ANALYSIS OF DALLAS ANIMAL SERVICES DATA

Frederick McCollum

University of Dallas

Table of Contents

INTRODUCTION

As a lover of math, and proud roommate of multiple animals, I am naturally interested in the statistics and numbers associated with pet ownership. In this analysis I will be diving into Animal Services Data from the Dallas metropolitan area. This includes recent incident reports, medical records, and surrender/adoption data from the year 2017 to the present. The purpose of this analysis is not to come to a single data-driven conclusion. Rather, the purpose of this analysis is to produce insight that will better enable long-term happiness in prospective pet owners who are looking into what type of dog they would like to own. I am hopeful that this analysis will help potential pet owners make a more informed decision regarding what type of dog they would like to adopt, with respect to the following factors:

- Identifying which dog breeds seem to be most aggressive in the Dallas area.
- Determining which dog breeds have the highest number of medical issues/treatments.
- Finding which areas of Dallas have the highest number/frequency of animal incidents.

Along with the above-mentioned primary purpose of this analysis, this topic interests me personally for multiple reasons. First, as mentioned, I am a pet owner who is currently living in the Dallas area. Therefore, the data used in this analysis is a direct representation of a population that I identify with. I am also interested in gaining a better understanding of various statistics around pet ownership. This includes, but is not limited to, which dog breeds have a higher cost to own, which breeds are most commonly involved in incidents, and which animal incidents are most common. I believe that others may be interested in this analysis for some of the same reasons mentioned above, as well as other motives that may apply at a more personal level for some.

To perform this analysis, I will be utilizing *Microsoft* cloud data tools including *Azure SQL Data Warehouse*, *SQL Server Management Studio*, *Azure Blob Storage*, and *PowerBI*.

DATA

CONTEXT

For this analysis, I will be using four different data sets containing data from the years 2017 to 2019. These four data sets will aggregate information regarding Dallas animal medical information, Dallas animal shelter field data, 311 service requests in the Dallas area, and Dallas animal shelter population data. A brief description of each of these data sets may be found below:

1. Dallas Animal Medical Records

   According to the *Dallas OpenData* website, this animal medical records data set contains information pertaining to operational processes carried out by medical personnel when providing care and attention to animals received at Dallas Animal Services. The information in this data set is collected and documented by medical personnel daily. The purpose of this data set is to both keep a record of operations, as well as better help citizens understand the processes performed daily by medical personnel.

2. Dallas Animals Field Data

   According to the *Dallas OpenData* website, the Dallas animals field data set contains details regarding operations performed by Animal Services Officers who respond to calls in the field throughout the City of Dallas. This includes animal incidents, complaints, and more. The information in this data set is collected and updated daily by the Dallas Animal Services Officers who are actively performing the documented operations. The purpose of this data set is to document operations, as well as help citizens gain a greater understanding of what Animal Services Officers are doing in the neighborhoods of the Dallas area.

3. Dallas 311 Service Requests

   According to the *Dallas OpenData* website, this data set includes information regarding 311 service requests that are reported due to systems that affect Dallas citizens quality of life. 311 service requests are a common way for citizens to request assistance regarding city services and information. This includes data such as animal services requests, high weeds, junk motor vehicles, street and pot-hole repair, and more. The data is collected and recorded by the Dallas 311 system.

4. Dallas Animal Shelter Data

   According to the *Dallas OpenData* website, this data set contains details pertaining to operations carried out by animal shelter personnel when assisting citizens. This includes receiving surrendered animals and stray animals, facilitating adoptions, transferring animals to rescue groups, and providing care to the animals in the shelter every day. The data is collected and documented by animal shelter personnel daily. The purpose of this data set is to help citizens better understand the operations performed by Dallas animal shelter personnel.

CHARACTERISTICS

Please find a brief description of each data set's characteristics in the table below:

Data Set Name	Number of Rows	Number of Columns	Variable Data Types
Dallas Animal Medical Records	2471	10	Plain Text: 8 Date & Time: 1 Number: 1
Dallas Animals Field Data	74500	10	Plain Text: 8 Date & Time: 1
Dallas 311 Service Requests	867000	8	Plain Text: 5 Date & Time: 1
Dallas Animal Shelter Data	84800	18	Plain Text: 13 Date & Time: 2

LIMITATIONS

With such a large, diverse amount of data, there are bound to be some limitations that one will encounter. This is common when working with any culmination of data. In this section, I will go into detail about some of the limitations that I have identified with this data set.

First and foremost, there is potential for some analysis to appear skewed due to certain animal breeds being more popular than others. For example, if Labrador Retriever dogs are more

popular than Dachshund dogs in the Dallas area, then it may seem that Labradors are more aggressive because they have been involved in more incidents. In reality, this may not be true. However, the data may reflect this assumption since there are more Labrador Retrievers in the Dallas Area than Dachshunds. In order to remediate this limitation, I plan to use percentages and ratios in my analysis, rather than raw counts. This will allow me to make comparisons between different classifications relative to the total amount, rather than rely solely on the count of a category.

The second limitation that I have identified is that there is no unique Animal ID in the "Dallas 311 Service Requests" data set. Unfortunately, this means that there is no way to directly link the animal shelter data to the 311 service request data. This will make it difficult to extract insight from the 311 service request data relative to the animal shelter data. However, I should still be able to make general comparisons between the 311 service request data and the high-level information provided in the animal services data. For example, I will be able to make comparisons between the total animal-related 311 service requests to the number of animals actively registered in the Dallas Animal Shelter system.

Finally, due to the high amount of data, there are likely to be duplicate animal records, as well as some animals entered twice under different IDs. Unfortunately, this is a limitation that I will be unable to fully resolve. However, I will be able to analyze duplicate Animal IDs and merge or drop records where possible. Through this process, I will be able to "clean" the data to the point where each Animal ID is unique. Additionally, for the purposes of this analysis, I will have assumed that the collectors of this data were meticulous and did not enter an animal twice under different IDs.

ANALYSIS

No analysis or interpretation of data is complete without associated driving questions. For this analysis, I have compiled a list of four driving questions. These questions serve as the backbone of the analysis. Please find details regarding each of the four driving questions below:

1. What dog breeds are most popular among Dallas area residents?

   To answer this question, I used the following variables from my data set:

Table Name	Variable Name	Variable Type
Dallas Animal Shelter Data	AnimalType	Discrete
Dallas Animal Shelter Data	AnimalBreed	Discrete

   Using the above variables, I was able to create a table that aggregated each of the different dog breeds that were registered in the Dallas Animal Shelter system. I was then able to come up with counts of each type of animal that was registered and create associated data visualizations with the results. Assuming this Dallas Animal Shelter data is representative of the Dallas-area population, I can use this aggregation and the associated data visualizations to answer this driving question.

   For complete documentation of the analysis steps performed for this driving question, please see the "Analysis Documentation" section of the Appendix.

2. Which dog breeds are most involved in aggression-related incidents in the Dallas area based on the ratio of incidents per animal?

   I used the following variables from my data set to answer this question:

Table Name	Variable Name	Variable Type
Dallas Animal Shelter Data	AnimalType	Discrete
Dallas Animal Shelter Data	AnimalBreed	Discrete
Dallas Animals Field Data	AnimalIntakeId	Continuous
Dallas Animal Shelter Data	AnimalId	Continuous
Dallas Animals Field Data	ActivitySubType	Discrete

With the above-mentioned variables, I created multiple different views in the database. These views contained filters to only display aggression-related incidents where the animal involved was also registered in the Dallas Animal Shelter system. The views also filtered down based on animal breeds that had a frequency of at least 100 in the Dallas Animal Shelter system. In this way, I was able to display the number of aggression-related incidents by dog breed, and then use this information to calculate a ratio value that represented the number of incidents per dog breed. Data visualizations were also created using this data grouping for better interpretation of the data.

For complete documentation of the analysis steps performed for this driving question, please see the "Analysis Documentation" section of the Appendix.

3. Which areas of Dallas experience the highest number of animal aggression-related incidents?

   For this driving question I used the following variables:

Table Name	Variable Name	Variable Type
Dallas Animals Field Data	ZipCode	Continuous
Dallas Animals Field Data	ActivitySubType	Discrete
Dallas 311 Service Requests	Department	Discrete
Dallas 311 Service Requests	Address	Continuous
Dallas 311 Service Requests	ServiceRequestType	Discrete

   Utilizing these variables, I created multiple database views that filtered reported animal incidents based on incident type and zip code. The resulting aggregation from these views was a table displaying Dallas area zip codes and the number of aggression-related animal incidents that occurred in that zip code during the time period that the data was collected (Oct. 2016 to Oct. 2019). This aggregation would then be used to create a heat map that will allow a visualization of the data based on location.

