

Exploring Dark Web Marketplaces

Jessica Guan

Abstract – This paper examines the hidden subsection of the internet most commonly known as the dark web or darknet. Most widely known for its illicit activities, the dark web hosts many marketplaces where vendors and consumers can list and buy any goods or services anonymously. We will use data-driven methods to gain insights on three former popular marketplaces, Dream Market, Agora, and Valhalla.

I. Introduction

The internet can be divided into three parts: surface, deep, and dark [1]. The surface is what most people see, as it is the part of the internet that can be accessed through search engines like Google, Yahoo, or Bing. However, 75% of the internet is considered the deep web. These are web pages that cannot be accessed through a search engine, such as internal websites. Then, the dark web is a subsection of the deep web, in which its contents are purposefully hidden, and can only be accessed with specialised tools. The most popular tool to access the dark web is TOR (The Onion Router), a software that enables anonymity when browsing the internet. The dark web usually contains activity such as drug trafficking or other illegal services [4].

Launching in 2011, Silk Road was the first modern dark web marketplace until 2013. Vendors on Silk Road sold a wide range of goods and services ranging from legal goods to illegal goods including drugs, forgeries, and hacking services. Many more dark web marketplaces have been created since Silk Road's demise. The nature of these dark web marketplaces are relatively unknown, as it is tucked away in this hidden part of the internet. In this paper, the dark web marketplaces that will be explored using scraped data from Dream Market, Agora, and Valhalla, three dark web marketplaces that are no longer operating.

II. Research Questions and Analysis Strategy

A. Research Questions

The goal of this paper is to explore and understand the nature of the three dark web marketplaces. This will be achieved by answering the following questions.

- Which countries are the listings coming from?
- What goods and services are the most listed?
- What is the comparison of cocaine prices and listings?
- How can the cocaine prices on the dark web be modelled and predicted?

These questions are interesting to explore because structured data and research on the dark web is sparse, as it is a purposefully hidden part of the internet due to its illicit activities.

B. Analysis Strategy

- Load and review the datasets
- Prepare the data
- Derive any necessary data
- Investigate relationships between the features
- Find the correlations between the features
- Create regression models to analyse the trends and further understand the relationship between the features
- Use data clustering to investigate how each country uses the dark web marketplaces
- Use text analysis to extract information from the listings

III. Data

A. Data domain

This analysis investigates three dark web marketplaces, Dream Market (active 2013 - 2017), Agora (active 2013 - 2015), and Valhalla (active 2013 - 2019). It will look at the nature of each marketplace and compare them to one another. Each dataset has information on costs, country of origin, and item descriptions. This will be used for the analysis. Because the dataset on Dream Market is focused on cocaine listings, there will also be a focus on cocaine listings from each marketplace in addition to investigating the general trends of each

platform. The Agora and Valhalla datasets have all types of listings, including cocaine. In addition to analysing the cocaine listings, these two datasets will be used to explore the general listings and nature of dark web marketplaces.

B. Datasets

This study utilises the following datasets:

- Dream Market Cocaine Listings (July 2017) [5]
- Agora Listings (2014 - 2015) [6]
- Valhalla Listings (October 2016) [7]

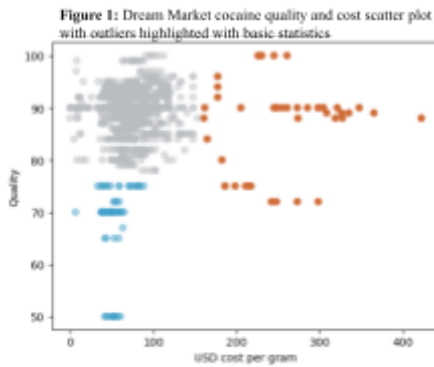
The Dream Market dataset [5] and Agora dataset [6] are from Kaggle, and the Valhalla dataset is from Sarah Jamie Lewis’s “Dark Web Data Dumps” [7]. The Dream Market dataset has 1504 rows and 63 columns, but 48 of the columns are booleans of where the product ships to which is already described in a “ships to” column. The Agora dataset has 109,689 rows and 8 columns. The Valhalla dataset has 16,511 and 4 columns.

IV. Data Preparation, Analysis, and Modelling

A. Data Preparation and Derivation

Dream Market

The dataset gives the cost of cocaine in Bitcoin. The USD cost per gram was derived using the conversion rate at the time of the dataset creation.



