DATE: 1 FEB 2021

AI Product/Service Prototype Ideation

(AI in economy)

AI FOR DEMAND FORECASTING IN FOOD INDUSTRY

REPORT PREPARED BY:

Surajkumar Sondagar

(ML INTERN AT FEYNN LABS)

INSTITUTE NAME:

L.D College Of Engineering, Ahmedabad

TABLE OF CONTENTS

1. Problem Statement
2. Market/Customer/Business Need Assessment
3. Target Specifications and Characterization
4. External Search
5. Bench marking alternate products
6. Applicable Patents
7. Applicable Regulations
8. Applicable Constraints
9. Business Opportunity
10. Concept Generation
11. Concept Development
12. Final Product Prototype
13. Product details
    - How does it work?
    - Data Sources
    - Algorithms, frameworks, software etc. needed
    - Team required to develop.
    - What does it cost? etc
14. Code Implementation/Validation on Small Scale

    What can be included?
    - Some Basic Visualizations on Real World or Augmented Data
    - Simple EDA
    - ML Modelling
    - Github link to the code implementation
15. Conclusion

INTRODUCTION

The success of a restaurant not only depends on taste, ambience 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

This Report describes the approach for forecasting methods using machine learning and statistical analytics.

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:

Weeks 1 to 145
Historical data of demand for a product-center combination

Information for fulfillment center
Center area, City Information etc.

Product(Meal) features
Category, Sub-category, Current Price and Discount

MARKET & BUSINESS NEED ASSESSMENT

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.

In order to solve such problems, we are researching how to predict the demand. Here we are researching food demand forecasting methods using internal data such as number of orders.

Business Benefits

The replenishment of raw materials is done only on weekly basis and since the raw material is perishable, the procurement planning is of utmost importance.

Therefore predicting the Demand helps in reducing the wastage of raw materials which would result in the reduced cost of operation. Increased customer satisfaction by timely fulfilling their expectations and requirements.

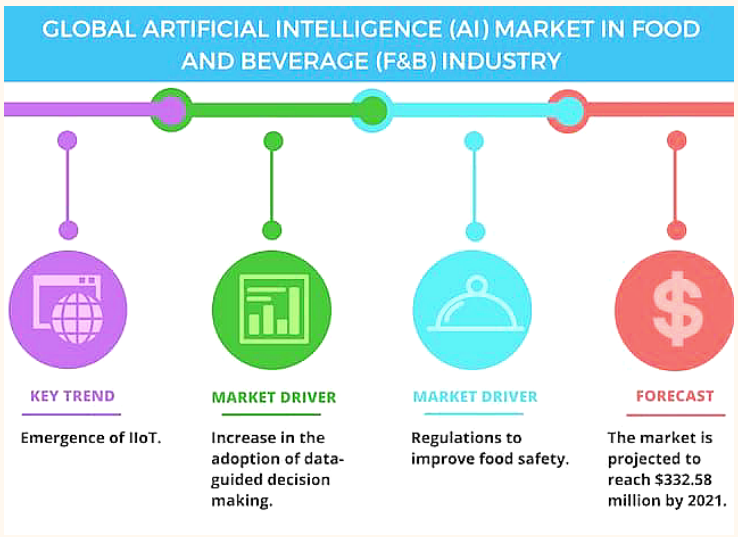
"Demand is an economic principle referring to a consumer's desire to purchase goods and services and willingness to pay a price for a specific good or service"

TARGET SPECIFICATIONS AND CHARACTERIZATION

Food Storage & Retention:
By analyzing Centre's Data as per stocks of weeks along with prices and Demand forecasting will have a significant influence on stock control management. It can help vendors better anticipate the purchasing habits of their customers and keep them satisfied.

For better results, Regressor methods should be used because they can accommodate as many features as possible in order to develop a generalized system. If we move in a more advanced direction, we may be able to mimic and understand exactly what the individual wants today, even if it is not in the dataset's pattern.

Increase in Demand/Supply Ratio
Using machine learning and data analysis Marketplace can visualize which products are being in Demand more and ordered frequently and which is then allocated numbers of orders to fulfilled. So they can focus on which items to preserve more for their inventory, so their supply management of loadout and finances are improved.



EXTERNAL SEARCH:

There are several datasets we have considered in order to do forecast the demand. We have taken the data from Kaggle, under the data of food demand forecasting. We used three csv files, fulfillment, train and meal info. We referred several other sites Benefit of AI in restaurant industry, Applying ML to demand forecasting, AI revolutionizing food processing business and even reviewed some papers as cited:

- Tanizaki, Takashi, et al. "Demand forecasting in restaurants using machine learning and statistical analysis." Procedia CIRP 79 (2019): 679-683.

- Lasek, Agnieszka, Nick Cercone, and Jim Saunders. "Restaurant sales and customer demand forecasting: Literature survey and categorization of methods." Smart City 360° (2016): 479-491.

- FOOD DEMAND PREDICTION USING MACHINE LEARNING K.Aishwarya1, Aishwarya.N.Rao2, Nikita Kumari3, Akshit Mishra4, Mrs.Rashmi M R5

BENCH MARKING ALTERNATE PRODUCTS:

There are several benchmarking models, and many large corporations have used food demand forecasting systems, well, why not? There are several advantages to using it. Here are a few examples:

Delicious Data offers Artificial Intelligence-based Demand Forecasting:

Delicious Data is a German firm that creates forecasting solutions for procurement and manufacturing using machine learning. Their service aids bakeries in forecasting sales in order to take demand-driven action and optimizing product display positioning in order to attract customers. The programmed estimates sales down to the minute, allowing staff to precisely optimize product manufacturing. As a result, there will be less food waste and overproduction, resulting in higher profitability.

Intuendi develops a demand forecasting platform that is cloud-based:

Intuendi is an Italian business that specializes in offering a cloud-based demand forecasting tool that uses artificial intelligence to improve inventory planning. Their system automates the data collecting and forecasting process from many sources, allowing them to derive insights from sales data. Promotions, stock shortages, stockouts, calendar events, seasonality, and other factors are all taken into consideration by the programmed.

The firm assists food companies in optimizing daily purchase orders, perishable item management, cost of dead inventory, and other aspects of their operations.

Alloy offers a Demand Planning Solution:

In the food industry, planning and execution of supply chains are essential to meet demand. However, in most cases, due to the lack of accurate demand forecasting, food inventories are overstocked and lead to food wastage. To avoid this, startups are creating solutions that use point of sale (PoS) data to build models that accurately predict the level of demand.

Digitory provides Demand Forecasting Platforms:

Companies and businesses in the food industry allocate a higher amount of their budget towards maintaining an excess of food inventory. This is done to ensure that there is always sufficient inventory to meet demand at any given time. Currently, managers estimate demand largely through experience and intuition, which results in inaccurate spending on inventory. For this reason, startups are providing forecasting models to optimize cash flows in inventory management.

Indian startup Digitory specializes in artificial intelligence (AI) and machine learning (ML) platforms to manage costs and profitability for the food industry. Their sales forecasting solution uses advanced computational techniques to identify sales patterns and create accurate models for prediction. Their solution increases the profits of restaurants through the optimization of production and inventory planning.

PredictHQ builds Event-Driven Demand Forecasting Software:

Grocery stores and supermarkets maintain a large range of products in high volumes to cater to diverse consumer demands. However, the demand for all products is not the same and some are subject to seasonality and fluctuations in demand. Apart from these, unscheduled events and other external factors negatively impact stock levels.

US-based startup **PredictHQ** develops demand forecasting solutions that use internal and external data to learn and adapt its models.

