Assessment Item 2: Report - Design business intelligence system and data warehouse.

Contents

[Table of Figures 3](#_Toc177568290)

[Introduction 4](#_Toc177568291)

[BI Architecture 5](#_Toc177568292)

[Data Sources 5](#_Toc177568293)

[Sales Data 5](#_Toc177568294)

[Customer Information 5](#_Toc177568295)

[Product Lines 6](#_Toc177568296)

[Customer Ratings 6](#_Toc177568297)

[Payment Methods 6](#_Toc177568298)

[ETL Process: 6](#_Toc177568299)

[Extraction 6](#_Toc177568300)

[Transformation 6](#_Toc177568301)

[Loading 7](#_Toc177568302)

[Data Storage 7](#_Toc177568303)

[Cloud-Based Storage 7](#_Toc177568304)

[Fact and Dimension Tables 7](#_Toc177568305)

[Data Access & Reporting 8](#_Toc177568306)

[Visualization Tools 8](#_Toc177568307)

[Real-Time Reporting 8](#_Toc177568308)

[Ad-Hoc Analysis: 8](#_Toc177568309)

[Security & Permissions 9](#_Toc177568310)

[User Access Control 9](#_Toc177568311)

[Data Encryption 9](#_Toc177568312)

[Data Privacy 9](#_Toc177568313)

[Dashboard 10](#_Toc177568314)

[Overall Overview Page 10](#_Toc177568315)

[Sales Overview Page 11](#_Toc177568316)

[Day of the Week Performance page 11](#_Toc177568317)

[Product Performance Page 12](#_Toc177568318)

[Profit analysis page 12](#_Toc177568319)

[Insights 13](#_Toc177568320)

[Branch Performance 13](#_Toc177568321)

[Top Selling Products 13](#_Toc177568322)

[Sales Trends 13](#_Toc177568323)

[Customer Type 13](#_Toc177568324)

[Recommendations 14](#_Toc177568325)

[Business Performance Recommendations 14](#_Toc177568326)

[Branch-Specific Strategies 14](#_Toc177568327)

[Market and Seasonal Adaptations 14](#_Toc177568328)

[Optimizing Product Offerings 14](#_Toc177568329)

[Customer Engagement and Loyalty Programs 14](#_Toc177568330)

[Business Intelligence System Recommendations 14](#_Toc177568331)

[Demographic Information 14](#_Toc177568332)

[Customer Loyalty Metrics 15](#_Toc177568333)

[Geolocation Data 15](#_Toc177568334)

[Promotion and Discount Information 15](#_Toc177568335)

[Customer Feedback 15](#_Toc177568336)

[Product-Specific Data 15](#_Toc177568337)

[Sales Channel 15](#_Toc177568338)

[Return Data 15](#_Toc177568339)

[Conclusion 16](#_Toc177568340)

# Table of Figures

[Figure 1 Overall Overview Page 10](#_Toc177568341)

[Figure 2 Sales Overview Page 11](#_Toc177568342)

[Figure 3 Day of the Week Performance page 11](#_Toc177568343)

[Figure 4 Product Performance Page 12](#_Toc177568344)

[Figure 5 Profit analysis page 12](#_Toc177568345)

# Introduction

Currently, with the stiff competition that has thrived in the retail industry, players depend on analyzing data in order to come up with wanted decisions, as well as managing the companies efficiently. The creation of a BI system that will integrate all aspects of business is vital in taking raw data and converting it into useful information for businesses to make changes fast in response to the market trends, customers and other problems are common in businesses. This type of assessment is designed to cover an approach to building the BI system and data warehouse structure for a retail store with the data gathered in its several branches to consider sales results, customers’ behavior, and product positions.

Based on the lessons learned from the first part of the project involving the retail dataset analysis, another goal of this project is to develop an efficient BI solution that would facilitate real-time data collection and improve the decision-making process and moreover, provide a solid platform for developing application-level visuals. The system will majorly feature a data warehouse to integrate and store the consolidated information from different branches’ sales, products, and customers and enable developing dashboards with the help of tools such as Power BI. The BI architecture and a good data warehouse system will help the retail business in the identification of growth prospects, customization of the products and services to fit the market and increase the level of customer satisfaction.

