

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













INTRODUCCIÓN

El gas natural es una fuente de energía fundamental en Colombia, utilizada tanto en sectores residenciales, industriales y de transporte. Se trata de un combustible fósil compuesto principalmente por metano, que se caracteriza por ser más limpio y eficiente en comparación con otros hidrocarburos como el carbón o el petróleo.

En el contexto colombiano, el gas natural ha cobrado especial relevancia en las últimas décadas debido a su disponibilidad local, su aporte a la diversificación de la matriz energética y su papel en la transición hacia fuentes de energía más sostenibles. Esta fuente energética no solo ha contribuido a mejorar la calidad del aire en las ciudades, sino también a fortalecer la seguridad energética del país.

En el caso de Colombia, el gas natural es un pilar fundamental de respaldo al sistema eléctrico, y su matriz energética se considera limpia, pues el 70% de la electricidad se genera con agua y el 30% restante con plantas térmicas, donde mayoritariamente se genera con gas. Sin embargo, estos porcentajes varían dependiendo de los niveles de los embalses.

A lo largo del siglo XX y lo que llevamos del XXI, el gas natural ha experimentado un crecimiento significativo en su adopción como fuente de energía para los hogares. Inicialmente utilizado principalmente para la iluminación y la cocción, su infraestructura se expandió para abarcar la calefacción y el agua caliente, transformando la vida doméstica y ofreciendo una alternativa a otras fuentes de energía.

Por ello, a través del análisis de una base de datos suministrada por la Comisión de Regulación de Energía y Gas CREG, este trabajo explora la evolución del crecimiento en la adquisición del servicio de gas natural a nivel nacional a lo largo de la década comprendida entre 2013 y 2022. Paralelamente, se examinan los costos vinculados a este servicio y sus fluctuaciones entre las distintas empresas que lo suministran.









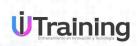




Tabla de contenido

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:	g
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 solido 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 Matplolib, 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.





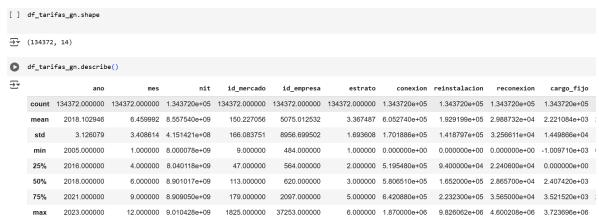




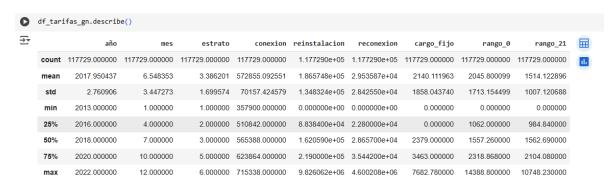


23





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



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.











() A



```
mode conexion = stats.mode(df_tarifas_gn['conexion'])
varianza_conexion = np.var(df_tarifas_gn['conexion'], ddof = 1)
                                                                print(f'La moda de la columna conexion es {mode conexion.mode}')
varianza_conexion
                                                                La moda de la columna conexion es 540798.0
4922064223.532464
                                                                mode_fijo = stats.mode(df_tarifas_gn['cargo_fijo'])
varianza fijo = np.var(df_tarifas_gn['cargo_fijo'], ddof = 1)
                                                                print(f'La moda de la columna cargo fijo es {mode_fijo.mode}'
varianza_fijo
                                                                La moda de la columna cargo fijo es 0.0
3452326.539569787
                                                                mode_rango0 = stats.mode(df_tarifas_gn['rango_0'])
varianza_rango0 = np.var(df_tarifas_gn['rango_0'], ddof = 1)
                                                                print(f'La moda de la columna rango 0 es {mode_rango0.mode}'
varianza_rango0
                                                                La moda de la columna rango 0 es 2015.64
2934898.336280019
varianza_rango21 = np.var(df_tarifas_gn['rango_21'], ddof = 1)
                                                                mode_rango21 = stats.mode(df_tarifas_gn['rango_21'])
                                                                print(f'La moda de la columna rango 21 es {mode_rango21.mode}'
varianza_rango21
1014292.0797878212
                                                                La moda de la columna rango 21 es 0.0
                      ## Curtosis cargo fijo
                      curtosis fijo = df tarifas gn["cargo fijo"].kurt()
                      print(f"La curtosis del cargo fijo es: {round(curtosis_fijo, 3)}")
                      La curtosis del cargo fijo es: -1.077
                      ## Curtosis conexion
                      curtosis_conexion = df_tarifas_gn["conexion"].kurt()
                      print(f"La curtosis del conexion es: {round(curtosis_conexion, 3)}")
                      La curtosis del conexion es: -0.29
                      ## Curtosis rango 0
```

```
La curtosis del rango 0 es: 3.826
```

