**Exploratory Data Analysis on Superstore Sales**

**Univariate Analysis:**

1. **Understand Dataset:**

Explore the structure of the dataset using df.info() and df.head() to get an overview of the columns and data.

1. **Data Cleaning:**

Check for missing values using df.isnull().sum().

Handle missing values appropriately, either by dropping or imputing.

Check for anomalies or outliers in the data.

1. **Data Types:**

Use df.dtypes to check the data types of each column.

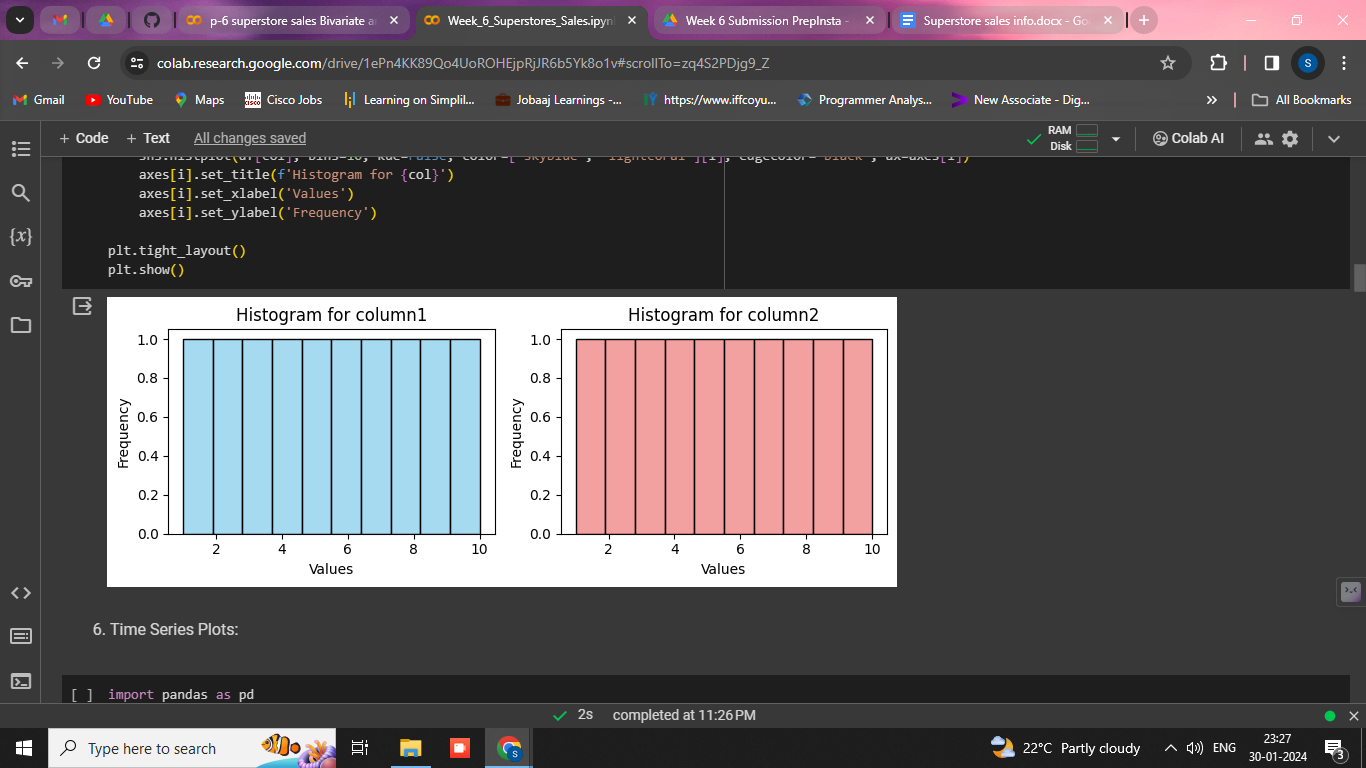
Convert data types if needed (e.g., converting date columns to datetime).

1. **Descriptive Statistics:**

Calculate basic descriptive statistics using df.describe() for key numeric variables like sales and quantity.

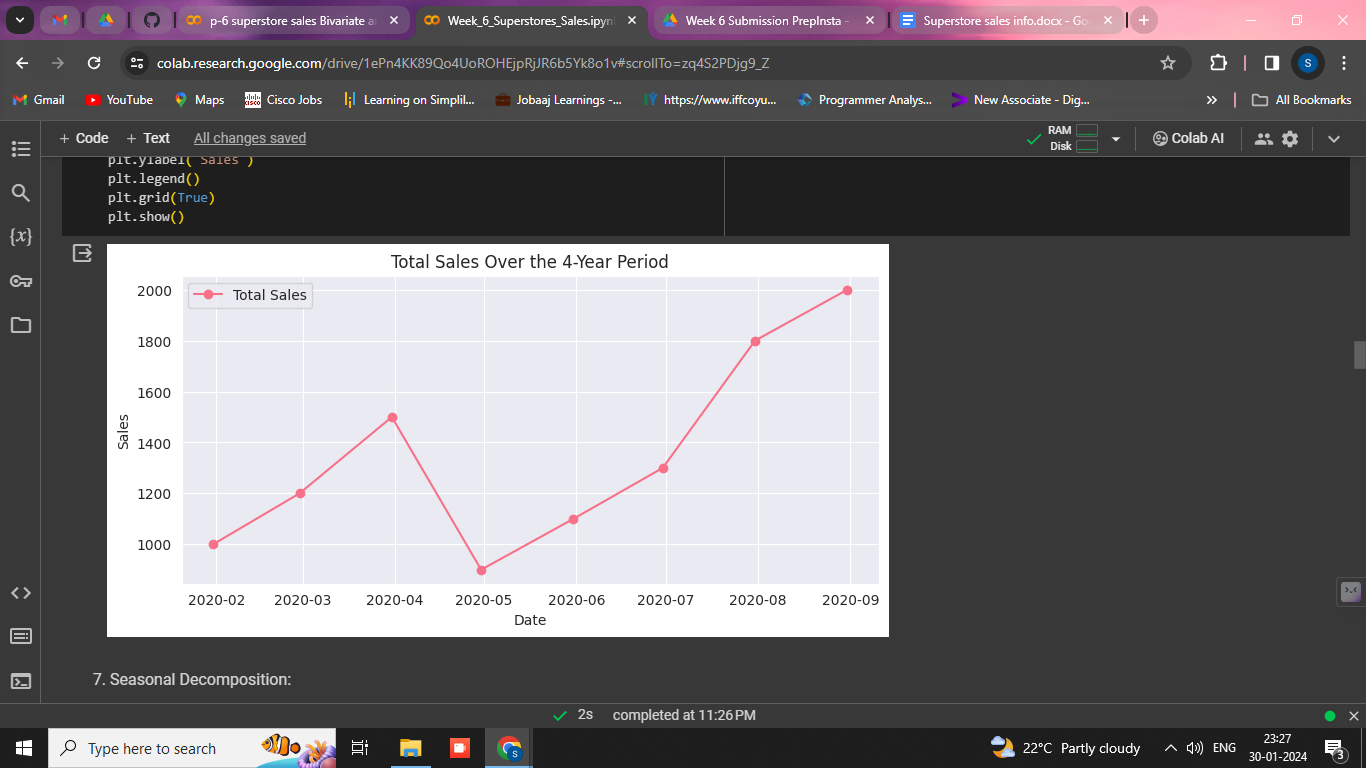
1. **Histograms:**

Create histograms for key numeric variables using df['column'].hist().



