**Project Title**

“What makes a hit”

**Team Members**

Aditya Bhatnagar

Amisha Patel

Andrew C Stoops

Sean Wayland

**Project Description/Outline**

To collate data from various sources and examine if they can be correlated with a hit song.

**Research Questions to Answer**

**“**What song characteristics go into a hit song”

“What other environmental factors like employment and weather affect which songs are a hit”

**Datasets to Be Used**

Spotify API song characteristics

**Sources:**

1. Spotify API for developers
2. Kaggle
3. Google Dataset Search <https://datasetsearch.research.google.com/>
4. Billboard 100 charts
5. <https://openweathermap.org/history>
6. US Bureau of labor statistics

**Rough Breakdown of Tasks**

Research data points

Select timeframe and number of data points.

Select variables from data sets

Collect data

Convert data into useable objects

Graph data

Perform statistical analysis

Come to conclusion

Write up results

Present results

**Procedure:**

1. Finding Dataset:

Find a dataset off a standard Spotify playlist of Top 100 Pop songs (or any other playlist with at least over 100 songs so that we can have a shot at getting a normally distributed data. We can possibly achieve this by hitting Spotify’s API to provide a song characteristic for an entire playlist. The characteristics can include features like: ‘BPM’, ‘Danceability’, ‘Duration’ and so on.

1. Cleaning of dataset:

Remove any missing data fields, organize the data frame to clearly identify key characteristics within the song.

1. Select variables from data sets:

We will need to do a regression analysis in order to figure this out. After doing the regression analysis to obtain the value of R to and then we can eliminate the non-relevant variables.

1. Charting the Data:

Scatter Plot: Create scatter plots with regression line (for each/most relevant/all characs. together) to show the relevance of each character.

Bar Chart: To compare the ‘averages’ of each song charac. together from the top 100 song list.

Pie chart: Will need to investigate data set to see if it can be used effective (only effective when comparing 2-4 variables).

Line Chart: Would be very apt for showing some kind of time series analysis. I’m thinking release\_date vs hit\_count. Perhaps get an idea on what time of the year produced max number of hits.

1. Analyze data using charts, conclude and make recommendations.