**Topic Idea 1: Music Industry Dashboard**

1. **Problem Statement**
   1. **Describe the problem you would like to tackle.**

There is no standard way to rank and identify trending genres, albums, and artists. The availability of vast amounts of user data on music-sharing platforms is increasing competitiveness between music industry companies, especially when signing artists and promoting appropriate content.

* 1. **What is the topic of your project?**

Retrieve data from Spotify API, other music related-datasets and possibly Youtube to create a searchable music data dashboard. The dashboard will include general information about the artist/genre/album/song and an analysis based on KPIs, like # of listens, # of favorites, common comments, and # of follows over time (using visuals and text).

* 1. **What do you want to learn about it?**
  + Trends in the music industry
  + Whether genre popularity is cyclical
  + Rising artists to listen to
  + Emotional connections to music and artists
  + How music recommendation algorithms can be optimized using our analysis
  + How the data available can help music companies get ahead and make good business decisions
  + (Kristie runs a music magazine on campus and is responsible for quoting singers to bring to Northeastern.) Kinds of artists would Northeastern students want to see on campus
  1. **What are some keywords for your topic to use in dataset search?**
  + Music
  + Entertainment
  + Songs
  + Artists
  + Concerts
  + Spotify
  + Soundcloud
  + Youtube
  + Singers
  + Albums

### **Significance of the Problem**

* 1. **Why is it important to tackle this problem in your project?**
  + Streaming services have increased music companies’ pressure to utilize big data/data science to analyze user behavior to adapt strategies.
  + Music is a huge industry. According to Forbes, the global music industry sales rose by 10% in 2018 to $19 billion. There is a lot of money to be made through satisfying market demands.<https://www.forbes.com/sites/hughmcintyre/2019/04/02/the-global-music-industry-hits-19-billion-in-sales-in-2018-jumping-by-almost-10/#174acd7a18a9>
  + Music is important to our society. It shapes our emotions, perception of life, and culture.
  1. **In what ways could the insights from this project be useful?**
  + Streaming companies need to recommend appropriate songs to ensure user retention and satisfaction. They can adapt their algorithm based on our findings on genre, album, song, artist, etc. trends.
  + There is a big gap between music that is consumed vs. created. This dashboard can also allow budding artists to understand the market demands and appropriate channels to promote.
  + Many independent artists are getting big through social media, like SoundCloud or Youtube. Labels need to offer deals to artists with significant social capital or potential for traction. They can use our dashboard to scout.

### **Potential Datasets**

* 1. **Look into some potential datasets using the online resources provided on Blackboard or other data sources you find online.**
  2. **Look through the datasets to identify any potential matches.**
  3. **The dataset should include at least 5-10 columns with at least 100-200 rows.**
  4. **Provide links to potential datasets here.**
  5. **If you are planning to scrape your own data from online resources, describe your data and online resources here. Provide links to the online pages/documents.**
  + Instructions to extract information from Spotify API:
    - <https://medium.com/@RareLoot/extracting-spotify-data-on-your-favourite-artist-via-python-d58bc92a4330>
    - <https://labs.spotify.com/2015/03/09/understanding-spotify-web-api/>
    - <https://tgel0.github.io/blog/spotify-data-project-part-1-from-data-retrieval-to-first-insights/>
    - See if we can pull from top 200 most streamed:<https://spotifycharts.com/regional>
      * Crawler:<https://github.com/edumucelli/spotify-worldwide-ranking>
    - Spotify’s Daily Song Ranking list:<https://www.kaggle.com/nnqkfdjq/spotifys-daily-song-ranking-music-released-date>
  + Million song dataset:
    - <https://www.rollingstone.com/music/music-news/welcome-to-rolling-stone-charts-853157/>
  + See if we can pull from Rolling Stone charts:
    - <https://www.rollingstone.com/music/music-news/welcome-to-rolling-stone-charts-853157/>
  + Exporting data from Youtube:
    - <https://support.google.com/youtube/answer/9088722?hl=en>
    - <https://developers.google.com/youtube/analytics/data_model>
    - <https://developers.google.com/youtube/reporting/v1/reports/>
    - <https://medium.com/greyatom/youtube-data-in-python-6147160c5833>
    - Gives us access to videos, channels, captions, comments, playlists
  + Soundcloud API
    - <https://blog.hellojs.org/june-25-2017-creating-a-searchable-music-site-using-the-soundcloud-api-1b6d1723610>
    - <https://developers.soundcloud.com/>
    - <https://www.programmableweb.com/api/soundcloud>
  + AMA Awards Tweets from 2018:
    - <https://www.kaggle.com/eliasdabbas/american_music_awards_tweets>
  + 18393 Pitchfork Reviews:
    - <https://www.kaggle.com/nolanbconaway/pitchfork-data>

## 

## **Topic Idea 2: Real Estate Sale Price Prediction**

1. **Problem Statement**
   1. **Describe the problem you would like to tackle.**

When real estate developers embark on a project, investment decisions are rarely based on a comprehensive analysis on past and present market statistics. In particular, most small development firms rely solely on the president’s personal intuition. This absence of data-driven decision-making exposes development firms and their projects to considerable financial risks.

* 1. **What is the topic of your project?**

A real estate investment calculator to predict the sale price of a development project based on historical sales data, regional trends, and the features of the underlying property and proposed structure. The calculator allows developers to input their projects specifications, output a predicted sales price, and explore the effects of what-if scenarios.

* 1. **What do you want to learn about it?**
  + Factors contributing to both successful and unsuccessful real estate investments
  + Importance weights of such factors on the financial outcome of a project
  + Impact of architectural and geospatial features on sales prices
  + Structural and architectural designs that optimize sales price
  + How investment decision-making can incorporate data driven models
  1. **What are some keywords for your topic to use in dataset search?**
  + Real estate
  + Development
  + Sales
  + Properties
  + Parcels

### **Significance of the Problem**

* 1. **Why is it important to tackle this problem in your project?**

The real estate industry is a massive economic component, contributing $1.15 trillion annually to US GDP. However, homeownership in the US has been in decline. This trend is primarily due to the high cost of homeownership proportional to other costs of living. The risks inherent to real estate development account for a significant portion of project costs. Impacts of such risks have been exacerbated by demographic change and the decline in homeownership, creating the start of a vicious cycle whereby home prices increase.

* 1. **In what ways could the insights from this project be useful?**
  + Identify risk factors in current project specifications
  + Identify risk mediation and cost reduction factors to update project specifications
  + Reduce home building costs associated with project risk
  + Help developers identify investment opportunities
  + Generate optimized project specifications

### **Potential Datasets**

* 1. **Look into some potential datasets using the online resources provided on Blackboard or other data sources you find online.**
  2. **Look through the datasets to identify any potential matches.**
  3. **The dataset should include at least 5-10 columns with at least 100-200 rows.**
  4. **Provide links to potential datasets here.**
  5. **If you are planning to scrape your own data from online resources, describe your data and online resources here. Provide links to the online pages/documents.**
  + Parcel Information
    - <https://data.boston.gov/dataset/parcels-2017-data-lite>
    - <https://data.boston.gov/dataset/boston-neighborhoods>
    - <https://data.boston.gov/dataset/approved-building-permits>
  + Parcel Assessments
    - <https://data.boston.gov/dataset/property-assessment>
  + Parcel Features
    - <https://data.boston.gov/dataset/fire-incident-reporting>
    - <https://data.boston.gov/dataset/trees>
    - <https://data.boston.gov/dataset/hydrography-polygon>
    - <https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system>
  + Parcels Sales
    - <http://boston.maps.arcgis.com/apps/webappviewer/index.html?id=e5ca1dbb90b845ffa997ef277ab134df>