Data Science with SQL Server and R Training

Duration: **4 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Overview

This {*training} introduces **R programming**, statistics, data mining, and machine learning, and shows how to apply data science within **SQL Server** and the **Microsoft BI stack**.

R is the most popular environment and language for statistical analysis, data mining, and machine learning. Its scalable version runs inside SQL Server, Power BI, and Azure ML. While the main focus of the course is **R**, it also covers how to integrate other Microsoft BI tools like**Python, T-SQL, Power BI, Azure ML, and Excel** for data science tasks.
Labs focus on R, but demos also feature other languages.

Why This Course?

- Compare R and Python
- Learn unsupervised learning techniques
- Perform matrix operations
- Visualize relationships between variables
- Prepare data for analytics
- Understand supervised learning models

Prerequisites

- Basic understanding of data analysis
- Familiarity with SQL Server tools

Data Science with SQL Server and R Training

Duration: **4 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

By the end of this {*training}, you will have gained knowledge and skills in the following areas:

- Program in R from scratch using R Engine and RStudio
- Understand the lifecycle of a data science project
- Perform data exploration and preparation
- Analyze relationships between variables using intermediate statistics
- Apply linear modeling and Bayesian inference
- Use R in SQL Server, Power BI, and Azure ML
- Explore how Python can be used across all tools via demos

Outline

Module 1: Introducing Data Science and R

- Definitions: statistics, data mining, machine learning
- Data science project lifecycle
- Introduction to R and its tools
- R data structures
- **Lab 1**

Module 2: Introducing Python

- Basic syntax and objects
- Data manipulation with NumPy and Pandas
- Visualizations with Matplotlib and Seaborn
- Machine learning with Scikit-Learn
- **Lab 2: R vs Python**

Module 3: Data Overview

- Datasets, cases, and variables
- Variable types
- Discrete and continuous variable statistics

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- Basic graphs and visualizations
- Sampling, confidence levels and intervals
- **Lab 2**

Module 4: Data Preparation

- Derived variables
- Handling missing values and outliers
- Smoothing and normalization
- Time series
- Training and test set creation
- **Lab 3**

Module 5: Associations Between Variables

- Covariance and correlation
- Contingency tables and chi-squared test
- T-tests and ANOVA
- Bayesian inference
- Linear modeling
- **Lab 4**

Module 6: Feature Selection and Matrix Operations

- Feature selection in linear models
- Basic matrix algebra
- Principal component analysis (PCA)
- Exploratory factor analysis
- **Lab 5**

Module 7: Unsupervised Learning

- Hierarchical clustering
- K-means clustering
- Association rules

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- **Lab 6**

Module 8: Supervised Learning

- Neural networks
- Logistic regression
- Decision and regression trees
- Random forests
- Gradient boosting
- K-nearest neighbors
- **Lab 7**

Module 9: Modern Topics

- Support vector machines
- Time series forecasting
- Text mining
- Deep learning
- Reinforcement learning
- **Lab 8**

Module 10: R in SQL Server and Microsoft BI

- ML Services (In-Database) architecture
- Executing external scripts in SQL Server
- Model storage and native prediction execution
- Using R in Azure ML and Power BI
- **Lab 9**