

Project Development Sprint 2

Team ID : PNT2022TMID35387

- Prepare data in **IBM COGNOS ANALYTICS**

IBM Cognos Analytics with Watson

Maintenance: Scheduled maintenance completed. Click [More Info](#) for details and to subscribe to future events.

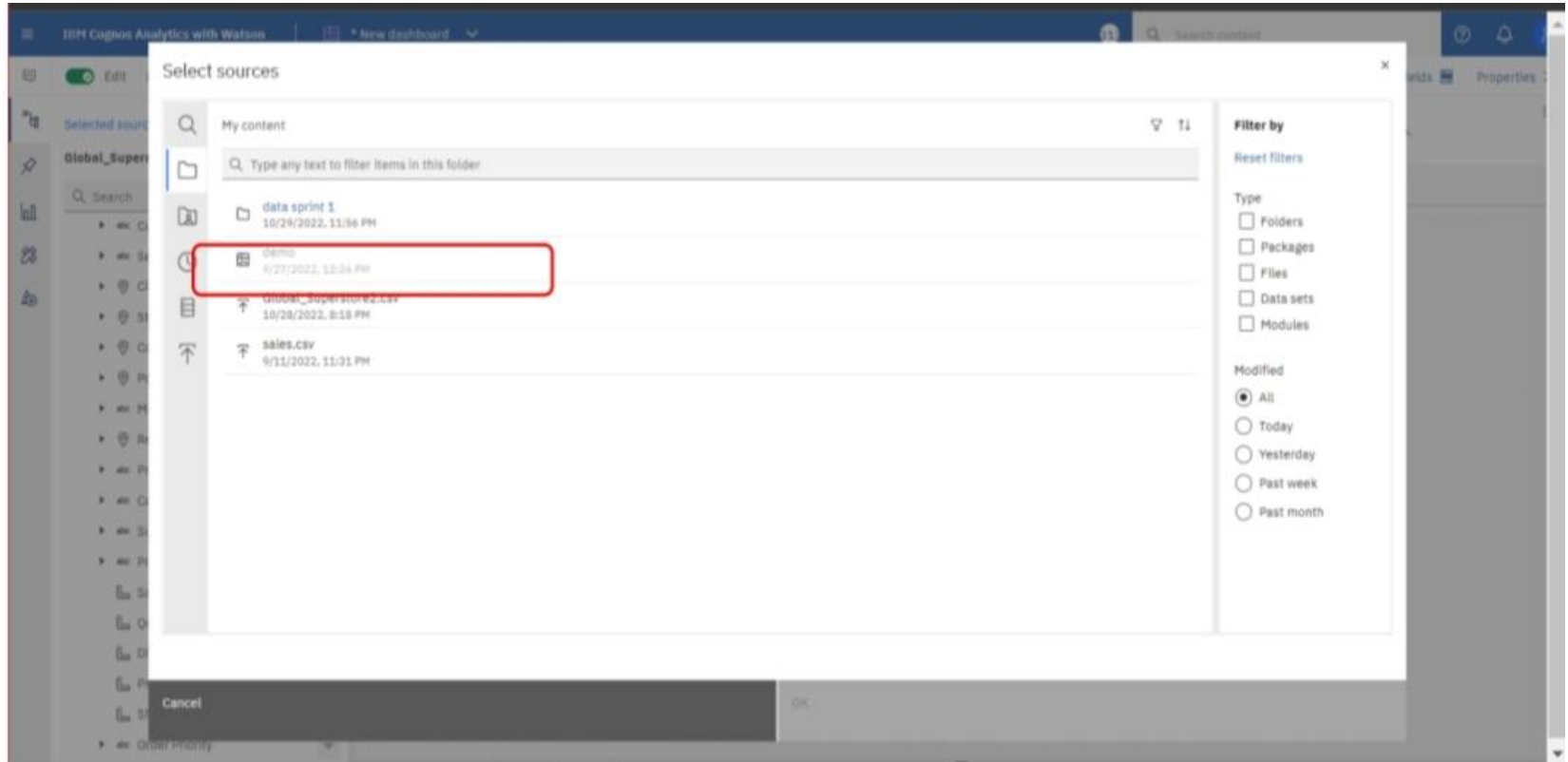
You can get started right away by taking a look at our introduction video, product tour and Getting Started tab.

[Watch video](#) [Take a product tour](#)

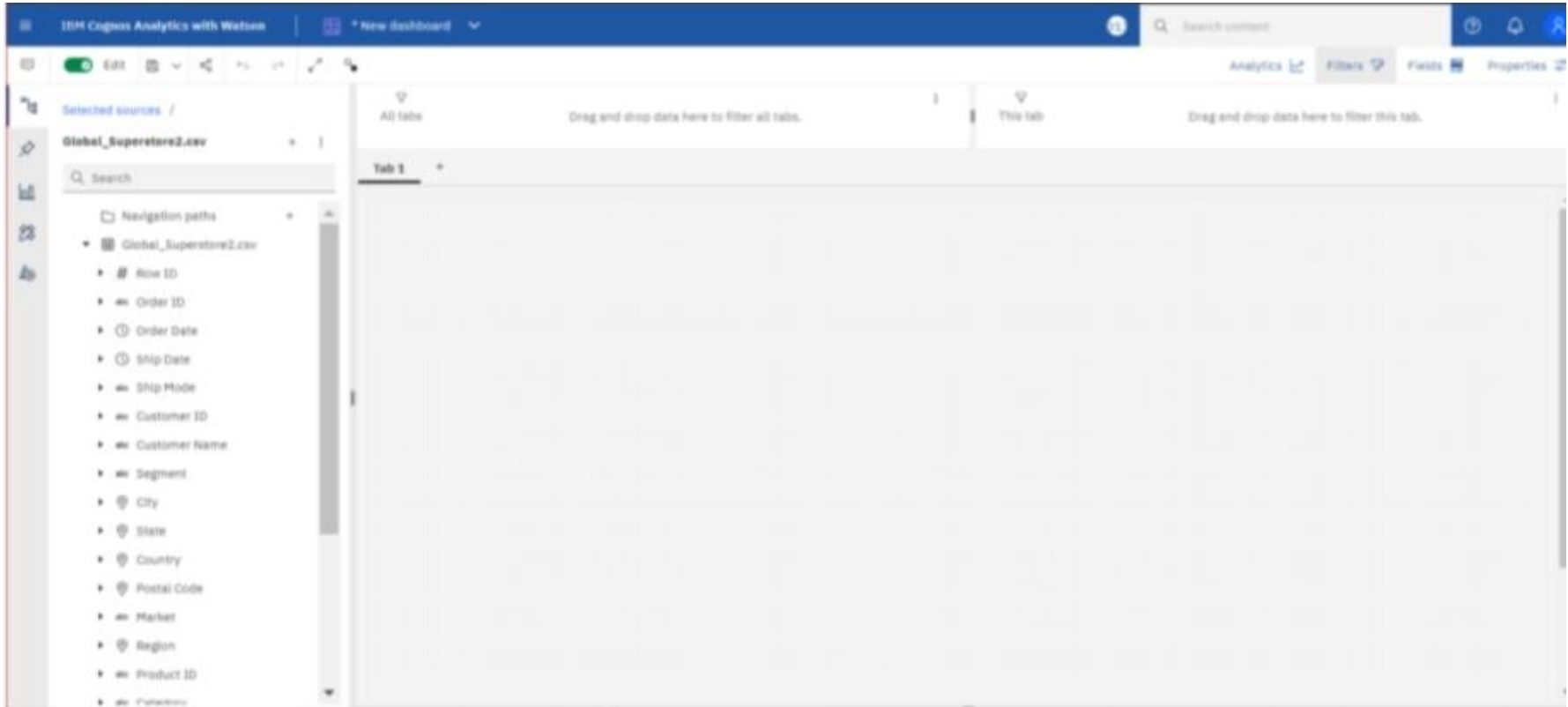
Quick launch

- Upload data**
Upload or drag and drop spreadsheets, csv files, and other data sources.
- Prepare data**
Use data modules to clean and connect data from multiple resources.
- Exploration**
Quickly find unbiased answers by identifying trends in your data with data exploration.
- Present data**
Create sophisticated, multi-page, multi-query dashboards, reports, or stories.

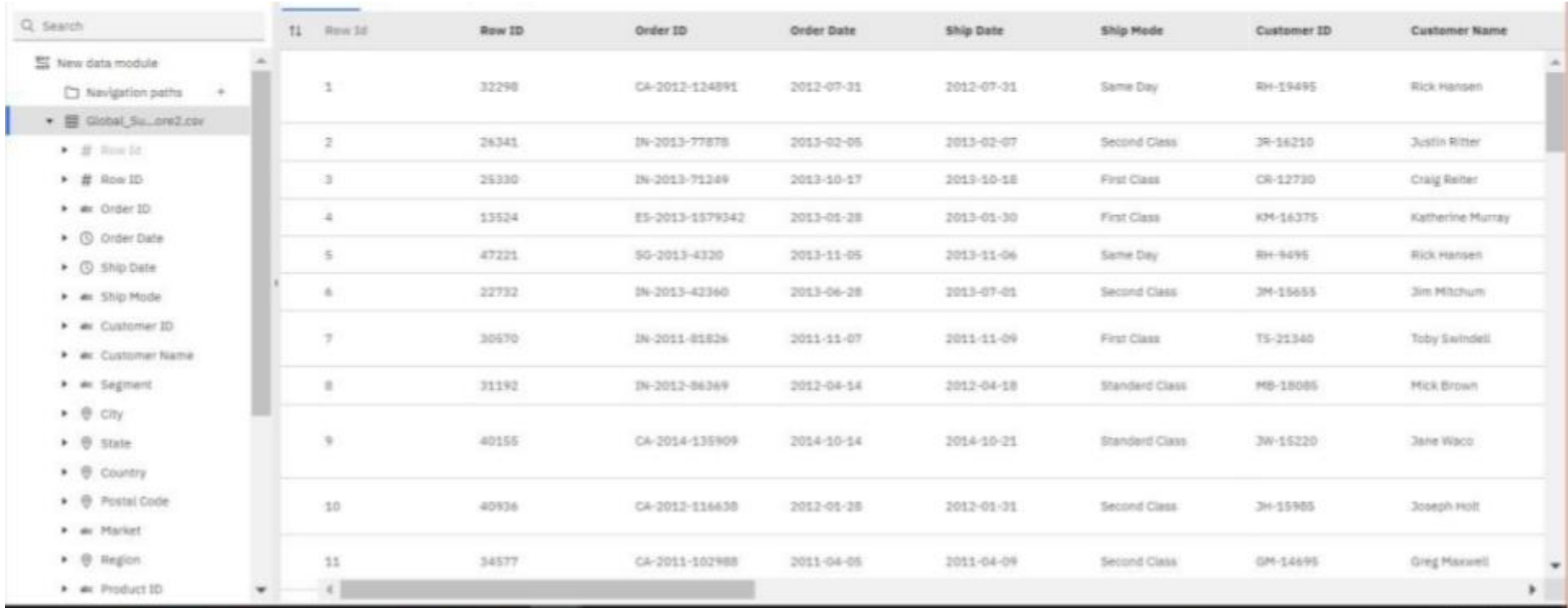
- Select source file in IBM COGNOS



- New data model for prepare data



- Loading dataset in Grid



Search

New data module

Navigation paths

Global_Sales2.csv

- Row ID
- Row ID
- Order ID
- Order Date
- Ship Date
- Ship Mode
- Customer ID
- Customer Name
- Segment
- City
- State
- Country
- Postal Code
- Market
- Region
- Product ID

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
1	CA-2012-124891	2012-07-31	2012-07-31	Same Day	RH-19495	Rick Hansen
2	IN-2013-77878	2013-02-05	2013-02-07	Second Class	JR-16210	Justin Ritter
3	IN-2013-71249	2013-10-17	2013-10-18	First Class	CR-12730	Craig Reiter
4	ES-2013-1579342	2013-01-28	2013-01-30	First Class	KM-16375	Katherine Murray
5	SG-2013-4320	2013-11-05	2013-11-06	Same Day	RH-9495	Rick Hansen
6	IN-2013-42340	2013-06-28	2013-07-01	Second Class	JM-15655	Jim Mitchum
7	IN-2011-81826	2011-11-07	2011-11-09	First Class	TS-21340	Toby Swindell
8	IN-2012-86349	2012-04-14	2012-04-18	Standard Class	MB-18085	Mick Brown
9	CA-2014-135909	2014-10-14	2014-10-21	Standard Class	JW-15220	Jahe Waco
10	CA-2012-116638	2012-01-28	2012-01-31	Second Class	JH-15985	Joseph Holt
11	CA-2011-102988	2011-04-05	2011-04-09	Second Class	GM-14695	Greg Maxwell

- Read the file data in Grid

The screenshot displays the IBM Cognos Analytics with Watson interface. The top navigation bar includes the product name, a 'New data module' button, a search bar, and user profile icons. The main workspace is divided into three sections: a left-hand 'Data module' pane, a central 'Grid' view, and a right-hand 'Properties' pane.

