

```
In [127]: # Data Manipulation
import pandas as pd

# Data Visualisation
import matplotlib.pyplot as plt
%matplotlib inline

import seaborn as sns
sns.set(color_codes=True)

In [61]: df = pd.read_csv('superstore_salessss.csv')
df

Out[61]:
   order_id  order_date  ship_date  ship_mode  customer_name  segment  state  country  market  region  ...  category  sub_category  product_name  sales  quantity  discount  profit  shipping_cost  order_pr
0      AG-2011-01-01-2040  01-01-2011  06-01-2011  Standard Class  Toby Braunhardt  Consumer  Constantine  Algeria  Africa  Africa  ...  Office Supplies  Storage  Tenex Lockers, Blue  408  2  0.0  106.1400  35.46  M
1      IN-2011-01-01-47883  01-01-2011  08-01-2011  Standard Class  Joseph Holt  Consumer  New South Wales  Australia  APAC  Oceania  ...  Office Supplies  Supplies  Acme Trimmer, High Speed  120  3  0.1  36.0360  9.72  M
2      HU-2011-01-01-1220  01-01-2011  05-01-2011  Second Class  Annie Thurman  Consumer  Budapest  Hungary  EMEA  EMEA  ...  Office Supplies  Storage  Tenex Box, Single Width  66  4  0.0  29.6400  8.17
3      IT-2011-01-01-3647632  01-01-2011  05-01-2011  Second Class  Eugene Moren  Home Office  Stockholm  Sweden  EU  North  ...  Office Supplies  Paper  Enermax Note Cards, Premium  45  3  0.5  -26.0550  4.82
4      IN-2011-01-01-47883  01-01-2011  08-01-2011  Standard Class  Joseph Holt  Consumer  New South Wales  Australia  APAC  Oceania  ...  Furniture  Furnishings  Eldon Light Bulb, Duo Pack  114  5  0.1  37.7700  4.70  M
...
51285  CA-2014-31-12-115427  31-12-2014  04-01-2015  Standard Class  Erica Bern  Corporate  California  United States  US  West  ...  Office Supplies  Binders  Cardinal Saint-D Ring Binder, Heavy Gauge Vinyl  14  2  0.2  4.5188  0.89  M
51286  MO-2014-31-12-2960  31-12-2014  05-01-2015  Standard Class  Liz Preis  Consumer  Souss-Massa-Draâ  Morocco  Africa  Africa  ...  Office Supplies  Binders  Wilson Jones Hole Reinforcements, Clear  4  1  0.0  0.4200  0.49  M
51287  MX-2014-31-12-110527  31-12-2014  02-01-2015  Second Class  Charlotte Melton  Consumer  Managua  Nicaragua  LATAM  Central  ...  Office Supplies  Labels  Hon Color Coded Labels, 5000 Label Set  26  3  0.0  12.3600  0.35  M
51288  MX-2014-31-12-114783  31-12-2014  06-01-2015  Standard Class  Tamara Dahlen  Consumer  Chihuahua  Mexico  LATAM  North  ...  Office Supplies  Labels  Hon Legal Exhibit Labels, Alphabetical  7  1  0.0  0.5600  0.20  M
51289  CA-2014-31-12-156720  31-12-2014  04-01-2015  Standard Class  Jill Matthias  Consumer  Colorado  United States  US  West  ...  Office Supplies  Fasteners  Bagged Rubber Bands  3  3  0.2  -0.6048  0.17  M

51290 rows x 21 columns

In [62]: # First ten rows of the dataset
df.head(10)

Out[62]:
   order_id  order_date  ship_date  ship_mode  customer_name  segment  state  country  market  region  ...  category  sub_category  product_name  sales  quantity  discount  profit  shipping_cost  order_priority
0      AG-2011-01-01-2040  01-01-2011  06-01-2011  Standard Class  Toby Braunhardt  Consumer  Constantine  Algeria  Africa  Africa  ...  Office Supplies  Storage  Tenex Lockers, Blue  408  2  0.0  106.140  35.46  Medium
1      IN-2011-01-01-47883  01-01-2011  08-01-2011  Standard Class  Joseph Holt  Consumer  New South Wales  Australia  APAC  Oceania  ...  Office Supplies  Supplies  Acme Trimmer, High Speed  120  3  0.1  36.036  9.72  Medium
