FOOD DEMAND FORECASTING FOR A FOOD DELIVERY COMPANY

A PROJECT REPORT

by

A.Sai Saran(18BEE1085)

Sajahan Shaik(18BEC1175)

Abbaraju Lakshmi Chenna Kesava(18BEC1320)

Rahul Reddy Putluri(18BEC1234)

B.venkata kushwanth(18bec1114)

under the guidance of

FOOD DEMAND FORECASTING FOR AFOOD DELIVERY COMPANY

TABLE OF CONTENTS

INTRODUCTION

The success of a restaurant not only depends on taste, ambiance but also on service. The most important part among the services is serving fresh food. In order to provide this, the restaurants need to prepare food daily, this requires buying some of fresh self life food products every day. The major task that one would face in this will be predicting the quantity of products to be bought and prepared. It is very difficult to predict the number of orders in a given restaurant on a given day. A wrong prediction may end up purchasing and preparing less amount of food which will cause shortage or purchasing and preparing more which will lead to wastage of food.

So, predicting the exact demand is a challenge because of uncertainty and fluctuations in consumer demand. These variations ad fluctuations in demand may be because of price change, promotions, change in customer's preferences and weather changes. All these factors imply that some dishes are sold mostly during

limited period of time. Although we know that some regular seasonal pattern is expected, the features that predict these seasons are not directly observed. Thus, drops and rises in orders because of these seasonal changes are difficult to predict.

Demand forecasting is a key component to every growing online business. It is the process in which historical data is used to estimate the quantity of product customer will purchase. This prediction activity is used in many fields like retailing, food industry etc. Without proper demand forecasting processes in place,it can be nearly impossible to have the right amount of stock on hand at any given time. A food delivery service has to deal with a lot of perishable raw materials which makes it all the more important for such a company to accurately forecast daily and weekly demand. Too much inventory in the warehouse means more risk of wastage,and not enough could lead to out-of-stocks - and push customers to seek solutions from your competitors.

PROBLEM STATEMENT

Your client is a meal delivery company which operates in multiple cities.They have various fulfillment centers in these cities for dispatching meal orders to their customers. The client wants you to help these centers with demand forecasting for upcoming weeks so that these centers will plan the stock of raw materials accordingly. The replenishment of majority of raw materials is done on weekly basis and since the raw material is perishable,the procurement planning is of utmost importance.Secondly, staffing of the centers is

also one area wherein accurate demand forecasts are really helpful.Given the following information,the task is to predict the demand for the next 10 weeks(Weeks: 146-155) for the center-meal combinations in the test set:

- **Historical data of demand for a product-center combination(Weeks:1 to 145)**
- **Product(Meal) features such as category,sub-category,current price and discount**
- **Information for fulfillment center like center area, city information etc.**

DATA DICTIONARY

1. **Weekly Demand data (train.csv):** Contains the historical demand data for all centers, test.csv contains all the following features except the target variable
2. **fulfillment_center_info.csv:** Contains information for each fulfillment center
3. **meal_info.csv:** Contains information for each meal being served

PROPOSED SOLUTION

Our solution aims at building a Machine Learning Model which takes in the relevant input features(independent variables) and predicts the total demand for a particular food item. Since the output value or the dependent variables is a continuous value the model is based on regression algorithm, Random Forest Regressor. The datasets where merged and data cleaning was performed before training the model. Exploratory Data Analysis and Visualization for the dataset was done using Pandas, Matplotlib and Seaborn. Then the irrelevant columns or feature variables are dropped, while the required columns used are:

- **week**
- **center_id**
- **meal_id**
- **checkout_price**
- **base_price**
- **city_code**
- **region_code**
- **op_area**
- **category**
- **cuisine**

Based on the following independent variables, the dependent variable **num_orders** is predicted. Then the categorical variables are encoded using label encoding. After which the values are scaled down to a common range by **StandardScaler()**, followed by splitting the dataset into training and testing set with random selection of data points. Then the model is built using Random Forest Regressor and the evaluation is done on the basis of **RMSLE**

or Root Mean Squared Log Error. Evaluation metric followed is **100*RMSLE**, based on which we have selected the particular model after trying out multiple regression algorithms. The model is integrated to a web app, made using **HTML, CSS** at the Frontend and **Flask** as Backend. The web app is a full-fledged website containing multiple pages based on the Food Company, food Swipe.

LITERATURE SURVEY

https://www.sciencedirect.com/science/article/pii/S1878450X13000231

This conceptual study specifically aims at reviewing the critical managerial issues of menu, and demonstrating the conceptual structure of menu management. Based on the conceptual and empirical findings of menu literature, the major menu management issues are menu planning, menu pricing, menu designing, menu operating and menu development.

http://studentsrepo.um.edu.my/1529/3/CH_2.pdf Talks about the problem part and how excessive inventory yields to low profit.And how to handle inventory properly.Additionally, it also proposes a model to examine the mediating effects of customer satisfaction in the relationship between service quality and behavioral intentions, between food quality and behavioral intentions and between perceived value and behavioural intentions.

http://eprints.utar.edu.my/3448/1/fyp_IA_2019_WSJ_1506513.pdf

An overview of the restaurant ordering system and how automated systems can generate more profits. Especially when there are a lot

of customers, the customers might think their order has been forgotten if their food still has not yet been served in a long time. It will be good if there is an estimated time to prepare the food shown to the customers. Furthermore, some of the customers might want to change their food order. https://www.agilecodex.com/restaurant-management Software for Stock/inventory management, Menu/Recipes management, Sales Reporting, Product & material management, transactions, Table Management & Reservations etc. It has Inventory Control features apart from general customer management features. Costs $350(25368.40 Rupees).

https://www.technodg.com/xenias-restaurant-software.html A tool for one screen order, kitchen-order-printing, food cost calculations, restaurant accounting, and table management options. Mostly suits standalone/one off restaurants. Handles table management, billing and data flow

EXPERIMENTAL INVESTIGATION

Importing the Libraries

The first step of starting the analysis on the dataset is by importing the libraries numpy, pandas, matplotlib. Numpy is the numerical python library used for all sorts of mathematical calculations.Matplotlib is used for visualization of data. Pandas is used for data manipulation.

Importing the Dataset

The dataset that was downloaded from kaggle is first uploaded in a .csv format and then imported in the file.

Exploratory Data Analysis

The raw data is visualized by plotting graphs of different input characteristics and the solution is proposed based on this analysis. This is an important step in analyzing which category of algorithms(regression or classification) are to be applied on the model.

Data Cleaning

- The data after visualization is then checked to find out the missing data in the dataset , and then filling them by using different mode and mean methods as per requirement. For numerical values we use mean and for categorical values, mode is used.
- Encoding in done for converting categorical columns into binary elements for easy acceptability by the algorithm.