This report identifies the foundational elements of the BI system which include data warehouse framework, real-time BI architecture, and visual analysis. It also indicates recommendations on the ways by which the manner of collecting data can be enhanced in the future and the different ways by which the efficiency of the retail store can be improved.

# BI Architecture

They provide real time information about the performance of the retail store through Business Intelligence (BI) architecture that intends to cover data collection processing and reporting techniques. This system utilizes a cloud-based database at the center to accumulate timely information about multiple store branches and helps in the further analysis with the help of tools like power BI. The architecture includes the following components:The architecture includes the following components:

## Data Sources

The BI system receives data from several branches of the store in real-time manner. Each store’s operational database feeds into the central system with key data categories such as:Each store’s operational database feeds into the central system with key data categories such as:

### Sales Data

Invoice number, Product name, Quantity, Unit price, Total amount of sales. This data will give information about the store operations, the products that are selling most, and revenue generated.

### Customer Information

Those are age information, gender, and their membership status within a given company. This will be helpful in targeting customers for a more effective marketing strategy or in socio behavior analysis.

### Product Lines

According to which products are grouped into certain categories like ‘Food & Beverages’, Electronics, Clothing etc so that sales can be analyzed at the product category level.

### Customer Ratings

Responses collected on the amount of satisfaction the customers have towards the service and the products offered which are based on a rating of 1-10.

### Payment Methods

Marketing and account records of the transactions stating whether the transaction was through credit cards, eWallets or cash. This will enable the examination of patterns of payments and clients’ preferences.

For collecting real-time data from these sources, **Kafka** or **Kinesis** (from amazon web service) can be used. These tools enable data stream from various stores’ operational systems including point of sale terminals, customer relationship management systems to the central cloud database allowing for continuous feed of data without breaks.

## ETL Process:

ETL: Extract, Transform, Load is the most critical and basic process of changing data from the real store format for suitable analysis. Another advantage of the real-time ETL process is that data is always available; in other words, data is ready for reporting and analysis.

### Extraction

Apache NiFi or AWS Glue will be used in this process for getting rid of the manual extraction of the data. These tools are designed for real-time data harvesting capacity, so that, data feeds in through the operational systems, like POS, CRM etc that are located at every store to the central cloud database. Information will be pulled from these systems perpetually to accrual transactional, customer and product information as it is produced.

### Transformation

Either way, Talend or Azure Data Factory will handle the transformation step. The transformation process includes:

#### Data Cleaning

Dealing with incomplete or missing values, for example missing customer’s demographic data or different definitions of products.

#### Standardization

This involves making sure that data brought into the database are in the proper format such as transforming one date format to another (for instance transforming a date format of DD-MM-YYYY to YYYY-MM-DD) and that the currency used has a constant format.

#### Normalization

Joining data where necessary, such as stacking sales data in form of a database having several branches’ data in a single file format for analysis. For example, making sure that product names and categories are in line with one another concerning the name standard.

#### Validation

Data sanitization or data cleansing where the objective is to enhance the quality of the data received so that it got sampled in a way that it complies with set standards such as the acceptable payment methods or product codes.

### Loading

The cleaning and transformation done to the data is then uploaded to the cloud data warehouse by use of platforms such as Amazon Redshift or Google big query. These platforms are ideal for raw data storage as well as Massive scale data handling as it includes real-time data ingestion. The data warehouse architecture accentuates on how easily query can be responded to as the data is sorted into fact and dimension tables. Star Schema design should be utilized to implement the system since fact tables hold transactional information (for instance, sales data) while, dimension tables hold details attributes (for instance, customer specifics, or products categories).

## Data Storage

There is the central cloud-based database that fulfils the role of the central database for the BI system.

### Cloud-Based Storage

Some of the popular Data Warehouse as a Service to be used are Amazon Redshift, Google BigQuery or Microsoft Azure SQL Data Warehouse to host and manage large datasets from the various stores. These platforms are highly scalable, available and can query results in a very short period of time.

