PLACEMENTDOST

Note:

Please be advised that these assessment inquiries are designed to accommodate interns with diverse skill levels, ranging from novices to seasoned analysts. Should you encounter any challenging questions, you are encouraged to seek solutions independently or reach out to us for assistance at intern@placementdost.com. Best wishes for success in completing the assessment!

Tools needed :- Hadoop, Sparks, Excel

1. Load the Customer Flight Activity, Customer Loyalty History, and Calendar tables into HDFS (Hadoop Distributed File System) and Spark DataFrames. Inspect the structure of each table.

2. Perform data cleaning on the Customer Flight Activity and Customer Loyalty History tables, addressing issues such as missing values, duplicates, and outliers using Spark transformations.

3. Conduct exploratory data analysis (EDA) using Spark to analyze the distribution of Total Flights, Distance, and Points Accumulated. Visualize the trends over the years.

4. Create a new column in the Customer Loyalty History table representing the ratio of Points Redeemed to Points Accumulated using Spark transformations.

5. Calculate the average Customer Lifetime Value (CLV) for customers in each Enrollment Type using Spark RDDs or DataFrames.

6. Analyze temporal trends in Total Flights and CLV over the given period using Spark time-series processing.

7. Merge the Customer Flight Activity and Customer Loyalty History tables using Spark to create a comprehensive dataset.

8. Explore the distribution of Marital Status and Loyalty Card status among customers using Spark SQL.

9. Examine the correlation between Salary and CLV using Spark SQL and DataFrame API.

10. Aggregate Total Flights and Points Redeemed at a monthly level using Spark and analyze trends.

11. Create a geographical map showing the distribution of customers based on their City using Spark SQL and GeoSpark.

12. Generate a heatmap to visualize the relationship between Enrollment Type and Marital Status using Spark.

13. Perform hypothesis testing using Spark to compare the CLV means between customers with different levels of education.

14. Apply clustering algorithms using Spark MLlib to segment customers based on their spending patterns (CLV).

15. Conduct an analysis of variance (ANOVA) using Spark to assess the impact of Enrollment Type on Points Redeemed.

16. Build a regression model using Spark MLlib to predict CLV based on relevant features.

17. Reshape the data to a long format for better visualization and analysis using Spark.

18. Create an interactive dashboard that displays key insights from the dataset using Spark and a suitable visualization tool.

19. Craft a data story summarizing the key findings, insights, and recommendations based on Spark analysis.

20. Optimize the Spark jobs for better performance, considering partitioning, caching, and other optimization techniques.