**Michael Chang**

**SID: 26028400**

**• Choose a name for your group.**

Team VGM

**• What is your overall topic? What question(s) will you address?**

Our overall topic is Movie data. We want to address questions such as - is there a way to predict how good a movie is other than through critics? I.E. total budget put into movie, director, # of reviews, or gross revenue. Who is the most successful Director? Actor?

1. Who are the most popular actors and directors?
2. What are the most popular genres?
3. Is there a relationship between Facebook likes and gross revenue?
4. Is there a relationship between the success of a movie and the budget allotted?
5. What content rated type of movie is most successful, in terms of revenue?
6. How does social media measure the popularity of a movie, in terms of revenue?

**• Where will you get the data? What is involved in obtaining it? What variables will you use?**

Kaggle. To obtain the data, we are creating a Kaggle account and then downloading the .csv file to our directories. We will most likely be using all the variables from the dataset.

**• Is there any initial data processing must be done to put the data in a form that is suitable for visualization or analysis?**

Yes, we are going to want to clean up the variable names so that people can actually understand the data. We also will have to only work with the cases that have data for all of the variables. For the Facebook likes variable, we will have to take into account all 0’s from the dataset so it does not throw off our statistical analysis such as median and mean likes for movies.

**• What plots will you make of the data? Will you perform any analyses, and if so, what?**

1. Plots of each director and actor’s total gross and median score to find the most popular actors and directors.
2. Plots showing the review scores and gross of each genre to find the most popular genre.
3. Plotting the gross revenue against the number of movie facebook likes to see if there is a correlation.
4. Plotting the budget against gross and review score to see if there is a correlation.
5. Plotting the content rating against the gross revenue to see which content rated movies are the most successful in terms of revenue.
6. Plotting total number of likes against revenue.

**• List the responsibilities of each person in the group in completing the rest of the assignment.**

1. Michael: Clean/Questions 4 and 6
2. Grant: Questions 3 and 5
3. Vedika: Questions 1 and 2

**• So that I may help you resolve any issues, what do you think will be the most difficult part of doing this project?**

I think the most difficult part of this project will be working with the SCF terminal processing. Also, collaboration on GitHub might initially be complicated.