**Real-Time Capstone Project: Retail Sales & Customer Insights Dashboard**

**Business Problem Statement**

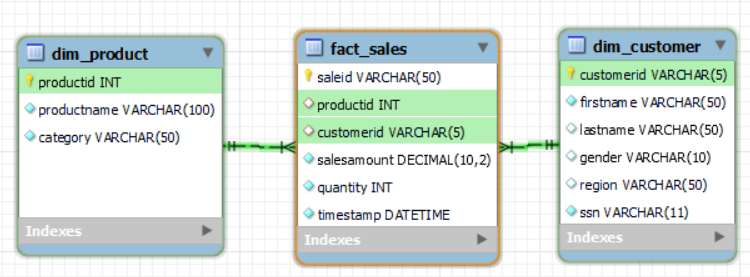
A retail chain with multiple stores is experiencing several challenges due to its extensive and varied customer base, wide range of products, and regional distribution. These challenges include:

1. **Identifying Top-Performing Products:**
2. **Understanding Complex Purchasing Patterns:**
   * Seasonality
   * Promotions
   * Customer Demographics
3. **Visualizing Regional Sales Trends:**
   * Inventory Optimization
   * Marketing Strategies

**Workflow:**

**Database Creation**:

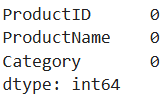
* Created a database ‘retail’ in MYSQL Workbench.
* **Star Schema Design –** Build a warehouse schema to optimize queries for inventory and supply chain management.
* I have created tables in the database using ‘star schema’ design.
* I created total 3 tables – 2 dimension tables and 1 fact table.



* Dim\_product, dim\_customer and fact\_sales are the tables.

**Importing data into Tables using Python**:

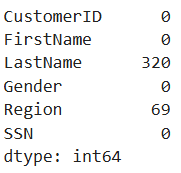
* Now I have connected the python and database tables using mysql connector library.
* The cursor is used to execute SQL queries.
* **product** data is residing in ‘products.csv’.
* loading data into product DataFrame.
* Checked for null values and duplicates in the DataFrame.

 Findings: There are no null and duplicate values.

* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_product table.
* Checked the dim\_product table in the MySQL Workbench all the records have been inserted.



* Findings: All the 50 records are imported.
* **customer** data is residing in ‘customers.json’.
* loading data into customer DataFrame.
* Checked for null values and duplicates in the DataFrame.



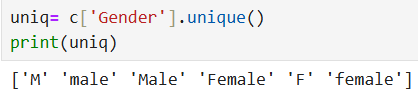
* Findings: There are null values in last name and region columns.

Before:  

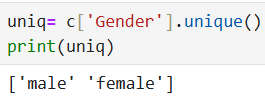

After:



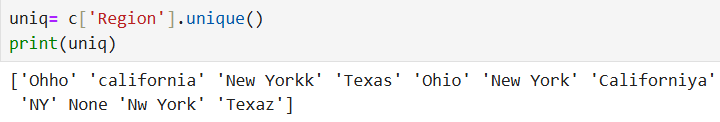
* Findings: There are 100 duplicate rows found and those are dropped.
* For gender column I found unique gender values.



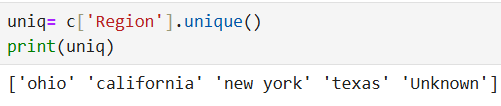
* Findings: same values are named differently.
* Converted the gender column values to lower and considered only ‘male’ and ‘female’ genders and mapped.

 Findings: after mapping there are only 2 genders.

* For region column I found unique region values.

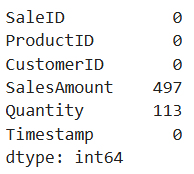


* Findings: same values are named differently.
* Converted the region column values to lower and considered only ‘ohio’, ‘california’, ‘new york’ and ‘Texas’ regions and mapped.
* And filled the null values as ‘unkwon’.



* Findings: after mapping there are only 5 regions.
* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_customer table.
* Checked the dim\_customer table in the MySQL Workbench all the records have been inserted.



