**INFORMATION SEEKING**

**Dataset 1: Underlying Cause of Death, 1999-2014**

CDC Wonder. *Underlying Cause of Death, 1994-2014*. 14 Sept 2016. Retrieved from <http://wonder.cdc.gov/controller/datarequest/D76;jsessionid=8EA562C914A76FECCF82F4AB074B886D>

Access Information: All of CDC Wonder’s datasets are covered by *CDC/ATSDR Policy on Releasing and Sharing Data.* Restriction Link: <http://wonder.cdc.gov/DataUse.html#>

This dataset is a country-level national mortality and population data spanning from 1999-2014. The data is based on the death certificates of U.S residents. Each death certificate has an underlying cause of death. The number of deaths and death rates can be obtained by place of residence, age group, race, ethnicity, gender, and year and cause-of-death. Statistics on cause of death can provide us information with developments over time on the underlying cause of death. They may be used to determine which preventive measure or in which field research needs to be done to increase the life expectancy of the population.

Potential Data users of this dataset could be medical researchers, doctors, WHO.

Questions:

* What is the crude death rate in one region of the country compared to another region?
* What is the common underlying cause of death among adolescents?
* Is the mortality rate due to health-related issues such as diabetes, hypertension or due to violence such as assault, suicide, rape etc.?

**Dataset 2: Crimes 2012-2015**

Data L.A (June, 2016). *Crimes 2012-2015.* 14 Sept 2016. Retrieved from <https://data.lacity.org/browse?category=A+Safe+City&utf8=%E2%9C%93&page=2>

Access Information: No License information was provided.

The dataset is the combined raw data for 2012 through 2015 for the City of Los Angeles. Analysis on crime data can help the police force tackle crimes in the city in a more effective manner. Through this dataset we can also get information regarding, in which area of the city large number of crimes have occurred.

Potential Data users of this dataset could be the police department, criminal justice, law-enforcement agencies, Housing department, potential future and current residents

Questions:

* Over the years, has the rate of crimes increased or decreased within the city?
* Which areas in the city does maximum number of crimes occur?
* What is the proportion of crimes that have been reported on the same day of incident to those not reported on the same day? What is the nature of the crimes that were not reported on the same day?

**Dataset 3: Speed Dating Experiment**

Kaggle (2016). *Speed Dating Experiment*. 15 Sept 2016. Retrieved from <https://www.kaggle.com/annavictoria/speed-dating-experiment>

Access Information: Unknown License

This dataset was compiled by Columbia Business School professors Ray Fisman and Sheena Iyengar for their paper ‘Gender Difference in Mate Selection: Evidence from a Speed Dating Experiment’. Data was gathered from participants in experimental speed dating events from 2002-2004. During the events, the attendees would have a four minute "first date" with every other participant of the opposite sex. At the end of their four minutes, participants were asked if they would like to see their date again. They were also asked to rate their date on six attributes: Attractiveness, Sincerity, Intelligence, Fun, Ambition, and Shared Interests. Through these datasets we can predict match between two persons depending on their interests and preferences.

Potential Data users of this dataset could be matrimonial and dating websites, psychologists.

Questions:

* Who had more number of dates? Men or women?
* What are the factors that play a role in match making? Is it based on liking or race?
* Do men and women in the same profession match better with each other than those in different professions?

Word count: 566