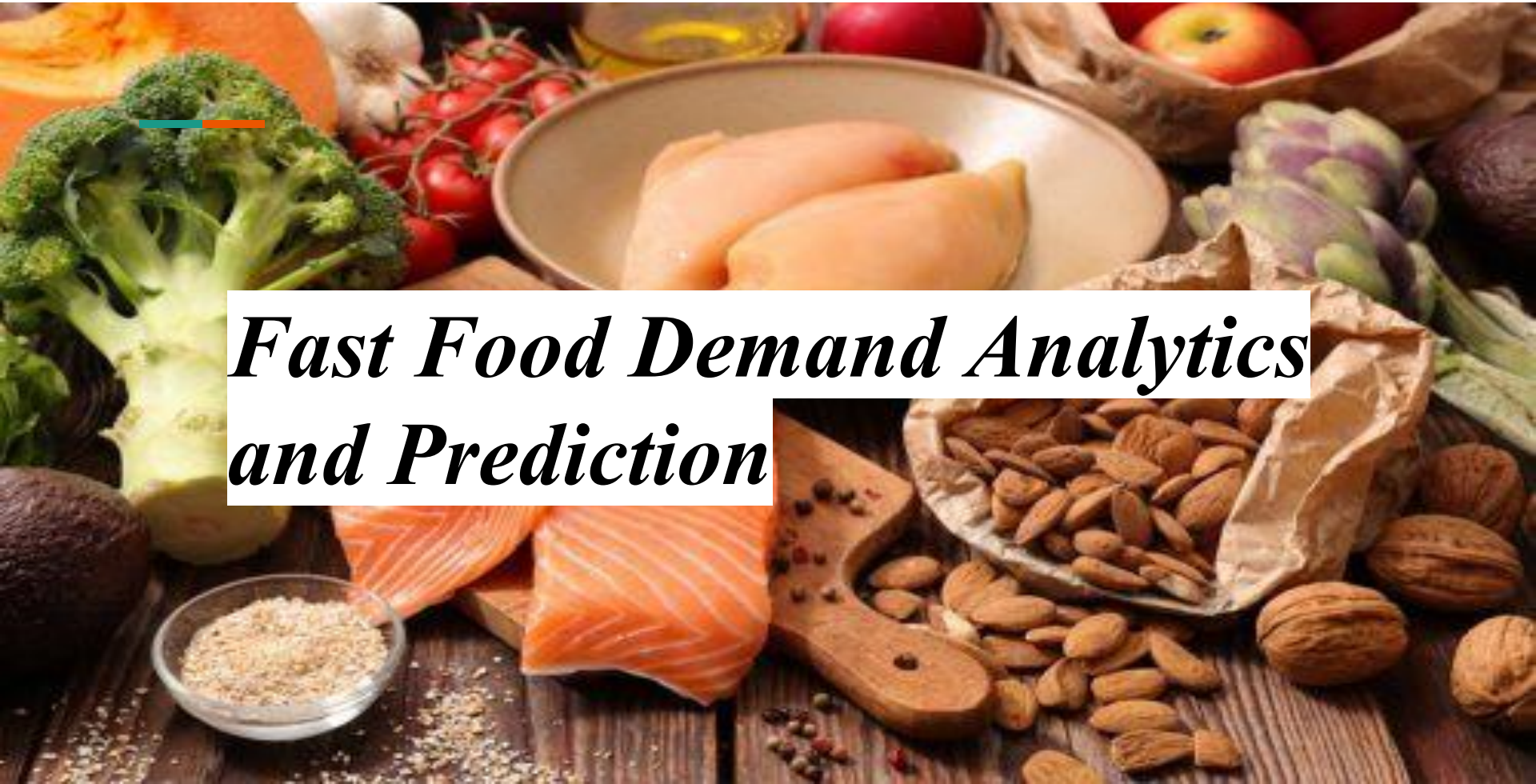


Fast Food Demand Analytics and Prediction

TEAM DETAILS :

1.	19MIA1005	Mohit More
2.	19MIA1066	Madasu Deepika
3.	19MIA1069	G. Harinisri

CSE3505 - Foundations of Data Analytics

***Class ID* - CH2022231000351**

***Slot* - L37+L38**

Type of the Project - C

PROBLEM STATEMENT

- A fast-food restaurant is a growing business and with time evolving the pattern of eating habits is changing. Therefore, many people have started opting for fast food which is an appealing phenomenon for them.
- In this fast-moving world where the expenses are day by day inclining, it is a must for all the members of the family to work for living. Also many times lot of people shift to other region away from the family in order to earn a living.
- This kind of a rush lifestyle gives a scope for consumption of quick meals which are easily prepared and quickly served which we term as fast-food.
- Lack of fast-food fulfillment to the consumer, excesses of fast food over the estimated demand, and business loss profit caused by inaccurate demand prediction are common nowadays in fast food centers and fast food-based businesses .
- Therefore Fast Food Demand Analytics and Prediction will give us an efficient solution to the problem we are facing.

