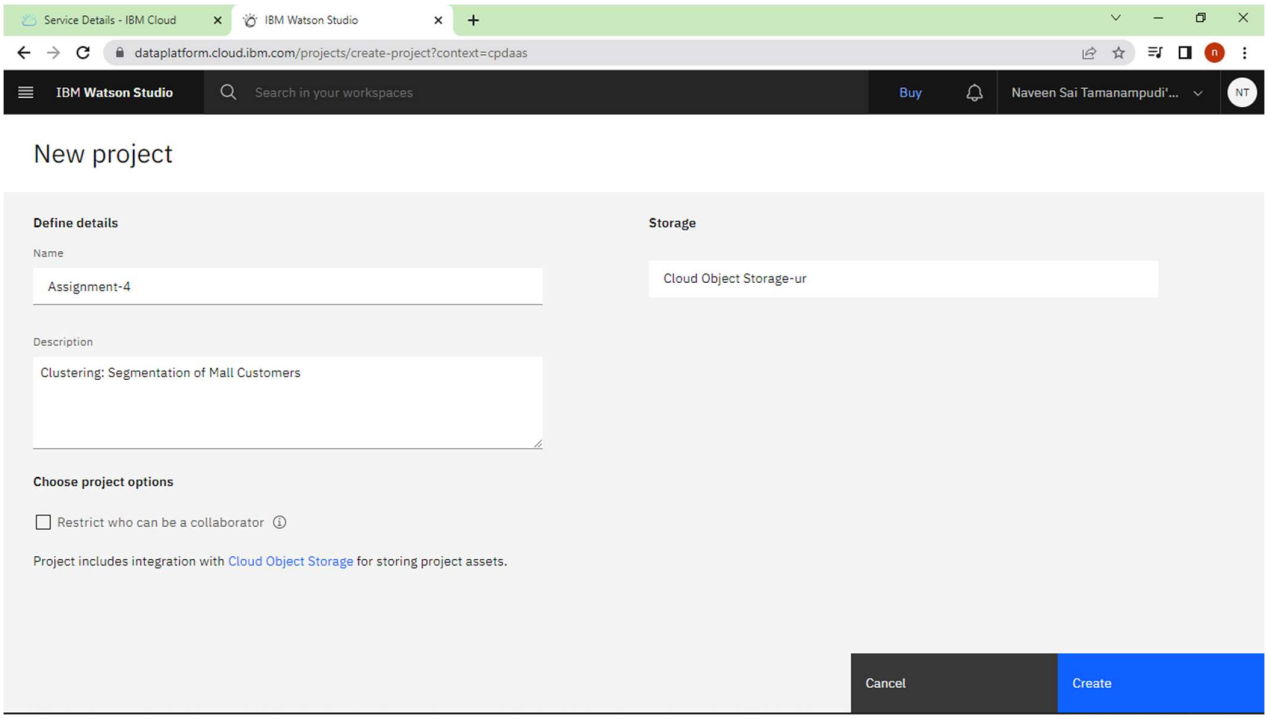
Data Analytics Externship Program
Assignment-4

Name: Tamanampudi Naveen Sai
Email ID: naveen.19bcd7003@vitap.ac.in

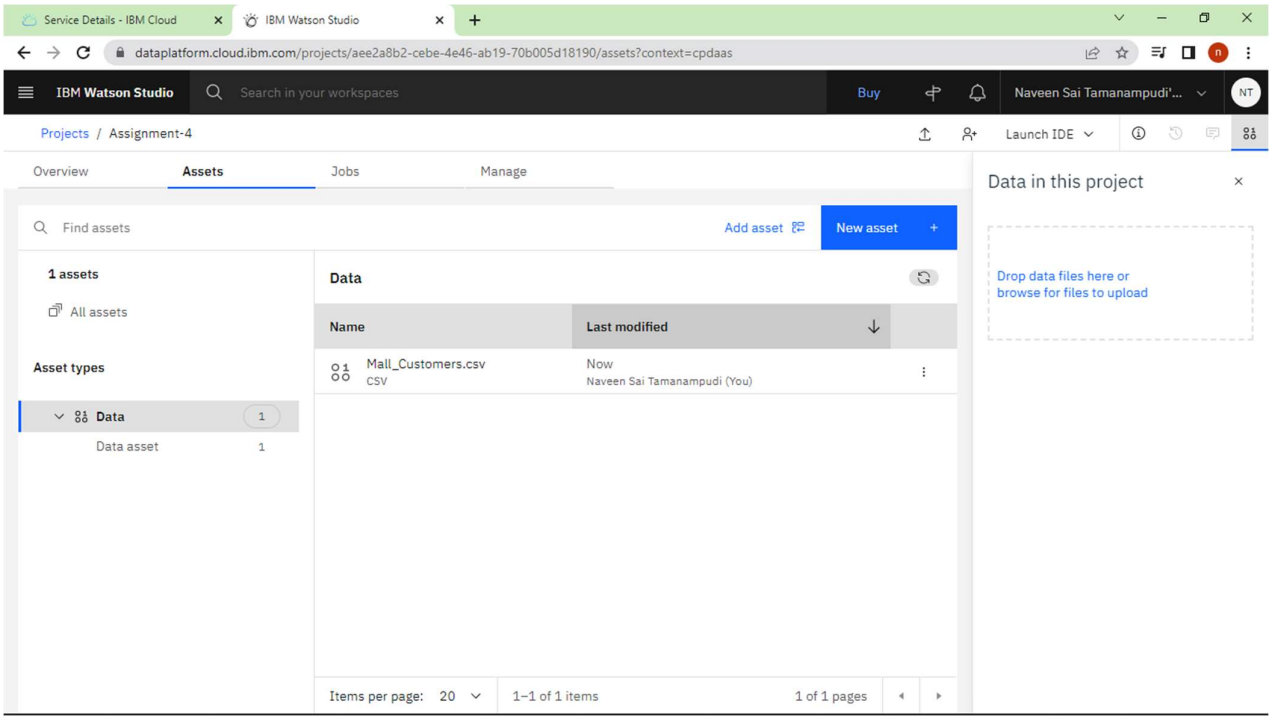
Clustering: Segmentation of Mall Customers

Dataset Used: Mall_Customers.csv

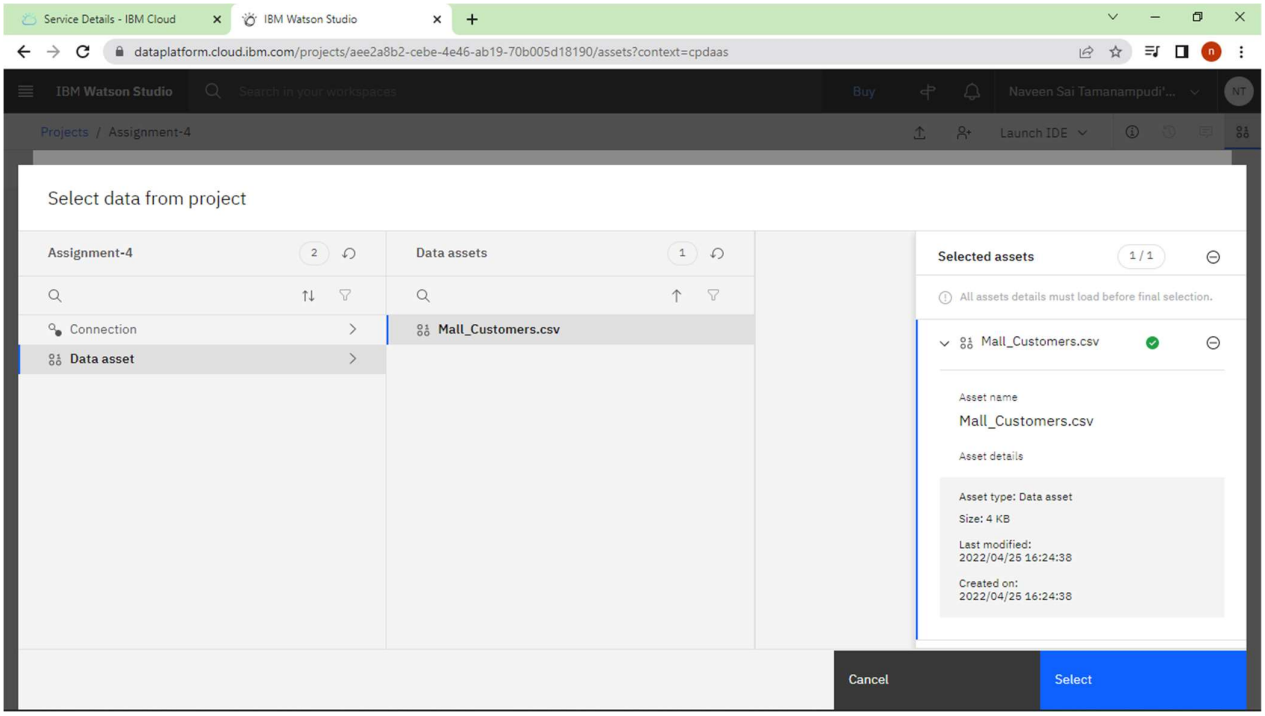
Creating the Project



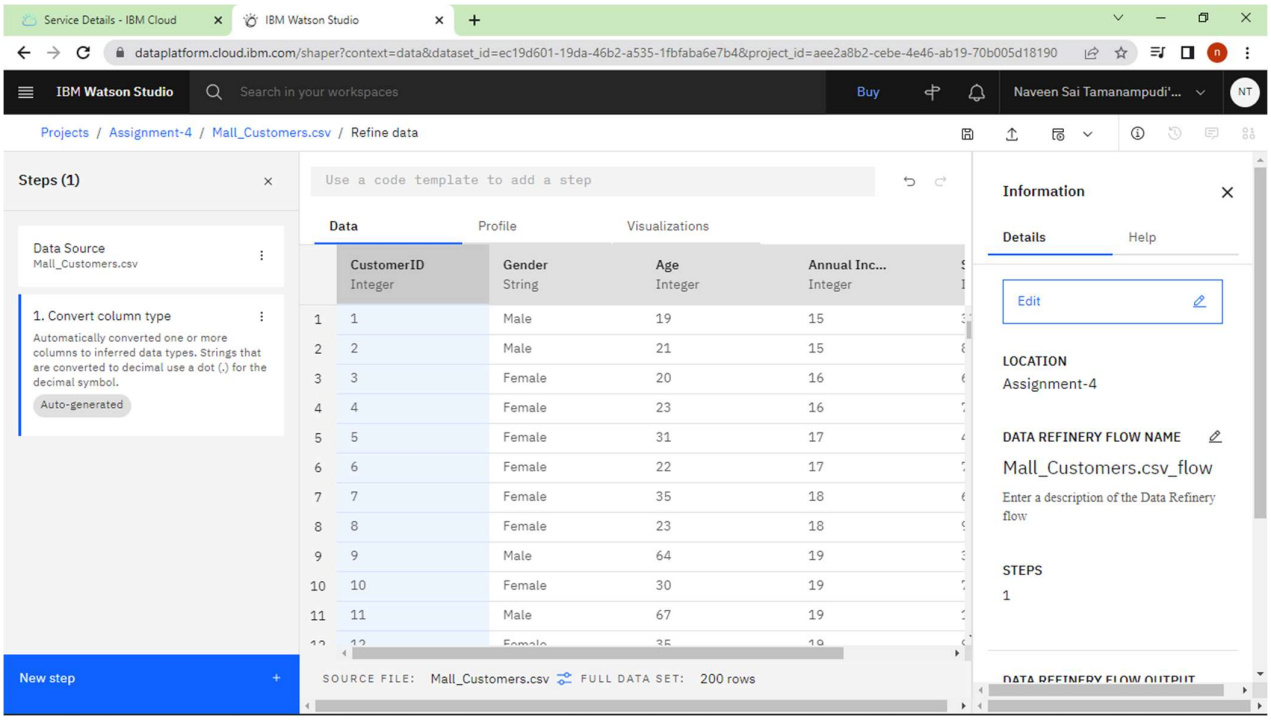
Uploading Data:



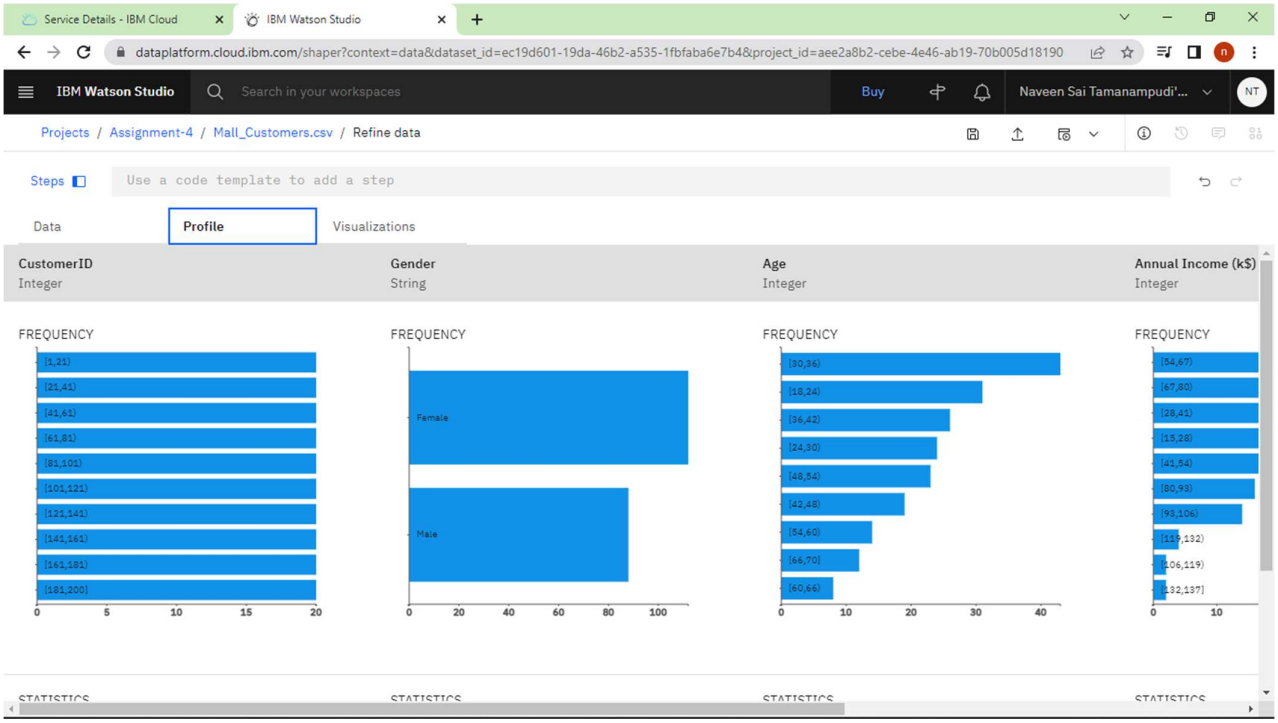
Creating a Data Refinery:

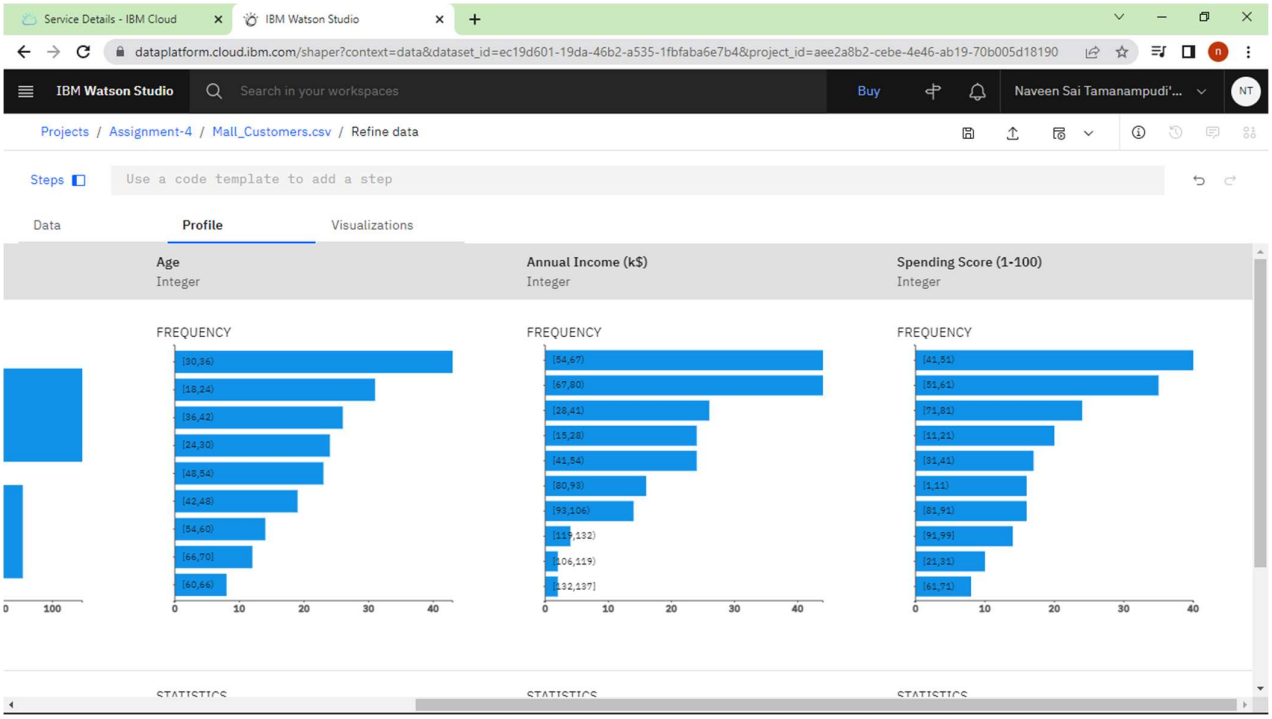


Viewing the output of refinery:

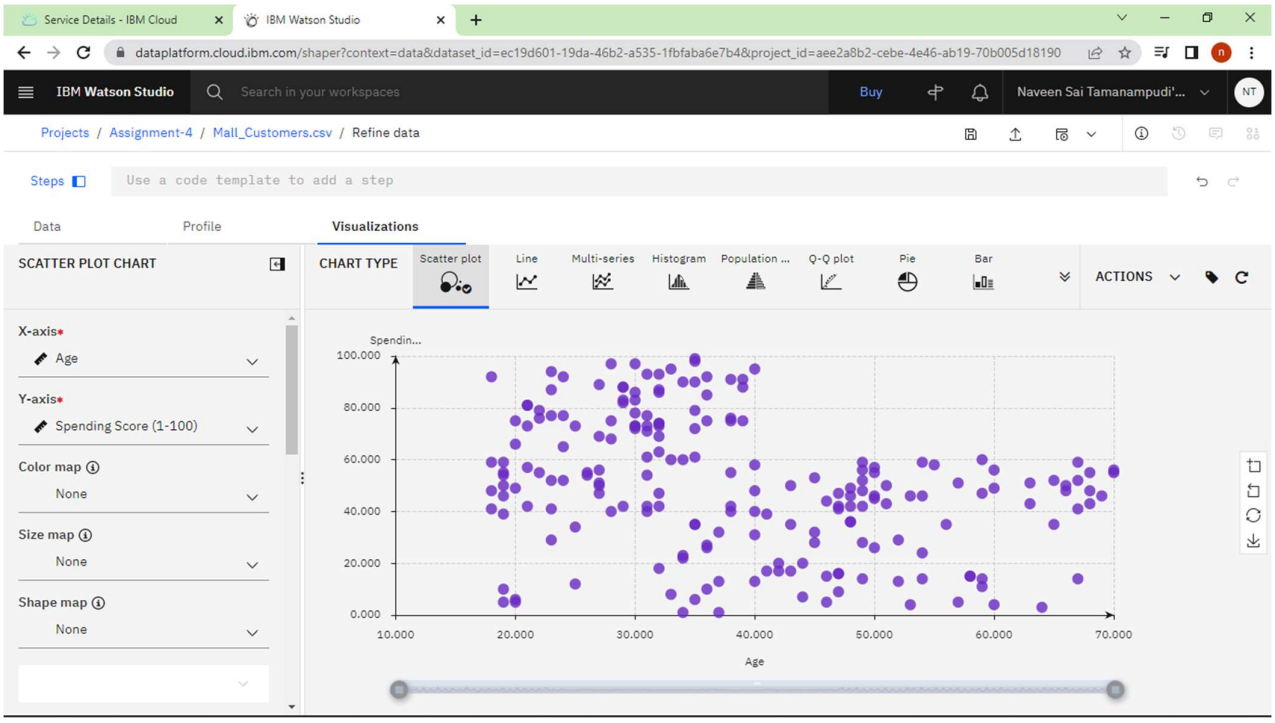


Viewing Profile:

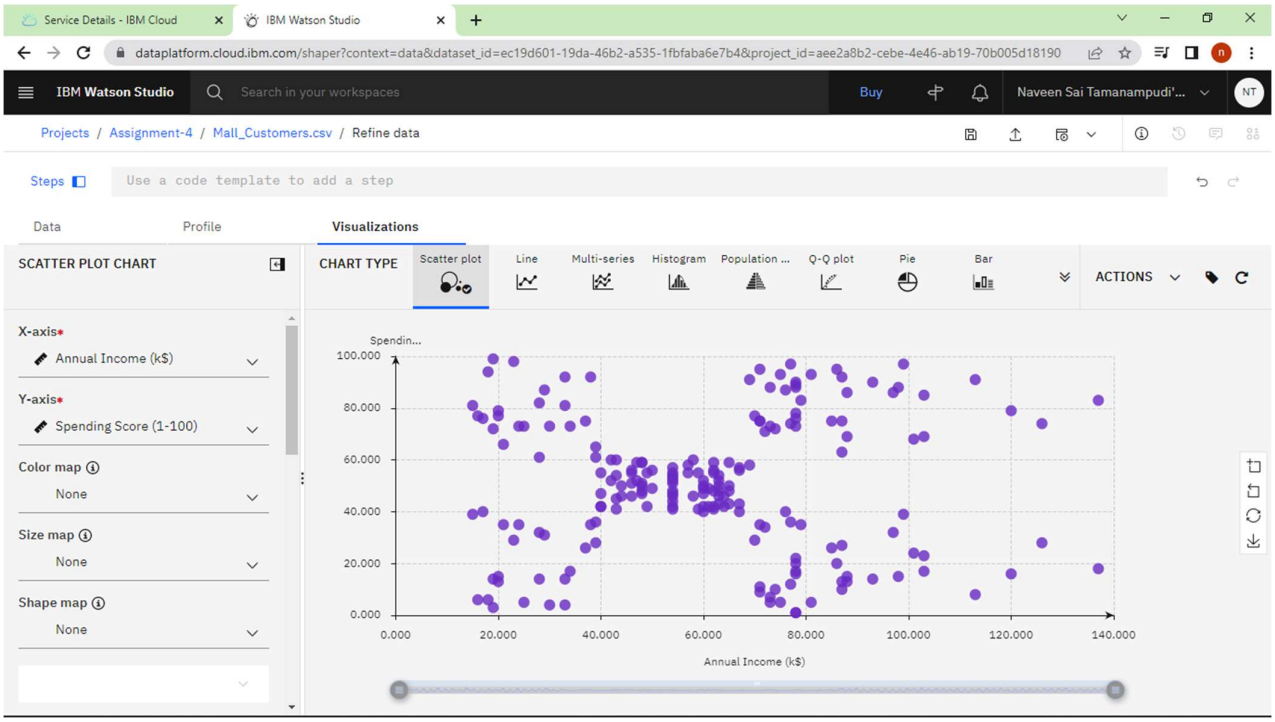




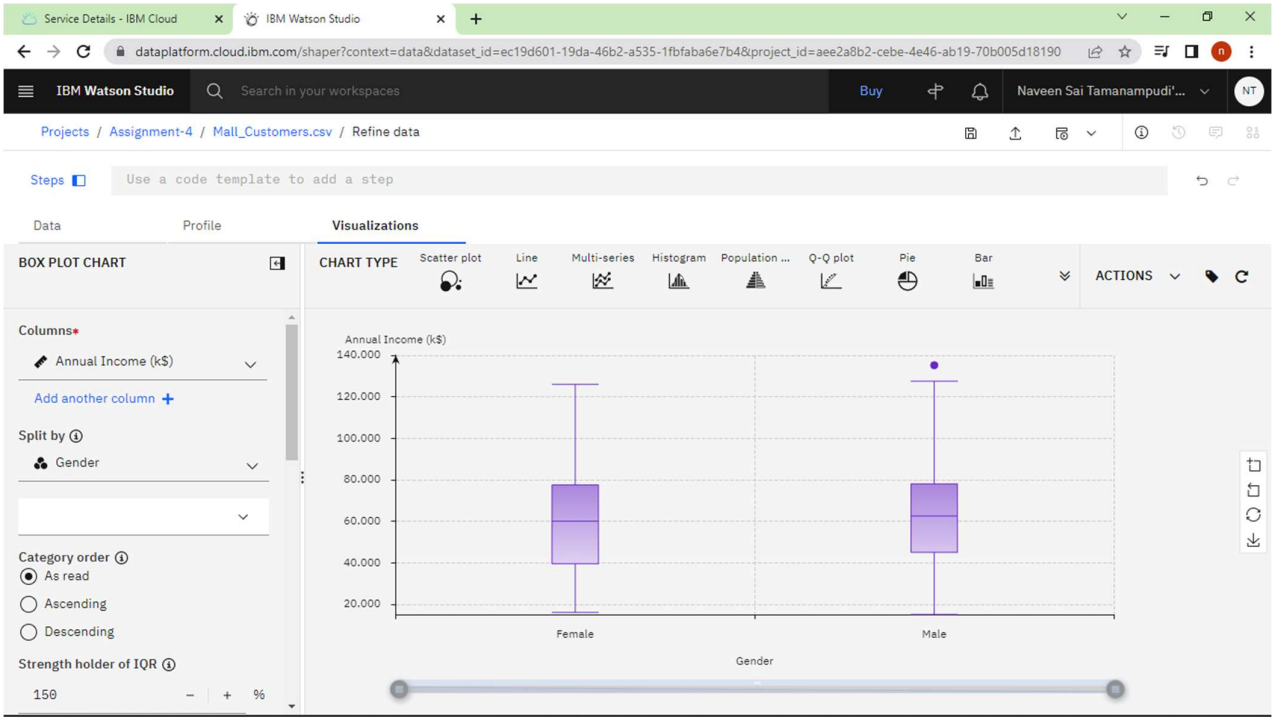
Viewing Scatterplot of Age vs Spending Score(1-100):



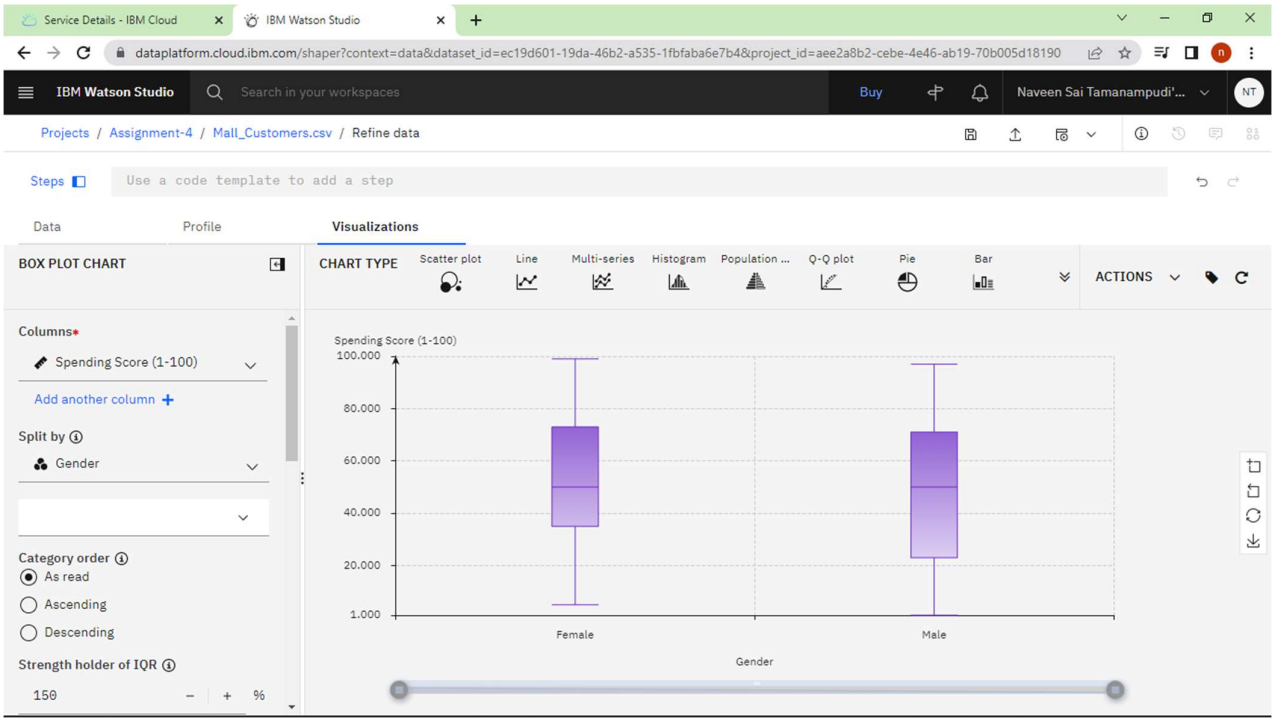
Viewing Scatterplot of Annual Income vs Spending Score(1-100):