Mean	Median	Mode
78.15	71.30	77.85

Agora

The listings ranged from hacking services and weapons to drugs and counterfeits. To compare the cocaine listings, rows under the category of ‘Drugs/Stimulants/Cocaine’ were extracted and stored in a new dataframe. The USD cost was

derived from the Bitcoin value using the conversion rate at the time of the dataset creation. The Agora dataset did not give the cost per gram, so this required some data derivation. The item name contained the amount in grams, so numeric values were extracted and used to compute the cost per gram. Figure 2 visualises the cost per gram and rating to check for outliers. We can observe the cost outliers in orange.

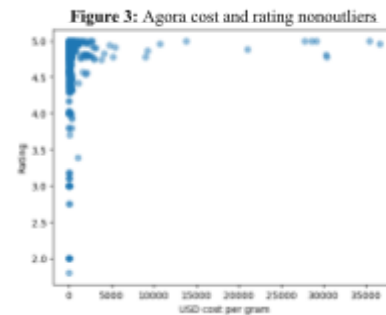
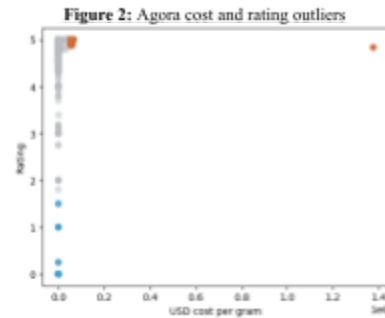


Figure 4: Agora cocaine statistics

Mean	Median	Mode
165.96	71.87	91.79, 92.21, 100.14

Figure 3 plots nonoutliers. There are still several listings in which the cost per gram is much higher. Figure 4 shows that the mean is much higher than the median, indicating that the existing outliers affect the results.

The average price for cocaine per gram at the end of 2009 was \$80.40 in the United States [2], approximately \$115.07 today. Based on this, the threshold for cost in the dataset was set to \$1000, outputting the data in Figure 5. The mean and median are comparable now.

Figure 5: Agora cost and rating with cost threshold of \$1000

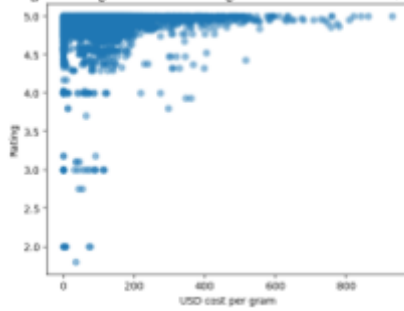


Figure 6: Agora cocaine statistics with cost threshold of \$1000

Mean	Median	Mode
88.30	70.01	91.79, 92.21, 100.14

Valhalla

The cost for each listing was given in Euros. To derive the USD cost of each listing, the estimated conversion rate at the time of the dataset creation was applied. Then, to focus on the cocaine listings, the rows that contained the word ‘cocaine’ in the item name were extracted and stored in a new dataframe. The numeric values were extracted and used to derive the ‘USD cost per gram’ column values.

Figure 7: Valhalla cost with constant Y

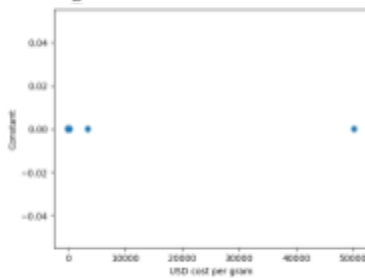
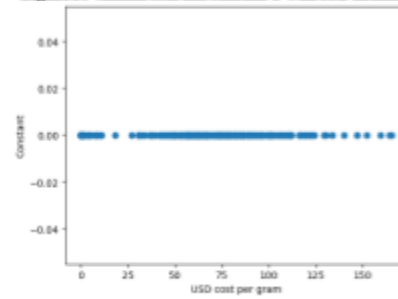


Figure 8: Valhalla cocaine statistics

Mean	Median	Mode
149.04	44.54	30.60, 75.72

In Figure 7 and 8, there are significant outliers. Based on the real prices, the threshold was set to \$1000, outputting the data in Figure 9. The mean and median are comparable now.

Figure 9: Valhalla cost and constant Y with cost threshold \$1000



Valhalla cocaine statistics with cost threshold \$1000

Mean	Median	Mode
41.54	44.38	30.60, 75.72

B. Analysis

Cocaine Listings across the Marketplaces

After shaping the data, we can compare the price of cocaine per gram across the dark web marketplaces. Agora had the highest average cost per gram across the listings for cocaine at \$88.30, followed by Dream Market at \$78.15. Based on the investigation, Valhalla had the lowest average cost at \$41.54 per gram. However, Valhalla had the least number of cocaine listings at 557, while Agora had the most at 6007. Dream market is in between the two with 1504.

Figure 10: Marketplaces statistics comparison

	Dream Market	Agora	Valhalla
Mean	78.15	88.30	41.54
Median	71.30	70.01	44.38
Mode	77.85	91.79, 92.21, 100.14	30.60, 75.72
Number of listings	1504	6007	557

Dream Market Cocaine Listings

Data clustering was utilised to investigate the data based on country. In the Dream Market data, most of the cocaine listings are dominated by vendors from the Netherlands, Denmark, Great Britain, and the United States, as shown in Figure 11. Despite the Netherlands leading in the highest number of listings, the UK had the highest number of successful transactions as shown in Figure 12. All four countries took the lead in where the listings are coming from and the number of successful transactions.

Figure 11: Dream Market - cocaine listings country of origin

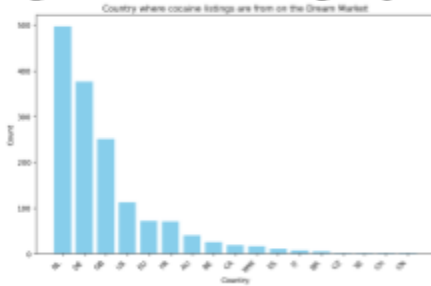


Figure 12: Dream Market - successful transactions by country



The following tables describe the average cost per gram, average quality (out of 100), average rating (out of 5), percentage of listings in which escrow is offered, and number of successful transactions. The Netherlands had the lowest average cost per gram, while the UK had the highest average cost per gram. Despite the cost per gram being more expensive on average from vendors in the UK, the UK still had the most successful transactions.

Figure 13: Dream Market - cocaine listings comparison of top countries

	Netherlands	Denmark	UK	USA
Average cost per gram	62.80	71.96	90.39	81.77
Average quality	87.77	89.57	87.31	89.96
Average rating	4.86	4.87	4.89	4.91
Escrow offered	78.83%	81.65%	85.26%	66.07%
Successful transactions	114,814	163,534	163,974	22,240
Number of listings	496	376	251	112

Agora Cocaine Listings

Due to the nature of this dataset, a word-cloud was used to investigate which countries had the most cocaine listings. Using text analysis, we can see which countries were most frequently listed. The countries were the USA, UK, Netherlands, and Australia. With this information, the data for each country was extracted and stored in a new dataframe. The dataset only gave the cost and rating information. We can see that the Netherlands had the

lowest cost per gram at \$52.04, while Australia had the highest at \$251.76.

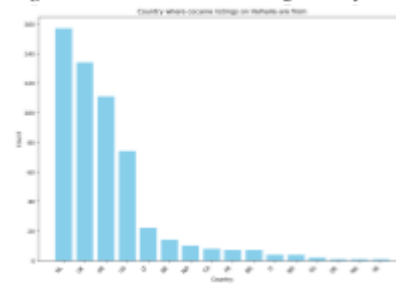
Figure 14: Agora - cocaine listings comparison of top countries

	USA	UK	Netherlands	Australia
Average cost per gram	77.92	88.85	52.04	251.76
Average rating	4.87	4.9	4.83	4.94
Number of listings	1273	842	701	501

Valhalla Cocaine Listings

Using data clustering in Valhalla's cocaine listings, the data shows that the cocaine listings were led by vendors from the Netherlands, UK, Germany, and USA, as shown in Figure 15.

Figure 15: Valhalla - cocaine listings country of origin



The dataset did not include data on the quality, rating, successful transactions, or escrow. We can only observe the average cost per gram by country. Figure 16 shows the average cost in the four leading countries of cocaine listings on Valhalla. The average cost was much lower than on Dream Market, with the Netherlands with the lowest cost per gram at \$28.65.

