

Data Analysis Project Using Python

ZOMATO

Zomato has an average of 17.5 million monthly transacting customers for its food delivery business. The average monthly active food delivery restaurant partners on Zomato's platform have also increased by 8.7% year-on-year, from 208,000 to 226,000. You are working in a data-driven role at Zomato. You have a dataset of customers. As a data professional, you need to analyze the data, perform EDA (Exploratory Data Analysis) and visualization, and answer the following questions:



Data Analysis Project Using Python

ZOMATO

	name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0	Jalsa	Yes	Yes	4.1/5	775	800	Buffet
1	Spice Elephant	Yes	No	4.1/5	787	800	Buffet
2	San Churro Cafe	Yes	No	3.8/5	918	800	Buffet
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	300	Buffet
4	Grand Village	No	No	3.8/5	166	600	Buffet
...
143	Melting Melodies	No	No	3.3/5	0	100	Dining
144	New Indraprasta	No	No	3.3/5	0	150	Dining
145	Anna Kuteera	Yes	No	4.0/5	771	450	Dining
146	Darbar	No	No	3.0/5	98	800	Dining
147	Vijayalakshmi	Yes	No	3.9/5	47	200	Dining

148 rows × 7 columns

ZOMATO

- 1) What type of restaurant do the majority of customers order from?
- 2 How many votes has each type of restaurant received from customers?
- 3) What are the ratings that the majority of restaurants have received?
- 4) Zomato has observed that most couples order most of their food online. What is their average spending on each order?
- 5) Which mode (online or offline) has received the maximum rating?
- 6) Which type of restaurant received more offline orders, so that Zomato can provide those customers with some good offers?

Connect with Us

 **theiscale**

The iScale Organization Handle



Instagram Handle

 **theiscale.founders**

Siblings - Nishant Dhote & Swati Dhote

Website – www.theiscale.com

Contact : 7880-113-112

[theiscale](#)[theiscale.founders](#)

Zomato Data Analysis Using Python

Step 1: Import necessary Python libraries.

```
In [10]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

pandas is used for data manipulation and analysis.
numpy is used for numerical operations.
matplotlib.pyplot and seaborn are used for data visualization.

Step 2: Create the data frame.

```
In [2]: dataframe = pd.read_csv("Zomato data .csv")  
print(dataframe.head())
```

	name	online_order	book_table	rate	votes	\
0	Jalsa	Yes	Yes	4.1/5	775	
1	Spice Elephant	Yes	No	4.1/5	787	
2	San Churro Cafe	Yes	No	3.8/5	918	
3	Addhuri Udupi Bhojana	No	No	3.7/5	88	
4	Grand Village	No	No	3.8/5	166	

	approx_cost(for two people)	listed_in(type)
0	800	Buffet
1	800	Buffet
2	800	Buffet
3	300	Buffet
4	600	Buffet

```
In [3]: dataframe = pd.read_csv("Zomato data .csv")
```

In [4]: dataframe

Out[4]:

		name	online_order	book_table	rate	votes	approx_cost(for two people)	listed_in(type)
0		Jalsa	Yes	Yes	4.1/5	775	800	Buffet
1		Spice Elephant	Yes	No	4.1/5	787	800	Buffet
2		San Churro Cafe	Yes	No	3.8/5	918	800	Buffet
3		Addhuri Udupi Bhojana	No	No	3.7/5	88	300	Buffet
4		Grand Village	No	No	3.8/5	166	600	Buffet
...	
143		Melting Melodies	No	No	3.3/5	0	100	Dining
144		New Indraprasta	No	No	3.3/5	0	150	Dining
145		Anna Kuteera	Yes	No	4.0/5	771	450	Dining
146		Darbar	No	No	3.0/5	98	800	Dining
147		Vijayalakshmi	Yes	No	3.9/5	47	200	Dining

148 rows × 7 columns

let's convert the data type of the “rate” column to float and remove the denominator.

```
In [6]: def handleRate(value):
    value=str(value).split('/')
    value=value[0];
    return float(value)

dataframe['rate']=dataframe['rate'].apply(handleRate)
print(dataframe.head())
```

		name	online_order	book_table	rate	votes	\
0		Jalsa	Yes	Yes	4.1	775	
1		Spice Elephant	Yes	No	4.1	787	
2		San Churro Cafe	Yes	No	3.8	918	
3		Addhuri Udupi Bhojana	No	No	3.7	88	
4		Grand Village	No	No	3.8	166	

	approx_cost(for two people)	listed_in(type)
0	800	Buffet
1	800	Buffet
2	800	Buffet
3	300	Buffet
4	600	Buffet

summary of the data frame

In [7]: `dataframe.info()`

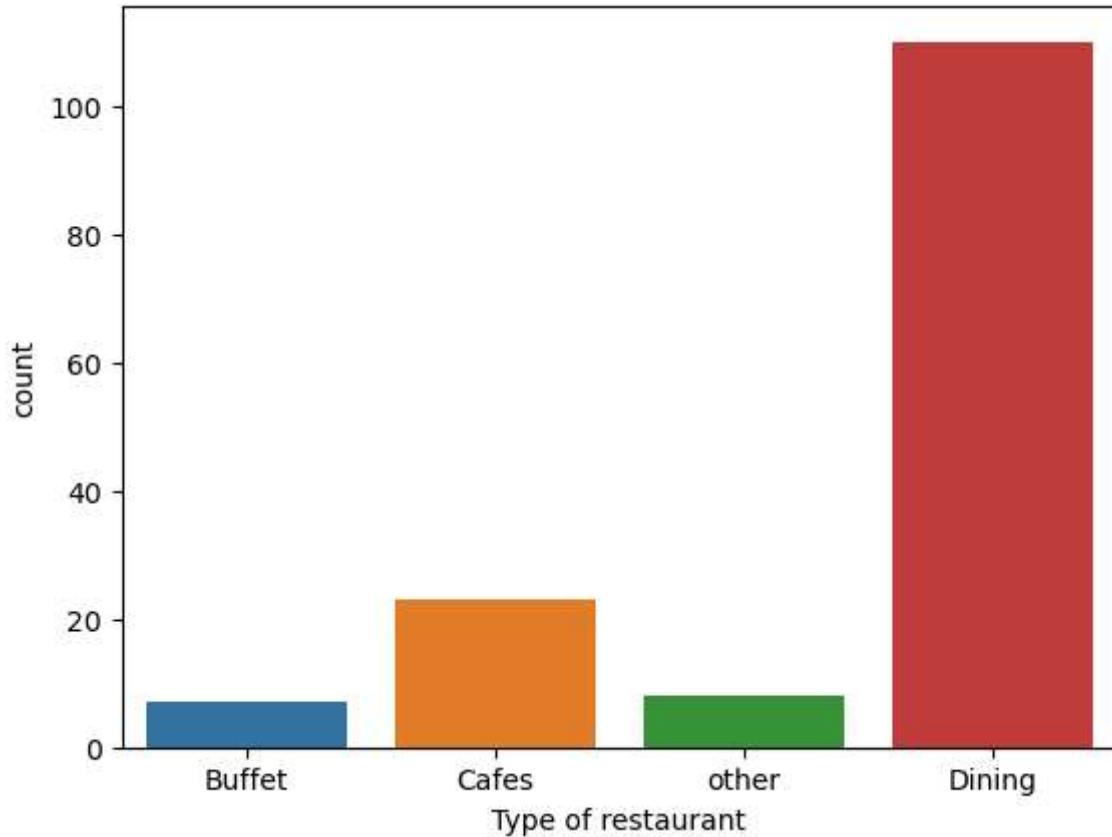
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148 entries, 0 to 147
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   name            148 non-null    object  
 1   online_order    148 non-null    object  
 2   book_table      148 non-null    object  
 3   rate            148 non-null    float64 
 4   votes           148 non-null    int64   
 5   approx_cost(for two people) 148 non-null    int64   
 6   listed_in(type) 148 non-null    object  
dtypes: float64(1), int64(2), object(4)
memory usage: 8.2+ KB
```

Conclusion - There is no NULL value in dataframe.

Type of Restaurant

```
In [8]: sns.countplot(x=dataframe['listed_in(type)'])  
plt.xlabel("Type of restaurant")
```

Out[8]: Text(0.5, 0, 'Type of restaurant')

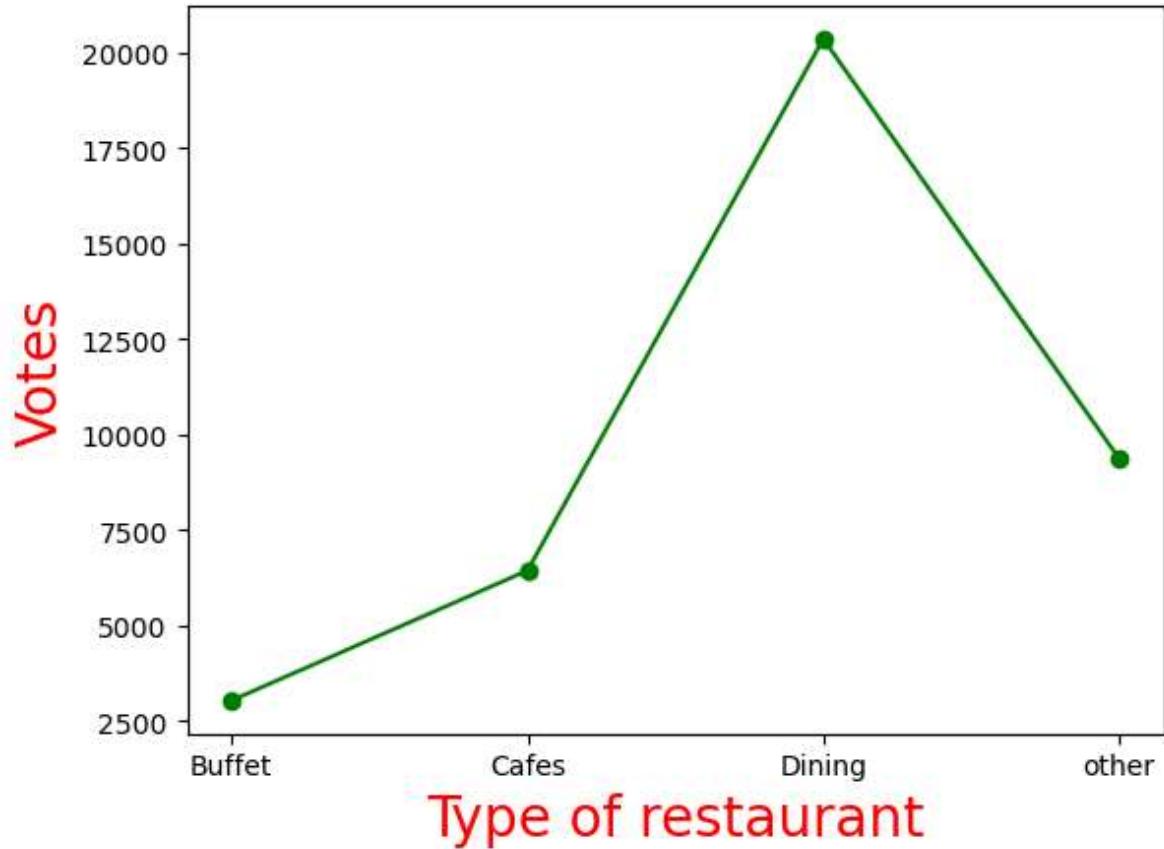


Conclusion: The majority of the restaurants fall into the dining category.

