

Python Web Scraping and Analysis of eBay Website (Laptop Section)



At Kapruka Global Shop via EBAY



DHRUV DATTA

045017 BDA

Project Title: eBay Laptops Section Analysis

Objective 1: Price-Review Relationship Analysis

Objective: Investigate the interdependence between laptop prices and customer reviews.

Key Metrics:

Examine how changes in laptop prices affect customer review sentiment and rating.

Identify price points that optimize positive reviews and ratings.

Objective 2: Shipping Cost and Price Correlation

Objective: Assess the relationship between shipping costs and laptop prices to enhance customer satisfaction.

Key Metrics:

DHRUV DATTA

045017 BDA

Analyze how variations in shipping costs influence laptop listing prices.

Determine if lower shipping costs lead to improved customer satisfaction and increased sales.

Objective 3: Review Count and Rating Monitoring

Objective: Monitor and manage customer reviews and ratings for laptop listings.

Key Metrics:

Continuously track the volume of customer reviews for each laptop listing.

Analyze the distribution of review ratings, including 5-star, 4-star, etc.



INTRODUCTION

Web Scraping, also known as web harvesting or web data extraction, is used to extract data from websites. Although it can be done manually by a user, it is generally done by a bot or an AI controlled web crawler. In this data extraction information of similar types is extracted from a web site and is stored in a spreadsheet or a central local database, which is later used for analysis.

Webpages are build using text-based mark-up language, i.e. HTML and XHTML, which generally contains useful data in text format.

However, as we see these pages do not display

DHRUV DATTA

045017 BDA

the HTML directly and are made for the end user and not made for automated use. As a result, the web scrapers come into play and make our lives easier.

MOTIVATION

There have been times when scrolling through websites is much more time consuming and there are things which are rather easier if we had a list to go through rather than scrolling through webpages. I personally faced this problem when I was myself searching to buy a mobile phone for myself. I spent endless times checking different products. With the help of web scraping, I can easily get the products I wish along with its rating and URL available to us the very same instant on a single page.

DHRUV DATTA

045017 BDA

TECHNIQUES

Web Scraping is the process of automatically mining data or collecting information from the World Wide Web. There are methods that some websites use to prevent web scraping, such as detecting and disallowing bots from crawling (viewing) their pages. In response, there are web scraping systems that rely on using techniques such as DOMS(Document Object Model), computer vision and natural language processing to simulate human browsing to enable gathering web page content for offline parsing. Current web scraping solutions range from the ad-hoc, requiring human effort, to fully automated systems that can convert entire websites into structured information, with limitations.

PROCEDURE

DHRUV DATTA

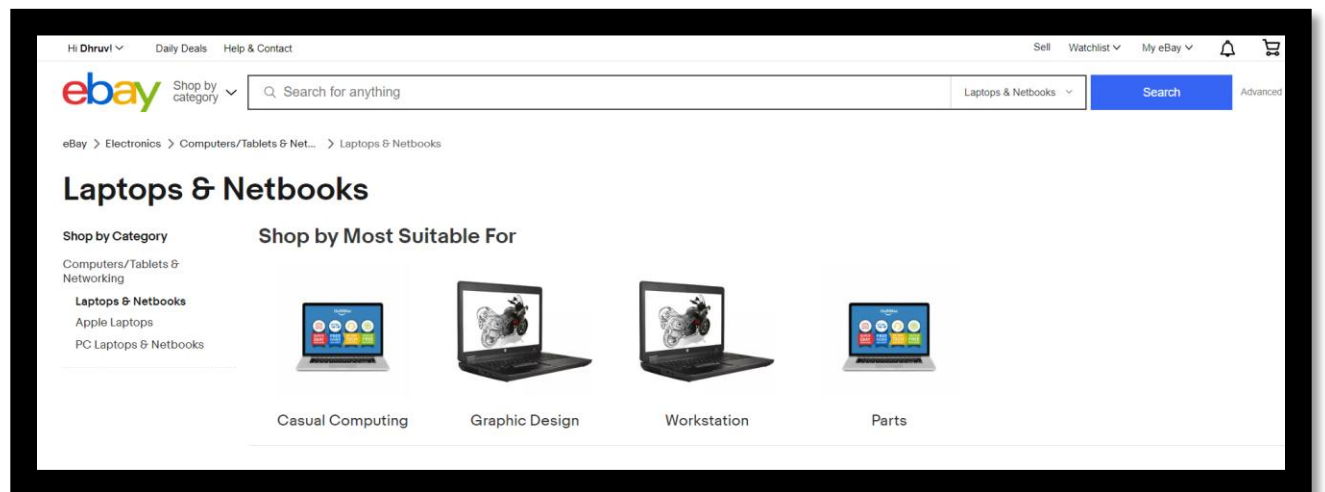
045017 BDA

The library of codes we are going to use for this project are:

