

Assignment 2

Dataset: Challengers.csv

Link: <https://dataplatform.cloud.ibm.com/projects/8965589a-daec-44f2-ad68-9a6d18b9e31f/assets?context=cpdaas>

The screenshot shows the 'Create a project' page in IBM Watson Studio. It features two main sections: 'Create an empty project' and 'Create a project from a sample or file'. The 'Create an empty project' section includes a circular icon with a plus sign and three boxes labeled A, B, and C, followed by a brief description and a 'NEW AutoAI experiment tool' button. The 'Create a project from a sample or file' section includes a circular icon with a plus sign and a document, followed by a brief description and a 'USE TO' section with three items: 'Prepare and visualize data', 'Analyze data in notebooks', and 'Train models'. Below these sections is a URL bar containing <https://dataplatform.cloud.ibm.com/projects/new-project?context=cpdaas#>.

Figure: Creating a project in IBM Watson Studio

The screenshot shows the 'challengers.csv' dataset in the 'Predictive Data Modelling' section of IBM Watson Studio. On the left, there's a 'Preview' tab showing a table with 5 columns: o_ring_ct, o_ring_failures, temperature, pressure, and launch_id. The table contains 11 rows of data. On the right, there's an 'Information' panel with sections for 'Data asset', 'challengers.csv', 'Description' (no description available), 'Tags' (no tags available), 'Creator' (Pranjal Gupta), 'Usage' (Created on Apr 17, 2022, 03:12 PM), and 'Size' (388 Bytes). The URL bar at the top contains <https://dataplatform.cloud.ibm.com/projects/8965589a-daec-44f2-ad68-9a6d18b9e31f/data-assets/15717e8f-c1b3-47d7-9df9-f1c2c7e66b4d/preview?context=cpdaas&wal...>.

Figure: Dataset added

The screenshot shows the IBM Watson Studio interface. On the left, a data preview for 'challengers.csv_flow' is displayed, showing a table with columns: o_ring_ct, o_ring_failu..., temperature, pressure, and launch_id. The table contains 23 rows of data. On the right, the 'Information' panel is open, showing the 'Details' tab. It includes fields for 'LOCATION' (Predictive Data Modelling), 'DATA REFINERY FLOW NAME' (challengers.csv_flow), and 'STEPS' (1). A large oval highlights the 'DATA REFINERY FLOW NAME' field.

