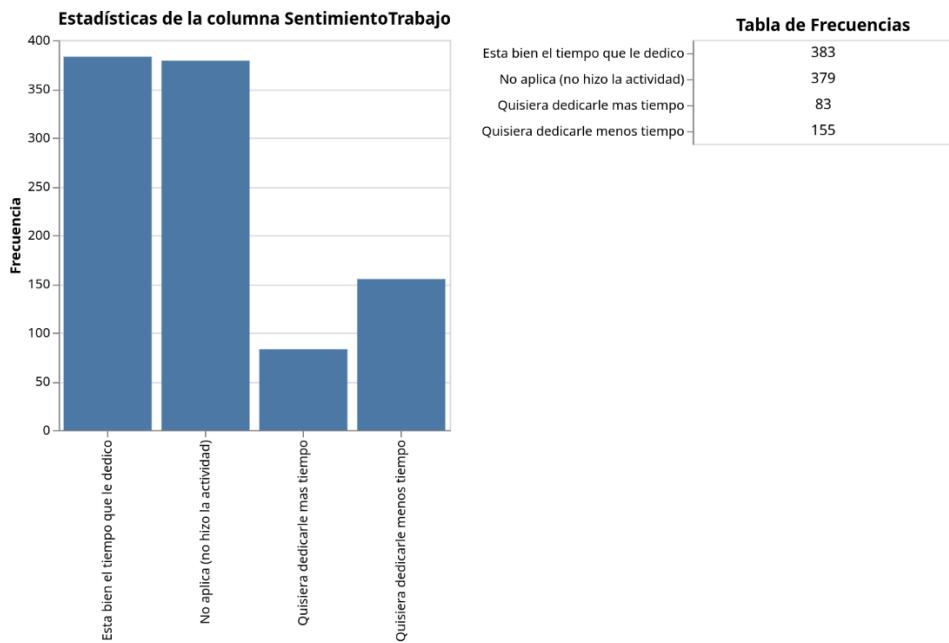
**Tabla de Frecuencias**

Esta bien el tiempo que le dedico	114
No aplica (no hizo la actividad)	823
Quisiera dedicarle mas tiempo	43
Quisiera dedicarle menos tiempo	20

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoEstudio**, que refleja cómo se sienten las personas respecto al tiempo que dedican a actividades de estudio. La categoría dominante es "No aplica (no hizo la actividad)" (823), lo que sugiere que la mayoría de las personas no están realizando actividades de estudio.

## Sentimiento

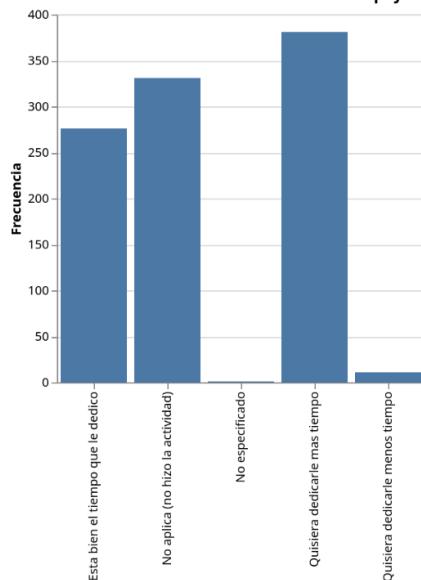
## Trabajo



La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoTrabajo**, que refleja cómo se sienten las personas respecto al tiempo que dedican a trabajar. Las categorías "Está bien el tiempo que le dedico" y "No aplica" tienen una cantidad similar de respuestas (~383 y ~379), indicando que el dataset cuenta con diferentes grupos de personas entre quienes trabajan y están satisfechos, que trabajan y desean dedicarle menos y quienes no trabajan.

## Sentimiento Cuidado Apoyo Familiares

Estadísticas de la columna SentimientoCuidadoApoyoFamiliares



## Apoyo

## Familiares

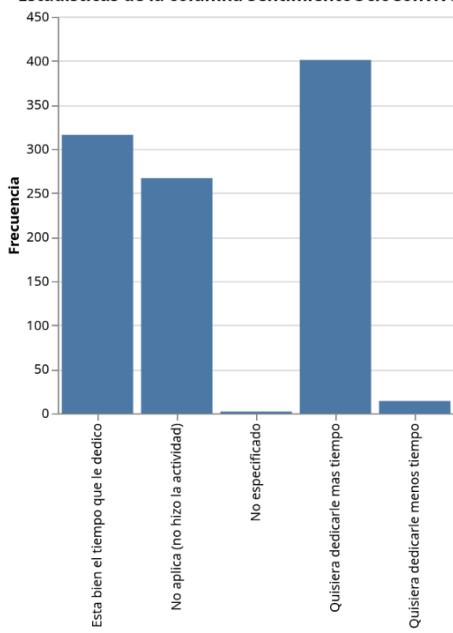
Tabla de Frecuencias

Esta bien el tiempo que le dedico	276
No aplica (no hizo la actividad)	331
No especificado	1
Quisiera dedicarle mas tiempo	381
Quisiera dedicarle menos tiempo	11

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoCuidadoApoyoFamiliares**, que refleja cómo se sienten las personas respecto al tiempo que dedican al cuidado o apoyo a familiares. La categoría dominante es "Quisiera dedicarle más tiempo" (381), seguida de "No aplica" (331), lo que indica que existe una necesidad percibida de dedicar más tiempo a esta actividad entre quienes participan en ella, además de existir personas que no realizan esta actividad debido a que no apoyan a familiares o no requieren apoyo sus familiares y por ello no aplica.

## Sentimiento

Estadísticas de la columna SentimientoOcioConvivencia



## Ocio

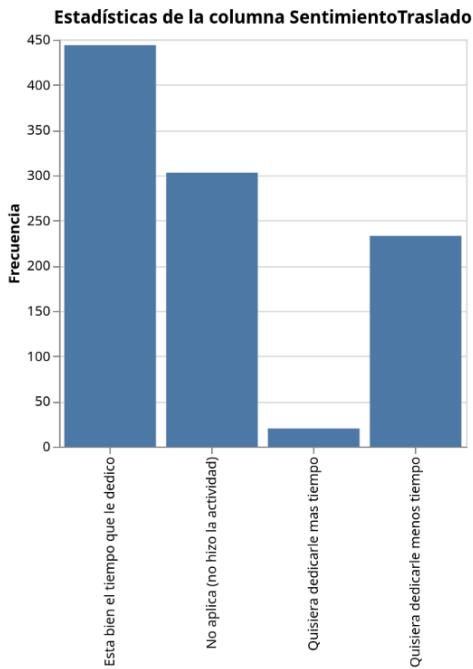
## Convivencia

Tabla de Frecuencias

Esta bien el tiempo que le dedico	316
No aplica (no hizo la actividad)	267
No especificado	2
Quisiera dedicarle mas tiempo	401
Quisiera dedicarle menos tiempo	14

La categoría dominante es "Quisiera dedicarle más tiempo" (381), seguida de "No aplica" (331), lo que indica que existe una necesidad percibida de dedicar más tiempo a esta actividad entre quienes participan en ella. Sin embargo, un grupo significativo no realiza esta actividad, posiblemente por falta de tiempo o realizar demasiadas responsabilidades que no les es posible disfrutar en este ámbito.

## Sentimiento



## Traslado

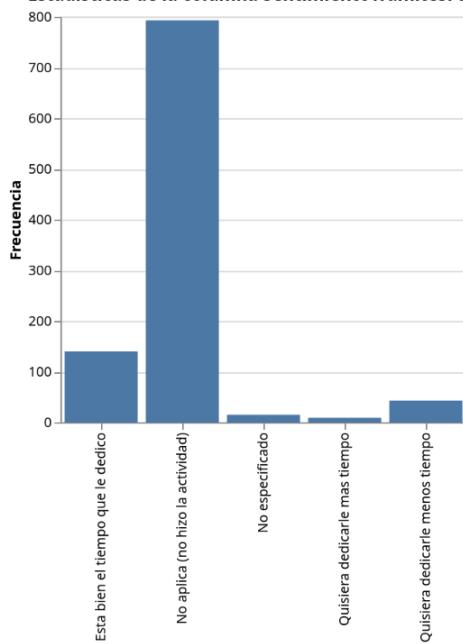
**Tabla de Frecuencias**

Esta bien el tiempo que le dedico	444
No aplica (no hizo la actividad)	303
Quisiera dedicarle mas tiempo	20
Quisiera dedicarle menos tiempo	233

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoTraslado**, que refleja cómo se sienten las personas respecto al tiempo que dedican a sus traslados (por ejemplo, de casa al trabajo, escuela, etc.). La mayoría de las personas está satisfecha con el tiempo que dedica a traslados, pero un grupo notable (233) considera que estos consumen demasiado tiempo posiblemente porque tienen sus actividades principales muy lejos de su vivienda.

## Sentimiento

Estadísticas de la columna **SentimientoTramitesPagos**



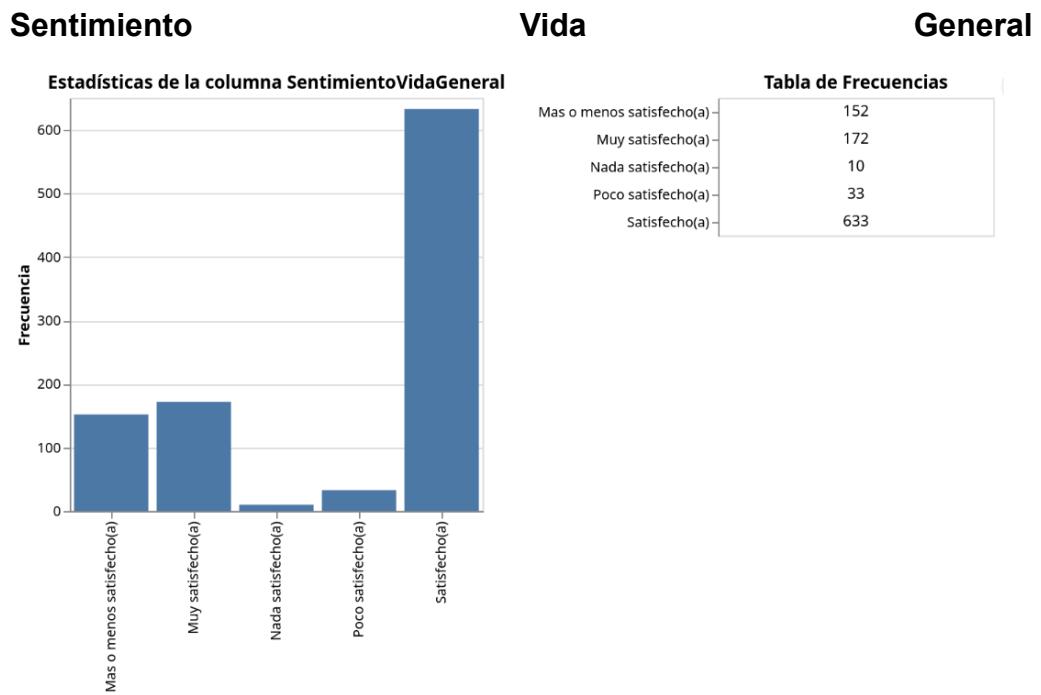
## Trámites

## Pagos

Tabla de Frecuencias

Esta bien el tiempo que le dedico	140
No aplica (no hizo la actividad)	793
No especificado	15
Quisiera dedicarle mas tiempo	9
Quisiera dedicarle menos tiempo	43

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoTramitesPagos**, que refleja cómo se sienten las personas respecto al tiempo que dedican a realizar trámites y pagos. La mayoría de las personas no participaron en actividades de trámites o pagos, mientras que un grupo más pequeño indicó satisfacción con el tiempo dedicado. Sin embargo, un porcentaje reducido siente que los trámites y pagos toman demasiado tiempo, lo que podría señalar que en los sitios donde realizan esa tarea son muy inefficientes o tienen muchos usuarios, lo que representa áreas de mejora en términos de eficiencia para los lugares en los que se realizan dichas actividades. Los valores opuestos reportados posiblemente se deban a que solo unas cuantas personas realizan esta actividad y es por ello esta diferencia.



La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoVidaGeneral**, que mide el nivel de satisfacción general de las personas con su vida. La mayoría de las personas tienen una percepción positiva de su calidad de vida, con más del 80% reportando satisfacción o alta satisfacción. Sin embargo, un pequeño grupo experimenta insatisfacción, lo que podría indicar áreas específicas de mejora en aspectos personales, laborales o sociales.

## Sentimiento

## Vida

## Familiar

Estadísticas de la columna **SentimientoVidaFamiliar**

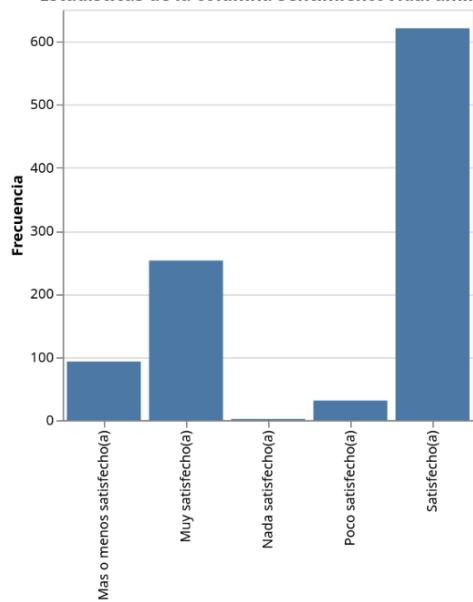


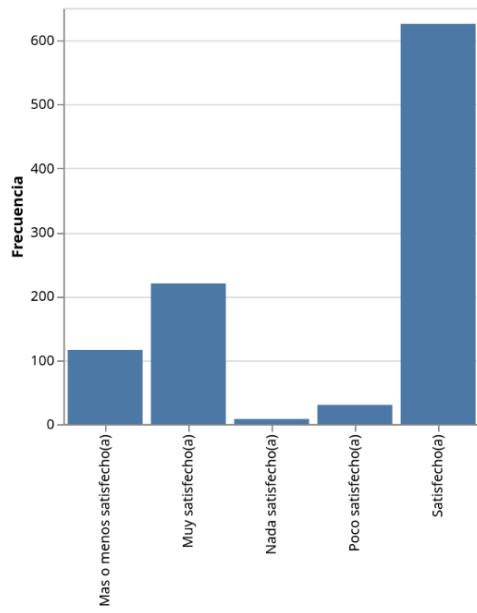
Tabla de Frecuencias

Mas o menos satisfecho(a)	93
Muy satisfecho(a)	253
Nada satisfecho(a)	2
Poco satisfecho(a)	31
Satisfecho(a)	621

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoVidaFamiliar**, que refleja cómo se sienten las personas respecto a su vida familiar. La mayoría de las personas reporta una alta satisfacción con su vida familiar, con muy pocos casos de insatisfacción. Esto podría reflejar una percepción positiva mayoritaria de las relaciones familiares en el grupo analizado, aunque también hay casos en los que puedan existir problemas familiares en los que requieran apoyo psicológico o problemas irreconciliables.

## Sentimiento

Estadísticas de la columna **SentimientoVidaAfectiva**



## Vida

## Afectiva

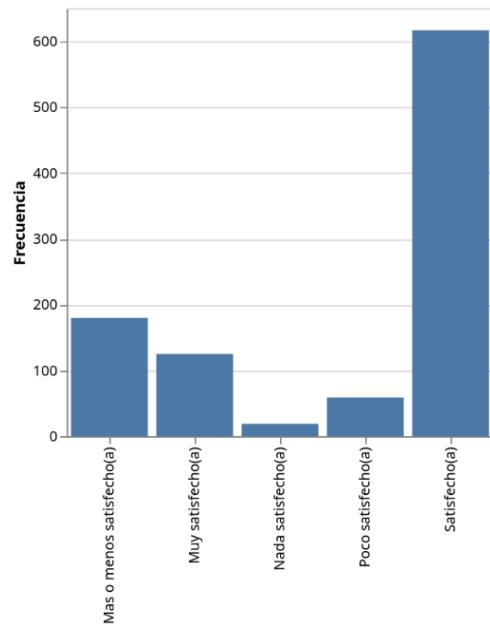
Tabla de Frecuencias

Mas o menos satisfecho(a)	116
Muy satisfecho(a)	220
Nada satisfecho(a)	8
Poco satisfecho(a)	30
Satisfecho(a)	626

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoVidaAfectiva**, que refleja cómo se sienten las personas respecto a su vida afectiva (relaciones emocionales o sentimentales). La mayoría de las personas reporta estar satisfechas o muy satisfechas con su vida afectiva, con una pequeña fracción indicando insatisfacción. Esto sugiere una percepción predominantemente positiva en el ámbito emocional y que en el caso de no estar satisfecho se esté pasando por algún problema particular que provoque insatisfacción.

## Sentimiento

Estadísticas de la columna **SentimientoVidaSocial**



## Vida

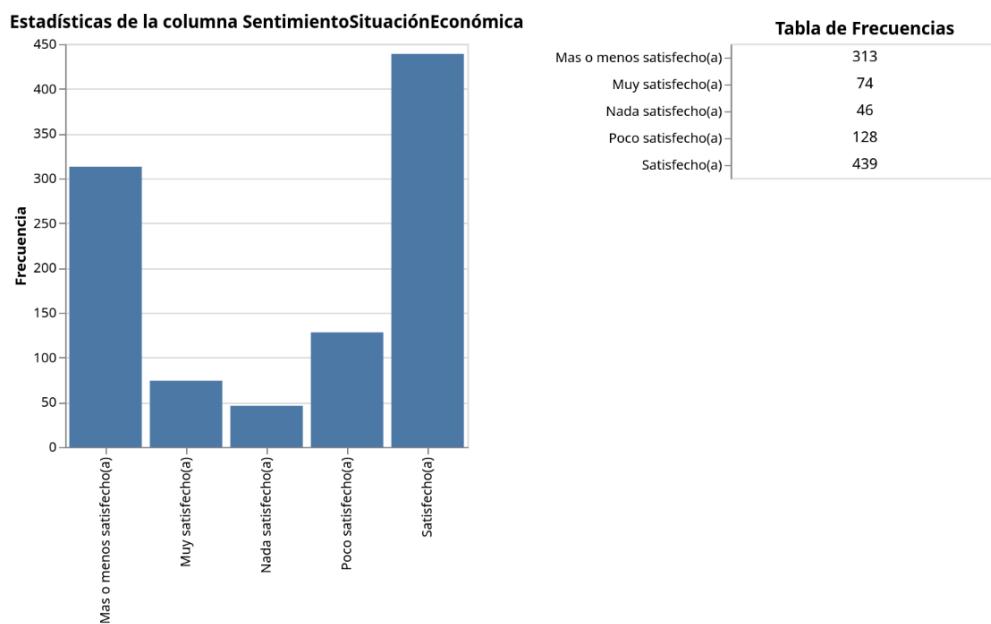
Tabla de Frecuencias

Mas o menos satisfecho(a)	180
Muy satisfecho(a)	125
Nada satisfecho(a)	19
Poco satisfecho(a)	59
Satisfecho(a)	617

## Social

La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoVidaSocial**, que refleja cómo se sienten las personas respecto a su vida social. La mayoría de las personas reporta satisfacción con su vida social, con solo un pequeño porcentaje que experimenta insatisfacción. Este resultado refleja una percepción positiva en términos de interacción y relaciones sociales. En el caso de estar "nada satisfecho" es posible que la persona presente pocas habilidades sociales o prefiera realizar actividades en solitario.

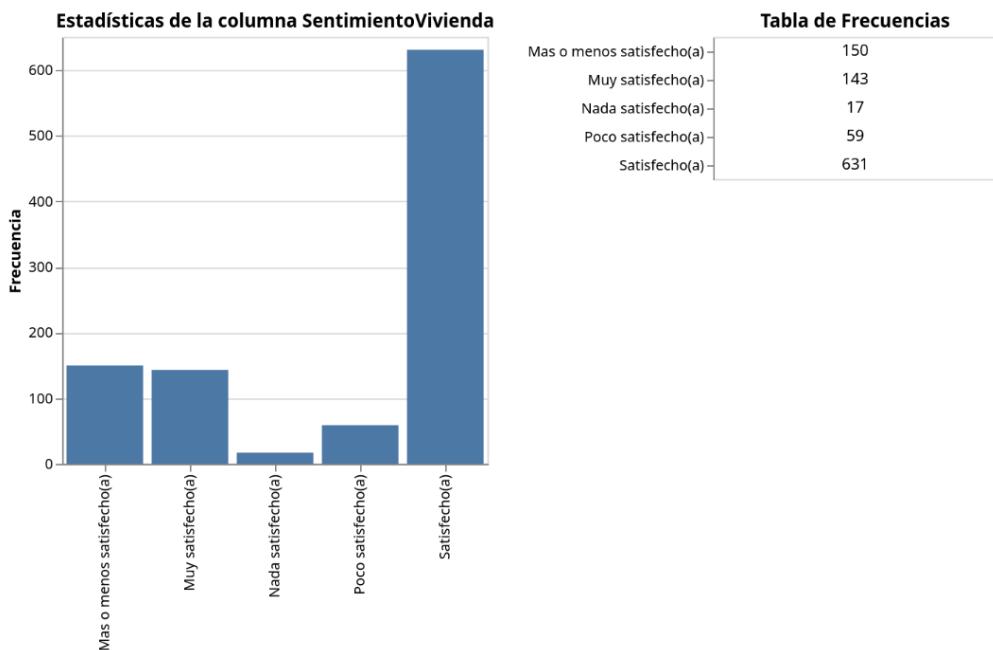
## Sentimiento Situación Económica



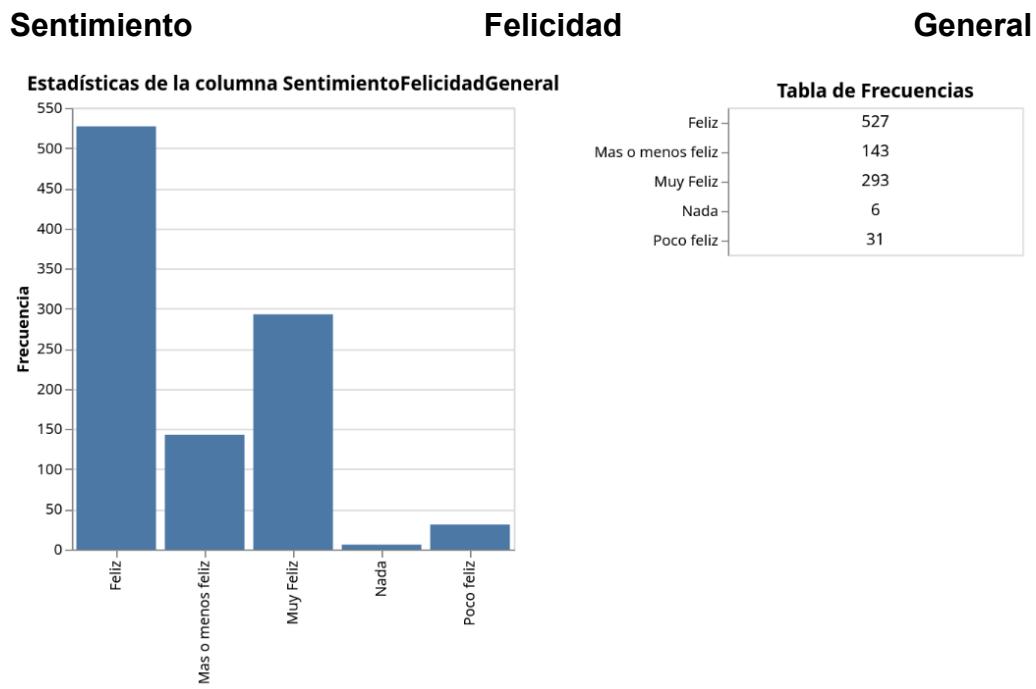
La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoSituaciónEconómica**, reflejando cómo las personas perciben su satisfacción respecto a su situación económica. Aunque la mayoría de las personas reporta satisfacción con su situación económica, hay un grupo significativo (174 personas) que indica insatisfacción. Este resultado sugiere desigualdades económicas o desafíos financieros en ciertos sectores.

## Sentimiento

## Vivienda



La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoVivienda**, que refleja cómo se sienten las personas respecto a su situación de vivienda. La mayoría de las personas reportan satisfacción con su situación de vivienda, mientras que un grupo pequeño expresa insatisfacción. Este resultado destaca una percepción general positiva, aunque existen áreas específicas que podrían necesitar atención en problemas de vivienda.



La gráfica y la tabla de frecuencias muestran información sobre la columna **SentimientoFelicidadGeneral**, que refleja cómo las personas perciben su nivel general de felicidad. La mayoría de las personas reporta sentirse felices o muy felices, con solo un pequeño porcentaje indicando infelicidad. Este resultado sugiere una percepción predominantemente positiva sobre la felicidad general en el grupo analizado. Las categorías negativas ("Poco feliz" y "Nada feliz") son numéricamente muy reducidas, mostrando que solo una minoría significativa reporta insatisfacción, lo que podría deberse a situaciones especiales que requieren atención posiblemente asociado a factores críticos como problemas de salud, económicos o sociales.

## ii y iv. Identificación de valores faltantes y fuera de dominio

Como se mostró en la limpieza de datos los valores faltantes fueron reemplazados por 0 o en el caso de las columnas del dataset original p5-7 eliminados.

## iii. Identificación de atributos correlacionadas

A continuación se enlistan los atributos que por su naturaleza guardan una correlación directa:

- TiempoTrabajoLV y TiempoTrabajoSD
- TiempoTrasladoLV y TiempoTrasladoSD
- TiempoDedicadoADormirLV y TiempoDedicadoADormirSD
- TiempoDedicadoAComerLV y TiempoDedicadoAComerSD
- TiempoDedicadoAAseoPersonalLV y  
TiempoDedicadoAAseoPersonalSD
- TiempoDedicadoAAActividadesEstudioLV y  
TiempoDedicadoAAActividadesEstudioSD
- TiempoDedicadoATrasladarEstudianteLV y  
TiempoDedicadoATrasladarEstudianteSD
- QuehaceresDomesticosLV y QuehaceresDomesticosSD
- TiempoGanaderiaAutoconsumoLV y  
TiempoGanaderiaAutoconsumoSD
- TiempoActividadesRecoleccionLV y  
TiempoActividadesRecoleccionSD
- TiempoConfeccionCaseraLV y TiempoConfeccionCaseraSD
- TiempoMantenimientoHogarLV y TiempoMantenimientoHogarSD
- TiempoComprasDelHogarLV y TiempoComprasDelHogarSD
- TiempoTramitesPagosLV y TiempoTramitesPagosSD
- TiempoOrganizacionHogarLV y TiempoOrganizacionHogarSD
- TiempoMantenimientoHogarTercerosLV y  
TiempoMantenimientoHogarTercerosSD
- TiempoCuidadosEspecialesOtraPersonaLV y  
TiempoCuidadosEspecialesOtraPersonaSD
- TiempoCrianzaMenores0a5LV y TiempoCrianzaMenores0a5SD
- TiempoCrianzaMenores0a14LV y  
TiempoCrianzaMenores0a14SD
- TiempoAcompañamientoFamiliar15a59LV y  
TiempoAcompañamientoFamiliar15a59SD

- TiempoAcompañamientoFamiliar60LV y  
TiempoAcompañamientoFamiliar60SD
- TiempoApoyoOtroHogarLV y TiempoApoyoOtroHogarSD
- TiempoActividadesRecreativasLV y  
TiempoActividadesRecreativasSD
- TiempoAtencionSaludLV y TiempoAtencionSaludSD

### **3. Procesamiento de datos**

#### **a. Integración de los datos**

##### **i. Fusión de datos**

Solo contamos con un conjunto de datos.

##### **ii. Adición de datos**

Solo contamos con un conjunto de datos.

#### **b. Selección de los datos**

##### **i. Creación de conjuntos específicos para cada tarea de descubrimiento de datos**

###### **Naive-Bayes**

###### **Conjunto de datos seleccionado:**

En este caso, se ha decidido utilizar todos los atributos disponibles en el conjunto de datos para la tarea de clasificación, ya que cada uno de ellos puede aportar información relevante para predecir la "Felicidad General". Naive Bayes funciona bien cuando se utilizan todos los atributos, siempre y cuando estos no sean excesivamente redundantes o estén demasiado correlacionados entre sí. El modelo evaluará la probabilidad de cada clase a partir de las combinaciones de estos atributos.

**Tabla con variables seleccionadas**

Variable	Tipo de Variable	Descripción	Rango/Categorías	Justificación de Selección
Entidad	Numérica (entera)	Código numérico que representa una entidad federativa.	1: Aguascalientes 2: Baja California Sur ... 32: Zacatecas	Puede influir en la felicidad general debido a diferencias socioeconómicas y culturales entre entidades.
Sexo	Categórica	Sexo de la persona.	['Mujer' 'Hombre']	El género puede estar relacionado con la felicidad general debido a roles de género y expectativas sociales.
Edad	Numérica (entera)	Edad de la persona en años.	12 - 95	La edad puede influir en la felicidad general debido a las diferentes etapas de la vida y las prioridades cambiantes.
TrabajoAlMenos1Hora	Categórica	Indica si la persona trabajó al menos una hora.	['No' 'Sí']	El trabajo puede ser una fuente de satisfacción o estrés, lo que afecta la felicidad general.
TiempoTrabajolV	Numérica (continua)	Tiempo dedicado al trabajo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 96.0	La cantidad de tiempo dedicado al trabajo puede influir en el equilibrio entre la vida laboral y personal, afectando la felicidad.
TiempoTrabajosoD	Numérica (continua)	Tiempo dedicado al trabajo en un día de fin de semana promedio en sábado y domingo (en horas).	0.0 - 48.0	Similar a TiempoTrabajolV, pero en fines de semana.
TiempoTrasladolV	Numérica (continua)	Tiempo dedicado al traslado al trabajo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 40.0	Los tiempos de traslado largos pueden afectar negativamente el bienestar y la felicidad.

TiempoTrasladoSD	Numérica (continua)	Tiempo dedicado al traslado al trabajo en un día de fin de semana promedio (en horas).	0.0 - 40.0	Similar a TiempoTrasladoLV, pero en fines de semana.
Salario	Numérica (continua)	Salario mensual de la persona.	0.0 - 780000.0	El nivel de ingresos puede influir en la satisfacción con la vida y la felicidad general.
PermisoEnfermedadMaternidadAccidente	Categórica	Indica si la persona tiene derecho a permiso por enfermedad, maternidad o accidente.	['No especificado' 'Sí' 'No']	La seguridad laboral y los beneficios pueden contribuir a la sensación de seguridad y bienestar.
PermisoVacaciones	Categórica	Indica si la persona tiene derecho a vacaciones.	['No especificado' 'Sí' 'No']	El acceso a vacaciones puede influir en el equilibrio entre la vida laboral y personal y la felicidad.
DerechoJubilación	Categórica	Indica si la persona tiene derecho a jubilación.	['No especificado' 'Sí' 'No']	La seguridad financiera a largo plazo puede afectar la felicidad general.
DerechoServiciosMedicos	Categórica	Indica si la persona tiene derecho a servicios médicos.	['No especificado' 'Sí' 'No']	El acceso a la atención médica puede influir en la salud y el bienestar general.
TiempoDedicadoADormirLV	Numérica (continua)	Tiempo dedicado a dormir en un día laboral promedio de lunes a viernes (en horas).	2.0 - 76.0	La cantidad y calidad del sueño son cruciales para la salud física y mental, que a su vez afectan la felicidad.
TiempoDedicadoADormirSD	Numérica (continua)	Tiempo dedicado a dormir en un día de fin de semana promedio (en horas).	0.0 - 35.0	Similar a TiempoDedicadoADormirLV, pero en fines de semana.

TiempoDedicadoAComerLV	Numérica (continua)	Tiempo dedicado a comer en un día laboral promedio de lunes a viernes (en horas). 0.0 - 77.5	Puede reflejar hábitos alimenticios y tiempo para el cuidado personal, que pueden influir en la felicidad.
TiempoDedicadoAComerSD	Numérica (continua)	Tiempo dedicado a comer en un día de fin de semana promedio de sábado y domingo(en horas). 0.0 - 45.0	Similar a TiempoDedicadoAComerLV, pero en fines de semana.
TiempoDedicadoAAseoPersonalLV	Numérica (continua)	Tiempo dedicado al aseo personal en un día laboral promedio de lunes a viernes(en horas). 0.0833333333333333 - 21.0	El tiempo dedicado al cuidado personal puede reflejar la autoestima y el bienestar.
TiempoDedicadoAAseoPersonalSD	Numérica (continua)	Tiempo dedicado al aseo personal en un día de fin de semana promedio en sábado y domingo (en horas). 0.0 - 20.0	Similar a TiempoDedicadoAAseoPersonalLV, pero en fines de semana.
TiempoDedicadoAActividadesEstudioLV	Numérica (continua)	Tiempo dedicado a actividades de estudio en un día laboral promedio de lunes a viernes (en horas). 0.0 - 80.0	La educación y el desarrollo personal pueden contribuir a la felicidad.
TiempoDedicadoAActividadesEstudioSD	Numérica (continua)	Tiempo dedicado a actividades de estudio en un día de fin de semana promedio (en horas). 0.0 - 20.0	Similar a TiempoDedicadoAActividadesEstudioLV, pero en fines de semana.

TiempoDedicadoATrasladarEstudianteLV	Numérica (continua)	Tiempo dedicado a trasladar a un estudiante en un día laboral promedio de lunes a viernes (en horas).	0.0 - 15.0	Las responsabilidades familiares pueden influir en el tiempo disponible y la felicidad.
TiempoDedicadoATrasladarEstudianteSD	Numérica (continua)	Tiempo dedicado a trasladar a un estudiante en un día de fin de semana promedio (en horas).	0.0 - 3.0	Similar a TiempoDedicadoATrasladarEstudianteLV, pero en fines de semana.
QuehaceresDomésticosLV	Numérica (continua)	Tiempo dedicado a quehaceres domésticos en un día laboral promedio de lunes a viernes (en horas).	0.0 - 84.58333333333333	- La distribución de las tareas domésticas puede influir en la satisfacción y el bienestar.
QuehaceresDomésticosSD	Numérica (continua)	Tiempo dedicado a quehaceres domésticos en un día de fin de semana promedio (en horas).	0.0 - 46.35	Similar a QuehaceresDomésticosLV, pero en fines de semana.
TiempoGanaderíaAutoconsumoLV	Numérica (continua)	Tiempo dedicado a la ganadería para autoconsumo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 35.0	Puede reflejar un estilo de vida o actividad que influye en la felicidad.
TiempoGanaderíaAutoconsumoSD	Numérica (continua)	Tiempo dedicado a la ganadería para autoconsumo en un día de fin de semana promedio (en horas).	0.0 - 18.0	Similar a TiempoGanaderíaAutoconsumoLV, pero en fines de semana.

TiempoActividadesRecolecciónLV	Numérica (continua)	Tiempo dedicado a actividades de recolección en un día laboral promedio de lunes a viernes (en horas).	0.0 - 20.0	Puede reflejar un estilo de vida o actividad que influye en la felicidad.
TiempoActividadesRecolecciónSD	Numérica (continua)	Tiempo dedicado a actividades de recolección en un día de fin de semana promedio (en horas).	0.0 - 8.0	Similar a TiempoActividadesRecolecciónLV, pero en fines de semana.
TiempoAgriculturaAutoconsumoLV	Numérica (continua)	Tiempo dedicado a la agricultura para autoconsumo en un día laboral promedio de lunes a viernes (en horas).	0.0 - 18.0	Puede reflejar un estilo de vida o actividad que influye en la felicidad.
TiempoAgriculturaAutoconsumoSD	Numérica (continua)	Tiempo dedicado a la agricultura para autoconsumo en un día de fin de semana promedio (en horas).	0.0 - 5.0	Similar a TiempoAgriculturaAutoconsumoLV, pero en fines de semana.
TiempoConfecciónCaseraLV	Numérica (continua)	Tiempo dedicado a la confección casera en un día laboral promedio de lunes a viernes (en horas).	0.0 - 25.0	Puede reflejar un pasatiempo o actividad que influye en la felicidad.
TiempoConfecciónCaseraSD	Numérica (continua)	Tiempo dedicado a la confección casera en un día de fin de semana promedio (en horas).	0.0 - 8.0	Similar a TiempoConfecciónCaseraLV, pero en fines de semana.

TiempoMantenimientoHogarLV	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar en un día laboral promedio de lunes a viernes (en horas). 0.0 - 28.0	Puede reflejar responsabilidades y tiempo dedicado al hogar, que pueden influir en la felicidad.
TiempoMantenimientoHogarSD	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar en un día de fin de semana promedio (en horas). 0.0 - 15.0	Similar a TiempoMantenimientoHogarLV, pero en fines de semana.
TiempoComprasDelHogarLV	Numérica (continua)	Tiempo dedicado a las compras del hogar en un día laboral promedio de lunes a viernes (en horas). 0.0 - 13.0	Puede reflejar responsabilidades y tiempo dedicado al hogar, que pueden influir en la felicidad.
TiempoComprasDelHogarSD	Numérica (continua)	Tiempo dedicado a las compras del hogar en un día de fin de semana promedio (en horas). 0.0 - 10.0	Similar a TiempoComprasDelHogarLV, pero en fines de semana.
TiempoTramitesPagosLV	Numérica (continua)	Tiempo dedicado a trámites y pagos en un día laboral promedio de lunes a viernes (en horas). 0.0 - 6.0	Puede reflejar el tiempo dedicado a tareas administrativas, que pueden influir en el estrés y la felicidad.
TiempoTramitesPagosSD	Numérica (continua)	Tiempo dedicado a trámites y pagos en un día de fin de semana promedio (en horas). 0.0 - 2.0	Similar a TiempoTramitesPagosLV, pero en fines de semana.

TiempoOrganizaciónHogarLV	Numérica (continua)	Tiempo dedicado a la organización del hogar en un día laboral promedio de lunes a viernes (en horas).	0.0 - 7.0	Puede reflejar responsabilidades y tiempo dedicado al hogar, que pueden influir en la felicidad.
TiempoOrganizaciónHogarSD	Numérica (continua)	Tiempo dedicado a la organización del hogar en un día de fin de semana promedio (en horas).	0.0 - 2.0	Similar a TiempoOrganizaciónHogarLV, pero en fines de semana.
TiempoMantenimientoHogarTercerosLV	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar por terceros en un día laboral promedio de lunes a viernes (en horas).	0.0 - 50.0	Puede reflejar el acceso a servicios y la delegación de tareas, que pueden influir en el tiempo libre y la felicidad.
TiempoMantenimientoHogarTercerosSD	Numérica (continua)	Tiempo dedicado al mantenimiento del hogar por terceros en un día de fin de semana promedio (en horas).	0.0 - 30.0	Similar a TiempoMantenimientoHogarTercerosLV, pero en fines de semana.
TiempoCuidadosEspecialesOtraPersonaLV	Numérica (continua)	Tiempo dedicado a cuidados especiales de otra persona en un día laboral promedio de lunes a viernes (en horas).	0.0 - 81.83333333333333 33	- Las responsabilidades de cuidado pueden influir en el tiempo disponible y la felicidad.

TiempoCuidadosEspecialesOtraPersonaSD	Numérica (continua)	Tiempo dedicado a cuidados especiales de otra persona en un día de fin de semana promedio (en horas).	0.0 - 46.0	Similar a TiempoCuidadosEspecialesOtraPersonaLV, pero en fines de semana.
TiempoCrianzaMenores0a5LV	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 5 años en un día laboral promedio (en horas).	0.0 - 85.0	Las responsabilidades de crianza pueden influir en el tiempo disponible y la felicidad.
TiempoCrianzaMenores0a5SD	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 5 años en un día de fin de semana promedio (en horas).	0.0 - 33.0	Similar a TiempoCrianzaMenores0a5LV, pero en fines de semana.
TiempoCrianzaMenores0a14LV	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 14 años en un día laboral promedio (en horas).	0.0 - 103.0	Las responsabilidades de crianza pueden influir en el tiempo disponible y la felicidad.
TiempoCrianzaMenores0a14SD	Numérica (continua)	Tiempo dedicado a la crianza de menores de 0 a 14 años en un día de fin de semana promedio (en horas).	0.0 - 39.0	Similar a TiempoCrianzaMenores0a14LV, pero en fines de semana.
TiempoAcompañamientoFamiliar15a59LV	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 15 a 59 años en un día laboral promedio (en horas).	0.0 - 60.0	El tiempo dedicado a la familia y las relaciones puede influir en la felicidad.

TiempoAcompañamientoFamiliar15a59SD	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 15 a 59 años en un día de fin de semana promedio (en horas).	0.0 - 24.0	Similar a TiempoAcompañamientoFamiliar15a59LV, pero en fines de semana.
TiempoAcompañamientoFamiliar60LV	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 60 años o más en un día laboral promedio (en horas).	0.0 - 62.0	El tiempo dedicado a la familia y las relaciones puede influir en la felicidad.
TiempoAcompañamientoFamiliar60SD	Numérica (continua)	Tiempo dedicado al acompañamiento familiar de personas de 60 años o más en un día de fin de semana promedio (en horas).	0.0 - 48.0	Similar a TiempoAcompañamientoFamiliar60LV, pero en fines de semana.
TiempoApoyoOtroHogarLV	Numérica (continua)	Tiempo dedicado a apoyar a otro hogar en un día laboral promedio de lunes a viernes (en horas).	0.0 - 40.0	El apoyo a otros puede influir en la felicidad y el sentido de comunidad.
TiempoApoyoOtroHogarSD	Numérica (continua)	Tiempo dedicado a apoyar a otro hogar en un día de fin de semana promedio (en horas).	0.0 - 32.0	Similar a TiempoApoyoOtroHogarLV, pero en fines de semana.
TiempoVoluntariadoLV	Numérica (continua)	Tiempo dedicado al voluntariado en un día laboral promedio de lunes a viernes (en horas).	0.0 - 10.0	El voluntariado y la participación comunitaria pueden contribuir a la felicidad.

TiempoVoluntariadoSD	Numérica (continua)	Tiempo dedicado al voluntariado en un día de fin de semana promedio (en horas).	0.0 - 26.0	Similar a TiempoVoluntariadoLV, pero en fines de semana.
TiempoActividadesRecreativasLV	Numérica (continua)	Tiempo dedicado a actividades recreativas en un día laboral promedio de lunes a viernes (en horas).	0.0 - 116.0	El tiempo dedicado al ocio y la recreación es importante para el bienestar y la felicidad.
TiempoActividadesRecreativasSD	Numérica (continua)	Tiempo dedicado a actividades recreativas en un día de fin de semana promedio (en horas).	0.0 - 71.5	Similar a TiempoActividadesRecreativasLV, pero en fines de semana.
TiempoAtenciónSaludLV	Numérica (continua)	Tiempo dedicado a la atención de la salud en un día laboral promedio de lunes a viernes (en horas).	0.0 - 99.0	El tiempo dedicado al cuidado de la salud puede influir en el bienestar físico y mental.
TiempoAtenciónSaludSD	Numérica (continua)	Tiempo dedicado a la atención de la salud en un día de fin de semana promedio (en horas).	0.0 - 48.0	Similar a TiempoAtenciónSaludLV, pero en fines de semana.
SentimientoQuehaceresHogar	Categórica	['Quisiera dedicarle mas tiempo' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle menos tiempo' 'No aplica (no hizo la trabajo doméstico actividad)']	'No la felicidad.'	La satisfacción con la división del trabajo doméstico puede influir en la felicidad.

			[especificado']	
SentimientoE studio	Categórica	Sentimiento general sobre el estudio.	['Quisiera dedicarle mas tiempo' 'No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle menos tiempo']	La satisfacción con el tiempo dedicado al estudio puede influir en la felicidad.
SentimientoTr abajo	Categórica	Sentimiento general sobre el trabajo.	['No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle menos tiempo' 'Quisiera dedicarle mas tiempo']	La satisfacción con el trabajo es un factor importante en la felicidad general.
SentimientoC uidadoApoyo Familiares	Categórica	Sentimiento general sobre el cuidado y apoyo a familiares.	['Quisiera dedicarle mas tiempo' 'No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle menos tiempo' 'No dedicado a la familia puede influir en la felicidad. especificado']	La satisfacción con el tiempo dedicado a la familia puede influir en la felicidad.

			[ 'Esta bien el tiempo que le dedico' 'No aplica (no hizo la actividad)' 'Quisiera dedicarle mas tiempo'	
SentimientoOcioConvivencia	Categórica	Sentimiento general sobre el ocio y la convivencia.	'Quisiera dedicarle La satisfacción con el tiempo menos tiempo' 'No dedicado al ocio y la convivencia especificado']	social puede influir en la felicidad.
SentimientoTraslado	Categórica	Sentimiento general sobre el traslado.	[ 'Esta bien el tiempo que le dedico' 'No aplica (no hizo la actividad)' 'Quisiera dedicarle mas tiempo' ]	La satisfacción con el tiempo dedicado al traslado puede influir en la felicidad.
SentimientoTramitesPagos	Categórica	Sentimiento general sobre los trámites y pagos.	[ 'No aplica (no hizo la actividad)' 'Esta bien el tiempo que le dedico' 'Quisiera dedicarle mas tiempo' 'Quisiera dedicarle menos tiempo' ]	La satisfacción con el tiempo dedicado a los trámites puede influir en la felicidad.
SentimientoVidaGeneral	Categórica	Sentimiento general sobre la vida en general.	[ 'Satisfecho(a)' 'Poco satisfecho(a)' 'Mas o menos satisfecho(a)' 'Muy satisfecho(a)' 'Nada' ]	Una medida general de satisfacción con la vida.

			satisfecho(a)']	
SentimientoVidaFamiliar	Categórica	Sentimiento general sobre la vida familiar.	['Muy satisfecho(a)' 'Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Poco satisfecho(a)' 'Nada satisfecho(a)']	La satisfacción con la vida familiar es un factor importante en la felicidad general.
SentimientoVidaAfectiva	Categórica	Sentimiento general sobre la vida afectiva.	['Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Poco satisfecho(a)' 'Muy satisfecho(a)' 'Nada satisfecho(a)']	La satisfacción con la vida afectiva es un factor importante en la felicidad general.
SentimientoVidaSocial	Categórica	Sentimiento general sobre la vida social.	['Mas o menos satisfecho(a)' 'Satisfecho(a)' 'Muy satisfecho(a)' 'Nada satisfecho(a)' 'Poco satisfecho(a)']	La satisfacción con la vida social es un factor importante en la felicidad general.
SentimientoSituaciónEconómica	Categórica	Sentimiento general sobre la situación económica.	['Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Poco satisfecho(a)' 'Muy satisfecho(a)' 'Nada satisfecho(a)']	La satisfacción con la situación económica puede influir en la felicidad general.

SentimientoVivienda	Categórica	Sentimiento general sobre la vivienda.	[ 'Muy satisfecho(a)' 'Satisfecho(a)' 'Mas o menos satisfecho(a)' 'Nada satisfecho(a)' 'Poco satisfecho(a)']	La satisfacción con la vivienda puede influir en la felicidad general.
---------------------	------------	--	--	--

### K-means

**Nombre de Conjunto:** Horas Laboradas.

**Objetivo del conjunto de datos:** Identificar grupos de personas con jornadas laborales nulas o ligeras, promedio o excesivas.

**Restricción:** Pocas variables para evitar que el Silhouette Score disminuya debido a la inclusión de información irrelevante o redundante.

Nombre del atributo	Dominio	Justificación de Selección
TiempoTrabajoLV	0.0 - 96.0	Proporciona las horas laboradas de Lunes a Viernes
TiempoTrabajoSD	0.0 - 48.0	Proporciona las horas laboradas de Sábados y Domingos.
SentimientoTrabajo	[ Esta bien el tiempo que le dedicó, No aplica (no hizo la actividad),	Proporciona los sentimientos de las personas que trabajan.

	Quisiera dedicarle más tiempo, Quisiera dedicarle menos tiempo]	
--	---	--

**Nombre de Conjunto:** Horas de Descanso.

**Objetivo del conjunto de datos:** Identificar grupos de personas en base al tiempo que dedican de sueño.

**Restricción:** Pocas variables para evitar que el Silhouette Score disminuya debido a la inclusión de información irrelevante o redundante.

Nombre del atributo	Dominio	Justificación de Selección
TiempoDedicadoADormirLV	4 - 80	Proporciona las horas de sueño de Lunes a Viernes
TiempoDedicadoADormirSD	0 - 26	Proporciona las horas de sueño de Sábados y Domingos.

**Nombre de Conjunto:** Horas de Traslado.

**Objetivo del conjunto de datos:** Identificar grupos de personas en base al tiempo que dedican de traslado.

**Restricción:** Pocas variables para evitar que el Silhouette Score disminuya debido a la inclusión de información irrelevante o redundante.

Nombre del atributo	Dominio	Justificación de Selección
TiempoTrasladoLV	0 - 26	Proporciona las horas de traslado de Lunes a Viernes
TiempoTrasladoSD	0 - 30	Proporciona las horas de traslado de Sábados y Domingos.
SentimientoTraslado	[Esta bien el tiempo que le dedicó, No aplica (no hizo la actividad), Quisiera dedicarle más tiempo, Quisiera dedicarle menos tiempo]	Proporciona los sentimientos de las personas que se trasladan.

**Nombre de Conjunto:** Horas de Estudio.

**Objetivo del conjunto de datos:** Identificar grupos de personas en base al tiempo que dedican en sus estudios.

**Restricción:** Pocas variables para evitar que el Silhouette Score disminuya debido a la inclusión de información irrelevante o redundante.

Nombre del atributo	Dominio	Justificación de Selección
TiempoDedicadoAActividadesEstudioLV	0 - 80	Proporciona las horas de estudio de Lunes a Viernes

TiempoDedicadoAAActividadesEstudioSD	0 - 29	Proporciona las horas de estudio de Sábados y Domingos.
SentimientoEstudio	[Esta bien el tiempo que le dedicó, No aplica (no hizo la actividad), Quisiera dedicarle más tiempo, Quisiera dedicarle menos tiempo]	Proporciona los sentimientos de las personas que estudian.

**Nombre de Conjunto:** Horas de Actividades Domésticas.

**Objetivo del conjunto de datos:** Identificar grupos de personas en base al tiempo que dedican en sus actividades domésticas.

**Restricción:** Pocas variables para evitar que el Silhouette Score disminuya debido a la inclusión de información irrelevante o redundante.

Nombre del atributo	Dominio	Justificación de Selección
QuehaceresDomesticos LV	0.0 - 101.5	Proporciona las horas en actividades domésticas de Lunes a Viernes
QuehaceresDomesticos SD	0.0 - 45.d133	Proporciona las horas en actividades domésticas de Sábados

		y Domingos.
SentimientoQuehaceres Hogar	[Esta bien el tiempo que le dedicó, No aplica (no hizo la actividad), Quisiera dedicarle más tiempo, Quisiera dedicarle menos tiempo]	Proporciona los sentimientos de las personas que realizan actividades domésticas

**Nombre de Conjunto:** Situación Económica.

**Objetivo del conjunto de datos:** Identificar grupos de personas en base al salario y el sentimiento que poseen acerca de su situación económica.

**Restricción:** Pocas variables para evitar que el Silhouette Score disminuya debido a la inclusión de información irrelevante o redundante.

Nombre del atributo	Dominio	Justificación de Selección
Salario	0 - 80	Proporciona las horas de estudio de Lunes a Viernes
SentimientoSituacionEconomica	0 - 29	Proporciona las horas de estudio de Sábados y Domingos.

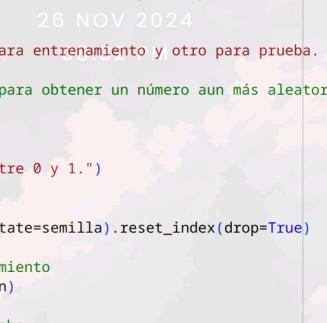
## ii. Selección de subconjuntos de datos (Muestreo)

### Naive-Bayes

Para evaluar el rendimiento del modelo Naive Bayes, dividiremos el conjunto de datos en dos subconjuntos: uno para el entrenamiento del modelo y otro para la evaluación del rendimiento. La división (como ya se había comentado en puntos anteriores) será de 70% de los datos para entrenamiento y 30% para prueba, asegurando que los conjuntos sean mutuamente excluyentes.

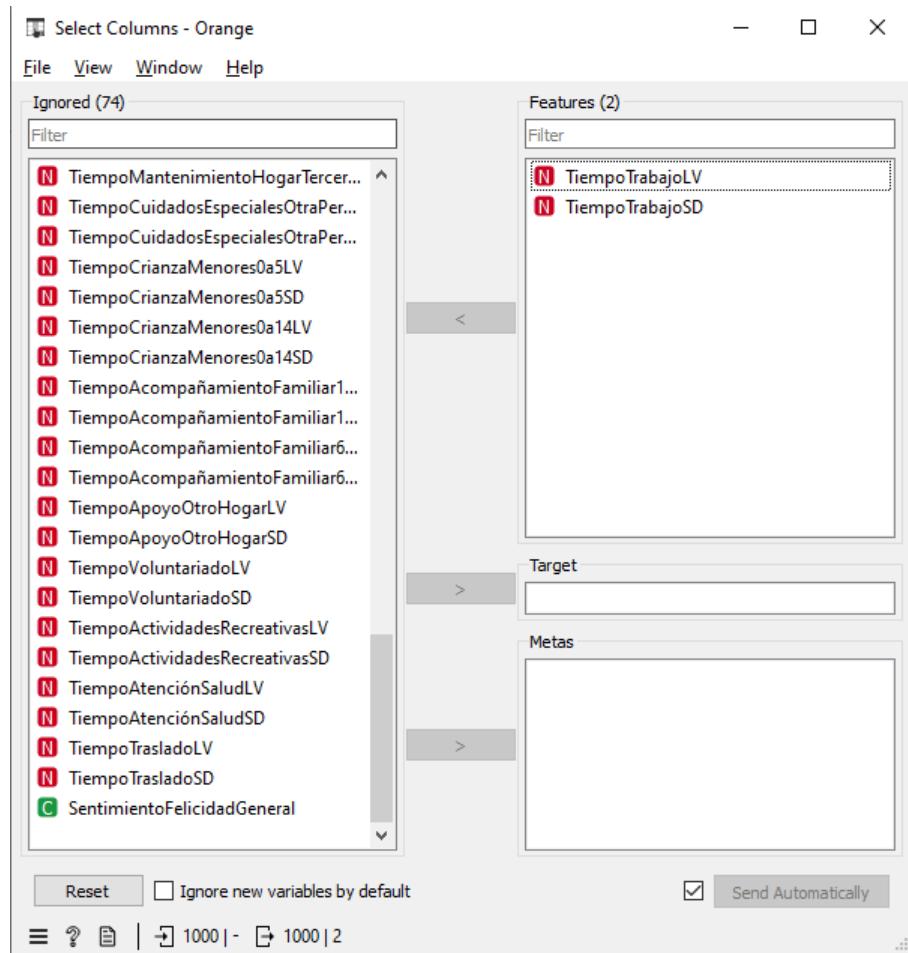
El código con el que se realiza la división se presenta a continuación

```
45 def generador_de_subconjuntos(datos: pd.DataFrame, fraccion: float, semilla: int):
46     """Divide el conjunto de datos en dos subconjuntos: entrenamiento y prueba de manera aleatoria.
47     Parámetros:
48         datos (pd.DataFrame): El conjunto de datos a dividir.
49         fraccion (float): La proporción del conjunto de datos a usar para el entrenamiento (entre 0 y 1).
50         semilla (int): La semilla para el generador de números aleatorios (para reproducibilidad).
51
52     Retorna:
53         pd.DataFrame, pd.DataFrame: Dos subconjuntos, uno para entrenamiento y otro para prueba.
54     """
55     # multiplicamos la semilla por un número aleatorio para obtener un número aun más aleatorio
56     semilla = semilla * random.randint(1, 1000)
57     try:
58         if not 0 < fraccion < 1:
59             raise ValueError("La fracción debe estar entre 0 y 1.")
60
61         # Barajar (shuffle) los datos para evitar sesgo
62         datos_barajados = datos.sample(frac=1, random_state=semilla).reset_index(drop=True)
63
64         # Calcular el tamaño del subconjunto de entrenamiento
65         tamano_entrenamiento = int(len(datos) * fraccion)
66
67         # Crear los subconjuntos de entrenamiento y prueba
68         datos_entrenamiento = datos_barajados[:tamano_entrenamiento]
69         datos_prueba = datos_barajados[tamano_entrenamiento:]
70
71         # Retornar los subconjuntos como una tupla
72         return datos_entrenamiento, datos_prueba
73
74     except Exception as e:
75         mensaje = f"Error al generar subconjuntos: {e}"
76         raise Exception(mensaje) from e
77
78
79
80
```



## K-Means

Debido a nuestro conjunto de datos, optamos por seleccionar sólo dos variables representativas para el problema principal, e iremos haciendo nuevos conjuntos. Seleccionamos las dos variables con la herramienta “Select Columns” en Orange, como se muestra a continuación.

