VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

DATA VISUALIZATION

REVIEW-1

FOOD DEMAND FORECASTING

Team Members

T.KARAN(19BCE1751)

S.V.SYAM NAGA PAVAN(19BCE1090)

Y.HEMANTH(19BCE1047)

G.PRAVEEN(19BCE1029)

ABSTRACT:

The vital aspect in the world of business is to have a proper analysis of their business outcomes. This outcome plays a major role in the development of the business. One of the expanding business spheres is food delivering companies.

The vital factor in running such a food delivery company that is located at various branches in the city is to maintain the stock properly and prepare the food in time for delivering to the customers.

The datasets used to train the model have the information regarding the meal for example, the type of meal, Week, center-id , base-price , category, cuisine etc. This prediction model helps to find out the popular meals and the least ordered type of meals, based on these results they can manage the purchase of stock and raw materials.

We have to deal with many perishable raw materials, if there is too much stock it would tend to waste, if the stock is insufficient, it would lead to out-of- stocks, it would lead to out of stock of the meals that in turn would decrease the number of orders for the company.

INTRODUCTION:

As in today's competitive life even the business has become more difficult. Demand for food is increasing day by day with the increase in the population of every country. Estimation of the demand in food consumption plays a vital role in supplying or generating resources to produce the required amount of food.

To meet this challenge, we need to predict the demand for food consumption for the future so that the hunger of everyone can be satisfied.

We will analyse all the year data on how the food demand has been through the out there restaurants, with that information we can readily produce the raw materials and also recruit the staff required.

PROBLEM STATEMENT:

Food demand forecasting is about to help meal delivery companies located at various centers of the city in the demand forecasting for upcoming weeks.

The majority of the raw materials are perishable, and the replenishment of them is done on the weekly basis and the procurement planning is of utmost importance. The demand forecast is helpful in the staffing of the centres too.

The Main Motto is to predict the demand for the upcoming weeks(in the challenge the data set containing the 10-days of meals for the customers)for the centre-meal combinations in the test set.

Here we will predict the number of orders and the demand of orders based on all the attributes given in the train data set.

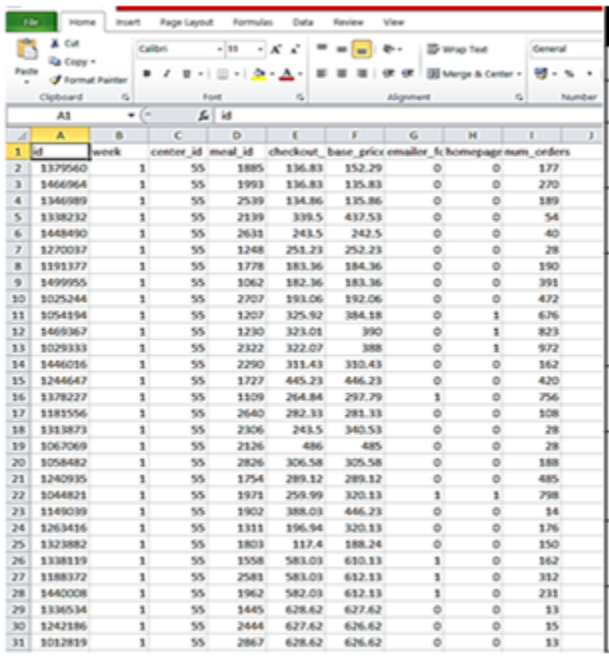
DATASET DESCRIPTION:

There are a total of 3 train data sets which are used in the processing. The three data sets are the train data set, fulfilment centre info data set and the meal info data set.

Train.csv data set includes all the attributes, the id's of the fulfilment centers and meals, and also the num_orders which is needed to be predicted, this is used to train the models, the fulfilment_center_info.csv file provides us with the details of the food centres which are providing the food.

We have merged the train , fulfilment center and the meal info datasets to form one train dataset and used that to build the model.

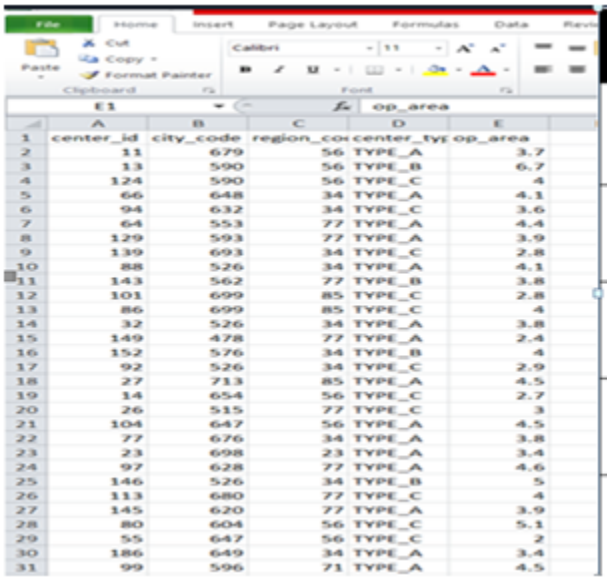
Train Dataset:



id	week	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	num_orders
1379560	1	55	1885	136.83	152.29	0	0	177
1466964	1	55	1993	136.83	135.83	0	0	270
1346989	1	55	2539	134.86	135.86	0	0	189
1338232	1	55	2139	339.5	437.53	0	0	54
1448490	1	55	2631	243.5	242.5	0	0	40
1270037	1	55	1248	251.23	252.23	0	0	28
1191377	1	55	1778	183.36	184.36	0	0	190
1499955	1	55	1062	182.36	183.36	0	0	391
1025244	1	55	2707	193.06	192.06	0	0	472
1054194	1	55	1207	325.92	384.18	0	1	676
1469367	1	55	1230	323.01	390	0	1	823
1029333	1	55	2322	322.07	388	0	1	972
1446016	1	55	2290	311.43	310.43	0	0	162
1244647	1	55	1727	445.23	446.23	0	0	420
1378227	1	55	1109	264.84	297.79	1	0	756
1181556	1	55	2640	282.33	283.33	0	0	108
1313873	1	55	2306	243.5	340.53	0	0	28
1067069	1	55	2126	486	485	0	0	28
1058482	1	55	2826	306.58	305.58	0	0	188
1240935	1	55	1754	289.12	289.12	0	0	485
1044821	1	55	1971	259.99	320.13	1	1	798
1149039	1	55	1902	388.03	446.23	0	0	14
1263416	1	55	1311	196.94	320.13	0	0	176
1323882	1	55	1803	117.4	188.24	0	0	150
1338139	1	55	1558	583.03	610.13	1	0	162
1188372	1	55	2581	583.03	612.13	1	0	312
1440008	1	55	1962	582.03	612.13	1	0	231
1336534	1	55	1445	628.62	627.62	0	0	13
1242186	1	55	2444	627.62	626.62	0	0	15
1012819	1	55	2867	628.62	626.62	0	0	13

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

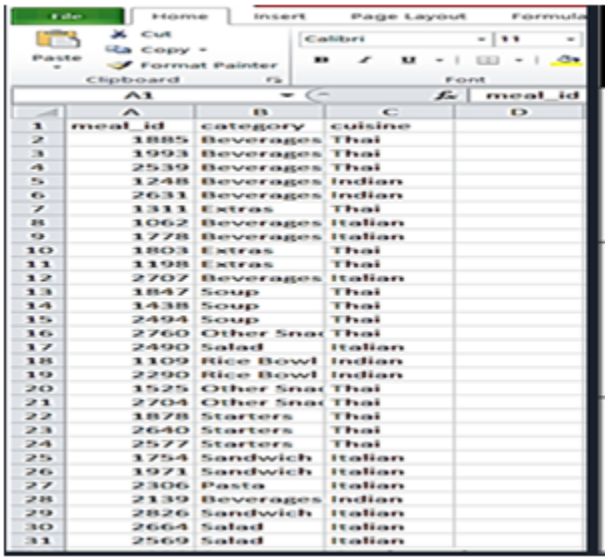
Fulfillment Dataset:



center_id	city_code	region_code	center_type	op_area
11	679	56	TYPE_A	3.7
13	590	56	TYPE_B	6.7
124	590	56	TYPE_C	4
66	648	34	TYPE_A	4.1
94	632	34	TYPE_C	3.6
64	553	77	TYPE_A	4.4
129	593	77	TYPE_A	3.9
139	693	34	TYPE_C	2.8
88	526	34	TYPE_A	4.1
143	562	77	TYPE_B	3.8
101	699	85	TYPE_C	2.8
86	699	85	TYPE_C	4
32	526	34	TYPE_A	3.8
149	478	77	TYPE_A	2.4
152	576	34	TYPE_B	4
92	526	34	TYPE_C	2.9
27	713	85	TYPE_A	4.5
14	654	56	TYPE_C	2.7
26	515	77	TYPE_C	3
104	647	56	TYPE_A	4.5
77	676	34	TYPE_A	3.8
23	698	23	TYPE_A	3.4
97	628	77	TYPE_A	4.6
146	526	34	TYPE_B	5
113	680	77	TYPE_C	4
145	620	77	TYPE_A	3.9
80	604	56	TYPE_C	5.1
55	647	56	TYPE_C	2
186	649	34	TYPE_A	3.4
99	596	71	TYPE_A	4.5

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)

Meal Info Dataset:



meal_id	category	cuisine
1885	Beverages	Thai
1993	Beverages	Thai
2539	Beverages	Thai
1248	Beverages	Indian
2631	Beverages	Indian
1311	Extras	Thai
1062	Beverages	Italian
1778	Beverages	Italian
1803	Extras	Thai
1198	Extras	Thai
2707	Beverages	Italian
1847	Soup	Thai
1438	Soup	Thai
2494	Soup	Thai
2760	Other Snac	Thai
2490	Salad	Italian
1109	Rice Bowl	Indian
2290	Rice Bowl	Indian
1525	Other Snac	Thai
2704	Other Snac	Thai
1878	Starters	Thai
2640	Starters	Thai
2577	Starters	Thai
1754	Sandwich	Italian
1971	Sandwich	Italian
2306	Pasta	Italian
2139	Beverages	Indian
2826	Sandwich	Italian
2664	Salad	Italian
2569	Salad	Italian

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

ALGORITHMS:

Linear Regression:

We will develop the 3 models using Linear Regression. We will import the Linear Regression from sklearn_linear_model package, split the dataset into train and test and perform linear regression directly on the dataset.

