Análisis Exploratorio de Datos:
Impacto del uso de redes sociales en la productividad laboral

Aprendices Grupo TAD 10

Instructor

ROYMER ROMERO ALGARÍN

TL PROGRAMACIÓN PARA LA ANALÍTICA DE DATOS

OCTUBRE DE 2025

TABLA DE CONTENIDO

INTRODUCCIÓN

El presente proyecto se enfoca en la aplicación del Análisis Exploratorio de Datos, integrado en el proceso de Análisis de datos del dataset “social-media-vs-productivity”, con el fin de aplicar lo aprendido acerca del proceso de limpieza de datos, su validación y visualización, para la comprensión profunda de las variables subyacentes que faciliten la comunicación de hallazgos y una toma de decisiones informada.

OBJETIVO GENERAL

Realizar el proceso de Análisis de Datos y Análisis Exploratorio de Datos que permitan encontrar la posible correlación y el impacto que tienen los patrones de uso de las redes sociales sobre la productividad laboral, de acuerdo con el conjunto de datos y las herramientas seleccionadas.

OBJETIVOS ESPECÍFICOS

- Realizar el proceso de Análisis de Datos utilizando herramientas de Power BI y librerías de Python para Análisis de Datos.
- Aplicar técnicas del Análisis Exploratorio de Datos para descubrir patrones, identificar anomalías y formar posibles hipótesis.
- Socializar los resultados del proceso de Análisis de Datos y el EDA.

COMPETENCIA

Integración de datos según técnicas de visualización y metodologías de análisis.

RESULTADOS DE APRENDIZAJE

RA2. Realizar el proceso de limpieza de datos de acuerdo con la herramienta informática seleccionada.

RA3. Validar la data de acuerdo con el proceso análisis de exploración de datos.

METODOLOGÍA

Herramientas / Tecnologías: Se utilizaron las herramientas Power BI Desktop (incluye Power BI, Power Pivot, Power Query y el lenguaje DAX), las librerías Pandas, Matplotlib y Seaborn de Python para la analítica de datos, Kaggle como repositorio del dataset, Google Colab como Cuaderno Interactivo.

Trabajo en Equipo: Los aprendices fueron distribuidos en los 4 grupos que corresponden con cada una de las fases del proceso de Análisis de Datos. En cada grupo un aprendiz tiene el rol de líder que gestiona el rendimiento y participación de cada integrante para cumplir con el cronograma propuesto. La salida de una fase es el insumo de entrada para la siguiente.

Cronograma: El proyecto tuvo una duración de X días de acuerdo con el cronograma propuesto.

Etapas para el Análisis: Para realizar el proceso de Análisis de Datos del dataset objetivo, se implementaron las siguientes fases del Análisis de Datos, incluyendo el Análisis Exploratorio de Datos:

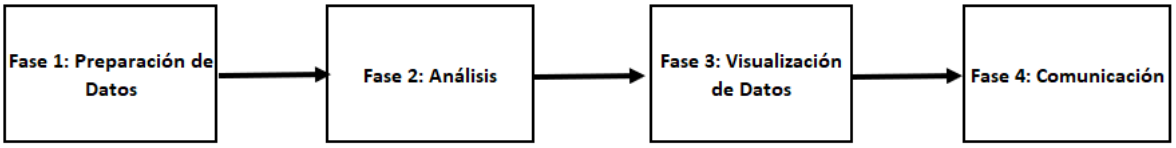


Figura 1. Fases Análisis Exploratorio de Datos (Elaboración propia)

A continuación, se describen las actividades realizadas:

Fase 1. Preparación de datos

- **Comprensión del Negocio (Business Understanding):** Se parte de la comprensión y análisis de las preguntas propuestas, además de las que surjan del análisis y comprensión inicial de los datos.
- **Preparación y Limpieza de Datos:**
    - Carga del Dataset.

- Tipos de Datos: Se verifica que el tipo de datos de cada columna sea el correcto (numérico, categórico, etc.).
- Duplicados: Se revisa la consistencia de los formatos y se eliminan las filas duplicadas.
- Identificación y tratamiento de Datos Faltantes: Se detectan valores nulos o incompletos. Si los hay, se decide cómo tratarlos (eliminarlos, imputarlos, etc.).
- Detección de outliers o valores atípicos.
- Se utilizaron las tecnologías Power Query, Power Pivot y Pandas.

Fase 2. Análisis

- **Análisis Unidireccional:** Se examina una variable a la vez. Esto incluye calcular la media, la mediana y la moda para variables numéricas, y la frecuencia de cada categoría para las variables categóricas.
- **Análisis Multidireccional:** Se examinan las relaciones entre dos o más variables. Se usa para encontrar correlaciones, tendencias y patrones. De igual forma, retroalimentar o complementar las imputaciones realizadas en el proceso de limpieza.
- **Modelado:** Después de la preparación, limpieza y análisis de datos, se llevará a cabo el modelad, relaciones, medidas y columnas calculadas con Power Pivot.
- Se utilizaron las tecnologías Power Query, Power Pivot, Pandas, Matplotlib y Seaborn.

Fase 3. Visualización

La visualización es muy importante para el EDA, porque ayuda a revelar patrones o problemas que los números por sí solos no muestran. Se implementarán diagramas para las variables principales, explicando cómo se usaron para identificar valores atípicos.

Para apoyar el proceso de comunicación, se desarrollará un informe paginado utilizando la herramienta Power BI Desktop y gráficas con Seaborn y Matplotlib de Python.

Fase 4. Comunicación

En esta etapa se dan a conocer los hallazgos y respuestas a las preguntas, aplicando la técnica de Storytelling: contando una historia con los datos. En conjunto con el dashboard diseñado con Power BI y los elementos gráficos con las librerías de Python para Análisis de Datos.

RESULTADOS

Para cada una de las fases anteriores, se describen a continuación, las acciones realizadas:

Fase 1. Preparación de datos

Comprensión del Negocio (Business Understanding): Inicialmente se plantean las siguientes preguntas para su respuesta por medio del análisis del dataset, no obstante surgirán otras como producto de las posibles relaciones entre los datos, tendencias, patrones, entre otros:

1. ¿Cuál es la diferencia promedio entre la productividad que las personas creen tener y la que realmente demuestran?

2. ¿Cómo varía el promedio de productividad real según el tiempo de uso diario de redes sociales?

3. ¿Cómo se relacionan las horas de sueño y los niveles de estrés con la productividad y el agotamiento laboral?