The 'Data module' pane on the left lists various data fields: Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, City, State, Country, Postal Code, Market, Region, Product ID, Category, Sub-Category, Product Name, Sales, and Quantity. The 'Order ID' field is currently selected.

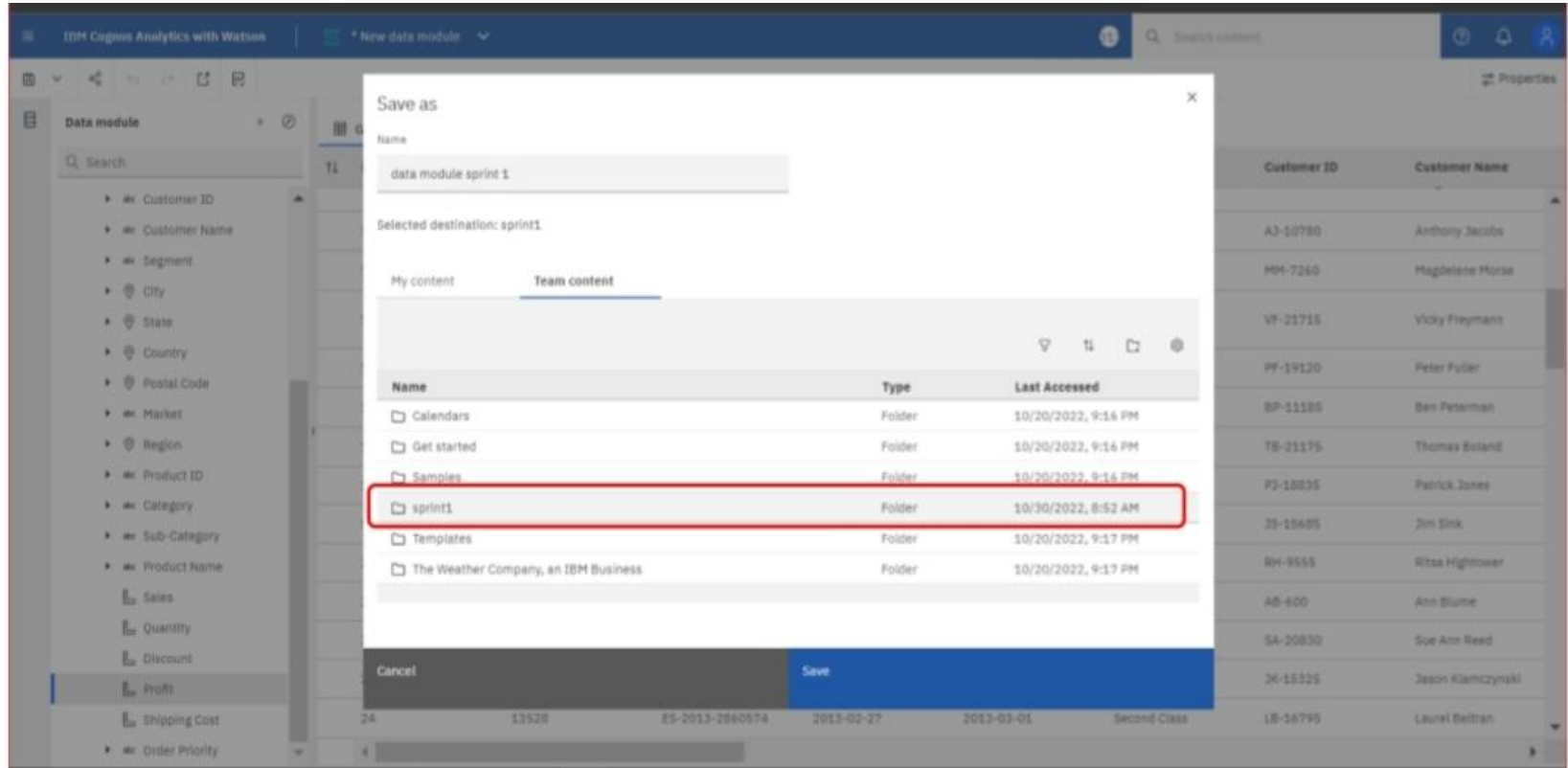
The central 'Grid' view shows a table with 11 rows of data. The columns are labeled: Row Id, Row ID, Order ID, Order Date, Ship Date, and Ship Mode. The data rows are as follows:

