.Experiment Title.

FRAUD DETECTION IN THE INSURANCE BUSINESS

Student Name: Hiral Jain **UID: 180167**

Branch: IIIrd yr CSE **Section/Group: C Semester: 6th**

Date of Performance: 24/02/2021

Subject Name: Predictive Modelling Analytics

Subject Code: CS 18.338

1. Aim/Overview of the practical: To implement the **hypothesis 3** into the Fraud Detection **(Data Refinery Visualization)**.

2. Task to be done: Discuss How IBM Watson Studio empowers us to scale analysis across your org to speed dev time and simplify collaboration with data scientists, risk analysts, investigators, and other subject matter experts while adhering to strong governance and security posture. In order to respond to new types of fraud, waste and abuse while minimizing false negatives and accelerating response, the platform continuously accommodates real-time data, monitors and detects fraudulent activities and adapts as the patterns change and spot anomalies.

3. Apparatus (For applied/experimental sciences/materials based labs)**:** The following apparatus we need are:

- Internet connectivity
- IBM cloud account
- IBM WATSON service
- NIC Data Set

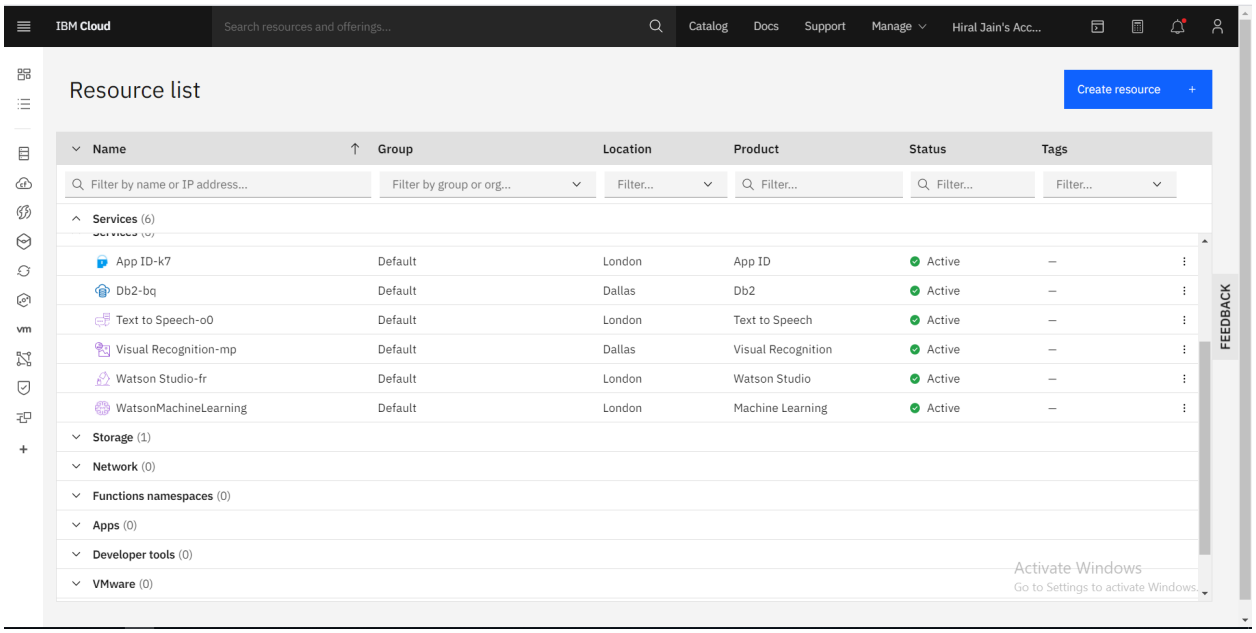
4. Hypothesis

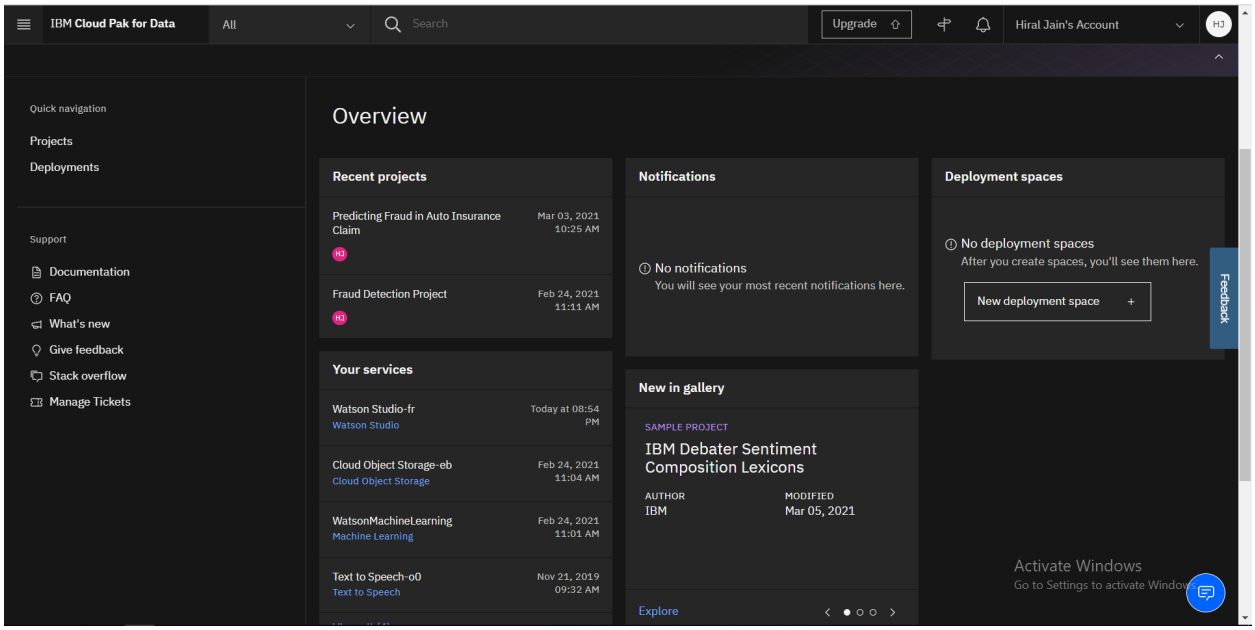
2. Claim filed after the license expiration date.
3. Excessive claim amount, which is over $10000 in value.

