Machine Learning Model Deployment with IBM Cloud Watson Studio

Phase – 5(Documentation)

Abstract

This project leverages IBM Cloud Watson Studio to develop and deploy predictive models for customer churn analysis in the telecommunications industry. By collecting and analyzing customer data, training machine learning models, and integrating them into operational systems, the project aims to proactively reduce churn rates and enhance customer retention strategies.

Problem Statement

Become a wizard of predictive analytics with IBM Cloud Watson Studio. Train machine learning models to predict outcomes in real-time. Deploy the models as web services and integrate them into your applications. Unlock the magic of data-driven insights and make informed decisions like never before!

Problem Definition

The project involves training a machine learning model using IBM Cloud Watson Studio and deploying it as a web service. The goal is to become proficient in predictive analytics by creating a model that can predict outcomes in real-time. The project encompasses defining the predictive use case, selecting a suitable dataset, training a machine learning model,deploying the model as a web service, and integrating it into applications.

DATASET COLLECTION:

Customers who left within the last month (Churn):

This column indicates whether a customer terminated their relationship with the company within the past month. This information is crucial for businesses, as it helps them understand customer attrition rates. Analyzing churn can provide insights into factors that lead to customer dissatisfaction or reasons for discontinuing services.

Services that each customer has signed up for:

This section details the specific services that each customer has subscribed to. These services may include phone lines, multiple phone lines, internet services, online security, online backup, device protection, tech support, and streaming TV and movie packages. Understanding which services are popular among customers can guide marketing efforts and product development.

Customer account information:

This section encompasses various details related to the customer's account and usage:

- ***Tenure (how long they've been a customer):*** Indicates the length of time a customer has been with the company. Long-tenured customers are often seen as more valuable due to their loyalty and potential for continued business.
- ***Contract:*** Specifies the type of contract the customer has (e.g., month-to-month, one-year, two-year). Different contract types may have implications on customer behavior and revenue stability.
- ***Payment method:*** Describes how the customer pays for the services (e.g., credit card, electronic transfer). This information is important for billing and financial management.
- ***Paperless billing:*** Indicates whether the customer opts for paperless billing, which can have environmental and cost-saving implications for the company.
- ***Monthly charges***: Specifies the amount the customer is billed on a monthly basis for the services they've subscribed to.
- ***Total charges:*** Represents the cumulative charges incurred by the customer over their tenure with the company.

Demographic info about customers:

This section provides additional information about the customers' characteristics:

- ***Gender:*** Indicates whether the customer is male, female, or of another gender identity. This information can be useful for targeted marketing campaigns.
- ***Age range:*** Categorizes customers into specific age groups (e.g., 18-24, 25-34, etc.). Understanding the age distribution can help tailor products and services to different demographic segments.
- ***Partners and dependents:*** Indicates whether the customer has a partner (spouse or significant other) and if they have dependents (children or other individuals they financially support). This information is valuable for family-oriented marketing strategies and understanding household dynamics.

Analyzing these aspects collectively can provide valuable insights into customer behavior, preferences, and potential areas for improvement in services or marketing strategies.

APPROACH:

Data Collection and Preprocessing: To build a customer churn prediction model, historical customer data is collected. This data typically includes information such as customer demographics, usage patterns, contract details, billing information, and records of customer service interactions. It is crucial to clean and preprocess this data, addressing issues like missing values and outliers.

Model Selection: The choice of machine learning algorithms plays a crucial role in building the predictive model. Common algorithms used for this purpose include logistic regression, decision trees, random forests, gradient boosting, and more. The selection of the appropriate algorithm depends on the specific characteristics of the dataset and the business objectives.

Model Training and Evaluation: The predictive model is trained using historical customer data, with a focus on identifying patterns and trends associated with customer churn. The model's performance is assessed using various metrics, including accuracy (the percentage of correct predictions), precision (the percentage of true positives among predicted positives), recall (the percentage of true positives identified), F1-score (a balance between precision and recall), and ROC-AUC (Receiver Operating Characteristic - Area Under the Curve).

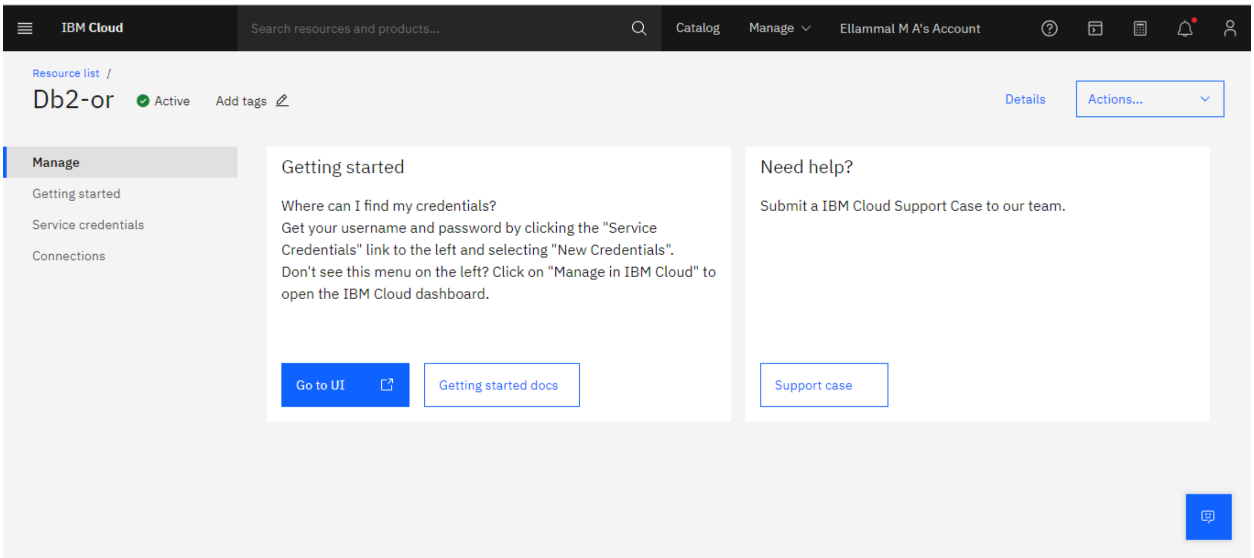
Churn Prediction: Once the model is trained and validated, it can be applied to new customer data. For each customer, the model calculates the probability of churn. Customers with high churn probabilities are flagged as at-risk churners.

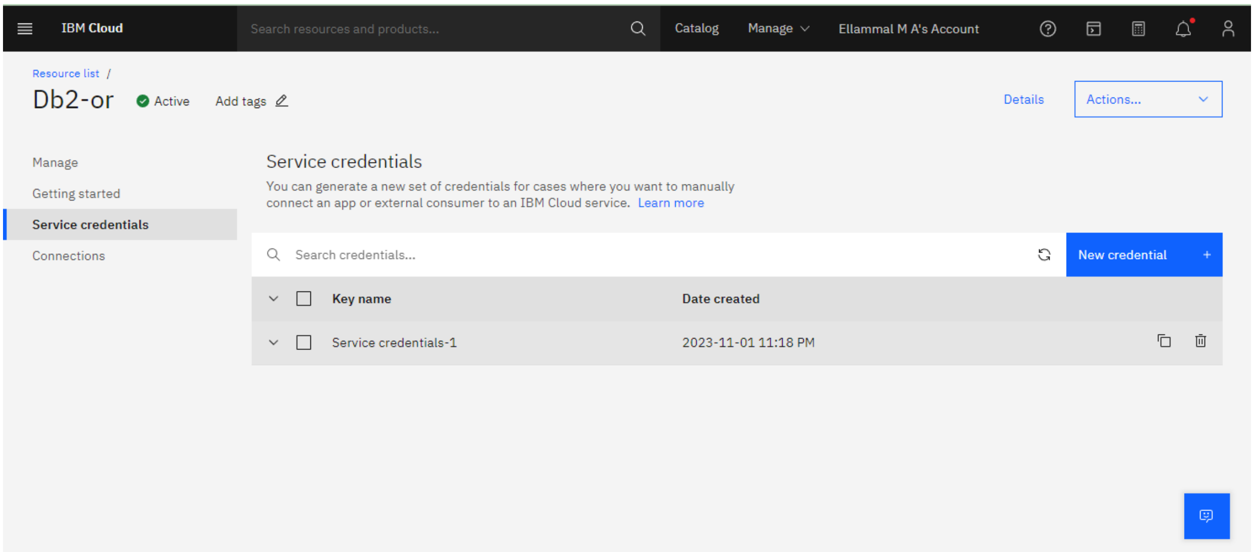
Retention Strategies: The predictions generated by the model are instrumental in devising retention strategies. For customers identified as high-risk churners, personalized offers, discounts, or incentives can be extended to encourage them to remain loyal to the telecom service.

Deployment: The churn prediction model is integrated into the telecom company's operational systems to enable real-time predictions and automate intervention processes. This ensures that retention efforts can be initiated promptly.

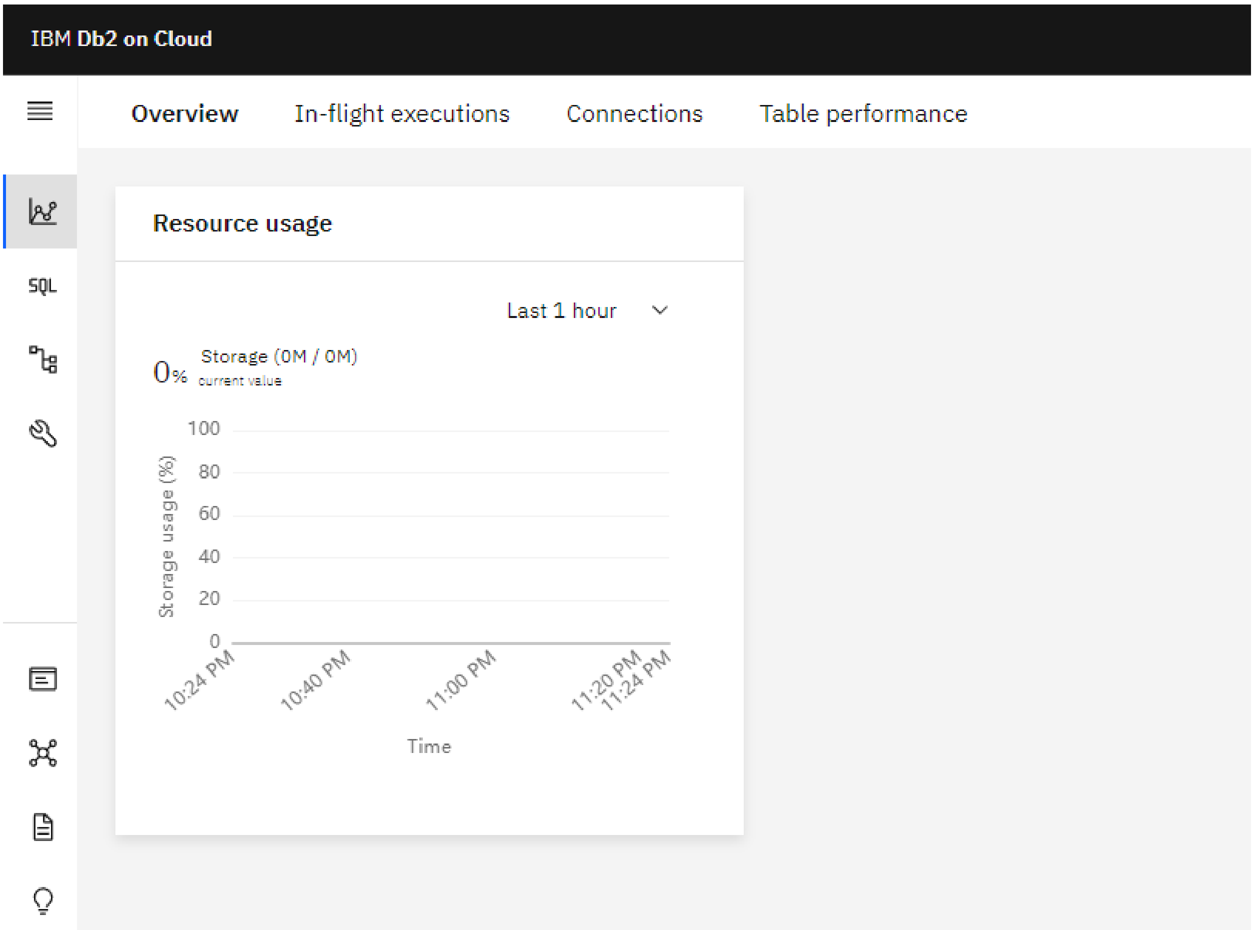
Monitoring and Iteration: The model's performance is continuously monitored, and it is regularly updated with new data. As customer behavior may change over time, this iterative process helps in maintaining the model's accuracy and relevance. Retention strategies are also adjusted based on feedback and evolving customer preferences.

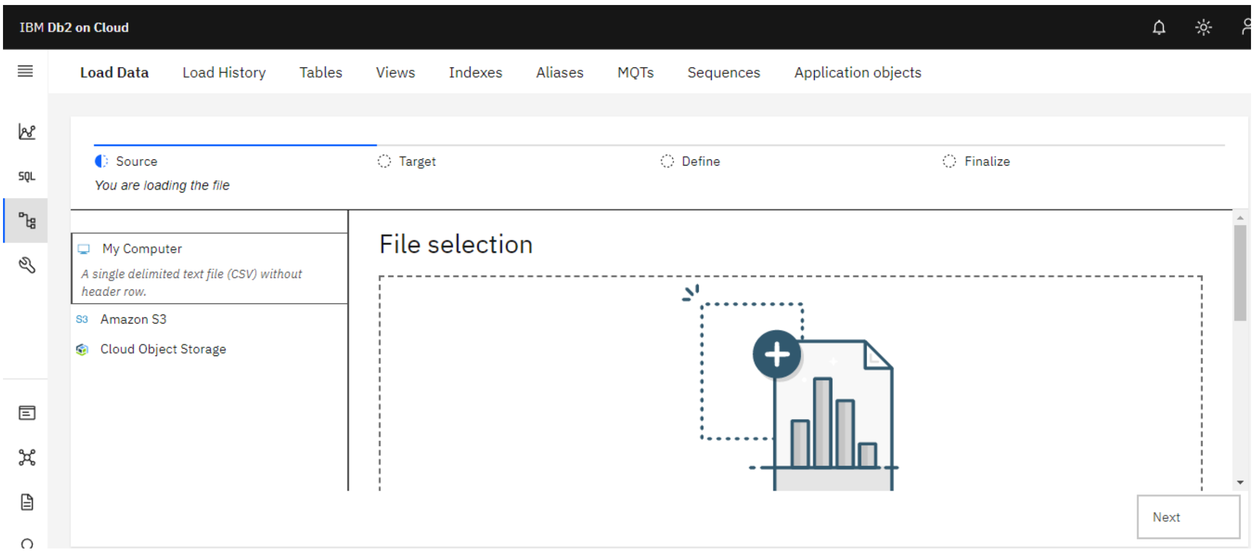
PHASE 3:

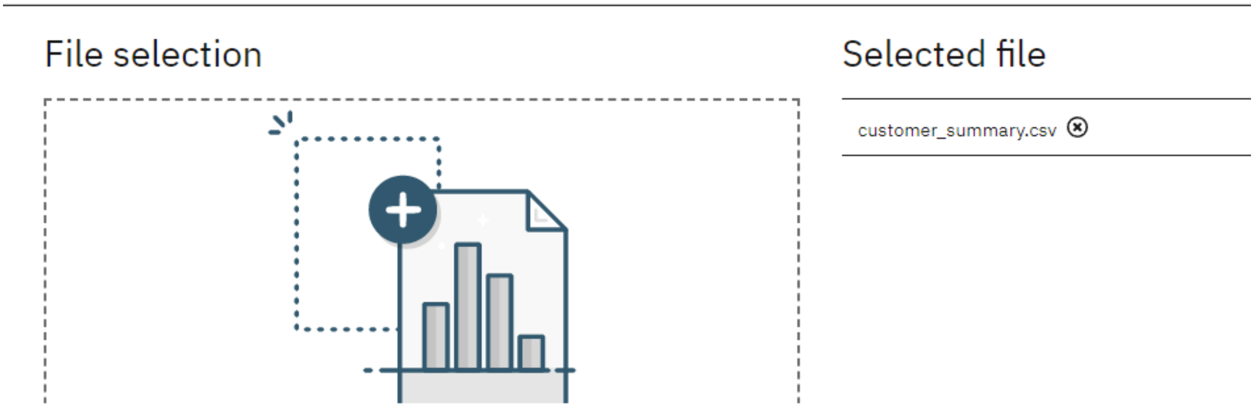


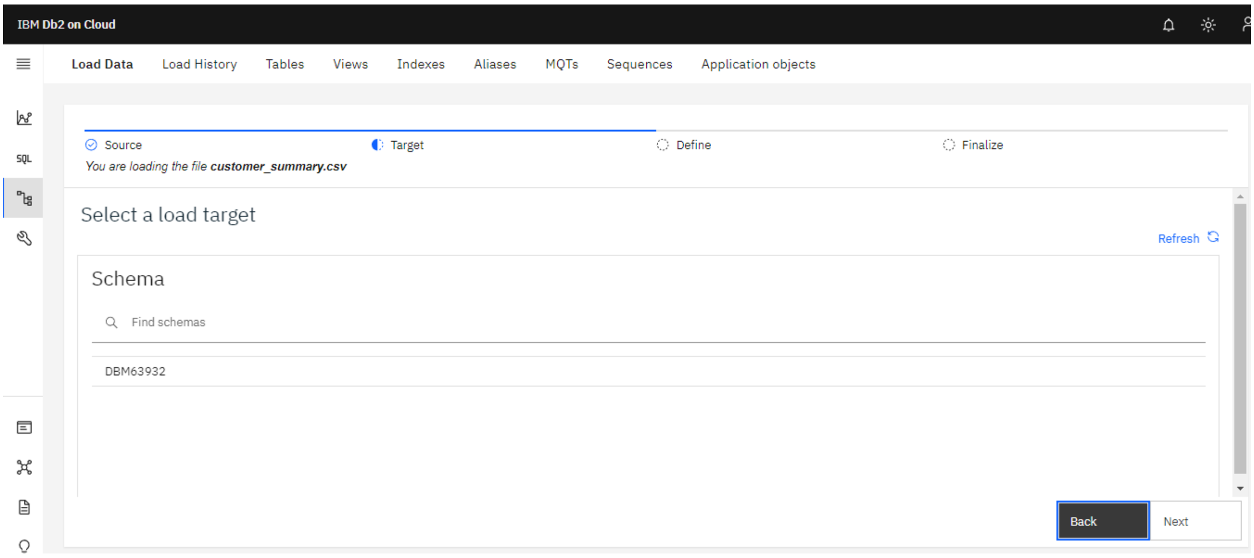


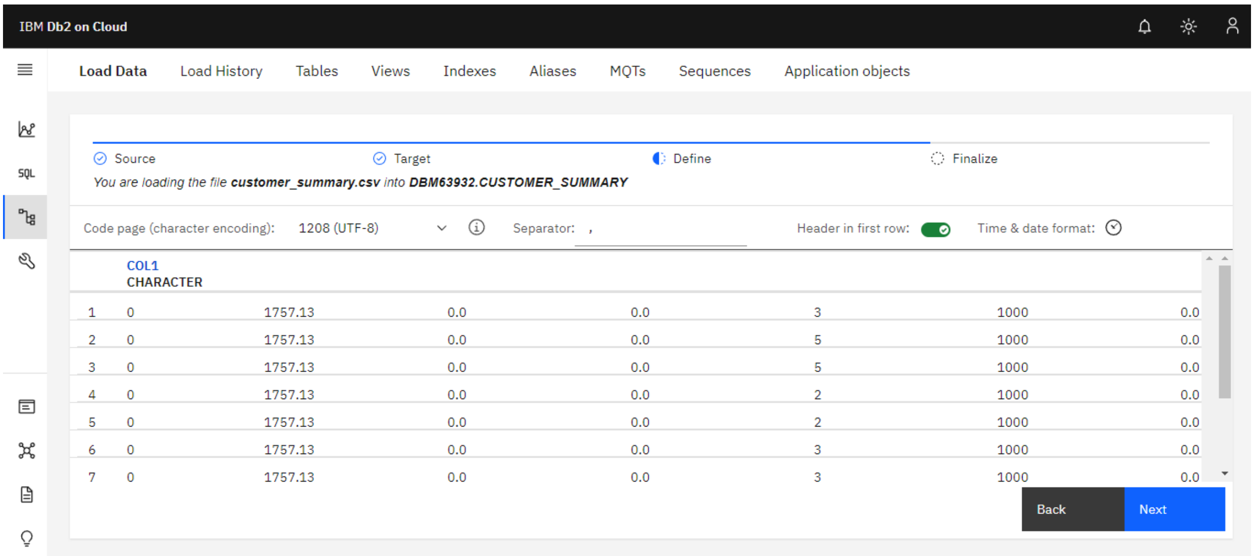

```
"-u",
        "dbm63932",
        "-p",
        "W3KPfJv9swuf699o",
        "--ssl",
        "--sslCAFile",
        "1dd14d0c-1b52-4f63-a606-53ecba28771d",
        "--authenticationDatabase",
        "admin",
        "--host",
```

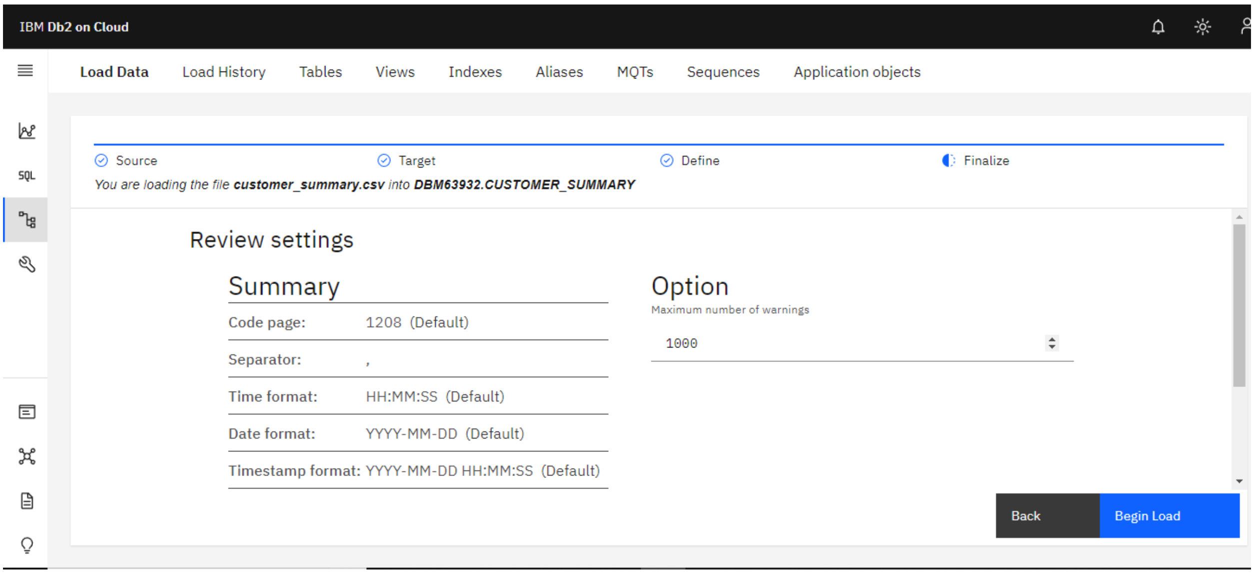

IBM Db2 on Cloud
Overview
In-flight executions
Connections
Table performance
Resource usage
Last 1 hour
0%
Storage (0M / 0M)
current value
Storage usage (%)
100
80
60
40
20
0
10:24 PM
10:40 PM
11:00 PM
11:20 PM
11:24 PM
Time

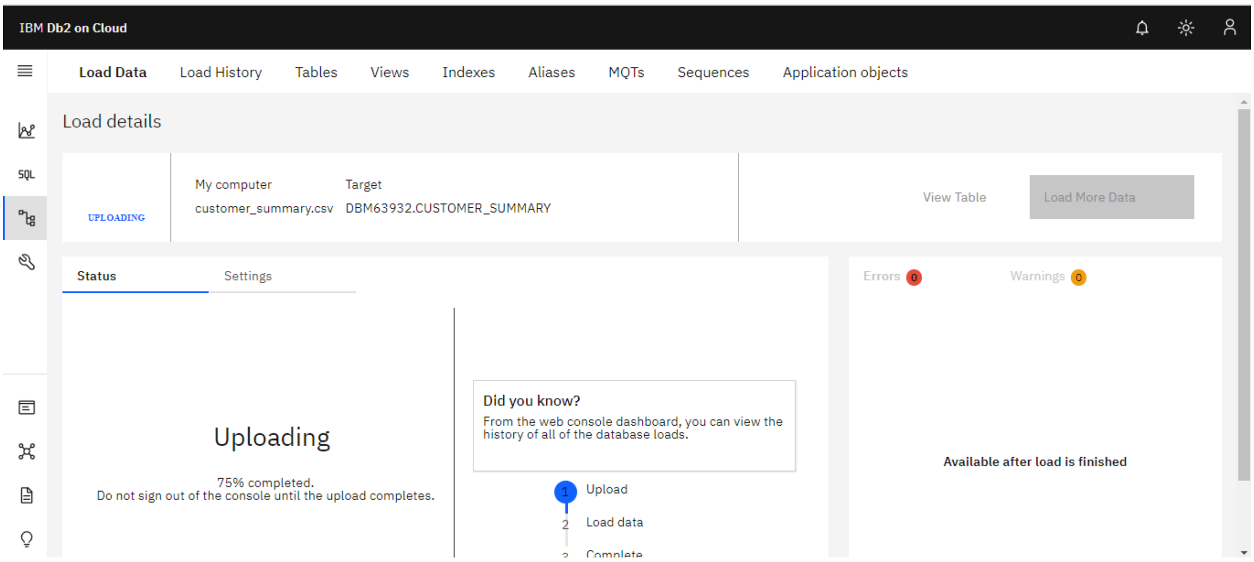

IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
Source
Target
Define
Finalize
You are loading the file
My Computer
A single delimited text file (CSV) without header row.
Amazon S3
Cloud Object Storage
File selection
Next

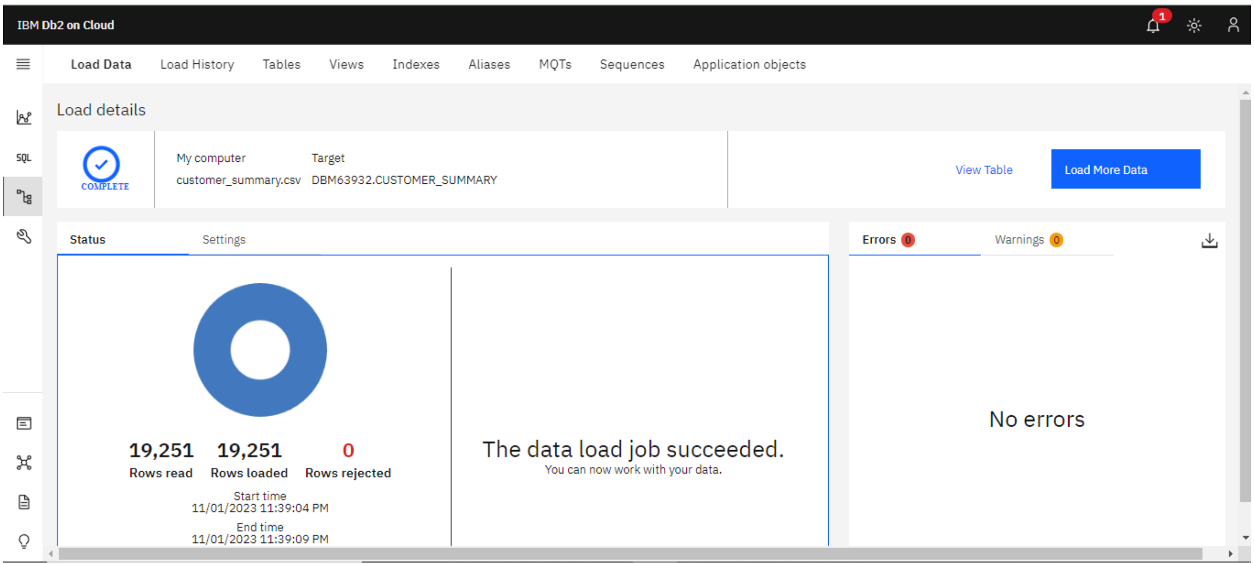

File selection
Selected file
customer_summary.csv


IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
Source
Target
Define
Finalize
You are loading the file customer_summary.csv
Select a load target
Refresh
Schema
Find schemas
DBM63932
Back
Next

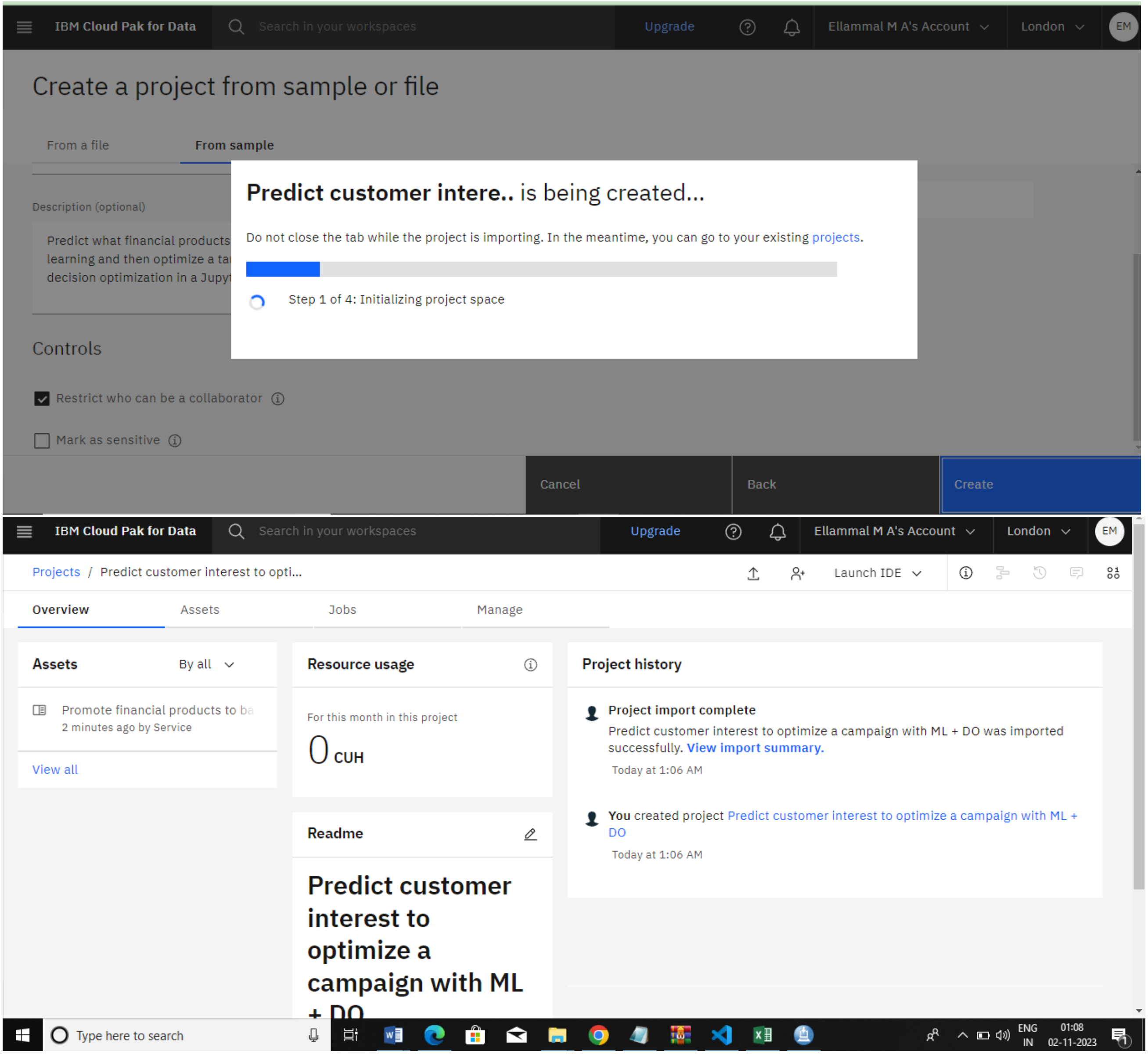

IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
Source
Target
Define
Finalize
You are loading the file customer_summary.csv into DBM63932.CUSTOMER_SUMMARY
Code page (character encoding):
1208 (UTF-8)
Separator:
Header in first row:
Time & date format:
COL1
CHARACTER
1 0 1757.13 0.0 0.0 3 1000 0.0
2 0 1757.13 0.0 0.0 5 1000 0.0
3 0 1757.13 0.0 0.0 5 1000 0.0
4 0 1757.13 0.0 0.0 2 1000 0.0
5 0 1757.13 0.0 0.0 2 1000 0.0
6 0 1757.13 0.0 0.0 3 1000 0.0
7 0 1757.13 0.0 0.0 3 1000 0.0
Back
Next

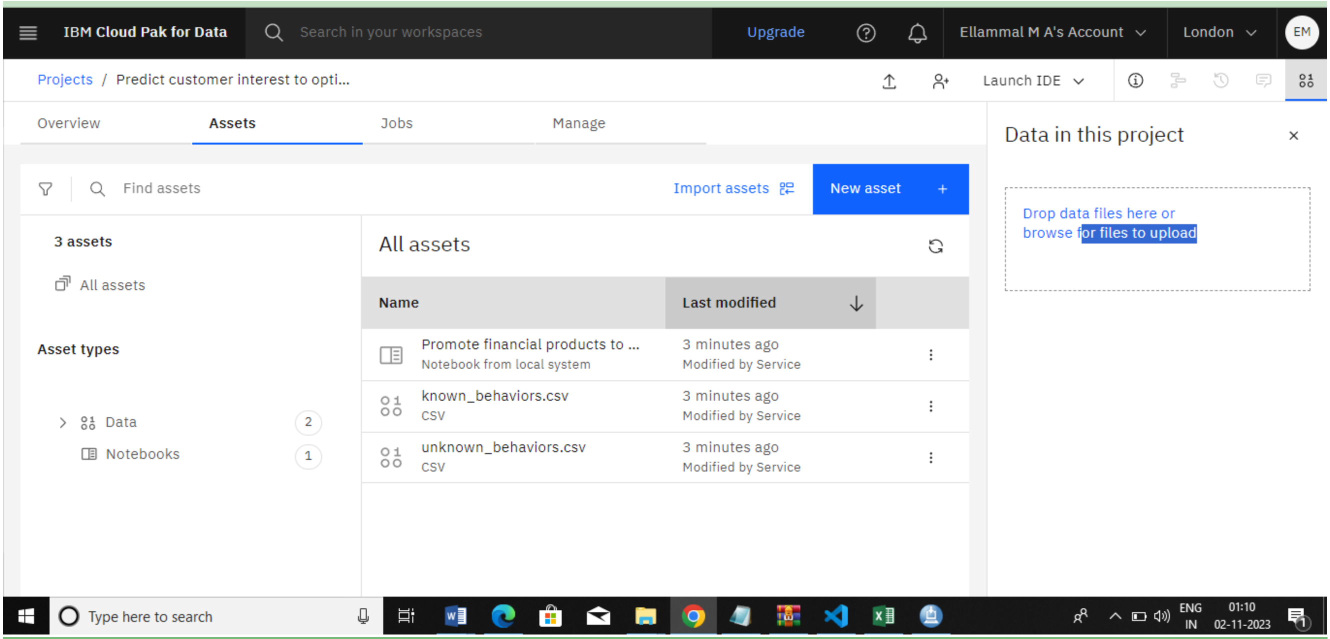

IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
SQL
Source
Target
Define
Finalize
You are loading the file customer_summary.csv into DBM63932.CUSTOMER_SUMMARY
Review settings
Summary
Code page: 1208 (Default)
Separator: ,
Time format: HH:MM:SS (Default)
Date format: YYYY-MM-DD (Default)
Timestamp format: YYYY-MM-DD HH:MM:SS (Default)
Option
Maximum number of warnings
1000
Back
Begin Load

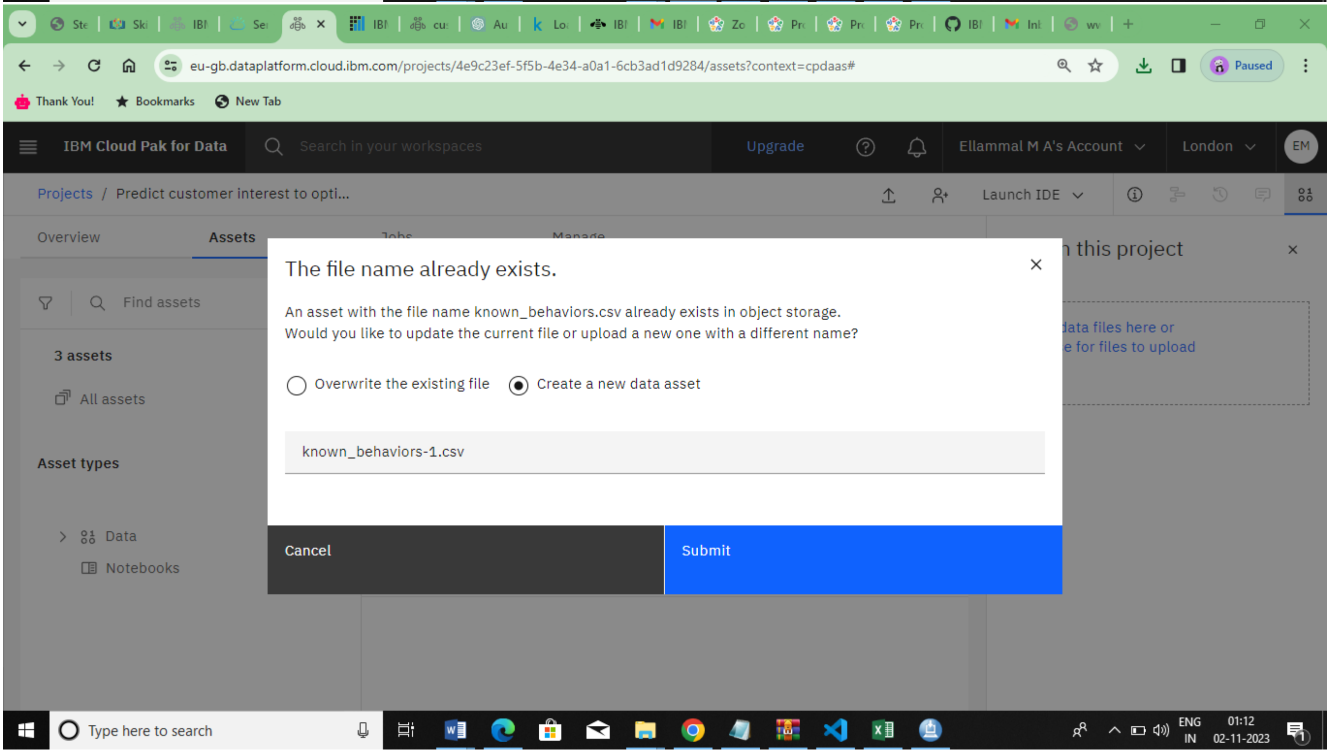

IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
SQL
Load details
UPLOADING
My computer
customer_summary.csv
Target
DBM63932.CUSTOMER_SUMMARY
View Table
Load More Data
Status
Settings
Errors 0
Warnings 0
Uploading
75% completed.
Do not sign out of the console until the upload completes.
Did you know?
From the web console dashboard, you can view the history of all of the database loads.
1 Upload
2 Load data
Available after load is finished

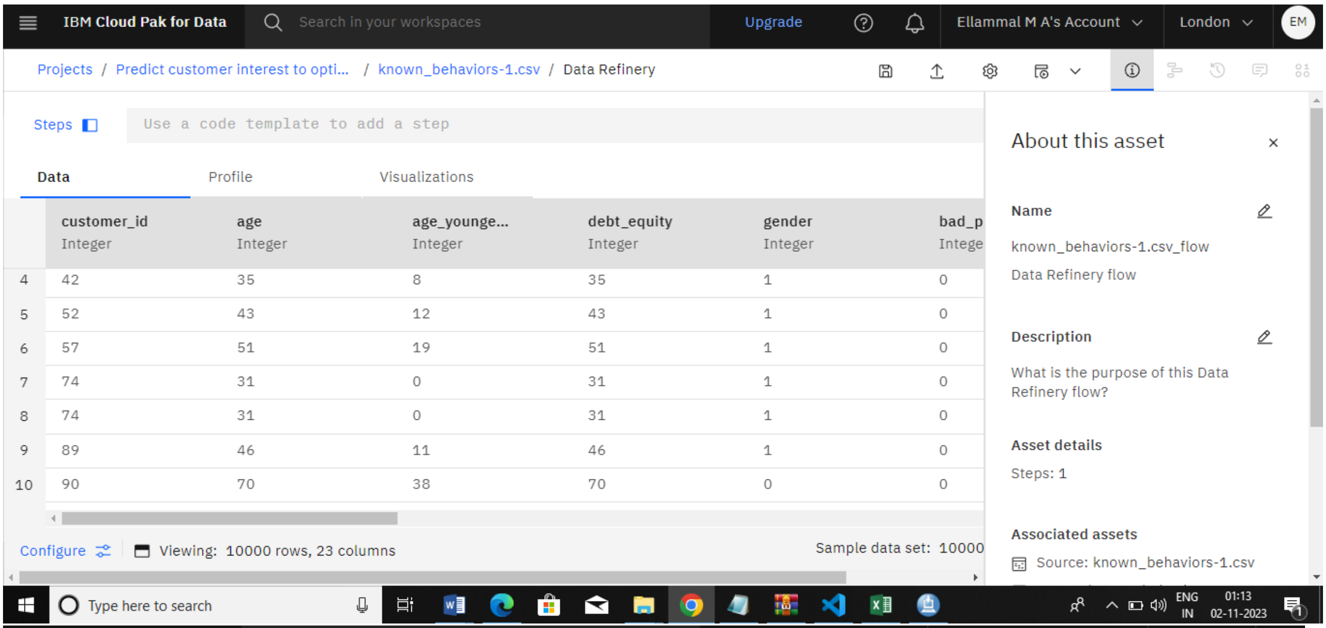

IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
SQL
Load details
COMPLETE
My computer
customer_summary.csv
Target
DBM63932.CUSTOMER_SUMMARY
View Table
Load More Data
Status
Settings
Errors 0
Warnings 0
19,251
Rows read
19,251
Rows loaded
0
Rows rejected
Start time
11/01/2023 11:39:04 PM
End time
11/01/2023 11:39:09 PM
The data load job succeeded.
You can now work with your data.
No errors

PHASE-4

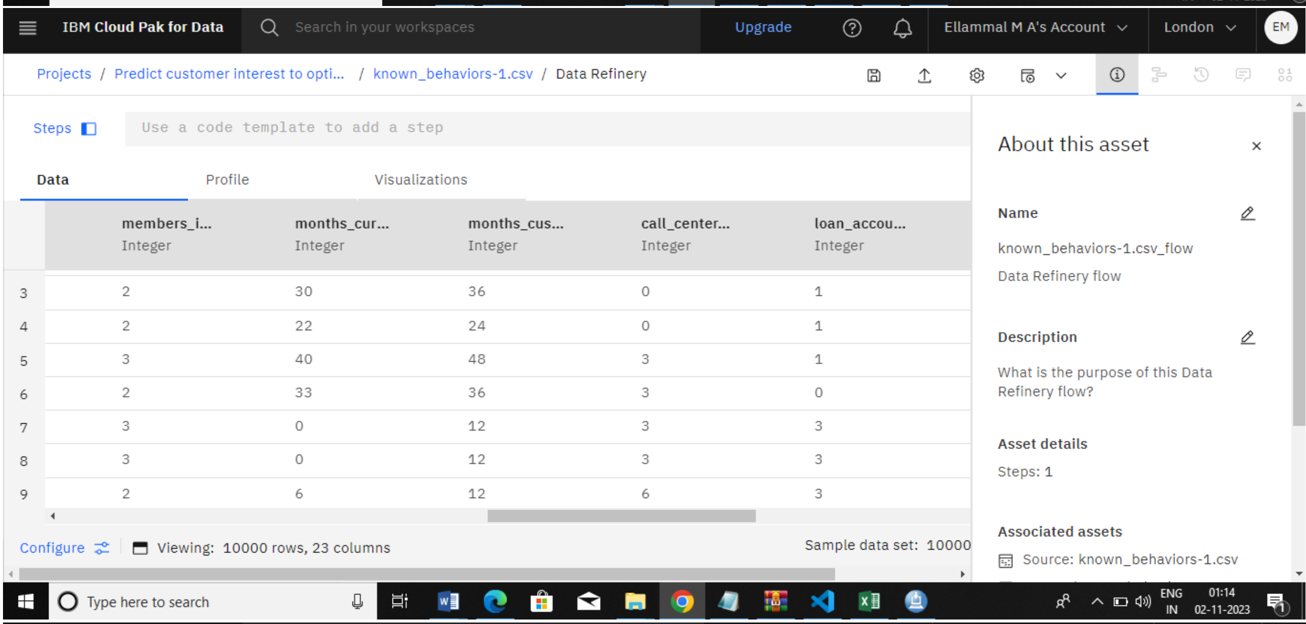

IBM Cloud Pak for Data
Create a project from sample or file
From a file
From sample
Predict customer intere.. is being created...
Do not close the tab while the project is importing. In the meantime, you can go to your existing projects.
Step 1 of 4: Initializing project space
Controls
Restrict who can be a collaborator
Mark as sensitive
Cancel
Back
Create
Projects / Predict customer interest to opti...
Overview
Assets
Jobs
Manage
Assets
Promote financial products to be
2 minutes ago by Service
View all
Resource usage
For this month in this project
0 CUH
Readme
Predict customer interest to optimize a campaign with ML + DO
Project history
Project import complete
Predict customer interest to optimize a campaign with ML + DO was imported successfully. View import summary.
Today at 1:06 AM
You created project Predict customer interest to optimize a campaign with ML + DO
Today at 1:06 AM


IBM Cloud Pak for Data
Search in your workspaces
Upgrade
Ellammal M A's Account
London
EM
Projects / Predict customer interest to opti...
Launch IDE
Overview
Assets
Jobs
Manage
Data in this project
Find assets
Import assets
New asset +
Drop data files here or
browse for files to upload
3 assets
All assets
All assets
Name
Last modified
Asset types
Promote financial products to ...
Notebook from local system
3 minutes ago
Modified by Service
known_behaviors.csv
CSV
3 minutes ago
Modified by Service
unknown_behaviors.csv
CSV
3 minutes ago
Modified by Service
Data
2
Notebooks
1
Type here to search
ENG
IN
01:10
02-11-2023

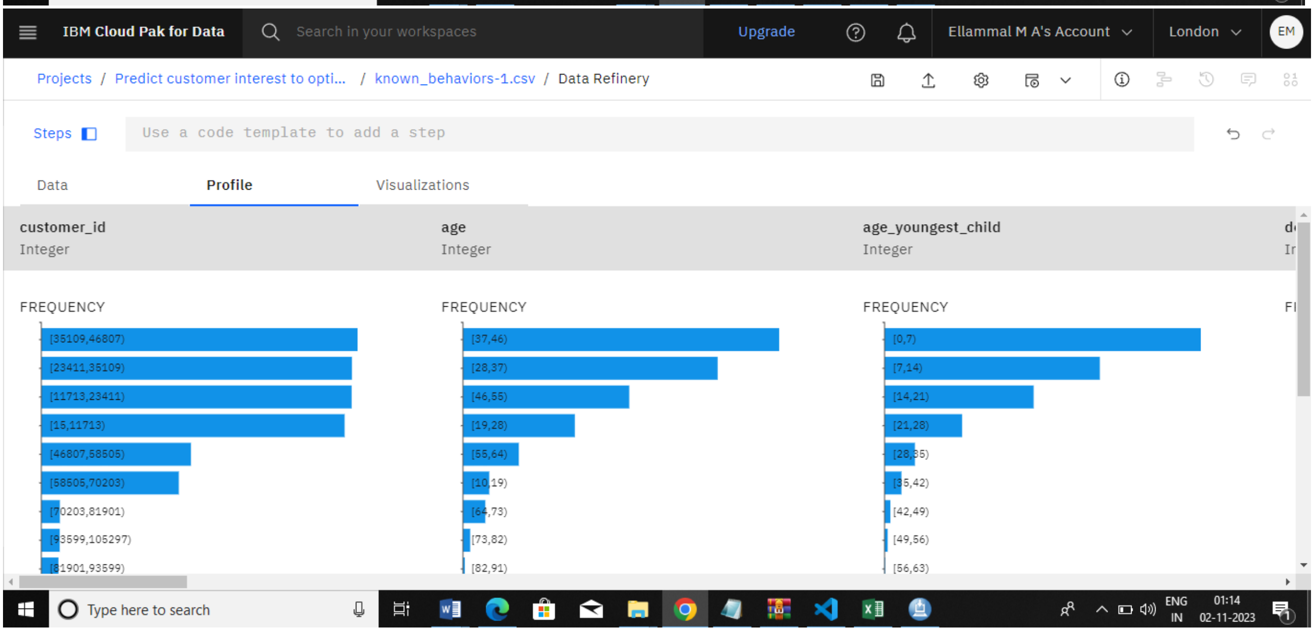

eu-gb.dataplatform.cloud.ibm.com/projects/4e9c23ef-5f5b-4e34-a0a1-6cb3ad1d9284/assets?context=cpdaas#
Paused
Thank You!
Bookmarks
New Tab
IBM Cloud Pak for Data
Search in your workspaces
Upgrade
Ellammal M A's Account
London
EM
Projects / Predict customer interest to opti...
Launch IDE
Overview
Assets
The file name already exists.
An asset with the file name known_behaviors.csv already exists in object storage.
Would you like to update the current file or upload a new one with a different name?
Overwrite the existing file
Create a new data asset
known_behaviors-1.csv
Cancel
Submit
Find assets
3 assets
All assets
Asset types
Data
Notebooks
Type here to search
ENG
IN
01:12
02-11-2023

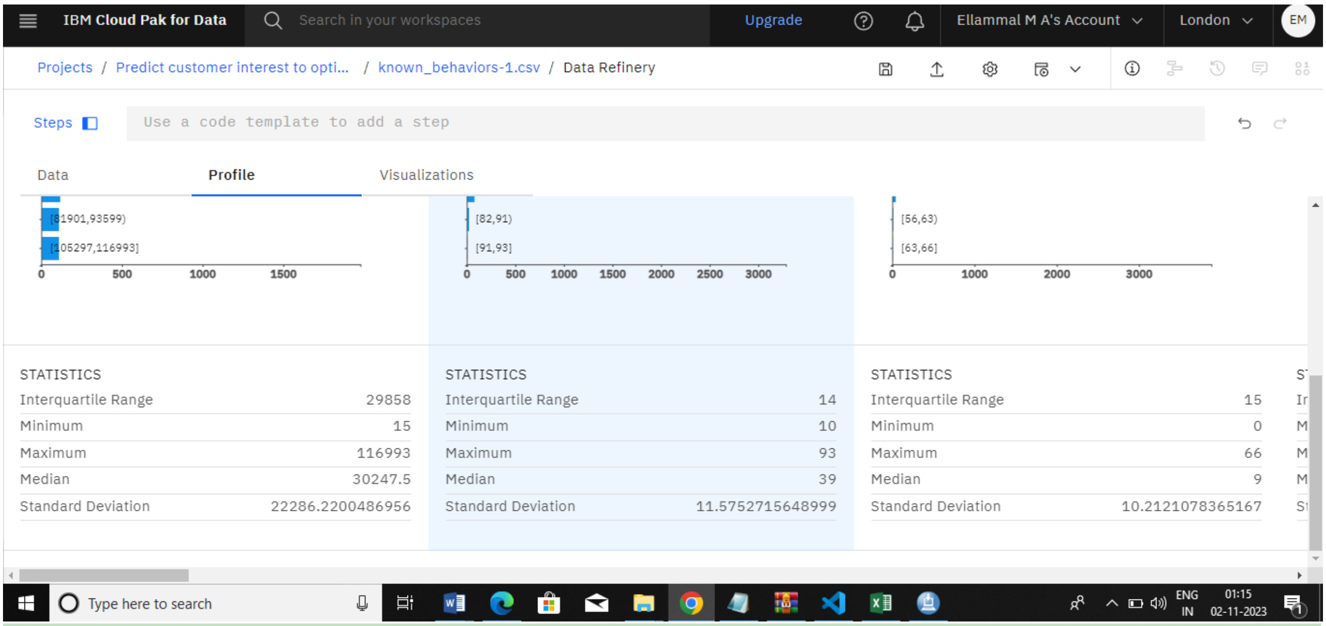


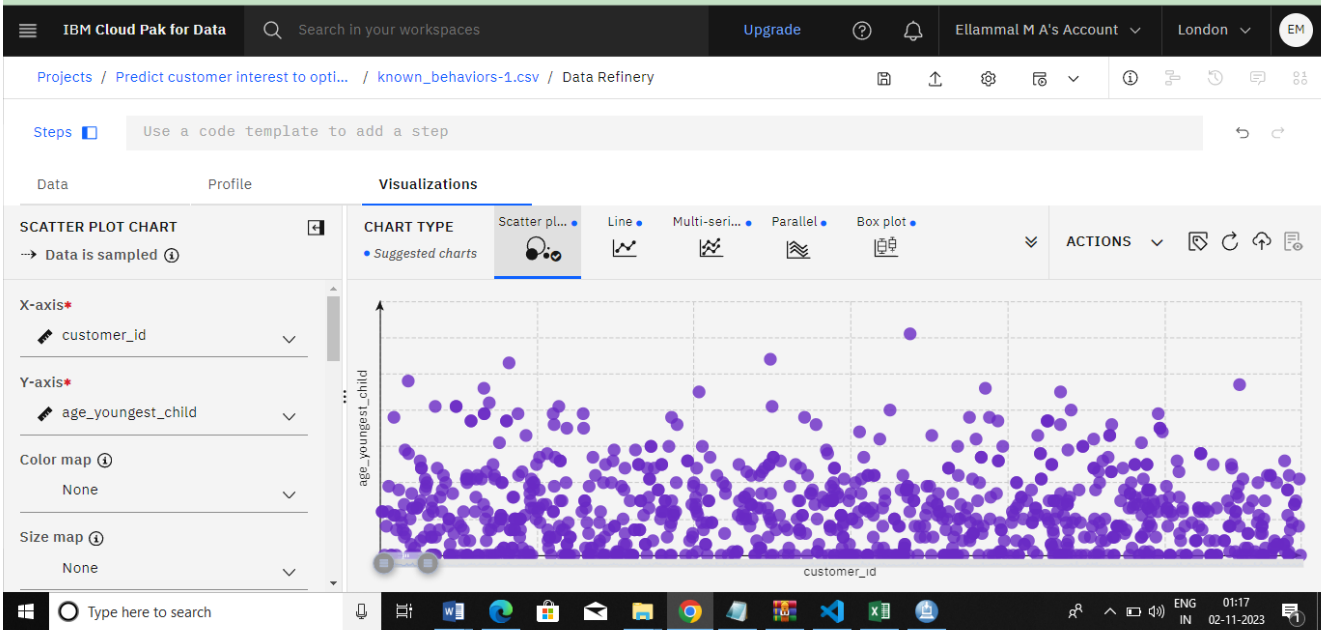
	customer_id Integer	age Integer	age_younge... Integer	debt_equity Integer	gender Integer	bad_p Intege
4	42	35	8	35	1	0
5	52	43	12	43	1	0
6	57	51	19	51	1	0
7	74	31	0	31	1	0
8	74	31	0	31	1	0
9	89	46	11	46	1	0
10	90	70	38	70	0	0

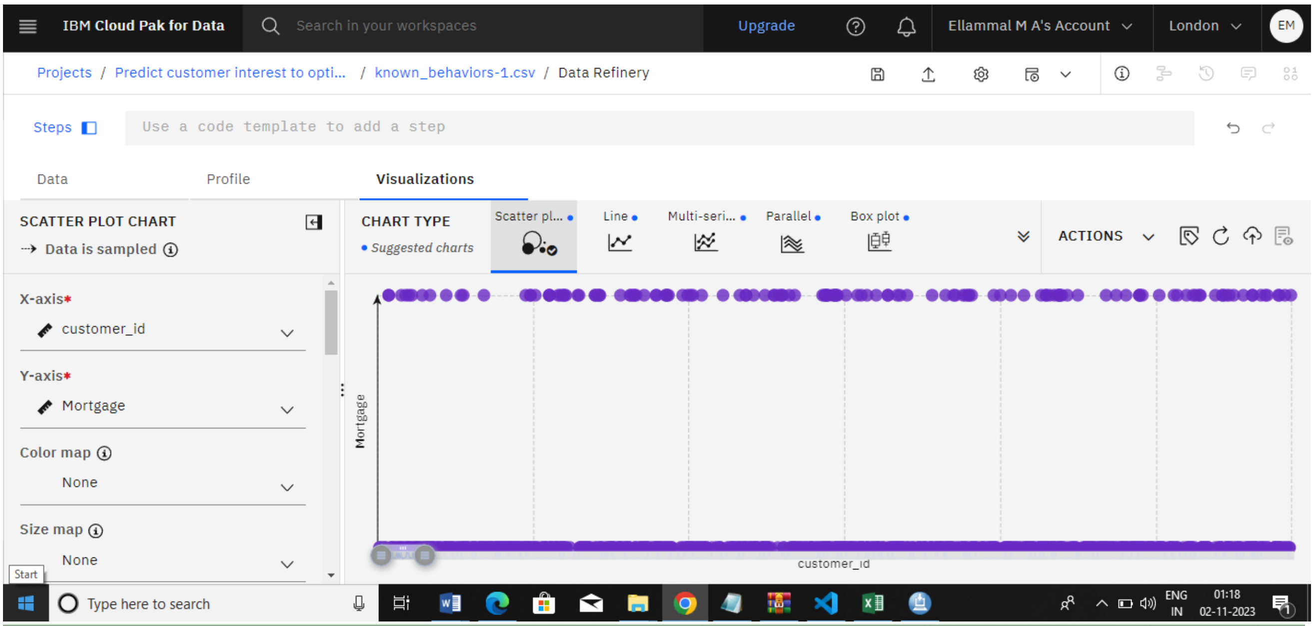


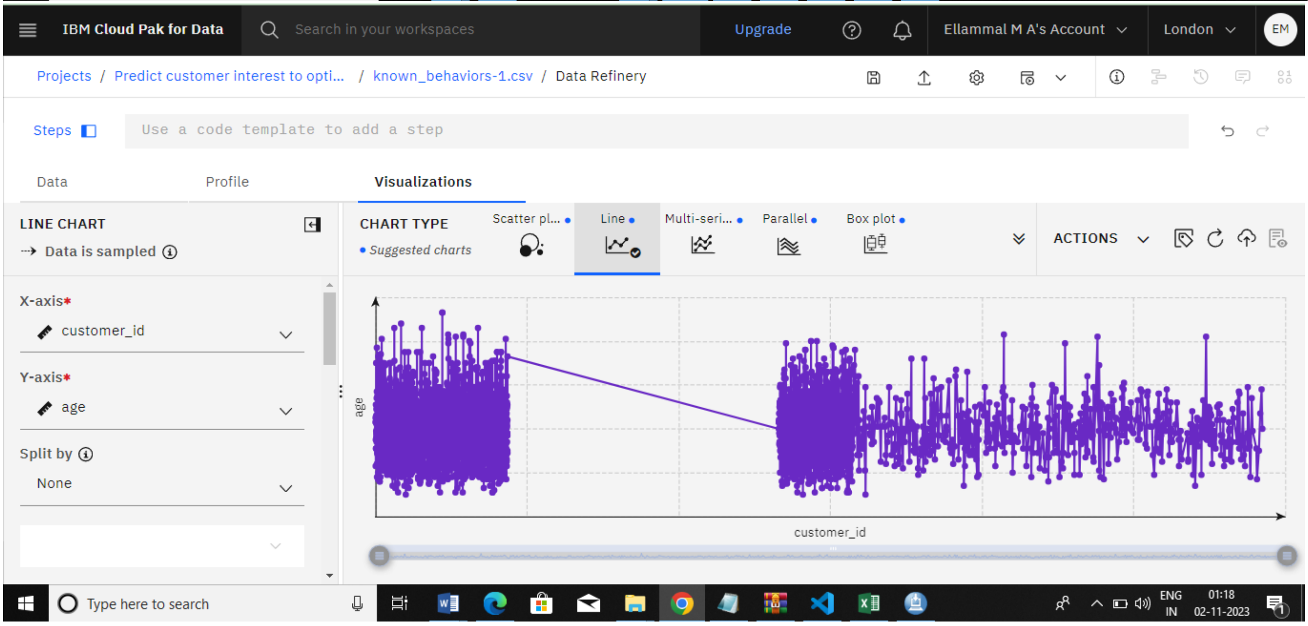
	members_i... Integer	months_cur... Integer	months_cus... Integer	call_center... Integer	loan_accou... Integer
3	2	30	36	0	1
4	2	22	24	0	1
5	3	40	48	3	1
6	2	33	36	3	0
7	3	0	12	3	3
8	3	0	12	3	3
9	2	6	12	6	3

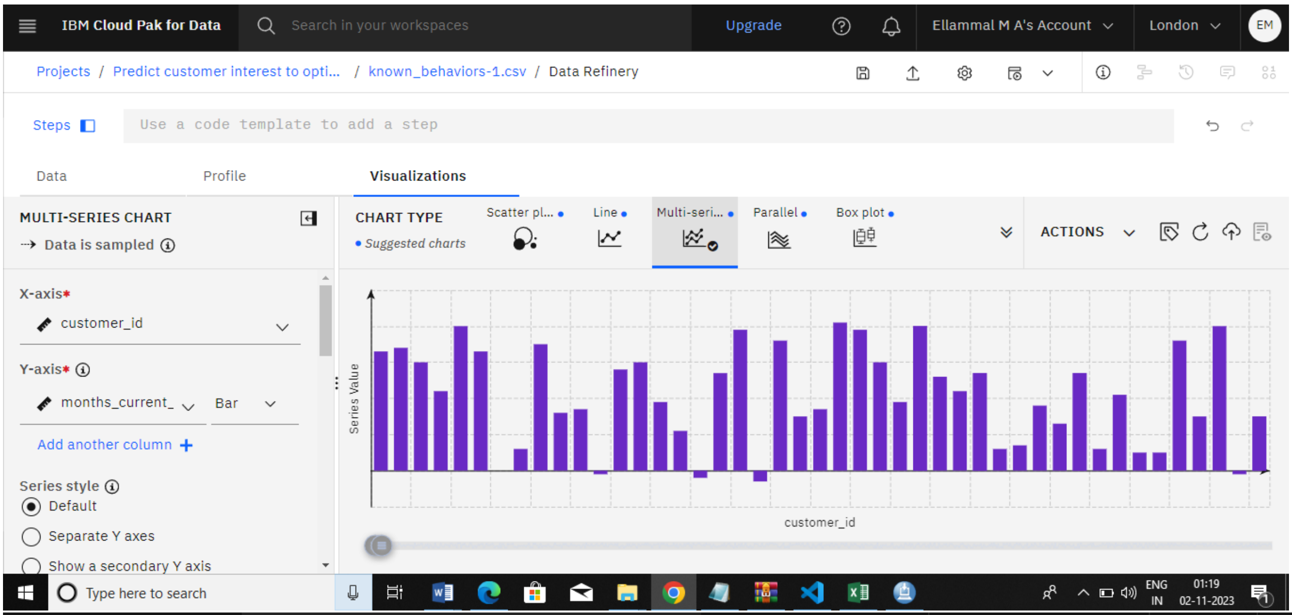


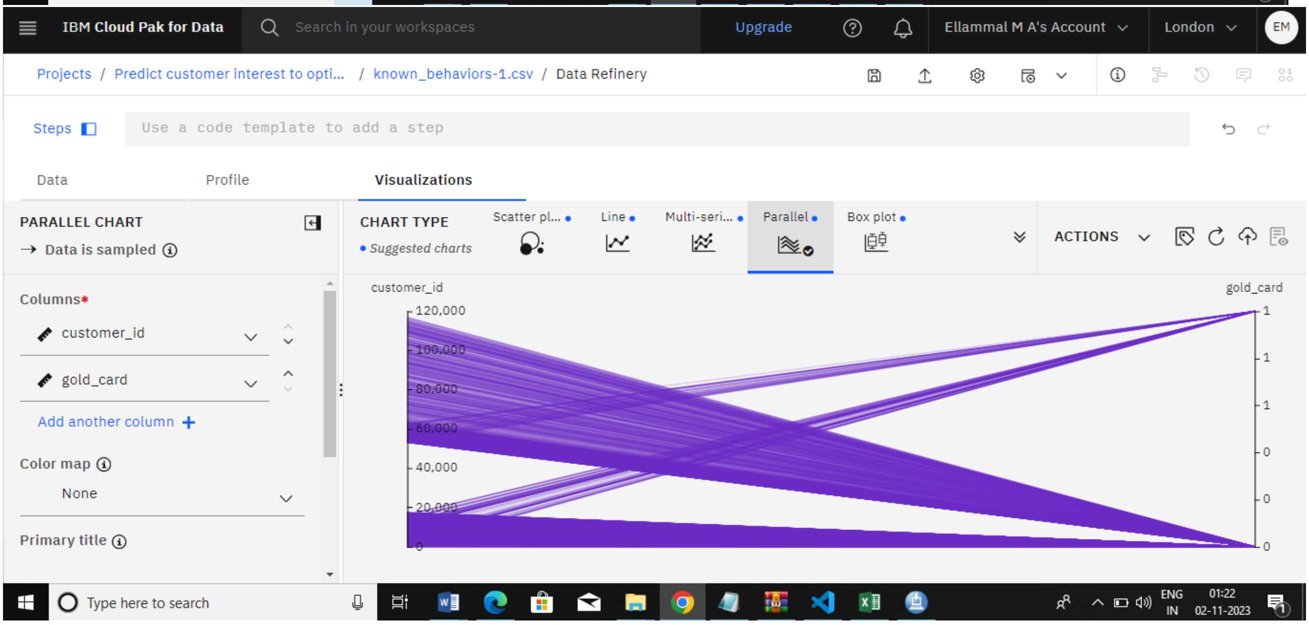

IBM Cloud Pak for Data
Search in your workspaces
Upgrade
Ellammal M A's Account
London
Projects / Predict customer interest to opti... / known_behaviors-1.csv / Data Refinery
Steps
Use a code template to add a step
Data
Profile
Visualizations
STATISTICS
Interquartile Range 29858
Minimum 15
Maximum 116993
Median 30247.5
Standard Deviation 22286.2200486956
STATISTICS
Interquartile Range 14
Minimum 10
Maximum 93
Median 39
Standard Deviation 11.5752715648999
STATISTICS
Interquartile Range 15
Minimum 0
Maximum 66
Median 9
Standard Deviation 10.2121078365167
Type here to search
01:15
02-11-2023


IBM Cloud Pak for Data
Search in your workspaces
Upgrade
Ellammal M A's Account
London
Projects / Predict customer interest to opti... / known_behaviors-1.csv / Data Refinery
Steps
Use a code template to add a step
Data
Profile
Visualizations
SCATTER PLOT CHART
Data is sampled
X-axis*
customer_id
Y-axis*
age_youngest_child
Color map
None
Size map
None
CHART TYPE
Suggested charts
Scatter pl...
Line
Multi-seri...
Parallel
Box plot
ACTIONS
age_youngest_child
customer_id
Type here to search
01:17
02-11-2023


IBM Cloud Pak for Data
Search in your workspaces
Upgrade
Ellammal M A's Account
London
Projects / Predict customer interest to opti... / known_behaviors-1.csv / Data Refinery
Steps
Use a code template to add a step
Data
Profile
Visualizations
SCATTER PLOT CHART
Data is sampled
CHART TYPE
Suggested charts
Scatter pl...
Line
Multi-seri...
Parallel
Box plot
ACTIONS
X-axis*
customer_id
Y-axis*
Mortgage
Color map
None
Size map
None
Mortgage
customer_id
Type here to search
ENG IN
01:18
02-11-2023


IBM Cloud Pak for Data
Search in your workspaces
Upgrade
Ellammal M A's Account
London
Projects / Predict customer interest to opti... / known_behaviors-1.csv / Data Refinery
Steps
Use a code template to add a step
Data
Profile
Visualizations
LINE CHART
Data is sampled
CHART TYPE
Suggested charts
Scatter pl...
Line
Multi-seri...
Parallel
Box plot
ACTIONS
X-axis*
customer_id
Y-axis*
age
Split by
None
age
customer_id
Type here to search
ENG IN
01:18
02-11-2023





Conclusion

In conclusion, the application of advanced machine learning models for customer churn prediction in telecommunication companies represents a pivotal step towards fostering customer retention, optimizing business operations, and maximizing profitability. the synergy of data-driven insights, predictive analytics, and proactive customer engagement offers a compelling avenue for telecommunications companies to mitigate the challenges of customer attrition. By harnessing the power of data, implementing robust machine learning frameworks, and continuously refining predictive models, these organizations can not only identify at-risk customers but also tailor strategic interventions that lead to enhanced customer satisfaction and long-term loyalty, ultimately propelling them towards sustained success in an ever-evolving industry landscape.