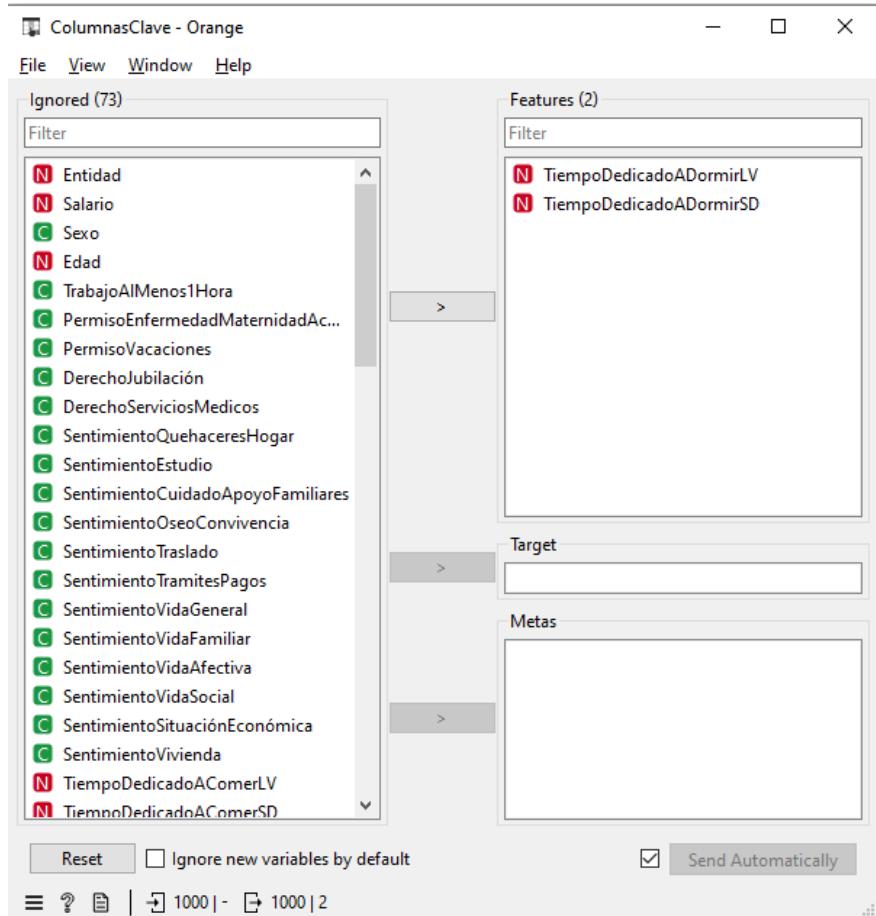


Selección de variables en la herramienta Select Columns en Orange

Data Table - Orange

	TiempoTrabajoLV	TiempoTrabajoSD
1	0	0
2	50	0
3	0	0
4	60	10
5	40	8
6	50	0
7	0	0
8	24	4
9	30	24
10	40	0
11	35	0
12	0	0
13	10	0
14	60	24
15	40	0
16	0	0
17	0	0
18	40	5
19	0	0
20	40	5
21	48	0
22	40	16
23	46.8333	0
24	40	16
25	0	0
26	40	9
27	0	0
28	40	6
29	0	6
30	15	0
31	32	16
32	0	0
33	0	0
34	75	24
35	0	0
36	32	0
37	50	10
38	37.5	7.5
39	0	0
40	0	0
41	40	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0

Resultado de la selección de variables en la herramienta Select Columns en Orange

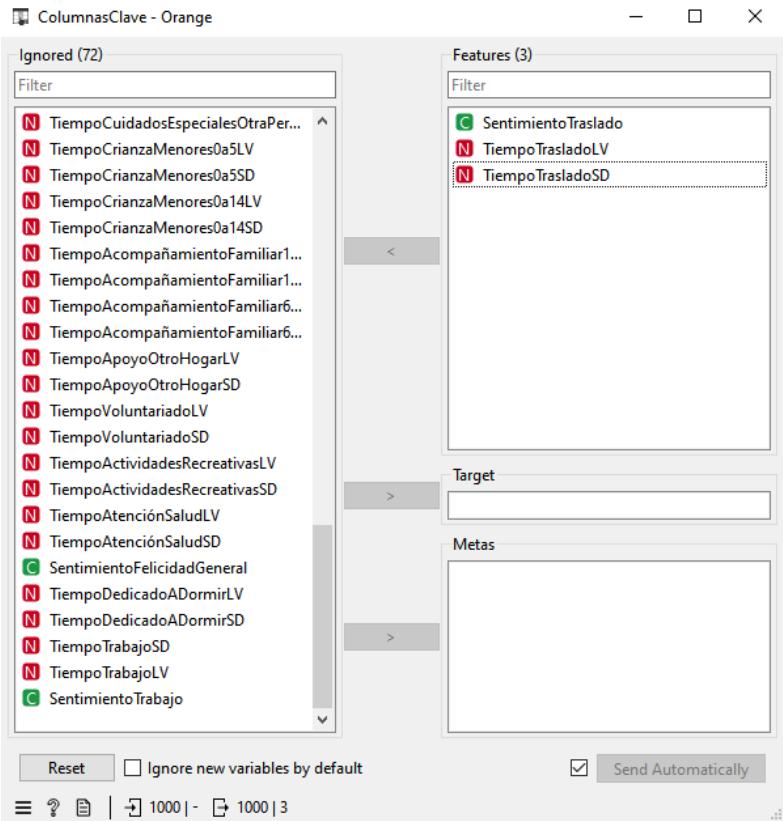


Selección de variables en la herramienta Select Columns en Orange

Info - Orange

	poDedicadoADorr	eoDedicadoADorr
35	80	32
17	60	24
39	60	24
150	60	24
241	60	21
588	60	24
723	60	17
812	60	24
883	60	24
884	60	24
918	56	20
221	55	22
296	55	22
314	55	18
406	55	0
437	55	20
618	55	17
851	55	16
970	55	22
1000	55	22
180	54	20
949	52.5	21
960	52	16
22	50	20
23	50	16
57	50	16
62	50	20
66	50	20
95	50	20
97	50	20
145	50	20
152	50	20
201	50	22
217	50	20
288	50	20
323	50	18
325	50	0
326	50	20
351	50	20
360	50	20
416	50	20
453	50	20

Resultado de la selección de variables en la herramienta Select Columns en Orange

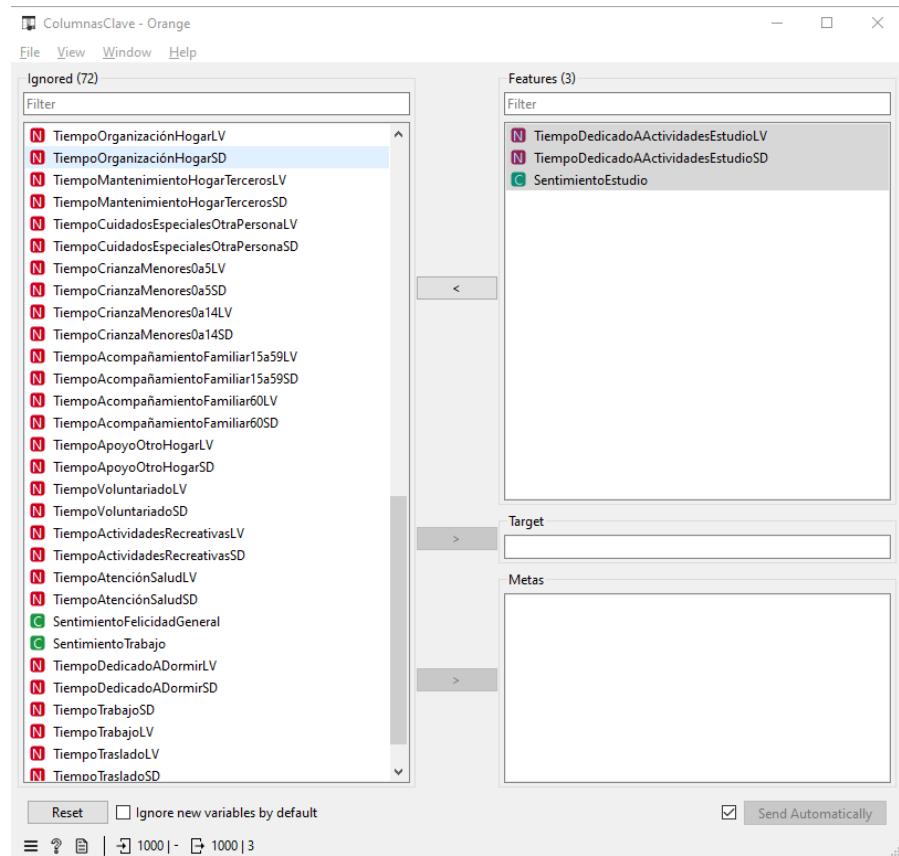


Selección de variables en la herramienta Select Columns en Orange

Info - Orange

	entimientoTraslad	TiempoTrasladoLV	TiempoTrasladoSD
1	Esta bien el tie...	2.5	0.5
2	Esta bien el tie...	1.66667	0.333333
3	Esta bien el tie...	0.333333	0.0666667
4	No aplica (no h...	0	0
5	Quisiera dedica...	0.5	0
6	Esta bien el tie...	10	2
7	Esta bien el tie...	5	0.5
8	No aplica (no h...	0	0
9	Esta bien el tie...	2	0
10	No aplica (no h...	0	0
11	No aplica (no h...	0	0
12	Quisiera dedica...	0	0
13	Esta bien el tie...	0.416667	0.166667
14	Esta bien el tie...	0	0
15	Quisiera dedica...	10	4
16	No aplica (no h...	0	0
17	No aplica (no h...	0	0
18	No aplica (no h...	0	0
19	Esta bien el tie...	0.0833333	0.0333333
20	No aplica (no h...	0	0
21	No aplica (no h...	0	0
22	No aplica (no h...	0	0
23	Quisiera dedica...	0	0
24	Esta bien el tie...	5	2
25	Quisiera dedica...	15	1.5
26	Esta bien el tie...	5	0
27	Esta bien el tie...	0.833333	0.333333
28	No aplica (no h...	0	0
29	Esta bien el tie...	10	2
30	Quisiera dedica...	5	1
31	Quisiera dedica...	0	0
32	No aplica (no h...	0	0
33	Quisiera dedica...	0	0.5
34	Esta bien el tie...	0.333333	0.0666667
35	No aplica (no h...	0	0
36	Esta bien el tie...	0	0
37	No aplica (no h...	0	0
38	Esta bien el tie...	5	1
39	Esta bien el tie...	1.33333	0
40	Esta bien el tie...	0	0
41	Quisiera dedica...	10	0
42	No aplica (no h...	0	0
43	No aplica (no h...	0	0
44	Quisiera dedica...	5	0
45	Esta bien el tie...	25	5
46	Quisiera dedica...	5	0
47	Esta bien el tie...	0	0
48	Esta bien el tie...	5	1
49	Esta bien el tie...	0	0
50	Esta bien el tie...	1	1

Resultado de la selección de variables en la herramienta Select Columns en Orange



Selección de variables en la herramienta Select Columns en Orange

Info - Orange

File Edit View Window Help

Info  
1000 instances (no missing data)  
3 features  
No target variable.  
No meta attributes.

Variables  
 Show variable labels (if present)  
 Visualize numeric values  
 Color by instance classes  
Selection  
 Select full rows

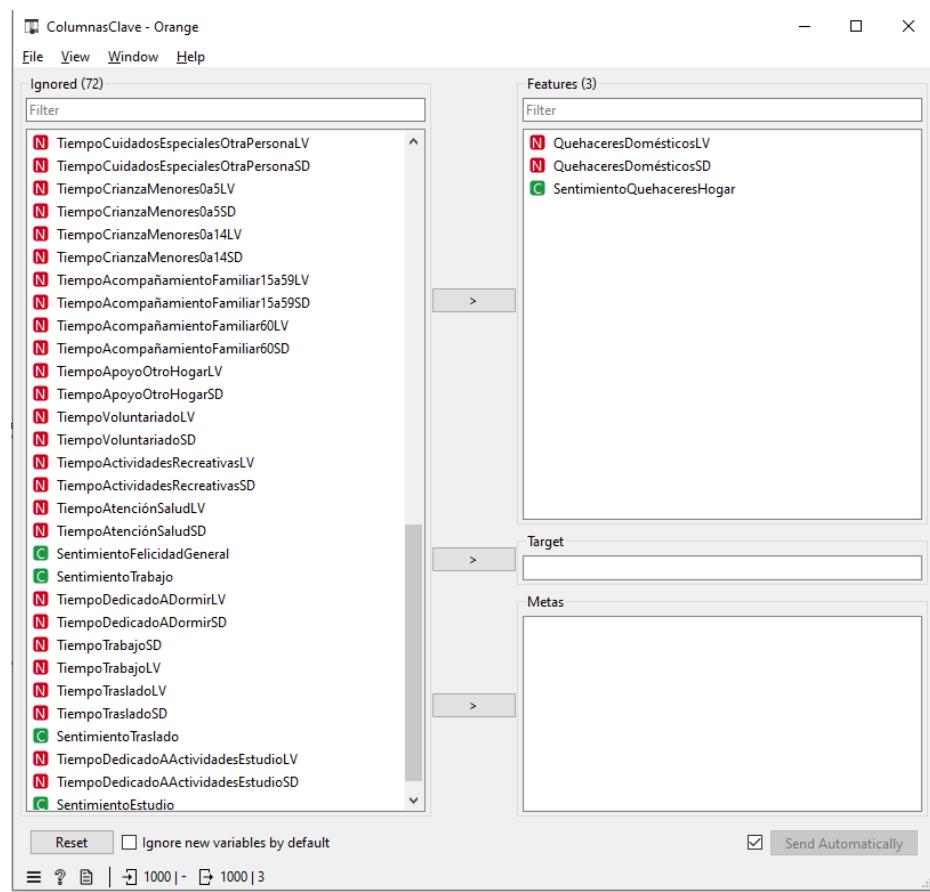
Restore Original Order  
 Send Automatically

icadoAAActividad1 icadoAAActividad2 entimientoEstudi1

1	0	0	No aplica (no h...
2	0	0	No aplica (no h...
3	0	0	No aplica (no h...
4	0	0	No aplica (no h...
5	0	0	No aplica (no h...
6	0	0	No aplica (no h...
7	0	0	No aplica (no h...
8	0	0	No aplica (no h...
9	0	0	No aplica (no h...
10	0	0	No aplica (no h...
11	0	0	No aplica (no h...
12	41	0	Esta bien el tie...
13	5	0	Quisiera dedicar...
14	50	0	Esta bien el tie...
15	0	0	No aplica (no h...
16	0	0	No aplica (no h...
17	0	0	No aplica (no h...
18	0	0	No aplica (no h...
19	0	0	No aplica (no h...
20	0	0	No aplica (no h...
21	0	0	No aplica (no h...
22	0	0	No aplica (no h...
23	45	0	Quisiera dedicar...
24	0	0	No aplica (no h...
25	0	0	No aplica (no h...
26	0	0	No aplica (no h...
27	0	0	No aplica (no h...
28	0	0	No aplica (no h...
29	0	0	No aplica (no h...
30	0	0	No aplica (no h...
31	32	0	Esta bien el tie...
32	37	1	Esta bien el tie...
33	43	0	Esta bien el tie...
34	0	0	No aplica (no h...
35	0	0	No aplica (no h...
36	16.3333	0	Esta bien el tie...
37	0	0	No aplica (no h...

⋮ ? ⌂ | ← 1000 → 1000 | 1000

Resultado de la selección de variables en la herramienta Select Columns en Orange



Selección de variables en la herramienta Select Columns en Orange

Info - Orange

File Edit View Window Help

Info  
1000 instances (no missing data)  
3 features  
No target variable.  
No meta attributes.

Variables  
 Show variable labels (if present)  
 Visualize numeric values  
 Color by instance classes  
 Select full rows

Selection  
 Select full rows

Restore Original Order

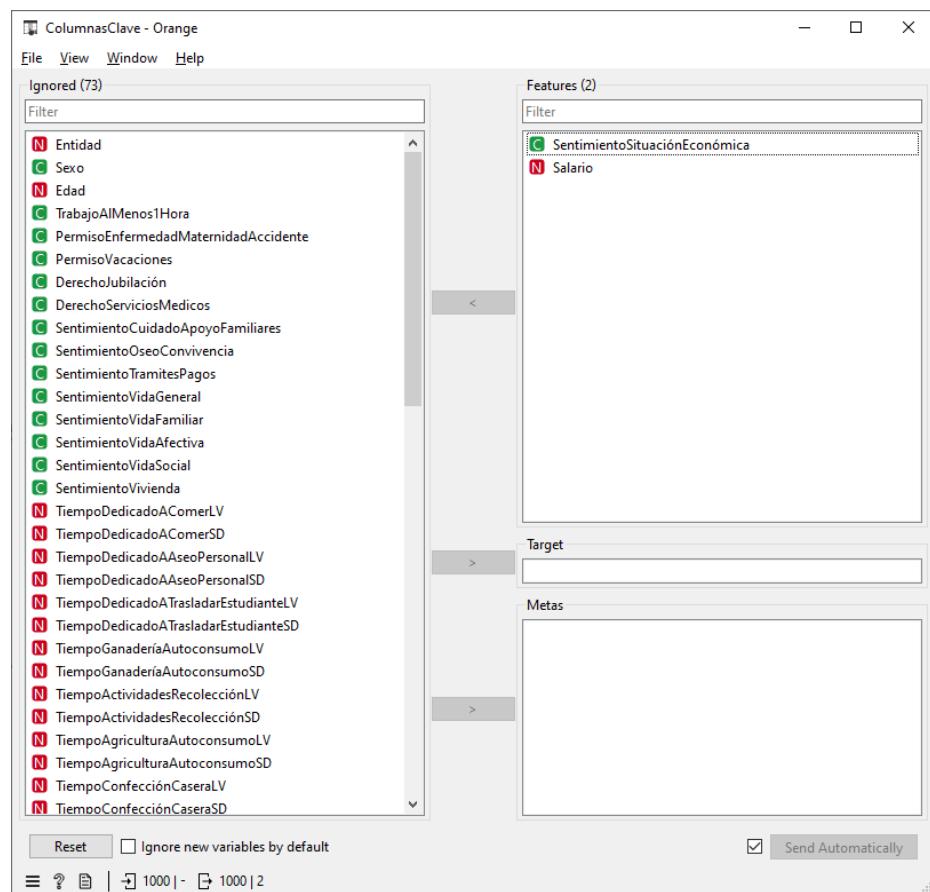
Send Automatically

1    2    3    4    5    6    7    8    9    10    11    12    13    14    15    16    17    18    19    20    21    22    23    24    25    26    27    28    29    30    31    32    33    34    35    36    37

haceresDomésticos    haceresDomésticos    mientoQuehaceres

	haceresDomésticos	haceresDomésticos	mientoQuehaceres
1	7.25	1.71667	Esta bien el tie...
2	0.416667	0.166667	No aplica (no h...
3	0.0833333	0.0333333	No aplica (no h...
4	3.58333	0.75	Esta bien el tie...
5	9.33333	2.06667	Esta bien el tie...
6	0	0	No aplica (no h...
7	4.41667	0.166667	Esta bien el tie...
8	18.6667	7.66667	Esta bien el tie...
9	37	15.25	Esta bien el tie...
10	40.75	14.5	Quisiera dedica...
11	26.25	15.3333	Esta bien el tie...
12	1.41667	0.666667	Esta bien el tie...
13	1.33333	0	Quisiera dedica...
14	3.41667	0.1	Esta bien el tie...
15	4.25	0.666667	Quisiera dedica...
16	26.5833	11.8333	Quisiera dedica...
17	0.5	0	Quisiera dedica...
18	17	3.08333	Esta bien el tie...
19	1.5	0.666667	Esta bien el tie...
20	24.25	4.16667	Quisiera dedica...
21	14.1167	4.66667	Esta bien el tie...
22	15.6667	4.66667	Esta bien el tie...
23	12.3333	10.3333	Quisiera dedica...
24	13.75	11.5833	Esta bien el tie...
25	2.38333	0	Esta bien el tie...
26	31.3333	13.3333	Esta bien el tie...
27	30.9167	12.9833	Esta bien el tie...
28	0.416667	0.166667	No aplica (no h...
29	25	6	Quisiera dedica...
30	20.3333	5.75	Esta bien el tie...
31	1.33333	0	Esta bien el tie...
32	9	1	Esta bien el tie...
33	3.5	1	Esta bien el tie...
34	2.5	10.6667	Esta bien el tie...
35	0.25	0	Esta bien el tie...
36	21.5833	5.18333	Esta bien el tie...
37	0.166667	0	Quisiera dedica...

Resultado de la selección de variables en la herramienta Select Columns en Orange



Selección de variables en la herramienta Select Columns en Orange

Info - Orange

File Edit View Window Help

**Info**  
1000 instances (no missing data)  
2 features  
No target variable.  
No meta attributes.

**Variables**  
 Show variable labels (if present)  
 Visualize numeric values  
 Color by instance classes

**Selection**  
 Select full rows

entoSituaciónEco Salario

1	Satisfecho(a)	1300
2	Mas o menos s...	1300
3	Satisfecho(a)	3000
4	Poco satisfecho...	0
5	Satisfecho(a)	5000
6	Mas o menos s...	0
7	Satisfecho(a)	0
8	Satisfecho(a)	0
9	Poco satisfecho...	600
10	Mas o menos s...	0
11	Poco satisfecho...	0
12	Satisfecho(a)	0
13	Poco satisfecho...	1500
14	Poco satisfecho...	0
15	Mas o menos s...	99999
16	Mas o menos s...	0
17	Satisfecho(a)	0
18	Mas o menos s...	0
19	Satisfecho(a)	4000
20	Nada satisfech...	1000
21	Muy satisfecho...	0
22	Nada satisfech...	0
23	Poco satisfecho...	0
24	Satisfecho(a)	1000
25	Mas o menos s...	1600
26	Muy satisfecho...	99999
27	Satisfecho(a)	1750
28	Satisfecho(a)	0
29	Mas o menos s...	20000
30	Poco satisfecho...	1800
31	Poco satisfecho...	500
32	Satisfecho(a)	0
33	Satisfecho(a)	200
34	Mas o menos s...	0
35	Poco satisfecho...	0
36	Satisfecho(a)	0
37	Mas o menos s...	1800

Restore Original Order  Send Automatically

≡ ? 🌐 | ⌛ 1000 ⌛ 1000 | 1000

Resultado de la selección de variables en la herramienta Select Columns en Orange

### c. Limpieza de datos:

Los puntos del i al v se abordan en un orden diferente al propuesto inicialmente, se toman en cuenta sin embargo por cómo se realiza la limpieza, no se sigue ese orden aunque se presenta un pequeño resumen en cada punto.

#### i. Eliminación de atributos con poca variabilidad

Aunque el conjunto de datos original contenía varios atributos con baja variabilidad, estos estaban relacionados con un grupo más amplio de atributos. Para preservar la información relevante,

en lugar de eliminarlos directamente, se optó por fusionarlos. Este enfoque se detalla más adelante en las secciones sobre "Quehaceres del hogar" y "Actividades recreativas" entre otras.

**ii. Identificación de valores erróneos (outliers y tipos)**

**iii. Detección de valores faltantes:**

Se identificaron instancias que carecían de un valor definido en el atributo de salario (P5\_7). Dado que estos valores faltantes podrían introducir sesgos en la ejecución del algoritmo K-means, se tomó la decisión de eliminar dichas instancias. Esta acción resultó en una reducción de 7,000 instancias en el conjunto de datos original.

**iv. Eliminación de falsos predictores**

Se examinaron las relaciones entre los atributos y la variable objetivo para identificar posibles falsos predictores. Estos son atributos que, aunque puedan parecer relevantes a primera vista, no aportan información significativa para predecir la variable objetivo o pueden introducir sesgos en el modelo y no se encontraron.

**v. Errores de dominio, tipo o formato.**

Se realizó una revisión minuciosa de los datos de la muestra seleccionada (más adelante en la limpieza) para detectar errores de dominio (valores que no tienen sentido en el contexto del problema), errores de tipo (atributos con tipos de datos incorrectos) y errores de formato (inconsistencias en la forma en que se representan los datos), no se encontraron inconsistencias.

## Conjunto de datos ENUT antes de la limpieza:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	UPM	VIV_SEL	HOGAR	NREN	ENT	SEXO	EDAD_V	PM_1	P4_1C	P4_2	NIV	GRAS	P4_4	P4_5	P5_1	P5_2	P5_3_1	P5_3_2	P5_3_3	P5_3_4	P5_4_1
2	0000001	1	01	01	2	64	2	1	2	3	4	2	2	2	7	40	50	20	50	01	
3	0000001	02	01	01	1	30	2	1	6	3	6	2	1	1	55	50	50	50	02	50	
4	0000001	02	01	01	1	25	2	1	3	3	6	2	1	1	22	50	18	50	00	50	
5	0000001	02	01	01	2	22	2	1	8	2	6	2	1	1	40	50	50	50	05	50	
6	0000001	03	01	01	2	48	2	1	7	3	2	2	1	1	40	50	50	50	00	50	
7	0000001	03	01	01	1	26	2	1	8	3	6	2	1	1	40	50	50	50	00	50	
8	0000001	03	01	01	1	17	2	1	6	6	2	2	2	7							
9	0000001	03	01	01	2	68	2	1	2	3	2	2	2	2	7	55	50	50	50	00	50
10	0000001	04	01	01	1	50	2	1	2	6	6	2	1	1							
11	0000001	04	01	01	2	21	2	1	3	3	2	2	2	7							
12	0000001	04	01	01	1	38	2	1	8	4	5	2	1	1	60	50	12	50	05	50	
13	0000001	04	01	01	2	38	2	1	8	4	5	2	1	1	40	50	50	50	02	50	
14	0000001	04	01	01	1	61	2	1	4	3	5	2	1	1	20	50	10	50	00	50	
15	0000002	01	01	01	2	37	2	1	6	3	5	2	1	1	55	50	21	50	00	50	
16	0000002	01	01	01	1	18	2	1	6	3	6	2	1	1							
17	0000002	02	01	01	2	68	2	1	4	3	4	2	2	2	7						
18	0000002	03	01	01	1	62	2	1	8	5	5	2	1	1	50	50	50	50	00	50	
19	0000002	03	01	01	2	60	2	1	3	3	5	2	2	7							
20	0000002	03	01	01	2	30	2	1	8	4	4	2	1	1	50	50	50	50	00	50	
21	0000002	03	01	01	1	37	2	1	6	3	6	2	1	1	40	50	30	50	05	50	
22	0000002	05	01	01	2	29	2	1	3	3	2	2	1	1	60	50	20	50	01	50	
23	0000003	01	01	01	2	57	2	1	2	6	4	2	1	1	20	50	50	50	00	50	
24	0000003	01	01	01	1	30	2	1	8	2	6	2	2	5	50	50	10	50	00	50	
25	0000003	01	01	01	2	27	2	1	8	4	4	2	1	1	45	50	50	50	02	50	
26	0000003	01	01	01	2	26	2	1	8	4	5	2	1	1	50	50	50	50	02	50	
27	0000003	01	01	01	1	35	2	1	9	2	1	2	1	1	50	50	50	50	00	50	
28	0000003	01	01	01	2	28	2	1	8	4	1	1	4	1	45	50	50	50	02	50	
29	0000003	01	01	01	1	34	2	1	6	3	6	1	4	1	40	50	50	50	00	50	
30	0000003	04	01	01	2	68	2	1	3	3	6	2	1	1	50	50	50	50	05	50	
31	0000003	04	01	01	2	63	2	1	3	3	6	2	1	1	50	50	50	50	00	50	
32	0000003	04	01	01	2	59	2	1	2	3	6	2	1	1	45	50	50	50	05	50	

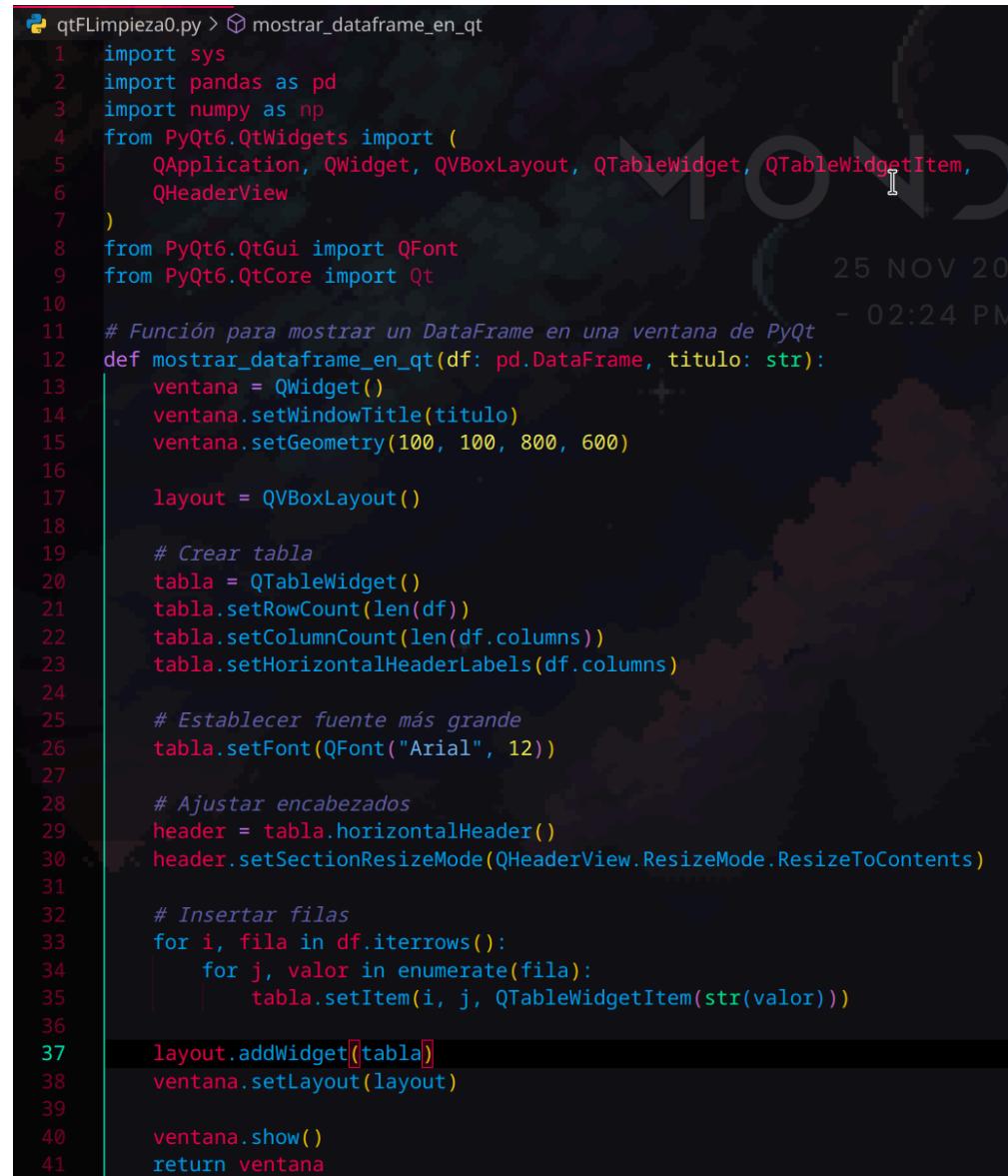
## Conjunto de datos ENUT antes después de la limpieza:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Entidad	Edad																			
2	29	Alta	17	01	01	0	0	0	0	Permitir	No	0	0	0	0	0	0	0	0	0	
3	9	Mujer	58	Si	50	0	7,5	0	No	Permitir	0	0	0	0	0	0	0	0	0	0	
4	15	Hombre	81	No	0	0	0	0	No	Permitir	0	0	0	0	0	0	0	0	0	0	
5	6	Hombre	60	0	10	1,666666667	0,25	104000	Si	Si	Si	Si	32	12	2,5	1	0	0	0	0	
6	16	Hombre	50	0	10	1,333333333	0,666666667	980000	No	No	No	No	35	14	2,5	1	0	0	0	0	
7	24	Hombre	60	Si	50	0	0	0	0	Permitir	0	0	0	0	0	0	0	0	0	0	
8	32	Mujer	16	No	0	0	0	0	No	Permitir	0	0	0	0	0	0	0	0	0	0	
9	20	Mujer	54	Si	24	4	1	0,333333333	130000	No	Permitir	0	0	0	0	0	0	0	0	0	0
10	28	Hombre	51	Si	30	24	10	6	156000	No	Permitir	0	0	0	0	0	0	0	0	0	0
11	40	Hombre	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	1	Mujer	19	Si	35	0	2,5	0	104000	No	No	No	No	35	16	2,5	1	51,5	9	1	
13	16	Mujer	50	No	0	0	0	0	No	Permitir	0	0	0	0	0	0	0	0	0	0	
14	31	Mujer	60	Si	10	0	0	0	0	52000	Si	Si	Si	30	12	7,5	5	3,333333333	0,666666667	0	
15	22	Hombre	50	0	24	3,333333333	1,333333333	620000	No	No	No	No	35	14	2,5	1	0	0	0	0	
16	31	Mujer	20	Si	40	1	0,666666667	0	90000	No	No	No	No	35	20	10	2	6	4	0,3333	
17	27	Mujer	57	No	0	0	0	0	No	Permitir	0	0	0	0	0	0	0	0	0	0	
18	13	Mujer	12	No	0	0	0	0	No	Permitir	0	0	0	0	0	0	0	0	0	0	
19	8	Mujer	50	Si	40	5	0,5	0,666666667	83200	Si	Si	Si	25	18	2,5	1	37,5	15	1,6666		
20	30	Mujer	37	No	0	0	0	0	No	Permitir	0	0	0	0	0	0	0	0	0	0	
21	3	Hombre	50	Si	40	5	5	1	78000	Si	Si	Si	60	24	6,5	3	2,5	1	0	0	
22	42	Hombre	48	Si	16	0,5	2,333333333	0,2	32000	Si	Si	Si	40	16	10	2	10,666666667	2	0	0	
23	18	Hombre	36	Si	40	16	0,5	0,833333332	0,2	32000	Si	Si	Si	40	16	2,5	1	1,666666667	1	0	
24	15	Hombre	45	Si	40	16	0	0	0	62400	Si	Si	Si	35	15	2,5	0,5	0	0	0	
25	29	Mujer	45	Si	40	16	0	0	0	31200	No	Permitir	0	40	18	11,25	2,5	2,916666667	1,666666667	0	
26	4	Mujer	23	No	0	0	0	0	41600	Si	Si	Si	20,666666667	10,5	5	2	10	4	0	0	
27	1	Mujer	15	Si	40	9	0,5	1	57200	No	Permitir	0	40	16	2,5	3	5	2	0	0	0
28	12	Hombre	40	No	0	0	0	0	No	Permitir	0	40	16	2,5	2	5	2	0	0	0	
29	2	Hombre	45	Si	40	6	2,5	0,5	156000	Si	Si	Si	35	14	2,5	1	0	0	0	0	
30	10	Mujer	14	Si	0	6	0	3	13000	No	No	No	45	4	3,333333333	0,5	2	41,5	0	1	
31	21	Hombre	63	Si	15	0	0	0	0	Permitir	0	0	0	40	16	3	2	5	2	0	0
32	20	Hombre	50	Si	32	16	3,333333333	1,066666667	57200	Si	Si	Si	40	16	7,5	3	7,5	5	0	0	0
33	22	Mujer	20	No	0	0	0	0	No	Permitir	0	0	0	40	16	8	2,9	4	2	40	0

El objetivo principal de la limpieza fue reducir en gran medida el número de atributos del dataset original, que era de 553, resultando en 75 atributos al final de la limpieza, además seleccionar una muestra pequeña de 1000 instancias, dado que el dataset contaba con más de 68000 instancias, para esto, optamos por utilizar python, con ayuda de la librería de pandas para leer archivos csv y utilizar sus versátiles data frames, la tarea de limpieza no solo consistió en lo anterior, además, el dataset contaba con múltiples diccionarios que optamos por pasar su significado al dataset, sabemos que esto conlleva aumentar el peso y por consecuencia el tiempo de ejecución de los modelos pero nos pareció una mejora considerable ya que facilita interpretar los datos.

## Interfaz gráfica

Para apreciar con comodidad el resultado de la limpieza de datos optamos por utilizar la librería de qt para python, no es igual de rápida que en c++ pero funciona para una cómoda visualización, creamos una función básica para mostrar una ventana con la tabla.



The screenshot shows a terminal window with a dark background. The code is displayed in white and blue text. The terminal window has a watermark in the background that reads "25 NOV 2020 - 02:24 PM". The code itself is a Python script named "qtFLimpieza0.py" which defines a function "mostrar\_dataframe\_en\_qt" that creates a Qt window with a QTableWidget containing a DataFrame.

```
qtFLimpieza0.py > ⌂ mostrar_dataframe_en_qt
 1  import sys
 2  import pandas as pd
 3  import numpy as np
 4  from PyQt6.QtWidgets import (
 5      QApplication, QWidget, QVBoxLayout, QTableWidget, QTableWidgetItem,
 6      QHeaderView
 7  )
 8  from PyQt6.QtGui import QFont
 9  from PyQt6.QtCore import Qt
10
11 # Función para mostrar un DataFrame en una ventana de PyQt
12 def mostrar_dataframe_en_qt(df: pd.DataFrame, titulo: str):
13     ventana = QWidget()
14     ventana.setWindowTitle(titulo)
15     ventana.setGeometry(100, 100, 800, 600)
16
17     layout = QVBoxLayout()
18
19     # Crear tabla
20     tabla = QTableWidget()
21     tabla.setRowCount(len(df))
22     tabla.setColumnCount(len(df.columns))
23     tabla.setHorizontalHeaderLabels(df.columns)
24
25     # Establecer fuente más grande
26     tabla.setFont(QFont("Arial", 12))
27
28     # Ajustar encabezados
29     header = tabla.horizontalHeader()
30     header.setSectionResizeMode(QHeaderView.ResizeMode.ResizeToContents)
31
32     # Insertar filas
33     for i, fila in df.iterrows():
34         for j, valor in enumerate(fila):
35             tabla.setItem(i, j, QTableWidgetItem(str(valor)))
36
37     layout.addWidget(tabla)
38     ventana.setLayout(layout)
39
40     ventana.show()
41     return ventana
```

Código de función salida gráfica.

## Orígenes de archivos csv

Como se mencionó en “**Orígenes de datos y su disponibilidad**” (Página 15) para interpretar el significado de cada parte del dataset es necesario acceder a una intrincada estructura de carpetas. para esto creamos múltiples variables con las rutas para cada catálogo, el diccionario y el dataset.

```

# Diccionario original
44 ruta_diccionario_original = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/diccionario_de_datos/
diccionario_datos_tmodulo_enut_2019.csv"
45 # Ruta dataset
46 ruta_dataset = "./conjunto_de_datos_tmodulo_enut_2019/conjunto_de_datos/conjunto_de_datos_tmodulo_enut_2019.csv"
47 # Rutas de catalogos
48 ruta_catalogo_sexo = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/exco.csv"
49 ruta_catalogo_permiso_enfermedad_internamiento = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p5_5_1.csv"
50 ruta_catalogo_permiso_cuidado_menores_hogar = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p5_1_1.csv"
51 ruta_catalogo_permiso_vacaciones = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p5_6_2.csv"
52 ruta_catalogo_derecho_jubilacion = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p5_6_3.csv"
53 ruta_catalogo_derecho_servicios_medicos = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p5_6_6.csv"
54 ruta_catalogo_sentimiento_quedaciones_hogar = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_1_1.csv"
55 ruta_catalogo_sentimiento_estudio = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_1_2.csv"
56 ruta_catalogo_sentimiento_trabajo = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_1_3.csv"
57 ruta_catalogo_sentimiento_cuidado_apoyo_familiares = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/
p7_1_4.csv"
58 ruta_catalogo_sentimiento_ocio_convivencia = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_1_5.csv"
59 ruta_catalogo_sentimiento_traslado = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_1_6.csv"
60 ruta_catalogo_sentimiento_ofertas_pagos = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_1_7.csv"
61 ruta_catalogo_sentimiento_vida_trabajo = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_2_1.csv"
62 ruta_catalogo_sentimiento_vida_familiar = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_2_2.csv"
63 ruta_catalogo_sentimiento_vida_afectiva = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_2_3.csv"
64 ruta_catalogo_sentimiento_vida_social = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_2_4.csv"
65 ruta_catalogo_sentimiento_situacion_economica = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_2_5.csv"
66 ruta_catalogo_sentimiento_vivienda = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_2_6.csv"
67 ruta_catalogo_sentimiento_felicidad_general = "./conjunto_de_datos_enut_2019_csv/conjunto_de_datos_enut_2019/conjunto_de_datos_tmodulo_enut_2019/catalogos/p7_3.csv"

```

## Selección, renombrado y fusión de atributos

Para esto cargamos el diccionario original con ayuda de pandas y lo mostramos en una ventana con ayuda de la función para salida gráfica.

```
def main():
    # Crear la aplicación una sola vez
    app = QApplication(sys.argv)

    # Cargar el diccionario original y mostrarlo en una ventana
    diccionario_original = pd.read_csv(ruta_diccionario_original)
    ventana1 = mostrar_dataframe_en_qt(diccionario_original, "Diccionario Original")
```

## Código

	Documento Original	longitud	tipo	nombre	catalogo	rango_cierre
1	Número del LPM	7	c	cpm	nan	0100017 - 3280769
2	Número de vivienda seleccionada	2	c	vvv_sel	nan	0-23
3	Número de Hogar	1	c	hogar	nan	0-5
4	Indicador de rango de la persona de la entrevista individual	2	c	ent	nan	0-20
5	Entidad	2	c	ent	nan	0-32
6	Sexo	1	c	sexo	sexo	1-2
7	Edad verificada	2	c	edad_v	edad	12-97
8	Pregunta 1.1.1. ¿Usted habla algún dialecto o lenguaje indígena?	1	c	indigena	indigena	0-1
9	Pregunta 1.1.2. ¿Usted sabe leer y escribir un recado?	4	c	leer_escr	nan	0200 - 7000_9999
10	Pregunta NV_Usté que año o grado aprobó usted en la escuela? - NIVEL	1	c	grado	nv	0-9
11	Pregunta NV_Hasta que año o grado asistió usted en la escuela? - GRADO	1	c	gra	gra	0-19
12	Pregunta NV_Está actualmente asistiendo	1	c	actual	actual	0-1
13	Pregunta P_En su cultura o tradición, ¿usted se considera indígena?	1	c	indigena	indigena	0-5
14	Pregunta P_6. Se considera indígena	1	c	p6_6	p6_6	1-2
15	Pregunta P_13 Durante la semana pasada, ¿usted trabajó al menos una hora?	1	c	p13_1	p13_1	0-1
16	Pregunta P_13 Si para ganar dinero o ayudar al gasto del hogar, la semana pasada usó...	1	c	p13_2	p13_2	0-1
17	Pregunta P_3.1 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (o actividad de lunes a viernes) - HORAS	2	c	p3_1	p3_1	0-998
18	Pregunta P_3.2 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (o actividad de lunes a viernes) - MINUTOS	2	c	p3_2	p3_2	0-998
19	Pregunta P_3.3 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (o actividad sábado y domingo) - HORAS	2	c	p3_3	p3_3	0-48
20	Pregunta P_3.4 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (o actividad sábado y domingo) - MINUTOS	2	c	p3_4	p3_4	0-96
21	Pregunta P_4.1 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de la y vuelta para trabajar (o su actividad de lunes a viernes) - HORAS	2	c	p4_1	p4_1	0-50
22	Pregunta P_4.2 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de la y vuelta para trabajar (o su actividad de lunes a viernes) - MINUTOS	2	c	p4_2	p4_2	0-95
23	Pregunta P_4.3 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de la y vuelta para trabajar (o su actividad sábado y domingo) - HORAS	2	c	p4_3	p4_3	0-48
24	Pregunta P_4.4 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de la y vuelta para trabajar (o su actividad sábado y domingo) - MINUTOS	2	c	p4_4	p4_4	0-96
25	Pregunta P_5.1 Usted no trabaja (actividad) (principal) la semana pasada, ¿cuál fue?	1	c	p5_1	p5_1	1-5
26	Pregunta P_5.2 Aunque no le use, ¿en su trabajo (principal) tiene derecho a licencia o permiso con sueldo por enfermedad, accidente o maternidad (paternidad)?	1	c	p5_2	p5_2	1-2
27	Pregunta P_6.2 Aunque no le use, ¿en su trabajo (principal) tiene derecho a vacaciones pagadas?	1	c	p6_2	p6_2	1-2
28	Pregunta P_6.3 Aunque no le use, ¿en su trabajo (principal) tiene derecho a jubilación, pension o sistema de ahorro para el retiro?	1	c	p6_3	p6_3	0-9
29	Pregunta P_6.4 Aunque no le use, ¿en su trabajo (principal) tiene derecho a pensiones o subsidios?	1	c	p6_4	p6_4	0-5

## Salida gráfica

Procedemos a cargar el dataset y crear un nuevo data frame vacío, en este vamos a colocar las columnas que nos sean de utilidad del dataset original.

```
77      # Cargar el dataset original
78      dataset = pd.read_csv(ruta_dataset)
79
80      # creamos un nuevo dataframe vacío
81      dataset_limpio = pd.DataFrame()
82
```

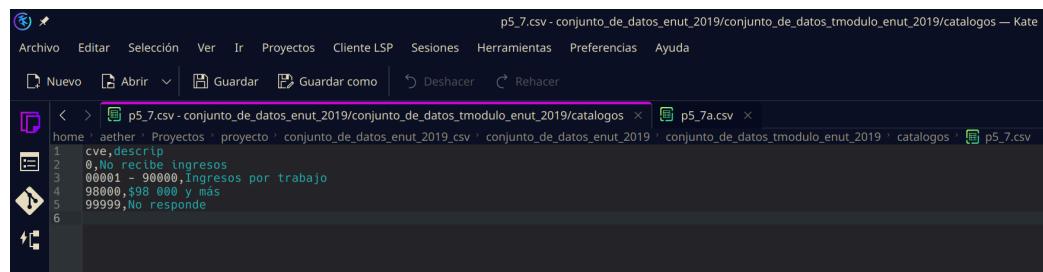
código para cargar el dataset y crear un nuevo data frame vacío.

Una vez cargado el dataset original procedemos a cambiar los valores “nan” por 0 dado que es mucho más fácil operar las fusiones con estos valores (las fusiones se detallan a continuación) además de quitar todas las instancias que no especificaron su salario.

verificamos el catálogo para la “P5\_7”

nombre_campo	diccionario	longitud	tipo	memoria	catalogo	rango_claves
26_Pregunta_P5_7	26_Pregunta_P5_7	1	C	p6_8	p6_8	1,2,9
27_Pregunta_P5_7	27_Pregunta_P5_7	1	C	p6_8,3	p6_8,3	1,2,9
28_Pregunta_P5_7	28_Pregunta_P5_7	1	C	p6_8,4	p6_8,4	1,2,9
29_Pregunta_P5_7	29_Pregunta_P5_7	1	C	p6_8,5	p6_8,5	1,2,9
30_Pregunta_P5_7	30_Pregunta_P5_7	1	C	p6_8,6	p6_8,6	1,2,9
31_Pregunta_P5_7	31_Pregunta_P5_7	1	C	p6_8,7	p6_8,7	1,2,9
32_Pregunta_P5_7	32_Pregunta_P5_7	1	C	p6_8,8	p6_8,8	1,2,9
33_Pregunta_P5_7	33_Pregunta_P5_7	1	C	p6_8,9	p6_8,9	1,2,9
34_Pregunta_P5_7	34_Pregunta_P5_7	1	C	p6_8,10	p6_8,10	1,2,9

## Buscamos y accedemos con kate



Las instancias que deseamos eliminar son las que tienen “99999” dado que no responde no es lo mismo que no recibe ingresos, entonces son instancias que pueden perjudicar la precisión de los modelos.

```
83      # eliminamos los valores nulos
84      dataset = dataset.fillna(0) # Reemplaza NaN con 0
85
86      # Eliminamos los valores 99999 para la pregunta P5_7 (Salario)
87      dataset = dataset[dataset["P5_7"] != 99999]
88      dataset.reset_index(drop=True, inplace=True)
```

código para reemplazar “nan” con “0” y eliminar las instancias que no especificaron su salario.

Posterior a esto con ayuda del diccionario seleccionamos los atributos que nos parecen interesantes del dataset original y los comenzamos a agregar al dataset limpio.

Diccionario Original					
1	Número de UPM				
2	Número de vivienda seleccionada				
3	Número de Hogar				
4	Número de renglón de la persona de la entrevista individual				
5					
6	Sexo				
7	Edad verificada				
8	Pregunta P4.1 ¿Usted habla algún dialecto o lenguaje indígena?				
9	Pregunta P4.1a ¿Cuál?				
10	Pregunta P4.2 ¿Usted sabe leer o escuchar un idioma?				
11	Pregunta P4.3 ¿Usted sabe leer o escuchar un idioma? (si no, ¿cuál?)				
12	Pregunta GRA ¿Hasta qué año o grado aprobado usted en la escuela?				
13	Pregunta P4.4 ¿Actualmente usted...				
14	Pregunta P4.5 ¿Usted considera indígena?				
15	Pregunta P4.6 ¿Usted considera indígena...				
16	Pregunta P5.1 Durante la semana pasada, ¿usted trabajó al menos una hora?				
17	Pregunta P5.2 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (su actividad) de lunes a viernes? - HORAS				
18	Pregunta P5.3 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (su actividad) de lunes a viernes? - MINUTOS				
19	Pregunta P5.3.3 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (su actividad) sábado y domingo? - HORAS				
20	Pregunta P5.3.4 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (su actividad) sábado y domingo? - MINUTOS				
21					

## Diccionario Original

Encontramos una buena cantidad de preguntas que aunque son interesantes no nos parecen de utilidad para nuestra problemática.

Omitimos **Número de UPM** ya que no pretendemos seleccionar por área geográfica, Tampoco no es útil el **número de vivienda seleccionada**, ni el **número de hogar**, **número de renglón de la persona de la entrevista individual**, no buscamos saber si la persona **habla un dialecto indígena**, ni **cual es**, aunque suena interesante tampoco nos es de utilidad **saber su nivel educativo**, tampoco si **se considera indigena**. por lo cual omitimos todas las preguntas relacionadas.

Nos interesa saber la **entidad federativa** de la muestra, el **sexo** de la muestra, su **edad** y si **trabaja al menos una hora**

```

95     dataset_limpio["Entidad"] = dataset["ENT"]
96
97     # Pregunta Sexo
98     dataset_limpio["Sexo"] = dataset["SEXO"]
99     catalogo_sexo = pd.read_csv(ruta_catalogo_sexo)
100    dataset_limpio["Sexo"] = dataset_limpio["Sexo"].replace(catalogo_sexo.set_index("cve").to_dict()["descrip"])
101
102    # Pregunta Edad
103    dataset_limpio["Edad"] = dataset["EDAD_V"]
104
105    # Pregunta P5_1
106    dataset_limpio["TrabajoAlMenos1Hora"] = dataset["P5_1"]
107    catalogo_trabajo_al_menos_1_hora = pd.read_csv(ruta_catalogo_trabajo_al_menos_1_hora, encoding="latin-1")
108    dataset_limpio["TrabajoAlMenos1Hora"] = dataset_limpio["TrabajoAlMenos1Hora"].replace(catalogo_trabajo_al_menos_1_hora.set_index("cve").to_dict()["descrip"])

```

Código que muestra la extracción de la información de entidad, sexo, edad y si trabaja al menos una hora del dataset original.

Nos encontramos con una estructura de preguntas que es interesante

Pregunta	Categoría	Número	Código	Línea	Rango	Rango_min	Rango_max	Categoría
16 Pregunta P5_1 Durante la semana pasada, ¿usted trabajó al menos una hora?								
17 Pregunta P5_2 Dígale en peso general dentro de su gasto de hogar, ¿a qué semana pasada usted:								
18 Pregunta P5_3 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (incluyendo el tiempo de lunes a viernes)? - HORAS								
19 Pregunta P5_3_1 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (incluyendo el tiempo de lunes a viernes)? - MINUTOS								
20 Pregunta P5_3_2 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (excluyendo sábado y domingo)? - HORAS								
21 Pregunta P5_3_3 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (excluyendo sábado y domingo)? - MINUTOS								
22 Pregunta P5_3_4 Durante la semana pasada, ¿cuánto tiempo en total dedicó a trabajar (excluyendo sábado y domingo)? - HORAS								
23 Pregunta P5_3_5 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de casa a vuelta para trabajar (a su actividad de lunes a viernes)? - HORAS								
24 Pregunta P5_3_6 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de casa a vuelta para trabajar (a su actividad de lunes a viernes)? - MINUTOS								
25 Pregunta P5_4 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de casa a vuelta para trabajar (a su actividad) sábado y domingo? - HORAS								
26 Pregunta P5_4_1 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de casa a vuelta para trabajar (a su actividad) sábado y domingo? - MINUTOS								
27 Pregunta P5_4_2 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de casa a vuelta para trabajar (a su actividad) sábado y domingo? - HORAS								
28 Pregunta P5_4_3 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de casa a vuelta para trabajar (a su actividad) sábado y domingo? - MINUTOS								
29 Pregunta P5_4_4 Durante la semana pasada, ¿cuánto tiempo utilizó en trasladarse de casa a vuelta para trabajar (a su actividad) sábado y domingo? - MINUTOS								

Es un grupo de 9 preguntas con relación, como mencionamos al comienzo de la limpieza, queremos reducir al mínimo el número de columnas perdiendo el mínimo de información.

Revisamos con ayuda de kate el catálogo de la P5\_2

```

< > p5_2.csv - conjunto_de_datos_enut_2019/conjunto_de_datos_modulo_enut_2019/catalogos < p5_7a.csv < p5_2.csv <
home aether Proyectos proyecto conjunto_de_datos_modulo_enut_2019 catalogos p5_2.csv
1 1, "Viví en un hogar familiar o no familiar?"
2 2, "¿Vivió en un negocio familiar o no familiar?"
3 3, "¿Vendió o hizo algún producto para vender? (tortillas, artesanías, ropa, textiles, comida, muebles, ladrillos, tabique, etcétera)"
4 4, "¿Ayudó en las labores del campo, cría de animales, caza o pesca? (arando o desyerbando, cuidando o alimentando cerdos, gallinas, vacas, etcétera)"
5 5, "Acarréo agua, recogió leña, plantas, hongos, flores o frutos silvestres para vender?"
6 6, "¿Vivió en un hogar familiar o no familiar?"
7 7, "Tenía trabajo, pero estuvo ausente? (vacaciones, enfermedad, huéija, para técnico, etcétera)"
8 8, "Entonces, ¿no trabajó?"
9

```

Notamos que básicamente pregunta por su tipo de trabajo y en las preguntas posteriores las horas dedicadas de lunes a viernes y de sábado a domingo, además de las horas de transporte en los mismos periodos, podemos simplificar esto a solo dos columnas para lunes a viernes y dos más para fines de semana, tomamos las horas de trabajo totales (horas y minutos) y las horas de transporte totales (horas y minutos) y en cuanto a la actividad realizada nos es indiferente, así que lo omitimos, a esta compactación de atributos del dataset lo llamaremos **fusión**.

```

110 # ESTA EN HORAS TOTALES - 04:51 PM -
111 dataset_limpio["TiempoTrabajoLV"] = dataset["P5_3_1"] + (dataset["P5_3_2"] / 60)
112 # ESTA EN HORAS TOTALES
113 dataset_limpio["TiempoTrabajoSD"] = dataset["P5_3_3"] + (dataset["P5_3_4"] / 60)
114 # ESTA EN HORAS TOTALES
115 dataset_limpio["TiempoTrasladoLV"] = dataset["P5_4_1"] + (dataset["P5_4_2"] / 60)
116 # ESTA EN HORAS TOTALES
117 dataset_limpio["TiempoTrasladoSD"] = dataset["P5_4_3"] + (dataset["P5_4_4"] / 60)

```

Código con la fusión de columnas

Diccionario Original					
	número_cargo	longitud	tipo	recorrido	rango_claves
37	Pregunta P5_1 * ¿Aunque no les use, ¿en su trabajo (principal) tiene derecho a licencia o permiso con sueldo por enfermedad, accidente o maternidad (paternidad)?	1	C	p5_8, 1	1, 2, 9
38	Pregunta P5_2 * Aunque no las use, ¿en su trabajo (principal) tiene derecho a licencias o vacaciones pagadas?	1	C	p5_8, 2	1, 2, 9
39	Pregunta P5_3 * Aunque no las use, ¿en su trabajo (principal) tiene derecho a jubilación, pensión o sistema de ahorro para el retiro?	1	C	p5_8, 3	1, 2, 9
40	Pregunta P5_4 * Aunque no las use, ¿en su trabajo (principal) tiene derecho a guardería o estancia infantil?	1	C	p5_8, 4	1, 2, 9
41	Pregunta P5_5 * Aunque no las use, ¿en su trabajo (principal) tiene derecho a licencia por cuidado maternal (paternidad)?	1	C	p5_8, 5	1, 2, 9
42	Pregunta P5_6 * Aunque no las use, ¿en su trabajo (principal) tiene derecho a servicios médicos (IMSS, ISSSTE, entre otros)?	1	C	p5_8, 6	1, 2, 9
33	Pregunta P5_7 * Aunque no las use, ¿en su trabajo (principal) tiene derecho a otras prestaciones?	1	C	p5_8, 7	1, 2, 9
34	Pregunta P5_8 * Cuánto gana o recibe usted por trabajar (su actividad)?	5	C	p5_7	00001-92000-98000-99999
35	Pregunta P5_9 * ¿Cuál es su edad?	1	C	p5_7a	1, 4

Las preguntas presentadas en esta sección (captura) tienen todas que ver con los derechos laborales de la persona así como su sueldo, por lo cual las pasamos sin más, notamos que cuentan con un catálogo por lo cual hacemos la conversión a los datos del mismo.

121	# Pregunta P5_7 (Salario)
122	# lo hacemos anual
123	dataset["P5_7"] = dataset.apply(lambda fila: fila["P5_7"] * 52 if fila["P5_7"] == 1 else fila["P5_7"] * 26 if fila["P5_7A"] == 2 else fila["P5_7"] * 12 if fila["P5_7A"] == 3 else fila["P5_7"] * 1 if fila["P5_7A"] == 4 else 0, axis=1)
124	dataset_limpio["Salario"] = dataset["P5_7"]
125	
126	# Pregunta P5_6_1
127	dataset_limpio["PermisoEnfermedadMaternidadAccidente"] = dataset["P5_6_1"]
128	catalogo_permiso_enfermedad_maternidad = pd.read_csv(ruta_catalogo_permiso_enfermedad_maternidad, encoding="latin-1")
129	dataset_limpio["PermisoEnfermedadMaternidadAccidente"] = dataset_limpio["PermisoEnfermedadMaternidadAccidente"].replace(catalogo_permiso_enfermedad_maternidad.set_index("cve").to_dict()["descrip"])
130	
131	# Pregunta P5_6_2
132	dataset_limpio["PermisoVacaciones"] = dataset["P5_6_2"]
133	catalogo_permiso_vacaciones = pd.read_csv(ruta_catalogo_permiso_vacaciones, encoding="latin-1")
134	dataset_limpio["PermisoVacaciones"] = dataset_limpio["PermisoVacaciones"].replace(catalogo_permiso_vacaciones.set_index("cve").to_dict()["descrip"])
135	
136	# Pregunta P5_6_3
137	dataset_limpio["DerechoJubilacion"] = dataset["P5_6_3"]
138	catalogo_derecho_jubilacion = pd.read_csv(ruta_catalogo_derecho_jubilacion, encoding="latin-1")
139	dataset_limpio["DerechoJubilacion"] = dataset_limpio["DerechoJubilacion"].replace(catalogo_derecho_jubilacion.set_index("cve").to_dict()["descrip"])
140	
141	# Pregunta P5_6_6
142	dataset_limpio["DerechoserviciosMedicos"] = dataset["P5_6_6"]
143	catalogo_derecho_servicios_medicos = pd.read_csv(ruta_catalogo_derecho_servicios_medicos, encoding="latin-1")
144	dataset_limpio["DerechoserviciosMedicos"] = dataset_limpio["DerechoserviciosMedicos"].replace(catalogo_derecho_servicios_medicos.set_index("cve").to_dict()["descrip"])
145	
146	

código con la copia de los atributos y la conversión basada en los catálogos del dataset

41	Pregunta P6_1 * Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) de lunes a viernes? - MINUTOS
42	Pregunta P6_1.1 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) de lunes a viernes? - HORAS
43	Pregunta P6_1.1.1 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - HORAS
44	Pregunta P6_1.1.2 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - MINUTOS
45	Pregunta P6_1.2 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - HORAS
46	Pregunta P6_1.2.1 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - MINUTOS
47	Pregunta P6_1.2.2 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - HORAS
48	Pregunta P6_1.2.3 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - MINUTOS
49	Pregunta P6_1.3 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - HORAS
50	Pregunta P6_1.3.1 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - MINUTOS
51	Pregunta P6_1.3.2 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - HORAS
52	Pregunta P6_1.3.3 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - MINUTOS
53	Pregunta P6_1.3.4 Durante la semana pasada, ¿cuánto tiempo dedicó en total a dormir (incluye sueño) sábado y domingo? - HORAS
54	Pregunta P6_2.1 Durante la semana pasada, ¿cuánto asistió a clase, tomó cursos o estudió? (Incluya sistema abierto o a distancia, diplomatos, etcétera)
55	Pregunta P6_2.1.1 Cuánto tiempo se dedictó de lunes a viernes? - HORAS
56	Pregunta P6_2.1.2 Cuánto tiempo se dedictó de lunes a viernes? - MINUTOS
57	Pregunta P6_2.1.3 Cuánto tiempo se dedictó de lunes a viernes? - HORAS
58	Pregunta P6_2.1.4 Cuánto tiempo se dedictó de lunes a viernes? - MINUTOS
59	Pregunta P6_2.2 Durante la semana pasada, ¿cuánto hizo tareas, prácticas escolares o alguna otra actividad de estudio?
60	Pregunta P6_2.2.1 Cuánto tiempo se dedictó de lunes a viernes? - HORAS
61	Pregunta P6_2.2.2 Cuánto tiempo se dedictó de lunes a viernes? - MINUTOS
62	Pregunta P6_2.2.3 Durante la semana pasada, ¿cuánto tiempo se dedictó sábado y domingo? - HORAS
63	Pregunta P6_2.2.4 Cuánto tiempo se dedictó sábado y domingo? - MINUTOS
64	Pregunta P6_2.3 Durante la semana pasada, ¿cuánto se quedó de lunes a viernes en la escuela?
65	Pregunta P6_2.3.1 Cuánto tiempo se quedó de lunes a viernes en la escuela? - HORAS
66	Pregunta P6_2.3.2 Cuánto tiempo se quedó de lunes a viernes? - HORAS
67	Pregunta P6_2.3.3 Cuánto tiempo se quedó sábado y domingo? - HORAS
68	Pregunta P6_2.3.4 Cuánto tiempo se quedó sábado y domingo? - MINUTOS

Los atributos que podemos ver en esta captura pueden simplificarse de la misma forma que los anteriores con hora y minuto, dado lo cual los fusionamos.

```

147 # Pregunta P6_1_1_1, P6_1_1_2, P6_1_1_3, P6_1_1_4      = 05:31 PM =
148 dataset_limpio["TiempoDedicadoADormirLV"] = dataset["P6_1_1_1"] + (dataset["P6_1_1_2"] / 60)
149 dataset_limpio["TiempoDedicadaADormirSD"] = dataset["P6_1_1_3"] + (dataset["P6_1_1_4"] / 60)
150
151 # Pregunta P6_1_2_1, P6_1_2_2, P6_1_2_3, P6_1_2_4
152 dataset_limpio["TiempoDedicadaAComerLV"] = dataset["P6_1_2_1"] + (dataset["P6_1_2_2"] / 60)
153 dataset_limpio["TiempoDedicadaAComerSD"] = dataset["P6_1_2_3"] + (dataset["P6_1_2_4"] / 60)
154
155 # Pregunta P6_1_3_1, P6_1_3_2, P6_1_3_3, P6_1_3_4
156 dataset_limpio["TiempoDedicadaAAseoPersonalLV"] = dataset["P6_1_3_1"] + (dataset["P6_1_3_2"] / 60) + dataset["P6_6A_5_1"] + (dataset["P6_6A_5_2"] / 60)
157 dataset_limpio["TiempoDedicadaAAseoPersonalSD"] = dataset["P6_1_3_3"] + (dataset["P6_1_3_4"] / 60) + dataset["P6_6A_5_3"] + (dataset["P6_6A_5_4"] / 60)
158
159 # Pregunta P6_2_1_1, P6_2_1_2, P6_2_1_3
160 dataset_limpio["TiempoDedicadaAActividadesEstudioLV"] = dataset["P6_2A_1_1"] + (dataset["P6_2A_1_2"] / 60) + dataset["P6_2A_2_1"] + (dataset["P6_2A_2_2"] / 60)
161 dataset_limpio["TiempoDedicadaAActividadesEstudioSD"] = dataset["P6_2A_1_3"] + (dataset["P6_2A_1_4"] / 60) + dataset["P6_2A_2_3"] + (dataset["P6_2A_2_4"] / 60)
162
163 # Pregunta P6_2_4, P6_2_5
164 dataset_limpio["TiempoDedicadoATrasladarEstudianteLV"] = dataset["P6_2A_3_1"] + (dataset["P6_2A_3_2"] / 60)
165 dataset_limpio["TiempoDedicadoATrasladarEstudianteSD"] = dataset["P6_2A_3_3"] + (dataset["P6_2A_3_4"] / 60)

```

Código para la selección y fusión de columnas, se redujo de 27 columnas a solo 10 (Tiempo dedicado a dormir LV, SD, comer LV, SD, aseo personal LV, SD, actividades estudio LV, SD, traslado estudiante LV, SD).

