

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Module Detail |  | Trainee's Detail | | |
| SECTOR: | ICT | Reg No: | 1.23RP00498  2. | |
| SUB-SECTOR: | Information Technology | Class: | Level 8 Information Technology | |
| Trainer's Detail | | |
| CERTIFICATE: | Bachelors of Technology | Name: | Eng. NYIRIMANA J.M Vianney | |
| MODULE  (Code &Title): | ITLDM801 – DATA MINING AND DATA WAREHOUSE | Additional info | | |
| Competence: | Apply Data Mining and Warehousing | Duration: |  | |
| Due date: | 30 March, 2025 | |
| Training Centre: | RP Ngoma College | Signature: |  | |
| Scored marks: |  | Decision: | Competent |  |
| Not Yet Competent |  |

**Store Books ETL and Data Warehouse Project Report**

1. **Summary**

This project aims to transform the existing Store Books Sales relational database into a data warehouse to enable advanced analytics, reporting, and business intelligence. The process involves Extract, Transform, Load **(ETL)** operations to migrate data from **Microsoft SQL Server 2022** into a structured warehouse model. The warehouse will support historical data analysis, trend identification, and decision-making for book sales management.

### ****Source Database Overview****

### The current database is set up as a relational database using Microsoft SQL Server 2022. It is managed with SQL Server Management Studio (SSMS), which helps in organizing, querying, and optimizing the data. The database includes several key tables that store information about authors, publishers, book languages, and books. These tables are connected to each other through relationships using foreign keys.

### Here are the main tables in the database:

### Author Table: This table contains information about the authors, with a unique ID for each author.

### Publisher Table: Stores information about the publishers, like the publisher’s name and ID.

### Book Language Table: Holds details about the languages of books, such as language codes and names.

### Book Table: This table has information about books, such as the title, ISBN number, number of pages, and publication date. It also connects to the Author, Publisher, and Language tables through foreign keys.

### Book Author Table: This table shows the many-to-many relationship between books and authors, linking each book to one or more authors.

### The following below is the overview of the Store Books Sales source database schema structure which shows us an illustration of tables, relationships, and key entities

