SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

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 file.

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.

SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

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."
6. **Occupation:** The individual's job title or occupation.
7. **MaritalStatus_Out:** The individual's marital status (e.g., Single, Married, Divorced).
8. **Salary_Out:** The individual's salary range, given in financial brackets (e.g., "< 50,000" or "100,000-149,999").
9. **Address:** The street address of the individual.
10. **City:** The city where the individual resides.
11. **State:** The state or region where the individual resides.
12. **Zip:** The postal code for the individual's address.
13. **Phone:** The contact phone number of the individual.
14. **Email:** The email address of the individual.
15. **SubDate:** The subscription or registration date in DD-MMM-YYYY format.
16. **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:

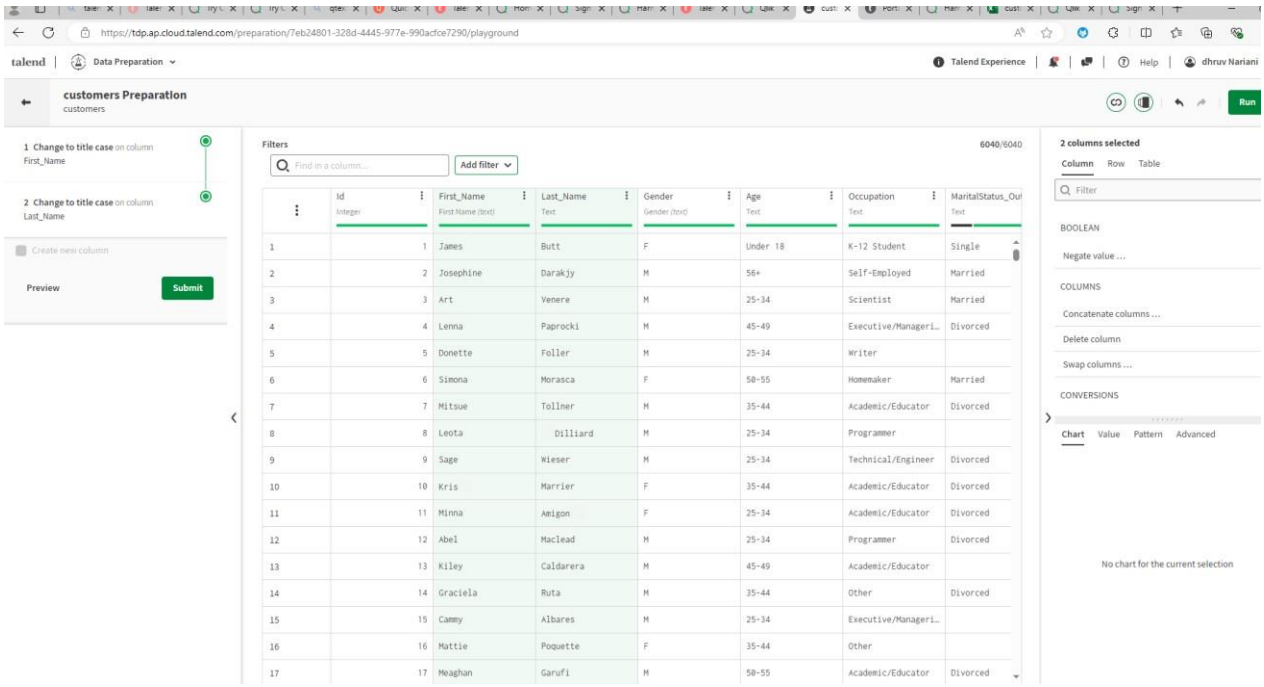
Talend Open Studio is an open-source software suite for data integration, data quality, and data management. It provides a user-friendly graphical interface for designing ETL (Extract, Transform, Load) processes, allowing you to connect, transform, and manage data from various sources. Key features include:

- 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.

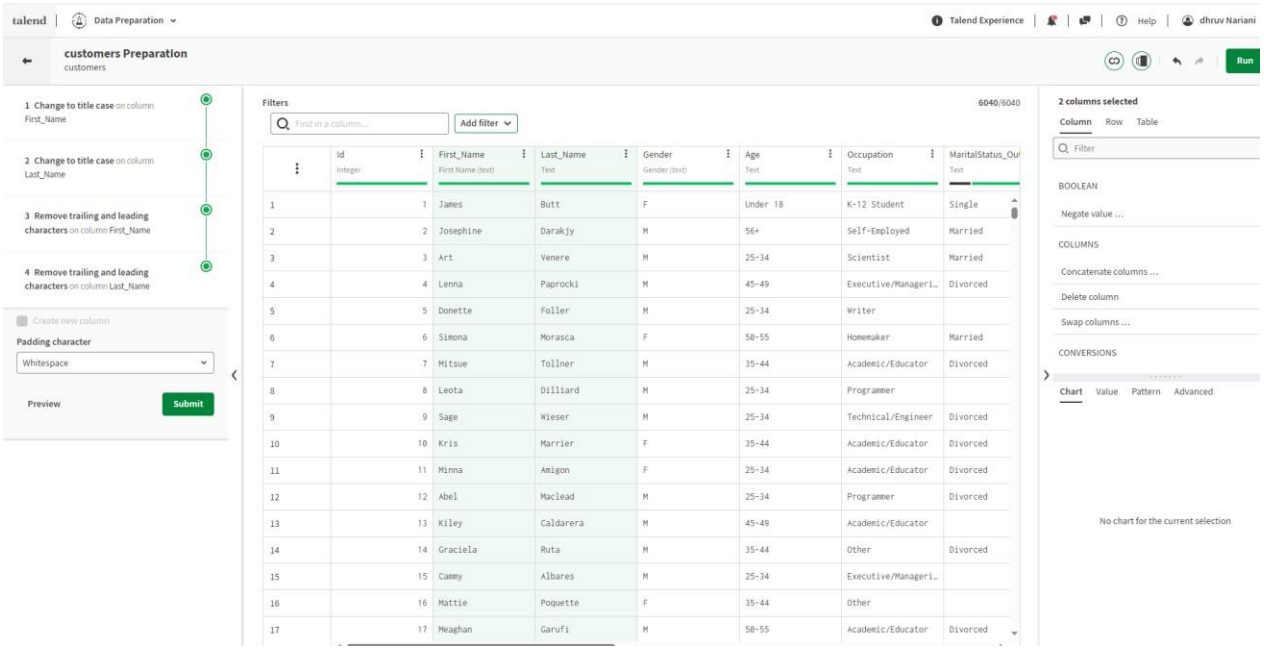
SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

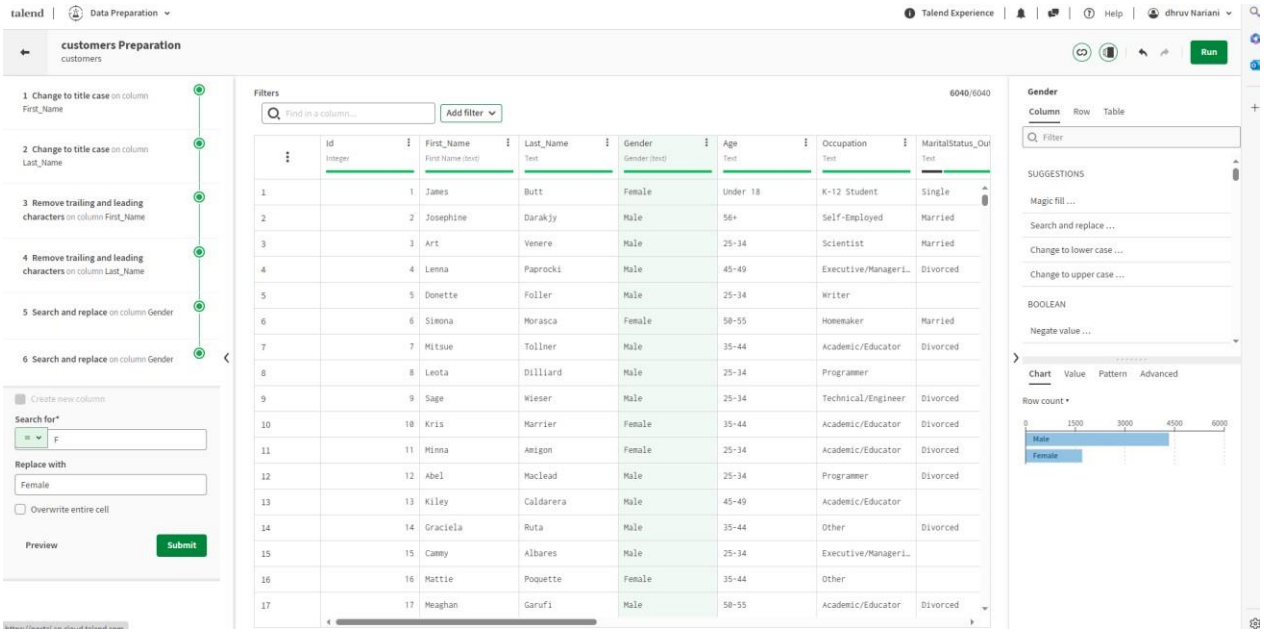
OBSERVATIONS / DISCUSSION OF RESULT:



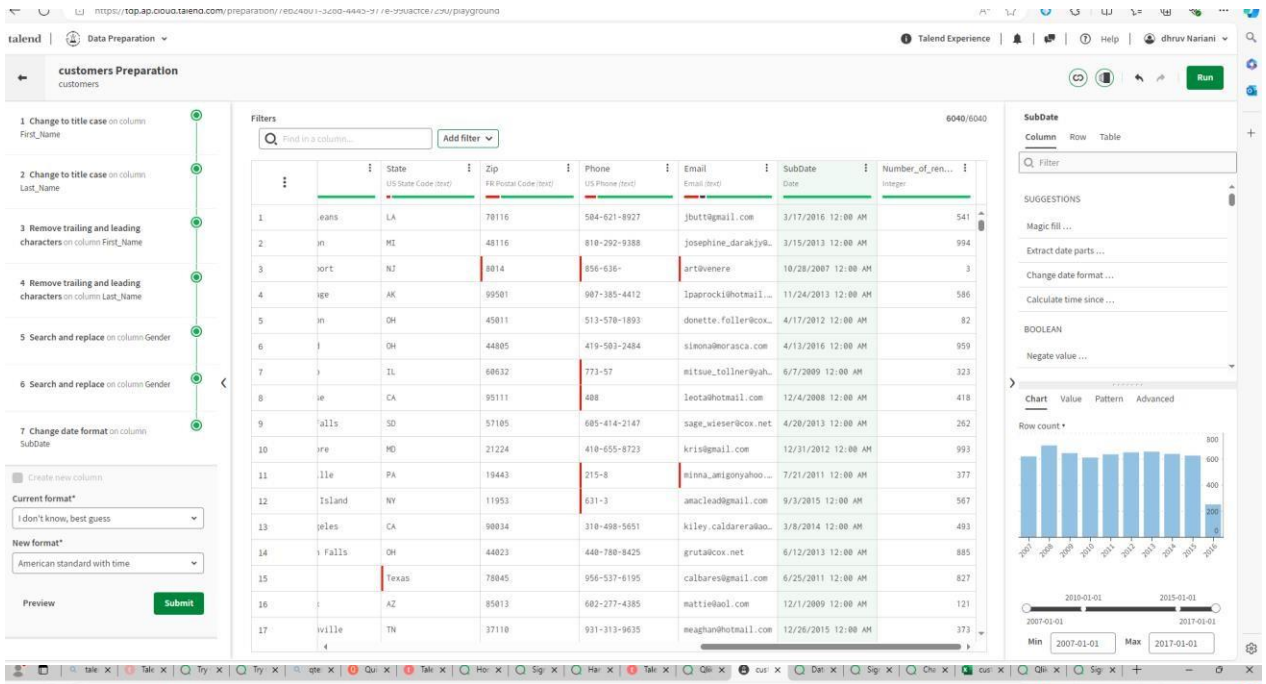
	Id	First_Name	Last_Name	Gender	Age	Occupation	MaritalStatus_Ou
1	1	James	Butt	F	Under 18	K-12 Student	Single
2	2	Josephine	Darakjy	M	56+	Self-Employed	Married
3	3	Art	Venere	M	25-34	Scientist	Married
4	4	Lenna	Paprocki	M	45-49	Executive/Managerі...	Divorced
5	5	Donette	Foller	M	25-34	Writer	
6	6	Simona	Morasca	F	50-55	Homemaker	Married
7	7	Mitsue	Tollner	M	35-44	Academic/Educator	Divorced
8	8	Leota	Dilliard	M	25-34	Programmer	
9	9	Sage	Wieser	M	25-34	Technical/Engineer	Divorced
10	10	Kris	Marrier	F	35-44	Academic/Educator	Divorced
11	11	Minna	Amigon	F	25-34	Academic/Educator	Divorced
12	12	Abel	Maclead	M	25-34	Programmer	Divorced
13	13	Kiley	Caldarera	M	45-49	Academic/Educator	
14	14	Graciela	Ruta	M	35-44	Other	Divorced
15	15	Cammy	Albares	M	25-34	Executive/Managerі...	
16	16	Mattie	Poquette	F	35-44	Other	
17	17	Meaghan	Garufi	M	50-55	Academic/Educator	Divorced

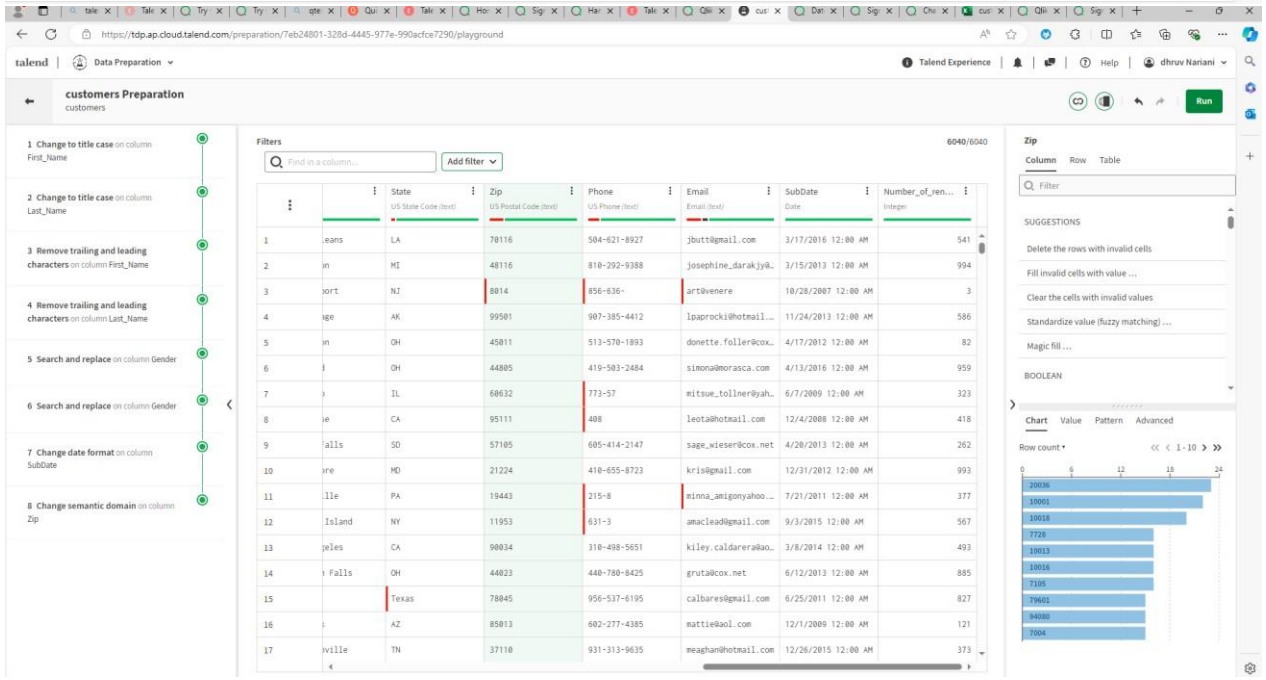
SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)



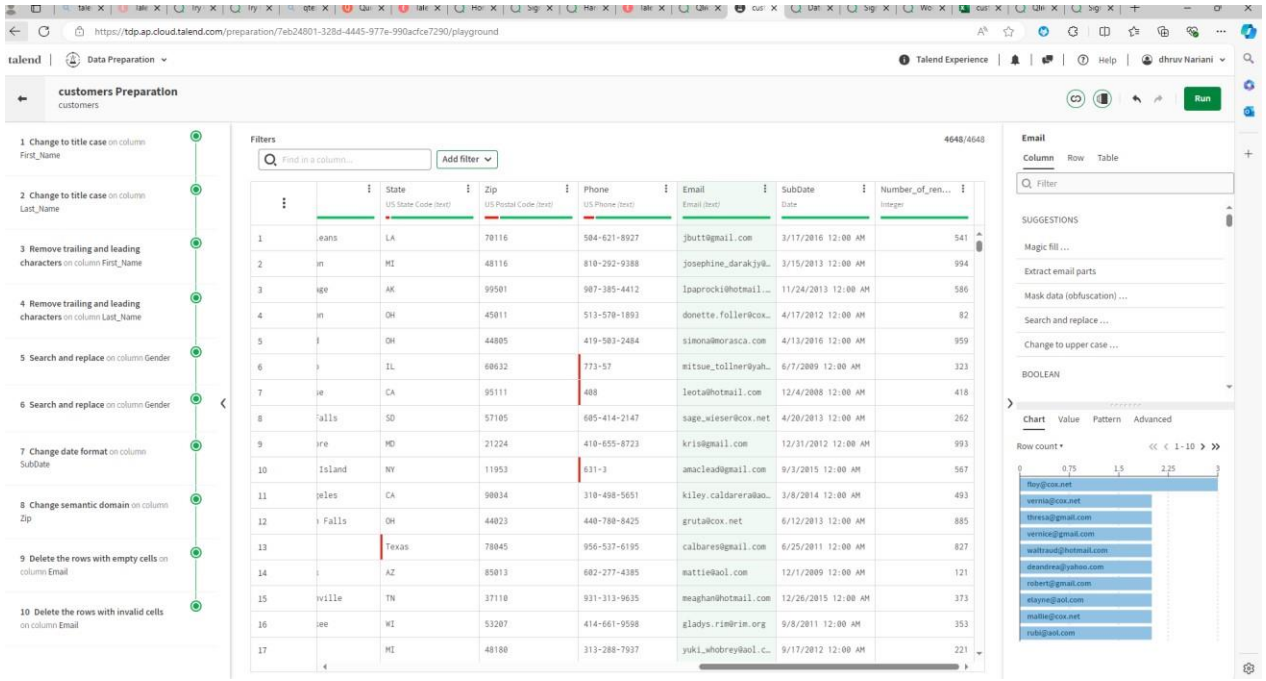


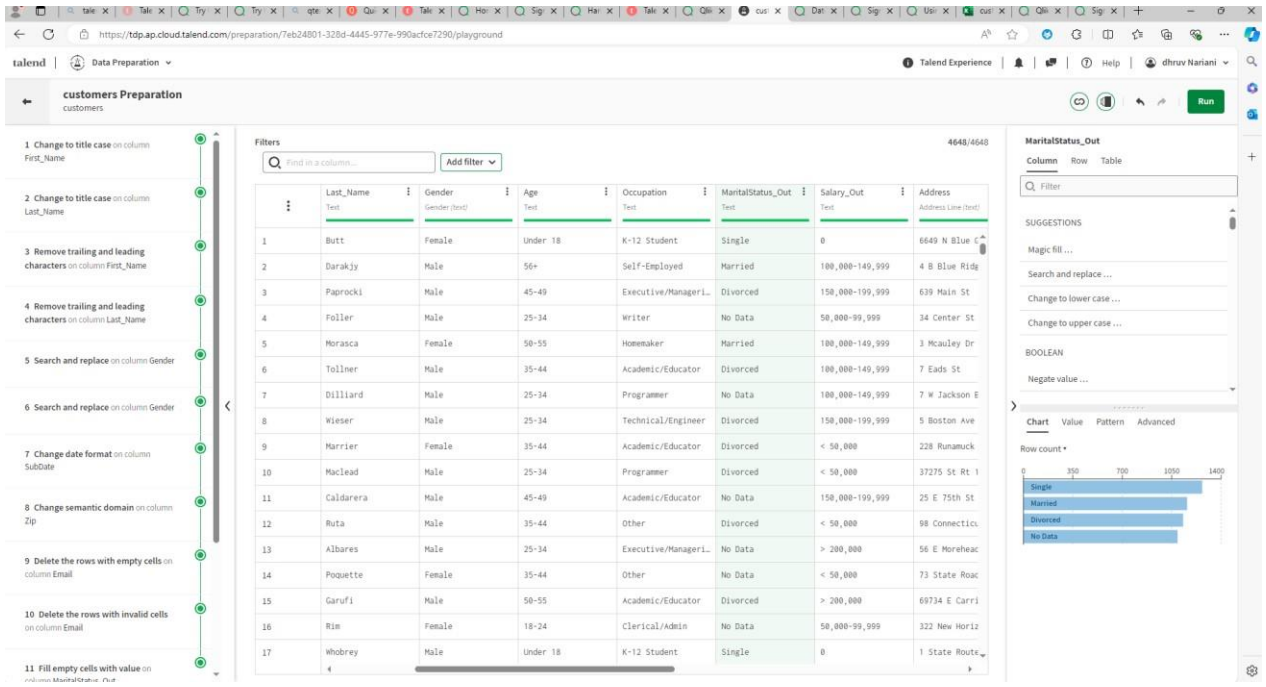
SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)



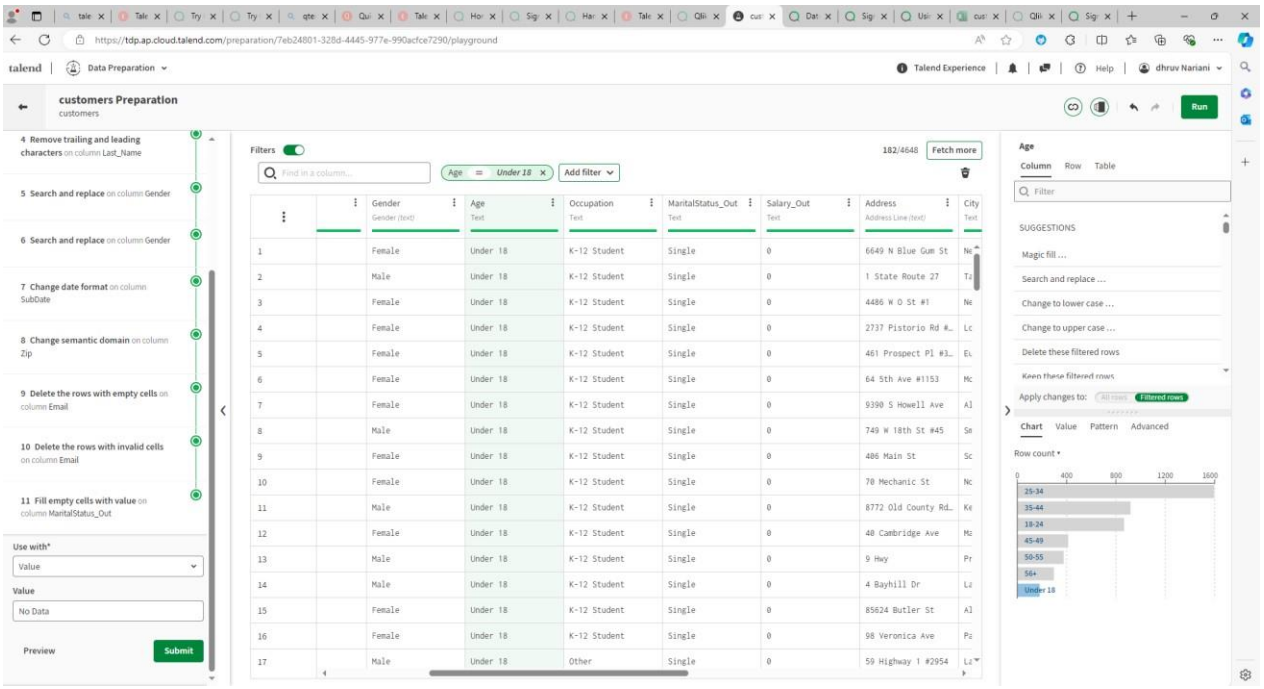


SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)





SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)



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