

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

I. Aim/Overview of the practical: To implement the **hypothesis 2** into the Fraud Detection (**Loss claim after Expired License**).

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:

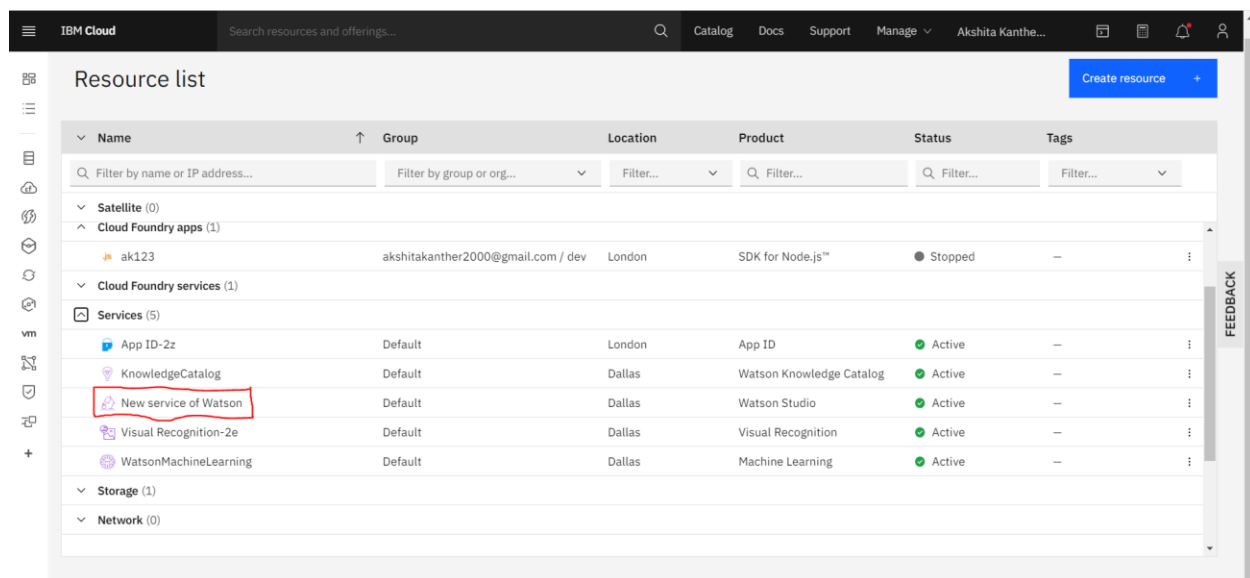
- Internet connectivity
- IBM cloud account
- IBM WATSON service
- NIC Data Set

IV. Hypothesis

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

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.



IBM Cloud Pak for Data All Search Upgrade Akshita Kanther's Account

Welcome, Akshita!

Watson Studio • Watson Knowledge Catalog • Watson Machine Learning

Learn by example

Step through solving a specific business problem in a sample project.

[Take a guided tutorial](#)

Work with data

Create a project for your team to prepare data, find insights, or build models.

[Create a project](#)

Extend your capabilities

Add tools, databases, or other features by creating services instances.

[Create a service](#)

Quick navigation

- Projects
- Catalogs
- Deployments

Support

- Documentation
- FAQ

Overview

Recent projects

Accident_and_Vandalism	Mar 03, 2021 11:20 AM
Predicting Fraud in Auto Insurance Claim	Mar 03, 2021 10:26 AM

Recent catalogs

No catalogs

Your catalogs show here after you create or join them. Click New catalog to get started.

[New catalog](#)

Notifications

Watson Visual Recognition model Accident training has completed

Mar 03, 2021 12:06 PM

New in gallery

3. Load the AutoInsClaims dataset on the cloud.

← Back

Create a project

Choose whether to create an empty project or to preload your project with data and analytical assets. Add collaborators and data, and then choose the right tools to accomplish your goals. Add services as necessary.

Create an empty project

Add the data you want to prepare, analyze, or model. Choose tools based on how you want to work: write code, create a flow on a graphical canvas, or automatically build models.

[AutoAI](#) AutoAI experiment tool Fully automated approach to building a classification or reg...

Create a project from a sample or file

Get started fast by loading existing assets. Choose a project file from your system, or choose a curated sample project.

USE TO

Prepare and visualize data
Analyze data in notebooks
Train models

USE TO

Learn by example
Build on existing work
Run tutorials

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

Projects / Predicting Fraud in Auto Insuran...

Overview

Predicting Fraud in Auto Insurance Claim

Last Updated: Mar 03, 2021

[Readme](#)

0

Assets

1

Collaborators

Overview

Date created

Mar 03, 2021

Description

No description available

Storage



0 Byte used
Cloud Object Storage

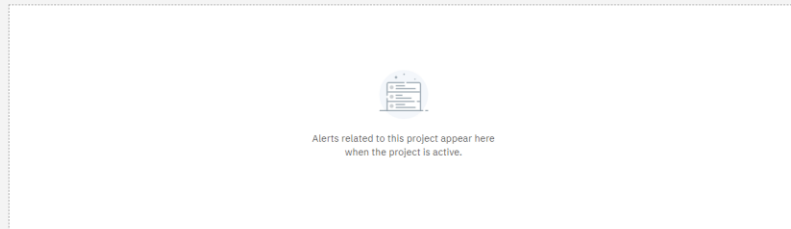
Collaborators



Akshita Kanther
Admin

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

Recent activity



Readme

[Back to top](#)

Projects / Predicting Fraud in Auto Insuran...

Overview

Assets

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.

IBM Cloud Pak for Data

Projects / Predicting Fraud in Auto Insuran...

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	Akshita Kanther	Mar 08, 2021, 04:57 PM

Drop files here or [browse](#) for files to upload.

Stay on the page until upload completes. Incomplete uploads are cancelled.

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

Projects / Fraud Detection / AutoInsClaims.csv

Preview Profile Activities

Schema: 38 Columns
Preview: 975 rows

Last refresh: just now

[Refine](#)

HOUSEHOLD_ID	DRIVER_ID	POLICY_ID	CLAIM_ID	INCIDENT_COUNT	DESCRIPTION	CLAIM_STATUS	ODOMETER_AT_MILE	LOSS_EVENT_DATE
CH42335	XZ32837	NW5567882	A-2017-UU907	3		1	157654.9	4/25/17
IH49805	VVR6423	UR4864804	A-2018-FJ481	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	ONMS465	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	XR3994270	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-Q0338	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

Operation + Code an operation to cleanse and shape your data

Data Profile Visualizations

	HOUSEHOLD_ID	DRIVER_ID	POLICY_ID	CLAIM_ID	INCIDENT_CAU...	DESCRIPTION	CLAIM_STATUS
1	CH42335	XZ32837	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-Z0863	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
13	CH81331	BO71566	704447870	A-2018-UT470	1		3

SOURCE FILE: AutoInsClaims.csv SAMPLE SIZE: First 50 rows

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

->After Converting

	INCIDENT_CAU...	CLAIM_STATUS	ODOMETER_AT...	LOSS_EVENT_TI...	CLAIM_INIT_TI...
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	2	201966.0	2018-07-27	2018-07-26

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.

	ODOMETER_AT...	LOSS_EVENT_TI...	CLAIM_INIT_TI...	POLICE_REPORT	CLAIMS_AT_LO...
1	157654.9	4/25/17	4/28/17	1	1
2	226154.5	8/26/18	8/31/18	0	1
3	83968.6	1/7/16	1/11/16	0	2
4	309570.3	12/11/16	12/18/16	0	1
5	136633.9	6/6/17	6/8/17	0	1
6	326514.1	3/6/18	3/19/18	0	1
7	58477.9	8/13/17	8/18/17	0	1
8	176476.9	2/10/17	2/18/17	0	1
9	277812.7	2/16/18	2/24/18	0	1
10	190541.3	12/28/16	1/11/17	0	1
11	290975.1	8/12/17	8/25/17	0	1
12	159873.2	7/30/18	7/30/18	1	5
13	201966.0	2/22/18	2/26/18	0	1

DATA REFINERY FLOW NAME
AutoInsClaims.csv_flow

Enter a description of the Data Refinery flow

STEPS
34

Location
Fraud Detection/Data assets

Data set name
Clean Claim Data

Enter a description of the resulting data set.

✓ If the data set already exists, overwrite the data in the existing data set with the Data Refinery flow output.