### 

### Data Warehouse Schema (StoreBooksDW)

### 

1. **Installation of Libraries**

! pip install sqlalchemy pyodbc

!pip install pandas sqlalchemy psycopg2

!pip install sqlalchemy pyodbc

1. **Import Necessary Libraries**

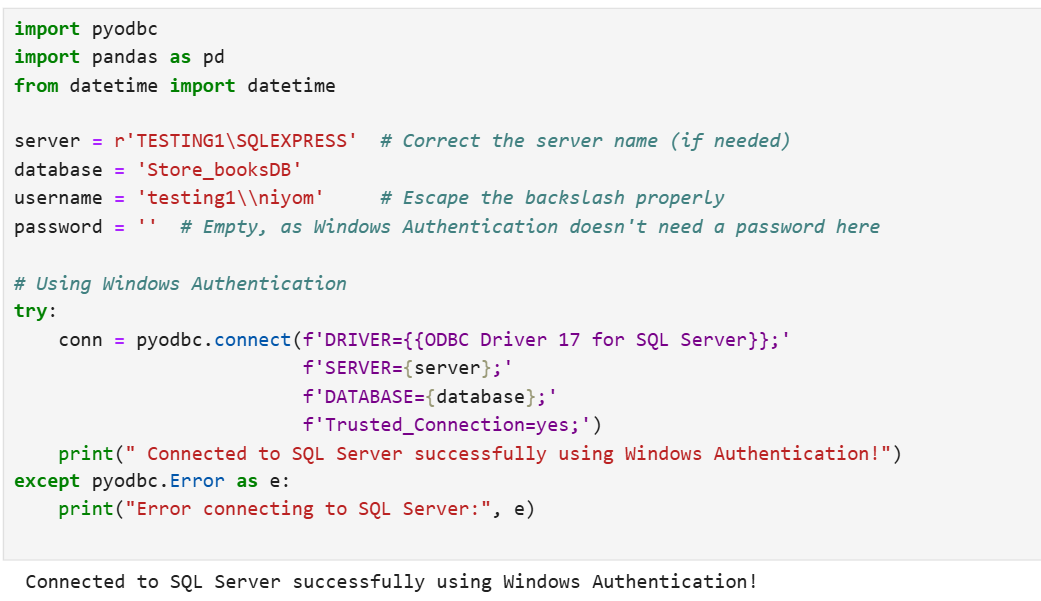
import pyodbc

import pandas as pd

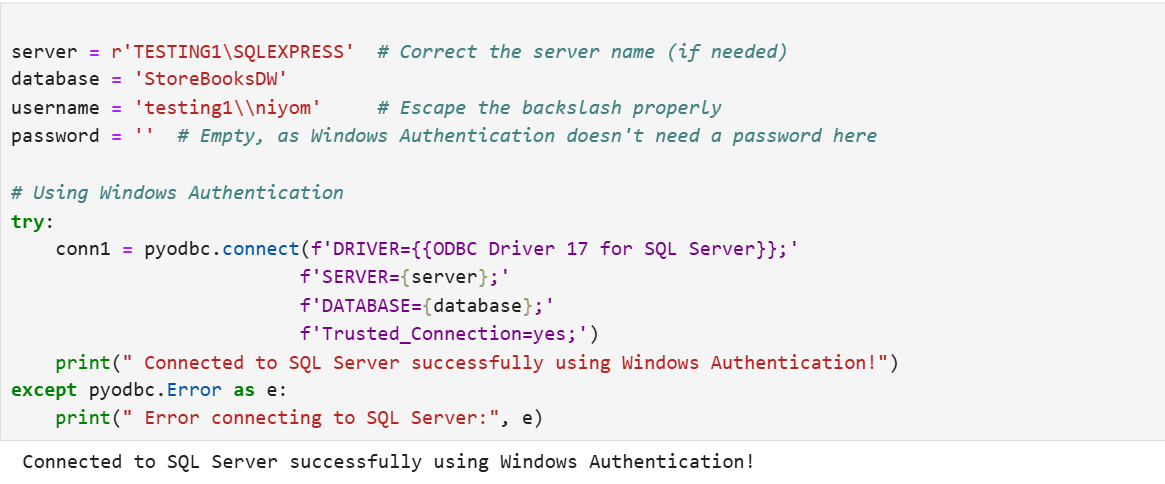
from datetime import datetime

1. **Database connection**

* **Database Connection**

****

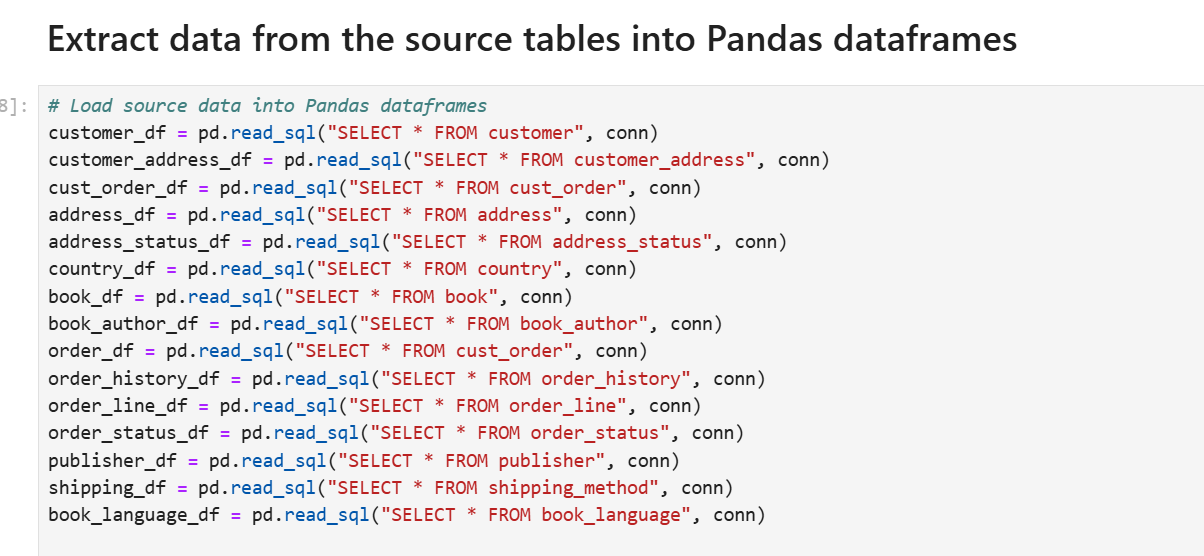
* **Data warehouse Connection**

****

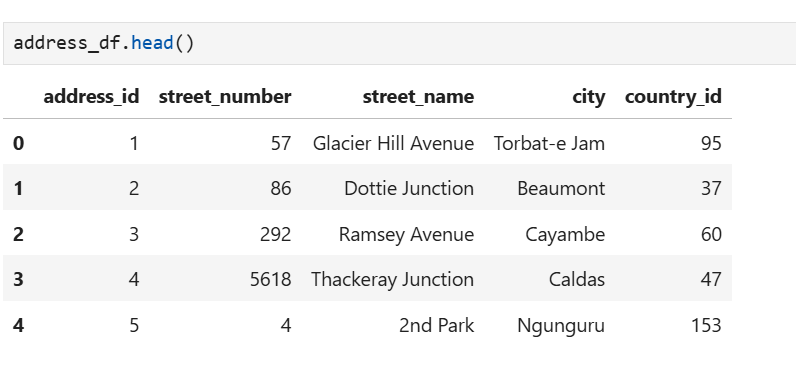
1. **ETL Process Overview**

* **Extraction Process** 
  + Extract data from the existing source database (Store\_booksDB) into staging tables.
