ASSIGNMENT POWER BI

THEORETICAL QUESTIONS

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 1: Power BI is a business analytics solution and a collection of software services, apps, and connectors by Microsoft that transforms raw data from disparate sources into cohesive, visually immersive, and interactive insights and reports. It enables users to analyze data, find trends, spot patterns, and share findings to make data-driven decisions.

The key components of the Power BI ecosystem specified in the query are:

- **Power BI Desktop:** This is a free, downloadable Windows desktop application used for the core tasks of data analysis and report creation. Users can connect to various data sources, clean and transform data using the built-in Power Query editor, model the data (using Power Pivot functionality), and create compelling visualizations and reports on a design canvas. It is where most of the heavy lifting of report creation occurs before publishing.
- **Power BI Service (Power BI Online):** This is the cloud-based, Software as a Service (SaaS) part of the Power BI ecosystem. It is primarily used for publishing, sharing, collaborating on, and viewing reports and dashboards from anywhere via a web browser. It allows for features like setting up automatic data refreshes, natural language querying (Power Q&A), and data alerts, which are not available in the Desktop version.

- **Power BI Mobile:** These are native apps available for iOS and Android devices that enable users to securely access and interact with their reports and dashboards on the go, whether the data is stored in the cloud or on-premise. The apps ensure that business leaders and stakeholders can monitor key metrics and receive real-time data alerts from their mobile devices.
- **Power BI Gateway:** This tool acts as a bridge between the cloud-based Power BI Service and on-premises data sources (such as SQL Server databases or local files) that need to stay behind a corporate firewall. The Gateway ensures that data remains secure and allows for scheduled or manual data refreshes so that cloud reports stay up-to-date with local data without moving the entire database to the cloud.

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 2: <u>Pie Chart vs. Donut Chart</u>

Both are used to show the proportion of categories as a percentage of the total.

Feature	Pie Chart	Donut Chart
Shape	Solid circle divided into slices.	Circle with a blank center.

Focus	Emphasizes the area/size of each slice.	Emphasizes the length/arc of each segment, with the center often used to display the total value or a key metric.

When to prefer one over the other:

- **Pie Chart is preferred** when you want viewers to focus strictly on the *relative size* of each slice.
- **Donut Chart is preferred** for better data legibility (human eyes are better at comparing arc lengths than areas) and when you need to display additional information (like the overall total percentage) in the central blank space [3].

Example:

- You might use a **pie chart** in a presentation to quickly show the market share of three major competitors to a general audience.
- You might use a **donut chart** on a dashboard to show a progress tracker, where the center of the chart clearly displays the current completion percentage.

Bar Chart vs. Column Chart

Both are used to compare categorical data using rectangular bars. The primary difference is their orientation.

Feature	**Bar Chart**	**Column Chart**

Orientation	Horizontal bars.	Vertical bars.
Axis	Categories on the vertical axis (Y-axis), values on the horizontal axis (X-axis).	Categories on the horizontal axis (X-axis), values on the vertical axis (Y-axis).

When to prefer one over the other:

- **Bar Chart is preferred** when category labels are long and descriptive, as horizontal space allows for easier reading without rotation or truncation [2]. It is also better for displaying negative values or a wide range of values.
- **Column Chart is preferred** for displaying data over a time series (e.g., monthly sales), as the flow of time is conventionally represented horizontally from left to right [2].

Example:

- You might use a **bar chart** to compare the population of the top 20 most populous cities in the world, where the city names are long labels.
- You might use a **column chart** to show monthly website traffic for a year, as the progression from January to December is clearly shown on the X-axis.

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 3: Star Schema vs Snowflake Schema

These are two common data warehousing schema styles used to organize data for analytical purposes.

Feature	Star Schema	Snowflake Schema
Structure	A central fact table surrounded by de-normalized dimension tables.	A central fact table connected to normalized dimension tables, which may further link to sub-dimension tables.
Normalization	Low normalization (de-normalized dimensions).	High normalization (dimensions are split into multiple related tables).
Performance	Generally faster query performance due to fewer joins required between tables.	Slower query performance than star schema due to the need for more complex joins across many dimension tables.

