Exploring Wildlife Trafficking Data

ADITYA GOUROJU, New York University, USA
VAMSHI NAIK VISLAVATH, New York University, USA
ARYAN PRASAD, New York University, USA

ACM Reference Format:
Aditya Gouroju, Vamshi Naik Vislavath, and Aryan Prasad. 2024. Exploring Wildlife Trafficking Data. 1, 1 (May 2024), 15 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The illegal wildlife trade, a multibillion-dollar criminal industry driven by demand for exotic pets, traditional medicines, fashion accessories and other animal products, has become a global crisis pushing numerous species towards extinction. Transnational trafficking networks are exploiting both physical markets and the online environment to facilitate this trade, which has surged in recent years due to increased connectivity and efficient supply chains.

The rise of e-commerce platforms, encrypted online marketplaces and social media has enabled wildlife traffickers to operate digitally with anonymity, direct access to consumers worldwide, and limited regulation compared to the physical trade. They can easily advertise contraband, use code words to avoid detection, and coordinate transactions over these digital channels with vast criminal networks involved in the illicit trade.

The online wildlife trade encompasses a wide range of species, from exotic pets to parts of endangered animals like illegal ivory from elephants and rhinos used for carved products. Despite its critical threat to conservation efforts, empirical research examining this digital dimension remains limited. There is an urgent need to better understand its scale, dynamics and evolving advertisement trends, as the digital footprints offer opportunities to monitor and combat the trafficking in real-time.

To address this knowledge gap, a scalable longitudinal study is being conducted at the NYU VIDA Center by systematically investigating wildlife product advertisements across diverse online marketplaces. A data collection pipeline gathers these ads potentially containing information on animals and derivatives. This dataset enables mapping the nature of digital wildlife trade, the key illegally traded species and products, and how this cybercrime environment evolves over time.

Advanced data techniques like natural language processing and machine learning are applied to uncover patterns, trends and hotspots of illegal online wildlife trade activity. The analysis aims to contribute evidence-based strategies for combating trafficking, promoting biodiversity conservation

Authors' addresses: Aditya Gouroju, New York University, USA; Vamshi Naik Vislavath, New York University, USA; Aryan Prasad, New York University, USA.

ACM XXXX-XXXX/2024/5-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

and safeguarding endangered species. Subsequent sections review related work, explain the data and methodology, present key results and analyses, and discuss broader implications for tackling wildlife cybercrime.

2 PROBLEM FORMULATION

The core dataset comprises ads extracted in real-time from e-commerce websites, offering a unique insight into the digital manifestations of illegal wildlife trade. Employing advanced analytics and distributed computing techniques with PySpark, our aim is to unveil crucial insights to combat online wildlife trafficking effectively. This analysis will equip stakeholders with evidence-based intelligence to address cybercrime endangering vulnerable species.

The dataset originates from a complex process involving ETL (Extract, Transform, Load) pipelines, meticulously crafted to collect information through web crawling methodologies. These pipelines are instrumental in capturing various details, serving as a foundational step in data acquisition. Leveraging a zero-shot classifier applied to the "predicted label" column, the dataset efficiently categorizes products based on their potential association with animals. Despite this categorization, the dataset's scope is somewhat limited, lacking in-depth insights into transactional nuances or comprehensive seller profiles. Consequently, the dataset necessitates supplementary analytical approaches to thoroughly explore the breadth of topics embedded within its contents.

Our primary objective is to conduct an in-depth analysis of the data collected through the VIDA Center's pipeline, revealing patterns, trends, and key characteristics of online wildlife trade. While the dataset includes ads from various online marketplaces potentially related to the sale of animals and animal-derived products, filtering out irrelevant ads poses a significant challenge. Therefore, our focus lies in developing a sophisticated pipeline capable of extracting pertinent information accurately, enabling us to address specific queries regarding wildlife-related ads.

We also try to answer these specifc queries:

(1) **Identification of Frequently Traded Species**: The primary goal is to determine the species most frequently traded in online wildlife trafficking advertisements. By analyzing the frequency of species mentioned across ads, we aim to identify the most targeted species and assess the scale of their exploitation in the digital marketplace.
(2) **Correlation between Ads and Derivatives**: We seek to explore whether there is a correlation between the number of ads and the diversity of derivatives associated with traded species. For instance, we aim to investigate whether certain species, such as sharks, are predominantly advertised for specific derivatives (e.g., teeth) compared to others like ostriches, which may have a broader range of products (e.g., skin, meat).
(3) **Geographical Distribution Analysis**: This aspect of the analysis involves mapping the geographical distribution of species trafficked based on the origin of the advertisements. By visualizing the spatial patterns of wildlife trade, we aim to identify hotspots and transit routes for illicit wildlife trafficking activities.
(4) **Identification of Potentially Illegal and Endangered Species**: Using data mining and classification techniques, we aim to identify potentially illegal and endangered species advertised on online platforms. We seek to uncover patterns of concentration for specific species sold on these platforms and assess the extent of their listing in relation to conservation status.
(5) **Price Range Analysis**: We intend to analyze the price range of species sold online to understand the economic incentives driving wildlife trafficking. By examining variations

in pricing across different species, we aim to identify high-value targets and assess their vulnerability to exploitation.

