**Write-up:**

Who let the Dogs Out?

**Project Summary:**

The aim of our project is to determine what dog breeds are the most convenient pets. We’ll examine relationships between genetic / behavioral traits related to noisiness, overall lifespan, and size.

**Step 1: Framing the Problem**

**Questions:**

1. Do smaller dogs bark more?
2. What are the life expectancies of different sized dogs?
3. Is there a correlation between a dog's height and weight?

**Step 2: Collecting the raw data to analyze**

**Data source**:

1. <https://api-ninjas.com/api/dogs>
2. https://tmfilho.github.io/akcdata

**Step 3: Processing the data**

**Cleaning Data**

1. Data from CSV was missing from API, run a try/except
2. Cleaned out duplicates using pandas
3. Grouped dogs by weight\*
4. API keys by every member of the group were generated to use in private. There are limitations on how much data can be pulled by the Dog API
5. Loop algorithm was developed to cycle through and retrieve all the dog data
6. Try-except code was used in order to filter our bad dog names or missing data
7. Dataframes were built to collect and store the data
8. Search for duplicates and removed them from the dataset
9. Created new calculated averaged fields to best represent the data
10. Created dog groupings to better analyze and summarize the data

Faceted Data by Size: Parameters set by American Kennel Club (AKC):: 4 sizes of dogs (Small, Medium, Large, X-Large).

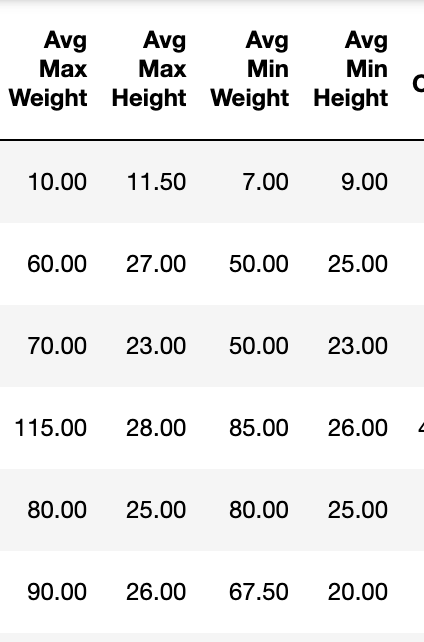
**Table 1: Size Facets**

| **Size** | **Weight** |
| --- | --- |
| **Small** | <= 22 |
| **Medium** | 22 < x < = 59 |
| **Large** | 59 < x < = 99 |
| **X-large** | 99 < x |

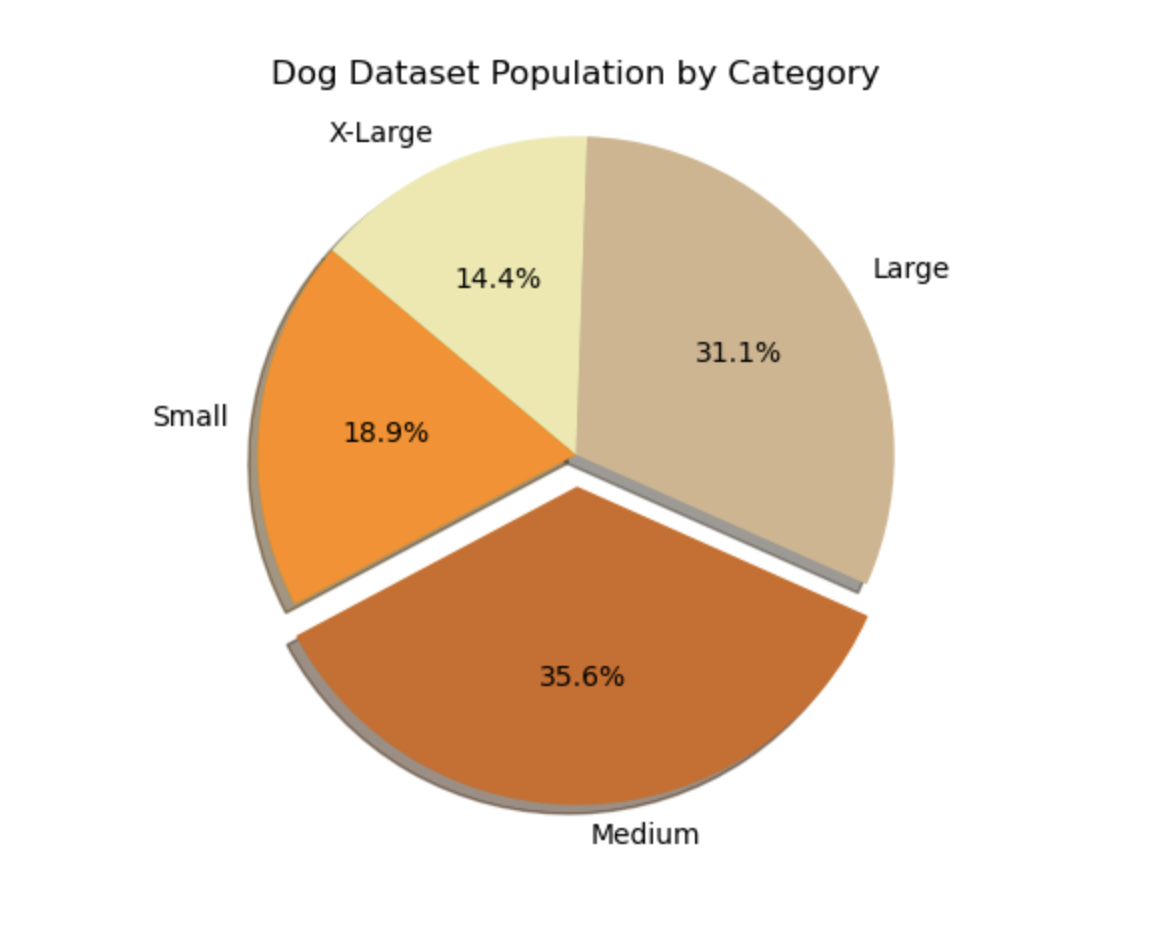
Gender Inefficiencies: Gender size differences made faceting original data inefficient (29 inefficiencies).

Solution: Average the weight of genders per breed to get congruent data.

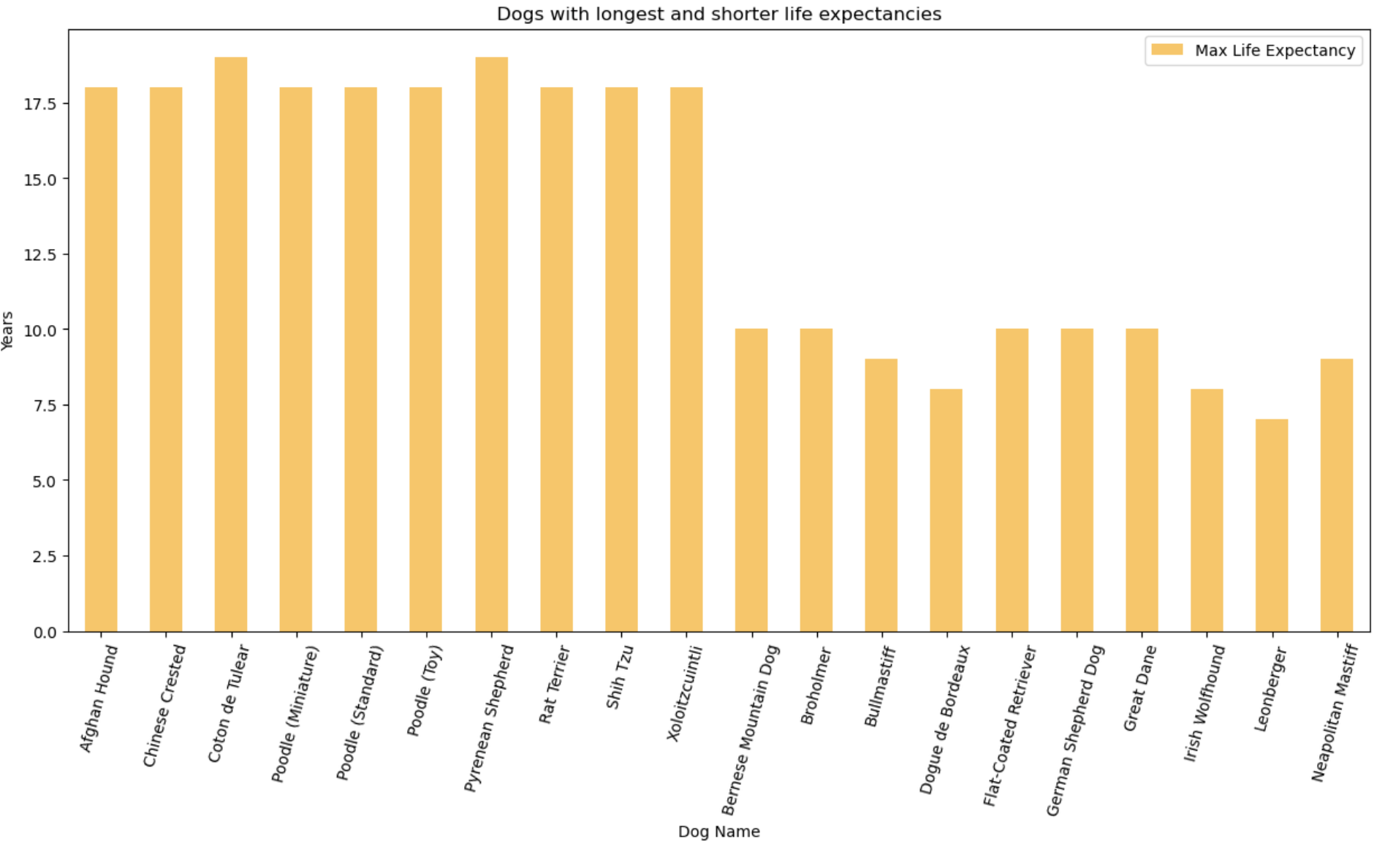
**Table 2: Average Genders:**



**Graph 1: Population of Dogs**

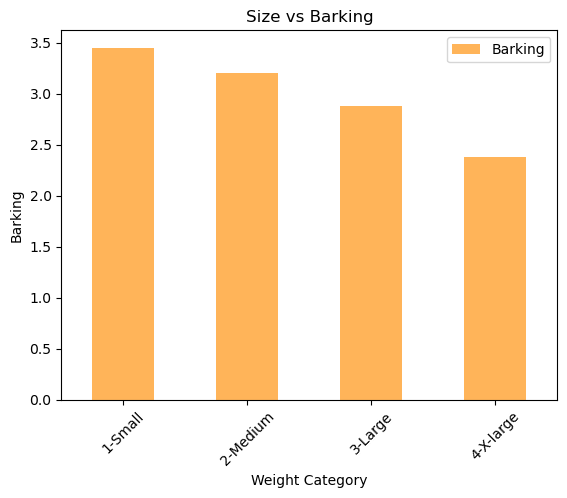
****

Population segmentation: Created populations of long-living / short-living dogs (10 top-tier and bottom-tier lifespans of dogs in the datatable.

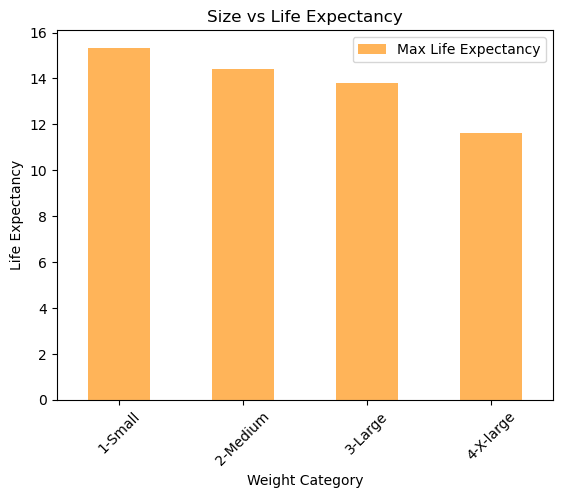
**Graph 2: Age vs Dog Name (Top 10 / Bottom 10)**

