**Food Demand Prediction**

Amulya Akinapuram (11558530)

(Email: [AmulyaAkinapuram@my.unt.edu](mailto:AmulyaAkinapuram@my.unt.edu))

**Abstract**

The demand forecasting/prediction in one of the key factors used in any growing business. In the food industries or restaurants, it is important to properly balance the stock and the sales of the items. By forecasting or predicting the sales of the restaurant one can predict the flow in the restaurant and by understanding the sales of items one can buy the raw materials. By purchasing less than the required amount may cause the business to lose sales. And by purchasing more stock, due to the decay of materials it will be a huge loss for the restaurants. So to balance the stock and consumption one can predict the sales of the products in a restaurant or any business firm to minimize the losses and maximize the profits. Also presents a demand forecasting algorithm that is capable of overcoming the wastage of short-life items. Used Regression algorithms can improve the performance of the forecasting process

**Data Specification**

The dataset taken for this project is from kaggle. The data is taken from 3 different files.By using the dataset we predict the number of sales by using the number of orders in the dataset. The predictor column in the dataset is the number of orders.

There are 3 different .csv files in the dataset i.e

* Food demand dataset
* Meal\_info
* Fulfilment\_center\_info

The Food Demand dataset consists of the following attributes:

1. Food id
2. Week
3. Center id
4. Meal id
5. Price at checkout
6. Base price
7. Number of orders
8. Category
9. Cuisine
10. City Code
11. Region Code

The main objective of the project is to predict the sales of a restaurant by using the above dataset. The key components of the dataset are Food id, Meal id, Base price, Price at checkout. Our target variable is number of orders. And the meal\_info and the fulfilcenter\_info gives us the detailed data about the meal and the area of the center respectively.

**Learning technique:** A machine learning technique known as regression is used to predict the continuous values of a given data set. It aims to plot a best-fit curve or a line between the various data points. The three main metrics used to evaluate the model's effectiveness are error, variance, and bias. We used regression algorithms to predict our model.

**Project Design**

The tools and techniques used in the project to achieve our goal to predict the sales of the food demand in a restaurant are:

1. Linear Regression
2. Gradient Boosting Regression
3. Support Vector Regression
4. Random Forest Regression

The models above are used for the analysis of data and producing the regression models which will be used in predicting the different regression values which can be used in the forecasting of different orders on various cuisines.

1. **Linear Regression:** The linear regression is used to build the relationship between the different variables i.e., dependent and independent variables. This helps us to find out the factors which affect the sales in the data. By analyzing the patterns, we can predict the orders based on the factors which affect it.
2. **Gradient Boosting Regression:** Gradient boosting Regression determines the discrepancy between the current prediction and the known correct target value. Relative refers to this discrepancy. Following that, a weak model that translates features to that residual is trained using gradient boosting regression.
3. **Support Vector Regression:** In order to predict discrete values, Support Vector Regression is a supervised learning approach. The same idea behind SVMs underlies support vector regression. To locate the line with the best fit is the fundamental tenet of SVR. The hyper plane with the most points in SVR is considered to be the greatest fit line.
4. **Random Forest Regression:** A supervised learning algorithm known as Random Forest Regression uses an ensemble learning method. This method combines the predictions made by different machine learning algorithms to provide a more accurate prediction.

**Performance Metrics:**

**Mean-Squared-error:**

The MSE is a measure of the effectiveness of a model by taking the difference between its predictions and the ground truth.

**Mean-Absolute-Error:**

Mean absolute error is calculated by finding the absolute difference between the actual values and the predicted values and taking the mean or average of those values. By doing so we can find the mean absolute error.

**Root- Mean-Squared-error:**

The error of a model in predicting quantitative data is typically measured using the Root-Mean-Squared-Error (RMSE). The square root of the residuals' variance yields the RMSE. It shows how closely the observed data points match the values predicted by the model, or how well the model fits the data in its entirety. RMSE is an absolute measure of fit whereas R-squared is a relative measure.

**R-Squared:**

R-Squared value is always between 0 and 1 (0% and 100%). Many data points are close to the linear regression function line if the R-Squared value is high. The linear regression function line does not adequately match the data if the R-Squared value is low.

**Accuracy:**

The accuracy ratio is a measure of how accurately a case is classified. The best value is 1 while the worst one is 0.

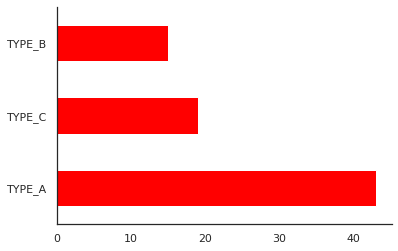
Performed all the above metrics to compare the 4 models of regression. Below are the result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Models** | MSE | MAE | RMSE | R-Squared |
| Linear Regression | 163393.50 | 195.19 | 404.22 | 0.14 |
| Gradient Boosting  Regression | 129247.91 | 156.06 | 359.51 | 0.32 |
| Support vector Regression | 201456.23 | 193.96 | 448.83 | -0.06 |
| Random  Forest  Regression | 15658.44 | 50.15 | 125.13 | 0.92 |

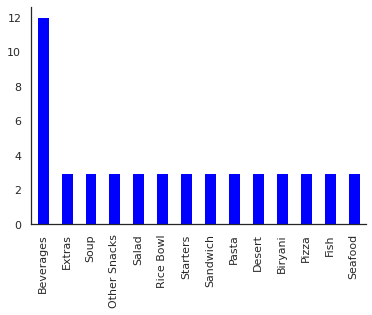
**Data Visualization:**

A data visualization is a type of graphical representation of the data collected. It can help people understand the various trends and patterns in the data.