(6) **Duplicate Advertisements ? Who are the biggest sellers? Are there connections between sellers that could characterize collaboration :** Identify top wildlife sellers online and explore potential collaborative networks. Analyze data to pinpoint dominant sellers. Employ data analysis to detect connections indicating collaboration, such as similar titles or pricing strategies. This investigation will unveil key market players and their interactions, providing insights into the dynamics of wildlife trade.

3 RELATED WORK

The proliferation of online marketplaces has significantly altered the landscape of wildlife trafficking, making it a pressing issue in conservation and law enforcement. The ease of access to a global market allows traffickers to expand their operations beyond traditional physical boundaries, leveraging the anonymity and reach of the internet. Studies have shown that digital platforms, including e-commerce sites and social media, are frequently utilized for advertising and selling wildlife products, ranging from live animals to derived products like ivory or fur.

Technological advancements play a dual role in this scenario. On one hand, they facilitate the illegal trade through simplified logistics and obscured identities; on the other, they provide new tools for monitoring and enforcement. Techniques such as data mining, natural language processing, and machine learning are increasingly applied to identify patterns, track online activities, and predict hotspots in wildlife trafficking.

Social media platforms have been identified as catalysts in the wildlife trade, not just for the sale of items, but also for the cultivation of consumer interest. Traffickers use these platforms to reach vast audiences, often employing coded language and imagery to evade detection by authorities and algorithms designed to flag illegal activities.

The effectiveness of existing legal and regulatory frameworks in combating online wildlife trafficking remains a significant challenge. International and local laws struggle to keep pace with the rapid pace of technological change, resulting in enforcement gaps. Moreover, the jurisdictional complexities of the internet make it difficult to prosecute offenders operating across multiple countries.

Conservation efforts are increasingly focusing on reducing demand through public awareness campaigns. Educating potential buyers about the ethical and legal ramifications of purchasing trafficked wildlife products is seen as a key strategy in decreasing market demand. Such efforts are crucial in altering consumer behavior, which is often driven by a lack of awareness about the species' conservation status or the impact of their extinction.

4 METHODS, ARCHITECTURE AND DESIGN

Fig. 1. gives a diagrammatic sketch of the pipeline that we will roughly use to get the results.

(a) Data is obtained from Minio buckets, which are stored in Parquet files.
(b) Subsequently, the data is uploaded to HDFS file storage to make it accessible for Spark programs.
(c) Utilizing the uploaded data, Spark offers a multitude of possibilities for data analysis and processing. This includes but is not limited to tasks such as data transformation,

Fig. 1. Description of the first result.

aggregation, machine learning model training, statistical analysis, and generating insights through exploratory data analysis.

As part of our analysis, we will try to answer the following questions. The questions and their formulated procedures are listed below:

4.1 What are the species most frequently traded?

To determine the species most commonly traded, we'll first use a zero-shot classification model to determine if the advertisement involves an animal or not, and if it does, it is classified the corresponding animal species in the ad using the title and description. We further utilize a large language model like GPT-3.5 to analyze the title and verify the result of the zero-shot classification model. So we have the data with what animal the ad corresponds to. We use this to calculate the frequency of each animal giving us the desired result.

4.2 Does the number of ads correlate with the number of derivatives, e.g., for sharks, we often find only teeth for sale, while for ostriches, people sell skin and meat?

We'll investigate whether there's a correlation between the number of advertisements and the presence of derivative products for certain species. To do this, we'll first refer to the list of unique animals that the crawler has utilized to scrape relevant ads from the internet, which is available in the root folder. Next, we'll tally the total number of advertisements associated with each animal in our dataset. Subsequently, we'll determine the count of ads that feature derivative products related to that particular animal. For instance, we'll examine if there are ads specifically offering shark teeth or other derivatives.

4.3 What is the geographical distribution of species trafficked?

We aim to ascertain the geographical distribution of trafficked species utilizing data from columns such as 'country', 'ships to', 'latitude', and 'longitude'. Our approach involves leveraging these values to map out the distribution of each species. To initiate this process, we will utilize the latitude and longitude coordinates provided in the dataset to create a comprehensive map of species distribution. This will entail interfacing with external APIs to determine the corresponding regions based on the coordinates obtained.

4.4 How many potentially illegal and endangered unique species can be identified, and is there a pattern of concentration for specific species sold on these online platforms?

- Firstly, we obtain the list of illegal and endangered unique species from reputable sources such as CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) or IUCN (International Union for Conservation of Nature).

- Subsequently, using this list, we gather the number of advertisements corresponding to each species and aggregate them based on the species.
- By leveraging data such as price, URLs, geographical coordinates, etc., we can discern various patterns within the dataset for these endangered species.
- These patterns might include the average price for a particular species and the types of products being sold. Additionally, we can explore the geographical distribution of the endangered species and their origins.
- If an animal name is absent from the list, we will utilize ChatGPT's aptitude to retrieve information regarding whether the animal is endangered or not.

4.5 What is the price range of the species sold online?

- To address this inquiry, we use the data from the 4.1 subpart having had the animal species defined.
- The data also contains the prices of the ads extracted from the websites, we convert this data to a standard currency (USD) using an API call to convert all the currencies to USD.
- The data is then grouped by the species, and the minimum and maximum price of the ad is calculated for each species.

4.6 Who are the biggest sellers? Are there connections between sellers that could characterize collaboration (or organized crime)?