* Findings: All the 1000 records are imported.
* **sales** data is residing in ‘sales 1.csv’.
* loading data into sales DataFrame.
* Checked for null values and duplicates in the DataFrame.
*  Findings: There are null values in sales amount and quantity columns and no duplicate rows.

**Handling null values**

* If both sales amount and quantity values are null in a row then drop it.

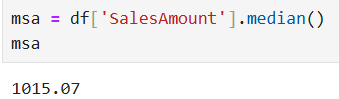
Before dropping:

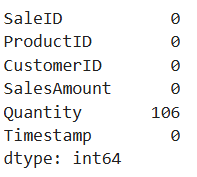


After dropping:

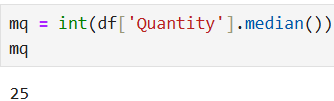
 Findings: There are 7 rows found and those are dropped.

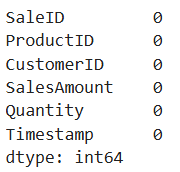
* For sales amount null values I calculated median and replaced null values.



 Findings: There are no null for sales amount.

* For quantity null values I calculated median and replaced null values.



 Findings: There are no null values.

* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the fact\_sales table.
* Checked the fact\_sales table in the MySQL Workbench all the records have been inserted.



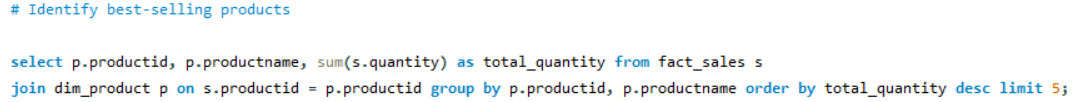
Findings: it says 1000 records are imported but we have only 9993 records.

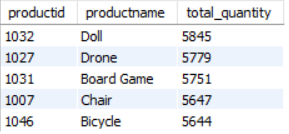


Findings: All the 9993 records are imported.

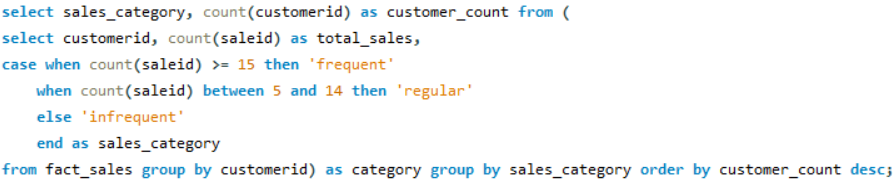
**SQL Tasks:**

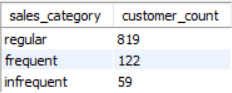
1. Identify best-selling products



 Findings: Doll is the best selling product.

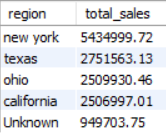
1. Segment customers based on purchase patterns



 Findings: regular customers who order 5 -14 times are more.

1. Analyze regional sales trends.



 Findings: New York region has the highest sales.

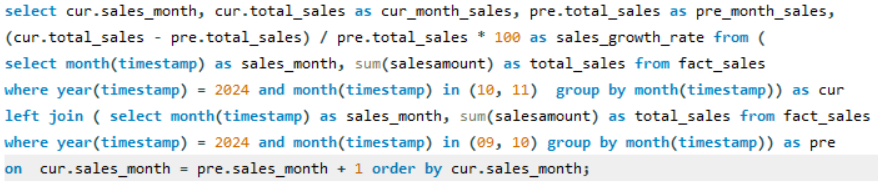
1. Total Sales Revenue

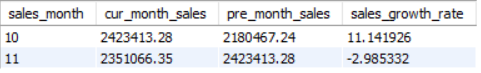


Findings: Total Sales Revenue is 1,41,53,194 crores.

