Comprehensive List of Topics to Prepare for Data Science Interviews

Mathematics

1. Linear Algebra

2. Matrix Theory

3. Calculus

Statistics

Descriptive Statistics

Sampling

- Simple Random Sampling
- Stratified Random Sampling
- Systematic Sampling
- Cluster Sampling
- Multi-stage Sampling

Parametric Hypothesis Testing

- 1 sample Z test
- 1 sample t test
- Z test for proportions
- 2 sample Z test
- 2 sample t test
- Paired samples t test
- t test for correlations
- F test for Homogeneity of variance

- One-way ANOVA
- Two-way ANOVA
- Multivariate ANOVA
- Design of Experiments

Non-Parametric Hypothesis Testing

- Chi-square test for Independence
- Chi-square test for Goodness of fit
- 1 sample Sign test
- 1 sample Wilcoxon test
- Mann-Whitney test
- Kruskal-Wallis test
- Mood's Median test
- Friedman test

Correlation Analysis

- Pearson
- Spearman
- Kendall Rank
- Point-Biserial

Regression Analysis

- Linear
- Lasso
- Ridge
- Polynomial
- Logistic

- Multinomial Logistic

Probability Distributions

- Normal
- Uniform
- Bernoulli
- Binomial
- Poisson
- Exponential

Bayesian Statistics

- Bayes Theorem
- Conditional Probability
- Priors
- Posteriors
- Maximum Likelihood

Multivariate Statistics

- Principal Component Analysis
- Linear Discriminant Analysis
- Factor Analysis

Time Series

- Moving Averages
- Exponential Smoothing
- ARMA
- ARIMA

Probability

- Theoretical Probability
- Experimental Probability
- Subjective Probability
- Or Probability Problems
- And Probability Problems
- At least one Probability Problems
- Conditional Probability Problems
- Probability Problems based on Normal Distribution
- Central Limit Theorem

SQL

- CREATE DATABASE statement
- DROP DATABASE statement
- BACKUP DATABASE statement
- CREATE TABLE statement
- DROP TABLE statement
- ALTER TABLE statement
- SELECT statement
- WHERE clause
- AND operator
- OR operator
- NOT operator

- ORDER BY keyword
- INSERT INTO statement
- NULL value
- UPDATE statement
- DELETE statement
- SELECT TOP clause
- MIN function
- MAX function
- COUNT function
- AVG function
- SUM function
- LIKE operator
- IN operator
- BETWEEN operator
- Aliases
- INNER JOIN
- LEFT JOIN
- RIGHT JOIN
- FULL JOIN
- Self JOIN
- UNION operator
- GROUP BY statement
- HAVING clause
- EXISTS operator
- ANY operator
- ALL operator

- Date Data Types
- SELECT INTO statement
- INSERT INTO SELECT statement
- CASE statement
- IFNULL function
- ISNULL function
- COALESCE function
- NVL function

Constraints in SQL

- NOT NULL
- UNIQUE
- PRIMARY KEY
- FOREIGN KEY
- CHECK
- DEFAULT
- INDEX

R Programming

- Objects
- Vectors
- Matrices
- Lists
- Factors

- Missing Values
- Data Frames
- read.table
- Reading EXCEL, CSV, TEXT files
- readr package
- Subsetting a vector
- Subsetting a Matrix
- Subsetting Lists
- dplyr package
- Operations on Dates and Times
- if else function
- for loop
- while loop
- next statement
- break statement
- debugging using traceback function
- ggplot2 package
- tidyverse package
- function creation
- lapply function
- sapply function
- split function
- tapply function
- apply function
- mapply function
- map function

Python

Python Fundamentals

- Operators
- Variables
- Data types
- Conditional statements
- Loops
- Functions

Data structures

- Lists
- Tuples
- Dictionary
- Sets

Reading CSV files

Indexing a Data Frame

Libraries

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Scipy

Machine Learning

K-Nearest Neighbour

Naïve Bayes Classification

Support Vector Machine

Random Forest

K-Means

Decision Trees

- ID3
- C4.5
- C5.0
- CART

Boosting

- AdaBoost
- Gradient Boosting
- XGBoost