**Step 4: Performing In-depth analysis:**

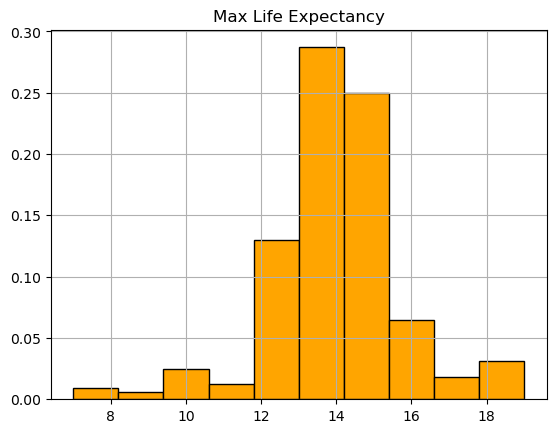
**Graph 3: Size vs Barking**



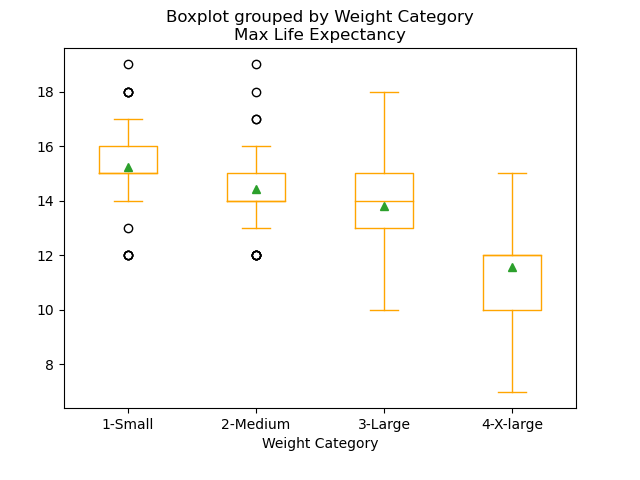
**Graph 4: Size vs Life Expectancy**

****

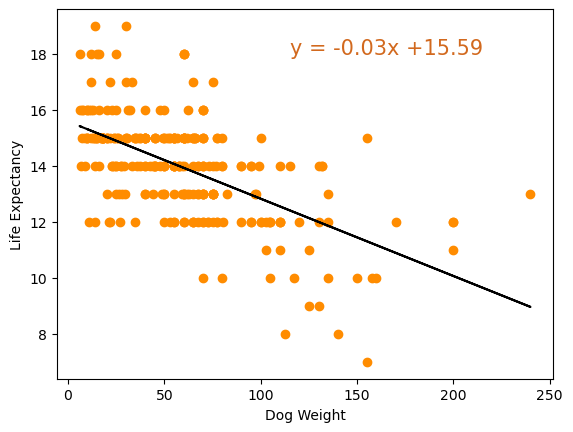
**Graph 5: Distribution of Lifespan**

****

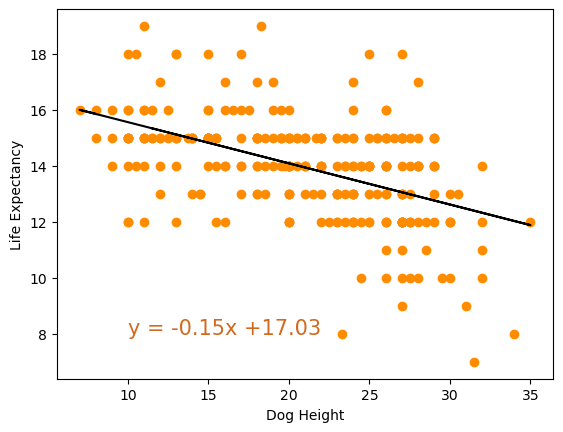
**Graph 6: Plot of Dogs & Life Expectancy**



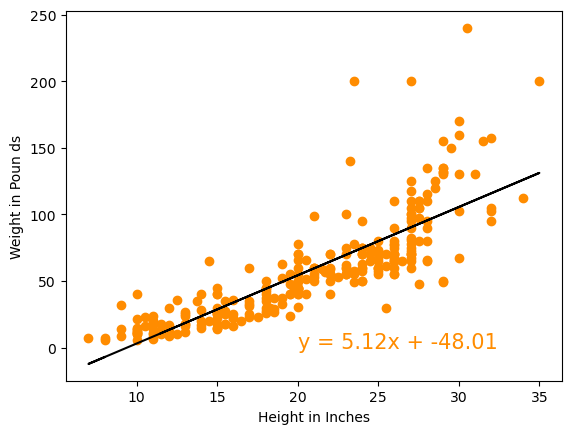
**Linear Regression 1: Life Expectancy (Max) and Weight (Avg. Max Weight)**



