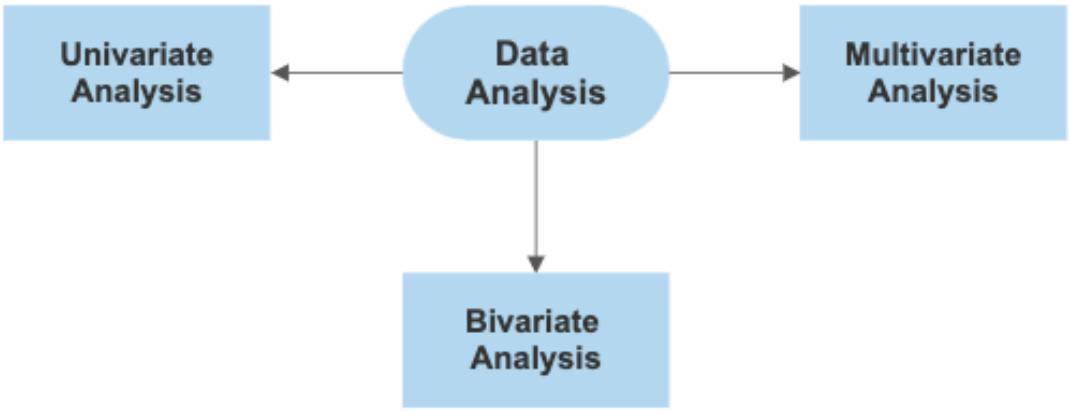
Data Science | 30 Days of Machine Learning | Day - 7

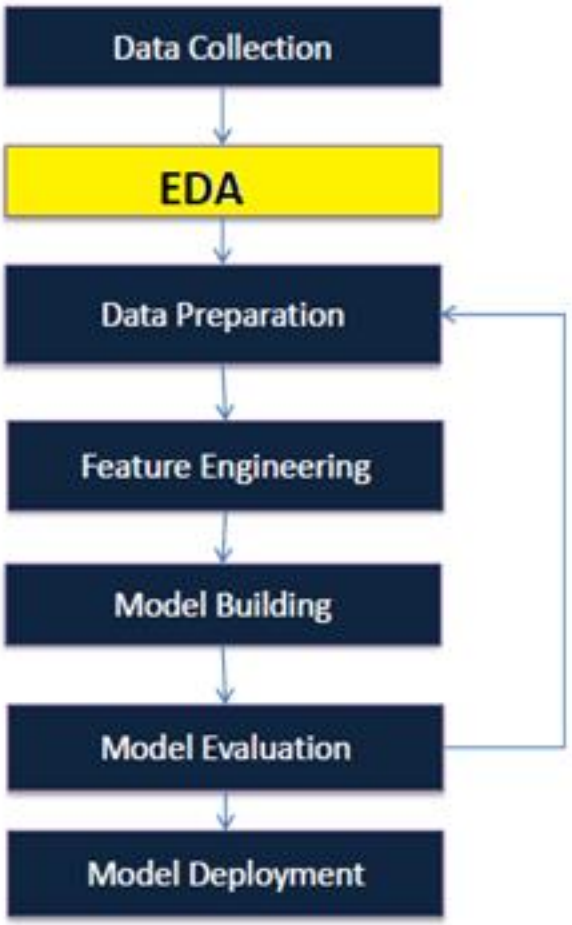
Educator Name: Nishant Dhote
Support Team: **+91-7880-113-112**

----Today Topics | Day 07----

- **EDA: Exploratory Data Analysis**

- **EDA Univariate Analysis (Day 06)**
- **EDA Bivariate Analysis**
- **EDA Multivariate Analysis**





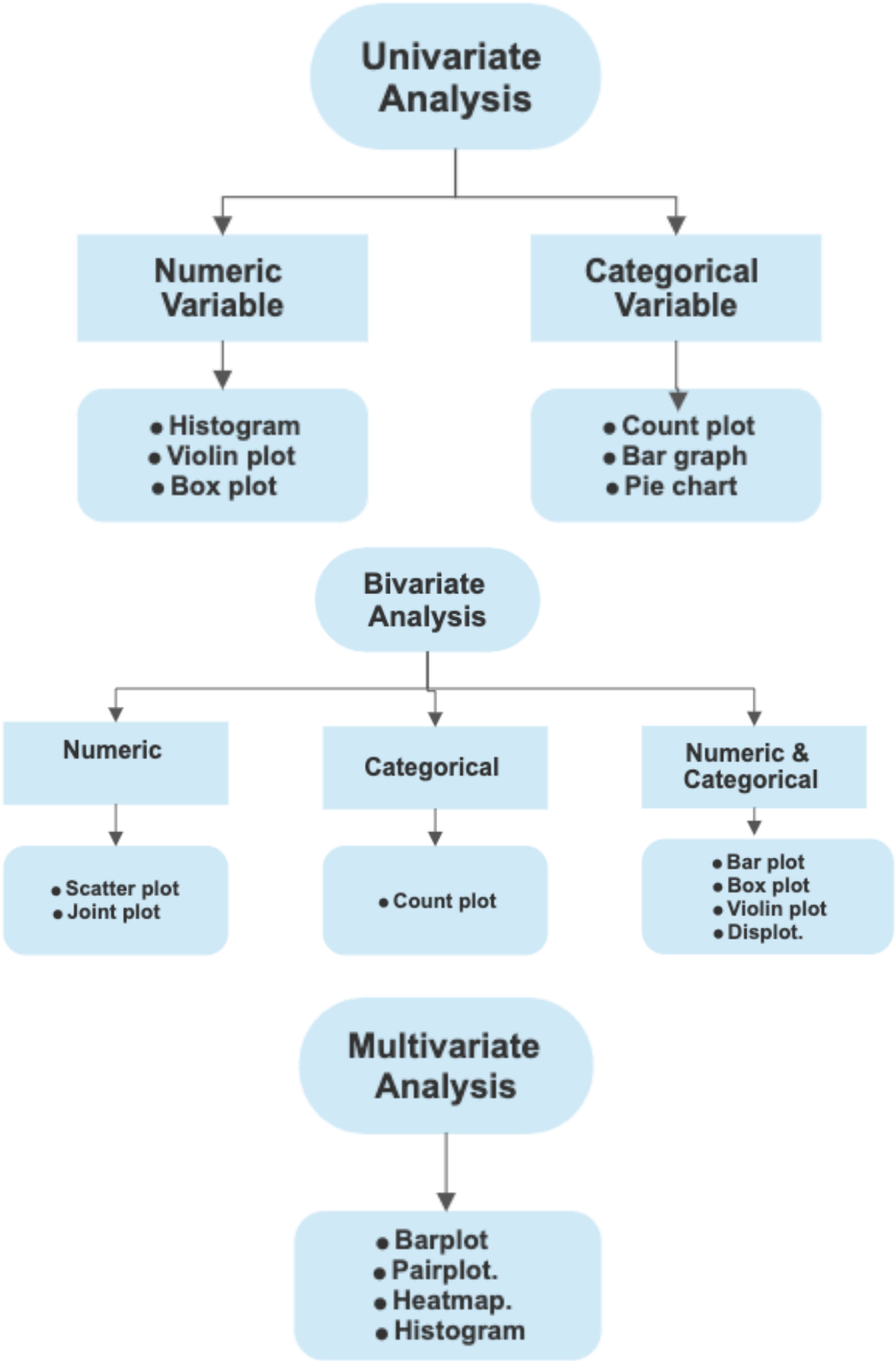
TYPES OF EXPLORATORY DATA ANALYSIS:

1. Univariate Analysis
2. Bivariate Analysis
3. Multivariate Analysis

Univariate EDA involves looking at a single variable at a time. Univariate EDA can help you understand the data distribution and identify any outliers.

Bivariate EDA involves looking at two variables at a time. Bivariate EDA can help you understand the relationship between two variables and identify any patterns that might exist.

Multivariate EDA involves looking at three or more variables at a time. Multivariate EDA can help you understand the relationships between several variables and identify any complex patterns or outliers that might exist.


Univariate Analysis
Numeric Variable
Categorical Variable
• Histogram
• Violin plot
• Box plot
• Count plot
• Bar graph
• Pie chart
Bivariate Analysis
Numeric
Categorical
Numeric & Categorical
• Scatter plot
• Joint plot
• Count plot
• Bar plot
• Box plot
• Violin plot
• Displot.
Multivariate Analysis
• Barplot
• Pairplot.
• Heatmap.
• Histogram

Dataset Link Kaggle: https://www.kaggle.com/competitions/titanic

GitHub Link:
https://github.com/TheiScale/30_Days_Machine_Learning/tree/main/Day%206%20ML

#Import Library

```
import pandas as pd
import seaborn as sns
```

#Import Datasets 1: Titanic

```
titanic = pd.read_csv('train.csv')
```

#Use Datasets 2: Hotel Bill & Tips

```
bill = sns.load_dataset('tips')
```

#Use Datasets 3: USA Flights

```
flights = sns.load_dataset('flights')
```

#Use Datasets 4: Iris Flower

```
irisflwr = sns.load_dataset('iris')
```

#1<Numerical – Numerical> | Scatterplot

```
sns.scatterplot(x=bill['total_bill'],y=bill['tip'])
```

```
----
```

```
sns.scatterplot(x=bill['total_bill'],y=bill['tip'],hue=bil
l['sex'],style=bill['smoker'],size=bill['size'])
```

#2<Numerical –Categorical> | Bar plot

```
sns.barplot(x=titanic['Pclass'],y=titanic['Age'])
----
```

```
sns.barplot(x=titanic['Pclass'],y=titanic['Fare'],
hue=titanic['Sex'])
```

#3<Numerical –Categorical> | Box Plot

```
sns.boxplot(x=titanic['Sex'],y=titanic['Age'])
```

```
sns.boxplot(x=titanic['Sex'],y=titanic['Age'],hue
 =titanic['Survived'])
```

#4<Numerical –Categorical> | Dist Plot

```
sns.distplot(titanic[titanic['Survived']==0]['Age'
],hist=False)

sns.distplot(titanic[titanic['Survived']==1]['Age'
],hist=False)
```

#5<Categorical – Categorical> | HeatMap

```
titanic.head(5)
```

```
pd.crosstab(titanic['Pclass'],titanic['Survived'])
```

```
 sns.heatmap(pd.crosstab(titanic['Pclass'],titanic
['Survived']))
```

#6<Categorical – Categorical> | ClusterMap

```
pd.crosstab(titanic['Parch'],titanic['Survived'])
```

```
sns.clustermap(pd.crosstab(titanic['Parch'],titani
c['Survived']))
```

#7<Numerical – Numerical -Categorical> | Pare Plot

```
irisflwr.head()

----
sns.pairplot(irisflwr)

----
sns.pairplot(irisflwr,hue='species')
```

EDA MCQ Question for Placement Practice:

https://www.examveda.com/data-science/practice-mcq-question-on-exploratory-data-analysis-(eda)/

Data Story Telling (Day 7): Curious Data Minds

Data science use cases in government?

How data science is used in national security?

Read Blog: https://activewizards.com/blog/top-12-data-science-use-cases-in-government/

Read Blog: https://www.iasparliament.com/article/data-analytics-and-national-security