6. Steps for experiment/practical: The following steps are:

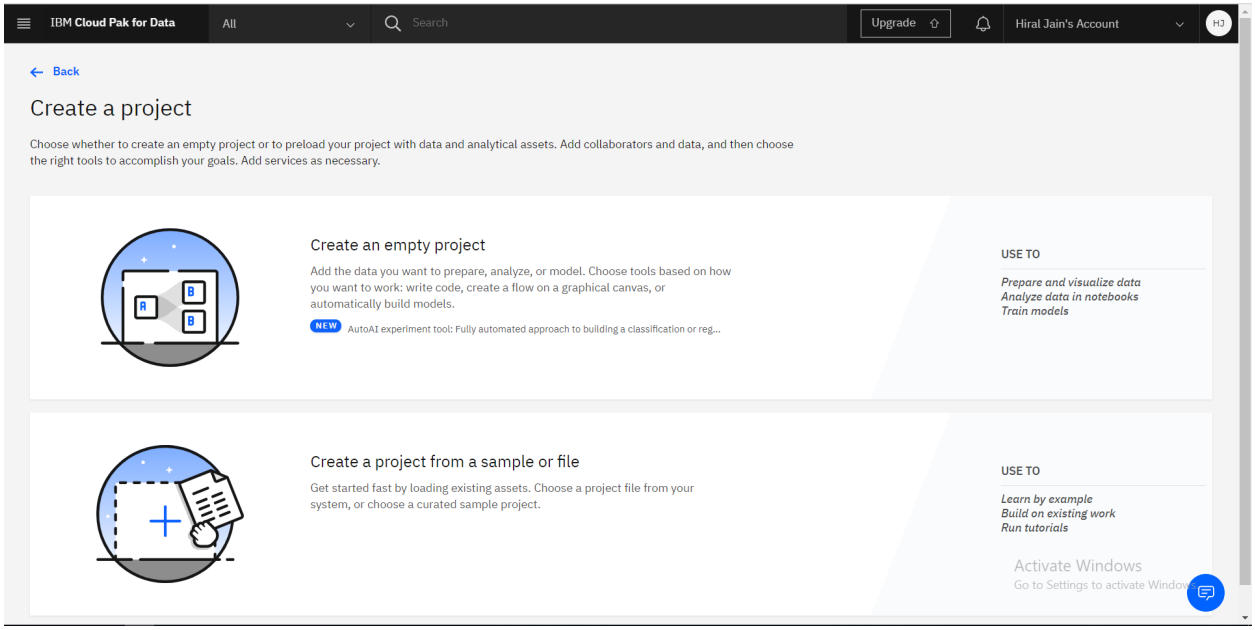
1. Login to your IBM Cloud account.

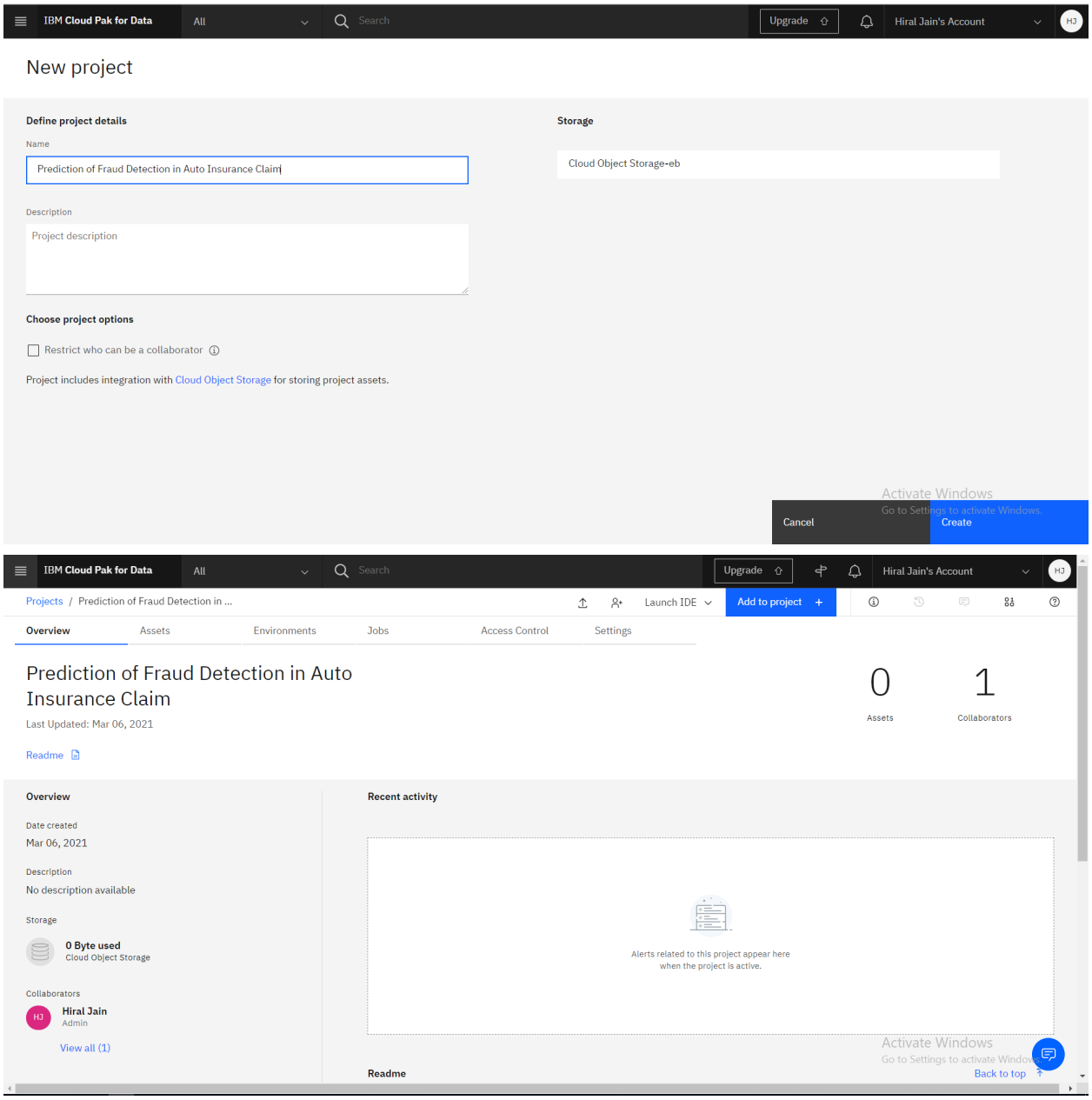
2. Go to your Watson Service and start a new project or start recent project which is already created.

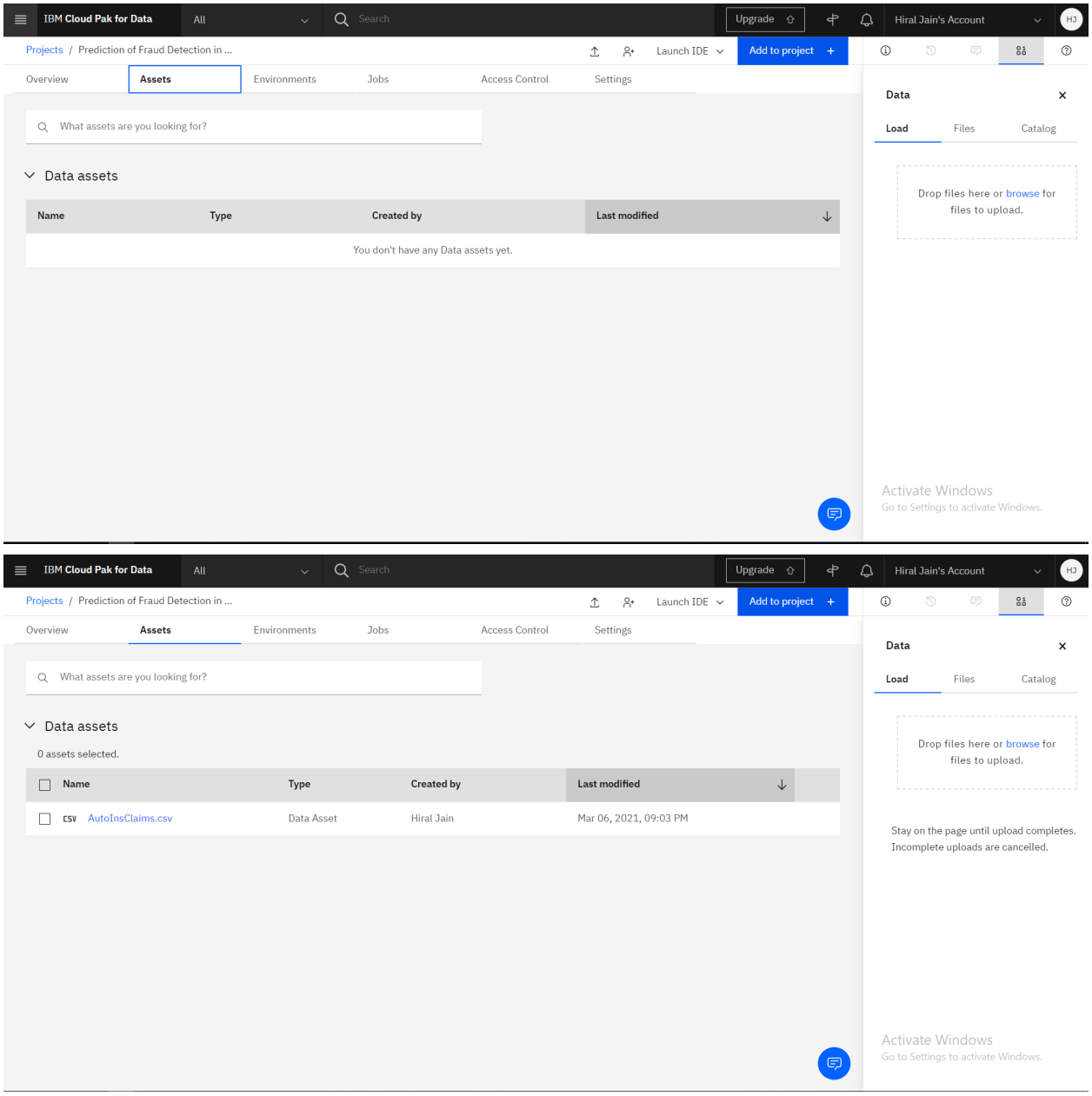




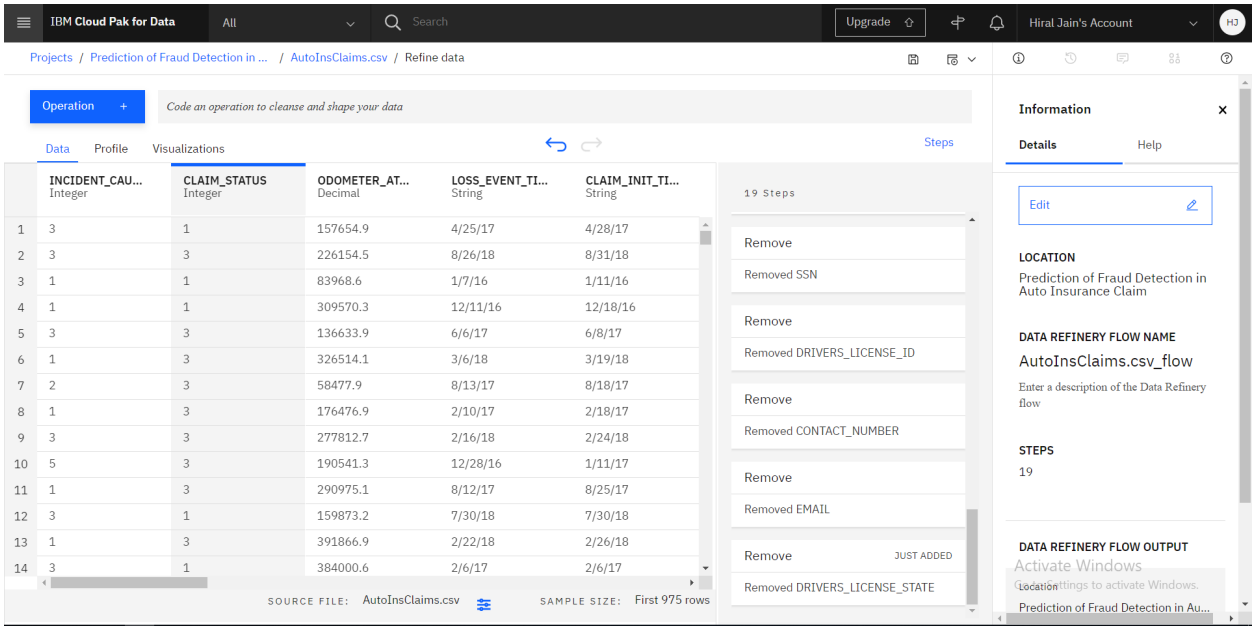
3. Load the AutoInsClaims dataset on the cloud.