4. ¿Existen diferencias significativas en la productividad promedio y el nivel de agotamiento entre hombres y mujeres, o entre distintos tipos de trabajo?

Preparación y Limpieza de Datos

- **Carga del Dataset:** Se utilizó el conjunto de datos llamado social-media-vs-productivity en formato CSV (Comma Separed Values), disponible en el repositorio de Kaggle (https://www.kaggle.com/datasets/mahdimashayekhi/social-media-vs-productivity).

  De manera inicial se ofrece información sobre los datos con relación al significado y número de las columnas de datos, el número de registros y calidad inicial.

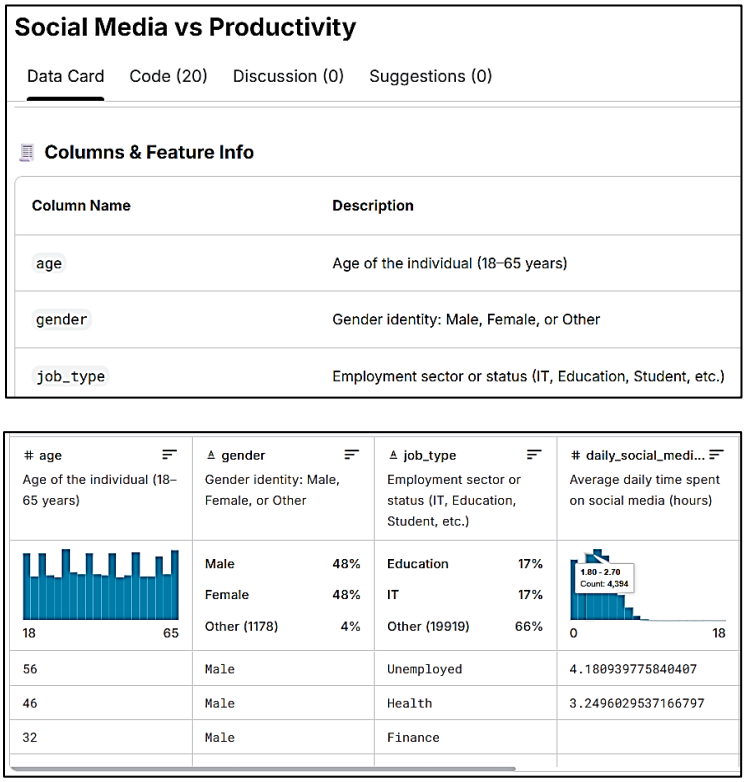


Figura 2. Información preliminar del Dataset Social Media vs Productivity

- **Tipos de Datos:** No se encontraron novedades en el tipo de dato de las columnas.
- **Duplicados:** No se encontraron registros duplicados.
- **Estandarización de nombres de columnas y formato de variables categóricas:** Se realizó la traducción de los nombres de las columnas al idioma español y el redondeo de la cantidad de números decimales de Horas_Diarias_En_Redes, Horas_Diarias_Trabajadas, Uso_Antes_De_Dormir, Puntuacion_Productividad_Autoevaluada, Puntuacion_De_Satisfaccion_Laboral, Puntuacion_Productividad_Real, Horas_De_Sueño, Horas_Semanales_Fuera_De_Linea.
- **Identificación de datos faltantes y outliers:** El objetivo de esta actividad consiste en la la imputación de los datos faltantes de acuerdo con su tipo. Para lograrlo, se analizaron las relaciones entre las columnas del conjunto de datos por medio de la observación

directa y la creación de un mapa de calor de la matriz de correlaciones. A continuación, se indican los datos faltantes del dataset:

COLUMNAS	CANTIDAD NULOS
HORA_DIARIA_EN_REDES	2765
PUNTUACION_PRODUCTIVIDAD_AUTOEVALUADA	1614
PUNTUACION_PRODUCTIVIDAD_REAL	2365
NIVEL_ESTRES	1904
HORAS_DE_SUEÑO	2598
USO_ANTES_DE_DORMIR	2211
PUNTUACION_DE_SATISFACCION_LABORAL	2730

Tabla 1. Resumen de datos faltantes generado con la librería Pandas

- **Tratamiento de los datos faltantes:**

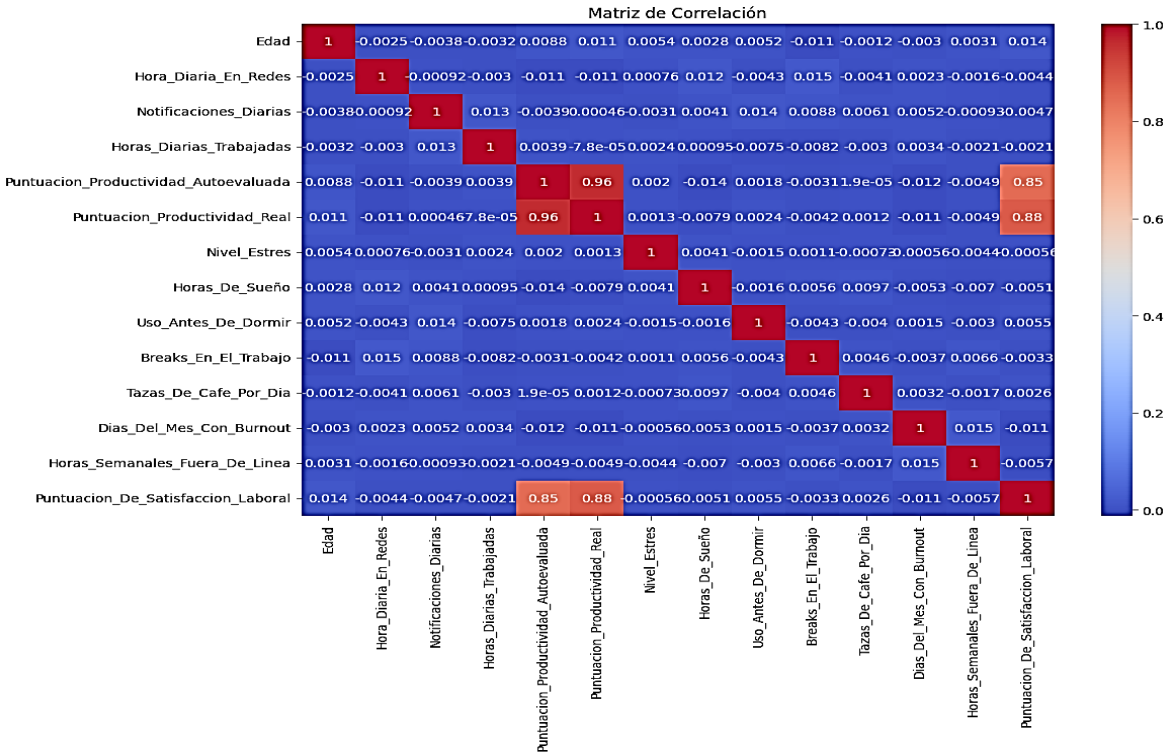


Figura 3. Mapa de calor de la matriz de correlación generada con Pandas

Para guiar la imputación de datos, se analizó de forma manual la existencia de relaciones entre las columnas del dataset. Después se generó una matriz de correlación utilizando la librería Pandas en Google Colab, y se visualizó con un mapa de calor (ver Figura 2), permitiendo la identificación de las variables con correlación alta, baja o nula.