  + Prepare data for transformation and loading into the data warehouse.
    1. **Extract data from author table**

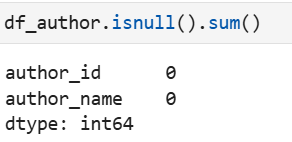




* + 1. **Reading data from address table**



* **Transformation**
* Data Cleaning: Handle any inconsistencies or missing values in the source data.
* Data Mapping: Map source tables to data warehouse structures.
  1. **Checking the null values**

****

* 1. **drop duplicates values**

****

* 1. **Selecting columns from pandas data frame with transformation**

****

**Rename and concatenate columns of DimCustomerData**

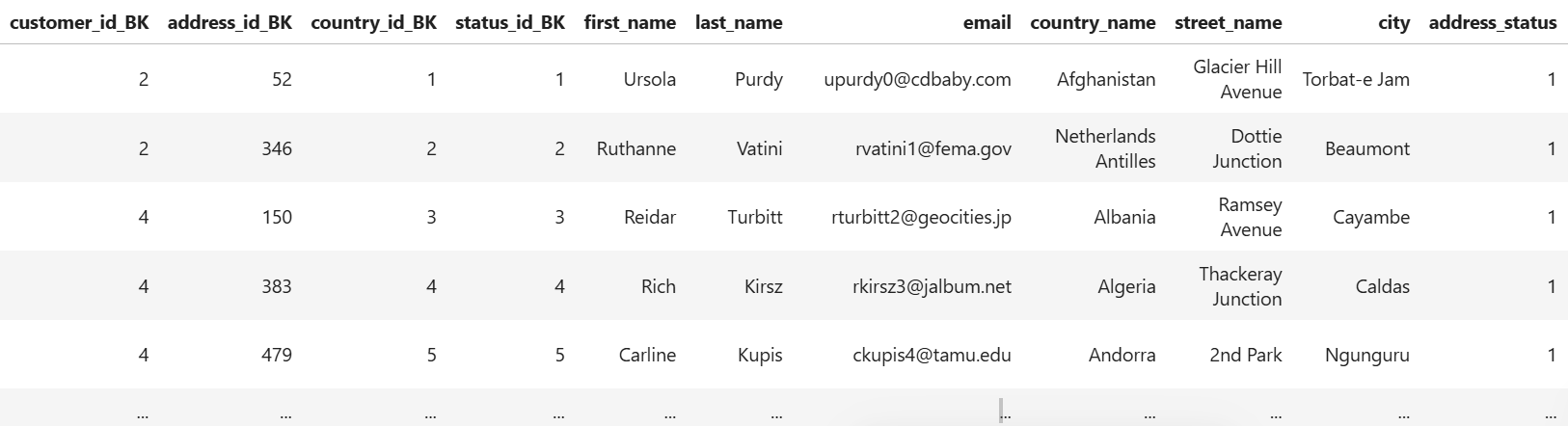


* **Loading**
* Load the transformed data into the data warehouse tables (StoreBooksDW).