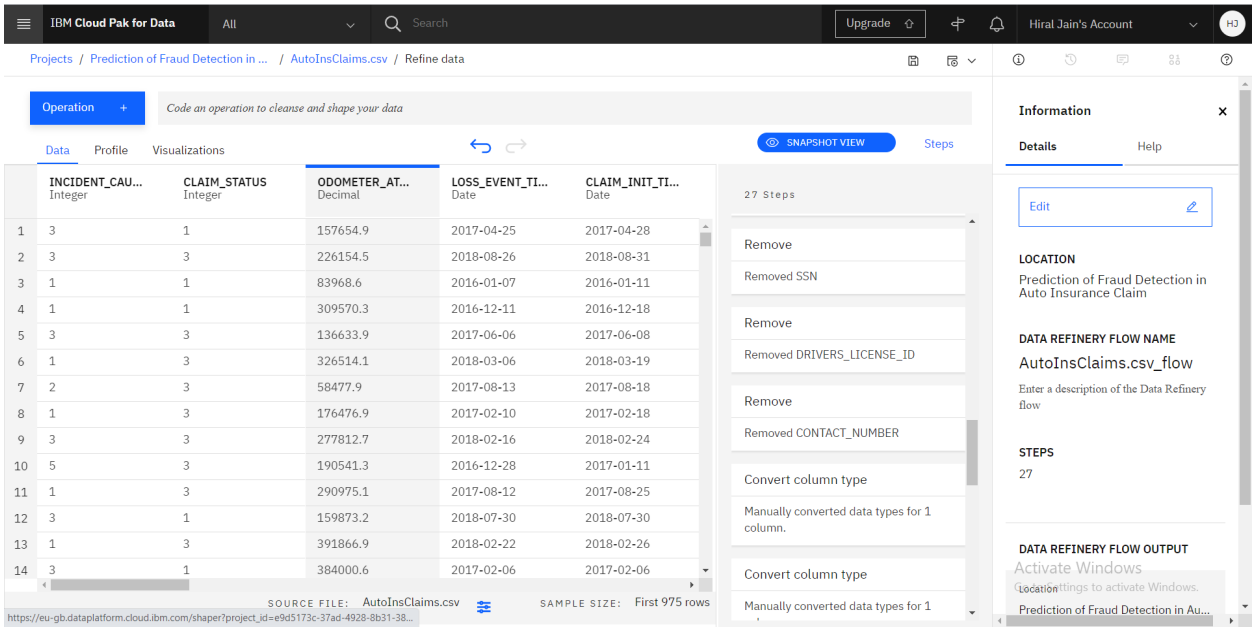

IBM Cloud Pak for Data
All
Search
Upgrade
Hiral Jain's Account
HJ
New project
Define project details
Name
Prediction of Fraud Detection in Auto Insurance Claim
Description
Project description
Choose project options
Restrict who can be a collaborator
Project includes integration with Cloud Object Storage for storing project assets.
Storage
Cloud Object Storage-eb
Activate Windows
Go to Settings to activate Windows.
Cancel
Create
IBM Cloud Pak for Data
All
Search
Upgrade
Hiral Jain's Account
HJ
Projects / Prediction of Fraud Detection in ...
Launch IDE
Add to project
Overview
Assets
Environments
Jobs
Access Control
Settings
Prediction of Fraud Detection in Auto Insurance Claim
Last Updated: Mar 06, 2021
Readme
0
Assets
1
Collaborators
Overview
Date created
Mar 06, 2021
Description
No description available
Storage
0 Byte used
Cloud Object Storage
Collaborators
HJ
Hiral Jain
Admin
View all (1)
Recent activity
Alerts related to this project appear here when the project is active.
Activate Windows
Readme
Back to top

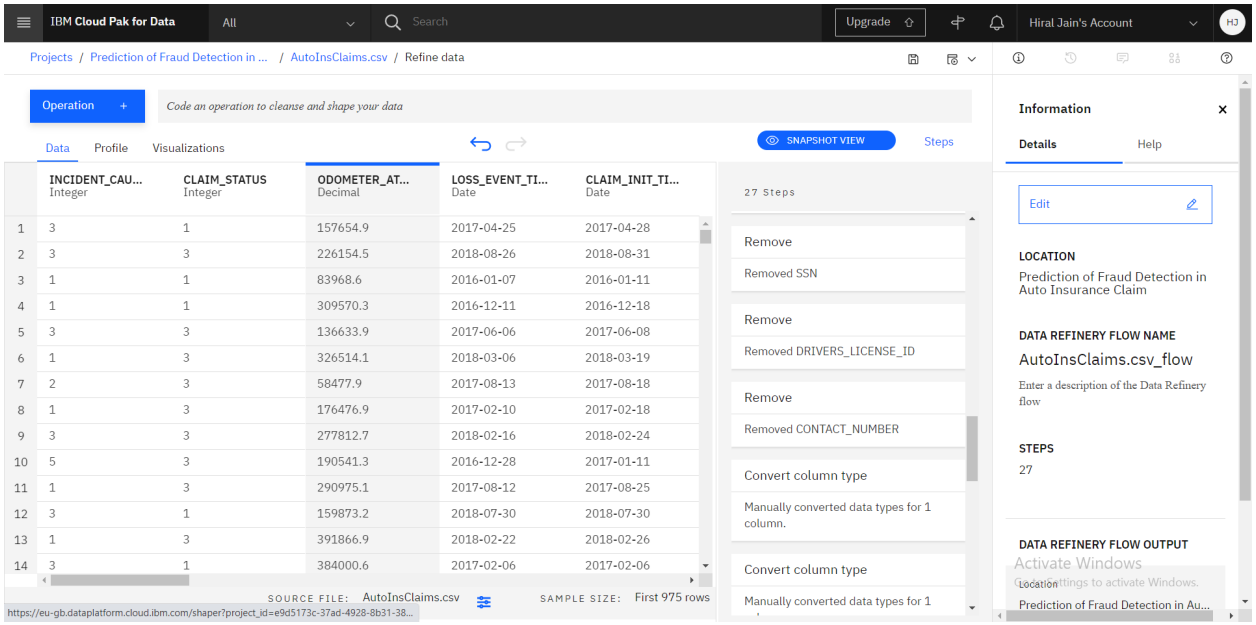


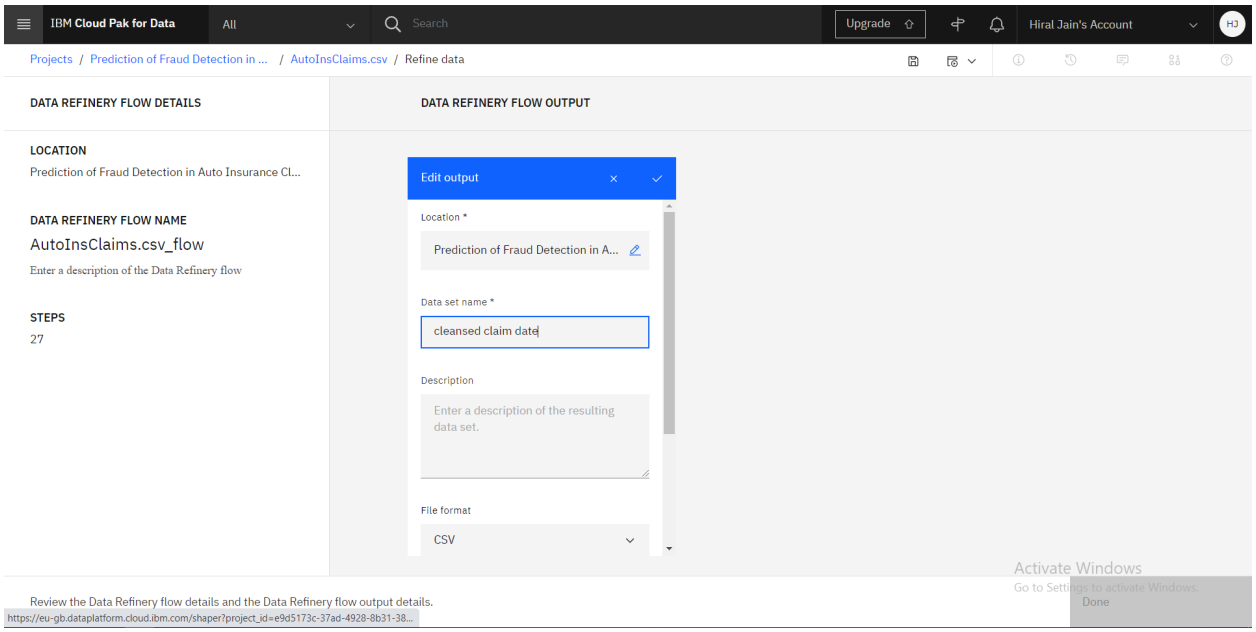
4. Clean the data, click on REFINE. Delete the following columns as they are not required for our hypothesis- household_id, driver_id, policy_id, claim_id, description, primary_driver_id, model_year, make, model, plate, color, first_name, ssn, last_name, driver_license_id, contact_number, e-mail, driver_license_state.



