Assignment-4

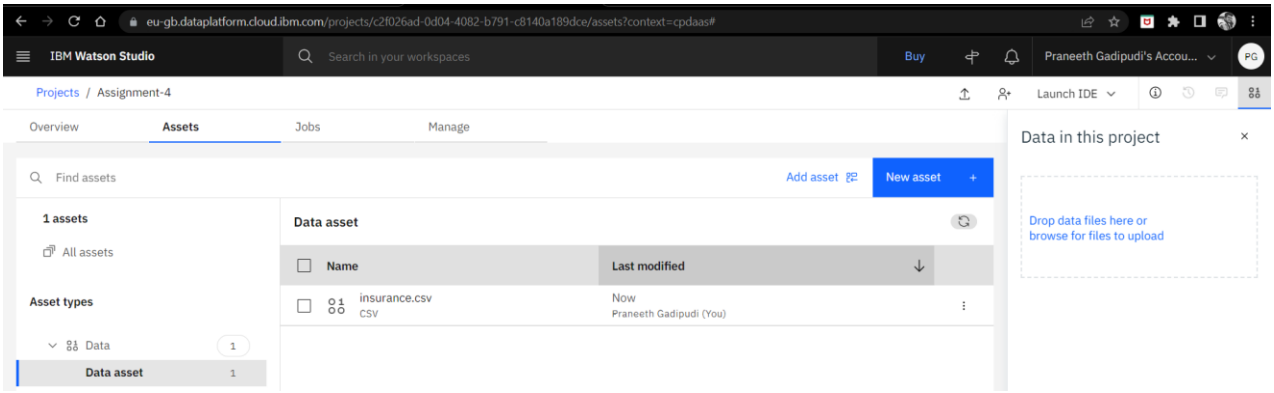
Name: G. Naga Sai Praneeth

Regd:19MIM10044

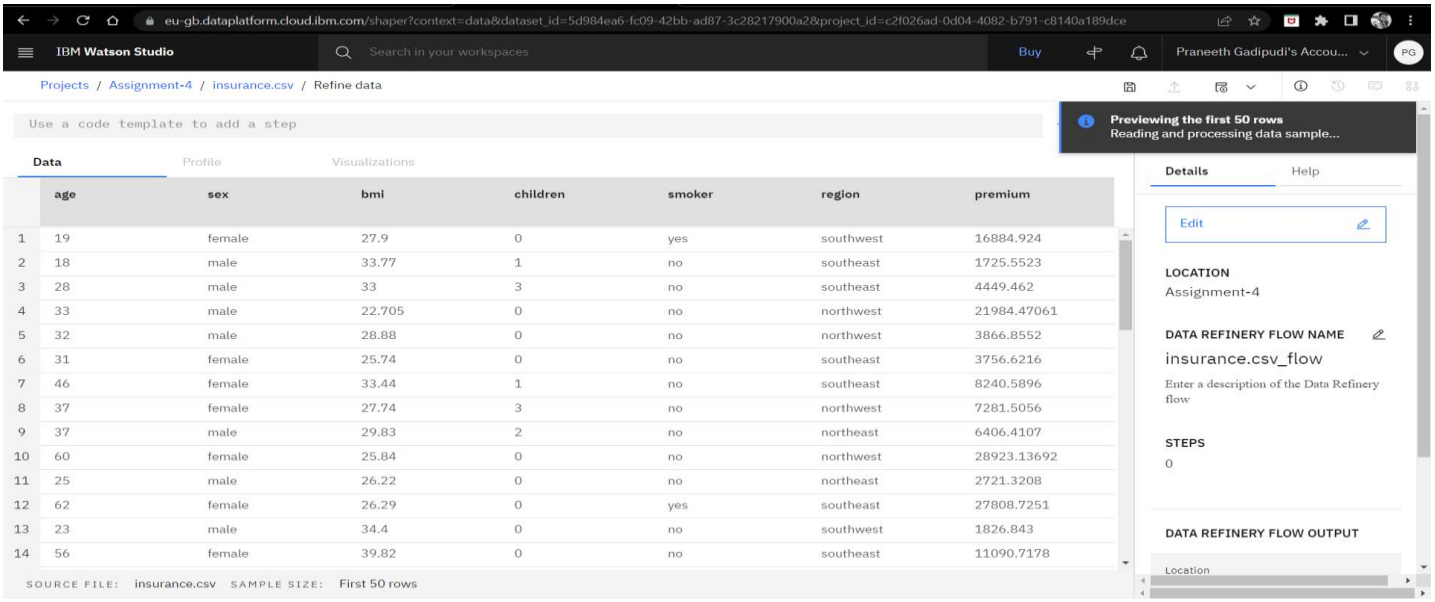
Clustering: Analysis of Medical Premium Charges for Insurance

