

## DATASET - 1

1. An APA-formatted data citation (including URL)

Chuan Sun, Kaggle (2016). *IMDB 5000 Movie Data Set* (Version 1) [movie\_metadata.csv]. Retrieved from <https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset/version/1>.

2. Details of the license or terms of use (include a link if needed)

Database released under [Open Database License](#), individual contents under [Database Contents License](#)

3. About one paragraph describing why these data are interesting

So, what do I like to do in my free time? – WATCH MOVIES!!

This has been true for as long as I can remember. Movies have always fascinated me and I have never missed an opportunity to see one. From the excitement of watching a preview show, to watching my favorite stars on the big screen, everything about a movie experience was and still is great. At first I was interested only in regional films. Eventually my interests broadened and I started to enjoy films from a wide range of countries and genres. The way movies have evolved, may it be by their use of more advanced technology or their content truly amazes me. Being a frequent IMDB user I was thrilled to find a dataset that I could relate to. The analytical possibilities that present themselves with respect to this dataset is vast for a variety of users, ranging from simple movie lovers (like me) to people in high posts whose decisions based on conclusions derived from studies conducted on this dataset can have a financial impact on an organization.

4. Potential data users and decision-makers for this data

Movie Studios/Producers, Movie Buffs

5. Three questions this data might help to answer; note additional sources needed if applicable

- Whether the total gross of the movie has a relation to the imdb score.
- Whether the popularity of a director affects how many users review the movie.
- Which genre of movies earns the most money?

## DATASET - 2

1. An APA-formatted data citation (including URL)

Baltimore Police Department, (2016). *Calls for Service* [Calls\_for\_Service.csv]. Retrieved from <https://data.baltimorecity.gov/Public-Safety/Calls-for-Service/xviu-ezkt>.

2. Details of the license or terms of use (include a link if needed)



-> <https://creativecommons.org/licenses/by/3.0/legalcode>

3. About one paragraph describing why these data are interesting

The 911 emergency call is a very important necessity and integrated part of society in general. This is the first time I have obtained a dataset from a police department. Having no prior knowledge whatsoever about the functioning of this service, the dataset at hand gives me the opportunity to take a closer look - at least at a general level - of the workings of this service. I'm curious as to whether I can find any patterns that may reconfirm an existing statistic, debunk what was considered to be a known statistic or find something completely new statistic.

4. Potential data users and decision-makers for this data

Police Department, General Public

5. Three questions this data might help to answer; note additional sources needed if applicable

- Whether there is a relation to 911 calls and weekends (Friday and Saturday)
- Whether the high priority cases peak during a certain time frame.
- Whether the number of 911 calls peak during a particular month and if so which cases.

## DATASET – 3

1. An APA-formatted data citation (including URL)

CDC, Data.gov (2016). *U.S. Chronic Disease Indicators (CDI)* [U.S.\_Chronic\_Disease\_Indicators\_\_CDI\_.csv]. Retrieved from <http://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi-e50c9>.

2. Details of the license or terms of use (include a link if needed)

Database released under [Open Database License](#).

3. About one paragraph describing why these data are interesting

This is a dataset of chronic diseases across the country. The data can be used to gain valuable insights into the health patterns of the population. Seeing the importance of the deductions that can be made from the information obtained from this dataset is why I chose it.

4. Potential data users and decision-makers for this data

Health Agencies, General Public.

5. Three questions this data might help to answer; note additional sources needed if applicable

- What are the percentages of occurrence of the chronic diseases?
- Which chronic disease occurs the most (state wise)?
- Is there a pattern in the occurrence of type of chronic disease as we move away from the coast towards the mainland?