**DEPARTMENT OF INFORMATION TECHNOLOGY**

**COURSE CODE: DJ19ITL503 DATE: 20-08-24**

**COURSE NAME: Data Warehousing and Mining CLASS: I1-Batch-1**

**NAME:Anish Sharma ROLL NO.: I011**

**LAB EXPERIMENT NO. 3**

**CO/LO:** Apply ETL steps for a given dataset

**AIM / OBJECTIVE:** Executing ETL operations on Talend Tool

**DESCRIPTION OF EXPERIMENT:**

ETL (Extract, Transform, Load) is used to integrate data from various sources into a unified format for analysis. It extracts data, transforms it by cleansing and enriching, and loads it into data warehouses or other systems, enabling efficient reporting, business intelligence, and improved data quality.

ETL processes work by first extracting data from multiple sources, then transforming it through cleansing, aggregation, and formatting to ensure consistency and accuracy. Finally, the transformed data is loaded into a data warehouse or database. This sequence enables efficient data integration, analysis, and reporting for informed decision-making.

**INPUT DATA / DATASET:**

[customers.xlsx f](https://help.talend.com/en-US/data-preparation-getting-started/7.3/Content/Resources/attachments/customers.xlsx)ile.

The dataset provided contains information about individuals and their associated details, presumably from a customer or client database. Here’s a brief description of each column and what it represents:

1. **Id:** A unique identifier for each individual in the dataset.
2. **First\_Name:** The individual's first name.
3. **Last\_Name:** The individual's last name.
4. **Gender:** The gender of the individual (e.g., Male, Female).
5. **Age:** The age group of the individual, categorized into ranges such as "Under 18" or "25-

34."

1. **Occupation:** The individual's job title or occupation.
2. **MaritalStatus\_Out:** The individual's marital status (e.g., Single, Married, Divorced).
3. **Salary\_Out:** The individual's salary range, given in financial brackets (e.g., "< 50,000" or "100,000-149,999").
4. **Address:** The street address of the individual.
5. **City:** The city where the individual resides.
6. **State:** The state or region where the individual resides.
7. **Zip:** The postal code for the individual's address.
8. **Phone:** The contact phone number of the individual.
9. **Email:** The email address of the individual.
10. **SubDate:** The subscription or registration date in DD-MMM-YYYY format.
11. **Number\_of\_rentals:** The total number of rentals or transactions associated with the individual.

**PROCEDURE / ALGORITHM:**

1. Open Talend Open Studio
2. Select data file for extraction process.
3. Perform transformation operations on the extracted file

