What Makes a Song Popular on Spotify?

ISYE 412 Project Report

Group 12: Ashley Hellenbrand, Kylie Hellenbrand, Connor Hull, & Ben Theobald

Table of Contents

1. Introduction

Spotify is an extremely popular music streaming service that continues to grow. In this report, we will compare Spotify's Top 100 songs from 2018 using Tableau and analyze how different song attributes affect the ranking of these songs. We thought this was an interesting topic due to the large number of people that listen to Spotify. According to Spotify's official Q3 2018 report, it had 191 million monthly active users. This is a huge number of people that influence Spotify's top charts. The data set we used was the Top Spotify Tracks of 2018. This is an original data source and the audio features for each song were extracted using the Spotify Web API and the Spotify Python library. Credit goes to Spotify for calculating the audio feature values.

2. Objective

While analyzing the data set, some questions started to arise regarding the relationships between different attributes and the popularity of the songs. Some of the questions we want to answer are:

1. How much of an impact does an artist's name have on a song's rank?
2. What attributes make a song popular?
3. For less popular artists, what makes their songs reach the Top 100?

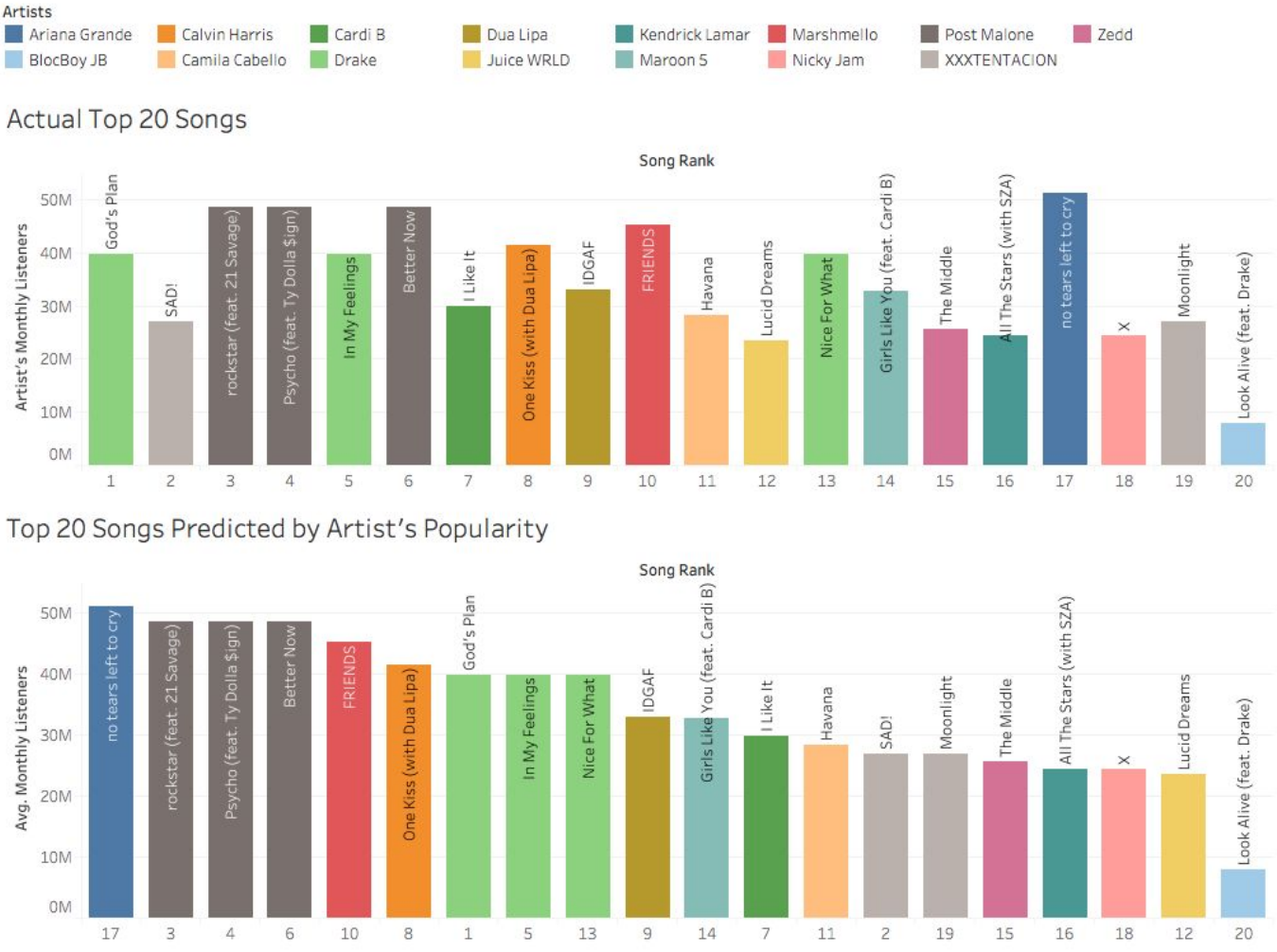
3. Approach

After determining the questions we wanted to answer, we had to come up with an approach to accurately represent the data and results. First, we manually recorded each artists' monthly listeners on Spotify and their ranking in the world for listeners because we needed a way to quantify artist popularity. Next, we normalized each attribute to fall within the range of zero to one to make it easier to draw conclusions on relationships between attributes. Data reduction was then used for some questions to allow for easier visualization. For example, we looked at only the top 20 songs to answer some questions so it would be easier to draw conclusions from the visualizations.

