CASE STUDY

ONLINE STORE ANALYSIS

INTRODUCTION

To understand the importance of visual resources in marketing strategies. For that, you have to determine the type of data that needs to be analyzed daily. Putting all this information in a simple visual format will help you to gain quicker insights and become a data-driven business.

Data visualization makes it possible by using shapes, charts, and information distributed in the visual space. The most important factor is the power of just taking a quick look and being able to understand many points of interest.

It shows trends, patterns, and correlations in your business. Besides, it helps managers and decision-makers to make data-driven choices. For instance, you may decide which products you'll need to have available in which seasons just by using data visualization in your supply chain strategy.

Or which media you may use to reach your audience efficiently by pulling out data from your previous campaigns.

IMPORTING REQUIRED LIBRARIES:

```python
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

LOADING THE DATASET:

```python
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
print(df)
```

```
              order_id  order_date ship_date           ship_mode  \
0          AG-2011-2040    1/1/2011  1/6/2011      Standard Class
1         IN-2011-47883    1/1/2011  1/8/2011      Standard Class
2           HU-2011-1220    1/1/2011  1/5/2011        Second Class
3        IT-2011-3647632    1/1/2011  1/5/2011        Second Class
4         IN-2011-47883    1/1/2011  1/8/2011      Standard Class
...                ...         ...       ...                 ...
51285    CA-2014-115427  12/31/2014  1/4/2015      Standard Class
51286      MO-2014-2560  12/31/2014  1/5/2015      Standard Class
51287    MX-2014-110527  12/31/2014  1/2/2015        Second Class
51288    MX-2014-114783  12/31/2014  1/6/2015      Standard Class
51289    CA-2014-156720  12/31/2014  1/4/2015      Standard Class

         customer_name      segment              state        country market
0        Toby Braunhardt     Consumer        Constantine        Algeria Africa
1           Joseph Holt     Consumer    New South Wales      Australia   APAC
2         Annie Thurman     Consumer           Budapest        Hungary   EMEA
3         Eugene Moren  Home Office          Stockholm         Sweden     EU
4           Joseph Holt     Consumer    New South Wales      Australia   APAC
...                ...          ...                ...            ...    ...
51285       Erica Bern    Corporate         California  United States     US
51286        Liz Preis     Consumer  Souss-Massa-Draâ        Morocco Africa
51287  Charlotte Melton     Consumer           Managua      Nicaragua  LATAM
51288     Tamara Dahlen     Consumer          Chihuahua         Mexico  LATAM
51289     Jill Matthias     Consumer           Colorado  United States     US

         region  ...          category sub_category  \
0        Africa  ...  Office Supplies        Storage
1       Oceania  ...  Office Supplies       Supplies
2          EMEA  ...  Office Supplies        Storage
3         North  ...  Office Supplies          Paper
4       Oceania  ...       Furniture  Furnishings
...         ...  ...              ...          ...
```

```
51285      West   ...  Office Supplies      Binders
51286    Africa   ...  Office Supplies      Binders
51287   Central   ...  Office Supplies       Labels
51288     North   ...  Office Supplies       Labels
51289      West   ...  Office Supplies    Fasteners
```

```
                                     product_name sales quantity  \
0                               Tenex Lockers, Blue   408        2
1                         Acme Trimmer, High Speed   120        3
2                          Tenex Box, Single Width    66        4
3                      Enermax Note Cards, Premium    45        3
4                        Eldon Light Bulb, Duo Pack   114        5
...                                            ...   ...      ...
51285  Cardinal Slant-D Ring Binder, Heavy Gauge Vinyl    14        2
51286         Wilson Jones Hole Reinforcements, Clear     4        1
51287          Hon Color Coded Labels, 5000 Label Set    26        3
51288          Hon Legal Exhibit Labels, Alphabetical     7        1
51289                          Bagged Rubber Bands     3        3
```

