Communicates the Insights

I analyzed the data by coding and visualizing them. The questions and the found answers are as:

Questions for codding section:

1. What is the population of each dog "stage" (doggo,floofer,pupper,puppo) in the data?

   Answer: The pupper has the most counts.

2. How the overal number of "favorite_count"s changes year by year?

   Answer: The "favorite_count"s of dogs has been increasing by time.

3. Is there any relationship between number of posted pictures and average rating_numerator and average favorite_count?

   Answer: By increasing the number of posted figures, both rating_numerator , favorite_count has been increases.

Questions for visualization section:

1. How confidence of image prediction can change by average number of posted pictures?

   Answer: definition of 4 groups of confidences.

   - *low: [0, 0.4]*
   - *moderate_low: [0.4, 0.6]*
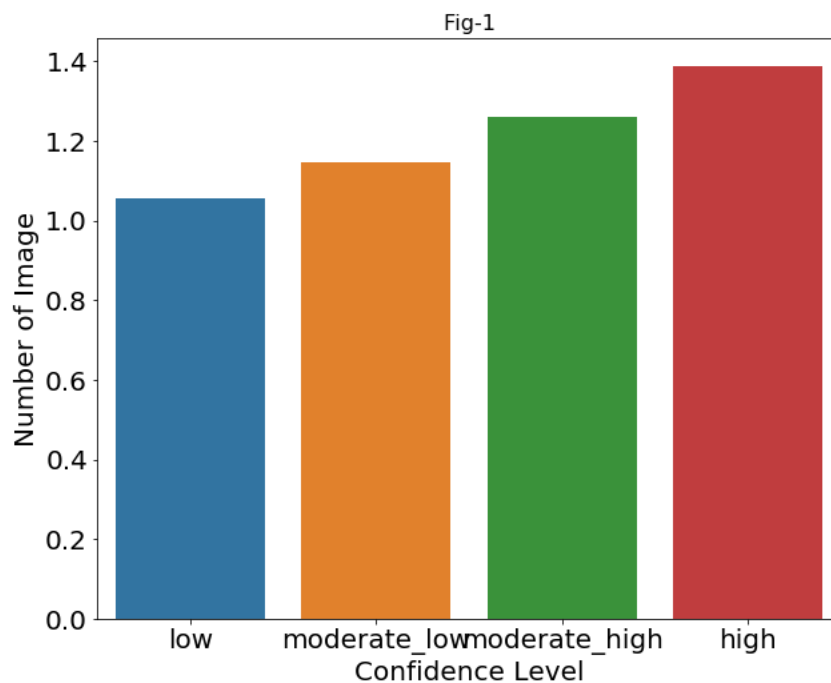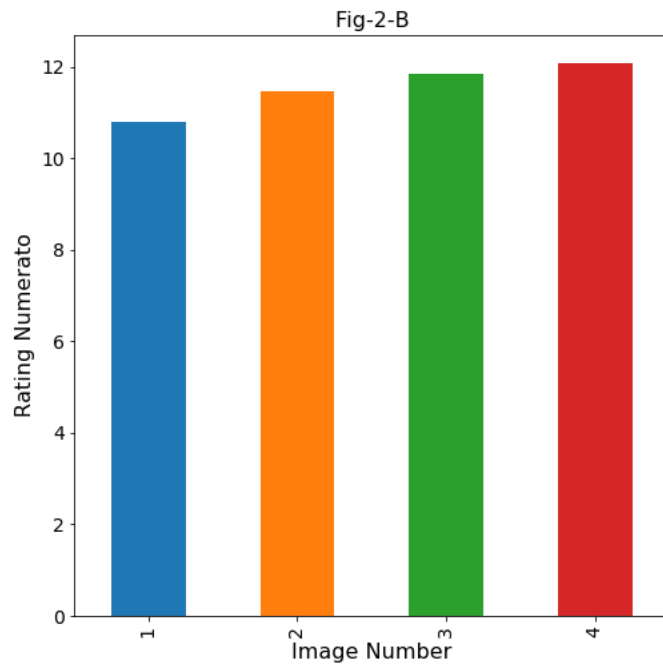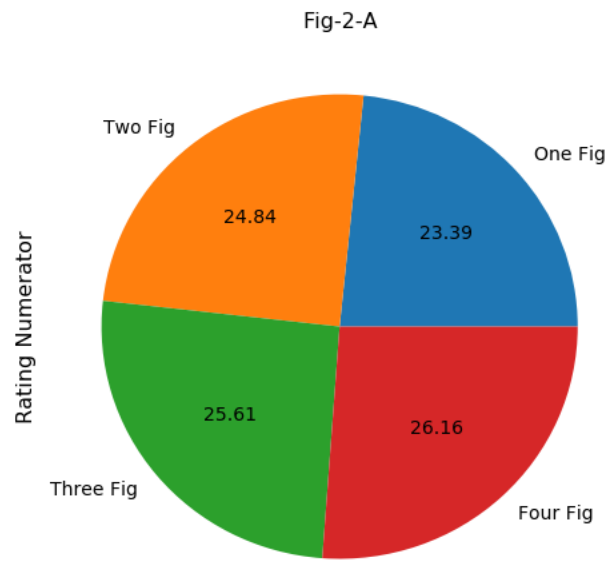   - *moderate_high: [0.6, 0.8]*
   - *high: [0.8, 1]*

