



.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:

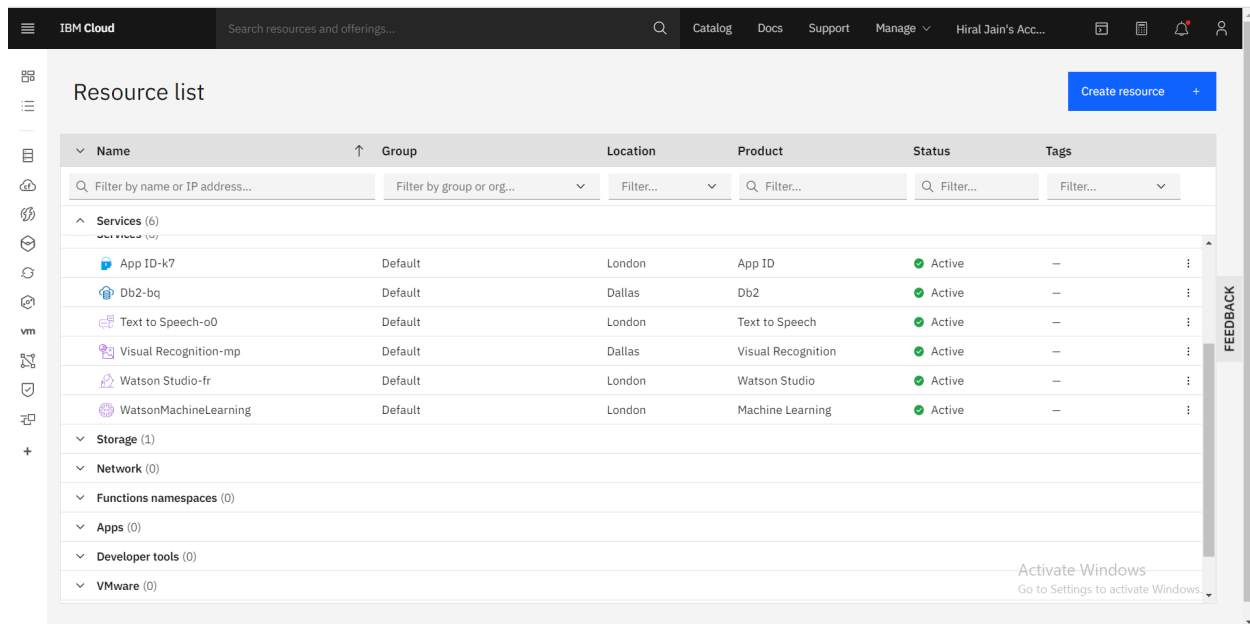
- Internet connectivity
- IBM cloud account
- IBM WATSON service
- NIC Data Set

