

## Project Development Phase

### Sprint 1

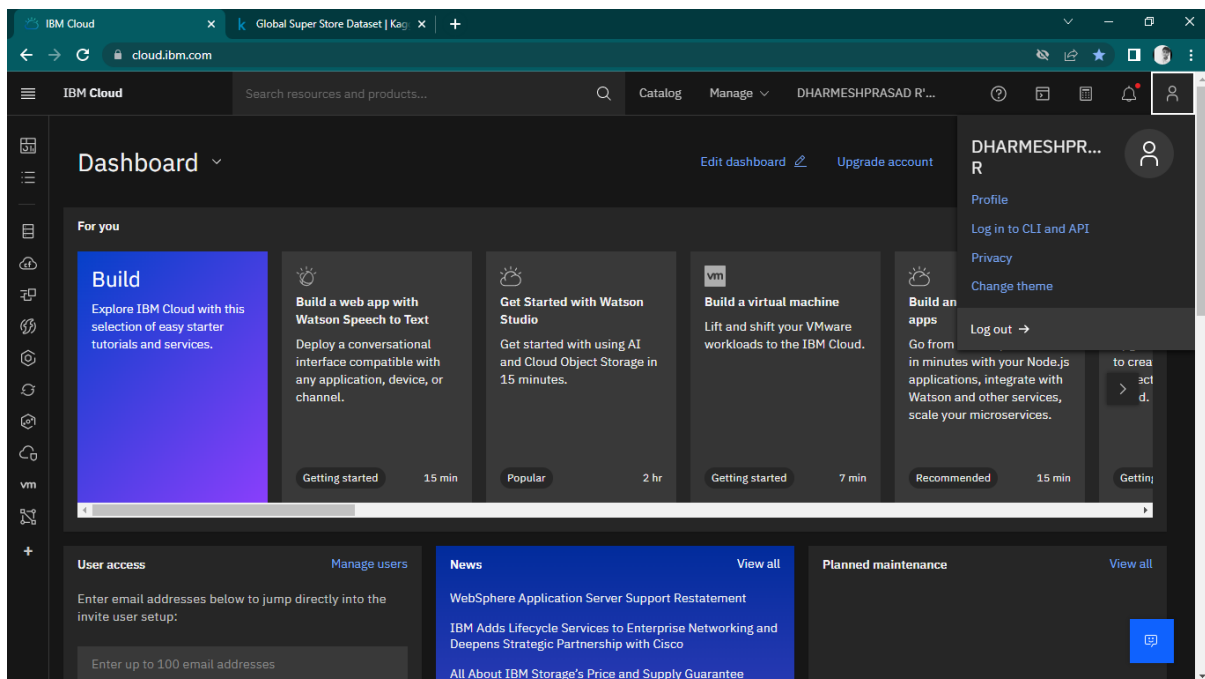
Date	25 October 2022
Team ID	PNT2022TMID41225
Project Name	Global Sales Data Analytics

#### Sprint 1 tasks:

- IBM Cloud Account creation
- IBM Cognos Account creation
- Fetch Data from external API
- Downloading dataset
- Data Collection
- Understanding dataset
- Create Data module
- Uploading dataset
- Login page creation

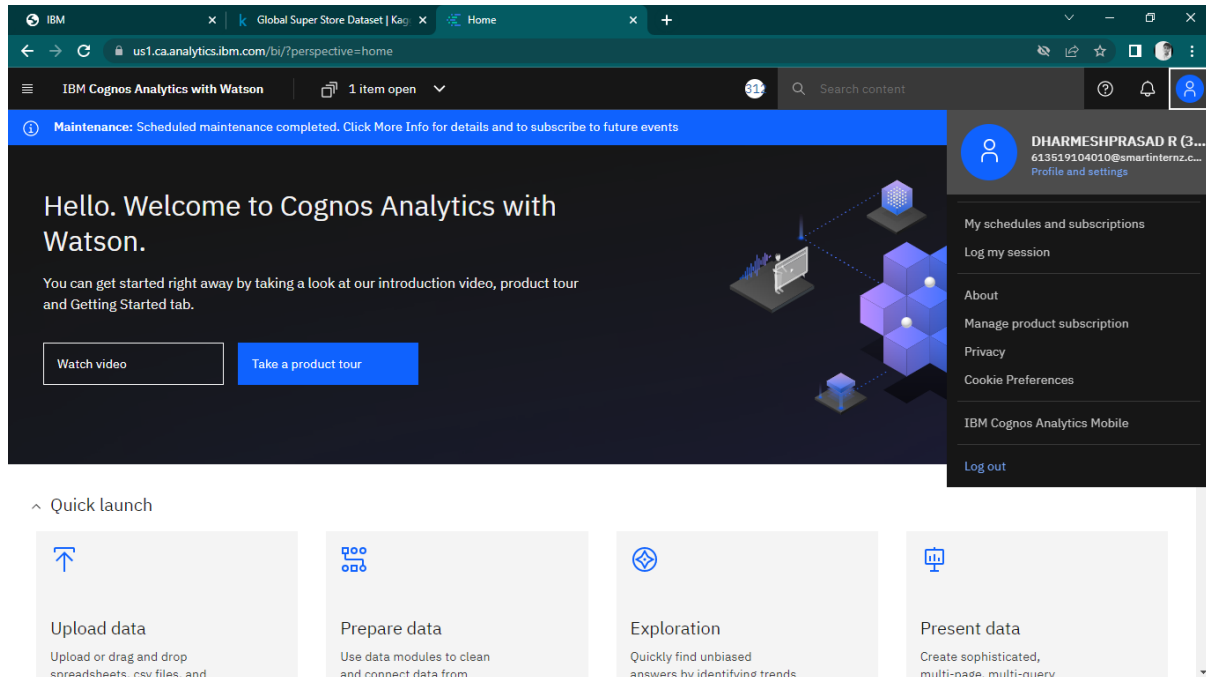
#### IBM Cloud Account:

Creating and login to IBM cloud account.



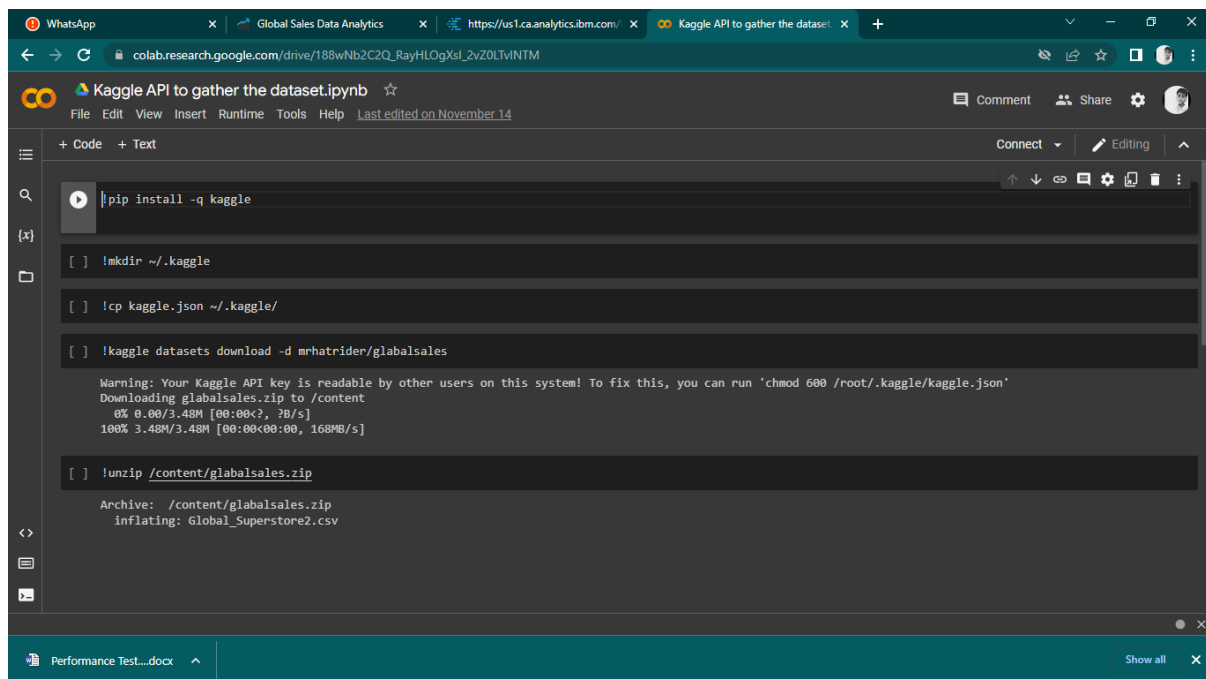
## IBM Cognos Account:

Creating and login to IBM Cognos account.



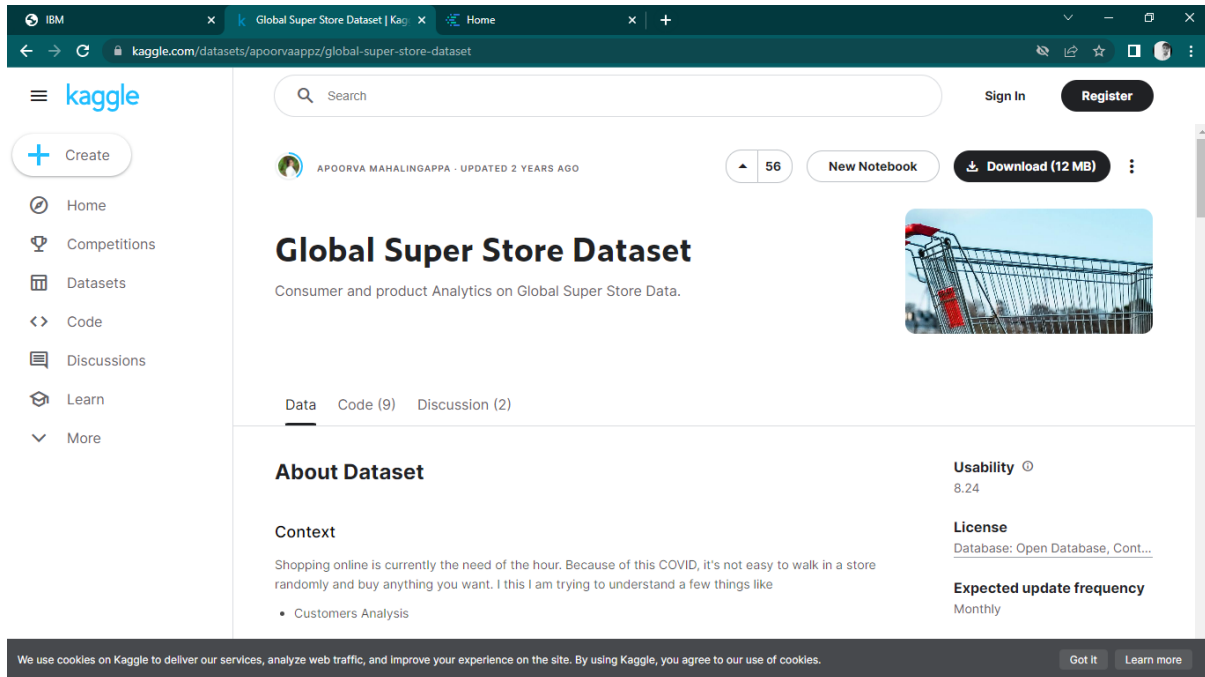
## Fetch Data from external API:

Fetching data from external kaggle API



## Downloading Dataset:

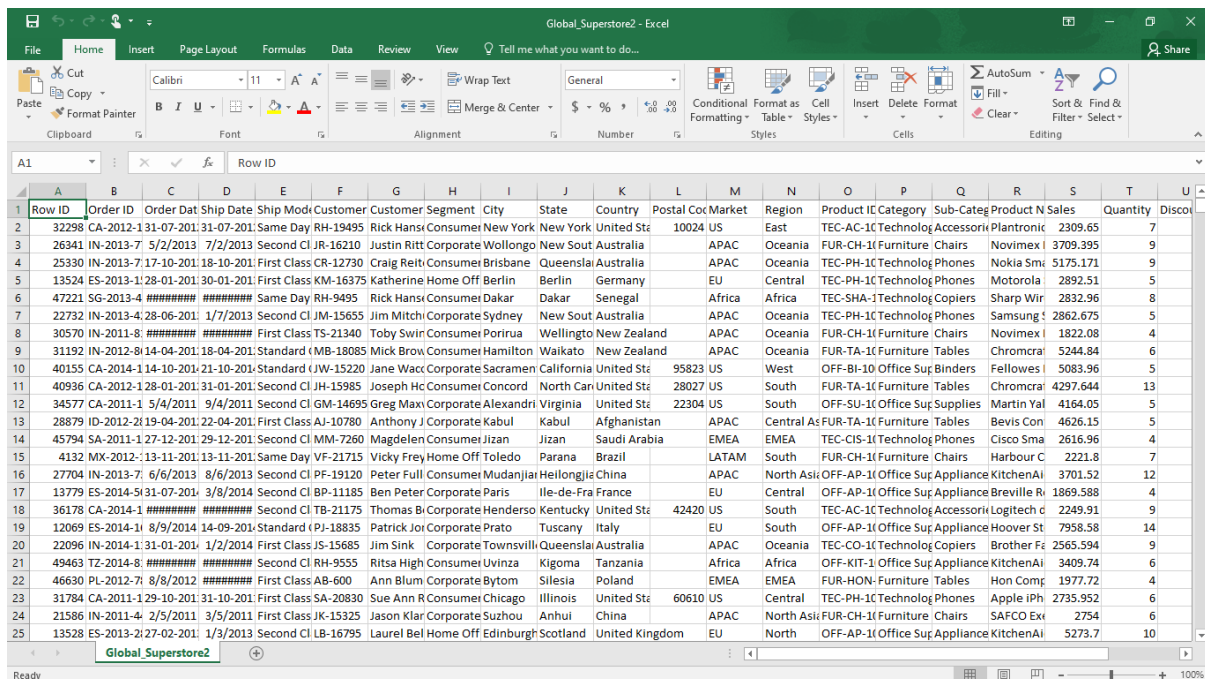
Downloading Global super store dataset from kaggle.



The screenshot shows the Kaggle dataset page for 'Global Super Store Dataset'. The page includes a search bar, navigation links (Home, Competitions, Datasets, Code, Discussions, Learn, More), and a sidebar with a 'Create' button. The main content area displays the dataset title, a description, and a 'Download (12 MB)' button. Below this, there are sections for 'About Dataset', 'Context', 'Usability', 'License', and 'Expected update frequency'. The 'Context' section mentions that shopping online is currently the need of the hour due to COVID-19. The 'Usability' section shows a score of 8.24. The 'License' section indicates the database is Open Database, Cont... and the 'Expected update frequency' is Monthly.

## Data collection:

Understanding the dataset and analysing the data requirements.



The screenshot shows an Excel spreadsheet titled 'Global\_Superstore2 - Excel'. The spreadsheet contains data for the Global Super Store Dataset. The columns are: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer, Customer Segment, City, State, Country, Postal Code, Market, Region, Product ID, Category, Sub-Category, Product Name, Sales, Quantity, and Discount. The data is organized into rows, with the first row being the header row. The spreadsheet shows a large number of rows of data, with the first few rows visible.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer	Customer Segment	City	State	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount
1	32298	CA-2012-1-31-07-2013	31-07-2013	Same Day	Rick Hansi	Consumer	New York	New York	United States	10024	US	East	TEC-AC-1C	Technology	Accessories	Plantronic	2309.65	7	
2	26341	IN-2013-7-5/2-2013	7/2-2013	Second Class	Justin Ritt	Corporate	Wollongong	New South Wales	Australia		APAC	Oceania	FUR-CH-1F	Furniture	Chairs	Novimex	3709.395	9	
3	25330	IN-2013-7-17-10-2013	18-10-2013	First Class	Craig Relti	Consumer	Brisbane	Queensland	Australia		APAC	Oceania	TEC-PH-1C	Technology	Phones	Nokia Smartphone	5175.171	9	
4	13524	ES-2013-1-28-01-2013	30-01-2013	First Class	Katherine Home	Off Berlin	Berlin	Germany			EU	Central	TEC-PH-1C	Technology	Phones	Motorola	2892.51	5	
5	47221	SG-2013-4-28-06-2013	1/7/2013	Second Class	Jim Mitche	Corporate	Sydney	New South Wales	Australia		APAC	Oceania	TEC-SHA-1T	Technology	Copiers	Sharp Wireless	2832.96	8	
6	22732	IN-2013-4-28-06-2013	1/7/2013	Second Class	Jim Mitche	Corporate	Sydney	New South Wales	Australia		APAC	Oceania	TEC-PH-1C	Technology	Phones	Samsung	2862.675	5	
7	30570	IN-2011-8-11-04-2011	11-04-2011	First Class	Toby Swin	Consumer	Porirua	Wellington	New Zealand		APAC	Oceania	FUR-CH-1F	Furniture	Chairs	Novimex	1822.08	4	
8	31192	IN-2012-8-14-04-2013	18-04-2013	Standard	Mick Brow	Consumer	Hamilton	Waikato	New Zealand		APAC	Oceania	FUR-TA-1F	Furniture	Tables	Chromcraft	5244.84	6	
9	40155	CA-2014-1-14-10-2013	21-10-2013	Standard	Jane Wacc	Corporate	Sacramento	California	United States	95823	US	West	OFF-BI-10	Office Supplies	Binders	Fellowes	5083.96	5	
10	40936	CA-2012-1-28-01-2013	31-01-2013	Second Class	Joseph Hc	Consumer	Concord	North Carolina	United States	28027	US	South	FUR-TA-1F	Furniture	Tables	Chromcraft	4297.644	13	
11	34577	CA-2011-1-5/4-2011	9/4-2011	Second Class	Greg Maxi	Corporate	Alexandria	Virginia	United States	22304	US	South	OFF-SU-1C	Office Supplies	Supplies	Martin's	4164.05	5	
12	28879	ID-2012-2-19-04-2013	22-04-2013	First Class	Anthony J	Corporate	Kabul	Kabul	Afghanistan		APAC	Central Asia	FUR-TA-1F	Furniture	Tables	Bevis Con	4626.15	5	
13	45794	SA-2011-1-27-12-2013	29-12-2013	Second Class	Magdalen	Consumer	Jizan	Jizan	Saudi Arabia		EMEA	EMEA	TEC-CIS-1F	Technology	Phones	Cisco Smartphone	2616.96	4	
14	4132	MX-2012-1-13-11-2013	13-11-2013	Same Day	Vicky Frey	Home Office	Toledo	Parana	Brazil		LATAM	South	FUR-CH-1F	Furniture	Chairs	Harbour Craft	2221.8	7	
15	27704	IN-2013-7-6/6-2013	8/6-2013	Second Class	Peter Full	Consumer	Mudanjiang	Heilongjiang	China		APAC	North Asia	OFF-AP-1F	Office Supplies	Appliance	KitchenAid	3701.52	12	
16	13779	ES-2014-5-31-07-2013	3/8-2014	Second Class	Ben Peter	Corporate	Paris	Ile-de-France	France		EU	Central	OFF-AP-1F	Office Supplies	Appliance	Breville	1869.588	4	
17	36178	CA-2014-1-11-09-2014	14-09-2014	Standard	Patrick Jor	Corporate	Prato	Tuscany	Italy		EU	South	TEC-AC-1C	Technology	Accessories	Logitech	2249.91	9	
18	12069	ES-2014-1-8/9-2014	14-09-2014	Standard	Patrick Jor	Corporate	Prato	Tuscany	Italy		EU	South	OFF-AP-1F	Office Supplies	Appliance	Hoover	7958.58	14	
19	22096	IN-2014-1-31-01-2013	1/2-2014	First Class	Jim Sink	Corporate	Townsville	Queensland	Australia		APAC	Oceania	TEC-CO-1C	Technology	Copiers	Brother	2565.594	9	
20	49463	TZ-2014-8-8/8-2012	8/8-2012	Second Class	Ritsa High	Consumer	Uvinza	Kigoma	Tanzania		AFRICA	Africa	OFF-KIT-1F	Office Supplies	Appliance	KitchenAid	3409.74	6	
21	46630	PL-2012-7-8/8-2012	8/8-2012	First Class	Ann Blum	Corporate	Bytom	Silesia	Poland		EMEA	EMEA	FUR-HON-1F	Furniture	Tables	Hon Comp	1977.72	4	
22	31784	CA-2011-1-29-10-2013	31-10-2013	First Class	Sue Ann R	Consumer	Chicago	Illinois	United States	60610	US	Central	TEC-PH-1C	Technology	Phones	Apple iPhone	2735.952	6	
23	21586	IN-2011-4-2/5-2011	3/5-2011	First Class	Jason Klar	Corporate	Suzhou	Anhui	China		APAC	North Asia	FUR-CH-1F	Furniture	Chairs	SAFCO	2754	6	
24	13528	ES-2013-2-27-02-2013	1/3-2013	Second Class	Laurel Bel	Home Office	Edinburgh	Scotland	United Kingdom		EU	North	OFF-AP-1F	Office Supplies	Appliance	KitchenAid	5273.7	10	

## Understanding the Dataset:

The dataset contains the details of the order done online by people across the globe in the time frame 1-jan-2011 to 31-dec-2014.

The screenshot shows the Kaggle website interface for the 'Global Super Store Dataset'. The page includes a search bar, navigation links (Home, Competitions, Datasets, Code, Discussions, Learn, More), and a sidebar with a 'Create' button. The main content area is titled 'Global Super Store Dataset' and features tabs for 'Data', 'Code (9)', and 'Discussion (2)'. A 'Download (12 MB)' button is visible. The 'About Dataset' section provides context on the dataset's relevance during the COVID-19 pandemic and lists analysis tasks under 'Customers Analysis' and 'Product Analysis'. On the right, 'Usability' (8.24), 'License' (Open Database, Cont...), and 'Expected update frequency' (Monthly) are displayed. A cookie consent banner is at the bottom.

### Global Super Store Dataset

Data Code (9) Discussion (2) 56 New Notebook Download (12 MB)

#### About Dataset

**Context**

Shopping online is currently the need of the hour. Because of this COVID, it's not easy to walk in a store randomly and buy anything you want. I this I am trying to understand a few things like

- Customers Analysis
  - 1. Profile the customers based on their frequency of purchase - calculate frequency of purchase for each customer
  - 2. Do the high frequent customers are contributing more revenue
  - 3. Are they also profitable - what is the profit margin across the buckets
  - 4. Which customer segment is most profitable in each year.
  - 5. How the customers are distributed across the countries -
- Product Analysis
  - 1. Which country has top sales?
  - 2. Which are the top 5 profit-making product types on a yearly basis
  - 3. How is the product price varying with sales - Is there any increase in sales with the decrease in price at a

**Usability** 8.24

**License**  
Database: Open Database, Cont...

**Expected update frequency**  
Monthly

We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies. Got it Learn more

## Create Data module:

The screenshot shows a Jupyter Notebook titled 'data\_process.ipynb' in Visual Studio Code. The notebook is running on a Jupyter Server. The first cell, 'Importing libraries', imports pandas, numpy, seaborn, matplotlib, and sklearn modules. The second cell, 'Loading Dataset', uses pandas to read a CSV file named 'Global.csv'. The third cell displays the 'data' variable, showing a preview of the dataset with columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, City, State, Product ID, Category, and Sub-Category. The bottom status bar indicates 'Jupyter Server: Local' and 'Cell 3 of 31'.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
```

[1] Python

#### Loading Dataset

```
data = pd.read_csv('Global.csv')
```

[2] Python

data

```
data
```

[3] Python

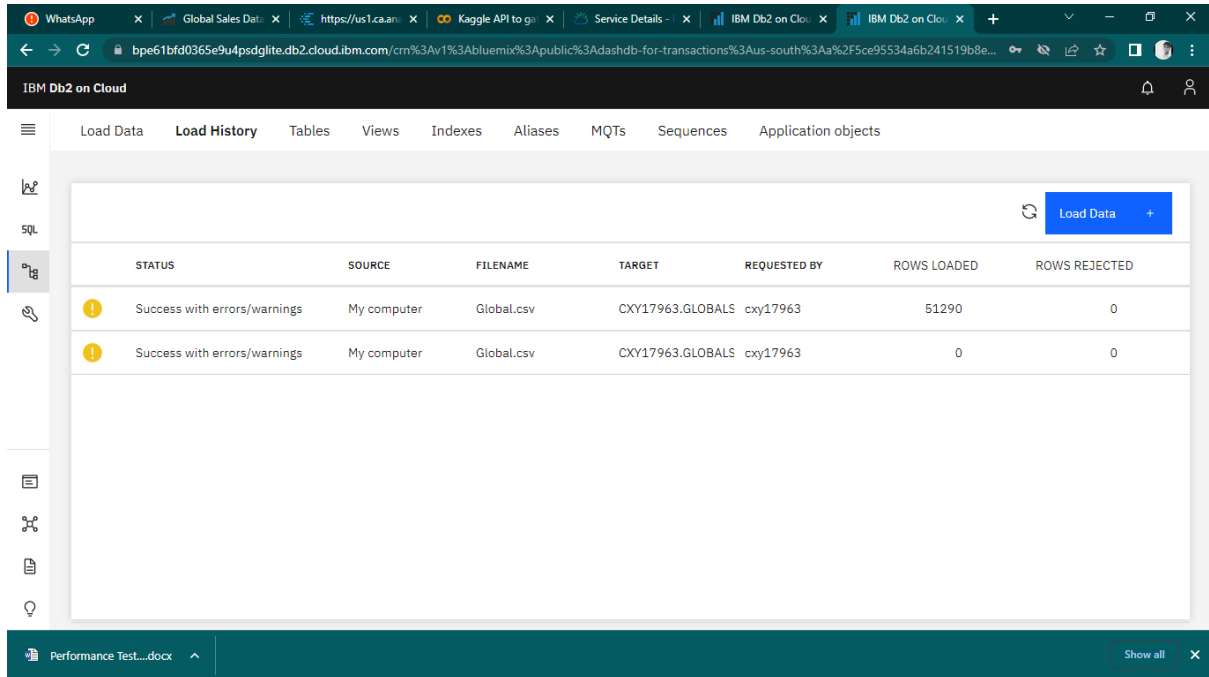
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Product ID	Category	Sub-Category
--------	----------	------------	-----------	-----------	-------------	---------------	---------	------	-------	------------	----------	--------------

Jupyter Server: Local Cell 3 of 31 Go Live

## Uploading dataset to IBM Cloud:

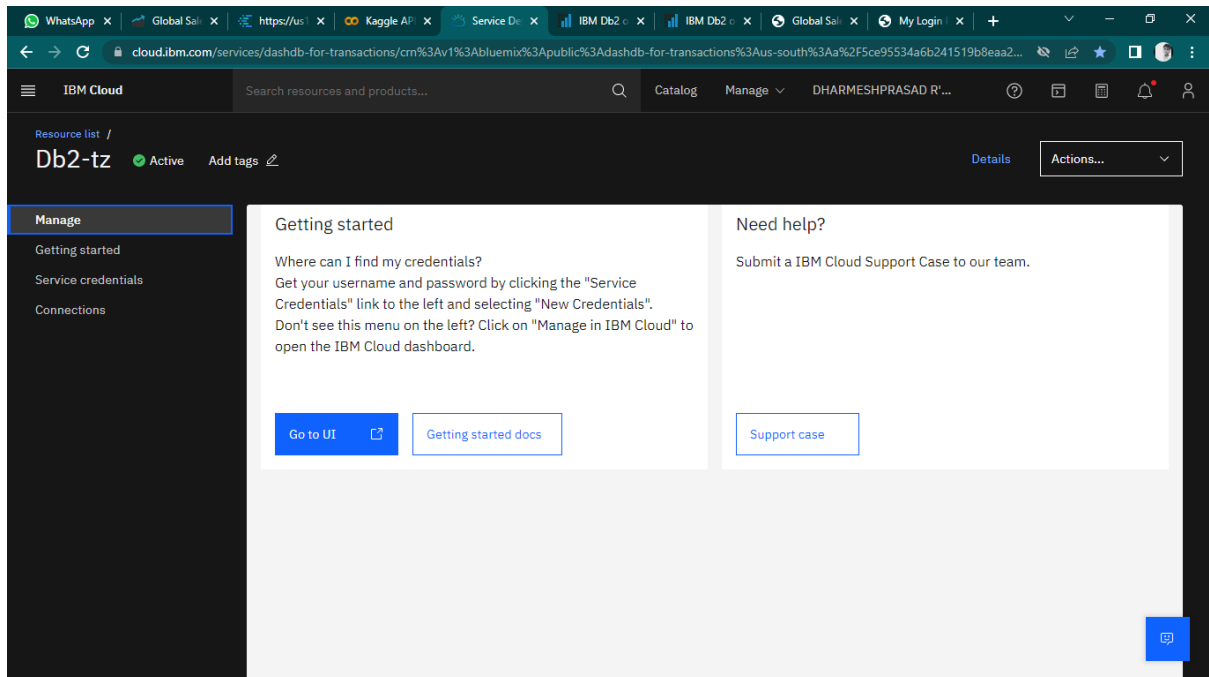
Uploading dataset into IBM cloud.

Using IBM Db2.



The screenshot shows the IBM Db2 on Cloud console interface. The top navigation bar includes tabs for Load Data, Load History, Tables, Views, Indexes, Aliases, MQTs, Sequences, and Application objects. The Load History table is displayed with the following columns: STATUS, SOURCE, FILENAME, TARGET, REQUESTED BY, ROWS LOADED, and ROWS REJECTED. The table contains two rows of data, both indicating successful uploads with errors/warnings.

STATUS	SOURCE	FILENAME	TARGET	REQUESTED BY	ROWS LOADED	ROWS REJECTED
Success with errors/warnings	My computer	Global.csv	CXY17963.GLOBALS	cxy17963	51290	0
Success with errors/warnings	My computer	Global.csv	CXY17963.GLOBALS	cxy17963	0	0



The screenshot shows the IBM Cloud console interface for the Db2-tz resource. The left sidebar contains a navigation menu with options: Manage, Getting started, Service credentials, and Connections. The main content area is divided into two sections: Getting started and Need help. The Getting started section provides instructions on how to find credentials and includes a button to go to the UI. The Need help section provides a link to submit a support case.

**Getting started**

Where can I find my credentials?  
Get your username and password by clicking the "Service Credentials" link to the left and selecting "New Credentials".  
Don't see this menu on the left? Click on "Manage in IBM Cloud" to open the IBM Cloud dashboard.

[Go to UI](#) [Getting started docs](#)

**Need help?**

Submit a IBM Cloud Support Case to our team.

[Support case](#)

WhatsApp Global Sales Data https://us1.ca... Kaggle API to go Service Details IBM Db2 on Cloud IBM Db2 on Cloud

bpe61bfd0365e9u4psdglite.db2.cloud.ibm.com/crm%3Av1%3Abluemix%3Apublic%3Adashdb-for-transactions%3Aus-south%3Aa%2F5ce95534a6b241519b8e...

IBM Db2 on Cloud

Load Data Load History **Tables** Views Indexes Aliases MQTs Sequences Application objects

CXY17963.GLOBALSALES

Back

Export to CSV

ROW_ID	ORDER_ID	ORDER_DATE	SHIP_DATE	SHIP_MODE	CUSTOMER_ID	CUSTOMER_NAME	SEGMENT	CITY	STATE
1	MX-2014-143658	02-10-2014	06-10-2014	Standard Class	SC-20575	Sonia Cooley	Consumer	Mexico City	Distrito Federal
2	MX-2012-155047	15-10-2012	20-10-2012	Standard Class	KW-16570	Kelly Williams	Consumer	Dos Quebra das	Risaralda
3	MX-2012-155047	15-10-2012	20-10-2012	Standard Class	KW-16570	Kelly Williams	Consumer	Dos Quebra das	Risaralda
4	MX-2012-155047	15-10-2012	20-10-2012	Standard Class	KW-16570	Kelly Williams	Consumer	Dos Quebra das	Risaralda
5	MX-2012-	15-10-2012	20-10-2012	Standard Class	KW-16570	Kelly Williams	Consumer	Dos Quebra das	Risaralda


Items per page: 50 1-50 items 1 page 1 Show all

Performance Test....docx

## Login page creation:

WhatsApp Global Sales Data https://us1.ca... Kaggle API to go Service Details IBM Db2 on Cloud IBM Db2 on Cloud Global Sales Data Global Sales Data

127.0.0.1:5500/index.html



**Login**

E-Mail Address

Password [Forgot Password?](#)

☐ Remember Me

Login

Don't have an account? [Create One](#)

Copyright © 2022 — PNT2022TMID41225

WhatsApp

Global Sal

https://us

Kaggle AI

Service De


IBM Db2 c

IBM Db2 c

Global Sal

My Login

127.0.0.1:5500/register.html



## Register

Name

E-Mail Address

Password

☐ I agree to the [Terms and Conditions](#)

Register

Already have an account? [Login](#)