

Project Description:

There are 4 source systems in this project from where we need to load data in python data frames,

All these source systems are some third party databases, and have similar schema. For one of the source that is CSV_s1, we also got source schema extract to understand the fields' definition.

Note – Same schema is applicable for all 4 source systems datasets.

Each data source system contains 3 types of datasets named Customers, Products, and Transactions.

Use cases:

Usecase 1 to Usecase 7 (about csv)

Usecase-1 Load all three csv datasets in Pandas Data frames and display first 5 records.

Usecase-2 Display only those customers from CSV_s1, who purchased more than 3 products

Usecase-3 Display top 5 most demanded products from CSV_s1

Usecase-4 Display top 5 transactions **amount** from CSV_s1

Usecase-5 Display distinct professions from CSV_s1

Usecase-6 Display highest age in CSV_s1 customer's dataset

Usecase-7 Display custno, gender, age, profession, contactno, productno, productName, txndate, spendby, amount from CSV_s1 for custno = 923

Usecase-8 Load all three PSV (pipe separated values) dataset in Pandas Data frames and display first 5 records.

Usecase-9 Load all three JSON datasets in Pandas Data frames and display first 5 records.

Usecase-10 Load all three XML datasets in Pandas Data frames and display first 5 records.