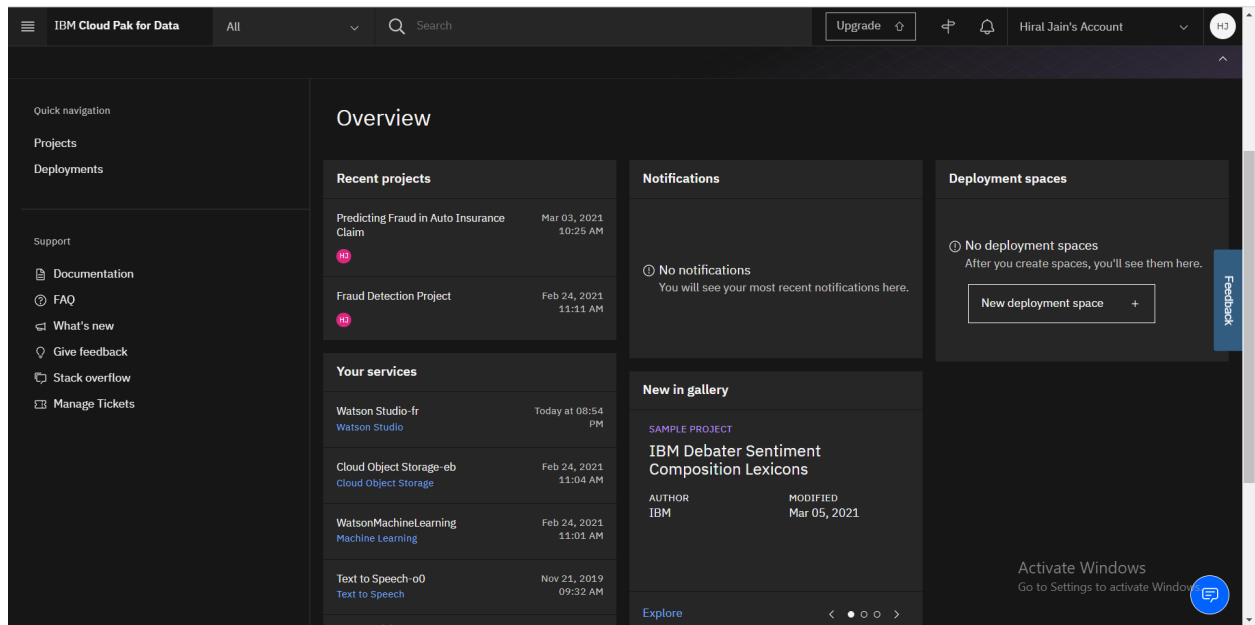
4. Hypothesis

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

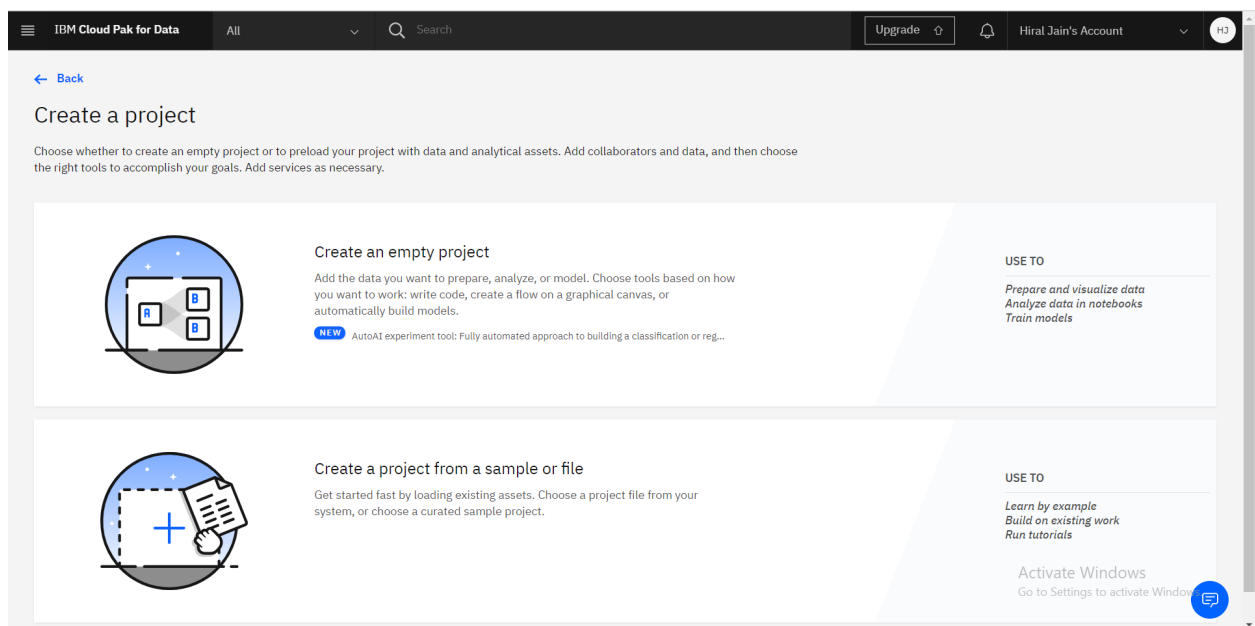
6. Steps for experiment/practical: The following steps are:

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.



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

Projects / Prediction of Fraud Detection in ...

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

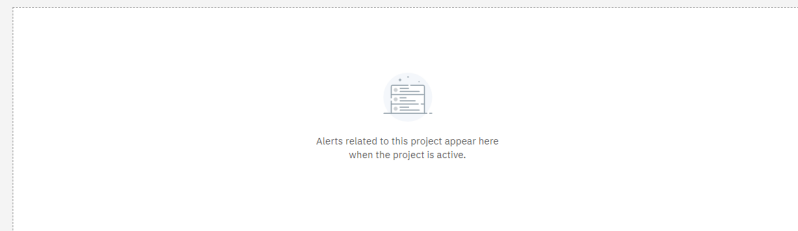
0 Byte used
Cloud Object Storage

Collaborators

Hiral Jain
Admin

[View all \(1\)](#)

Recent activity



[Readme](#)

Activate Windows

Go to Settings to activate Windows.

[Back to top](#)

The top screenshot shows the IBM Cloud Pak for Data interface with the 'Assets' tab selected. The 'Data assets' section is empty, displaying the message 'You don't have any Data assets yet.' The right sidebar shows the 'Data' panel with the 'Load' tab active, indicating a file upload area.

The bottom screenshot shows the same interface, but the 'Data assets' table now contains one asset:

<input type="checkbox"/>	Name	Type	Created by	Last modified
<input type="checkbox"/>	CSV AutoInsClaims.csv	Data Asset	Hiral Jain	Mar 06, 2021, 09:03 PM

The right sidebar remains the same, showing the 'Data' panel with the 'Load' tab active.

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

Hiral Jain's Account

Projects / Prediction of Fraud Detection in ... / AutoInsClaims.csv / Refine data

Operation +

Code an operation to cleanse and shape your data

Data

Profile

Visualizations

19 Steps

Remove

Removed SSN

Remove

Removed DRIVERS_LICENSE_ID

Remove

Removed CONTACT_NUMBER

Remove

Removed EMAIL

Remove JUST ADDED

Removed DRIVERS_LICENSE_STATE

	INCIDENT_CAU...	CLAIM_STATUS	ODOMETER_AT...	LOSS_EVENT_TI...	CLAIM_INIT_TI...
	Integer	Integer	Decimal	String	String
1	3	1	157654.9	4/25/17	4/28/17
2	3	3	226154.5	8/26/18	8/31/18
3	1	1	83968.6	1/7/16	1/11/16
4	1	1	309570.3	12/11/16	12/18/16
5	3	3	136633.9	6/6/17	6/8/17
6	1	3	326514.1	3/6/18	3/19/18
7	2	3	58477.9	8/13/17	8/18/17
8	1	3	176476.9	2/10/17	2/18/17
9	3	3	277812.7	2/16/18	2/24/18
10	5	3	190541.3	12/28/16	1/11/17
11	1	3	290975.1	8/12/17	8/25/17
12	3	1	159873.2	7/30/18	7/30/18
13	1	3	391866.9	2/22/18	2/26/18
14	3	1	384000.6	2/6/17	2/6/17

SOURCE FILE: AutoInsClaims.csv

SAMPLE SIZE: First 975 rows

Information

Details

Help

Edit

LOCATION

Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME

AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS

19

DATA REFINERY FLOW OUTPUT

Activate Windows

Location settings to activate Windows.

Prediction of Fraud Detection in Au...

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.

IBM Cloud Pak for Data

All

Search

Upgrade

Hiral Jain's Account

Projects / Prediction of Fraud Detection in ... / AutoInsClaims.csv / Refine data

Operation +

Code an operation to cleanse and shape your data

Data

Profile

Visualizations

27 Steps

Remove

Removed SSN

Remove

Removed DRIVERS_LICENSE_ID

Remove

Removed CONTACT_NUMBER

Convert column type

Manually converted data types for 1 column.

Convert column type

Manually converted data types for 1

	INCIDENT_CAU...	CLAIM_STATUS	ODOMETER_AT...	LOSS_EVENT_TI...	CLAIM_INIT_TI...
	Integer	Integer	Decimal	Date	Date
1	3	1	157654.9	2017-04-25	2017-04-28
2	3	3	226154.5	2018-08-26	2018-08-31
3	1	1	83968.6	2016-01-07	2016-01-11
4	1	1	309570.3	2016-12-11	2016-12-18
5	3	3	136633.9	2017-06-06	2017-06-08
6	1	3	326514.1	2018-03-06	2018-03-19
7	2	3	58477.9	2017-08-13	2017-08-18
8	1	3	176476.9	2017-02-10	2017-02-18
9	3	3	277812.7	2018-02-16	2018-02-24
10	5	3	190541.3	2016-12-28	2017-01-11
11	1	3	290975.1	2017-08-12	2017-08-25
12	3	1	159873.2	2018-07-30	2018-07-30
13	1	3	391866.9	2018-02-22	2018-02-26
14	3	1	384000.6	2017-02-06	2017-02-06

SOURCE FILE: AutoInsClaims.csv

SAMPLE SIZE: First 975 rows

Information

Details

Help

Edit

LOCATION

Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME

AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS

27

DATA REFINERY FLOW OUTPUT

Activate Windows

Location settings to activate Windows.