**Bar plot:**

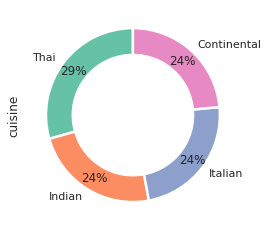


The above plot shows the different types of centers and their count. It shows the most restaurants present in certain city type. By plotting it we can the most profitable centers and can access the sales of the restaurants in the particular center.



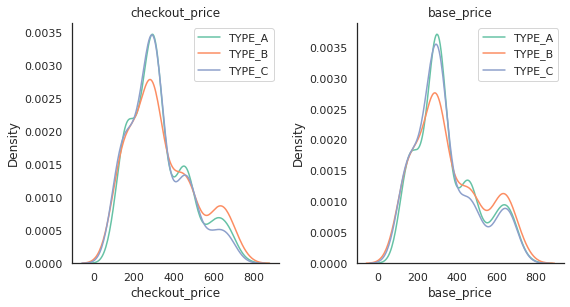
The above plot depicts the count of various categories of products consumed on average. By the above plot we can say that the beverages are the most ordered or the most liked product among other categories. By analyzing the plot, we can predict the different consumption requirements of the customers and can collect the stock according to the requirements.

**Donut chart:**

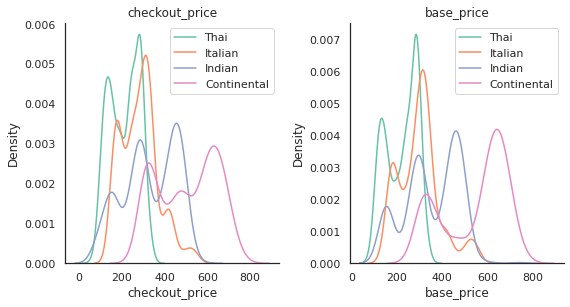


The above doughnut chart shows the various cuisines and the percentage of orders done on by the customers on type of cuisine. This shows the requirements of the customers and their interest in those cuisines. This helps us understand the requirement of the stock which can be imported based on the consumption.

**Line Plot:**



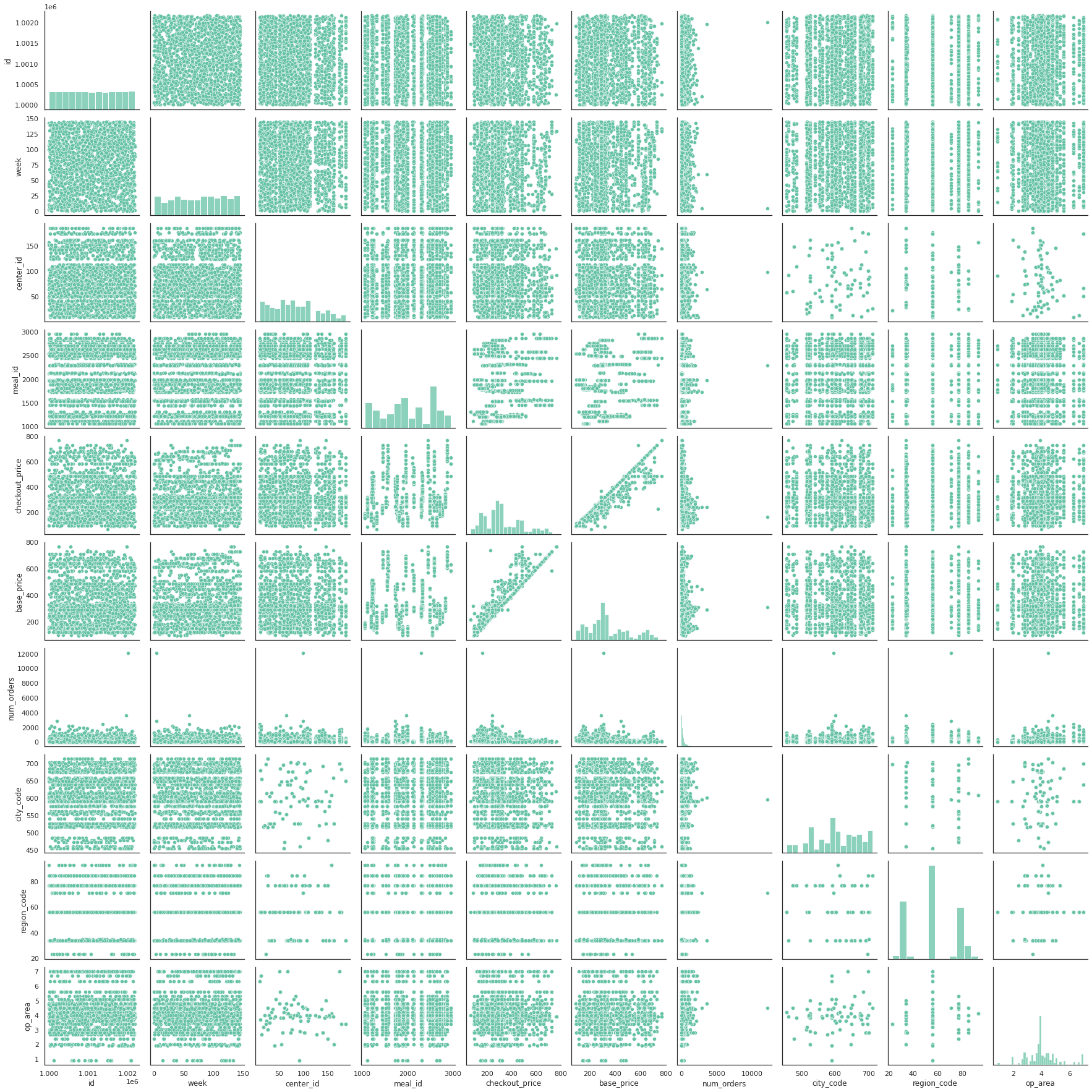
The above plot is based on the prices according to different center types. The above plot has both base price and the checkout price of the products on an average in different centers. By analyzing the plot, we can understand the total profit and loss in different centers. By analyzing the profits and losses we can understand the flow in certain centers and can predict the margin of each center.