Dining restaurants are preferred by a larger number of individuals.

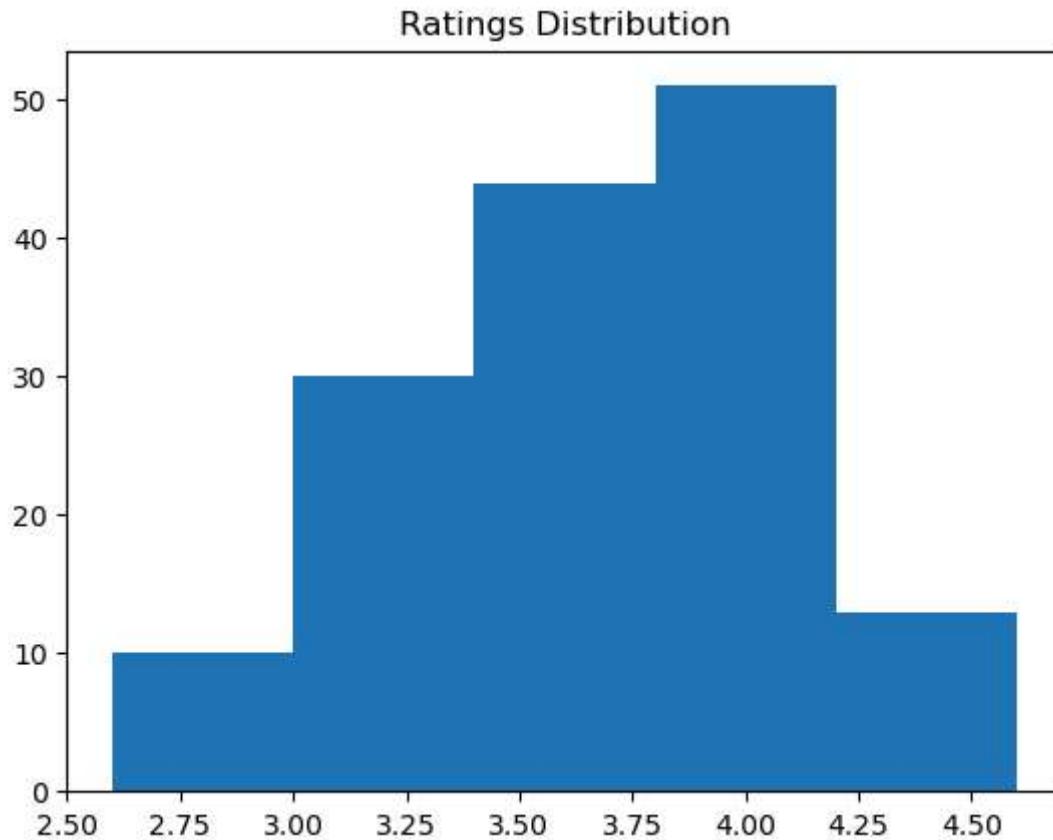
```
In [9]: grouped_data = dataframe.groupby('listed_in(type))['votes'].sum()
result = pd.DataFrame({'votes': grouped_data})
plt.plot(result, c="green", marker="o")
plt.xlabel("Type of restaurant", c="red", size=20)
plt.ylabel("Votes", c="red", size=20)
```

Out[9]: Text(0, 0.5, 'Votes')



The majority of restaurants received ratings

```
In [11]: plt.hist(dataframe['rate'],bins=5)
plt.title("Ratings Distribution")
plt.show()
```

