Name: Aditya Yadav

Data Analytics

VIT Vellore, Tamil Nadu

Assignment-3

Dataset: diabetes.csv

Link to dashboard: https://dataplatform.cloud.ibm.com/projects/238bd258-17ba-45ba-9897-7c7c3543e496/assets?context=cpdaas
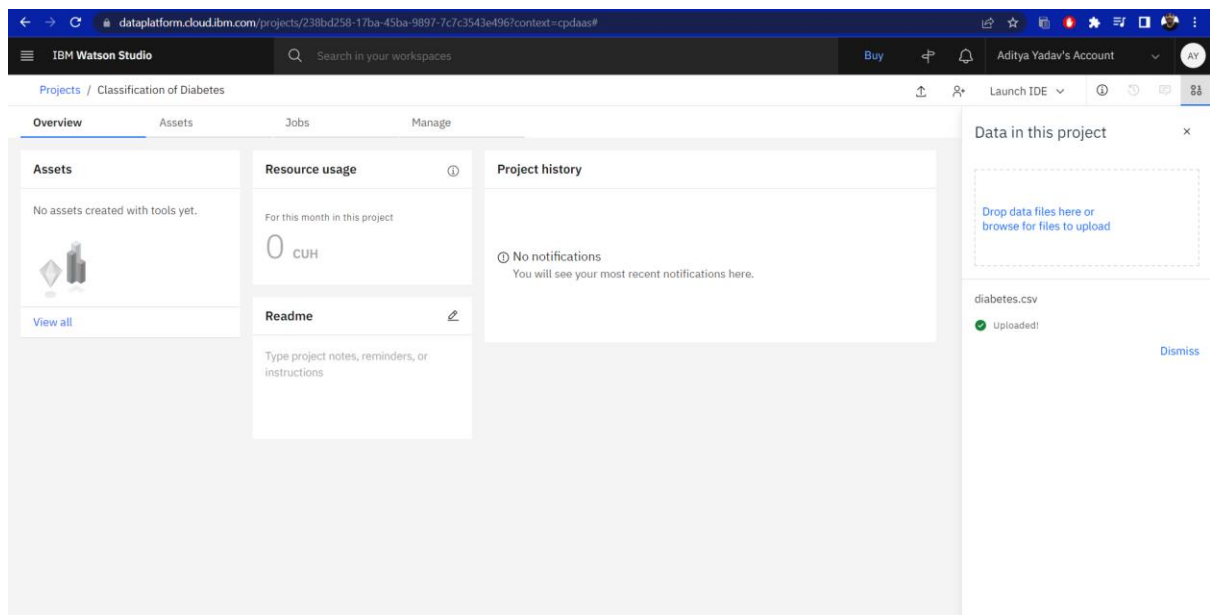

Fig: Create a new Project
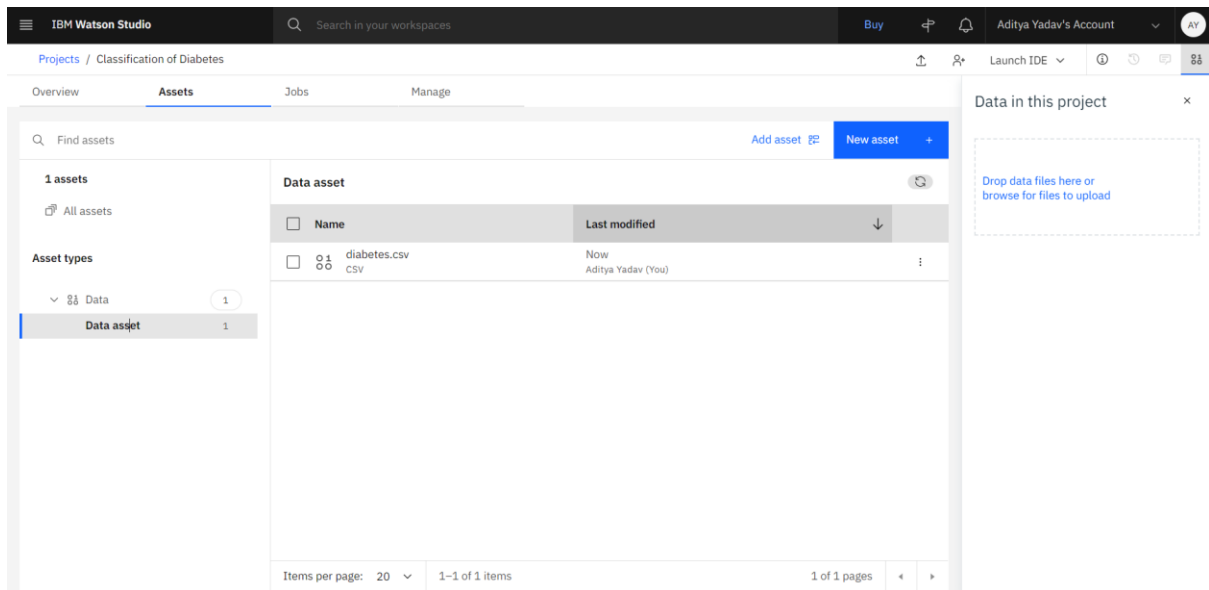

Fig: upload the dataset to the project
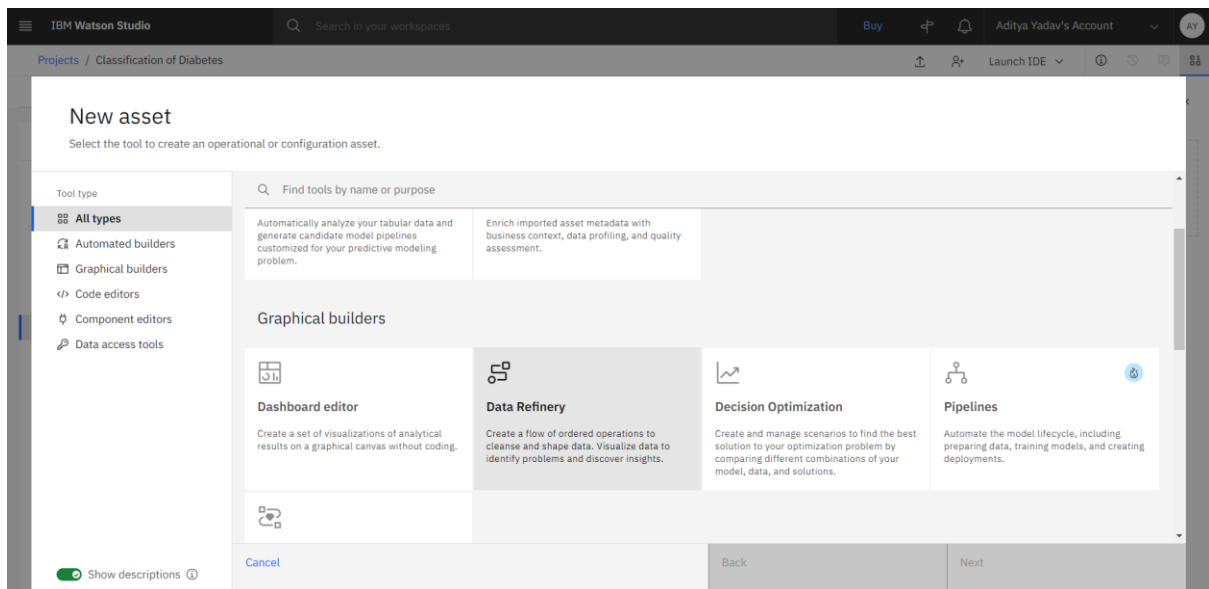
Fig: data asset available
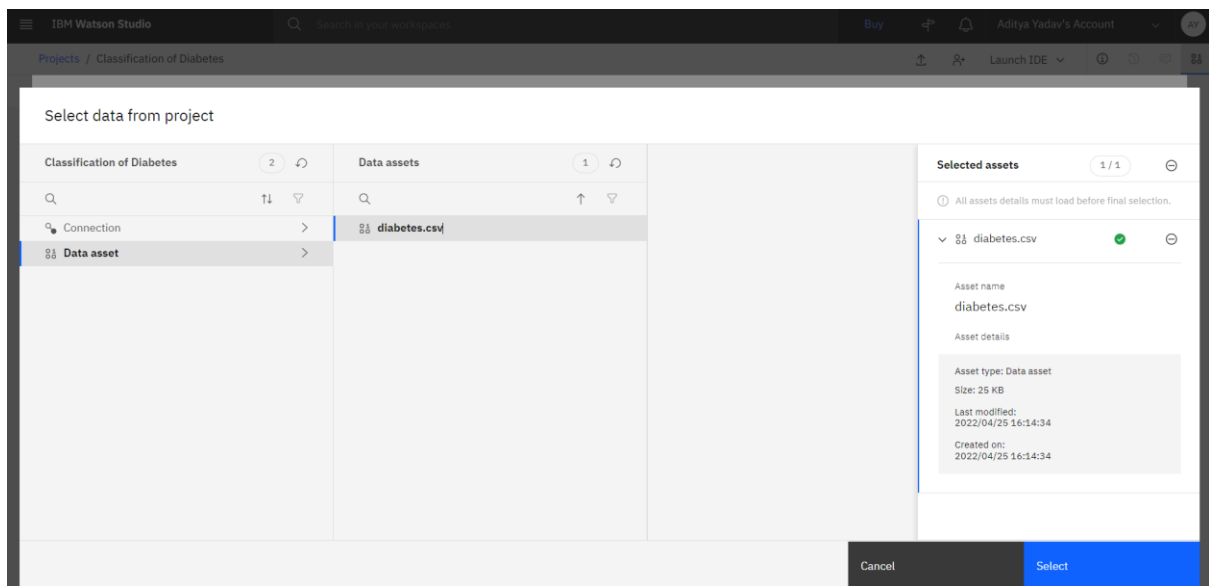

Fig: create data refinery
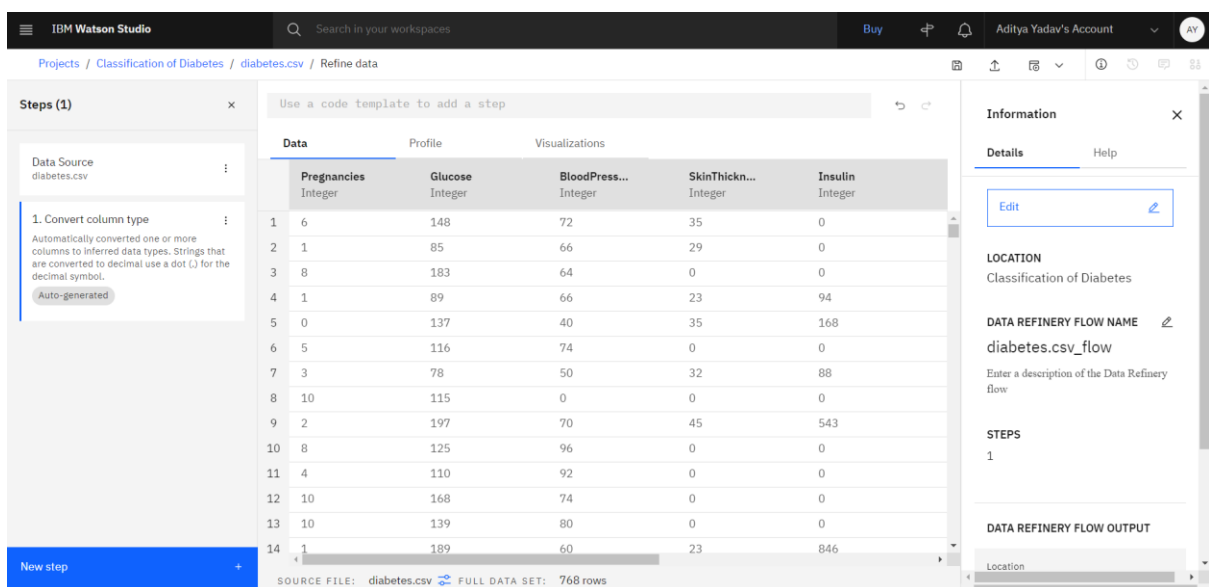
Fig: add the dataset for data refining