Ahora en el dataset hay muchas columnas que podríamos simplificar como **quehaceres domésticos, ganadería de autoconsumo, agricultura de autoconsumo, tiempo confección casera, tiempo mantenimiento del hogar, tiempo compras del hogar, tiempo trámites pagos, tiempo organización hogar.**

Diccionario Original				
	original	rops	removido	catalogo
64	C	p6_3a_1_1	nan	1,2
65	C	p6_3a_1_1,1	nan	00:00
66	C	p6_3a_1,2	nan	00:55
67	C	p6_3a_1,3	nan	00:40
68	C	p6_3a_1,4	nan	00:50
69	C	p6_3a_2,2	p6_3,2	1,2
70	C	p6_3a_2,1	nan	00:00
71	C	p6_3a_2,2	nan	00:55
72	C	p6_3a_2,3	nan	00:40
73	C	p6_3a_2,4	nan	00:50
74	C	p6_3a_3,3	nan	00:33
75	C	p6_3a_3,4	nan	00:48
76	C	p6_3a_3,2	nan	00:50
77	C	p6_3a_3,1	nan	00:24
78	C	p6_3a_3,4	nan	00:50
79	C	p6_3a_4,4	p6_3,4	1,2
80	C	p6_3a_4,1	nan	00:00
81	C	p6_3a_4,2	nan	00:55
82	C	p6_3a_4,3	nan	00:40
83	C	p6_3a_4,4	nan	00:50
84	C	p6_3a_5,5	p6_3,5	1,2
85	C	p6_3a_5,1	nan	00:50
86	C	p6_3a_5,2	nan	00:55
87	C	p6_3a_5,3	nan	00:40
88	C	p6_3a_5,4	nan	00:50

Documento Original						
	pregunta	tempo	tempo	tempo	tempo	tempo
93	Pregunta PE-3.6 Durante la semana pasada, SOLO para el consumo de su hogar, ¿usted elaboró o tejió ropa, manteles, cortinas o textiles, otros?					
94	Pregunta PE-3A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS					
95	Pregunta PE-3A.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS					
96	Pregunta PE-3A.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS					
97	Pregunta PE-3A.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS					
98	Pregunta PE-3.7 Durante la semana pasada, SOLO para el consumo de su hogar, ¿usted elaboró alimentos para conservarse o almacenarse? (mermeladas, conservas, encurtidos, salar carne, chocolate, pan; quesos u otros)					
99	Pregunta PE-3A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS					
100	Pregunta PE-3A.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS					
101	Pregunta PE-3A.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS					
102	Pregunta PE-3A.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS					
103	Pregunta PE-3.8 Durante la semana pasada, SOLO para el consumo de su hogar, ¿usted hizo muebles, utensilios de cocina, bloques, adobes u otros productos?					
104	Pregunta PE-3A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS					
105	Pregunta PE-3A.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS					
106	Pregunta PE-3A.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS					
107	Pregunta PE-3A.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS					
108	Pregunta PE-3.9 Durante la semana pasada, SOLO para el consumo de su hogar, ¿usted arregló o remodeló su vivienda o la está viendo/constuyendo?					
109	Pregunta PE-3A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS					
110	Pregunta PE-3A.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS					
111	Pregunta PE-3A.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS					
112	Pregunta PE-3A.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS					
113	Pregunta PE-4.1 Durante la semana pasada, ¿usted degustó mate, café o té, o tomó el mate/mate o hizó mate o té para su hogar?					
114	Pregunta PE-4A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS					
115	Pregunta PE-4A.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS					
116	Pregunta PE-4A.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS					
117	Pregunta PE-4A.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS					
1	C p6_3.6	p6_3.6	1.2			
2	C p6_3a_6_1	nan	00:00:00			
2	C p6_3a_6_2	nan	00:00:50			
2	C p6_3a_6_3	nan	00:40:00			
2	C p6_3a_6_4	nan	00:00:50			
1	C p6_3.7	p6_3.7	1.2			
2	C p6_3a_7_1	nan	00:00:00			
2	C p6_3a_7_2	nan	00:00:50			
2	C p6_3a_7_3	nan	00:16:00			
2	C p6_3a_7_4	nan	00:45:00			
1	C p6_3.8	p6_3.8	1.2			
2	C p6_3a_8_1	nan	00:00:48			
2	C p6_3a_8_2	nan	00:50:00			
2	C p6_3a_8_3	nan	00:39:00			
2	C p6_3a_8_4	nan	00:34:00			
1	C p6_3.9	p6_3.9	1.2			
2	C p6_3a_9_1	nan	00:00:00			
2	C p6_3a_9_2	nan	00:50:00			
2	C p6_3a_9_3	nan	00:36:00			
2	C p6_3a_9_4	nan	00:50:00			
1	C p6_4.1	p6_4.1	1.2			
2	C p6_4a_1_1	nan	00:00:00			
2	C p6_4a_1_2	nan	00:50:00			
2	C p6_4a_1_3	nan	00:40:00			
2	C p6_4a_1_4	nan	00:55:00			

	Indicador	Tipo	Comisión	Categoría	Mínimo	Máximo	Último	
118	Pregunta P6_A.2.2	Durante la semana pasada, ¿usted encendió el fogón, hornos o arafes de leña o carbón para preparar o calentar alimentos?			1	C	p6_A_2_2	1
119	Pregunta P6_A.2.1	- Cuanto tiempo le dedicó de lunes a viernes? - HORAS			2	C	p6_A_2_2_1	00-60
120	Pregunta P6_A.2.2	- Cuanto tiempo le dedicó de lunes a viernes? - MINUTOS			2	C	p6_A_2_2_2	nan
121	Pregunta P6_A.2.3	Cuanto tiempo le dedicó sábado y domingo? - HORAS			2	C	p6_A_2_2_3	00-40
122	Pregunta P6_A.2.4	- Cuanto tiempo le dedicó sábado y domingo? - MINUTOS			2	C	p6_A_2_2_4	00-54
123	Pregunta P6_A.3.1	Durante la semana pasada, ¿usted cocinó, preparó y/o sirvió alimentos o bebidas?			1	C	p6_A_3_1	12
124	Pregunta P6_A.3.1	- Cuanto tiempo le dedicó de lunes a viernes? - HORAS			2	C	p6_A_3_1	00-92
125	Pregunta P6_A.3.2	- Cuanto tiempo le dedicó de lunes a viernes? - MINUTOS			2	C	p6_A_3_2	nan
126	Pregunta P6_A.3.3	Cuanto tiempo le dedicó sábado y domingo? - HORAS			2	C	p6_A_3_3	00-45
127	Pregunta P6_A.3.4	- Cuanto tiempo le dedicó sábado y domingo? - MINUTOS			2	C	p6_A_3_4	00-59
128	Pregunta P6_A.4.4	Durante la semana pasada, ¿usted sirvió la comida, recogió, lavó, secó o acomodó los trastes?			1	C	p6_A_4_4	12
129	Pregunta P6_A.4.1	- Cuanto tiempo le dedicó de lunes a viernes? - HORAS			2	C	p6_A_4_4	00-92
130	Pregunta P6_A.4.2	- Cuanto tiempo le dedicó de lunes a viernes? - MINUTOS			2	C	p6_A_4_5	nan
131	Pregunta P6_A.4.3	Cuanto tiempo le dedicó sábado y domingo? - HORAS			2	C	p6_A_4_5	00-44
132	Pregunta P6_A.4.4	- Cuanto tiempo le dedicó sábado y domingo? - MINUTOS			2	C	p6_A_4_5	00-55
133	Pregunta P6_A.5	Durante la semana pasada, ¿usted llevó comida a algún integrante de su hogar a la escuela, trabajo u otro lugar?			1	C	p6_A_5	p6_A_5 12
134	Pregunta P6_A.5.1	Cuanto tiempo le dedicó de lunes a viernes? - HORAS			2	C	p6_A_5_1	nan
135	Pregunta P6_A.5.2	Cuanto tiempo le dedicó de lunes a viernes? - MINUTOS			2	C	p6_A_5_2	00-59
136	Pregunta P6_A.5.3	Cuanto tiempo le dedicó sábado y domingo? - HORAS			2	C	p6_A_5_3	nan
137	Pregunta P6_A.5.4	- Cuanto tiempo le dedicó sábado y domingo? - MINUTOS			2	C	p6_A_5_4	00-50
138	Pregunta P6_A.6	Durante la semana pasada, ¿usted bendijo la bandera de su país o pidió de su vivienda?			1	C	p6_A_6	p6_A_6 1
139	Pregunta P6_A.6.1	- Cuanto tiempo le dedicó de lunes a viernes? - HORAS			2	C	p6_A_6_1	00-99
140	Pregunta P6_A.6.2	- Cuanto tiempo le dedicó de lunes a viernes? - MINUTOS			2	C	p6_A_6_2	nan
141	Pregunta P6_A.6.3	Cuanto tiempo le dedicó sábado y domingo? - HORAS			2	C	p6_A_6_3	nan
142	Pregunta P6_A.6.4	- Cuanto tiempo le dedicó sábado y domingo? - MINUTOS			2	C	p6_A_6_4	00-50

	Diccionario Original						
	entidad	concepto	definición	valor	referencia	origen	rango_cierre
143	Pregunta_P6.S.2.2	Durante la semana pasada, ¿usted limpia o recogió el interior de su vivienda? (ordenar objetos, tender camas, barrer, trapear, sacudir, lavar la cocina, el baño, entre otros)				1	C p6_s2_5_2 man .5-2 1,2
144	Pregunta_P6.S.2.1_1	Cuánto tiempo le dedica de lunes a viernes? - HORAS				2	C p6_s4_2_1 man .00-.80
145	Pregunta_P6.S.2.2_2	Cuánto tiempo le dedica de lunes a viernes? - MINUTOS				2	C p6_s4_2_2 man .00-.55
146	Pregunta_P6.S.2.3_1	Cuánto tiempo le dedica sábado y domingo? - HORAS				2	C p6_s4_2_3 man .00-.40
147	Pregunta_P6.S.2.4_2	Cuánto tiempo le dedica sábado y domingo? - MINUTOS				2	C p6_s4_2_4 man .00-.58
148	Pregunta_P6.S.3	Durante la semana pasada, ¿usted recogió, separó, tiró o quemó la basura?				1	C p6_s3_3 man p6_s3_3 1,2
149	Pregunta_P6.SA.3_1	Cuánto tiempo le dedicó de lunes a viernes? - HORAS				2	C p6_s4_3_1 man .00-.99
150	Pregunta_P6.SA.3_2	Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS				2	C p6_s4_3_2 man .00-.99
151	Pregunta_P6.S.3_3	Cuánto tiempo le dedicó sábado y domingo? - HORAS				2	C p6_s4_3_3 man .00-.40
152	Pregunta_P6.SA.4_1	Cuánto tiempo le dedicó sábado y domingo? - MINUTOS				2	C p6_s4_3_4 man .00-.55
153	Pregunta_P6.S.4_2	Durante la semana pasada, ¿cuál cultivo o regó manzanas y plantas de su patio o jardín?				1	C p6_s4_4 man p6_s4_4 1,2
154	Pregunta_P6.SA.4_1_1	Cuánto tiempo le dedicó de lunes a viernes? - HORAS				2	C p6_s4_4_1 man .00-.99
155	Pregunta_P6.SA.4_2_2	Cuánto tiempo le dedicó sábado y domingo? - HORAS				2	C p6_s4_4_2 man .00-.55
156	Pregunta_P6.SA.4_3_3	Cuánto tiempo le dedicó sábado y domingo? - MINUTOS				2	C p6_s4_4_3 man .00-.40
157	Pregunta_P6.SA.4_4_4	Cuánto tiempo le dedicó sábado y domingo? - MINUTOS				2	C p6_s4_4_4 man .00-.55
158	Pregunta_P6.S.5	Durante la semana pasada, ¿usted limpia, alimentó o cuidó a la(s) mascota(s) (animales de compañía) de su hogar?				1	C p6_s5_5 man p6_s5_5 1,2
159	Pregunta_P6.SA.5_1_1	Cuánto tiempo le dedicó de lunes a viernes? - HORAS				2	C p6_s6_5_1 man .00-.99
160	Pregunta_P6.SA.5_2_2	Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS				2	C p6_s6_5_2 man .00-.55
161	Pregunta_P6.SA.5_3_3	Cuánto tiempo le dedicó sábado y domingo? - HORAS				2	C p6_s6_5_3 man .00-.40
162	Pregunta_P6.SA.5_4_4	Cuánto tiempo le dedicó sábado y domingo? - MINUTOS				2	C p6_s6_5_4 man .00-.50
163	Pregunta_P6.I.1	¿Cuánto tiempo le dedica a su trabajo, estudio, tarea, hobby o a hacer la casa? (si lo hizo con máquina, quite el tiempo de operación)				1	C p6_i_1_1 man p6_i_1_1 1,2
164	Pregunta_P6.IA.1_1	Cuánto tiempo le dedica de lunes a viernes? - HORAS				2	C p6_i_1_1_1 man .00-.90
165	Pregunta_P6.IA.1_2_2	Cuánto tiempo le dedica de lunes a viernes? - MINUTOS				2	C p6_i_1_2 man .00-.55
166	Pregunta_P6.IA.1_3_3	Cuánto tiempo le dedica sábado y domingo? - HORAS				2	C p6_i_1_3 man .00-.45
167	Pregunta_P6.IA.1_4_4	Cuánto tiempo le dedica sábado y domingo? - MINUTOS				2	C p6_i_1_4 man .00-.56

	nombre_campo	longitud	tipo	referencia	catalogo	rango_descrip
168	Pregunta_P6.2.2 Durante la semana pasada, ¿usted planchó la ropa?	1	C	p6_8_2	p6_8_2	1,2
169	Pregunta_P6.2.1 ¿Cuánto tiempo le dedico a lunes y viernes? - HORAS	2	C	p6_8_2_1	nan	00-55
170	Pregunta_P6.6.2.2 ¿Cuánto tiempo le dedico de lunes a viernes? - MINUTOS	2	C	p6_6_2_2	nan	00-55
171	Pregunta_P6.6.2.3 ¿Cuánto tiempo le dedico sábado y domingo? - HORAS	2	C	p6_6_2_3	nan	00-40
172	Pregunta_P6.6.2.4 ¿Cuánto tiempo le dedico sábado y domingo? - MINUTOS	2	C	p6_6_2_4	nan	00-50
173	Pregunta_P6.6.3 Durante la semana pasada, ¿usted separó, dobló, arrolló o guardó la ropa?	1	C	p6_6_3	p6_6_3	1,2
174	Pregunta_P6.6.3.1 ¿Cuánto tiempo le dedico de lunes a viernes? - HORAS	2	C	p6_6_3_1	nan	00-50
175	Pregunta_P6.6.3.2 ¿Cuánto tiempo le dedico de lunes a viernes? - MINUTOS	2	C	p6_6_3_2	nan	00-59
176	Pregunta_P6.6.3.3 ¿Cuánto tiempo le dedico sábado y domingo? - HORAS	2	C	p6_6_3_3	nan	00-45
177	Pregunta_P6.6.3.4 ¿Cuánto tiempo le dedico sábado y domingo? - MINUTOS	2	C	p6_6_3_4	nan	00-45
178	Pregunta_P6.6.4 Durante la semana pasada, ¿usted arregló o remendó la ropa, mantelería, cortinas o sábanas (excluya confección)?	1	C	p6_6_4	p6_6_4	1,2
179	Pregunta_P6.6.4.1 ¿Cuánto tiempo le dedico de lunes a viernes? - HORAS	2	C	p6_6_4_1	nan	00-30
180	Pregunta_P6.6.4.2 ¿Cuánto tiempo le dedico de lunes a viernes? - MINUTOS	2	C	p6_6_4_2	nan	00-55
181	Pregunta_P6.6.4.3 ¿Cuánto tiempo le dedico sábado y domingo? - HORAS	2	C	p6_6_4_3	nan	00-09
182	Pregunta_P6.6.4.4 ¿Cuánto tiempo le dedico sábado y domingo? - MINUTOS	2	C	p6_6_4_4	nan	00-55
183	Pregunta_P6.6.5 Durante la semana pasada, ¿usted limpó, bocinó o pintó el cabello? (tenis, uarrachones, botas, etcétera)	1	C	p6_6_5	p6_6_5	1,2
184	Pregunta_P6.6.5.1 ¿Cuánto tiempo le dedico de lunes a viernes? - HORAS	2	C	p6_6_5_1	nan	00-30
185	Pregunta_P6.6.5.2 ¿Cuánto tiempo le dedico de lunes a viernes? - MINUTOS	2	C	p6_6_5_2	nan	00-59
186	Pregunta_P6.6.5.3 ¿Cuánto tiempo le dedico sábado y domingo? - HORAS	2	C	p6_6_5_3	nan	00-40
187	Pregunta_P6.6.5.4 ¿Cuánto tiempo le dedico sábado y domingo? - MINUTOS	2	C	p6_6_5_4	nan	00-50
188	Pregunta_P6.7.1 Durante la semana pasada, ¿usted reparó o hizo alguna instalación menor a su vivienda?	1	C	p6_7_1	p6_7_1	1,2
189	Pregunta_P6.7.1.1 ¿Cuánto tiempo le dedico de lunes a viernes? - HORAS	2	C	p6_7_1_1	nan	00-40
190	Pregunta_P6.7.1.2 ¿Cuánto tiempo le dedico de lunes a viernes? - MINUTOS	2	C	p6_7_1_2	nan	00-51
191	Pregunta_P6.7.1.3 ¿Cuánto tiempo le dedico sábado y domingo? - HORAS	2	C	p6_7_1_3	nan	00-20
192	Pregunta_P6.7.1.4 ¿Cuánto tiempo le dedico sábado y domingo? - MINUTOS	2	C	p6_7_1_4	nan	00-50

Diccionario Original					
nombre_campo	longitud_campo	tipo	referencia	ultimo	rango_cierre
194 Pregunta P6.TA.2.2 Durante la semana pasada, ¿usted reparó muebles, juguetes, aparatos domésticos o computadora de su hogar?	1	C	p6_7a_2_2	p6_7a_2_2	1,2
194 Pregunta P6.TA.2.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_7a_2_1	nan	00-30
195 Pregunta P6.TA.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_7a_2_2	nan	00-50
196 Pregunta P6.TA.2.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_7a_2_3	nan	00-30
197 Pregunta P6.TA.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_7a_2_4	nan	00-45
198 Pregunta P6.TA.3.1 ¿Cuánto tiempo le dedicó de lunes a viernes o algún medio de transporte de su hogar? (bicicleta, moto, camioneta, automóvil)	1	D	p6_7a_3_1	p6_7a_3_1	1
198 Pregunta P6.TA.3.2 ¿Cuánto tiempo le dedicó de lunes a viernes o algún medio de transporte de su hogar? (bicicleta, moto, camioneta, automóvil)	2	C	p6_7a_3_1	nan	00-33
199 Pregunta P6.TA.3.3 ¿Cuánto tiempo le dedicó de lunes a viernes o algún medio de transporte de su hogar? (bicicleta, moto, camioneta, automóvil)	2	C	p6_7a_3_2	nan	00-50
200 Pregunta P6.TA.3.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_7a_3_3	nan	00-40
201 Pregunta P6.TA.3.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_7a_3_4	nan	00-55
202 Pregunta P6.TA.3.6 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_7a_4	nan	00-55
203 Pregunta P6.TA.4.1 Durante la semana pasada, ¿usted reparó o dio mantenimiento a algún medio de transporte de su hogar? (bicicleta, moto, camioneta, automóvil)	1	C	p6_7a_4_1	p6_7a_4_1	1,2
204 Pregunta P6.TA.4.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_7a_4_1	nan	00-30
205 Pregunta P6.TA.4.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_7a_4_2	nan	00-50
206 Pregunta P6.TA.4.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_7a_4_3	nan	00-12
207 Pregunta P6.TA.4.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_7a_4_4	nan	00-50
208 Pregunta P6.BA.1.1 Durante la semana pasada, ¿usted buscó o compró refacciones, llantas, herramientas o materiales de construcción, automóvil, casa o terreno?	1	C	p6_Ba_1_1	p6_Ba_1_1	1
209 Pregunta P6.BA.1.2 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_Ba_1_1	nan	00-30
210 Pregunta P6.BA.1.3 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_Ba_1_2	nan	00-55
211 Pregunta P6.BA.1.4 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_Ba_1_3	nan	00-45
212 Pregunta P6.BA.1.5 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_Ba_1_4	nan	00-55
213 Pregunta P6.BA.2.1 Durante la semana pasada, ¿usted buscó o compró refacciones, llantas, herramientas o materiales de construcción, automóvil, casa o terreno?	1	C	p6_Ba_2_1	p6_Ba_2_1	1,2
214 Pregunta P6.BA.2.2 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_Ba_2_1	nan	00-30
215 Pregunta P6.BA.2.3 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_Ba_2_2	nan	00-55
216 Pregunta P6.BA.2.4 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_Ba_2_3	nan	00-45
217 Pregunta P6.BA.2.5 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_Ba_2_4	nan	00-55

Diccionario Original					
nombre_campo	longitud_campo	tipo	referencia	ultimo	rango_cierre
218 Pregunta P6.BA.3.3 Durante la semana pasada, ¿usted buscó o compró artículos o bienes para su hogar como trastes, manteles, muebles, ropa, calzado u otros?	1	C	p6_Ba_3_3	p6_Ba_3_3	1,2
219 Pregunta P6.BA.3.1 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_Ba_3_1	nan	00-45
220 Pregunta P6.BA.3.2 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_Ba_3_2	nan	00-55
221 Pregunta P6.BA.3.3 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_Ba_3_3	nan	00-30
222 Pregunta P6.BA.3.4 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_Ba_3_4	nan	00-40
223 Pregunta P6.BA.1.1 Durante la semana pasada, ¿usted hizo pagos o trámites de servicios para su hogar? (incluya si lo hizo por internet) (tenencia, predial, agua, luz, credencial de elector, colegiatura, crédito hipotecario, caja de ahorro, renta, actas, CURP, pasaporte, denuncias, etcétera)	1	C	p6_Ba_1_1	p6_Ba_1_1	1,2
224 Pregunta P6.BA.1.2 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_Ba_1_1	nan	00-30
225 Pregunta P6.BA.1.3 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_Ba_1_2	nan	00-55
226 Pregunta P6.BA.1.4 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_Ba_1_3	nan	00-40
227 Pregunta P6.BA.1.5 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_Ba_1_4	nan	00-55
228 Pregunta P6.BA.2.1 Durante la semana pasada, ¿usted llevó o supervisó la reparación de su vivienda?	1	C	p6_Ba_2_1	p6_Ba_2_1	1,2
229 Pregunta P6.BA.2.2 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_Ba_2_1	nan	00-50
230 Pregunta P6.BA.2.3 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_Ba_2_2	nan	00-55
231 Pregunta P6.BA.2.4 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_Ba_2_3	nan	00-30
232 Pregunta P6.BA.2.5 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_Ba_2_4	nan	00-50
233 Pregunta P6.BA.3.1 Durante la semana pasada, ¿usted hizo pagos o trámites de servicios para su hogar? (incluya si lo hizo por internet) (tenencia, predial, agua, luz, credencial de elector, colegiatura, crédito hipotecario, caja de ahorro, renta, actas, CURP, pasaporte, denuncias, etcétera)	1	C	p6_Ba_3_1	p6_Ba_3_1	1,2
234 Pregunta P6.BA.3.2 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_Ba_3_1	nan	00-30
235 Pregunta P6.BA.3.3 Incluyendo el tránsito, ¿cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_Ba_3_2	nan	00-50
236 Pregunta P6.BA.3.4 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_Ba_3_3	nan	00-50
237 Pregunta P6.BA.3.5 Incluyendo el tránsito, ¿cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_Ba_3_4	nan	00-45
238 Pregunta P6.10.1.1 Durante la semana pasada, ¿usted llevó o repartió algo rosas o pétalos a algún lugar para su limpieza o reparación?	1	C	p6_10a_1_1	p6_10a_1_1	1,2
239 Pregunta P6.10.1.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_10a_1_2	nan	00-10
240 Pregunta P6.10.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_10a_1_4	nan	00-50
241 Pregunta P6.10.1.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_1_3	nan	00-30
242 Pregunta P6.10.1.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_1_4	nan	00-50

Diccionario Original					
nombre_campo	longitud_campo	tipo	referencia	ultimo	rango_cierre
243 Pregunta P6.10.2 Durante la semana pasada, ¿usted supervisó la construcción, reparación o mantenimiento de su vivienda?	1	C	p6_10a_2	p6_10a_2	1,2
244 Pregunta P6.10a.2.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_10a_2_1	nan	00-50
245 Pregunta P6.10a.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_10a_2_2	nan	00-50
246 Pregunta P6.10a.2.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_10a_2_3	nan	00-24
247 Pregunta P6.10a.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_2_4	nan	00-45
248 Pregunta P6.10.3 Durante la semana pasada, ¿usted realizó o supervisó la reparación de muebles, juguetes, aparatos domésticos o computadoras de su hogar?	1	C	p6_10a_3	p6_10a_3	1,2
249 Pregunta P6.10a.3.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_10a_3	nan	00-50
250 Pregunta P6.10a.3.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_10a_3_2	nan	00-55
251 Pregunta P6.10a.3.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_10a_3_3	nan	00-10
252 Pregunta P6.10a.3.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_3_4	nan	00-45
253 Pregunta P6.10.4 Durante la semana pasada, ¿usted llevó o supervisó la reparación de su auto?	1	C	p6_10a_4	p6_10a_4	1,2
254 Pregunta P6.10a.4.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_10a_4_1	nan	00-48
255 Pregunta P6.10a.4.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_10a_4_2	nan	00-55
256 Pregunta P6.10a.4.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_10a_4_3	nan	00-40
257 Pregunta P6.10a.4.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_4_4	nan	00-50
258 Pregunta P6.10.5 Durante la semana pasada, ¿usted realizó o supervisó otras medidas para proteger sus bienes y su vivienda? (guardó el auto, encendió la alarma)	1	C	p6_10a_5	p6_10a_5	1,2
259 Pregunta P6.10a.5.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_10a_5_1	nan	00-30
260 Pregunta P6.10a.5.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_10a_5_2	nan	00-59
261 Pregunta P6.10a.5.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_10a_5_3	nan	00-40
262 Pregunta P6.10a.5.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_5_4	nan	00-51
263 Pregunta P6.10.6 Durante la semana pasada, ¿usted organizó o repartió los quehaceres de su hogar? (indicó qué hacer de comer, supervisó la limpieza de su vivienda, etcétera)	1	C	p6_10a_6	p6_10a_6	1,2
264 Pregunta P6.10a.6.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_10a_6_1	nan	00-30
265 Pregunta P6.10a.6.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_10a_6_2	nan	00-51
266 Pregunta P6.10a.6.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_10a_6_3	nan	00-30
267 Pregunta P6.10a.6.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_6_4	nan	00-50

Diccionario Original					
nombre_campo	longitud_campo	tipo	referencia	ultimo	rango_cierre
268 Pregunta P6.10.7 Durante la semana pasada, ¿usted organizó o repartió los quehaceres de su hogar? (indicó qué hacer de comer, supervisó la limpieza de su vivienda, etcétera)	1	C	p6_10a_7	p6_10a_7	1,2
269 Pregunta P6.10a.7.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_10a_7_1	nan	00-60
270 Pregunta P6.10a.7.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_10a_7_2	nan	00-55
271 Pregunta P6.10a.7.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_10a_7_3	nan	00-40
272 Pregunta P6.10a.7.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_10a_7_4	nan	00-50

Los coloque todos juntos porque la parte de los quehaceres del hogar toma mnemónicos muy variados

preguntas fusionadas en quehaceres domésticos LV y SD:

- Pregunta P6.4.1 Durante la semana pasada, ¿usted desgranó maíz, coció o molvió el nixtamal o hizo tortillas de maíz o trigo para su hogar?
- Pregunta P6.4.2 Durante la semana pasada, ¿usted encendió el fogón, horno o anafré de leña o carbón para preparar o calentar alimentos?

- Pregunta P6.4.3 Durante la semana pasada, ¿usted cocinó, preparó o calentó alimentos o bebidas?
- Pregunta P6.4.4 Durante la semana pasada, ¿usted sirvió la comida, recogió, lavó, secó o acomodó los trastes?
- Pregunta P6.5.1 Durante la semana pasada, ¿usted barrió la banqueta, cochera o patio de su vivienda?
- Pregunta P6.5.2 Durante la semana pasada, ¿usted limpió o recogió el interior de su vivienda? (ordenar objetos, tender camas, barrer, trapear, sacudir, lavar la cocina, el baño, entre otros)
- Pregunta P6.5.3 Durante la semana pasada, ¿usted recogió, separó, tiró o quemó la basura?
- Pregunta P6.5.4 Durante la semana pasada, ¿usted cuidó o regó macetas y plantas de su patio o jardín?
- Pregunta P6.6.1 Durante la semana pasada, ¿usted lavó, tendió o puso a secar la ropa? (si lo hizo con máquina, quite el tiempo de operación)
- Pregunta P6.6.2 Durante la semana pasada, ¿usted planchó la ropa?
- Pregunta P6.6.3 Durante la semana pasada, ¿usted separó, dobló, acomodó o guardó la ropa?
- Pregunta P6.6.4 Durante la semana pasada, ¿usted arregló o remendó la ropa, manteles, cortinas o sábanas (excluya confección)?
- Pregunta P6.10.5 Durante la semana pasada, ¿usted cerró puertas, ventanas, puso candados u otras medidas para proteger sus bienes y su vivienda? (guardó el auto, encendió la alarma)
- Pregunta P6.10.6 Durante la semana pasada, ¿usted esperó el gas, la pipa de agua, el camión de basura u otro servicio sin hacer otra actividad?

- Pregunta P6.3.7 Durante la semana pasada, SÓLO para el consumo de su hogar, ¿usted elaboró alimentos para conservarse o almacenarse? (mermeladas, conservas, encurtidos, salar carne, chocolate, pan, quesos u otros)

```

167 # Pregunta P6_4_1, P6_4_2, P6_4_3, P6_4_4, P6_5_1, P6_5_2, P6_5_3, P6_5_4, P6_6_1, P6_6_2, P6_6_3, P6_6_4, P6_10_5, P6_10_6, P6_3_7
168 dataset_limpio["QuehaceresDomesticosLV"] = dataset["P6_4A_1_1"] + (dataset["P6_4A_1_2"] / 60) + dataset["P6_4A_2_1"] + (dataset["P6_4A_2_2"] / 60) + dataset
169 ["P6_4A_3_1"] + (dataset["P6_4A_3_2"] / 60) + dataset["P6_4A_4_1"] + (dataset["P6_4A_4_2"] / 60) + dataset["P6_5A_1_1"] + (dataset["P6_5A_1_2"] / 60) + dataset
170 ["P6_5A_2_1"] + (dataset["P6_5A_2_2"] / 60) + dataset["P6_5A_3_1"] + (dataset["P6_5A_3_2"] / 60) + dataset["P6_5A_4_1"] + (dataset["P6_5A_4_2"] / 60) + dataset
171 ["P6_5A_5_1"] + (dataset["P6_5A_5_2"] / 60) + dataset["P6_6A_1_1"] + (dataset["P6_6A_1_2"] / 60) + dataset["P6_6A_2_1"] + (dataset["P6_6A_2_2"] / 60) + dataset
172 ["P6_6A_3_1"] + (dataset["P6_6A_3_2"] / 60) + dataset["P6_6A_4_1"] + (dataset["P6_6A_4_2"] / 60) + dataset["P6_10A_5_1"] + (dataset["P6_10A_5_2"] / 60) + dataset
173 ["P6_10A_6_1"] + (dataset["P6_10A_6_2"] / 60) + dataset["P6_10A_7_1"] + (dataset["P6_10A_7_2"] / 60)
174 dataset_limpio["QuehaceresDomesticosSD"] = dataset["P6_4A_3_1"] + (dataset["P6_4A_3_2"] / 60) + dataset["P6_4A_4_1"] + (dataset["P6_4A_4_2"] / 60) + dataset
175 ["P6_5A_1_1"] + (dataset["P6_5A_1_2"] / 60) + dataset["P6_5A_2_1"] + (dataset["P6_5A_2_2"] / 60) + dataset["P6_5A_3_1"] + (dataset["P6_5A_3_2"] / 60) + dataset
176 ["P6_5A_4_1"] + (dataset["P6_5A_4_2"] / 60) + dataset["P6_5A_5_1"] + (dataset["P6_5A_5_2"] / 60) + dataset["P6_6A_1_1"] + (dataset["P6_6A_1_2"] / 60) + dataset
177 ["P6_6A_2_1"] + (dataset["P6_6A_2_2"] / 60) + dataset["P6_6A_3_1"] + (dataset["P6_6A_3_2"] / 60) + dataset["P6_6A_4_1"] + (dataset["P6_6A_4_2"] / 60) + dataset
178 ["P6_10A_5_1"] + (dataset["P6_10A_5_2"] / 60) + dataset["P6_10A_6_1"] + (dataset["P6_10A_6_2"] / 60)
179 dataset_limpio["TiempoGanaderiaAutoconsumoLV"] = dataset["P6_3A_1_1"] + (dataset["P6_3A_1_2"] / 60)
180 dataset_limpio["TiempoGanaderiaAutoconsumoSD"] = dataset["P6_3A_1_3"] + (dataset["P6_3A_1_4"] / 60)
181 dataset_limpio["TiempoActividadesRecolecciónLV"] = dataset["P6_3A_2_1"] + (dataset["P6_3A_2_2"] / 60) + dataset["P6_3A_3_1"] + (dataset["P6_3A_3_2"] / 60) +
182 dataset["P6_3A_5_1"] + (dataset["P6_3A_5_2"] / 60)
183 dataset_limpio["TiempoActividadesRecolecciónSD"] = dataset["P6_3A_2_3"] + (dataset["P6_3A_2_4"] / 60) + dataset["P6_3A_3_3"] + (dataset["P6_3A_3_4"] / 60) +
184 dataset["P6_3A_5_3"] + (dataset["P6_3A_5_4"] / 60)
185
186 dataset_limpio["TiempoAgriculturaAutoconsumoLV"] = dataset["P6_3A_4_1"] + (dataset["P6_3A_4_2"] / 60)
187 dataset_limpio["TiempoAgriculturaAutoconsumoSD"] = dataset["P6_3A_4_3"] + (dataset["P6_3A_4_4"] / 60)
188
189 # pregunta p6_3_6
190 # para la p6_3a_6_1, p6_3a_6_2, p6_3a_6_3, p6_3a_6_4
191 dataset_limpio["TiempoConfecciónCaseríaLV"] = dataset["P6_3A_6_1"] + (dataset["P6_3A_6_2"] / 60)
192 dataset_limpio["TiempoConfecciónCaseríaSD"] = dataset["P6_3A_6_3"] + (dataset["P6_3A_6_4"] / 60)
193
194
195
196

```

La operación de fusión para la columna de quehaceres domésticos a pesar de parecer desafiante realmente fue simple solo que se fusionaron demasiadas preguntas del dataset original, el resto es una fusión normal de 2 a 4 columnas.

Entre las preguntas que se mostraron para la fusión de quehaceres del hogar se encuentran algunas que se fusionaron en

- Tiempo mantenimiento del hogar LV y SD:
  - Pregunta P6.3.8 Durante la semana pasada, SÓLO para el consumo de su hogar, ¿usted hizo muebles, utensilios de cocina, blocks, adobes u otros productos?
  - Pregunta P6.3.9 Durante la semana pasada, SÓLO para el consumo de su hogar, ¿usted amplió o remodeló usted mismo(a) su vivienda o la estuvo construyendo?
  - Pregunta P6.7.1 Durante la semana pasada, ¿usted reparó o hizo alguna instalación menor a su vivienda?
  - Pregunta P6.7.2 Durante la semana pasada, ¿usted reparó muebles, juguetes, aparatos domésticos o computadora de su hogar?

- Pregunta P6.7.3 Durante la semana pasada, ¿usted lavó o limpió algún medio de transporte de su hogar? (bicicleta, moto, camioneta, automóvil)
  - Pregunta P6.7.4 Durante la semana pasada, ¿usted reparó o dio mantenimiento a algún medio de transporte de su hogar? (bicicleta, moto, camioneta, automóvil)
- Tiempo compras del hogar LV y SD:
  - Pregunta P6.8.1 Durante la semana pasada, ¿usted buscó o compró refacciones, llantas, herramientas o materiales de construcción, automóvil, casa o terreno?
  - Pregunta P6.8.2 Durante la semana pasada, ¿usted buscó o hizo las compras del mandado, la despensa, papelería, medicinas o artículos de limpieza?
  - Pregunta P6.8.3 Durante la semana pasada, ¿usted buscó o compró artículos o bienes para su hogar como trastes, manteles, muebles, ropa, calzado u otros?
- Tiempo trámites pagos LV y SD
  - Pregunta P6.9.1 Durante la semana pasada, ¿usted hizo pagos o trámites de servicios para su hogar? (incluya si lo hizo por internet) (tenencia, predial, agua, luz, credencial de elector, colegiatura, crédito hipotecario, caja de ahorro, renta, actas, CURP, pasaporte, denuncias, etcétera)
  - Pregunta P6.9.3 Durante la semana pasada, ¿usted tramitó o cobró algún programa social? (pensión para adultos mayores, tarjeta LICONSA, etcétera)
- Tiempo organización hogar LV y SD:
  - Pregunta P6.9.2 Durante la semana pasada, ¿usted llevó los gastos (cuentas) de su hogar?

- Pregunta P6.10.7 Durante la semana pasada, ¿usted organizó o repartió los quehaceres de su hogar? (indicó qué hacer de comer, supervisó la limpieza de su vivienda, etcétera)
- Tiempo mantenimiento hogar terceros LV y SD:
  - Pregunta P6.10.1 Durante la semana pasada, ¿usted llevó o recogió ropa o calzado a algún lugar para su limpieza o reparación?
  - Pregunta P6.10.2 Durante la semana pasada, ¿usted supervisó la construcción, reparación o mantenimiento de su vivienda?
  - Pregunta P6.10.3 Durante la semana pasada, ¿usted llevó o supervisó la reparación de muebles, juguetes, aparatos domésticos o computadora de su hogar?

```

187 # pregunta p6_3_8, p6_3_9, p6_7_1, p6_7_2, p6_7_3, p6_7_4
188 # para la p6_3a_7_1, p6_3a_7_2, p6_3a_7_3, p6_3a_7_4, p6_3a_8_1, p6_3a_8_2, p6_3a_8_3, p6_3a_8_4, p6_3a_9_1, p6_3a_9_2, p6_3a_9_3, p6_3a_9_4, p6_3a_10_1,
189 p6_3a_10_2, p6_3a_10_3, p6_3a_10_4
190 dataset_limpio["TiempoMantenimientoHogarLV"] = dataset["P6_3A_8_1"] + (dataset["P6_3A_8_2"] / 60) + dataset["P6_3A_9_1"] + (dataset["P6_3A_9_2"] / 60) + dataset
191 ["P6_7A_1_1"] + (dataset["P6_7A_1_2"] / 60) + dataset["P6_7A_2_1"] + (dataset["P6_7A_2_2"] / 60) + dataset["P6_7A_3_1"] + (dataset["P6_7A_3_2"] / 60) + dataset
192 ["P6_7A_4_1"] + (dataset["P6_7A_4_2"] / 60)
193 dataset_limpio["TiempoMantenimientoHogarSD"] = dataset["P6_3A_8_3"] + (dataset["P6_3A_8_4"] / 60) + dataset["P6_3A_9_3"] + (dataset["P6_3A_9_4"] / 60) + dataset
194 ["P6_7A_1_3"] + (dataset["P6_7A_1_4"] / 60) + dataset["P6_7A_2_3"] + (dataset["P6_7A_2_4"] / 60) + dataset["P6_7A_3_3"] + (dataset["P6_7A_3_4"] / 60) + dataset
195 ["P6_7A_4_3"] + (dataset["P6_7A_4_4"] / 60)
196
197 # pregunta p6_8_1, p6_8_2, p6_8_3
198 # para la p6_8a_1_1, p6_8a_1_2, p6_8a_1_3, p6_8a_1_4, p6_8a_2_1, p6_8a_2_2, p6_8a_2_3, p6_8a_2_4, p6_8a_3_1, p6_8a_3_2, p6_8a_3_3, p6_8a_3_4
199 dataset_limpio["TiempoComprasDelHogarLV"] = dataset["P6_8A_1_1"] + (dataset["P6_8A_1_2"] / 60) + dataset["P6_8A_2_1"] + (dataset["P6_8A_2_2"] / 60) + dataset
200 ["P6_8A_3_1"] + (dataset["P6_8A_3_2"] / 60)
201 dataset_limpio["TiempoComprasDelHogarSD"] = dataset["P6_8A_1_3"] + (dataset["P6_8A_1_4"] / 60) + dataset["P6_8A_2_3"] + (dataset["P6_8A_2_4"] / 60) + dataset
202 ["P6_8A_3_3"] + (dataset["P6_8A_3_4"] / 60)
203
204 # pregunta p6_9_1, p6_9_3
205 # para la p6_9a_1_1, p6_9a_1_2, p6_9a_1_3, p6_9a_1_4, p6_9a_3_1, p6_9a_3_2, p6_9a_3_3, p6_9a_3_4
206 dataset_limpio["TiempoTramitesPagosLV"] = dataset["P6_9A_1_1"] + (dataset["P6_9A_1_2"] / 60) + dataset["P6_9A_3_1"] + (dataset["P6_9A_3_2"] / 60)
207 dataset_limpio["TiempoTramitesPagosSD"] = dataset["P6_9A_1_3"] + (dataset["P6_9A_1_4"] / 60) + dataset["P6_9A_3_3"] + (dataset["P6_9A_3_4"] / 60)
208
209 # pregunta p6_10_1, p6_10_2, p6_10_3
210 # para la p6_10a_1_1, p6_10a_1_2, p6_10a_1_3, p6_10a_1_4, p6_10a_2_1, p6_10a_2_2, p6_10a_2_3, p6_10a_2_4, p6_10a_3_1, p6_10a_3_2, p6_10a_3_3, p6_10a_3_4,
211 p6_10a_5_1, p6_10a_5_2, p6_10a_5_3, p6_10a_5_4
212 dataset_limpio["TiempoMantenimientoHogarTercerosLV"] = dataset["P6_10A_1_1"] + (dataset["P6_10A_1_2"] / 60) + dataset["P6_10A_2_1"] + (dataset["P6_10A_2_2"] / 60)
213 + dataset["P6_10A_3_1"] + (dataset["P6_10A_3_2"] / 60)
214 dataset_limpio["TiempoMantenimientoHogarTercerosSD"] = dataset["P6_10A_1_3"] + (dataset["P6_10A_1_4"] / 60) + dataset["P6_10A_2_3"] + (dataset["P6_10A_2_4"] / 60)
215 + dataset["P6_10A_3_3"] + (dataset["P6_10A_3_4"] / 60)

```

Código con la fusión de tiempo mantenimiento hogar, compras del hogar, trámites pagos, organización hogar, mantenimiento hogar.

Pasamos a la siguiente sección del diccionario original, en esta se toma nuevamente una enorme cantidad de preguntas para fusionar.

Diccionario Original					
Nombre campo	longitud	tipo	memoria	catalogo	range clave
278 Pregunta P6.11.01 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted le(s) dio de comer o ayudó a hacerlo?	1	C	p6_11_01	p6_11_01	1,2
278 Pregunta P6.11A.01.2 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_01_1	p6_11_01	00-95
278 Pregunta P6.11A.01.3 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_01_2	p6_11_01	00-95
278 Pregunta P6.11A.01.4 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_01_3	p6_11_01	00-95
278 Pregunta P6.11A.01.5 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_01_4	p6_11_01	00-95
278 Pregunta P6.11.02 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) bañó, aseó, vistió, arregló o ayudó a hacerlo?	1	C	p6_11_02	p6_11_02	1,2
278 Pregunta P6.11A.02.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_00_1	p6_11_02	00-25
278 Pregunta P6.11A.02.2 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_00_2	p6_11_02	00-50
278 Pregunta P6.11A.02.3 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_00_3	p6_11_02	00-40
278 Pregunta P6.11A.02.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_00_4	p6_11_02	00-50
278 Pregunta P6.11.03 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) cargo, acostó o le(s) ayudó a hacerlo?	1	C	p6_11_03	p6_11_03	1,2
278 Pregunta P6.11A.03.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_03_1	p6_11_03	00-54
278 Pregunta P6.11A.03.2 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_03_2	p6_11_03	00-62
278 Pregunta P6.11A.03.3 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_03_4	p6_11_03	00-50
278 Pregunta P6.11A.03.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_03_5	p6_11_03	00-50
278 Pregunta P6.11.04 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted le(s) preparó remedios caseros o algún alimento especial?	1	C	p6_11_04	p6_11_04	1,2
278 Pregunta P6.11A.04.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_04_1	p6_11_04	00-30
278 Pregunta P6.11A.04.2 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_04_2	p6_11_04	00-50
278 Pregunta P6.11A.04.3 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_04_3	p6_11_04	00-12
278 Pregunta P6.11A.04.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_04_4	p6_11_04	00-55
278 Pregunta P6.11.05 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted le(s) dio medicamentos o checó sus síntomas? (temperatura, presión, otros)	1	C	p6_11_05	p6_11_05	1,2
278 Pregunta P6.11A.05.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_05_1	p6_11_05	00-60
278 Pregunta P6.11A.05.2 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_05_2	p6_11_05	00-50
278 Pregunta P6.11A.05.3 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_05_3	p6_11_05	00-34
278 Pregunta P6.11A.05.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_05_4	p6_11_05	00-50

Diccionario Original					
Nombre campo	longitud	tipo	memoria	catalogo	range clave
299 Pregunta P6.11.06 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) llevó, recogió o esperó para que recibieran(án) atención de salud (exámenes, visita al médico, etc.) o alguna terapia especial?	1	C	p6_11_06	p6_11_06	1,2
300 Pregunta P6.11A.06.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_06_1	p6_11_06	00-63
300 Pregunta P6.11A.06.2 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_06_2	p6_11_06	00-50
300 Pregunta P6.11A.06.3 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_06_3	p6_11_06	00-24
300 Pregunta P6.11A.06.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_06_4	p6_11_06	00-50
304 Pregunta P6.11.07 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted le(s) dio terapia especial o ayudó a realizar ejercicios?	1	C	p6_11_07	p6_11_07	1,2
305 Pregunta P6.11A.07.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_07_1	p6_11_07	00-45
305 Pregunta P6.11A.07.2 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_07_2	p6_11_07	00-55
305 Pregunta P6.11A.07.3 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_07_3	p6_11_07	00-35
305 Pregunta P6.11A.07.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_07_4	p6_11_07	00-50
305 Pregunta P6.11A.07.5 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	1	C	p6_11_08	p6_11_08	1,2
305 Pregunta P6.11A.08.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_08_1	p6_11_08	00-25
311 Pregunta P6.11A.08.2 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_08_2	p6_11_08	00-50
312 Pregunta P6.11A.08.3 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_08_3	p6_11_08	00-10
313 Pregunta P6.11A.08.4 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_08_4	p6_11_08	00-40
314 Pregunta P6.11A.08.5 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) llevó y recogió de las tareas de la escuela o trabajo?	1	C	p6_11_09	p6_11_09	1,2
315 Pregunta P6.11A.09.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_09_1	p6_11_09	00-32
315 Pregunta P6.11A.09.2 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_09_2	p6_11_09	00-50
317 Pregunta P6.11A.09.3 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_09_3	p6_11_09	00-05
318 Pregunta P6.11A.09.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_09_4	p6_11_09	00-50
319 Pregunta P6.11A.09.5 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted asistió a juntas, festivales o actividades de apoyo escolar?	1	C	p6_11_10	p6_11_10	1,2
320 Pregunta P6.11A.10.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_10_1	p6_11_10	00-12
321 Pregunta P6.11A.10.2 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_10_2	p6_11_10	00-50
322 Pregunta P6.11A.10.3 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_10_3	p6_11_10	00-07
323 Pregunta P6.11A.10.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_10_4	p6_11_10	00
324 Pregunta P6.11A.11.1 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted mientras hacía otra cosa, lo(s) cuidó o estuvo al pendiente?	1	C	p6_11_11	p6_11_11	1,2
325 Pregunta P6.11A.11.1 „Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_11a_11_1	p6_11_11	00-99
326 Pregunta P6.11A.11.2 „Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_11a_11_2	p6_11_11	00-59
327 Pregunta P6.11A.11.3 „Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_11a_11_3	p6_11_11	00-48
328 Pregunta P6.11A.11.4 „Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_11a_11_4	p6_11_11	00

- Fusión para tiempo cuidados especiales otra persona LV Y SD:
  - Pregunta P6.11.01 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted le(s) dio de comer o ayudó a hacerlo?
  - Pregunta P6.11.02 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) bañó, aseó, vistió, arregló o ayudó a hacerlo?
  - Pregunta P6.11.03 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) cargó, acostó o le(s) ayudó a hacerlo?
  - Pregunta P6.11.04 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted le(s) preparó remedios caseros o algún alimento especial?
  - Pregunta P6.11.05 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital

u otro lugar, ¿usted le(s) dio medicamentos o checó sus síntomas? (temperatura, presión, otros)

- Pregunta P6.11.06 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) llevó, recogió o esperó para que recibiera(n) atención de salud (exámenes, visitas al médico, etc.) o alguna terapia especial?
- Pregunta P6.11.07 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted le(s) dio terapia especial o ayudó a realizar ejercicios?
- Pregunta P6.11.08 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) llevó y/o recogió de clases, trabajo u otro lugar?
- Pregunta P6.11.09 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted lo(s) ayudó o apoyó en las tareas de la escuela o trabajo?
- Pregunta P6.11.10 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted asistió a juntas, festivales o actividades de apoyo escolar?
- Pregunta P6.11.11 (NOMBRE(S)) necesitó(aron) cuidados de otra persona Durante la semana pasada, sea en la casa, hospital u otro lugar, ¿usted mientras hacía otra cosa, lo(s) cuidó o estuvo al pendiente?

```

214      # Tiempo cuidados especiales para otra persona
215      # pregunta p6_11a_01, p6_11a_02, p6_11a_03, p6_11a_04, p6_11a_05, p6_11a_06, p6_11a_07, p6_11a_08, p6_11a_09, p6_11a_10, p6_11a_11
216      # para la p6_11a_01_1, p6_11a_01_2, p6_11a_01_3, p6_11a_01_4, p6_11a_02_1, p6_11a_02_2, p6_11a_02_3, p6_11a_02_4, p6_11a_03_1, p6_11a_03_2, p6_11a_03_3,
217      # p6_11a_03_4, p6_11a_04_1, p6_11a_04_2, p6_11a_04_3, p6_11a_04_4, p6_11a_05_1, p6_11a_05_2, p6_11a_05_3, p6_11a_05_4, p6_11a_06_1, p6_11a_06_2, p6_11a_06_3,
218      # p6_11a_06_4, p6_11a_07_1, p6_11a_07_2, p6_11a_07_3, p6_11a_07_4, p6_11a_08_1, p6_11a_08_2, p6_11a_08_3, p6_11a_08_4, p6_11a_09_1, p6_11a_09_2, p6_11a_09_3,
```

```

dataset_limpio("TiempoCuidadosEspecialesOtraPersonasSD") = dataset["P6_11a_01_1"] + (dataset["P6_11a_01_2"] / 60) + dataset["P6_11a_02_1"] + (dataset["P6_11a_02_2"] / 60) + dataset["P6_11a_03_1"] + (dataset["P6_11a_03_2"] / 60) + dataset["P6_11a_04_1"] + (dataset["P6_11a_04_2"] / 60) + dataset["P6_11a_05_1"] + (dataset["P6_11a_05_2"] / 60) + dataset["P6_11a_06_1"] + (dataset["P6_11a_06_2"] / 60) + dataset["P6_11a_07_1"] + (dataset["P6_11a_07_2"] / 60) + dataset["P6_11a_08_1"] + (dataset["P6_11a_08_2"] / 60) + dataset["P6_11a_09_1"] + (dataset["P6_11a_09_2"] / 60) + dataset["P6_11a_10_1"] + (dataset["P6_11a_10_2"] / 60) + dataset["P6_11a_11_1"] + (dataset["P6_11a_11_2"] / 60)
dataset_limpio("TiempoCuidadosEspecialesOtraPersonasLV") = dataset["P6_11a_01_1"] + (dataset["P6_11a_01_2"] / 60) + dataset["P6_11a_02_1"] + (dataset["P6_11a_02_2"] / 60) + dataset["P6_11a_03_1"] + (dataset["P6_11a_03_2"] / 60) + dataset["P6_11a_04_1"] + (dataset["P6_11a_04_2"] / 60) + dataset["P6_11a_05_1"] + (dataset["P6_11a_05_2"] / 60) + dataset["P6_11a_06_1"] + (dataset["P6_11a_06_2"] / 60) + dataset["P6_11a_07_1"] + (dataset["P6_11a_07_2"] / 60) + dataset["P6_11a_08_1"] + (dataset["P6_11a_08_2"] / 60) + dataset["P6_11a_09_1"] + (dataset["P6_11a_09_2"] / 60) + dataset["P6_11a_10_1"] + (dataset["P6_11a_10_2"] / 60) + dataset["P6_11a_11_1"] + (dataset["P6_11a_11_2"] / 60)
dataset_limpio("TiempoCuidadosEspecialesOtraPersonasDV") = dataset["P6_11a_01_3"] + (dataset["P6_11a_01_4"] / 60) + dataset["P6_11a_02_3"] + (dataset["P6_11a_03_1"] / 60) + dataset["P6_11a_03_4"] + (dataset["P6_11a_04_3"] / 60) + dataset["P6_11a_04_4"] + (dataset["P6_11a_05_3"] / 60) + dataset["P6_11a_05_4"] + (dataset["P6_11a_06_3"] / 60) + dataset["P6_11a_06_4"] + (dataset["P6_11a_07_3"] / 60) + dataset["P6_11a_07_4"] + (dataset["P6_11a_08_3"] / 60) + dataset["P6_11a_08_4"] + (dataset["P6_11a_09_3"] / 60) + dataset["P6_11a_09_4"] + (dataset["P6_11a_10_3"] / 60) + dataset["P6_11a_11_3"] + (dataset["P6_11a_11_4"] / 60)

```

Código de la fusión para tiempo cuidados especiales otra persona

Pasamos a la siguiente sección del diccionario original, en esta se toman solo 3 preguntas y se fusionan en una.

Diccionario Original					
	lenguaje	tipos	nombre	catalogo	origen
339 Preguntas P6.12.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_12a_1_1	nan	00:00
340 Preguntas P6.12.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_12a_1_2	nan	00:00
341 Preguntas P6.12.1 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_12a_1_3	nan	00:40
342 Preguntas P6.12.1 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_12a_1_4	nan	00:30
343 Preguntas P6.12.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) bañó, aseó (cambió pañales), vistió o amamantó?	1	C	p6_12_a_2	00:00	00:55
344 Preguntas P6.12.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_12a_2_1	nan	00:00
345 Preguntas P6.12.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_12a_2_2	nan	00:00
346 Preguntas P6.12.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) bañó, aseó (cambió pañales), vistió o arregló?	2	C	p6_12a_2_3	nan	00:00
347 Preguntas P6.12.2 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_12a_2_4	nan	00:00
348 Preguntas P6.12.2 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	1	C	p6_12_a_3	p6_12_a_2_1	00:00
349 Preguntas P6.12.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) cargó o acostó?	2	C	p6_12a_3_1	nan	00:00
350 Preguntas P6.12.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_12a_3_2	nan	00:00
351 Preguntas P6.12.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_12a_3_3	nan	00:00
352 Preguntas P6.12.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) cargó o acostó?	2	C	p6_12a_3_4	nan	00:00
353 Preguntas P6.12.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_12a_3_5	nan	00:00
354 Preguntas P6.12.3 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_12a_3_6	nan	00:00

- Fusión para Tiempo Crianza Menores 0 a 5 LV y SD:

- Pregunta P6.12.1 Durante la semana pasada, ¿usted a (NOMBRE(S)) le(s) dio de comer (amamantó) o dio de beber?
- Pregunta P6.12.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) bañó, aseó (cambió pañales), vistió o arregló?
- Pregunta P6.12.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) cargó o acostó?