LITERATURE SURVEY

S. no	TITLE	AUTHOR & YEAR OF PUBLICATION	FINDINGS
1	Demand forecasting in restaurants using machine learning and statistical analysis	TakashiTanizaki TomohiroHoshino TakeshiShimmura TakeshiTakenaka 2019	Demand forecasting in restaurants using machine learning was proposed. Forecasting rate for boosted decision tree is low and other algorithms did not have much difference.The forecast rate did not exceed 85%.
2	Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network	Krzysztof Lutoslawski Marcin Hernes Joanna Radomska Monika Hajdas Ewa Walaszczyk Agata Kozina 2021	Data science methods, including artificial intelligence methods, was used. The aim of this research was to develop models for food demand prediction based on a nonlinear autoregressive exogenous neural network. The architectures of the developed models differed in the number of hidden layers and the number of neurons in the hidden layers, as well as with different sizes of the delay line, were tested for a given product.

S. no	Title	Author & Year of publication	Findings
3	Predicting food demand in food courts by decision tree approaches	Ahmet SelmanBozkir Ebru AkcapinarSezer 2010	Three decision tree methods CART, CHAID and Microsoft Decision Trees are utilized. As a result, prediction accuracies up to 0.83 in R2 are achieved.
4	Data mining on time series: an illustration using fast-food restaurant franchise data	Lon-MuLiu SiddharthaBhattacharyya L.Sclove RongChen William J.Lattyak 2001	This shows how data mining can be applied to such time series, and help the franchise reap the benefits of such an effort. Time series data mining at both the store level and corporate level are discussed. Outlier detection also leads to information that can be used not only for better inventory management and planning, but also to identify potential sales opportunities as a part of results.

S. no	Title	Author & Year of publication	Findings
5	Predicting consumer preference for fast-food franchises: a data mining approach	Y Hayashi, M-H Hsieh R Setiono 2009	They evaluated the adequacy of two data mining techniques, decision tree and neural network in analysing consumer preference for a fast-food franchise and to examine the sufficiency of the criteria selected in understanding this preference. The generated rules show that while both decision tree and neural network models can achieve predictive accuracy of more than 80% on the training data samples and more that 70% on the cross-validation data samples
6	FOOD DEMAND PREDICTION USING MACHINE LEARNING	K.Aishwarya,Aishwarya.N.Rao, Nikita Kumari, AkshitMishra, Mrs.Rashmi M R 2020	The demands depend upon many explicit and hidden context such as season, region etc.The number of order is used to forecast stock of items, using machine learning with internal and external data. They used an appropriate algorithm for demand forecasting which is capable of overpowering the wastage of short life items.Algorithms like Bayesian Linear Regression, LASSO, XGBoost algorithm are used that considerably improves the forecasting performance.

OBJECTIVE

The core agenda of this project is to evaluate the factors of customer satisfaction highlighting the fast-food trend. Furthermore, this project will allow us to know which factor leads to customer satisfaction and also to determine the various factors which influence visiting fast-food outlets. In addition, to this we can discover how satisfied customers decide the trend of fast-food restaurants, what factors make customers satisfied and why they prefer a specific fast-food outlet.

To create an app which will predict the fast food prices and the data analysis of overall consumption.

DATASET

	A	B	C	D	E	F	G	H
1	id	week	center_id	meal_id	checkout_price	base_price	num_orders	
2	1379560	1	55	1885	136.83	152.29	177	
3	1466964	1	55	1993	136.83	135.83	270	
4	1346989	1	55	2539	134.86	135.86	189	
5	1338232	1	55	2139	339.5	437.53	54	
6	1448490	1	55	2631	243.5	242.5	40	
7	1270037	1	55	1248	251.23	252.23	28	
8	1191377	1	55	1778	183.36	184.36	190	
9	1499955	1	55	1062	182.36	183.36	391	
10	1025244	1	55	2707	193.06	192.06	472	
11	1054194	1	55	1207	325.92	384.18	676	
12	1469367	1	55	1230	323.01	390	823	
13	1029333	1	55	2322	322.07	388	972	
14	1446016	1	55	2290	311.43	310.43	162	
15	1244647	1	55	1727	445.23	446.23	420	
16	1378227	1	55	1109	264.84	297.79	756	
17	1181556	1	55	2640	282.33	281.33	108	

ALGORITHMS

Using a forecasting algorithm known as CatBoost with a data categorization technique. Fast food demand is affected by several independent variables such as seasonality, trend, price fluctuation, and length of historical data.

A combination of these selected variables was used to calculate demand prediction using parameter tuning in the CatBoost algorithm and other algorithms used for the experiment such as Linear Regression, LGBM, and XGBoost.

NOVELTY

- We will discover how satisfied customers decide the trend of fast-food restaurants, what factors make customers satisfied and why they prefer a specific fast-food outlet.
- To create an app which will predict the fast food prices and the data analysis of overall consumption and understand how visualization of a data is very important .
- To enhance the results with help of demand forecasting and prediction.

TIMELINE CHART

Review - 1	Introduction to project(week1), literature review(week2), Data Extraction.(week3)
Review - 2	Basic Data Analytics (week5-after cat1) Visualization(week6) Model Planning, Model Building(week7&8)
Review - 3	Model Evaluation(Week8), Creating Website(week9-11) , Upload in Github,Project Report Submission(week12)

MODULES

1) **Data Preprocessing from the Extracted Data:** When the user inputs files to the system, all information on the submitted files is merged into a separate dataset. Therefore, not having null values or missing values is compulsory. After this, validation process will initiate. Therefore, merging information is required to be validated. As an example, when merging file called “trainForLearnInformation”, specific variable name and their contents should be matched to the same name and contents in the other file. Such as trainToLearn file “meal_id” should match in the mealInfromation file “meal_id” variable.

