



Análisis Exploratorio de Datos Productividad y Redes Sociales, ¿amigos o enemigos?

Grupo TAD – 10



@SENACOMUNICA

www.sena.edu.co

Organigrama de Equipos de Trabajo

Líder de Proyecto:
Juan Camilo De La Cruz Montes

Instructor:
Roymer Antonio Romero Algarin

Fase 1 : Preparación de Datos

Líder: José Suarez Gonzales
Luis Carlos Castro Rojas
Camilo Andrés Visbal Polo
Iham David Donado Gutiérrez
Wuil Rafael De La Cruz Méndez

Fase 2 : Análisis Descriptivo

Líder: Juan De La Cruz Montes
Jean Carlos Arevalo Cervera
Julián Ricardo Laverde Meza
Jean Pierr David Pulido Carrasquilla
Jesús Daniel De La Hoz Carranza

Fase 3 : Visualización De Datos

Líder: Jons Sandoval Pérez
Brayan José Torres Arellano
Daniel José Reyes Betancourt
Luis Yerson Herrera Ruiz

Fase 4 : Storytelling y Presentación de Hallazgos

Líder: Josh Carroll Solís
David Antonio Rivero León
Robinson Maza Velásquez
Luis Ángel Chinchilla Moreno

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.

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



Herramientas



Power BI (Power Pivot, Power Query, Power Service, Lenguaje DAX)



Pandas, Numpy, Matplotlib, Seaborn, Sklearn



Google Colaboratory



GitHub



Fase de Preparación y Limpieza de Datos



Esta fase se centró en acondicionar el dataset para su análisis.

Dataset: Social Media vs Productivity - Impact of Social Media Usage on Individual Productivity

Disponible en: **kaggle**

<https://www.kaggle.com/datasets/mahdimashayekhi/social-media-vs-productivity>

Detalle:

About this file

This CSV file contains 30,000 rows and 19 columns, simulating real-world behavioral data on individuals' social media usage and its impact on productivity and well-being.

- Includes missing values, outliers, and correlated features
- Ideal for data cleaning, feature engineering, and predictive modeling
- Covers a range of demographic, digital behavior, and mental health indicators
- Ready to be used for EDA, ML pipelines, or statistical research.

Fase de Preparación y Limpieza de Datos



Estandarización de la base de datos.

Se tradujeron los nombres de las columnas al español y se redondearon los números decimales a solo dos dígitos.

No se encontraron inconsistencias en los tipos de dato ni registros .

Detección y análisis de datos faltantes.

COLUMNAS	CANTIDAD NULOS	% DATOS NULOS
HORA_DIARIA_EN_REDES	2765	9.2%
PUNTUACION_PRODUCTIVIDAD_AUTOEVALUADA	1614	5.3%
PUNTUACION_PRODUCTIVIDAD_REAL	2365	7.8%
NIVEL_ESTRES	1904	6.3%
HORAS_DE_SUEÑO	2598	8.6%
USO_ANTES_DE_DORMIR	2211	7.3%
PUNTUACION_DE_SATISFACCION_LABORAL	2730	9.1%
TOTAL NULOS	16187	53.6%



[Limpieza de Datos.ipynb](#)

