

PROJECT DEVELOPMENT PHASE

SPRINT-2 : DATA PRE PROCESSING AND EXPLORATORY ANALYSIS

DATE	03/11/2022
TEAM ID	PNT2022TMID34610
PROJECT NAME	Retail Store Stock Inventory Analysis.

Team Leader	BAGIEO A
Team member	AROCKIYA LIBIN V
Team member	BENILO CHRISTIN ROY A
Team member	ASWIN SUDEER S

- Reading the CSV file :

The screenshot shows a Jupyter Notebook interface. The code cell contains the following Python code:

```
import pandas as pd
import numpy as np
store = pd.read_csv('prob1-US-Superstore-data.csv')
store.head()
```

The output of the code is a preview of the first three rows of the CSV file:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	
0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	42420	South	FUR-BO-10001798	Furniture	Bookcases	Somerset Bookcase
1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	42420	South	FUR-CH-10000454	Furniture	Chairs	Honoring Upholstered Chair
2	3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	90036	West	OFF-LA-10000240	Office Supplies	Labels	Adhesive Labels Typewriter

- Splitting the date to year month and day wise for better readability:

```
[8] store["Order year"]=store["Order Date"].dt.year
store["Order month"]=store["Order Date"].dt.month
store["Order day"]=store["Order Date"].dt.day
store.head()
```

Order ID	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	Sales	Quantity	Discount	Profit	year	month	day	Order year	Order month	Order day
6-08	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	261.9600	2	0.00	41.9136	2016	11	8	2016	11	8
6-08	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	731.9400	3	0.00	219.5820	2016	11	8	2016	11	8
6-12	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	14.6200	2	0.00	6.8714	2016	6	12	2016	6	12
5-11	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	957.5775	5	0.45	-383.0310	2015	10	11	2015	10	11
5-11	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	22.3680	2	0.20	2.5164	2015	10	11	2015	10	11

Split_date.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- US.csv
- prob1-US-Superstore-da...

```
[8] store["Order year"]=store["Order Date"].dt.year
store["Order month"]=store["Order Date"].dt.month
store["Order day"]=store["Order Date"].dt.day
store.head()
```

er	te	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Sales	Quantity	Discount	Profit	year	month	day	Order year	Order month	Order day
6-	08	11/11/2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	...	261.9600	2	0.00	41.9136	2016	11	8	2016	11	8
6-	08	11/11/2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	...	731.9400	3	0.00	219.5820	2016	11	8	2016	11	8
6-	12	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	14.6200	2	0.00	6.8714	2016	6	12	2016	6	12
5-	11	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	957.5775	5	0.45	-383.0310	2015	10	11	2015	10	11
5-	11	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	22.3680	2	0.20	2.5164	2015	10	11	2015	10	11

0s completed at 1:10 AM

- Reindexing and removing the unwanted columns in the data frame:

Split_date.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- US.csv
- prob1-US-Superstore-da...

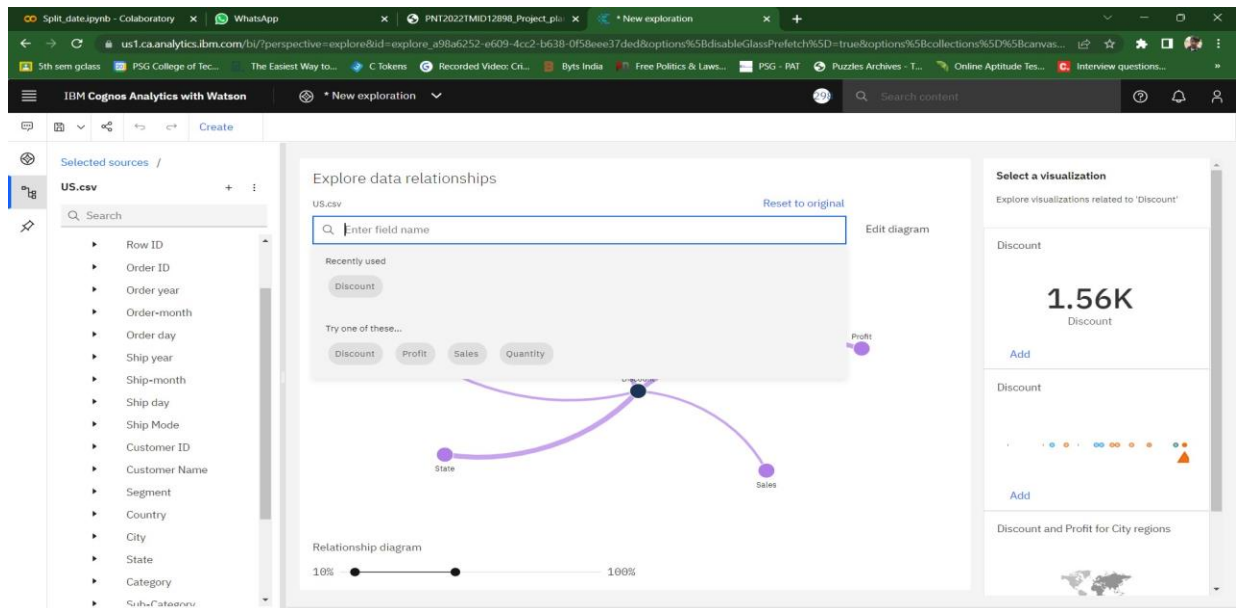
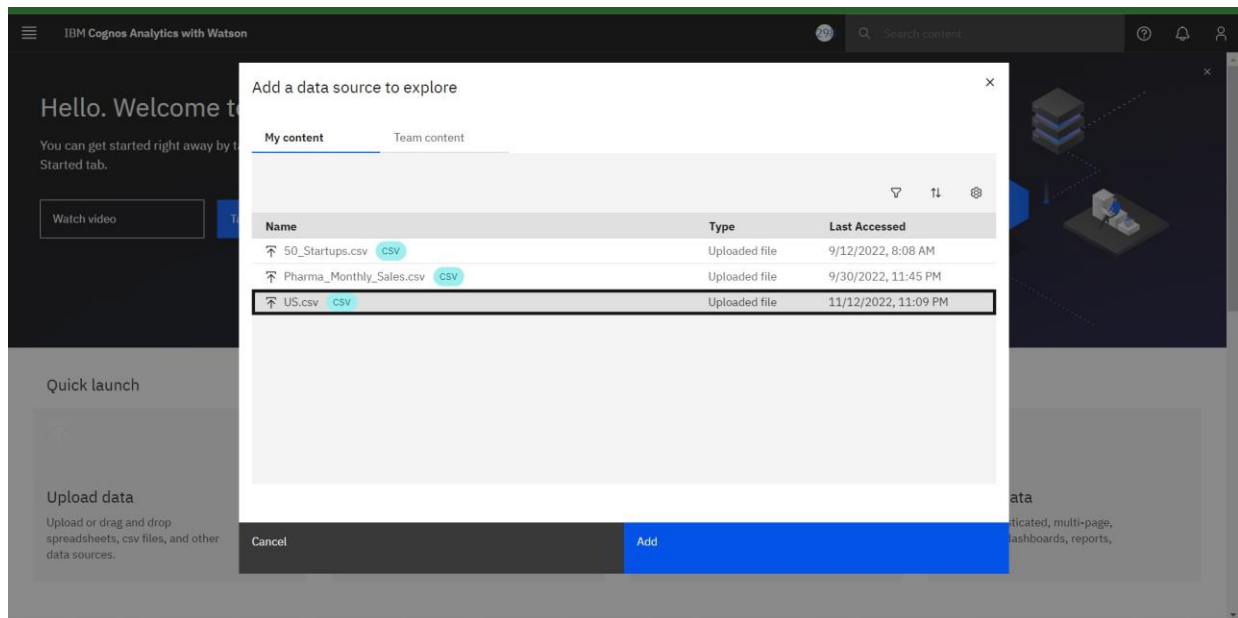
```
[36] new_cols = ["Row ID", "Order ID", "Order year", "Order month", "Order day", "Ship year", "Ship month", "Ship day", "Ship Mode", "Customer ID", "Customer Name", "Segment"]
store1=store1.reindex(columns=new_cols)
store1.head()
```

Row ID	Order ID	Order year	Order month	Order day	Ship year	Ship month	Ship day	Ship Mode	Customer ID	...	Country	City	state	Category	Sub-Category	Product Name	Quantity	
0	1	CA-2016-152156	2016	Nov	8	2016	Nov	11	Second Class	CG-12520	...	United States	Henderson	Kentucky	Furniture	Bookcases	Bush Somerset Collection Bookcase	2
1	2	CA-2016-152156	2016	Nov	8	2016	Nov	11	Second Class	CG-12520	...	United States	Henderson	Kentucky	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...	3
2	3	CA-2016-139688	2016	Jun	12	2016	Jun	16	Second Class	DV-13045	...	United States	Los Angeles	California	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	2
3	4	US-2015-108966	2015	Oct	11	2015	Oct	18	Standard Class	SO-20335	...	United States	Fort Lauderdale	Florida	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	5 95
4	5	US-2015-108966	2015	Oct	11	2015	Oct	18	Standard Class	SO-20335	...	United States	Fort Lauderdale	Florida	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	2

5 rows x 22 columns

0s completed at 1:10 AM

- **Uploading the dataset into cognos:**



Collab file link:

<https://colab.research.google.com/drive/1DlxXJarog5u84rdq0B2oUyPi2Cw9BRu3#scrollTo=CdoHwYOA93B>