2) **Exploratory Data Analysis for the Dataset:** This is a process, performing an initial investigation on data. Such as identify patterns, identify anomalies, test hypotheses and validate assumptions with the help of graphical statistics representation or statistical information. At the beginning of the exploratory data analysis process, it is compulsory to identify and remove unnecessary variables that are not contributing to the prediction process. Therefore, four variables were removed. Such as region_code, op_area, emailer_for_promotion and homepage_featured. emailer_for_promotion and homepage_featured column with data were dropped by updating the file. However, region_code, op_area kept in the dataset for identifying its necessity for implementation.The next process of the exploratory data analysis is the standardization of features. Standardization is a technique. It changes the values of numeric columns in the dataset to a common scale, without altering differences in the ranges of values.

3) Feature Engineering from the Extracted Information

This is a method to create features according to the domain knowledge that enables to enhancement performance and accuracy of the machine learning models using the dataset.

4) Data Transformation for Eliminate Outliers

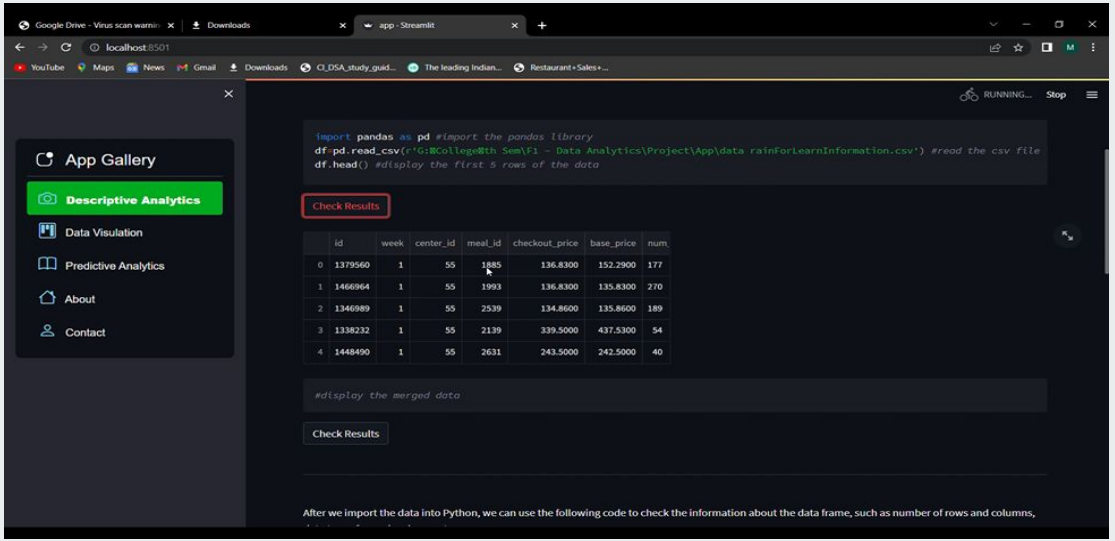
In the demand prediction context, it is compulsory to outlier data to be 0% on a targeted variable called “num_orders”. Therefore, this necessity is achieved by using the Interquartile range method. Log transformation is the most popular among the different types of transformations used to transform skewed data to approximately conform to normality in feature engineering. Therefore, the target variable called “num_orders” is not aligned with normality and non-use of transformation methods will reduce the performance of the data model. Therefore, it was decided to include log transformation on the targeted variable “num_orders”.

5) Machine Learning Algorithms for Demand Prediction: Multiple data modeled using gradient boosting algorithms like CatBoost algorithm. Those algorithms are implemented with feature extraction, data transformation and data preprocessing for achieving better accuracy on the predicted result.After the above process, it was decided to categorize dataset “week” values into a created feature called “Quarter” and “Year” . Reason for categorizing, train dataset contains 146 weeks of data which is approximately 11 quarters and one-quarter consists of approximately 13 weeks. Then manipulated those data accordingly for the detection outlier purpose.

6) Dataset Splitting as Test set and Trains set: It was necessary to drop some variables that are not affecting the prediction to improve the prediction result. Such as variables “id” and “city_code” are identified as irrelevant variables for the train, “num_orders” is a target variable for prediction, “special price” variable calculation of base price and checkout price. But identified there is lack of correlation with the target variable, “week” variable categorized with quarter/year wise and “special price percent” also removed. After removing irrelevant variables, it was decided to fit catboostRegressor model to the training data using the fit method. Therefore, it was able to predict result based on this data using predict method .

7) Model Training and Data Prediction

The predicted result was evaluated according to the implemented standard evaluation metrics .



Google Drive - Virus scan warnin | Downloads | app - Streamlit

localhost:8501

YouTube | Maps | News | Gmail | Downloads | CI_DSA_study_guid... | The leading Indian... | Restaurant+Sales+...

