

Business intelligence and Database Management System

Business Intelligence Research on Starbucks

Prepared by:

Rayen Rouaissi

Mohamed Aziz Belhadj

Achref Khairi

Maha Jdidi

Contents

1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2 Implementation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.1 Data Gathering . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.2 Data Preparation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.3 Data Storage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

2.3.1 Storage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

2.3.2 Fact . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.3.3 Dimensions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.4 Data Visualization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

3 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

1 Introduction  
This business intelligence project is focused on analyzing the characteristics of Starbucks customers and their ratings based on quality, service, and promotions.

This study aims to give insights into how different factors such as age, income, and gender can influence customer ratings.

After understanding the relationships and interactions between the facts and dimensions we succeeded in having a full view of preferences, which can help Starbucks improve their offerings and consequently drive more business.

Throughout this project, we aim to improve overall customer satisfaction and build better implementation strategies for Starbucks in the United States.

2 Implementation

2.1 Data Gathering  
We extracted the Starbucks stores’ and customers’ ratings basically from GitHub and Kaggle datasets. All the raw data is stored in the Raw Data folder.

2.2 Data Preparation  
For the data preparation, we used Python to manipulate data and configure it for the data warehouse. First, we started by converting all the data into one format (JSON) from different formats (CSV, Excel, JSON).

We employed the use of pandas’ library in Python and Talend to make necessary transformations and apply the ETL process. All the code employed is stored in the Transformations folder.

Then we moved to change. Firstly, we created a script that allows us to clean and manipulate the customer file because it contains some undesired columns, and we fixed their IDs, so they match with the payment method. Additionally, we created a transfrom\_income function that helped us to transform the format of average spent per visit to the dollar sign ($) since we are dealing with Starbucks stores based in the United States. Moreover, we created a new dataset called working hours where we calculated the total number of working hours per store in each day. For the map and sort, we used Talend. All the transformations are stored in the result folder. Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquementUne image contenant texte, Police, capture d’écran, blanc

Description générée automatiquementUne image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

Une image contenant texte, capture d’écran, Police

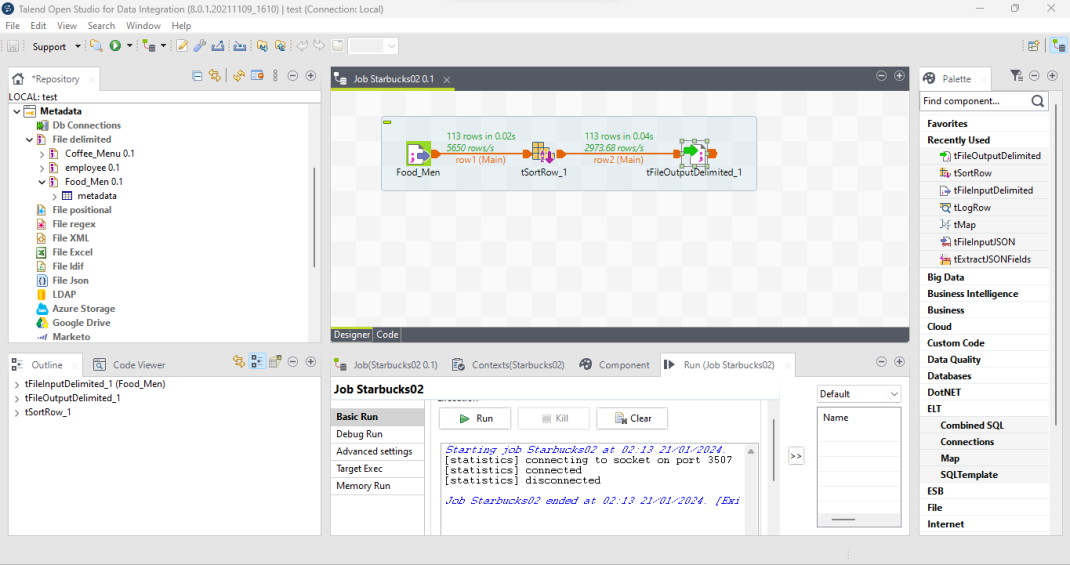
Description générée automatiquementUne image contenant texte, Police, capture d’écran

Description générée automatiquementUne image contenant texte, capture d’écran, Police, nombre

Description générée automatiquementUne image contenant texte, capture d’écran, Police, nombre

Description générée automatiquementUne image contenant texte, logiciel, Icône d’ordinateur, Logiciel multimédia

Description générée automatiquement



2.3 Data Storage  
2.3.1 Storage  
For data storage, we employed the use of SQL to insert and map tables in our code and PostgreSQL as the database management system.

We have 2 main tables in the database, Customers, and Stores. the other tables are Ratings, Menu Category, Employees, Working Hours, Customer Payment. Attached here is the code for the SQL.

