**INFM 600 - INFORMATION ENVIRONMENTS  
Assignment - Info Seeking**

1. **DC Metro Crime Rate[1]**

This dataset contains details of crimes that have happened in DC Metro area through the years of 2011 - 2016. There is a lot of potential in this dataset for *Crime Analysis.* There are around 27 attributes on basis of which the crime incidents are recorded.

**Terms of License**

License for the database is unknown. Since, it has been extracted from dc.gov website it comes under Creative Common 3.0 United States <https://creativecommons.org/licenses/by/3.0/us/> which states that it can be shared and adapted for any use.

**Potential data users and decision makers**

* The Police department and Crime Analysts can use this data to prevent future crime from happening.
* The public can use it in general to avoid potential crime areas.
* Real estate agents can use this information to analyze areas and determine which all are the ‘safe neighborhoods’ and use that for marketing purposes.

**Questions that Data might answer**

* Which district/ward/neighborhood has the maximum number crimes reported?
* Which type of offense is the most in DC metro area?
* What is the trend of various crimes (based on either offense of method) from the years of 2011-2016? Through this we can calculate whether there has been an increase in crime or decrease in crime.

1. **IMDB Movie Dataset[2]**

This dataset consists of 5000+ movies from IMDB website. The movies are categorized by 28 variables which include actors in the movie, number of images in the poster, IMDB rating and Facebook likes among a few.

**Terms of License**

The database is *Open Database License* <http://opendatacommons.org/licenses/odbl/1.0/> .

**Potential data users and decision makers**

* Movie Directors and Producers can use this database to decide which actors or genre will generate more popularity. In general, they can use it to make sound decisions before starting the making of the movie.
* Actors can use this to identify which type of movies gain them maximum praise or popularity.

**Questions that Data might answer**

* Which year had maximum number of hits?
* Which language movies have the maximum hits?
* Which director has the maximum hits and maximum flops?

1. **House Sales in King County, USA[3]**

This database contains all house prices for King County sold between May 2014 and May 2015. The details such as the type of house sold, area of house, year it was built etc. can help determine property buying trend in King County area.

**Terms of License**

The database is released under *CC0: Public Domain License* <https://creativecommons.org/publicdomain/zero/1.0/> .

**Potential data users and decision makers**

* Real estate brokers can use this information to study the market trend.
* People wanting to relocate in the area can get an idea as to what houses are sold at what rates.
* People willing to sell their houses can study this data and put an estimate value on their house.

**Questions that Data might answer**

* How many bedroom set houses are sold the most and least?
* What is the relationship between the year a house was built in comparison to the price it was sold at?
* What is the average size of a house in a particular area code?

**REFERENCES**

[1] DC Metro Crime Data | Kaggle. (n.d.). Retrieved September 14, 2016, from https://www.kaggle.com/vinchinzu/dc-metro-crime-data

[2] @. (2016). NYC Data Science Academy. Retrieved September 14, 2016, from https://blog.nycdatascience.com/student-works/machine-learning/movie-rating-prediction/

[3] House Sales in King County, USA | Kaggle. (n.d.). Retrieved September 14, 2016, from https://www.kaggle.com/harlfoxem/housesalesprediction