- Begin by aggregating the data by sellers, allowing us to obtain a clear overview of the number of listings per seller and the range of species they sell.
- Following this, extract the relevant data containing information about the sellers, such as their name, URL, and location.
- Subsequently, identify commonalities among sellers, such as shared locations or similar product listings, which may indicate potential connections or collaboration.
- Additionally, consider factors such as high sales volume, calculated based on the sum of prices of listings by a seller, to gauge the magnitude of their operations. Assess the frequency of listings by each seller and whether they engage in the trade of illegal species to further evaluate their prominence and practices.
- Employing fuzzywuzzy similarity analysis enables us to discern the likeness between provided titles, facilitating the correlation of sellers based on this similarity metric. This method aids in identifying potential connections or collaborations among sellers, contributing to a deeper understanding of their interactions within the marketplace.

5 RESULTS

By visualising data with various levels and filters, we discovered patterns and factual information that can only be obtained from datasets. We created dynamic dashboards, each depicting a story and trend that we can observe. Figures below show information integrated in multiple formats for easy understanding and visualisation. Filtering elements/graphs allows for dynamic display of findings in other graphs.

(a) **Frequency of Traded Species**: The analysis of advertisement data reveals a striking prominence of **Ostrich-related** ads, suggesting a substantial commercial interest in this species. This observation prompts investigation into the underlying drivers of this phenomenon. Understanding the economic, socio-cultural, and regulatory factors shaping the

prevalence of Ostrich-centric advertising offers valuable insights into consumer behavior, market dynamics, and ethical considerations surrounding wildlife trade.
Exploring the economic significance of Ostrich trade, alongside societal perceptions and regulatory frameworks, provides a comprehensive understanding of the commercial landscape. It highlights the interplay between market forces, cultural preferences, and legal constraints in shaping advertising practices related to Ostriches. Moreover, assessing the environmental implications of heightened commercial interest in Ostriches underscores the importance of sustainable resource management and biodiversity conservation in the face of increasing human-animal interactions.
By delving into these interconnected dimensions, researchers can gain deeper insights into the complex relationship between advertising, wildlife trade, and conservation efforts. This nuanced understanding is essential for devising effective policies and strategies that balance economic interests with environmental sustainability and ethical considerations in the commercial exploitation of wildlife resources.

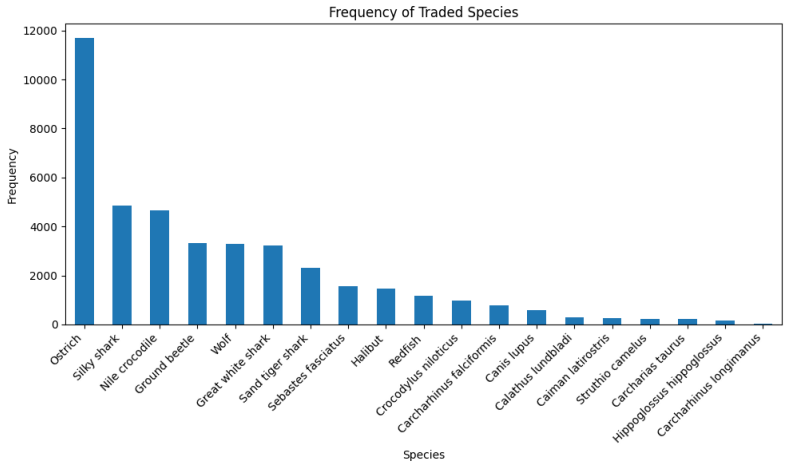

Fig. 2. Frequency of Traded Species

(b) **Top seller distribution:** Our further analysis methodology delves into the seller landscape within the wildlife trade, focusing on two key metrics: the number of listings and the total price of their offerings. Firstly, examining the number of listings attributed to each seller allows us to identify prolific actors within the trade network. Sellers with a high volume of listings indicate significant involvement in the trade, potentially operating at a large scale across various regions. This insight helps law enforcement agencies and conservation authorities prioritize their efforts to disrupt major trafficking networks effectively.
Secondly, by assessing the total price of listings associated with each seller, we gain insights into the economic dimensions of wildlife trafficking. Sellers commanding higher prices for their offerings may be involved in trading high-value or endangered species. This observation highlights the presence of lucrative markets for rare or exotic wildlife products, providing valuable intelligence for targeting enforcement actions and disrupting illicit trade flows.
Moreover, analyzing the species and market trends revealed through this approach offers additional insights into wildlife trafficking dynamics. Understanding which species are

being trafficked and their market value helps stakeholders develop targeted interventions to combat trafficking networks and protect vulnerable species. By tracking shifts in market demand and supply, authorities can adapt their strategies to effectively combat wildlife crime and preserve biodiversity in the face of evolving challenges.