Row Id	Row ID	Order ID	Order Date	Ship Date	Ship Mode
1	32298	CA-2012-124891	2012-07-31	2012-07-31	Same Day
2	26341	IN-2013-77878	2013-02-06	2013-02-07	Second Class
3	26330	IN-2013-71249	2013-10-17	2013-10-18	First Class
4	13624	ES-2013-1579342	2013-01-28	2013-01-30	First Class
5	47221	SG-2013-4320	2013-11-06	2013-11-06	Same Day
6	22732	IN-2013-42360	2013-06-28	2013-07-01	Second Class
7	30670	IN-2011-81826	2011-11-07	2011-11-09	First Class
8	31192	IN-2012-86369	2012-04-14	2012-04-18	Standard Class
9	40156	CA-2014-138909	2014-10-14	2014-10-21	Standard Class
10	40936	CA-2012-116638	2012-01-28	2012-01-31	Second Class
11	34577	CA-2011-102988	2011-04-06	2011-04-09	Second Class

The 'Properties' pane on the right is open for the 'Order ID' field. It shows the following details:

- Label:** Order ID
- Hide from users:** Enabled (toggle switch)
- Expression:** View or edit >
- Usage:** Identifier (dropdown menu)
- Aggregate:** Count Distinct (dropdown menu)
- Data type:** Text
- Represents:** Default (dropdown menu)
- Description:** (empty text area)
- Comments:** (empty text area)

- Save as the preparing data



((Data Preparation, Data Cleaning, Data Format))

The screenshot displays the Visual Studio Code interface with a Jupyter Notebook titled 'global_sales_data_analysis.ipynb'. The notebook contains three cells of Python code:

```
import numpy as np
import pandas as pd
import seaborn as sb
from matplotlib import pyplot as plt
import sklearn
```

Execution time: 0.1s

global sale data analysis

```
dat=pd.read_csv('sales.csv', encoding='latin-1')
```

Execution time: 0.3s

```
dat.head()
```

Execution time: 0.1s

The output of the third cell is a table showing the first few rows of the 'sales.csv' dataset:

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	Ship...
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 CSS10 - Over-the-Head monaural Wir...	2309.650	7	0.0	762.1845	93
IN-...	05-...	07-...										Novimex Executive					

FileEditSelectionViewGoRunTerminalHelpglobal_sales_data_analysis.ipynb - Visual Studio Code

assignment TL - 4.ipynbglobal_sales_data_analysis.ipynb01-Introduction to Pandas.ipynb

C:\Users> admin> Desktop> itm workspace> completed work doc> completed assignment data> global_sales_data_analysis.ipynb> global sale data analysis> dat.head()

+ Code + Markdown | ▶ Run All | Clear Outputs of All Cells | Restart | Variables | Outline | Python 3.10.4 64-bit

describe the excel

```
dat.describe()
```

[x] ✓ 0.1s Python

	Row ID	Postal Code	Sales	Quantity	Discount	Profit	Shipping Cost
count	51290.00000	9994.000000	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000
mean	25645.50000	55190.379428	246.490581	3.476545	0.142908	28.610982	26.375915
std	14806.29199	32063.693350	487.565361	2.278766	0.212280	174.340972	57.296804
min	1.00000	1040.000000	0.444000	1.000000	0.000000	-6599.978000	0.000000
25%	12823.25000	23223.000000	30.758625	2.000000	0.000000	0.000000	2.610000
50%	25645.50000	56430.500000	85.053000	3.000000	0.000000	9.240000	7.790000
75%	38467.75000	90008.000000	251.053200	5.000000	0.200000	36.810000	24.450000
max	51290.00000	99301.000000	22638.480000	14.000000	0.850000	8399.976000	933.570000

removes the rows that contains NULL values

```
dat.dropna()
```

[x] ✓ 0.2s Python

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
													Plantronics CS510 -			

FileEditSelectionViewGoRunTerminalHelpglobal_sales_data_analysis.ipynb - Visual Studio Code

assignment 71 - 4.ipynbglobal_sales_data_analysis.ipynb01-Introduction to Pandas.ipynb

C:\Users\admin\Desktop> ibm workspace > completed work doc > completed assignment data > global_sales_data_analysis.ipynb > global sale data analysis > dat.head()

+ Code + Markdown | ▶ Run All | Clear Outputs of All Cells | Restart | Variables | Outline

Python 3.10.4 64-bit

removes the rows that contains NULL values

```
dat.dropna()
```

[14] ✓ 0.2s

Python

	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
0	32298	CA-124891	2012-07-2012	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 Wir...	2309.650	7	0.0
8	40155	CA-135909	2014-10-2014	14-10-2014	Standard Class	JW-15220	Jane Waco	Corporate	Sacramento	California	-	OFF-BI-10003527	Office Supplies	Binders	Fellowes PB500 Electric Punch Plastic Comb Bin...	5083.960	5	0.2
9	40936	CA-116638	2012-01-2012	28-01-2012	Second Class	JH-15985	Joseph Holt	Consumer	Concord	North Carolina	-	FUR-TA-10000198	Furniture	Tables	Chromcraft Bull-Nose Wood Oval Conference Tabl...	4297.644	13	0.4
10	34577	CA-102988	2011-04-2011	05-04-2011	Second Class	GM-14695	Greg Maxwell	Corporate	Alexandria	Virginia	-	OFF-SU-10002881	Office Supplies	Supplies	Martin Yale Chadless Opener Electric Letter Op...	4164.050	5	0.0

FileEditSelectionViewGoRunTerminalHelpglobal_sales_data_analysis.ipynb - Visual Studio Code

assignment TL - 4.ipynbglobal_sales_data_analysis.ipynb01-Introduction to Pandas.ipynb

C:\Users\> admin > Desktop > IBM workspace > completed work doc > completed assignment data > global_sales_data_analysis.ipynb > global sale data analysis > dat.head()

+ Code + Markdown | ▶ Run All | Clear Outputs of All Cells | Restart | Variables | Outline | Python 3.10.4 64-bit

4 columns

dat.tail()

✓ 0.2s

Python Python

	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
51285	29002	IN-2014-06-19-06-2366	19-06-2014	19-06-2014	Same Day	KE-16420	Katrina Edelman	Corporate	Kure	Hiroshima	OFF-FA-10000746	Office Supplies	Fasteners	Advantus Thumb Tacks, 12 Pack	65.100	5	0.0	4.5000
51286	35398	US-2014-06-20-06-102288	20-06-2014	24-06-2014	Standard Class	ZC-21910	Zuschuss Carroll	Consumer	Houston	Texas	OFF-AP-10002906	Office Supplies	Appliances	Hoover Replacement Belt for Commercial Guardsm...	0.444	1	0.8	-1.1100
51287	40470	US-2013-12-02-12-155768	02-12-2013	02-12-2013	Same Day	LB-16795	Laurel Beltran	Home Office	Oxnard	California	OFF-EN-10001219	Office Supplies	Envelopes	#10- 4 1/8" x 9 1/2" Security-Tint Envelopes	22.920	3	0.0	11.2308
51288	9596	MX-2012-02-18-02-140767	02-02-2012	22-02-2012	Standard Class	RB-19795	Ross Baird	Home Office	Valinhos	São Paulo	OFF-BI-10000806	Office Supplies	Binders	Acco Index Tab, Economy	13.440	2	0.0	2.4000
51289	6147	MX-2012-05-22-05-134460	22-05-2012	26-05-2012	Second Class	MC-18100	Mick Crebagga	Consumer	Tipitapa	Managua	OFF-PA-10004155	Office Supplies	Paper	Eaton Computer Printout Paper, 8.5 x 11	61.380	3	0.0	1.8000

FileEditSelectionViewGoRunTerminalHelpglobal_sales_data_analysis.ipynb - Visual Studio Code

assignment TL - 4.ipynbglobal_sales_data_analysis.ipynb01-Introduction to Pandas.ipynb

C > Users > admin > Desktop > itm workspace > completed work doc > completed assignment data > global_sales_data_analysis.ipynb > global sale data analysis > dat.head()

+ Code + Markdown | Run All Clear Outputs of All Cells Restart | Variables Outline ...Python 3.10.4 64-bit

finding duplicates

```
print(df.duplicated())
```

[11] ✓ 0.7sPython

```
...
0      False
8      False
9      False
10     False
16     False
...
51270   False
51276   False
51277   False
51286   False
51287   False
Length: 9994, dtype: bool
```

Empty markdown cell, double click or press enter to edit.

