

In []:

Day-18(21-08-2024)PANDAS -- framework to handle data frames

PANDAS -- framework to handle data frames

fun() = arg

data analysis --> np, plt,sns,pandas---(family) datascience model building

packages = framework=library

PANDAS-->np+pandas+plt

excel sheet if any values are missing (you plot -- graph dont has meaning) pandas will handle missing data
NUMPY - MULTIDIMENSIONAL ARRAY
PANDAS - DATAFRAME
MATPLOTLIB - VISUALIZATION OF THE EXCEL SHEET WHICH WE FEED TO MACHINE
SEABORN -- STATISTIC VISUALIZATION (ADVANCED VISUALIZATION)

COLUMN NAME == ATTRIBUTE == FEATURE
EXCEL SHEET == DATASET
RAW DATA -- DATASET MUST HAVE SOME MISSING VALUE
CLEAN DATA -- DATA SET DOESNOT HAVE ANY MISSING VALUE
TEXT DATA -- CATEGORICAL DATA
NUMBER - NUMERICAL DATA

DATA ANALYST OR DATASCIENCE MOST OF DATASET (EXCEL SHEET) WE WILL WORK WITH CSV COMMA SEPARATED VALUE
len() in pd .columns .shape .isnull() .isnull().sum() .isna() = .i

statistic 3 part --> - descriptive stats inferential stats advanced statistics

```
In [2]: import pandas as pd
```

```
In [4]: pd.__version__
```

```
Out[4]: '2.2.2'
```

```
In [6]: store = pd.read_csv(r'E:\Data Science & AI\Sample - Superstore_Orders.csv')
```

```
In [8]: store
```

Out[8]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 19 columns



In [10]: store.shape

Out[10]: (10194, 19)

```
In [82]: store.columns # columns of the dataset
```

```
Out[82]: 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')
```

```
In [30]: len(store)
```

Out[30]: 10194

```
In [32]: len(store.columns)
```

Out[32]: 19

```
In [34]: store.isnull()
```

```
Out[34]:
```

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
10189	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False

10194 rows × 19 columns



```
In [36]: store.isnull().sum()
```

```
Out[36]: 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
```

```
In [38]: store.dtypes
```

```
Out[38]: 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
```

```
In [40]: store.isna()
```

Out[40]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
10189	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False

10194 rows × 19 columns



In [42]:

```
store.head()
```

Out[42]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19

In [44]: store.tail()

Out[44]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Post Code
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143	
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115	
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156	
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143	
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143	

In [46]: store.head(2)

Out[46]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Post Code
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	7705
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	6054

In [48]: store[:]

Out[48]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 19 columns



In [50]: store[:,:]

Out[50]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 19 columns



In [52]: store[:, :-1]

Out[52]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
...
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103

10194 rows × 19 columns



In [54]: store[1:10:3]

Out[54]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	30

In [56]: store.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
```

In [58]: store['Category']

```
Out[58]: 0      Office Supplies
1      Office Supplies
2      Office Supplies
3      Office Supplies
4      Office Supplies
...
10189   Office Supplies
10190   Office Supplies
10191   Office Supplies
10192      Technology
10193   Office Supplies
Name: Category, Length: 10194, dtype: object
```

```
In [60]: store[['Category', 'Customer Name']]
```

```
Out[60]:
```

	Category	Customer Name
0	Office Supplies	Darren Powers
1	Office Supplies	Phillina Ober
2	Office Supplies	Phillina Ober
3	Office Supplies	Phillina Ober
4	Office Supplies	Mick Brown
...
10189	Office Supplies	Patrick O'Donnell
10190	Office Supplies	Erica Bern
10191	Office Supplies	Jill Matthias
10192	Technology	Patrick O'Donnell
10193	Office Supplies	Harry Olson

10194 rows × 2 columns

```
In [62]: len(store.dtypes)
```

```
Out[62]: 19
```

```
In [70]: store.columns
```

```
Out[70]: 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')
```

```
In [72]: store_categorical = store[['Category', 'City', 'Country/Region', 'Customer Name',
                                   'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
                                   'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category',
                                   'Discount', 'Profit', 'Quantity', 'Sales']]
```

