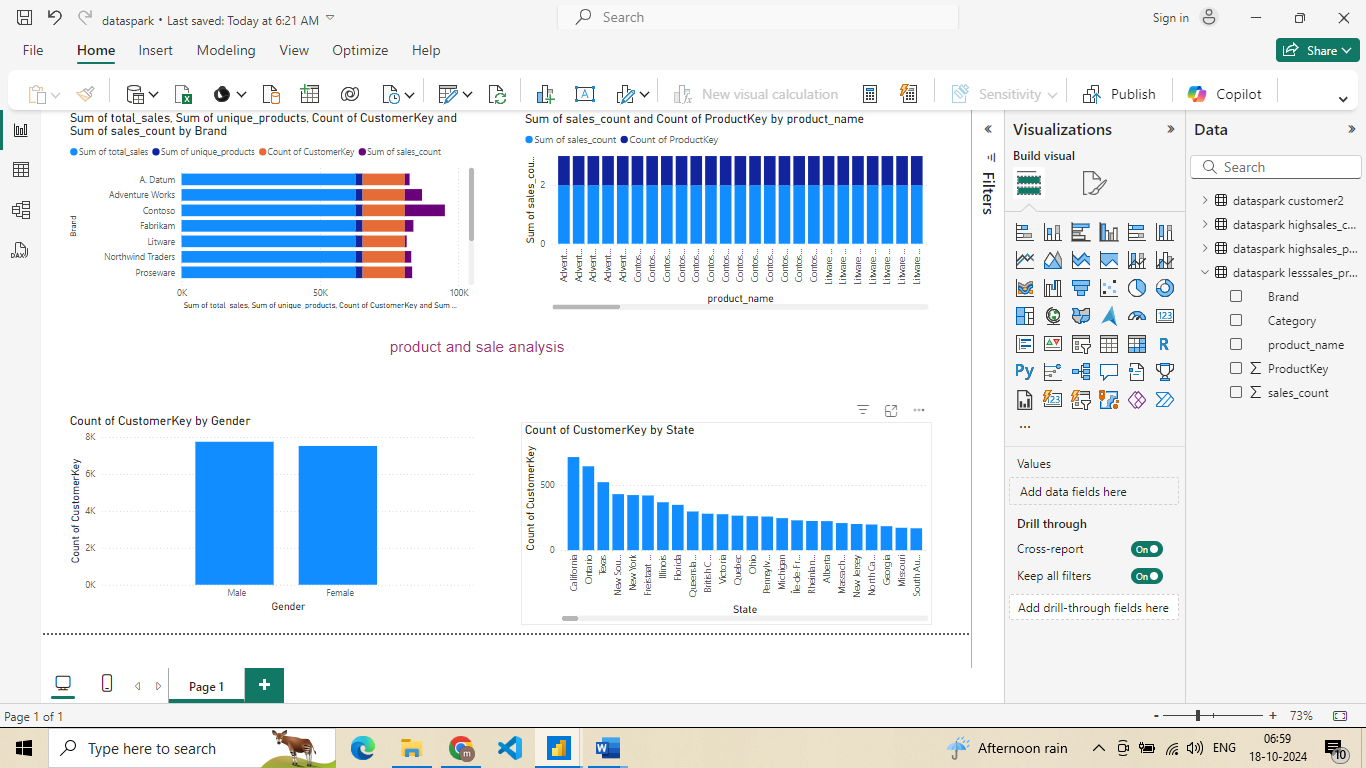
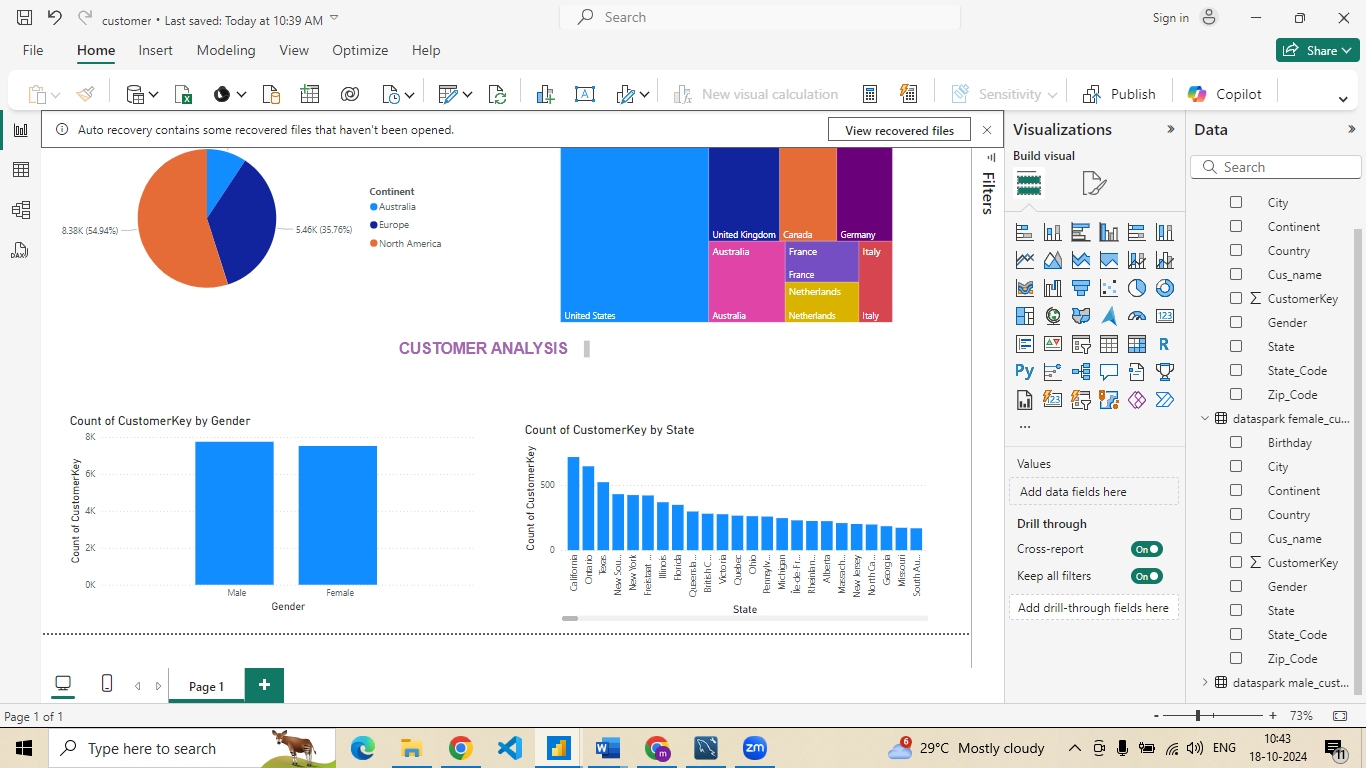
* **project overview:**

