# Data Collection and Wrangling

# List of Registered Veterinary Medicines

SUBMITTED BY:
Nawaf Abdulrhman Alageel
Albadar Ibrahim Almaymani
Mohammed Khalid Altufayhi
Abdullah Mansour Habit

Supervised by: Dr. Hassan Ruwayziq Rudayd Alhuzali

DEPARTMENT OF (DATA SCIENCE)
COLLEGE OF COMPUTERS
UMM AL-QURA UNIVERSITY

2024

# ABSTRACT

This document presents an in-depth analysis of the "List of Registered Veterinary Medicines Products," a critical endeavour in understanding the complexities of the veterinary pharmaceutical industry. The project is centered around evaluating the diverse array of veterinary drugs, their compositions, applications, and the regulatory frameworks governing them. This analysis is pivotal for multiple stakeholders, including veterinarians, pharmaceutical companies, regulators, and policymakers, offering insights crucial for informed decision-making.

The primary focus of the project is to dissect the intricate details of veterinary medicine composition. It closely examines the active ingredients, their combinations, and the intended therapeutic uses, shedding light on the evolving landscape of veterinary pharmaceuticals. This aspect of the study is particularly important for understanding how drugs are tailored to meet specific health needs of various animal species, and how these needs change over time.

Moreover, the project delves into the trends and patterns prevalent in the veterinary medicine industry. This includes an analysis of the most common types of medicines, the frequency of their usage, and emerging trends in drug development. Such insights are invaluable in predicting future market shifts and understanding the current state of veterinary healthcare.

Compliance with regulatory standards is another crucial aspect covered in the document. It explores how veterinary medicines are regulated, the processes involved in ensuring their safety and efficacy, and the implications of these regulations for manufacturers and users. This is crucial in an industry where the health and wellbeing of animals, and indirectly humans, are at stake. The project also examines the market dynamics of veterinary medicines. This involves an analysis of the distribution networks, market shares of different manufacturers, and the withdrawal periods necessary for ensuring that animal products entering the human food chain are safe. Understanding these dynamics is essential for stakeholders to make strategic decisions that align with market trends and regulatory requirements. Furthermore, the implications of veterinary medicines on global and local markets, animal welfare, and public health are critically assessed. The study underscores the interconnectedness of animal health with public health, emphasizing the need for a holistic approach to veterinary medicine regulation and use.

**Problem Questions for Analysis:**

1. Market Dynamics: How do the types of veterinary medicines, their forms and administration methods vary across different animal types, and what does this indicate about market trends?

2. Geographical and Manufacturing Insights: What patterns emerge in the distribution of manufacturers and marketing companies globally, and how does this affect the veterinary medicine market?

3. Regulatory Compliance and Safety: Are there any notable trends or outliers in withdrawal periods across different medicines, and what implications might these have for animal and public health safety?

# INTRODUCTION

The dataset titled "List of Registered Veterinary Medicines Products" is a comprehensive collection of data pertaining to veterinary medicines registered for use. The choice of this dataset is rooted in its significance for various stakeholders in the veterinary pharmaceutical industry, including veterinarians, pharmaceutical companies, regulatory bodies, and pet owners. The insights gleaned from this dataset can drive informed decisions, ensure compliance with regulatory standards, and ultimately contribute to the well-being of animals.

## Rationale for Selecting the Dataset

1- Regulatory Compliance and Safety: In the context of increasing scrutiny over animal healthcare products, understanding the composition, usage, and regulatory status of veterinary medicines is critical. This dataset offers an extensive view of these aspects, assisting in ensuring compliance with regulatory standards and safety protocols.

2- Market Analysis and Development: For pharmaceutical companies, this dataset provides a window into current market trends, popular compounds, and competitive analysis. It can aid in strategic decision-making for product development and marketing.

3- Veterinary Practice and Animal Health: Veterinarians can leverage this dataset to stay updated with available medicines, their indications, and withdrawal periods, thereby enhancing the quality of animal healthcare.

4- Public Health and Research: Given the interconnection between animal health and public health, especially in the context of zoonotic diseases, this dataset is a valuable resource for researchers and public health officials. It helps in monitoring and managing the use of antibiotics and other critical drugs in animals.

5- Ensuring Availability: The dataset contributes to assessing the global supply chain resilience of veterinary medicines. This is essential for understanding potential disruptions, ensuring the continuous availability of crucial medications, and mitigating supply chain risks.

6- Strategic Planning: Stakeholders, including pharmaceutical companies and regulatory bodies, can use this information to strategically plan for contingencies, ensuring that animal healthcare products are accessible even in challenging circumstances, such as pandemics or geopolitical disruptions.

## Insights Gained from the Dataset

1- Drug Composition and Use: Analysis of the generic names and strengths of the drugs provides insights into the most commonly used compounds in veterinary medicine, their dosages, and their formulations.

2- Trends in Veterinary Medicine: By examining the types of medicines registered, one can discern trends in veterinary healthcare, such as the rise in the use of certain classes of drugs or shifts in treatment protocols.

3- Regulatory Compliance: Information on the registration number, authorization status, and withdrawal periods are vital for ensuring that the medicines used are legally compliant and safe for animal consumption.

4- Global and Local Market Insights: The dataset includes information about the country of manufacture and the marketing company, offering a perspective on both global and local market dynamics.

5- Impact on Animal and Public Health: Understanding the withdrawal periods and routes of administration is crucial for animal welfare and public health, as it affects the residues of drugs in animal products consumed by humans.

