

super-store-data-analysis

November 21, 2024

1 Super Store Data Analysis

1.1 Pandas Workshop

1.1.1 Mohd Sameer Hussain - mohdsameerhussain28@gmail.com

1.1.2 <https://github.com/SameerHussain128?tab=repositories>

```
[1]: import pandas as pd
```

```
[2]: pd.__version__
```

```
[2]: '2.2.2'
```

```
[3]: store = pd.read_csv('dataset.csv')
```

```
[4]: store
```

```
[4]:
```

	Category	City	Country/Region	Customer Name \
0	Office Supplies	Houston	United States	Darren Powers
1	Office Supplies	Naperville	United States	Phillina Ober
2	Office Supplies	Naperville	United States	Phillina Ober
3	Office Supplies	Naperville	United States	Phillina Ober
4	Office Supplies	Philadelphia	United States	Mick Brown
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell
10190	Office Supplies	Fairfield	United States	Erica Bern
10191	Office Supplies	Loveland	United States	Jill Matthias
10192	Technology	New York City	United States	Patrick O'Donnell
10193	Office Supplies	Charlottetown	Canada	Harry Olson

	Manufacturer	Order Date	Order ID	Postal Code \
0	Message Book	03-01-2020	US-2020-103800	77095
1	GBC	04-01-2020	US-2020-112326	60540
2	Avery	04-01-2020	US-2020-112326	60540
3	SAFCO	04-01-2020	US-2020-112326	60540
4	Avery	05-01-2020	US-2020-141817	19143
...
10189	Wilson Jones	30-12-2023	US-2023-143259	10009

10190	GBC	30-12-2023	US-2023-115427	94533
10191	Other	30-12-2023	US-2023-156720	80538
10192	Other	30-12-2023	US-2023-143259	10009
10193	Wilson Jones	30-12-2023	CA-2023-143500	COA

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

	Segment	Ship Date	Ship Mode	State/Province \
0	Consumer	07-01-2020	Standard Class	Texas
1	Home Office	08-01-2020	Standard Class	Illinois
2	Home Office	08-01-2020	Standard Class	Illinois
3	Home Office	08-01-2020	Standard Class	Illinois
4	Consumer	12-01-2020	Standard Class	Pennsylvania
...
10189	Consumer	03-01-2024	Standard Class	New York
10190	Corporate	03-01-2024	Standard Class	California
10191	Consumer	03-01-2024	Standard Class	Colorado
10192	Consumer	03-01-2024	Standard Class	New York
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island

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

[10194 rows x 19 columns]

```
[5]: len(store)
```

```
[5]: 10194
```

```
[6]: store.columns
```

```
[6]: 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')
```

```
[7]: store.shape
```

```
[7]: (10194, 19)
```

```
[8]: store.isnull().sum()
```

```
[8]: 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
```

2 Dataset is clean No missing values

```
[9]: store.dtypes
```

```
[9]: 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

```

```
[10]: store.isna()
```

```

[10]:
   Category  City  Country/Region  Customer Name  Manufacturer  \
0      False  False              False          False          False
1      False  False              False          False          False
2      False  False              False          False          False
3      False  False              False          False          False
4      False  False              False          False          False
...      ...    ...              ...            ...            ...
10189  False  False              False          False          False
10190  False  False              False          False          False
10191  False  False              False          False          False
10192  False  False              False          False          False
10193  False  False              False          False          False

   Order Date  Order ID  Postal Code  Product Name  Region  Segment  \
0      False    False    False        False      False    False
1      False    False    False        False      False    False
2      False    False    False        False      False    False
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

   Ship Date  Ship Mode  State/Province  Sub-Category  Discount  Profit  \
0      False    False    False          False        False    False
1      False    False    False          False        False    False

```

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

	Quantity	Sales
0	False	False
1	False	False
2	False	False
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]

```
[11]: store.head()
```

```
[11]:
```

	Category	City	Country/Region	Customer Name	Manufacturer \
0	Office Supplies	Houston	United States	Darren Powers	Message Book
1	Office Supplies	Naperville	United States	Phillina Ober	GBC
2	Office Supplies	Naperville	United States	Phillina Ober	Avery
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery

	Order Date	Order ID	Postal Code \
0	03-01-2020	US-2020-103800	77095
1	04-01-2020	US-2020-112326	60540
2	04-01-2020	US-2020-112326	60540
3	04-01-2020	US-2020-112326	60540
4	05-01-2020	US-2020-141817	19143

	Product Name	Region	Segment \
0	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer
1	GBC Standard Plastic Binding Systems Combs	Central	Home Office
2	Avery 508	Central	Home Office
3	SAFCO Boltless Steel Shelving	Central	Home Office
4	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer

	Ship Date	Ship Mode	State/Province	Sub-Category	Discount	Profit \
0	07-01-2020	Standard Class	Texas	Paper	0.2	5.5512
1	08-01-2020	Standard Class	Illinois	Binders	0.8	-5.4870
2	08-01-2020	Standard Class	Illinois	Labels	0.2	4.2717
3	08-01-2020	Standard Class	Illinois	Storage	0.2	-64.7748
4	12-01-2020	Standard Class	Pennsylvania	Art	0.2	4.8840

	Quantity	Sales
0	2	16.448
1	2	3.540
2	3	11.784
3	3	272.736
4	3	19.536

```
[12]: store.tail()
```

```
[12]:
```