2      HU-2011-01-01-1220  01-01-2011  05-01-2011  Second Class  Annie Thurman  Consumer  Budapest  Hungary  EMEA  EMEA  ...  Office Supplies  Storage  Tenex Box, Single Width  66  4  0.0  29.640  8.17  High
3      IT-2011-01-01-3647632  01-01-2011  05-01-2011  Second Class  Eugene Moren  Home Office  Stockholm  Sweden  EU  North  ...  Office Supplies  Paper  Enermax Note Cards, Premium  45  3  0.5  -26.055  4.82  High
4      IN-2011-01-01-47883  01-01-2011  08-01-2011  Standard Class  Joseph Holt  Consumer  New South Wales  Australia  APAC  Oceania  ...  Furniture  Furnishings  Eldon Light Bulb, Duo Pack  114  5  0.1  37.770  4.70  Medium
5      IN-2011-01-01-47883  01-01-2011  09-01-2011  Standard Class  Joseph Holt  Consumer  New South Wales  Australia  APAC  Oceania  ...  Office Supplies  Paper  Eaton Computer Printout Paper, 8.5 x 11  55  2  0.1  15.342  1.80  Medium
6      CA-2011-02-01-1510  02-01-2011  06-01-2011  Standard Class  Magdelene Morse  Consumer  Ontario  Canada  Canada  Canada  ...  Technology  Machines  Okidata Inkjet, Wireless  314  1  0.0  3.120  24.10  Medium
7      IN-2011-03-01-79397  03-01-2011  03-01-2011  Same Day  Kean Nguyen  Corporate  New South Wales  Australia  APAC  Oceania  ...  Office Supplies  Appliances  Hoover Microwave, White  276  1  0.1  110.412  125.32  Critical
8      ID-2011-03-01-80230  03-01-2011  09-01-2011  Standard Class  Ken Lonsdale  Consumer  Auckland  New Zealand  APAC  Oceania  ...  Technology  Copiers  Hewlett Wireless Fax, Laser  912  4  0.4  -319.464  107.10  Low
9      IZ-2011-03-01-4680  03-01-2011  07-01-2011  Standard Class  Lindsay Williams  Corporate  Ninawa  Iraq  EMEA  EMEA  ...  Furniture  Chairs  Novimex Swivel Stool, Set of Two  667  4  0.0  253.320  81.26  High

10 rows x 21 columns

In [63]: # Last five rows of the dataset
df.tail()

Out[63]:
   order_id  order_date  ship_date  ship_mode  customer_name  segment  state  country  market  region  ...  category  sub_category  product_name  sales  quantity  discount  profit  shipping_cost  order_priority
51285  CA-2014-31-12-115427  31-12-2014  04-01-2015  Standard Class  Erica Bern  Corporate  California  United States  US  West  ...  Office Supplies  Binders  Cardinal Saint-D Ring Binder, Heavy Gauge Vinyl  14  2  0.2  4.5188  0.89  Medium
51286  MO-2014-31-12-2960  31-12-2014  05-01-2015  Standard Class  Liz Preis  Consumer  Souss-Massa-Draâ  Morocco  Africa  Africa  ...  Office Supplies  Binders  Wilson Jones Hole Reinforcements, Clear  4  1  0.0  0.4200  0.49  Medium
51287  MX-2014-31-12-110527  31-12-2014  02-01-2015  Second Class  Charlotte Melton  Consumer  Managua  Nicaragua  LATAM  Central  ...  Office Supplies  Labels  Hon Color Coded Labels, 5000 Label Set  26  3  0.0  12.3600  0.35  Medium
51288  MX-2014-31-12-114783  31-12-2014  06-01-2015  Standard Class  Tamara Dahlen  Consumer  Chihuahua  Mexico  LATAM  North  ...  Office Supplies  Labels  Hon Legal Exhibit Labels, Alphabetical  7  1  0.0  0.5600  0.20  Medium
51289  CA-2014-31-12-156720  31-12-2014  04-01-2015  Standard Class  Jill Matthias  Consumer  Colorado  United States  US  West  ...  Office Supplies  Fasteners  Bagged Rubber Bands  3  3  0.2  -0.6048  0.17  Medium

5 rows x 21 columns

In [64]: # Columns present in the dataset
