

# Data and Features

Types of Data and Distribution  
Features, Inference and Visualization

# Types of Data

## Based on Structure

### STRUCTURED, UNSTRUCTURED AND SEMI-STRUCTURED

Structured Data



Semi-structured Data



Unstructured Data



# Types of Data

## Structured Data

- Standardised format for providing information about a page
- Classifying the page content in a organised manner
- Databases
- Tables
- Arrays
- So on...

# Types of Data

## Semi-structured Data

- Standardised format for providing information about a page
- Classifying the page content in a organised manner
- JSON files
- Lists
- Bullet Points
- So on...

# Types of Data

## Unstructured Data

- Unstructured data is the data which does not conform to a data model
- has no easily identifiable structure such that it can not be used
- Text-files
- Rich media
- Surveillance data like Weather Data, Geo-spatial Data
- So on...

# Types of Data

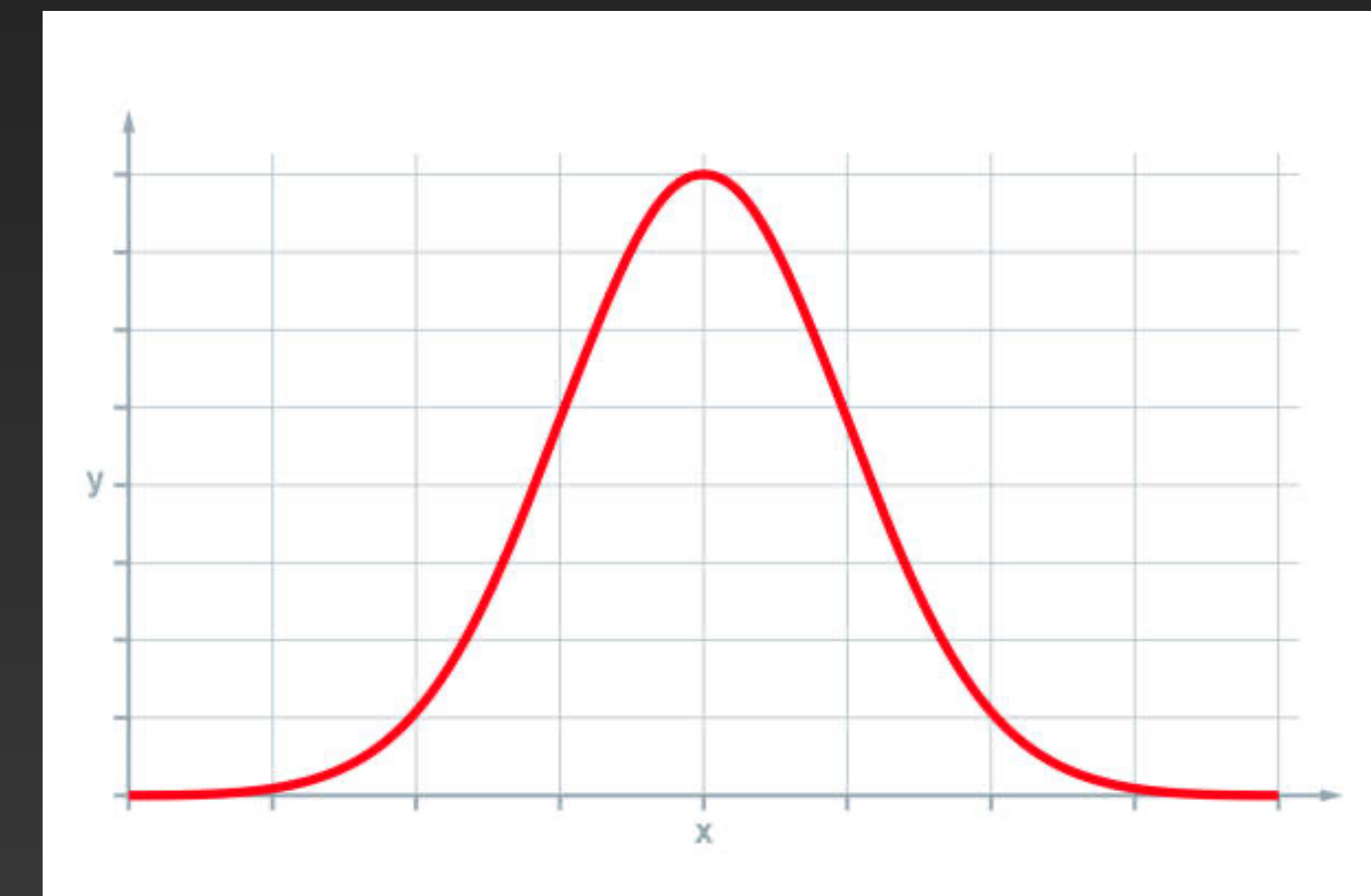
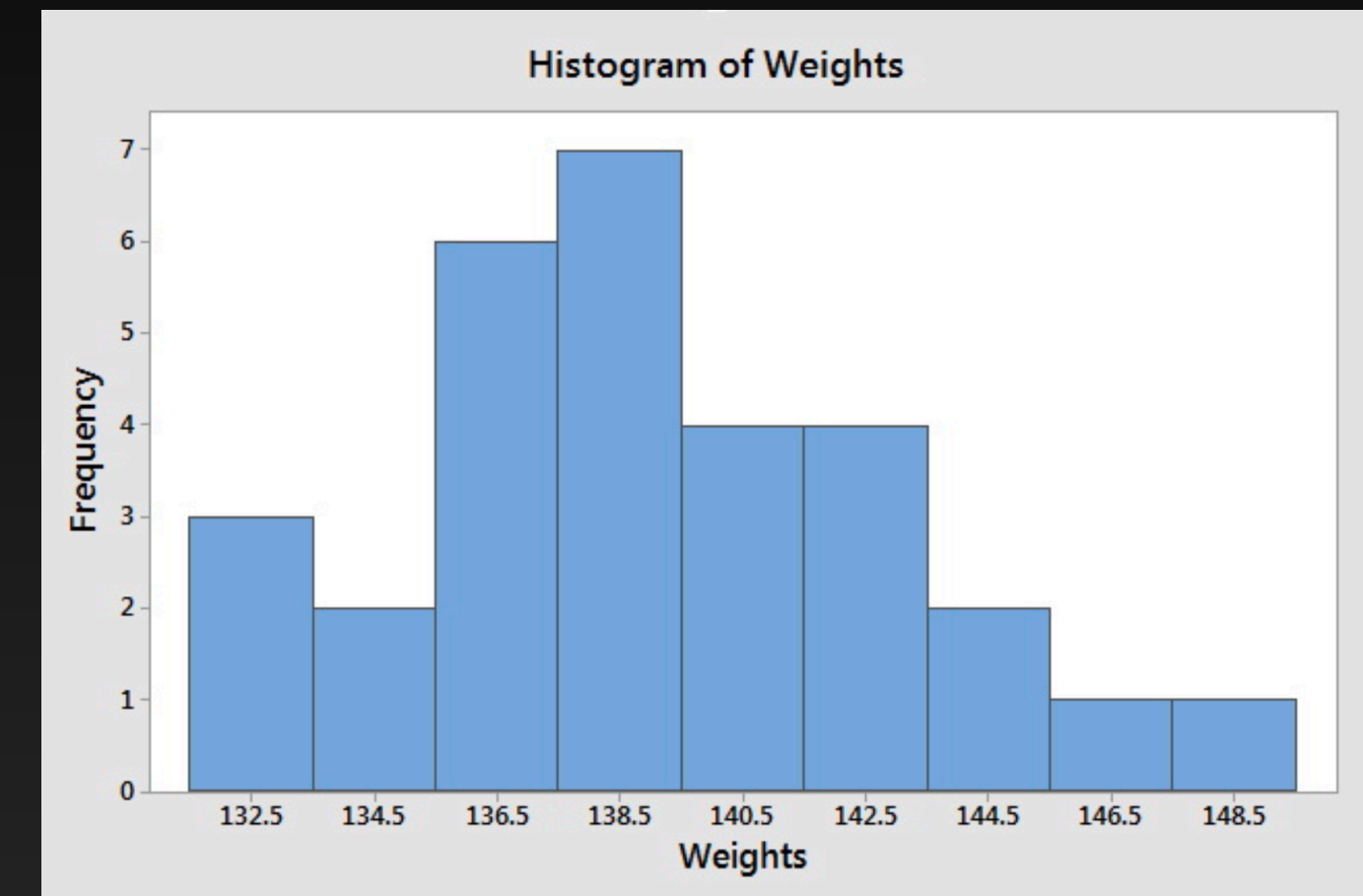
## Based on Distribution