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.

The screenshot shows the 'Assets' tab in a data management interface. Under 'Data Refinery flows', the flow 'AutoInsClaims.csv_flow' is listed and highlighted with a red rectangular box. The interface includes a search bar, a table of assets, and a sidebar for data upload options.

Name	Type	Created by	Last modified
AutoInsClaims.csv	Data Asset	Akshita Kanther	Feb 14, 2021, 10:35 PM
AutoInsClaims.csv_flow	Data Refinery flow	Akshita Kanther	Mar 01, 2021, 08:13 PM

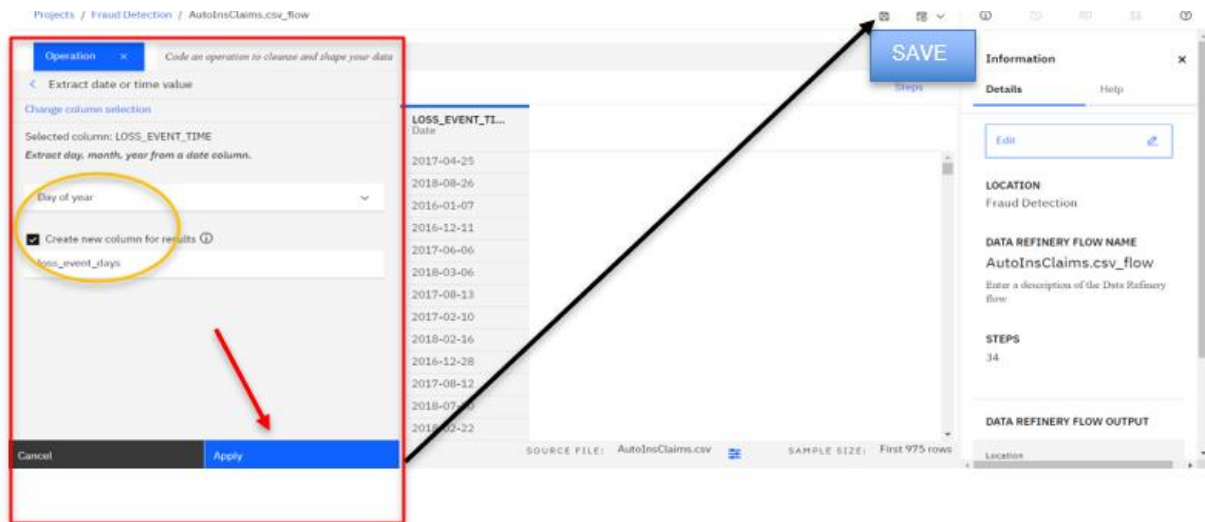
The screenshot displays the 'Operation' tab of a Data Refinery flow configuration. The 'LOSS_EVENT_TIME' column is highlighted with a red box. The 'Steps' panel on the right shows a list of conversion steps. The data table below shows the first 13 rows of the dataset.

	INCIDENT_CAUSE...	CLAIM_STATUS	ODOMETER_AT...	LOSS_EVENT_TIME...	CLAIM_INITI...
	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	201844.0	2018-07-13	2018-07-16

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

34 Steps

- Convert column type
- Manually converted data types for 1 column.
- Convert column type
- Manually converted data types for 1 column.
- Convert column type
- Manually converted data types for 1 column.
- Convert column type
- Manually converted data types for 1 column.



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

	DRIVERS_LICE...	DATE_AT_CURR...	COMMUTE_DIS...	loss_event_days	Expiry_days
	Date	Date	Integer	Decimal	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

->After Saving

Operation + Code an operation to cleanse and shape your data

Data Profile Visualizations

	COMMUTE_DIS...	LOSS_EVENT_D...	EXPIRY_DAYS	Suspicious_Clai...	Suspicious_Clai...	Days_of_L_Expiry	Excessive_Clai...
	Integer	Decimal	Decimal	Decimal	Integer	Integer	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
13	0	53	10	-17	1	0	0