4. Visualizations

4.1 Artist Name and Its Influence on Song Rank

When listening to the top charts on the radio, genuine curiosity often arises. Are these songs popular because of the artist's popularity, or do they have very specific qualities that resonate well with audiences?



In the top half of this visualization, the Top 20 streamed songs on Spotify are displayed in order. The rankings are on the x-axis, and each color represents an artist. You can see that Drake has the top song, God's Plan, in addition to 2 other Top 20 songs. On the y-axis, we have the artist's monthly listeners in millions. This is the number of unique listeners that play their music within a rolling 28-day period. You can see that Ariana Grande, in blue, has the most at about 50 million, and the artist holding the 20th spot,

BlocBoy JB, has the least at around 8 million listeners. As explained earlier, this measure will be used to quantify artist popularity.

If we instead wanted to take these Top 20 songs and rank them based solely on the artist's monthly listeners, we get the bottom half of the visualization. The original song rank is represented under each bar in the graph. For the most part, the actual rankings are very comparable to the rankings based on the artist's monthly listeners. However, Ariana Grande has the most monthly listeners not only out of this Top 20, but out of every artist on Spotify in the world. Yet, she has the 17th ranked song. On the other hand, XXXTENTACION is ranked 14th in listeners, yet he has the 2nd most streamed song on Spotify for 2018. Why is this so?

The monthly listeners measure we created was as of April 19th, 2019. Given our limited access to Spotify's data as regular users, there was no way for us to calculate an average of this value for 2018. Ariana Grande released her hugely successful album Thank U, Next on February 8th, 2019. Her popularity has skyrocketed since then, explaining the disparity between her monthly listeners in 2019 and her top song ranking in 2018.

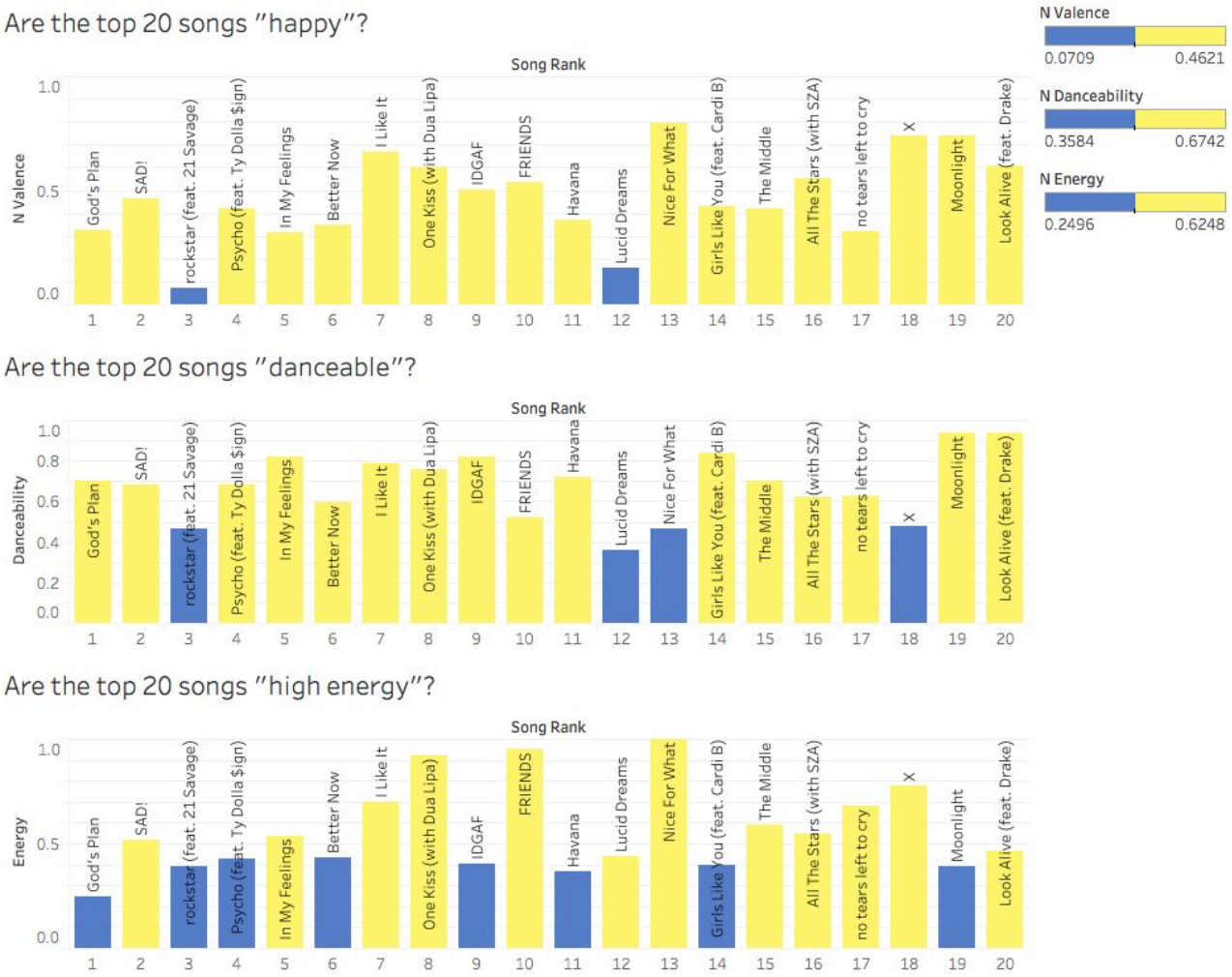
XXXTENTACION released his single Sad! on March 2nd, 2018, and was murdered a few months later on June 18th, 2018 at age 20. In the week following his death, Sad! moved from 52nd to 1st on the Billboard Hot 100, making him the first artist to top the Hot 100 after passing away since the Notorious B.I.G. in 1997. Other young artists like Mac Miller, Avicii, and Nipsey Hussle have had similar success with topping the charts following their deaths in the past year. From this, we can conclude that artists who pass away at a young age tend to increase in popularity following their passing.

4.2 Important Attributes of the Top 20 Songs

For our second question, we wanted to explore what specific attributes stand out in the most popular songs. Our dataset had 13 different audio features measured for each song, but we thought it would be best to focus on a few attributes that we found most interesting. We decided to compare danceability, energy, and valence among the top 20 songs. Danceability measures how suitable a track is for dancing based on tempo, rhythm stability, beat strength, and overall regularity. Energy measures the song's intensity and activity based on dynamic range, perceived loudness, onset rate, and general entropy. Finally, valence describes the song's overall "positiveness" (happy, cheerful, etc.).

After normalizing the data to fall between zero and one, we also determined the median value for each attribute. For example, we sorted the data from most danceable to

least danceable and then recorded the 50th ranked song's danceability. The 50th value for each attribute would be the cutoff point for marking a song as happy/sad, danceable/not danceable, and high/low energy. Through this methodology, songs were easily determined as better than the majority or worse than the majority. The results of the analysis can be shown in the visualization below with yellow meaning happy, danceable, and high energy and blue meaning sad, not danceable, and low energy.