The above plot is based on the prices according to various cuisines. The above plot has both base price and the checkout price of the products on an average in different cuisines. By analyzing the plot, we can understand the total profit and loss on cuisines. By analyzing the profits and losses we can understand the flow in certain centers and can predict the margin of each cuisine which can help us in collecting the stock on different cuisines.

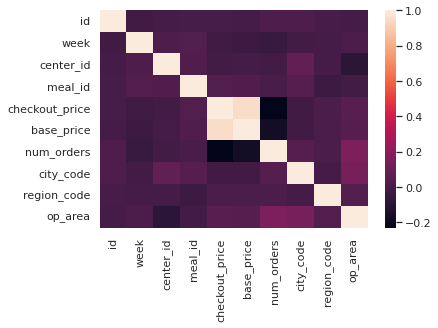
**Pair Plots:**

A pair plot is a statistical method that allows us to see the distribution of single and two variable relationships. It can also identify trends in follow-up analysis. For our dataset the variables relation is shown below how it is distributed according to other(variable).



**Correlation:**

The correlation coefficient is a measure of how two or more variables move in relation to one another. A positive correlation means that the two are moving in the same direction, while a negative correlation shows that the relationship is negative. A zero correlation means that there is no linear correlation.



From the above correlation which shows that number of orders is highly correlated with op\_area, meal\_id and city\_code.

**Project Milestones:**

**Week 1:** Explored various research papers and the one which intrigued us was-Food Demand Prediction. Collected the dataset from Kaggle.

**Week 2:** Preprocessing the data and Performed Data visualization.

**Week 3:** Training the dataset into train and test. Parallelly, we have worked on Presentation as well.

**Week 4:** Built four regression model and performance metrics for each model and their comparisons then performed hyperparameter tuning for best model to improve its performance.

**Week 5:** Lastly, we worked on Project report and also the final presentation.

**Project Results:**

We analyzed the dataset, merge the 3 datasets to get required results. Performed data visualizations. Encoded categorical variables. Trained the data to build 4 different models i.e., Linear Regression, Gradient Boosting Regression, Support Vector Regression and Random Forest Regression. To compare the model, we used performance metrics such as Mean Squared Error, Root Mean Squared Error, Mean Absolute Error and R-Squared Score. By observing this metrics, we conclude that Random Forest Regression is best model. And also performed hyperparameter tuning using RandomizedSearchCV got a RMSE is 182 and improvement from the previous Random Forest Regressor model.

**Repository / Archive:**

<https://github.com/amulyaakinapuram/Food_demand_prediction/blob/main/Food_demand_Prediction.ipynb>

**Reference:**

* Dataset - https://www.kaggle.com/code/swetarajsinha/food-demand-forecasting/data
* Code For Reference - https://www.kaggle.com/code/swetarajsinha/food-demand-forecasting/notebook
* Dataset: https://www.kaggle.com/datasets/gauravsahani/food-demand-prediction-dataset
* Aishwarya,K et al. (2020).Food Demand. International Research Journal of Engineering and Technology. <https://tinyurl.com/23hdxatw>

**How this model is different from reference:**

Performed Hyperparameter tuning to improve our model performance.

And The model will be able to predict the future needs of demand for the food without the wastage of stock item.