

Delivery of Sprint 2

Date	05th November
Team ID	PNT2022TMID43650
Project Name	Project - Global Sales Data Analytics

Cleaning Data in Python

```
Untitled0.ipynb
File Edit View Insert Runtime Tools Help

+ Code + Test

[1] import numpy as np
import sklearn as sk
import seaborn as sns
import pandas as pd

[2] data = pd.read_csv('Global_Superstore2.csv', encoding = 'unicode_escape', engine = 'python')
data

data.info()

Removing The Rows that contains NULL VALUES

[5] data.dropna()

Row ID      Order ID      Order Date      Ship Date      Ship Mode      Customer ID      Customer Name      Segment      City      State      ...      Product ID      Category      Sub-Category      Product Name      Sales      Quantity      Discount      Profit      Shipping Cost      Order Priority
0      32296      CA-2012-124891      31-07-2012      31-07-2012      Same Day      RH-19495      Rick Hansen      Consumer      New York City      New York      ...      TEC-AC-10003033      Technology      Accessories      Plantronics CS510 - Over-the-Head monaural W...      2309.6500      7.0      0.00      762.1845      933.57      Critical
8      49155      CA-2014-125909      14-10-2014      21-10-2014      Standard Class      JW-15220      Jane Waco      Corporate      Sacramento      California      ...      OFF-BI-10003527      Office Supplies      Binders      Followes PB500 Electric Punch Plastic Comb Bin...      5083.9600      5.0      0.20      1906.4850      867.69      Low
9      40936      CA-2012-119638      28-01-2012      31-01-2012      Second Class      JH-15985      Joseph Holt      Consumer      Concord      North Carolina      ...      FUR-TA-10001198      Furniture      Tables      Chromcraft Bull-Nose Wood Oval Conference Tabl...      4297.6440      13.0      0.40      -1862.3124      865.74      Critical
10     34577      CA-2011-102988      05-04-2011      09-04-2011      Second Class      GM-14695      Greg Maxwell      Corporate      Alexandria      Virginia      ...      OFF-SU-10002881      Office Supplies      Supplies      Martin Yale Chadless Opener Electric Letter Op...      4164.0500      5.0      0.00      83.2810      846.54      High
16     36178      CA-2014-143567      03-11-2014      06-11-2014      Second Class      TB-21175      Thomas Boland      Corporate      Henderson      Kentucky      ...      TEC-AC-10004145      Technology      Accessories      Logitech dNovo Edge Keyboard      2249.9100      9.0      0.00      517.4793      789.70      Critical
...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
12684  38622      US-2014-128447      11-11-2014      18-11-2014      Standard Class      MC-17845      Michael Chen      Consumer      Pasco      Washington      ...      OFF-AP-10004540      Office Supplies      Appliances      Eureka The Boss Lite 10-Amp Upright Vacuum, Blue      400.8000      5.0      0.00      112.2240      23.36      Medium
12687  33700      US-2013-110170      28-09-2013      04-10-2013      Standard Class      HM-14860      Harry Maile      Corporate      Huntsville      Texas      ...      FUR-BO-10000780      Furniture      Bookcases      O'Sullivan Plantations 2-Door Library in Lambr...      956.6648      7.0      0.32      -225.0976      23.35      Medium
13275  34788      CA-2012-157322      02-07-2012      06-07-2012      Standard Class      RH-19600      Rob Haberlin      Consumer      Carol Stream      Illinois      ...      OFF-ST-10003208      Office Supplies      Storage      Adjustable Depth LetterLegal Cart      435.5040      3.0      0.20      48.9942      23.32      Medium
13284  37178      CA-2013-133935      19-09-2013      23-09-2013      Standard Class      JW-15220      Jane Waco      Corporate      San Diego      California      ...      OFF-PA-10000100      Office Supplies      Paper      Xerox 1945      368.9100      9.0      0.00      180.7659      23.30      High
13292  33830      CA-2011-111500      17-08-2011      21-08-2011      Standard Class      DJ-13510      Don Jones      Corporate      Little Rock      Arkansas      ...      TEC-AC-10008444      Technology      Accessories      Logitech Gaming G510s - Keyboard      339.9600      4.0      0.00      122.3856      23.28      Medium
2248 rows x 24 columns
```

```
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Test

Finding AND REMOVING DUPLICATES

print(data.duplicated())

0      False
1      False
2      False
3      False
4      False
...
13292  False
13293  False
13294  False
13295  False
13296  False
length: 13297, dtype: bool

The Rows with empty cells were removed

[10] data.drop_duplicates(inplace=True)
print(data.to_string())

bound method DataFrame.to_string of
0      32296      CA-2012-124891      31-07-2012      31-07-2012      Same Day      RH-19495      Rick Hansen      Consumer      New York City      New York      ...      TEC-AC-10003033      Technology      Accessories      Plantronics CS510 - Over-the-Head monaural W...      2309.6500      7.0      0.00      762.1845      933.57      Critical
1      26341      IN-2013-77878      05-02-2013      07-02-2013      Second Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
2      25394      IN-2013-72949      17-10-2013      18-10-2013      First Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
3      13528      ES-2013-1209342      28-04-2013      30-04-2013      First Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
4      47221      SE-2013-4320      05-11-2013      06-11-2013      Same Day      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13292  13630      CA-2011-131580      17-08-2011      21-08-2011      Standard Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13293  44020      IN-2014-6460      03-08-2014      07-08-2014      Standard Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13294  7493      IN-2014-142786      27-09-2014      02-10-2014      Standard Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13295  11746      ES-2013-295595      08-06-2013      29-06-2013      Second Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13296  11899      IT-2011-2396333      24-08-2011      28-08-2011      Standard Class      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13292  01-13538      Don Jones      Corporate      Little Rock      Arkansas      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13293  58-18179      Sarah Bern      Consumer      Little Rock      Arkansas      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13294  06-28958      Suzanne McElair      Corporate      Little Rock      Arkansas      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13295  06-17778      Maxwell Schwartz      Consumer      Jerez de la Frontera      Spain      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13296  09-23345      Dennis Newton      Corporate      Little Rock      Arkansas      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
2248 rows x 24 columns
```

```
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Test

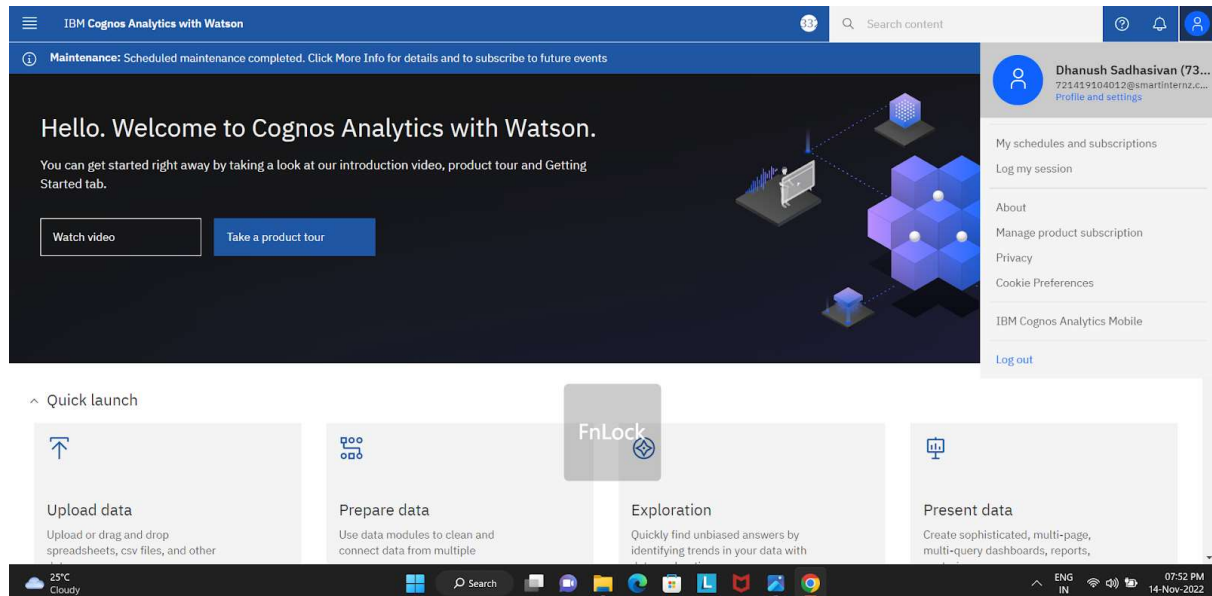
Removing The Rows that contains NULL VALUES

[3] data.dropna()

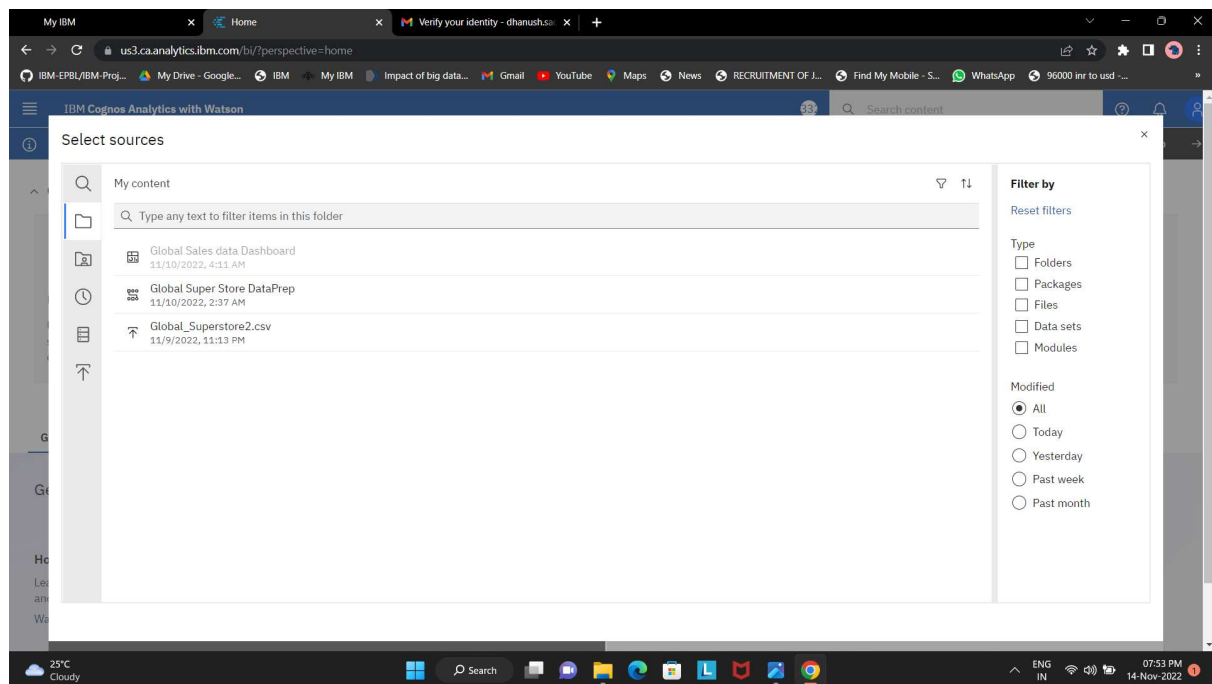
Row ID      Order ID      Order Date      Ship Date      Ship Mode      Customer ID      Customer Name      Segment      City      State      ...      Product ID      Category      Sub-Category      Product Name      Sales      Quantity      Discount      Profit      Shipping Cost      Order Priority
0      32296      CA-2012-124891      31-07-2012      31-07-2012      Same Day      RH-19495      Rick Hansen      Consumer      New York City      New York      ...      TEC-AC-10003033      Technology      Accessories      Plantronics CS510 - Over-the-Head monaural W...      2309.6500      7.0      0.00      762.1845      933.57      Critical
8      49155      CA-2014-125909      14-10-2014      21-10-2014      Standard Class      JW-15220      Jane Waco      Corporate      Sacramento      California      ...      OFF-BI-10003527      Office Supplies      Binders      Followes PB500 Electric Punch Plastic Comb Bin...      5083.9600      5.0      0.20      1906.4850      867.69      Low
9      40936      CA-2012-119638      28-01-2012      31-01-2012      Second Class      JH-15985      Joseph Holt      Consumer      Concord      North Carolina      ...      FUR-TA-10001198      Furniture      Tables      Chromcraft Bull-Nose Wood Oval Conference Tabl...      4297.6440      13.0      0.40      -1862.3124      865.74      Critical
10     34577      CA-2011-102988      05-04-2011      09-04-2011      Second Class      GM-14695      Greg Maxwell      Corporate      Alexandria      Virginia      ...      OFF-SU-10002881      Office Supplies      Supplies      Martin Yale Chadless Opener Electric Letter Op...      4164.0500      5.0      0.00      83.2810      846.54      High
16     36178      CA-2014-143567      03-11-2014      06-11-2014      Second Class      TB-21175      Thomas Boland      Corporate      Henderson      Kentucky      ...      TEC-AC-10004145      Technology      Accessories      Logitech dNovo Edge Keyboard      2249.9100      9.0      0.00      517.4793      789.70      Critical
...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...      ...
13264  38622      US-2014-128447      11-11-2014      18-11-2014      Standard Class      MC-17845      Michael Chen      Consumer      Pasco      Washington      ...      OFF-AP-10004540      Office Supplies      Appliances      Eureka The Boss Lite 10-Amp Upright Vacuum, Blue      400.8000      5.0      0.00      112.2240      23.36      Medium
12687  33700      US-2013-110170      28-09-2013      04-10-2013      Standard Class      HM-14860      Harry Maile      Corporate      Huntsville      Texas      ...      FUR-BO-10000780      Furniture      Bookcases      O'Sullivan Plantations 2-Door Library in Lambr...      956.6648      7.0      0.32      -225.0976      23.35      Medium
13275  34788      CA-2012-157322      02-07-2012      06-07-2012      Standard Class      RH-19600      Rob Haberlin      Consumer      Carol Stream      Illinois      ...      OFF-ST-10003208      Office Supplies      Storage      Adjustable Depth LetterLegal Cart      435.5040      3.0      0.20      48.9942      23.32      Medium
13284  37178      CA-2013-133935      19-09-2013      23-09-2013      Standard Class      JW-15220      Jane Waco      Corporate      San Diego      California      ...      OFF-PA-10000100      Office Supplies      Paper      Xerox 1945      368.9100      9.0      0.00      180.7659      23.30      High
13292  33830      CA-2011-111500      17-08-2011      21-08-2011      Standard Class      DJ-13510      Don Jones      Corporate      Little Rock      Arkansas      ...      TEC-AC-10008444      Technology      Accessories      Logitech Gaming G510s - Keyboard      339.9600      4.0      0.00      122.3856      23.28      Medium
2248 rows x 24 columns
```

