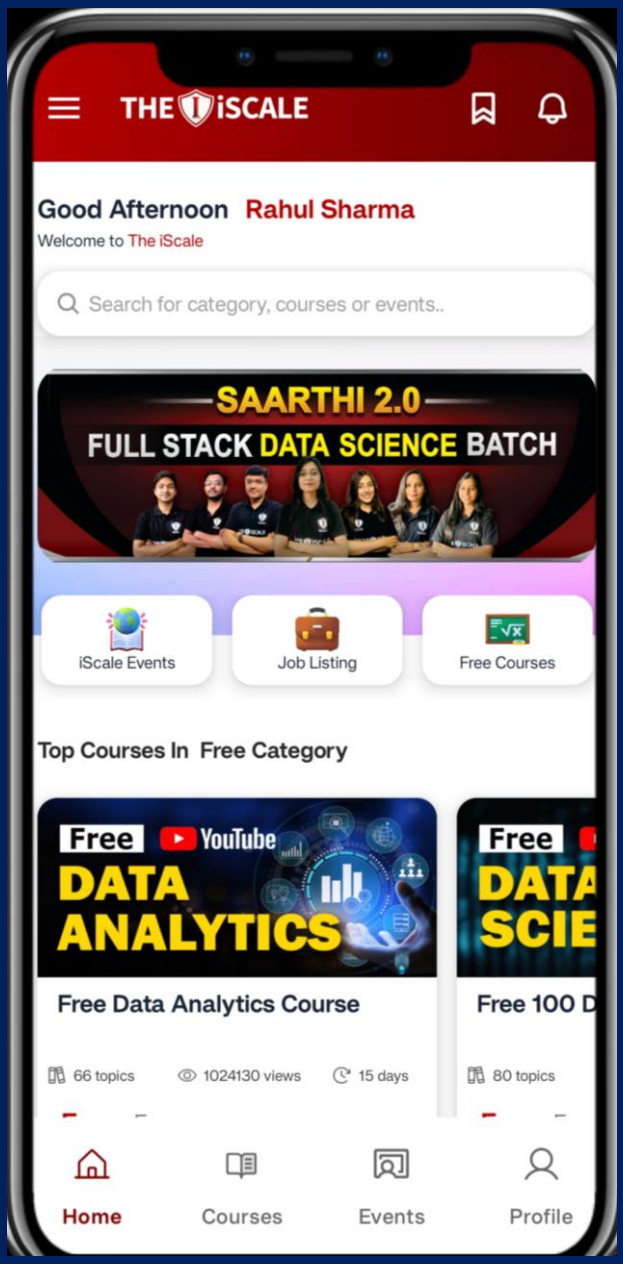
The iSCALE

Statistics Full Course Notes



Contact - 78801-13112

www.theiscale.com

Statistics

- What is Statistics
- Types of statistics
- Descriptive statistics
- Inferential Statistics
- Measure of central tendency
- Mean, mode, median
- Population mean, Sample mean
- Finding mean – manual calculation
- Mean using numpy
- Mean using stats
- Average of the data set
- Median
- Why Median is different from mean
- Median using numpy
- Median using Statistics
- Median using Pandas
- Evaluating median in a dataset
- Mode
- Scipy Library
- Mode using statistics module
- Uisng Scipy library
- Handling multimodal datasets
- Measure of Dispersion
- Range
- Variance
- Population & sample variance
- Why sample variance formula is different than population
- Variance using Numpy
- Variance using Pandas
- Variance using Scipy
- Standard Deviation
- Standard Deviation using numpy

iSCALE

Statistics

- Standard Deviation using pandas
- Deviation using Scipy
- Data Distribution
- Normal/Gaussian Disribution
- Gaussian distribution - code
- Skewness
- Right skewed curve
- Positive skew curve - code
- Left skewed curve
- Negative skew curve - code
- Uniform Distribution
- Code for uniform distribution
- Bi-modal distribution
- Code for Bi-modal curve
- Multimodal distribution
- Code for Histogram multimodal
- Inferential statistics
- Estimate
- Point estimate
- Interval estimate
- Confidence interval
- Significance value
- Hypothesis testing
- Mechanism for hypothesis testing
- P-test
- P- test use case via code
- T –test
- T-test use case via code
- Z – test
- Z- table
- Z- test use case via code
- Two test trail
- Practice questions

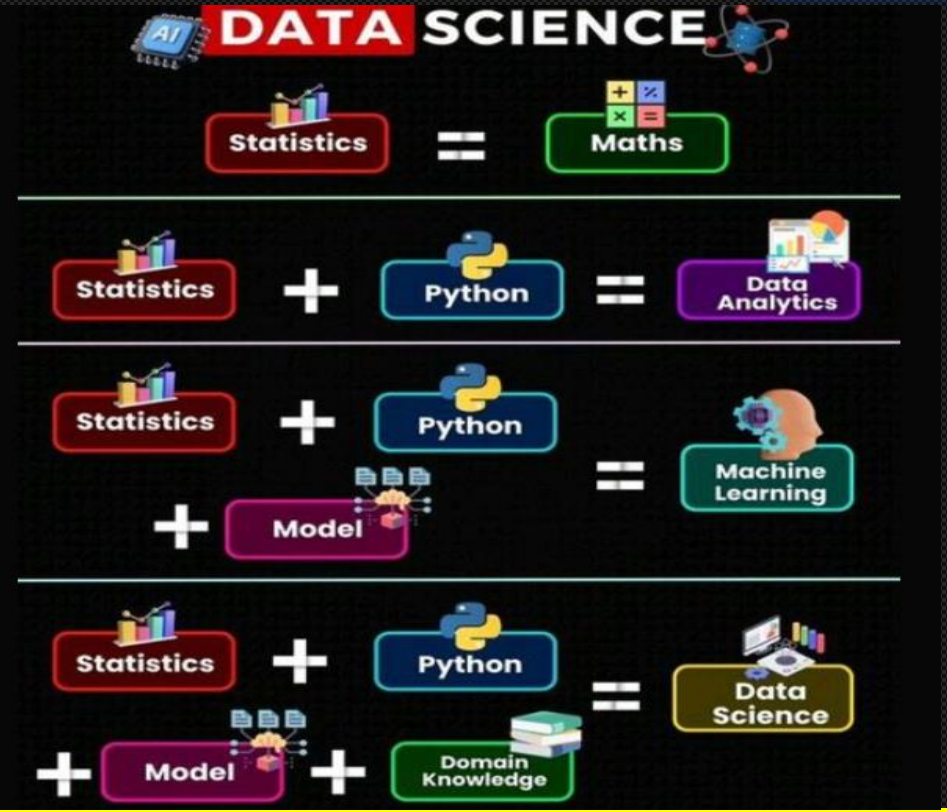
iSCALE

Statistics

For which Job Roles can this statistics course be helpful?

- Data Scientist
- Data Analyst
- Business Analyst
- Machine Learning Engineers
- AI Engineers
- BI Developers

Statistics

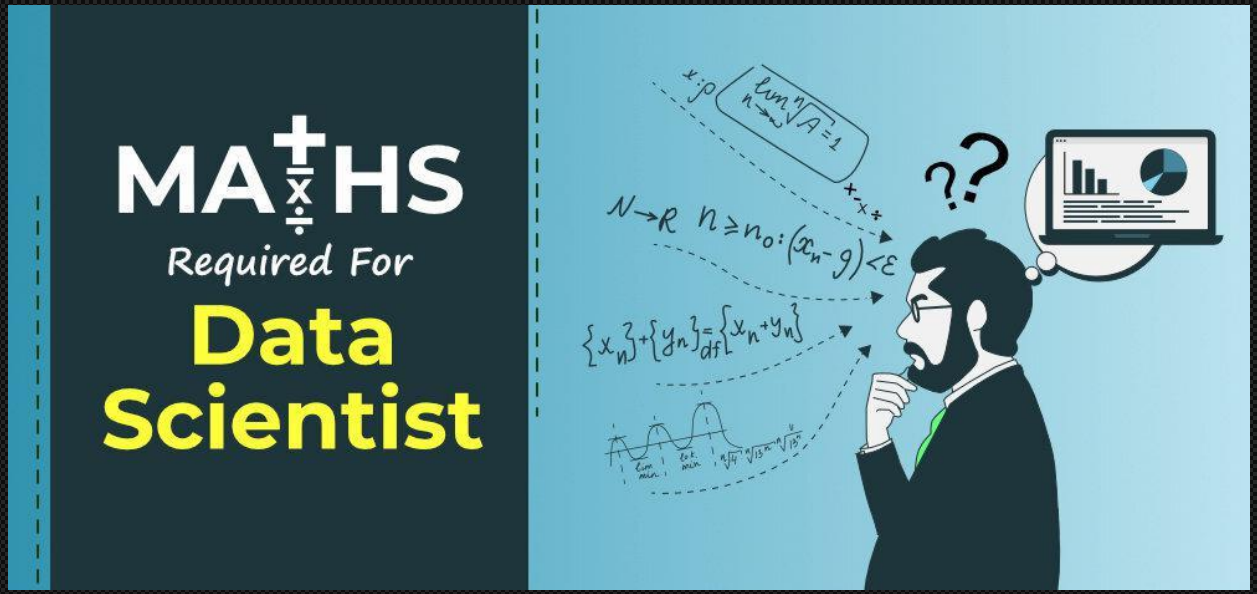


Why are we studying statistics for data science?

Statistics

Data Science & Analytics

Why are we studying statistics for data science?



Statistics

Why are we studying statistics for data science?

Descriptive Statistics

Regression Analysis

k-means clustering

time series analysis

A/B testing

Types of Statistics

Descriptive Statistics

Descriptive statistics ka kaam hai kisi bhi dataset ke main features ko summarize, analyze, visualize aur describe karna.

Inferential Statistics

Inferential statistics ka use karke hum ek chhoti si sample data se puri population ke baare mein predictions ya inferences bana sakte hain.

Descriptive Statistics

Descriptive statistics ka kaam hai kisi bhi dataset ke main features ko summarize, analyze, visualize aur describe karna.

Descriptive Statistics

Descriptive statistics ka kaam hai kisi bhi dataset ke main features ko summarize, analyze, visualize aur describe karna. Iska main purpose yeh hota hai ki hum data ke andar chhipe hue patterns ko samajh sakein aur unhe simple terms mein explain kar sakein.

Common Techniques

Measures of Central Tendency - Mean (average), Median, aur Mode

Measures of Variability - Range, Variance, Standard Deviation

Data Representation - Charts, graphs, tables, frequency distributions

Inferential Statistics

Inferential Statistics (Inferential Statistics) ka purpose hota hai kisi sample data ka use karke population ke baare mein conclusions aur predictions banana. Yeh statistics ke ek branch hai jo data analysis ke baad generalized statements aur hypotheses test karne mein madad karta hai.

Common Techniques

- Hypothesis Testing
- Confidence Intervals
- Regression Analysis
- Correlation

Practice Question 1

What is Descriptive Statistics?

A. A method to predict future trends
B. A method to summarize and organize data
C. A method to test hypotheses
D. A method to establish a relationship between two variables

Practice Question 2

Which of the following techniques is used in Inferential Statistics?

A. Mean
B. Hypothesis Testing
C. Frequency Distribution
D. Median

Practice Question 3

What is the main difference between Descriptive and Inferential Statistics?

A. Descriptive statistics summarize data, while inferential statistics make predictions or inferences about a population
B. Descriptive statistics are based on probability, while inferential statistics are not
C. Inferential statistics focus on organizing data, while descriptive statistics focus on making predictions
D. Both deal with making predictions about data

Measure of Central Tendency

Measures of Central ka matlab hota hai data set ke center ko samajhna. Inka use karke hum jaan sakte hain ki data ke values kis jagah ke aas-paas clustered hain. Yeh three main types ke hote hain: Mean, Median, aur Mode.

Mean

Mean ek statistical measure hai jo central tendency ko represent karta hai. Yeh ek dataset ke sabhi values ka average hota hai. Hum mean ko do tarikon se calculate karte hain: **Population Mean** aur **Sample Mean**.

Population Mean

Population Mean tab calculate kiya jata hai jab aapke paas poore population ka data available ho. Population mean ko μ se denote kiya jata hai.

Formula: $\mu = \frac{\Sigma X_i}{N}$

- ΣX_i: Sabhi data points ka sum (i = 1 se N tak).
- **N**: Total number of observations in the population.