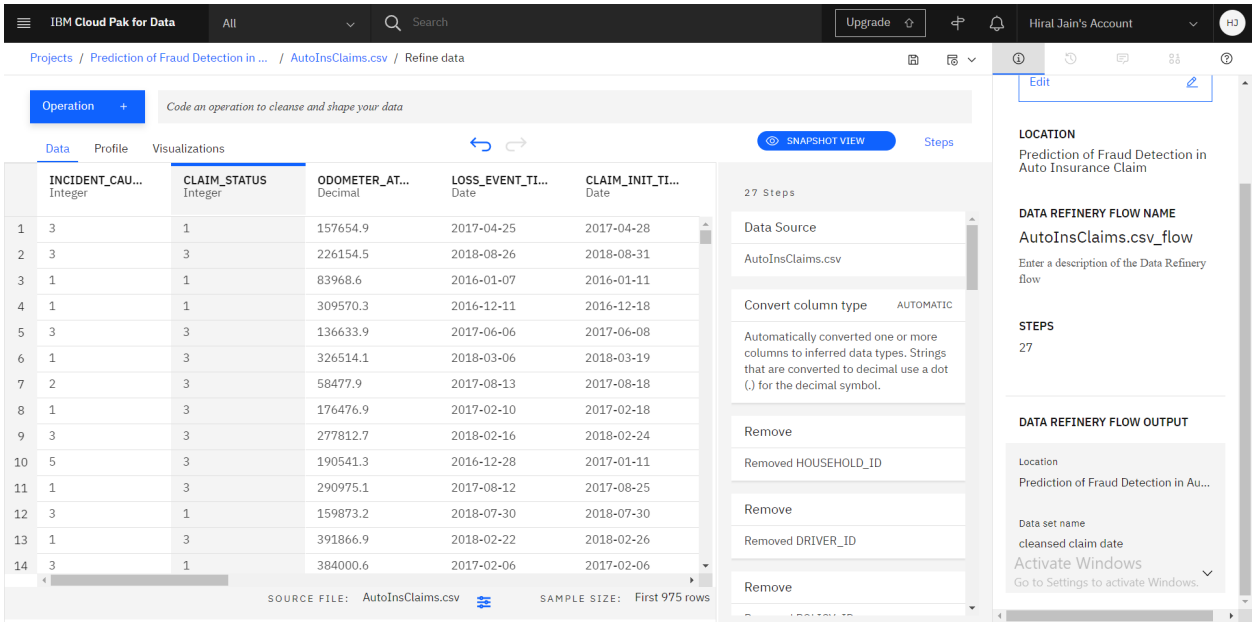
5. Convert all the date columns into MDY format. For the same, select the column, click on convert and select date, select current order as (mdy) and apply.



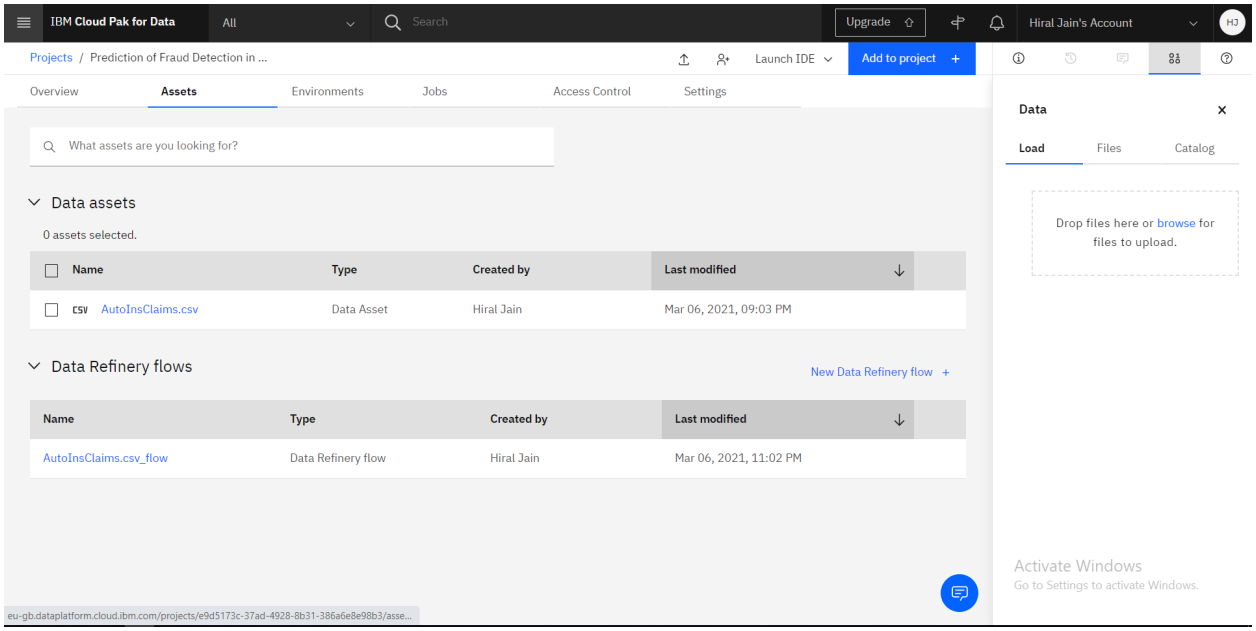


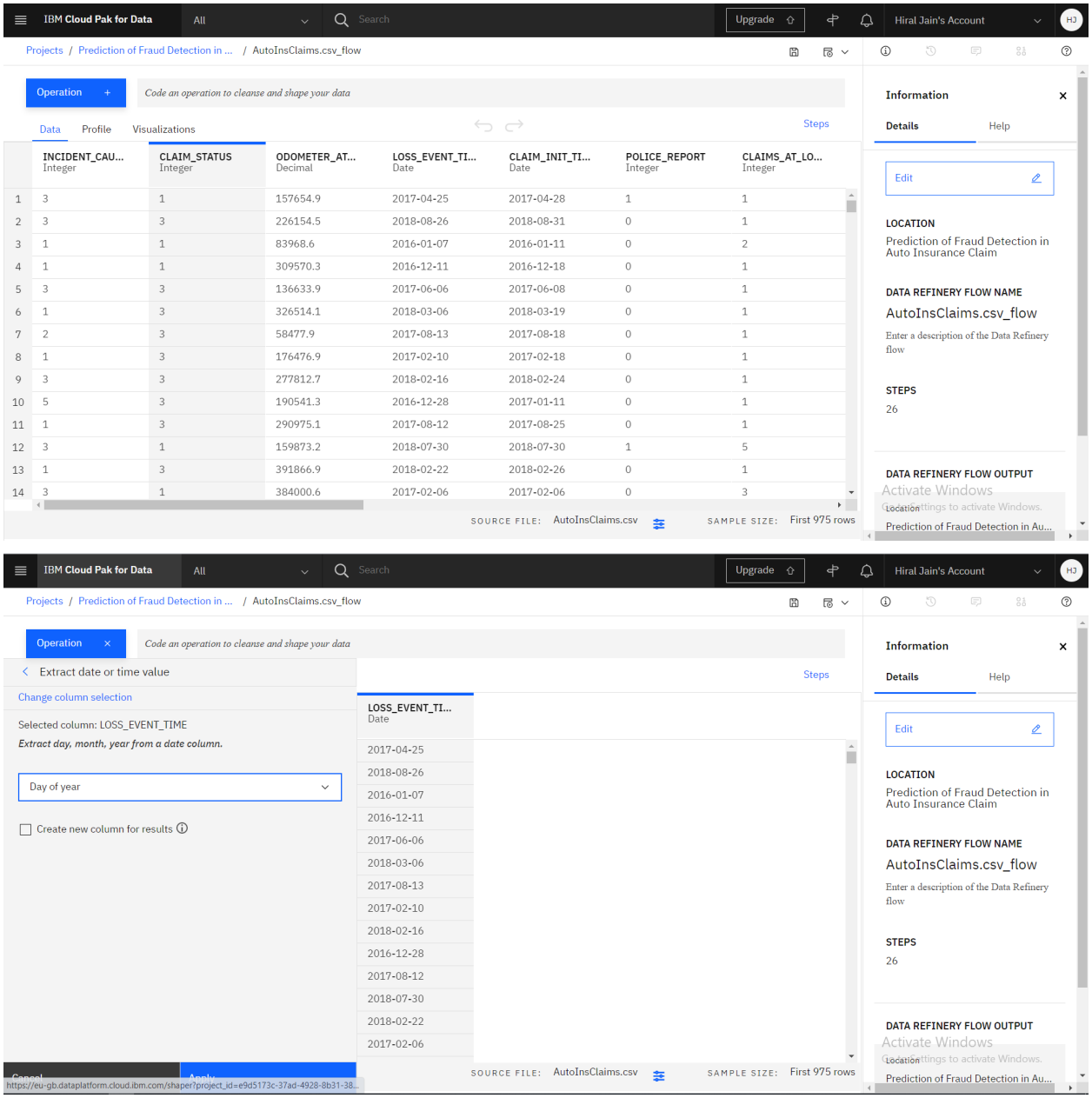


6. Save the refined data, click on Details then click on edit. Click on edit output and name the data set as (cleansed claim date.csv). click on done.

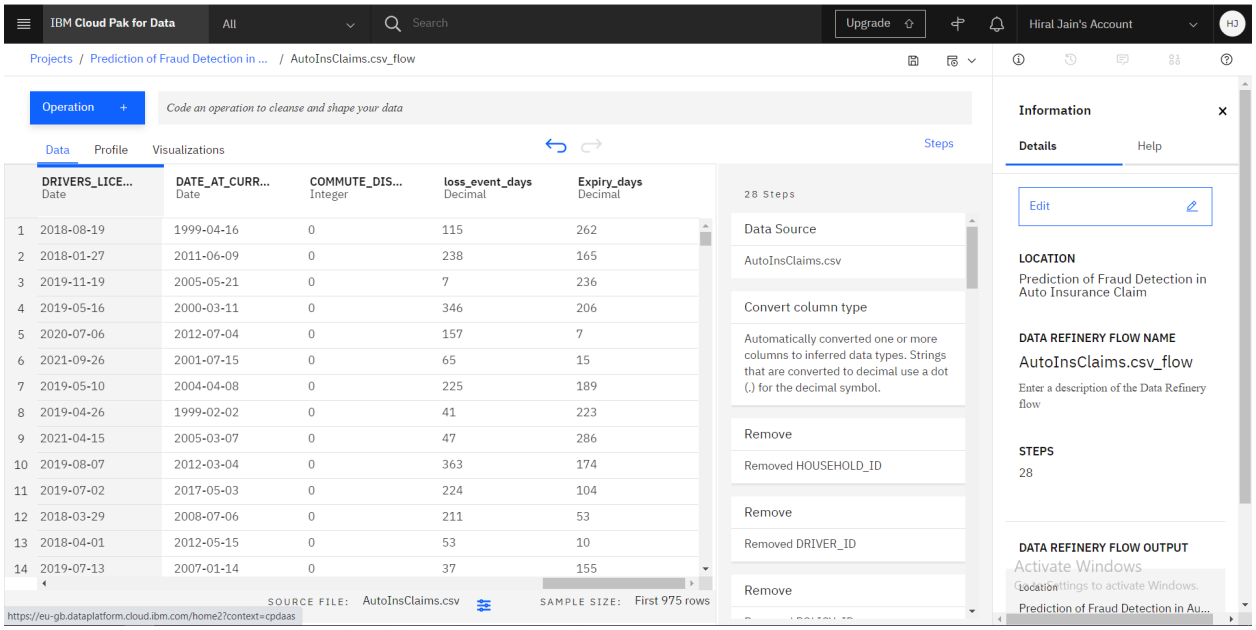


7. Go to Data Refinery Flow and select AutoInsClaims.csv_flow. Select Loss_event_time column, go on operations and select ext (extract date or time value), select Day Of The Year and create a new column with column name as (loss_event_days) and click on apply. Save the data.





8. Similarly do for expiry_date column and name the new column as (expiry_days).



2. Claim filed after the license expiration date

This hypothesis tells if the claim for the car was filled till the validity of driver's license, or if it was filed after the expiration of driver's license. Steps for the hypothesis-

- Select driver_license_expiry column and add mutate operation. Select code

  Mutate (provide_new_column='<column>'<operator>'<column>').