APPLICABLE PATENTS:

- Real time demand prediction in a fast service restaurant environment by Kerien W. Fitzpatrick, R. Craig Coulter, Henning M. Pangels (US6842719B1) in 2003

- Real-time demand prediction in a fast service restaurant environment by Abhishek Sharma, Robert Craig Coulter, Kerien W. Fitzpatrick (US8712822B2) in 2007

APPLICABLE REGULATIONS:

The applicable laws, applicable codes and guidelines can be:

- Customers' data protection and privacy legislation.
- Employment Laws.
- Government regulations for small businesses.
- Food service license.
- Regulations against false advertising.
- Antitrust Regulations.
- Open source, academic and research community License for educational purposes.

APPLICABLE CONSTRAINTS:

- Data Collection from customers and workers
- Lack of technical knowledge for the user (workers)
- Continuous data collection and maintenance.
- Convincing the restaurant holders to implement the system in their restaurant.
- To deliver more precise and dependable findings, a large amount of study is required to compile a worldwide historical record of shipment prices.
- Necessitates current demand pattern and effecting features in order to train the model with current patterns.
- Application development and integration on a continuous basis.

Any project has restrictions and hazards that must be handled in order for the project to be successful in the end. Time, scope, and money are the three key restrictions that project managers should be aware of. These are sometimes known as the project management triangle or the three restrictions.

BUSINESS OPPORTUNITY:

Matching customer tastes with your business strategy

It is definitely an arguable statement, but there is some truth to the fact that data science can help optimize different processes and the way the business operates. The company is relying heavily on data science to predict demand and correctly manage correctly the acquisition of supplies. Basically, it needs to connect information on current food preferences in their menus, customer behavior, and purchasing history to their production. This is a very exciting, demand-driven example of data science in the food industry that could become the blueprint for similar businesses.

Profitable in Ventures

Because the aforementioned approach has only been employed by major corporations, it may now be used to small enterprises, including not only restaurants and takeaways, but also other potential industries. As a result, there's a good probability that this service will be a profitable venture. Every small business that relies on sales may and should use this service to ensure that they are always aware of what their consumers want. As a result, the birth of every small firm represents a rather significant commercial potential for the services we supply.

Introducing new recipes

The number of ways we can combine ingredients in recipes is limitless. Adding the fact that you can cook those ingredients in multiple ways, makes cooking dishes an area of endless possibilities. Today we have huge online databases for recipes, which allows the analysis of ingredients in different cuisines.

Reinventing Food Delivery

Machine Learning Applications in the Restaurant Business

- Analytical solutions for a better customer experience
- Food-selling sites and applications
- AI for online restaurant search
- Restaurant Revenue Prediction Using Machine Learning

The Benefits of AI in the Food Industry:

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

CONCEPT GENERATION:

First and foremost, we began by gaining an understanding of the problem. The restaurant business has long been one of the most popular and active. The increased need in this field, as more people choose to eat outside, has pushed us to make a few advancements as the globe rapidly digitalizes for the betterment.

Why not this sector? There have been numerous ideas previously, and they are still being executed at a basic level. For garbage management, anything from chatbots to smart bins is available. These advantages benefit consumers and businesses straight away, as people visit and receive the desired service, and the firm is able to give it by gaining greater popularity and funds, therefore balancing the market scenario.

The forecasting system is one of the most desirable concepts of all, as the human tendency is to always foresee the future. We implemented this in restaurants as well. It will benefit the business because they will be able to manage the amount of raw materials they need to buy for a specific time and preprepared a meal for that time. As the model's scope expands, it will be able to include more and more elements to make it more appealing to each and every visitor.

CONCEPT DEVELOPMENT:

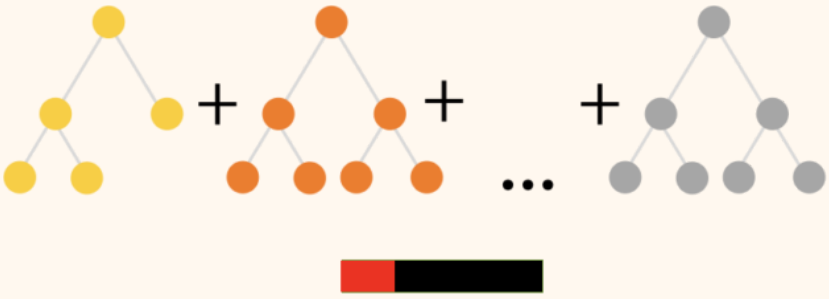
The term CatBoost is an acronym that stands for "Category" and “Boosting.” Does this mean the “Category’ in CatBoost means it only works for categorical features?

According to the CatBoost documentation, CatBoost supports numerical, categorical, and text features but has a good handling technique for categorical data. The CatBoost algorithm has quite a number of parameters to tune the features in the processing stage "Boosting" in CatBoost refers to the gradient boosting machine learning. Gradient boosting is a machine learning technique for regression and classification problems. Which produces a prediction model in an ensemble of weak prediction models, typically decision trees.

Gradient boosting is a robust machine learning algorithm that performs well when used to provide solutions to different types of business problems such as

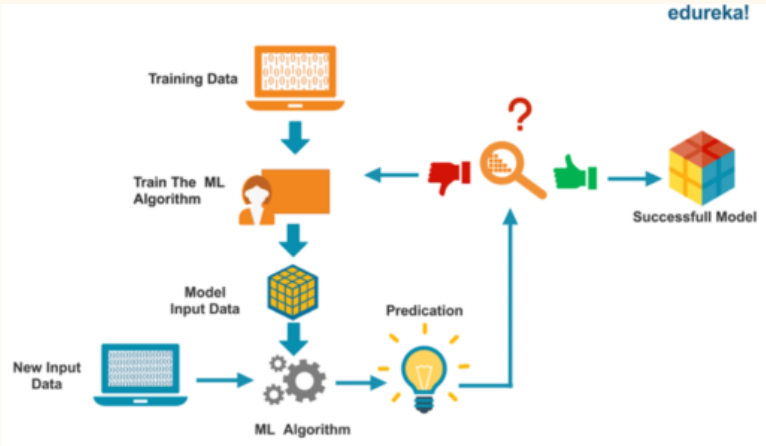
- Fraud detection,
- Recommendation system,
- Forecasting.

Again, it can return an outstanding result with relatively fewer data. Unlike other machine learning algorithms that only perform well after learning from extensive data.
We would suggest you read the article How the gradient boosting algorithms works if you want to learn more about the gradient boosting algorithms functionality.



CatBoosting Model

For the development most important thing we need is data. It can be collected from the Market and restaurant or can be collected by surveying. Then we will be applying Regression methods which is an supervised machine learning technique, that predict the data item maps accordingly. Then using Data cleansing and pre-processing methods data can be classify as for the feature engineering and using EDA technique we acquire hidden insights from the data.

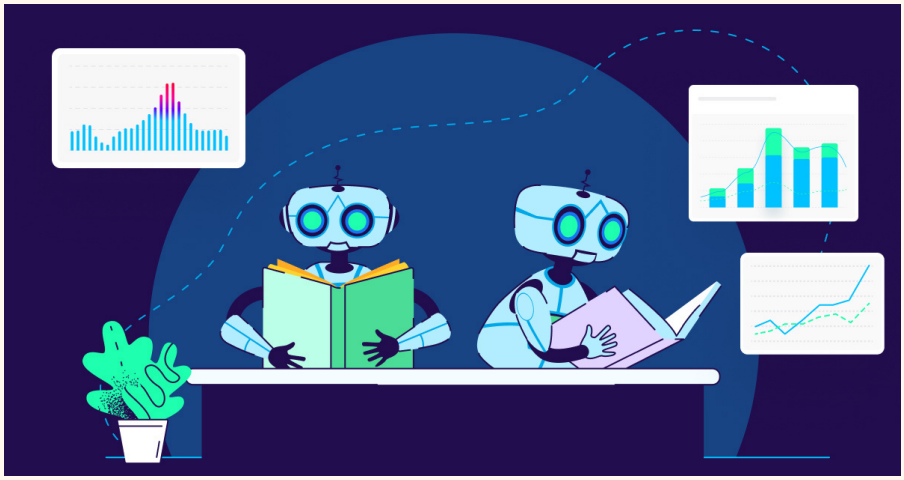


FINAL PRODUCT PROTOTYPE:

Abstract

Demand forecasting 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. In Restaurants, prediction play a vital role as most of the basic ingredients have short-shelf life. The demands depend upon many explicit and hidden context such as season, region etc. In this paper, number of order is used to forecast stock of items, using machine learning with internal and external data. In this we provide an appropriate algorithm for demand forecasting which is capable of overpowering the wastage of short life items. Proposed algorithm like Basic Linear regression model, Boosting models such as XGBoost, LIghtBoost and CatBoost regressor are used that considerably improves the forecasting performance.