RUNNING... Stop

App Gallery

- Descriptive Analytics
- Data Visulation
- Predictive Analytics
- About
- Contact

```
#display the merged data
```

Check Results

	id	week	center_id	meal_id	checkout_price	base_price	num	city_code	regic	center_type	op_area	category	cuisine
456543	1271326	145	61	1543	484.0900	484.0900	68	473	77	TYPE_A	4.5000	Desert	Indian
456544	1062036	145	61	2304	482.0900	482.0900	42	473	77	TYPE_A	4.5000	Desert	Indian
456545	1110849	145	61	2664	237.6800	321.0700	501	473	77	TYPE_A	4.5000	Salad	Italian
456546	1147725	145	61	2569	243.5000	313.3400	729	473	77	TYPE_A	4.5000	Salad	Italian
456547	1361984	145	61	2490	292.0300	290.0300	162	473	77	TYPE_A	4.5000	Salad	Italian

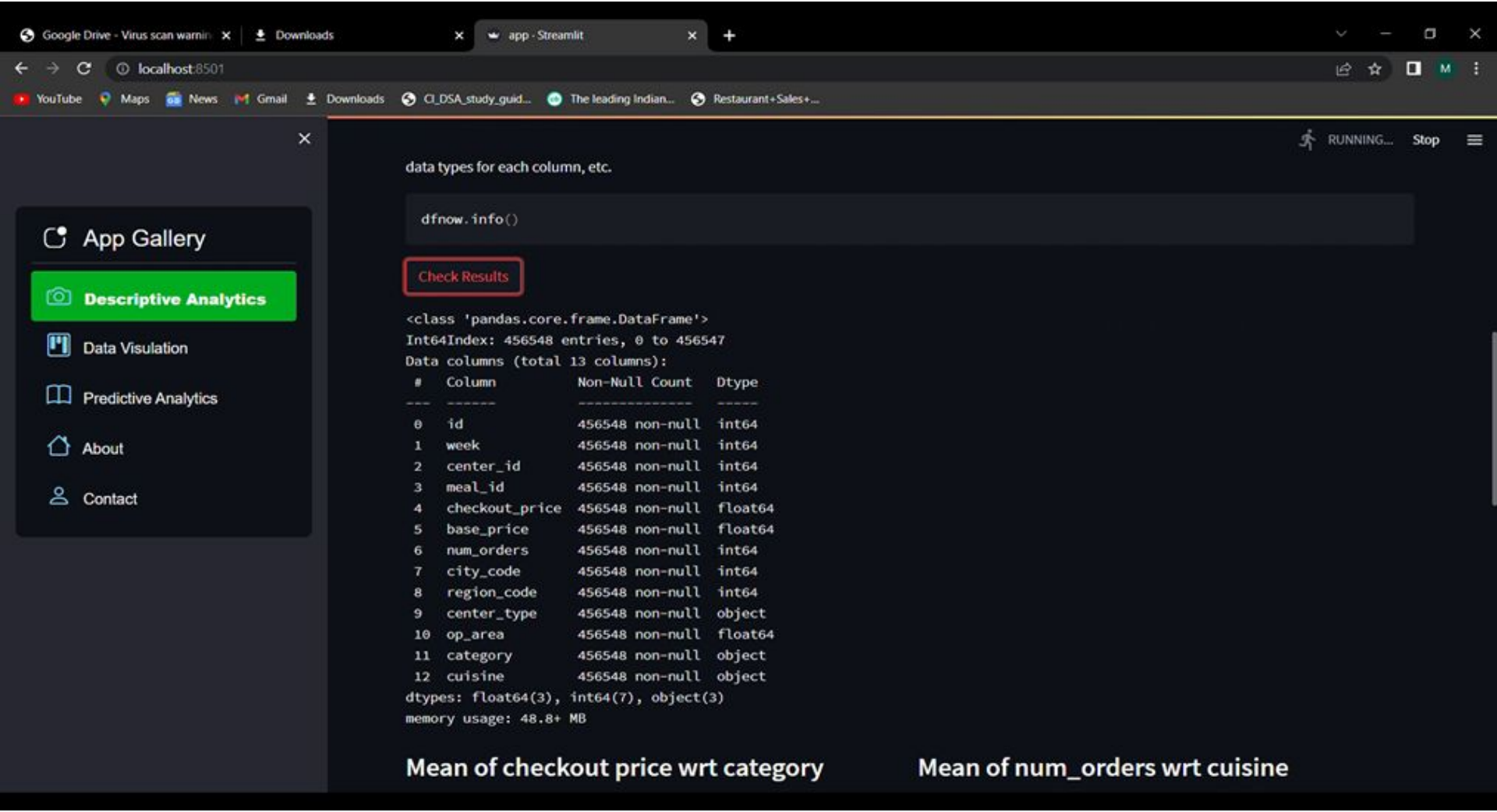
After we import the data into Python, we can use the following code to check the information about the data frame, such as number of rows and columns, data types for each column, etc.

```
dfnow.info()
```

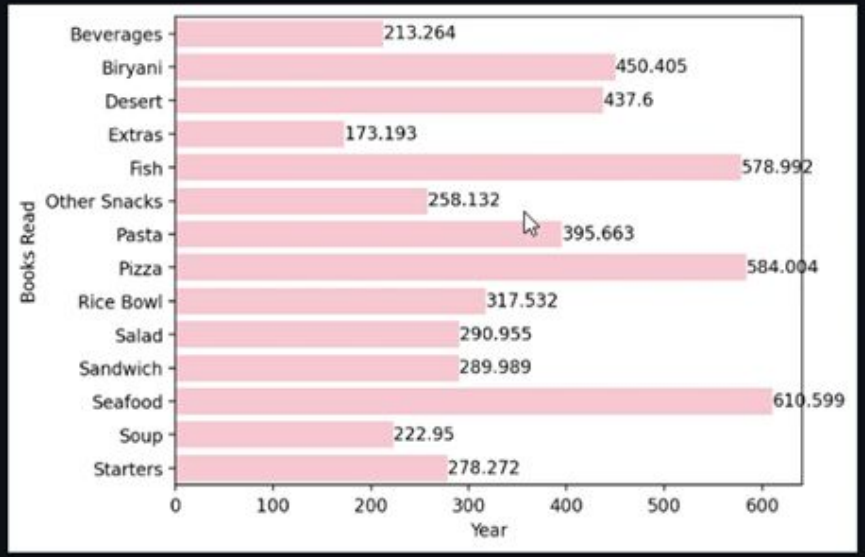
Check Results

Mean of checkout price wrt category

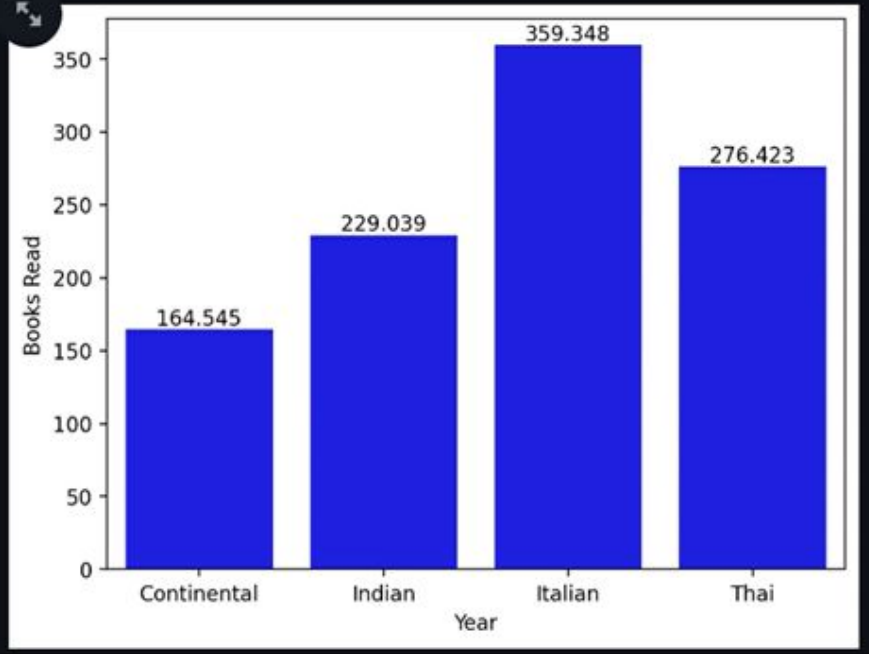
Mean of num_orders wrt cuisine


Google Drive - Virus scan warnin
Downloads
app - Streamlit
localhost:8501
YouTube
Maps
News
Gmail
Downloads
CI_DSA_study_guid...
The leading Indian...
Restaurant+Sales+...
RUNNING...
Stop
data types for each column, etc.
dfnow.info()
Check Results
App Gallery
Descriptive Analytics
Data Visulation
Predictive Analytics
About
Contact
<class 'pandas.core.frame.DataFrame'>
Int64Index: 456548 entries, 0 to 456547
Data columns (total 13 columns):
Column Non-Null Count Dtype
0 id 456548 non-null int64
1 week 456548 non-null int64
2 center_id 456548 non-null int64
3 meal_id 456548 non-null int64
4 checkout_price 456548 non-null float64
5 base_price 456548 non-null float64
6 num_orders 456548 non-null int64
7 city_code 456548 non-null int64
8 region_code 456548 non-null int64
9 center_type 456548 non-null object
10 op_area 456548 non-null float64
11 category 456548 non-null object
12 cuisine 456548 non-null object
dtypes: float64(3), int64(7), object(3)
memory usage: 48.8+ MB
Mean of checkout price wrt category
Mean of num_orders wrt cuisine

Mean of checkout price wrt category



Mean of num_orders wrt cuisine



Google Drive - Virus scan warnin | Downloads | app - Streamlit

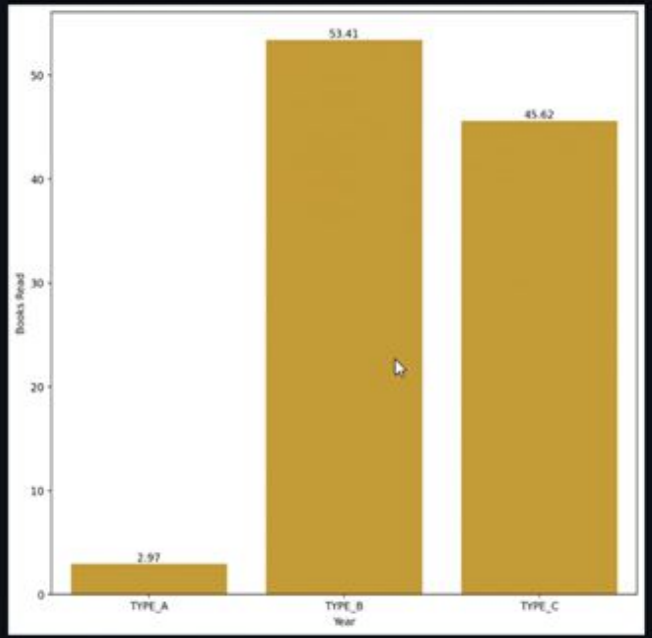
localhost:8501

YouTube | Maps | News | Gmail | Downloads | CI_DSA_study_guid... | The leading Indian... | Restaurant+Sales+...

