Introduction to Power BI, Charts, DAX & Creating Reports

Question 1 : Define Power BI and What are the key components of the Power BI ecosystem? Briefly explain:
● Power BI Desktop
● Power BI Service
● Power BI Mobile
● Power BI Gateway

Answer :

- **Power BI Desktop**: This is a free Windows desktop application used to connect to data, transform and model it, and create reports and visualizations. It's the primary authoring tool for Power BI.
- **Power BI Service**: This is a cloud-based SaaS (Software as a Service) where users can publish, share, and collaborate on their reports and dashboards. It also allows for refreshing data, creating apps, and managing access to content.
- **Power BI Mobile**: These are native mobile applications for iOS, Android, and Windows devices that allow users to view and interact with reports and dashboards created in Power BI Desktop and published to the Power BI Service, enabling access to data insights on the go.
- **Power BI Gateway**: This is a software component that acts as a bridge, securely connecting Power BI cloud services to on-premises data sources (like SQL Server databases, SharePoint, or Excel files). It allows for refreshing data that resides behind a firewall and sending queries from the Power BI Service to these on-premises sources.

Question 2 :

Compare the following Power BI visuals:

● Pie Chart vs Donut Chart

● Bar Chart vs Column Chart

When would you prefer one over the other? Give one example for each pair.

Answer :

- **Pie Chart**: A circular statistical graphic, which is divided into slices to illustrate numerical proportion. The arc length of each slice, and consequently its area, is proportional to the quantity it represents.
- **Donut Chart**: Similar to a pie chart, but with a blank center, allowing for additional information (like a total value) to be displayed within the hole. This often makes them slightly less cluttered for comparison of slices, as the focus is on the arc length rather than the area.

When to prefer one over the other:

- **Pie Chart**: Prefer a pie chart when you want to quickly show the proportion of a few categories relative to a whole, especially if one category is significantly larger or smaller than others. It's generally less effective for many categories or very similar values.
- **Donut Chart**: Prefer a donut chart when you want to show proportions while also displaying a central metric (e.g., total sales) or when you feel the hole improves readability by reducing the visual weight of the center, making comparisons of arc lengths slightly easier for some users.

Example:

- **Pie Chart Example**: Displaying the market share of 3-4 major smartphone brands in a country. The focus is on how each brand contributes to the total market.
- **Donut Chart Example**: Showing the percentage breakdown of different product categories in total revenue, with the total revenue amount displayed in the center of the donut.

Bar Chart vs. Column Chart

- **Bar Chart**: A chart that represents categorical data with rectangular bars with lengths or heights proportional to the values that they represent. A bar chart can be plotted horizontally or vertically. In Power BI, a 'Bar Chart' specifically refers to the horizontal orientation.
- **Column Chart**: A chart that represents categorical data with rectangular bars with lengths or heights proportional to the values that they represent. A column chart is typically plotted vertically. In Power BI, a 'Column Chart' specifically refers to the vertical orientation.

When to prefer one over the other:

- **Bar Chart (Horizontal)**: Prefer a horizontal bar chart when you have many categories, or when the category names are long. The horizontal orientation allows for easier reading of long labels without rotation, and it can be more effective for comparing values across many items.
- **Column Chart (Vertical)**: Prefer a vertical column chart when comparing values across a small number of categories, or when visualizing time-series data (e.g., monthly sales), as time naturally flows from left to right. It's also often used when the magnitude of the values is the primary focus.

Example:

- **Bar Chart Example**: Comparing the sales performance of the top 15 product categories, where the category names might be lengthy (e.g., 'Portable Bluetooth Speakers with Noise Cancellation'). The horizontal layout prevents label overlap.
- **Column Chart Example**: Showing the monthly sales figures for the past 12 months. The vertical bars naturally represent the progression of time and allow for easy comparison of sales volume month-over-month.

Question 3 : Explain the significance of:

● Star schema vs Snowflake schema

● Primary key vs Foreign key in relationships (Power BI)

Why is cardinality important?

