Retail_orders

June 5, 2024

```python
[1]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```python
[2]: df = pd.read_csv('orders.csv',na_values=['Not Available','unknown'])
     df.head()
```

```
[2]:    Order Id  Order Date         Ship Mode     Segment          Country  \
     0         1  2023-03-01     Second Class     Consumer  United States
     1         2  2023-08-15     Second Class     Consumer  United States
     2         3  2023-01-10     Second Class    Corporate  United States
     3         4  2022-06-18   Standard Class     Consumer  United States
     4         5  2022-07-13   Standard Class     Consumer  United States

                  City        State  Postal Code Region         Category  \
     0       Henderson     Kentucky        42420  South        Furniture
     1       Henderson     Kentucky        42420  South        Furniture
     2     Los Angeles   California        90036   West  Office Supplies
     3  Fort Lauderdale      Florida        33311  South        Furniture
     4  Fort Lauderdale      Florida        33311  South  Office Supplies

       Sub Category       Product Id  cost price  List Price  Quantity  \
     0    Bookcases  FUR-BO-10001798         240         260         2
     1       Chairs  FUR-CH-10000454         600         730         3
     2       Labels  OFF-LA-10000240          10          10         2
     3       Tables  FUR-TA-10000577         780         960         5
     4      Storage  OFF-ST-10000760          20          20         2

        Discount Percent
     0                 2
     1                 3
     2                 5
     3                 2
     4                 5
```

```python
[3]: df.shape
```

```
[3]: (9994, 16)
```

```
[4]: df.dtypes
```

```
[4]: Order Id           int64
     Order Date         object
     Ship Mode          object
     Segment            object
     Country            object
     City               object
     State              object
     Postal Code        int64
     Region             object
     Category           object
     Sub Category       object
     Product Id         object
     cost price         int64
     List Price         int64
     Quantity           int64
     Discount Percent   int64
     dtype: object
```

```
[5]: df['Ship Mode'].unique()
```

```
[5]: array(['Second Class', 'Standard Class', nan, 'First Class', 'Same Day'],
           dtype=object)
```

```
[6]: df.columns=df.columns.str.lower()
     df.columns=df.columns.str.replace(' ','_')
     df.head(5)
```

```
[6]:    order_id  order_date        ship_mode      segment          country  \
     0         1  2023-03-01    Second Class     Consumer    United States
     1         2  2023-08-15    Second Class     Consumer    United States
     2         3  2023-01-10    Second Class    Corporate    United States
     3         4  2022-06-18  Standard Class     Consumer    United States
     4         5  2022-07-13  Standard Class     Consumer    United States

                  city        state  postal_code region          category  \
     0        Henderson     Kentucky        42420  South          Furniture
     1        Henderson     Kentucky        42420  South          Furniture
     2      Los Angeles   California        90036   West    Office Supplies
     3  Fort Lauderdale      Florida        33311  South          Furniture
     4  Fort Lauderdale      Florida        33311  South    Office Supplies

       sub_category       product_id  cost_price  list_price  quantity  \
     0    Bookcases  FUR-BO-10001798         240         260         2
     1       Chairs  FUR-CH-10000454         600         730         3
```

```
2      Labels  OFF-LA-10000240              10          10          2
3      Tables  FUR-TA-10000577             780         960          5
4     Storage  OFF-ST-10000760              20          20          2

   discount_percent
0                 2
1                 3
2                 5
3                 2
4                 5
```

[7]:
```python
df['order_date']=pd.to_datetime(df['order_date'],format="%Y-%m-%d")
df.dtypes
```

[7]:
```
order_id                     int64
order_date          datetime64[ns]
ship_mode                   object
segment                     object
country                     object
city                        object
state                       object
postal_code                  int64
region                      object
category                    object
sub_category                object
product_id                  object
cost_price                   int64
list_price                   int64
quantity                     int64
discount_percent             int64
dtype: object
```

[ ]:

[8]:
```python
df.head(2)
```

[8]:
```
   order_id order_date      ship_mode   segment        country       city  \
0         1 2023-03-01  Second Class  Consumer  United States  Henderson
1         2 2023-08-15  Second Class  Consumer  United States  Henderson

      state  postal_code region    category sub_category    product_id  \
0  Kentucky        42420  South   Furniture    Bookcases  FUR-BO-10001798
1  Kentucky        42420  South   Furniture       Chairs  FUR-CH-10000454

   cost_price  list_price  quantity  discount_percent
0         240         260         2                 2
1         600         730         3                 3
```

```
[9]: df.to_csv("retail_orders.csv",index = False)
```

```
[10]: df.columns
```

```
[10]: Index(['order_id', 'order_date', 'ship_mode', 'segment', 'country', 'city',
             'state', 'postal_code', 'region', 'category', 'sub_category',
             'product_id', 'cost_price', 'list_price', 'quantity',
             'discount_percent'],
           dtype='object')
```

```
[12]: df.to_csv("retail_orders_data.csv",index = False)
```