1. Sales Growth Rate

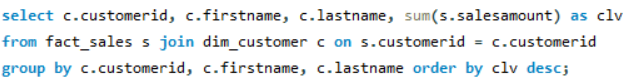
Formula: (cur.total\_sales - pre.total\_sales) / pre.total\_sales \* 100

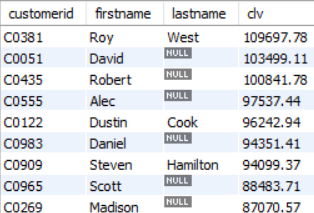




* Findings: Sales growth I calculated for 10 and 11month. There is increase in sales in 9-10 month. But there is decrease in sales in 10-11 month.

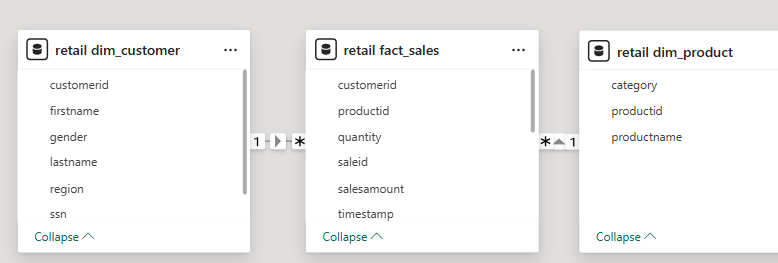
1. **Customer Lifetime Value (CLV)**



Findings: Roy west has the highest customer lifetime value.

**Power BI Reporting and Analytics**

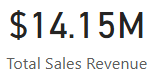
* The 3 Tables in MYSQL Workbench are loaded into power query for transformation.
* The data types are changed for ID columns and item number from number to text as there is no need of performing mathematical calculations.
* After transformation data is loaded into report.
* Goto Model view and check the relationships.



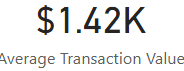
* I added title ‘**Retail Sales & Customer Insights Dashboard**’ using text box.
* Created measure ‘**total sales’**. Displayed using card visual under retail fact\_sales table.

 Findings: total number of sales are 9993.

* Created measure ‘**Total Sales Revenue’**. Displayed using card visual under retail fact\_sales table.

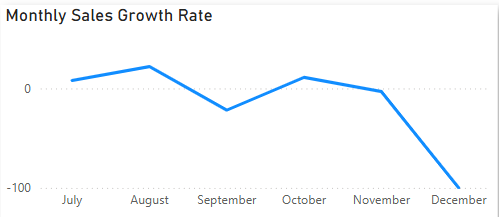
 Findings: total revenue generated from sales is 14.15 million.

* Created measure ‘**Average Transaction Value’**. Displayed using card visual under retail fact\_sales table. By calculating total sales revenue/ total sales.

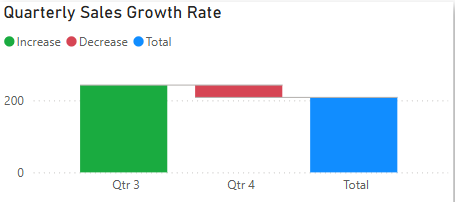
 Findings: Average Transaction value or amount is 1,420.

**Sales Growth Rate**

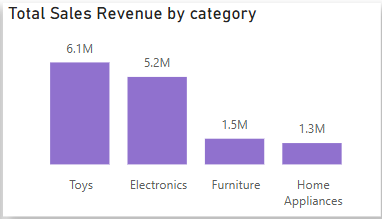
* Created measure ‘**Monthly Sales Growth Rate’** under retail fact\_sales table. Using (curmonthsales - premonthsales) / premonthsales \* 100 formula.
* Displayed using line chart visual.

Findings: Sales growth rate is highest in August and lowest in September we cannot consider December as that month data is not available. Sales data available is from August – November.

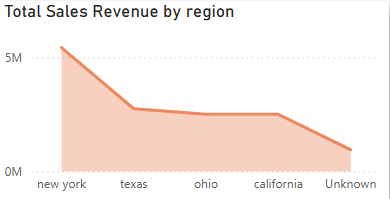
* Created measure ‘**Quarterly Sales Growth Rate’** under retail fact\_sales table. Using (curmonthsales - premonthsales) / premonthsales \* 100 formula but grouping for each quarter.
* Displayed using waterfall chart visual.