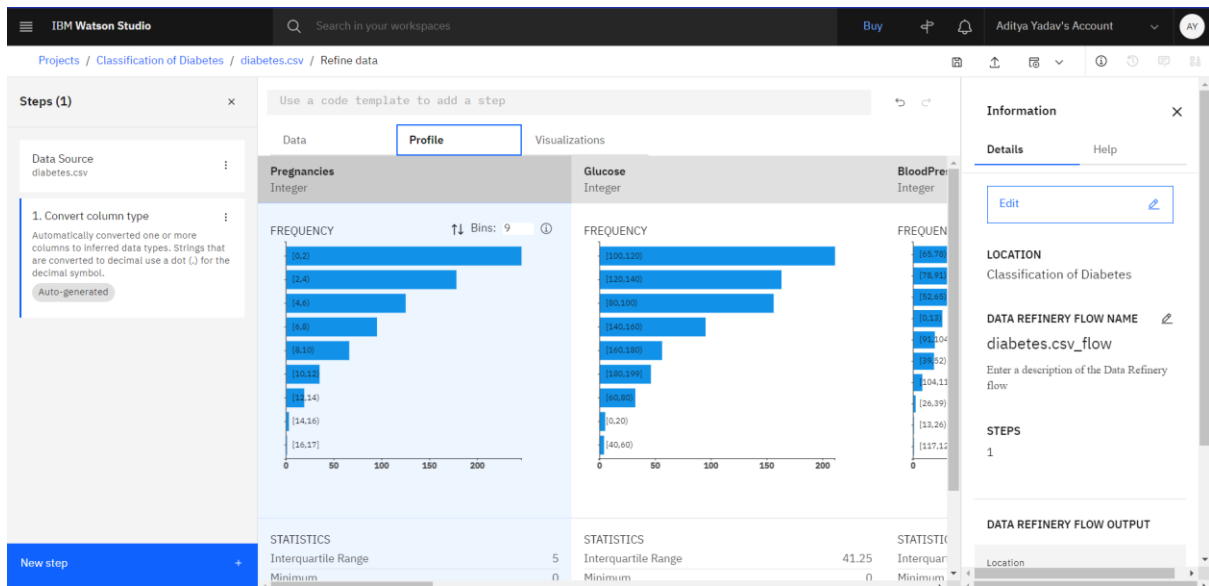

Fig: descriptive details of the dataset

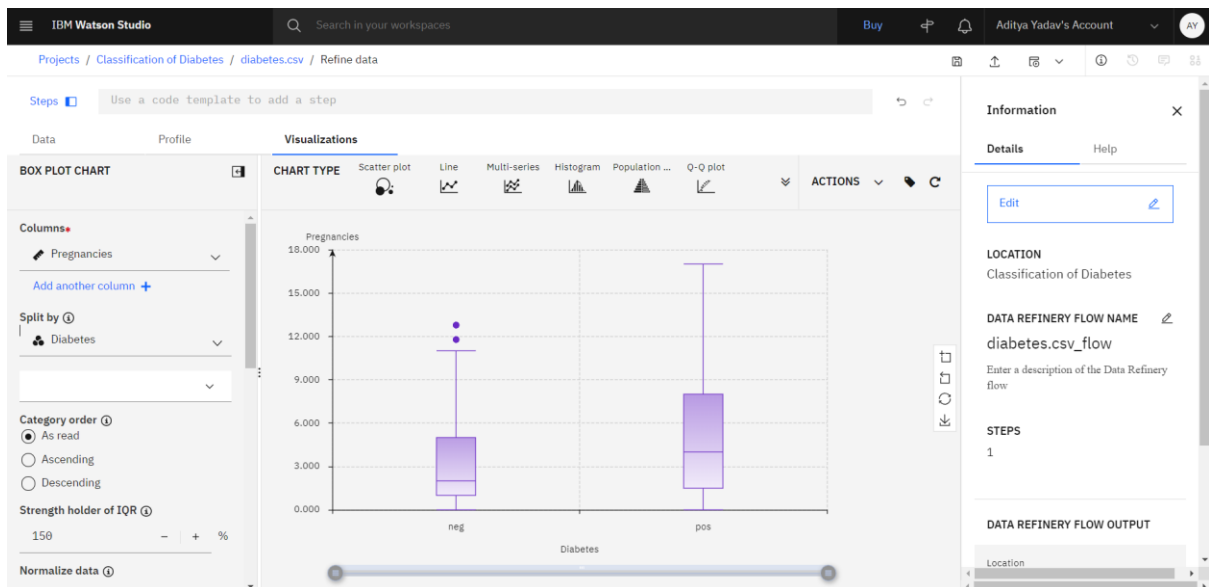Fig: profile of our data

Visualize on the basis of box plot chart on pregnancies and diabetes.
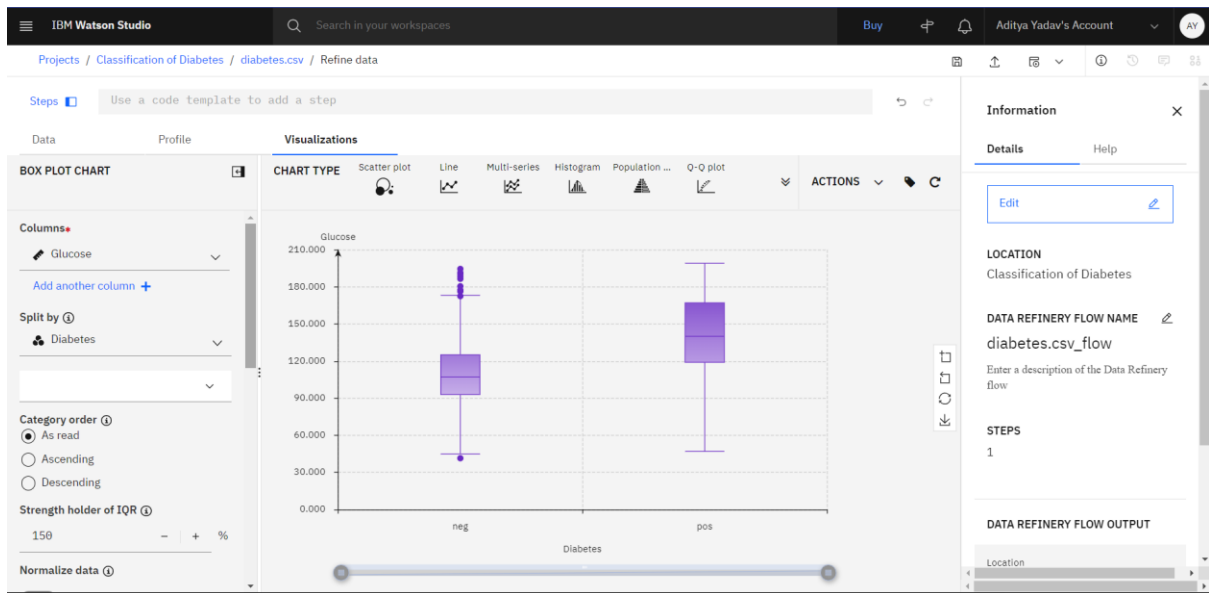
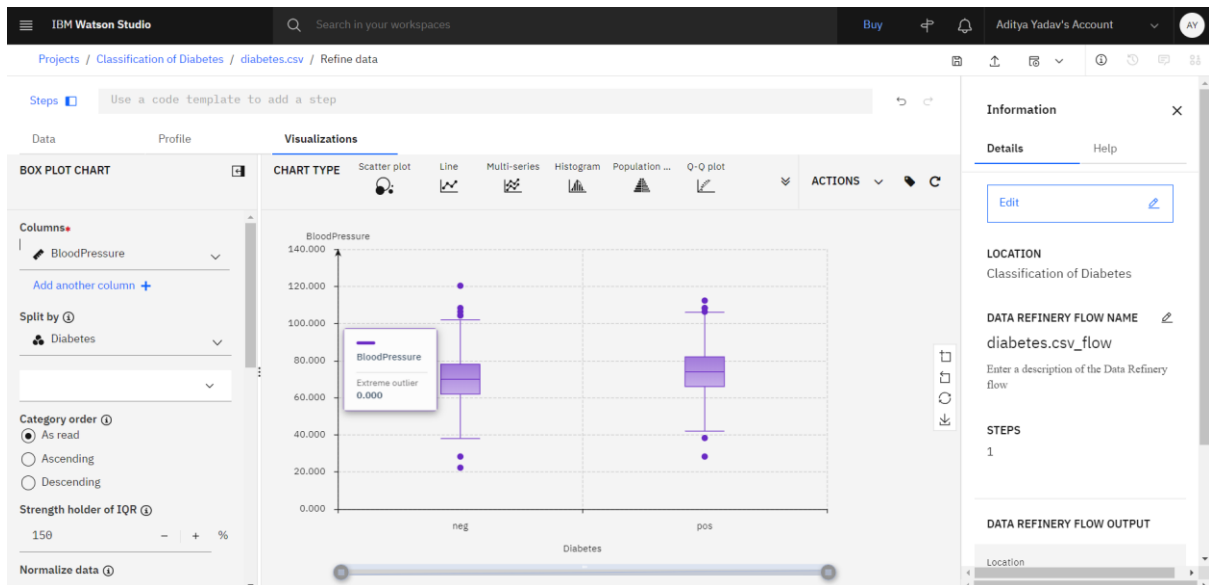Fig: Visualize on the basis of box plot chart on glucose and diabetes.

Fig: graph for BloodPressure and diabetes
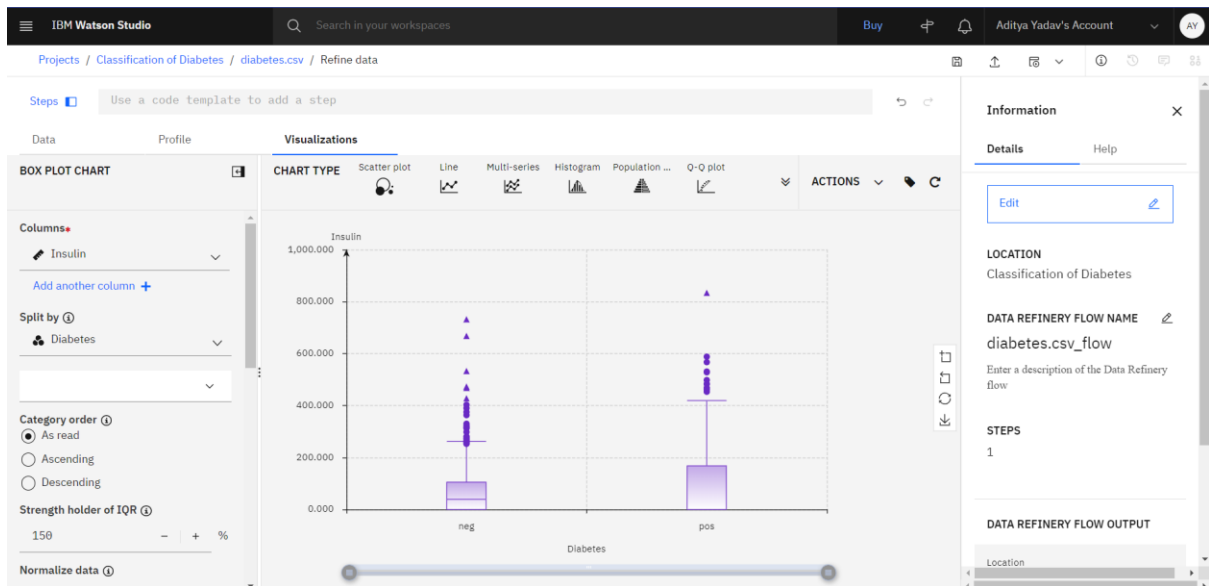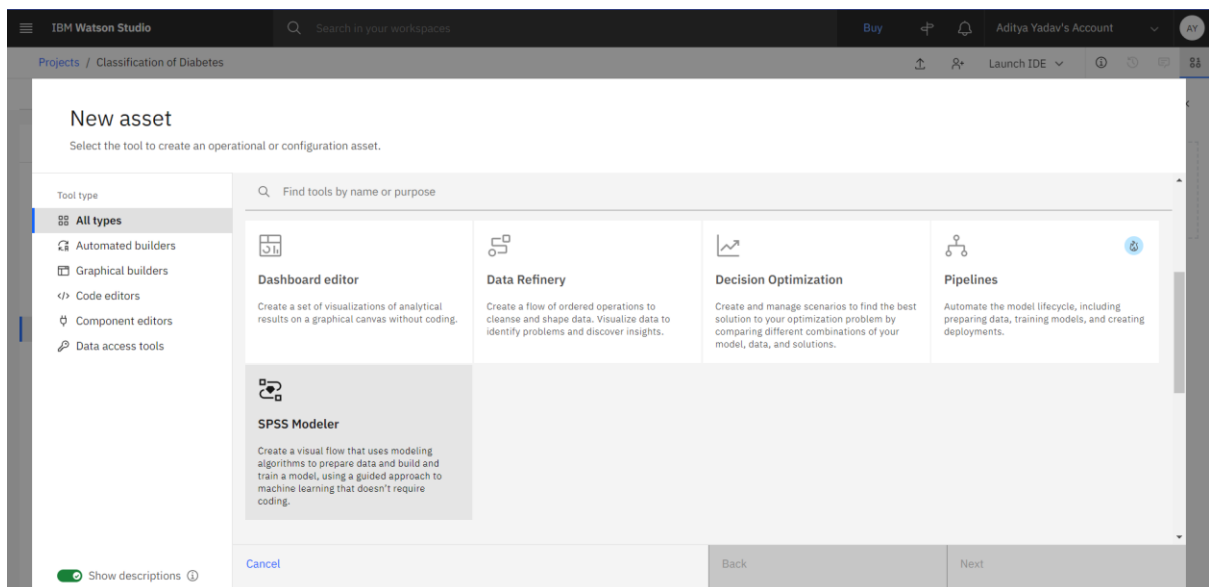
Fig: graph for Insulin and diabetes
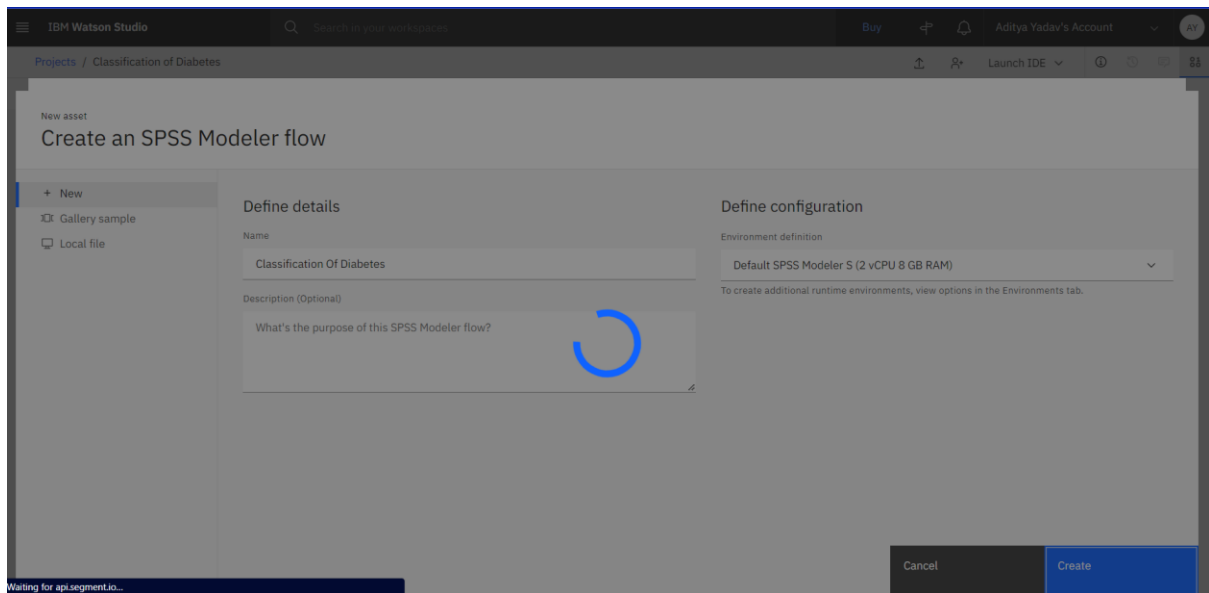