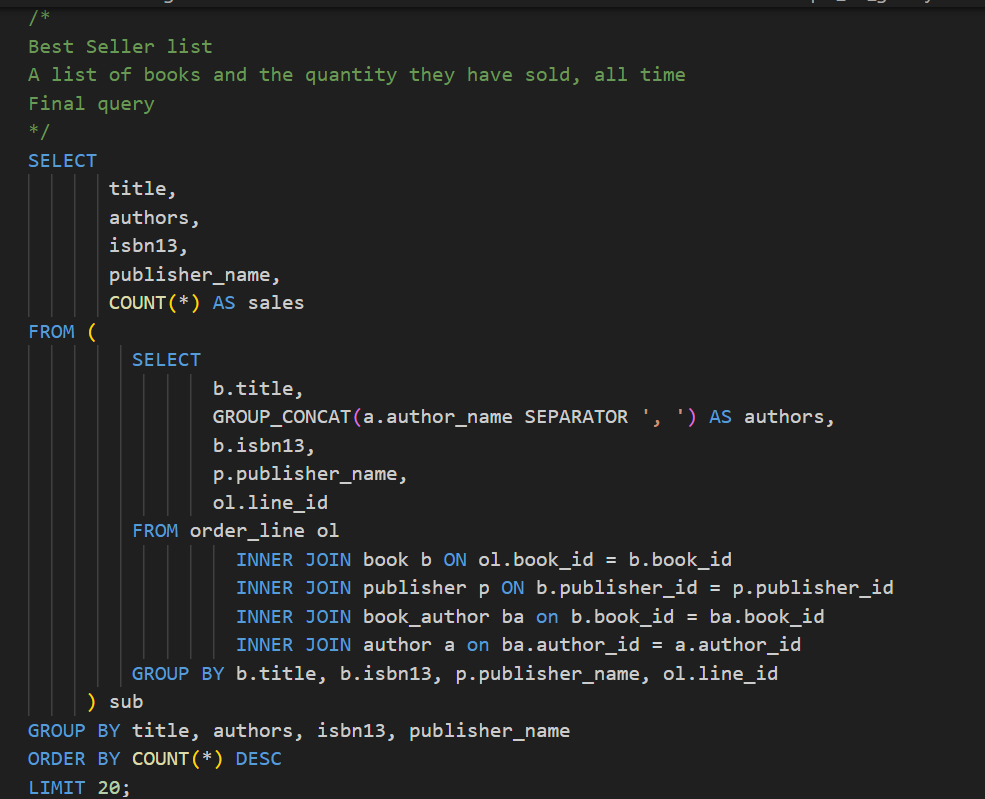
**Loading Data to DimCustomer Table**



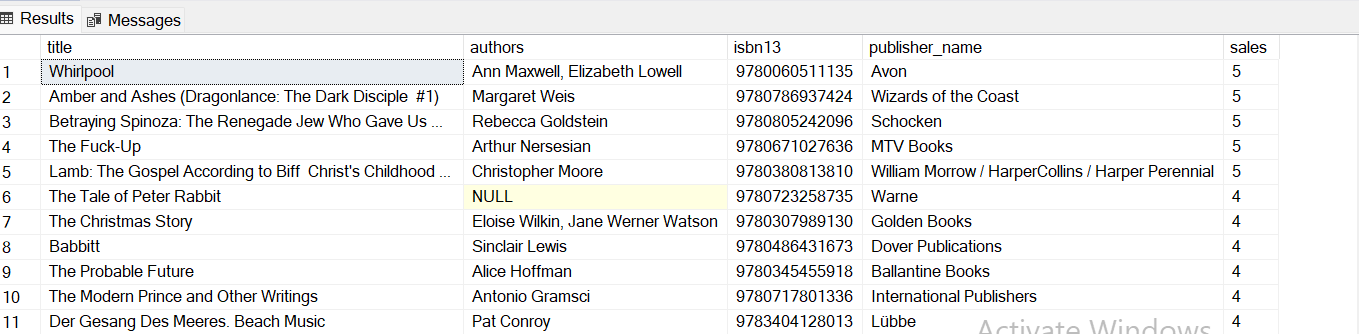
**OUTPUT**

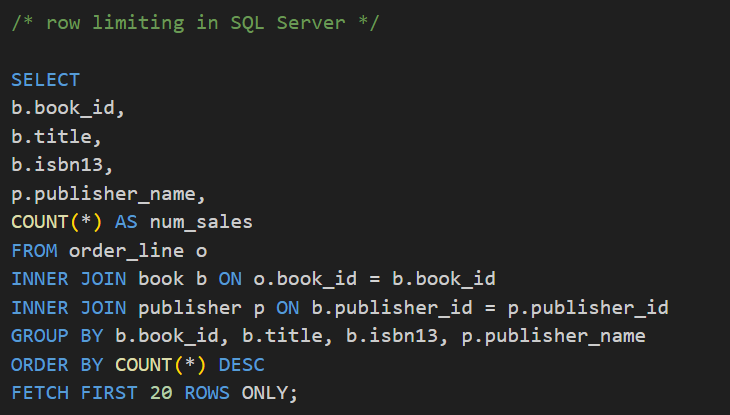