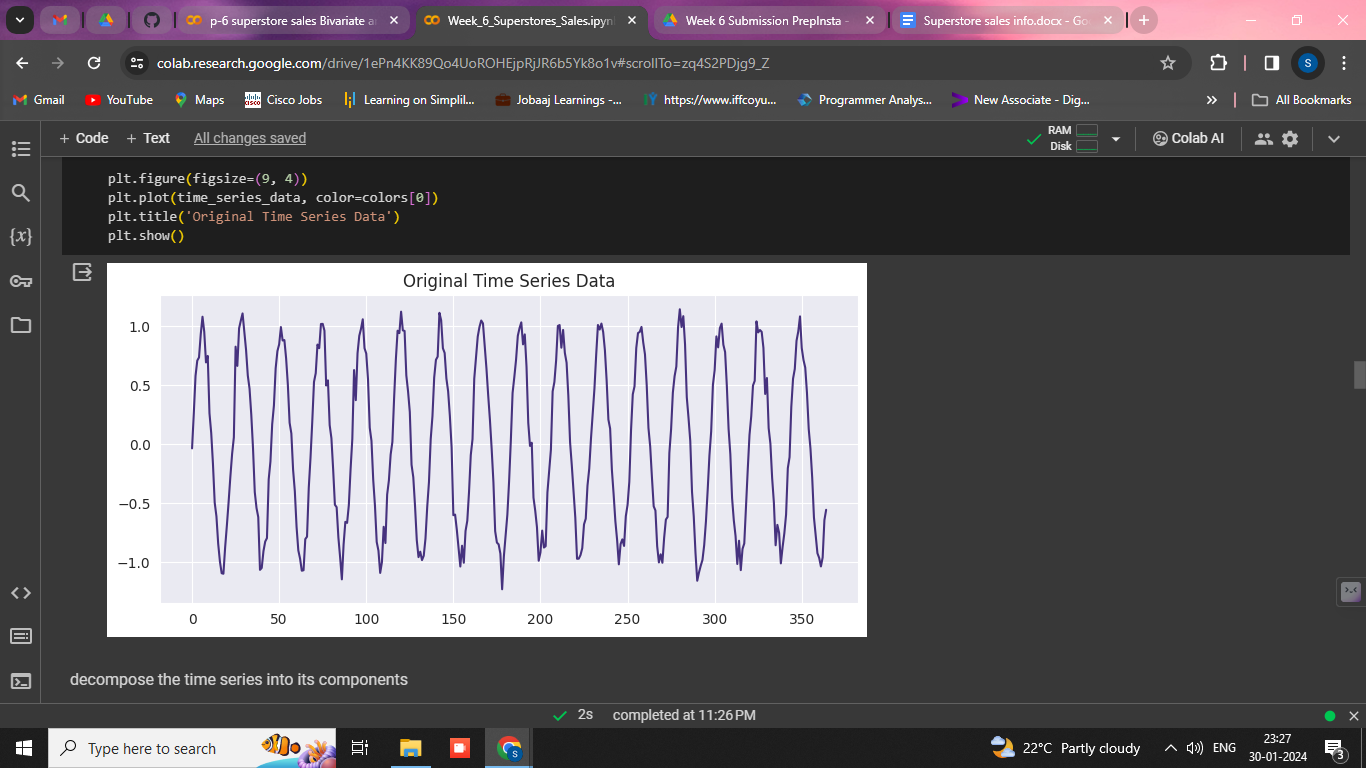
1. **Time Series Plots:**

Plot time series graphs for sales using the matplotlib or seaborn library.

****

**Seasonal Decomposition:**

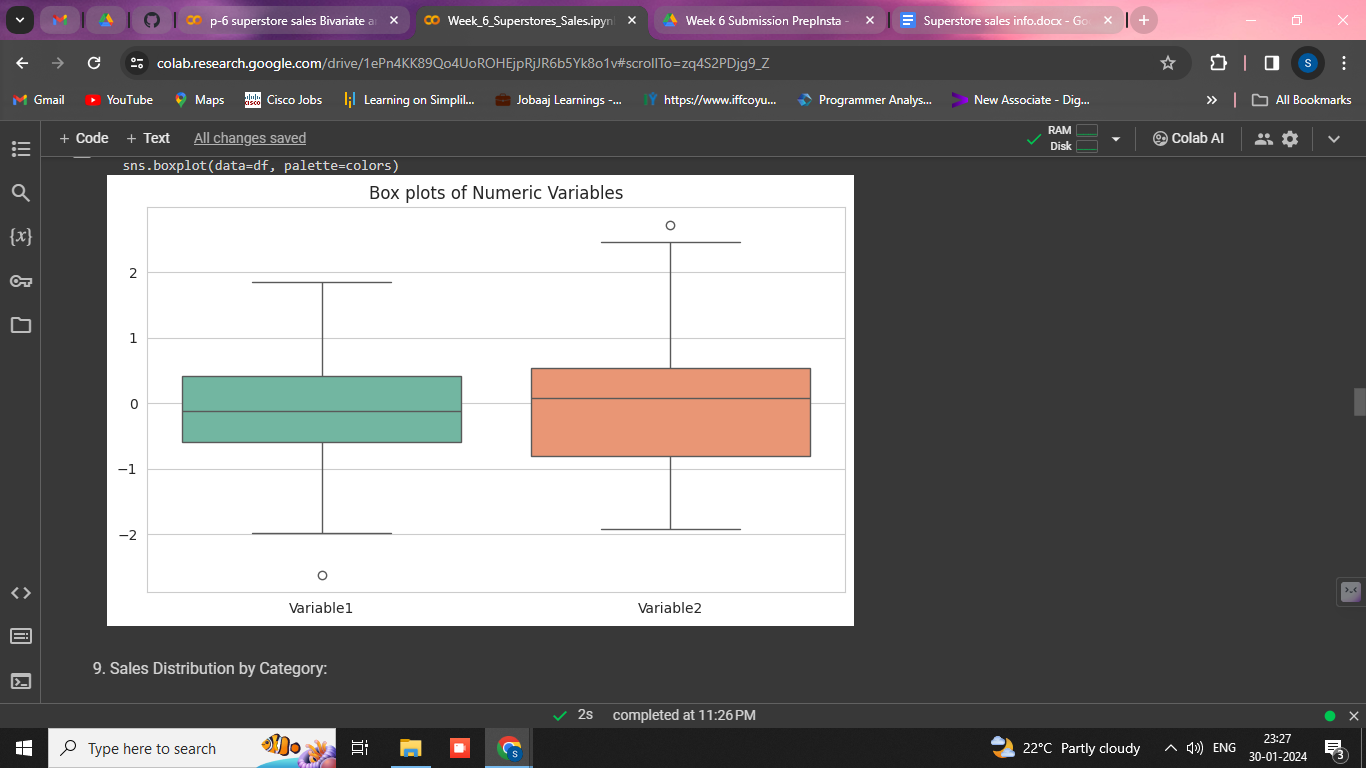
Use seasonal decomposition techniques (e.g., statsmodels.tsa.seasonal.seasonal\_decompose) to decompose time series data.



1. **Box Plots:**

Use box plots to identify outliers and understand the distribution of numeric variables.

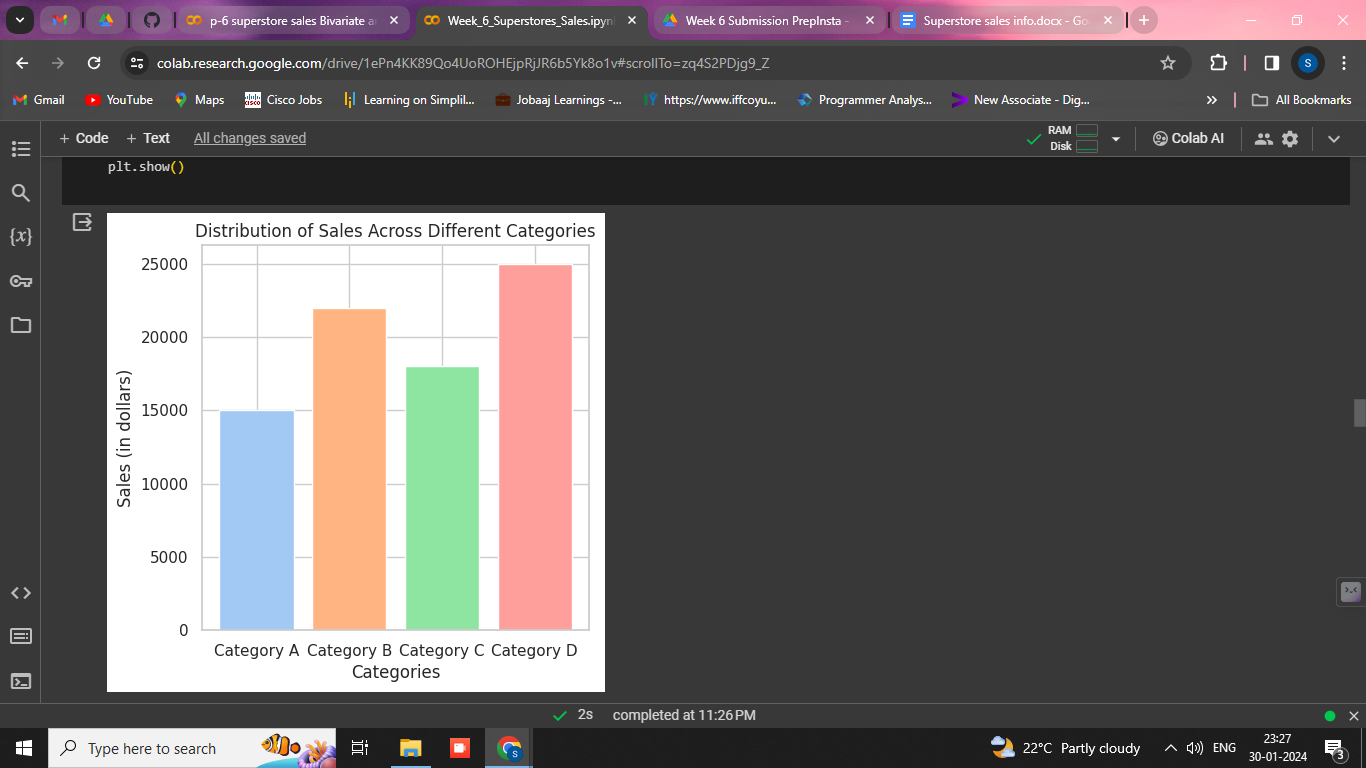
Sales Distribution by Category:

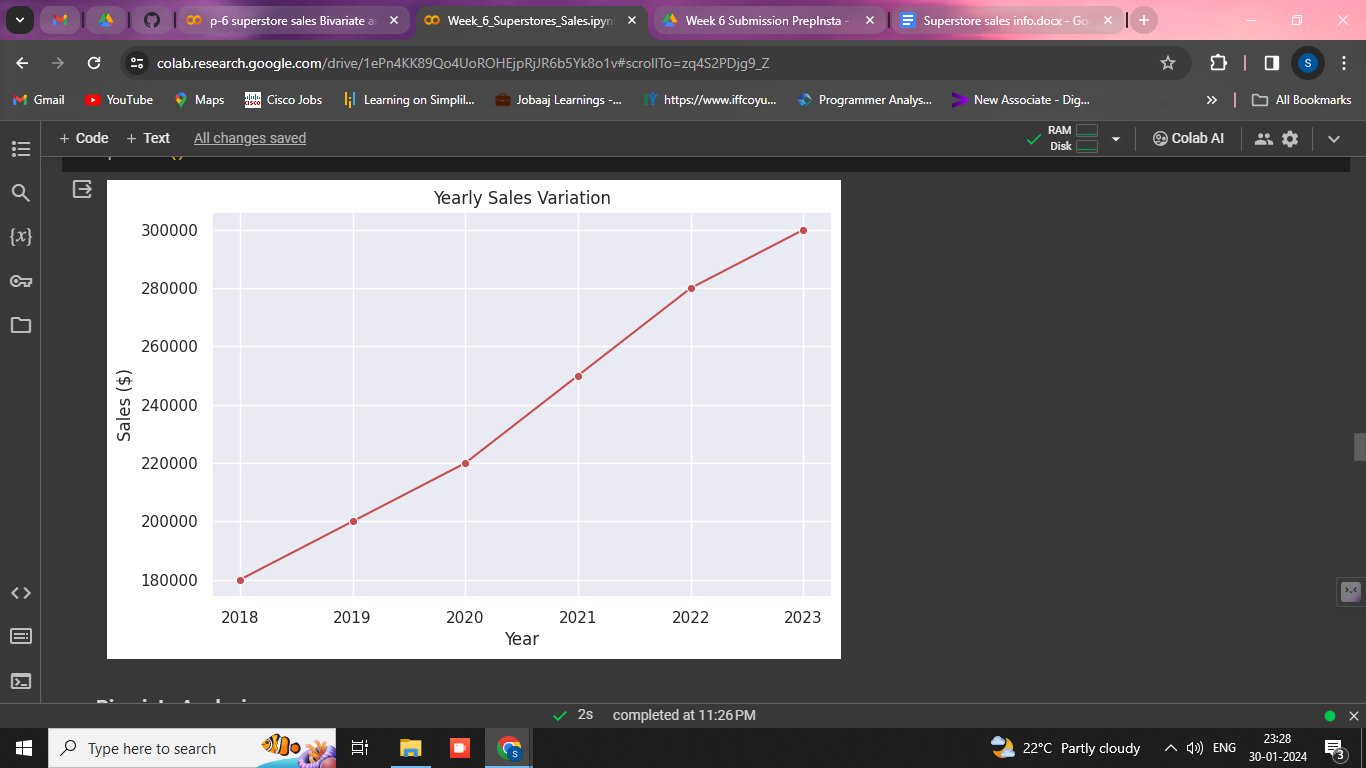
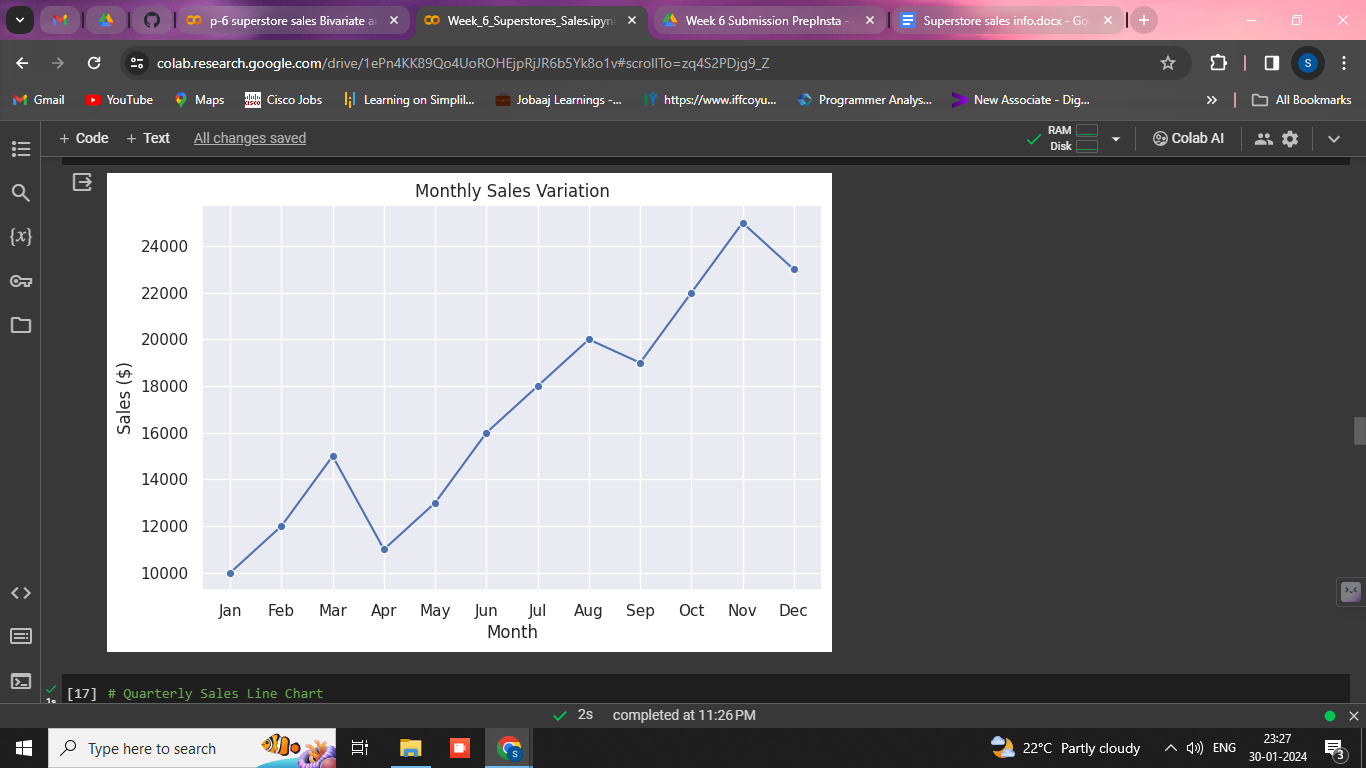
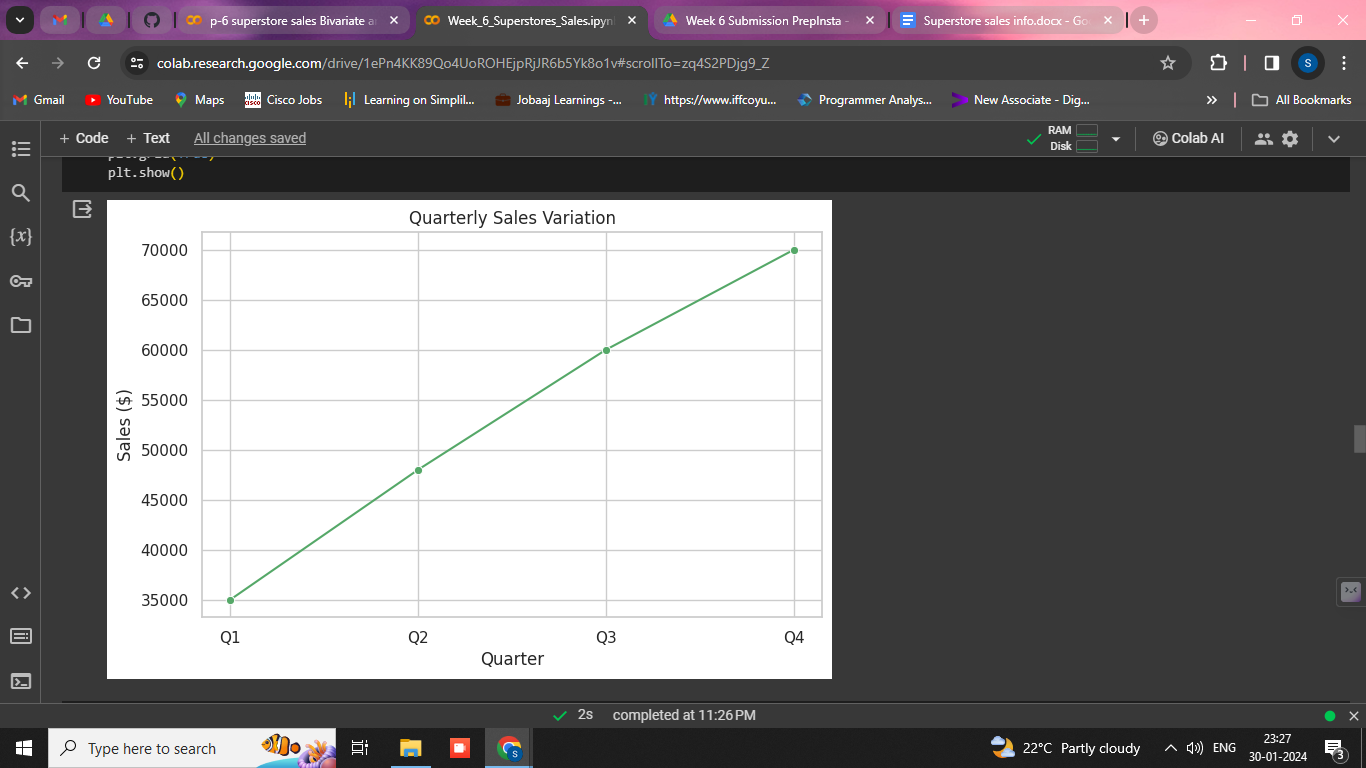
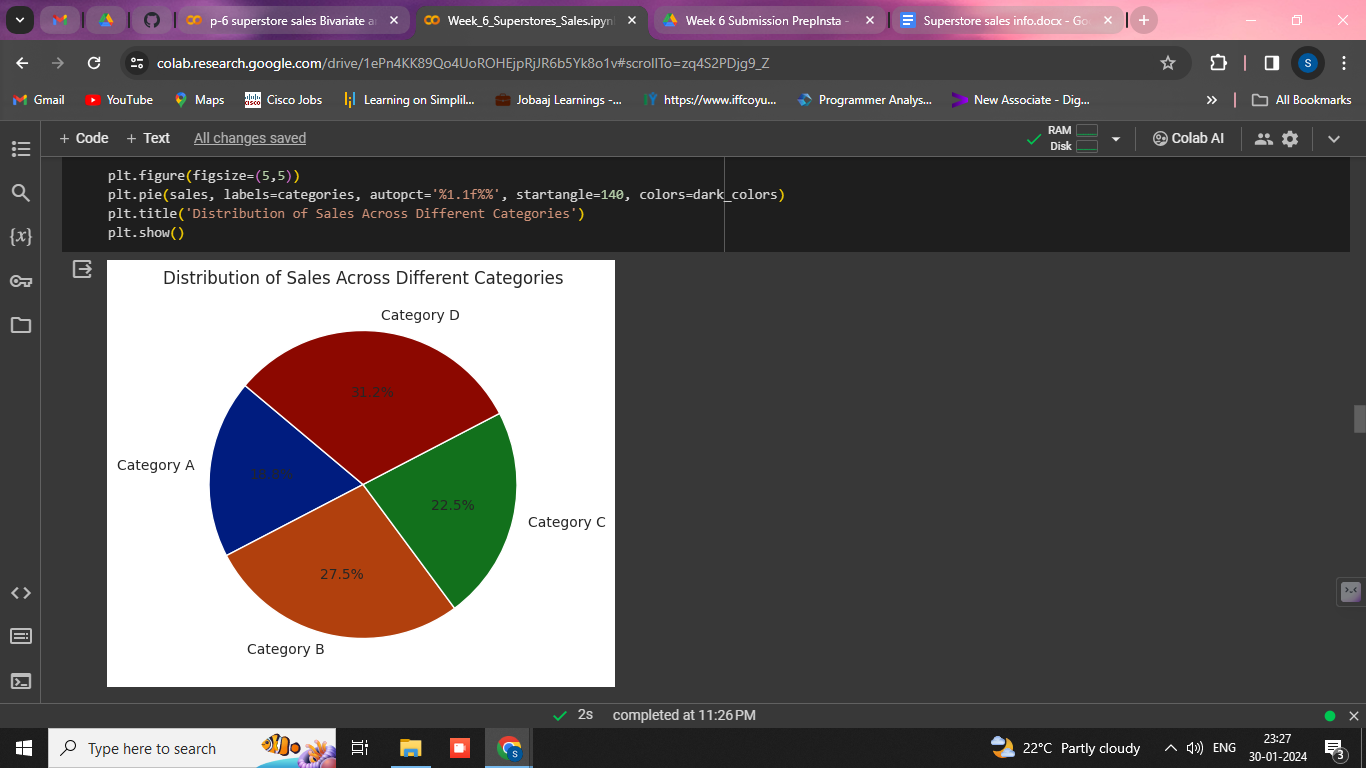