```
In [74]: store_categorical.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
```

```
In [76]: store_numerical = store[['Discount','Profit','Quality','Sales']]
store_numerical.dtypes
```

```
-----
KeyError                                Traceback (most recent call last)
Cell In[76], line 1
----> 1 store_numerical = store[['Discount','Profit','Quality','Sales']]
      2 store_numerical.dtypes

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:4108, in DataFrame.__getitem__(self, key)
    4106     if is_iterator(key):
    4107         key = list(key)
-> 4108     indexer = self.columns._get_indexer_strict(key, "columns")[1]
    4110 # take() does not accept boolean indexers
    4111 if getattr(indexer, "dtype", None) == bool:

File ~\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:6200, in Index._get_indexer_strict(self, key, axis_name)
    6197 else:
    6198     keyarr, indexer, new_indexer = self._reindex_non_unique(keyarr)
-> 6200 self._raise_if_missing(keyarr, indexer, axis_name)
    6202 keyarr = self.take(indexer)
    6203 if isinstance(key, Index):
    6204     # GH 42790 - Preserve name from an Index

File ~\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:6252, in Index._raise_if_missing(self, key, indexer, axis_name)
    6249     raise KeyError(f"None of [{key}] are in the [{axis_name}]")
    6251 not_found = list(ensure_index(key)[missing_mask.nonzero()[0]].unique())
-> 6252 raise KeyError(f"{not_found} not in index")

KeyError: "['Quality'] not in index"
```

```
In [78]: id(store) #address location
```

```
Out[78]: 2863272692864
```

```
In [80]: len(store) # no. of rows from the dataset
```

```
Out[80]: 10194
```

```
In [84]: len(store.columns) # len of the columns
```

```
Out[84]: 19
```

```
In [86]: store.shape # dimension of the dataset ( rows, column )
```

```
Out[86]: (10194, 19)
```

```
In [88]: store.dtypes
```

```
Out[88]: 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
```

```
In [90]: store.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

```

In [92]: `type(store)`

Out[92]: `pandas.core.frame.DataFrame`

In [94]: `store.isnull()`

Out[94]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
10189	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False

10194 rows × 9 columns




```
In [96]: store.isnull()
```

Out[96]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
10189	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False

10194 rows × 19 columns



```
In [98]: store.isnull().sum()
```

Out[98]:

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

```
In [102... store_1 = pd.read_csv(r'E:\Data Science & AI\Sample - Superstore_Orders_1.csv')
```

```
In [104... store_1
```


Out[104...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 19 columns



In [106...

store_1.isnull()

Out[106...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
10189	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False

10194 rows × 19 columns



In [108...

```
store_1.isnull().sum()
```

Out[108...

```
Category      1
City          0
Country/Region 1
Customer Name 1
Manufacturer   0
Order Date    0
Order ID      1
Postal Code   0
Product Name  0
Region        1
Segment       0
Ship Date     0
Ship Mode     0
State/Province 0
Sub-Category  1
Discount      0
Profit        0
Quantity      1
Sales         0
dtype: int64
```

In [110...

```
store_1.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10194 entries, 0 to 10193
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Category              10193 non-null  object
1   City                  10194 non-null  object
2   Country/Region        10193 non-null  object
3   Customer Name         10193 non-null  object
4   Manufacturer           10194 non-null  object
5   Order Date            10194 non-null  object
6   Order ID              10193 non-null  object
7   Postal Code           10194 non-null  object
8   Product Name          10194 non-null  object
9   Region                10193 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          10193 non-null  object
15  Discount              10194 non-null  float64
16  Profit                10194 non-null  float64
17  Quantity              10193 non-null  float64
18  Sales                 10194 non-null  float64
dtypes: float64(4), object(15)
memory usage: 1.5+ MB

```

In [112... `store.columns`

Out[112... `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')`

In [114... `store.info()`*# categorical & numreical data*

