

Name: Aditya Yadav

Data Analytics

VIT Vellore, Tamil Nadu

Assignment-2

Dataset: challengers.csv

Link to Dashboard: <https://dataplatform.cloud.ibm.com/projects/56e6a2b6-9271-4cde-a504-bafd222f2e4e/assets?context=cpdaas>

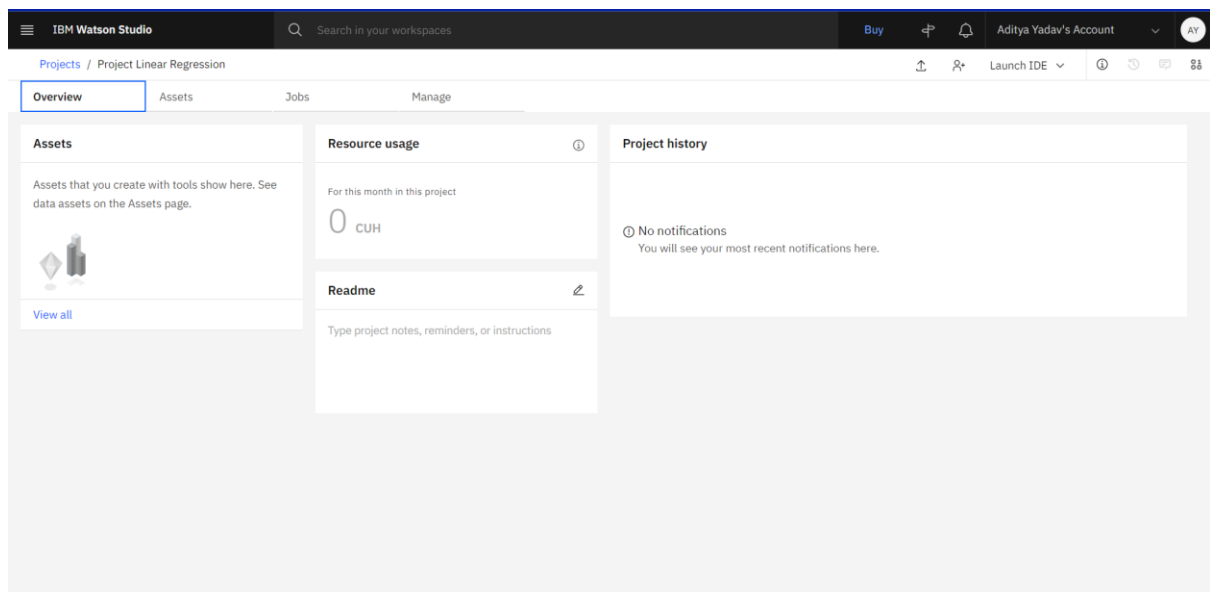


Fig: Creating a new project and the overview screen of the project

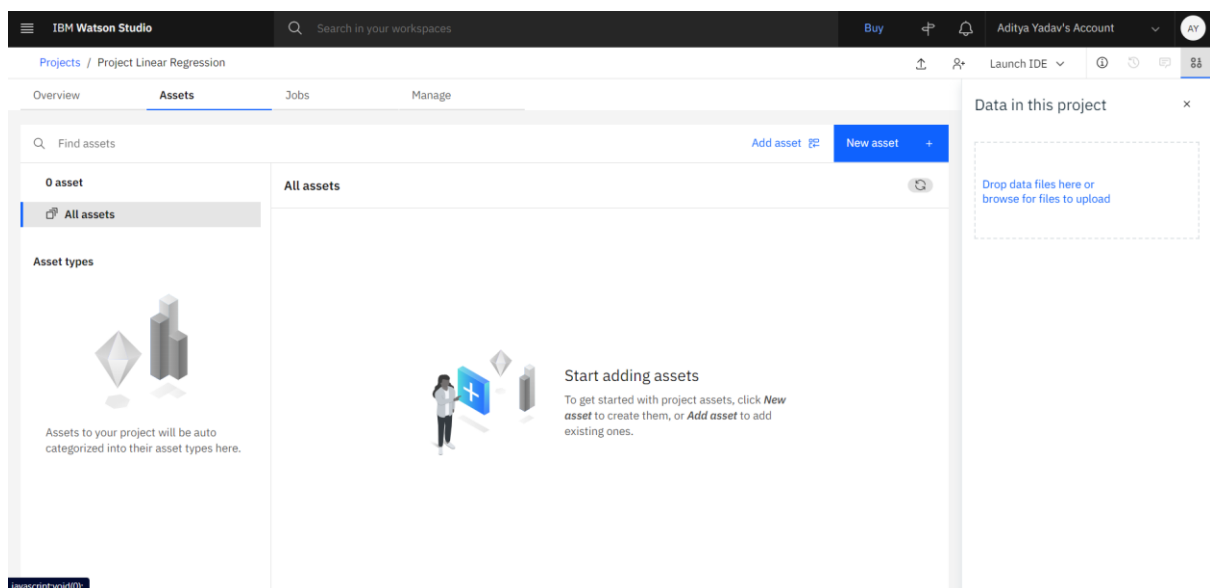


Fig: Go to assets and load your data.

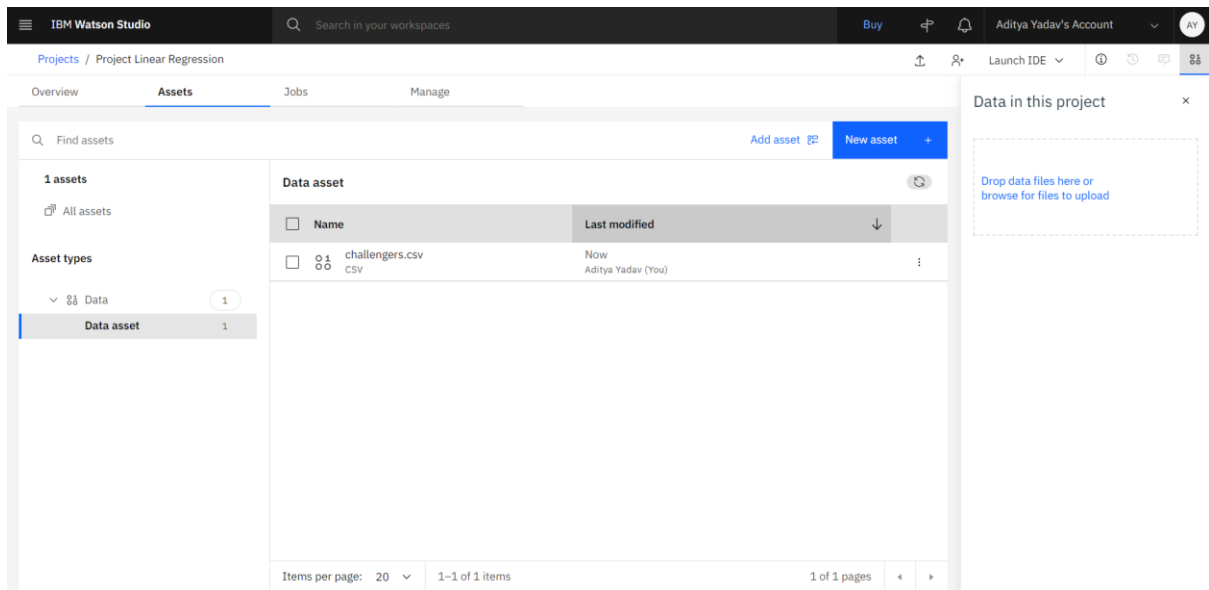


Fig: uploaded dataset appears in the data asset section.

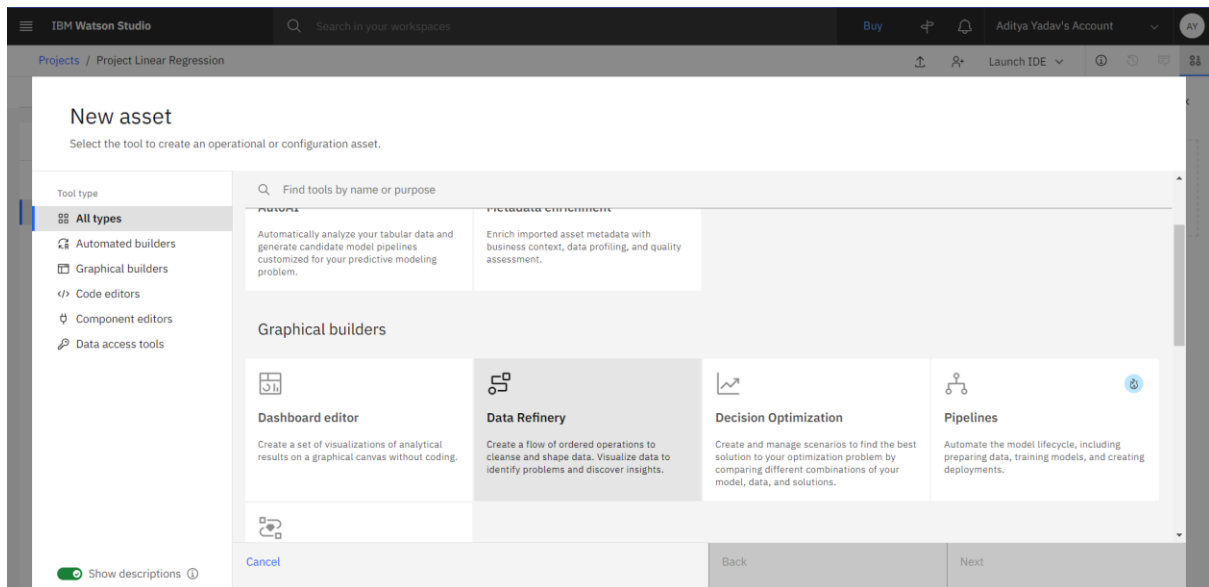


Fig: Add new asset namely data refinery

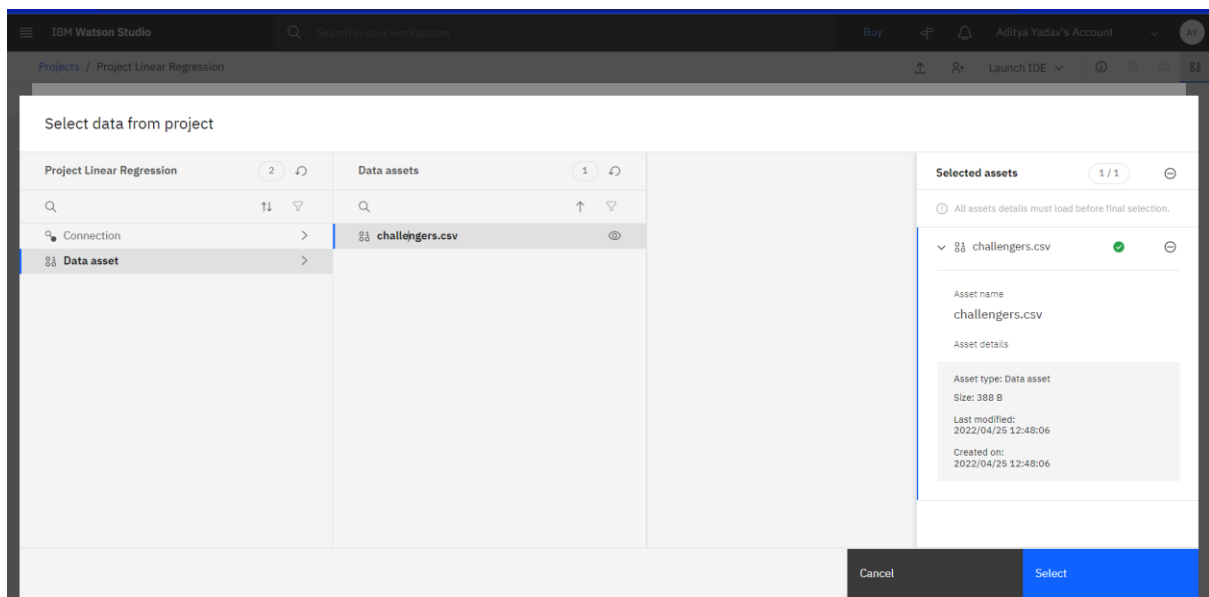


Fig: select dataset of the project after adding data refinery.

	o_ring_ct	o_ring_fallu...	temperature	pressure	launch_id
1	6	0	66	50	1
2	6	1	70	50	2
3	6	0	69	50	3
4	6	0	68	50	4
5	6	0	67	50	5
6	6	0	72	50	6
7	6	0	73	100	7
8	6	0	70	100	8
9	6	1	57	200	9
10	6	1	63	200	10
11	6	1	70	200	11
12	6	0	78	200	12
13	6	0	67	200	13
14	6	2	53	200	14

SOURCE FILE: challengers.csv SAMPLE SIZE: First 23 rows

Details

LOCATION
Project Linear Regression

DATA REFINERY FLOW NAME
challengers.csv_flow

STEPS
0

DATA REFINERY FLOW OUTPUT
Location

Fig: descriptive statistics of the data that we imported.

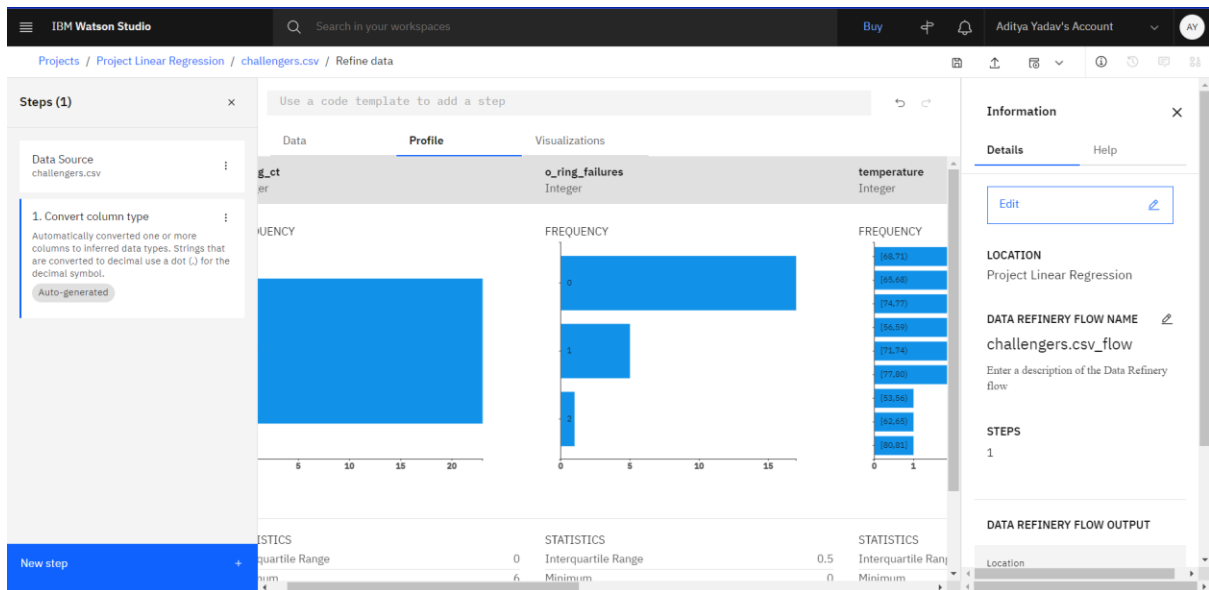


Fig: generated profiles of our data.

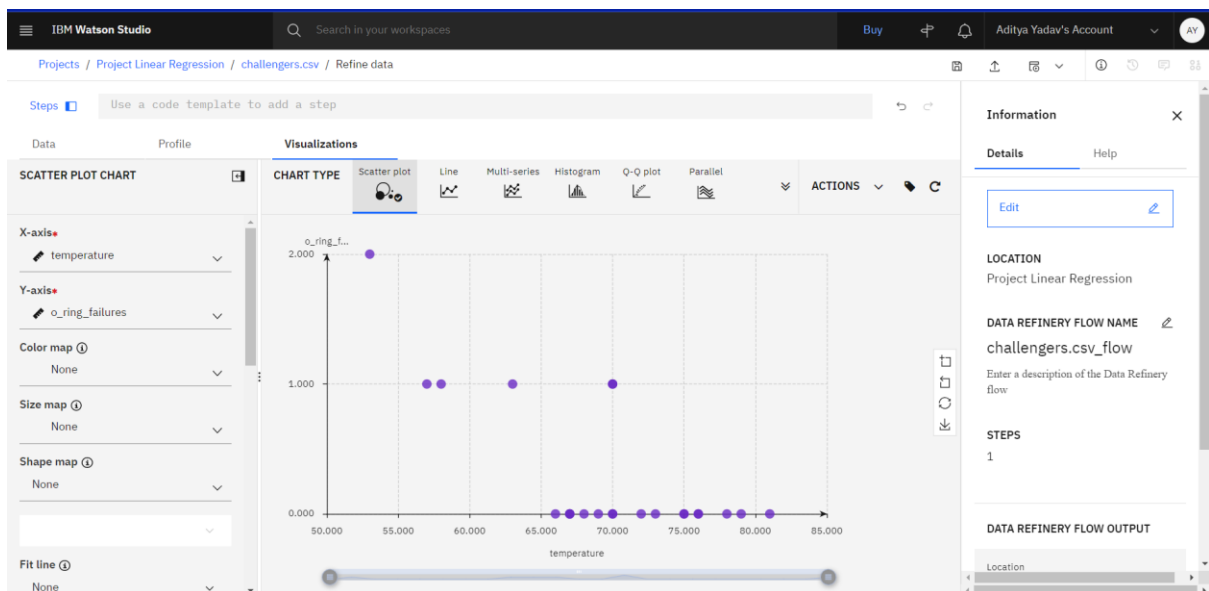


Fig: Visualizing temperature with respect to o_ring_failures.

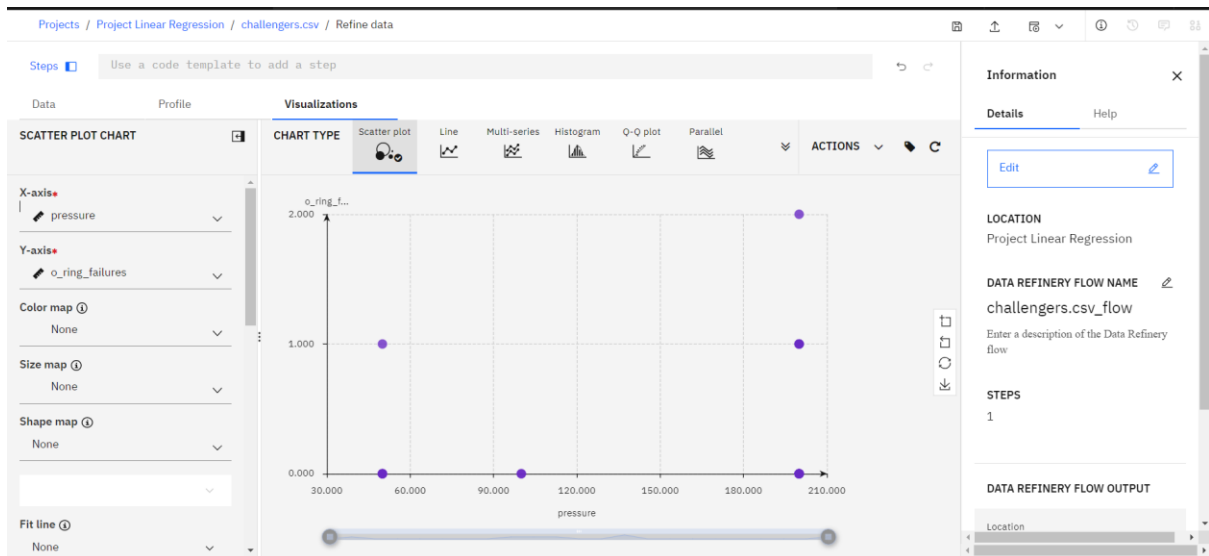


Fig: Visualizing pressure with respect to o_ring_failures and save.

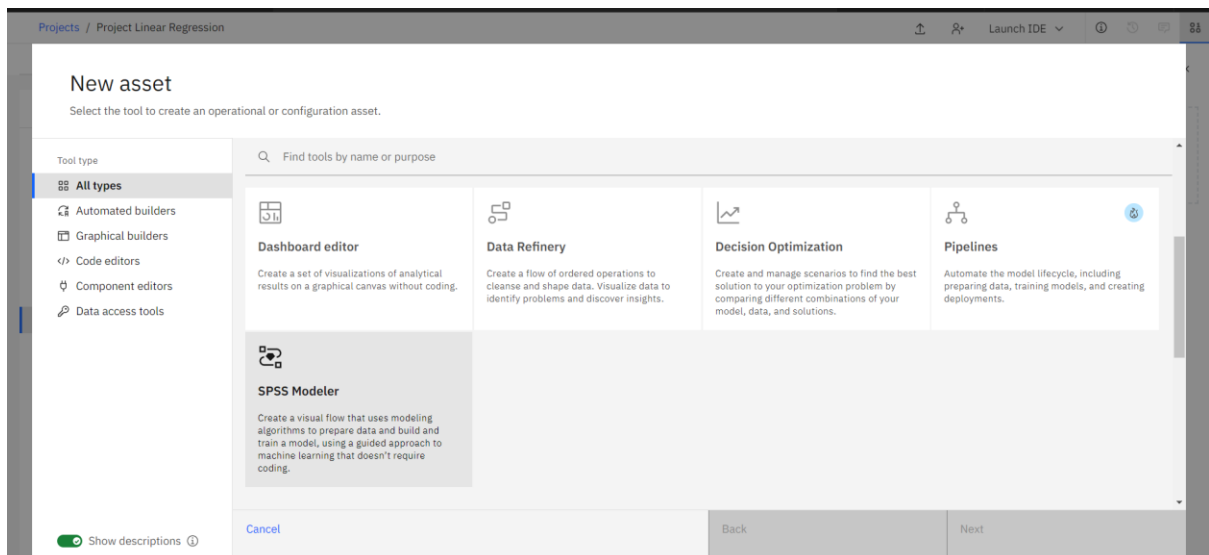


Fig: Add a new asset for SPSS modeler.

