Project Report

Course IT-300

Business Intelligence and Database Management Systems

Convenience Store Chain Analysis

Prepared By:

Ahmed Kharrat

Wael Zaier

Skander Challouf

Mohamed Daoud

Contents

Contents

1 Introduction

This business intelligence project revolves around the examination of data related to Convenience stores in the United States of America. The objective of the study is to help the decision makers of a convenience store chain company to make decisions regarding the opening of new stores or closing certain stores and which items to focus on in the upcoming marketing campaign by investigating various factors, including Sales, Operational Expenses, Regions, and the most selling items on each store. By analyzing these data, we seek to acquire valuable insights about stores' performance and their potential. This understanding can assist the company in their expansion. Through the implementation of this project, our goal is to determine the most optimal decisions regarding the stores and to enhance overall profitability of the company and formulate more effective strategies for it.

2 Implementation

2.1 Data Gathering

We extracted the convenience stores data of the United States of America from the website Kaggle.com

2.2 ETL Process

For the data preparation, we used Talend to manipulate data and configure it for the data warehouse. First, we sorted the file "stores1.csv" using the tSortRow component in Talend and removed 2 irrelevant columns from the schema which are "Store surface area"

and “available items” because in this analysis the size of the store is unimportant and the available items information will be replaced by the “inventory turnover rate” in the other data file “storesdetailed.xml”.

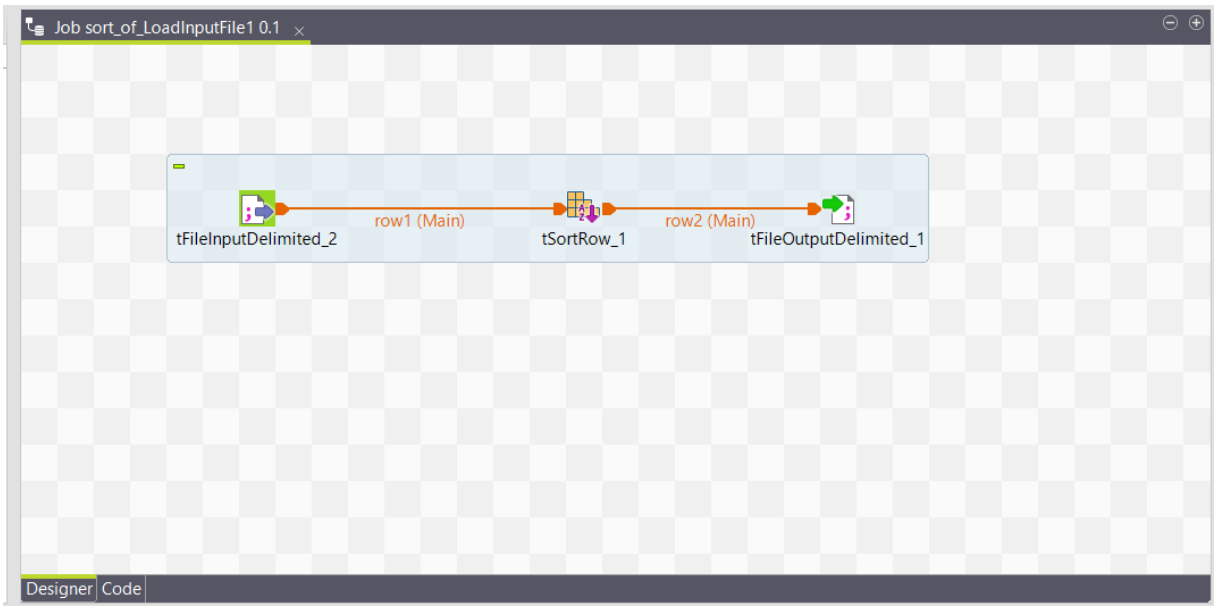

Figure 1: Sorting Job

The Talend code can be found in the “Sorting Job Talend.txt”.

Then we imported 3 files “outsort.csv” (that is the output of sorting “stores1.csv”), “storesdetailed.xml” and “store ratings.csv” and we joined them together using the tMap component and it generated the dataset “Talend output.csv”.