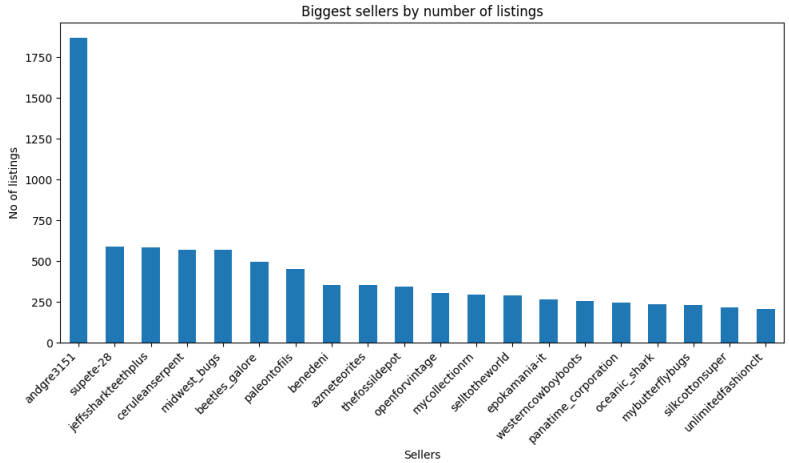

Fig. 3. Biggest Sellers by number of ads

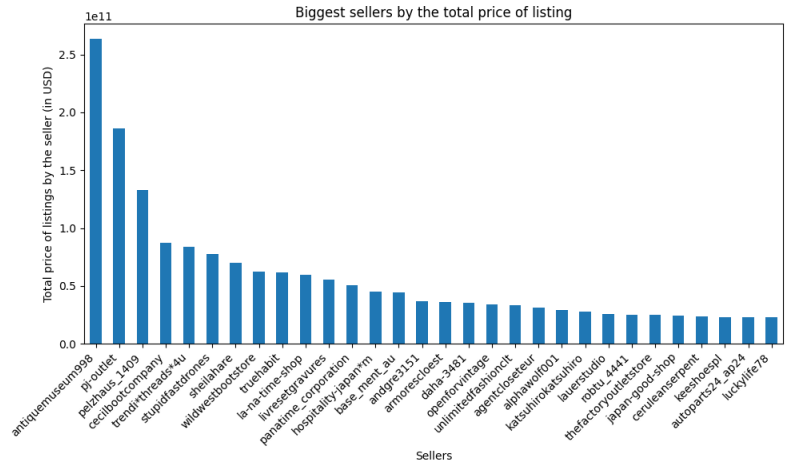

Fig. 4. Biggest Sellers by total price of listings

The variance in seller names within the top sellers list unveils the nuanced nature of wildlife trafficking. While some sellers overtly signal their involvement in the trade of animal products, others adopt less conspicuous aliases like "*benedeni*." This dichotomy suggests a spectrum of operations, with overtly named sellers likely specializing in distinct niches while those with ambiguous names **may** engage in more covert or diversified activities, potentially evading regulatory scrutiny.

Understanding these naming conventions offers critical insights for enforcement strategies. Specialized sellers may represent high-value targets for disruption, while those with

ambiguous names necessitate sophisticated monitoring to unveil clandestine operations. By comprehensively analyzing the nomenclature of wildlife traffickers, authorities can develop targeted interventions to combat illicit trade networks effectively.

(c) **Analysis on the domain distribution:**

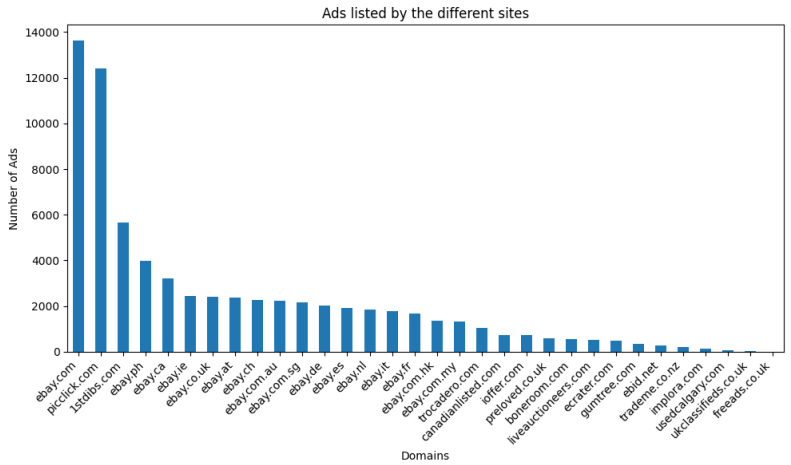

Fig. 5. Domain Distribution

We are able to observe that high amount of ads are concentrated on Ebay. We will further analyze by granularly analyzing the correlation of websites with the sellers.

(d) **Analysing domain distribution of the top sellers:** The concentration of top sellers predominantly on *eBay.com* suggests significant market dominance for wildlife products on this platform compared to others. This finding underscores the platform's prominence as a preferred marketplace for wildlife trafficking activities. Moreover, it raises questions about the regulatory landscape governing online wildlife trade, with potential implications for enforcement effectiveness and conservation efforts.

Further insights can be gleaned by examining the specific characteristics of eBay.com that might facilitate its attractiveness to wildlife traffickers. One plausible explanation could be the relative leniency of eBay's regulations compared to other popular e-commerce platforms. Differences in seller verification processes, listing policies, and enforcement mechanisms across platforms may create varying levels of risk and opportunity for traffickers. Additionally, the accessibility, user base, and international reach of eBay.com may offer traffickers a broader audience and higher transaction volumes, further incentivizing their preference for the platform.

