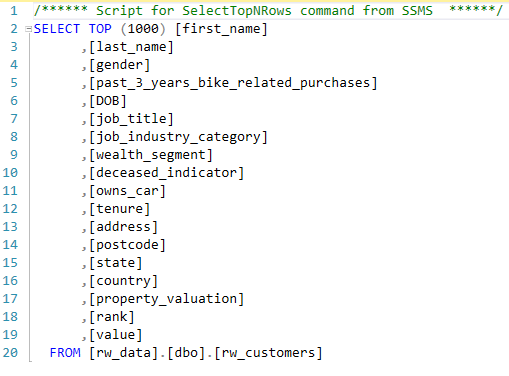
**Findings**

As I was assigned to bike data and need to make raw, master and staging database of the same, I came across multiple bugs and similarities in the data which was accounting to reluctancies.

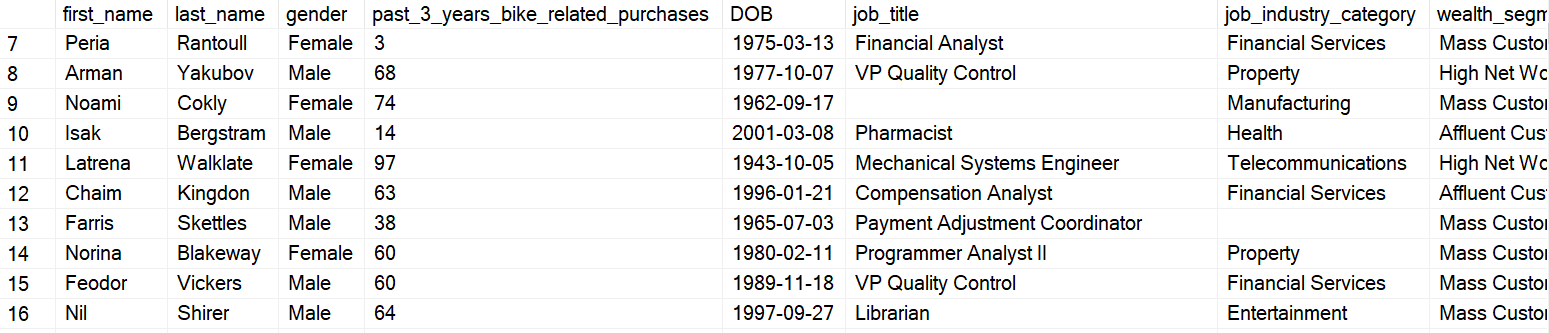
I was working with 4 CSV files which was as follows:

1. **Customers**:



In this CSV files, one of the major bugs was that there were no Customer\_ID Field. This made this data very much tricky to work upon.

This data tells majorly revolves about a customer’s location and the wealth he owns, and which segment he/she belongs to. The work he does and the assests he has.



1. **Customer Demographics:**

A screenshot of a computer code

Description automatically generated with medium confidence

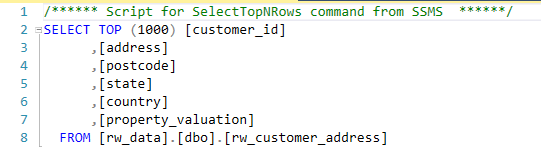
This data is a subset of customer data which includes Customer\_ID. This makes this data more reliable but on the other hand it also makes it tricky as it need to be joined with some other table otherwise it will give reluctancy when joined to extract the output.

A screenshot of a computer

Description automatically generated with medium confidence

1. **Customer Addresses:**

This data is also a subset of customer data as it also has Customer\_ID as a field.



It let us know only the address of a customer.

A screenshot of a computer

Description automatically generated with low confidence

1. **Transactions:**

A screenshot of a computer

Description automatically generated with medium confidence

This data is quite different from the previous 3 data as it let us know about the transactions done by a customer and the product and brand he was after.

A screenshot of a computer

Description automatically generated with low confidence

**Creation of RAW DATABASE:**

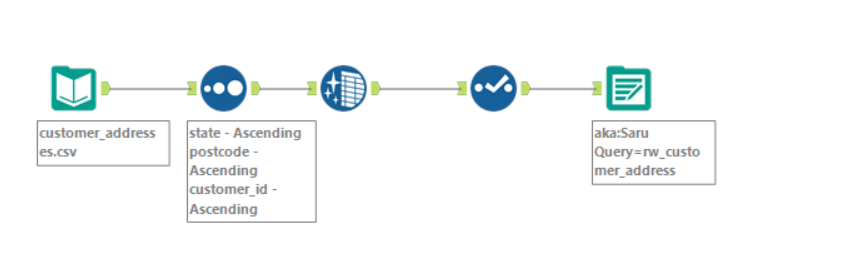
A screenshot of a computer

Description automatically generated

Creation of Raw Database was not as difficult as we just need to clean the data and sort it. Also, need to align the fields to the correct data types and putting each field into it.

A picture containing text, font, line, diagram

Description automatically generated



A picture containing text, font, line, diagram

Description automatically generated

A picture containing text, font, line, diagram

Description automatically generated

The above figures show the working of cleaning and manipulation of data in the Alteryx for finally putting it in raw database.

**Creation of Master DATABASE:**

There were lot of issues while making Master database because of the bug in Customer.csv.

When I joined customer table with customer addresses table and customer demographics table based on first name, last name and DOB, there were a smaller number of data that I got which was less than 100. This joining was very essential because it will be joined to transactions table to make the complete master database for Bike data.

As there were lot of bugs and data inconsistency, I did outer join on customer demography and customer addresses to get a wholesome of data and to get an idea of the complete data. Yes, it has a lot of null values, but this will the master database for further investigation.

A screenshot of a computer

Description automatically generated with medium confidence

This shows that how much data we do not have to preform adequate analysis.

Because of this I made 2 Master Databases, one of customer and the other of Transaction.

**Transaction Master Database**

A picture containing screenshot, software, diagram, text

Description automatically generated

**Customer Master Database**

A picture containing text, diagram, plot, line

Description automatically generated

The data has lot of null values because of the lack of complete data that I got.

A screenshot of a computer

Description automatically generated with medium confidence

Finally, with the 2 Master Databases I have made Staging Database of Transaction.

**Transaction Staging Database:**

A picture containing text, diagram, screenshot, design

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated with medium confidence