Disk Space	Uses more storage space due to data redundancy in the larger dimension tables.	Uses less storage space due to efficient storage of normalized data.
Complexity	Simpler design; easier for business users to understand and navigate.	More complex design and maintenance.

Significance: The star schema is typically preferred in Power BI and data warehousing for its simplicity and performance, making data retrieval for reports fast and efficient. The snowflake schema is used when data integrity and storage efficiency are paramount, at the cost of slightly increased complexity and query time.

Primary Key vs Foreign Key in Relationships (Power BI)

Keys are essential for defining how tables relate to one another, forming the backbone of a data model.

- Primary Key (PK): A column or set of columns in a table that uniquely identifies each row (record) in that table.
    - Significance: It ensures data integrity by guaranteeing that every record is unique and can be accurately referenced. In Power BI modeling, the primary key usually resides on the "one" side of a one-to-many relationship (e.g., in a DimProduct table, a ProductID is unique for each product).
- Foreign Key (FK): A column in one table that refers to the primary key of another table.

- Significance: The foreign key creates the link (relationship) between two tables. It enforces referential integrity, ensuring that a value in the foreign key column corresponds to an existing value in the primary key table. In Power BI, foreign keys reside on the "many" sides of a relationship (e.g., the ProductID in a FactSales table links back to the primary key ProductID in the DimProduct table).

Significance in Power BI: Correctly identifying and defining these keys is crucial for creating accurate and functional relationships, allowing data to be filtered and aggregated correctly across different tables.

Why Cardinality is Important

Cardinality defines the numerical nature of the relationship between two tables (e.g., one-to-one, one-to-many, many-to-many).

Its importance stems from several factors:

- Relationship Direction and Filtering: Cardinality determines how filters propagate through the data model. In a one-to-many relationship (the most common type), a filter applied to the "one" side (e.g., filtering for a specific date in a Dim Date table) automatically filters the data on the "many" sides (e.g., the Fact Sales table).
- Performance Optimization: Correct cardinality helps the Power BI engine optimize query paths and utilize resources efficiently. The engine handles different cardinalities in specific ways; for instance, many-to-many relationships require special handling (often via bridge tables) to avoid performance issues or ambiguity.
- Data Accuracy: Setting the correct cardinality prevents ambiguity and ensures that aggregations and calculations return accurate results. An incorrectly defined cardinality can lead to incorrect data filtering or double-counting of measures.

- Model Validation: It serves as a data quality check. If you expect a one-to-many relationship but the model detects many-to-many, it indicates potential duplicates in what you intended to be the unique key column.

Question 4: Differentiate between:

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

Answer 4: <u>Calculated Column vs. Measure</u>

Feature	Calculated Column	Measure
Calculation Time	Calculated at the time of data refresh or loading into the data model.	Calculated on-the-fly when added to a visual (e.g., a table, chart, or slicer).
Storage	Stored physically in every row of the table, increasing the size of the data model.	Not stored physically in the model; it only exists when referenced in a visualization.
Context Used	Primarily operates within **Row context**.	Primarily operates within **Filter context**.
Usage	Best for static values needed row-by-row, such as a full name	Best for aggregations (sums, averages, counts) that change

	(combining first and last) or an age group category.	based on user interactions and filters in reports.
Example Formula	Full Name = [FirstName] & " " & [LastName]	Total Sales = SUM ([Sales Amount])

Context Definitions and Examples

Row Context

Row context refers to the calculation being evaluated one row at a time within a table. It is the default context for calculated columns.

- **Simple Example:**
  Imagine you have a sales table with columns for Quantity and Unit Price. You want to calculate the Line Total for each individual transaction.
  - The Line Total is calculated by looking only at the Quantity and Unit Price of the *current* row being processed.
  - The DAX formula would be Line Total = [Quantity] * [Unit Price].
  - This is a perfect use case for a **calculated column** because the result is a static value tied to that specific row.