### Fact and Dimension Tables

In support of some of these models, the data warehouse will be formed using Star Schema or Snowflake schema.

#### Fact Tables

It is these tables that will store reoccuring data of such a business, for instance, sales records, number of products sold, and total income accrued.

#### Dimension Tables

As for these tables, they will be containing some contextual information including the customer’s details, store locations, products offered in the store, and the options for payment.

Storing the data in the cloud will help the system to be real-time, which is paramount important in making sure real-time reports are compiled and analyzed.

## Data Access & Reporting

BI system will be integrated with tools such as Power BI and Tableau as these tools will be the main tools used in data access, data visualization and reporting.

### Visualization Tools

The other tool that will be used is Power BI for creating engaging and interactive dashboards as well as visualization. It will provide the users with ways to navigate and interact with the data, which include the use of graphical user interfaces like those of sale force ;for real time analysis of performance, customer, and product trends. Visualizations may include:

* Sales pitch within a given duration for instance the number of sales made within a given period such as a day, week or a month.
* Performance of products by category wise or store wise.
* Analysis of customers characteristics distribution according to demographic information.

### Real-Time Reporting

That is because the data is to be stored in an environment hosted cloud with real-time updates, which means users would get instant and updated reports, and dashboard. DirectQuery feature in the Power BI will enable the tool to take data directly from the cloud database rather than retrieving it in the initial analysis and then refreshing the report daily.

### Ad-Hoc Analysis:

It is also possible to do chad analysis using Excel from the cloud database thus giving users more flexibility in reporting and data analysis.

## Security & Permissions

It is important especially when it comes to access control specifically ensuring that data is well secured.

### User Access Control

Such tools as Azure Active Directory or AWS IAM will be applied in order to handle roles and permissions of the user. These tools allow access to persons who are allowed access to specific data and materials only. Speaking about access rights, they will vary depending on the customer’s status, being a store manager, regional manager or a member of the board.

### Data Encryption

Any information that may be transferred from stores to the cloud and also between the cloud database and Power BI shall be encryptedREDIT through the SSL/TLS protocols.

### Data Privacy

In a bid to protect user data and meeting the GDPR or CCPA requirements, the customer data will be masked or anonymized where and when required. For example those fields that require extra protection for instance, customer’s name and contact details will be stored in encrypted form.

