

19BCE2647

Aadarshbhushan Singh

VIT Vellore

Smartinternz Assignment 02

Data Analytics

23rd April 2022

Dataset Name: Challenger.csv

Link to the dashboard:

<https://eu-gb.dataplatform.cloud.ibm.com/projects/ce0f8d61-8715-49ec-ac76-0870af48764f/assets?context=cpdaas>

The screenshot shows the IBM Cloud dashboard with the 'Dashboard' selected in the top-left corner. The main area is titled 'For you' and features several cards:

- Build**: Explore IBM Cloud with this selection of easy starter tutorials and services. (Getting Started, 10 min)
- Get started with machine learning + Watson Studio**: Build, run and manage AI models. Prepare data and build models anywhere using open source code or visual modeling. Predict and optimize your outcomes. (Getting started, 2 min)
- Explore IBM Cloud Shell**: Try a command-driven approach for creating, developing, and deploying a web project. (Getting started, 3 min)
- Create a custom dashboard**: Create a shareable dashboard that you can customize with widgets, scope, and your own layout. (Recommended, 5 min)
- Watson Discovery Youtube Channel**: Check out the demos that we have published on YouTube and explore the features Watson Discovery has to offer. (Recommended, 1 hr)
- Build a data lake**: Create a data lake using Object Storage and query it in its native format with SQL Query and Watson Studio. (Recommended, 10 min)
- Search Skill Demo**: Check out this Search Skill Demo to learn how Watson Assistant and Watson Discovery work together to provide better, more accurate results. (Popular)
- IBM Watson Machine Learning**: Deploy, monitor and optimize machine models quickly and leverage auto-generated APIs to infuse AI into applications. (Popular)

Below the 'For you' section, there are three cards:

- Resource summary**: 2 Resources (Services and software, Storage). Add resources +
- Planned maintenance**: Clear skies! You can view your scheduled maintenance events here. View all
- IBM Cloud status**: No issues. View all

At the bottom, there are links for News, Recent support cases, User access, and Manage users.

Fig: Login to cloud.ibm.com and go to services and software

Watson Studio in Cloud Pak for Data

Watson Studio is one of the core services in Cloud Pak for Data as a Service. Build, deploy and manage AI models, and optimize decisions on IBM Cloud Pak for Data.

IBM Watson Studio in Cloud Pak for Data
IBM Cloud Pak for Data Unifying platform
IBM Cloud Base cloud infrastructure

Launch in IBM Cloud Pak for Data

Helpful links

- Documentation**: Learn about tools, features, and how to perform a wide variety of Data and AI tasks.
- Learning path**: Start a step-by-step tutorial to get up and running quickly.
- Videos**: Watch videos to learn about Watson Studio and Cloud Pak for Data as a Service.

How to use Watson Studio

Build, deploy, and trust AI models

Build, deploy and manage AI models, and optimize decisions anywhere on IBM Cloud Pak® for Data. Get started with uniting teams, automating AI lifecycles and speed time to value on an open multicloud architecture.

Fig: Go to Watson studio and launch it

Now create a new project and allocate the hardware requirements.

New asset

Select the tool to create an operational or configuration asset.

Tool type

- All types
- Automated builders
- Graphical builders
- Code editors
- Component editors
- Data access tools

Find tools by name or purpose

Automated builders

- AutoAI**: Automatically analyze your tabular data and generate candidate model pipelines customized for your predictive modeling problem.
- Metadata enrichment**: Enrich imported asset metadata with business context, data profiling, and quality assessment.

Graphical builders

- Dashboard editor**: Create a set of visualizations of analytical results on a graphical canvas without coding.
- Data Refinery**: Create a flow of ordered operations to cleanse and shape data. Visualize data to identify problems and discover insights.
- Decision Optimization**: Create and manage scenarios to find the best solution to your optimization problem by comparing different combinations of your model, data, and solutions.
- SPSS Modeler**: Create a visual flow that uses modeling algorithms to prepare data and build and train a model, using a guided approach to machine learning that doesn't require coding.

Code editors

- Federated Learning**: Create a federated learning experiment to train a common model on a set of remote data sources. Share training results without sharing data.
- Jupyter notebook editor**: Create a notebook in which you run Python, R, or Scala code to prepare, visualize, and analyze data, or build a model.