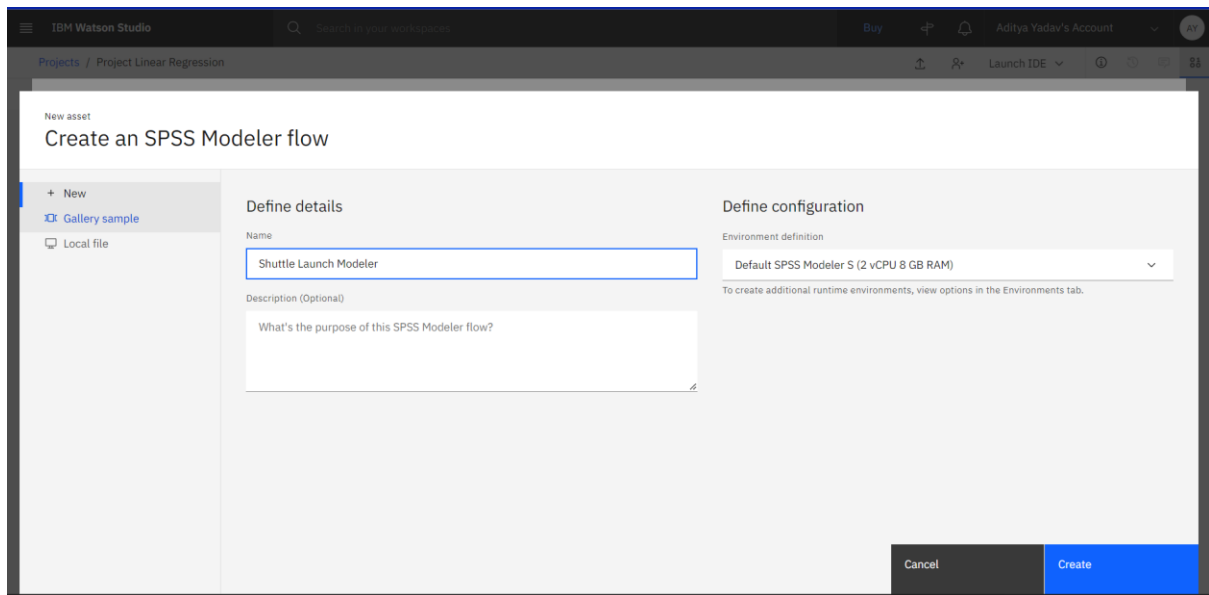


Fig: create an SPSS modeler flow

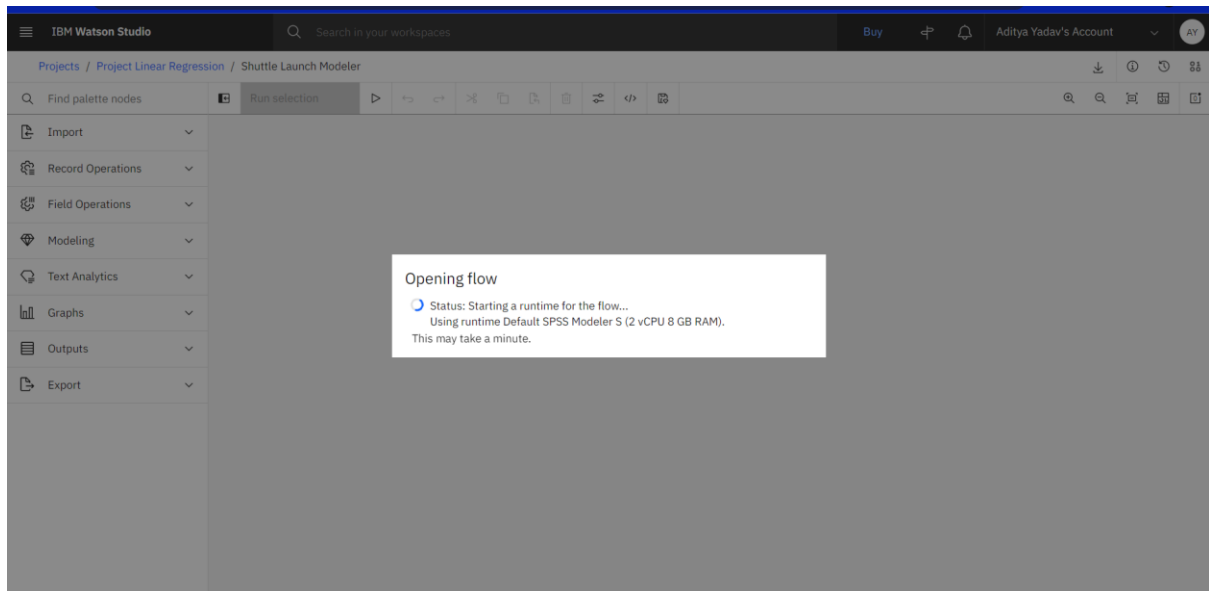


Fig: Workbench of the modeler is being created.

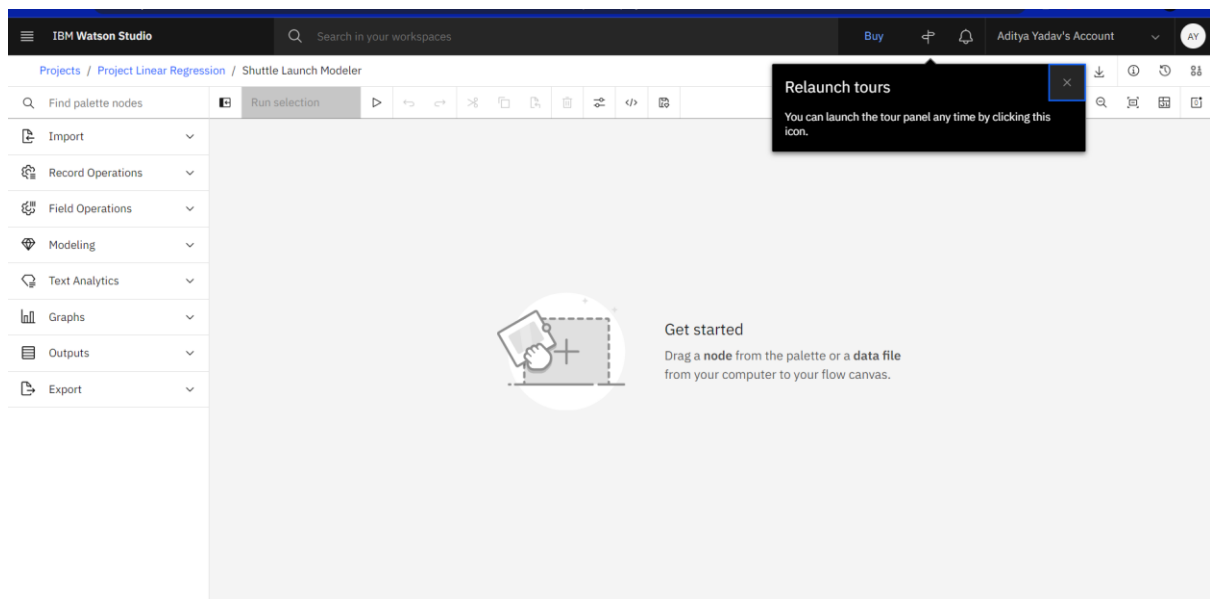


Fig: work-bench

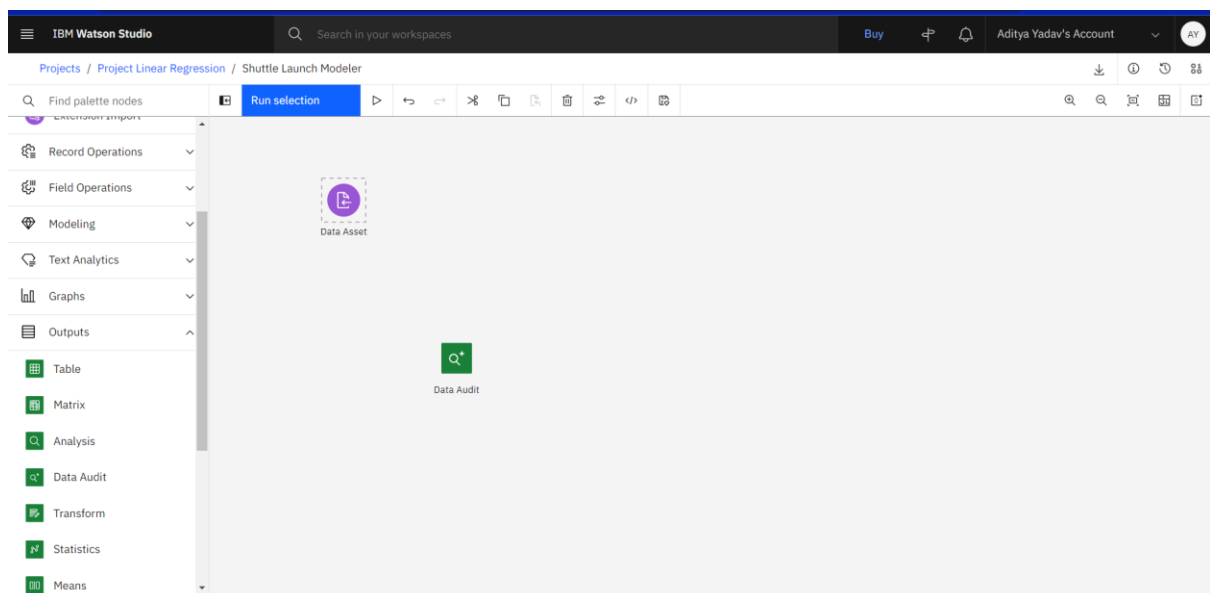


Fig: drag and drop data asset from import and connect dataset to it and drop data audit.

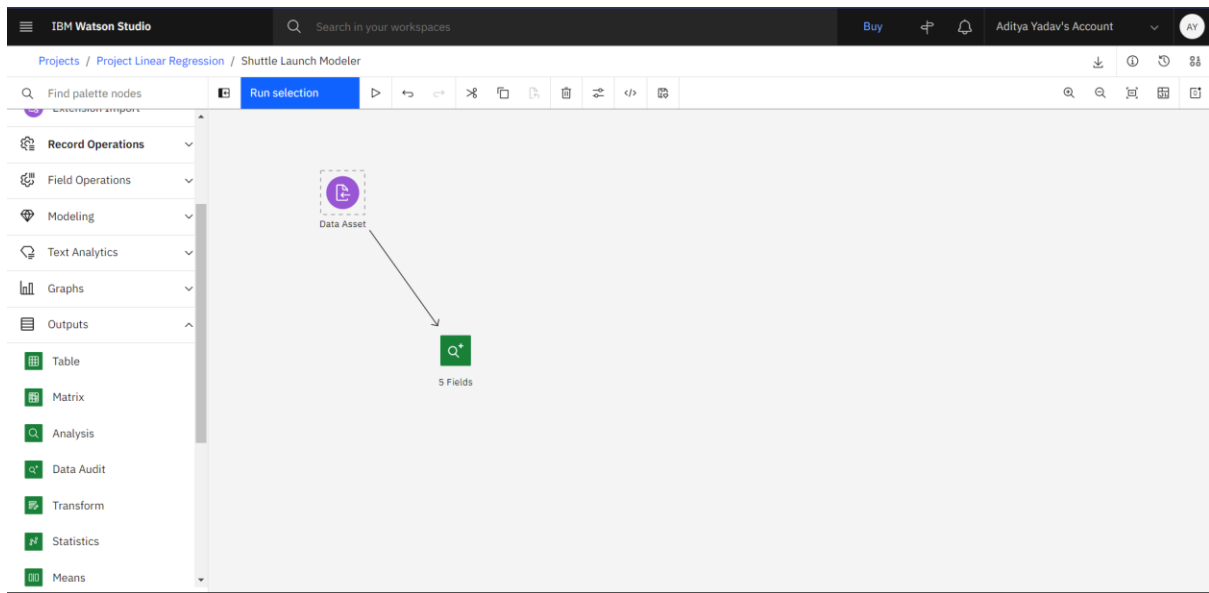


Fig: link data asset to the data audit.