# Tiempo Crianza de menores 0-5					
220 # preguntas p6_12_a_1, p6_12_a_2, p6_12_a_3					
221 # para la p6_12a_1_1, p6_12a_1_2, p6_12a_1_3, p6_12a_1_4, p6_12a_2_1, p6_12a_2_2, p6_12a_2_3, p6_12a_2_4, p6_12a_3_1, p6_12a_3_2, p6_12a_3_3, p6_12a_3_4					
222 dataset_limpiado["TiempoCrianzaMenores0a5LV"] = dataset["P6_12A_1_1"] + dataset["P6_12A_1_2"] / 60 + dataset["P6_12A_2_1"] + dataset["P6_12A_2_2"] / 60 + dataset					
223 ["P6_12A_3_1"] * (dataset["P6_12A_3_2"] / 60)					
dataset_limpiado["TiempoCrianzaMenores0a5SD"] = dataset["P6_12A_1_3"] + (dataset["P6_12A_1_4"] / 60) + dataset["P6_12A_2_3"] + (dataset["P6_12A_2_4"] / 60) + dataset					
["P6_12A_3_3"] + (dataset["P6_12A_3_4"] / 60)					

código para fusión tiempo crianza menores 0 a 5

Pasamos a la siguiente sección del diccionario y tomamos las preguntas para tiempo crianza menores 0 a 14

Diccionario Original					
	lenguaje	tipos	nombre	catalogo	origen
346 Preguntas P6.13.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	1	C	p6_13_a_1	00:00	00:00
347 Preguntas P6.13.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_13a_1_2	nan	00:00
348 Preguntas P6.13.1 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_13a_1_3	nan	00:00
349 Preguntas P6.13.1 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_13a_1_4	nan	00:00
350 Preguntas P6.13.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) asistió a jardines, fiestas o actividades de apoyo a realizar ejercicios?	1	C	p6_13_a_2	p6_13_a_1_2	00:00
351 Preguntas P6.13.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_13a_2_1	nan	00:02
352 Preguntas P6.13.2 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_13a_2_2	nan	00:00
353 Preguntas P6.13.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) asistió a jardines, fiestas o actividades de apoyo a realizar ejercicios?	1	C	p6_13_a_3	p6_13_a_2_1	00:00
354 Preguntas P6.13.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_13a_3_1	nan	00:02
355 Preguntas P6.13.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_13a_3_2	nan	00:00
356 Preguntas P6.13.4 Durante la semana pasada, ¿usted a (NOMBRE(S)) asistió a jardines, fiestas o actividades de apoyo a realizar ejercicios?	2	C	p6_13a_3_3	nan	00:00
357 Preguntas P6.13.4 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	1	C	p6_13_a_4	p6_13_a_3_2	00:00
358 Preguntas P6.13.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_13a_4_1	nan	00:00
359 Preguntas P6.13.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_13a_4_2	nan	00:00
360 Preguntas P6.13.4 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_13a_4_3	nan	00:00
361 Preguntas P6.13.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_13a_4_4	nan	00:00
362 Preguntas P6.13.5 Durante la semana pasada, ¿usted a (NOMBRE(S)) asistió a jardines, fiestas o actividades de apoyo a realizar ejercicios?	1	C	p6_13_a_5	p6_13_a_4_3	00:00
363 Preguntas P6.13.5 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_13a_5_1	nan	00:00
364 Preguntas P6.13.5 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_13a_5_2	nan	00:00
365 Preguntas P6.13.5 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_13a_5_3	nan	00:00
366 Preguntas P6.13.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_13a_5_4	nan	00:00
367 Preguntas P6.13.6 Durante la semana pasada, ¿usted a (NOMBRE(S)) vivió hasta otra casa, trajo a otro o vivió el penderijo?	1	C	p6_13_a_6	p6_13_a_5_4	00:00
368 Preguntas P6.13.6 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_13a_6_1	nan	00:00
369 Preguntas P6.13.6 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_13a_6_2	nan	00:00
370 Preguntas P6.13.6 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_13a_6_3	nan	00:00
371 Preguntas P6.13.6 Durante la semana pasada, ¿usted a (NOMBRE(S)) vivió hasta otra casa, trajo a otro o vivió el penderijo?	2	C	p6_13a_6_4	nan	00:00
372 Preguntas P6.13.7 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_13a_7	nan	00:00
373 Preguntas P6.13.7 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_13a_7_1	nan	00:00
374 Preguntas P6.13.7 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_13a_7_2	nan	00:00
375 Preguntas P6.13.7 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_13a_7_3	nan	00:00

- Fusión tiempo crianza menores 0 a 14 LV y SD

- Pregunta P6.13.1 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) llevó y/o recogió de la guardería, de clases, de la casa de algún familiar o amigo para ser cuidado?
- Pregunta P6.13.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) le(s) dio terapia especial o ayudó a realizar ejercicios?
- Pregunta P6.13.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) ayudó en las tareas de la escuela?
- Pregunta P6.13.4 Durante la semana pasada, ¿usted a (NOMBRE(S)) asistió a juntas, festivales o actividades de apoyo en la guardería o escuela?
- Pregunta P6.13.5 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) llevó, recogió o esperó para que recibiera(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera)
- Pregunta P6.13.6 Durante la semana pasada, ¿usted a (NOMBRE(S)) mientras hacía otra cosa, lo(s) cuidó o estuvo al pendiente?

```

225 # Tiempo Crianza de menores 0-14
226 # preguntas p6_13a_1_1, p6_13a_1_2, p6_13a_1_3, p6_13a_1_4, p6_13a_1_5, p6_13a_2_1, p6_13a_2_2, p6_13a_2_3, p6_13a_2_4, p6_13a_3_1, p6_13a_3_2, p6_13a_3_3, p6_13a_3_4,
227 p6_13a_4_1, p6_13a_4_2, p6_13a_4_3, p6_13a_4_4, p6_13a_5_1, p6_13a_5_2, p6_13a_5_3, p6_13a_5_4, p6_13a_6_1, p6_13a_6_2, p6_13a_6_3, p6_13a_6_4
228 dataset["TiempoCrianzaMenores0a15D"] = dataset["P6_13A_1_1"] + (dataset["P6_13A_1_2"] / 60) + dataset["P6_13A_1_3"] + (dataset["P6_13A_1_4"] / 60) + dataset["P6_13A_2_1"] + (dataset["P6_13A_2_2"] / 60) +
229 dataset["P6_13A_3_1"] + (dataset["P6_13A_3_2"] / 60) + dataset["P6_13A_4_1"] + (dataset["P6_13A_4_2"] / 60) + dataset["P6_13A_5_1"] + (dataset["P6_13A_5_2"] / 60) + dataset["P6_13A_6_1"] + (dataset["P6_13A_6_2"] / 60)
dataset["limpio["TiempoCrianzaMenores0a15D"]"] = dataset["P6_13A_1_3"] + (dataset["P6_13A_1_4"] / 60) + dataset["P6_13A_2_3"] + (dataset["P6_13A_2_4"] / 60) + dataset["P6_13A_3_3"] + (dataset["P6_13A_3_4"] / 60) + dataset["P6_13A_4_3"] + (dataset["P6_13A_4_4"] / 60) + dataset["P6_13A_5_3"] + (dataset["P6_13A_5_4"] / 60) + dataset["P6_13A_6_3"] + (dataset["P6_13A_6_4"] / 60)

```

Código para tiempo crianza menores 0 a 14 LV y SD ( fusión )

Pasamos a la siguiente sección del diccionario

Original	limpio
377 Pregunta P6.14.1 Durante la semana pasada, ¿usted a (NOMBRE(S)) llevó, recogió o asistió en el uso de la computadora, celular, internet o actividades relacionadas con sus cursos o clases?	1 C p6_14_1 p6_14_1 1 12
378 Pregunta P6.14.1.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_14a_1_1 nan 00:50
379 Pregunta P6.14.1.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_14a_1_2 nan 00:51
380 Pregunta P6.14.1.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_14a_1_3 nan 00:50
381 Pregunta P6.14.1.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_14a_1_4 nan 00:50
382 Pregunta P6.14.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) llevó, recogió o esperó para que recibiera(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera)	1 C p6_14_2 p6_14_2 1 12
383 Pregunta P6.14.2.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_14a_2_1 nan 00:48
384 Pregunta P6.14.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_14a_2_2 nan 00:49
385 Pregunta P6.14.2.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_14a_2_3 nan 00:48
386 Pregunta P6.14.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_14a_2_4 nan 00:51
387 Pregunta P6.14.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) llevó y/o recogió de clases, trabajo, de algún trámite u otro lugar?	1 C p6_14_3 p6_14_3 1 12
388 Pregunta P6.14.3.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_14a_3_1 nan 00:48
389 Pregunta P6.14.3.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_14a_3_2 nan 00:49
390 Pregunta P6.14.3.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_14a_3_3 nan 00:30
391 Pregunta P6.14.3.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_14a_3_4 nan 00:30

Fusión para tiempo acompañamiento familiar 15 a 59 LV y SD

- Pregunta P6.4.5 Durante la semana pasada, ¿usted llevó comida a algún integrante de su hogar a la escuela, trabajo u otro lugar?
- Pregunta P6.14.1 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) apoyó o asesoró en el uso de la computadora, celular, internet o actividades relacionadas con sus cursos o clases?
- Pregunta P6.14.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) llevó, recogió o esperó para que recibiera(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera)
- Pregunta P6.14.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) llevó y/o recogió de clases, trabajo, de algún trámite u otro lugar?

```

231 # Tiempo Acompañamiento familiar 15 - 59
232 # preguntas p6_4_5, p6_14_1, p6_14_2, p6_14_3
233 # para la p6_4_5, p6_4_5_2, p6_4_5_3, p6_4_5_4, p6_14a_1_1, p6_14a_1_2, p6_14a_1_3, p6_14a_1_4, p6_14a_2_1, p6_14a_2_2, p6_14a_2_3, p6_14a_2_4, p6_14a_3_1,
234 p6_14a_3_2, p6_14a_3_3, p6_14a_3_4
235 dataset_limpio[TiempoAcompañamientoFamiliar15a59LV] = dataset["P6_4A_5_1"] + (dataset["P6_4A_5_2"] / 60) + dataset["P6_14A_1_1"] + (dataset["P6_14A_1_2"] / 60)
+ dataset["P6_14A_2_1"] + (dataset["P6_14A_2_2"] / 60) + dataset["P6_14A_3_1"] + (dataset["P6_14A_3_2"] / 60)
dataset_limpio[TiempoAcompañamientoFamiliar15a59SD] = dataset["P6_4A_5_4"] / 60 + dataset["P6_14A_1_3"] + (dataset["P6_14A_1_4"] / 60)
+ dataset["P6_14A_2_3"] + (dataset["P6_14A_2_4"] / 60) + dataset["P6_14A_3_3"] + (dataset["P6_14A_3_4"] / 60)

```

Código para tiempo acompañamiento 15 a 59 LV y SD ( fusión)

### Preguntas relevantes para la siguiente fusión

Nombre Categoría	longitud	tipo	minimico	maximo	codigos	rango_categoría
395 Pregunta P6_15_1 Durante la semana pasada, ¿usted a (NOMBRE(S)) apoyó o asesoró en el uso de los computadores, celular, internet o actividades relacionadas con sus cursos o clases?	1	C	p6_15_1	p6_15_1	1,2	
396 Pregunta P6_15_2 ¿Cuánto tiempo le dedicó a las vacaciones? - MINUTOS	2	C	p6_15a_2_1	p6_15a_2_1	00 - 50	
395 Pregunta P6_15a_1_2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_15a_1_2	p6_15a_1_2	00 - 50	
396 Pregunta P6_15a_1_3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_15a_1_3	p6_15a_1_3	00 - 20	
397 Pregunta P6_15A_1_4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_15a_1_4	p6_15a_1_4	00 - 50	
398 Pregunta P6_15_3 Durante la semana pasada, ¿usted a (NOMBRE(S)) llevó, recogió o esperó para que recibieran(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera)	1	C	p6_15_3	p6_15_3	1,2	
399 Pregunta P6_15_4 Durante la semana pasada, ¿usted a (NOMBRE(S)) llevó, recogió o esperó para que recibieran(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera)	2	C	p6_15a_4_1	p6_15a_4_1	00 - 48	
400 Pregunta P6_15_2_2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_15a_2_2	p6_15a_2_2	00 - 50	
401 Pregunta P6_15A_2_2 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_15a_2_3	p6_15a_2_3	00 - 18	
402 Pregunta P6_15A_2_4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_15a_2_4	p6_15a_2_4	00 - 30	
403 Pregunta P6_15_3_1 Durante la semana pasada, ¿usted a (NOMBRE(S)) llevó, recogió o esperó para que recibieran(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera) y/o recogió del trabajo, de algún trámite u otro lugar?	1	C	p6_15_3_1	p6_15_3_1	1,2	
404 Pregunta P6_15A_3_1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_15a_3_1	p6_15a_3_1	00 - 20	
405 Pregunta P6_15A_3_2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_15a_3_2	p6_15a_3_2	00 - 50	
406 Pregunta P6_15A_3_3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_15a_3_3	p6_15a_3_3	00 - 20	
407 Pregunta P6_15A_3_4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_15a_3_4	p6_15a_3_4	00 - 30	
408 Pregunta P6_15_4 Durante la semana pasada, ¿usted a (NOMBRE(S)) llevó, recogió o esperó para que recibieran(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera)	1	C	p6_15_4	p6_15_4	1,2	
409 Pregunta P6_15A_4_1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_15a_4_1	p6_15a_4_1	00 - 99	
410 Pregunta P6_15A_4_2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_15a_4_2	p6_15a_4_2	00 - 50	
411 Pregunta P6_15A_4_3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_15a_4_3	p6_15a_4_3	00 - 48	
412 Pregunta P6_15A_4_4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_15a_4_4	p6_15a_4_4	00 - 50	

- Fusión Tiempo acompañamiento familiar 60 LV y SD:

- Pregunta P6.15.1 Durante la semana pasada, ¿usted a (NOMBRE(S)) apoyó o asesoró en el uso de la computadora, celular, internet o actividades relacionadas con sus cursos o clases?

- Pregunta P6.15.2 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) llevó, recogió o esperó para que recibiera(n) atención de salud? (vacunas, dentista, chequeo médico, etcétera)
- Pregunta P6.15.3 Durante la semana pasada, ¿usted a (NOMBRE(S)) lo(s) llevó y/o recogió del trabajo, de algún trámite u otro lugar?
- Pregunta P6.15.4 Durante la semana pasada, ¿usted a (NOMBRE(S)) mientras hacía otra cosa, lo(s) cuidó o estuvo al pendiente?

```

237 # Tiempo Acompañamiento familiar 60
238 # preguntas p6_15_1, p6_15_2, p6_15_3, p6_15_4
239 # para la p6_15a_1_1, p6_15a_1_2, p6_15a_1_3, p6_15a_1_4, p6_15a_2_1, p6_15a_2_2, p6_15a_2_3, p6_15a_2_4, p6_15a_3_1, p6_15a_3_2, p6_15a_3_3, p6_15a_3_4,
240 p6_15a_4_1, p6_15a_4_2, p6_15a_4_3, p6_15a_4_4
241 dataset_limpio["TiempoAcompañamientofamiliar60LV"] = dataset["P6_15A_1_1"] + (dataset["P6_15A_1_2"] / 60) + dataset["P6_15A_2_1"] + (dataset["P6_15A_2_2"] / 60) +
242 dataset["P6_15A_3_1"] + (dataset["P6_15A_3_2"] / 60) + dataset["P6_15A_4_1"] + (dataset["P6_15A_4_2"] / 60)
243 dataset_limpio["TiempoAcompañamientofamiliar60SP"] = dataset["P6_15A_1_3"] + (dataset["P6_15A_1_4"] / 60) + dataset["P6_15A_2_3"] + (dataset["P6_15A_2_4"] / 60) +
244 dataset["P6_15A_3_3"] + (dataset["P6_15A_3_4"] / 60) + dataset["P6_15A_4_3"] + (dataset["P6_15A_4_4"] / 60)
245
246
247

```

Código para tiempo acompañamiento familiar 60 LV y SD (fusión).

## Preguntas relevantes para la siguiente función

Indicador	Indicador	Indicador	Indicador
Indicador temporal			
413 Pregunta P6.16.1 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en los quehaceres domésticos? (preparación de alimentos, limpieza de la vivienda, lavado o planchado de ropa, etcétera)	2 C p6_16a_1_1_men 00-92	2 C p6_16a_1_2_men 00-54	2 C p6_16a_1_3_men 00-38
414 Pregunta P6.16A.1.2 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_1_4_men 00-56	2 C p6_16a_2_1_men 00-55	2 C p6_16a_2_2_men 00-55
415 Pregunta P6.16A.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? HORAS	2 C p6_16a_2_3_men 00-55	2 C p6_16a_2_4_men 00-55	2 C p6_16a_2_5_men 00-55
416 Pregunta P6.16A.1.4 ¿Cuánto tiempo le dedicó viernes y jueves? HORAS	2 C p6_16a_3_1_men 00-55	2 C p6_16a_3_2_men 00-55	2 C p6_16a_3_3_men 00-55
417 Pregunta P6.16A.2 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en las compras, pagos, trámites, reparaciones de esa vivienda?	2 C p6_16a_3_4_men 00-55	2 C p6_16a_3_5_men 00-55	2 C p6_16a_3_6_men 00-55
418 Pregunta P6.16A.2.1 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_3_7_men 00-55	2 C p6_16a_3_8_men 00-55	2 C p6_16a_3_9_men 00-55
419 Pregunta P6.16A.2.2 ¿Cuánto tiempo le dedicó de horas y minutos? MINUTOS	2 C p6_16a_3_10_men 00-55	2 C p6_16a_3_11_men 00-55	2 C p6_16a_3_12_men 00-55
420 Pregunta P6.16A.2.3 ¿Cuánto tiempo le dedicó sábado y domingo? HORAS	2 C p6_16a_3_13_men 00-55	2 C p6_16a_3_14_men 00-55	2 C p6_16a_3_15_men 00-55
421 Pregunta P6.16A.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? MINUTOS	2 C p6_16a_3_16_men 00-55	2 C p6_16a_3_17_men 00-55	2 C p6_16a_3_18_men 00-55
422 Pregunta P6.16A.3 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en la atención de personas que necesitaron cuidados por discapacidad o enfermedad? (darles de medicinas, llevarlos al doctor, entre otras)	2 C p6_16a_3_19_men 00-55	2 C p6_16a_3_20_men 00-55	2 C p6_16a_3_21_men 00-55
423 Pregunta P6.16A.3.1 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_3_22_men 00-55	2 C p6_16a_3_23_men 00-55	2 C p6_16a_3_24_men 00-55
424 Pregunta P6.16A.3.2 ¿Cuánto tiempo le dedicó sábado y domingo? HORAS	2 C p6_16a_3_25_men 00-55	2 C p6_16a_3_26_men 00-55	2 C p6_16a_3_27_men 00-55
425 Pregunta P6.16A.3.3 ¿Cuánto tiempo le dedicó sábado y domingo? MINUTOS	2 C p6_16a_3_28_men 00-55	2 C p6_16a_3_29_men 00-55	2 C p6_16a_3_30_men 00-55
426 Pregunta P6.16A.3.4 ¿Cuánto tiempo le dedicó sábado y domingo? MINUTOS	2 C p6_16a_3_31_men 00-55	2 C p6_16a_3_32_men 00-55	2 C p6_16a_3_33_men 00-55
427 Pregunta P6.16A.4 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en el cuidado de menores de 6 años? (lavando o recogiendo a la guardería o escuela, cargarlos, bañarlos, ayudando en las tareas escolares, etcétera)	2 C p6_16a_3_34_men 00-55	2 C p6_16a_3_35_men 00-55	2 C p6_16a_3_36_men 00-55
428 Pregunta P6.16A.4.1 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_3_37_men 00-55	2 C p6_16a_3_38_men 00-55	2 C p6_16a_3_39_men 00-55
429 Pregunta P6.16A.4.2 ¿Cuánto tiempo le dedicó de horas y minutos? MINUTOS	2 C p6_16a_3_40_men 00-55	2 C p6_16a_3_41_men 00-55	2 C p6_16a_3_42_men 00-55
430 Pregunta P6.16A.4.3 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_3_43_men 00-55	2 C p6_16a_3_44_men 00-55	2 C p6_16a_3_45_men 00-55
431 Pregunta P6.16A.4.4 ¿Cuánto tiempo le dedicó sábado y domingo? HORAS	2 C p6_16a_3_46_men 00-55	2 C p6_16a_3_47_men 00-55	2 C p6_16a_3_48_men 00-55
432 Pregunta P6.16A.4.5 ¿Cuánto tiempo le dedicó sábado y domingo? MINUTOS	2 C p6_16a_3_49_men 00-55	2 C p6_16a_3_50_men 00-55	2 C p6_16a_3_51_men 00-55
433 Pregunta P6.16A.5 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_3_52_men 00-55	2 C p6_16a_3_53_men 00-55	2 C p6_16a_3_54_men 00-55
434 Pregunta P6.16A.5.1 ¿Cuánto tiempo le dedicó de horas y minutos? MINUTOS	2 C p6_16a_3_55_men 00-55	2 C p6_16a_3_56_men 00-55	2 C p6_16a_3_57_men 00-55
435 Pregunta P6.16A.5.2 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_3_58_men 00-55	2 C p6_16a_3_59_men 00-55	2 C p6_16a_3_60_men 00-55
436 Pregunta P6.16A.5.3 ¿Cuánto tiempo le dedicó de horas y minutos? MINUTOS	2 C p6_16a_3_61_men 00-55	2 C p6_16a_3_62_men 00-55	2 C p6_16a_3_63_men 00-55
437 Pregunta P6.16A.5.4 ¿Cuánto tiempo le dedicó sábado y domingo? MINUTOS	2 C p6_16a_3_64_men 00-55	2 C p6_16a_3_65_men 00-55	2 C p6_16a_3_66_men 00-55
438 Pregunta P6.16 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en el cuidado de personas de 60 años y más? (lavando, recogiendo o esperando para hacer costuras, trámites, etcétera)	2 C p6_16a_4_1_men 00-55	2 C p6_16a_4_2_men 00-55	2 C p6_16a_4_3_men 00-55
439 Pregunta P6.16A.6.1 ¿Cuánto tiempo le dedicó de horas y minutos? HORAS	2 C p6_16a_4_4_men 00-55	2 C p6_16a_4_5_men 00-55	2 C p6_16a_4_6_men 00-55
440 Pregunta P6.16A.6.2 ¿Cuánto tiempo le dedicó sábado y domingo? HORAS	2 C p6_16a_4_7_men 00-55	2 C p6_16a_4_8_men 00-55	2 C p6_16a_4_9_men 00-55
441 Pregunta P6.16A.6.3 ¿Cuánto tiempo le dedicó de horas y minutos? MINUTOS	2 C p6_16a_4_10_men 00-55	2 C p6_16a_4_11_men 00-55	2 C p6_16a_4_12_men 00-55
442 Pregunta P6.16A.6.4 ¿Cuánto tiempo le dedicó sábado y domingo? MINUTOS	2 C p6_16a_4_13_men 00-55	2 C p6_16a_4_14_men 00-55	2 C p6_16a_4_15_men 00-55

## ● Fusión Tiempo apoyo otro hogar:

- Pregunta P6.16.1 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en los quehaceres domésticos? (preparación de alimentos, limpieza de la vivienda, lavado o planchado de ropa, etcétera)

- Pregunta P6.16.2 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en las compras, pagos, trámites, reparaciones de esa vivienda?
- Pregunta P6.16.3 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en la atención de personas que necesitaron cuidados por discapacidad o enfermedad? (darles su medicina, llevarlos al doctor, entre otras)
- Pregunta P6.16.4 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en el cuidado de menores de 6 años? (llevárselos o recogerlos a la guardería o estancia, cargarlos, bañarlos, ayudarlos en las tareas escolares, etcétera)
- Pregunta P6.16.5 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en el cuidado de personas de 6 a 59 años? (llevárselos o recogerlos de clases, trabajo, ayudarlos en las tareas escolares, acompañarlos durante la atención de salud, etcétera)
- Pregunta P6.16.6 Durante la semana pasada, ¿usted ayudó de manera gratuita a otro hogar en el cuidado de personas de 60 años y más? (llevárselos, recogerlos o esperarlos para hacer cobros, trámites, etcétera)

```

243 # Tiempo Apoyo a otro hogar
244 # preguntas p6_16_1, p6_16_2, p6_16_3, p6_16_4, p6_16_5, p6_16_6
245 # para la p6_16a_1_1, p6_16a_1_2, p6_16a_1_3, p6_16a_1_4, p6_16a_2_1, p6_16a_2_2, p6_16a_2_3, p6_16a_2_4, p6_16a_3_1, p6_16a_3_2, p6_16a_3_3, p6_16a_3_4,
246 p6_16a_4_1, p6_16a_4_2, p6_16a_4_3, p6_16a_4_4, p6_16a_5_1, p6_16a_5_2, p6_16a_5_3, p6_16a_5_4, p6_16a_6_1, p6_16a_6_2, p6_16a_6_3, p6_16a_6_4
dataset_limpio["TiempoApoyoOtroHogarLV"] = dataset["P6_16A_1_1"] + (dataset["P6_16A_1_2"] / 60) + dataset["P6_16A_2_1"] + (dataset["P6_16A_2_2"] / 60) + dataset
["P6_16A_3_1"] + (dataset["P6_16A_3_2"] / 60) + dataset["P6_16A_4_1"] + (dataset["P6_16A_4_2"] / 60) + dataset["P6_16A_5_1"] + (dataset["P6_16A_5_2"] / 60) +
dataset["P6_16A_6_1"] + (dataset["P6_16A_6_2"] / 60)
247 dataset_limpio["TiempoApoyoOtroHogarSD"] = dataset["P6_16A_1_4"] + (dataset["P6_16A_1_4"] / 60) + dataset["P6_16A_2_3"] + (dataset["P6_16A_2_4"] / 60) + dataset
["P6_16A_3_3"] + (dataset["P6_16A_3_4"] / 60) + dataset["P6_16A_4_3"] + (dataset["P6_16A_4_4"] / 60) + dataset["P6_16A_5_3"] + (dataset["P6_16A_5_4"] / 60) +
dataset["P6_16A_6_3"] + (dataset["P6_16A_6_4"] / 60)
248

```

Código para tiempo apoyo otro hogar LV Y SD ( fusión)

## Preguntas relevantes para la siguiente fusión

nombre_campo	longitud	tipo	nomérico	catalogo	rango_clave
43) Pregunta P6.17.1 Durante la semana pasada, ¿usted hizo actividades o servicios gratuitos como voluntario en la Cruz Roja, asilos, casa hogar, DIF, hospitales, iglesias, Alcohólicos Anónimos, partidos políticos, etcétera?	1	C	p6_17_1	p6_17_1	12
44) Pregunta P6.17A.1.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_17a_1_1	nan	00-76
45) Pregunta P6.17A.1.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_17a_1_2	nan	00-30
46) Pregunta P6.17A.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_17a_1_3	nan	00-35
47) Pregunta P6.17A.1.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_17a_1_4	nan	00-50
48) Pregunta P6.17A.2 Durante la semana pasada, ¿usted hizo actividades o servicios gratuitos para la comunidad como tequio, faena, mano vuelta, mayordomía, fiestas patronales o sembrar árboles, limpiar calles, ríos, mercados, etcétera?	1	C	p6_17_2	p6_17_2	12
49) Pregunta P6.17A.2.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_17a_2_1	nan	00-56
50) Pregunta P6.17A.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_17a_2_2	nan	00-50
51) Pregunta P6.17A.2.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_17a_2_3	nan	00-22
52) Pregunta P6.17A.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_17a_2_4	nan	00-45

- Fusión tiempo voluntariado:

- Pregunta P6.17.1 Durante la semana pasada, ¿usted hizo actividades o servicios gratuitos como voluntario en la Cruz Roja, asilos, casa hogar, DIF, hospitales, iglesias, Alcohólicos Anónimos, partidos políticos, etcétera?
- Pregunta P6.17.2 Durante la semana pasada, ¿usted hizo actividades o servicios gratuitos para la comunidad como tequio, faena, mano vuelta, mayordomía, fiestas patronales o sembrar árboles, limpiar calles, ríos, mercados, etcétera?

```

249 # Tiempo Voluntariado
250 # preguntas 6.17.1, p6_17.2
251 # para la p6_17a_1_1, p6_17a_1_2, p6_17a_1_3, p6_17a_1_4, p6_17a_2_1, p6_17a_2_2, p6_17a_2_3, p6_17a_2_4
252 dataset_limpio["TiempoVoluntariadoLV"] = dataset["P6_17A_1_1"] + (dataset["P6_17A_1_2"] / 60) + dataset["P6_17A_2_1"] + (dataset["P6_17A_2_2"] / 60)
253 dataset_limpio["TiempoVoluntariadoSD"] = dataset["P6_17A_1_3"] + (dataset["P6_17A_1_4"] / 60) + dataset["P6_17A_2_3"] + (dataset["P6_17A_2_4"] / 60)
254

```

## Código para tiempo voluntariado LV y SD (fusión)

### Preguntas relevantes para la siguiente fusión

nombre_campo	longitud	tipo	nomérico	catalogo	rango_clave
45) Pregunta P6.18 Durante la semana pasada, en su tiempo libre, ¿usted hizo deporte o ejercicio físico? (fútbol, basquetbol, natación, box, karate, caminar, correr, andar en bicicleta, yoga, zumba)	1	C	p6_18	p6_18	12
46) Pregunta P6.18A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_18a_1	nan	00-57
47) Pregunta P6.18A.1.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_18a_1_1	nan	00-55
48) Pregunta P6.18A.1.2 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_18a_1_2	nan	00-40
49) Pregunta P6.18A.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_18a_1_3	nan	00-55
50) Pregunta P6.18A.2 Durante la semana pasada, ¿usted realizó actividades artísticas o culturales? (tocar un instrumento musical, pintar o realizar artes plásticas, gráficas, literarias o espirituales; incluye danza)	1	C	p6_18_2	p6_18_2	12
51) Pregunta P6.18A.2.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_18a_2_1	nan	00-70
52) Pregunta P6.18A.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_18a_2_2	nan	00-30
53) Pregunta P6.18A.2.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_18a_2_3	nan	00-30
54) Pregunta P6.18A.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_18a_2_4	nan	00-30
55) Pregunta P6.19 Durante la semana pasada, ¿usted participó en juegos de azar (cartas, ajedrez, dominó, ruleta, lotería), videojuegos, aficiones o pasatiempos (manualidades)?	1	C	p6_19	p6_19	12
56) Pregunta P6.19A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_19a_1	nan	00-50
57) Pregunta P6.19A.1.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_19a_1_1	nan	00-50
58) Pregunta P6.19A.1.2 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_19a_1_2	nan	00-50
59) Pregunta P6.19A.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_19a_1_3	nan	00-30
60) Pregunta P6.19A.2 Durante la semana pasada, ¿usted dedicó tiempo a hobbies? - HORAS	2	C	p6_19a_2	nan	00-50
61) Pregunta P6.19A.2.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_19a_2_1	nan	00-50
62) Pregunta P6.19A.2.2 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_19a_2_2	nan	00-50
63) Pregunta P6.19A.2.3 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_19a_2_3	nan	00-30
64) Pregunta P6.19A.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_19a_2_4	nan	00-50
65) Pregunta P6.20 Durante la semana pasada, ¿usted asistió a parques, riñes, ferias, estadios o canchas, museos u otros sitios culturales o de entretenimiento?	1	C	p6_20	p6_20	12
66) Pregunta P6.20A.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_20a_1	nan	00-60
67) Pregunta P6.20A.1.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_20a_1_1	nan	00-55
68) Pregunta P6.20A.1.2 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_20a_1_2	nan	00-48
69) Pregunta P6.20A.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_20a_1_3	nan	00-56
70) Pregunta P6.21.1 Durante la semana pasada, ¿usted dedicó tiempo especial (en hacer otra actividad) a los integrantes de su hogar para platicar de las actividades diarias, contar o aconsejar?	1	C	p6_21_1	p6_21_1	12
71) Pregunta P6.21A.1.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_21a_1_1	nan	00-40
72) Pregunta P6.21A.1.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_21a_1_2	nan	00-55
73) Pregunta P6.21A.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_21a_1_3	nan	00-48
74) Pregunta P6.21A.1.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_21a_1_4	nan	00-55
75) Pregunta P6.21.2 Durante la semana pasada, ¿usted asistió o participó en actividades religiosas? (actividades ceremoniales en casa u otro lugar, misas, rosarios u otro tipo de oraciones grupales, funerales, fiestas patronales, kermés de la iglesia)	1	C	p6_21_2	p6_21_2	12
76) Pregunta P6.21A.2.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2	C	p6_21a_2_1	nan	00-51
77) Pregunta P6.21A.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2	C	p6_21a_2_2	nan	00-55
78) Pregunta P6.21A.2.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2	C	p6_21a_2_3	nan	00-48
79) Pregunta P6.21A.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2	C	p6_21a_2_4	nan	00-58

Documento Original			
numero_campo	original	revisor	ultimo_cambio
458 Pregunta P6.21.3.3 Durante la semana pasada, ¿usted asistió a celebraciones civiles o políticas? (desfiles, mitines, marchas, reuniones)	1 C p6_21a_3_1, nan 00:50		
458 Pregunta P6.21.3.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_21a_3_2, nan 00:50		
458 Pregunta P6.21.3.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_21a_3_3, nan 00:50		
458 Pregunta P6.21.3.4 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_21a_3_4, nan 00:50		
458 Pregunta P6.21.4.1 Durante la semana pasada, ¿usted visitó a familiares, amigos o conocidos? (visitar o recatar a alguien; ir al arco, llorar, hablar por teléfono, escribir correspondencia)	1 C p6_21a_4_1, nan 00:50		
458 Pregunta P6.21.4.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_21a_4_1, nan 00:50		
458 Pregunta P6.21.4.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_21a_4_3, nan 00:50		
458 Pregunta P6.21.4.4 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_21a_4_4, nan 00:50		
458 Pregunta P6.21.5.1 Durante la semana pasada, ¿usted realizó alguna actividad para ENTRETENERSE, como ver hacer otra actividad (vídeos, novelas, series, programas, videos o documentales en televisión, tablet, celular o computadora)?	1 C p6_22_1, nan 00:50		
458 Pregunta P6.22.1.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_22a_1_1, nan 00:50		
458 Pregunta P6.22.1.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_22a_1_2, nan 00:50		
458 Pregunta P6.22.1.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_22a_1_3, nan 00:50		
458 Pregunta P6.22.1.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_22a_1_4, nan 00:50		
458 Pregunta P6.22.2.1 Durante la semana pasada, ¿usted realizó alguna actividad para HACER OíR ALGO, como escuchar una canción, escuchar o leer un libro, escuchar o leer algún tipo de audio?	1 C p6_22a_2_1, nan 00:50		
458 Pregunta P6.22.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_22a_2_1, nan 00:50		
458 Pregunta P6.22.2.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_22a_2_2, nan 00:50		
458 Pregunta P6.22.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_22a_2_3, nan 00:50		
458 Pregunta P6.22.2.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_22a_2_4, nan 00:50		
458 Pregunta P6.22.3.1 Durante la semana pasada, ¿usted realizó alguna actividad para HACER OíR ALGO, como escuchar una canción, escuchar o leer un libro, escuchar o leer algún tipo de audio?	1 C p6_22a_3_1, nan 00:50		
458 Pregunta P6.22.3.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_22a_3_1, nan 00:50		
458 Pregunta P6.22.3.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_22a_3_2, nan 00:50		
458 Pregunta P6.22.3.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_22a_3_3, nan 00:50		
458 Pregunta P6.22.3.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_22a_3_4, nan 00:50		

Documento Original			
numero_campo	original	revisor	ultimo_cambio
515 Pregunta P6.22.5 Durante la semana pasada, PARA ENTRETENERSE, ¿usted realizó alguna otra actividad relacionada con el uso de internet como descargar archivos o consultar información en cualquier aparato o dispositivo? (excluye si es por trabajo o estudio)	1 C p6_22_5_1, nan 00:50		
516 Pregunta P6.22.5.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_22a_5_1, nan 00:50		
516 Pregunta P6.22.5.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_22a_5_2, nan 00:50		
516 Pregunta P6.22.5.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_22a_5_3, nan 00:50		
516 Pregunta P6.22.5.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_22a_5_4, nan 00:50		
517 Pregunta P6.22.6 Durante la semana pasada, PARA ENTRETENERSE, ¿usted leyó algún libro, revista, periódico o artículo mediante algún dispositivo digital o impresor? (excluye si es por trabajo o estudio)	1 C p6_22_6_1, nan 00:50		
517 Pregunta P6.22.6.1 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_22a_6_1, nan 00:50		
517 Pregunta P6.22.6.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_22a_6_2, nan 00:50		
517 Pregunta P6.22.6.3 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_22a_6_3, nan 00:50		
517 Pregunta P6.22.6.4 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_22a_6_4, nan 00:50		
518 Pregunta P6.23.1.1 Durante la semana pasada, ¿usted realizó otra actividad que no se haya mencionado anteriormente?	1 C p6_23_1_1, nan 00:50		
518 Pregunta P6.23.1.2 ¿Cuánto?	2 C p6_23a_1_2, nan 00:50		
518 Pregunta P6.23.1.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_23a_1_3, nan 00:50		
518 Pregunta P6.23.1.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_23a_1_4, nan 00:50		
518 Pregunta P6.23.2.1 Durante la semana pasada, ¿usted realizó otra actividad que no se haya mencionado anteriormente? (relajarse, dormir, leer, escuchar música, escuchar o leer un libro, escuchar o leer algún grupo de ayuda o se recuperó de alguna enfermedad)	1 C p6_23_2_1, nan 00:50		
518 Pregunta P6.23.2.2 ¿Cuánto tiempo le dedicó de lunes a viernes? - HORAS	2 C p6_23a_2_1, nan 00:50		
518 Pregunta P6.23.2.3 ¿Cuánto tiempo le dedicó de lunes a viernes? - MINUTOS	2 C p6_23a_2_2, nan 00:50		
518 Pregunta P6.23.2.4 ¿Cuánto tiempo le dedicó sábado y domingo? - HORAS	2 C p6_23a_2_3, nan 00:50		
518 Pregunta P6.23.2.5 ¿Cuánto tiempo le dedicó sábado y domingo? - MINUTOS	2 C p6_23a_2_4, nan 00:50		

### ● Fusión tiempo actividades recreativas:

- Pregunta P6.18 Durante la semana pasada, en su tiempo libre ¿usted hizo deporte o ejercicio físico? (fútbol, basquetbol, natación, box, karate, caminar, correr, patinar, andar en bicicleta, yoga, zumba)
- Pregunta P6.19.1 Durante la semana pasada, ¿usted realizó actividades artísticas o culturales? (tocar un instrumento musical, pintar o realizar artes plásticas, gráficas, literarias o escénicas; incluye danza)
- Pregunta P6.19.2 Durante la semana pasada, ¿usted participó en juegos de mesa o azar (cartas, ajedrez, dominó, ruleta, etcétera), videojuegos, aficiones o pasatiempos (manualidades)?
- Pregunta P6.20 Durante la semana pasada, ¿usted asistió a parques, cines, ferias, estadios o canchas, museos u otros sitios culturales o de entretenimiento?
- Pregunta P6.21.1 Durante la semana pasada, ¿usted dedicó tiempo especial (sin hacer otra actividad) a los

integrantes de su hogar para platicar de las actividades diarias, consolar o aconsejar?

- Pregunta P6.21.2 Durante la semana pasada, ¿usted asistió o participó en actividades o celebraciones religiosas? (actividades ceremoniales en casa u otro lugar, misas, rosarios u otro tipo de oraciones grupales, funerales, fiestas patronales, kermés de la iglesia)
- Pregunta P6.21.3 Durante la semana pasada, ¿usted asistió a celebraciones cívicas o políticas? (desfiles, mítines, marchas, reuniones)
- Pregunta P6.21.4 Durante la semana pasada, ¿usted asistió a fiestas, atendió visitas de familiares, amigos o conocidos? (visitar o recibir a alguien; ir al antro, bar, hablar por teléfono, escribir correspondencia)
- Pregunta P6.22.1 Durante la semana pasada, PARA ENTRETENERSE ¿usted sin hacer otra actividad vio películas, novelas, series, programas, videos o documentales en televisión, tablet, celular o computadora?
- Pregunta P6.22.2 Durante la semana pasada, PARA ENTRETENERSE ¿usted sin hacer otra actividad escuchó música, noticias u otro programa de radio en cualquier dispositivo o aparato de audio?
- Pregunta P6.22.3 Durante la semana pasada, PARA ENTRETENERSE ¿usted sin hacer otra actividad revisó su correo electrónico, chateó o consultó redes sociales? (whatsapp, facebook, twitter, entre otros)
- Pregunta P6.22.4 Durante la semana pasada, PARA ENTRETENERSE ¿usted leyó algún libro, revista, periódico o artículo mediante algún dispositivo digital o impreso? (excluya si es por trabajo o estudio)

- Pregunta P6.22.5 Durante la semana pasada, PARA ENTRETENERSE ¿usted realizó alguna otra actividad relacionada con el uso de internet como descargar archivos o consultar información en cualquier aparato o dispositivo? (excluya si es por trabajo o estudio)
  - Pregunta P6.23.1 Durante la semana pasada, ¿usted rezó, meditó o descansó sin hacer otra actividad?

Código para tiempo actividades recreativas LV y SD ( fusión )

Preguntas relevantes para los siguientes atributos

Diccionario Original						
nombre_campo	label	tipos	menudeos	catalog	range_low	range_high
\$53 Pregunta P7.1.1 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a las actividades domésticas que hace en su hogar?					1	C p7_1.1 p7_1.1 1389
\$55 Pregunta P7.1.2 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a sus clases, cursos o estudios?					1	C p7_1.2 p7_1.2 1389
\$56 Pregunta P7.1.3 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a su trabajo, remunerado o actividad económica?					1	C p7_1.3 p7_1.3 1389
\$57 Pregunta P7.1.4 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a cuidar y apoyar a las personas de su hogar?					1	C p7_1.4 p7_1.4 1389
\$58 Pregunta P7.1.5 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a convivir con familiares y amigos?					1	C p7_1.5 p7_1.5 1389
\$59 Pregunta P7.1.6 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a los trámites o a su trabajo o escuela?					1	C p7_1.6 p7_1.6 1389
\$61 Pregunta P7.1.7 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a hacer trámites, pagar o cobrar algún programa social que recibe o recibe?					1	C p7_1.7 p7_1.7 1389
\$62 Pregunta P7.1.8 Por favor digame, ¿cómo se siente con el tiempo que le dedica la semana pasada a hacer lo que realmente le gusta?					1	C p7_1.8 p7_1.8 138
\$62 Pregunta P7.2.1 Por favor digame, ¿cómo se siente con su vida en general?					1	C p7_2.1 p7_2.1 15
\$64 Pregunta P7.2.2 Por favor digame, ¿cómo se siente con su vida familiar?					1	C p7_2.2 p7_2.2 15
\$64 Pregunta P7.2.3 Por favor digame, ¿cómo se siente con su vida afectiva (con el cónyuge que da y recibe)?					1	C p7_2.3 p7_2.3 15
\$65 Pregunta P7.2.4 Por favor digame, ¿cómo se siente con su vida social?					1	C p7_2.4 p7_2.4 15
\$66 Pregunta P7.2.5 Por favor digame, ¿cómo se siente con su situación económica?					1	C p7_2.5 p7_2.5 15
\$67 Pregunta P7.2.6 Por favor digame, ¿cómo se siente con su salud?					1	C p7_2.6 p7_2.6 15
\$68 Pregunta P7.3 En general, ¿qué le felicita más que se use?					1	C p7_3 p7_3 15

En este caso no es necesario realizar ninguna fusión, solo colocamos las preguntas con un nombre de atributo representativo y cambiando los significados con los de sus respectivos catálogos, se abordará con más detalle en el código.

```

268 # preguntas p7_1_1
269 dataset_limpio["SentimientoQuehaceresHogar"] = dataset["P7_1_1"]
270 catalogo_sentimiento_quehaceres_hogar = pd.read_csv(ruta_catalogo_sentimiento_quehaceres_hogar, encoding="latin1")
271 dataset_limpio["SentimientoQuehaceresHogar"] = dataset_limpio["SentimientoQuehaceresHogar"].replace(catalogo_sentimiento_quehaceres_hogar,
272 set_index("cve").to_dict()["descrip"])
273
274 # Sentimiento actividades estudio
275 # preguntas p7_1_2
276 dataset_limpio["SentimientoEstudio"] = dataset["P7_1_2"]
277 catalogo_sentimiento_estudio = pd.read_csv(ruta_catalogo_sentimiento_estudio, encoding="latin1")
278 dataset_limpio["SentimientoEstudio"] = dataset_limpio["SentimientoEstudio"].replace(catalogo_sentimiento_estudio.set_index("cve").to_dict()
279 ["descrip"])
280
281 # Sentimiento trabajo
282 # preguntas p7_1_3
283 dataset_limpio["SentimientoTrabajo"] = dataset["P7_1_3"]
284 catalogo_sentimiento_trabajo = pd.read_csv(ruta_catalogo_sentimiento_trabajo, encoding="latin1")
285 dataset_limpio["SentimientoTrabajo"] = dataset_limpio["SentimientoTrabajo"].replace(catalogo_sentimiento_trabajo.set_index("cve").to_dict()
286 ["descrip"])
287
288 # Sentimiento cuidado y apoyo a familiares
289 # preguntas p7_1_4
290 dataset_limpio["SentimientoCuidadoApoyoFamiliares"] = dataset["P7_1_4"]
291 catalogo_sentimiento_cuidado_apoyo_familiares = pd.read_csv(ruta_catalogo_sentimiento_cuidado_apoyo_familiares, encoding="latin1")
292 dataset_limpio["SentimientoCuidadoApoyoFamiliares"] = dataset_limpio["SentimientoCuidadoApoyoFamiliares"].replace(
293 catalogo_sentimiento_cuidado_apoyo_familiares.set_index("cve").to_dict()["descrip"])
294
295 # Sentimiento actividades de ocio convivencia
296 # preguntas p7_1_5
297 dataset_limpio["SentimientoOcioConvivencia"] = dataset["P7_1_5"]
298 catalogo_sentimiento_ocio_convivencia = pd.read_csv(ruta_catalogo_sentimiento_ocio_convivencia, encoding="latin1")
299 dataset_limpio["SentimientoOcioConvivencia"] = dataset_limpio["SentimientoOcioConvivencia"].replace(catalogo_sentimiento_ocio_convivencia,
300 set_index("cve").to_dict()["descrip"])
301
302 # Sentimiento traslado
303 # preguntas p7_1_6
304 dataset_limpio["SentimientoTraslado"] = dataset["P7_1_6"]
305 catalogo_sentimiento_traslado = pd.read_csv(ruta_catalogo_sentimiento_traslado, encoding="latin1")
306 dataset_limpio["SentimientoTraslado"] = dataset_limpio["SentimientoTraslado"].replace(catalogo_sentimiento_traslado.set_index("cve").to_dict()
307 ["descrip"])
308
309 # Sentimiento trámites y pagos
310 # preguntas p7_1_7
311 dataset_limpio["SentimientoTramitesPagos"] = dataset["P7_1_7"]
312 catalogo_sentimiento_tramites_pagos = pd.read_csv(ruta_catalogo_sentimiento_tramites_pagos, encoding="latin1")
313 dataset_limpio["SentimientoTramitesPagos"] = dataset_limpio["SentimientoTramitesPagos"].replace(catalogo_sentimiento_tramites_pagos.set_index(
314 "cve").to_dict()["descrip"])
315
316 # Sentimiento vida general
317 # preguntas p7_2_1
318 dataset_limpio["SentimientoVidaGeneral"] = dataset["P7_2_1"]
319 catalogo_sentimiento_vida_general = pd.read_csv(ruta_catalogo_sentimiento_vida_general, encoding="latin1")
320 dataset_limpio["SentimientoVidaGeneral"] = dataset_limpio["SentimientoVidaGeneral"].replace(catalogo_sentimiento_vida_general.set_index("cve"),
321 to_dict()["descrip"])
322
323 # Sentimiento vida familiar
324 # preguntas p7_2_2
325 dataset_limpio["SentimientoVidaFamiliar"] = dataset["P7_2_2"]
326 catalogo_sentimiento_vida_familiar = pd.read_csv(ruta_catalogo_sentimiento_vida_familiar, encoding="latin1")
327 dataset_limpio["SentimientoVidaFamiliar"] = dataset_limpio["SentimientoVidaFamiliar"].replace(catalogo_sentimiento_vida_familiar.set_index(
328 "cve").to_dict()["descrip"])
329
330 # Sentimiento vida afectiva
331 # preguntas p7_2_3
332 dataset_limpio["SentimientoVidaAfectiva"] = dataset["P7_2_3"]
333 catalogo_sentimiento_vida_afectiva = pd.read_csv(ruta_catalogo_sentimiento_vida_afectiva, encoding="latin1")
334 dataset_limpio["SentimientoVidaAfectiva"] = dataset_limpio["SentimientoVidaAfectiva"].replace(catalogo_sentimiento_vida_afectiva.set_index(
335 "cve").to_dict()["descrip"])
336
337 # Sentimiento vida social
338 # preguntas p7_2_4
339 dataset_limpio["SentimientoVidaSocial"] = dataset["P7_2_4"]
340 catalogo_sentimiento_vida_social = pd.read_csv(ruta_catalogo_sentimiento_vida_social, encoding="latin1")
341 dataset_limpio["SentimientoVidaSocial"] = dataset_limpio["SentimientoVidaSocial"].replace(catalogo_sentimiento_vida_social.set_index("cve"),
342 to_dict()["descrip"])
343
344 # Sentimiento situación económica
345 # preguntas p7_2_5
346 dataset_limpio["SentimientoSituaciónEconómica"] = dataset["P7_2_5"]
347 catalogo_sentimiento_situacion_economica = pd.read_csv(ruta_catalogo_sentimiento_situacion_economica, encoding="latin1")
348 dataset_limpio["SentimientoSituaciónEconómica"] = dataset_limpio["SentimientoSituaciónEconómica"].replace(
349 catalogo_sentimiento_situacion_economica.set_index("cve").to_dict()["descrip"])
350
351 # Sentimiento con tu vivienda
352 # preguntas p7_2_6
353 dataset_limpio["SentimientoVivienda"] = dataset["P7_2_6"]
354 catalogo_sentimiento_vivienda = pd.read_csv(ruta_catalogo_sentimiento_vivienda, encoding="latin1")
355 dataset_limpio["SentimientoVivienda"] = dataset_limpio["SentimientoVivienda"].replace(catalogo_sentimiento_vivienda.set_index("cve"),
356 to_dict()["descrip"])
357
358 # Sentimiento felicidad general
359 # preguntas p7_3
360 dataset_limpio["SentimientoFelicidadGeneral"] = dataset["P7_3"]
361 catalogo_sentimiento_felicidad_general = pd.read_csv(ruta_catalogo_sentimiento_felicidad_general, encoding="latin1")
362 dataset_limpio["SentimientoFelicidadGeneral"] = dataset_limpio["SentimientoFelicidadGeneral"].replace(catalogo_sentimiento_felicidad_general,
363 set_index("cve").to_dict()["descrip"])

```

En el código de las últimas preguntas solo se encarga de copiarlas y reemplazar sus valores por los de los catálogos correspondientes.

Con esto terminamos la limpieza ahora mostramos el dataset limpio en una nueva ventana y guardamos el resultado

```

353 # extraemos la cantidad de columnas
354 cantidad_columnas = dataset_limpio.shape[1]
355 ventana2 = mostrar_dataframe_en_qt(dataset_limpio, "Nuevo diccionario " + str(cantidad_columnas))
356
357 # vamos a quitar todos los acentos y caracteres especiales
358
359 # Guardar el nuevo dataset
360 dataset_limpio.to_csv("nuevo_dataset.csv", index=False, encoding="latin1")
361
362 # Ejecutar la aplicación
363 sys.exit(app.exec())
364
365
366 if __name__ == "__main__":
367     main()

```

*Salida gráfica*

Línea	Sexo	Edad	TiendaAhorroInteres	TiendaAhorroSal	TiendaAhorroDif	TiendaAhorroCierre	TiendaAhorroG	TiendaAhorroDiferencia	TiendaAhorroDiferenciaSal	TiendaAhorroDiferenciaDif	TiendaAhorroDiferenciaCierre	TiendaAhorroDiferenciaG	TiendaAhorroDiferenciaDiferencia	TiendaAhorroDiferenciaDiferenciaSal	TiendaAhorroDiferenciaDiferenciaDif	TiendaAhorroDiferenciaDiferenciaCierre	TiendaAhorroDiferenciaDiferenciaG	TiendaAhorroDiferenciaDiferenciaDiferencia
1	Mujer	89	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	22.0	10.0	5.0	7.5	3.0		
2	Hombre	80	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	15.0	7.5	2.5	5.0	1.0		
3	Hombre	30	Si	40.0	29.0	1.3333333333333333	0.0	0.0	0.0	0.0	40.0	18.0	3.333333333333333	1.3333333333333333	5.0	2.0		
4	Hombre	25	Si	55.0	0.0	2.0	0.0	104000.0	Si	Si	Si	40.0	20.0	5.0	20.0	5.0	3.0	
5	Mujer	40	Si	40.0	5.0	5.0	2.0	72000.0	Si	Si	Si	18.0	9.0	2.0	5.0	2.0		
6	Hombre	20	Si	40.0	0.0	0.0	0.0	96000.0	Si	Si	Si	40.0	15.0	7.5	10.0	2.0		
7	Mujer	50	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	35.0	18.0	5.0	6.0	2.5	1.0	
8	Mujer	68	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	40.0	18.0	7.5	9.0	2.0		
9	Hombre	50	No	55.0	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	40.0	18.0	7.5	3.0	7.5	3.0
10	Mujer	21	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	40.0	18.0	7.5	3.0	10.0	4.0	
11	Mujer	55	No	60.0	0.0	2.0	0.0	434000.0	Si	Si	Si	40.0	14.0	5.0	5.0	2.5	1.0	
12	Mujer	38	Si	40.0	9.0	2.0	0.0	416000.0	Si	Si	Si	40.0	16.0	5.0	2.0	7.5	3.0	
13	Hombre	61	Si	55.0	21.0	0.0	0.0000000000000000	240000.0	No especificado	No especificado	No especificado	No especificado	30.0	12.0	2.5	1.0	2.5	1.0
14	Mujer	10	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	35.0	17.0	10.0	5.0	2.0		
15	Mujer	61	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	35.0	14.0	5.0	5.0	2.0		
16	Mujer	55	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	35.0	14.0	5.0	5.0	2.0		
17	Hombre	37	Si	40.0	30.0	5.0	2.0	120000.0	No especificado	No especificado	No especificado	No especificado	40.0	20.0	5.0	5.0	3.0	
18	Mujer	57	Si	20.0	0.0	0.0	0.0	52000.0	No especificado	No especificado	No especificado	No especificado	40.0	20.0	5.0	2.0	2.5	1.0
19	Hombre	30	No	50.0	10.0	0.0	0.0	132000.0	Si	Si	Si	40.0	16.0	2.0	5.0	1.0		
20	Mujer	40	Si	40.0	5.0	0.0	0.0	134000.0	Si	Si	Si	40.0	17.0	5.0	7.5	3.0		
21	Mujer	28	Si	18.0	4.0	2.0	0.0	312000.0	No	No	No	45.0	22.0	7.5	10.0	2.0		
22	Hombre	35	Si	50.0	0.0	3.0	0.0	130000.0	Si	Si	Si	35.0	16.0	5.0	2.0	1.0		
23	Mujer	28	Si	45.0	5.0	2.0	0.0	720000.0	Si	Si	Si	30.0	12.0	2.0	5.0	1.0	10.333333333333334	
24	Hombre	34	Si	40.0	8.0	0.0	0.0000000000000007	105000.0	No	No	No	40.0	16.0	7.5	10.0	2.0		
25	Mujer	34	Si	25.0	0.0	0.0	0.0	120000.0	No	No	No	45.0	16.0	7.5	10.0	2.0		
26	Mujer	63	Si	50.0	0.0	0.0	0.0	48000.0	No	No	No	60.0	24.0	5.0	6.0	2.5	1.0	
27	Mujer	59	Si	45.0	5.0	5.0	1.0	64000.0	No	No	No	45.0	16.0	7.5	2.0	5.0	2.0	
28	Hombre	55	Si	40.0	0.0	5.15	1.0	64000.0	No	No	No	30.0	12.0	5.0	2.0	2.5	1.0	
29	Hombre	40	No	0.0	0.0	0.0	0.0	No especificado	No especificado	No especificado	No especificado	35.0	20.0	5.0	7.5	2.0		

Ahora contamos con un dataset limpio, sin embargo, aún cuenta con demasiados elementos, con ayuda de un CLI desarrollado por nosotros reducimos el número de instancias a 1000

## Código del CLI



```
CLI_DIVIDIR_DATASET.PY × muestra.csv

1 import pandas as pd
2 import random
3 import os
4 import pyfiglet
5 import encodings
6
7 # VARIABLES GLOBALES PARA COLORES ANSI
8 RESET = "\033[0m"
9 NEGRO = "\033[30m"
10 ROJO = "\033[31m"
11 VERDE = "\033[32m"
12 AMARILLO = "\033[33m"
13 AZUL = "\033[34m"
14 MAGENTA = "\033[35m"
15 CYAN = "\033[36m"
16 BLANCO = "\033[37m"
17 NEGRITA = "\033[1m"
18 SUBRAYADO = "\033[4m"
19
20 # Función para limpiar la pantalla según el sistema operativo
21 def limpiar_pantalla():
22     if os.name == 'nt':
23         os.system('cls')
24     else:
25         os.system('clear')
26
27 def mostrar_datos_actuales(ruta_dataset: str, dataset: pd.DataFrame):
28     if any([ruta_dataset, dataset]):
29         if ruta_dataset:
30             print(MAGENTA + "Ruta dataset: " + RESET, ruta_dataset)
31         if dataset is not None:
32             print(MAGENTA + "Dataset cargado." + RESET)
33     else:
34         print(AMARILLO + "No hay datos agregados." + RESET)
35
36 def mostrar_menu_principal(limpiar_pantalla_activado):
37     """Muestra el menú principal del programa"""
38     print(MAGENTA + NEGRITA + "MENU PRINCIPAL" + RESET)
39     print(VERDE + "1. Cargar datos")
40     print("2. Tipo de particionado")
41     print("3. Particionar")
42     if limpiar_pantalla_activado:
43         print(AMARILLO + "4. Desactivar limpieza de pantalla")
```

```

56     def mostrar_menu_principal(limpiar_pantalla_activado):
57         """Muestra el menú principal del programa"""
58         print(MAGENTA + NEGRITA + "MENU PRINCIPAL" + RESET)
59         print(VERDE + "1. Cargar datos")
60         print("2. Tipo de particionado")
61         print("3. Particionar")
62         if limpiar_pantalla_activado:
63             print(AMARILLO + "4. Desactivar limpieza de pantalla")
64         else:
65             print(AMARILLO + "4. Activar limpieza de pantalla")
66         print(VERDE + "5. Guardar datasets resultantes")
67         print(ROJO + "0. Salir" + RESET)
68
69
70     return input(AZUL + NEGRITA + "Seleccione una opción: " + RESET)
71
72
73 # Función para mostrar los encodings disponibles
74 def mostrar_encodings_disponibles():
75     """Muestra encodings disponibles en páginas, destacando los comunes en la primera página."""
76     encodings_comunes = ['utf-8', 'latin-1', 'ascii', 'utf-16', 'cp1252']
77     encodings_disponibles = list(encodings.aliases.aliases.keys())
78     encodings_disponibles = [e for e in encodings_disponibles if e not in encodings_comunes]
79
80     # Insertar los encodings comunes al principio
81     encodings_disponibles = encodings_comunes + encodings_disponibles
82     tamano_pagina = 10
83     total_paginas = (len(encodings_disponibles) + tamano_pagina - 1) // tamano_pagina
84
85     pagina_actual = 0
86     while True:
87         limpiar_pantalla()
88         print(f"\n{MAGENTA}{NEGRITA}ENCODINGS DISPONIBLES{RESET}")
89         inicio = pagina_actual * tamano_pagina
90         fin = inicio + tamano_pagina
91
92         # Mostrar la página actual
93         print(f"\n{CYAN}{NEGRITA}Página {pagina_actual + 1} de {total_paginas}:{RESET}")
94         for i, encoding in enumerate(encodings_disponibles[inicio:fin], start=1):
95             # Encodings comunes resaltados
96             if encoding in encodings_comunes:
97                 print(f"\n{VERDE}{NEGRITA}{inicio + i}. {encoding}{RESET}")
98             else:
99                 print(f"\n{BLANCO}{inicio + i}. {encoding}{RESET}")
100
101         # Opciones para la navegación
102         print(f"\n{AMARILLO}{NEGRITA}Opciones:{RESET}")
103         print(f"\n{AZUL}[n]{RESET} Siguiente página")
104         print(f"\n{AZUL}[p]{RESET} Página anterior")
105         print(f"\n{AZUL}[q]{RESET} Salir")
106         print(f"\n{AZUL}[0]{RESET} elige un {SUBRAYADO}número{RESET} para seleccionar un encoding.")
107
108         opcion = input(f"\n{CYAN}Elige una opción: {RESET}").strip().lower()
109         if opcion == 'n' and pagina_actual < total_paginas - 1:
110             pagina_actual += 1
111         elif opcion == 'p' and pagina_actual > 0:
112             pagina_actual -= 1
113         elif opcion == 'q':
114             print(f"\n{ROJO}Saliendo...{RESET}")
115             return None
116         else:
117             try:
118                 eleccion = int(opcion)
119                 if inicio + 1 <= eleccion <= min(fin, len(encodings_disponibles)):
120                     encoding_elegido = encodings_disponibles[eleccion - 1]
121                     print(f"\n{VERDE}Has elegido: {NEGRITA}{encoding_elegido}{RESET}")
122                     return encoding_elegido
123                 else:
124                     print(f"\n{ROJO}Número inválido. Por favor, elige un número de la lista.{RESET}")
125             except ValueError:
126                 print(f"\n{ROJO}Entrada inválida. Por favor, intenta de nuevo.{RESET}")

```

```

105
106
107 # Función para cargar los datos
108 def cargar_datos():
109     """Carga un dataset en formato CSV."""
110     limpiar_pantalla()
111     while True:
112         print(AZUL + NEGRITA + "CARGAR DATOS" + RESET)
113         ruta_dataset = input("Ingrese la ruta del dataset: ")
114         encoding = mostrar_encodings_disponibles()
115         if encoding is None:
116             print(ROJO + "No se ha seleccionado un encoding. Por favor, intenta de nuevo." + RESET)
117             input("Presione Enter para continuar...")
118             return None, None
119     try:
120         dataset = pd.read_csv(ruta_dataset, encoding=encoding)
121         return ruta_dataset, dataset, encoding
122     except Exception as e:
123         limpiar_pantalla()
124         print(ROJO + "Error al cargar el dataset" + RESET)
125         print(AMARILLO + str(e) + RESET)
126
127
128 def particionar_dataset(dataset, tipo_particionado):
129     """Realiza el particionado del dataset según el tipo seleccionado."""
130     # arreglo de datasets retornados
131     datasets = []
132     if dataset is None:
133         print(ROJO + "No se ha cargado ningún dataset." + RESET)
134         input("Presione Enter para continuar...")
135         return
136
137     if tipo_particionado == 1:
138         # División por porcentaje
139         porcentaje = float(input("Ingrese el porcentaje de entrenamiento (0-100): "))
140         train = dataset.sample(frac=porcentaje / 100, random_state=42)
141         test = dataset.drop(train.index)
142         print(f"\{VERDE}Dataset dividido: {len(train)} entrenamiento y {len(test)} prueba.\{RESET}\")
143         datasets.append(train)
144         datasets.append(test)
145     elif tipo_particionado == 2:
146         # División por k-folds
147         from sklearn.model_selection import KFold
148         k = int(input("Ingrese el número de folds (k): "))
149         kf = KFold(n_splits=k, shuffle=True, random_state=42)
150         for fold, (train_index, test_index) in enumerate(kf.split(dataset), start=1):
151             print(f"\{Fold {fold}: {len(train_index)} entrenamiento y {len(test_index)} prueba.\")
152             train = dataset.iloc[train_index]    "iloc": Unknown word.
153             test = dataset.iloc[test_index]    "iloc": Unknown word.
154             datasets.append(train)
155             datasets.append(test)
156     elif tipo_particionado == 3:
157         # División por leave-one-out
158         from sklearn.model_selection import LeaveOneOut
159         loo = LeaveOneOut()
160         for train_index, test_index in loo.split(dataset):
161             print(f"\{Entrenamiento: {len(train_index)}, Prueba: {len(test_index)}\")
162             train = dataset.iloc[train_index]    "iloc": Unknown word.
163             test = dataset.iloc[test_index]    "iloc": Unknown word.
164             datasets.append(train)
165             datasets.append(test)
166     elif tipo_particionado == 4:
167         # Muestra aleatoria de tamaño determinado
168         print(VERDE + "Tomar muestra aleatoria de tamaño determinado" + RESET)
169         print(AMARILLO + "El tamaño de la muestra debe ser menor o igual a " + str(len(dataset)) + RESET)
170         tamano_muestra = int(input("Ingrese el tamaño de la muestra: "))
171         # verificamos que el tamaño de la muestra sea menor o igual al tamaño del dataset
172         if tamano_muestra < len(dataset):
173             muestra = dataset.sample(n=tamano_muestra, random_state=42)
174             print(f"\{VERDE}Muestra tomada: {len(muestra)} elementos.\{RESET}\")
175         else:
176             print(ROJO + "El tamaño de la muestra debe ser menor o igual al tamaño del dataset." + RESET)
177             datasets.append(muestra)
178
179     else:
180         print(ROJO + "Tipo de particionado no válido." + RESET)
181
182     input("Presione Enter para continuar...")
183     return datasets

```

```

194 def guardar_datasets_resultantes(datasets_resultantes, tipo_particionado, codificacion):
195     print(AMARILLO + "Guardar datasets resultantes" + RESET)
196     if not datasets_resultantes:
197         print(ROJO + "No hay datasets para guardar." + RESET)
198         input("Presione Enter para continuar...")
199         return
200     if tipo_particionado == 1:
201         # División por porcentaje
202         train, test = datasets_resultantes
203         train.to_csv("train.csv", index=False, encoding=codificacion)
204         test.to_csv("test.csv", index=False, encoding=codificacion)
205         print(f"\n{VERDE}Datasets guardados: train.csv y test.csv{RESET}\n")
206     elif tipo_particionado == 2:
207         # División por k-folds
208         for i, dataset in enumerate(datasets_resultantes, start=1):
209             if i % 2 == 1:
210                 dataset.to_csv(f"train_fold_{i}.csv", index=False, encoding=codificacion)
211             else:
212                 dataset.to_csv(f"test_fold_{i}.csv", index=False, encoding=codificacion)
213                 #dataset.to_csv(f"fold_{i}.csv", index=False)
214         print(f"\n{VERDE}Datasets guardados: fold_1.csv, fold_2.csv, ...{RESET}\n")
215     elif tipo_particionado == 3:
216         # División por leave-one-out
217         for i, dataset in enumerate(datasets_resultantes, start=1):
218             dataset.to_csv(f"leave_one_out_{i}.csv", index=False, encoding=codificacion)
219         print(f"\n{VERDE}Datasets guardados: leave_one_out_1.csv, leave_one_out_2.csv, ...{RESET}\n")
220     elif tipo_particionado == 4:
221         # Muestra aleatoria de tamaño determinado
222         muestra = datasets_resultantes[0]
223         muestra.to_csv("muestra.csv", index=False, encoding=codificacion)
224         print(f"\n{VERDE}Muestra guardada: muestra.csv{RESET}\n")
225
226
227 if __name__ == "__main__":
228     limpiar_pantalla_activado = True
229     ruta_data_set = None
230     dataset = None
231     codificacion = None
232     tipo_particionado = None
233     datasets_resultantes = []
234     while True:
235         try:
236             if limpiar_pantalla_activado:
237                 limpiar_pantalla()
238                 print(AZUL + NEGRITA + "Bienvenido " + os.getlogin() + "!")
239                 mostrar_datos_actuales(ruta_data_set, dataset)
240                 seleccion = int(mostrar_menu_principal(limpiar_pantalla_activado))
241                 if seleccion == 1:
242                     ruta_data_set, dataset, codificacion = cargar_datos()
243                 elif seleccion == 2:
244                     tipo_particionado = menu_tipo_particionado()
245                 elif seleccion == 3:
246                     datasets_resultantes = particionar_dataset(dataset, tipo_particionado)
247                     input("Presione Enter para continuar...")
248                 elif seleccion == 4:
249                     limpiar_pantalla_activado = not limpiar_pantalla_activado
250                 elif seleccion == 5:
251                     guardar_datasets_resultantes(datasets_resultantes, tipo_particionado, codificacion)
252                 elif seleccion == 0:
253                     limpiar_pantalla()
254                     print(AMARILLO + "Saliendo..." + RESET)
255                     ascii_art = pyfiglet.figlet_format("Desarrollado por:")
256                     print(ROJO + ascii_art + RESET)
257                     ascii_art = pyfiglet.figlet_format("Aether")
258                     print(CYAN + ascii_art + RESET)
259                     break
260                 else:
261                     print(ROJO + "Opción inválida" + RESET)
262             except Exception as e:
263                 limpiar_pantalla()
264                 print(ROJO + "Informe de error" + RESET)
265                 print(str(e))
266                 input("Presione Enter para continuar...")

