



# INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043

## COMPUTER SCIENCE AND ENGINEERING

### TUTORIAL QUESTION BANK

Course Name	DATA PREPARATION AND ANALYSIS
Course Code	BCSB13
Class	M.Tech II Semester
Branch	CSE
Year	2018-2019
Course Coordinator	Ms. G Sulakshana, Assistant Professor, CSE Dept

#### COURSE OBJECTIVES (COs):

The course should enable the students to:

I	Develop strategies for dealing with imperfect real world data
II	Read data from databases and clean the data for statistical analysis in SAS.
III	Prepare data marts for statistical analysis using SAS software.
IV	Convert the data for analysis and develop meaningful Data Visualizations.

#### COURSE LEARNING OUTCOMES (CLOs):

Students, who complete the course, will have demonstrated the ability to do the following:

BCSB13.01	Identify and understand the difference between data and information with formats.
BCSB13.02	Gain knowledge to identify the data parsing and transformations.
BCSB13.03	Describe the fundamentals of scalability with a desired real time issues.
BCSB13.04	Explain the basic concept of data cleaning for valuable information with a minimum consistency checking.
BCSB13.05	Learn data transformations and segmentation to solve statistical problems.
BCSB13.06	Understand statistical exploratory analysis with hypothesis generation.
BCSB13.07	Distinguish Clustering and association and apply them in solving statistical problems.
BCSB13.08	Designing visualizations for exploratory analysis.
BCSB13.09	Understand the concept of correlations and connections for geo located data.
BCSB13.10	Visualize the basic hierarchies in a network for interactivity.

## TUTORIAL QUESTION BANK

UNIT – I			
DATA GATHERING AND PREPARATION			
PART – A (SHORT ANSWER QUESTIONS)			
S. No	Question	Blooms Taxonomy Level	Course Learning Outcome (CLOs)
MODULE – I INTRODUCTION			
1.	Identify what is Data?	Remember	BCSB13.01
2.	Identify what is Information?	Remember	BCSB13.03
3.	Find out the actual information from data?	Remember	BCSB13.02
4.	List out the various data formats?	Understand	BCSB13.03
5.	Explain what is parsing?	Remember	BCSB13.02
6.	List out the parsing techniques?	Understand	BCSB13.02
7.	Find out what is transformation?	Understand	BCSB13.02
8.	Find out what is Scalability?	Remember	BCSB13.05
9.	List out various scalability techniques?	Understand	BCSB13.05
10.	Identify the real time issues in data preparation?	Understand	BCSB13.06
PART – B (LONG ANSWER QUESTIONS)			
1.	Explain the comparison between data gathering and data preparation?	Understand	BCSB13.01
2.	Distinguish the difference between data and Information?	Understand	BCSB13.03
3.	Identify different data formats in data preparation?	Remember	BCSB13.04
4.	Explain briefly about parsing techniques?	Understand	BCSB13.03
5.	Explain in detail about data transformation and its methods?	Remember	BCSB13.03
6.	Write the functionality of structured, unstructured and semi-structured?	Understand	BCSB13.02
7.	List out the various data formats with different examples?	Remember	BCSB13.02
8.	Explain in detail about scalability?	Understand	BCSB13.05
9.	List out various scalability methods in preparation?	Remember	BCSB13.05
10.	Write a brief note on various scalability in real time issues?	Remember	BCSB13.06
PART – C (CRITICAL THINKING QUESTIONS)			
1.	List out what technology might help us reduce bottlenecks to the conversion and gathering of information to make it more accessible?	Remember	BCSB13.01
2.	Explain how can we change our processes and technology to eliminate or cut across silos of information?	Understand	BCSB13.03
3.	Study how to learn a semantic parser of state- of-the-art accuracy with less supervised training data?	Remember	BCSB13.04
4.	Write what is more crucial for the system: fault tolerance or high performance?	Understand	BCSB13.03
5.	Write Where does the need for scalability appear?	Understand	BCSB13.03
MODULE-II			
DATA CLEANING			
PART – A (SHORT ANSWER QUESTIONS)			
1.	Identify what is data cleaning?	Remember	BCSB13.02
2.	List out the sources of errors?	Remember	BCSB13.02
3.	Describe screening data?	Remember	BCSB13.02
4.	Write what is missing values?	Remember	BCSB13.02
5.	Find out what is imputation?	Remember	BCSB13.03

6.	List out various data integration methods?	Remember	BCSB13.02
7.	Interpret what is data transformation?	Remember	BCSB13.04
8.	Identify what is segmentation?	Understand	BCSB13.04
9.	Describe hierarchical segmentation?	Understand	BCSB13.05
10.	Identify the consistency for data?	Understand	BCSB13.03
<b>PART – B (LONG ANSWER QUESTIONS)</b>			
1.	Explain in detail about data cleaning?	Understand	BCSB13.02
2.	List out the various data cleaning tools?	Understand	BCSB13.03
3.	What is the difference between heterogeneous and homogeneous data sets?	Understand	BCSB13.04
4.	Find in dealing with missing values in heterogeneous data using k-nearest neighbors?	Understand	BCSB13.04
5.	Explain briefly about consistency checking after data cleaning technique?	Remember	BCSB13.05
6.	Find out missing values after consistency checking?	Remember	BCSB13.05
7.	Explain about data transformation in detail?	Remember	BCSB13.06
8.	State the overview of selected segmentation approaches?	Remember	BCSB13.06
9.	Comparison of segmentation methods based on actual data?	Understand	BCSB13.06
10.	Explain various segmentation methods?	Understand	BCSB13.06
<b>PART – C (CRITICAL THINKING QUESTIONS)</b>			
1.	State data transformation functions for expanded search spaces in geographic sample supervised segment generation?	Understand	BCSB13.02
2.	State out the implications for marketing and research?	Remember	BCSB13.03
3.	State that can we transform or clean up old data to new formats?	Understand	BCSB13.03
4.	Solve the missing value estimation and impact on heterogeneous data set?	Understand	BCSB13.04
5.	State the heterogeneous missing data in real time issues?	Remember	BCSB13.05
<b>MODULE– III</b>			
<b>EXPLORATORY ANALYSIS</b>			
<b>PART – A (SHORT ANSWER QUESTIONS)</b>			
1.	Identify what is exploratory analysis?	Remember	BCSB13.03
2.	Find out the terminology of analysis in business environment?	Remember	BCSB13.05
3.	Find out the terminology of analytics in business environment?	Understand	BCSB13.04
4.	State the term statistics and its methods?	Remember	BCSB13.05
5.	Write down about descriptive statistics?	Understand	BCSB13.04
6.	Write down about comparative statistics?	Remember	BCSB13.05
7.	Explain what is clustering?	Understand	BCSB13.05
8.	List out the various types of clustering?	Remember	BCSB13.05
9.	Explain about association in analysis?	Understand	BCSB13.06
10.	Write down about hypothesis generation?	Understand	BCSB13.06
<b>PART – B (LONG ANSWER QUESTIONS)</b>			
1.	Explain what are the similarities between descriptive and comparative statistics?	Remember	BCSB13.05
2.	Explain what are the four types of descriptive statistics?	Understand	BCSB13.06
3.	Solve the most appropriate strategy for data cleaning before performing clustering analysis?	Remember	BCSB13.06
4.	Explain how can clustering be used to improve the accuracy of linear regression model?	Understand	BCSB13.07
5.	State which are valid iterative strategy for treating missing values before clustering analysis?	Remember	BCSB13.07
6.	Explain what are the association rules and the apriori algorithm?	Understand	BCSB13.06

7.	Identify what are the pre-requisites for generating an association rule?	Remember	BCSB13.06
8.	Explain about the statistical hypothesis and its types?	Remember	BCSB13.06
9.	Write down the four main steps to follow the process of hypothesis testing?	Understand	BCSB13.07
10.	Explain in brief about the hypothesis testing?	Understand	BCSB13.07
<b>PART – C (CRITICAL THINKING QUESTIONS)</b>			
1.	Solve which clustering algorithms suffers from the problem of convergence at local optima?	Understand	BCSB13.07
2.	State what is confidence in association rule mining?	Understand	BCSB13.06
3.	Demonstrate how to run the k-means clustering algorithm in analysis?	Understand	BCSB13.06
4.	Determine the process of statistics, whether to reject a null hypothesis based on sample data?	Remember	BCSB13.07
5.	Explain about the Directed Acyclic Graph[DAG] for code generation by hypothesis?	Remember	BCSB13.07
<b>MODULE– IV</b>			
<b>VISUALIZATION-1</b>			
<b>PART – A (SHORT ANSWER QUESTIONS)</b>			
1.	Identify what is visualization?	Remember	BCSB13.05
2.	Implement how to design a visualization?	Remember	BCSB13.06
3.	List out various types of visualization?	Understand	BCSB13.07
4.	Write down what is geospatial data?	Understand	BCSB13.07
5.	Solve how to geo code your data?	Remember	BCSB13.06
6.	Explain about time series analysis?	Understand	BCSB13.07
7.	Identify the term time series regression?	Remember	BCSB13.07
8.	Find out the term indexing data?	Understand	BCSB13.06
9.	Identify how do you find correlation of data?	Remember	BCSB13.06
10.	Does correlation imply linear relationship?	Understand	BCSB13.07
<b>PART – B (LONG ANSWER QUESTIONS)</b>			
1.	Explain about the benefits of data visualization?	Understand	BCSB13.07
2.	Explain about the visualization techniques in data mining?	Remember	BCSB13.06
3.	Determine the factors that influence data visualization choices?	Understand	BCSB13.07
4.	Explain the basic method of hierarchical data visualization?	Remember	BCSB13.06
5.	Write in detail about network data models?	Remember	BCSB13.07
6.	Explain the goals in time series analysis?	Understand	BCSB13.06
7.	Explain about the descriptive analysis under time series?	Remember	BCSB13.07
8.	Explain what are data correlations and why is it important to identify correlations?	Remember	BCSB13.06
9.	Identify the basic purpose of regression analysis?	Understand	BCSB13.07
10.	Differentiate the difference between correlation and regression analysis?	Understand	BCSB13.06
<b>PART – C (CRITICAL THINKING QUESTIONS)</b>			
1.	Explain about the purpose of using a dashboard for data visualization?	Understand	BCSB13.07
2.	Explain how does visualization of big data help in interpreting information?	Remember	BCSB13.07
3.	Determine what does correlation tell us about two variables?	Understand	BCSB13.06
4.	Empower the basic tools for complex data visualization and analytics?	Remember	BCSB13.07
5.	Explain about the big data visualization and analytics on elastic stack?	Understand	BCSB13.06
<b>MODULE– V</b>			
<b>VISUALIZATION-1</b>			
<b>PART – A (SHORT ANSWER QUESTIONS)</b>			
1.	Identify what is data visualization and its techniques?	Remember	BCSB13.07
2.	Is eclipse a data visualization tool?	Remember	BCSB13.08
3.	Write down the best data visualization tools?	Understand	BCSB13.07

4.	Identify the term data visualization dashboard?	Understand	BCSB13.09
5.	List out the diagrams and matrices for visualization?	Remember	BCSB13.08
6.	Write down about multidimensional data visualizations?	Understand	BCSB13.09
7.	Identify what is network visualization?	Remember	BCSB13.08
8.	Explain how social network analysis used for?	Understand	BCSB13.10
9.	Find out the term centrifuge?	Remember	BCSB13.08
10.	Write down about supervised learning?	Remember	BCSB13.10
<b>PART – B (LONG ANSWER QUESTIONS)</b>			
1.	Explain what do you mean by network analysis?	Understand	BCSB13.09
2.	Explain about the novel method for multivariate data projection and structure visualization?	Remember	BCSB13.08
3.	Implement how the functionality of data visualization to interactive data analysis?	Understand	BCSB13.10
4.	Explain about the three main uses of data visualization?	Remember	BCSB13.09
5.	Write down how does interactive data analysis work?	Understand	BCSB13.10
6.	Explain what are the important aspects of data analysis?	Remember	BCSB13.09
7.	Write down about quo vadis interaction?	Understand	BCSB13.09
8.	Explain what are the challenges of interactive visual data analysis?	Remember	BCSB13.10
9.	Explain about the multiple analysis perspectives?	Understand	BCSB13.10
10.	Write in detail about simple spectral analysis?	Remember	BCSB13.09
<b>PART – C (CRITICAL THINKING QUESTIONS)</b>			
1.	Identify why is text added to a data visualization presentation?	Understand	BCSB13.10
2.	Visualize and analyze hierarchical neural projections for data mining?	Understand	BCSB13.10
3.	Visualize the method of hybridizing multidimensional scaling and self organizing map?	Remember	BCSB13.10
4.	Interpret how fancy a visualization need to be in order to be useful for data analysis?	Understand	BCSB13.09
5.	Evaluate the interpretation for various visualizations and models?	Understand	BCSB13.10

**Prepared by:**  
Ms. G Sulakshana, Assistant Professor, CSE Dept.

**HOD, CSE**











---



---





















