6- Continuous Monitoring: The dataset supports ongoing monitoring of adverse events, contributing to the improvement of veterinary medicines' safety profiles over time. This iterative process helps in addressing emerging issues and maintaining the overall safety of animal healthcare products.

7- Enhancing Safety Protocols: Insights into adverse reactions enable the development and enhancement of safety protocols in veterinary practice. Veterinarians and regulatory bodies can use this information to educate practitioners and pet owners, promoting responsible and safe medication use.

## Important Features of the Dataset

1- Registration No and Trade Name: Essential for identifying specific products and their legal status.

2- Generic Name and Strength Value: Provides information on the active ingredients and their concentration.

3- Dosage Form and Route of Administration: Indicates how the medicine is to be used, which is critical for effectiveness and safety.

4- Withdrawal Period: Important for food-producing animals to ensure that drug residues do not enter the human food chain.

5- Manufacturer and Marketing Company: Offers insights into the pharmaceutical companies involved and their market reach.

6- Authorization Status: Indicates whether the product is legally permitted for use.

7- Marketed Status: Provides information on the availability of the product in the market.

# RESULT AND ANALYSIS

**Imports Libraries**: Pandas for data manipulation, seaborn for statistical data visualization, NumPy for numerical operations, and matplotlib for creating plots.

**Loads the Dataset**: Uses the reads CSV function from pandas to load your dataset from a CSV file. Replace 'your_dataset.csv' with the actual file path or URL.

**Displays Information**: Prints the number of rows and columns in the dataset when we print it.

After **Imports**, **Loads** and **Displays.** will cleaning the data
we used "isnll" to check the null values, and with sum function to get the sum of the null values for each column.

1. **Strength value:** 65 missing values
2. **Unit of strength:** 100 missing values
3. **Dosage Form:** 4 missing values
4. **Route of Administration:** 58 missing values
5. **Volume:** 2 missing values
6. **Unit of volume:** 5 missing values
7. **Package type:** 0 missing values
8. **Shelf-life (mon):** 1 missing value
9. **Package size:** 128 missing values
10. **Legal status:** 0 missing values
11. **Product control:** 0 missing values
12. **Storage conditions:** 11 missing values
13. **ATCvet:** 119 missing values
14. **Target Animals:** 19 missing values
15. **Type:** 163 missing values
16. **Withdrawal Period (days):** 138 missing values
17. **Manufacturer name:** 0 missing values
18. **Country of Manufacturer:** 0 missing values
19. **Marketing Company:** 0 missing values
20. **Nationality:** 0 missing values
21. **MAH (Agent name):** 0 missing values
22. **Authorization status:** 0 missing values
23. **Marketed:** 209 missing values
24. **Marketing status:** 1 missing value

The dataset has a total of 1,205 missing values across various columns.

There are 14 columns that contain missing values:
1. **Strength value**
2. **Unit of strength**
3. **Dosage Form**
4. **Route of Administration**
5. **Volume**
6. **Unit of volume**
7. **Shelf-life (mon)**
8. **Package size**
9. **Storage conditions**
10. **ATCvet**
11. **Target Animals**
12. **Type**
13. **Withdrawal Period (days)**
14. **Marketed**

There are 10 columns that do not have any missing values:
1. **Registration No**
2. **Generic name**
3. **Trade name**
4. **Package type**
5. **Legal status**
6. **Product control**
7. **Manufacturer name**
8. **Country of Manufacturer**
9. **Marketing Company**
10. **Nationality**
11. **MAH (Agent name)**
12. **Authorization status**
13. **Marketing status**

Understanding which columns have missing values is essential for determining the appropriate strategies for handling missing data during data preprocessing.

I used the **dropna()** method to ensure the quality of the dataset for analysis. By removing rows with missing values, I aimed to create a cleaner dataset that is more suitable for statistical analysis and modeling. This process helps in maintaining the integrity of the data, ensuring that the analysis is based on complete and reliable information.
However, it's important to acknowledge that this approach might lead to a reduction in the dataset size, and careful consideration should be given to the potential impact on the representativeness of the data. The decision to drop missing values should be made based on the specific goals of the analysis and the nature of the missing data. In cases where the missing

Top 10 Dosage Forms
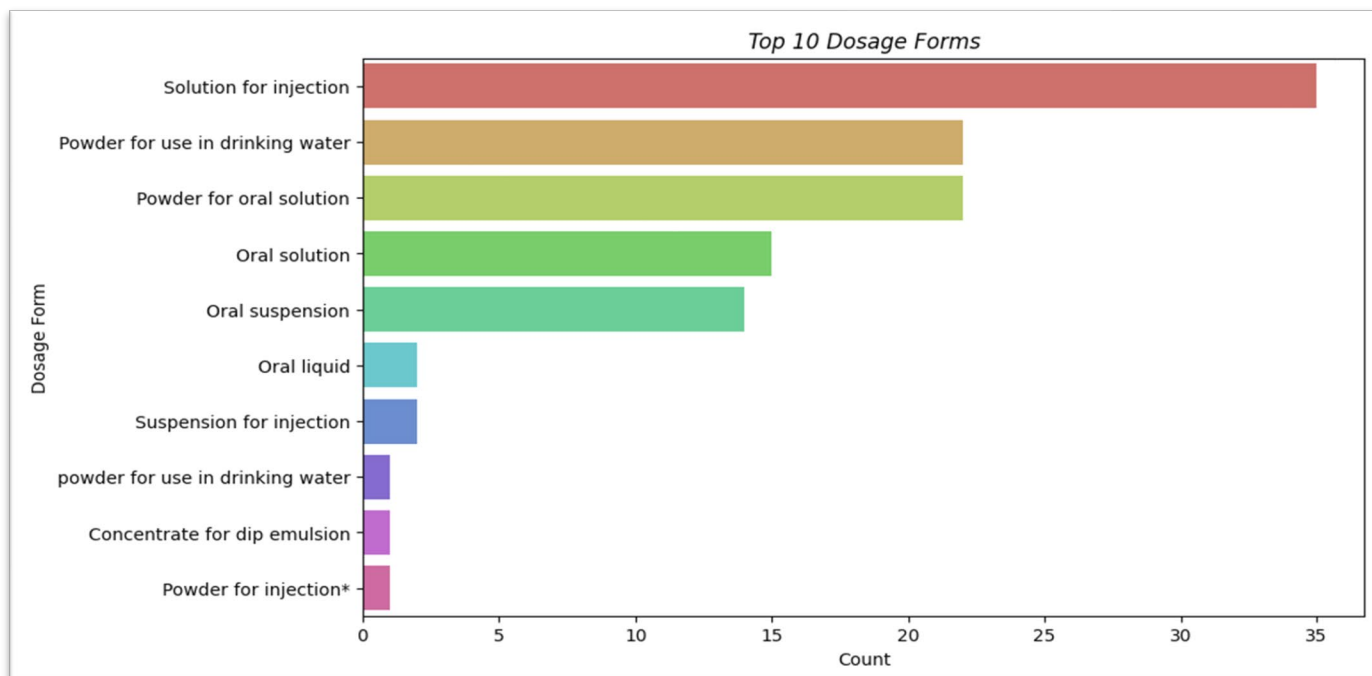
The image is a bar chart with horizontal bars of varying lengths. Each bar represents a different category of a dosage form, and the length of the bar indicates the count or frequency of each dosage form. The chart is intended to compare these frequencies at a glance.

values are critical or systematically related to the variables of interest, alternative methods like imputation or more advanced data cleaning techniques may be considered.

Pharmaceutical Dosage Forms in Veterinary Medicine:
An Analytical Bar Chart Depicting the Top 10 Dosage Forms Used in the Industry. This chart provides a clear quantitative comparison.
emphasizing the dominance of injection solutions and the substantial utilization of powders for drinking water and oral solutions.
 The distribution of these forms reveals preferences in veterinary practices and may reflect.
on the ease of administration, cost-effectiveness,
 or efficacy in delivering treatment to animals.
 Understanding these trends is vital for optimizing drug formulation and addressing the specific needs of various animal species in veterinary care.

_____

Top 10 Animal Types by Number of Unique Medicine Types
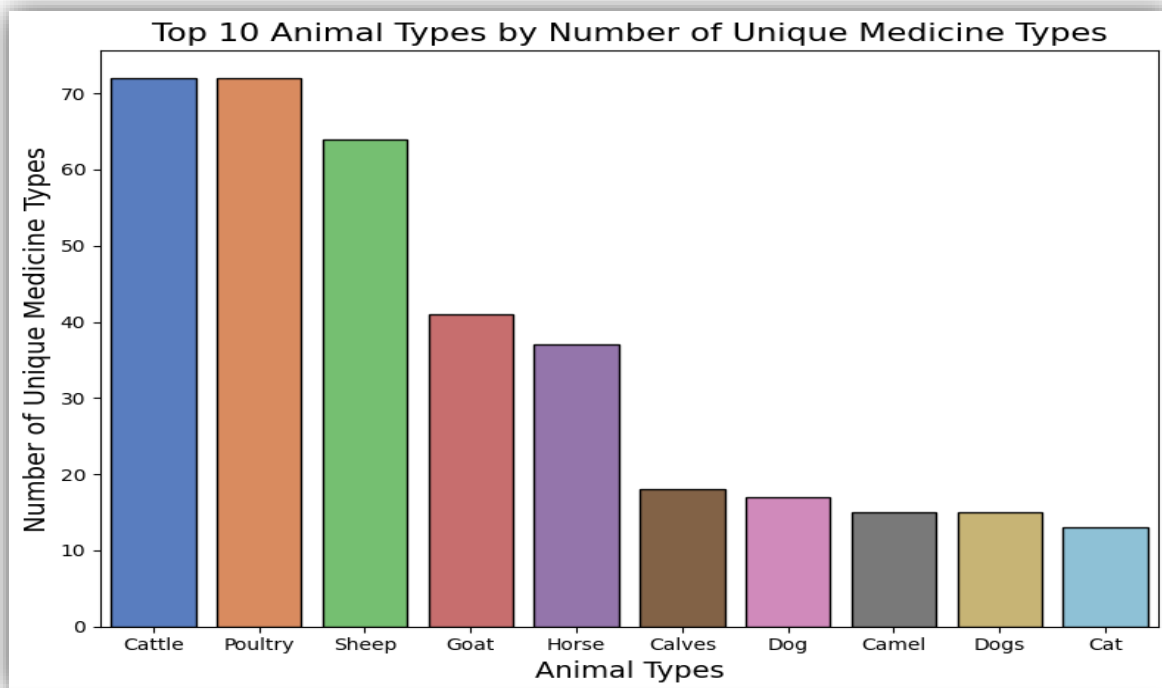
The image is a vertical bar chart. Each bar represents a different animal type, and the height of the bar corresponds to the number of unique medicine types associated with each animal. The chart aims to show a comparison across various animal categories based on the number of unique medicines for each.

