Department of Commerce

Osmania University

Computer Lab – Practical Question Bank

B.Com (Computer Applications)

Semester VI

Data Analytics

Time: 60 Minutes

Record	: <u>10</u>
Skill Test	: <u>15</u>
Viva - Voce	: 10
Total Marks	: 35

Use Excel to answer the following questions.
Download Telco Customer Churn data from Kaggle to answer the following questions.

1. Give a detail description of the data and anticipate what factors are influencing customer churn value.

2. Extract data for customers with Churn Value = 1 and create a frequency distribution for the contract type.
Compute the mean and variance of Monthly charges for customers who have churned and those who have not

3. Calculate the sampling distribution of the variance of monthly charges for churned customers. Use samples of size 50. Visualize the results using a histogram.

4. Perform stratified sampling based on churn value and state. For each stratun, compute the median CLTV. Visualize these medians using a bar chart.

5. Develop a function in Excel to randomly select 20% of the customers who have churned the churn value = 1 . Compute the mean Monthly charges and compare it to the population mean using statistical inference.

6. Analyze the relationship between Churn score and Tenure months using quadratic regression model. Evaluate the residuals and calculate the prediction error.

7. Implement resampling (eg., bootstrapping) to estimate the average Tenure Months for churned customers.

8. Test if the mean Total charges for churned and non-churned customers differs significantly using a t-test.

9. Create a pivot table showing the average churn score for customers grouped by internet setvice and payment method.

10. Analyze the relationship between Monthly charges and churn value by plotting a box plot and interpreting the results.

11. Build a histogram of CLTV and calculate the skewness and kurtosis.

12. Simulate random samples of size 100 from the total charges column and compute the sampling distribution of the mean.

13. Using the resamples data from Q.7, calculate the 95% confidence interval for the mean Tenure Months.

14. Create a scatter plot of tenure months Vs. Monthly charges and calculate the correlation coefficient.

15. Use Linear Regression to predict total charges based on tenure months and calculate the prediction error.

16. Analyze the impact of internet service on churn value using logistic regression.

17. Compare the variance of Monthly charges for customers with and without paperless billing using F-test.

18. Implement a resampling based hypothesis test to determine if the churn value distribution is independent of Gender.

19. Build a prediction model in excel to predict Churn value using Tenure Months, Monthly Charges, and CLTV. Evaluate the model's accuracy using cross validation.

20. Create a pivot table to analyze the average churn score and CLTV by state and internet service. Perform hypothesis testing to compare CLTV across states.

21. Create a pivot table to calculate the total and average Monthly charges for each internet service type. Highlight the highest and lowest averages.

22.   Select any company's stock prices. Download five Year data.
   a. Analyse the stock data by calculating key descriptive statistics.
   b. Give all possible visualizations.
   c. What do these statistics reveal about the stock's performance over the selected period.

23.   Plot a stock's closing price over time. Identify any visible trends, seasonality, or volatility in the data. Are there periods of rapid growth, decline or stability?

24.   Calculate the daily percentage change in stock prices and visualize it. Identify periods of high and low volatility. How does the stock's volatility compare to its average daily return?

25.   Analyse the relationship between stock price movements and trading volume. Does a high trading volume correlate with significant price changes? Provide examples from the dataset.

26.   Simulate a stock price stream for a selected company. Create a program that processes one data point (Price) at a time.

27.   Build a real time dash board showing the current stock price, average price, and price change percentage.

28.   Use a rolling window of past stock prices to predict the next price using linear regression.

29.   For the stock data selected, measure the past price fluctuations to estimate risk.

30.   Plot the stock's daily closing prices over the last five years. Give your insights on
   a. Identify overall trends
   b. Detect periods of volatility or stability

31.   Overlay 3-day and 20day moving averages on the closing price chart and predict the next few days price.

32.   Plot daily trading volume alongside stock prices. Detect the relationships between price changes and trading volume.

33.   Using a data set of grocery transactions, identify frequent item sets with a minimum support threshold of 0.2. Generate association rules with

confidence above 0.7.

Use Market Basket Analysis (MBA) from Kaggle to answer the following.

34. Calculate and visualize the frequency of individual items purchased in a data set. Rank the top 10 items by frequency.

35. Identify the top 10 most frequently purchased items in the data set. Visualise the results using a bar chart.

36. Analyse which day of the week has the highest number of transactions.

37. Using the Apriori algorithm, find frequent item sets with minimum support of 0.05. Generate Association rules with confidence above 0.7.

38. Perform market basket analysis for transactions on weekends (Saturday and Sunday) Only. Compare frequent item sets with weekdays.

39. Identify pairs of items that are frequently purchased together. Visualize them using a heat map.

40. Analyse which months have the highest sales for a selected item.

41. Cluster customers based on the number of transactions they made and the diversity of items purchased. Visualize clusters using a scatter plot.

42. Identify customers with unusual purchase behaviour. (eg., purchasing only a single item multimer times). Highlight these outliers in the clustering results.

43. List the steps involved to install Hadoop on your local machine or set it up using a virtual machine.

44. Prepare a Presentation on installation and interface of Hadoop Distributed File System.
Explain how to create a directory, upload a file and view its replication details.

45. Prepare a presentation on Apache Hive and the steps to connect it to HDFS.

46. Write the steps to install MongoDB on your machine. Explain its interface to create a database, insert data and retrieve it.

47. Write steps to create an Amazon S3 bucket. Write steps to upload a file, set permissions.

48. Explain the concept of Sharding in NoSQL databases. Illustrate with an example.

49. With a presentation show the types of visualizations possible in excel.

50. With a presentation, who the types of visualizations possible in Python.