**Data Warehouse Report**

**Prepared by:** Ahmed Al-Yazouri, Ahmed Al-Khatib, and Fares AL-Namla , Anas Al-Shaqra

**Overview**

This report provides a comprehensive overview of the data warehousing process, including **data ingestion**, **transformation**, **modeling**, and **statistical analysis**. The data was stored and processed using **PostgreSQL**, and the statistics were generated from various sales and customer datasets.

**1. Data Ingestion**

**Script:** *Ingestion.py*

The data ingestion process involved loading multiple **CSV files** into the **PostgreSQL** database. The files included sales data from **2015 to 2017**, customer details, product categories, and more.

**Key Steps:**

* **Loading DB Configuration:** The database credentials were loaded from a **JSON** configuration file.
* **File Ingestion:** CSV files were read and ingested into respective tables in the PostgreSQL database. Encoding issues were handled by switching between **UTF-8** and **ISO-8859-1**.

**Ingested Files:**

* **Sales Data** (2015, 2016, 2017)
* **Territories**
* **Calendar**
* **Customers**
* **Products**

**2. Data Transformation**

**Script:** *transformation.py*

The data transformation phase ensured that the ingested data was **clean** and **consistent**.

**Process Details:**

* **Products Data:**
  + Removed **unnamed columns**.
  + Ensured essential columns like **ProductKey** and **ProductSubcategoryKey** were present.
  + Saved cleaned data into the **'products'** table and as a **CSV**.
* **Sales Data:**
  + Combined sales data from **2015, 2016, and 2017**.
  + Removed rows with missing essential fields.
  + Converted **OrderDate** to datetime format.
  + Saved the cleaned data into the **'cleaned\_sales'** table and as a **CSV**.
* **Customers Data:**
  + Filled missing **Gender** values with the most common gender.
  + Filled missing **Prefix** values based on gender.
  + Saved the cleaned data into the **'customers'** table and as a **CSV**.

**3. Data Modeling**

**Script:** *Modeling.py*

Data modeling involved **creating** and **populating** the data warehouse schema.

**Key Activities:**

* **Schema Creation:** Read and executed SQL schema from the **'dw\_schema.sql'** file.
* **Database Connection:** Established connection using **SQLAlchemy** with PostgreSQL.
* **Schema Population:** Dropped, recreated, and populated the data warehouse schema successfully.

**4. Statistical Analysis**

**Script:** *Stats.py*

The statistical analysis was performed to derive **insights** from the cleaned data.

**Key Insights:**

* **Total Sales:** $24,914,586.82
* **Unique Customers:** 17,416

**Top 5 Products by Sales:**

| **Product Name** | **Sales Count** |
| --- | --- |
| Water Bottle - 30 oz. | 3,983 |
| Patch Kit/8 Patches | 2,952 |
| Mountain Tire Tube | 2,846 |
| Road Tire Tube | 2,173 |
| Sport-100 Helmet, Red | 2,099 |

**Top 5 Regions by Revenue:**

| **Territory Name** | **Total Revenue** |
| --- | --- |
| Australia, Australia | 7,416,460 |
| Southwest, United States | 4,822,790 |
| Northwest, United States | 3,095,070 |
| United Kingdom, United Kingdom | 2,902,560 |
| Germany, Germany | 2,524,680 |

**Total Sales by Year:**

| **Year** | **Total Sales** |
| --- | --- |
| 2015 | 6,404,930 |
| 2016 | 9,324,200 |
| 2017 | 9,185,450 |

**Sales by Weekday:**

| **Weekday** | **Total Sales** |
| --- | --- |
| Wednesday | 3,660,070 |
| Tuesday | 3,581,060 |
| Friday | 3,574,590 |
| Sunday | 3,469,720 |
| Monday | 3,625,050 |
| Thursday | 3,484,860 |
| Saturday | 3,519,250 |

**Conclusion**

The data warehousing project successfully **ingested**, **transformed**, and **analyzed** sales and customer data, providing valuable insights into **product performance** and **regional sales trends**. The robust data pipeline ensures **data consistency** and **reliability**, supporting strategic business decisions.

**Prepared by:**

* Ahmed Al-Yazouri
* Ahmed Al-Khatib
* Fares AL-Namla
* Anas Al-Shaqra