Project Title - Superstore orders Dataset Analysis

The dataset contains 51290 rows and 21 columns that will be used in analysing and creating visualisation . The dataset is gotten from kaggle.com/datasets, the dataset is used for analysis and visualisation using pandas, matplotlib and seaborn.

The dataset includes data for sales of multiple products sold by the store with other information related to geography, product categories, subcategories, sales and profits. The course zero to pandas has been so helpful and impactful. I have learnt alot in programing from using the if statement, dictionary, lists, numpy for numerical large data set ,pandas for converting data into tabular form, data cleaning, retrieving of information from a data frame and matplotlib seaborn for ploting of different graphs like bar charts, histogram, heatmap and so on.

Downloading the Dataset

The dataset is downloaded from kaggle using the opendatasets in csv format.

In [128]:
```
pip install jovian opendatasets
```

Requirement already satisfied: jovian in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (0.2.45)
Requirement already satisfied: opendatasets in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (0.1.22)
Requirement already satisfied: requests in c:\users\adesoji awobajo\appdata\loc
al\programs\python\python311\lib\site-packages (from jovian) (2.28.1)
Requirement already satisfied: uuid in c:\users\adesoji awobajo\appdata\local\p
rograms\python\python311\lib\site-packages (from jovian) (1.30)
Requirement already satisfied: pyyaml in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (from jovian) (6.0)
Requirement already satisfied: click in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (from jovian) (8.1.3)
Requirement already satisfied: tqdm in c:\users\adesoji awobajo\appdata\local\p
rograms\python\python311\lib\site-packages (from opendatasets) (4.64.1)
Requirement already satisfied: kaggle in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (from opendatasets) (1.5.12)
Requirement already satisfied: colorama in c:\users\adesoji awobajo\appdata\loc
al\programs\python\python311\lib\site-packages (from click->jovian) (0.4.6)
Requirement already satisfied: six>=1.10 in c:\users\adesoji awobajo\appdata\lo
cal\programs\python\python311\lib\site-packages (from kaggle->opendatasets) (1.
16.0)
Requirement already satisfied: certifi in c:\users\adesoji awobajo\appdata\loca
l\programs\python\python311\lib\site-packages (from kaggle->opendatasets) (202
2.9.24)
Requirement already satisfied: python-dateutil in c:\users\adesoji awobajo\appd
ata\local\programs\python\python311\lib\site-packages (from kaggle->opendataset
s) (2.8.2)
Requirement already satisfied: python-slugify in c:\users\adesoji awobajo\appda
ta\local\programs\python\python311\lib\site-packages (from kaggle->opendataset
s) (7.0.0)
Requirement already satisfied: urllib3 in c:\users\adesoji awobajo\appdata\loca
l\programs\python\python311\lib\site-packages (from kaggle->opendatasets) (1.2
6.13)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\adesoji awo
bajo\appdata\local\programs\python\python311\lib\site-packages (from requests->
jovian) (2.1.1)
Requirement already satisfied: idna<4,>=2.5 in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (from requests->jovian) (3.
4)
Requirement already satisfied: text-unidecode>=1.3 in c:\users\adesoji awobajo
\appdata\local\programs\python\python311\lib\site-packages (from python-slugify
->kaggle->opendatasets) (1.3)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip available: 22.3 -> 22.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip

In [129]:
```
dataset_url ='https://www.kaggle.com/laibaanwer/superstore-sales-dataset'
```

In [130]:
```python
import opendatasets as od
od.download(dataset_url)
```

Skipping, found downloaded files in ".\superstore-sales-dataset" (use force=Tru
e to force download)

In [131]:
```python
data_dir='./superstore-sales-dataset'
```

In [132]:
```python
import os
os.listdir(data_dir)
```

Out[132]: ['.ipynb_checkpoints']

In [133]:
```python
project_name= 'zerotopandas-course-project-starter'
```

In [134]:
```python
pip install jovian
```

Requirement already satisfied: jovian in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (0.2.45)
Requirement already satisfied: requests in c:\users\adesoji awobajo\appdata\loc
al\programs\python\python311\lib\site-packages (from jovian) (2.28.1)
Requirement already satisfied: uuid in c:\users\adesoji awobajo\appdata\local\p
rograms\python\python311\lib\site-packages (from jovian) (1.30)
Requirement already satisfied: pyyaml in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (from jovian) (6.0)
Requirement already satisfied: click in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (from jovian) (8.1.3)
Requirement already satisfied: colorama in c:\users\adesoji awobajo\appdata\loc
al\programs\python\python311\lib\site-packages (from click->jovian) (0.4.6)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\adesoji awo
bajo\appdata\local\programs\python\python311\lib\site-packages (from requests->
jovian) (2.1.1)
Requirement already satisfied: idna<4,>=2.5 in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (from requests->jovian) (3.
4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\adesoji awobaj
o\appdata\local\programs\python\python311\lib\site-packages (from requests->jov
ian) (1.26.13)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\adesoji awobajo\a
ppdata\local\programs\python\python311\lib\site-packages (from requests->jovia
n) (2022.9.24)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip available: 22.3 -> 22.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip

In [136]:
```
jovian.commit(project=project_name)
```

<IPython.core.display.Javascript object>

[jovian] Updating notebook "adetounawobajo026/zerotopandas-course-project-start
er" on https://jovian.ai/ (https://jovian.ai/)
[jovian] Committed successfully! https://jovian.ai/adetounawobajo026/zerotopand
as-course-project-starter (https://jovian.ai/adetounawobajo026/zerotopandas-cou
rse-project-starter)

Out[136]: 'https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter'

In [ ]:

Data Preparation and Cleaning

Data preparation and Cleaning is the process of dealing with messy, disordered data and
eliminating incorrect, missing, duplicated values in your dataset.it improves the accuracy of the
dataset for analysis in putting in right shape and quality.