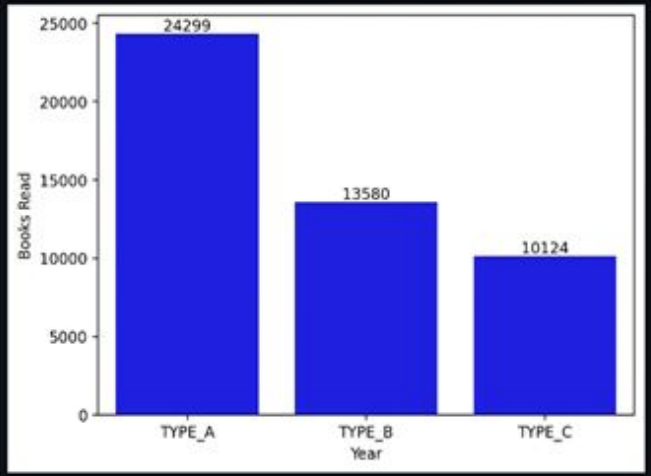
App Gallery

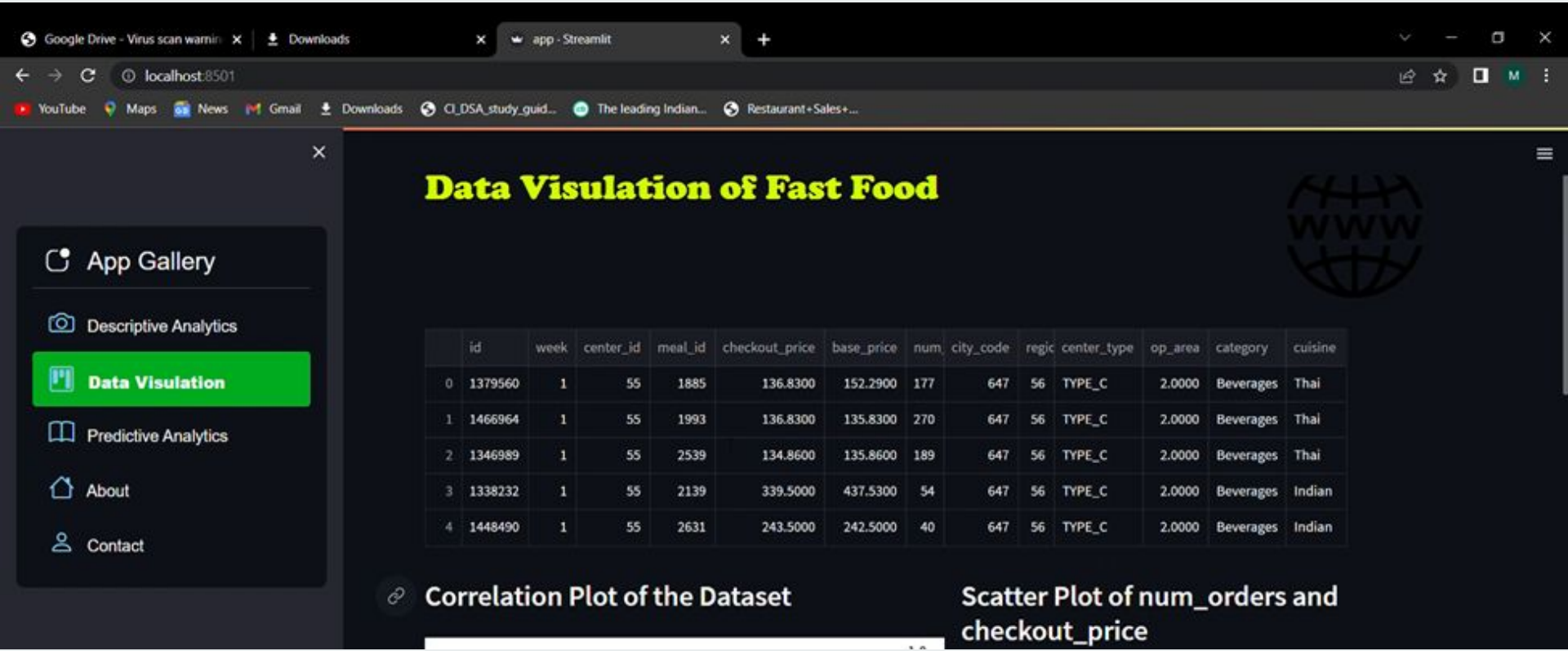
- Descriptive Analytics
- Data Visulation
- Predictive Analytics
- About
- Contact

Minimum of checkout price wrt center



Maximum of no of orders wrt center





	id	week	center_id	meal_id	checkout_price	base_price	num	city_code	regic	center_type	op_area	category	cuisine
0	1379560	1	55	1885	136.8300	152.2900	177	647	56	TYPE_C	2.0000	Beverages	Thai
1	1466964	1	55	1993	136.8300	135.8300	270	647	56	TYPE_C	2.0000	Beverages	Thai
2	1346989	1	55	2539	134.8600	135.8600	189	647	56	TYPE_C	2.0000	Beverages	Thai
3	1338232	1	55	2139	339.5000	437.5300	54	647	56	TYPE_C	2.0000	Beverages	Indian
4	1448490	1	55	2631	243.5000	242.5000	40	647	56	TYPE_C	2.0000	Beverages	Indian

