

# PROYECTO FINAL DE ANÁLISIS DE DATOS NIVEL EXPLORADOR

## Comportamiento de las tarifas aplicadas al gas natural en Colombia durante los años 2013 - 2022

**ELABORADO POR**

**Cristián David Perafán Agudelo**

**Diana Carolina Rangel Fuentes**

**Juan David Gómez**

**María Luisa Moreno Gutierrez**

**MINISTERIO DE LAS TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES –  
MINTIC**

## TALENTO TECH

**MEDELLIN**

**2025**



## Tabla de contenido

Introducción	2
Objetivos del Proyecto	4
Objetivo general	4
Objetivos específicos	4
Desarrollo del Proyecto:	5
1. Identificación de la Problemática:	5
2. Recolección de Datos:	5
3. Exploración y Limpieza de Datos:	6
4. Análisis Descriptivo:	9
5. Interpretación de Resultados:	14
Conclusiones	16
Recomendaciones	17
Repositorio GitHub	18

## Objetivos del Proyecto

### Objetivo general

Analizar el crecimiento en la adquisición del servicio de gas natural a nivel nacional durante el periodo 2013-2022, y examinar la dinámica de los costos asociados a este servicio y sus variaciones entre las diferentes empresas proveedoras.

### Objetivos específicos

- Cuantificar el crecimiento en la adquisición del servicio de gas natural a nivel nacional durante el periodo 2013-2022, identificando las tasas de crecimiento anual y el aumento acumulado en el número de usuarios o conexiones.
- Evaluar la posible relación entre el nivel socioeconómico de los hogares y la adopción del servicio de gas natural a nivel nacional durante el periodo 2013-2022.

## Desarrollo del Proyecto:

## 1. Identificación de la Problemática:

La necesidad del proyecto surge como requerimiento para implementar y sustentar las habilidades adquiridas durante el Bootcamp: Análisis de datos nivel explorador, además de contribuir con un análisis sólido acerca del crecimiento adquisitivo por parte de los usuarios y comportamiento tarifario del gas natural en Colombia, como fuente ecológicamente sostenible, una alternativa de energía que apunta hacia el futuro.

## 2. Recolección de Datos:

Se investigaron en diferentes fuentes virtuales, temas relacionados con transición energética que abordará temas como el consumo de energía, impactos de fuentes de energía renovable, y que además contarán con una base de datos confiable que permitiera ser descargada y analizada desde Python, que fue el lenguaje afianzado para desarrollar esta problemática. Una vez seleccionada la base de datos sólida en cuanto a cantidad de datos y confiabilidad de la fuente, fue cargada a un Drive en Google Colab, el cual, permitió su posterior manipulación desde Python.

### 3. Exploración y Limpieza de Datos:

## Importación de librerías:

Fue indispensable para el análisis de la base de datos descargar librerías como: Pandas, Numpy, -Statsmodels.api para el procesamiento de los datos desde su limpieza hasta agrupaciones o procesos estadísticos. Así como Matplotlib, Seaborn para la parte gráfica de la información que se ilustra a lo largo del estudio.

A continuación, se puede hallar una descripción de lo realizado en el desarrollo del código, plasmado en un registro secuencial con su respectiva explicación.

Inicialmente se descargaron las librerías en Phyton para poder realizar los análisis pertinentes.

```
[ ] ## Librerias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
```

```
[ ] !pip install mlcroissant
```

 [Mostrar salida oculta](#)

```
[ ] import mlcroissant as mlc
from scipy import stats
```

## Base de datos:

Se cargó la URL de la cual fueron extraídos los datos analizados.

```
[ ] ##Carga de archivos desde una URL
url = 'https://www.datos.gov.co/resource/ek3f-5wn4.json?$limit=134372'
#el parametro sep indica que los valores estan separados por ;
df_tarifas_gn = pd.read_json(url)
```

La parte de \$limit.json= fue usada para delimitar la fila hasta la cual se iban a cargar los datos, en este caso, este límite coincide con el último dato registrado del año 2022.

