

Data Science using SAS and Python

Data Science is a study of data. We believe Data is everything in today's world that can be treated as 'Information' about different aspects of Business.

In this course, we will cover all the components of Data Science like Data Management, Programming, Visualization, and Analytics.

All the components of Data Science mentioned above are equally important to successfully become a Data Scientist.

This course focuses mainly on SAS and Python tools. It starts with programming using SAS and Python followed by analytical and visualization techniques. It includes multiple hands-on exercises and project work in the domains of marketing, banking, finance, risk management, retail, clinical, manufacturing/inventory management etc.

The ultimate purpose of this course is to cover all the basics and advance predictive modeling techniques which are a must in today's competitive edge.

This course will cover all the programming, visualization and analytical models to be implemented in Python and SAS.

Key Features:

- Online/Classroom Training
- Flexible Schedule
- 35 Days (2 Hrs. per day)
- Certification and Job Assistance
- Project work & Exercises
- Study Materials (Software /Reference Book/Interview Preparation/ Project Data / SAS & Python Codes)

Faculty Profile

Our Faculty has 9+ years of experience in the field of Data and Analytics, Data Science, Training, and Analytical application development.

Our Faculty is a certified Predictive Modeler and Programmer from SAS Institute Inc.

Our Faculty worked on SAS (All analytical modules), Python (Data Science and Data Management libraries), and Big Data (Hadoop and Hive) during his 9+ years of tenure.

He has successfully implemented analytics and developed analytical applications in the field of AML, Risk Management, Healthcare Fraud, Retail, and Market Research.

Former employee of TCS and SAS Institute Inc.

Course Content:

Topic 1: Introduction to Data Science

- Welcome/General Discussion about the expectation from course
- Definition of Data
- Difference between data management and data analytics
- Data Science components

Topic 2: Programming using SAS

Q&A/Project discussion based on previous days

- Base SAS Overview
- Data Step and Proc Step processing:
- Concept of PDV, Input Buffer
- Concept of SAS library and SAS Catalog
- Variable Types in SAS
- Reading Data stored external to SAS
- Importing Data by using Proc Import
- Data Step SAS statements
- SAS Functions
- Appending and Merging using SAS
- SAS Procedures like proc means, Proc Univariate, proc append proc freq, and proc export.
- SAS SQL
- SAS Macros

Topic 3: Programming using Python

Q&A/Project discussion based on previous days

- Python Overview
- Python Data Types
- Python operations using Numbers, String, Logical, Arithmetic and so on
- Python Strings
- Python Lists
- Python Tuple
- Python Dictionary
- FOR and WHILE loops
- IF/THEN/ELSE in Python
- Data Manipulation using Numpy and Pandas



Topic 4: Statistics using SAS and Python

Q&A/Project discussion based on previous days

- Levels of Measurement and Variable types
- Descriptive Statistics and Picturing Distributions
- Confidence Interval for the Mean

Topic 5: Hypothesis Testing and ANOVA using SAS and Python

Q&A/Project discussion based on previous days

- One Sample t-test of comparing means
- Two Sample t-test of comparing means
- One Way ANOVA
- Assumptions of ANOVA Modeling
- n-way ANOVA
- ANOVA Post Hoc Studies

Topic 6: Exploratory Data Analysis using SAS and Python

Q&A/Project discussion based on previous days

- Data Exploration by using Scatter Plots
- Pearson and Spearman Correlations

Topic 7: Linear Regression using SAS and Python

Q&A/Project discussion based on previous days

- Fit Simple Linear Regression Model
- Assumptions of Linear Regression Model
- Analyze the output of the Linear Regression
- Producing Predicted Values
- Difference between Simple Linear Regression and Multiple Linear Regression Models
- Fit Multiple Linear Regression Model
- Stepwise Regression/Model Selection Techniques



Topic 8: Regression Diagnostics using SAS and Python

Q&A/Project discussion based on previous days

- Residual Analysis
- Influential Observation
- Difference between Influential Observation and Outliers
- Collinearity Diagnostics

Model Building Process using SAS and Python

Topic 9: Categorical Data Analysis using SAS and Python

Q&A/Project discussion based on previous days

- Examining Distributions
- Test of Associations by using the chi-square test
- Fisher's Exact p-values for Pearson Chi-square test

Topic 10: Logistic Regression using SAS and Python

Q&A/Project discussion based on previous days

- Odds and Odds Ratio
- Simple Logistic Regression
- Multiple Logistic Regression with categorical predictors
- Analyze the output of Logistic Regression

Topic 11: Measure Model Performance using SAS and Python

Q&A/Project discussion based on previous days

- Apply the principles of honest assessment to model performance measurement
- Rare event adjustments
- Assess classifier performance using the confusion matrix
- Model selection and validation using training and validation data
- Create and interpret graphs (ROC, lift, and gains charts) for model comparison and selection
- Establish effective decision cut-off values for scoring

Topic 12: Decision Tree Modeling and XGBOOST using SAS and Python

Decision Tree Modeling

- Introduction to Decision Tree Modeling
- Model essential for decision tree models
- Decision Tree model development by using CHAID, Entropy/Information gain and Gini
- Decision Tree Model tuning

Gradient Boosting (XGBOOST)

- Introduction to Boosting
- Example of Boosting
- Regression Decision Tree
- Gradient Boosted Trees Regression.

