



Tecnológico de Monterrey
Escuela de Ingeniería y Ciencias

Campus Puebla

**Reporte técnico de desarrollo de proyectos de análisis
de datos**

Course:

IN1002B.302

Germán Reyes Hernández
Rodrigo Jalil Aguilar Martínez
José Carlos Trueba Martinez
Emilio Arturo Kayary Ramos
Pablo Moreno Cagigas

27/11/2023

Gabriela Guadalupe Reyes Zarate
Alfredo García Suárez

TABLA DE CONTENIDOS

1. Capítulo 1. Project Charter
 - Descripción de la necesidad y justificación del proyecto
 - Objetivo del proyecto
 - Beneficios del proyecto
 - Alcance
 - Metas intermedias
 - Descripción de los entregables finales del proyecto
2. Capítulo 2. Preparación de base de datos
 - Descripción de la metodología utilizada
 - Descripción de los pasos y criterios establecidos en la toma de decisiones
 - Interpretación de los análisis exploratorios gráficos
 - Resultados y conclusiones de preparación de datos
3. Capítulo 3. Modelos estadísticos de datos
 - Metodología del análisis de datos
 - Modelo inicial de regresión
 - Análisis numéricos y gráficos de los residuales
 - Análisis de los índices de correlación y potencia
 - Análisis de normalidad de los datos
 - Optimización del modelo. Descripción e interpretación
4. Conclusiones finales
5. Lecciones aprendidas (individuales)
6. Referencias bibliográficas
7. Anexos

ENTREGABLE 1. PROJECT CHARTER

NOMBRE DEL PROYECTO: PROJECT CHARTER

EQUIPO DEL PROYECTO Y
CORREO ELECTRÓNICO:

Germán Reyes Hernández
A01737467@tec.mx
Rodrigo Jalil Aguilar Martínez
A01425281@tec.mx
José Carlos Trueba Martinez
A01733813@tec.mx
Emilio Arturo Kayary Ramos
A01552526@tec.mx
Pablo Moreno Cagigas
a01737838@tec.mx

PROPÓSITO DEL PROYECTO

Descripción de la Necesidad y Justificación del proyecto

Necesidad:

El canal tradicional de retail en México representa el 90% del total de puntos de venta en el país. Sin embargo, se enfrenta a diversos retos, como la competencia de los grandes retailers, la digitalización de los consumidores y la necesidad de mejorar su eficiencia operativa. Para superar estos retos, el canal tradicional debe innovar su modelo de negocios adoptando nuevas tecnologías, desarrollando nuevos productos y servicios, y mejorando su experiencia de compra.

Justificación:

La recolección de datos por medio de encuestas a microempresas en la ciudad de Puebla es necesaria para comprender las necesidades y desafíos de este segmento del canal tradicional. Estos datos permitirán identificar las áreas de oportunidad para la innovación, desarrollar soluciones personalizadas y promover el crecimiento de las microempresas.

Objetivo del Proyecto

Realizar, recopilar y analizar encuestas detalladas a microempresas del canal tradicional de retail en Puebla, con la finalidad de obtener información específica sobre sus operaciones, desafíos y necesidades. La meta cuantificable es llevar a cabo un mínimo de 60 encuestas, asegurando una muestra representativa. Para lograrlo, se establecerá una colaboración activa entre todos los integrantes del equipo de encuestadores. La relevancia de este esfuerzo radica en obtener insights clave que destacan áreas de oportunidad para la innovación, abordando desafíos cruciales como la competencia, la digitalización y la eficiencia operativa en el canal tradicional de retail. El proyecto tiene un plazo definido de 4 semanas para completar la recopilación y análisis de datos, proporcionando resultados accionables que guiarán estrategias de innovación en las microempresas del canal tradicional en Puebla.

Beneficios del Proyecto

- Obtener información sobre las necesidades y desafíos del segmento
- Identificar áreas de oportunidad para la innovación en los negocios
- Desarrollar soluciones personalizadas para apoyar el crecimiento de las microempresas
- Mejorar la competitividad de las microempresas del canal tradicional de retail en la ciudad de Puebla

DESCRIPCIÓN DEL PROYECTO

Alcance

- Objetos de estudio: Microempresas del canal tradicional de retail en la ciudad de Puebla.
- Ámbito geográfico: Zona Capu, 72045 Heroica Puebla de Zaragoza, Pue.
- Período de ejecución: Del 1 de noviembre al 1 de diciembre
- Metodología: Encuesta en línea por medio de Google Forms
- Entregables:
 - Informe de resultados de la encuesta por Google Forms
 - Identificación de áreas de oportunidad para la innovación por medio de modelos matemáticos.

METAS INTERMEDIAS

Descripción

- Meta 1: Distribución de la encuesta a un número representativo de microempresas en la ciudad de Puebla.
- Meta 2: Recopilación de datos de al menos el 50% de las microempresas encuestadas.

DESCRIPCIÓN DE LOS ENTREGABLES AL FINAL DEL PROYECTO

NOMBRE Y TIPO DE DOCUMENTO	CRITERIO DE ACEPTACIÓN
Repositorio de GitHub	Repositorio con varios documentos acerca de la recopilación de datos por medio de modelos matemáticos basados en las encuestas.