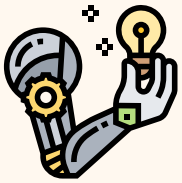
In addition, papers relevant to the research issue were examined. Under different forecasting methods, the results reveal how demand forecasting influences store inventory replenishment decisions and supplier production decisions. The simulation result indicates that the forecasting model used has an impact on the performance and demand patterns experienced by merchants. The findings also assist managers in selecting appropriate forecasting models in order to increase performance.



PRODUCT DETAILS:

What is the mechanism behind it?

Well, it's an interactive user system; if you click on forecast now, it'll predict the order specifics in terms of day, time, weather, and so on, and display the outcome on screen in layman's terms so that employees can grasp it on a daily basis.



What is the source of the data?

The data is given by a meal kit company. As food is perishable, planning and demand prediction is extremely important. For real-time data, they'll need to contact other restaurants that have used this system, or they'll have to build their own in the coming years by capturing all of the data on a data sheet.

Algorithms, frameworks, and software, to name a few. Needed?

Machine learning methods are used in this product. The implementation of this needs the Python programming language. NumPy, pandas, matplotlib, seaborn, Plotly for Visualization and other libraries such as Scikit Learn, XGBoost, Lightgbm, CatBoost can be used for forecasting. For the integrating API, the Flask framework will be utilized. The editors are Visual Studio Code and Google Colab. And for the primary interactive web page, django.

Is a development team required?

This would necessitate the hiring of a project manager, a resource manager, a team leader, and team members such as programmers, data analysts, testers, stakeholders, project sponsors, and business assistants, among others.

How much does it cost?

By arranging each phase, you can determine the overall cost of labour, materials, equipment, services, software, hardware, facilities, contingency charges, and so on.

CODE IMPLEMENTATION ON SMALL SCALE

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

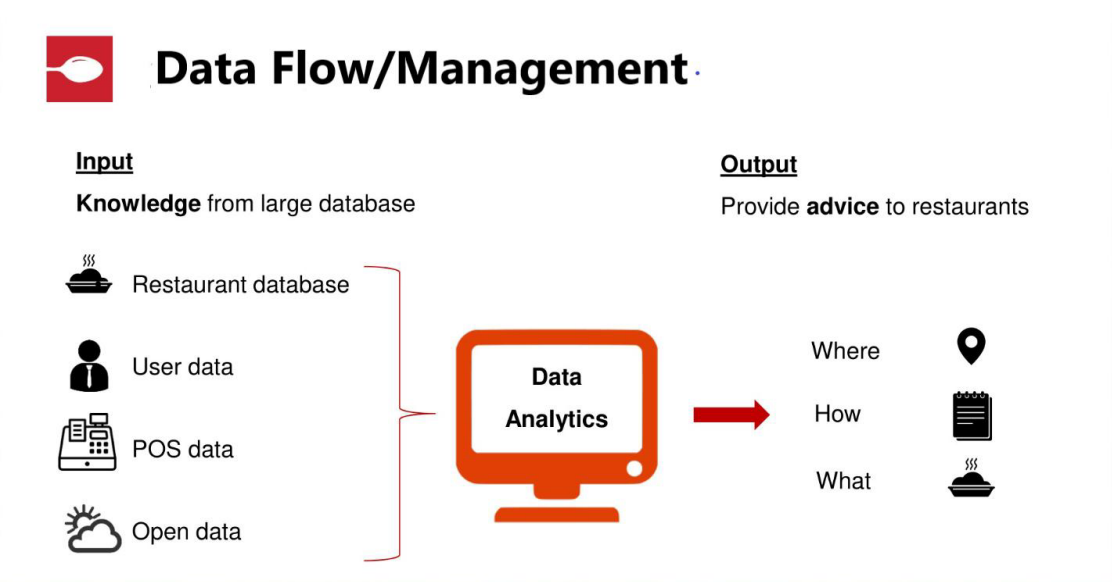
Variable	Definition
id	Unique ID
week	Week No
center_id	Unique ID for fulfillment center
meal_id	Unique ID for Meal
checkout_price	Final price including discount, taxes & delivery charges
base_price	Base price of the meal
emailer_for_promotion	Emailer sent for promotion of meal
homepage_featured	Meal featured at homepage
num_orders	(Target) Orders Count

2. **fulfilment_center_info.csv:** Contains information for each fulfilment center

Variable	Definition
center_id	Unique ID for fulfillment center
city_code	Unique code for city
region_code	Unique code for region
center_type	Anonymized center type
op_area	Area of operation (in km^2)

3. **meal_info.csv:** Contains information for each meal being served

Variable	Definition
meal_id	Unique ID for the meal
category	Type of meal (beverages/snacks/soups....)
cuisine	Meal cuisine (Indian/Italian/...)



Libraries:

Data Pre-Processing:

- There are no Missing/Null Values in any of the three datasets.
- Before proceeding with the prediction process, all the three datasheets need to be merged into a single dataset. Before performing the merging operation, primary feature for combining the datasets needs to be validated.
- The number of Center IDs in train dataset is matching with the number of Center IDs in the Centers Dataset i.e 77 unique records. Hence, there won't be any missing values while merging the datasets together.
- The number of Meal IDs in train dataset is matching with the number of Meal IDs in the Meals Dataset i.e 51 unique records. Hence, there won't be any missing values while merging the datasets together.
- As checked earlier, there were no Null/Missing values even after merging the datasets.

Importing the Libraries

Importing Dependencies & Extracting Datasets

```
import numpy as np
import pandas as pd
pd.set_option('display.max_columns', None)
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
%matplotlib inline
import seaborn as sns
import plotly.io as pio
import plotly.graph_objects as go
from plotly.offline import init_notebook_mode, iplot
from sklearn.preprocessing import StandardScaler
import warnings
warnings.filterwarnings('ignore')
```

```
data=pd.read_csv('train.csv')
center=pd.read_csv('fulfilment_center_info.csv')
meal=pd.read_csv('meal_info.csv')
test=pd.read_csv('test.csv')
```

Dataset shape

There are 4 datasets: Data, Center, Meal, Test.