Al interpretar el gráfico, se confirmó que existían pocas correlaciones significativas entre las columnas. Esta baja interdependencia sugiere que la mayoría de los valores ausentes son del tipo **MCAR (Missing Completely At Random) o Faltantes Completamente al Azar.**

Por lo anterior, la imputación de los datos nulos se simplificó y todos los valores ausentes en las columnas identificadas fueron imputados utilizando la **media aritmética** de la columna correspondiente, la cual fue calculada con la función mean() de la librería **Pandas**; como se identifican en la siguiente tabla:

COLUMNAS	PROMEDIO
HORA_DIARIA_EN_REDES	3.11
NIVEL_ESTRES	5.51
HORAS_DE_SUEÑO	6.50
USO_ANTES_DE_DORMIR	1.02

Tabla 2. Columnas imputadas con valores faltantes imputados con la media

La **excepción** a esta regla fueron las siguientes columnas, que mostraron tener una **correlación positiva** entre sí:

Puntuacion_Productividad_Real: Puntuación de productividad simulada (escala: 0-10)

Puntuacion_Productividad_Autoevaluada: Puntuación de productividad autoevaluada (escala: 0-10)

Puntuacion_De_Satisfaccion_Laboral: Satisfacción con el trabajo/responsabilidades de vida (escala: 0-10)

Se identificó una relación directa entre la **puntuación real** y la **autoevaluada** de productividad. Para cuantificar esta relación, se implementó una estrategia de imputación en dos fases:

1. Cuantificación e Imputación por Desviación con Power Query

Se creó una columna calculada en **Power Query** para calcular la **diferencia** entre la productividad percibida y la real.

La mayoría de los datos arrojó una diferencia constante, oscilando entre **0.5 y 1.0**. Se calculó el promedio de esta diferencia, resultando en **0.55**.

Este promedio se utilizó para crear una **columna personalizada** que imputó los valores **nulos** en la **Puntuacion_Productividad_Autoevaluada**: si la puntuación autoevaluada

estaba ausente, se igualó a la **Puntuacion_Productividad_Real más 0.55** (el promedio de diferencia).

2. Imputación Avanzada con Regresión Lineal con Google Colab

Pese a la imputación anterior, persistían **129 registros** con datos faltantes en **ambas** columnas de puntuaciones.

Para tratar estos casos sin comprometer la relación entre las columnas, se utilizó **Google Colab** para implementar una **regresión lineal** utilizando la librería Pandas como marco de datos. Este modelo fue empleado para **imputar los datos nulos restantes**, preservando la correlación observada entre las columnas.

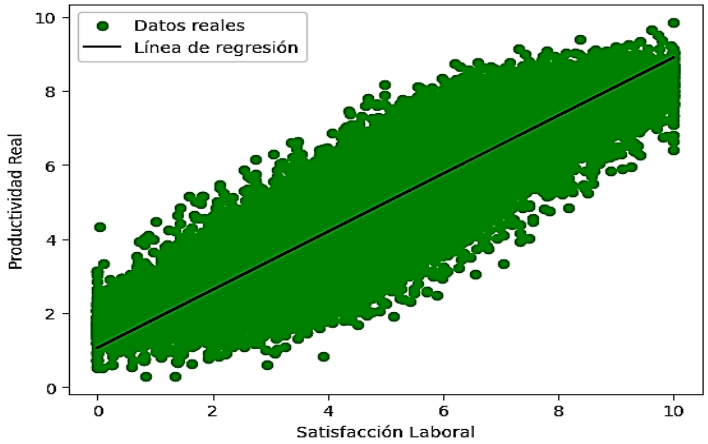


Figura 4. Regresión lineal calculada para imputar datos faltantes.

- **Deteccion de outliers**

Los valores atípicos (*outliers*) se identificaron en las columnas Horas_Trabajadas_Por_Dia, utilizando el Diagrama de Cajas y Bigotes (*Boxplot*) generado con la librería Pandas, lo cual facilitó su visualización.

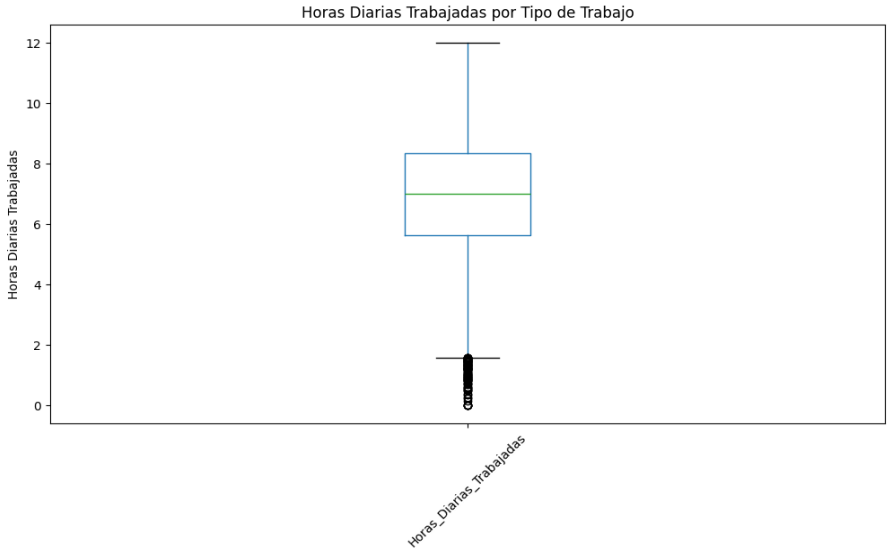


Figura 5. Diagrama de Caja de la columna Horas_Trabajadas_Por_Dia

El diagrama de cajas (**boxplot**) confirmó la existencia de **95 valores atípicos** (*outliers*) en la columna **Horas_Trabajadas_Por_Dia**, los cuales se encontraban **por debajo de 1.55 horas** trabajadas en promedio. Para corregir y minimizar la distorsión estadística causada por estos valores extremos, los 95 registros atípicos fueron imputados utilizando la **media aritmética** de la columna, que es de **6.99 horas** promedio.