Presentación de PowerPoint	Presentación recopilando todos los datos acerca de la investigación y la metodología utilizada, presentando información final.
ProjectLibre	Esquema acerca de todos los procesos de recopilación de información y el proyecto en general.
Project Charter	Validación del desarrollo de análisis de datos mediante el Project Charter.

ENTREGABLE 2. PREPARACIÓN DE LA BASE DE DATOS

NOMBRE DEL PROYECTO: ASESORÍA PARA ALUMNOS

EQUIPO DEL PROYECTO Y
CORREO ELECTRÓNICO:

Germán Reyes Hernández
A01737467@tec.mx
Rodrigo Jalil Aguilar Martínez
A01425281@tec.mx
José Carlos Trueba Martinez
A01733813@tec.mx
Emilio Arturo Kayary Ramos
A01552526@tec.mx
Pablo Moreno Cagigas
a01737838@tec.mx

PROPÓSITO DEL PROYECTO

Descripción de la Necesidad y Justificación del proyecto

Elaborar y validar el project charter del desarrollo de análisis de datos de los valores de la base de datos.

Descripción de la metodología utilizada

El enfoque general de la investigación fue un estudio mixto con un diseño experimental. Se utilizaron técnicas de recopilación de datos primarios por medio de una encuesta en Google Forms. También se utilizaron métodos estadísticos como la regresión lineal múltiple, la filtración de datos cualitativos y cuantitativos para realizar diferentes gráficas.

Descripción de los pasos y criterios establecidos en la toma de decisiones

Para la preparación de la base de datos, tuvimos que seguir 5 pasos.

1. Descargar el archivo .csv y revisar los datos.
2. Realizar el procedimiento para los valores nulos.
3. Filtrar los datos cuantitativos y cualitativos.
4. Realizar el procedimiento de los outliers a los datos cuantitativos.
5. Realizar gráficas para visualizar datos obtenidos.

Los criterios para la preparación son los siguientes:

1. Datos completos y concisos, eliminando valores duplicados o faltantes e igualarlos a cero.
2. Datos relevantes para el propósito de la investigación recabados por Google Forms.

Interpretación de los análisis exploratorios gráficos

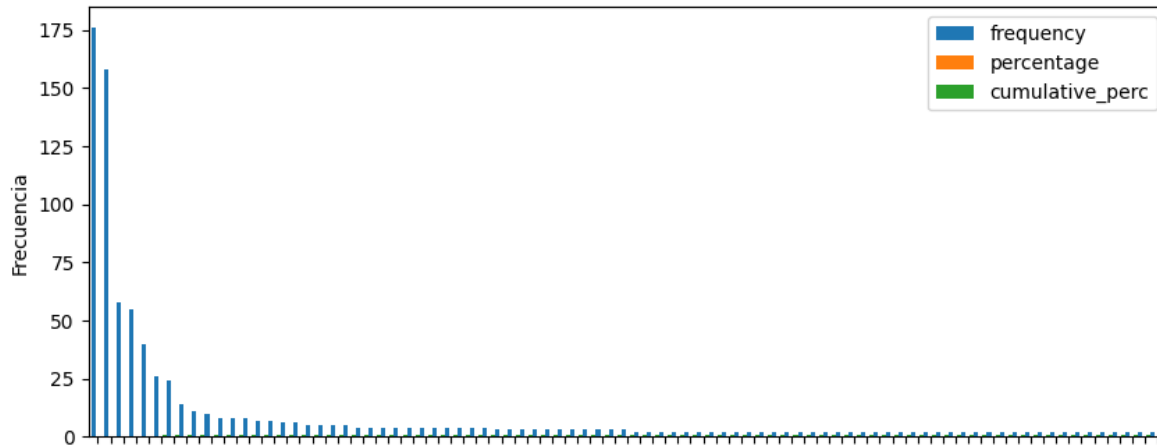


Figura 1. Tipo de local

Aquí podemos visualizar los diferentes tipos de negocios que se entrevistaron, con un valor de 175 de estos siendo pequeños restaurantes o fondas y varios negocios no establecidos en la encuesta con una frecuencia menor a 5.