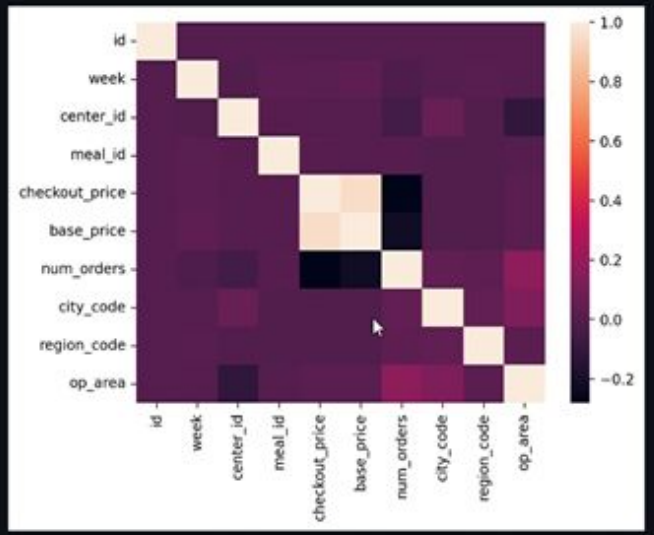
Google Drive - Virus scan warnin | Downloads | app - Streamlit

localhost:8501

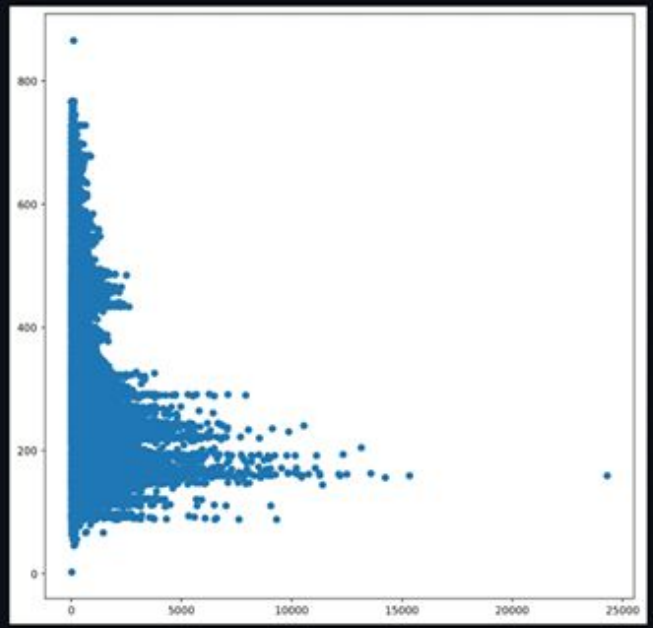
YouTube | Maps | News | Gmail | Downloads | CI_DSA_study_guid... | The leading Indian... | Restaurant+Sales+...

App Gallery

- Descriptive Analytics
- **Data Visulation**
- Predictive Analytics
- About
- Contact



eckout_price



Google Drive - Virus scan warnin | Downloads | app - Streamlit

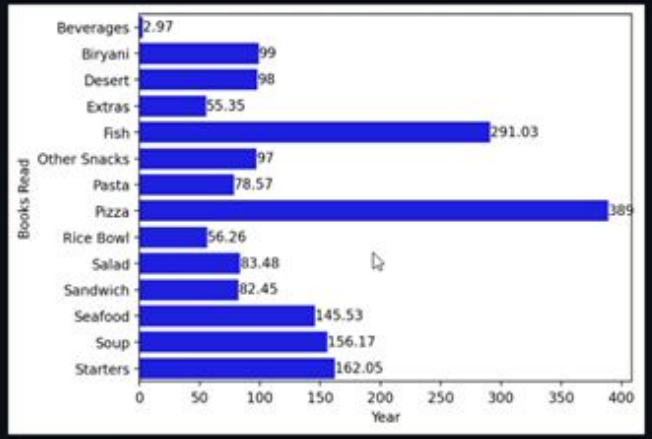
localhost:8501

YouTube | Maps | News | Gmail | Downloads | CI_DSA_study_guid... | The leading Indian... | Restaurant+Sales+...

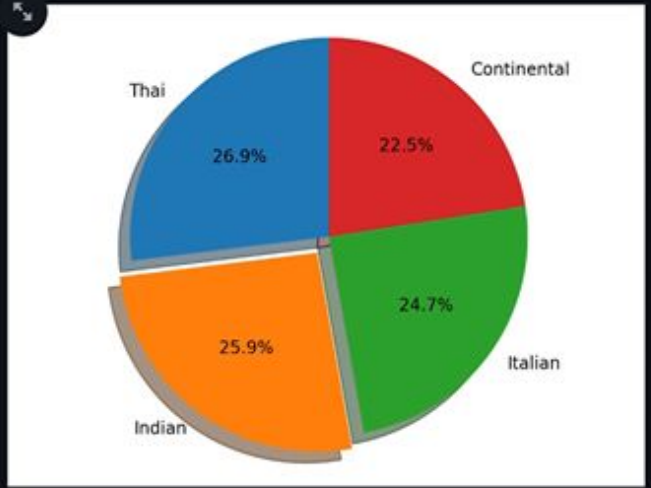
App Gallery

- Descriptive Analytics
- **Data Visulation**
- Predictive Analytics
- About
- Contact