In [137]:
```
pip install pandas
```

Requirement already satisfied: pandas in c:\users\adesoji awobajo\appdata\local
\programs\python\python311\lib\site-packages (1.5.2)Note: you may need to resta
rt the kernel to use updated packages.

[notice] A new release of pip available: 22.3 -> 22.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip

Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\adesoji awoba
jo\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2.
8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (from pandas) (2022.6)
Requirement already satisfied: numpy>=1.21.0 in c:\users\adesoji awobajo\appdat
a\local\programs\python\python311\lib\site-packages (from pandas) (1.23.5)
Requirement already satisfied: six>=1.5 in c:\users\adesoji awobajo\appdata\loc
al\programs\python\python311\lib\site-packages (from python-dateutil>=2.8.1->pa
ndas) (1.16.0)

In [138]:
```
import pandas as pd
```

In [139]:
```
superstoreorders_df =pd.read_csv('SuperStoreOrders.csv')
```

In [140]: `superstoreorders_df`

Out[140]:

	order_id	order_date	ship_date	ship_mode	customer_name	segment	state	cour
0	AG-2011-2040	1/1/2011	6/1/2011	Standard Class	Toby Braunhardt	Consumer	Constantine	Alge
1	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	Joseph Holt	Consumer	New South Wales	Austr
2	HU-2011-1220	1/1/2011	5/1/2011	Second Class	Annie Thurman	Consumer	Budapest	Hung
3	IT-2011-3647632	1/1/2011	5/1/2011	Second Class	Eugene Moren	Home Office	Stockholm	Swe
4	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	Joseph Holt	Consumer	New South Wales	Austr
...	...	...	...	...	...	...	...	
51285	CA-2014-115427	31-12-2014	4/1/2015	Standard Class	Erica Bern	Corporate	California	Uni Sta
51286	MO-2014-2560	31-12-2014	5/1/2015	Standard Class	Liz Preis	Consumer	Souss-Massa-Draâ	Moro
51287	MX-2014-110527	31-12-2014	2/1/2015	Second Class	Charlotte Melton	Consumer	Managua	Nicara
51288	MX-2014-114783	31-12-2014	6/1/2015	Standard Class	Tamara Dahlen	Consumer	Chihuahua	Mex
51289	CA-2014-156720	31-12-2014	4/1/2015	Standard Class	Jill Matthias	Consumer	Colorado	Uni Sta

51290 rows × 21 columns
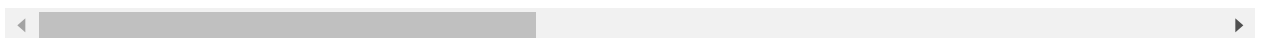
In [141]: `superstoreorders_df.shape`

Out[141]: `(51290, 21)`

In [142]: `superstoreorders_df.describe()`

Out[142]:

	quantity	discount	profit	shipping_cost	year
count	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000
mean	3.476545	0.142908	28.641740	26.375915	2012.777208
std	2.278766	0.212280	174.424113	57.296804	1.098931
min	1.000000	0.000000	-6599.978000	0.000000	2011.000000
25%	2.000000	0.000000	0.000000	2.610000	2012.000000
50%	3.000000	0.000000	9.240000	7.790000	2013.000000
75%	5.000000	0.200000	36.810000	24.450000	2014.000000
max	14.000000	0.850000	8399.976000	933.570000	2014.000000

In [143]: `superstoreorders_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 [144]: `superstoreorders_df.columns`

Out[144]:
```
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')
```

In [145]:
```
superstoreorders_df['order_date']=superstoreorders_df['order_date'].astype('datet
superstoreorders_df['ship_date']=superstoreorders_df['ship_date'].astype('datetin
superstoreorders_df['sales']= pd.to_numeric(superstoreorders_df['sales'],errors='
```

```
C:\Users\Adesoji Awobajo\AppData\Local\Temp\ipykernel_7704\2739567199.py:1: Use
rWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default)
was specified. This may lead to inconsistently parsed dates! Specify a format t
o ensure consistent parsing.
  superstoreorders_df['order_date']=superstoreorders_df['order_date'].astype('d
atetime64[ns]')
C:\Users\Adesoji Awobajo\AppData\Local\Temp\ipykernel_7704\2739567199.py:2: Use
rWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default)
was specified. This may lead to inconsistently parsed dates! Specify a format t
o ensure consistent parsing.
  superstoreorders_df['ship_date']=superstoreorders_df['ship_date'].astype('dat
etime64[ns]')
```