Dataset: insurance.csv

Adding the dataset to the project

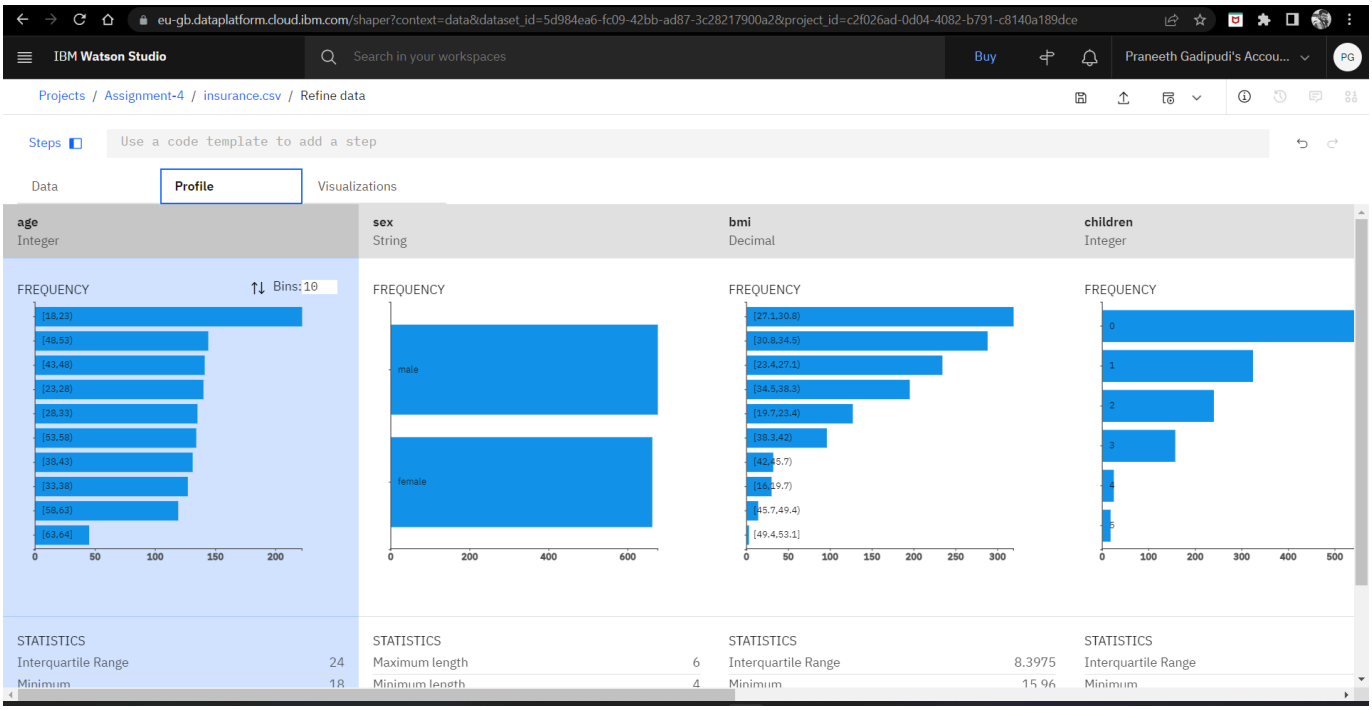


Data Refinery

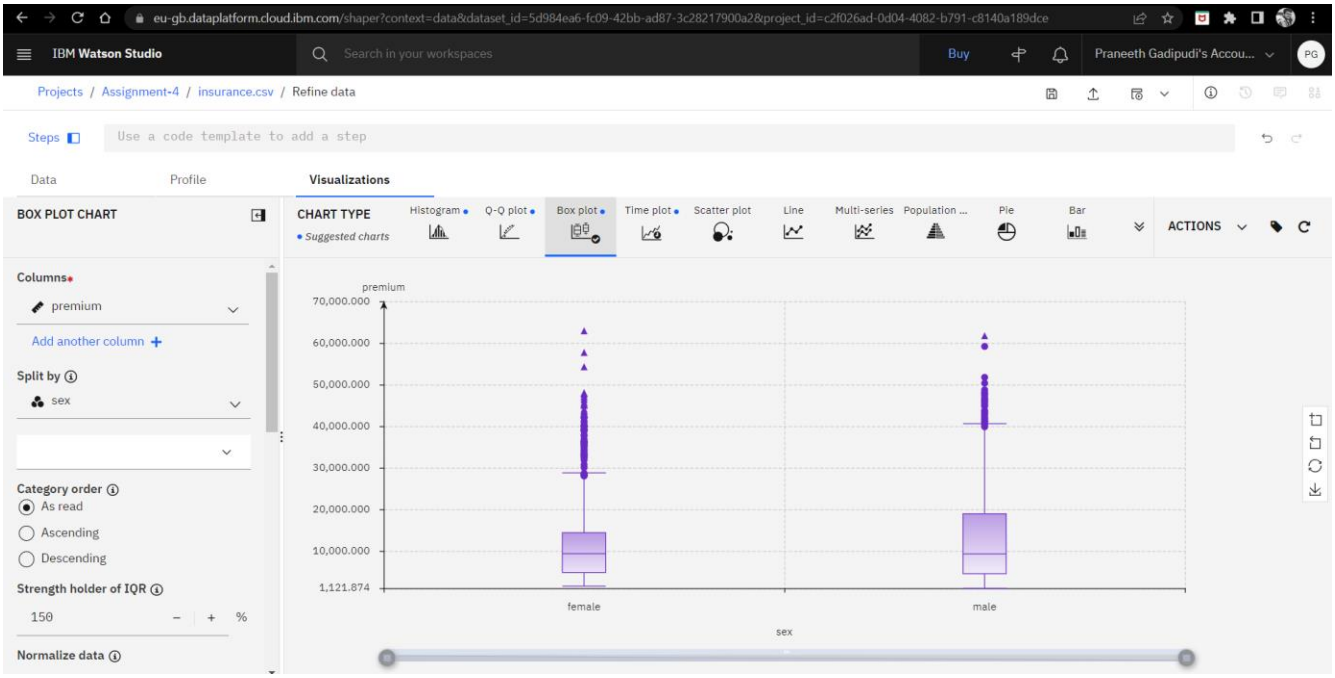


	age	sex	bmi	children	smoker	region	premium
1	19	female	27.9	0	yes	southwest	16884.924
2	18	male	33.77	1	no	southeast	1725.5523
3	28	male	33	3	no	southeast	4449.462
4	33	male	22.705	0	no	northwest	21984.47061
5	32	male	28.88	0	no	northwest	3866.8552
6	31	female	25.74	0	no	southeast	3756.6216
7	46	female	33.44	1	no	southeast	8240.5896
8	37	female	27.74	3	no	northwest	7281.5056
9	37	male	29.83	2	no	northeast	6406.4107
10	60	female	25.84	0	no	northwest	28923.13692
11	25	male	26.22	0	no	northeast	2721.3208
12	62	female	26.29	0	yes	southeast	27808.7251
13	23	male	34.4	0	no	southwest	1826.843
14	56	female	39.82	0	no	southeast	11090.7178

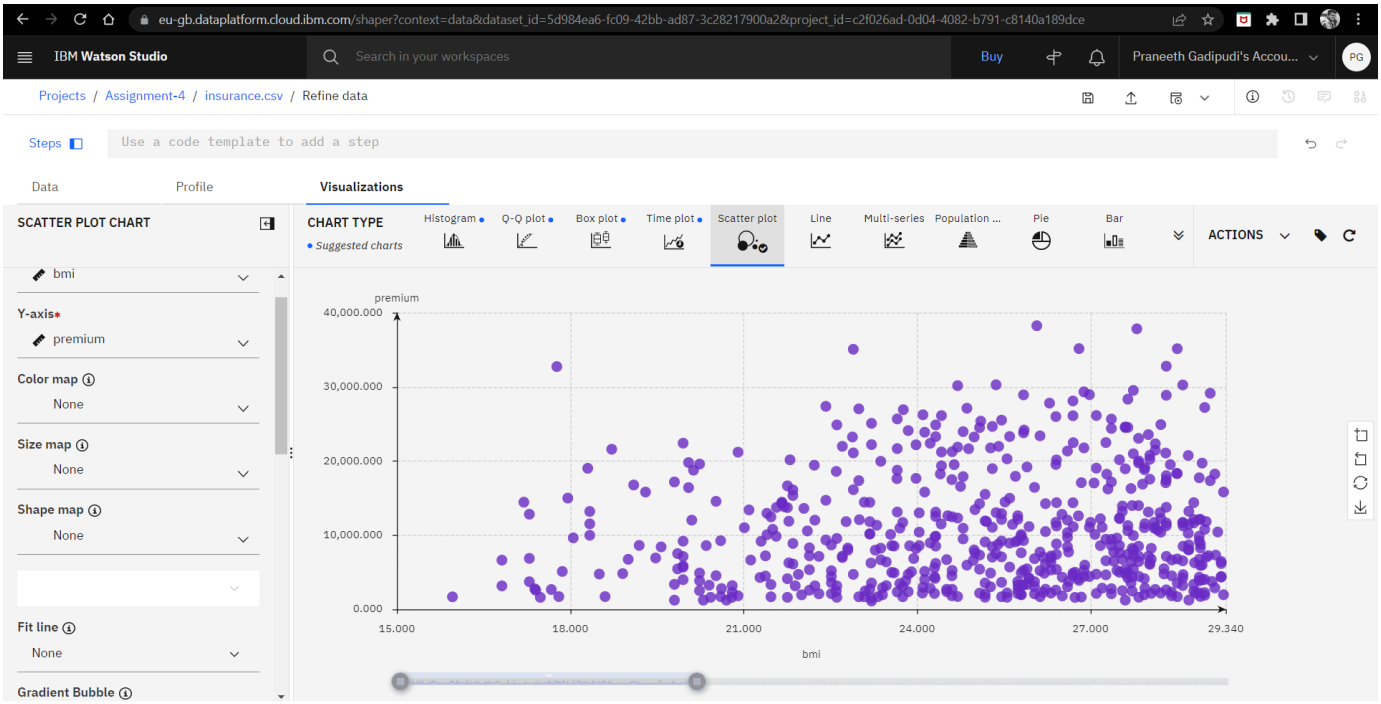
Profile:



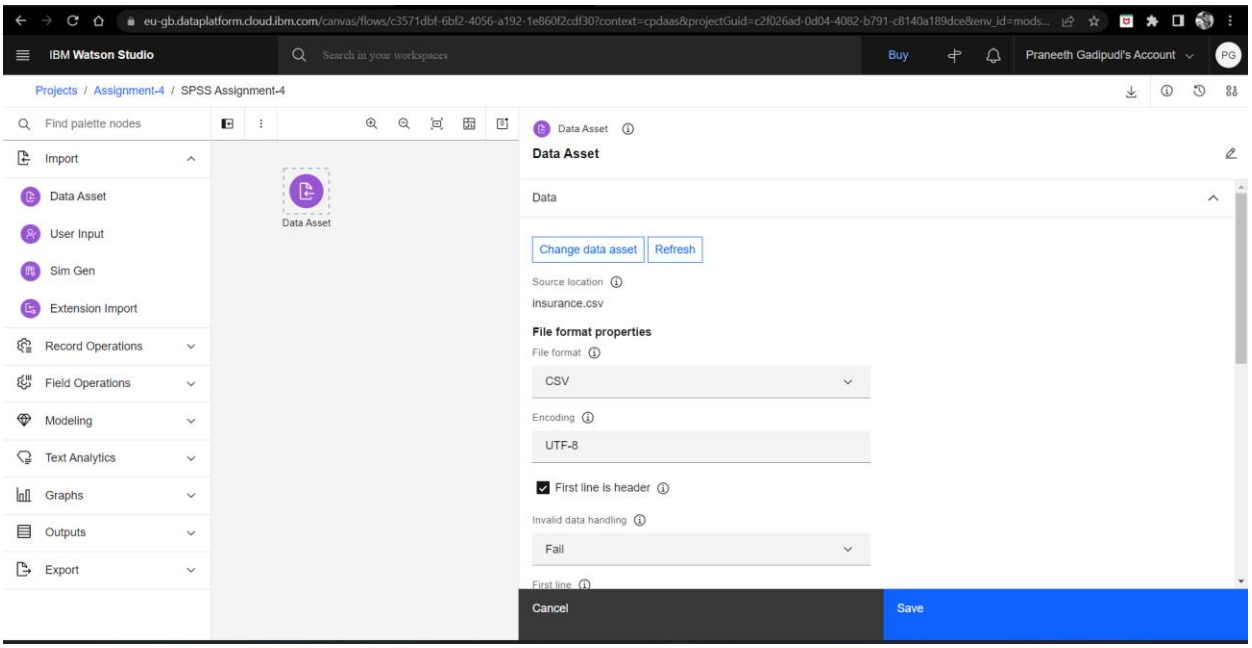
Visualization: Viewing Boxplot of Sex vs Premium



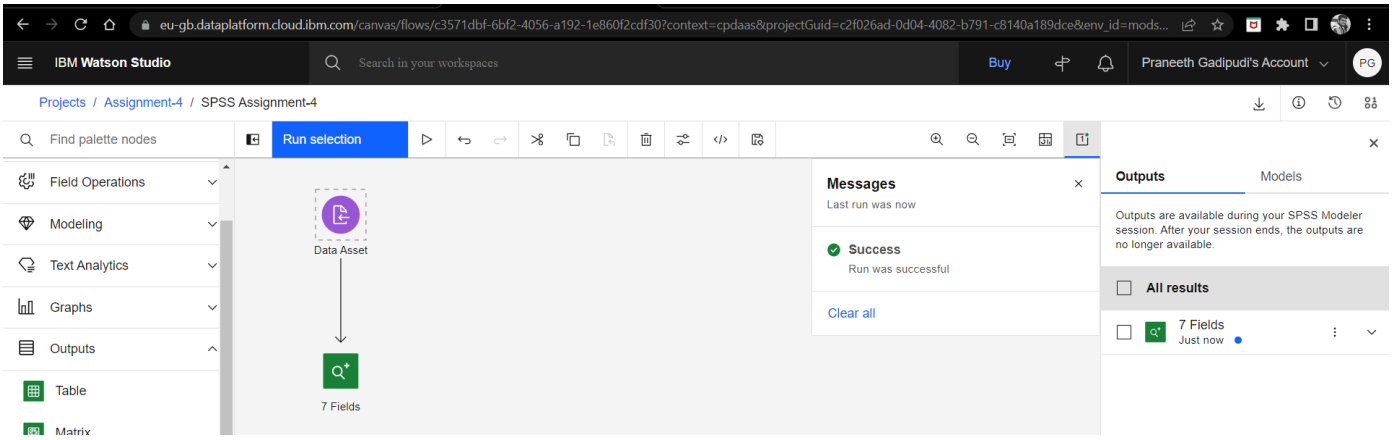
Scatterplot of BMI vs Premium:



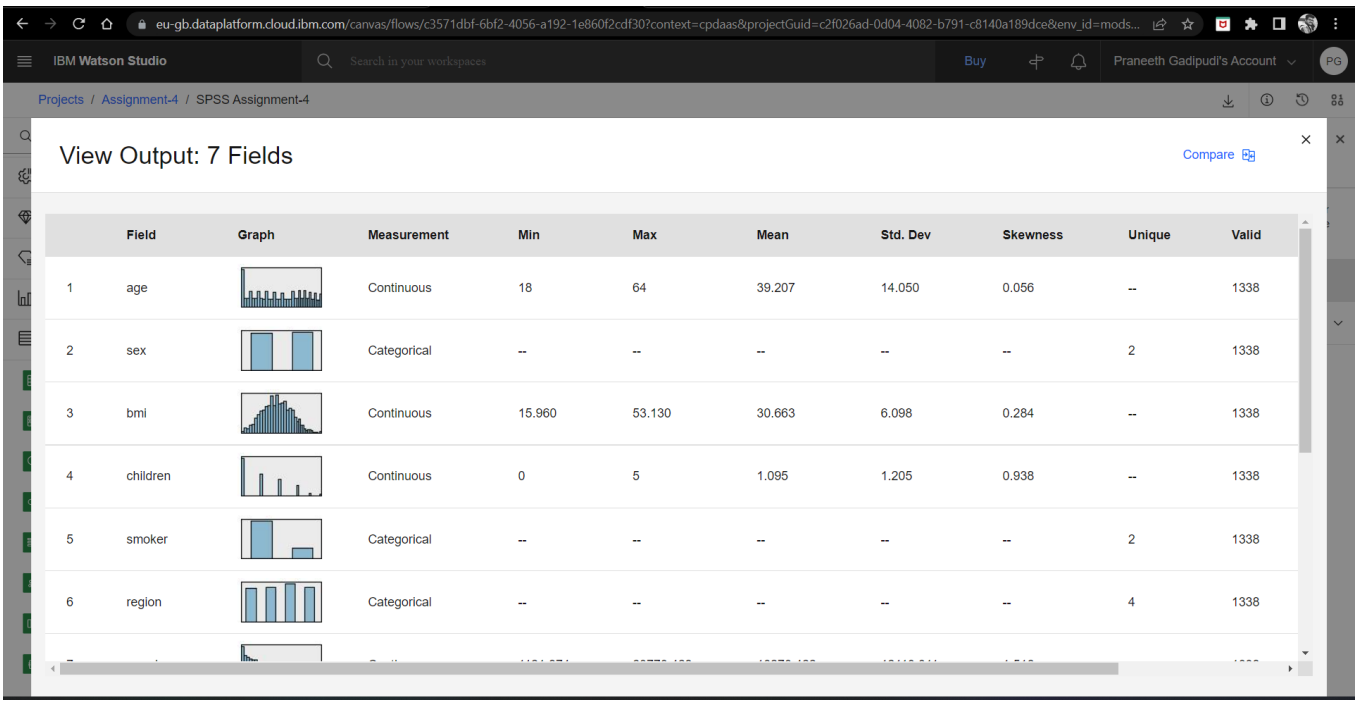
Creating an SPSS Modeler and uploading the dataset:



Data Audit:

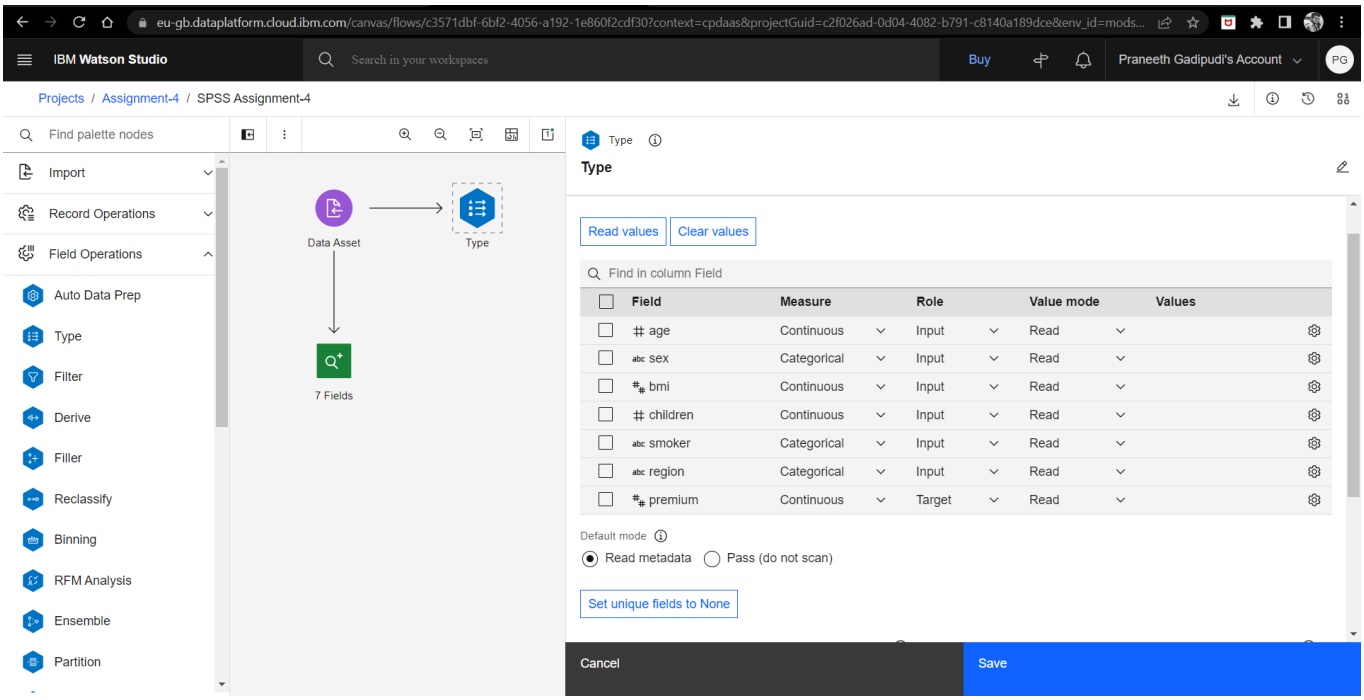


Data Audit output:

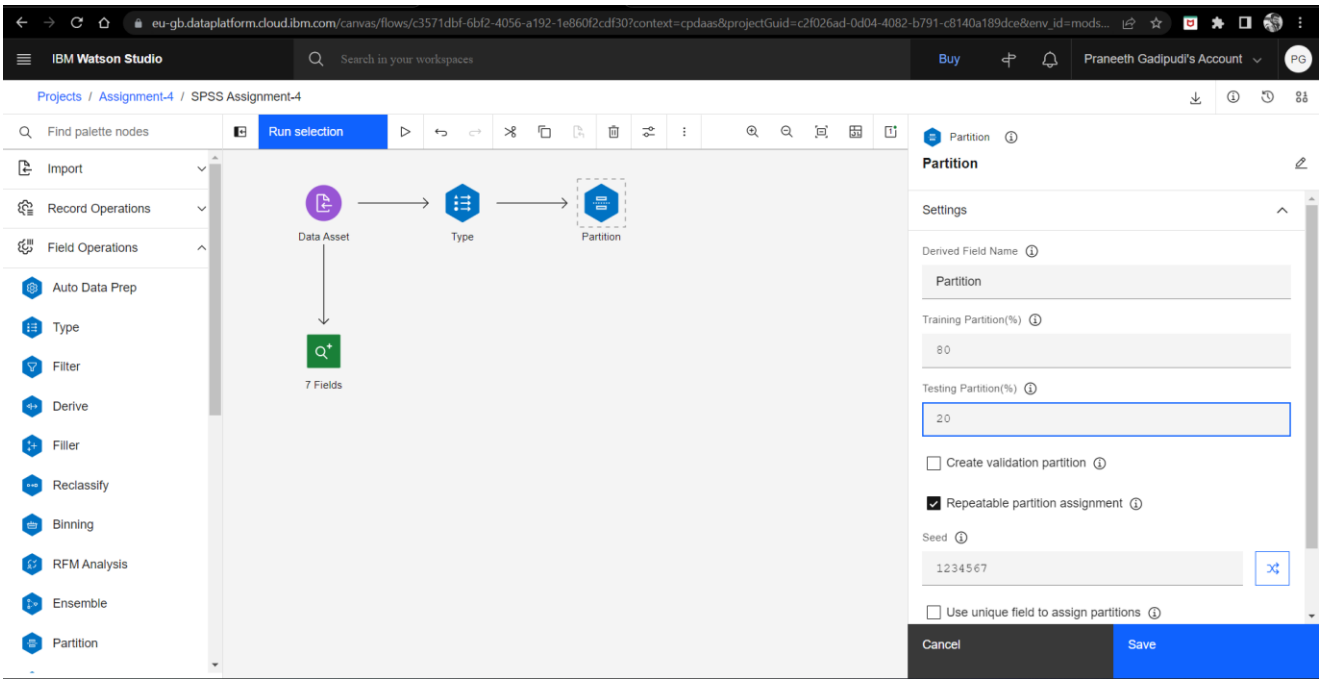


	Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	age		Continuous	18	64	39.207	14.050	0.056	--	1338
2	sex		Categorical	--	--	--	--	--	2	1338
3	bmi		Continuous	15.960	53.130	30.663	6.098	0.284	--	1338
4	children		Continuous	0	5	1.095	1.205	0.938	--	1338
5	smoker		Categorical	--	--	--	--	--	2	1338
6	region		Categorical	--	--	--	--	--	4	1338

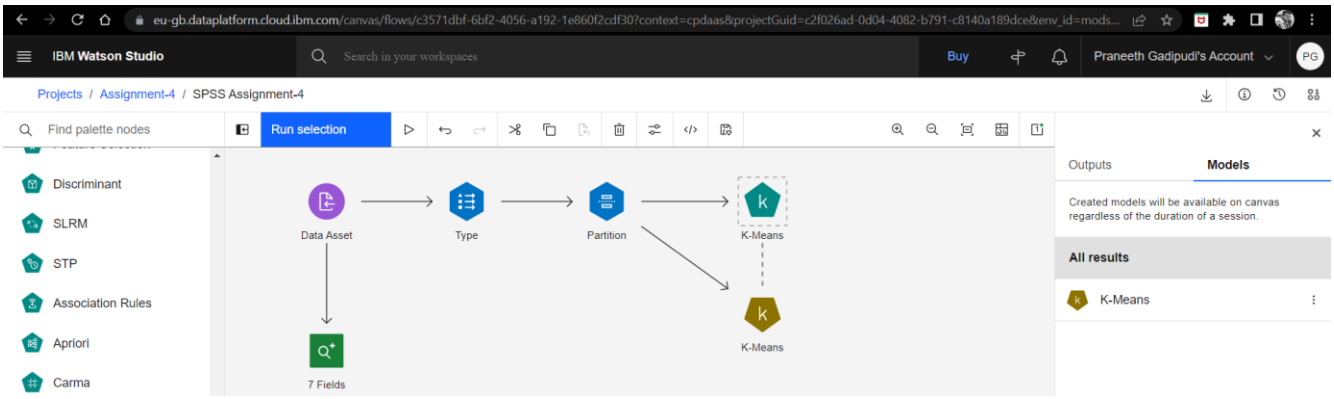
Type Node:



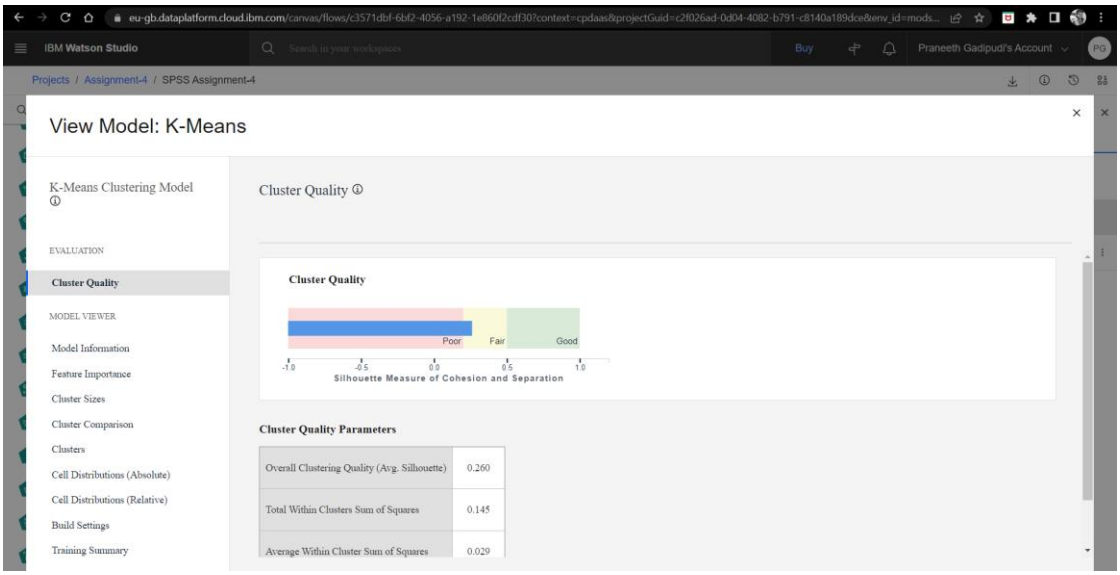
Partition Node:

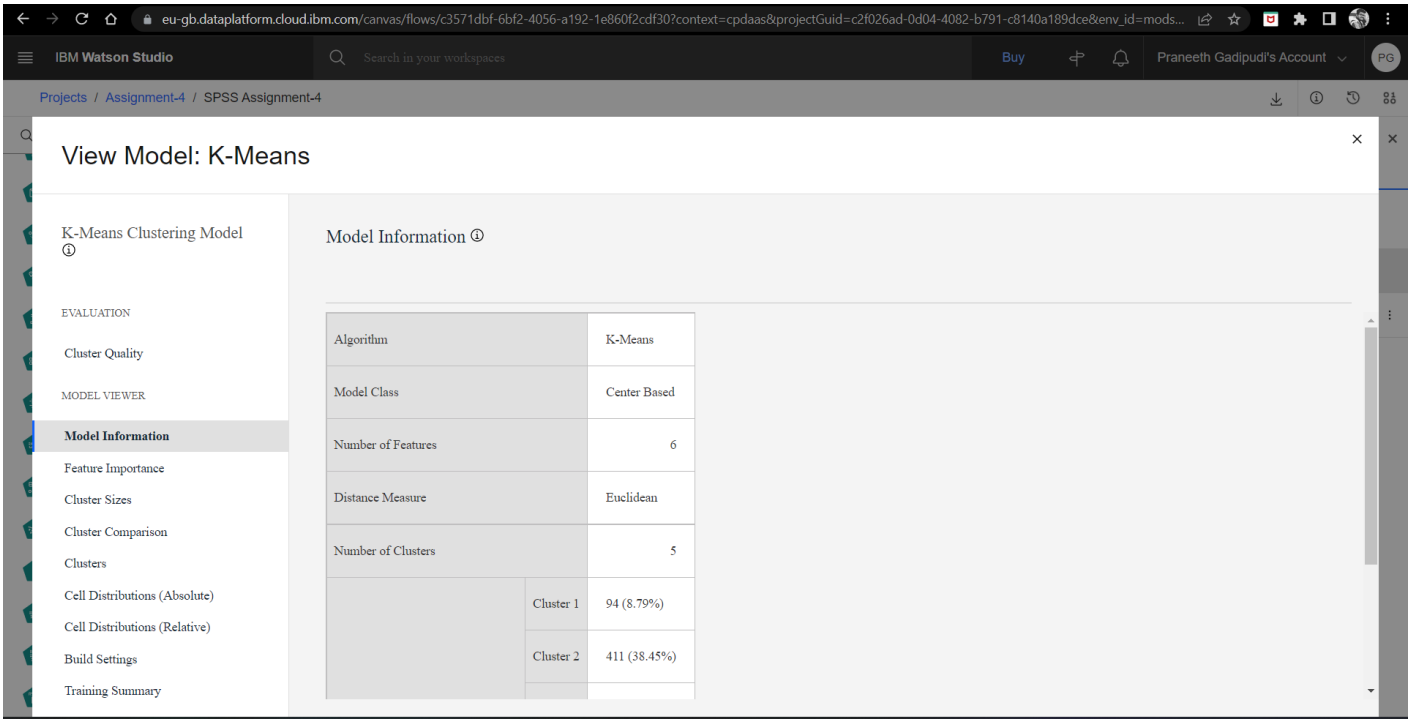


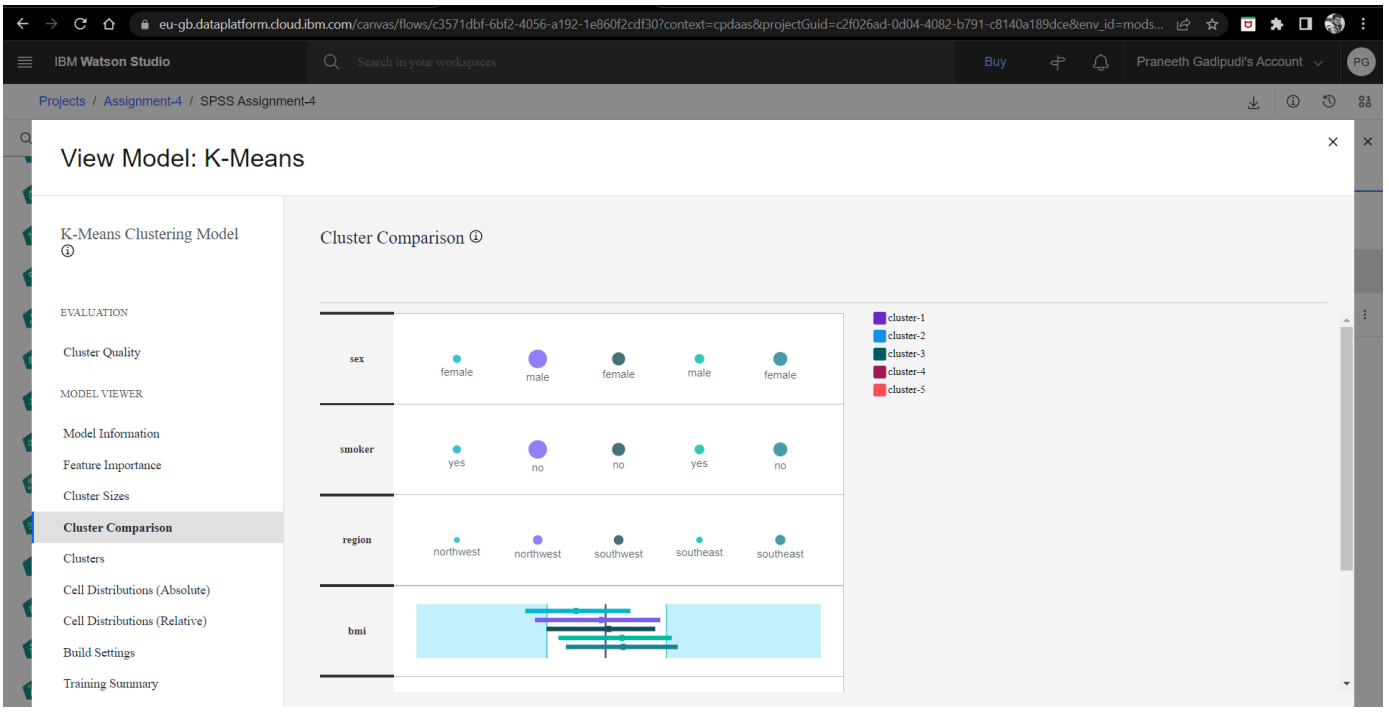
K-Means Cluster Model:



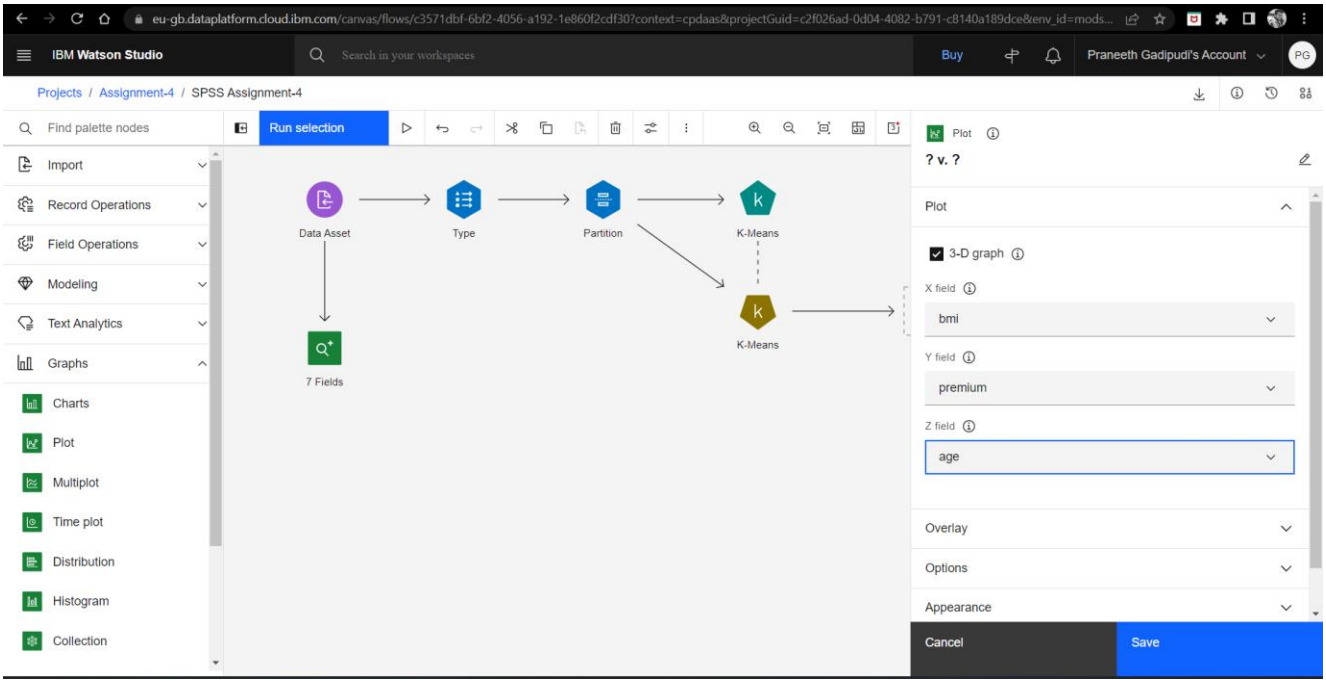
K-means Output:



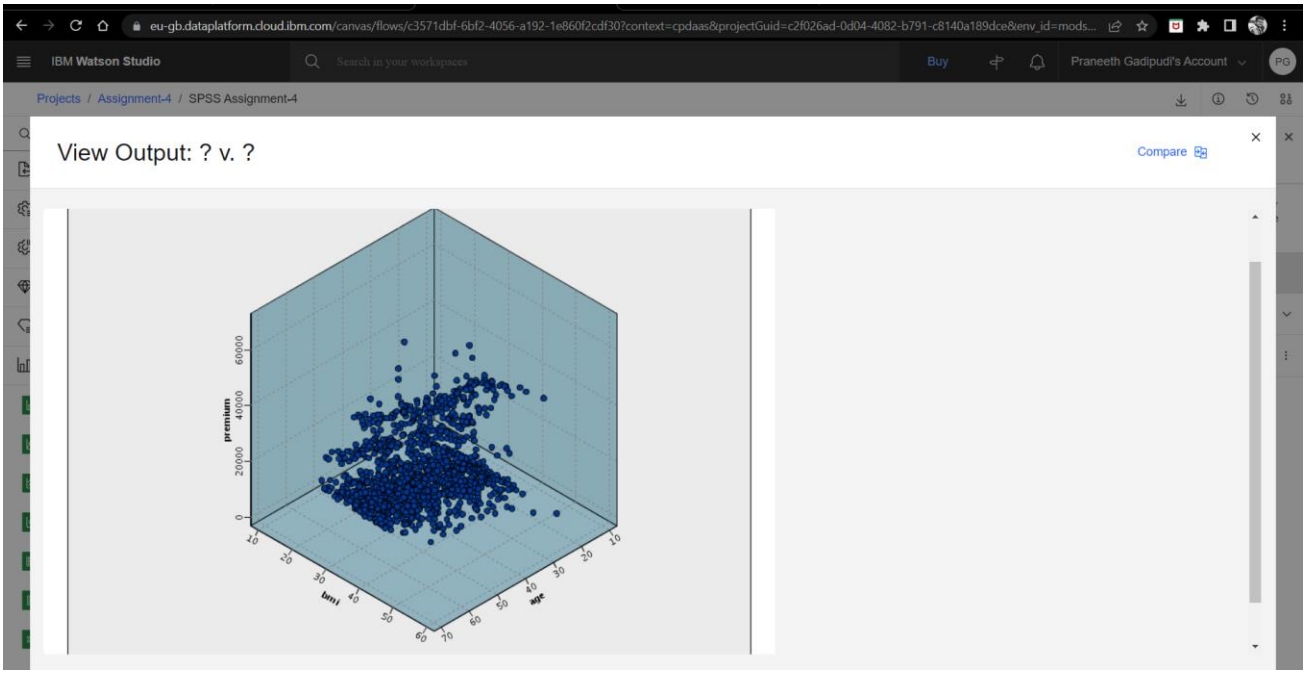

eu-gb.dataplatform.cloud.ibm.com/canvas/flows/c3571dbf-6bf2-4056-a192-1e860f2cdf30?context=cpdaas&projectGuid=c2f026ad-0d04-4082-b791-c8140a189dce&env_id=mods...
IBM Watson Studio
Search in your workspaces
Buy
Praneeth Gadipudi's Account
PG
Projects / Assignment-4 / SPSS Assignment-4
View Model: K-Means
K-Means Clustering Model
EVALUATION
Cluster Quality
MODEL VIEWER
Model Information
Feature Importance
Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary
Model Information
Algorithm
K-Means
Model Class
Center Based
Number of Features
6
Distance Measure
Euclidean
Number of Clusters
5
Cluster 1
94 (8.79%)
Cluster 2
411 (38.45%)


eu-gb.dataplatform.cloud.ibm.com/canvas/flows/c3571dbf-6bf2-4056-a192-1e860f2cdf30?context=cpdaas&projectGuid=c2f026ad-0d04-4082-b791-c8140a189dce&env_id=mods...
IBM Watson Studio
Search in your workspaces
Buy
Praneeth Gadipudi's Account
PG
Projects / Assignment-4 / SPSS Assignment-4
View Model: K-Means
K-Means Clustering Model
EVALUATION
Cluster Quality
MODEL VIEWER
Model Information
Feature Importance
Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary
Cluster Comparison
sex
female
male
female
male
female
smoker
yes
no
no
yes
no
region
northwest
northwest
southwest
southeast
southeast
bmi
cluster-1
cluster-2
cluster-3
cluster-4
cluster-5

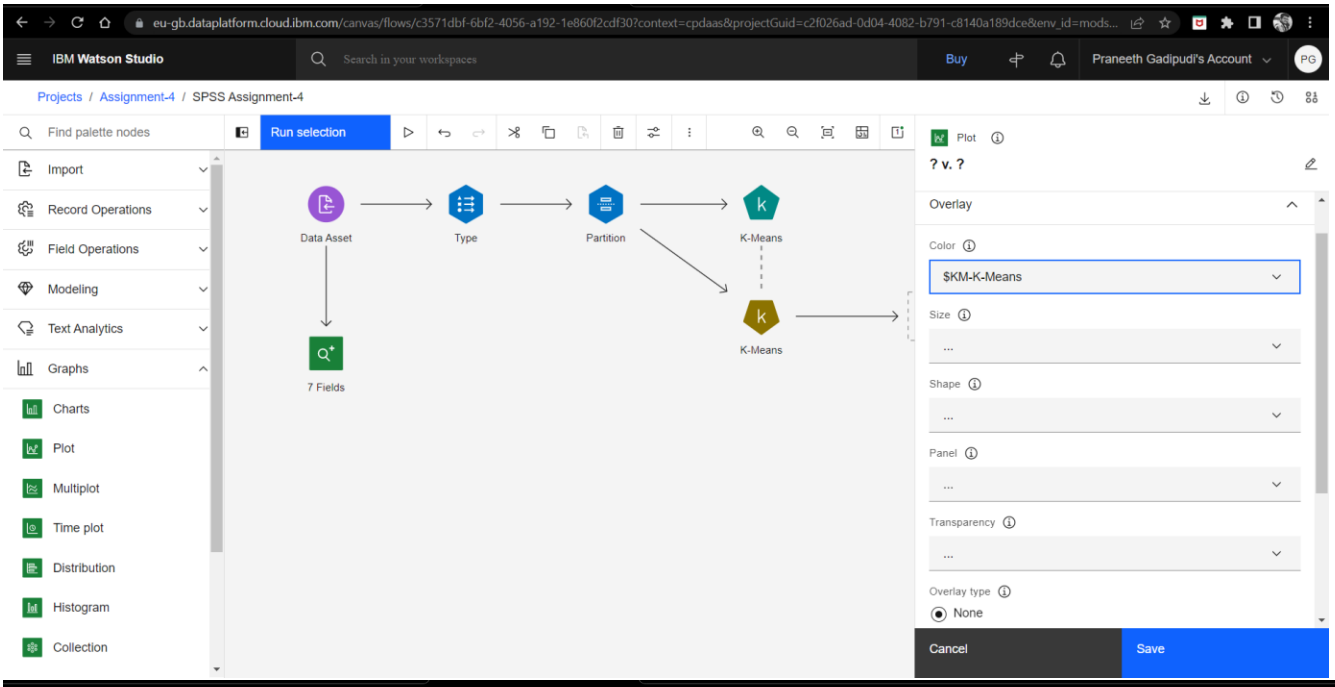
Plot Node:

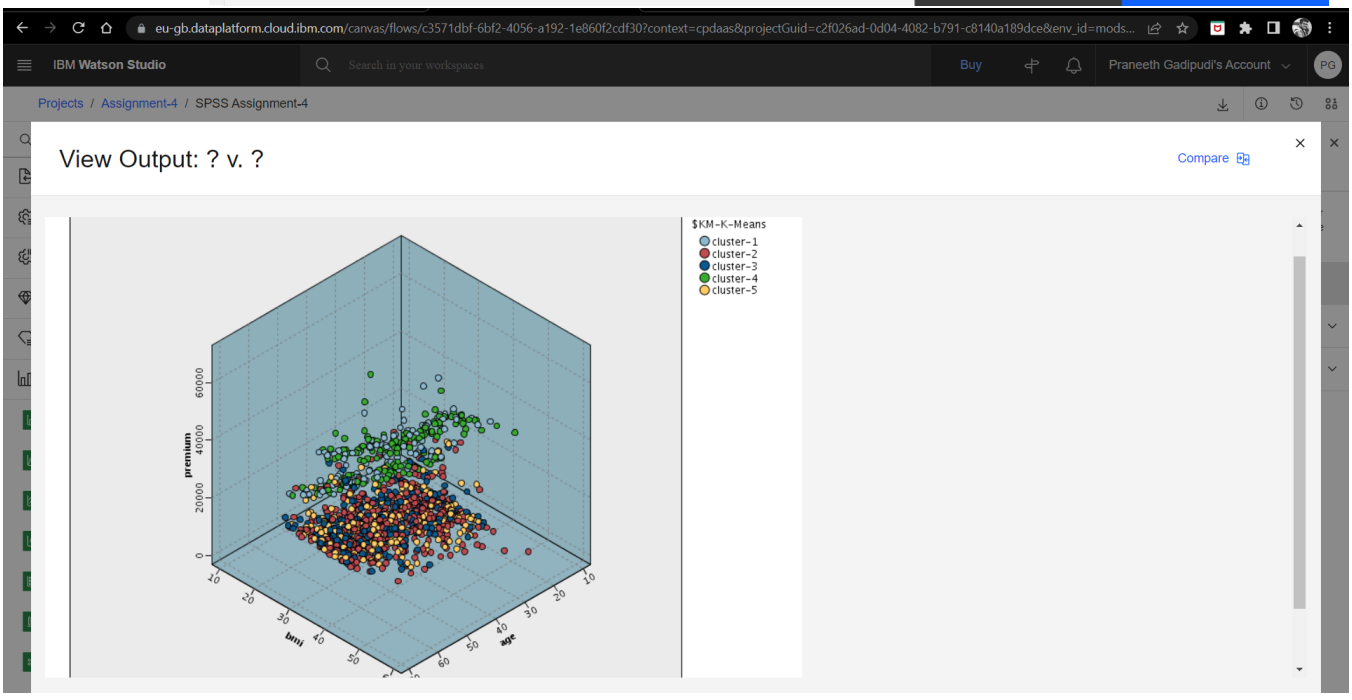


Output:

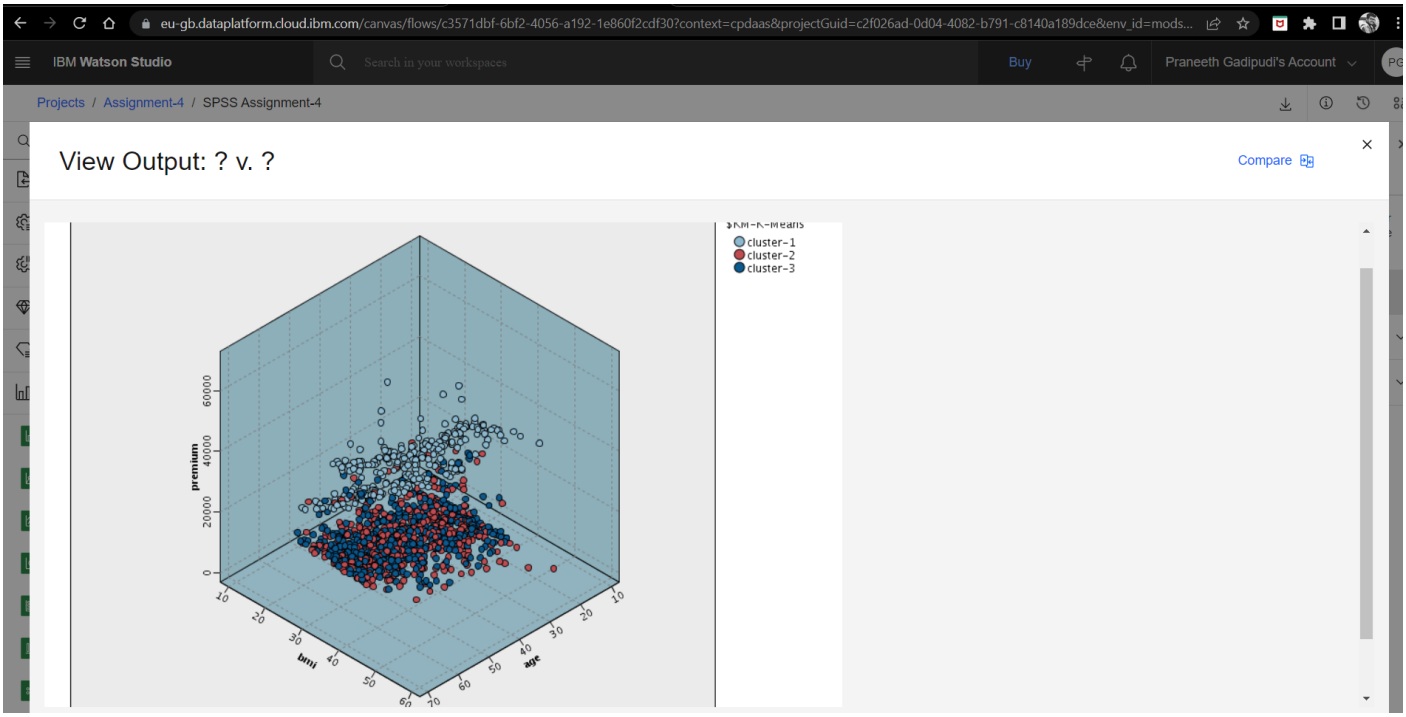


Assigning Color:





Cluster =3:


eu-gb.dataplatform.cloud.ibm.com/canvas/flows/c3571dbf-6bf2-4056-a192-1e860f2cdf30?context=cpdaas&projectGuid=c2f026ad-0d04-4082-b791-c8140a189dce&env_id=mods...
IBM Watson Studio
Search in your workspaces
Buy
Praneeth Gadipudi's Account
Projects / Assignment-4 / SPSS Assignment-4
View Output: ? v. ?
Compare
cluster-1
cluster-2
cluster-3
premium
bmi
age