Prediction of Fraud Detection in Au...

Operation + Code an operation to cleanse and shape your data

Data Profile Visualizations

	INCIDENT_CAUSE Integer	CLAIM_STATUS Integer	ODOMETER_AT_MILE Decimal	LOSS_EVENT_TIMESTAMP Date	CLAIM_INIT_TIMESTAMP Date
1	3	1	157654.9	2017-04-25	2017-04-28
2	3	3	226154.5	2018-08-26	2018-08-31
3	1	1	83968.6	2016-01-07	2016-01-11
4	1	1	309570.3	2016-12-11	2016-12-18
5	3	3	136633.9	2017-06-06	2017-06-08
6	1	3	326514.1	2018-03-06	2018-03-19
7	2	3	58477.9	2017-08-13	2017-08-18
8	1	3	176476.9	2017-02-10	2017-02-18
9	3	3	277812.7	2018-02-16	2018-02-24
10	5	3	190541.3	2016-12-28	2017-01-11
11	1	3	290975.1	2017-08-12	2017-08-25
12	3	1	159873.2	2018-07-30	2018-07-30
13	1	3	391866.9	2018-02-22	2018-02-26
14	3	1	384000.6	2017-02-06	2017-02-06

SOURCE FILE: AutoInsClaims.csv SAMPLE SIZE: First 975 rows

27 Steps

- Remove
- Removed SSN
- Remove
- Removed DRIVERS_LICENSE_ID
- Remove
- Removed CONTACT_NUMBER
- Convert column type
- Manually converted data types for 1 column.
- Convert column type
- Manually converted data types for 1

Information

Details Help

Edit

LOCATION

Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME

AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS

27

DATA REFINERY FLOW OUTPUT

Activate Windows

Go to Settings to activate Windows.

Prediction of Fraud Detection in Au...

DATA REFINERY FLOW DETAILS

LOCATION

Prediction of Fraud Detection in Auto Insurance CL...

DATA REFINERY FLOW NAME

AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS

27

DATA REFINERY FLOW OUTPUT

Edit output

Location *

Prediction of Fraud Detection in A...

Data set name *

cleansed claim date

Description

Enter a description of the resulting data set.

File format

CSV

Activate Windows

Go to Settings to activate Windows.

Done

Review the Data Refinery flow details and the Data Refinery flow output details.

https://eu-gb.dataplatform.cloud.ibm.com/shaper?project_id=e9d5173c-37ad-4928-8b31-38...

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.

IBM Cloud Pak for Data | All | Search | Upgrade | Hiral Jain's Account

Projects / Prediction of Fraud Detection in ... / AutoInsClaims.csv / Refine data

Operation + Code an operation to cleanse and shape your data

Data Profile Visualizations

	INCIDENT_CAU...	CLAIM_STATUS	ODOMETER_AT...	LOSS_EVENT_TI...	CLAIM_INIT_TI...
	Integer	Integer	Decimal	Date	Date
1	3	1	157654.9	2017-04-25	2017-04-28
2	3	3	226154.5	2018-08-26	2018-08-31
3	1	1	83968.6	2016-01-07	2016-01-11
4	1	1	309570.3	2016-12-11	2016-12-18
5	3	3	136633.9	2017-06-06	2017-06-08
6	1	3	326514.1	2018-03-06	2018-03-19
7	2	3	58477.9	2017-08-13	2017-08-18
8	1	3	176476.9	2017-02-10	2017-02-18
9	3	3	277812.7	2018-02-16	2018-02-24
10	5	3	190541.3	2016-12-28	2017-01-11
11	1	3	290975.1	2017-08-12	2017-08-25
12	3	1	159873.2	2018-07-30	2018-07-30
13	1	3	391866.9	2018-02-22	2018-02-26
14	3	1	384000.6	2017-02-06	2017-02-06

SOURCE FILE: AutoInsClaims.csv | SAMPLE SIZE: First 975 rows

27 Steps

Data Source: AutoInsClaims.csv

Convert column type: AUTOMATIC

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Remove: Removed HOUSEHOLD_ID

Remove: Removed DRIVER_ID

Remove:

LOCATION: Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME: AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS: 27

DATA REFINERY FLOW OUTPUT

Location: Prediction of Fraud Detection in Au...

Data set name: cleansed claim date

Activate Windows: Go to Settings to activate Windows.

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 | All | Search | Upgrade | Hiral Jain's Account

Projects / Prediction of Fraud Detection in ... | Launch IDE | Add to project +

Overview Assets Environments Jobs Access Control Settings

Q What assets are you looking for?

▼ Data assets

0 assets selected.

Name	Type	Created by	Last modified
CSV AutoInsClaims.csv	Data Asset	Hiral Jain	Mar 06, 2021, 09:03 PM

▼ Data Refinery flows

New Data Refinery flow +

Name	Type	Created by	Last modified
AutoInsClaims.csv_flow	Data Refinery flow	Hiral Jain	Mar 06, 2021, 11:02 PM

Drop files here or browse for files to upload.

Activate Windows: Go to Settings to activate Windows.

eu-gb.dataplatform.cloud.ibm.com/projects/e9d5173c-37ad-4928-8b31-386a6e8e98b3/asse...

IBM Cloud Pak for Data All Search Upgrade Hiral Jain's Account

Projects / Prediction of Fraud Detection in ... / 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...	POLICE_REPORT	CLAIMS_AT_LO...
	Integer	Integer	Decimal	Date	Date	Integer	Integer
1	3	1	157654.9	2017-04-25	2017-04-28	1	1
2	3	3	226154.5	2018-08-26	2018-08-31	0	1
3	1	1	83968.6	2016-01-07	2016-01-11	0	2
4	1	1	309570.3	2016-12-11	2016-12-18	0	1
5	3	3	136633.9	2017-06-06	2017-06-08	0	1
6	1	3	326514.1	2018-03-06	2018-03-19	0	1
7	2	3	58477.9	2017-08-13	2017-08-18	0	1
8	1	3	176476.9	2017-02-10	2017-02-18	0	1
9	3	3	277812.7	2018-02-16	2018-02-24	0	1
10	5	3	190541.3	2016-12-28	2017-01-11	0	1
11	1	3	290975.1	2017-08-12	2017-08-25	0	1
12	3	1	159873.2	2018-07-30	2018-07-30	1	5
13	1	3	391866.9	2018-02-22	2018-02-26	0	1
14	3	1	384000.6	2017-02-06	2017-02-06	0	3

SOURCE FILE: AutoInsClaims.csv SAMPLE SIZE: First 975 rows

Information

Details Help

Edit

LOCATION
Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow
Enter a description of the Data Refinery flow

STEPS
26

DATA REFINERY FLOW OUTPUT
Activate Windows
Go to Settings to activate Windows.
Prediction of Fraud Detection in Au...

IBM Cloud Pak for Data All Search Upgrade Hiral Jain's Account

Projects / Prediction of Fraud Detection in ... / AutoInsClaims.csv_flow

Operation x 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_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
2018-07-30
2018-02-22
2017-02-06

SOURCE FILE: AutoInsClaims.csv SAMPLE SIZE: First 975 rows

Information

Details Help

Edit

LOCATION
Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow
Enter a description of the Data Refinery flow

STEPS
26

DATA REFINERY FLOW OUTPUT
Activate Windows
Go to Settings to activate Windows.
Prediction of Fraud Detection in Au...

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

Operation + Code an operation to cleanse and shape your data

Data Profile Visualizations Steps

	DRIVERS_LICENSE Date	DATE_AT_CURRENT Date	COMMUTE_DISTANCE Integer	loss_event_days Decimal	Expiry_days Decimal
1	2018-08-19	1999-04-16	0	115	262
2	2018-01-27	2011-06-09	0	238	165
3	2019-11-19	2005-05-21	0	7	236
4	2019-05-16	2000-03-11	0	346	206
5	2020-07-06	2012-07-04	0	157	7
6	2021-09-26	2001-07-15	0	65	15
7	2019-05-10	2004-04-08	0	225	189
8	2019-04-26	1999-02-02	0	41	223
9	2021-04-15	2005-03-07	0	47	286
10	2019-08-07	2012-03-04	0	363	174
11	2019-07-02	2017-05-03	0	224	104
12	2018-03-29	2008-07-06	0	211	53
13	2018-04-01	2012-05-15	0	53	10
14	2019-07-13	2007-01-14	0	37	155

SOURCE FILE: AutoInsClaims.csv SAMPLE SIZE: First 975 rows

28 Steps

- Data Source: AutoInsClaims.csv
- Convert column type: Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.
- Remove: Removed HOUSEHOLD_ID
- Remove: Removed DRIVER_ID
- Remove

Information

Details Help

Edit

LOCATION: Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME: AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS: 28

DATA REFINERY FLOW OUTPUT: Activate Windows

Location settings to activate Windows.

Prediction of Fraud Detection in Au...

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.

IBM Cloud Pak for Data

All

Search

Upgrade

Hiral Jain's Account

H3

Projects / Prediction of Fraud Detection in ... / AutoInsClaims.csv_flow

Operation +

Code an operation to cleanse and shape your data

DataProfileVisualizations

Steps

	DATE_AT_CURR... Date	COMMUTE_DIS... Integer	loss_event_days Decimal	Expiry_days Decimal	days_from_lice... Integer
1	1999-04-16	0	115	262	0
2	2011-06-09	0	238	165	1
3	2005-05-21	0	7	236	0
4	2000-03-11	0	346	206	0
5	2012-07-04	0	157	7	0
6	2001-07-15	0	65	15	0
7	2004-04-08	0	225	189	0
8	1999-02-02	0	41	223	0
9	2005-03-07	0	47	286	0
10	2012-03-04	0	363	174	0
11	2017-05-03	0	224	104	0
12	2008-07-06	0	211	53	1
13	2012-05-15	0	53	10	0
14	2007-01-14	0	37	155	0

SOURCE FILE: AutoInsClaims.csvSAMPLE SIZE: First 975 rows

https://eu-gb.dataplatform.cloud.ibm.com/home?context=cpdaas

30 Steps

Data Source

AutoInsClaims.csv

Convert column type

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Remove

Removed HOUSEHOLD_ID

Remove

Removed DRIVER_ID

Remove

Information

DetailsHelp

Edit

LOCATION

Prediction of Fraud Detection in Auto Insurance Claim

DATA REFINERY FLOW NAME

AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS

30

DATA REFINERY FLOW OUTPUT

Activate Windows

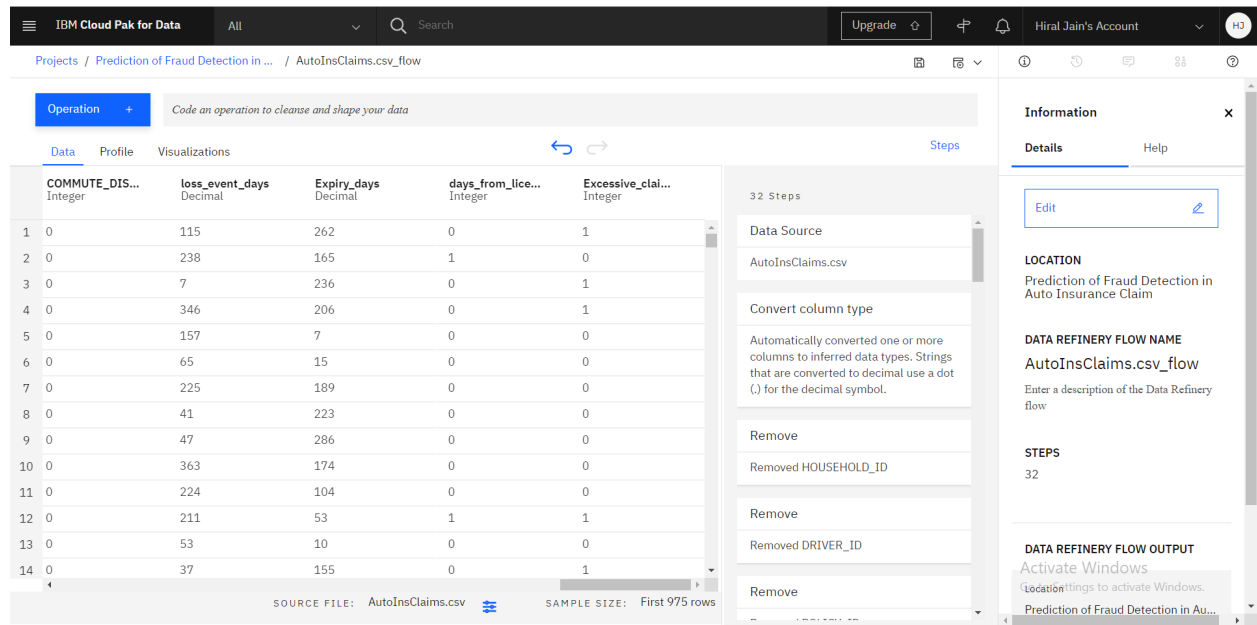
Location: to activate Windows.

Prediction of Fraud Detection in Au...

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.

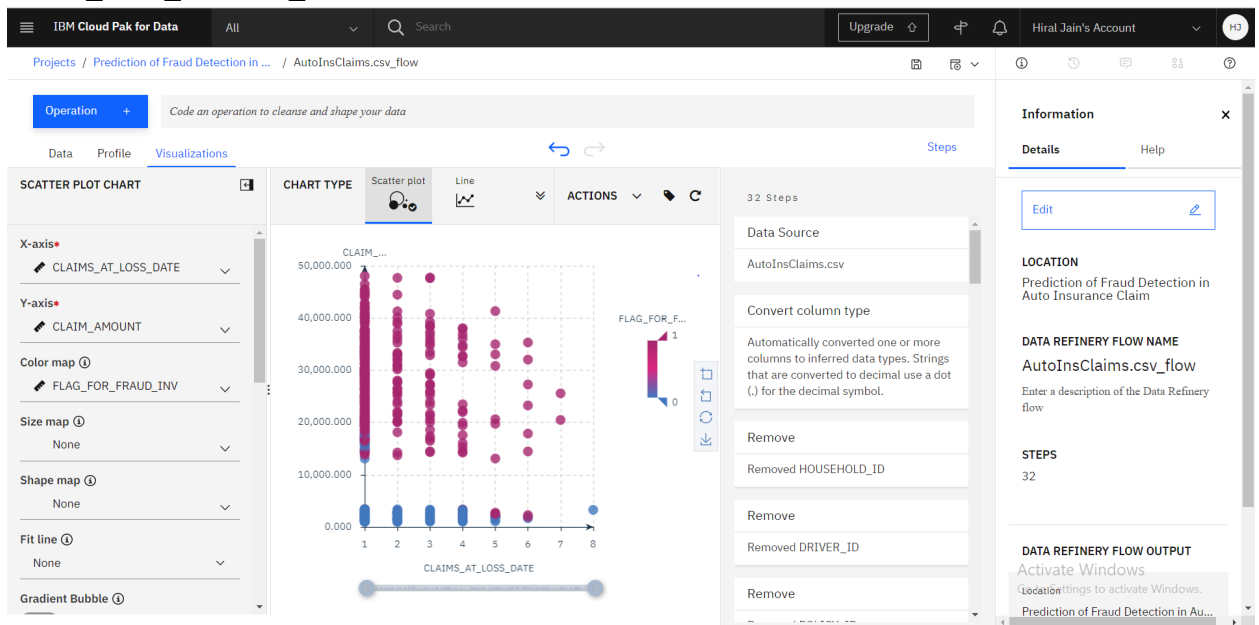


4. Data Refinery Visualization-

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

1. CLAIM_AT_LOSS_DATE

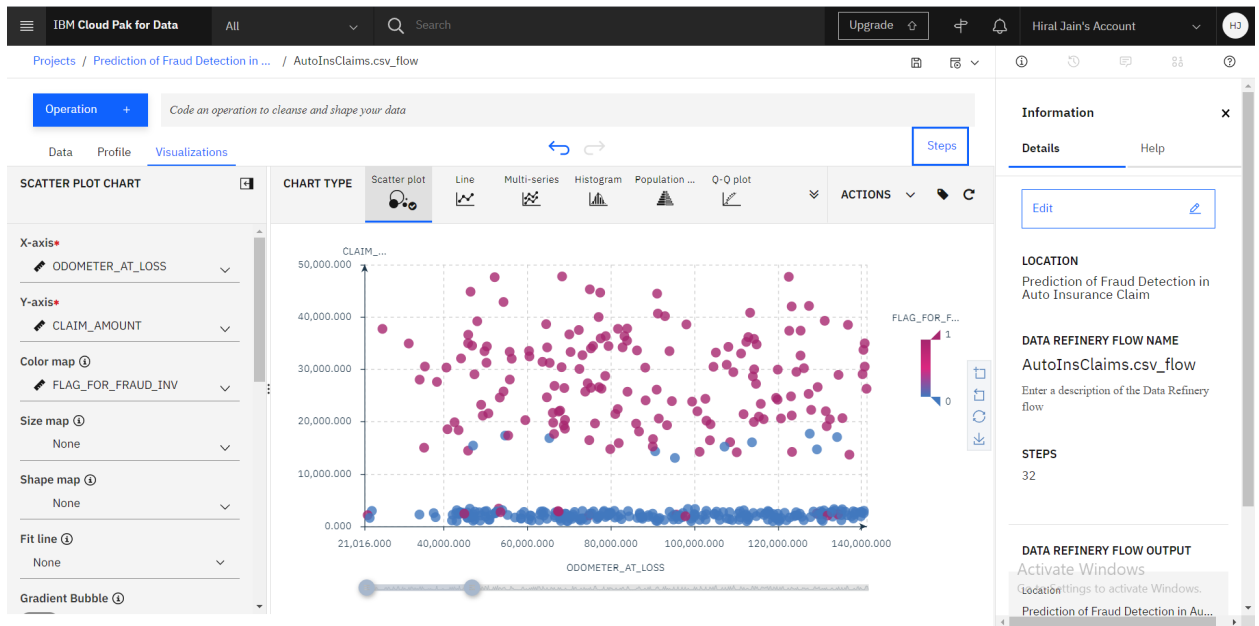
Select the scatter plot and for the X-axis select