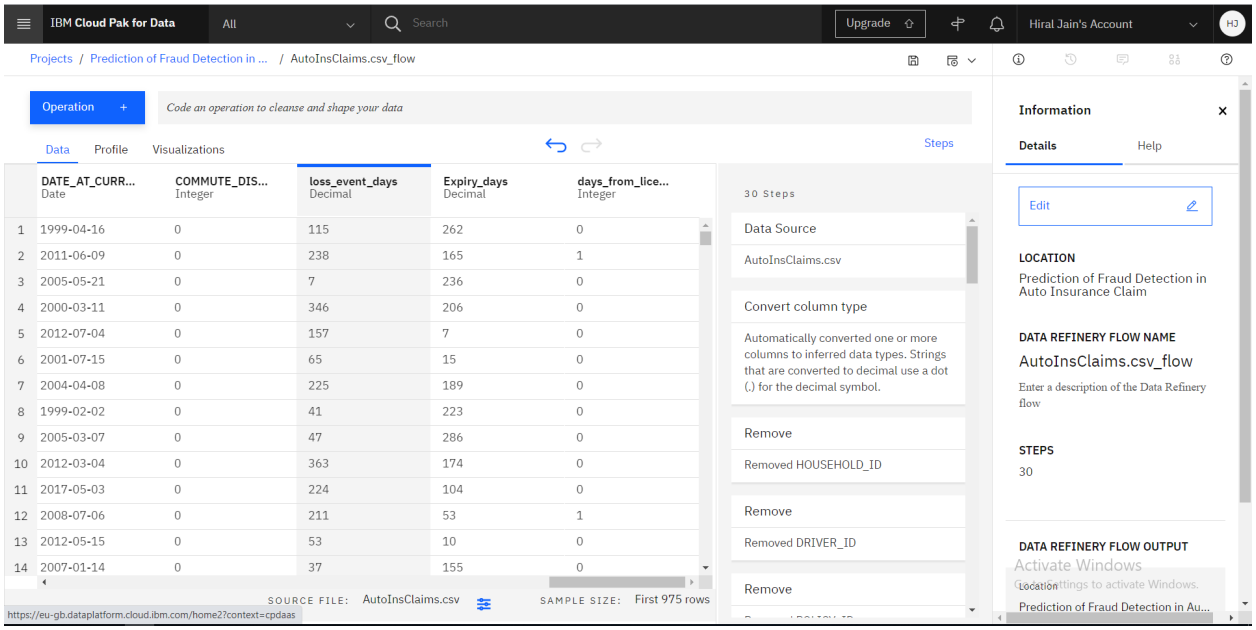
  Provide_new_column = days_from_license_expiry

  Column = loss_event_time

  Operator = >

  Column = driver_license_expiry

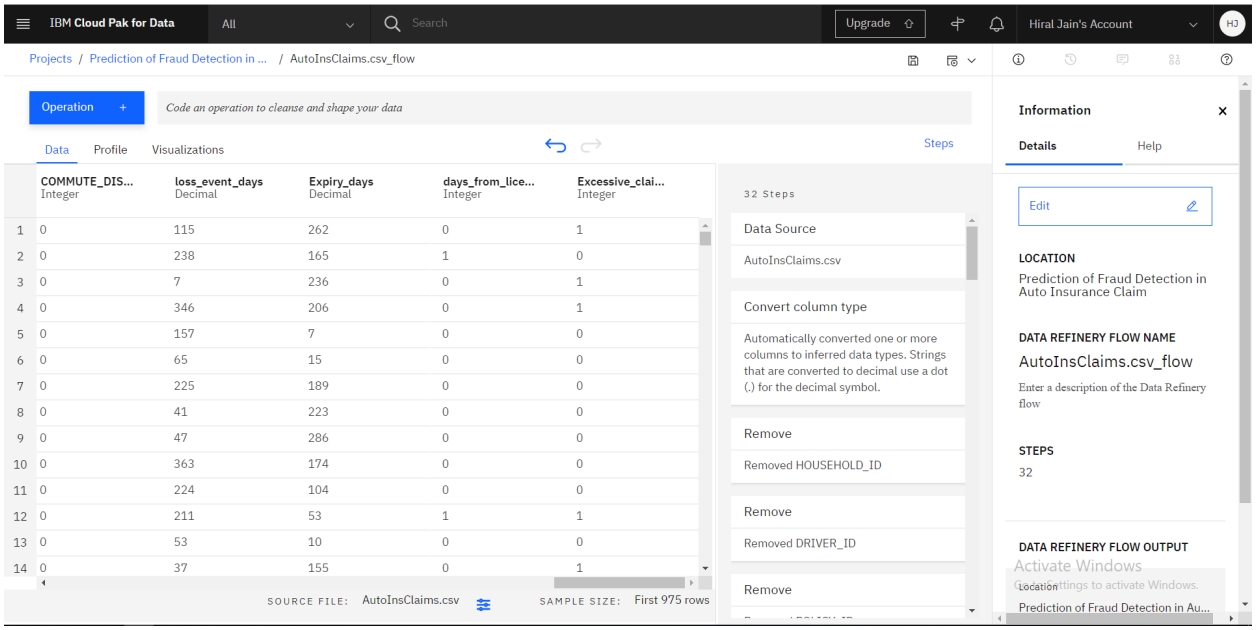
  Apply. You will receive output in Boolean. Convert the Boolean type into Integer type.



3. Excessive claim amount, which is over $10000 in value

This hypothesis says that the car accidents who claim for over $10000 will not be given the entire claim amount. But the claims under $10000 will be fully claimed. Steps for the hypothesis-

- Select claim_amount column and select calculate operator. Select the greater than operator and enter the value as 10000. Create a new column with column name as (excessive_claim_amount). • Convert the column into Integer type of data from Boolean type.

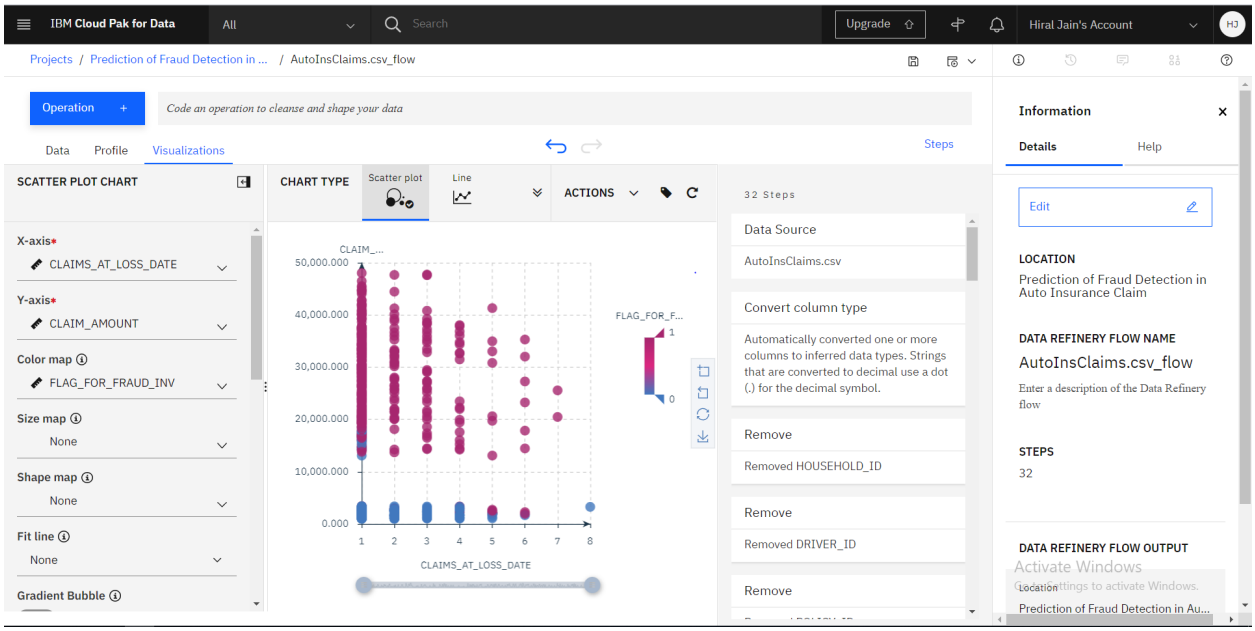


4. Data Refinery Visualization-

From the Data Refinery flow,click the **Visualization** tab.

1. **CLAIM_AT_LOSS_DATE**

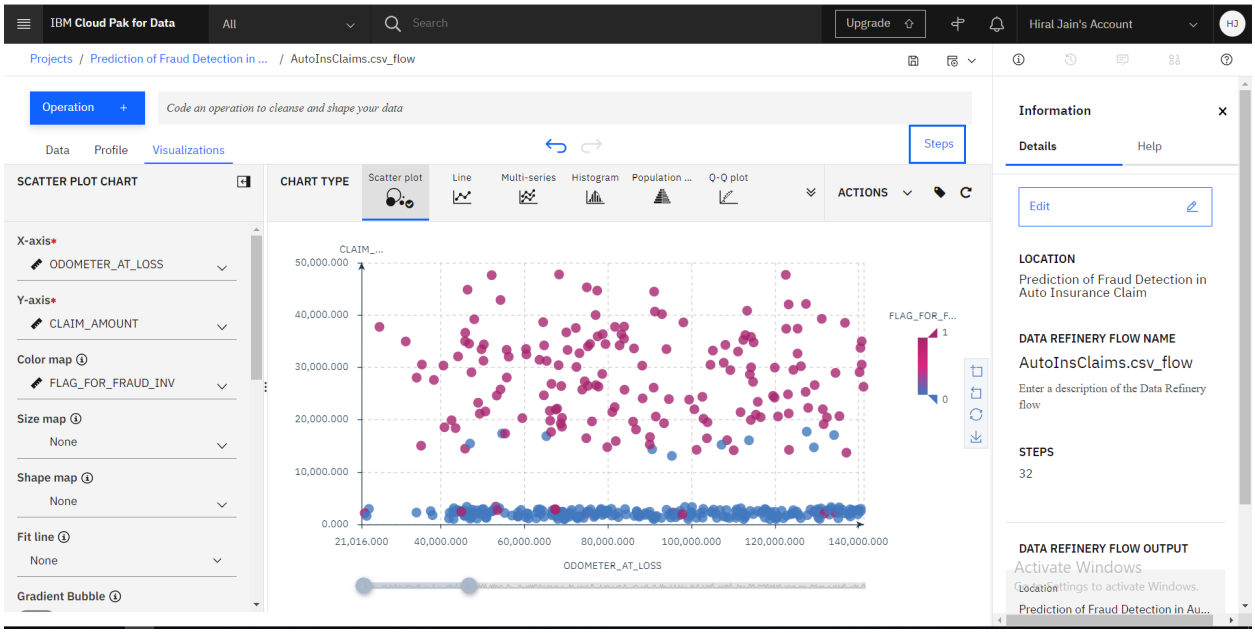
   Select the scatter plot and for the X-axis select CLAIMS_AT_LOSS_DATE, for the Y-axis select CLAIM_AMOUNT and for the Color map, select FLAG_FOR_FRAUD_INV.



2. **ODOMETER_AT_LOSS**

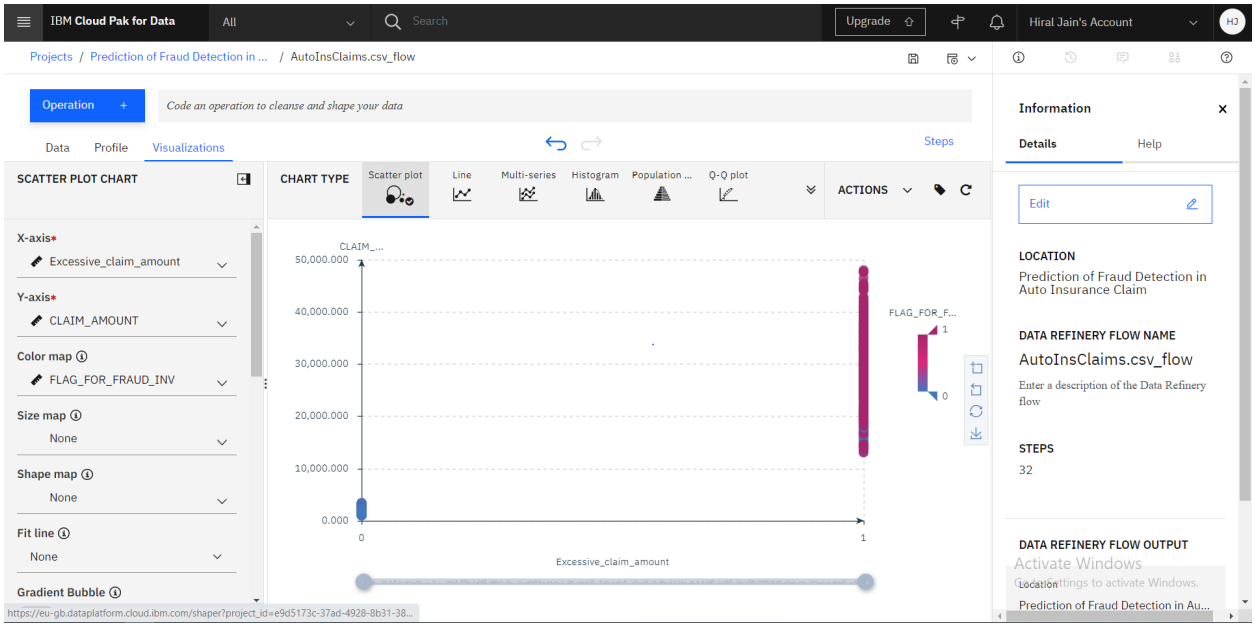
   Select the scatter plot and for the X-axis select ODOMETER_AT_LOSS, for the Y-axis select CLAIM_AMOUNT and for the Color map, select

FLAG_FOR_FRAUD_INV.



3. EXCESSIVE CLAIM AMOUNT

Change the X-axis to EXCESSIVE_CLAIM_AMOUNT and keep the remaining parameters the same.



7. Observations/Discussions (For applied/experimental sciences/materials-based labs): After conducting the above practical, we came to know about the IBM Watson service. We were able to generate the hypothesis3(Visualization) for our data analysis according to the requirement of the customer.

Learning outcomes (What I have learnt): After conducting the practical, we now know how to work on IBM Cloud and its Watson service. We came to know how to work on a data set, how to clean the data and analyse the data.

Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.			
2.			
3.			