

Luego de haber preparado las planillas utilizando Excell, por ejemplo ordenando las columnas por año y trimestre de menor a mayor y eliminando aquellas columnas que no son relevantes, se realizó el análisis de los datos de cada una mediante el programa EDA.ipynb. También se hicieron predicciones de los datos, para esto se utilizó el programa prediccion.ipynb, donde se usó el método Holt-Winters. Para más información de como funcionan estos programas, ver los archivos ya que además del código se encontrará una explicación detallada de los mismos.

Estos fueron los resultados:

Planilla "para EDA internet.xlsx"

Hojas: "totales VMD", "Totales Accesos Por Tecnología", "Penetración totales", Totales Accesos por rango", "Ingresos"

Limpieza de las hojas: Resultado de la búsqueda de datos faltantes o duplicados:

--- Análisis de la hoja: Totales VMD ---

Cantidad de valores faltantes: 0

Cantidad de registros duplicados: 0

--- Análisis de la hoja: Totales Accesos Por Tecnología ---

Cantidad de valores faltantes: 0

Cantidad de registros duplicados: 0

--- Análisis de la hoja: Penetracion-totales ---

Cantidad de valores faltantes: 0

Cantidad de registros duplicados: 0

--- Análisis de la hoja: Totales Accesos por rango ---

Cantidad de valores faltantes: 0

Cantidad de registros duplicados: 0

--- Análisis de la hoja: Ingresos ---

Cantidad de valores faltantes: 0

Cantidad de registros duplicados: 0

Valores Atípicos:

Antes quiero aclarar que, al ser todas series temporales las estadísticas básicas (media, varianza, cuartiles, etc) no tiene sentido ser analizadas. Ya que lo que se busca con las series temporales es ver el comportamiento de una variable a lo largo del tiempo. Las estadísticas básicas nos brindan información general y resumida de un conjunto de datos, por ejemplo, si quiero saber cuál fue la cantidad de descarga promedio en determinado mes en determinada provincia. O cual fue el rendimiento de las empresas que brindan el servicio en determinada región y si existe alguna empresa que sobresalga del resto. Pero para responder estas preguntas necesitaríamos otro tipo de planillas las cuales no tenemos disponibles para este trabajo.

Sin embargo, es útil que veamos y realicemos unos gráficos tipo box plot, para ver si existen datos atípicos y cuáles pueden ser su origen.

El programa utilizado para el EDA brinda los gráficos para cada una de las columnas de cada hoja de la planilla, pero para este informe vamos a analizar los que arrojaron resultados positivos, es decir los que presentan estos datos atípicos

En Python la función que calcula el boxplot utiliza el método del Rango Inter cuartil (IQR) para identificar los valores atípicos. Este se calcula de la siguiente manera:

- Q1: Es el primer cuartil, también llamado percentil 25%, que representa el valor por debajo del cual se encuentra el 25% de los datos.
- Q3: Es el tercer cuartil (percentil 75%), que representa el valor por debajo del cual se encuentra el 75% de los datos.
- $IQR=Q3-Q1$: Este es el rango Inter cuartil, que mide la dispersión del 50% central de los datos.

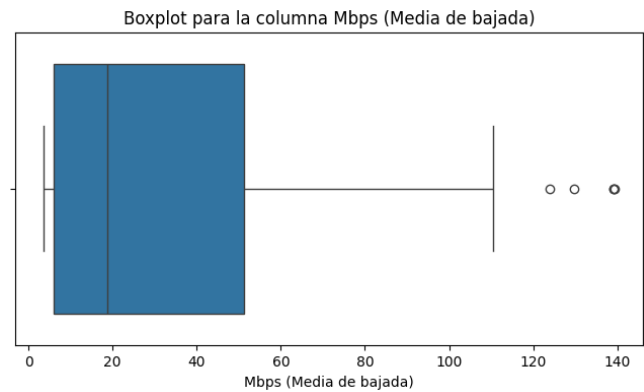
Para detectar los valore atípicos se calculan los limites inferior y superior de la siguiente manera:

- Límite inferior: $Q1-1.5 \times IQR$
- Límite superior: $Q3+1.5 \times IQR$

Finalmente, un dato se considera atípico si:

- Es menor que el límite inferior.
- Es mayor que el límite superior.

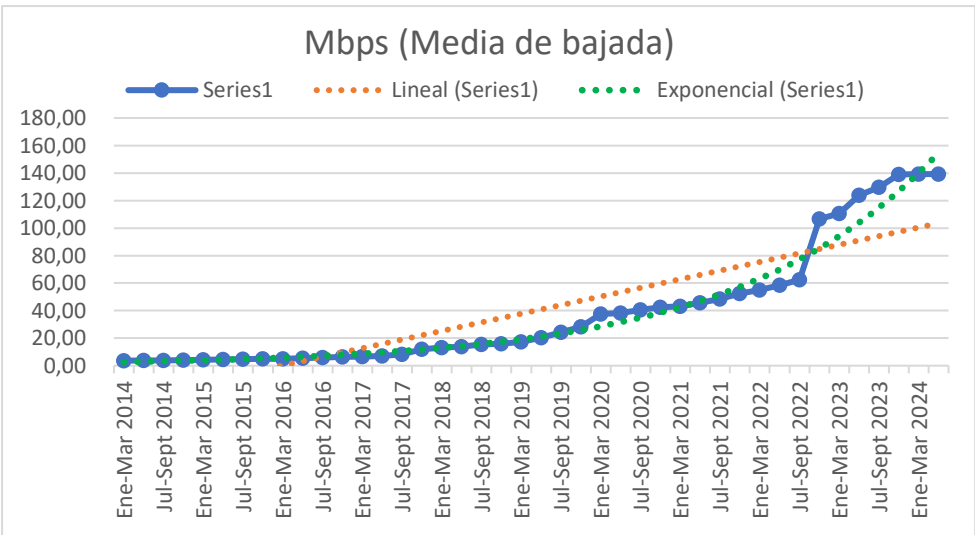
Cantidad de valores atípicos en la columna 'Mbps (Media de bajada)': 5



Analizando estos datos los valores atípicos tenemos que corresponden a los últimos años:

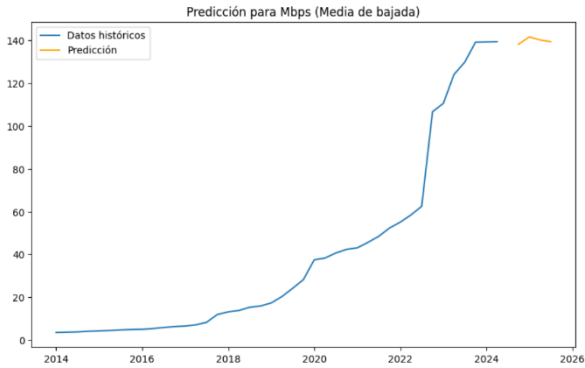
Año	Trimestre	Mbps
2024	2	139,25
2024	1	139,15
2023	4	139,04
2023	3	129,67
2023	2	123,95

Si vemos la tendencia de estos datos



Observamos que siguen una tendencia exponencial, con un aumento marcado entre el tercer trimestre y el cuarto trimestre del 2022, casi duplicando su valor. Por lo que no serán eliminados del análisis.

Haciendo una predicción de los mismos, nos sugiere que esta tendencia se va a estabilizar alrededor de 140 Mbps durante los próximos 4 trimestres



Predicciones para Mbps (Media de bajada):

2024-09-30	138.071042
2024-12-31	141.486250
2025-03-31	140.079309
2025-06-30	139.250000

--- Análisis de la hoja: Totales Accesos Por Tecnología ---
Estos son los resultados del análisis de los valores atípicos

Cantidad de valores atípicos en la columna 'ADSL': 0
Cantidad de valores atípicos en la columna 'Cablemodem': 0
Cantidad de valores atípicos en la columna 'Fibra óptica': 0
Cantidad de valores atípicos en la columna 'Wireless': 0
Cantidad de valores atípicos en la columna 'Otros': 0
Cantidad de valores atípicos en la columna 'Total': 0

Del análisis realizado se observan que no hay valores atípicos en esta hoja. Se presenta los gráficos BoxPlot

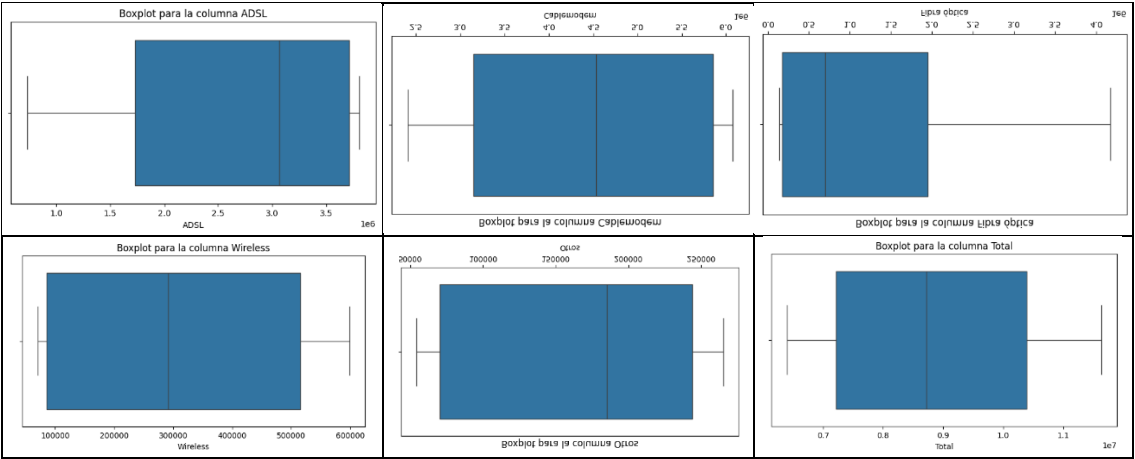


Ilustración 1: gráficas boxplot de las columnas de la hoja Totales Accesos Por Tecnología para ver si existen datos atípicos.

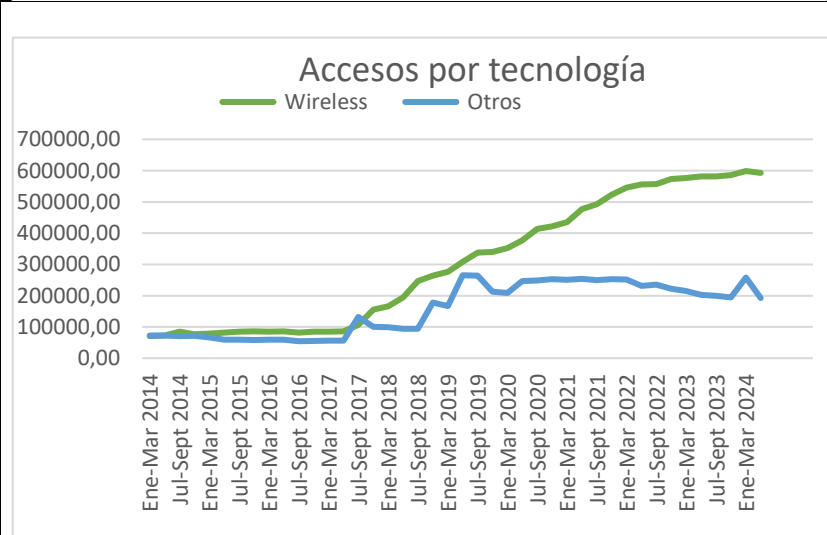
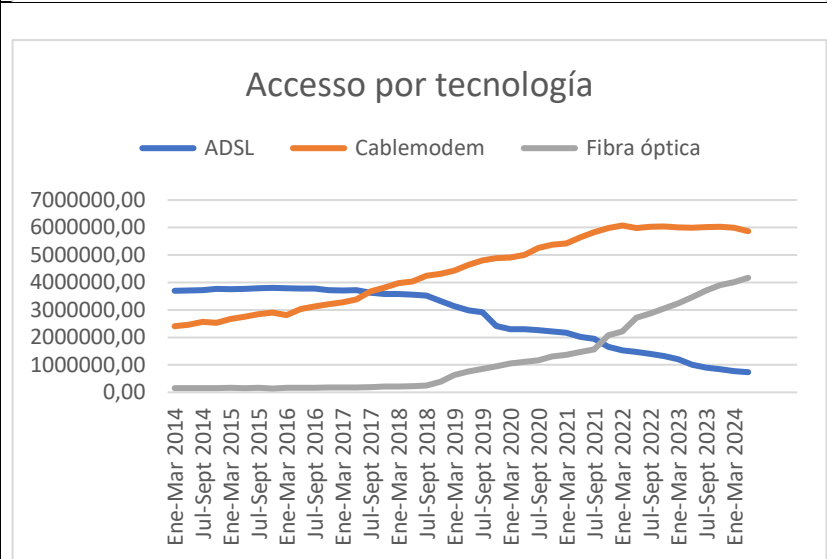
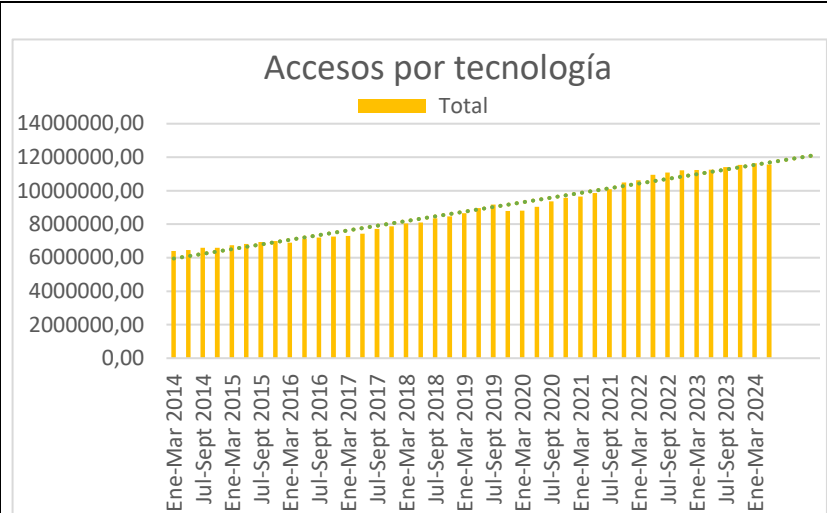
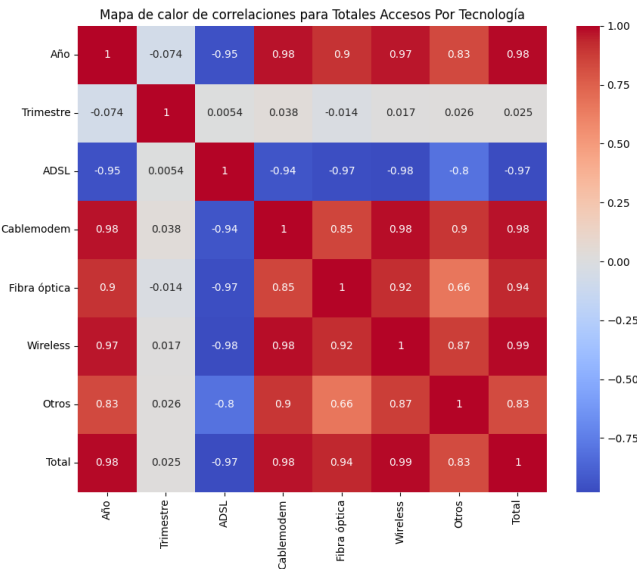
Observemos ahora las series temporales saquemos algunas conclusiones. Para las gráficas se utilizó el programa Excel y se agrupó las variables de acuerdo a la magnitud de sus valores. Si observamos el gráfico que corresponde al Total de accesos (Ilustración 4 arriba) vemos como los accesos por tecnología aumentan de manera lineal. En los gráficos se observa como la tecnología ADSL está perdiendo terreno respecto a las otras tecnologías. Esto es lógico si recordamos que ADSL corresponde a internet por telefonía fija, la cuales está en desuso desde

la aparición de los celulares. Se observa que tanto el cablemódem, la Fibra óptica como el Wireless están aumentando el número de accesos, la diferencia es que el cablemódem parece haber alcanzado un estancamiento alrededor de los 6 millones de accesos, mientras que la conexión a través de Wireless parece seguir en alza, aunque con una desaceleración. La tecnología que está creciendo es la fibra óptica.

La explicación es que esta es una tecnología reciente, comienza a desarrollarse en el país a partir del segundo semestre del 2018 y es la que proporciona un ancho de banda mayor y la que brinda mayor velocidad.

Respecto a Otras tecnologías, hay que destacar la conexión a través de satélites, la primer empresa en ofrecer este tipo de tecnología fue Orbith (El Cronista, 2024) y desde el 2024 la conocida empresa Starlink de Elon Musk comenzó a ofrecer su servicio en la Argentina (Clarín, 2023). Por lo que de ofrecer el servicio en Argentina lo recomendable es ofrecer internet mediante fibra óptica, en el caso de las zonas urbanas y Satelital para las zonas rurales.

Para terminar la perdida de conexión de ADSL respecto a las otras tecnologías se ve reflejada en el Mapa de Calor donde se representan las correlaciones. Si observamos la ilustración que sigue:



Vemos que la columna (o fila) correspondiente al ADSL tiene correlación negativa respecto a las demás tecnologías, esto indica que cuando las demás variables crecen, el ADSL toma valores menores. También vemos que las demás tecnologías tienen un comportamiento similar, es decir todas aumentan de la misma manera. Para explicar esto miremos que la

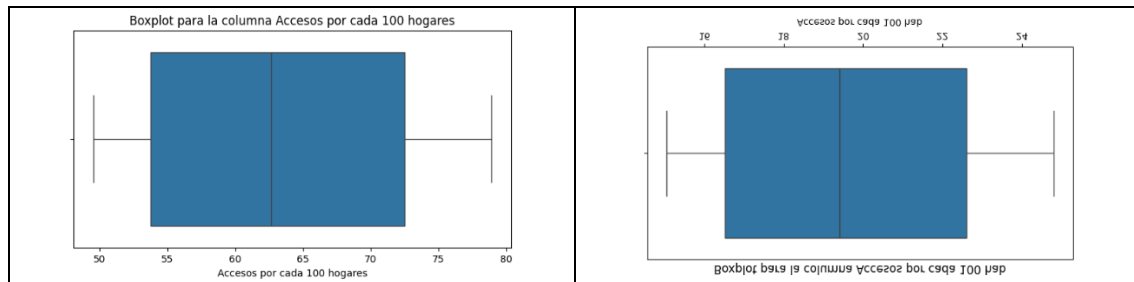
gráficas de Wireless y Cablemódem son muy parecidas, y su índice de correlación es de 0.98. El de estos con Fibra Óptica es de 0.9 aproximadamente, lo que indica que si bien la tendencia es la misma sus gráficas no son tan parecidas.

--- Análisis de la hoja: Penetración-totales ---

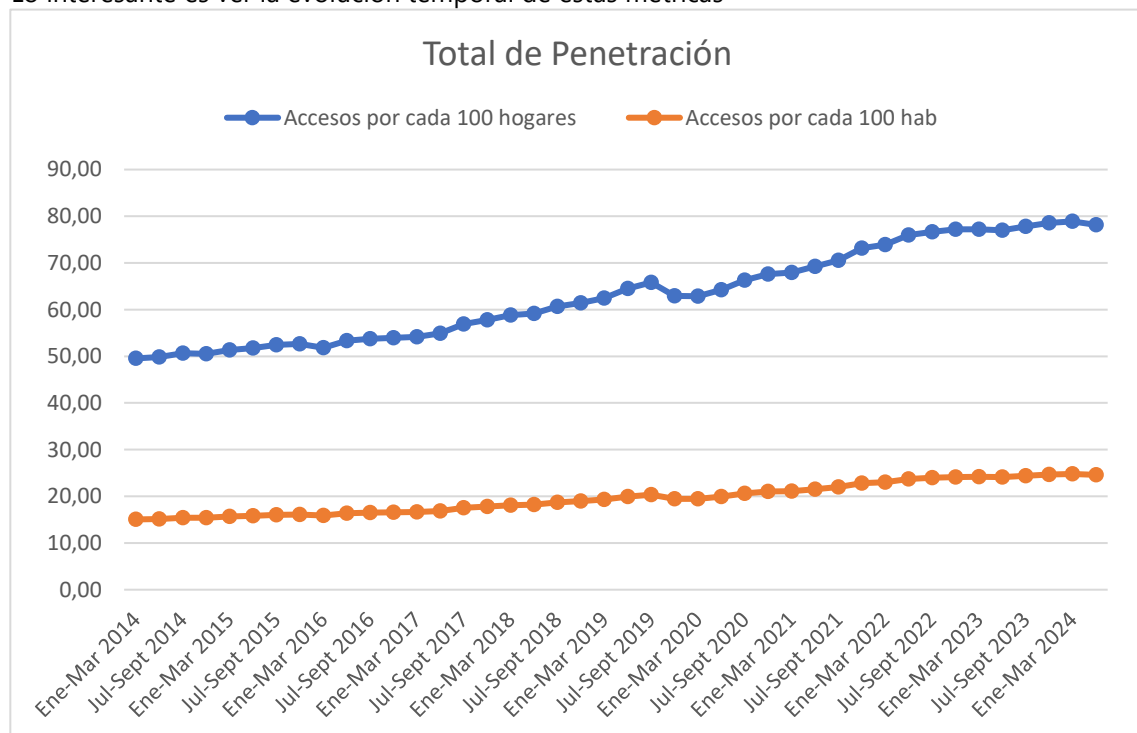
Cantidad de valores atípicos en la columna 'Accesos por cada 100 hogares': 0

Cantidad de valores atípicos en la columna 'Accesos por cada 100 habitantes': 0

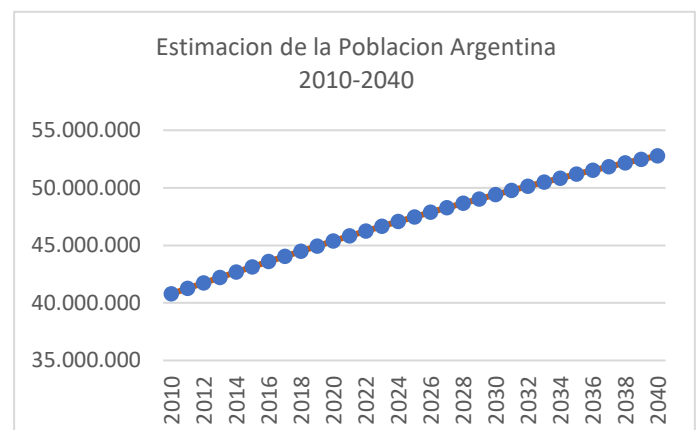
Veamos los Boxplot



Lo interesante es ver la evolución temporal de estas métricas



En ambas gráficas se observa un aumento constante de accesos tanto por hogares como por habitantes. Si hacemos una comparación del crecimiento de la población en la Argentina, la página del INDEC brinda una de planilla con la población estimada al 1 de julio de cada año calendario desde 2010 a 2040 (<https://www.indec.gob.ar/indec/web/Nivel4-Tema-2-24-84>) que indica que la población va en aumento de manera similar a lo que lo hace la penetración. Si calculamos el coeficiente de



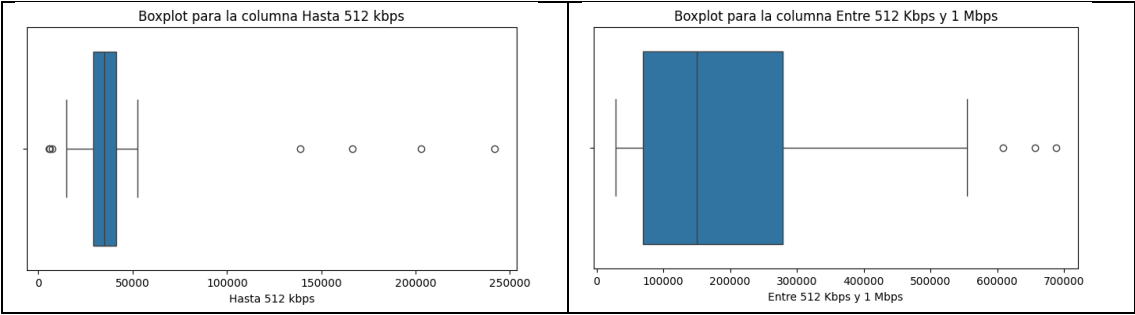
correlación entre la población y la penetración, tomando los años en común y el tercer trimestre en el caso de la penetración tenemos que el coeficiente de correlación tanto con la penetración por hogares como de la población es de 0,99, por lo que se espera que el acceso a internet siga en aumento. Por último cabe aclarar que las estimaciones y proyecciones fueron elaboradas en base a resultados del Censo Nacional de Población, Hogares y Viviendas 2010, si tenemos en cuenta que la estimación para el 2022 es de 46234830 y el resultado del censo del 2022 nos da que la población Argentina es de 45892285 tenemos que los valores pudieron ser sobre estimados, aunque esto no incluye en el análisis realizado.

--- Análisis de la hoja: Totales Accesos por rango ---

Presentamos las columnas que tienen valores atípicos, junto a su BoxPlot, luego haremos un análisis de cada uno de ellos.

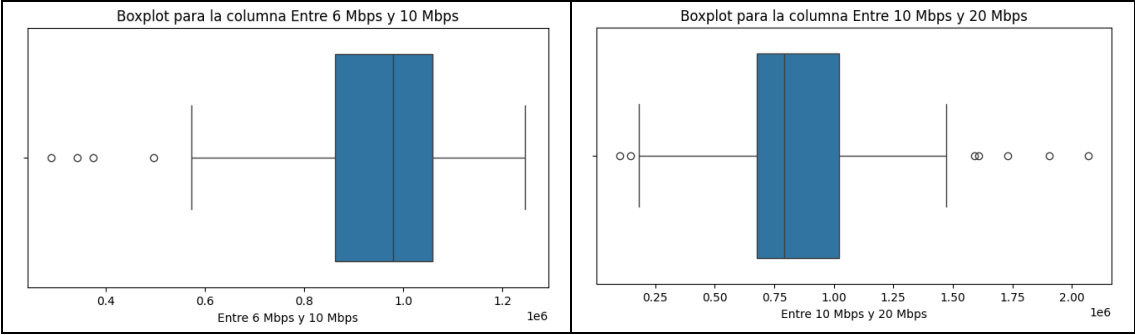
Cantidad de valores atípicos en la columna 'Hasta 512 kbps': 7

Cantidad de valores atípicos en la columna 'Entre 512 Kbps y 1 Mbps': 3

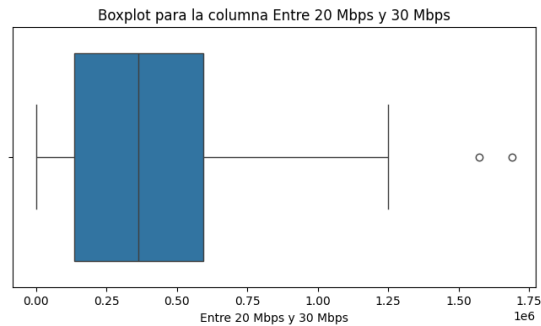


Cantidad de valores atípicos en la columna 'Entre 6 Mbps y 10 Mbps': 4

Cantidad de valores atípicos en la columna 'Entre 10 Mbps y 20 Mbps': 7



Cantidad de valores atípicos en la columna 'Entre 20 Mbps y 30 Mbps': 2



Cantidad de valores atípicos en la columna 'Entre 1 Mbps y 6 Mbps': 0

Cantidad de valores atípicos en la columna 'Más de 30 Mbps': 0

Cantidad de valores atípicos en la columna 'OTROS': 0

Cantidad de valores atípicos en la columna 'Total': 0

Veamos cuales son los valores atípicos y en qué fecha se dieron, en rojo están los que se encuentran por debajo del límite inferior y en azul oscuro los que están por encima del límite superior.

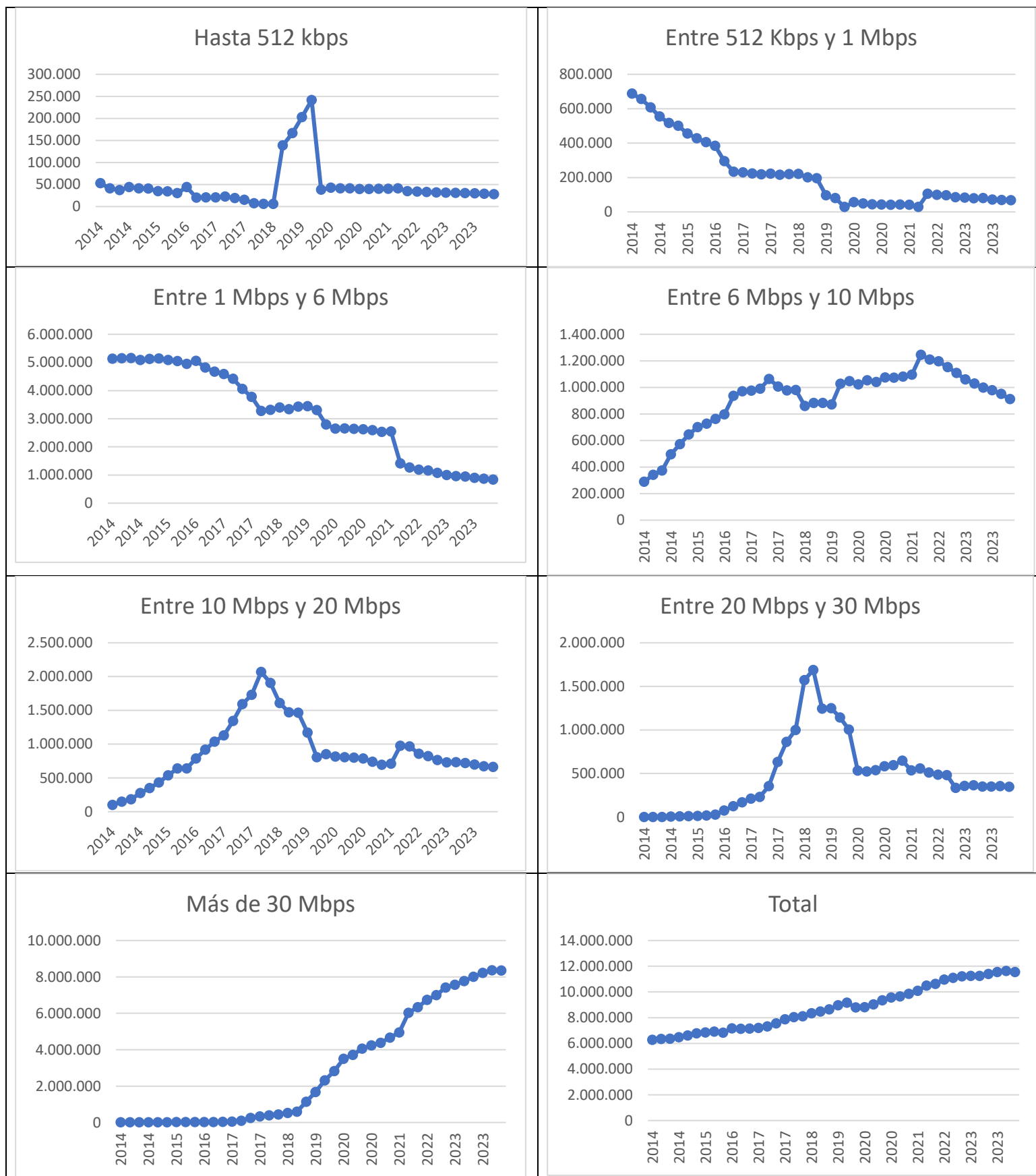
Año	Trimestre	Hasta 512 kbps	Año		Trimestre	Entre 512 Kbps y 1 Mbps
2018	3	5675	2014		3	608018
2018	2	5972	2014		2	656408
2018	1	7309	2014		1	687619
2018	4	138740				
2019	1	166573				
2019	2	202790				
2019	3	241713				

Año	Trimestre	Entre 6 Mbps y 10 Mbps	Año	Trimestre	Entre 10 Mbps y 20 Mbps
2014	1	289182	2014	1	101127
2014	2	341689	2014	2	147273
2014	3	373372	2017	3	1592304
2014	4	496251	2018	3	1607137
			2017	4	1730353
			2018	2	1905839
			2018	1	2068087

Año	Trimestre	Entre 20 Mbps y 30 Mbps
2018	3	1571692
2018	4	1688748

Haciendo un análisis de las tablas, los datos que dan mayor duda son los de la columna “Hasta 520 Mbps” ya que los datos que están por debajo del límite inferior se produjeron en el primero segundo y tercer trimestre del año 2018 mientras que en el 4 trimestre el mismo año se produce el primer de los cuatro datos que están por encima del límite superior. Esto puede sugerir que haya habido un error en la carga de los mismos. Respecto a los valores atípicos de las columnas “Entre 512 Kbps y 1 Mbps” y “Entre 6 Mbps y 10 Mbps” vemos que los valores máximos de la primera corresponden a los valores mínimos de la segunda, esto se podría explicar a que en el año 2014 se mejoró la comunicación por lo los usuarios pudieron “navegar” más rápido por internet. Algo parecido sucede en los años 2018, donde se producen los máximos de acceso de internet “Entre 10 Mbps y 20 Mbps” y entre “Entre 20 Mbps y 30 Mbps”, si vemos el análisis de la hoja “Totales Accesos Por Tecnología” en este año comenzó a popularizarse el Wireless y la Fibra Óptica, lo que permitió mayor velocidad de conexión.

Pasemos a analizar las series temporales para ver si van en línea con el párrafo anterior.

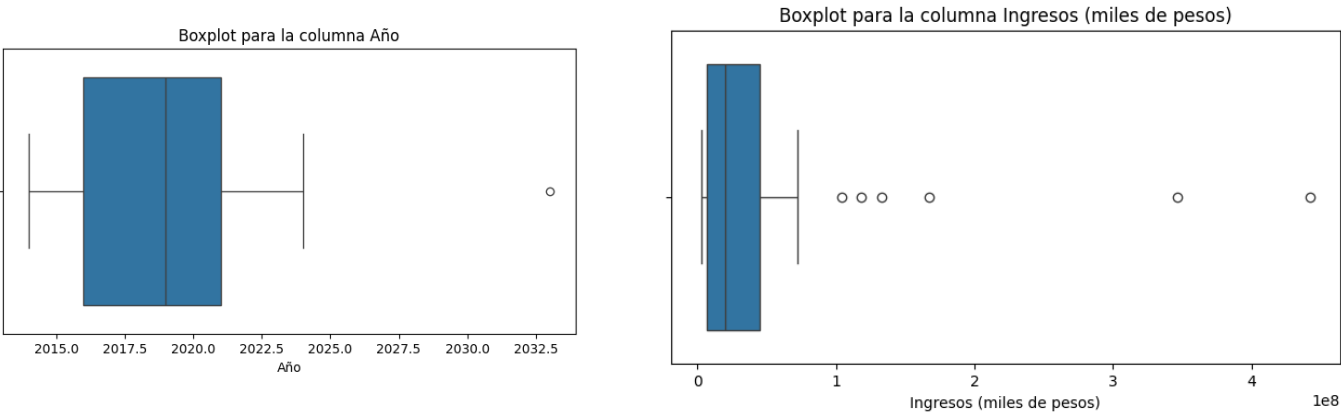


Vemos en las gráficas que se corrobora lo analizado en el párrafo de los datos atípicos ya que a partir de 2017 – 2018 los accesos tienen un crecimiento, podríamos decir lineal con una

pendiente muy pronunciada, haciendo que los demás rangos bajen de manera considerable, principalmente el rango que corresponde a los 20 a 30 Mbps que en el 2017 tuvo un aumento significativo pero ese mismo aumento se reflejo en caída hasta el año 2020. De una nota del diario “La Nación” se puede leer lo siguiente “En 2016, la velocidad era de aproximadamente 5 Mbps y en 2015 de alrededor de 4 Mbps; mientras que las proyecciones para el cierre del corriente año estiman que se alcanzará un promedio de 20 Mbps.” (La Nación, 2019). Sin embargo, no encontré datos que puedan explicar los valores atípicos observados en el análisis de la columna “hasta 512 Kbps” por lo que no hay que descartar la posibilidad de que hubo un error al cargar la planilla.

--- Análisis de la hoja: Ingresos ---

Cantidad de valores atípicos en la columna 'Año': 1
Cantidad de valores atípicos en la columna 'Ingresos (miles de pesos)': 6



Lo primero para decir es que este análisis me permitió descubrir un error en la tabla, ya que el valor atípico en la columna Año corresponde a que cargaron el año 2033 en vez de 2023 como se observa en la tabla a continuación.

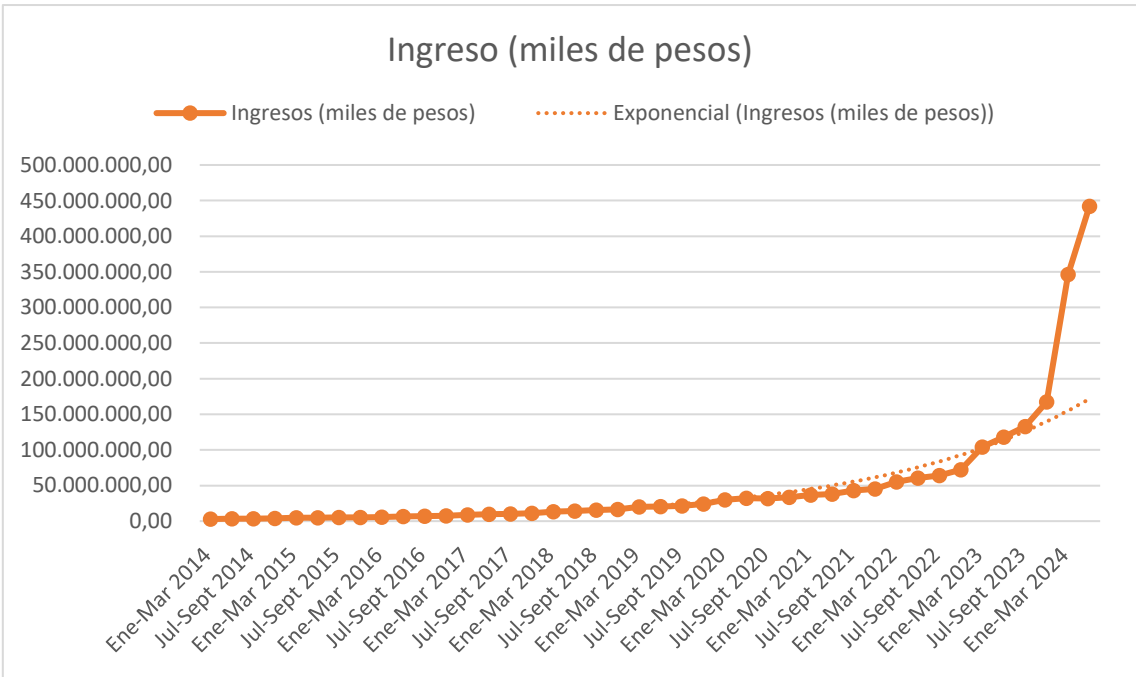
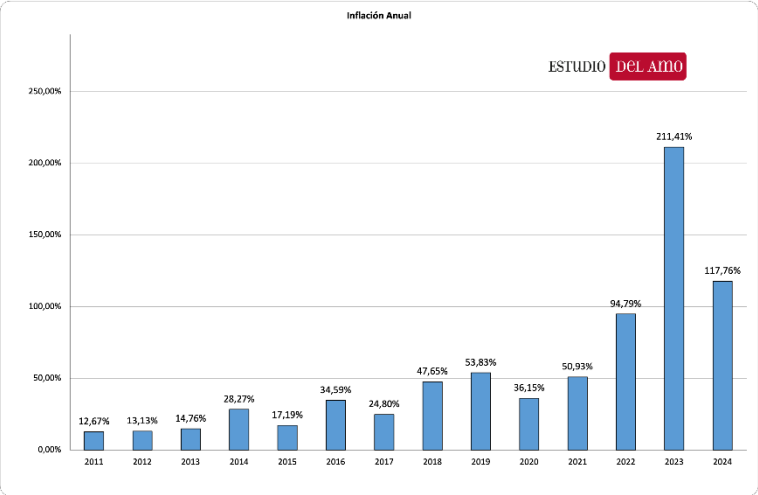
2023	4	167.376.014,80	Oct-Dic 2023
2023	3	133.106.593,41	Jul-Sept 2023
2023	2	118.060.280,28	Jun-Mar 2023
2033	1	104.134.992,81	Ene-Mar 2023
2022	4	72.443.482,99	Oct-Dic 2022
2022	3	64.086.784,33	Jul-Sept 2022
2022	2	60.419.500,62	Abr-Jun 2022
2022	1	55.219.313,01	Ene-Mar 2022

Corregido de manera manual este error sigamos analizando la columna Ingresos, vemos que tiene seis valores atípicos por encima del límite superior. Veamos a que fechas corresponden

2024	2	442.032.166,69	Abr-Jun 2024
2024	1	346.198.986,13	Ene-Mar 2024
2023	4	167.376.014,80	Oct-Dic 2023
2023	3	133.106.593,41	Jul-Sept 2023
2023	2	118.060.280,28	Jun-Mar 2023
2033	1	104.134.992,81	Ene-Mar 2023

Todos corresponden a los últimos años analizados en el dataset. Esto es algo esperable si se considera que la ganancia está en miles de pesos y considerando el aumento constante del acceso a internet analizado anteriormente y los datos de inflación vemos un aumento de la inflación en el año 2023, (ver imagen derecha, fuente: Estudio Del Amo) lo que llevo a un aumento de las tarifas de los servicios de los que internet no quedo afuera de este aumento.

Veamos la serie temporal de los ingresos



Vemos que, a pesar de lo analizado en el párrafo anterior, los ingresos tienen un crecimiento exponencial. Por lo que es altamente recomendable una inversión en este negocio.

Continuemos ahora con las otras planillas utilizadas, aclarando que el análisis no será tan exhaustivo ya que estas son utilizadas como complemento del análisis anterior.

Análisis para la Planilla "Portabilidad.xlsx"

Esta planilla solo tiene una hoja, analicemos el EDA de la misma.

--- Análisis de la hoja: Portin ---

Cantidad de valores faltantes antes de la limpieza: 0

Cantidad de registros duplicados antes de la limpieza: 0

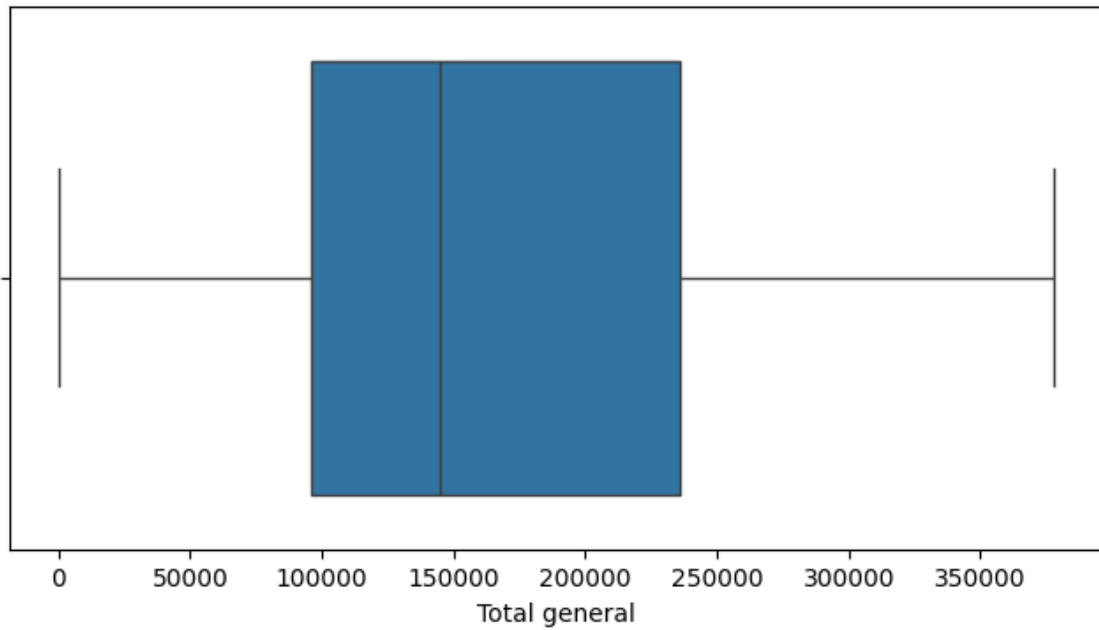
Cantidad de valores faltantes eliminados: 0

Cantidad de valores atípicos en la columna 'Año': 0

Cantidad de valores atípicos en la columna 'Mes': 0

Cantidad de valores atípicos en la columna 'Total general': 0

Boxplot para la columna Total general

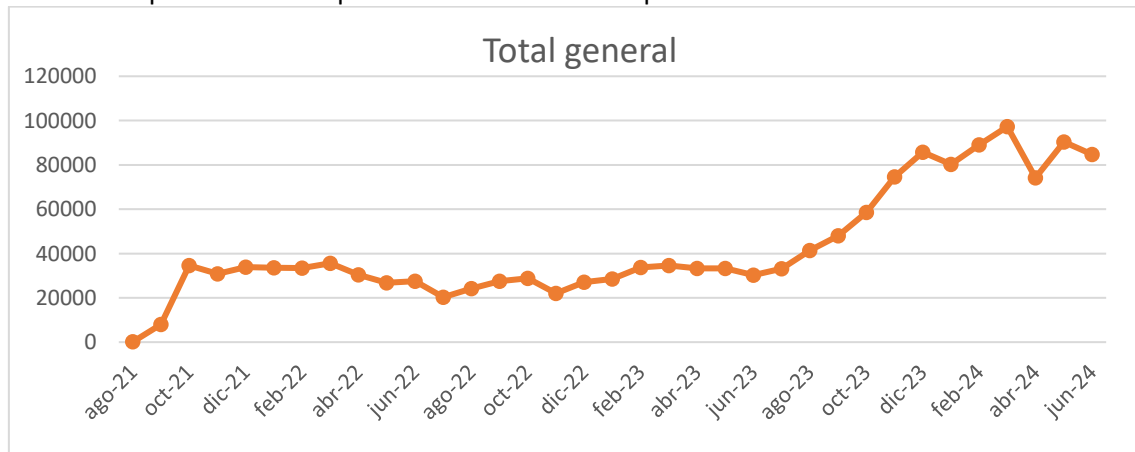


Para esta variable si tiene sentido analizar los estadísticos básicos:

Resumen estadístico:

```
general
count    148
mean    159261
std     95330
min      153
25%    95727.25
50%   144530.50
75%   235856.75
max   377682
```

Se observa que en promedio mensualmente hay unos 160000 cambios de compañía y miramos la línea temporal tenemos que en los últimos años la portabilidad va en aumento



Haciendo una predicción para los próximos meses tenemos que esta tendencia sigue en alza

Predicciones para Total general:

```
2024-02-29  169073
2024-03-31  161526
2024-04-30  184107
2024-05-31  186081
```

Esto es positivo para una empresa que quiere desarrollarse en el país ya que teniendo en cuenta que el acceso casa 100 habitantes es del 25%, es decir 25 personas de cada 100 tienen acceso a internet y teniendo en cuenta que, la población de Argentina, según el censo del 2022, es de 46234830, tenemos que alrededor de 11.558.707 tienen acceso a internet, de estas cerca del 2% cambia de compañía cada mes.

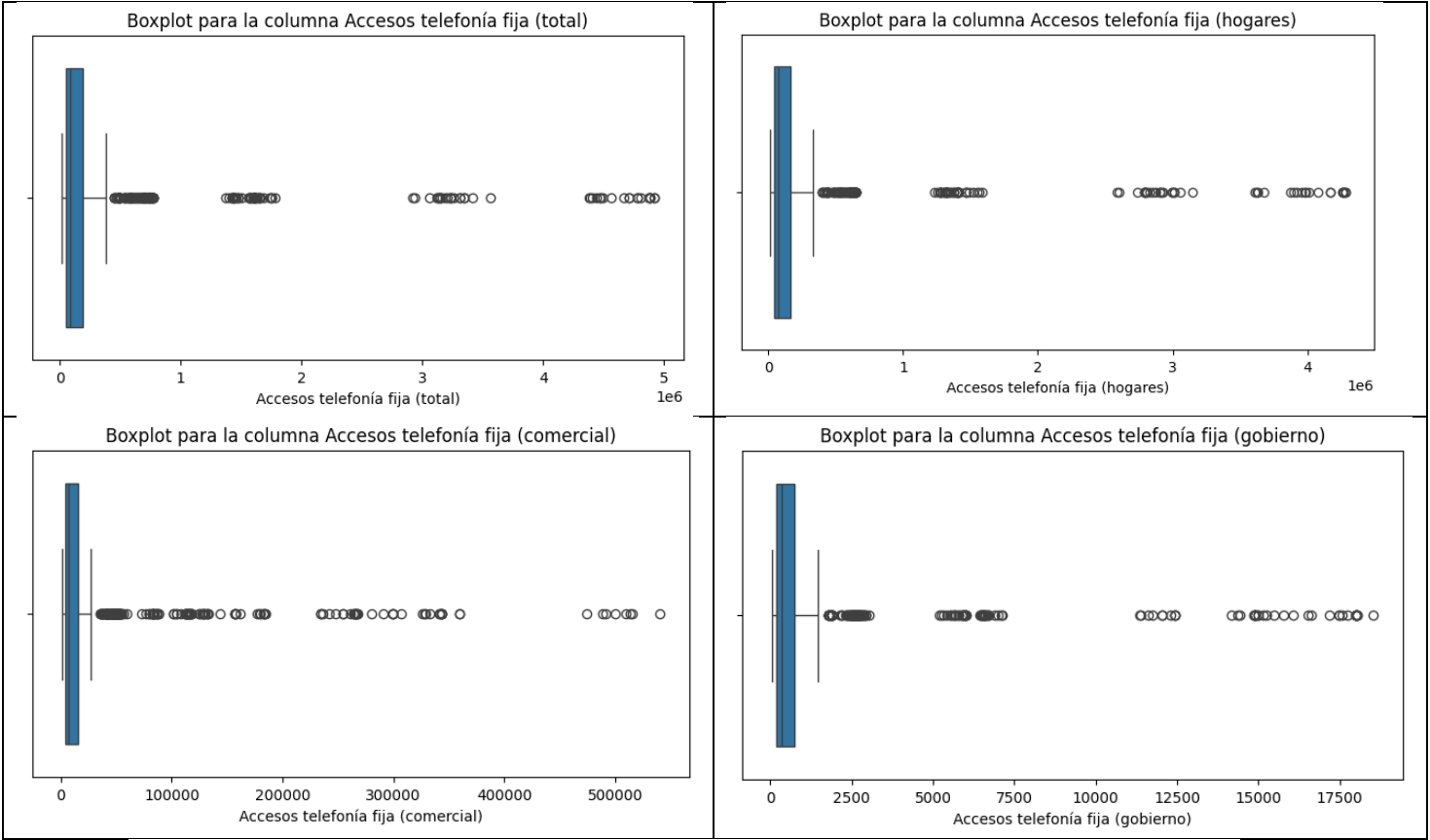
Planilla Telefonía Fija

Esta planilla tiene solo la hoja “Fija_prov_tipo” la cual contiene siete columnas: Año, Trimestre, Provincia, Accesos telefonía fija (total), Accesos telefonía fija (hogares), Accesos telefonía fija (comercial), Accesos telefonía fija (gobierno), Accesos telefonía fija (otros). Para este análisis se borro la columna Provincia ya que no es relevante para el análisis. El resto de columnas está formada por números enteros y representa a la cantidad de accesos respecto a entidad, es decir Hogares, comercial, gobierno, la columna total es la suma de las demás.

--- Análisis de la hoja: Fija_prov_tipo ---
Cantidad de valores faltantes antes de la limpieza: 0
Cantidad de registros duplicados antes de la limpieza: 0
Cantidad de valores faltantes eliminados: 0

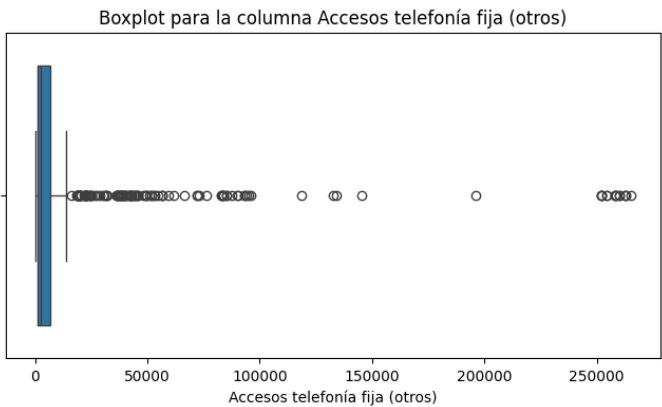
Con los valores atípicos hay una particularidad, veamos el análisis realizado y junto con la ayuda de algunas gráficas saquemos conclusiones:

Cantidad de valores atípicos en la columna 'Accesos telefonía fija (total)': 168
Cantidad de valores atípicos en la columna 'Accesos telefonía fija (hogares)': 168
Cantidad de valores atípicos en la columna 'Accesos telefonía fija (comercial)': 176
Cantidad de valores atípicos en la columna 'Accesos telefonía fija (gobierno)': 168
Cantidad de valores atípicos en la columna 'Accesos telefonía fija (otros)': 143



Tenemos que estos valores atípicos corresponden a la provincia de Buenos Aires y a la Capital Federal principalmente pero también a Córdoba y Santa Fe. Estas son las provincias más pobladas de la Argentina

Jurisdicción	Año 2020	Porcentaje
República Argentina	45.376.763	
Buenos Aires	17.541.141	38,7
Córdoba	3.760.450	8,3
Santa Fe	3.536.418	7,8
ad Autónoma de Buenos A	3.075.646	6,8
Mendoza	1.990.338	4,4
Tucumán	1.694.656	3,7
Salta	1.424.397	3,1
Entre Ríos	1.385.961	3,1
Misiones	1.261.294	2,8
Chaco	1.204.541	2,7
Corrientes	1.120.801	2,5
Santiago del Estero	978.313	2,2
San Juan	781.217	1,7
Jujuy	770.881	1,7
Río Negro	747.610	1,6
Neuquén	664.057	1,5
Chubut	618.994	1,4
Formosa	605.193	1,3
San Luis	508.328	1,1
Catamarca	415.438	0,9
La Rioja	393.531	0,9
Santa Cruz	365.698	0,8
La Pampa	358.428	0,8
Tierra del Fuego	173.432	0,4
Fuente IGN		



Plantilla Televisión:

Esta es la última planilla utilizada para el análisis además de Año y Trimestre tiene dos columnas de datos significativos, estas son el acceso de TV por suscripción y por satélite. Esta tabla se uso para ver el progreso temporal de este medio de comunicación y compararlo con internet.

EL análisis EDA es el siguiente.

--- Análisis de la hoja: Accesos_totales_TV ---

Cantidad de valores faltantes antes de la limpieza: 0

Cantidad de registros duplicados antes de la limpieza: 0

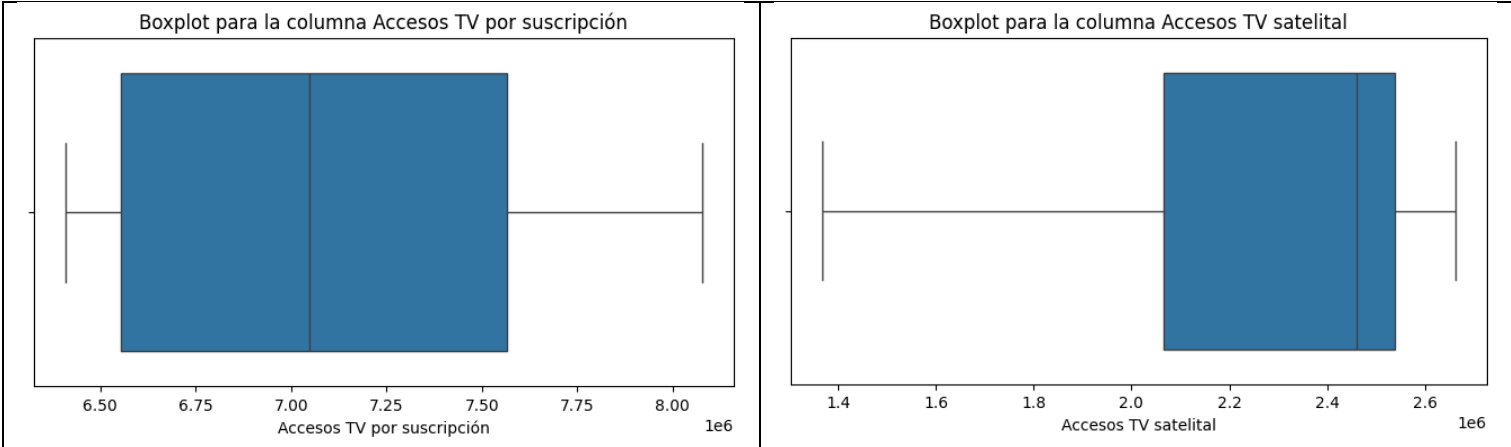
Cantidad de valores faltantes eliminados: 0

Cantidad de valores atípicos en la columna 'Año': 0

Cantidad de valores atípicos en la columna 'Trimestre': 0

Cantidad de valores atípicos en la columna 'Accesos TV por suscripción': 0

Cantidad de valores atípicos en la columna 'Accesos TV satelital': 0



De los gráficos de Box Plot ya se observa que la cantidad de accesos por suscripción es mayor que el de la TV satelital, esto se apoya también con el análisis de los estadísticos básicos

Resumen estadístico:

	Accesos TV por suscripción \	Accesos TV satelital
count	4.100000e+01	4.100000e+01
mean	7.086233e+06	2.297208e+06
std	4.832267e+05	3.572491e+05
min	6.409504e+06	1.367489e+06
25%	6.553801e+06	2.064699e+06
50%	7.048709e+06	2.460177e+06
75%	7.564827e+06	2.537610e+06
max	8.076400e+06	2.662031e+06

Estudio Del Amo. <https://estudiodelamo.com/inflacion-argentina-anual-mensual/>

El Cronista online 14 de Mayo de 2024 "Internet satelital en la Argentina: cómo son y cuánto cuestan los nuevos planes que Orbith lanzó para hogares" <https://www.cronista.com/infotechnology/internet/internet-satelital-en-la-argentina-como-son-y-cuanto-cuestan-los-nuevos-planes-que-lanzo-orbith-para-hogares/> . Visto el 19-01-2024

Clarín.com 22/12/2023 "Starlink confirmó su llegada a la Argentina: precio del servicio y cuándo estará disponible en el país" https://www.clarin.com/tecnologia/starlink-confirmando-llegada-argentina-precio-servicio-disponible-pais_0_UrYoebCfxD.html . Visto el 19-01-2024

La Nación.com 17 de mayo de 2019 "Día de Internet: en dos años la velocidad de conexión fija se triplicó en la Argentina" <https://www.lanacion.com.ar/tecnologia/en-dos-anos-velocidad-conexion-internet-fija-nid2248828/> Visto el 20-01-2025