working with JSON

API → JSON → JS / python

code:

```
import pandas as pd
pd.read_json('__.json')
pd.read_json('url.com/json')
```

extract data from an API

working with SQL

* !pip install mysql.connector
* import mysql.connector
* con=mysql.connector.connect(host='localhost', user='root', password='', database='world')
* pd.read_sql_query('SELECT * FROM city', conn)

city table

API Application program Interface

Rapid API

code:

```
import pandas as pd
import requests
response=requests.get('__url__')
pd.DataFrame(response.json()['results'])
```

- when the site doesn't provide api's go for web scraping.
  * (use beautiful soup for web scraping)

Understanding your data

01. How big the data is?
   - df.shape
02. How does data look like?
   - df.head() | df.sample(5)
03. what is the data type of cols?
   - df.info()

04. Are there any missing values?

- df.isnull().sum()

05- How does data look mathematically?

- df.describe()

06. Are there any duplicate values?

- df.duplicated().sum()

07. How is the correlation b/w cols?

- df.corr()

EDA (Exploratory Data Analysis) (Titanic dataset)

univariate analysis: Analysis of single column

01. categorical data (plotting data)

sns.countplot(df['survived'])
df['survived'].value_counts()
] count how many have survived.
df['survived'].value_counts().plot(kind='pie')

02. Numerical data

Histogram

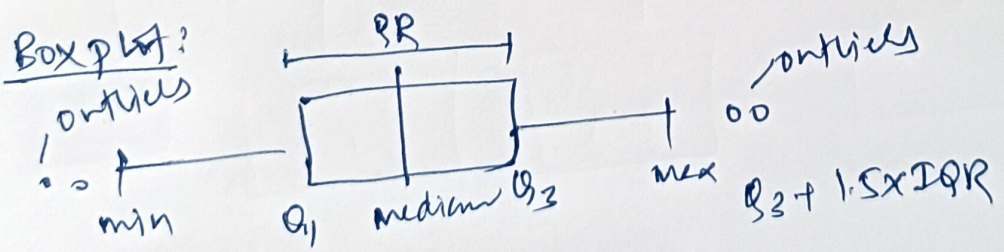
plt.hist(df['Age'])

- the distribution of data in that numerical column.

Distplot: improvement of histogram.

sns.distplot(df['Age'])

Boxplot:



$$IQR = Q_3 - Q_1$$

sns.boxplot(df['Fare'])