



Data Collection and Preprocessing Phase

Date	4 th July 2024
Team ID	739754
Project Title	Food demand forecasting for food delivery company
Maximum Marks	6 Marks

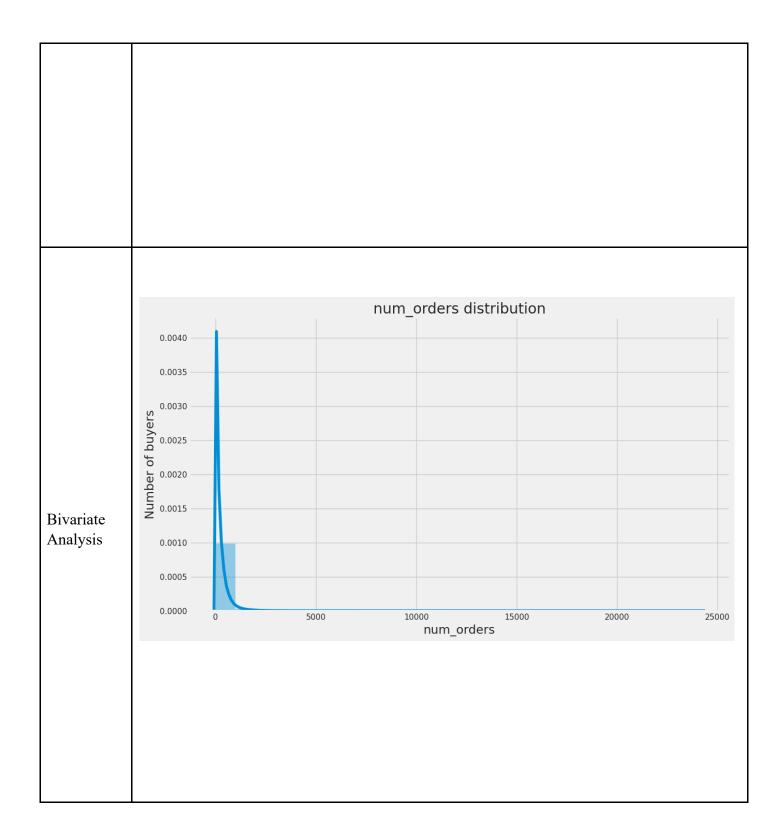
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Dim										
<u>Dimension:</u> 32574 rows × 10 columns <u>Descriptive statistics:</u>										
	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	
	0 1 2 3	id 0 1379560 1 1466964 2 1346989 3 1338232 4 1448490	id week 0 1379560 1 1 1466964 1 2 1346989 1 3 1338232 1	id week center_id 0 1379560 1 55 1 1466964 1 55 2 1346989 1 55 3 1338232 1 55	id week center_id meal_id 0 1379560 1 55 1885 1 1466964 1 55 1993 2 1346989 1 55 2539 3 1338232 1 55 2139	id week center_id meal_id checkout_price 0 1379560 1 55 1885 136.83 1 1466964 1 55 1993 136.83 2 1346989 1 55 2539 134.86 3 1338232 1 55 2139 339.50	id week center_id meal_id checkout_price base_price 0 1379560 1 55 1885 136.83 152.29 1 1466964 1 55 1993 136.83 135.83 2 1346989 1 55 2539 134.86 135.86 3 1338232 1 55 2139 339.50 437.53	id week center_id meal_id checkout_price base_price emailer_for_promotion 0 1379560 1 55 1885 136.83 152.29 0 1 1466964 1 55 1993 136.83 135.83 0 2 1346989 1 55 2539 134.86 135.86 0 3 1338232 1 55 2139 339.50 437.53 0	id week center_id meal_id checkout_price base_price emailer_for_promotion homepage_featured 0 1379560 1 55 1885 136.83 152.29 0 0 0 1 1466964 1 55 1993 136.83 135.83 0 0 0 2 1346989 1 55 2539 134.86 135.86 0 0 0 3 1338232 1 55 2139 339.50 437.53 0 0 0	











	num_orders	1.00	0.29	0.28	0.18	0.04	0.03	0.03	-0.02	- 1.0
	homepage_featured	0.29	1.00	0.39	0.04	0.01	0.00	0.03	-0.01	- 0.8
	emailer_for_promotion	0.28	0.39	1.00	-0.02	-0.01	-0.01	0.08	-0.00	
	op_area	0.18	0.04	-0.02	1.00	0.13	0.02	-0.01	0.00	- 0.6
	city_code	0.04	0.01	-0.01	0.13	1.00	0.04	-0.00	0.00	- 0.4
Multivariate	region_code	0.03	0.00	-0.01	0.02	0.04	1.00	-0.00	0.00	
Analysis	category	0.03	0.03	0.08	-0.01	-0.00	-0.00	1.00	0.03	- 0.2
	week	-0.02	-0.01	-0.00	0.00	0.00	0.00	0.03	1.00	- 0.0
		num_orders	homepage_featured	emailer_for_promotion	op_area	city_code	region_code	category	week	

Outliers and Anomalies

Data Preprocessing Code Screenshots





Loading Data	# Convert the encoded 'category' column to numeric type trainfinal['category'] = trainfinal['category'].astype(int) trainfinal.head() id week city_code region_code center_type op_area category cuisine checkout_price base_price email 0 1379560 1 647 56 2 2.0 0 Thai 136.83 152.29 1 1018704 2 647 56 2 2.0 0 Thai 135.83 152.29 2 1196273 3 647 56 2 2.0 0 Thai 132.92 133.92 3 1116527 4 647 56 2 2.0 0 Thai 135.86 134.86 4 1343872 5 647 56 2 2.0 0 Thai 146.50 147.50	Pyth
Handling Missing Data Data Transformation	<pre>trainfinal = pd.merge(train, meal_info, on="meal_id",how="outer") trainfinal = pd.merge(trainfinal, center_info, on="center_id",how="outer") trainfinal.head()</pre>	
	<pre>from sklearn.preprocessing import LabelEncoder lb1 = LabelEncoder() trainfinal['center_type'] = lb1.fit_transform(trainfinal['center_type']) lb2 = LabelEncoder() trainfinal['category'] = lb2.fit_transform(trainfinal['category']) # Convert the encoded 'category' column to numeric type trainfinal['category'] = trainfinal['category'].astype(int)</pre>	





Feature Engineering	<pre>features = columns.drop(['num_orders']) trainfinal3 = trainfinal[features] x = trainfinal3.values y = trainfinal['num_orders'].values</pre>
Save Processed Data	<pre>import pickle pickle.dump(DT,open('fdemand.pkl','wb')) testfinal = pd.merge(test, meal_info, on="meal_id",how="outer") testfinal = pd.merge(testfinal, center_info, on="center_id",how="outer") testfinal = testfinal.drop(['center_id','meal_id'], axis=1)</pre>