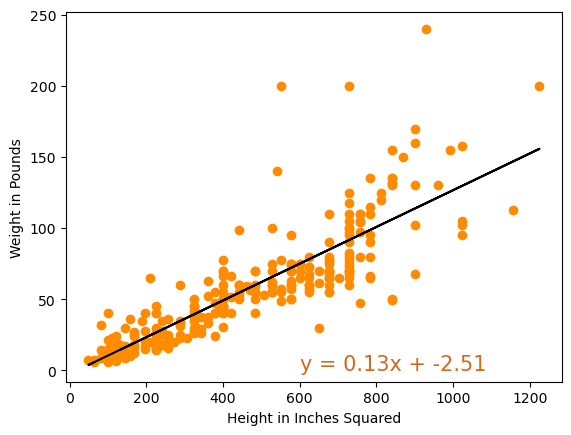
**Linear regression 2: Life Expectancy (Max) & Height (Avg. Max Height)**

****

**Linear Regression 3: Dog Height (Avg. Max Height) and Weight (Avg Max Weight)**



**Linear Regression 4: Dogs Height (Avg. Max height) and Weight (Avg. Max weight) (R^2)**

****

**Step 5: Communicating results:**

Questions:

1. Do smaller dogs bark more?
2. There is evidence of less noise with larger bred dogs.(Graph 3)
3. What are the life expectancies of different sized dogs?
4. There is evidence to suggest the size of dogs has a direct impact on lifespan. (Graph 4, Graph 5, Graph 6, Linear Regression 1 & 2)
5. Is there a correlation between a dog's height and weight?
6. There is evidence of correlation to dogs height and weight Graph 2, Linear Regression 3 & 4)

**Graph 1: Population of Dogs**

The graph is a representation of dogs for our 4 facets.

**Graph 2: Age vs Dog Name (Top 10/ Bottom 10)**

Demonstrates the top tier and bottom tier of the lifespan of dogs in the datatable.

**Graph 3: Size vs Barking**

Conclusion: Based on our findings there is a relationship with as the size of dogs gets larger the noise level decreases.

**Graph 4: Size vs Life Expectancies**

Conclusion: Based on our findings there is a relationship between the size of a dog and its lifespan.

**Graph 5: Distribution of Lifespan Graph:**

This graph shows the distribution of the lifespan of the dogs.

**Graph 6: Plot of all Dogs & Life Expectancy**

Explains different sizes of dogs and the range of their lifespans. There are outliers and partial skews of the data in the small and medium dogs categories. We attribute this partly due to the larger population sizes of the animals in our data set.. We determined that due to little amounts of outliers that the skew is not significant enough to label our analysis insignificant.

**Linear Regression Graphs**

**Linear Regression 1:** Life expectancy(Max Life Expectancy) vs Dog Weight (Avg. Max Weight)

**Linear Equation:** Y = - 0.03x +15.59

**R- Value:** - 0.586

Conclusion:

There is sufficient evidence to conclude that there is a significant linear relationship between life expectancyand *weight. This graph shows there is a moderate strength of correlation fit for our model.*

**Linear Regression 2:** Life Expectancy (Max) vs Dog Height (Avg.Max Height)

**Linear equation:** y = - 0.15 x + 17.03

**R- value:** - 0.49

Conclusion:

There is sufficient evidence to conclude that there is a slight linear relationship between life expectancyand max *height.*The r-value tells at .49 that it is a weak / moderate fit for our model.

**Linear Regression 3:** Dogs Height (Avg. Max Height) vs Weight (Avg. Max Weight)

**Linear equation:** y = 5.12 x + -48.3

**R- value:** 0.666

Conclusion:

There is sufficient evidence to conclude that there is a significant linear relationship between max heightand max *weight.**This graph shows there is a moderate strength of correlation fit for our model.*

**Linear Regression 4:** Dog Height (Avg. Max Height)\*\*2 vs Dog Weight (Avg. Max Weight) (R^32)

**Linear Equation:** Y = 0.13x + -2.51

**R- Value:** 0.698

Conclusion:

There is sufficient evidence to conclude that there is a significant linear relationship between life expectancyand *weight. This graph shows there is a strong strength of correlation fit for our model.*