   For complete documentation of the analysis steps performed for this driving question, please see the "Analysis Documentation" section of the Appendix.

4. Which dog breeds encounter the highest ratio of medical incidents?

   To provide insight pertaining to this question, I analyzed these variables:

Table Name	Variable Name	Variable Type
Dallas Animal Shelter Data	AnimalType	Discrete
Dallas Animal Shelter Data	AnimalBreed	Discrete
Dallas Animals Medical Data	AnimalId	Continuous
Dallas Animal Shelter Data	AnimalId	Continuous
Dallas Animals Medical Data	TreatmentSubType	Discrete
Dallas Animals Medical Data	TreatmentType	Discrete

   The above variables were used to aggregate medical incidents reported in the Dallas Animal Shelter system by animal type and breed. The views created from these variables were then used to produce ratios and data visualizations to assist in answering this driving question. This data was filtered down based on dog breeds that had a frequency of at least 100 in the Dallas Animal Shelter system.

For complete documentation of the analysis steps performed for this driving question, please see the "Analysis Documentation" section of the Appendix.

RESULTS

Please find details regarding the results of this analysis below. The following results are divided up based on the four driving questions which were introduced in the previous section.

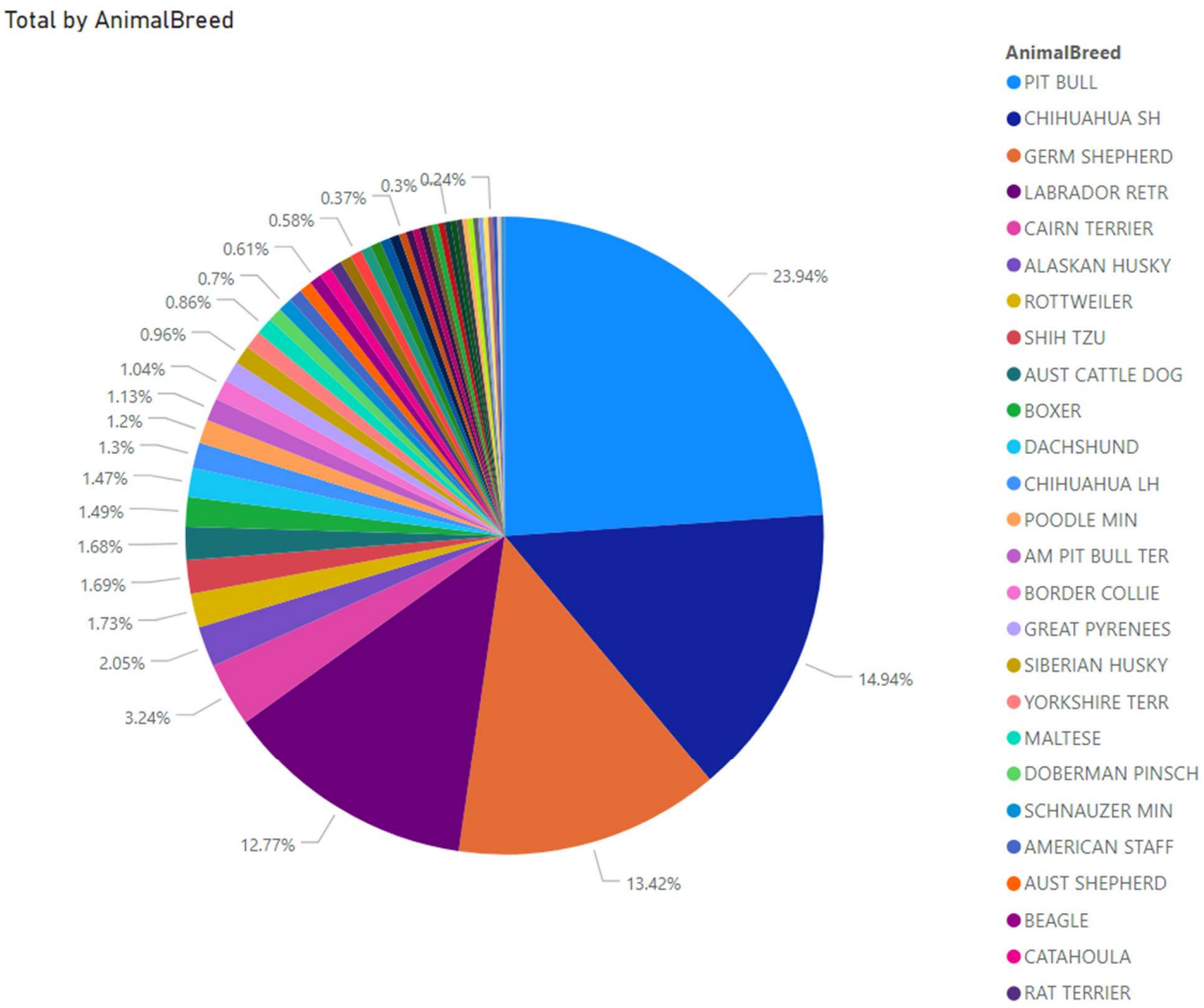
1. What dog breeds are most popular among Dallas area residents?

   I have included the resulting aggregated view below, which shows the number of registered dogs in the Dallas Animal Shelter system based on dog breed. The table below has been limited to only show the top 10 results from this view.

AnimalBreed	Total
PIT BULL	13447
CHIHUAHUA SH	8389
GERM SHEPHERD	7540
LABRADOR RETR	7170
CAIRN TERRIER	1822
ALASKAN HUSKY	1153
ROTTWEILER	969
SHIH TZU	949
AUST CATTLE DOG	941
BOXER	839

   As one can see, the animal breed that was registered the most in the Dallas Animal Shelter system is the Pit Bull, followed by the short-haired Chihuahua, and the German Shepherd.

   Below is a pie chart showing dog breeds as a percentage of the total registered dog population:



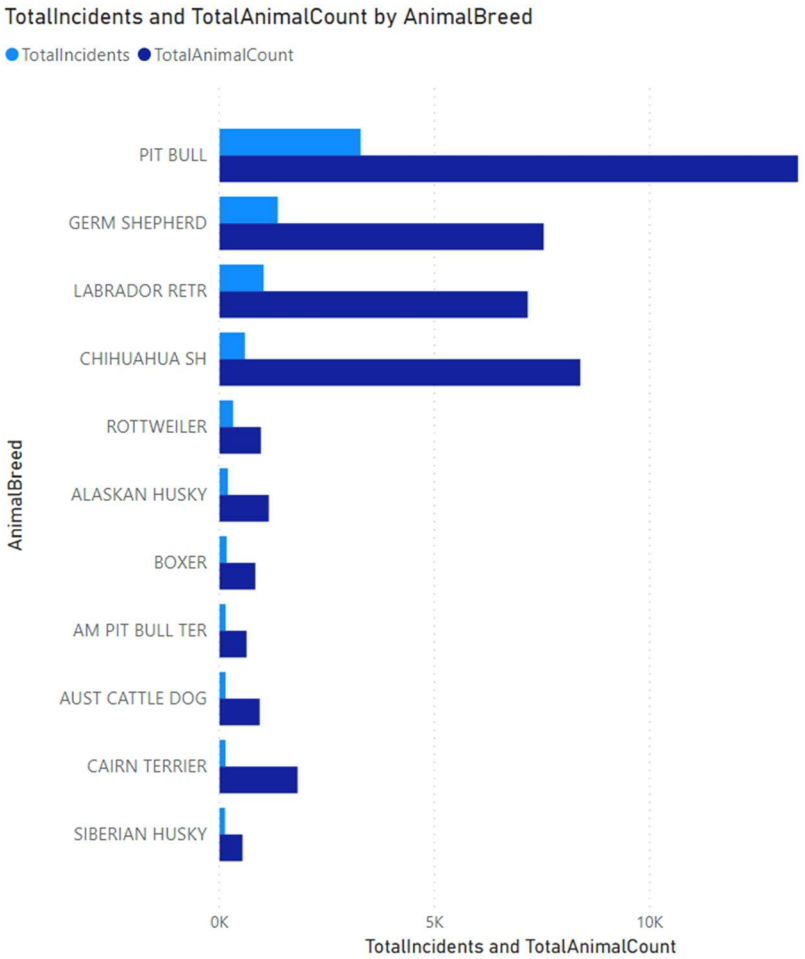
2. Which dog breeds are most involved in aggression-related incidents in the Dallas area based on the ratio of incidents per animal?

   The table below contains aggregated data which displays the number of aggression-related incidents per dog breed, the frequency of that dog breed in the Dallas Animal Shelter system, as well as the ratio of incidents per registered dog of that breed.

AnimalBreed	IncidentsRatio	TotalAnimalCount	TotalIncidents
AKITA	0.55	141	78
BELG MALINOIS	0.39	120	47
STAFFORDSHIRE	0.35	193	68
GREAT DANE	0.34	170	57
ROTTWEILER	0.33	969	324
BULL TERRIER	0.32	169	54
AMER BULLDOG	0.31	294	90
MASTIFF	0.28	201	56
AMERICAN STAFF	0.26	385	99
SIBERIAN HUSKY	0.25	541	136

Based on the above results, the Akita is involved in the most aggression-related incidents per registered animal in the Dallas area. The incident ratio for this dog breed is slightly over 0.55 per registered animal.

Furthermore, below I have included a bar graph that shows dog breeds' number of incidents alongside the number of registered dogs for that breed. This graph is ordered by number of incidents.



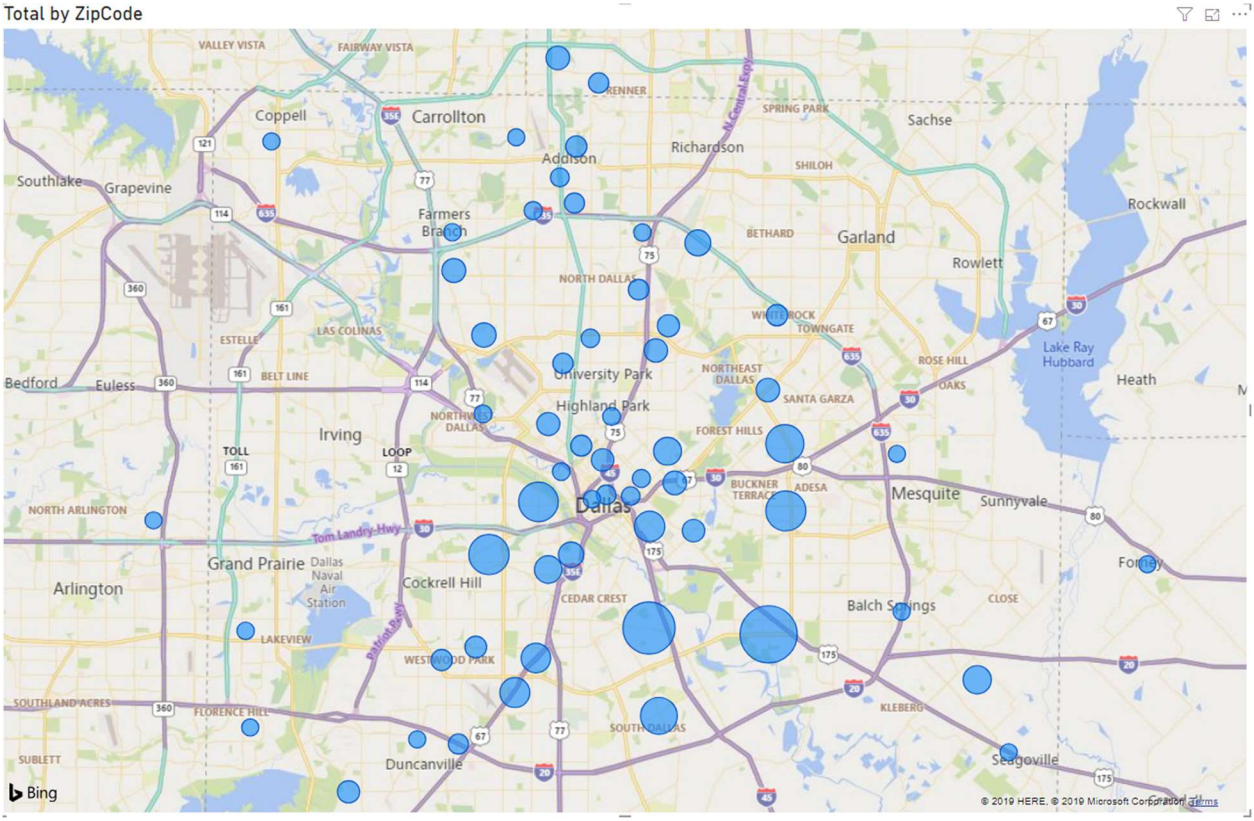
3. Which areas of Dallas experience the highest number of animal aggression-related incidents?

   The following table shows Dallas area zip codes, and the corresponding number of aggression-related incidents that occurred in that zip code from October of 2016 to October of 2019.

ZipCode	Total
75217	6160
75216	5397
75211	3553
75227	3526
75212	3451

   As can be seen, the 75217 zip code in the Dallas area has the highest number of reported aggression-related animal incidents over the specified time period (Oct. 2016 to Oct. 2019).

   Additionally, below is a map that visualizes these zip codes and their associated animal incident frequencies:

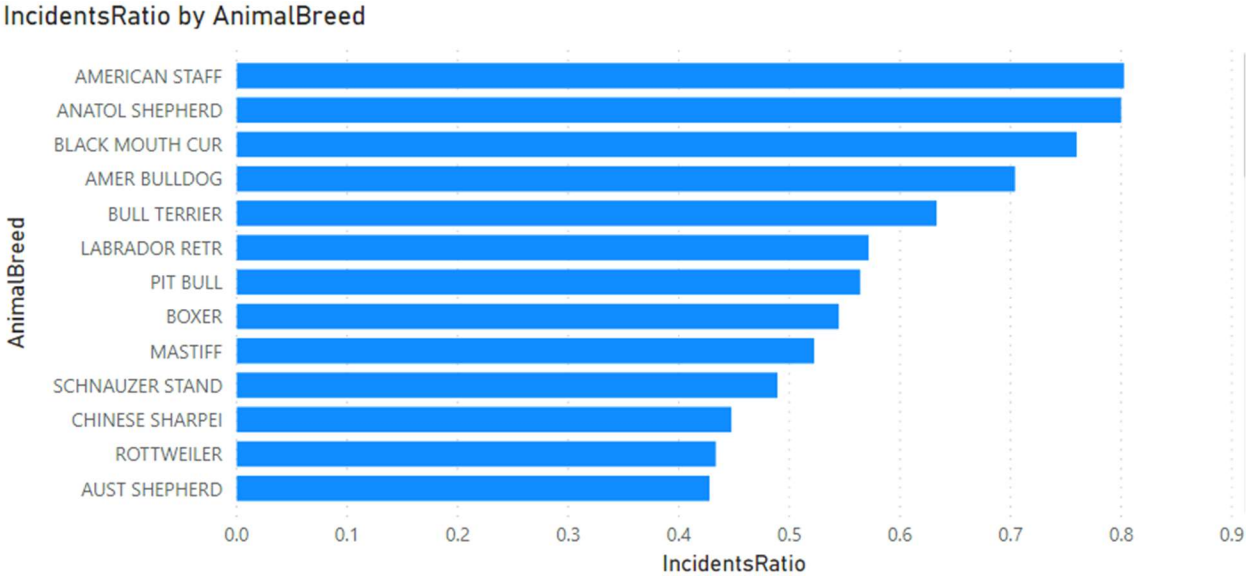


4. Which dog breeds encounter the highest percentage of medical incidents?

   The provided aggregated table below presents the number of recorded medical incidents per dog breed, the frequency of that dog breed in the Dallas Animal Shelter system, as well as the ratio of medical incidents per registered animal of that breed.

AnimalBreed	IncidentsRatio	TotalAnimalCount	TotalIncidents
AMERICAN STAFF	0.80	385	309
ANATOL SHEPHERD	0.80	105	84
BLACK MOUTH CUR	0.76	179	136
AMER BULLDOG	0.70	294	207
BULL TERRIER	0.63	169	107
LABRADOR RETR	0.57	7170	4099
PIT BULL	0.56	13447	7584
BOXER	0.54	839	457
MASTIFF	0.52	201	105
SCHNAUZER STAND	0.49	139	68

   The American Staffordshire, Anatolian Shepherd, and Black Mouth Cur dog breeds encounter the most medical incidents per registered animal. The medical incident ratios for these animal breeds are 0.80, 0.80, and 0.76, respectively. This can also be seen below in the associated data visualization:



   Additionally, I have included below a table which shows dog breeds, and the percentage of heartworm tests that returned positive for that breed. A positive heartworm test means that the dog does have heartworms.

AnimalBreed	CountNegative	CountPositive	Total	PercentPositive
BOXER	13	13	26	50.00
LABRADOR RETR	109	45	154	29.22
CHIHUAHUA SH	44	17	62	27.42
GERM SHEPHERD	123	42	165	25.45
PIT BULL	290	89	380	23.42

DISCUSSION

INTERPRETATION OF RESULTS

I have included my interpretation of the above results below:

1. What dog breeds are most popular among Dallas area residents?

   Based on the table provided, we can see that the Pit Bull is far and away the most popular dog breed based on the Dallas Animal Shelter data. This is interesting to note because the Pit Bull is a breed that is notoriously involved in controversy around potential aggressive behavior. Following the Pit Bull, we can see that the second, third, and fourth most popular dog breeds are all somewhat close in popularity. However, there is a large drop off after Labrador Retrievers in fourth. Therefore, if someone is interested in getting a dog based on the popularity of breeds in general, the Pit Bull, short-haired Chihuahua, German Shepherd, or Labrador Retriever would be the top choices.

2. Which dog breeds are most involved in aggression-related incidents in the Dallas area based on ratio?

   Based on the table, we can see that the Akita dog breed has the highest number of aggression-related incidents per registered dog in the Dallas area. This is interesting to me, as from my experience it is widely believed that the Pit Bull is involved in the most aggressive incidents.

   In the associated data visualization, we can see that the Pit Bull does have the highest number of incidents recorded. However, this appears to be solely due to the Pit Bull being the most popular breed. The better indicator of aggression is the ratio of incidents per registered animal. When it comes to this, the Pit Bull is not even listed in the top ten.

3. Which areas of Dallas experience the highest number of animal aggression-related incidents?

   Based on the provided table of incidents per zip code, we can see that the top two zip codes (75217 and 75216) are very close to each other geographically. This is also evident when viewing the geo-spatial map with indicators showing frequency of incidents. The bigger the circles are in the map, the more incidents occurred in that area. In this map, the most incidents appear in the southeast area of Dallas (containing zip codes 75217 and 75216).

   Therefore, based on these results, it appears that the southeast area of Dallas has the highest risk of encountering an aggressive animal.

4. Which dog breeds encounter the highest percentage of medical incidents?

   Based on the table provided, we can see that the American Staffordshire and Anatolian Shepherd dog breeds seem to encounter the highest number of medical incidents in the Dallas Area. Additionally, the Boxer has the highest ratio of positive heartworm test per test performed. Therefore, based on these results, the Boxer has the highest risk of having heartworms in the Dallas area.

   An interesting point is that the Labrador Retriever and Pit Bull are both included in the top seven of medical incidents ratio, as well as the top five when looking at ratio of positive heartworm tests. Therefore, my interpretation of each of these results would be that the Boxer, Labrador Retriever, and Pit Bull seem to have the highest risk of encountering more frequent medical incidents.

RECOMMENDATIONS FOR ADDITIONAL RESEARCH

Looking beyond this analysis, I believe that there is potential for additional research that could build upon these results. I have provided an overview of some gaps in this analysis, and the potential for additional research in those areas, below:

1. Comparison of resident demographics and crime data to aggressive animal incidents.

   I believe that it would be beneficial to compile/gather data regarding the demographic and crime information of Dallas area residents. This geo-spatial data could then be compared with the geo-spatial data in this analysis. This would improve the value of this analysis by allowing comparisons of aggressive animal incidents to Dallas resident income, race, crime rates, etc. in that area.

2. Compare this research to the national average for similar statistics.

   I also believe that it would be interesting to compare the aggression and medical results of this analysis to a similar analysis performed on a national scale. This would allow a comparison of the aggression and medical data from this analysis to a national average. This comparison would be beneficial to identify potential trends and/or outliers in the Dallas area specifically that potential owners should be aware of.

CONCLUSION

In conclusion, the purpose of this analysis is to produce insight that will better enable long-term happiness in prospective pet owners who are looking into what type of dog they would like to own. During this analysis, I dug deep into Animal Services Data from the Dallas metropolitan area. This included general Animal Shelter data, Animal Medical data, Animal Incident data, and Dallas 311 Service Request data.

During this analysis, I first cleaned and formatted the data. Following this, the data was imported into Azure Blob Storage. From there, it was integrated into Azure SQL Data Warehouse using *PolyBase*. Then I created multiple database views which were used to gather insights. These database views were also used to create rich data visualizations in Power BI.

The results of this analysis indicated that the Pit Bull is easily the most popular dog breed in the Dallas area. This is according to the number of registered dogs in the Dallas Animal Shelter system. However, the Pit Bull is also relevant when discussing the animals with the highest

ratio of medical incidents and positive heartworm test results. Specifically, the Boxer, Labrador Retriever, and Pit Bull appear to be the dog breeds that have the highest risk to encounter more frequent medical incidents. The Akita dog breed has the highest ratio of aggression-related incidents per registered dog. Additionally, the southeast area of Dallas (specifically the 75217 and 75216 zip codes) experiences the highest number of aggression-related animal incidents.

Looking forward, there are two gaps in this analysis that I feel warrant potential future research. First, I feel that a comparison of resident demographics and crime data to the geo-spatial aggression incident data in this analysis would be insightful and beneficial. Secondly, a comparison of this analysis to similar statistics at a national level would be insightful, as this would allow the identification of potential trends and/or outliers in the Dallas area.

APPENDIX

ENVIRONMENT CREATION

1. Create an *Azure Resource Group*.
2. In the above *Azure Resource Group*, create an *Azure Blob Storage* account.

   More information may be found at the *Microsoft Azure Blob Storage* documentation link in the references of this report.
3. In the *Azure Blob Storage* account, create four different containers.

   One container will be needed for each of the four data sets that will be imported. I named the containers the same as the table schema names: AnimalShelterData, ServiceRequests, AnimalFieldData, and AnimalMedicalData.
4. Create an *Azure SQL Server* and an *Azure SQL Data Warehouse*.

   More information may be found at the *Microsoft Azure SQL Data Warehouse* documentation link in the references of this report.
5. Create an external connection from the *Azure SQL Data Warehouse* to the *Blob Storage* account using the following SQL:

```
CREATE MASTER KEY;
CREATE DATABASE SCOPED CREDENTIAL AzureStorageCredential
WITH
    IDENTITY = 'dallasanimals',
    SECRET = '<AccessKeyHere>'
;
CREATE EXTERNAL DATA SOURCE AnimalShelterData
WITH (
    TYPE = HADOOP,
    LOCATION = '<ContainerLocationHere>',
    CREDENTIAL = AzureStorageCredential
);
CREATE EXTERNAL DATA SOURCE AnimalFieldData
WITH (
    TYPE = HADOOP,
    LOCATION = '<ContainerLocationHere>',
    CREDENTIAL = AzureStorageCredential
);
CREATE EXTERNAL DATA SOURCE AnimalMedicalData
WITH (
    TYPE = HADOOP,
    LOCATION = '<ContainerLocationHere>',
    CREDENTIAL = AzureStorageCredential
);
CREATE EXTERNAL DATA SOURCE ServiceRequests
WITH (
    TYPE = HADOOP,
    LOCATION = '<ContainerLocationHere>',
    CREDENTIAL = AzureStorageCredential
);
```

DATA CLEANING & IMPORT

Please find below a detailed overview of the steps performed to clean and import each data set into the *Microsoft Azure* cloud environment.

Dallas Animals Field Data:

1. Import each of the three .csv files into *OpenRefine* application (FY2017, FY2018, FY2019).
2. Remove the following columns, as their data was either not needed or it was too ambiguous:

   Street Number, Street Direction, Street Name, Street Type, Apartment, City, State, Census Tract, Council District, Mapsco Page, Animal Description, Tag Number, Activity Sequence, Activity Quantity 1, Activity Result 1, Activity Quantity 2, Activity Result 2, Activity Quantity 3, Activity Result 3, Activity Quantity 4, Activity Result 4, Activity Quantity 5, Activity Result 5, Final Activity Quantity, Final Activity Result, Activity Comment, Officer ID, Activity Status, Cancel Reason, Radio Signal Code, Call Date/Time, Dispatch Date/Time, Working Date/Time, Completed Date/Time, User ID, Month, Year, Service Request Number
3. Export the *OpenRefine* project as a .csv file.
4. Import the .csv file into an *Azure Blob Storage* container.
5. Run the following SQL scripts in the *Azure SQL Data Warehouse* to import the data:

```
CREATE EXTERNAL TABLE dbo.TempFieldData (
    [ActivityNumber] nvarchar(20) NULL,
    [ActivityType] nvarchar(50) NULL,
    [ActivitySubType] nvarchar(50) NULL,
    [ActivityPriority] nvarchar(20) NULL,
    [StreetAddress] nvarchar(100) NULL,
    [ZipCode] nvarchar(20) NULL,
    [AnimalIntakeId] nvarchar(20) NULL,
    [AnimalType] nvarchar(20) NULL,
    [CreatedOn] nvarchar(50) NULL
)
WITH (
    LOCATION='/',
    DATA_SOURCE=AnimalFieldData,
    FILE_FORMAT=TextFile
);
CREATE TABLE [dbo].[AnimalFieldData]
WITH (
    CLUSTERED COLUMNSTORE INDEX,
    DISTRIBUTION = ROUND_ROBIN
)
AS
SELECT * FROM [dbo].[TempFieldData];

DROP EXTERNAL TABLE [dbo].[TempFieldData];
```