	Category	City	Country/Region	Customer Name \
10189	Office Supplies	New York City	United States	Patrick O'Donnell
10190	Office Supplies	Fairfield	United States	Erica Bern
10191	Office Supplies	Loveland	United States	Jill Matthias
10192	Technology	New York City	United States	Patrick O'Donnell
10193	Office Supplies	Charlottetown	Canada	Harry Olson

	Manufacturer	Order Date	Order ID	Postal Code \
10189	Wilson Jones	30-12-2023	US-2023-143259	10009
10190	GBC	30-12-2023	US-2023-115427	94533
10191	Other	30-12-2023	US-2023-156720	80538
10192	Other	30-12-2023	US-2023-143259	10009
10193	Wilson Jones	30-12-2023	CA-2023-143500	COA

	Product Name	Region	Segment	Ship Date \
10189	Wilson Jones Legal Size Ring Binders	East	Consumer	03-01-2024
10190	GBC Binding covers	West	Corporate	03-01-2024
10191	Bagged Rubber Bands	West	Consumer	03-01-2024
10192	Gear Head AU3700S Headset	East	Consumer	03-01-2024
10193	Wilson Jones Impact Binders	East	Consumer	03-01-2024

	Ship Mode	State/Province	Sub-Category	Discount	Profit \
10189	Standard Class	New York	Binders	0.2	19.7910
10190	Standard Class	California	Binders	0.2	6.4750
10191	Standard Class	Colorado	Fasteners	0.2	-0.6048
10192	Standard Class	New York	Phones	0.0	2.7279
10193	Standard Class	Prince Edward Island	Binders	0.2	-0.6048

	Quantity	Sales
10189	3	52.776

10190	2	20.720
10191	3	3.024
10192	7	90.930
10193	3	3.024

```
[13]: store.head(2)
```

```
[13]:
```

	Category	City	Country/Region	Customer Name	Manufacturer	\
0	Office Supplies	Houston	United States	Darren Powers	Message Book	
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	

	Order Date	Order ID	Postal Code	\
0	03-01-2020	US-2020-103800	77095	
1	04-01-2020	US-2020-112326	60540	

	Product Name	Region	Segment	\
0	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	
1	GBC Standard Plastic Binding Systems Combs	Central	Home Office	

	Ship Date	Ship Mode	State/Province	Sub-Category	Discount	Profit	\
0	07-01-2020	Standard Class	Texas	Paper	0.2	5.5512	
1	08-01-2020	Standard Class	Illinois	Binders	0.8	-5.4870	

	Quantity	Sales
0	2	16.448
1	2	3.540

```
[14]: store[:]
```

```
[14]:
```

	Category	City	Country/Region	Customer Name	\
0	Office Supplies	Houston	United States	Darren Powers	
1	Office Supplies	Naperville	United States	Phillina Ober	
2	Office Supplies	Naperville	United States	Phillina Ober	
3	Office Supplies	Naperville	United States	Phillina Ober	
4	Office Supplies	Philadelphia	United States	Mick Brown	
...	
10189	Office Supplies	New York City	United States	Patrick O'Donnell	
10190	Office Supplies	Fairfield	United States	Erica Bern	
10191	Office Supplies	Loveland	United States	Jill Matthias	
10192	Technology	New York City	United States	Patrick O'Donnell	
10193	Office Supplies	Charlottetown	Canada	Harry Olson	

	Manufacturer	Order Date	Order ID	Postal Code	\
0	Message Book	03-01-2020	US-2020-103800	77095	
1	GBC	04-01-2020	US-2020-112326	60540	
2	Avery	04-01-2020	US-2020-112326	60540	
3	SAFCO	04-01-2020	US-2020-112326	60540	

4	Avery	05-01-2020	US-2020-141817	19143
...
10189	Wilson Jones	30-12-2023	US-2023-143259	10009
10190	GBC	30-12-2023	US-2023-115427	94533
10191	Other	30-12-2023	US-2023-156720	80538
10192	Other	30-12-2023	US-2023-143259	10009
10193	Wilson Jones	30-12-2023	CA-2023-143500	COA

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

	Segment	Ship Date	Ship Mode	State/Province \
0	Consumer	07-01-2020	Standard Class	Texas
1	Home Office	08-01-2020	Standard Class	Illinois
2	Home Office	08-01-2020	Standard Class	Illinois
3	Home Office	08-01-2020	Standard Class	Illinois
4	Consumer	12-01-2020	Standard Class	Pennsylvania
...
10189	Consumer	03-01-2024	Standard Class	New York
10190	Corporate	03-01-2024	Standard Class	California
10191	Consumer	03-01-2024	Standard Class	Colorado
10192	Consumer	03-01-2024	Standard Class	New York
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island

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

[10194 rows x 19 columns]

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

```
[15]:
```

	Category	City	Country/Region	Customer Name	\
10193	Office Supplies	Charlottetown	Canada	Harry Olson	
10192	Technology	New York City	United States	Patrick O'Donnell	
10191	Office Supplies	Loveland	United States	Jill Matthias	
10190	Office Supplies	Fairfield	United States	Erica Bern	
10189	Office Supplies	New York City	United States	Patrick O'Donnell	
...	
4	Office Supplies	Philadelphia	United States	Mick Brown	
3	Office Supplies	Naperville	United States	Phillina Ober	
2	Office Supplies	Naperville	United States	Phillina Ober	
1	Office Supplies	Naperville	United States	Phillina Ober	
0	Office Supplies	Houston	United States	Darren Powers	

	Manufacturer	Order Date	Order ID	Postal Code	\
10193	Wilson Jones	30-12-2023	CA-2023-143500	C0A	
10192	Other	30-12-2023	US-2023-143259	10009	
10191	Other	30-12-2023	US-2023-156720	80538	
10190	GBC	30-12-2023	US-2023-115427	94533	
10189	Wilson Jones	30-12-2023	US-2023-143259	10009	
...	
4	Avery	05-01-2020	US-2020-141817	19143	
3	SAFCO	04-01-2020	US-2020-112326	60540	
2	Avery	04-01-2020	US-2020-112326	60540	
1	GBC	04-01-2020	US-2020-112326	60540	
0	Message Book	03-01-2020	US-2020-103800	77095	

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

	Segment	Ship Date	Ship Mode	State/Province	\
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island	
10192	Consumer	03-01-2024	Standard Class	New York	
10191	Consumer	03-01-2024	Standard Class	Colorado	

10190	Corporate	03-01-2024	Standard Class	California
10189	Consumer	03-01-2024	Standard Class	New York
...
4	Consumer	12-01-2020	Standard Class	Pennsylvania
3	Home Office	08-01-2020	Standard Class	Illinois
2	Home Office	08-01-2020	Standard Class	Illinois
1	Home Office	08-01-2020	Standard Class	Illinois
0	Consumer	07-01-2020	Standard Class	Texas

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

[10194 rows x 19 columns]

```
[16]: store[1:10:3]
```

```
[16]:      Category      City Country/Region Customer Name Manufacturer \
1 Office Supplies  Naperville United States Phillina Ober      GBC
4 Office Supplies Philadelphia United States Mick Brown      Avery
7 Office Supplies Athens United States Jack O'Briant      Dixon
```

	Order Date	Order ID	Postal Code	\
1	04-01-2020	US-2020-112326	60540	
4	05-01-2020	US-2020-141817	19143	
7	06-01-2020	US-2020-106054	30605	

	Product Name	Region	Segment	\
1	GBC Standard Plastic Binding Systems Combs	Central	Home Office	
4	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer	
7	Dixon Prang Watercolor Pencils, 10-Color Set w...	South	Corporate	

	Ship Date	Ship Mode	State/Province	Sub-Category	Discount	Profit	\
1	08-01-2020	Standard Class	Illinois	Binders	0.8	-5.4870	
4	12-01-2020	Standard Class	Pennsylvania	Art	0.2	4.8840	
7	07-01-2020	First Class	Georgia	Art	0.0	5.2398	

Quantity	Sales
----------	-------

```

1          2    3.540
4          3   19.536
7          3   12.780

```

```
[17]: 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

```

```
[18]: store.columns
```

```

[18]: 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')

```

```
[19]: store['Category']
```

```

[19]: 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

```

```
[20]: store['City']
```

```

[20]: 0      Houston
      1      Naperville
      2      Naperville
      3      Naperville
      4      Philadelphia
...
10189   New York City
10190       Fairfield
10191       Loveland
10192   New York City
10193   Charlottetown
Name: City, Length: 10194, dtype: object

```

```
[21]: store['City'].value_counts()
```

```

[21]: City
New York City    915
Los Angeles      747
Philadelphia     537
San Francisco    510
Seattle          428
...
Manhattan        1
Iowa City         1
Cedar Rapids     1
Baytown          1
Danbury          1
Name: count, Length: 542, dtype: int64

```

```
[22]: store[['Category', 'City']]
```

```

[22]:      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]

```
[23]: store[['Category', 'City', 'Order ID']]
```

```

[23]:
      Category      City      Order ID
0      Office Supplies  Houston  US-2020-103800
1      Office Supplies  Naperville  US-2020-112326
2      Office Supplies  Naperville  US-2020-112326
3      Office Supplies  Naperville  US-2020-112326
4      Office Supplies  Philadelphia  US-2020-141817
...
10189 Office Supplies  New York City  US-2023-143259
10190 Office Supplies  Fairfield  US-2023-115427
10191 Office Supplies  Loveland  US-2023-156720
10192      Technology  New York City  US-2023-143259
10193 Office Supplies  Charlottetown  CA-2023-143500

```

[10194 rows x 3 columns]

```
[24]: store.dtypes
```

```

[24]: 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
```

```
[25]: len(store.dtypes)
```

```
[25]: 19
```

```
[26]: store.columns
```

```
[26]: 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')
```

```
[27]: store_categorical=store[['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']]
```

```
[28]: store_categorical
```

```
[28]:
```

	Category	City	Country/Region	Customer Name \
0	Office Supplies	Houston	United States	Darren Powers
1	Office Supplies	Naperville	United States	Phillina Ober
2	Office Supplies	Naperville	United States	Phillina Ober
3	Office Supplies	Naperville	United States	Phillina Ober
4	Office Supplies	Philadelphia	United States	Mick Brown
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell
10190	Office Supplies	Fairfield	United States	Erica Bern
10191	Office Supplies	Loveland	United States	Jill Matthias
10192	Technology	New York City	United States	Patrick O'Donnell
10193	Office Supplies	Charlottetown	Canada	Harry Olson

	Manufacturer	Order Date	Order ID	Postal Code \
0	Message Book	03-01-2020	US-2020-103800	77095
1	GBC	04-01-2020	US-2020-112326	60540
2	Avery	04-01-2020	US-2020-112326	60540
3	SAFCO	04-01-2020	US-2020-112326	60540
4	Avery	05-01-2020	US-2020-141817	19143
...
10189	Wilson Jones	30-12-2023	US-2023-143259	10009
10190	GBC	30-12-2023	US-2023-115427	94533
10191	Other	30-12-2023	US-2023-156720	80538
10192	Other	30-12-2023	US-2023-143259	10009
10193	Wilson Jones	30-12-2023	CA-2023-143500	COA

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

	Segment	Ship Date	Ship Mode	State/Province	\
0	Consumer	07-01-2020	Standard Class	Texas	
1	Home Office	08-01-2020	Standard Class	Illinois	
2	Home Office	08-01-2020	Standard Class	Illinois	
3	Home Office	08-01-2020	Standard Class	Illinois	
4	Consumer	12-01-2020	Standard Class	Pennsylvania	
...	
10189	Consumer	03-01-2024	Standard Class	New York	
10190	Corporate	03-01-2024	Standard Class	California	
10191	Consumer	03-01-2024	Standard Class	Colorado	
10192	Consumer	03-01-2024	Standard Class	New York	
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island	

	Sub-Category
0	Paper
1	Binders
2	Labels
3	Storage
4	Art
...	...
10189	Binders
10190	Binders
10191	Fasteners
10192	Phones
10193	Binders

[10194 rows x 15 columns]

```
[29]: 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

```
[30]: store_categorical.dtypes #Text data
```

```
[30]: 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
      dtype: object
```

```
[31]: store_numerical=store[['Discount', 'Profit', 'Quantity', 'Sales']] #Numbers
```

```
[32]: store_numerical
```

```
[32]:      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 x 4 columns]

```
[33]: store_numerical.dtypes
```

```
[33]: Discount    float64
      Profit      float64
      Quantity     int64
      Sales       float64
      dtype: object
```

```
[34]: store.columns
```

```
[34]: 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')
```

```
[35]: len(store.columns)
```

```
[35]: 19
```

```
[36]: store_categorical.columns
```

```
[36]: 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'],
          dtype='object')
```

```
[37]: len(store_categorical.columns)
```

```
[37]: 15
```

```
[38]: store_numerical.columns
```

```
[38]: Index(['Discount', 'Profit', 'Quantity', 'Sales'], dtype='object')
```

```
[39]: len(store_numerical.columns)
```

[39]: 4

```
[40]: print(len(store.columns))
      print(len(store_categorical.columns))
      print(len(store_numerical.columns))
```

19
15
4

```
[41]: store.head()
```

```
[41]:      Category      City Country/Region Customer Name Manufacturer \
0  Office Supplies      Houston    United States  Darren Powers  Message Book
1  Office Supplies  Naperville    United States  Phillina Ober      GBC
2  Office Supplies  Naperville    United States  Phillina Ober      Avery
3  Office Supplies  Naperville    United States  Phillina Ober      SAFCO
4  Office Supplies  Philadelphia  United States    Mick Brown      Avery
```

```
      Order Date      Order ID Postal Code \
0  03-01-2020  US-2020-103800      77095
1  04-01-2020  US-2020-112326      60540
2  04-01-2020  US-2020-112326      60540
3  04-01-2020  US-2020-112326      60540
4  05-01-2020  US-2020-141817      19143
```

```
      Product Name      Region      Segment \
0  Message Book, Wirebound, Four 5 1/2" X 4" Form...  Central      Consumer
1      GBC Standard Plastic Binding Systems Combs  Central  Home Office
2      Avery 508  Central  Home Office
3      SAFCO Boltless Steel Shelving  Central  Home Office
4  Avery Hi-Liter EverBold Pen Style Fluorescent ...  East      Consumer
```

```
      Ship Date      Ship Mode State/Province Sub-Category Discount Profit \
0  07-01-2020  Standard Class      Texas      Paper      0.2    5.5512
1  08-01-2020  Standard Class      Illinois  Binders      0.8   -5.4870
2  08-01-2020  Standard Class      Illinois  Labels      0.2    4.2717
3  08-01-2020  Standard Class      Illinois  Storage      0.2  -64.7748
4  12-01-2020  Standard Class  Pennsylvania      Art      0.2    4.8840
```

```
      Quantity      Sales
0          2    16.448
1          2     3.540
2          3    11.784
3          3   272.736
4          3    19.536
```

```
[42]: store.tail()
```

```
[42]:
```

	Category	City	Country/Region	Customer Name	\
10189	Office Supplies	New York City	United States	Patrick O'Donnell	
10190	Office Supplies	Fairfield	United States	Erica Bern	
10191	Office Supplies	Loveland	United States	Jill Matthias	
10192	Technology	New York City	United States	Patrick O'Donnell	
10193	Office Supplies	Charlottetown	Canada	Harry Olson	

	Manufacturer	Order Date	Order ID	Postal Code	\
10189	Wilson Jones	30-12-2023	US-2023-143259	10009	
10190	GBC	30-12-2023	US-2023-115427	94533	
10191	Other	30-12-2023	US-2023-156720	80538	
10192	Other	30-12-2023	US-2023-143259	10009	
10193	Wilson Jones	30-12-2023	CA-2023-143500	C0A	

		Product Name	Region	Segment	Ship Date	\
10189	Wilson Jones	Legal Size Ring Binders	East	Consumer	03-01-2024	
10190		GBC Binding covers	West	Corporate	03-01-2024	
10191		Bagged Rubber Bands	West	Consumer	03-01-2024	
10192		Gear Head AU3700S Headset	East	Consumer	03-01-2024	
10193	Wilson Jones	Impact Binders	East	Consumer	03-01-2024	

	Ship Mode	State/Province	Sub-Category	Discount	Profit	\
10189	Standard Class	New York	Binders	0.2	19.7910	
10190	Standard Class	California	Binders	0.2	6.4750	
10191	Standard Class	Colorado	Fasteners	0.2	-0.6048	
10192	Standard Class	New York	Phones	0.0	2.7279	
10193	Standard Class	Prince Edward Island	Binders	0.2	-0.6048	

	Quantity	Sales
10189	3	52.776
10190	2	20.720
10191	3	3.024
10192	7	90.930
10193	3	3.024

```
[43]: store[:]
```

```
[43]:
```

	Category	City	Country/Region	Customer Name	\
0	Office Supplies	Houston	United States	Darren Powers	
1	Office Supplies	Naperville	United States	Phillina Ober	
2	Office Supplies	Naperville	United States	Phillina Ober	
3	Office Supplies	Naperville	United States	Phillina Ober	
4	Office Supplies	Philadelphia	United States	Mick Brown	
...	
10189	Office Supplies	New York City	United States	Patrick O'Donnell	

10190	Office Supplies	Fairfield	United States	Erica Bern
10191	Office Supplies	Loveland	United States	Jill Matthias
10192	Technology	New York City	United States	Patrick O'Donnell
10193	Office Supplies	Charlottetown	Canada	Harry Olson

	Manufacturer	Order Date	Order ID	Postal Code	\
0	Message Book	03-01-2020	US-2020-103800	77095	
1	GBC	04-01-2020	US-2020-112326	60540	
2	Avery	04-01-2020	US-2020-112326	60540	
3	SAFCO	04-01-2020	US-2020-112326	60540	
4	Avery	05-01-2020	US-2020-141817	19143	
...	
10189	Wilson Jones	30-12-2023	US-2023-143259	10009	
10190	GBC	30-12-2023	US-2023-115427	94533	
10191	Other	30-12-2023	US-2023-156720	80538	
10192	Other	30-12-2023	US-2023-143259	10009	
10193	Wilson Jones	30-12-2023	CA-2023-143500	COA	

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

	Segment	Ship Date	Ship Mode	State/Province	\
0	Consumer	07-01-2020	Standard Class	Texas	
1	Home Office	08-01-2020	Standard Class	Illinois	
2	Home Office	08-01-2020	Standard Class	Illinois	
3	Home Office	08-01-2020	Standard Class	Illinois	
4	Consumer	12-01-2020	Standard Class	Pennsylvania	
...	
10189	Consumer	03-01-2024	Standard Class	New York	
10190	Corporate	03-01-2024	Standard Class	California	
10191	Consumer	03-01-2024	Standard Class	Colorado	
10192	Consumer	03-01-2024	Standard Class	New York	
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island	

	Sub-Category	Discount	Profit	Quantity	Sales
0	Paper	0.2	5.5512	2	16.448
1	Binders	0.8	-5.4870	2	3.540

2	Labels	0.2	4.2717	3	11.784
3	Storage	0.2	-64.7748	3	272.736
4	Art	0.2	4.8840	3	19.536
...
10189	Binders	0.2	19.7910	3	52.776
10190	Binders	0.2	6.4750	2	20.720
10191	Fasteners	0.2	-0.6048	3	3.024
10192	Phones	0.0	2.7279	7	90.930
10193	Binders	0.2	-0.6048	3	3.024

[10194 rows x 19 columns]

```
[44]: store[:, :-1]
```

```
[44]:
```

	Category	City	Country/Region	Customer Name \
10193	Office Supplies	Charlottetown	Canada	Harry Olson
10192	Technology	New York City	United States	Patrick O'Donnell
10191	Office Supplies	Loveland	United States	Jill Matthias
10190	Office Supplies	Fairfield	United States	Erica Bern
10189	Office Supplies	New York City	United States	Patrick O'Donnell
...
4	Office Supplies	Philadelphia	United States	Mick Brown
3	Office Supplies	Naperville	United States	Phillina Ober
2	Office Supplies	Naperville	United States	Phillina Ober
1	Office Supplies	Naperville	United States	Phillina Ober
0	Office Supplies	Houston	United States	Darren Powers

	Manufacturer	Order Date	Order ID	Postal Code \
10193	Wilson Jones	30-12-2023	CA-2023-143500	COA
10192	Other	30-12-2023	US-2023-143259	10009
10191	Other	30-12-2023	US-2023-156720	80538
10190	GBC	30-12-2023	US-2023-115427	94533
10189	Wilson Jones	30-12-2023	US-2023-143259	10009
...
4	Avery	05-01-2020	US-2020-141817	19143
3	SAFCO	04-01-2020	US-2020-112326	60540
2	Avery	04-01-2020	US-2020-112326	60540
1	GBC	04-01-2020	US-2020-112326	60540
0	Message Book	03-01-2020	US-2020-103800	77095

	Product Name	Region \
10193	Wilson Jones Impact Binders	East
10192	Gear Head AU3700S Headset	East
10191	Bagged Rubber Bands	West
10190	GBC Binding covers	West
10189	Wilson Jones Legal Size Ring Binders	East
...