Filter Context

Filter context refers to the set of filters applied to the data model before a calculation is performed. These filters come from slicers, report filters, interactions between visuals, or specific functions within the DAX expression itself (like CALCULATE).

- **Simple Example:**
  You have a sales report and want to display the Total Sales for each Region.
  - When a user clicks on the "North" region in a slicer, the filter context changes.
  - The measure Total Sales = SUM ([Sales Amount]) automatically recalculates to sum *only* the sales where the region is "North".
  - If the user then filters by a specific year and product category, the filter context accumulates those filters, and the measure recalculates again based on the intersection of all applied filters. This is a perfect use case for a measure.

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

Answer 5: In Power BI, a dashboard is a single-page, high-level overview of KPIs (Key Performance Indicators), often combining visuals from multiple reports/datasets for quick monitoring (like a snapshot), while a report is a multi-page, detailed, interactive environment for deep data exploration, built from a single dataset with extensive filtering and drill-down capabilities, acting as the 'application' for analysis. Dashboards offer alerts and are service-only; reports provide in-depth insights and are built in Desktop, acting as the source for dashboard tiles.

Power BI Report

- Purpose: Detailed analysis, answering "why" something happened, discovering trends.

- Pages: Multi-page, allowing for complex data stories.
- Data Source: Based on a single dataset.
- Interactivity: Highly interactive (slicers, filters, drill-down, cross-filtering).
- Creation: Built in Power BI Desktop.
- Features: Allows viewing underlying data, supports bookmarks, subscriptions.

Power BI Dashboard

- Purpose: High-level monitoring, quick performance checks, at-a-glance insights.
- Pages: Single-page canvas (like a Windows desktop).
- Data Source: Can pull visuals (tiles) from multiple reports and datasets.
- Interactivity: Less interactive (static view, limited filtering/cross-filtering within dashboard).
- Creation: Created in the Power BI Service (web).
- Features: Supports email alerts, can be marked as 'featured', acts as a launching pad to reports, no underlying data view.

