#01 Sales Project

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Load and describe data

In [2]:
```
sales = pd.read_excel("superstore_sales.xlsx")
```

In [3]:
```
sales.describe()
```

Out[3]:

	sales	quantity	discount	profit	shipping_cost	year
count	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000
mean	246.490581	3.476545	0.142908	28.641740	26.375818	2012.777208
std	487.565361	2.278766	0.212280	174.424113	57.296810	1.098931
min	0.444000	1.000000	0.000000	-6599.978000	0.002000	2011.000000
25%	30.758625	2.000000	0.000000	0.000000	2.610000	2012.000000
50%	85.053000	3.000000	0.000000	9.240000	7.790000	2013.000000
75%	251.053200	5.000000	0.200000	36.810000	24.450000	2014.000000
max	22638.480000	14.000000	0.850000	8399.976000	933.570000	2014.000000

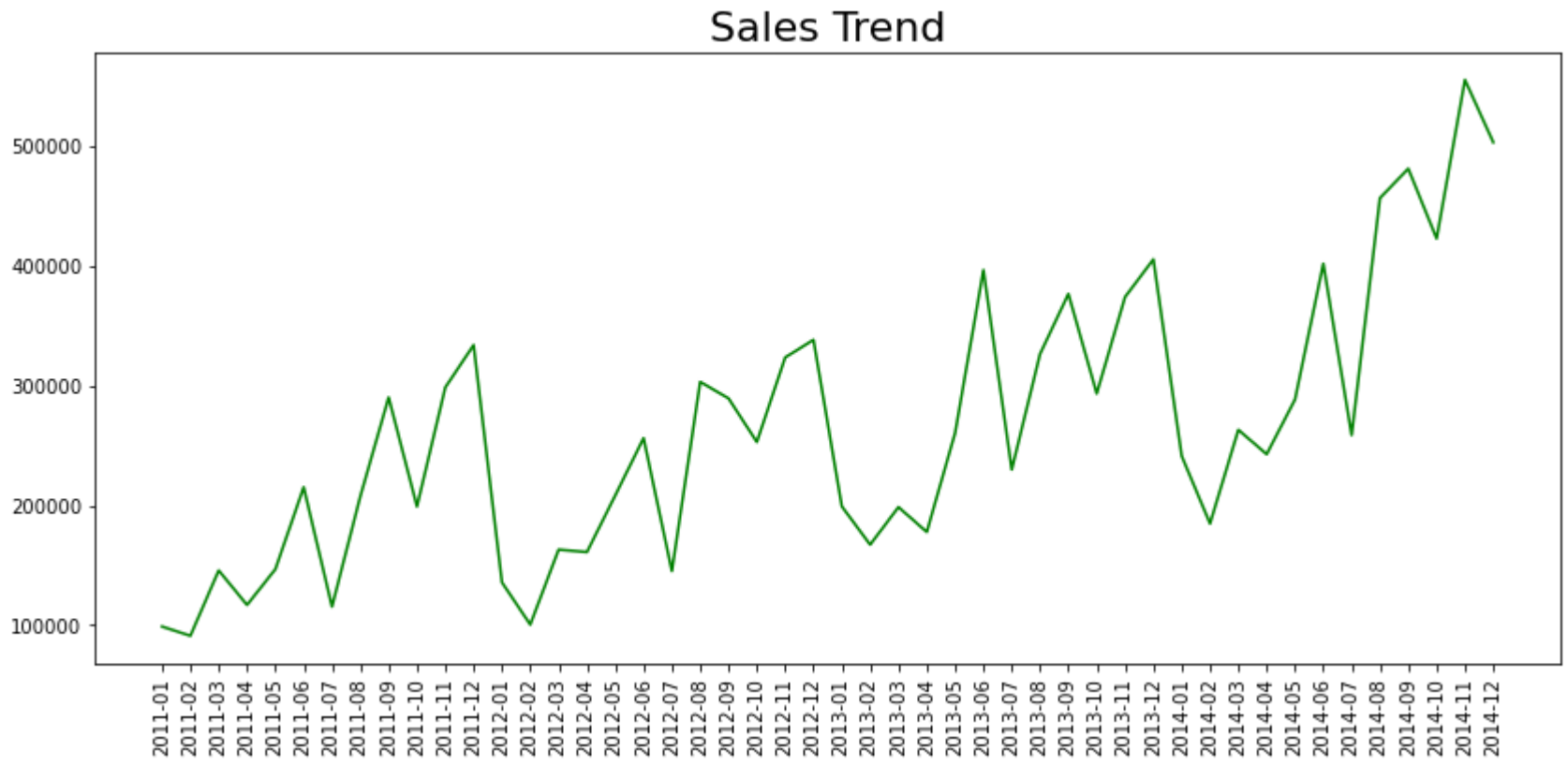
In [7]:
```
sales.info()
```

```
<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  datetime64[ns]
 2   ship_date       51290 non-null  datetime64[ns]
 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  float64
 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: datetime64[ns](2), float64(4), int64(2), object(13)
memory usage: 5.7+ MB
```

WHAT IS THE OVERALL SALES TREND?

In [4]:
```
to_ym = lambda x: str(x)[:7]
sales['year_month'] = sales['order_date'].apply(to_ym)
sales_trend = sales.groupby('year_month')['sales'].sum()
```

In [5]:
```
plt.figure(figsize=(14,6))
plt.plot(sales_trend.index , sales_trend.values , color='green')
plt.xticks(rotation='vertical' , size=10)
plt.title("Sales Trend" , size=22)
plt.show()
```

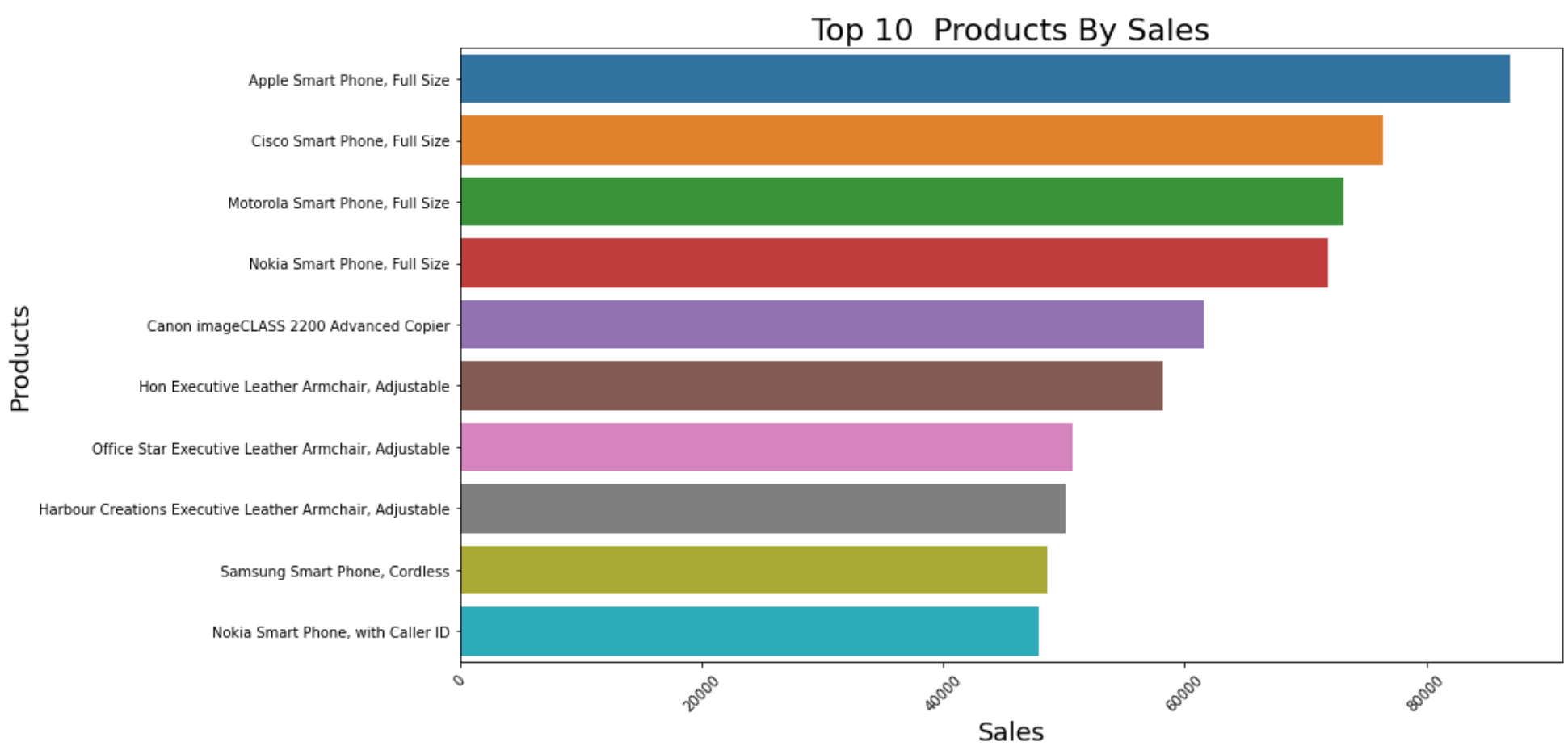


WHICH ARE THE TOP 10 PRODUCTS BY SALES

In [6]:
```
top_10_product = sales.groupby('product_name').sum()['sales'].sort_values(ascending=0)[:10]
top_10_product
```

Out[6]:
```
product_name
Apple Smart Phone, Full Size                          86935.7786
Cisco Smart Phone, Full Size                          76441.5306
Motorola Smart Phone, Full Size                       73156.3030
Nokia Smart Phone, Full Size                          71904.5555
Canon imageCLASS 2200 Advanced Copier                 61599.8240
Hon Executive Leather Armchair, Adjustable            58193.4841
Office Star Executive Leather Armchair, Adjustable    50661.6840
Harbour Creations Executive Leather Armchair, Adjustable    50121.5160
Samsung Smart Phone, Cordless                         48653.4600
Nokia Smart Phone, with Caller ID                     47877.7857
Name: sales, dtype: float64
```

In [7]:
```
plt.figure(figsize=(14,8))
sns.barplot(y = top_10_product.index , x = top_10_product.values)
plt.xticks(rotation=45, size=10)
plt.title("Top 10  Products By Sales" , size=22)
plt.xlabel('Sales', size=18)
plt.ylabel('Products', size=18)
plt.show()
```

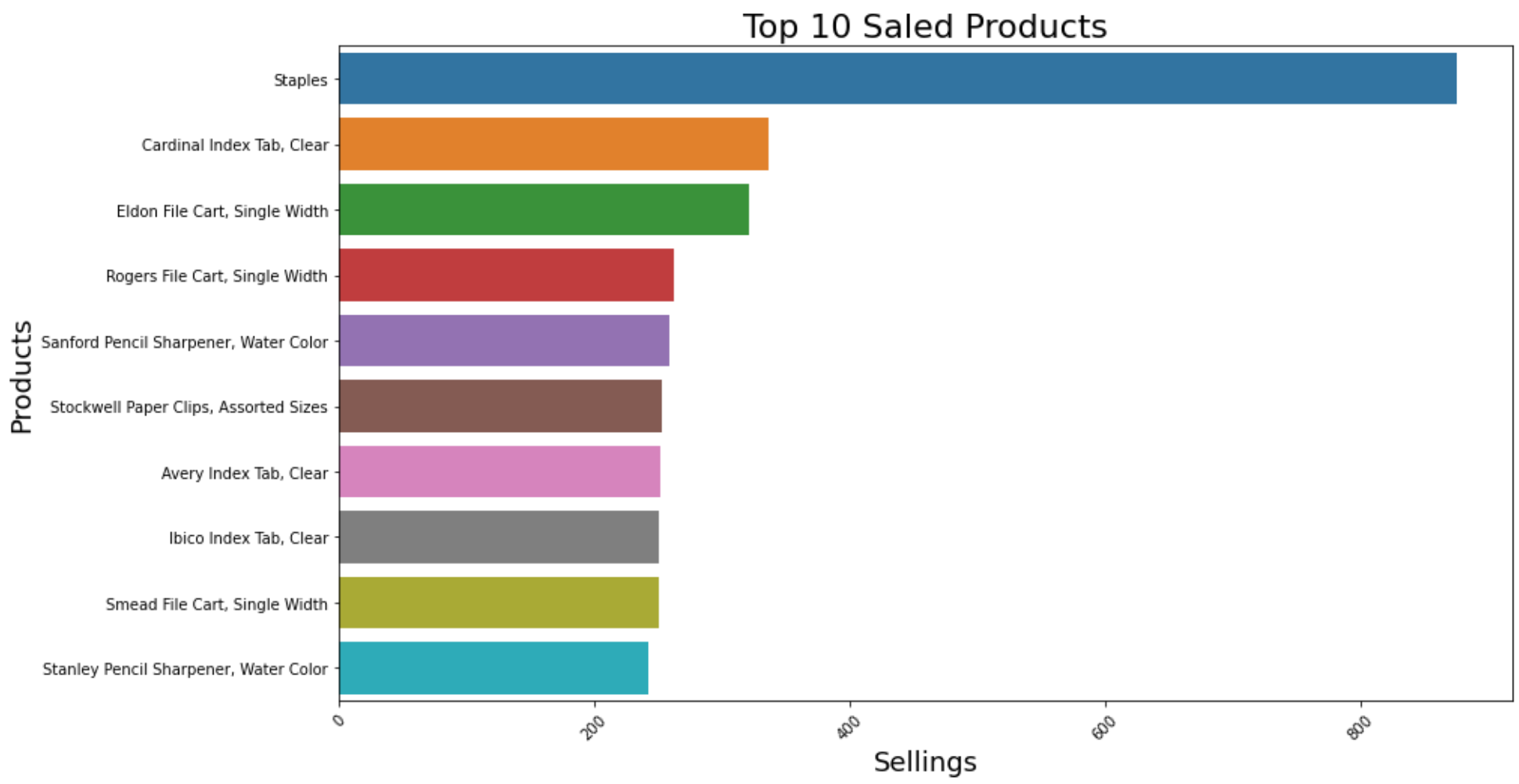


WHICH ARE THE MOST SELLING PRODUCTS

In [8]:
```
most_selling = sales.groupby("product_name")['quantity'].sum().sort_values(ascending = 0)[:10]
most_selling
```

Out[8]:
```
product_name
Staples                                 876
Cardinal Index Tab, Clear               337
Eldon File Cart, Single Width           321
Rogers File Cart, Single Width          262
Sanford Pencil Sharpener, Water Color   259
Stockwell Paper Clips, Assorted Sizes   253
Avery Index Tab, Clear                  252
Ibico Index Tab, Clear                  251
Smead File Cart, Single Width           250
Stanley Pencil Sharpener, Water Color   242
Name: quantity, dtype: int64
```

In [63]:
```
plt.figure(figsize=(14,8))
sns.barplot(y = most_selling.index , x = most_selling.values)
plt.xticks(rotation=45, size=10)
plt.title("Top 10 Saled Products" , size=22)
plt.xlabel('Sellings', size=18)
plt.ylabel('Products', size=18)
plt.show()
```

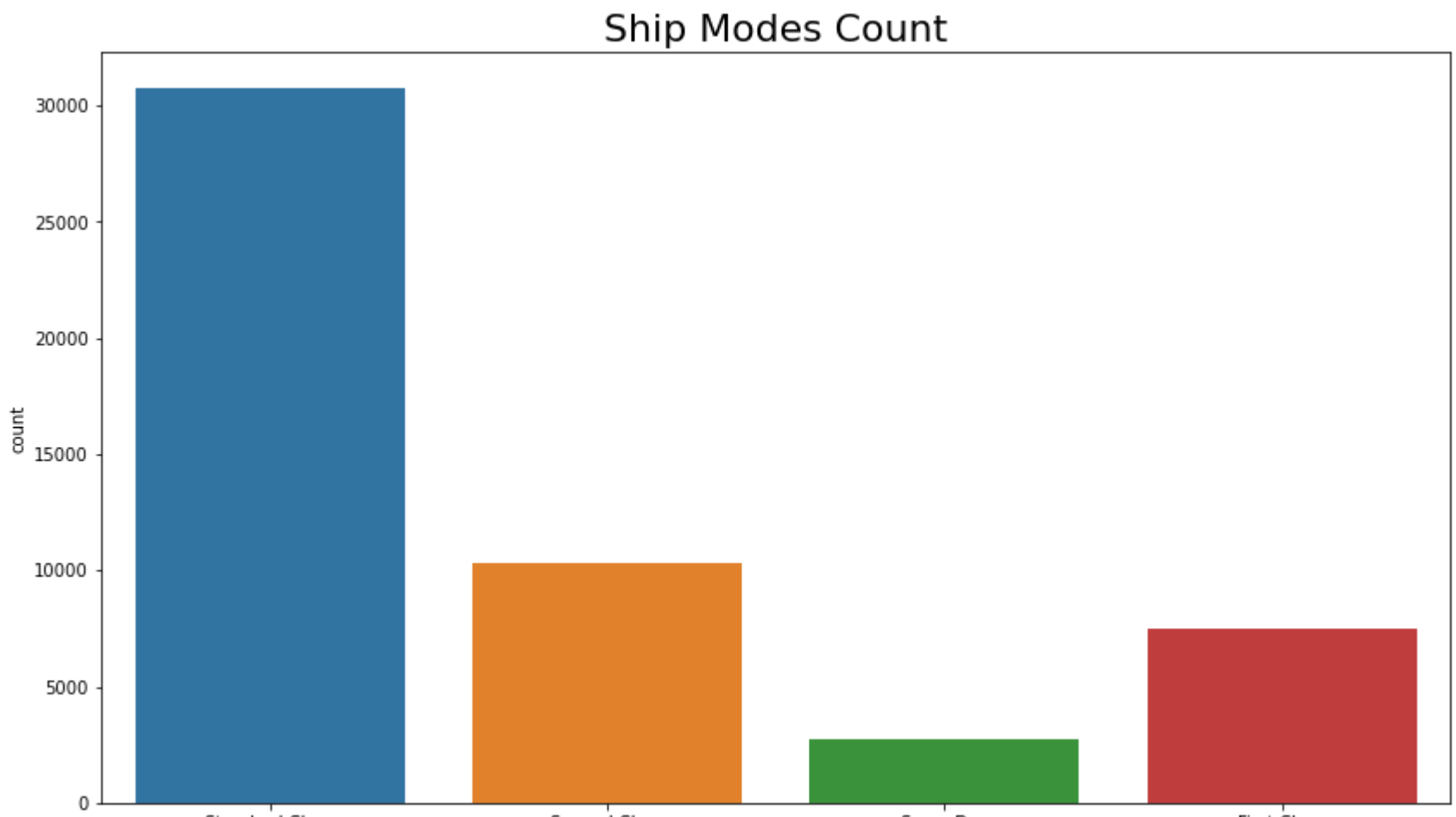


WHAT IS THE MOST PREFERRED SHIP MODE

In [9]:
```
sales_count = sales.copy()
sales_count['count'] = 1
most_ship_mode = sales_count.groupby("ship_mode")['count'].sum()
most_ship_mode
```

Out[9]:
```
ship_mode
First Class        7505
Same Day           2701
Second Class      10309
Standard Class    30775
Name: count, dtype: int64
```

In [10]:
```
plt.figure(figsize=(14,8))
sns.countplot(x='ship_mode' , data=sales)
plt.title("Ship Modes Count", size=22)
plt.xlabel('Ship Mode', size=10)
plt.show()
```

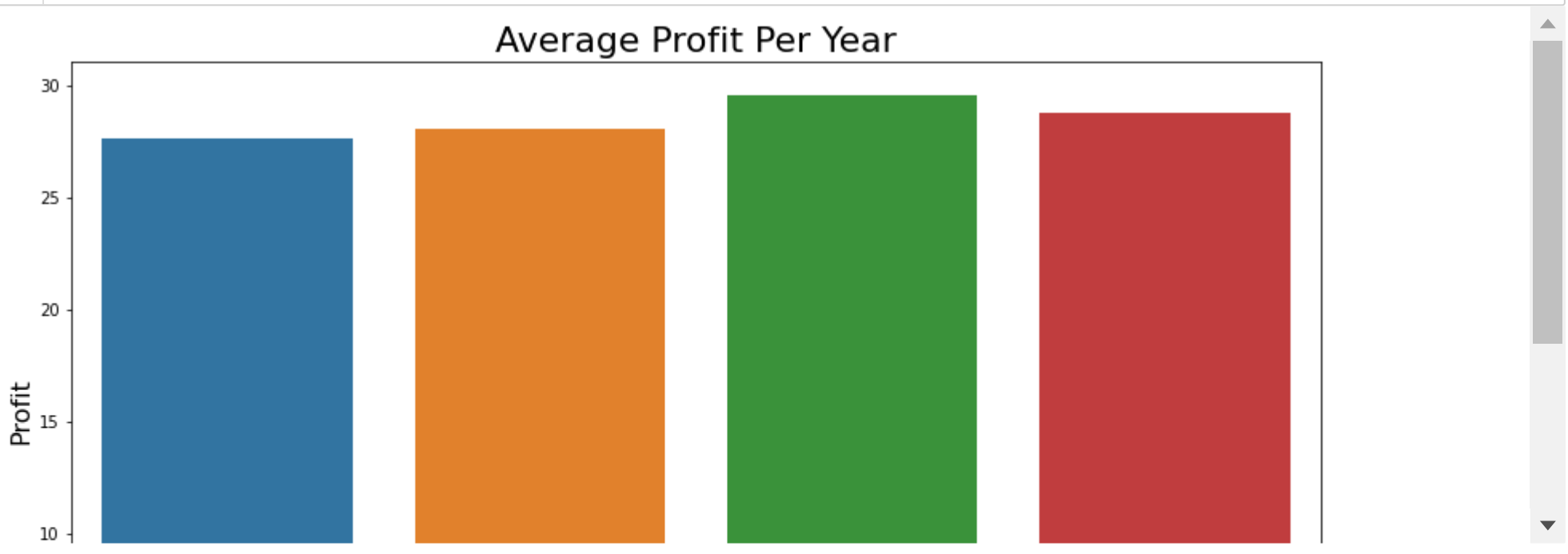


WHAT IS AVERAGE PROFIT FOR EACH YEAR

In [11]:
```
average_profit_per_year = sales.groupby('year')['profit'].mean()
average_profit_per_year
```

Out[11]:
```
year
2011    27.666238
2012    28.043722
2013    29.604519
2014    28.758540
Name: profit, dtype: float64
```

In [12]:
```
plt.figure(figsize=(14,8))
sns.barplot(x=average_profit_per_year.index , y= average_profit_per_year.values)
plt.title("Average Profit Per Year", size=22)
plt.xlabel('Year', size=16)
plt.ylabel('Profit', size=16)
plt.show()
```



By : Musab Shifu