Sample Mean

Sample Mean tab calculate kiya jata hai jab aap population se ek chhota sample lete hain. Sample mean ko $\bar{x}$ se denote kiya jata hai.

Formula: $\bar{X} = \frac{\Sigma X_i}{n}$

- ΣX_i: Sample data points ka sum (i = 1 se n tak).
- **n**: Total number of observations in the sample.

Median

The **median** is a measure of central tendency that represents the middle value of a dataset when it is arranged in ascending or descending order. Unlike the mean, the median is not affected by outliers or skewed data, making it a more robust measure for datasets with extreme values.

Steps to Calculate the Median

1. **Sort the dataset** in ascending or descending order.
2. **Determine the number of observations** n.
3. **Apply the appropriate formula** depending on whether n is odd or even.

Mode

In statistics, the **mode** is a measure of central tendency that represents the value that appears most frequently in a dataset. Unlike the mean (average) and median (middle value), the mode focuses solely on the frequency of values.

Consider the following dataset of exam scores:
[85, 90, 90, 95, 85, 80, 85, 90, 80]

List Frequencies:

85 appears 3 times
90 appears 4 times
95 appears 1 time
80 appears 2 times
Determine Mode:

The value 90 appears most frequently (4 times), so 90 is the mode of this dataset.

Scipy Library

The scipy library is a powerful open-source Python library used for scientific and technical computing. It builds on the numpy library and provides a wide range of functionalities for different types of scientific and engineering tasks.

The scipy library is widely used in various fields, including:

- Engineering: For optimizing designs and control systems.
- Finance: To optimize investment portfolios and forecast economic trends.
- Medical Imaging: Enhances and analyzes diagnostic images.
- Data Science: Fills missing data and optimizes machine learning models.
- Signal Processing: Analyzes audio and vibration signals for better quality and maintenance.

Practice Question 1

If the mode of a data set is 10, what does this imply?

a) 10 is the most frequently occurring value in the data set
b) 10 is the average of the data set
c) 10 is the middle value when the data is sorted
d) 10 is the sum of all values divided by the number of values

Practice Question 2

If a data set is perfectly symmetrical, which of the following statements is true?

a) The mean is less than the median
b) The mean is greater than the median
c) The mean and median are equal
d) The mode and median are always equal

Practice Question 3

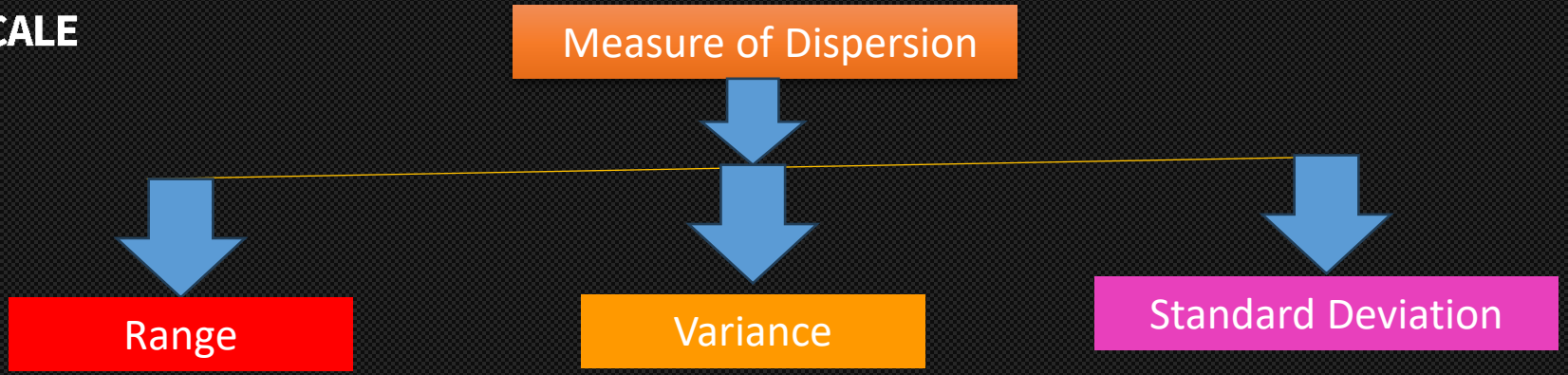
In which situation is the mean not a good measure of central tendency?

a) When the data set has extreme outliers
b) When the data set is normally distributed
c) When the data set has a small number of values
d) When the data set is uniformly distributed

Measure of Dispersion

Measure of dispersion , statistics mein data ke spread ya spread out hone ki extent ko measure karta hai. Ye bataata hai ki data points ek mean value se kitne door hain. Inka use data ke distribution ko samajhne aur compare karne ke liye hota hai.

Agar hum sirf mean (average) dekhte hain, to humein yeh nahi pata chalta ki data points mean se kitna door hain. Isliye dispersion ki measures help karti hain humein yeh samajhne mein ki data kitna spread out hai.


Measure of Dispersion
Range
Variance
Standard Deviation

Range

In descriptive statistics, **range** is a simple measure of **variability** or **spread** in a data set. It indicates how far the data points are spread out by showing the difference between the **maximum** and **minimum** values in the dataset.

Formula for Range:

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}$$

Key Points:

1.Range gives a quick sense of dispersion: If the range is small, it indicates that the data points are close to each other. A larger range suggests that the data points are more spread out.

2.Range is sensitive to outliers: Since the range only looks at the two extreme values (maximum and minimum), any outlier can significantly increase the range, even if the rest of the data is closely packed.

Variance

Variance ek statistical measure hai jo data set ki values ke beech ki variability ya dispersion ko quantify karta hai. Yani, yeh batata hai ki data ke individual values mean (average) se kitna door hain, aur yeh data set ka spread kitna hai.

Population Variance and **Sample Variance** are measures of dispersion that quantify how much data points differ from the mean.

1. Population Variance

Formula: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

Where:

- σ^2 = Population Variance
- N = Number of data points in the population
- x_i = Each data point
- μ = Mean of the population

2. Sample Variance

Formula: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Where:

- s^2 = Sample Variance
- n = Number of data points in the sample
- x_i = Each data point
- $\bar{x}$ = Mean of the sample

Standard Deviation

Standard Deviation ek important measure of dispersion hai statistics mein. Yeh batata hai ki aapke data points average ke aas-paas kitna spread (faila) hue hain. Simple terms mein, standard deviation se aapko pata chalta hai ki aapke data values mean (average) se kitni door ya paas hain.

Population ke liye: $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$

Sample ke liye: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Yahan:

- σ = Population Standard Deviation
- s = Sample Standard Deviation
- N = Population mein total data points
- n = Sample mein total data points
- x_i = Each data point
- μ = Population Mean
- $\bar{x}$ = Sample Mean

Practice Question- 1

Which measure of dispersion is affected the least by outliers?

a) Range
b) Variance
c) Standard Deviation
d) Median

Practice Question- 2

Which of the following statements is true about the standard deviation?

a) It can never be zero
b) It is always larger than the range
c) It is the square root of the variance
d) It is less informative than the variance

Practice Question- 3

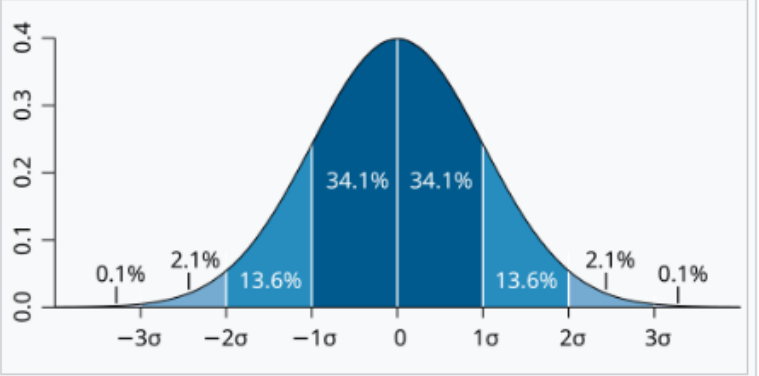
If a data set has a range of 30 and a standard deviation of 6, what can you infer about the data?

a) The standard deviation is greater than the range
b) The range is likely to be larger than the standard deviation
c) The data set is normally distributed
d) The variance is 36

Data Distribution

Normal Distribution (Gaussian Distribution)

Ye distribution bell-shaped hoti hai, jahan data ka majority center ke aas-paas concentrated hota hai aur gradually decrease hota hai dono taraf. Iska mean, median, aur mode sab ek hi point par hota hai.



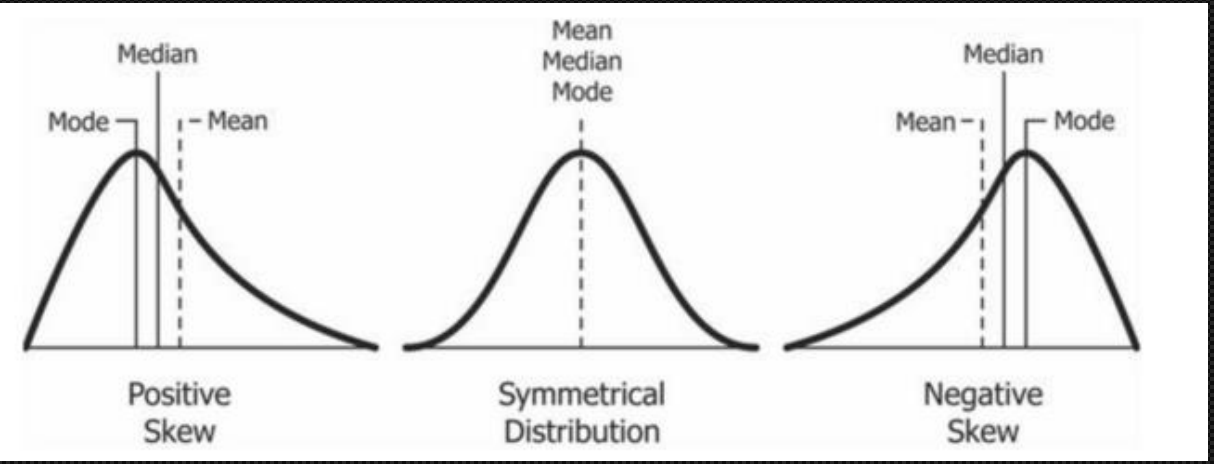
For the normal distribution, the values less than one standard deviation from the mean account for 68.27% of the set; while two standard deviations from the mean account for 95.45%; and three standard deviations account for 99.73%.

iSCALE

Data Distribution

Skewed Distribution

Skewness ek statistical measure hai jo data distribution ki symmetry ko describe karta hai. Iska use distribution ke shape ko samajhne ke liye kiya jata hai. Skewness ko positive, negative, ya zero ke roop me categorize kiya jata hai.



Positive Skewness (Right-Skewed)

Mean > Median > Mode

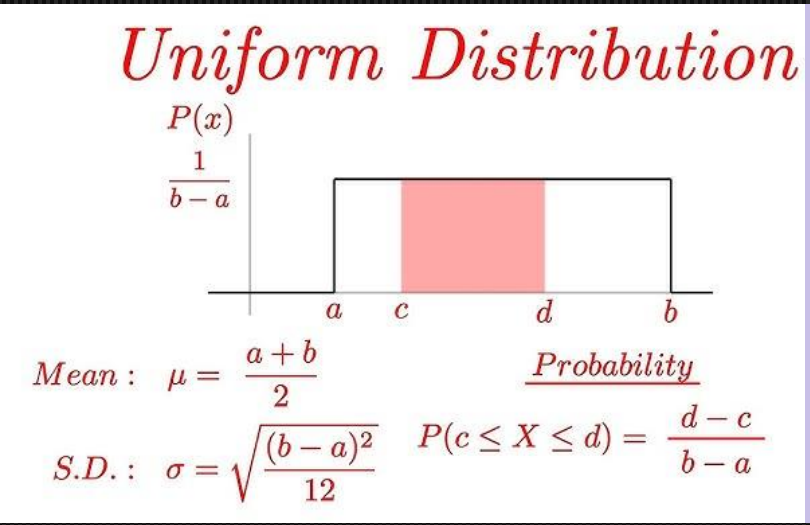
Negative Skewness (Left-Skewed)

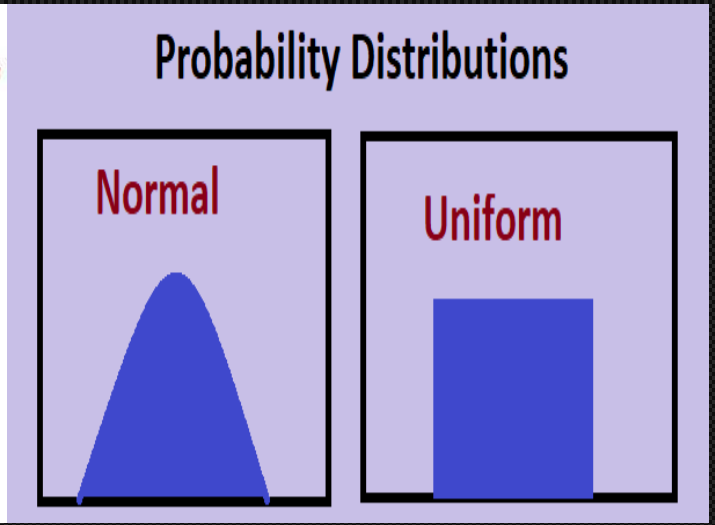
Mean < Median < Mode

Data Distribution

Uniform Distribution

Isme sabhi values ki probability equal hoti hai. Graph ek straight horizontal line hoti hai jo indicate karti hai ki kisi bhi value ki occurrence equal chance ke saath hoti hai.