Visualize the distribution of sales across different categories using bar charts or pie charts.

**Sales Variation Over Time:**

Plot line charts to observe how sales vary over different time periods (months, quarters, years).

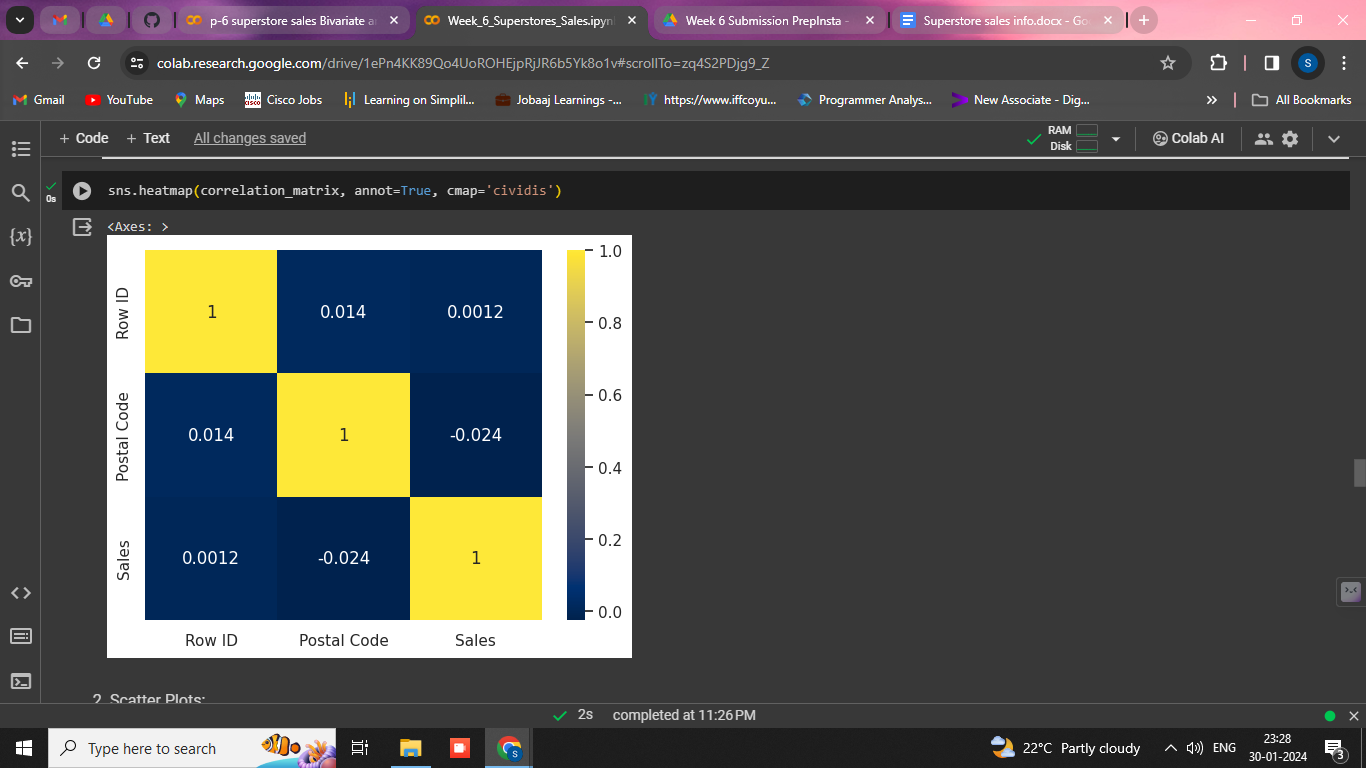


****

**Bivariate Analysis:**

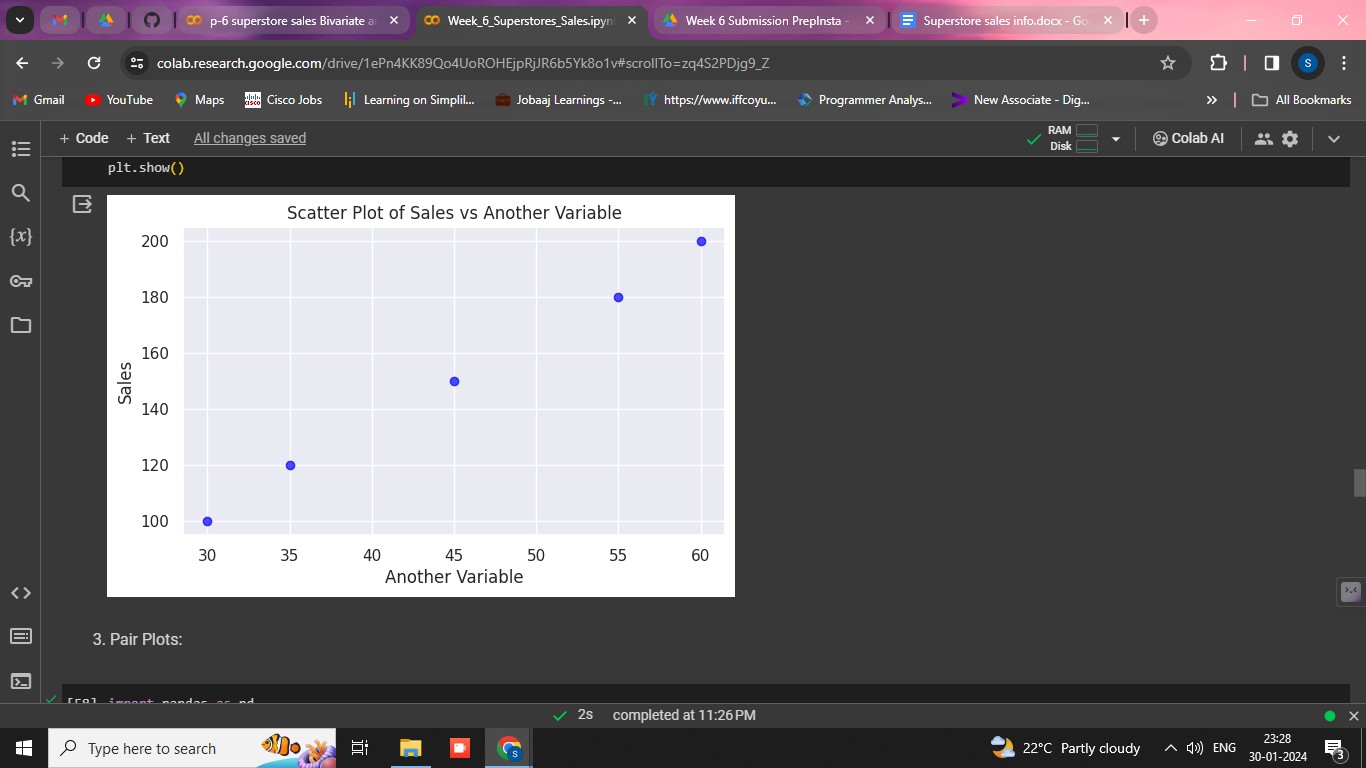
**Correlation Analysis:**

Calculate correlation coefficients between sales and other relevant numeric variables using df.corr().

