**Walmart Dataset –Uncovered**

This document contains my observations at each step of analyzing the dataset. It includes details about the dataset's structure, data quality checks, unique values, and other key findings. Each step's output is recorded along with insights derived from the analysis.

**Understanding Dataset:**

**data.columns :**

We have following columns in dataset.

* invoice\_id
* Branch
* City
* category
* unit\_price
* quantity
* date
* time
* payment\_method
* rating
* profit\_margin

**data.info() :**

* Null values in unit\_price and quantity
* unit\_price is in object data type have to convert it to float and remove dollar sign.

**data.nunique():**

* There are 100 branches in 98 cities.
* Six different categories of products.
* 3 types of payment methods.
* Data is from 1st Jan 2019 to 31st Dec 2023 mean 4 years. Total 1460 Days.
* 1460 unique dates are there in dataset Sales were recorded daily.

**data.nunique():**

* 31 null value in unit\_price and quantity each out of 10051.

**data.duplicated().sum():**

* 51 duplicates entries.

**Cleaning Dataset:**

* Dropping duplicate rows.
* Removing dollar sign from unit\_price column
* Changing Data type of unit\_price from object to float.
* Changing Data type of date column from object to datetime.
* Renaming City and Branch column to completely lower case.
* Dropping null values columns.
* Deriving new columns (total and profit).
* Dropping unnecessary columns (unit\_price, profit\_margin)
* **Loading Cleaned Data to MS SQL SERVERS and New csv file.**

**EDA through SQL**

Business problems to answer through SQL analysis.

1. What are the different payment methods, and how many transactions and items were sold with each method?
2. Which category received the highest average rating in each branch?
3. What is the busiest day of the week for each branch based on transaction volume?
4. How many items were sold through each payment method?
5. What are the average, minimum, and maximum ratings for each category in each city?
6. What is the total profit for each category, ranked from highest to lowest?
7. What is the most frequently used payment method in each branch?
8. How many transactions occur in each shift (Morning, Afternoon, Evening) across branches?
9. Which branches experienced the largest decrease in revenue compared to the previous

year?

Answers are in sql file.