Understanding the dynamics driving the concentration of wildlife trafficking activities on eBay.com is crucial for developing targeted interventions to combat illicit trade. Assessing the effectiveness of existing regulatory measures and identifying gaps in enforcement can inform policy reforms and platform-specific interventions aimed at mitigating the facilitation of wildlife crime online. Furthermore, collaboration between law enforcement agencies, regulatory authorities, and e-commerce platforms is essential for enhancing monitoring, detection, and deterrence mechanisms to safeguard against the exploitation of online marketplaces for illicit wildlife trade.

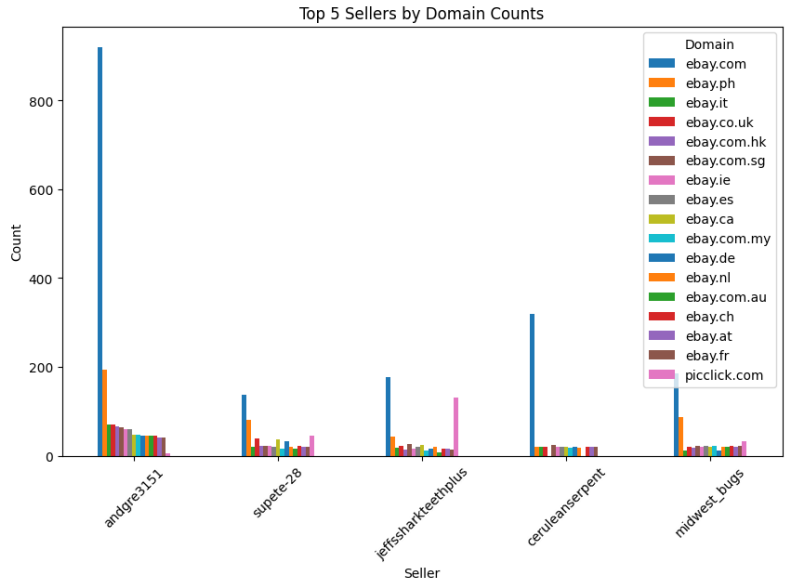

Fig. 6. Top sellers with their ad distribution across platforms

(e) **Product distribution by animal:** To get in more detail what products are mainly sold for a particular animal, we find the top products for an animal:
Identifying the top-selling product types corresponding to specific animals can yield valuable insights into consumer preferences, market demand, and trafficking trends within the wildlife trade. By analyzing the correlation between animal species and the associated product types, several key insights can be derived:

(i) *Consumer Preferences and Market Demand:* Understanding which product types are most commonly associated with specific animals provides crucial insights into consumer preferences and market demand. For example, if the top-selling product type for a particular animal is its skin or fur, it suggests a high demand for fashion and luxury goods derived from that species. Conversely, if the top-selling product type is meat or body parts used in traditional medicine, it indicates demand for consumable or medicinal products.

(ii) *Species Vulnerability and Conservation Priorities:* The identification of top-selling product types can highlight the most exploited or vulnerable species within the wildlife trade. Animals associated with high-demand product types may face greater conservation threats due to overexploitation or illegal harvesting. This information enables conservationists and policymakers to prioritize conservation efforts and implement targeted interventions to protect endangered species and their habitats.

(iii) *Trafficking Trends and Enforcement Strategies:* Analyzing the correlation between animal species and top-selling product types provides valuable intelligence on trafficking trends and modus operandi employed by wildlife traffickers. For example, if certain product types consistently rank as top sellers across multiple species, it may indicate organized trafficking networks specializing in the trade of specific wildlife products. This insight can inform law enforcement agencies and conservation authorities in devising effective

strategies to combat illicit trade, disrupt trafficking networks, and dismantle supply chains.

(iv) *Policy and Regulatory Interventions:* The identification of top-selling product types can inform policy and regulatory interventions aimed at curbing wildlife trafficking and protecting biodiversity. For instance, if certain product types are driving the exploitation of endangered species, policymakers may consider implementing stricter regulations, trade bans, or consumer awareness campaigns to reduce demand and deter illegal trade activities.

In conclusion, analyzing the correlation between animal species and top-selling product types provides valuable insights into consumer behavior, species vulnerability, trafficking trends, and policy interventions within the wildlife trade. By leveraging this information, stakeholders can develop targeted strategies to address conservation challenges, combat illegal wildlife trade, and promote sustainable use of natural resources.