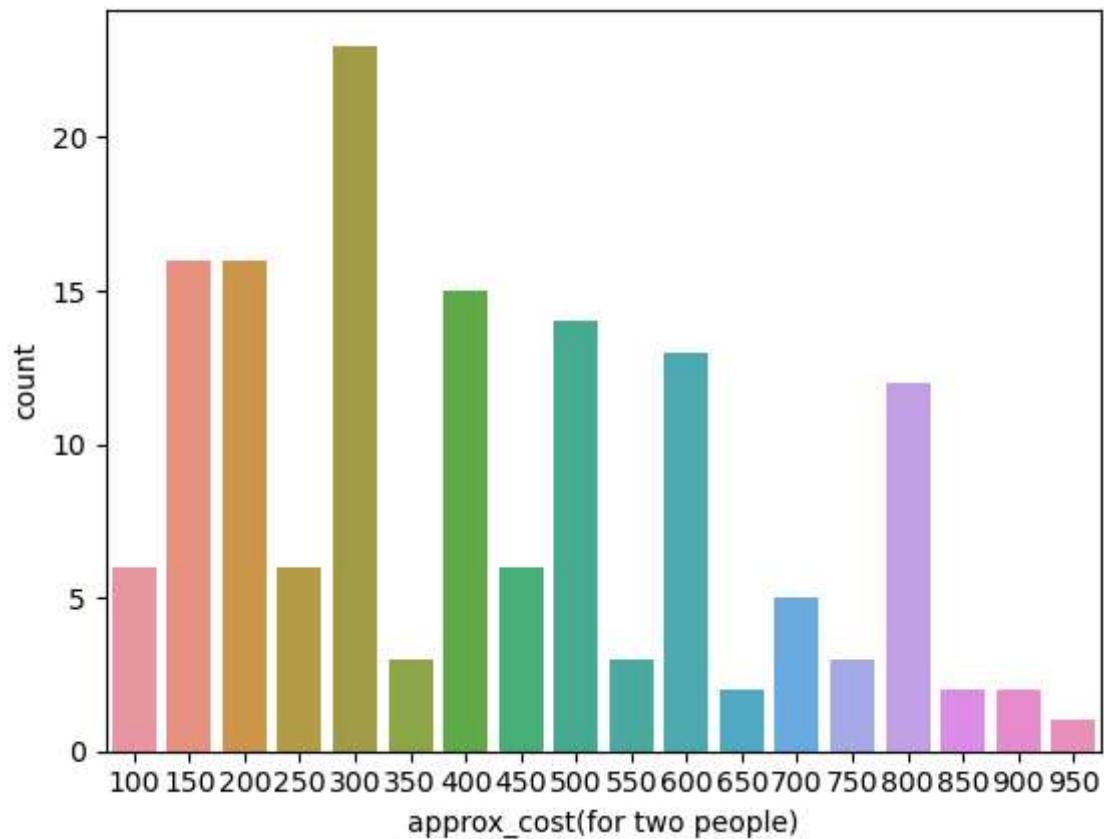


Conclusion: The majority of restaurants received ratings ranging from 3.5 to 4.

The majority of couples prefer restaurants with an approximate cost of 300 rupees.

```
In [15]: couple_data=dataframe['approx_cost(for two people)']
sns.countplot(x=couple_data)
```

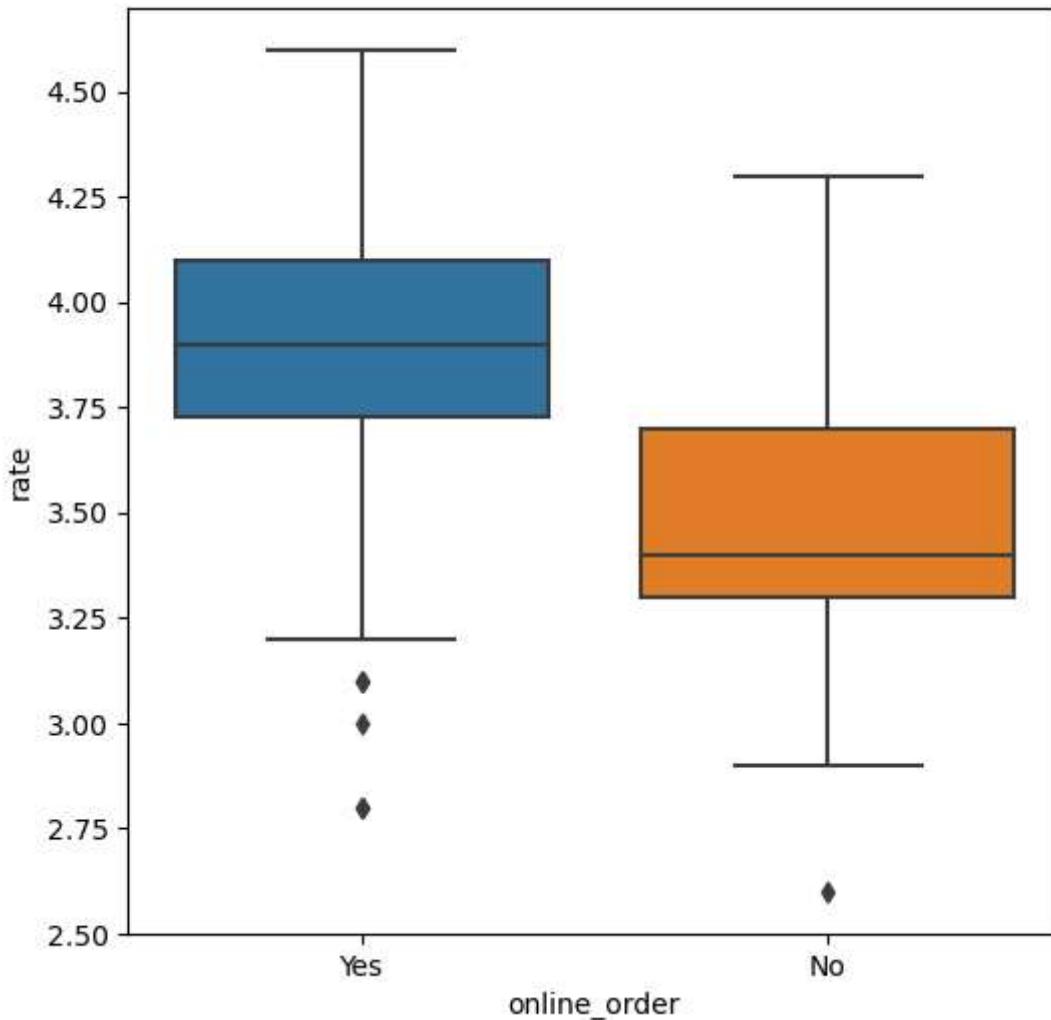
```
Out[15]: <Axes: xlabel='approx_cost(for two people)', ylabel='count'>
```



whether online orders receive higher ratings than offline orders.

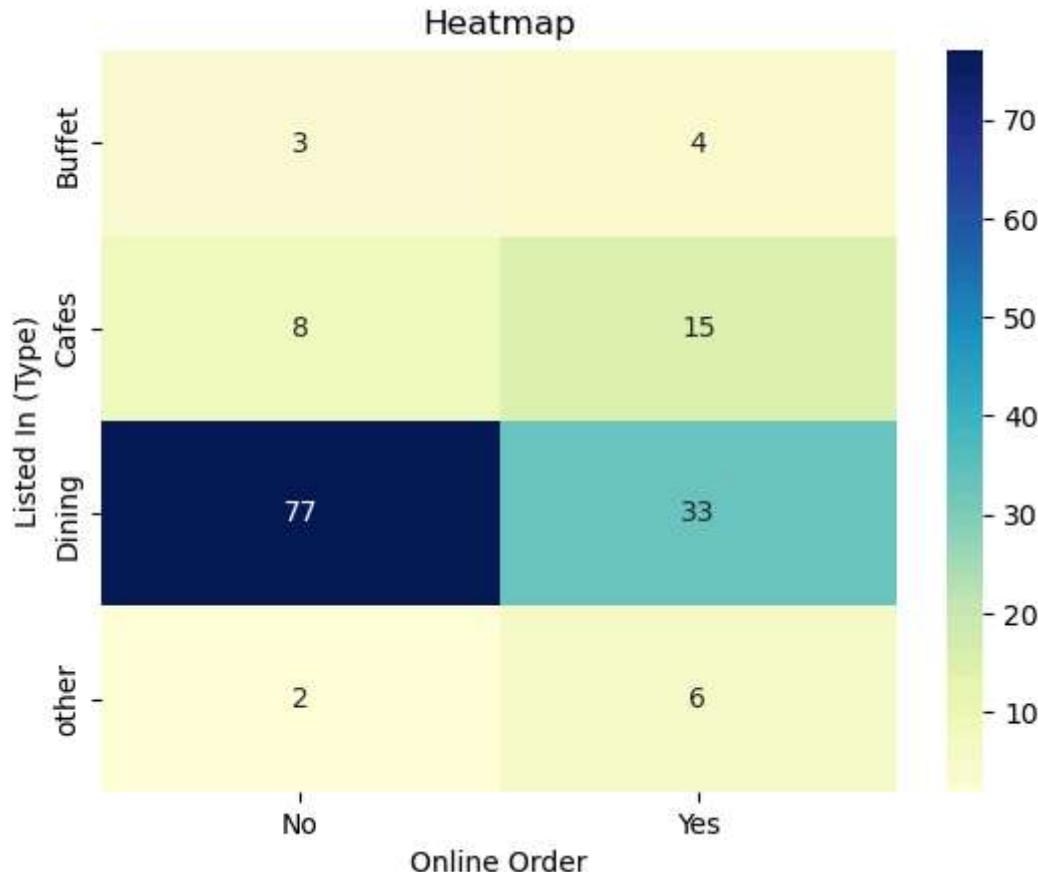
```
In [16]: plt.figure(figsize = (6,6))
sns.boxplot(x = 'online_order', y = 'rate', data = dataframe)
```

```
Out[16]: <Axes: xlabel='online_order', ylabel='rate'>
```



CONCLUSION: Offline orders received lower ratings in comparison to online orders, which obtained excellent ratings.

```
In [17]: pivot_table = datafram.pivot_table(index='listed_in(type)', columns='online_order', values='id', aggfunc='count')
sns.heatmap(pivot_table, annot=True, cmap="YlGnBu", fmt='d')
plt.title("Heatmap")
plt.xlabel("Online Order")
plt.ylabel("Listed In (Type)")
plt.show()
```



CONCLUSION: Dining restaurants primarily accept offline orders, whereas cafes primarily receive online orders. This suggests that clients prefer to place orders in person at restaurants, but prefer online ordering at cafes.

```
In [ ]:
```

FULL STACK DATA SCIENCE & DATA ANALYTICS PROGRAM

- Practical Oriented Program
- Industry Oriented Projects
- Business Case Studies
- Real World Portfolio

**Master Your Data Science & Analytics Skills
and take your career to the next level**

INDEX

- 1** Hottest Job of the 21st Century
- 2** Diverse Role in Industry
- 3** Top Companies
- 4** About the Program
- 5** Program Highlights
- 6** Dedicated Career Support
- 7** Program Outcomes
- 8** Tools & Frameworks
- 9** Learning Path
- 10** Hands-on Industry Oriented Projects
- 11** Academic Team
- 12** Connected CEOs & CHROs
- 13** Student Testimonials
- 14** Talk with Community
- 15** Curriculum

WHY DATA SCIENTIST & ANALYST?

Hottest Job of the 21st Century

Data will change the world, and becoming data-literate will pay dividends for your career. Data scientist & analytics is a fast-growing field and it will continue to grow over the next decade. Since many companies are now working based on data, its analysis is crucial. This is why the demand for data scientist & analysts will grow more and more in the future.



Data analysts and scientists will become the number 1 emerging role in the world.



Only 25% of new employees feel confident enough to use organizational data.



The global demand for data analysts & science professionals is expected to increase by **11.5 million by 2026**.



The data scientist position is the **highest-paying entry-level job** in the world.



54% of the working population need significant reskilling & upskilling.



The **average ROI** for enterprises using business intelligence and analytics is **1300%**

Diverse Role In Industry



Data Scientist



Business Analyst



Data Analyst



Marketing Analyst



Product Analyst



ML Engineer



BI Developer



Data Research Scientist

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:

Deloitte.   Microsoft   

  HSBC     

 Adobe     

About the Program

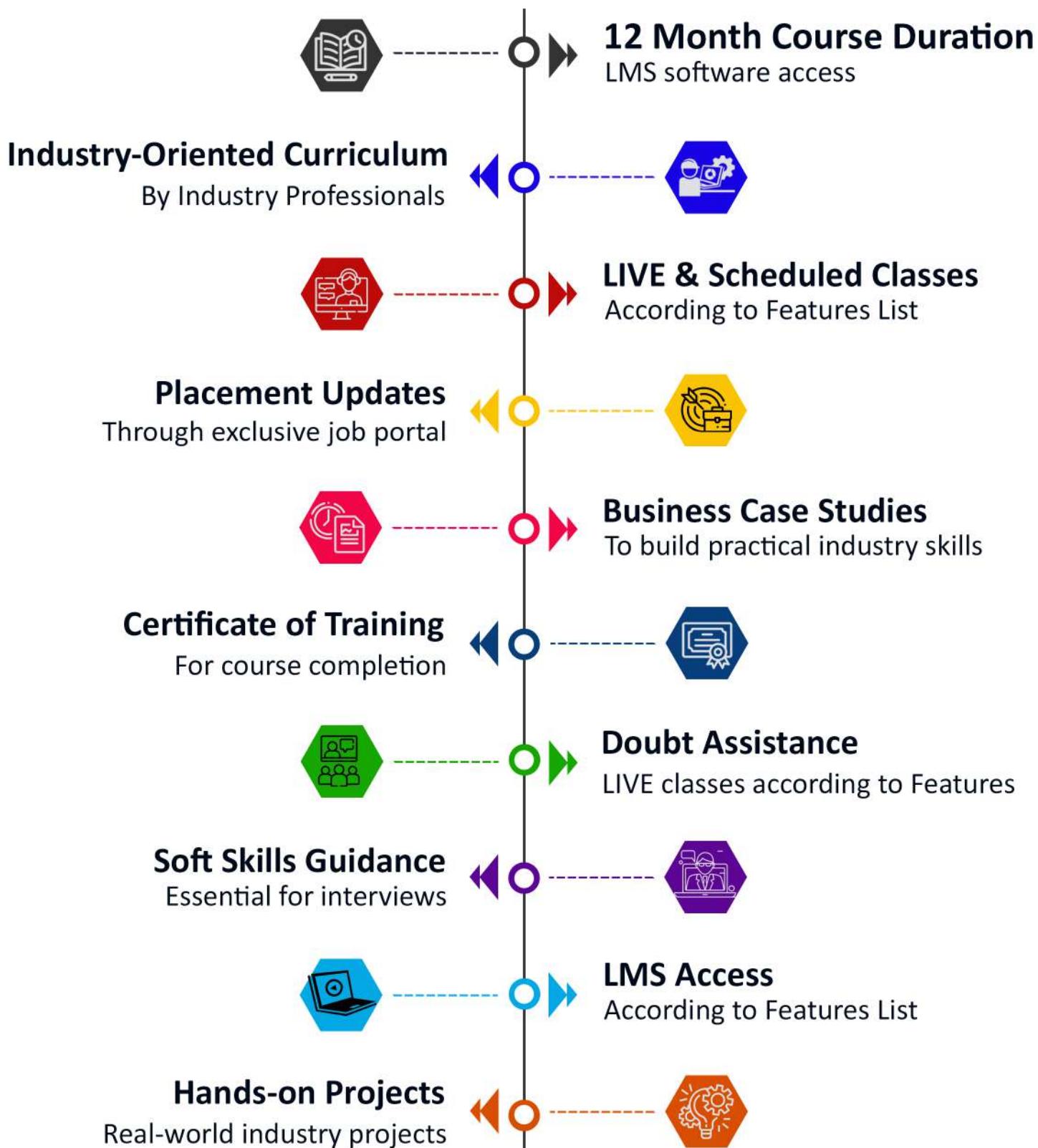


Embark on a transformative learning experience with our online program tailored for aspiring data scientists and analytics enthusiasts. This comprehensive curriculum is designed to instill proficiency in industry-standard tools, statistical analysis, and machine learning techniques. Through hands-on projects and expert-led instruction, you'll not only build a strong foundation in data science but also gain practical skills to excel in real-world scenarios. With a focus on flexibility and career development support, this program empowers you to launch a successful career in the dynamic and rapidly growing field of Data Analytics and Data Science.

Our program prioritizes flexibility, allowing you to shape your learning experience to fit your schedule. Helps in career development, preparing you for success in competitive landscape of Data Science and Analytics.

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.



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.



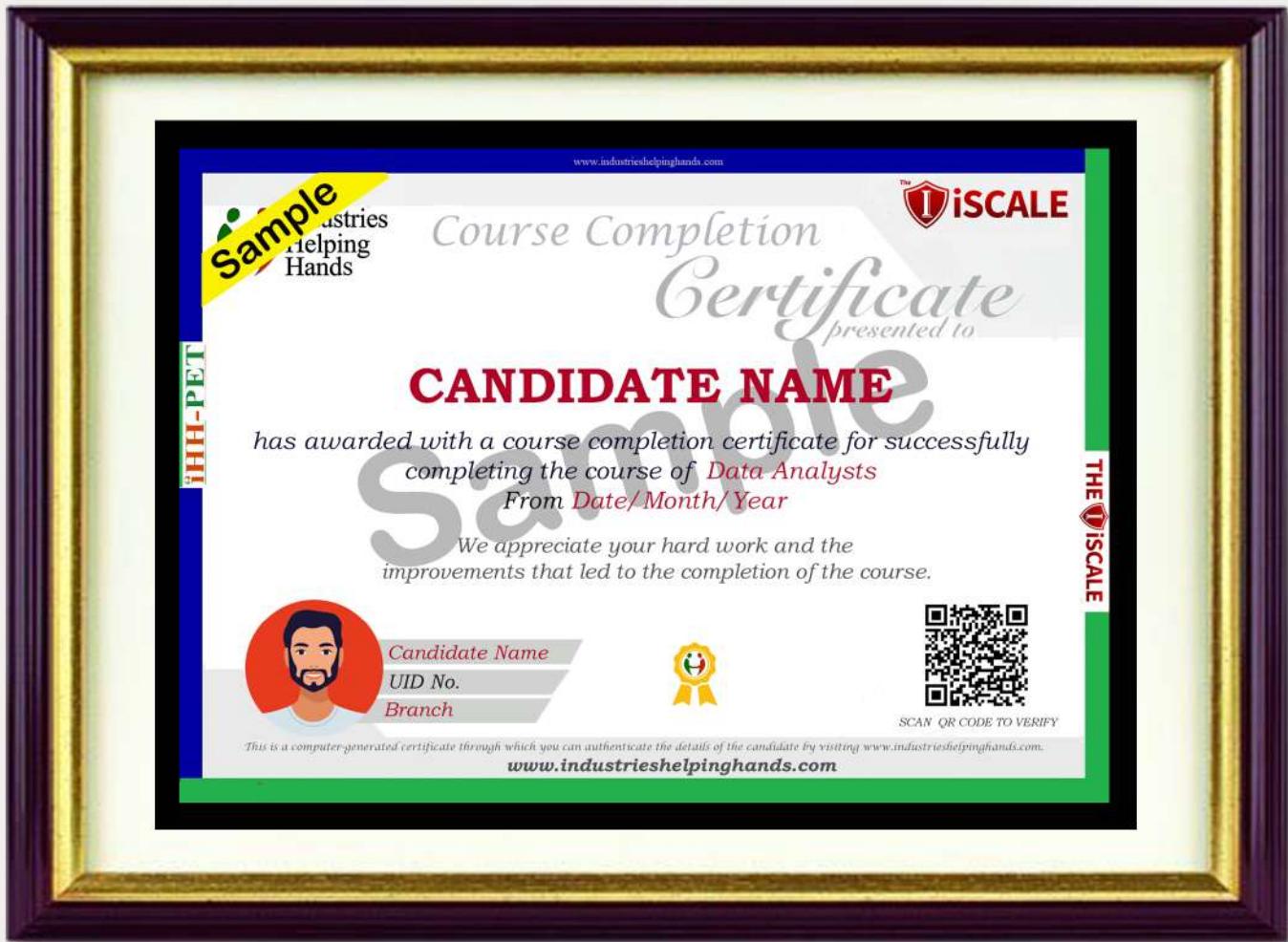
Program Outcomes

- 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



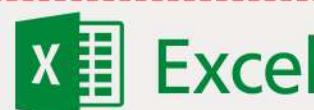
Get the Certificate of Course Completion

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

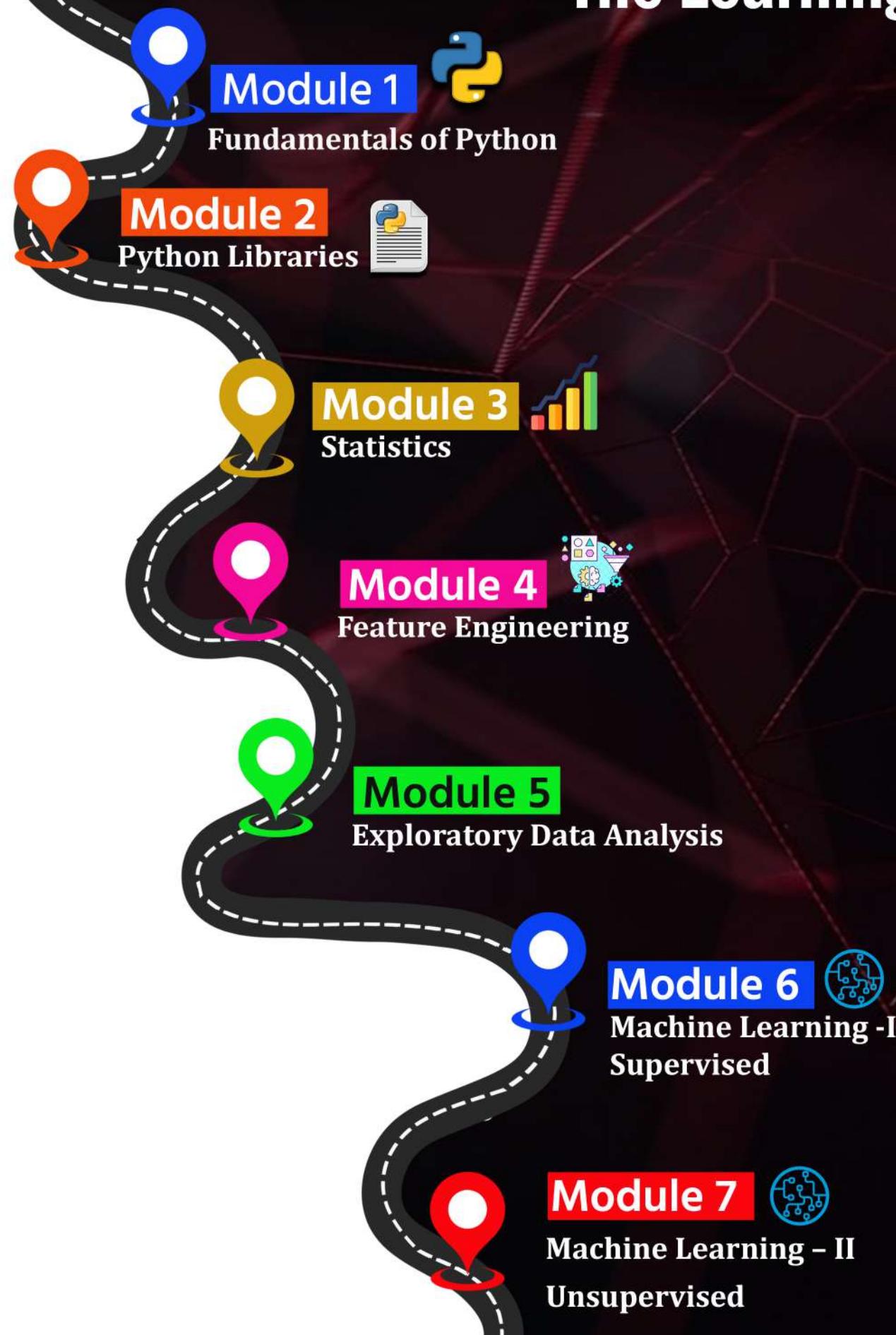


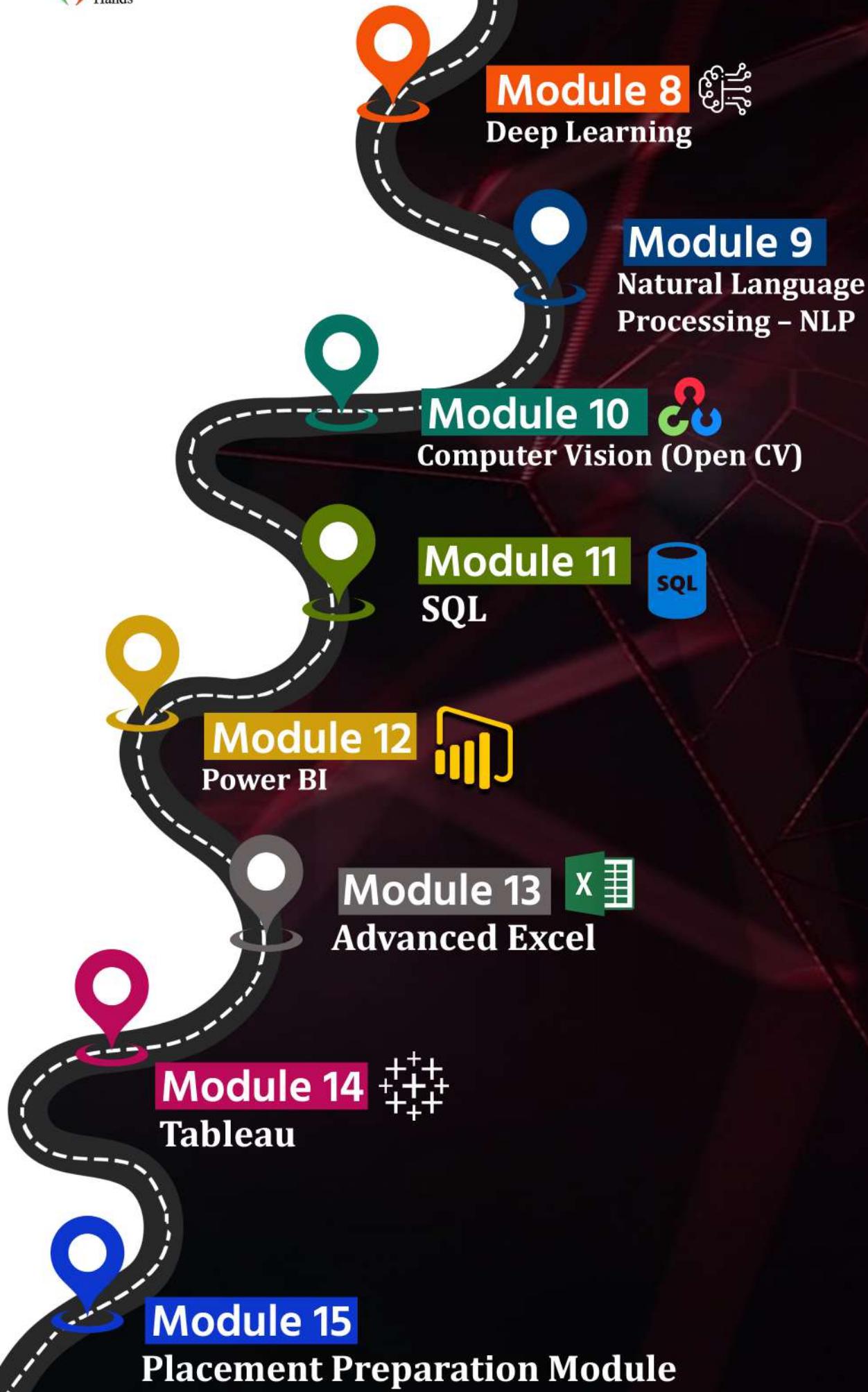
Tools & Frameworks

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.



The Learning Path





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

Hands on Industry Oriented Projects

10 PROJECT



Credit Card
Fault Detection

11 PROJECT



Object
Racking

12 PROJECT



Image
Classification

13 PROJECT



Image to
Text

14 PROJECT



Indian Premier
League

15 PROJECT



Shipping
E-Commerce

16 PROJECT



Virat Kohli ODI T20
Performance Index

17 PROJECT



Netflix Stock
Prediction

18 PROJECT



Moon Mission
Analysis Project

Academic Team

For helping you in your learning Journey

Nishant Dhote

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



Swati

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



Vibhor Sahu

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



Bhavesh

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



Bijay

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



Sweta Agrawal

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



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

Jainendra 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

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



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



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



Anand



Jashan



Ayushi



Roshan



Anindya



Vaibhav



Ankush



Nishant



Lokesh



Harshit



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.

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 Reqr Implementation

Module 7

Machine Learning - II Unsupervised

PART 1: PCA

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

PART 2: Clustering Algorithms

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

UNIT 3: Anomaly Detection

- Anomaly Detection Isolation Forests
- DBSCAN Anomaly Detection
- Local Outlier Factor Anomaly Detection

PART 4: Time Series

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

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 clause
- 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
- Reference 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	✗	✓
3. Basic Python Libraries	✓	✓
4. Advance Python Libraries	✗	✓
5. Statistics	✗	✓
6. Feature Engineering	✗	✓
7. Exploratory Data Analysis	✗	✓
8. Machine Learning -I Supervised	✗	✓
9. Machine Learning -II Unsupervised	✗	✓
10. Deep Learning	✗	✓
11. Natural Language Processing – NLP	✗	✓
12. Computer Vision (Open CV)	✗	✓
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



Download Our Android App

Have more questions?

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