**EXPLORATORY DATA ANALYSIS OF NIKE FOOTWEAR AND APPAREL SALES**



**Introduction**

This report presents an **Exploratory Data Analysis (EDA)** of Nike's footwear and apparel sales. The dataset contains detailed transaction records, including sales methods, product categories, pricing, and regional distribution. The objective of this analysis is to uncover key patterns, trends, and insights that can help understand sales performance and customer preferences.

**Aim**

The aim of this report is to conduct an **Exploratory Data Analysis (EDA)** of Nike's footwear and apparel sales to identify key trends, patterns, and insights. This analysis will help understand **sales performance**, **customer preferences**, **regional variations**, and the impact of **pricing and sales methods** on overall revenue. The findings will provide valuable insights for decision-making and strategic planning.

**Objectives**

The primary goals of this analysis include:

1. Identifying **seasonal trends** in sales.
2. Understanding the **best-performing product categories** and consumer preferences by gender.
3. Analysing **regional sales variations** and retailer performance.
4. Evaluating the impact of **pricing strategies** on total sales and revenue.
5. Gaining insights into **sales methods** (online vs. offline) and their effectiveness.

**Data Overview**

**Columns**

Invoice Date datetime64[ns]

Day Name object

Month Name object

Month int32

Year int32

Product object

Product\_category object

Gender object

Region object

Retailer object

Sales Method object

State object

Price per Unit int64