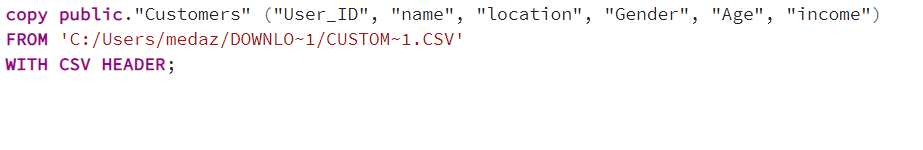
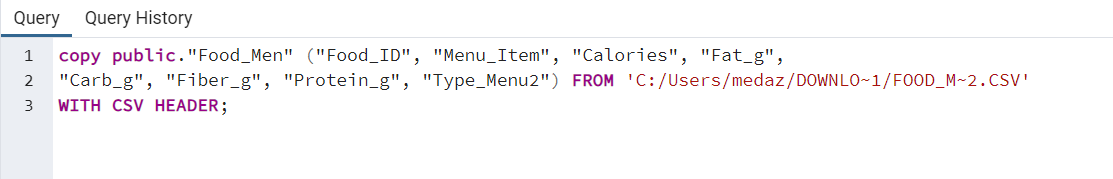
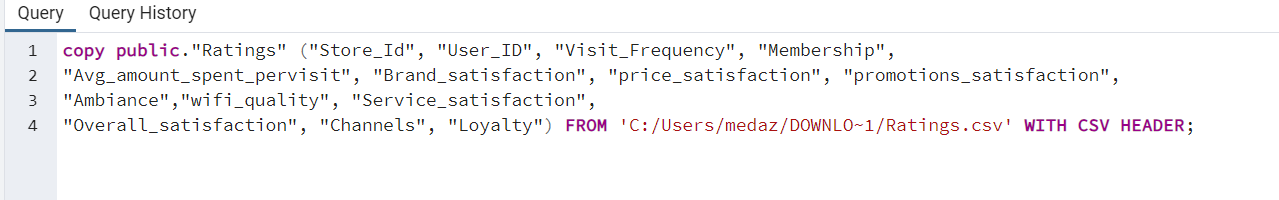
Une image contenant texte, capture d’écran, Police, document

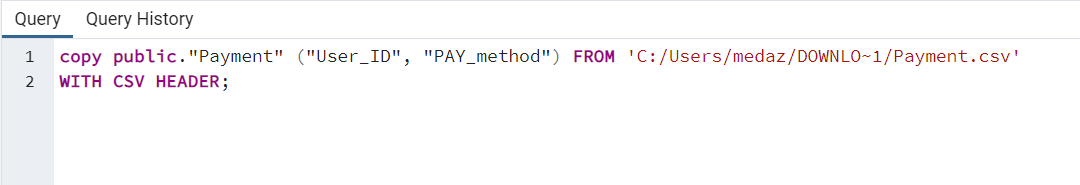
Description générée automatiquementUne image contenant texte, capture d’écran, Police

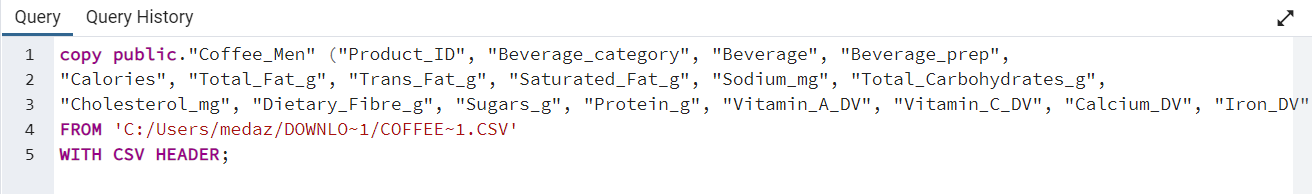
Description générée automatiquement

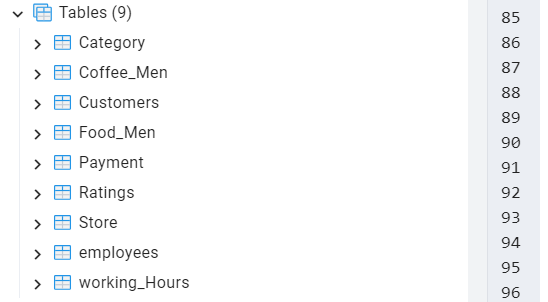


Une image contenant texte, Police, ligne, capture d’écran

Description générée automatiquementUne image contenant texte, Police, ligne, capture d’écran

Description générée automatiquementUne image contenant texte, Police, capture d’écran

Description générée automatiquement



* **Customers:** represents the customer information.
* **Store:** represents the store information (store ID, name, city, and country).
* **Ratings:** represents the service, promotions, Wi-Fi, and overall satisfaction. Additionally, the loyalty and membership of each customer to the store.
* **Category:**represents the type of menu of each store.
* **Coffee Menu:** represents the drinks available on the menu and their characteristics.
* **Food Menu:**represents the foods available on the menu and their characteristics.
* **Working Hours:** represents the number of working hours for each store in each day.
* **Employees:** represents the information of employees including their salary working in a given store.
* **Payment:** represents the payment method desired for each customer.

Une image contenant texte, capture d’écran, Rectangle, conception

Description générée automatiquement

2.3.2 Fact  
The fact in this data is the ratings given by the customer to each store. The types of ratings include:  
**• Ambiance Rating  
• Brand Rating  
• price Rating**

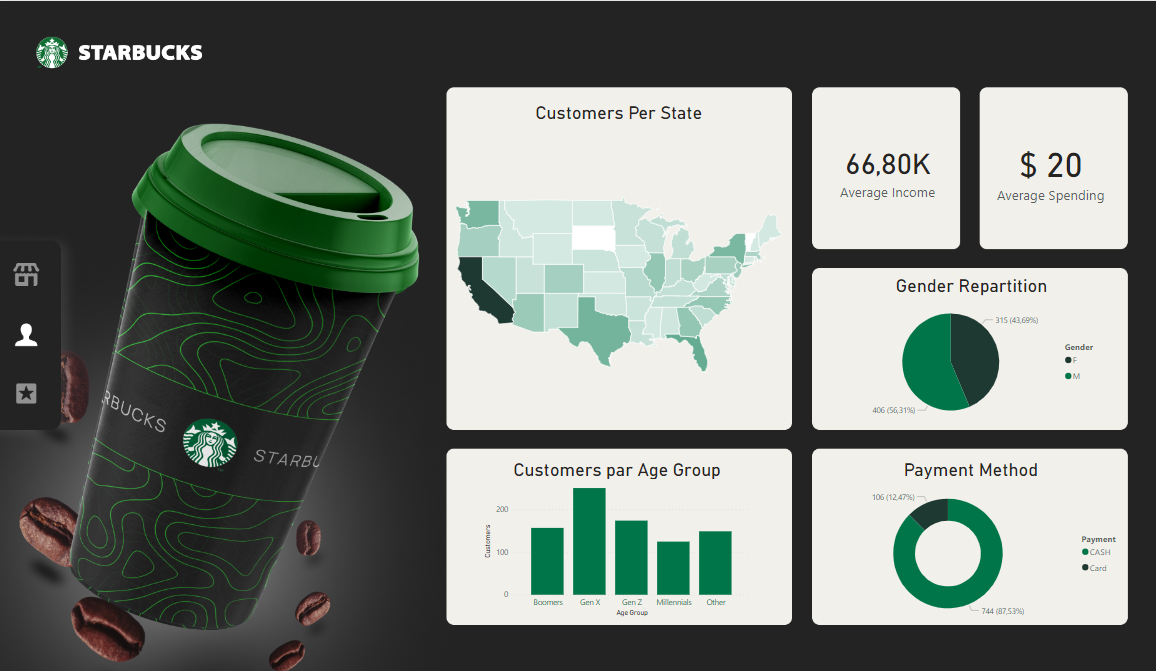
**• Wifi Rating  
• Promotions Rating**

**• Overall Rating**

2.3.3 Dimensions  
Dimensions included in this data are:  
**• Store  
• Customer**  
Dimensions that are derived from the customer dimensions include:  
**• Payment Method**Dimensions that are derived from the restaurant’s dimensions include:  
**• Working Days and hours  
• Menu type.  
• Employees**

Finaly, we made the Snowflake Schema as Data Warehouse Schema because of one-to-many relationships between customers’ and stores’ values in the customer.

2.4 Data Visualization  
For the Data visualization, we used power bi to understand customer behaviours and their satisfaction about the offerings of Starbucks. It also helped us to get insights on some points and channels that Starbucks can work on to attract more customers and serve them better.



• Most customers in the dataset live in California.

• The average income of customers is 66.800$ and it ranges from 30 to 120K.

• Most of the customers are males.

• Most of the customers belong to Generation X and Z.

• Most of the customers use cash as a payment method.

• Most customers are satisfied with service, ambiance, and Brand.

• Customers are fairly satisfied with the price and Wi-Fi.

• Each customer spends 20$ on average for each visit.

• 77% of current customers are satisfied with Starbucks and consider staying loyal in the future.

• 49.18% of current customers have a membership card.

• The average monthly salary for employees is 6180$.

• Most of the employees are in the low- and medium-income category.

3 Conclusion

In conclusion, the research proved that most of the customers in the dataset are living in California. They are mostly males who have a medium level of income and belong to generation X and Z. Most of them are satisfied with what Starbucks is offering and consider staying loyal to the brand in the future. Wi-Fi and price could be improved to meet the customers’ needs.