- Data contains 456548 Rows and 9 columns
- Center contains 77 Rows and 5 columns
- Meal contains 51 Rows and 3 columns
- Test contains 32573 Rows and 8 columns

Data Pre-Processing

```
print("The Shape of Demand dataset :",data.shape)
print("The Shape of Fulmilment Center Information dataset :",center.shape)
print("The Shape of Meal information dataset :",meal.shape)
print("The Shape of Test dataset :",test.shape)
```

```
The Shape of Demand dataset : (456548, 9)
The Shape of Fulmilment Center Information dataset : (77, 5)
The Shape of Meal information dataset : (51, 3)
The Shape of Test dataset : (32573, 8)
```

```
data.head()
```

[4]:

	id	week	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	num_orders
0	1379560	1	55	1885	136.83	152.29	0	0	177
1	1466964	1	55	1993	136.83	135.83	0	0	270
2	1346989	1	55	2539	134.86	135.86	0	0	189
3	1338232	1	55	2139	339.50	437.53	0	0	54
4	1448490	1	55	2631	243.50	242.50	0	0	40

```
center.head()
```

	center_id	city_code	region_code	center_type	op_area
0	11	679	56	TYPE_A	3.7
1	13	590	56	TYPE_B	6.7
2	124	590	56	TYPE_C	4.0
3	66	648	34	TYPE_A	4.1
4	94	632	34	TYPE_C	3.6

```
meal.head()
```

	meal_id	category	cuisine
0	1885	Beverages	Thai
1	1993	Beverages	Thai
2	2539	Beverages	Thai
3	1248	Beverages	Indian
4	2631	Beverages	Indian

Feature Engineering

Feature engineering is the process of using domain knowledge of the data to create features that improves the performance of the machine learning models.

Wi**th the given data, We have derived the below features to improve our model performance.**

- **Discount Amount :** This defines the difference between the "base_Price" and "checkout_price".
- **Discount Percent :** This defines the % discount offer to customer.
- **Discount Y/N :** This defines whether Discount is provided or not - 1 if there is Discount and 0 if there is no Discount.
- **Compare Week Price :** This defines the increase / decrease in price of a Meal for a particular center compared to the previous week.
- **Compare Week Price Y/N :** Price increased or decreased - 1 if the Price increased and 0 if the price decreased compared to the previous week.
- **Quarter :** Based on the given number of weeks, derived a new feature named as Quarter which defines the Quarter of the year.
- **Year :** Based on the given number of weeks, derived a new feature named as Year which defines the Year.

Data After Merging With Derived New features

```
data.head()
```

id	num_orders	city_code	region_code	center_type	op_area	category	cuisine	discount amount	discount percent	discount y/n	compare_week_price	compare_week_price y/n
0	177	647	56	TYPE_C	2.0	Beverages	Thai	15.46	10.151684	1	0.0	0
0	270	647	56	TYPE_C	2.0	Beverages	Thai	-1.00	-0.736214	0	0.0	0
0	189	647	56	TYPE_C	2.0	Beverages	Thai	1.00	0.736052	1	0.0	0
0	54	647	56	TYPE_C	2.0	Beverages	Indian	98.03	22.405321	1	0.0	0
0	40	647	56	TYPE_C	2.0	Beverages	Indian	-1.00	-0.412371	0	0.0	0

```
data.isnull().sum()
```

```
id                     0
week                   0
center_id              0
meal_id                0
checkout_price         0
base_price             0
emailer_for_promotion  0
homepage_featured      0
num_orders             0
city_code              0
region_code            0
center_type            0
op_area                0
category               0
cuisine                0
discount amount        0
discount percent       0
discount y/n           0
compare_week_price     0
compare_week_price y/n 0
dtype: int64
```

Feature Extraction (Train Test Split)

The Shape of Train dataset : (456548, 20)
The Shape of Test dataset : (32573, 20)

```
train=data[data['week'].isin(range(1,146))]
test=data[data['week'].isin(range(146,156))]
```

```
print("The Shape of Train dataset :",train.shape)
print("The Shape of Test dataset :",test.shape)
```

```
The Shape of Train dataset : (456548, 20)
The Shape of Test dataset : (32573, 20)
```

Encoding City

As per our observation from our bar chart of the City against the number of orders. There the high significant difference between the Top 3 cities which has the highest number of orders. Therefore, in our first approach we will encode the City with Highest No. of Orders as CH1, City with 2nd Highest No. of Orders as CH2 and City with 3rd Highest No. of Orders as CH3 and rest all of the cities which does not have much significant differences between the number of orders as CH4.

```
city4={590:'CH1', 526:'CH2', 638:'CH3'}
data['city_enc_4']=data['city_code'].map(city4)
data['city_enc_4']=data['city_enc_4'].fillna('CH4')
```

```
data['city_enc_4'].value_counts()
```

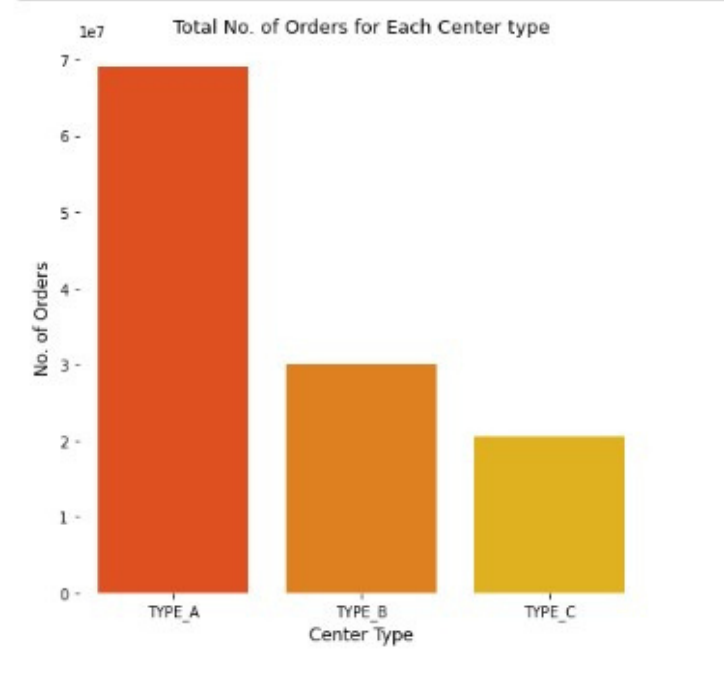
```
CH4    362341
CH1     58708
CH2     46589
CH3     21483
Name: city_enc_4, dtype: int64
```