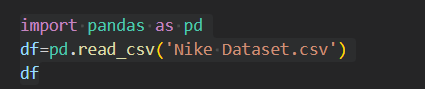
Price in Indian Rupees float64

Price\_category object

Total Sales int64

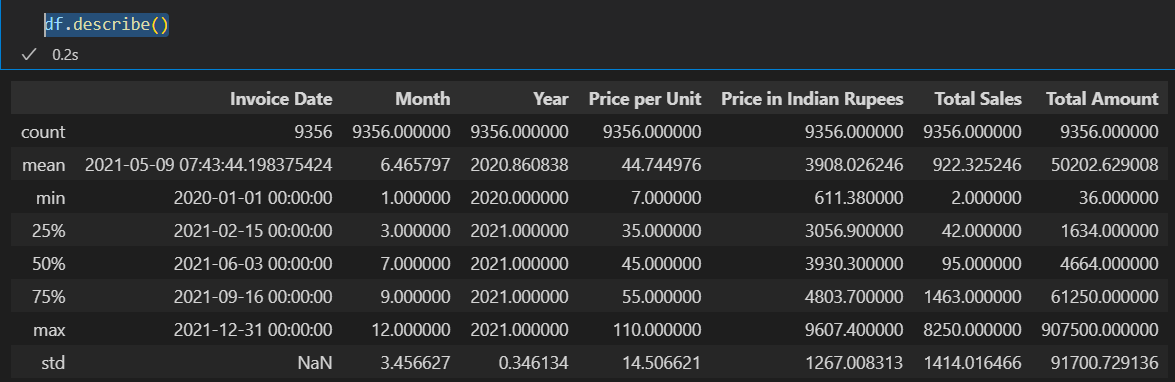
Total Amount int64

**Reading the csv file**



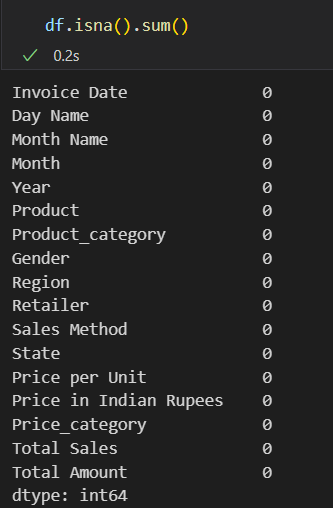
**EDA METHODS**

1. Statistical Summary of the dataset

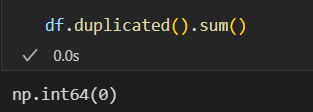


1. Data Cleaning

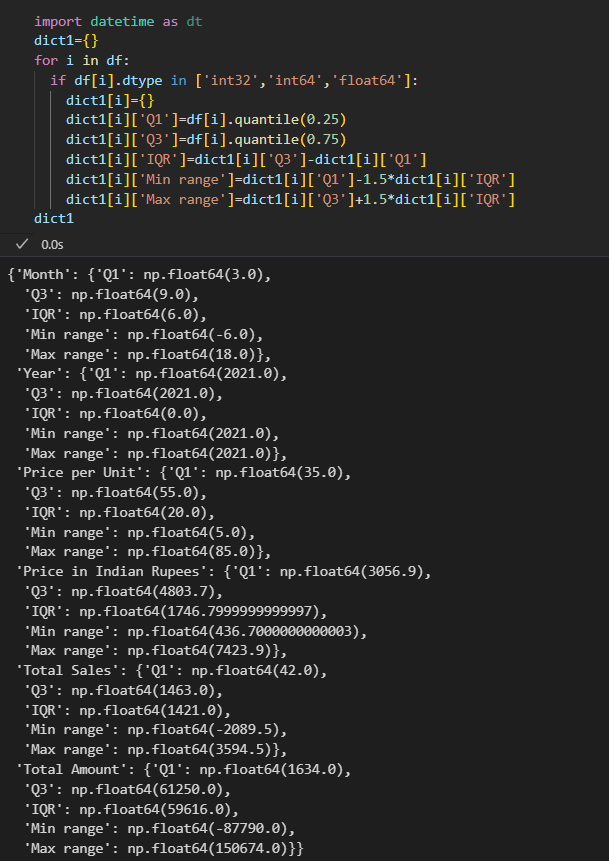
* Missing value handling:- It is noted that there is no missing values while checking.



* Duplicates removal:- It is also noted that there is no duplicate rows in the dataset.

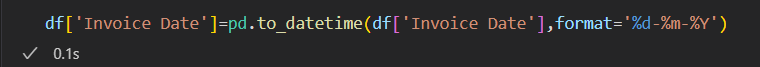


* IQR Method:



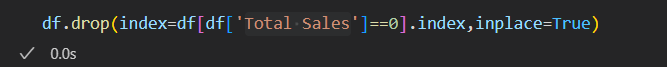
When analysing the **minimum and maximum range** of a dataset, blindly deleting data points based on the **Interquartile Range (IQR) method** can be an illogical approach. This method identifies outliers by filtering values beyond a specific threshold, but in doing so, it may **unintentionally remove crucial data points** that represent natural variations rather than actual anomalies. Therefore, much cleaning using IQR method is needed for this dataset.

* Converting datatypes of columns if needed:-
* Converting the datatype of the column ‘Invoice Date’ from ‘object’ to ‘datetime64[ns].



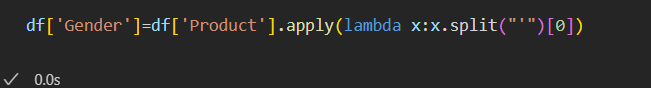
* Other errors:

It is noted that some values in the column ‘Total Sales’ is 0. Thus, removing the rows by the total sales is zero.

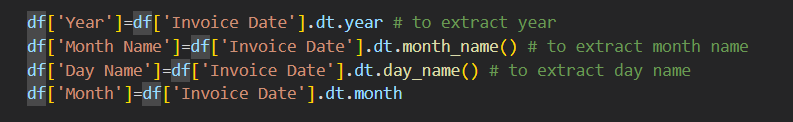


1. Feature Engineering

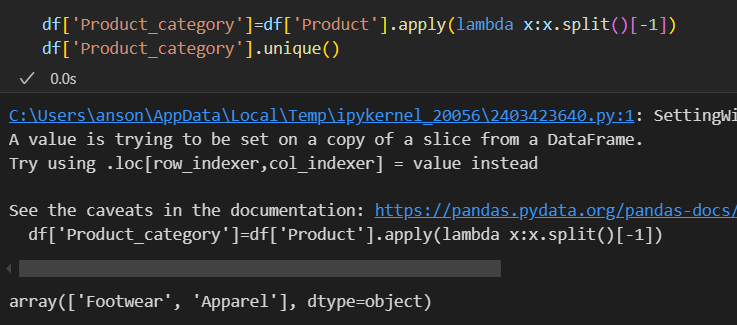
* Created a new column ‘Gender’ which is extracted from the column ‘Product’