Fig: create a SPSS modeler

Fig: creating workbench
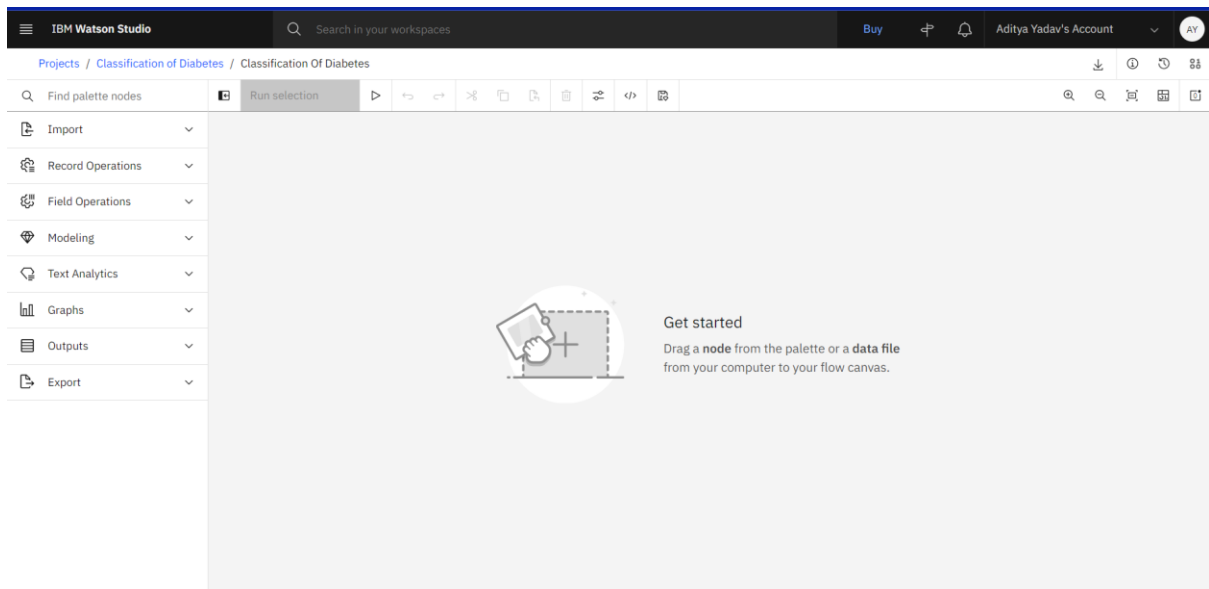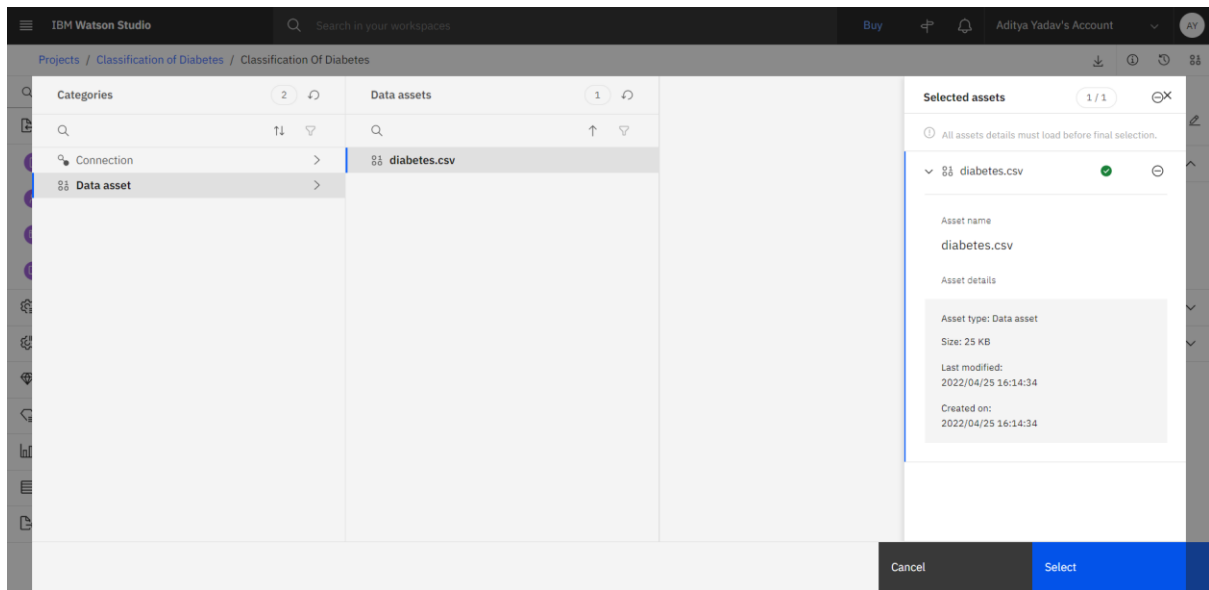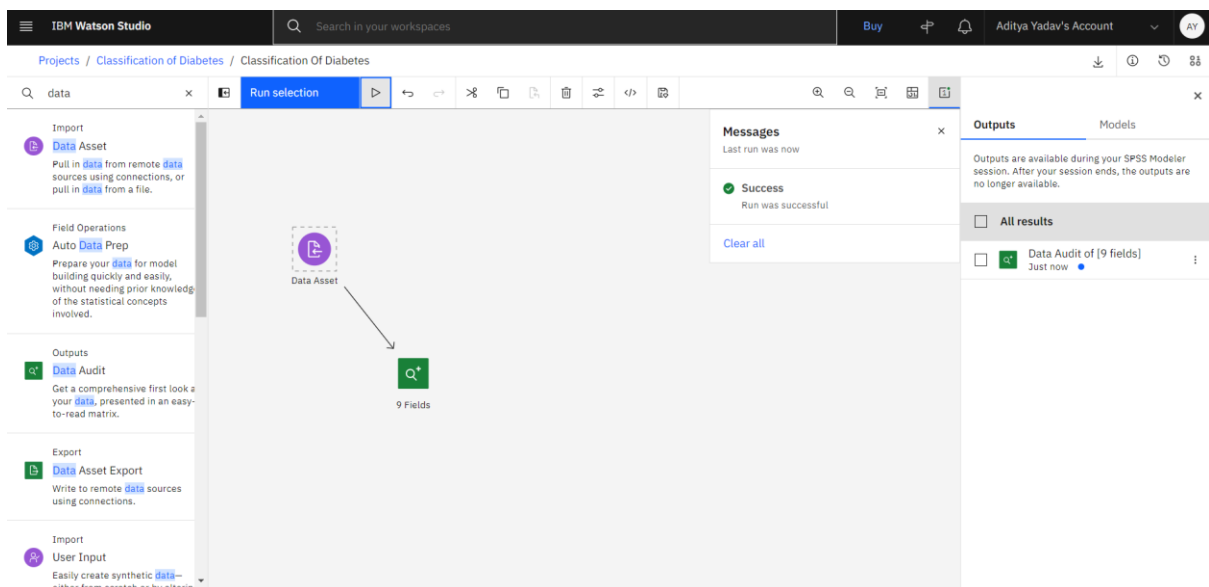

Fig: workbench

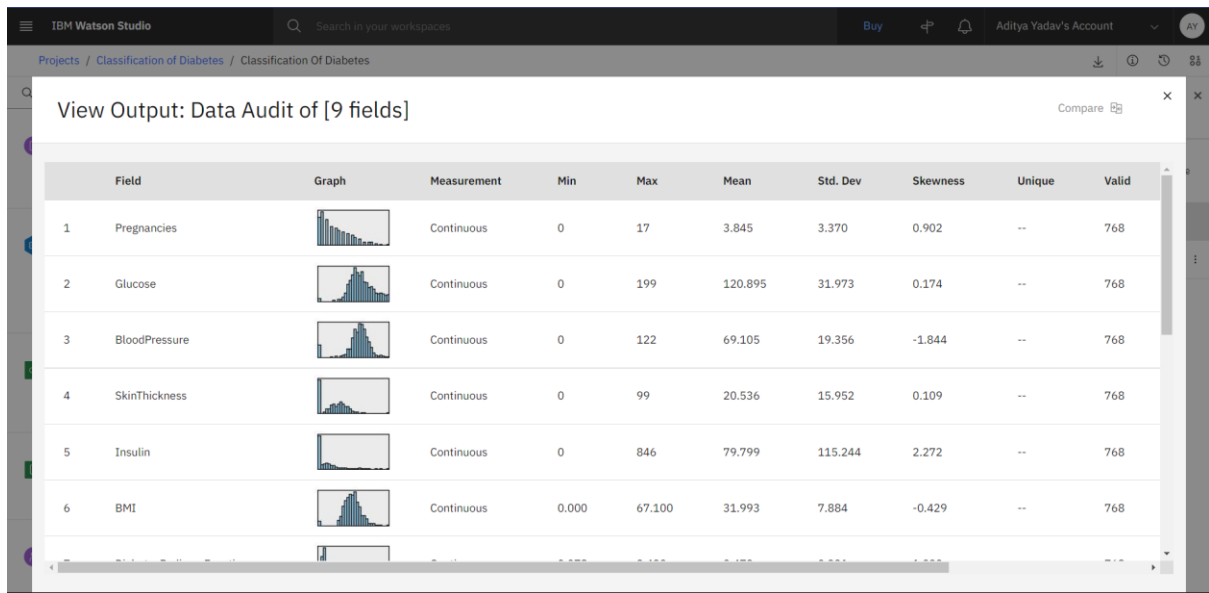add dataset node to the workplace

Add data audit node and run the model.

Fig: output of audit

Fig: drag type node and update with the target and input data.