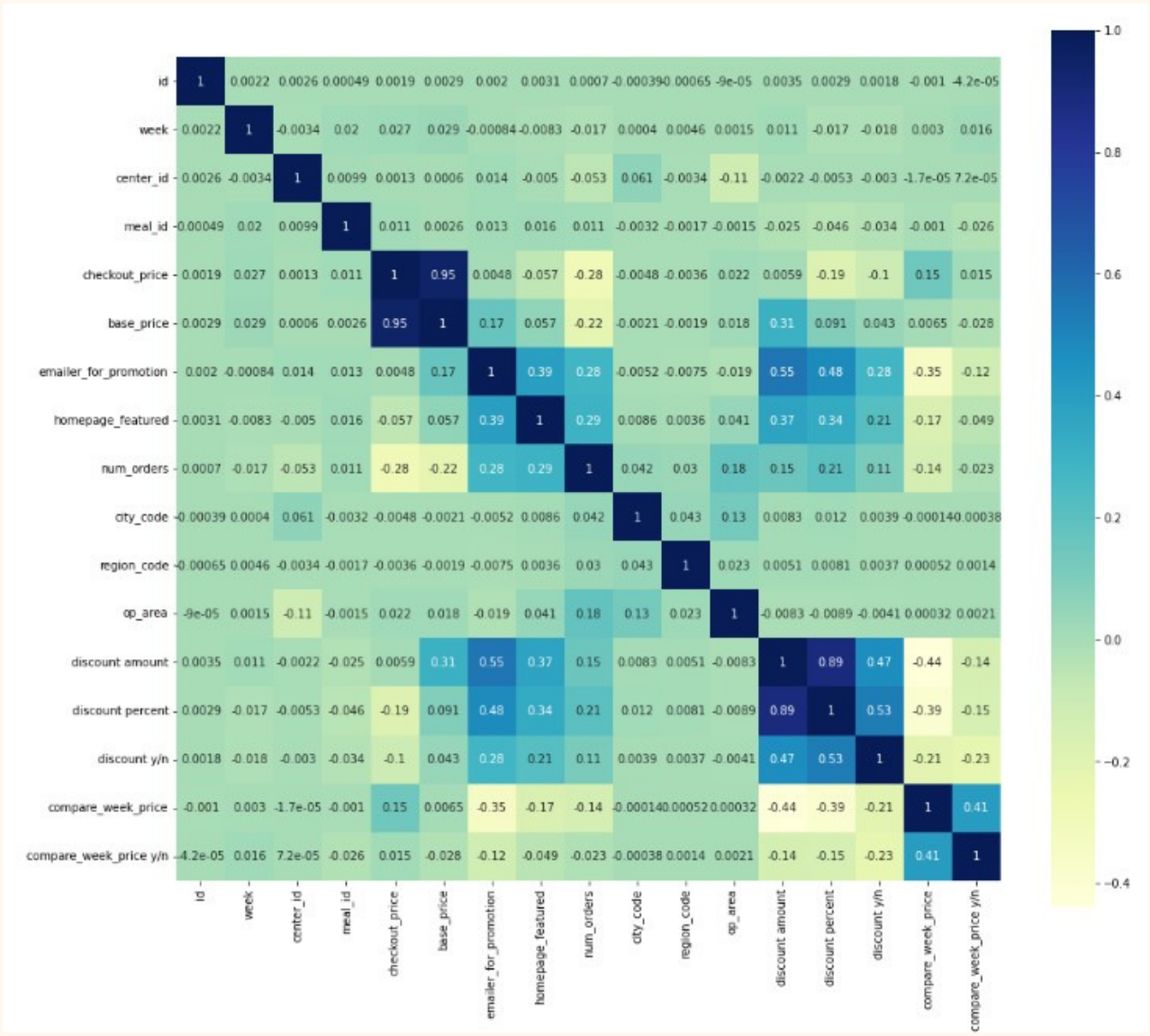
```
data.head()
```

ers	city_code	region_code	center_type	op_area	category	cuisine	discount amount	discount percent	discount y/n	compare_week_price	compare_week_price y/n	city_enc_4
177	647	56	TYPE_C	2.0	Beverages	Thai	15.46	10.151684	1	0.0	0	CH4
270	647	56	TYPE_C	2.0	Beverages	Thai	-1.00	-0.736214	0	0.0	0	CH4
189	647	56	TYPE_C	2.0	Beverages	Thai	1.00	0.736052	1	0.0	0	CH4
54	647	56	TYPE_C	2.0	Beverages	Indian	98.03	22.405321	1	0.0	0	CH4
40	647	56	TYPE_C	2.0	Beverages	Indian	-1.00	-0.412371	0	0.0	0	CH4

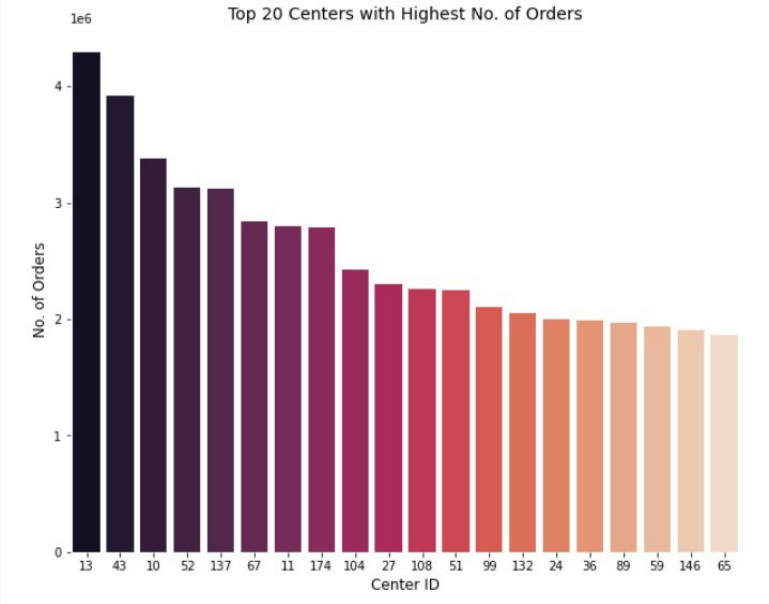
EDA (Exploratory Data Analysis)

Type_A Centers have the highest number of Orders placed and **Type_C** has the least.

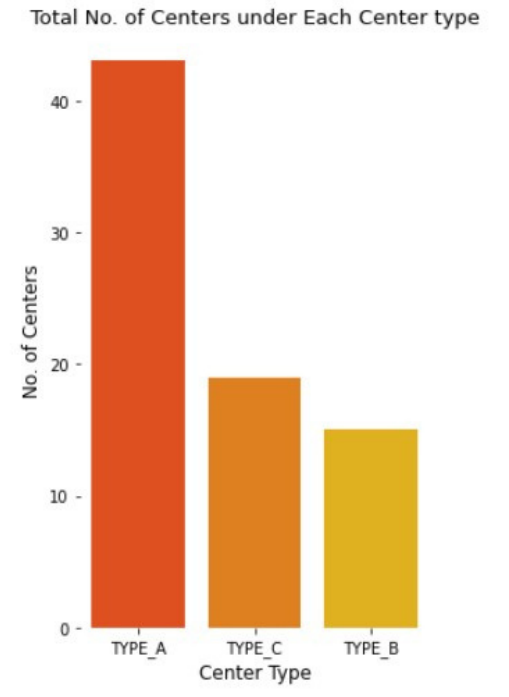




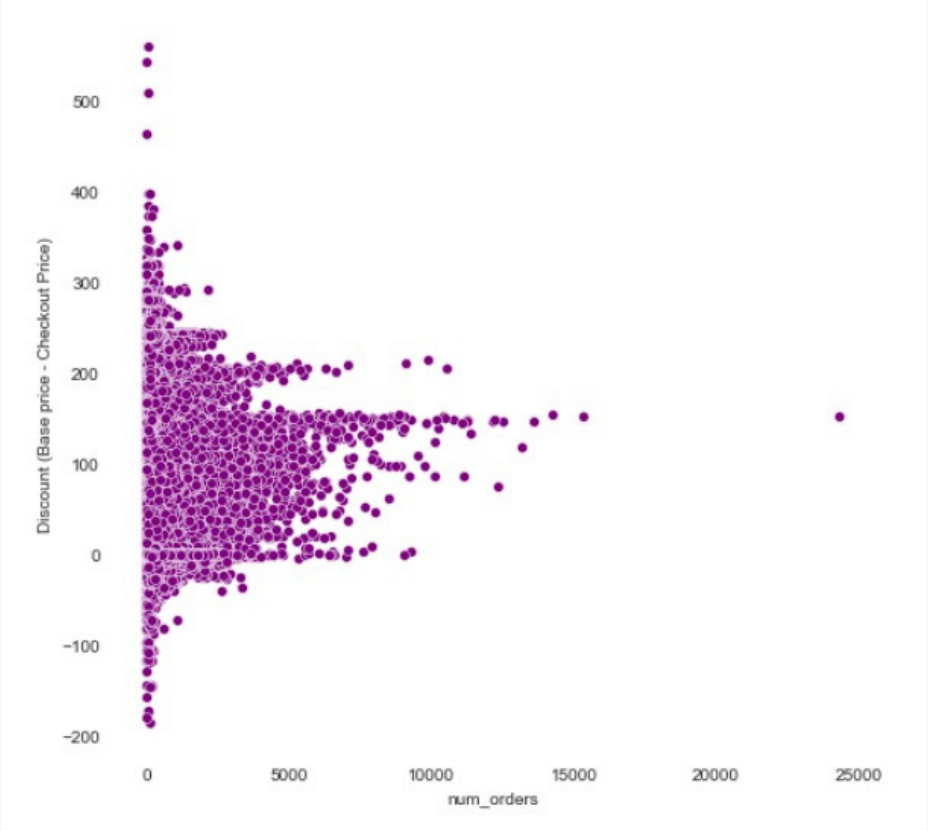
Initially, when we checked, which Center Type has the highest number of Orders, We found that Center **Type_A** has the highest number of orders, but now when we check individually, we could see that Center 13 of **Type_B** has the highest number of Orders.