La curtosis del rango 0 es: 9.761

Curtosis rango 21

curtosis_rango0 = df_tarifas_gn["rango_0"].kurt()

curtosis_rango21 = df_tarifas_gn["rango_21"].kurt()

print(f"La curtosis del rango 0 es: {round(curtosis_rango0, 3)}")

print(f"La curtosis del rango 0 es: {round(curtosis_rango21, 3)}")













() A



5. Interpretación de Resultados:

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

T _T 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 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.



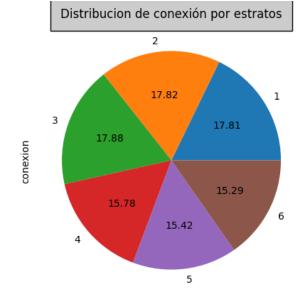


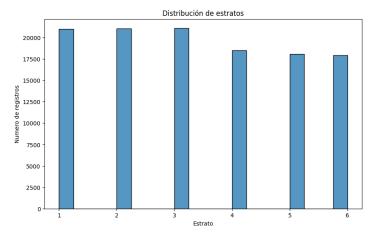












Gráfica 1: Distribución de conexión de gas por estrato y registro

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.











ස ල _අ



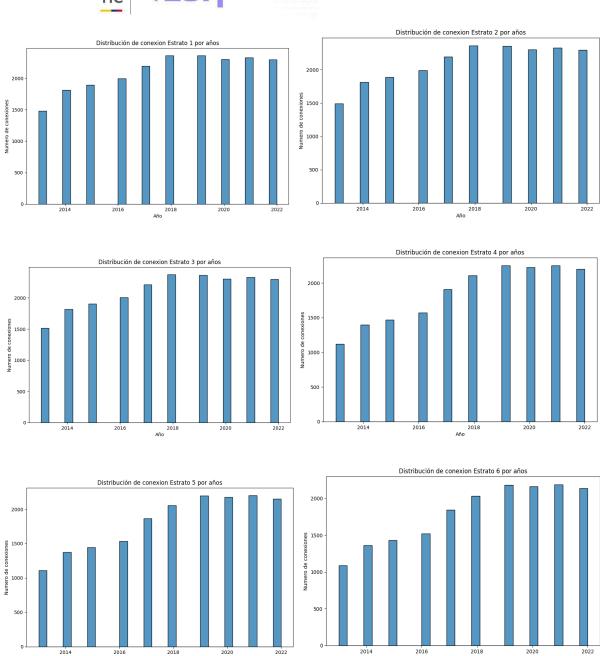


Gráfico 2: Conexión de gas por estrato y año

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.











ස ල _අ



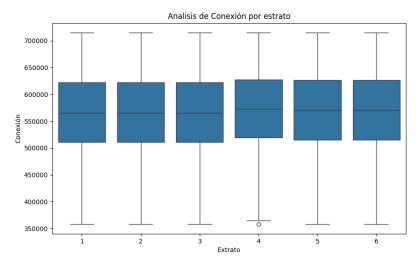


Gráfico 3: Conexión de gas por estrato

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.