- Requests Library
- BeautifulSoup Library
- Pandas
- Autoviz

The use of each of these will be explained with the code

Web Scrapping from a Searched eBay product

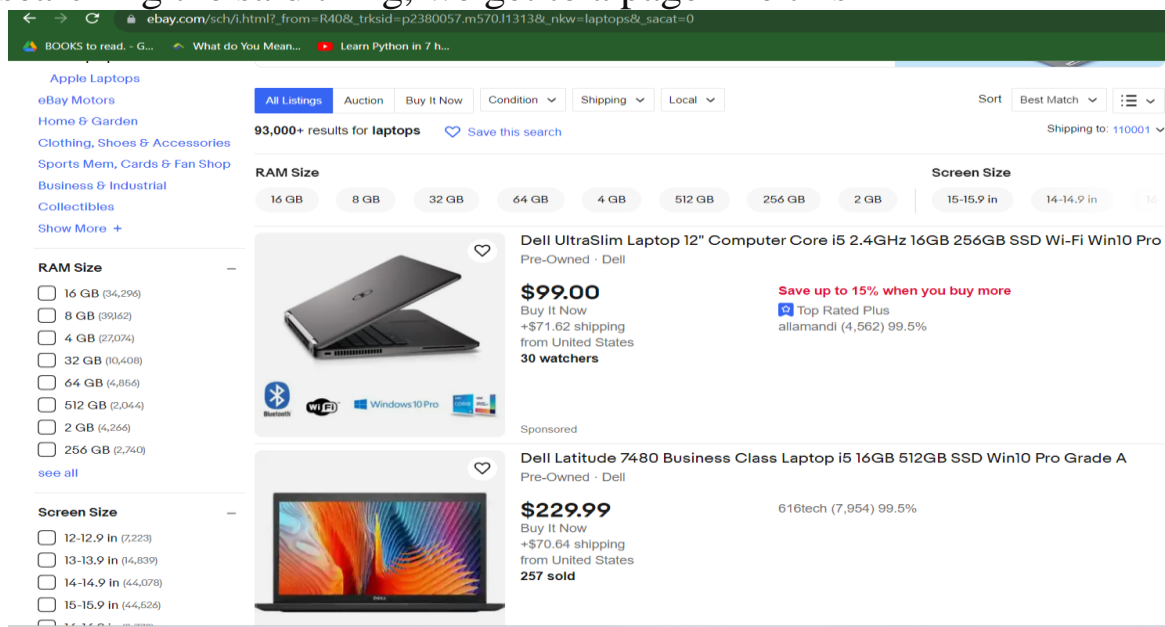


DHRUV DATTA

045017 BDA

When we go to the homepage of eBay, we see a page similar to what we see above. On the very top centre of the page, we can see the search handle as highlighted in the above image.

Next, we search for the products we wish to scrape the data for, in this case we need variety, so I search for mobile phones under 20000, which gives us a plenty of data to be analyzed and scraped from. After searching the said thing, we get to a page like this

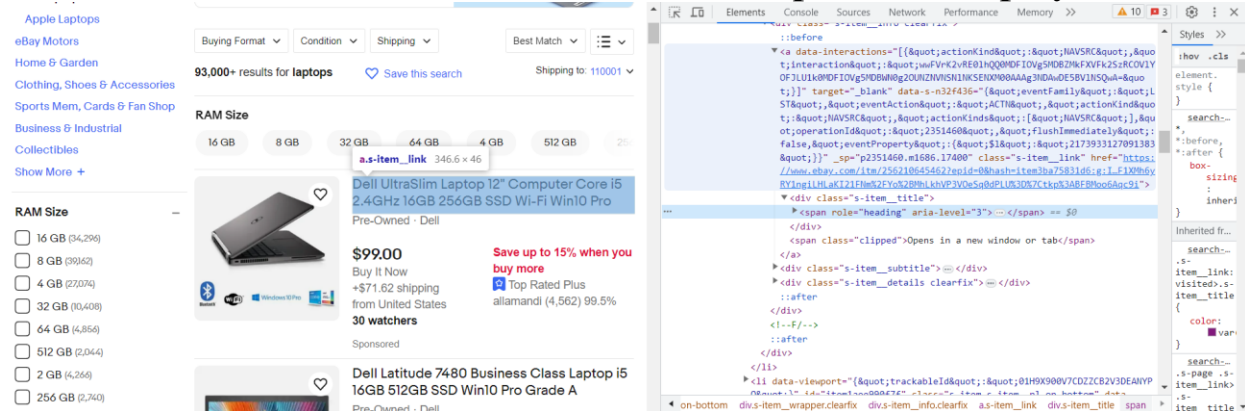


Here we see that there are similar types of data such as the name of the product, the price of the product, number of reviews and ratings and the number of stars received by the product. Using the inspector, we will see if the similar type of data has similar HTML classes too, by which we

DHRUV DATTA

045017 BDA

can easily select the classes we want to extract from this webpage. We will check the classes of the names of the two products displayed above