	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty S
1	Pregnancies	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
2	Glucose	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
3	BloodPressure	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
4	SkinThickness	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
5	Insulin	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
6	BMI	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
7	DiabetesPedigreeFunction	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
8	Age	Continuous	0	768	None	Never	Fixed	100.000	768	0	0
9	Diabetes	Categorical	--	--	--	Never	Fixed	100.000	768	0	0

Updated outliers

Add partition node and give training and test as 90% and 10%

Select auto classifier and run the model

Run the auto classifier model

examine the model by view model
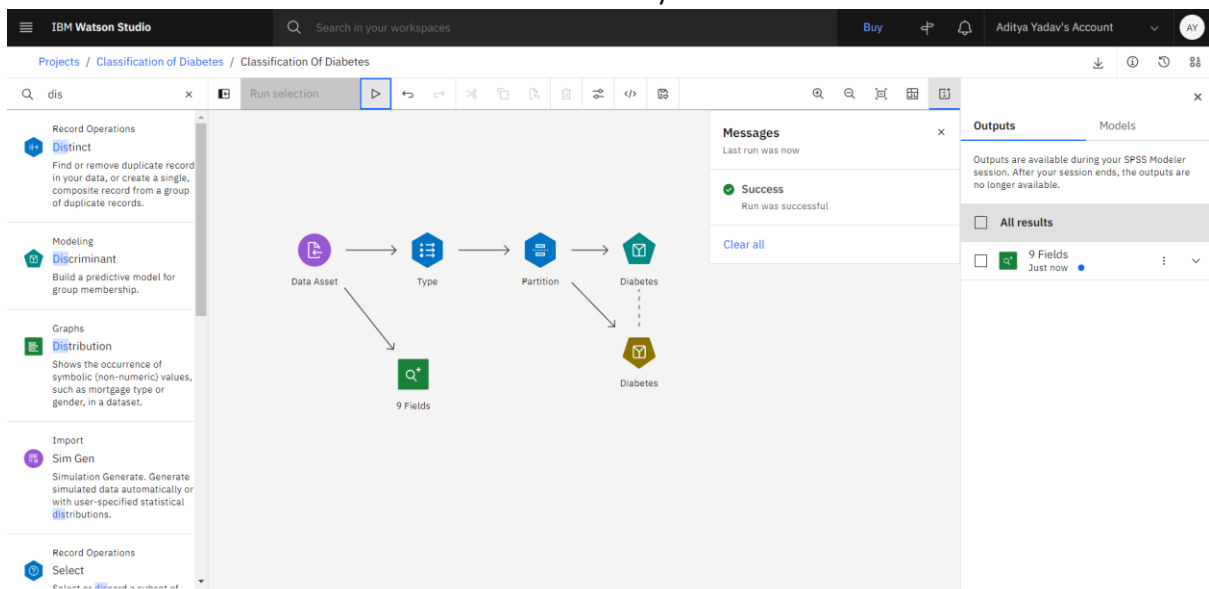
Perform again for discriminant model
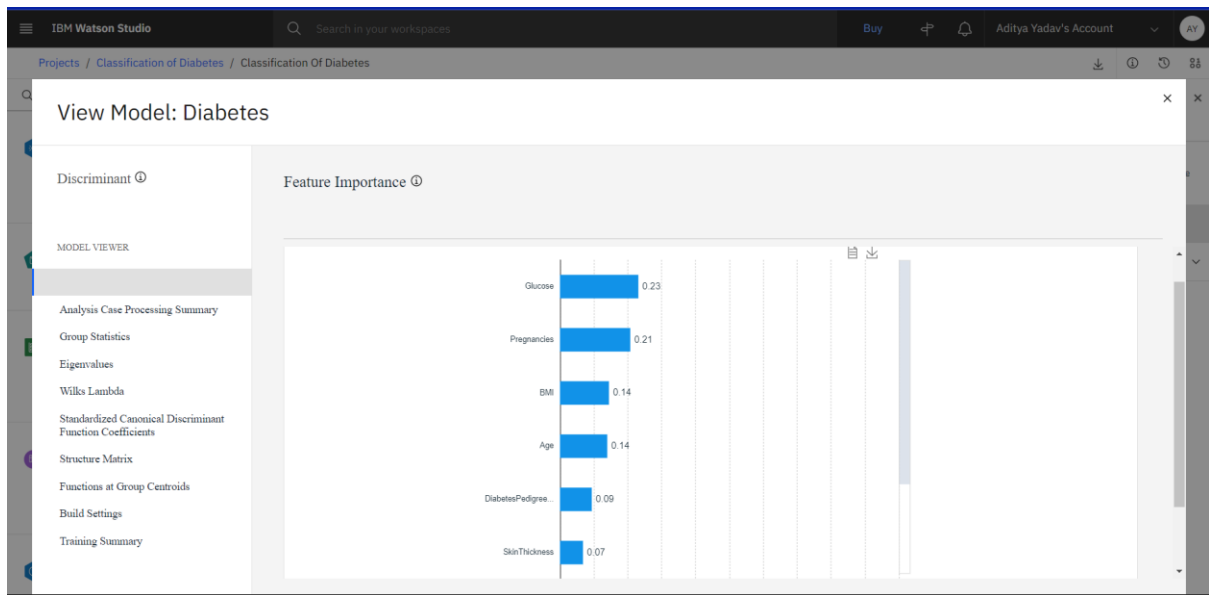
View model

Create output in form of table and analysis form

View Output: Table (12 fields, 768 records)

Compare
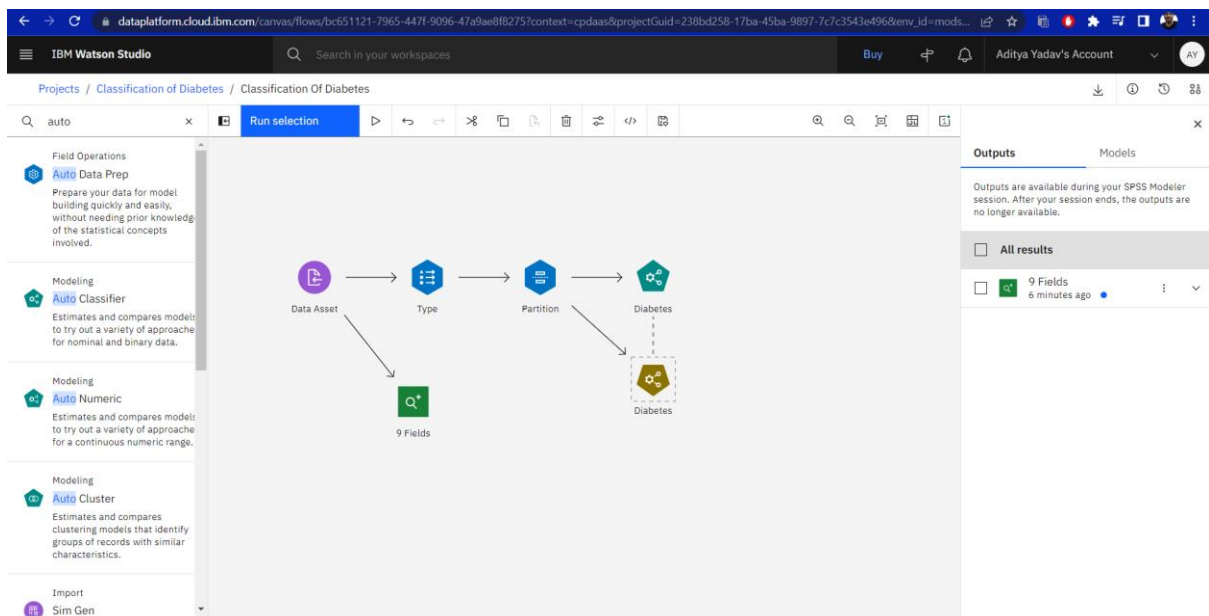
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Diabetes	Partition	$D-Diabetes	$DP-Diabetes
1	6	148	72	35	0	33.600	0.627	50	pos	1_Training	pos	0.842
2	1	85	66	29	0	26.600	0.351	31	neg	1_Training	neg	0.932
3	8	183	64	0	0	23.300	0.672	32	pos	1_Training	pos	0.908
4	1	89	66	23	94	28.100	0.167	21	neg	2_Testing	neg	0.943
5	0	137	40	35	168	43.100	2.288	33	pos	1_Training	pos	0.938
6	5	116	74	0	0	25.600	0.201	30	neg	1_Training	neg	0.777
7	3	78	50	32	88	31.000	0.248	26	pos	1_Training	neg	0.908
8	10	115	0	0	0	35.300	0.134	29	neg	1_Training	neg	0.775
9	2	197	70	45	543	30.500	0.158	53	pos	1_Training	pos	0.873
10	8	125	96	0	0	0.000	0.232	54	pos	1_Training	neg	0.911
11	4	110	92	0	0	37.600	0.191	30	neg	1_Training	neg	0.726
12	10	168	74	0	0	38.000	0.537	34	pos	1_Training	pos	0.947
13	10	139	80	0	0	27.100	1.441	57	neg	2_Testing	pos	0.889
14	1	189	60	23	846	30.100	0.398	59	pos	1_Training	pos	0.830
15	5	166	72	19	175	25.800	0.587	51	pos	1_Training	pos	0.803

Table

View Output: Analysis of [Diabetes]

Compare

Collapse All

⊖ **Results for output field Diabetes**

⊖ **Comparing $D-Diabetes with Diabetes**

'Partition'	1_Training		2_Testing	
Correct	528	76.97%	60	73.17%
Wrong	158	23.03%	22	26.83%
Total	686		82	