```
       discount    profit  shipping_cost  order_priority  year
0           0.0  106.1400          35.46          Medium  2011
1           0.1   36.0360           9.72          Medium  2011
2           0.0   29.6400           8.17            High  2011
3           0.5  -26.0550           4.82            High  2011
4           0.1   37.7700           4.70          Medium  2011
...         ...       ...            ...             ...   ...
51285       0.2    4.5188           0.89          Medium  2014
51286       0.0    0.4200           0.49          Medium  2014
51287       0.0   12.3600           0.35          Medium  2014
51288       0.0    0.5600           0.20          Medium  2014
51289       0.2   -0.6048           0.17          Medium  2014

[51290 rows x 21 columns]
```

PROCESSING THE DATASET :

```python
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
df.shape
```

(51290, 21)

```python
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
df.columns
```

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

```python
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
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
```

```python
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
df.isnull().sum()
```

```
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
```

```
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
df.describe().round()
```

	quantity	discount	profit	shipping_cost	year
count	51290.0	51290.0	51290.0	51290.0	51290.0
mean	3.0	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

EXPLORATORY DATA ANALYSIS :

1.DISPLAY CATEGORY WISE SALES DISTRIBUTION.

```python
import matplotlib.pyplot as plt
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
sns.barplot(x='category',y='sales',data=df)
plt.xticks(rotation='vertical', size=10)
plt.show()
```
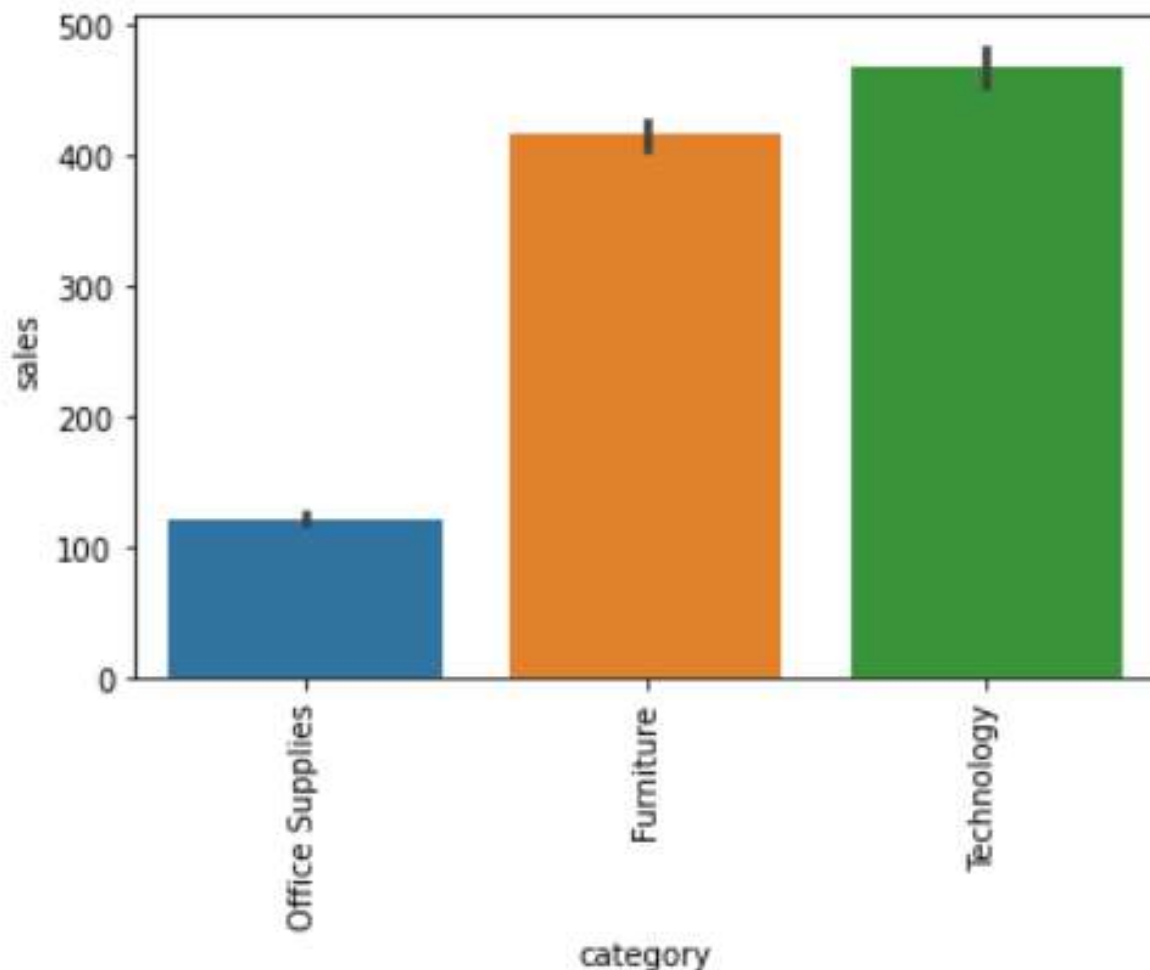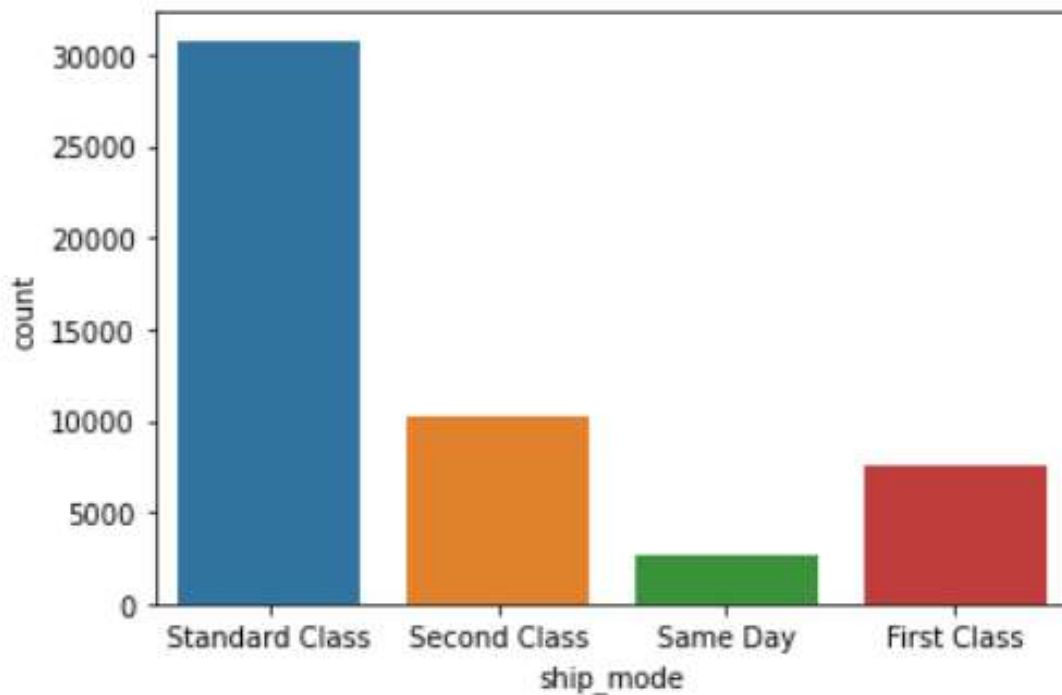

According to the above bar plot , Technology category has most of the sales whereas Office Supplies has the least sales.

2.DISPLAY THE MOST PREFERRED SHIP MODE.

```python
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
sns.countplot(x='ship_mode', data=df)
```

```
<AxesSubplot:xlabel='ship_mode', ylabel='count'>
```
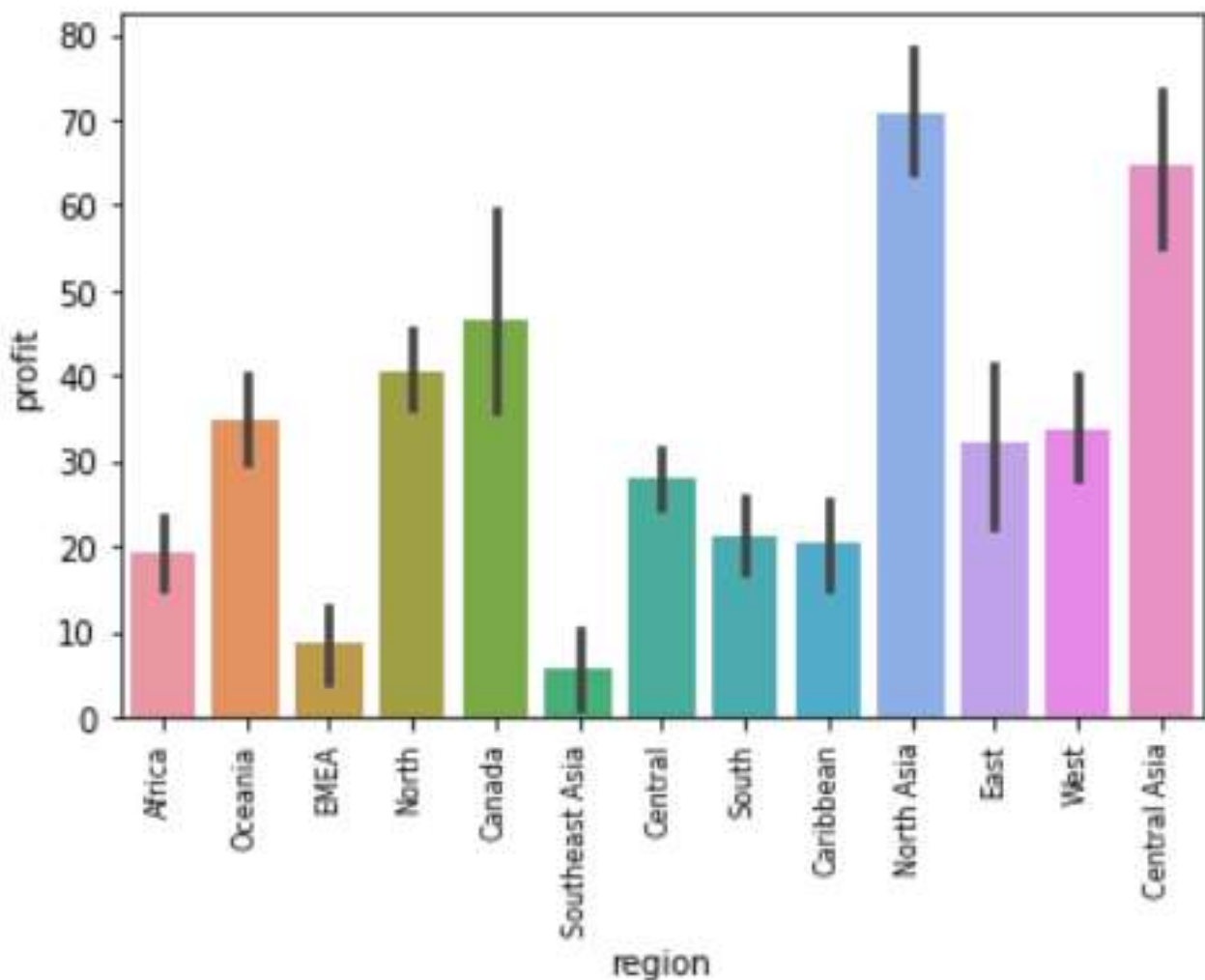

According to the above count plot , the most preferred shipping mode is Standard class .

3. DISPLAY REGION WISE PROFIT DISTRIBUTION .

```python
import matplotlib.pyplot as plt
import seaborn as sns
# Importing dataset
df = sns.load_dataset('superstore_sales')
sns.barplot(x='region',y='profit',data=df)
plt.xticks(rotation='vertical', size=8)
plt.show()
```


According to the above bar plot , the most of the profit is earned in the North Asian region whereas the least profit is earned in the Southeast Asian region.