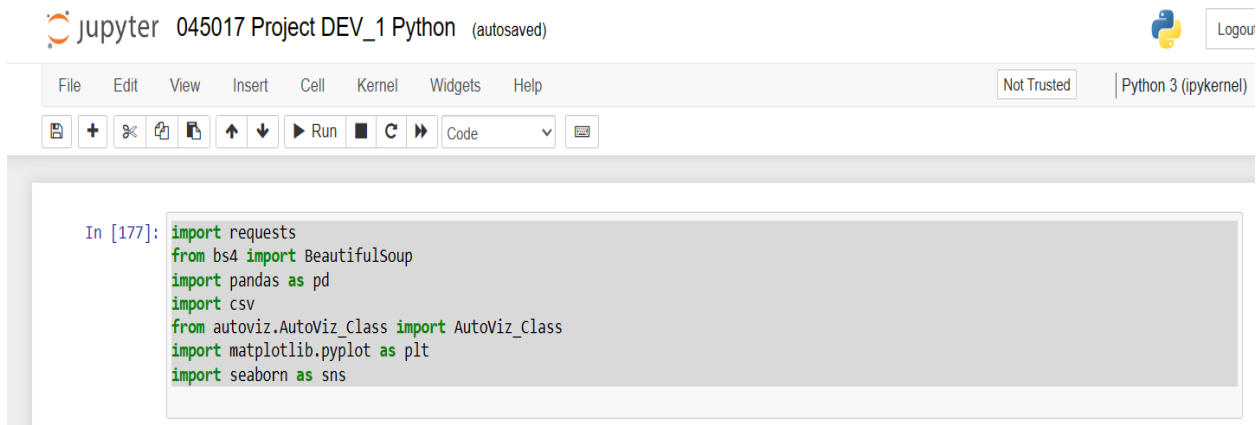
If we notice the highlighted area, we can see that the classes under which the names of both the product's name is given, is the same. Similarly, we check all the other data that we wish to extract and then proceed to building the web scraper

For this project I've used Jupyter Notebook because data because they are great for showcasing my work. We can see the code and its result together. We can run cell by cell for better understanding and we can also insert markdown which is better than normal comments because we can use things like bullet points and makes the project look organized.

First, we install the Requests and BeautifulSoup library and store the URL of the webpage in a variable 'products_url' and requests we GET from the webpage discussed above.

DHRUV DATTA

045017 BDA



```
In [177]: import requests
from bs4 import BeautifulSoup
import pandas as pd
import csv
from autoviz.AutoViz_Class import AutoViz_Class
import matplotlib.pyplot as plt
import seaborn as sns
```

After examining the HTML of the webpage we choose the classes of the data we wish to extract that is name of the product, price of the product, the stars received by the product, the number of ratings and reviews received by the product and the URL of the product (the URL is not a separate entity that is displayed on the webpage like the other data that is extracted). In the name on inspecting, we see that it is stored in a <div> so using BeautifulSoup we use the doc we created above and use the findAll function. The findAll function takes 2 arguments i.e., the tag the data is stored under (in the case of the name of the product it is stored under the <div>) and the name of the class of the data we wish to extract (in this case which we have stored under 'heading_class') and store the data received in a list called 'heading_tags'.

After getting all the heading tags we analyze the data we got (since there is a lot of data under the same class, we only check the first four because the structure of the rest will be same)

DHRUV DATTA

045017 BDA

```
Jupyter 045017 Project DEV_1 Python (autosaved) Python 3 (ipykernel)
File Edit View Insert Cell Kernel Widgets Help Not Trusted
In [180]: lap_name = []
for n in soupobj.findAll('div',{'class':'s-item__title'}):
    lap_name.append(n.text.strip())

offer = []
for o in soupobj.findAll('div',{'class':'s-item__detail s-item__detail--primary'}):
    offer.append(o.text.strip())

Condition = []
for c in soupobj.findAll('div',{'class':'s-item__subtitle'}):
    Condition.append(c.text.strip())

Price = []
for p in soupobj.findAll('span',{'class':'s-item__price'}):
    Price.append(p.text.strip())

Shipping = []
for sh in soupobj.findAll('span',{'class':'s-item__shipping s-item__logisticsCost'}):
    Shipping.append(sh.text.strip())

Status = []
for s in soupobj.findAll('span',{'class':'BOLD'}):
    Status.append(s.text.strip())

Rating = []
for ra in soupobj.findAll('div',{'class':'x-star-rating'}):
    Rating.append(ra.text.strip())

Review_Count = []
for rc in soupobj.findAll('span',{'class':'s-item__reviews-count'}):
    Review_Count.append(rc.text.strip())

Review_Detail = []
for rd in soupobj.findAll('span',{'class':'s-item__reviews-count'}):
    Review_Detail.append(rd.text.strip())
```

This Data Frame presents a comprehensive overview of Laptops available in the market, with the data collected from web scraping. It includes detailed information on each Laptop's product name, current price, offers, customer reviews, the number of ratings and reviews, Shipping, Status, Condition, Review count and detail.

Column Descriptions:

- Product Name: The name of the Laptop model.
- Current Price: The current selling price of the Laptop.
- Offers: The Offers which are being offered by various dealers
- Condition: If the product is brand new or pre-owned

DHRUV DATTA

045017 BDA

- Reviews: The average customer rating for the laptops.
- Shipping: The shipping cost per product.
- Status: how many laptops are getting sold or not

Description: A concise description of the mobile phone's specifications and features. This DataFrame serves as a valuable resource for market analysis, price comparison, and understanding customer sentiment towards various mobile phone models

```
dict = {'Laptop Names':lap_name,'Offers':offer,'Condition':Condition,
        'Price':Price,'Shipping':Shipping,'Status':Status,'Rating':Rating,
        'Review Count':Review_Count,'Review Detail':Review_Detail}
```

```
In [181]: Data = pd.DataFrame.from_dict(dict, orient='index')
          Data = Data.transpose()
```

```
In [182]: Data.head(50)
```

	Laptop Names	Offers	Condition	Price	Shipping	Status	Rating	Review Count	Review Detail
0	Shop on eBay	\$20.00	Brand New	\$20.00	+\$70.64 shipping	91,000	4.5 out of 5 stars.	20 product ratings - Asus Chromebook C202 11.6...	20 product ratings - Asus Chromebook C202 11.6...
1	Dell Latitude 7490 Business Class Laptop i5 16...	or Best Offer	Pre-Owned - Dell	\$229.99	+\$112.29 shipping	laptops	4.5 out of 5 stars.	89 product ratings - Acer Chromebook C720 11.6...	89 product ratings - Acer Chromebook C720 11.6...
2	Cheap Dell HP Lenovo Fujitsu Windows 10 Laptop...	Sponsored	Pre-Owned - Lenovo	\$87.33	+\$61.79 shipping	255 sold	5.0 out of 5 stars.	2 product ratings - Lenovo Ideapad Gaming 3 15...	2 product ratings - Lenovo Ideapad Gaming 3 15...
	Asus							1 product...	1 product...

```
In [183]: data1 = Data
```

DHRUV DATTA

045017 BDA

```

In [183]: data1 = Data

In [184]: data1['Price'] = data1.Price.str.extract('(\\d+)')
data1['Shipping'] = data1.Shipping.str.extract('(\\d+)')
data1['Rating'] = data1['Rating'].apply(lambda x: x.split(' ')[0] if isinstance(x, str) else x)
data1['Review Count'] = data1['Review Count'].apply(lambda x: x.split(' ')[0] if isinstance(x, str) else x)

In [185]: data1["Price"] = pd.to_numeric(data1["Price"])
data1["Shipping"] = pd.to_numeric(data1["Shipping"])
data1["Rating"] = pd.to_numeric(data1["Rating"])
data1["Review Count"] = pd.to_numeric(data1["Review Count"])

In [186]: data1.head()
Out[186]:

```

	Laptop Names	Offers	Condition	Price	Shipping	Status	Rating	Review Count	Review Detail
0	Shop on eBay	\$20.00	Brand New	20.0	70.0	91,000	4.5	20.0	20 product ratings - Asus Chromebook C202 11.6...
1	Dell Latitude 7490 Business Class Laptop i5 16...	or Best Offer	Pre-Owned · Dell	229.0	112.0	laptops	4.5	89.0	89 product ratings - Acer Chromebook C720 11.6...
2	Cheap Dell HP Lenovo Fujitsu Windows 10 Laptop...	Sponsored	Pre-Owned · Lenovo	87.0	61.0	255 sold	5.0	2.0	2 product ratings - Lenovo Ideapad Gaming 3 15...
3	Asus Chromebook C202 11.6" Intel 1.6 GHz 4GB R...	\$229.99	TOP QUALITY FAST SHIPPING	39.0	77.0	129 sold	5.0	1.0	1 product rating - Lenovo LOQ 15 15.6" Gaming ...
4	Dell Latitude Laptop Computer 13.3" Core i7 16...	Buy It Now	Pre-Owned · ASUS · 16 GB	358.0	66.0	Asus Chromebook C202 11.6" Intel 1.6 GHz 4GB R...	5.0	1.0	1 product rating - DELL LATITUDE 3500 Laptop I...

```

In [187]: data1.to_csv('Ebay Data.csv', index=False, header=True)

```

Analysis:

So analysis starts with knowing how to read a csv file which is first done by importing pandas and numpy. They help in working with datasets and gathering information out of them. Then I used “pfd.read_csv(‘file name’)” to read the csv file of used in extracting eBay Website. After that I used head function to get the first few values of the dataset as loading all the values would not be feasible and it is actually not required. Following this it is important that we know how big our dataset is. This would help us get a basic idea on what we are working on and this is achieved by using the “shape ()” function. After getting this brief idea about the dataset I tried to get a count of the number of cars differentiated with respect to fuel type. This was done by the use of “value counts()” function

DHRUV DATTA

045017 BDA

VARIOUS WAYS OF ANALYSING THE DATA:-

1. Scatter Plot: *A scatter plot is a graphical representation of data points in a two-dimensional space, typically on a Cartesian plane. It is used to visualize the relationship or correlation between two variables. Each data point on the plot represents the value of one variable on the x-axis and the value of another variable on the y-axis.*

- Price and Shipping Cost Relationship:

Inference: There appears to be a positive correlation between price and shipping cost within the specified range.

Managerial Action: Consider adjusting shipping cost strategies based on price points. For higher-priced items, customers may be more willing to accept slightly higher shipping costs, while for lower-priced items, competitive shipping rates may be more critical.

- Density Concentration:

Inference: The scatter plot density is concentrated in the \$50-\$150 shipping cost range and \$0-\$300 price range.

Managerial Action: Focus marketing efforts and promotional strategies on laptops within this specific range, as they likely represent a significant portion of your sales. Offer bundle deals, discounts, or shipping incentives to attract more customers.

- Price Sensitivity:

Inference: Within the \$0-\$300 price range, examine if there are clusters of data points at certain price points.

Managerial Action: Identify the price points where customer demand is highest and consider expanding product offerings or featuring laptops at those price points prominently.

- Outliers:

Inference: The scatter plot now clearly shows the presence of an outlier with a \$900 price and \$310 shipping cost.

DHRUV DATTA

045017 BDA

Managerial Action: Scrutinize this outlier's performance, customer reviews, and sales figures. If it's a high-value product that attracts a specific customer segment, consider promoting it separately and optimizing its shipping cost to improve profitability

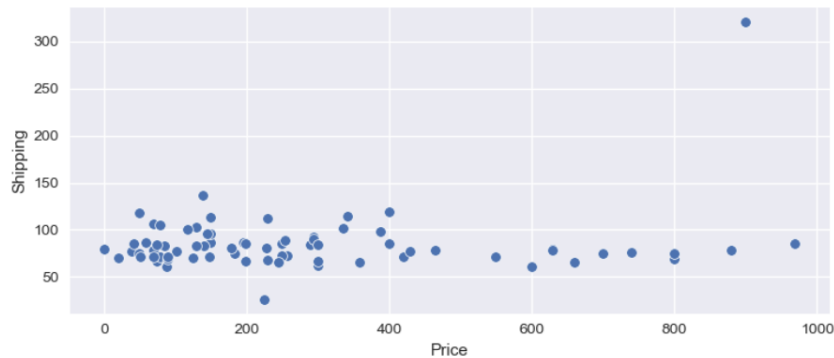
- Competitive Analysis:

Inference: Compare your scatter plot data with competitors to see if your price and shipping cost relationships align with market trends.

Managerial Action: If competitors offer similar products with lower shipping costs or more attractive pricing, consider adjusting your pricing and shipping strategies to remain competitive.

```
No categorical or boolean vars in data set, hence no bar charts.  
Nothing to add Plot not being added  
Time to run AutoViz (in seconds) = 0.170
```

Pair-wise Scatter Plot of all Continuous Variables



Histograms (KDE plots) of all Continuous Variables

2. Heatmap of all continuous

variables: A heat map is a two-dimensional representation of data in which various values are represented by colors. A simple heat map provides an immediate visual summary of information across two axes, allowing users to quickly grasp the most important or relevant data points. More elaborate heat maps allow the viewer to understand complex data sets

DHRUV DATTA

045017 BDA

Price and Shipping Cost Proportionality:

- Insight: The proportionate relationship between price and shipping cost indicates that as the price of laptops increases, the shipping cost also tends to rise proportionally.
- Managerial Implication: Consider optimizing shipping cost strategies to align with price points. For higher-priced laptops, customers may be more accepting of slightly higher shipping costs, while for lower-priced laptops, competitive shipping rates may be crucial to maintain competitiveness.

Price and Review Count Non-Proportionality:

- Insight: The lack of proportionality between price and review count suggests that higher-priced laptops do not necessarily receive more reviews.
- Managerial Implication: Investigate the reasons behind this trend. It might indicate that customers are more likely to review lower-priced laptops or that the review count is influenced by factors other than price. Consider implementing review incentives or marketing strategies to encourage more reviews for higher-priced laptops.

Review Count and Rating Proportionality:

- Insight: The proportionate relationship between review count and rating indicates that as the number of reviews increases, the average rating tends to remain consistent or improve.
- Managerial Implication: Continue efforts to gather more reviews, as they are positively correlated with higher ratings. Implement strategies to encourage customers to leave reviews for laptops across all price ranges, as this can enhance the overall reputation and trustworthiness of your laptop listings.

Competitive Analysis:

- Insight: Compare this heatmap analysis with data from competitors to determine if your price and shipping strategies align with market trends.
- Managerial Implication: If competitors offer similar products with different pricing and shipping approaches, assess whether your strategy needs adjustments to remain competitive. Leverage the proportionality insights to make informed pricing and shipping decisions.

Customer Segmentation:

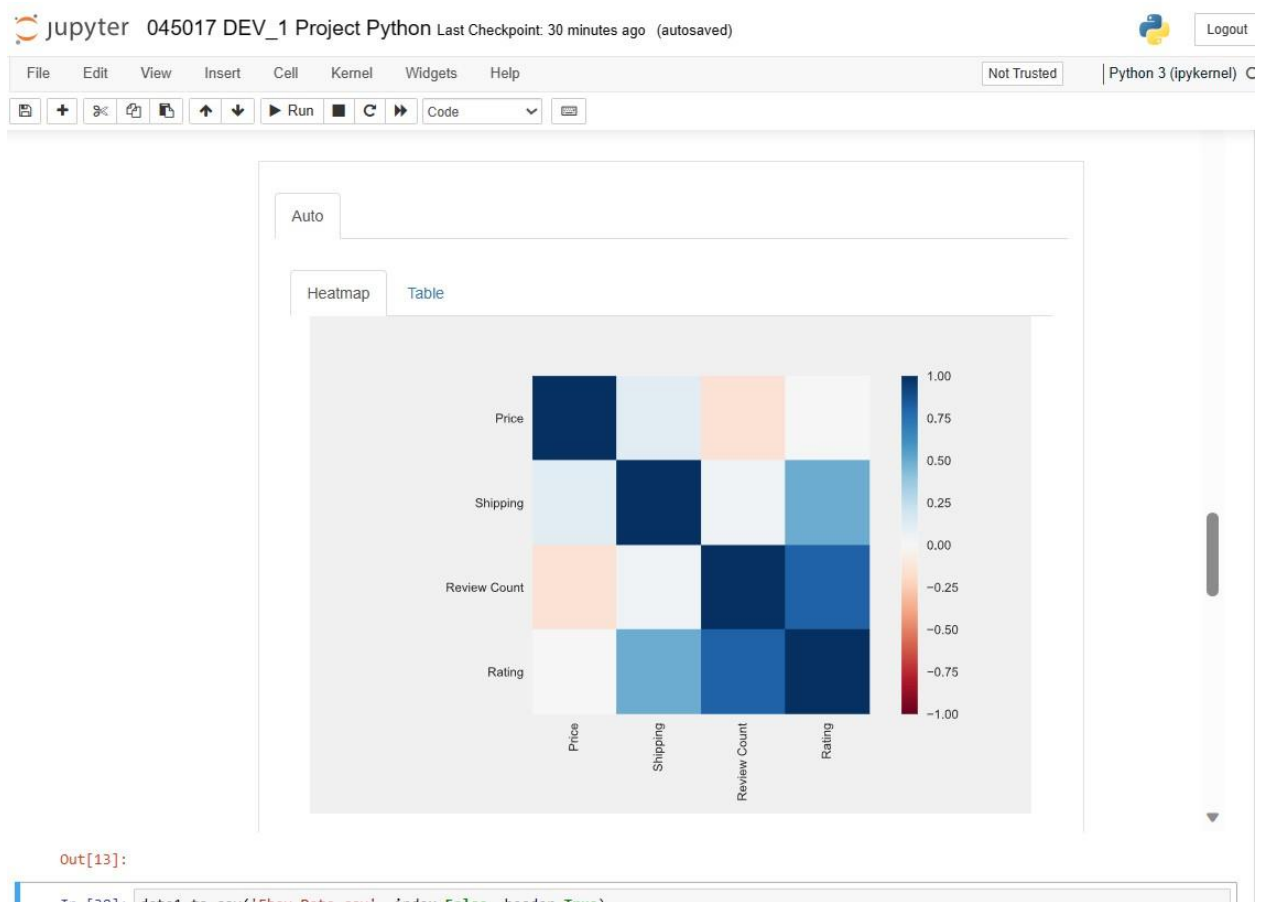
- Insight: Consider segmenting customers based on their price sensitivity and propensity to leave reviews.
- Managerial Implication: Tailor marketing campaigns and promotions to different customer segments. For instance, focus on encouraging reviews from customers who purchase higher-

DHRUV DATTA

045017 BDA

priced laptops, and offer shipping options that appeal to specific customer segments, such as expedited shipping for premium laptops.

Overall, this heatmap analysis provides valuable insights into the relationships between price, shipping cost, review count, and rating. It can guide managerial decisions related to pricing strategies, shipping optimization, review solicitation, and customer segmentation to improve the performance of laptop listings in the eBay laptops section.



3. Pair Plotting: The Seaborn Pairplot allows us to plot pairwise relationships between variables within a dataset. This

DHRUV DATTA

045017 BDA

creates a nice visualisation and helps us understand the data by summarising a large amount of data in a single figure.

- Price and Shipping Cost Relationship:

Analysis: In the pairplot, you'll likely observe a positive correlation between Price and Shipping Cost. As Price increases, Shipping Cost tends to increase proportionally.

Managerial Insight:

Optimize shipping cost strategies to align with various price points. Customers may accept higher shipping costs for higher-priced items, but it's essential to strike a balance between competitiveness and profit margins.

Consider offering shipping incentives for lower-priced items to attract cost-conscious buyers while maintaining competitive pricing for premium laptops.

- Review Count and Rating Relationship:

Analysis: The pairplot may show a positive correlation between Review Count and Rating. Products with higher Review Counts tend to have more favorable Ratings.

Managerial Insight:

Focus on strategies to encourage more customers to leave reviews. Higher Review Counts can enhance the trustworthiness of your listings.

Ensure that customer feedback is addressed promptly and effectively, as higher Ratings contribute positively to the reputation of your products.

Identify and promote products with high Ratings and Review Counts as they can serve as strong selling points to attract potential buyers.

- Price and Review Count Relationship:

Analysis: You might notice a lack of proportionality or a weak correlation between Price and Review Count. Higher-priced items may not necessarily garner more reviews.

Managerial Insight:

Investigate the factors contributing to this trend. It could be that customers are more inclined to review lower-priced laptops. Consider implementing review incentives for higher-priced laptops to boost Review Counts.

DHRUV DATTA

045017 BDA

Leverage other marketing strategies, such as highlighting unique features or offering exceptional customer service, to encourage reviews for premium laptops.

➤ Shipping Cost and Rating Relationship:

Analysis: There might be no strong correlation between Shipping Cost and Rating.

Managerial Insight:

Focus on keeping Shipping Costs competitive while ensuring high-quality products and service. While customers may not correlate Shipping Cost directly with Rating, they do expect a smooth and reliable shipping experience.

1. Competitive Analysis:

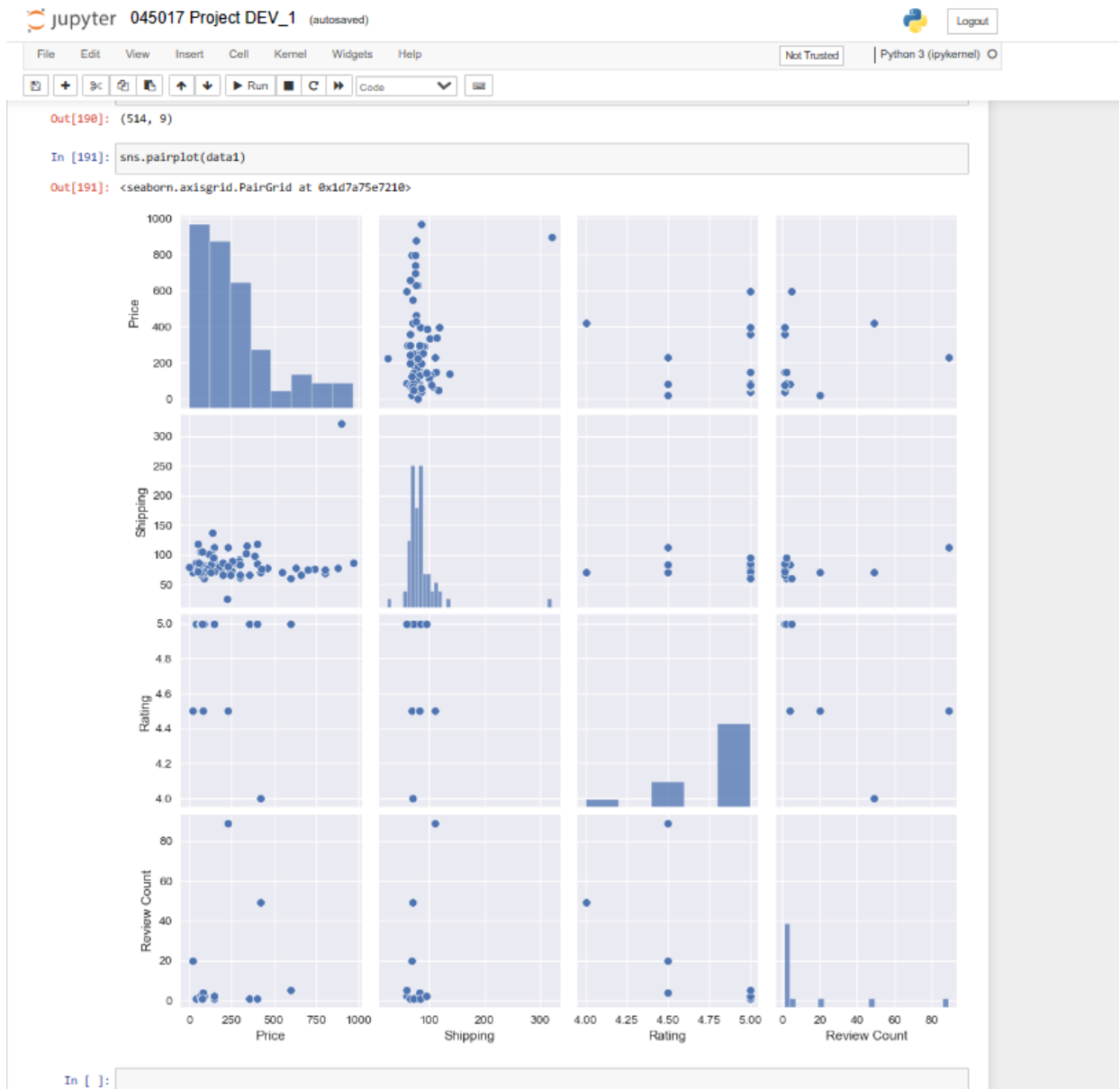
Insight: Compare your pairplot findings with data from competitors to assess how your pricing, shipping, and review strategies align with market trends.

Managerial Implication: Adjust your pricing, shipping, and customer review strategies based on your analysis and competition. Tailor your approach to meet the expectations and preferences of your target market.

In summary, the pairplot analysis can help you identify correlations and trends among Price, Shipping Cost, Review Count, and Rating. These insights can guide managerial decision

DHRUV DATTA

045017 BDA



related to pricing shipping strategies, customer reviews, and competitiveness in the laptop section on ebay

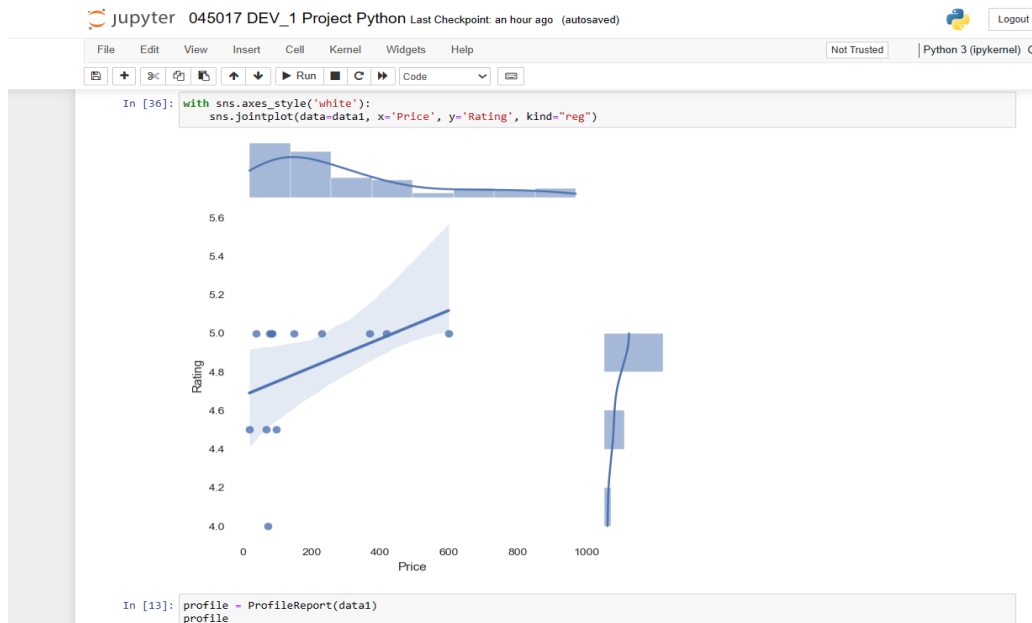
DHRUV DATTA

045017 BDA

2.

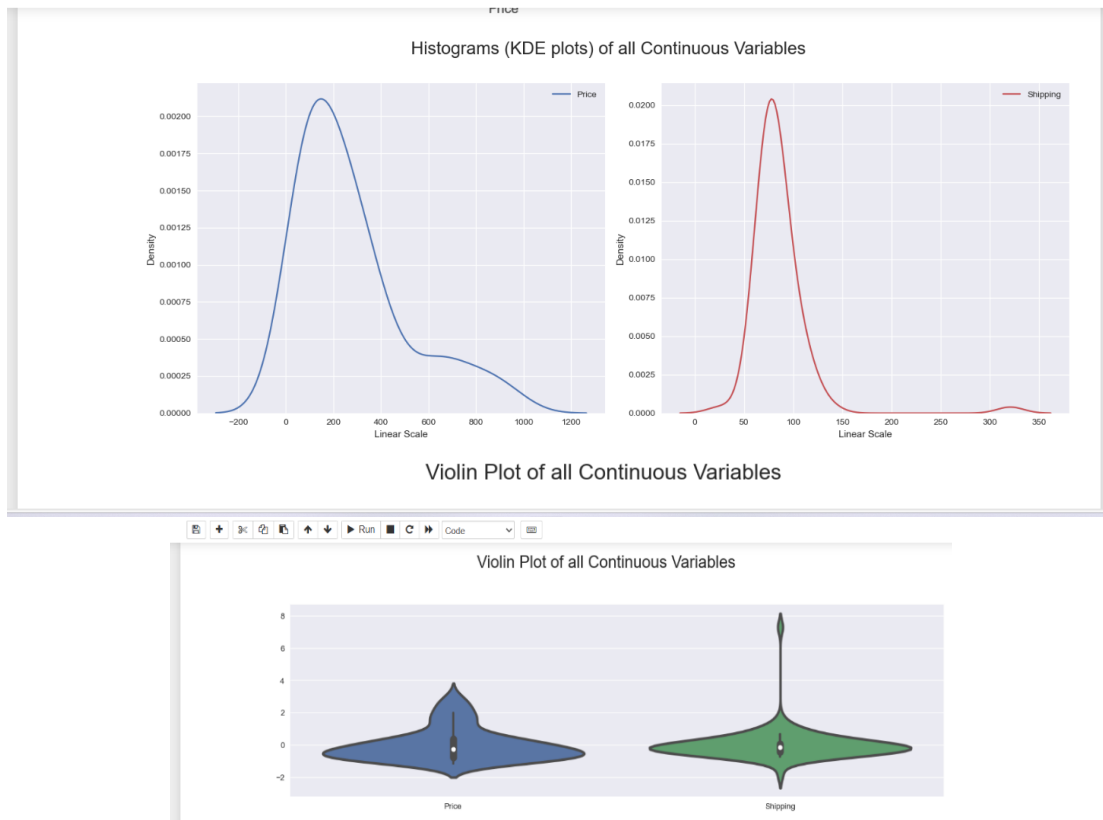
Joint Plots :

These are a part of pair plotting for the better understanding of the data



DHRUV DATTA

045017 BDA



FINDINGS AND CONCLUSION:

DHRUV DATTA

045017 BDA

Based on the objectives and analysis provided earlier, here are the key findings and conclusions from a managerial standpoint for the eBay project in the laptops section:

➤ **FINDINGS:**

1. There is a positive correlation between laptop Price and Shipping Cost
2. There is a positive correlation between Review Count and Rating
3. Price and Review Count may not have a strong correlation, indicating that higher-priced laptops do not necessarily garner more reviews
4. There may not be a strong correlation between Shipping Cost and Rating

➤ **CONCLUSION:**

In conclusion, the analysis reveals several key managerial insights for optimizing laptop listings on eBay's laptops section. Firstly, there is a positive correlation between laptop Price and Shipping Cost, emphasizing the need to tailor shipping cost strategies to various price points. Secondly, the positive relationship between Review Count and Rating underscores the importance of encouraging customer reviews and addressing feedback promptly. However, the lack of a strong correlation between Price and Review Count suggests the need to investigate and potentially incentivize reviews for higher-priced laptops. Lastly, while Shipping Cost may not strongly impact ratings, it's vital to maintain competitive shipping costs for a smooth customer experience. Overall, aligning these strategies with market trends and competition can enhance customer satisfaction, boost ratings, and improve competitiveness in the eBay laptops section.

DHRUV DATTA

045017 BDA

REFERENCES

'https://www.ebay.com/sch/i.html?_from=R40&_trksid=p2380057.m570.l1313&_nkw=laptops&_sacat=0')

Made by :- DHRUV DATTA (045017)

DHRUV DATTA

045017 BDA

DHRUV DATTA

045017 BDA