

```
import pandas as pd
pd.__version__
```

```
↗ '2.1.4'
```

**Pandas** It is a python package that provides fast, flexible, and expressive data structures designed to make working with relational/labeled data both easy and initiative.

The most powerful and flexible open source data analysis/manipulation tool available in any language.

- Easy handling of missing data
- Size mutability(columns can be inserted and deleted)
- 

```
from google.colab import drive
drive.mount('/content/drive')
```

```
↗ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
#read a csv file as a data frame df
```

```
df=pd.read_csv("/content/drive/MyDrive/FSDS @Kodi Senapati/Sample - Superstore_Orders.csv")
```

```
#find the columns
print(df.columns)
```

```
#length of data frame
print(len(df))
```

```
#find shape of df
print(df.shape)
```

```
↗ Index(['Category', 'City', 'Country/Region', 'Customer Name', 'Manufacturer',
        'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
        'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category',
        'Discount', 'Profit', 'Quantity', 'Sales'],
        dtype='object')
10194
(10194, 19)
```

```
#find if any null values in df
print(df.isnull())
```

```
#By above one we cant find exact null value.
#Summarize the count of null values column wise
print('\n')
print(df.isnull().sum())
```

```
↗
```

3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
10189	False	False	False	False	False	False
10190	False	False	False	False	False	False
10191	False	False	False	False	False	False
10192	False	False	False	False	False	False
10193	False	False	False	False	False	False


```
8/28/24, 8:28 PM
Pandas(Super store project).ipynb - Colab

3      False  False
4      False  False
...      ...    ...
10189   False  False
10190   False  False
10191   False  False
10192   False  False
10193   False  False

[10194 rows x 19 columns]

Category      0
City           0
Country/Region 0
Customer Name  0
Manufacturer   0
Order Date    0
Order ID      0
Postal Code   0
Product Name   0
Region        0
Segment       0
Ship Date     0
Ship Mode     0
State/Province 0
Sub-Category  0
Discount      0
Profit        0
Quantity      0
Sales         0
dtype: int64
```

```
df.head()
#defaultly prints 5 rows
```



	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date	Ship Mode	Stat
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77015	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	07-01-2020	Standard Class	
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office	08-01-2020	Standard Class	
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office	08-01-2020	Standard Class	
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office	08-01-2020	Standard Class	
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer	12-01-2020	Standard Class	F


```
df.head(2)
#prints 2 rows
```



	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date	Ship Mode	State/
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77015	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	07-01-2020	Standard Class	
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office	08-01-2020	Standard Class	

```
df[1:10:3]


# 1 to 10 with step 3
```





	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date	Ship Mode	Stat
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office	08-01-2020	Standard Class	
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer	12-01-2020	Standard Class	F
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	30605	Dixon Prang Watercolor Pencils, 10-Color Set w...	South	Corporate	07-01-2020	First Class	

```
df.describe()


#Gives overview of data of numbered columns
```



	Discount	Profit	Quantity	Sales	
count	10194.000000	10194.000000	10194.000000	10194.000000	
mean	0.155385	28.673417	3.791838	228.225854	
std	0.206249	232.465115	2.228317	619.906839	
min	0.000000	-6599.978000	1.000000	0.444000	
25%	0.000000	1.760800	2.000000	17.220000	
50%	0.200000	8.690000	3.000000	53.910000	
75%	0.200000	29.297925	5.000000	209.500000	
max	0.800000	8399.976000	14.000000	22638.480000	

```
#The info() method prints information about the DataFrame.
#The information contains the number of columns, column labels, column data types, memory usage, range index, and the number of cells in eac

df.info()
```



<class 'pandas.core.frame.DataFrame'>			
RangeIndex: 10194 entries, 0 to 10193			
Data columns (total 19 columns):			
#	Column	Non-Null Count	Dtype
---	-----	-----	-----

```

0    Category      10194 non-null object
1    City          10194 non-null object
2    Country/Region 10194 non-null object
3    Customer Name  10194 non-null object
4    Manufacturer   10194 non-null object
5    Order Date     10194 non-null object
6    Order ID       10194 non-null object
7    Postal Code    10194 non-null object
8    Product Name   10194 non-null object
9    Region         10194 non-null object
10   Segment       10194 non-null object
11   Ship Date      10194 non-null object
12   Ship Mode      10194 non-null object
13   State/Province 10194 non-null object
14   Sub-Category   10194 non-null object
15   Discount       10194 non-null float64
16   Profit         10194 non-null float64
17   Quantity       10194 non-null int64
18   Sales          10194 non-null float64
dtypes: float64(3), int64(1), object(15)
memory usage: 1.5+ MB

```

```
#get column wise details
```

```
print(df[['Category','City']])
```

```
print('\n')
```

```
print(df[['Order Date','Order ID','Product Name']])
```

```

Category      City
0    Office Supplies  Houston
1    Office Supplies  Naperville
2    Office Supplies  Naperville
3    Office Supplies  Naperville
4    Office Supplies  Philadelphia
...
10189 Office Supplies  New York City
10190 Office Supplies  Fairfield
10191 Office Supplies  Loveland
10192 Technology      New York City
10193 Office Supplies  Charlottetown

```

```
[10194 rows x 2 columns]
```

```

Order Date      Order ID \
0    03-01-2020  US-2020-103800
1    04-01-2020  US-2020-112326
2    04-01-2020  US-2020-112326
3    04-01-2020  US-2020-112326
4    05-01-2020  US-2020-141817
...
10189 30-12-2023  US-2023-143259
10190 30-12-2023  US-2023-115427
10191 30-12-2023  US-2023-156720
10192 30-12-2023  US-2023-143259
10193 30-12-2023  CA-2023-143500

```

```

Product Name
0    Message Book, Wirebound, Four 5 1/2" X 4" Form...
1    GBC Standard Plastic Binding Systems Combs
2    Avery 508
3    SAFCO Boltless Steel Shelving
4    Avery Hi-Liter EverBold Pen Style Fluorescent ...
...
10189 Wilson Jones Legal Size Ring Binders
10190 GBC Binding covers
10191 Bagged Rubber Bands
10192 Gear Head AU3700S Headset
10193 Wilson Jones Impact Binders

```

```
[10194 rows x 3 columns]
```

```

# Find data types of each column
print(df.dtypes)
print('\n')
# Find the count of columns
print(len(df.dtypes))

```

```

Category      object
City           object
Country/Region object
Customer Name  object
Manufacturer   object
Order Date     object
Order ID       object
Postal Code    object
Product Name   object
Region         object
Segment        object
Ship Date      object
Ship Mode      object
State/Province object
Sub-Category   object
Discount       float64
Profit         float64
Quantity       int64
Sales          float64
dtype: object

```

19

```
#split the categorial and number columns
```

```
df_categorial=df[['Category', 'City', 'Country/Region', 'Customer Name', 'Manufacturer',
                  'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
                  'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category']]
```

```
df_numerical=df[['Discount', 'Profit', 'Quantity', 'Sales']]
```

```
print(len(df.columns))
print(len(df_categorial.columns))
print(len(df_numerical.columns))
```

```

19
15
4

```

Start coding or [generate](#) with AI.

## Data Cleaning


```
# Drop rows if any column value is NULL
```

```
df.dropna()
```

```
#Else you can replace the blank spaces with any default value(ex:100)
```

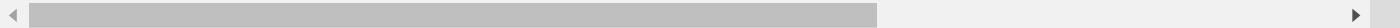
```
#df.fillna(100,inplace=True)
```

```
#df['Column'].fillna(100,inplace=True)
```




	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date	
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77015	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	07-01-2020	Stan C
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office	08-01-2020	Stan C
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office	08-01-2020	Stan C
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office	08-01-2020	Stan C
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer	12-01-2020	Stan C
...	...	...	...	...	...	...	...	...	...	...	...	...	
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259	10009	Wilson Jones Legal Size Ring Binders	East	Consumer	03-01-2024	Stan C
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427	94533	GBC Binding covers	West	Corporate	03-01-2024	Stan C
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720	80538	Bagged Rubber Bands	West	Consumer	03-01-2024	Stan C
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259	10009	Gear Head AU3700S Headset	East	Consumer	03-01-2024	Stan C
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500	C0A	Wilson Jones Impact Binders	East	Consumer	03-01-2024	Stan C

10194 rows × 19 columns



# Drop duplicate rows

```
df.drop_duplicates()
```



	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date	
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77015	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	07-01-2020	Stan C
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office	08-01-2020	Stan C
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office	08-01-2020	Stan C
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office	08-01-2020	Stan C
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer	12-01-2020	Stan C
...	...	...	...	...	...	...	...	...	...	...	...	...	