Análisis De Datos Faltantes

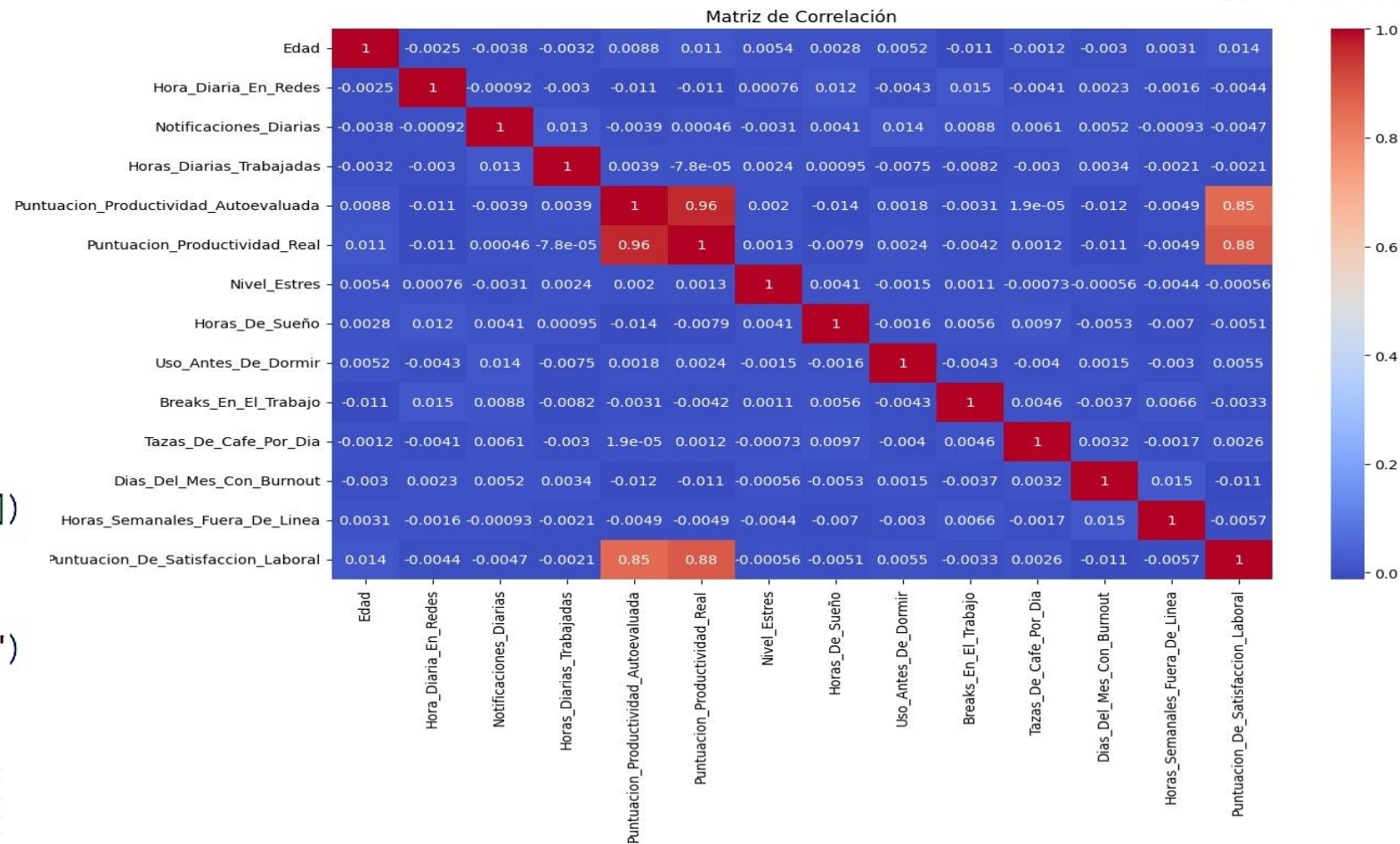


Se creo una matriz de correlación que mostró pocas correlaciones significativas entre las columnas, sugiriendo que la mayoría de los datos faltantes eran del tipo MCAR (Missing Completely At Random), permitiendo una imputación por media aritmética en las columnas de la tabla, a excepción de Puntuacion_De_Satisfaccion_Laboral, Puntuacion_Productividad_Real y Puntuacion_Productividad_Autoevaluada

COLUMNAS	PROMEDIO
HORA_DIARIA_EN_REDES	3.11
NIVEL_ESTRES	5.51
HORAS_DE_SUEÑO	6.50
USO_ANTES_DE_DORMIR	1.02

```
numeric_df = df.select_dtypes(include=['float64', 'int64'])
```

```
plt.figure(figsize=(14,8))
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm")
plt.title("Matriz de Correlación")
plt.show()
```

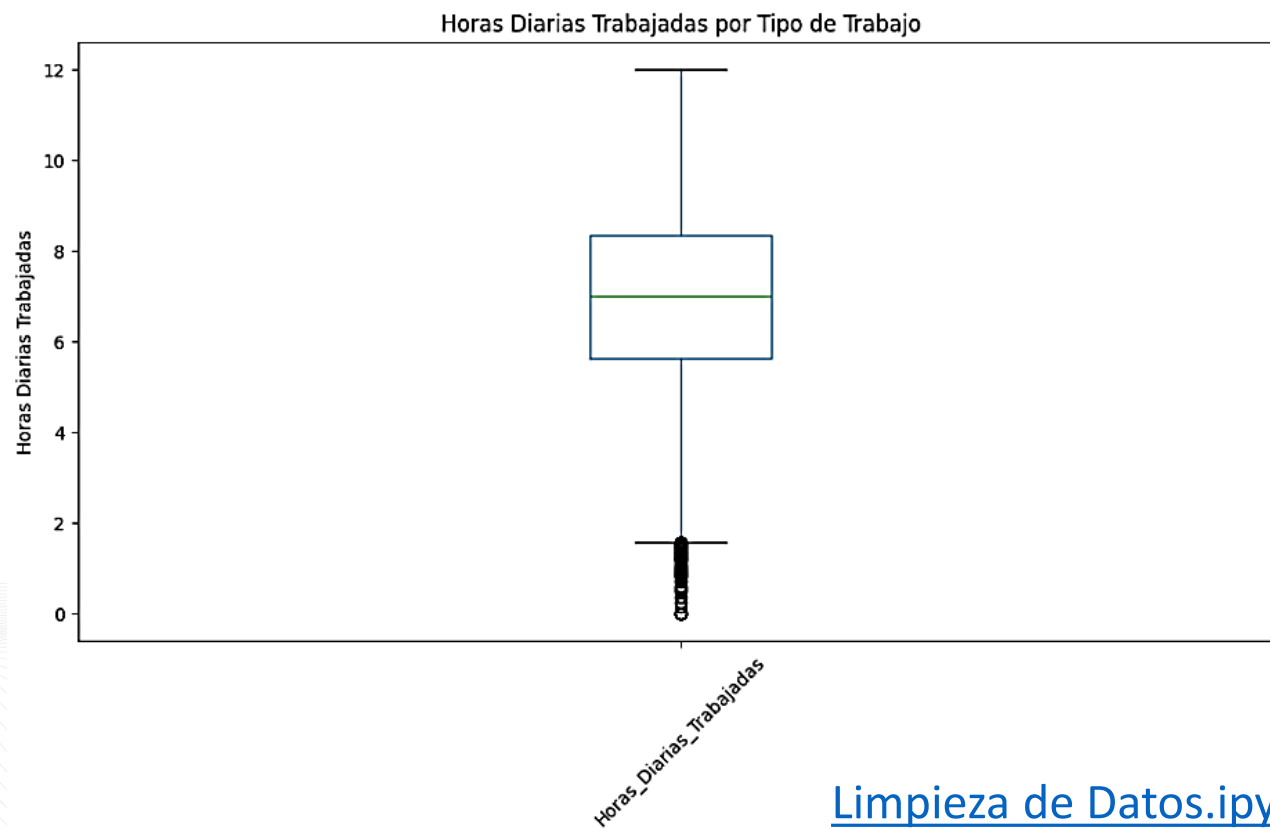
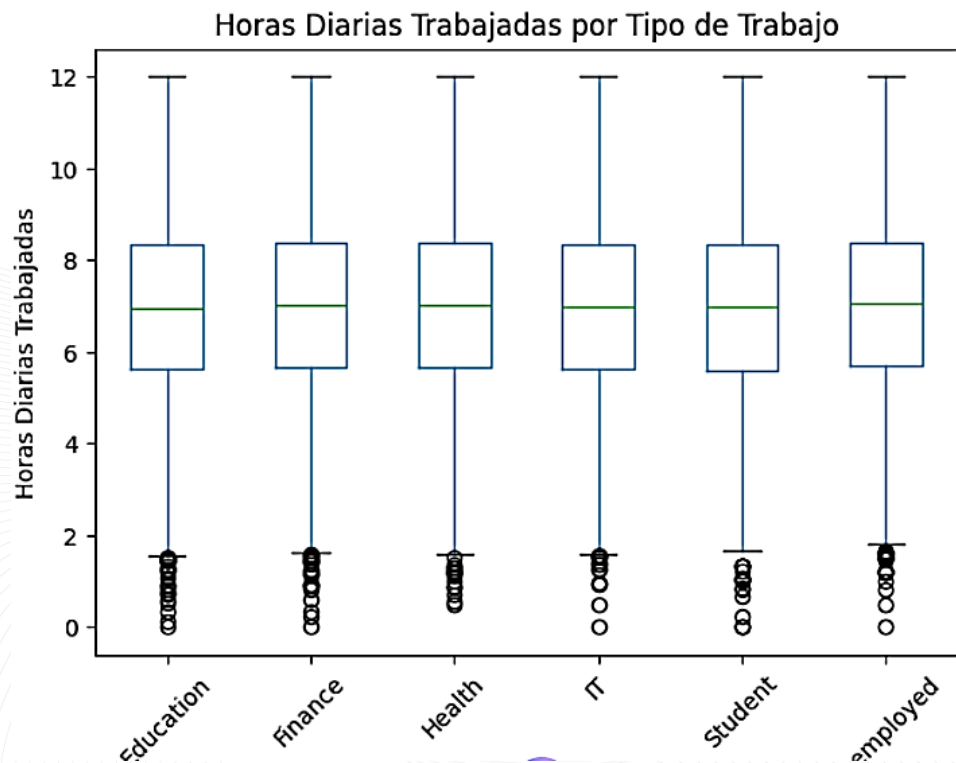


Detección De Outliers



Se identificaron 95 valores atípicos en la columna **Horas_Trabajadas_Por_Dia** (por debajo de 1.55 horas) utilizando un Diagrama de Cajas y Bigotes (Boxplot). Estos se corrigieron imputándolos con la de la columna (6.99 horas promedio media aritmética).

```
df.loc[df['Horas_Diarias_Trabajadas'] < 1.575, 'Horas_Diarias_Trabajadas'] = 6.99
```

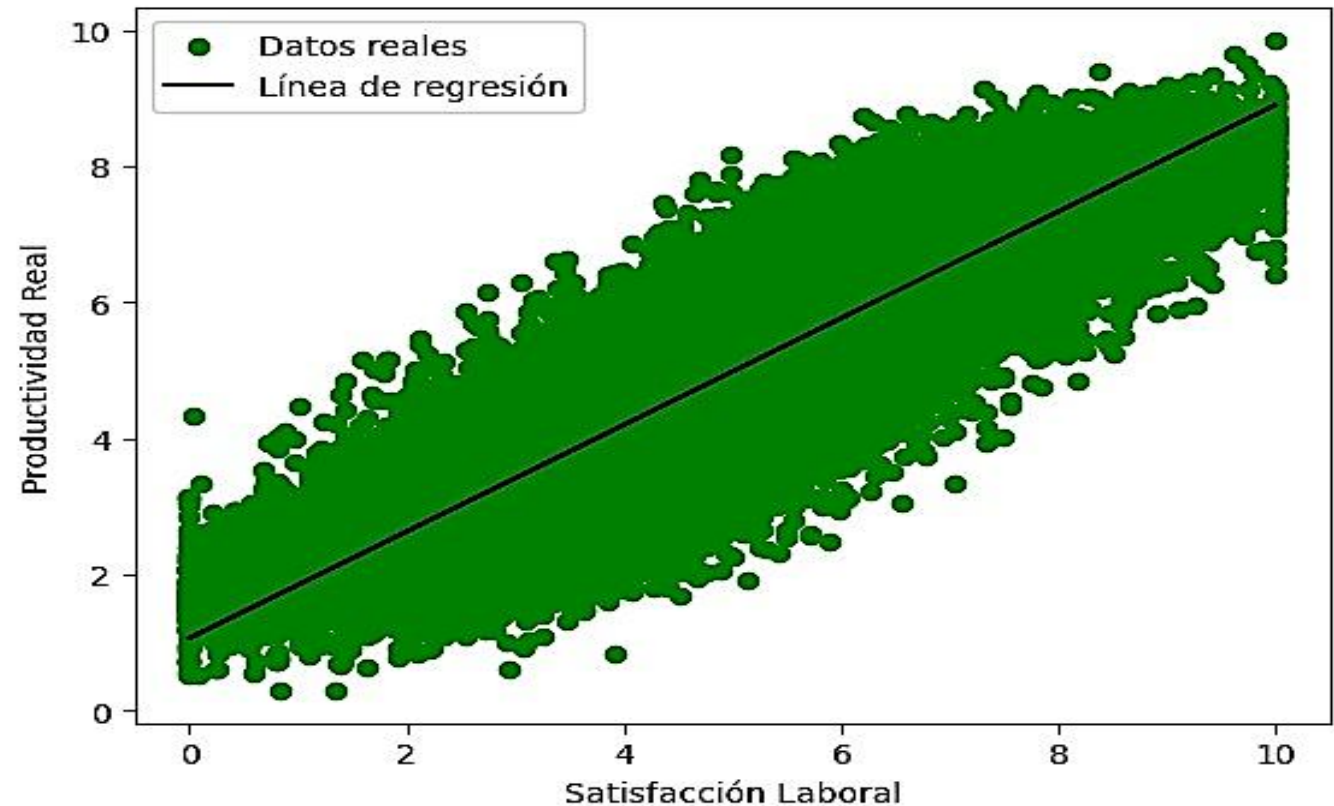


Imputación Con Regresión Lineal



Los valores faltantes de `Puntuacion_Productividad_Real` y `Puntuacion_Productividad_Autoevaluada` se trataron de forma especial debido a su correlación positiva: Primero, se imputó la **autoevaluada** usando la **real más 0.55** (el promedio de diferencia constante encontrado entre ellas). Los 129 registros restantes con datos faltantes en ambas columnas se imputaron mediante **regresión lineal** en Google Colab para preservar la correlación observada.

$$y = 0.7834 * x + 1.0669$$



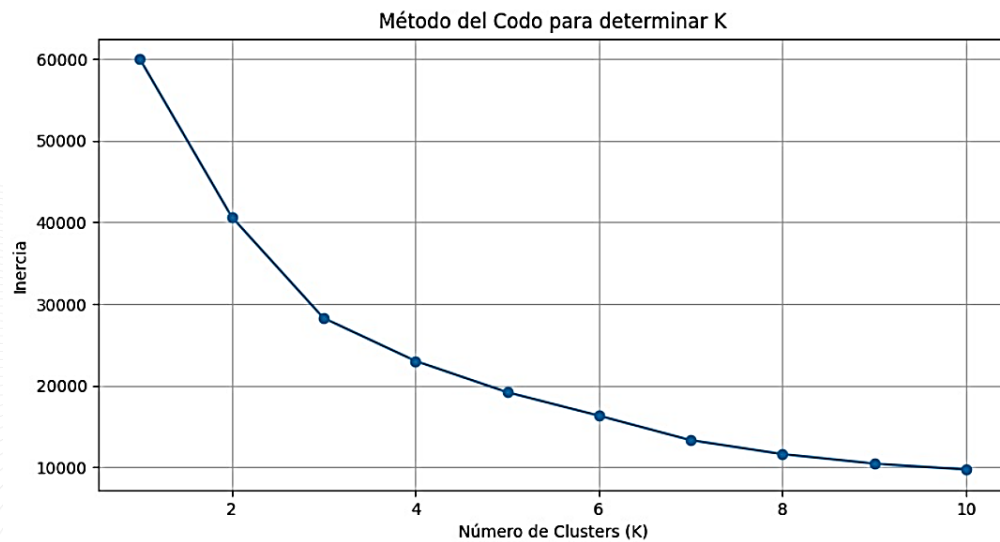
```
df['Puntuacion_Productividad_Real'] = df['Puntuacion_Productividad_Real'].fillna(0.7834316671585180 * df['Puntuacion_De_Satisfaccion_Laboral'] + 1.066996038794188)
```

Clusterizacion K-Means



Método Para Determinar K

Por medio del método de codo entre las variables `Horas_Diarias_En_Redes` y `Horas_Diarias_Trabajadas` se observa la mayor variación en $K = 3$



[Análisis y clustering.ipynb](#)