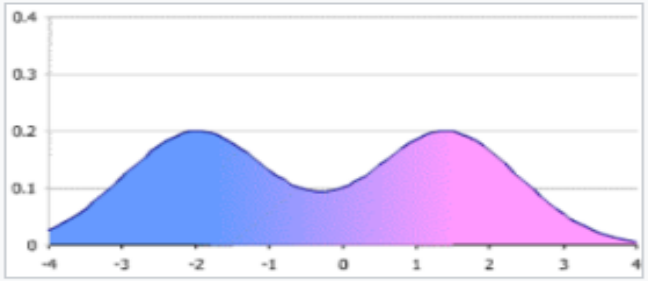


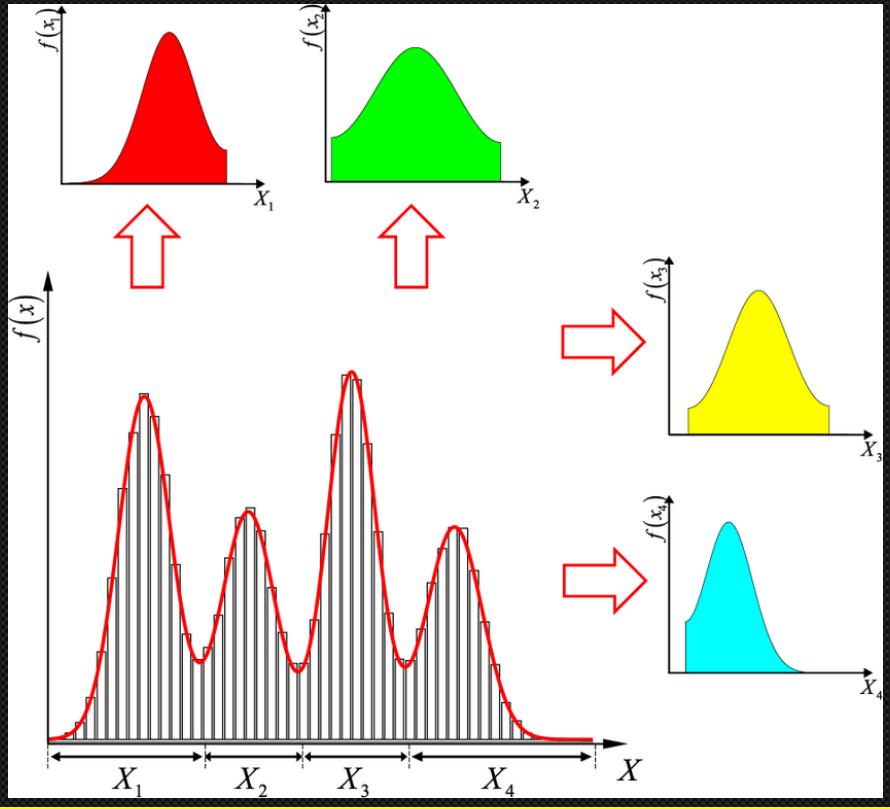
Figure 1. A simple bimodal distribution, in this case a mixture of two normal distributions with the same variance but different means. The figure shows the probability density function (p.d.f.), which is an equally-weighted average of the bell-shaped p.d.f.s of the two normal distributions. If the weights were not equal, the resulting distribution could still be bimodal but with peaks of different heights.

Data Distribution

Bimodal Distribution

Bimodal Distribution ek statistical distribution hai jisme data ke do distinct peaks ya modes hote hain. Iska matlab hai ki data do alag-alag clusters ya groups me divided hota hai, aur in clusters ki frequencies alag-alag peaks ki form me show hoti hain.

Data Distribution



Multimodal Distribution ek statistical distribution hai jisme data ke do se zyada distinct peaks ya modes hote hain. Ye distribution complex data sets ko represent karta hai jahan data multiple distinct groups ya clusters me divided hota hai.

Inferential Statistics

Estimate

Inferential Statistics me "estimate" ka matlab hota hai kisi population parameter (jaise population ka mean ya proportion) ko ek sample data ke basis pe approximate karna. Jab bhi hum kisi bade group (population) ke baare me kuch conclude karte hain sample data ke use se, to usse "estimation" kehte hain.

Types of Estimates in Inferential Statistics:

Point Estimate

Interval Estimate

Inferential Statistics

Estimate

Types of Estimates

Point Estimate:

Definition: Point estimate ek single value hoti hai jo sample data se kisi population parameter ka best guess hoti hai.

Example: Agar aapke sample ka mean 50 hai, toh aapka point estimate population mean ke liye 50 hoga.

Interval Estimate:

Definition: Interval estimate ek range provide karta hai jismein population parameter hone ki probability hoti hai. Yeh zyada reliable hota hai kyunki yeh ek range provide karta hai, na ki ek single value.

Example: Agar aapke sample data se aapka interval estimate 45 se 55 hai, toh aap keh sakte hain ki population mean 45 se 55 ke beech ho sakta hai with a certain level of confidence.

Inferential Statistics

Confidence Interval (CI) ek range hota hai jo estimate karta hai ki kisi population parameter ki value kis range ke andar fall kar sakti hai, aur yeh ek confidence level ke sath hota hai. Confidence level yeh batata hai ki agar aap multiple samples lenge aur unke confidence intervals calculate karenge, toh usme se kitne intervals population parameter ko cover karenge.

For Example:

Agar aapke sample mean ke liye 95% confidence interval 45 se 55 hai, iska matlab hai ki aapke population mean ka 95% chance hai ki yeh range (45 se 55) ke andar hoga.

Inferential Statistics

Hypothesis Testing

inferential statistics mein ek fundamental method hai jo help karta hai data ke basis pe decisions lene aur conclusions draw karne mein. Yeh process decide karta hai ki sample data ke basis pe population ke baare mein koi specific claim ya hypothesis true hai ya nahi.

Hypothesis Testing Kya Hai?

Hypothesis Testing ek systematic method hai jisme hum ek hypothesis (assumption) ko test karte hain aur decide karte hain ki sample data us hypothesis ko support karta hai ya nahi. Yeh statistical inference ka ek part hai jo ki probability theory pe based hota hai.

Inferential Statistics

Hypothesis Testing Mechanism

1) Null Hypothesis (H0)
The assumptions you are beginning with

2) Alternative Hypothesis (H1)
Opposite of Null Hypothesis

3) Experiments
Use Statistical methods (such as t-tests, z-tests, ANOVA, chi-square tests,etc) to analyze the data.

4)Accept the null hypothesis or reject the null hypothesis
It helps researchers and analysts assess the validity of claims, evaluate the effectiveness of interventions and draw meaningful conclusions about population

Inferential Statistics

p-Value Test

p-Value Test inferential statistics mein ek key concept hai jo hypothesis testing ke process ka essential part hota hai. Yeh ek measure hai jo batata hai ki sample data ko null hypothesis ke saath kitna consistent hai. Specifically, p-value se yeh assess kiya jata hai ki null hypothesis ko reject karne ke liye aapke sample data ke basis pe kitni evidence hai.

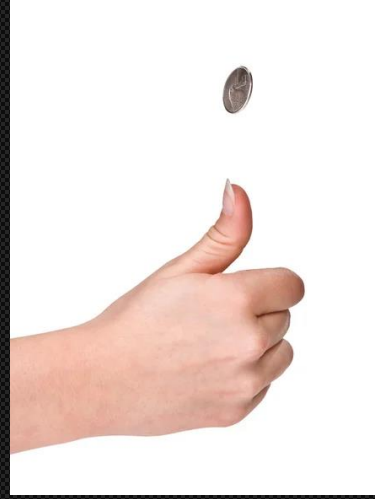
p-Value Kya Hai?

p-Value ek probability hai jo aapke sample data ya usse extreme results aane ki probability ko represent karta hai, assuming ki null hypothesis sach hai. Yeh batata hai ki aapke observed results (ya aur extreme results) ko observe karna kitna unlikely hai agar null hypothesis sach hoti.

iSCALE

Inferential Statistics

p-Value Test



In statistics, a t-test is a hypothesis test used to determine if there is a significant difference between the means of two groups. It's especially useful when dealing with small sample sizes or when the population standard deviation is unknown. Here’s a breakdown of what a t-test involves:

Inferential Statistics

T Test

In statistics, a t-test is a hypothesis test used to determine if there is a significant difference between the means of two groups. It's especially useful when dealing with small sample sizes or when the population standard deviation is unknown.

Inferential Statistics

Z - Test

Z-Value Test (Z-Test) inferential statistics ka ek method hota hai, jo tab use hota hai jab aapko test karna ho ki sample ka mean population mean ke barabar hai ya nahi. Z-test ko tab use kiya jata hai jab sample size bada ho (usually n>30n > 30n>30) ya jab population ka standard deviation σ\sigmaσ pata ho.

Inferential Statistics

Z - Test

The average height of all the residents in a city is 168 cm, with a standard deviation = 3.9. A doctor believes the mean to be different. He measured the height of 36 individuals and found the average height to be 169.5 cm.

1) State null and alternate hypothesis
2) At a 95% confidence level, is there enough evidence to reject the null hypothesis

iSCALE

Practice Question – 1

What is the purpose of a confidence interval in statistics?

A. To determine the exact value of a population parameter.
B. To provide a range of values within which the population parameter is expected to lie with a certain level of confidence.
C. To calculate the probability of a Type II error.
D. To test the null hypothesis.

Practice Question – 2

Which of the following is true about a z-test?

A. It is used when the population standard deviation is unknown.
B. It is used for hypothesis testing when the sample size is small and the population variance is known.
C. It is appropriate when the sample size is large and the population standard deviation is known.
D. It is used to test the relationship between two categorical variables.

Practice Question – 3

If the sample mean is significantly different from the population mean based on a z-test, what can be inferred?

A. The sample is not representative of the population.
B. The null hypothesis can be rejected, suggesting a significant difference.
C. The sample size is too small for the z-test.
D. The population variance is not known.

How to connect with us?

