**CUSTOMER SHOPPING DATA DOCUMENTATION**

**ANALYSIS PROCESS:**

* The initial step involved downloading the dataset in CSV format.
* Subsequently, the data was imported into Power Query Editor in Excel to initiate the cleaning process.
* A thorough dataset review was conducted to detect and rectify any null values, duplicates, or blank entries.
* Before proceeding with the cleaning process, the total number of rows in the dataset was determined to be 99458.
* A check for duplicate entries was performed to ensure their authenticity, revealing no duplicates.
* No blank entries were found in the "Invoice Number" and "Customer ID" columns.
* The data was then converted into a Pivot table for further analysis.
* Using the Pivot table, the total number of invoices was calculated to be 99457.
* Employing both 'COUNTA' and 'UNIQUE' functions, the total number of unique customers was determined to be 99457.
* The average age of customers was calculated using the 'AVERAGE' function, yielding approximately 43.4 years.
* To identify the unique categories of items in the dataset, the 'SORT' and 'UNIQUE' functions were utilized, revealing 8 categories: Books, Clothing, Cosmetics, Food and Beverage, Shoes, Souvenirs, Technology, and Toys.
* Employing a combination of 'INDEX', 'MATCH', and 'MAX' functions, the category with the highest quantity sold was determined to be 'Clothing'.
* A Pivot table was utilized to calculate the average price per category.
* To examine the distribution of payment methods, a Pivot table was employed, and a pie chart was generated to visualize the data.
* The shopping mall with the highest number of transactions was identified as 'Forum Istanbul' using a Pivot table.
* Utilizing a Pivot table, the average quantity of items purchased in each mall was determined.
* Gender distribution within the dataset was analyzed using a Pivot table, and the results were represented using a bar chart.
* Total revenue was calculated by inserting a new column, dividing the quantity by price, and summing the total amount, resulting in $251,505,794.25.
* Customers were grouped into different age brackets using the 'COUNTIFS' function to count the number of customers in each group.
* A bar chart was generated to illustrate the distribution of customers across different age groups.
* A scatter plot was employed to visualize the relationship between age and the quantity of items purchased.
* The top 5 customers were identified using filters, and their data was represented using a bar chart