CLAIMS_AT_LOSS_DATE, for the Y-axis select CLAIM_AMOUNT and for the Color map, select FLAG_FOR_FRAUD_INV.



2. ODOMETER_AT_LOSS

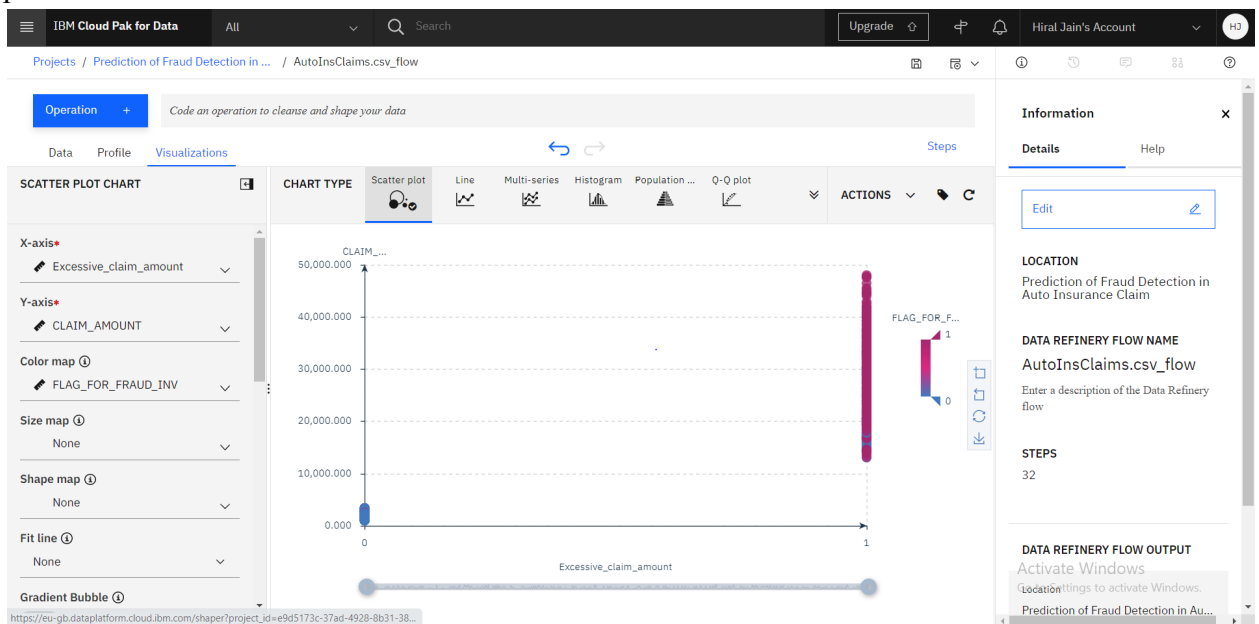
Select the scatter plot and for the X-axis select ODOMETER_AT_LOSS, for the Y-axis select CLAIM_AMOUNT and for the Color map, select

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.			