Digital Egypt Pioneers Initiative – DEPI

Graduation Project

Supply Chain Dataset Analysis

Group Code:

RS-2368 ALX1_DAT1_G3e Data Analyst Specialist 3

Supervised by:

Mayar Ayman

Submitted by:

Emad Maarouf Abd Elfattah

Tarek Ibrahim Osman

Hossam Mohamed Abd Elfattah

Shehab El-Din Mohamed Ramadan

May Gamaleldin Mohamed

Hania Said Zakarya

October 2024

Table of Contents

Part One:

Introduction

1.1. Introduction:

This Supply Chain dataset belongs to a Fashion and Beauty startup business. The dataset is based on a supply chain of haircare, skincare and cosmetic products. It aims to deliver the products after manufacturing them in the needed quality to ensure that the consumer gets the value they were promised.

1.2. Objective:

The main objective of this dataset is to support the startup business with valuable information and insights to help with its growth and improvement.

Part Two:

Data Preprocessing

Data Preprocessing is an essential initial step of data preparation, where data is cleaned then transformed to be prepared for analysis.

2.1. Data Exploration:

During the exploration phase, all data types were checked using python, the headers were defined, descriptive statistics were done, as well as boxplots to check for any outliers.

2.2. Data Cleaning:

This step ensures that the data is free of errors and inaccuracies through thorough exploration. In this dataset, we scanned for the following:

2.2.1. Missing Data:

This dataset was already free of any missing data, and it only contains 100 records.

2.2.2. Redundancy:

There were no redundant records found in this dataset.

2.2.3. Outliers:

There were no extreme values found in this dataset.

2.2. Data Transformation:

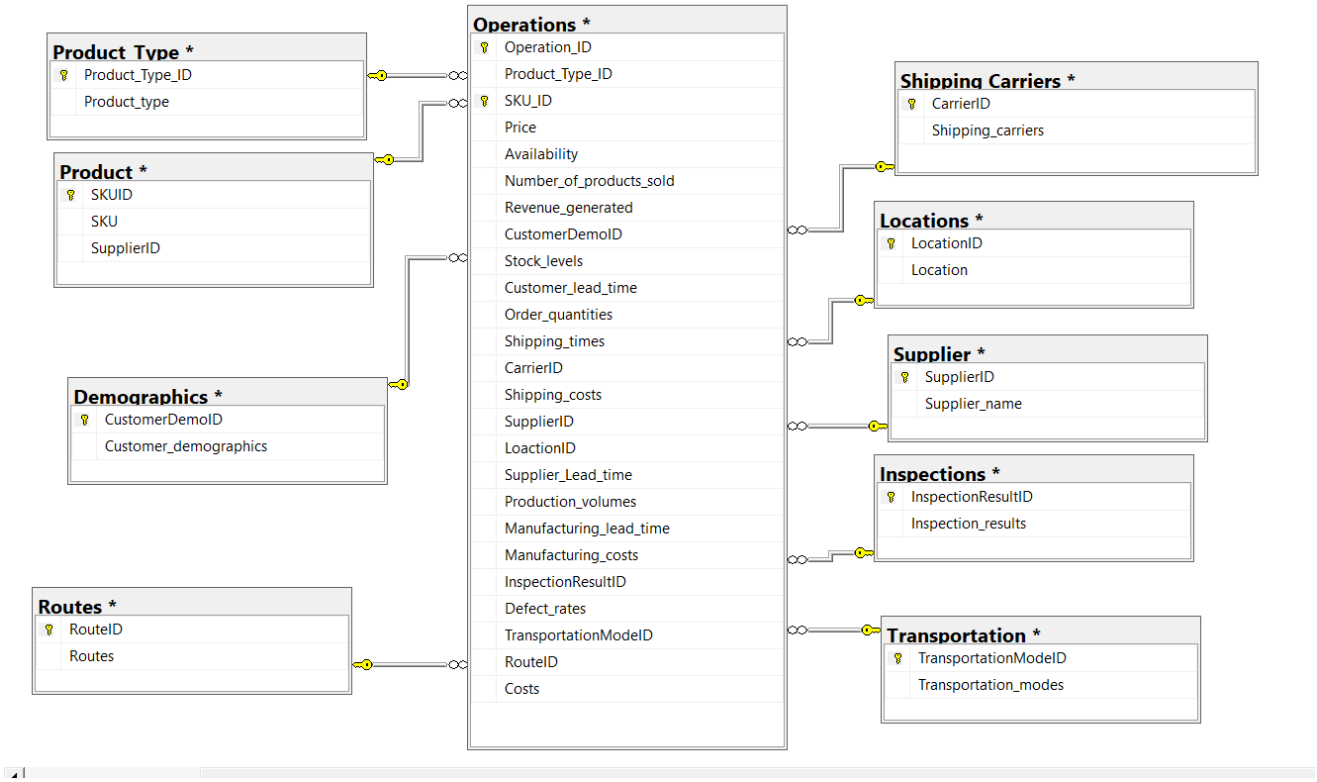
As per the requirements, this dataset had to be prepared to be put in a database. Nine dimension tables were extracted from the fact table, then all of these tables' extensions were converted into CSV. Then these files were imported to a DBMS for the database to be created and for further modelling.

