Project 4-Group 2 consisting of Shahla Shahnawaz, Asia Byrne, Aaron Suarez, William Berry, and Cassidy Bell.

Project 4 Overview:

This project aims to classify darknet cocaine listings into Low, Medium, and High price categories using data from [two different datasets] multiple sources including Agora, Silk Road, and MiDarknet. By combining structured and unstructured features, we will train a classification model and analyze vendor behavior to flag suspicious listings, helping uncover potential criminal trends. Tableau dashboards will be used to visualize pricing patterns and risk indicators across darknet markets. We will address pricing trends and what factors influence them, if machine learning can predict the price or category of a listing, if there are behavioral or linguistic patterns that indicate suspicious or high-risk listing.

Background and Introduction:

First, what is the darknet(darkweb)? According to the United Nations Office on Drugs and Crime (UNODC) "Darknets (darkweb), or overlay networks within the Internet that can only be accessed with specific software, configurations, or authorization, and often use a unique customized communication protocol. Two typical darknet types are social networks (usually used for file hosting with a peer-to-peer connection), and anonymity proxy networks such as Tor (special software) via an anonymized series of connections".

According to the U.S. Department of the Treasury press release from March 4, 2025, the Office of Foreign Assets Control (OFAC) helped take down "Nemesis" in 2024. Per the press release, "Nemesis, an online darknet marketplace, which was subject of an international law enforcement operation and was taken down in 2024. Prior to its takedown by law enforcement, narcotics traffickers and cybercriminals openly traded in illegal drugs and services on Nemesis, which was designed with built-in money laundering features. Nemesis had over 30,000 active users and 1,000 vendors and facilitated the sale of nearly $30 million worth of drugs around the world between 2021 and 2024, including to the United States".

The press release went on to say "Nemesis was established in 2021 and operated as a criminal marketplace on the darknet, an encrypted network within the Internet that can only be accessed with special anonymity-enhancing browsers. Drug traffickers active on Nemesis sold fentanyl around the world, both on its own and surreptitiously laced into other drugs. In addition to offering narcotics for sale, Nemesis facilitated the sale of a wide variety of other goods and services such as false identification documents and professional hacking services that enabled buyers to hire hackers to illegally seize control of the online accounts and communications of selected victims". Per the story "US sanctions crypto addresses linked to Nemesis darknet marketplace" by Cointelegraph and Stephen Katte on MSN.com which is in relation to the U.S. Department of the Treasury press release, "Darknet marketplaces generated over $1.7 billion in revenue in 2024, only a slight increase from the previous year, according to blockchain intelligence firm TRM Labs 2025 Crypto Crime Report".

Data Cleaning:

First order of business was to clean our two datasets. Our main data set was rather big and cumbersome. There were many columns of information to contend with.

```
df.columns
```

Out[3]:

```
Index(['product_title', 'ships_from_to', 'grams', 'quality', 'btc_price',
       'cost_per_gram', 'cost_per_gram_pure', 'vendor_name',
       'successful_transactions', 'rating', 'ships_from', 'ships_to',
       'ships_to_US', 'ships_from_US', 'ships_to_NL', 'ships_from_NL',
       'ships_to_FR', 'ships_from_FR', 'ships_to_GB', 'ships_from_GB',
       'ships_to_CA', 'ships_from_CA', 'ships_to_DE', 'ships_from_DE',
       'ships_to_AU', 'ships_from_AU', 'ships_to_EU', 'ships_from_EU',
       'ships_to_ES', 'ships_from_ES', 'ships_to_N. America',
       'ships_from_N. America', 'ships_to_BE', 'ships_from_BE', 'ships_to_WW',
       'ships_from_WW', 'ships_to_SI', 'ships_from_SI', 'ships_to_IT',
       'ships_from_IT', 'ships_to_DK', 'ships_from_DK', 'ships_to_S. America',
       'ships_from_S. America', 'ships_to_CH', 'ships_from_CH', 'ships_to_BR',
       'ships_from_BR', 'ships_to_CZ', 'ships_from_CZ', 'ships_to_SE',
       'ships_from_SE', 'ships_to_CO', 'ships_from_CO', 'ships_to_CN',
       'ships_from_CN', 'ships_to_PL', 'ships_from_PL', 'ships_to_GR',
       'ships_from_GR', 'usd_price', 'vendor_performance'],
      dtype='object')
```

We were able to reduce the columns to the information we felt was most valuable for our project.

```
features = ['vendor_name', 'grams', 'quality', 'successful_transactions',
'rating', 'ships_from', 'ships_to', 'product_title_sentiment',
'usd_price_per_gram']
df2 = df.loc[:, features]
print(df2.shape)
```

```
df2.head()
```

```
(1504, 9)
```

Out[8]:

	vendor_name	grams	quality	successful_transactions	rating	ships_from	ships_to	product_title_sentiment	usd_price_per_gram
0	Mister-Molly	1.0	90.0	90	4.63	NL	EU	0.000000	2225.964410
1	Mister-Molly	2.0	90.0	90	4.63	NL	EU	0.000000	2224.236847
2	0ldamsterdamm	0.5	89.0	620	4.94	NL	EU	0.200000	2848.750727
3	lhomme-masquer	1.0	89.0	15	5.00	FR	EU	0.000000	3558.778956
4	SMOOTHCRIMINAL007	1.0	87.0	28	4.78	NL	WW	-0.142857	2936.856420

In [9]:

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1504 entries, 0 to 1503
Data columns (total 9 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   vendor_name              1504 non-null   object
 1   grams                    1504 non-null   float64
 2   quality                  1504 non-null   float64
 3   successful_transactions  1504 non-null   int64
 4   rating                   1504 non-null   float64
```