****

**Final query**

****

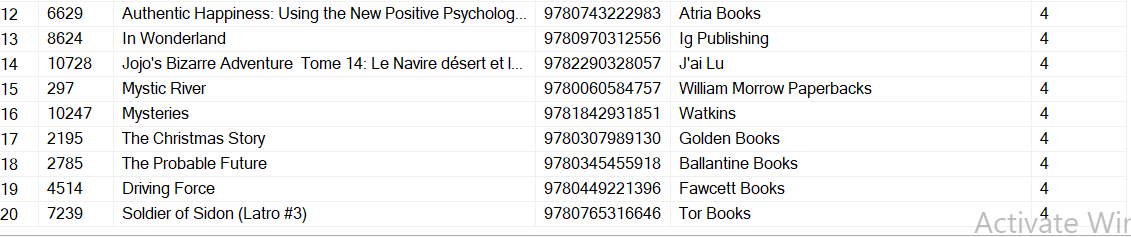
**Output**

****

****

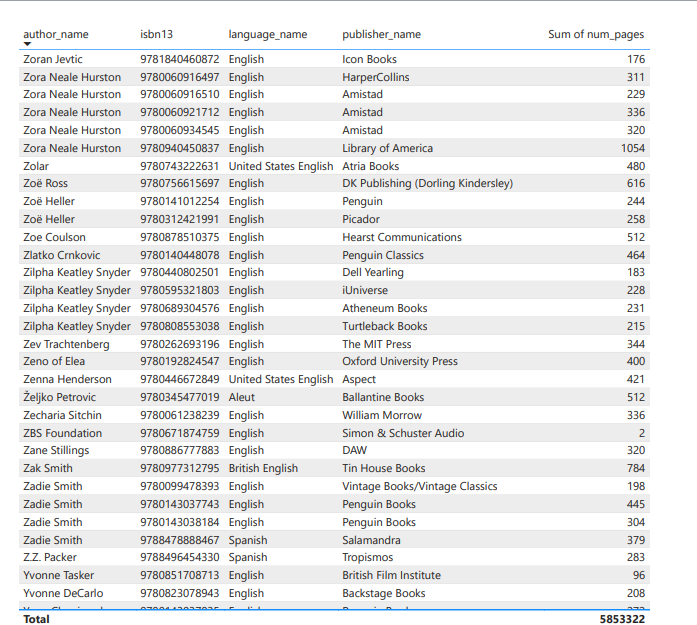
**Output**

****

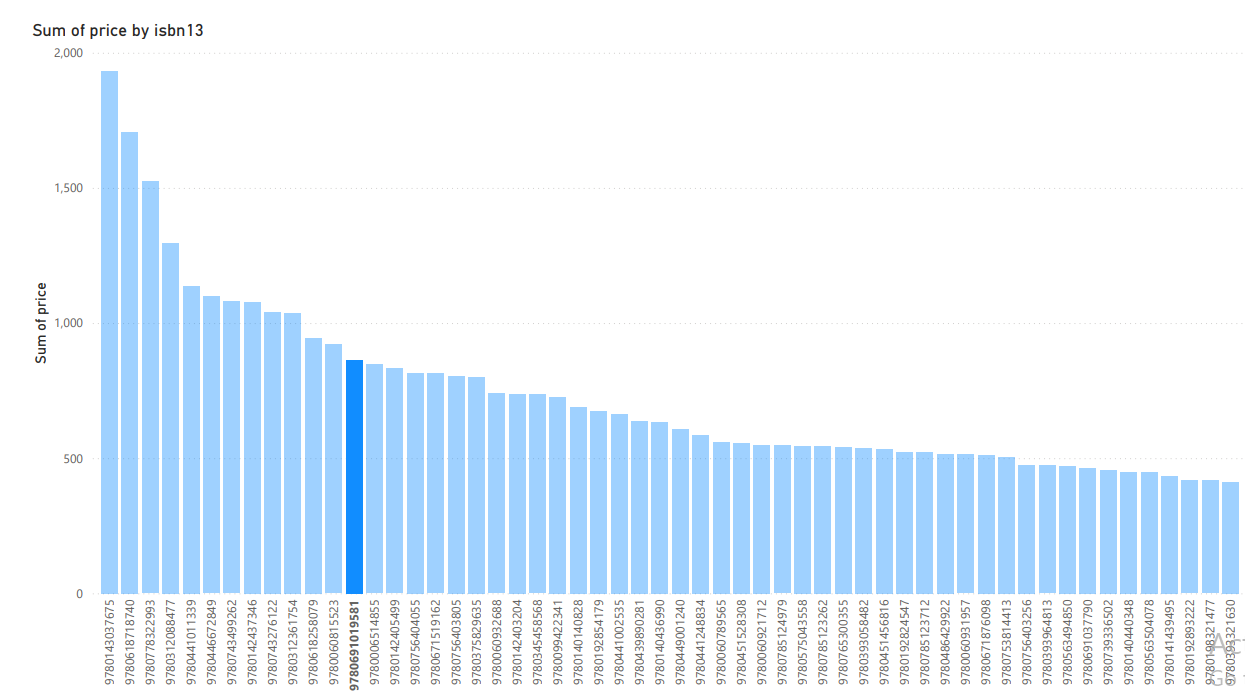