## Limpieza:

Después de un primer análisis de la base de datos, se decidió que algunas columnas que no eran prácticas para el análisis, tales como [nit, id\_mercado, id\_empresa] ya que solo tenían como función identificar la empresa prestadora del servicio por medio del código establecido por la CREG, que no era afín al análisis principal. Así mismo se encontraron valores negativos y atípicos, los cuales fueron respectivamente procesados y limpiados.

	ano	mes	nit	id_mercado	id_empresa	estrato	conexion	reinсталacion	reconexion	cargo_fijo
count	134372.000000	134372.000000	1.343720e+05	134372.000000	134372.000000	134372.000000	1.343720e+05	1.343720e+05	1.343720e+05	1.343720e+05
mean	2018.102946	6.459992	8.557540e+09	150.227056	5075.012532	3.367487	6.052740e+05	1.929199e+05	2.988732e+04	2.221084e+03
std	3.126079	3.408614	4.151421e+08	166.083751	8956.699502	1.693608	1.701886e+05	1.418797e+05	3.256611e+04	1.449866e+04
min	2005.000000	1.000000	8.000078e+09	9.000000	484.000000	1.000000	0.000000e+00	0.000000e+00	0.000000e+00	-1.009710e+03
25%	2016.000000	4.000000	8.040118e+09	47.000000	564.000000	2.000000	5.195480e+05	9.400000e+04	2.240600e+04	0.000000e+00
50%	2018.000000	6.000000	8.901017e+09	113.000000	620.000000	3.000000	5.806510e+05	1.652000e+05	2.865700e+04	2.407420e+03
75%	2021.000000	9.000000	8.909050e+09	179.000000	2097.000000	5.000000	6.420880e+05	2.232300e+05	3.565000e+04	3.521520e+03
max	2023.000000	12.000000	9.010428e+09	1825.000000	37253.000000	6.000000	1.870000e+06	9.826062e+06	4.600208e+06	3.723696e+06

Una vez realizado todo el proceso de limpieza y organización de datos, se obtuvo el nuevo DataFrame para analizarlo.

	año	mes	estrato	conexion	reinсталacion	reconexion	cargo_fijo	rango_0	rango_21
count	117729.000000	117729.000000	117729.000000	117729.000000	1.177290e+05	1.177290e+05	117729.000000	117729.000000	117729.000000
mean	2017.950437	6.548353	3.386201	572855.092551	1.865748e+05	2.953587e+04	2140.111963	2045.800099	1514.122896
std	2.760906	3.447273	1.699574	70157.424579	1.348324e+05	2.842550e+04	1858.043740	1713.154499	1007.120688
min	2013.000000	1.000000	1.000000	357900.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000
25%	2016.000000	4.000000	2.000000	510842.000000	8.838400e+04	2.280000e+04	0.000000	1062.000000	984.840000
50%	2018.000000	7.000000	3.000000	565388.000000	1.620590e+05	2.865700e+04	2379.000000	1557.260000	1562.690000
75%	2020.000000	10.000000	5.000000	623864.000000	2.190000e+05	3.544200e+04	3463.000000	2318.868000	2104.080000
max	2022.000000	12.000000	6.000000	715338.000000	9.826062e+06	4.600208e+06	7682.780000	14388.800000	10748.230000

#### 4. Análisis Descriptivo:

Se realiza un análisis descriptivo estadístico básico utilizando “.describe”, para poder así conocer cuales son los valores minimos y maximos, cual es la media, la mediana, desviación estándar y los cuartiles 25%, 50% y 75%, esto con el fin de saber que tan ordenados están los datos y si hay valores atípicos en la base de datos.

Con las líneas de código que se muestran a continuación, se realizó el cálculo de la moda, la varianza, y la curtosis.