```

*Ejecutamos el cli*

```
Bienvenido aether!
No hay datos agregados.
MENU PRINCIPAL
1. Cargar datos
2. Tipo de particionado
3. Particionar
4. Desactivar limpieza de pantalla
5. Guardar datasets resultantes
0. Salir
Seleccione una opción: 1
```

*Cargamos el dataset*

```
CARGAR DATOS
Ingrese la ruta del dataset: /home/aether/Proyectos/proyecto/nuevo_dataset.csv
```

*Damos enter*

```
ENCODINGS DISPONIBLES
Página 1 de 34:
1. utf-8
2. latin-1
3. ascii
4. utf-16
5. cp1252
6. 646
7. ansi_x3.4_1968
8. ansi_x3_4_1968
9. ansi_x3.4_1986
10. cp367

Opciones:
[n] Siguiente página
[p] Página anterior
[q] Salir
0 elige un número para seleccionar un encoding.

Elige una opción: 2
```

*Elegimos el tipo de codificación*

```
Bienvenido aether!
Ruta dataset: /home/aether/Proyectos/proyecto/nuevo_dataset.csv
Dataset cargado.
MENU PRINCIPAL
1. Cargar datos
2. Tipo de particionado
3. Particionar
4. Desactivar limpieza de pantalla
5. Guardar datasets resultantes
0. Salir
Seleccione una opción: 2
```

*Elegimos el tipo de particionado*

```
TIPO DE PARTICIONADO
1. División por porcentaje
2. División por k-folds
3. División por leave-one-out
4. Tomar muestra aleatoria de tamaño determinado
Seleccione el tipo de particionado: 4
```

*Seleccionamos el tipo de muestra*

```
Bienvenido aether!
Ruta dataset: /home/aether/Proyectos/proyecto/nuevo_dataset.csv
Dataset cargado.
MENU PRINCIPAL
1. Cargar datos
2. Tipo de particionado
3. Particionar
4. Desactivar limpieza de pantalla
5. Guardar datasets resultantes
0. Salir
Seleccione una opción: 3
```

## *Particionamos*

```
Bienvenido aether!
Ruta dataset: /home/aether/Proyectos/proyecto/nuevo_dataset.csv
Dataset cargado.
MENU PRINCIPAL
1. Cargar datos
2. Tipo de particionado
3. Particionar
4. Desactivar limpieza de pantalla
5. Guardar datasets resultantes
0. Salir
Seleccione una opción: 3
Tomar muestra aleatoria de tamaño determinado
El tamaño de la muestra debe ser menor o igual a 68538
Ingrese el tamaño de la muestra: 1000
Muestra tomada: 1000 elementos.
Presione Enter para continuar...
```

## *Colocamos el número de muestras*

```
Bienvenido aether!
Ruta dataset: /home/aether/Proyectos/proyecto/nuevo_dataset.csv
Dataset cargado.
MENU PRINCIPAL
1. Cargar datos
2. Tipo de particionado
3. Particionar
4. Desactivar limpieza de pantalla
5. Guardar datasets resultantes
0. Salir
Seleccione una opción: 5
```

## *Guardamos la muestra*

```
Bienvenido aether!
Ruta dataset: /home/aether/Proyectos/proyecto/nuevo_dataset.csv
Dataset cargado.
MENU PRINCIPAL
1. Cargar datos
2. Tipo de particionado
3. Particionar
4. Desactivar limpieza de pantalla
5. Guardar datasets resultantes
0. Salir
Seleccione una opción: 0
```

## *Salimos*

```
Saliendo...
[...]
[...]
[...]
(variado) (base) ➔ scriptsVariados
```

Para este punto ya tenemos la muestra de 1000 elementos, sin embargo nos falta quitar los acentos ya que el dataset se encuentra en formato “latin1” y aunque para python está bien complica el funcionamiento en orange.

Para esto se usó el siguiente programa en python

```
qtFLimpieza0.py      conjunto_de_datos_tmodulo_enut_2019.csv      quitasacentos.py      QUITARCOLUMNAEAFAPD.py
qtquitasacentos.py ...
1 import pandas as pd
2 import unicodedata
3
4 # Leer el archivo CSV
5 try:
6     df = pd.read_csv("/home/aether/Proyectos/python/scriptsVariados/muestra.csv",
7                      encoding="latin1")
7 except UnicodeDecodeError:
8     print("Error de codificación, intenta con otro encoding como 'utf-8' o 'latin1'.")
9     raise
10
11 # Función para eliminar acentos
12 def eliminar_acentos(texto):
13     if isinstance(texto, str):
14         return ''.join(c for c in unicodedata.normalize('NFKD', texto) if not unicodedata.
15                         combining(c))
16         return texto
17
18 # Aplicar la función a todo el DataFrame
19 df_sin_acentos = df.applymap(eliminar_acentos)    "applymap": Unknown word.
20
21 # Guardar el resultado en un archivo CSV
22 df_sin_acentos.to_csv("muestraFinal.csv", index=False, encoding="utf-8")
23
24 print("Archivo procesado y guardado como 'muestraFinal.csv'.")
25 print(df.head())
26 print(df_sin_acentos.head())
27
```

Limpieza completada.

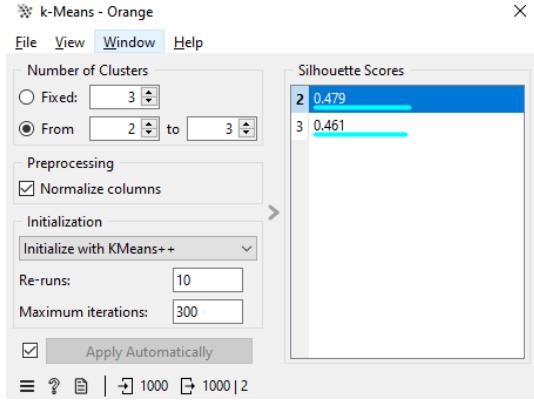
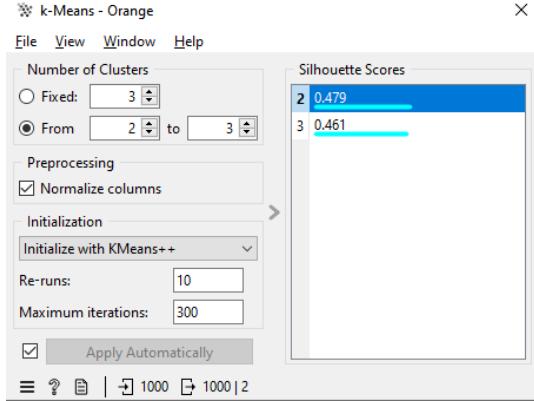
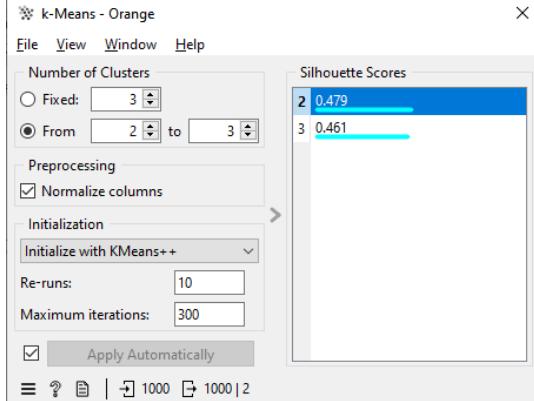
Vista en Orange

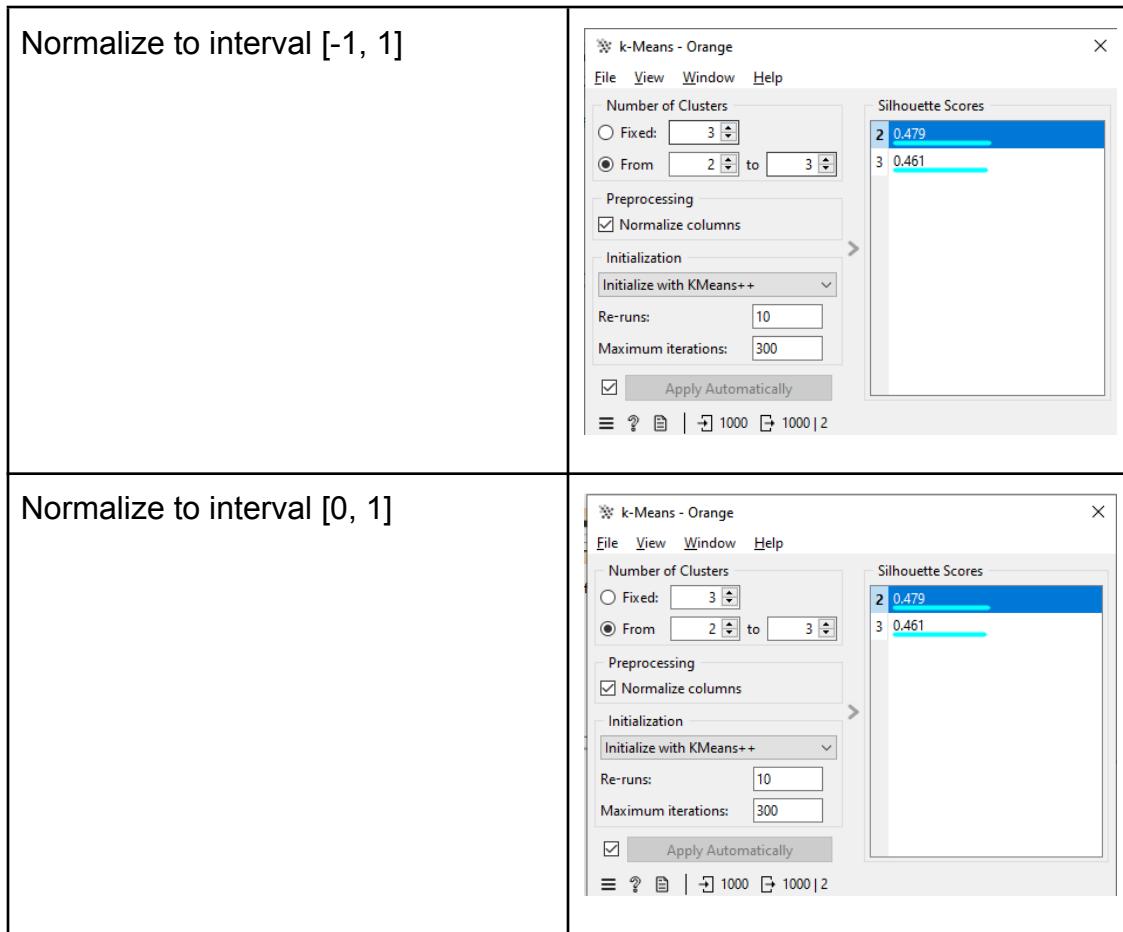
Entidad	Sexo	Edad	díasAlMenos14k	TiempoTrabajoV	TiempoTrabajoSD	TiempoTrabajoL1	TiempoTrabajoSF	Salario	medidaMaterial	remiso/actuacion	terreno/obstaculo	rechazos/obstaculo	rechazos/actuacion	rechazos/obstaculo	rechazos/actuacion
1	21 Hombre	40 Si	50,0	11,00	2,5	0,5		1300	No especificado	No especificado	No especificado	40	16	5	
2	16 Hombre	66 Si	50,0	14,00	1,666667	0,333333		1300	No especificado	No especificado	No especificado	25	10	3,75	
3	1 Hombre	60 Si	70,0	14,00	0,333333	0,666667		1000	No especificado	No especificado	No especificado	35	14	6,25	
4	16 Hombre	56 No	0,0	0,00	0,5	0	0	0	No especificado	No especificado	No especificado	40	16	2,5	
5	20 Hombre	52 Si	8,0	0,00	0,5	0	0	5000	No especificado	No especificado	No especificado	45	18	3,333333	
6	21 Hombre	53 Si	10,0	4,00	10	2	0	0	No especificado	No especificado	No especificado	40	20	7,5	
7	23 Hombre	50 Si	20,0	6,00	5	0,5	0	0	No especificado	No especificado	No especificado	40	20	5	
8	29 Mujer	63 No	0,0	0,00	0	0	0	600	No especificado	No especificado	No especificado	40	18	7,5	
9	28 Mujer	47 No	8,0	0,00	2	0	0	600	No especificado	No especificado	No especificado	35	7	2,5	
10	15 Mujer	36 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	35	20	10	
11	21 Mujer	36 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	40	17	4,5	
12	15 Hombre	15 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	30	18	5	
13	27 Mujer	24 Si	60,0	24,00	0,416667	0,166667		1500	No	No	No	40	16	5	
14	10 Hombre	17 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	35	14	6	
15	15 Hombre	49 Si	24,0	12,00	10	4	99999	1000	No especificado	No especificado	No especificado	40	16	5	
16	8 Mujer	61 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	35	16	3	
17	11 Mujer	68 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	60	24	15	
18	13 Mujer	52 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	40	16	3	
19	2 Hombre	57 Si	20,0	10,00	0,033333	0,033333		4000	No especificado	No especificado	No especificado	40	16	5	
20	23 Mujer	63 Si	40,0	9,00	0	0	0	1000	No especificado	No especificado	No especificado	35	14	7,5	
21	19 Mujer	59 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	30	12	7,5	
22	4 Mujer	77 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	50	20	7,5	
23	10 Mujer	16 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	50	16	7,5	
24	15 Mujer	35 Si	35,0	14,00	5	2	1000	Si	Si	Si	Si	45	18	7,5	
25	10 Hombre	20 Si	40,0	4,00	15	1,5	1600	Si	Si	Si	Si	35	16	15	
26	24 Mujer	47 Si	25,0	0,00	5	0	99999	No especificado	No especificado	No especificado	No especificado	40	16	7,5	
27	21 Mujer	51 Si	30,0	12,00	0,033333	0,033333		1750	No especificado	No especificado	No especificado	35	14	2,5	
28	19 Hombre	70 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	25	10	5	
29	19 Hombre	68 Si	30,0	4,00	10	2	2000	No especificado	No especificado	No especificado	No especificado	30	16	40	
30	15 Mujer	59 Si	40,0	6,00	5	1	1800	Si	Si	Si	Si	45	18	7,5	
31	20 Hombre	15 Si	25,0	16,00	0	0	500	No	No	No	No	35	12	7,5	
32	20 Hombre	16 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	40	16	6	
33	1 Hombre	17 Si	0,0	6,00	0	0,5	200	No	No	No	No	42	16	4,5	
34	24 Mujer	27 Si	50,0	10,00	0,333333	0,666667		0	No especificado	No especificado	No especificado	30	12	0	
35	30 Hombre	65 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	60	32	15	
36	25 Mujer	15 No	0,0	0,00	0	0	0	0	No especificado	No especificado	No especificado	40	20	5	
37	22 Hombre	43 Si	50,0	10,00	0	0	0	1800	No especificado	No especificado	No especificado	40	16	10	
38	1 Hombre	22 Si	30,0	12,00	5	1	2500	No	No	No	No	35	10	7	
39	6 Hombre	26 Si	6,0	0,00	1,333333	0	0	0	No especificado	No especificado	No especificado	60	24	7,666667	
40	21 Hombre	12 No	0,0	0,00	0	0	0	1300	No especificado	No especificado	No especificado	45	14	3,333333	
41	16 Hombre	22 Si	40,0	0,00	10	0	0	0	No	No	No	30	24	1,666667	

## d. Construcción de datos

### i. Normalización de valores numéricos

Se evaluaron varias técnicas de normalización para preparar los datos antes de aplicar el modelo K-Means, con el objetivo de garantizar que las variables tuvieran la misma escala y, por ende, una contribución equitativa al cálculo de distancias. Las técnicas consideradas fueron:

Técnica de Normalización	Silhouette Score
Standardize to $\mu=0, \sigma^2=1$	
Center to $\mu=0$	
Scale to $\sigma^2=1$	



Calculamos el Silhouette Score, una métrica que mide la calidad del agrupamiento al evaluar la cohesión interna de los clusters y su separación respecto a otros clusters.

Todas las técnicas de normalización produjeron el mismo Silhouette Score, lo que indica que el rendimiento del modelo no fue afectado significativamente por el método de normalización aplicado.

Al no haber una diferencia notable, optamos por usar la técnica de Standardize to  $\mu=0$ ,  $\sigma^2=1$  por ser una de las técnicas más comunes y recomendables en K-Means, ya que iguala tanto la media como la dispersión de todas las variables, es decir, garantiza que todas las variables tengan una contribución igual en el cálculo de distancias.

- ii. Transformación de valores categóricos a numéricos
- iii. Transformación de valores numéricos a categóricos

## 4. Modelado

### a. Selección de técnicas de modelado

#### i. Lista de las técnicas que se utilizarán para la evaluación del modelo.

##### K-means:

- Silhouette Score: Mide la calidad de la agrupación de los puntos dentro de sus clusters. Valores cercanos a 1 indican buenos clusters, mientras que valores cercanos a 0 sugieren posibles errores.
- Visualización de los clusters: Mediante las herramientas gráficas de Orange interpretaremos y evaluaremos si los clusters generados tienen sentido práctico con el problema.

##### Naive-Bayes:

- Se utilizará la **precisión** como métrica principal para medir el desempeño del modelo. La precisión es una medida que evalúa qué tan bien el modelo clasifica correctamente las instancias positivas en relación con todas las instancias que ha predicho como positivas. Se calcula como:

$$\text{Precisión} = \frac{\text{Verdaderos Positivos (VP)}}{\text{Verdaderos Positivos (VP)} + \text{Falsos Positivos (FP)}}$$

Esto significa que, de todas las instancias que el modelo predijo como pertenecientes a una clase en particular (en este caso, "**SentimientoFelicidadGeneral**"), la precisión

indica qué proporción de ellas realmente pertenece a esa clase.

- **Técnica de evaluación:**

- Dividiremos el dataset en dos subconjuntos:
  - **Subconjunto de entrenamiento:** 70% de los datos, utilizado para entrenar el modelo.
  - **Subconjunto de pruebas:** 30% de los datos, utilizado para evaluar el modelo.
- Este procedimiento permitirá medir la capacidad del modelo para predecir correctamente la clase "**SentimientoFelicidadGeneral**".

- **Interpretación de resultados:**

- El modelo será evaluado realizando 30 iteraciones para garantizar la estabilidad y confiabilidad de los resultados.
- La precisión promedio obtenida será analizada para determinar si el modelo es adecuado para resolver el problema planteado.

## ii. Descripción de las razones por las que fueron seleccionados dichos modelos.

El problema a tratar puede ser abordado a través de algoritmos de agrupamiento, como el de K-Means, que permite identificar patrones ocultos en los datos y agrupar a las personas con características similares. De forma que se propone agrupar a las personas mediante el tiempo que invierten en sus actividades cotidianas, como el trabajo, traslado, estudios, alimentación, etc. Esto permitirá analizar cómo distribuyen sus hábitos diarios y detectar grupos con estilos de vida similares. Una vez agrupadas, explorar cómo estas

diferencias en la distribución de tiempo afectan sus niveles de bienestar o felicidad.

Naive Bayes fue seleccionado como clasificador supervisado debido a su rapidez, manejo eficiente de datos categóricos y robustez frente a datos desbalanceados. En el contexto del problema, donde buscamos determinar el nivel de felicidad de la población mexicana según sus actividades (esenciales y otras), este clasificador es especialmente adecuado ya que puede manejar categorías claras de felicidad (como 'Muy feliz', 'Feliz', 'Indiferente', etc.) y asociarlas de manera eficiente con patrones en las actividades diarias.

Aunque técnicas como Decision Trees o SVM también fueron consideradas, estas presentaron limitaciones específicas que las hicieron menos adecuadas para los objetivos del análisis. En el caso de Decision Trees, si bien su capacidad de segmentar datos en reglas comprensibles es altamente interpretativa, estas tienden a sobreajustarse a los datos, especialmente si no se aplican métodos de poda adecuados. Además, los árboles de decisión generan estructuras altamente específicas que, aunque útiles en ciertos contextos, no son ideales para problemas con relaciones más difusas o complejas entre las variables, como la relación entre el tiempo dedicado a actividades y los niveles de felicidad.

Por otro lado, SVM es una técnica poderosa para clasificación, especialmente en problemas donde los datos son linealmente separables o donde se requiere manejar límites de decisión complejos. Sin embargo, SVM tiende a ser computacionalmente costoso, particularmente en datasets grandes o multiclase como este, donde el análisis incluye múltiples categorías de felicidad y datos numéricos sobre actividades. Además, ajustar los hiperparámetros de SVM, como el kernel, agrega una capa de complejidad que puede no ser práctica en flujos de trabajo rápidos o exploratorios.

**b. Generación de plan de prueba**

**i. Lista de las diferentes pruebas que se realizarán para cada uno de los conjuntos de datos:**

Esto deberá estar en concordancia con los objetivos de la minería de datos definidos en etapas anteriores. Las pruebas incluyen:

- **Naive-Bayes:** Evaluación de la precisión del modelo para la clasificación de la clase "**SentimientoFelicidadGeneral**".
- **K-Means:** Evaluación de la calidad de los clusters generados utilizando la métrica **Silhouette Score** y herramientas de visualización como Scatter Plot, Distributions y Heat Map para interpretar los resultados de los clusters.

**ii. Indicar los números de iteraciones y técnicas de validación a ser utilizadas (por cada conjunto de datos):**

● **Naive-Bayes:**

- **Número de iteraciones:** Se realizaron 30 iteraciones para garantizar una evaluación confiable y reducir la influencia de la variabilidad.
- **Técnica de validación:** Validación cruzada con **submuestreo aleatorio (Random Subsampling)**.
  - Los datos se dividen aleatoriamente en cada iteración en un 70% para entrenamiento y un 30% para prueba.
  - En cada iteración, se entrena y evalúa el modelo, y al final se calcula la métrica de precisión promedio a lo largo de todas las iteraciones.

● **Estabilidad de los resultados:**

Para garantizar estabilidad, se evaluó la dispersión de las métricas obtenidas mediante el cálculo de la desviación estándar entre las iteraciones. Resultados consistentes presentan baja

variabilidad, lo que indica que el modelo no depende significativamente de una partición específica de los datos.

### **Criterios para interpretar la estabilidad basados en la desviación estándar**

#### **1. Muy estable: Desviación estándar < 0.01**

**Justificación estadística:** Una desviación estándar muy baja significa que la mayoría de las precisiones están extremadamente cercanas a la media. Esto indica que las diferencias en las particiones de los datos no afectan significativamente el rendimiento del modelo.

**Significado práctico:** Un modelo con alta estabilidad es ideal para implementaciones reales, ya que su desempeño será consistente incluso cuando se aplique a diferentes subconjuntos de datos.

#### **2. Moderadamente estable: Desviación estándar entre 0.01 y 0.02**

**Justificación estadística:** Este rango permite cierta variabilidad, pero dentro de límites razonables. Si bien los resultados no son perfectamente consistentes, la dispersión sigue siendo lo suficientemente baja como para considerar el modelo confiable.

**Significado práctico:** Un modelo con estabilidad moderada puede ser adecuado en aplicaciones donde alguna variabilidad es tolerable, como en

problemas con datos más ruidosos o en contextos exploratorios.

### 3. Inestable: Desviación estándar > 0.02

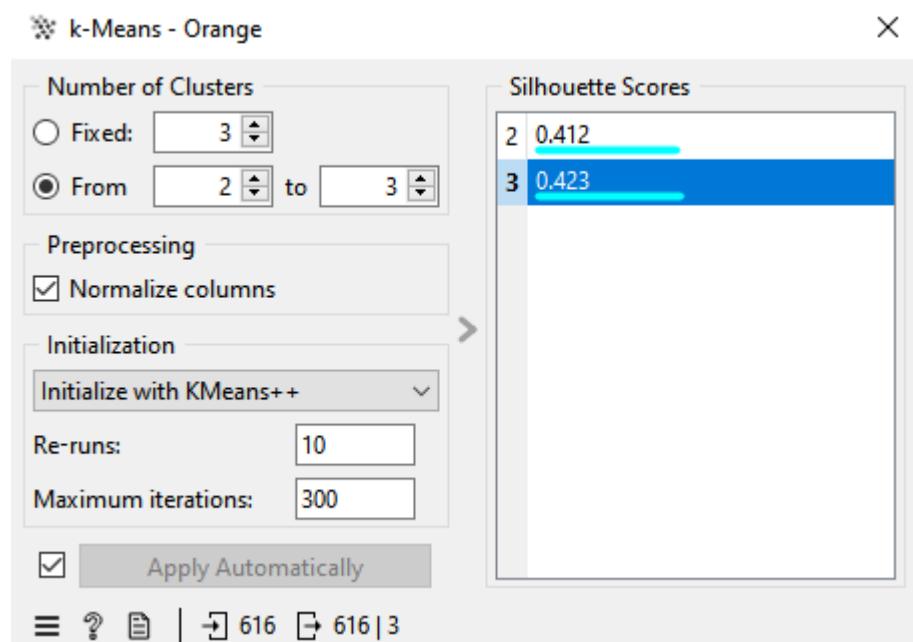
**Justificación estadística:** Una desviación estándar más alta sugiere que el rendimiento del modelo depende significativamente de cómo se dividen los datos. Esto podría deberse a problemas con el conjunto de datos, como insuficiencia de tamaño, ruido, o características desequilibradas.

	Precisión	Desviación estándar	Interpretación
1	0.9933333333333333	0.011502012706317812	Moderadamente estable

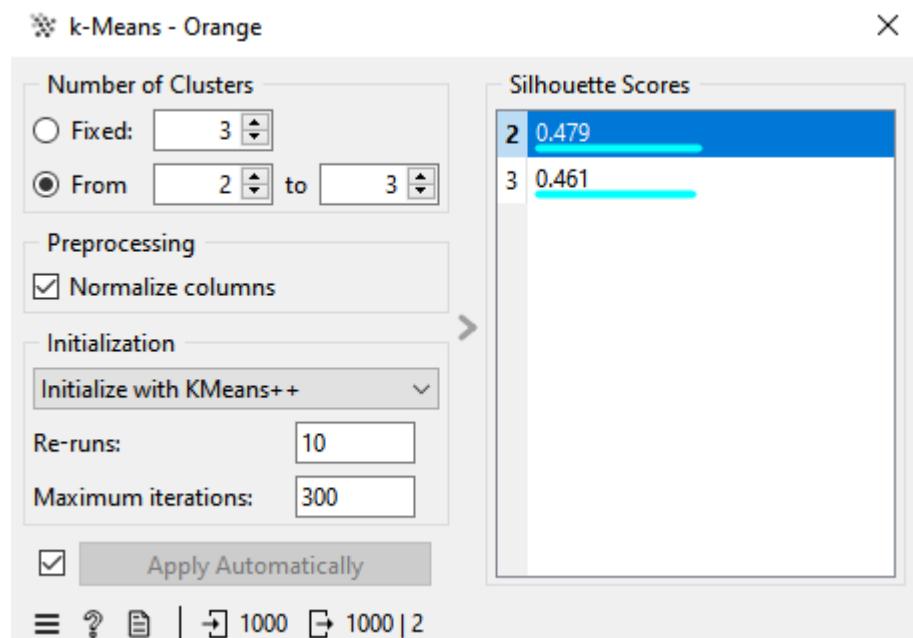
Ejemplo de visualización del resultado

- K-Means:

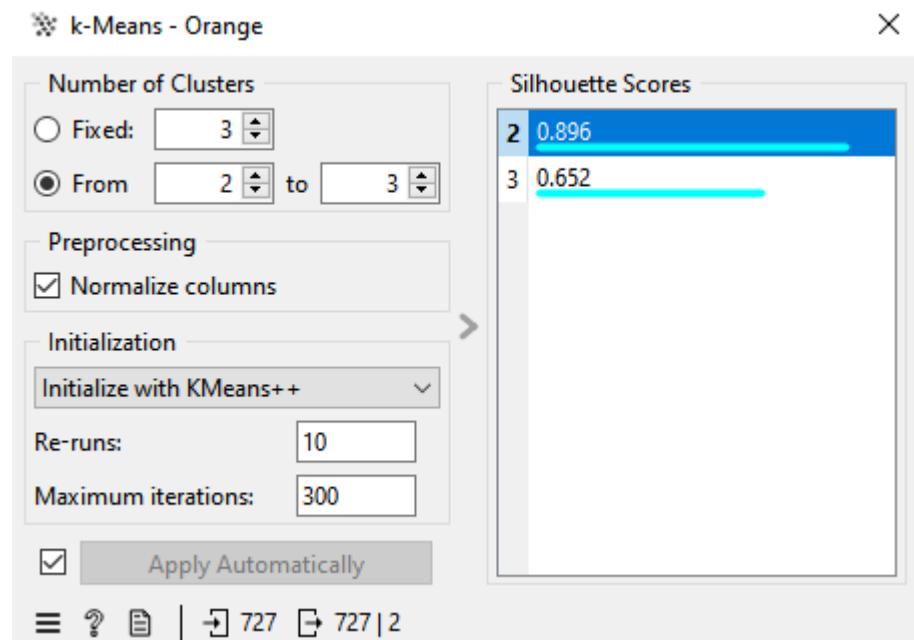
- Conjunto Horas Laborales



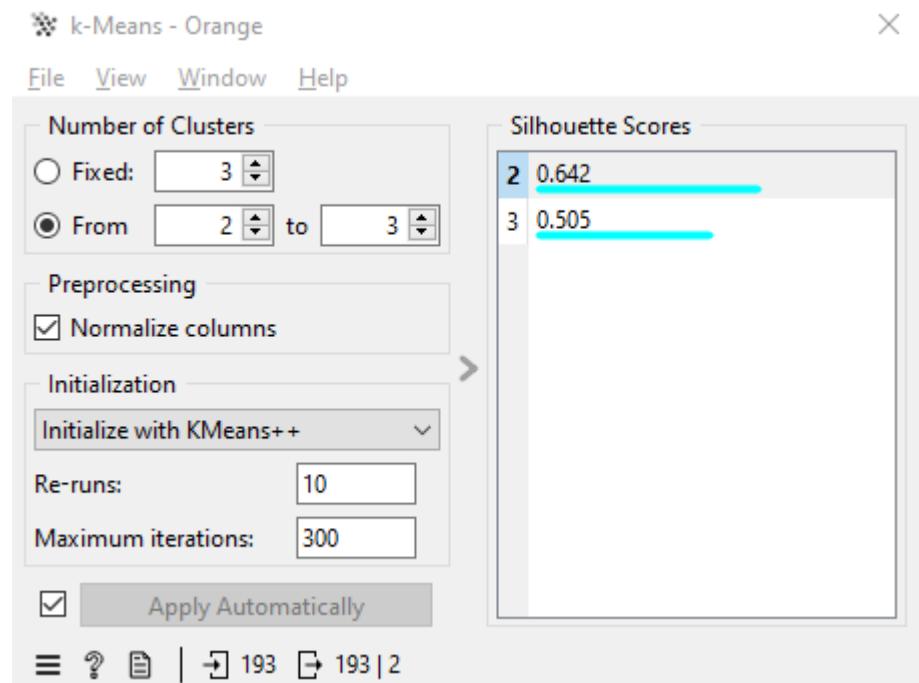
- Conjunto Horas de Descanso



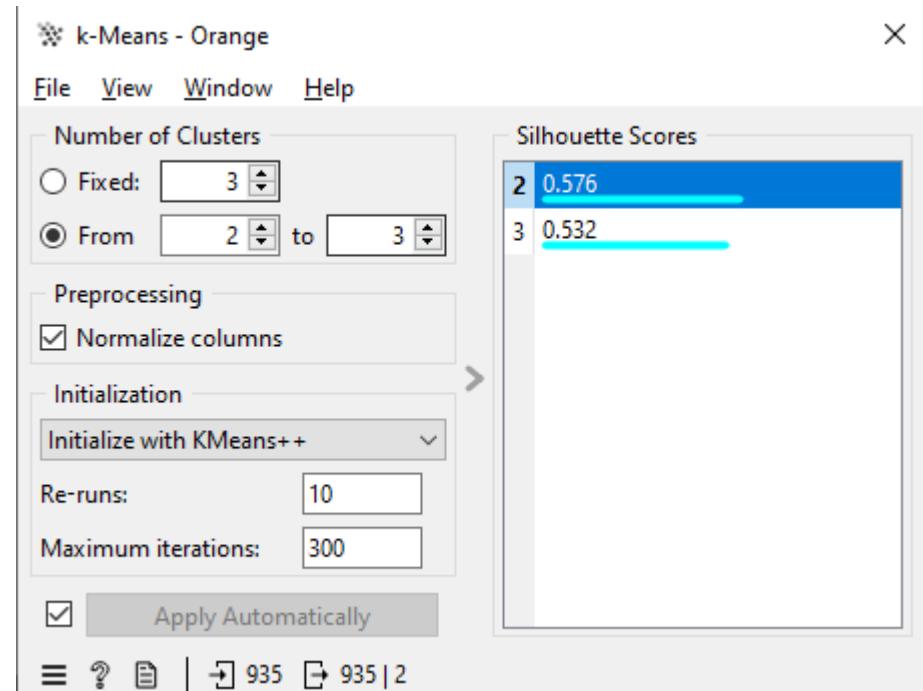
- **Conjunto Horas de Traslado**



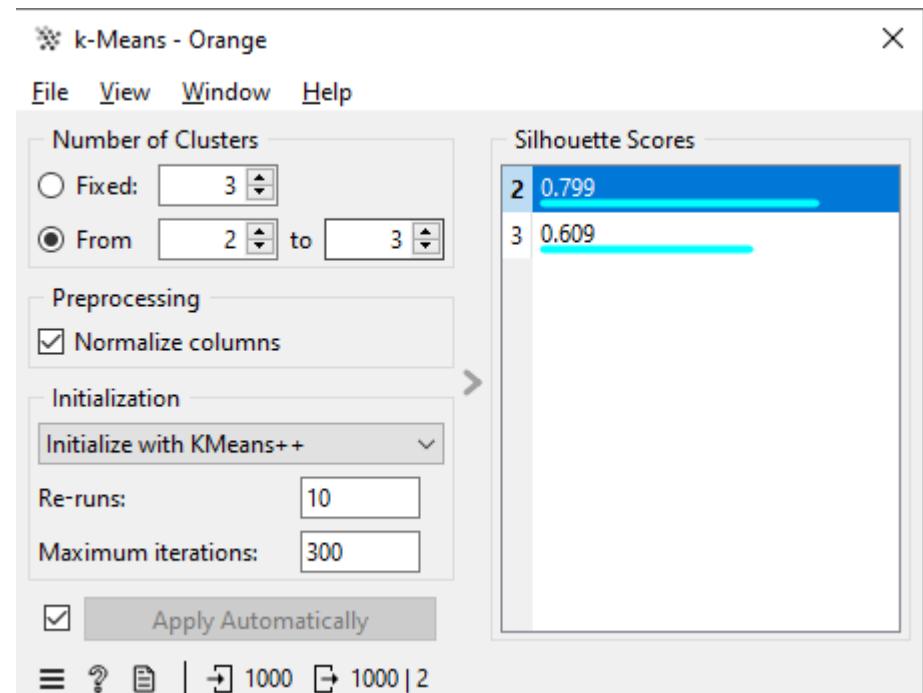
- **Conjunto Horas de Estudio**



○ **Conjunto Horas de Actividades Domésticas.**



○ **Conjunto Situación Económica.**



### c. Construcción de modelos

i. Selección de parámetros: Indicar los parámetros seleccionados para cada uno de los modelos. Esto por cada modelo y conjunto de datos utilizado, si se realizan varias pruebas (diferentes selecciones de parámetros), indicar los parámetros para cada prueba realizada.

#### Naive-Bayes

##### Tipo de Naive Bayes: Gaussiano

Justificación: La elección del Naive Bayes Gaussiano se basa en la naturaleza de los datos, que incluyen una gran cantidad de atributos numéricos continuos. Este tipo de Naive Bayes asume que estos atributos siguen una distribución normal (gaussiana) dentro de cada clase.

Suavizado (Alpha): Aunque el modelo se basa en distribuciones gaussianas para los atributos numéricos, es importante tener en cuenta que el código implementa el suavizado de Laplace ( $\alpha=1$ ) para las variables categóricas presentes en el conjunto de datos. Este suavizado evita probabilidades cero en el cálculo de las probabilidades condicionales, lo que podría mejorar la robustez del modelo. Se podría explorar la optimización de este valor de alpha utilizando técnicas como la validación cruzada para evaluar su impacto en el rendimiento del modelo.

#### K-Means:

- Se realizaron pruebas con diferentes valores de clusters (de 2 a 5), la selección final del número de clusters se basó en el índice de Silhouette, priorizando un equilibrio entre un mayor score y la

interpretabilidad de los grupos (un menor número de clusters puede facilitar interpretaciones claras).

- El número de iteraciones fue configurado en 300 siguiente el estándar determinado, realizamos pruebas con mayor cantidad de iteraciones y no obtuvimos cambios significativos en el aumento del índice de Score para cada conjunto.
- Se utilizó la inicialización k-means++ para garantizar que los centroides iniciales están distribuidos de manera estratégica, evitando resultados sensibles a los puntos iniciales.
- Se utilizaron 1000 instancias normalizadas con *Standardize to*  $\mu=0$ ,  $\sigma=1$  y sin valores faltantes.
- Se utilizó la distancia euclídea que es la métrica estándar utilizada por K-Means en Orange.

## ii. Ejecución de cada una de las pruebas definidas en el punto anterior.

### Naive-Bayes

Para llevar a cabo las pruebas, se desarrolló una implementación personalizada del algoritmo Naive Bayes en Python, sin el uso de bibliotecas externas como scikit-learn. Esta implementación proporciona un control detallado sobre el cálculo de probabilidades y el manejo de datos categóricos y numéricos. A continuación, se explica el proceso con referencias al código correspondiente:

#### 1. Cálculo de Probabilidades:

##### ○ Variables categóricas:

El cálculo de probabilidades condicionales para atributos categóricos se realiza en la función “calcular\_probabilidades\_categoricas”

```

12     def calcular_probabilidades_categoricas(datos: pd.DataFrame, columna: str, clase: str = None):
13         if columna not in datos.columns:
14             raise ValueError(f"La columna '{columna}' no se encuentra en el DataFrame.")
15         if clase is None:
16             return datos[columna].value_counts(normalize=True).to_dict()
17         tabla_frecuencias = datos.groupby([clase, columna]).size().unstack(fill_value=0)
18         tabla_probabilidades = tabla_frecuencias.div(tabla_frecuencias.sum(axis=1), axis=0)
19         if (tabla_probabilidades == 0).any().any():
20             tabla_frecuencias += 1
21             tabla_probabilidades = tabla_frecuencias.div(tabla_frecuencias.sum(axis=1)) + len(
22                 tabla_frecuencias.columns), axis=0)
23     return tabla_probabilidades.to_dict()

```

- Variables numéricas

Las probabilidades condicionales para atributos numéricos, asumiendo una distribución normal, se calculan mediante la función “calcular\_probabilidades\_numericas”, que obtiene la media y desviación estándar de cada clase.

```

24     def calcular_probabilidades_numericas(datos: pd.DataFrame, columna: str, clase: str):
25         if columna not in datos.columns or clase not in datos.columns:
26             raise ValueError(f"Columna '{columna}' o '{clase}' no encontrada.")
27         return datos.groupby(clase)[columna].agg(["mean", "std"]).to_dict()
28

```

## 2. Cálculo de Probabilidades:

El modelo se construye en la función “modelo\_naive\_bayes” que genera un diccionario que almacena:

- Probabilidades a priori de cada clase
- Probabilidades condicionales para cada atributo, categórico o numérico.

```

29 def modelo_naive_bayes(datos: pd.DataFrame, clase: str):
30     if clase not in datos.columns:
31         raise ValueError(f"La columna '{clase}' no se encuentra en el DataFrame.")
32
33     start_time = time.perf_counter()
34
35     # Cálculo de la probabilidad a priori (rápido, un solo hilo)
36     with ThreadPoolExecutor(max_workers=1) as executor:
37         probabilidad_clase_future = executor.submit(calcular_probabilidades_categoricas, datos,
38                                                       clase, None)
39         probabilidad_clase = probabilidad_clase_future.result()
40
41     # Cálculo de probabilidades numéricas (máximo 10 hilos)
42     start_numeric = time.perf_counter()
43     probabilidades_numericas_futures = {}
44     with ThreadPoolExecutor(max_workers=10) as executor:
45         for columna in datos.select_dtypes(include=["number"]).columns:
46             if columna != clase:
47                 probabilidades_numericas_futures[columna] = executor.submit(
48                     calcular_probabilidades_numericas, datos, columna, clase)
49
50     probabilidades_numericas = {columna: future.result() for columna, future in
51     probabilidades_numericas_futures.items()}
52     end_numeric = time.perf_counter()
53
54     # Cálculo de probabilidades categóricas (rápido, un solo hilo)
55     start_categorical = time.perf_counter()
56     probabilidades_categoricas_futures = {}
57     with ThreadPoolExecutor(max_workers=1) as executor:
58         for columna in datos.select_dtypes(include=["object"]).columns:
59             probabilidades_categoricas_futures[columna] = executor.submit(
60                 calcular_probabilidades_categoricas, datos, columna, clase)
61
62     probabilidades_categoricas = {columna: future.result() for columna, future in
63     probabilidades_categoricas_futures.items()}
64     end_categorical = time.perf_counter()
65
66     end_time = time.perf_counter()
67
68     # Impresión de tiempos
69     # print(f"Tiempo total: {(end_time - start_time):.2f} segundos")
70     # print(f"Tiempo en cálculos numéricos: {(end_numeric - start_numeric):.2f} segundos")
71     # print(f"Tiempo en cálculos categóricos: {(end_categorical - start_categorical):.2f} segundos")
72
73     modelo = {
74         "clase": clase,
75         "probabilidad_clase": probabilidad_clase,
76         "probabilidades_numericas": probabilidades_numericas,
77         "probabilidades_categoricas": probabilidades_categoricas
78     }
79
80     return modelo
81

```

### 3. Evaluación del modelo:

La función “evaluar\_naive\_bayes” es responsable de realizar las predicciones sobre el conjunto de pruebas. En este bloque

- Se calcula la probabilidad logarítmica acumulada por clase.
- Se usa la fórmula de la distribución gaussiana “calcular\_probabilidad\_gaussinana” para atributos numéricos.
- Se selecciona la clase con la mayor probabilidad como predicción.

```

1  import pandas as pd
2  import numpy as np
3  from concurrent.futures import ThreadPoolExecutor, as_completed
4  import time
5
6  def calcular_probabilidad_gaussiana(x, mean, std):
7      if std == 0:
8          return 1 if x == mean else 0
9      exponent = np.exp(-(x - mean) ** 2) / (2 * (std ** 2))
10     return (1 / (np.sqrt(2 * np.pi) * std)) * exponent
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77  def evaluar_naive_bayes(modelo, prueba: pd.DataFrame):
78      """Evaluar el modelo Naive Bayes en un conjunto de prueba."""
79      clase = modelo["clase"]
80      probabilidad_clase = modelo["probabilidad_clase"]
81      probabilidades_numericas = modelo["probabilidades_numericas"]
82      probabilidades_categoricas = modelo["probabilidades_categoricas"]
83      columnas_categoricas = prueba.select_dtypes(include=["object"]).columns
84      columnas_numericas = prueba.select_dtypes(include=["number"]).columns
85
86      # Función para evaluar cada fila
87      def evaluar_fila(fila):
88          prob_clases = {}
89
90          for c, prob_c in probabilidad_clase.items():
91              prob_total = np.log(prob_c)
92
93              for columna, prob_columna in probabilidades_categoricas.items():
94                  if columna in columnas_categoricas:
95                      valor = fila[columna]
96                      if valor in prob_columna.get(c, {}):
97                          prob_total += np.log(prob_columna[c].get(valor, 1e-6))
98                      else:
99                          prob_total += np.log(1e-6)
100
101             elif columna in columnas_numericas:
102                 valor = fila[columna]
103                 if c in prob_columna:
104                     mean = prob_columna[c]["mean"]
105                     std = prob_columna[c]["std"]
106                     prob_total += np.log(calcular_probabilidad_gaussiana(valor, mean, std))
107
108             prob_clases[c] = prob_total
109
110         return max(prob_clases, key=prob_clases.get)
111
112     # Cronometrar el tiempo de evaluación de todas las filas
113     start_time = time.perf_counter()
114
115     # Usar multihilos para evaluar cada fila
116     with ThreadPoolExecutor() as executor:
117         predicciones = list(executor.map(evaluar_fila, [fila for _, fila in prueba.iterrows()]))
118
119     end_time = time.perf_counter()
120
121     # Imprimir el tiempo total de evaluación
122     #print(f"Tiempo de evaluación: {end_time - start_time:.4f} segundos")
123
124     return predicciones

```

## Flujo de prueba

- **División de Datos:**

En cada iteración, el conjunto de datos se divide aleatoriamente en dos subconjuntos disjuntos: 70% para

entrenamiento y 30% para prueba. Esta división es aleatoria pero estratificada, garantizando que la distribución de clases se mantenga constante en ambos conjuntos.

- **Ejecución y Visualización:**

Se diseñó una interfaz CLI para ejecutar las pruebas. Los resultados se muestran al finalizar cada iteración en una tabla generada mediante PyQt, donde se resume la precisión obtenida por el modelo en cada ejecución.

*A continuación se explica la interfaz CLI*

## 1. Importaciones y Dependencias Externas

Este bloque carga todas las bibliotecas y módulos necesarios para el funcionamiento del programa.

```
1 import pandas as pd # Importamos la librería pandas para trabajar con dataframes
2 import datos.manejo as manejo
3 import os # Importamos la librería os para limpiar la pantalla
4 import pyfiglet
5 import numpy as np
6
7 from modelos.naive_bayes import modelo_naive_bayes
8 from modelos.naive_bayes import evaluar_naive_bayes
9
10 from PyQt6.QtWidgets import (
11     QApplication, QWidget, QVBoxLayout, QTableWidget, QTableWidgetItem,
12     QHeaderView
13 )
14 from PyQt6.QtGui import QFont
15 from PyQt6.QtCore import Qt
```

## 2. Función para Mostrar DataFrames en PyQt

```
18 # Función para mostrar un DataFrame en una ventana de PyQt
19 def mostrar_dataframe_en_qt(df: pd.DataFrame, titulo: str):
20     ventana = QWidget()
21     ventana.setWindowTitle(titulo)
22     ventana.setGeometry(100, 100, 800, 600)
23
24     layout = QVBoxLayout()
25
26     # Crear tabla
27     tabla = QTableWidget()
28     tabla.setRowCount(len(df))
29     tabla.setColumnCount(len(df.columns))
30     tabla.setHorizontalHeaderLabels(df.columns)
31
32     # Establecer fuente más grande
33     tabla.setFont(QFont("Arial", 12))
34
35     # Ajustar encabezados
36     header = tabla.horizontalHeader()
37     header.setSectionResizeMode(QHeaderView.ResizeMode.ResizeToContents)
38
39     # Insertar filas
40     for i, fila in df.iterrows():
41         for j, valor in enumerate(fila):
42             tabla.setItem(i, j, QTableWidgetItem(str(valor)))
43
44     layout.addWidget(tabla)
45     ventana.setLayout(layout)
46
47     ventana.show()
48     return ventana
```

26 NOV 2024  
- 11:26 PM -

## 3. Variables Globales de Colores ANSI

Define colores y estilos para el texto en la terminal.

```
50 # VARIABLES GLOBALES PARA COLORES ANSI
51 RESET = "\u001b[0m"
52 NEGRO = "\u001b[30m"
53 ROJO = "\u001b[31m"
54 VERDE = "\u001b[32m"
55 AMARILLO = "\u001b[33m"
56 AZUL = "\u001b[34m"
57 MAGENTA = "\u001b[35m"
58 CYAN = "\u001b[36m"
59 BLANCO = "\u001b[37m"
60 NEGRITA = "\u001b[1m"
61 SUBRAYADO = "\u001b[4m"
```

## 4. Funciones Utilitarias

- **Limpiar pantalla:** Se adapta según el sistema operativo.

```

64  # Función para limpiar la pantalla según el sistema operativo
65  def limpiar_pantalla():
66      if os.name == 'nt':
67          os.system('cls')
68      else:
69          os.system('clear')
70

```

- **Mostrar datos actuales:** Presenta los datos cargados en el programa.

```

70  # Función para mostrar los datos actuales
71  def mostrar_datos_actuales(ruta_dataset=None, dataset=None, tam_dataset_prueba=None, clase=None):
72      """Muestra los datos agregados al programa."""
73      if any([ruta_dataset, dataset, tam_dataset_prueba, clase]):
74          print(CYAN + NEGRITA + "DATOS ACTUALMENTE AGREGADOS" + RESET) 26 NOV 2024
75          if ruta_dataset:
76              print(MAGENTA + "Ruta dataset: " + RESET, ruta_dataset)
77          if dataset is not None:
78              print(MAGENTA + "Dataset cargado." + RESET)
79          if tam_dataset_prueba:
80              print(MAGENTA + "Tamaño dataset prueba: " + RESET, tam_dataset_prueba)
81          if clase:
82              print(MAGENTA + "Clase seleccionada: " + RESET, clase)
83          else:
84              print(AMARILLO + "No hay datos agregados." + RESET)
85

```

## 5. Funciones para la Interacción con el Usuario

- **Cargar datos:** Solicita al usuario la ruta de un dataset y lo carga.

```

# Función para cargar los datos
def cargar_datos():
    """Carga un dataset en formato CSV."""
    limpiar_pantalla()
    while True:
        print(AZUL + NEGRITA + "CARGAR DATOS" + RESET)
        ruta_dataset = input("Ingrese la ruta del dataset: ")
        try:
            dataset = manejo.cargar_datos(ruta_dataset)
            return ruta_dataset, dataset
        except Exception as e:
            limpiar_pantalla()
            print(ROJO + "Error al cargar el dataset" + RESET)
            print(AMARILLO + str(e) + RESET)

```

- **Seleccionar tamaño del dataset de prueba:** Permite al usuario definir el porcentaje del dataset para pruebas.

```

101 # Función para seleccionar el tamaño del dataset de prueba
102 def seleccionar_tam_dataset_prueba():
103     """Selecciona el tamaño del dataset de prueba."""
104     limpiar_pantalla()
105     while True:
106         try:
107             print(AZUL + NEGRITA + "TAMAÑO DEL DATASET DE PRUEBA" + RESET)
108             print("Por ejemplo, si se selecciona 0.2, el 20% del dataset se usará para pruebas.")
109             tam_dataset_prueba = float(input("Ingrese el tamaño del dataset de prueba (entre 0 y 1): "))
110             if 0 < tam_dataset_prueba < 1:
111                 return tam_dataset_prueba
112             else:
113                 limpiar_pantalla()
114                 print(ROJO + "El valor debe estar entre 0 y 1" + RESET)
115         except ValueError:
116             limpiar_pantalla()
117             print(ROJO + "Error: Debe ingresar un número válido entre 0 y 1." + RESET)
118

```

26 NOV 2024  
- 11:36 PM -

- **Mostrar menú principal:** Despliega las opciones principales.

```

119 # Función para mostrar opciones del menú principal
120 def mostrar_menu_principal():
121     """Muestra el menú principal del programa."""
122     print(AZUL + NEGRITA + "MENU PRINCIPAL" + RESET)
123     print(VERDE + "1. Cargar datos")
124     print("2. Porcentaje del dataset para pruebas")
125     print("3. Seleccionar clase")
126     print("4. Evaluar Modelo")
127     print("5. Salir" + RESET)
128     return input("Seleccione una opción: ")
129

```

- **Seleccionar clase:** Permite elegir la columna objetivo del dataset.

```

# Función para seleccionar la clase
def seleccionar_clase(dataset):
    """Función para seleccionar la clase."""
    limpiar_pantalla()
    if dataset is None:
        raise ValueError("No se ha cargado ningún dataset.")
    print(AZUL + NEGRITA + "SELECCIONAR CLASE" + RESET)
    print(AMARILLO + "Columnas disponibles:" + RESET)
    contador = 1
    for columna in dataset.columns:
        print(f"{contador}. {columna}")
        contador += 1
    seleccion = int(input("Seleccione una columna: "))
    return dataset.columns[seleccion - 1]

```

## 6. Evaluación del Modelo Naive Bayes

```
146 # Función para evaluar el modelo (vacía)
147 def evaluar_modelo(dataset: pd.DataFrame, tam_dataset_prueba, clase):
148     """Función para evaluar el modelo (A implementar)."""
149     # Verificamos si se han cargado los datos correctamente
150     if dataset.empty:
151         raise ValueError("El dataset está vacío o no se ha cargado.")
152     if tam_dataset_prueba is None or not (0 < tam_dataset_prueba < 1):
153         raise ValueError("El tamaño del dataset de prueba no es válido o no ha sido seleccionado.
154     ")
155     if not clase:
156         raise ValueError("No se ha especificado la clase a predecir.")
157     print(AZUL + NEGRITA + "EVALUAR MODELO" + RESET)
158
159     # Preguntamos si desea continuar
160     continuar = input("¿Desea continuar con la evaluación del modelo? (s/n): ")
161     if continuar.lower() != "s" and continuar.lower() != "si" and continuar.lower() != "y":
162         return
163     # Pedimos el número de iteraciones
164     while True:
165         try:
166             iteraciones = int(input("Ingrese el número de iteraciones: "))
167             break
168         except ValueError:
169             print(ROJO + "Error: Debe ingresar un número válido." + RESET)
170
171     # Evaluamos el modelo
172     arregloResultados = []
173     arregloPrediccionExitosa = []
174     arregloPrueba = []
175     for i in range(iteraciones):
176         datos_entrenamiento, datos_prueba = manejo.generador_de_subconjuntos(dataset,
177             tam_dataset_prueba, i)
178         modelo = modelo_naive_bayes(datos_entrenamiento, clase)
179         #print(f"Modelo Naive Bayes: {modelo}")
180         resultado = evaluar_naive_bayes(modelo, datos_prueba)
181         arregloResultados.append(resultado)
182         # Calcular la precisión del modelo usando los datos de prueba
183         prediccion_exitosa = datos_prueba[clase] == resultado
184         precision = prediccion_exitosa.sum() / len(prediccion_exitosa)
185         arregloPrediccionExitosa.append(precision)
186         print(f"Precision: {precision}") # Desviación estándar: 0.01381044514350382 Moderadamente estable
187         # Guardar los datos de prueba de la clase
188         arregloPrueba.append(datos_prueba[clase])
189         # presione cualquier tecla para continuar 0.01381044514350382 Moderadamente estable
190         input("Presione cualquier tecla para continuar...")
191     return arregloPrediccionExitosa
```

Incluye:

- Validación de datos.
- Solicitud de número de iteraciones.
- Entrenamiento y evaluación del modelo.
- Cálculo de precisión y desviación estándar.
- Interpretación de resultados.