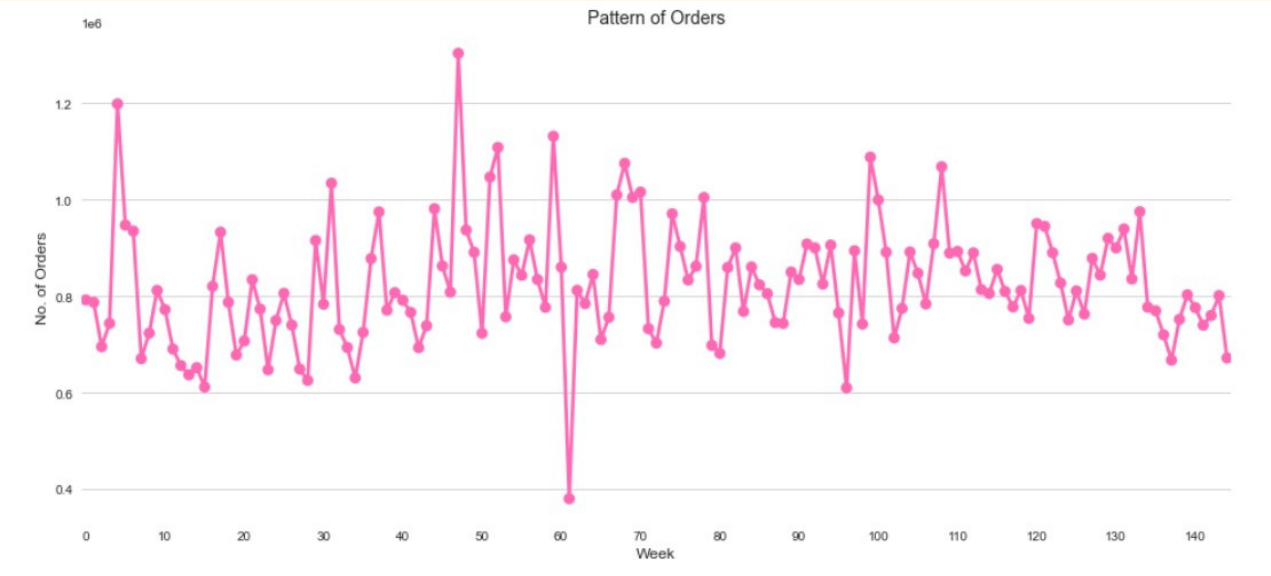
Type_A has the most number of orders because, **Type_A** has the most number of Centers : 43 Centers.



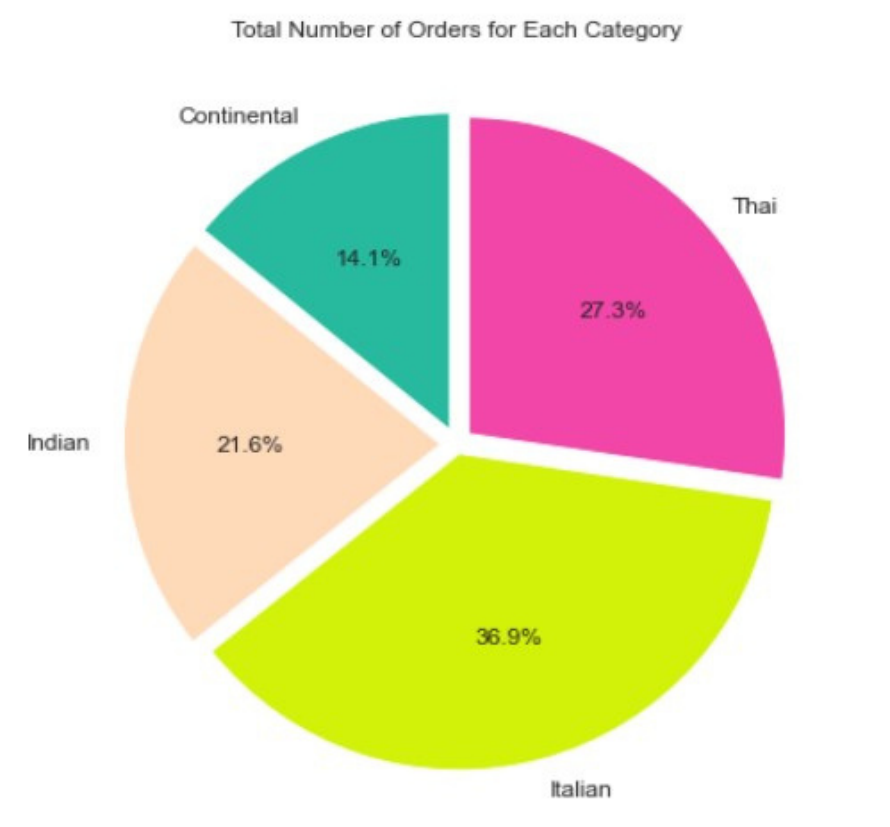
We created a new feature: Discount which is the difference of base price and checkout price and tried to find out if there is any relationship between the discount and the number of orders. But surprisingly there are no good correlation between the discount and the number of orders.



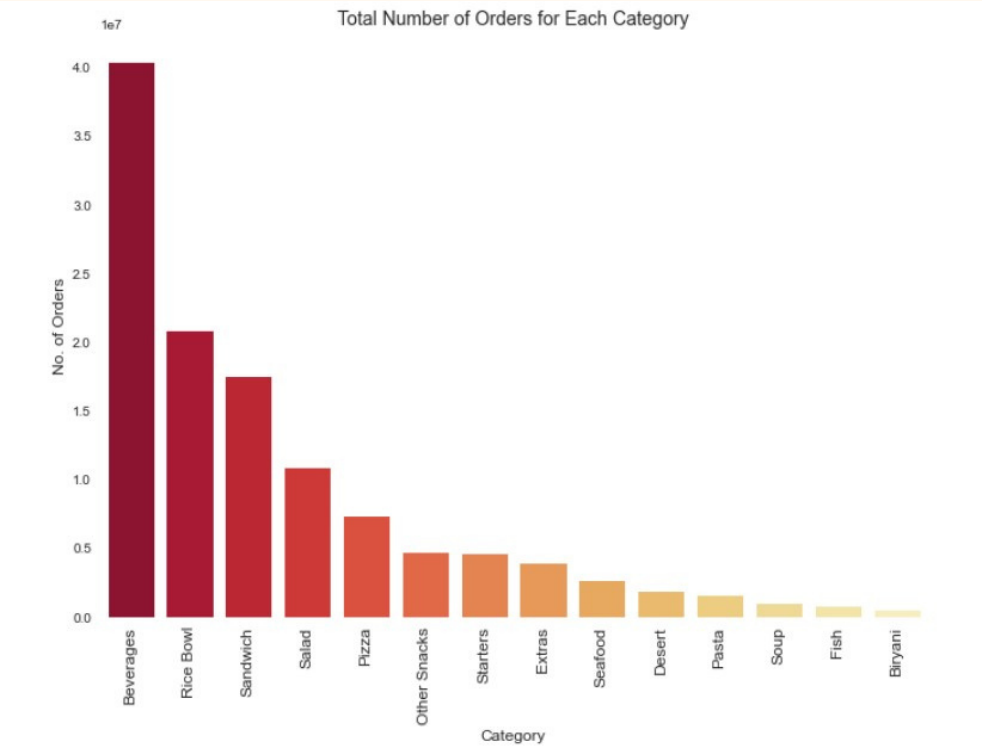
When we analyzed the trend of order placed over the weeks, we could see that the highest number of orders were received in week 48 and the lowest in week 62.



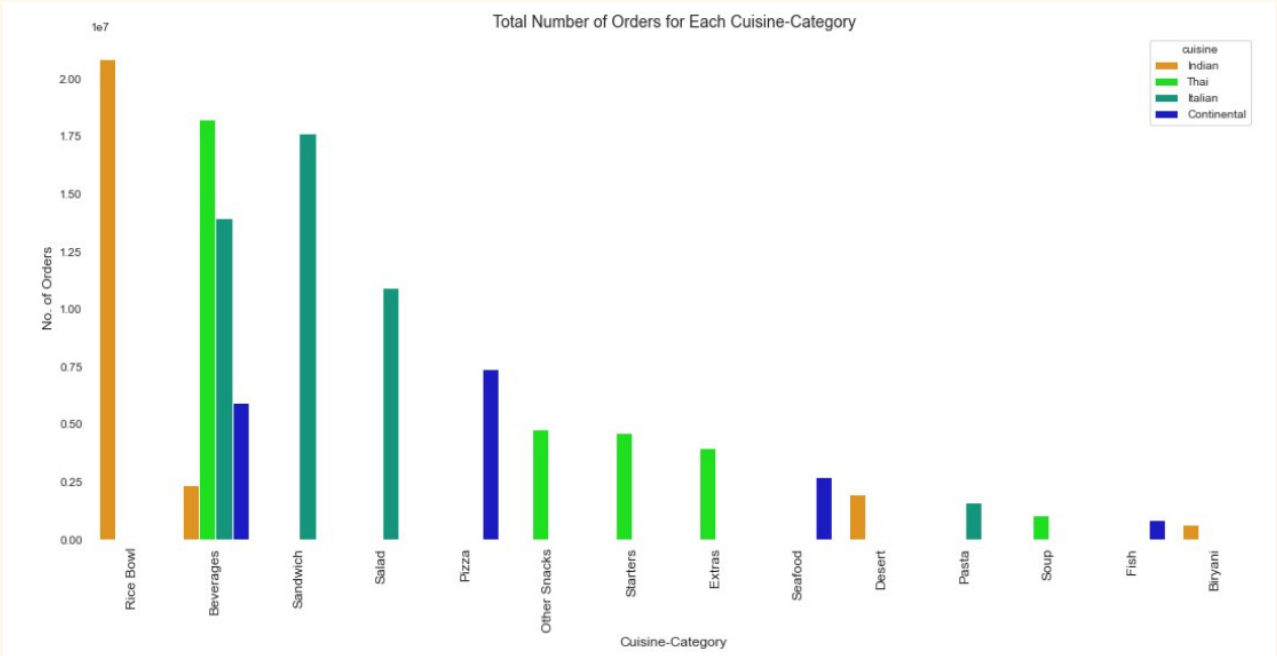
Italian Cuisine has the highest number of orders with Continental cuisine being the least.



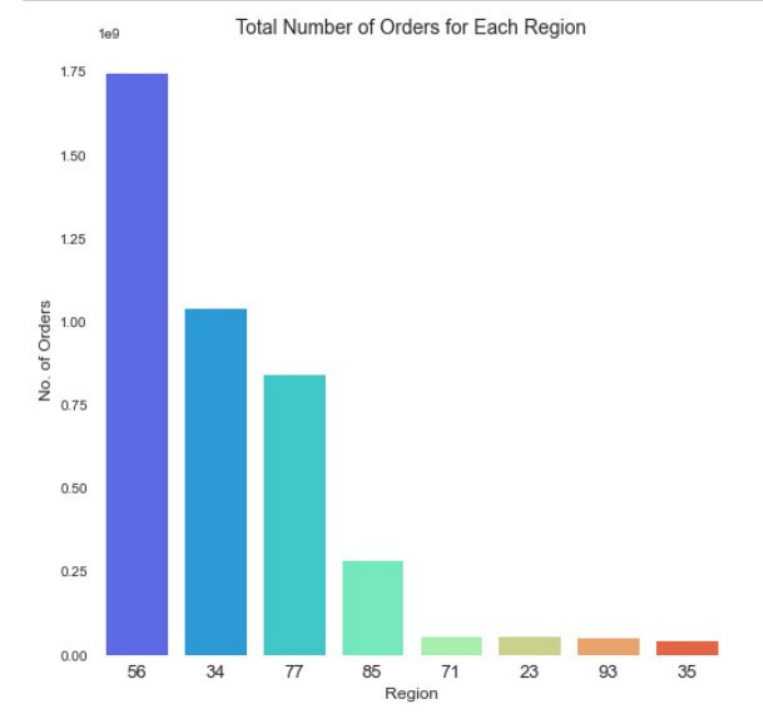
We could see that **Beverages** are the food category which has the highest number of orders and **Biriyani** is the food category with least number of orders.



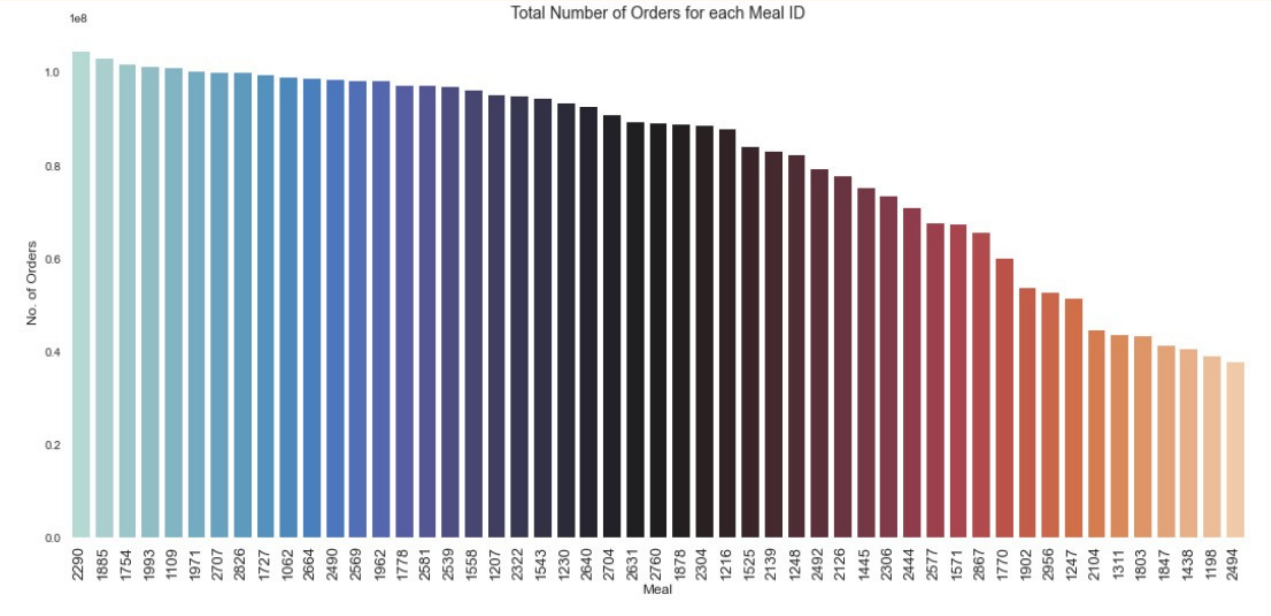
Similarly when we checked which specific cuisine-food category has the highest number of orders, we could see that **Indian-Rice Bowl** has the highest number of orders and Indian-**Biriyani** has the least.



