

ORIGINAL RESEARCH ARTICLE

# Harnessing Big Data in the Animal Welfare Industry: Utilizing Data Science to Improve Regulatory Oversight of Commercial Dog Breeding

Clinton Ross Mauck , Jyothi Vinnakota Robertson and Marjorie Robin Vincent

JVR Shelter Strategies, Belmont, CA, United States

## Abstract

**Introduction:** In the age of Big Data, the animal welfare industry stands to benefit from data-driven decision making, particularly in commercial dog breeding. Despite its potential, many organizations and regulatory bodies, such as the United States Department of Agriculture (USDA), face significant challenges in organizing and using it effectively. The existing challenges limit the extent to which the vast amount of data collected by the USDA can be used to improve regulatory oversight and promote animal welfare. This study explored the potential of leveraging publicly-available inspection report data to inform animal welfare standards and identify areas of improvement.

**Methods:** We formulated an innovative approach for extracting, cleaning, and structuring data from the Public Search Tool (PST) database. Our approach involved the use of customized web-scraping tools and data manipulation techniques, including automatic data retrieval, transformation of inspection reports into a text-friendly format, and pattern recognition for collating pertinent data elements. We conducted descriptive statistical analyses on the assembled dataset to set the stage for a comprehensive exploration of inspection reports from Class 'A' commercial dog breeding facilities.

**Results:** Our study produced an extensive dataset detailing compliance with animal welfare standards at Class 'A' commercial dog breeding facilities across the United States from 2014 to 2023. Preliminary analysis revealed prevalent areas of non-compliance, such as inadequate veterinary care and substandard housing conditions. The dataset facilitated a deep analysis of animal welfare practices within the commercial dog breeding industry, providing insights across geographical locations and facility sizes.

**Conclusion:** Our study underscores the potential of harnessing Big Data to inform regulatory decisions and improve animal welfare within commercial dog breeding. It introduces a method to transform publicly available data into an accessible format. This allows us to go beyond anecdotal evidence into comprehensive assessments, facilitating constructive dialogue and effective policy-making. Further research leveraging advancements is recommended to deepen insights and encourage collaborative efforts to elevate animal welfare standards.

**Keywords:** *Commercial Dog Breeding; USDA Inspection Compliance; Canine Welfare; Big Data; Data Science*

Received: 18 August 2023  
Revised: 9 October 2023  
Accepted: 16 October 2023  
Published: 1 December 2023

## Correspondence

Clinton Ross Mauck  
Email: ross@shelterstrategies.com

## Reviewers

Sloane Hawes  
Tennille Lamon

In the age of Big Data, the acquisition and evaluation of data has become a critical component of business and industry.<sup>1-4</sup> Big data refers to extremely large datasets that may be analyzed computationally to reveal patterns, trends, and associations. Big data analytics has proven effective in diverse fields where it has revolutionized processes, improved operational efficiency, and enabled evidence-based decision-making.<sup>3,5</sup> For the animal welfare industry, big data analytics offers unprecedented opportunities for organizations and regulatory bodies to apply data-driven approaches in the formation

of policy, enhancement of regulatory oversight, and the facilitation of constructive dialogue among stakeholders by supplementing anecdotal evidence with more comprehensive insights in a collective effort to address animal welfare concerns.<sup>6,7</sup>

One area of concern within animal welfare which would benefit from harnessing big data is the regulation of commercial dog breeders by the United States Department of Agriculture (USDA).<sup>8</sup> Commercial dog breeding is a socially contentious topic due to concerns about the welfare of dogs in breeding facilities and the

quality of care they receive.<sup>9–11</sup> There are ongoing discussions within the animal welfare community about the effectiveness of existing regulatory measures as prescribed by the Animal Welfare Act (AWA).<sup>12,13</sup> Concerns have been raised about enforcement, and transparency in reporting and inspections.<sup>14–16</sup> These concerns highlight the importance of collecting and analyzing data to obtain evidence-based insights and exploring data-driven approaches to ensure the welfare of dogs in commercial breeding situations.

Under the AWA, the USDA is responsible for regulating these facilities and enforcing the established standards for animal care.<sup>17</sup> To support these efforts, the USDA Animal & Plant Health Inspection Service (APHIS) maintains a repository of published inspection reports and descriptive metadata which represent a significant resource for evaluating and improving animal welfare practices in commercial dog breeding facilities. However, this data is currently not readily accessible for broad analysis. Most of the information can be found in the USDA APHIS Public Search Tool.<sup>18</sup> Although it is publicly accessible, its user interface makes it difficult to efficiently extract the needed data. In addition, the information contained in the reports is stored as PDF documents, posing obstacles to automated text analysis to enable efficient processing of large volumes of text data by extracting valuable insights, patterns, and trends. Despite the immense potential of the USDA's database, the existing challenges in accessing and analyzing the public information hinder its use for analyzing commercial breeder compliance with AWA regulations and addressing concerns related to the necessity of regulation improvements.

The study encompassed all types of inspections conducted within the commercial dog breeding facilities. A **routine inspection** is a normal periodic, unannounced complete inspection of the facility. A **pre-license inspection** is performed to determine compliance prior to issuance of a new USDA license. A **re-license inspection** is performed prior to re-issuance of an existing USDA license. A **new site inspection** is performed on existing licensees prior to the use of a new facility site. A **focused inspection** is an unannounced partial inspection of a facility including reinspections following specific non-compliant items (NCI) or a public complaint. Focused inspections were not categorized separately from routine inspections until September 22, 2016.<sup>19</sup>

The severity of NCIs are designated as 'direct', 'critical' or 'non-critical'. The 'direct' designation is used when (at the time of inspection) the non-compliance is having a severe adverse effect on the welfare of an animal or has the high potential to have that effect in the immediate future. A 'critical' non-compliance is one that had a severe adverse effect at some point but at the time of inspection is

no longer having that effect. The 'critical' non-compliance was not categorized separately from 'direct' non-compliance until September 22, 2016.<sup>19</sup>

Persisting NCIs are designated as 'repeat' when the non-compliance was cited in the same section and subsection on the last inspection (routine, focused or re-license) or cited at least three times within the past 3 years (including the current citation).

In the following sections, we outline the detailed methodology employed to extract, process, and analyze the data, providing a framework for leveraging this valuable resource towards enhancing animal welfare practices in the commercial dog breeding industry. We discuss some of the innovative research methods applied, such as the use of pattern matching and keyword extraction techniques to capture information from the text and metadata of inspection reports. Finally, we explore some of the ways in which advanced technologies could improve the quality and quantity of animal welfare research in commercial breeding facilities.

## Methods

### Data acquisition

The primary data source for this study was the USDA Public Search Tool. This tool provides access to a wealth of information via inspection reports of commercial dog breeding facilities collected from January 30, 2014 to February 9, 2023. This timeline represents the entire scope of information available in the Public Search Tool on February 9, 2023 when the data was accessed. For data extraction, we developed a web scraping process to gather the necessary data directly from PDF files available on the Public Search Tool.

We first established a selection criteria for searching inspection reports based on 'License/Registration Type' and 'Animal Category', choosing breeders and dogs respectively. This allowed us to focus our search on Class 'A' licensed commercial dog breeders. To cover all relevant data on this group, we divided the search by iterating over all available states using the search filters, as the Public Search Tool was limited to the first 2,100 reports per search. Four states—Missouri, Ohio, Indiana, and Iowa—had a volume exceeding 2,100 inspection reports and were further segmented into searches by all applicable zip codes within them.

We then identified download links for PDF files within the search results, using a programmatically-controlled method to manage the 'next page' button on the website and ensure that all results were viewed. With these components identified and managed, we were able to fully automate the process of collecting available inspection reports.

For navigation and interaction with the Public Search Tool, we employed Python's Selenium library's Web

Driver. This tool proved instrumental in automating the task. Further, we utilized the Python Requests module to download each inspection report from the retrieved download links.<sup>20,21</sup>

#### *Data processing*

The resulting 21,667 inspection reports required conversion from PDF to text for analysis. A number of open-source Python libraries were evaluated but ‘pdfplumber’ was found to be the most accurate and was sufficient for our purposes.<sup>22</sup> The conversion from PDF to text interpreted the content correctly. Furthermore, the outcomes retained attributes enabling the extraction of the desired data points from each report.

The raw data extracted from the PDF files required significant cleaning and structuring. The data’s complexity and variety—encompassing both numerical and textual elements—necessitated the implementation of particular methods to achieve uniformity. The data cleaning process involved the removal of irrelevant or incomplete data, standardization of text fields, and consolidation of redundant entries.

The textual information from each inspection report was processed using tailor-made patterns crafted to distinguish different data components. The Python Re module for regular expression operations was utilized to accurately locate elements within the text files.<sup>21</sup>

Base data that was captured in each report:

- File Name: The name of the inspection report file.
- Search Criteria: The search filters selected to access the inspection report.
- Inspection Date: The date the inspection was conducted.
- Inspection Type: Type of inspection report.
- Certificate ID: The certificate ID of the facility.
- Customer ID: The customer ID of the facility, tied to personal unique identifiers.
- Name: Name linked to the Customer ID.
- Dog Adult: The number of adult dogs at the facility at the time of inspection.
- Dog Puppy: The number of puppies at the facility at the time of inspection.
- Total Animals: The total number of animals at the facility at the time of inspection.
- Site Number: The facility number to track customers with multiple facilities.
- Street Address: Location of the facility to be used for further segmenting location.
- Non-Compliance: Whether the inspection report identified any NCIs.

Recognizing and accommodating the changes in format, language, and approach that the USDA APHIS has implemented in their inspection reports over time was essential. This step was crucial in maintaining the

consistency and reliability of the data throughout the entire research period. For example, before 2016, the reports didn’t differentiate between ‘focused’ and ‘routine’ inspection types. To address this, we analyzed the report text to determine the nature of each inspection, ensuring accurate categorization.

We further enriched this dataset using external data sources. The ‘uszipcode’ Python library was used to assign each report to a county based on the zip code.<sup>23</sup> In addition to these standard data points, specific phrases and identified Code of Federal Regulations (CFR) were also captured in the text from each inspection report.

#### *Data analysis*

The data analysis phase aimed to demonstrate the ability to generate insights from the cleaned data, providing an overarching examination of the available information. Descriptive statistics, using the Python PANDAS module, were conducted to give an overview of facility sizes, geographical distribution, inspection types, and compliance rates.<sup>24</sup> Further, NCIs were identified and categorized based on their CFR sections.

The compliance rate was calculated using the proportion of inspections where facilities fully complied with the required standards. The compliance rate was then evaluated across the entire study period, as well as for the more recent period from January 1, 2021 to February 9, 2023. The field of commercial dog breeding is dynamic and continually evolving, with regulations and standards subject to change and adaptation. By examining a more recent period, specifically from January 1, 2021 to February 9, 2023, we aimed to provide insights that are more reflective of the present state of compliance and adherence to the prevailing standards and regulations.

Data visualization techniques were employed to highlight key findings and patterns in the data. A geographic map was created to represent the distribution of commercial dog breeding facilities across the states and the rate of compliance within each county. Histograms were used to represent the variation in the number of dogs housed in each facility.

### **Results**

For the initial demonstration of the data’s potential, a comprehensive set of descriptive statistics are presented, covering aspects such as facility size, geographical distribution, and compliance rates. These metrics offer a high-level overview of the industry, providing a foundation for further analyses.

#### *Population overview*

Our analysis identified 3,903 unique customer IDs, representing commercial dog breeding facilities subjected to routine inspections between January 30, 2014, and

February 9, 2023. A segment of this population, comprising 2,478 unique customer IDs, reflected facilities inspected between January 1, 2021 and February 9, 2023.

**Geographical distribution**

Analysis of geographical distribution showed a state-wise variation in the presence of unique customer IDs between January 1, 2021 and February 9, 2023. Notably, Missouri ( $n = 886$ ) housed the most customer IDs that underwent routine inspections, followed by Indiana ( $n = 354$ ), Ohio ( $n = 252$ ), Iowa ( $n = 210$ ), Oklahoma ( $n = 141$ ) and Kansas ( $n = 116$ ). Figure 1 further visualizes this geographical distribution, displaying the compliance rate and the number of unique licensed breeders in each county.

**Inspection type**

During the entire study period, the most common type of inspection was the Routine Inspection ( $n = 17,202$ ). This was followed by Pre-License Inspections ( $n = 2,896$ ), Re-License Inspections ( $n = 927$ ), Focused Inspections ( $n = 520$ ), New Site Inspections with ( $n = 98$ ) reports (Table 1).

From January 1, 2021 to February 9, 2023, the count of inspection types included: Routine Inspections

( $n = 3,955$ ), Re-License Inspections ( $n = 927$ ), Pre-License Inspections ( $n = 766$ ), New Site Inspections ( $n = 90$ ), and Focused Inspections ( $n = 63$ ).

**Facility size distribution**

We investigated the average size of adult dog and puppy populations based on customer IDs in commercial dog breeding facilities. To provide insights into the recent distribution and range of breeder site sizes, we focused on routine inspections conducted between January 1, 2021, and February 9, 2023.

We observed considerable variation with the number of total canines housed in each facility with 27.8% of facilities, denoted by their customer IDs, as having less than 30 canines. 19.2% of customer IDs had more than 100 canines on average at facilities during routine inspections. For this analysis, because they were extreme outliers, we removed 42 customer IDs (1.7%) that had more than 300 canines, on average. The highest count of canines in a facility was 21,283. The inclusion of these customer IDs would mask the characteristics of the average dog breeder.

Figure 2 shows the wide distribution of the number of total canines in each facility, demonstrating the diversity in the size of commercial dog breeding operations.

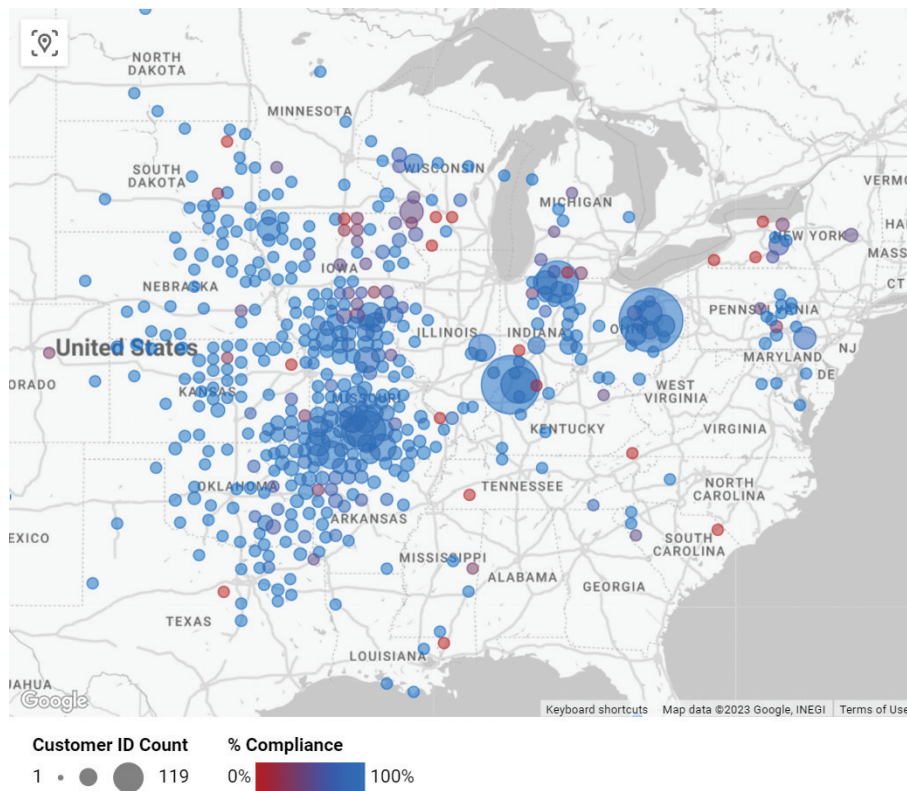


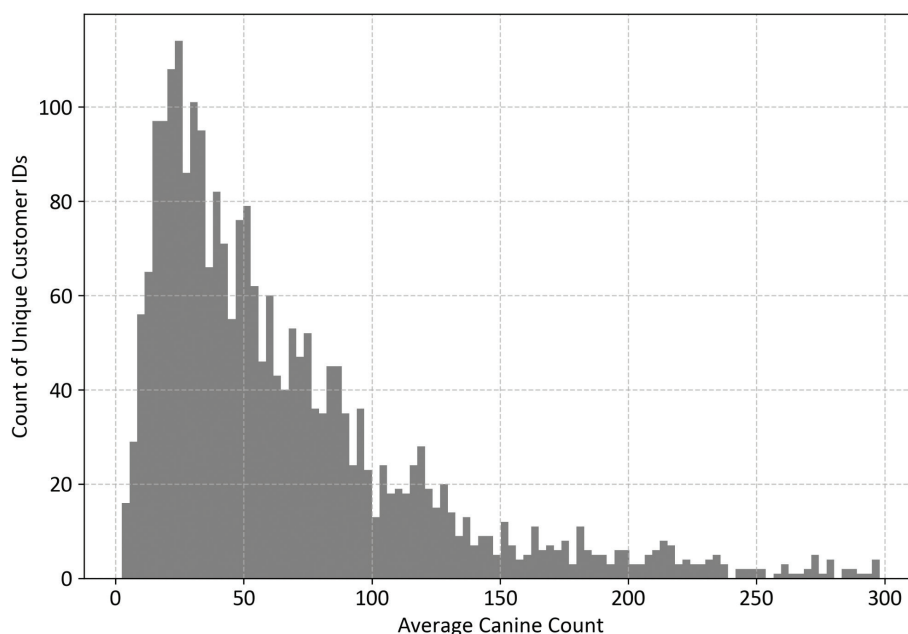
Figure 1. Compliance Rate and Count of Licensed Breeders by County During Routine Inspections between Jan 1, 2021 and Feb 9, 2023.



**Table 1.** Counts and Percentages of Inspection Reports by Inspection Type for Jan 30, 2014 to Feb 9, 2023 ( $N = 21,677$  reports) and Jan 1, 2021 to Feb 9, 2023 ( $n = 5,801$  reports)

Inspection type	Count of inspection reports (Jan 30, 2014 to Feb 9, 2023) $N = 21,677$ (100%)	Count of inspection reports (Jan 1, 2021 to Feb 9, 2023) $n = 5,801$ (100%)
Routine	17,202 (79)	3,955 (68)
Pre-license	2,896 (13)	766 (13)
Re-license <sup>a</sup>	927 (4)	927 (16)
Focused	520 (2)	63 (1)
New site	98 (<1)	90 (2)

<sup>a</sup>Re-License inspections were not categorized before October 1, 2020.



**Figure 2.** Average Canines in Facility by Customer ID During Routine Inspections from January 1, 2021 to February 9, 2023 Histogram.

Averages, medians, and standard deviations by animal age group are presented in Table 2.

#### Compliance rate

From the total of 17,202 routine inspections conducted between January 31, 2014 and February 9, 2023, a compliance rate of 85% was found. However, the rate showed an improvement in the more recent time frame from January 1, 2021 to February 9, 2023, with a compliance rate of 91%.

#### Degree and persistence of non-compliance

The severity associated with each instance of non-compliance was identified in 2,704 inspection reports, from the entire study period, comprising 6,391 total NCIs. The subset of routine inspections included 2,505 reports with 5,937 total NCIs. Direct NCIs formed 8.3% of non-compliant instances, critical ones formed only 0.5%, and the remaining 91.2% were non-critical.

**Table 2.** Average Number of Animals per Customer ID Facility, Standard Deviation and Median

Animal age group	Average animals per customer ID facility and standard deviation	Median animals per customer ID facility
Adult dogs	42.0 ± 34.8 SD	31
Puppies	26.7 ± 25.4 SD	19
Total canines	66.1 ± 54.0 SD	50

For this same subset of routine inspections over the entire study period repeat NCIs accounted for about 20.4% of all non-compliance instances.

#### Non-compliant items

The analysis of routine inspections conducted from January 30, 2014 to February 9, 2023 highlighted varied non-compliance rates across CFRs. These rates were

grouped by section and measured by total count of NCIs as a percentage of 17,202 inspection reports, count of repeat NCIs, and count of direct NCIs (Table 3).

The section with the highest non-compliance rate was ‘9 CFR 2.40: Attending veterinarian and adequate veterinary care’ with nine unique subsection items, 1,387 total NCIs, of which 390 (28%) were repeat NCIs and 340 (25%) were direct NCIs. Following closely was ‘9 CFR 3.1: Housing facilities, general’ with 15 unique subsection items and 1,305 total NCIs. This section saw 298 (23%) repeat NCIs, but only 12 (1%) were direct NCIs.

Further down, the section ‘9 CFR 3.11: Cleaning, sanitization, housekeeping, and pest control’ showed a slightly lower but still significant number of 879 total NCIs from 13 unique subsection items, 171 (20%) of which were repeat NCIs, and 18 (2%) were direct NCIs.

Sections like ‘9 CFR 2.75: Records’ and ‘9 CFR 2.50: Time and method of identification’ had notable rates of non-compliance, with 335 and 295 total NCIs respectively, but none of these NCIs in these sections were direct.

Finally, the table also includes the category ‘Others’ representing 36 additional sections that weren’t individually listed due to the ranked structure of the table. For example, ‘Others’ included 3.10, 3.2, 3.8, and 2.131. Together the category of Others accounted for 358 NCIs. Repeat NCIs made up 8% of the Others group, while 11% were direct NCIs.

*Table 3.* Non-compliance Rate by CFR and CFR Section Identified in Routine Inspections from Jan 30, 2014 to Feb 9, 2023

CFR section	All NCIs (%)	Repeat NCIs (%)	Direct NCIs (%)	Customer IDs with NCIs (%)
9 CFR 2.40	1387 (23)	390 (28)	340 (25)	763 (20)
9 CFR 3.1	1305 (22)	298 (23)	12 (1)	536 (14)
9 CFR 3.11	879 (15)	171 (20)	18 (2)	475 (12)
9 CFR 3.6	609 (10)	94 (15)	17 (3)	355 (9)
9 CFR 3.4	356 (6)	52 (15)	37 (10)	210 (5)
9 CFR 2.75	335 (6)	71 (21)	0 (0)	219 (6)
9 CFR 2.50	295 (5)	56 (19)	0 (0)	210 (5)
9 CFR 3.9	153 (3)	25 (16)	15 (10)	110 (3)
9 CFR 3.13	150 (3)	17 (11)	6 (4)	91 (2)
9 CFR 3.3	110 (2)	8 (7)	4 (4)	89 (2)
36 Others	358 (6)	28 (8)	41 (11)	242 (6)

Section titles:

- 9 CFR 2.40: Attending veterinarian and adequate veterinary care
- 9 CFR 3.1: Housing facilities, general
- 9 CFR 3.11: Cleaning, sanitization, housekeeping, and pest control
- 9 CFR 3.6: Primary enclosures
- 9 CFR 3.4: Outdoor housing facilities
- 9 CFR 2.75: Records
- 9 CFR 2.50: Time and method of identification
- 9 CFR 3.9: Feeding
- 9 CFR 3.3: Sheltered housing facilities

## Discussion

The present study delved into the opportunities and challenges posed by the vast and ever-increasing quantity of data available today. While this data holds valuable insights, it requires extensive cleaning, preparation, and processing to be usable. In our specific case, we encountered these challenges while dealing with a substantial dataset extracted from PDF documents over an extended period, focusing on various aspects of commercial dog breeding facilities and their compliance with animal welfare standards.

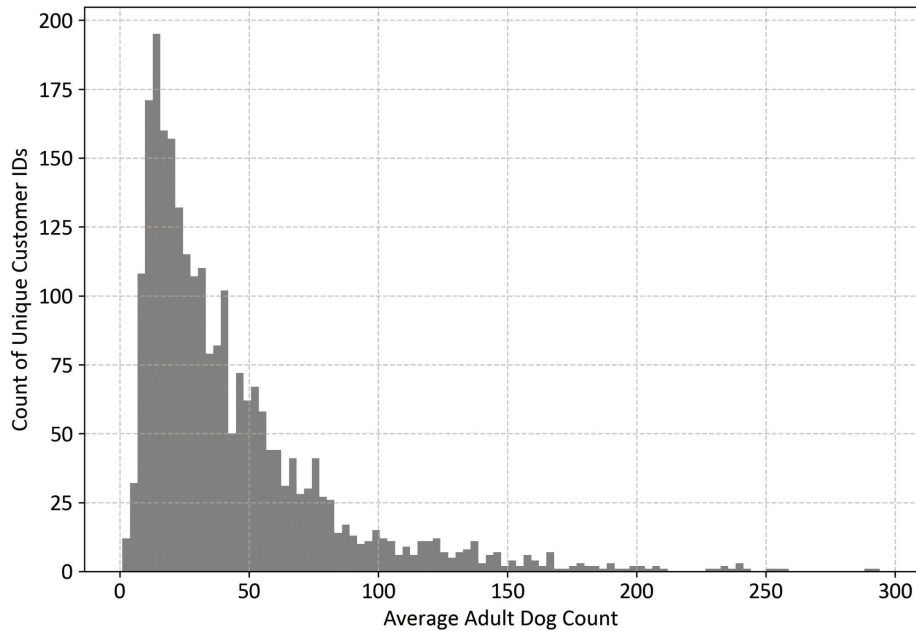
The dataset we obtained not only included quantitative variables but also incorporated textual data, allowing for a more comprehensive analysis using statistical methods and broader analytical approaches. Through our statistical analysis, critical insights into the state of commercial dog breeding facilities in the United States emerged. The geographical distribution of these facilities and their compliance rates highlighted potential variations in practices and regulations across different states, as well as the diversity of facility size within the commercial dog breeding industry. This dataset gives us the ability to track these trends over time.

We compared counties based on the number of licensed breeders and the average compliance rate of all routine inspections (Figure 1). However, these metrics can be interchanged with other parameters such as specific NCIs or occurrences of certain phrases that might indicate unique circumstances. The granularity offered by the location-based data casts light on regional patterns within the U.S. commercial dog breeding industry. Understanding these regional patterns is crucial as states have different regulations and enforcement practices for commercial dog breeding facilities. Additionally, discerning regional patterns allows for the identification of localized common practices of breeders, which may be indicative of regional norms, preferences, or breeding challenges.

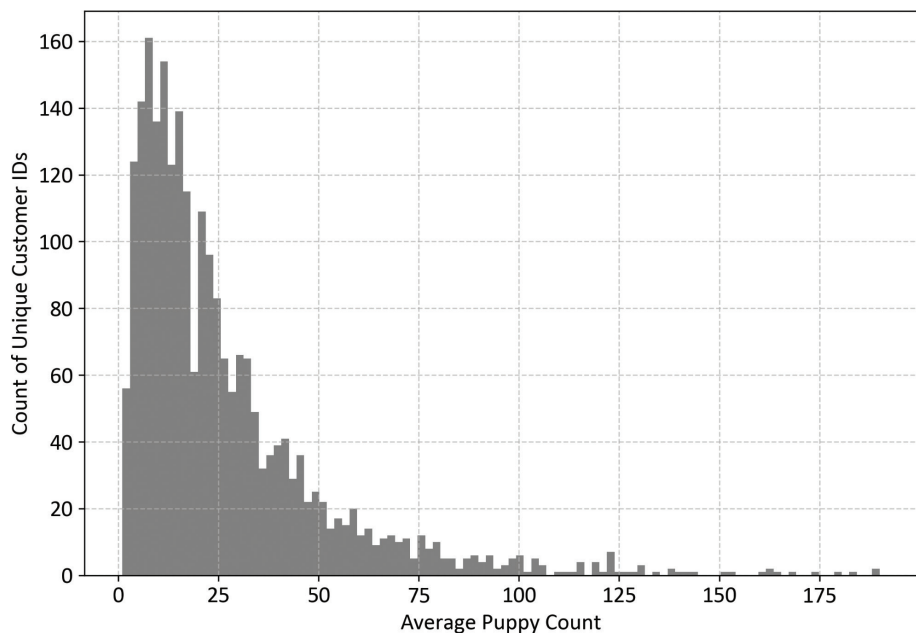
Our findings indicated that both commercial breeding operations and smaller-scale operations coexist in the industry (Figures 2, 3, and 4). We identified specific areas of non-compliance, such as inadequate veterinary care and substandard housing conditions, emphasizing the need for targeted improvements in these aspects.

The sections dealing with veterinary care and housing facilities generally saw higher rates of non-compliance. It’s noteworthy that the sections relating to records and identification methods had a significant number of NCIs but none of these were direct, implying that these infractions may not immediately impact the animals under regulation. However, it’s evident that there is room for improvement across all these areas to ensure better adherence to regulations.

Despite the patterns revealed by our analysis, certain limitations should be acknowledged. The study’s reliance



*Figure 3.* Average Adult Dogs in Facility by Customer ID During Routine Inspections from January 1, 2021 to February 9, 2023 Histogram.



*Figure 4.* Average Puppies in Facility by Customer ID During Routine Inspections from January 1, 2021 to February 9, 2023 Histogram.

on the accuracy and completeness of USDA inspection reports may not account for variations in enforcement across all facilities. Another significant limitation was the impact of the COVID-19 pandemic on the USDA inspection process. In the study dataset, there were notably fewer inspections in 2020 as the USDA halted routine inspections for a significant period due to the pandemic. This reduction in inspections would have influenced

the non-compliance ratings for the time periods in our study, potentially leading to an underrepresentation of non-compliance incidents during this year. This limitation underscores the need for cautious interpretation of the compliance trends observed in our study, especially for the year 2020.

Our automated data retrieval process encountered some limitations. The search tool, for instance, would

only load the initial 21 pages (2,100 reports). This posed a significant challenge which is why we subdivided the search criteria by state and zip code.

Nevertheless, the study showcased the potential of big data and advanced analytical tools to inform and enhance animal welfare. Techniques like web scraping, textual analysis, and data visualization proved invaluable for regulators, animal welfare organizations, and researchers in making informed decisions and devising effective strategies for improving animal welfare in commercial breeding facilities.

In our study, we chose to focus on Class ‘A’ licensee commercial dog breeders. A Class ‘A’ licensee is anyone, who owns at least five breeding females, meeting the definition of ‘dealer’ whose business consists only of animals acquired for the sole purpose of maintaining or enhancing the breeding colony and animals that are bred and raised on the premises.<sup>25</sup> This group represents the vast majority of commercial dog breeders, but not the entire industry. A Class ‘B’ licensee is anyone meeting the definition of ‘dealer’ whose whole business includes the purchase and/or resale of any animal. Some Class ‘B’ breeders may primarily be breeders that occasionally sell other’s animals and some may be exclusively brokers. Our methodology could be directly applied to inspection data on Class ‘B’ breeders.

To maximize the value of the data, we developed a comprehensive methodology for data extraction, cleaning, and structuring from the Public Search Tool, addressing user interface limitations and PDF document challenges. This methodology can serve as a blueprint for future research on large-scale data analysis in animal welfare and other related areas.

Our findings have practical implications for businesses seeking to understand their suppliers’ breeding practices and researchers investigating the health of bred animals. Moreover, by adopting and refining our methods, future research can delve deeper into the root causes of non-compliance, the impact on animal health, and the effectiveness of regulatory interventions, leading to actionable insights that directly improve animal welfare.

Looking forward, collaborative efforts between regulatory bodies, the scientific community, and welfare organizations will be essential in fully leveraging data analysis tools to enhance animal welfare in commercial breeding facilities. As we continue to advance in data accessibility and analysis, there is a significant opportunity to elevate the standards of animal welfare, promoting ethical considerations and public confidence in the industry. As patterns emerge from this newly created dataset, further research is necessary to deepen our understanding of the underlying causes of non-compliance and to develop and test solutions to these issues. By capitalizing on technological advances and collective efforts, we can strive for a higher standard of animal welfare and ensure the well-being of animals in commercial breeding facilities.

## Conclusion

This study demonstrated the ability to generate a useful dataset from publicly available inspection reports and the potential of data analysis to inform regulatory decisions within the animal welfare industry, particularly concerning commercial dog breeding facilities. By incorporating both numerical and textual data, we set the groundwork to achieve a detailed understanding of factors influencing compliance in order to identify potential areas for improvement. These insights could inform the creation of more targeted, effective inspection strategies in the future.

Considering the public attention drawn to specific incidents of animal welfare violations, it is critical to note that these individual cases, while important, only represent a snapshot of a much broader landscape. This study offers an approach for harnessing the power of big data to supplement anecdotal evidence with more comprehensive insights over a longer period. Critics, advocates and the industry alike can use these methods to test their hypotheses on a broader scale and promote a more constructive dialogue. By focusing not only on the negatives but also on areas of success, a more balanced view of the situation can emerge, leading to more effective strategies and policy recommendations for the benefit of animal welfare.

## Author credit statement

C.R. Mauck: Conceptualization, Data Collection, Data Analysis, Writing—Original Draft.

J. V. Robertson, DVM, DABVP(Shelter Medicine): Conceptualization, Writing—Review & Editing, Supervising Author.

M.R. Vincent: Conceptualization, Writing—Original Draft, Review & Editing.

## Statement of ethics

The inspection report dataset collection and evaluation used in the study was limited to publicly available information.

## Conflict of interest statement

The authors declare no potential conflicts of interest.

## Acknowledgements

We would like to thank Life of Riley, a signature program of Spring Point Partners, for funding this research.

## References

1. Cao L. Data science: a comprehensive overview. *ACM Comput Surv.* 2017;50(3):43:1–43:42. doi: 10.1145/3076253
2. Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: from big data to big impact. *MIS Quarterly.* 2012;36(4):1165–1188. doi: 10.2307/41703503



3. Elgendy N, Elragal A. Big data analytics: a literature review paper. *Lect Notes Comput Sci*. 2014;8557:214–227. doi: 10.1007/978-3-319-08976-8\_16
4. Ajah IA, Nweke HF. Big data and business analytics: trends, platforms, success factors and applications. *Big Data Cogn Comput*. 2019;3(2):32. doi: 10.3390/bdcc3020032
5. Grover V, Chiang RHL, Liang TP, Zhang D. Creating strategic business value from big data analytics: a research framework. *J Manag Inf Syst*. 2018;35(2):388–423. doi: 10.1080/07421222.2018.1451951
6. Widmar NO. 'Big Data' in animal industries. Purdue University Center for Food and Agricultural Business. Published September 13, 2021. [https://agribusiness.purdue.edu/consumer\\_corner/big-data-in-animal-industries/](https://agribusiness.purdue.edu/consumer_corner/big-data-in-animal-industries/). Accessed October 3, 2023.
7. White BJ, Amrine DE, Larson RL. big data analytics and precision animal agriculture symposium: data to decisions. *J Anim Sci*. 2018;96(4):1531–1539. doi: 10.1093/jas/skx065
8. Widmar NO. 'Big Data' provides insights to public perceptions of USDA. *Choices*. Published 2019;(Quarter4). <http://www.choicesmagazine.org/choices-magazine/submitted-articles/big-data-provides-insights-to-public-perceptions-of-usda>. Accessed October 3, 2023.
9. Croney C. Turning up the volume on man's best friend: ethical issues associated with commercial dog breeding. *J Appl Anim Ethics Res*. 2019;1:1–22. doi: 10.1163/25889567-12340011
10. Stella J, Shreyer T, Ha J, Croney C. Improving canine welfare in commercial breeding (CB) operations: evaluating rehoming candidates. *Appl Anim Behav Sci*. 2019;220:104861. doi: 10.1016/j.applanim.2019.104861
11. McMillan FD, Duffy DL, Serpell JA. Mental health of dogs formerly used as 'breeding stock' in commercial breeding establishments. *Appl Anim Behav Sci*. 2011;135(1–2):86–94. doi: 10.1016/j.applanim.2011.09.006
12. Bir C, Croney CC, Widmar NJO. US residents' perceptions of dog welfare needs and canine welfare information sources. *J Appl Anim Welfare Sci*. 2019;22(1):42–68. doi: 10.1080/10888705.2018.1476862
13. Marceau JF. How the animal welfare act harms animals. Published online May 9, 2018. <https://papers.ssrn.com/abstract=3176101>. Accessed July 20, 2023.
14. Animal Care Program Oversight Of Dog Breeders. USDA Office of Inspector General; 2021. [https://usdaoig.oversight.gov/sites/default/files/reports/2023-07/33601-0002-31finaldistribution\\_0.pdf](https://usdaoig.oversight.gov/sites/default/files/reports/2023-07/33601-0002-31finaldistribution_0.pdf). Accessed July 28, 2023.
15. United States Government Accountability Office. Animal and plant health inspection service animal care program inspections of problematic dealers. Washington, DC, USA:USDA Office of Inspector General; 2010.
16. Audit of Animal Care Facilities Inspection Program. The Office of State Auditor of Missouri Claire McCaskill; 2001. [www.auditor.state.mo.us](http://www.auditor.state.mo.us). Accessed July 28, 2023.
17. Animal Welfare Act | National Agricultural Library. <https://www.nal.usda.gov/animal-health-and-welfare/animal-welfare-act>. Accessed May 31, 2023.
18. Animal and plant health inspection service inspection reports search. USDA. <https://aphis-efile.force.com/PublicSearchTool/s/inspection-reports>. Accessed July 27, 2023.
19. USDA APHIS. New terms will appear on USDA inspection reports. Published September 22, 2016. <https://content.govdelivery.com/accounts/USDAAPHIS/bulletins/16620a4>. Accessed October 3, 2023.
20. Singer-Vine J. Selenium Python (Version 4.4.3) [Computer software]. GitHub; 2023. <https://github.com/baijum/selenium-python>. Accessed February 9, 2023.
21. Python Software Foundation. Python Language Reference, version 3.9.6 [Computer software]. 2021. <https://www.python.org>. Accessed February 9, 2023.
22. Muthukadan B. pdfplumber (Version 0.7.1) [Computer software]. GitHub; 2023. <https://github.com/jsvine/pdfplumber>. Accessed February 9, 2023.
23. Hu S. uszipcode (Version 1.0.1) [Computer software]. GitHub; 2023. <https://github.com/MacHu-GWU/uszipcode-project>. Accessed February 9, 2023.
24. McKinney W. Data structures for statistical computing in python. In *Proceedings of the 9th python in science conference* (Vol. 445, pp. 51–56). Austin, Texas, USA: SciPy Conference 2010.
25. USDA APHIS. Animal care inspection guide. Published 2023. [https://www.aphis.usda.gov/animal\\_welfare/downloads/Animal-Care-Inspection-Guide.pdf](https://www.aphis.usda.gov/animal_welfare/downloads/Animal-Care-Inspection-Guide.pdf). Accessed October 3, 2023.