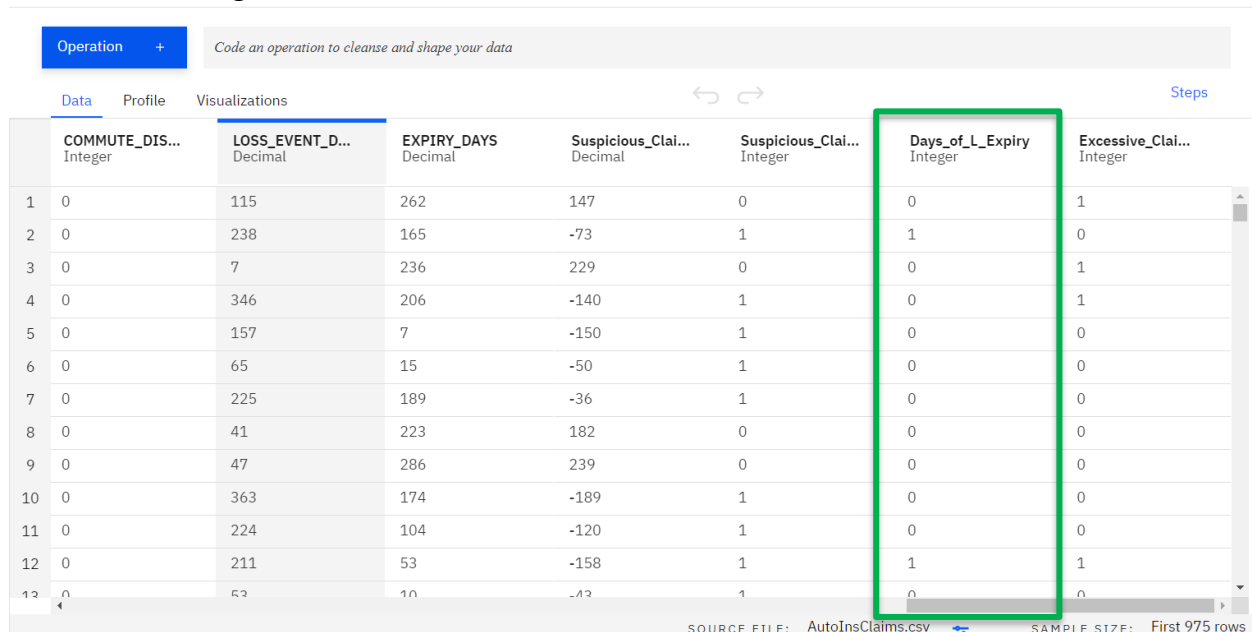
SOURCE FILE: AutoInsClaims.csv SAMPLE SIZE: First 975 rows

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



The screenshot shows a data manipulation interface with a table of 13 rows and 7 columns. The columns are: COMMUTE_DIS... (Integer), LOSS_EVENT_D... (Decimal), EXPIRY_DAYS (Decimal), Suspicious_Clai... (Decimal), Suspicious_Clai... (Integer), Days_of_L_Expiry (Integer), and Excessive_Clai... (Integer). The 'Days_of_L_Expiry' column is highlighted with a green box. The interface includes a 'Data' tab, a 'Profile' tab, and a 'Visualizations' tab. A 'Steps' button is visible in the top right corner. The source file is 'AutoInsClaims.csv' and the sample size is 'First 975 rows'.

	COMMUTE_DIS...	LOSS_EVENT_D...	EXPIRY_DAYS	Suspicious_Clai...	Suspicious_Clai...	Days_of_L_Expiry	Excessive_Clai...
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
13	0	53	10	-13	1	0	0

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.

The screenshot displays the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and user information. Below this, the breadcrumb shows 'Projects / Fraud Detection / AutoInsClaims.csv_flow'. The main area is divided into 'Operation' (with a code editor placeholder) and 'Data' (showing a table). The table has columns: 'COMMUTE_DIS...', 'LOSS_EVENT_D...', 'EXPIRY_DAYS', 'Suspicious_Clai...', 'Suspicious_Clai...', 'Days_of_L_Expiry', and 'Excessive_Clai...'. The right sidebar, titled 'Information', contains 'Details' and 'Help' tabs, an 'Edit' button, 'LOCATION' (Fraud Detection), 'DATA REFINERY FLOW NAME' (AutoInsClaims.csv_flow), 'STEPS' (34), and 'DATA REFINERY FLOW OUTPUT' (Location).

	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
13	0	53	10	-12	1	0	0

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

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 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.			