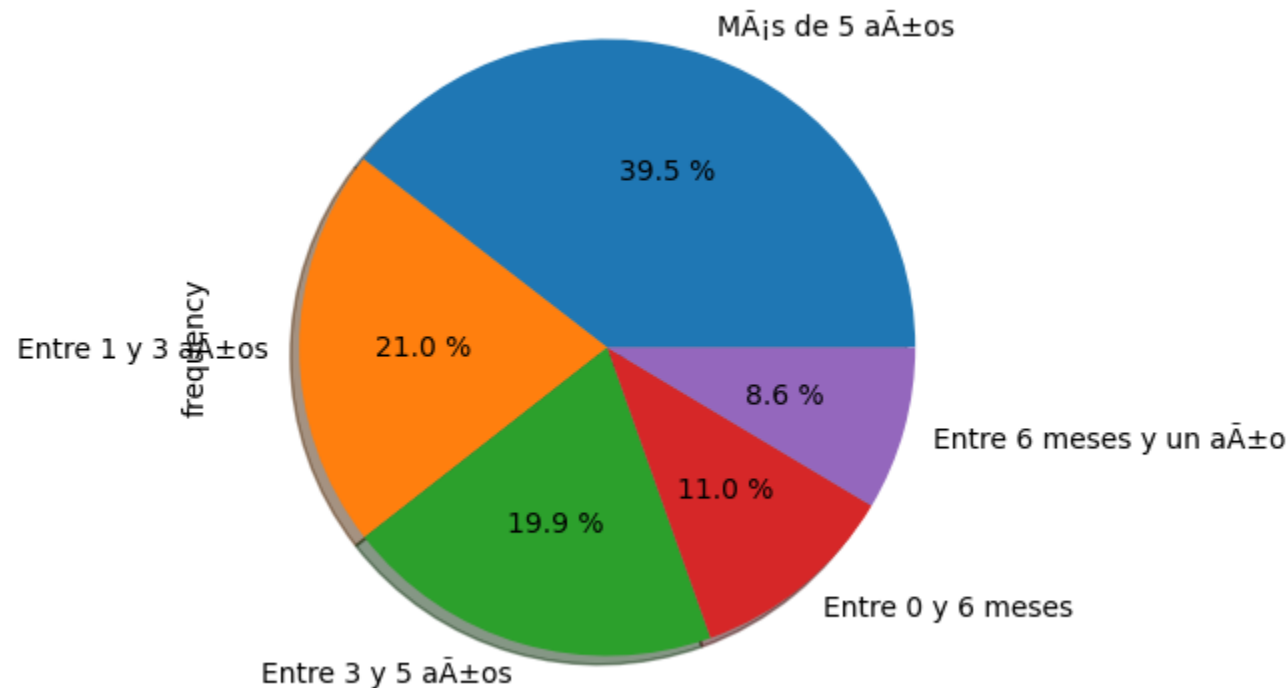


Figura 2. Antigüedad de negocios

En esta gráfica podemos observar la antigüedad de los negocios, con un porcentaje mayor pasando de los 5 años y un porcentaje menor de 8.6% entre los 6 meses y un año de ser establecidos.



A pie chart illustrating the distribution of satisfaction levels. The chart is divided into four segments: a large blue segment for 'Muy satisfecho' (49.6%), a large orange segment for 'Algo satisfecho' (37.0%), a small green segment for 'Algo insatisfecho' (9.2%), and a small red segment for 'Muy insatisfecho' (4.2%). The labels for the categories are placed around the chart, and the percentages are displayed within their respective segments.

Satisfaction Level	Percentage
Muy satisfecho	49.6 %
Algo satisfecho	37.0 %
Algo insatisfecho	9.2 %
Muy insatisfecho	4.2 %

Figura 4. Grado de satisfacción sobre soluciones tecnológicas

En este apartado de gráfica, comparamos las respuestas acerca del grado de satisfacción con soluciones tecnológicas implementadas en las tiendas. Podemos observar que casi un 50% de encuestados concuerdan en que están muy satisfechos con lo mismo y solo un bajo porcentaje está muy insatisfecho.

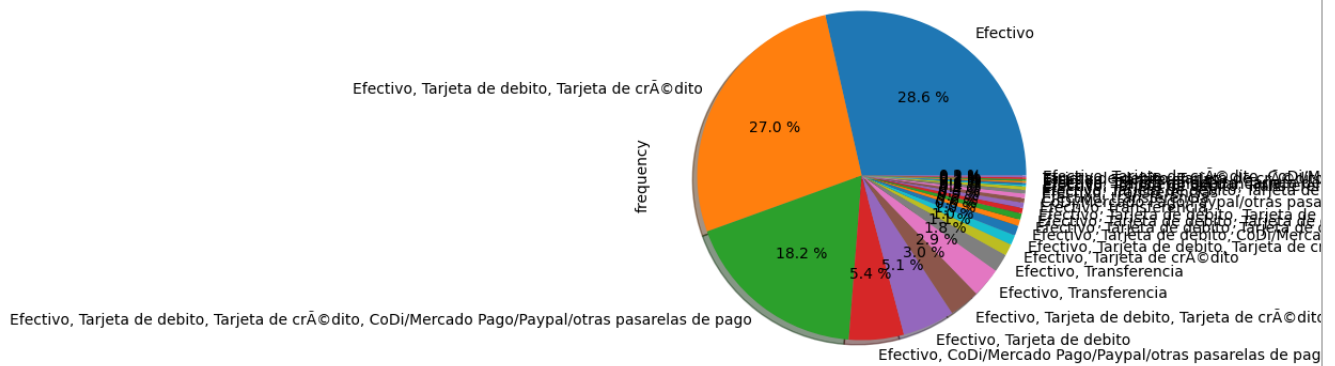


Figura 5. Métodos de pago admitidos

En esta gráfica podemos comparar los diferentes métodos de pago permitidos por las tiendas para el pago de los clientes. Podemos decir que una gran parte admite efectivo, tarjetas de crédito y débito y PayPal.

