

Final Project Ideas

Abdullahi Ahmed

Krish Pandya

Mark Feng

Steven Volovar

Abdullahi's Section:

1. <https://grouplens.org/datasets/movielens/latest/> is the source of the datasets.

One potential data set is movies_metadata.csv and contains information on 45,000 movies featured in the Full MovieLens dataset. Features include posters, backdrops, budget, revenue, release dates, languages, production.

2. My customers would be individuals who are looking for or information on movies that were released on or before 2017 to watch or otherwise. Movies can be found by searching for keywords or by sorting. Customers can compare revenues of movies among other features.
3. The proposed project makes it easy to find movies and information on movies that were produced on or before 2017 by searching for keywords. The project also provides 26 million ratings by 270,000 users on all the 45,000 movies in the dataset; this is helpful in utilizing reviews to select movies to watch or recommend to a friend or loved one. This project is needed as people increasingly watch movies for entertainment or information.
4. The proposed project is for customers searching for movies they are interested in to watch or recommend movies to their friends or family. This product could be useful as it utilizes a wide variety of movies to select from and provides reviews by over 270,000 to provide guidance on movie selection. It also provides more features such as comparing movies by revenue, search by keywords, cast and crew, among other search criteria. This product makes it easier to find any movie as it provides a relatively larger search parameters.

5. Major features

- Top movies based on ratings
- Search movies based on keywords, cast and crew, language,
- Compare movies based on revenue, ratings and reviews

Krish's Section:

1. Source: [Zillow Research](#). This dataset is a more in depth look into the popular real estate company Zillow. Zillow offers the data they use/the research they conduct; to the public and this link provides information into typical home values in each region in the US as well as rental rates. In addition to this, many other datasets are provided as well, but perhaps not directly correlating to the prices of real estate.
2. Customers in this case would be individuals looking to either purchase a rental property or house in the city they are moving to.
3. The problem this project solves is a lack of holistic view on the entire housing market and subsequently the provision of an easy to navigate, visual platform for customers. This project will simplify data that isn't necessarily important to customers when making decisions. On the other hand, it will also add other key factors that competitors do not provide when picking a house. A key example of this is crime rate in the neighborhood, another factor is quality of education. In essence, the project will tie together the best aspects from each individual concept into a unified platform.
4. The problem this project solves is a lack of holistic view on the entire housing market and subsequently the provision of an easy to navigate, visual platform for customers. In essence, the project will tie together the best aspects from each individual real estate site into a unified platform. This creates a go-to application for anyone looking to buy a new house in any area anywhere.
5. Major features list:
 - Housing prices
 - Crime rates in neighborhoods

- Average weather in area
- School quality in the area

Mark's Section:

- 1) Source: [Data.gov](https://data.gov) This dataset provides an enormous amount of detailed information regarding the coronavirus as well as how humans are affected by it. It details metrics like vaccinations, cases, deaths, and government responses. The data is incredibly up to date and new information is added daily. This allows us to compare different datasets to see if correlations can be found between the sets.
- 2) Users would be those who seek up-to-date data for specific areas of the world rather than just seeing broad figures like total cases.
- 3) The problem this project seeks to solve is the lack of detailed information that is easily accessible. Google has a decent amount of data but lacks detailed analysis. For example, comparatively sized states in the South (for example, Alabama) tend to have higher mortality as well as there being an significant uptick of cases very recently whereas in the North, cases are currently trending down.
- 4) This project can solve this problem by collecting and displaying metrics together so that they can be compared and analyzed. That way, users can have a more complete picture of what is happening in a certain region of the world.
- 5) Major features list:
 - Metric Graphs
 - Overlaid Graphs for Comparison

Steven's Section: