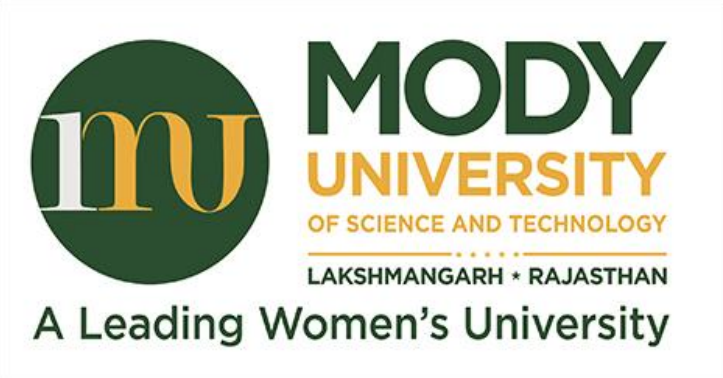



# Experiment Title

## FRAUD DETECTION IN THE INSURANCE BUSINESS

**Student Name: Akshita Kanther** **UID: 180161**

**Branch: IIIrd yr CSE** **Section/Group: C**

**Semester: 6th**

**Date of Performance: 22/02/2021**

**Subject Name: Predictive Modelling Analytics**

**Subject Code: CS 18.338**

**I. Aim/Overview of the practical:** To implement the **hypothesis1** into the Fraud Detection **(Loss claim after Expired Policy)**.

**II. 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.

**III. 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

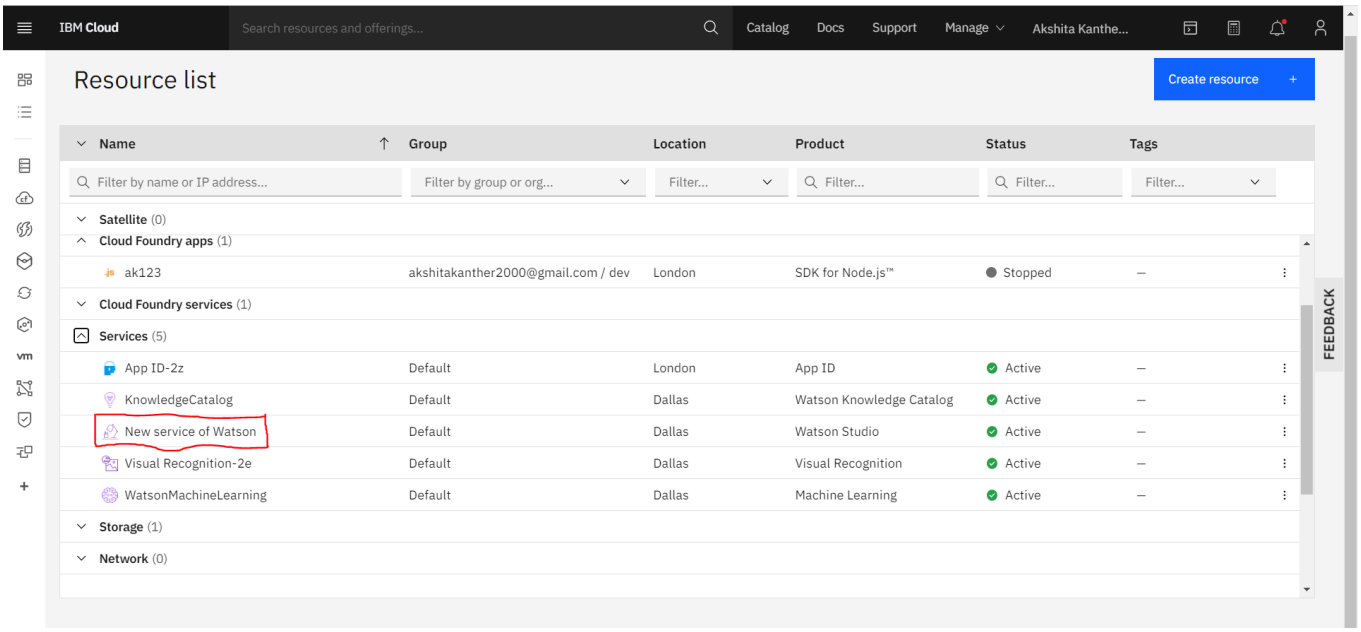
## IV. Hypothesis

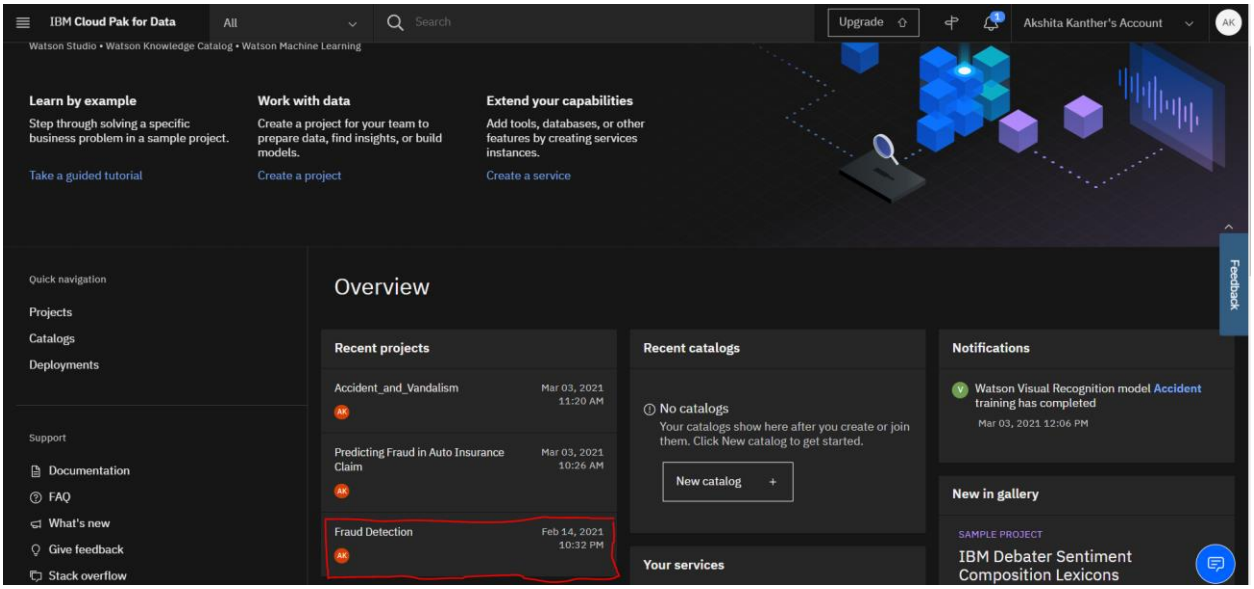
2. Claim filed after the license expiration date.
3. Excessive claim amount, which is over $10000 in value.