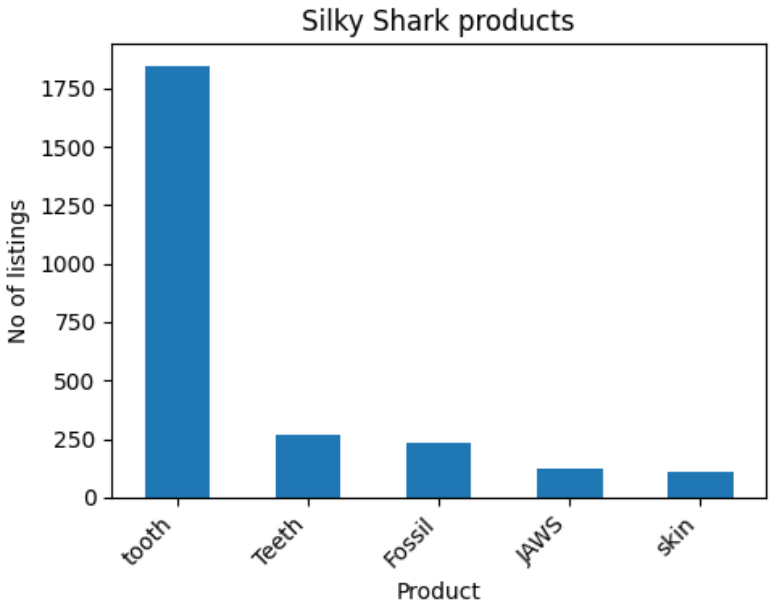

Fig. 7. Silky shark top products

(f) **Similar Sellers:** The depicted process entails a systematic journey towards unraveling similarities among sellers, focusing on their product offerings. Commencing with data segmentation based on the type of animal being sold, the subsequent phase employs a sophisticated fuzzy matching technique. This technique harnesses the power of the Levenshtein Distance metric to meticulously compare the textual compositions of sellers' listings. A stringent similarity threshold, calibrated at 95%, acts as a gatekeeper to ensure the fidelity of matches. Following this, an aggregation mechanism tallies the number of sellers meeting this stringent criterion. The resultant dataset finds its visual manifestation within a Tableau dashboard, a dynamic canvas where sellers boasting the highest semblance in their listings emerge as focal points. This meticulously crafted methodology not only

expedites the process of categorizing sellers but also lays a robust foundation for informed decision-making in the realm of sales and marketplace management.

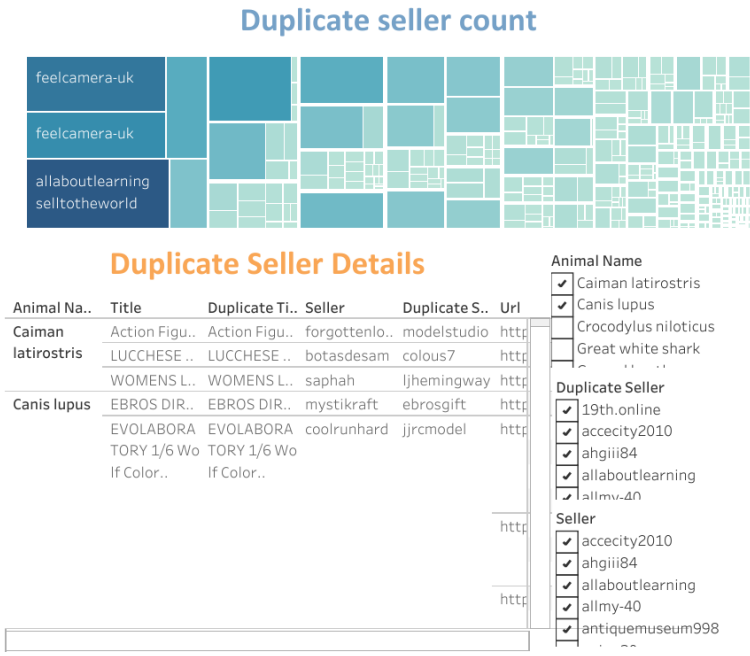

Fig. 8. Seller Similarity

(g) **Distribution of Endangered Species and Geographic data:** The image illustrates the conservation status of animals sourced from the IUCN Red List of Threatened Species, indicating whether a species is endangered. Additionally, it includes a geographic map pinpointing the locations represented in the dataset. This presentation provides valuable insights into the status of various animal species, crucial for conservation efforts and biodiversity management. By visually depicting the distribution of endangered species and their geographic context, stakeholders gain a clearer understanding of areas requiring heightened conservation measures. This visualization aids researchers, policymakers, and conservationists in identifying priority regions for intervention and resource allocation. Through effective communication of complex data, such visualizations serve as powerful tools for raising awareness and fostering informed decision-making towards the protection of endangered species and their habitats.

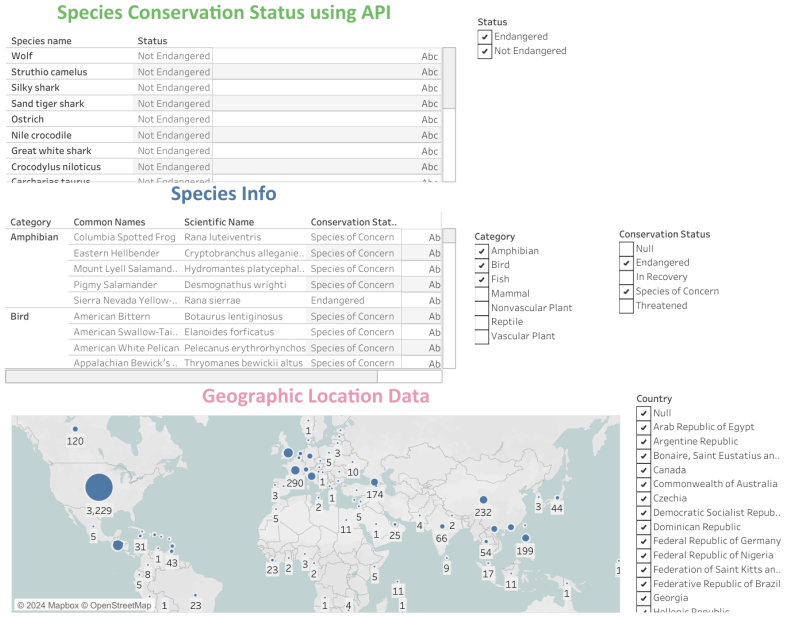

Fig. 9. Endangered species and Geographic data

(h) **Analysis of price range of animals** : This analysis provides the price distribution of various animals. The data is presented in terms of minimum, maximum, and average prices in USD. The provided data offers an insightful look into the price range of various animals. Each animal is associated with a minimum, maximum, and average price, based on market values.

For instance, the Halibut has a minimum price of approximately $1.18, a maximum price of around $1515.11, and an average price of about $79.48. This suggests a wide range of prices for this particular species, indicating a significant variation in the market value of individual specimens, possibly due to factors such as size, age, or other quality indicators. On the other hand, the Nile Crocodile and the great white shark exhibits a much larger price range, with the the maximum price reaching up to $480,000 and $1,200,000 respectively, and an average price of approximately $3299.79 and $3977.68 respectively. This high values compared to other animals could be indicative of the high demand the products of these specific animals have in the market.

In contrast, species like the Sebastes Fasciatus and the Calathus Lundbladi have relatively lower price ranges and averages, suggesting that these species might be more common to obtain or less in demand in the market. When we combine the results from frequency of traded species with price ranges The data reveals a fascinating dynamic in the wildlife trade market. The Ostrich, despite its high trading frequency, has a relatively lower average cost compared to species like sharks and crocodiles. These latter species, while traded less frequently, command significantly higher average and maximum prices. This discrepancy in trading frequency and price points underscores the market demand for specific species. It suggests that the allure of high profits may incentivize hunters or sellers to procure

products from species like sharks and crocodiles, even in the face of legal prohibitions. Consequently, these species with high-value products but lower trading frequencies face the highest threat of extinction/endangerment.

This analysis highlights the complex interplay between market forces and conservation efforts. It underscores the need for robust regulatory frameworks to curb illegal trade and protect endangered species. Furthermore, it emphasizes the importance of raising public awareness about the ecological implications of wildlife trade. By fostering a market that values sustainability, we can help ensure the preservation of biodiversity while still accommodating economic interests.

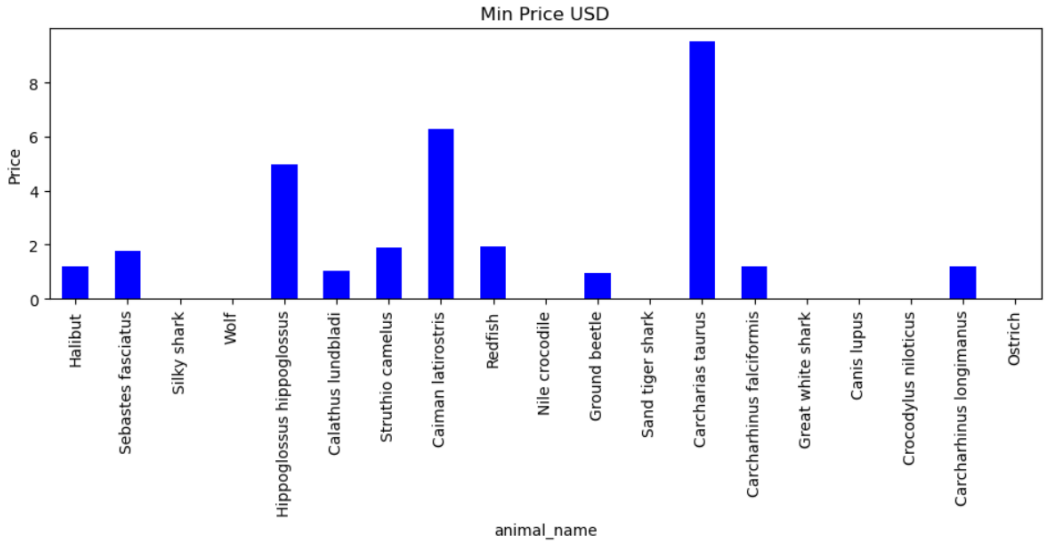

Fig. 10. animals by their minimum price

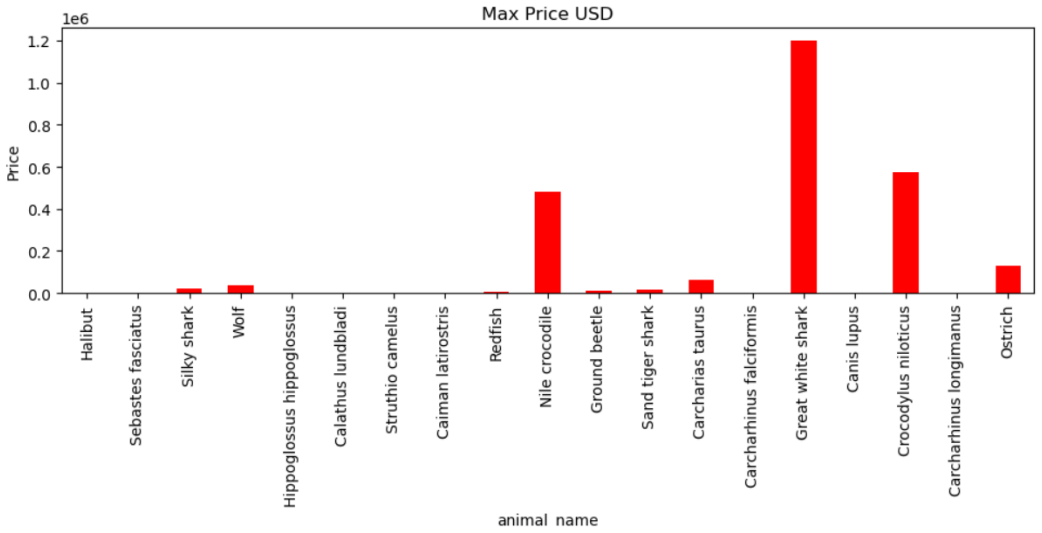

Fig. 11. animals by their maximum price

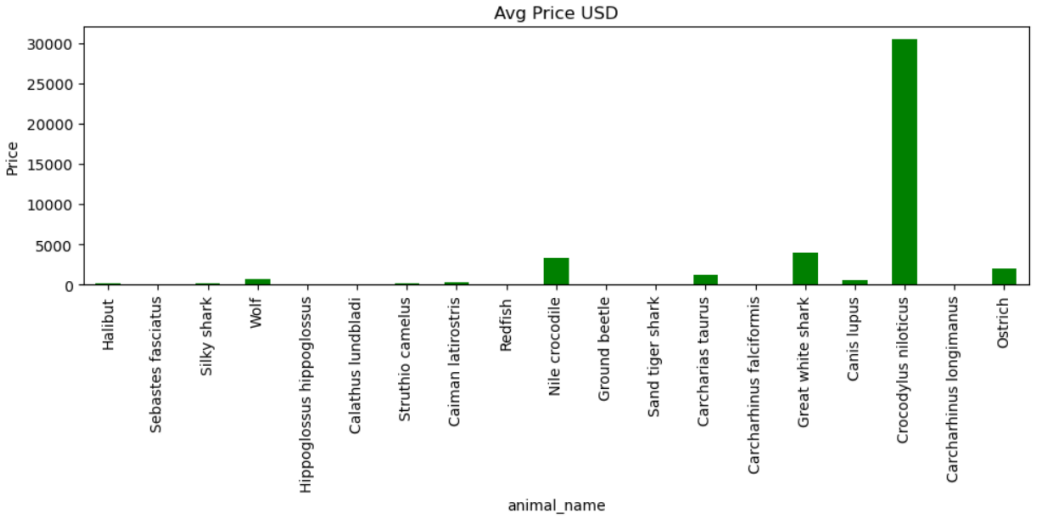

Fig. 12. animals by their average price

(i) **Distribution by Animal Product type** : The distribution can help us understand the market demand for different types of animal products. For instance, tooth of animals is more prevalent, it suggests a higher demand for that product in the market. The distribution can guide regulatory focus. Authorities can prioritize monitoring and regulation efforts based on the prevalence of different product types. The main product types that are traded are tooth, leather, skin and feathers among other types.

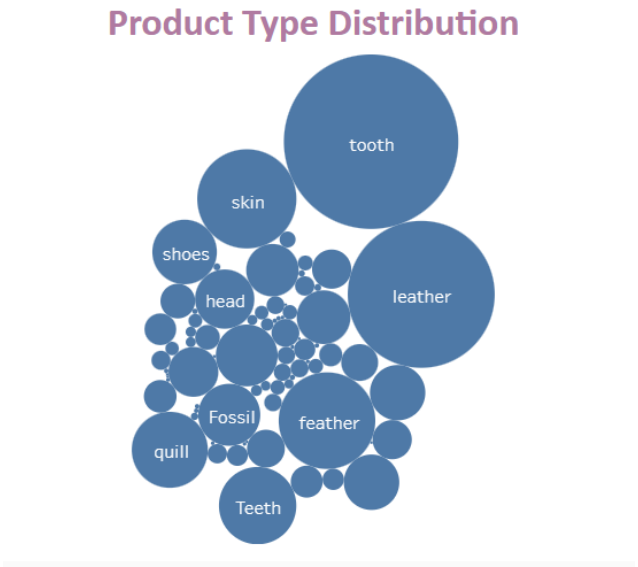

Fig. 13. Distribution by Animal Product type

6 ACKNOWLEDGEMENTS

Thank you NYU Tandon and Prof. Juliana Freire for providing us with the datasets, dataproc resources, pipeline information and implementations. We really appreciate the help and guidance provided to us through out this project.

7 REFERENCES

[1] Juliana Barbosa, Sunandan Chakraborty, "The illegal wildlife digital market: an analysis of Chinese wildlife marketing and sale on Facebook," InfoWild '23, October 22, 2023. https://drive.google.com/file/d/1tCHJdNtie96Fkm9NSvUTZiq30s3wlJ/view

[2] Q. Xu, M. Cai, and T. K. Mackey, "A Flexible and Scalable Approach for Collecting Wildlife Advertisements on the Web" Environmental Conservation, vol. 47, no. 3, pp. 206–212, 2020. doi:10.1017/S0376892920000235

[3] Reuter, P. and O'Regan, D. (2017) 'Smuggling wildlife in the Americas: scale, methods, and links to other organised crimes', Global Crime, 18(2), pp. 77–99. doi: 10.1080/17440572.2016.1179633.

[4] Stringham OC, Moncayo S, Hill KGW, Toomes A, Mitchell L, Ross JV, et al. (2021) Text classification to streamline online wildlife trade analyses. PLoS ONE 16(7): e0254007. https://doi.org/10.1371/journal.pone.0254007