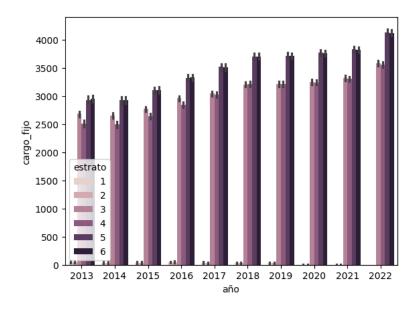


Gráfico 4: Cargo fijo de la conexión de gas natural. Año 2013-2022











ස ල _අ



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.

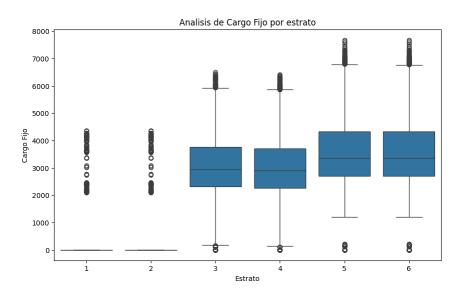


Gráfico 5: Cargo fijo de gas por estrato

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.











(C)



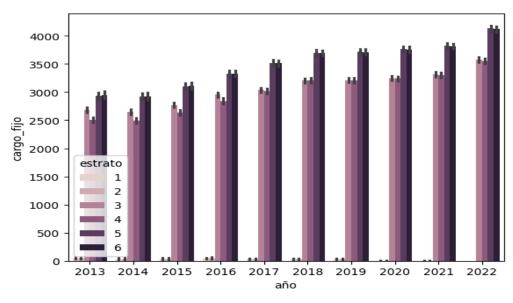


Gráfico 6: Cargo fijo de gas por estrato por cada año

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.



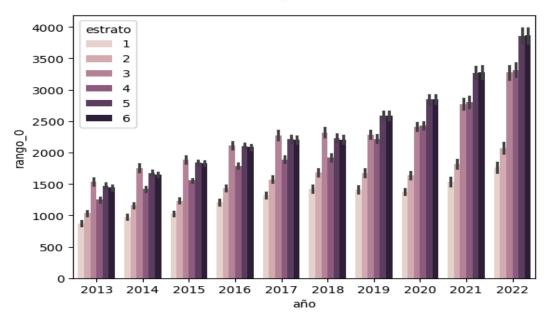












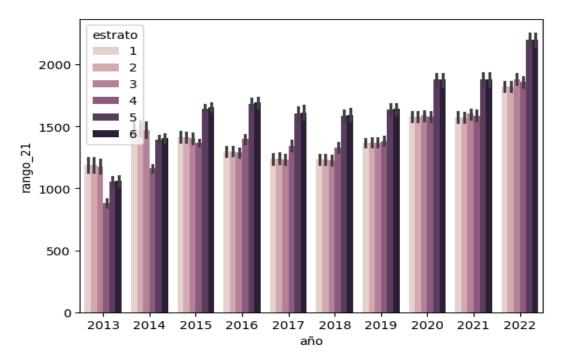


Gráfico 7. Consumo de gas por m³. Rango_0 (menor a 20 m³). Rango_21 (mayor a 20 m³)













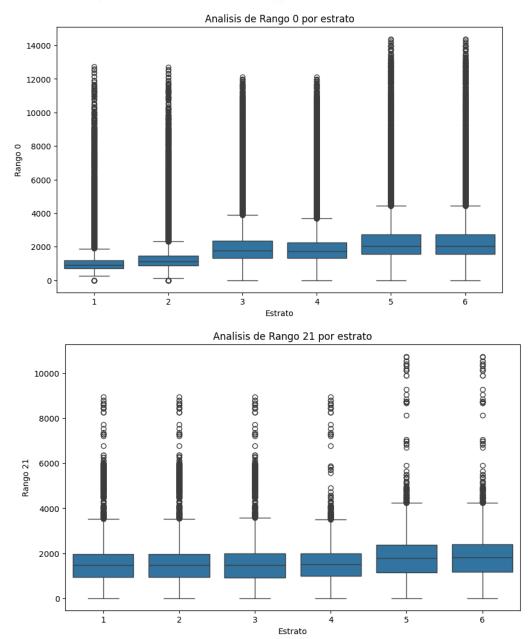


Gráfico 8. Caja de bigotes para gas por m³. Rango_0 (menor a 20 m³). Rango_21 (mayor a 20 m³)