## V. Steps for experiment/practical: The following steps are:

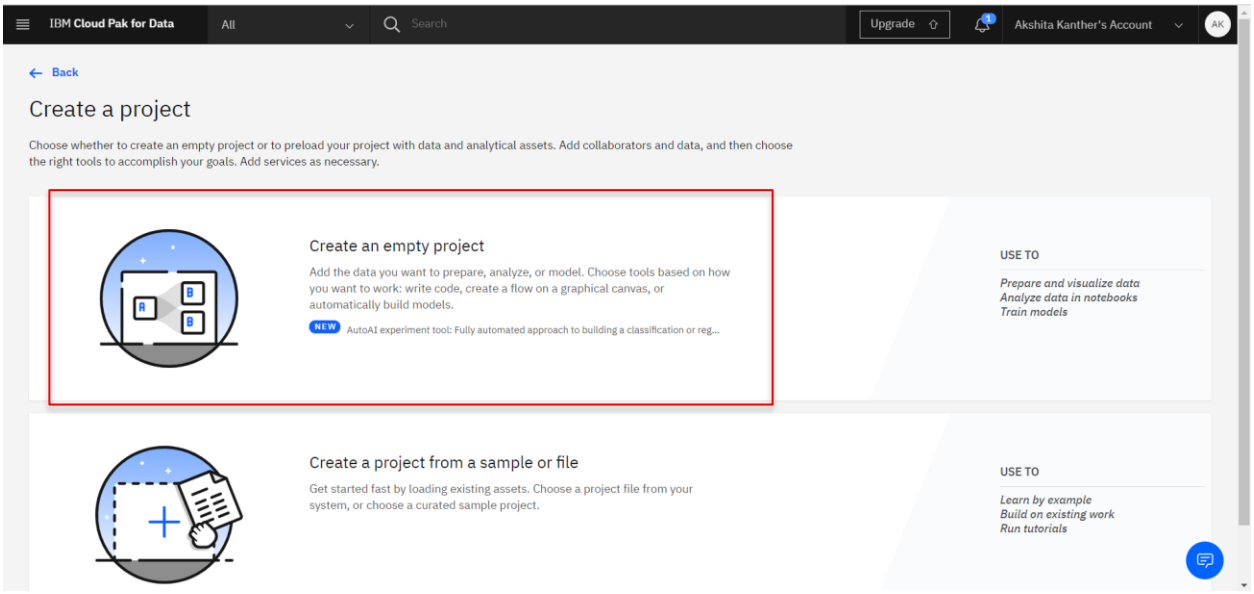
**1.** Login to your IBM Cloud account.

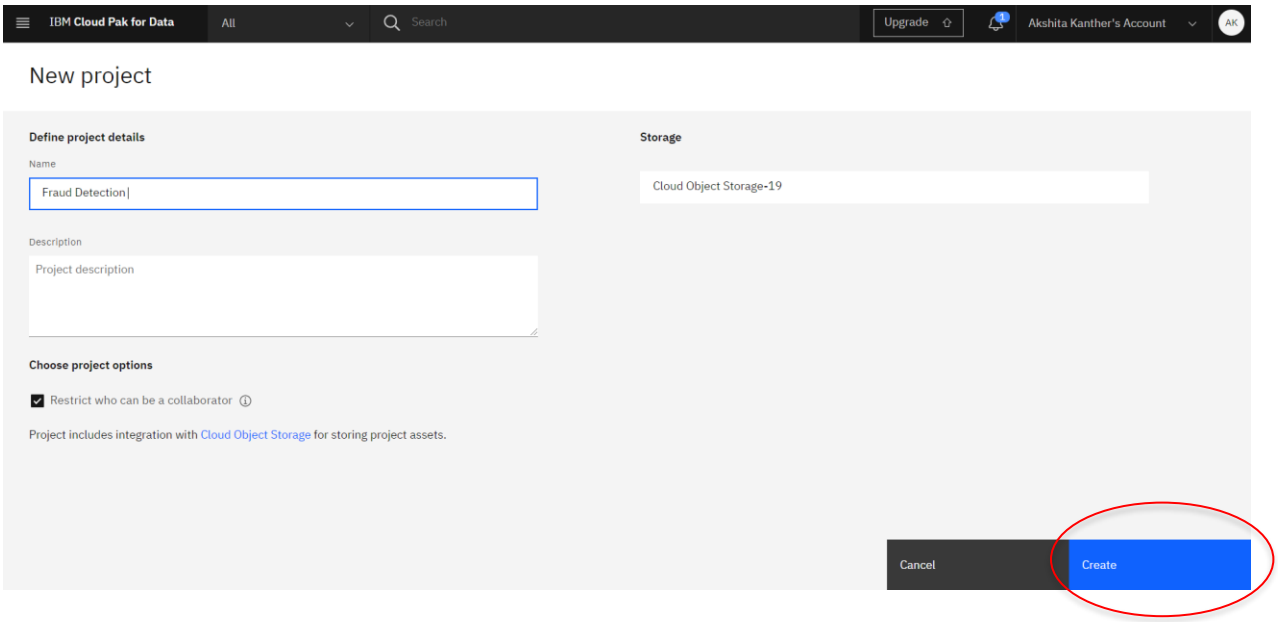
**2.** Go to your Watson Service and start a new project or start a new project which is already created.