Answer :

- **Star Schema**:
    - **Structure**: It consists of a central **fact table** surrounded by multiple **dimension tables**, forming a 'star' shape. The fact table contains quantitative data (measures) and foreign keys that link to the primary keys of the dimension tables.
    - **Normalization**: Dimension tables are typically denormalized, meaning they contain all attributes related to a dimension (e.g., a 'Product' dimension might have product name, category, subcategory, brand, etc., all in one table).
    - **Advantages**: Simpler to understand and navigate, faster query performance due to fewer joins required, and easier for Power BI to optimize query paths.
    - **Disadvantages**: Can lead to data redundancy in dimension tables if attributes are repeated across multiple rows.
    - **Significance**: Preferred in Power BI for its simplicity and performance benefits, especially for analytical queries.
- **Snowflake Schema**:

    - **Structure**: It's an extension of the star schema where dimension tables are further normalized into sub-dimension tables. This means a dimension table can have its own dimension tables, resembling a 'snowflake' pattern.
    - **Normalization**: Highly normalized, reducing data redundancy.
    - **Advantages**: Reduces data redundancy and data storage, easier to maintain data integrity.
    - **Disadvantages**: More complex to understand and navigate, potentially slower query performance due to more joins required between dimension tables, which can impact Power BI's query engine.
    - **Significance**: Used when data redundancy is a significant concern or when dimensions have very deep hierarchies that benefit from normalization, though generally less common than star schemas in typical Power BI implementations for performance reasons.

Primary Key vs. Foreign Key in Relationships (Power BI)

These terms are fundamental to establishing relationships between tables in a data model, which is crucial for Power BI to correctly filter and aggregate data across different tables.

- **Primary Key (PK)**:
    - **Definition**: A column or a set of columns in a table that uniquely identifies each row in that table. It ensures that every record in the table is unique.
    - **Characteristics**: Cannot contain NULL values and must contain unique values for each row.
    - **Role in Power BI**: In Power BI, the column designated as the 'one' side of a one-to-many relationship is typically the primary key (or a unique identifier) in that table.
    - **Example**: In a 'Products' table, 'ProductID' would be the primary key, uniquely identifying each product.
- **Foreign Key (FK)**:
    - **Definition**: A column or a set of columns in a table that refers to the primary key in another table. It establishes a link between two tables.
    - **Characteristics**: Can contain duplicate values and NULL values (depending on referential integrity rules).
    - **Role in Power BI**: The foreign key in one table links to the primary key in another table, forming the 'many' side of a one-to-many relationship. This link allows filtering and calculations to propagate between tables.
    - **Example**: In a 'Sales' table, 'ProductID' would be a foreign key, linking each sales transaction to the corresponding product details in the 'Products' table.

Why is Cardinality Important?

Cardinality in Power BI relationships describes the nature of the relationship between two tables, specifically how many instances of an entity in one table are related to instances of an entity in another table. It's crucial for several reasons:

1. **Correct Data Filtering and Aggregation**: Cardinality defines how filters applied to one table propagate to another. If cardinality is incorrectly defined, Power BI might produce incorrect aggregations or filters, leading to misleading insights.
    - **Example**: If you set a 'Many-to-Many' relationship where it should be 'One-to-Many', filters might not propagate as expected, or you might encounter ambiguous filter contexts.

2. **Performance Optimization**: Power BI's VertiPaq engine optimizes query performance based on the cardinality of relationships. A well-defined cardinality allows the engine to create efficient query plans. 'One-to-Many' relationships are generally the most performant.
3. **Ambiguity Resolution**: Incorrect cardinality can lead to ambiguous relationships, especially in complex models, causing Power BI to choose an arbitrary path for filtering or even resulting in errors. Correct cardinality helps in resolving these ambiguities.
4. **Enabling DAX Calculations**: Many DAX functions (like `RELATED`, `RELATEDTABLE`, `CALCULATE` with filter context modifications) rely heavily on correctly defined relationships and their cardinalities to retrieve values or propagate filters across tables.
5. **Model Clarity and Integrity**: Explicitly defining cardinality improves the clarity of your data model and helps maintain data integrity. It ensures that the relationships reflect the true nature of your business data.