En el **Gráfico 8**, se puede observar el rango 0 como la media de datos está en el rango de 1500-2200 para los diferentes estratos, sin embargo, con una cantidad considerable de valores atípicos fuera de los rangos Intercuartiles, lo cual está relacionado con costos más elevados por algunas empresas prestadoras del servicio o por zonas del país donde aumenta el costo del metro cúbico por reglamentación de la CREG y variables relacionadas al costo que conlleva prestar el servicio. Así mismo en el rango 21 la media está concentrada para los diferentes estratos, en un rango de











(D)



1700-2200, con media similar desde el estrato 1 al 4, y un aumento del estrato 5 al 6, por lo cual hay un costo medio de rango 21 similar entre los estratos en general.

El siguiente gráfico de correlación, **Gráfico 9**, nos muestra una relación entre las diferentes variables que hemos analizado. En la matriz, los colores cálidos (rojo, naranja) indican correlaciones positivas fuertes, los colores fríos (azul) indican correlaciones negativas fuertes, y los colores claros (blanco) indican correlaciones cercanas a cero.

Teniendo en cuenta lo anterior, podemos decir que hay una correlación positiva moderada a fuerte (color naranja) entre el estrato socioeconómico y el cargo fijo. Esto sugiere que los estratos socioeconómicos más altos tienden a tener cargos fijos más elevados lo cual se corrobora en los análisis anteriores. También se evidencia que hay una correlación positiva moderada (color azul claro) entre el estrato y el costo de conexión; aunque la correlación no es tan fuerte como con el cargo fijo, indica que los estratos más altos también tienden a tener costos de conexión ligeramente mayores.

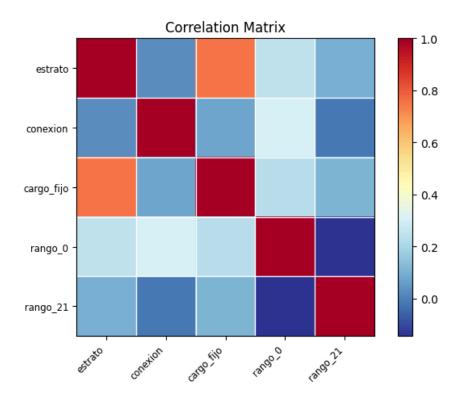


Gráfico 9. Matriz de correlación





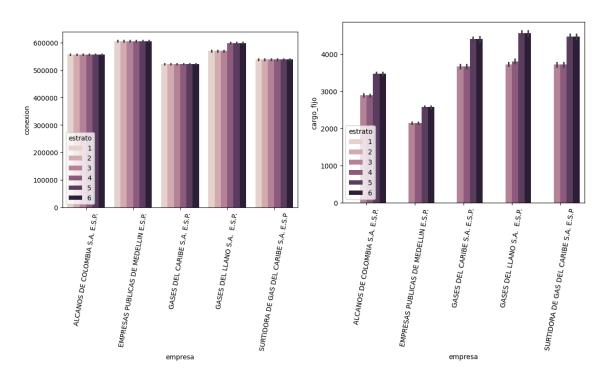








Hemos realizado un análisis de las 5 mejores empresas prestadoras del servicio de gas natural en diferentes regiones del país con relación a los costos asociados a conexión, cargo fijo, cargo 0 y cargo 21. El gráfico 10, nos da un contexto de lo mencionado anteriormente donde podemos observar diferencias 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.













(D)

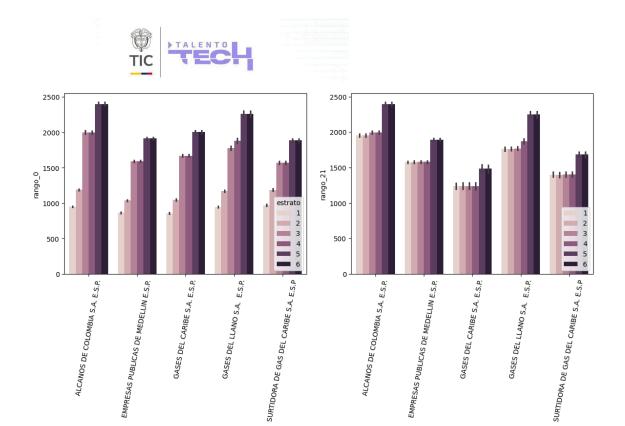


Gráfico 10. Relación del top 5 de empresas con respecto a la conexión, cargo fijo, rango 0 y rango 21

empresa

Realizando el respectivo análisis estadístico, se encuentran valores elevados correspondientes a la desviación estándar y a la varianza ya que se están analizando precios, los cuales han tenido diferentes cambios por la inflación del mercado anualmente, y algunas tarifas especiales aplicadas a estratos bajos (específicamente el estrato 1 y 2).

Como se puede evidenciar en la tabla, la curtosis para las variables: conexión y cargo fijo son menores a 3, por lo que son tipo platicúrticas, lo que significa que sus valores son màs cercanos a la media entre sí, por lo que sus colas son más delgadas, sin ser una distribución normal. Mientras que para las variables: Rango 0 y Rango 21 sus curtosis son mayores a 3, por tanto, son tipo leptocúrticas, lo que representa que sus valores están más dispersos y alejados de la media, por ende, sus colas son más pesadas que una distribución normal; Uno de los motivos por los que se obtuvieron estos resultados, es debido a que estas variables son precios, que cambian con respecto al tiempo, la inflación del mercado y el estrato socioeconómico.





empresa







(D)



	Conexión	Cargo fijo	Rango 0	Rango 21
media	572830.405	2138.905	2045.309	1513.792
desviación estándar	70153.847	1858.641	1712.978	1007.069
25 %	510842	0	1061.47	983.57
50 %	565388	2377.41	1556.82	1562.33
75 %	623864	3462	2318.77	2103.86
Moda	540798	0	2015.64	0
Varianza	4922064223.532	3452326.539	2934898.336	1014292.079
Curtosis	-0.29	-1.077	9.761	3.826

Tabla 1. Análisis estadístico de variables.

Conclusiones

- El aumento en el número de conexiones en todos los estratos refleja un crecimiento generalizado en la conectividad a lo largo del tiempo. Esto podría estar relacionado con la expansión de la infraestructura, la accesibilidad a la tecnología o el aumento de la demanda de servicios que requieren conexión.
- El estrato socioeconómico no es un factor determinante en el costo de la conexión de gas natural para las empresas en Colombia, lo que sugiere una política de precios de conexión relativamente uniforme.
- El gráfico 4, sugiere que el cargo fijo del gas natural para empresas en Colombia está fuertemente influenciado por el estrato socioeconómico, probablemente debido a políticas de subsidios, diferencias en el tipo de empresa o costos de infraestructura.
- El gráfico 6, proporciona una visión completa de la evolución de los costos de gas natural para diferentes rangos de consumo y estratos socioeconómicos; lo cual concluye en que, desde el año 2013 al 2022 estos se han visto caracterizados por el alza en sus precios debido a factores económicos y políticos del país.
- La matriz de correlación proporciona una visión cuantitativa de las relaciones entre las diferentes variables, respaldando y profundizando en las observaciones realizadas en los análisis previos de los gráficos de barras.
- El mercado de gas natural en Colombia se ve fuertemente diferenciado por el estrato socioeconómico, jugando un papel importante en la determinación de las tarifas, con variaciones significativas entre empresas que podrían reflejar diferentes realidades operativas y estratégicas.













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