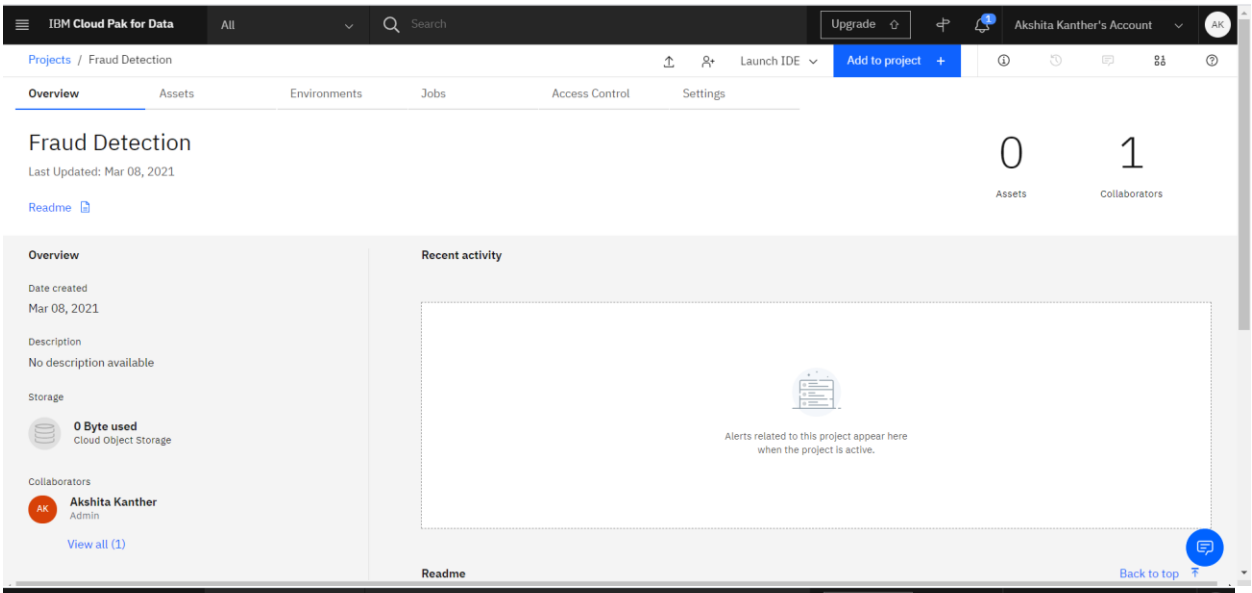


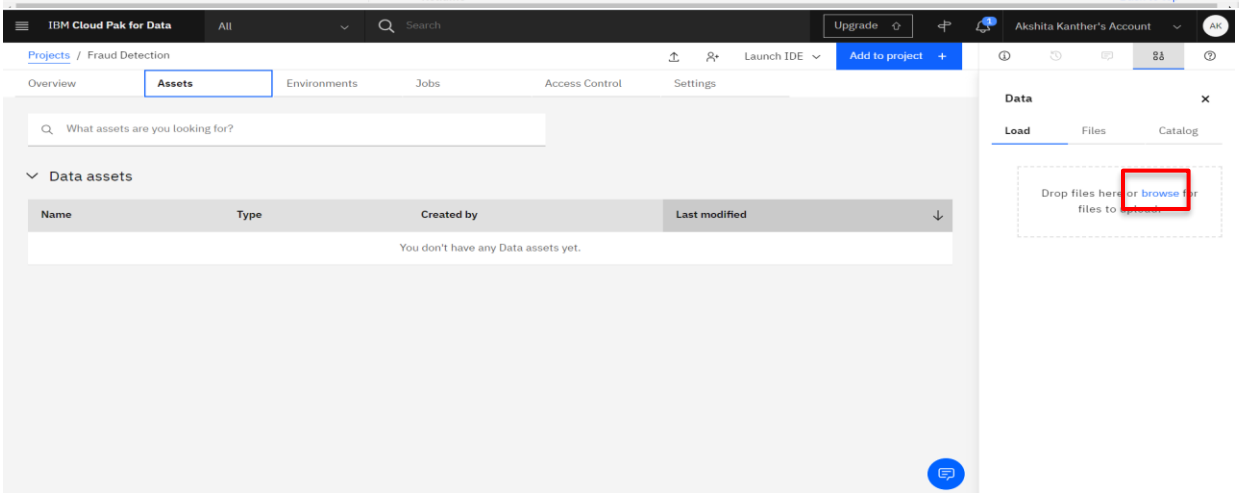


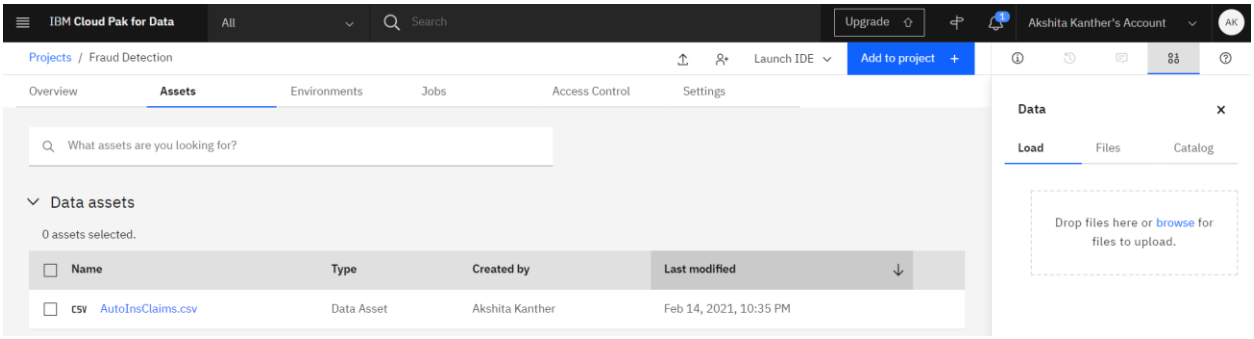

**3.** Load the AutoInsClaims dataset on the cloud.



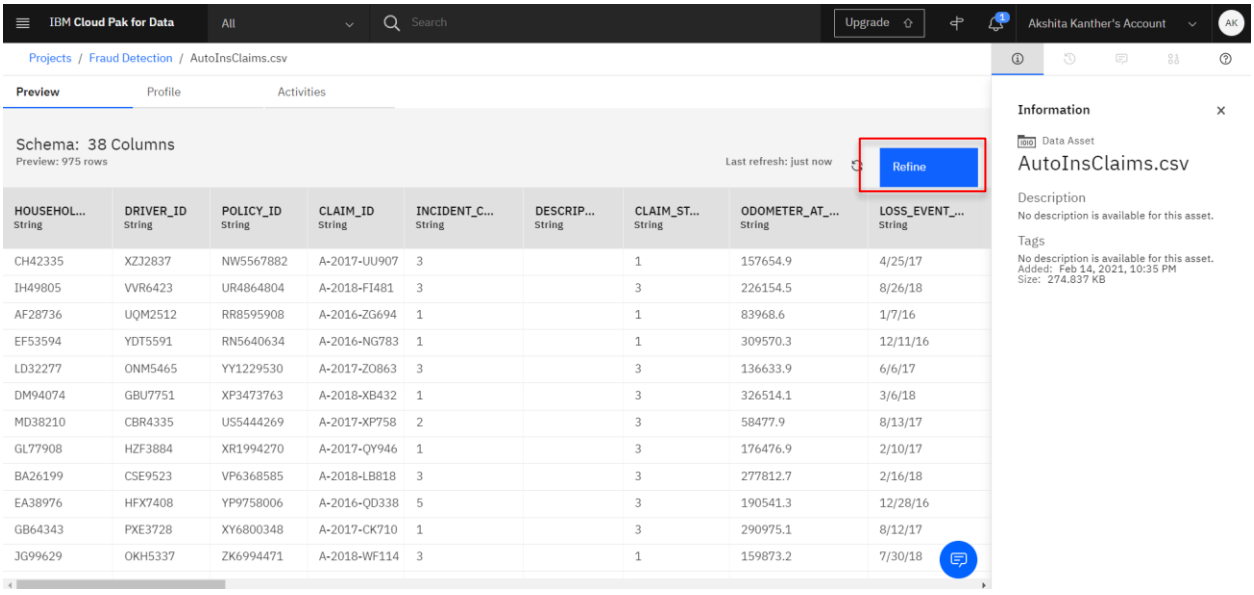

IBM Cloud Pak for Data
All
Search
Upgrade
Akshita Kanther's Account
AK
New project
Define project details
Name
Fraud Detection
Description
Project description
Storage
Cloud Object Storage-19
Choose project options
Restrict who can be a collaborator
Project includes integration with Cloud Object Storage for storing project assets.
Cancel
Create

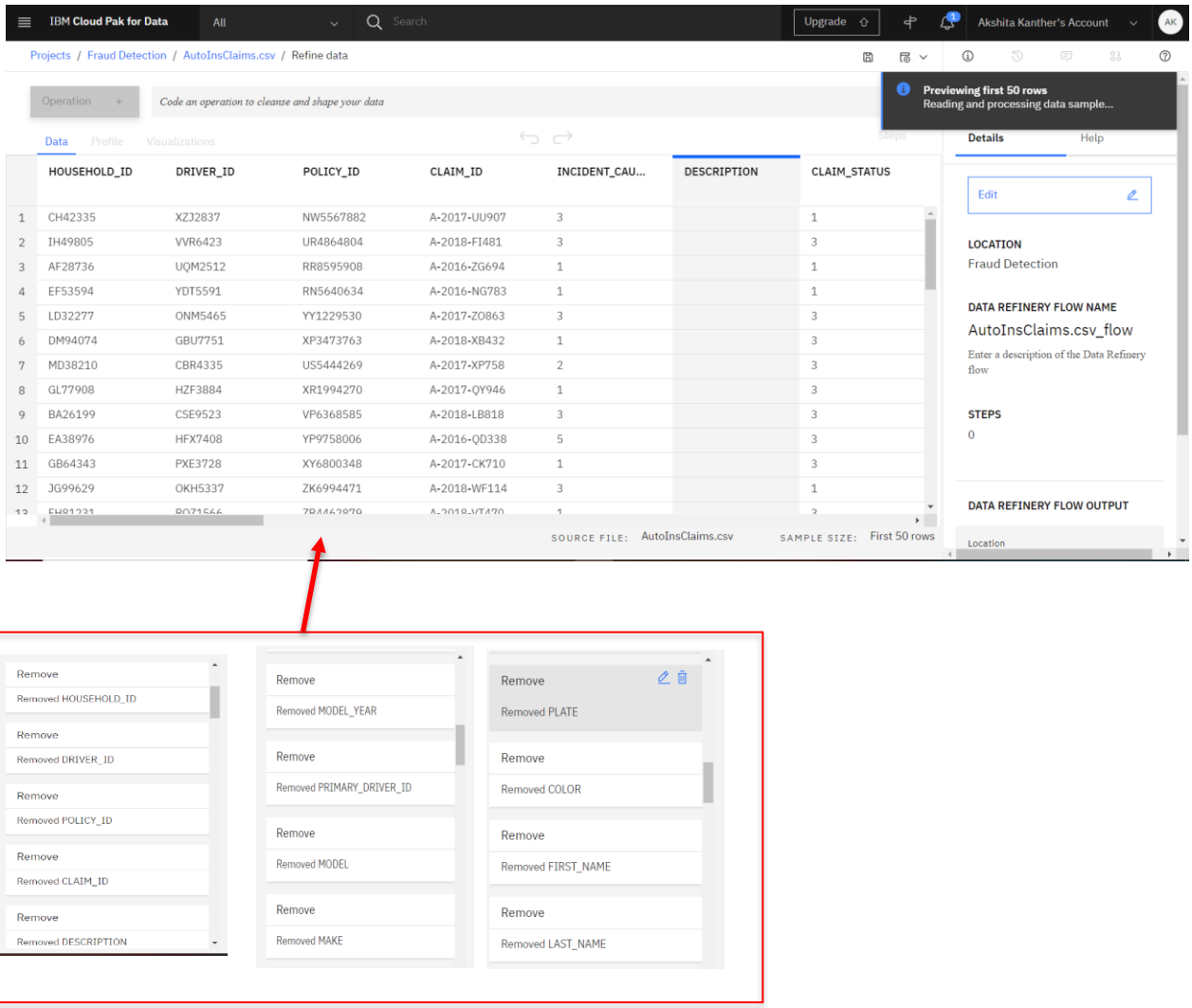

IBM Cloud Pak for Data
All
Search
Upgrade
Akshita Kanther's Account
AK
Projects / Fraud Detection
Launch IDE
Add to project
Overview
Assets
Environments
Jobs
Access Control
Settings
Fraud Detection
Last Updated: Mar 08, 2021
Readme
0
Assets
1
Collaborators
Overview
Date created
Mar 08, 2021
Description
No description available
Storage
0 Byte used
Cloud Object Storage
Collaborators
AK
Akshita Kanther
Admin
View all (1)
Recent activity
Alerts related to this project appear here
when the project is active.
Readme
Back to top


IBM Cloud Pak for Data
All
Search
Upgrade
Akshita Kanther's Account
AK
Projects / Fraud Detection
Launch IDE
Add to project
Overview
Assets
Environments
Jobs
Access Control
Settings
What assets are you looking for?
Data assets
Name
Type
Created by
Last modified
You don't have any Data assets yet.
Data
Load
Files
Catalog
Drop files here or browse for
files to upload

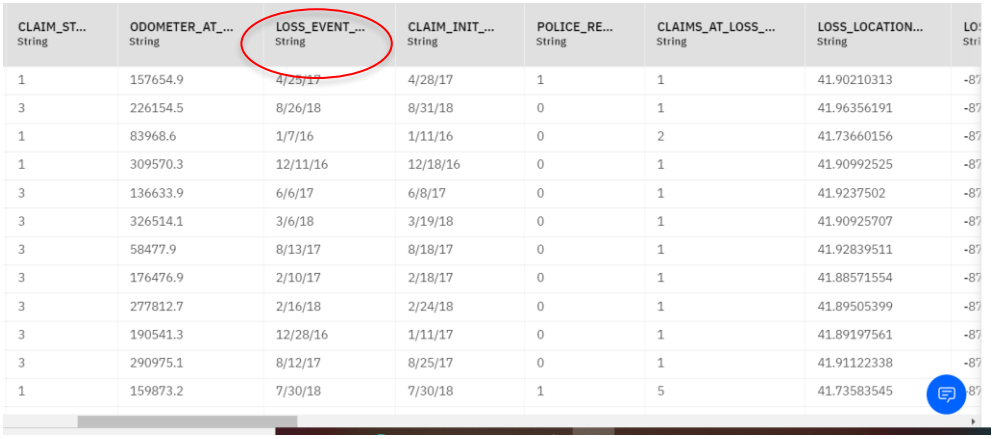


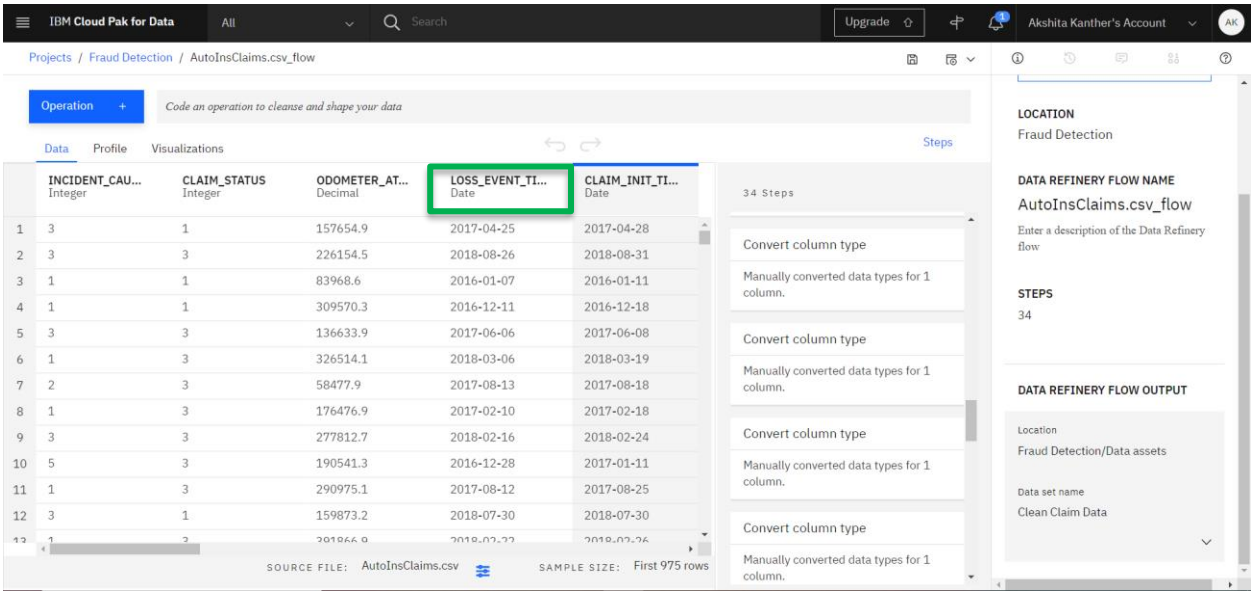

**4.** Clean the data, click on REFINE. Delete the following columns as they are not required for our hypothesis- household_id, driver_id, policy_id, claim_id, description, primary_driver_id, model_year, make, model, plate, color, first_name, ssn, last_name, driver_license_id, contact_number, e-mail,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



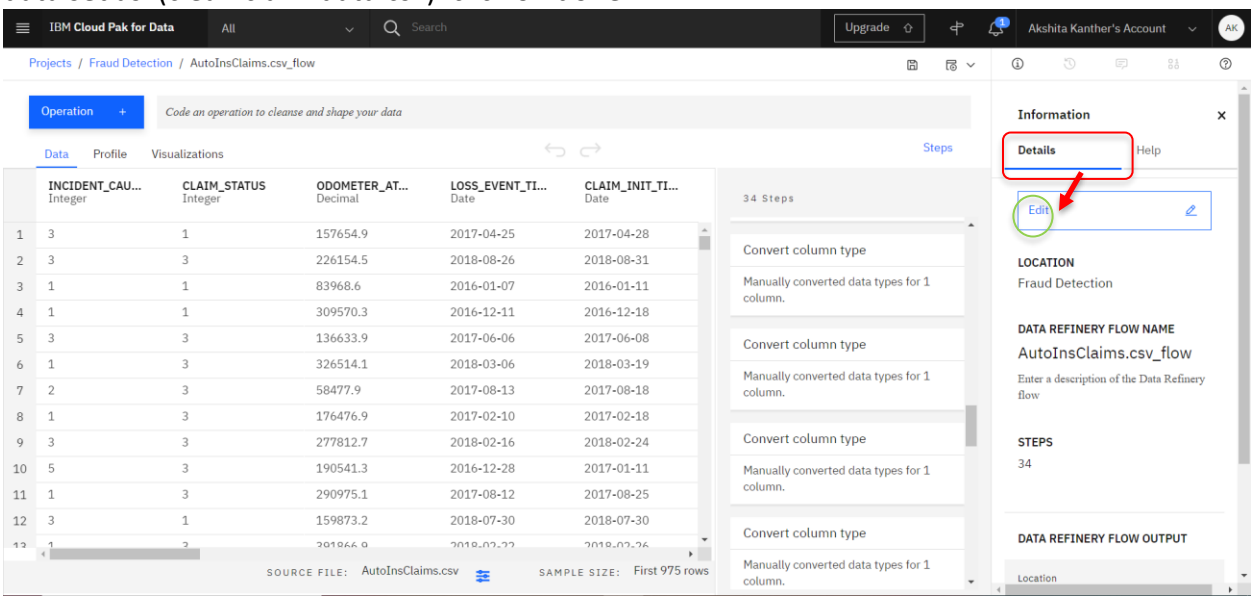


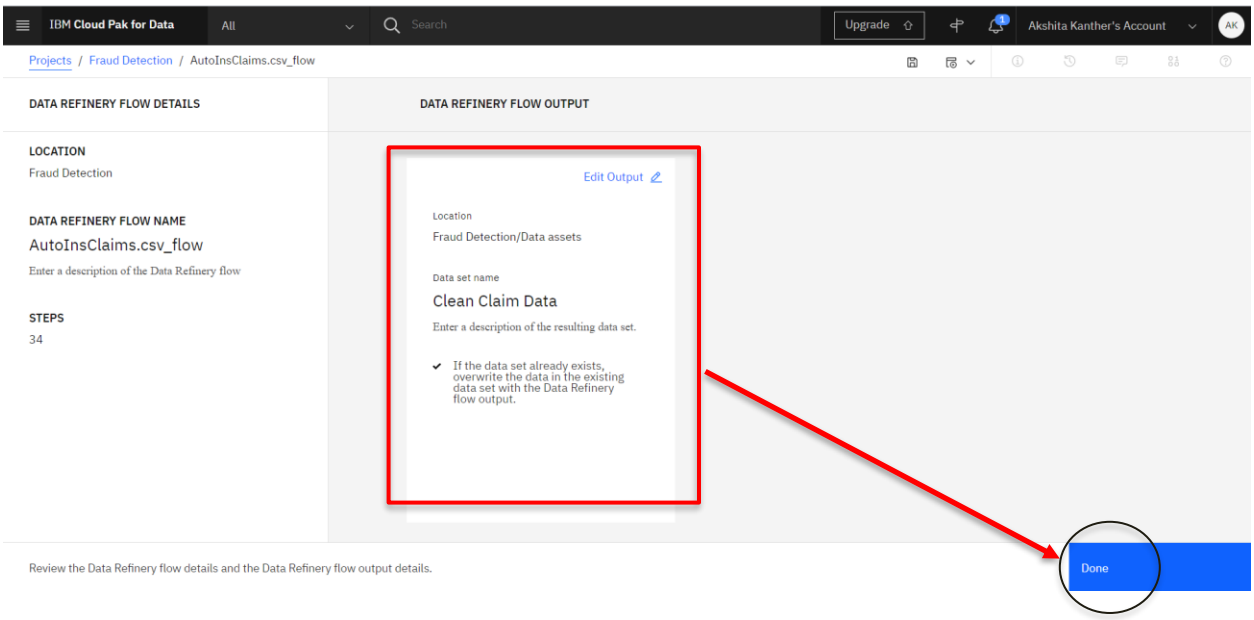

**5.** Convert all the date columns into MDY format. For the same, select the column, click on convert and select date, select current order as (mdy) and apply.



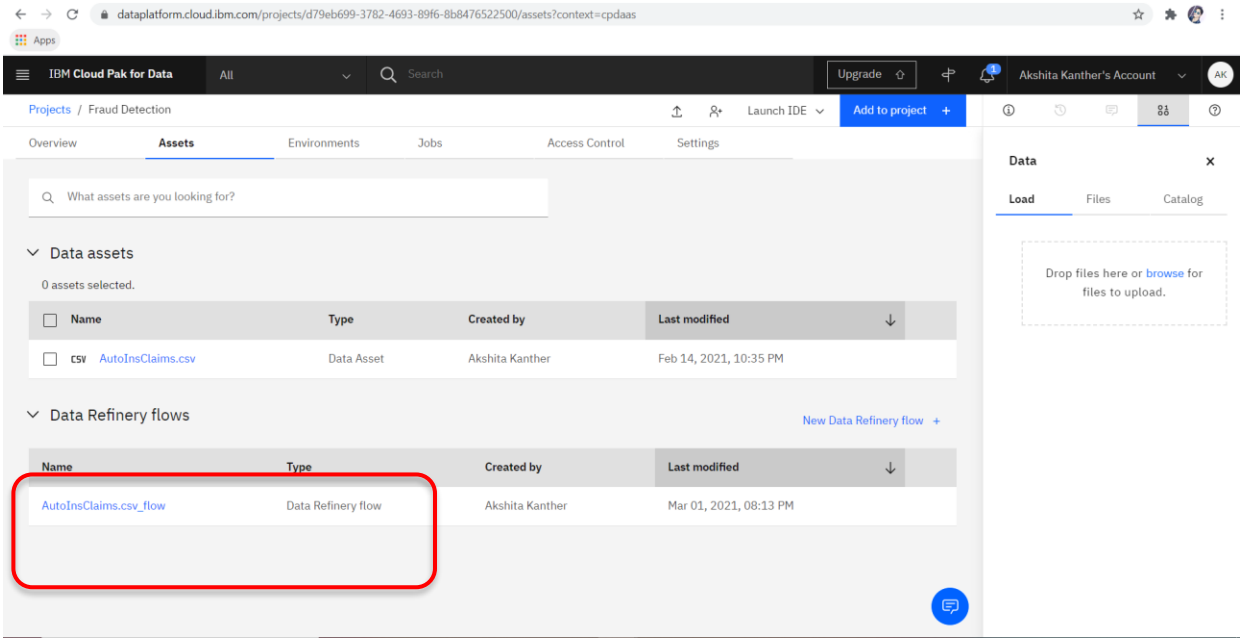


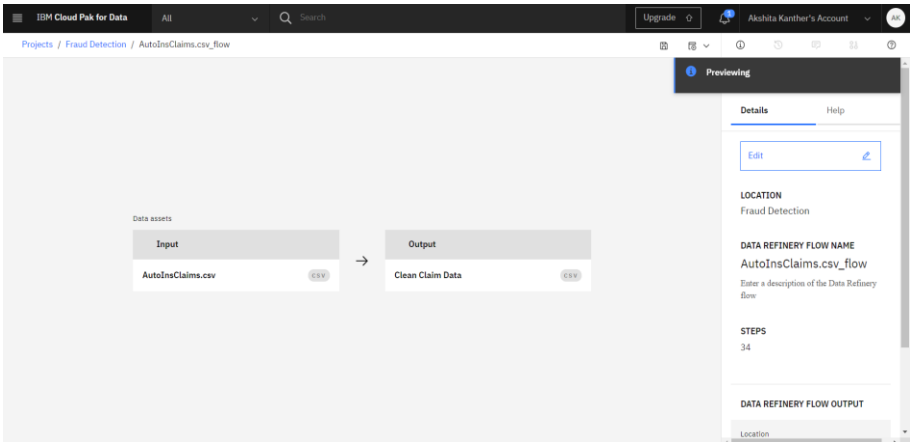

**6.** Save the refined data, click on Details then click on edit. Click on edit output and name the data set as  (clean claim data.csv). click on done.

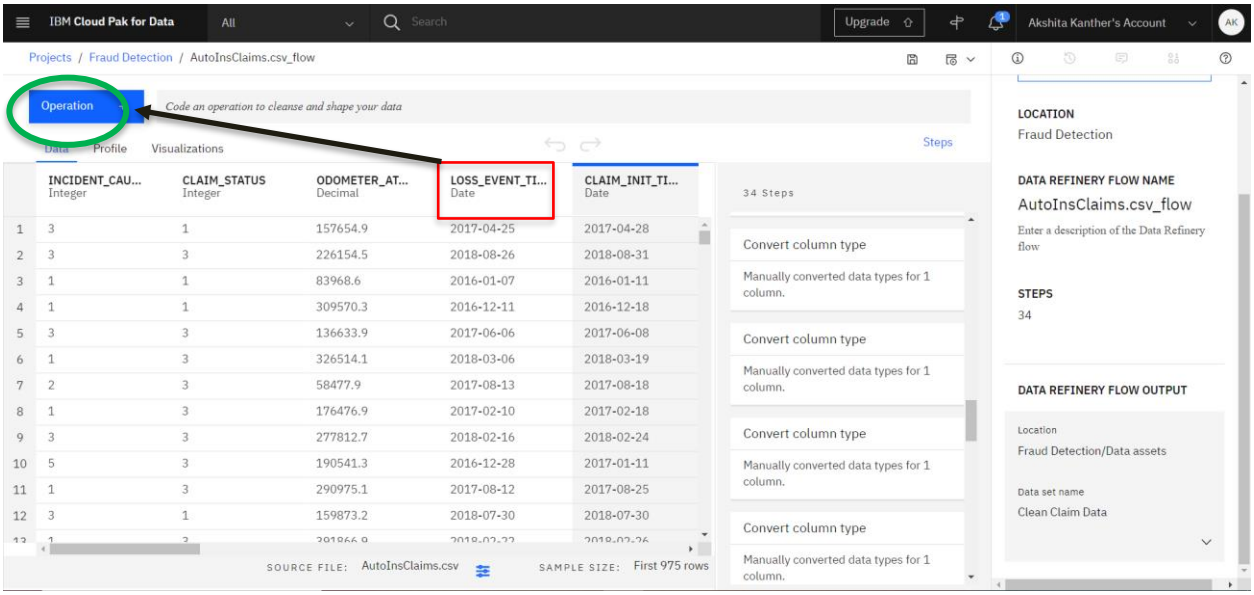


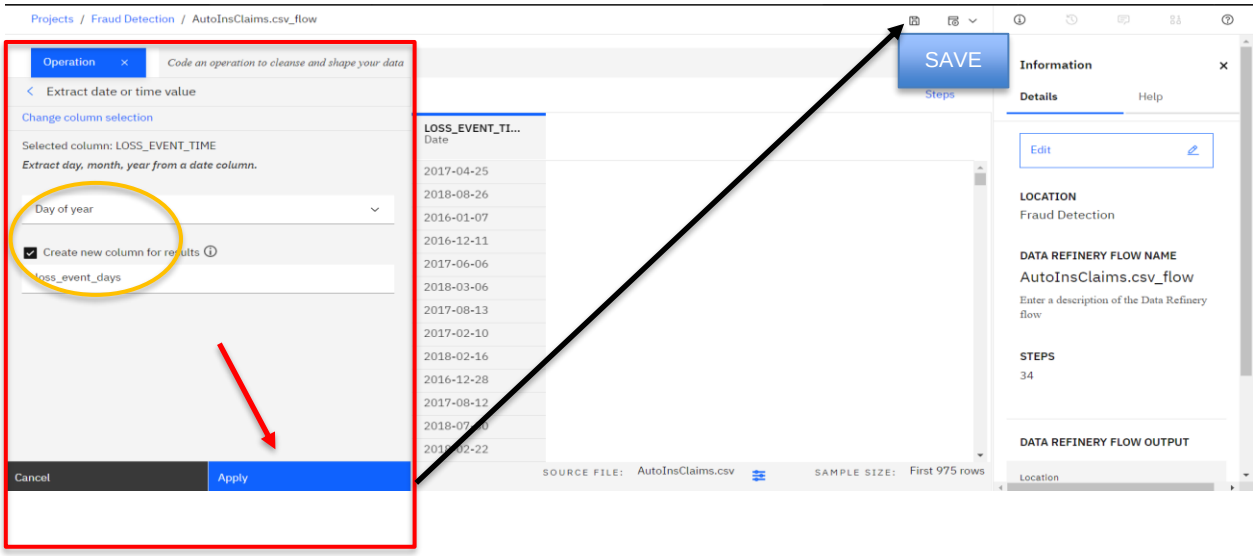



**7.** Go to Data Refinery Flow and select AutoInsClaims.csv_flow. Select Loss_event_time column, go on operations and select ext (extract date or time value), select Day Of The Year and create a new column with column name as (loss_event_days) and click on apply. Save the data.

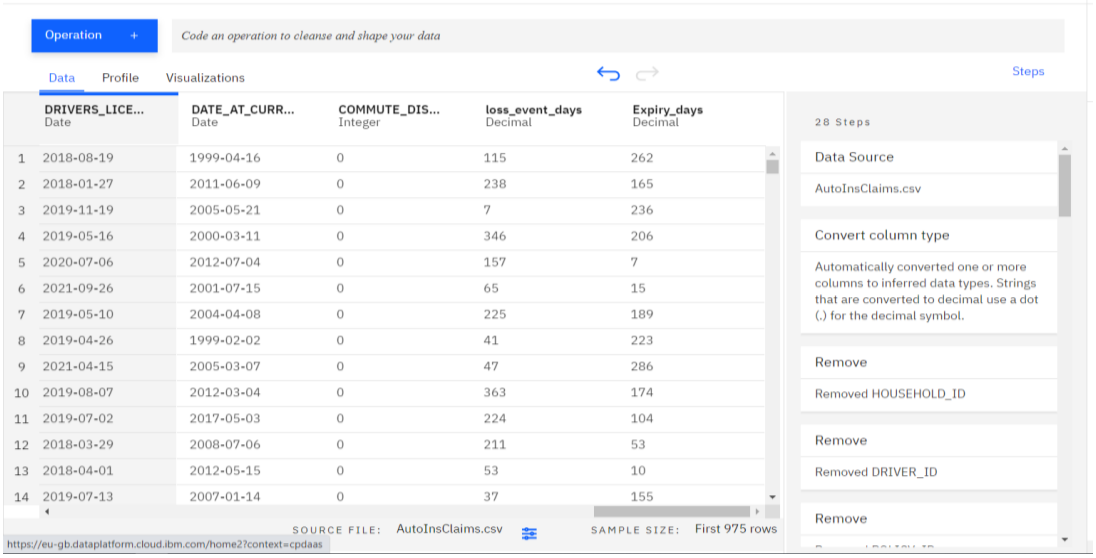



IBM Cloud Pak for Data
Upgrade
Akshita Kanther's Account
Projects / Fraud Detection / AutoInsClaims.csv_flow
Previewing
Details
Help
Edit
LOCATION
Fraud Detection
DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow
Enter a description of the Data Refinery flow
STEPS
34
DATA REFINERY FLOW OUTPUT
Data assets
Input
AutoInsClaims.csv
Output
Clean Claim Data

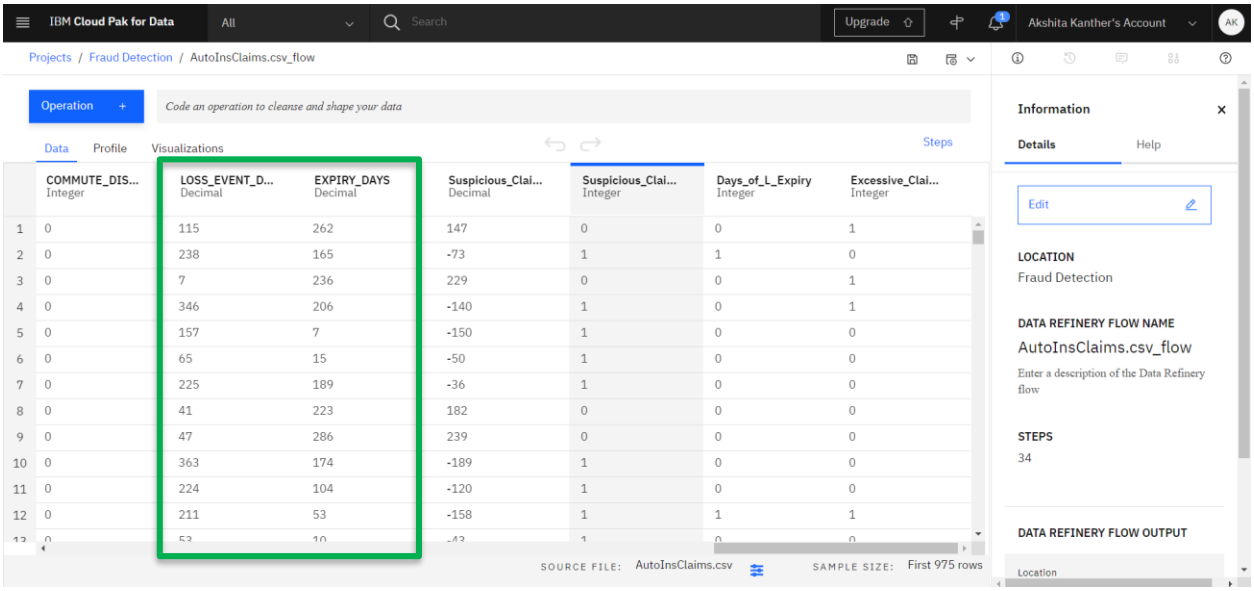

IBM Cloud Pak for Data
Upgrade
Akshita Kanther's Account
Projects / Fraud Detection / AutoInsClaims.csv_flow
Operation
Code an operation to cleanse and shape your data
Data
Profile
Visualizations
Steps
INCIDENT_CAU... Integer
CLAIM_STATUS Integer
ODOMETER_AT... Decimal
LOSS_EVENT_TI... Date
CLAIM_INIT_TI... Date
34 Steps
Convert column type
Manually converted data types for 1 column.
SOURCE FILE: AutoInsClaims.csv
SAMPLE SIZE: First 975 rows
LOCATION
Fraud Detection
DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow
Enter a description of the Data Refinery flow
STEPS
34
DATA REFINERY FLOW OUTPUT
Location
Fraud Detection/Data assets
Data set name
Clean Claim Data


Projects / Fraud Detection / AutoInsClaims.csv_flow
Operation
Code an operation to cleanse and shape your data
Extract date or time value
Change column selection
Selected column: LOSS_EVENT_TIME
Extract day, month, year from a date column.
Day of year
Create new column for results
loss_event_days
Cancel
Apply
SAVE
Steps
LOSS_EVENT_TI... Date
2017-04-25
2018-08-26
2016-01-07
2016-12-11
2017-06-06
2018-03-06
2017-08-13
2017-02-10
2018-02-16
2016-12-28
2017-08-12
SOURCE FILE: AutoInsClaims.csv
SAMPLE SIZE: First 975 rows
Information
Details
Help
Edit
LOCATION
Fraud Detection
DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow
Enter a description of the Data Refinery flow
STEPS
34
DATA REFINERY FLOW OUTPUT
Location

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




->After Saving

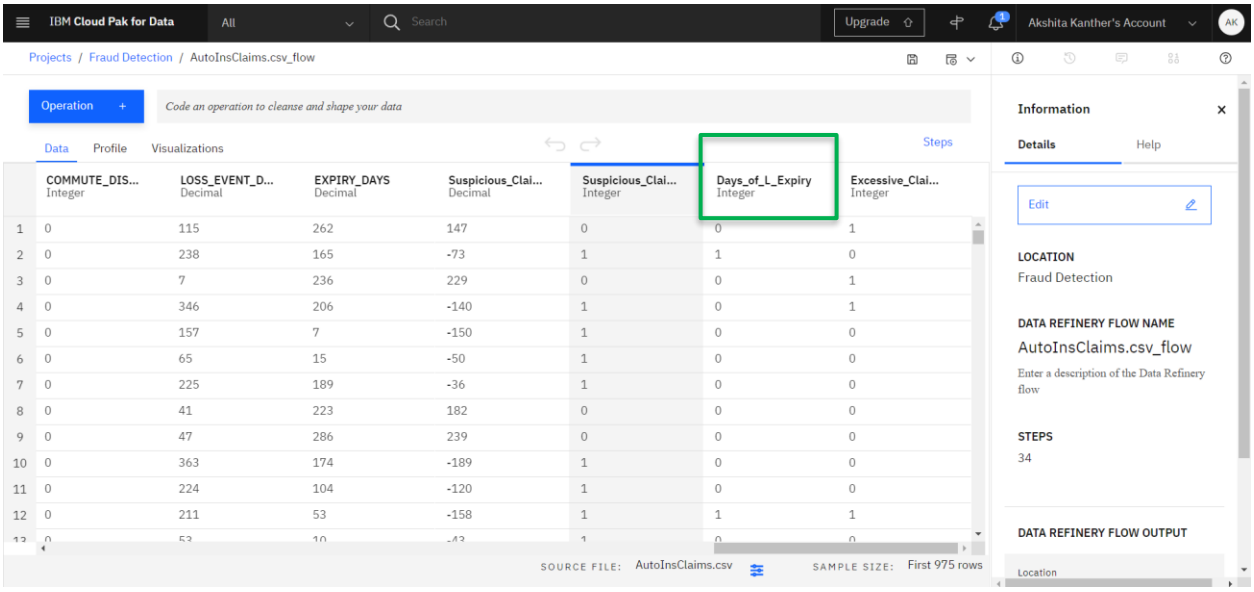


## 2. Claim filed after the license expiration date

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.

After converting boolean to integer type




**VI. 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 hypothesis 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. | | | |
| | | | |