

---

# Retail Sales: Design Document

---



**FEBRUARY 25**

---

**Student Name – Richik Mukherjee**  
**Student ID - 922176**

---

## Table of Contents

1. Introduction .....	3
1.1 Purpose .....	3
1.2 Scope .....	3
2. System Overview .....	3
2.1 System Architecture.....	3
2.2 System Components .....	3
3. Data Management.....	3
3.1 Data Sourcing.....	3
3.2 ETL Process .....	4
4. Data Modelling .....	4
4.1 Entity-Relationship Diagram.....	4
4.2 Schema Design .....	5
5. Visualization & User Interface .....	5
5.1 Layout.....	5
5.2 Features & Functionalities .....	7
6. Technical Requirements.....	7
6.1 Software & Tools .....	7
6.2 Hardware Requirements .....	7
7. Milestones & Timeline .....	7
8. Conclusion .....	9

## List of Figures

Figure 1: Data Model .....	4
Figure 2: Dashboard 1 .....	6
Figure 3: Dashboard 2.....	6
Figure 4: Gantt Chart.....	8

## List of Tables

Table 1: Project Timeline .....	8
---------------------------------	---

---

# 1. Introduction

## 1.1 Purpose

The purpose of choosing this dataset was to understand the 3 main fundamentals of retail chain operations globally namely, revenue, profit and customer demographics. These elements help analysts forecast the demand curve of products based on the highest and lowest sold and profitable items.

## 1.2 Scope

The project covers the in-depth analysis of 3 main product categories namely accessories, bikes and clothing globally. These categories are evaluated step wise in Power BI highlighting the trend of product categories and the subcategories across different countries. This dataset does not contain any specific information about a company or its relevant annual report, it is a generalized dataset sourced online.

# 2. System Overview

## 2.1 System Architecture

The project will feature the in depth analysis of the Retail Sales Data sourced from [Kaggle](#), the dataset will be primarily stored in a laptop storage in separate folder under BI documents in Local Disk C. The dataset is in a raw CSV format and is extracted and transformed in Power BI using DAX queries to create a snow flake schema. The data model illustrates the queries and the entities, finally the visualizations are performed using interactive and immersive dashboards to create a clear understanding of the objectives making it open for collaborations within the corporate dynamics.

## 2.2 System Components

- **User Interface:** The User Interface for this project will feature the use of interactive Power BI dashboard, there will be slicers, filters, and cards present to navigate through the different segments of the data.
- **Backend Processing:** The backend processing will be performed in Power BI firstly extracting the data and using DAX queries in Power BI, here measures will be created to transform the data after it is extracted. This makes a clear and concise approach towards visualizing data later through dashboards.
- **Database:** The dataset will be stored in raw CSV format in the internal laptop storage under Local Disk C. The manipulation and remodeling of the data can be done by directly accessing it through MS Excel to justify its accessibility prior to importing the data in Power BI.

# 3. Data Management

## 3.1 Data Sourcing

The dataset had been sourced through this website '[Kaggle](#)', this platform provides an exhaustive overview of different types of datasets that can be used for data analytics and relevant data modelling or ETL operations.

### 3.2 ETL Process

- **Extract:** The data is extracted from this website '[Kaggle](#)' and stored in internal storage in a laptop, later this initial step follows the CSV format being imported in Power BI which includes the information of the product categories, sub-categories, cost, revenue, quantity and profit.
- **Transform:** The initial transformation is done before importing the file to Power BI, here additional columns are added in MS Excel to complete the data. Further, the dataset is then imported in Power BI where additional cleansing and transformation is done through data transformation tab in Power BI to remove null values, blank rows and duplicates. Moreover, the final transformation includes the creation of DAX queries for additional calculations or formatting.
- **Load:** The final step is completed by loading the finished dataset into Power BI after creation of measures table by clicking apply and load data. This then creates a data model schema which includes all the relational schemas, later the entities in the data model tab are arranged in a snow flake schema to illustrate the proper flow of data.

## 4. Data Modelling

### 4.1 Entity-Relationship Diagram

The relational data model is created in the format of a snow flake schema, through this it creates better understandability by showcasing the order table at the center, it connects the other three tables customer, calendar lookup and product tables to justify the visualization capabilities of the data. The fourth entity is the measures table that is created with DAX queries to increase accessibility of visualization and further calculations in the dataset.

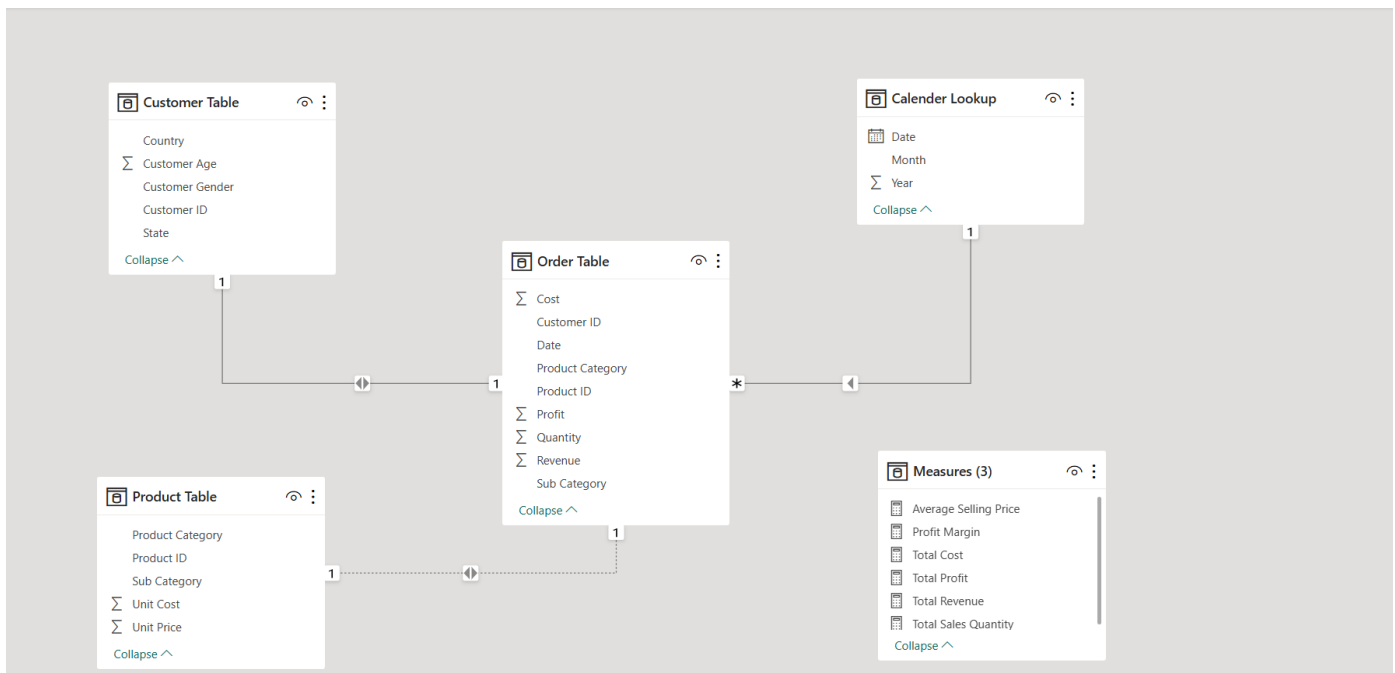


Figure 1: Data Model

---

## 4.2 Schema Design

The relational schema is constructed to show the extraction and transformation of the Retail Sales Data. This dataset contains 8 columns based on which four different tables are created namely, order table, customer table, calendar lookup and product table. The measures table is created for further remodeling or in-depth calculations. The 4 main tables include the information about the product categories, cost, customer ID, total sales quantity, revenue and sub-categories, these entities are connected with the other independent tables through one to many and many to one relationship.

- **Order Table** – This table is the main interface which connects the other three through Date, Customer ID, and Product Category. This table defines the total sales quantities, product categories and the subcategories followed by the cost of the units, price and the related customer IDs involved in each transaction.
- **Calendar Lookup** – This table includes the Date, Month and the Year for clear date wise assortment of the data upon visualization.
- **Customer Table** – This table includes the information about the customer, mainly customer demographics can be understood through the navigation of this table, it includes the Customer ID, Country, Customer Age and Gender.
- **Product Table** – This table shows the different product categories and its sub-categories along with the Product IDs, unit cost and price. This helps to navigate the different types of products in demand based on the cost and revenue.
- **Measures Table** – This table is created using DAX queries for a quick reference or remodeling the data upon visualization.

## 5. Visualization & User Interface

### 5.1 Layout

The first visualization dashboard includes the differentiation of profit and revenue based on product categories, country and sales quantity. The analysis starts from the total revenue and total profit based on the product categories by the use of pie charts, this visualization is followed by the next one that shows total profit and total revenue by country with the use of a line graph and a bar graph respectively. The total sales quantity by country is shown next with the use of a funnel chart. Upon hovering above, the summarized results of total revenue, total profit, total sales quantity and the highest and least sold products are shown with cards. There is also a slicer of sub-categories on the right corner for viewer navigation across different possibilities.

Respectively, the customer demographics includes comprehensive analysis of the product category based on the customer age, and gender, they are evaluated against the profit and revenue through the visualization of bar charts and line graphs. There is also the slicer for month to assort the trends based on the preferred timeline.

Please view the figures below.

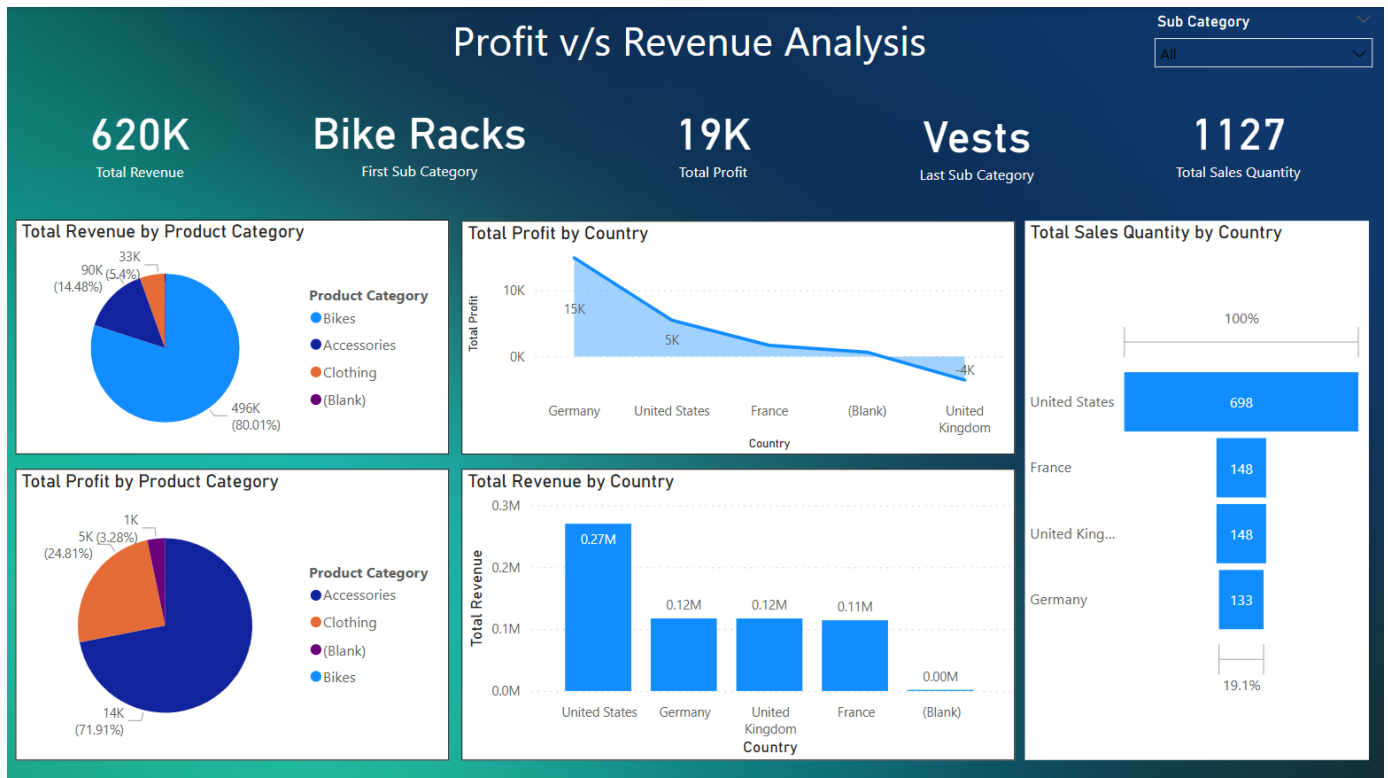


Figure 2: Dashboard 1

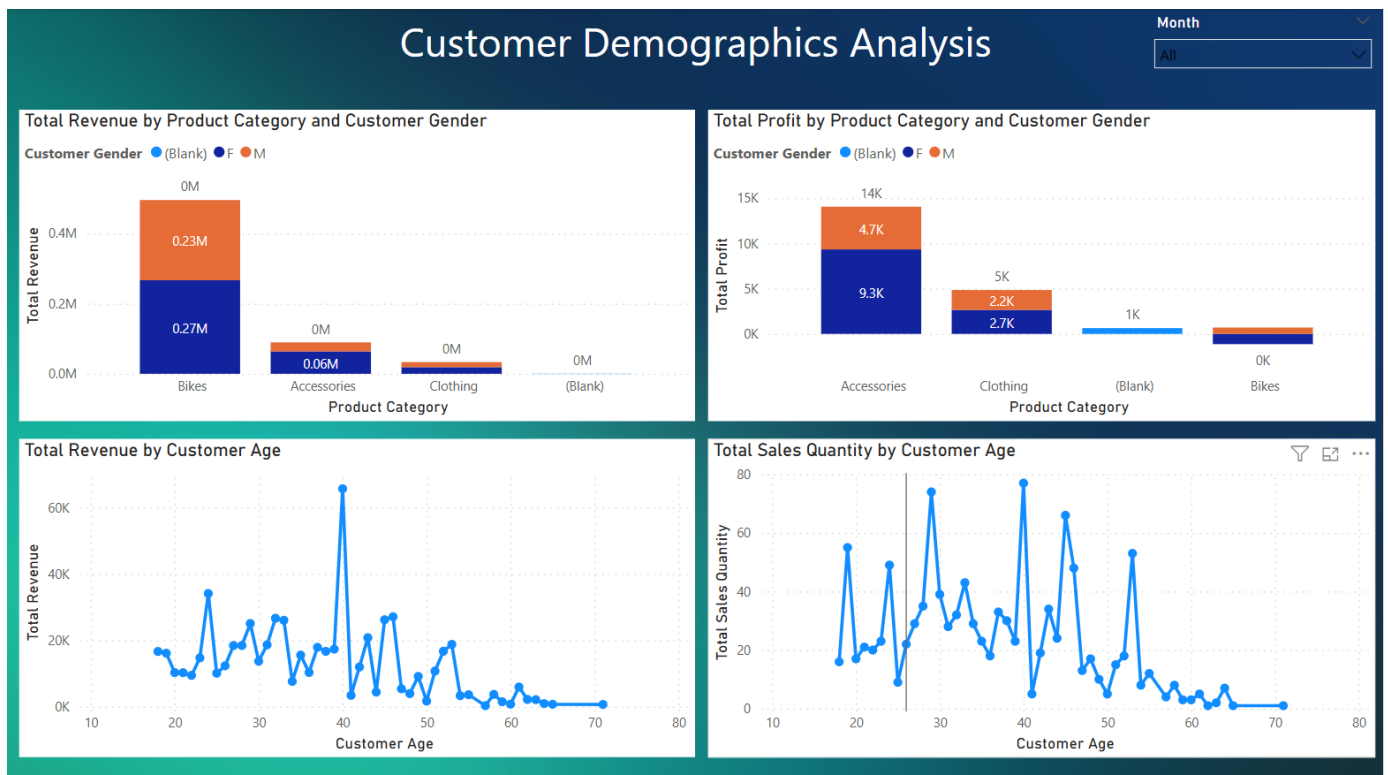


Figure 3: Dashboard 2

---

## 5.2 Features & Functionalities

The following are the features and functions of the dashboard

- **Interactive Elements:** The user can enjoy the use of slicers, immersive and interactive charts and graphs, and cards to navigate through the data based on individual preferences.
- **Filter & Search Option:** Users can use slicers, operate through the interactive graphs and charts and filter through information like analyzing product category based on profits grouped by customer age/gender for a specific evaluation.
- **Navigation:** The presence of sequence wise graphs and charts with the slicers will help the viewers to navigate through the visualizations, moreover the goals and objectives of the data will provide additional clarity to the viewers.

## 6. Technical Requirements

### 6.1 Software & Tools

The following are the tools used for the project

- **MS Excel** – The initial data cleansing was done in MS Excel itself to incorporate additional columns.
- **Power Query (DAX)** – Power Queries were used to further evaluate and transform data to create separate measures for the efficient data navigation and modelling. The data transformation function was used for cleansing the data and removing null and blank values from the dataset, making it leaner and ready for visualization.
- **Power BI Desktop** – This was used for creating immersive and interactive dashboards for viewers to get a fuller understanding of the data being presented.

### 6.2 Hardware Requirements

The hardware requirements for running Power BI smoothly are listed below, the list contains both minimum and recommended specializations for better utilization.

Processor (CPU):

- Minimum: 13th Gen Intel(R) Core(TM) i5-1340P 1.90 GHz
- Recommended: Intel Core i7 (10th Gen) / AMD Ryzen 7 or higher

Memory (RAM):

- Minimum: 8GB
- Recommended: 16GB (for medium datasets)
- Ideal: 32GB+ (for large datasets and complex calculations)

Storage (SSD Recommended):

- Minimum: 256GB SSD
- Recommended: 512GB SSD

Graphics Processing Unit (GPU):



- Recommended: Dedicated GPU with DirectX 11 support (e.g., NVIDIA GeForce GTX 1660, RTX 3060, or equivalent for Power BI Desktop)
- Required for AI & ML features: NVIDIA RTX Series with CUDA support

## 7. Milestones & Timeline

The timeline for the project is described below with the use of a Gantt Chart. The tasks start from dataset selection which complied with the goals and objectives for the project. The next stage is the resource gathering where the general knowledge behind the dataset is gathered through surfing online about sales data and how comparisons are made. Furthermore, the data extraction is the next phase which took 3 days due to cleaning the dataset and adding needed columns. The next phase is the cleansing or transformation, here the in-depth changes were made and data was further cleaned through Power BI and DAX Queries. The final stage is the creation of a dashboard, the final stage took 2 days. Therefore, in total the duration of the project was of 11 days.

Task	Start Date	End Date	Duration (Days)
Dataset Selection	1/23/2025	1/26/2025	3
Resource Gathering	1/26/2025	1/26/2025	1
Data Extraction	1/26/2025	1/29/2025	3
Data Cleaning/transformation	1/31/2025	2/1/2025	2
Loading into Dashboard	2/2/2025	2/3/2025	2
<b>TOTAL</b>			<b>11</b>

Table 1: Project Timeline



Figure 4: Gantt Chart



---

## 8. Conclusion

In conclusion, the Retail Sales Data can be effectively used by corporates willing to expand businesses or any think tank organizations to better educate the public regarding market research and product marketing. Viewers can take reference of the dataset and cleanse the data in their own way using Python, Power Query or MS Excel. The Retail Sales Data is a common dataset, however for every company the insights about business development is important, based on that analysis of the market and products must be conducted, therefore this project assists entrepreneurs, companies and other corporate entities get a know-how about data analytics and forecasting revenue for the next year.