DEMAND FORECASTING OF A FOOD STORE ANALYSIS: AN INDIA MEAL DELIVERY COMPANY

By:

Karunashree NAGARAJ

Supervised by: Dr.NYAWA Serge

A research dissertation
Submitted for the partial fulfillment
for the degree of

MSc
Artificial Intelligence and Business Analytics

At

Toulouse Business School
Toulouse, France

20 June 2022

ABSTRACT

The food and beverage industry are one of the most important sectors in any country's economy. Predicting food demand is a daunting task. Especially when online and in-store orders generate a wealth of data that needs to be efficiently stored, analysed, understood, and ultimately runnable in a very short time. The project uses a variety of machine learning algorithms, including web applications that allow users to predict food demand in the coming weeks, to predict different food demands in different food delivery businesses in different regions.

Keywords: Demand Forecasting, Food delivery, Raw materials,

AKNOWLEDGEMENT

First and foremost, I would like to thank my academic advisor, Dr. NYAWA Serge, for constructive insights and guidelines in my master's thesis.

I would like to thank the freshmen for providing the dataset for their research.

Second, I would like to thank my professor and the director of the master's program. Artificial Intelligence and Business for master's degree Lectures and Continuous Support

I would also like to thank all the professors and staff of Toulouse Business. Almighty God for supporting me in all my achievements for helping me continue this master's course.

Table of Contents

INTRODUCTION

1.0 Background

Food delivery is one of the world's most popular services today. This concept is not novel, and it allows customers to order from the restaurant whenever they want. Ordering at home has become commonplace for everyone, from ordering over the phone to using an online food platform linked to the restaurant of your choice. Many companies around the world have focused on strategy and begun to compete in the food delivery industry over the last seven years. Nonetheless, it has greatly expanded its options, giving consumers access to a vast array of restaurants and millions of dishes.

The global food delivery market is € 83 billion, or 1% of the total food market and 4% of food sold in restaurants, according to McKinsey & Company's "Changing Food Delivery Market" (Hirschberg et al., 2016). Equivalent. And fast-food restaurants. Most countries have reached maturity, and it is expected to grow at a compound annual growth rate of only 3.5 percent over the next five years.

The current market value of food delivery in India is estimated to be $13.99 billion, with a projected increase to around $21.5 billion by 2026. The annual growth rate is expected to be around 11.92 percent. Although traditional phone orders continue to dominate the food delivery industry, advances in digital technology are reforming the market and redefining how the industry's players compete. The current market is accustomed to shopping online through websites or apps.

As a result, the transparency and convenience of digital platforms enhance the consumer's experience of ordering lunch, dinner, or snacks whenever and wherever they want. Investments have continued to pour into the digital market, fuelling its rapid growth. Hirschberg et al. estimate that 26 percent of standard by-phone. Swiggy, UberEATS, Zomato, and Box8 are among the major global players in the digital food delivery platform. These players are concentrated in various market regions throughout the country. To survive in this highly competitive market, these companies, for example, must have a distinct and robust strategy for meeting global consumer demand.

For a company having the right technology in place to scale in the timeframe desired. As a result, when a company's demand increases, it is ready for growth and follows the predetermined technology strategy. Having a business forecasting process in place in the company is one way to predict future growth. In this context, the forecasting process is how a company can effectively predict how many orders will be placed, particularly to prepare scalable support for operations.

1.1 The Use of Demand Forecasting in freshmenu

Fresh menu, founded in 2014 by Rashmi Daga, is an Indian company known for its online food-ordering platform that operates 44 cloud kitchens. In these cities, they have several fulfilment centres where they ship meal orders to customers. To meet the demand for meal delivery orders in various cities, having enough supply is critical.

Forecasts of frustrating processes can be found in traditional businesses. Forecasting necessitates the use of historical data, which is frequently dispersed throughout the organization. The method for obtaining forecast results is also ambiguous. There is usually a disconnect between someone who knows the formula and the company's business. Because so many stakeholders require forecast data, competing incentives complicate the forecasted metrics / objectives. To make matters worse, the predictions may be politically influenced and therefore suboptimal. As a result, it is not beneficial to the decision maker and will cause the decision maker to restart and trust his or her intuition. It is also difficult to make accurate forecasts from the start because decision makers require time to review and trust the forecast results.

Accepting that nothing exists is one of the benefits of prediction for decision makers. There is no right or wrong way to make or use predictions; only a good or bad way to create or use them (Kolassa & Siemsen, 2016). If the result that predicted differs from the actual result and allows to use of all available information. It can be considered an accident if it is used effectively. On the contrary, if the prediction is accurate. Is correct, but it lacks the information required for the process, so you cannot use it to determine uncertainty; instead, you must focus more on the process that caused it. The choice as well as the quality of the decision itself, considering the actual results.

Despite the risks and uncertainties of forecasts, many businesses use them to plan their strategies (Shim, 2000). It is caused due to the inherent uncertainty of operations in business. The predicted numbers are unlikely to accurately reflect the actual results. Because no forecast is perfect, the company must face, quantify, and balance the remaining risks. The difference in uncertainty between actual and predicted numbers. for instance, Based on previous historical data, a partial offset (x percent plus or minus) of the forecast. Using the error rate approach sets decision makers' expectations and allows them to meet them. Furthermore, the predictions reveal the possibility that the decision will be influenced.

Designing a good demand forecast system may appear difficult at first, but the benefits of getting the method right are enormous. The difficulty in developing the forecasting process is not only in utilizing advanced algorithms or hiring experienced individuals, but also in managing cross-functional communications among stakeholders (Smith 2009). The benefits of overcoming these obstacles are enormous. For example, Simon Clarke (Clarke, 2006) described a major overhaul of dope IN forecasting's method that resulted in a significant reduction of worker-days in their inventory warehouse. Similarly, these forecasts on demand will considerably improve however a corporation should prepare their provide and be treated as a major chance. Therefore, as one of India’s top food delivery stores Fresh menu has been unrelentingly making effective and efficient processes internal.

1.3 Problem Statement

- The goal is to assist these centers in forecasting demand for the coming weeks so that they can plan their raw material stock accordingly.
- Most raw materials are replenished on a weekly basis, and because the raw materials are perishable, procurement planning is critical. • Second, accurate demand forecasts are useful in staffing the centers.
- Given the following information, predict the demand for the following center-meal combinations in the test set for the next 10 weeks (Weeks: 146-155):

1. Demand history for a product-center combination (Weeks: 1 to 145)
2. Meal product characteristics such as category, sub-category, current price, and discount Information for fulfillment centers such as center area, city information, and so on.

LITERATURE REVIEW

Sales forecasting is an important field in the grocery and food industry, and it has recently received a lot of attention because of new technologies to improve business operations and profitability. However, while the industry has historically relied on traditional statistical models, more advanced machine learning methods have gained popularity in recent years.

As a result, when it comes to sales forecasting, various studies are examined in order to comprehend the current methodology used by various organizations in the industry, and some of the best practices, such as Linear regression, Xgboost, and weighted moving averages, are discovered and may be considered for the purposes of this paper.

