BUDGET SALES ANALYSIS

Siddhartha Borgohain

Sachin Verma

OBJECTIVE

- Extract various information such as Sales, budget, variance.

- comparing sales, profit, income and budget with various attributes.

- Extract necessary information about Products and Customers.

- Do ETL : Extract, transform and Load the data

- Use various visualization and features and make the best dashboard.

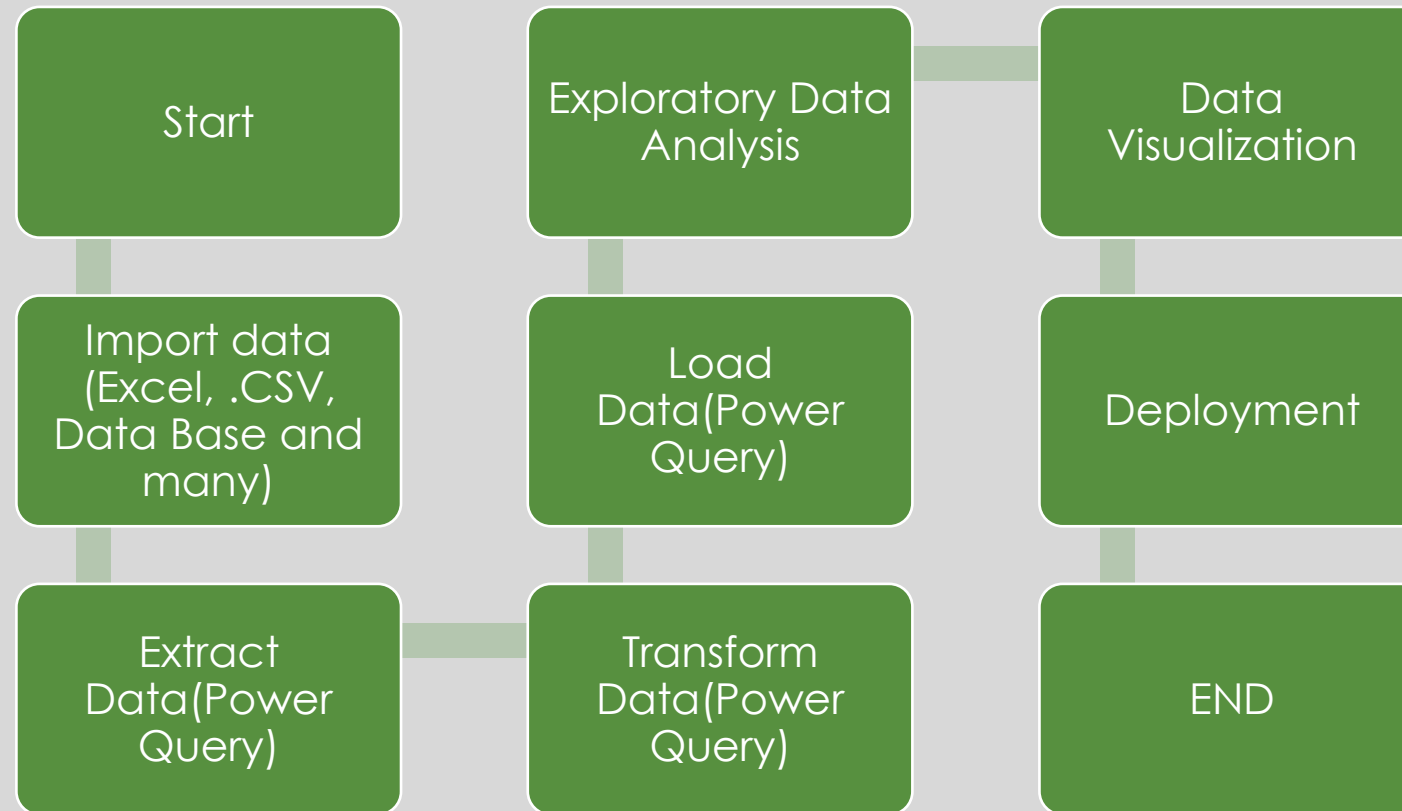- Find key metrics and factors and show the meaningful relationships between attributes.

Benefits

◦ Analyzing the Revenue and Sales of various items.

◦ Determining the budget in the year of 2016

◦ Extracting the flows and achievements

◦ Detecting the trends of Profit in each Fiscal years

Data Sharing Agreement :

• Sample file name

• Length of date stamp(8 digits)

• Length of time stamp(6 digits)

• Number of Columns

• Column names

• Column data type

Architecture

Data Validation and Data Transformation :

◦ Name Validation - Validation of files name as per the DSA. We have created a

◦ regex pattern for validation. After it checks for date format and time format if

◦ these requirements are satisfied, we move such files to "Good_Data_Folder" else

◦ "Bad_Data_Folder."

◦ Number of Columns – Validation of number of columns present in the files, and if

◦ it doesn't match then the file is moved to "Bad_Data_Folder."

◦ Name of Columns - The name of the columns is validated and should be the same

◦ as given in the schema file. If not, then the file is moved to "Bad_Data_Folder".

◦ Data type of columns - The data type of columns is given in the schema file. It is

◦ validated when we insert the files into Database. If the datatype is wrong, then the

◦ file is moved to "Bad_Data_Folder".

◦ Null values in columns - If any of the columns in a file have all the values as

◦ NULL or missing, we discard such a file and move it to "Bad_Data_Folder".

Data Insertion in Database:

◦ Table creation :- Table name "t_motorpv_fraud" is created in the database for

◦ inserting the files. If the table is already present then new files are inserted in the

◦ same table.

◦ Insertion of files in the table - All the files in the "Good_Data_Folder" are inserted

◦ in the above-created table. If any file has invalid data type in any of the columns,

◦ the file is not loaded in the table

Q & A:

◦ Q1) What's the source of data?

◦ The data for training is provided by the client in multiple batches and each batch contain multiple files

◦ Q 2) What was the type of data?

◦ The data was the combination of numerical and Categorical values.

◦ Q 3) After the File validation what you do with incompatible file or files which didn't pass the validation?

◦ Files like these are moved to the Achieve Folder and a list of these files has been shared with the client and we removed the bad data folder.

◦ Q 4) How logs are managed?

◦ We are using different logs as per the steps that we follow in validation and modeling like File validation log , Data Insertion , Data Visualization etc.

- Q 6) What techniques were you using for data pre-processing?

- Removing unwanted attributes

- Visualizing relation of independent variables with each other and output variables

- Checking and changing Distribution of continuous values

- Removing outliers

- cleaning data and imputing if null values are present.

- Converting categorical data into numeric values.

- Scaling the data

- Q 9) What are the different stages of deployment?

- When the model is ready we deploy it in Fire environment .Where SIT and UAT is performed over it . Once We get Sign off from Fire we deploy in Earth and UAT is performed over it. After getting the sign off from Earth we deploy in production