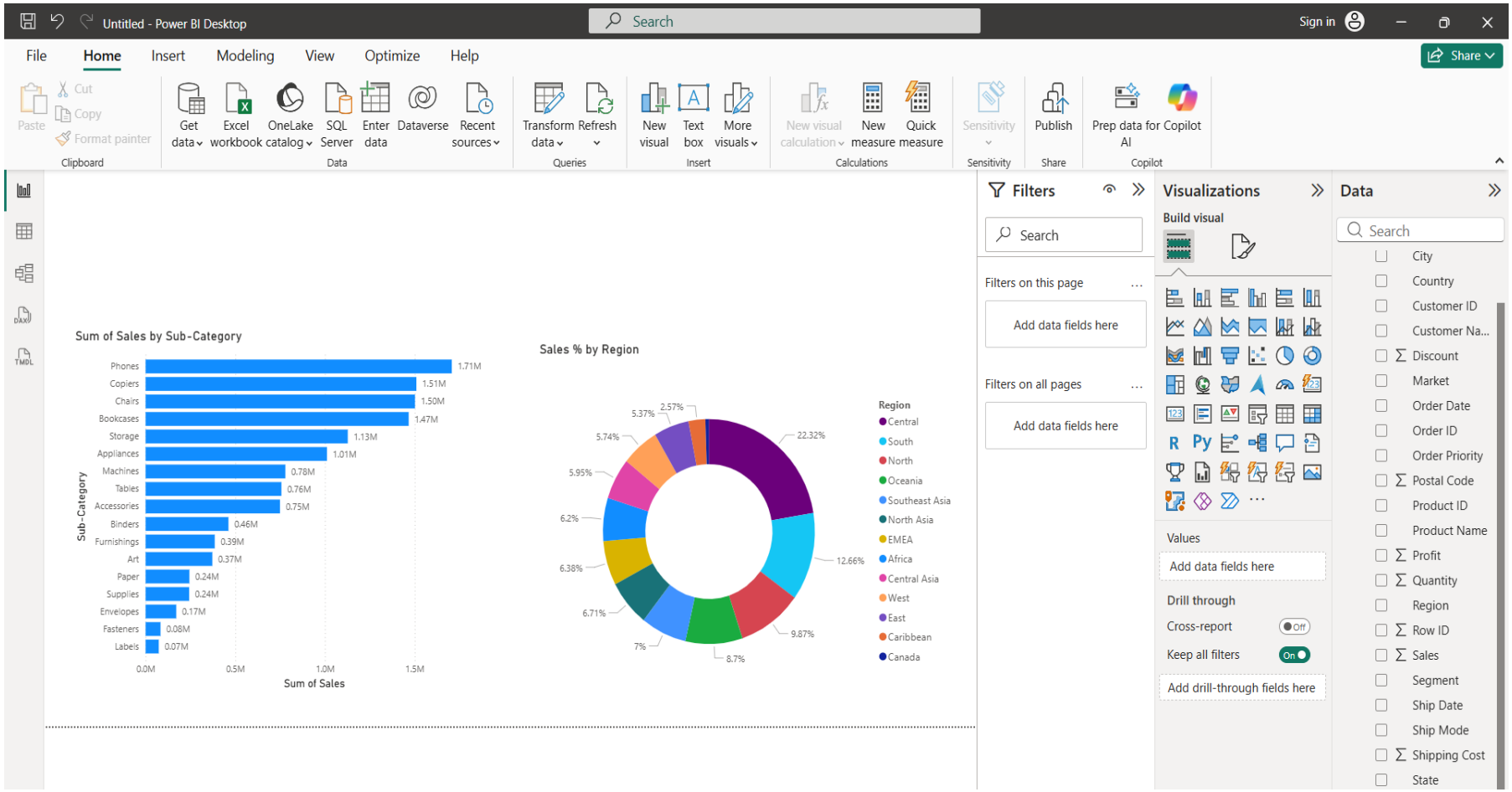
PRACTICAL QUESTIONS

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