Maximum of price wrt Category



Percentage of Cuisines Available



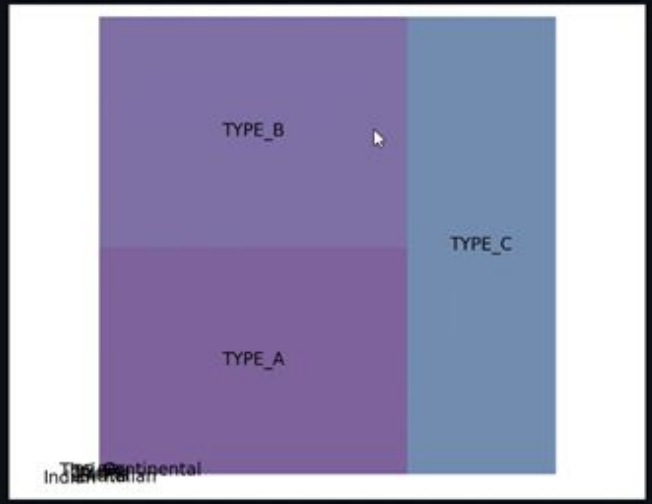
Descriptive Analytics

Data Visulation

Predictive Analytics

About

Contact



Google Drive - Virus scan warnin | Downloads | app - Streamlit

localhost:8501

YouTube | Maps | News | Gmail | Downloads | CI_DSA_study_guid... | The leading Indian... | Restaurant+Sales+...

RUNNING... Stop

App Gallery

- Descriptive Analytics
- Data Visulation
- **Predictive Analytics**
- About
- Contact

Predictive Analytics

CatBoostRegressor

CatBoost builds upon the theory of decision trees and gradient boosting. The main idea of boosting is to sequentially combine many weak models (a model performing slightly better than random chance) and thus through greedy search create a strong competitive predictive model.

Here we will predict the number of orders for a given week.

Prediction Going ON!!!!!!!!!!!!!! Please Wait

Prediction Done!!!!!!!!!!!!!! Thankyou For Your Patience!!!!!

Accuracy Metrics

Mean Absolute Error: 0.3975117975986247

Mean Squared Error: 0.26790889209376845

Root Mean Squared Error: 0.5175991616045842

Predicted Results

Google Drive - Virus scan warnin | Downloads | app - Streamlit

localhost:8501

YouTube | Maps | News | Gmail | Downloads | CI_DSA_study_guid... | The leading Indian... | Restaurant+Sales+...

App Gallery

- Descriptive Analytics
- Data Visulation
- **Predictive Analytics**
- About
- Contact

Here we will predict the number of orders for a given week.

Prediction Going ON!!!!!!!!!!!!! Please Wait

Prediction Done!!!!!!!!!!!!! Thankyou For Your Patience!!!!!

Accuracy Metrics

Mean Absolute Error: 0.3975117975986247

Mean Squared Error: 0.26790889209376845

Root Mean Squared Error: 0.5175991616045842

Predicted Results

	id	num_week	city_code	center_id	meal_id	checkout_price	base_price
423727	1017495	196	136	647	55	1885	148.4400	148.4400
423728	1395634	155	136	647	55	1993	151.3800	151.3800
423729	1007493	123	136	647	55	2539	152.3500	151.3500
423730	1042952	97	136	647	55	2631	96.0300	165.9300
423731	1022147	57	136	647	55	1248	97.0000	165.9300

CONCLUSION

This research is conducted to evaluate the pattern of fast food and customer satisfaction encompassing various factors. The analysis of customer satisfaction encompasses independent variables which lead to consumer satisfaction and showcase the fast-food trend. A computer system which can predict consumer demand for fast food sector would be highly recommended for many applications in food industry.The aim is to understand as much as possible about consumer demand statistics prediction in the fast-food sector.This research is conducted to evaluate the pattern of fast food and customer satisfaction encompassing various factors. The analysis of customer satisfaction encompasses independent variables which lead to consumer satisfaction and showcase the fast-food trend. The actual value tends to increase as the predicted values increases. Therefore, it is possible to say there is a linear positive correlation between those variables with a little number of outliers. At last, it was decided to use this model for the demand prediction process.

REFERENCES

1) Tanizaki, Takashi & Hoshino, Tomohiro & Shimmura, Takeshi & Takenaka, Takeshi. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. Procedia CIRP. 79. 679-683. 10.1016/j.procir.2019.02.042.
2) Lutosławski, Krzysztof & Hernes, Marcin & Radomska, Joanna & Hajdas, Monika & Walaszczyk, Ewa & Kozina, Agata. (2021). Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network. IEEE Access. 10.1109/ACCESS.2021.3123255.
3) Bozkir, Ahmet & Sezer, Ebru. (2011). Predicting food demand in food courts by decision tree approaches. Procedia CS. 3. 759-763. 10.1016/j.procs.2010.12.125.
4) Liu, Lon-Mu & Bhattacharyya, Siddhartha & Sclove, Stanley & Chen, Rong & Lattyak, William. (2001). Data mining on time series: An illustration using fast-food restaurant franchise data. Computational Statistics & Data Analysis. 37. 455-476. 10.1016/S0167-9473(01)00014-7.
5) Hayashi, Yoichi & Hsieh, M-H & Setiono, Rudy. (2009). Predicting consumer preference for fast-food franchises: A data mining approach. Journal of the Operational Research Society. 60. 1221-1229. 10.1057/palgrave.jors.2602646.
6) K.Aishwarya1, Rao, A. N., Kumari, N., Mishra, A., & R, R. M. (2020). *FOOD DEMAND PREDICTION USING MACHINE LEARNING*, *07*(06). https://doi.org/International Research Journal of Engineering and Technology (IRJET)

THANK YOU !