In [146]:
```python
superstoreorders_df.value_counts()
```

Out[146]:
```
order_id        order_date  ship_date   ship_mode       customer_name       seg
ment    state                   country                 market  region
product_id      category        sub_category  product_name
sales   quantity  discount  profit        shipping_cost  order_priority  year
AE-2011-9160    2011-03-10  2011-07-10  Standard Class  Patrick O'Donnell  Con
sumer      'Ajman                   United Arab Emirates  EMEA    EMEA
OFF-FEL-10001405  Office Supplies  Storage      Fellowes File Cart, Industrial
83.0   2        0.700    -157.08600  5.69            Medium          2011    1
MX-2011-132066  2011-07-07  2011-07-13  Standard Class  Charles McCrossin  Con
sumer      Buenos Aires             Argentina             LATAM   South
TEC-CO-10002759  Technology       Copiers      HP Copy Machine, Laser
194.0  2        0.402    -123.94488  13.92           Medium          2011    1
MX-2011-131933  2011-07-12  2011-11-12  Second Class    Trudy Glocke       Con
sumer      Santiago                 Chile                 LATAM   South
OFF-LA-10001800  Office Supplies  Labels       Hon Shipping Labels, Laser Pri
nter Compatible  53.0   7        0.000    3.64000    2.27            Medium
2011    1
MX-2011-131996  2011-03-08  2011-07-08  Standard Class  Craig Molinari     Cor
porate     Holguín                  Cuba                  LATAM   Caribbean
OFF-AR-10002486  Office Supplies  Art          Sanford Pencil Sharpener, Wate
r Color          77.0   4        0.000    22.40000   10.39           High
2011    1

OFF-AR-10002564  Office Supplies  Art          Boston Pens, Easy-Erase
29.0   3        0.000    0.84000    2.54            High            2011    1

..
ES-2014-3977451  2014-05-27  2014-05-31  Standard Class  Jamie Kunitz      Con
sumer      Provence-Alpes-Côte d'Azur  France                EU      Central
FUR-CH-10002610  Furniture        Chairs       Novimex Swivel Stool, Black
437.0  3        0.100    -34.02900  18.15           Medium          2014    1

OFF-LA-10001546  Office Supplies  Labels       Avery Shipping Labels, Adjusta
ble              39.0   4        0.000    12.84000   1.58            Medium
2014    1
                 2014-08-12  2014-09-12  First Class     Steven Roelle     Hom
e Office   Ile-de-France            France                EU      Central
OFF-SU-10002379  Office Supplies  Supplies     Fiskars Letter Opener, High Sp
eed              82.0   3        0.000    3.24000    22.91           High
2014    1

TEC-PH-10004915  Technology       Phones       Nokia Speaker Phone, Cordless
214.0  2        0.150    70.56000   55.31           High            2014    1
ZI-2014-9650    2014-05-21  2014-05-23  Second Class    Henry Goldwyn      Cor
porate     Manicaland               Zimbabwe              Africa  Africa
OFF-AVE-10000418  Office Supplies  Labels      Avery Color Coded Labels, 5000
Label Set        4.0    1        0.700    -4.80600   0.80            High
2014    1
Length: 48660, dtype: int64
```

In [147]:
```python
superstoreorders_df.nunique()
```

Out[147]:
```
order_id          25035
order_date         1430
ship_date          1464
ship_mode             4
customer_name       795
segment               3
state              1094
country             147
market                7
region               13
product_id        10292
category              3
sub_category         17
product_name       3788
sales               993
quantity             14
discount             27
profit            24575
shipping_cost     10037
order_priority        4
year                  4
dtype: int64
```

In [148]:
```python
superstoreorders_df.isnull().sum()
```

Out[148]:
```
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             2630
quantity             0
discount             0
profit               0
shipping_cost        0
order_priority       0
year                 0
dtype: int64
```

In [149]:
```python
superstoreorders_df['sales']=superstoreorders_df['sales'].fillna(0)
```

In [153]:
```python
superstoreorders_df['ship_date'] = pd.to_datetime(superstoreorders_df.ship_date)
```

In [154]: `superstoreorders_df['month']= pd.DatetimeIndex(superstoreorders_df.ship_date).mon`

In [155]: `superstoreorders_df['day']= pd.DatetimeIndex(superstoreorders_df.ship_date).day`

In [156]: `superstoreorders_df['weekday']= pd.DatetimeIndex(superstoreorders_df.ship_date).w`

In [157]: `superstoreorders_df`

Out[157]:

	order_id	order_date	ship_date	ship_mode	customer_name	segment	state	cour
0	AG-2011-2040	2011-01-01	2011-06-01	Standard Class	Toby Braunhardt	Consumer	Constantine	Alge
1	IN-2011-47883	2011-01-01	2011-08-01	Standard Class	Joseph Holt	Consumer	New South Wales	Austr
2	HU-2011-1220	2011-01-01	2011-05-01	Second Class	Annie Thurman	Consumer	Budapest	Hung
3	IT-2011-3647632	2011-01-01	2011-05-01	Second Class	Eugene Moren	Home Office	Stockholm	Swed
4	IN-2011-47883	2011-01-01	2011-08-01	Standard Class	Joseph Holt	Consumer	New South Wales	Austr
...	...	...	...	...	...	...	...	
51285	CA-2014-115427	2014-12-31	2015-04-01	Standard Class	Erica Bern	Corporate	California	Uni Sta
51286	MO-2014-2560	2014-12-31	2015-05-01	Standard Class	Liz Preis	Consumer	Souss-Massa-Draâ	Moro
51287	MX-2014-110527	2014-12-31	2015-02-01	Second Class	Charlotte Melton	Consumer	Managua	Nicarag
51288	MX-2014-114783	2014-12-31	2015-06-01	Standard Class	Tamara Dahlen	Consumer	Chihuahua	Mex
51289	CA-2014-156720	2014-12-31	2015-04-01	Standard Class	Jill Matthias	Consumer	Colorado	Uni Sta

51290 rows × 24 columns
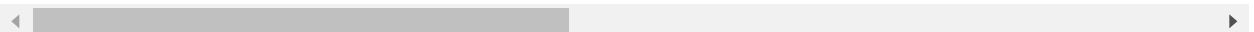
In [158]: `import jovian`

In [159]: `jovian.commit()`

```
<IPython.core.display.Javascript object>

[jovian] Updating notebook "adetounawobajo026/zerotopandas-course-project-start
er" on https://jovian.ai/ (https://jovian.ai/)
[jovian] Committed successfully! https://jovian.ai/adetounawobajo026/zerotopand
as-course-project-starter (https://jovian.ai/adetounawobajo026/zerotopandas-cou
rse-project-starter)
```

Out[159]: `'https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter'`

Exploratory Analysis and Visualisation

Exploratory data analysis and visualisation is the process used by data scientists to analyse,investigate datasets , and summarize their main characteristics, often employing data visualisation methods.

In [160]:
```
pip install matplotlib seaborn
```

```
Requirement already satisfied: matplotlib in c:\users\adesoji awobajo\appdata\l
ocal\programs\python\python311\lib\site-packages (3.6.2)
Requirement already satisfied: seaborn in c:\users\adesoji awobajo\appdata\loca
l\programs\python\python311\lib\site-packages (0.12.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\adesoji awobajo\app
data\local\programs\python\python311\lib\site-packages (from matplotlib) (1.0.
6)
Requirement already satisfied: cycler>=0.10 in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\adesoji awobajo\ap
pdata\local\programs\python\python311\lib\site-packages (from matplotlib) (4.3
8.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\adesoji awobajo\ap
pdata\local\programs\python\python311\lib\site-packages (from matplotlib) (1.4.
4)
Requirement already satisfied: numpy>=1.19 in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (from matplotlib) (1.23.5)
Requirement already satisfied: packaging>=20.0 in c:\users\adesoji awobajo\appd
ata\local\programs\python\python311\lib\site-packages (from matplotlib) (21.3)
Requirement already satisfied: pillow>=6.2.0 in c:\users\adesoji awobajo\appdat
a\local\programs\python\python311\lib\site-packages (from matplotlib) (9.3.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\adesoji awobajo\app
data\local\programs\python\python311\lib\site-packages (from matplotlib) (3.0.
9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\adesoji awobajo
\appdata\local\programs\python\python311\lib\site-packages (from matplotlib)
(2.8.2)
Requirement already satisfied: pandas>=0.25 in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (from seaborn) (1.5.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\adesoji awobajo\appdata
\local\programs\python\python311\lib\site-packages (from pandas>=0.25->seaborn)
(2022.6)
Requirement already satisfied: six>=1.5 in c:\users\adesoji awobajo\appdata\loc
al\programs\python\python311\lib\site-packages (from python-dateutil>=2.7->matp
lotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.


[notice] A new release of pip available: 22.3 -> 22.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [161]:
```python
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size']=14
matplotlib.rcParams['figure.figsize']=(9,5)
matplotlib.rcParams['figure.facecolor']='#00000000'
```

TODO

Which market made the highest profit per year

```
In [162]:  plt.figure(figsize=(12,6))
           plt.title('profit year market dimensions ')
           sns.scatterplot(x='year',
                           y='profit',
                           hue='market',
                           s=100,
                           data=superstoreorders_df);
```
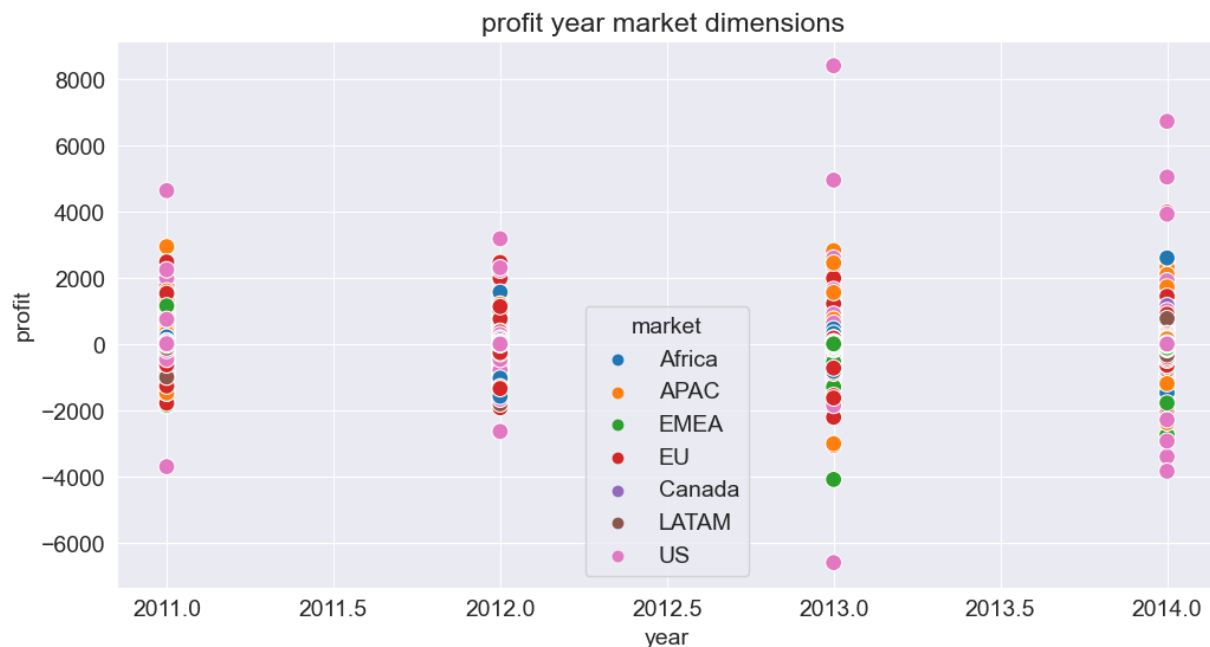

This is a scatter plot graph that shows the profit made by each market and the year using a seaborn which indicates each market with a identification of colours which made it unique.From the graph, it is shown that US made more profit in 2011, 2012,2013 and in 2014.

TODO

which sub_category of goods was mostly demanded for

In [163]:
```python
plt.figure(figsize=(12,6))
plt.title('Distribution of count sub_category ',fontsize=16,fontweight='bold')
plt.hist(superstoreorders_df.sub_category.value_counts(), bins=5, color='purple')
plt.show()
```

Distribution of count sub_category
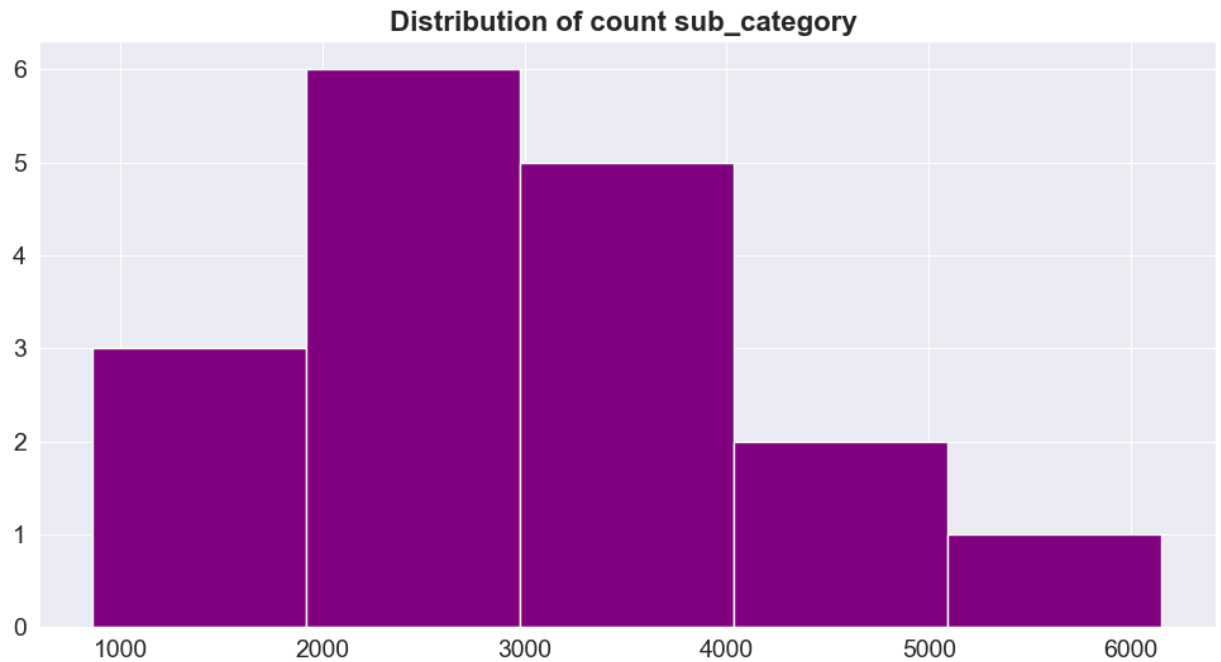
This is a histogram that shows the distibution value count of sub_category into five bins. from the graph plot above it shows that the highest count is between 2000-3000

TODO

which top ten countries made most sales

In [164]:
```python
plt.figure(figsize=(12,6))
superstoreorders_df.groupby('country')[['sales']].sum().sort_values('sales',ascen
plt.title('Ten countries with highest sales');
```
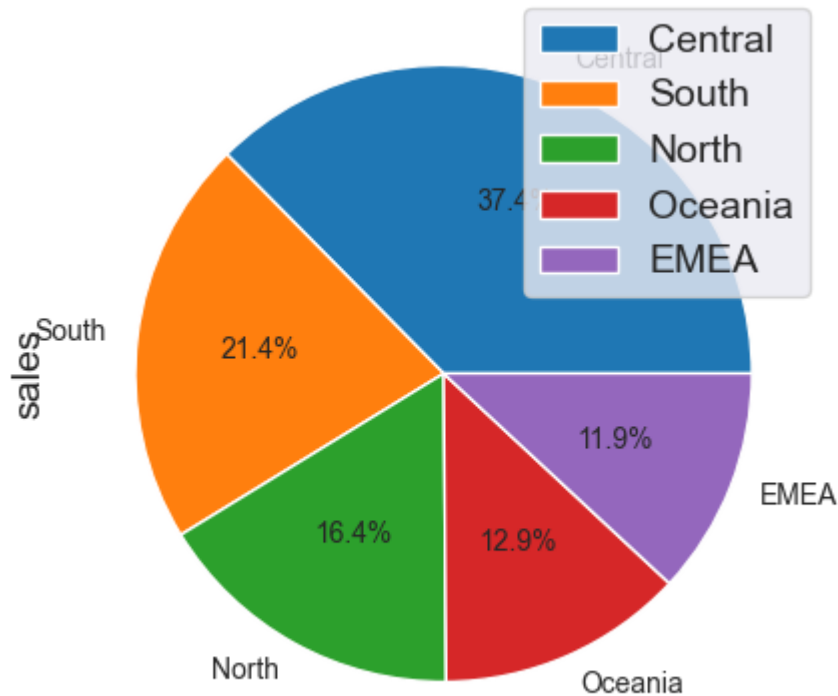
`<Figure size 1200x600 with 0 Axes>`
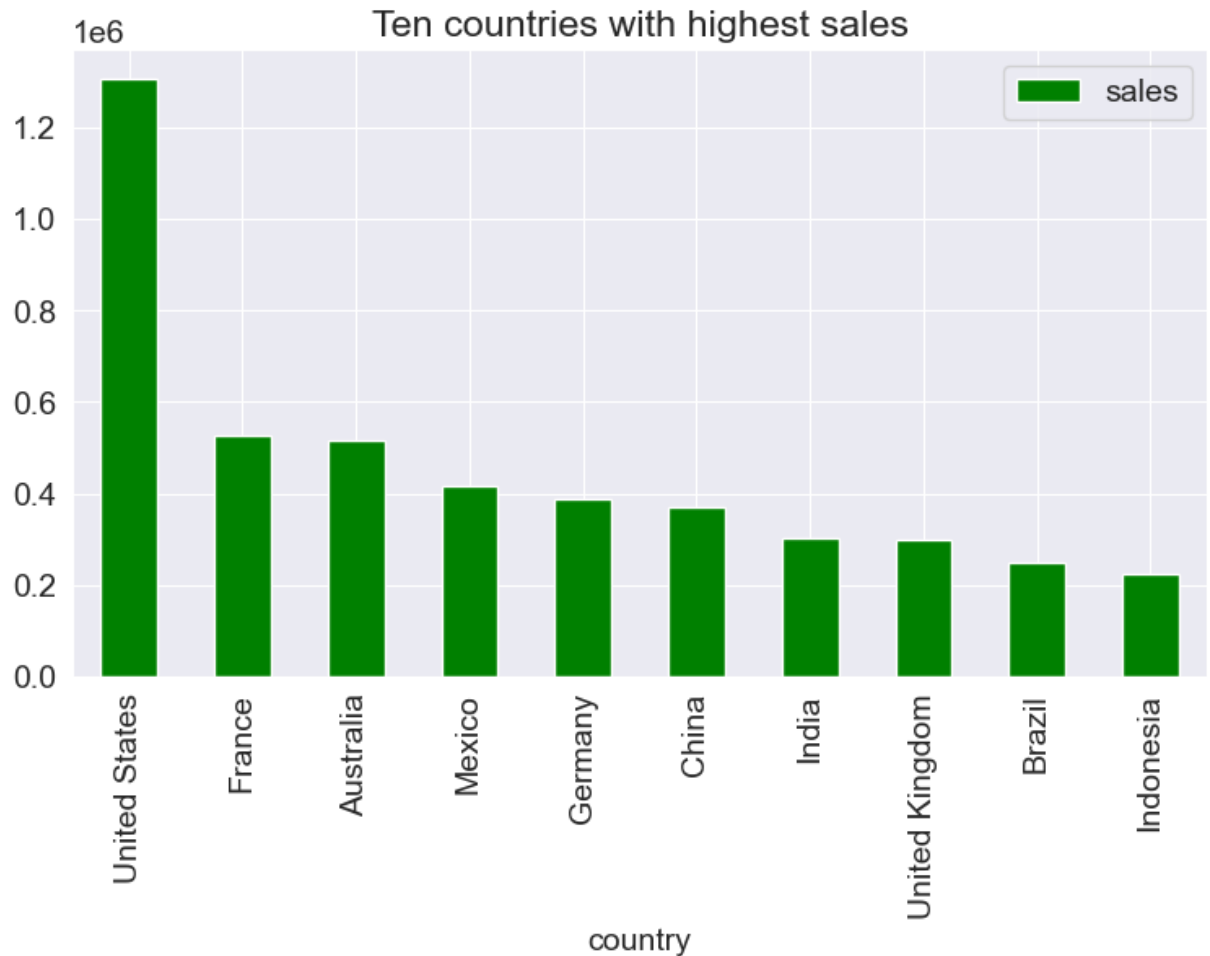
This is a bar plot, that shows the ten top countries with highest sales. from the graph above indicates that UNITED STATE made more sales than other countries.

TODO

which top 5 region has the highest and lowest percentage of sales

In [174]:
```python
plt.figure(figsize=(30,30))
superstoreorders_df.groupby('region')[['sales']].sum().sort_values('sales',ascend
```
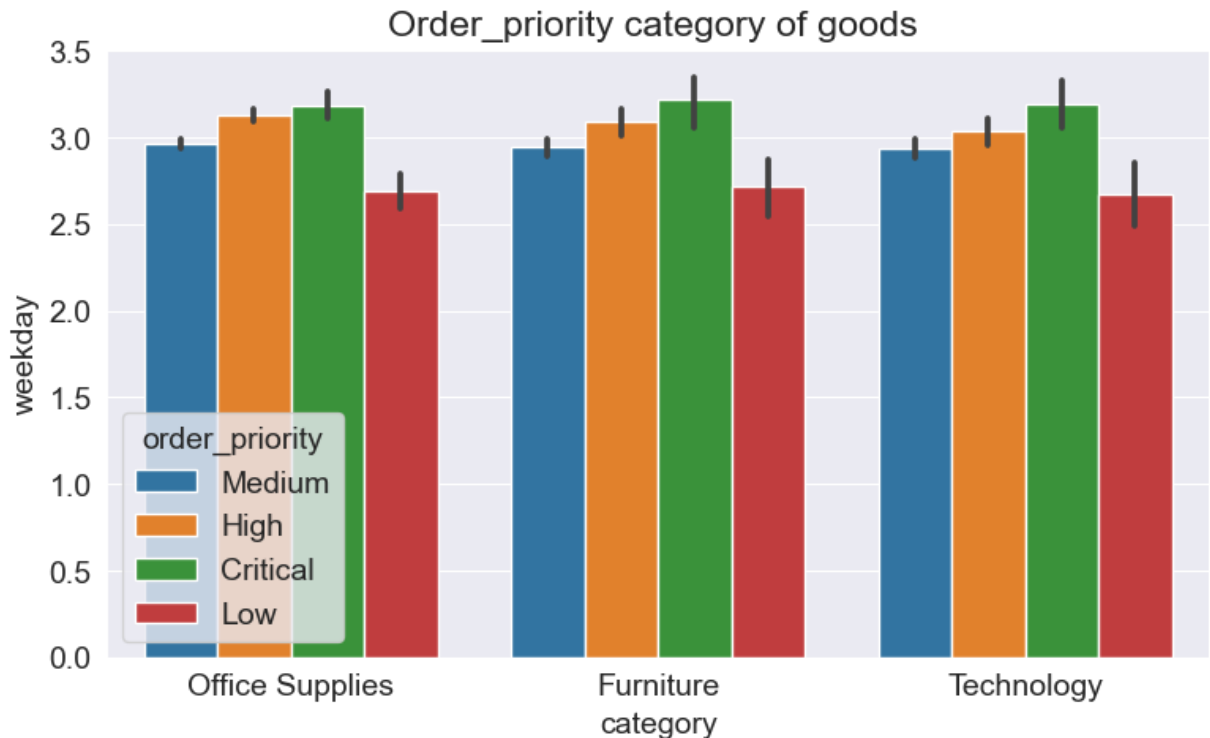
`<Figure size 3000x3000 with 0 Axes>`

This is a pie chart that shows the percentage of sales of top Five regions. The graph shows CENTRAL has the highest percentage of sales and EMEA has the lowest.

TODO

what is the order priority of each category of Goods

In [175]:
```
sns.barplot(x='category',y='weekday',hue='order_priority',data=superstoreorders_c
plt.title('Order_priority category of goods ');
```


This graph shows the order priority of category of goods per week using a bar chart with seaborn.The graph shows averagely which category of goods is mostly ordered.

In [176]:
```
import jovian
```

In [178]:
```
jovian.commit()
```

```
<IPython.core.display.Javascript object>

[jovian] Updating notebook "adetounawobajo026/zerotopandas-course-project-start
er" on https://jovian.ai/ (https://jovian.ai/)
[jovian] Committed successfully! https://jovian.ai/adetounawobajo026/zerotopand
as-course-project-starter (https://jovian.ai/adetounawobajo026/zerotopandas-cou
rse-project-starter)
```

Out[178]:
```
'https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter'
```

Asking and Answering Questions

TODO-

we have gain several insights on the sales store orders data frame by exploring columns from the dataset.,let's ask five questions and provide answers to them using pandas, seaborn and matplotlib for graphs, data cleaning when necessary.

Q1: TODO- what is the percentage of Grossprofit?

```
In [179]: percentage_of_grossprofit=superstoreorders_df.profit.sum()/superstoreorders_df.sa
```

```
In [180]: print('The percentage of grossprofit is {:.2f}%.'.format(percentage_of_grossprofi
```

```
The percentage of grossprofit is 18.75%.
```

finding the grossprofit percentage from the data frame using pandas by summing the profit and dividing the summing of the sales multiply by 100 and placing in two decimal place.

Q2: TODO - Which ship mode is mostly used

```
In [181]: sns.countplot(x='ship_mode',data=superstoreorders_df)
          plt.title('ship_mode counts');
```
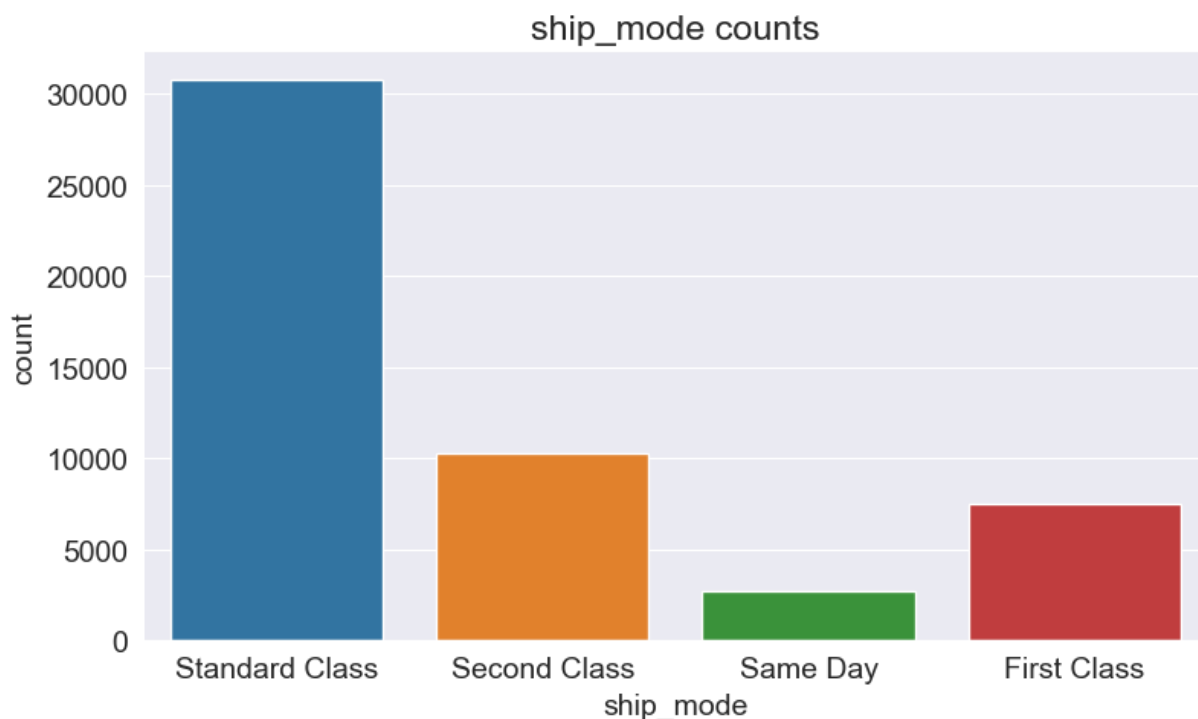
This graph shows that STANDARD CLASS shipmode is mostly used. seaborn is used here to make the graph look more beautiful and presentable by writing one line of code to count the number of times each ship mode is been used.

Q3: TODO-Which Top Ten customers made the highest profit

In [182]: `superstoreorders_df.groupby('customer_name')[['profit']].sum().sort_values('profi`

Out[182]:

|	profit
customer_name	
Tamara Chand	8672.89890
Raymond Buch	8453.04950
Sanjit Chand	8205.37990
Hunter Lopez	7816.56778
Bill Eplett	7410.00530
Harry Marie	6958.28640
Susan Pistek	6484.40726
Mike Gockenbach	6458.67620
Adrian Barton	6417.28450
Tom Ashbrook	6311.97910

Retrieving of the top ten customers with the highest profit from the dataframe using pandas groupby and sort values to place according to the highest customer and profit.From the information retrieved, it is shown that TAMARA CHAND is the customer with the highest profit.

Q4: TODO- What Ten state has the most used subcategory from the least

In [187]:
```python
plt.figure(figsize=(30,15))
superstoreorders_df.groupby('sub_category')[['state']].count().sort_values('state
```

Out[187]: <AxesSubplot: ylabel='state'>

<Figure size 3000x1500 with 0 Axes>
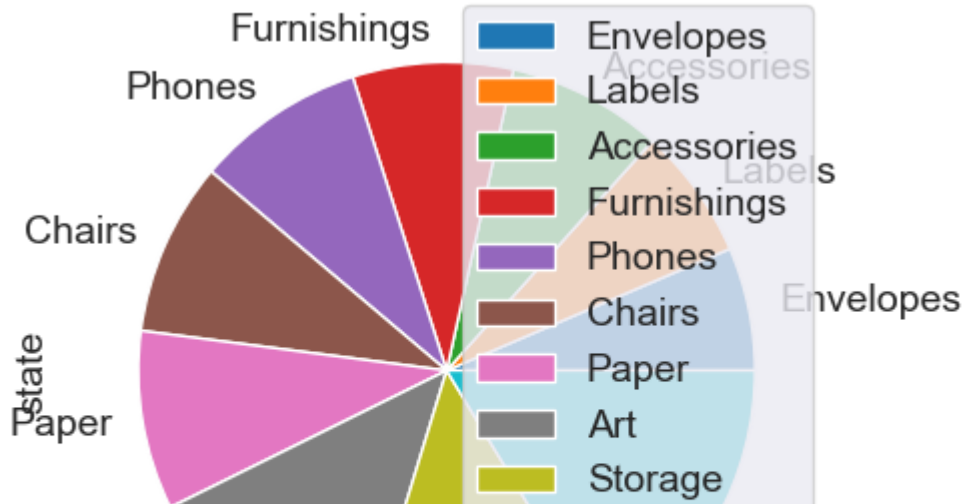
This is a pie chart that shows the least of ten sub category of goods.in this pandas is used to group and sort each columns from the dataframe and then plot the pie chart.

Q5: TODO- Which Ten product name has the highest sales

In [188]:
```python
superstoreorders_df.groupby('product_name')[['sales']].sum().sort_values('sales'
```

Out[188]:

product_name	sales
Eldon File Cart, Single Width	31319.0
Rogers File Cart, Single Width	22645.0
Tenex File Cart, Single Width	20778.0
Smead File Cart, Single Width	20775.0
Office Star Executive Leather Armchair, Adjustable	19355.0
Fellowes Lockers, Industrial	19172.0
Smead Lockers, Industrial	18648.0
Hewlett Copy Machine, Color	16849.0
Rogers Lockers, Blue	16494.0
Fellowes Lockers, Wire Frame	16470.0

Retrieving of the top ten product with the highest sales from the dataframe using pandas groupby and sort values to place according to the highest product_name and sales.From the information retrieved, it is shown that ELDON FILE CART,SINGLE WIDTH is the Product_name with the highest sales.

In [189]: `import jovian`

In [191]: `jovian.commit()`

`<IPython.core.display.Javascript object>`

[jovian] Updating notebook "adetounawobajo026/zerotopandas-course-project-starter" on https://jovian.ai/ (https://jovian.ai/)
[jovian] Committed successfully! https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter (https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter)

Out[191]: `'https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter'`

Inferences and Conclusion

These are some of the inferences from the analysis. Here is a summary of them: . The market is not too competitive in nature based on the fact that, from the scatter plot graph above shows that a particular market took the lead throughout the years and profit made is high. . Even in the country and the region only one particular country has the highest sales, the competition is low. the remaining countries and regions are struggling to get there. .There is demand for all the sub_category of goods in a margin of minimal level ,almost at the same level . .The order priority of each category are in average level, actually almost the same level. .There is a high gap between the ship mode mostly used, it is the standard class that is the mostly used ship mode from the rest. . The profit made by each customer is competitive.There is a little margin between them . . There is also high competition between each sales of product name.

In [192]: `import jovian`

In [193]: `jovian.commit()`

`<IPython.core.display.Javascript object>`

[jovian] Updating notebook "adetounawobajo026/zerotopandas-course-project-starter" on https://jovian.ai/ (https://jovian.ai/)
[jovian] Committed successfully! https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter (https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter)

Out[193]: `'https://jovian.ai/adetounawobajo026/zerotopandas-course-project-starter'`

References and Future Work

Check out the following resources to learn more about the dataset and tools used in this notebook:

. Opendatasets python library: https://github.com/jovianML/opendatasets (https://github.com/jovianML/opendatasets)

. superstore sales dataset: https://www.kaggle.com/laibaanwer/superstore-sales-dataset (https://www.kaggle.com/laibaanwer/superstore-sales-dataset)

. Pandas user guide: https://pandas.pydata.org/docs/user_guide/index.html (https://pandas.pydata.org/docs/user_guide/index.html)

. Matplotlib user guide https://matplotlib.org/3.3.1/users/index.html (https://matplotlib.org/3.3.1/users/index.html)

. Seaborn user guide & tutorial: https://seaborn.pydata.org/tutorial.html (https://seaborn.pydata.org/tutorial.html)

As a next step, you can try out a project on another dataset of your choice: https://www.kaggle.com/datasets?fileType=csv (https://www.kaggle.com/datasets?fileType=csv)

In [ ]: