

```
# Proyecto de Análisis de Datos con Python - Dataset "Flights"

## 1. Título del Proyecto
# Análisis de patrones de pasajeros en vuelos comerciales (1949-1960)

## 2. Objetivo del análisis
# Identificar tendencias anuales y estacionales en la cantidad de pasajeros utilizando el dataset 'flights' de Seaborn.

## 3. Carga de datos
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = sns.load_dataset('flights')

## 4. Exploración inicial
print(df.head())
print(df.info())
print(df.describe())

## 5. Análisis por año y por mes
# Total de pasajeros por año
pasajeros_por_anio = df.groupby('year')['passengers'].sum()
print("Pasajeros por año:")
print(pasajeros_por_anio)

# Total de pasajeros por mes (promedio de todos los años)
df_months = df.groupby('month')['passengers'].sum().sort_values(ascending=False)
print("Pasajeros por mes:")
print(df_months)

# Análisis específico del año 1958
df_1958 = df[df['year'] == 1958]
df_1958_ordenado = df_1958.groupby('month')['passengers'].sum().sort_values(ascending=False)
print("Pasajeros por mes en 1958:")
print(df_1958_ordenado)

## 6. Visualizaciones

# Heatmap de pasajeros por mes y año
tabla = df.pivot(index='month', columns='year', values='passengers')
plt.figure(figsize=(12, 6))
sns.heatmap(tabla, annot=True, fmt=".0f", cmap="YlGnBu")
plt.title("Cantidad de pasajeros por mes y año")
plt.xlabel("Año")
plt.ylabel("Mes")
plt.show()

# Gráfico de barras: pasajeros por año
plt.figure(figsize=(7, 5))
sns.barplot(x="year", y="passengers", data=df, ci="sd")
plt.title("Pasajeros por año")
plt.xlabel("Año")
plt.ylabel("Pasajeros")
plt.show()

# Gráfico de barras: pasajeros por mes
plt.figure(figsize=(7, 5))
sns.barplot(x="month", y="passengers", data=df, ci="sd")
plt.title("Pasajeros por mes")
plt.xlabel("Mes")
plt.ylabel("Pasajeros")
plt.show()

## 7. Conclusiones

# - La cantidad de pasajeros aumentó progresivamente cada año entre 1949 y 1960.
# - Julio y agosto son consistentemente los meses con mayor número de pasajeros.
# - Esta concentración estacional se acentúa en los últimos años de la serie.

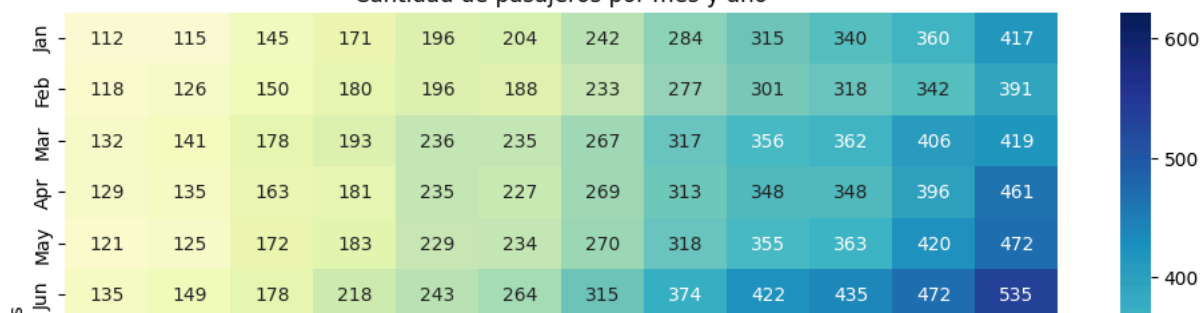
## 8. Interpretación final

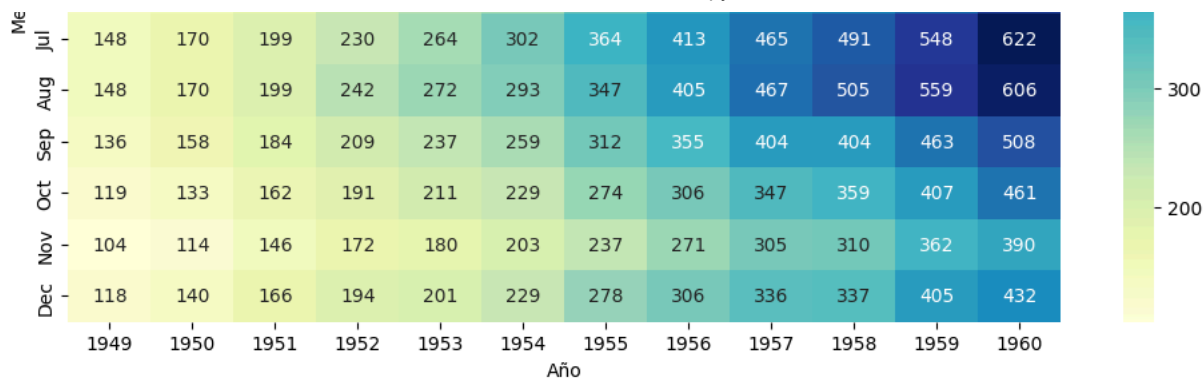
# A medida que avanzaron los años se incrementó significativamente la cantidad de pasajeros, posiblemente debido a mejoras tecnológicas,
# La estacionalidad marcada en julio y agosto sugiere una fuerte influencia del turismo en verano (hemisferio norte) sobre el uso de vuel
```

```
↵ year month passengers
0 1949 Jan 112
1 1949 Feb 118
2 1949 Mar 132
3 1949 Apr 129
4 1949 May 121
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144 entries, 0 to 143
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   year        144 non-null   int64
1   month       144 non-null   category
2   passengers  144 non-null   int64
dtypes: category(1), int64(2)
memory usage: 2.9 KB
None

      year  passengers
count  144.000000  144.000000
mean   1954.500000  280.298611
std      3.464102  119.966317
min    1949.000000  104.000000
25%    1951.750000  180.000000
50%    1954.500000  265.500000
75%    1957.250000  360.500000
max    1960.000000  622.000000
Pasajeros por año:
year
1949    1520
1950    1676
1951    2042
1952    2364
1953    2700
1954    2867
1955    3408
1956    3939
1957    4421
1958    4572
1959    5140
1960    5714
Name: passengers, dtype: int64
Pasajeros por mes:
month
Jul      4216
Aug      4213
Jun      3740
Sep      3629
May      3262
Mar      3242
Apr      3205
Oct      3199
Dec      3142
Jan      2901
Feb      2820
Nov      2794
Name: passengers, dtype: int64
Pasajeros por mes en 1958:
month
Aug      505
Jul      491
Jun      435
Sep      404
May      363
Mar      362
Oct      359
Apr      348
Jan      340
Dec      337
Feb      318
Nov      310
Name: passengers, dtype: int64
<ipython-input-1-fe8216b84db2>:28: FutureWarning: The default of observed=False is deprecated and will be changed to True in a futur
df_months = df.groupby('month')['passengers'].sum().sort_values(ascending=False)
<ipython-input-1-fe8216b84db2>:34: FutureWarning: The default of observed=False is deprecated and will be changed to True in a futur
df_1958_ordenado = df_1958.groupby('month')['passengers'].sum().sort_values(ascending=False)
```

Cantidad de pasajeros por mes y año

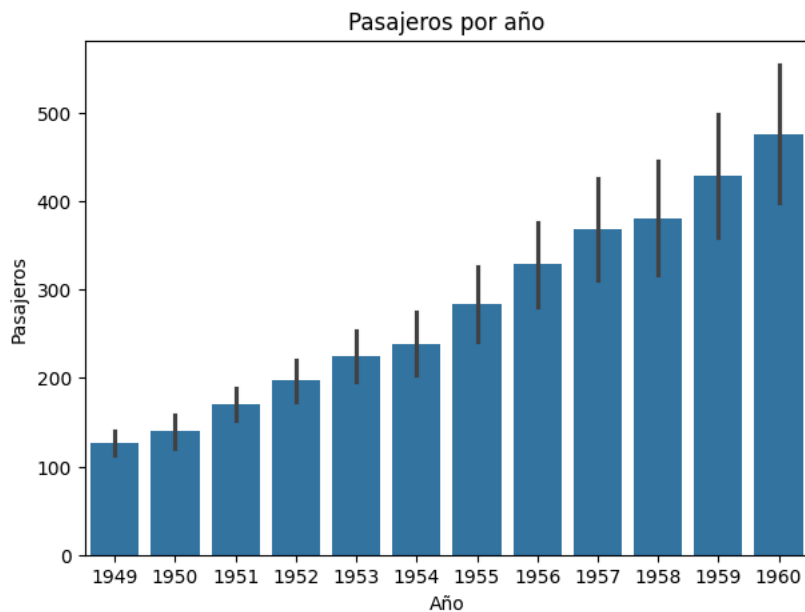




```
<ipython-input-1-fe8216b84db2>:51: FutureWarning:
```

The `ci` parameter is deprecated. Use `errorbar='sd'` for the same effect.

```
sns.barplot(x="year", y="passengers", data=df, ci="sd")
```



```
<ipython-input-1-fe8216b84db2>:59: FutureWarning:
```

The `ci` parameter is deprecated. Use `errorbar='sd'` for the same effect.

```
sns.barplot(x="month", y="passengers", data=df, ci="sd")
```