Splitting the Dataset

The dataset is then divided into two sets:

1. **Training Data-** It mainly comprises of 80% of the data. It is used to train the data to acquire desired output.

2. **Testing Data-** It comprises of the other 20% of the data that is tested on the model and on the basis of which accuracy of the model is predicted.

Model Building

The main work done here is of applying the different algorithms on the training set and checking the accuracy of the model on the basis of these algorithms. The evaluation is done using RMSE values.

We made models using the following algorithms:

1. **Linear Regression - 100*RMSLE = 156**
2. **Decision Tree Regressor - 100*RMSLE = 70**
3. **Random Forest Regressor - 100*RMSLE = 55**
4. **Lasso Regressor - 100*RMSLE = 154**
5. **Ridge Regressor - 100*RMSLE = 156**
6. **XGBoost Regressor - 100*RMSLE = 57(After Tuning)**
7. **KNN - 100*RMSLE = 57**
8. **Gradient Boost - 100*RMSLE = 100**
9. **ElasticNet - 100*RMSLE = 124**
10. **AdaBoost - 100*RMSLE = 259**

This was done to find the best model which produces the least possible RMSLE value, which was the criterion followed.

Hyperparameter Tuning

Hyperparameter tuning is **choosing a set of optimal hyper parameters for** a learning algorithm. A hyper parameter is a model argument whose value is set before the learning process begins. The key to machine learning algorithms is hyper parameter tuning.

Hyper parameter tuning was done one XGBoost and KNN Algorithms to find the optimal parameters to produce the best possible results.

Evaluation

RMSLE is used when **y has long tail distribution**, or we are interested in the ratio of true value and predicted value. 1 is added to avoid divergence when y is zero. It is an extension on Mean Squared Error (MSE) that is mainly used when predictions have large deviations, which is the case with this energy prediction competition. Values range from 0 up to millions and we don't want to punish deviations in prediction as much as with MSE.

The Root Mean Squared Log Error (RMSLE) can be defined using a slight modification on sklearn's mean_squared_log_error function, which itself a modification on the familiar Mean Squared Error (MSE) metric.

The formula for RMSLE is represented as follows:

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i+1)-\log(a_i+1))^2}$$

Where:

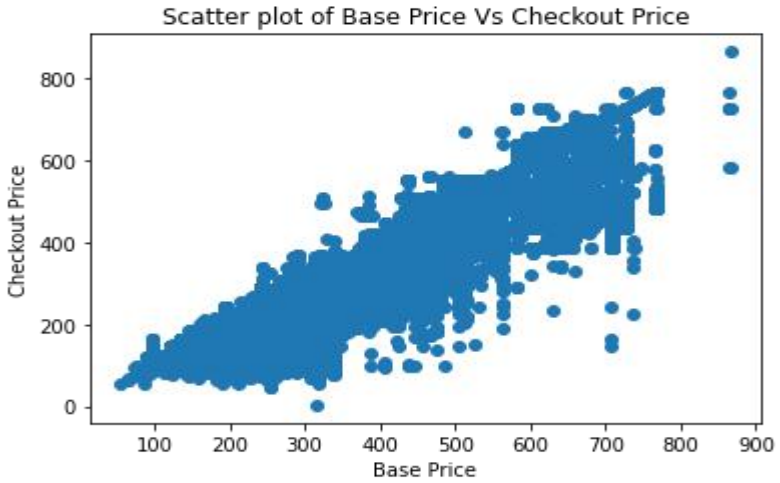
n is the total number of observations in the (public/private) data set, pi is your prediction of target, and ai is the actual target for i.

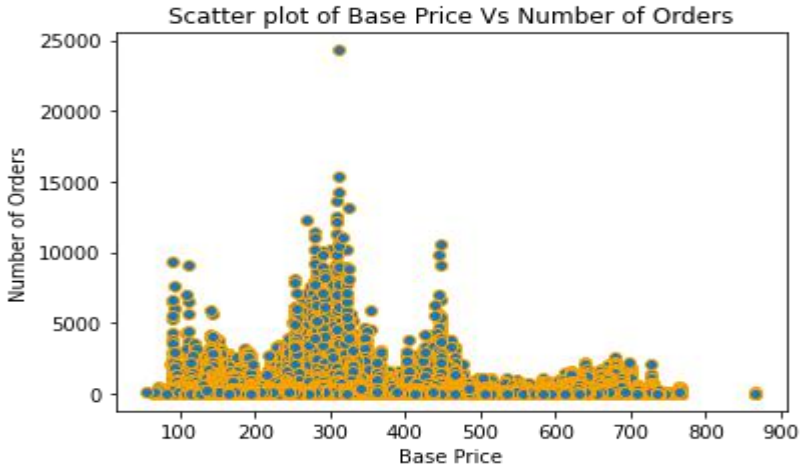
log(x) is the natural logarithm of x (loge(x).

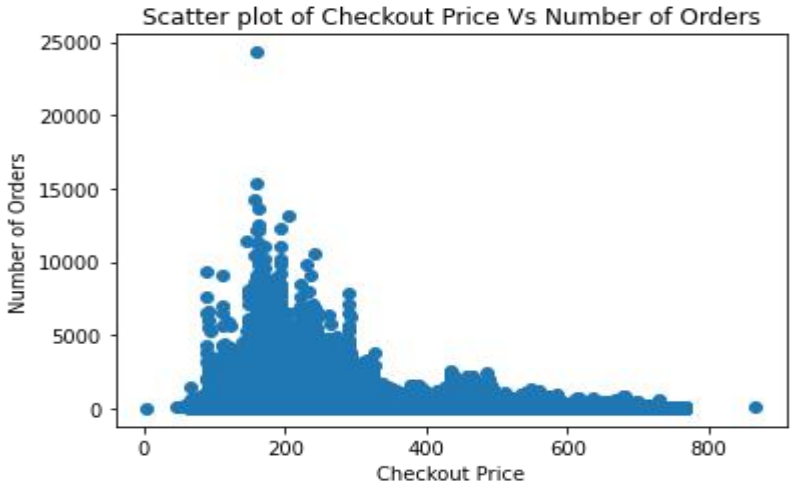
Saving the Model

Pickle is the standard way of serializing objects in Python. We can use the pickle operation to serialize your machine learning algorithms and save the serialized format to a file. Later we can load this file to deserialize your model and use it to make new predictions.