“Demand forecasting in retail grocery stores in the Czech Republic”

Research Problem	Methodology	Results	Further analysis
In the Czech Republic, grocery store retailers forecast demand using their own intuition and retailing experience, so qualitative forecasting methods are most used).	In the Czech Republic, quantitative research was conducted in 75 selected retail stores. Only retail stores where groceries predominated in the range of goods were included in the sample.	Judgmental method (40%), (Moving) Average (21%), Naive method (19%), Customer expectations (9%), Unknown methods implemented in software (9%). (9 percent). (5%), Analogy method (4%), simple regression (3%), time series decomposition (1%), exponential smoothing (0%), ARIMA models (0 percent) Advanced forecasting models (0 percent).	More research should be conducted to identify the causes of the current level of demand forecasting in the retail industry, including the possibility of removing barriers to the implementation of more appropriate approaches to demand forecasting in the surveyed retailers.

“Flexible Demand Forecasting in Intelligent Food Supply Chain Management”

Research Problem	Methodology	Results	Further analysis
In this work, three modules dealing with Food Supply Chain Management are discussed: Demand	This study provides a new Demand Forecasting module algorithm that combines an outlier detection technique	All three modules are general and may be used to any domain of need. They can be integrated and	More methods could be investigated to improve accuracy..

Forecasting, Food Tracing System, and Information Sharing Module for Suppliers, Warehouses, and Restaurants to connect with one another.	with the LightGBM Regressor, which handles stated targets, and the SARIMA Algorithm, which handles data seasonality.	deployed to manage any supply chain, assisting in improved supply chain management to handle the supply chain's flexibility and unforeseeable events.	

“Designing a Demand Forecasting Service in a Food delivery Platform”

Research Problem	**Methodology**	**Results**	**Further Analysis**
The goal of this study is to learn about Wolt's present demand forecasting service and to apply the user-centric design approach to create a service that effectively supports the forecasting process**.**	The user-centric design methods, such as semi-structured interviews, affinity diagramming, stakeholder mapping, persona, user journey mapping, and service blueprint, were employed in this thesis.	In rethinking how the demand forecast process should move ahead in the future, the service design approaches helped uncover the pain areas, wishes, and gaps. The method describes the underlying experience of the OMs when anticipating demand and uncovers the hidden needs and insights that may be used to improve future forecasting service development.	The service of demand forecasting could be expanded to include supply forecasting. The OMs must anticipate how much courier supply they will need to be online and deliver the food demand in each time frame based on demand forecasts. If machine learning is used to automate both components of the demand and supply forecasting calculations, OMs will be more likely to monitor ad hoc or exceptional occurrences**.**

“Peek to Peak: Time Series Forecasting to predict demand for meal kits”

Research Problem	Methodology	Results	Further Analysis
The paper focuses on to develop a model that will help reduce ‘losses’ for a meal kit service. This was done by quantifying monetary losses which lead to saving thousands of dollars on a weekly basis.	Applied Machine learning to bring in profits for a company and showed how such a model could be lucrative to this unique industry.	Random forest was the chosen model which gave comparable results across both Mean Squared Error and R squared and can be tuned further.	Further analysis can be done on

“Models for predicting perishable products demands in food trading companies”

Research problem	Methodology	Results	Further Analysis
The primary goal of this thesis is to create a machine learning-based decision support system to forecast sales of a self-service restaurant in Bangalore, India**.**	On the dataset, apply machine learning techniques such as regression and KNN decision tree**.**	Positive Coef. determines a positive association. The relevance of a variable in a model is determined by its P-value score. Data from Google Trends, which shows the user search trend for restaurant searches, can be used as a key variable in the model.	Another technique, such as a neural network or LSTM, can be examined with the availability of more detailed and exact data on an hourly factor. The time factor for predicting sales can be reduced by integrating the model over the cloud for continuous analysis of POS data.

RESEARCH METHODOLOGY

Several machine learning approaches are frequently used in demand forecasting features. Several factors influence the selection of machine learning models, including business goal, data type, data amount and quality, forecasting period, and so on.

3.0 Data collection

Data was collected from a meal delivery company which operates in multiple cities across India. From freshmen and analytics Vidhya. Where Dataset size – 55.7 MB (Number of observations 456548).

3.1 Data Description:

Data Files:

Weekly Demand data (train.csv)

fulfilmentcenterinfo.csv

meal_info.csv

3.2 Evaluation Metric:

A evaluation metric is a quantifiable measure used to track and evaluate the status of a particular business process. The metric to be used in this project is RMSE (Root Mean Square Value), and its formula is

$$RMSE = \sqrt{\frac{1}{n} \sum e_t^2}$$

where, error, $e = (y_i - \hat{y})$, actual value = y_i, predicted value = $\hat{y}$

The standard deviation of the error is represented by the Root Mean Square Error (residual error). It's always good, and a lower value means greater results. The optimum value is 0, however this is never reached.

The RMSE penalizes large errors. It yields the same unit as the outcome variables, which is smoothly differentiable and simplifies mathematical operations. Because RMSE is highly sensitive to outliers, it should not be used before removing outliers.

RMSE is the most used metric in modern ML and Data science problems, such as meteorology, bioinformatics, economics, hydrogeology, imaging science, computational neuroscience, Netflix prize submission, and simulation of building energy consumption.

The coefficient of determination (R2) is used in this research to measure how well a model fits a dataset. R2 ranges from 0.0 to 1.0. A value of 1.0 denotes a perfect fit and thus a highly reliable model.

EXPLORATORY DATA ANALYSIS

4.0 Univariate and Bivariate Feature Analysis:

Histograms, distribution plots, scatter plots, line plots, and boxplots are examples of feature analysis used to understand data insights. These are the most effective strategies for generating beautiful data insights, particularly when it comes to the relationship between two variables.

However, it fails to give complete information on which variable has the most important feature to give highest accurate prediction.

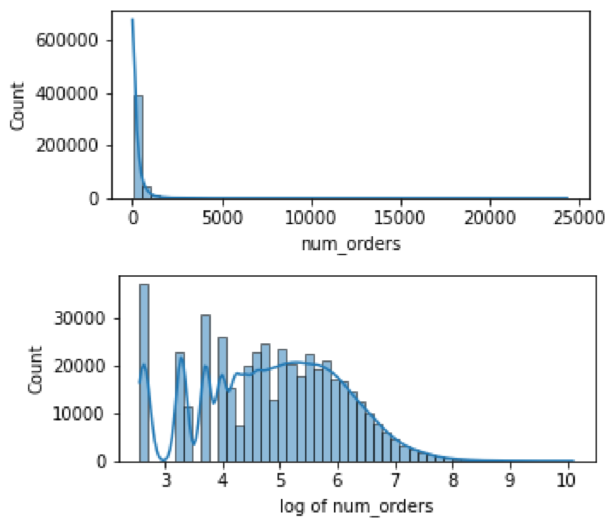
We plotted some of these to know more about the behavior of our data.

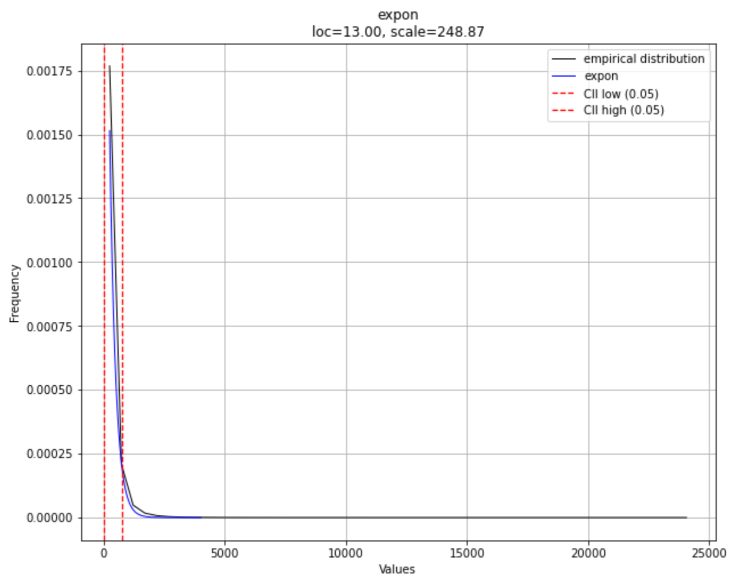
The pattern of number of orders follows exponential distribution and can be seen from the graph below along with its histogram of "log of num_orders" with its kernel density estimation (kde).