Preparation Of Data In IBM COGNOS ANALYTICS

LOGIN to IBM COGNOS ANALYTICS



Select The Source As the dataset (Global_Superstore2.csv)



Analyze the Data in the Grid

IBM Cognos Analytics with Watson | New data module

Search content

Properties

Data module

Grid Relationships Custom tables

T1	Row Id	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
	1	32298	CA-2012-124891	2012-07-31	2012-07-31	Same Day	RH-19495	Rick Hansen
	2	26341	IN-2013-77878	2013-02-05	2013-02-07	Second Class	JR-16210	Justin Ritter
	3	25330	IN-2013-71249	2013-10-17	2013-10-18	First Class	CR-12730	Craig Reiter
	4	13524	ES-2013-1579342	2013-01-28	2013-01-30	First Class	KM-16375	Katherine Murra
	5	47221	SG-2013-4320	2013-11-05	2013-11-06	Same Day	RH-9495	Rick Hansen
	6	22732	IN-2013-42360	2013-06-28	2013-07-01	Second Class	JM-15655	Jim Mitchum
	7	30570	IN-2011-81826	2011-11-07	2011-11-09	First Class	TS-21340	Toby Swindell
	8	31192	IN-2012-86369	2012-04-14	2012-04-18	Standard Class	MB-18085	Mick Brown
	9	40155	CA-2014-135909	2014-10-14	2014-10-21	Standard Class	JW-15220	Jane Waco
	10	40936	CA-2012-116638	2012-01-28	2012-01-31	Second Class	JH-15985	Joseph Holt
	11	34577	CA-2011-102988	2011-04-05	2011-04-09	Second Class	GM-14695	Greg Maxwell

CREATE Calculations

1. ORDER_DAY
2. ORDER_MONTH
3. ORDER_YEAR
4. ALL
5. MIN_SALES
6. TARGET_SALES
7. MAX_SALES
8. MID SALES

IBM Cognos Analytics with Watson | Global Super Store DataPrep

Search content

Maintenance: Scheduled maintenance completed. Click More Info for details and to subscribe to future events

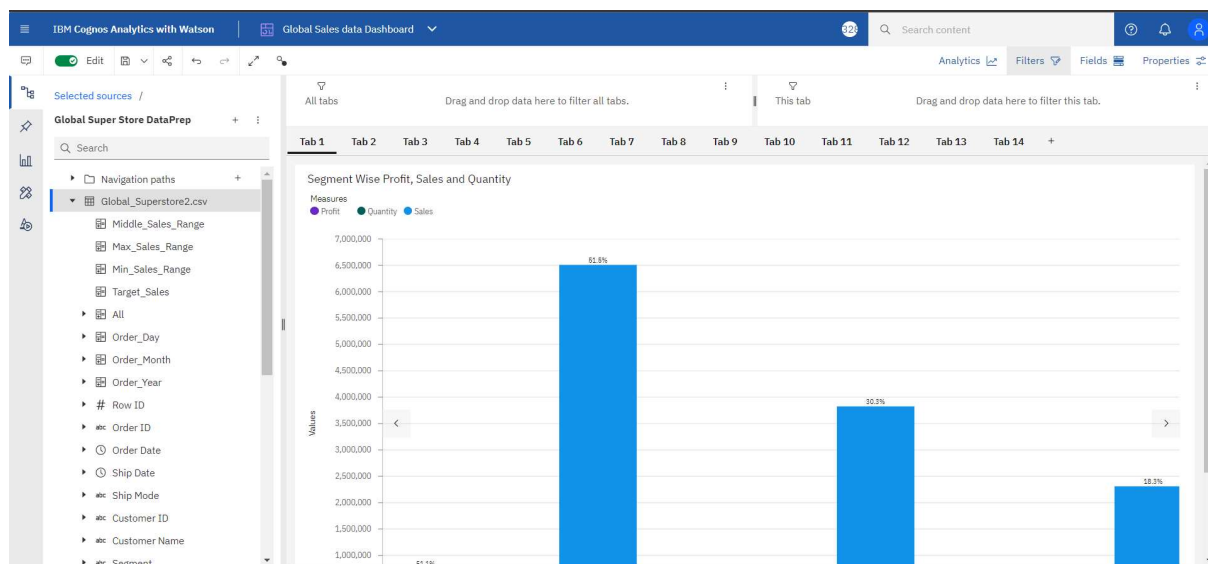
Dismiss More info

Properties

Data module

Grid Relationships Custom tables

T1	Row Id	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
	1	32298	CA-2012-124891	2012-07-31	2012-07-31	Same Day	RH-19495	Rick Hansen
	2	26341	IN-2013-77878	2013-02-05	2013-02-07	Second Class	JR-16210	Justin Ritter
	3	25330	IN-2013-71249	2013-10-17	2013-10-18	First Class	CR-12730	Craig Reiter
	4	13524	ES-2013-1579342	2013-01-28	2013-01-30	First Class	KM-16375	Katherine Murra
	5	47221	SG-2013-4320	2013-11-05	2013-11-06	Same Day	RH-9495	Rick Hansen
	6	22732	IN-2013-42360	2013-06-28	2013-07-01	Second Class	JM-15655	Jim Mitchum
	7	30570	IN-2011-81826	2011-11-07	2011-11-09	First Class	TS-21340	Toby Swindell
	8	31192	IN-2012-86369	2012-04-14	2012-04-18	Standard Class	MB-18085	Mick Brown
	9	40155	CA-2014-135909	2014-10-14	2014-10-21	Standard Class	JW-15220	Jane Waco
	10	40936	CA-2012-116638	2012-01-28	2012-01-31	Second Class	JH-15985	Joseph Holt



SAVED The Prepared DATASET AS Global Super Store DataPrep