Figure 16: Valhalla - cocaine listings comparison of top countries

	Netherlands	UK	Germany	USA
Average cost per gram	28.65	57.23	39.36	57.03
Number of listings	149	131	101	62

C. Regression Modelling

Because the Dream Market dataset has the most features on cocaine listings, this dataset was used to perform regression modelling to predict the cost per gram for cocaine on the dark web.

Feature Selection and Correlations

We must look at the correlations between the features (quality, escrow, successful transactions,

rating) and the independent variable (cost per gram). In Figure 21, we can see the Pearson and Spearman correlation matrix. Looking at the second row of both graphs, the features do not show a monotonic relationship.

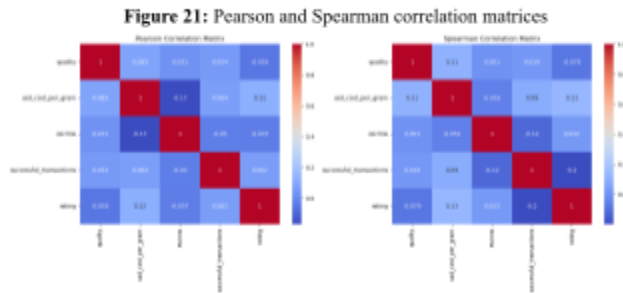
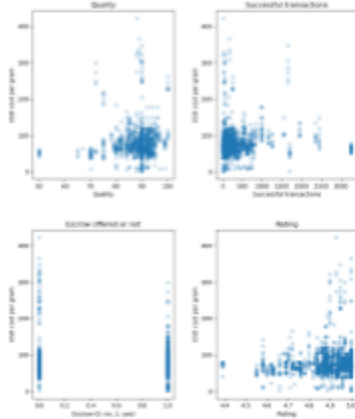


Figure 22 visualises the data. The data is mostly clustered in one area of the graph with some outliers.

Figure 22: Scatter plots for features in Dream Market data

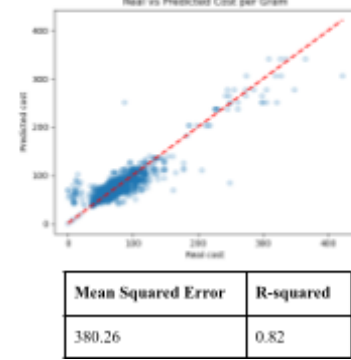


Three models were built, multilinear regression, polynomial regression, and decision tree. Due to the lack of monotonic relationship, multilinear and polynomial regression had poor results, so we will focus on decision tree.

Decision Tree

The decision tree model gave the best performance results with a mean squared error of 380.26 and a r-squared number of 0.82, capturing 82% of the variability of the dependent variables. Figure 25 shows that the predicted values versus real values follow the line more closely. Because decision trees can handle nonlinear relationships, the performance was significantly better than the multilinear and polynomial regression models' performance results.

Figure 25: Decision tree regression with MSE and r-squared results



V. Findings, reflections, and further work

A. Findings

Cocaine listings on each marketplace

Valhalla had the lowest average cost per gram and the lowest number of listings. Agora had the highest average cost per gram. However, the same three countries, Netherlands, UK, and USA dominated the cocaine listings on all three platforms. Comparing the mean cost for these three countries on Dream Market and Agora, Agora actually has a lower cost. The overall average cost for Agora's listings can be attributed to Australia being the fourth country that was most active in Agora's cocaine market. The average cost per gram for Australian vendors on Agora was \$251.76. This aligns with real costs because the price for cocaine in Australia ranges from \$168.21 to \$269.14 [3], while the average cost in the USA is approximately \$115.07.

Using word-clouds, the description of the cocaine listings on all platforms was explored. Cocaine listings across all three marketplaces shared the most frequent word, "pure". "Pure" tells buyers that the product is uncut with other substances. Dream Market and Agora listings also shared their second most used word, "fishscale". "Fishscale" is a term used to describe cocaine when it is pure. Valhalla's second most frequently used word in cocaine listings was "flake". "Flake" is also used to describe cocaine when it is pure.

Figure 17: Most frequent words on all marketplaces in cocaine listings

	Dream Market	Agora	Valhalla
1.	Pure	Pure	Pure
2.	Fishscale	Fishscale	Flake

Top active countries on each marketplace

While we do not have data on Dream Market's listings outside of its cocaine market, the most active countries on Agora and Valhalla align with the most active countries in their cocaine listings.