Fase 2. Análisis

- **Análisis Multidireccional**

Mediante un análisis con Google Colab, se intentó encontrar qué otras relaciones podrían existir para realizar las medidas de las preguntas planteadas.
El análisis se centró en segmentar a los empleados basándose en sus horas diarias dedicadas a las redes sociales y sus horas diarias trabajadas, utilizando el algoritmo de clustering K-Means. Los pasos principales fueron:

Carga y exploración de datos: Se cargó la base de datos, se verificó la información general y la presencia de valores nulos, y se obtuvieron estadísticas descriptivas.

Preparación de datos: Se seleccionaron las variables relevantes para el clustering (Horas_Diarias_En_Redes ' y 'Horas_Diarias_Trabajadas') y se escalaron para asegurar que tuvieran una influencia similar en el algoritmo.

Determinación del número de clusters (K): Se utilizó el método del codo para identificar el número óptimo de clústeres para K-Means.

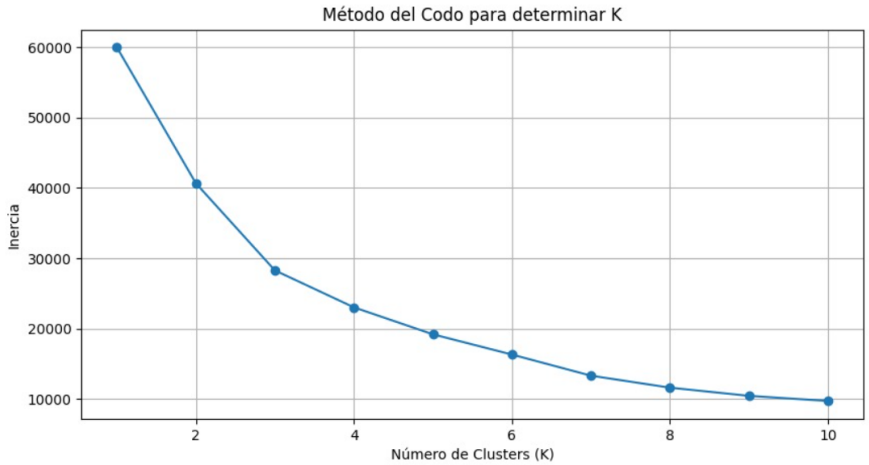


Figura 6. Cálculo de K clusters con Google Colab

Aplicación de K-Means y segmentación: Se aplicó el algoritmo K-Means con el número óptimo de clusters (3 de acuerdo con la Figura 5) a los datos escalados. Se asignó a cada empleado un clúster basado en sus patrones de horas en redes sociales y horas trabajadas, obteniéndose el resultado de la Figura 6.

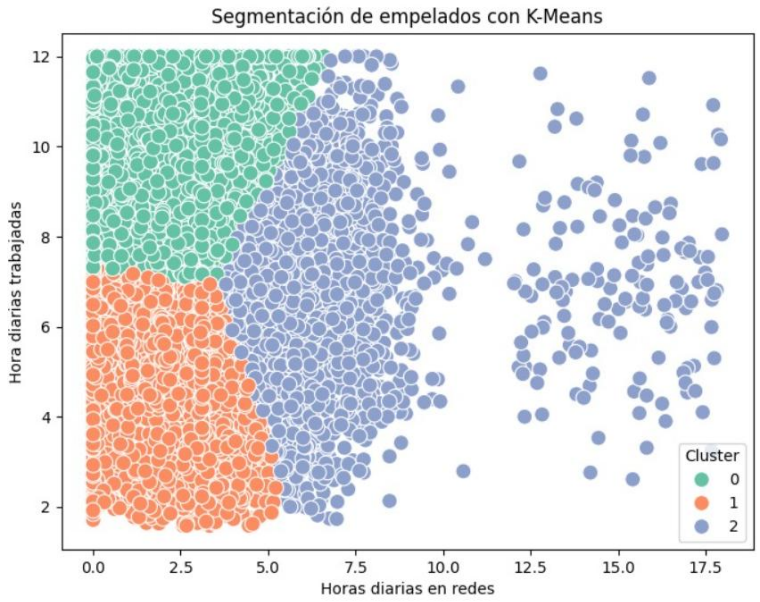


Figura 7. Diagrama de dispersión de segmentación de clusters con Google Colab

Visualización y análisis de clusters: Los clústers se llevaron a un gráfico de dispersión y se obtuvo un resumen de las estadísticas principales (media, mediana, desviación estándar, mínimo y máximo) para cada clúster, lo que permite entender las características de cada segmento de empleados.

Guardar resultados: Se guardó el DataFrame con la asignación de clústers en un nuevo archivo Excel.

Resultado del Análisis de correlación: Se exploraron las relaciones entre las variables numéricas mediante gráficos pairplot y un nuevo mapa de calor de la matriz de correlación (Figura 7).

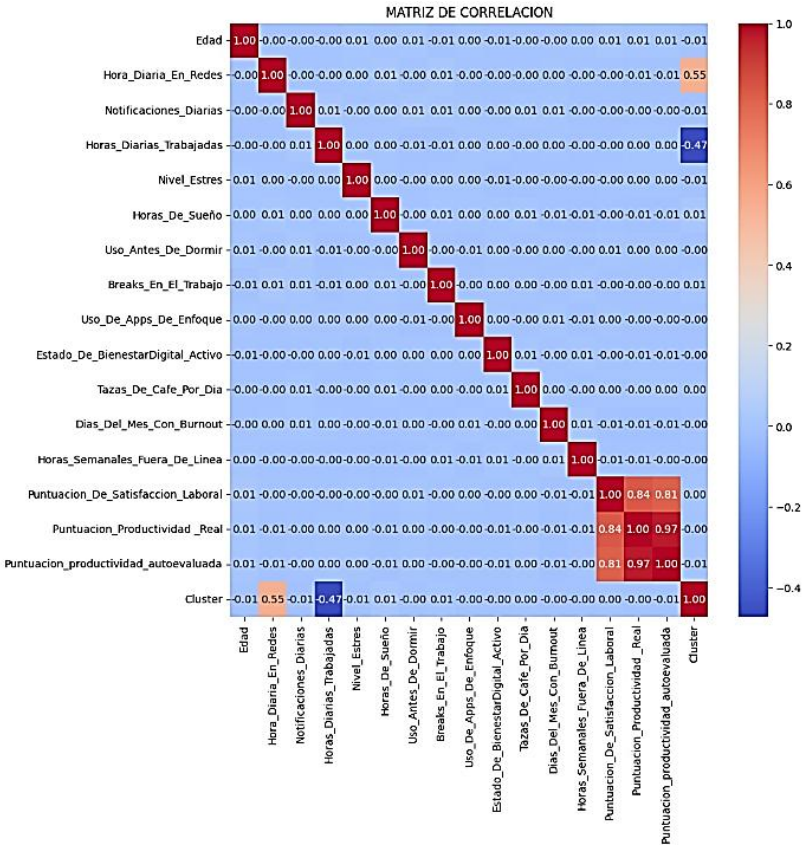


Figura 8. Mapa de calor de correlación de Clusters creados.

Se realizó un análisis de segmentación utilizando K-Means para identificar grupos distintos de empleados basados en su uso de redes sociales y horas de trabajo, y se exploraron las características de estos grupos.

Con base a esto y los análisis previos, se crearon las medidas para dar contexto y respuesta a las preguntas presentadas. Las medidas se dividieron en:

Promedios: Se realizaron medidas de promedio, entre estas se encuentran promedios calculados para columnas cómo Horas_De_Sueño, Horas_Diarias_Trabajadas, Puntuacion_De_Satisfaccion_Laboral, entre otros; que contribuyeron a contemplar la información necesaria para dar una respuesta a las preguntas.

Conteos: Se realizaron medidas de conteo para conocer la cantidad de usuarios en columnas específicas, por ejemplo, en columna "Género" el conteo permitió identificar tres tipos a saber: Hombre, Mujer u Otros, Específicamente, la columna "Género" se usó para contar el número exacto de usuarios en cada una de sus tres categorías. Este mismo método se empleó con la "App de enfoque" y el "Estado de bienestar digital" para verificar su uso entre los participantes.

Porcentajes: Se realizaron medidas para el cálculo de tasas, cuyo resultado es más evidente a nivel visual (dashboard), generando un contexto de filtro que afecta a otras columnas . En este caso las medidas fueron realizadas para las columnas de "Género", "Red_Social_Favorita" y "Tipo_Trabajo"

- **Modelado**

En esta etapa utilizaron las herramientas **Power Pivot y el lenguaje DAX (Data Expression Analysis)** para crear columnas condicionales. La primera columna se llama Clúster, con base al proceso K-MEANS realizado en Google Colab. Y la columna "Consumo_Café", creada con la intención de saber qué tanto puede afectar el consumo de café en los usuarios.

Se crearon las medidas que dan respuesta a las preguntas planteadas, por ejemplo:

La medida "Diferencia promedio productividad", brinda respuesta a la pregunta "¿Cuál es la diferencia promedio entre la productividad que las personas creen tener y la que realmente demuestran?". La fórmula es:

```
diferencia_Promedio_productividad = [Promedio_puntuacion_productiva_autoeluavada]-[Promedio_puntuacion_productiva_real]
```

También se crearon las siguientes medidas para efectuar un análisis cuantitativo de diversas métricas contenidas en la Tabla_de_Hechos. Específicamente, se calcula el valor promedio para cuatro variables distintas:

La medida Promedio_horas_de_sueño determina la media aritmética de las horas de descanso registradas.

```
Promedio_horas_de_sueño = AVERAGE(Tabla_de_Hechos[Horas_De_Sueño])
```

El Promedio_puntuacion_productiva_autoevaluada cuantifica el promedio de la productividad desde una perspectiva subjetiva, según la autoevaluación de los individuos.

```
Promedio_puntuacion_productiva_autoeluavada = AVERAGE(Tabla_de_Hechos[Puntuacion_productividad_autoevaluada])
```

La medida Promedio_puntuacion_productiva_real establece la media de la productividad objetiva, incorporando una operación de suma (+0) para asegurar que un resultado nulo (BLANK()) sea tratado como un valor cero.

```
Promedio_puntuacion_productiva_real = AVERAGE(Tabla_de_Hechos[Puntuacion_Productividad _Real])+0
```

Finalmente, Promedio_Puntuacion_de_la_sastifacion_laboral, calcula el promedio de las puntuaciones de satisfacción laboral.

```
promedio_Puntuacion_de_la_sastifascion_laboral = AVERAGE(Tabla_de_Hechos[Puntuacion_De_Satisfaccion_Laboral])
```

En su totalidad, estas medidas constituyen indicadores clave de rendimiento (KPIs) que permiten la evaluación y correlación sistemática entre el descanso, la satisfacción y la productividad, tanto percibida como objetiva.

Fase 3. Visualización

La fase de visualización representa el punto culminante del análisis, donde los datos procesados se transforman en conocimiento tangible y comprensible. El objetivo principal de esta etapa no fue solo representar gráficamente las respuestas a las preguntas iniciales, sino también explorar las relaciones entre las variables para descubrir patrones ocultos. Precisamente, uno de los hallazgos más significativos de este proceso fue identificar y ratificar la notable ausencia de correlación en casos donde intuitivamente se esperaba encontrar una conexión fuerte, demostrando así el valor de la visualización para desafiar suposiciones y revelar la verdadera dinámica de los datos.

Utilizando la herramienta Power BI, se elaboró un informe paginado con el propósito de dar contexto a los datos y ofrecer información adicional que trasciende el alcance de las preguntas originales.