Findings: Sales growth rate in Qtr 3 is increasing and Qtr 4 is decreasing. There are only 2 quarters data.

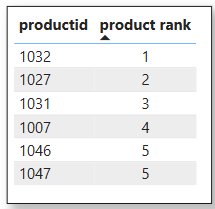
* **Total Sales Revenue by category** plotted total sales revenue and category using clustered column chart.

 Findings: Toys category generates highest revenue.

* **Total Sales Revenue by region** plotted total sales revenue and region using area chart.

 Findings: New York generates highest revenue.

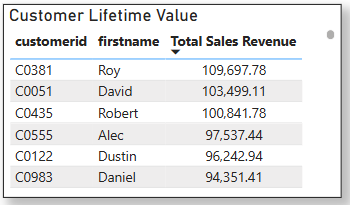
* Created measure ‘**Total Quantity Sold**.**’** under retail fact\_sales table.
* Created measure ‘**Product rank’** under retail fact\_sales table. Based on Total Quantity Sold. Displayed using table visual.
* To get only top 5 products I added filter on product id top 5.

 Findings: 1032 product is the highest sold product by quantity.

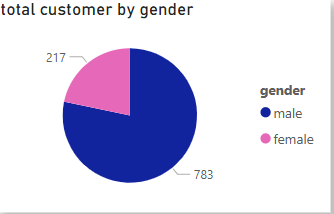
* Created measure ‘**total customers’**. Displayed using card visual under retail dim\_customer table.

Findings: total number of customers are 1000.

* Customer Lifetime Value - plotted customer id, customer id and total sales revenue using table visual.

 Findings: C0381 Roy is has the highest customer lifetime value.

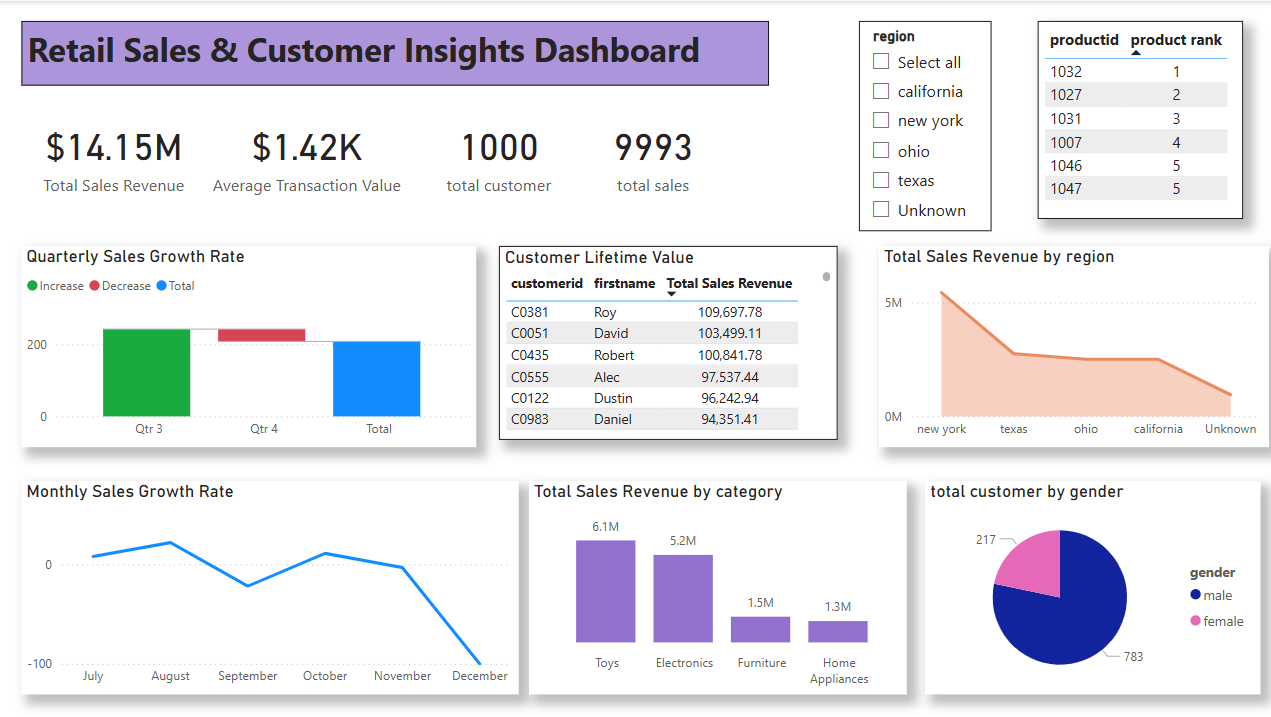
* **Total customers by gender** – plotted gender and total customers by pie chart.

