**1. Approach**

**Data Cleaning and Preparation**

The data cleaning and preparation phase ensures that the dataset is structured and formatted correctly for further analysis. It involves the following steps:

**Data Loading**

1. **Insert Data into SQL Database:**
   * Create SQL tables for each data source (e.g., customers, sales, products).
   * Use CREATE TABLE and INSERT INTO statements to load the cleaned data.

Example to create and insert into a customer table:  
MySQL :

mycursor.execute("""

CREATE TABLE newcustomers (

CustomerKey INT,

Gender TEXT,

Name TEXT,

City TEXT,

`State Code` VARCHAR(255),

State TEXT,

`Zip Code` TEXT,

Country TEXT,

Continent VARCHAR(255),

Birthday DATE

)

""")

INSERT INTO newcustomers (CustomerKey, Gender, Name, City, `State Code`, State, `Zip Code`, Country, Continent, Birthday)

VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s, %s)

**Power BI Visualization**

1. **Connect SQL to Power BI:**
   * Connect the SQL database to Power BI.
   * Import the cleaned data from the SQL tables into Power BI.
2. **Create Interactive Dashboards:**
   * Develop interactive dashboards for visualizing insights from the data.
   * Use charts like bar graphs, line charts, and heat maps for better understanding.

**2. SQL Queries for Key Insights**

Using SQL queries, we extract valuable insights from the dataset, which play a crucial role in analyzing raw data and transforming it into meaningful visualizations. Below, we outline the types of insights derived from the data, along with example queries.

**Customer Analysis**

**Demographic Distribution:**

* + Analyze the customers by gender, age (calculated from birthdate), and location (city, state, country, continent).

SELECT

YEAR(now()) - YEAR(`BirthDay`) AS CustomerAge,

COUNT(CustomerKey) AS CustomerCount

FROM newcustomers

GROUP BY CustomerAge

ORDER BY CustomerAge ASC;

**Purchase Patterns:**

Calculate the total number of products, grouped by product key, on a yearly basis.

SELECT YEAR(`Order Date`) AS OrderYear, ProductKey, COUNT(Quantity) AS ProductCount

FROM sales

GROUP BY OrderYear, ProductKey;

**Customer Segmentation:**

Segment customers based on demographics and purchasing behavior and determine the total number of customers over time.

SELECT YEAR(`Order Date`) AS OrderYear, COUNT(CustomerKey) AS CustomerCount

FROM sales

GROUP BY OrderYear

ORDER BY OrderYear DESC

**Sales Analysis**

**Sales by Product:**

Sales by Currency: Examine how different currencies impact sales figures, considering exchange rates.

select year(Date) as year\_,Currency,avg(Exchange) as sum\_exchange from exchange\_rates group by Currency,year\_

**Store:**

New stores opened date based on different countries.

SELECT

Country,

YEAR(`Open Date`) AS OpenYear,

COUNT(StoreKey) AS TotalStoresOpened

FROM stores

GROUP BY Country, OpenYear

ORDER BY Country ASC, OpenYear ASC;

New stores opened date based on Yearly wise.

SELECT

Country,

YEAR(`Open Date`) AS OpenYear,

COUNT(StoreKey) AS TotalStoresOpened

FROM stores

GROUP BY Country, OpenYear

ORDER BY Country ASC, OpenYear ASC;

**Product Analysis**

**Product Popularity:**

Identify the most and least popular products based on sales data.

(SELECT s.ProductKey, COUNT(s.Quantity) AS total\_sales, p.`Product Name`

FROM sales AS s

JOIN products AS p ON p.ProductKey = s.ProductKey

GROUP BY s.ProductKey, p.`Product Name`

ORDER BY total\_sales DESC

LIMIT 1)

union all

(SELECT s.ProductKey, COUNT(s.Quantity) AS total\_sales, p.`Product Name`

FROM sales AS s

JOIN products AS p ON p.ProductKey = s.ProductKey

GROUP BY s.ProductKey, p.`Product Name`

ORDER BY total\_sales ASC

LIMIT 1)

**Profitability Analysis:**

Calculate profit margins for products by comparing unit cost and unit price.

(SELECT s.ProductKey, COUNT(s.Quantity) AS total\_sales, p.`Product Name`, p.`Unit Cost USD`,

p.`Unit Price USD`,(p.`Unit Price USD`-p.`Unit Cost USD`) as profit

FROM sales AS s

JOIN products AS p ON p.ProductKey = s.ProductKey

GROUP BY s.ProductKey, p.`Product Name`, p.`Unit Cost USD`,p.`Unit Price USD`, profit

ORDER BY total\_sales DESC

LIMIT 10)

**3. Analysis Steps**

**Customer Analysis**

1. **Gender:**
   * Percentage breakdown: e.g., Male (55%), Female (45%).
   * Insights: Target marketing strategies based on dominant gender groups.
2. **Age:**
   * Age groups distribution: e.g., 19-30, 31-40, 41–50, 51-60, 61-70, 71-80, 80+...
   * Insights: Focused campaigns for the largest age group.
3. **Customers count by Order year:**
   * 2016: 6,905 customers — Initial base, with limited geographic reach.
   * 2017: 7,942 customers (+15.02%) — Growth driven by online marketing and promotional campaigns.
   * 2018: 14,188 customers (+78.63%) — Expansion into new regions significantly increased reach.
   * 2019: 21,611 customers (+52.31%) — Introduction of new product categories attracted diverse customer segments.
   * 2020: 8,632 customers (-60.05%) — Decline attributed to external factors or shifts in market demand.

**Sales Analysis**

1. **Overall Sales Performance**: Track sales over time to identify seasonal patterns and long-term trends.
2. **Sales by Product**: Identify top-performing products and optimize inventory and marketing efforts.
3. **Sales by Store**: Understand regional differences in sales performance and manage resources accordingly.
4. **Sales by Currency**: Analyze how currency fluctuations impact overall sales and identify strategies for international markets.

**Product Analysis**

1. **Product Popularity**: Identify the best-selling and least popular products for inventory and marketing strategy.
2. **Profitability Analysis**: Calculate profit margins for each product to identify high-margin items.
3. **Category Analysis**: Evaluate which product categories are performing best to optimize product offerings.

**Store Analysis**

1. **Store Performance**: Measure sales performance in relation to store size and operational metrics.
2. **Geographical Analysis**: Analyze regional sales performance to guide expansion or closing decisions.