## 7. Ciclo Principal del Programa (Menú)

```
193 def main():
194     dataset = None
195     tam_dataset_prueba = None
196     clase = None
197
198     # Limpiar pantalla inicial
199     limpiar_pantalla()
200
201     # Ciclo principal del menú
202     while True:
203         try:
204             # Mostrar datos actuales
205             mostrar_datos_actuales(ruta_data_set, dataset, tam_dataset_prueba, clase)
206
207             # Mostrar menú y obtener la opción seleccionada
208             seleccion = mostrar_menu_principal()
209
210             # Procesar la selección del usuario
211             if seleccion == "1":
212                 ruta_data_set, dataset = cargar_datos()
213                 limpiar_pantalla()
214
215             elif seleccion == "2":
216                 tam_dataset_prueba = seleccionar_tam_dataset_prueba()
217                 limpiar_pantalla()
218
219             elif seleccion == "3":
220                 clase = seleccionar_clase(dataset)
221                 limpiar_pantalla()
222
223             elif seleccion == "4":
224                 limpiar_pantalla()
225                 mostrar_datos_actuales(ruta_data_set, dataset, tam_dataset_prueba, clase)
226                 #evaluar_modelo(dataset, tam_dataset_prueba, modelo_selec, clase)
227                 predicciones = evaluar_modelo(dataset, tam_dataset_prueba, clase)
228                 # creamos un DataFrame con los resultados
229                 df = pd.DataFrame(predicciones, columns=["Precisión"])
230                 desviacion_estandar = np.std(predicciones)
231                 # agregamos la desviación estándar al DataFrame como una columna nueva
232                 df["Desviación estándar"] = desviacion_estandar
233                 # Agregar la interpretación basada en criterios
234                 if desviacion_estandar < 0.01:
235                     interpretacion = "Muy estable"
236                 elif 0.01 <= desviacion_estandar <= 0.02:
237                     interpretacion = "Moderadamente estable"
238                 else:
239                     interpretacion = "Inestable"
240
241             # Resultados de las iteraciones
242             df["Interpretación"] = interpretacion
243
244             # Mostraremos el DataFrame en una ventana de PyQt
245             app = QApplication([])
246             ventana = mostrar_dataframe_en_qt(df, "Resultados de las iteraciones")
247             app.exec_()
248             limpiar_pantalla()
249
250             elif seleccion == "5":
251                 print(AMARILLO + "Saliendo..." + RESET)
252                 ascii_art = pyfiglet.figlet_format("Desarrollado por:")
253                 print(ROJO + ascii_art + RESET)
254                 ascii_art = pyfiglet.figlet_format("Aether")
255                 print(CYAN + ascii_art + RESET)
256                 break
257
258             else:
259                 limpiar_pantalla()
260                 print(ROJO + "Opción no válida" + RESET)
261
262             0.99 0.01381044514350382 Moderadamente estable
263         except Exception as e:
264             # En caso de cualquier excepción, limpiamos la pantalla y mostramos un error legible
265             limpiar_pantalla()
266             print(ROJO + "Informe de error" + RESET)
267             print(str(e)) # Convertimos la excepción a cadena para mostrarla correctamente
268
269         return 0 0.01381044514350382 Moderadamente estable
270
271         0.99666666666666667 0.01381044514350382 Moderadamente estable
272
```

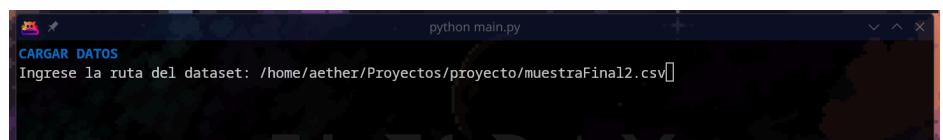
## 8. Punto de Entrada Principal

```
362     # Iniciar programa
363     if __name__ == "__main__":
364         main()
365
```

*Ejecución del CLI*



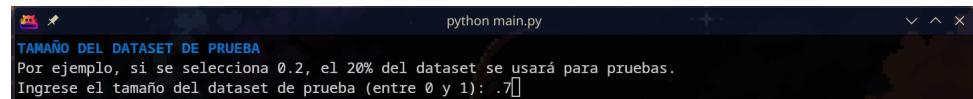
*Procedemos a cargar el dataset con 1*



*Ingresamos la ruta y damos enter*

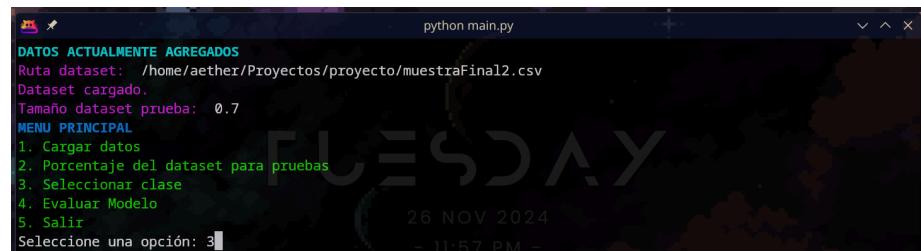


*Colocamos 2 para colocar el porcentaje del dataset para pruebas y damos enter*



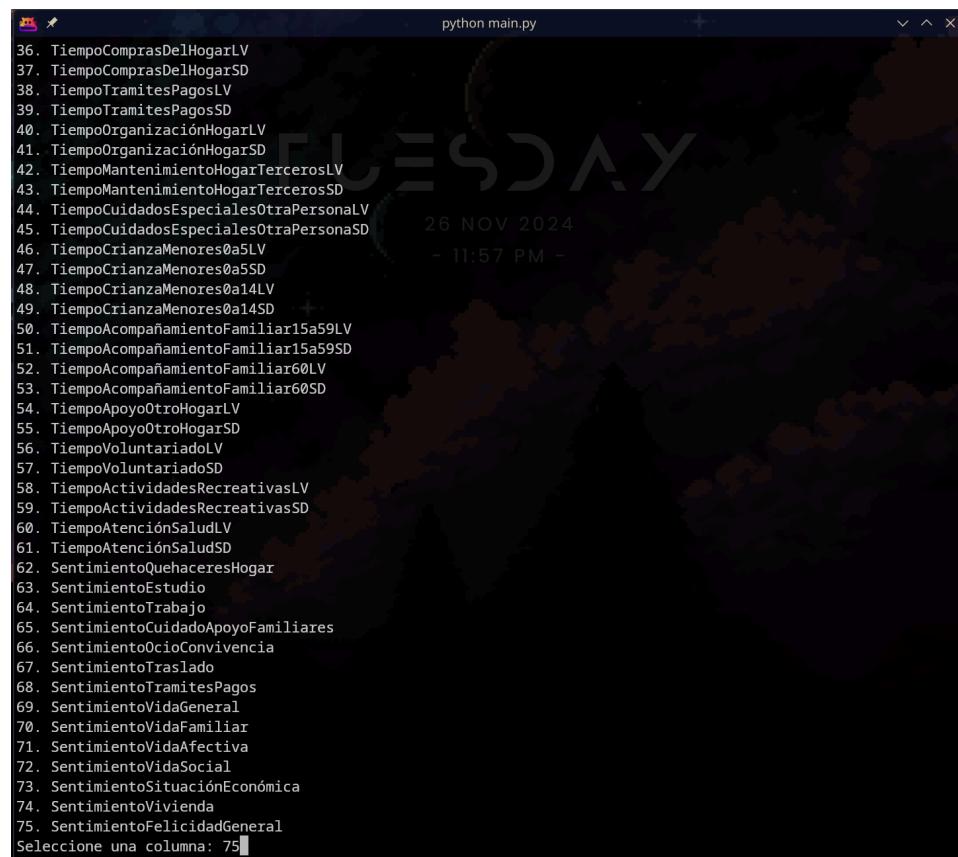
```
python main.py
TAMAÑO DEL DATASET DE PRUEBA
Por ejemplo, si se selecciona 0.2, el 20% del dataset se usará para pruebas.
Ingrese el tamaño del dataset de prueba (entre 0 y 1): .7
```

*Colocamos el tamaño y damos enter*



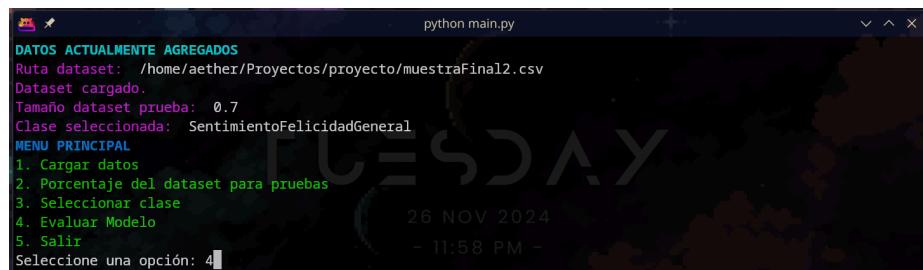
```
DATOS ACTUALMENTE AGREGADOS
Ruta dataset: /home/aether/Proyectos/proyecto/muestraFinal2.csv
Dataset cargado.
Tamaño dataset prueba: 0.7
MENU PRINCIPAL
1. Cargar datos
2. Porcentaje del dataset para pruebas
3. Seleccionar clase
4. Evaluar Modelo
5. Salir
Seleccione una opción: 3
```

*Colocamos 3 para seleccionar la clase objetivo*



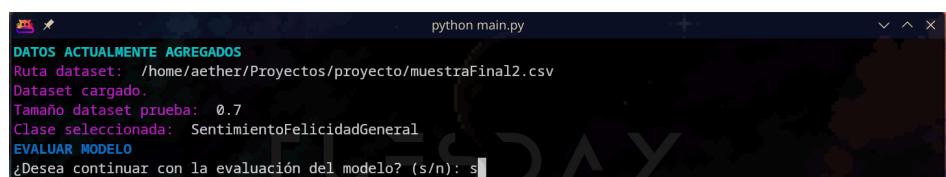
```
36. TiempoComprasDelHogarLV
37. TiempoComprasDelHogarSD
38. TiempoTramitesPagosLV
39. TiempoTramitesPagosSD
40. TiempoOrganizaciónHogarLV
41. TiempoOrganizaciónHogarSD
42. TiempoMantenimientoHogarTercerosLV
43. TiempoMantenimientoHogarTercerosSD
44. TiempoCuidadosEspecialesOtraPersonaLV
45. TiempoCuidadosEspecialesOtraPersonaSD
46. TiempoCrianzaMenores0a5LV
47. TiempoCrianzaMenores0a5SD
48. TiempoCrianzaMenores0a14LV
49. TiempoCrianzaMenores0a14SD
50. TiempoAcompañamientoFamiliar15a59LV
51. TiempoAcompañamientoFamiliar15a59SD
52. TiempoAcompañamientoFamiliar60LV
53. TiempoAcompañamientoFamiliar60SD
54. TiempoApoyoOtroHogarLV
55. TiempoApoyoOtroHogarSD
56. TiempoVoluntariadoLV
57. TiempoVoluntariadoSD
58. TiempoActividadesRecreativasLV
59. TiempoActividadesRecreativasSD
60. TiempoAtenciónSaludLV
61. TiempoAtenciónSaludSD
62. SentimientoQuehaceresHogar
63. SentimientoEstudio
64. SentimientoTrabajo
65. SentimientoCuidadoApoyoFamiliares
66. SentimientoOcioConvivencia
67. SentimientoTraslado
68. SentimientoTramitesPagos
69. SentimientoVidaGeneral
70. SentimientoVidaFamiliar
71. SentimientoVidaAfactiva
72. SentimientoVidaSocial
73. SentimientoSituaciónEconómica
74. SentimientoVivienda
75. SentimientoFelicidadGeneral
Seleccione una columna: 75
```

*Colocamos 75 para seleccionar SentimientoFelicidadGeneral*



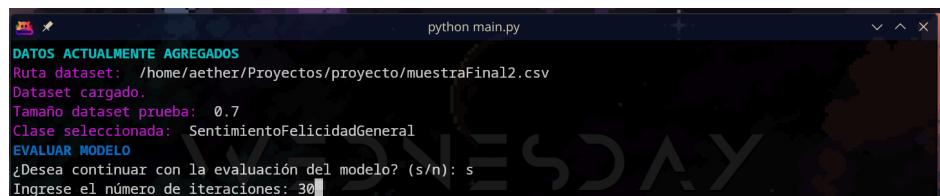
```
python main.py
DATOS ACTUALMENTE AGREGADOS
Ruta dataset: /home/aether/Proyectos/proyecto/muestraFinal2.csv
Dataset cargado.
Tamaño dataset prueba: 0.7
Clase seleccionada: SentimientoFelicidadGeneral
MENU PRINCIPAL
1. Cargar datos
2. Porcentaje del dataset para pruebas
3. Seleccionar clase
4. Evaluar Modelo
5. Salir
Seleccione una opción: 4
```

*Colocamos 4 para evaluar el modelo*



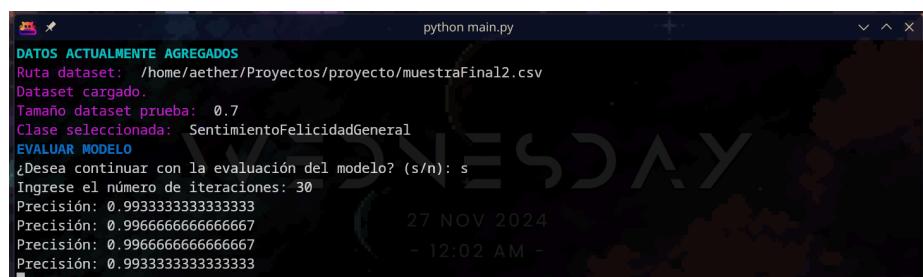
```
python main.py
DATOS ACTUALMENTE AGREGADOS
Ruta dataset: /home/aether/Proyectos/proyecto/muestraFinal2.csv
Dataset cargado.
Tamaño dataset prueba: 0.7
Clase seleccionada: SentimientoFelicidadGeneral
EVALUAR MODELO
¿Desea continuar con la evaluación del modelo? (s/n): s
```

Se muestra un resumen de la selección y nos pregunta si deseamos continuar, damos s y enter



```
python main.py
DATOS ACTUALMENTE AGREGADOS
Ruta dataset: /home/aether/Proyectos/proyecto/muestraFinal2.csv
Dataset cargado.
Tamaño dataset prueba: 0.7
Clase seleccionada: SentimientoFelicidadGeneral
EVALUAR MODELO
¿Desea continuar con la evaluación del modelo? (s/n): s
Ingrese el número de iteraciones: 30
```

Nos pide el número de iteraciones, para nuestra evaluación necesitamos 30 así que colocamos 30, damos enter y esperamos a que termine de evaluar el modelo



```
python main.py
DATOS ACTUALMENTE AGREGADOS
Ruta dataset: /home/aether/Proyectos/proyecto/muestraFinal2.csv
Dataset cargado.
Tamaño dataset prueba: 0.7
Clase seleccionada: SentimientoFelicidadGeneral
EVALUAR MODELO
¿Desea continuar con la evaluación del modelo? (s/n): s
Ingrese el número de iteraciones: 30
Precisión: 0.9933333333333333
Precisión: 0.9966666666666667
Precisión: 0.9966666666666667
Precisión: 0.9933333333333333
```

*Conforme termine cada iteración va mostrando la precisión de la misma*

```
python main.py
DATOS ACTUALMENTE AGREGADOS
Ruta dataset: /home/aether/Proyectos/proyecto/muestraFinal2.csv
Dataset cargado.
Tamaño dataset prueba: 0.7
Clase seleccionada: SentimientoFelicidadGeneral
EVALUAR MODELO
¿Desea continuar con la evaluación del modelo? (s/n): s
Ingrese el número de iteraciones: 30
Precisión: 0.9933333333333333
Precisión: 0.9966666666666667
Precisión: 0.9966666666666667
Precisión: 0.9933333333333333
Precisión: 0.9966666666666667
Precisión: 0.9966666666666667
Precisión: 0.9933333333333333
Precisión: 0.9866666666666667
Precisión: 0.99
Precisión: 0.9966666666666667
Precisión: 0.9533333333333334
Precisión: 0.99
Precisión: 0.99
Precisión: 0.9933333333333333
Precisión: 0.9933333333333333
Precisión: 0.9933333333333333
Precisión: 0.9966666666666667
Precisión: 0.9933333333333333
Precisión: 0.9866666666666667
Precisión: 0.99
Precisión: 0.9466666666666667
Precisión: 1.0
Precisión: 0.9933333333333333
Precisión: 1.0
Precisión: 0.9933333333333333
Precisión: 0.9966666666666667
Presione cualquier tecla para continuar...[]
```

*Una vez termina damos enter*

	Precisión	Desviación estándar	Interpretación
1	0.9933333333333333	0.011376822876595733	Moderadamente estable
2	0.99666666666666667	0.011376822876595733	Moderadamente estable
3	0.99666666666666667	0.011376822876595733	Moderadamente estable
4	0.9933333333333333	0.011376822876595733	Moderadamente estable
5	0.99666666666666667	0.011376822876595733	Moderadamente estable
6	0.99666666666666667	0.011376822876595733	Moderadamente estable
7	0.9933333333333333	0.011376822876595733	Moderadamente estable
8	0.98666666666666667	0.011376822876595733	Moderadamente estable
9	0.99	0.011376822876595733	Moderadamente estable
10	0.99666666666666667	0.011376822876595733	Moderadamente estable
11	0.9533333333333334	0.011376822876595733	Moderadamente estable
12	0.99	0.011376822876595733	Moderadamente estable
13	0.99	0.011376822876595733	Moderadamente estable
14	0.9933333333333333	0.011376822876595733	Moderadamente estable
15	0.9933333333333333	0.011376822876595733	Moderadamente estable
16	0.9933333333333333	0.011376822876595733	Moderadamente estable
17	0.9933333333333333	0.011376822876595733	Moderadamente estable
18	0.99666666666666667	0.011376822876595733	Moderadamente estable
19	0.9933333333333333	0.011376822876595733	Moderadamente estable
20	0.98666666666666667	0.011376822876595733	Moderadamente estable
21	0.98666666666666667	0.011376822876595733	Moderadamente estable
22	0.99	0.011376822876595733	Moderadamente estable
23	0.94666666666666667	0.011376822876595733	Moderadamente estable
24	1.0	0.011376822876595733	Moderadamente estable
25	0.9933333333333333	0.011376822876595733	Moderadamente estable
26	1.0	0.011376822876595733	Moderadamente estable
27	0.9933333333333333	0.011376822876595733	Moderadamente estable
28	0.99666666666666667	0.011376822876595733	Moderadamente estable
29	0.99	0.011376822876595733	Moderadamente estable
30	0.9933333333333333	0.011376822876595733	Moderadamente estable

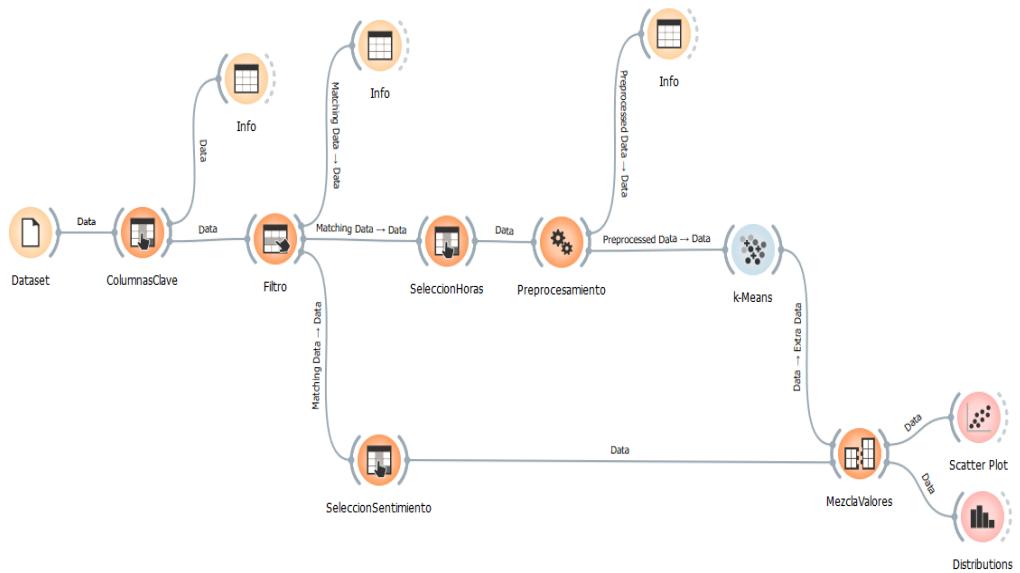
Nos genera esta tabla con los resultados de cada iteración, la desviación estándar y la estabilidad del modelo.

### K-Means:

Las pruebas fueron implementadas en Orange haciendo uso de las herramientas, el workspace y el dataset previamente limpio y transformado. Se realizaron diferentes modelos con diferentes variables y la misma metodología con el objetivo de que se tengan los valores más coherentes y certeros.

## Conjuntos:

- **Horas Laboradas**
- **Horas de Descanso**
- **Horas de Traslado**
- **Horas de Estudio**
- **Horas de Actividades Domésticas**
- **Situación Económica.**



Flujo de trabajo para los conjuntos en Orange

-iii. Interpretación de los resultados, desde el punto de vista de minería de datos, no del negocio.

### Naive Bayes

Muestra 1000 elementos

	Precisión	Desviación estándar	Interpretación
1	0.9933333333333333	0.011376822876595733	Moderadamente estable
2	0.9966666666666666	0.011376822876595733	Moderadamente estable
3	0.9966666666666666	0.011376822876595733	Moderadamente estable
4	0.9933333333333333	0.011376822876595733	Moderadamente estable
5	0.9966666666666666	0.011376822876595733	Moderadamente estable
6	0.9966666666666666	0.011376822876595733	Moderadamente estable
7	0.9933333333333333	0.011376822876595733	Moderadamente estable
8	0.9866666666666666	0.011376822876595733	Moderadamente estable
9	0.99	0.011376822876595733	Moderadamente estable
10	0.9966666666666666	0.011376822876595733	Moderadamente estable
11	0.9533333333333334	0.011376822876595733	Moderadamente estable
12	0.99	0.011376822876595733	Moderadamente estable
13	0.99	0.011376822876595733	Moderadamente estable
14	0.9933333333333333	0.011376822876595733	Moderadamente estable
15	0.9933333333333333	0.011376822876595733	Moderadamente estable
16	0.9933333333333333	0.011376822876595733	Moderadamente estable
17	0.9933333333333333	0.011376822876595733	Moderadamente estable
18	0.9966666666666666	0.011376822876595733	Moderadamente estable
19	0.9933333333333333	0.011376822876595733	Moderadamente estable
20	0.9866666666666666	0.011376822876595733	Moderadamente estable
21	0.9866666666666666	0.011376822876595733	Moderadamente estable
22	0.99	0.011376822876595733	Moderadamente estable
23	0.9466666666666666	0.011376822876595733	Moderadamente estable
24	1.0	0.011376822876595733	Moderadamente estable
25	0.9933333333333333	0.011376822876595733	Moderadamente estable
26	1.0	0.011376822876595733	Moderadamente estable
27	0.9933333333333333	0.011376822876595733	Moderadamente estable
28	0.9966666666666666	0.011376822876595733	Moderadamente estable
29	0.99	0.011376822876595733	Moderadamente estable
30	0.9933333333333333	0.011376822876595733	Moderadamente estable

#### Consistencia del Modelo (Desviación Estándar de 0.0114):

En minería de datos, una baja desviación estándar como la observada (**0.0114**) indica que el modelo es **consistente** en sus predicciones. Esto

sugiere que el modelo mantiene un rendimiento similar a través de distintas particiones del conjunto de datos. Este comportamiento es deseable, ya que significa que el modelo no es excesivamente sensible a la selección aleatoria de datos de entrenamiento y prueba.

#### **Precisión Alta y Distribuida (Rango 0.9466 a 1.0):**

La mayoría de las precisiones se encuentran entre **0.99 y 1.0**, lo que evidencia una **alta capacidad predictiva** del modelo. En minería de datos, este resultado indica que el modelo ha aprendido correctamente las relaciones entre los atributos y las clases objetivo. Las precisiones más bajas (**0.9466 y 0.9533**) podrían estar asociadas a particiones del conjunto de datos con características más desbalanceadas o complejas.

#### **Estabilidad Moderada del Modelo:**

El modelo se clasifica como **moderadamente estable** debido a que, aunque la mayoría de los resultados están agrupados en un rango estrecho, algunas iteraciones muestran ligeras variaciones. Esto puede indicar que ciertos subconjuntos de datos presentan patrones atípicos o ruidos que afectan ligeramente el rendimiento del modelo.

#### **Generalización Adecuada:**

El modelo parece tener una buena capacidad de **generalización**, es decir, no se sobreajusta a los datos de entrenamiento y mantiene un rendimiento robusto en los conjuntos de prueba. Esto es fundamental en minería de datos, ya que un modelo que generaliza bien es capaz de hacer predicciones precisas en datos nuevos y no vistos.

#### **Posibles Causas de Variación:**

- **Distribución de clases:** Aunque se garantiza la estratificación, ciertas clases podrían tener menos representación, lo que podría afectar el cálculo de probabilidades.
- **Datos numéricos:** La suposición de distribución normal para atributos numéricos podría ser una fuente de error si los datos no siguen esta distribución.

- **Ruido o outliers:** La presencia de datos atípicos puede influir en la precisión en algunas iteraciones.

### *Dataset completo*

	Precisión	Desviación estándar	Interpretación
1	1.0	0.0	Muy estable
2	1.0	0.0	Muy estable
3	1.0	0.0	Muy estable
4	1.0	0.0	Muy estable
5	1.0	0.0	Muy estable
6	1.0	0.0	Muy estable
7	1.0	0.0	Muy estable
8	1.0	0.0	Muy estable
9	1.0	0.0	Muy estable
10	1.0	0.0	Muy estable
11	1.0	0.0	Muy estable
12	1.0	0.0	Muy estable
13	1.0	0.0	Muy estable
14	1.0	0.0	Muy estable
15	1.0	0.0	Muy estable
16	1.0	0.0	Muy estable
17	1.0	0.0	Muy estable
18	1.0	0.0	Muy estable
19	1.0	0.0	Muy estable
20	1.0	0.0	Muy estable
21	1.0	0.0	Muy estable
22	1.0	0.0	Muy estable
23	1.0	0.0	Muy estable
24	1.0	0.0	Muy estable
25	1.0	0.0	Muy estable
26	1.0	0.0	Muy estable
27	1.0	0.0	Muy estable
28	1.0	0.0	Muy estable
29	1.0	0.0	Muy estable
30	1.0	0.0	Muy estable

### **Precisión Perfecta (1.0 en todas las iteraciones):**

El modelo ha mostrado precisión perfecta en todas las ejecuciones, lo que significa que ha clasificado correctamente todas las instancias en el conjunto de datos completo.

### **Riesgo de Sobreajuste:**

Aunque la precisión perfecta es una excelente señal de rendimiento, también existe el riesgo de sobreajuste si el modelo ha sido entrenado de manera excesiva sobre el conjunto de datos. En este caso, dado que la precisión es del 100% en todas las iteraciones, podríamos preguntarnos si el modelo ha aprendido detalles muy específicos de los datos, lo que podría no generalizar bien a nuevos datos fuera de este conjunto.

### **Desviación Estándar de 0:**

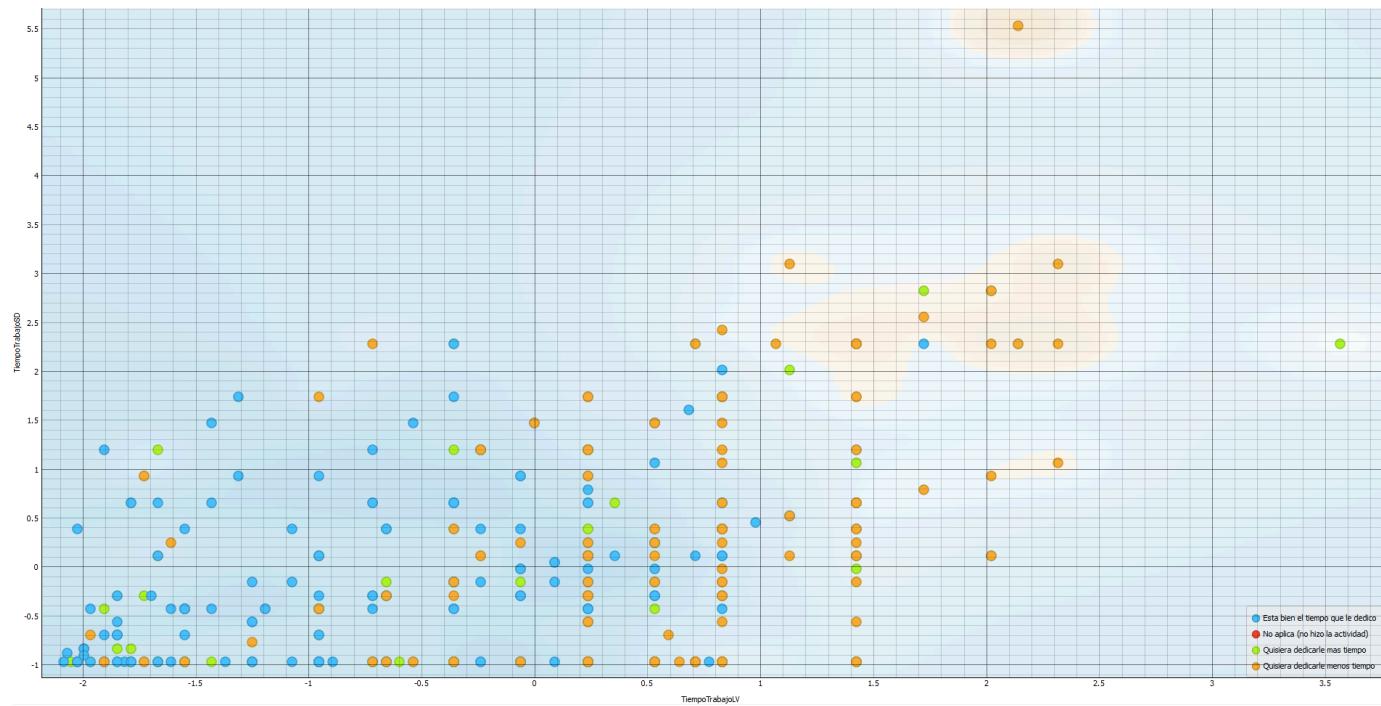
Una desviación estándar de 0 indica que el modelo tiene una estabilidad total en sus predicciones, sin ninguna variabilidad entre las ejecuciones. Esto es generalmente una buena señal de que el modelo no está siendo influenciado por el azar o la variabilidad inherente a las particiones del conjunto de datos.

### **Estabilidad Máxima:**

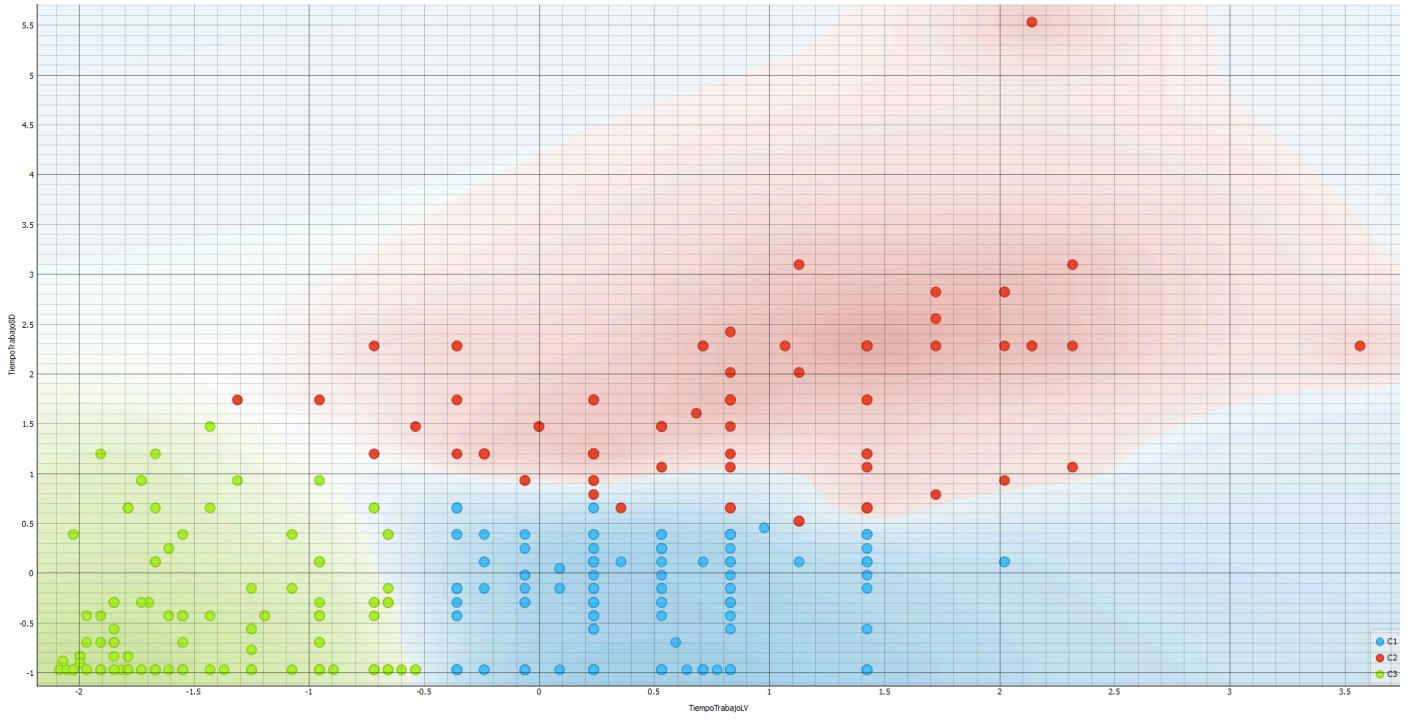
La falta de variabilidad sugiere que el modelo se comporta de manera muy consistente, lo que podría ser esperado si el conjunto de datos no contiene ruido significativo o si está completamente equilibrado en términos de clases. Esta estabilidad también sugiere que el modelo es robusto para este conjunto específico de datos.

## K-Means

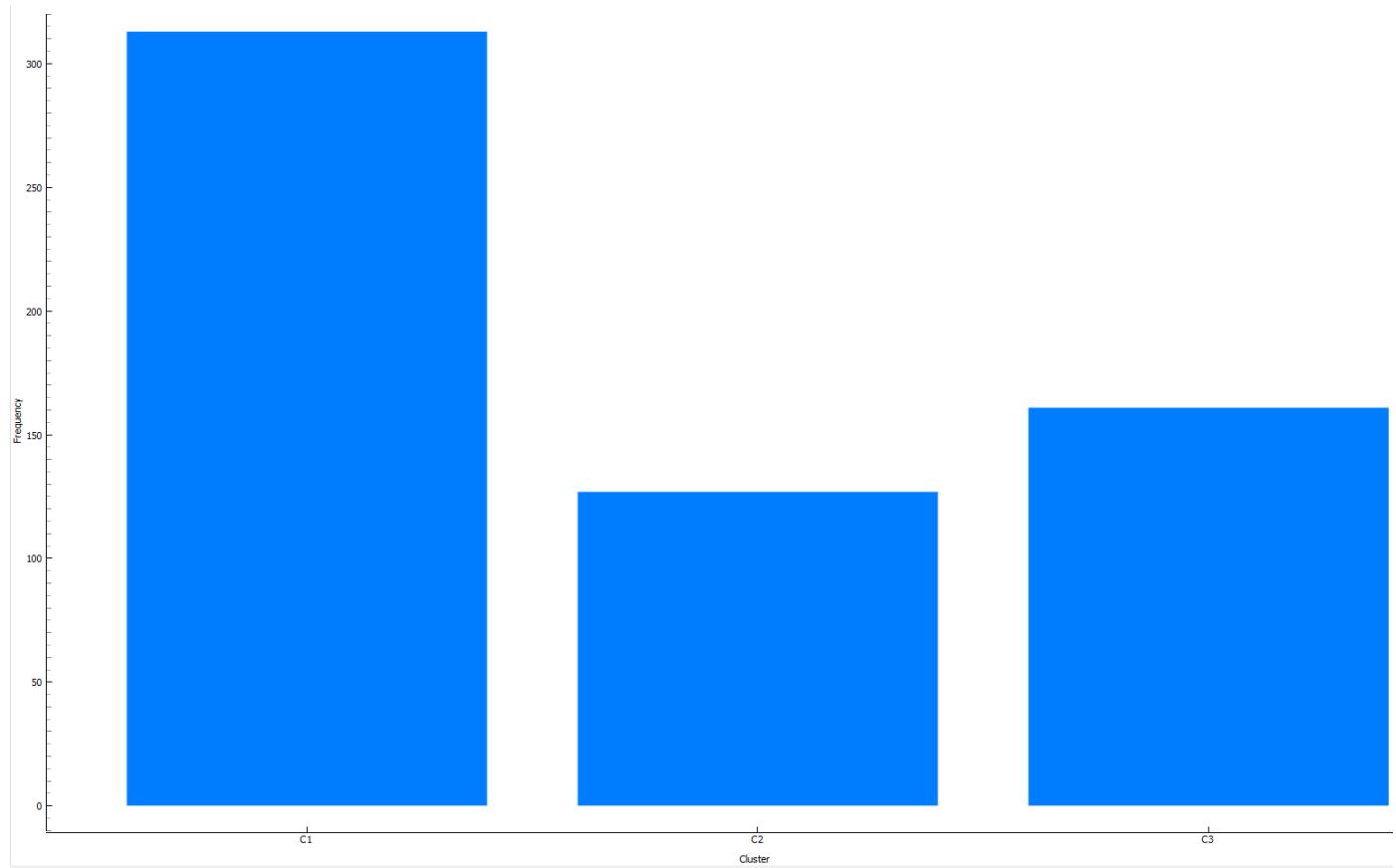
### • Horas Laboradas



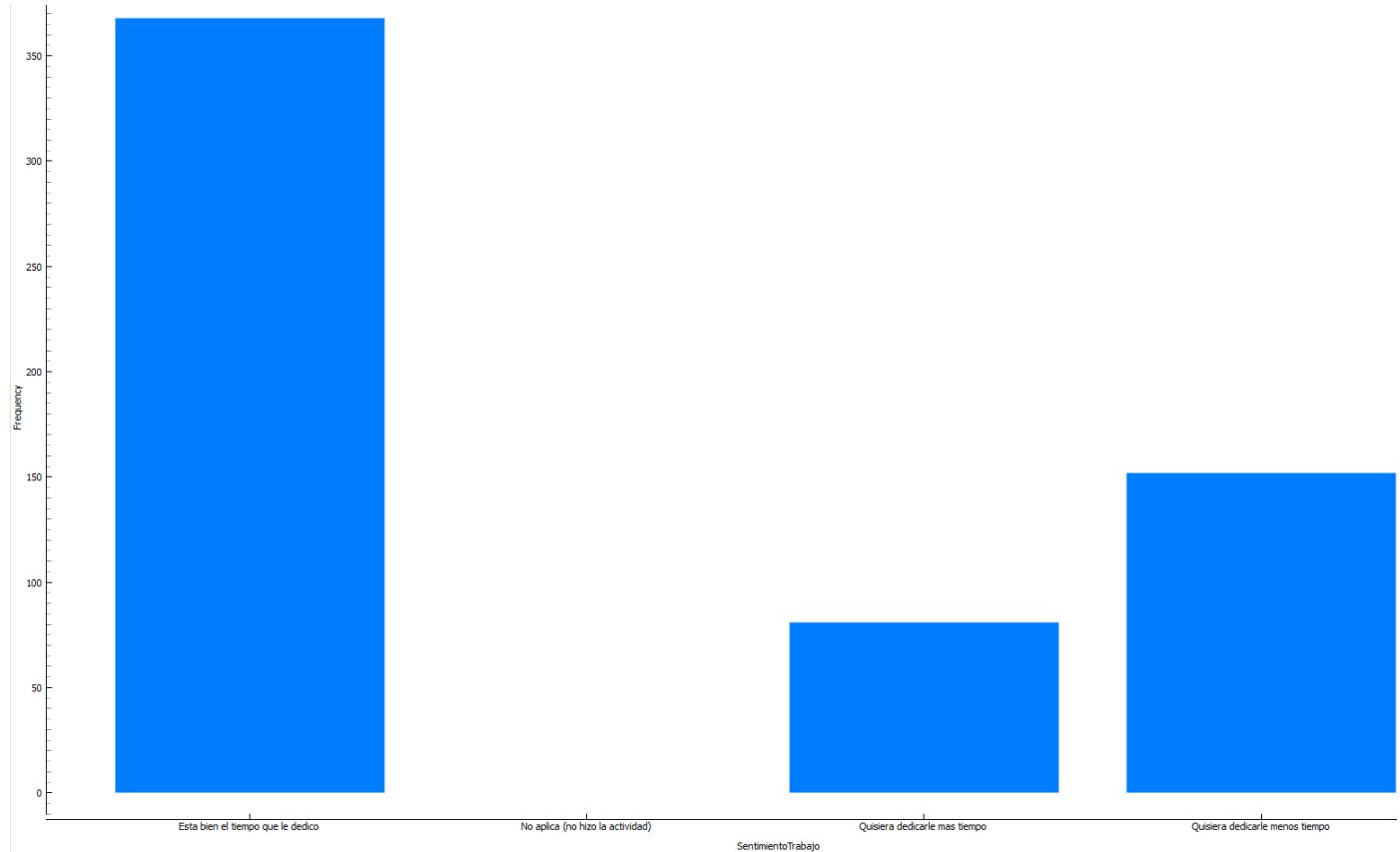
Relación entre el tiempo dedicado al trabajo entre semana (eje X) y los fines de semana (eje Y), agrupado por la percepción del tiempo invertido en estas actividades. Los colores indican la percepción: azul (conforme con el tiempo dedicado), rojo (no aplica), naranja (quisiera dedicarle más tiempo), y verde (quisiera dedicarle menos tiempo). Los puntos representan individuos agrupados por estas categorías.



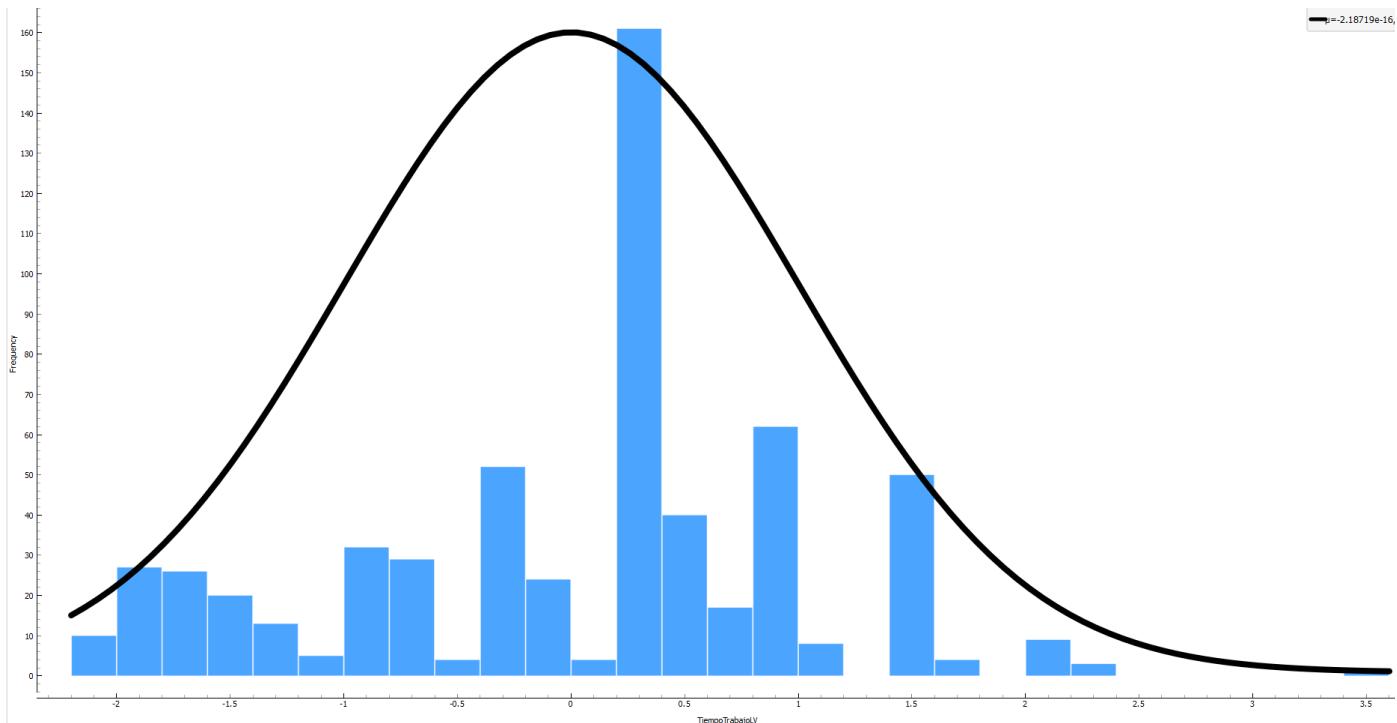
Relación entre el tiempo dedicado al trabajo entre semana (eje X) y los fines de semana (eje Y), agrupado por la percepción del tiempo invertido en estas actividades. Los colores indican la percepción: azul (conforme con el tiempo dedicado), rojo (no aplica), naranja (quisiera dedicarle más tiempo), y verde (quisiera dedicarle menos tiempo). Los puntos representan individuos agrupados por estas categorías.



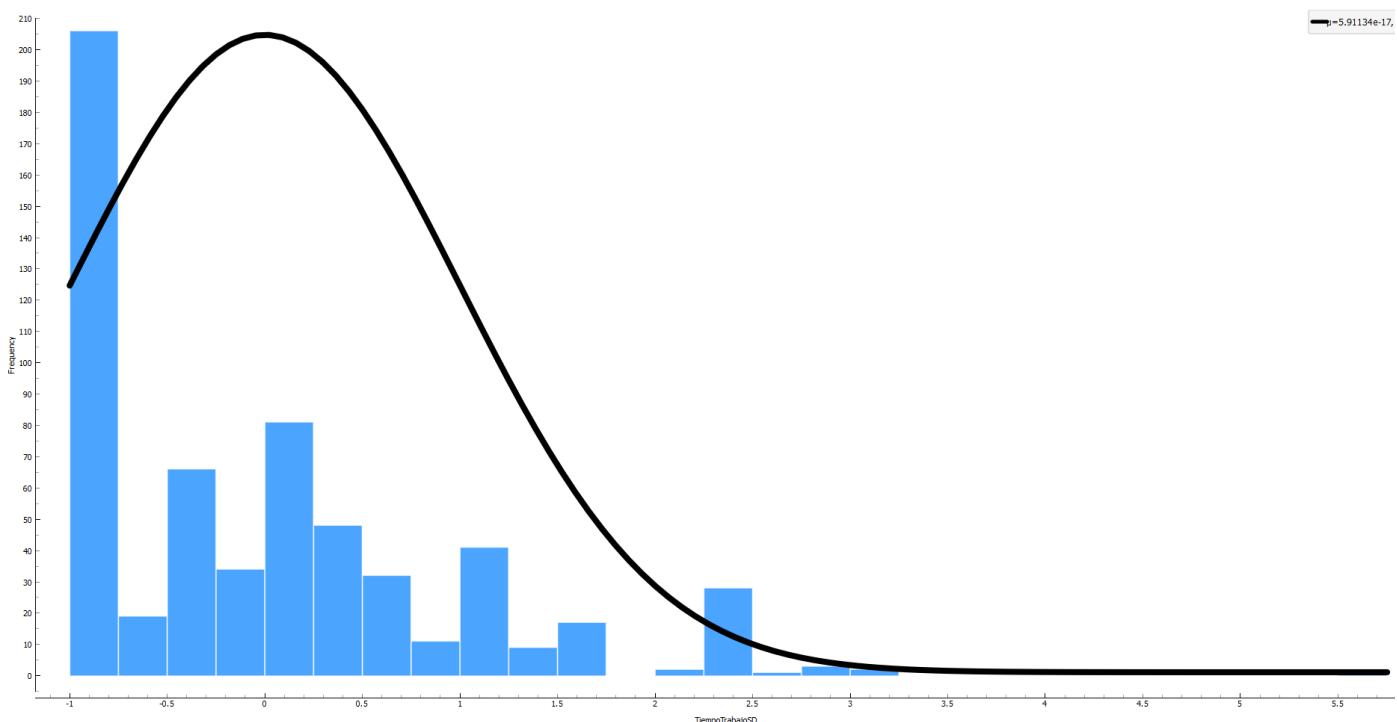
Distribución de la cantidad de personas en cada cluster generado por el modelo K-Means. Los clusters (C1, C2, C3) representan grupos de personas con características similares en función de las variables analizadas. El eje X indica los clusters, mientras que el eje Y muestra la frecuencia (número de personas) en cada uno.



Distribución de los sentimientos respecto al tiempo dedicado al trabajo. El eje X representa las diferentes percepciones de las personas sobre su tiempo laboral: 'Está bien el tiempo que le dedico', 'No aplica (no hizo la actividad)', 'Quisiera dedicarle más tiempo', y 'Quisiera dedicarle menos tiempo'. El eje Y muestra la frecuencia (número de personas) en cada categoría.

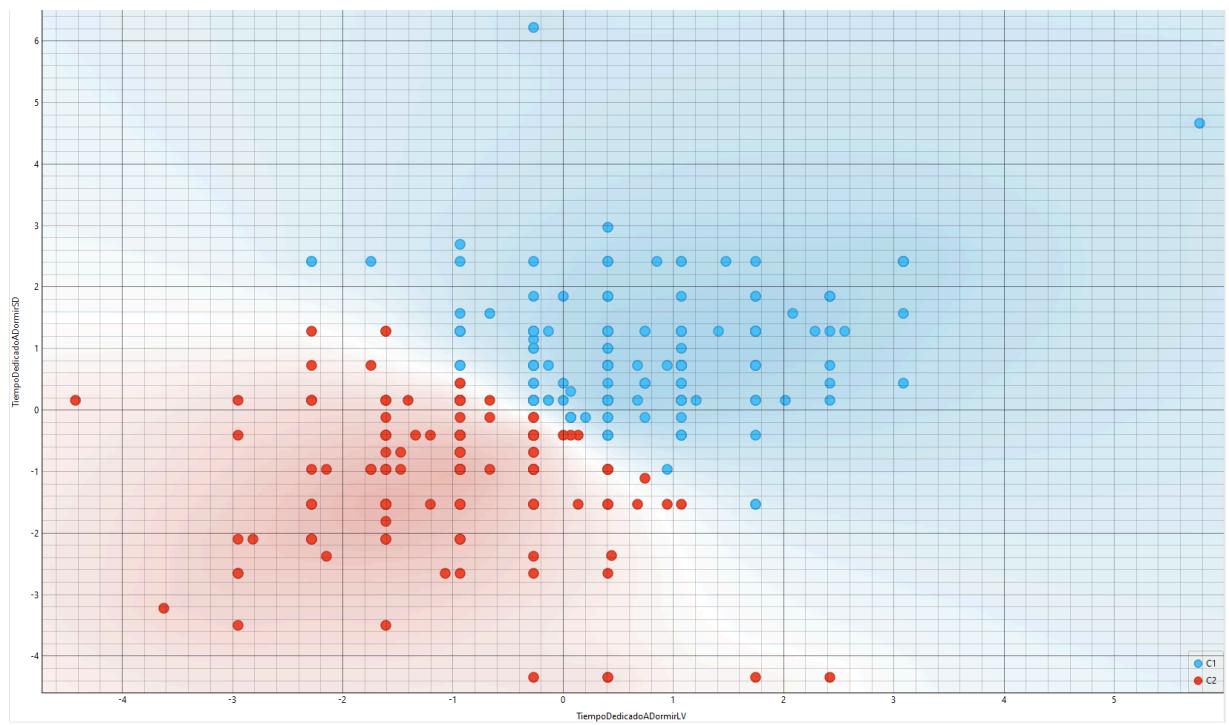


Distribución del tiempo dedicado al trabajo entre semana (TiempoTrabajoLV). El histograma muestra la frecuencia de personas en diferentes rangos de tiempo, mientras que la curva negra representa la estimación de densidad (ajuste a una distribución normal) para interpretar la tendencia general de los datos.

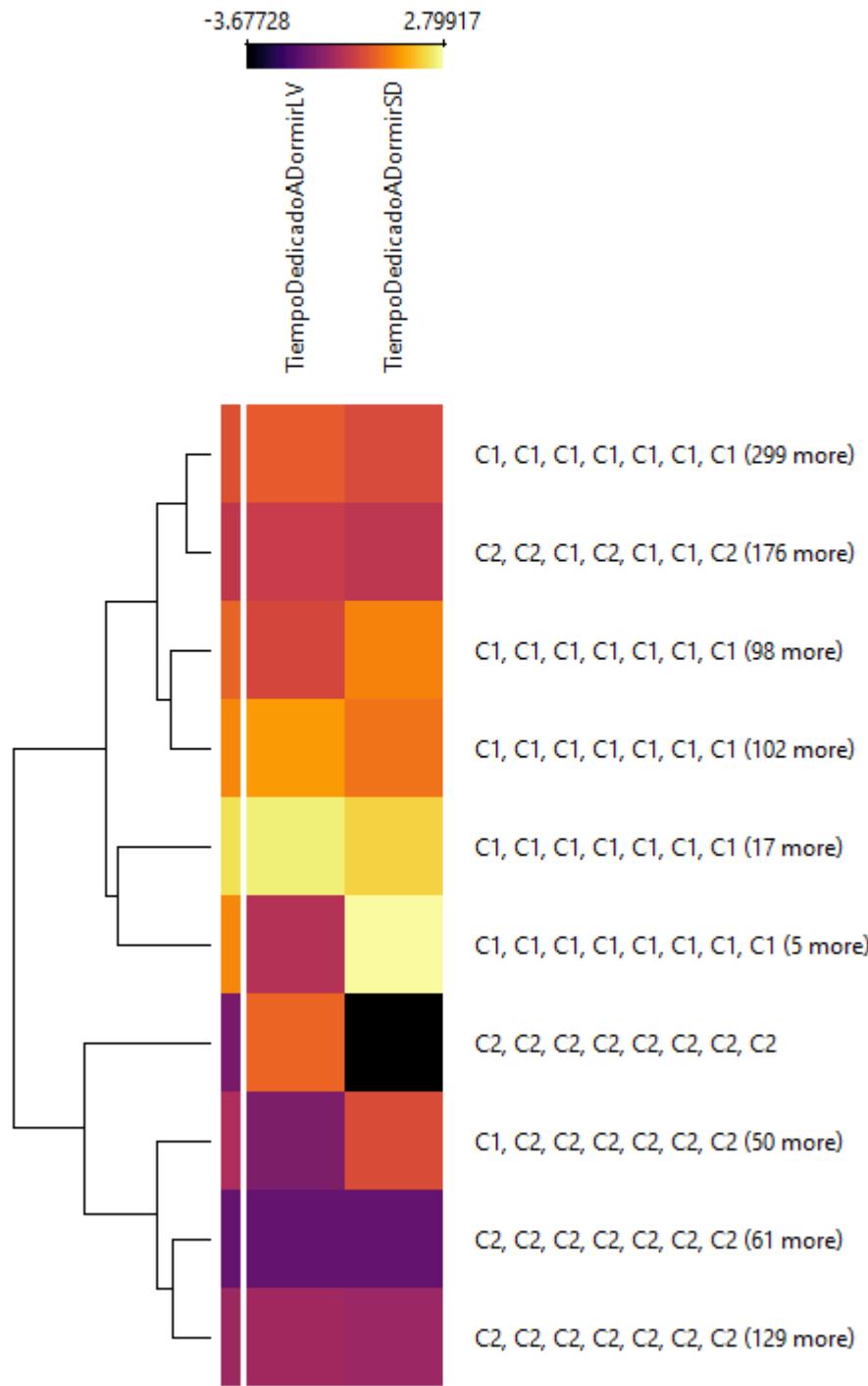


Distribución del tiempo dedicado al trabajo los fines de semana (TiempoTrabajoSD). El histograma muestra la frecuencia de personas en diferentes rangos de tiempo, y la curva negra representa una estimación de densidad (posible ajuste a una distribución normal).

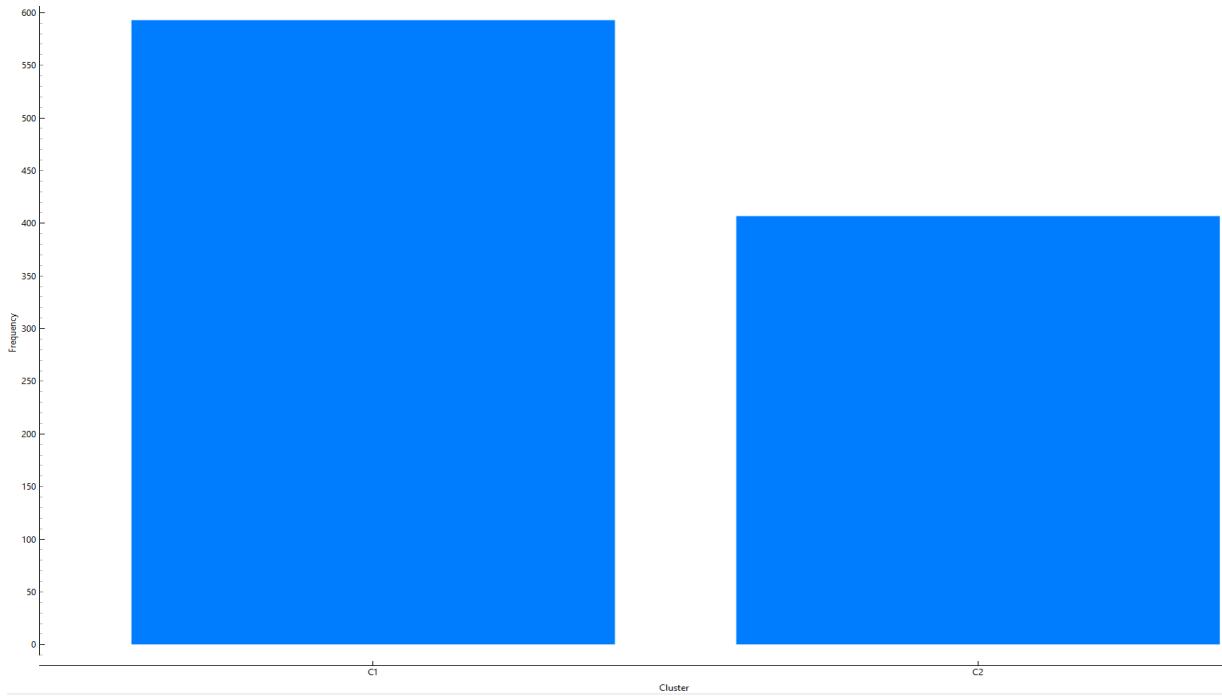
## ● Horas de Descanso



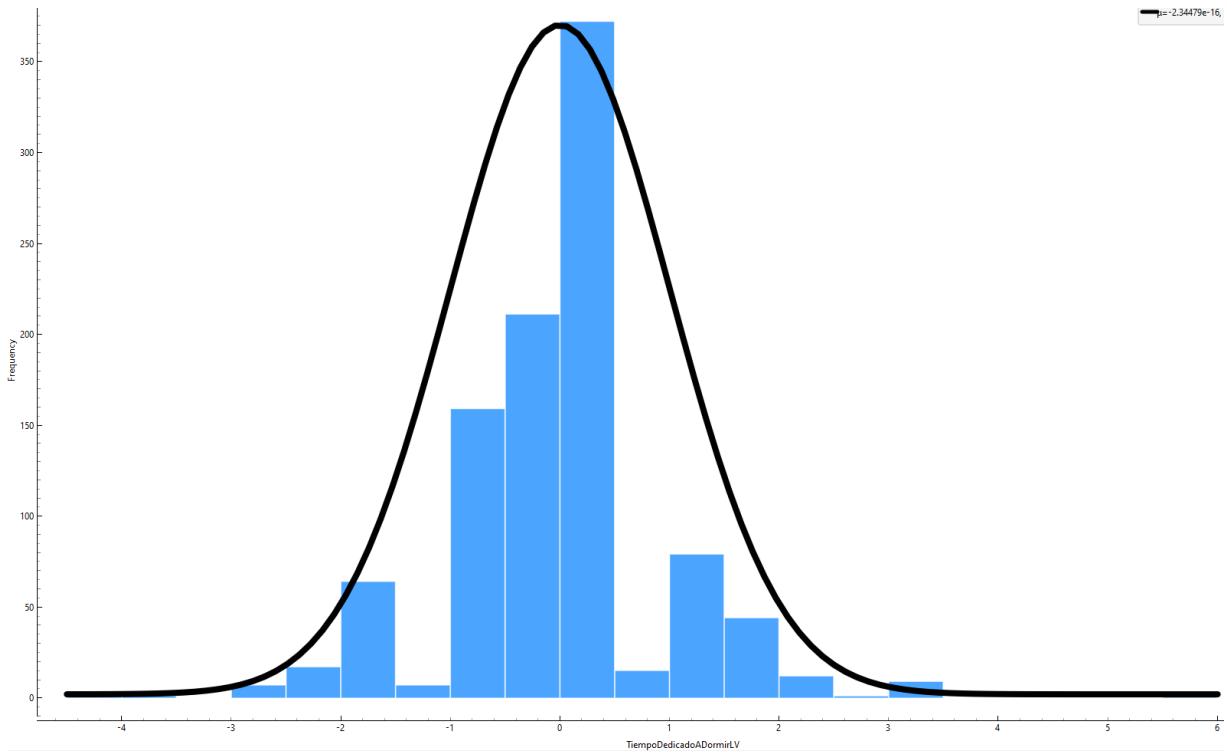
Clasificación de clusters en relación con el tiempo dedicado a dormir entre semana (TiempoDedicadoADormirLV) y fines de semana (TiempoDedicadoADormirSD). El clúster C1 (azul) representa a personas con mayor tiempo dedicado al sueño en ambos períodos, mientras que el clúster C2 (rojo) agrupa a aquellos con menos horas de descanso.



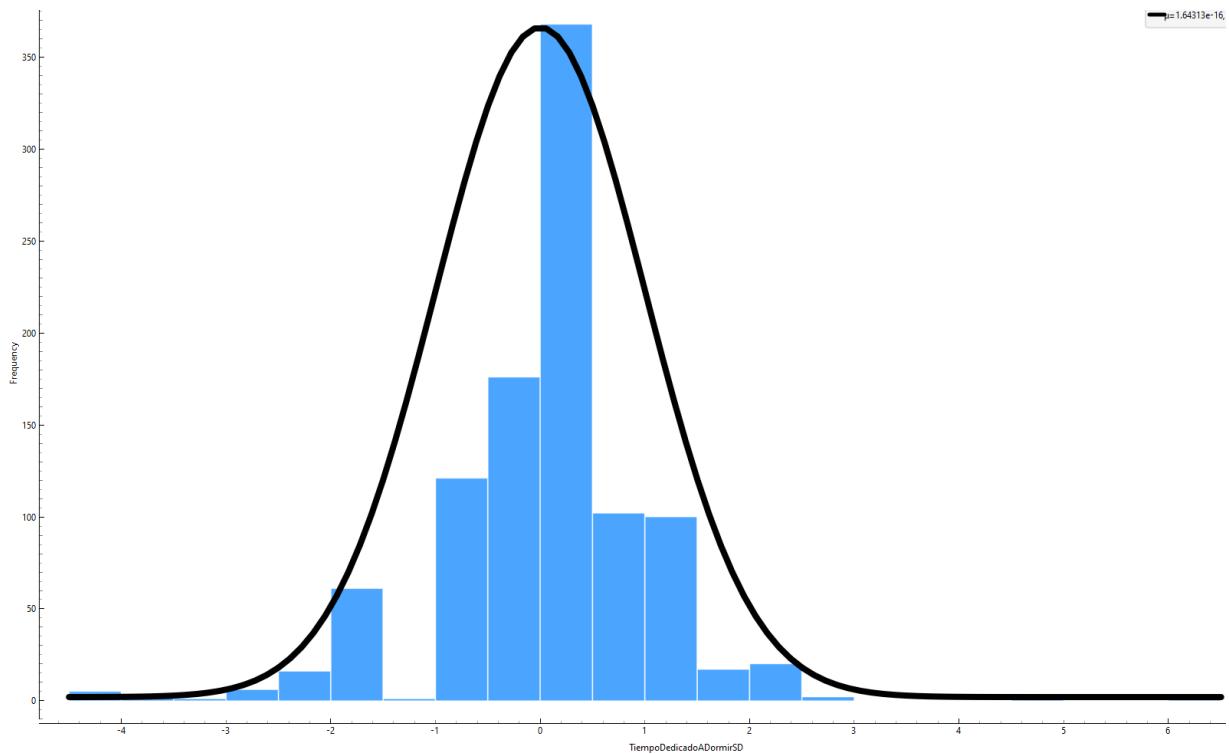
Mapa de calor de patrones de sueño por clusters: Este gráfico muestra cómo se agrupan las personas en función del tiempo dedicado a dormir entre semana (TiempoDedicadoADormirLV) y los fines de semana (TiempoDedicadoADormirSD), con valores normalizados entre -3.67728 y 2.79917. Los colores más claros (amarillo) indican mayor tiempo dedicado al sueño, mientras que los colores oscuros (púrpura/negro) representan menos horas de sueño. El clúster C1 incluye mayoritariamente personas con tiempos más altos de descanso, mientras que el clúster C2 concentra aquellos con tiempos significativamente menores. El dendrograma sugiere relaciones jerárquicas entre los subgrupos dentro de los clusters.



Frecuencia de instancias por clúster: El clúster C1 tiene un mayor número de instancias (alrededor de 600), representando a la mayoría de las personas con características similares en los patrones de sueño. Por otro lado, el clúster C2 agrupa a un número menor de instancias (aproximadamente 400), indicando un grupo más reducido pero distintivo en sus hábitos de descanso.

