SI 507 Final Project Proposal [Revised]

Red=changes

*Overview*

* I will scrape the American Kennel Club website.
* ~~The reason for doing this is to take the text from the blurbs and count the frequency of words used in the descriptions for the animal.~~
* ~~Using the NLTK module, the most frequent words will be returned in a dictionary.~~
* ~~I will limit the amount of words to 10 and these words will be added into a database.~~
* From the website I will collect:
  + Table 1: BREEDS
    - Primary key == breed rank
    - Dog name 🡪 1 entry
    - Height (males and females) 🡪 2 entries
    - Weight 🡪 1 entry
    - Life expectancy 🡪 1 entry
    - Group 🡪 1 entry, foreign key
  + Table 2: GROUP
    - Primary key 🡪 ENUMERATE (digit 1-7)
    - Group Name 🡪 1 entry
* Every dog breed in the AKC website will be included as an entry into the database (Table 1).
* ~~I will also take the words from the temperament heading and sort the dogs into 4 different temperament categories.~~
* The dog’s popularity rankings will also be scraped and included into the database (Table 1, Primary Key) and finally the “Group” section of the website will also be scraped and added into the database (“Group” =Terrier, Hound, Sporting, etc)[Table 2].

*Data Sources*

* Crawling [and scraping] multiple pages in a site you haven’t used before: <http://www.akc.org/dog-breeds/>
  + **Challenge Score = 8**

*Presentation Options*

* I will make an interactive interface where the person can enter a dog breed (they will be given the list of possible dog breed options).
  + ~~Based on the User’s input of the dog breed they will be shown a graph of the most common words that are used to describe the dog.~~
  + ~~Another option the interface has is to show a graph that displays the number of dogs in each different temperament category.~~
  + ~~The third display option will show a graph of the top 5 most popular dog breeds.~~
  + Enter a dog breed:
    - Get a list of stats associated with that breed
  + Enter value: height females/ or height males
    - Get a chart that returns a graph of the average female or male heights (grouped by “Group”)
  + Enter value: weight
    - Get a chart that returns a graph of the average weights (group dogs by “Group”)
  + Enter value: life expectancy
    - Get a chart that returns a graph of average life expectancy (group dogs by “Group”)
  + Enter value: distribution
    - The final graph option will show a pie chart where each slice of the pie represents the number of dog breeds based on their “Group”.

*Presentation Tools*

* Plotly