ETL Assignment

# Data Collection

There were multiple options to choose data from but I choose the Store Sales Data from the Kaggle which can be found [here](https://www.kaggle.com/datasets/tanayatipre/store-sales-forecasting-dataset). This dataset contains following columns.

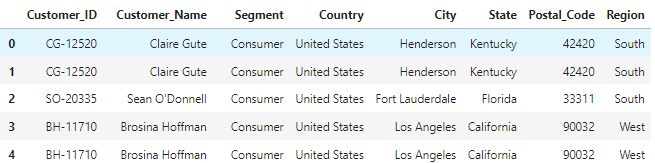
|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Row ID | Identifier for each row in the table |
| Order ID | Identifier for each order |
| Order Date | Date when the order was made |
| Ship Date | Date when the order was shipped |
| Ship Mode | Method used for shipping the order |
| Customer ID | Identifier for each customer |
| Customer Name | Name of the customer |
| Segment | Market segment the customer belongs to |
| Country | Country where the order was shipped |
| City | City where the order was shipped |
| State | State or province where the order was shipped |
| Postal Code | Postal code of the shipping destination |
| Region | Geographical region of the shipping destination |
| Product ID | Identifier for each product |
| Category | General category of the product |
| Sub-Category | Specific category of the product |
| Product Name | Name of the product ordered |
| Sales | Total sales amount for the order |
| Quantity | Number of items ordered |
| Discount | Discount applied to the order |
| Profit | Profit generated from the order |

# Splitting the Data

I have decided to split the data into 3 parts the first one will be loaded into Mysql Database second one will be requested from API and third one will be read from a local file. The Api is locally hosted because unfortunately I could not able to find the suitable option according to my dataset instead I build it locally and hosted.

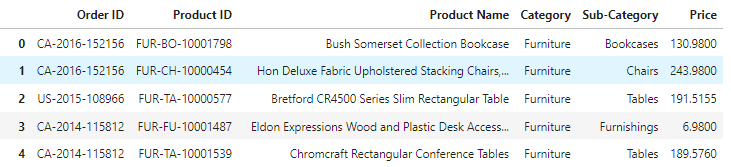
## Extracting from Mysql

The data is extracted from Mysql using the mysql connector in python following are the columns which were loaded from the Mysql.

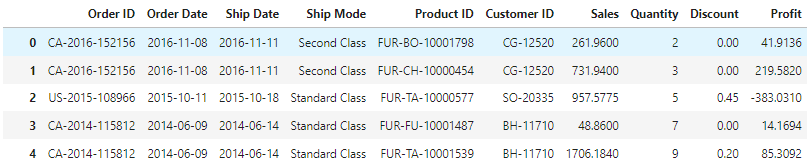


### Extracting from API

Following are the columns which were extracted from the API.

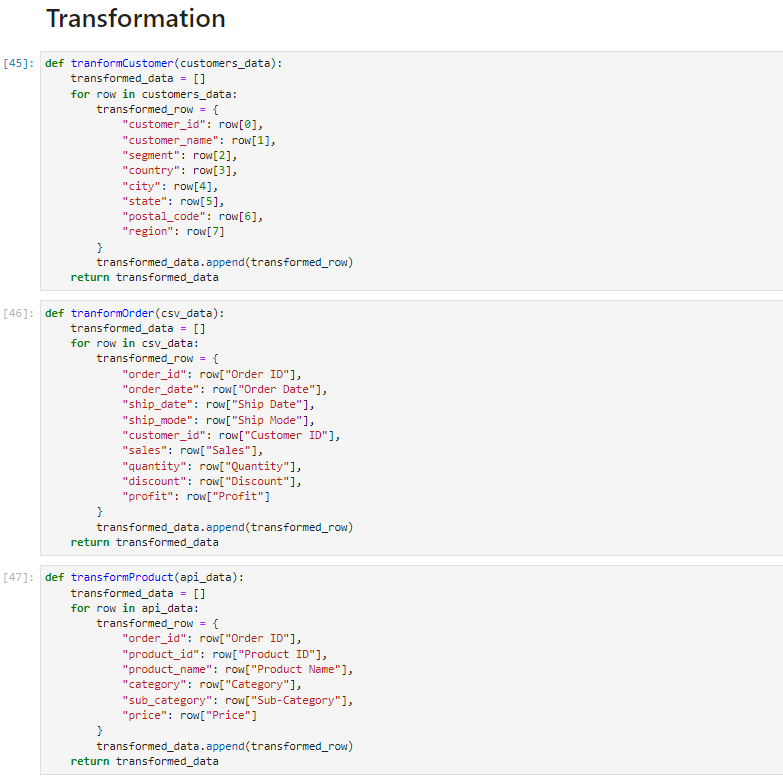


### Extracting from CSV



# Transformation

Transformation is done by assigning the appropriate columns and separating the columns for the right tables.



# Database Schema

