Hank Zhang

**Project Idea #1:**

* **Topic (Title):** Analysis of Billboard Hot 100 Music Trends
* **High level description of project:**  
  This dataset contains audio features of the Billboard Top 100 songs. There are 28090/317795 entries in this dataset. The dataset is mostly numerical data of various ratings. The columns we are analyzing are mainly the features of the songs themselves. Some features of the dataset include genre, danceability, energy, speechiness, and acousticness. The objective of this project is to analyze the trends in the top 100 songs and determine the most common traits that compose a popular song.
* **Data Source(s):** <https://data.world/kcmillersean/billboard-hot-100-1958-2017/workspace/file?filename=Hot+100+Audio+Features.xlsx>
* *Optional* **Potential Future Employer:** Music streaming companies (Spotify, YouTube, Apple, Google, Amazon, Vevo, Pandora, Microsoft), music production companies (Warner Music Group, SM Entertainment, YG Entertainment, JYP Entertainment)

**Project Idea #2:**

* **Topic (Title):** Analysis of Daily Traffic
* **High level description of project:**  
  This dataset contains traffic count data for various counties in New York over several years. There are 21994 entries in this dataset. The dataset is mostly text data of road names and directions grouped by county, detailing the amount of traffic that has passed within that year. The columns we are analyzing are mainly average annual daily traffic (AADT). Some features of the dataset include the start and end locations and direction of the road. The objective of this project is to analyze traffic trends and see some characteristics of what causes traffic and possibly how to mitigate or prevent it.
* **Data Source(s):** <https://data.world/buffalony/y93c-u65y/workspace/file?filename=annual-average-daily-traffic-volume-counts-1.csv>
* *Optional* **Potential Future Employer:** Government, Uber/Lyft, transit companies

**Project Idea #3:**

* **Topic (Title):** Analysis of Characteristics of Bike Sharing Data
* **High level description of project:**  
  This dataset contains the number of registered bikes by hour over a period of 2 years. There are 17380 entries in this dataset. The dataset is numerical data of various measures. The columns we are analyzing are mainly count of bikes used. Some features of the dataset that may affect whether a given user rents a bike are weekday, whether it’s a working day, temperature, humidity, wind speed, and registration status of the user. The objective of this project is to analyze the characteristics of renting bikes and see how the various day or weather conditions affect the number of bikes rented.
* **Data Source(s):** <https://archive.ics.uci.edu/ml/datasets/bike%2Bsharing%2Bdataset>
* *Optional* **Potential Future Employer:** Lime, Mobike