Figure 18: Most active countries on the marketplaces (except Dream Market)

	Agora	Valhalla	Dream Market (cocaine listings only)
1.	USA	Netherlands	Netherlands
2.	UK	UK	Denmark
3.	Netherlands	Germany	UK
4.	Australia	USA	USA

Most frequently listed goods and services

The general listings were explored for Agora and Valhalla. Using word-clouds, the most commonly listed goods and services were investigated. Figure 19 shows the most frequent category on Agora.

Figure 19: Most frequent category of items on Agora

	Agora
1.	Cannabis/Weed
2.	Ecstasy
3.	Stimulants
4.	Psychedelics

The most listed items from the top four countries on Agora were also explored. Shown in Figure 20, USA listed the most hacking services, UK listed the most cocaine, Netherlands listed the most MDMA, and Germany listed the most decanoate (nandrolone decanoate), a synthetic testosterone commonly used to build muscle mass.

Figure 20: Most sold items by country on Agora

USA	UK	Netherlands	Germany
Hacking	Cocaine	MDMA	Decanoate

As the Valhalla dataset did not have a column that categorised the items, text analysis was used on the item listings and the most relevant frequently used word was "pure". Although we cannot conclude what exactly the items were, the term "pure" is often used to describe drugs. With this knowledge, the most common items were drugs.

B. Reflections

This analysis answers the research questions through investigating the data to compare the cocaine listings

on all three marketplaces, explore the nature of each marketplace, find which countries were most active on each platform, and what goods and services are most listed with data preparation, data derivation, inferential statistics, data clustering, and text analysis. In addition, regression modelling was performed to understand the relationship between the features and the cost per gram. The decision tree model yielded the best results, which were validated by the mean squared error and r-squared results, and can be further used to predict cocaine costs on the dark web.

C. Further Work

Given the time constraints, the research questions were answered. However, we can improve the research by exploring data on dark web marketplaces that are currently operating. This would be useful in examining how trends have changed or developed since 2014 - 2017, which is the time period of the three datasets used in this study. We can also improve the accuracy of the insights in this study by applying further data manipulation, as all three datasets varied greatly in their structures and data provided.

References

- [1] A. Gupta, S. Maynard, and A. Ahmad, "The Dark Web Phenomenon: A Review and Research Agenda Full Paper," 2019. Accessed: Dec. 18, 2023. [Online]. Available: <https://arxiv.org/pdf/2104.07138.pdf>
- [2] "The Global Cocaine Market," 2010. Available: https://www.unodc.org/documents/wdr/WDR_2010/1.3_The_global_cocaine_market.pdf
- [3] P. Duffin, "Cocaine Cartels Cash in on sky-high Prices in Australia," *The Sydney Morning Herald*, Jul. 15, 2023. <https://www.smh.com.au/national/nsw/cocaine-cartels-cash-in-on-sky-high-prices-in-australia-20230715-p5dohc.html> (accessed Dec. 18, 2023).
- [4] Sultana, J. & Jilani, Abdulkhadar. (2021). Exploring and Analysing Surface, Deep, Dark Web and Attacks. 10.1007/978-3-030-69174-5_5.
- [5] D. Everling, "Darknet Market Cocaine Listings," *www.kaggle.com*. <https://www.kaggle.com/datasets/everling/cocaine-listings> (accessed Dec. 23, 2023).
- [6] P. James, "Dark Net Marketplace Data (Agora 2014-2015)," *www.kaggle.com*. <https://www.kaggle.com/datasets/philipjames11/dark-net-marketplace-drug-data-agora-20142015/data> (accessed Dec. 23, 2023).
- [7] S. J. Lewis, "Dark Web Data Dumps," *Gwern.net*, 2016. <https://gwern.net/doc/darknet-market/2016-sarahjamilewis-darkwebdatadumps-hansavalhalla.zip> (accessed Dec. 23, 2023).

Word Counts

Abstract: 63

- I. Introduction: 233
- II. Research Questions and Analysis Strategy: 165
- III. Data: 228
- IV. Data Preparation, Analysis, and Modelling: 992
- V. Findings, reflections, and further work: 592