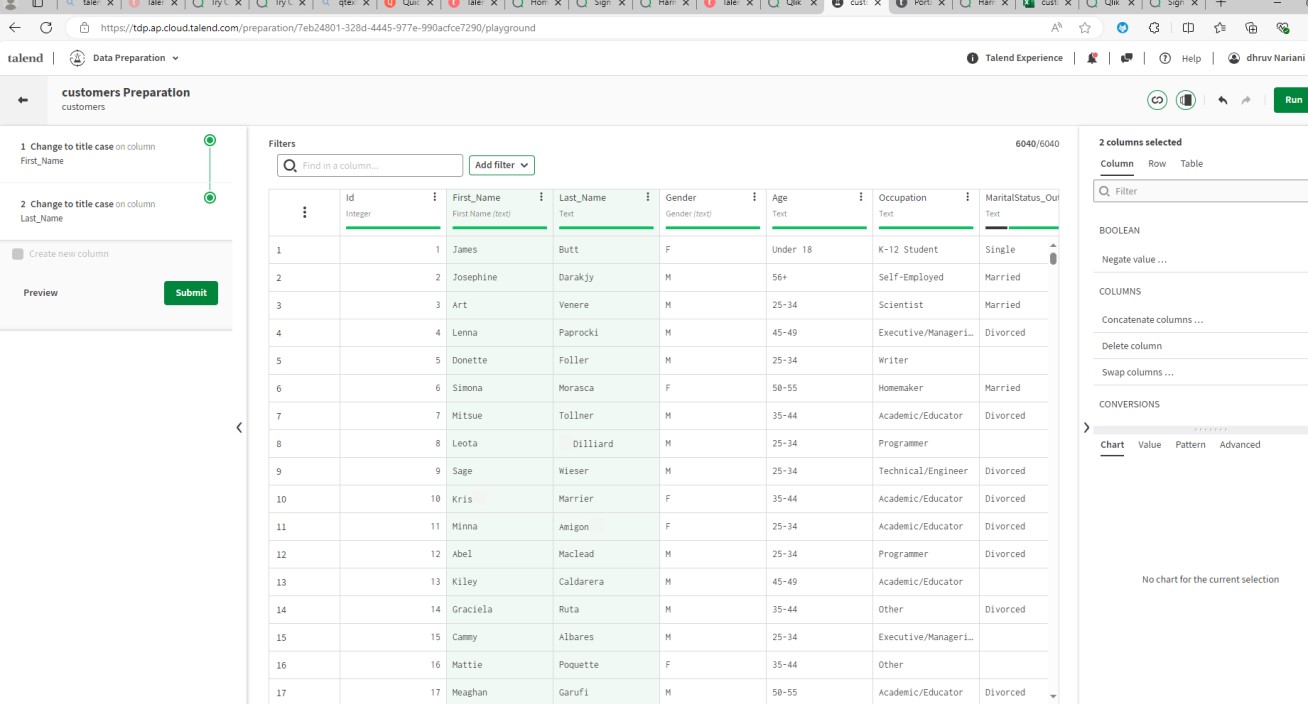
**TECHNOLOGY STACK USED:**

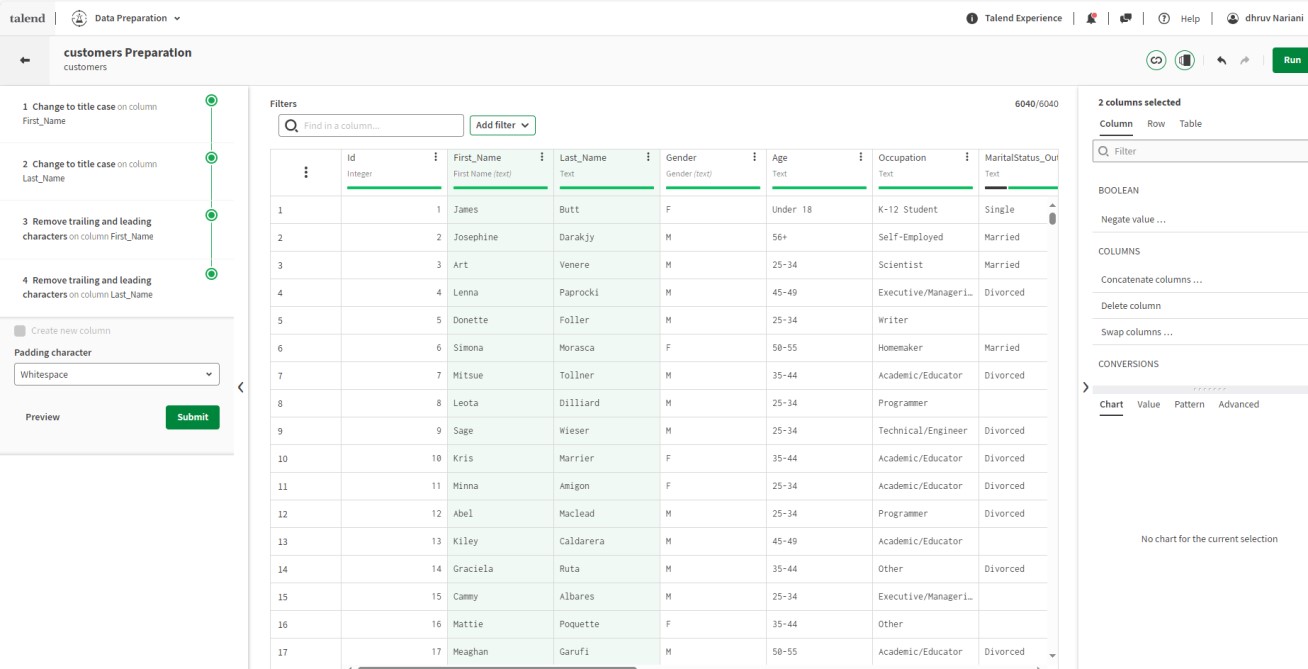
Talend Open Studio is an open-source software suite for data integration, data quality, and data management. It provides a user-friendly graphical interface for designing ETL (Extract, Transform, Load) processes, allowing you to connect, transform, and manage data from various sources. Key features include:

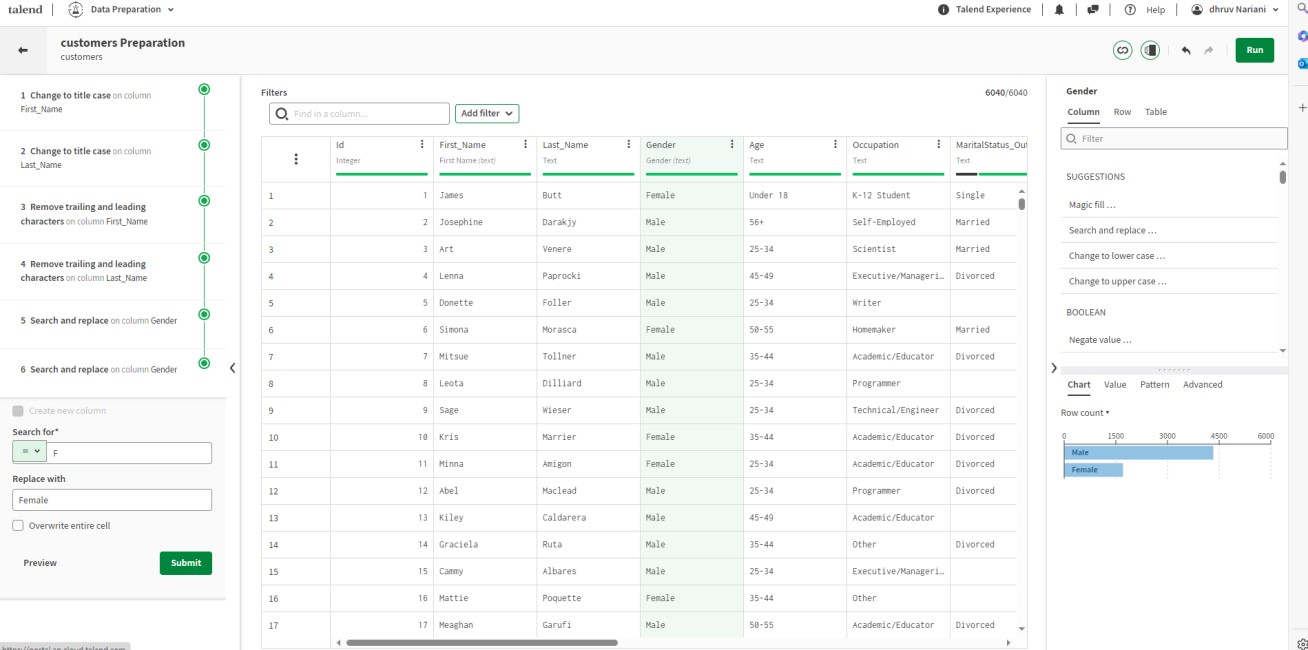
* + Graphical design interface
  + Wide range of data source connectivity
  + Data quality tools
  + Extensible and open-source

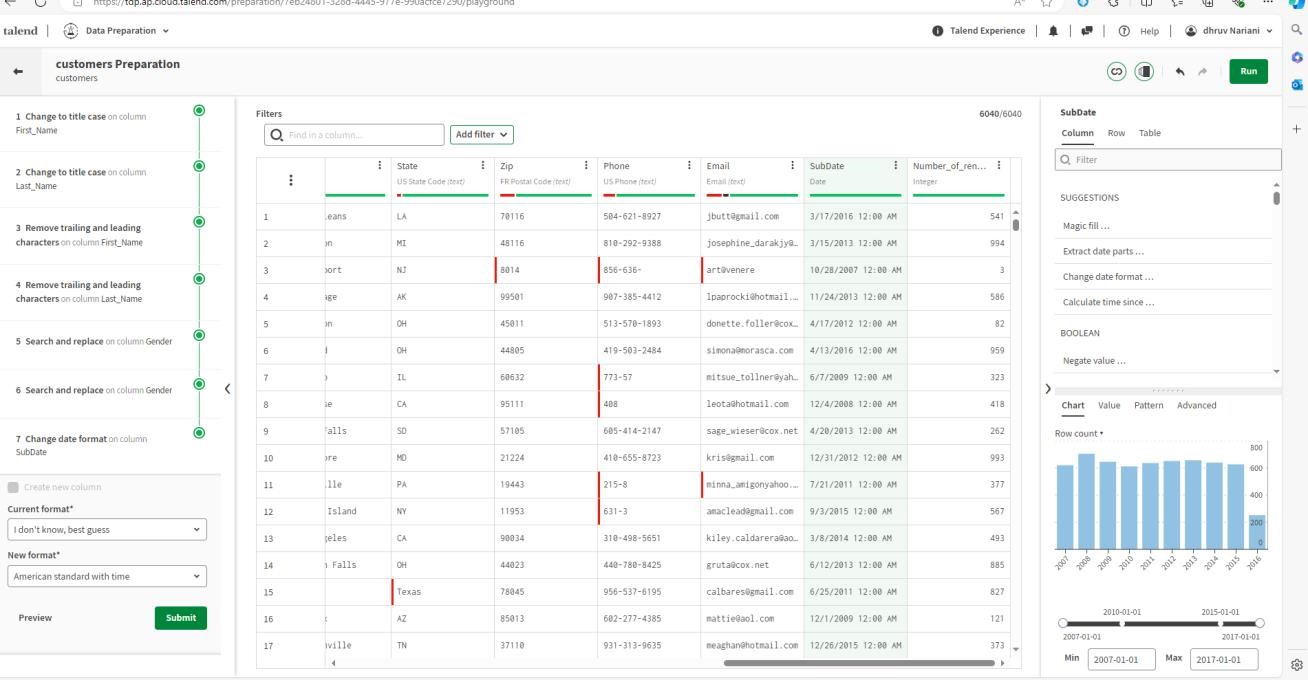
It's used for tasks like data warehousing, ETL processes, and data migration.

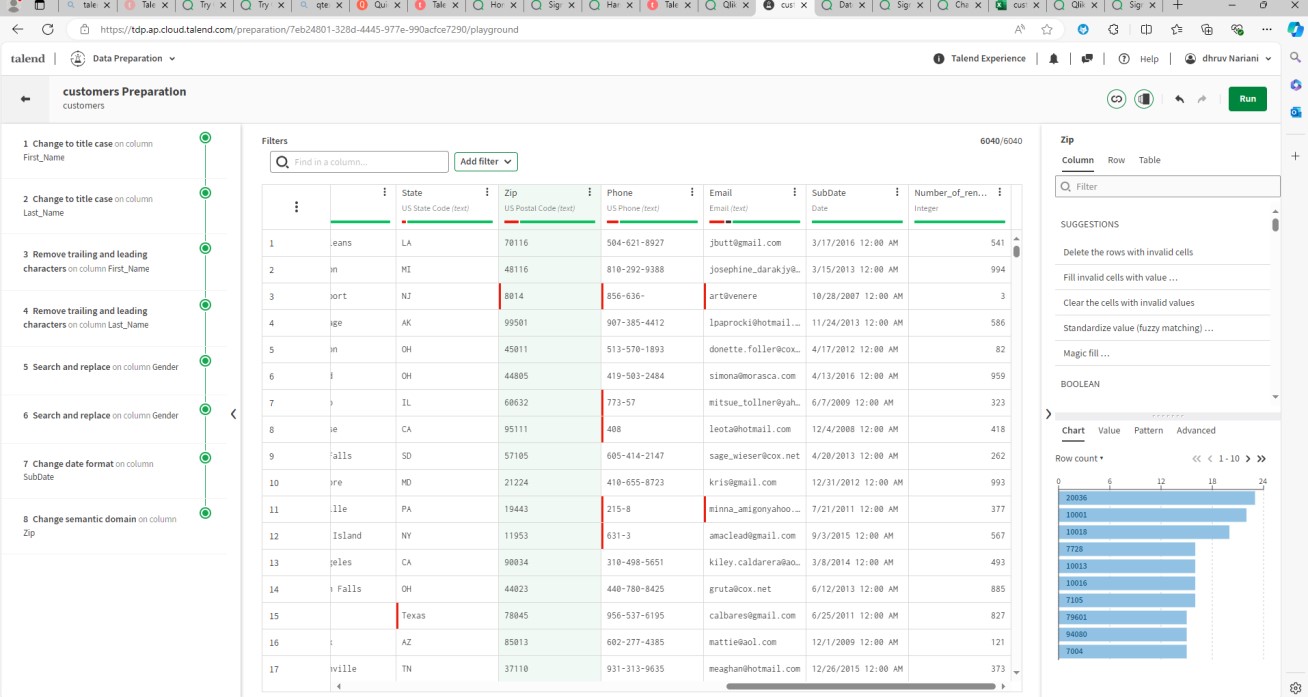
**OBSERVATIONS / DISCUSSION OF RESULT:**

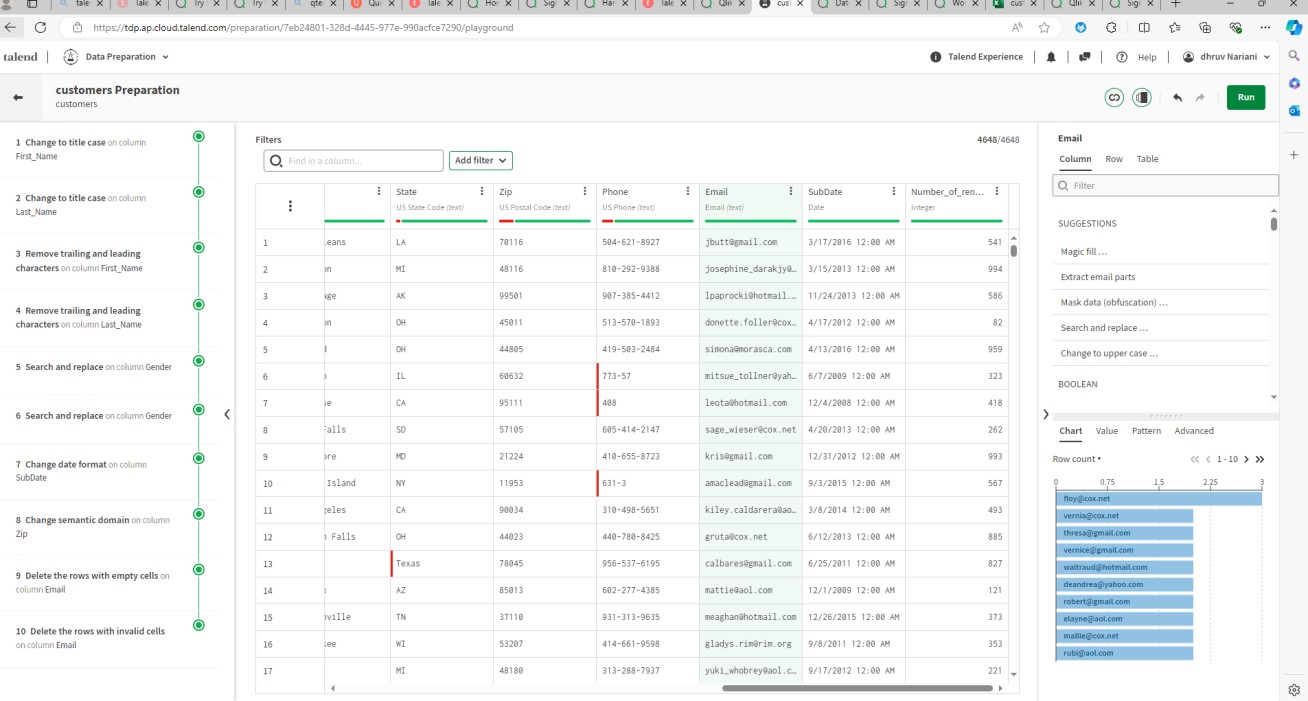


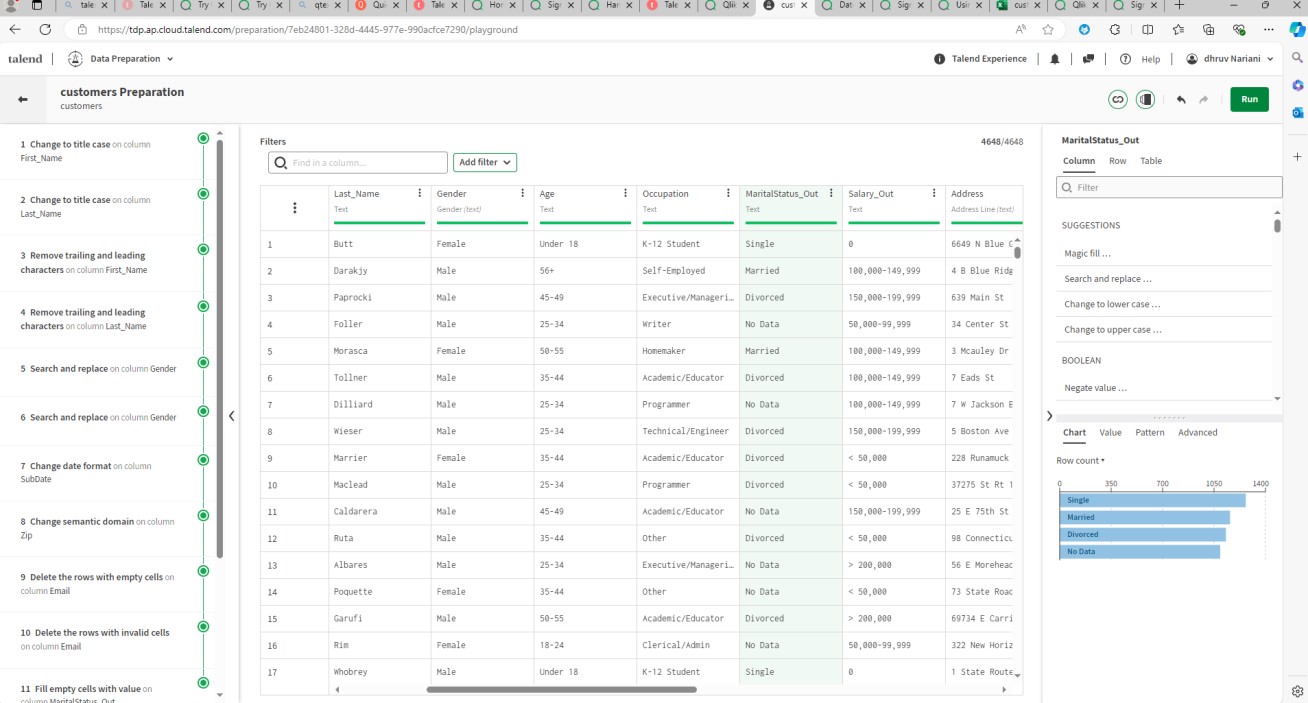


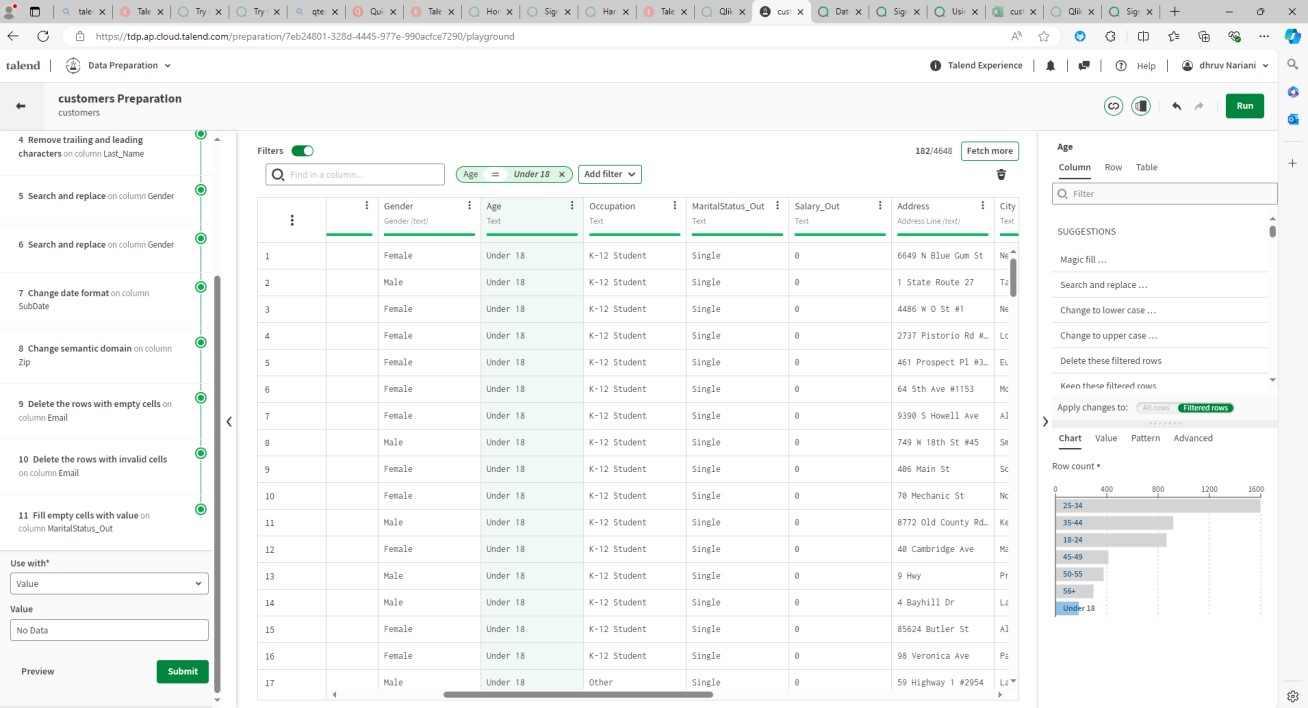












**CONCLUSION:**

In this experiment, we learned how to apply ETL steps for a given dataset

**Website References:**

1. https://www.talend.com/products/data-preparation/
2. https://help.talend.com/r/en-US/7.3/data-preparation-getting-starte