E-Commerce Supply Chain Analysis

July 31, 2023

Project Portfolio: E-Commerce Supply Chain Analysis using Python Python libraries: Pandas, Matplotlib, and Seaborn!

Karimjon Mullaboev

Introduction:

This project focuses on analyzing the E-Commerce Supply Chain Analysis. As data scientists, we aim to provide valuable insights and information to help Company ABC's stakeholders make informed decisions. While the business problem and task might vary in real-world scenarios, this project will showcase various Python data science techniques and visualizations to gain meaningful insights from the dataset. Let's investigate the analysis and explore the data to uncover valuable information for Company ABC's supply chain.

Project Objectives: The Project aim is to:

1. Analyzing Revenue by Product Type: - Investigating the revenue generated by different product types to understand their contribution to overall sales. - Identifying valuable patterns and trends in customer preferences to inform marketing and product strategies.

2. Unraveling Sales by Product Type: - Analyzing sales data to determine each product type's sales share and volumes. - Uncovering valuable market trends to assist businesses in making data-driven decisions for maximizing profits.

3. Exploring Revenue from Shipping Carriers: - Evaluating the revenue generated by different shipping carriers to identify efficient and cost-effective options for shipping and delivery.

4. Analyzing Revenue by SKU (Stock Keeping Unit): - Examining SKU-specific data to gain insights into the performance of individual products. - Empowering businesses to optimize their inventory strategies based on SKU performance.

5. Understanding Shipping Costs of Carriers: - Analyzing shipping costs to identify areas for cost optimization while maintaining high-quality service.

6. Cost Distribution by Transportation Modes: - Breaking down costs across different transportation modes to assess logistics efficiency.

7. Assessing Defect Rates during Shipping: - Investigating defect rates during shipping to ensure improved customer satisfaction and identify areas of improvement in the shipping process.

[1]:
```
#Importing Libraries
import numpy as np
import pandas as pd
```

```
import matplotlib.pyplot as plt
import seaborn as sns
```

Data

The data was accessed here: https://github.com/areeb399/E-Commerce-Sales-Analysis it contains 100 raw and 24 columns, including numerical, categorical and binary variables.

[106]:
```
#Importing Data
df = pd.read_csv("supply_chain_data.csv")
df
```

[106]:
```
   Product type    SKU      Price  Availability  Number of products sold  \
0      haircare   SKU0  69.808006            55                      802
1      skincare   SKU1  14.843523            95                      736
2      haircare   SKU2  11.319683            34                        8
3      skincare   SKU3  61.163343            68                       83
4      skincare   SKU4   4.805496            26                      871
..          …      …          …           …                      …
95     haircare  SKU95  77.903927            65                      672
96    cosmetics  SKU96  24.423131            29                      324
97     haircare  SKU97   3.526111            56                       62
98     skincare  SKU98  19.754605            43                      913
99     haircare  SKU99  68.517833            17                      627

    Revenue generated Customer demographics  Stock levels  Lead times  \
0         8661.996792            Non-binary            58           7
1         7460.900065                Female            53          30
2         9577.749626               Unknown             1          10
3         7766.836426            Non-binary            23          13
4         2686.505152            Non-binary             5           3
..                …                     …           …         …
95        7386.363944               Unknown            15          14
96        7698.424766            Non-binary            67           2
97        4370.916580                  Male            46          19
98        8525.952560                Female            53           1
99        9185.185829               Unknown            55           8

    Order quantities  …  Location Lead time  Production volumes  \
0                 96  …    Mumbai        29                 215
1                 37  …    Mumbai        23                 517
2                 88  …    Mumbai        12                 971
3                 59  …   Kolkata        24                 937
4                 56  …     Delhi         5                 414
..               …   …       …        …                  …
95                26  …    Mumbai        18                 450
96                32  …    Mumbai        28                 648
97                 4  …    Mumbai        10                 535
```

```
98                27  …   Chennai        28                 581
99                59  …   Chennai        29                 921

   Manufacturing lead time Manufacturing costs  Inspection results  \
0                       29           46.279879             Pending
1                       30           33.616769             Pending
2                       27           30.688019             Pending
3                       18           35.624741                Fail
4                        3           92.065161                Fail
..                     …                  …                 …
95                      26           58.890686             Pending
96                      28           17.803756             Pending
97                      13           65.765156                Fail
98                       9            5.604691             Pending
99                       2           38.072899                Fail

    Defect rates  Transportation modes   Routes       Costs
0       0.226410                  Road  Route B  187.752075
1       4.854068                  Road  Route B  503.065579
2       4.580593                   Air  Route C  141.920282
3       4.746649                  Rail  Route A  254.776159
4       3.145580                   Air  Route A  923.440632
..           …                    …       …         …
95      1.210882                   Air  Route A  778.864241
96      3.872048                  Road  Route A  188.742141
97      3.376238                  Road  Route A  540.132423
98      2.908122                  Rail  Route A  882.198864
99      0.346027                  Rail  Route B  210.743009

[100 rows x 24 columns]
```

[107]:
```
#Top 10 Data
df.head(10)
```

[107]:
```
  Product type   SKU      Price  Availability  Number of products sold  \
0     haircare  SKU0  69.808006            55                      802
1     skincare  SKU1  14.843523            95                      736
2     haircare  SKU2  11.319683            34                        8
3     skincare  SKU3  61.163343            68                       83
4     skincare  SKU4   4.805496            26                      871
5     haircare  SKU5   1.699976            87                      147
6     skincare  SKU6   4.078333            48                       65
7    cosmetics  SKU7  42.958384            59                      426
8    cosmetics  SKU8  68.717597            78                      150
9     skincare  SKU9  64.015733            35                      980

   Revenue generated Customer demographics  Stock levels  Lead times  \
```

0	8661.996792	Non-binary	58	7
1	7460.900065	Female	53	30
2	9577.749626	Unknown	1	10
3	7766.836426	Non-binary	23	13
4	2686.505152	Non-binary	5	3
5	2828.348746	Non-binary	90	27
6	7823.476560	Male	11	15
7	8496.103813	Female	93	17
8	7517.363211	Female	5	10
9	4971.145988	Unknown	14	27

	Order quantities	…	Location	Lead time	Production volumes	\
0	96	…	Mumbai	29	215	
1	37	…	Mumbai	23	517	
2	88	…	Mumbai	12	971	
3	59	…	Kolkata	24	937	
4	56	…	Delhi	5	414	
5	66	…	Bangalore	10	104	
6	58	…	Kolkata	14	314	
7	11	…	Bangalore	22	564	
8	15	…	Mumbai	13	769	
9	83	…	Chennai	29	963	

	Manufacturing lead time	Manufacturing costs	Inspection results	\
0	29	46.279879	Pending	
1	30	33.616769	Pending	
2	27	30.688019	Pending	
3	18	35.624741	Fail	
4	3	92.065161	Fail	
5	17	56.766476	Fail	
6	24	1.085069	Pending	
7	1	99.466109	Fail	
8	8	11.423027	Pending	
9	23	47.957602	Pending	

	Defect rates	Transportation modes	Routes	Costs
0	0.226410	Road	Route B	187.752075
1	4.854068	Road	Route B	503.065579
2	4.580593	Air	Route C	141.920282
3	4.746649	Rail	Route A	254.776159
4	3.145580	Air	Route A	923.440632
5	2.779194	Road	Route A	235.461237
6	1.000911	Sea	Route A	134.369097
7	0.398177	Road	Route C	802.056312
8	2.709863	Sea	Route B	505.557134
9	3.844614	Rail	Route B	995.929461

```
[10 rows x 24 columns]
```

[108]:
```python
#Bottom 10 Data
df.tail(10)
```

[108]:
```
   Product type    SKU      Price  Availability  Number of products sold  \
90     skincare  SKU90  13.881914            56                      320
91    cosmetics  SKU91  62.111965            90                      916
92    cosmetics  SKU92  47.714233            44                      276
93     haircare  SKU93  69.290831            88                      114
94    cosmetics  SKU94   3.037689            97                      987
95     haircare  SKU95  77.903927            65                      672
96    cosmetics  SKU96  24.423131            29                      324
97     haircare  SKU97   3.526111            56                       62
98     skincare  SKU98  19.754605            43                      913
99     haircare  SKU99  68.517833            17                      627

    Revenue generated Customer demographics  Stock levels  Lead times  \
90        9592.633570            Non-binary            66          18
91        1935.206794                  Male            98          22
92        2100.129755                  Male            90          25
93        4531.402134               Unknown            63          17
94        7888.356547               Unknown            77          26
95        7386.363944               Unknown            15          14
96        7698.424766            Non-binary            67           2
97        4370.916580                  Male            46          19
98        8525.952560                Female            53           1
99        9185.185829               Unknown            55           8

    Order quantities  …   Location Lead time  Production volumes  \
90                96  …  Bangalore         8                 585
91                85  …      Delhi         5                 207
92                10  …     Mumbai         4                 671
93                66  …    Chennai        21                 824
94                72  …      Delhi        12                 908
95                26  …     Mumbai        18                 450
96                32  …     Mumbai        28                 648
97                 4  …     Mumbai        10                 535
98                27  …    Chennai        28                 581
99                59  …    Chennai        29                 921

   Manufacturing lead time Manufacturing costs  Inspection results  \
90                       8           85.675963                Pass
91                      28           39.772883             Pending
92                      29           62.612690                Pass
93                      20           35.633652                Fail
94                      14           60.387379                Pass
```

```
95                        26         58.890686            Pending
96                        28         17.803756            Pending
97                        13         65.765156               Fail
98                         9          5.604691            Pending
99                         2         38.072899               Fail

    Defect rates  Transportation modes   Routes       Costs
90      1.219382                  Rail  Route B  990.078473
91      0.626002                  Rail  Route B  996.778315
92      0.333432                  Rail  Route B  230.092783
93      4.165782                   Air  Route A  823.523846
94      1.463607                  Rail  Route B  846.665257
95      1.210882                   Air  Route A  778.864241
96      3.872048                  Road  Route A  188.742141
97      3.376238                  Road  Route A  540.132423
98      2.908122                  Rail  Route A  882.198864
99      0.346027                  Rail  Route B  210.743009

[10 rows x 24 columns]
```

[109]:
```python
#Statistical Summary of the DataFrame
df.describe()
```

[109]:
```
            Price  Availability  Number of products sold  Revenue generated  \
count  100.000000    100.000000               100.000000         100.000000
mean    49.462461     48.400000               460.990000        5776.048187
std     31.168193     30.743317               303.780074        2732.841744
min      1.699976      1.000000                 8.000000        1061.618523
25%     19.597823     22.750000               184.250000        2812.847151
50%     51.239831     43.500000               392.500000        6006.352023
75%     77.198228     75.000000               704.250000        8253.976921
max     99.171329    100.000000               996.000000        9866.465458

       Stock levels  Lead times  Order quantities  Shipping times  \
count    100.000000  100.000000        100.000000      100.000000
mean      47.770000   15.960000         49.220000        5.750000
std       31.369372    8.785801         26.784429        2.724283
min        0.000000    1.000000          1.000000        1.000000
25%       16.750000    8.000000         26.000000        3.750000
50%       47.500000   17.000000         52.000000        6.000000
75%       73.000000   24.000000         71.250000        8.000000
max      100.000000   30.000000         96.000000       10.000000

       Shipping costs   Lead time  Production volumes  \
count      100.000000  100.000000          100.000000
mean         5.548149   17.080000          567.840000
std          2.651376    8.846251          263.046861
```

```
min           1.013487    1.000000          104.000000
25%           3.540248   10.000000          352.000000
50%           5.320534   18.000000          568.500000
75%           7.601695   25.000000          797.000000
max           9.929816   30.000000          985.000000

       Manufacturing lead time  Manufacturing costs  Defect rates       Costs
count                100.00000           100.000000    100.000000  100.000000
mean                  14.77000            47.266693      2.277158  529.245782
std                    8.91243            28.982841      1.461366  258.301696
min                    1.00000             1.085069      0.018608  103.916248
25%                    7.00000            22.983299      1.009650  318.778455
50%                   14.00000            45.905622      2.141863  520.430444
75%                   23.00000            68.621026      3.563995  763.078231
max                   30.00000            99.466109      4.939255  997.413450
```

[110]:
```python
#Checking the dimensions or shape of a DataFrame
df.shape
```

[110]:
```
(100, 24)
```

[111]:
```python
#Removing rows containing any missing values (NaN values)
print(df.dropna(inplace = True))
```

```
None
```

[112]:
```python
#Checking Missing Values
df.isna()
#True: If cell has a missing value (NaN), and False : If it's not a missing␣
 ↪value.
```

[112]:
```
    Product type    SKU  Price  Availability  Number of products sold  \
0          False  False  False         False                    False
1          False  False  False         False                    False
2          False  False  False         False                    False
3          False  False  False         False                    False
4          False  False  False         False                    False
..           …      …      …           …                        …
95         False  False  False         False                    False
96         False  False  False         False                    False
97         False  False  False         False                    False
98         False  False  False         False                    False
99         False  False  False         False                    False

    Revenue generated  Customer demographics  Stock levels  Lead times  \
0               False                  False         False       False
1               False                  False         False       False
2               False                  False         False       False
```

```
3               False                  False         False       False
4               False                  False         False       False
..                …                      …            …          …
95              False                  False         False       False
96              False                  False         False       False
97              False                  False         False       False
98              False                  False         False       False
99              False                  False         False       False

    Order quantities  …  Location  Lead time  Production volumes  \
0              False  …     False      False               False
1              False  …     False      False               False
2              False  …     False      False               False
3              False  …     False      False               False
4              False  …     False      False               False
..               …   …        …        …                   …
95             False  …     False      False               False
96             False  …     False      False               False
97             False  …     False      False               False
98             False  …     False      False               False
99             False  …     False      False               False

    Manufacturing lead time  Manufacturing costs  Inspection results  \
0                     False                False               False
1                     False                False               False
2                     False                False               False
3                     False                False               False
4                     False                False               False
..                      …                    …                 …
95                    False                False               False
96                    False                False               False
97                    False                False               False
98                    False                False               False
99                    False                False               False

    Defect rates  Transportation modes  Routes  Costs
0          False                 False   False  False
1          False                 False   False  False
2          False                 False   False  False
3          False                 False   False  False
4          False                 False   False  False
..           …                    …       …      …
95         False                 False   False  False
96         False                 False   False  False
97         False                 False   False  False
98         False                 False   False  False
99         False                 False   False  False
```

```
[100 rows x 24 columns]
```

[113]:
```
#Count of number of missing(Nan) values in each column
df.isna().sum()
```

[113]:
```
Product type               0
SKU                        0
Price                      0
Availability               0
Number of products sold    0
Revenue generated          0
Customer demographics      0
Stock levels               0
Lead times                 0
Order quantities           0
Shipping times             0
Shipping carriers          0
Shipping costs             0
Supplier name              0
Location                   0
Lead time                  0
Production volumes         0
Manufacturing lead time    0
Manufacturing costs        0
Inspection results         0
Defect rates               0
Transportation modes       0
Routes                     0
Costs                      0
dtype: int64
```

[114]:
```
#Creating a Visualization on Price of Products and Revenue Generated by them:
plt.figure(figsize = (10,10))
sns.lmplot(x='Price', y='Revenue generated', data=df, hue='Product type',␣
 ↪ci=None, palette = 'magma')
plt.title("Price of Products and Revenue Generated", fontweight = 'bold')
plt.show()
```

```
<Figure size 1000x1000 with 0 Axes>
```

CONCLUSIONS

1.Skincare Products: There is a positive correlation between the price of skincare products and the revenue generated. As the price of skincare products increases, the revenue generated also increases significantly. This indicates that customers might be willing to spend more on skincare products, leading to a substantial increase in revenue for the company.

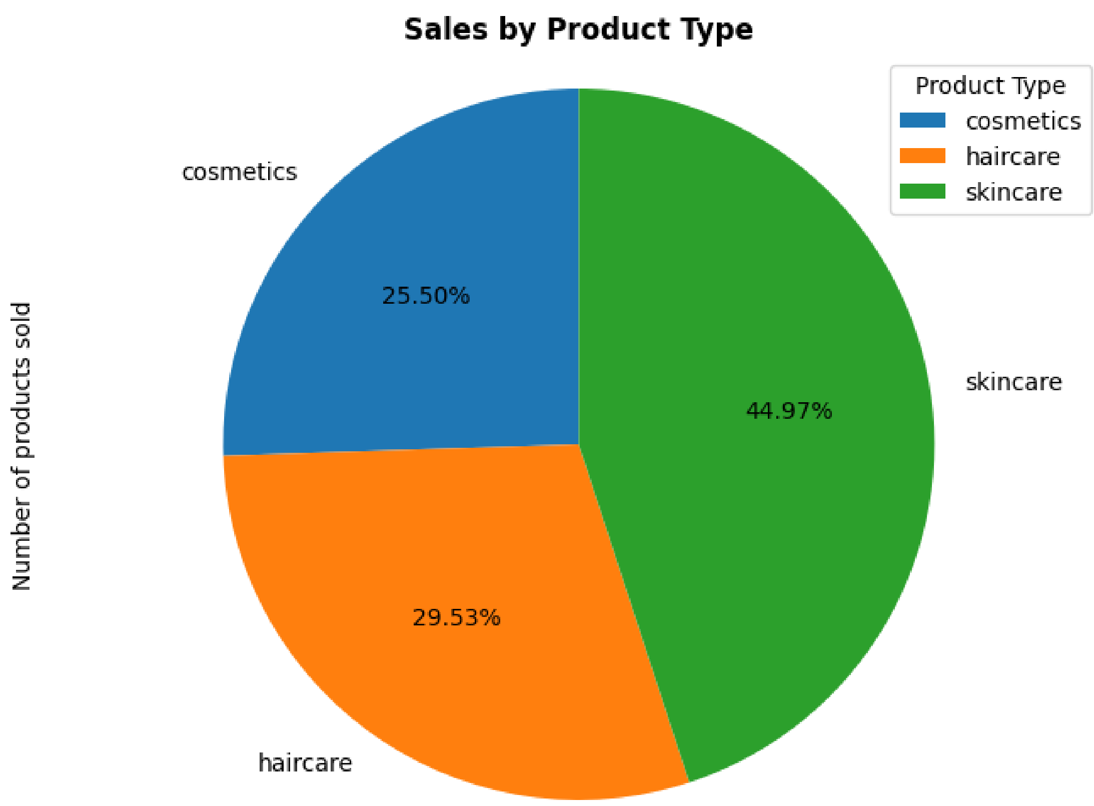
2.Cosmetics Products: There is a negative correlation between the price of cosmetics products and the revenue generated. As the price of cosmetics products increases, the revenue generated tends to decrease. This suggests that customers might be more price-sensitive when it comes to cosmetics products, and higher prices could lead to reduced sales and lower revenue.

3.Hair Care Products: Similar to cosmetics products, there is a negative correlation between the price of hair care products and the revenue generated. However, the negative impact of price on revenue is less pronounced compared to cosmetics. While higher prices may still affect revenue negatively, the slope of the relationship is minor, indicating that the effect is not as strong as for cosmetics products.'

[18]:
```
# Grouping the DataFrame by 'Product type' and calculate the sum of 'Number of
 ↪products sold'
sales_data = df.groupby('Product type')['Number of products sold'].sum()
print(sales_data)
```

```
Product type
cosmetics    11757
haircare     13611
skincare     20731
Name: Number of products sold, dtype: int64
```

[24]:
```
#Creating Visualization Based on Sales Data
plt.figure(figsize = (8,6))
sales_data.plot(kind='pie', autopct='%0.2f%%', startangle = 90)
plt.axis('equal')
plt.legend(sales_data.index , title='Product Type', loc='upper right')
plt.title("Sales by Product Type", fontweight = 'bold')
plt.show()
```



CONCLUSIONS 1. **Skincare** products have the highest sales share, accounting for approximately 44.97% of the total products sold. This corresponds to a total of 20,731 product items sold.

2. **Haircare** products constitute approximately 29.53% of the total product sales, with a total of 13,611 product items sold.

3. **Cosmetics** make up approximately 25.5% of the total product sales, amounting to 11,757 product items sold.

[27]:
```
df.head()
```

[27]:
```
  Product type   SKU      Price  Availability  Number of products sold  \
0     haircare  SKU0  69.808006            55                      802   
1     skincare  SKU1  14.843523            95                      736   
2     haircare  SKU2  11.319683            34                        8   
3     skincare  SKU3  61.163343            68                       83   
4     skincare  SKU4   4.805496            26                      871   

   Revenue generated Customer demographics  Stock levels  Lead times  \
0        8661.996792            Non-binary            58           7   
1        7460.900065                Female            53          30   
2        9577.749626               Unknown             1          10   
3        7766.836426            Non-binary            23          13   
4        2686.505152            Non-binary             5           3   

   Order quantities  …  Location Lead time  Production volumes  \
0                96  …    Mumbai        29                 215   
1                37  …    Mumbai        23                 517   
2                88  …    Mumbai        12                 971   
3                59  …   Kolkata        24                 937   
4                56  …     Delhi         5                 414   

  Manufacturing lead time Manufacturing costs  Inspection results  \
0                      29           46.279879             Pending   
1                      30           33.616769             Pending   
2                      27           30.688019             Pending   
3                      18           35.624741                Fail   
4                       3           92.065161                Fail   

   Defect rates  Transportation modes   Routes       Costs  
0      0.226410                  Road  Route B  187.752075  
1      4.854068                  Road  Route B  503.065579  
2      4.580593                   Air  Route C  141.920282  
3      4.746649                  Rail  Route A  254.776159  
4      3.145580                   Air  Route A  923.440632  

[5 rows x 24 columns]
```

[37]:
```python
#Groupby operation for calculating the sum of the "Revenue generated" for each␣
↪"Shipping carriers"
total_revenue = df.groupby('Shipping carriers')['Revenue generated'].sum().
↪reset_index()
print(total_revenue)
```

```
  Shipping carriers  Revenue generated
0         Carrier A      142629.994607
1         Carrier B      250094.646988
2         Carrier C      184880.177143
```

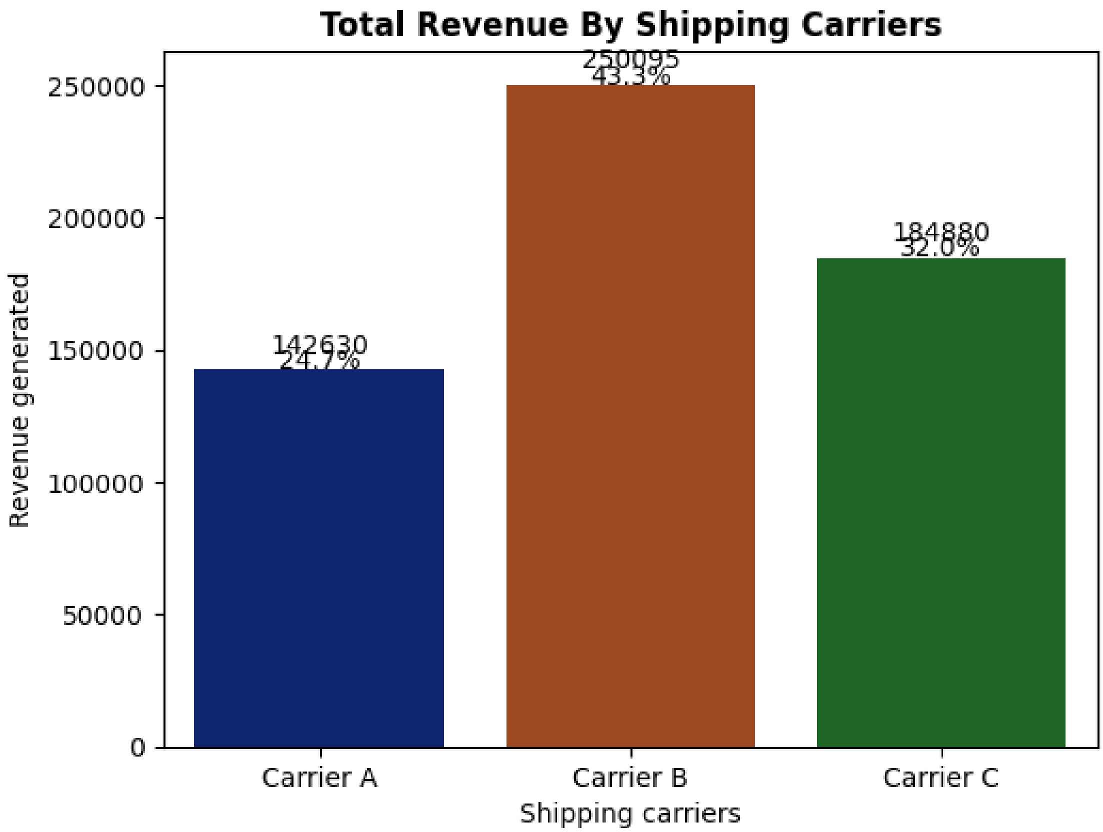
[38]:
```python
# Calculate the total generated revenue across all "Shipping carriers"
total_generated_revenue = total_revenue['Revenue generated'].sum()
print(total_generated_revenue)
```

```
577604.8187380086
```

[54]:
```python
# Calculate the percentage of revenue generated for each "Shipping carriers"
total_revenue['Percentage'] = (total_revenue['Revenue generated'] /␣
↪total_generated_revenue) * 100
print(total_revenue[['Shipping carriers', 'Percentage']])
```

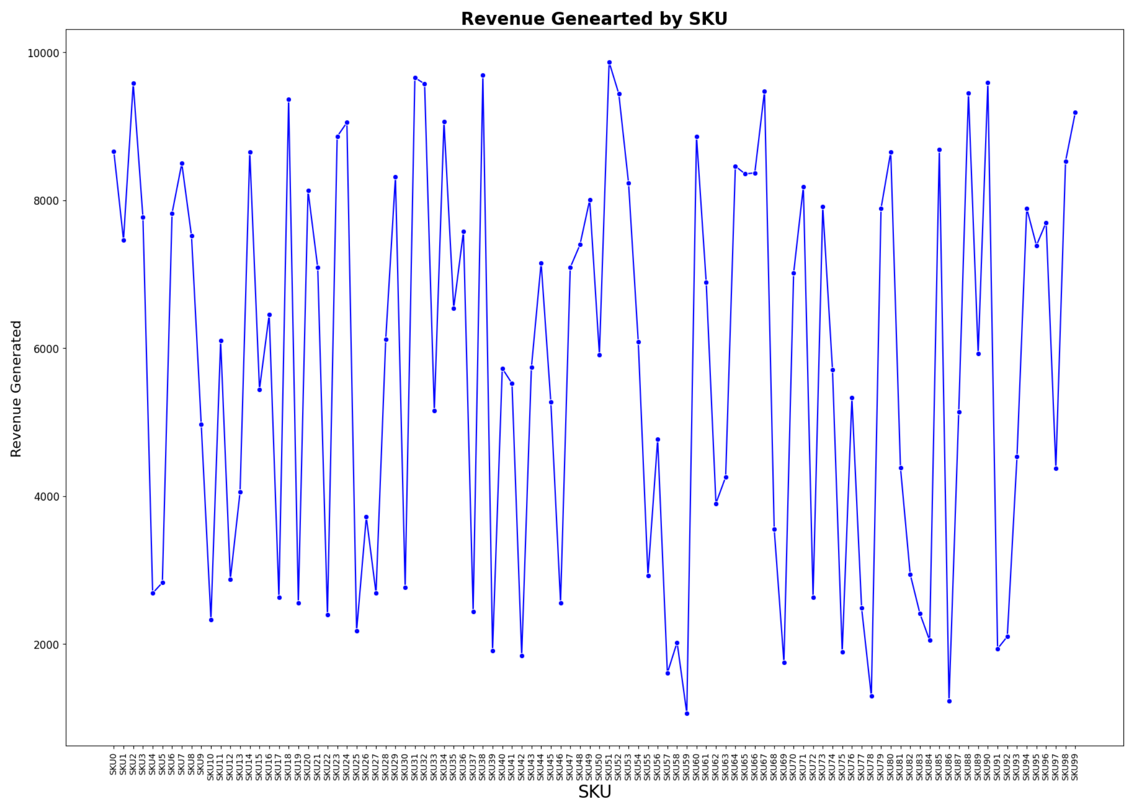
```
  Shipping carriers  Percentage
0         Carrier A   24.693353
1         Carrier B   43.298574
2         Carrier C   32.008074
```

[40]:
```python
# Creating Visualization on Revenue Generated through Shipping Carriers
fig = sns.barplot(x=total_revenue['Shipping carriers'],␣
↪y=total_revenue['Revenue generated'], palette='dark')

# Add count labels on top of each bar
for bars in fig.containers:
    fig.bar_label(bars, label_type='edge', fontsize=10, padding=4)

# Add percentage labels on top of each bar
for index, value in enumerate(total_revenue['Percentage']):
    fig.text(index, total_revenue['Revenue generated'][index] + 200, f'{value:.
↪1f}%', color='black', ha="center")

plt.title("Total Revenue By Shipping Carriers", fontweight='bold')
plt.show()
```



CONCLUSIONS: - **Carrier B** has the highest share in generating revenue, accounting for approximately 43.3%. This corresponds to a total revenue of 250,094. - **Carrier C** accounts for approximately 32% of the revenue, with a total revenue of 184,880.2. - **Carrier A** makes up approximately 24.7% of the revenue, amounting to 142,630.'

[41]:
```
#Creating Visualization on Revenue Generated through Shipping Carriers
plt.figure(figsize=(22, 15))
sns.lineplot(x = 'SKU', y = 'Revenue generated', data = df , marker='o',␣
 ↪color='b')
plt.xticks(fontsize = 10, rotation = 90)
plt.yticks(fontsize = 12)
plt.xlabel('SKU', fontsize = 20)
plt.ylabel('Revenue Generated', fontsize = 16)
plt.title('Revenue Genearted by SKU',fontsize = 20, fontweight = 'bold')
plt.show()
```



[57]:
```
#Groupby operation for calculating the sum of the "Shipping costs" for each "Shipping carriers#
shipping_cost = df.groupby('Shipping carriers')['Shipping costs'].sum().reset_index()
print(shipping_cost)
```

```
  Shipping carriers  Shipping costs
0         Carrier A      155.537831
1         Carrier B      236.897620
2         Carrier C      162.379457
```

[58]:
```
# Calculate the total shipping costs across all "Shipping carriers"
total_shipping_costs = shipping_cost['Shipping costs'].sum()
print(total_shipping_costs)
```

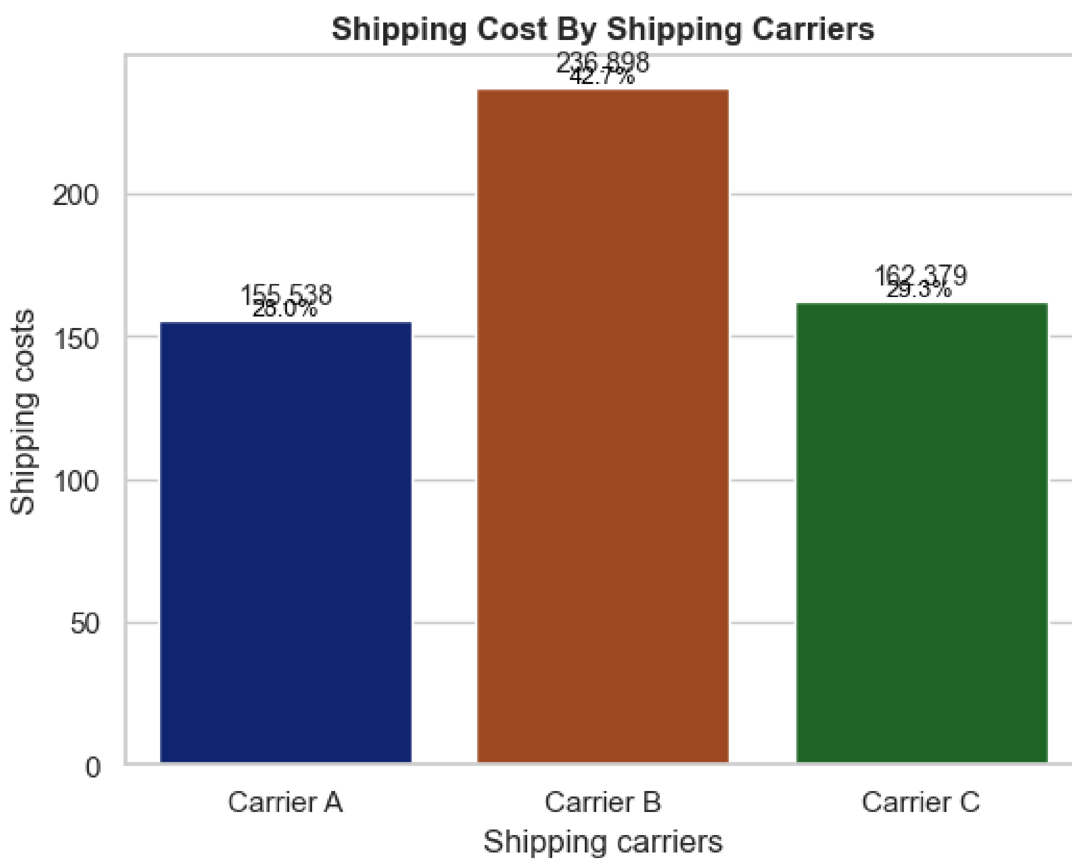
```
554.8149072019587
```

[65]:
```
# Calculate the percentage of shipping costs for each "Shipping carriers"
shipping_cost['Percentage'] = (shipping_cost['Shipping costs'] / total_shipping_costs) * 100
print(shipping_cost[['Shipping carriers', 'Percentage']])
```

```
  Shipping carriers  Percentage
```

```
0         Carrier A   28.034184
1         Carrier B   42.698496
2         Carrier C   29.267320
```

[79]:
```python
# Creating Visualization on Shipping Cost through Shipping Carriers
fig = sns.barplot(x=shipping_cost['Shipping carriers'],␣
 ↪y=shipping_cost['Shipping costs'], palette='dark')
# Add count labels on top of each bar
for bars in fig.containers:
    fig.bar_label(bars, label_type='edge', fontsize=10, padding=4)
# Add percentage labels on top of each bar
for index, value in enumerate(shipping_cost['Percentage']):
    fig.text(index, shipping_cost['Shipping costs'][index] +0, f'{value:.1f}%',␣
 ↪color='black', ha="center", va="bottom",  fontsize=9)
plt.title("Shipping Cost By Shipping Carriers", fontweight='bold')
plt.show()
```



CONCLUSIONS: - **Carrier B** has the highest share 42.7%. This corresponds to a total shipping costs of 236.8. - **Carrier C** accounts for 29.3% of the shipping costs, with a total cost of 162.3. - **Carrier A** makes up 28.7% of the shipping costs, amounting to 155.5.

[88]:
```
#Grouping the "Transportation modes" column and then calculating the sum of the␣
↪"Costs" for each group
transportation_cost = df.groupby('Transportation modes')['Costs'].sum().
↪reset_index()
print(transportation_cost)
```

```
  Transportation modes         Costs
0                  Air  14604.527498
1                 Rail  15168.931559
2                 Road  16048.193639
3                  Sea   7102.925520
```

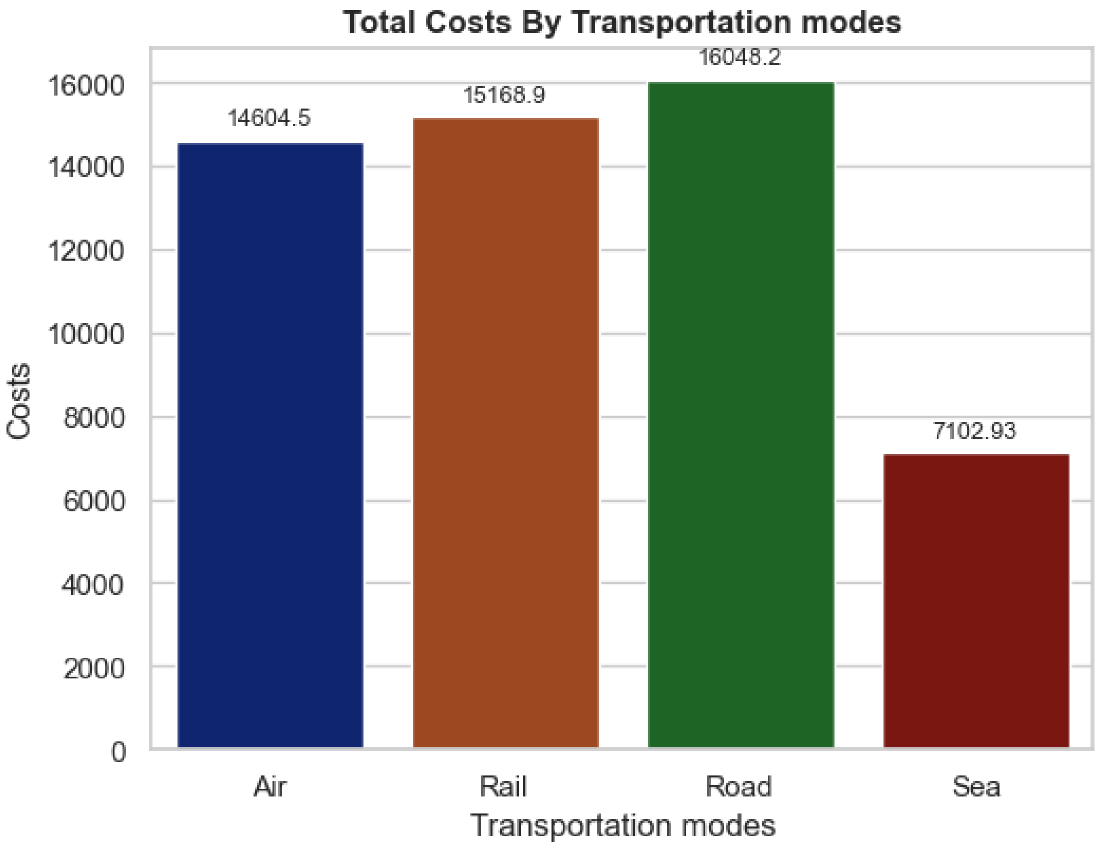
[89]:
```
# Calculate the total transportation costs for Transportation modes
total_transportation_costs = transportation_cost['Costs'].sum()
print(total_transportation_costs)
```

```
52924.57821581411
```

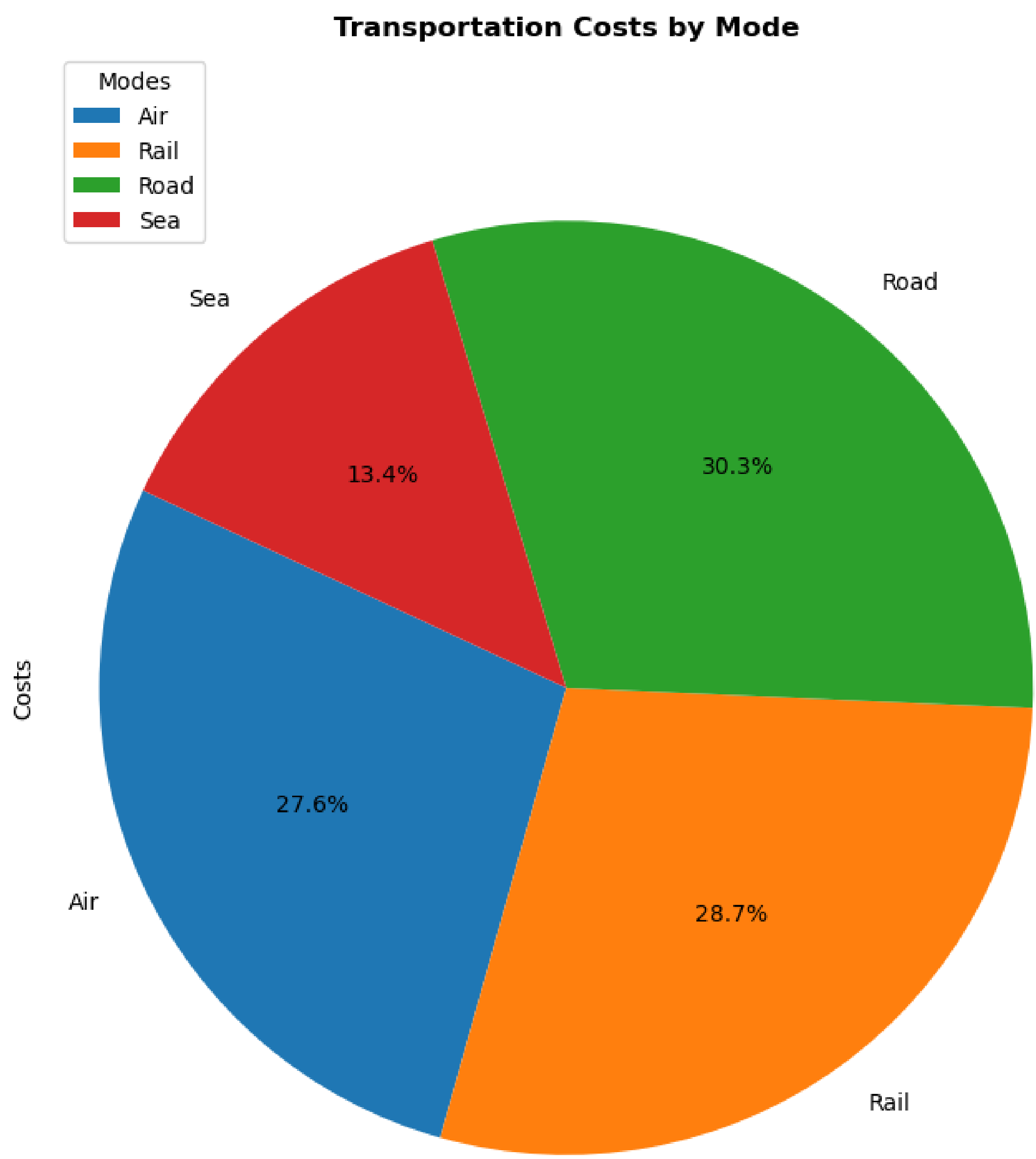
[90]:
```
# Calculate the percentage of transportation mode to costs
transportation_cost['Percentage'] = (transportation_cost['Costs'] /
↪total_transportation_costs) * 100
print(transportation_cost[['Transportation modes', 'Percentage']])
```

```
  Transportation modes  Percentage
0                  Air   27.594981
1                 Rail   28.661412
2                 Road   30.322762
3                  Sea   13.420845
```

[96]:
```
#Creating Visualization on Transportation Costs by differnt Transportation Modes
fig = sns.barplot(x = transportation_cost['Transportation modes'] , y =␣
↪transportation_cost['Costs'], palette = 'dark')
# Add count labels on top of each bar
for bars in fig.containers:
    fig.bar_label(bars, label_type='edge', fontsize=9, padding=4)
plt.title("Total Costs By Transportation modes", fontweight='bold')
plt.show()
```



[34]:
```
# Plotting the pie chart for transportation_cost
plt.figure(figsize=(8, 10))
transportation_cost.plot(kind='pie', autopct='%0.1f%%', startangle=155)
plt.axis('equal')
plt.legend(transportation_cost.index, title='Modes', loc='upper left')
plt.title("Transportation Costs by Mode", fontweight='bold')
plt.show()
```



CONCLUSIONS: Based on the aggregated data visualization, the transportation modes can be ranked by their cost as follows: - **Road:** 30.3% (16,048.2) - **Rail:** 28.7% (15,168.9) - **Air:** 27.6% (14,604.5) - **Sea:** 13.4% (7,102.93)

[98]:
```
df.head()
```

[98]:

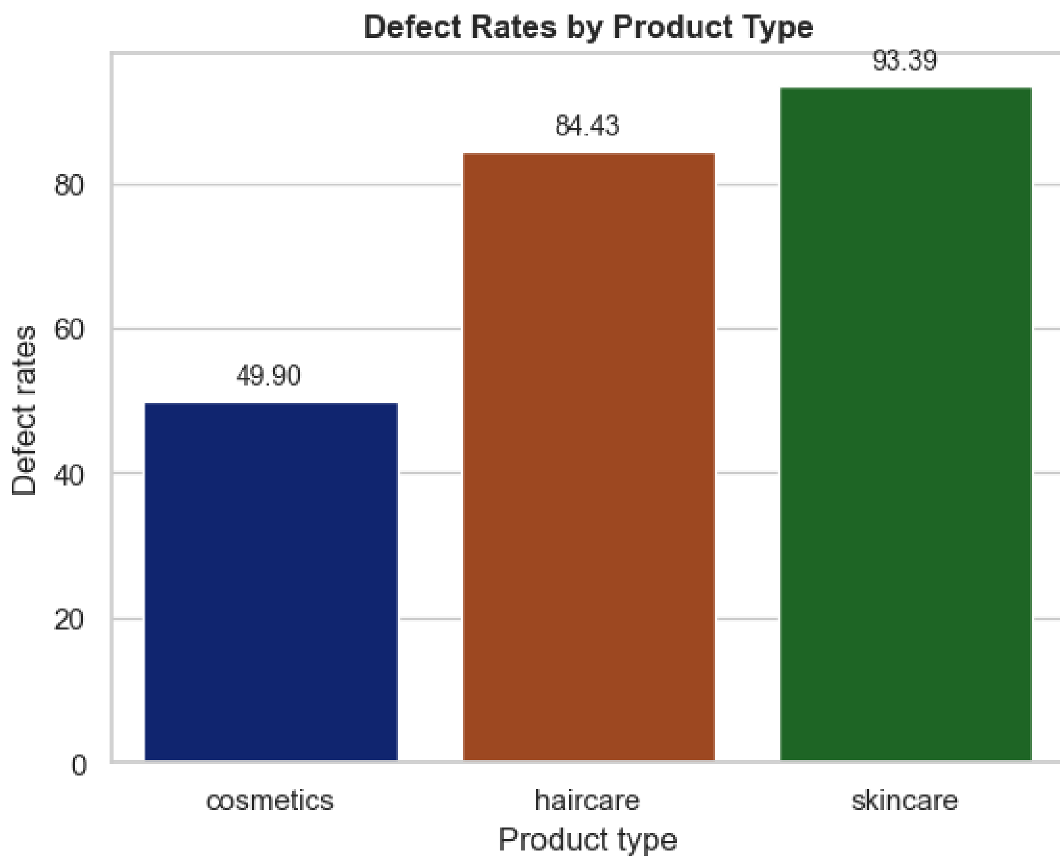
```
  Product type   SKU      Price  Availability  Number of products sold  \
0     haircare  SKU0  69.808006            55                      802
1     skincare  SKU1  14.843523            95                      736
2     haircare  SKU2  11.319683            34                        8
3     skincare  SKU3  61.163343            68                       83
4     skincare  SKU4   4.805496            26                      871

   Revenue generated Customer demographics  Stock levels  Lead times  \
0        8661.996792            Non-binary            58           7
1        7460.900065                Female            53          30
2        9577.749626               Unknown             1          10
3        7766.836426            Non-binary            23          13
4        2686.505152            Non-binary             5           3

   Order quantities  …  Location Lead time  Production volumes  \
0                96  …    Mumbai        29                 215
1                37  …    Mumbai        23                 517
2                88  …    Mumbai        12                 971
3                59  …   Kolkata        24                 937
4                56  …     Delhi         5                 414

  Manufacturing lead time Manufacturing costs  Inspection results  \
0                      29           46.279879             Pending
1                      30           33.616769             Pending
2                      27           30.688019             Pending
3                      18           35.624741                Fail
4                       3           92.065161                Fail

   Defect rates  Transportation modes   Routes        Costs
0      0.226410                  Road  Route B   187.752075
1      4.854068                  Road  Route B   503.065579
2      4.580593                   Air  Route C   141.920282
3      4.746649                  Rail  Route A   254.776159
4      3.145580                   Air  Route A   923.440632

[5 rows x 24 columns]
```

[99]:

```python
#Grouping the "Product type" column and then calculating the sum of "Defect␣
↪rates" for each group
Total_Defect_Rates = df.groupby('Product type')['Defect rates'].sum().
↪reset_index()
print(Total_Defect_Rates)
```

```
  Product type  Defect rates
0    cosmetics     49.901461
1     haircare     84.427107
2     skincare     93.387231
```

[101]:
```
#Creating Visualization on Defect Rates by Product Type
fig = sns.barplot(x = Total_Defect_Rates['Product type'] , y =␣
 ↪Total_Defect_Rates['Defect rates'], palette = 'dark')
# Add count labels on top of each bar
for bars in fig.containers:
    fig.bar_label(bars, label_type='edge', fontsize=10, padding=4, fmt='%.2f')
plt.title("Defect Rates by Product Type", fontweight = 'bold')
plt.show()
```



CONCLUSIONS: Based on the bar plot visualization, it is evident that: - **Skincare** products have the highest defect rate value (93.39), - **Haircare** products in the second position (84.43), - **Cosmetics** with the lowest defect rate (49.90).

SUMMARY OF CONCLUSSIONS AND FINDINGS

1. Analyzing Revenue by Product Type:

a). Skincare Products: There is a positive correlation between the price of skincare products and the revenue generated. As the price of skincare products increases, the revenue generated also increases significantly. This indicates that customers might be willing to spend more on skincare products, leading to a substantial increase in revenue for the company.

b). Cosmetics Products: There is a negative correlation between the price of cosmetics products and the revenue generated. As the price of cosmetics products increases, the revenue generated tends to decrease. This suggests that customers might be more price-sensitive when it comes to cosmetics products, and higher prices could lead to reduced sales and lower revenue.

c). Hair Care Products: Similar to cosmetics products, there is a negative correlation between the price of hair care products and the revenue generated. However, the negative impact of price on revenue is less pronounced compared to cosmetics. While higher prices may still affect revenue negatively, the slope of the relationship is minor, indicating that the effect is not as strong as for cosmetics products.

2. Unraveling Sales by Product Type:

a). Skincare products have the highest sales share, accounting for approximately 44.97% of the total products sold. This corresponds to a total of 20,731 product items sold.

b). Haircare products constitute approximately 29.53% of the total product sales, with a total of 13,611 product items sold.

c). Cosmetics make up approximately 25.5% of the total product sales, amounting to 11,757 product items sold.

3. Exploring Revenue from Shipping Carriers:

a). Carrier B has the highest share in generating revenue, accounting for approximately 43.3%. This corresponds to a total revenue of 250,094.

b). Carrier C accounts for approximately 32% of the revenue, with a total revenue of 184,880.2.

c). Carrier A makes up approximately 24.7% of the revenue, amounting to 142,630.'.

4. Analyzing Revenue by SKU (Stock Keeping Unit): - Examining SKU-specific data to gain insights into the performance of individual products. - Empowering businesses to optimize their inventory strategies based on SKU performance.

5. Understanding Shipping Costs of Carriers:

a). Carrier B has the highest share 42.7%. This corresponds to a total shipping costs of 236.8.

b). Carrier C accounts for 29.3% of the shipping costs, with a total cost of 162.3.

c). Carrier A makes up 28.7% of the shipping costs, amounting to 155.5.

6. Cost Distribution by Transportation Modes:

Conclusions based on the aggregated data visualization, the transportation modes can be ranked by their cost as follows:

a). Road: 30.3% (16,048.2)

b). Rail: 28.7% (15,168.9)

c). Air: 27.6% (14,604.5)

d). Sea: 13.4% (7,102.93)

7. Assessing Defect Rates during Shipping:

Conclusions based on the bar plot visualization, it is evident that:

a). **Skincare** products have the highest defect rate value (93.39),

b). **Haircare** products in the second position (84.43),

c). **Cosmetics** with the lowest defect rate (49.90).