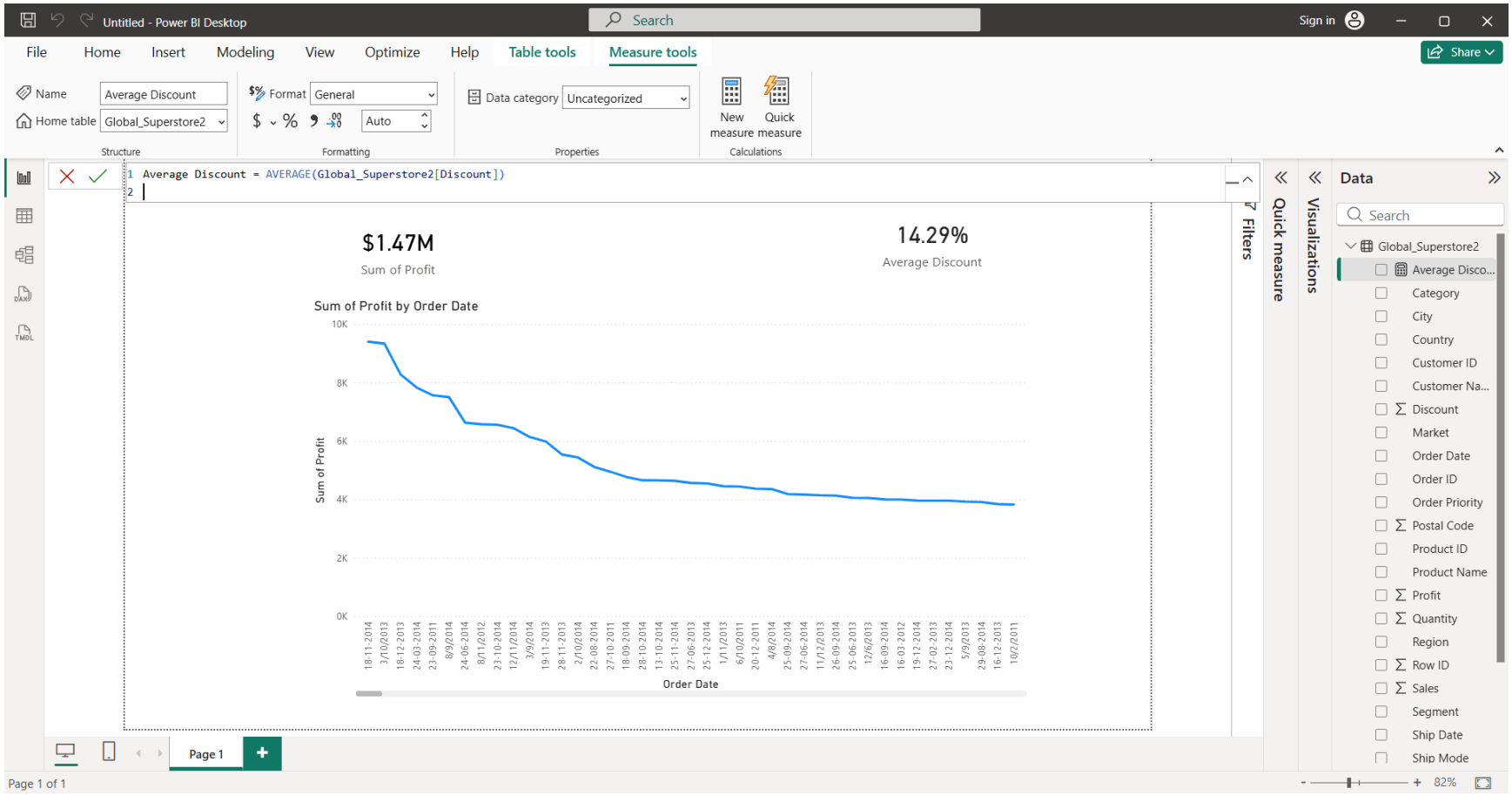
Provide screenshots of both visuals.