****

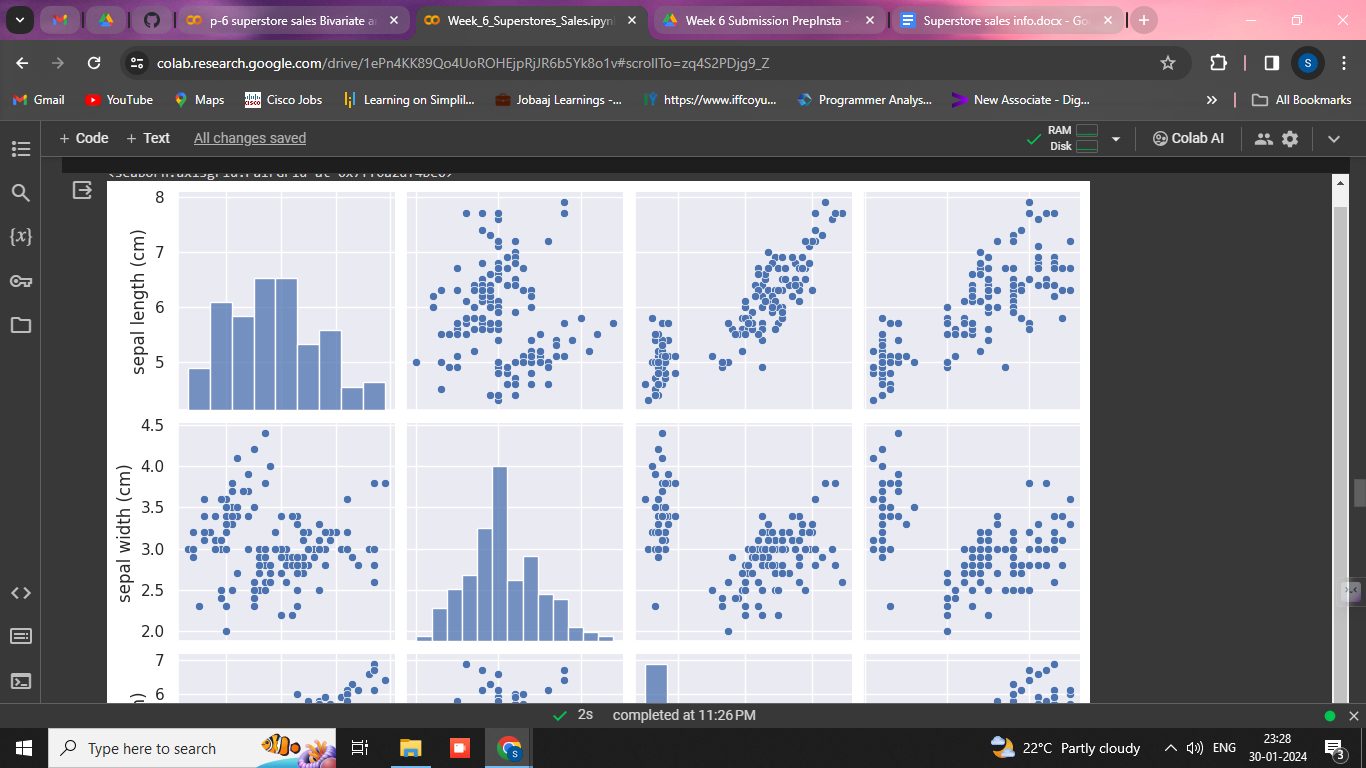
**Scatter Plots:**

Plot scatter plots to explore the relationship between sales and another numeric variable using plt.scatter() or sns.scatterplot().



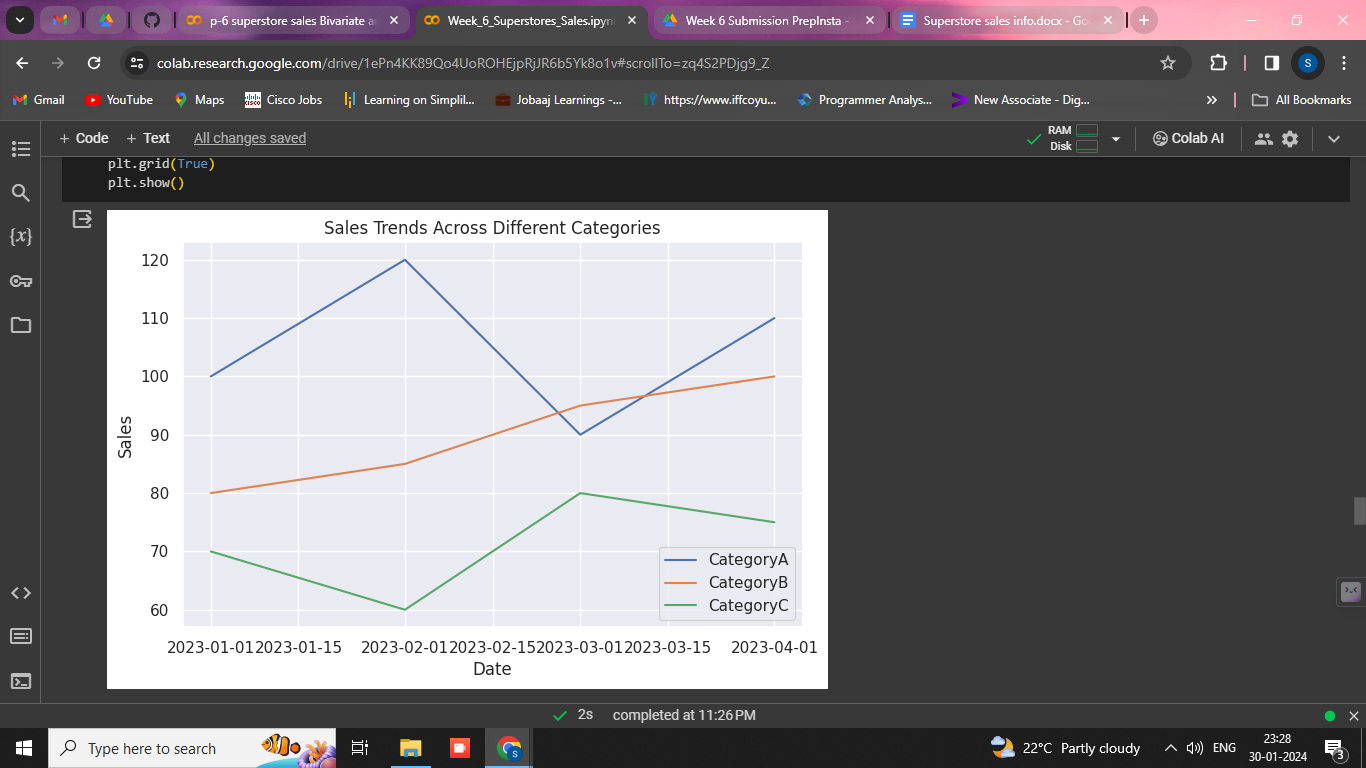
**Pair Plots:**

Use pair plots for a quick overview of relationships between multiple numeric variables using sns.pairplot().



