Colab: https://colab.research.google.com/drive/1_8kAoExECvH09c_A5ApycfS9CJ3siLO2?usp=sharing

```
import numpy as np
import pandas as pd


# Matplotlib
# Seaborn


# 1. Exploratory - EDA, looking for patterns, analysing the data
# 2. Explanatory - Storytelling, Dashboarding


# Science - anatomy of plot, choosing the right plot
# Art - right scale, labels, axis ticks, remove clutter, highlight some information


import matplotlib.pyplot as plt
import seaborn as sns
# matplotlib+pandas
# why not plotly - is creates dynamic plots
# more code to write
# more difficult to grasp for a beginner
# not used a lot in Industry
# M+S, Tableau


!wget https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/021/299/origin
```

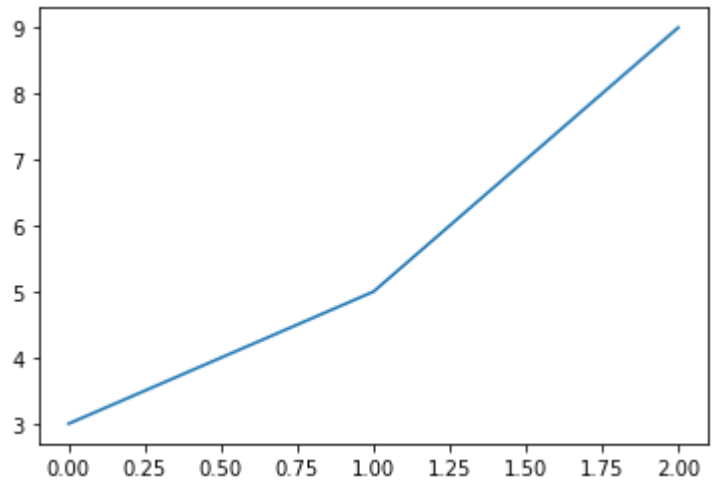
```
--2022-12-12 16:21:42--  https://d2beiqkhq929f0.cloudfront.net/public_assets/a
Resolving d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)... 108
Connecting to d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)|10
HTTP request sent, awaiting response... 200 OK
Length: 2041483 (1.9M) [text/plain]
Saving to: 'final_vg.csv'

final_vg.csv        100%[===================>]   1.95M  7.49MB/s    in 0.3s

2022-12-12 16:21:42 (7.49 MB/s) - 'final_vg.csv' saved [2041483/2041483]
```

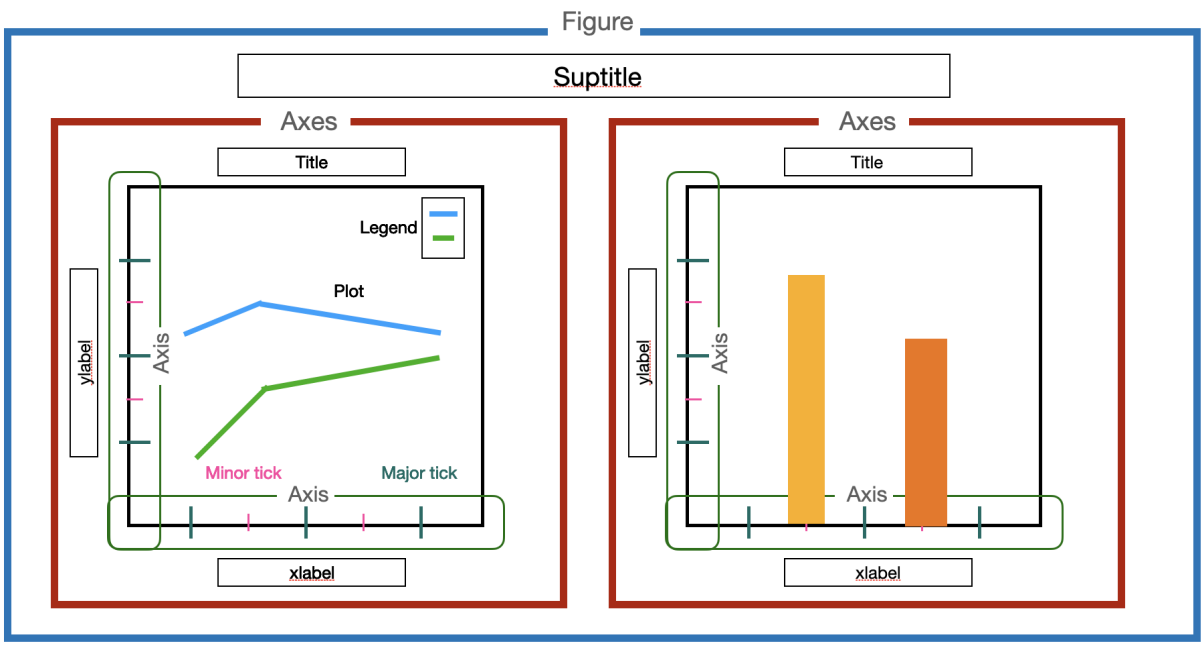
```
data = pd.read_csv('final_vg.csv')
data.head()
```

Rank	Name	Platform	Year	Genre	Publishe

```
# (0, 3), (1, 5), (2, 9)
x_val = [0, 1, 2]
y_val = [3, 5, 9]
plt.plot(x_val, y_val)
```

```
[<matplotlib.lines.Line2D at 0x7f96778f3a30>]
```






```
# Jupyter Cell - shown after the code
# Terminal - figure will be displayed as a sep window
# IDE - Seperate very small window will pop in the IDE itself


# Choosing is the right plot?
# Number of variables involved in answering a question
# Q1- How many variables of interest are involved?
# Q2 - Whether these variables are numerical or categorical?
```

```
# How many variables of interest are involved?
# 1 Variable - Univariate Data Visualisation
# 2 Variable - Bi-variate Data Visualisation
# 2+ Variables - Multivariate Data Visualisation


# Univariate
  # Numercical
  # Categorical
# Bivariate
  # Num, Num
  # Num, Cat
  # Cat, Cat
# Multivariate - n-dimensional, 3D
 # Num, Num, Cat
 # Cat, Cat, Num
 # Cat, Cat, Cat
 # Num, Num, Num
# Subplots


# Categorical - count of each categories, share/fraction component of each category


# How can you find the top-N Genres?


data["Genre"].value_counts()
# whenever you see a cat variable, start thinking about placing some bars
```

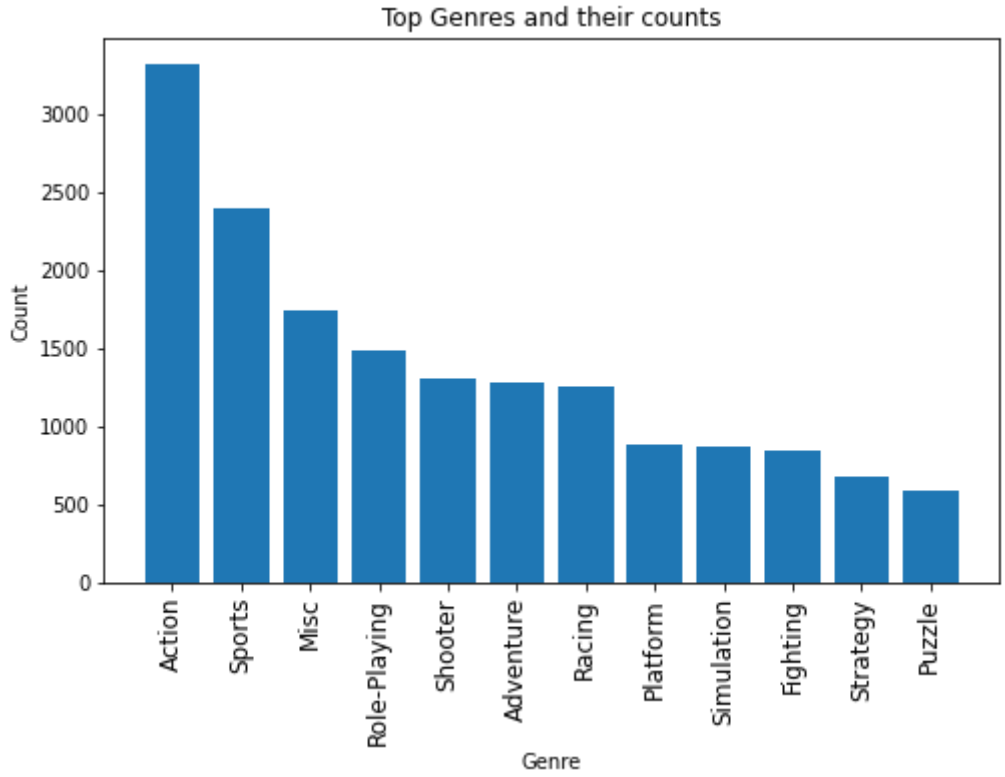
```
Action          3316
Sports          2400
Misc            1739
Role-Playing    1488
Shooter         1310
Adventure       1286
Racing          1249
Platform         886
Simulation       867
Fighting         848
Strategy         681
Puzzle           582
Name: Genre, dtype: int64
```

```
x_val = data["Genre"].value_counts().index
y_val = data["Genre"].value_counts().to_list()
plt.bar(x_val, y_val)
```

```
<BarContainer object of 12 artists>
```


```
plt.figure(figsize=(8,5))
plt.bar(x_val, y_val) # sns.____plot()
plt.xticks(rotation=90, fontsize=12)
plt.xlabel("Genre")
plt.ylabel("Count")
plt.title("Top Genres and their counts",  fontsize=12)
```

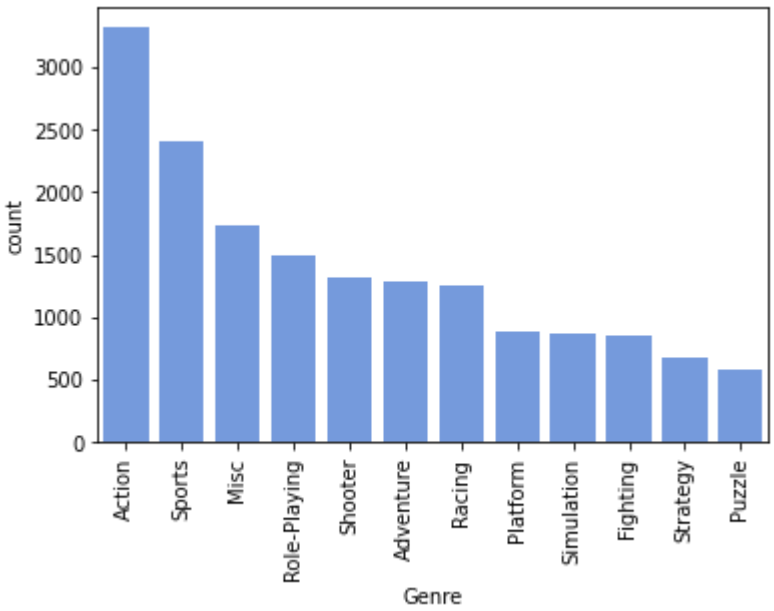
```
Text(0.5, 1.0, 'Top Genres and their counts')
```



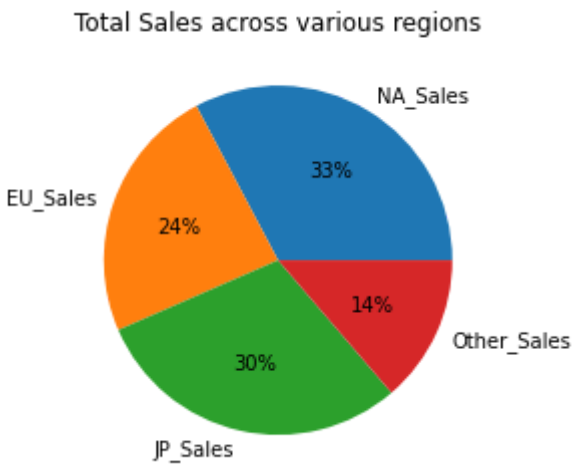

```
plt.figure(figsize=(8,5))
plt.bar(x_val, y_val, width=0.2, color="orange") # sns.____plot()
plt.xticks(rotation=90, fontsize=12)
```

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11],
 <a list of 12 Text major ticklabel objects>)
```


```
sns.countplot(data=data,·
··············x="Genre",
··············order=data["Genre"].value_counts().index,
··············color="cornflowerblue")
plt.xticks(rotation=90)
plt.show() # telling Python, that, hey now you should display all of the stuff
```




```
# Pie chart - it is not very well received by scientific, seaborn doesn't piechart


# piecharts in matplotlib, verbose - post-read
```



Colab paid products - Cancel contracts here

✓ 0s completed at 23:01