+ Code + Markdown

Removing Duplicates

```
df.drop_duplicates(inplace = True)
print(df.to_string())
```

[11] ✓ 2.9sPython

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State
	Country	Postal Code	Market	Region	Product ID	Category	Sub-Category			
	Product Name	Sales	Quantity	Discount	Profit	Shipping Cost	Order Priority			
0	32298	CA-2012-124891	31-07-2012	31-07-2012	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York
States	10024.0	US	East	TEC-AC-10003033	Technology	Accessories	Plantronics CS510			United
8	40155	CA-2014-135909	14-10-2014	21-10-2014	Standard Class	JW-15220	Jane Waco	Corporate	Sacramento	California
States	95823.0	US	West	OFF-BI-10003527	Office Supplies	Binders	Fellowes PB500			United
9	40936	CA-2012-116638	28-01-2012	31-01-2012	Second Class	JH-15985	Joseph Holt	Consumer	Concord	North Carolina
States	28027.0	US	South	FUR-TA-10000198	Furniture	Tables	Chromcraft			United
10	34577	CA-2011-102988	05-04-2011	09-04-2011	Second Class	GM-14695	Greg Maxwell	Corporate	Alexandria	Virginia
States	22304.0	US	South	OFF-SU-10002881	Office Supplies	Supplies				United
16	36178	CA-2014-143567	03-11-2014	06-11-2014	Second Class	TB-21175	Thomas Boland	Corporate	Henderson	Kentucky
States	42420.0	US	South	TEC-AC-10004145	Technology	Accessories				United
21	31784	CA-2011-154627	29-10-2011	31-10-2011	First Class	SA-20830	Sue Ann Reed	Consumer	Chicago	Illinois
States	60610.0	US	Central	TEC-PH-10001363	Technology	Phones				United
28	37311	CA-2013-159016	11-03-2013	12-03-2013	First Class	KF-16285	Karen Ferguson	Home Office	Los Angeles	California
States	90008.0	US	West	TEC-PH-10002885	Technology	Phones				United
32	32735	CA-2012-139731	15-10-2012	15-10-2012	Same Day	JE-15745	Joel Eaton	Consumer	Amarillo	Texas
States	20100.0	US	South	TEC-SU-10002003	Technology	Accessories				United

File Edit Selection View Go Run Terminal Help global_sales_data_analysis.ipynb - Visual Studio Code

assignment TL - 4.ipynb global_sales_data_analysis.ipynb 01-Introduction to Pandas.ipynb

C:\Users> admin> Desktop> itm workspace> completed work doc> completed assignment data> global_sales_data_analysis.ipynb> global sale data analysis> dat.head()

+ Code + Markdown | ▶ Run All | Clear Outputs of All Cells | Restart | Variables | Outline ... Python 3.10.4 64-bit

These rows had cells with empty values removed

```
df = dat
emp=df.dropna(inplace = True)
print(df.to_string())
```

(11) ✓ 0.1s Python

... Output exceeds the [size limit](#). Open the full output data [in a text editor](#)

<bound method DataFrame.to_string of

	Row ID	Order ID	Order Date	Ship Date	Ship Mode
0	32298	CA-2012-124891	31-07-2012	31-07-2012	Same Day
8	40155	CA-2014-135909	14-10-2014	21-10-2014	Standard Class
9	40936	CA-2012-116638	28-01-2012	31-01-2012	Second Class
10	34577	CA-2011-102988	05-04-2011	09-04-2011	Second Class
16	36178	CA-2014-143567	03-11-2014	06-11-2014	Second Class
...
51270	38414	CA-2011-143168	18-10-2011	23-10-2011	Second Class
51276	31558	US-2014-155299	09-06-2014	13-06-2014	Standard Class
51277	37361	CA-2012-111780	25-12-2012	30-12-2012	Second Class
51286	35398	US-2014-102288	20-06-2014	24-06-2014	Standard Class
51287	40470	US-2013-155768	02-12-2013	02-12-2013	Same Day

	Customer ID	Customer Name	Segment	City
0	RH-19495	Rick Hansen	Consumer	New York City
8	JW-15220	Jane Waco	Corporate	Sacramento
9	JH-15985	Joseph Holt	Consumer	Concord
10	GM-14695	Greg Maxwell	Corporate	Alexandria
16	TB-21175	Thomas Boland	Corporate	Henderson
...
51270	IG-15085	Ivan Gibson	Consumer	Seattle