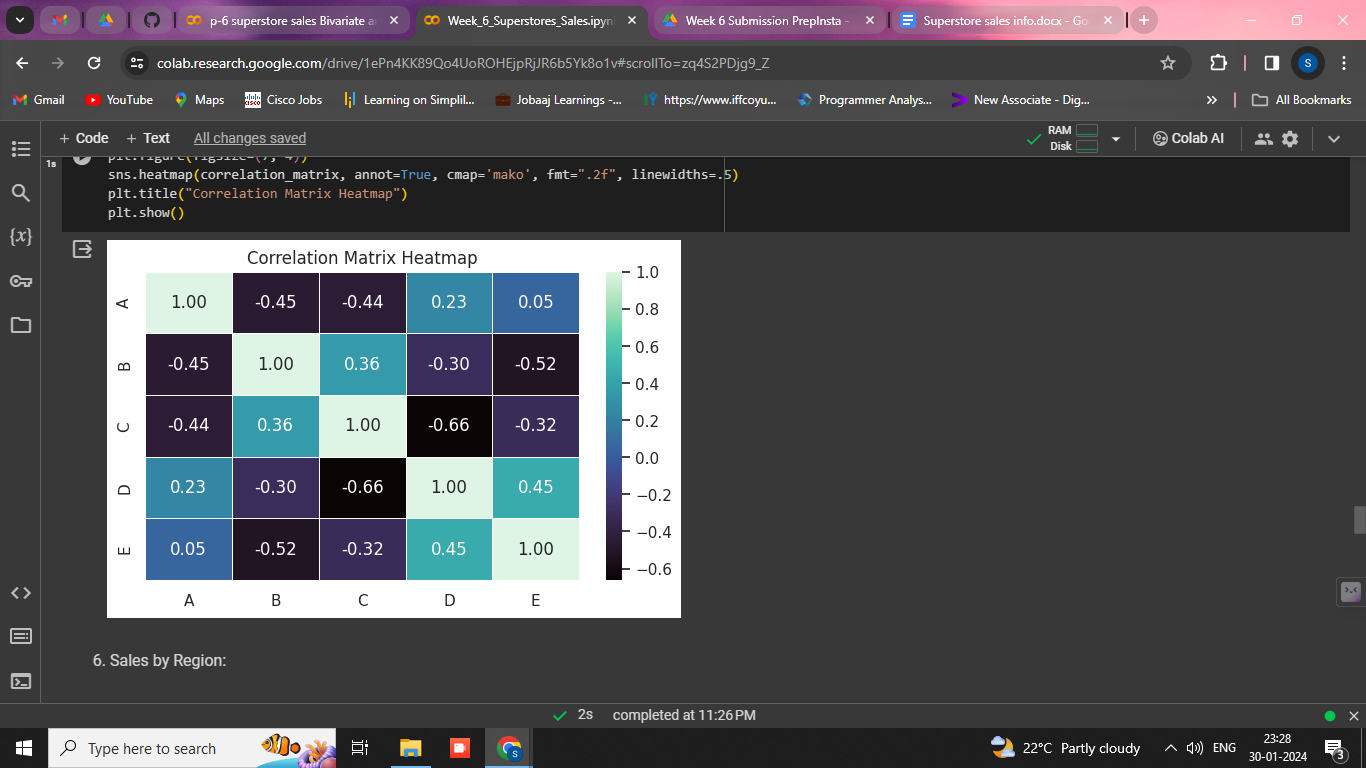
**Category-wise Sales Trends:**

Analyze how sales trends differ across different categories using line charts.



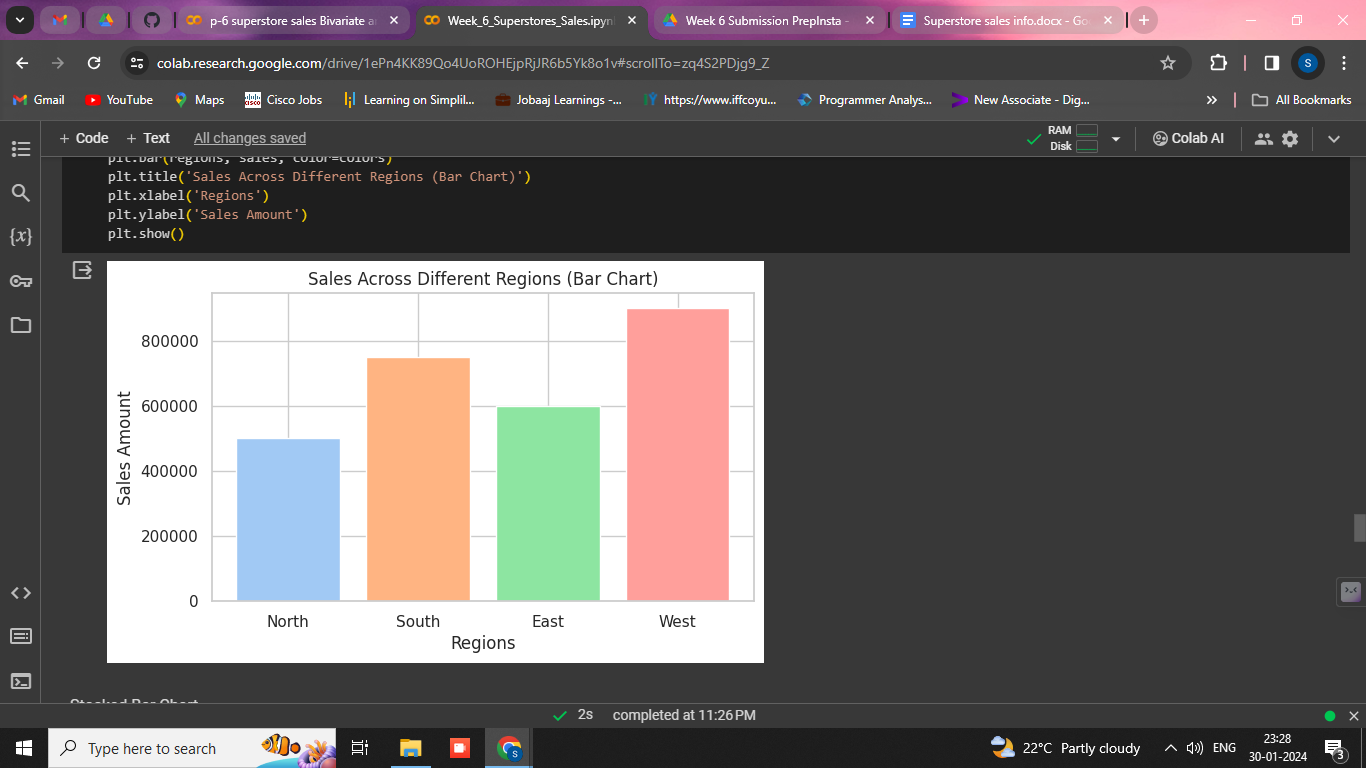
**Heatmaps:**

Create heatmaps to visualize the correlation matrix for better insights.



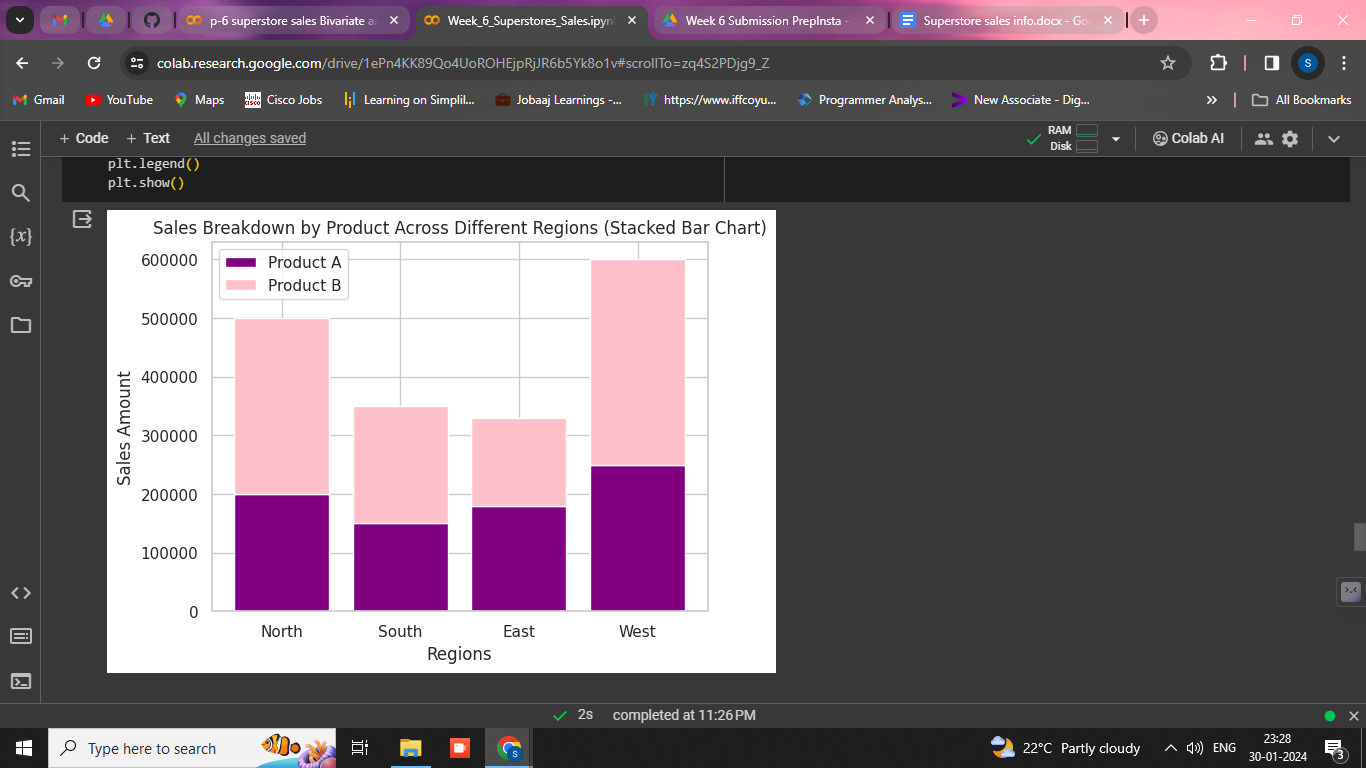
**Sales by Region:**

Compare sales across different regions using bar charts or stacked bar charts.