Global Electronics' data analytics team, you are tasked with conducting a comprehensive Exploratory Data Analysis (EDA) to uncover valuable insights from the company’s data. Your goal is to provide actionable recommendations that can enhance customer satisfaction, optimize operations, and drive overall business growth.

* Skills need:
* Data Cleaning and Preprocessing, EDA,Python, Data Management using SQL, Power Bi/Tableau

|  |  |
| --- | --- |
| * **Project Title** | * **DataSpark: Illuminating Insights for Global Electronics** |
| * **Skills take away From This Project** | * **Data Cleaning and Preprocessing, EDA,Python, Data Management using SQL, Power Bi/Tableau** |
| * **Domain** | * **Retail Analytics in the Electronics Industry** |

* **Business Use Cases:**
* By analyzing Global Electronics' customer, product, sales, and store data, we aim to identify key insights that will enhance marketing strategies, optimize inventory management, and improve sales forecasting. This will help tailor marketing campaigns, develop better products, plan effective promotions, and decide on store expansions and optimizations. Additionally, understanding the impact of currency exchange rates on sales will allow for better international pricing strategies. Overall, these insights will help Global Electronics increase customer satisfaction and drive business growth.
* **Approach:**
* **Data Cleaning and Preparation**
* Check for missing values and handle them appropriately.
* Convert data types where necessary (e.g., dates, numerical values).
* Merge datasets where necessary for analysis (e.g., linking sales data with product and customer data).
* **Load Data**
* Insert the preprocessed data into an SQL database by creating relevant tables for each data source and using SQL INSERT statements to load the data.
* **Power BI Visualization**
* Connect SQL to Power BI/Tableau, import the data, and create interactive dashboards.
* **Develop 10 SQL Queries**
* Formulate and execute 10 SQL queries to extract key insights from the data. These queries should address important business questions and support the analysis steps below.
* **Analysis Steps:**
* **Customer Analysis**
* Demographic Distribution: Analyze the distribution of customers based on gender, age (calculated from birthday), location (city, state, country, continent).
  + 1.Find male and female customers
  + SELECT
  + Gender,
  + COUNT(\*) AS TotalCustomers
  + FROM
  + customer2
  + GROUP BY
* Gender;
  + o/p=
  + Female 7514
* Male 7742
* 2.Total unitprice\_usd
* SELECT
* SUM(Unit\_Price\_USD) AS TotalUnitPrice
* FROM
* products2;
* o/p
* TotalUnitPrice;;;
* '898141.4399999968'
* 3.AverageUnitPrice
* SELECT
* AVG(Unit\_Price\_USD) AS AverageUnitPrice
* FROM
* products2;
* o/p;;;
* 356.83013110846116
* **Product Analysis**
* 1….Product Popularity: Identify the most and least popular products based on sales data.
* CREATE TABLE lesssales\_products AS
* SELECT
* p.ProductKey,
* p.product\_name,
* p.Brand,
* p.Category,
* COUNT(s.ProductKey) AS sales\_count
* FROM
* dataspark.products2 AS p
* LEFT JOIN
* dataspark.sales2 AS s ON p.ProductKey = s.ProductKey
* GROUP BY
* p.ProductKey, p.product\_name, p.Brand, p.Category
* HAVING
* sales\_count < 3;
* 2..Profitability Analysis: Calculate profit margins for products by comparing unit cost and unit price.
* 3…Category Analysis: Analyze sales performance across different product categories and subcategories.
* CREATE TABLE highsales\_category\_analysis AS
* SELECT
* p.Category,
* p.Subcategory,
* COUNT(s.ProductKey) AS total\_sales,
* COUNT(DISTINCT p.ProductKey) AS unique\_products
* FROM
* dataspark.products2 AS p
* LEFT JOIN
* dataspark.sales2 AS s ON p.ProductKey = s.ProductKey
* GROUP BY
* p.Category, p.Subcategory
* HAVING
* total\_sales >= 10;
* **Power BI Visualization**
* 
* 
* **Develop 10 SQL Queries**