PROJECTS

Project: 1 Portfolio Data Analysis

Domain: Risk Management

Problem Statement: As an analyst, you need to advise your client to decide which mutual fund risk category should invest in.

Topics: Descriptive Analytics, Distributions, and Visualization

Project: 2 Effectiveness of Production Process

Domain: Manufacturing/Inventory Management

Problem Statement: As a manager/supervisor of a company, you need to measure the effectiveness of the production of cereal boxes. The aim is to analyze whether or not the cereal boxes' weight is as per company specifications.

Topics: Hypothesis Testing (One-Sample tests)

Project: 3 Product Assortment Strategies in Retail Store

Domain: Marketing/Retail

Problem Statement: As a regional sales manager of a company, you need to analyze mean sales comparison between two types of displays of products in the retail store. The aim is to decide whether or not the Promotional display of the product is more effective than the Normal display of the product. This helps management to decide the display location of the product in a store that will maximize sales

Topics: Hypothesis Testing (Two-Sample tests)

Project 04: Drug Analysis

Domain: Clinical

Problem Statement: Before you launch the new drug in the market, you need to analyze the effect of new drug and its different doses on blood pressure of the human body

Topics: Analysis of Variance (ANOVA Models)

Project 05: Analyzing consumption of oxygen in the human body during running

Domain: Physiology

Problem Statement: In exercise physiology, an objective measure of aerobic fitness is how effectively the body can absorb and use oxygen during their 1.5 miles run. Factors affecting oxygen consumption are runtime, age, and gender, run pulse, rest pulse, and so on. The aim is to identify the key factors affecting the oxygen consumption during a run.

Topics: Analysis of Variance (EDA and Linear Regression Models)

Project 06: Titanic Event Analysis

Domain: Event Analysis

Problem Statement: On the 14th of April, the Titanic hit an iceberg and sank. There were 1517 fatalities from different age groups, class (1, 2, and 3) and gender. The objective is to measure how all these factors are associated with the survival status of passengers.

Topics: Odds, Odds Ratio, Chi-Square tests, Ordinal associations, and Logistic Regression Model

Project 07: Marketing Campaign for a Bank

Domain: Marketing

Problem Statement: A target marketing campaign for a bank was undertaken to identify a segment of customers who are likely to respond to an insurance product. Here, the target variable is whether or not the customers bought insurance product and it depends on factors like Product usage in three months, demographics, transaction patterns as like deposit amount, checking account, a branch of the bank, Residential information (like urban, rural) and so on.

Topics: Classification, Categorical Data Analysis, Logistic regression, Decision Tree and Gradient Boosting (XGBOOST)

Course Certification

The whole Data Science course content is planned by industry experts for you to land the best positions in top MNCs.

Toward the finish of each topic, there will be project associated with that topic. The project contains live data and codes from SAS and Python.

Techinfoplace course fruition authentication will be granted to you when you complete the task work.

After completion of the projects from each topic, your skills will be equal to 6 months of rigorous industry experience in each topic.

FAQ:

1. Why should I learn Data Science from Techinfoplace?

- Techinfoplace offers exclusive Data Science online/classroom courses for professional/fresher who wants to expand their knowledge base and start a career in this field. There are many reasons for choosing Techinfoplace.
- A personal mentor to track your progress
- Techinfoplace online/classroom sessions conducted by experienced professionals in this field
- Real-time exercises, assignments, and projects
- Study materials, reference books for every topic.
- 24/7 learning support
- The large community of like-minded learners

2. What are the different modes of training that Techinfoplace provides?

- Online Session
- Classroom Session

3. Who can attend this course?

- Anyone who wants to play with Data can enter into this field.

4. What kind of projects are included as part of the training?

- This course is known for projects involved in it. This is purely Job Oriented training.
- You will work on highly exciting projects in the domains of high technology, Retail, Banking, Marketing, Clinical, Manufacturing, and so on.
- After completing the projects successfully, your skills will be equal to 6 months of rigorous industry experience in each topic.
- You will also encounter a different interview question for each topic which will help in cracking Data Science interviews

5. Is there any discount for this course?

- Yes. The fees mentioned for this course is for one candidate. If you come with at least 3 enrollments, a discount of 20% on base price would be given to that group.