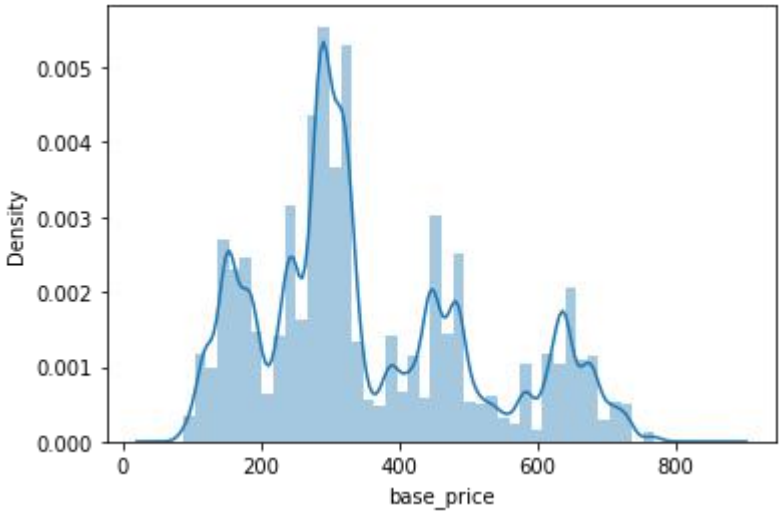
SCATTER PLOT

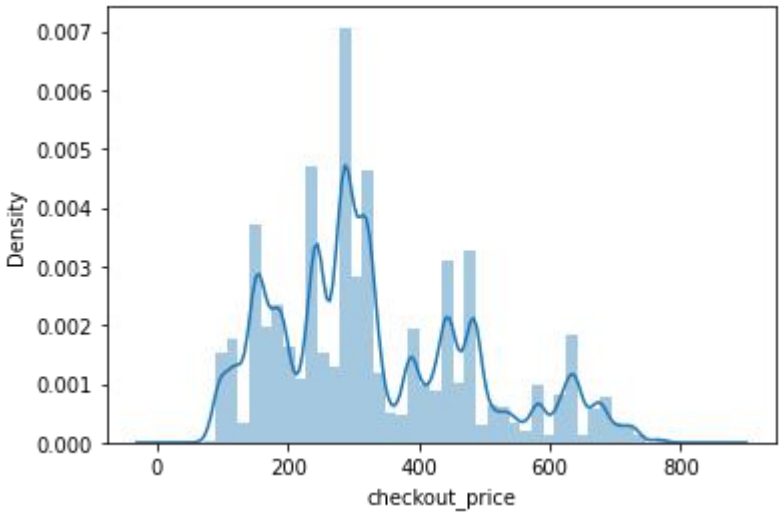


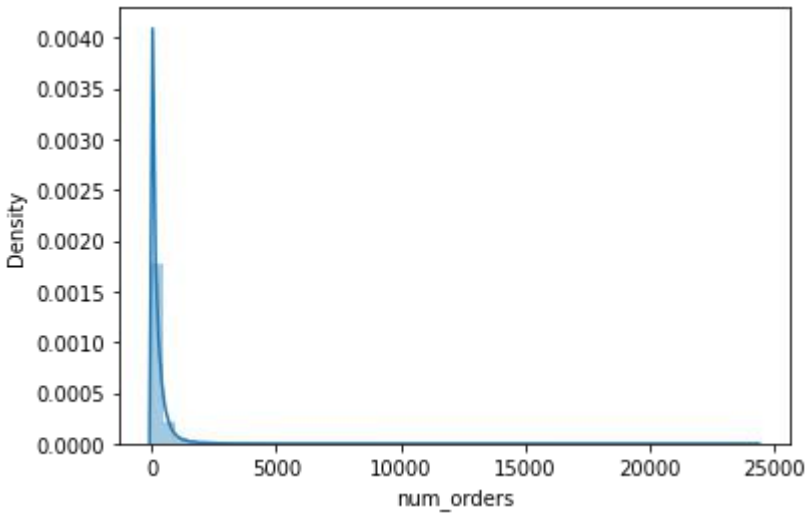




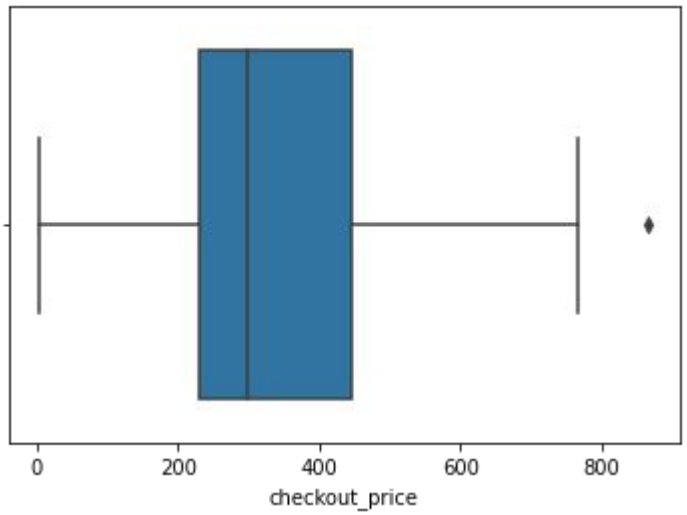
DISTPLOT

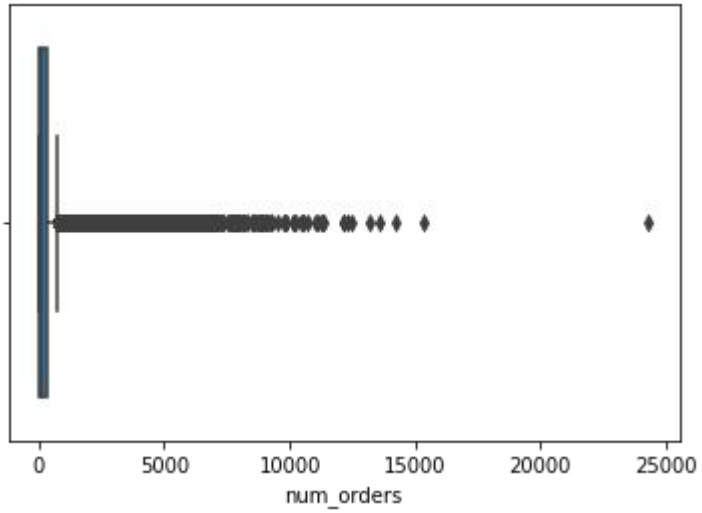


