Exploratory Data Analytics

UNIT-I

Short Questions

1. What are the different types of attributes
2. Write a short note on Survey method for data collection
3. What are the different types of data sets
4. What is a Semi structure data
5. Write about mode in descriptive statistics
6. Write about mean in descriptive statistics
7. What is Z-score
8. What is Skewness and Kurtosis
9. What are the different steps in EDA
10. What are the different datatypes in R programming

Long Questions

1. Explain about structured, unstructured and semi structured data
2. Explain about different attribute values
3. What are the different data collection strategies explain them
4. Explain about data similarity measures
5. Draw and explain data analytics life cycle
6. Explain different steps in the EDA
7. Compare Mode and class
8. Explain Numeric datatypes in R programming
9. What are the EDA motivations
10. Explain about Inferential statistics

UNIT-I I

Short Questions

1. Name the different types of databases
2. Advantages of centralized databases
3. Advantages of Distributed Database
4. Properties of Relational Database
5. Advantage of NoSQL Database
6. What is Data Warehousing
7. What are the different elements of Data Warehouse
8. Write a short note on Data cube
9. What are the different Map Reduce Components
10. What are the types of cloud storage

Long Questions

1. Explain about centralized databases
2. Explain about NoSQL Database
3. Explain about ETL process in Data Ware housing
4. Draw and Explain Three Tier Architecture of the data warehouse
5. Explain about cloud storage classes
6. Draw and explain cloud storage system .
7. Draw and Explain how to retrieve data from the cloud
8. How HDFS used for Big data processing
9. How to store Bigdata into No SQL explain with a block diagram
10. Describe different Map Reduce Components

UNIT-III

Short Questions

1. What is Missing Completely at Random (MCAR)
2. What is List Wise Deletion
3. What is Regression Imputation
4. What is pair wise deletion
5. Advantages of K-Mean clustering
6. Write a short note on Kernal Density Plots
7. Write a short note on Correlation Heatmaps
8. Write a short note on Principle component Analysis
9. Regression Imputation dis advantages
10. What is Not Missing at Random(NMAR)

Long Questions

1. Explain the traditional methods of dealing with missing data
2. Explain about List Wise deletion
3. What are the different methods of Handling Missing data
4. Explain how to handle missing data with a example
5. What are the practical issues in multiple Imputation
6. Explain different data modelling techniques
7. Explain about data cleaning and transformation
8. Explain about schema design
9. Explain about data relationships and joins
10. Explain about Bivariate data Modelling

UNIT-IV

Short Questions

1. Write a short note on data cleaning
2. Write a short note on data integration
3. Write a short note on normalization and standardization
4. Write a short note on Feature Engineering
5. Write a short note on Ingesting data
6. Write a short note on Trifacta Wrangler
7. What are the different data quality improvement measures
8. How to facilitate exploratory data analytics
9. Write a short note on describing data
10. Write a short note on Ad hoc Reporting

Long Questions

1. Importance of Data wrangling
2. Explain about data transformation techniques
3. How to handle outliers
4. Explain different tasks of data wrangling
5. Explain about design and building refined data
6. Explain about data wrangling tools
7. Explain how Excel is used for data wrangling
8. Explain how Trifacta Wrangler is used for data wrangling
9. How the data wrangling is used to convert raw data into machine ready format (CSV,JSON,XML)
10. What is a structure in a dataset

Unit-V

Short Questions

1. Write a short note on describe() in pandas
2. Write a short note on value\_counts() in pandas
3. Write a short note on info() in pandas
4. Write about head() function in pandas
5. Write about tail() function in pandas
6. Write about isnull() function in pandas
7. Write about tail() function in pandas
8. Write about mean() function in pandas
9. Write about median() function in pandas
10. Write about boxplot() function in pandas

Long Questions

1. Demonstrate Range by using EDA
2. Explain about Summary() function in EDA
3. Explain about Describe() function in EDA
4. Explain about seaborn pairplot()
5. Explain about seaborn heatmap()
6. Explain about scipy describe() function
7. Explain about matplotlib boxplot()
8. Explain about matplotlib histogram()
9. Explain numpy mean()
10. Explain numpy std()