

ABOUT THE COURSE

| | |
|----------------------------------|--|
| COURSE NAME: | Data Science with Python. |
| TOTAL DURATION: | 45 Hrs |
| MODE OF DELIVERY | PHYSICAL CLASSROOM TRAINING AT RESPECTIVE COLLEGES |
| TRAINER TO STUDENT RATIO: | 1:50 |
| TOTAL MARKS: | 75 |

Table 1

| | |
|----------------------------------|--|
| OVERALL COURSE OBJECTIVE: | Data science is rapidly transforming various industries, from healthcare and finance to marketing and education. As a result, the demand for skilled data scientists is skyrocketing. This course will help students comprehend the usage of Python and its libraries for creation of applications in the data science area for various real world problems. The field of data science is experiencing explosive growth, leading to numerous job opportunities in various sectors. Learning data science with Python can significantly enhance a student's skill and industry readiness. |
|----------------------------------|--|

| | |
|--------------------------|--|
| LEARNING OUTCOME: | <ol style="list-style-type: none">1. Comprehend applications of data science to real world problems and industries2. Use python libraries for crafting simple models as solutions for data science use cases3. Use techniques to collect & organize data.4. Use techniques for plotting and visualizing data. |
|--------------------------|--|

| TABLE 2: MODULE WISE COURSE CONTENT AND OUTCOME | | | | |
|--|---|--|--|-----------------------|
| SL. NO | MODULE NAME | MODULE CONTENT | MODULE LEARNING OUTCOME | DURATION (HRS) |
| 1 | Introduction to Data Science TOC - Introduction to Data Science Infosys Springboard | Data Revolution, Components of Data Science, Class & Object, Data Science Process & Architecture, Data Science Project Life cycle, Data Science Ecosystems | To understand what data science is to recognize why data science is gaining importance in today's business world to comprehend where data science can be applied in different scenarios across industry domains to understand major components of data science stack to learn how a data science project is implemented step-by-step in a given business use-case | 5 |
| 2 | Python for DS TOC - Python for Data Science Infosys Springboard | Introduction to Numpy, Introduction to Pandas, Introduction to Pandas objects, Working with datasets, Operations in Pandas | <ol style="list-style-type: none"> 1. Explain the need of python libraries 2. Use Numpy to work with arrays 3. Use Pandas to load, explore, manipulate, analyze and process data 4. Derive statistical outcomes of a real dataset 5. Visualize data 6. Create a machine learning model for predictive analysis | 11 |
| 3 | Python for DS TOC - Python | Introduction to Data Visualization, | Python is a high-level programming | 11 |

| | | | | |
|---|---|---|--|---|
| | for Data Science Infosys Springboard | Introduction to Matplotlib, Object Oriented Interface in Matplotlib, Machine learning using sklearn | language which has gained a remarkable growth in the field of data science. This course aims to introduce you to some of the powerful, open- source python libraries used in Data Science including Numeric- Python (Numpy), Pandas, Matplotlib and Scikit-learn. | |
| 4 | Exploratory Data analysis - TOC - Exploratory Data Analysis Infosys Springboard | Collecting and Organizing Data, Importing Data, Preprocessing Data, Developing Insights from data | What is Exploratory Data Analysis Realize why is it significant in the analysis of data Grasp the techniques to collect & organize data Import data from various types of files into the Python environment Learn various data preprocessing techniques Explore and summarize data Develop insights from data | 9 |
| 5 | Data visualization - TOC - Data Visualisation using Python Infosys Springboard | Data Visualisation, Variety in Data, Data Visualisation Stakeholders, Data Visualisation, Visualisation Constructs | To distinguish between different plotting techniques for different data types. To plot different kinds of graphs - box plot, scatter plot, line | 9 |

| | | | | |
|--|--|--|---|--|
| | | | chart, dot chart, bar graph, and histogram. To plot different types of data on diverse kinds of graphs. | |
|--|--|--|---|--|