In the Second Model we will perform the Standard Scaling and log transformation on the target variable.

For the Third model we will create some attributes namely quarter and year, clear the outliers using quantile method.

XG-Boost:

XG-Boost is a high-speed and high-performance implementation of gradient boosted decision trees. The algorithm's implementation is designed to maximize computation time and memory resources. One of the design goals is to make the most of the resources available to train the model.

The two reasons to use XG-Boost are Execution speed and Model performance. XG-Boost dominates structured or tabular datasets on classification and regression predictive modeling problems. In general, XG Boost is quick.

Cat-Boost:

The Cat Boost algorithm can be used to solve a wide range of problems and can deal with many types of data with unique parameters which gives us accurate results. Categorical data in particular yields better accuracy when the cat boost algorithm is applied. With many parameters to tune with we can easily arrive at the best result out of all. Cat Boost algorithm also uses the gradient boosting algorithm which can yield very good results even with very small data.

This algorithm can be used without any preprocessing which is required in other algorithms. The other qualities include robustness, more accuracy and easy to implement. Most of the datasets can be processed by the default settings of the parameters without tuning them.

Random-Forest:

Random Forest Algorithm : It is an algorithm with a decision tree Model, in which the sub trees are learned.

Resulting predictions from all the sub-trees will have less correlation. Here it tries to build multiple CART models with different samples and initial variables. It will repeat the same process and then can make a final prediction.

This final prediction can simply be mean of each prediction. For example, if we have1000 observations in the complete population with 10 variables. It will take a random sample of 200 observations and 5 randomly chosen initial variables to build a CART model. It will repeat the process say 10 to 12 times and then make a final prediction on each observation.

REFERENCES:

1.)Demand forecasting for production planning in a food company by Nathalia Barbosa , Kelly Alonso Costa, Eliane da Silva

https://www.researchgate.net/profile/Nathalia-Barbosa2/publication/285219852_Demand_forecasting_for_production_planning_in_a_food_company/links/59b930fca6fdcc687230e27b/Demand-forecasting-for-production-planning-in-a-food-company.pdf

The food and beverage industry is one of the most important sectors of the Brazilian economy, with a significant participation in GDP index. The Brazilian economy has been showing relative stability in the last decades, which makes the sales demand to be more predictable. Due to this scenario of economic stability, the companies have been worried about investing in planning their operations, making use, mainly, of forecasting methods in order to become more competitive in the market. In the case of the food industry, the seasonal and the short perishability factors are a limitation to the maintenance of stocks, requiring a forecast with a high accuracy level. The present work consists in applying methods to forecast the demand for products of a food industry, which directs its sales to the food service market, in order to base the short to medium term production planning. Posteriorly, the forecasts will be evaluated using the error measure MAPE and compared to the demand currently considered by the company. The proposed methods feature a reduction of the error approximately 5%.

2.)Demand forecasting for the Ration companies by Langdon (Landon) Hollingsworth and Junlin (Shawn) Xiang

https://dspace.mit.edu/bitstream/handle/1721.1/131054/Hollingsworth_Xiang_project_Demand%20Planning%20for%20United%20Nations%20Food%20Rations_REVISED.pdf?sequence=1&isAllowed=y

Demand planning is a challenging component for organizations across a broad spectrum of industries. A key element of a successful demand plan is accurate forecasting, due in part to the operational decisions that are made based on the results of forecasting models. This is what our capstone project sponsor, Agility, has come to realize during their time-sensitive operations. Agility supplies food rations to the United Nations (UN) peacekeeping missions around the world.

3.) Predicting food demand in food courts by decision tree approaches by Ahmet Selman Bozkir & Ebru Akçapınar Sezer

https://reader.elsevier.com/reader/sd/pii/S1877050910005004?token=CA588BA39A21BEFB787869A8A4C0E79AA2DC036441BA5DB86EE8E070592CF5192A50D5492CF3C2EAC595F07677374719&originRegion=eu-west-1&originCreation=20210930130748

Fluctuations and unpredictability in food demand generally cause problems from an economic point of view in public food courts. In this study, to overcome this problem and predict actual consumption demand for a specified menu on a selected date, three decision tree methods (CART, CHAID and Microsoft Decision Trees) are

utilized. A two year period dataset which is gathered from food courts of Hacettepe University in Turkey is used during the analyses. As a result, prediction accuracies up to 0.83 in R2are achieved. By this study, it's shown that decision tree methodology is suitable for food consumption prediction.