# **Final Project: Project ideas**

**CIS Vikings: David Hogg, Austin Lowery, Justin Melhorn, and Grace Jones**

In this activity each member of the team needs to do the following tasks

1. **Create a GitHub repo for your project. In that repo add a "Documentation" folder. In that folder create a document called "Project ideas"** **(2 points)  
   A.** https://github.com/jtmelhorn/ITSC3155-FinalProject.git
2. **Include the name of all of your team members in the document to be submitted** **(2 points)  
   A.** David Hogg, Austin Lowery, Justin Melhorn, and Grace Jones
3. **List the source of and describe at least 4 potential data sets** **(8 points)**
   1. [**https://www.kaggle.com/shivamb/netflix-shows**](https://www.kaggle.com/shivamb/netflix-shows)
      1. This is a database that includes the list of every TV Show and Movie on Netflix. This data allows people to see which content is available in each country based on directors, genre, actors and other factors.
   2. [**https://data.census.gov/cedsci/all?q=United States**](https://data.census.gov/cedsci/all?q=United%20States)
      1. This dataset contains a large collection of data based on past census records. Information such as housing esitmates, total population, race, age, sex, and much more is detailed in this dataset.
   3. [**https://data.fivethirtyeight.com/**](https://data.fivethirtyeight.com/)
      1. This dataset contains data about different topics such as sports and politics. In regards to sports, it has game predictions for the NBA, MLB, and for Club Soccer. For politics, it contains information about governor races and about voting laws.
   4. [**https://www.kaggle.com/datasnaek/youtube-new**](https://www.kaggle.com/datasnaek/youtube-new)
      1. This dataset contains data about a list of the top trending videos on the platform. Data is included for the many different countries such as the United States, Germany and France, with up to 200 listed trending videos per day. This dataset was collected using the YouTube API. This includes things such as Age, Race and Gender.
4. **Identify who your customer/s would be (8points).**
   1. A customer for data set A would be somebody who is interested in learning about geological differences in accessible information. This could include research into censorship due to movie availability due to religion or politics.
   2. A customer for data set B would be somebody who is interested in learning of various data related to people’s housing situations and information on each individual in these homes. This could include research into housing prices in a general area and comparing that to various factors such as race, sex, and age.
   3. A customer for data set C would be a person who is a sports fanatic and who is very political. This could include research on different sports statistics as well as statistics on politics, which could be compared to sports and politics in other countries.
   4. A customer for data set D would be a person who is interested in learning about demographics of popular media within different countries. This would allow them to sort out different races, age ranges, genders and categories of videos. This is significant because it allows a lower division of cultural barriers in the world.
5. **Describe the problem that each proposed project solves and justify its need (10 points)**
   1. A problem that data set A could solve would be to shed light on countries that suffer from a lack of information. This could allow for more people to be able to use activism to provide freedom of speech and information for everybody.
   2. A problem that data set B could solve would be to determine race, sex, and income inequality that may be prevelent in certain areas of the United States. This could allow for more awareness and research to go into this to see why they exist and what can be done to counteract these inequalities.
   3. A problem that data set C would solve would be to eliminate any uncertainty that person would have when watching sports or reading about politics. This would allow a person to enjoy watching their favorite game without the ups and downs of emotions, and it also would allow a person to have a scholarly political debate without participants becoming angry.
   4. A problem that data set D could solve would be to view popularity over time, at any time. This would allow someone to track how long a video is popular on Youtube, or to track what types of videos are popular the longest. This could help smaller Youtube channels grow, as they could see what type of videos trend the best, and trend for the longest amount of time.
6. **State the product vision and how your web based project would be useful to society (8 points)**
   1. Data Set A:
      1. This product would be made for people who are looking to learn information about censorship between different countries.
      2. There is a need for this information because the general public does not know what information isn’t available to many people of the world.
      3. The name of the product would be “Censorship and Netflix”.
      4. A key benefit of this would be that it would take public information, and organize it in a way that anybody could learn and sort it in an interactive way.
      5. There are not any competitors that do this that I know of.
      6. A difference between other websites and ours is that it is using free information and will provide it to other people in a open and interesting way.
   2. Data Set B:
      1. This product would be made for people who are looking to learn information about various data related to people’s homing situations and information on each individual in these homes.
      2. There is a need for this information because it is important that awareness of unequal housing is made aware to the public.
      3. The name of the product would be “Is Housing Inequality Prevalent Today?”
      4. A key benefit of this would be that it would take information from the census and show if someone’s local area is suffering from housing inequality based on certain factors.
      5. There are not any competitors that do this that I know of.
      6. A difference between other websites and ours is that it is this dataset in a way that addresses individual local areas to see if a user of the app has housing inequality in their area.
   3. Data Set C:
      1. This product would be made for people who want to know the outcome of a NBA game, a MLB game or a Club Soccer game. Furthermore, it would be for people who want to learn about statistics on political parties.
      2. There is a need for this information because it is important for individuals to know the statistics about sports and politics so that they can be able to educate people around the world.
      3. The name of this product would be POLI-SPORT.
      4. The key benefit of this product would be to use information from this census and find countries that are similar in their sports and politics.
      5. There are no competitors.
      6. A difference between other websites and ours would be that it would be designed specifically for the individual and that it would only obtain information about sports and politics from three similar countries.
   4. Data Set D:
      1. This product is for people who are interested in finding out about the different things that trend within different countries of the world. It will be presented neatly so that anybody has the ability to learn what they want to know.
      2. There is a need for this information because it will show what the differences and similarities are between countries and their culture of popularity. This could show many things such as music tastes, comedic tastes and where most of the entertainment in the world comes from. You can learn if people have a bias towards content from their own country, or if they consume from others.
      3. The name of this product would be Organized Video Charts.
      4. The key benefit of this product would be to easily and neatly sort information from each country.
      5. A competitor for this product could be the YouTube app itself. All of the data is available on the website, it just is not organized.
      6. A difference between other websites and ours would be that it would be designed specifically for people with easy user interfaces. This would allow different things to be sorted.

1. **Identify the major features of each of the proposed projects. (12 points)**
   1. **Data Set A:**
      1. The ability to sort number of movies available of each genre in each country.
      2. The ability to sort the number of movies directed by female directors in each country.
      3. The ability to show what movies aren’t available in a country.
      4. All of the above, but with TV also.
   2. **Data Set B:**
      1. The ability to put in a specific local area and see housing data, likely using small dots on the map
      2. The ability for these dots to show information on age, race, and other factors about individuals in these areas.
      3. The ability to show where housing inequality is most prevalent and for what group of people it is most common
      4. Perhaps some Google Maps plugin that will use their map for the purposes of this project.
   3. **Data Set C:**
      1. The ability to search for statistics about any sports game and political party.
      2. The ability to sort different sports teams and politicians based on the individual’s likes.
      3. The ability to find similar teams and politicians across the globe and make comparisons.
      4. The ability to directly message a sports analyst and/or political analyst about upcoming sports games and/or elections.
   4. **Data Set D:**
      1. The ability to sort the similarities and differences of popular media within Youtube. This will show what videos trend in the most countries, and which videos show up the least.
      2. This would show which categories of youtube videos are the most popular in each country.
      3. This would show which countries contribute to the most popular videos, and where most of the views come from.
      4. This would show the different demographics of each country. Whether the people who are watching the videos are Male, Female, and their race and age range.