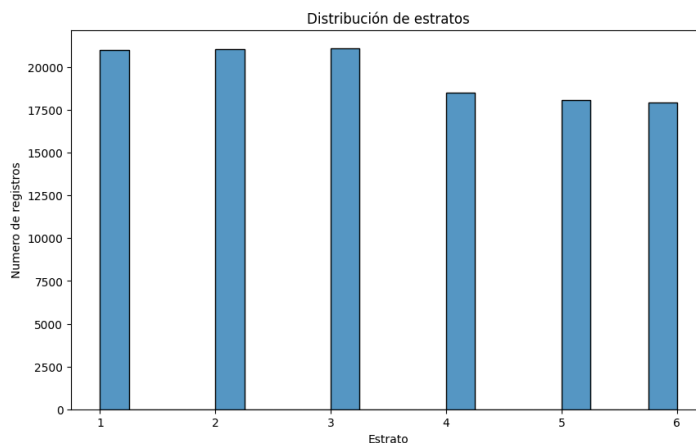
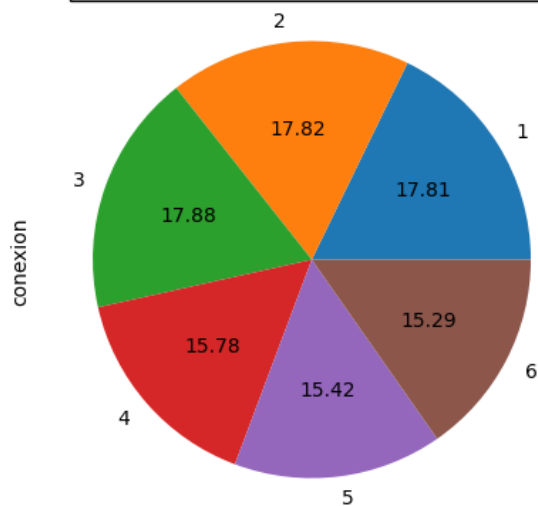
## 5. Interpretación de Resultados:

Una vez agrupados los datos, las variables analizadas fueron las siguientes:

Tt	ANO		ano
#	MES		mes
#	ESTRATO	Estrato socio-economico residencial de los usuarios de Gas Natural	estrato
#	CONEXION	Valor (cantidad monetaria) de conexión del servicio de Gas Natural para un usuario, según su estrato	conexion
#	CARGO_FIJO	Valor (cantidad monetaria) de la tarifa fija del servicio de Gas Natural para un usuario, según su estrato	cargo_fijo
#	RANGO_0	Valor (cantidad monetaria) del pago servicio de Gas Natural para un usuario, según su estrato y su consumo (si es menor a 20M3)	rango_0
#	RANGO_21	Valor (cantidad monetaria) del pago servicio de Gas Natural para un usuario, según su estrato y su consumo (si es mayor a 20M3)	rango_21

