



Explorational Analysis and Visualization of Retail Sales Dataset; Using SQL, POWER BI and EXCEL



Peace Ojukwu

5 min read · 2 days ago



Retail Sales data exploration

INTRODUCTION

Data exploration is an initial approach to data analysis, whereby a data analyst uses visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems. These characteristics can include size or amount of data, completeness of the data, correctness of the data, possible relationships amongst data elements or files/tables in the data.

Data exploration is typically conducted using a combination of automated and manual activities. Automated activities can include data profiling or data visualization or tabular reports to give the analyst an initial view into the data and an understanding of key characteristics. This is often followed by manual drill-down or filtering of the data to identify anomalies or patterns identified through the automated actions. Data exploration can also require manual scripting and queries into the data (e.g. using languages such as SQL or R) or using spreadsheets or similar tools to view the raw data.

ABOUT DATASET

Retail sales dataset is a collection of data that contains information about the sales of products or services in the retail industry. Retail sales dataset can be used for various purposes, such as analyzing customer behavior, forecasting demand, optimizing pricing, improving inventory management, and enhancing marketing strategies.

This dataset contains historical sales data from a store, along with information about customer ID, Gender, Transaction ID, Product Category,

Age, Quantity, Price and Product Value. This dataset can be used to analyze the most valued customers, impact of gender to the store and external factors on sales performance.

TASK BREAKDOWN

The idea of exploration of this dataset was brought about by the quest for perfection in data analysis with constant study and practice. This dataset was sourced from kaggle and is made up of few columns and a thousand row.

TOOLS

The exploration and visualization of this data set gave head way to the practice of 3 data analysis tools (Excel, SQL and Power BI).

Excel spreadsheet was used to clean data set by finding duplicates and missing values.

SQL work tool was used for the modification and querying of the data set to understand the behavior and objective of the analysis.

Power BI was the work tool used to get this analysis visual done. This dynamic tool serves as the pictorial display where data, insights, and creativity converge. Its ability to ingest, process, and present data is unparalleled, making it the perfect companion for this journey.

AIM AND OBJECTIVES

This article aims at understanding the relationship between customers age groups and their preferred product category.

Objectives:

1. To perform cleaning and transformation of Retail dataset.
2. To design and develop interactive dashboard using Power BI, showcasing relationships and actionable insight.
3. To ask data driven questions and develop queries to give actionable results.

PROCESS

Data summarization:

Summary of the data to ascertain average, mean and median of table fields.

Data cleaning:

The tool used for this process was Microsoft excel. The dataset was cleaned column after column to find duplicates using filter. No duplicate was found and no missing values.

Data transformation and modification

The dataset was modified using SQL, for altering and organizing the dataset. This is to make data useable for analysis. Adding columns of year, month and age groups. The process was done by modifying and inserting the modifications to a new table.

```

--Data modification
Use [Data exploration]
go

select
Distinct
[Transaction ID], cast(Date as date) as Date,
--New column; Year
year(Date) as Year,
--New column; Month of the year
DATENAME(Month,Date) as Month,
-----New column; Day of the month
DATENAME(Day,Date) as Day, [Customer ID],Gender, Age,
---Age group range
case when Age <18 then 'Under18'
      when Age between 18 and 30 then '18-30'
      when Age between 31 and 40 then '31-40'
      when Age between 41 and 50 then '41-50'
      when Age between 51 and 60 then '51-60'
      when Age between 61 and 70 then '61-70'
      when Age between 71 and 80 then '71-80'
      else 'old' end as AgeGroup,
[Product Category],Quantity, [Price per Unit],[Total Amount]
into [Data exploration]..retailSales
from [Data exploration]..retail_sales_dataset

```

INSIGHTS

The Retail sales Analysis Dashboard was a compass, displaying sales performance by customers and product preference. In this article, I charted a course through multiple analytical avenues.

— CUSTOMER SALES ANALYSIS

1. How does Customer gender affect sales?

```
select Gender, Count([Total Amount]) as NumberOfSales
from [Data exploration]..retailSales
group by Gender
order by NumberOfSales desc
```

```
-----
select Gender, sum([Total Amount]) as TotalSales
from [Data exploration]..retailSales
group by Gender
order by TotalSales desc
```

```
select Gender, [Product Category], sum([Total Amount]) as TotalSales
from [Data exploration]..retailSales
where Gender = 'female'
group by Gender, [Product Category]
order by TotalSales desc
```

```
-----
select Gender, [Product Category], sum([Total Amount]) as TotalSales
from [Data exploration]..retailSales
where Gender = 'male'
group by Gender, [Product Category]
order by TotalSales desc
```

Females top the chart when it comes to purchases with total purchase of 232.84 dollars.

2. How does customer age affect Sales?

```
select AgeGroup, Count([Total Amount]) as NumberOfSales
from [Data exploration]..retailSales
group by AgeGroup
order by NumberOfSales desc
```

```
select AgeGroup, sum([Total Amount]) as Total_sales
from [Data exploration]..retailSales
```

```
group by AgeGroup
order by Total_sales desc
```

According to the analysis, consumers between the ages of 18 and 30 make more purchases.

3. Most preferred customer product?

```
select [Product Category], count([Customer ID]) as NumberOfCustomers
from [Data exploration]..retailSales
group by [Product Category]
```

Electronics are the most frequently purchased product, but clothing came in second place on the chart given the prevalence of women in the market.

4. Relationship between Age groups and product preference

```
select AgeGroup, [Product Category], count([Customer ID]) as NumberOfCustomers
from [Data exploration]..retailSales
where [Product Category] = 'Clothing'
group by AgeGroup, [Product Category]
order by NumberOfCustomers DESC
-----Beauty
select AgeGroup, [Product Category], count([Customer ID]) as NumberOfCustomers
from [Data exploration]..retailSales
where [Product Category] = 'Beauty'
group by AgeGroup, [Product Category]
order by NumberOfCustomers DESC
-----Electronics
select AgeGroup, [Product Category], count([Customer ID]) as NumberOfCustomers
from [Data exploration]..retailSales
where [Product Category] = 'Electronics'
```

```
group by AgeGroup, [Product Category]
order by NumberOfCustomers DESC
```

According to our data, those between the ages of 31 and 60 preferred to invest more on clothing, while those in the younger demographic (18–30) bought more cosmetics. Those between the ages of 61 to 70, on the other hand, tended to buy electronics.

5. Customers shopping trends

```
select Month, sum([Total Amount]) as TotalSales
from [Data exploration]..retailSales
group by Month
order by TotalSales
-----
select Year, sum([Total Amount]) as TotalSales
from [Data exploration]..retailSales
group by Year
```

From the time of data collection, there has been 100% decrease in sales from 2022 to 2023.

THE ART OF VISUALIZATION

The Art of Visualization is a technique that involves creating a mental image of something you want to achieve or accomplish. It is a powerful tool that can help you to focus your mind and achieve your goals. The explored questions on the retail sales analysis can be visualized using the tool POWER BI, this helps with good decision making.



CONCLUSION

Businesses benefit from the potential advantages of an analyst through exploratory investigation. Through inventory optimization and market strategy implementation, the store may better adapt to changing client needs and sustain a steady flow or increase of revenue all year round. This will help with question-related insights and data-driven decision-making.



Written by Peace Ojukwu

Edit profile