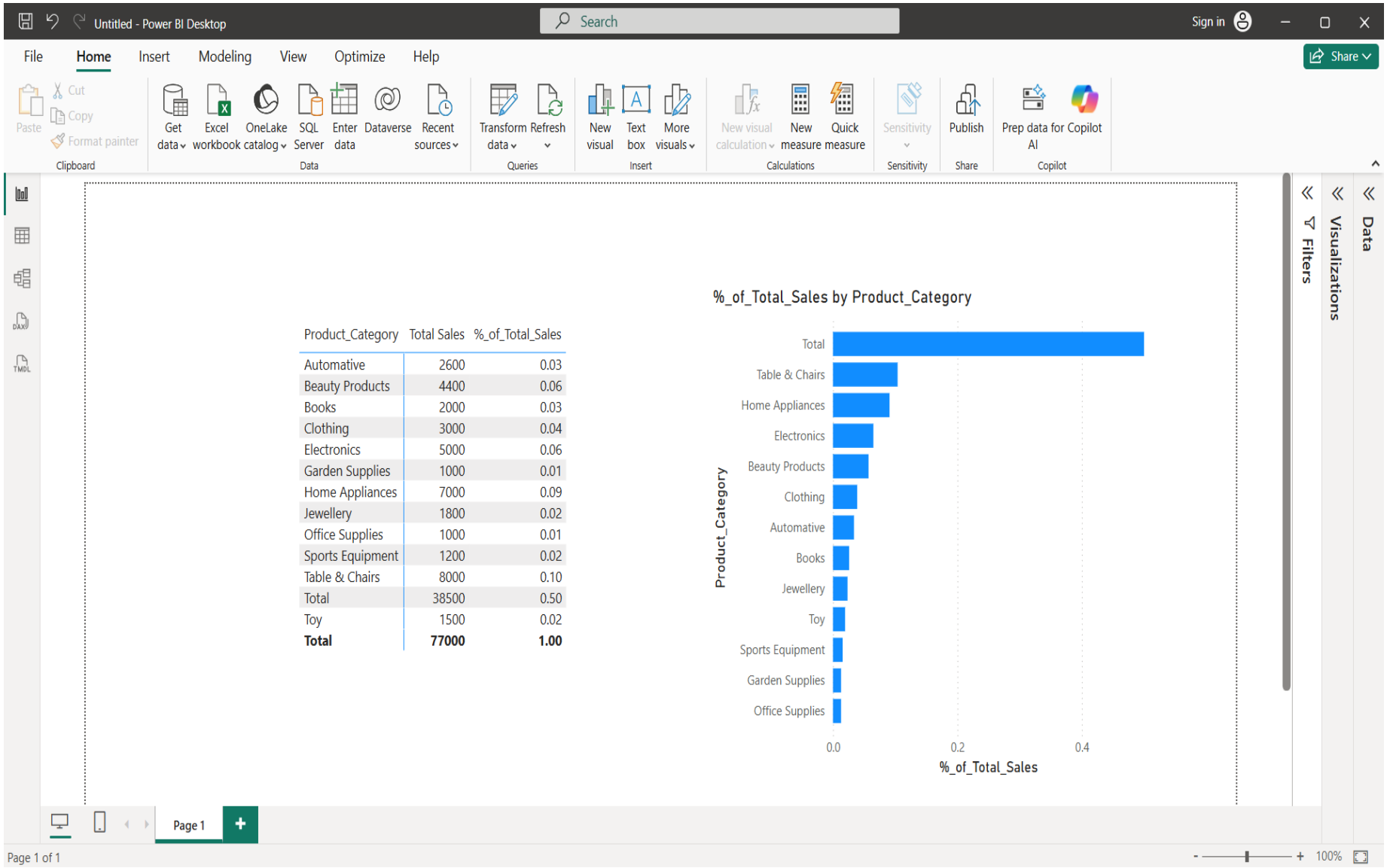
Question 7: Write and apply the following measures:

● 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.