The chart is a graphical representation detailing the distribution of veterinary pharmaceuticals tailored to various livestock. The striking prevalence of medications for cattle, poultry, and sheep indicates a significant emphasis on these animals in veterinary pharmaceutical science, likely due to their substantial roles in the agricultural industry. This variety points to a comprehensive approach to animal healthcare, ensuring that the specific needs of these animals are met with precise medical interventions.

Lesser heights of bars for goats, horses, calves, and dogs imply fewer medication types, which could reflect the relative scale of their populations or the specifics of their health issues in comparison to the leading categories. Camels and cats, with even fewer unique medicines, might represent niche areas within veterinary care, indicating a more focused demand for pharmaceuticals.

The chart functions as a concise yet powerful tool, encapsulating the scope of medicinal preparations available to livestock and serving as a critical asset for stakeholders in the veterinary field. It enables informed decision-making for the development of future treatments and the effective management of animal health, reflecting a dedicated effort to cater to the diverse needs of animal husbandry.
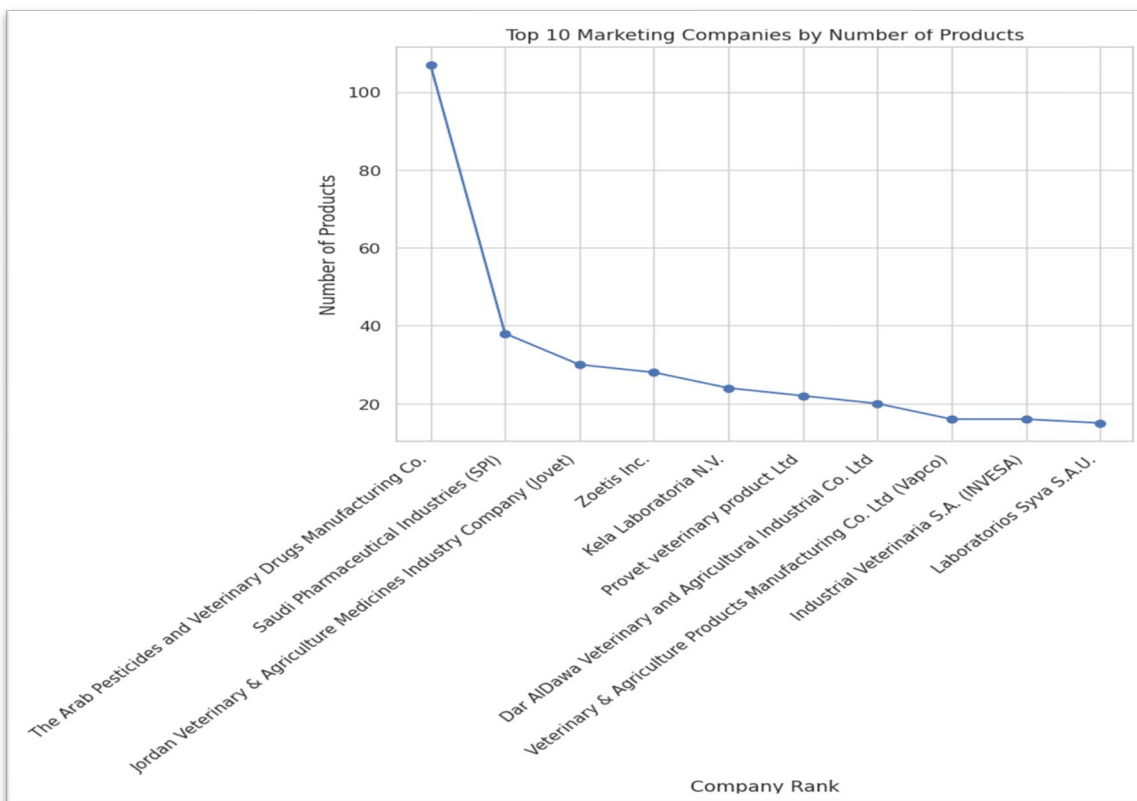
The image is a line graph depicting the number of products offered by the top 10 marketing companies. The graph shows a sharp decline between the first and second company and a gradual decrease thereafter.

The line graph illustrates a stark contrast in the scale of product offerings among the top 10 marketing companies in a particular industry. The leading company's product count is significantly higher than its competitors, surpassing the 90-product mark, which could imply a monopolistic stance or a diversified portfolio that eclipses the others. This outlier suggests that the company has a substantial market share, which may be the result of an aggressive expansion strategy, a wide array of product types, or a combination of both.

Descending from the peak, the product counts taper off sharply to the second-ranked company and then show a more gradual decrease from one company to the next. This pattern reveals a concentration of market presence at the top, followed by a more evenly distributed competition among the remaining companies. The latter companies, while less dominant, still demonstrate a significant presence, each contributing a range of products to the market.

Top 10 Manufacturers by Number of Products
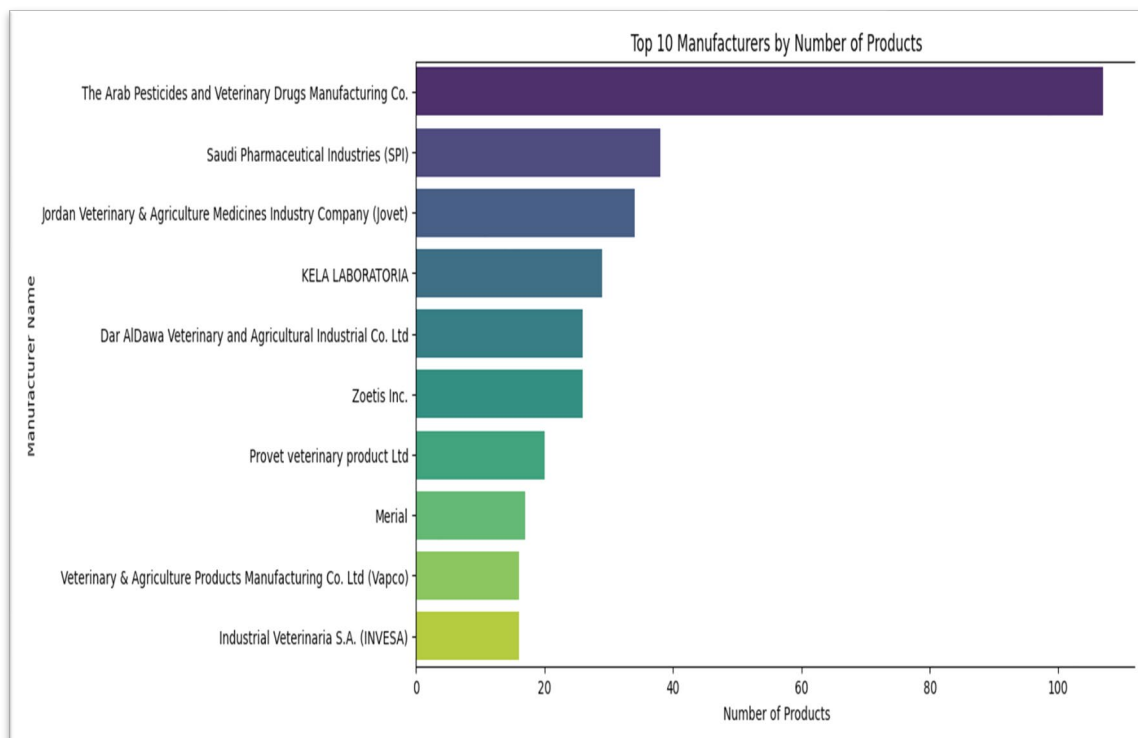
The image shows a horizontal bar chart. Each bar's length is proportional to the number of products offered by each manufacturer listed on the y-axis. The chart is designed to compare the product portfolios of the top 10 manufacturers, with the x-axis serving as a numerical scale from 0 to 100 to indicate the quantity of products.

The horizontal bar chart illustrates the number of products offered by the top 10 manufacturers in a certain sector, likely the veterinary pharmaceutical industry given the context. The manufacturer with the highest number of products, possibly over 90, occupies the top position, suggesting a significant market presence or a diversified product range. This could reflect a strong research and development focus, a wide distribution network, or a variety of product offerings to meet different market needs.

Following the leader, the other manufacturers show a descending order of product counts, with each subsequent bar representing a slightly lower number of products. This indicates a competitive market where several companies have substantial product offerings, but none match the scale of the leading company.

The manufacturers at the lower end of the chart, while still among the top 10, have significantly fewer products compared to the top manufacturer, which may point to a more specialized focus or a smaller market share.

Top 10 Marketing Companies by Number of Products
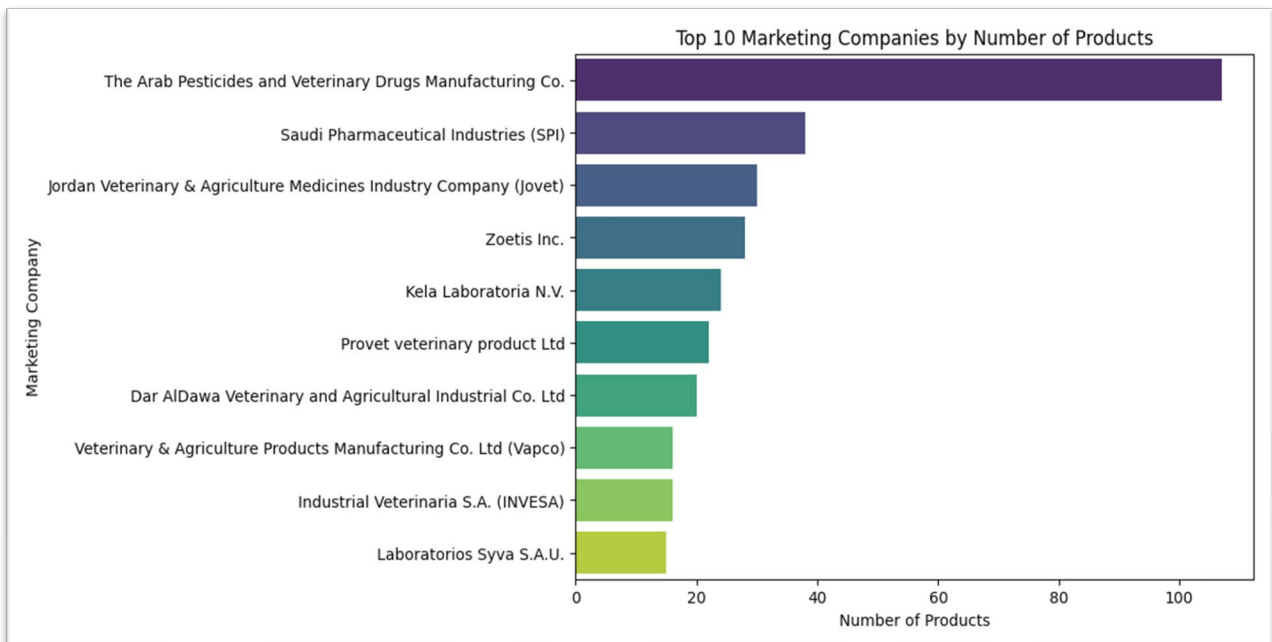
The image is a horizontal bar chart. Each horizontal bar represents a different marketing company, and the length of the bar corresponds to the number of products that company markets. The companies are listed in descending order from top to bottom based on the number of products, with the company having the most products at the top. The x-axis is a scale indicating the number of products, which ranges from 0 to 100. The chart is used to compare the product portfolios of the top 10 companies visually.