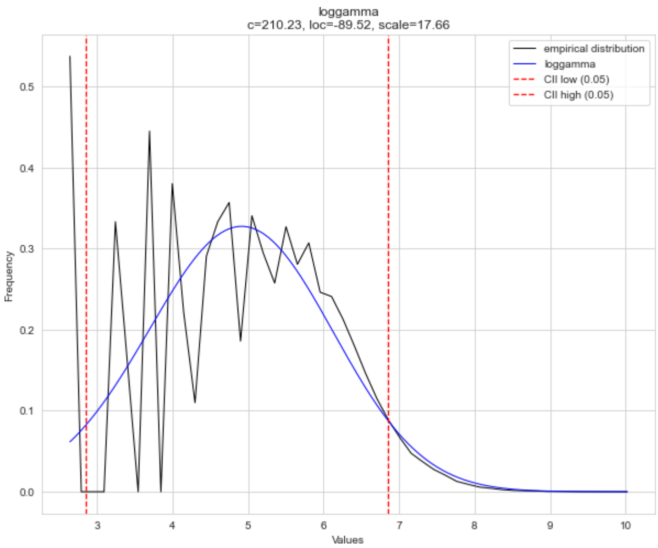
Log is taken to handle the wide range of data (13 to 24299). It is multimodal as it has more than one peak point. It is right skewed with a value of 6.929966065.

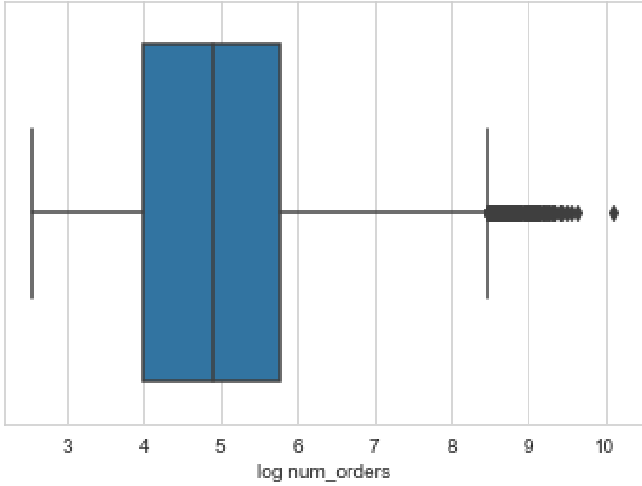
The higher number of orders is concentrated in between 1 to 6. The plots below depict the difference on the histogram of num_orders with and without log.

The log of num_orders follow log gamma distribution and from the boxplot of the log of num_orders, it is seen that there are some outliers that needs to be removed.



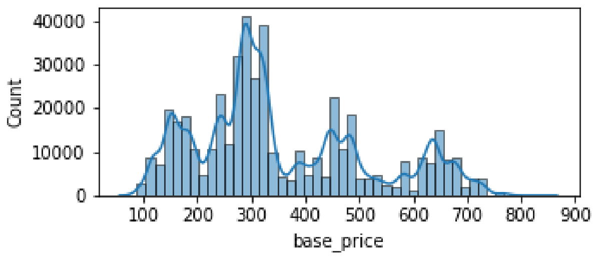


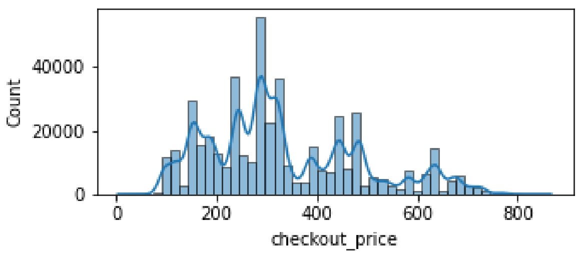


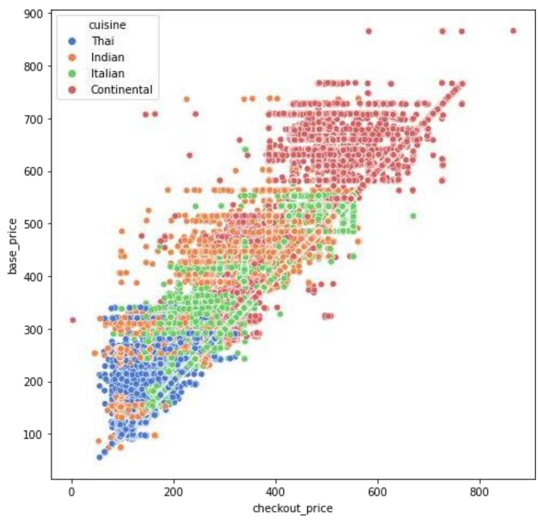


From the histogram of base_price and checkout_price below, it is found that they both are in the range of 100 to 800 with 300 as the highly concentrated range.

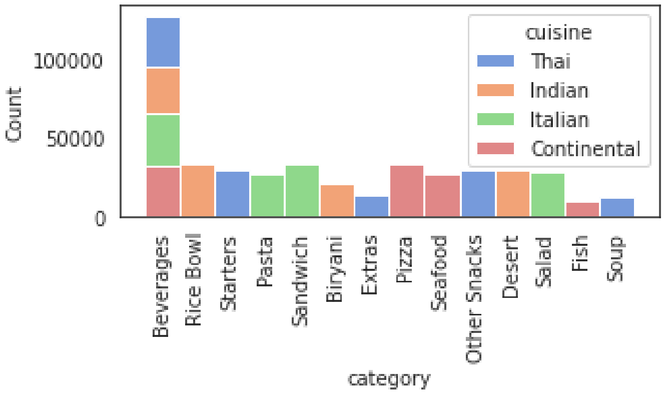
As the scatter plot between base_price and checkout_price depicts they have linear relation where majority of the base_price are on the higher side although there are also some points with higher checkout_price than its base_price. This means that there is a greater number of discounts

