[Seattle Pet Licenses | City of Seattle Open Data portal]

[Github Link]

☐ Data and Audience

Depending on the specifics of your project, you should outline the answers to these (and perhaps other) questions:

- What is the dataset you'll be working with? Please include background on who collected the data, where you accessed it, and any additional information we should know about how this data came to be.
  - We would like to present a set of information about the eligibility registration of pets in the Seattle area. This data table was found by Mengqi Shi in the City of Seattle's public database.
- Who is your target audience? Depending on the domain of your data, there may be a variety of audiences interested in using the dataset. You should home in on one of these audiences.
  - Our target audience is anyone who is interested in pets and considering having a pet. It is also fun for us to read all of this data, and we believe that those pet lovers can also take pleasure in discovering information about popular cat/dog breeds in Seattle and how pet-friendly each area is(depending on the quantity of each zip code). Moreover, it could provide some suggestions for people who are interested in having a pet in the future, for example, the name, species, and breed they should choose.

    If someone finds a lost pet on the street, they could easily find the pet's license on it.

- What does your audience want to learn from your data? Please list out at least 3 specific questions that your project will answer for your audience.

-What are the most popular cat/dog breeds in Seattle?

-What is the distribution pattern of pets in Seattle?

-Which area is the most pet friendly?

-Which area is the most regulated for pets?

-Which animal species have both primary and secondary breeds?

-What are the top 10 popular names for pets?

- Are there any correlation between the animal breed and the location of the animal?

☐ Technical Description

This section of your proposal asks you to think through the specific analytical steps you'll need to complete throughout the project.

- How will you be reading in your data (i.e., are you using an API, or is it a static .csv/.json file)?
  - static.csv
- What types of data-wrangling (reshaping, reformatting, etc.) will you need to do to your data?
  - Data integration: we will combine multiple variable columns into one single data graph, for example merging the pet's name and breed based on the numbers.
  - Data reformatting: we might convert the animal's breed to a numeric formation to perform the calculation on it.
  - Data reshaping: we may need to transform rows into columns, and vice versa. Also, we need to aggregate the data to summarize the data using statistical functions. We will be adding new variables to facilitate analysis.

- What (major/new) libraries will be using in this project (no need to list common libraries that are used in many projects such as dplyr)

library(gglot2)

library(dplyr)

library(Shiny)

R markdown

- What questions, if any, will you be answering with statistical analysis/machine learning?

    - Information on the extent to which pets are regulated in each area, as well as information on popular pets, can also be provided to Animal Control.
    - The relationship between breeds and zip codes.
- What major challenges do you anticipate?
    - What information do we decide to display and analyze, and how to aesthetically present them through the UI.

☑ ~~Project Set-up (Done)~~

In addition to outlining the steps you'll be taking to complete the project, you'll also be setting up your project on GitHub by doing the following:

- Create a public repository on GitHub in which you'll be completing your final projects. All group members should be added to the repo so they can push directly to it
- Create 5 GitHub Issues as your first set of steps to take in the project. You should assign these to individual group members to complete
- Optional: Create a slack/discord/… channel for your group to use for communication.