Show descriptions

Cancel Back Next

Fig: Upload the data and go to Data Refinery

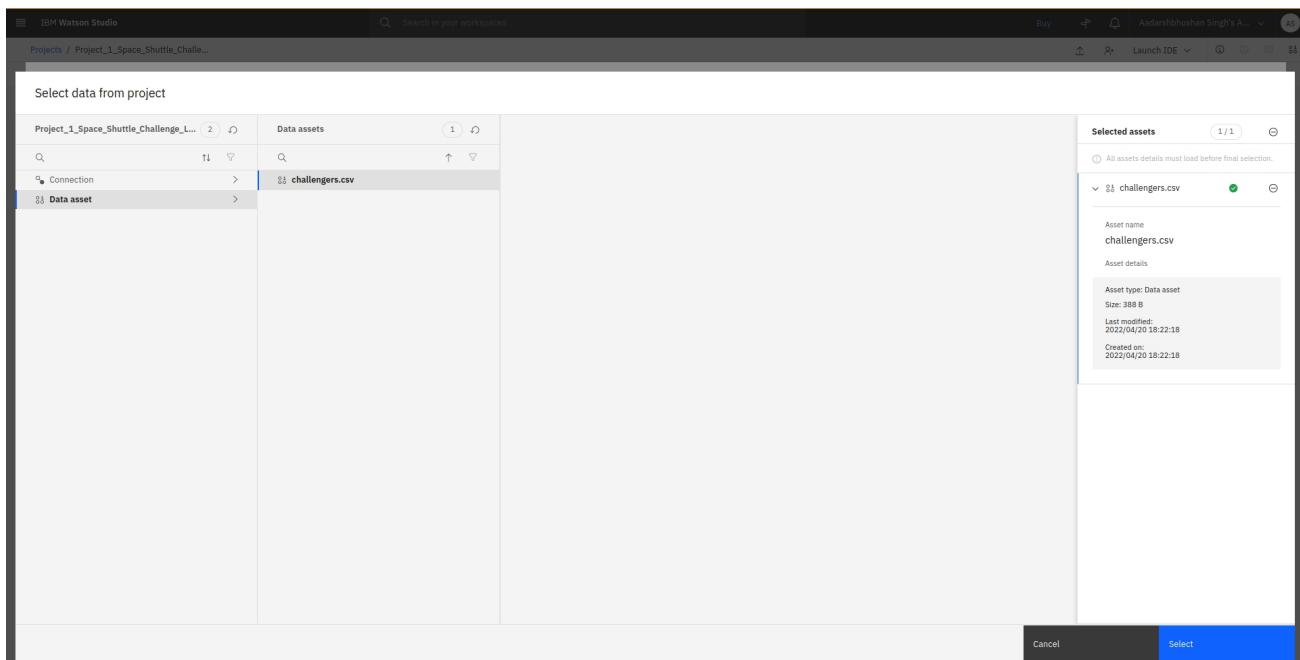


Fig: Select Challengers.csv

The screenshot shows the 'Refine data' dialog for the 'challengers.csv' file. On the left, a sidebar lists various data operations: CLEANSE, COMPUTE, and ORGANIZE. The 'Data' tab is selected, 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' tab displays details: LOCATION (Project_1_Space_Shuttle_Challen...), DATA REFINERY FLOW NAME (challengers.csv_flow), and STEPS (1). Below the table, it says SOURCE FILE: challengers.csv FULL DATA SET: 23 rows.

Fig: Showing the data and performing some experimental actions in new Assets

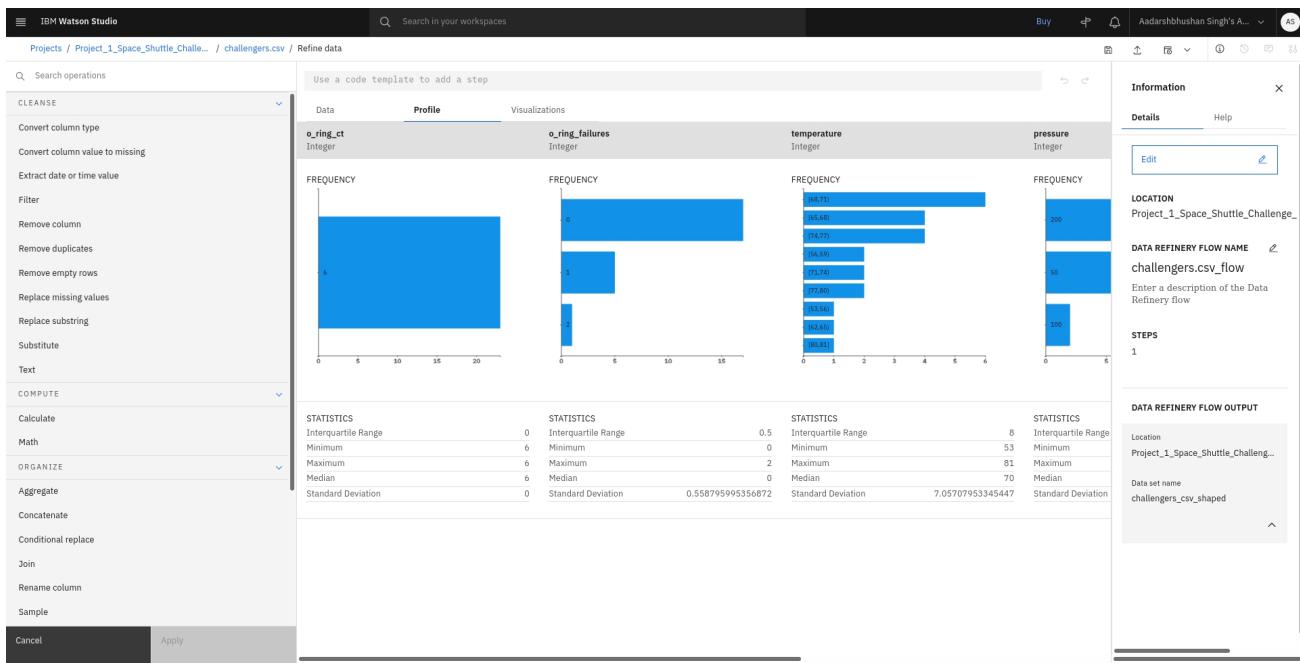


Fig: Viewing the profile and other details of data

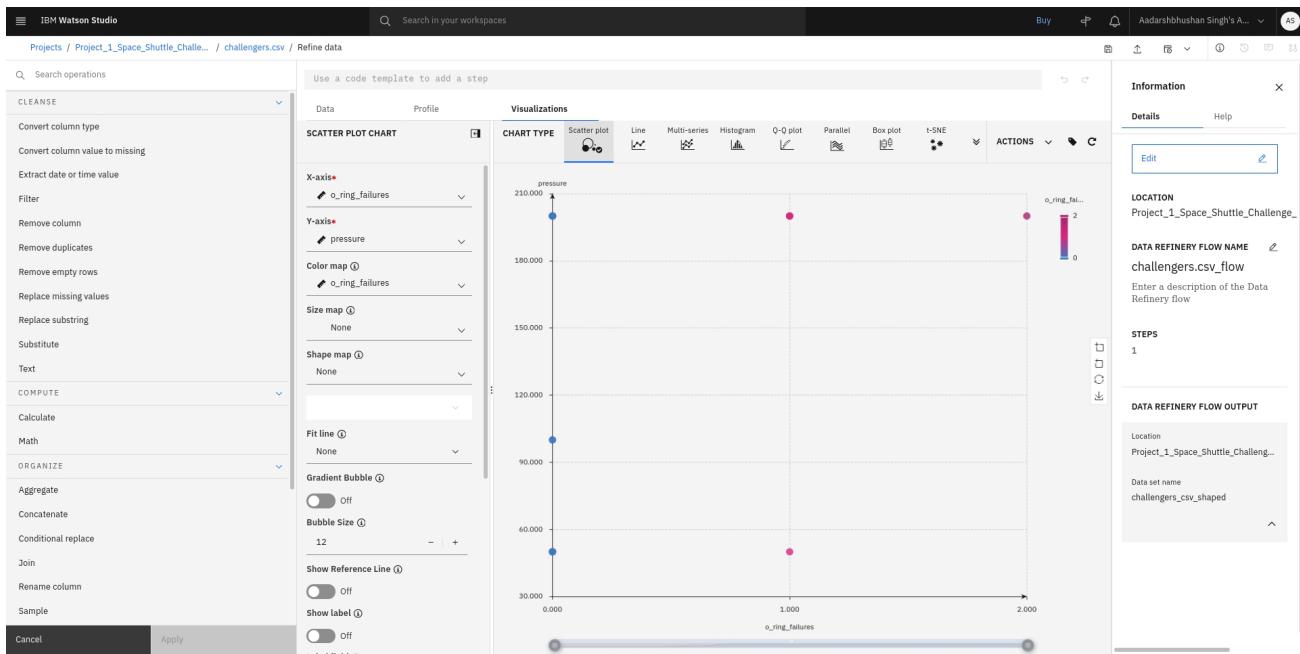


Fig: Visualizing the data of ring failure vs pressure

Now Create the SSPS Modeler and perform the actions.

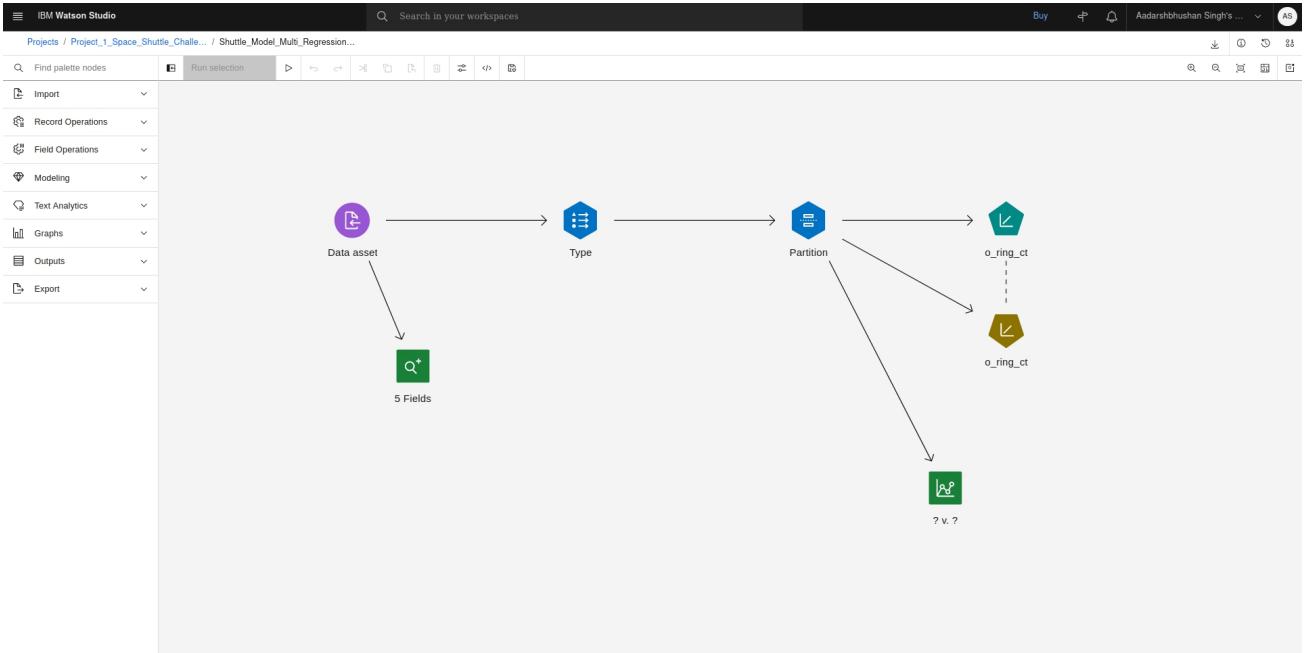


Fig: SSPS Modeler of the data

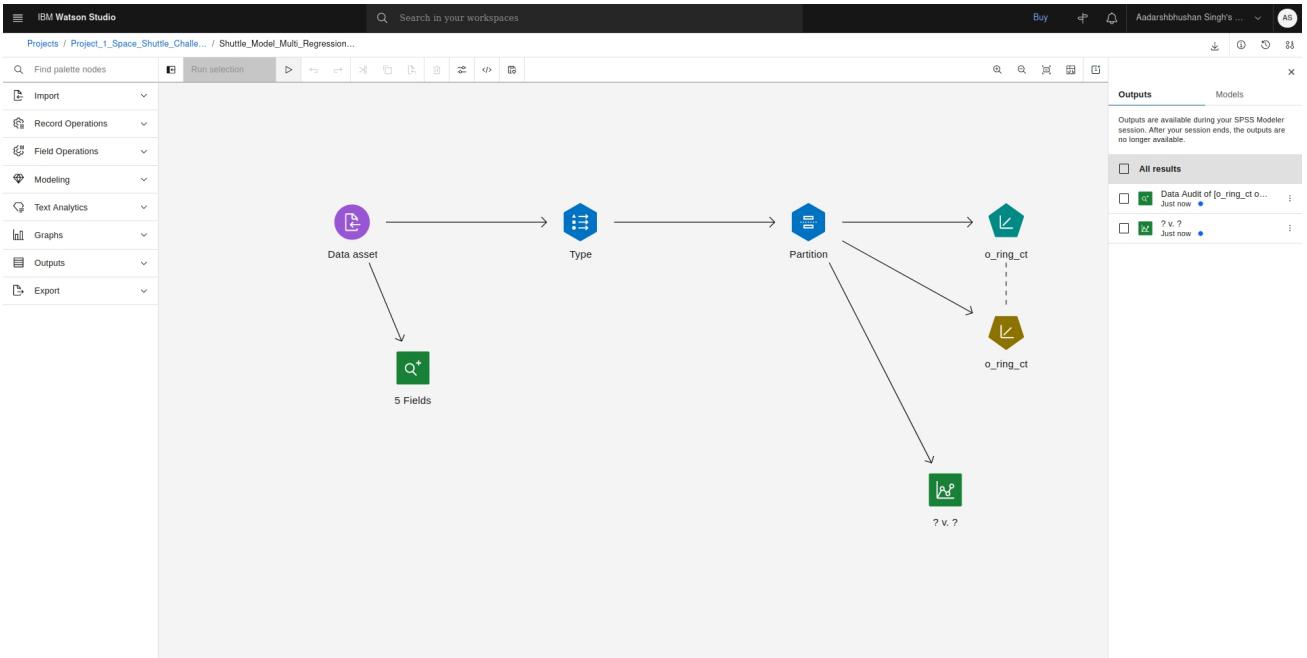


Fig: Running the model

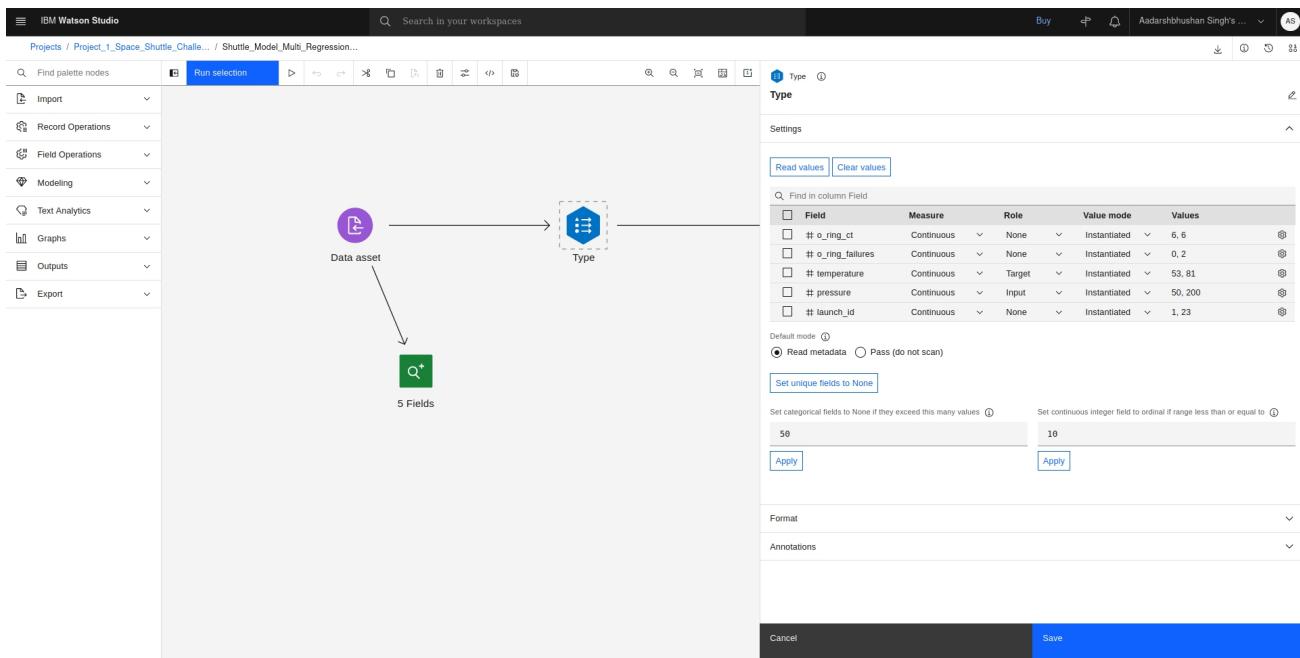


Fig: Fixing the type of attribute

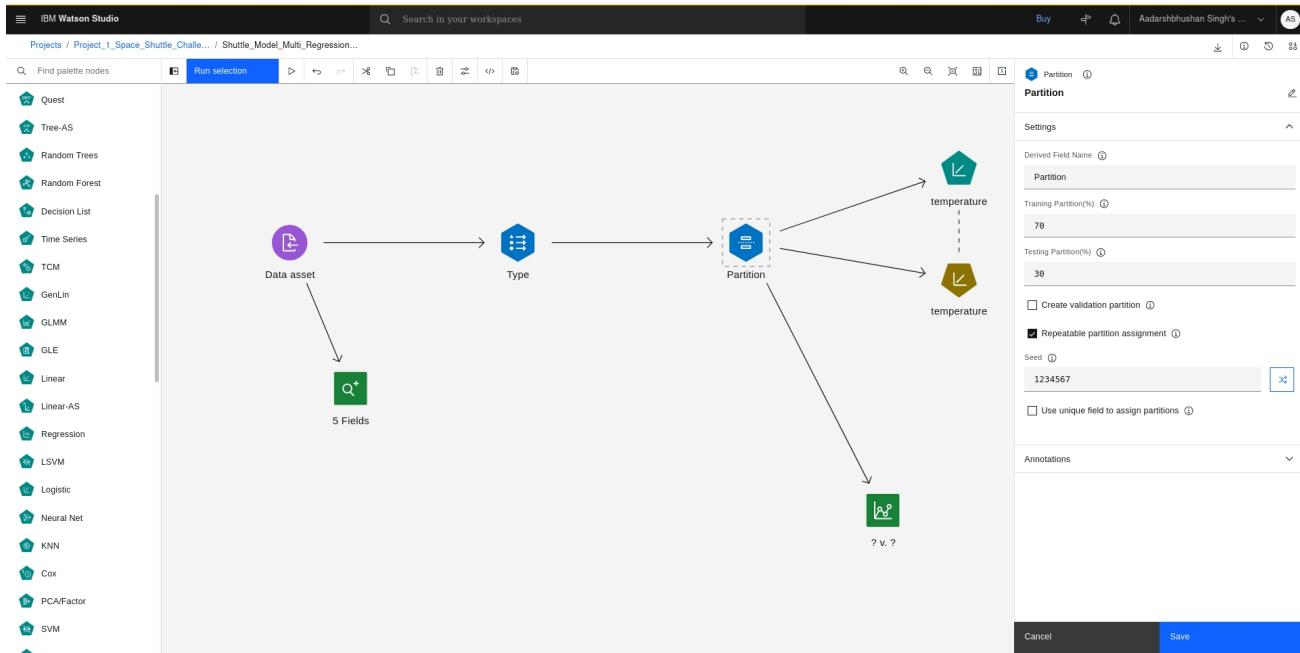


Fig: Setting up 70 train and 30 test model

IBM Watson Studio Search in your workspaces Buy Compare

Projects / Project_1_Space_Shuttle_Challe... / Shuttle_Model_Multi_Regression...

View Output: Data Audit of [o_ring_ct o_ring_failures temperature pressure launch_id]

Compare

Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid			
1 o_ring_ct		Continuous	6	6	6	0	--	--	23			
2 o_ring_failures		Continuous	0	2	0.304	0.559	1.735	--	23			
3 temperature		Continuous	53	81	69.565	7.057	-0.654	--	23			
4 pressure		Continuous	50	200	152.174	68.221	-0.791	--	23			
5 launch_id		Continuous	1	23	12	6.782	0	--	23			
Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
1 o_ring_ct	Continuous	0	0	None	Never	Fixed	100.000	23	0	0	0	0
2 o_ring_failures	Continuous	1	0	None	Never	Fixed	100.000	23	0	0	0	0
3 temperature	Continuous	0	0	None	Never	Fixed	100.000	23	0	0	0	0
4 pressure	Continuous	0	0	None	Never	Fixed	100.000	23	0	0	0	0
5 launch_id	Continuous	0	0	None	Never	Fixed	100.000	23	0	0	0	0

Fig: Output of data audit of the data

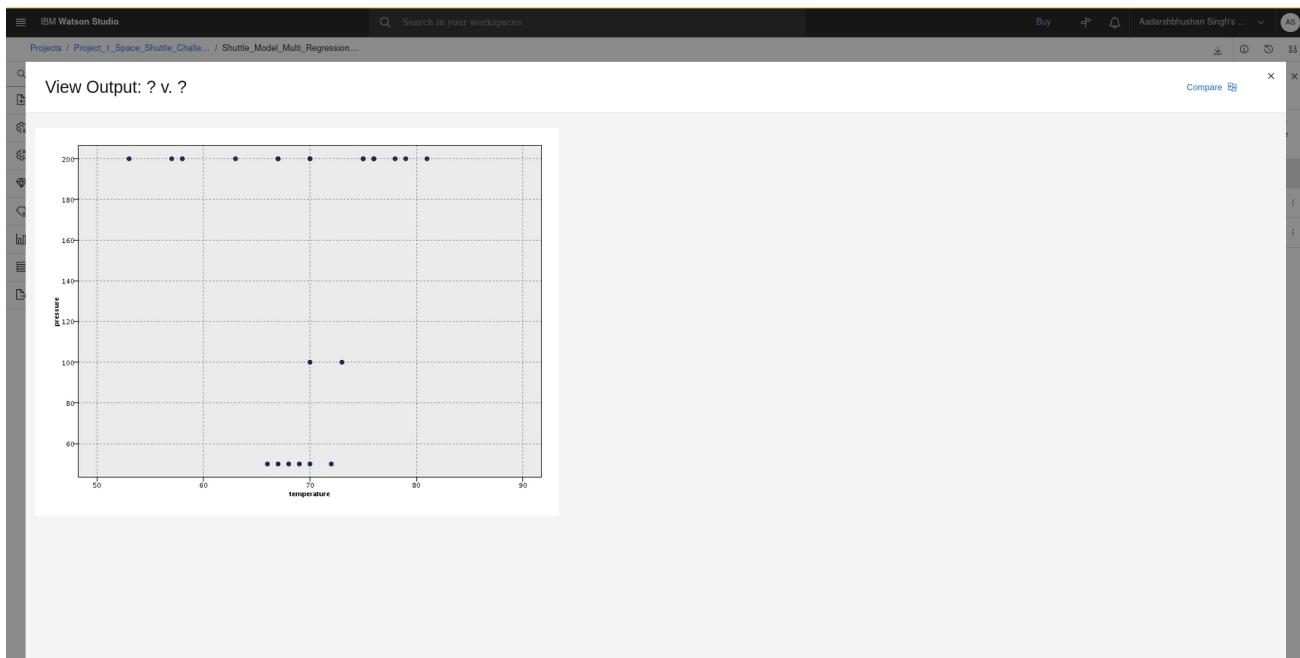


Fig: Output of plot

Dataset Name: bank.csv

Link to the dashboard:

<https://eu-gb.dataplatfrom.cloud.ibm.com/projects/0fc5e9d2-c170-4210-a250-cea347dd3818/assets?context=cpdaas>

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'IBM Watson Studio', a search bar, and user information. Below the header is a large banner with the text 'Welcome, Aadarshbhushan!' and three main sections: 'Take a tutorial', 'Work with data', and 'Extend your capabilities'. To the right of these is a decorative graphic of a computer monitor displaying a 3D cube grid with a magnifying glass over it. The main content area has several sections: 'Quick start' (with links to build dashboards, data pipelines, customer profiles, catalog/governance, ML models, and query anywhere), 'Overview' (showing a list of projects like 'Classification_Bank_Data_Prediction', 'Predicting_Diabetes_21-04-2022', and 'Project_1_Space_Shuttle_Challenge_Linear_Regression'), 'Notifications' (listing a 'Project export complete' message for a specific project), and 'Deployments' (which is currently empty). On the left side, there's a sidebar with 'Support' links for documentation, FAQ, share ideas, and support center.

Fig: Create New Project

This screenshot shows a modal dialog titled 'Select data from project' within the IBM Watson Studio interface. The dialog lists data assets under the project 'Classification_Bank_Data_Prediction'. A single asset, 'bank.csv', is selected and highlighted in blue. On the right side of the dialog, there's a detailed view of the selected asset: 'Selected assets' (1/1), 'bank.csv', Asset name (bank.csv), Asset details (Asset type: Data asset, Size: 897 KB, Last modified: 2023/04/22 11:07:01, Created on: 2023/04/22 11:07:01). At the bottom of the dialog are 'Cancel' and 'Select' buttons.

Fig: Selecting data for data refining

IBM Watson Studio

Search in your workspaces

Buy  Aadarshbhusan Singh's A... 

Projects / Classification_Bank_Data_Pred... / bank.csv / Refine data

Steps (1)

Data Source: bank.csv

1. Convert Column type

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Auto-generated

Data Profile Visualizations

	age	job	marital	education	default	balance	housing	loan	contact
1	59	admin.	married	secondary	no	2343	yes	no	unknown
2	56	admin.	married	secondary	no	45	no	no	unknown
3	41	technician	married	secondary	no	1270	yes	no	unknown
4	55	services	married	secondary	no	2476	yes	no	unknown
5	54	admin.	married	tertiary	no	184	no	no	unknown
6	42	management	single	tertiary	no	0	yes	yes	unknown
7	56	management	married	tertiary	no	830	yes	yes	unknown
8	60	retired	divorced	secondary	no	545	yes	no	unknown
9	37	technician	married	secondary	no	1	yes	no	unknown
10	28	services	single	secondary	no	5090	yes	no	unknown
11	38	admin.	single	secondary	no	100	yes	no	unknown
12	30	blue-collar	married	secondary	no	309	yes	no	unknown
13	29	management	married	tertiary	no	199	yes	yes	unknown
14	46	blue-collar	single	tertiary	no	460	yes	no	unknown
15	31	technician	single	tertiary	no	703	yes	no	unknown
16	35	management	divorced	tertiary	no	3837	yes	no	unknown
17	32	blue-collar	single	primary	no	611	yes	no	unknown
18	49	services	married	secondary	no	-8	yes	no	unknown
19	41	admin.	married	secondary	no	55	yes	no	unknown
20	49	admin.	divorced	secondary	no	168	yes	yes	unknown
21	28	admin.	divorced	secondary	no	785	yes	no	unknown
22	43	management	single	tertiary	no	2067	yes	no	unknown
23	43	management	divorced	tertiary	no	388	yes	no	unknown
24	43	blue-collar	married	primary	no	-192	yes	no	unknown

New step

SOURCE FILE: bank.csv SAMPLE SIZE: First 10000 rows

Information

Details Help

LOCATION Classification_Bank_Data_Prediction

DATA REFINERY FLOW NAME bank.csv_flow

Enter a description of the Data Refinery flow

STEPS 1

DATA REFINERY FLOW OUTPUT

Location Classification_Bank_Data_Prediction...

Data set name bank_csv_shaped

Fig: Data Refining

IBM Watson Studio

Search in your workspaces

Buy  Aadarshbhusan Singh's A... 

Projects / Classification_Bank_Data_Pred... / bank.csv / Refine data

Steps (1)

Data Source: bank.csv

1. Convert column type

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Auto-generated

Data Profile Visualizations

Profile

age job marital education

FREQUENCY FREQUENCY FREQUENCY FREQUENCY

STATISTICS STATISTICS STATISTICS STATISTICS

Interquartile Range	17	Maximum length	13	Maximum length
Minimum	18	Minimum length	6	Minimum length
Maximum	95	Mean length	9.3258	Mean length
Median	39	Unique	12	Unique
Standard Deviation	12.1068977800832			

New step

Information

Details Help

LOCATION Classification_Bank_Data_Prediction

DATA REFINERY FLOW NAME bank.csv_flow

Enter a description of the Data Refinery flow

STEPS 1

DATA REFINERY FLOW OUTPUT

Location Classification_Bank_Data_Prediction...

Data set name bank_csv_shaped

Fig: Understanding Data though profile section

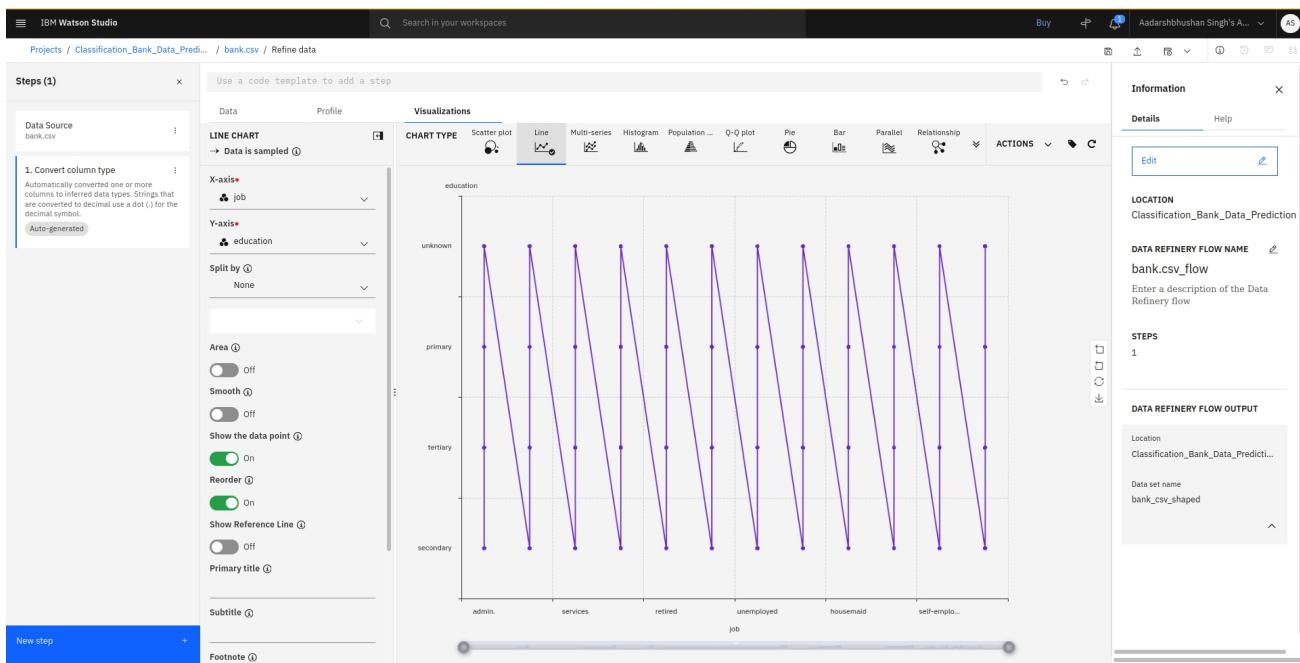


Fig: Data Visualization

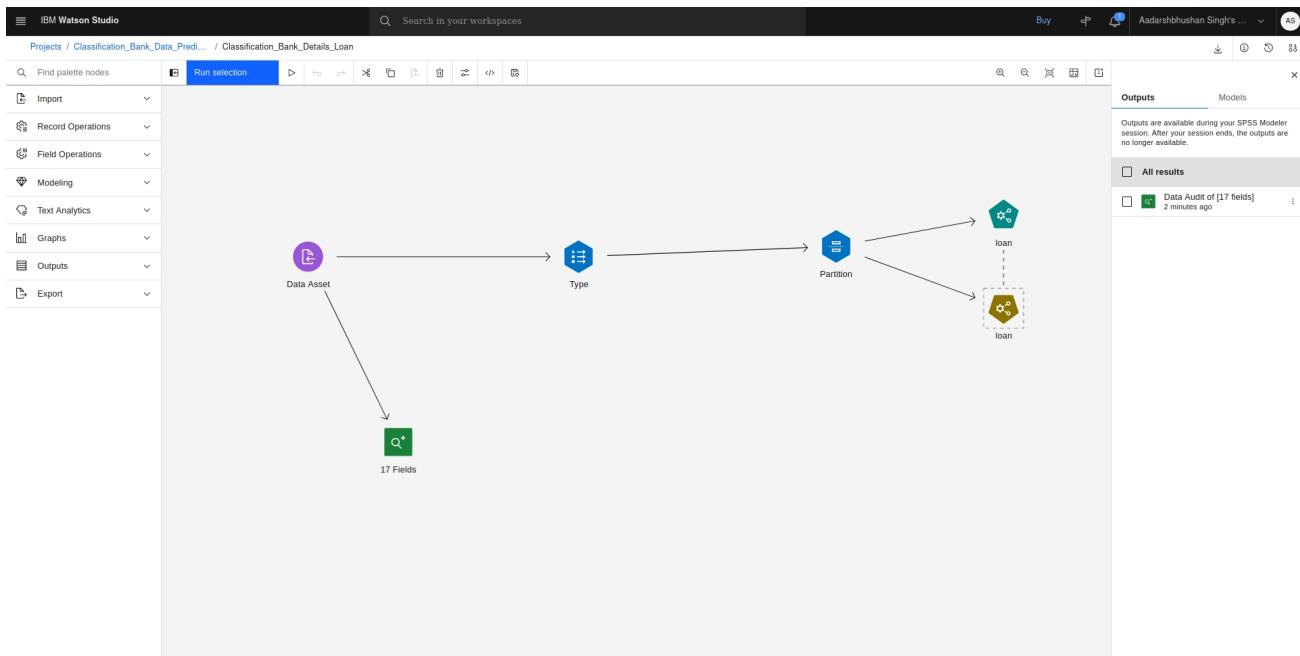


Fig: SSMP Modeling

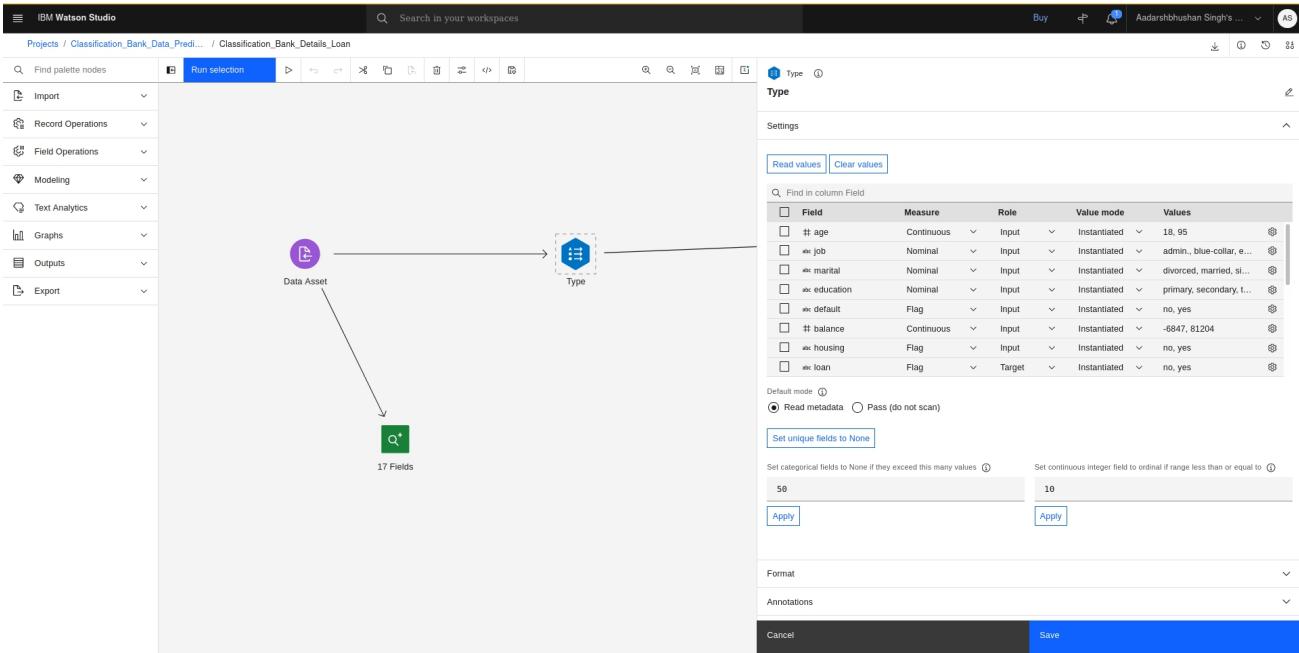


Fig: Fixing the attributes input and target value

USE	MODEL NAME	ESTIMATOR	BUILD TIME (MINS)	NO. FIELDS USED	ACCURACY	ACCUMULATED ACCURACY	AREA UNDER CURVE	ACCUMULATED AUC	RECALL	PRECISION	ACTIONS
<input checked="" type="checkbox"/>	XGBoost Linear 1	XGBoost Linear 1	< 1	13	87.189	87.189	0.725	0.725	0.018	0.800	