Part Three:

Data Modelling

3.1. Database Diagram (Schema):

Each one of these entities has their own primary key. In the fact table, however, we have a composite key composed of operation_id and SKU_ID. The following diagram was made by SQL Server Management Studio:



Part Four:

SQL Queries

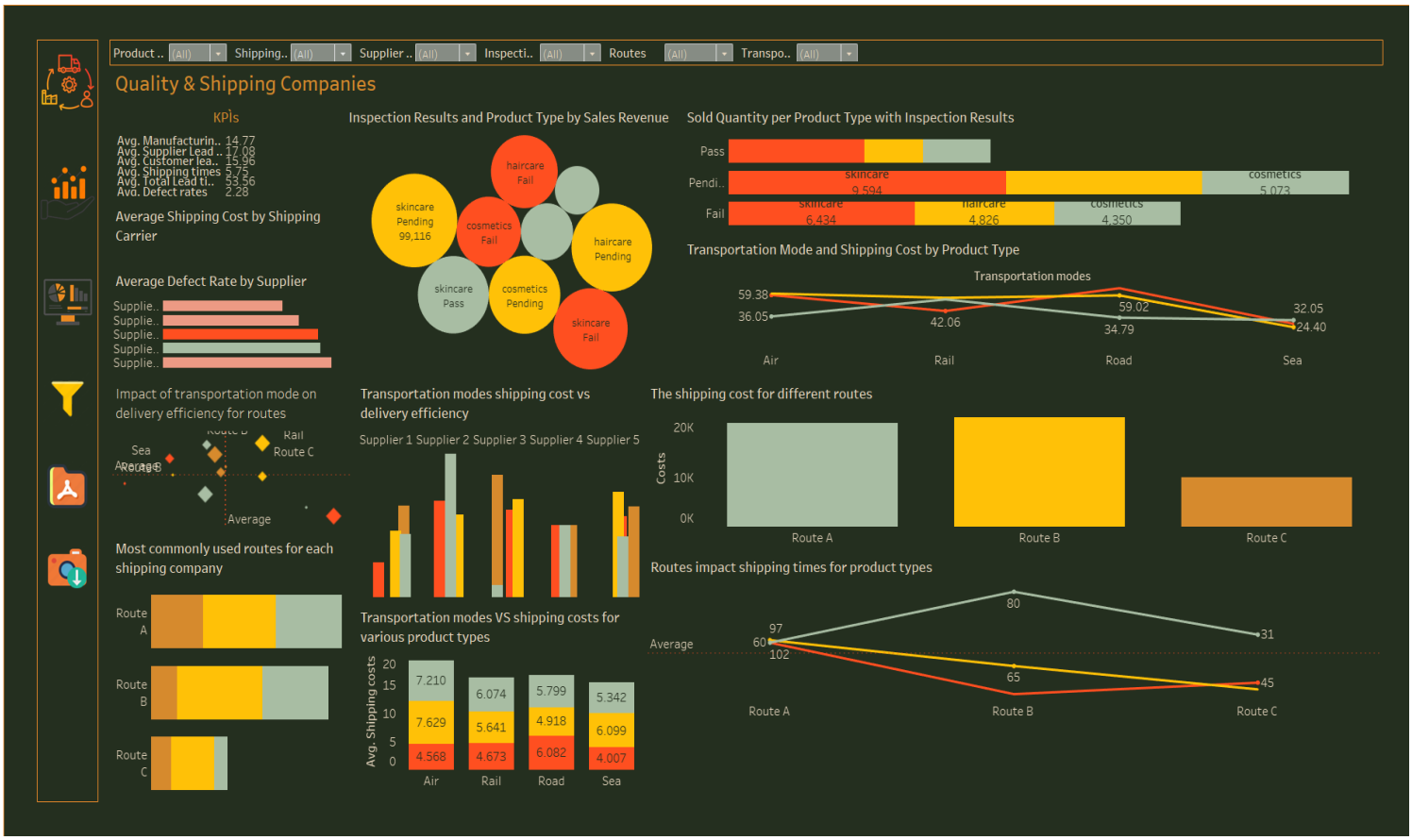
Multiple SQL queries were executed to produce insights. These queries were all gathered in a separate file.