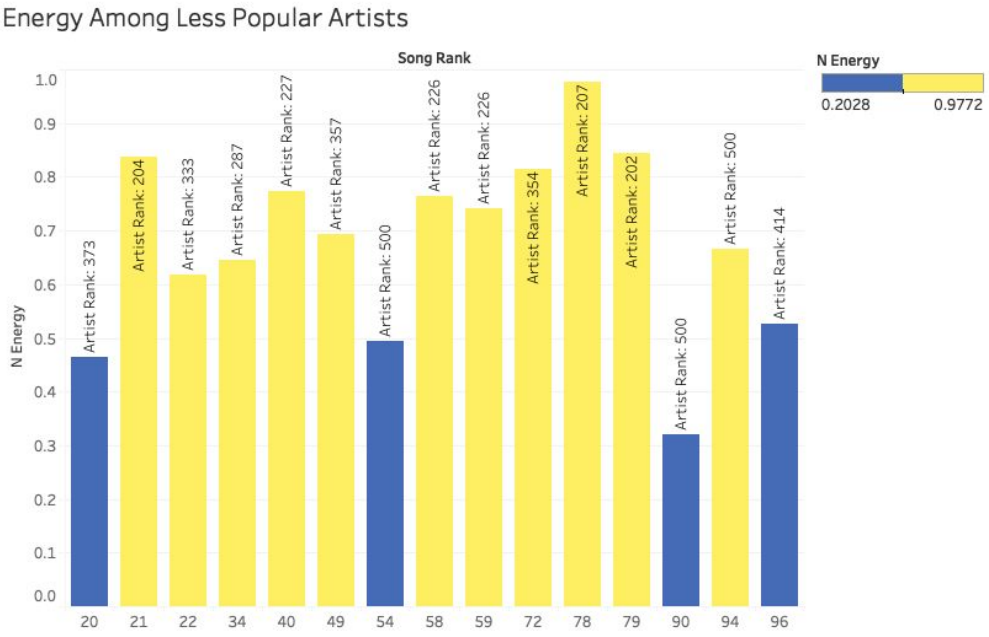


Looking at the top (happy) visualization, the vast majority of the top 20 songs were happy with the exception of two songs, rockstar and Lucid Dreams. The middle (danceable) visualization produced similar results with 16 of the top 20 songs marked as danceable. The bottom (high energy) visualization showed a different trend with only about half of the top 20 songs marked as high energy.

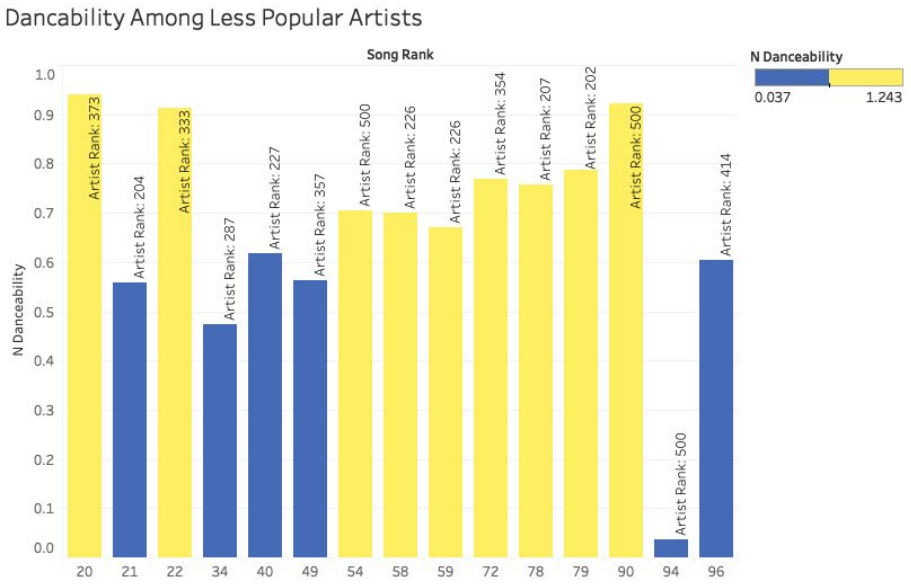
After analyzing these results, it can be concluded that the most popular songs tend to be happy and danceable. There was not a clear trend on whether the top 20 songs were high energy or not, so we concluded that energy doesn't seem to play a role in a song's popularity. Stacking all three visualizations on top of each other, we are also able to see that valence (happiness) and danceability might be correlated. With this being said, we think that a happy song might be more likely to be danceable, and vice versa.