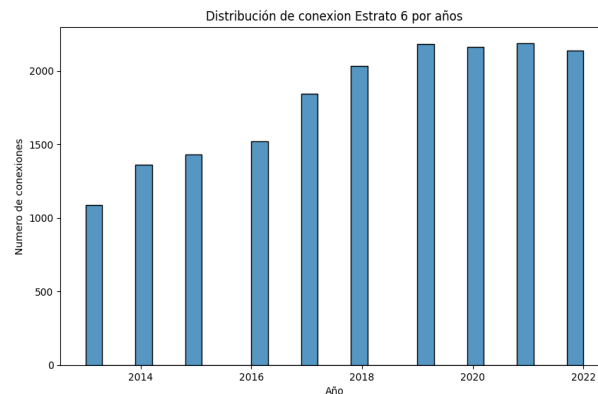
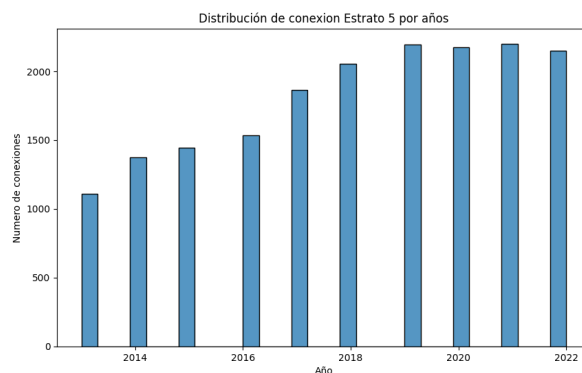
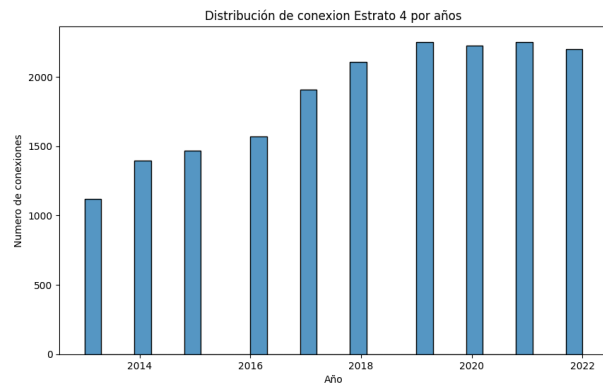
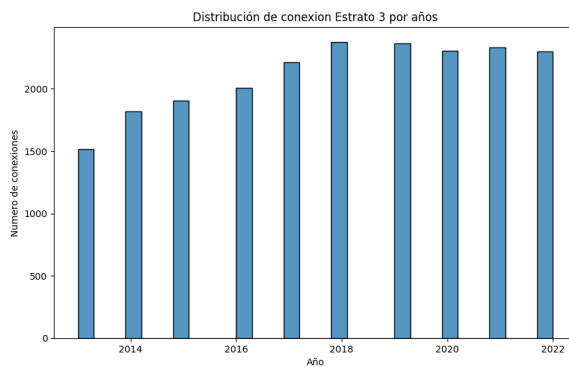
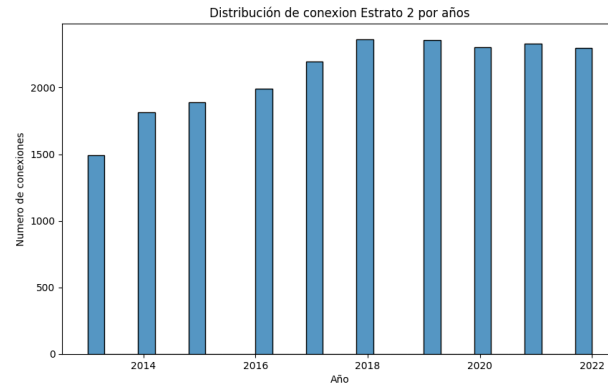
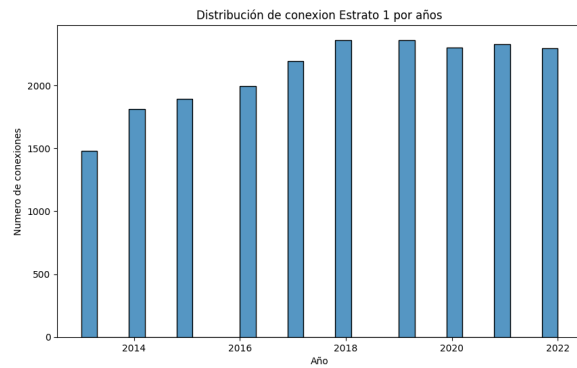
La primera agrupación realizada fue para analizar la cantidad de servicios instalados y cómo estaban repartidos a lo largo de los 6 diferentes estratos, lo que se refleja en el **Gráfico 1**, el cual representa la distribución de conexión de gas por estratos; el cual, evidencia que los porcentajes de conexión son bastante similares entre los estratos, oscilando entre 15.29% y 17.88%. Esto sugiere una distribución relativamente equitativa de las conexiones entre los diferentes grupos; sin embargo, los estratos 1, 2 y 3 tienen porcentajes ligeramente más altos alrededor del 17.8%, mientras los estratos 4, 5 y 6 tienen porcentajes ligeramente más bajos entre 15.3% y 15.8%. Lo anterior se relaciona con el número de registros obtenidos por estratos, ya que el número de registros es bastante similar entre los estratos 1, 2 y 3 (alrededor de 20,800). Los estratos 4, 5 y 6 tienen un número ligeramente menor de registros (alrededor de 18,000 - 18,500); sin embargo, todos los estratos muestran un crecimiento de las conexiones a lo largo del tiempo.