Develop 10 SQL Queries

1.Find male and female customers count

SELECT

Gender,

COUNT(\*) AS TotalCustomers

FROM

customer2

GROUP BY

Gender;

SELECT

SUM(CASE WHEN Gender = 'Male' THEN 1 ELSE 0 END) AS MaleCount,

SUM(CASE WHEN Gender = 'Female' THEN 1 ELSE 0 END) AS FemaleCount

FROM

customer2;

|  |  |
| --- | --- |
|  |  |

2..high productsales

CREATE TABLE highsales\_products AS

SELECT

p.ProductKey,

p.product\_name,

p.Brand,

p.Category,

COUNT(s.ProductKey) AS sales\_count

FROM

dataspark.products2 AS p

LEFT JOIN

dataspark.sales2 AS s ON p.ProductKey = s.ProductKey

GROUP BY

p.ProductKey, p.product\_name, p.Brand, p.Category

HAVING

sales\_count >= 10;

3.CREATE TABLE female\_customers AS

SELECT

CustomerKey,

Gender,

Cus\_name,

City,

State\_Code,

State,

Zip\_Code,

Country,

Continent,

Birthday

FROM

dataspark.customer2

WHERE

Gender = 'Female'; -- Filter for female customers

4.find less product sales

CREATE TABLE lesssales\_products AS

SELECT

p.ProductKey,

p.product\_name,

p.Brand,

p.Category,

COUNT(s.ProductKey) AS sales\_count

FROM

dataspark.products2 AS p

LEFT JOIN

dataspark.sales2 AS s ON p.ProductKey = s.ProductKey

GROUP BY

p.ProductKey, p.product\_name, p.Brand, p.Category

HAVING

sales\_count < 3;

5.CREATE TABLE male\_customers AS

SELECT

CustomerKey,

Gender,

Cus\_name,

City,

State\_Code,

State,

Zip\_Code,

Country,

Continent,

Birthday

FROM

dataspark.customer2

WHERE

Gender = 'male'; -- Filter for male customers

6.. highsales\_category\_analysis

CREATE TABLE highsales\_category\_analysis AS

SELECT

p.Category,

p.Subcategory,

COUNT(s.ProductKey) AS total\_sales,

COUNT(DISTINCT p.ProductKey) AS unique\_products

FROM

dataspark.products2 AS p

LEFT JOIN

dataspark.sales2 AS s ON p.ProductKey = s.ProductKey

GROUP BY

p.Category, p.Subcategory

HAVING

total\_sales >= 10;

7.get large stores in size

CREATE TABLE large\_stores AS

SELECT

StoreKey,

Country,

State,

Square\_Meters,

Open\_Date

FROM

dataspark.store2

WHERE

Square\_Meters > 300;

8..

CREATE TABLE store\_currency\_analysis AS

SELECT

StoreKey,

Currency\_Code,

COUNT(DISTINCT Order\_Number) AS order\_count,

SUM(Quantity) AS total\_quantity

FROM

dataspark.sales2

GROUP BY

StoreKey, Currency\_Code

ORDER BY

StoreKey, Currency\_Code;

9.. CREATE TABLE fast\_moving\_colors AS

SELECT

p.Color,

SUM(s.Quantity) AS total\_sales

FROM

dataspark.products2 AS p

JOIN

dataspark.sales2 AS s ON p.ProductKey = s.ProductKey

GROUP BY

p.Color

ORDER BY

total\_sales DESC;

10. CREATE TABLE categorized\_products AS

SELECT

p.Category,

p.ProductKey,

p.Product\_Name,

p.Brand,

p.Color,

p.Unit\_Cost\_USD,

p.Unit\_Price\_USD,

p.SubcategoryKey,

p.Subcategory

FROM

dataspark.products2 AS p

ORDER BY

p.Category, p.Product\_Name;

**conclusion:**

In this project, we successfully completed the following key steps:

* **Data Cleaning and Preparation**: All datasets have been thoroughly cleaned and integrated, ensuring they are ready for analysis. This step involved addressing missing values, correcting data types, and standardizing formats to maintain data integrity.
* **Exploratory Data Analysis (EDA)**: Comprehensive EDA was conducted to uncover patterns, trends, and insights within the data. This analysis facilitated a deeper understanding of the dataset and guided subsequent analytical processes.
* **Visualizations**: Business Intelligence (BI) visualizations were created to effectively communicate the key findings from our analysis. These visualizations provide clear and impactful representations of the data, aiding stakeholders in decision-making.