**ASSIGNMENT 2: INFO SEEKING**

**DATA SET 1**

1. IMDB 5000 Movie Dataset. (2016). Version 1. [Dataset of *5000+ movie data from 1916 – 2016 scraped from IMDB website*]*.* Retrieved from <https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset> on 9/11/2016.
2. Dataset released under Open Database License, individual contents under Database Contents License.
3. The IMDB move dataset represents movie data from the last 100 years. It contains data on genre, sales, actor, movie poster details (i.e. b/w, color, number of faces), Facebook likes on variables such as director, cast, movie. It would be interesting to see how movie trends have changed over the years—what genres have become more/least popular, is there a growth/decline in certain movie ratings.
4. Potential users of this data might be decision makers of the movie industry wanting to identify trends in movie genres. Advertising executives would benefit from knowing how consumers respond to poster, director, actor information, etc when designing marketing material. The information could help theatre owners select movies based on genre and sales.
5. Three questions this data might help to answer:
6. Does the number of Facebook likes (i.e. director, cast) correspond to gross sales?
7. Does the number of faces in a poster affect gross sales? How has this changed over time?
8. Which genres have generated the highest/lowest gross sales over time?

**DATA SET 2**

1. Sports Illustrated Covers.(February 12, 2015) [Dataset of Sports Illustrated Covers since 1955.]. Retrieved from <https://www.crowdflower.com/data-for-everyone/> on 9/11/2016.
2. Database retrieved from Crowdflower , a provider of free open source data sets for anyone to use.
3. The Sports Illustrated Cover data identifies by year the number of instances a sport was featured on a cover. Of interest would be to identify what sports have gained/diminished in popularity over time and those that have remained consistent.
4. Potential users of this data might be magazine publishers interested in creating a new media outlet based solely on the sport with the most covers (an indicator of reader interest).
5. Three questions this data might help to answer:
   1. What sport has consistently been a featured cover?
   2. Which sport(s) were most popular in each decade since 1955?
   3. What new sports were identified in each decade?

**DATA SET 3**

1. Consumer Complaint Database (September 26, 2015). [Dataset of approximately 629,000 consumer complaints on financial products between 2011 and 2016]. Retrieved from <http://catalog.data.gov/dataset/consumer-complaint-database#topic=consumer_navigation> on 9/13/2016.
2. The Access and Use information for this dataset states it is available for public access and use. No license information was provided.
3. This dataset is interesting as it provides a log of consumer complaints by product, state, company, issue, the manner of how the complaint was filed (web, phone, postal mail, etc) and how the company responded to consumer. From the data, one could determine which financial products receive the most complaints, which companies receive the most/least complaints, and identify the number of resolved to non-resolved.
4. Potential users of this data would be consumers deciding on a financial institution from which to purchase a particular product based on the fewest complaints. Financial institutions could use this information to identify issues with their products and areas of improvement. The Better Business Bureau could use this information to identify companies conducting deceptive sales and practices.
5. Three questions this data might help answer:
   1. Which products receive the most complaints?
   2. Is there a correlation between how the consumer files a complaint and how the issue is resolved, i.e. are complaints received via postal mail resolved in the consumer’s favor more often?
   3. Did the consumer being identified as an “Older American and/or Servicemember” make a difference in how the issue was resolved compared to the consumer not being specifically identified?

**WORD COUNT: 610**