Part Five:

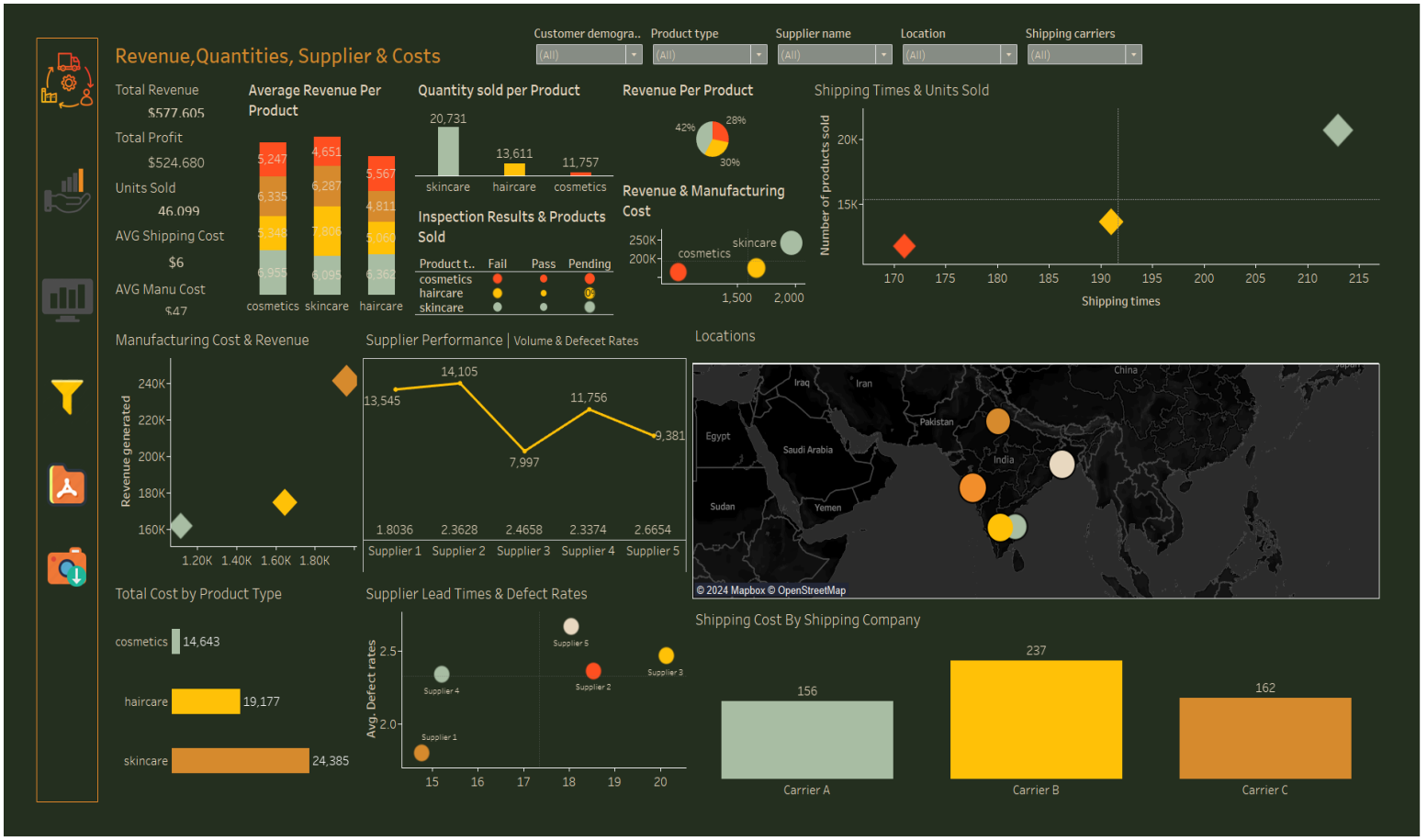
Data Visualisation

All insights were put in two dashboards. One related to Quantity and shipping company, the other related to revenue, quantities, supplier and costs.

1- Quantity and shipping company dashboard:



2- Revenue, quantities, supplier and costs



Part Six:

Tools Used

Multiple tools were used during the making of this project, which include:

1. Programming Languages: Python
2. IDE: Google Colab
3. Database Management System: SQL Server Management Studio
4. Business Intelligence and Analytics Software: Tableau

Part Seven:

Conclusion

7.1. Conclusion:

The business seems to be at a good start, however, there should be more attention paid to the cost reduction. The business could also add new lines of products to help increase the amount of quantity they sell.

7.2. Recommendations:

1. The business should focus on the category that generates the most revenue (skincare) since the costs for all of the categories are relatively similar.
2. Investigate the reason why products fail inspection to reduce the cost