The bar chart provides a hierarchical representation of the top 10 marketing companies within a particular sector, ranked by their respective product counts. The leading company markedly eclipses others, showcasing a bar that nearly reaches the chart's maximum scale, implying a dominant market position through an extensive product lineup.

Descending through the ranks, the other companies exhibit decreasing product counts. These shorter bars represent a less extensive, yet still significant, market presence. Their strategic positioning might focus on a narrower range of products, specialized markets, or targeted therapeutic areas.

Notably, the bottom-tier companies, while still among the elite top 10, have considerably smaller product ranges. Their shorter bars might suggest a focus on niche markets or a strategic decision to maintain a lean product portfolio.

This bar chart serves as a diagnostic tool for analyzing market distribution, strategic business planning, and competitive dynamics within the industry. It visually narrates the story of market leadership, strategic diversity, and competitive positioning, encapsulating the broad spectrum of company approaches to product marketing.

Top 10 Countries by Number of Products Manufactured
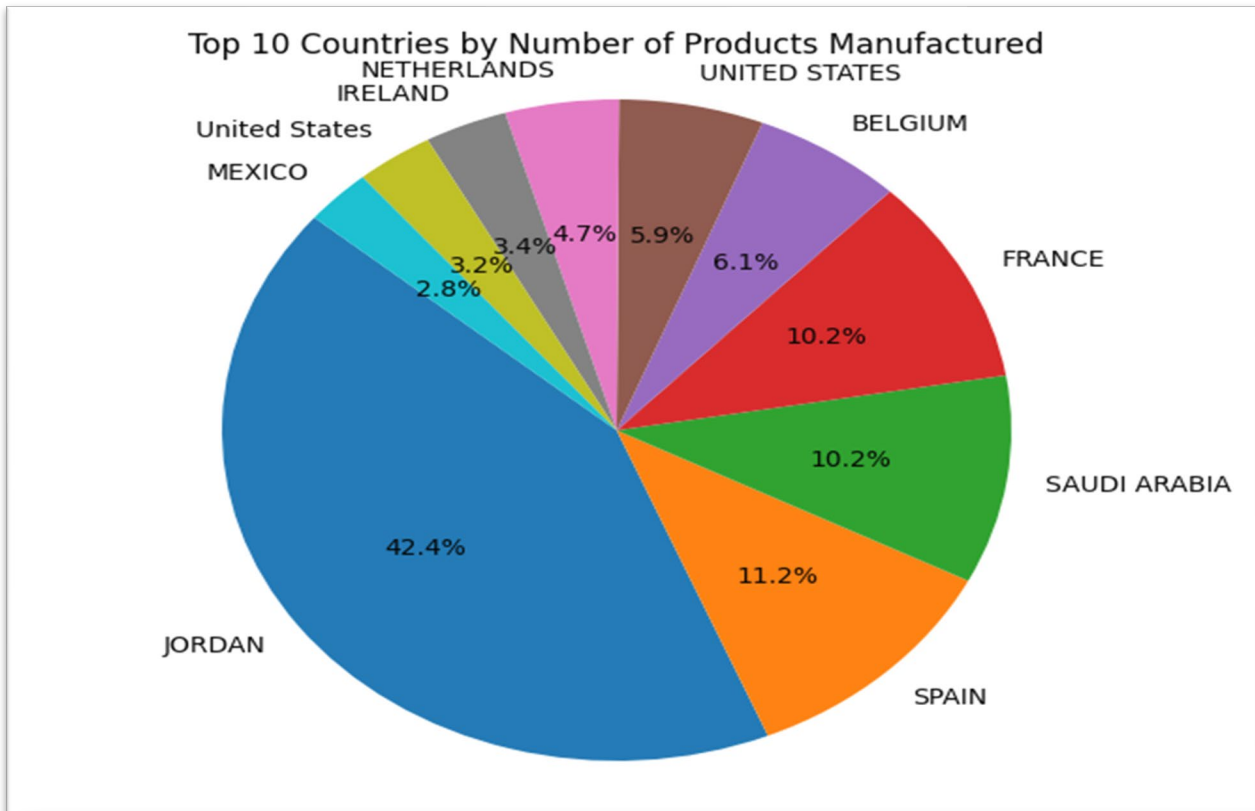
The image is a pie chart with colored slices indicating the proportion of products manufactured by the top 10 countries. Jordan has the largest slice, and other countries have smaller portions, each labeled with a percentage.

The pie chart displays the distribution of product manufacturing across the top 10 countries, with each slice representing a country's share of the total number of products manufactured.

Jordan dominates the chart with a substantial 42.4% share, indicating it as a major hub for manufacturing in this sector. Such a significant portion suggests a concentrated industry presence or specialization in product manufacturing within Jordan.

Spain and Saudi Arabia follow, each with 11.2% of the products manufactured, which positions them as substantial players, though far behind Jordan in terms of volume. The relatively equal shares between these two countries may point to a competitive balance in their manufacturing capabilities or market strategies.

France and Belgium hold the next tier of shares with 10.2% and 6.1%, respectively, signifying their strong, yet notably smaller, manufacturing presence compared to Jordan.

The United States, Ireland, Netherlands, Mexico, and one country with an obscured name due to overlapping text, each have less than 6% of the total share, ranging from 5.9% down to 2.8%. These smaller slices indicate a more modest manufacturing output compared to the leading countries.

This chart reflects the global landscape of manufacturing within the industry, revealing which countries are the most prominent producers and providing insights into global market dynamics, production capacities, and potential areas of manufacturing dominance.

Top Medicines with Longest Withdrawal Periods
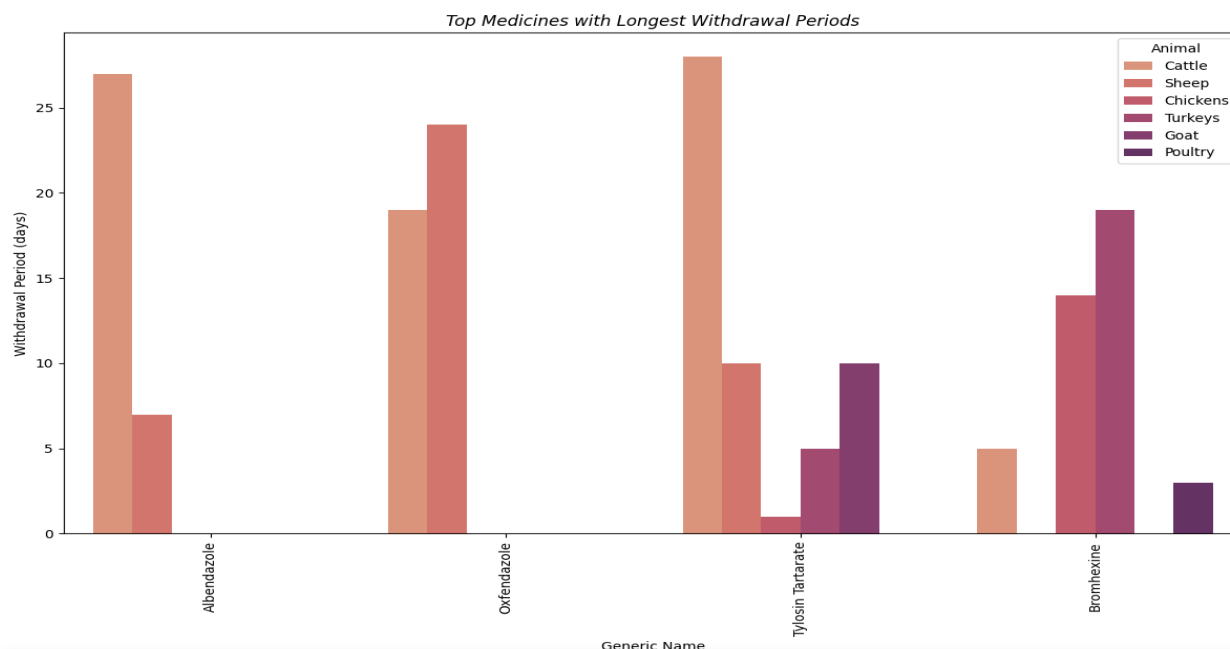
> The image is a clustered bar chart with vertical bars representing withdrawal periods in days for various medicines, categorized by animal type. Each cluster of bars corresponds to a different generic medicine, and each bar within a cluster represents a different animal, with colors distinguishing the animal types. The height of each bar indicates the length of the withdrawal period for that specific medicine-animal combination.

The vertical clustered bar chart provides a detailed analysis of the withdrawal periods for top veterinary medicines, categorized by animal species. The generic names of the medicines, such as Albendazole, Oxytetracycline, Tylosin Tartrate, and Ivermectine, are positioned along the horizontal axis, while the vertical axis quantifies the withdrawal period in days.

Albendazole's bars indicate that cattle require a significantly longer withdrawal period compared to other animals, which may be reflective of bovine metabolism rates or the dosage used. Oxytetracycline shows a notable extended withdrawal for sheep, which might be due to the pharmacodynamics of the drug in ovine systems or regulatory standards for sheep products.

The variations in withdrawal periods across the medicines for different animals such as chickens, turkeys, goats, and the combined category of poultry highlight the importance of understanding pharmacological effects in various species. These differences underscore the necessity for tailored veterinary practices and the meticulous regulation of food safety.

This chart is a vital tool for veterinarians, animal husbandry professionals, and regulatory authorities. It underscores the critical nature of adhering to withdrawal times to ensure that animal-derived food products are safe for human consumption, free from residual medications, and compliant with food safety regulations. It also speaks to the complexity of veterinary pharmacology and the need for comprehensive guidelines for medication use in food-producing animals.

## DISCUSSION

**Problem 1:** Variation Across Animal Types: The types of veterinary medicines, their dosage forms, and administration methods exhibit significant variation across different animal types. For instance, cattle have a wider variety of unique medicine types compared to birds or broilers, which may reflect the economic value and larger scale of cattle farming. Dosage Forms: The top dosage forms, as indicated by the visualizations, include solutions for injection, powders for oral solution, and oral suspensions. These forms are likely preferred due to their ease of administration, effectiveness in drug delivery, and stability of the medication. Administration Methods: The most common routes of administration are oral and injectable routes (subcutaneous, intramuscular, and intravenous). The choice of administration method depends on factors such as the speed of action required, the animal's tolerance, and practical considerations related to animal handling. Market Trends: The variations in medicine types and administration methods suggest market trends that prioritize certain animals and treatments. The presence of a wide variety of medicines for animals like cattle and poultry indicates a robust market driven by the need for livestock health management and the economic importance of these animals. In contrast, less variety in medicines for other animals could suggest a smaller market or less focus on those animal types within the veterinary pharmaceutical industry.

**Problem 2: Patterns and Impacts:**
Centralization in certain regions and companies may reduce competition, impacting prices and innovation in the veterinary medicine market.
Geographical concentration can result from economic policies, raw material availability, or pharmaceutical investments.
Distribution and marketing reach affect global accessibility, influencing animal treatment in different regions.
 - Top 10 Manufacturers, Marketing Companies, and Countries:
Identifies major players in manufacturing, marketing, and leading countries in veterinary medicine production.
**Part 2: How it Affects the Veterinary Medicine Market:**
Dominance in certain regions may lead to variations in product availability, impacting market accessibility and diversity.
Concentration in specific regions might result in cost variations, affecting pricing and affordability in regions reliant on imports.
Regions with a high concentration of manufacturers are likely centers for research and innovation, influencing global trends.
Dominant manufacturing countries may influence global regulatory standards, affecting product safety and quality.
A diverse and globally distributed manufacturing base contributes to a stable supply chain, while over-reliance on specific regions can pose vulnerabilities.

**Part 3 - Market Impact:**

Concentration of manufacturers suggests advanced infrastructure and potential market leadership in specific countries.
Diverse marketing companies contribute to a competitive market, influencing product diversity and pricing strategies.
Countries with many manufacturers are likely centers for innovation, affecting global trends in veterinary medicine.
Concentration of manufacturers in certain regions affects the global supply chain, influencing product availability and pricing.
A competitive environment among marketing companies can drive product diversity and pricing strategies. Regions with fewer manufacturers may depend more on imports, affecting access and cost of veterinary products. Understanding these dynamics is crucial for navigating the veterinary medicine market, shaped by economic forces and the global imperative to address animal health comprehensively.

**Problem 3:** **The vertical bar chart demonstrates some clear trends and potential outliers in the withdrawal periods for different medicines and animals:**
Albendazole shows the longest withdrawal period for cattle at 27 days, which is quite high compared to its withdrawal period for sheep, which is only 7 days. This disparity might suggest a different metabolism or sensitivity between these two species, which could have implications for dosing and safety protocols.
Oxfendazole presents an interesting case where the withdrawal period for sheep is significantly longer (24 days) than for cattle (19 days). Additionally, the data mentions a withdrawal period for cattle at 3.5 days, which is an outlier and seems anomalously low compared to other values. This might be due to a different formulation or application of the medicine that wasn't clearly captured in the OCR text.
**Tylosin Tartarate** has a broad spectrum of withdrawal periods ranging from just 1 day for chickens to 28 days for cattle, indicating a substantial variation in how different animals process this medication. This wide range could have implications for how the medicine is used in mixed farming operations and the monitoring of food products from these animals.
Bromhexine shows a shorter withdrawal period for poultry (3 days) compared to turkeys (19 days) and chickens (14 days). This suggests that different poultry species metabolize Bromhexine at different rates, which could impact treatment decisions and the timing of product processing for market.
From a public health perspective, adherence to withdrawal periods is crucial to ensure that no residues of these medicines are present in animal products that could pose risks to consumers.
For animal health, understanding these variations is important to prevent either underdosing or overdosing, which could lead to ineffective treatment or drug resistance.
For veterinary and farm management, this data highlights the need for careful planning and record-keeping to ensure that withdrawal times are strictly followed. It also underscores the necessity for species-specific research and guidelines to ensure both animal welfare and food safety.

**CONCLUSION AND LIMITATIONS**

**Conclusion**: The project successfully analyzed the "List of Registered Veterinary Medicines Products" dataset, providing valuable insights into the veterinary pharmaceutical industry. It highlighted trends in drug composition, regulatory compliance, market dynamics, and impacts on animal and public health. The data facilitated a deeper understanding of product diversity, manufacturing distribution, and withdrawal periods for various animal species, crucial for informed decision-making in veterinary practices, pharmaceutical companies, and regulatory bodies.

**Limitations:**
1. Data Scope and Completeness: The dataset may not cover all available veterinary medicines globally, limiting the generalizability of findings.
2. Data Quality and Accuracy: Potential errors or inconsistencies in the dataset can affect the accuracy of analysis and conclusions.
3. Temporal Dynamics: The dataset represents a snapshot in time and might not reflect current trends or future developments in the veterinary pharmaceutical sector.
4. Regional Variability: The findings may be more representative of certain regions, given the dataset's source, and may not fully capture global market dynamics.
5. Specificity to Veterinary Field: The insights are specific to veterinary medicines and may not be applicable to other pharmaceutical sectors.

# CONTRIBUTIONS

## Team member parts:
1. **Nawaf Abdulrhman Alageel: Title, Names, Introduction, Abstract Results and Analysis (problem1).**

2. **Mohammed Khalid Altufayhi: Results and Analysis(problem3), conclusion and limitations, Discussion**

3. **Albadar Ibrahim Almaymani: Results and Analysis(problem2), References**

4. **Abdullah Mansour Habit: Results and Analysis(problem2), Discussion, Attached code.**

## REFRENCES
**List of Registered Veterinary Products Datasets - Saudi Open Data**
**List of Registered Veterinary Products - SAHPRA**
**The source code Hare in colab Google**
**The Data Set we use in the Project**