Project Proposal: Project 4-Group 3 (choice 1)

Topic:

- Image recognition for common house pets
- User will be able to upload a picture of any type of pet that they want to identify
- The model will predict whether if it is one of the following pet type
- List of categories
    - Dog x 500
    - Cat x 500
    - Bird x 500
    - Fish x 500
    - Hamsters x 500
    - Reptile (Bearded Dragon) x 500

Reason for choosing topic/purpose of this project:

- We are interested in learning more about the CNN learning model.

Datasets:

- Scrape images off internet (At least 500+ for each type of pet)
- If 500 is not enough for training we will increase this amount in order make the model to work
- Look for data sets off free image libraries.
- Google Images/Bing Images/Imgur

Task breakdown:

EDA:

- See Datasets Above

ETL:

- Get images. (500 pixels max, .jpeg, 24kb max)
- Clean up images to correct format / size / quality / correct labelling and sorting
- Categorise images

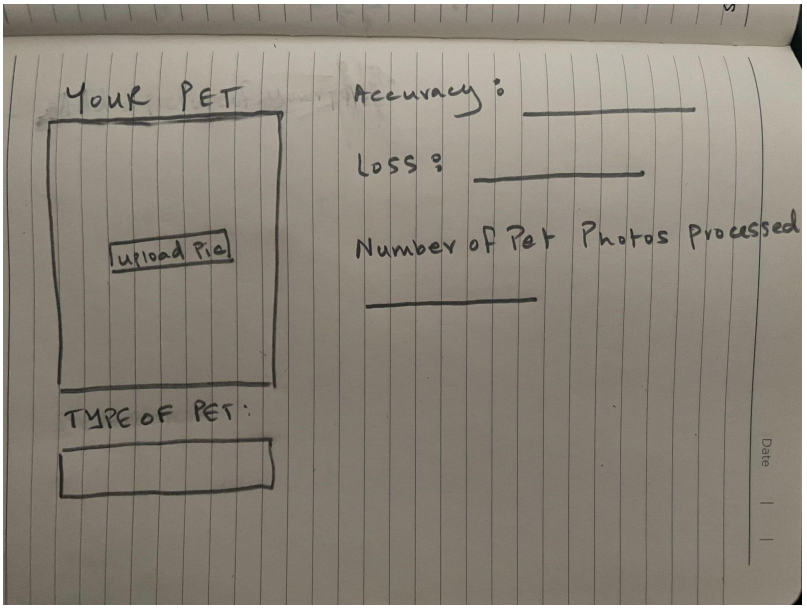
ML:

- Creation of model to process SQL data

Frontend:

- Live website query to ML predictions: User uploads image or inputs some basic information, prediction is output

Napkin drawing:



Bonus stretch goal:

- What colour is your pet?
- Add more pet types.
- Add in Red Herring items and label them as “Other”

Project Proposal: Project 4-Group 3 (choice2)

Topic:

- Using School Data to predict if the average SAT or other test scores of highschools is above average

Reason for choosing topic/purpose of this project:

- Ability to learn and utilise a correlation matrix model (i.e. Seaborn)
- A correlation matrix and machine learning model can be interesting in predicting school test performance and metrics that are responsible for affecting test scores

Datasets:

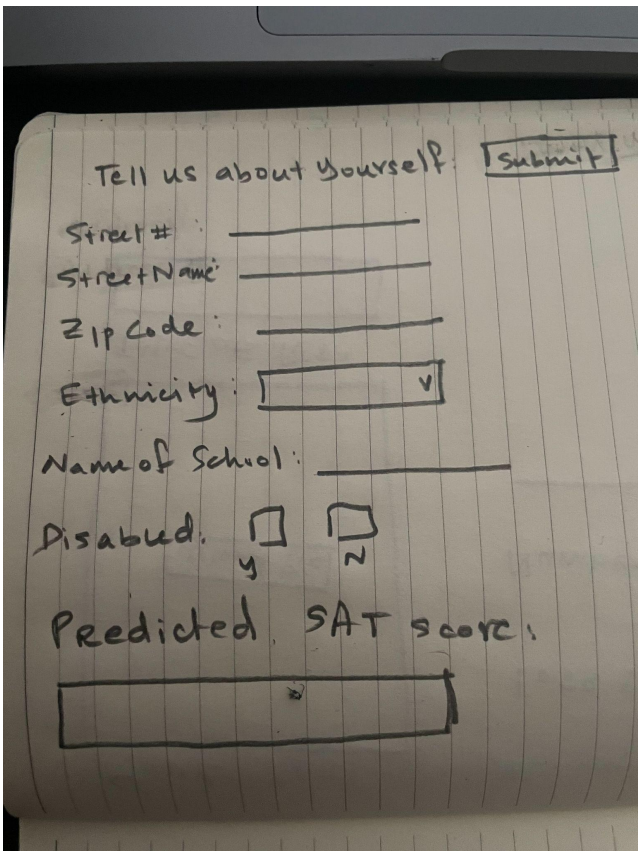
- https://data.cityofnewyork.us/Education/2013-2014-School-Quality-Reports-Results-for-High-/bm9v-cvch
- https://data.cityofnewyork.us/Education/2013-2015-New-York-State-Mathematics-Exam-by-Schoo/gcvr-n8qw (alternative)

Task breakdown:

- EDA: Correlation Matrix to determine if dataset can be used for ML
- ETL: SQL table with data that has been cleaned based on EDA
- ML: Creation of model to process SQL data

- Frontend: Live website query to ML predictions: User inputs some basic information, prediction is output.

Napkin drawing:



Bonus:
Dataset and Correlation Matrix dependent.