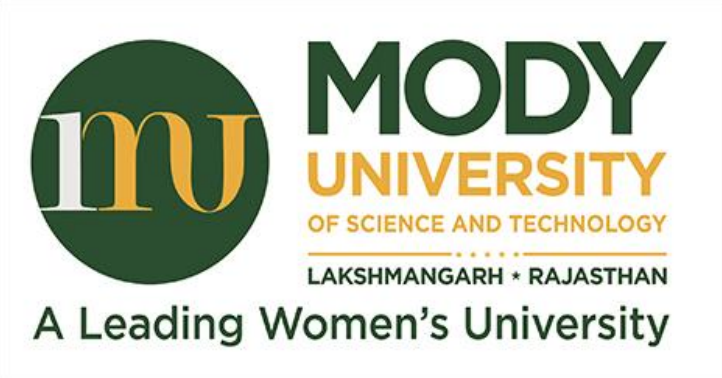


.Experiment Title.

FRAUD DETECTION IN THE INSURANCE BUSINESS

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

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

Semester: 6th

Date of Performance: 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

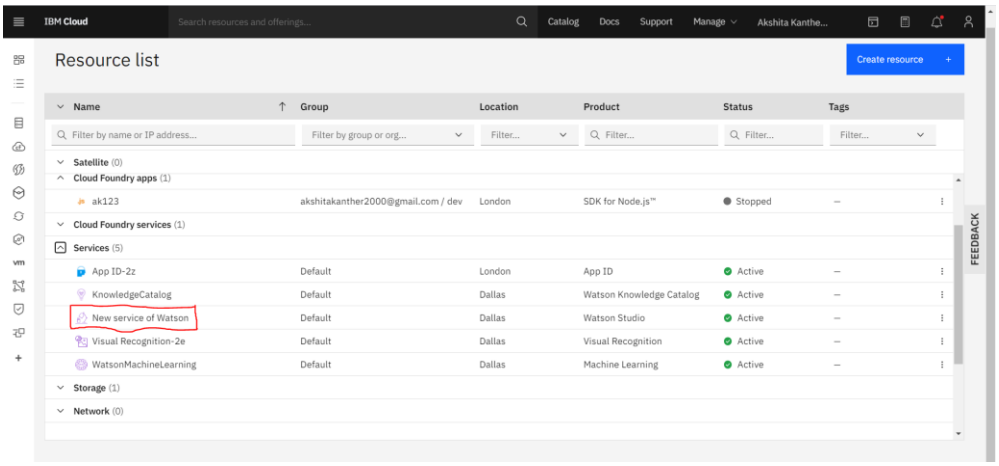
4. Hypothesis

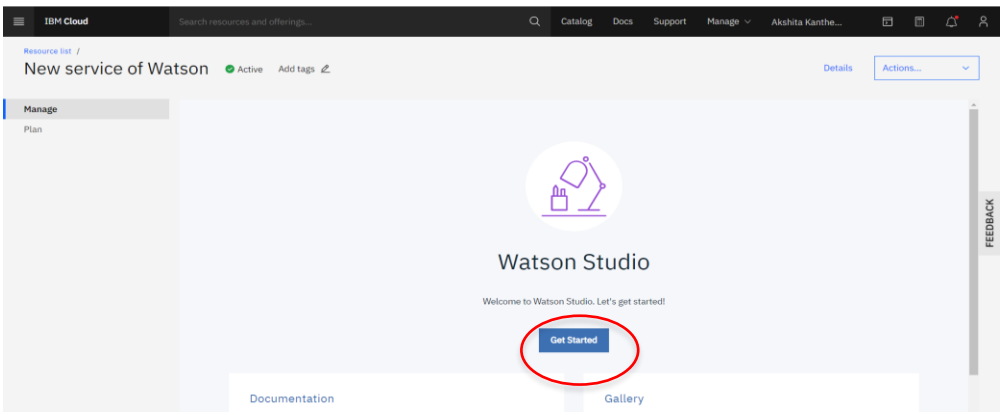
2. Claim filed after the license expiration date.
3. Excessive claim amount, which is over $10000 in value.

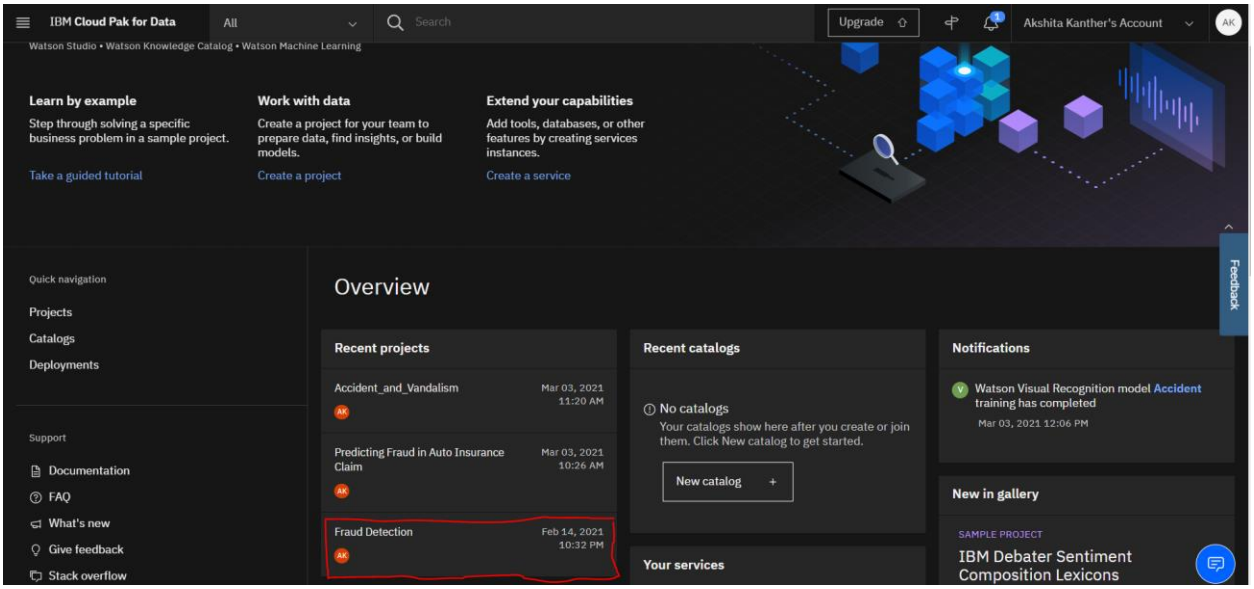
6. Steps for experiment/practical: The following steps are:

1. Login to your IBM Cloud account.

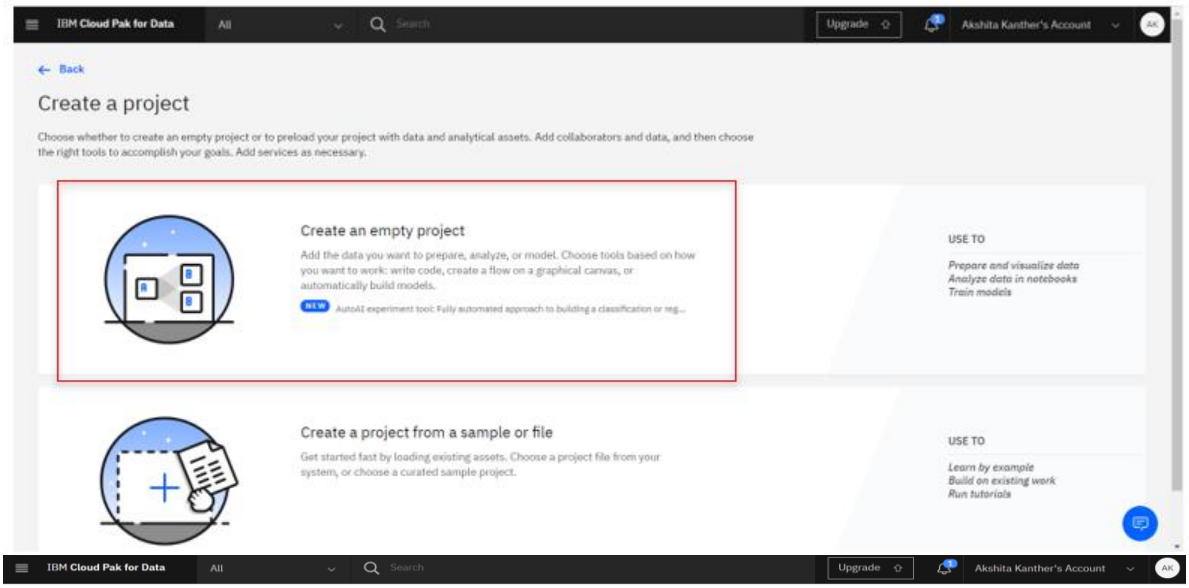
2. Go to your Watson Service and start a new project or start recent project which is already created.

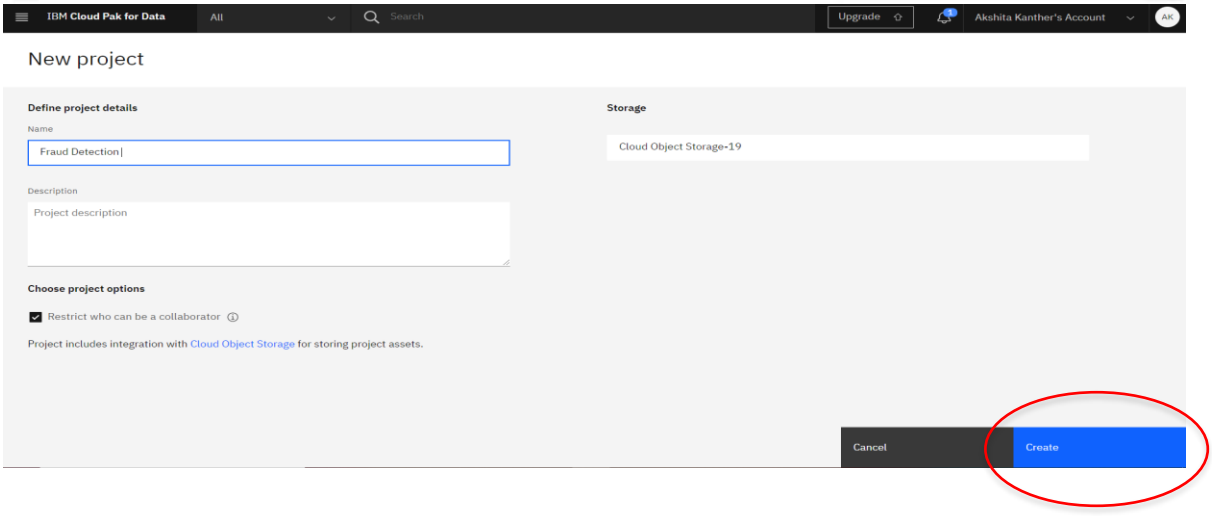


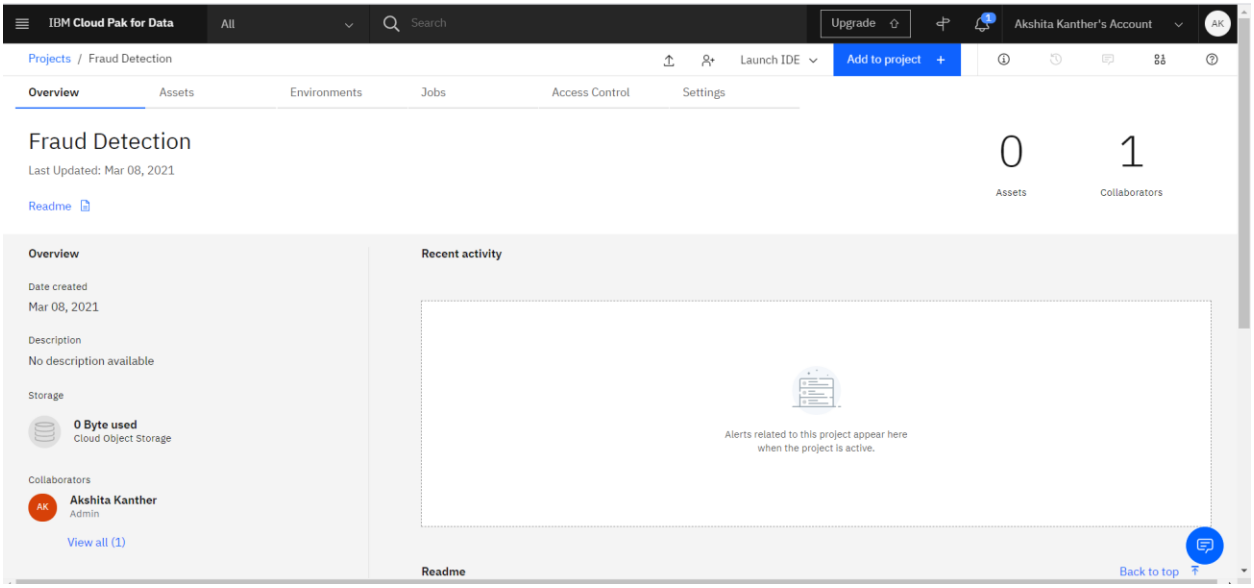


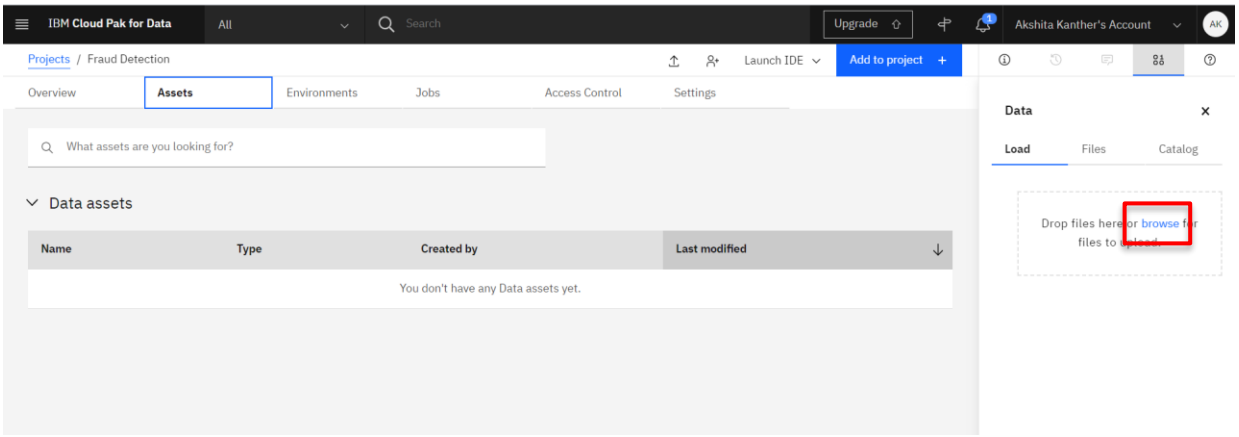


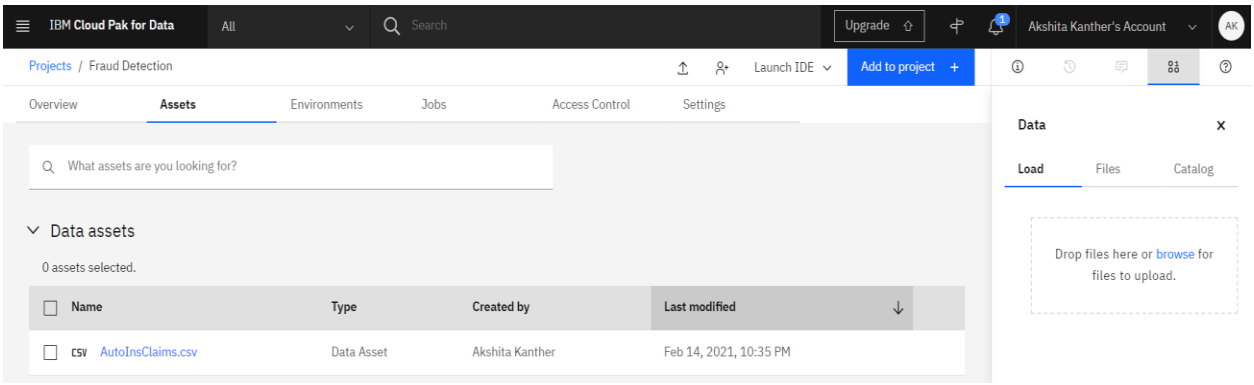
3. Load the AutoInsClaims dataset on the cloud.



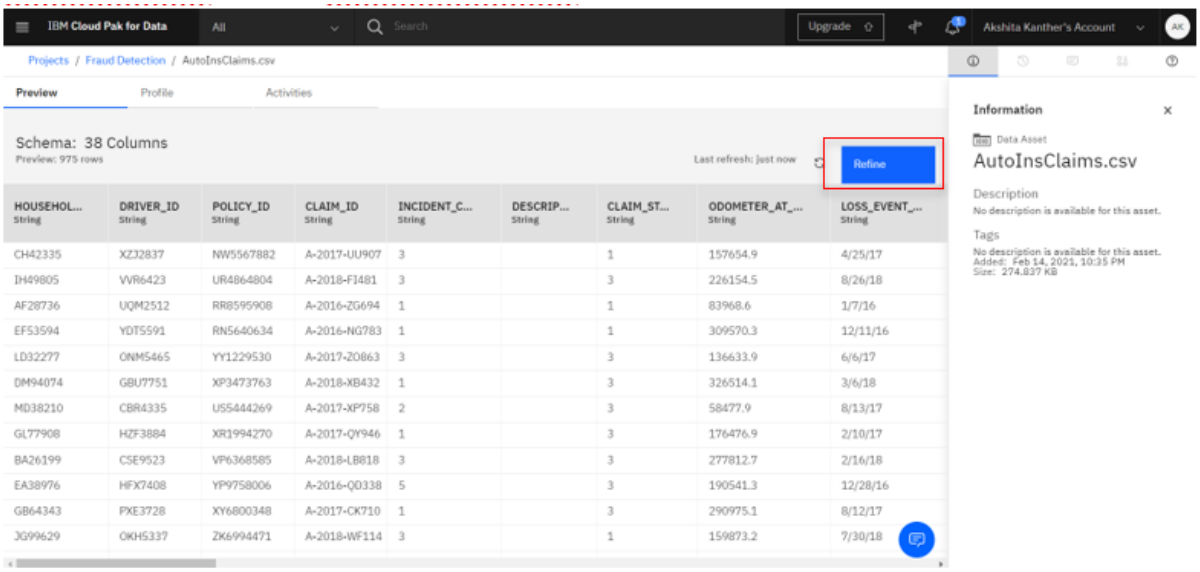


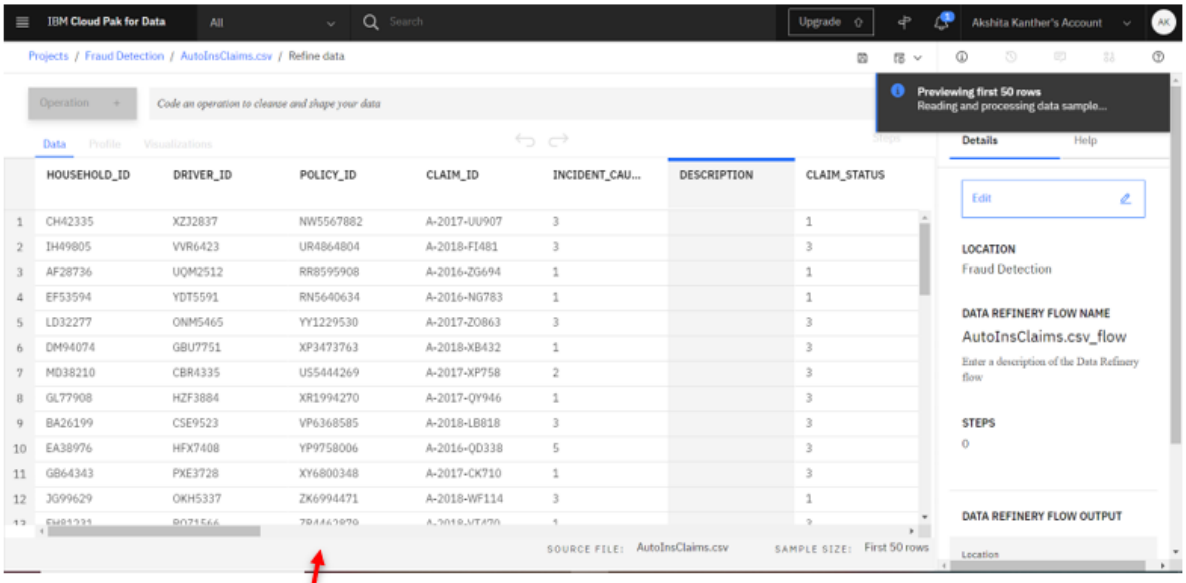


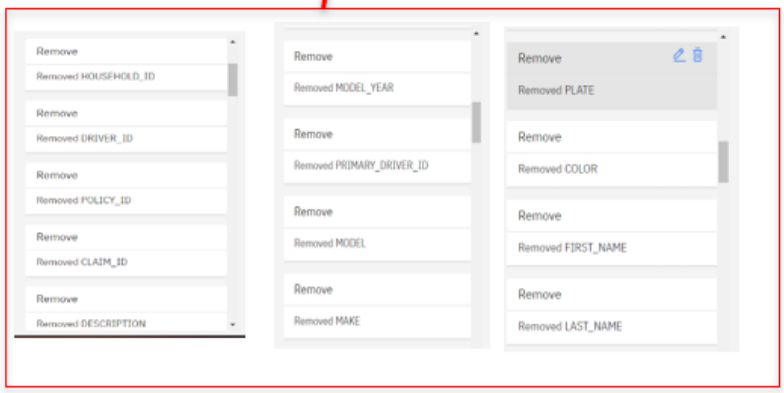




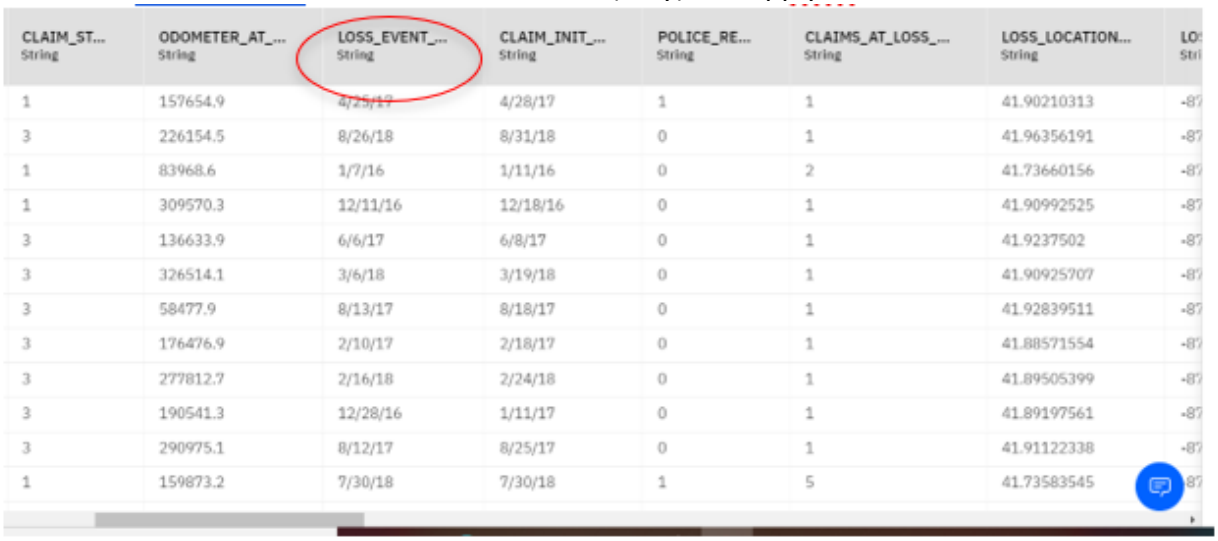
4. Clean the data, click on REFINE. Delete the following columns as they are not required for our hypothesis- household_id, driver_id, policy_id, claim_id, description, primary_driver_id, model_year, make, model, plate, color, first_name, ssn, last_name, driver_license_id, contact_number, e-mail, driver_license_state.


IBM Cloud Pak for Data
All
Search
Upgrade
Akshita Kanther's Account
Projects / Fraud Detection / AutoInsClaims.csv
Preview
Profile
Activities
Schema: 38 Columns
Preview: 975 rows
Last refresh: just now
Refine
HOUSEHOL... String
DRIVER_ID String
POLICY_ID String
CLAIM_ID String
INCIDENT_C... String
DESCRIP... String
CLAIM_ST... String
ODOMETER_AT_... String
LOSS_EVENT_... String
CH42335 XZJ2837 NW5567882 A-2017-UU907 3 1 157654.9 4/25/17
IH49805 VVR6423 UR4864804 A-2018-FI481 3 3 226154.5 8/26/18
AF28736 UQM2512 RR8595908 A-2016-ZG694 1 1 83968.6 1/7/16
EF53594 YDT5591 RN5640634 A-2016-NG783 1 1 309570.3 12/11/16
LD32277 ONM5465 YY1229530 A-2017-ZO863 3 3 136633.9 6/6/17
DM94074 GBU7751 XP3473763 A-2018-XB432 1 3 326514.1 3/6/18
MD38210 CBR4335 US5444269 A-2017-XP758 2 3 58477.9 8/13/17
GL77908 HZF3884 XR1994270 A-2017-QY946 1 3 176476.9 2/10/17
BA26199 CSE9523 VP6368585 A-2018-LB818 3 3 277812.7 2/16/18
EA38976 HFX7408 YP9758006 A-2016-QD338 5 3 190541.3 12/28/16
GB64343 PXE3728 XY6800348 A-2017-CK710 1 3 290975.1 8/12/17
JG99629 OKH5337 ZK6994471 A-2018-WF114 3 1 159873.2 7/30/18
Information
Data Asset
AutoInsClaims.csv
Description
No description is available for this asset.
Tags
No description is available for this asset.
Added: Feb 14, 2021, 10:35 PM
Size: 274.837 KB

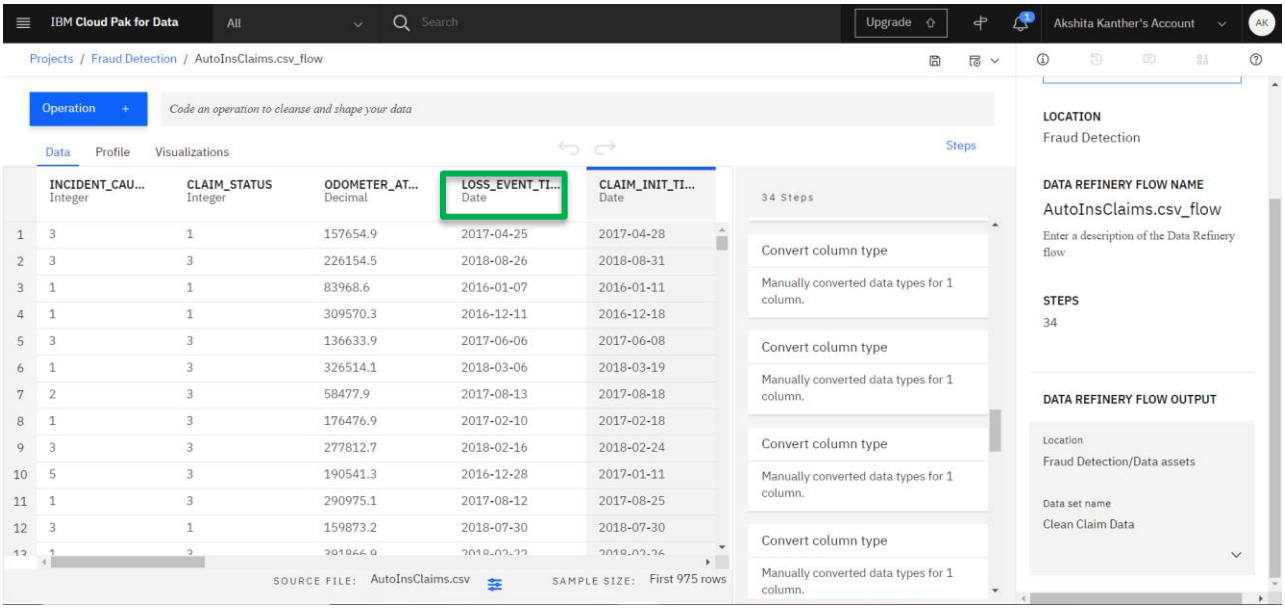

IBM Cloud Pak for Data
All
Search
Upgrade
Akshita Kanther's Account
Projects / Fraud Detection / AutoInsClaims.csv / Refine data
Operation
Code an operation to cleanse and shape your data
Previewing first 50 rows
Reading and processing data sample...
Data
Profile
Visualizations
Steps
Details
Help
HOUSEHOLD_ID DRIVER_ID POLICY_ID CLAIM_ID INCIDENT_CAU... DESCRIPTION CLAIM_STATUS
1 CH42335 XZJ2837 NW5567882 A-2017-UU907 3 1
2 IH49805 VVR6423 UR4864804 A-2018-FI481 3 3
3 AF28736 UQM2512 RR8595908 A-2016-ZG694 1 1
4 EF53594 YDT5591 RN5640634 A-2016-NG783 1 1
5 LD32277 ONM5465 YY1229530 A-2017-ZO863 3 3
6 DM94074 GBU7751 XP3473763 A-2018-XB432 1 3
7 MD38210 CBR4335 US5444269 A-2017-XP758 2 3
8 GL77908 HZF3884 XR1994270 A-2017-QY946 1 3
9 BA26199 CSE9523 VP6368585 A-2018-LB818 3 3
10 EA38976 HFX7408 YP9758006 A-2016-QD338 5 3
11 GB64343 PXE3728 XY6800348 A-2017-CK710 1 3
12 JG99629 OKH5337 ZK6994471 A-2018-WF114 3 1
SOURCE FILE: AutoInsClaims.csv
SAMPLE SIZE: First 50 rows
Edit
LOCATION
Fraud Detection
DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow
Enter a description of the Data Refinery flow
STEPS
0
DATA REFINERY FLOW OUTPUT
Location


Remove
Removed HOUSEHOLD_ID
Remove
Removed DRIVER_ID
Remove
Removed POLICY_ID
Remove
Removed CLAIM_ID
Remove
Removed DESCRIPTION
Remove
Removed MODEL_YEAR
Remove
Removed PRIMARY_DRIVER_ID
Remove
Removed MODEL
Remove
Removed MAKE
Remove
Removed PLATE
Remove
Removed COLOR
Remove
Removed FIRST_NAME
Remove
Removed LAST_NAME

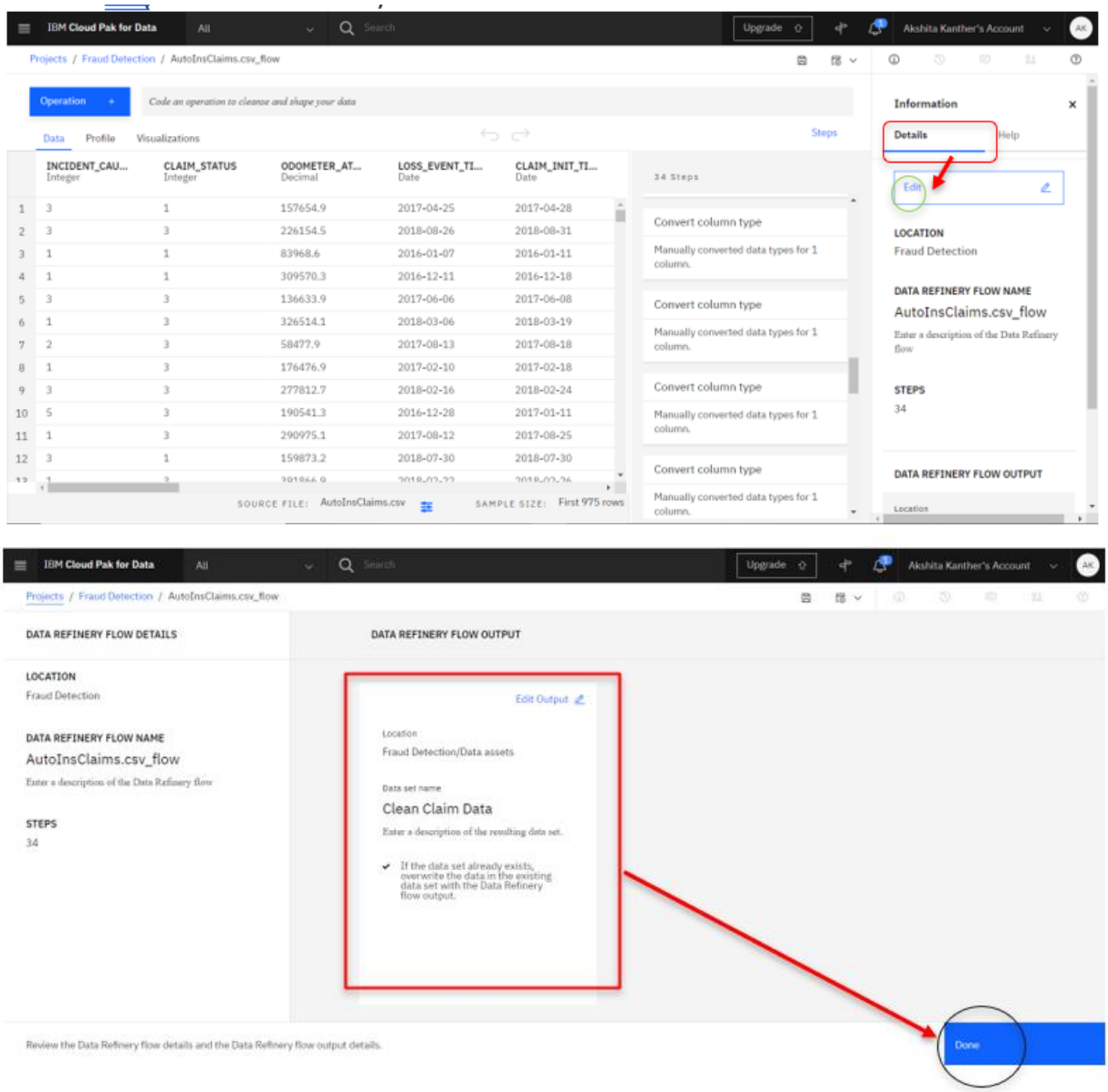
5. Convert all the date columns into MDY format. For the same, select the column, click on convert and select date, select current order as (mdy) and apply.



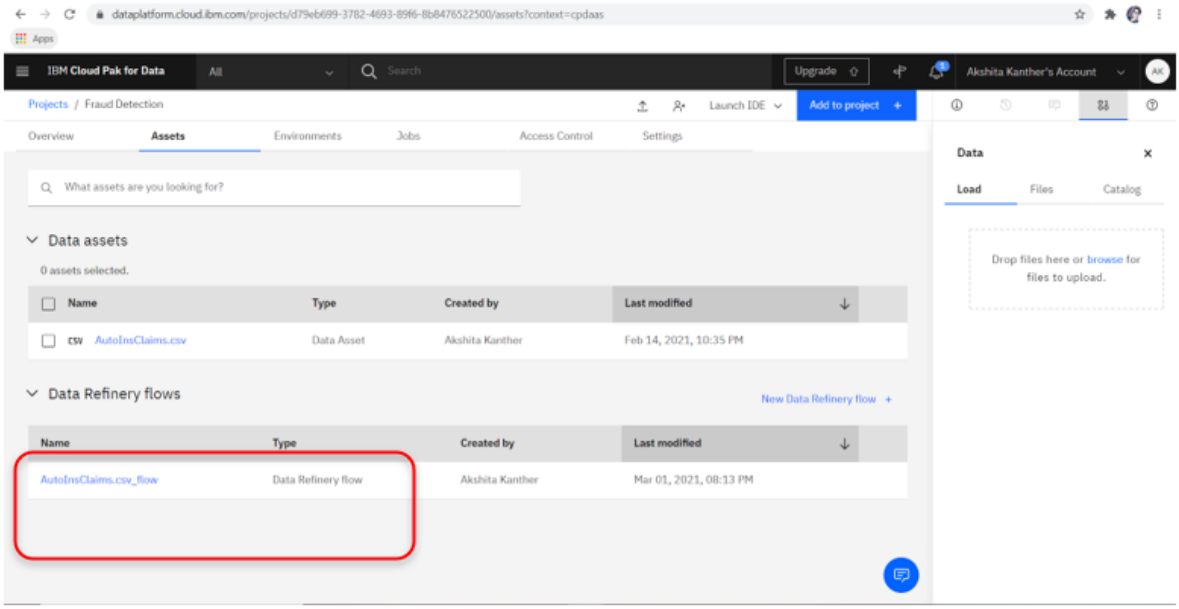
CLAIM_ST... String	ODOMETER_AT_... String	LOSS_EVENT_... String	CLAIM_INIT_... String	POLICE_RE... String	CLAIMS_AT_LOSS_... String	LOSS_LOCATION... String	LO... Stri
1	157654.9	4/25/17	4/28/17	1	1	41.90210313	-87
3	226154.5	8/26/18	8/31/18	0	1	41.96356191	-87
1	83968.6	1/7/16	1/11/16	0	2	41.73660156	-87
1	309570.3	12/11/16	12/18/16	0	1	41.90992525	-87
3	136633.9	6/6/17	6/8/17	0	1	41.9237502	-87
3	326514.1	3/6/18	3/19/18	0	1	41.90925707	-87
3	58477.9	8/13/17	8/18/17	0	1	41.92839511	-87
3	176476.9	2/10/17	2/18/17	0	1	41.88571554	-87
3	277812.7	2/16/18	2/24/18	0	1	41.89505399	-87
3	190541.3	12/28/16	1/11/17	0	1	41.89197561	-87
3	290975.1	8/12/17	8/25/17	0	1	41.91122338	-87
1	159873.2	7/30/18	7/30/18	1	5	41.73583545	-87

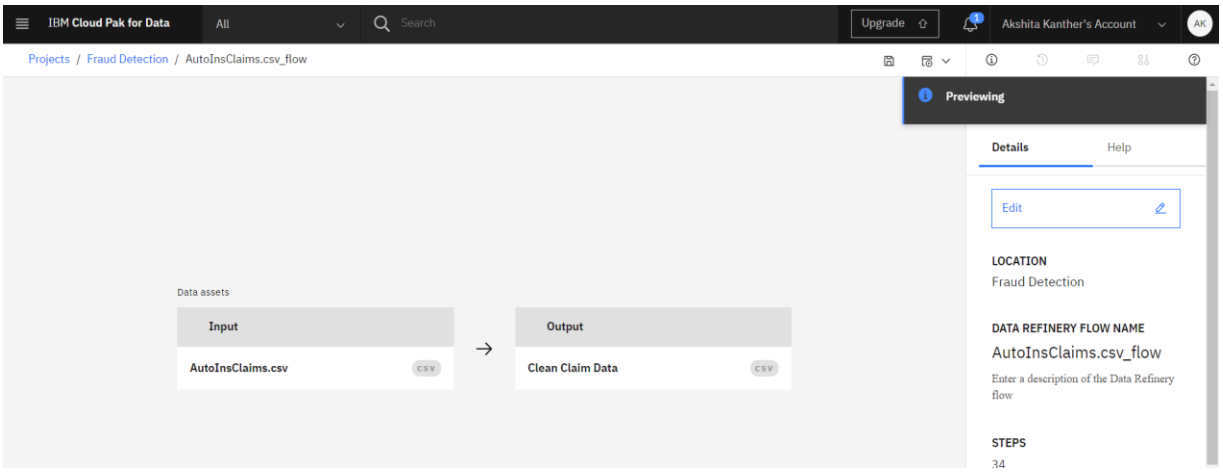


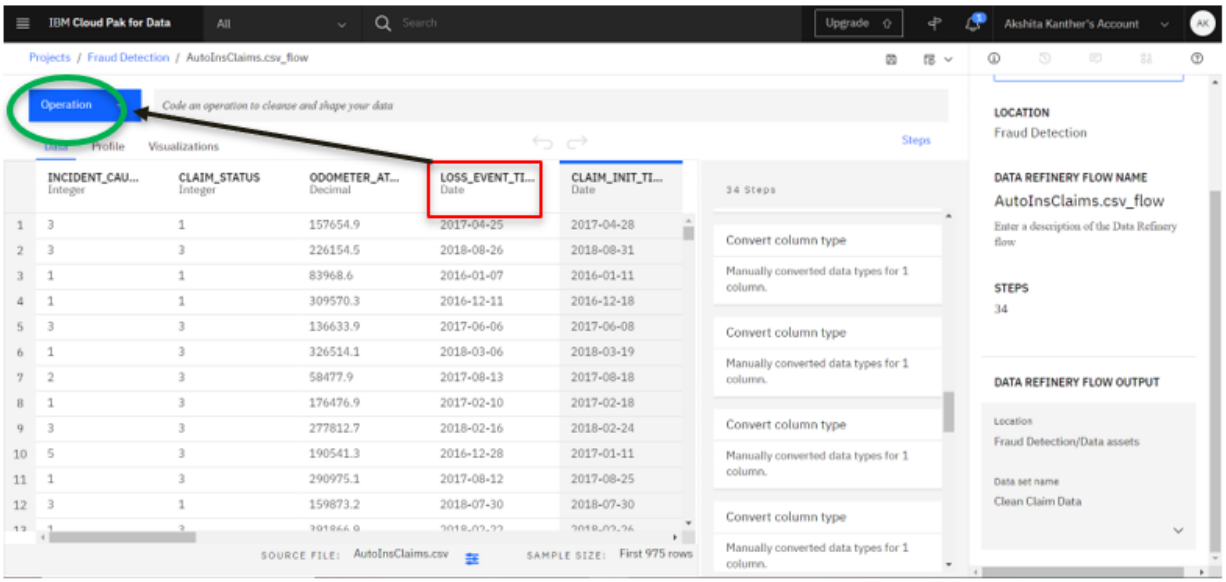
6. Save the refined data, click on Details then click on edit. Click on edit output and name the data set as (cleansed claim date.csv). click on done.

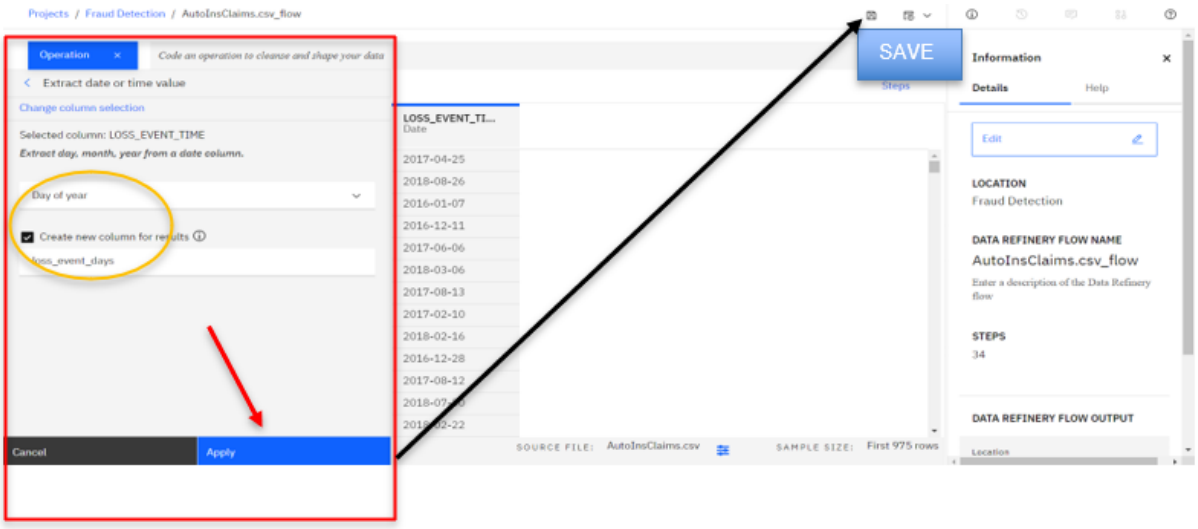


7. Go to Data Refinery Flow and select AutoInsClaims.csv_flow. Select Loss_event_time column, go on operations and select ext (extract date or time value), select Day Of The Year and create a new column with column name as (loss_event_days) and click on apply. Save the data.

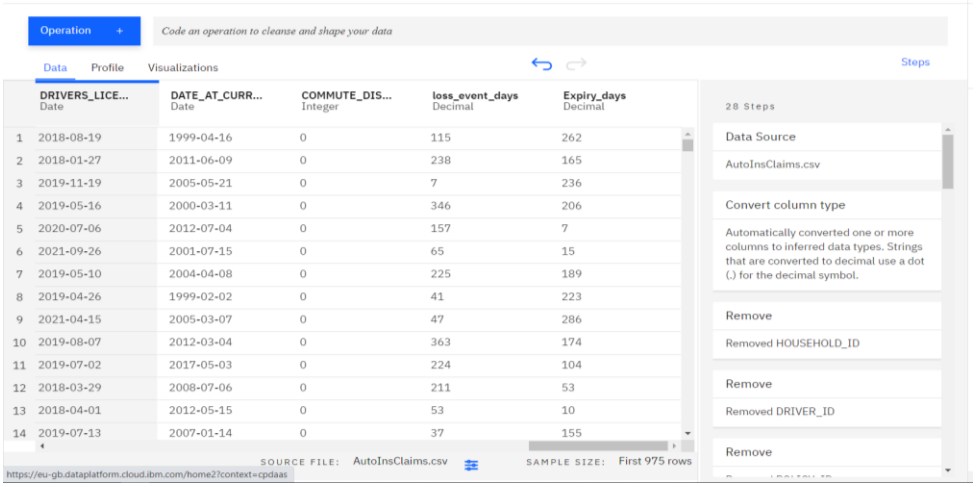

IBM Cloud Pak for Data
Projects / Fraud Detection
Overview
Assets
Environments
Jobs
Access Control
Settings
Data assets
Name
Type
Created by
Last modified
AutoInsClaims.csv
Data Asset
Akshita Kanther
Feb 14, 2021, 10:35 PM
Data Refinery flows
New Data Refinery flow
AutoInsClaims.csv_flow
Data Refinery flow
Mar 01, 2021, 08:13 PM
Data
Load
Files
Catalog
Drop files here or browse for files to upload.


IBM Cloud Pak for Data
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 assets
Input
AutoInsClaims.csv
Output
Clean Claim Data

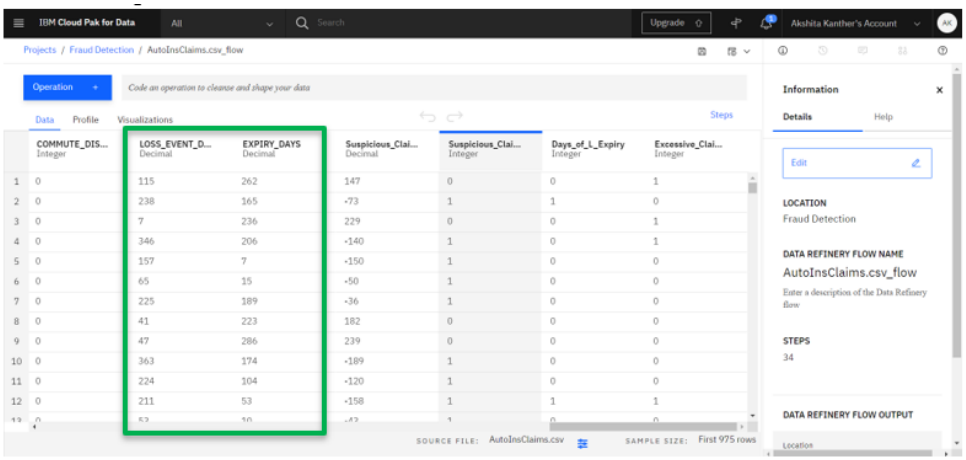

IBM Cloud Pak for Data
Projects / Fraud Detection / AutoInsClaims.csv_flow
Operation
Code an operation to cleanse and shape your data
Data
Profile
Visualizations
Steps
INCIDENT_CAU...
CLAIM_STATUS
ODOMETER_AT...
LOSS_EVENT_TI...
CLAIM_INIT_TI...
34 Steps
Convert column type
Manually converted data types for 1 column.
LOCATION
Fraud Detection
DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow
STEPS
34
DATA REFINERY FLOW OUTPUT
Location
Fraud Detection/Data assets
Data set name
Clean Claim Data
SOURCE FILE: AutoInsClaims.csv
SAMPLE SIZE: First 975 rows



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



->After Saving

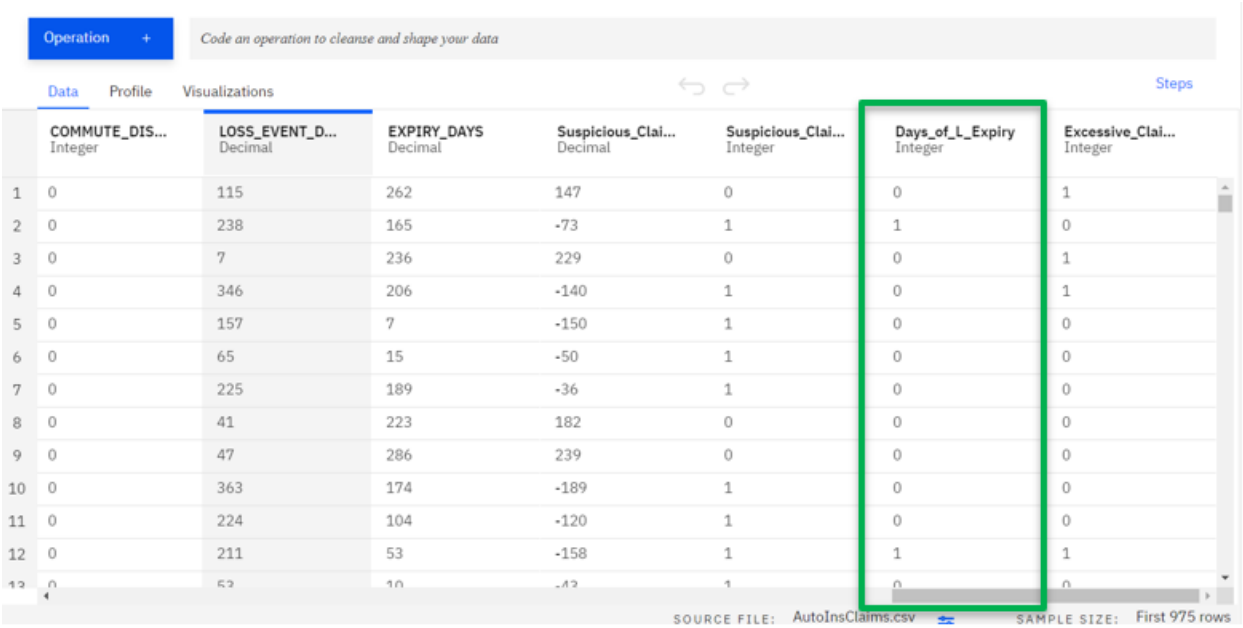


2. Claim filed after the license expiration date

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

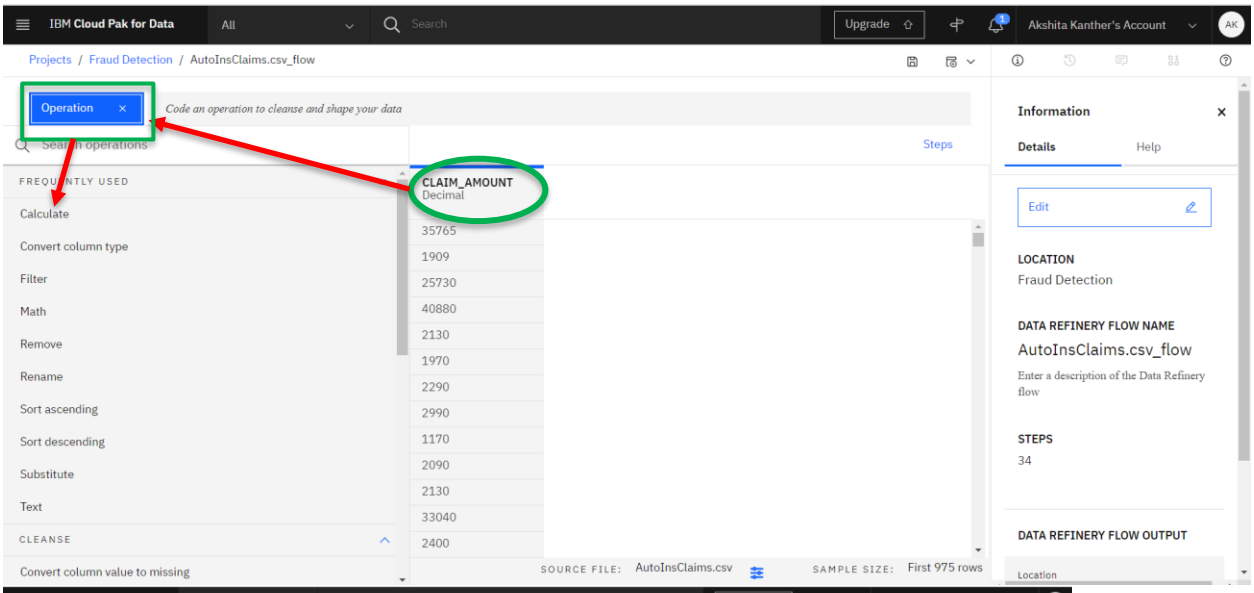


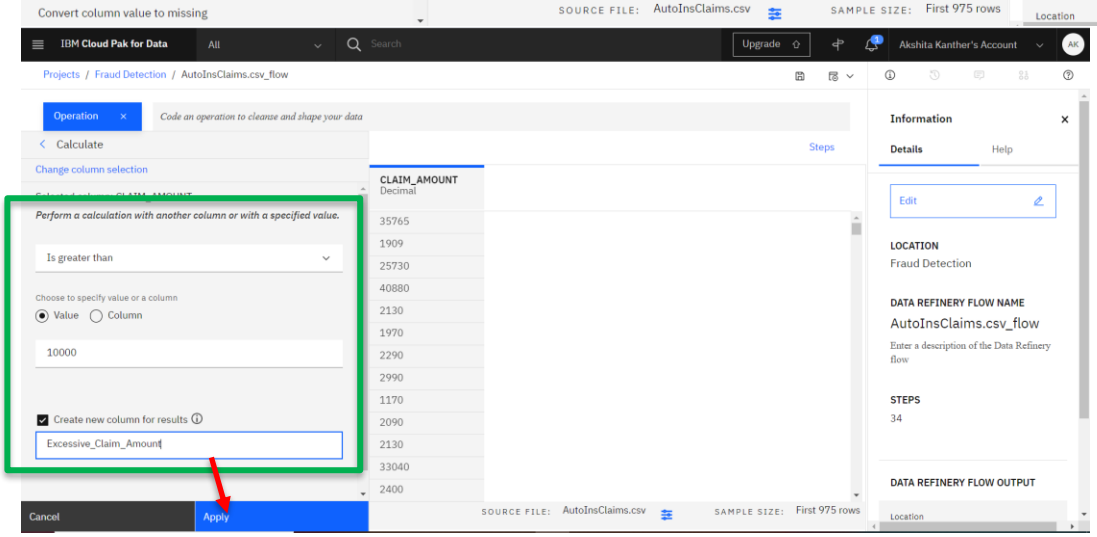
	COMMUTE_DIS... Integer	LOSS_EVENT_D... Decimal	EXPIRY_DAYS Decimal	Suspicious_Clai... Decimal	Suspicious_Clai... Integer	Days_of_L_Expiry Integer	Excessive_Clai... Integer
1	0	115	262	147	0	0	1
2	0	238	165	-73	1	1	0
3	0	7	236	229	0	0	1
4	0	346	206	-140	1	0	1
5	0	157	7	-150	1	0	0
6	0	65	15	-50	1	0	0
7	0	225	189	-36	1	0	0
8	0	41	223	182	0	0	0
9	0	47	286	239	0	0	0
10	0	363	174	-189	1	0	0
11	0	224	104	-120	1	0	0
12	0	211	53	-158	1	1	1

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

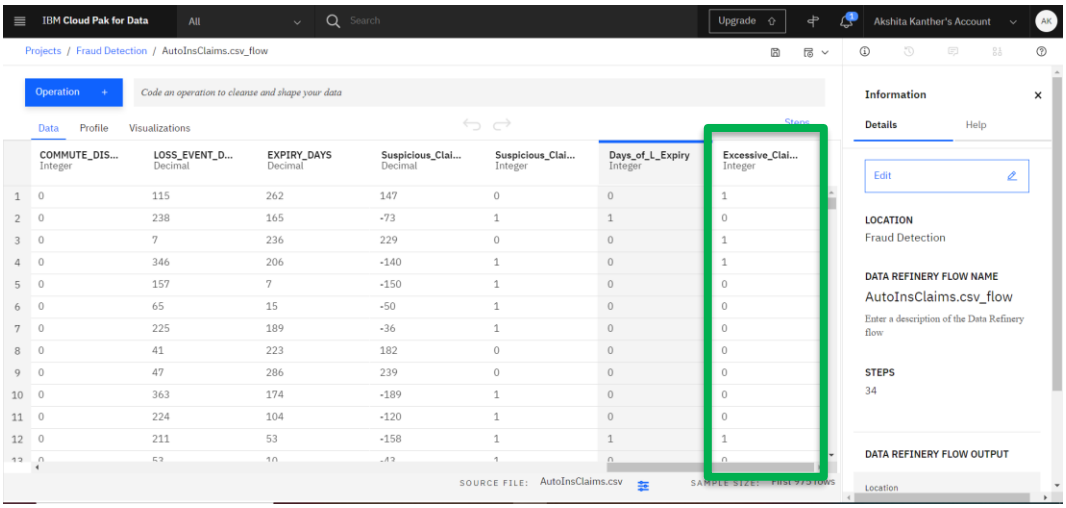
This hypothesis says that the car accidents who claim for over $10000 will not be given the entire claim amount. But the claims under $10000 will be fully claimed. Steps for the hypothesis-

- 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.





->After converting from boolean to integer

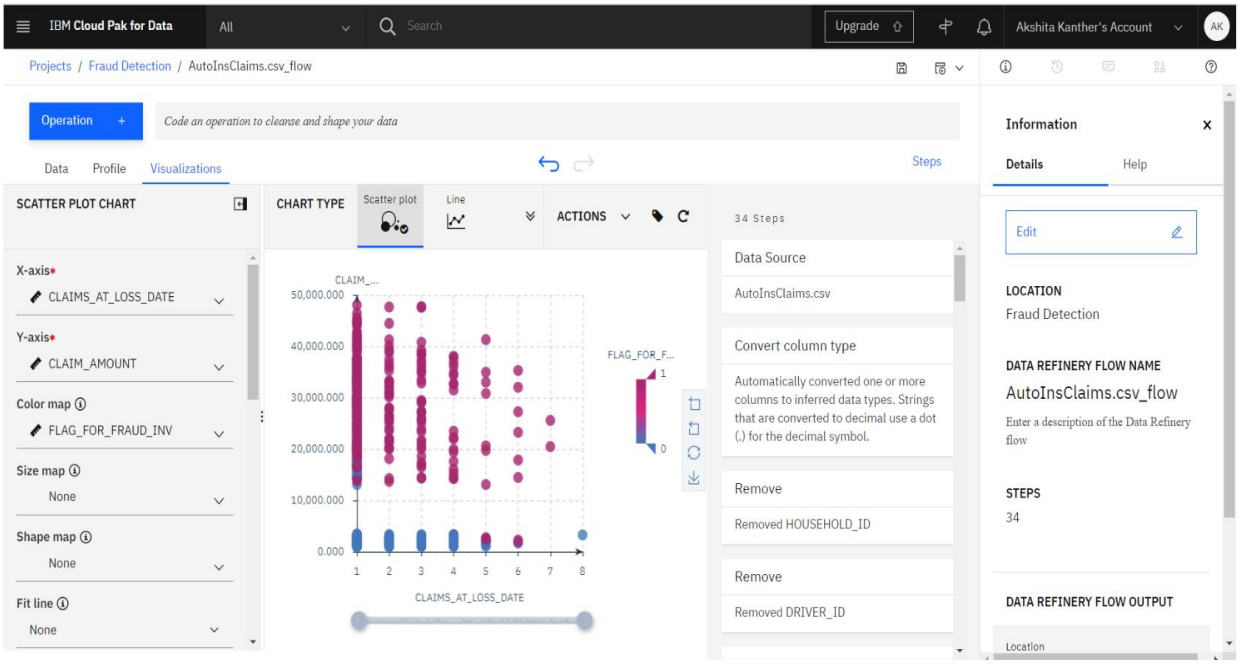


4. Data Refinery Visualization-

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

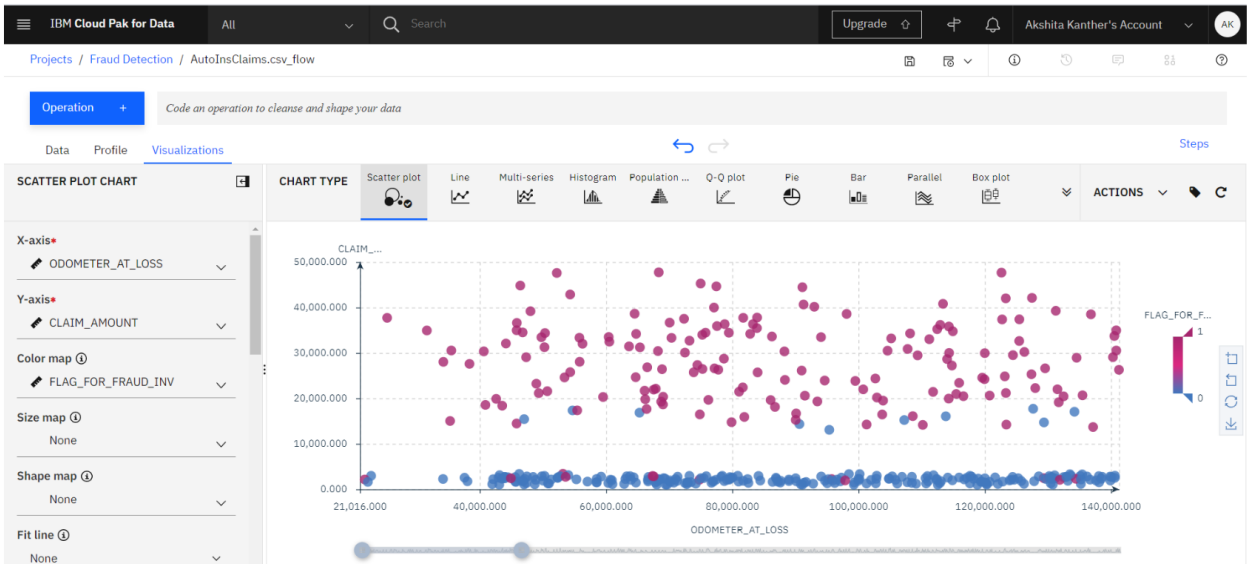
1. **CLAIM_AT_LOSS_DATE**

   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.



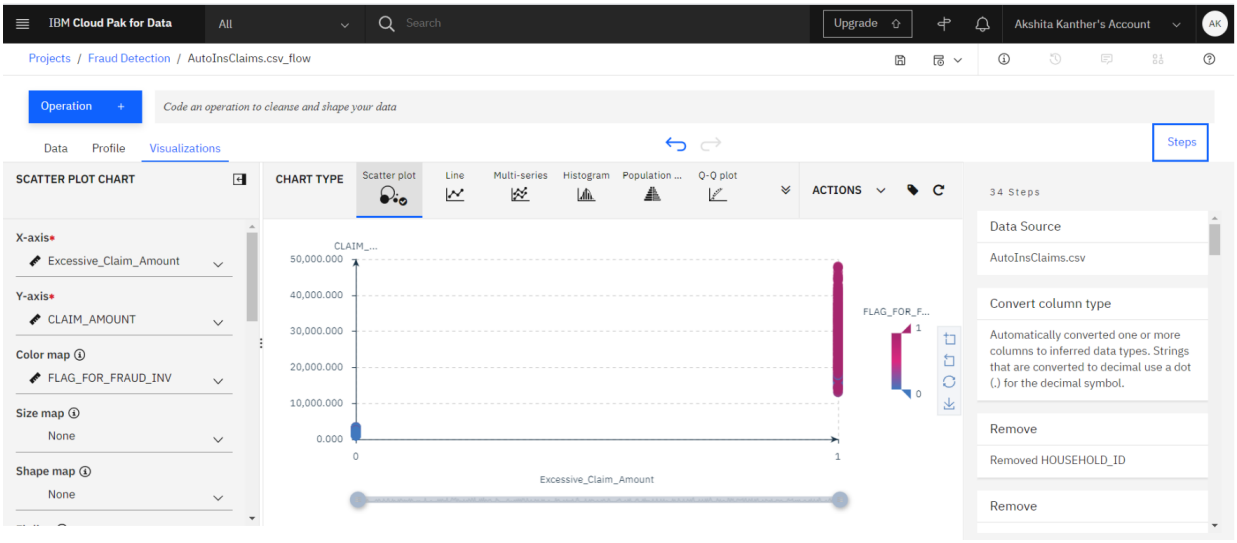
2. **ODOMETER_AT_LOSS**

   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.			