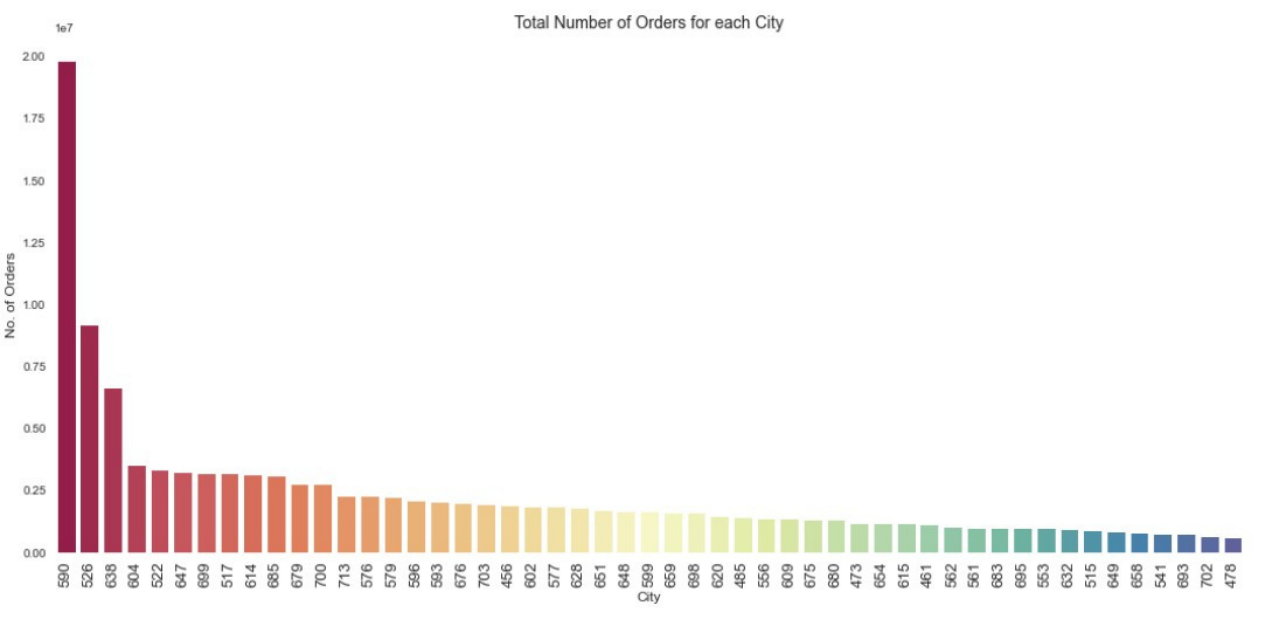
Also when we checked the number of orders with respect to Region, we could see that Region - 56 has the highest number of orders : 60.5M orders which is almost 35M orders higher than the Region with second highest number of orders : Region 34 : 24M orders.



Meal ID 2290 has the highest number of Orders. There is not much significant differences between number of orders for different Meal IDs.



Also when we checked the number of orders with respect to City, we could see that City - 590 has the highest number of orders - 18.5M orders which is almost 10M orders higher than the City with second highest number of orders - City 526 - 8.6M orders.



Data Transformation

- Logarithm transformation (or log transform) is one of the most commonly used mathematical transformations in feature engineering. It helps to handle skewed data and after transformation, the distribution becomes more approximate to normal.
- In our data, the target variable 'num_orders' is not normally distributed. Using this without applying any transformation techniques will downgrade the performance of our model.
- Therefore, we have applied Logarithm transformation on our Target feature 'num_orders' post which the data seems to be more approximate to normal distribution.
- After Log transformation, We have observed 0% of Outlier data being present within the Target Variable - num_orders using 3 IQR Method.

Evaluation Metric

The evaluation metric for this competition is 100*RMSLE where RMSLE is Root of Mean Squared Logarithmic Error across all entries in the test set.

Initial Approach

- Simple Linear Regression model without any feature engineering and data transformation which gave a RMSE : 194.402
- Without feature engineering and data transformation, the model did not perform well and could'nt give a good score.
- Post applying feature engineering and data transformation (log and log1p transformation), Linear Regression model gave a RMSLE score of 0.634.

CatBoost Regressor

```
train=datay[datay['week'].isin(range(1,136))]
test=datay[datay['week'].isin(range(136,146))]


X_train=train.drop(['id','num_orders','week','discount amount','city_code','Quarter_Q2','base_price','discount percent'],
                   axis=1)
y_train=train['num_orders']

X_test=test.drop(['id','num_orders','week','discount amount','city_code','Quarter_Q2','base_price','discount percent'],
                 axis=1)
y_test=test['num_orders']

CGB=CatBoostRegressor(learning_rate=0.3,loss_function='RMSE',max_depth=9,verbose=False)

CGB.fit(X_train, y_train)
CGBpred=CGB.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test,CGBpred))
print("RMSLE :",rmse)
```

```
RMSLE : 0.4968591322290948
```

Advanced Models

- With improvised feature engineering, built advanced models using Ensemble techniques and other Regressor algorithms.
- CatBoost and LightGBM Regressors performed well on the model which gave much reduced RMSLE.
- With proper hyper-parameter tuning, CatBoost Regressor performed well on the model and gave the lease RMSLE of 0.5237

CatBoost Regressor

Model 3 ¶

```
Result=pd.DataFrame(CGBpred)
Result=np.expm1(Result).astype('int64')
Submission = pd.DataFrame(columns=['id', 'num_orders'])
Submission['id'] = test['id']
Submission['num_orders'] = Result.values
Submission.to_csv('Result Cat 1.csv', index=False)
```

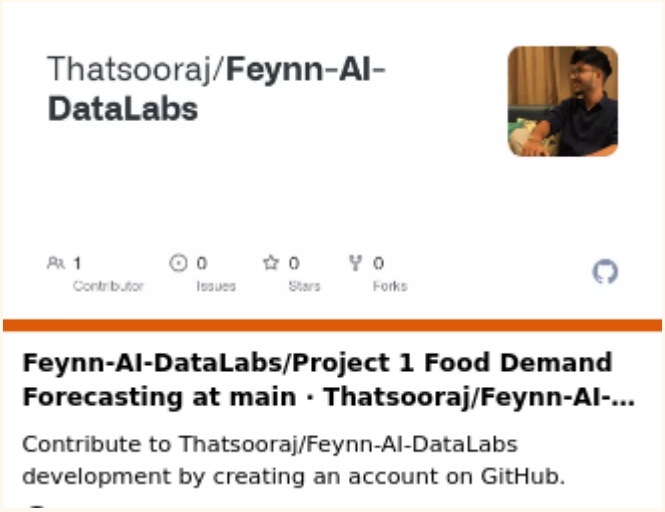
```
Submission.head()
```

[75]:

	id	num_orders
423727	1017495	179
423728	1395634	171
423729	1007493	132
423730	1042952	77
423731	1022147	45

Github: Code Implementation

Double click to interact
Github Link



CONCLUSION

In this Report, we are using external and internal data for the prediction consisting of different factors like region ID, week etc. Food demand prediction is an important and challenging problem.

Demand forecasting is critical in restaurant management's operations planning. Other analyses are based on having a realistic estimate of a menu item's future demand. Various forecasting strategies have been created, each with its own set of benefits and drawbacks when compared to other approaches. AI is altering our behavior. As a result, we must go to more complex solutions in order to keep up with the rest of the globe. Many firms are shifting their applications to the AI module. With the correct tools, software, and programmed, we can create an automated process that increases customer demand forecasting and improve the customer experience. As a result, I've described the prototype and concept for food demand forecasting in restaurant industry. It has a lot of application in today's environment. This will undoubtedly generate a fantastic opportunity for businesses in logistics, customer demand forecasting, and other related fields.