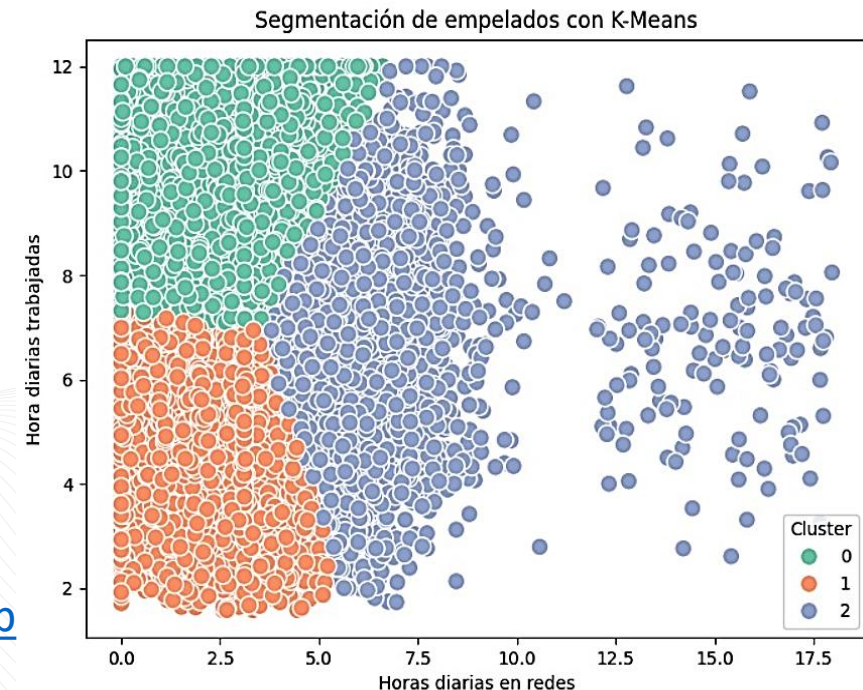
Análisis Clusterizacion

Observando la grafica de dispersión podemos identificar los 3 tipos diferentes de empleado:

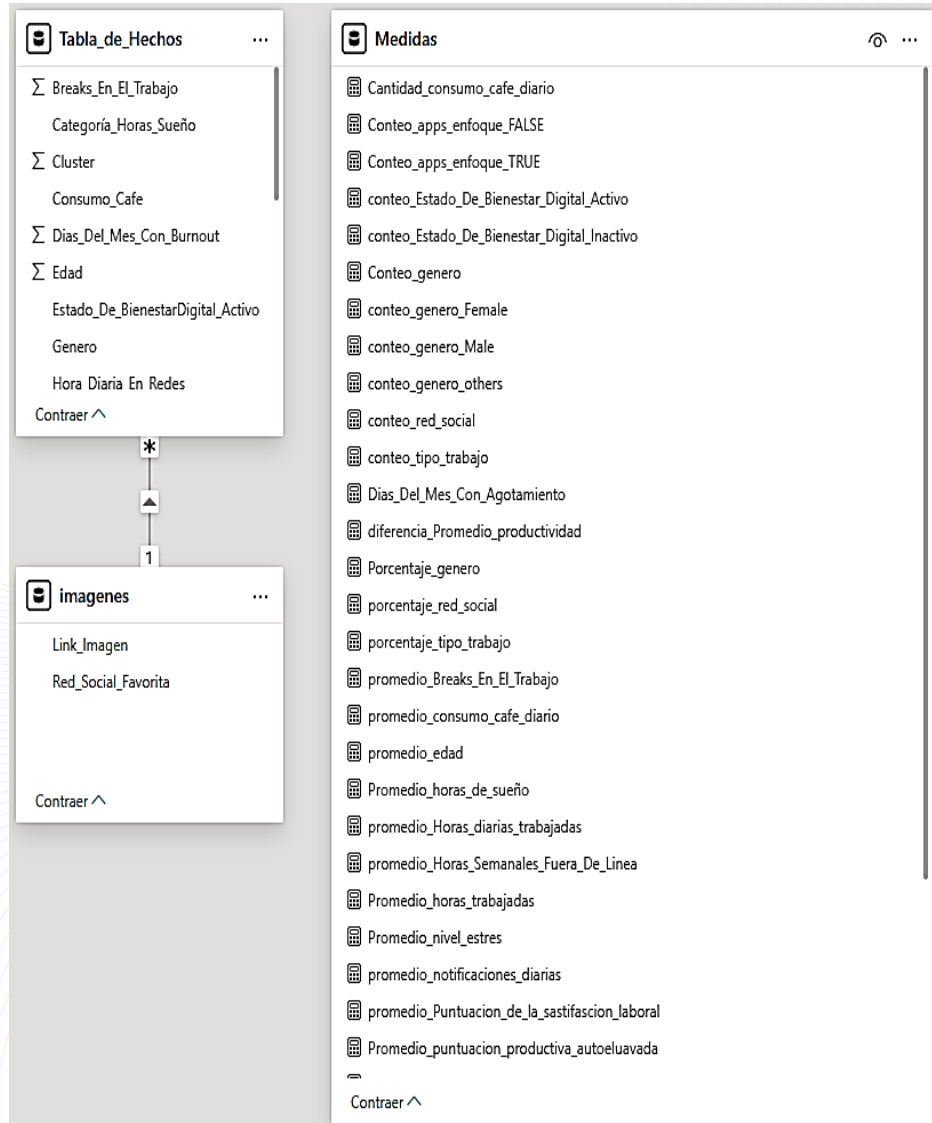
Cluster0 = Empleados con alta intensidad laboral y bajo uso de redes

Cluster1 = Empleados con baja intensidad laboral y bajo uso de redes

Cluster2 = Empleados con alto uso de redes



Fase de Modelado



Medidas Principales Usadas

`diferencia_Promedio_productividad = [Promedio_puntuacion_productiva_autoeluvada] - [Promedio_puntuacion_productiva_real]`

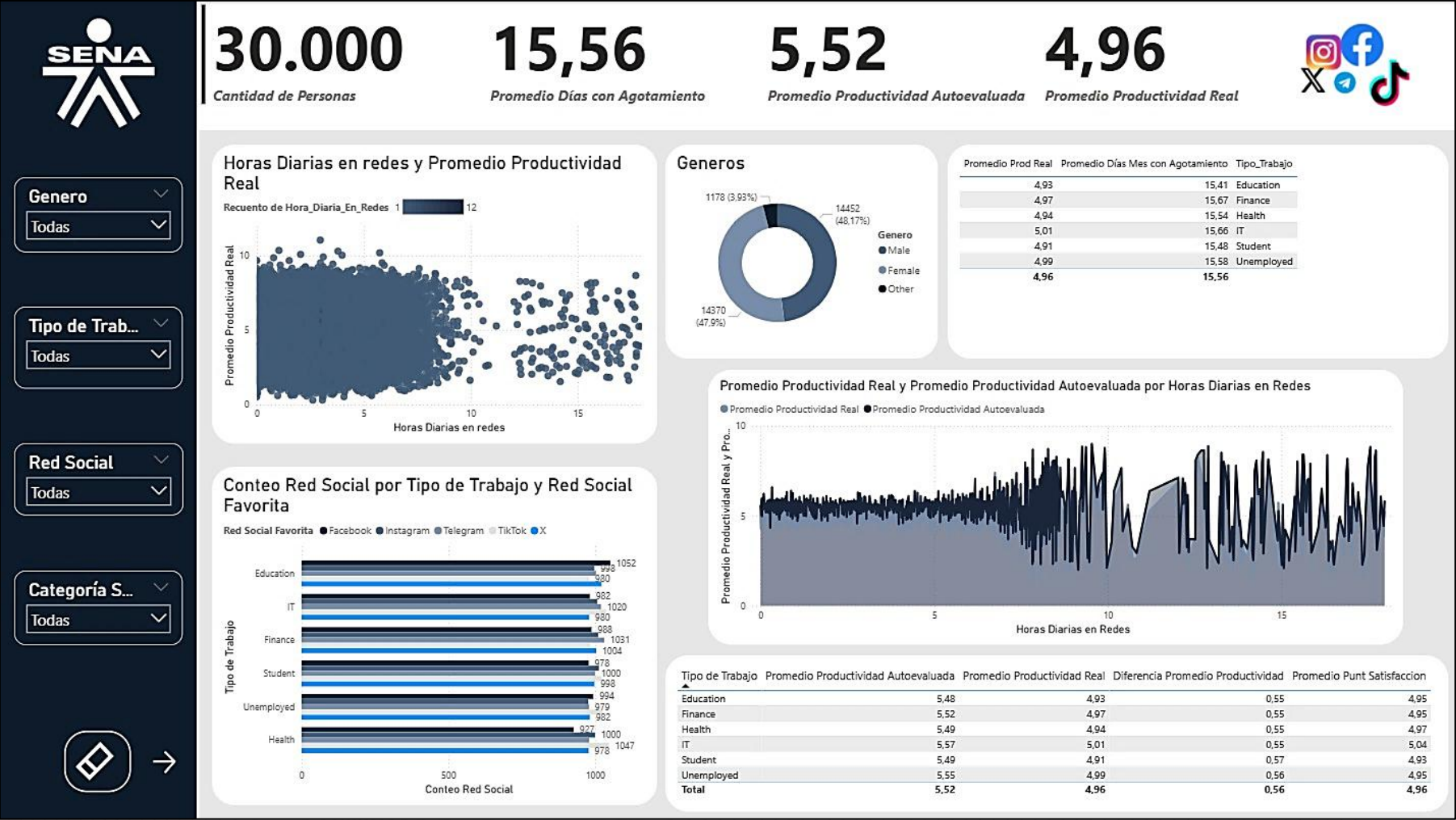
`Promedio_horas_de_sueño = AVERAGE(Tabla_de_Hechos[Horas_De_Sueño])`

`Promedio_puntuacion_productiva_autoeluvada = AVERAGE(Tabla_de_Hechos[Puntuacion_productividad_autoevaluada])`

`Promedio_puntuacion_productiva_real = AVERAGE(Tabla_de_Hechos[Puntuacion_Productividad_Real]) + 0`

`promedio_Puntuacion_de_la_sastifascion_laboral = AVERAGE(Tabla_de_Hechos[Puntuacion_De_Satisfaccion_Laboral])`

Fase de Visualización



Escanea



O visita:
<https://tinyurl.com/2vpyhj7a>

Respuesta a Preguntas Base



Se establecieron cuatro preguntas principales para guiar el análisis

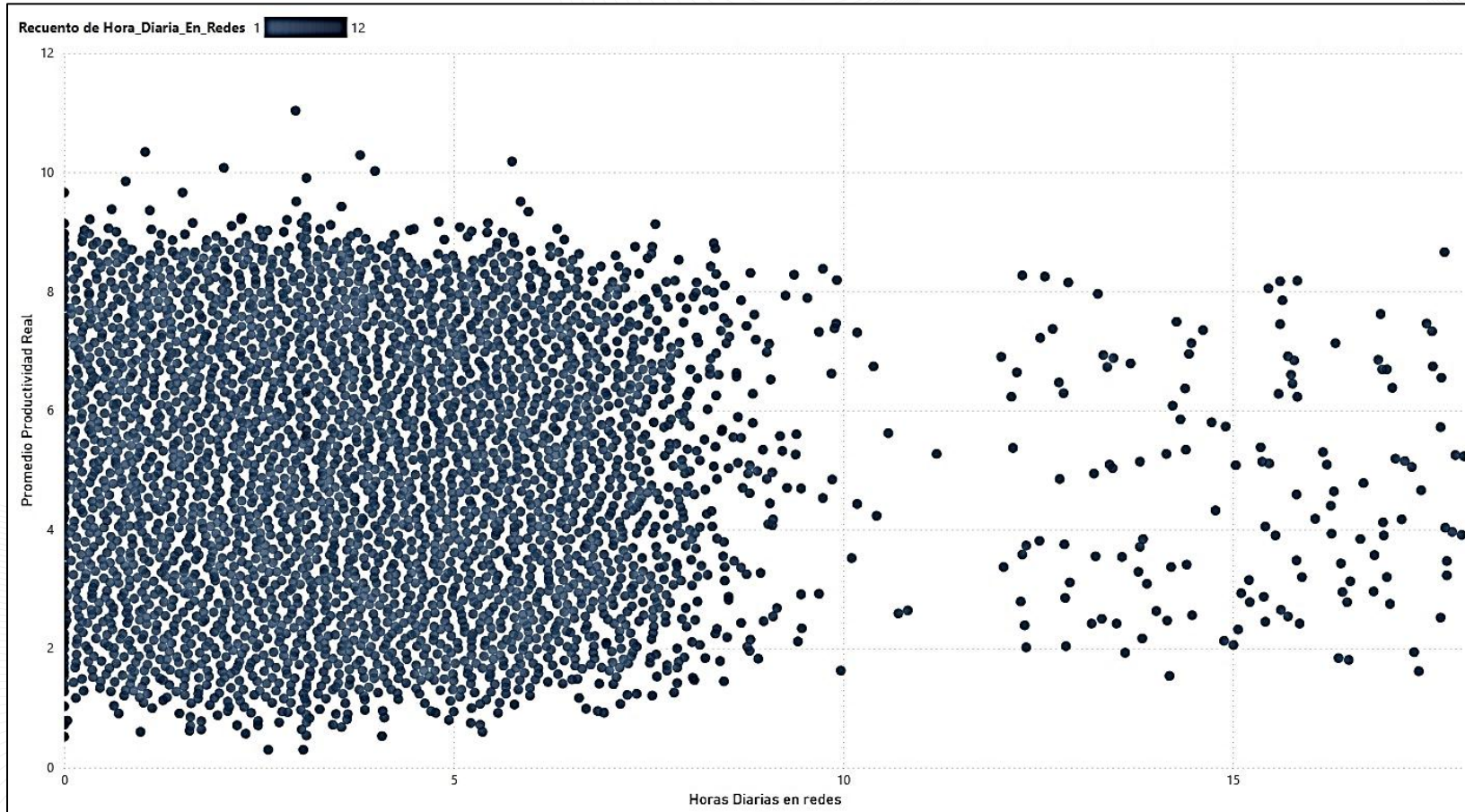
1. ¿Cuál es la diferencia promedio entre la productividad que las personas creen tener y la que realmente demuestran?

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



Los trabajadores perciben con bastante precisión su productividad: la diferencia entre su rendimiento real y el autoevaluado es mínima (0,56), lo que indica que lo que creen producir coincide casi totalmente con lo que realmente logran.

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



El gráfico muestra que no hay una correlación clara entre el tiempo en redes sociales y la productividad real. Los datos son muy variables: con las mismas horas de uso se observan productividades tanto altas como bajas, lo que indica que el uso de redes no determina el rendimiento.

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

☐ Deficit de Sueño
☐ Exceso de Sueño
☐ Sueño Recome...
 Todas

El análisis muestra que ni el nivel de estrés ni las horas de sueño influyen de forma significativa en la productividad real ni en el agotamiento laboral. La variabilidad en los datos es mínima: la productividad promedio se mantiene entre 4.90 y 5.03, y los días de agotamiento entre 15.47 y 15.62 al mes.

Nivel de Estrés	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

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

El análisis muestra que no hay diferencias significativas en el nivel de agotamiento promedio entre géneros ni tipos de trabajo, ya que 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

El análisis evidencia una alta consistencia en la productividad entre los distintos grupos y tipos de trabajo, con diferencias mínimas y valores concentrados en un rango estrecho, ubicado en el centro de la escala (0 a 10).

Fase de Comunicación: Storytelling



Entre Likes, Sueños y Productividad

Imagina un día cualquiera en la vida de miles de trabajadores, estudiantes y profesionales.

Cada mañana comienza con una taza de café, una notificación en el celular y la esperanza de que, esta vez, las redes sociales no roben más minutos de concentración.

Pero... ¿realmente afectan tanto nuestra productividad como creemos?

Así nació este análisis: una travesía de datos para descubrir si los clics, los “me gusta” y los mensajes nocturnos son los verdaderos villanos del rendimiento laboral.



Fase de Comunicación: Storytelling



Capítulo 1: El mito de la percepción

Primero, quisimos saber: ¿Las personas saben cuán productivas son realmente?

Los números revelaron una verdad sorprendente.

Aunque muchos creen rendir más de lo que producen, la diferencia promedio entre su productividad real y la autoevaluada fue de solo **0.56 puntos**.

Es decir, somos buenos jueces de nuestro propio desempeño, aunque quizás un poco optimistas.

En el fondo, cada “me siento productivo” no era una ilusión, sino una intuición bastante acertada.



Fase de Comunicación: Storytelling



Capítulo 2: Redes sociales, ¿culpables o inocentes?

Luego miramos hacia las pantallas. ¿Más horas en redes significan menos productividad?

La intuición decía que sí.

Los datos, en cambio, contaron una historia distinta.