| TABLE 3: OVERALL COURSE LEARNING OUTCOME ASSESSMENT CRITERIA AND USE CASES | | | |
|---|--------------------------------------|---|---|
| LEARNING OUTCOME | ASSESSMENT CRITERIA | Performance Criteria | USE CASES |
| Apply Python libraries for data analysis and modeling. | Completion of practical assignments. | Construct clean and organized datasets; implement Python libraries for statistical and predictive modeling. | Automate data wrangling for a retail sales dataset. |
| Develop effective data visualizations. | Project-based evaluation. | Design and present advanced visualizations that support clear decision-making. | Create dashboards for stock market analysis using Python visualization libraries. |
| Build machine learning models for predictive | Model performance | Propose suitable models and justify their efficiency | Predict customer churn using classification |

| | | | |
|---|---|---|---|
| analytics. | evaluations. | through accuracy and precision metrics. | models. |
| Summarize data insights using EDA techniques. | Hands-on exercises on preprocessing data. | Develop clean datasets and identify patterns or anomalies using Python libraries. | Prepare datasets for sales forecasting through preprocessing and summarization. |
| Design an end-to-end data science pipeline. | Capstone project evaluation. | Create and deploy a data pipeline integrating EDA, visualization, and predictive modeling for a real-world problem. | Design a credit scoring system for a financial institution. |

TABLE 4: LIST OF FINAL PROJECTS (PROJECTS THAT COMPREHENSIVELY COVER ALL THE LEARNING OUTCOME)

| SL.NO | FINAL PROJECT |
|-------|---|
| 1 | <p>A company wants to study the demographic data to make predictions about the earning potential of the population. However, the data gathered is not clean for analysis. The company requests you, as a data scientist, to perform the following operations and gain some insights from the data for data driven competitive advantage.</p> <ol style="list-style-type: none"> 1. Remove data with missing values 2. Remove outliers 3. Establish the importance of the weekly working hours on earning potential 4. Find the features that are highly correlated with the earning potential 5. Find the relation between the number of years spent to get the degree and earning potential |

| | |
|--|--|
| | <p>6. Find the relationship between age and earning potential</p> <p>Data Source</p> <p>The dataset for the problem is available here.</p> <p>Contents of the folder are as follows.</p> <ul style="list-style-type: none"> • Dataset • Dataset Description |
|--|--|

| TABLE 5: COURSE ASSESSMENT RUBRICS (TOTAL MARKS: 75) | | | | | |
|---|--|---|---|--|--------------------|
| ASSESSMENT CRITERIA | Learning Outcome | Fair (1–5) | Good (6–10) | Excellent (11–15) | TOTAL MARKS |
| Data Preprocessing and Organization | Summarize data insights using EDA techniques. | Demonstrates limited ability to preprocess and organize datasets. | Prepares datasets with minor inconsistencies and partial preprocessing. | Prepares clean, well-structured datasets with comprehensive preprocessing. | 15 |
| Python Coding and Modeling | Apply Python libraries for data analysis and modeling. | Implements basic coding solutions with limited functionality. | Produces functional scripts using Python libraries with minor errors. | Develops optimized scripts for efficient data analysis and modeling. | 15 |

| | | | | | |
|------------------------------|---|--|--|--|----|
| Visualization Design | Develop effective data visualizations. | Creates basic visualizations with minimal insights or clarity. | Designs clear visuals with adequate insights and logical presentation. | Develops advanced, impactful visualizations that communicate key insights effectively. | 15 |
| Predictive Model Development | Build machine learning models for predictive analytics. | Builds basic models with low accuracy or limited application. | Creates functional models with moderate accuracy and reliability. | Develops optimized, highly accurate models, demonstrating advanced techniques. | 15 |
| Capstone Project Execution | Design an end-to-end data science pipeline integrating EDA, visualization, and predictive modeling. | Demonstrates limited integration of techniques into the project. | Integrates most techniques with some inconsistencies or gaps. | Seamlessly integrates all techniques, creating a cohesive and effective data science solution. | 15 |