Figure: Data after refining

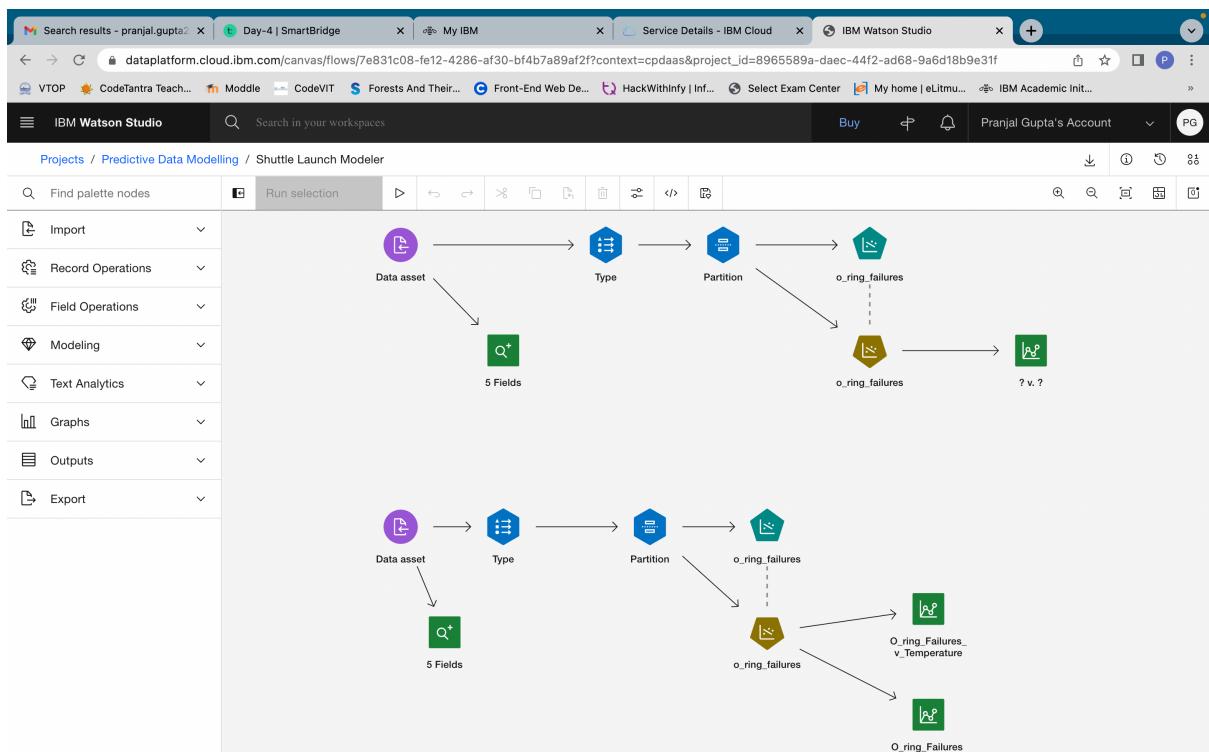


Figure: Scenario after creating SSP Model

The screenshot shows the 'Type' configuration dialog in IBM Watson Studio. On the left, a sidebar lists project categories like Import, Record Operations, Field Operations, Modeling, Text Analytics, Graphs, Outputs, and Export. The main area displays a data flow diagram with a 'Data asset' node connected to a 'Type' node, which then connects to a 'Partition' node. The 'Partition' node has two output branches, each ending in an 'o_ring_failure' node. Below the flow diagram, a table lists five fields: '# o_ring_ct', '# o_ring_failures', '# temperature', '# pressure', and '# launch_id'. Each field is defined as Continuous, with various roles and value modes. A 'Default mode' section includes options for 'Read metadata' (selected) and 'Pass (do not scan)'. Other settings include 'Set categorical fields to None if they exceed this many values' (50) and 'Set continuous integer field to ordinal if range less than or equal to' (10). Buttons for 'Cancel' and 'Save' are at the bottom.

Figure: Type of attributes of the data

The screenshot shows the 'Partition' configuration dialog in IBM Watson Studio. The sidebar and flow diagram are identical to the previous 'Type' dialog. The right panel is titled 'Partition' and contains settings for training and testing partitions. It shows 'Training Partition(%)' set to 70 and 'Testing Partition(%)' set to 30. There are checkboxes for 'Create validation partition', 'Repeatable partition assignment' (which is checked), and 'Seed' (set to 1234567). There is also a checkbox for 'Use unique field to assign partitions'. Buttons for 'Cancel' and 'Save' are at the bottom.