Analysis

Use custom fields

After custom input

Projects / Classification of Diabetes / Classification Of Diabetes

View Output: Analysis of [Diabetes] #1

Compare

Collapse All

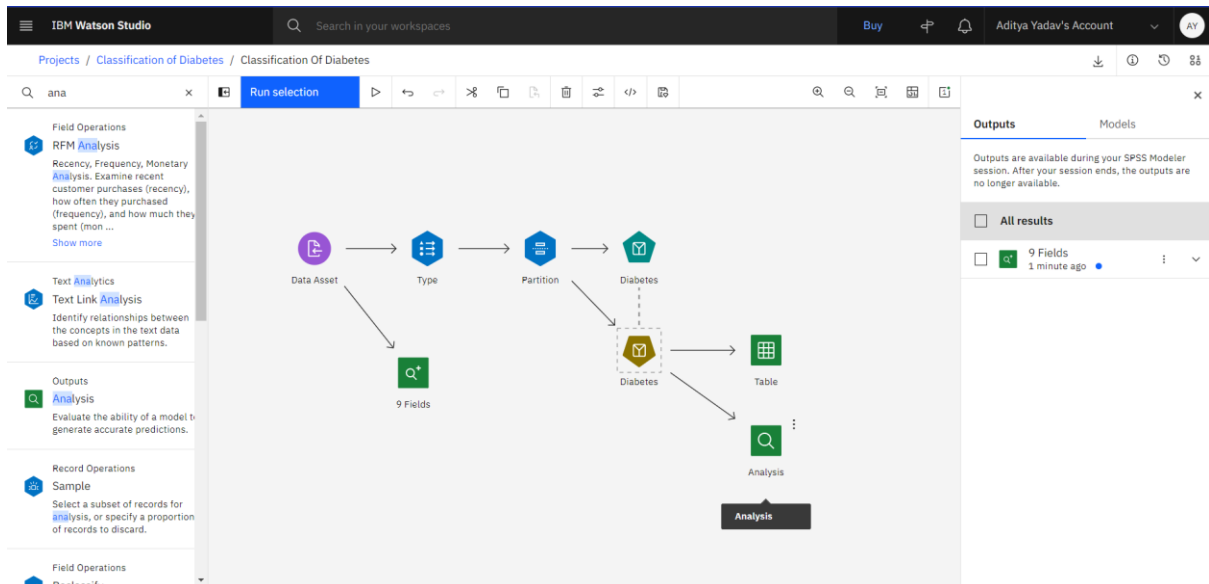⊖ **Results for output field Diabetes**

⊖ **Comparing $D-Diabetes with Diabetes**

'Partition'	1_Training		2_Testing	
Correct	515	75.07%	59	71.95%
Wrong	171	24.93%	23	28.05%
Total	686		82	

After custom input

Dataset: Insurance.csv

Link to Dashboard: https://dataplatform.cloud.ibm.com/projects/04d013ce-26a3-40b3-be32-306ac7004ed6/assets?context=cpdaas
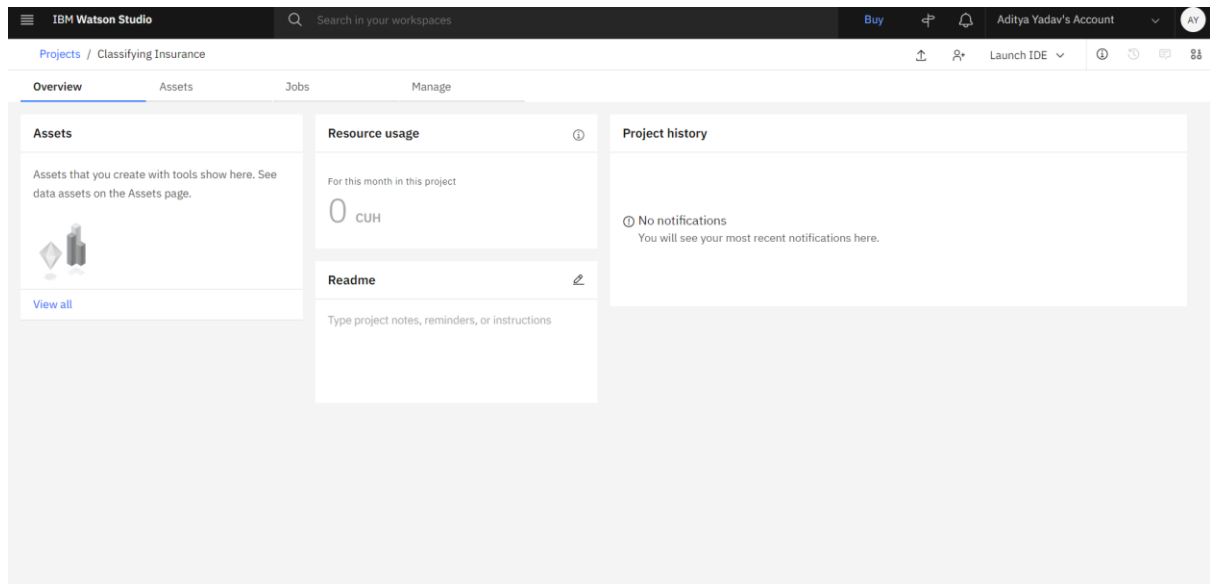
Create a new project
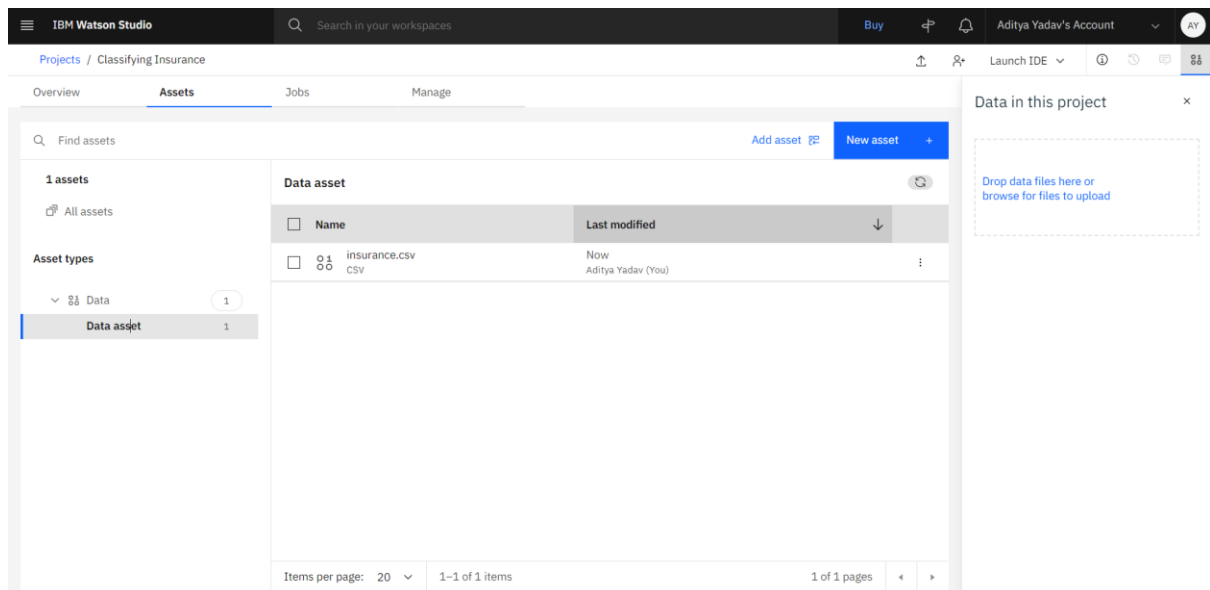
Upload dataset
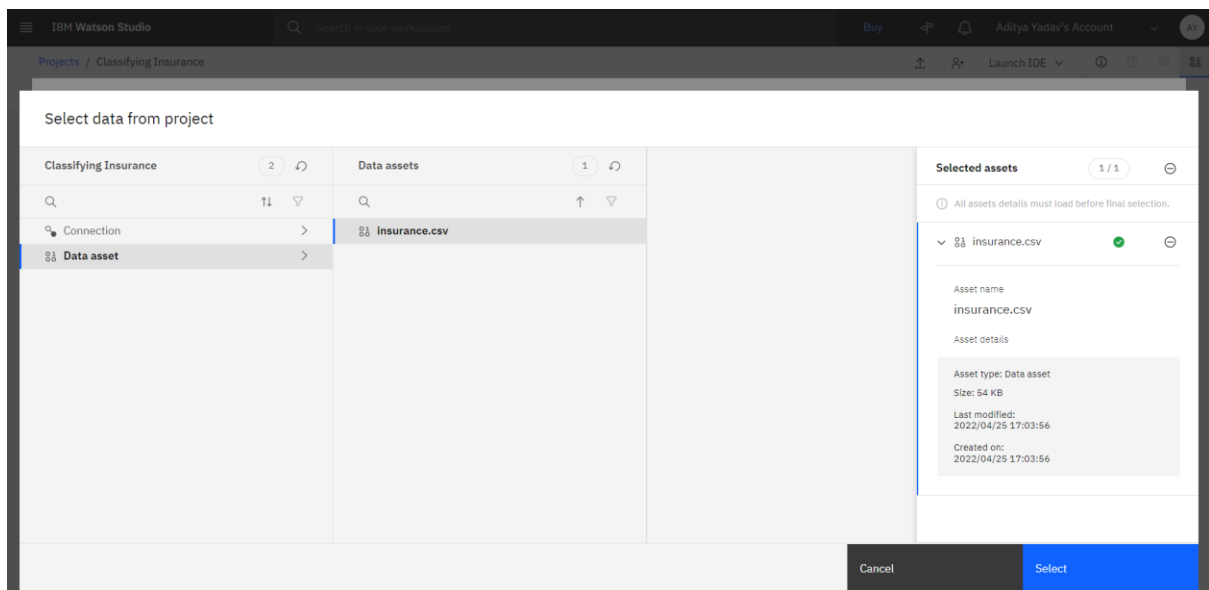
Go for data refining for the given dataset

Profile for dataset

Visualize the data

Visualize with other parameters

Create a SPSS modeler

Workspace

Import data asset and connect dataset to it

Connect data audit to data asset and run model

Data Audit output

Auto Classifier model

View Model: smoker

Auto Classifier ⓘ

Models

Auto Classifier - Models ⓘ
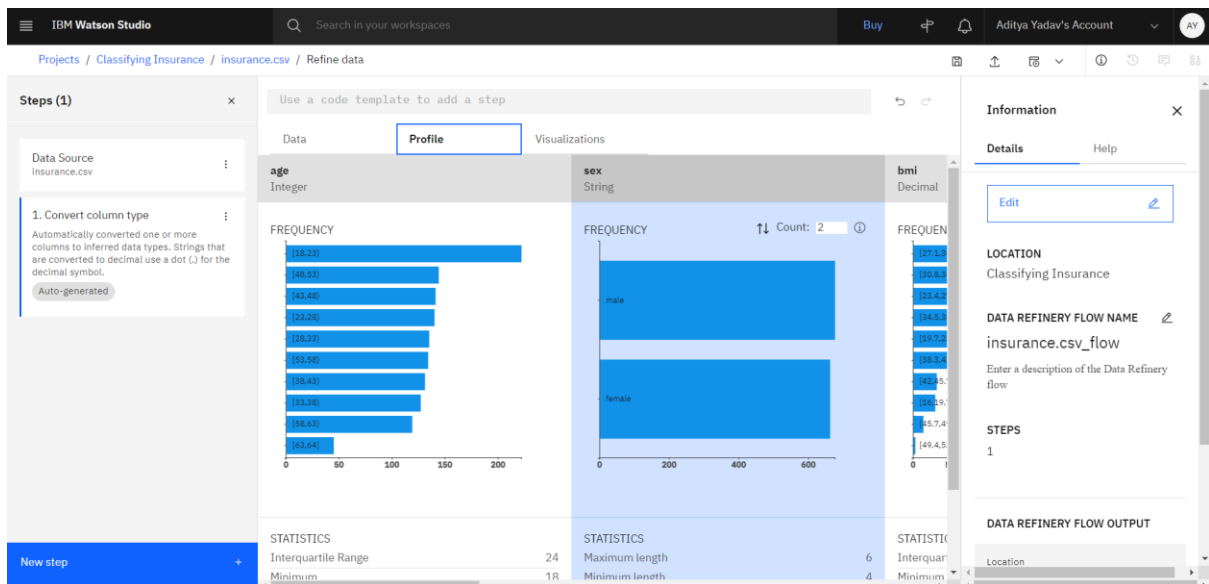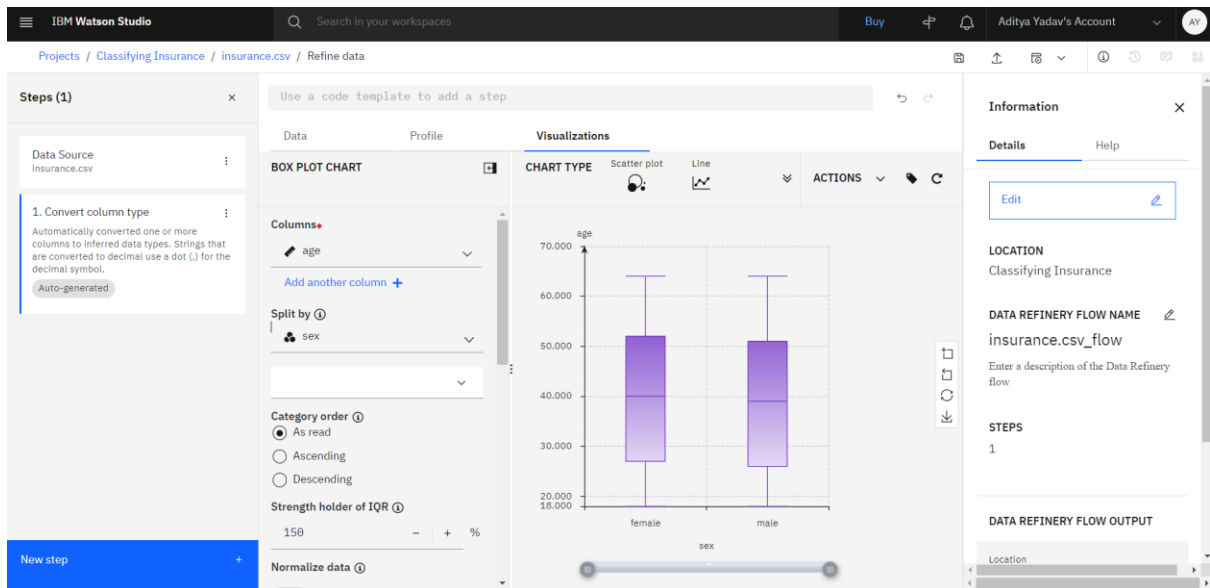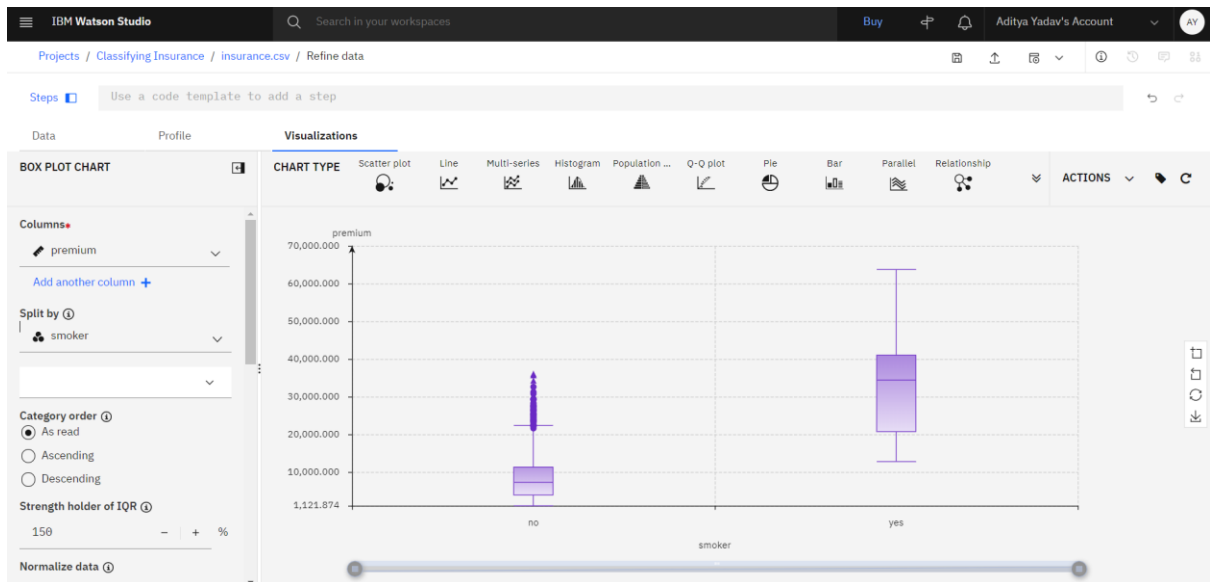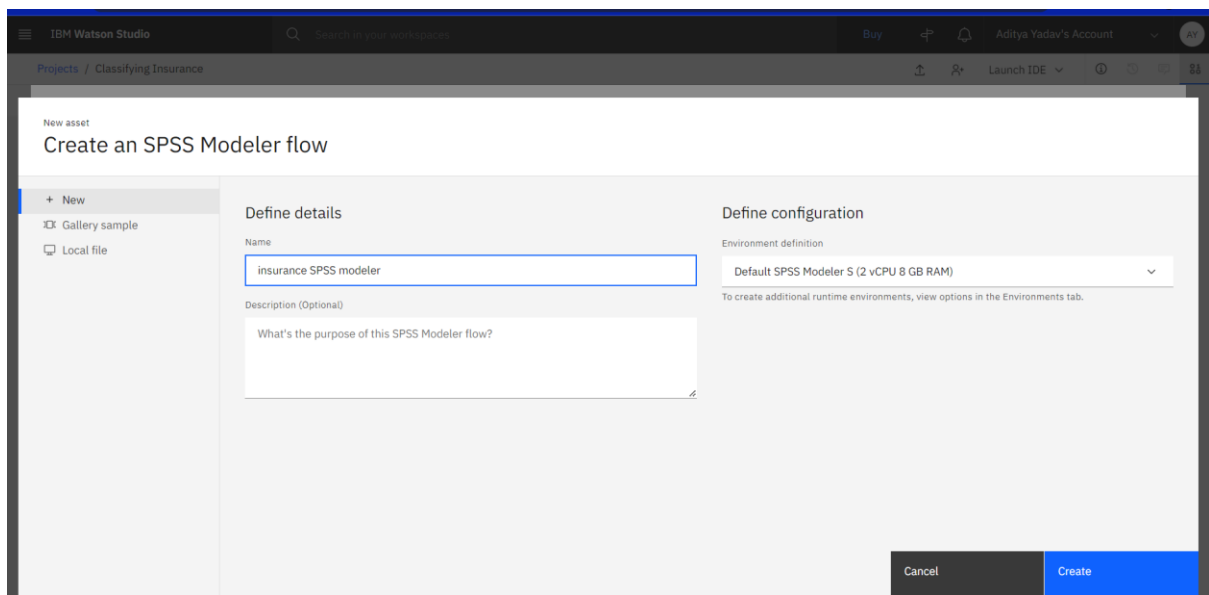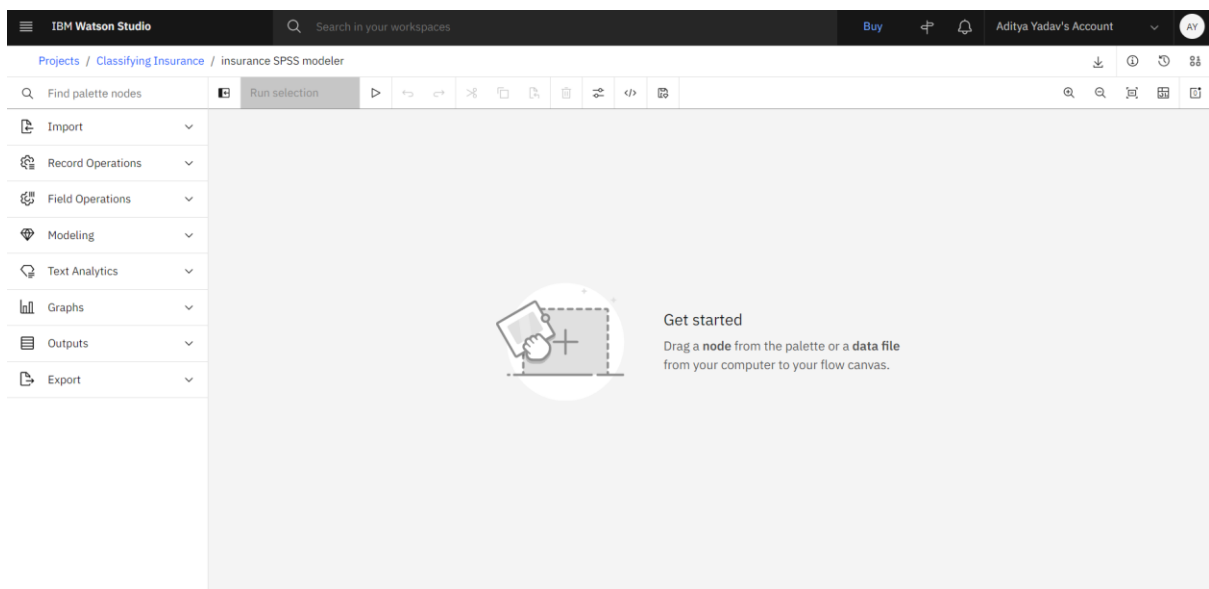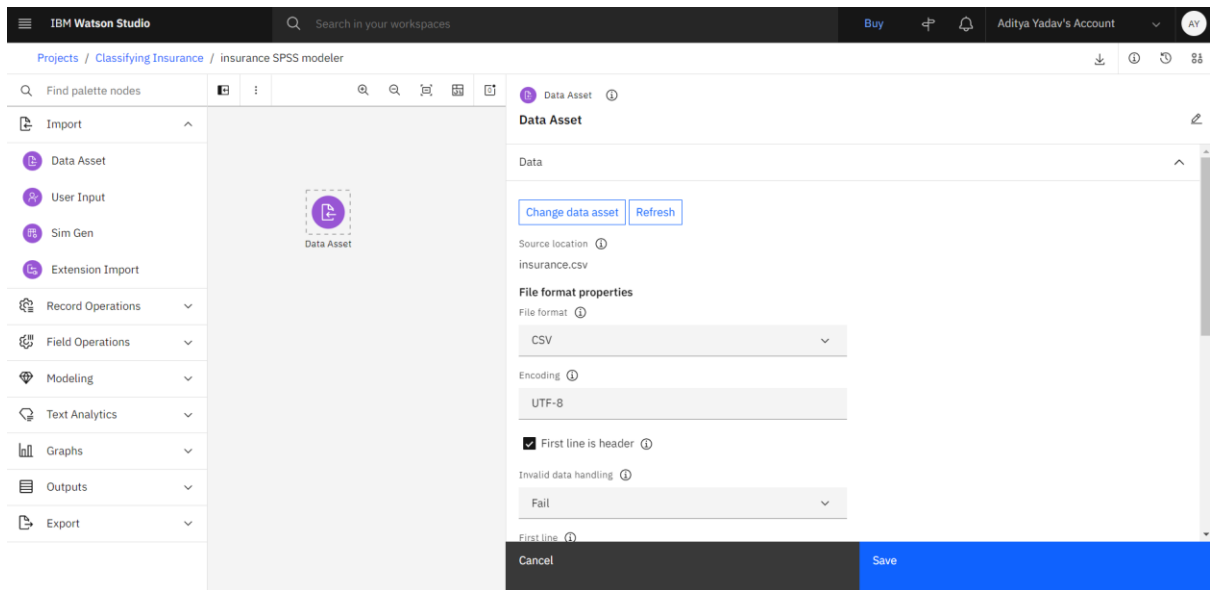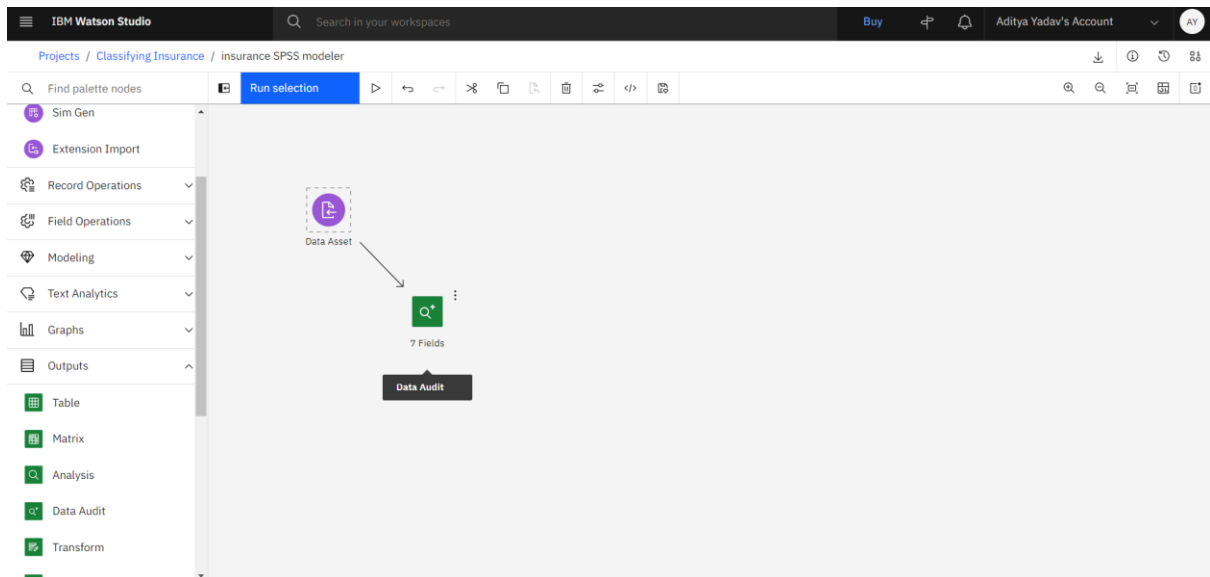
TARGET : SMOKER