df.columns

Out[64]:
Index(['order_id', 'order_date', 'ship_date', 'ship_mode', 'customer_name', 'segment', 'state', 'country', 'market', 'region', 'category', 'sub_category', 'product_name', 'sales', 'quantity', 'discount', 'profit', 'shipping_cost', 'order_priority', 'year'],
      dtype='object')

In [65]: # A concise summary of the dataset
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51290 entries, 0 to 51289
Data columns (total 21 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   order_id    51290 non-null  object
1   order_date  51290 non-null  object
2   ship_date   51290 non-null  object
3   ship_mode   51290 non-null  object
4   customer_name  51290 non-null  object
5   segment     51290 non-null  object
6   state       51290 non-null  object
7   country     51290 non-null  object
8   market      51290 non-null  object
9   region      51290 non-null  object
10  product_id  51290 non-null  object
11  category    51290 non-null  object
12  sub_category  51290 non-null  object
13  product_name  51290 non-null  object
14  sales       51290 non-null  object
15  quantity    51290 non-null  int64
16  discount    51290 non-null  float64
17  profit      51290 non-null  float64
18  shipping_cost  51290 non-null  float64
19  order_priority  51290 non-null  object
20  year        51290 non-null  int64
dtypes: float64(3), int64(2), object(16)
memory usage: 8.2+ MB

In [66]: # Checking missing values
df.isnull().sum()

Out[66]:
order_id      0
order_date    0
ship_date     0
ship_mode     0
customer_name  0
segment       0
state         0
country       0
market        0
region        0
product_id    0
category      0
sub_category  0
product_name  0
sales         0
quantity      0
discount      0
profit        0
shipping_cost  0
order_priority 0
year          0
dtype: int64

In [67]: # Generating descriptive statistics summary
df.describe().round()

Out[67]:
   quantity  discount    profit  shipping_cost    year
count  51290.0  51290.0  51290.0         51290.0  51290.0
mean     3.00    0.0    29.0          26.0    2013.0
std      2.0    0.0   174.0           57.0     1.0
min      1.0    0.0 -6600.0           0.0   2011.0
25%      2.0    0.0    0.0           3.0   2012.0
50%      3.0    0.0    9.0           8.0   2013.0
75%      5.0    0.0   37.0          24.0   2014.0
max     14.0    1.0  8400.0          934.0   2014.0

In [109]: df.dtypes

Out[109]:
order_id      object
order_date    datetime64[ns]
ship_date     object
ship_mode     object
customer_name  object
segment       object
state         object
country       object
market        object
region        object
product_id    object
category      object
sub_category  object
product_name  object
sales         object
quantity      int64
discount      float64
profit        float64
shipping_cost  float64
order_priority object
year          int64
month_year    object
dtype: object

In [109]: ## TO CHECK PROFIT TREND

In [104]: df['order_date'].min()

Out[104]:
Timestamp('2011-01-01 00:00:00')

In [72]: df['order_date'].max()

Out[72]:
Timestamp('2014-12-31 00:00:00')

In [73]: # Getting month year from order_date
df['month_year'] = df['order_date'].apply(lambda x: x.strftime('%Y-%m'))

In [74]: df['month_year']

Out[74]:
0      2011-01
1      2011-01
2      2011-01
3      2011-01
4      2011-01
...
51285  2014-12
51286  2014-12
51287  2014-12
51288  2014-12
51289  2014-12
Name: month_year, Length: 51290, dtype: object

In [110]: # grouping month_year by sales
df_temp = df.groupby('month_year').sum()[['profit']].reset_index()

In [111]: df_temp

Out[111]:
   month_year    profit
0      2011-01  13457.23302
1      2011-02  17588.83726
2      2011-03  16169.36062
3      2011-04  13405.46924
4      2011-05  14777.45792
5      2011-06  25932.87796
6      2011-07  10631.84406
7      2011-08  19650.67124
8      2011-09  32313.25458
9      2011-10  30745.54166
10     2011-11  21261.40536
11     2011-12  33006.85862
12     2012-01  19627.42058
13     2012-02  17828.18244
14     2012-03  22752.87684
15     2012-04  20804.00532
16     2012-05  22865.99572
17     2012-06  34358.03962
18     2012-07  21725.18808
19     2012-08  36511.44996
20     2012-09  25039.93498
21     2012-10  27773.87454
22     2012-11  26160.60696
23     2012-12  31968.30416
24     2013-01  29052.40252
25     2013-02  31553.99756
26     2013-03  34873.71938
27     2013-04  26008.05144
28     2013-05  40630.83946
29     2013-06  43298.85000
30     2013-07  27019.91294
31     2013-08  32977.75576
32     2013-09  18850.09084
33     2013-10  27871.97470
34     2013-11  51720.88568
35     2013-12  47654.27990
36     2014-01  31447.74660
37     2014-02  29454.01876
38     2014-03  50097.04458
39     2014-04  36709.53320
40     2014-05  34335.60404
41     2014-06  40869.02108
42     2014-07  26450.70856
43     2014-08  46443.47934
44     2014-09  52533.83284
45     2014-10  52342.49740
46     2014-11  55561.72406
47     2014-12  48920.76000

In [113]: # Setting the figure size
plt.figure(figsize=(16, 5))
plt.plot(df_temp['month_year'], df_temp['profit'], color='red')
plt.xticks(rotation='vertical', size=8)
plt.show()



In [114]: ## TO FIND TOP 10 PROFITABLE PRODUCTS

In [117]: # Grouping products by PROFIT
prod_profit = pd.DataFrame(df.groupby('product_name').sum()[['profit']])

# Sorting the dataframe in descending order
prod_profit.sort_values(by=['profit'], inplace=True, ascending=False)

# Top 10 products by sales
prod_profit[:10]

Out[117]:
   product_name  profit
0  Canon imageCLASS 2200 Advanced Copier  25199.9280
1  Cisco Smart Phone, Full Size  17238.5206
2  Motorola Smart Phone, Full Size  17027.1130
3  Hoover Stove, Red  11807.9690
4  Sauder Classic Bookcase, Traditional  10672.0730
5  Harbour Creations Executive Leather Armchair, Adjustable  10427.3260
6  Nokia Smart Phone, Full Size  9938.1955
7  Cisco Smart Phone, with Caller ID  9786.6408
8  Nokia Smart Phone, with Caller ID  9465.3257
9  Belkin Router, USB  8955.0180

In [118]: ## ANALYZING MOST SELLING PRODUCTS

In [119]: # Grouping products by quantity
best_selling_prods = pd.DataFrame(df.groupby('product_name').sum()[['quantity']])

# Sorting the dataframe in descending order
best_selling_prods.sort_values(by=['quantity'], inplace=True, ascending=False)

# Most selling products
best_selling_prods[:10]

Out[119]:
   product_name  quantity
0  Staples  876
1  Cardinal Index Tab, Clear  337
2  Eldon File Cart, Single Width  321
3  Rogers File Cart, Single Width  262
4  Sanford Pencil Sharpener, Water Color  259
5  Stockwell Paper Clips, Assorted Sizes  253
6  Avery Index Tab, Clear  252
7  Ibico Index Tab, Clear  251
8  Smead File Cart, Single Width  250
9  Stanley Pencil Sharpener, Water Color  242

In [120]: ## MOST PREFERRED SHIP MODE

In [121]: # Setting the figure size
plt.figure(figsize=(10, 8))

# countplot: Show the counts of observations in each categorical bin using bars
sns.countplot(x='ship_mode', data=df)

# Display the figure
plt.show()



In [122]: ## MOST PROFITABLE CATEGORY AND SUB-CATEGORY

In [123]: # Grouping products by Category and Sub-Category
cat_subcat = pd.DataFrame(df.groupby(['category', 'sub_category']).sum()[['profit']])

# Sorting the values
cat_subcat.sort_values(['category', 'profit'], ascending=False)

Out[123]:
   category  sub_category  profit
0  Technology  Copiers  258567.54818
1  Technology  Phones  216717.00580
2  Technology  Accessories  129626.30620
3  Technology  Machines  58867.87300
4  Office Supplies  Appliances  141680.58940
5  Office Supplies  Storage  108461.48980
6  Office Supplies  Binders  72449.84600
7  Office Supplies  Paper  59297.68270
8  Office Supplies  Art  57953.91090
9  Office Supplies  Envelopes  29601.11630
10 Office Supplies  Supplies  22583.26310
11 Office Supplies  Labels  15010.51200
12 Office Supplies  Fasteners  11525.42410
13 Office Supplies  Bookcases  161924.41950
14 Office Supplies  Chairs  141973.79750
15 Office Supplies  Furnishings  46967.42550
16 Office Supplies  Tables  -64083.38870

In [ ]: ## THE END
## THANK YOU
```