Supermarket Sales Analysis illustrates the sales made in a Supermarket starting from 2021 to 2023. Additionally, key factors are highlighted and they are enumerated below-

1. Central Tendency – It refers to the statistical measure that describe the center or typical value of a dataset.

* There are three values which I have calculated Average, Median and Mode.
* By performing this analysis, we can generate the insights on the factors affecting the process along with identifying the outliers present in the process.
* Formulas used –
* Average/Mean - =AVERAGE (NUMBER1, NUMBER2, …)
* Median - =MEDIAN (NUMBER1, NUMBER2, …)
* MODE - =MODE.SNGL(NUMBER1, NUMBER2, …)

1. Measures of Dispersion – They are the statistical tools that quantify the extent to which data points in a dataset differ from each other and from a central value.

* In the measures of dispersion there are two types which I have measured; Variation and Standard Deviation.
* In this dataset it is specifically used for the year 2021.
* Formulas used –
* Variance - =VAR (NUMBER1, NUMBER2, …)
* Standard Deviation - =STDEV.S(NUMBER1, NUMBER2, …)

1. Minimum and Maximum – This function is used to determine the smallest and largest values in the dataset.

* This function is used in this dataset to evaluate two factors affecting the process. Function is used to calculate the minimum and maximum value of the revenues and calculated the difference between the revenue generated in the year 2021 and the revenue generated in the year 2023.
* Formulas used –
* Minimum - =MIN (NUMBER1, NUMBER2, …)
* Maximum - =MAX (NUMBER1, NUMBER2, …)

1. Quartiles and Percentiles – Quartiles and Percentiles are measure used to describe the distribution of data by dividing it into two parts

* Quartiles – It divides a dataset into four equal parts, each representing a quarter of the data. It is further bifurcated into three quartiles-
* First Quartile – The median of the lower half of the dataset (25th percentile). It represents the value below which 25% of the data falls.
* Second Quartile – The median of the dataset (50th percentile). It divides the data into two equal halves.
* Third Quartile – The median of the upper half of the dataset (75th percentile). It represents the value below which 75% of the data falls.
* Percentiles – It divides a dataset into 100 equal parts, with each percentile representing 1% of the data.
* Here the Quartiles and Percentiles are used to divided the data, specifically for the year 2021.
* Formulas used –
* Quartile - =QUARTILE (NUMBER1, NUMBER2, …)
* Percentile - =PERCENTILE (NUMBER1, NUMBER2, …)

1. Histogram – It is a graphical representation of the distribution of a dataset. It is used to visualize the frequency of the data points within specific intervals or bins.

* In this dataset, histogram is used to analyze the data for the revenue generated in the year 2021 -2023

1. Pivot Table for each city and country –

* This pivot table provides a clear overview of revenue generated by each city within its respective country. By analyzing this data, you can identify which country generates the least revenue based on the aggregated figures.
* Steps –
* Select your data range.
* Go to "Insert" > "PivotTable."
* Drag "Country" to Rows, "City" to Rows below Country, and "Revenue" to Values.

1. Pivot Table (Average Revenue for Each Product Family in Each City) –

* This pivot table allows for an analysis of average revenue per product family across different cities and countries. It highlights which product families are performing better in specific locations.
* Steps –
* Drag "Country" and "City" to Rows.
* Drag "Product Family" to Rows below City.
* Drag "Revenue" to Values and set it to show average.

1. Multiple Linear Regression – It is a statistical technique used to model the relationship between one dependent variable and two or more independent variables.

* In a regression analysis:
* Dependent Variable: This is typically your outcome variable; in this case, it could be "2023 Sales."
* Independent Variables: These might include "2021 Sales," "2022 Sales," and other relevant predictors.
* You can perform regression analysis using Excel's Data Analysis Tool Pak:
* Go to "Data" > "Data Analysis."
* Select "Regression."
* Define your Y Range (dependent variable) and X Range (independent variables).

1. Regression Analysis Results –

* After running the regression analysis:
* R-squared Value: This indicates how much variance in the dependent variable is explained by the independent variables.
* Significance F: This tests whether your model is statistically significant overall.
* p-values for ‘2021 Sales’ and ‘2022 Sales’: These indicate whether each independent variable significantly contributes to predicting the dependent variable.
* Explanation
  + The R-squared value tells you what percentage of variance in sales for the year 2023 can be explained by changes in sales from previous years.
  + A high R-squared value suggests a strong relationship between your independent variables and the dependent variable.
  + The Significance F-value indicates whether your overall regression model is statistically significant; a value less than a chosen alpha level (typically .05) suggests significance.
  + Finally, p-values for individual predictors indicate whether each independent variable significantly contributes to explaining variations in sales; p-values less than .05 typically suggest significance.