# Dashboard

## Overall Overview Page

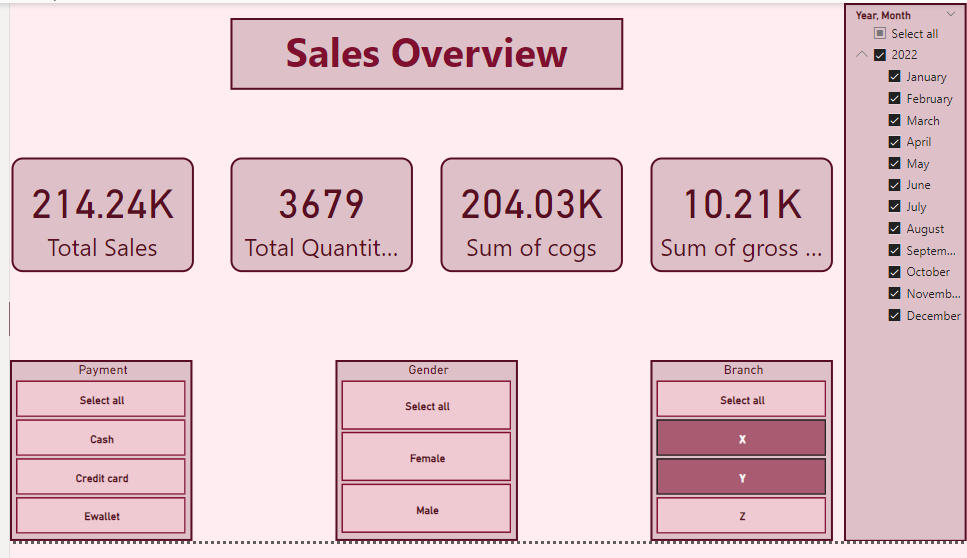


Figure 1 Overall Overview Page

## Sales Overview Page

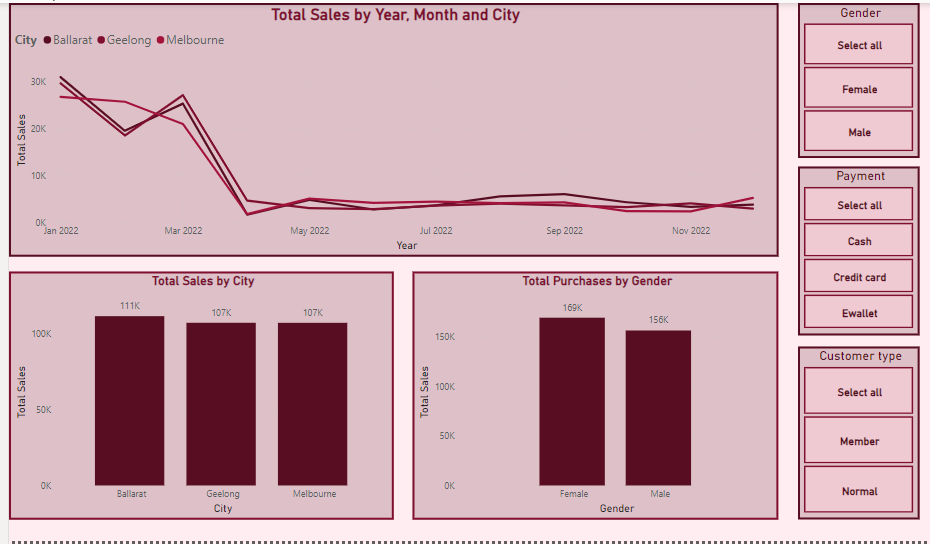


Figure 2 Sales Overview Page

## Day of the Week Performance page

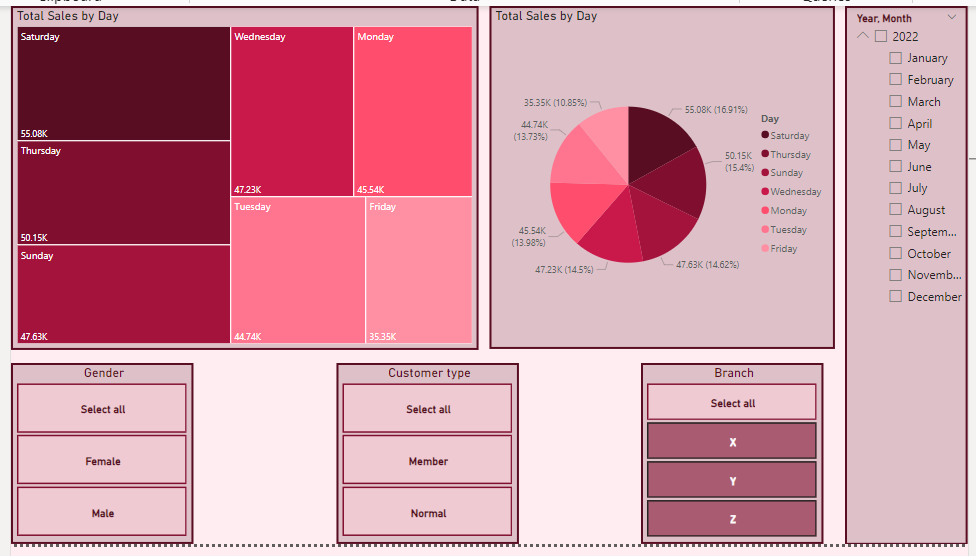


Figure 3 Day of the Week Performance page

## Product Performance Page

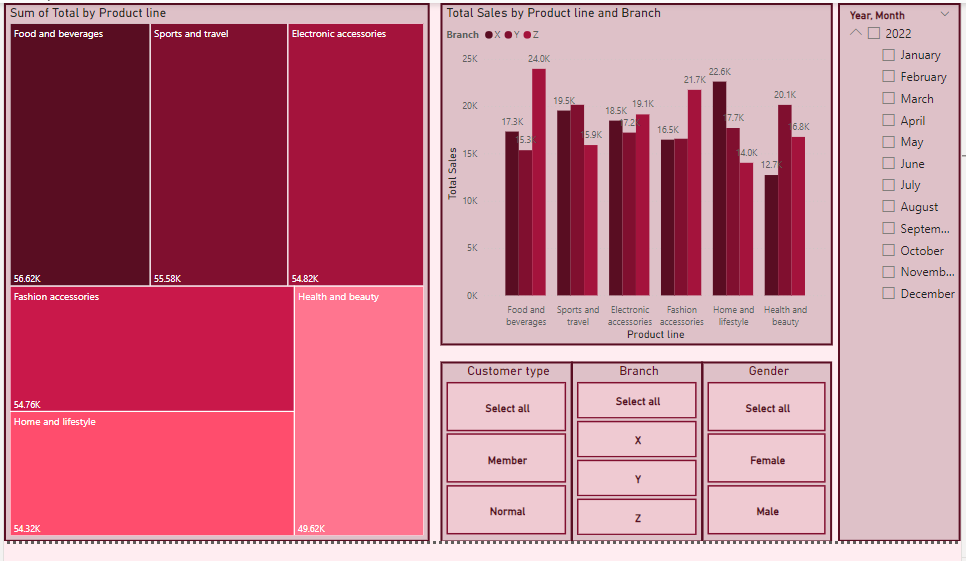


Figure 4 Product Performance Page

## Profit analysis page

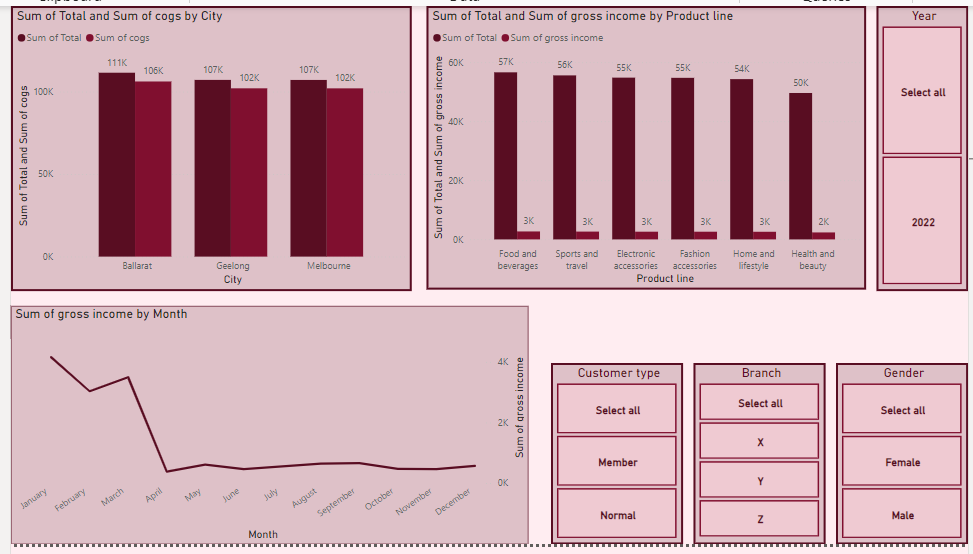


Figure 5 Profit analysis page

# Insights

## Branch Performance

In total sales, Branch Z had the highest recorded performance all through the year and especially in August and September On the other hand, all the branches faced the set back of reduced sales in the second quarter of the year. Changes in trends could be tackled, especially if one has to formulate promotions and discounts based on the seasons.

## Top Selling Products

While in Branch X it was possible to identify a clear specialty in “Home and lifestyle,” in Branch Y it was “Electronic accessories,” and in Branch Z it was “Food and beverages.” The results indicate the necessity of branch-specific product orientations.

