Data Collection and Preprocessing Phase

Date	22 June 2024
Team ID	739768
Project Title	To Predict The consumer Price Index
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

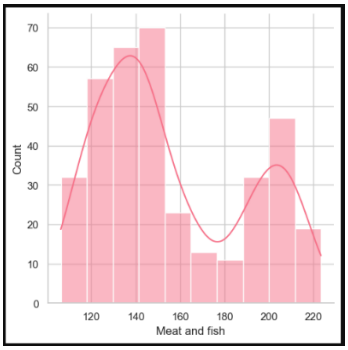
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong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

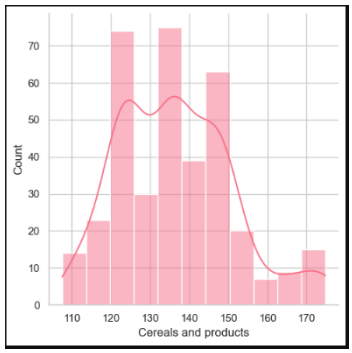
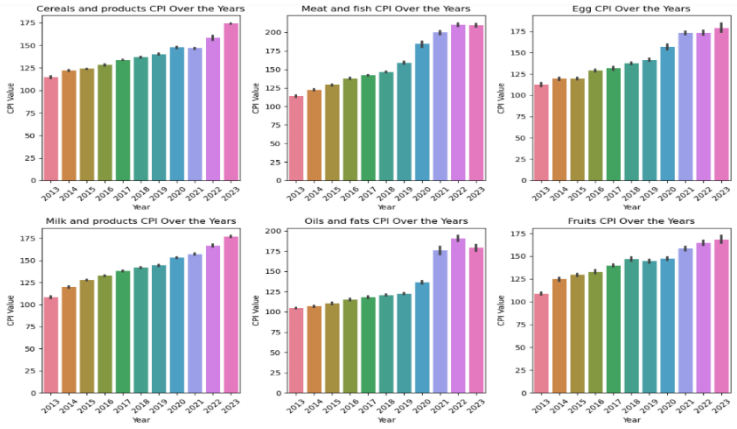
Section	Description

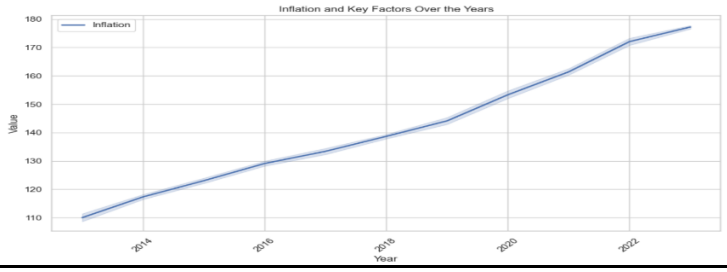
Data Overview	Descriptive statistics:
	`#Load the dataset`
	`cpi_data=pd.read_csv("All_India_Index_july2019_20Aug2020_dec20_2.csv")`
	`cpi_data`

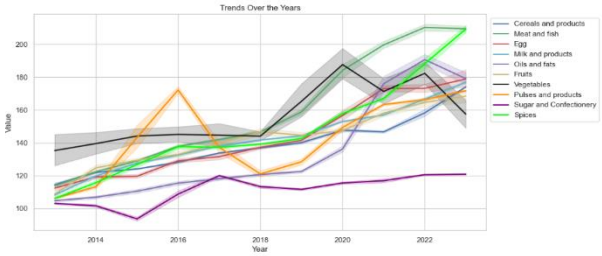
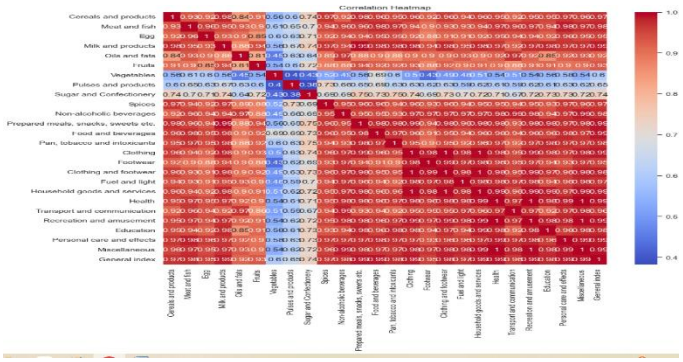
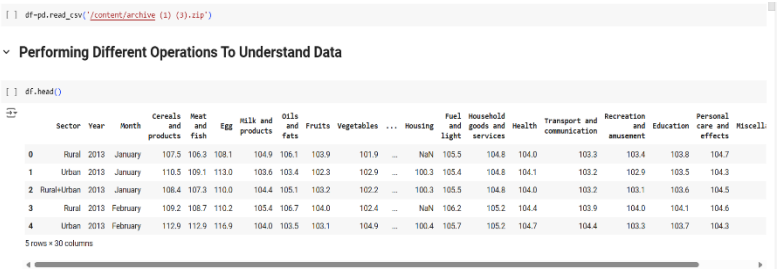
Univariate Analysis	