- Discrete and Continuous
- Symmetrical or Skewed
- Left Skewed or Positively Skewed Distribution
- Right Skewed or Negatively Skewed Distribution

# Types of Data

## Discrete and Continuous

- Discrete Distributions include
  - Binomial Distribution
  - Poisson Distribution
- Continuous Distributions include
  - Normal Distribution
  - Students T Distribution



# Features Of Data

## Introduction.

- Each feature, or column, **represents a measurable piece of data that can be used for analysis**: Name, Age, Sex, Fare, and so on.
- Features are also sometimes referred to as “variables” or “attributes.”
- Depending on what you're trying to analyse, the features you include in your dataset can vary widely.



# Features Of Data

## Example

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

df = pd.read_csv('iris_EDA.csv')
df.head(5)
```

- Features of Iris DataSet

	sepallength	sepalwidth	petallength	petalwidth	class	Name	Score	Color
0	5.1	3.5	1.4	0.2	Iris-setosa	F1	12.0	Red
1	4.9	3.0	1.4	0.2	Iris-setosa	F2	NaN	Blue
2	4.7	3.2	1.3	0.2	Iris-setosa	F3	18.0	Orange
3	4.6	3.1	1.5	0.2	Iris-setosa	F4	14.0	Purple
4	5.0	3.6	1.4	0.2	Iris-setosa	F5	22.0	Red