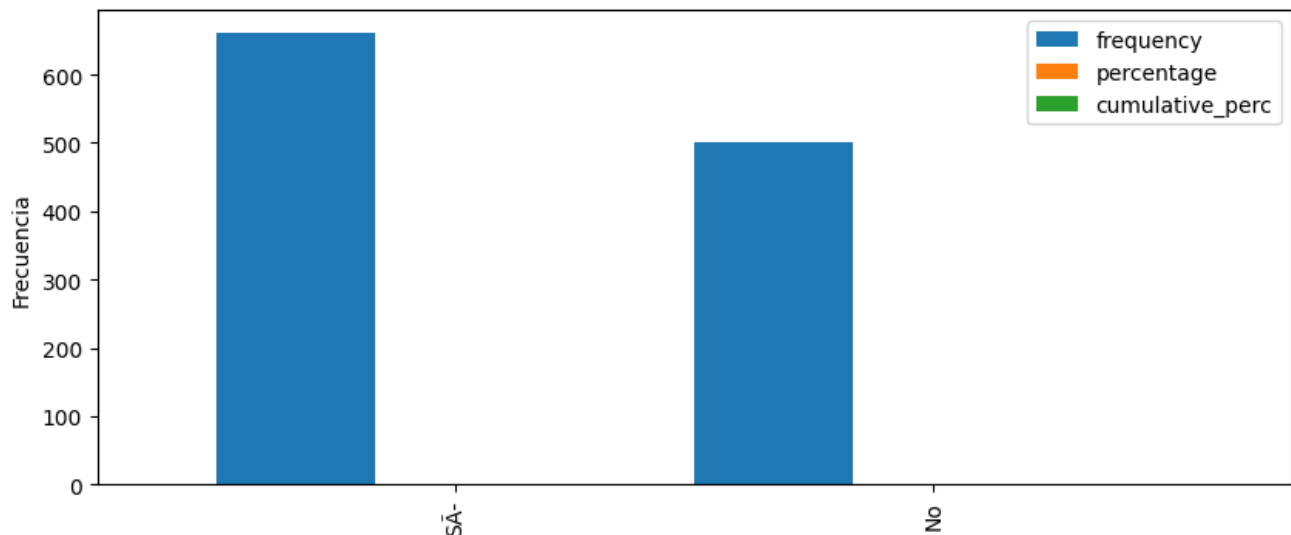


Figura 6. Implementación de tecnología en los últimos 6 meses

En esta gráfica podemos observar que más de 600 encuestados han implementado tecnologías en sus negocios en los últimos 6 meses y 500 no lo han hecho.

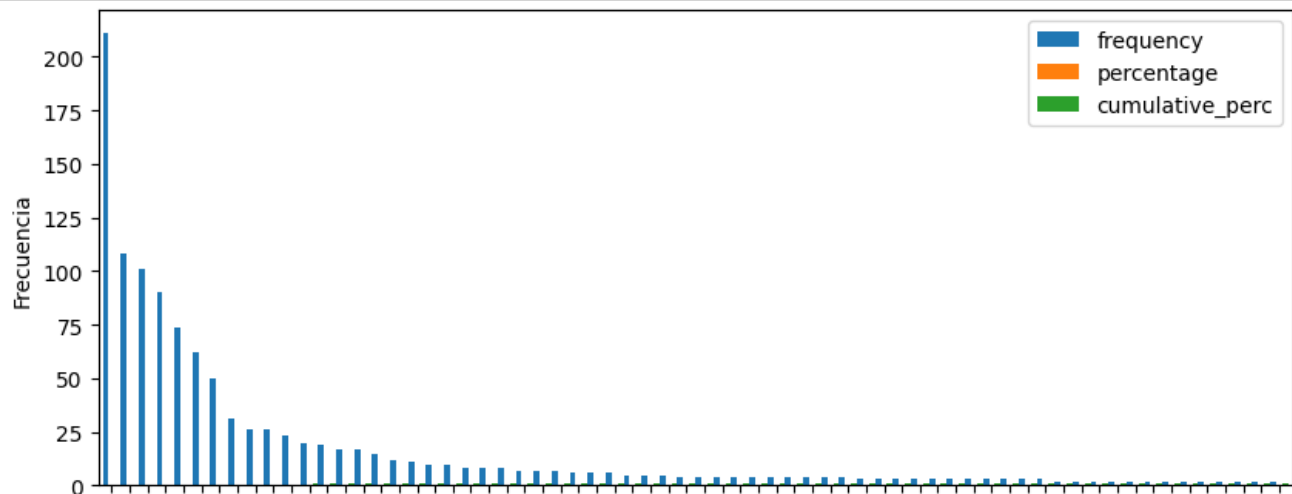


Figura 7. Formas de comunicación con proveedores

Podemos observar en esta gráfica que la mayoría de locales se comunican con sus proveedores por medio de visitas personales (200) y llamadas telefónicas (110).