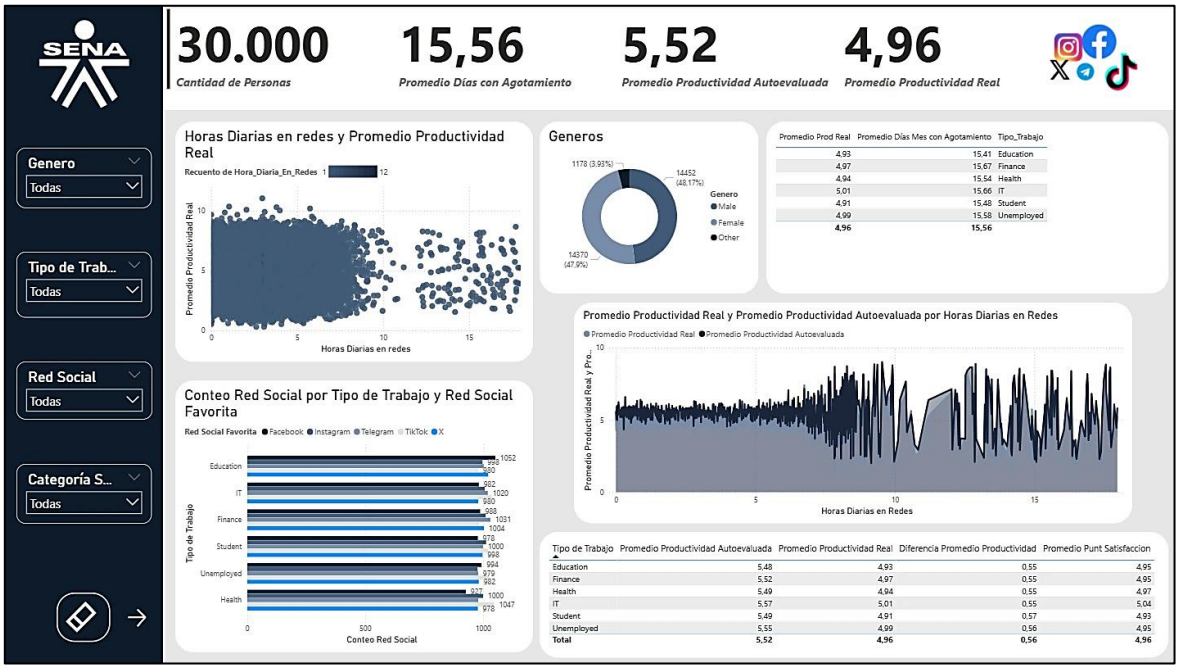
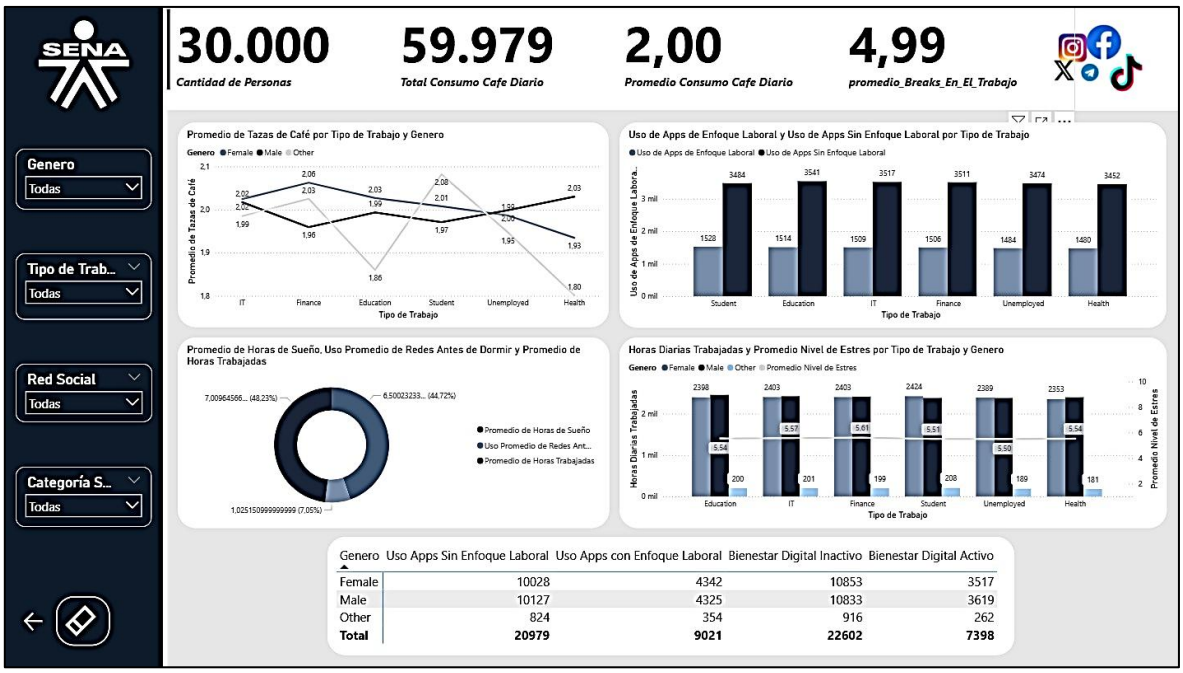


Figura 9. Informe en Power BI – Página 1



Genero	Uso Apps Sin Enfoque Laboral	Uso Apps con Enfoque Laboral	Bienestar Digital Inactivo	Bienestar Digital Activo
Female	10028	4342	10853	3517
Male	10127	4325	10833	3619
Other	824	354	916	262
Total	**20979**	**9021**	**22602**	**7398**

Figura 10. Informe en Power BI – Página 2

La finalidad del dashboard consiste en analizar y visualizar el estilo de vida de una población, conectando sus comportamientos personales y hábitos diarios con su vida profesional.

Su intención es identificar patrones y correlaciones entre estas dos áreas para entender cómo factores como el sueño, el uso de la tecnología y otras rutinas cotidianas se relacionan con el estrés, el bienestar y el rendimiento en el entorno laboral.