Aunque quienes usan más redes sociales tienden a tener una productividad levemente menor (4.94 vs. 4.97), la diferencia es mínima.

Los gráficos mostraron una dispersión caótica, sin una tendencia clara. **Al final**, el tiempo en redes no es el villano principal, sino un personaje secundario en la trama de la productividad.



Fase de Comunicación: Storytelling



Capítulo 3: Entre sueños y estrés

También exploramos el equilibrio entre cuerpo y mente: ¿Duermes poco? ¿Estás estresado? ¿Te afecta eso en el trabajo?

Los datos fueron tajantes: ni el sueño ni el estrés alteraron significativamente la productividad ni el agotamiento laboral.

Quienes dormían más no rendían mucho mejor, y quienes decían estar más estresados no se agotaban más.

Pareciera que la mayoría ha aprendido a mantener un equilibrio estable, sin importar las tormentas externas.



<https://www.manabiba-s.com/wp-content/uploads/2022/05/nebusoku.png>

Fase de Comunicación: Storytelling



Capítulo 4: Género, café y profesión

¿Y si el secreto estaba en las diferencias entre personas?

Ni el género, ni el tipo de trabajo marcaban una brecha real.

La productividad y el agotamiento fueron uniformes en todos los grupos.

Lo único constante era la taza de café, fiel compañera diaria: 2 por persona, cada día, como si el café fuese el punto de encuentro universal entre los que trabajan y los que sueñan.



Fase de Comunicación: Storytelling



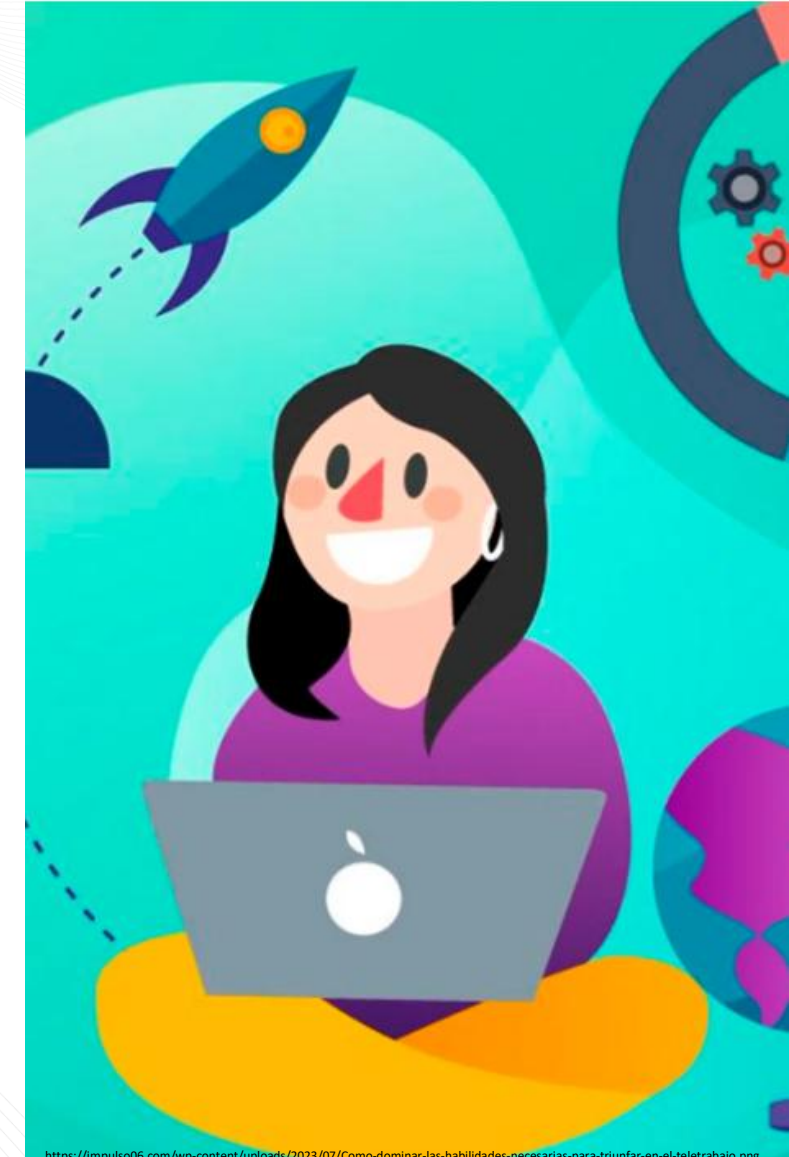
Epílogo: Una historia de equilibrio

En un mundo que vive conectado, descubrimos que la productividad no depende de cuántas horas pasas en redes, sino de cómo equilibras tus hábitos.

Las diferencias son mínimas, los patrones se repiten y, al final, todos compartimos la misma lucha silenciosa entre foco y distracción.

La historia que cuentan los datos es simple pero poderosa:

- No somos tan improductivos como creemos, ni las redes tan culpables como pensamos.
- El verdadero reto no está en desconectarnos, sino en encontrar armonía en medio de la conexión.



Repositorio GitHub



En el QR podrás visitar nuestro repositorio en GitHub y descargar nuestro contenido.



 Visítame



GRACIAS



@SENAcomunica

www.sena.edu.co

Líneas de atención al ciudadano, empresarios y PQRS:

Bogotá: +(57) 601 736 60 60

Línea gratuita resto del país: 018000 91 02 70

Línea nacional: +(57) 601 546 15 00