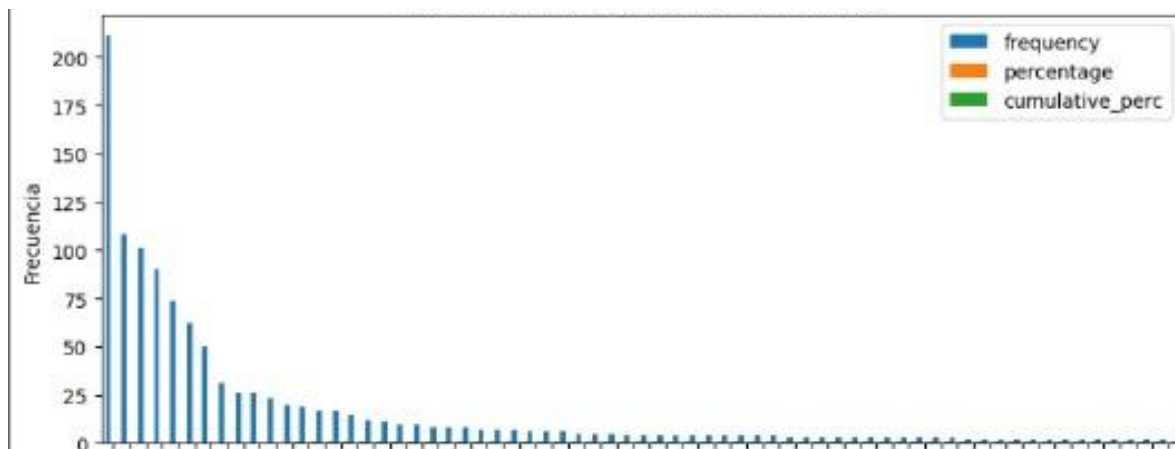


Figura 8. Factores para adoptar tecnologías

Podemos observar que en esta gráfica la mayoría de locales lo hicieron para aumentar las ventas

Resultados y conclusiones de preparación de datos

Los resultados obtenidos en la preparación de datos fueron diversos, ya que muchos negocios han empezado a implementar la tecnología en sus locales como forma de evolución hacia un mercado más digitalizado. De igual manera, varios aceptan métodos de pago de dicha índole para facilitar la experiencia del cliente y ganar más ventas debido a que mucha gente no maneja efectivo y ahora solo tiene tarjeta. Aún así, estos micro negocios siguen recibiendo sus visitas por medio personal y también se comunican con sus proveedores por el mismo, siendo una gran mayoría los locales de comida como pequeños restaurantes o fondas.

Los resultados de la preparación de datos muestran que los micro negocios en México están adoptando la tecnología, pero a su propio ritmo. La mayoría de los negocios han implementado métodos de pago digital, pero siguen recibiendo visitas y comunicándose con sus proveedores de manera personal. Este resultado es comprensible, ya que los micro negocios suelen ser de propiedad familiar y tienen recursos limitados. La adopción de la tecnología puede ser costosa y requerir tiempo y esfuerzo, además de la existente preocupación por perder el contacto personal con sus clientes y proveedores.

ENTREGABLE 3. GENERACIÓN DE LOS MODELOS DE ANÁLISIS DE DATOS

NOMBRE DEL PROYECTO: GENERACIÓN DE LOS MODELOS DE ANÁLISIS DE DATOS

EQUIPO DEL PROYECTO Y
CORREO ELECTRÓNICO:

Rodrigo Jalil Aguilar Martínez
A01425281@tec.mx
José Carlos Trueba Martinez
A01733813@tec.mx
Emilio Arturo Kayary Ramos
A01552526@tec.mx
Pablo Moreno Cagigas
a01737838@tec.mx
Germán Reyes Hernández
A01737467@tec.mx

PROPÓSITO DEL PROYECTO

Generar los modelos estadísticos que nos permitan predecir los precios de renta en Airbnb y poder tomar decisiones para resolver el reto.

Metodología del análisis de datos

1. Limpieza de Datos: Realiza una limpieza inicial de los datos para abordar valores atípicos y valores nulos.
2. Analizar correlación de datos: Seleccionar características relevantes para el análisis.
3. Exposición de datos: Implementar modelos de regresión lineal simple y múltiple.

Modelo inicial de regresión

Modelo Lineal 1: $y = 0.00036257x + 50.620807029$

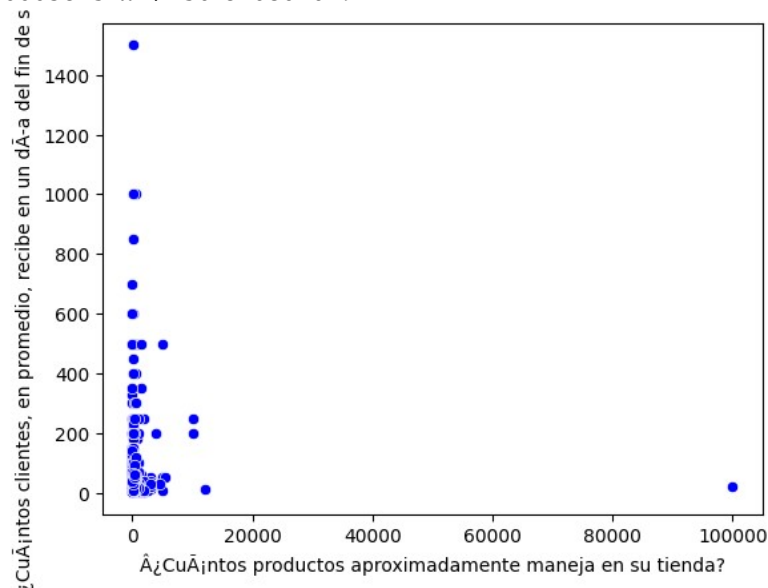


Figura 9. Gráfico de correlación

El modelo presenta un valor de correlación muy bajo, esto quiere decir que las variables no tienen casi nada en común y esto no es óptimo para realizar una predicción del número de clientes que llegarán en dichos días.

Análisis numérico y gráficos de los residuales

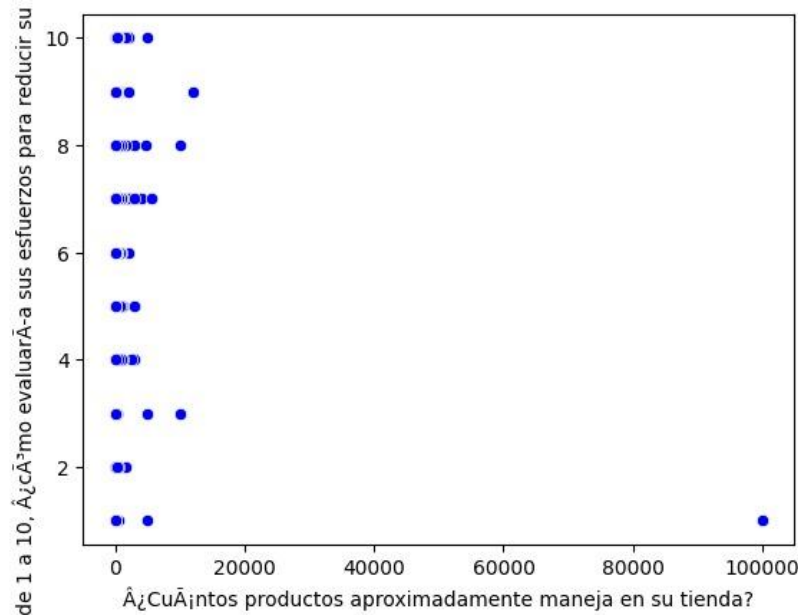


Figura 10. Gráfico de correlación

Los puntos están bastante dispersos, esto significa que el residuo es mayor y en general, con las variables que escogimos en el filtro no son muy útiles para realizar análisis de regresión, sin embargo obtuvimos 3 modelos diferentes que se pueden ver en el código anexo y de estos, el mejor es el de la Figura 10.

Análisis de los índices de correlación y potencia

Correlación Modelo Lineal 1: 0.01093499288982266
 Correlación Modelo Lineal 2: 0.027706000827707068
 Correlación Modelo Lineal 3: 0.06693151708350578

Al analizar estos índices de correlación, se puede concluir que el Modelo Lineal 3 muestra la correlación más alta entre los tres, ya que tiene un valor de 0.06693151708350578. Aunque es la correlación más fuerte de los tres modelos, aún es relativamente baja y sugiere que la relación lineal entre las variables en el Modelo Lineal 3 también es débil. Ninguno de los modelos presenta una correlación lo suficientemente fuerte como para considerarse un buen ajuste, lo que indica que la capacidad predictiva de cada modelo puede ser limitada debido a la débil relación lineal entre las variables.

Análisis de normalidad de los datos

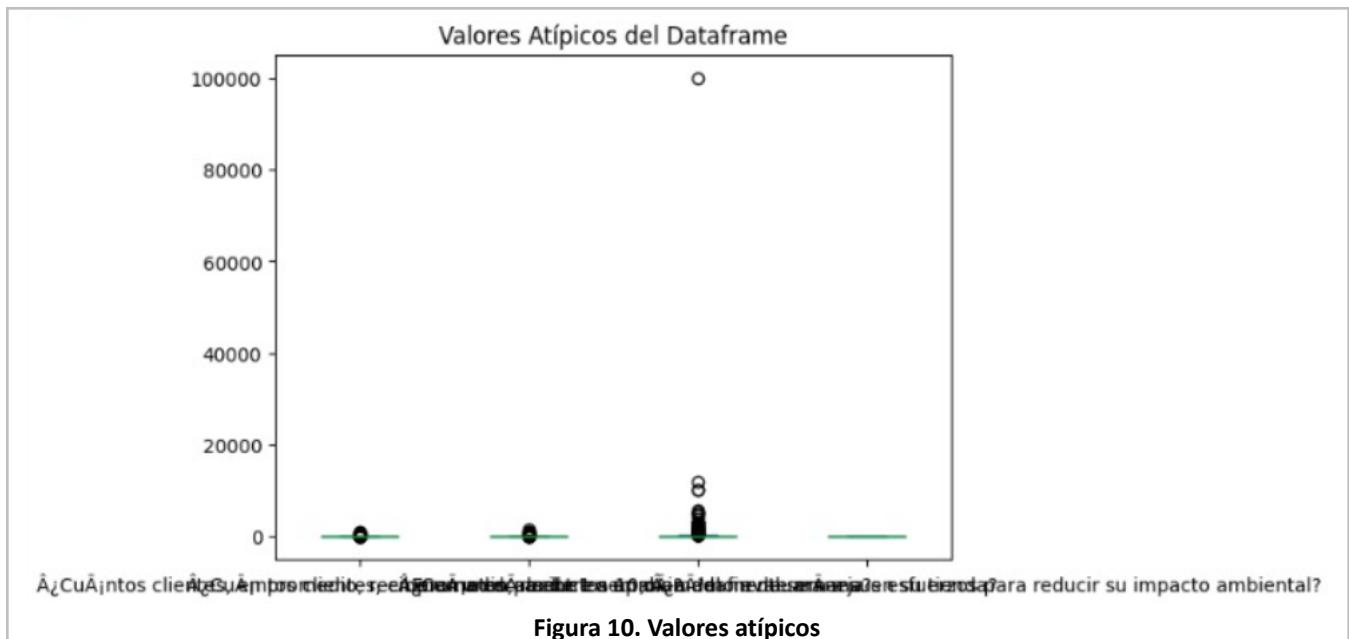


Figura 10. Valores atípicos

Optimización del modelo. Descripción e interpretación

El modelo más óptimo es el tercer modelo que se realizó en el código el cual muestra la relación entre los productos que se tienen en los negocios y el impacto ambiental que producen. Es el modelo con mejor R cuadrada y es el más indicado para realizar predicciones de si se tiene cierto número de productos, cuál será el impacto ambiental a futuro.

Conclusiones finales

El análisis de los proyectos presentados en este artículo muestra la convergencia de la necesidad general de innovación en los sectores comerciales tradicionales, especialmente en el canal minorista mexicano. Project Charter destaca por su enfoque en las microempresas de Puebla, reconociendo la importancia de comprender las complejidades y limitaciones de este sector. La recopilación y análisis de datos se considera una herramienta esencial para identificar áreas de oportunidad, adaptando así soluciones específicas y estrategias de innovación para mejorar la competitividad de estas.

La preparación de datos para el Proyecto de Consejería Estudiantil se presenta como un paso crítico en el proceso de análisis utilizando una metodología mixta que incluye tanto la recopilación de datos primarios como el uso de métodos estadísticos. Los resultados visualizados en el diagrama muestran la tendencia de adopción gradual de tecnología por parte de las microempresas, demostrando una respuesta pragmática a los desafíos modernos sin perder de vista las prácticas comerciales tradicionales.

El campo de la generación de modelos de análisis de datos destaca la importancia de la evaluación crítica de la capacidad predictiva de los modelos implementados. Aunque se han implementado modelos de regresión lineal, los análisis de correlación muestran relaciones débiles entre las variables, lo que sugiere que es necesario mejorar la precisión y validez de dichos modelos.

Lecciones aprendidas

Rodrigo Jalil: Aprendí acerca del manejo de bases de datos, en lo personal encuentro muy útil el uso de estas herramientas digitales para el análisis y la planeación de proyectos; la lección más importante para mí fue el procesamiento de los datos y el análisis para realizar predicciones. Estoy seguro de que no solo a lo largo de mi carrera, sino en la vida profesional seguiré utilizando estas herramientas para optimizar mis proyectos.

José Carlos Trueba Martínez: Lo que aprendí en esta unidad de formación fue lo que representa tener en orden y el trabajo que cuesta realizar correctamente una base de datos, y que principalmente muchas veces se cometen errores al

momento de alimentarlas y que corregir esos errores muchas veces es muy tedioso, pero funciona para futuros proyectos y así ya saber cómo manipularlos.

Germán Reyes Hernández: A través del análisis de proyectos, se destaca la importancia de adaptar soluciones y estrategias de innovación a las complejidades del comercio minorista, especialmente en microempresas de Puebla. La preparación de datos, utilizando métodos estadísticos y recopilación de datos primarios, emerge como un paso crítico en el análisis, revelando la tendencia de adopción gradual de tecnología por microempresas. En el campo de generación de modelos de análisis de datos, se aprende la necesidad de una evaluación crítica de la capacidad predictiva de los modelos implementados, reconociendo que, aunque se emplean modelos de regresión lineal, la mejora continua es esencial para la precisión y validez.

Emilio Arturo Kayary Ramos A partir de la experiencia en proyectos analíticos, se comprende que el enfoque centrado en microempresas permite una comprensión más profunda de las dinámicas comerciales locales. La metodología mixta para la preparación de datos en el Proyecto de Consejería Estudiantil resalta la necesidad de abordar la complejidad mediante la combinación de datos primarios y métodos estadísticos. Además, la debilidad en las relaciones identificadas por los análisis de correlación subraya la importancia de evolucionar continuamente los modelos de regresión lineal para mejorar su precisión y validez, demostrando que la adaptabilidad y la mejora constante son cruciales en el análisis de datos y la toma de decisiones estratégicas.

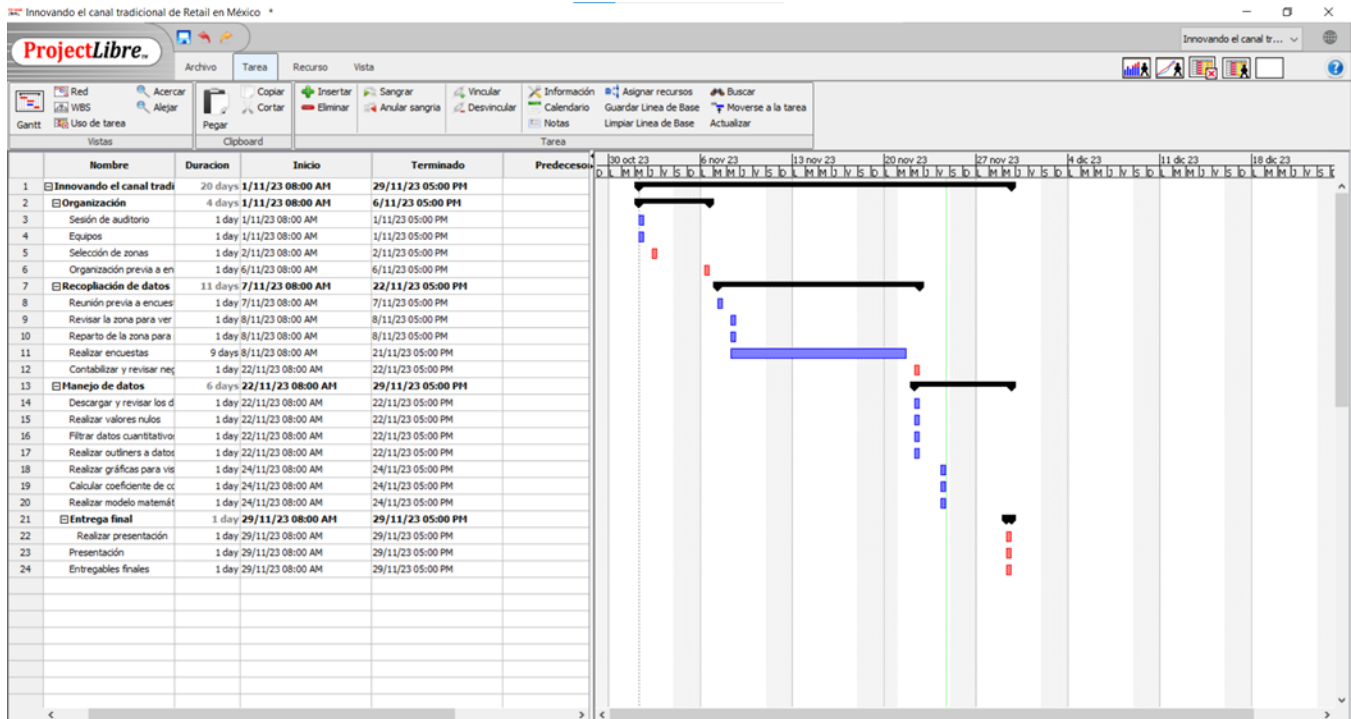
Pablo Moreno Cagigas: El análisis de proyectos destaca la importancia crítica de entender los desafíos específicos en el comercio minorista mexicano, especialmente en microempresas. La recopilación y análisis de datos son herramientas esenciales para la innovación adaptada, y la mejora continua de modelos de análisis es crucial para abordar relaciones débiles identificadas.

REFERENCIAS

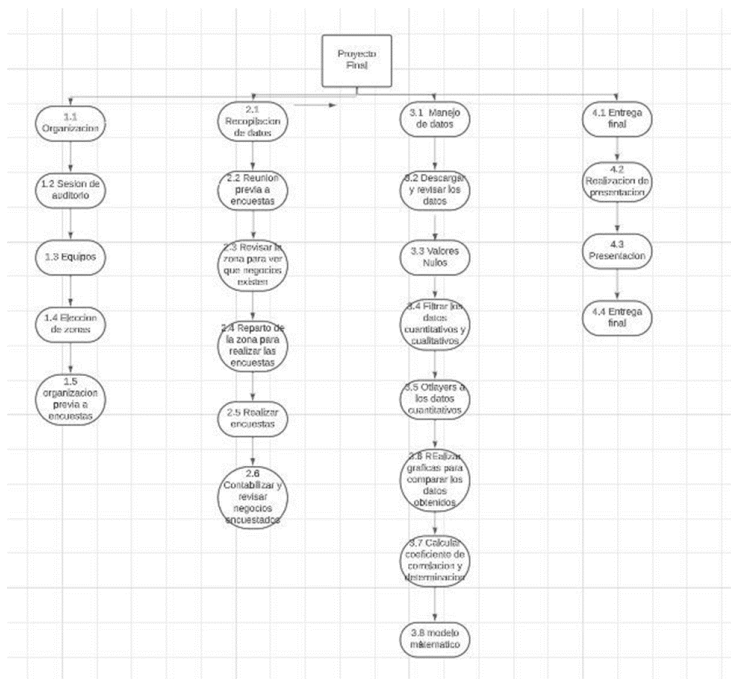
conocimientos aprendidos en la clase de “Desarrollo de proyectos de análisis de datos”, tec de monterrey campus puebla, periodo otoño invierno 2023, por los profesores: Gabriela Guadalupe Reyes Zárte y Alfredo García Suárez

ANEXOS

ProjectLibre



EDT



REPARTO DE RESPONSABILIDADES

	A	B	C	D	E	F	G	H	I
1	INNOVANDO EL CANAL TRADICIONAL DE RETAIL EN MÉXICO	EMILIO	JOSE CARLOS	GERMÁN	PABLO	RODRIGO			
2	Organización								
3	Sesión de auditorio	P	P	P	P	P			
4	Equipos	P	P	P	P	P			
5	Selección de zonas	P			A				
6	Organización previa a las encuestas		P	A					
7	Recopilación de datos								
8	Reunión previa a las encuestas	P	A	A	A	P			
9	Revisar la zona		P	P		A			
10	Reparto de zona	A	P	A	A	A			
11	Realizar encuestas	P	P	P	P	P			
12	Contabilizar y revisar encuestas		A	P					
13	Manejo de datos								
14	Descargar y revisar los datos	A				P			
15	Realizar valores nulos	P	A	A	A	P			
16	Realizar outliners a datos cuantitativos	P	A	A	A	P			
17	Realizar gráficas para visualizar resultados	A	P						
18	Calcular coeficiente de correlación y determinación			A	A	P			
19	Realizar modelo matemático	P	A	A	A	P			
20	Entrega final								
21	Realizar presentación	P	P	P	P	P			
22	Presentación	P	P	P	P	P			
23	Entregables finales	P	P	P	P	P			

Simbología	
P	Principal
A	Ayuda