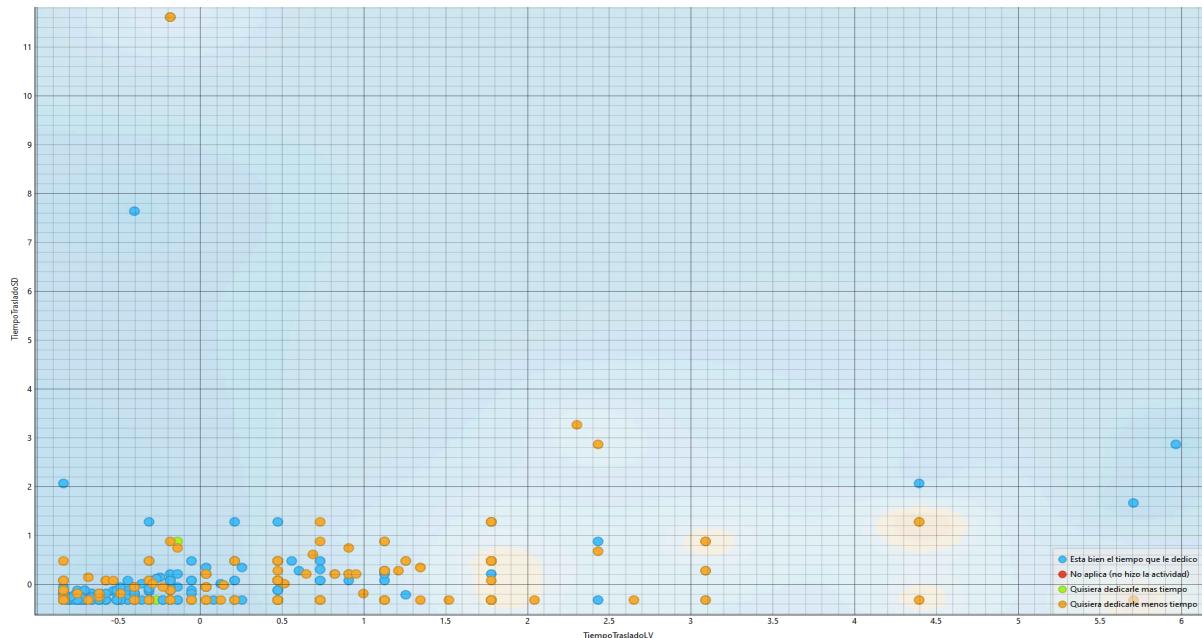


Distribución del tiempo dedicado a dormir entre semana (TiempoDedicadoADormirLV): El histograma muestra una distribución normalizada del tiempo de sueño entre semana, donde la mayor parte de las personas dedican un tiempo promedio cercano a 0 (en valores normalizados). La curva negra representa la aproximación a una distribución normal, indicando que la mayoría de los datos se concentran alrededor del promedio, con menos personas en los extremos, ya sea con tiempos muy altos o muy bajos.

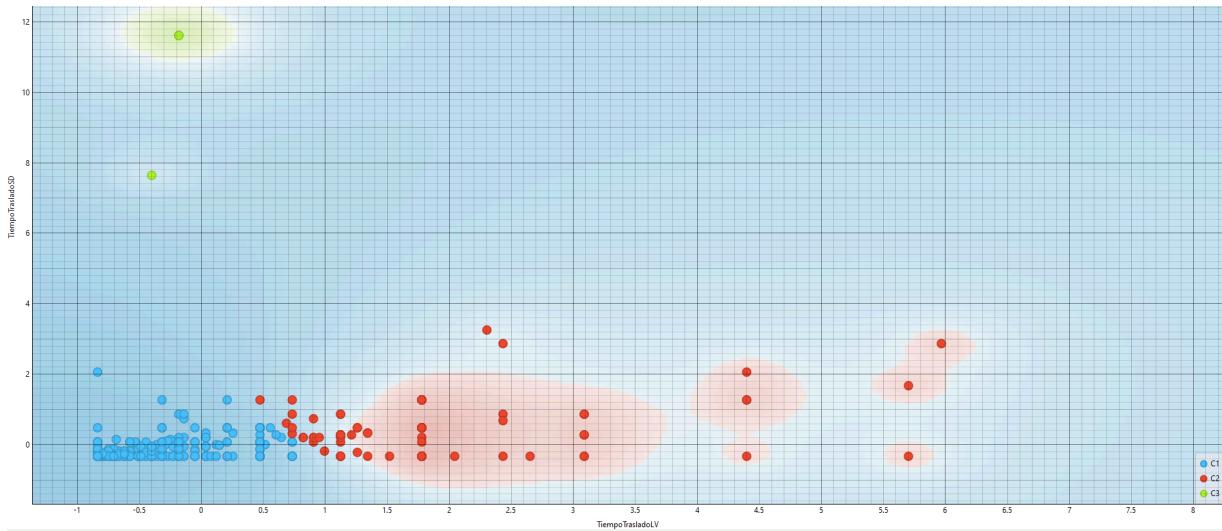


Distribución del tiempo dedicado a dormir los fines de semana (TiempoDedicadoADormirSD): Este histograma muestra la frecuencia de valores normalizados del tiempo de sueño durante los fines de semana. La mayoría de los datos se concentran alrededor de 0 (en la escala normalizada), representando un promedio de sueño consistente. La curva negra de distribución normal confirma la simetría en la concentración de datos, mientras que los valores extremos a ambos lados reflejan hábitos menos comunes de sueño, ya sea mucho tiempo o poco tiempo dedicado a dormir.

### ● Horas de Traslado



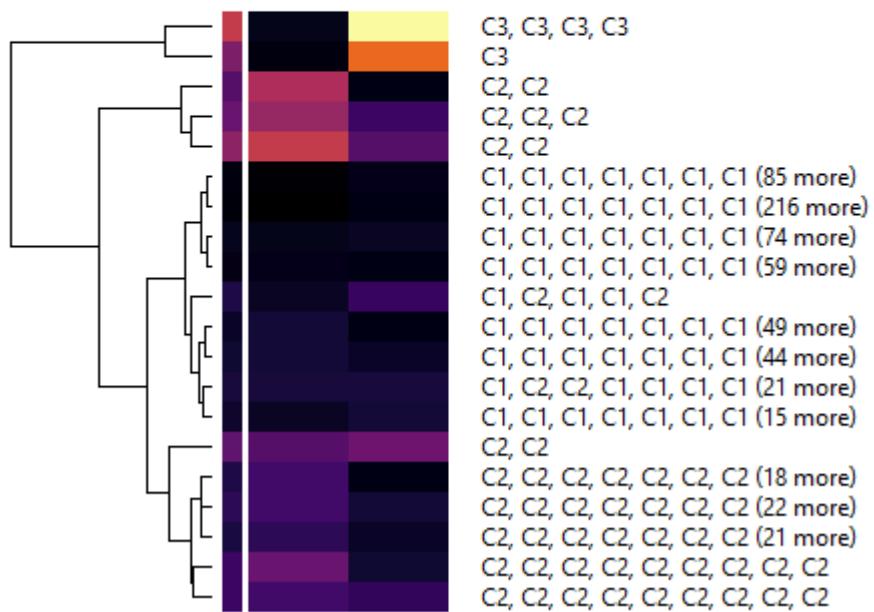
*Distribución del tiempo de traslado entre semana y fines de semana según percepción personal. Este gráfico muestra cómo los participantes perciben el tiempo que dedican a traslados. Los puntos en azul representan a quienes consideran adecuado el tiempo invertido, mientras que los puntos amarillos reflejan a quienes desean dedicarle menos tiempo.*



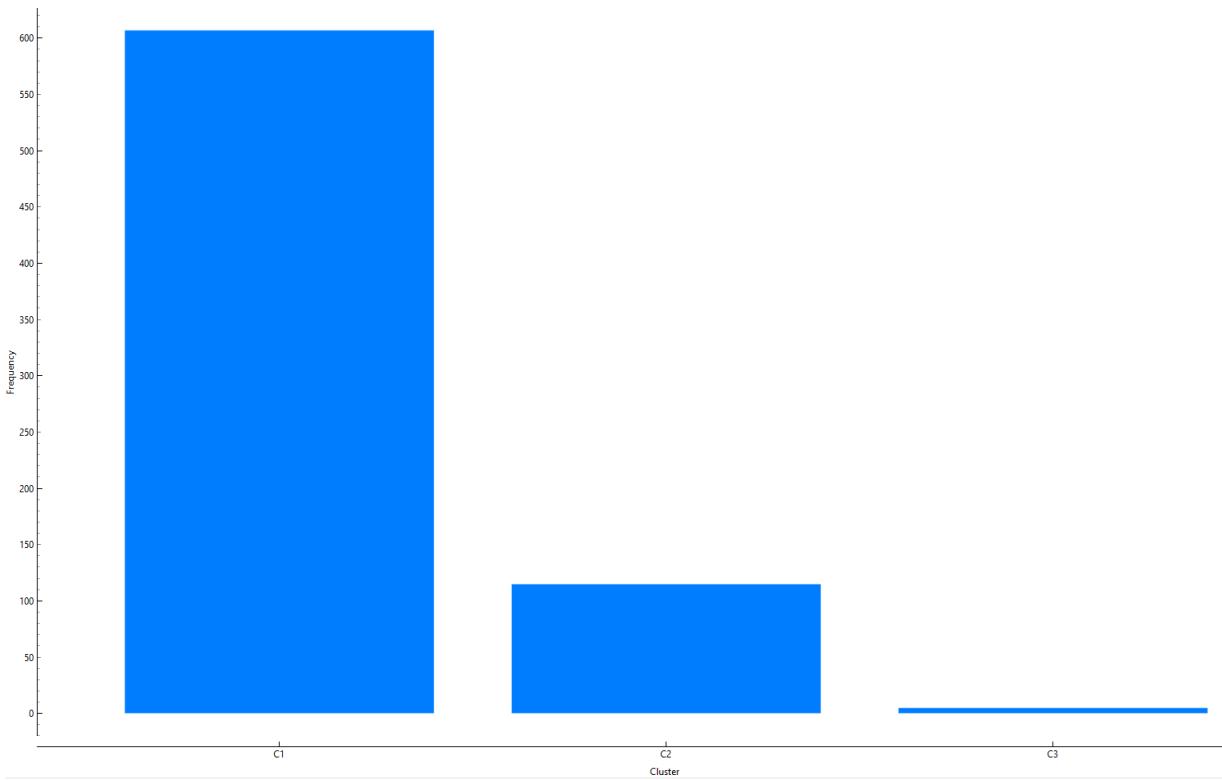
Distribución de los cluster según el tiempo de traslado entre semana y su desviación estándar. Cluster 1(Azul), Cluster 2(Rojo), Cluster 3(Verde)



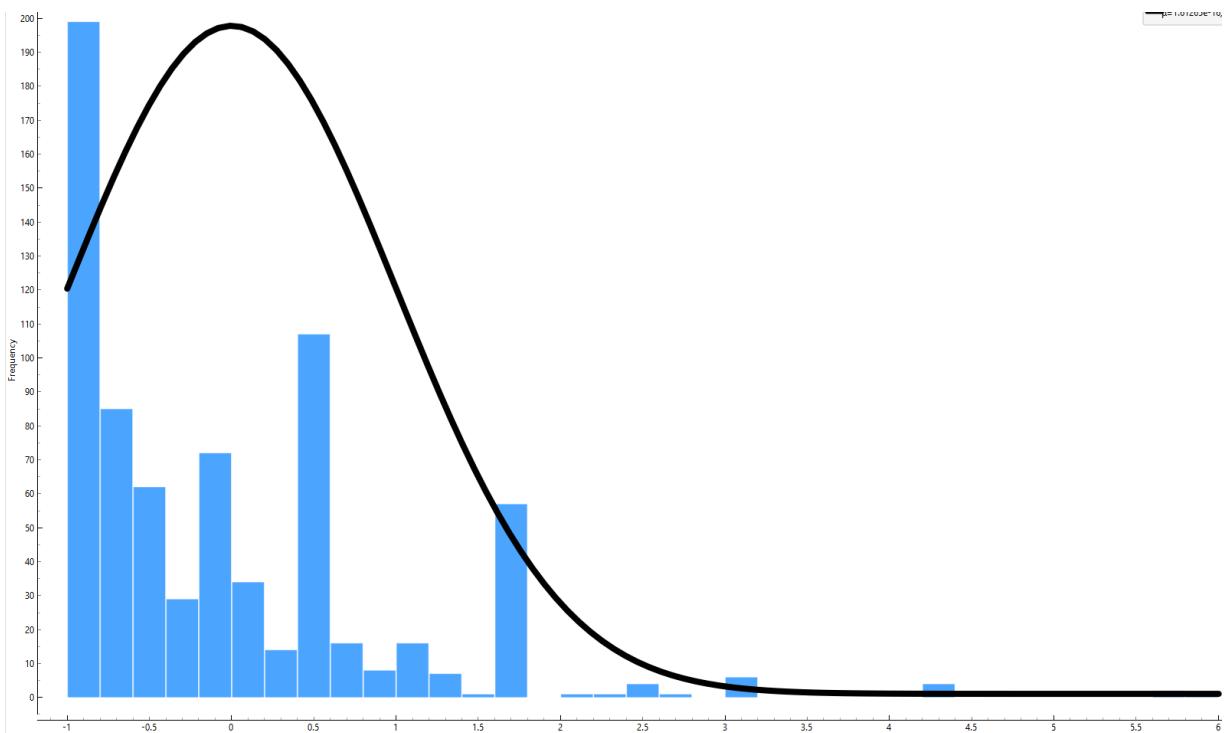
TiempoTrasladoLV  
TiempoTrasladoSD



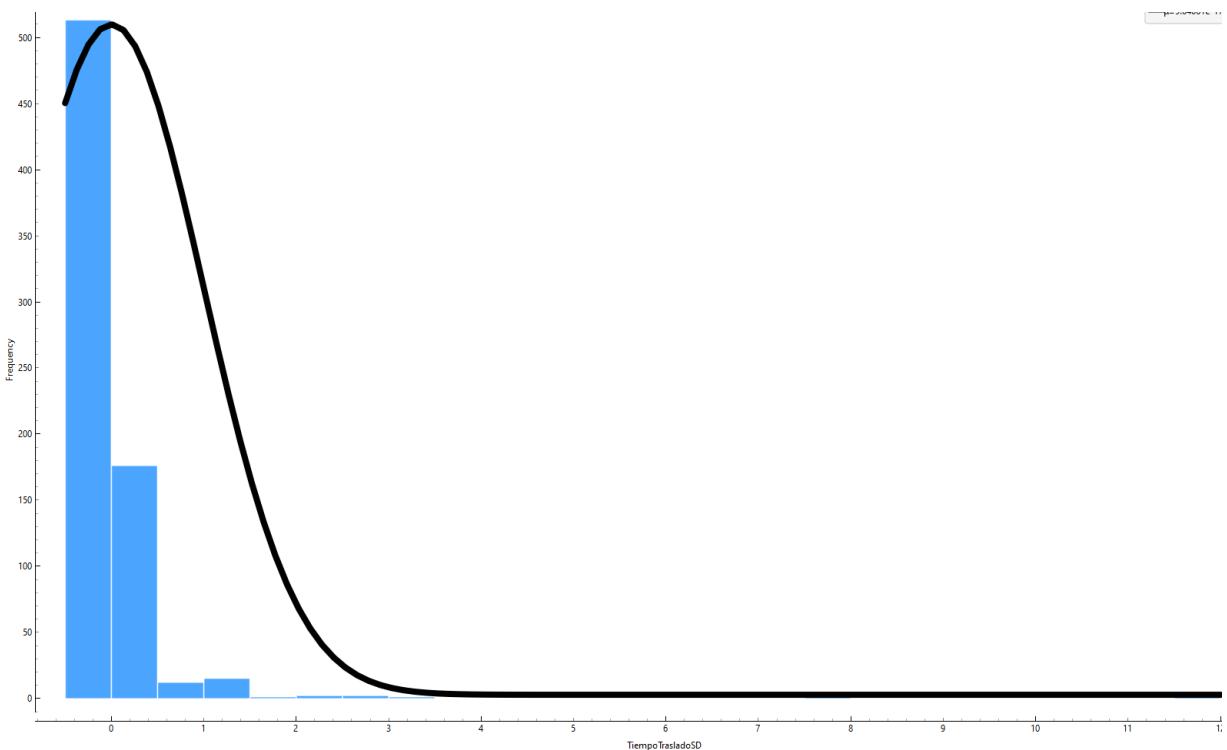
Mapa de calor de los clusters basado en tiempo de traslado y su desviación estándar. Cluster C1 (color predominante en la parte inferior): Agrupa a individuos con tiempos de traslado consistentemente bajos y baja variabilidad. Cluster C2 (color intermedio): Representa a quienes tienen tiempos de traslado moderados con mayor variabilidad. Cluster C3 (color claro en la parte superior): Incluye a personas con los tiempos de traslado más altos y una desviación estándar significativa, lo que indica una gran dispersión en los patrones de traslado.



*Distribución de las instancias entre los tres clusters generados. C1 contiene la mayoría de las instancias.*

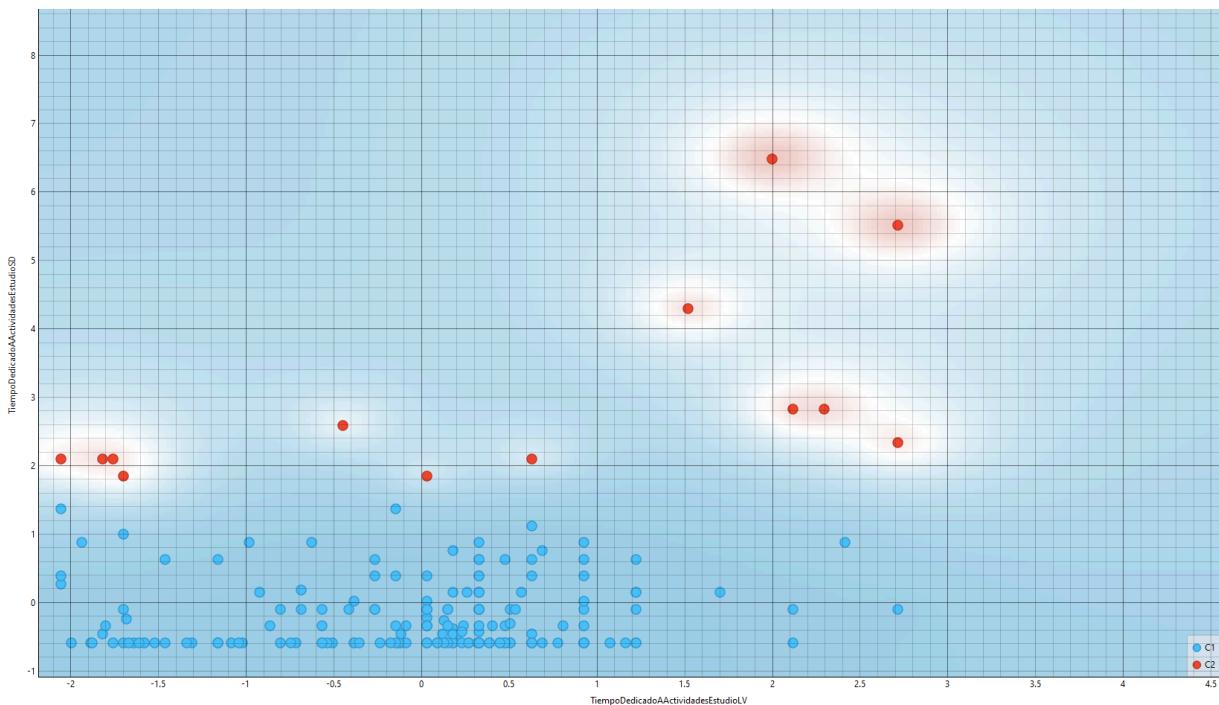


*Distribución del tiempo de traslado entre semana, la mayoría de las instancias se concentran cerca de valores más bajos.*

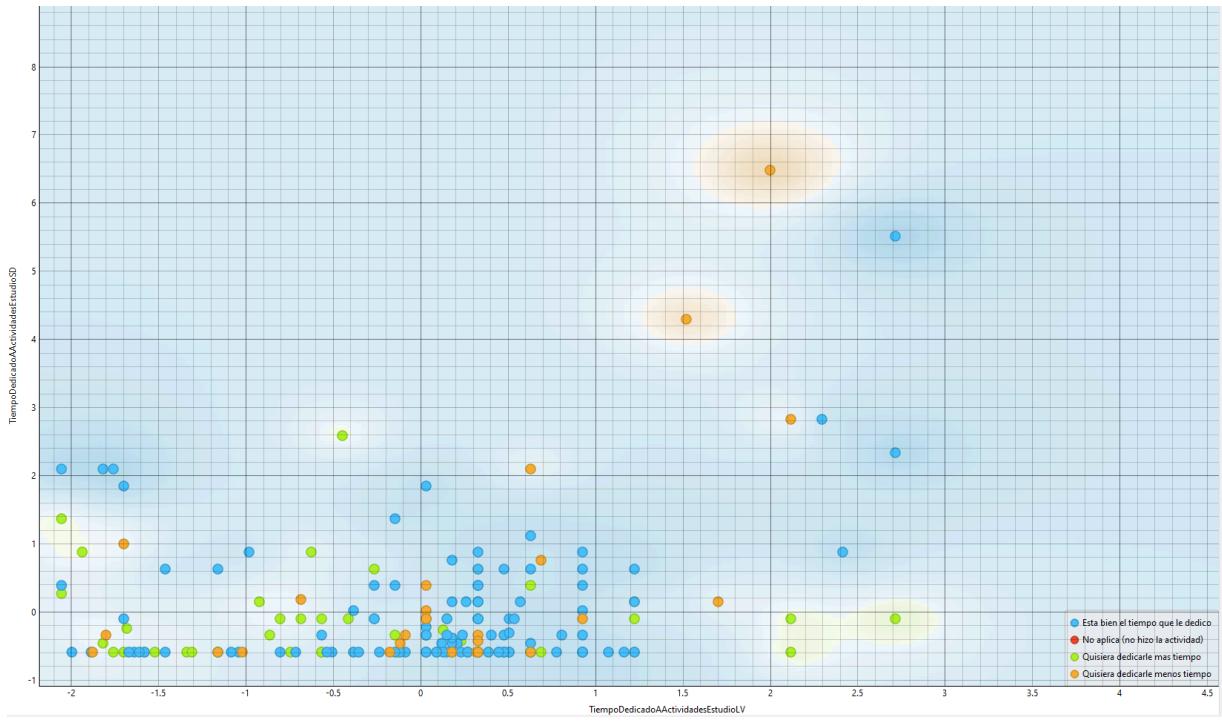


*Distribución del tiempo de traslado sábados y domingos, la mayor parte de las personas tiene poca variabilidad en su tiempo de traslados.*

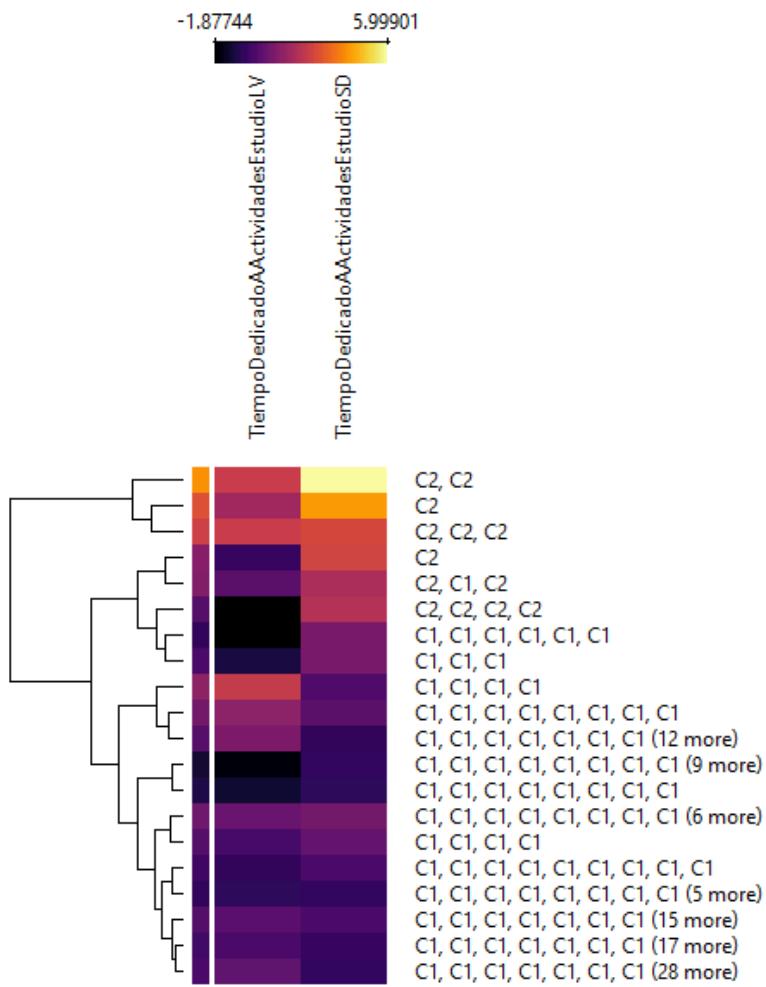
### ● Horas de Estudio



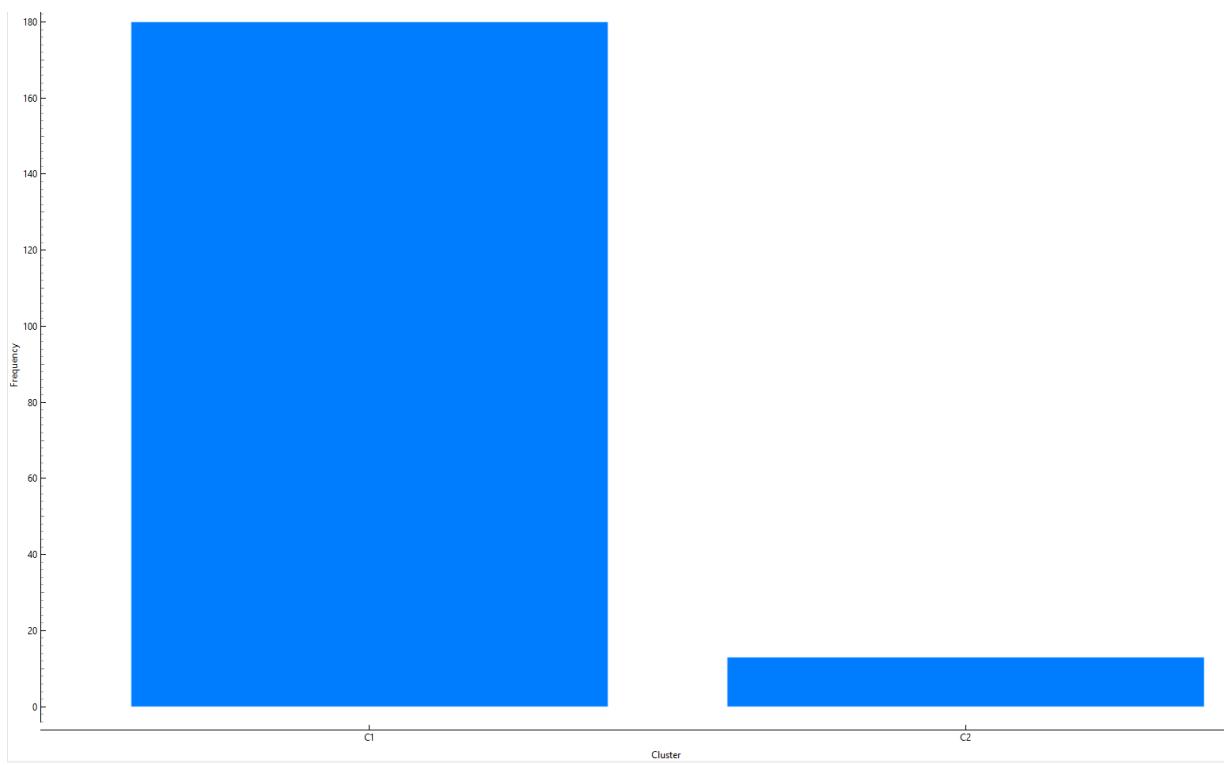
*Distribución de los clústeres en relación con el tiempo dedicado a actividades de estudio*



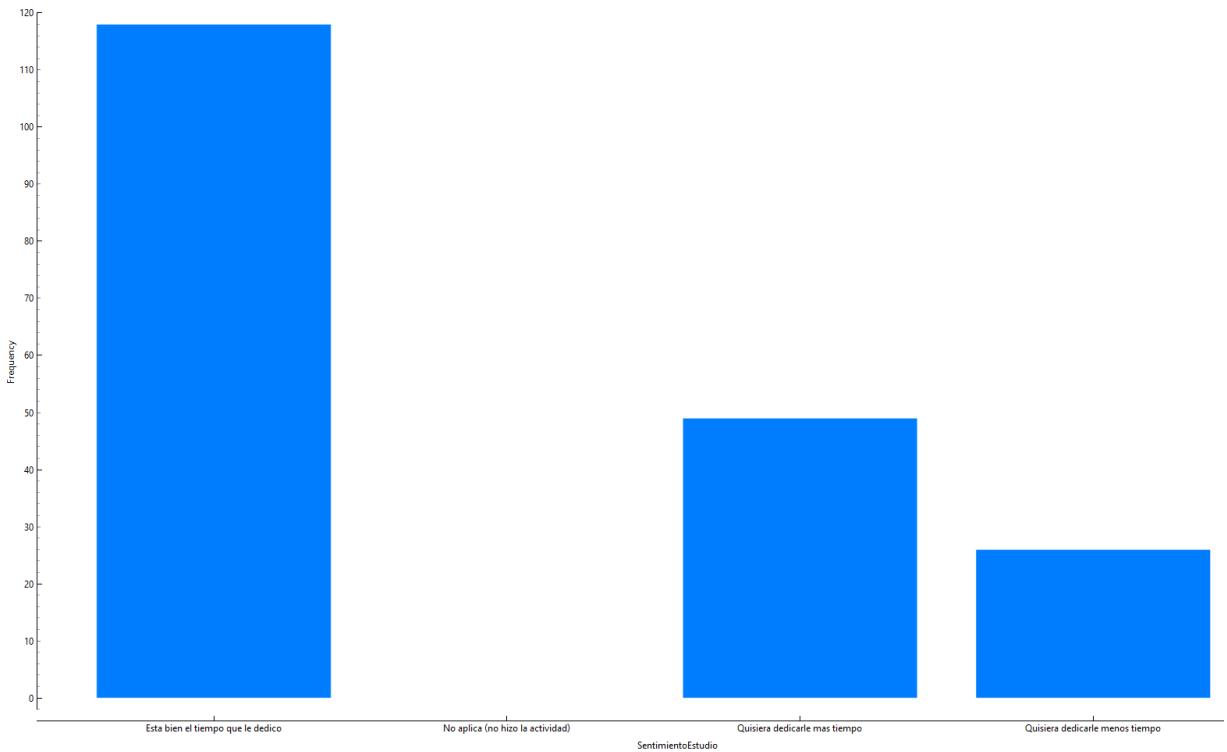
*Distribución del tiempo dedicado a actividades de estudio y el sentimiento de las personas.*



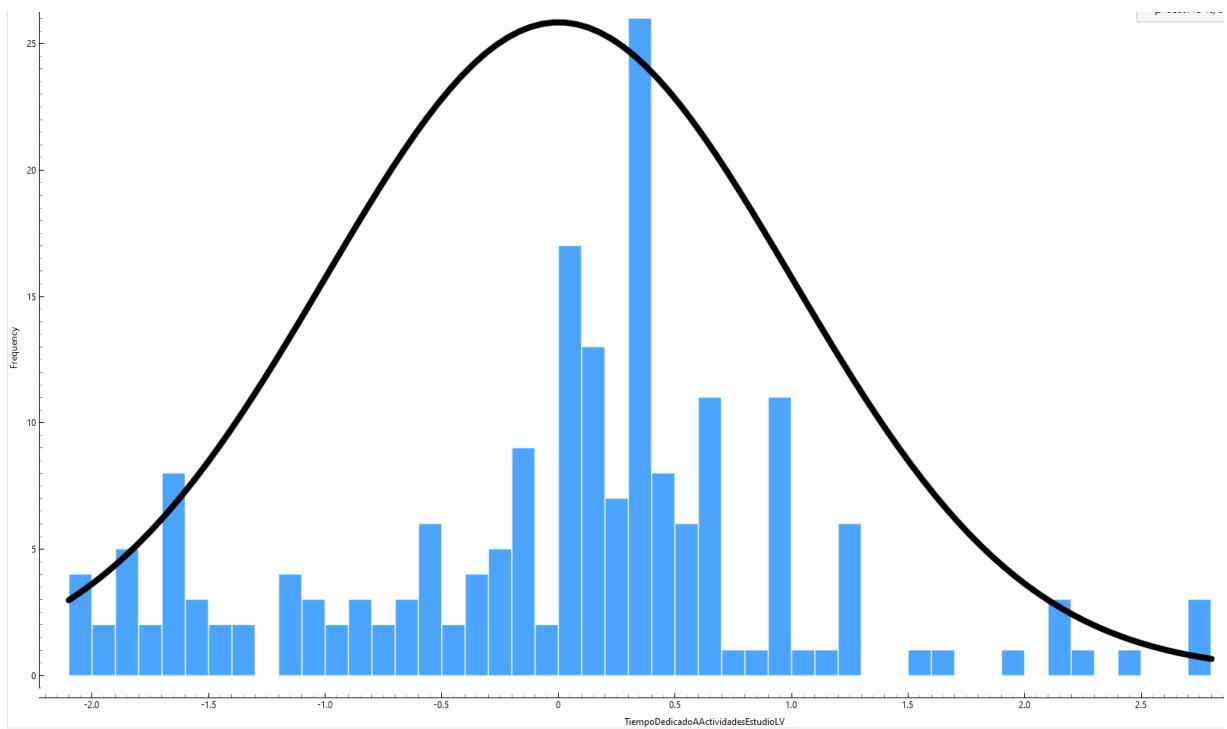
*Mapa de Calor de Actividades de Estudio*



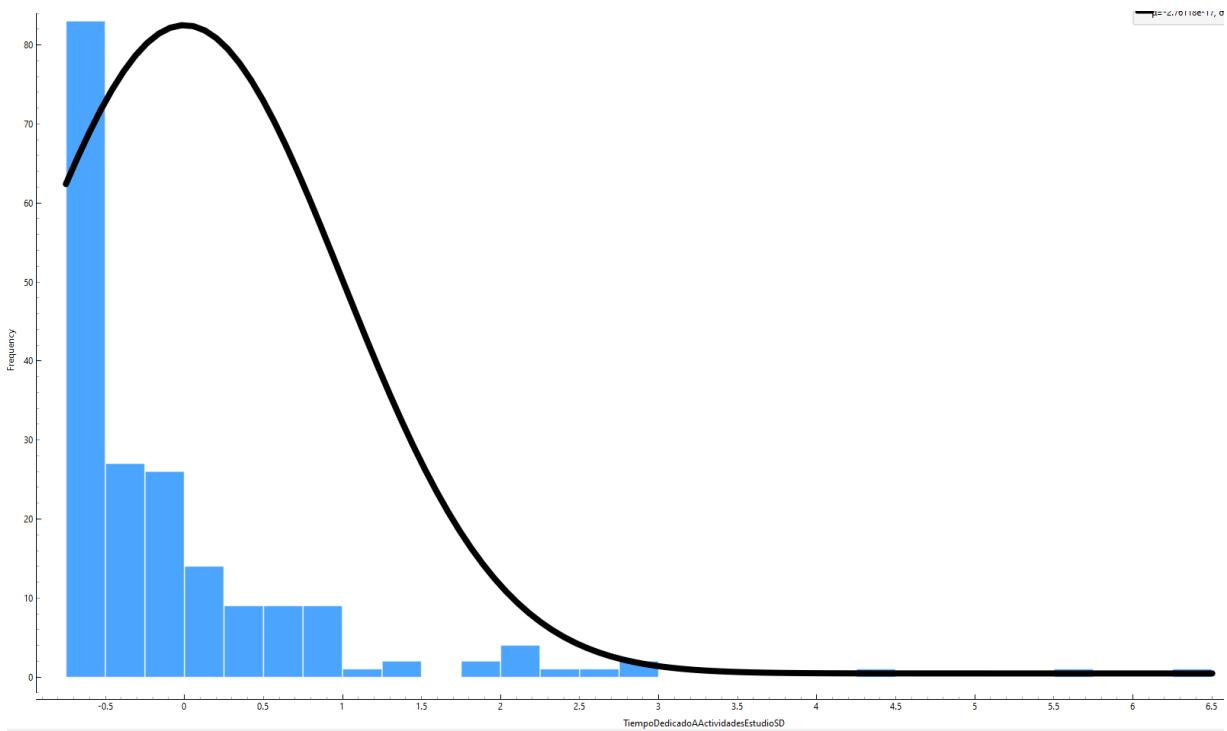
*Distribución de Clústeres en Actividades de Estudio*



*Distribución del Sentimiento Respecto al Tiempo Dedicado a las Actividades de Estudio*

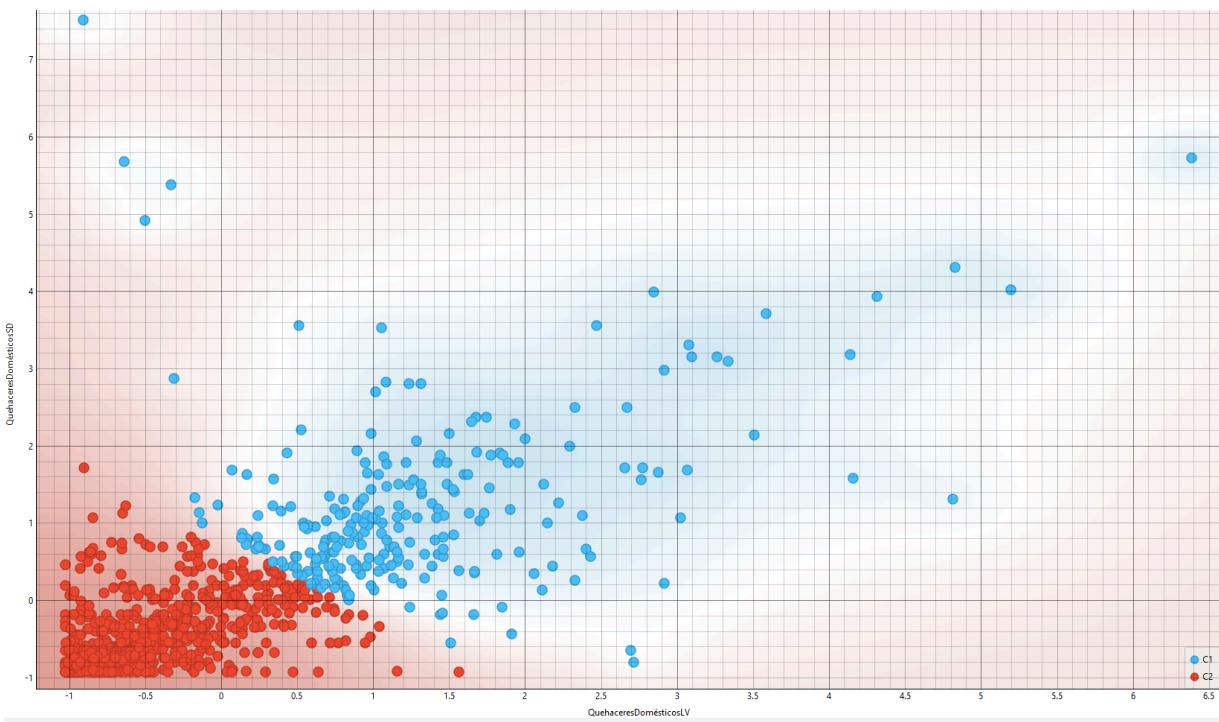


Distribución del Tiempo Dedicado a Actividades de Estudio de Lunes a Viernes



Distribución del Tiempo Dedicado a Actividades de Estudio Durante los Fines de Semana

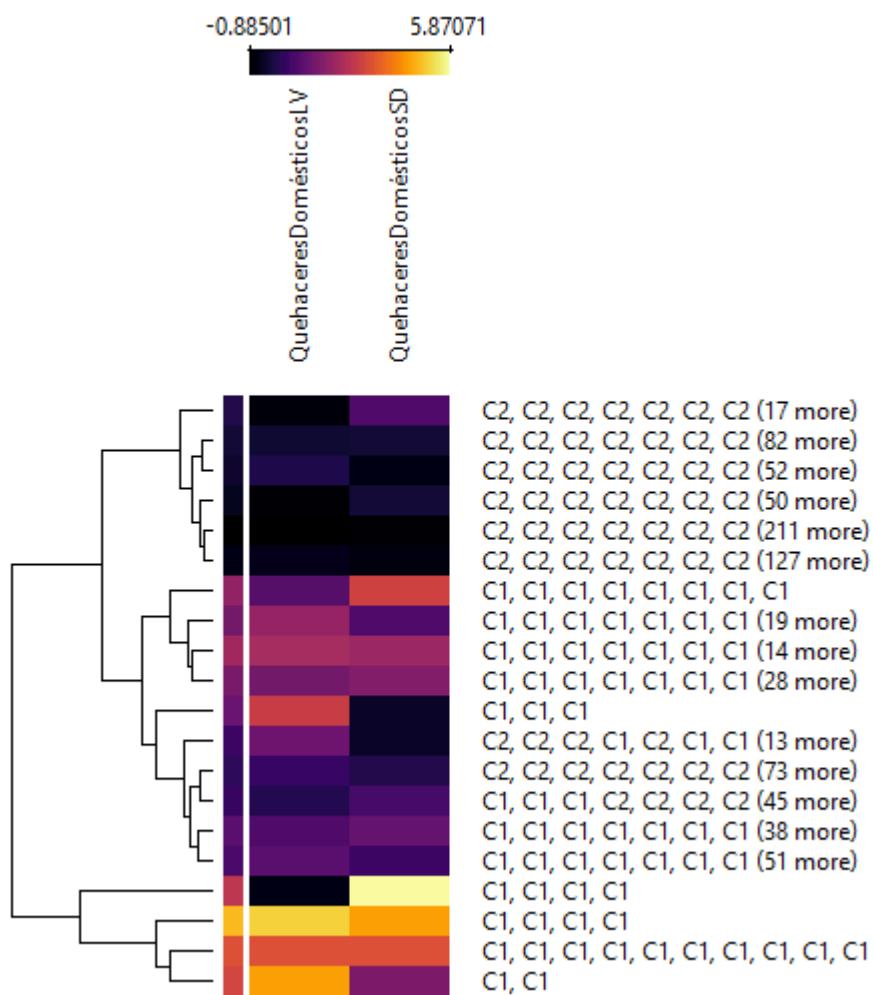
## ● Horas de Actividades Domésticas.



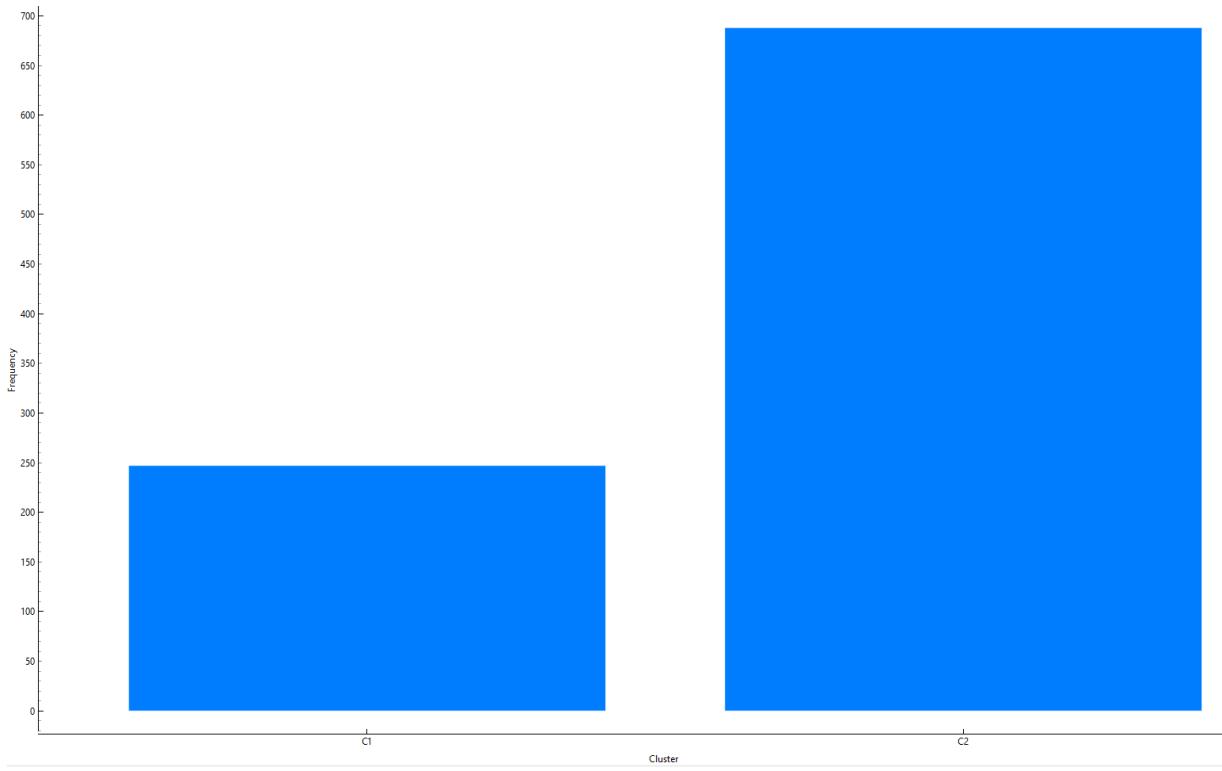
Distribución de los clústeres en relación con el tiempo dedicado a actividades domésticas



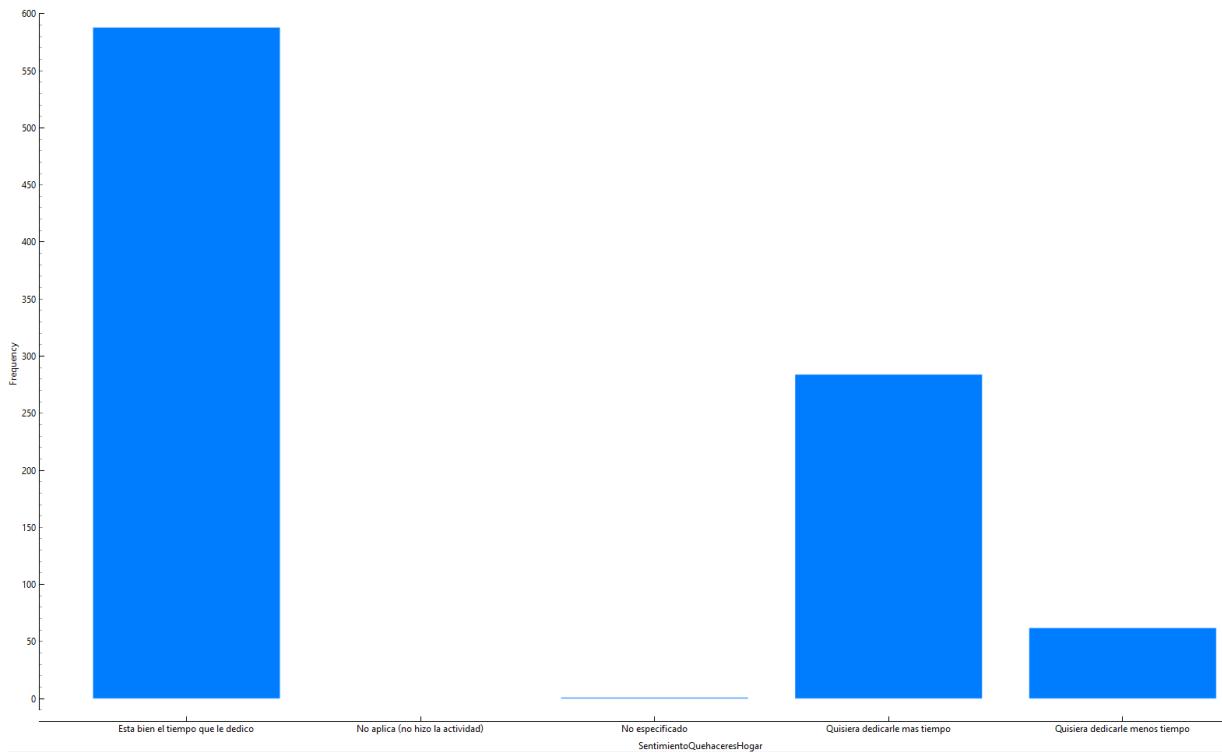
Distribución del tiempo dedicado a actividades domésticas por categoría de sentimiento



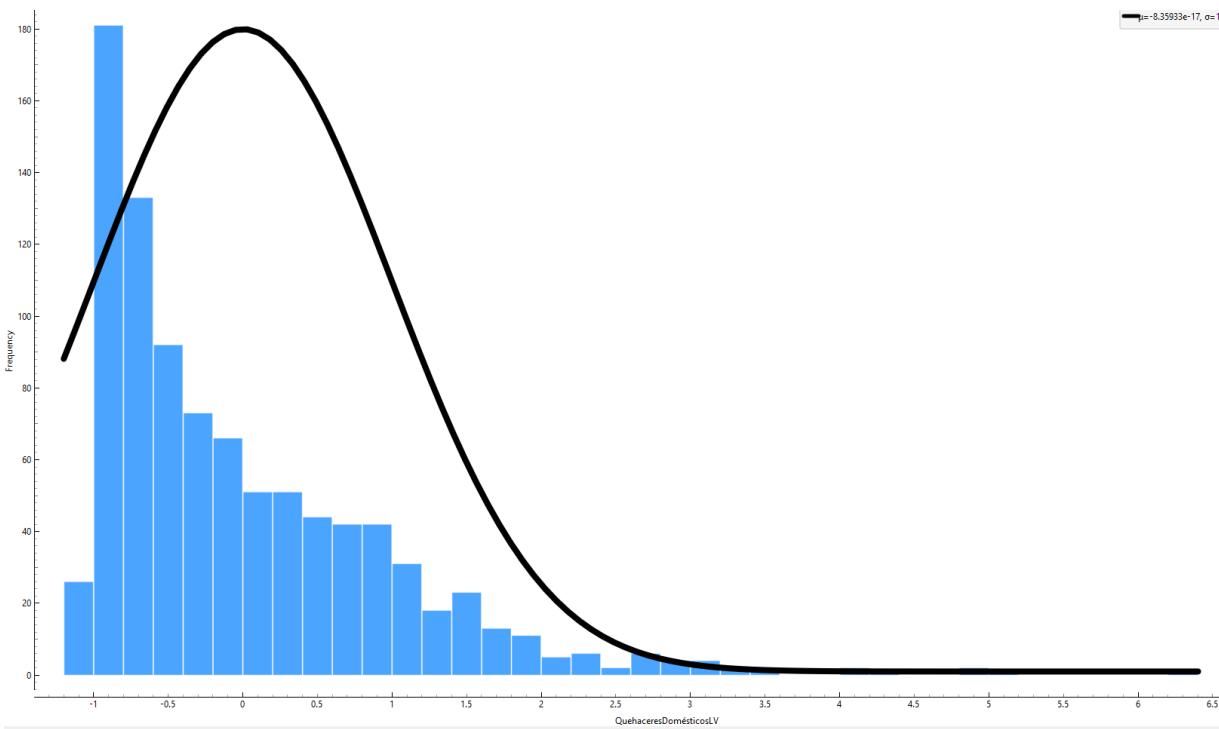
mapa de calor de actividades domésticas



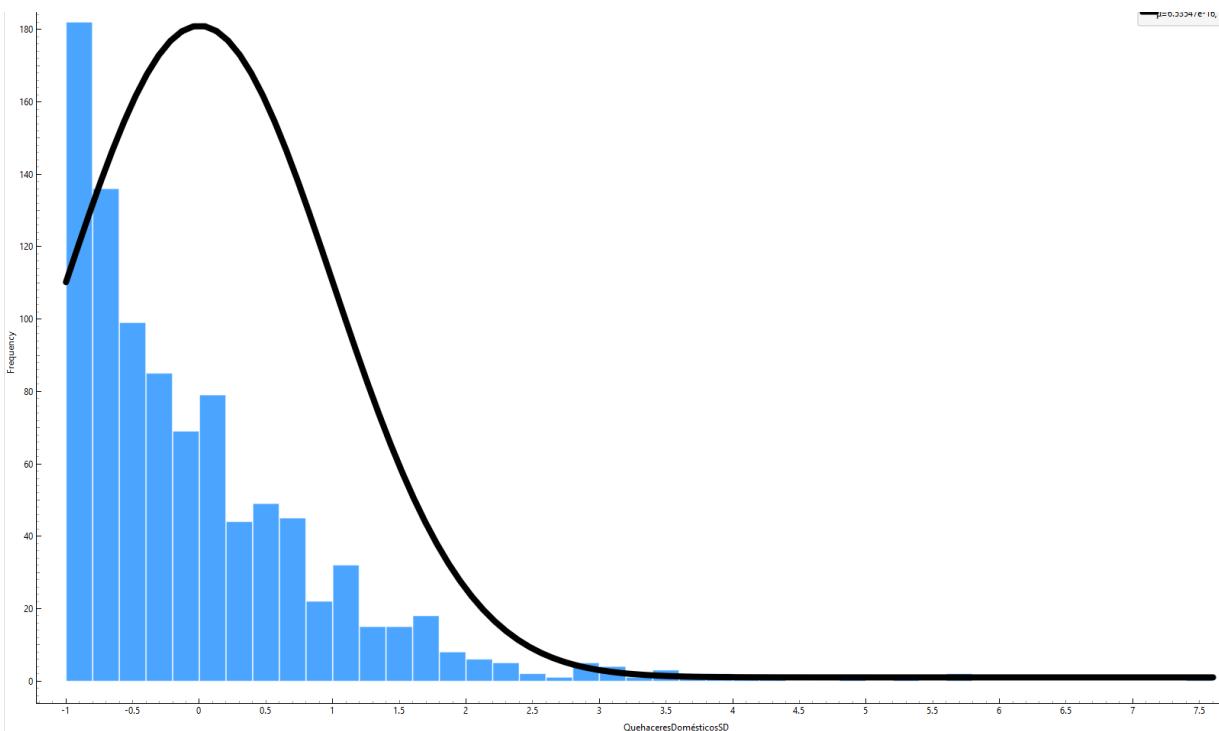
*Distribución de Clústeres en actividades domésticas*



*Distribución del Sentimiento Respecto al Tiempo Dedicado a las actividades domésticas*

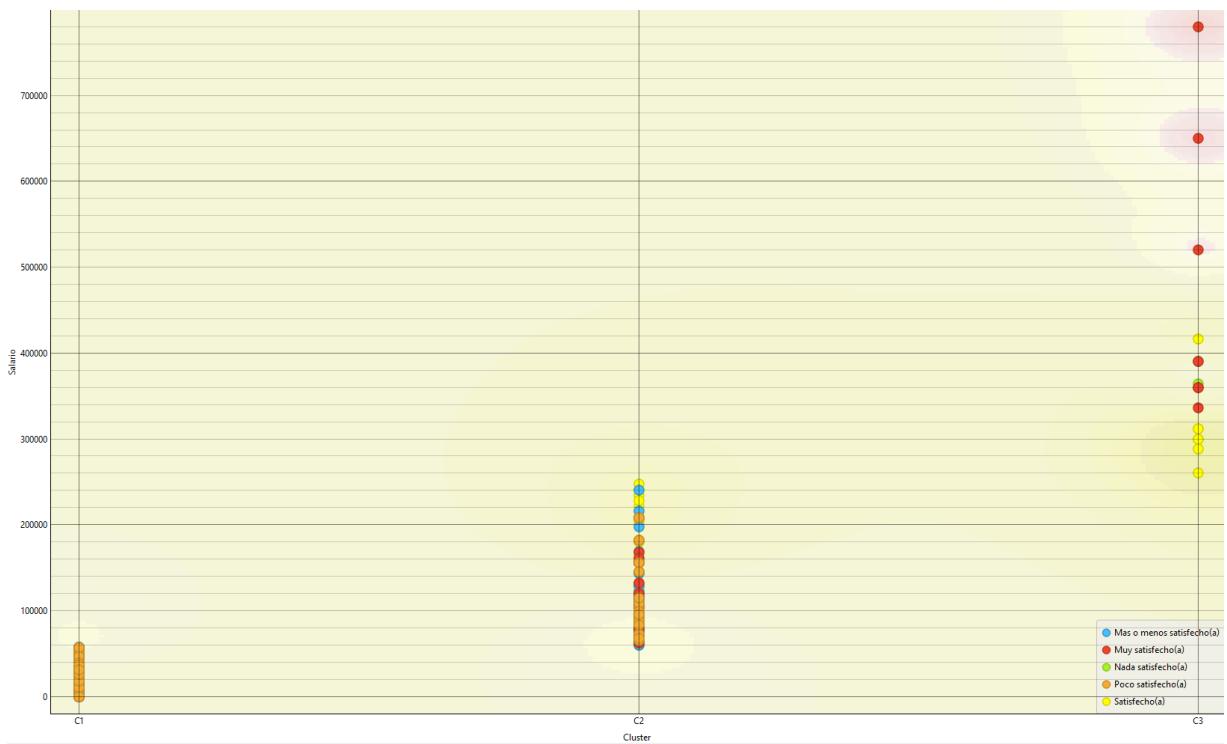


Distribución del Tiempo Dedicado a Actividades domésticas de Lunes a Viernes

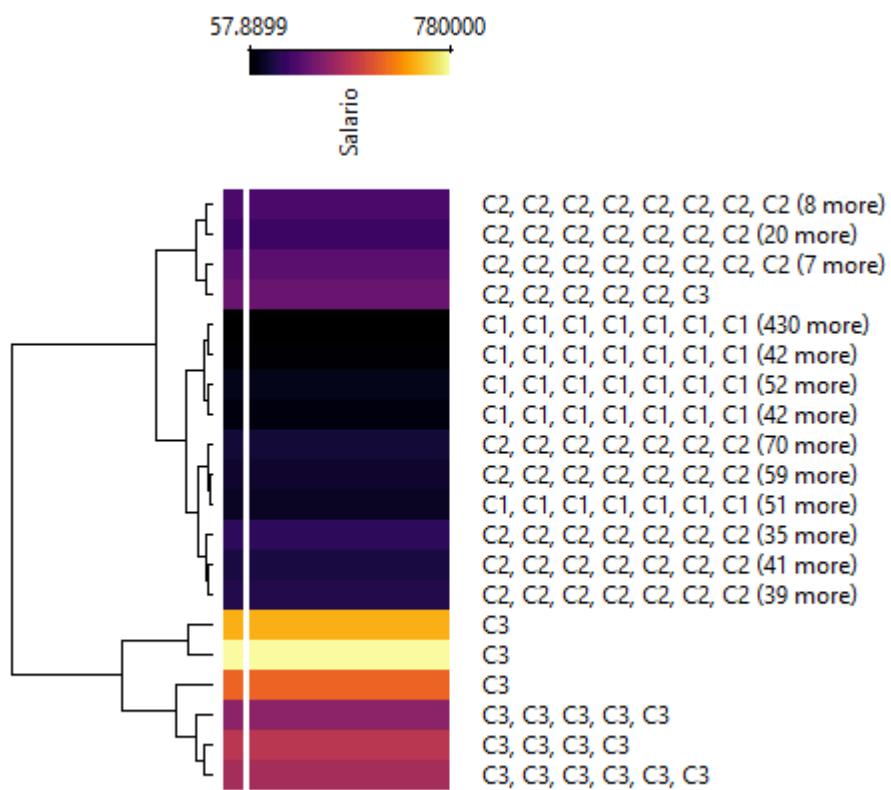


Distribución del Tiempo Dedicado a Actividades domésticas durante los Fines de Semana

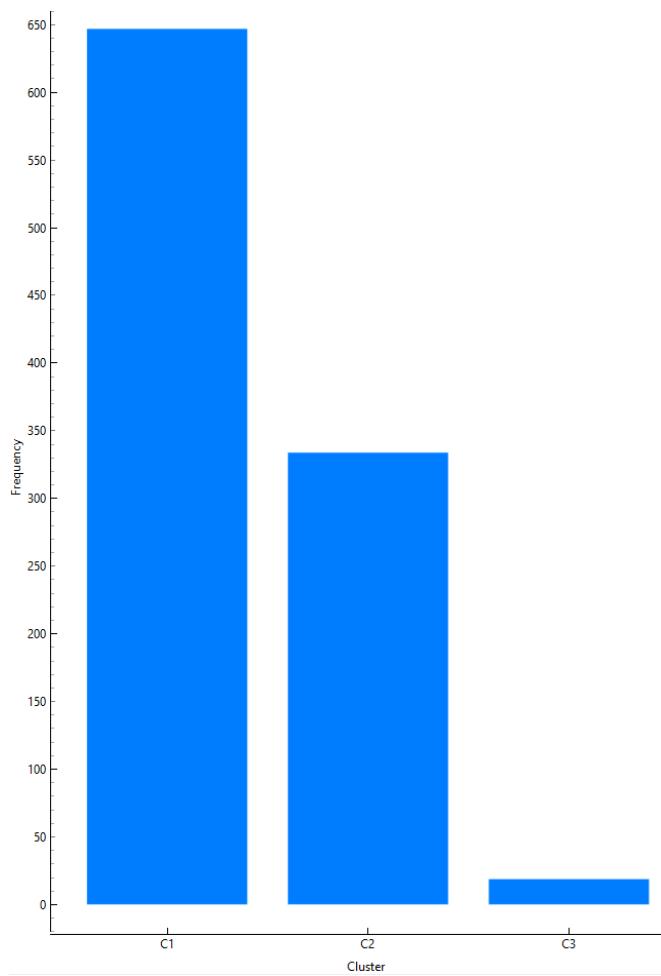
- **Situación Económica.**



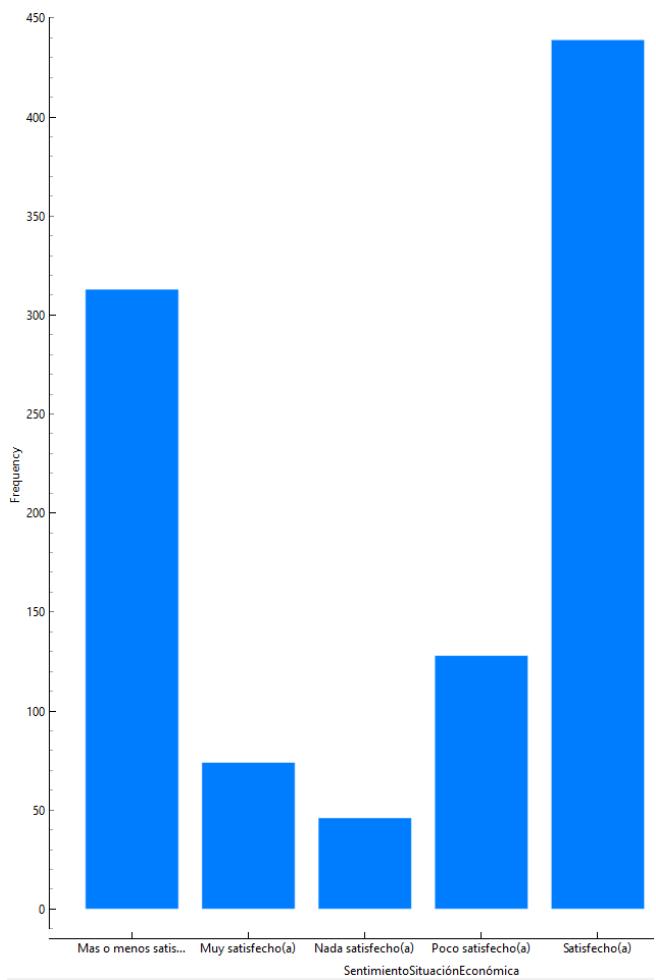
*Distribución de los clusters generados en función del salario y el nivel de satisfacción económica. Cada cluster (C1, C2 y C3) refleja segmentos diferenciados de personas agrupadas según sus ingresos y percepción económica.*



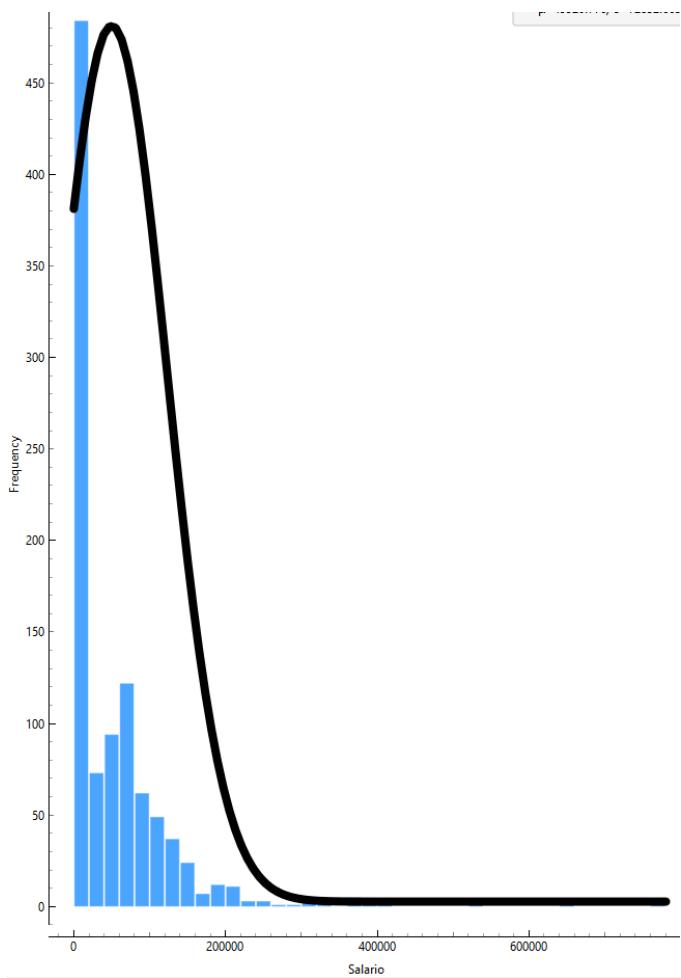
Mapa de calor jerárquico que representa la agrupación de individuos según sus salarios y la pertenencia a clusters (C1, C2 y C3).