4.3 Important Attributes for Less Popular Artists

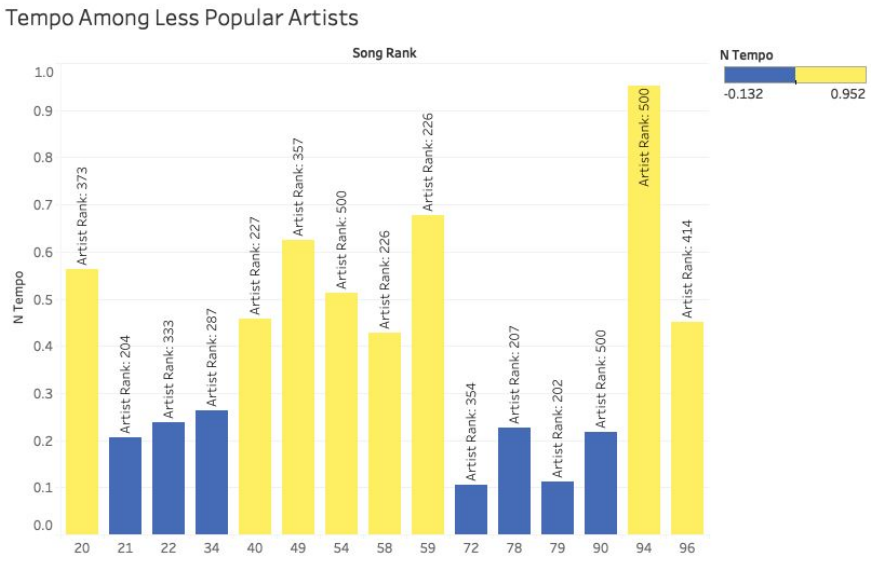
The third question we looked to answer is for less popular artists: what makes their songs reach the Top 100? We have taken a closer look into trends that occur in the top 20 hits, which are mostly comprised of songs by relatively well-known artists. We wanted to see if similar trends happen among artists who might not have as high name recognition or overall popularity. For these visualizations, we have filtered only artists with a world rank greater than 200. There are 15 songs with artists who fit this filter. On the x-axis, the artists are ordered by rank within the top 100, and the y-axis shows their energy values. 11 of the 15 songs have energy values higher than the average for the top 100 as a whole (0.59). This shows that these artists are likely to have more success if they produce higher energy songs.



Similarly, the same visualization was put together using the danceability attribute instead of energy. Here we notice the same trend with the majority of the songs having a higher danceability value than the average of 0.64, as seen on the next page.



Next, we wanted to look at tempo because we had assumed that high energy, danceable songs would have a relatively fast tempo. However, this trend did not transition because it is about a 50/50 split above and below the average tempo of 0.41.



Given these visualizations, we can conclude that lesser known artists are more likely to create a hit song if they incorporate elements with high energy and danceability, yet tempo seems to have no impact on popularity. This could be because these attributes are

more likely to attract attention from first-time listeners. Catchy, upbeat songs could attract popularity if the artist's name alone does not.

5. Conclusion

Throughout the data analysis, there were many interesting results. Overall the results were pretty significant. We determined that there is a high correlation between artist popularity and their songs making it in the top 100. Also, songs by artists who passed away at a young age tend to increase in popularity and streams immensely. Some of these observations may seem fairly obvious, so we looked into some lesser known relationships. The most popular songs have high danceability and valence, but energy does not seem to influence popularity. Also, danceability and energy help make lesser known artists' songs popular, but tempo has little impact.

6. Sources

"Stats – FAQ – Spotify for Artists." *Stats – FAQ – Spotify for Artists*, artists.spotify.com/faq/stats#can-i-compare-my-stats-with-other-artists-stats.

Tamer, Nadin. "Top Spotify Tracks of 2018." *Kaggle*, 7 Feb. 2019, www.kaggle.com/nadintamer/top-spotify-tracks-of-2018.

"Thank U, Next." *Wikipedia*, 25 Apr. 2019, https://en.wikipedia.org/wiki/Thank_U,_Next.

"XXXTentacion." *Wikipedia*, 25 Apr. 2019, https://en.wikipedia.org/wiki/XXXTentacion.

Fundamentals of industrial data analytics

Spring 2019

Instructor: Kaibo Liu

Team Project

For the project, you need to work in a team of 3 or 4 students. Each group only needs to submit one final presentation file and one final report in the end of the semester. This project is to help you to better practice the Tableau software learned in class, and also learn how to formulate and solve a real data analytics problem starting from messy raw data. Please start planning your project as early as possible and be aware of the following deadlines:

1. The project proposal (10 points) is due on **3/25/2019 (Monday)**. The purpose of the proposal is to get you started. It shall be **one page** (or possibly two pages if necessary). You will need to provide the following information:

(a) Your names in the group

(b) The dataset that you decide to use

(c) The dataset description

(d) Three to five scientific research questions that you may want to address

You will receive full credit on the project proposal if you provide all the above information. Each group please upload the project proposal to the designated space in the assigned team folder as a Word or pdf file: ISyE412_Group number_Project Proposal.