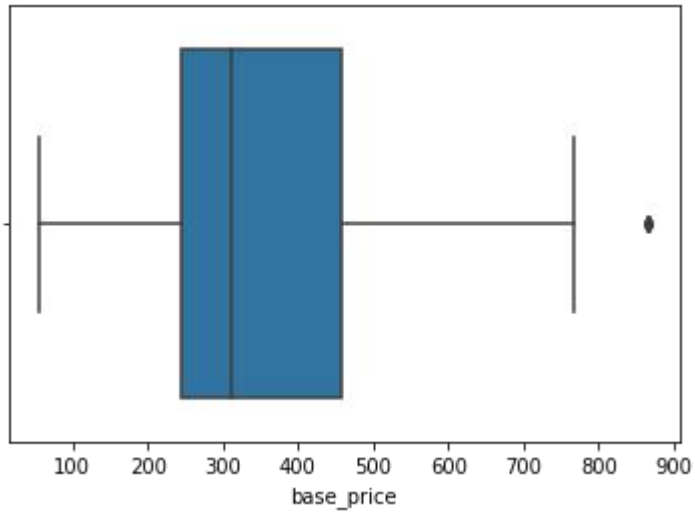




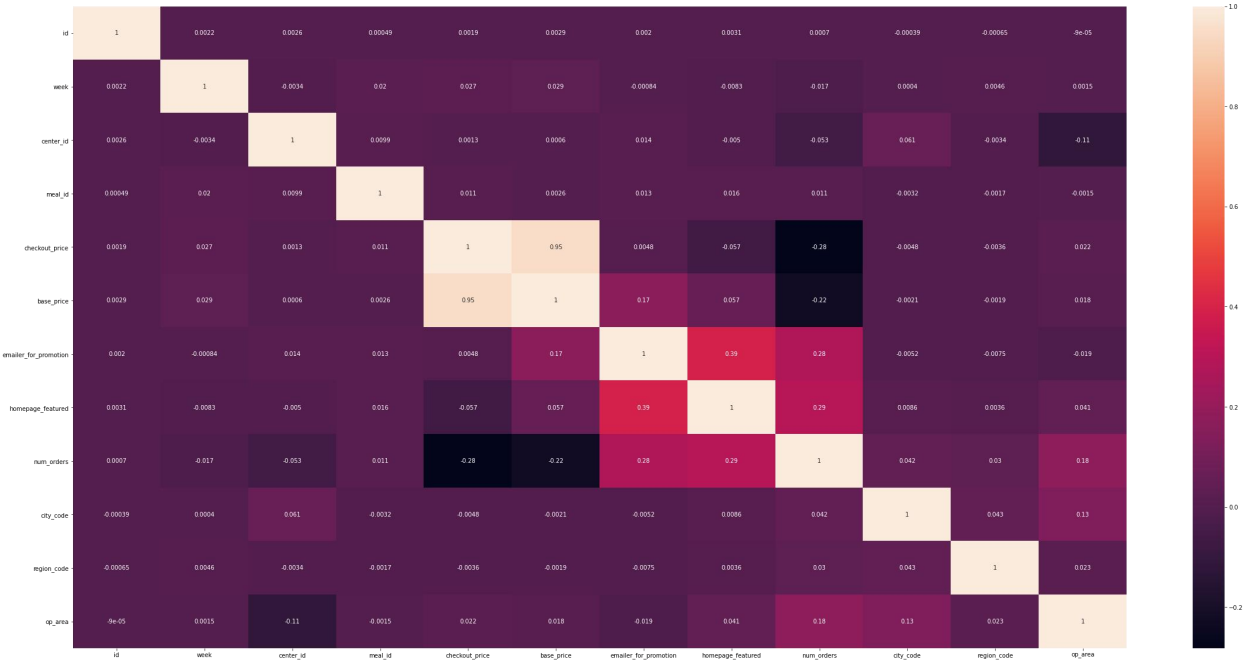
BOXPLOT


0
200
400
600
800
checkout_price

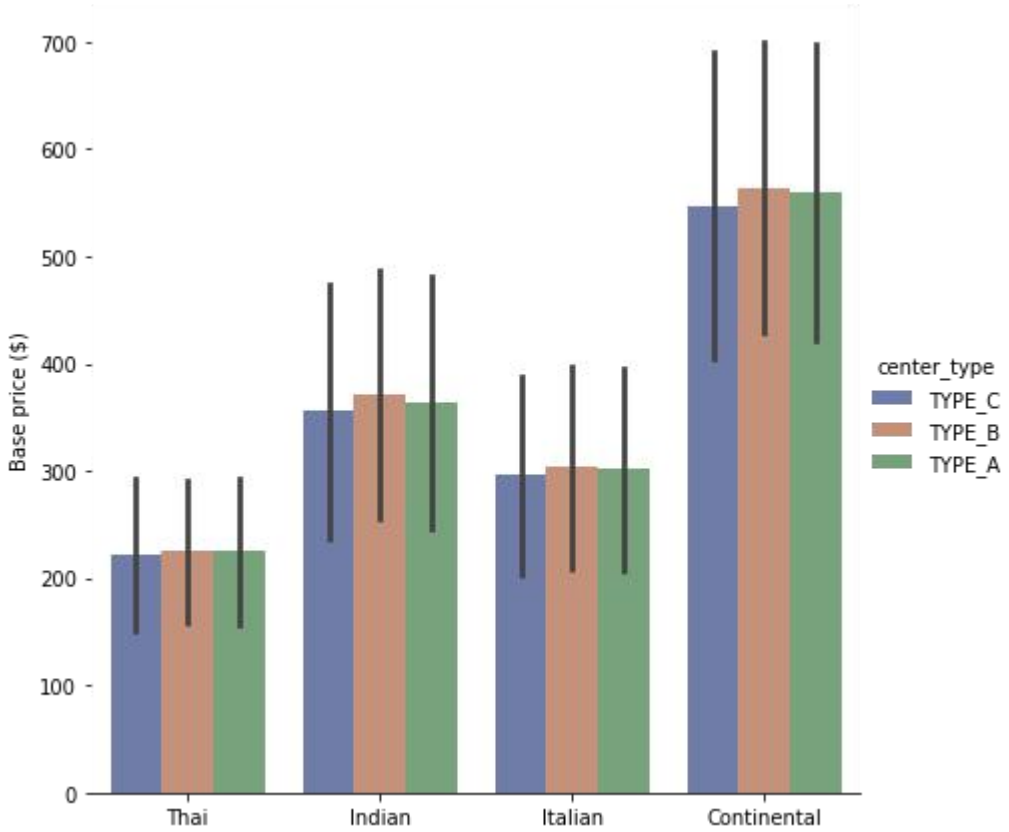

0
5000
10000
15000
20000
25000
num_orders

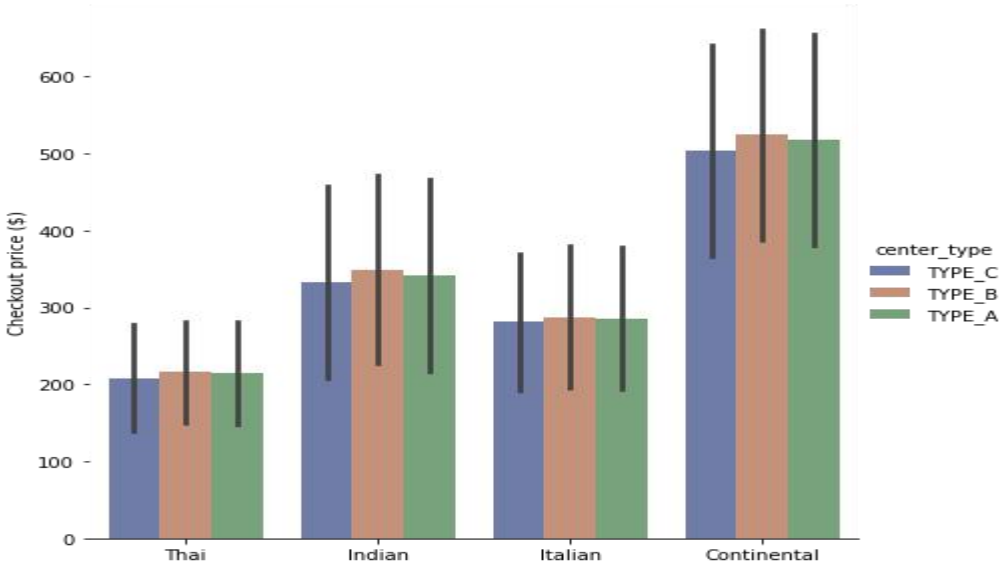

100
200
300
400