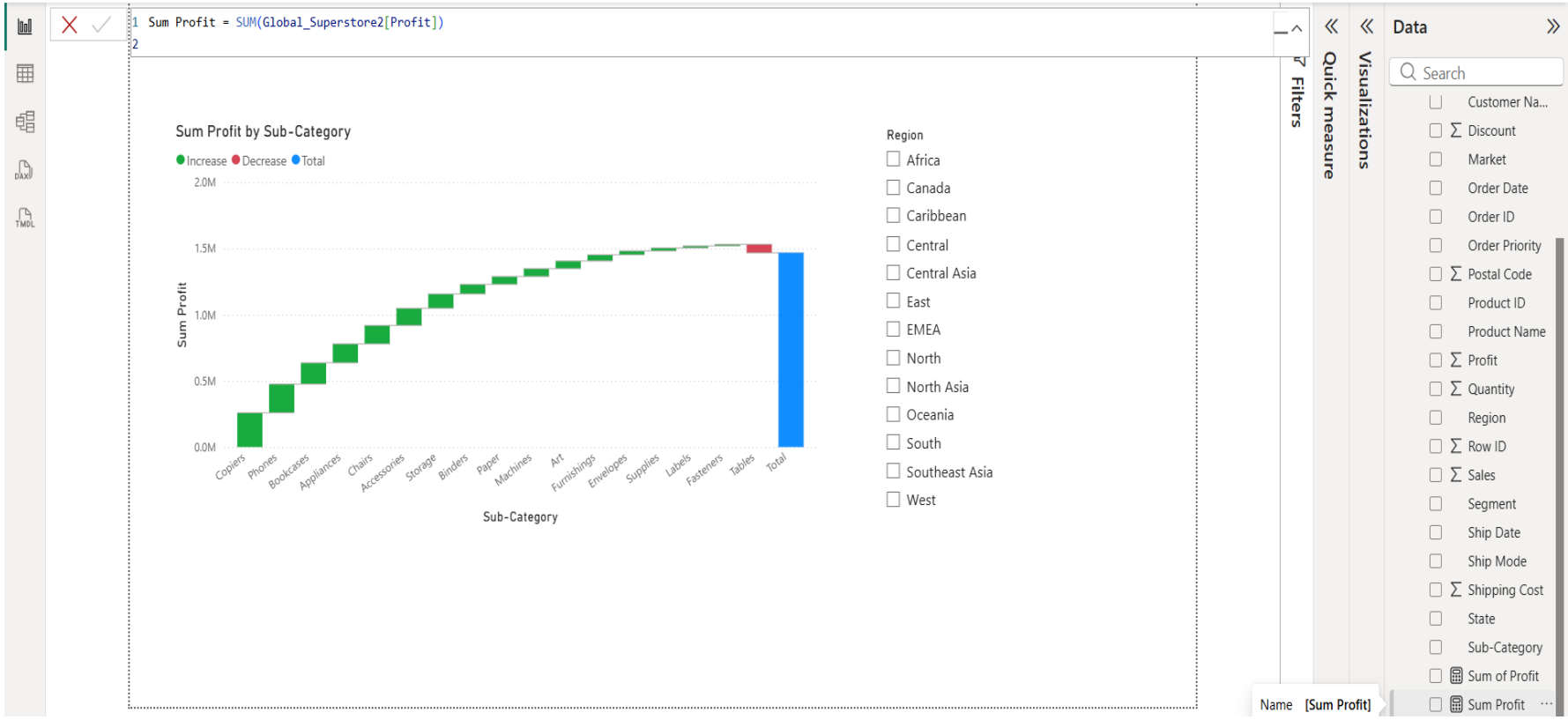
Question 8: Implement a DAX measure that calculates the percentage of total sales by product category.



Product_Category	Total Sales	%_of_Total_Sales
Automative	2600	0.03
Beauty Products	4400	0.06
Books	2000	0.03
Clothing	3000	0.04
Electronics	5000	0.06
Garden Supplies	1000	0.01
Home Appliances	7000	0.09
Jewellery	1800	0.02
Office Supplies	1000	0.01
Sports Equipment	1200	0.02
Table & Chairs	8000	0.10
Total	38500	0.50
Toy	1500	0.02
Total	**77000**	**1.00**

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-driven recommendations to improve profit.



Business Insights

- Copiers and Phones contribute the highest positive profit, making them key revenue drivers.
- Tables show a significant negative contribution, reducing overall profitability.
- Mid-range categories like Accessories and Storage provide steady profit support.
- Profit distribution varies noticeably when filtered by region, indicating regional demand differences.
- A few loss-making sub-categories offset gains from high-performing products.

Data-Driven Recommendations

- Optimize or discontinue loss-making sub-categories like Tables through cost control or price revision.
- Increase marketing focus on high-profit products (Copiers, Phones) in high-performing regions.
- Apply region-specific strategies, stocking profitable sub-categories based on regional demand patterns.

Question 10: Scenario:

VitaTrack Wellness, a digital health company in FitZone, has collected data on users' daily habits and health vitals. The analytics team is tasked with drawing actionable insights from this data to improve lifestyle suggestions and prevent heart-related risks.

