

Project Development Phase

Sprint 1

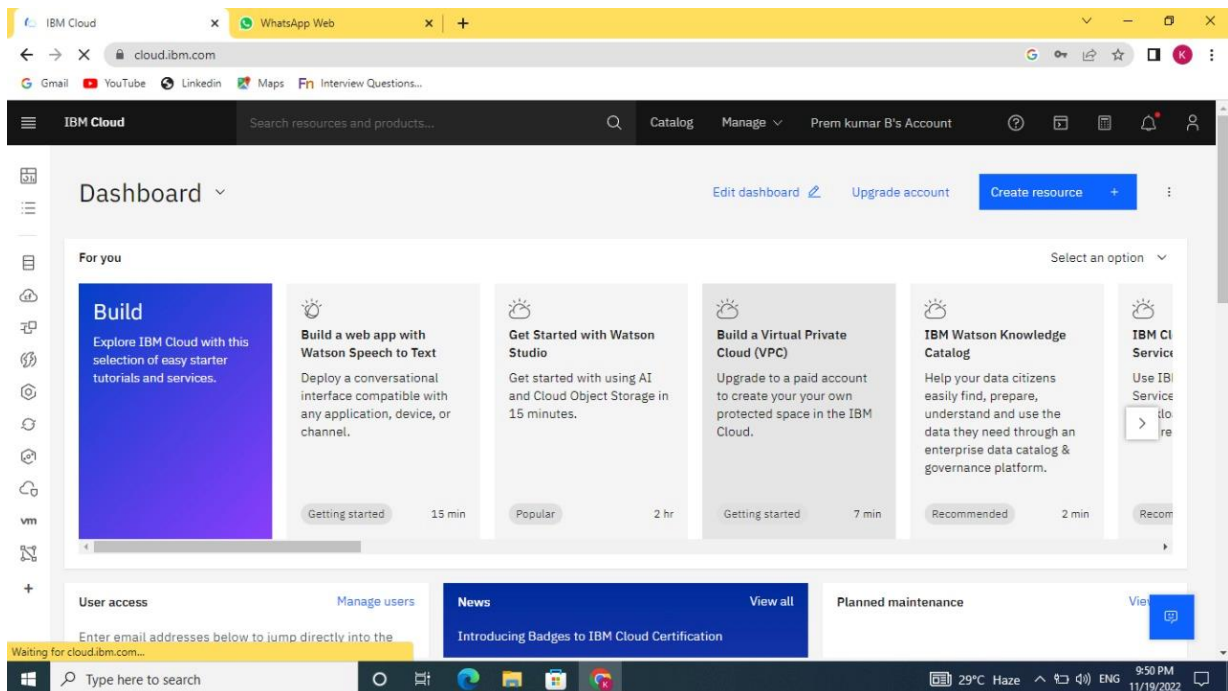
Date	25 October 2022
Team ID	PNT2022TMID24517
Project Name	Global Sales Data Analytics

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

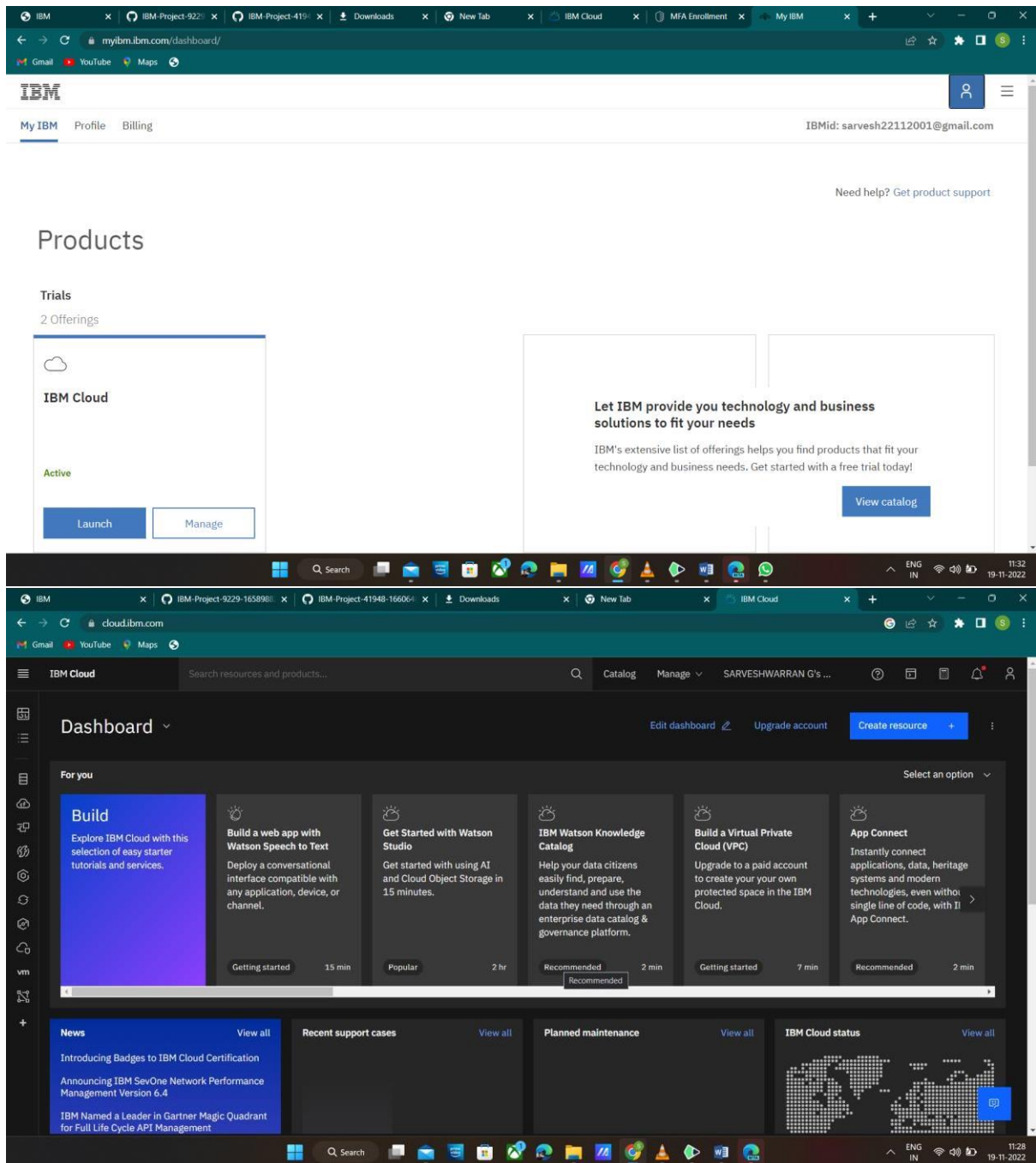
IBM Cloud Account:

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

```
Kaggle API to gather the dataset.ipynb
File Edit View Insert Runtime Tools Help Last edited on November 14

+ Code + Text
Connect Editing

!pip install -q kaggle

[ ] !mkdir ~/.kaggle

[ ] !cp kaggle.json ~/.kaggle/

[ ] !kaggle datasets download -d mrhatrider/global-sales

Warning: Your Kaggle API key is readable by other users on this system! To fix this, you can run 'chmod 600 /root/.kaggle/kaggle.json'
Downloading global-sales.zip to /content
0% 0.00/3.48M [00:00<, ?B/s]
100% 3.48M/3.48M [00:00<00, 168MB/s]

[ ] !unzip /content/global-sales.zip

Archive: /content/global-sales.zip
  inflating: global_superstore2.csv
```

Downloading Dataset:

Downloading Global super store dataset from kaggle.

The screenshot shows the Kaggle website interface. The top navigation bar includes the Kaggle logo, a search bar, and links for 'Sign In' and 'Register'. The left sidebar contains a 'Create' button and a list of navigation options: Home, Competitions, Datasets, Code, Discussions, Learn, and More. The main content area displays the 'Global Super Store Dataset' page, which was updated 2 years ago by Apoorva Mahalingappa. The page features a 'Download (12 MB)' button and a 'New Notebook' button. The dataset is described as 'Consumer and product Analytics on Global Super Store Data.' Below this, there are tabs for 'Data', 'Code (9)', and 'Discussion (2)'. The 'About Dataset' section provides context, stating that the dataset is used for understanding customer analysis. The 'Usability' section shows a rating of 8.24. The 'License' section indicates that the database is open and the content is available under a specific license. The 'Expected update frequency' is listed as 'Monthly'. A footer notice states that cookies are used to deliver services and improve the user experience.

Data collection:

Understanding the dataset and analysing the data requirements.

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
32298	CA-2012-1	31-07-2011	31-07-2011	Same Day	RH-19495	Rick Hansi Consumer	New York	New York	United States	10024	US	East	TEC-AC-1C	Technology	Accessories	Plantronic	2309.65	7	
26341	IN-2013-7	5/2/2013	7/2/2013	Second Class	JR-16210	Justin Ritt Corporate	Wollongong	New South Wales	Australia		APAC	Oceania	FUR-CH-1C	Furniture	Chairs	Novimex	3709.395	9	
25330	IN-2013-1	17-10-2011	18-10-2011	First Class	CR-12730	Craig Reitt Consumer	Brisbane	Queensland	Australia		APAC	Oceania	TEC-PH-1C	Technology	Phones	Nokia Smartphone	5175.171	9	
13524	ES-2013-1	28-01-2011	30-01-2011	First Class	KM-16375	Katherine Home	Off Berlin	Berlin	Germany		EU	Central	TEC-PH-1C	Technology	Phones	Motorola	2892.51	5	
47221	SG-2013-4	#####	#####	Same Day	RH-9495	Rick Hansi Consumer	Dakar	Dakar	Senegal		Africa	Africa	TEC-SHA-1C	Technology	Copiers	Sharp Wireless	2832.96	8	
22732	IN-2013-4	28-06-2011	1/7/2013	Second Class	JM-15655	Jim Mitchell Corporate	Sydney	New South Wales	Australia		APAC	Oceania	TEC-PH-1C	Technology	Phones	Samsung	2862.675	5	
30570	IN-2011-8	#####	#####	First Class	TS-21340	Toby Swirl Consumer	Porirua	Wellington	New Zealand		APAC	Oceania	FUR-CH-1C	Furniture	Chairs	Novimex	1822.08	4	
31192	IN-2012-8	14-04-2011	18-04-2011	Standard	MB-18085	Mick Brown Consumer	Hamilton	Waikato	New Zealand		APAC	Oceania	FUR-TA-1C	Furniture	Tables	Chromcraft	5244.84	6	
40155	CA-2014-1	14-10-2011	21-10-2011	Standard	JW-15220	Jane Wack Corporate	Sacramento	California	United States	95823	US	West	OFF-BI-10	Office Supplies	Binders	Fellowes	5083.96	5	
40936	CA-2012-1	28-01-2011	31-01-2011	Second Class	JH-15985	Joseph Hc Consumer	Concord	North Carolina	United States	28027	US	South	FUR-TA-1C	Furniture	Tables	Chromcraft	4297.644	13	
34577	CA-2011-1	5/4/2011	9/4/2011	Second Class	GM-14695	Greg Maxi Corporate	Alexandria	Virginia	United States	22304	US	South	OFF-SU-1C	Office Supplies	Supplies	Martin Yale	4164.05	5	
28879	ID-2012-2	19-04-2011	22-04-2011	First Class	AJ-10780	Anthony J Corporate	Kabul	Kabul	Afghanistan		APAC	Central Asia	FUR-TA-1C	Furniture	Tables	Bevis Com	4626.15	5	
45794	SA-2011-1	27-12-2011	29-12-2011	Second Class	MM-7260	Magdalen Consumer	Jizan	Jizan	Saudi Arabia		EMEA	EMEA	TEC-CIS-1C	Technology	Phones	Cisco Smartphone	2616.96	4	
4132	MX-2012-1	13-11-2011	13-11-2011	Same Day	VF-21715	Vicky Frey Home	Off Toledo	Parana	Brazil		LATAM	South	FUR-CH-1C	Furniture	Chairs	Harbour C	2221.8	7	
27704	IN-2013-7	6/6/2013	8/6/2013	Second Class	PF-19120	Peter Full Consumer	Mudanjiang	Heilongjiang	China		APAC	North Asia	OFF-AP-1C	Office Supplies	Appliance	KitchenAid	3701.52	12	
13779	ES-2014-5	31-07-2011	3/8/2014	Second Class	BP-11185	Ben Peter Corporate	Paris	Ile-de-France	France		EU	Central	OFF-AP-1C	Office Supplies	Appliance	Breville	1869.588	4	
36178	CA-2014-1	#####	#####	Second Class	TB-21175	Thomas B Corporate	Henderso	Kentucky	United States	42420	US	South	TEC-AC-1C	Technology	Accessories	Logitech	2249.91	9	
12069	ES-2014-1	8/9/2014	14-09-2011	Standard	PJ-18835	Patrick Jor Corporate	Prato	Tuscany	Italy		EU	South	OFF-AP-1C	Office Supplies	Appliance	Hoover	7958.58	14	
22096	IN-2014-1	31-01-2011	1/2/2014	First Class	JS-15685	Jim Sink Corporate	Townsville	Queensland	Australia		APAC	Oceania	TEC-CO-1C	Technology	Copiers	Brother	2565.594	9	
49463	TZ-2014-8	#####	#####	Second Class	RH-9555	Ritsa High Consumer	Uvinza	Kigoma	Tanzania		Africa	Africa	OFF-KIT-1C	Office Supplies	Appliance	KitchenAid	3409.74	6	
46630	PL-2012-7	8/8/2012	#####	First Class	AB-600	Ann Blum Corporate	Bytom	Silesia	Poland		EMEA	EMEA	FUR-HON-1C	Furniture	Tables	Hon Comp	1977.72	4	
31784	CA-2011-1	29-10-2011	31-10-2011	First Class	SA-20830	Sue Ann R Consumer	Chicago	Illinois	United States	60610	US	Central	TEC-PH-1C	Technology	Phones	Apple iPhone	2735.952	6	
21586	IN-2011-4	2/5/2011	3/5/2011	First Class	JK-15325	Jason Klar Corporate	Suzhou	Anhui	China		APAC	North Asia	FUR-CH-1C	Furniture	Chairs	SAFCO	2754	6	
13528	ES-2013-2	27-02-2011	1/3/2013	Second Class	LB-16795	Laurel Bel Home	Off Edinburgh	Scotland	United Kingdom		EU	North	OFF-AP-1C	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.

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:

File Edit Selection View Go Run Terminal Help data_process.ipynb - Global Sales Data Analytics - Visual Studio Code

EXPLORER GLOBAL SALES DATA ANALYTICS... archive (5).zip data_process.ipynb eda.ipynb Global_Superstore2.csv Global_Superstore2.xlsx Global.csv

data_process.ipynb > M>Loading Dataset

Importing libraries

```
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.

IBM Db2 on Cloud

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

Load Data

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

Performance Test....docx

IBM Db2 on Cloud

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

CXY17963.GLOBALSALES

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

Items per page: 50 1-50 items

1 page 1

Performance Test....docx

Login page creation:

WhatsApp

Global Sal

https://us1

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](#)