**Project Requirements:**

**Data validation and Preprocessing:** Clean andpreprocess the data. This step includes handling any missing values, outliers and ensuring that the data is ready for analysis.

**Chart 1. Distribution of Store Area**

* Use ML algorithms to analyse the distribution of store areas.
* Visualize the distribution using a histogram to show the frequency of different store area sizes.

**Chart 2. Relationship between Store Area and Daily Customer Count (Scatter Plot)**

* Perform predictive analysis to understand the relationship between store area and daily customer count using ML regression techniques.
* Develop a predictive model to estimate the daily customer count based on the store area.
* Create a scatter plot visualization to depict the relationship between store area and daily customer count.

**Chart 3. Distribution of Items Available**

* Aggregate stores by the number of items available using ML techniques.
* Visualize the distribution using a bar chart, where each bar represents a range of items available and the height represents the number of stores within each range.

**Chart 4. Relationship between Daily Customer Count and Store Sales**

* Analyse the relationship between daily customer count and store sales using machine learning regression techniques.
* Develop a predictive model to estimate store sales based on the daily customer count.
* Visualize the relationship using a line chart, where daily customer count is on the x-axis and store sales is on the y-axis.

**Chart 5. Distribution of Store Sales**

* Analyse the distribution of store sales using machine learning techniques.
* Visualize the distribution using a box plot to show the median, quartiles, and outliers of store sales.

**Chart 6. Average Daily Customer Count by Store Area**

* Aggregate the daily customer count by different ranges of store areas using machine learning techniques.
* Visualize the average daily customer count for each range using a bar chart, where each bar represents a range of store areas.

**Chart 7. Store Sales Trend Over Time**

* Analyse the trend of store sales over time using machine learning time series analysis techniques.
* Visualize the trend using a line chart, where the x-axis represents time (e.g., months or years) and the y-axis represents store sales.

**Chart 8. Correlation Matrix Heatmap**

* Generate a correlation matrix using machine learning techniques to explore the relationships between different variables (store area, items available, daily customer count, store sales).
* Visualize the correlation matrix using a heatmap, where each cell represents the correlation coefficient between two variables.

**Note:** *Please use ML models wherever necessary even if not specified – apply you critical thinking and analytical skills*

**TECHNICAL EXPECTATIONS:**

* Use appropriate tools and libraries when necessary
* Data Analysis – Pandas, NumPy, etc.
* Visualization – Matplotlib, Seaborn, Plotly
* Interactive Dashboard – Dash/Streamlit

**CAPSTONE PRESENTATION:**

* Include the preprocessing methods employed in your project
* Provide the insights and interpretations of the visualizations
* Explain in detail the analysis and the visualization techniques used and why?
* Provide explanation for the choice of the ML algorithm