# Features Of Data Inferences

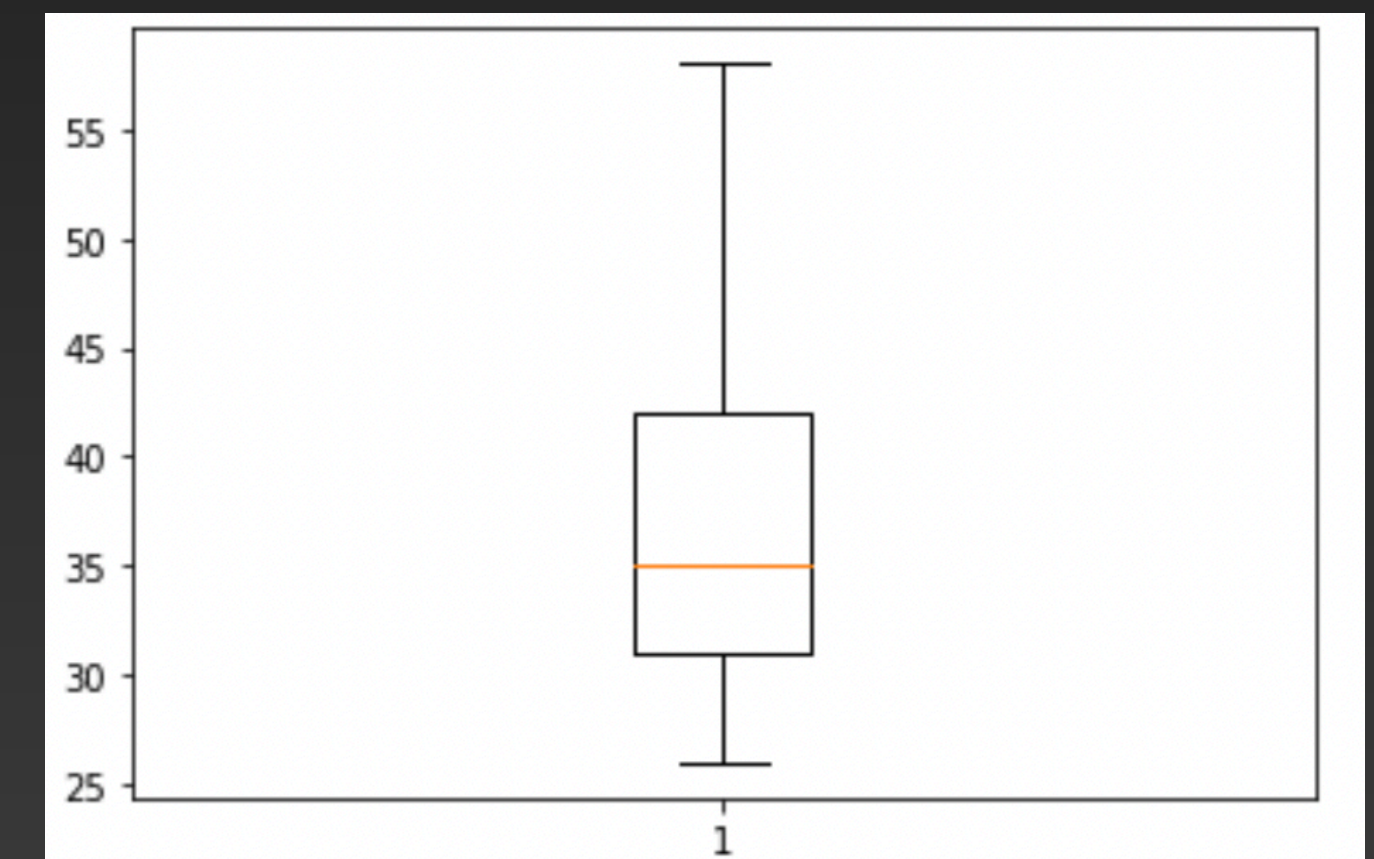
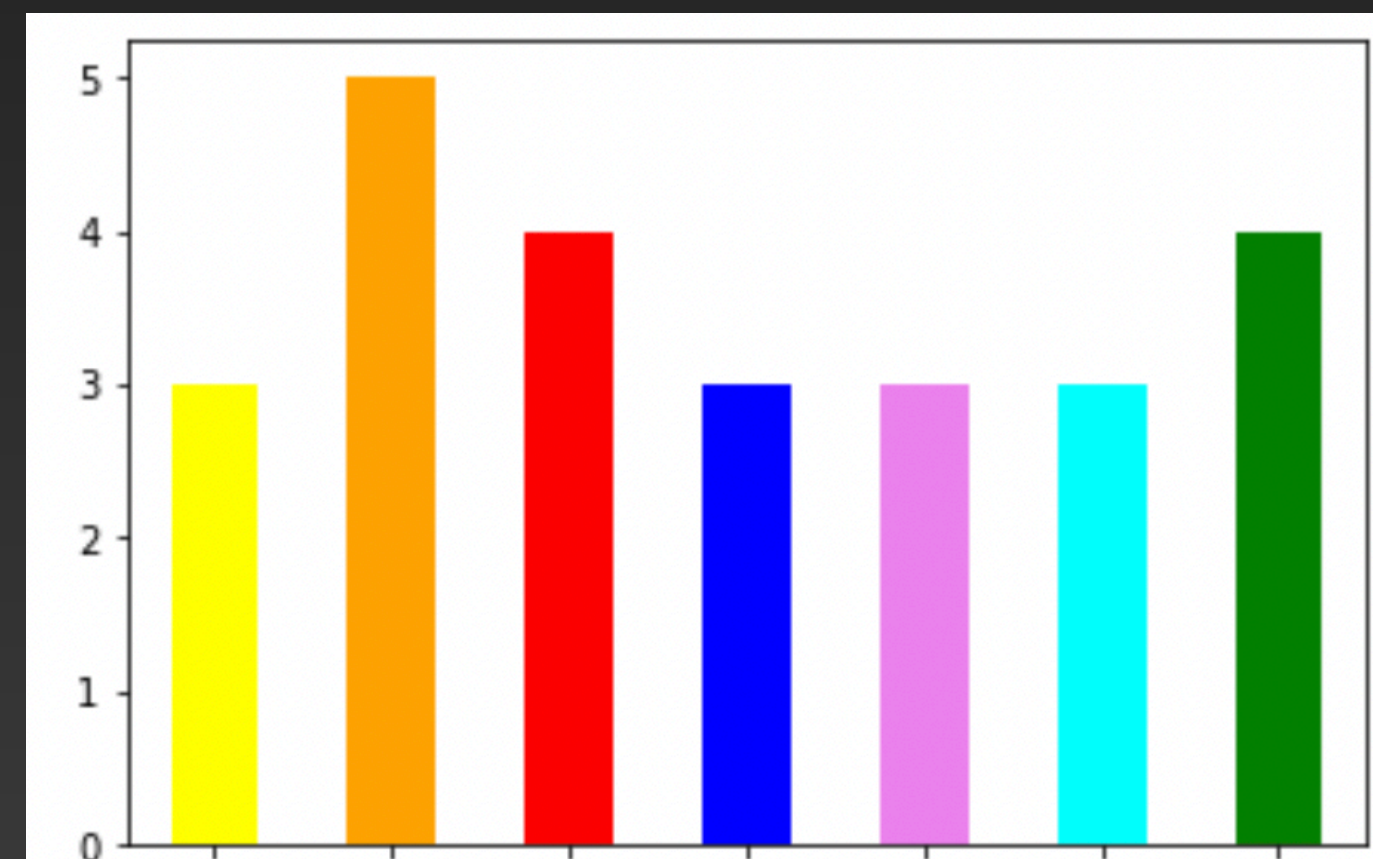
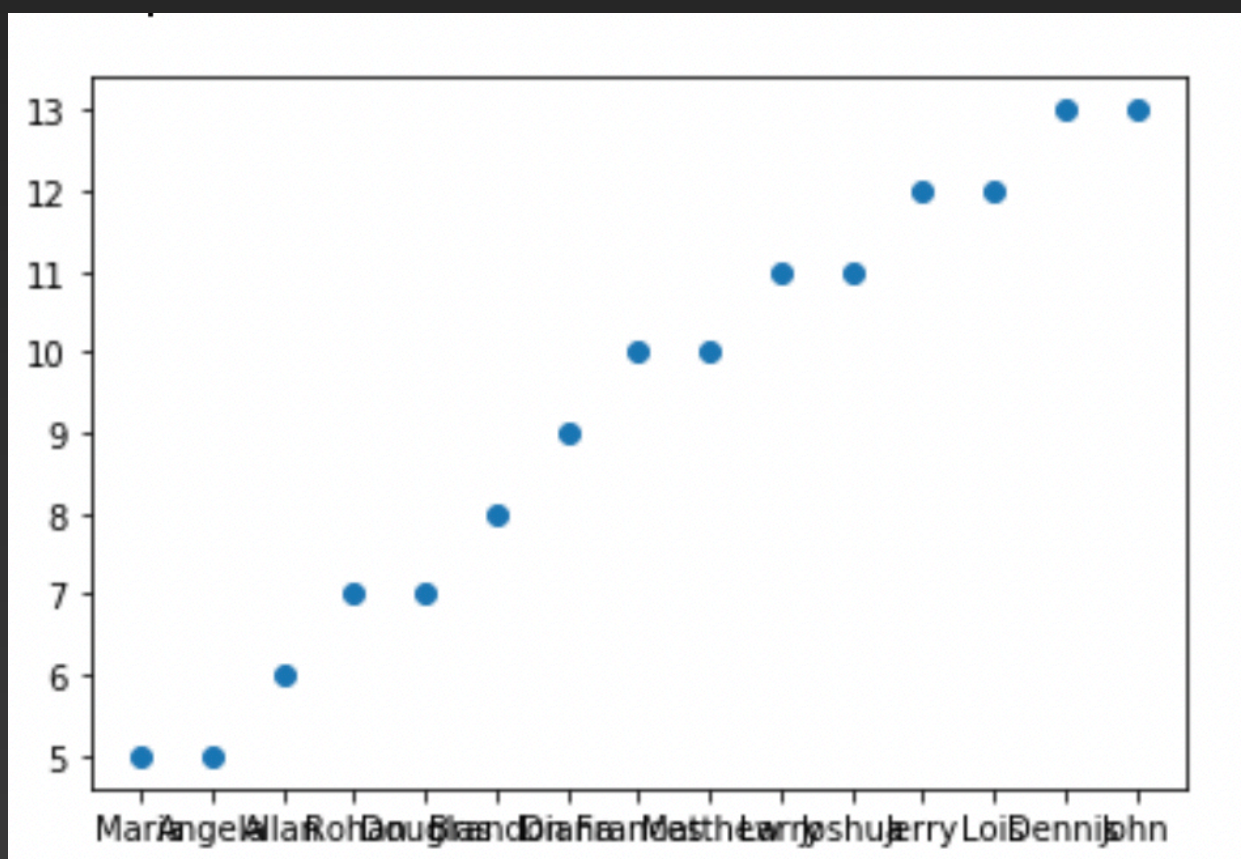
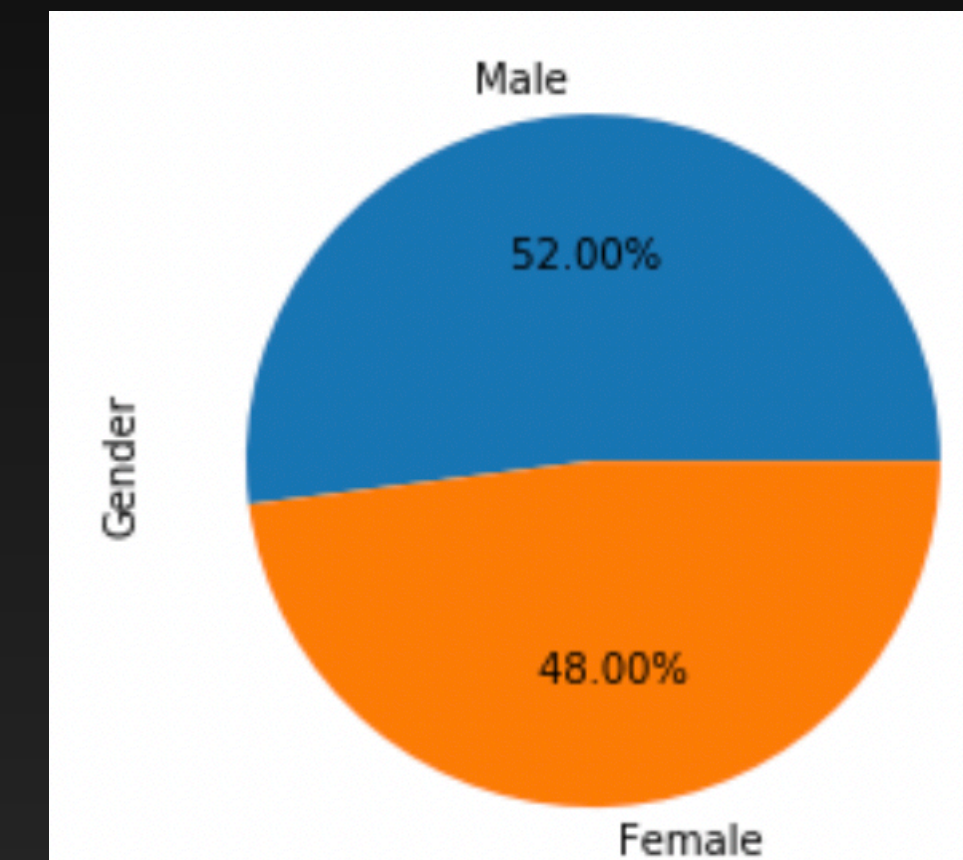
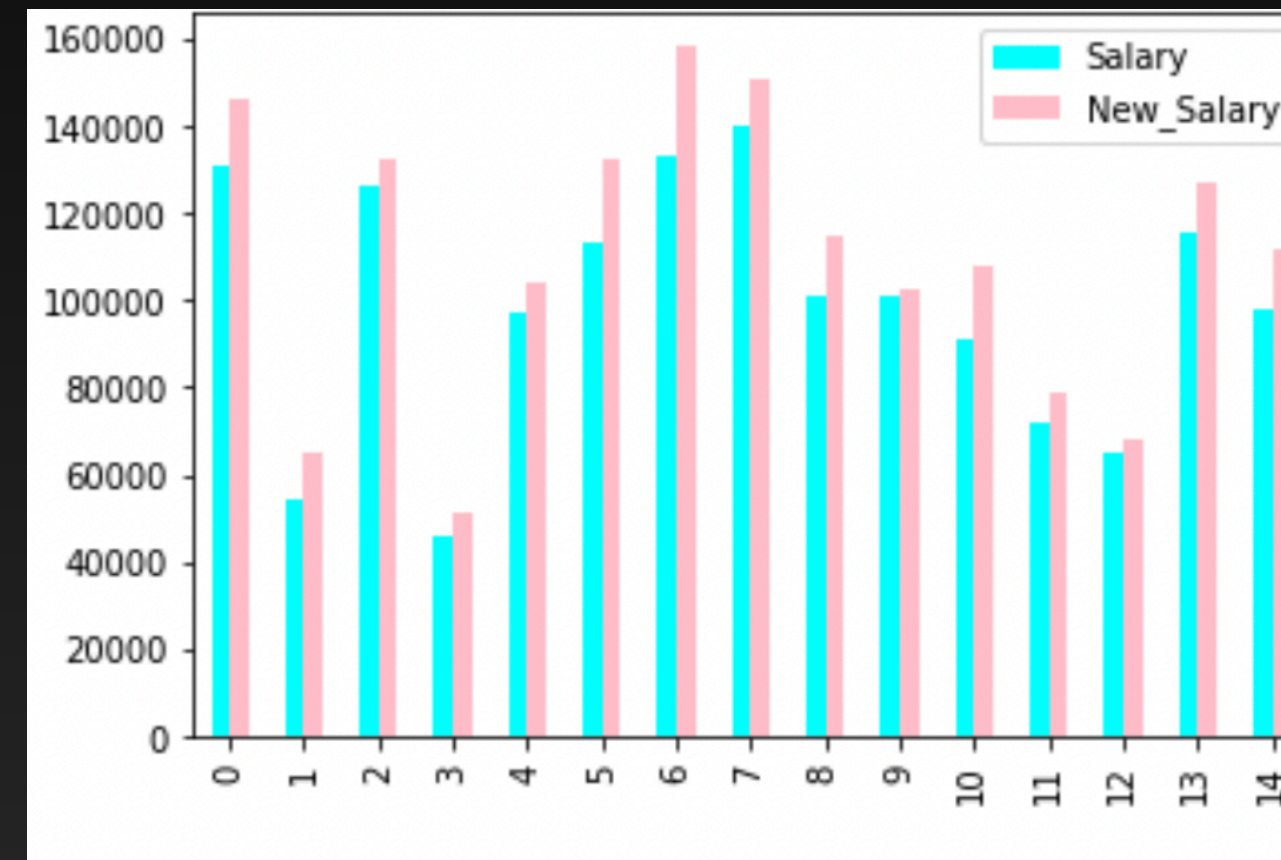
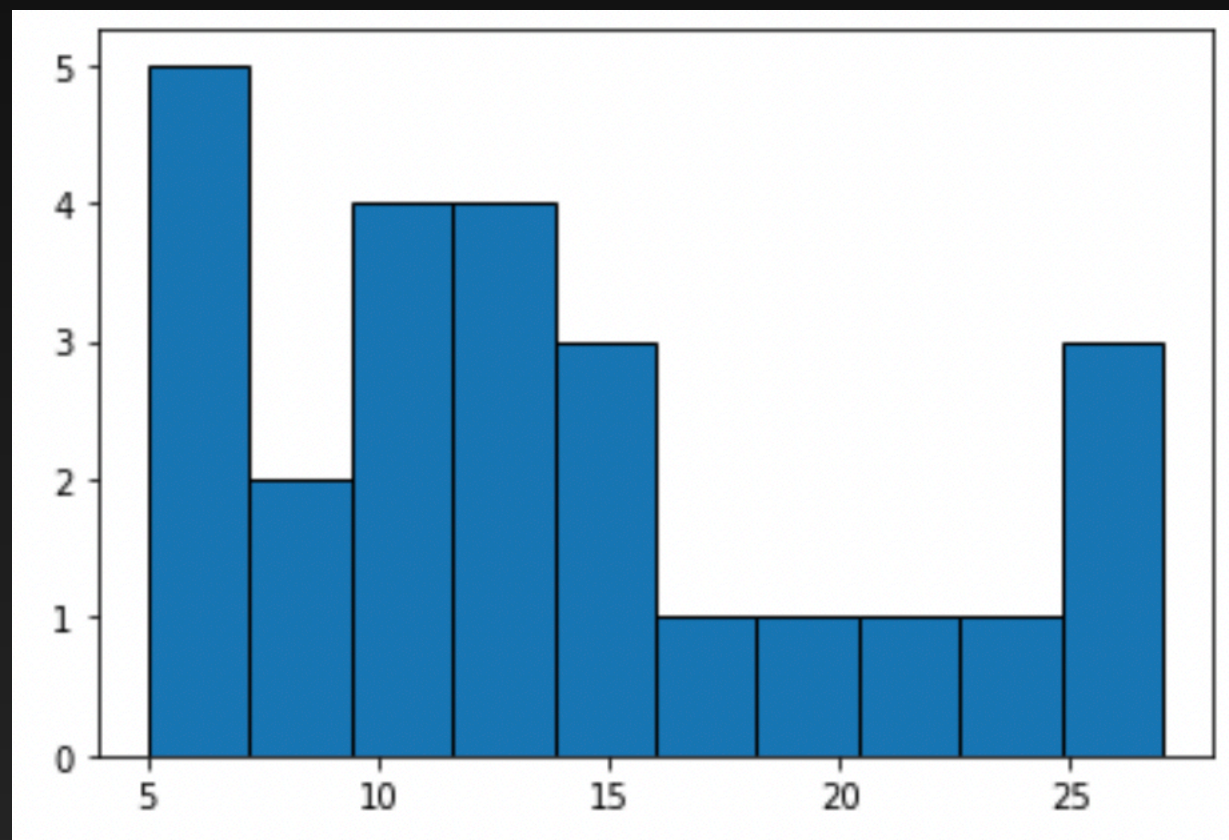
- `Dataframe.head()`
- `Dataframe.tail()`
- `Dataframe.describe()`
- `Dataframe.info`
- `Dataframe.shape()`
- `Dataframe.unique()`
- `Dataframe.loc[:,:]`
- `Dataframe.iloc[:,:]`

# Features Of Data Inferences.

Lets See Some  
Solved Examples



# Features Of Data Visualisation.



THE END