1. Basic Insights:

* Change the date to short date as date is not visible properly by changing its format to short date.
* Total No. Of Rows: 5001
* Total No. Of Columns: 8
* Total No. Of Features: 8
* Missing values in dataset: 0 (Uses COUNT and COUNTA to find missing values of each column).
* Duplicates: There are no duplicate values in the dataset. Select all dataset and use remove duplicates.
* Unique values: In date column there are 581 unique values. In Product ID column there are 888 unique values. Store ID, Price per unit and discount rate have 0 Unique values. Sales volume has 148 unique values. Sales revenue has 1000 unique values. Customer ID has 903 unique values. All values are seen in the power query by checked column distribution in VIEW section.

1. Statistical Summary:

* First select all dataset then in data tab go to data analysis select descriptive statistics select range and checked options like labels in the first row, summary statistics, confidence interval for mean, Kth largest, Kth smallest then press OK.
* It gives statistical values for our dataset which helps to find out the distribution of our data or how our data spread out.
* Date column indicates that our data contain Negative or Left Skewed Data because the value of mean is less than median. The kurtosis value of date column indicate that it contains less outliers because it’s value is less than 3. The skewness value indicates that our column contains larger values because it is negative.
* Product ID column indicates that our data contain Positive or Right skewed data because the value of mean is greater than median. Kurtosis value shows that it contain less outliers because it’s value is less than 3. The skewness value indicates that our dataset is good for analysis.
* Store ID column indicates that our data contain Negative or left skewed data because the value of mean is less than median. Kurtosis value shows that it contain less outliers because it’s value is less than 3. The skewness value indicates that our dataset is good for analysis.
* Sales volume column indicates that our data contain Positive or Right skewed data because the value of mean is greater than median. Kurtosis value shows that it contain less outliers because it’s value is less than 3. The skewness value indicates that our dataset is good for analysis.
* Sales revenue column indicates that our data contain Negative or Left skewed data because the value of mean is less than median. Kurtosis value shows that it contain less outliers because it’s value is less than 3. The skewness value indicates that our dataset is good for analysis.
* Price per unit column indicates that our data contain Negative or left skewed data because the value of mean is less than median. Kurtosis value shows that it contain less outliers because it’s value is less than 3. The skewness value indicates that our dataset is good for analysis.
* Discount rate column indicates that our data contain Positive or Right skewed data because the value of mean is greater than median. Kurtosis value shows that it contains less outliers because it’s value is less than 3. The skewness value indicates that our dataset is good for analysis.
* Customer ID column indicates that our data contain Negative or Left skewed data because the value of mean is less than median. Kurtosis value shows that it contains less outliers because it’s value is less than 3. The skewness value indicates that our dataset is good for analysis.
* The skewness of all columns are closed to 0 so it indicates that our data is not too much tailed towards left or right so our data distribution is Symmetrical or Normal.

1. Feature Wise Interpretation:

* 1) Date: It gives the date for the sales transaction. It gives time-wise analysis.
* 2) Product ID: A unique value for each product. It is crucial for understanding which product are selling well.
* 3) Store ID: A unique value for each store location. It helps to analyze sales across different stores.
* 4) Sales Volume: It shows quantity of products sold in a transaction. It is essential for understanding the customer demands for a product.
* 5) Sales Revenue: It shows the total revenue generated from a transaction. It is used to measure the financial performance of sales generated.
* 6) Price Per Unit: The price of a single unit of the product. It gives the prices of the different products.
* 7) Discount Rate: The percentage discount applied to the sale. It helps to understand the impact of discount on sales.
* 8) Customer ID: A unique value for each customer. It is crucial to understand customer behaviour.

1. Transformation:

* From the date column we extract year, month, quarter, week of year, and day which are used in time-wise analysis of sales data. From this we can analyze how sales volume and revenue changes over time.
* In Product Id and Customer ID we extract number by using replace value. It contains both text and numbers which affect our analysis and also it is not possible to calculate the descriptive statistics as it is calculate only for numeric values.
* We change the data type to whole number because after replacing value it is consider as a text and we need number to calculate descriptive statistics.
* Descriptive Statistics useful to find out how our data is distributed.

1. Insights and Patterns:

* Highest No. of products sold in a year = 366 in 2024, 2028, 2032, 2036.
* Lowest No. of products sold in a year = 119 in 2037.
* Which quarter has the highest no. of products sold = Quarter3 and Quarter4.
* Which quarter has the lowest no. of products sold = Quarter1.
* Which month has the highest products sold = January, March, May, July, August, October, December.
* Which month has the lowest products sold = February.
* Which year has highest average sales revenue = 51,840 in 2036.
* Which year has lowest average sales revenue = 42,524 in 2037.
* Which quarter has highest average sales revenue = Quarter4.
* Which quarter has the lowest average sales revenue = Quarter1.
* Which month has the highest average sales revenue = March.
* Which month has the lowest average sales revenue = September.