Types of Cardinality in Power BI:

- *One-to-many (1:)**: The most common and recommended type. One row in the 'one' side table can relate to multiple rows in the 'many' side table.
- **Many-to-one (*:1)**: Functionally the same as one-to-many, just viewed from the opposite direction.
- **One-to-one (1:1)**: Less common. One row in one table relates to exactly one row in another table.
- **Many-to-many (*:*)**: Often requires a bridge table in between for best practice, as direct many-to-many relationships can sometimes be complex and impact performance or filtering behavior. While Power BI supports direct many-to-many, it's often more robust to model them as two one-to-many relationships with an intermediary table.

Question 4 : Differentiate between:

● Calculated column vs Measure Also, define Row context and Filter context with simple examples.

Answer :

Both Calculated Columns and Measures are formulas created using Data Analysis Expressions (DAX) in Power BI, but they serve different purposes and behave differently.

- **Calculated Column**:
    - **Definition**: A column that you add to an existing table in your data model by defining a DAX formula. The result of the formula is computed row by row and stored in the model, occupying memory.
    - **Calculation Time**: Calculated at data refresh time (when data is loaded or refreshed) or when the model is processed.
    - **Storage**: Stored physically in the model alongside other columns. This means it consumes RAM.
    - **Context**: Evaluated in **row context** (for each row) by default.
    - **Use Cases**: Best for calculations that need to be performed row-by-row and then used for filtering, slicing, or displaying on rows/columns in visuals. Examples include concatenating text columns, creating age groups from birthdates, or extracting parts of a date.

Example: Creating a `Full Name` column by concatenating `First Name` and `Last Name`.
Full Name = [First Name] & " " & [Last Name]

    -

- **Measure**:
    - **Definition**: A dynamic calculation whose result is not stored in the model but is computed on the fly when it's used in a visual, based on the **filter context** applied to that visual.
    - **Calculation Time**: Calculated at query time (when a visual requests data).
    - **Storage**: Not stored physically; consumes memory only when evaluated in a query.
    - **Context**: Evaluated in **filter context** by default, but can transition to row context using iterators.
    - **Use Cases**: Best for aggregations (sums, averages, counts, distinct counts) or complex calculations that change based on what's selected in a report. Examples include Total Sales, Average Order Value, Year-to-Date Sales, or Percentage of Grand Total.

    Example: Calculating `Total Sales`.
    Total Sales = SUM('Sales'[SalesAmount])

    -

Key Difference Summary: Calculated columns add new columns to your table that exist at a row level, while measures calculate aggregated values dynamically based on what's visible in your report.

Row Context and Filter Context

Understanding these two contexts is fundamental to writing correct and efficient DAX.

- **Row Context**:
    - **Definition**: The context in which a formula is evaluated one row at a time. When a DAX expression is evaluated in row context, it 'sees' the values of the current row for each column in the table.
    - **When it occurs**: Primarily used by calculated columns and by iterator functions (functions ending with 'X', like `SUMX`, `AVERAGEX`, `MAXX`, `COUNTX`, `FILTER`), which iterate through a table row by row.
    - **Simple Example**: If you have a `Sales` table with `Quantity` and `UnitPrice` columns, and you create a calculated column `LineTotal = [Quantity] * [UnitPrice]`. When Power BI calculates `LineTotal` for a specific row, it takes the `Quantity` and `UnitPrice` from *that particular row* to compute the `LineTotal` for *that same row*.
        - *Scenario*: In a table visual displaying each transaction, `LineTotal` shows the total for each individual transaction line.
- **Filter Context**:
    - **Definition**: The set of filters currently applied to the data model before a DAX expression is evaluated. These filters come from various sources: slicers, filters

on visuals, rows/columns of a matrix or table, filters passed from other visuals, or filters explicitly defined within DAX functions (like `CALCULATE`, `FILTER`).