*Distribución de Clústeres en Situación Económica.*



*Distribución del Sentimiento Respecto a Situación Económica.*



*Distribución del salario en Situación Económica.*

**iv. Presentar conclusiones sobre los patrones encontrados.**

**K-Means**

**Horas Laboradas**

- Se concluye que:
  - Cluster C1 (Verde)
    - Personas que dedican poco o nada de tiempo al trabajo entre semana y los fines de semana.
    - Este grupo puede estar compuesto por personas que no tienen trabajo formal.
    - Estas personas suelen estar bien con el tiempo que le dedican.
  - Cluster 2 (Rojo)
    - Personas que dedican una cantidad significativa de tiempo al trabajo tanto entre semana como los fines de semana (valores altos en ambos ejes).
    - Puede incluir profesionales, emprendedores o trabajadores con jornadas extendidas.
    - Las personas que trabajan sábados y domingos están bien con lo que dedican.
    - Las personas que trabajan de lunes a viernes quisieran dedicar menos tiempo.
    - Las personas que trabajan toda la semana quisieran dedicar menos tiempo.
  - 3. Clúster 3 (Azul)
    - Personas que dedican una cantidad moderada (promedio) de tiempo al trabajo entre semana y muy poco tiempo los fines de semana (valores intermedios en el eje X y bajos en el eje Y).
    - Personas que parecen tener un equilibrio en sus actividades laborales con jornadas promedio.

- En su mayoría están bien con el tiempo que le dedican y una proporción considerable preferiría dedicar menos tiempo.

### **Horas de Descanso**

- Se concluye que:
  - Cluster 1 (Azul)
    - Agrupa a las personas que dedican un tiempo de sueño promedio o por encima, tanto entre semana como durante el fin de semana.
    - Es el cluster con mayor número de personas, por lo que se infiere que en general las personas tiene un manejo aceptable de las horas de sueño
  - Cluster 2 (Rojo)
    - Agrupa a las personas que dedican menos tiempo de sueño promedio, podría estar vinculado a un estilo de vida más demandante o factores externos que limitan el sueño, como largas jornadas laborales o traslados prolongados.

### **Horas de Traslado**

- Se concluye que:
  - Cluster 1 (Azul)
    - Es el cluster con mayor número de personas, agrupa a las personas que dedican un tiempo promedio y bajo del promedio a trasladarse.
    - Indica que la mayoría de los individuos tienen patrones similares.
  - Cluster 2 (Rojo)
    - Agrupa a las personas que dedican tiempo de traslado mayor al promedio a trasladarse entre semana. Aunque sea un cluster de menor tamaño al Cluster 1, es de

considerar que son personas que exceden demasiado el promedio de horas entre semana.

- Cluster 3 (Verde)
  - Es el cluster de menor tamaño, agrupa a personas que dedican tiempo mayor al promedio a trasladarse sábados y domingos, posee agrupamientos muy altos al promedio.
- En su mayoría, predominan las personas que dicen sentirse bien con el tiempo que dedican aunque también hay una alta presencia de personas que dicen que desearían dedicar menos tiempo e incluyen personas que están debajo del promedio.

### **Horas de Estudio**

- Se concluye que:
  - Cluster 1 (Azul)
    - Representa la mayor cantidad de personas.
    - Personas que forman parte del promedio de horas dedicadas a estudiar toda la semana.
  - Cluster 2 (Rojo)
    - Representa una minoría.
    - Personas que dedican más tiempo a estudiar.
    - Estudian los fines de semana o toda la semana.
  - En promedio las personas dicen que están bien con el tiempo que dedican.
  - Existen casos aislados donde las personas dedican demasiado tiempo y aún quisieran dedicar más. Así como casos donde los alumnos dedican poco tiempo y quisieran dedicar menos tiempo.

### **Horas de Actividades Domésticas.**

- Se concluye que:
  - Cluster 1
    - Representa menos de la mitad del total de personas

- Personas que dedican menor e igual tiempo promedio a la actividad, podemos considerar que no es la principal actividad de esas personas.
- Cluster 2
  - Representa la mayor cantidad de personas.
  - Personas que dedican tiempo mayor al promedio a la actividad, sugiere que podrían ser personas que realizan dicha labor tanto para su propio beneficio como para terceros.
- Es una actividad que disfruta la mayoría aunque se le dedique mucho o poco tiempo.

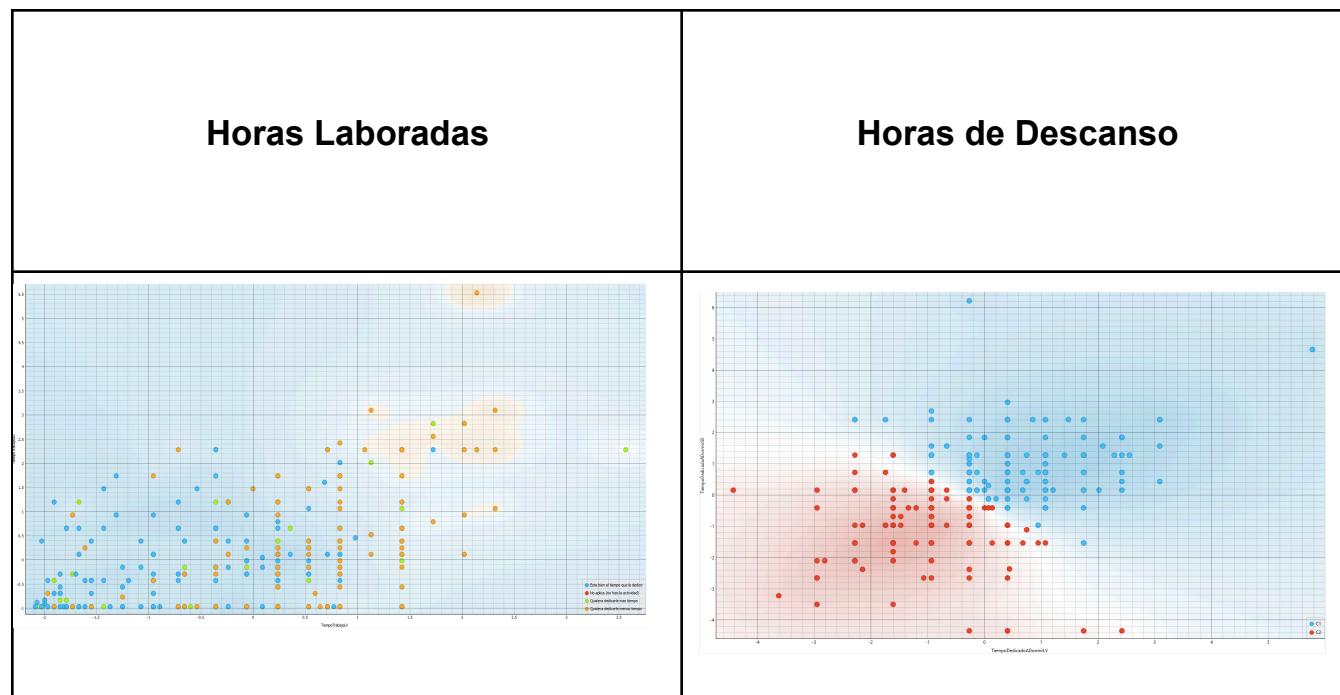
### **Situación Económica.**

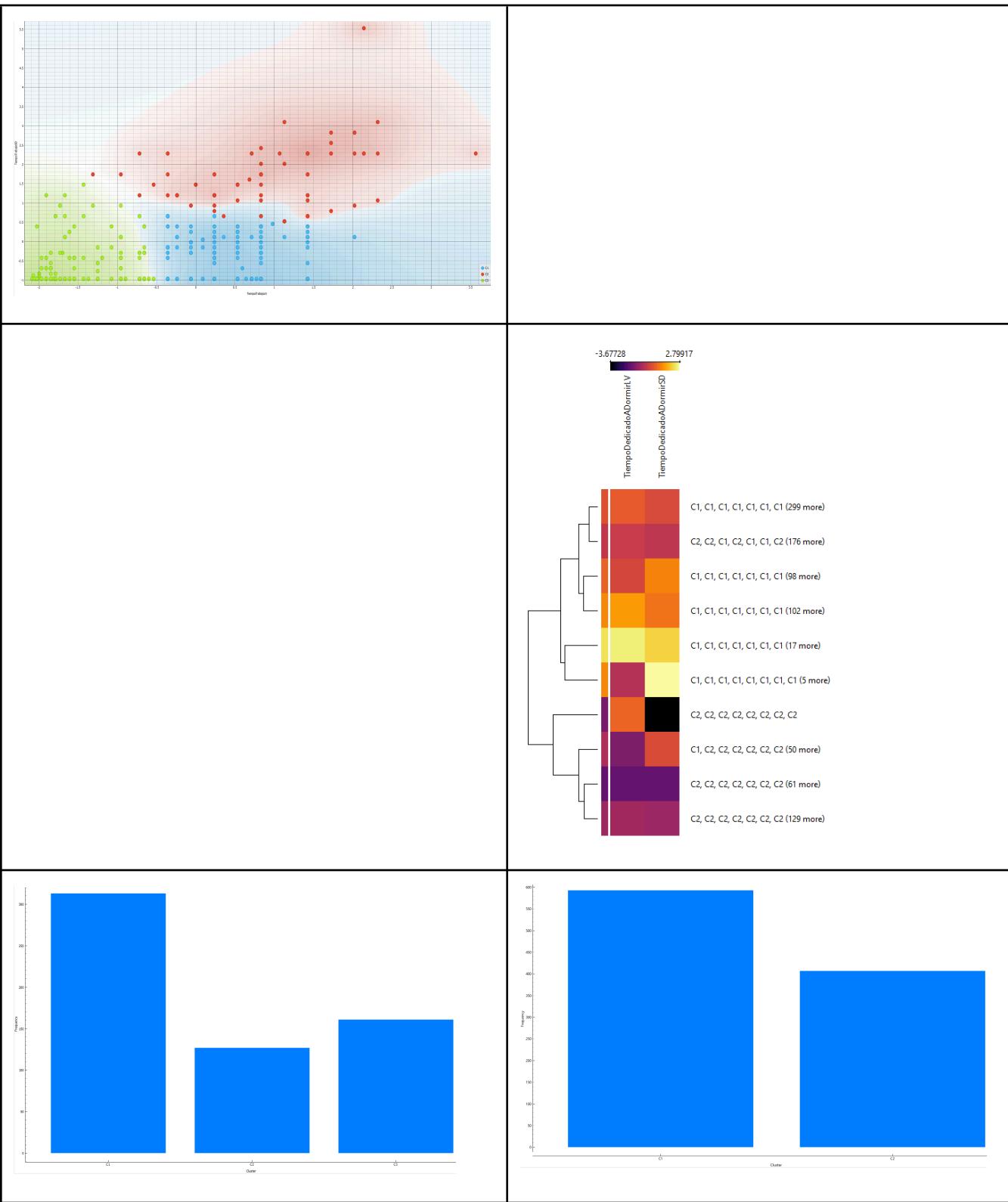
- Se concluye que:
  - Cluster 1
    - Representa la mayor cantidad de personas
    - Personas que perciben salarios anuales menores a \$80,000, un poco mayor al salario mínimo definido por el gobierno en ese mismo año.
    - Se sienten poco satisfechos con el salario que perciben
  - Cluster 2
    - Representa la cantidad media de personas
    - Personas que perciben salarios anuales mayores a \$80,000 y menores a \$250,000
    - Las personas se sienten poco satisfechas pero hay presencia de personas que se sienten nada satisfecho, más o menos satisfechos y pocos se sienten satisfechos.
  - Cluster 3
    - Representa la menor cantidad de las personas,

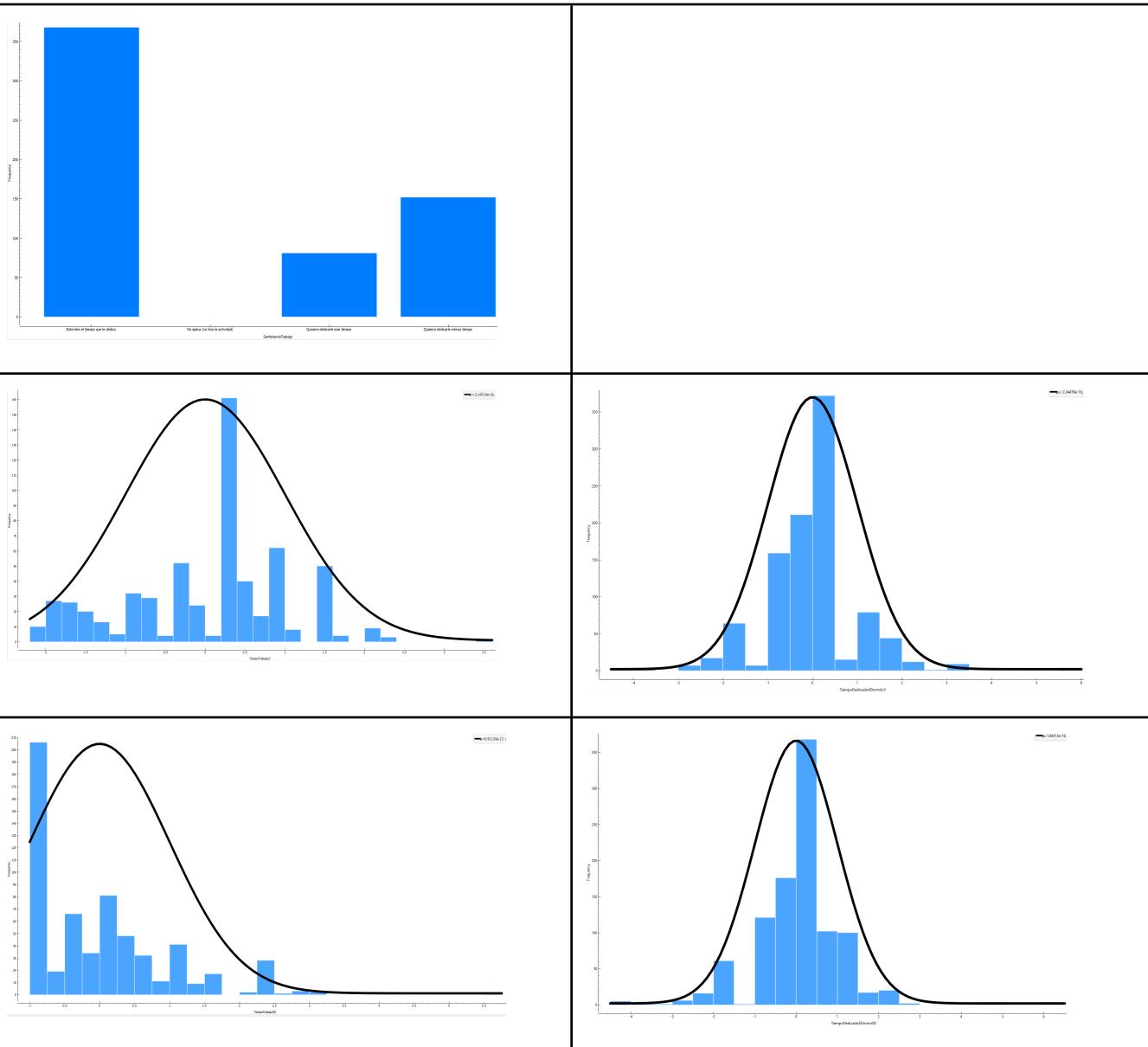
- Personas que perciben salarios anuales mayores a \$250,000
- Las personas se sienten satisfechas y muy satisfechas con el sueldo que perciben.
- Sugiere que la satisfacción de las personas aumenta en base al salario percibido, aunque existen casos donde las personas suelen sentirse satisfechas sin ganar tanto, se insinúa que depende de las aspiraciones de cada individuo.

#### d. Evaluación de modelos

i. Presentar los resultados anteriores en una tabla comparativa para ver el desempeño de todos.

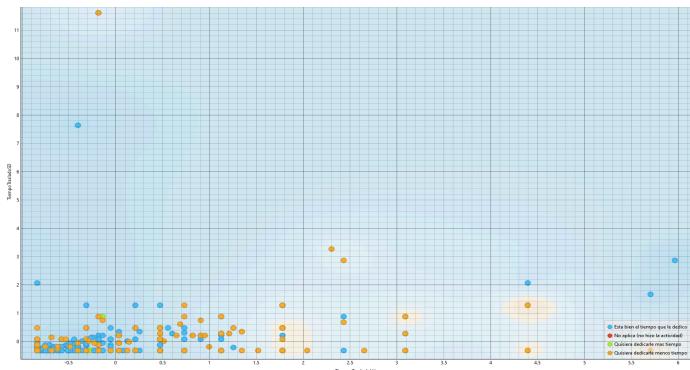




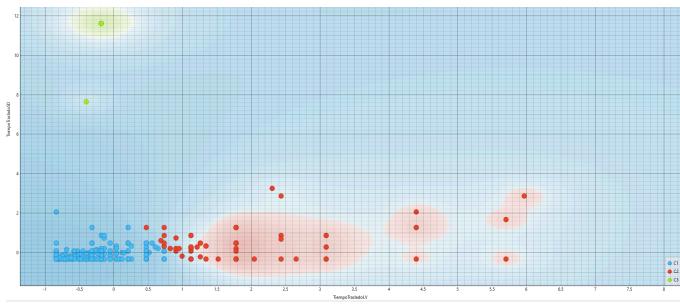


**Horas de Traslado**

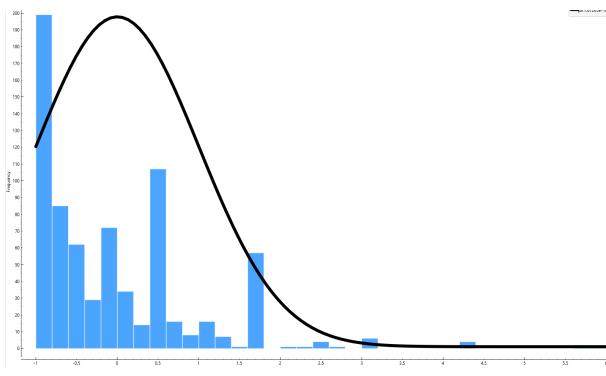
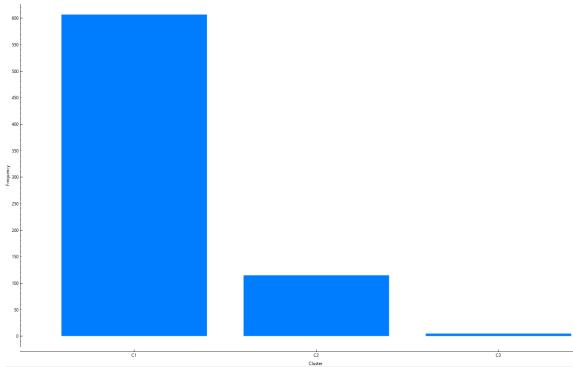
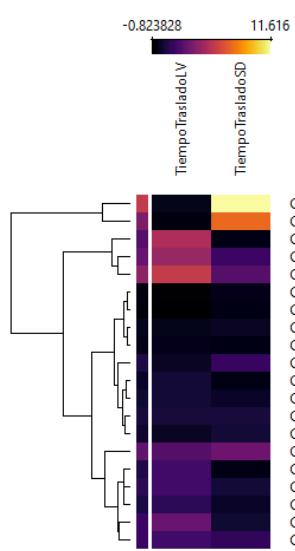
**Naive-Bayes**

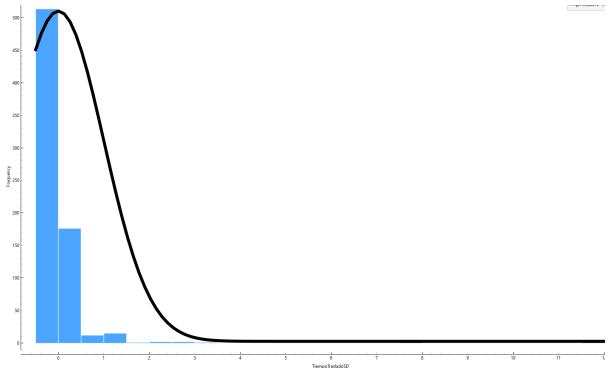


Resultados de las iteraciones			
	Precisión	Desviación estándar	Interpretación
1	0.9933333333333333	0.011376822876595733	Moderadamente estable
2	0.99666666666666667	0.011376822876595733	Moderadamente estable
3	0.99666666666666667	0.011376822876595733	Moderadamente estable
4	0.9933333333333333	0.011376822876595733	Moderadamente estable
5	0.99666666666666667	0.011376822876595733	Moderadamente estable
6	0.99666666666666667	0.011376822876595733	Moderadamente estable
7	0.9933333333333333	0.011376822876595733	Moderadamente estable
8	0.98666666666666667	0.011376822876595733	Moderadamente estable
9	0.99	0.011376822876595733	Moderadamente estable
10	0.99666666666666667	0.011376822876595733	Moderadamente estable
11	0.953333333333334	0.011376822876595733	Moderadamente estable
12	0.99	0.011376822876595733	Moderadamente estable
13	0.99	0.011376822876595733	Moderadamente estable
14	0.9933333333333333	0.011376822876595733	Moderadamente estable
15	0.9933333333333333	0.011376822876595733	Moderadamente estable
16	0.9933333333333333	0.011376822876595733	Moderadamente estable
17	0.9933333333333333	0.011376822876595733	Moderadamente estable
18	0.99666666666666667	0.011376822876595733	Moderadamente estable
19	0.9933333333333333	0.011376822876595733	Moderadamente estable
20	0.98666666666666667	0.011376822876595733	Moderadamente estable
21	0.98666666666666667	0.011376822876595733	Moderadamente estable
22	0.99	0.011376822876595733	Moderadamente estable
23	0.94666666666666667	0.011376822876595733	Moderadamente estable
24	1.0	0.011376822876595733	Moderadamente estable
25	0.9933333333333333	0.011376822876595733	Moderadamente estable
26	1.0	0.011376822876595733	Moderadamente estable
27	0.9933333333333333	0.011376822876595733	Moderadamente estable
28	0.99666666666666667	0.011376822876595733	Moderadamente estable
29	0.99	0.011376822876595733	Moderadamente estable
30	0.9933333333333333	0.011376822876595733	Moderadamente estable

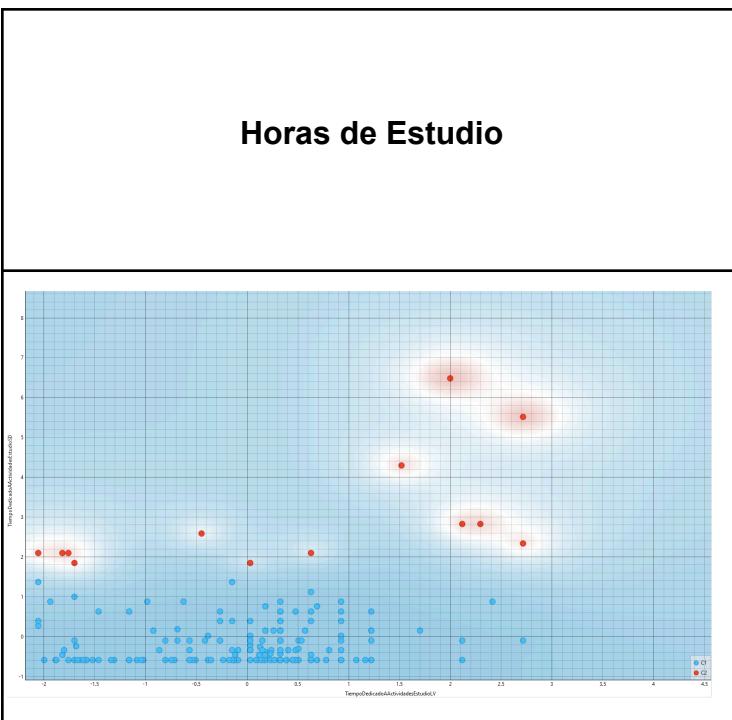


	Precisión	Desviación estándar	Interpretación
1	1.0	0.0	Muy estable
2	1.0	0.0	Muy estable
3	1.0	0.0	Muy estable
4	1.0	0.0	Muy estable
5	1.0	0.0	Muy estable
6	1.0	0.0	Muy estable
7	1.0	0.0	Muy estable
8	1.0	0.0	Muy estable
9	1.0	0.0	Muy estable
10	1.0	0.0	Muy estable
11	1.0	0.0	Muy estable
12	1.0	0.0	Muy estable
13	1.0	0.0	Muy estable
14	1.0	0.0	Muy estable
15	1.0	0.0	Muy estable
16	1.0	0.0	Muy estable
17	1.0	0.0	Muy estable
18	1.0	0.0	Muy estable
19	1.0	0.0	Muy estable
20	1.0	0.0	Muy estable
21	1.0	0.0	Muy estable
22	1.0	0.0	Muy estable
23	1.0	0.0	Muy estable
24	1.0	0.0	Muy estable
25	1.0	0.0	Muy estable
26	1.0	0.0	Muy estable
27	1.0	0.0	Muy estable
28	1.0	0.0	Muy estable
29	1.0	0.0	Muy estable
30	1.0	0.0	Muy estable

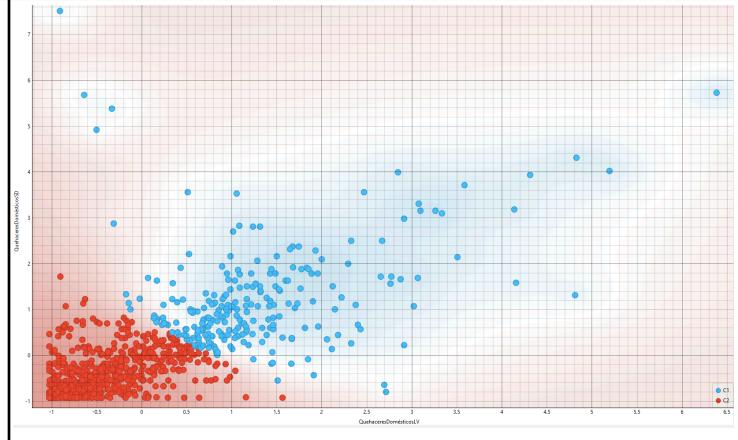


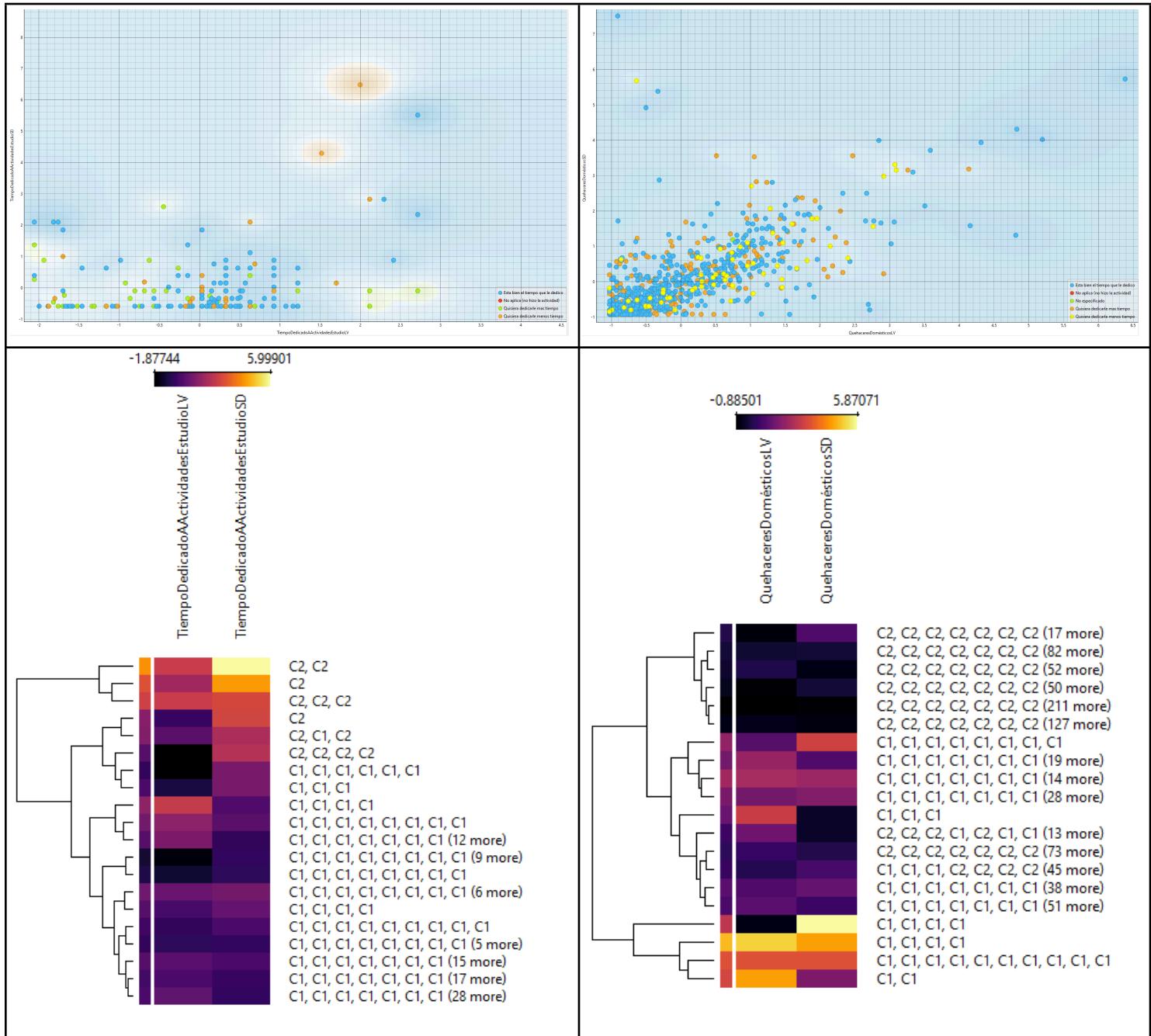


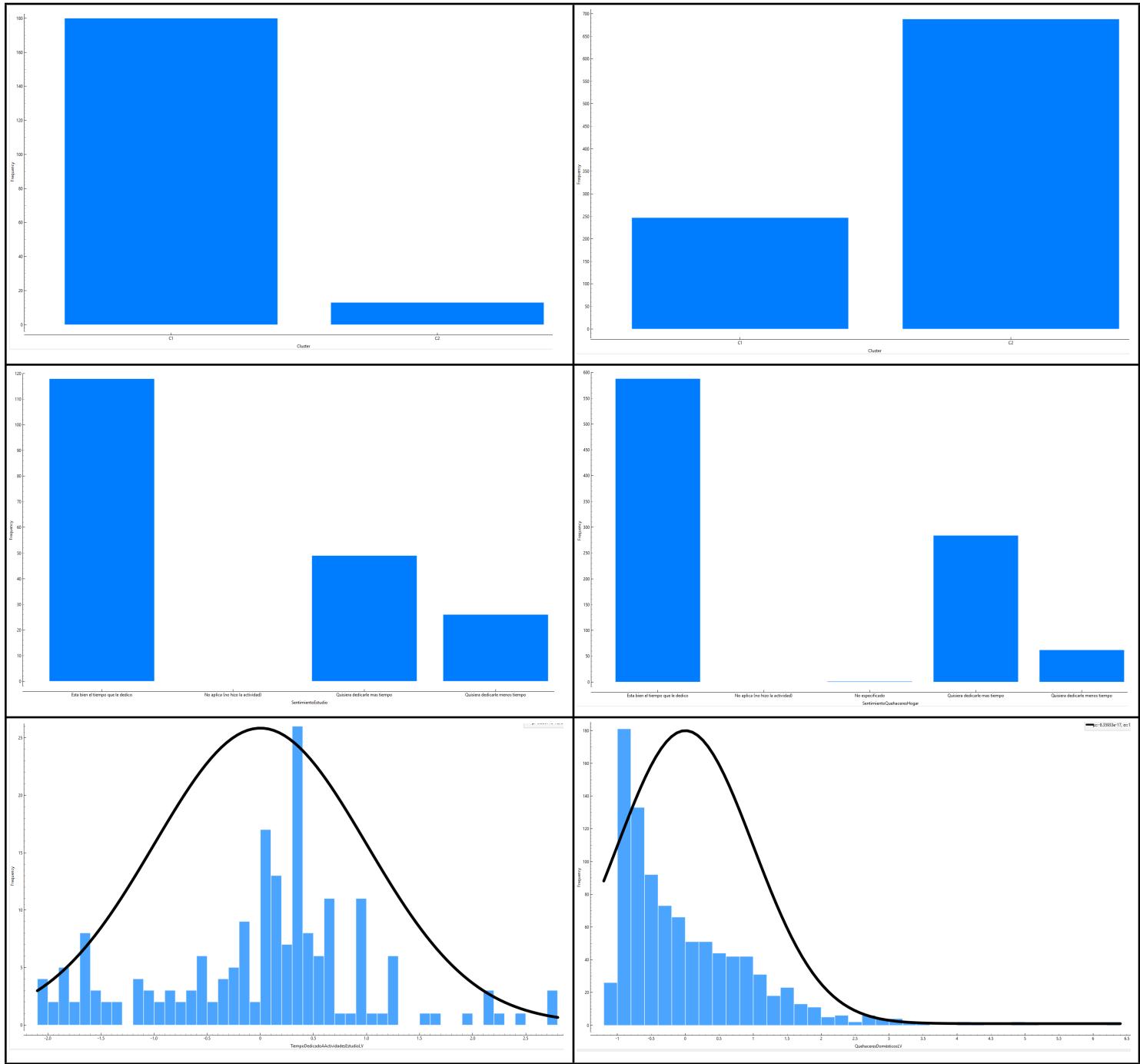
**Horas de Estudio**

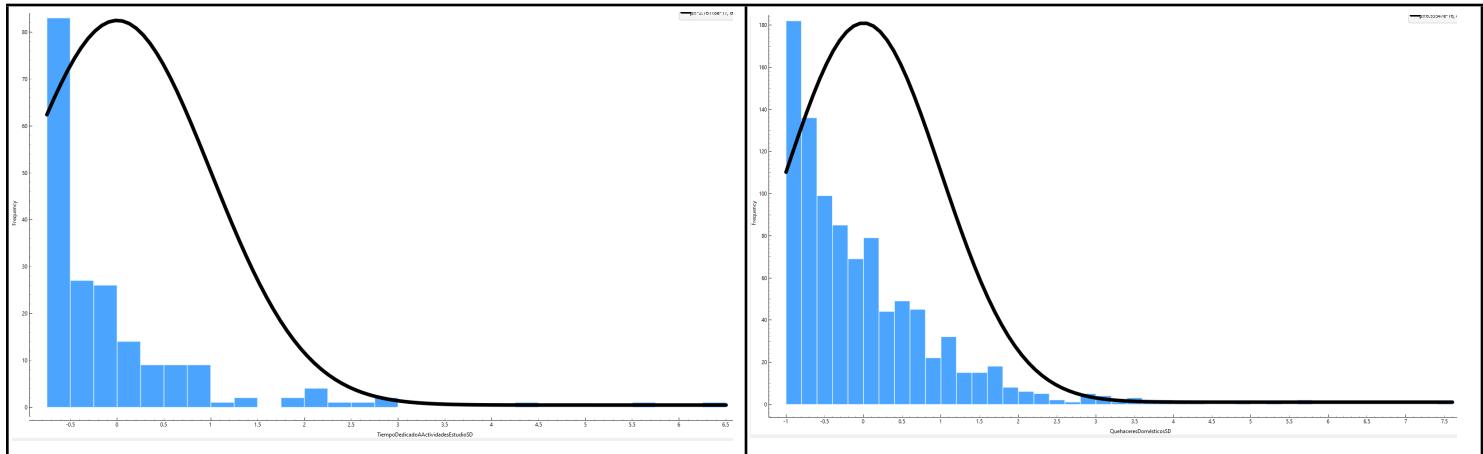


**Horas de Actividades Domésticas**

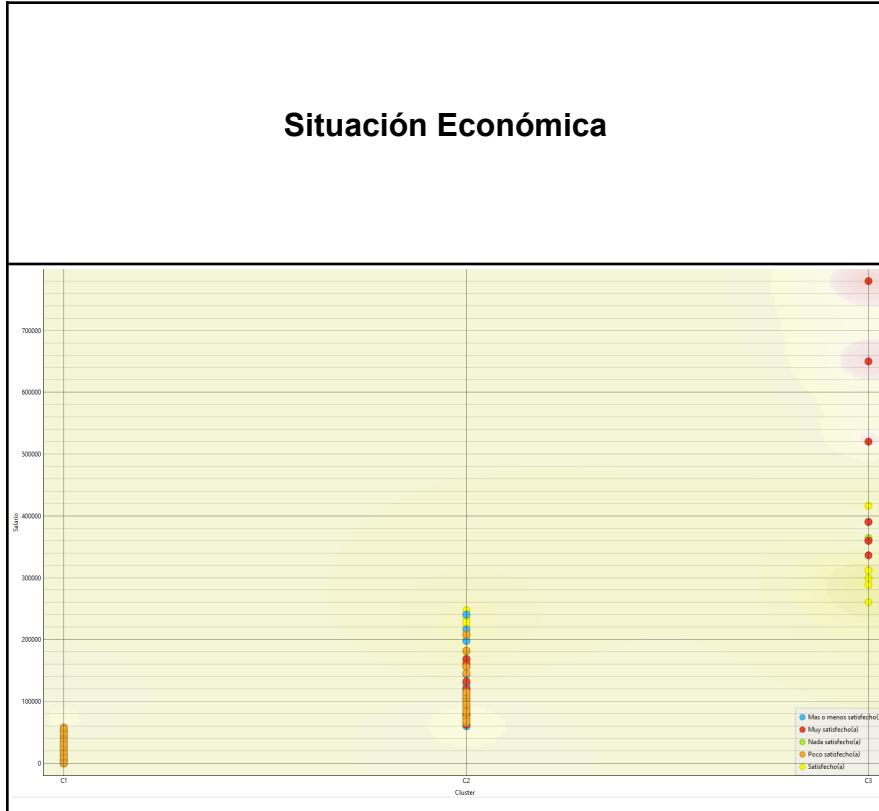


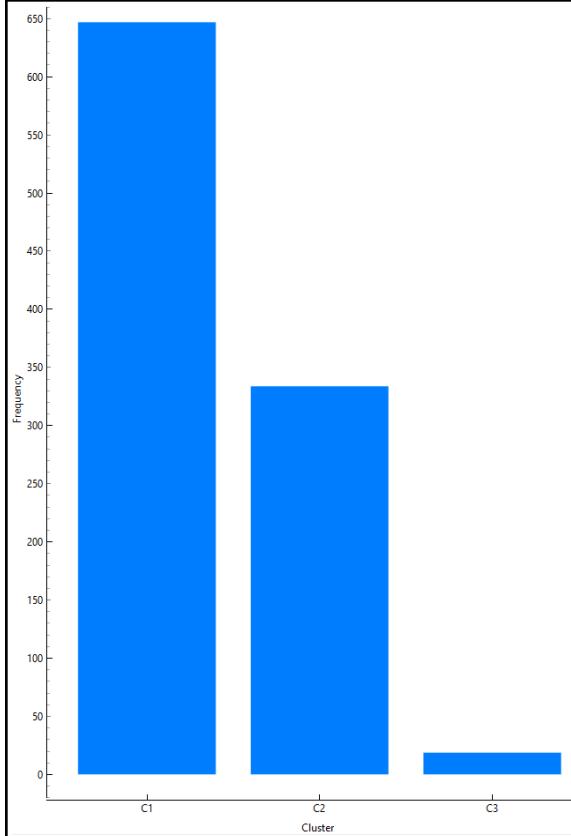
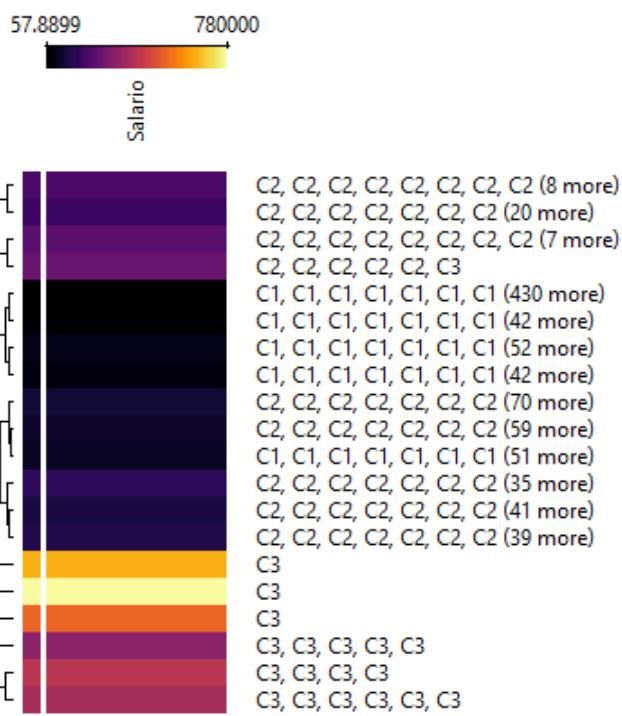


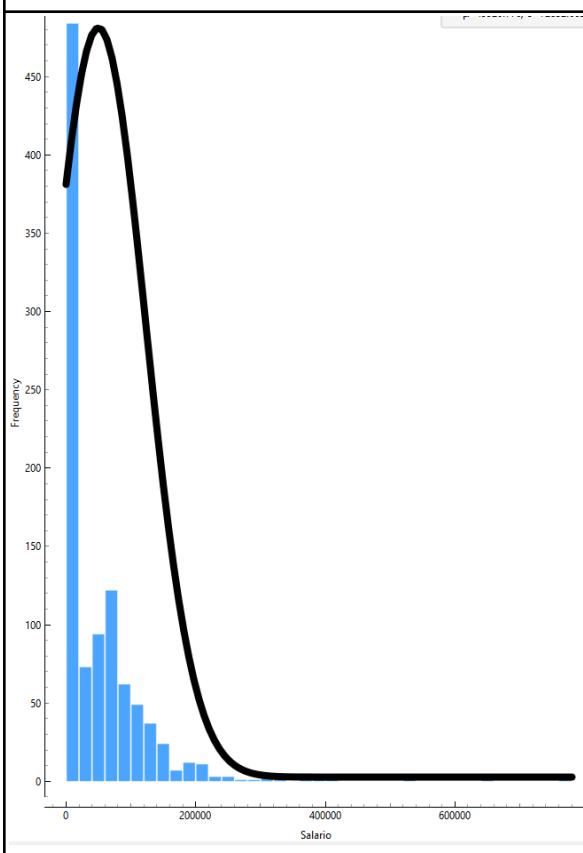
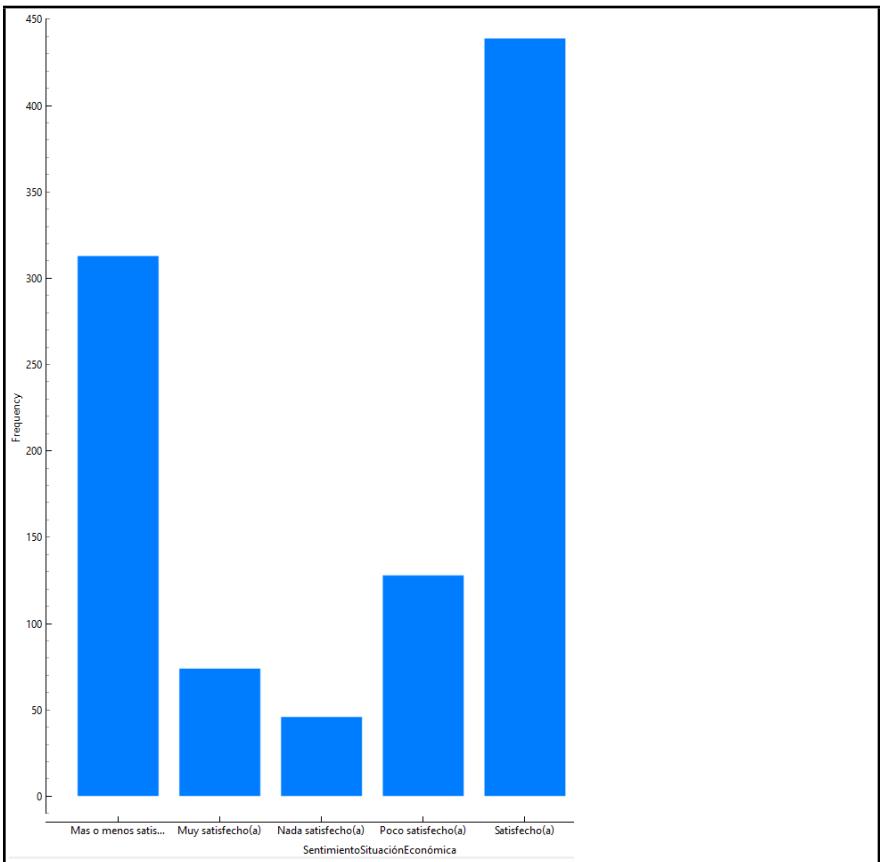




## Situación Económica







Para evaluar si Naive Bayes supera a K-Means en términos de utilidad para este problema, se pueden considerar las siguientes métricas clave:

Métrica	Naive Bayes	K-Means	Consideraciones
Precisión	Alta ▾	No aplica directamente ▾	La precisión en la clasificación de la "Felicidad General" es esencial. Naive Bayes proporciona probabilidades de clase que facilitan la interpretación.
Interpretabilidad	Alta ▾	Moderada ▾	Naive Bayes ofrece una comprensión clara de cómo cada variable contribuye a la predicción. K-Means identifica grupos, pero su interpretación requiere un análisis adicional.
Complejidad del Modelo	Baja ▾	Moderada ▾	Naive Bayes es un modelo simple y rápido de entrenar. K-Means puede requerir ajuste de hiperparámetros y evaluación de múltiples soluciones.
Capacidad para Manejar Datos Categóricos	Alta ▾	Baja ▾	Naive Bayes maneja naturalmente variables categóricas. K-Means requiere transformación o codificación de estas variables.

<b>Escalabilidad</b>	Alta ▾	Moderada ▾	Naive Bayes es eficiente con grandes conjuntos de datos. K-Means puede ser más lento con un gran número de instancias o dimensiones.
----------------------	--------	------------	--

**ii. Identificar el o los mejores modelos encontrados en la etapa anterior y explicarlos en términos del negocio.**

### **Naive Bayes**

El objetivo de este proyecto es analizar la relación entre las actividades de la población mexicana y su nivel de felicidad percibida, utilizando datos de la Encuesta Nacional sobre Uso del Tiempo (ENUT). Durante el análisis, se implementaron dos versiones del modelo Naive Bayes:

- **Modelo 1:** Aplicado al dataset completo (después de limpieza), compuesto por 68,000 instancias y 75 atributos, el cual logró una precisión del 100%.
- **Modelo 2:** Aplicado a una muestra aleatoria de 1,000 instancias con los mismos 75 atributos, obteniendo una precisión del 99%.

A primera vista, el resultado del Modelo 1 puede generar dudas sobre un posible sobreentrenamiento debido a la perfección aparente en la predicción. Sin embargo, el desempeño del modelo está justificado mediante una evaluación rigurosa con la técnica de Random Subsampling.

### **K-Means**

El mejor modelo de K-Means fue seleccionado utilizando el coeficiente de Silhouette Score, que permitió identificar el número óptimo de clusters para segmentar los datos. Este modelo generó dos clusters que agrupan a los

individuos según su comportamiento o características en relación a las horas de traslado.

## **Validación y Justificación**

### **Naive Bayes**

Para validar el Modelo 1 y reducir la posibilidad de sobreajuste, se llevaron a cabo 30 iteraciones usando validación cruzada con submuestreo aleatorio. En cada iteración:

- Los datos se dividieron en un 70% para entrenamiento y un 30% para prueba.
- Se entrenó y evaluó el modelo, calculando la precisión promedio y la desviación estándar de las métricas.

Los resultados mostraron una baja variabilidad en las métricas, lo que indica que el modelo es consistente y no depende de una partición específica de los datos. Esto valida que el 100% de precisión no es resultado de sobreajuste, sino del tamaño y calidad del dataset completo.

### **K-Means**

Al escoger el valor más alto de Silhouette Score, aseguramos la calidad de los clusters generados, evaluando la cohesión interna de cada cluster así como la separación entre ellos. Por lo que la segmentación es más precisa y significativa.

## **Implicaciones para el Negocio**

### **Naive Bayes**

Dado que ambos modelos ofrecen resultados sobresalientes, es importante no alarmarse por el 100% de precisión del Modelo 1. Este resultado ha sido validado y puede considerarse confiable gracias al enfoque de evaluación

utilizado. No obstante, optar por el Modelo 2 con 99% de precisión también es una alternativa válida si se busca un balance entre eficiencia y robustez, especialmente en escenarios donde el tiempo de ejecución es crítico.

Ambos modelos ofrecen herramientas valiosas para entender cómo las actividades de la población afectan su felicidad percibida. Esto puede contribuir a la creación de políticas públicas y estrategias que fomenten actividades relacionadas con el bienestar, basadas en predicciones confiables y bien fundamentadas.

### **K-Means**

El usar este modelo permitió dividir a la población en segmentos bien definidos para analizar y entender sus necesidades y comportamientos. Cada cluster puede tener productos, servicios o estrategias diseñados específicamente para sus necesidades. Brindando una visión amplia para la generación de ideas y el desarrollo de proyectos que ayuden a la sociedad.

Al analizar cómo se distribuyen los grupos, podemos optimizar recursos enfocando esfuerzos en los grupos más grandes, logrando impactos potenciales de gran magnitud. Desde campañas publicitarias hasta iniciativas o inversiones de alto nivel.

**iii. Comentarios sobre el desempeño de cada uno de los modelos e indicar sus puntos de vista sobre por qué tiene el desempeño que tienen (bueno o malo).**

Naive Bayes: Dataset completo posterior a limpieza (68, 000 instancias, 75 atributos)

- Desempeño: Precisión del 100%.
- Comentario: Este modelo logró predecir correctamente todas las instancias del dataset completo, alcanzando una precisión perfecta. Si bien esto puede generar inquietudes sobre un posible

sobreentrenamiento, el resultado fue validado mediante 30 iteraciones de validación cruzada con submuestreo aleatorio (Random Subsampling), mostrando baja variabilidad en las métricas y confirmando la consistencia del modelo.

- Razones del buen desempeño:
  - Tamaño y calidad del dataset: El uso de 68,000 instancias proporcionó una muestra representativa que permitió al modelo capturar patrones complejos y generalizables.
  - Optimización del código: Las mejoras implementadas, especialmente el uso de procesamiento multihilo, permitieron manejar eficientemente el gran volumen de datos y reducir significativamente los tiempos de ejecución.
- Consideración: El tiempo de ejecución inicial de más de 5 horas por iteración se redujo a 5 segundos gracias a estas optimizaciones, lo que hizo viable realizar todas las iteraciones necesarias para validar el modelo. Este modelo es ideal para escenarios donde la máxima precisión y estabilidad son prioritarias, aunque puede ser menos práctico en situaciones con restricciones de tiempo.

#### Modelo 2: Muestra aleatoria (1,000 instancias, 75 atributos)

- Desempeño: Precisión del 99%.
- Comentario: A pesar de trabajar con una muestra significativamente más pequeña, este modelo mantuvo un rendimiento excelente, logrando una precisión muy alta. La ligera disminución en comparación con el Modelo 1 es razonable debido al menor volumen de datos disponibles para el entrenamiento.
- Razones del buen desempeño:
  - Representatividad de la muestra: La muestra aleatoria conservó la diversidad y los patrones clave del dataset completo, permitiendo al modelo generalizar bien.

- Optimización y eficiencia: Gracias a las optimizaciones aplicadas, el tiempo de ejecución del modelo pasó de 18 segundos a menos de medio segundo por iteración, haciendo posible su uso en análisis rápidos y frecuentes.
- Consideración: Aunque el modelo es más rápido y eficiente, es importante tener en cuenta que su capacidad para generalizar puede ser ligeramente menor en comparación con el Modelo 1 debido al tamaño reducido de la muestra.

### **Optimización y uso de multihilo:**

Uno de los aspectos más destacados en la mejora del desempeño fue la implementación de procesamiento multihilo, que permitió distribuir la carga computacional entre varios núcleos del procesador. Esto fue clave para:

- Acelerar cálculos intensivos, como el manejo de grandes volúmenes de datos y el cálculo de probabilidades.
- Mejorar la escalabilidad, permitiendo que ambos modelos se ejecuten de manera eficiente incluso en escenarios más exigentes.

### **K-Means**

Se facilitó mucho el trabajo debido a que la estructura implementada en Orange fue la misma en cada modelo, cada uno presentó un buen desempeño según el coeficiente de Silhouette Score, el haber realizado varios modelos con las relaciones de atributos definidos, ayudó a mejorar bastante el score a comparación de solo haber creado uno. Aportó a mejorar no solo la segmentación sino también la interpretación de cada uno de los resultados.

## Referencias:

- Campo, M<sup>a</sup> Teresa del. (2018). Bienestar y trabajo. Revista de la Asociación Española de Especialistas en Medicina del Trabajo, 27(4), 195-196. Epub 00 de abril de 2019. Recuperado de: [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S3020-11602018000400001&lng=es&tlng=es](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S3020-11602018000400001&lng=es&tlng=es).
- Follis, K. (2024). Conceptos de minería de datos. Microsoft Learn. Recuperado de: <https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>
- Follis, K. (2023). Modelos de minería de datos (Analysis Services - Minería de datos). Microsoft Learn. Recuperado de: <https://learn.microsoft.com/es-es/analysis-services/data-mining/mining-models-analysis-services-data-mining?view=asallproducts-allversions>
- Granada D H (2016). Una técnica de clasificación con variables categóricas. Ciencia y desarrollo. 7. 15-20. 10.19053/01217488.4226. Recuperado de: [https://www.researchgate.net/publication/307509532\\_Una\\_tecnica\\_de\\_clasificacion\\_con\\_variables\\_categoricas](https://www.researchgate.net/publication/307509532_Una_tecnica_de_clasificacion_con_variables_categoricas)
- IBM. (2024a). Introducción: Clústeres. IBM Documentation. <https://www.ibm.com/docs/es/was-zos/9.0.5?topic=servers-introduction-cluster-s>
- IBM. (2024). Machine Learning. <https://www.ibm.com/mx-es/topics/machine-learning>.
- INE - Instituto Nacional de Estadística. (2024). Glosario de conceptos. INE. Recuperado de: <https://www.ine.es/DEFIne/es/concepto.htm?c=4484#:~:text=Seg%C3%BAn%20la%20OMS%20el%20%22sexo,apropriados%20para%20hombres%20y%20mujeres.>

- Instituto Nacional de Estadística y Geografía (INEGI). (1999). *INEGI, Quiénes somos.* www.inegi.org.mx. Recuperado de: <https://www.inegi.org.mx/inegi/contenido/instituto.html>
- Instituto Nacional de Estadística y Geografía(INEGI). (2019). Encuesta Nacional sobre Uso del Tiempo (ENUT) 2019. Recuperado de: www.inegi.org.mx. <https://www.inegi.org.mx/programas/enut/2019/>
- López, J. F. (2024, 18 marzo). Muestra estadística: Qué es, tipos y ejemplos. Economipedia. Recuperado de: <https://economipedia.com/definiciones/muestra-estadistica.html>
- Marcela, Julieth & Gil Quemba, Julieth & Francisco, Juan & Muñoz-Olano, Juan. (2017). La felicidad: Conceptos, teorías, formas de medición y discusiones. Happiness: Concepts, theories, measurement and discussions. Recuperado de: [https://www.researchgate.net/publication/315744708\\_La\\_felicidad\\_Conceptos\\_teorias\\_formas\\_de\\_medicion\\_y\\_discusiones\\_Happiness\\_Concepts\\_theories\\_measurement\\_and\\_discussions](https://www.researchgate.net/publication/315744708_La_felicidad_Conceptos_teorias_formas_de_medicion_y_discusiones_Happiness_Concepts_theories_measurement_and_discussions)
- Mahmoud, M. H. (2022). What is a kaggle? Kaggle.com. Recuperado de: <https://www.kaggle.com/discussions/general/328265>
- NumPy documentation — NumPy v2.1 manual. (2024). Numpy.org. Recuperado de: <https://numpy.org/doc/stable/>
- OMS. (2023). Lograr el bienestar. Who.int. <https://iris.who.int/bitstream/handle/10665/376906/9789240091597-spa.pdf>
- ORACLE. (2019). Interfaz de línea de comandos (CLI). Documentación de Oracle Cloud Infrastructure. <https://docs.oracle.com/es-ww/iaas/Content/API/Concepts/cliconcepts.htm>
- Roy-García, I., Rivas-Ruiz, R., Pérez-Rodríguez, M., & Palacios-Cruz, L. (2019). Correlación: no toda correlación implica causalidad. redalyc.org. Recuperado de: <https://www.redalyc.org/journal/4867/486761439011/html/>
- Python Software Foundation. (2024). Tutorial de Python. Python documentation. Recuperado de: <https://docs.python.org/es/3.13/tutorial/>

- Perez Prieto, J. A. (2019). La librería científica Scipy. research.iac.es. Recuperado de: <https://research.iac.es/sieinvents/python-course/scipy.html>
- The United Nations. (2024). UNSD - welcome to UNSD. Unstats.un.org. Recuperado de: <https://unstats.un.org/UNSDWebsite/>
- Universidad de Antioquia. (2020). PyQt — Tutorial de PyQt. Recuperado de: <https://juan-suarezp.github.io/PyQt/intro.html>
- Universidad Latina. (2023). ¿Qué es un Algoritmo en Informática? UNILA. Unila. Recuperado de: <https://www.unila.edu.mx/que-es-un-algoritmo-en-informatica/>