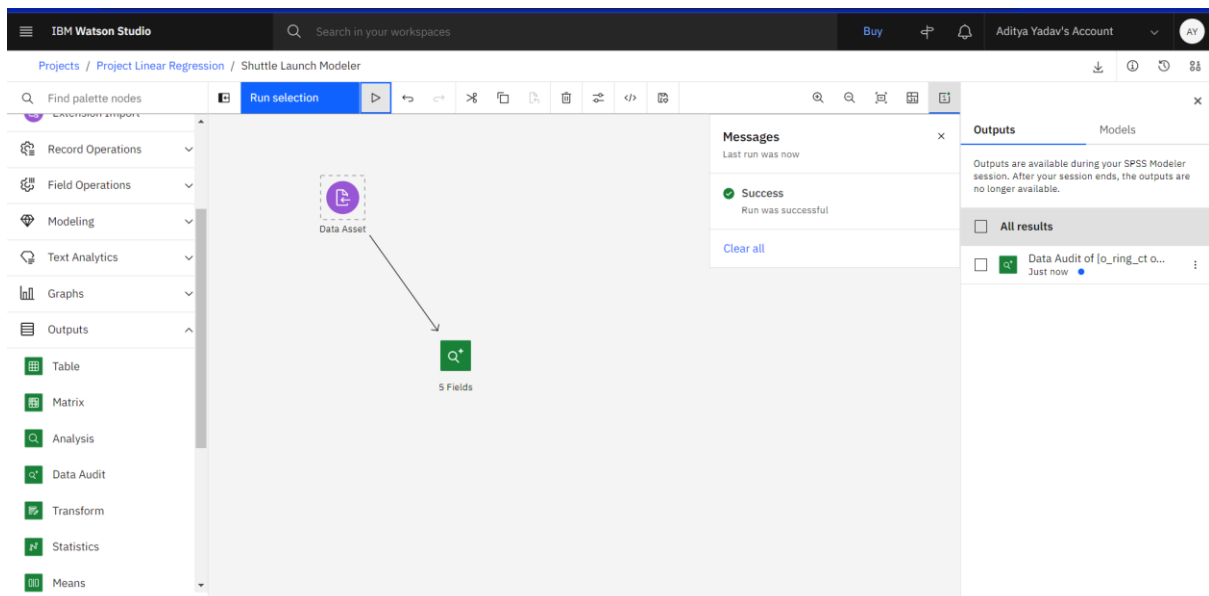


Fig: run the created model

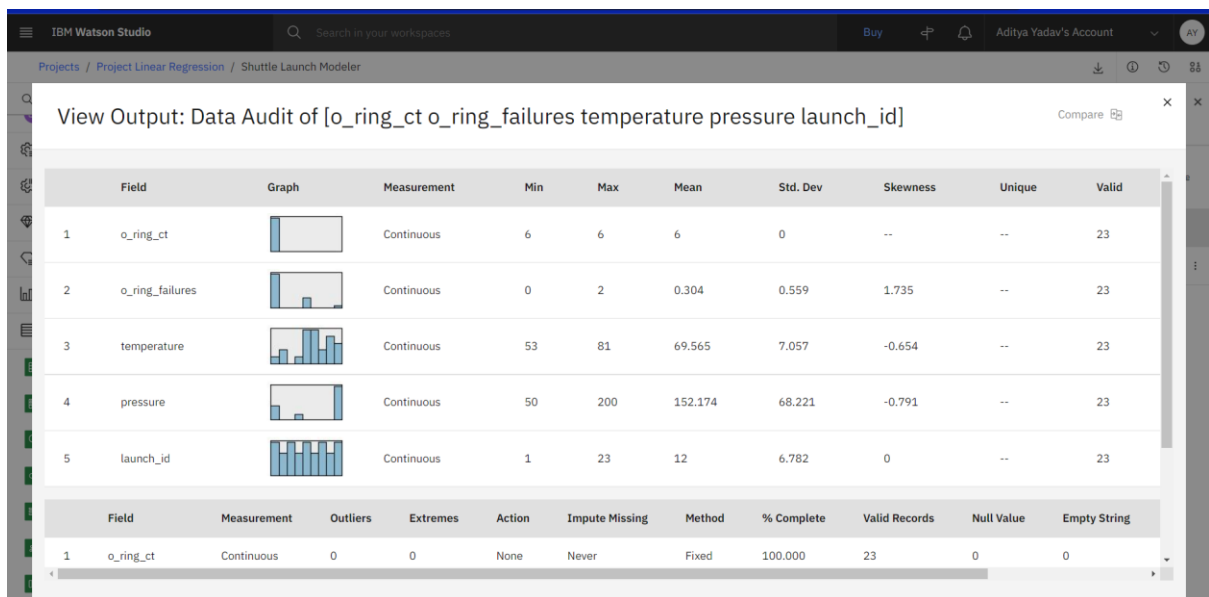


Fig: descriptive analytics of the data audit.

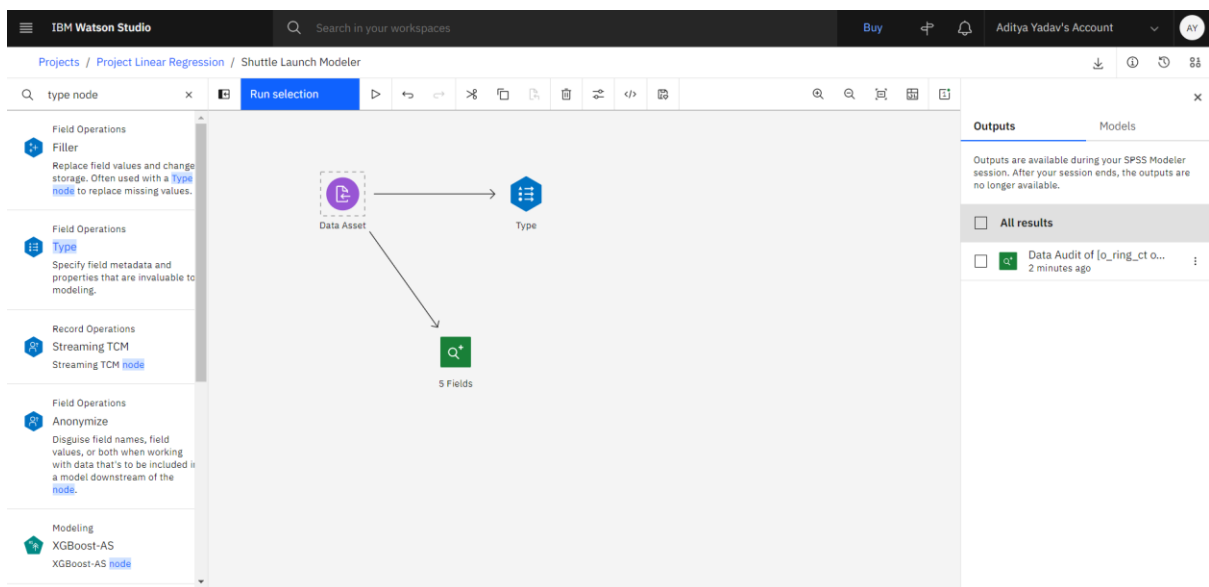


Fig: Add type node and connect it with the data asset.

IBM Watson Studio

Search in your workspaces

Buy

Aditya Yadav's Account

Projects / Project Linear Regression / Shuttle Launch Modeler

type node

Field Operations

Filler

Replace field values and change storage. Often used with a [Type node](#) to replace missing values.

Field Operations

Type

Specify field metadata and properties that are invaluable to modeling.

Record Operations

Streaming TCM

Streaming TCM [node](#)

Field Operations

Anonymize

Disguise field names, field values, or both when working with data that's to be included in a model downstream of the [node](#).

Modeling

XGBoost-AS

XGBoost-AS [node](#)

Data Asset

Type

5 Fields

Type

Settings

[Read values](#) [Clear values](#)

Find in column Field

Field	Measure	Role	Value mode	Values
<input type="checkbox"/> # o_ring_ct	Continuous	Input	Read	
<input type="checkbox"/> # o_ring_failures	Continuous	Input	Read	
<input type="checkbox"/> # temperature	Continuous	Input	Read	
<input type="checkbox"/> # pressure	Continuous	Input	Read	
<input type="checkbox"/> # launch_id	Continuous	Input	Read	

Default mode

☒ Read metadata ☐ Pass (do not scan)

[Set unique fields to None](#)

Set categorical fields to None if they exceed this many values

Set continuous integer field to ordinal if range less than or equal to

Cancel

Save

Fig: Properties of the type node.

IBM Watson Studio

Search in your workspaces

Buy

Aditya Yadav's Account

Projects / Project Linear Regression / Shuttle Launch Modeler

type node

Field Operations

Filler

Replace field values and change storage. Often used with a [Type node](#) to replace missing values.

Field Operations

Type

Specify field metadata and properties that are invaluable to modeling.

Record Operations

Streaming TCM

Streaming TCM [node](#)

Field Operations

Anonymize

Disguise field names, field values, or both when working with data that's to be included in a model downstream of the [node](#).

Modeling

XGBoost-AS

XGBoost-AS [node](#)

Data Asset

Type

5 Fields

Type

Settings

[Read values](#) [Clear values](#)

Find in column Field

Field	Measure	Role	Value mode	Values
<input type="checkbox"/> # o_ring_ct	Continuous	None	Read	
<input type="checkbox"/> # o_ring_failures	Continuous	Target	Read	
<input type="checkbox"/> # temperature	Continuous	Input	Read	
<input type="checkbox"/> # pressure	Continuous	None	Read	
<input type="checkbox"/> # launch_id	Continuous	None	Read	

Default mode

☒ Read metadata ☐ Pass (do not scan)

[Set unique fields to None](#)

Set categorical fields to None if they exceed this many values

Set continuous integer field to ordinal if range less than or equal to

Cancel

Save

Fig: make o_ring_failure as target and temperature as input.

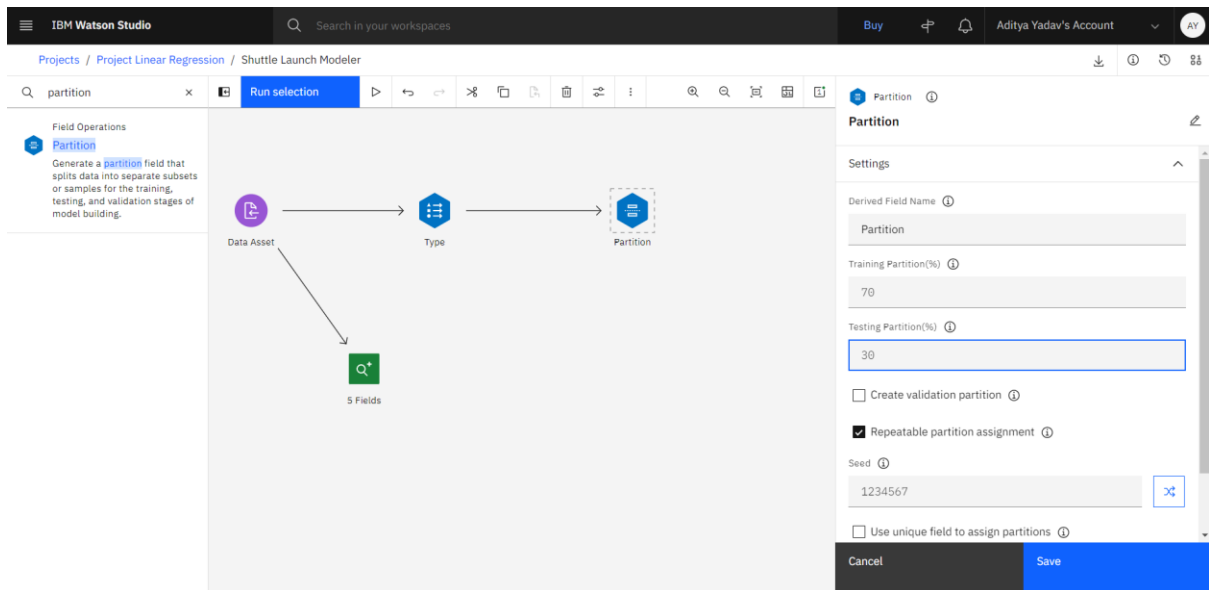


Fig: add and connect partition node and update training and testing percentage.

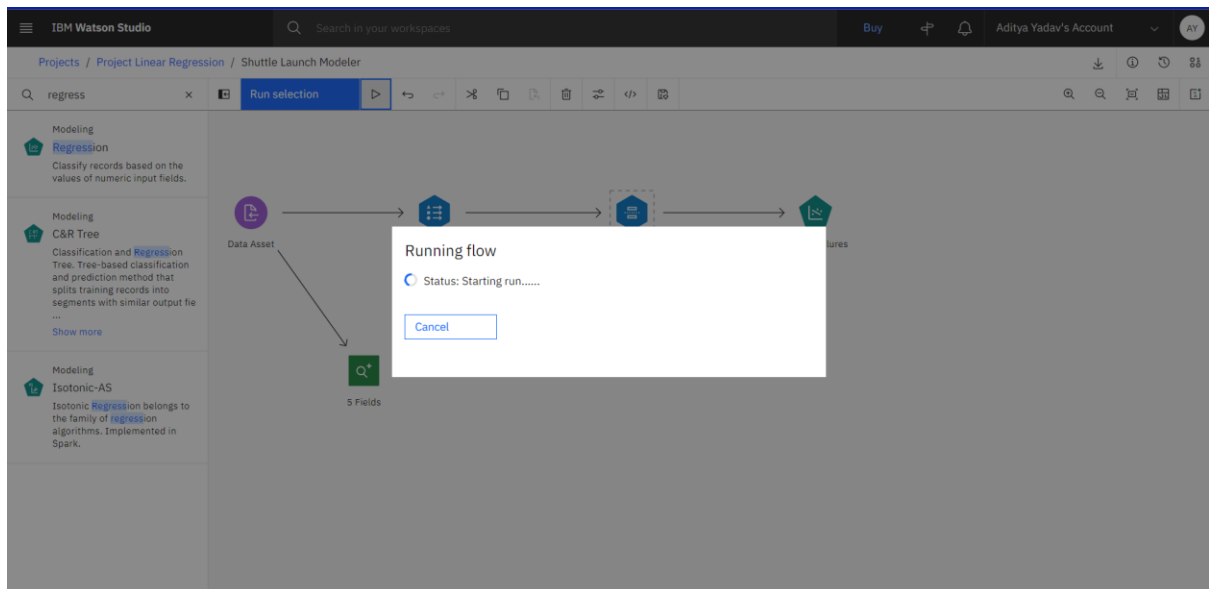


Fig: add regression model and connect to the partition node and run the model.

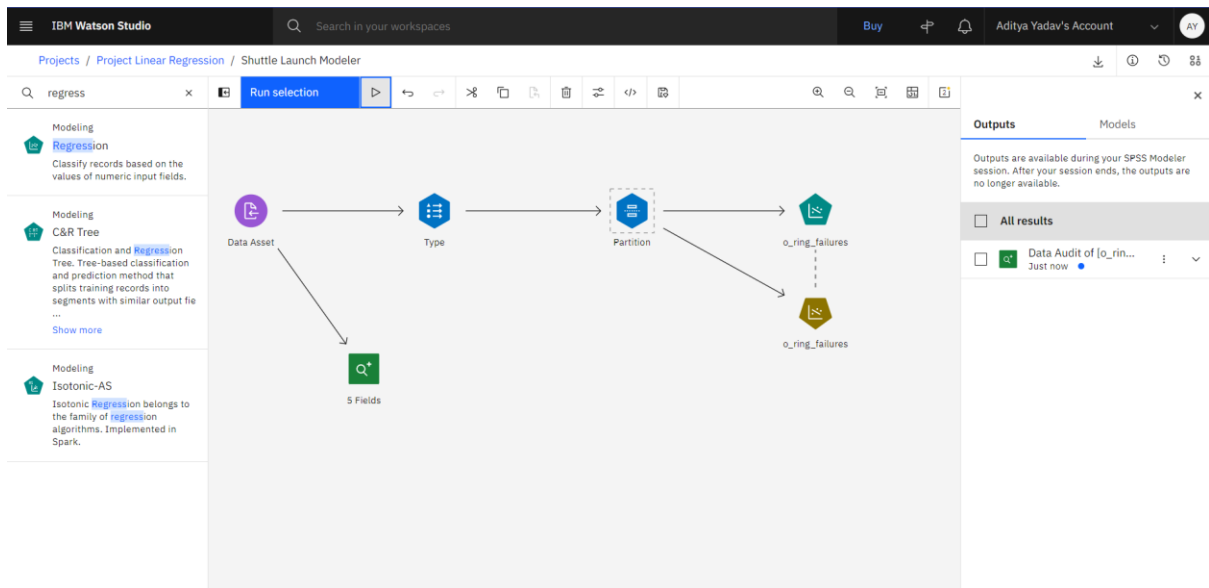


Fig: output after running the model after adding the regression model.

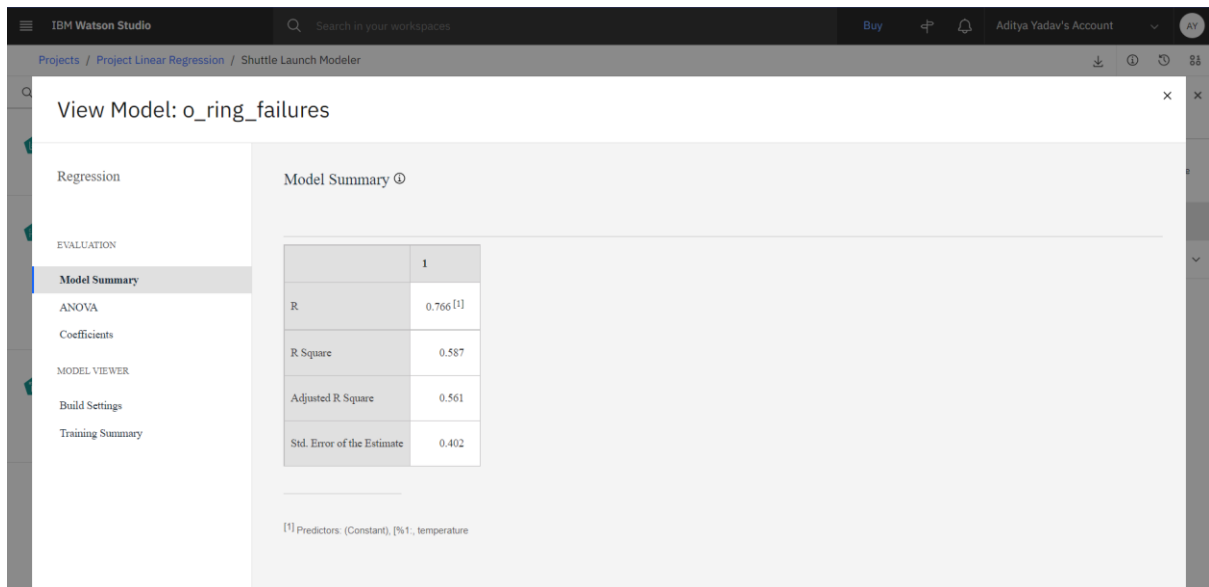


Fig: View model of the generated output.

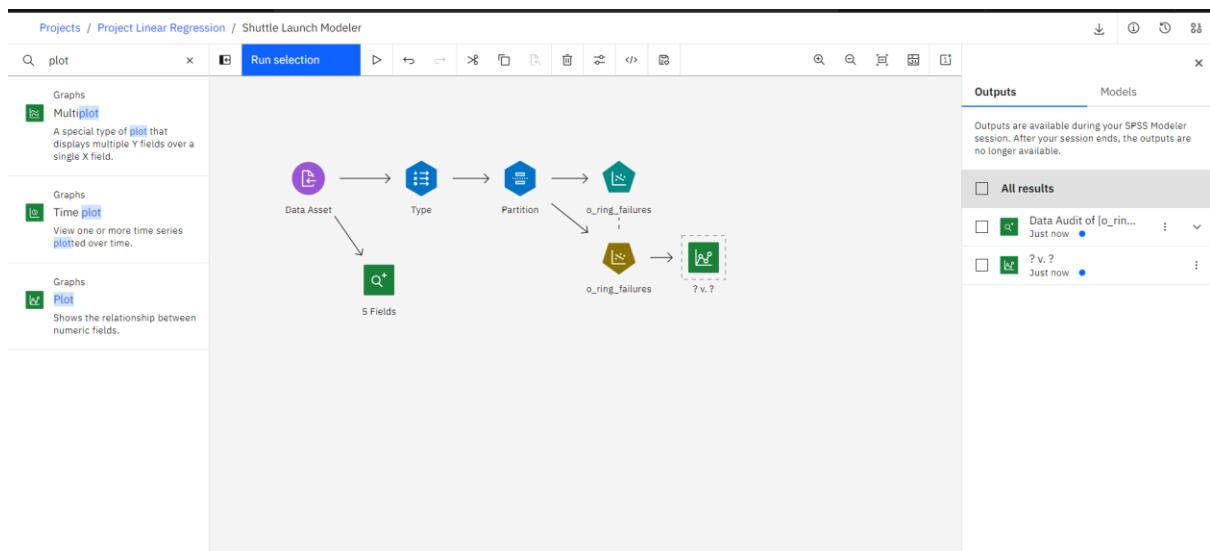


Fig: Add plot node to the model and connect to the regression model and give the parameter and run the model.