```
 5   ships_from               1504 non-null   object
 6   ships_to                 1504 non-null   object
 7   product_title_sentiment  1504 non-null   float64
 8   usd_price_per_gram       1504 non-null   float64
dtypes: float64(5), int64(1), object(3)
memory usage: 105.9+ KB
```

Machine Learning:

With some assistance from Professor Booth, we were able to jumpstart our machine learning code. We are working with two datasets. The first dataset deals specifically with cocaine on the darkweb. This dataset was used for our machine learning. Not only was there a considerable amount of information about the product i.e. grams, quality, and price along with a tremendous amount of shipping data. Where the product is shipping from and where the product is shipping to. For starters, we worked to see what may be a good predictive feature. Initially, "Vendor Quality" seemed promising. However, "Product Title Sentiment" did not work out well.

Working on Processing Pipelines and Correlation:

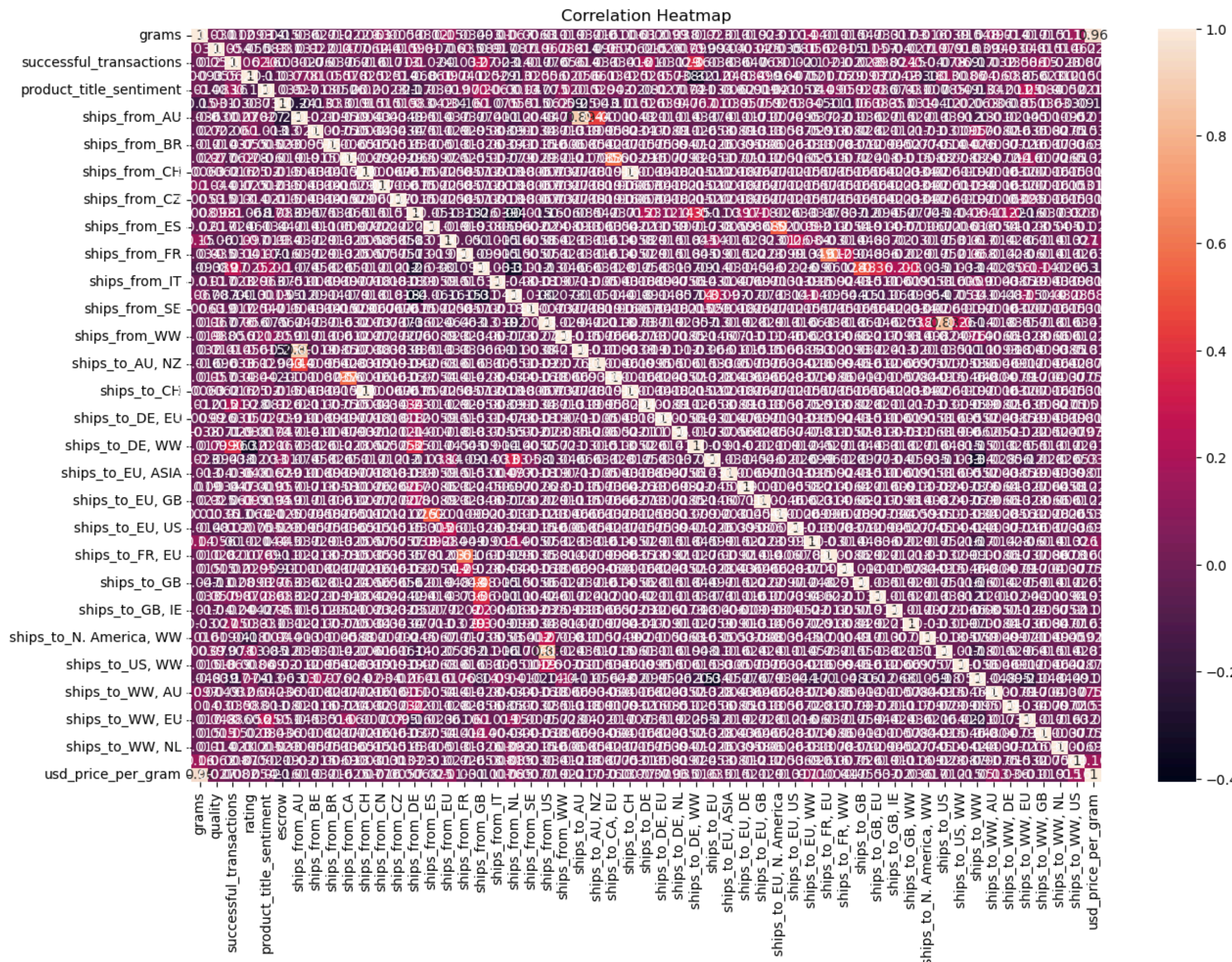
```
# Define Preprocessing Pipelines

# Define preprocessing for numeric features (quality, ships_from, ships_to)
numeric_features = ['grams', 'quality', 'successful_transactions', 'rating',
'product_title_sentiment'] # You can also do this in a loop, select the
numeric columns
numeric_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='median')),
    ('scaler', StandardScaler())])

# Define preprocessing for the binary features
# binary_features = ['escrow']
# binary_transformer = Pipeline(steps=[
#     ('imputer', SimpleImputer(strategy='most_frequent',
missing_values=pd.NA)),
#     ('label', OrdinalEncoder())])  # Label encode for binary feature

# Define preprocessing for categorical features
categorical_features = ['ships_from', 'ships_to',]
categorical_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='most_frequent', missing_values=pd.NA)),
    ('onehot', OneHotEncoder(handle_unknown='ignore'))])

# Combine preprocessing for numeric and categorical features
preprocessor = ColumnTransformer(
```

```
    transformers=[
        ('num', numeric_transformer, numeric_features),
        # ('binary', binary_transformer, binary_features),
        ('cat', categorical_transformer, categorical_features)])
```

In [394]:

```
# Correlation Analysis (this is optional and not needed for the ML Experiment)
# It just shows what the data looks like after transformation before training
# We will still declare a full pipeline of preprocessing + training

# Use only preprocessing pipeline to transform the data
preprocessed_X_train = preprocessor.fit_transform(df).toarray()

# Convert preprocessed data to a DataFrame
# Get the feature names after one-hot encoding
encoded_feature_names = (numeric_features + binary_features +

list(preprocessor.transformers_[2][1]['onehot'].get_feature_names_out(categoric
al_features)))

df_final = pd.DataFrame(preprocessed_X_train, columns=encoded_feature_names)
df_final["usd_price_per_gram"] = df.usd_price
df_final.head()
```

Below is our Correlation Heatmap:



Below is one of our Regressions:

TRAIN METRICS

R2: 0.9283289685365075

MSE: 132482.06725879977

RMSE: 363.9808611160754

MAE: 207.96061551160312

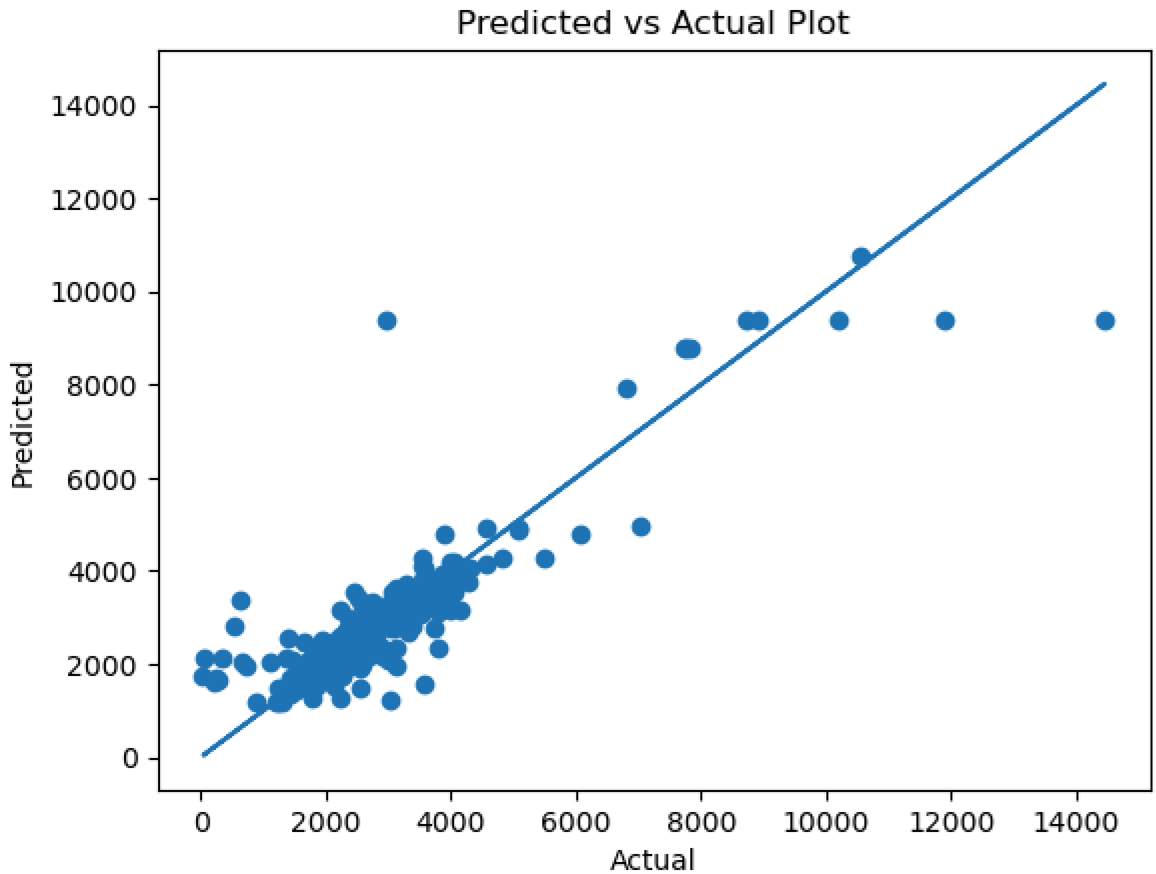
TEST METRICS

R2: 0.7945958134005243

MSE: 438486.3102823003

RMSE: 662.1829885177513

MAE: 342.44823471455874


Predicted vs Actual Plot
Predicted
Actual
0
2000
4000
6000
8000
10000
12000
14000

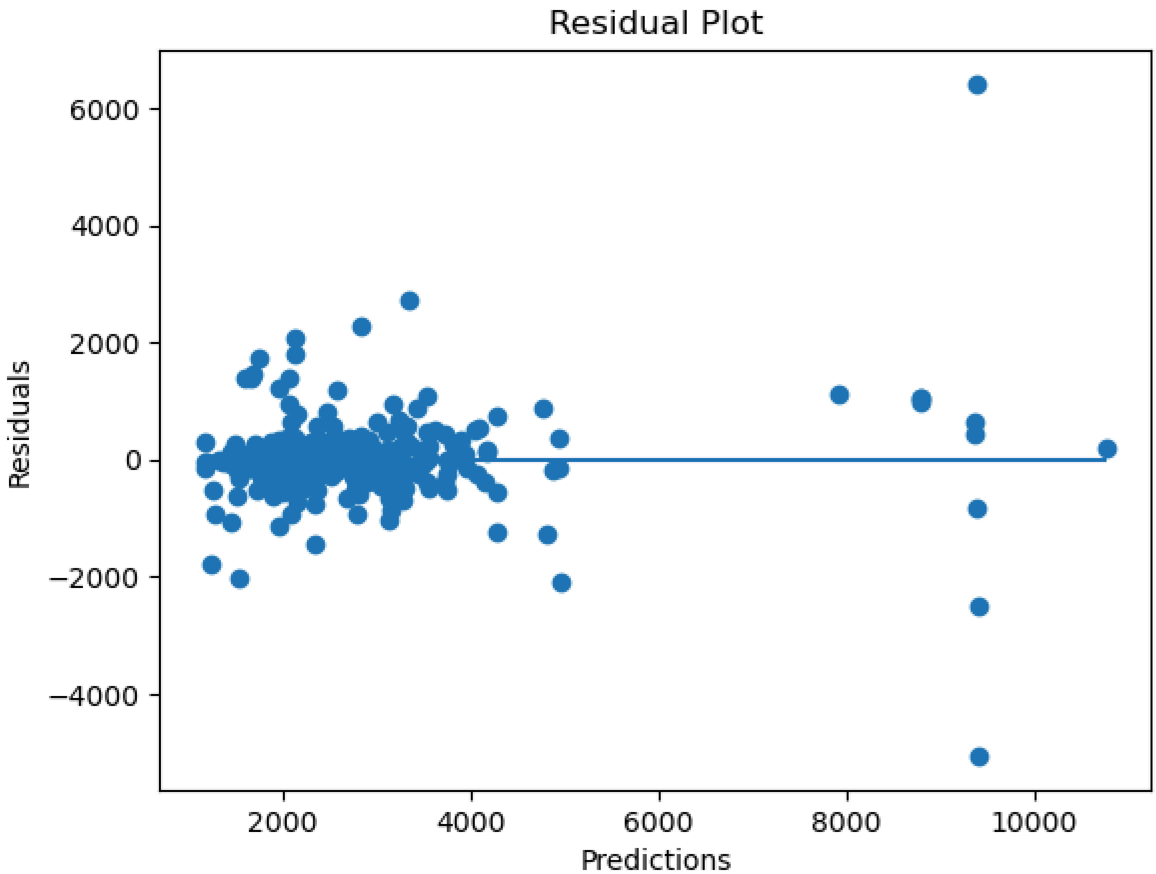
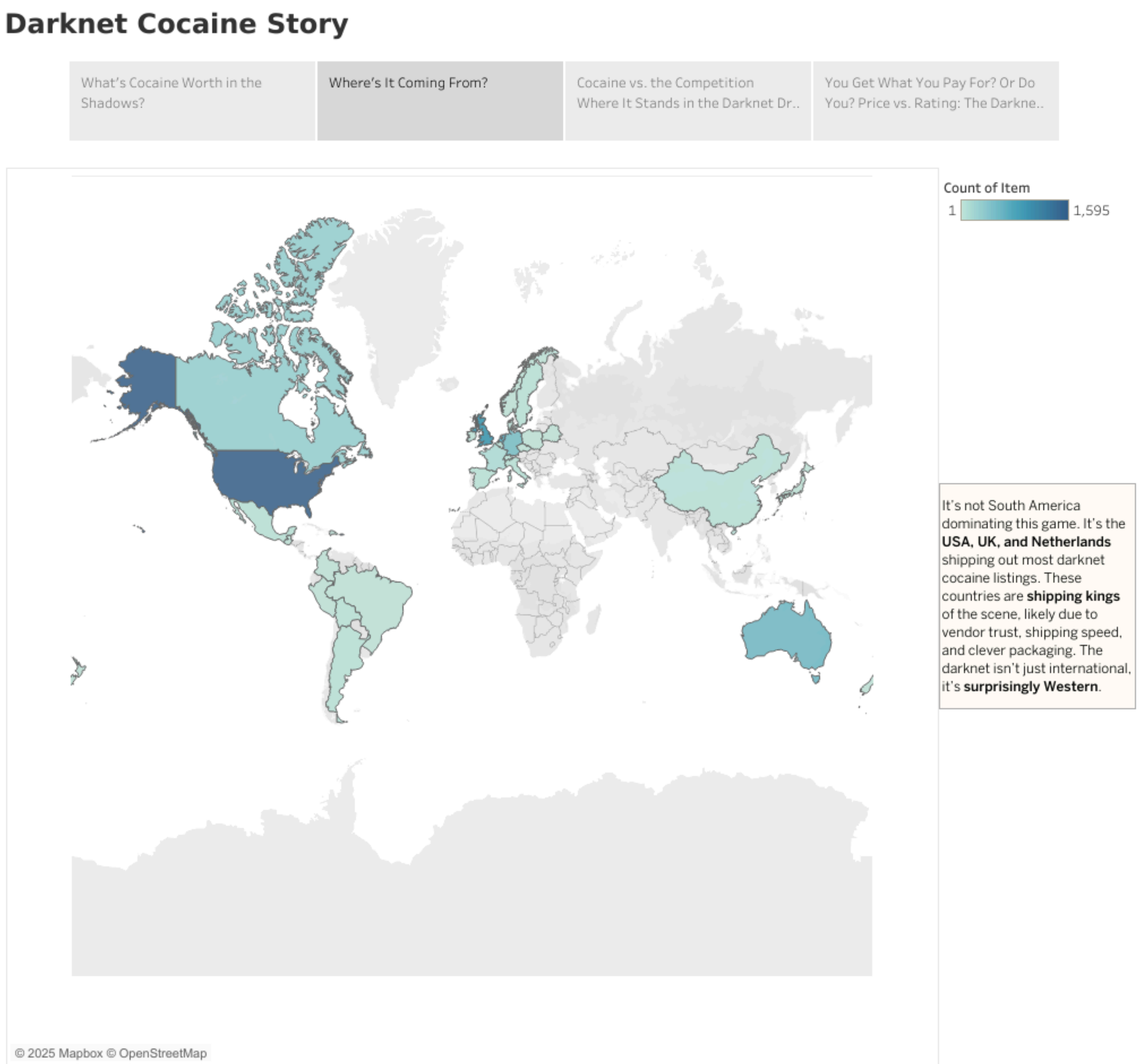
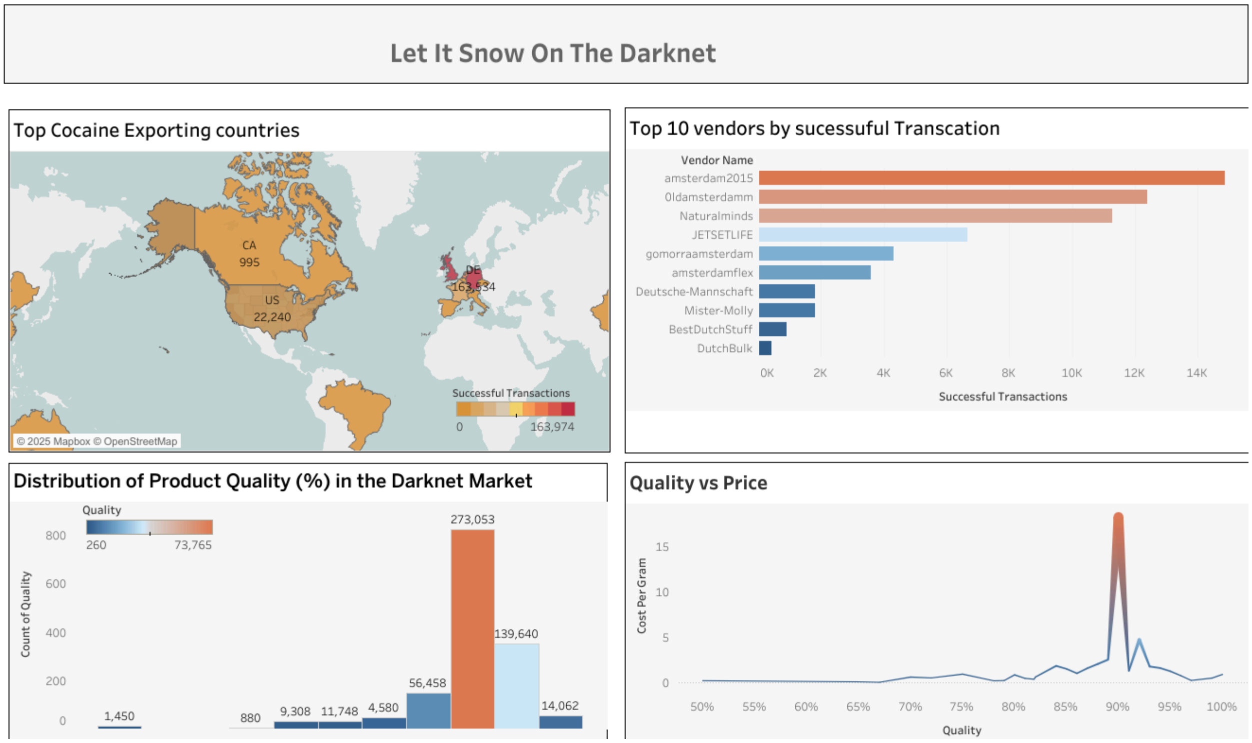


Tableau:

We incorporated our second dataset for the group's Tableau dashboards. Below is our first dashboard.



Here is our second dashboard:



Our website:

We built a website containing a home page which explains what the darknet is, a modeling page, our Tableau dashboards, and information about the team members. The website has additional pages as well.

Below is a sample of the website code:

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <!-- Bootstrap -->
    <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootswatch/4.5.2/lux/bootstrap.min.css">
    <!-- Font Awesome -->
    <link rel="stylesheet"
href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/6.0.0-beta3/css/all.min.css">
```

```
    <!-- JQuery -->
    <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
    <script type="text/javascript"
src="https://public.tableau.com/javascripts/api/tableau-2.min.js"></script>
    <script type="text/javascript" src="static/js/tableau_api1.js"></script>
    <!DOCTYPE html>

        <!-- boilerplate -->
        <meta charset="UTF-8">
        <meta name="viewport" content="width=device-width, initial-scale=1.0">

        <!-- content -->
        <title>Darkweb MarketPlace Analysis</title>
        <link rel="stylesheet"
href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-alpha1/dist/css/bootstrap.min.css">

        <!-- scripts -->
        <script defer
src="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-alpha1/dist/js/bootstrap.bundle.min.js"></scri
pt> <!-- includes Popper.js -->

      </head>

    <body>

      <!-- header section -->
      <div class="container-fluid bg-dark text-white text-center py-4">
        <h1>DarkWeb MarketPlace Analysis</h1>
      </div>

      <!-- navbar section -->
      <nav class="navbar navbar-expand-lg bg-primary" data-bs-theme="dark">
        <ul class="navbar-nav me-auto">
          <li class="nav-item">
            <a href="/" class="nav-link active">Home</a>
          </li>
          <li class="nav-item">
            <a href="/tableau1" class="nav-link">Tableau1</a>
```

```
          </li>
          <li class="nav-item">
            <a href="/tableau2" class="nav-link">Tableau2</a>
          </li>
          <li class="nav-item">
            <a href="/model" class="nav-link">Model</a>
          </li>
          <li class="nav-item">
            <a href="/report" class="nav-link">Report</a>
          </li>
          <li class="nav-item">
            <a href="/about_us" class="nav-link">About Us</a>
          </li>
          <li class="nav-item">
            <a href="/sources" class="nav-link">Sources</a>
          </li>
        </ul>
      </nav>

      <!-- content section -->
      <div class="container mt-4">
      </div>

      <h2>What is the DarkWeb?</h2>
      <p>Darknets are special parts of the Internet that require certain accesses in order to use
them. They often use unique communication platforms such as peer-to-peer social networks and
encrypted connections.</p>

      <div class="col-md-6">
        <img
src="https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2023/01/27122247/shu
tterstock_721521202-2-scaled-1-800x450.jpg" class="img-fluid"/>
      </div>
    </body>

</html>
```

Conclusion and Summary:

Our group discussed drugs and the dark web via Kaggle.com datasets. We defined the dark web and provided examples of the extent of some of the recent illegal activities on the dark web. We created a model to predict the cost of cocaine sales. We looked at cocaine price versus the drug quality. We created tableau dashboards showing how cocaine sales compare to other illicit drug sales. We even have a map showing top countries of dark web drug sales along with top vendors. With this being a capstone project, we used many of the things we learned in the bootcamp including data cleaning, machine learning, flask apps, building the front end and back end, tableau visualizations, and building a website.

There is more information needed for further research. Obviously, it would be extremely difficult to fully determine the true extent to dark web illegal drug sales. Also, we would need more information in regards to each country's illegal drug activity via the dark web.

Bibliography:

1. James, P. (2015). *Dark Net Marketplace Data (Agora 2014-2015)*. Kaggle. https://www.kaggle.com/datasets/philipjames11/dark-net-marketplace-drug-data-agora-20142015

2. Everling, A. (2025). *Darknet Market Cocaine Listings*. Kaggle. https://www.kaggle.com/datasets/everling/cocaine-listings

3. Nicapotato. (2025). *DARK NET: Light in the Dark* [Kaggle Notebook]. https://www.kaggle.com/code/nicapotato/dark-net-light-in-the-dark

4. OpenAI. (2025). *ChatGPT*. https://openai.com/chatgpt/overview/

5. United Nations Office on Drugs and Crime. (2023). *Use of the dark web and social media for drug supply*. World Drug Report 2023. https://www.unodc.org/res/WDR-2023/WDR23_B3_CH7_darkweb.pdf

6. U.S. Department of the Treasury. (2025, March 4). *Press release: US sanctions crypto addresses linked to Nemesis darknet marketplace*. https://home.treasury.gov/news/press-releases/sb0040

7. Katte, S. (2025, March 4). *US sanctions crypto addresses linked to Nemesis darknet marketplace*. Cointelegraph, via MSN. https://www.msn.com/en-us/money/markets/us-sanctions-crypto-addresses-linked-to-nemesis-darknet-marketplace/ar-AA1Al56B

8. TRMLabs. (2025, March 31). *Category deep-dive: Illicit drug sales grew and expanded outside of darknet marketplaces in 2024*. https://www.trmlabs.com/resources/blog/category-deep-dive-illicit-drug-sales-grew-and-expanded-outside-of-darknet-marketplaces-in-2024