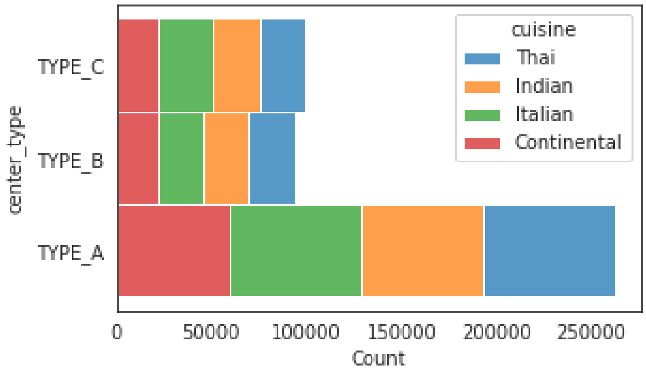






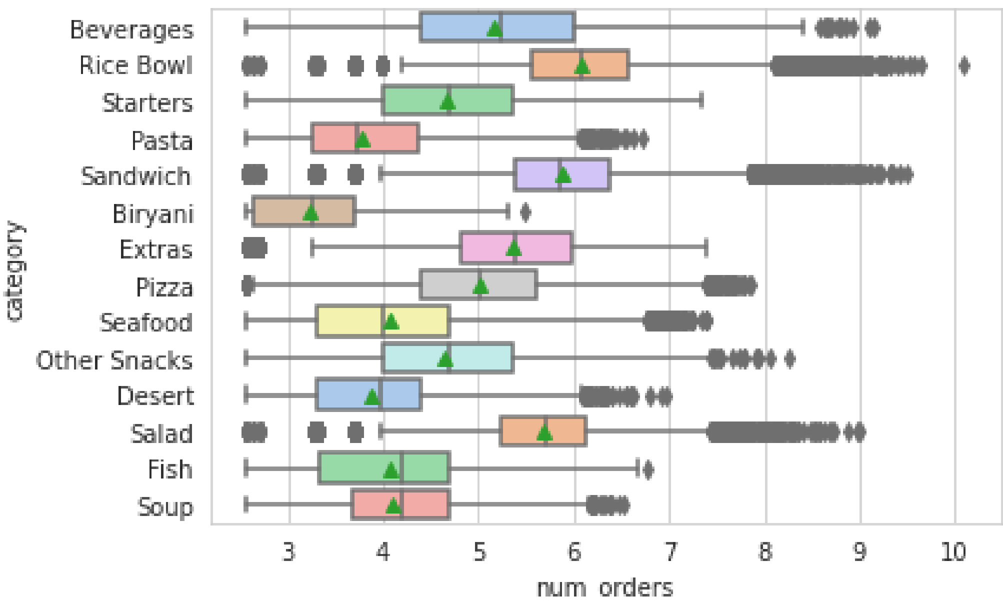
There are 14 meals and 4 cuisines in the given dataset. From the below histogram of variable “category”, “Beverages” can be seen in all the “cuisine”, consequently having the highest number of orders compare to other meals.



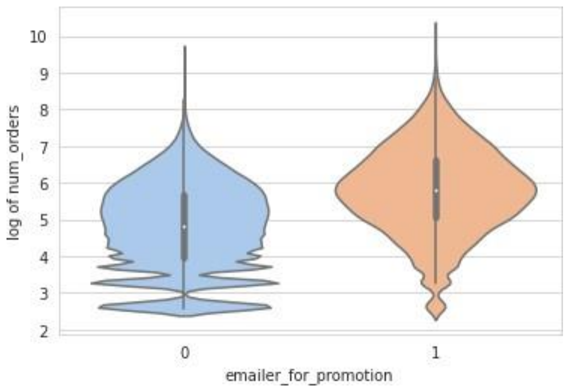


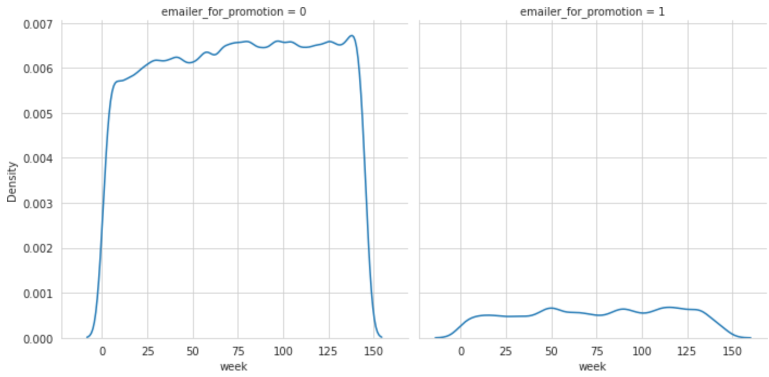
There are 3 centre types given in the dataset. From the above histogram of variable “center_type” colour encoded with “cuisine”; it is seen that there is equal distributions of all cuisines in all types of centre.

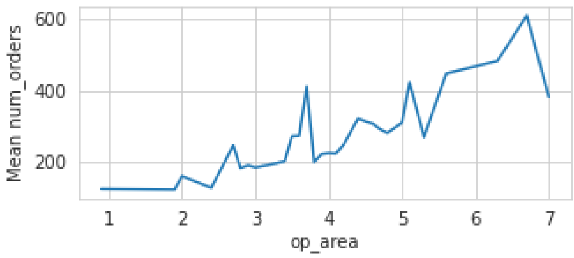
From the below boxplot of variable “category” with the number of orders, it is found that “Rice Bowl”, “Sandwich” and “Salad” has higher outliers and mean number of orders compare to other meals, while “Biryani” has the lowest mean order value.



From the distribution and violin plots below, it can be seen that “emailer_for_promotion” with value “1” has a smooth distribution curve compared to the one with value “0”, even though it was lasted for very a smaller number of weeks. Likewise, variable “homepage_featured” with value “1” has same effect on the number of orders.





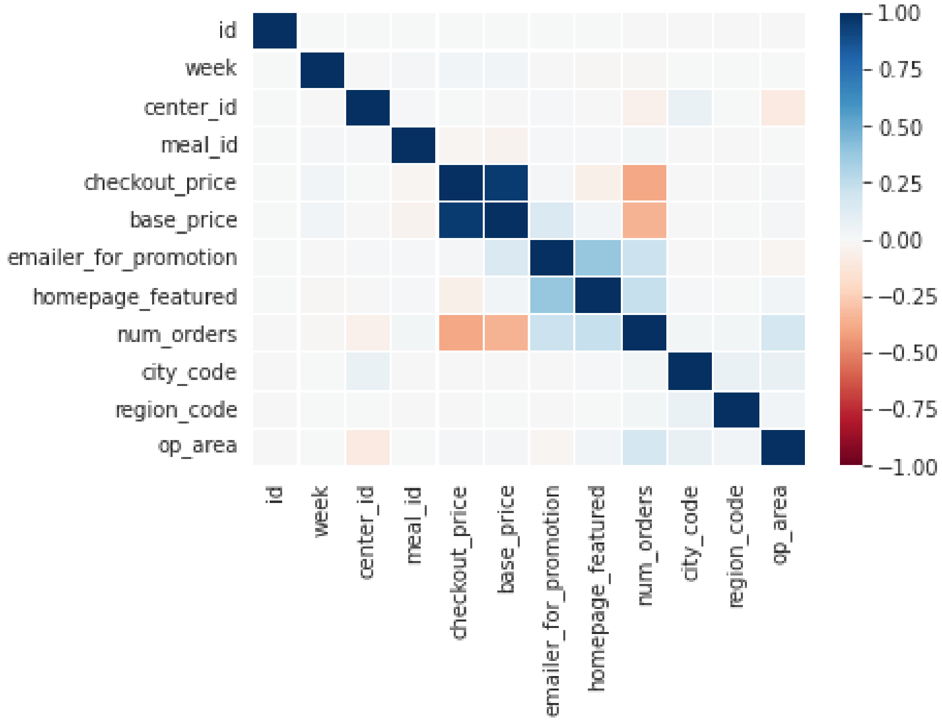


4.1 Multivariate Feature analysis:

Finding correlation coefficients is one of the most effective multivariate analyses. Pearson correlation is the most common one which gives measure of how numerical variables are correlated to each other. So as Spearman and Kendall coefficients.

However, Spearman correlation coefficient is more robust to outliers and handles sparse matrix more effectively. Below is a heatmap plot for spearman coefficient. It gives values between -1 to 1. 1 being linearly related, -1 being inversely related while 0 being no relations.