Distribucion de conexión por estratos



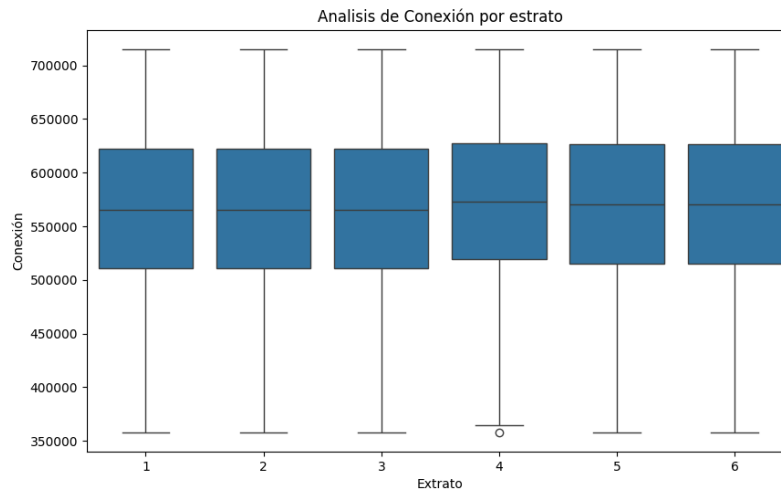
**Gráfica 1:** Distribución de conexión de gas por estrato y registro

Los histogramas del **Gráfico 2**, relacionan la distribución de conexión por años para diferentes estratos socioeconómicos; en general, todos los estratos muestran una clara tendencia ascendente en el número de conexiones a lo largo de los años (2013 a 2022). Esto indica un crecimiento constante en la adopción o el registro de conexiones en todos los grupos y reflejan un patrón muy similar; el cual, muestran un aumento gradual al principio, luego una estabilización o un aumento más lento en los últimos años. Si bien la tendencia es similar, la escala del número de conexiones varía entre los estratos, ya que para los estratos más altos (4, 5 y 6) comienzan con un número inicial de conexiones más bajo en 2013, pero alcanzan niveles similares a los estratos 1, 2 y 3 al final del período; en este caso, para el año 2023. Estos primeros tres estratos con un pico de número de conexiones para el año 2018, mientras para los últimos tres estratos, esto fue reflejado para el año 2019.



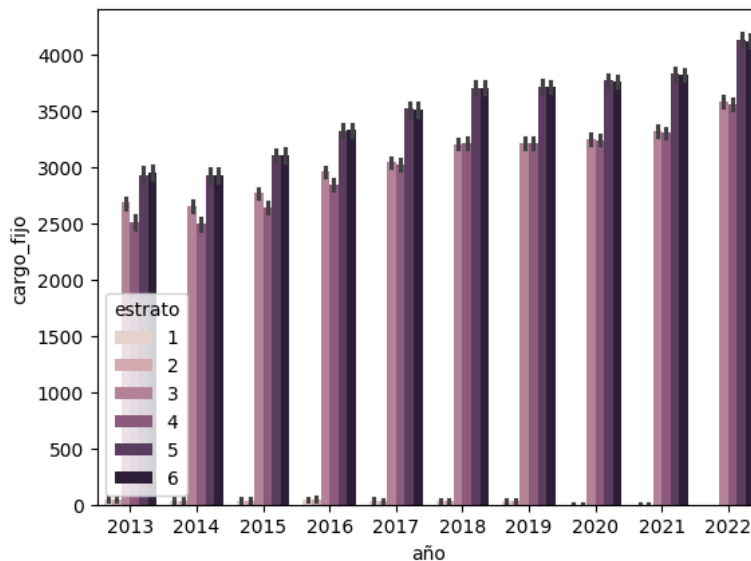
**Gráfico 2: Conexión de gas por estrato y año**