```

<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

```

In [116... `store.columns`

Out[116... 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')

In [118... `store['Category']`

Out[118... 0 Office Supplies
1 Office Supplies
2 Office Supplies
3 Office Supplies
4 Office Supplies
...
10189 Office Supplies
10190 Office Supplies
10191 Office Supplies
10192 Technology
10193 Office Supplies
Name: Category, Length: 10194, dtype: object

In [120... `store[['Customer Name','City']]`

Out[120...

	Customer Name	City
0	Darren Powers	Houston
1	Phillina Ober	Naperville
2	Phillina Ober	Naperville
3	Phillina Ober	Naperville
4	Mick Brown	Philadelphia
...
10189	Patrick O'Donnell	New York City
10190	Erica Bern	Fairfield
10191	Jill Matthias	Loveland
10192	Patrick O'Donnell	New York City
10193	Harry Olson	Charlottetown

10194 rows × 2 columns

In [122...

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

In [124...

```
store_categorical
```

Out[124...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 15 columns



In [126...

store_categorical.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10194 entries, 0 to 10193
Data columns (total 15 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
dtypes: object(15)
memory usage: 1.2+ MB

```

```

In [128... store_numerical = store[['Discount', 'Profit', 'Quantity', 'Sales']]
store_numerical

```

Out[128...

	Discount	Profit	Quantity	Sales
0	0.2	5.5512	2	16.448
1	0.8	-5.4870	2	3.540
2	0.2	4.2717	3	11.784
3	0.2	-64.7748	3	272.736
4	0.2	4.8840	3	19.536
...
10189	0.2	19.7910	3	52.776
10190	0.2	6.4750	2	20.720
10191	0.2	-0.6048	3	3.024
10192	0.0	2.7279	7	90.930
10193	0.2	-0.6048	3	3.024

10194 rows × 4 columns

```

In [130... store.head() # top 5 rows

```

Out[130...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19



In [132...

```
store.tail() # bottom 5 records
```


Out[132...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143



In [134...

```
store[:]
```

Out[134...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 19 columns



In [136...

store[0:10]

Out[136...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77
1	Office Supplies	Naperville	United States	Phyllina Ober	GBC	04-01-2020	US-2020-112326	60
2	Office Supplies	Naperville	United States	Phyllina Ober	Avery	04-01-2020	US-2020-112326	60
3	Office Supplies	Naperville	United States	Phyllina Ober	SAFCO	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199	42
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	42
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	30
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020-167199	42
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020-167199	42

Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C

In [138...

store[0:10:2]

Out[138...

Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Pos	
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	42
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020-167199	42



In [140...

store[:, -1]

Out[140...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
...
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103

10194 rows × 19 columns



In [142...

```
store[0:10:2]
```

Out [142...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	42
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020-167199	42

In [144...

```
store[:: -1]
```

Out[144...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
...
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103

10194 rows × 19 columns



In [146...

```
store[6:7]
```

Out[146...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Pos Co
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	424



In [148...

```
store[0:100:10]
```


	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-2020-167199
20	Furniture	Dover	United States	Seth Vernon	DAX	11-01-2020	US-2020-130092
30	Office Supplies	San Francisco	United States	Brian Dahlen	Tennsco	13-01-2020	US-2020-157147
40	Office Supplies	Scottsdale	United States	Toby Swindell	GBC	19-01-2020	US-2020-146591
50	Office Supplies	Jonesboro	United States	Hunter Lopez	Newell	20-01-2020	US-2020-147627
60	Furniture	Miami	United States	Tom Boeckenhauer	Ultra	21-01-2020	US-2020-110422
70	Furniture	Alexandria	United States	Shirley Daniels	GE	26-01-2020	US-2020-155502
80	Office Supplies	New York City	United States	Jasper Cacioppo	Pressboard	28-01-2020	US-2020-100328
90	Office Supplies	Escondido	United States	Mark Haberlin	GBC	04-02-2020	US-2020-104808