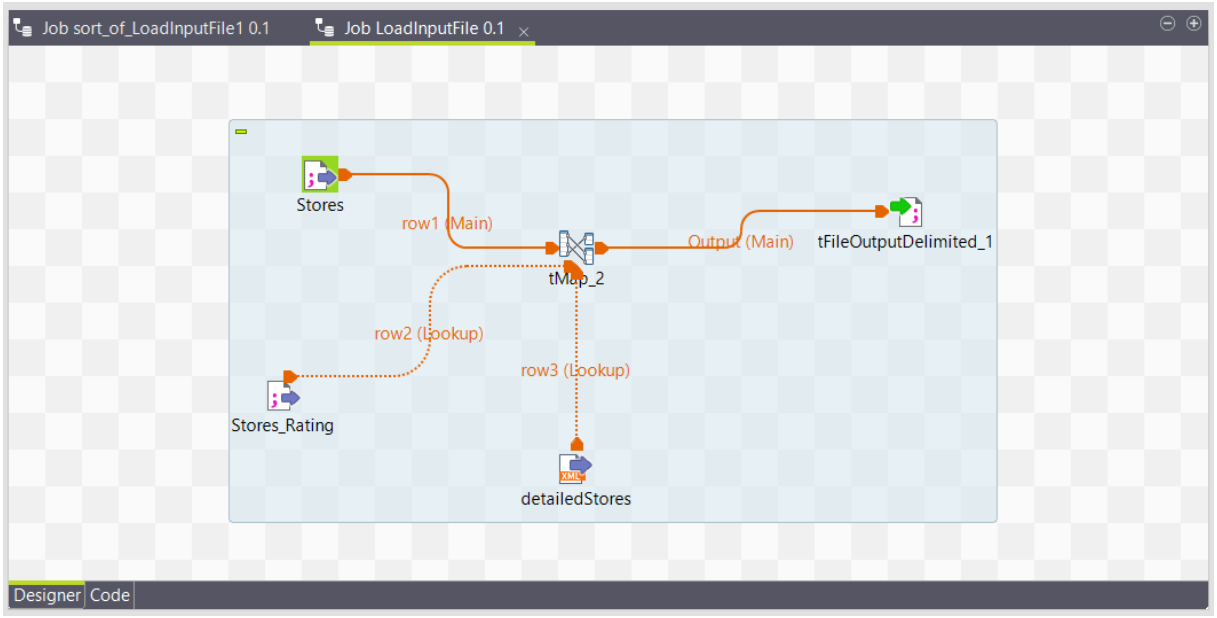

Figure 2: Merging Job

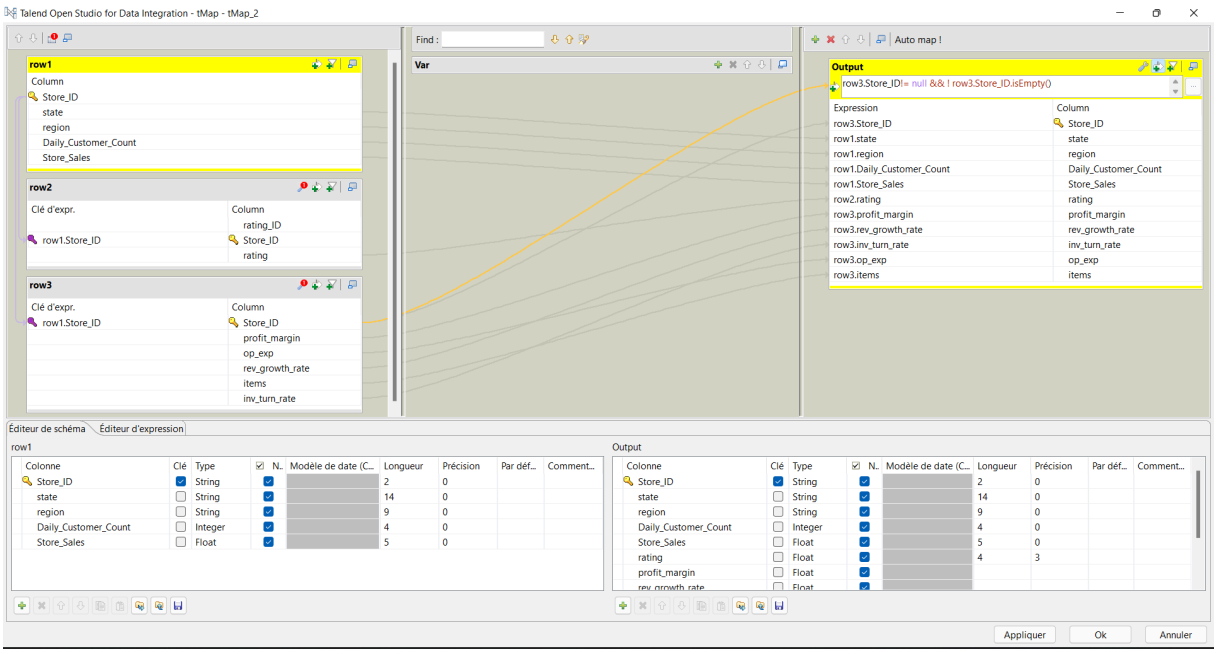

Figure 3: tMap

The Talend code can be found in the “Merging job Talend.txt”.