Ya analizando el costo de conexión del servicio a la red de gas natural, se obtuvo el **Gráfico 3**, de caja de bigotes, en el cual se evidencia una uniformidad en los costos de conexión entre los estratos socioeconómicos. Lo cual, explica que las medianas en los costos son muy similares indicando que el costo típico de conexión no varía significativamente según el nivel socioeconómico. También, al observar el rango intercuartil, nos podemos dar cuenta que los factores que influyen en el costo de la conexión de gas.



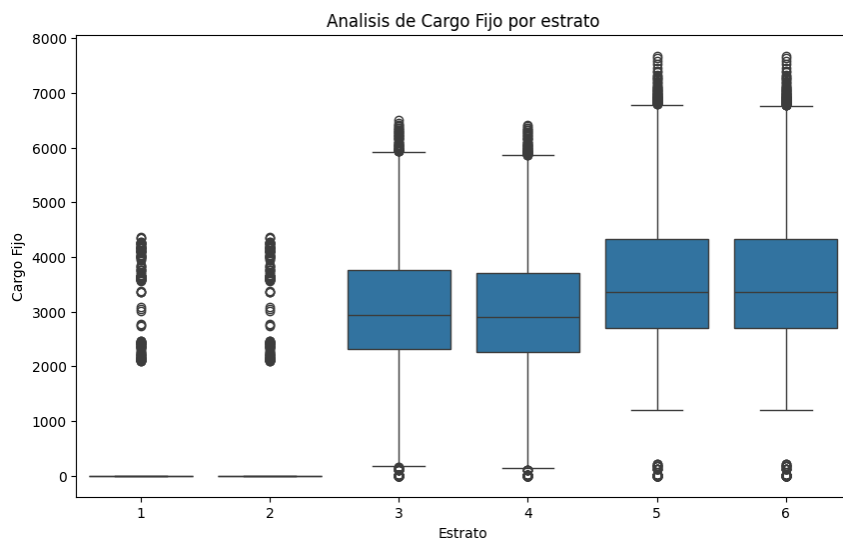
**Gráfico 3: Conexión de gas por estrato**

En el **Gráfico 4** se ilustra la evolución del cargo fijo por la conexión de gas natural en hogares a lo largo de los años 2013 a 2022; a partir de la gráfica, se puede observar una clara tendencia general al aumento del cargo fijo de conexión de gas a lo largo del tiempo para todos los estratos socioeconómicos. Esto sugiere que, independientemente del estrato, el costo de la conexión ha ido incrementándose año tras año y que los estratos socioeconómicos más altos, en este caso el estrato 5 y 6 consistentemente presentan los cargos fijos más elevados en comparación con los estratos más bajos 1 y 2. También se pueden notar ligeras variaciones en la magnitud del incremento de un año a otro, algunos años muestran aumentos más pronunciados que otros, lo que podría estar relacionado con factores económicos, regulatorios o de mercado.



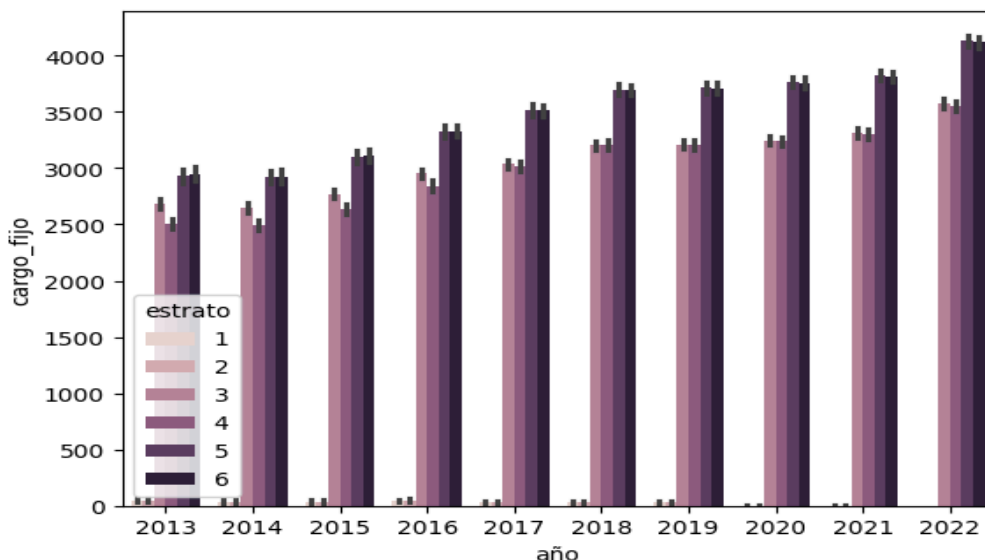
**Gráfico 4: Cargo fijo de la conexión de gas natural. Año 2013-2022**

Por otro lado, al realizar un comparativo del costo en el cargo fijo que se maneja por estrato socioeconómico; podemos encontrar una gran diferencia como se observa en el **Gráfico 5**, ya que los estratos 1 y 2 tienen un cargo fijo consistentemente muy bajo, donde su moda como se ve en el análisis descriptivo es de valor cero ya que una gran cantidad de usuarios en estos dos estratos están beneficiados por subsidios reglamentarios establecidos por la CREG, lo que justifica esta variación en la forma en que se da la caja, con una gran cantidad de valores atípicos dispersos. Mientras los estratos 3 al 6 muestran cargos fijos significativamente más altos y con una distribución más concentrada. También, podemos decir que a partir del estrato 3, se observa una tendencia general al aumento del cargo fijo promedio a medida que se avanza hacia los estratos superiores, donde los estratos 5 y 6 tienden a tener los cargos fijos más altos en general.



**Gráfico 5:** Cargo fijo de gas por estrato

Lo anteriormente descrito se puede confirmar en el siguiente gráfico de barras, que demuestra particularmente para el estrato 1 y 2 un costo de cargo fijo de valor cero a través de los años, sin embargo, a partir del estrato 3 en adelante, todos han tenido un costo que ha aumentado gradualmente año por año.



**Gráfico 6:** Cargo fijo de gas por estrato por cada año

Sin embargo, un costo de cargo fijo de valor cero, no representa que el servicio por completo es gratuito, ya que el consumo por metro cúbico del gas natural tiene costo para todos los estratos como se observa a continuación.

Teniendo en cuenta en el **Gráfico 7** el rango\_0 representa el costo para consumos de gas inferiores a 20 metros cúbicos y rango\_21 representa el costo para consumos de gas superiores a 20 metros cúbicos; podemos decir que hay un crecimiento de los costos asociados al consumo de gas natural entre los años 2013 y 2022 independiente del estrato socio económico, lo anterior puede atribuirse a muchos factores que pueden estar relacionados con la economía del país. Claramente se observa que los estratos 5 y 6, consistentemente pagan montos más elevados, tanto para consumos altos como para consumos bajos de gas.

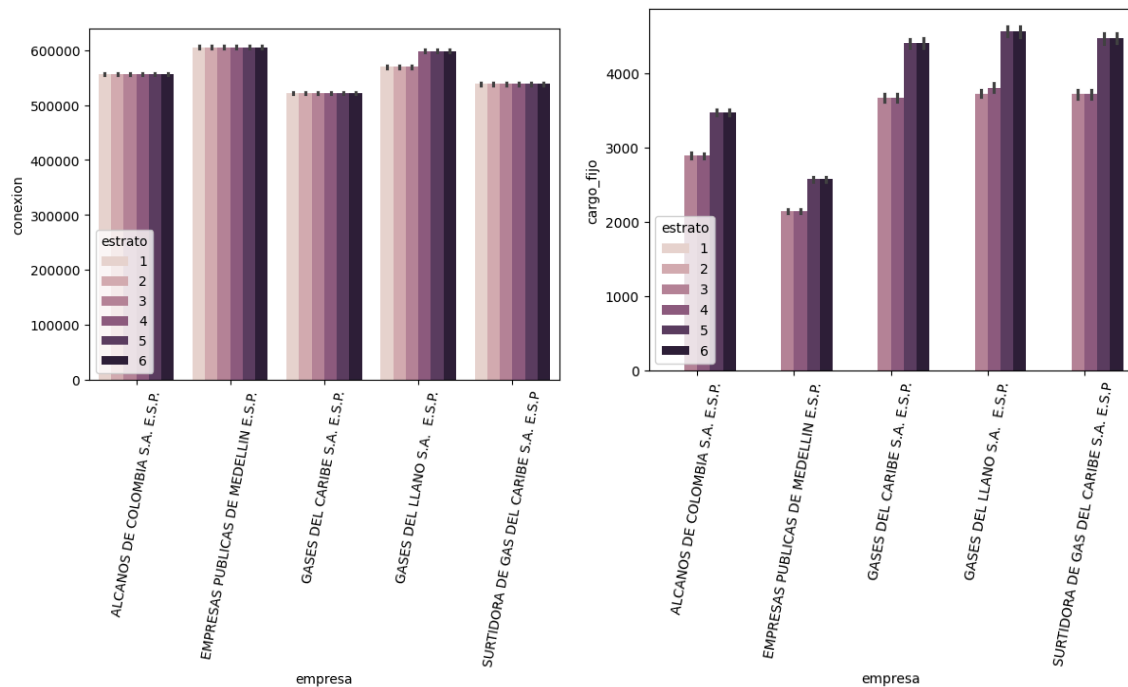








Hemos realizado un análisis de las 5 mejores empresas prestadoras del servicio de gas natural en diferentes regiones del país con relación a los costos asociados a conexión, cargo fijo, cargo 0 y cargo 21. El gráfico 10, nos da un contexto de lo mencionado anteriormente donde podemos observar diferencias notables en los niveles de precios entre las distintas empresas proveedoras de gas. La métrica de "conexión" no muestra una correlación clara con el estrato socioeconómico, ya que se mantiene relativamente constante entre los estratos dentro de cada empresa, lo que sugiere que este factor se rige por dinámicas diferentes, como la cobertura de la red o la densidad de población. Por otro lado, algunas empresas, como empresas públicas de Medellín, tienden a tener tarifas más bajas en general, mientras que otras, como alcanos de colombia y gases del llano, presentan tarifas más elevadas, especialmente para los estratos más altos. Estas diferencias podrían reflejar variaciones en los costos operativos, la eficiencia de la distribución, las políticas de inversión o las estrategias de precios de cada empresa. Lo anterior, explica que el mercado del gas natural en Colombia se ve fuertemente diferenciado por el estrato socioeconómico.







## Recomendaciones

- En general EPM(Empresas Públicas de Medellín) es la empresa que maneja los precios más accesibles para el consumidor, lo que convierte a las zonas donde EPM preste servicio, en zonas de alto interés para industrias que realicen alto consumo de gas natural para situarse.
- Para viviendas o ciertas empresas, es más estratégico económicamente, ubicarse en sectores o terrenos que se clasifiquen entre el estrato 1 y 2, debido a que son los estratos mayormente favorecidos por subsidios del gobierno para obtener costo cero en su cargo fijo y conexión, lo que a largo plazo representa una ganancia o un ahorro para el consumidor.
- En caso de formar una prestadora de servicio de gas natural, buscando mayor precio de venta, es estratégico ubicarse en las zonas donde presta el servicio Alcanos de Colombia (como los departamentos del Tolima, Huila, Caquetá, Cauca y la ciudad de Puerto Boyacá), debido a que en estas zonas el costo del servicio para el consumidor es el más elevado a nivel nacional.

## Enlace GitHub