Figure: Partitioning the data for test and training

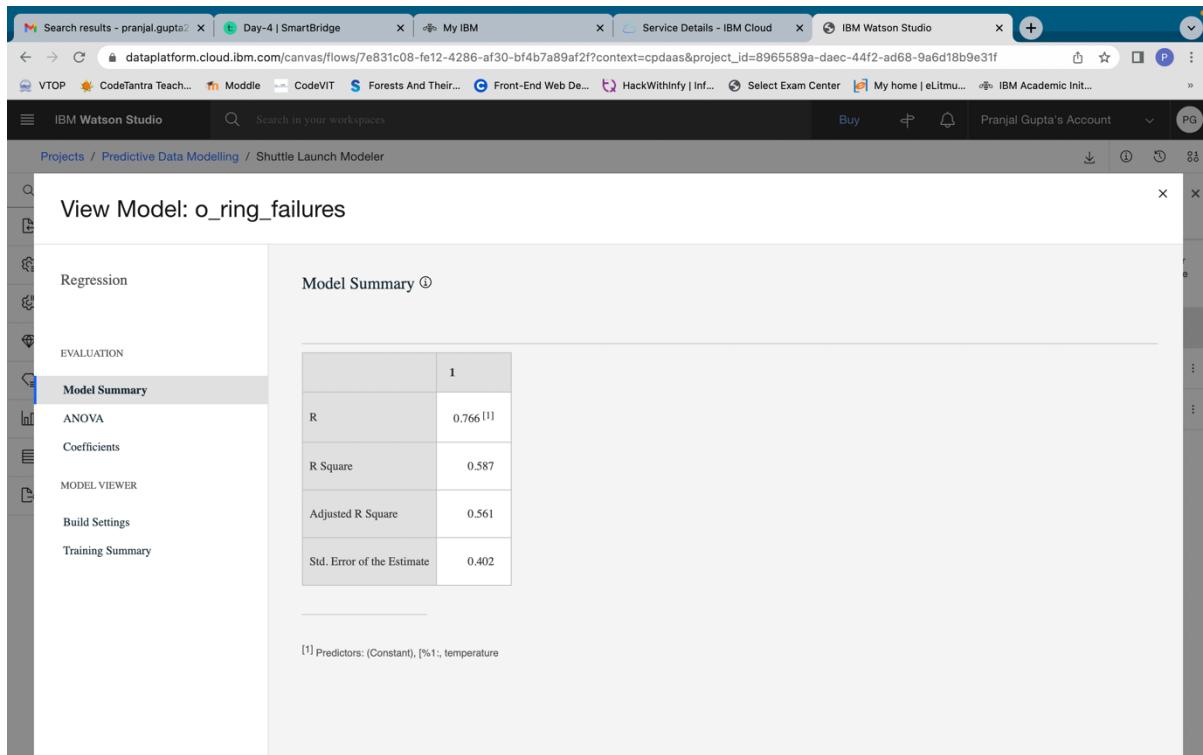


Figure: Output after applying Regression Model

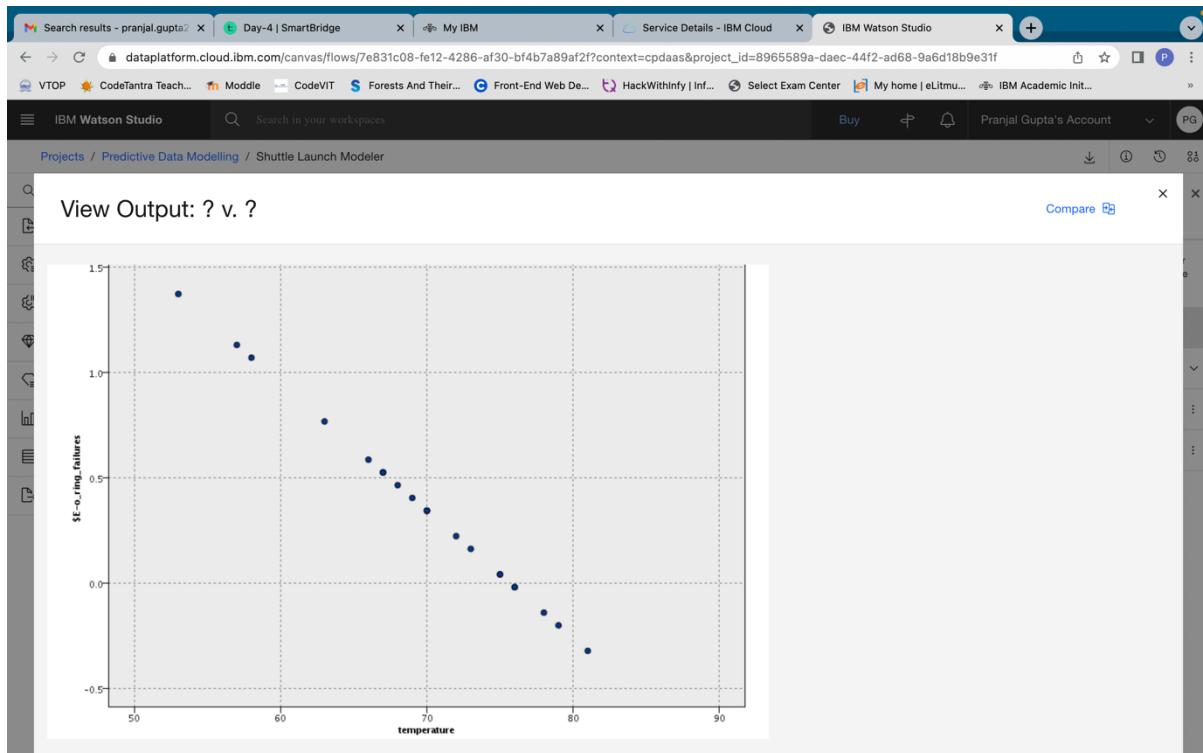


Figure: Visualizing the regression output using plot graph (For Temperature)

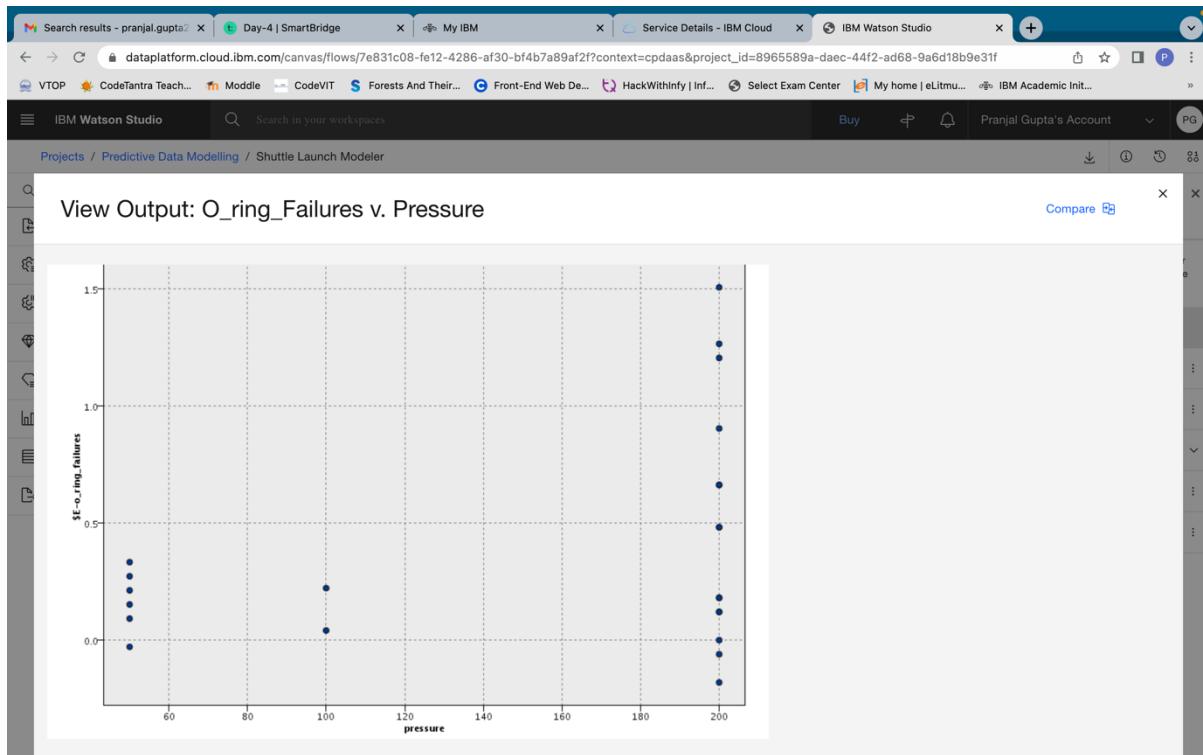


Figure: Visualizing the regression output (For Pressure)

Dataset: Bank.csv

Link: https://dataplatform.cloud.ibm.com/canvas/flows/cfc3ee5e-a7d0-4ed0-8e45-9e80842dc99c?context=cpdaas&project_id=145a6d4f-5398-4e47-806c-dd139854af60

The figure shows the "New project" creation dialog in the IBM Watson Studio interface. The "Define details" section includes fields for "Name" (Predictive_Analysis_BankData) and "Storage" (Cloud Object Storage-ma). The "Description" field contains "Predictive Analysis for Bank". The "Choose project options" section includes a checkbox for "Restrict who can be a collaborator" and a note about integration with Cloud Object Storage. At the bottom are "Cancel" and "Create" buttons.

Figure: Creating Project for bank data set

Pranjal Gupta

19BCE2619

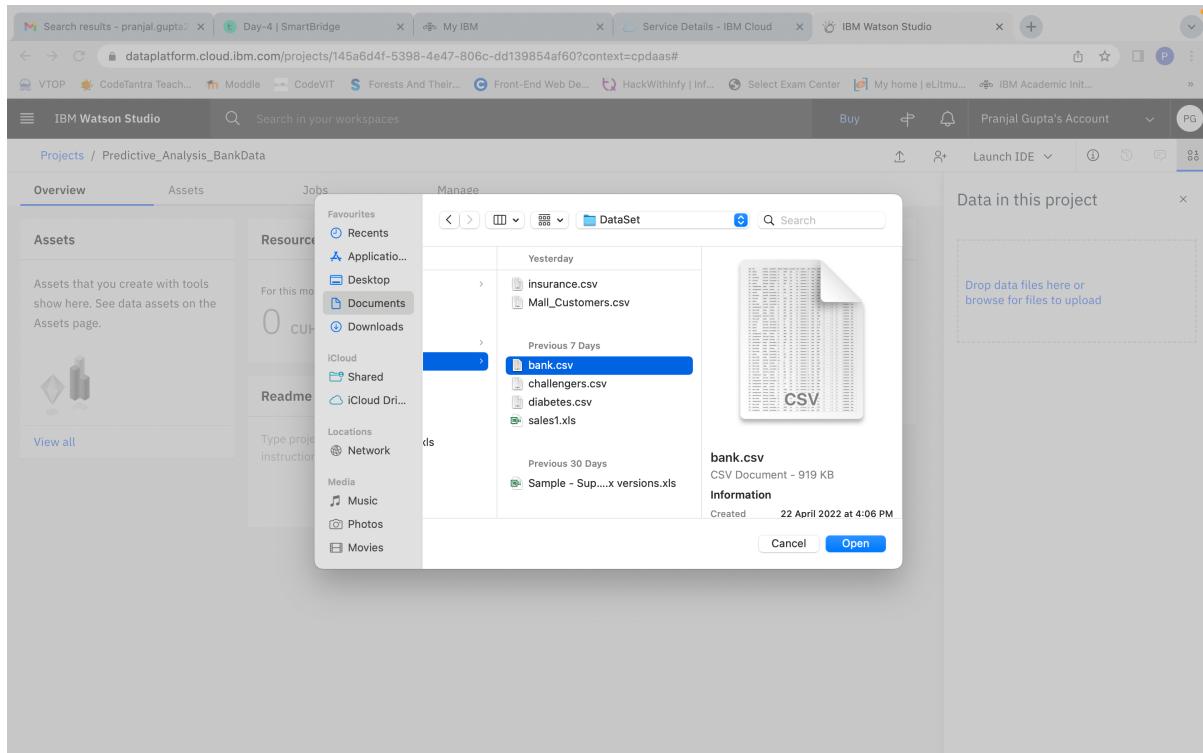


Figure: Adding Dataset

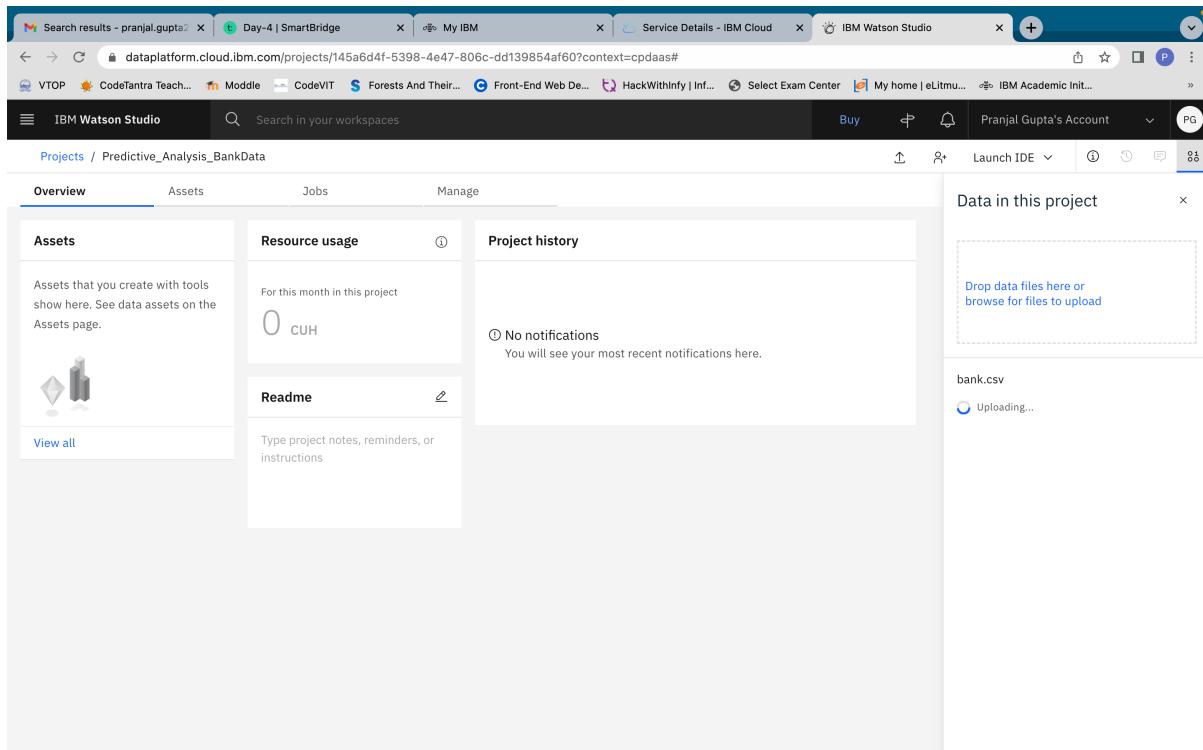


Figure: Data uploading

Pranjal Gupta
19BCE2619

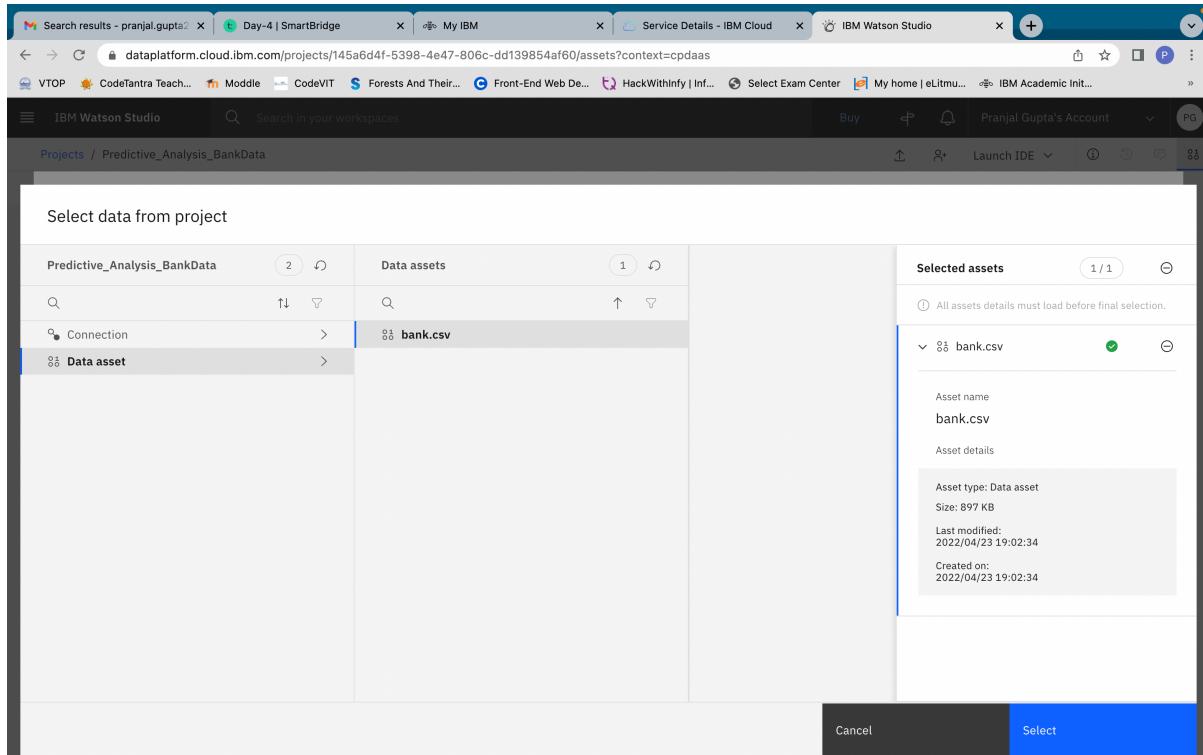


Figure: Selecting Dataset for data refining

The screenshot shows the 'Refine data' step in the 'Predictive_Analysis_BankData' project. On the left, a preview of the first 50 rows of the 'bank.csv' dataset is displayed in a table format. The columns are labeled: age, job, marital, education, default, balance, and housing. The data shows various categories and numerical values. On the right, the 'Details' tab of the data refinement flow is visible, showing the flow name 'bank.csv_flow', a description field, and a 'STEPS' section indicating 0 steps. Below this, the 'DATA REFINERY FLOW OUTPUT' section shows the location 'Predictive_Analysis_BankData' and the data set name 'Data a...'. A status message at the top right says 'Previewing the first 50 rows'.

Figure: Data Refining

Pranjal Gupta
19BCE2619

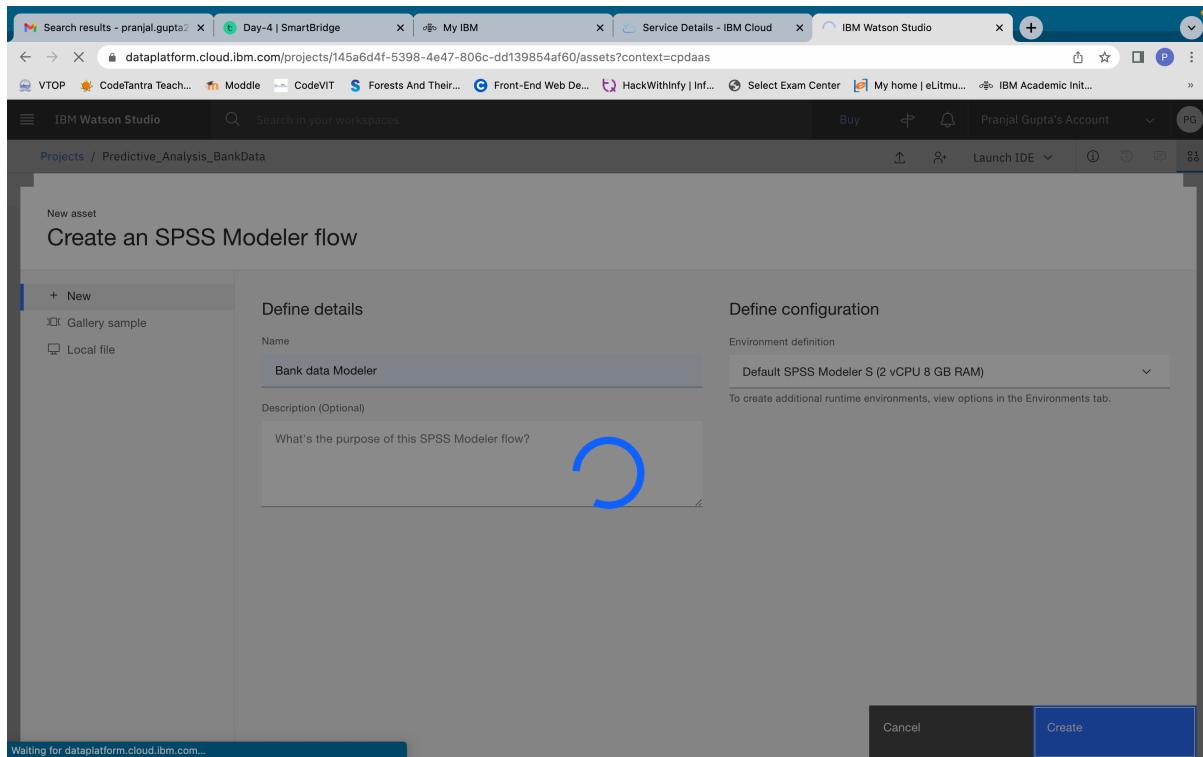


Figure: Creating SSP Modeler

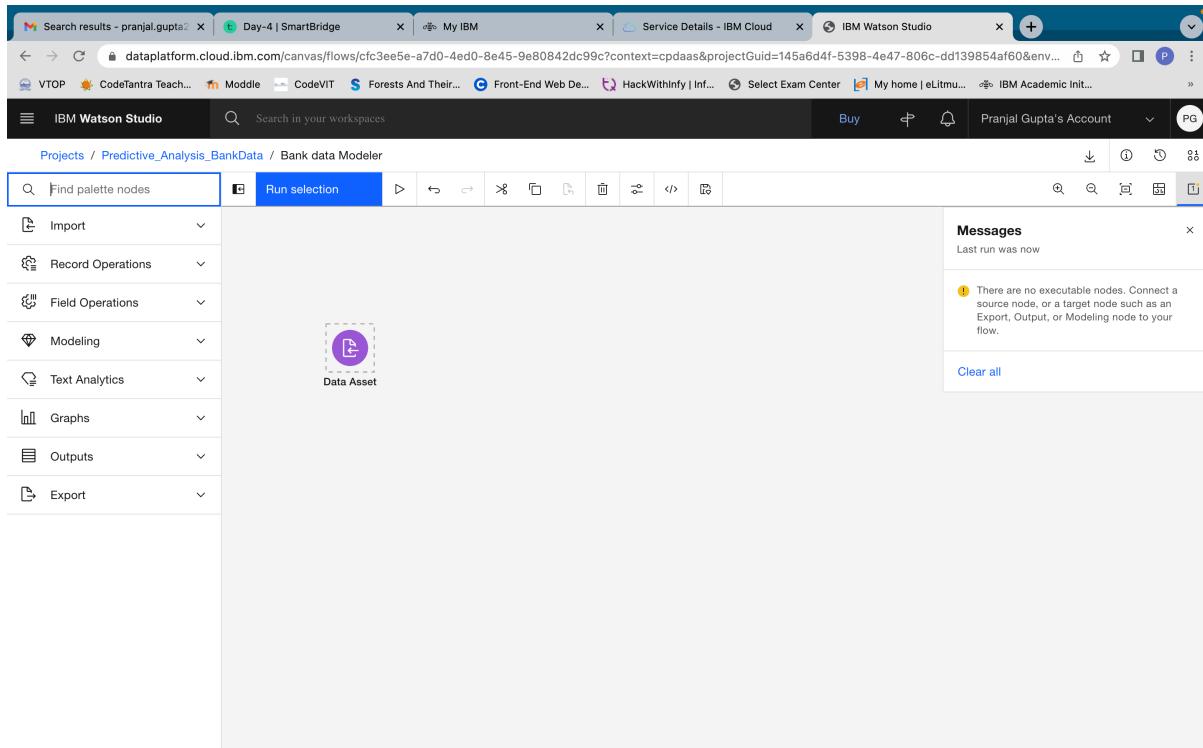


Figure: Data Set Loaded

The screenshot shows the IBM Watson Studio interface. On the left, there's a sidebar with a search bar and a list of output types: Table, Matrix, Analysis, Data Audit, Transform, Statistics, Means, Report, Set Globals, Sim Fit, Sim Evaluation, Extension Output, and KDE Simulation. The main workspace displays a flowchart: a 'Data Asset' node connects to a 'Type' node, which then connects to a '17 Fields' node. On the right, there's a panel titled 'Outputs' with a message about session availability and a list of audit results.

Figure: Performed Data Audit

This screenshot shows the 'Type' configuration dialog in IBM Watson Studio. It lists various fields with their properties: # age (Continuous, Input, Read), abc job (Categorical, Input, Read), abc marital (Categorical, Input, Read), abc education (Categorical, Input, Read), abc default (Categorical, Input, Read), # balance (Continuous, Input, Target, Read), abc housing (Categorical, Input, Read), and abc loan (Categorical, Input, Read). The '# balance' field is circled in red to indicate it has been assigned as the target attribute.

Field	Measure	Role	Value mode	Values
# age	Continuous	Input	Read	
abc job	Categorical	Input	Read	
abc marital	Categorical	Input	Read	
abc education	Categorical	Input	Read	
abc default	Categorical	Input	Read	
# balance	Continuous	Input	Read	
abc housing	Categorical	Input	Read	
abc loan	Categorical	Input	Read	

Figure: Assigned the target attribute

Pranjal Gupta

19BCE2619

Find palette nodes

Run selection

Data Asset → Type → Partition

17 Fields

Partition

Settings

Derived Field Name: Partition

Training Partition(%): 70

Testing Partition(%): 30

Create validation partition

Repeatable partition assignment

Seed: 1234567

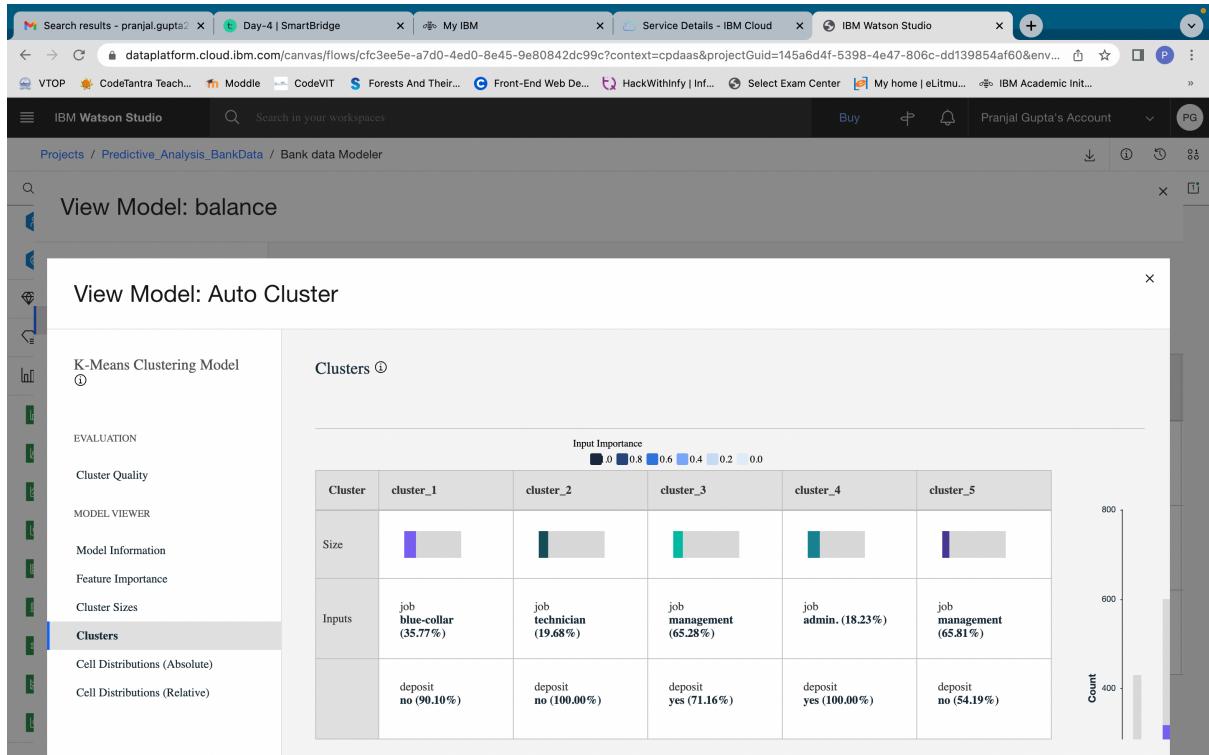
Use unique field to assign partitions

Cancel Save

Figure: Specified the training partition and test partition

USE	MODEL_NAME	ESTIMATOR	GRAPH	SILHOUETTE	BUILD TIME (MINS)	NUMBER OF CLUSTERS	SMALLEST CLUSTER (N)	SMALLEST CLUSTER (%)	LARGEST CLUSTER (N)	LARGEST CLUSTER (%)
<input checked="" type="radio"/>	K-means	KMeans	[Bar Chart]	0.156	< 1	5	965	0.124	2,026	100%
<input type="radio"/>	Kohonen	Kohonen	[Bar Chart]	0.057	< 1	12	70	0.009	1,367	100%
<input type="radio"/>	TwoStep	TwoStep	[Bar Chart]	0.115	< 1	7	527	0.068	1,730	100%

Figure: Model View after performing auto cluster



Figure; K-Means Cluster Model

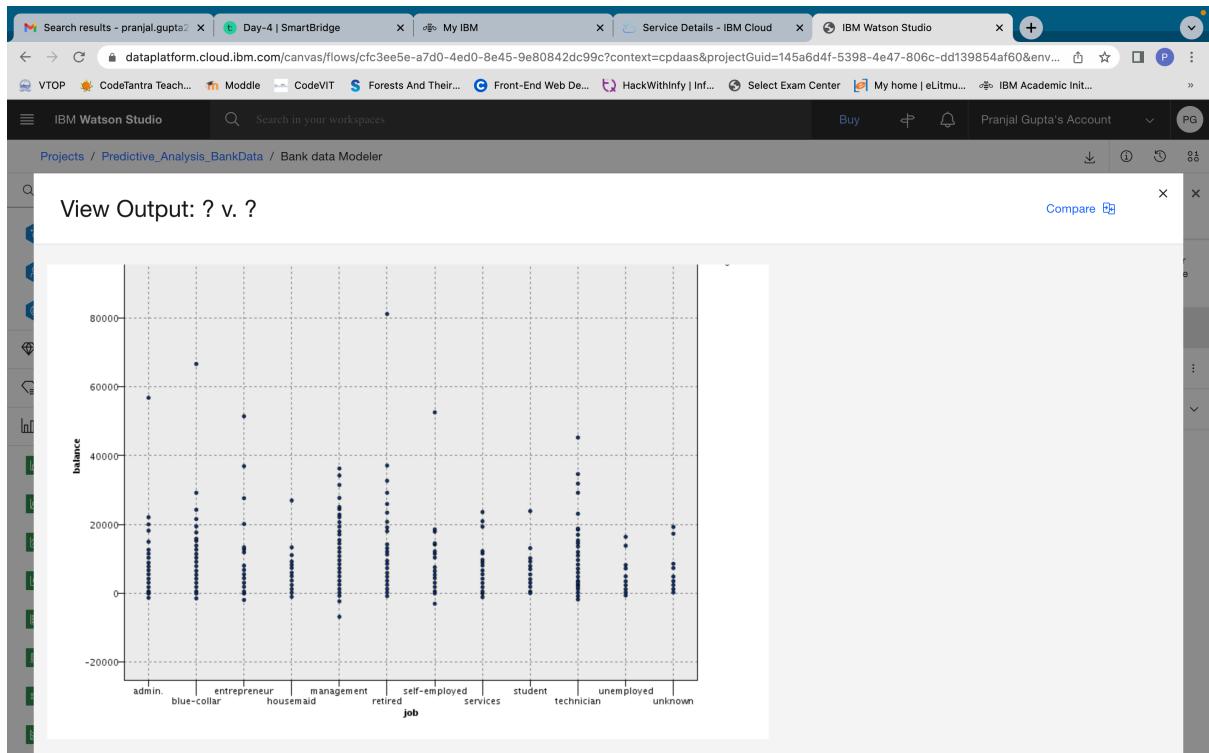


Figure: visualization of Balances over jobs