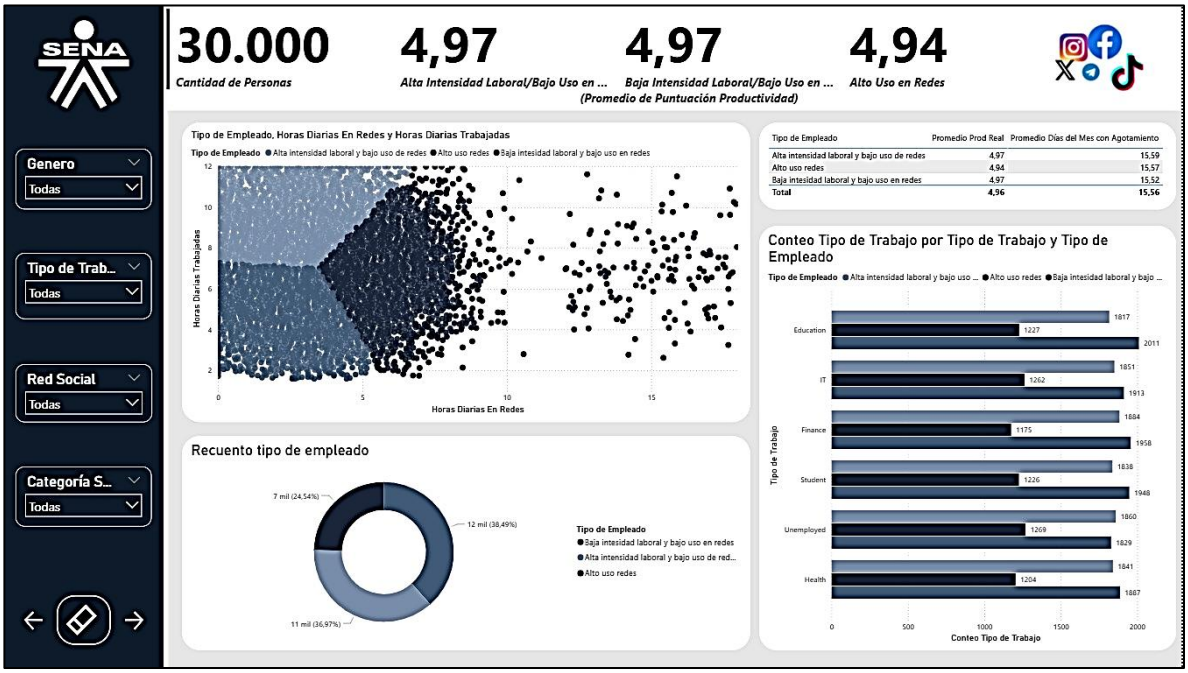


Figura 11. Informe en Power BI – Página 3

El informe paginado en Power BI, visible en la Figura 11, detalla un análisis de clústeres sobre 30,000 individuos. La segmentación, basada en las variables de "Horas Diarias Trabajadas" y "Horas Diarias en Redes", clasifica a los empleados en tres grupos distintos. Como se observa en el gráfico de anillos, el grupo de "Alta intensidad laboral y bajo uso en redes" compone el 38.5% de la muestra, seguido por el de "Baja intensidad laboral y bajo uso en redes" con un 36.9%. El grupo de "Alto uso en redes" representa el 24.6% restante.

El propósito de esta visualización es correlacionar estos perfiles con su "Puntuación de Productividad Real"

Respuesta a las Preguntas Base:

1. ¿Cuál es la diferencia promedio entre la productividad que las personas creen tener y la que realmente demuestran?

Tipo de Trabajo	Promedio Productividad Autoevaluada	Promedio Productividad Real	Diferencia Promedio Productividad	Promedio Punt Satisfaccion
Education	5,48	4,93	0,55	4,95
Finance	5,52	4,97	0,55	4,95
Health	5,49	4,94	0,55	4,97
IT	5,57	5,01	0,55	5,04
Student	5,49	4,91	0,57	4,93
Unemployed	5,55	4,99	0,56	4,95
Total	**5,52**	**4,96**	**0,56**	**4,96**

Tabla 3. Diferencia promedio productividad real y autoevaluada

Un hallazgo clave del análisis es que los trabajadores tienen una percepción bastante precisa de su propia productividad. La diferencia promedio entre el rendimiento real y el autoevaluado es mínima (0.56), lo que confirma que, en general, la productividad que los empleados creen tener es muy similar a la que demuestran.

2. ¿Cómo varía el promedio de productividad real según el tiempo de uso diario de redes Sociales?

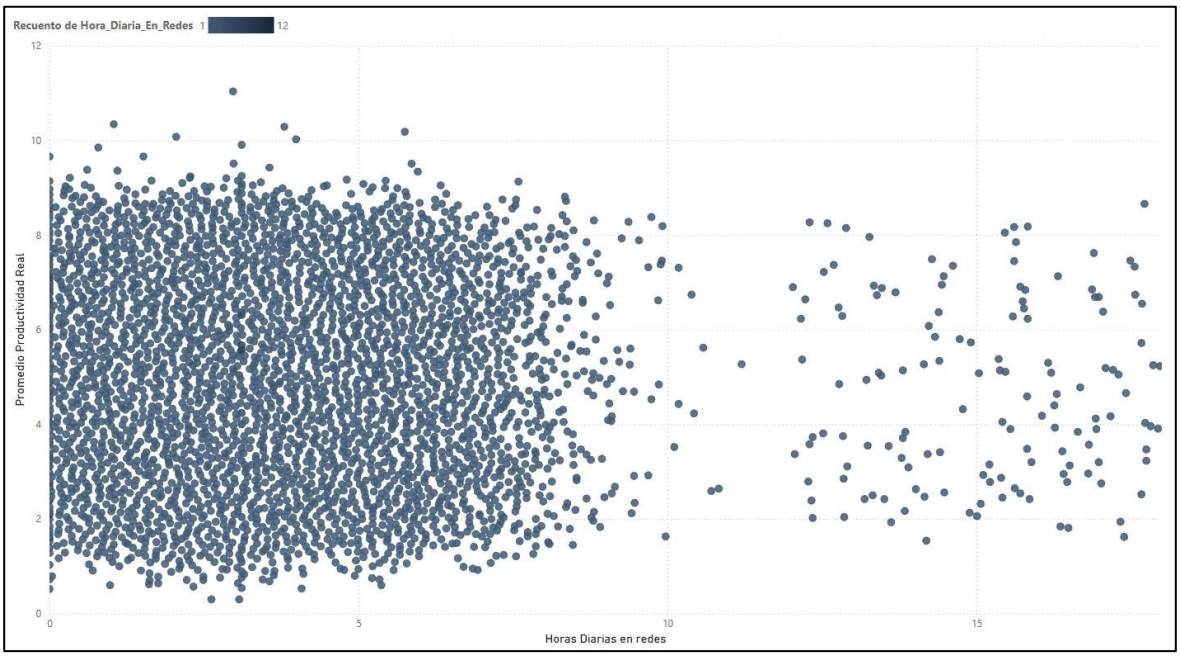


Figura 12. Diagrama de dispersión variación productividad media

El análisis del gráfico de indica que no existe una correlación evidente entre el tiempo de uso diario de redes sociales y el promedio de productividad real.

La distribución de los datos muestra una alta variabilidad, lo que significa que, para un mismo número de horas invertidas en redes sociales, se observan individuos con niveles de productividad tanto muy altos como muy bajos. Esta falta de una tendencia definida sugiere que el tiempo de uso no es un factor determinante del rendimiento.

3. ¿Cómo se relacionan las horas de sueño y los niveles de estrés con la productividad y el agotamiento laboral?

Nivel de Estres	Promedio Prod Real	Promedio Horas de Sueño	Promedio Días Mes con Agotamiento
1	4,94	6,50	15,57
2	4,97	6,46	15,62
3	4,96	6,50	15,57
4	4,97	6,50	15,53
5	4,95	6,54	15,59
6	4,96	6,51	15,48
7	4,96	6,48	15,52
8	4,90	6,46	15,58
9	5,03	6,53	15,70
10	4,94	6,51	15,47
Total	**4,96**	**6,50**	**15,56**

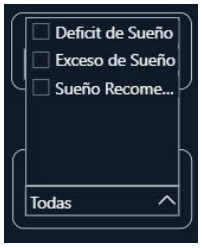


Tabla 4. Visualización tabla multivariable

Según el análisis realizado, se agruparon las horas de sueño dependiendo las horas de sueño de cada persona teniendo en cuenta que el sueño recomendado para personas entre 18 a 65 años es de 7 a 9 horas, esto con el fin de poder relacionar las variables Horas de sueño y Nivel de Estrés con respecto a la productividad real y los días de agotamiento en dónde se obtuvo que el Nivel de Estrés de los individuos influye de manera moderada en su productividad promedio, sus horas de sueño o la cantidad de días que experimentan agotamiento laboral. En todas las métricas la variabilidad de los datos es mínima, en productividad real el valor mínimo y máximo promedio son 4.90 y 5.03 y en Promedio días mes con Agotamiento el valor mínimo y máximo promedio son 15.47 y 15.62.

Por lo anterior, de acuerdo con la data, aunque el estrés aumente o disminuya, las personas no se ven afectadas en su productividad ni evidenciando agotamiento laboral. Respecto a las horas de sueño, su variabilidad es poca, demostrando que no afecta la productividad y agotamiento.

4. ¿Existen diferencias significativas en la productividad promedio y el nivel de agotamiento entre hombres y mujeres, o entre distintos tipos de trabajo?

Agotamiento:

Tipo De Trabajo	Hombre	Mujer	Otros
Education	15,20	15,56	16,09
Finance	15,83	15,58	14,67
Health	15,54	15,60	14,80
IT	15,54	15,71	16,55
Student	15,25	15,78	14,78
Unemployed	15,70	15,41	16,41

Tabla 5. Visualización tabla comparativa Agotamiento – Géneros - Grupos

El análisis de la tabla de agotamiento por tipo de trabajo y género permite concluir que no existen diferencias significativas en el nivel de agotamiento promedio entre los grupos estudiados, porque la variabilidad de los datos es mínima.

Productividad Real:

Tipo De Trabajo	Hombre	Mujer	Otros
Education	4,91	4,96	4,78
Finance	4,97	4,97	5,07
Health	4,94	4,97	4,67
IT	5,04	5,01	4,81
Student	4,93	4,91	4,86
Unemployed	4,98	5,02	4,92

Tabla 6. Visualización tabla comparativa Productividad – Géneros - Grupos

En el análisis observamos que los datos muestran una notable consistencia en la productividad entre todos los grupos y tipos de trabajo. Las diferencias, aunque existentes, son mínimas y todas las puntuaciones se agrupan en un rango muy estrecho justo en el centro de la escala de productividad real (0 a 10).

CONCLUSIONES

Productividad y percepción

La productividad real promedio es 4.96, mientras que la autoevaluada es 5.52, con una diferencia de sólo 0.56 puntos.

Esto indica que las personas perciben rendir ligeramente más de lo que realmente producen, pero la brecha no es significativa.

Agotamiento

El promedio de días de agotamiento al mes es 15.56 en todos los grupos (Education, Finance, Health, Student, Unemployed).

Esto evidencia que el cansancio es un factor constante y generalizado, sin grandes diferencias por tipo de trabajo.

Redes sociales y productividad

A mayor número de horas en redes, la productividad disminuye levemente.

Por ejemplo, quienes tienen alto uso de redes muestran una productividad promedio de 4.94, mientras que quienes tienen alta intensidad laboral y bajo uso de redes mantienen 4.97.

La diferencia máxima es de solo 0.03 puntos, lo que confirma que el impacto, aunque presente, no es drástico.

Consumo de café y hábitos

Se consumen en total 59.979 tazas de café, con un promedio de 2 tazas diarias por persona.

El promedio de horas de sueño es 7.00, el uso de redes antes de dormir es 5.5 horas y las horas trabajadas son 1.05 (valores cercanos entre todos los géneros y tipos de trabajo).

Tipos de empleado

La mayoría se concentra en baja intensidad laboral y bajo uso de redes (11.000 personas, 36.97%) y en alta intensidad laboral y bajo uso de redes (12.000 personas, 38.49%).

Esto significa que cerca del 75% de la población tiene un uso moderado o bajo de redes sociales, con impactos mínimos en su productividad.

Los datos muestran que las diferencias entre grupos son mínimas. La productividad real (4.96) y autoevaluada (5.52) son muy cercanas, el agotamiento mensual se mantiene en 15.56 días para todos, y el consumo de café se estabiliza en 2 tazas diarias.

Aunque se observa que un mayor uso de redes sociales reduce la productividad, la variación es de apenas 0.03 puntos entre los extremos. En síntesis, los problemas de agotamiento, productividad y hábitos digitales son generales en toda la población, sin que un sector o grupo presente diferencias significativas.

Referencias

Google. (s.f.). *Colab*. Recuperado el 1 de octubre de 2025, de https://colab.research.google.com/

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90-95. https://doi.org/10.1109/MCSE.2007.55

Mashayekhi, M. (2024). *Social Media vs Productivity*. Kaggle. https://www.kaggle.com/datasets/mahdimashayekhi/social-media-vs-productivity

Microsoft Corporation. (2025). *Microsoft Excel* (Versión 16.0) [Software de computadora]. https://office.microsoft.com/excel

Microsoft. (s.f.). *Power BI*. Recuperado el 1 de octubre de 2025, de https://powerbi.microsoft.com/

Python Software Foundation. (s.f.). *Python* (Versión 3.12.5) [Software de computadora]. https://www.python.org

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

LISTA DE TABLAS

LISTA DE FIGURAS