Your Task: 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)

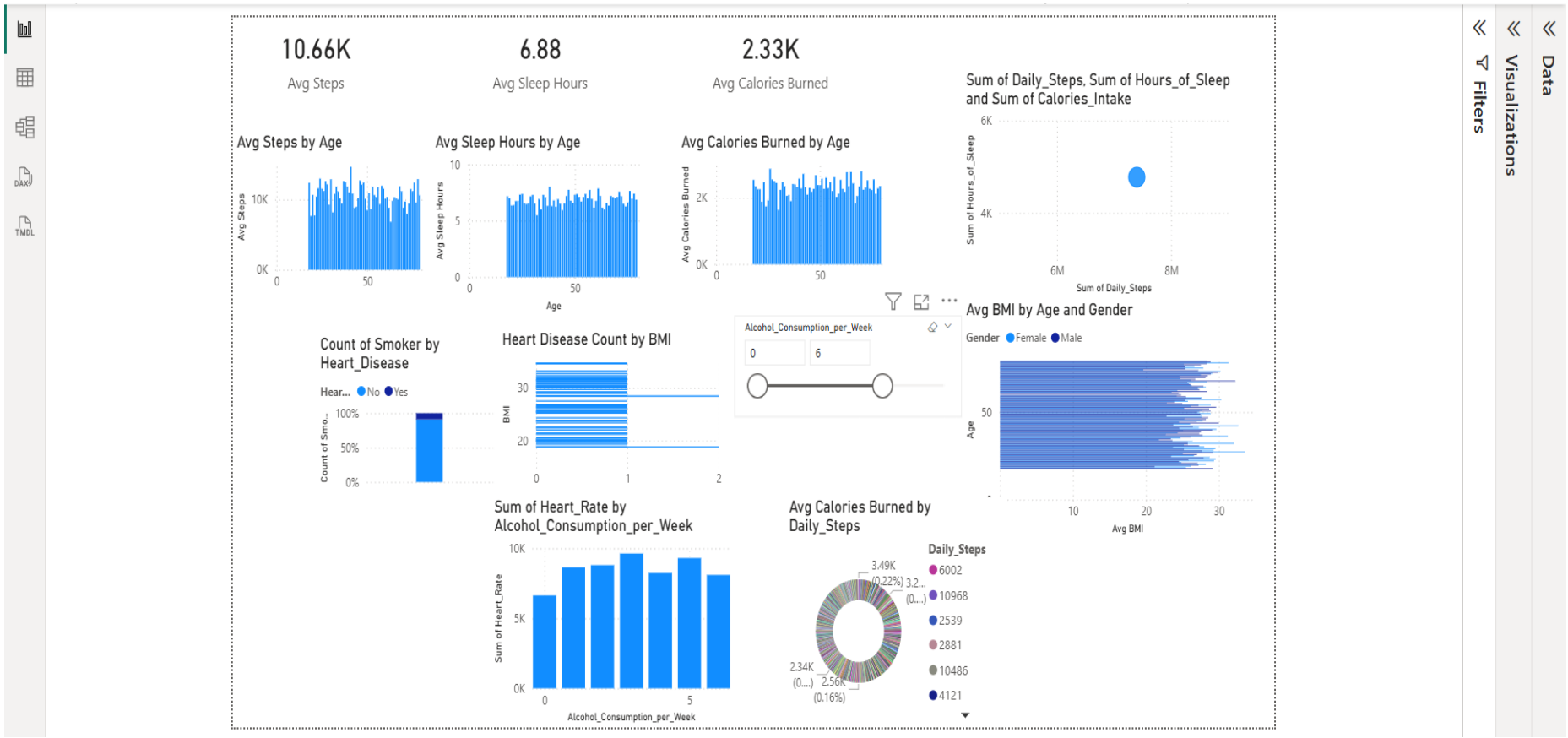
2. What lifestyle patterns (Smoking, Alcohol, BMI, etc.) indicate heart disease risk?

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.



INSIGHTS

(A) Users meeting 7–8 hrs. sleep and 7k–10k steps indicate balanced lifestyle; deviations highlight improvement areas.

(B) Higher BMI, smoking, and alcohol intake strongly correlate with heart disease incidence.

(C) Active users tend to sleep better. Low activity + poor sleep clusters show higher heart risk.

(D) BMI increases with age, especially among males in middle age—early intervention needed.

(E) Smokers and alcohol consumers show elevated heart rate and blood pressure levels.