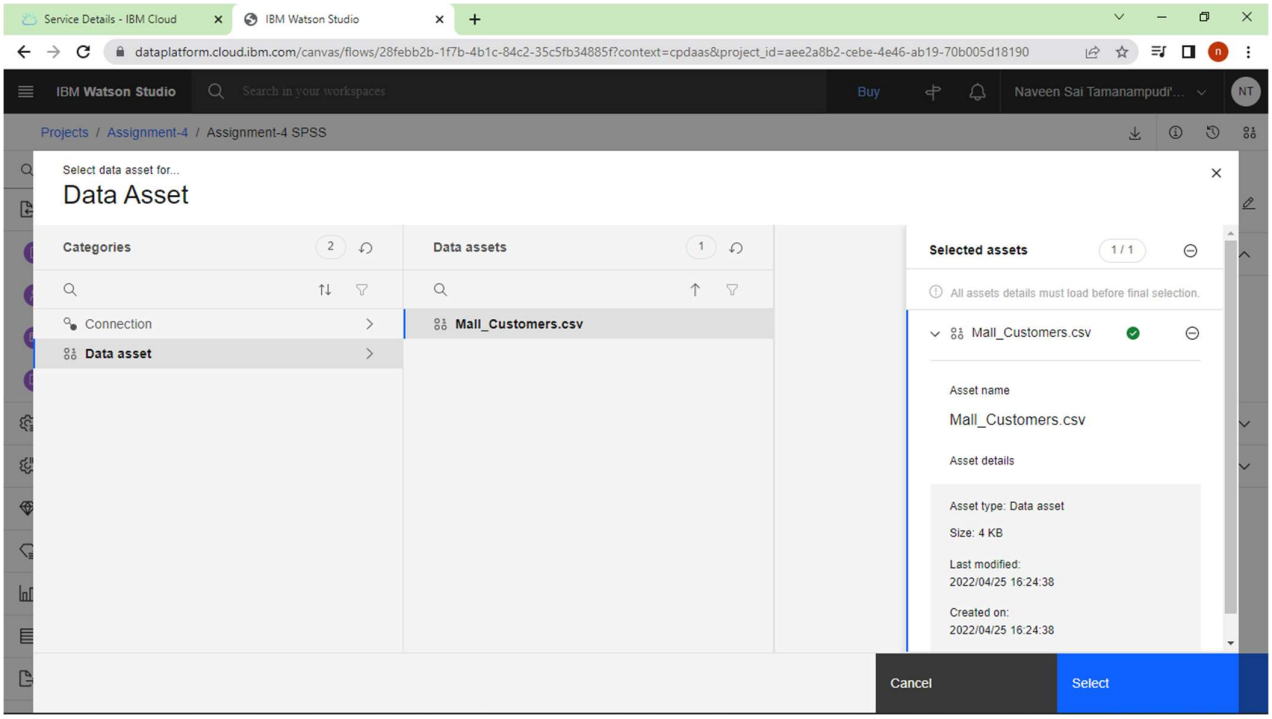
Viewing Boxplot of Annual Income vs Gender:



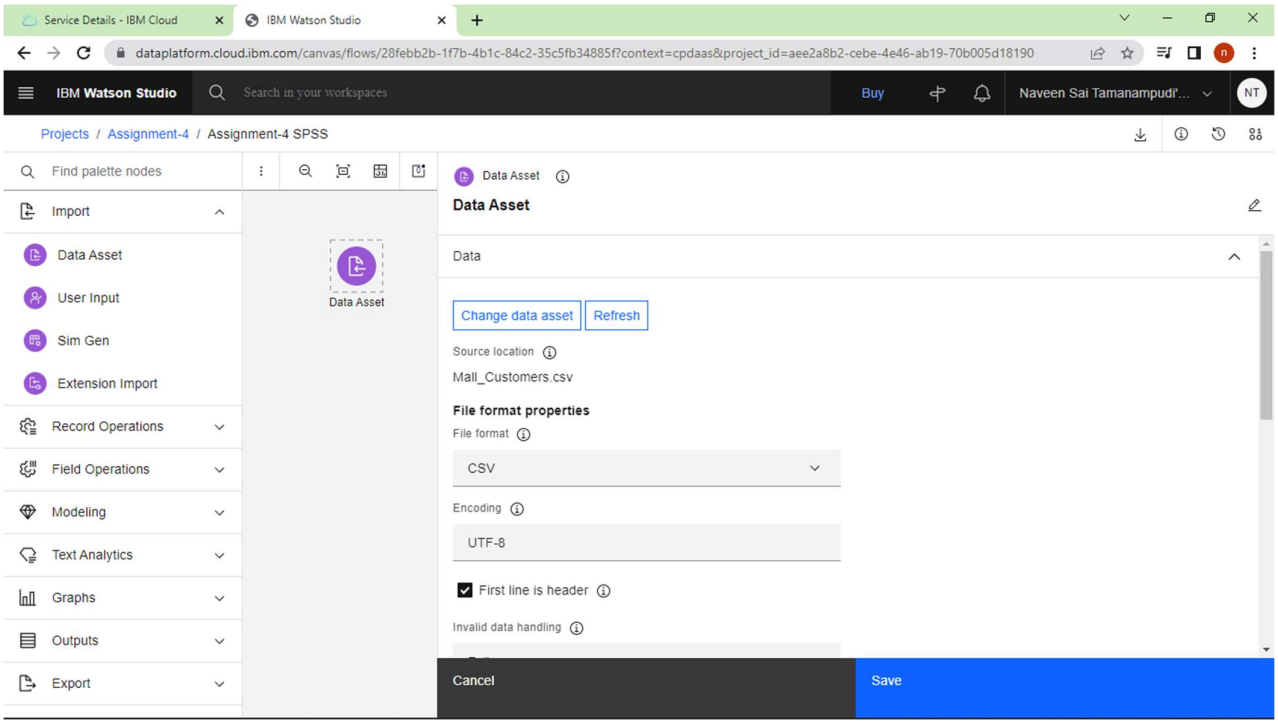
Viewing Boxplot of Spending Score(1-100) vs Gender:



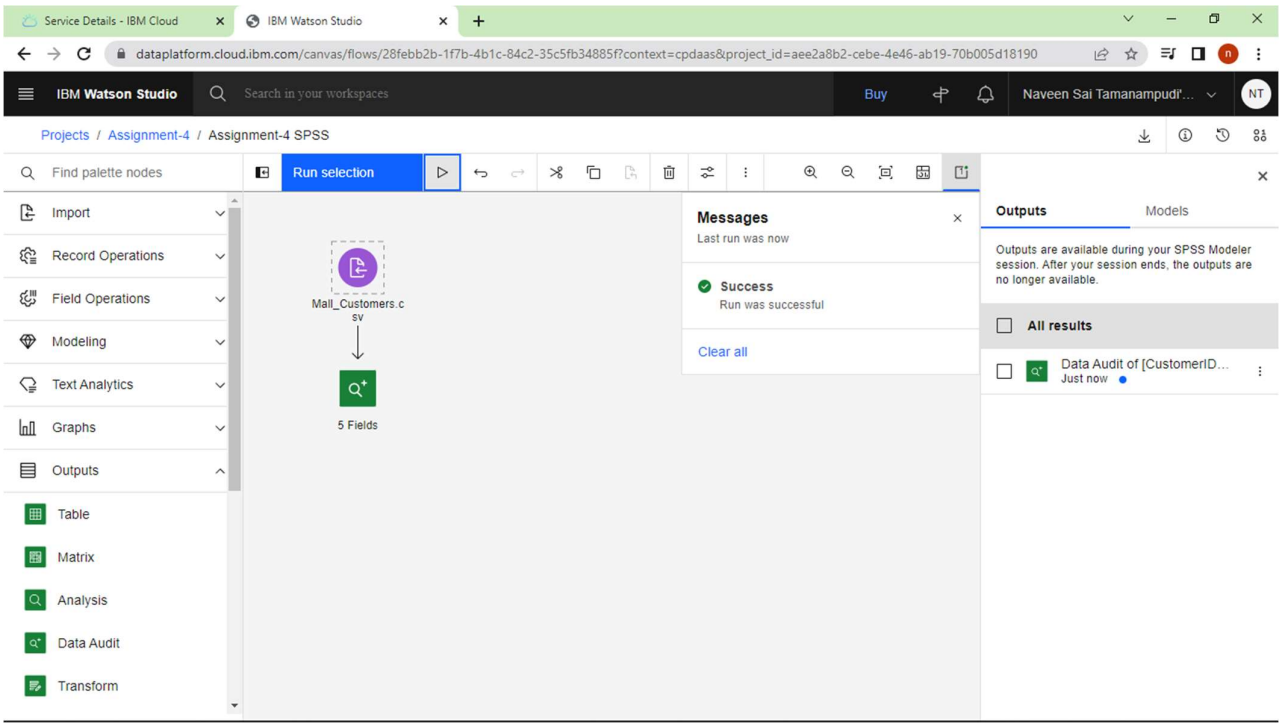
Creating an SPSS Modeler:



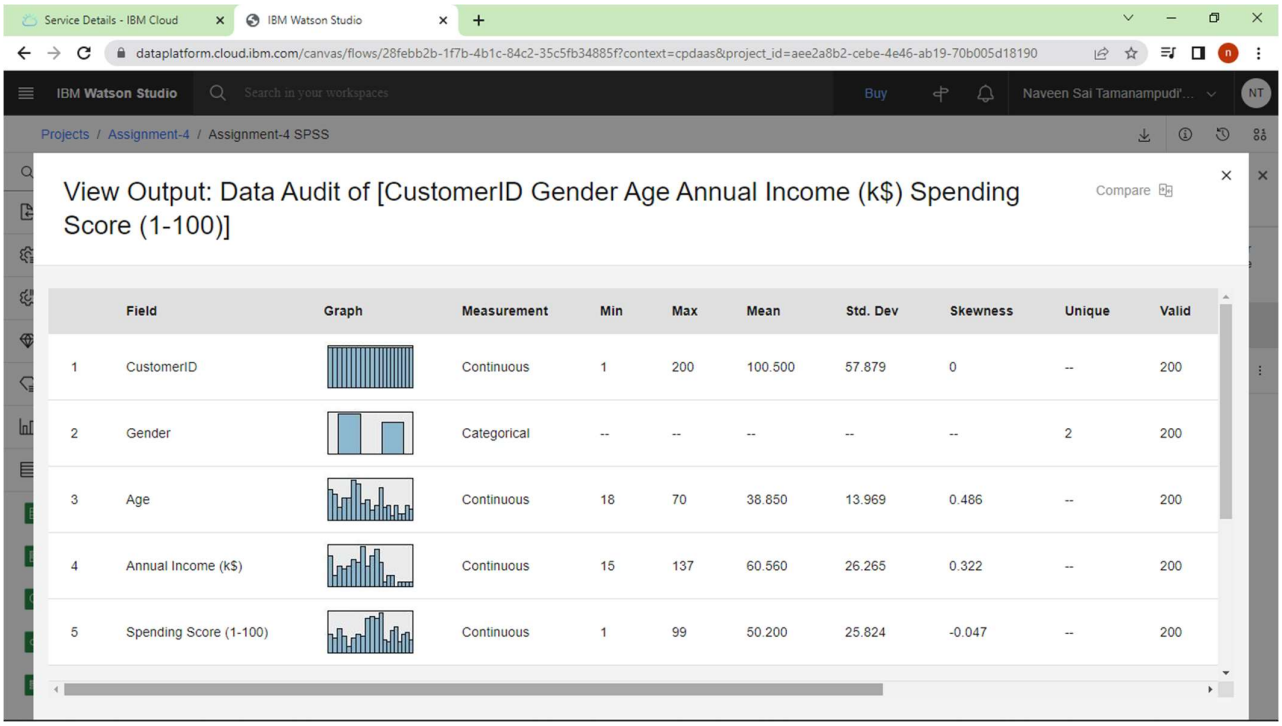
Uploading Dataset to Data Asset Node:



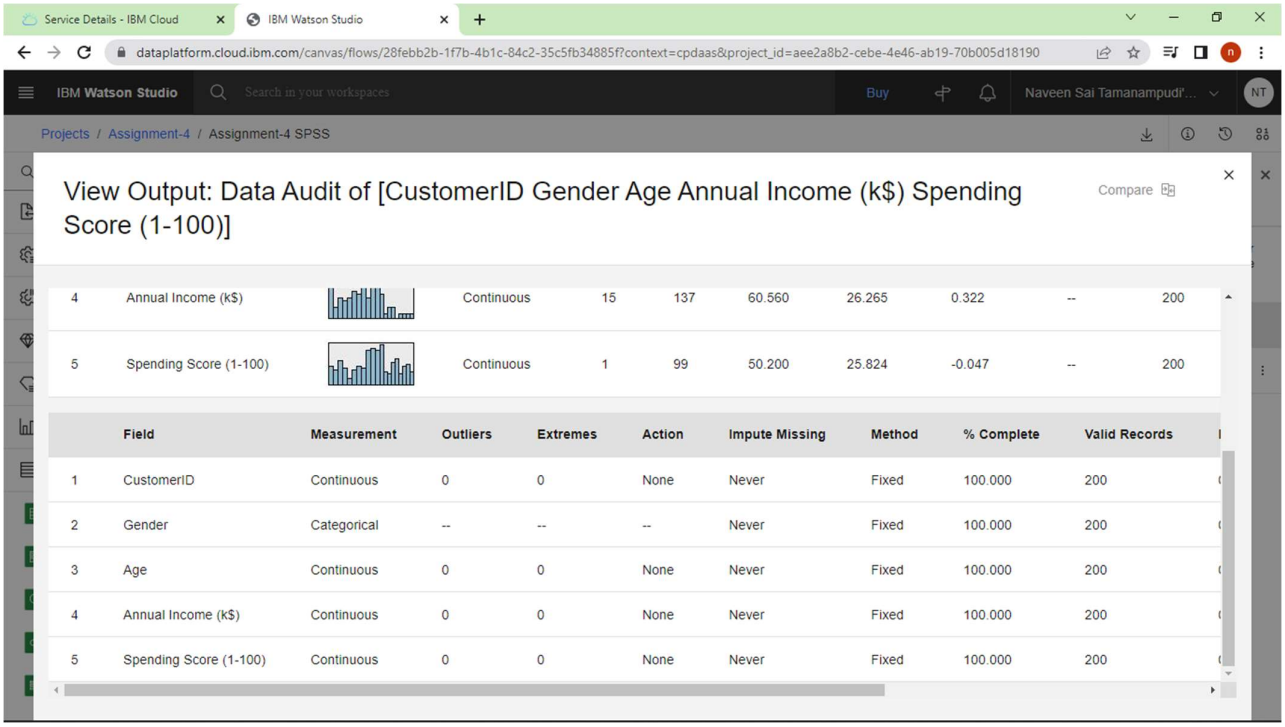
Creating Data Audit Node:



Output from Data Audit:

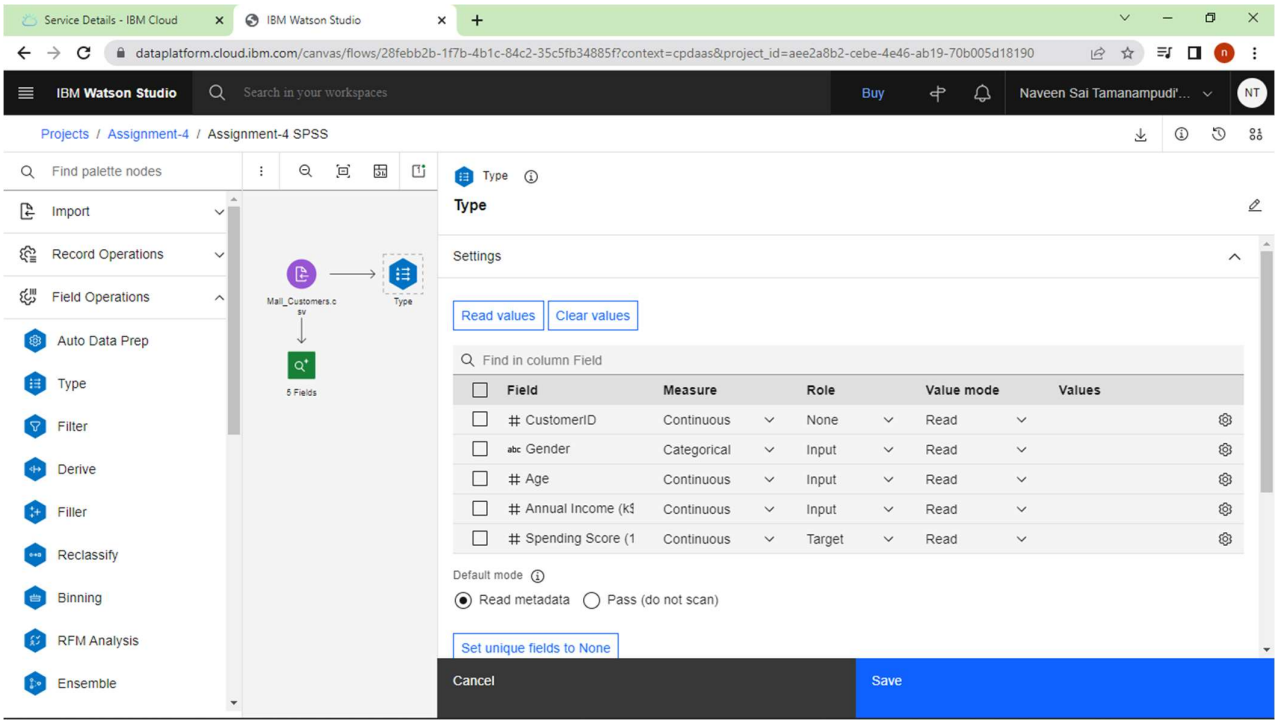


	Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	CustomerID		Continuous	1	200	100.500	57.879	0	--	200
2	Gender		Categorical	--	--	--	--	--	2	200
3	Age		Continuous	18	70	38.850	13.969	0.486	--	200
4	Annual Income (k$)		Continuous	15	137	60.560	26.265	0.322	--	200
5	Spending Score (1-100)		Continuous	1	99	50.200	25.824	-0.047	--	200

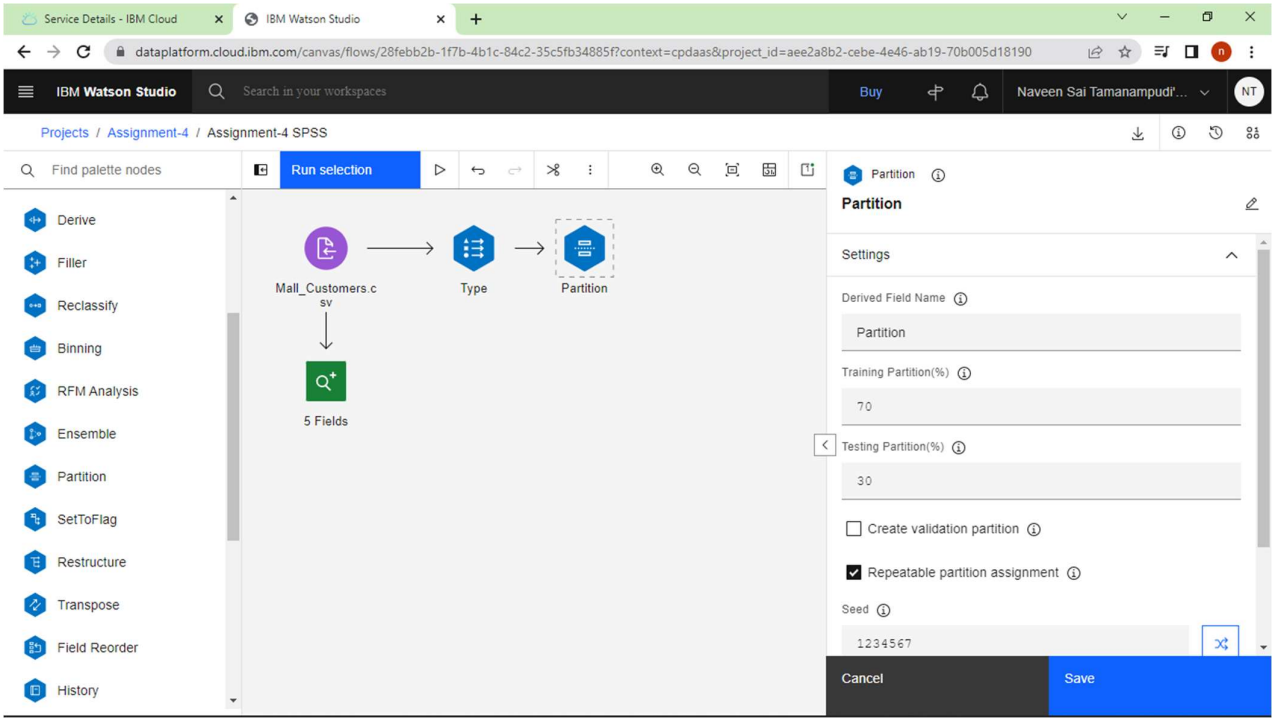


	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records
1	CustomerID	Continuous	0	0	None	Never	Fixed	100.000	200
2	Gender	Categorical	--	--	--	Never	Fixed	100.000	200
3	Age	Continuous	0	0	None	Never	Fixed	100.000	200
4	Annual Income (k$)	Continuous	0	0	None	Never	Fixed	100.000	200
5	Spending Score (1-100)	Continuous	0	0	None	Never	Fixed	100.000	200

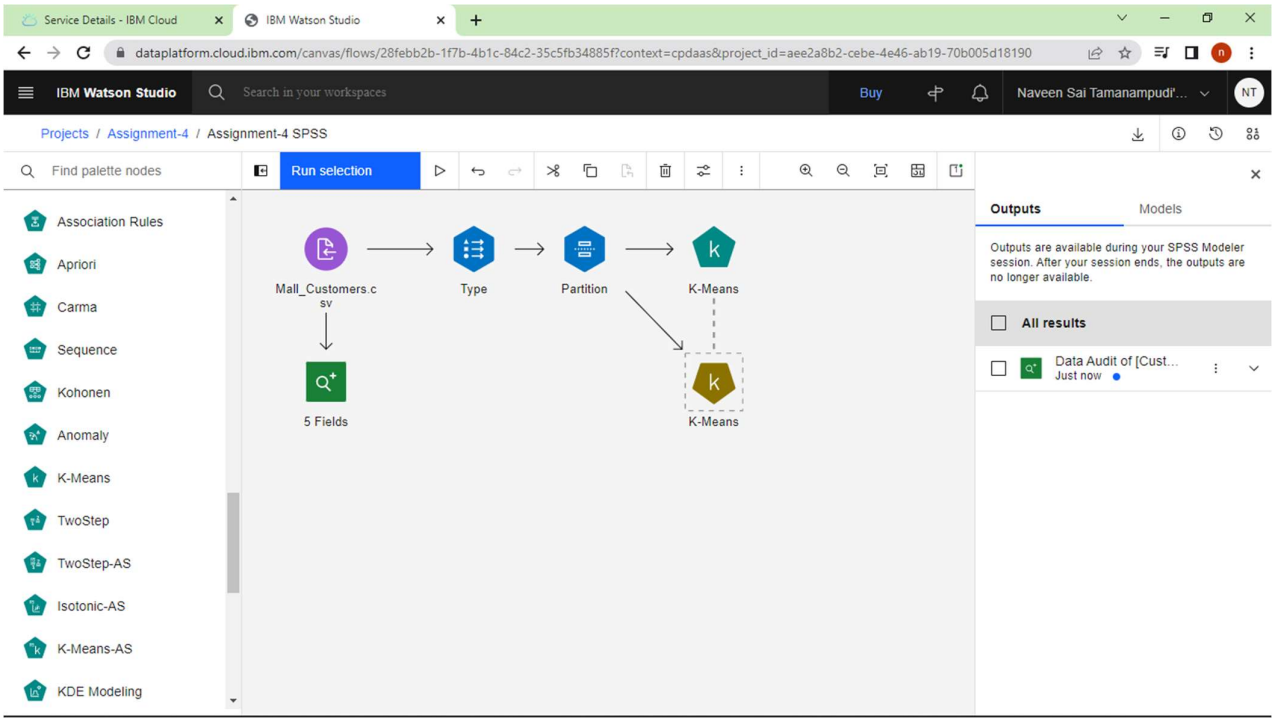
Creating a Type Node and setting Inputs, Targets:



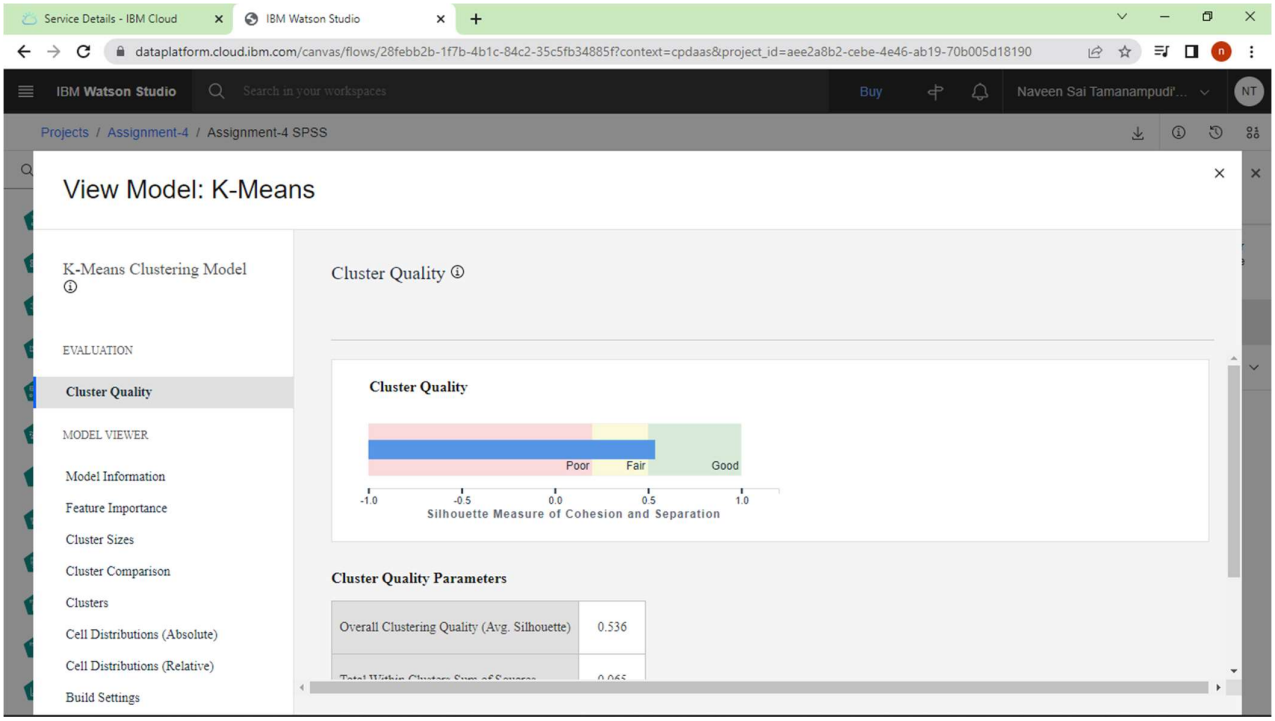
Creating a Partition Node with 70:30 split:

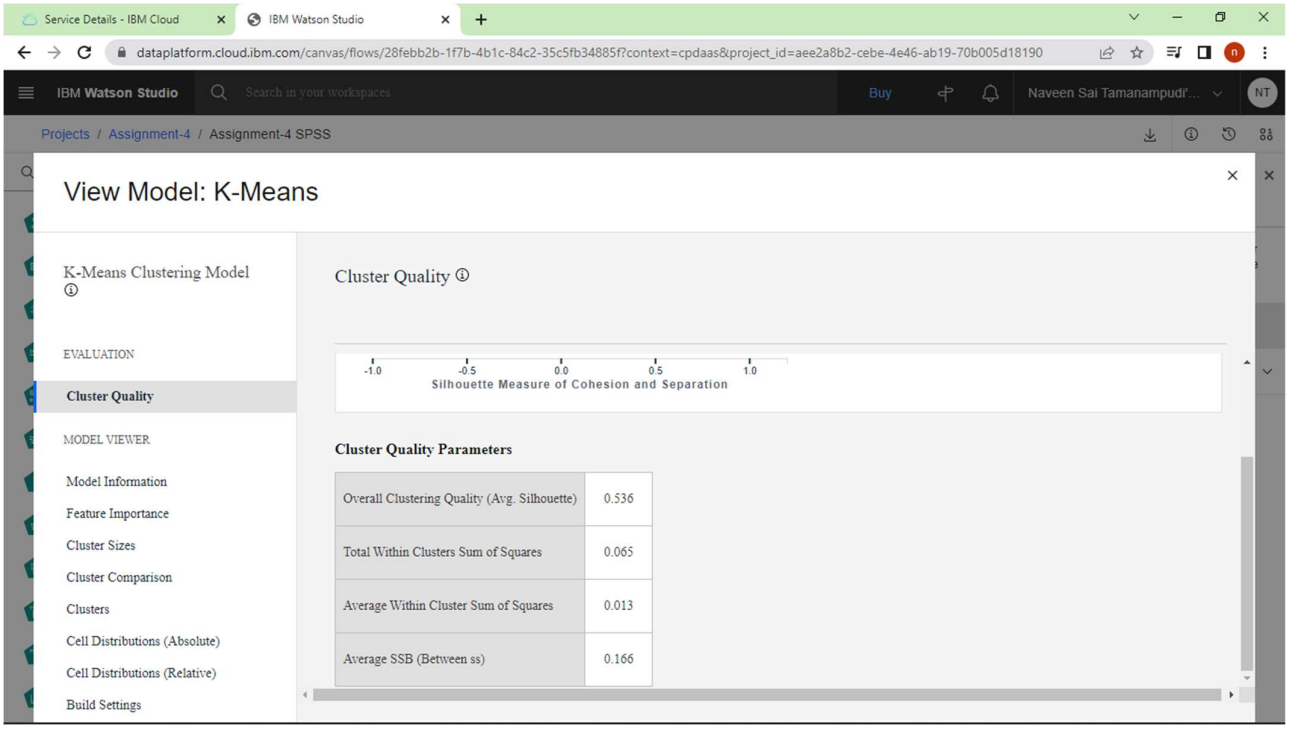


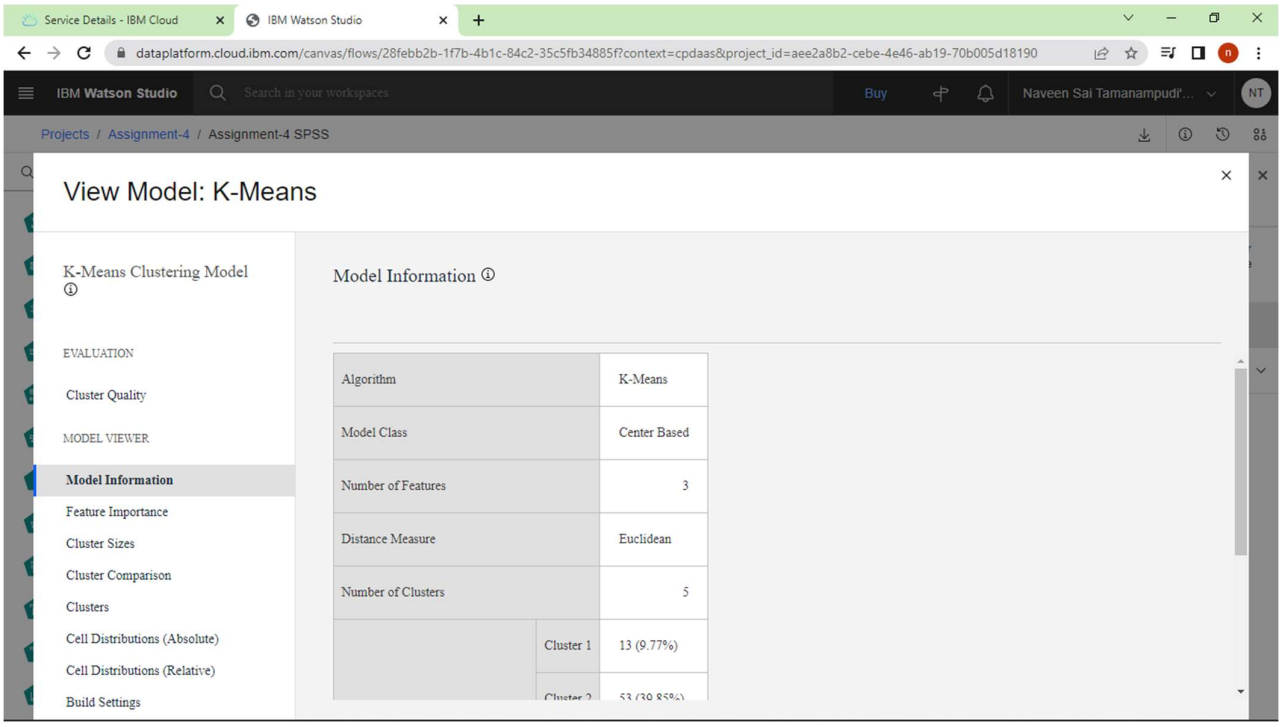
Creating an K-Means Clustering Model:

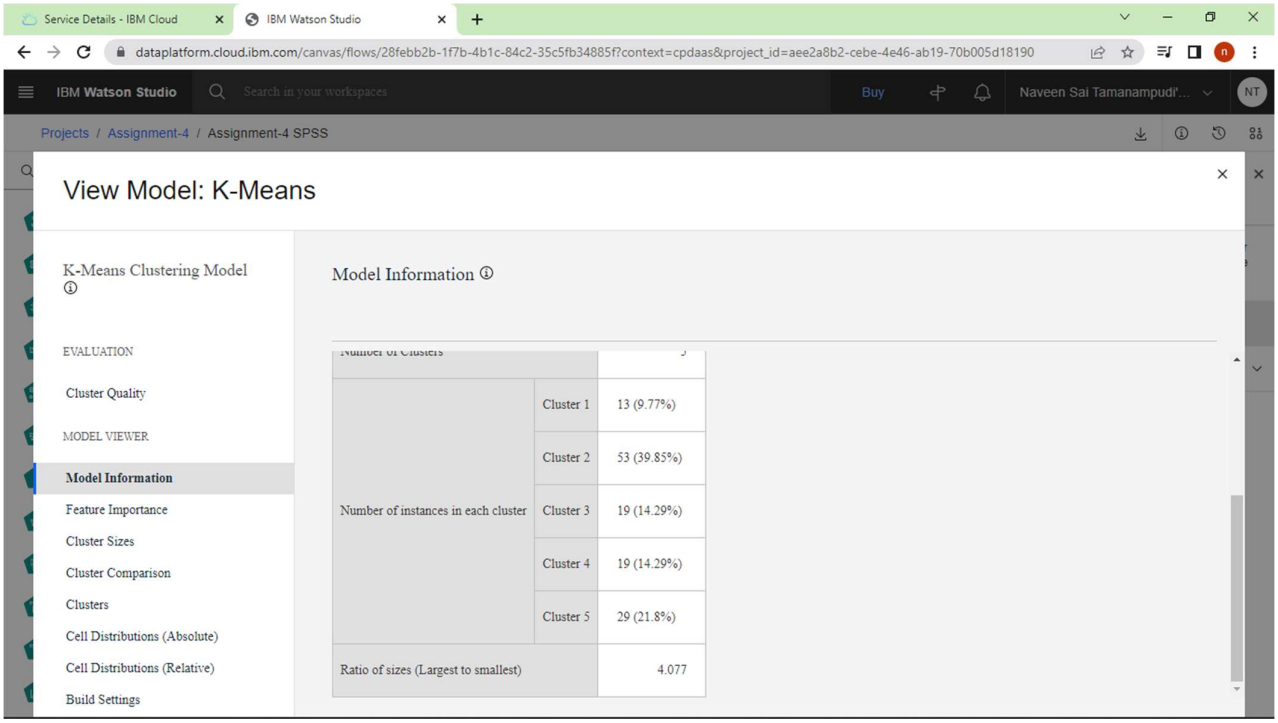


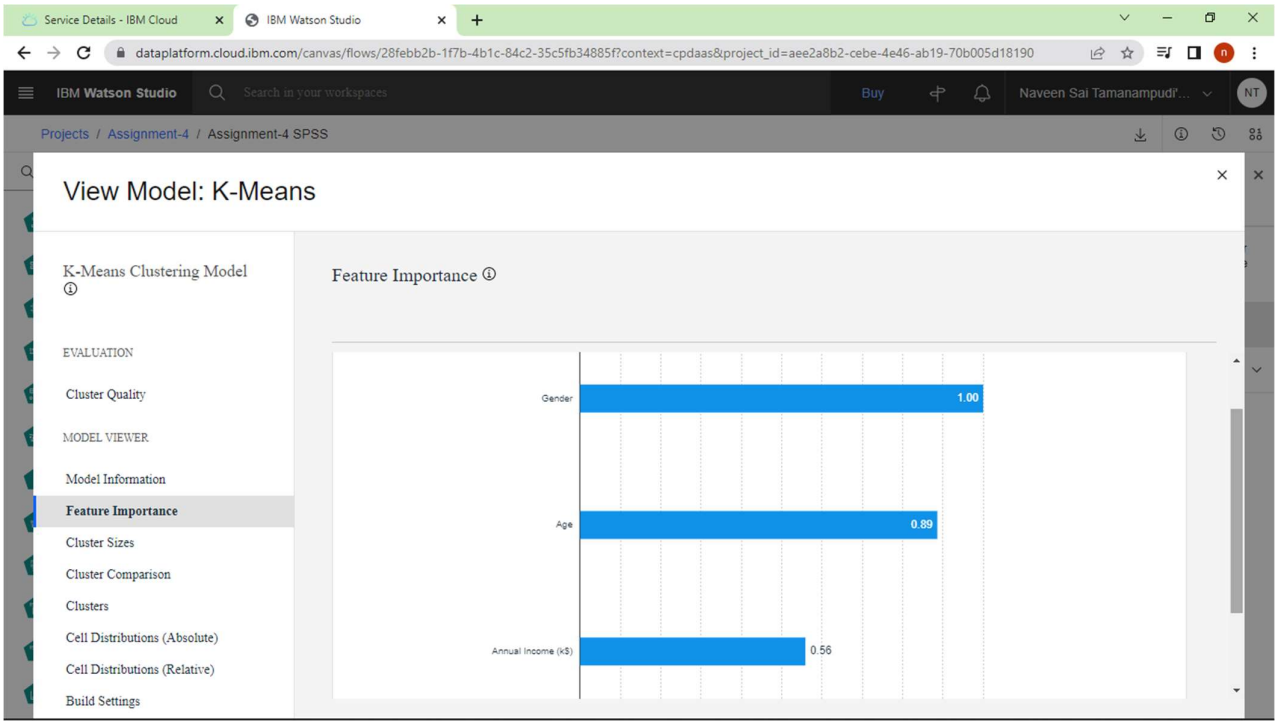
Output:

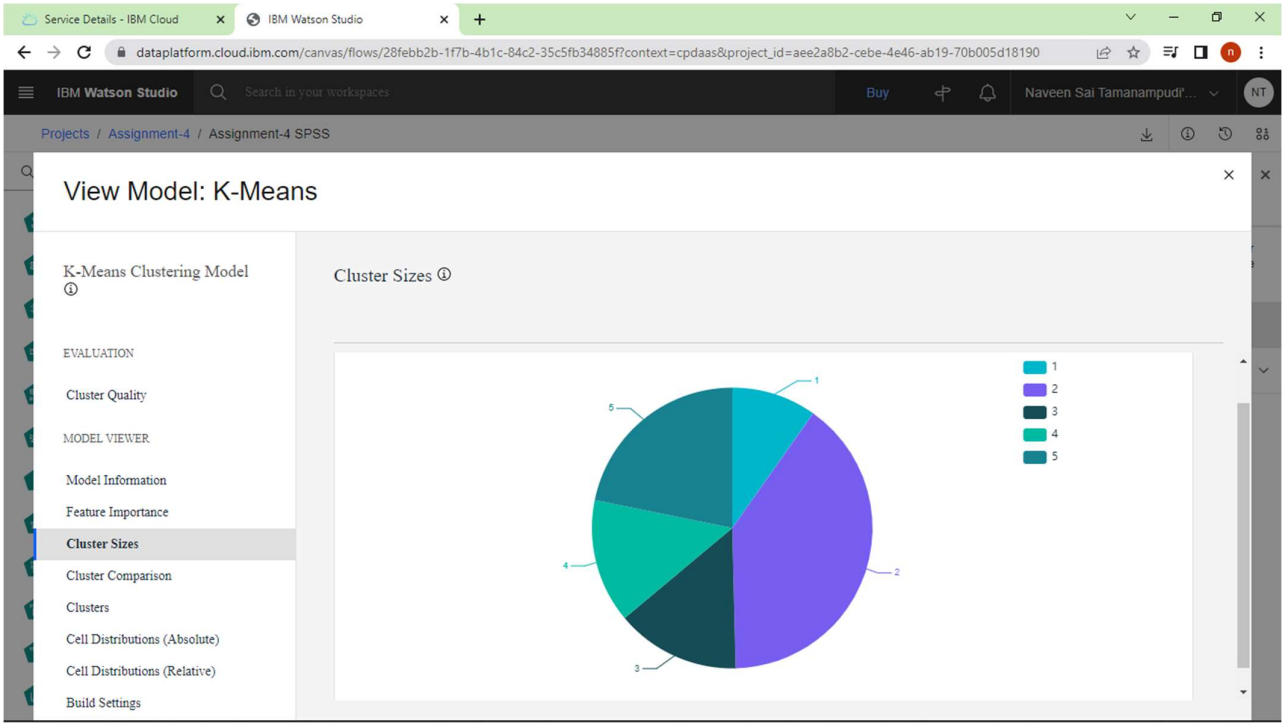


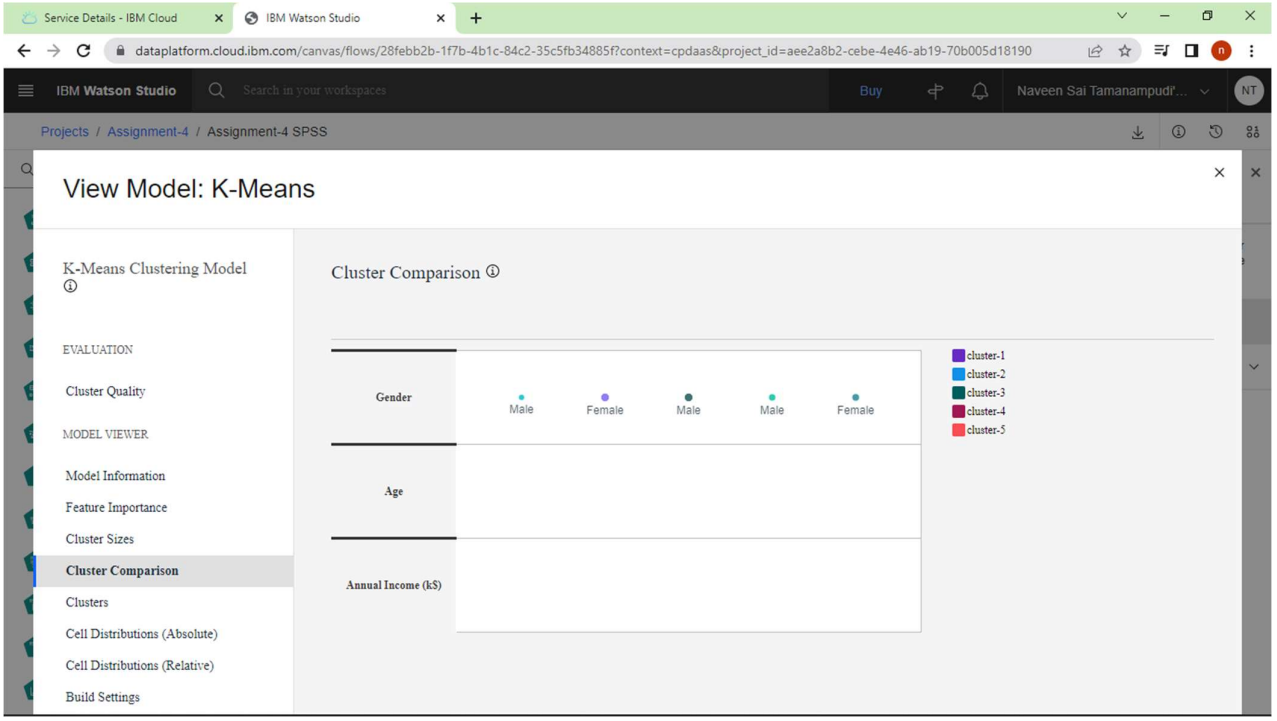

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Cluster Quality
-1.0
-0.5
0.0
0.5
1.0
Silhouette Measure of Cohesion and Separation
Cluster Quality Parameters
Overall Clustering Quality (Avg. Silhouette) 0.536
Total Within Clusters Sum of Squares 0.065
Average Within Cluster Sum of Squares 0.013
Average SSB (Between ss) 0.166

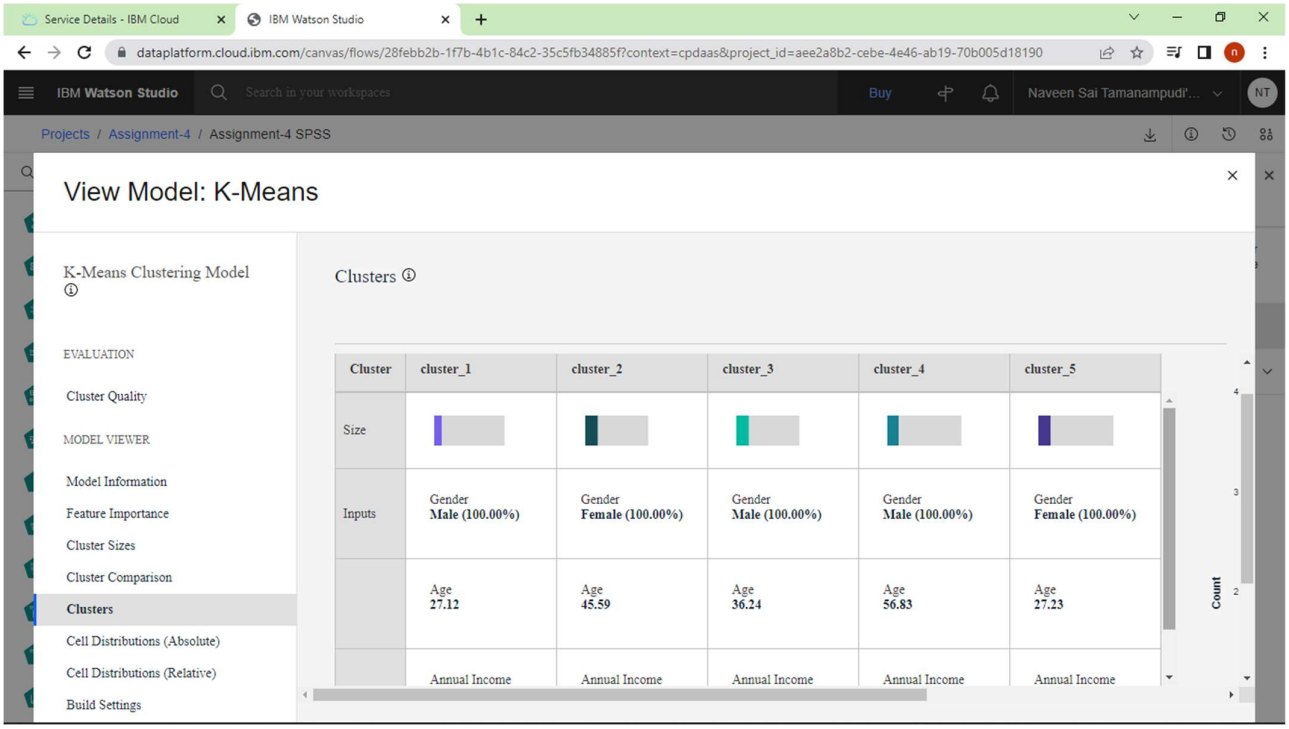

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Model Information
Algorithm K-Means
Model Class Center Based
Number of Features 3
Distance Measure Euclidean
Number of Clusters 5
Cluster 1 13 (9.77%)

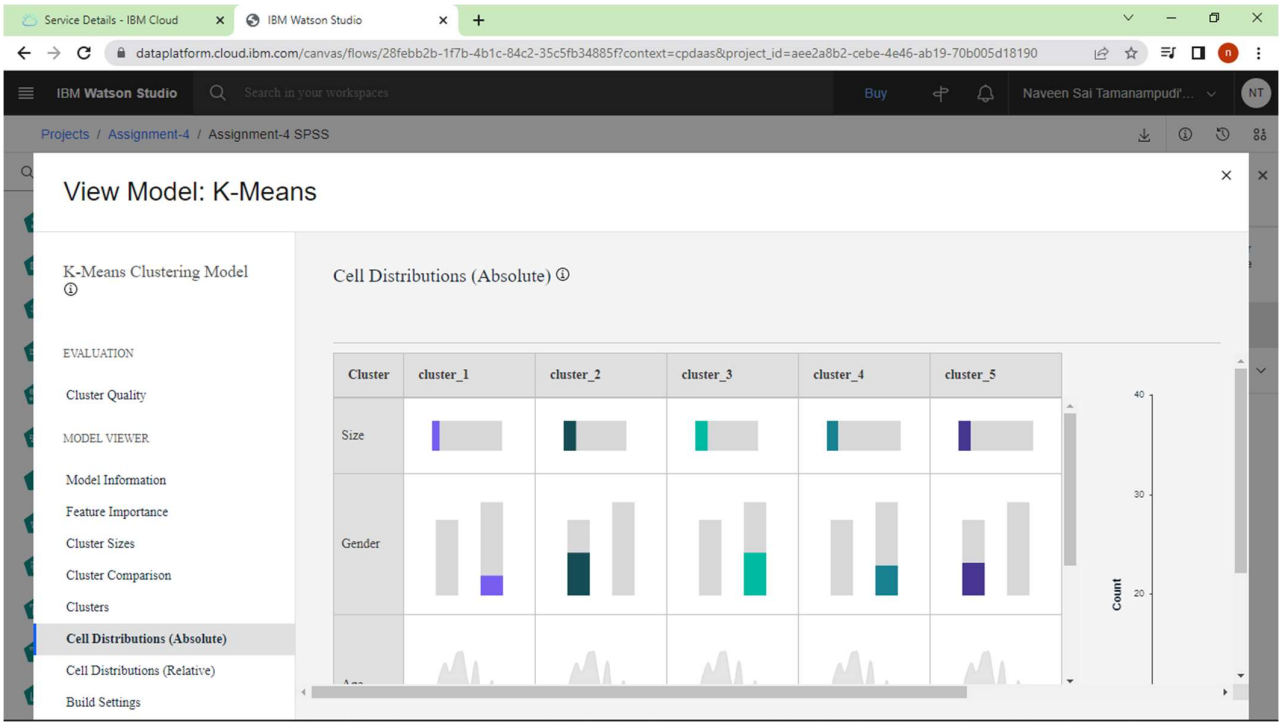

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Model Information
Number of instances in each cluster
Cluster 1
13 (9.77%)
Cluster 2
53 (39.85%)
Cluster 3
19 (14.29%)
Cluster 4
19 (14.29%)
Cluster 5
29 (21.8%)
Ratio of sizes (Largest to smallest)
4.077

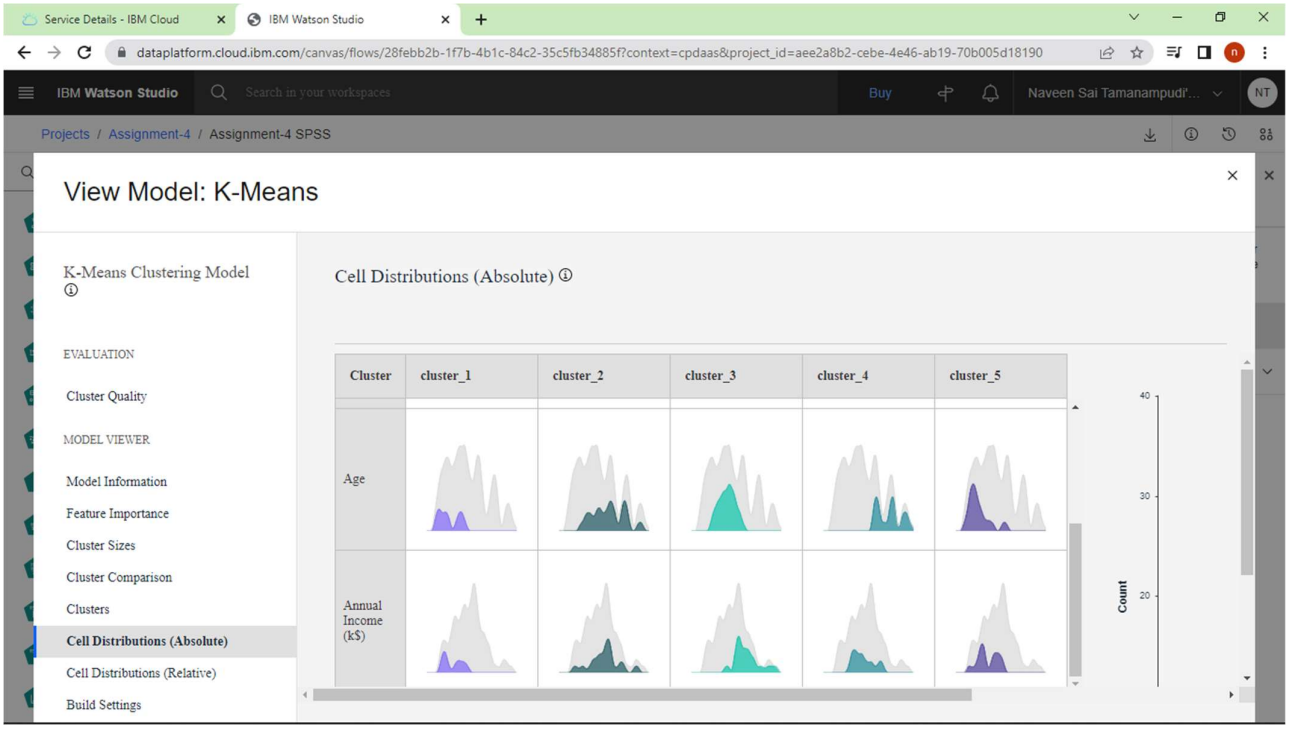

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Feature Importance
Gender
1.00
Age
0.89
Annual Income (k$)
0.56

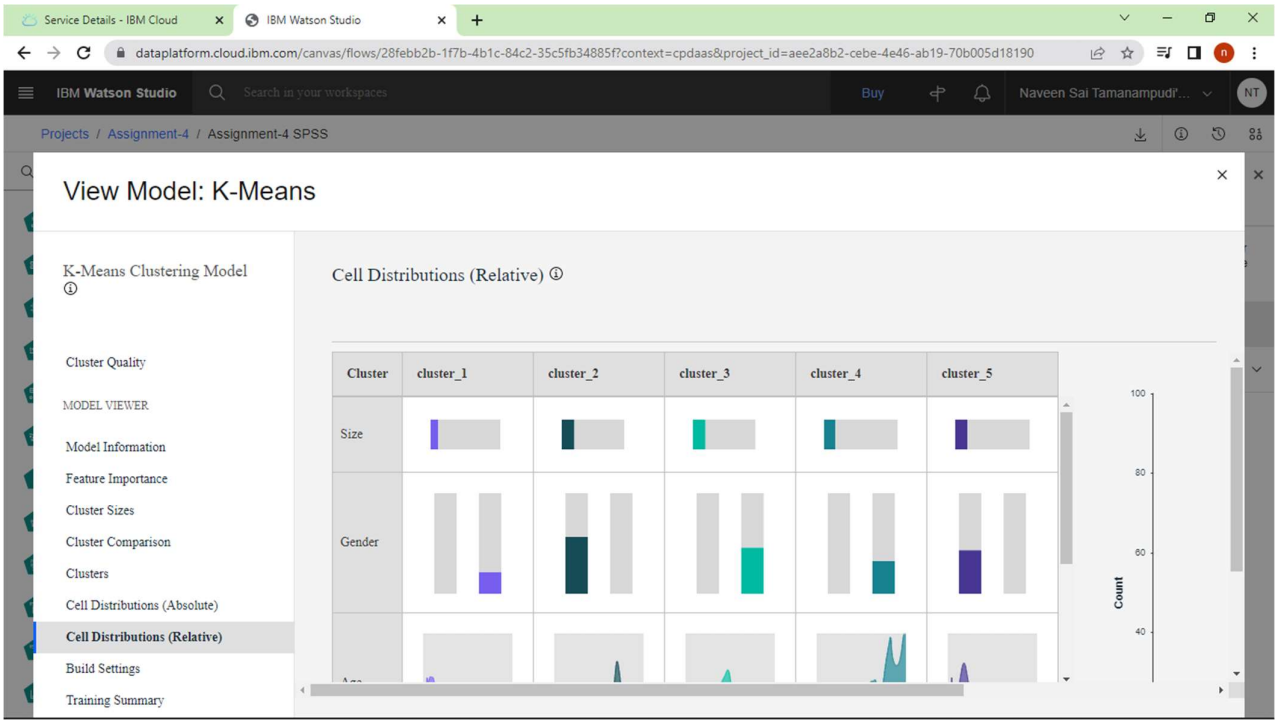

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Cluster Sizes
1
2
3
4
5

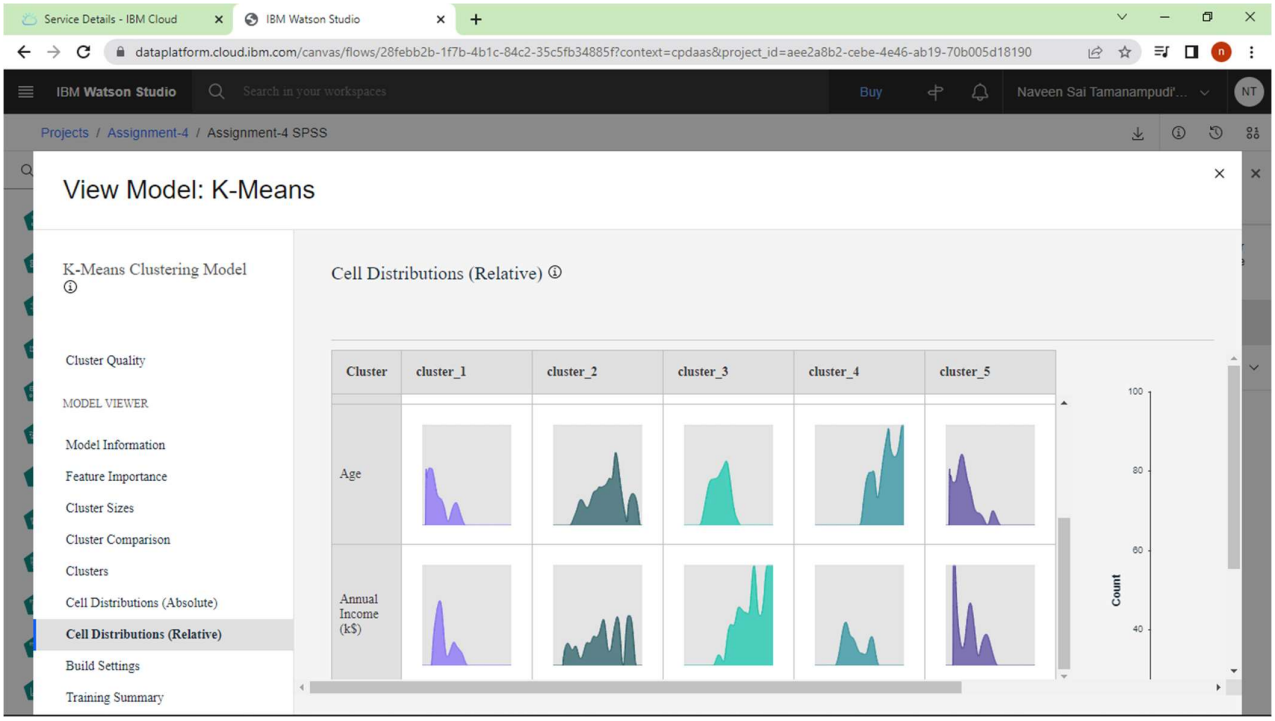

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Cluster Comparison
Gender
Male
Female
Male
Male
Female
Age
Annual Income (k$)
cluster-1
cluster-2
cluster-3
cluster-4
cluster-5

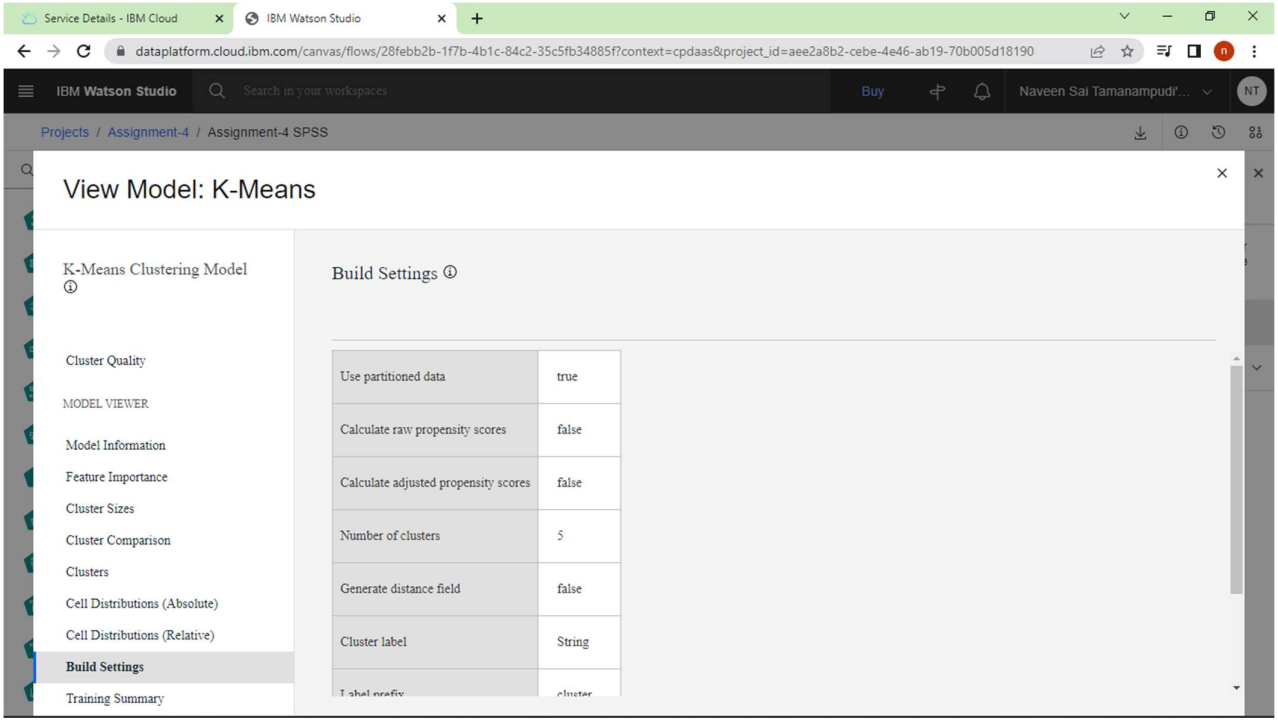

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Clusters
Cluster
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5
Size
Inputs
Gender Male (100.00%)
Gender Female (100.00%)
Gender Male (100.00%)
Gender Male (100.00%)
Gender Female (100.00%)
Age 27.12
Age 45.59
Age 36.24
Age 56.83
Age 27.23
Annual Income
Count

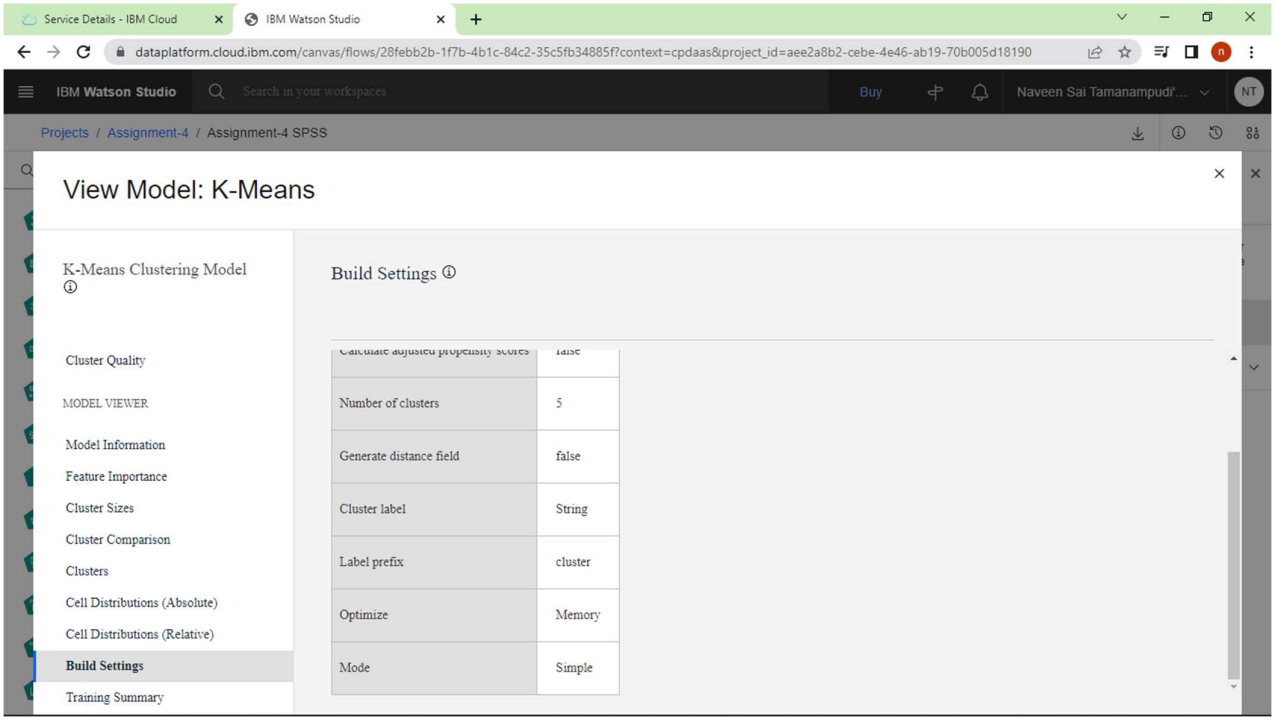

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Cell Distributions (Absolute)
Cluster
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5
Size
Gender
Count

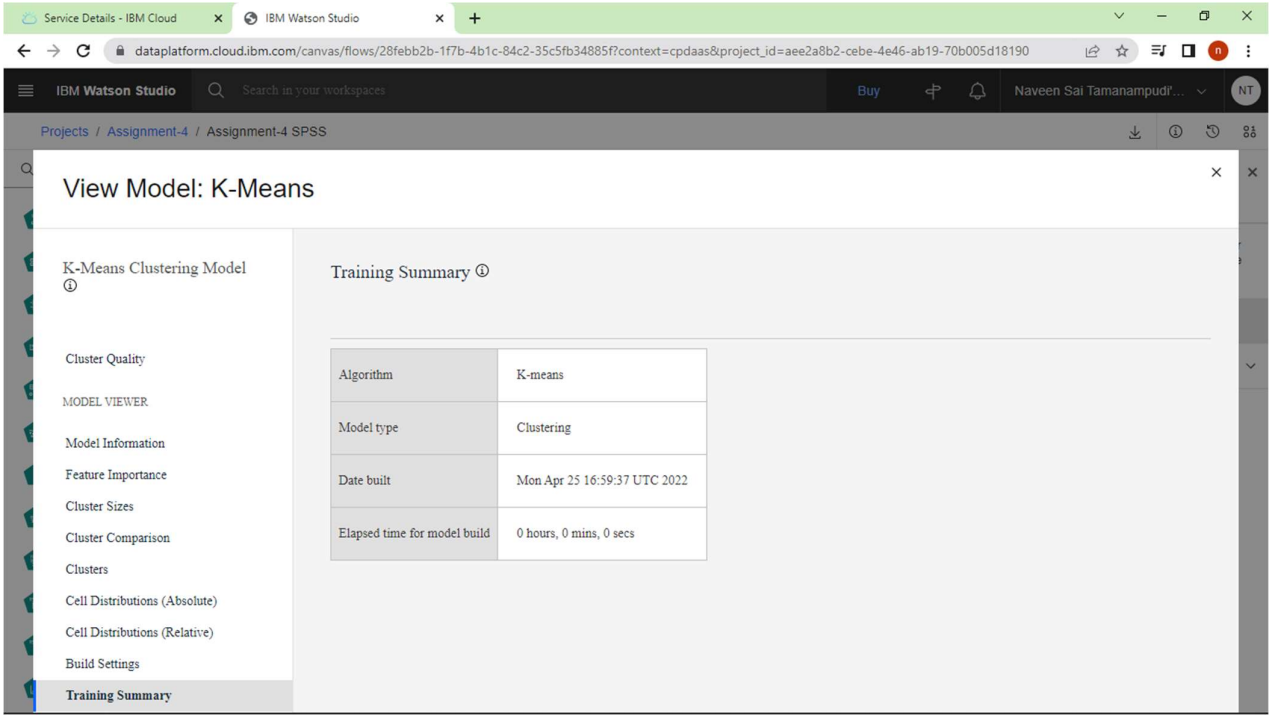

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
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
Cell Distributions (Absolute)
Cluster
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5
Age
Annual Income (k$)
Count
40
30
20


Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
View Model: K-Means
K-Means Clustering Model
Cluster Quality
MODEL VIEWER
Model Information
Feature Importance
Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary
Cell Distributions (Relative)
Cluster
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5
Size
Gender
Count
100
80
60
40

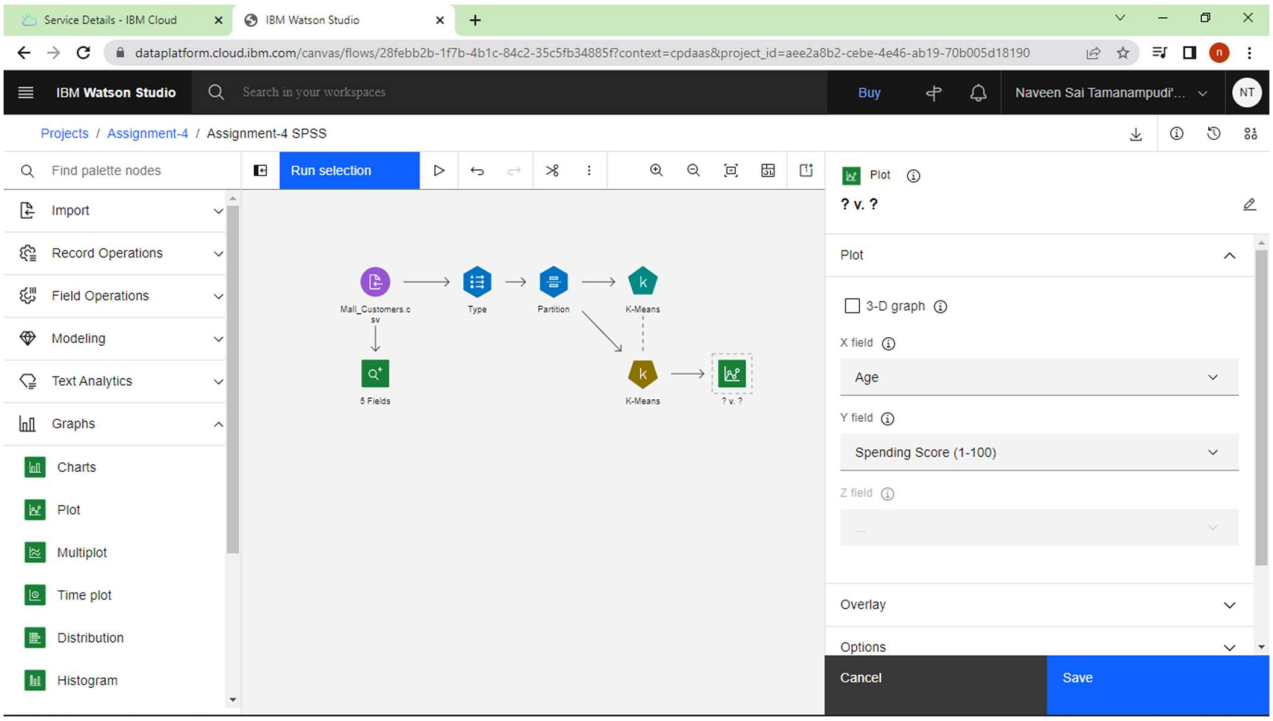

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
View Model: K-Means
K-Means Clustering Model
Cluster Quality
MODEL VIEWER
Model Information
Feature Importance
Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary
Cell Distributions (Relative)
Cluster
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5
Age
Annual Income (k$)
Count
100
80
60
40


Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
View Model: K-Means
K-Means Clustering Model
Cluster Quality
MODEL VIEWER
Model Information
Feature Importance
Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary
Build Settings
Use partitioned data
true
Calculate raw propensity scores
false
Calculate adjusted propensity scores
false
Number of clusters
5
Generate distance field
false
Cluster label
String
Label prefix
cluster

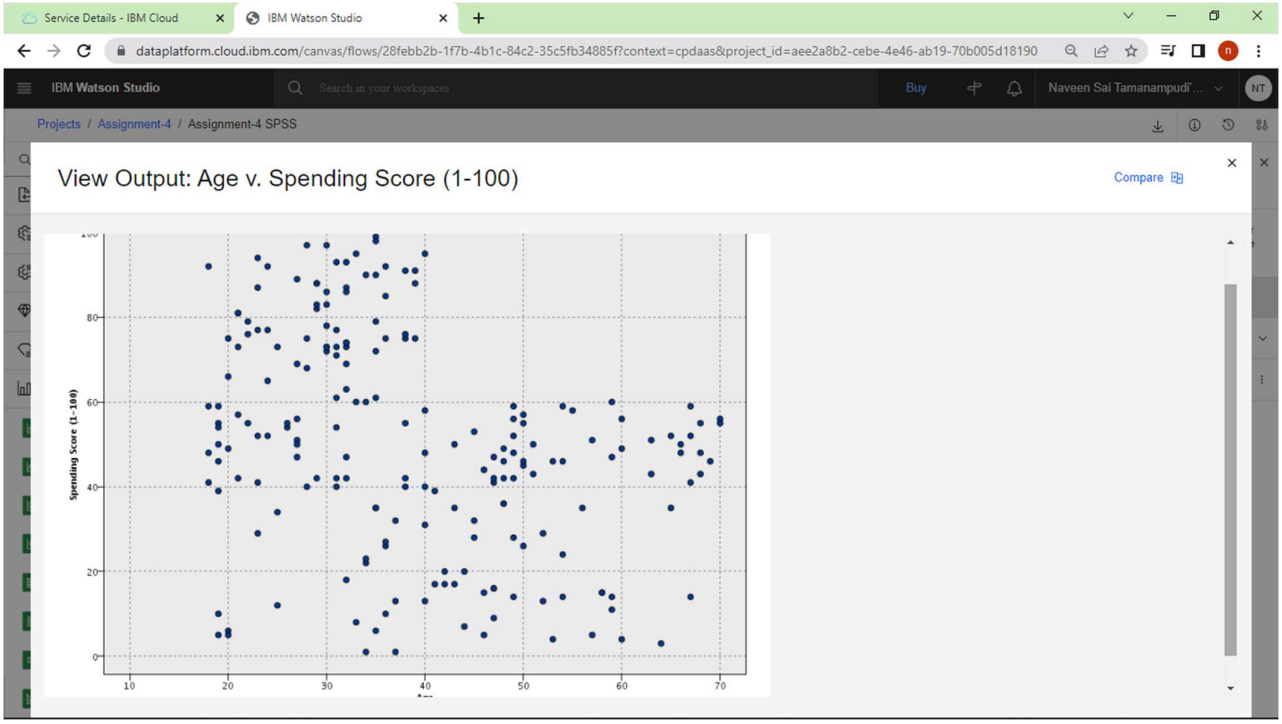

Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
View Model: K-Means
K-Means Clustering Model
Cluster Quality
MODEL VIEWER
Model Information
Feature Importance
Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary
Build Settings
Number of clusters 5
Generate distance field false
Cluster label String
Label prefix cluster
Optimize Memory
Mode Simple


Service Details - IBM Cloud
IBM Watson Studio
dataplatform.cloud.ibm.com/canvas/flows/28febb2b-1f7b-4b1c-84c2-35c5fb34885f?context=cpdaas&project_id=aee2a8b2-cebe-4e46-ab19-70b005d18190
IBM Watson Studio
Search in your workspaces
Buy
Naveen Sai Tamanampudi'...
Projects / Assignment-4 / Assignment-4 SPSS
View Model: K-Means
K-Means Clustering Model
Cluster Quality
MODEL VIEWER
Model Information
Feature Importance
Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary
Training Summary
Algorithm K-means
Model type Clustering
Date built Mon Apr 25 16:59:37 UTC 2022
Elapsed time for model build 0 hours, 0 mins, 0 secs

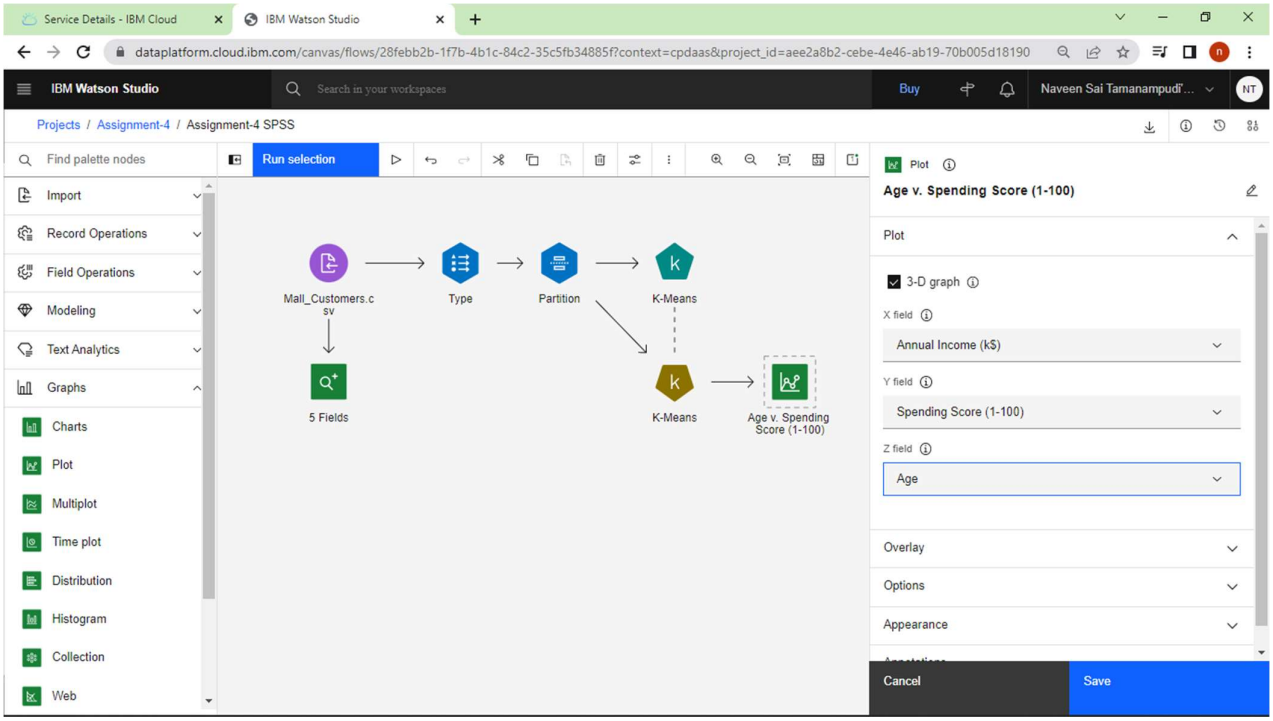
Creating a Plot Node with Age vs Spending Score(1-100):



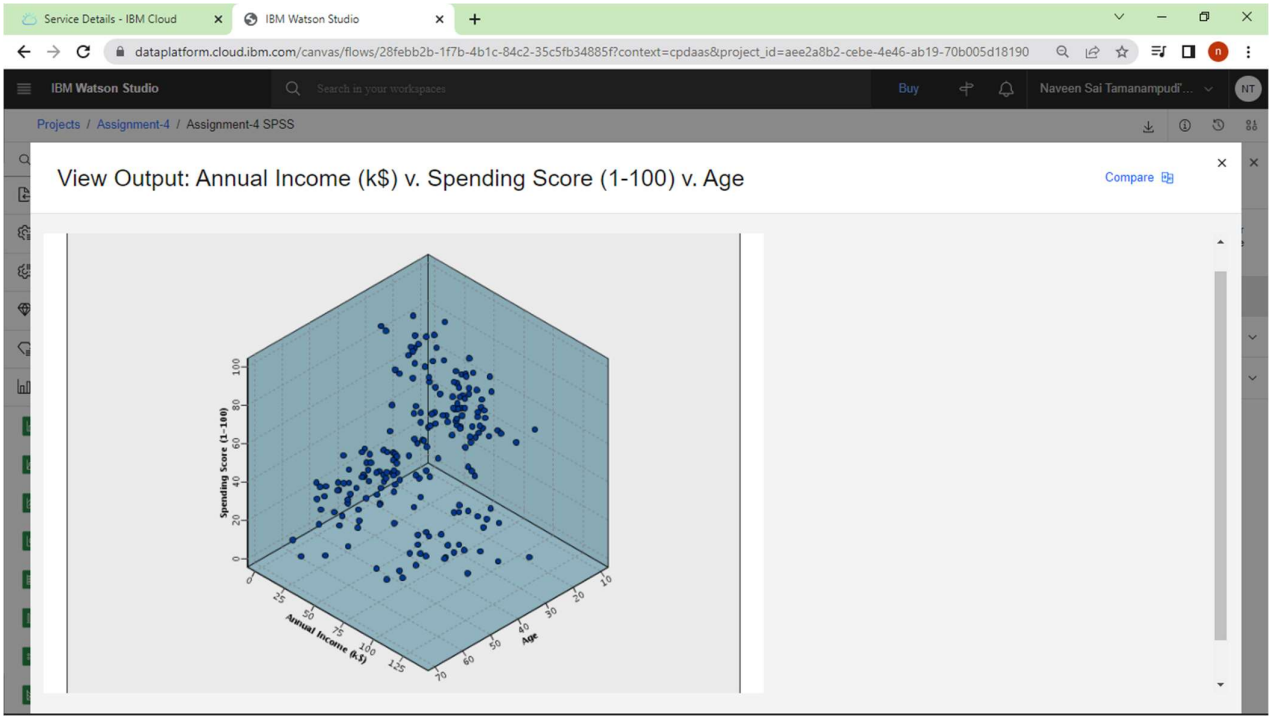
Viewing Plot:



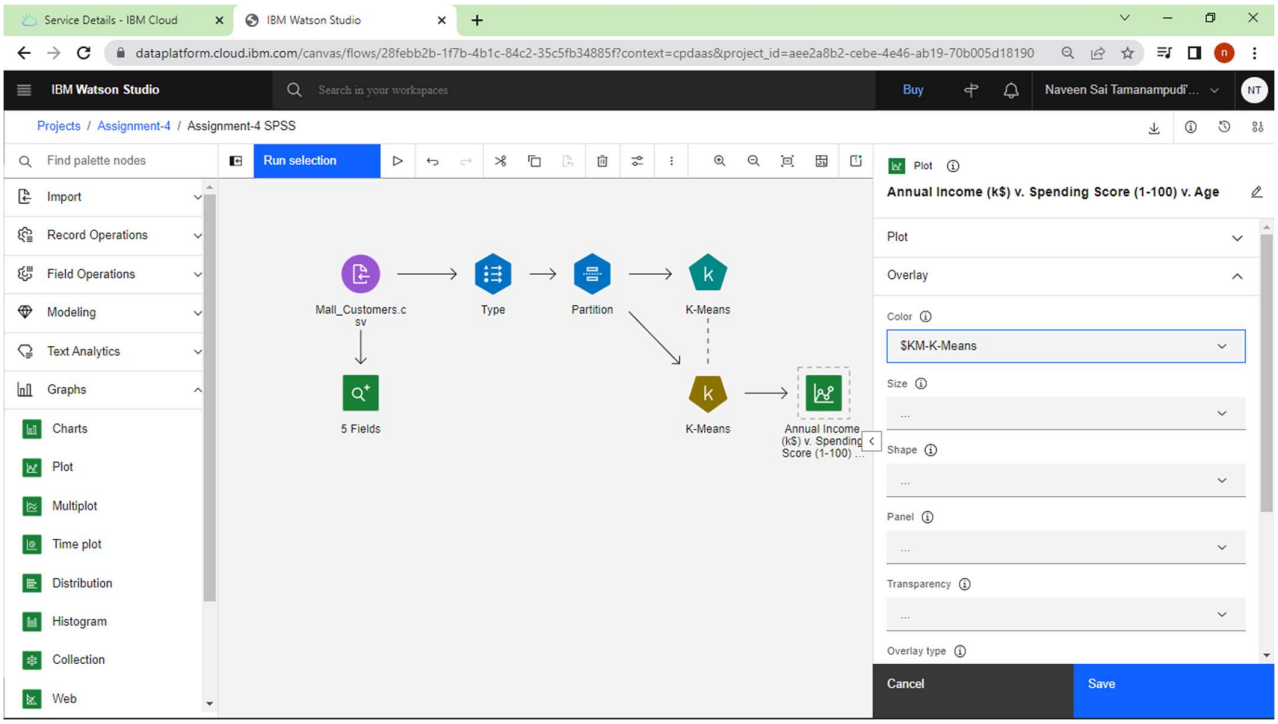
Creating a Plot Node with Annual Income vs Spending Score(1-100) vs Age:



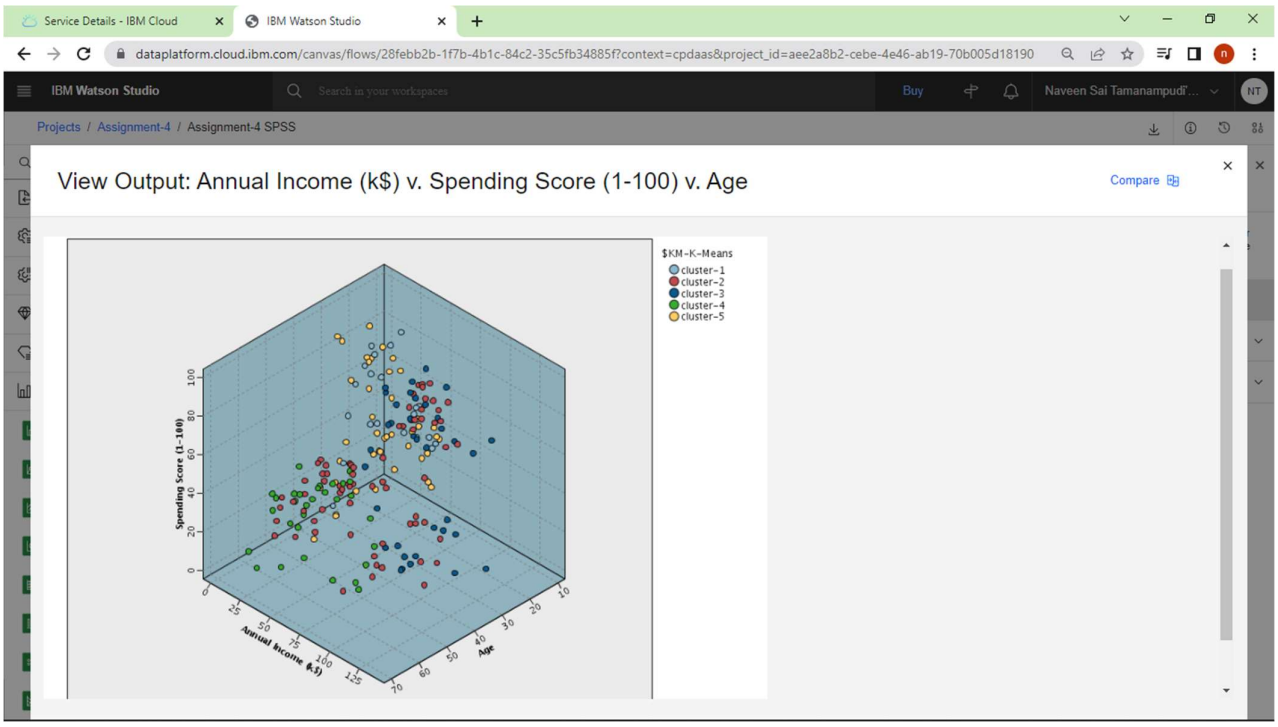
Output:



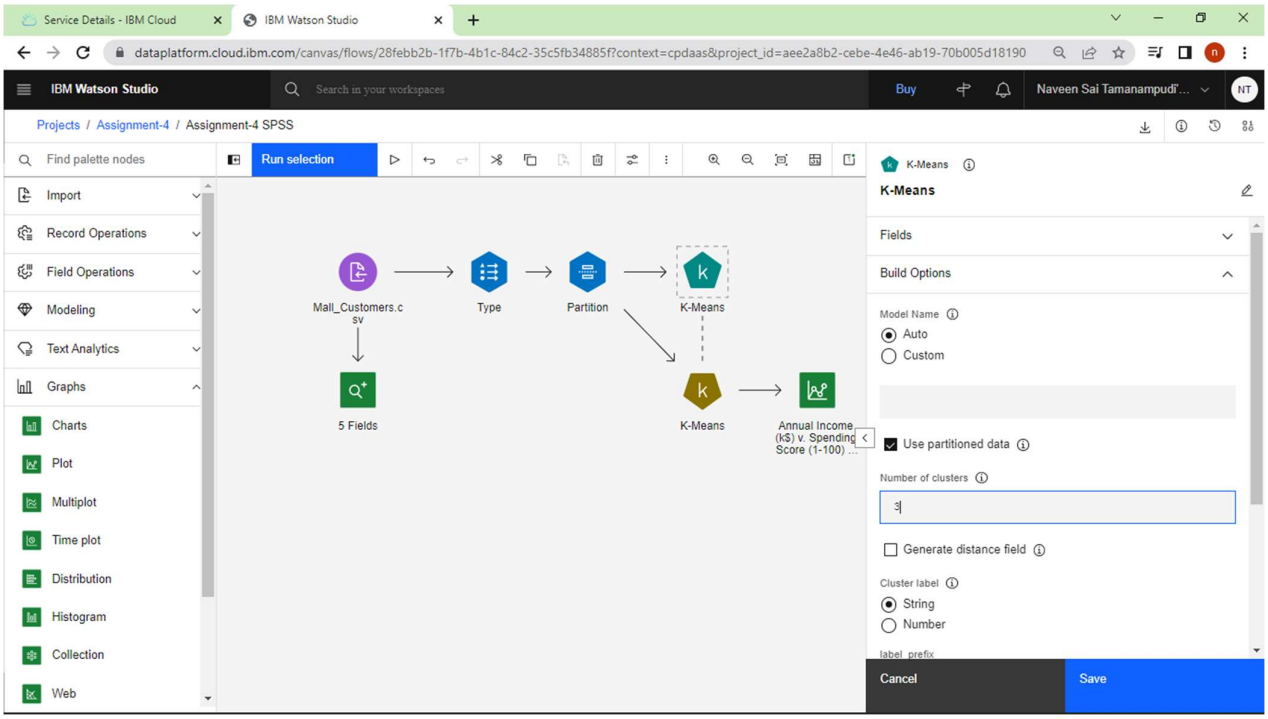
Assigning Colors:



Output:



Changing Number of clusters to 3:



Output: