

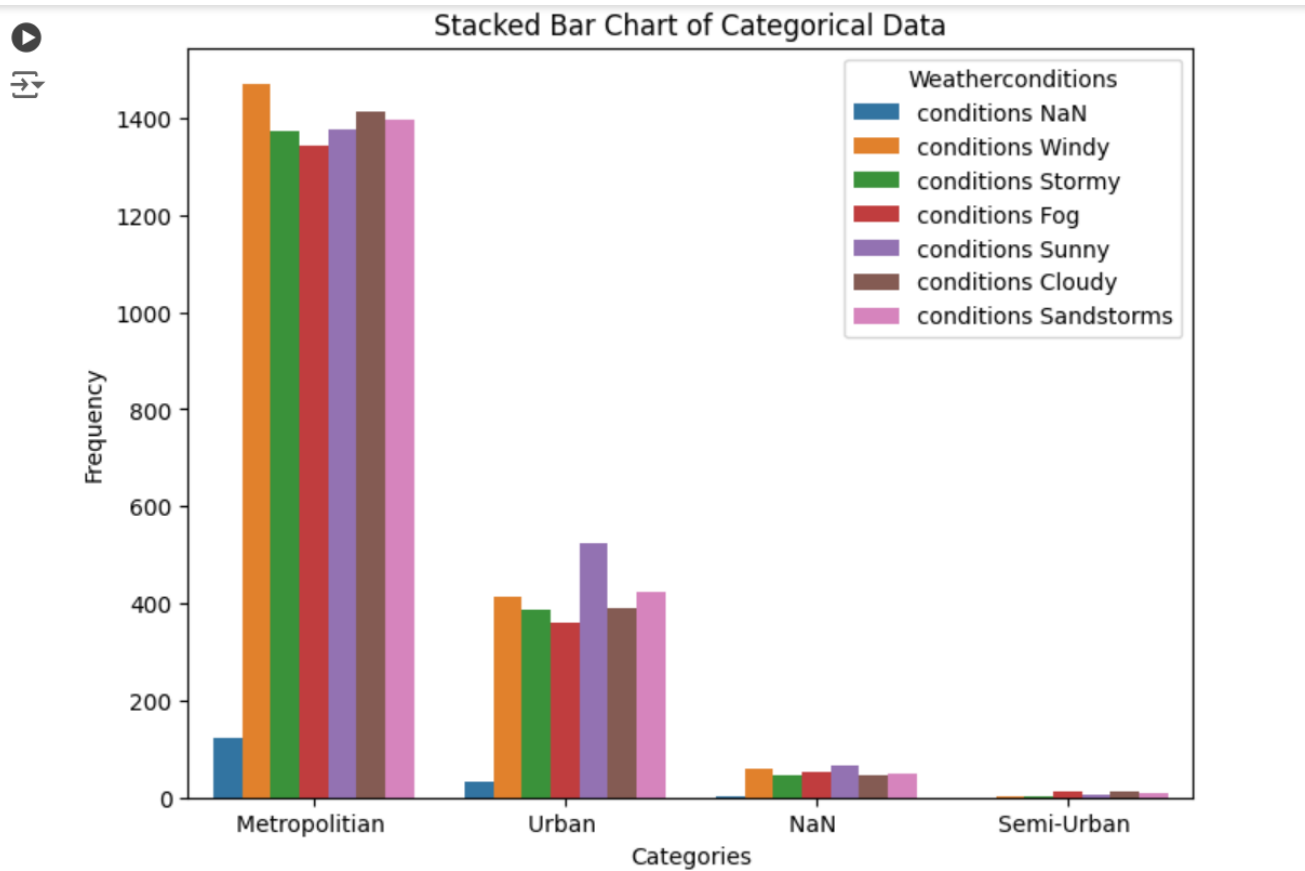
Data Collection and Preprocessing Phase

Date	07 JULY 2024
Team ID	739828
Project Title	Optimising Food Delivery Using Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

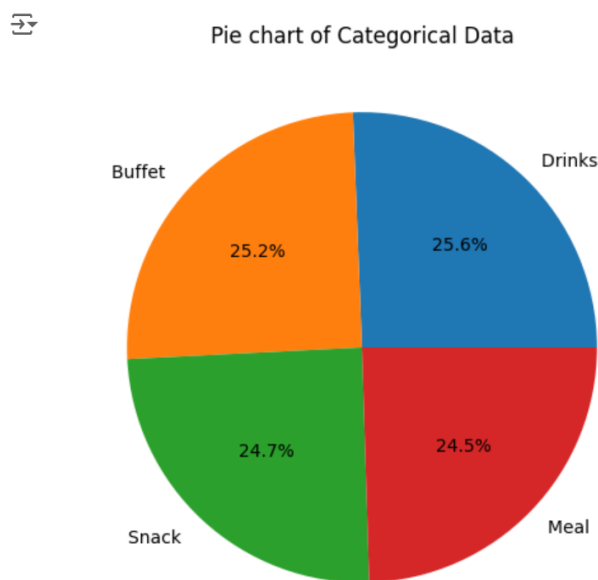
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis

Section	Description																																																																																																												
Data Overview	<div><pre>[] train_data.head()</pre><table><thead><tr><th></th><th>ID</th><th>Delivery_person_ID</th><th>Delivery_person_Age</th><th>Delivery_person_Ratings</th><th>Restaurant_latitude</th><th>Restaurant_longitude</th><th>Delivery_location_latitude</th><th>Del</th></tr></thead><tbody><tr><td>0</td><td>0x4607</td><td>INDORES13DEL02</td><td>37</td><td>4.9</td><td>22.745049</td><td>75.892471</td><td>22.765049</td><td></td></tr><tr><td>1</td><td>0xb379</td><td>BANGRES18DEL02</td><td>34</td><td>4.5</td><td>12.913041</td><td>77.683237</td><td>13.043041</td><td></td></tr><tr><td>2</td><td>0x5d6d</td><td>BANGRES19DEL01</td><td>23</td><td>4.4</td><td>12.914264</td><td>77.678400</td><td>12.924264</td><td></td></tr><tr><td>3</td><td>0x7a6a</td><td>COIMBRES13DEL02</td><td>38</td><td>4.7</td><td>11.003669</td><td>76.976494</td><td>11.053669</td><td></td></tr><tr><td>4</td><td>0x70a2</td><td>CHENRES12DEL01</td><td>32</td><td>4.6</td><td>12.972793</td><td>80.249962</td><td>13.012793</td><td></td></tr></tbody></table><pre>[] train_data.tail()</pre><table><thead><tr><th></th><th>ID</th><th>Delivery_person_ID</th><th>Delivery_person_Age</th><th>Delivery_person_Ratings</th><th>Restaurant_latitude</th><th>Restaurant_longitude</th><th>Delivery_location_latitude</th><th>Del</th></tr></thead><tbody><tr><td>45588</td><td>0x7cd9</td><td>JAPRES04DEL01</td><td>30</td><td>4.8</td><td>26.902328</td><td>75.794257</td><td>26.912328</td><td></td></tr><tr><td>45589</td><td>0xd641</td><td>AGRRES16DEL01</td><td>21</td><td>4.6</td><td>0.000000</td><td>0.000000</td><td>0.070000</td><td></td></tr><tr><td>45590</td><td>0x4f8d</td><td>CHENRES08DEL03</td><td>30</td><td>4.9</td><td>13.022394</td><td>80.242489</td><td>13.052394</td><td></td></tr><tr><td>45591</td><td>0x5eee</td><td>COIMBRES11DEL01</td><td>20</td><td>4.7</td><td>11.001753</td><td>76.986241</td><td>11.041753</td><td></td></tr><tr><td>45592</td><td>0x5fb2</td><td>RANCHIRES09DEL02</td><td>23</td><td>4.9</td><td>23.351058</td><td>85.325731</td><td>23.431058</td><td></td></tr></tbody></table></div> <p>Basic statistics, dimensions, and structure of the data.</p>		ID	Delivery_person_ID	Delivery_person_Age	Delivery_person_Ratings	Restaurant_latitude	Restaurant_longitude	Delivery_location_latitude	Del	0	0x4607	INDORES13DEL02	37	4.9	22.745049	75.892471	22.765049		1	0xb379	BANGRES18DEL02	34	4.5	12.913041	77.683237	13.043041		2	0x5d6d	BANGRES19DEL01	23	4.4	12.914264	77.678400	12.924264		3	0x7a6a	COIMBRES13DEL02	38	4.7	11.003669	76.976494	11.053669		4	0x70a2	CHENRES12DEL01	32	4.6	12.972793	80.249962	13.012793			ID	Delivery_person_ID	Delivery_person_Age	Delivery_person_Ratings	Restaurant_latitude	Restaurant_longitude	Delivery_location_latitude	Del	45588	0x7cd9	JAPRES04DEL01	30	4.8	26.902328	75.794257	26.912328		45589	0xd641	AGRRES16DEL01	21	4.6	0.000000	0.000000	0.070000		45590	0x4f8d	CHENRES08DEL03	30	4.9	13.022394	80.242489	13.052394		45591	0x5eee	COIMBRES11DEL01	20	4.7	11.001753	76.986241	11.041753		45592	0x5fb2	RANCHIRES09DEL02	23	4.9	23.351058	85.325731	23.431058	
	ID	Delivery_person_ID	Delivery_person_Age	Delivery_person_Ratings	Restaurant_latitude	Restaurant_longitude	Delivery_location_latitude	Del																																																																																																					
0	0x4607	INDORES13DEL02	37	4.9	22.745049	75.892471	22.765049																																																																																																						
1	0xb379	BANGRES18DEL02	34	4.5	12.913041	77.683237	13.043041																																																																																																						
2	0x5d6d	BANGRES19DEL01	23	4.4	12.914264	77.678400	12.924264																																																																																																						
3	0x7a6a	COIMBRES13DEL02	38	4.7	11.003669	76.976494	11.053669																																																																																																						
4	0x70a2	CHENRES12DEL01	32	4.6	12.972793	80.249962	13.012793																																																																																																						
	ID	Delivery_person_ID	Delivery_person_Age	Delivery_person_Ratings	Restaurant_latitude	Restaurant_longitude	Delivery_location_latitude	Del																																																																																																					
45588	0x7cd9	JAPRES04DEL01	30	4.8	26.902328	75.794257	26.912328																																																																																																						
45589	0xd641	AGRRES16DEL01	21	4.6	0.000000	0.000000	0.070000																																																																																																						
45590	0x4f8d	CHENRES08DEL03	30	4.9	13.022394	80.242489	13.052394																																																																																																						
45591	0x5eee	COIMBRES11DEL01	20	4.7	11.001753	76.986241	11.041753																																																																																																						
45592	0x5fb2	RANCHIRES09DEL02	23	4.9	23.351058	85.325731	23.431058																																																																																																						
Univariate Analysis	<p>Exploration of individual variables (mean, median, mode, etc.).</p> <div><pre>[]</pre><p>Distribution of Restaurant_latitude</p></div>																																																																																																												
Bivariate Analysis	<p>Relationships between two variables (correlation, scatter plots).</p>																																																																																																												



Multivariate
Analysis

Patterns and relationships involving multiple variables.



Outliers and Anomalies

Identification and treatment of outliers.

```
import math
def haversine(lat1, lon1, lat2, lon2):
    lat1 = math.radians(lat1)
    lon1 = math.radians(lon1)
    lat2 = math.radians(lat2)
    lon2 = math.radians(lon2)

    radius = 6371.0

    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = math.sin(dlat/2)**2 + math.cos(lat1) * math.cos(lat2) * math.sin(dlon/2)**2
    c = 2 * math.atan2(math.sqrt(a), math.sqrt(1-a))
    distance = radius * c

    return distance

test_data['distance'] = test_data.apply(lambda row: haversine(row['Restaurant_latitude'], row['R

[ ] for column in test_data.columns:
    if test_data[column].dtype != 'O':
        sns.displot(test_data[column])
        plt.title(f'Distribution of {column}')
        plt.show()
```

Data Preprocessing Code Screenshots

Loading Data

Code to load the dataset into the preferred environment (e.g., Python, R).

[] x_train

	Delivery_person_Ratings	Restaurant_latitude	Restaurant_longitude	Delivery_location_latitude	Delivery_location_longitude	Time_Orderd	Weathr
29044	24	4.723145	73.167753	22.318096	73.177753	36	
41736	18	3.512546	76.616792	12.367978	76.646792	160	
17874	22	4.770911	75.886362	22.831593	75.956362	101	
42093	16	4.723381	73.169083	22.350329	73.209083	51	
22952	24	0.000000	0.000000	0.110000	0.110000	132	
...
11284	25	3.310868	76.971082	11.071850	77.081082	100	
44732	25	5.185672	75.802083	26.981191	75.892083	141	
38158	25	4.350584	72.832585	18.967584	72.872585	70	
860	25	3.617518	80.220672	13.136439	80.270672	100	
15795	22	4.172574	78.437225	17.470371	78.497225	117	

36474 rows x 11 columns

Handling Missing Data	<p>Code for identifying and handling missing values.</p> <pre> ▶ Delivery_person_Ratings 0 ↔ Restaurant_latitude 431 ↔ Restaurant_longitude 0 Delivery_location_latitude 0 Delivery_location_longitude 0 Time_Orderd 0 Weatherconditions 0 Road_traffic_density 0 Type_of_order 0 Festival 0 City 0 dtype: int64 </pre> <pre>[] X['Restaurant_latitude'].fillna(X['Restaurant_latitude'].mean(),inplace=True)</pre> <pre>[] X.isnull().sum()</pre> <pre> ↔ Delivery_person_Ratings 0 Restaurant_latitude 0 Restaurant_longitude 0 Delivery_location_latitude 0 Delivery_location_longitude 0 Time_Orderd 0 Weatherconditions 0 Road_traffic_density 0 Type_of_order 0 Festival 0 City 0 dtype: int64 </pre>
Data Transformation	Code for transforming variables (scaling, normalization).
Feature Engineering	Code for creating new features or modifying existing ones.

```
[ ] #Get feature importances

importances = rf.feature_importances_

#Create a DataFrame for visualization

feature_importances = pd.DataFrame({'feature': X.columns, 'importance': importances})
```

```
[ ] feature_importances
```

	feature	importance
0	Delivery_person_Ratings	0.229907
1	Restaurant_latitude	0.062626
2	Restaurant_longitude	0.060808
3	Delivery_location_latitude	0.090019
4	Delivery_location_longitude	0.088612
5	Time_Orderd	0.117300
6	Weatherconditions	0.125084
7	Road_traffic_density	0.129196
8	Type_of_order	0.027626
9	Festival	0.046998