```

4 Avery Hi-Liter EverBold Pen Style Fluorescent ... East
3 SAFCO Boltless Steel Shelving Central
2 Avery 508 Central
1 GBC Standard Plastic Binding Systems Combs Central
0 Message Book, Wirebound, Four 5 1/2" X 4" Form... Central

```

	Segment	Ship Date	Ship Mode	State/Province \
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island
10192	Consumer	03-01-2024	Standard Class	New York
10191	Consumer	03-01-2024	Standard Class	Colorado
10190	Corporate	03-01-2024	Standard Class	California
10189	Consumer	03-01-2024	Standard Class	New York
...
4	Consumer	12-01-2020	Standard Class	Pennsylvania
3	Home Office	08-01-2020	Standard Class	Illinois
2	Home Office	08-01-2020	Standard Class	Illinois
1	Home Office	08-01-2020	Standard Class	Illinois
0	Consumer	07-01-2020	Standard Class	Texas

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

[10194 rows x 19 columns]

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

```

[45]:
      Category      City Country/Region  Customer Name \
0 Office Supplies Houston United States  Darren Powers
10 Office Supplies Henderson United States  Maria Etezadi
20 Furniture Dover United States  Seth Vernon
30 Office Supplies San Francisco United States  Brian Dahlen
40 Office Supplies Scottsdale United States  Toby Swindell
50 Office Supplies Jonesboro United States  Hunter Lopez
60 Furniture Miami United States  Tom Boeckenhauer
70 Furniture Alexandria United States  Shirley Daniels
80 Office Supplies New York City United States  Jasper Cacioppo
90 Office Supplies Escondido United States  Mark Haberin

```

	Manufacturer	Order Date	Order ID	Postal Code	\
0	Message Book	03-01-2020	US-2020-103800	77095	
10	Southworth	06-01-2020	US-2020-167199	42420	
20	DAX	11-01-2020	US-2020-130092	19901	
30	Tennsco	13-01-2020	US-2020-157147	94109	
40	GBC	19-01-2020	US-2020-146591	85254	
50	Newell	20-01-2020	US-2020-147627	72401	
60	Ultra	21-01-2020	US-2020-110422	33180	
70	GE	26-01-2020	US-2020-155502	22304	
80	Pressboard	28-01-2020	US-2020-100328	10024	
90	GBC	04-02-2020	US-2020-104808	92025	

	Product Name	Region	Segment	\
0	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	
10	Southworth 25% Cotton Granite Paper & Envelopes	South	Home Office	
20	DAX Value U-Channel Document Frames, Easel Back	East	Consumer	
30	Tennsco 6- and 18-Compartment Lockers	West	Consumer	
40	GBC Standard Recycled Report Covers, Clear Pla...	West	Consumer	
50	Newell 327	South	Consumer	
60	Ultra Door Pull Handle	South	Consumer	
70	GE General Use Halogen Bulbs, 100 Watts, 1 Bul...	South	Home Office	
80	Pressboard Covers with Storage Hooks, 9 1/2" x...	East	Consumer	
90	GBC Standard Recycled Report Covers, Clear Pla...	West	Corporate	

	Ship Date	Ship Mode	State/Province	Sub-Category	Discount	\
0	07-01-2020	Standard Class	Texas	Paper	0.2	
10	10-01-2020	Standard Class	Kentucky	Paper	0.0	
20	14-01-2020	First Class	Delaware	Furnishings	0.0	
30	18-01-2020	Standard Class	California	Storage	0.0	
40	20-01-2020	First Class	Arizona	Binders	0.7	
50	26-01-2020	Standard Class	Arkansas	Art	0.0	
60	23-01-2020	Second Class	Florida	Furnishings	0.2	
70	31-01-2020	Standard Class	Virginia	Furnishings	0.0	
80	03-02-2020	Standard Class	New York	Binders	0.2	
90	08-02-2020	Second Class	California	Binders	0.2	

	Profit	Quantity	Sales
0	5.5512	2	16.448
10	3.0084	1	6.540
20	3.0814	2	9.940
30	238.6530	5	1325.850
40	-23.7160	10	32.340
50	1.7901	3	6.630
60	4.1028	3	25.248
70	30.7818	3	62.820
80	1.3257	1	3.928

90 6.0368 2 17.248

[46]: store[10:11]

```
[46]:      Category      City Country/Region Customer Name Manufacturer \
10 Office Supplies Henderson United States Maria Etezadi Southworth

      Order Date      Order ID Postal Code \
10 06-01-2020 US-2020-167199 42420

      Product Name Region Segment \
10 Southworth 25% Cotton Granite Paper & Envelopes South Home Office

      Ship Date      Ship Mode State/Province Sub-Category Discount Profit \
10 10-01-2020 Standard Class Kentucky Paper 0.0 3.0084

      Quantity Sales
10 1 6.54
```

[47]: store_categorical[:4]

```
[47]:      Category      City Country/Region Customer Name Manufacturer \
0 Office Supplies Houston United States Darren Powers Message Book
1 Office Supplies Naperville United States Phillina Ober GBC
2 Office Supplies Naperville United States Phillina Ober Avery
3 Office Supplies Naperville United States Phillina Ober SAFCO

      Order Date      Order ID Postal Code \
0 03-01-2020 US-2020-103800 77095
1 04-01-2020 US-2020-112326 60540
2 04-01-2020 US-2020-112326 60540
3 04-01-2020 US-2020-112326 60540

      Product Name Region Segment \
0 Message Book, Wirebound, Four 5 1/2" X 4" Form... Central Consumer
1 GBC Standard Plastic Binding Systems Combs Central Home Office
2 Avery 508 Central Home Office
3 SAFCO Boltless Steel Shelving Central Home Office

      Ship Date      Ship Mode State/Province Sub-Category
0 07-01-2020 Standard Class Texas Paper
1 08-01-2020 Standard Class Illinois Binders
2 08-01-2020 Standard Class Illinois Labels
3 08-01-2020 Standard Class Illinois Storage
```

[48]: store_numerical[:, :-1]


```
[48]:
```

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

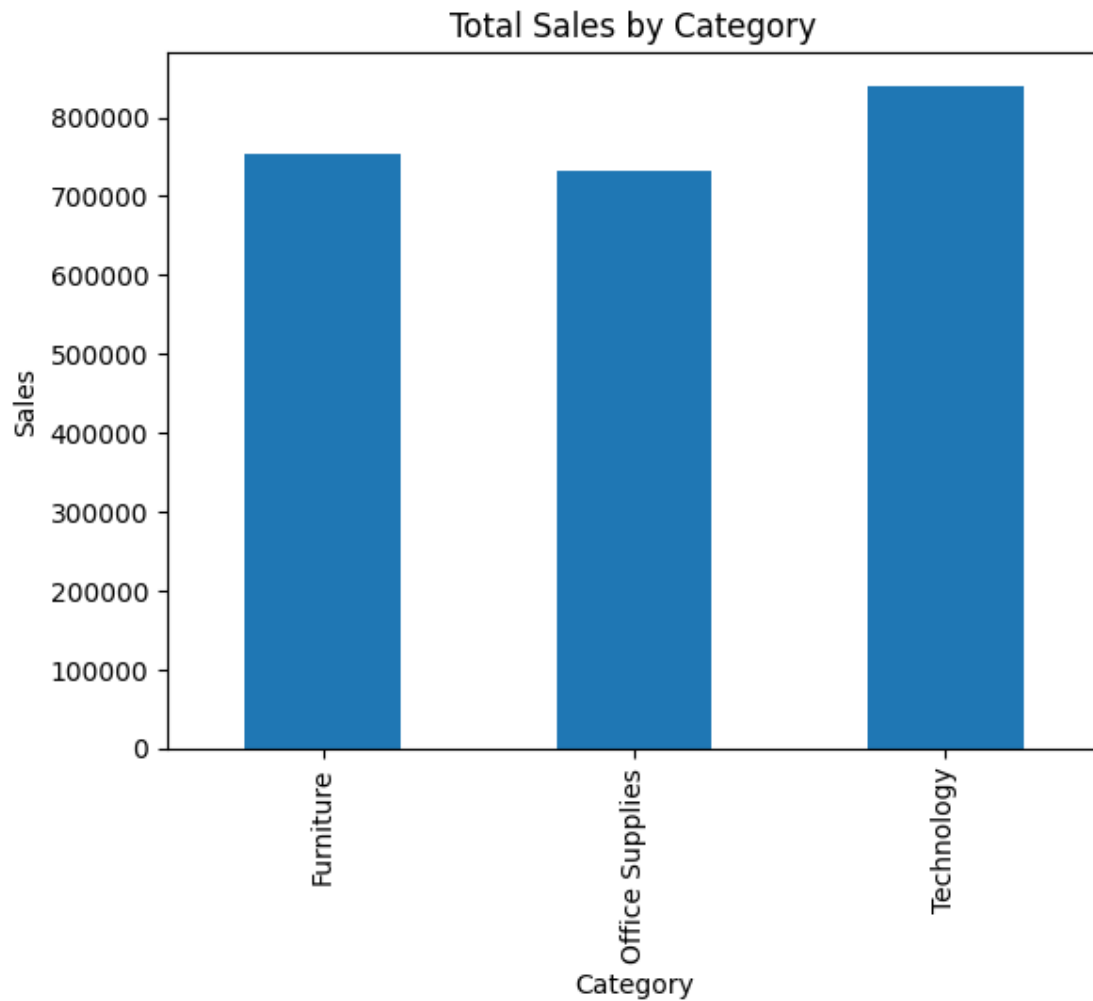
[10194 rows x 4 columns]

3 VISUALIZATION

```
[49]: # 1. Bar chart of sales by category
import matplotlib.pyplot as plt

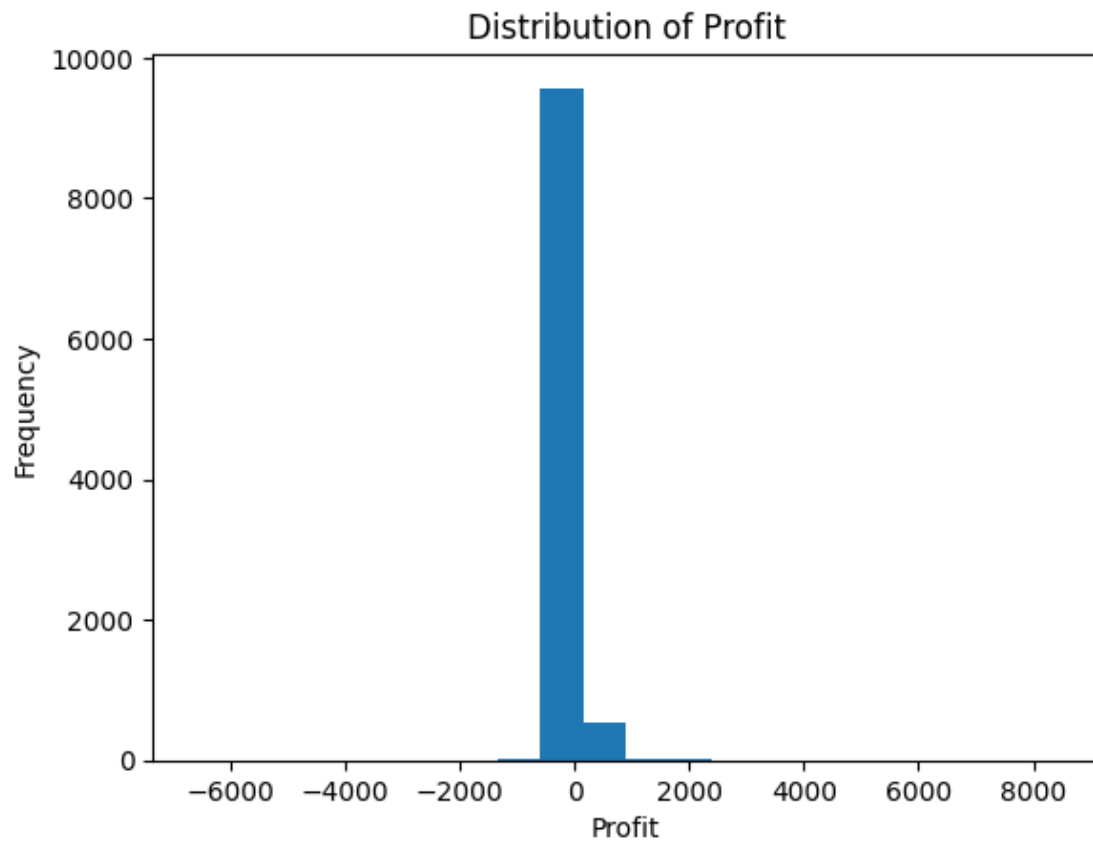
# Group sales by category and sum
category_sales = store.groupby('Category')['Sales'].sum()

# Create a bar chart
category_sales.plot(kind='bar')
plt.title('Total Sales by Category')
plt.xlabel('Category')
plt.ylabel('Sales')
plt.show()
```

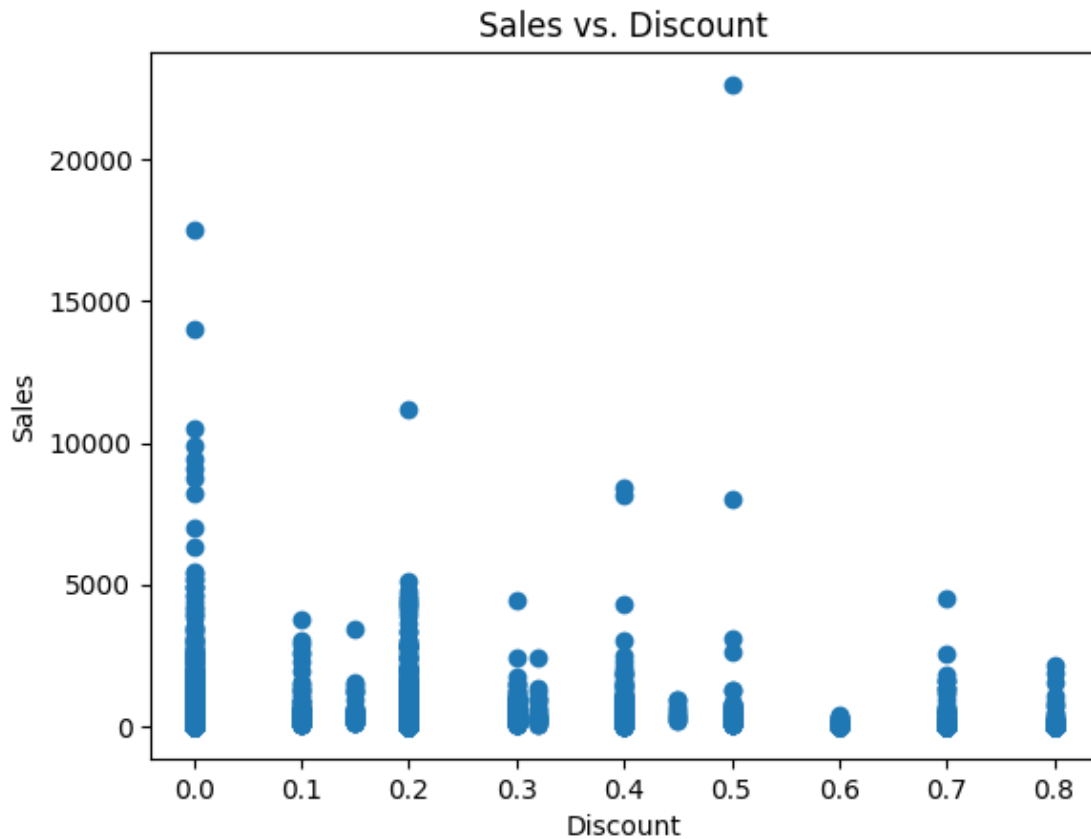


```
[50]: # 2. Histogram of profit

# Create a histogram of profit
plt.hist(store['Profit'], bins=20)
plt.title('Distribution of Profit')
plt.xlabel('Profit')
plt.ylabel('Frequency')
plt.show()
```



```
[51]: # 3. Scatter plot of sales vs discount
      # Create a scatter plot
      plt.scatter(store['Discount'], store['Sales'])
      plt.title('Sales vs. Discount')
      plt.xlabel('Discount')
      plt.ylabel('Sales')
      plt.show()
```



```
[52]: # 4. Pie chart of sales by region
# Group sales by region and sum
region_sales = store.groupby('Region')['Sales'].sum()

# Create a pie chart
region_sales.plot(kind='pie', autopct='%1.1f%%')
plt.title('Sales Distribution by Region')
plt.ylabel('') # Remove the default ylabel
plt.show()
```

Sales Distribution by Region