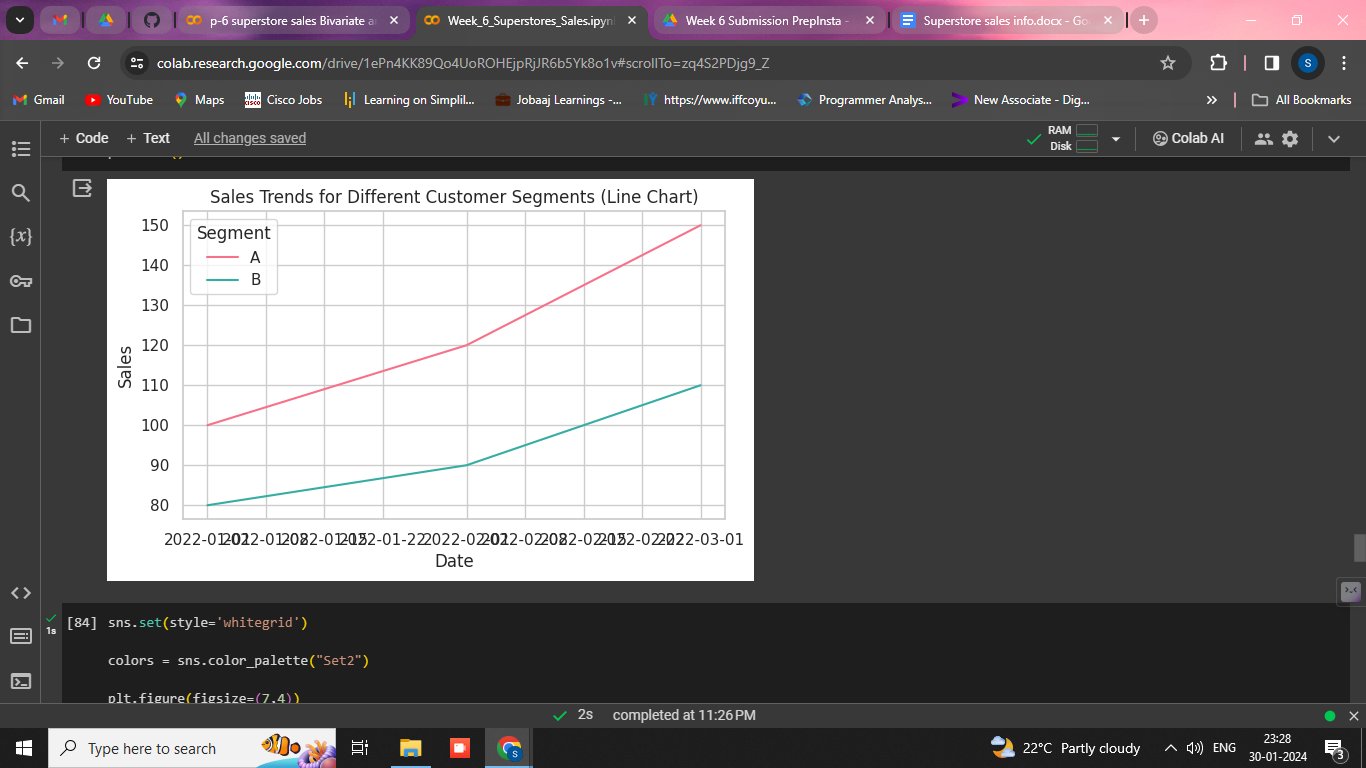
**Customer Segment Analysis:**

Explore sales trends for different customer segments using line charts or bar charts.



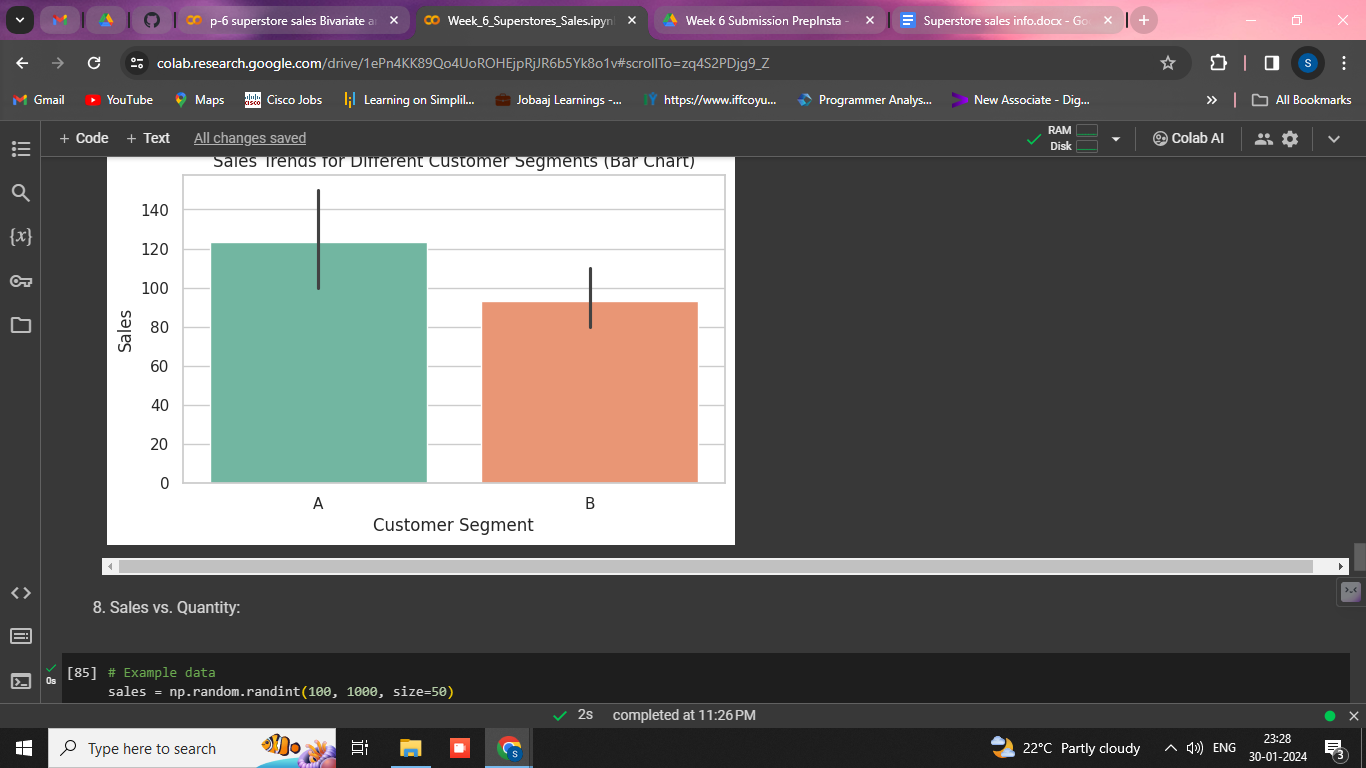
**Sales vs. Quantity:**

Investigate the relationship between sales and quantity sold using scatter plots.



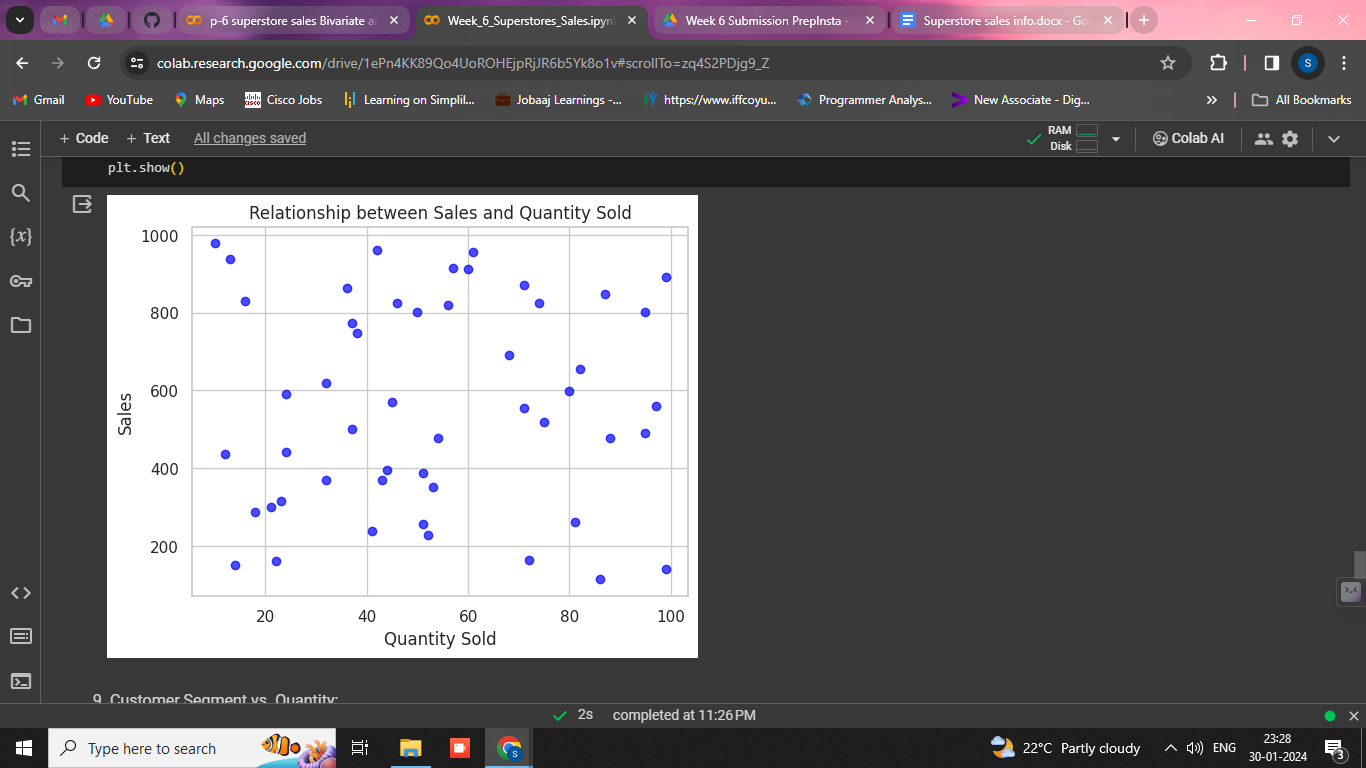
**Customer Segment vs. Quantity:**

Compare the quantity sold across different customer segments using bar charts.



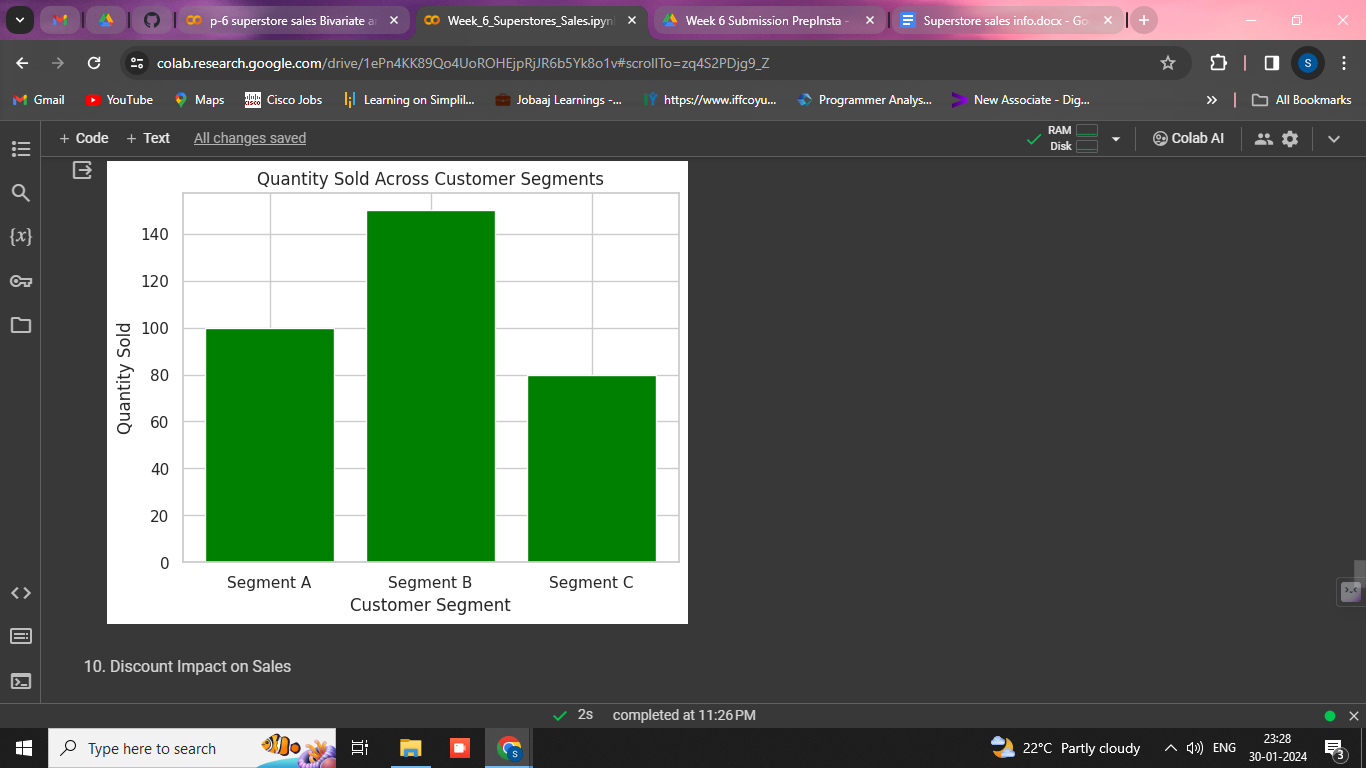
**Discount Impact on Sales:**

Analyze the impact of discounts on sales using bar charts or line charts.



**Shipping Mode vs. Sales:**

Explore the relationship between shipping modes and sales using categorical plots.



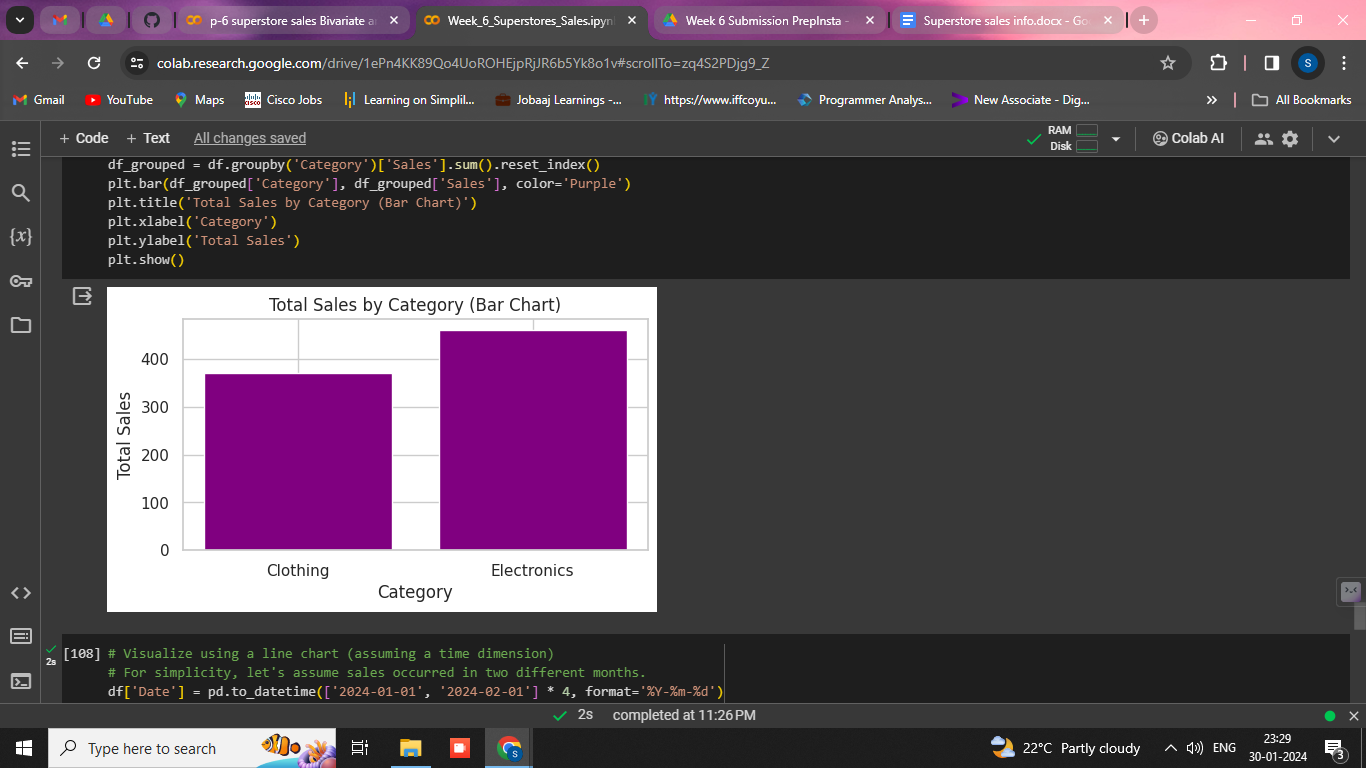
**Product-wise Sales Analysis:**

Investigate how sales vary for different products or categories using bar charts or line charts.



**Region-wise Product Sales:**

Compare product sales across different regions using grouped bar charts.



**Document Findings:**

Summarize key findings from both univariate and bivariate analyses.

Highlight any notable trends or insights.