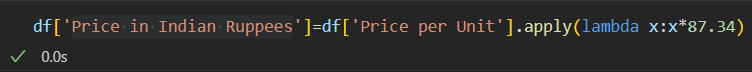
* Created new columns ‘Year’, ‘Month name’, ‘Day Name’ and ‘Month’ and is extracted from the column ‘Year’



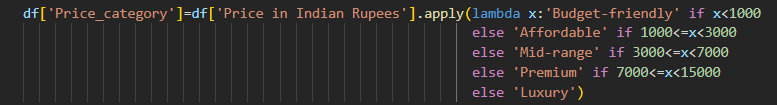
* Created a new column 'Product\_category' extracted from the column ‘Product’



* Created a new column ‘Price in Indian Rupees’ by converting the price in dollars to Indian rupees.



* Created a new column ‘Price\_category’ by analysing the column ‘Price in Indian Rupees’

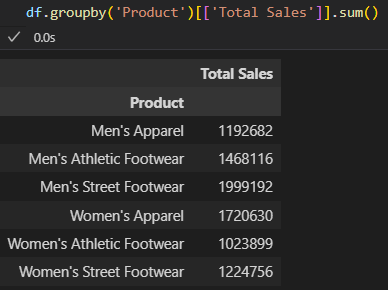


* Created a new column 'Total Amount' which represents total sale amount of each products by multiplying the total sale and price per unit

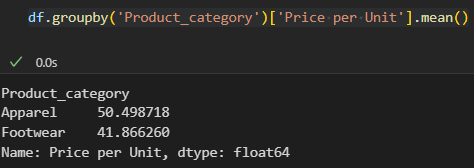


1. Descriptive Analysis

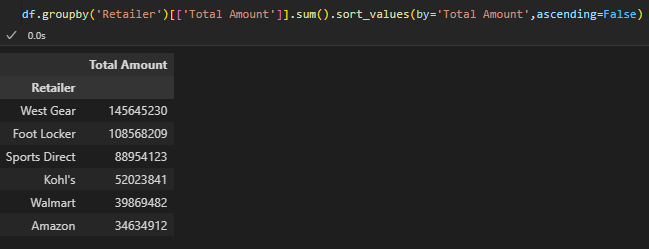
* Univariate analysis
* Total sales among each products



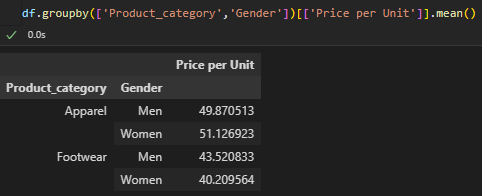
* Average price per unit for each category



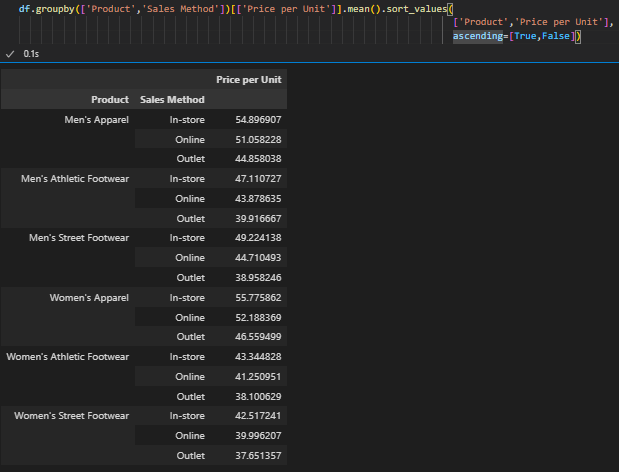
* Total amount generated per retailers



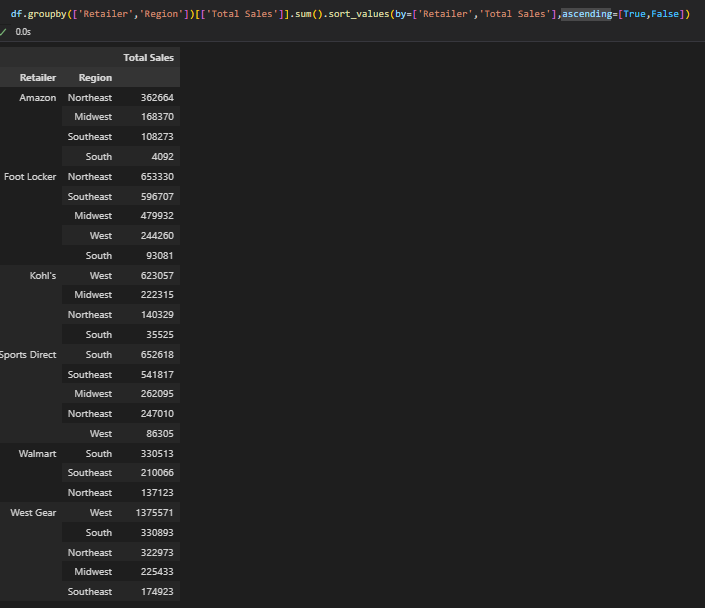
* Bivariate Analysis
* Average price per unit of each product among each genders



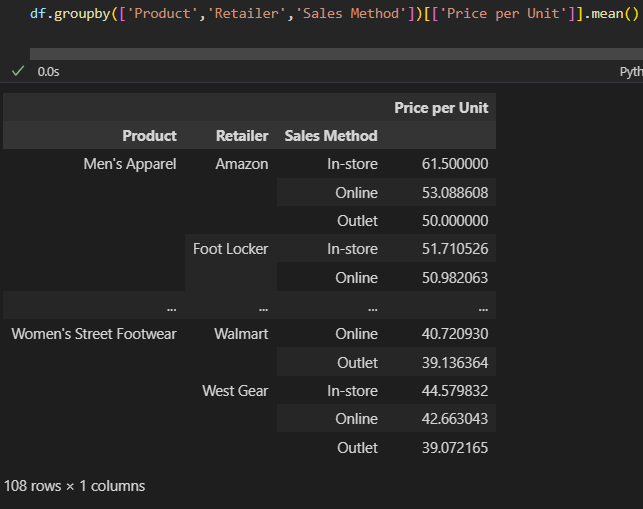
* Average price of each products among each sales methods



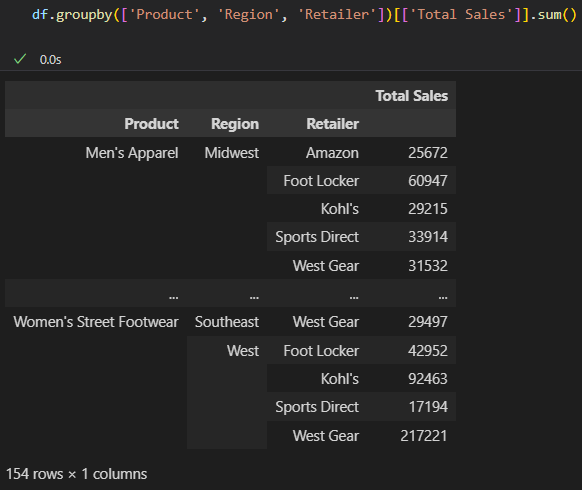
* Total sales of each Retailers among each regions



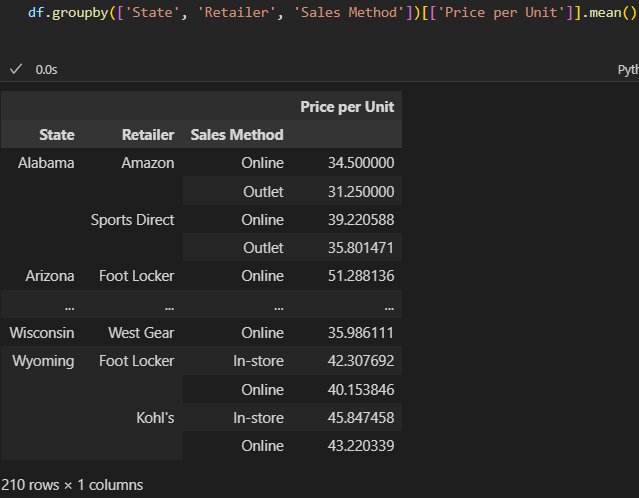
* Multivariate Analysis
* Average price per unit of each products among each retailers with different sales method



* Total sales per products, regions and retailers

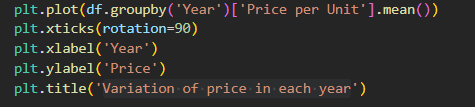


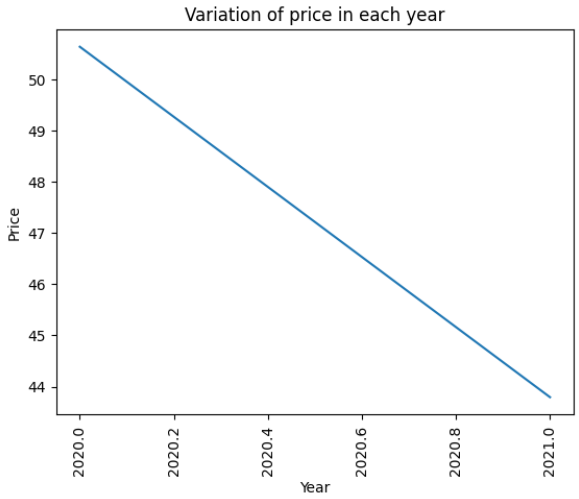
* Average price generated per state, retailers, and sales methods



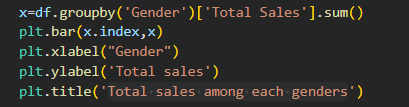
1. Visualisation

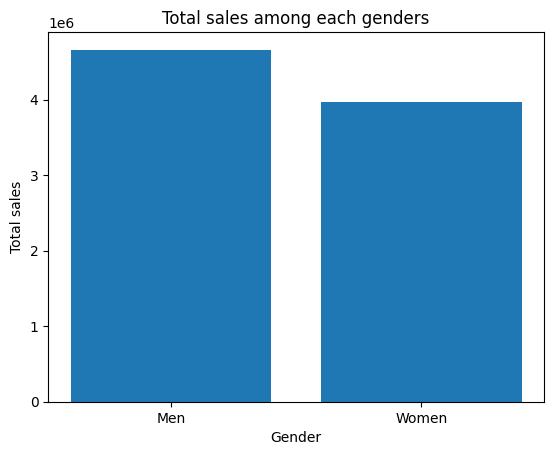
* Variation of price in each year



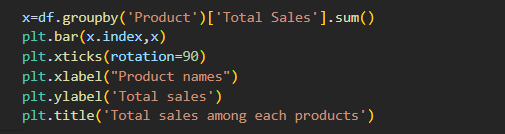


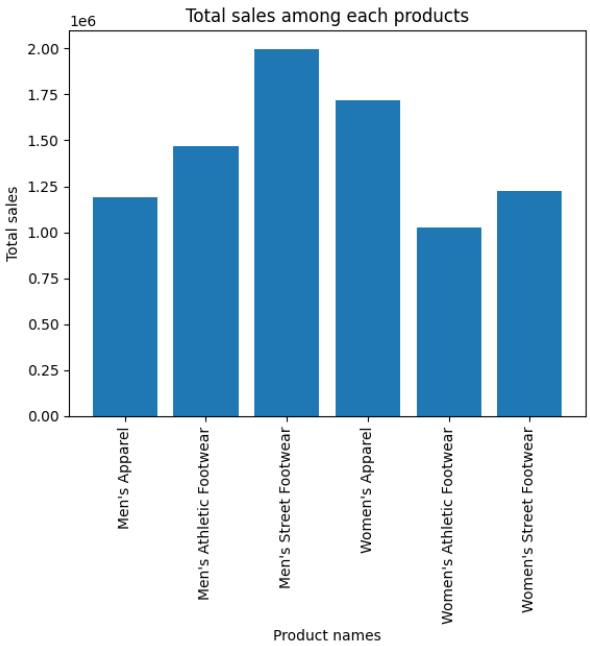
* Total sales among each genders



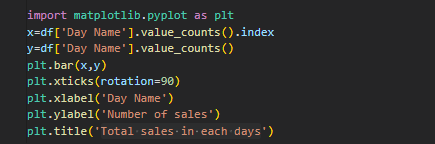


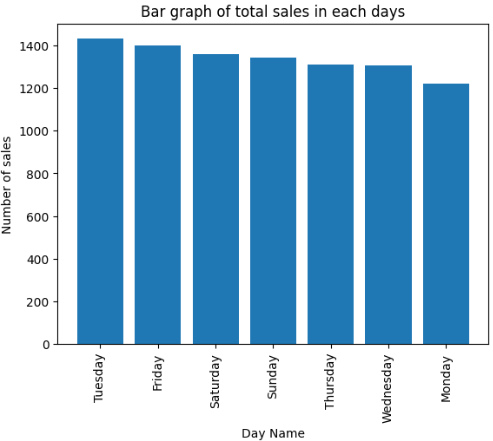
* Total sales among each products





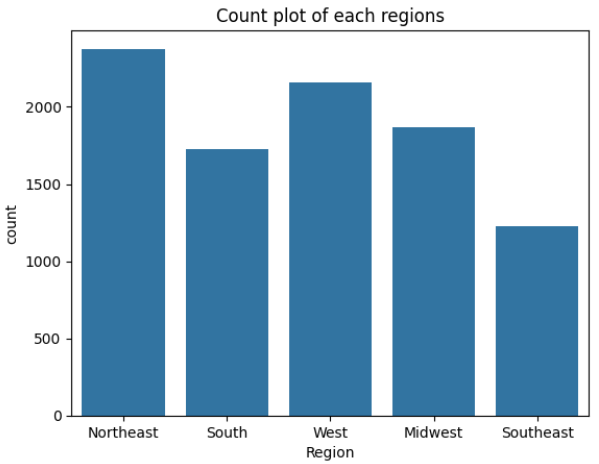
* Total sales in each days



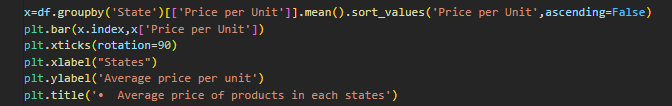


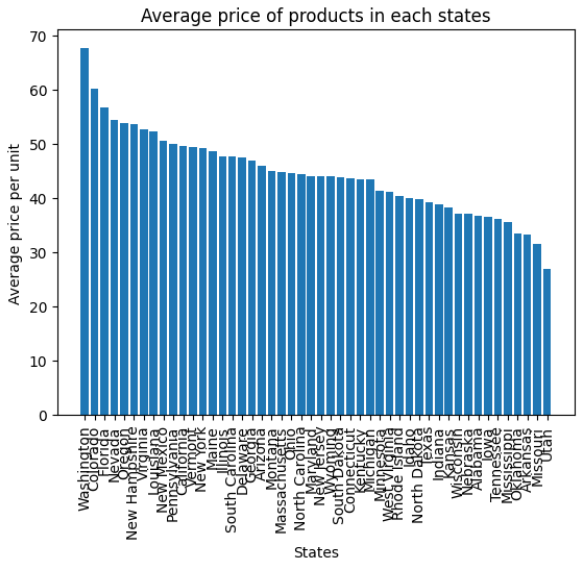
* Count plot of each Regions



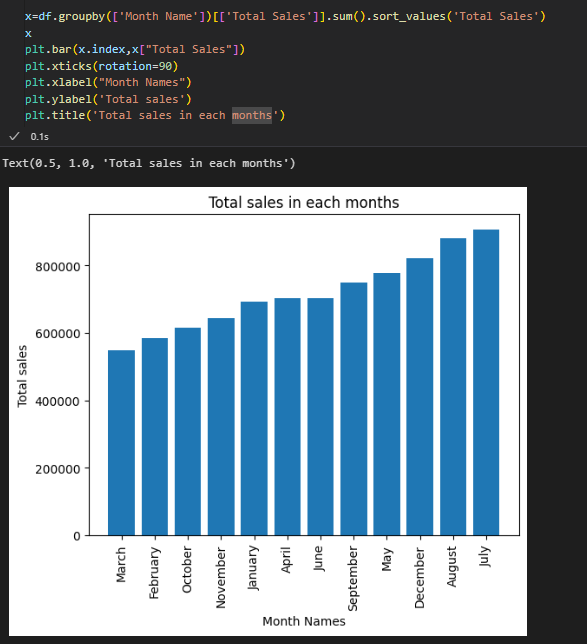


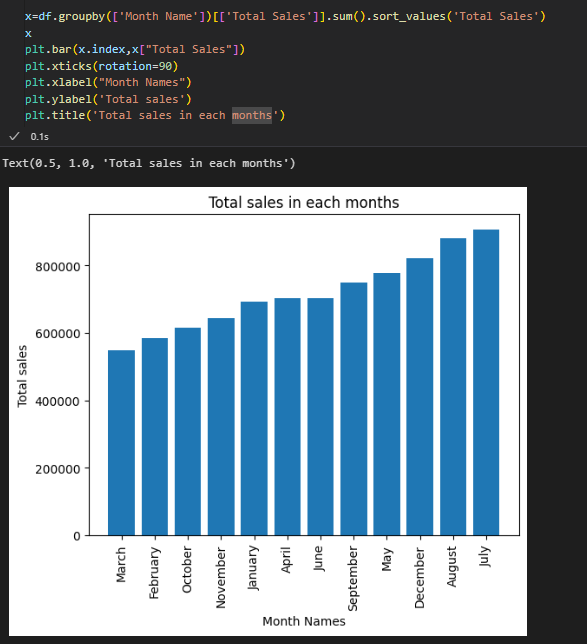
* Average price of products in each states.



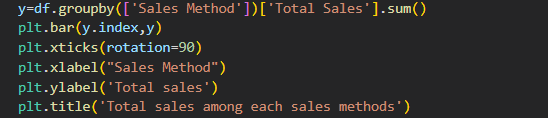


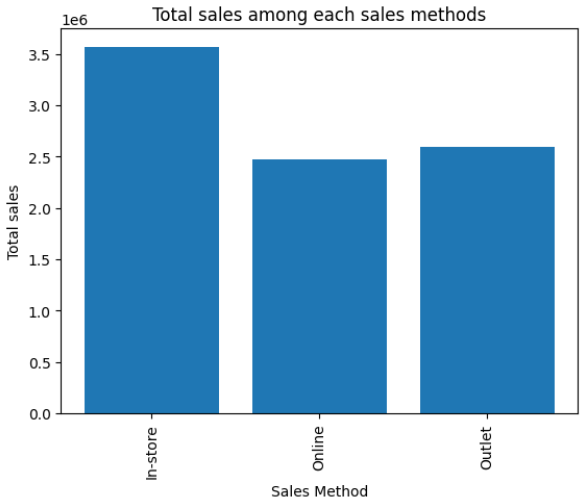
* Total sales in each month



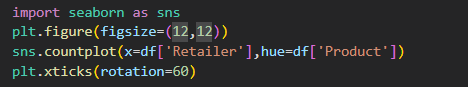


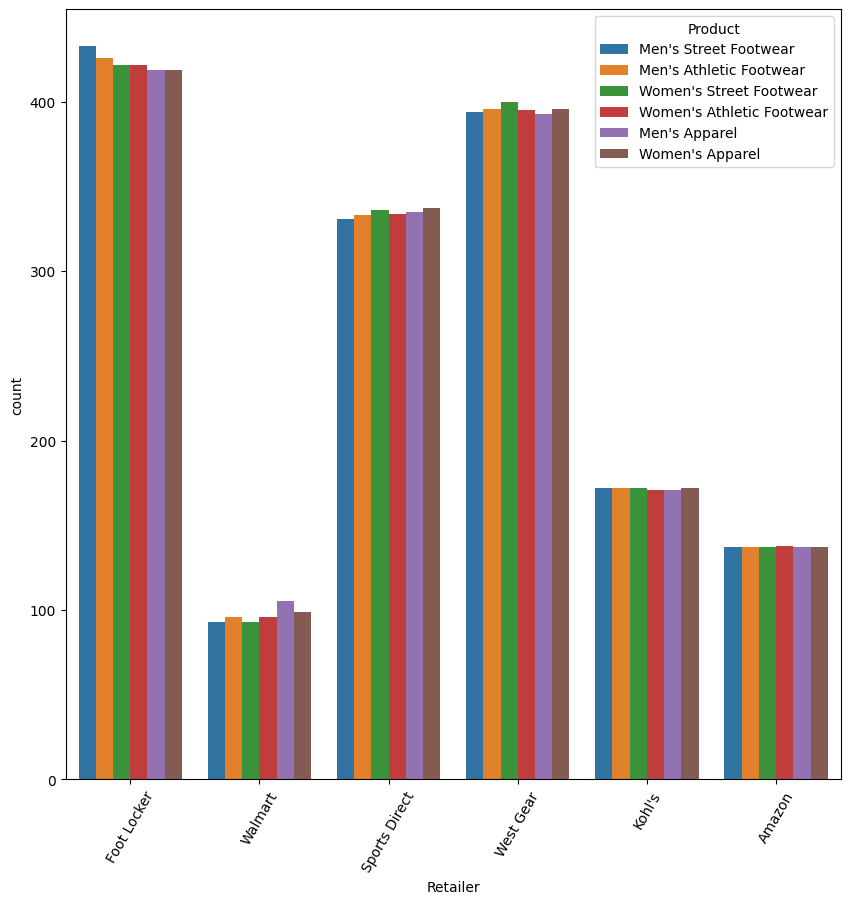
* Total sales among each Sales Methods





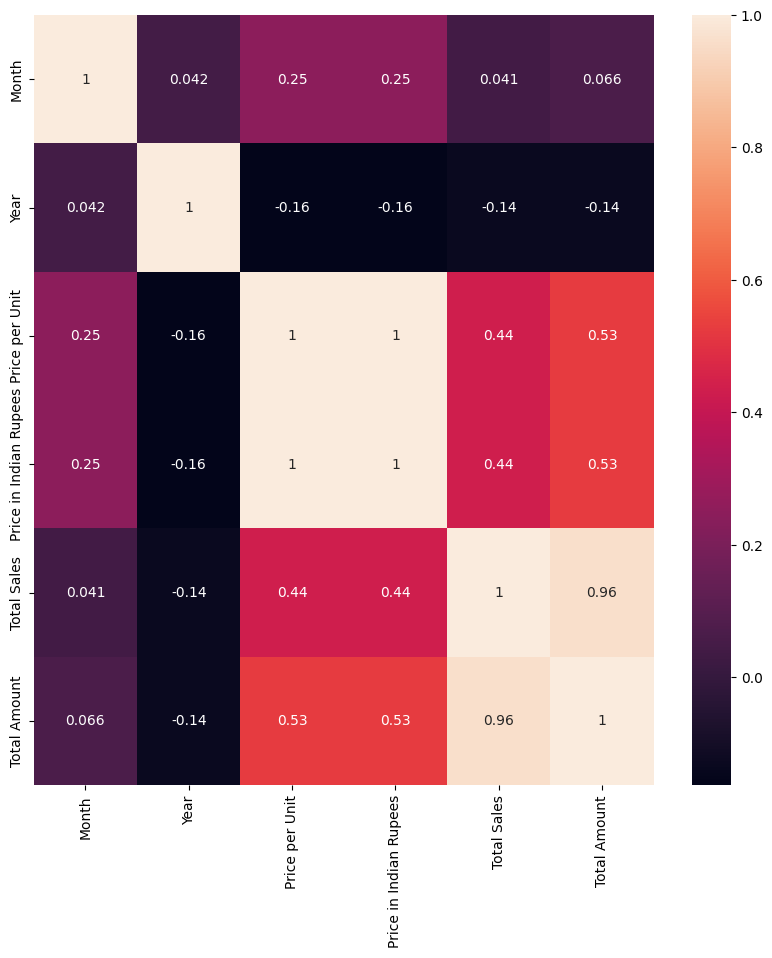
* Count plot of products among each Retailers





* Heatmap of corelation matrix





**RESULT**

1. It is noted that Men's Street Footwear had the highest total sales, while the lowest was for Women's Athletic Footwear.
2. Apparels had the highest average price than Footwears.
3. West Gear was the retailer in which generated the maximum total amount while the least one is Amazon.
4. The variation among the year and price per unit was inversely proportional.
5. Total sales was highest in the case of men than women.
6. The northeast regions had the highest purchases and the least one was for southeast regions.
7. The average price was highest at Washington and lowest at Utan.
8. July had the higher number of sales and lower one in march.
9. More people purchase the products using In-store than Outlet than Online method.