500
600
700
800
900
base_price

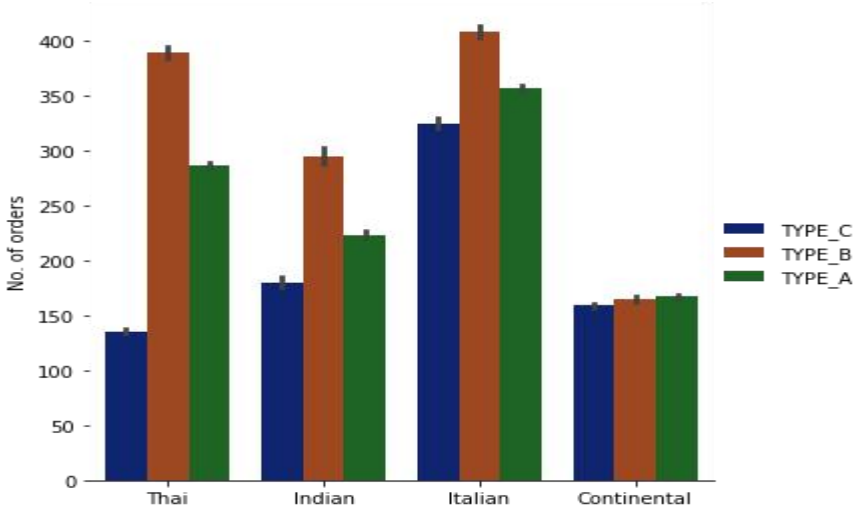
HEATMAP

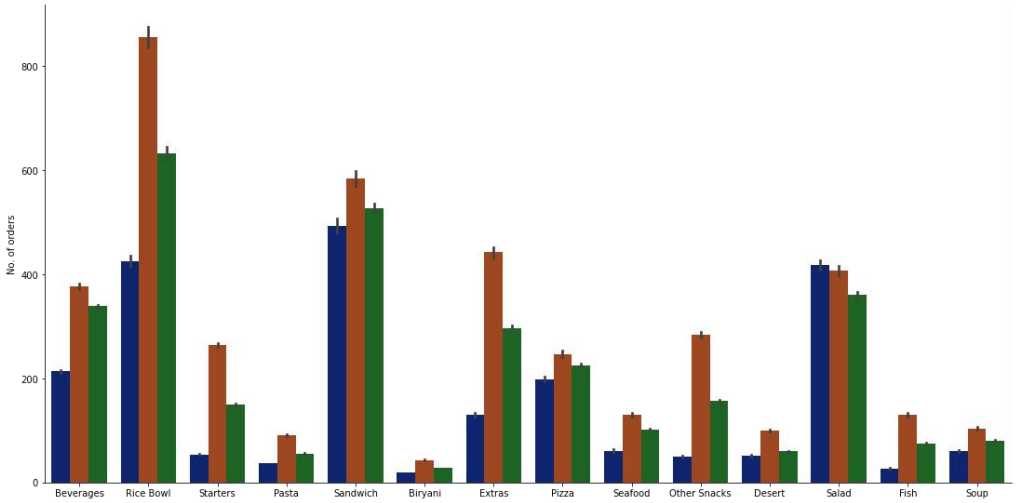

id
week
center_id
meal_id
checkout_price
base_price
emailer_for_promotion
homepage_featured
num_orders
city_code
region_code
op_area

