**DAT 205 Meeting Minutes Feb 28, 2021 at 1300hrs ET**

**Topic:** Team meeting – Project proposal and dataset selection

**Participants:**

Bhavika Patil

Cindy Guo

Dennis Hung

**Discussion**

* Review dataset options but additional research is required before finalizing what dataset to use for the project. Group wants to investigate further due to issues with some of the APIs or complexity of some dataset/topics
  + Default option (if nothing else) NYC Taxis
    - Data
      * <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>
      * 20 features
      * Approximately 72GB of data from 2009 to 2020 June
      * New York City taxis only
    - Suggested problem: Predict trends and how it correlates with the economy or demographics.
  + Option: NBA API
    - <https://www.playingnumbers.com/2019/12/how-to-get-nba-data-using-the-nba_api-python-module-beginner/>
    - Issues: API yet to work for group
    - Suggested problem: Determine improvements to team rosters to produce more wins (ie Moneyball angle) and maybe correlate to team budgets
      * Player salaries <https://hoopshype.com/salaries/players/>
  + Option NASA
    - python-nasa-api Simple NASA Api Wrapper written in Python!pypi.org
    - <https://pypi.org/project/python-nasa-api/>
    - Lots of data and narrowing down problem/requirements with assumptions will be important
    - Suggest problem: Find new [star/planet/asteroid/UFO]
  + Option Twitter
    - Twitter Data Mining: A Guide to Big Data Analytics Using Python. We will take a look at how to use Tweepy to gather data from Twitter using Python, followed by various usage examples.www.toptal.com
    - <https://www.toptal.com/python/twitter-data-mining-using-python>
    - Data pulls are limited (ie 500K records per month)
    - Suggested problem: find hash tags that could go viral
  + Other options?

Timelines

|  |  |
| --- | --- |
| **Date** | **Milestone** |
| Mon, Mar 1 | Finalized topic selection |
| Tue, Mar 2 | Finalized Initial proposal |
| Thu, Mar 4 | Submit initial proposal |

**Action Item**

Group:

* Investigate further into other possible datasets and finalize selection on a call Mar 1st 9:30pm

Bhavika

* Research further into API’s for NBA, NASA, Twitter, and others
* Draft proposal ideals for what is of interest

Cindy

* Research NYC Taxi dataset
* Start a draft proposal or ideas for this angle
* Test NBA API

Dennis

* Test NBA API
* Start a draft proposal NYC Taxi and/or ideas for NBA data
* Issue draft project plan for group (EoD Feb 28)