

AFLUENCIA DIARIA DEL METRO DE LA CDMX

importamos las librerías que utilizaremos para nuestro análisis:

- pandas
- matplotlib
- os

```
In [2]: import pandas as pd
import matplotlib.pyplot as plt
import os
```

Importamos nuestros datos puros para empezar el proceso de ETL (Extract, Transform and Load)

```
In [49]: os.listdir(r"C:\Users\malet\Metro_Dataset")
```

```
Out[49]: ['afluenciastc_desglosado_07_2024.csv', 'afluenciastc_simple_02_2024.csv']
```

```
In [50]: metro_simple = pd.read_csv(r"C:\Users\malet\Metro_Dataset/afluenciastc_simple_02_2024.csv")
metro_simple.head()
```

```
Out[50]:
```

	fecha	anio	mes	linea	estacion	afluencia
0	01/01/2010	2010	Enero	Linea 1	Zaragoza	20227
1	01/01/2010	2010	Enero	Linea 1	Isabel la Católica	6487
2	01/01/2010	2010	Enero	Linea 1	Moctezuma	10304
3	01/01/2010	2010	Enero	Linea 1	Pino Suárez	8679
4	01/01/2010	2010	Enero	Linea 1	Gómez Farías	19499

```
In [51]: metro_simple = metro_simple.rename(columns={"anio": "año"})
metro_simple.head()
```

```
Out[51]:
```

	fecha	año	mes	linea	estacion	afluencia
0	01/01/2010	2010	Enero	Linea 1	Zaragoza	20227
1	01/01/2010	2010	Enero	Linea 1	Isabel la Católica	6487
2	01/01/2010	2010	Enero	Linea 1	Moctezuma	10304
3	01/01/2010	2010	Enero	Linea 1	Pino Suárez	8679
4	01/01/2010	2010	Enero	Linea 1	Gómez Farías	19499

```
In [184]: metro_simple = metro_simple.apply(lambda x: x.str.strip() if x.dtype == "object" else x)
```

```
In [53]: metro_simple.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1008735 entries, 0 to 1008734
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   fecha      1008735 non-null  object
1   año        1008735 non-null  int64
2   mes        1002690 non-null  object
3   linea      1008735 non-null  object
4   estacion   1008735 non-null  object
5   afluencia  1008735 non-null  int64
dtypes: int64(2), object(4)
memory usage: 46.2+ MB
```

```
In [54]: metro_simple.duplicated().sum()
```

```
Out[54]: 0
```

```
In [55]: metro_simple = metro_simple.dropna(how="all")
```

```
In [56]: metro_simple["fecha"] = pd.to_datetime(metro_simple["fecha"], format="%d/%m/%Y")
```

```
In [57]: metro_simple.dtypes
```

```
Out[57]: fecha          datetime64[ns]
año              int64
mes              object
linea            object
estacion         object
afluencia        int64
dtype: object
```

```
In [58]: metro_simple.nunique()
```

```
Out[58]: fecha          5173
año              15
mes              12
linea            24
estacion         163
afluencia        90354
dtype: int64
```

```
In [59]: metro_simple["linea"].unique()
```

```
Out[59]: array(['Linea 1', 'Linea 6', 'Linea 9', 'Linea 8', 'Linea 5', 'Linea 7',
                'Linea 3', 'Linea 4', 'Linea 2', 'Linea B', 'Linea 12', 'Linea A',
                'Línea 1', 'Línea 12', 'Línea 2', 'Línea 3', 'Línea 4', 'Línea 5',
                'Línea 6', 'Línea 7', 'Línea 8', 'Línea 9', 'Línea A', 'Línea B'],
              dtype=object)
```

```
In [60]: metro_simple.replace(
        {
            "Línea 1": "Linea 1",
            'Línea 2': "Linea 2",
            'Línea 3': "Linea 3",
            'Línea 4': "Linea 4",
            'Línea 5': "Linea 5",
            'Línea 6': "Linea 6",
            'Línea 7': "Linea 7",
            'Línea 8': "Linea 8",
            'Línea 9': "Linea 9",
            'Línea 12': "Linea 12",
            'Línea A': "Linea A",
            'Línea B': "Linea B"
        },
        inplace=True)
```

```
In [61]: metro_simple["linea"].unique()
```

```
Out[61]: array(['Linea 1', 'Linea 6', 'Linea 9', 'Linea 8', 'Linea 5', 'Linea 7',
                'Linea 3', 'Linea 4', 'Linea 2', 'Linea B', 'Linea 12', 'Linea A'],
              dtype=object)
```

```
In [77]: metro_simple.head()
```

Out[77]:

	fecha	año	mes	linea	estacion	afluencia	dia_semana
0	2010-01-01	2010	January	Linea 1	Zaragoza	20227	Friday
1	2010-01-01	2010	January	Linea 1	Isabel la Católica	6487	Friday
2	2010-01-01	2010	January	Linea 1	Moctezuma	10304	Friday
3	2010-01-01	2010	January	Linea 1	Pino Suárez	8679	Friday
4	2010-01-01	2010	January	Linea 1	Gómez Farías	19499	Friday

```
In [82]: metro_simple["dia_semana"] = metro_simple["fecha"].dt.day_name()
```

```
In [83]: metro_simple["mes"] = metro_simple["fecha"].dt.month_name()
```

```
In [84]: metro_simple.head()
```

Out[84]:

	fecha	año	mes	linea	estacion	afluencia	dia_semana
0	2010-01-01	2010	January	Linea 1	Zaragoza	20227	Friday
1	2010-01-01	2010	January	Linea 1	Isabel la Católica	6487	Friday
2	2010-01-01	2010	January	Linea 1	Moctezuma	10304	Friday
3	2010-01-01	2010	January	Linea 1	Pino Suárez	8679	Friday
4	2010-01-01	2010	January	Linea 1	Gómez Farías	19499	Friday

```
In [85]: metro_simple.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1008735 entries, 0 to 1008734
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype  
---  -
0   fecha       1008735 non-null  datetime64[ns]
1   año         1008735 non-null  int64   
2   mes         1008735 non-null  object  
3   linea       1008735 non-null  object  
4   estacion    1008735 non-null  object  
5   afluencia   1008735 non-null  int64   
6   dia_semana  1008735 non-null  object  
dtypes: datetime64[ns](1), int64(2), object(4)
memory usage: 53.9+ MB
```

```
In [86]: metro_simple.to_csv(r"C:\Users\malet\Metro_Dataset\afluenciastc_simple_02_2024_clean.csv")
```

```
In [209]: metro_desglosado = pd.read_csv(r"C:\Users\malet\Metro_Dataset\afluenciastc_desglosado_07_2024.csv")
metro_desglosado.head()
```

```
Out[209]:
```

	fecha	mes	año	linea	estacion	tipo_pago	afluencia
0	01/01/2021	Enero	2021	Línea 1	Balbuena	Boleto	1848
1	01/01/2021	Enero	2021	Línea 1	Balbuena	Prepago	790
2	01/01/2021	Enero	2021	Línea 1	Balbuena	Gratuidad	325
3	01/01/2021	Enero	2021	Línea 1	Balderas	Boleto	1103
4	01/01/2021	Enero	2021	Línea 1	Balderas	Prepago	922

```
In [210]: metro_desglosado.nunique()
```

```
Out[210]:
```

fecha	1308
mes	12
año	4
linea	24
estacion	163
tipo_pago	3
afluencia	37477

dtype: int64

```
In [211]: metro_desglosado.groupby("tipo_pago")["afluencia"].sum()
```

```
Out[211]:
```

tipo_pago	
Boleto	770641003
Gratuidad	407800479
Prepago	2420348652

Name: afluencia, dtype: int64

Ahora sigue el proceso del análisis

Intentaremos responder a las siguientes preguntas:

- ¿Cuáles son los meses con mayor y menor afluencia?
- ¿Cómo ha cambiado la afluencia diaria a lo largo del tiempo?
- ¿Cuáles son los días de la semana con mayor afluencia en el metro?
- ¿Qué líneas del Metro tienen la mayor y menor afluencia?
- ¿Cuáles son las estaciones más concurridas?
- ¿Qué tipo de pago(Boleto, Prepago o Gratuidad) fue el más utilizado en 2023?

```
In [87]: metro_simple.head()
```

```
Out[87]:
```

	fecha	año	mes	linea	estacion	afluencia	dia_semana
0	2010-01-01	2010	January	Línea 1	Zaragoza	20227	Friday
1	2010-01-01	2010	January	Línea 1	Isabel la Católica	6487	Friday
2	2010-01-01	2010	January	Línea 1	Moctezuma	10304	Friday
3	2010-01-01	2010	January	Línea 1	Pino Suárez	8679	Friday
4	2010-01-01	2010	January	Línea 1	Gómez Farías	19499	Friday

```
In [127]: metro_simple
```

Out[127..

	fecha	año	mes	linea	estacion	afluencia	dia_semana
0	2010-01-01	2010	January	Linea 1	Zaragoza	20227	Friday
1	2010-01-01	2010	January	Linea 1	Isabel la Católica	6487	Friday
2	2010-01-01	2010	January	Linea 1	Moctezuma	10304	Friday
3	2010-01-01	2010	January	Linea 1	Pino Suárez	8679	Friday
4	2010-01-01	2010	January	Linea 1	Gómez Farías	19499	Friday
...
1008730	2024-02-29	2024	February	Linea B	Romero Rubio	16403	Thursday
1008731	2024-02-29	2024	February	Linea B	Río de los Remedios	19512	Thursday
1008732	2024-02-29	2024	February	Linea B	San Lázaro	17981	Thursday
1008733	2024-02-29	2024	February	Linea B	Tepito	20752	Thursday
1008734	2024-02-29	2024	February	Linea B	Villa de Aragón	18203	Thursday

1008735 rows × 7 columns

¿Cuáles son los meses con mayor y menor afluencia?

In [190..

```
afluencia_mes = metro_simple.groupby("mes")["afluencia"].sum().sort_values(ascending=False)
afluencia_mes
```

Out[190..

```
mes
October      1798689961
August       1748054105
January      1731777711
November     1721338276
March        1713793233
February     1704009465
December     1679193192
May          1669266326
September    1628585228
July         1618303742
June         1611161876
April        1567995549
Name: afluencia, dtype: int64
```

In [197..

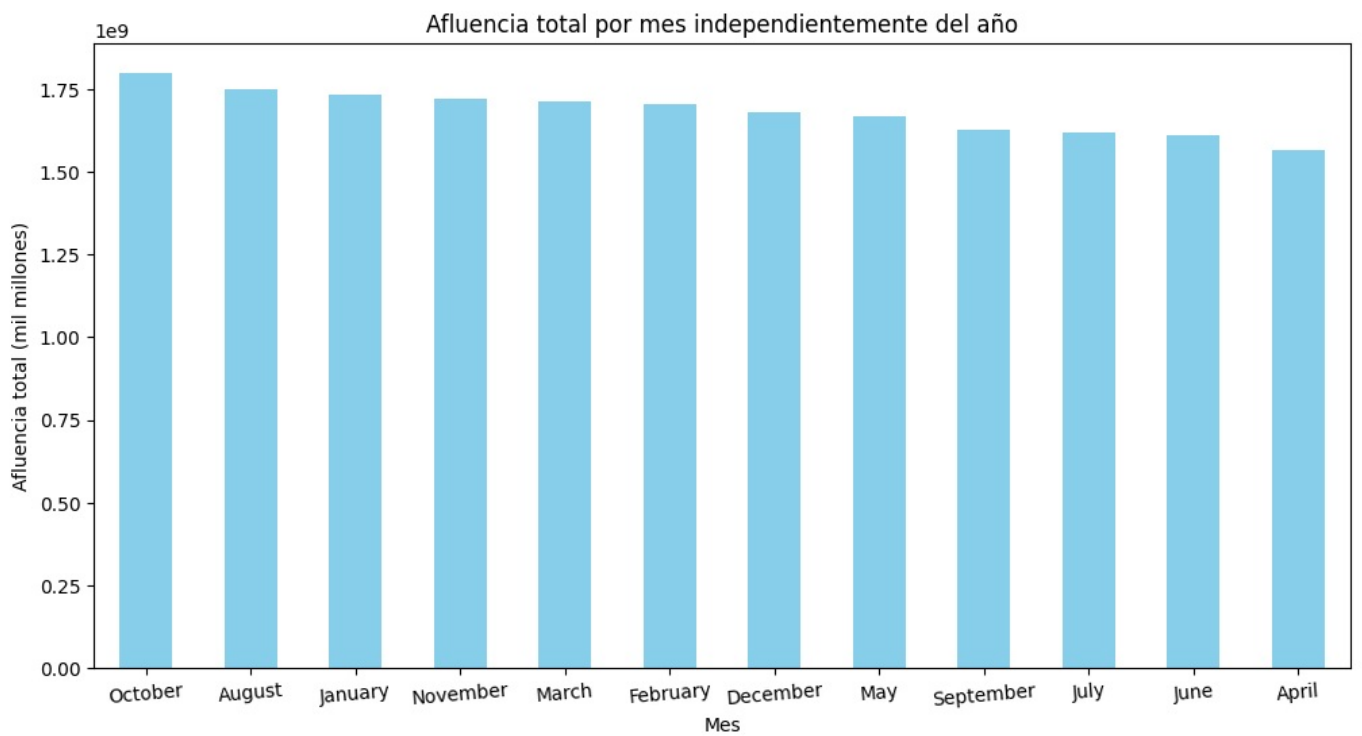
```
mes_max_afluencia = afluencia_mes.idxmax()
max_afluencia = afluencia_mes.max()
```

In [198..

```
mes_min_afluencia = afluencia_mes.idxmin()
min_afluencia = afluencia_mes.min()
```

In [199..

```
plt.figure(figsize=(12,6))
afluencia_mes.plot(kind="bar", color="skyblue")
plt.title("Afluencia total por mes independientemente del año")
plt.xlabel("Mes")
plt.ylabel("Afluencia total (mil millones)")
plt.xticks(rotation=5);
```



```
In [202...] print(f"{mes_max_afluencia} ha sido el mes con la máxima afluencia acumulando desde 2010 un total de {max_afluencia}")
```

October ha sido el mes con la máxima afluencia acumulando desde 2010 un total de 1798689961 de personas. mientras que April fue el mes que presentó la menor afluencia, con aproximadamente 1567995549 de personas. Debemos considerar que este mes incluye Semana Santa, un periodo en el que muchas personas toman vacaciones y la movilidad dentro de la ciudad disminuye

¿Cómo ha cambiado la afluencia diaria a lo largo del tiempo?

```
In [150...] metro_simple["año"] = metro_simple["fecha"].dt.year
```

```
In [151...] filtrado_año = metro_simple[metro_simple["año"] < 2024]
filtrado_año
```

```
Out[151...]

```

	fecha	año	mes	linea	estacion	afluencia	día_semana
0	2010-01-01	2010	January	Linea 1	Zaragoza	20227	Friday
1	2010-01-01	2010	January	Linea 1	Isabel la Católica	6487	Friday
2	2010-01-01	2010	January	Linea 1	Moctezuma	10304	Friday
3	2010-01-01	2010	January	Linea 1	Pino Suárez	8679	Friday
4	2010-01-01	2010	January	Linea 1	Gómez Farías	19499	Friday
...
997030	2023-12-31	2023	December	Linea B	Romero Rubio	10118	Sunday
997031	2023-12-31	2023	December	Linea B	Río de los Remedios	8540	Sunday
997032	2023-12-31	2023	December	Linea B	San Lázaro	10278	Sunday
997033	2023-12-31	2023	December	Linea B	Tepito	26312	Sunday
997034	2023-12-31	2023	December	Linea B	Villa de Aragón	4228	Sunday

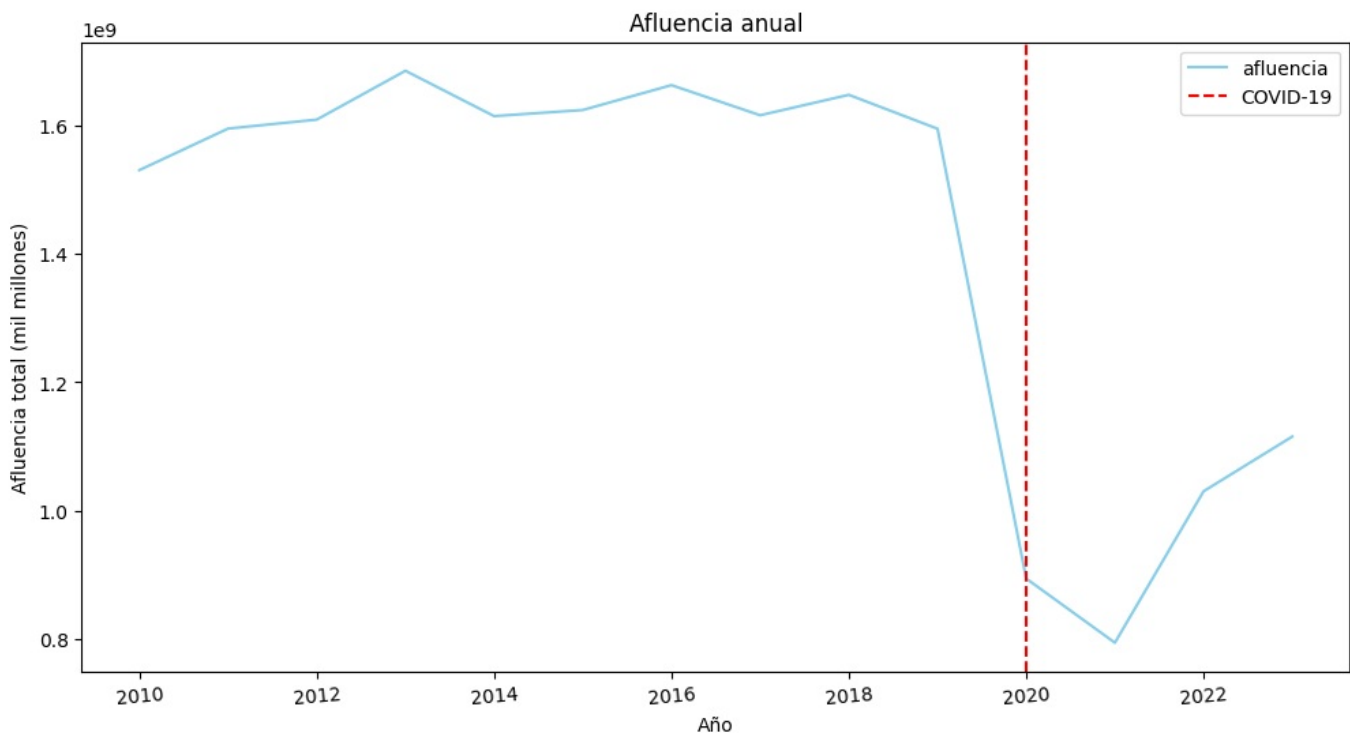
997035 rows × 7 columns

```
In [152...] afluencia_año = filtrado_año.groupby("año")["afluencia"].sum()
afluencia_año
```

```
Out[152... año
2010    1530248597
2011    1594903897
2012    1608865175
2013    1684936618
2014    1614333594
2015    1623828642
2016    1662562714
2017    1615652411
2018    1647475013
2019    1594651279
2020     894224342
2021     794299549
2022    1030108295
2023    1115326009
Name: afluencia, dtype: int64
```

```
In [203... plt.figure(figsize=(12,6))
afluencia_año.plot(kind="line", color="skyblue")
plt.title("Afluencia anual")
plt.xlabel("Año")
plt.ylabel("Afluencia total (mil millones)")
plt.axvline(x=2020, color='red', linestyle='--', label='COVID-19')
plt.legend()

plt.xticks(rotation=5);
```



De 2014 a 2019, la afluencia se mantuvo relativamente estable con ligeras fluctuaciones en donde los años con mayor afluencia fueron 2013 y 2016. En 2020, se observa una drástica caída en la afluencia. Este descenso coincide con la pandemia de COVID-19, cuando muchas actividades se suspendieron y la movilidad urbana disminuyó drásticamente. En 2020, vemos una recuperación de afluencia en el metro en un 30% con respecto a 2021.

¿Cuáles son los días de la semana con mayor afluencia en el metro?

```
In [160... metro_simple.head()
```

```
Out[160...
```

	fecha	año	mes	linea	estacion	afluencia	dia_semana
0	2010-01-01	2010	January	Linea 1	Zaragoza	20227	Friday
1	2010-01-01	2010	January	Linea 1	Isabel la Católica	6487	Friday
2	2010-01-01	2010	January	Linea 1	Moctezuma	10304	Friday
3	2010-01-01	2010	January	Linea 1	Pino Suárez	8679	Friday
4	2010-01-01	2010	January	Linea 1	Gómez Farías	19499	Friday

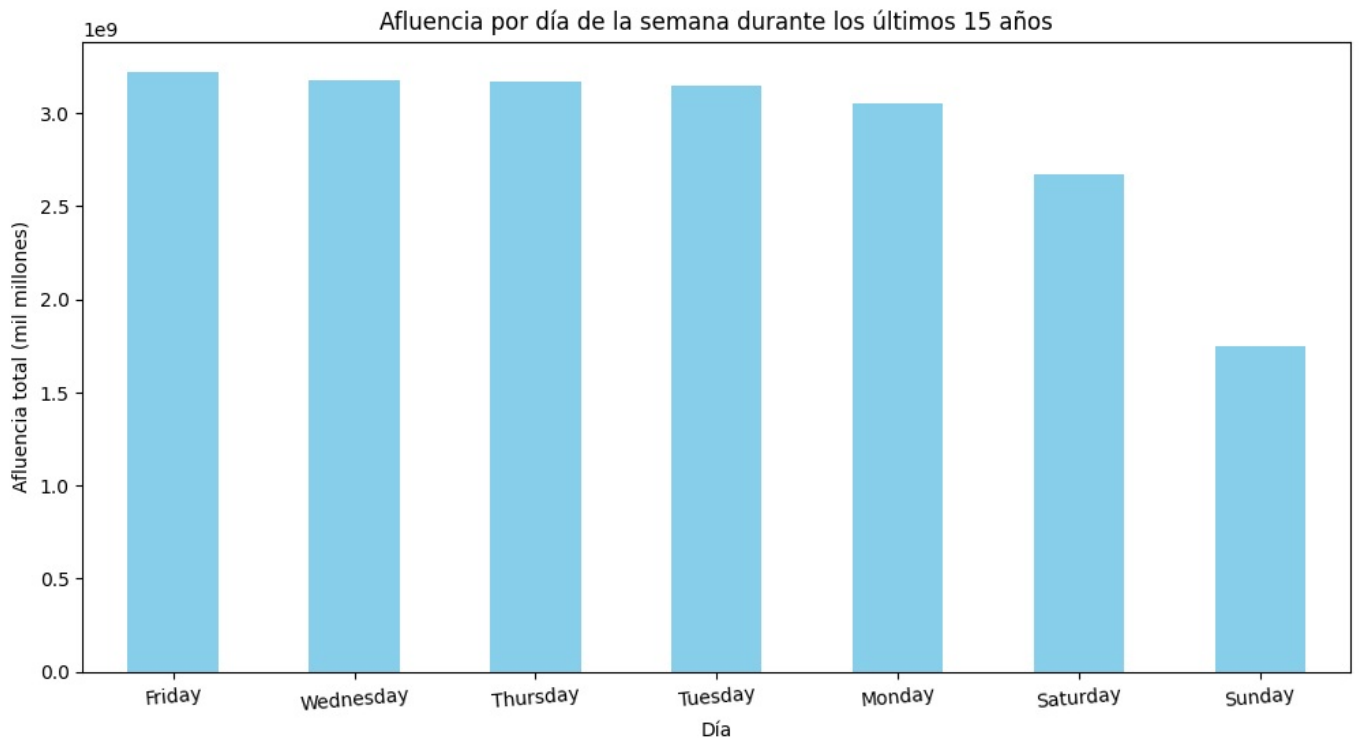
```
In [161... afluencia_dia = metro_simple.groupby("dia_semana")["afluencia"].sum().sort_values(ascending=False)
```

```
afluencia_dia
```

```
Out[161]: dia_semana
Friday      3222826103
Wednesday   3179390505
Thursday     3167074235
Tuesday      3146517261
Monday       3051244722
Saturday     2674202779
Sunday       1750913059
Name: afluencia, dtype: int64
```

```
In [204]: plt.figure(figsize=(12,6))
afluencia_dia.plot(kind="bar", color="skyblue")
plt.title("Afluencia por día de la semana durante los últimos 15 años")
plt.xlabel("Día")
plt.ylabel("Afluencia total (mil millones)")

plt.xticks(rotation=5);
```



En el gráfico anterior se puede apreciar que el día con mayor afluencia es Viernes. Esto puede deberse a que, al ser el último día de la semana, muchas personas realizan más actividades, como compras o actividades sociales, lo que incrementa el uso del metro en ese día. El domingo es el día con la menor afluencia. Esto es normal debido a que, como es un día de descanso para la mayoría, la movilidad es mucho menor.

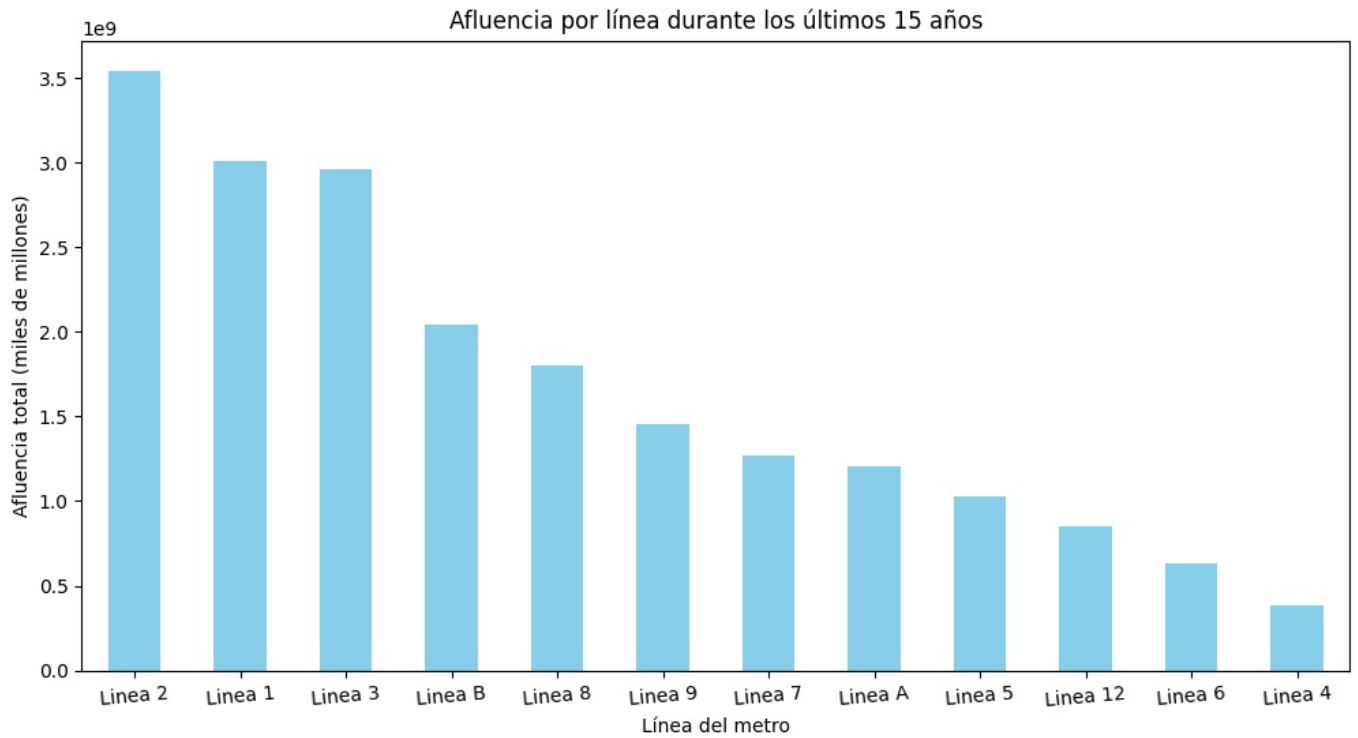
¿Qué líneas del Metro tienen la mayor y menor afluencia?

```
In [164]: afluencia_linea = metro_simple.groupby("linea")["afluencia"].sum().sort_values(ascending=False)
afluencia_linea
```

```
Out[164]: linea
Linea 2      3544946782
Linea 1      3013189412
Linea 3      2959979511
Linea B      2041443602
Linea 8      1800199882
Linea 9      1454223126
Linea 7      1268056677
Linea A      1204193137
Linea 5      1030407553
Linea 12     854514583
Linea 6      635408597
Linea 4      385605802
Name: afluencia, dtype: int64
```

```
In [171]: plt.figure(figsize=(12,6))
afluencia_linea.plot(kind="bar", color="skyblue")
plt.title("Afluencia por línea durante los últimos 15 años")
plt.xlabel("Línea del metro")
plt.ylabel("Afluencia total (miles de millones)")
```

```
plt.xticks(rotation=5);
```



La Línea 2 es la más concurrida del metro de la CDMX debido a que, conecta áreas desamente pobladas con zonas comerciales y administrativas clave. También su alta afluencia se debe a que conecta múltiples líneas del metro y otros sistemas de transporte público.

La línea 4 es la que tiene la menor afluencia de todo el sistema del metro. Esto puede deberse a que es la línea más corta y atraviesa áreas con menor densidad poblacional.

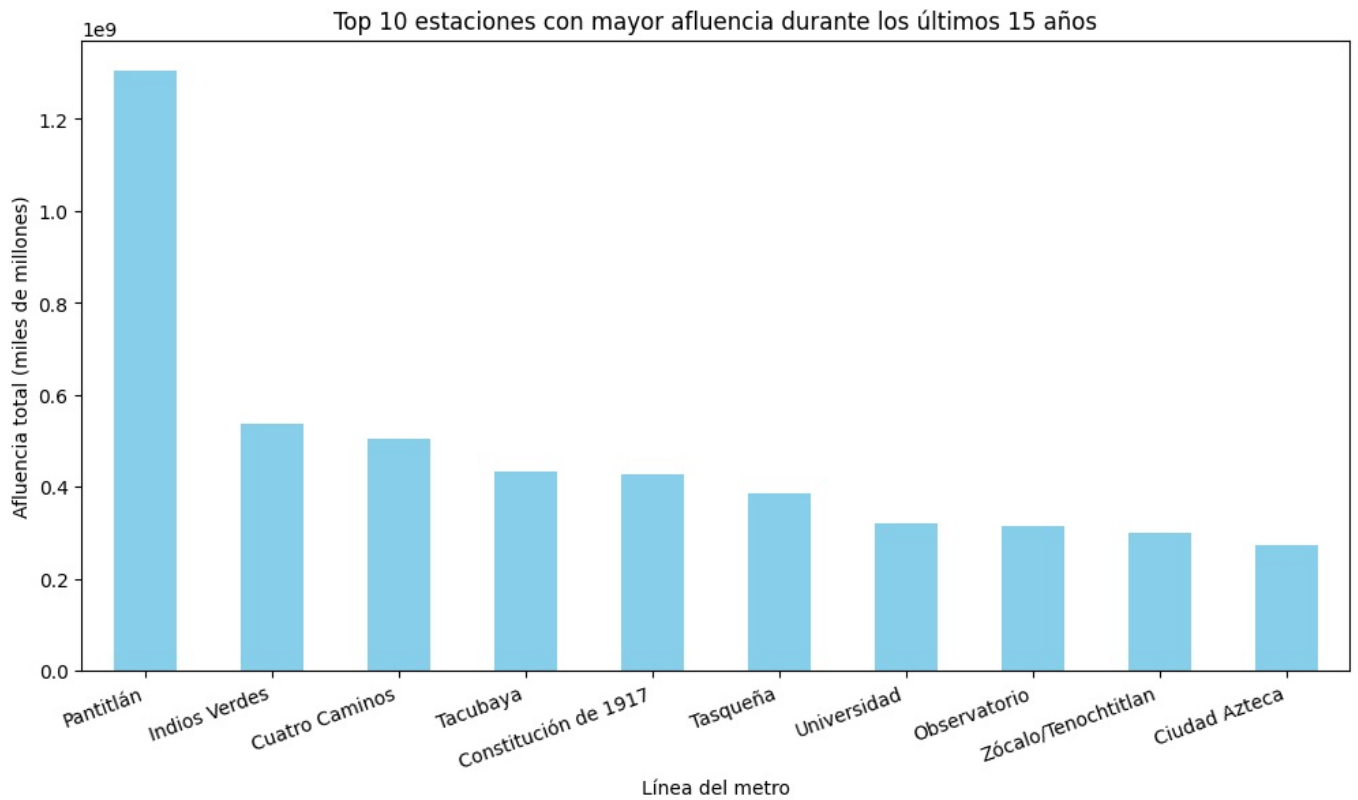
¿Cuáles son las estaciones más concurridas?

```
In [174.. top_estaciones = metro_simple.groupby('estacion')['afluencia'].sum().nlargest(10)
top_estaciones
```

```
Out[174.. estacion
Pantitlán          1304802317
Indios Verdes      535827451
Cuatro Caminos     503152429
Tacubaya           434248526
Constitución de 1917 427527596
Tasqueña           385170618
Universidad        320292067
Observatorio       312863116
Zócalo/Tenochtitlan 298348138
Ciudad Azteca      273243864
Name: afluencia, dtype: int64
```

```
In [183.. plt.figure(figsize=(12,6))
top_estaciones.plot(kind="bar", color="skyblue")
plt.title("Top 10 estaciones con mayor afluencia durante los últimos 15 años")
plt.xlabel("Línea del metro")
plt.ylabel("Afluencia total (miles de millones)")

plt.xticks(rotation=20, ha="right");
```

Pantitlán es la estación más transitada del sistema debido a que es un importante nodo de intercambio, sirviendo a cuatro líneas del metro (1, 5, 9, A) y conectando con otros sistemas de transporte público, como el Metrobús y autobuses.

La estación Cuatro caminos como terminal de la Línea 2, atrae a numerosos usuarios provenientes del Estado de México. Además, su papel como punto de acceso a la Ciudad de México desde municipios industriales cercanos contribuye a su alta afluencia.

¿Qué tipo de pago(Boleto, Prepago o Gratuidad) fue el más utilizado en 2023?

```
In [213...] metro_desglosado.head()
```

```
Out[213...]
      fecha  mes  anio  linea  estacion  tipo_pago  afluencia
0  01/01/2021  Enero  2021  Línea 1  Balbuena    Boleto      1848
1  01/01/2021  Enero  2021  Línea 1  Balbuena    Prepago      790
2  01/01/2021  Enero  2021  Línea 1  Balbuena    Gratuidad     325
3  01/01/2021  Enero  2021  Línea 1  Balderas    Boleto     1103
4  01/01/2021  Enero  2021  Línea 1  Balderas    Prepago      922
```

```
In [218...] metro_2023 = metro_desglosado[metro_desglosado["anio"] == 2023]
```

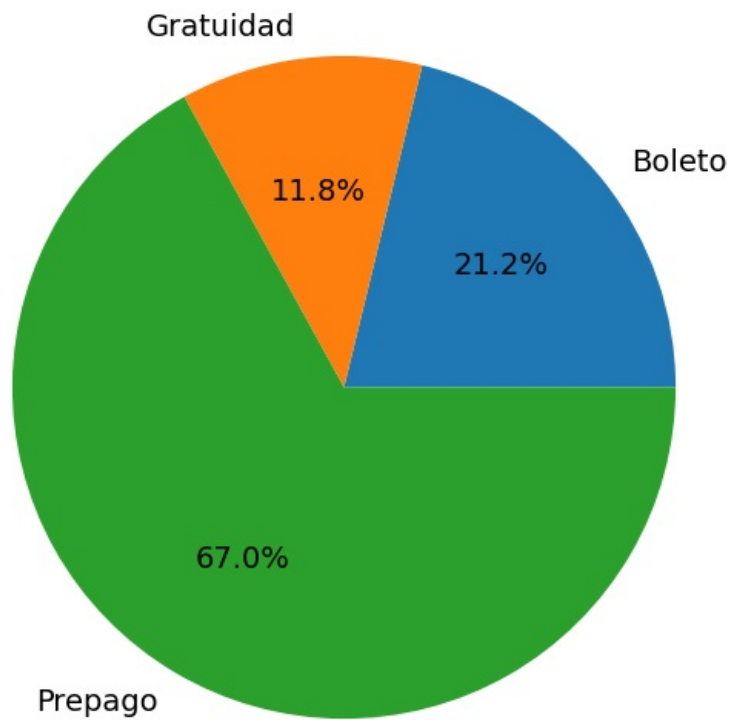
```
In [219...] tipo_pago = metro_2023.groupby("tipo_pago")["afluencia"].sum()
tipo_pago
```

```
Out[219...]
tipo_pago
Boleto      236532337
Gratuidad   131511385
Prepago     747282287
Name: afluencia, dtype: int64
```

```
In [241...] plt.figure(figsize=(19,7))
tipo_pago.plot(kind="pie",autopct='%1.1f%', color="skyblue",textprops={'fontsize': 14})
plt.title("Tipo de pago más utilizado en 2023",fontsize=20)
plt.ylabel("")

plt.xticks(rotation=5);
```

Tipo de pago más utilizado en 2023



La tarjeta de prepago fue el método de pago más utilizado en 2023. Esto puede deberse a su conveniencia, rapidez y al hecho de que es un método recargable que permite a los usuarios acceder a múltiples medio de transporte con un solo pago.

El de gratuidad fue el tipo de "pago" menos utilizado debido a que solo incluye el acceso gratuito a ciertos grupos de personas, como mayores de edad, personas con discapacidad, niños menores de 5 años y uniformados.

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js