USE	MODEL NAME	ESTIMATOR ▼	BUILD TIME (MINS)	NO. FIELDS USED	ACCURACY	ACCUMULATED ACCURACY	AREA UNDER CURVE	ACCUMULATED AUC	RECALL	PRECISION	
☑	XGBoost Tree 1	XGBoost Binary Classification Model	< 1	3	78.870	78.870	0.501	0.501	0.025	17.000	
☑	Neural Net 1	MLP Neural Network	< 1	3	80.098	80.098	0.540	0.540	0.000	0.000	
☑	C5 1	C5.0	< 1	3	80.098	80.098	0.500	0.500	0.000	15.000	
☑	C&R Tree 1	C&RT	< 1	3	80.098	80.098	0.500	0.500	0.000	16.000	

View model

View Output: Table (10 fields, 1,338 records)

Compare ▣

	age	sex	bmi	children	smoker	region	premium	Partition	$XF-smoker	$XFC-smoker
1	19	female	27.900	0	yes	southwest	16884.924	1_Training	no	0.751
2	18	male	33.770	1	no	southeast	1725.552	1_Training	no	0.803
3	28	male	33.000	3	no	southeast	4449.462	1_Training	no	0.790
4	33	male	22.705	0	no	northwest	21984.471	2_Testing	no	0.787
5	32	male	28.880	0	no	northwest	3866.855	1_Training	no	0.803
6	31	female	25.740	0	no	southeast	3756.622	1_Training	no	0.799
7	46	female	33.440	1	no	southeast	8240.590	1_Training	no	0.792
8	37	female	27.740	3	no	northwest	7281.506	1_Training	no	0.758
9	37	male	29.830	2	no	northeast	6406.411	1_Training	no	0.818
10	60	female	25.840	0	no	northwest	28923.137	1_Training	no	0.829
11	25	male	26.220	0	no	northeast	2721.321	1_Training	no	0.821
12	62	female	26.290	0	yes	southeast	27808.725	1_Training	no	0.638
13	23	male	34.400	0	no	southwest	1826.843	2_Testing	no	0.803
14	56	female	39.820	0	no	southeast	11090.718	1_Training	no	0.826
15	27	male	42.130	0	yes	southeast	39611.758	2_Testing	no	0.783

Table output

Analysis output