- **When it occurs**: Every measure calculation in Power BI is evaluated within a filter context. This context determines the subset of data that the measure will operate on.
- **Simple Example**: Using the `Total Sales` measure: `Total Sales = SUM('Sales'[SalesAmount])`.
  - *Scenario 1*: If you drag `Total Sales` into a card visual without any other filters, it shows the sum of `SalesAmount` for the *entire* `Sales` table (the initial filter context).
  - *Scenario 2*: If you put `Total Sales` in a table visual with `Product Category` on rows, then for each `Product Category`, the `Total Sales` measure is evaluated within a filter context where only sales belonging to that specific product category are considered. For example, for the 'Electronics' row, the filter context is `Product Category = "Electronics"`, and the measure sums sales only for Electronics.
  - *Scenario 3*: If a slicer is applied to `Date = "2023"`, the `Total Sales` measure will sum sales only for the year 2023, as the filter context now includes `Date = "2023.`

Question 5: What is the difference between a report and a dashboard in Power BI?

Answer:

Power BI Report

- **Definition**: A multi-perspective view into a dataset, presenting various findings and insights from that dataset. A report can have multiple pages, each containing different visuals, tables, and text boxes.
- **Data Source**: Connected directly to one or more underlying datasets. All visuals on a report page are typically derived from the same data model.
- **Interactivity**: Highly interactive. Users can drill down, drill through, cross-filter, slice, and dice the data to explore it in depth. Reports are designed for detailed analysis and exploration.
- **Creation**: Created using Power BI Desktop, where you connect to data, model it, and design your visuals and pages. Reports are then published to the Power BI Service.
- **Features**: Supports many types of visuals, including tables, matrices, charts (bar, line, pie, etc.), maps, custom visuals, and more. Also supports DAX for complex calculations.
- **Purpose**: To provide in-depth analysis, answer specific business questions, and allow users to interact with and explore data.

Power BI Dashboard

- **Definition**: A single-page canvas that tells a story by consolidating a high-level overview of key metrics and visuals from one or more reports and even other dashboards. It's often referred to as a 'canvas' or 'pinboard'.
- **Data Source**: Visuals (tiles) on a dashboard are *pinned* from existing reports or Q&A results. This means a single dashboard can display data from multiple different underlying datasets and reports.
- **Interactivity**: Limited interactivity. Clicking on a tile on a dashboard usually takes you to the underlying report or Q&A that the tile came from. You cannot directly filter or drill down within the dashboard itself.
- **Creation**: Created in the Power BI Service by pinning visuals from one or more published reports.
- **Features**: Consists of 'tiles' which are snapshots of visuals from reports. Can also include images, videos, text boxes, and web content. Supports natural language Q&A for quick insights.
- **Purpose**: To provide a quick, at-a-glance view of key performance indicators (KPIs) and metrics, monitor real-time data, and serve as an entry point for further exploration into detailed reports.

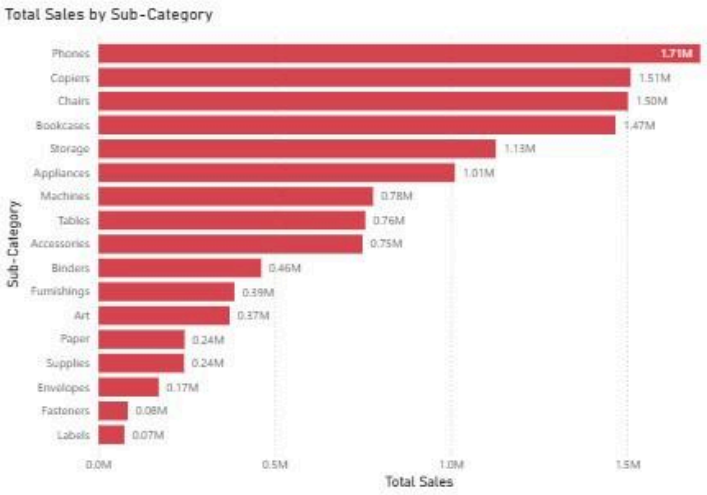
Question 6 : Using the Sample Superstore dataset:

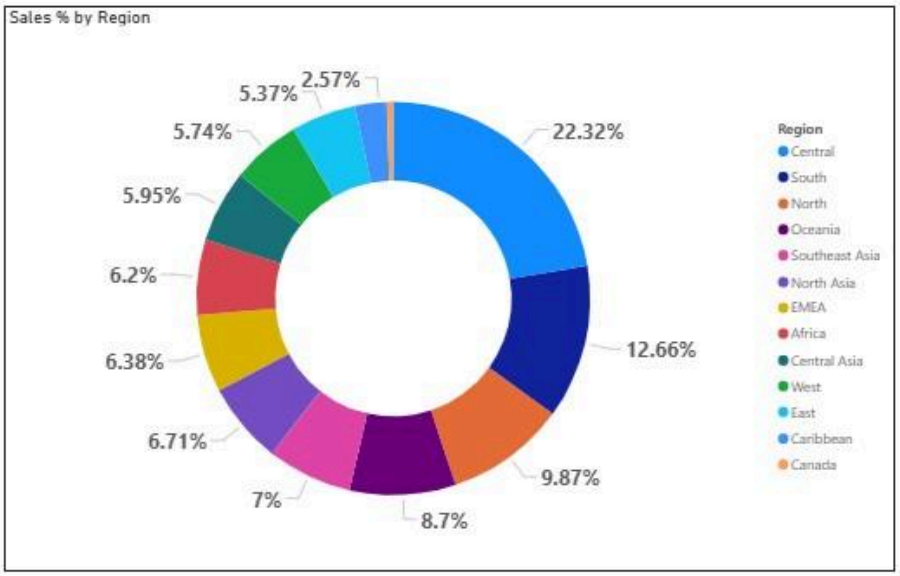
● Create a Clustered Bar Chart to display Total Sales by Sub-Category

● Create a Donut Chart for Sales % by Region Provide screenshots of both visuals.

DATASET LINK : Global_superstore2

Answer :





Question 7 : Write and apply the following measures:

● Total Profit = SUM([Profit])

● Average Discount = AVERAGE([Discount])

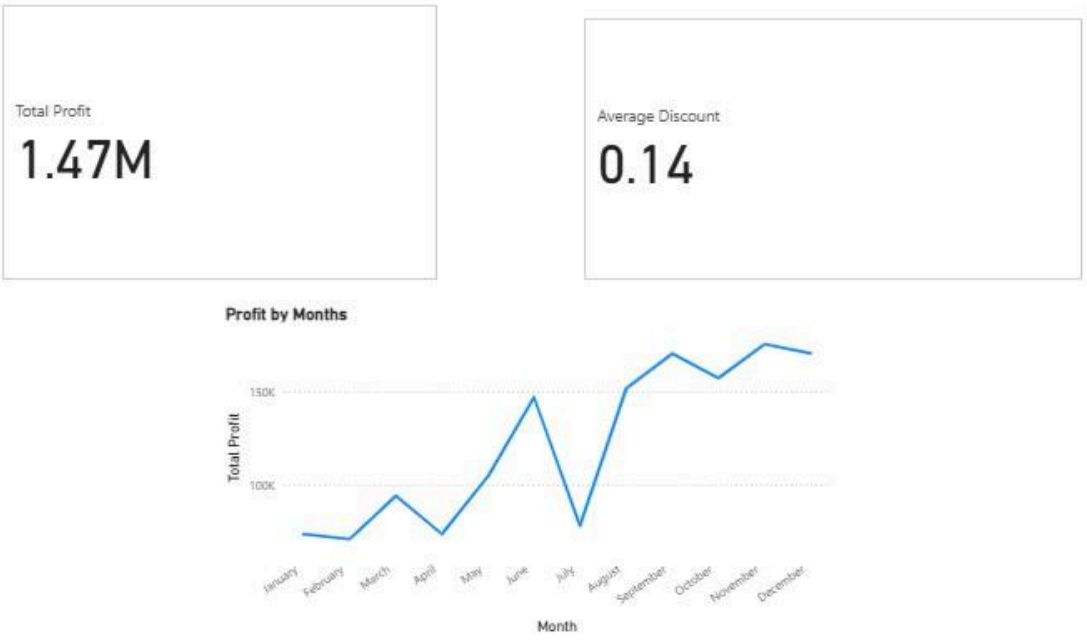
Display both in a KPI Card, and use a Line Chart to show profit trend over months. Add visuals and DAX formulas.