Instagram Handle

theiscale

The iScale Organization Handle

theiscale.founders

Siblings - Nishant Dhote & Swati Dhote

Website – www.theiscale.com **Contact : 7880-113-112** **theiscale** **theiscale.founders**

Statistics - One Shot Youtube

Calculating Mean ¶

method 1 - Manual Calculation

In [3]:
```
# Population data
population_data = [150, 160, 170, 165, 155]

# Sample data
sample_data = [150, 170, 165]

# Calculate population mean
population_mean = sum(population_data) / len(population_data)
print(f"Population Mean: {population_mean}")

# Calculate sample mean
sample_mean = sum(sample_data) / len(sample_data)
print(f"Sample Mean: {sample_mean}")
```

```
Population Mean: 160.0
Sample Mean: 161.66666666666666
```

Method 2 - Using numpy

In [5]:
```
import numpy as np

# Define the dataset
data = [85, 90, 78, 92, 88]

# Calculate the mean using numpy
arithmetic_mean = np.mean(data)

# Output the result
print(f"Arithmetic Mean using numpy: {arithmetic_mean}")
```

```
Arithmetic Mean using numpy: 86.6
```

Method 3 - Using statistics

In [12]:
```
import statistics as stats

# Sample dataset
data = [10, 20, 30, 40, 50]

# Calculate the mean using statistics module
mean_stats = stats.mean(data)

print(f"Mean using statistics module: {mean_stats}")
```

```
Mean using statistics module: 30
```

finding mean - data set

In [13]:
```
import seaborn as sns
import numpy as np

# Load the Iris dataset
iris = sns.load_dataset('iris')

# Display the first few rows of the dataset
print(iris.head())

# Calculate the mean of the 'sepal_length' column
mean_sepal_length = np.mean(iris['sepal_length'])

# Output the result
print(f"The mean of the 'sepal_length' column is: {mean_sepal_length}")
```

```
   sepal_length  sepal_width  petal_length  petal_width species
0           5.1          3.5           1.4          0.2  setosa
1           4.9          3.0           1.4          0.2  setosa
2           4.7          3.2           1.3          0.2  setosa
3           4.6          3.1           1.5          0.2  setosa
4           5.0          3.6           1.4          0.2  setosa
The mean of the 'sepal_length' column is: 5.843333333333334
```

Calculating median

Method 1 - Using numpy

In [14]:
```
import numpy as np

# Sample dataset
data = [20, 22, 24, 26, 28, 100]

# Calculate the median using numpy
median_np = np.median(data)

print(f"Median using numpy: {median_np}")
```

```
Median using numpy: 25.0
```

Method 2 - Using statistics module

In [15]:
```
import statistics as stats

# Sample dataset
data = [20, 22, 24, 26, 28, 100]

# Calculate the median using statistics module
median_stats = stats.median(data)

print(f"Median using statistics module: {median_stats}")
```

```
Median using statistics module: 25.0
```

Method 3 - Using pandas

In [16]:
```
import pandas as pd

# Create a sample DataFrame
df = pd.DataFrame({
    'values': [20, 22, 24, 26, 28, 100]
})

# Calculate the median using pandas
median_pandas = df['values'].median()

print(f"Median using pandas: {median_pandas}")
```

```
Median using pandas: 25.0
```

median using data set

In [17]:
```
import seaborn as sns
import numpy as np

# Load the Iris dataset
iris = sns.load_dataset('iris')
print(iris.head())

# Calculate the median of the 'sepal_length' column
median_sepal_length = np.median(iris['sepal_length'])

print(f"The median of the 'sepal_length' column is: {median_sepal_length}")
```

```
   sepal_length  sepal_width  petal_length  petal_width species
0           5.1          3.5           1.4          0.2  setosa
1           4.9          3.0           1.4          0.2  setosa
2           4.7          3.2           1.3          0.2  setosa
3           4.6          3.1           1.5          0.2  setosa
4           5.0          3.6           1.4          0.2  setosa
The median of the 'sepal_length' column is: 5.8
```

Calculating Mode

Using the statistics module

In [39]:
```
import statistics

data = [85, 90, 90, 95, 85, 85, 96, 80]

mode_result = statistics.mode(data)
print(f"Mode using statistics module: {mode_result}")
```

```
Mode using statistics module: 85
```

Using the scipy library

In [23]:
```
from scipy import stats

data = [85, 90, 90, 95, 85, 80, 85, 91, 80]

mode_result = stats.mode(data)
print(f"Mode using scipy library: {mode_result.mode[0]}")
```

```
Mode using scipy library: 85

C:\Users\lenovo\AppData\Local\Temp\ipykernel_5964\4025025990.py:5: FutureWarn
ing: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default
behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.
0, this behavior will change: the default value of `keepdims` will become Fal
se, the `axis` over which the statistic is taken will be eliminated, and the
value None will no longer be accepted. Set `keepdims` to True or False to avo
id this warning.
  mode_result = stats.mode(data)
```

In [37]:
```
from scipy import stats

data = [90,85, 92, 90, 95, 85, 80, 85, 90, 80]

mode_result = stats.mode(data)
print(f"Mode using scipy library: {mode_result.mode[0]}")
```

```
Mode using scipy library: 85

C:\Users\lenovo\AppData\Local\Temp\ipykernel_5964\3910792679.py:5: FutureWarn
ing: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default
behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.
0, this behavior will change: the default value of `keepdims` will become Fal
se, the `axis` over which the statistic is taken will be eliminated, and the
value None will no longer be accepted. Set `keepdims` to True or False to avo
id this warning.
  mode_result = stats.mode(data)
```

stats.mode returns the smallest mode when multiple modes exist.

Handling Multimodal Datasets

In [38]:
```
import statistics

data = [4, 5, 6, 6, 7, 7, 8, 8, 9]

modes = statistics.multimode(data)
print(f"The mode(s) of the dataset are: {modes}")
```

```
The mode(s) of the dataset are: [6, 7, 8]
```

calculating variance

Using Numpy

In [4]:
```
import numpy as np

data = [2, 4, 4, 4, 5, 5, 7, 9]

# Sample variance
sample_variance = np.var(data, ddof=1)
print("Sample Variance:", sample_variance)

# Population variance
population_variance = np.var(data, ddof=0)
print("Population Variance:", population_variance)
```

```
Sample Variance: 4.571428571428571
Population Variance: 4.0
```

Using Pandas

In [5]:
```
import pandas as pd

data = [2, 4, 4, 4, 5, 5, 7, 9]
series = pd.Series(data)

# Sample variance
sample_variance = series.var()
print("Sample Variance:", sample_variance)

# Population variance
population_variance = series.var(ddof=0)
print("Population Variance:", population_variance)
```

```
Sample Variance: 4.571428571428571
Population Variance: 4.0
```

Using SciPy

In [6]:
```
from scipy import stats

data = [2, 4, 4, 4, 5, 5, 7, 9]

# Sample variance
sample_variance = stats.tvar(data)
print("Sample Variance:", sample_variance)

# Population variance
population_variance = stats.tvar(data, ddof=0)
print("Population Variance:", population_variance)
```

```
Sample Variance: 4.571428571428571
Population Variance: 4.0
```

calculate standard deviation

Using NumPy

In [7]:
```
import numpy as np

data = [60, 65, 70, 75, 80]

# Sample standard deviation
sample_std_dev = np.std(data, ddof=1)
print("Sample Standard Deviation:", sample_std_dev)

# Population standard deviation
population_std_dev = np.std(data, ddof=0)
print("Population Standard Deviation:", population_std_dev)
```

```
Sample Standard Deviation: 7.905694150420948
Population Standard Deviation: 7.0710678118654755
```

Using Pandas

In [8]:
```
import pandas as pd

data = [60, 65, 70, 75, 80]
series = pd.Series(data)

# Sample standard deviation
sample_std_dev = series.std()
print("Sample Standard Deviation:", sample_std_dev)

# Population standard deviation
population_std_dev = series.std(ddof=0)
print("Population Standard Deviation:", population_std_dev)
```

```
Sample Standard Deviation: 7.905694150420948
Population Standard Deviation: 7.0710678118654755
```

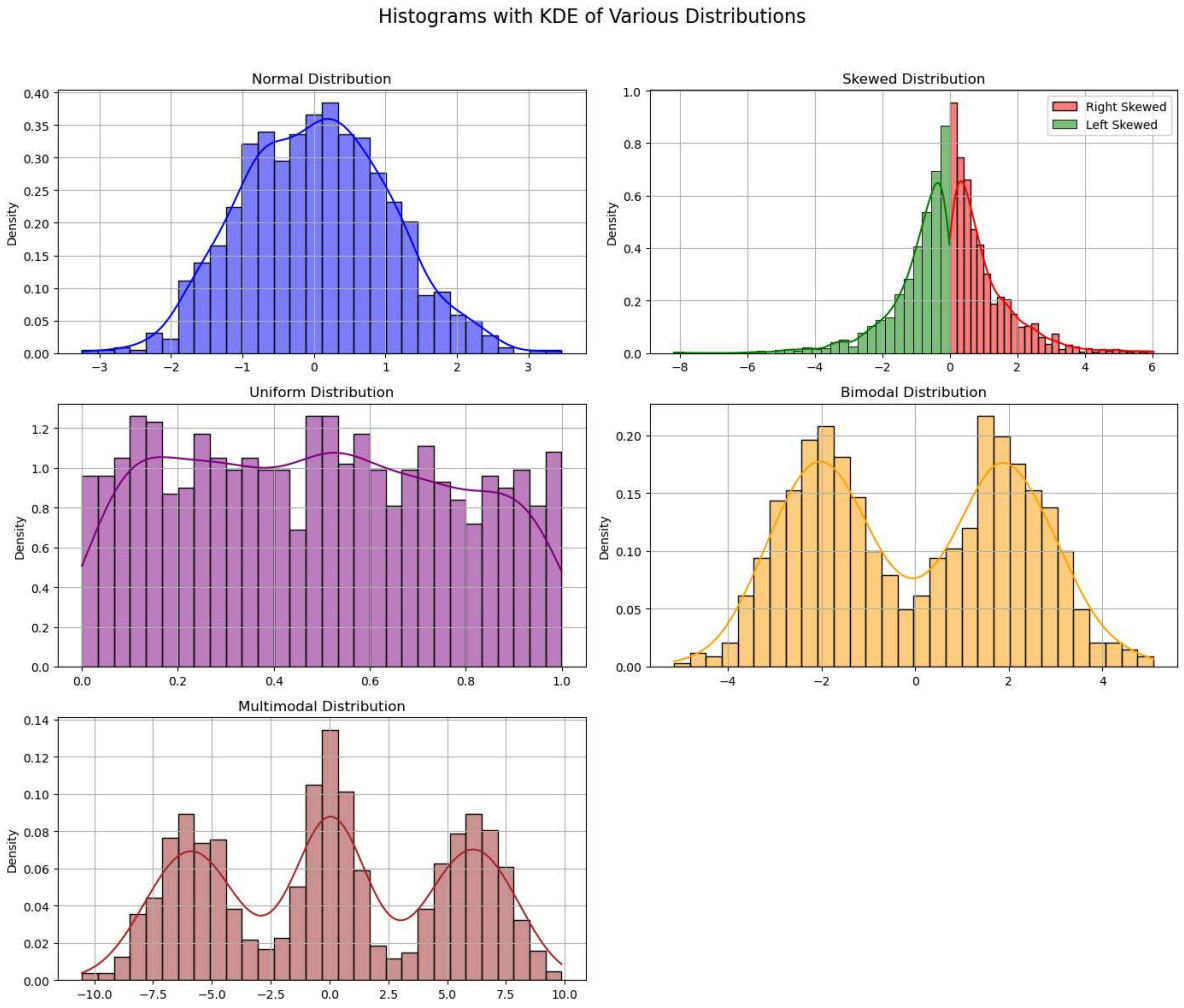
Using SciPy

In [9]:
```
from scipy import stats
import math

data = [60, 65, 70, 75, 80]

# Sample standard deviation
sample_std_dev = math.sqrt(stats.tvar(data))
print("Sample Standard Deviation:", sample_std_dev)

# Population standard deviation
population_std_dev = math.sqrt(stats.tvar(data, ddof=0))
print("Population Standard Deviation:", population_std_dev)
```

```
Sample Standard Deviation: 7.905694150420948
Population Standard Deviation: 7.0710678118654755
```

In [ ]:
```

```

In [14]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Create a figure and axes
fig, axs = plt.subplots(3, 2, figsize=(14, 12))
fig.suptitle('Histograms with KDE of Various Distributions', fontsize=16)

# Normal Distribution (Gaussian Distribution)
data_normal = np.random.normal(loc=0, scale=1, size=1000)  # Mean=0, Std Dev=1
sns.histplot(data_normal, kde=True, ax=axs[0, 0], color='blue', stat='density'
axs[0, 0].set_title('Normal Distribution')
axs[0, 0].grid(True)

# Skewed Distribution
data_right_skewed = np.random.exponential(scale=1, size=1000)  # Right skewed
data_left_skewed = -np.random.exponential(scale=1, size=1000)  # Left skewed

# Right skewed histogram
sns.histplot(data_right_skewed, kde=True, ax=axs[0, 1], color='red', stat='den
# Left skewed histogram
sns.histplot(data_left_skewed, kde=True, ax=axs[0, 1], color='green', stat='de
axs[0, 1].set_title('Skewed Distribution')
axs[0, 1].legend()
axs[0, 1].grid(True)