<input checked="" type="checkbox"/>	XGBoost Tree 1	XGBoost Binary Classification Model	< 1	13	86.892	86.892	0.751	0.751	0.037	0.444	
<input checked="" type="checkbox"/>	Logistic Regression 1	Nominal Regression	< 1	13	87.159	87.159	0.726	0.726	0.016	0.778	
<input checked="" type="checkbox"/>	LSVM 1	Linear SVM	< 1	13	87.070	87.070	0.722	0.722	0.007	0.750	
<input checked="" type="checkbox"/>	Bayesian Network 1	BayesNet	< 1	13	87.129	87.129	0.719	0.719	0.114	0.532	

Fig: Clustering of Data

IBM Watson Studio Search in your workspaces Buy Aadeershushan Singh's ... AS

Projects / Classification_Bank_Data_Pred... / Classification_Bank_Details_Loan

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

Field Sample Graph Measurement Min Max Mean Std. Dev Skewness Unique Valid

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
7 housing		Categorical	--	--	--	--	--	2	11162
8 loan		Categorical	--	--	--	--	--	2	11162
9 contact		Categorical	--	--	--	--	--	3	11162
10 day		Continuous	1	31	15.658	8.421	0.111	--	11162

https://eu-gb.dataplatform.cloud.ibm.com/canvasstreams/0fc5e9d2-c170-4210-a250-cea347dd3818:4cfbf010-425b-4853-99db-3...0f/nodes/id34BVXKT22TB/models?project_id=0fc5e9d2-c170-4210-a250-cea347dd3818&context=cpdaas&tearsheet_mode=true#view

Fig: Output of Data Audit