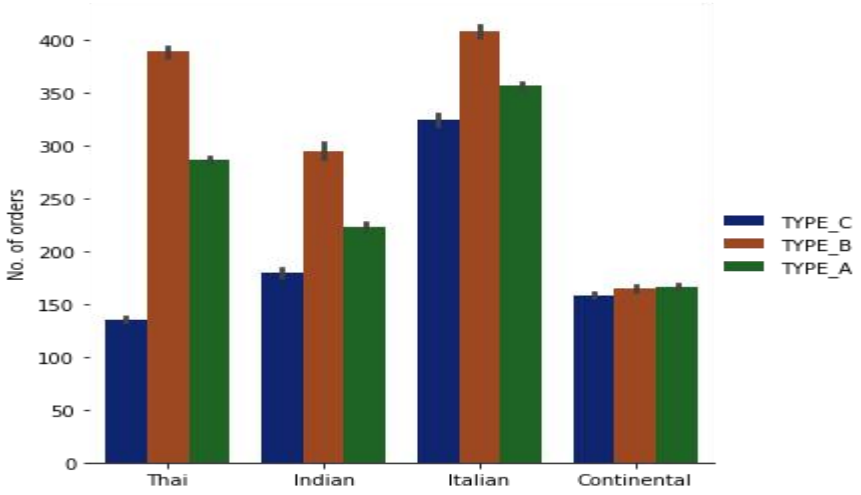
CATPLOT

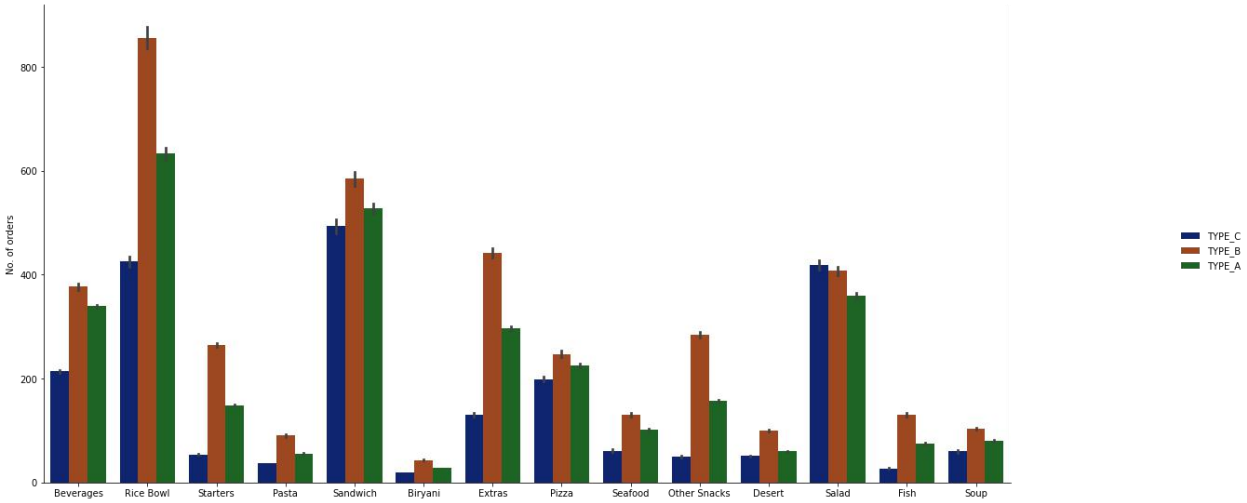

Base price ($)
0
100
200
300
400
500
600
700
Thai
Indian
Italian
Continental
center_type
TYPE_C
TYPE_B
TYPE_A

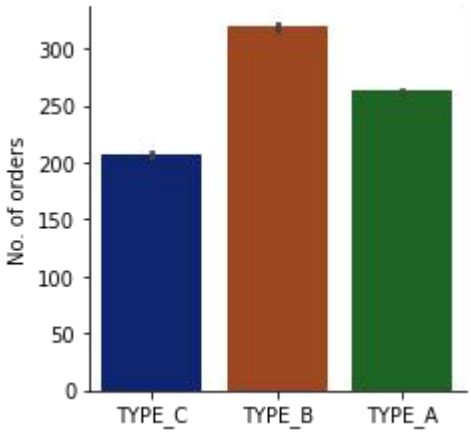

Checkout price ($)
600
500
400
300
200
100
0
Thai
Indian
Italian
Continental
center_type
TYPE_C
TYPE_B
TYPE_A

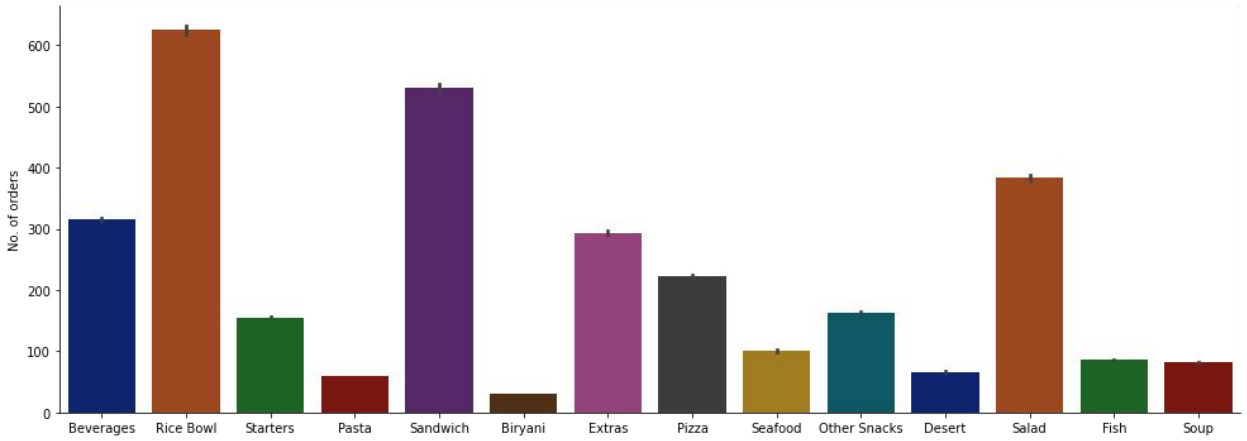

No. of orders
400
350
300
250
200
150
100
50
0
Thai
Indian
Italian
Continental
TYPE_C
TYPE_B
TYPE_A

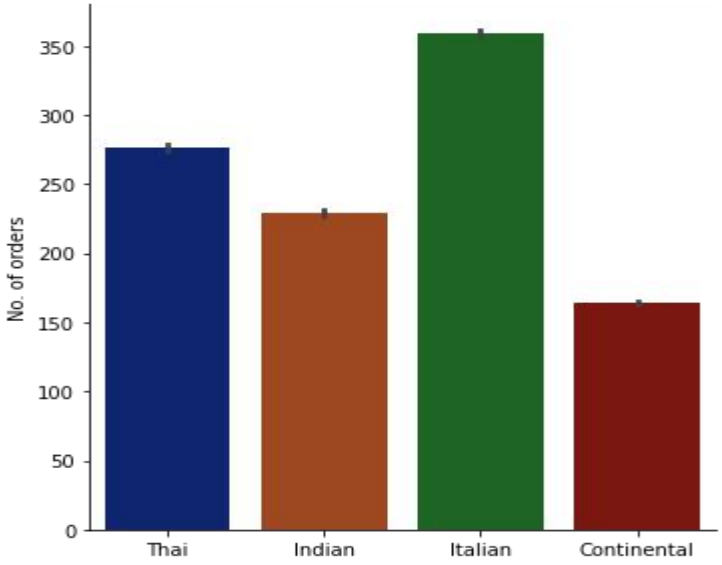

No. of orders
800
600
400
200
0
Beverages
Rice Bowl
Starters
Pasta
Sandwich
Biryani
Extras
Pizza
Seafood
Other Snacks
Desert
Salad
Fish
Soup
TYPE_C
TYPE_B
TYPE_A


No. of orders
400
350
300
250
200
150
100
50
0
Thai
Indian
Italian
Continental
TYPE_C
TYPE_B
TYPE_A

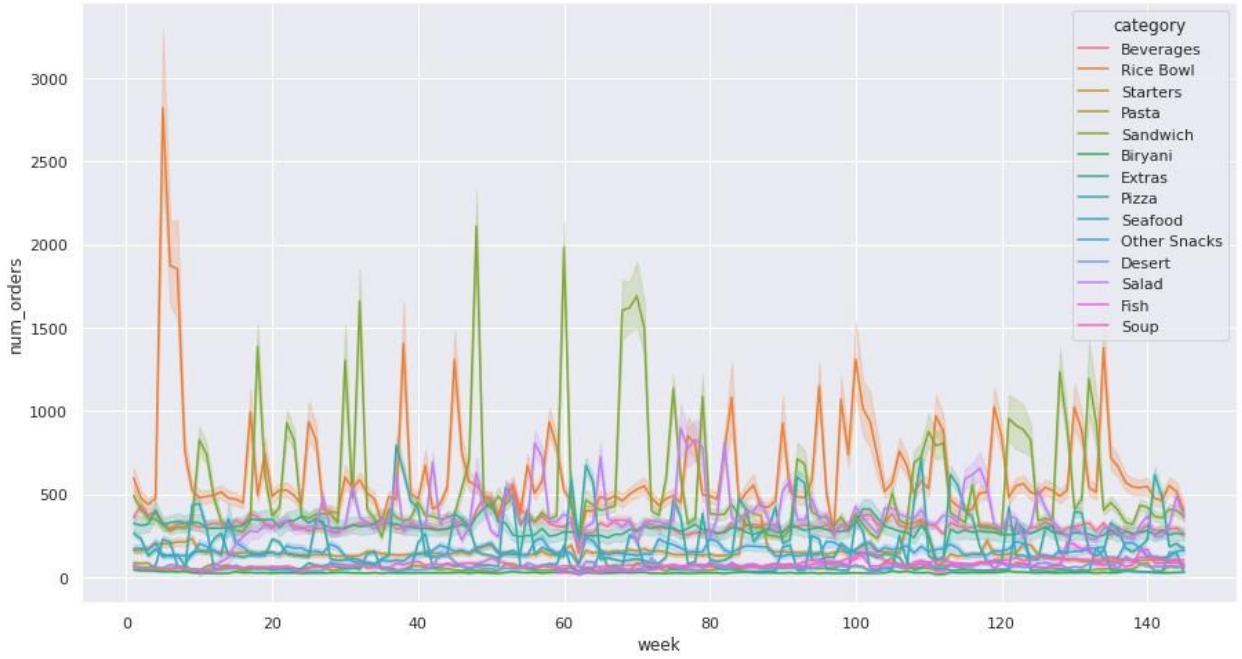

No. of orders
800
600
400
200
0
Beverages
Rice Bowl
Starters
Pasta
Sandwich
Biryani
Extras
Pizza
Seafood
Other Snacks
Desert
Salad
Fish
Soup
TYPE_C
TYPE_B
TYPE_A

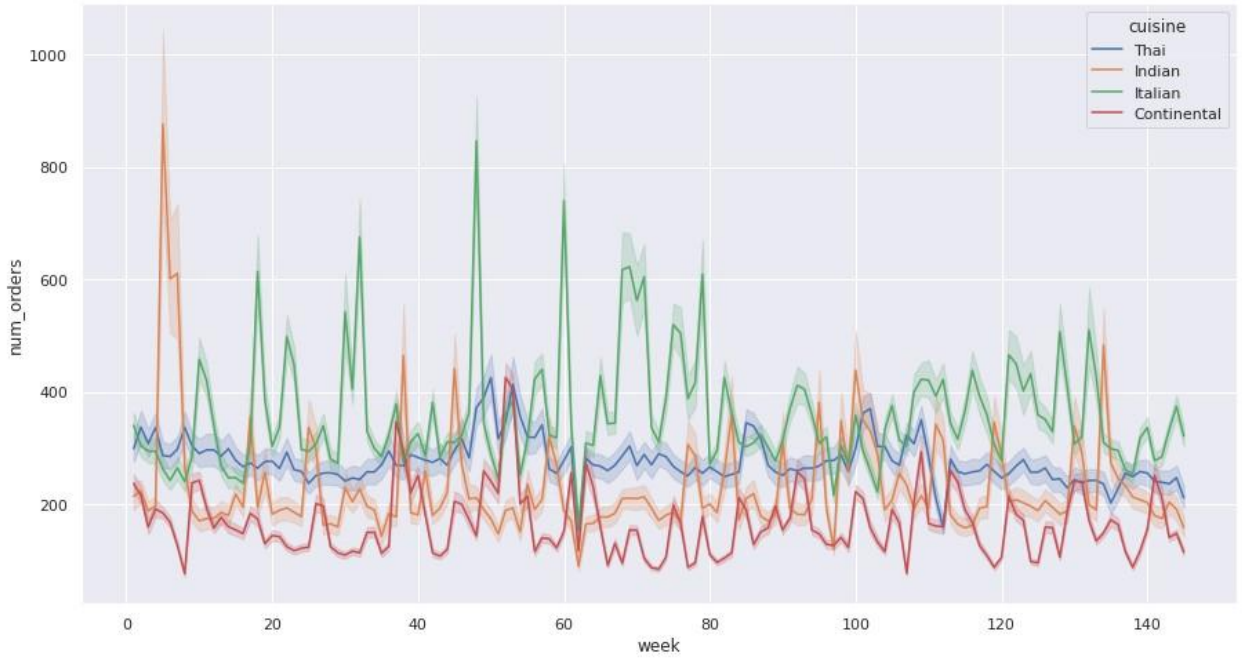

No. of orders
300
250
200
150
100
50
0
TYPE_C
TYPE_B
TYPE_A


No. of orders
0
100
200
300
400
500
600
Beverages
Rice Bowl
Starters
Pasta
Sandwich
Biryani
Extras
Pizza
Seafood
Other Snacks
Desert
Salad
Fish
Soup

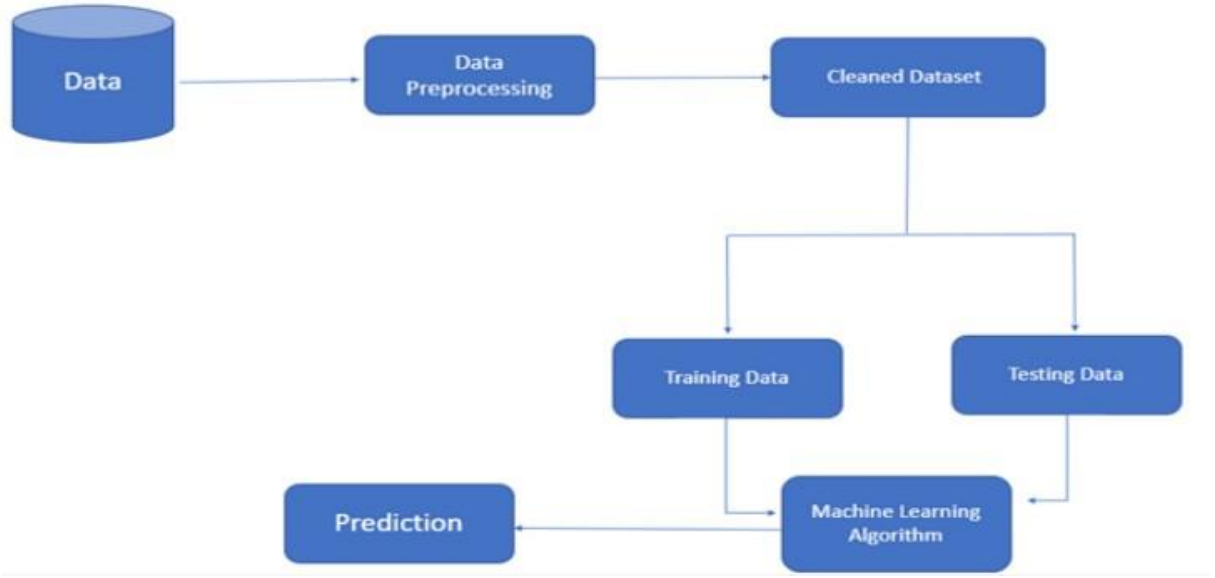

No. of orders
0
50
100
150
200
250
300
350
Thai
Indian
Italian
Continental

LINE PLOT


category
Beverages
Rice Bowl
Starters
Pasta
Sandwich
Biryani
Extras
Pizza
Seafood
Other Snacks
Desert
Salad
Fish
Soup
num_orders
week


cuisine
Thai
Indian
Italian
Continental
num_orders
week

BLOCK DIAGRAM

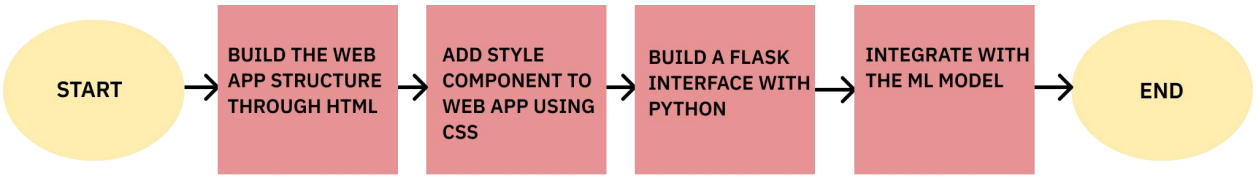


SOFTWARE DESIGN

A User Interface for Predicting the Food Demand using Machine Learning was created in software designing using following software technologies:

- **Flask -** Backend Part
- **HTML -** Webpage Outline
- **CSS -** Styling Part
- **Zoho Writer -** For Documentation ● **Github -** Code Repository

FLOWCHART OF UI DEVELOPMENT



APPLICATION BUILDING

In this section we have built a web application to integrate the ML model which we built. A UI is built for the user to enter the input details ie; feature variables for predicting the total order volume of the particular product at a center and locality. Once the input is received it is then fed into the model for prediction purpose after which the result will be printed in the same page.

HTML Page

- We have used HTML to create the structure of webpages.
- A total of 9 webpages were made for the Food Delivery Company Website.
- In which the index.html file shows the overall UI of the website.
- Users will be able to navigate to different pages with the help of hyperlink.
- Pages include registration page, about page, contact page, login page, admin page and feedback page.

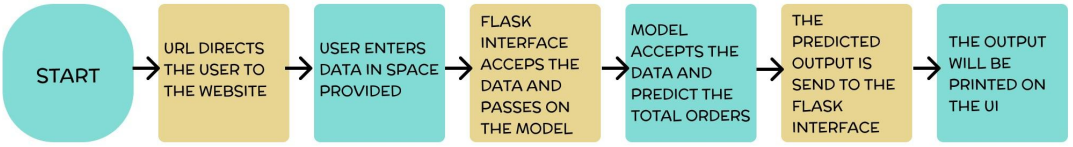
CSS

- CSS has been used for the styling purpose of the webpages.
- A total of 4 pages with extension .css were made.
- The colors and theme used are friendly for eyes.
- External CSS consist of these 4 pages made.
- While Internal CSS was used for some.

Flask

- Build flask file "app.py" which is a web framework written in python for server-side scripting.
- App starts running when "__name__" constructer is called in main.
- render_template is used to render the HTML pages. ● "GET" method is used to get input from the user.
- "POST" method is used to display output to the user.

WORKFLOW OF SOFTWARE



RESULT AND CONCLUSION

In this paper, we have used dataset collected from Kaggle for the forecast of food demand for a food delivery company. Food demand prediction is an important and challenging problem. In this paper we presented various regression methods which we used and how we evaluated each in order to reach into a final proposed model. As we go through different algorithms for prediction the performance rate keeps on varying.

There was not big difference other than precision rate of forecasting. It was found during the experimental study that the feature engineering techniques are the ones which can cater towards better result. Hyper parameter tuning was a necessity to aid the algorithms in the feature extraction so that the performance can be made optimum.

Because of the statistical characteristic of the dataset we have used RMSLE or Root Mean Squared Log Error as our evaluation metric. This evaluation is used practically for restaurants. Furthermore, in future more refined prediction can be done based on many other factors like cultural habits, religious holiday, consumer preferences etc. In future, this method can be used for predicting work force requirement, automated food ordering based on forecasting results.

FUTURE SCOPE

The implementation of AI and ML in food manufacturing and restaurant businesses is already moving the industry to a new level, enabling fewer human errors and less waste of abundant products; lowering costs for storage/delivery and transportation; and creating happier customers, quicker service, voice searching, and more personalized orders. Robotics is still quite a subtle thing to introduce, even for big factories and restaurant businesses, but it will occupy its niche very soon, bringing an obvious benefit in the long run.

The Benefits of AI in the Food Industry

1. Recently, more and more companies are trusting Artificial Intelligence to improve supply chain management thorough logistics and predictive analytics as well as to add transparency.
2. Digitization of the supply chain ultimately drives revenue and provides a better understanding of the situation. AI can

analyze enormous amounts of data that are beyond human capability.

3. Artificial Intelligence helps businesses to reduce time to market and better deal with uncertainties.
4. Automated sorting will definitely reduce labor costs, increase the speed of the process, and improve the quality of yields.

With AI, the food industry will ultimately be better in the area of safety standards.

BIBILIOGRAPHY

- How to Deploy Machine Learning Models using Flask (with Code) (analyticsvidhya.com)
- https://spd.group/machine-learning/machine-learning-and-ai-in-food-industr y/#The_Benefits_of_AI_in_the_Food_Industry
- https://www.analyticsvidhya.com/blog/2020/03/beginners-guide-random-for est-hyperparameter-tuning/
- https://www.kaggle.com/kannanaikkal/food-demand-forecasting
- https://datahack.analyticsvidhya.com/contest/genpact-machine-learning-ha ckathon-1/