2.3 Data Storage

2.3.1 Storage

For the data storage, we used SQL Server Management Studio (SMSS) as the database management system to store the data in SQL Server Express.

Location: represents the location of each store by state and region.

Item: includes existing items in the stores.

Store: represents a reference of stores in each location and items associated to it by id.

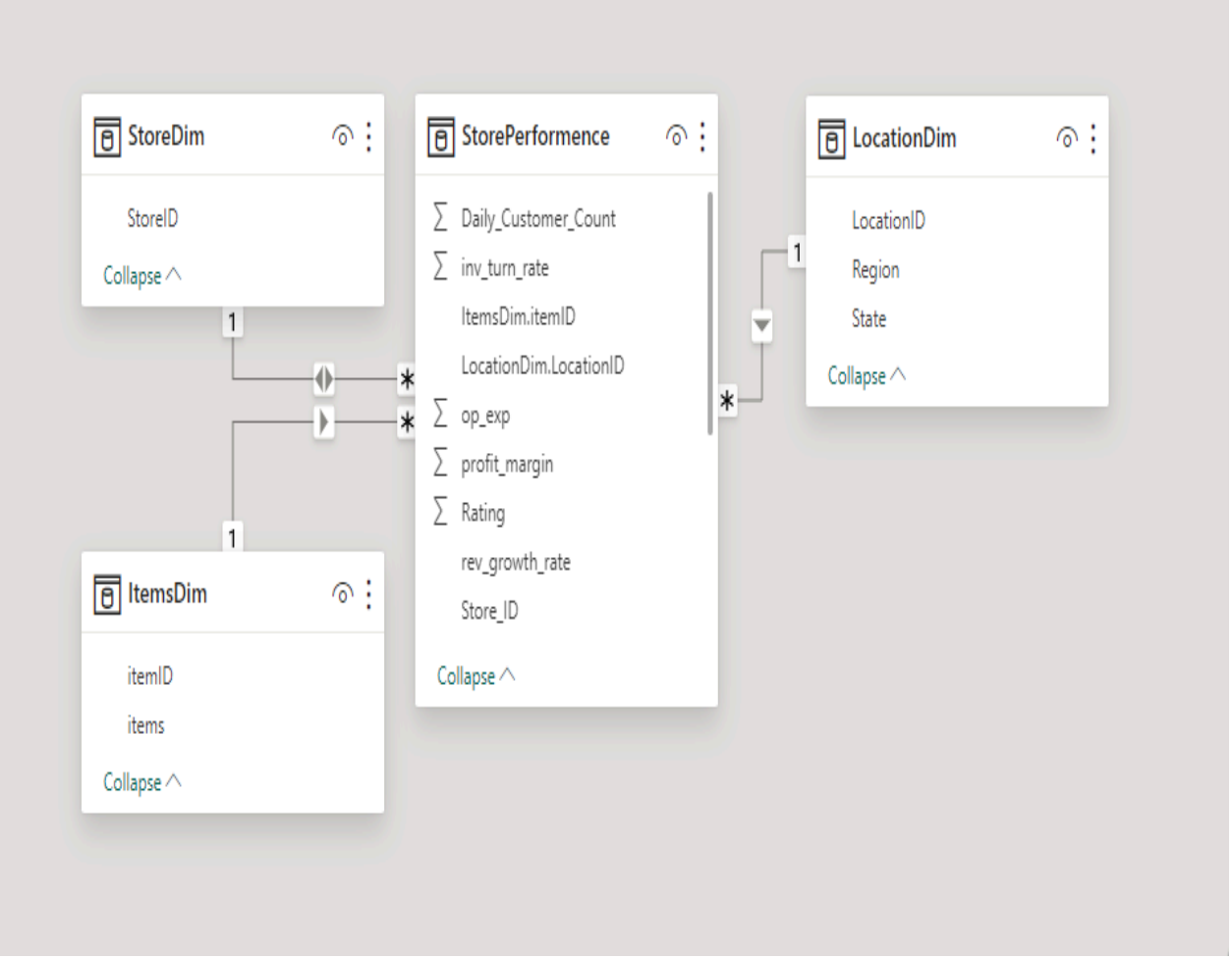

Figure 4: Star Schema

2.3.1 Fact

The fact in this data is the stores' performance in terms of top selling items, stores rating and overall sales.

2.3.1 Dimensions

Dimensions included in this data are:

- Location
- Items
- Stores

Eventually, we made the data warehouse schema to be a star schema because of the many to one relationship between stores performance and their location and items placed for selling.

2.4 Data Visualization

We used Microsoft Power BI as our OLAP and data visualization tool, We linked our warehouse from MS SQL Server to Power BI we used several metrics to understand the data and get several insights from it including:

- The most rated stores are located in the west and the least are located in the northeast.
- The average profit margin of the stores is 14%
- The average customer rating overall is 3.084
- "Chips", "Coffee", and "Slurpee" are the most sold items in most stores per region with 25 stores each and the least most sold items in stores are Pizza Slices, Energy Drinks
- Georgia and Kentucky are the states with the most sales revenue of $1.5M in total.
- The north region holds the highest growth rate of 0.11

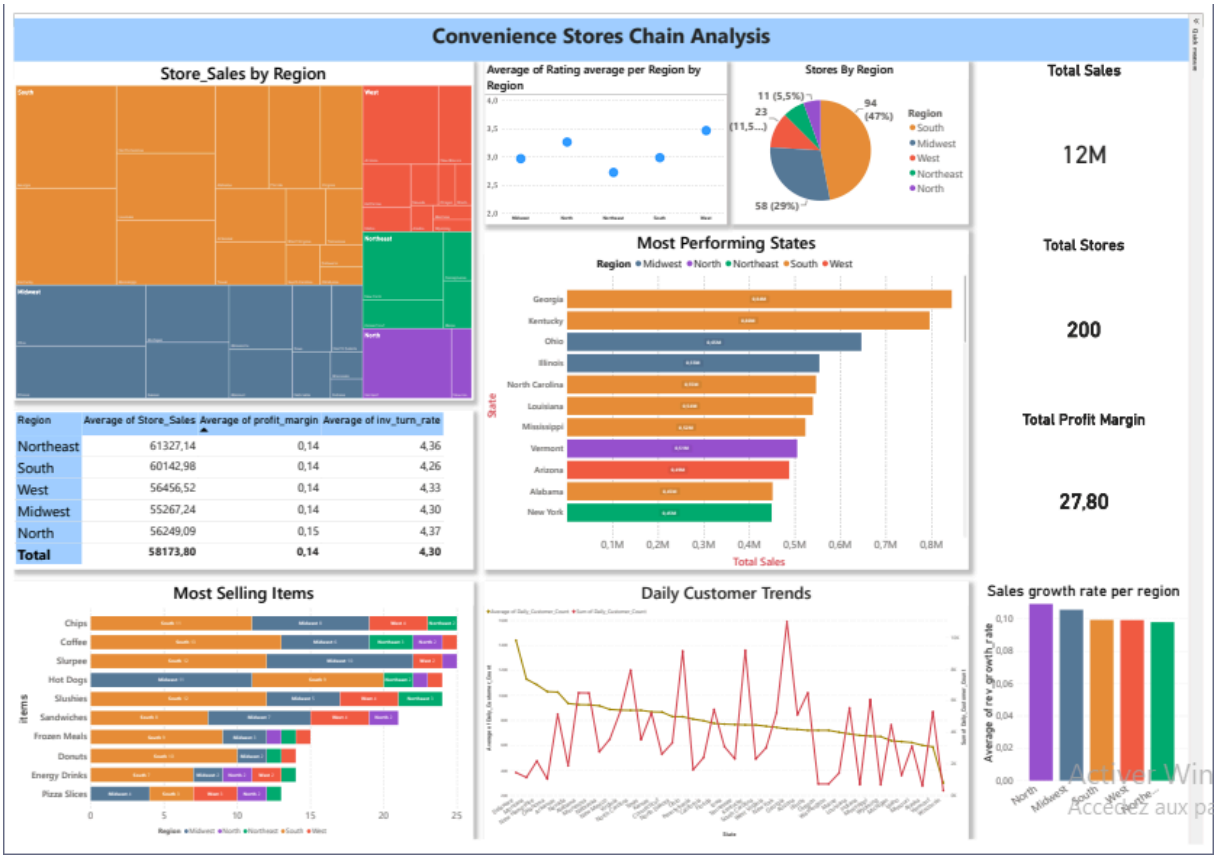

Figure 5: Dashboard

3 Conclusion

In conclusion, the analysis proved that the stores with the most sales are located in the south and to be more specific in Georgia, Kentucky and North Carolina which implies that more stores should be opened in those areas. However, we found that stores with the least sales are located in the north and northeast so it is not recommended to open new stores in those areas. Also, on the upcoming marketing campaigns the company should focus on discounting items that are not included in the list of most sold items in stores and should also capitalize the items that are most sold including chips and coffee by continually making sure that the inventory for them is enough for the

demand, promoting and offering bundle offers, and optimize pricing for higher profit.