

Project Meeting Minutes 3 - Week 5, 6 & 7

Project Name: Multi Mart Revenue forecasting

Below is the summary of the analysis/visualization project we completed in 3rd semester. The project focused on visualizing and analyzing sales and revenue data of Multi Mart Retail store spanning the period of 2019 to 2023. The primary objective was to provide a comprehensive understanding of sales performance, revenue generation, and customer behavior analysis to support informed decision-making, particularly regarding the potential expansion of its Loyalty Card program into new regions.

Git Hub Project Repository: [DAB Grp7 Capstone Project](#)

Week 5,6 & 7 Minutes of meeting are as follows:

Week 5 & 6:

Ikram Patel & Sujata Biswas

- Converted below text columns to numerical using one hot encoding:
 - referralsource
 - responsetolastcampaign
 - preferredpaymentmethod
- Identified new correlation using the above converted columns.
- Made list of supervised regression models
- Created base model for Random Forest and Linear Regression
- Evaluated model performance using R2 score.
- Feature importance analysis to identify the features which are important for predicting target variable.

Alisha James & Gayathri Manju Jayasena Kurup

- Split the data to train and test set.
- Feature importance analysis to identify the features which are important for predicting target variable.
- Created models using Random Forest, Linear Regression and Lasso
- Evaluated and compared the three model's performance on test set using R2 score.
- Created new attribute purchases_category by categorizing the totalpurchases into three categories High/Medium/low.
- Identified correlation between the new attribute and existing columns.

Shrikant

- Analysis of dataset and code
- Discussion with team members to catchup on works done during the absence.

Group Work

Attached is the analysis document.

- [Grp7_Draft_Analysis_doc.docx](#)

Week 7

Date and Time	Location	Attendees	Professor
20th Feb'24 Tuesday 10:00 AM – 10:30 AM	In person meeting	Alisha James (0811919) Gayathri Manju Jayasena Kurup (0836679) Ikram Patel (0822315) Sujata Biswas (0832706) Shrikant Ayyalasomayajula (0808545)	Abiodun Sodiq Shofoluwe

Discussion: Following points were discussed as part of the above meeting

- Run models using all the features in dataset.
- Run models using only the top 4 or 5 features in dataset after doing feature analysis.
- Evaluate and compare models using R2, MSE, MAE and RMSE
- Research on Principal component Analysis (PCA)
- Perform Principal component Analysis (PCA) on the model as it will give a new dimension that will capture the variations in target and thus can be used for selecting the best features.
- compare outputs of PCA with feature importance as this comparison will help select the features.
- If the outcomes for above comparison are different then build different models for both
- Run 4-5 models using PCA and feature selection.
- use models like Gradient Boost
- Research on the library that compares all possible models at once.

Next Meeting schedule: 27th Feb'24, Tuesday