## Sales Trends

The sales records further highlighted that certain months had higher amounts of sale as compared to other months and out of operating days of the week, the weekends recorded the highest number of sales across various branches of the organization. There are thus opportunities when the sales might seem low but focusing on weekends or avoiding the low performing weekdays can improve traffic.

## Customer Type

They read in the statistics that members contributed more into sales than non-members to justify the need to have a loyalty program.

It is possible to note that the improvement of these aspects can bring a higher level of sales and customer experience in all branches.

# Recommendations

## Business Performance Recommendations

### Branch-Specific Strategies

* Branch X: On the Home and Lifestyle category, it is recommended that this segment should be marketed most in the early part of the year while using seasonal selling promotions to overcome declining sales.
* Branch Y: Maximize the periodicity promotion of “Electronic Accessories” in mid-year and make the sales fluctuations in May and December less significant.
* Branch Z: More emphasis should be put on the “Over‐the‐counter‐sales” especially during the low seasons; introduce “Bags and Wallets” under “Fashion Accessories”.

### Market and Seasonal Adaptations

* Depending on such variations, it is recommended to launch off-peak promotions, for example, Black Friday.
* Both Branches X and Z should be focused to Weekend promotions while weekday promotions should be targeted towards Branch Y.

### Optimizing Product Offerings

* Add on to the high-selling categories of the store (Electronics Accessories, Food and Beverages) and combine products together to offer sale offers.
* Sustain unstable sales volume in the shifting subcategories such as “Fashion Accessories,” ”Health and Beauty” through promotional strategies.

### Customer Engagement and Loyalty Programs

* Customer Engagement and Loyalty are some of the vital segments that should be highlighted.
* Build on the concepts of buyer loyalty by providing special discounts to only those customers who are enrolled with membership programs and adopt techniques that can encourage other customers to subscribe to such membership programs.

## Business Intelligence System Recommendations

### Demographic Information

Finally, it may be useful to add customer age and income level as to target the individual customer with the most suitable marketing options as well as product recommendations.

### Customer Loyalty Metrics

Guarantee that membership duration and different levels of repeat purchase are good loyalty segmentation indicators.

### Geolocation Data

Collect the postal codes so that you can get further broken-down regional sales data.

### Promotion and Discount Information

Make a check on whether or not discounts were used and the type of promotion that was used for further analysis on the sales.

### Customer Feedback

Collect qualitative data from customers in order to increase customer satisfaction and upgrade service quality.

### Product-Specific Data

Include bar codes for improved stock control and numbers or codes for the suppliers for assessment purposes.

### Sales Channel

To have channel specifics, distinguish between online and in-store for the evaluation.

### Return Data

Collect and document product return status and the causes, to evaluate product quality and customers’ levels of satisfaction.

# Conclusion

The given report is to present the business intelligence system and the concept of data warehouse that can be helpful in enhancing the involved decision making in the retail store. In the context of BI, the proposed architecture focuses on how to obtain real-time data and analyze them to respond to market demands and customers’ behaviors quickly. To enforce the compliance of different branches in processing and inputting data, the tools used are cloud based storages, ETL processes, power BI, and many more.

The design of the fact and dimension tables and also use of and visualization tools provides the store with useful information on sales data, customer characteristic and product characteristics. The data access limitation and using encrypted protocols make it possible to align the system with some of the privacy laws and protect the customer’s details. The features that allow for the real-time reporting also improve the store’s capacity to act based on the most current information available.

Furthermore, recommendations on branch strategies, customer interaction, and loyalty programme are integrated as the findings from data analysis. When these strategies are applied, fruitful results will be obtained such as efficient branch performance, increased customer retention and great product output to fit the market. As previously recommended integrating the demographic details, and geolocation details will also help to strengthen the targeted marketing as well as increasing the chances of attaining higher sales.

In conclusion, the BI system and the data warehouse framework are capable of enhancing operation efficiency, customer satisfaction, and hence increase the sale of the retail store. It allows the store to have real-time data and employ higher levels of analytics that will enable it to respond to the competition and changes in the market in order to create personal connections with its clientele.