Checkout_price and base_price are closely related and inversely related to the num_orders. Emailer_for_promotion and homepage_featured are also closely related and to num_orders as well. Op_area is also somehow related to num_orders.



MODELLING AND ERROR ANALYSIS

5.0 Feature encoding:

Before we did pre-processing methods to make the data machine-readable before moving on to the modelling part. Hot encoding is one of the most basic strategies for working with categorical information. The technique converts categorical variables into binary digits, making input variables more machine-readable. All categorical variables are converted to a single column, resulting in a sparsely high dimensional dataset.

“category”, “cuisine” and “center_type” columns are one hot encoded and “category”, “cuisine” and “center_type” columns are removed as they will not be needed anymore.

As it is a time series problem, whole data frame is arranged in ascending order of “week” column and splits into train and test data with Timeseries Split (n_splits=5). To begin with, we tried different Machine algorithms to check which algorithm fits the best to our dataset.

Here we have taken data for only 4 weeks because some of the algorithm might take forever to run the whole 456548 data. We have also done grid search hyper-parameter tuning to find thebest hyper-parameter on each algorithm. After find the best algorithm we would train our whole data on it.

5.1 Models

Baseline model:

As a simple baseline model, mean value of y_train is assumed to be predicted value. RMSE value and R2_score is being calculated as shown below. RMSE turns out to be 360.93 and R2_score with negative value. This shows that the baseline model is quite a bad model.

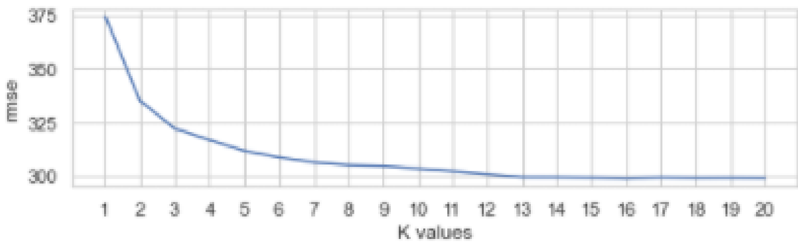
```
RMSE: 360.930
  R2:  -0.003
```

KNN regression:

KNN regression is a non-parametric method that approximates the relationship between independent variables and continuous outcomes by averaging observations in the same neighborhood.

Here, distance measured is Minkowski by giving more weight to the closer neighbors of the querypoints we tried for different k values (1 to 20). 16 came out to be the best k value with RMSE value of 289.829. Codes below shows how we have implemented KNN algorithm and plotted RMSE vs. K graph for.

```
Best k: 16.0
RMSE: 298.829
R2: 0.039
21.05 sec
```



Linear regression:

A supervised learning algorithm is the linear regression model. Based on a collection of independent factors, linear regression is used to predict the value of a dependent variable (y) (x1 through xp). As a result, this regression technique determines that x (input) and y (output) have a linear relationship (output). As a result, the name Linear Regression was chosen. Linear regression attempts to fit data into a straight line by minimizing variance in the form of the sum of squares of all errors. The linear regression formula is:

$$y = X\beta + \varepsilon$$

Where y is the dependent variable's predicted or expected value, X is the N*K matrix's independent or predictor variables, is the value of y when all the independent variables are equal to zero (error) and is the estimated regression coefficients by ordinary least squares, as given by the formula.:

$$\beta = \arg min_b \sum_{i=1}^{N}(y_i - x_i b^2)$$

Linear regression assumes that the data follows a normal distribution and fits a straight line whichis not the case in our problem; hence linear regression gives RMSE error of 222.157.

```
RMSE: 222.157
R2: 0.469
Time taken: 0.03 sec
```

SGD regression:

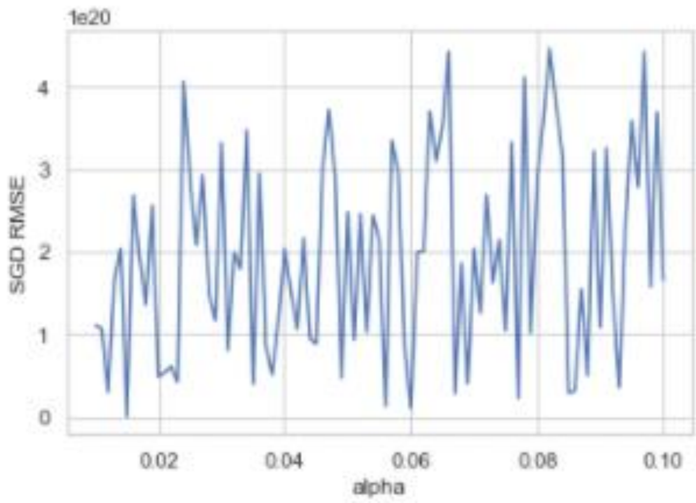
Stochastic gradient descent is a machine learning optimization approach that is frequently used to discover the model parameters that best fit the expected and actual outputs. It's a simple but effective method.

The Stochastic Gradient Descent (SGD) regressor is a simple SGD learning process that can fit linear regression models with a variety of loss functions and penalties.

SGD with default parameters: loss= “squared_error” and eta0=0.01, with the best alpha=0.23, RMSE result with a large value and R2 with negative value.

```
Best alpha: 0.23
RMSE: 6.1861234326198714e+19
R2: -2.5763897362568915e+35
Time taken: 99.28 sec
```

.

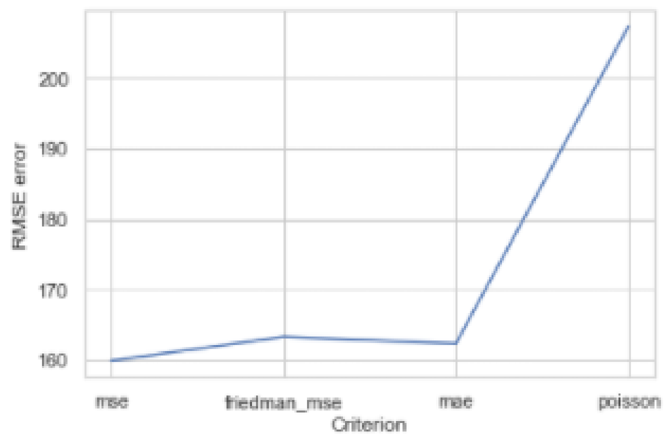



```
Time taken: 26.75 sec
```

Decision Tree regression:

To develop regression or classification models, decision trees employ a tree structure. It breaks a dataset into smaller and smaller subsets over time while also creating a decision tree. As a result, a tree is generated containing decision nodes and leaf nodes. RMSE=159.866 with criterion="mse" was found using decision tree regression.

```
Best parameters: mse
RMSE: 159.866
R2: 0.7249691405527896
Time taken: 15.49 sec
```

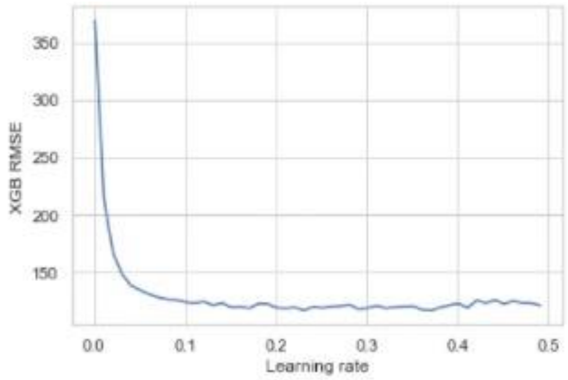



```
Time taken: 5.34 sec
```

XGBoost regression:

For improved performance and execution speed, XGBoost (eXtreme Gradient Boosting), an ensemble learning method that provides a systematic solution for combining the predictive power of multiple learners, is used.
XGBoost with best learning of 0.311, RMSE value is 121.009

```
Best learning rate: 0.311
RMSE: 121.009
R2: 0.843
Time taken: 665.98 sec
```




```
Time taken: 82.47 sec
```

Comparing all the error values of all the algorithms, we could see that XGBoost performs bestwith RMSE value of 121.009. we assume that XGBoost can perform well with our whole dataset, so we go forward with this powerful algorithm.

	Model	RMSE value	R2 value
1	Baseline	360.930	-0.003
2	KNN regression	298.829	0.039
3	Linear regression	222.157	0.469
4	Support Vector regression	6.19e+19	-2.57e+3
5	Decision Tree regression	159.866	0.724
6	eXtreme Gradient Boosting	121.009	0.843

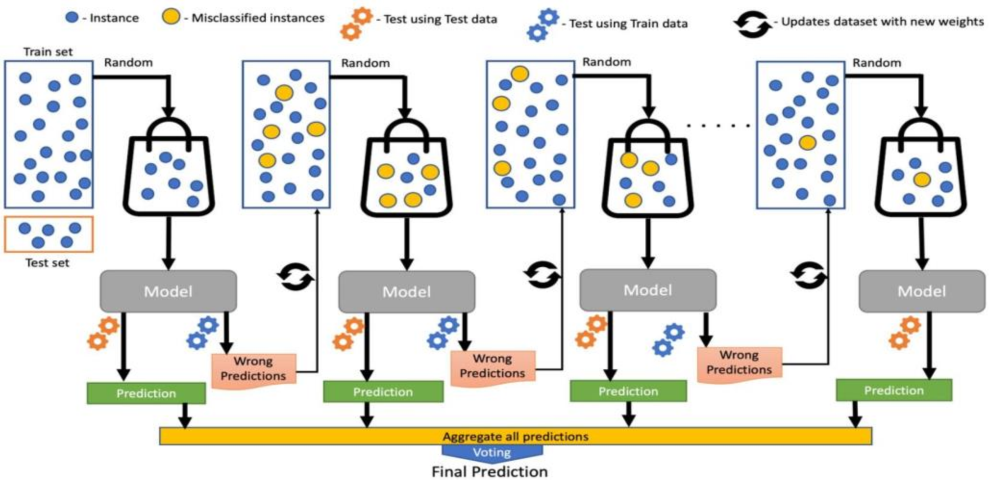
5.2 Deep dive into best model:

Regardless of the sort of prediction job at hand, regression, or classification, XGBoost is one of the most prominent machine learning algorithms these days. XGBoost is a method of ensemble learning. The above performance of various models shows that relying on the results of just one machine learning model may not be sufficient, and XGBoost performs the best of all models. Ensemble learning is a method for combining the predictive power of numerous learners in a systematic way. As a result, a single model is created that aggregates the output of several models.

Boosting trees are built in a sequential order, with each tree seeking to reduce the errors of the one before it. Each tree updates the residual errors by learning from the trees that came before it. As a result, an updated version of the residuals will be used to train the next tree in the sequence.

In boosting, the base learners are poor learners with a strong bias and predictive power that is only marginally better than random guessing. Each of these weak learners provides vital information for prediction, allowing the boosting strategy to efficiently combine these weak learners into a strong learner. The ultimate powerful learner lowers the bias as well as the variance.

The scalability of this powerful algorithm is its main advantage. The beauty of this powerful algorithm is its scalability, which allows for fast learning via parallel and distributed computing while also providing efficient memory usage. The method is similar to that of a neural network.



Three simple steps make up the boosting ensemble technique:
To predict the target variable y, an initial model F0 is defined. A residual (y – F0) will be associated with this model.
The residuals from the previous phase are fitted to a new model h1.
Now, F_0 and h_1 are combined to give F_1, the boosted version of F_0. The mean squared error from
F_1will be lower than that from F_0:

$$F_1(x) < -F_0(x) + h_1(x)$$

To improve the performance of F1, we could model after the residuals of F_1 and create a newmodel F_2:
$F_2(x) < -F_1(x) + h_2(x)$

This can be repeated for 'm' iterations till the residuals are as low as possible:

$$F_m(x) < -F_{m-1}(x) + h_m(x)$$

The additive learners in this case do not interfere with the functions created in the previous steps. Instead, they provide their own information to reduce the errors.

Having a lot of trees may result in overfitting. As a result, the stopping criteria for boosting must be carefully chosen.

Unlike traditional gradient descent algorithms, which minimize output error with each boosting round, gradient boosting predicts the additive model's ideal gradient.

Where t is the number of trees, d is their height, and x is the number of data points in the training set, XGBoost training takes O(tdxlogn). A new sample prediction is made.

Table below shows some advantages and disadvantages of XGBoost along with its applications.

<table>
<tr><th>Advantages</th><th>Disadvantages</th></tr>
<tr><td>1) Inbuilt L1 and L2 regularization prevents itfrom overfitting.
2) It can handle sparse and weighted data.
3) Inbuilt capability to handle missing values.
4) Better accuracy than single decision tree models
5) Better prediction performance as it uses boosting ensemble learning algorithm.</td><td>1) Training time is high for largedataset.
2) More likely to overfit than bagging.</td></tr>
<tr><th colspan="2">Applications</th></tr>
<tr><td colspan="2">1) XGBoost is intensively used in business intelligence modeling because of its ability to process scattered data efficiently, resulting in overall excellent model performance.
2) By optimizing CPU memory use with parallel computing, XGBoost becomes the appropriate model for extensive data models, classification, and feature recognition data.</td></tr>
</table>

ADAVANCED MODELING AND FEATURE ENGEINEERING

In the earlier section we were using data of only 4 weeks. After knowing XGBoost performs bestin those data we go forward training our whole data with XGBoost algorithm.

6.0 Hyperparameter tuning:
To begin with, we run our whole dataset with default parameters. Default parameters looks like:

```
{'max_depth': 6,
 'min_child_weight': 1,
 'eta': 0.3,
 'subsample': 1,
 'colsample_bytree': 1,
 'eval_metric': 'rmse',
 'objective': 'reg:squarederror'}
```

Running the model with the default hyperparameters, we see that minimum RMSE value converges unto 183.77 in 33 rounds. Below snippet of output shows how it converges.

```
[23]    Test-rmse:190.34909
[24]    Test-rmse:188.46718
[25]    Test-rmse:187.97455
[26]    Test-rmse:187.35979
[27]    Test-rmse:186.69882
[28]    Test-rmse:186.07811
[29]    Test-rmse:185.19665
[30]    Test-rmse:184.92970
[31]    Test-rmse:184.03519
[32]    Test-rmse:183.76558
[33]    Test-rmse:184.96480
[34]    Test-rmse:185.00926
[35]    Test-rmse:185.00470
[36]    Test-rmse:184.83537
[37]    Test-rmse:184.24490
[38]    Test-rmse:184.06963
[39]    Test-rmse:190.84912
[40]    Test-rmse:190.71968
[41]    Test-rmse:190.70969
Best RMSE with default parameters: 183.77 with 33 rounds
Time taken: 7.8748619556427 sec
```

With the help of gridsearch, tuning the value of max_depth and min_child_weight from 6 to 9 and 5 to 6 respectively, we see that minimum RMSE value is 133.69 when max_depth=7 andmin_child_weight=6. Then we update these values to our parameter. Codes are given below.

```
CV with max_depth=6, min_child_weight=5
        RMSE 135.2981322 for 998 rounds
CV with max_depth=6, min_child_weight=6
        RMSE 135.4826172 for 824 rounds
CV with max_depth=6, min_child_weight=7
        RMSE 136.07359639999999 for 997 rounds
CV with max_depth=7, min_child_weight=5
        RMSE 134.5116884 for 785 rounds
CV with max_depth=7, min_child_weight=6
        RMSE 133.6908294 for 784 rounds
CV with max_depth=7, min_child_weight=7
        RMSE 135.72905559999998 for 485 rounds
CV with max_depth=8, min_child_weight=5
        RMSE 135.8453614 for 395 rounds
CV with max_depth=8, min_child_weight=6
        RMSE 135.050038 for 448 rounds
CV with max_depth=8, min_child_weight=7
        RMSE 135.30952440000002 for 343 rounds
Best params: 7, 6, RMSE: 133.6908294
Time taken: 4369 sec
```

Then we again tune the value of sub_sample and colsample from 9 to 11 and 6 to 11 respectively. From the output of below codes we could see that minimum RMSE valueconverges to 132.61 when sub_sample=1.0 and colsample=0.6.

```
CV with subsample=1.0, colsample=1.0
        RMSE 133.6908294 for 784 rounds
CV with subsample=1.0, colsample=0.9
        RMSE 135.0003188 for 723 rounds
CV with subsample=1.0, colsample=0.8
        RMSE 133.8671446 for 857 rounds
CV with subsample=1.0, colsample=0.7
        RMSE 133.3833678 for 823 rounds
CV with subsample=1.0, colsample=0.6
        RMSE 132.6128082 for 964 rounds
CV with subsample=0.9, colsample=1.0
        RMSE 136.0344422 for 548 rounds
CV with subsample=0.9, colsample=0.9
        RMSE 134.58330099999998 for 934 rounds
CV with subsample=0.9, colsample=0.8
        RMSE 134.5635894 for 718 rounds
CV with subsample=0.9, colsample=0.7
        RMSE 136.70336 for 623 rounds
CV with subsample=0.9, colsample=0.6
        RMSE 135.207083 for 760 rounds
Best params: 1.0, 0.6, RMSE: 132.6128082
Time taken: 5377 sec
```

Last but not the least, we checked RMSE values for every values of “eta” = [0.3, 0.2, 0.1,0.01, 0.05]. “eta” with value 0.1 gives minimum RMSE of 130.27.

```
Wall time: 0 ns
CV with eta=0.3
Wall time: 9min 21s
	RMSE 132.6128082 for 964 rounds

CV with eta=0.2
Wall time: 8min 4s
	RMSE 132.56892720000002 for 838 rounds

CV with eta=0.1
Wall time: 9min 31s
	RMSE 133.5884828 for 998 rounds

CV with eta=0.01
Wall time: 9min 58s
	RMSE 160.86943060000002 for 998 rounds

CV with eta=0.05
Wall time: 9min 38s
	RMSE 139.5577792 for 998 rounds

Best params: 0.2, RMSE: 132.56892720000002
Time taken: 2794 sec
```

Finally, the updated parameters look like:

```
{'max_depth': 7,
 'min_child_weight': 6,
 'eta': 0.2,
 'subsample': 1.0,
 'colsample_bytree': 0.6,
 'eval_metric': 'rmse',
 'objective': 'reg:squarederror'}
```

Now we retrain our model with the updated parameters and see RMSE value.

	Train	Test
RMSE:	125.626	172.927
R2:	0.884	0.687

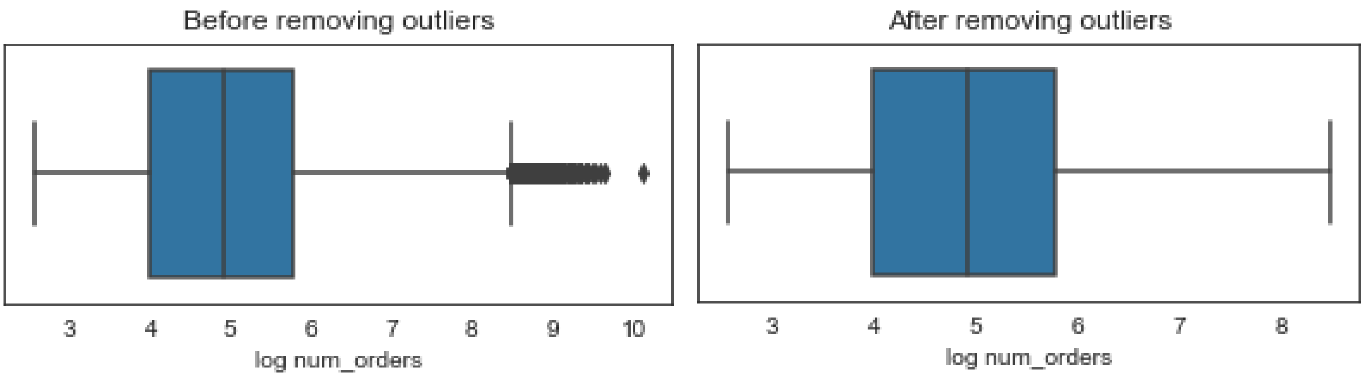
With proper hyper-parameter tuning we could decrease the error from 183.77 to 172.24, which is a good progress. Now to further decrease the error we went through some feature engineering.

6.1 Feature engineering:

We began with removing outliers from the “num_orders”. Quartile1 ($q1$), quartile2 ($q2$) and Inter Quartile Region (iqr) on log of “num_orders" are being calculated and the points which are less than $q1 - 1.5 * iqr$, and points which are greater than $q2 + 1.5 * iqr$ are being removed.

Log is being taken on the “num_orders” as it is exponentially distributed which is right skewed hence there are more outliers on the right side of the graph which we have already seen from theboxplot. This may lead to removal of non-outlier points from the dataset so it is necessary to take log to convert the data into normal like distribution before removal of outliers.

This is how the boxplot of log of “num_orders” looks before and after removing outliers.



Outliers for the feature “base_price” and “checkout_price” are also being removed in the same way. Finally, we have 456241 data points after removing 306 outliers.

Feature named “discount_precentage” is formed and discount% is calculated for each data pointwith the formula

Columns “checkout_price” and “base_price” are of no use now. Hence are removed. Column “id” containing unique values of each order would not give much value to our model so it is also beingremoved. Columns “city_code”, and “region_code” are also removed.

Now we are left with 456241 rows and 29 columns of data.

6.2 Fine tuning:

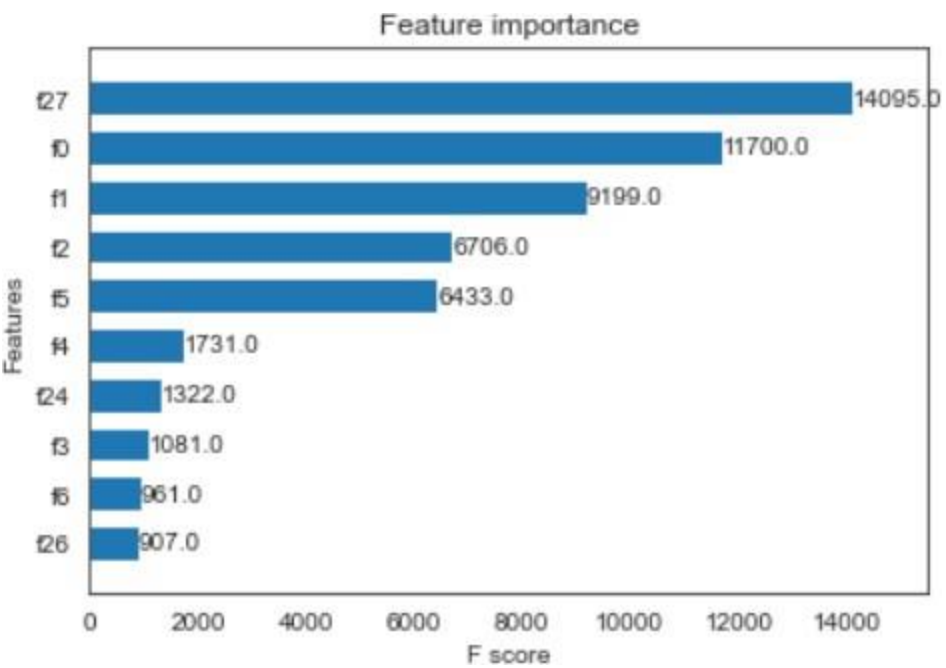
After the completion of feature engineering, we fine tune the XGBoost model to get the best parameters for these newly engineered features. Fine tuning is done the same way we did hyper-parameter tuning but with newly formed dataset. Below shows are the parameters we got after fine tuning.

```
{'max_depth': 10,
 'min_child_weight': 7,
 'eta': 0.1,
 'subsample': 1,
 'colsample_bytree': 0.7,
 'eval_metric': 'rmse',
 'objective': 'reg:squarederror'}
```

We finally retrain the model with the above parameters and got RMSE= , which is reduced by %. Thus by giving some time we increase our models performance by just hyper-parameter tuning. The model is saved as a pickle file.

	Train	Test
RMSE:	108.388	146.267
R2:	0.909	0.804

The following shows feature importance graph for the first 10 features.



RECOMANDATIONS

7.0 Recommendations

The future of online food delivery is exciting, and to ensure that the world develops in a very sustainable manner that serves the interests of all stakeholders involved, we tend to still reflect on what's happening, and question whether things can be done better.

AI and machine learning are used to deliver up-to-date food sector insights, projections, and new food trends by analyzing billions of data points. This allows food delivery companies to adapt new flavors and ingredients that customers want and successfully meet trending needs.

Estimate Delivery Time: In the online food delivery industry, machine learning allows businesses to quantify the time spent on previous deliveries, which aids in future deliveries to customers.

CONCLUSION

8.0 Conclusion

Demand forecasting is critical in restaurant management's operations planning. The dataset loaded from freshmenu is analyzed, understood, and fitted into various machine learning algorithms, with the benefits and drawbacks of each discussed. No single method is best in every case, according to the forecasting literature; however, we extensively fine-tuned the XGBoost algorithm to give the lowest possible mistakes in order to provide the best predictions. Throughout the process, we discovered that the meal demand is highly dependent on the meal discount, the type of meal, and the location of the store; promotions on web pages and emails improve on this.

8.1 **Limitations:**

The system predicts with an accuracy of 80.4 percent; it can be improved with more feature engineering and the assistance of domain experts.

Because there are so many hyperparameters in XGBoost, tuning them all takes hours. And if not done carefully, it could go wrong.

BIBLOGRAPHY

Hirschberg, C., Rajko, A., Schumacher, T., & Wrulich, M. (2016). The Changing Market for food delivery.
(http://dln.jaipuria.ac.in:8080/jspui/bitstream/123456789/2874/1/The-changing-market-for-food-delivery.pdf)

https://www.statista.com/outlook/dmo/eservices/online-food-delivery/india#key-players

Khan, M. A., Saqib, S., Alyas, T., Rehman, A. U., Saeed, Y., Zeb, A., ... & Mohamed, E. M. (2020). Effective demand forecasting model using business intelligence empowered with machine learning. *IEEE Access*, *8*, 116013-116023.

Kolassa, S., & Siemsen, E. (2016). *Demand forecasting for managers*. Business Expert Press.

Shim, J. K. (2000), Strategic business forecasting: the complete guide to forecasting real world company performance. CRC Press.
(https://books.google.com/books?hl=en&lr=&id=tVtdbIHTy-QC&oi=fnd&pg=PA1&dq=Shim,+J.+K.+(2000).+Strategic+business+forecasting:+the+complete+guide+to+forecasting+real+world+company+performance.+CRC+Press&ots=a2mVvgp14e&sig=eWecvXIuazLsrQ5oQW_XX7oV2uE)

Clarke, S. (2006). Transformation Lessons from Coca-Cola Enterprises Inc.: Managing the Introduction of a Structured Forecast Process. *Foresight: The International Journal of Applied Forecasting*, (4), 21-25. (https://ideas.repec.org/a/for/ijafaa/y2006i4p21-25.html)

Paták, M., Branska, L., & Pecinova, Z. (2015). Demand forecasting in retail grocery stores in the Czech Republic. In *2nd International Multidisciplinary Scientific Conference on Social Sciences & Arts SGEM* (pp. 693-700).
Url(https://www.researchgate.net/profile/MichalPatak/publication/312208849_DEMAND_FORECASTING_IN_RETAIL_GROCERY_STORES_IN_THE_CZECH_REPUBLIC/links/590c353aa6fdcc5d421ee75e/DEMAND-FORECASTING-IN-RETAIL-GROCERY-STORES-IN-THE-CZECH-REPUBLIC.pdf)

Ravisankar, S., Mahendran, K., Arulmurugan, S., & Sumalatha, M. R. (2022). Flexible Demand Forecasting in Intelligent Food Supply Chain Management. *Available at SSRN 4119151*. Url(https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4119151)

Pramudita, K. E. (2020). Designing a Demand Forecasting Service in a Food-delivery Platform. URL(https://www.diva-portal.org/smash/record.jsf?pid=diva2:1554149)

Llanes, R. P., Sala, H. V., & García, A. O. (2020). Models for predicting perishable products demands in food trading companies. *Revista Cubana de Ciencias Informáticas*, *14*(1), 110-135. (https://www.redalyc.org/journal/3783/378365895005/378365895005.pdf)

Know The Best Evaluation Metrics for Your Regression Model! By **Raghav Agrawal** — **May 19, 2021**(Evaluation Metrics for Your Regression Model - Analytics Vidhya)

Evaluation Metrics for Regression models- MAE Vs MSE Vs RMSE vs RMSLE by march 20, 2019 (Evaluation Metrics for Regression models- MAE Vs MSE Vs RMSE vs RMSLE (akhilendra.com).

https://xgboost.readthedocs.io/en/latest/tutorials/model.html

https://www.aionlinecourse.com/tutorial/machine-learning/evaluating-regression-models-performance

https://www.statisticshowto.com/probability-and-statistics/regression-analysis/

https://www.relexsolutions.com/resources/demand-forecasting/

ABOUT THE AUTHOR

Karunashree Nagaraj graduated with Bachelor of Computer Applications, Acharya Institutes, India and Completed her MBA in Operation Management from REVA university, India. Following her passion analytics, she started to purse her master's degree in Artificial Intelligence and Business Analytics from Toulouse Business School, France.