Answer :

DAX Measures Used:

Total Profit = SUM('Global_Superstore2'[Profit])

Average Discount = AVERAGE('Global_Superstore2'[Discount])



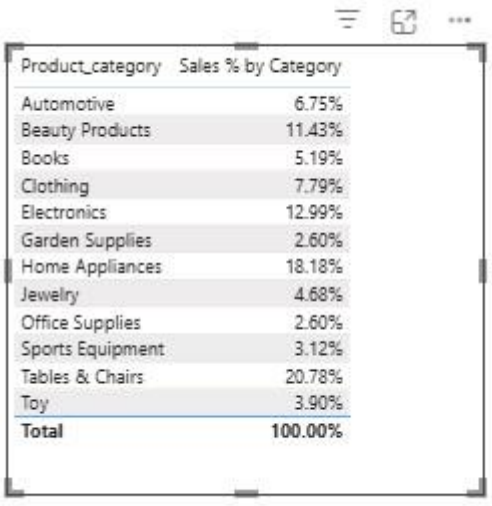
Question 8 : Implement a DAX measure that calculates the percentage of total sales by product category.

Answer:

DAX Measures Used:

Total_Sales_ = SUM('Sales'[Sales_Amount])

Sales % by Category = DIVIDE([Total_Sales_], CALCULATE([Total_Sales_], ALL('Sales')))



Product_category	Sales % by Category
Automotive	6.75%
Beauty Products	11.43%
Books	5.19%
Clothing	7.79%
Electronics	12.99%
Garden Supplies	2.60%
Home Appliances	18.18%
Jewelry	4.68%
Office Supplies	2.60%
Sports Equipment	3.12%
Tables & Chairs	20.78%
Toy	3.90%
Total	**100.00%**

Question 9 :

● Create a DAX Measure for Total Profit

● Use it in a Waterfall Chart to analyze how different Sub-Categories contribute to overall profit

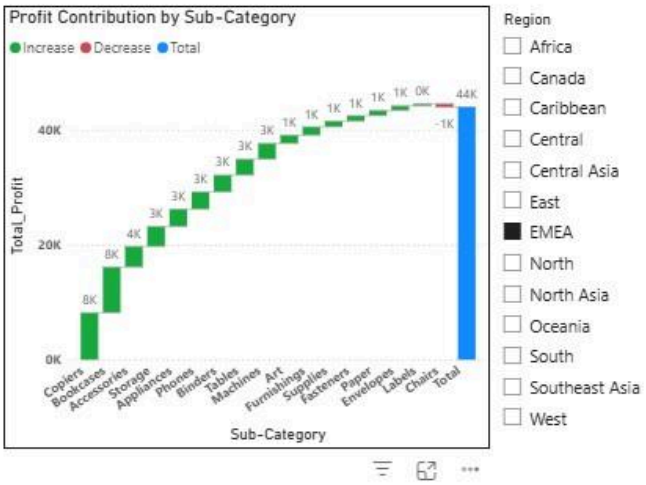
● Add a Slicer for Region to filter the visual

● Write brief business insights (4–5 lines) from the chart and provide 2–3 data-drivenrecommendations to improve profit.

Answer:

DAX Measures Used:

Total_Profit = SUM('Global_Superstore2'[Profit])



Brief Business Insights

- A few sub-categories act as major profit drivers, contributing positively to overall profit.
- Some sub-categories generate negative profit, which reduces overall profitability.
- Profit contribution varies significantly when filtered by Region, indicating regional performance differences.
- High sales volume does not always result in high profit margins.

Data-Driven Recommendations

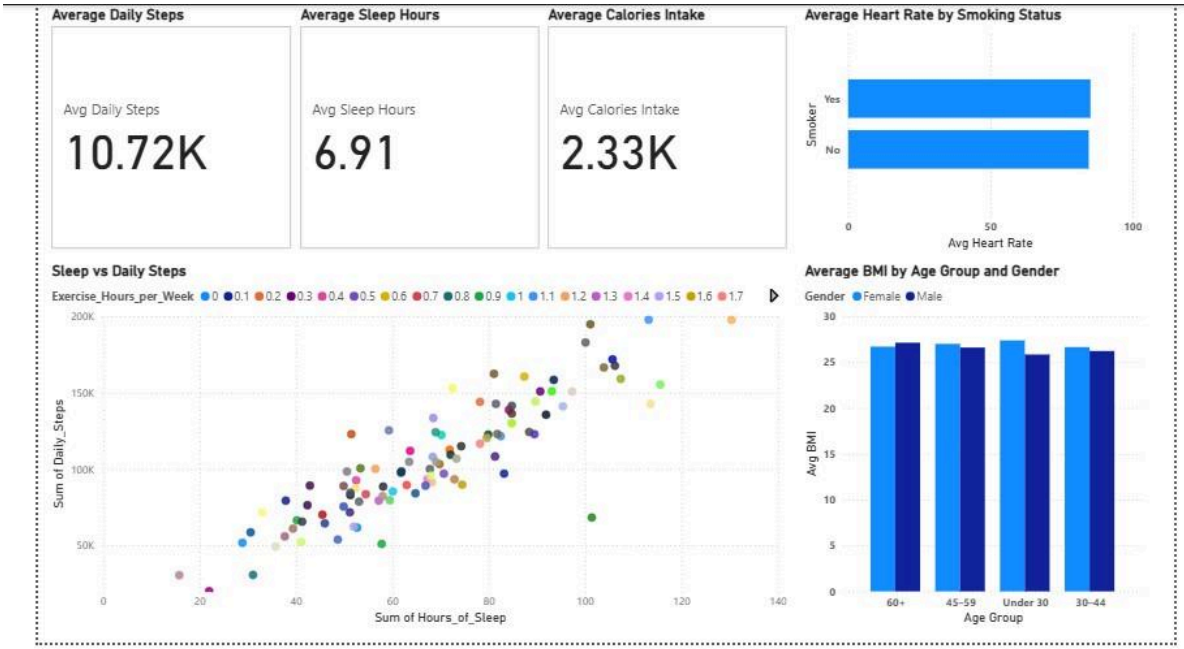
- Increase focus on high-profit sub-categories through targeted promotions and optimized inventory planning.
- Reassess pricing, discount strategies, and supplier costs for sub-categories generating losses.
- Implement region-specific strategies for sub-categories that show strong profitability in certain regions.

Question 10 : Using the provided dataset (includes Age, Gender, BMI, Steps, Calories, Sleep, Heart Rate, Blood Pressure, Smoking, Alcohol, Exercise, Diabetic & Heart Disease status):

Build a one-page Power BI dashboard that answers:

1. Are users maintaining a balanced lifestyle (Steps, Sleep, Calories)
2. What lifestyle patterns (Smoking, Alcohol, BMI, etc.) indicate heart disease risk?
3. Is there any visible relationship between Sleep and Physical Activity?
4. How does BMI vary across Age Groups and Genders?
5. What is the impact of smoking and alcohol on heart rate and blood pressure?

6. Segment people based on their health activity to suggest lifestyle changes



People can be segmented based on their health activity using indicators such as **daily steps, sleep hours, BMI, and exercise habits**.

- **Active & Healthy Group:** High daily steps, adequate sleep, normal BMI, and regular exercise.
  Suggestion: Maintain current lifestyle and encourage consistency.
- **Moderately Active Group:** Average steps and sleep with slightly elevated BMI.
  Suggestion: Increase physical activity and adopt balanced diet practices.
- **Sedentary / High-Risk Group:** Low steps, poor sleep, high BMI, and little or no exercise.
  Suggestion: Introduce gradual physical activity, improve sleep routines, and reduce unhealthy habits such as smoking or excessive alcohol intake.