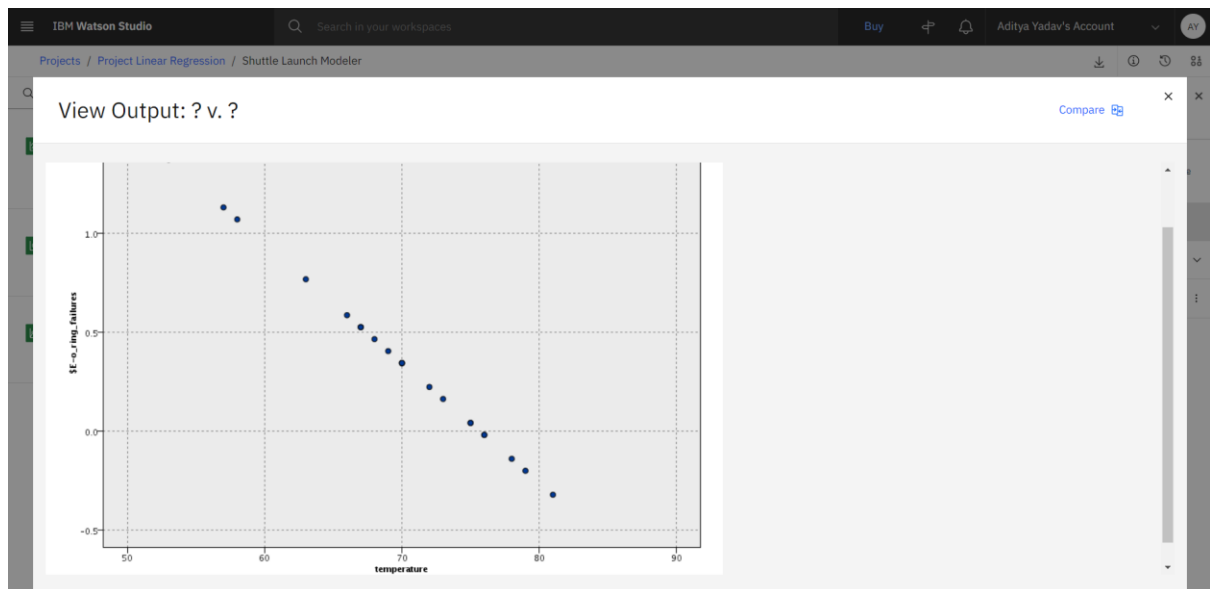


Fig: view the plotted graph

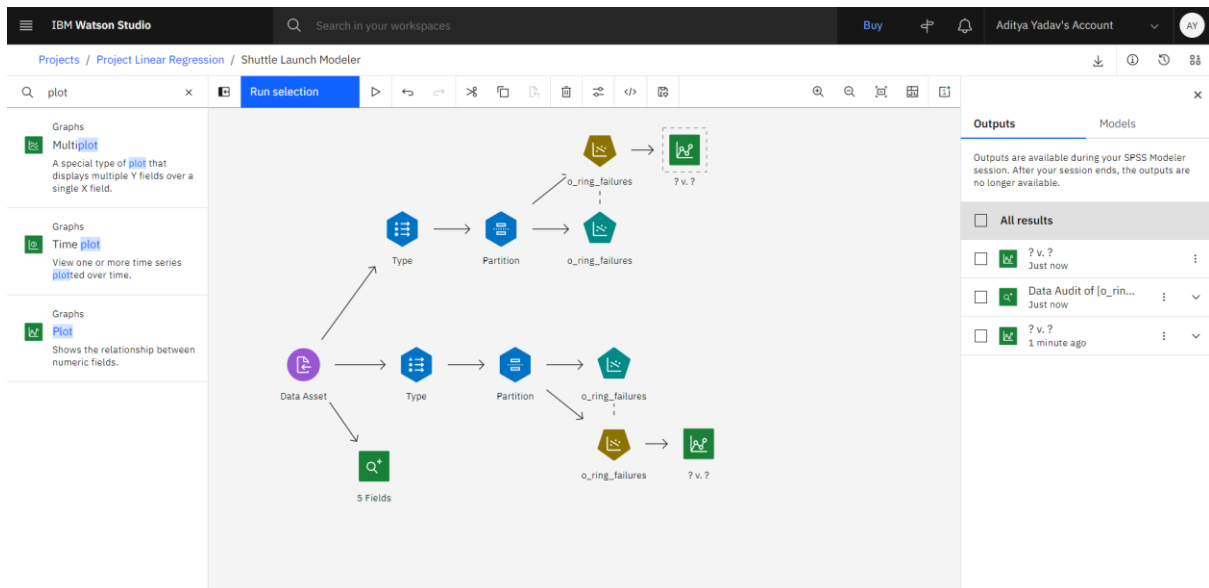


Fig: perform same steps by taking pressure in the above part of the model and plot the graph.

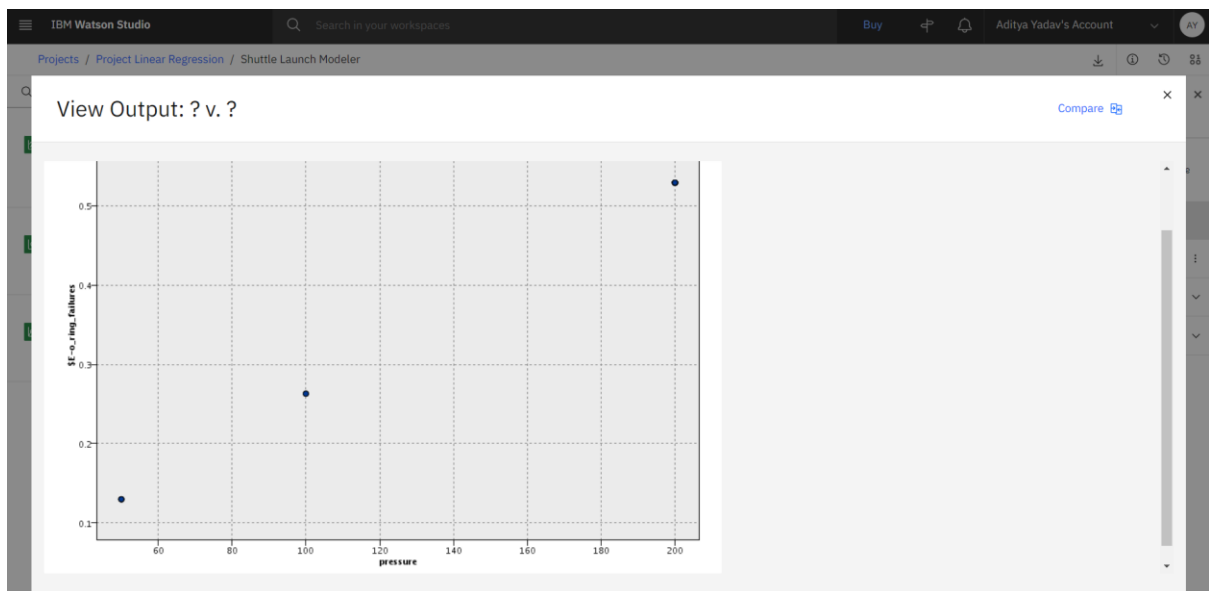


Fig: plot the graph for pressure vs o_ring_failure.

Dataset: bank.csv

Link to dashboard: <https://dataplatform.cloud.ibm.com/projects/469ce38c-0dfd-473f-952f-aaf46a4d8e69/assets?context=cpdaas>

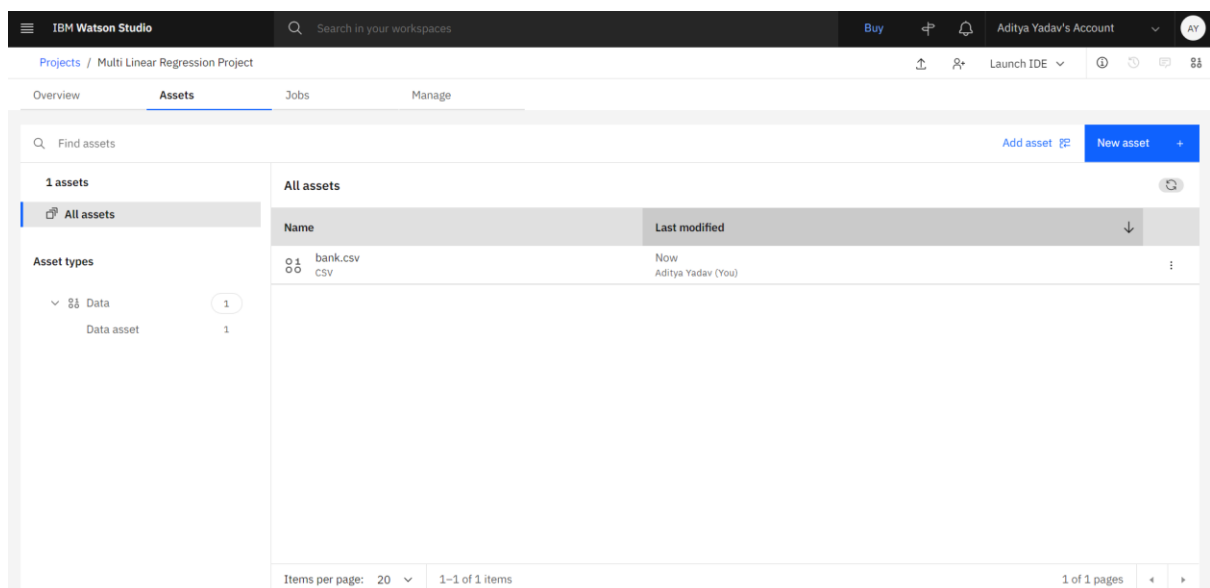


Fig: create new project and add dataset in the asset.

The screenshot shows the 'Preview' page for the 'bank.csv' dataset. The table displays 10 columns: age, job, marital, education, default, balance, housing, loan, contact, and day. The data is presented in a grid format with 10 rows. A sidebar on the right provides 'Information' about the dataset, including its name, description, tags, creator, usage, and size.

age	job	marital	education	default	balance	housing	loan	contact	day
59	admin.	married	secondary	no	2343	yes	no	unknown	5
56	admin.	married	secondary	no	45	no	no	unknown	5
41	technician	married	secondary	no	1270	yes	no	unknown	5
55	services	married	secondary	no	2476	yes	no	unknown	5
54	admin.	married	tertiary	no	184	no	no	unknown	5
42	management	single	tertiary	no	0	yes	yes	unknown	5
56	management	married	tertiary	no	830	yes	yes	unknown	6
60	retired	divorced	secondary	no	545	yes	no	unknown	6
37	technician	married	secondary	no	1	yes	no	unknown	6

Information sidebar details:

- Data asset
- bank.csv
- Description: No description available for this asset
- Tags: No tags available for this asset
- Creator: Aditya Yadav
- Usage: Created on Apr 25, 2022, 07:28 PM
- Size: 918.96 KB

Fig: descriptive view of the dataset.

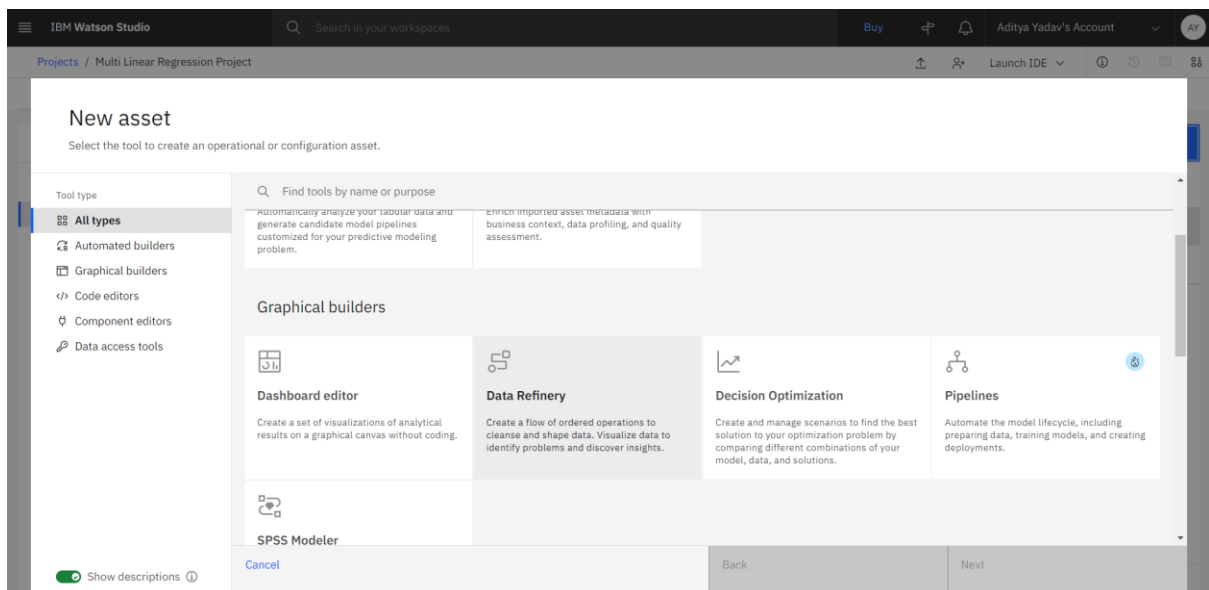


Fig: add new data asset for data refinery

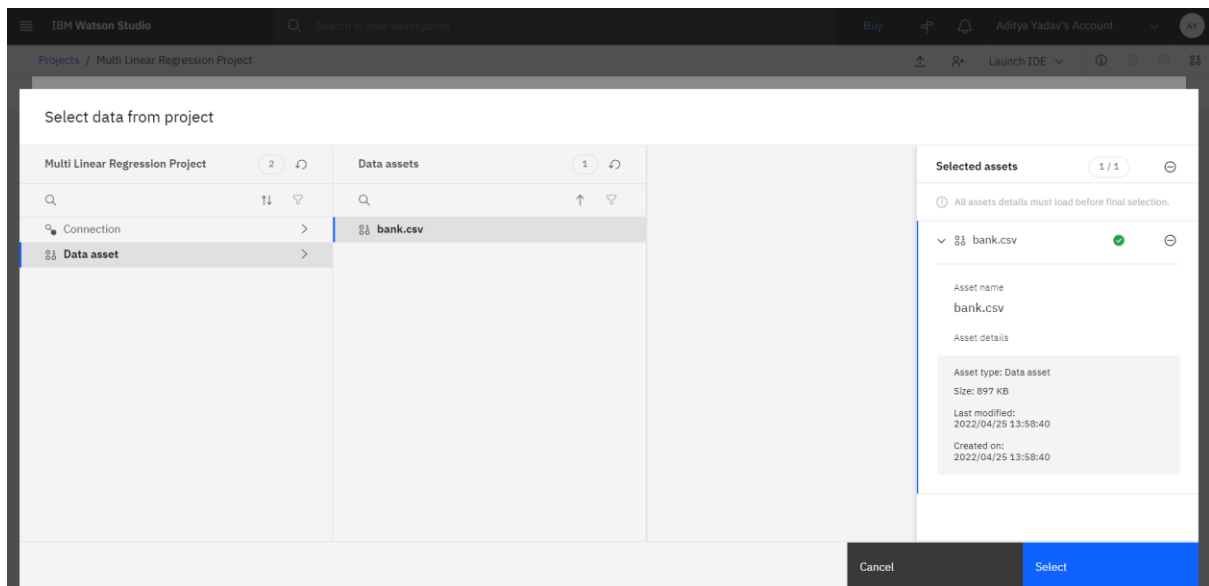


Fig: select dataset to be refined.

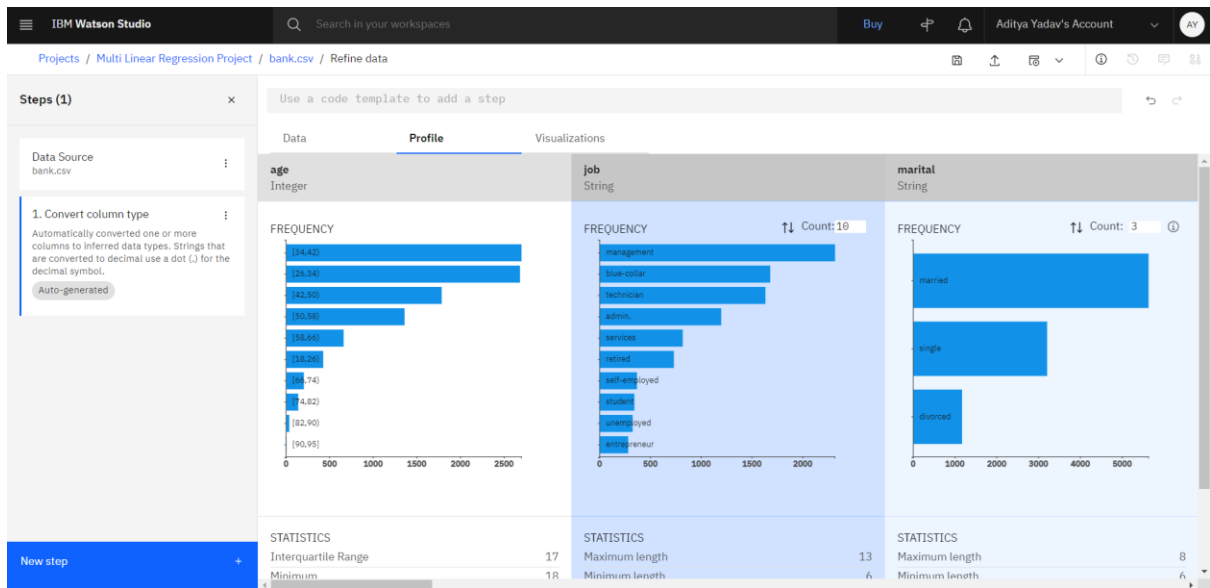


Fig: generate profile for our dataset

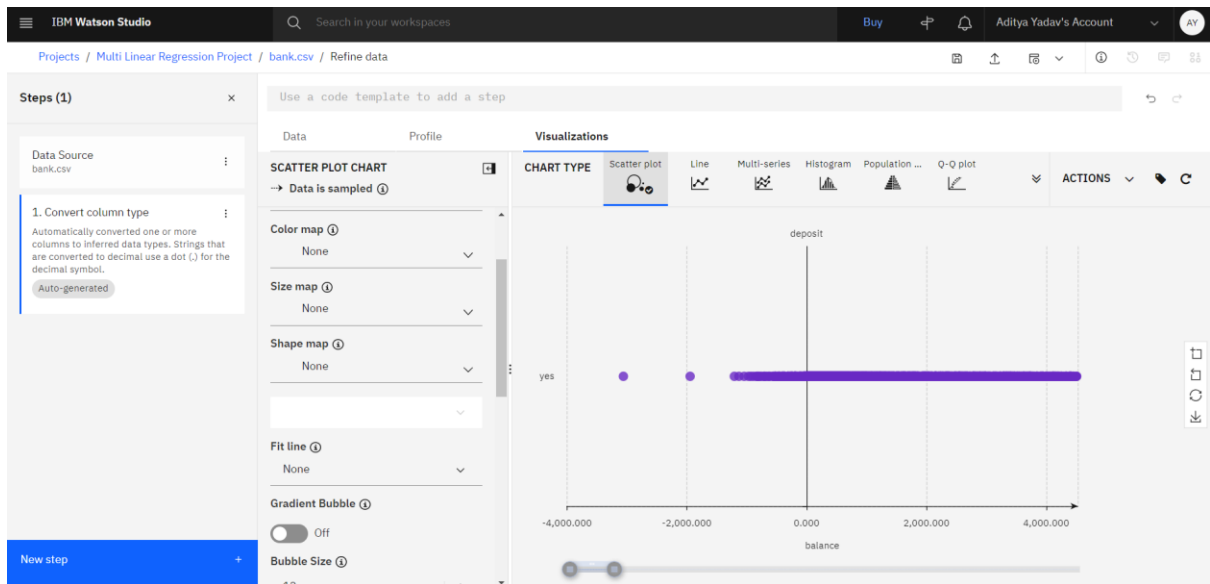


Fig: scatter plot your data.

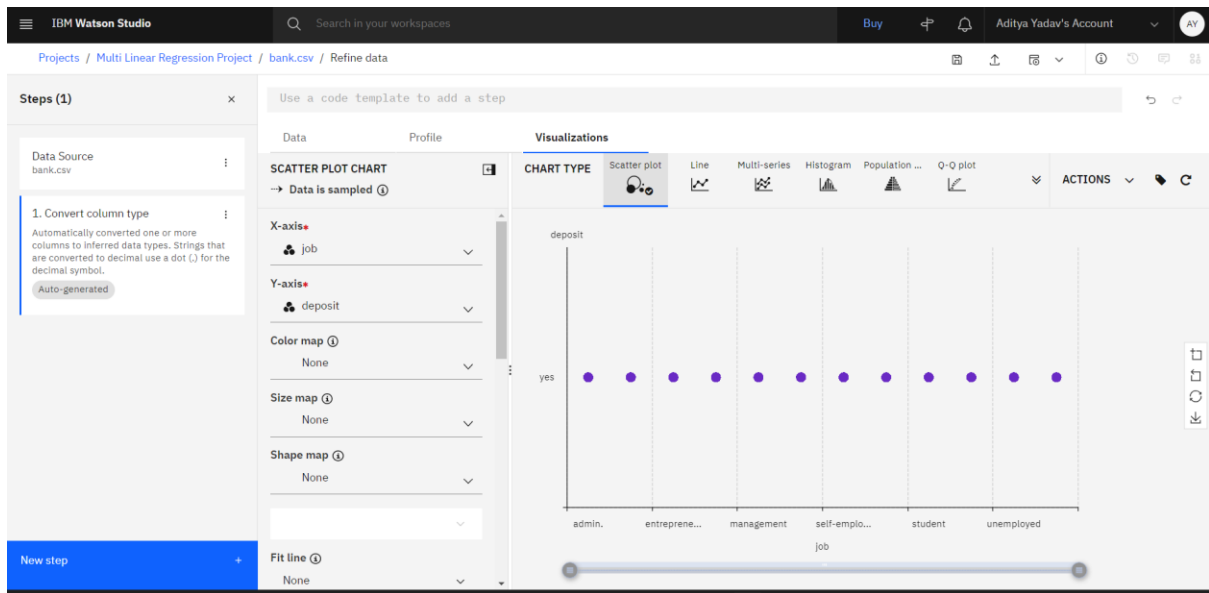


Fig: scatter-plot with job

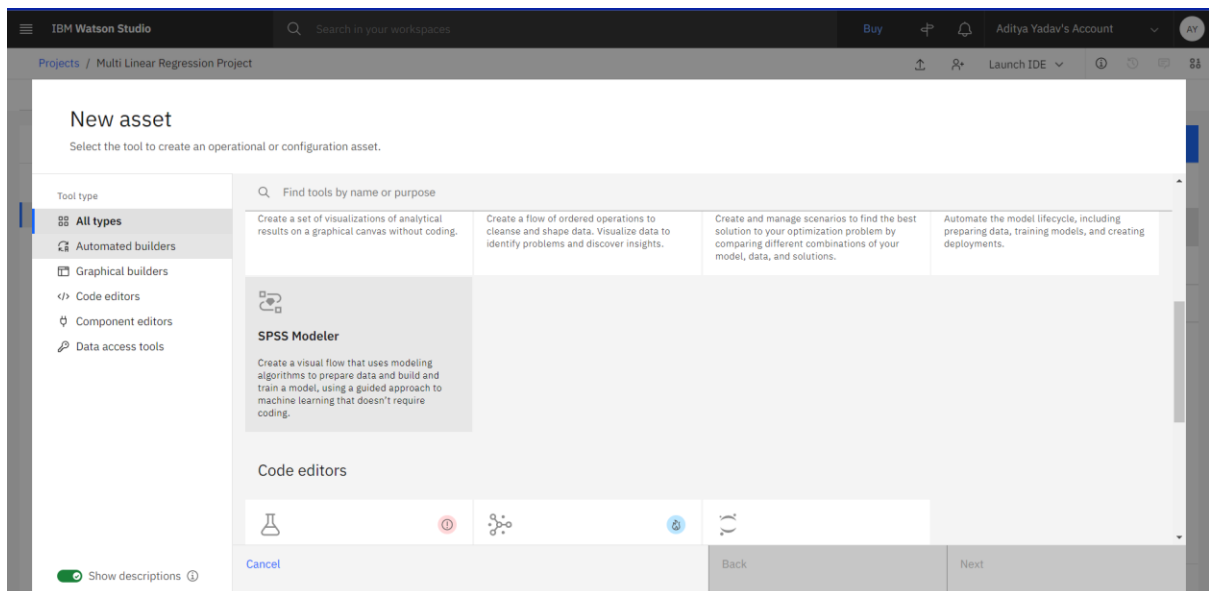


Fig: Create a SPSS modeler

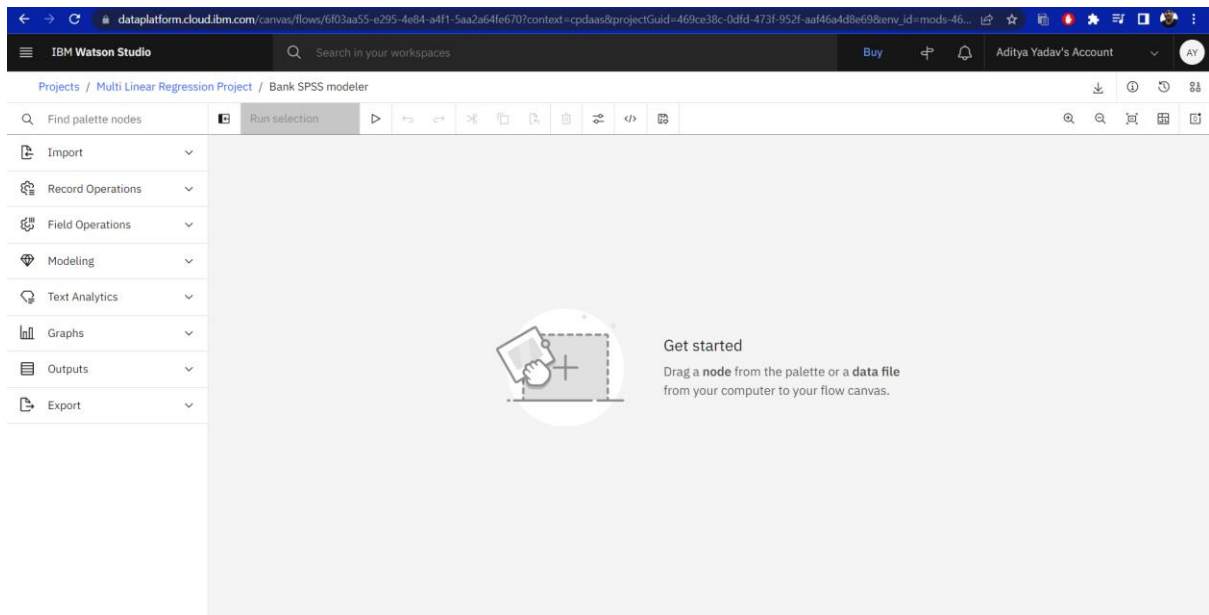


Fig: create a workspace

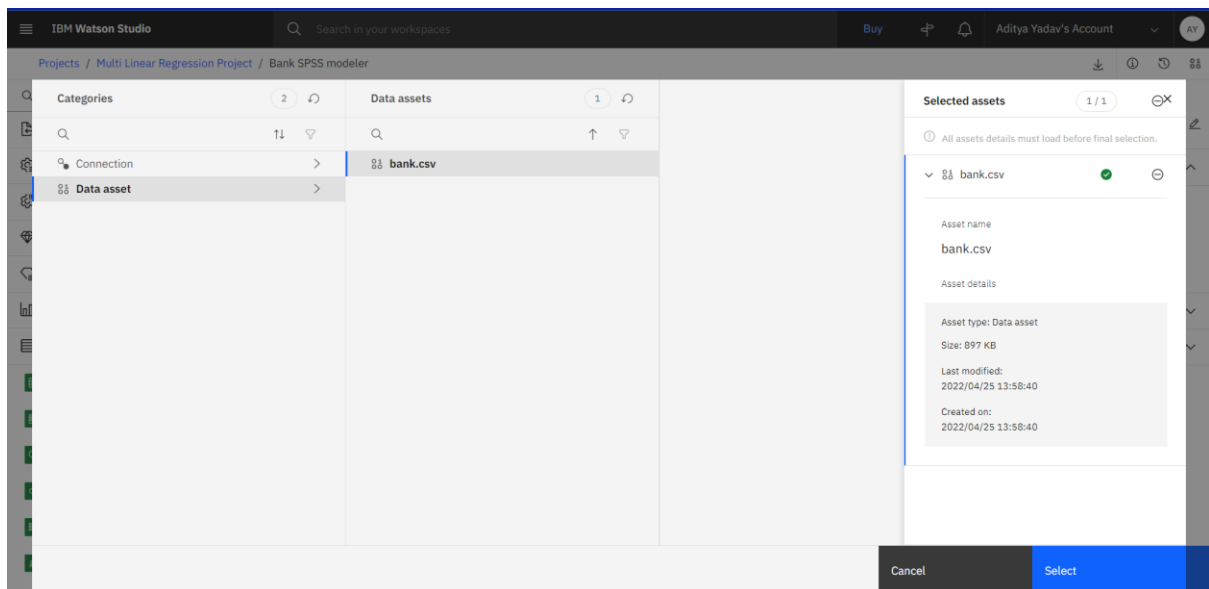


Fig: connect data set to the data asset.

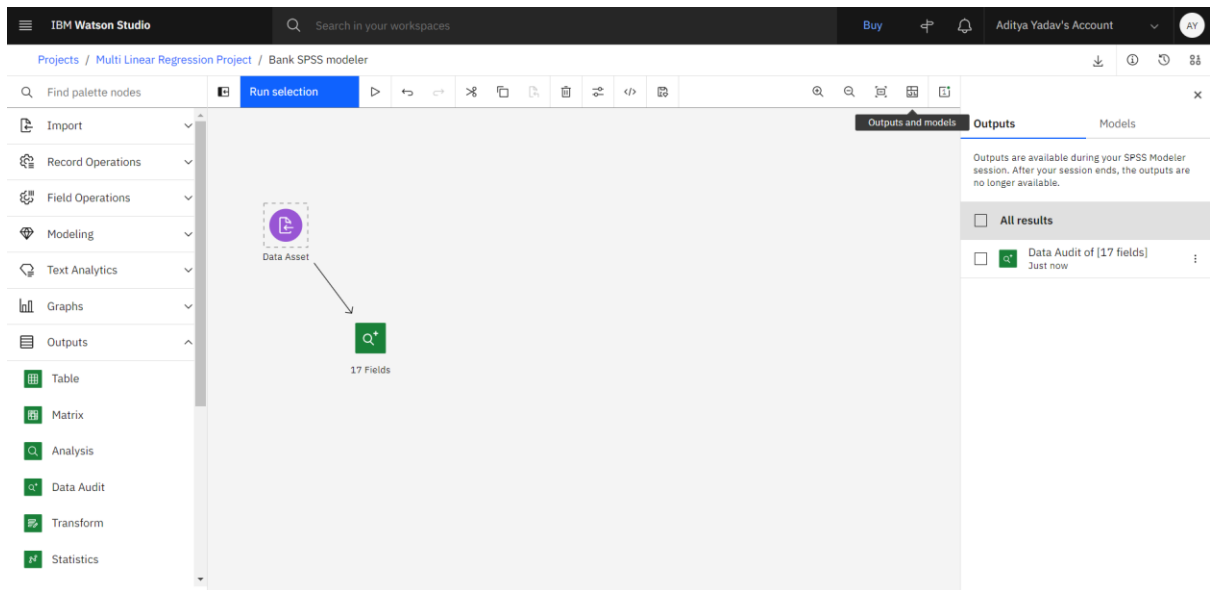


Fig: connect with data audit

View Output: Data Audit of [17 fields]

	Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	age		Continuous	18	95	41.232	11.913	0.863	--	11162
2	job		Categorical	--	--	--	--	--	12	11162
3	marital		Categorical	--	--	--	--	--	3	11162
4	education		Categorical	--	--	--	--	--	4	11162
5	default		Categorical	--	--	--	--	--	2	11162
6	balance		Continuous	-6847	81204	1528.539	3225.413	8.225	--	11162

Fig: Data audit output.

Field Operations

- Type**
Specify field metadata and properties that are invaluable to modeling.
- Filler**
Replace field values and change storage. Often used with a **Type** node to replace missing values.
- Multiplot**
A special **type** of plot that displays multiple Y fields over a single X field.
- Distribution**
Shows the occurrence of symbolic (non-numeric) values, such as mortgage **type** or gender, in a dataset.

Type

Settings

Read values Clear values

Find in column Field

Field	Measure	Role	Value mode	Values
# age	Continuous	None	Read	
abc job	Categorical	None	Read	
abc marital	Categorical	None	Read	
abc education	Categorical	None	Read	
abc default	Categorical	None	Read	
# balance	Continuous	Input	Read	
abc housing	Categorical	None	Read	
abc loan	Categorical	None	Read	

Default mode

☒ Read metadata ☐ Pass (do not scan)

Cancel Save

Fig: add and update type node

Field Operations

- Partition**
Generate a **partition** field that splits data into separate subsets or samples for the training, testing, and validation stages of model building.

Partition

Settings

Derived Field Name

Partition

Training Partition(%)

70

Testing Partition(%)

30

☐ Create validation partition

☒ Repeating partition assignment

Seed

1234567

☐ Use unique field to assign partitions

Cancel Save

Fig: add partition and update training and testing data.

IBM Watson Studio interface showing a regression modeling workflow. The workflow includes a Data Asset, Type, Partition, and a Regression model node. The Regression model node is highlighted with a dashed box and labeled "previous". The left sidebar shows modeling options like Regression, C&R Tree, and Isotonic-AS. The right sidebar shows the Outputs section with a message "Success: Run was successful" and a list of outputs including "Data Audit of [17 fields]".

Fig: perform regression modelling.

View Output: Data Audit of [17 fields] #1

Index	Name	Distribution	Type	Min	Q1	Median	Q3	Max	Count
1	age		Continuous	18	95	41.232	11.913	0.863	--
2	job		Categorical	--	--	--	--	--	12
3	marital		Categorical	--	--	--	--	--	3
4	education		Categorical	--	--	--	--	--	4
5	default		Categorical	--	--	--	--	--	2
6	balance		Continuous	-6847	81204	1528.539	3225.413	8.225	--
7	housing		Categorical	--	--	--	--	--	2

Fig: get the output of regression model.

IBM Watson Studio

Search in your workspaces

Buy

Aditya Yadav's Account

Projects / Multi Linear Regression Project / Bank SPSS modeler

plot

Run selection

Graphs

Multiplot

A special type of **plot** that displays multiple Y fields over a single X field.

Graphs

Time plot

View one or more time series **plotted** over time.

Graphs

Plot

Shows the relationship between numeric fields.

Data Asset

Type

Partition

previous

17 Fields

previous

? v. ?

Outputs

Models

Outputs are available during your SPSS Modeler session. After your session ends, the outputs are no longer available.

☐ All results

☐ Data Audit of [17 fie... Just now

☐ ? v. ? Just now

Fig: Add plot node

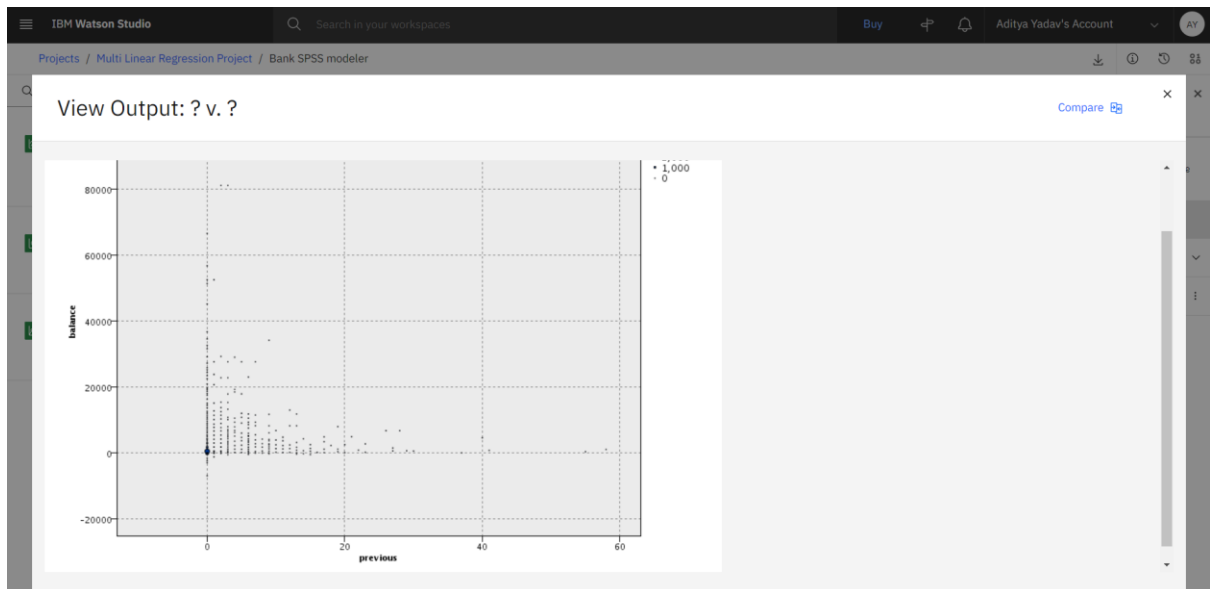


Fig: view the graph