Fig.1: Increasing the number of posted figures, the confidence of prediction increases from low to high.

- And 3: How the number of posted figures affect the rating_numerator and favorite_counts?
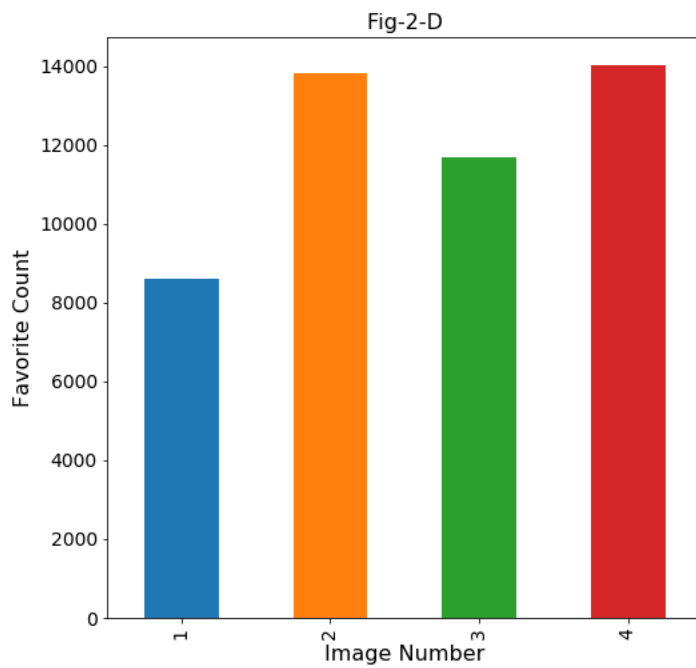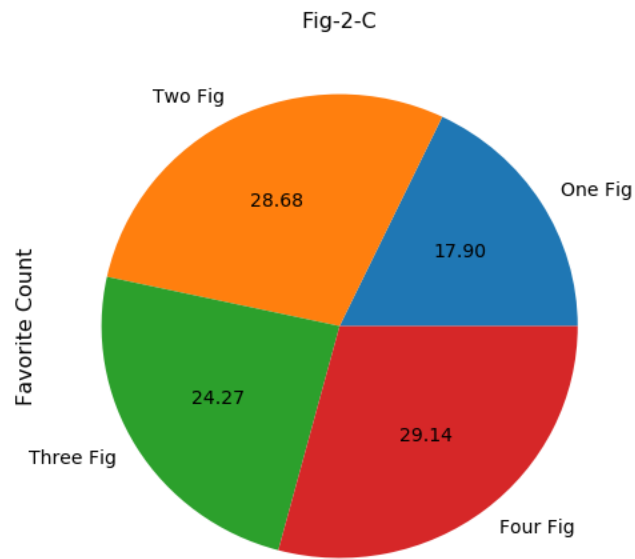
Fig-2-A

Rating Numerator

Two Fig — 24.84
One Fig — 23.39
Three Fig — 25.61
Four Fig — 26.16

Fig-2-B

Rating Numerato (y-axis)
Image Number (x-axis): 1, 2, 3, 4

Fig.2: By increasing the number of posted figures, both rating_numerator (A and B) and favorite_count has been increased (C and D).

Limitations: I do not have a good understanding about the collected data and the prediction models. This may bring many uncertainties. However, the project was a good practice for wrangling including reading, assessing, and cleaning data. Besides, I observed that, there is no information about the countries of people who voted for these dogs. For example, we do not know which dog is the most popular one in each country? how many people from each country are participating in this tweet page? How popular are these dogs among men and women? How popular are these dogs among different ages of voters?