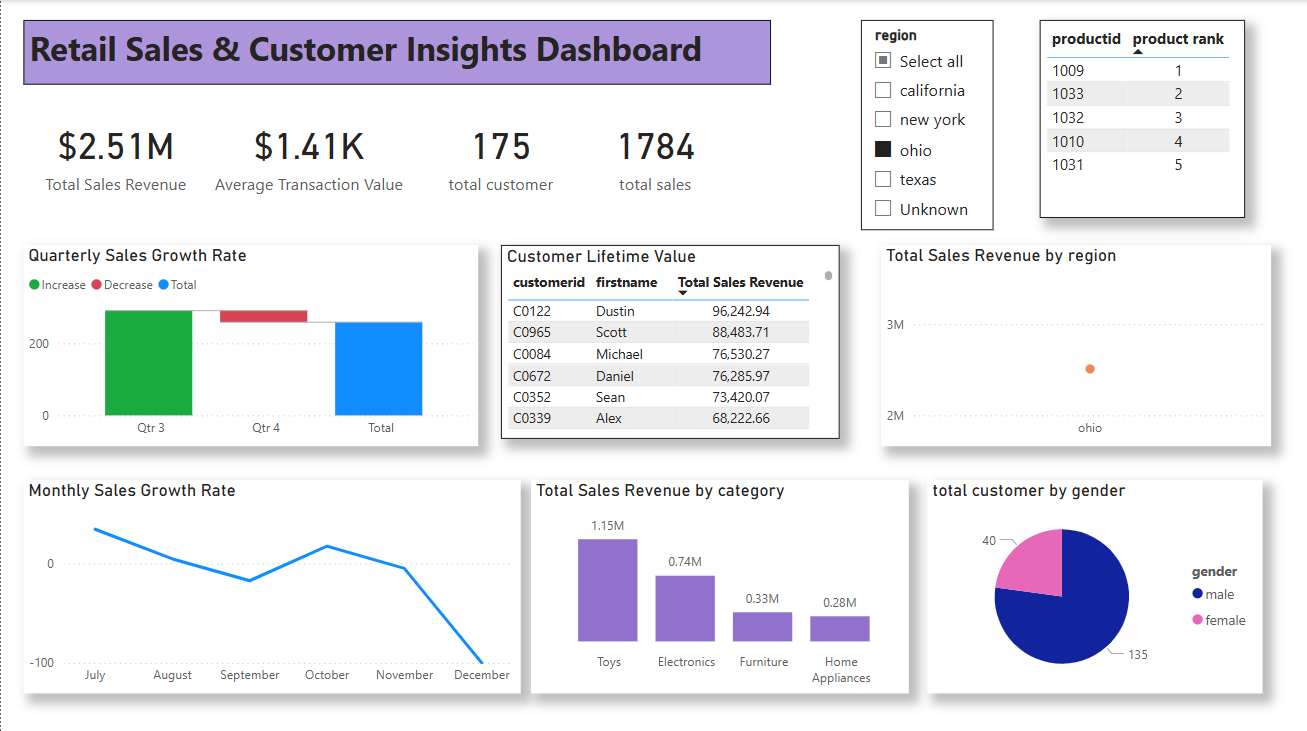
 Findings: Male customers are more in number.

* Created slicer for region to filter all the visuals.



**Complete Report**





The project focuses on improving decision-making, understanding complex purchasing patterns, visualizing regional sales trends.