****

* **Power BI tool analysis**

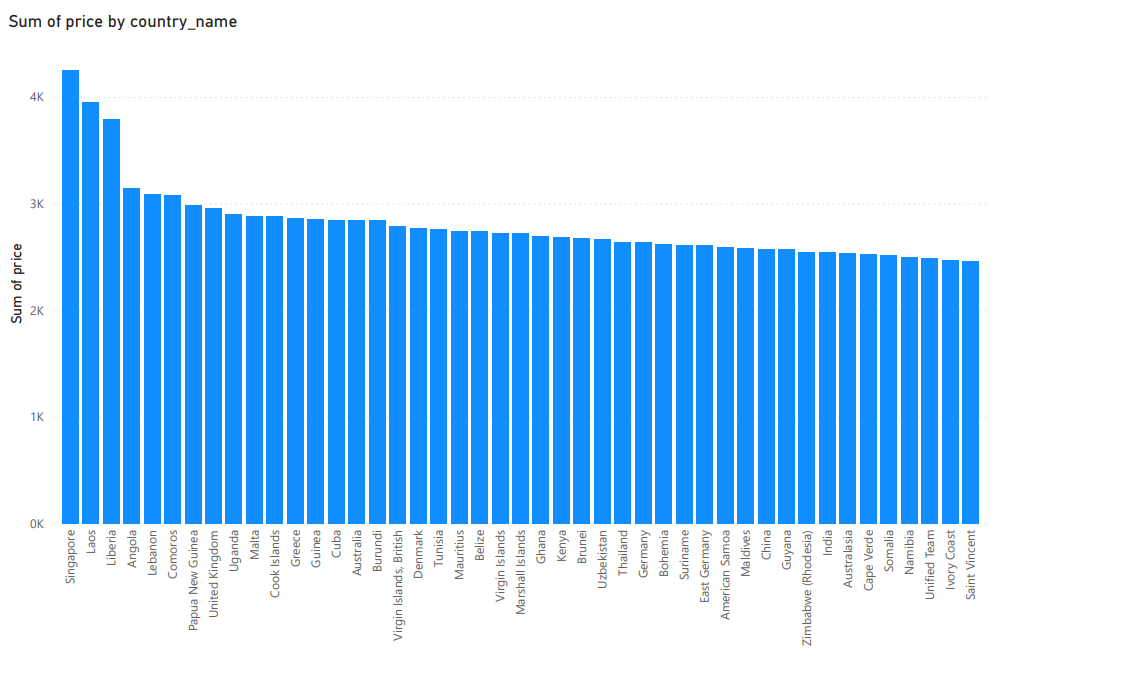
1. **Book details**



1. **Top most book**

****

1. **Country vs Price**

****