**Data Sources**

**Raw Data Files (Original Format)**

The raw data files used in this project were uploaded to Snowflake under the **coffee\_shop\_stage** stage. The files include:

1. sales\_receipts.csv – Contains transactional sales data.
2. customer.csv – Contains customer details.
3. product.csv – Contains product information.
4. sales\_outlet.csv – Contains store details.
5. staff.csv – Contains staff details.
6. Dates.csv – Contains date-related information.

**Documentation of Data Sources & Relationships**

Each data source contributes to a specific aspect of the **dimensional model**, and they are linked as follows:

|  |  |  |
| --- | --- | --- |
| **Data Source** | **Purpose** | **Maps To** |
| sales\_receipts.csv | Sales transactions | fact\_sales (fact table) |
| customer.csv | Customer demographics | dim\_customer (dimension) |
| product.csv | Product details | dim\_product (dimension) |
| sales\_outlet.csv | Store information | dim\_store (dimension) |
| staff.csv | Employee information | dim\_staff (dimension) |
| Dates.csv | Calendar dates | dim\_date (dimension) |

**Key Relationships:**

* fact\_sales references **customers**, **products**, **stores**, **staff**, and **dates** through **foreign keys**.
* Each **dimension table** stores detailed information about its respective entity.

**Data Dictionary**

A **data dictionary** provides details on each field within our sources.

**1. sales\_receipts.csv (Transactional Data)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Description** | **Mapped Table** | **Mapped Column** |
| transaction\_id | Unique identifier for a sale | fact\_sales | transaction\_id |
| transaction\_date | Date of sale | fact\_sales | date\_id |
| transaction\_time | Time of transaction | Not stored | - |
| sales\_outlet\_id | Store ID | fact\_sales | store\_id |
| staff\_id | Staff who handled transaction | fact\_sales | staff\_id |
| customer\_id | Customer involved in transaction | fact\_sales | customer\_id |
| instore\_yn | Whether transaction was in-store (Y/N) | Not stored | - |
| order | Order number | Not stored | - |
| line\_item\_id | Line item identifier | Not stored | - |
| product\_id | Product purchased | fact\_sales | product\_id |
| quantity | Quantity purchased | fact\_sales | quantity\_sold |
| line\_item\_amount | Total sales amount | fact\_sales | total\_sales\_amount |
| unit\_price | Unit price of product | Not stored | - |
| promo\_item\_yn | Whether the item was a promo item | Not stored | - |

**2. customer.csv (Customer Data)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Mapped Table** |
| customer\_id | Unique customer identifier | dim\_customer |
| first\_name | First name of customer | dim\_customer |
| last\_name | Last name of customer | dim\_customer |
| email | Email address | dim\_customer |
| phone\_number | Contact number | dim\_customer |
| loyalty\_member\_yn | Loyalty program membership (Y/N) | dim\_customer |
| birth\_date | Date of birth | dim\_customer |
| gender | Gender | dim\_customer |

**3. product.csv (Product Data)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Mapped Table** |
| product\_id | Unique product identifier | dim\_product |
| product\_group | Product category (e.g., Beverages) | dim\_product |
| product\_category | Subcategory within product group | dim\_product |
| product\_type | Type of product | dim\_product |
| product | Product name | dim\_product |
| product\_description | Description of product | dim\_product |
| unit\_of\_measure | Measurement unit (e.g., oz, single) | dim\_product |
| current\_wholesale\_price | Wholesale price of product | dim\_product |
| current\_retail\_price | Retail price of product | dim\_product |
| tax\_exempt\_yn | Whether tax applies (Y/N) | dim\_product |
| promo\_yn | Whether product is on promo (Y/N) | dim\_product |
| new\_product\_yn | Whether product is newly introduced (Y/N) | dim\_product |

**4. sales\_outlet.csv (Store Data)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Mapped Table** |
| store\_id | Unique store ID | dim\_store |
| store\_name | Name of the store | dim\_store |
| location | Store location | dim\_store |
| store\_manager | Manager’s name | dim\_store |

**5. staff.csv (Staff Data)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Mapped Table** |
| staff\_id | Unique staff identifier | dim\_staff |
| first\_name | First name of staff | dim\_staff |
| last\_name | Last name of staff | dim\_staff |
| position | Job title | dim\_staff |
| start\_date | Employment start date | dim\_staff |
| assigned\_store\_id | Assigned store ID | dim\_staff |

**6. Dates.csv (Date Information)**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Mapped Table** |
| full\_date | Actual date | dim\_date |
| year | Year value | dim\_date |
| month | Month value | dim\_date |
| day | Day value | dim\_date |
| weekday | Name of the weekday | dim\_date |