# Uniform Distribution
data_uniform = np.random.uniform(low=0, high=1, size=1000)  # Uniform distribu
sns.histplot(data_uniform, kde=True, ax=axs[1, 0], color='purple', stat='densi
axs[1, 0].set_title('Uniform Distribution')
axs[1, 0].grid(True)

# Bimodal Distribution
data_bimodal = np.concatenate([np.random.normal(loc=-2, scale=1, size=500),
                               np.random.normal(loc=2, scale=1, size=500)])
sns.histplot(data_bimodal, kde=True, ax=axs[1, 1], color='orange', stat='densi
axs[1, 1].set_title('Bimodal Distribution')
axs[1, 1].grid(True)

# Multimodal Distribution
data_multimodal = np.concatenate([
    np.random.normal(loc=-6, scale=1.5, size=500),
    np.random.normal(loc=0, scale=1, size=500),
    np.random.normal(loc=6, scale=1.5, size=500)
])
sns.histplot(data_multimodal, kde=True, ax=axs[2, 0], color='brown', stat='den
axs[2, 0].set_title('Multimodal Distribution')
axs[2, 0].grid(True)

# Hide empty subplot
axs[2, 1].axis('off')

# Adjust layout
plt.tight_layout(rect=[0, 0, 1, 0.96])
plt.show()
```


Histograms with KDE of Various Distributions
Normal Distribution
Skewed Distribution
Right Skewed
Left Skewed
Uniform Distribution
Bimodal Distribution
Multimodal Distribution
Density

p value test

In [1]:
```
import numpy as np
from scipy import stats

# Step 1: Create two random sample datasets
# Suppose these are test scores from two groups (e.g., group A and group B)
group_A = np.random.normal(75, 10, 30)  # Group A with mean 75, std 10, and 30
group_B = np.random.normal(80, 12, 30)  # Group B with mean 80, std 12, and 30

# Step 2: Perform an independent t-test
t_statistic, p_value = stats.ttest_ind(group_A, group_B)

# Step 3: Display the p-value and t-statistic
print(f"T-statistic: {t_statistic}")
print(f"p-value: {p_value}")

# Step 4: Interpretation based on p-value
alpha = 0.05  # Significance level
if p_value <= alpha:
    print("Reject the null hypothesis (Significant difference between groups)"
else:
    print("Fail to reject the null hypothesis (No significant difference betwe
```

```
T-statistic: -1.2171673720859824
p-value: 0.22846990920807717
Fail to reject the null hypothesis (No significant difference between groups)
```

In [5]:
```
import numpy as np
from scipy import stats

# Step 1: Create two random sample datasets
# Suppose these are test scores from two groups (e.g., group A and group B)
group_A = np.random.normal(75, 10, 30)  # Group A with mean 75, std 10, and 30
group_B = np.random.normal(80, 12, 30)  # Group B with mean 80, std 12, and 30

# Step 2: Perform an independent t-test
t_statistic, p_value = stats.ttest_ind(group_A, group_B)

# Step 3: Display the p-value and t-statistic
print(f"T-statistic: {t_statistic}")
print(f"p-value: {p_value}")

# Step 4: Interpretation based on p-value
alpha = 0.05  # Significance level
if p_value <= alpha:
    print("Reject the null hypothesis (Significant difference between groups)"
else:
    print("Fail to reject the null hypothesis (No significant difference betwe
```

```
T-statistic: -2.459033444858363
p-value: 0.016935924531047187
Reject the null hypothesis (Significant difference between groups)
```

Z-Test

Imagine a company claims that the average weight of their product is 500 grams. We want to test if this claim is true by collecting a sample of 30 products. The sample mean is 505 grams, and the population standard deviation is 10 grams. We will conduct a Z-test to see if the difference in the mean is significant or not.

In [6]:
```
import numpy as np
from scipy import stats

# Step 1: Define the known parameters
population_mean = 500  # Claimed population mean (mu)
sample_mean = 505  # Sample mean (X̄)
population_std = 10  # Population standard deviation (sigma)
sample_size = 30  # Sample size (n)

# Step 2: Calculate the Z-value
z_value = (sample_mean - population_mean) / (population_std / np.sqrt(sample_s
print(f"Z-value: {z_value}")

# Step 3: Calculate the p-value (two-tailed test)
p_value = 2 * (1 - stats.norm.cdf(abs(z_value)))
print(f"p-value: {p_value}")

# Step 4: Set significance level
alpha = 0.05

# Step 5: Compare p-value with significance level
if p_value < alpha:
    print("Reject the null hypothesis (significant difference)")
else:
    print("Fail to reject the null hypothesis (no significant difference)")
```

```
Z-value: 2.7386127875258306
p-value: 0.0061698993205441255
Reject the null hypothesis (significant difference)
```

In [ ]:
```
```

The iSCALE

Check out the

Premium Courses of

Data Science & Data Analyst

Contact -> 78801-13112

Top Companies for
Data Scientist & Analyst Jobs

Learn alongside goal-oriented individuals & build a strong network in top firms. This is your chance to grow with professionals who work at revered organizations. Take a look at some of the top companies:

Program Highlights

Immerse yourself in a curriculum that seamlessly integrates theoretical knowledge with hands-on projects, ensuring a comprehensive skill set in data analysis and data science.

12 Month Course Duration
LMS software access

Industry-Oriented Curriculum
By Industry Professionals

LIVE & Scheduled Classes
According to Features List

Placement Updates
Through exclusive job portal

Business Case Studies
To build practical industry skills

Certificate of Training
For course completion

Doubt Assistance
LIVE classes according to Features

Soft Skills Guidance
Essential for interviews

LMS Access
According to Features List

Hands-on Projects
Real-world industry projects

FORMERLY Industries Helping Hands

The iSCALE

Dedicated Career Support

- According to features list

Resume Building - An excellent resume has the power to open doors of success. We help you to build your resume to highlight your skills and your previous professional experience.

Mock Interviews - Mock Interviews help candidates to reduce their stress and help boost their confidence. You will also learn to crack interviews with our interview preparation sessions.

Portfolio - A portfolio is a snapshot of all the projects done and skills required during the program that is shareable across media channels. This will help you to showcase your expertise to potential recruiters.

Exclusive Job Portal – Unlock unparalleled career prospects with our exclusive job portal, facilitating seamless connections with industry-leading companies. Dive into a world of opportunities and fast-track your journey to professional success.

Linkedin Profile Building - Strong LinkedIn profile can contribute to career development. We help you in making a professional Linkedin Profile.

Program Outcomes

- According to features list

• Acquire proficiency in industry-standard tools such as Python, SQL, and data visualization libraries to manipulate and analyze diverse datasets effectively.

• Apply theoretical knowledge to practical scenarios through hands-on projects, building a robust portfolio that demonstrates your ability to tackle diverse challenges encountered in the field.

• Master the art of conveying insights through compelling data visualizations, using tools like Matplotlib, Seaborn, PowerBI and Tableau to communicate complex findings in a clear and impactful manner.

• Complete the program with a portfolio showcasing practical projects, a comprehensive grasp of data science methodologies, and the expertise needed to launch a successful career journey as a Data Scientist or Data Analyst.

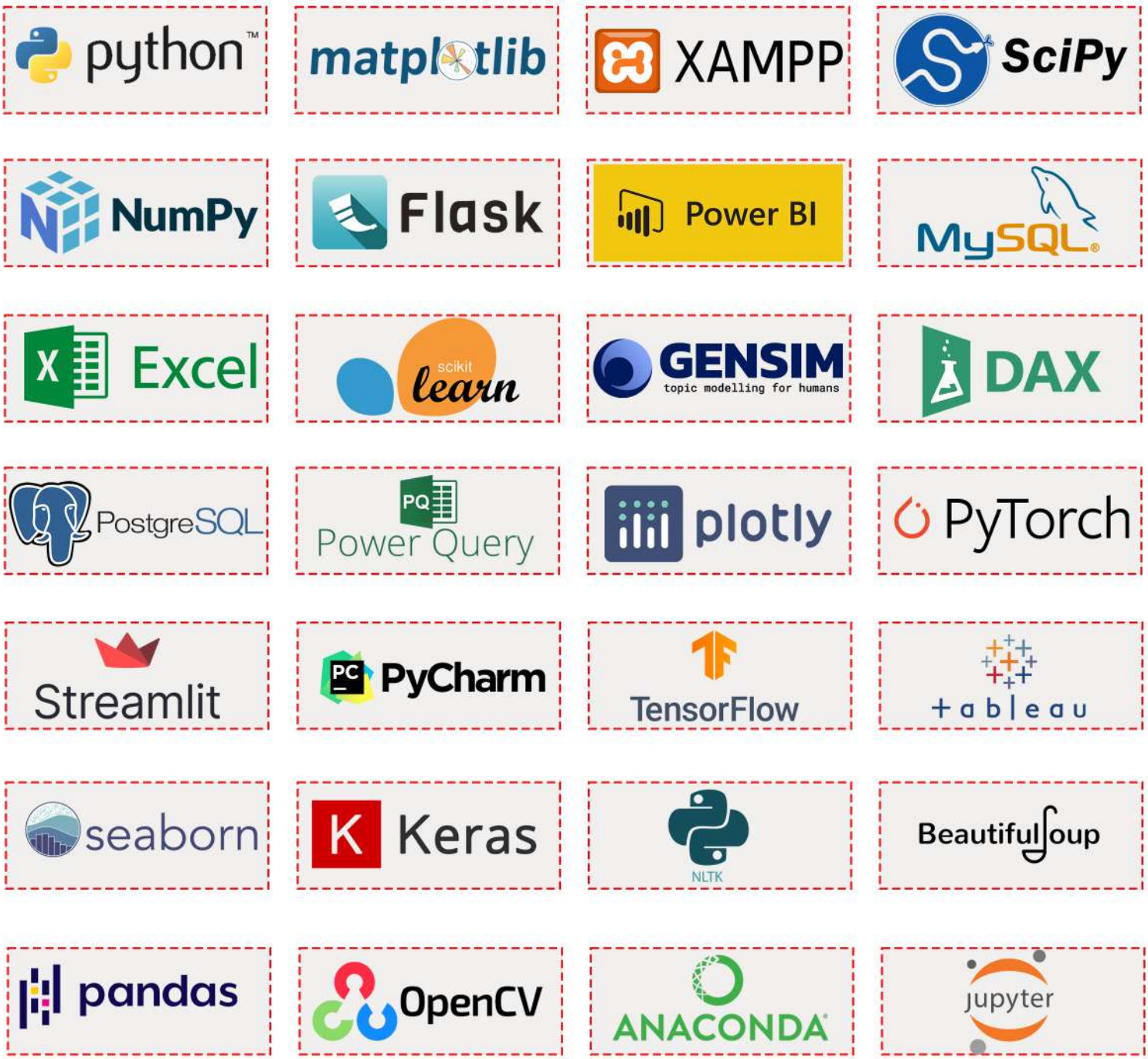
CERTIFICATE

Get the Certificate of Course Completion

Upon completion of this Course, you will receive a verified, industry - recognized certificate from our training platform.

Tools & Frameworks

Learners build real-world portfolio projects that help them master industry-standard tools, frameworks & libraries that are used daily in thousands of companies around the world.



FORMERLY Industries Helping Hands

The iSCALE

The Learning Path

Module 1
Fundamentals of Python

Module 2
Python Libraries

Module 3
Statistics

Module 4
Feature Engineering

Module 5
Exploratory Data Analysis

Module 6
Machine Learning -I
Supervised

Module 7
Machine Learning – II
Unsupervised

FORMERLY Industries Helping Hands

The iSCALE

Module 8
Deep Learning

Module 9
Natural Language Processing – NLP

Module 10
Computer Vision (Open CV)

Module 11
SQL

Module 12
Power BI

Module 13
Advanced Excel

Module 14
Tableau

Module 15
Placement Preparation Module

FORMERLY Industries Helping Hands

The iSCALE

Hands on Industry Oriented Projects

Google Image Scrapping with Flask

IMDB Web Scrapping with Flask

Movie Recommendation Project

Covid 19 Analysis Using Dash Library

Whatsapp Chat Analysis

Weather Prediction

Diabetes Prediction

Medical Prediction

Diamond Price Prediction

FORMERLY Industries Helping Hands

The iSCALE

Hands on Industry Oriented Projects

Credit Card Fault Detection

Object Racking

Image Classification

Image to Text

Indian Premier League

Shipping E-Commerce

Virat Kohli ODI T20 Performance Index

Netflix Stock Prediction

Moon Mission Analysis Project

FORMERLY Industries Helping Hands

The iSCALE

Academic Team

For helping you in your learning Journey

Nishant Dhote

- Founder & Serial Entrepreneur
- Industries Helping Hands
- Industry Exp – 7+ Years

Industries Helping Hands

Swati

- CBO – The iSCALE
- Ex Byjus – Product Analyst
- Social Projects (Ministry of Youth Affairs)

BYJU'S

Vibhor Sahu

-IIM - Raipur
-Associate Manager -Tredence Inc.
-Ex-Companies – TCS & Pixel

Tredence
Connect the Dots

Bhavesh

-IIT - Dhanbad
-Placement Head
-Industry Exp – 8+ Years

Bijay

- Berlin School of Business Germany
- Marketing Manager –Allmyhomes
- Ex-Companies Sygns, Internetwarriors, Felmo

ALLMYHOMES

Sweta Agrawal

- 5Years+ Work Experience
- Data Scientist
- Projects with Int. Clients – Zigtal, Vintage Wine Estates

BenchPrep

Connected CEOs & CHROs

Prasad Menon
CHRO

amagi Flipkart

Uday Narang
CEO

OSM OMEGA SEIKI MOBILITY

Harjeet Khanduja
Vice President

RELIANCE Jio

Rahil Gupta
CTO

hop

Jaibir Siwach
CEO

Kabira Mobility

Jainendera Anand
CEO

REVOLT

Ashwani Jaiswal
Vice President

OKINAWA Power the Change

Pushkaraj Salunke
CTO

REVAMP MOTO

Dheeraj Shetty
CHRO

Ultraviolette

Srikanth Reddy
CEO

hala!

Javed Khatri
CEO

clanx

Hyder Khan
CEO

eblu

FORMERLY Industries Helping Hands

The iSCALE

Students Testimonials

The Iscale (Formerly Industries Helping Hands) assisted me in securing a placement within three months. The guidance and support from educators were incredible during my journey to become a data analyst.

Manas Jyoti Bohra ★★★★★

My learning experience has been great because, despite coming from a non-IT background, educators helped me understand the concepts and secure a Data Analyst job.

Shubham ★★★★★

Real-time, industry-oriented end-to-end projects are important and beneficial for me to mention on my resume.

Riya Garg ★★★★★

As a working professional, I easily manage my studies and job because the lectures are easy to understand, and the doubt assistance is really helpful for me

Suryakant ★★★★½

Transitioning from a non-IT BBA background to learning data analytics here is amazing because mock interviews and project sessions are very beneficial.

Anwar Khan ★★★★★

My learning experience is going well because I am gaining confidence through the interview preparation sessions, and the teaching method is incredible.

Abhinandan ★★★★★

I can learn anytime and anywhere because the lectures and study materials are easily accessible through the mobile app and web.

Vipin Verma ★★★★★

Here, educators skillfully integrate technical concepts with real-time, industry-oriented projects, enhancing our understanding through practical applications.

Bhoopendra ★★★★½

Talk With Community

We have conducted the Interview Experience Talk of the Achievers. Focused, hard work is the real key to success. Keep your eyes on the goal, and just keep taking the next step towards completing it.

TIGER ANALYTICS

Ankush

Cognizant

Sharmine

CATERPILLAR

Siya

accenture

Manish

Deloitte.

Shreya

Itron

Kishan

accenture strategy

Aman

ORACLE

Ashish

PayPal

Vrishank

MAERSK

Venus

IBM

Jahanvi

FiftyFive TECHNOLOGIES

Vishal

CISCO

Anand

OLA

Jashan

HSBC

Ayushi

SWIGGY

Roshan

Tech Mahindra

Anindya

Microsoft

Vaibhav

Disney+ hotstar

Ankush

Walmart

Save money. Live better.

Nishant

amazon

Lokesh

amazon

Harshit

TCS TATA CONSULTANCY SERVICES

Pushpanjali

There are approximately 400,000 bytes of data for every grain of sand on earth. For managing this huge data, Data Scientist & Analysts' demands are growing globally.

FORMERLY Industries Helping Hands

The iSCALE

CURRICULUM

Module 1
Fundamentals of Python

UNIT 1: The Basics of Python Programming

- Understanding Problem Solving
- Why Python
- What is an IDE (Integrated Development Environment)
- Installation of Python – Anaconda & Launching of Jupiter Notebook
- Write your first Python Program

UNIT 2: Variables and Operators in Python

- Python Arithmetic Operators
- Python Assignment Operators
- Python Boolean & Logical Operators
- Python Comparison Operators
- Python Bitwise Operator

UNIT 3: Built-in Functions of Python

- Round Function & Divmod Function
- Isinstance Function & Pow Function,
- Input Function , int function, Float Function
- Len, Min and Max function
- Sorted Function and help function

UNIT 4: Control Flow Statements

- If Condition
- If else Condition
- If-Elif-else Condition
- Nested if

UNIT 5: Concepts of Loops

- While Loop
- Break & Continue Statements
- For Loop & Range Function
- Nested Loops in Python

UNIT 6: User-Defined Functions in Python

- Introduction & Doc String
- Functions Input Arguments
- Functions Order of input Arguments
- Functions Return Statement
- Function Variable Number of Input Arguments
- *args and **kwargs in Function
- Lambda function

UNIT 7: Python Strings

- Introduction to String
- Indexing & Slicing
- Methods – strip, lower, upper
- Methods – replace, split, capitalize
- Methods – title, swapcase, find, index

- Methods – isalnum, isalpha, isdigit
- Methods – isupper, islower, istitle, startswith, endswith
- Multiline String & checking of substring in String
- Converting String into another Data Type
- Loops in string

UNIT 8: Python List

- Introduction to List
- Slicing on List
- Indexing on List
- Methods - Append, Extend & Insert in List
- Methods - Remove, Pop, Clear
- Methods – Count, Max,Min
- Reverse, Index Count, Copy
- Using Loops in List
- Using conditional statement in List

UNIT 9: Python Tuples

- Introduction, Concatenation & Nesting
- length, count, index method, indexing & slicing
- The tuple() Constructor
- Duplicates in tuples
- max, min, unpacking of tuples
- conditional statement and loops in tuple

UNIT 10: Python Sets

- Introduction to Set

- Methods - add, update, remove, discard, pop, clear, del, union
- Methods - intersection_update, intersection, difference_ update, difference
- Methods - symmetric_difference_update, symmetric_ difference
- Methods - isdisjoint, issuperset, issubset
- Methods - set(), max, min, copy, loops in sets

UNIT 11: Python Dictionary

- Introduction to Dictionary
- Methods-, len, dict constructor fromkeys, setdefault
- Methods - get, keys, values, items, update
- Methods - pop, popitem, del, clear, copy
- Methods - Nested Dictionaries, loops in dictionary

UNIT 12: Object Oriented Programming

- OOPS Basic concept & creating classes
- polymorphism and encapsulation
- Types of Inheritance - single, multilevel, multiple, hierarchical
- abstraction & decorator
- Types of methods - Class method, static method
- Special-magic/dunder method
- property decorators – getters, setters and deletes

UNIT 13: Files Handling

- Working with Files

- Reading & writing files
- Buffered Read and Write
- Other File Methods
- Logging & Debugger

UNIT 14: Exception Handling

- Introduction
- Exception handling with try, except, else, and finally block of code.
- Custom Exception Handling
- List of general use exceptions
- Best Practice Exception Handling

Module 2
Python Libraries

UNIT 1: NumPy

- Introduction to Numpy
- Numpy - Nd Array Object.
- Numpy - Data Types.
- Numpy - Array Attributes.
- Numpy - Array Creation Routines.
- Numpy - Array From Existing.
- Data Array From Numerical Ranges.
- Numpy - Indexing & Slicing.
- Numpy – Advanced Indexing.
- Numpy – Broadcasting.
- Numpy - Iterating Over Array.
- Numpy - Array Manipulation.
- Numpy - Binary Operators.
- Numpy - String Functions.
- Numpy - Mathematical Functions.
- Numpy - Arithmetic Operations.
- Numpy - Statistical Functions.
- Sort, Search & Counting Functions.
- Numpy - Byte Swapping.
- Numpy - Copies &Views.
- Numpy - Matrix Library.
- Numpy - Linear Algebra.

UNIT 2: Pandas

- Introduction to Pandas
- Reading Data From Different File Systems
- Python Pandas – Data Frame
- Python Pandas - operations
- Python Pandas - Basic Functionality
- Reading Data From Different File Systems
- Python Pandas – Re-Indexing Python
- Pandas – Iteration
- Python Pandas – Sorting.
- Working With Text Data Options & Customization
- Indexing & Selecting
- Data Statistical Functions
- Python Pandas - Window Functions
- Python Pandas - Date Functionality
- Python Pandas –Time Delta
- Python Pandas - Categorical Data
- Python Pandas – Visualization
- Python Pandas – Tools

UNIT 3: Matplotlib

- Introduction to Matplotlib
- Bar chart
- Scatter Plot
- Bar Graph
- Histogram
- Pie Chart
- Stem Plots
- Box Vs Whisker Plot
- Area Vs Stack Plot
- Step Plot
- Fill_Between Plotots

- subplot
- Savefig
- Axis matplotlib plots
- Text in matplotlib plots

UNIT 4: Seaborn

- Introduction to Seaborn
- Heatmap plot method
- Count plot method
- Violin Plot method
- Pair plot method
- Strip Plot method
- Box plot method
- Factor plot method
- Cat Plot method
- Styling plot method
- Facet Grid Method
- Line Plot method
- Histogram plot method
- Bar plot method
- Scatter plot method

UNIT 5: Plotly

- Introduction to Plotly
- Bubble Charts
- Bar charts
- Pie Charts
- Creating maps
- Time Series

Python Projects

Project 1: Google Image Scrapping - Scraping Google Images, involves extracting data from website the project involved sending requests to the Google Images website, parsing the HTML content, and extracting relevant image URLs using BeautifulSoup. This project showcases technical skills in Python programming and data analytics.

Project 2: Covid-19 Impacts Analysis - Covid-19 impacted the global economy. This project is to analyze the spread of Covid-19 cases and all the impacts of covid-19 on the economy.

Project 3: Movie-Recommendation System Project -
Recommendation systems are among the most popular applications of data science. This project is about the recommendation of movies.

Project 4: WhatsApp Chats Analysis - WhatsApp is one of the most used messenger applications today so we can use WhatsApp chats for analyzing our chat with a friend, customer, or a group of people.

Project 5: Image Scrapping with Flask - Image Scraping with Flask is a Python data science project that utilizes the Flask framework. It involves extracting and collecting images from various sources, demonstrating the integration of web scraping capabilities within a Flask web application for effective data retrieval and analysis.

Module 3
Statistics

UNIT 1:

- Introduction to Statistics
- Types of Statistics
- Types of Data
- Levels of Measurement

UNIT 2:

- Measures of Central Tendency
- Measures of Central Tendency in Python
- Measures of Dispersion
- Measures of Dispersion in Python
- Random Variables
- Sets
- Histogram Skewness
- Covariance Correlation
- Covariance Correlation in Python

UNIT 3:

- Probability Density Distribution Function
- PDF, PMF CDF
- Types of Probability Distribution
- Bernoulli Distribution

- Binomial Distribution
- Poisson Distribution
- Normal or Gaussian Distribution

UNIT 4:

- Uniform Distribution
- Z Stats Z Table
- Central Limit Theorem

UNIT 5:

- Estimate
- Hypothesis Testing Mechanism
- P value
- Z Test with Examples
- Student T Distribution
- T Stats T Test
- When to use T test vs Z test

UNIT 6:

- Type 1 Type 2 Error
- Confidence Interval Margin of Error
- Bayes Theorem

UNIT 7:

- What is Chi Square Test
- Chi Square for Goodness of Fit
- Chi Square Test with Python

UNIT 8:

- F Distribution
- Variance Ratio Test (F test)
- F test with Python

UNIT 9:

- Anova
- Assumptions in Anova
- Types of Anova
- Partition of variance in Anova

Module 4
Feature Engineering

UNIT 1:

- AI vs ML vs DL vs DS
- Supervised, Unsupervised Reinforcement Learning
- Train, Test Validation
- Variance, Bias, Overfitting Underfitting

UNIT 2:

- Handling Missing Values
- Handling Imbalanced Dataset
- SMOTE
- Data Interpolation
- Handling Outliers

UNIT 3:

- Feature Extraction
- Feature Scaling Normalization
- Normalization Min Max Scaling
- Unit Vectors Feature Scaling
- PCA

UNIT 4:

- Data Encoding
- Nominal or One Hot Encoding

UNIT 5:

- Label Ordinal Encoder
- Target Guided Ordinal Encoding

Module 5
Exploratory Data Analysis

UNIT 1:

- **EDA With Red Wine Data**

Exploratory Data Analysis (EDA) With Red Wine Data" is focusing on the analysis of a dataset related to red wine. Through statistical and visual techniques, this aims to uncover patterns, trends, and insights within the data.

- **EDA Student Performance Indicator**

EDA Student Performance Indicator involves conducting Exploratory Data Analysis (EDA) on a dataset related to student performance.

- **EDA Forest Fires**

EDA Forest Fires refers to an Exploratory Data Analysis project focusing on the examination of a dataset related to forest fires.

UNIT 2:

- **EDA Flight Price**

EDA Flight Price focuses on Exploratory Data Analysis for a dataset related to flight prices.

- **EDA Forest Fires**

EDA Google Playstore" involves conducting Exploratory Data Analysis on a dataset related to the Google Play Store.

Module 6
Machine Learning – I Supervised

Part 1 – Regression

UNIT 1:

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Linear Regression

UNIT 2:

- R squared Adjusted R squared
- MSE,MAE RMSE
- Simple Linear Regression With Python
- Multiple Linear Regression

UNIT 3:

- Ridge Regression
- Lasso Regression

UNIT 4:

- Elastic Net Regression
- EDA with Algerian Forest Fire
- Model Training for Ridge, Lasso Elastic Net

Project 1: End to End ML Project :

Weather Prediction - Scraping Google Images, involves extracting data from website the project involved sending requests to the Google Images website, parsing the HTML content, and extracting relevant image URLs using BeautifulSoup. This project showcases

PART 2: Logistics Regression

- Logistic Regression Indepth Intuition
- Logistic Regression with Regularization
- Performance Metrics Confusion Matrix, Accuracy, Precision Recall
- Cross Validation Types
- Hyperparameter Tuning
- Logistic Regression Implementation
- Logistic Regression Multiclass Classification

Project 2: End to End ML Project :

Diabetes Prediction - Diabetes Prediction is a machine learning project for data science designed to forecast the likelihood of diabetes based on relevant health data. Using predictive modeling, the project aims to assist in early detection and proactive management of diabetes risk factors.technical skills in Python programming and data analytics.

PART 3: Decision Tree

- Decision Tree Classifier Intuition
- Decision Tree For Numerical Split

- Post Pruning And Prepruning Decision Trees
- Decision Tree Classifier Implementation
- Decision Tree Post Pruning
- Decision Tree Regressor In depth Intuition
- Decision Tree Regressor Implementation

PART 4: Support Vector Machines

- Support Vector Classifier Indepth Intuition
- Support Vector Machines Classifier
- Support Vector Regressor Indepth Intuition
- Support Vector Regressor Implementation
- SVM Kernels Intuition
- SVM Kernels Implementation

PART 5: Naive Bayes

- Naive Bayes Indepth Intuition
- Variants Of Naive Bayes Algorithms
- Naive Bayes Practical Implementation

Project 3: End to End ML Project :

Medical Prediction - • Medical Prediction employs machine learning for data science to forecast health outcomes based on relevant medical data. This project aims to enhance predictive analytics in the medical field for improved diagnosis and patient care technical skills in Python programming and data analytics.

PART 6: Ensemble Techniques & its types

- Ensemble Techniques And Bagging
- Random Forest Classifier And Regressor
- Out Of Bag Score Decision Trees
- Random Forest Practical Implementation

PART 7: Boosting

- Boosting Technique
- Gradient Bosting Indepth Intuition
- Xgboost Classification Algorithms
- Xgboost Regresor Algorithm

PART 8: KNN Algorithm

- KNN Classification And Regression
- Variants Of KNN
- KNN Classifier and Regr Implementation

Module 7

Machine Learning – II Unsupervised

PART 1: PCA

- Curse Of Dimensionality
- Geometric Intuition Behind PCA
- MAthemathical Intuition Of PCA
- PCA Practical Implementation

PART 2: Clustering Algorithms

- K Means Indepth Intuition
- Hierarichal Clustering Intuition
- K means vs Hierarichal Clustering
- DBSCAN Clustering
- Silhouette Score Clustering
- K Means Clustering Implementation
- Hierarichal Clustering Implementation
- DBSCAN Algorithms Implementation

UNIT 3: Anomaly Detection

- Anomaly Detection Isolation Fores
- DBSCAN Anomaly Detection
- Local Outlier Factor Anamoly Detection

PART 4: Time Series

- Introduction of time series
- components of time series
- moving average
- stationary and non stationary
- acf pacf
- arima
- time sereis model building
- Time sereis EDA

Project 1: Diamond Price Prediction - Diamond Price Prediction is a machine learning project for data science that predicts diamond prices based on various features. Using algorithms, it aims to provide accurate and data-driven insights for pricing within the diamond industry.

Module 8
Deep Learning

PART 1: Neural Network & Its Perception

- Intro to Deep Learning Usecases
- Neural Network, Perceptron Mathematical Explanation

PART 2: Neural Network

- Mathematical Concepts
- Activation Functions
- Forward Back Propagation
- Implementation of ANN using Keras
- Callback Functions

UNIT 3: ANN

- Regression using ANN
- Loss Function
- Batch Normalisation
- Regularization

UNIT 4: Tensor Flow & Pytorch

- Regularisation in Deep Learning
- Weight Initialisation
- Optimizers

- Tensorflow
- Pytorch

Project 1: Credit card Fault detection : Credit Card Fault Detection is a deep learning project for data science focused on identifying fraudulent transactions within credit card data. Leveraging neural networks, the project aims to enhance the accuracy of fraud detection algorithms for secure financial transactions

PART 5: CNN

- Convolutional Neural Networks
- CNN Foundation 1
- CNN 2
- CNN 3 (Explainer)
- CNN 4 (LENET)
- CNN 5 (Alex Net)
- CNN 6
- CNN 7
- VGG NET
- Resnet
- Inception Net
- RCNN
- Fast RCNN
- Faster RCNN (Object Detection)
- Non Maximum Suppression

PART 6: YOLO

- YOLO
- YOLO V2
- YOLO V3
- YOLO V4
- YOLO V4 Part 2
- YOLO V5

PART 7: GAN

- GAN Introduction
- Training using GAN
- DC GAN
- Style GAN
- W Gan
- GAN Practical

Module 9
NLP

PART 1:

- Introduction to NLP
- History of NLP
- Why NLP
- Use of NLP
- web scrapping
- Text processing
- Understanding regex

PART 2:

- String Tokenization
- Sentence Processing
- Word Embedding
- Lemmatization in text processing
- Frequency Distribution
- Annotator creation
- Word Count
- Text Normalization
- Word 2 Vec
- Co Occurance Vector

PART 3: NLP

- Doc 2 Vec
- Text blob
- NLTK
- Genism
- RNN
- LSTM
- Bi LSTM

Module 10
Computer Vision (Open CV)

Computer Vision

- **Computer Vision** in data science focuses on enabling machines to interpret and make decisions based on visual data. It involves developing algorithms and models to analyze images or videos, extracting meaningful information for applications like image segmentation, object detection, and scene recognition.

- **Detection of Objects** detection in computer vision is the process of identifying and locating specific objects within images or videos. Using advanced algorithms, this technology enables machines to recognize and draw bounding boxes around objects.

- **Scene Recognition** in computer vision involves teaching machines to understand and categorize the overall context or setting within an image. Through sophisticated algorithms, this technology enables systems to identify and classify scenes.

- **Image Segmentation** in computer vision is the process of dividing an image into meaningful segments or regions. This technique, facilitated by advanced algorithms, allows for precise identification and extraction of objects within an image.

Project 1: Object racking - Object Tracking in computer vision for data science involves developing algorithms to follow and analyze the movement of objects. This project aims to enhance real-time tracking capabilities for applications such as surveillance and autonomous systems.

Project 2: Image Classification - Image Classification" in computer vision for data science focuses on developing algorithms to categorize and label images based on their content. This project aims to improve automated image recognition systems for various applications.

Project 3: image to text - Image to Text in computer vision for data science involves converting visual content into text using machine learning algorithms. This project aims to enhance accessibility and information retrieval by enabling automated extraction of textual information from images

Module 11
SQL

UNIT 1: Introduction and Installation Software

- Roadmap to learn SQL & syllabus discussion
- Download & Install MySQL
- Download & Install XAMPP
- Download & Install PostgreSQL & Pgadmin

UNIT 2: Fundamentals of SQL

- Introduction to SQL
- Database Fundamental
- Learning Resources - (Cheat Code |Practice Dataset| PDF Notes & Books |Learning Website | Interview Questions)

UNIT 3: Case Study with Example

- Create Database | Drop Database | Create Table | Insert Table | Drop Table
- Case study with XAMPP Software
- Case study with MySQL software
- Case study with PostgreSQL

UNIT 4: Data Types and Keys

- All Data Type in SQLs)
- ER / Database Diagram Example: Amazon Ecommerce
- Database Keys
- Cardinality of Relationships

UNIT 5: DDL Commands for SQL

- DDL Commands for Databases: Create | Drop
- DDL Commands for Table: Create | Truncate | Drop
- Data Indignity

UNIT 6: Constraints

- Constraints in MySQL (NOT NULL | UNIQUE | PRIMARY KEY)
- Constraints in MySQL (AUTO INCRIMENT | CHECK | DEFAULT | FOREIGN KEY)
- Referential Actions (Restrict | CASCADE | SET NULL | SET DEFAULT

UNIT 7: Alter Table

- ADD Columns in Alter Table
- DELETE Columns in Alter Table
- MODIFY Columns in Alter Table

UNIT 8: ADD, Editing and Deleting Constraints

- ADD Constraints
- DELETE Constraints
- EDIT Problems in MySQL, Drop Constrains

UNIT 9: INSERT Query

- INSERT query
- INSERT query variation
- INSERT multiple values

UNIT 10: SELECT Query

- SELECT all columns
- Filter columns
- Alias | renaming columns
- Create expression using columns
- Constant value
- DISTINCT(unique) values from a column
- DISTINCT Combinations
- Filter rows WHERE claus
- BETWEEN operator
- IN and NOT IN Operator

UNIT 11: UPDATE Query

- Update Query to update row(s)
- Update Multiple Columns

UNIT 12: DELETE

- DELETE Query to delete row(s)
- Deletion based on multiple conditions

UNIT 13: Aggregate Functions

- MAX() & MIN()
- AVG()
- SUM()

UNIT 14: Scalar Functions

- ABS()
- ROUND()
- CEIL() & FLOOR()

UNIT 15: SQL Joins

- What are SQL joins?
- Types of Joins
- Cross Joins --- Cartesian Products
- Inner Joins
- Right Joins
- Left Joins
- Full Outer Join
- Set Operations
- Self-Join
- Filtering Columns
- Filtering Rows

UNIT 16: Subquery

- What is Subquery?
- Scalar Subquery
- Row Subquery
- Table Subquery
- Correlated Subquery

UNIT 17: Sorting

- Fundamental of SQL Sorting
- Sorting top 10 data
- Sorting Descending Orders
- Sorting functions in datasets

UNIT 18: Grouping

- Fundamental of SQL Grouping
- Grouping Average Function
- Group By on multiple columns
- Sorting Descending Orders
- Sorting functions in datasets

UNIT 19: Project & Case Study:

Project 1: Indian Premier League (IPL Dataset)

- Case Study in Indian Premier League
- Find the top 5 batsmen in IPL
- Find the 2nd highest 6 hitters in IPL
- Find Batsman (M.S Dhoni, Virat Kohli) performance against all IPL team
- Find the top 10 batsmen with centuries in IPL using Sub-Quires
- Find the top 5 batsmen with the highest strike rate

UNIT 20: Project & Case Study:

Project 2: Shipping E-Commerce (Sorting: Smart Phone Datasets)

- Find the top 5 Samsung phones with the biggest screen size
- Sort all the phones in descending order of the number of total cameras
- Sort data on the basis of ppi in decreasing order
- Find the phone with 2nd largest battery
- Find the name and rating of the worst rated apple phone
- Sort phones alphabetically and then on the basis of rating in desc order
- Sort phones alphabetically and then on the basis of price in ascending order

Module 12
Power BI

UNIT 1: Introduction Power BI

- How to Download, Install and upgrade features in Power BI
- User Interface and Navigation Features in Power BI
- Import First Sample Data in Power BI

UNIT 2: Charts

- Introduction to Basic Charts in Power BI
- Create a Column Chart in Power BI
- Create a Stacked Column Chart in Power BI
- Create a Pie Chart in Power BI
- Create a Donut Chart in Power BI
- Create a Funnel Chart in Power BI
- Create a Ribbon Chart in Power BI
- What is included & Exclude in Power BI
- View Data and Export in CSV from Power BI Visuals
- Create a Ribbon Chart in Power BI
- What is included & Exclude in Power BI
- View Data and Export in CSV from Power BI Visuals

UNIT 3: Tables & Matrix

- Introduction Tables & Matrix in Power BI

- Creating a Table in Power BI
- How to Format a Table in Power BI
- How to Apply Conditional Formatting in Power BI
- How to Change Aggregations in Power BI
- Create a Matrix in Power BI
- Apply Conditional Formatting in Matrix in Power BI
- Create Hierarchies in Power BI Matrix
- Add Totals & Sub Totals in Matrix in Power BI
- How to Change Number Formatting in Power BI

UNIT 4: Slicers

- Create Slicers in Power BI
- Create Text Slicers in Power BI
- How to format Text Slicers in Power BI
- Create Date Slicers in Power BI
- How to format a Date Slicer Power BI
- Create a Number of Slicers in Power BI

UNIT 5: Charts Visualizations Tools

- Introduction to Other Charts in Power BI
- Create a Waterfall Chart in Power BI
- Create a Gauge Chart in Power BI
- Create Line Chart in Power BI
- How to Drill Down in Line Chart in Power BI
- Create an Area Chart in Power BI
- Create a Stacked Area Chart in Power BI

- Create a Line vs Stacked Column Chart in Power BI
- Create a Line vs Cluster Column Chart in Power BI
- Create Scatter Plot in Power BI
- Create Tree Map in Power BI

UNIT 6: Map

- Introduction Map Visualization
- Create a Basic Map in Power BI
- Creating a Filled Map in Power BI
- Creating a Map with Pie Chart in Power BI
- Use Slicer with Map of India
- Learn Formatting of Map in Power BI
- Change the Background of Maps in Power BI
- Create a Map of the International Continent Vs Donut in Power BI

UNIT 7: Cards & Filter

- What is Cards & Filters in Power BI
- Create a Number Card in Power BI
- Create a Date Card in Power BI
- Create a Relative Date filter in Power BI
- Create a Text Card in Power BI
- How to Format a Card in Power BI
- Create a Multi - Row Card in Power BI
- Create a Filter on Visual in Power BI
- Create a Filter On This Page in Power BI

- How to Use Filter on All Pages in Power BI
- How to Use Drill through in Power BI

UNIT 8: Insert & Action Functions

- Introduction Insert & Action
- How to Insert Image in Power BI
- Create Insert Text in Power BI
- How to Insert Shapes in Power BI
- Create Insert Buttons in Power BI
- Use Web URL Action in Power BI
- Work on Page Navigation Action in Power BI
- Use Bookmark Action in Power BI
- How to create Drill through Action in Power BI
- Create Power BI Account

UNIT 9: Advanced Charts

- Introduction to Advanced Charts in Power BI Desktop
- Create Scroller in Power BI
- Create a Word Cloud in Power BI
- Create Infographics in Power BI
- Create Drill-Down Donut Chart in Power BI
- Create Drill-Waterfall Chart in Power BI
- Create Play Axis Slicer in Power BI
- Create an Animated Bar Chart Race in Power BI
- Create Sunburst Chart in Power BI
- Create Sankey Chart in Power BI
- Create Calendar & Timeline Slicers

UNIT 10: KPI & Other Functions

- Introduction to Advance Function
- How to use KPI Visualization Tool
- Learn Updates in MS Power BI
- Create a Key influencer in Power BI
- Create a Decomposition Tree in Power BI
- Create a Python & R Scripting in Power BI
- Third-Party Extensions 3 Icons Introduction in Power BI

UNIT 11: Create a Superstore Report

- Create reports and dashboards documentation.
- Collaborate, share and integrate across products documentation.
- Get Power BI samples.
- Use report themes.
- Add conditional table formatting.
- Ways to share your work.
- Organize work in the new workspaces.
- Publish to web.

UNIT 12: Dashboard Functions

- How to Publish Report to Power BI Service
- Export Power BI Report into PPT, PDF & PBIX format
- Create a Pin Live Dashboard in Power BI
- What is Comment, Share & Subscribe in Power BI Reports
- How to Refresh Reports Automatically in Power BI Service
- Free Data Files Source for Practice

Power BI Projects

Project 1: Project Title: Virat Kohli ODI T20 Performance Index:

Description: Solve a problem statement with the help of Virat Kohli ODI T20 Performance Index Power Bi Dashboard and discuss multiple case studies.

Project 2: Project Title: Netflix Stock Prediction & runtime by genres

Description: Solve a problem statement with the help of Netflix Stock Prediction Datasets & analysis the runtime by genres in Power Bi Dashboard and discuss multiple case studies.

Project 3: Project Title: Moon Mission Analysis Project

Description: Solve a problem statement with the help of moon mission datasets & analyse the multiple functions in Power Bi Dashboard and discuss multiple case studies.

Module 13
Advance Excel

UNIT 1: Fundamental of Excel: Beginner Level

- Basic Features - Introduction
- Basic Formatting
- Using Formulas
- Save File
- Filter & Sorting
- Conditional Formatting
- Insert & Delete Columns or Rows
- Find & Remove Duplicates
- Merge and Centre

UNIT 2: Intermediate Level of Excel

- Rounding of Numbers
- Autofill in Excel
- Add or Edit Comment
- Filters for Data Manipulation
- Sorting on multiple columns
- Insert Table
- Slicers

UNIT 3: Moderate Level of Excel

- Creating an Excel Column Chart
- Working with Excel Pie Charts
- Chart Group (Line, Area, Waterfall)
- Insert Picture or Shapes
- Add link or hyperlink
- Excel Sparkline's
- Add Drop Down List
- Split Text
- Page Layout Tab and Ribbon
- Printing Excel Worksheet

UNIT 4: Advance Level of Excel

- Formula Tabs & Library
- Define Names
- Formula Auditing
- Calculation Options & Watch Window
- Get and Transform Data
- Queries and Connections
- Excel data validation
- Excel PivotTables
- Excel conditional functions
- Excel Text based Functions
- Excel "What if?" Tools
- Conditional IF Function IN [EXCEL]
- Result Format with Nested-IF [Excel]
- XOR Formula (Exclusive OR) EXCEL

- Excel VLOOKUP Function
- Advance VLOOKUP Formula

UNIT 5: Data Processing and Create Visualization Dashboard

Project Title: Analysis the Online Store Annual Report Like Mantra | AJio | Amazon | Etc.

Description: Importing data from text files Create a fully functional online store dashboard with the analysis the store annual reports and learn data cleaning, processing, analysis and dashboard creation in excel.

Module 14
Tableau

Tableau is a business intelligence tool which is widely used in the analytics industry. It is one of the top data visualization tools used for data analysis and getting data driven business insights. Learn various techniques to extract meaning from large datasets, detailed approaches to design dashboards. Learn to create storyboards using Tableau which enables us to transform complex data into meaningful insights.

UNIT 1: Creating Visualizations

- Text Table and Formatting Visuals
- Heatmap and Highlight table
- Lines Area Chart
- Circle, Side by Side Circle View and Shapes
- WordCloud
- Scatter Plot
- Histogram Bins
- Gantt view and Waterfall Chart
- Refrence lines and bands, Distribution band Bullet graphs
- Dual Axes Visuals Part 1
- Dual Axes Part 2

UNIT 2: Filtering, Sorting & Connecting

- Animated Charts

- Funnel Charts its types
- Donut Chart Multiple Donut Chart
- Dumbbell, Barbell or DNA Chart
- Butterfly Chart
- Box Whisker Plot
- Filters in tableau part 1
- Filters in tableau part 2

UNIT 3: Dashboard & Actions

- Parameters in tableau
- Sets in tableau
- Table Calculation
- Joins & Unions
- Data Blending
- Tableau dashboard
- Tableau Server
- User Security

UNIT 4: Review of Tableau Desktop

- Bar chart
- Line chart
- Scatter plot
- Geographic map
- Filled map
- Crosstab report
- Side by side bar

- Combines axis
- Stacked bar
- Dual axis
- Bar in bar
- Heat map
- Highlight table
- Pie chart
- Tree map
- Motion chart
- Histogram
- Dashboard
- Groups
- Sets
- Dates
- Calculations
- Quick table
- Parameters
- Reference lines

Module 15
Placement Preparation

UNIT 1:

- **Crafting an impressive Resume -**

A resume is a summary of your career, whether yours is just getting started or has been going on for years. In this program, we will help you in crafting an impressive resume.

- **Portfolio Building -**

A portfolio helps you to organize information so that you're always ready to apply for jobs. Your portfolio creates an image of you in an employer's mind. In this curriculum, you will learn about building the best portfolio.

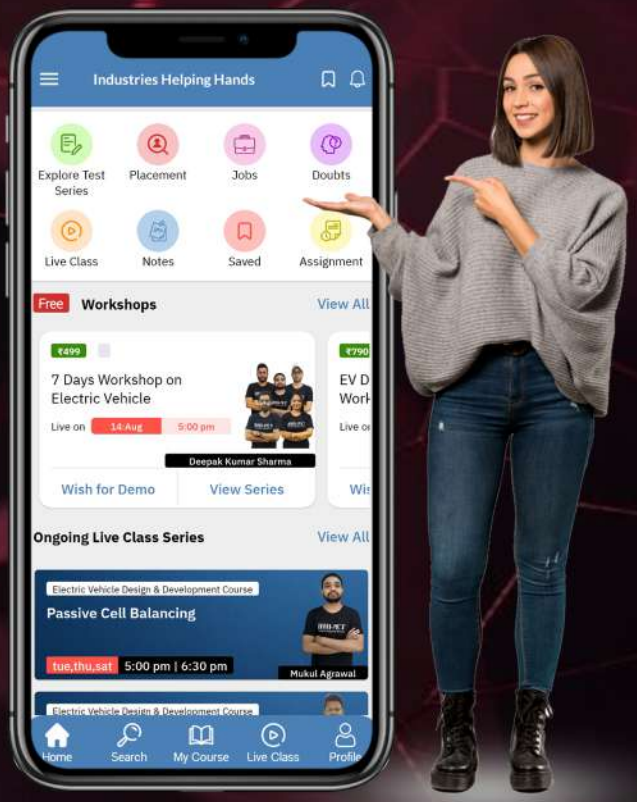
- **Interview Tips -**

Interview tips will help you gain valuable insights into effective communication and key strategies to showcase your skills confidently.

Online face-to-face mock interviews help you train & prepare for job interviews in a no-pressure, stress-free environment simulating a real job interview. Mock interviews will be conducted for you in this course.

Subjects	Master of Data Analytics Program	Full Stack Data Science and Data Analytics Program
1. Fundamental of Python	✓	✓
2. Advance Python	X	✓
3. Basic Python Libraries	✓	✓
4. Advance Python Libraries	X	✓
5. Statistics	X	✓
6. Feature Engineering	X	✓
7. Exploratory Data Analysis	X	✓
8. Machine Learning -I Supervised	X	✓
9. Machine Learning -II Unsupervised	X	✓
10. Deep Learning	X	✓
11. Natural Language Processing – NLP	X	✓
12. Computer Vision (Open CV)	X	✓
13. SQL	✓	✓
14. Power BI	✓	✓
15. Advanced Excel	✓	✓
16. Tableau	✓	✓
17. Placement Preparation Module	✓	✓

TAKE THE FIRST STEP TOWARDS YOUR DREAM CAREER IN DATA SCIENCE & ANALYTICS



ANDROID APP ON Google play

Download Our Android App

Have more questions?

Reach out to us at
WhatsApp at +91-7880-113-112