2. **The final presentation file** (45 points) is due **on 5/3/2019 (Friday)**. Each presentation is around 15 minutes (= 12 + 3 QA), and presentations will take place during the four lectures (4/17-4/29).

The contents of the presentation must include:

- Introduction: background of the dataset
- Objective: why do you analyze this dataset and what kind of questions that you would like to answer from the dataset. Please be creative!
- Your approaches to answer the identified questions
- Data visualization results and interpretations

- Prepare a dashboard: this should be the take-home message that you want to send to the audience (Imagine you are presenting the result in front of your boss)
- Summary and comments

All group members need to present at least one content part during presentation. All students are required to attend those presentation classes and a peer evaluation score sheet will be provided then. Each student will give a score to all the other groups' presentation but not including his/her own. Each student should complete this peer evaluation score sheet and give it to the instructor at the end of each presentation class. Each team's project score will be finally determined based on the average of the given scores from other teams.

3. **A final summary report of your class project** (45 points) is due **on 5/3/2018 (Friday)**. The excellent project work and report will be selected and published in the online IERA platform. One advantage for posting your project online is that you can easily share your fantastic work with others via social networks, which can help you better disseminate your knowledge and become more competitive in the job hunting.

When preparing the final report, please draft the materials in word file and follow the general guideline here: http://iera.name/guidelines-for-writing-articles/. Here is an example report based on a course project in previous year: http://iera.name/data-visualization-worldwide-small-business-analysis/.

Your group presentation file and report must be uploaded to the designated space in your assigned team folder with name: ISyE412_Group number_Project presentation, ISyE412_Group number_Project report, etc. (Including the ppt, the report, the tableau file, and the dataset or a statement with link to download the data) before the deadline.

Grading: This project is worth 30% of your final grade, and all team members will receive the same grade on the project. Your grade on the project presentation and report will depend on

"Problem statement (15%) _____

Correctness of the approach (20%) _____

Creativity and significance of the result (20%) _____

Logic and organization of the presentation/report (15%) _____

Quantity of the work (15%) _____

Visualization and clarity (15%) _____"

Possible Topics of Your Project

You are free to use any datasets that you are interested in. Here are some examples of online datasets that you could use for this project:

1. OlympicAthletes.xlsx (source: http://www.sports-reference.com/olympics): This sample data set has medals and athletes from each Olympic Games since the 2000 Games in Sydney, Australia.

2. World Bank Indicators.xlsx (source http://data.worldbank.org/indicator): This sample data set has 18 indicators from the World Bank on life, health and business between 2000 and 2010.

3. Bird Strikes.xlsx (source: http://wildlife-mitigation.tc.faa.gov/wildlife/default.aspx): This data set example has reported records of US flights between 2000-2011 that have been struck by a bird while in flight.

4. DataDownload_USDA.xlsx (source: http://ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads.aspx#.VCoMGBa-Y5C): United States Department of Agriculture---Food Environment data file and documentation

There are three data source initiatives:

1. World Bank's Open Data initiative (source: http://datacatalog.worldbank.org/): The data catalog is a listing of available World Bank datasets, including databases, pre-formatted tables, reports, and other resources.

2. NYC open data initiative: (source: https://data.cityofnewyork.us/): NYC Open Data makes the wealth of public data generated by various New York City agencies and other City organizations available for public use.

3. Seattle data initiative (source: https://data.seattle.gov/browse?sortBy=most_accessed&sortPeriod=week).

4. UCI Machine Learning: http://archive.ics.uci.edu/ml/index.php: This is an incredible collection of over 350 different datasets specifically curated for practicing machine learning. You can search by task (i.e. regression, classification, or clustering), industry, dataset size, and more.

5. Kaggle: https://www.kaggle.com/datasets: is most famous for hosting data science competitions, but the site also houses over 180 community datasets for fun topics ranging from Pokemon data to European Soccer matches.

6. Data: https://www.data.gov/: If you're looking for social science or government-related datasets, look no further than Data.gov, a collection of the U.S. government's open data. You can search over 190,000 datasets.

How to be creative?

There are many ways to be creative? Such as:

1. Answer a question which is not realized as the main initiative when collecting the dataset

2. Achieve a surprising conclusion that contradicts to our intuition

3. Combine multiple data sources to release new information

4. Use tableau techniques that are not covered in class

…

Please feel free to discuss your project with the instructor if you have any question.