	Sector	Year	Month	Cereals and products	Meat and fish	Egg	Milk and products	Oils and fats	Fruits	Vegetables
0	Rural	2013	January	107.5	106.3	108.1	104.9	106.1	103.9	101.9
1	Urban	2013	January	110.5	109.1	113.0	103.6	103.4	102.3	102.9
2	Rural+Urban	2013	January	108.4	107.3	110.0	104.4	105.1	103.2	102.2
3	Rural	2013	February	109.2	108.7	110.2	105.4	106.7	104.0	102.4
4	Urban	2013	February	112.9	112.9	116.9	104.0	103.5	103.1	104.9

<table>
<tr>
<td></td>
<td> 

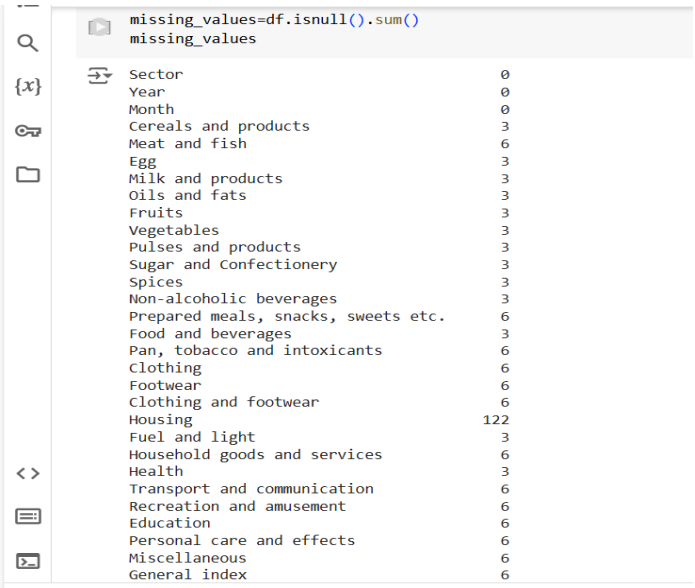

</td>
</tr>
<tr>
<td>Bivariate Analysis</td>
<td> 


</td>
</tr>
</table>

Multivariate Analysis	 
Outliers and Anomalies	-
Data Preprocessing Code Screenshots	
Loading Data	

	Sector	Year	Month	Cereals and products	Meat and fish	Egg	Milk and products	Oils and fats	Fruits	Vegetables	...	Housing	Fuel and light	Household goods and services	Health	Transport and communication	Recreation and amusement	Education	Personal care and effects	Miscell
0	Rural	2013	January	107.5	106.3	108.1	104.9	106.1	103.9	101.9	...	NaN	105.5	104.8	104.0	103.3	103.4	103.8	104.7	
1	Urban	2013	January	110.5	109.1	113.0	103.6	103.4	102.3	102.9	...	100.3	105.4	104.8	104.1	103.2	102.9	103.5	104.3	
2	Rural+Urban	2013	January	108.4	107.3	110.0	104.4	105.1	103.2	102.2	...	100.3	105.5	104.8	104.0	103.2	103.1	103.6	104.5	
3	Rural	2013	February	109.2	108.7	110.2	105.4	106.7	104.0	102.4	...	NaN	106.2	105.2	104.4	103.9	104.0	104.1	104.6	
4	Urban	2013	February	112.9	112.9	116.9	104.0	103.5	103.1	104.9	...	100.4	105.7	105.2	104.7	104.4	103.3	103.7	104.3	

5 rows × 30 columns

Handling Missing Data	
Feature Engineering	Enhance the accuracy and the robustness of the CPI predictions
Save Processed Data	-