Project Proposal

Prepared for: Vahid Behzadan, PhD.
Prepared by: Ariga Tejasimha , Srimanikantha Arjun Karimalammanavar
Rahul Akkineni
4 November 2019

STATEMENT OF OBJECTIVE

The goal of our project is to provide the company to improve their business with insights like

- which products to promote,
- which products to place near other products,
- which products to stock
- which products to recommend to specific customers

We will recommend the company throughout the year by considering the dataset for 3 months of sales at 3 different locations.By considering on time series modelling of sales which has patterns of sales based on time and location analysis based on the store present which will give insight to company to knew it's target audience or customers.

BACKGROUND RESEARCH

Data exploration methods, statistical measures, significance testing and some of the algorithms in machine learning like Decision Tree, Regression Models and Bayesian network. we will study the weather conditions at the time line shown in the data set when we are considering the location analysis(Dec 2018 to Mar 2019) and holidays. There is no much of research needed other than learning the functionality and procedures of machine learning algorithms.

DATASET DETAILS

We will consider the below dataset which is present in kaggle.

https://www.kaggle.com/aungpyaeap/supermarket-sales

In this dataset there are 17 columns which is gives complete insight of super market sales for three months from December 2018 - March 2019.It has data of supermarket which is located in three different locations. It has columns like payment methods, types of customers , date of purchase, time of purchase, ratings of customers and product line which is categorised groups like food and beverages, fashion accessories and electronic accessories. The above-mentioned data set contains limited entries which might not be

sufficient for optimization of our model. We might look for other datasets with a better description of the sales of the products with better parameters.

APPROACH

By using the data exploration techniques first we will clean the data set. After that we will see the date and time of purchase of product and find the patterns of it thought out the 3 months and each and everyday. Date is important to consider as to see whether it is a holiday or not. Time is also important field to consider as the weather of the day will effect the purchase of product by the customer. We can also the time patterns when are the customers are highly active in purchase.Now we will the categorise the product line and do analysis which products needed to be promoted well and which products to be arranged. We will also Same follows by considering the location analysis. We will see the payment methods and analysis the customers. Finally we will build a regression model using these both columns or variables(Time series and location).By this we will recommend the company how to manage the inventory and products based on the 3 months analysis.

VALIDATION

Only validation and baseline condition here is considering the dataset is perfect and there are no errors in the columns.By studying the date and time columns from the dataset we will study the pattern of the particular product .Based on the patterns we will suggest the products and segregate the ascending order of the products to be promoted based on the sales.If sales are low for a product then promotion will be high and will be tagged or combined with related category. Based on the consumption we will also recommend the stock recommendations and suggestion based on the weather and holidays.When there is tentative holiday or weekend or long weekend.if there is any climatic warning or hazardous thing there what kind of product line should be loaded first.Is there any kind flu break at the particular location then we have keep some flu related thing over the counter.By taking these two sets and developing the model which will recommend the product to company.(detail description of procedure of what is happening and what will be happeing)

NEXT STEPS

The next steps include everything mentioned above. Currently only the plan of action has been formulated, no actual work has yet been done with the dataset. These steps include cleaning the data, exploring the data, identifying descriptive statistics, separating and categorising the columns and manipulating the methods. First 2 weeks we will clean the data set and finish the time series modelling set before the progress date and in process of gathering the data and research for location analysis. Once we are done with recommending based on time series modelling and location analysis then we will combine both and build a optimise model which will give the best results or recommendations to company.

BIBLIOGRAPHY

https://www.kaggle.com/aungpyaeap/supermarket-sales

https://blackboard.newhaven.edu/webapps/blackboard/content/listContent.jsp?course_id=_47382_1&content_id=_1544777_1&mode=reset

https://www.geeksforgeeks.org/machine-learning/

www.google.com

https://www.khanacademy.org/math/statistics-probability

https://docs.python.org/