6. Run the following SQL queries to further clean the data:

```
update AnimalFieldData
```

```
set ActivityType = 'LOOSE'
where ActivityType = 'ANILOOSE';

update AnimalFieldData
set ActivityType = 'OTHER'
where ActivityType = 'ANIOTHER';

update AnimalFieldData
set ActivitySubtype = 'NOISE'
where ActivityType = 'BARKING';

update AnimalFieldData
set ActivitySubtype = 'BITE'
where ActivityType = 'ANMLBITE';
```

Dallas Animals Medical Data:

1. Import each of the three .csv files into *OpenRefine* application (FY2017, FY2018, FY2019).
2. Remove the following columns, as their data was either not needed or it was too ambiguous:

   Treatment_Time, Treatment_By, Number_of_Days, Temperature, Month, Year
3. Export the *OpenRefine* project as a .csv file.
4. Import the .csv file into an *Azure Blob Storage* container.
5. Run the following SQL scripts in the *Azure SQL Data Warehouse* to import the data:

```
CREATE EXTERNAL TABLE dbo.TempMedicalData (
    [TreatmentNumber] nvarchar(20) NULL,
    [AnimalId] nvarchar(20) NULL,
    [ConditionType] nvarchar(100) NULL,
    [TreatmentDate] nvarchar(20) NULL,
    [TreatmentType] nvarchar(50) NULL,
    [TreatmentSubtype] nvarchar(20) NULL,
    [Weight] decimal NULL,
    [VisitNumber] nvarchar(20) NULL,
    [VisitType] nvarchar(50) NULL,
    [VisitReason] nvarchar(20) NULL
)
WITH (
    LOCATION='/',
    DATA_SOURCE=AnimalMedicalData,
    FILE_FORMAT=TextFile
);
CREATE TABLE [dbo].[AnimalMedicalData]
WITH (
    CLUSTERED COLUMNSTORE INDEX,
    DISTRIBUTION = ROUND_ROBIN
)
AS
SELECT * FROM [dbo].[TempMedicalData];
DROP EXTERNAL TABLE [dbo].[TempMedicalData];
```

6. Run the following SQL queries to further clean the data:

```
update AnimalMedicalData
set TreatmentSubtype = 'NEGATIVE'
where TreatmentSubtype like '%NEG%'

update AnimalMedicalData
set TreatmentSubtype = 'POSITIVE'
where TreatmentSubtype like '%POS%'

update AnimalMedicalData
set TreatmentSubtype = 'POSITIVE'
where TreatmentType like '%POS%'

update AnimalMedicalData
set TreatmentSubtype = 'NEGATIVE'
where TreatmentType like '%NEG%'

update AnimalMedicalData
set TreatmentType = 'HEARTWORM TEST'
where TreatmentType like '%HEARTWORM%'

update AnimalMedicalData
set TreatmentType = 'MICROCHIP'
where TreatmentType like '%MICROCHIP%' or TreatmentType like
'%MICRO CHIP%'

update AnimalMedicalData
set TreatmentSubtype = 'POSITIVE'
where TreatmentType = 'MICROCHIP'
```

Dallas 311 Service Requests:

1. Import each of the two .csv files into *OpenRefine* application (Oct. 2016 and Oct. 2018).
2. Remove the following columns, as their data was either not needed or it was too ambiguous:

   City Council District, ERT (Estimated Response Time), Overall SR Due Date, Status, Update Date, Closed Date, Outcome, Method Received, Lat Long Location, Unique Key, Location 1
3. Export the *OpenRefine* project as a .csv file.
4. Import the .csv file into an *Azure Blob Storage* container.
5. Run the following SQL scripts in the *Azure SQL Data Warehouse* to import the data:

```
CREATE EXTERNAL TABLE dbo.TempServiceRequests (
    [ServiceRequestNumber] nvarchar(20) NULL,
    [Address] nvarchar(100) NULL,
    [Department] nvarchar(50) NULL,
    [ServiceRequestType] nvarchar(100) NULL,
    [CreatedDate] nvarchar(50) NULL,
    [Priority] nvarchar(20) NULL
```

```
)
WITH (
    LOCATION='/',
    DATA_SOURCE=ServiceRequests,
    FILE_FORMAT=TextFile
);
CREATE TABLE [dbo].[ServiceRequests]
WITH (
    CLUSTERED COLUMNSTORE INDEX,
    DISTRIBUTION = ROUND_ROBIN
)
AS
SELECT * FROM [dbo].[TempServiceRequests];
DROP EXTERNAL TABLE [dbo].[TempServiceRequests];
```

6. Run the following SQL queries to further clean the data:

```
update ServiceRequests
set ServiceRequestType = 'Animal - Bite'
where Department = 'Dallas Animal Services' and
ServiceRequestType like '%Bite%'

update ServiceRequests
set ServiceRequestType = 'Animal - Aggressive'
where Department = 'Dallas Animal Services' and
ServiceRequestType like '%Aggressive%'

update ServiceRequests
set ServiceRequestType = 'Animal - Attack'
where Department = 'Dallas Animal Services' and
ServiceRequestType like '%Attack%'
```

Dallas Animal Shelter Data:

1. Import the single .csv file into *OpenRefine* application.
2. Remove the following columns, as their data was either not needed or it was too ambiguous:

   Kennel Number, Tag Type (all blank), Activity Number, Activity Sequence, Source ID, Census Tract, Council District, Intake Total, Staff ID, Intake Time, Due Out, Outcome Time, Receipt Number, Impound Number, Service Request Number (majority blank), Additional Information, Month, Year
3. Make the following edits in *OpenRefine*:
   - Animal Breed, changed "Dachshund LH" to "Dachshund"
   - Animal Breed, changed "Dachshund WH" to "Dachshund"
   - Animal Breed, changed "Domestic LH" to "Domestic"
   - Animal Breed, changed "Domestic MH" to "Domestic"
   - Animal Breed, changed "Domestic SH" to "Domestic"
   - Animal Breed, changed "DUTCH" to "DUTCH SHEPHERD"

- Animal Type, changed "D" to "DOG"
- Hold Request, changed "RESCU ADOP" to "ADOP RESCU"
- Hold Request, changed "FOSTER" to "RESCU FOSTER"
- Hold Request, changed "RESC FOSTR" to "RESCU FOSTER"
- Hold Request, changed "AGG" to "DANGEROUS"

4. Export the *OpenRefine* project as a .csv file.
5. Import the .csv file into an *Azure Blob Storage* container.
6. Run the following SQL scripts in the *Azure SQL Data Warehouse* to import the data:

```
CREATE EXTERNAL TABLE dbo.TempAnimalShelter (
    [AnimalId] nvarchar(20) NULL,
    [AnimalType] nvarchar(20) NULL,
    [AnimalBreed] nvarchar(50) NULL,
    [KennelStatus] nvarchar(50) NULL,
    [IntakeType] nvarchar(50) NULL,
    [IntakeSubtype] nvarchar(50) NULL,
    [Reason] nvarchar(50) NULL,
    [IntakeDate] nvarchar(50) NULL,
    [IntakeCondition] nvarchar(100) NULL,
    [HoldRequest] nvarchar(50) NULL,
    [OutcomeType] nvarchar(50) NULL,
    [OutcomeSubtype] nvarchar(50) NULL,
    [OutcomeDate] nvarchar(50) NULL,
    [OutcomeCondition] nvarchar(100) NULL,
    [ChipStatus] nvarchar(50) NULL,
    [AnimalOrigin] nvarchar(50) NULL
)
WITH (
    LOCATION='/',
    DATA_SOURCE=AnimalShelterData,
    FILE_FORMAT=TextFile
);
CREATE TABLE [dbo].[AnimalShelter]
WITH (
    CLUSTERED COLUMNSTORE INDEX,
    DISTRIBUTION = ROUND_ROBIN
)
AS
SELECT * FROM [dbo].[TempAnimalShelter];
DROP EXTERNAL TABLE [dbo].[TempAnimalShelter];
```

7. Run the following SQL queries to further clean the data:

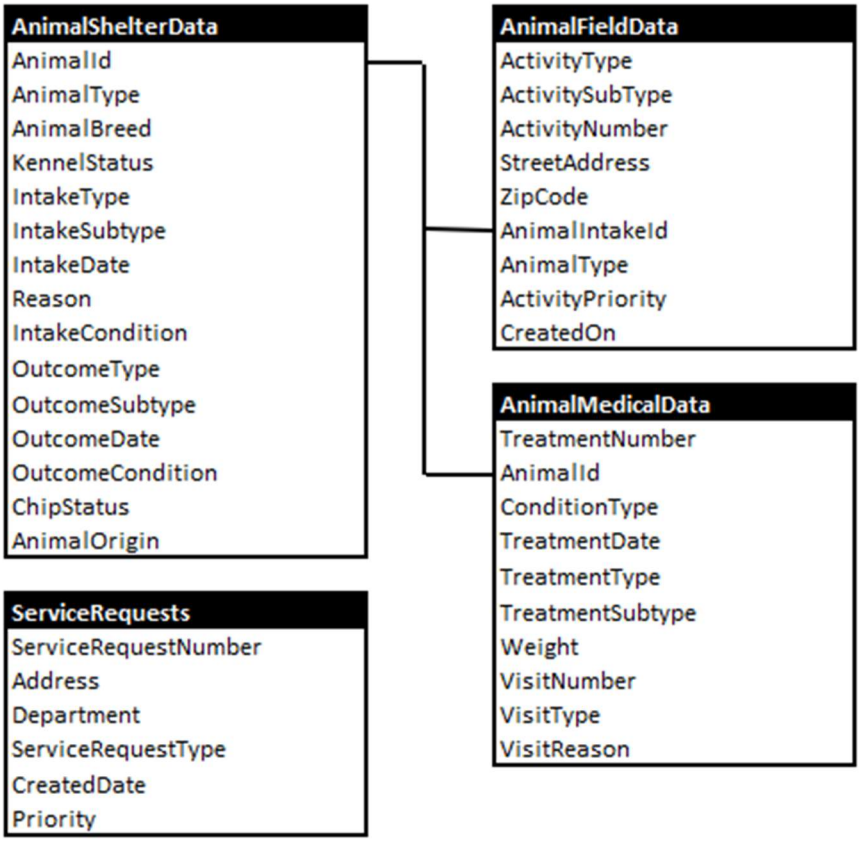
```
update AnimalShelterData
set AnimalBreed = 'PIT BULL'
where AnimalBreed = 'PITBULL'
```

DATA DICTIONARY

Dallas Animal Medical Record (Table Schema Name: AnimalMedicalData)		
Variable Name	Data Type	Description
TreatmentNumber	Plain Text	Treatment number assigned to the patient.
AnimalId	Plain Text	ID of the animal receiving treatment.
ConditionType	Plain Text	Condition of the animal receiving treatment.
TreatmentDate	Date & Time	Date of treatment.
TreatmentType	Plain Text	Type of treatment being administered.
TreatmentSubtype	Plain Text	Sub-type of treatment being administered.
Weight	Number	Weight of the animal receiving treatment.
VisitNumber	Plain Text	Visit number (can have multiple treatment #s associated).
VisitType	Plain Text	Type of visit.
VisitReason	Plain Text	Reason for visit.
Dallas Animals Field Data (Table Schema Name: AnimalFieldData)		
Variable Name	Data Type	Description
ActivityNumber	Plain Text	Number assigned to the activity record.
ActivityType	Plain Text	Type of activity record.
ActivitySubType	Plain Text	Sub-type of the activity type.
StreetAddress	Plain Text	Street address where the activity occurred.
ZipCode	Plain Text	Zip code where the activity occurred.
AnimalIntakeId	Plain Text	ID of the animal involved in the activity.
AnimalType	Plain Text	Type of animal involved in the activity.
ActivityPriority	Plain Text	Priority of the activity.
CreatedOn	Date & Time	Date the activity occurred.
Dallas 311 Service Requests (Table Schema Name: ServiceRequests)		
Variable Name	Data Type	Description
ServiceRequestNumber	Plain Text	Number assigned to the 311 service request.
Address	Plain Text	Address where the 311 service request was reported.
Department	Plain Text	Department that the 311 service request was opened for.

ServiceRequestType	Plain Text	Type of 311 service request opened.
CreatedDate	Date & Time	Date the 311 service request was created.
Priority	Plain Text	Priority assigned to the 311 service request.
Dallas Animal Shelter Data (Table Schema Name: AnimalShelterData)		
Variable Name	Data Type	Description
AnimalId	Plain Text	ID of the animal at the shelter.
AnimalType	Plain Text	Type of animal.
AnimalBreed	Plain Text	Breed of the animal.
KennelStatus	Plain Text	Residency status of the animal in the Animal Shelter system.
IntakeType	Plain Text	Type of intake for this animal.
IntakeSubtype	Plain Text	Sub-type of the intake type for this animal.
Reason	Plain Text	Reason the animal was brought to the shelter.
IntakeDate	Date & Time	Date the animal was received by the shelter.
IntakeCondition	Plain Text	Condition of the animal upon reception by the shelter.
OutcomeType	Plain Text	Type of outcome associated with the animal leaving shelter.
OutcomeSubtype	Plain Text	Sub-type for the outcome type for this animal.
OutcomeDate	Date & Time	Date the outcome occurred for this animal.
OutcomeCondition	Plain Text	Condition of the animal upon outcome.
ChipStatus	Plain Text	Whether the animal was micro-chipped or not.
AnimalOrigin	Plain Text	Origin of the animal.

DATABASE DESIGN



ANALYSIS DOCUMENTATION

1. What animal types and breeds are most popular among Dallas area residents?

```
Create view CountAnimals
As
select AnimalType, AnimalBreed, Count(1) as Total
from AnimalShelterData
group by AnimalType, AnimalBreed
order by Total desc
go;
```

   The above was used to create the "Popular Dashboard" in the "Dashboard Visualizations" section of the Appendix.

2. Which animal types and breeds are most commonly involved in aggression-related incidents in the Dallas area based on percentage?

```
Create view FieldIncidentsAnimalType
as
select ASD.AnimalType,
      ASD.AnimalBreed,
      count(1) as Total
from AnimalFieldData AFD
left outer join AnimalShelterData ASD
```

```
on AFD.AnimalIntakeId = ASD.AnimalId
where (ActivitySubType = 'AGGRESSION' or ActivitySubType = 'BITE')
      and ASD.AnimalType is not null
Group by ASD.AnimalType, ASD.AnimalBreed
Go;
Create view FieldIncidentsRatio
as
      select  CONCAT(FI.AnimalType,  '  -  ',  FI.AnimalBreed)  AS
Animal,
            FI.Total as TotalIncidents,
            CA.Total as TotalAnimalCount,
            (CAST(FI.Total AS decimal) / CAST(CA.Total AS decimal))
as IncidentsRatio
      from FieldIncidentsAnimalType FI
      inner join CountAnimals CA
            on FI.AnimalBreed = CA.AnimalBreed
      where CA.Total >= 100
go;
```

The above was used to create the "Aggression Dashboard" in the "Dashboard Visualizations" section of the Appendix.

3. Which areas of Dallas experience the highest number of animal aggression-related incidents?

```
Create view IncidentsByZipCode
As
SELECT ZipCode, count(1) as Total
from AnimalFieldData
where (ActivitySubType = 'AGGRESSION' or ActivitySubType = 'BITE')
and (ZipCode <> 'UNK' and ZipCode is not NULL and ZipCode <>
'00000')
Group by ZipCode
go;
Create view SRIncidentsZipCode
as
select RIGHT(Address, 5) as ZipCode, count(1) as Total
from ServiceRequests
where Department = 'Dallas Animal Services' and LEFT(RIGHT(Address,
5), 1) = '7' and
    (ServiceRequestType = 'Animal - Bite'
    or ServiceRequestType = 'Animal - Attack'
    or ServiceRequestType = 'Animal - Aggressive')
group by RIGHT(Address, 5), ServiceRequestType
go;
create view TotalIncidentsByZipCode
as
      select ZipCode, Sum(Total) as Total
      from (select *
            from IncidentsByZipCode
            union all
            select *
            from SRIncidentsZipCode) as Incidents
```

```
      group by ZipCode
go;
```

The above was used to create the "Location Dashboard" in the "Dashboard Visualizations" section of the Appendix.

4. Which animal types and breeds encounter the highest percentage of medical incidents?

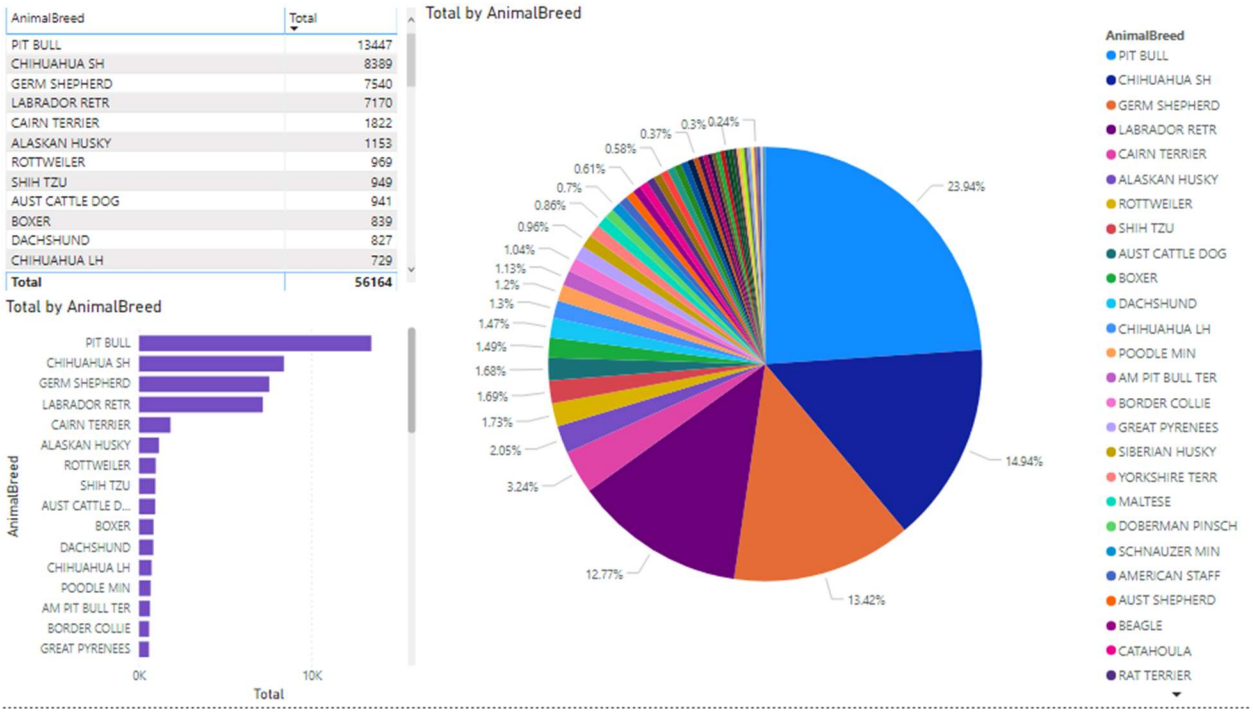
```
Create view MedicalCountBreed
As
select AnimalType, AnimalBreed, count(1) as TotalIncidents
from AnimalMedicalData AMD
left outer join AnimalShelterData ASD
    on AMD.AnimalId = ASD.AnimalId
where AnimalType is not NULL
group by AnimalType, AnimalBreed
order by TotalIncidents desc
go;
create view MedicalIncidentsRatio
as
      select CONCAT(m.AnimalType, ' - ', m.AnimalBreed) as
Animal,
            m.AnimalType,
            m.AnimalBreed,
            m.TotalIncidents as TotalIncidents,
            c.Total as TotalAnimalCount,
            (CAST(m.TotalIncidents AS decimal) / CAST(c.Total AS
decimal)) as IncidentsRatio
      from MedicalCountBreed m
      inner join CountAnimals c
            on m.AnimalBreed = c.AnimalBreed
            and m.AnimalType = c.AnimalType
      where c.Total >= 100
go;
create view DogHeartwormTestResultsBreed
as
SELECT asd.AnimalBreed,
    sum(
        CASE TreatmentSubtype
            WHEN 'POSITIVE' THEN 0
            WHEN 'NEGATIVE' THEN 1
        END
    ) as CountNegative,
    count(1) as Total,
    (cast(sum(
        CASE TreatmentSubtype
            WHEN 'POSITIVE' THEN 0
            WHEN 'NEGATIVE' THEN 1
        END
    ) as decimal) / cast(count(1) as decimal)) * 100 as
PercentNegative
from AnimalMedicalData amd
left outer join AnimalShelterData asd
      on amd.AnimalId = asd.AnimalId
```

```
where TreatmentType = 'HEARTWORM TEST'
group by asd.AnimalBreed;
go;
```

The above was used to create the "Medical Dashboard" in the "Dashboard Visualizations" section of the Appendix.

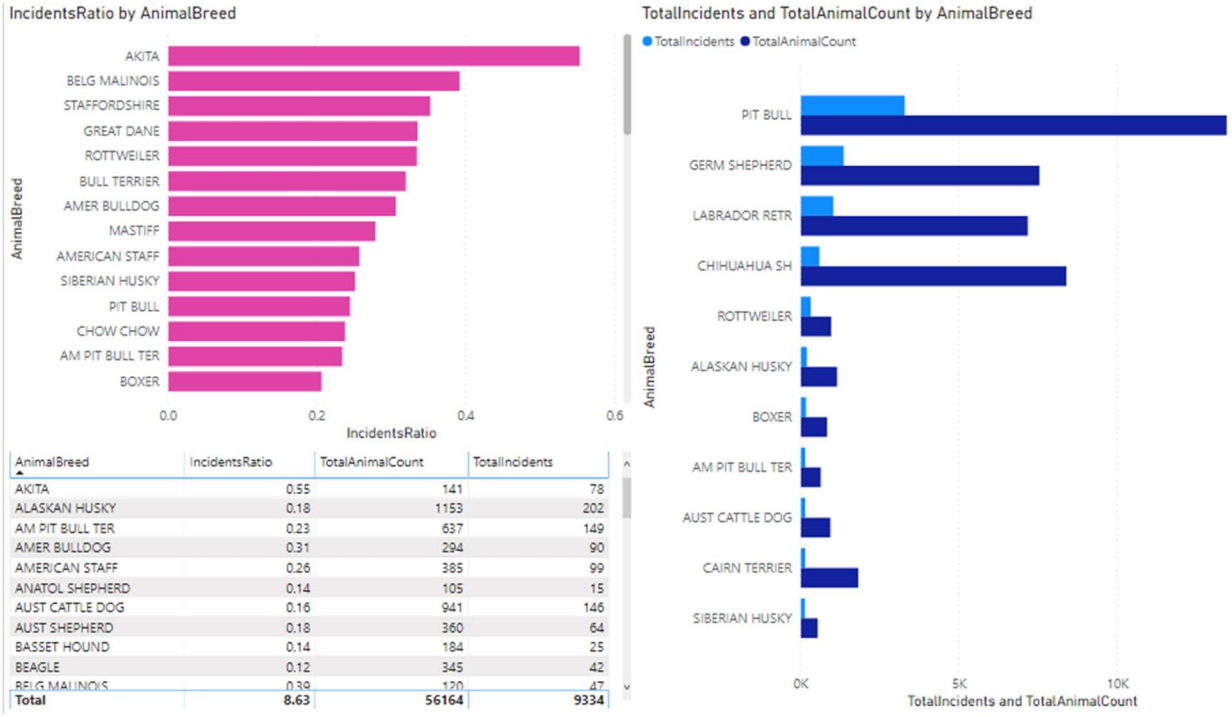
DASHBOARD VISUALIZATIONS

1. Popular Dashboard



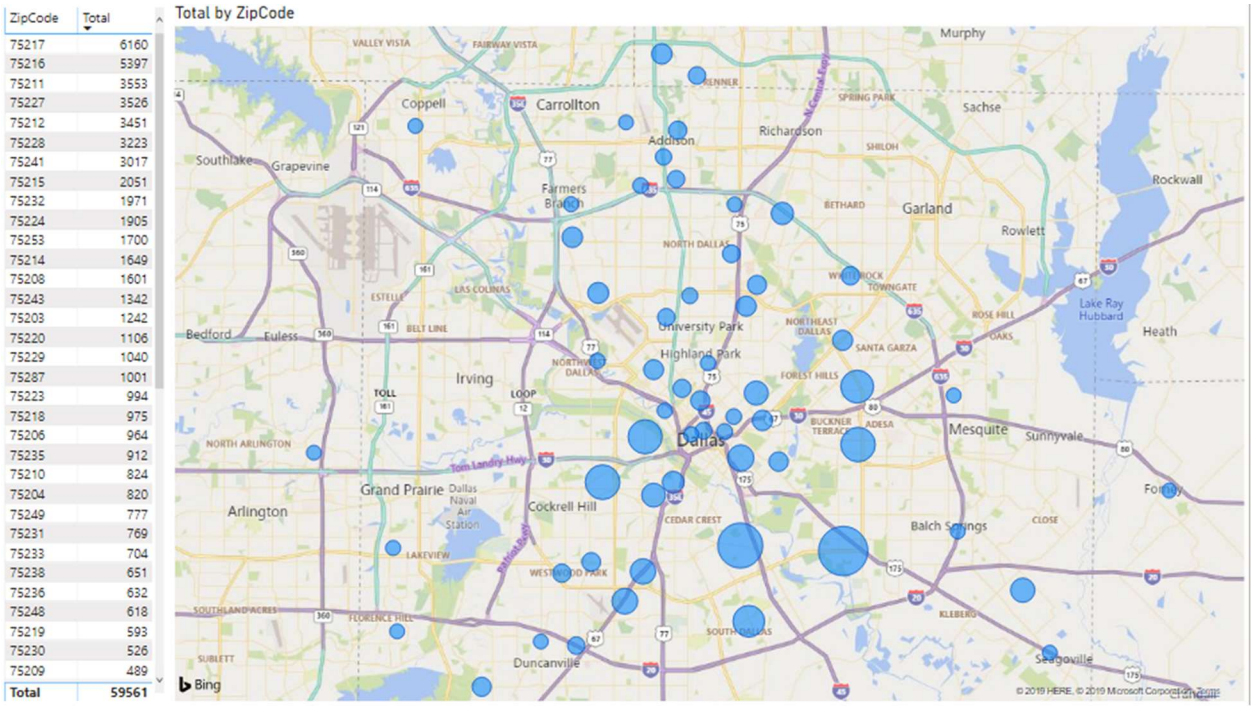
AnimalBreed	Total
PIT BULL	13447
CHIHUAHUA SH	8389
GERM SHEPHERD	7540
LABRADOR RETR	7170
CAIRN TERRIER	1822
ALASKAN HUSKY	1153
ROTTWEILER	969
SHIH TZU	949
AUST CATTLE DOG	941
BOXER	839
DACHSHUND	827
CHIHUAHUA LH	729
Total	**56164**

2. Aggression Dashboard



AnimalBreed	IncidentsRatio	TotalAnimalCount	TotalIncidents
AKITA	0.55	141	78
ALASKAN HUSKY	0.18	1153	202
AM PIT BULL TER	0.23	637	149
AMER BULLDOG	0.31	294	90
AMERICAN STAFF	0.26	385	99
ANATOL SHEPHERD	0.14	105	15
AUST CATTLE DOG	0.16	941	146
AUST SHEPHERD	0.18	360	64
BASSET HOUND	0.14	184	25
BEAGLE	0.12	345	42
BELG MALINOIS	0.39	120	47
Total	**8.63**	**56164**	**9334**

3. Location Dashboard

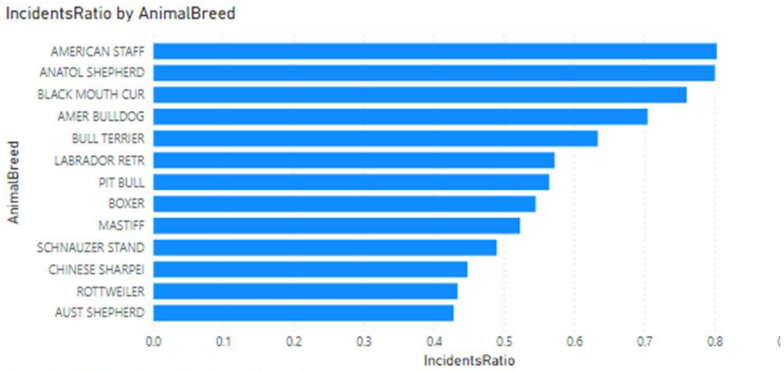


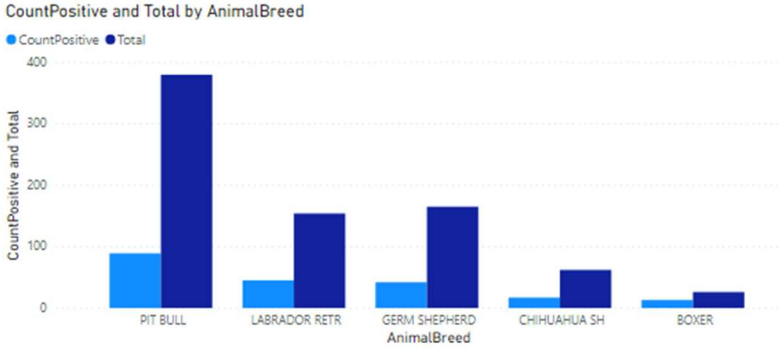
ZipCode	Total
75217	6160
75216	5397
75211	3553
75227	3526
75212	3451
75228	3223
75241	3017
75215	2051
75232	1971
75224	1905
75253	1700
75214	1649
75208	1601
75243	1342
75203	1242
75220	1106
75229	1040
75287	1001
75223	994
75218	975
75206	964
75235	912
75210	824
75204	820
75249	777
75231	769
75233	704
75238	651
75236	632
75248	618
75219	593
75230	526
75209	489
Total	**59561**

4. Medical Dashboard

AnimalBreed	IncidentsRatio	TotalAnimalCount	TotalIncidents
AMERICAN STAFF	0.80	385	309
ANATOL SHEPHERD	0.80	105	84
BLACK MOUTH CUR	0.76	179	136
AMER BULLDOG	0.70	294	207
BULL TERRIER	0.63	169	107
LABRADOR RETR	0.57	7170	4099
PIT BULL	0.56	13447	7584
BOXER	0.54	839	457
MASTIFF	0.52	201	105
SCHNAUZER STAND	0.49	139	68
CHINESE SHARPEI	0.45	143	64
ROTTWEILER	0.43	969	420
AUST SHEPHERD	0.43	360	154
MIN PINSCHER	0.41	180	73
BORDER COLLIE	0.38	585	224
GERM SHEPHERD	0.38	7540	2831
BEAGLE	0.37	345	128
STAFFORDSHIRE	0.36	193	70
CHOW CHOW	0.25	328	81
POINTER	0.24	323	78
WELSH CORGI CAR	0.24	152	36
AM PIT BULL TER	0.22	637	141
SHIH TZU	0.17	949	163
Total	**12.54**	**55211**	**19578**

AnimalBreed	CountNegative	CountPositive	Total	PercentPositive
BOXER	13	13	26	50.00
LABRADOR RETR	109	45	154	29.22
CHIHUAHUA SH	44	17	62	27.42
GERM SHEPHERD	123	42	165	25.45
PIT BULL	290	89	380	23.42
Total	**579**	**206**	**787**	**155.52**





REFERENCES

Dallas OpenData. 311 Service Requests from October 1, 2016 to October 1, 2018. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/311-Service-Requests-October-1-2016-to-September-3/shgm-yzbp

Dallas OpenData. 311 Service Requests from October 1, 2018 to Present. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/311-Service-Requests-October-1-2018-to-Present-/m36q-vtbr

Dallas OpenData. Dallas Animal Shelter Data. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/Dallas-Animal-Shelter-Data/7h2m-3um5

Dallas OpenData. FY17 Dallas Animal Medical Records. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/FY-2017-Dallas-Animal-Medical-Record/c24f-adak

Dallas OpenData. FY17 Dallas Animals Field Data. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/FY-2017-Dallas-Animals-Field-Data/xy7j-y59g

Dallas OpenData. FY18 Dallas Animal Medical Records. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/FY-2018-Dallas-Animal-Medical-Record/utjn-mhn6

Dallas OpenData. FY18 Dallas Animals Field Data. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/FY-2018-Dallas-Animals-Field-Data/rjkn-q4kc

Dallas OpenData. FY19 Dallas Animal Medical Records. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/FY-2019-Dallas-Animal-Medical-Record/6ucs-kgep

Dallas OpenData. FY19 Dallas Animals Field Data. DallasOpenData.com. Retrieved from: https://www.dallasopendata.com/City-Services/FY2019-Dallas-Animals-Field-Data/rgsr-utra

Microsoft. Introduction to Azure Blob Storage. Docs.Microsoft.com. Retrieved from: https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction

Microsoft. What is Azure SQL Data Warehouse? Docs.Microsoft.com. Retrieved from: https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is