***IBM***

***NAAN MUDHALVAN***

***APPLIED DATA SCIENCE***

***PROJECT TITLE: Future Sales***

***Prediction***

***REG.NO : 712221205013***

***NAME :Madhumitha.K***

***DEPT : INFORMATION TECHNOLOGY***

***YEAR & SEM : III & 05***

***COLLEGE : PARK COLLEGE OF ENGINEERING AND TECHNOLOGY***

***Phase 2***

***Project definition:***

*** The problem is to implement data science techniques to segment customers based on their behavior, preferences, and demographic attributes. The goal is to enable businesses to personalize marketing strategies and enhance customer satisfaction. This project involves data collection, data preprocessing, feature engineering, clustering algorithms, visualization, and interpretation of results.***

***Flow chart:***

*Start*

*Data collection*

*Data processing*

*End*

*Decision making*

*Result Analysis*

*Data splitting*

*Feature selection*

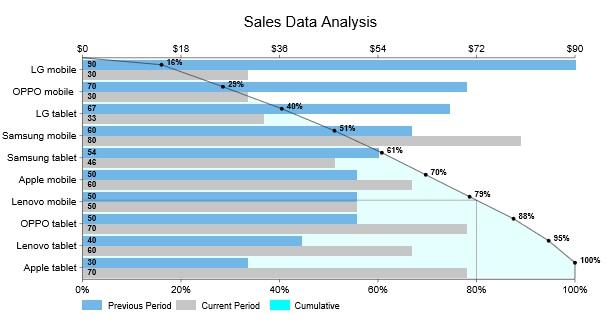
*Model evaluation*

*Prediction*

*Model selection*

***Data Set:***

[***https://in.docworkspace.com/d/sIPuvxMfQAcvMmakG***](https://in.docworkspace.com/d/sIPuvxMfQAcvMmakG)

******

***Project overview:***

***1. \*\*Start\*\*: Begin the process of sales prediction.***

***2. \*\*Data Collection\*\*:***

***- Collect historical sales data from the database.***

***- Gather additional relevant data such as market trends, customer behavior, and economic indicators.***

***3. \*\*Data Preprocessing\*\*:***

***- Clean the data to remove duplicates and errors.***

***- Handle missing values and outliers appropriately.***

***- Transform categorical data into numerical values if necessary.***

***4. \*\*Feature Selection\*\*:***

***- Identify the relevant features (variables) that influence sales.***

***- Select features based on their correlation with sales data.***

***5. \*\*Data Splitting\*\*:***

***- Divide the dataset into training and testing sets.***

***- Typically, 70-80% of the data is used for training and the rest for testing.***

***6. \*\*Model Selection\*\*:***

***- Choose an appropriate predictive model (e.g., regression, time series analysis, machine learning algorithms) based on the nature of your data.***

***- Train the selected model using the training data.***

***7. \*\*Model Evaluation\*\*:***

***- Evaluate the model's performance using the testing data.***

***- Use metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) to assess the accuracy of the model.***

***8. \*\*Prediction\*\*:***

***- Use the trained and validated model to predict future sales based on new data inputs.***

***9. \*\*Result Analysis\*\*:***

***- Analyze the predicted sales data.***

***- Compare predictions with actual sales figures to assess the model's accuracy.***

***10. \*\*Decision Making\*\*:***

***- Make informed business decisions based on the predicted sales data.***

***- Adjust strategies, marketing efforts, or inventory levels as needed.***

***11. \*\*End\*\*: Complete the process of sales prediction.***

***Please note that the above flowchart provides a general overview.***