**Project Ideas**

1. "George Burns Effect"-- How can we predict which individuals are likely to be long-lived (85+ at time of death) despite smoking throughout their entire lives? Look at data obtainable from data.gov and the CDC regarding smokers’ age at death, cause of death, exercise frequency, junk food consumption percentage, psychological treatment/overall mental health, climate where they reside, race, gender, socioeconomic level, genetic markers. This topic would be of interest to health insurance companies in setting rates on customers who are known to be smokers and to tobacco companies who can focus their marketing efforts on those who will likely not suffer the ill effects of smoking.

2. Jury selection--Which individuals are more likely to side with personal injury plaintiffs? Look at socioeconomic background, race, gender, education level, criminal record, age, party affiliation.  Obtain information concerning whether a case resulted in a verdict for plaintiff or defendant and the size of the verdict from verdict reports available through Lexis. Demographic information may be harder to come by as specifically applied to juries since there is some controversy as to whether this data should be collected. Perhaps we could extrapolate from demographics as a whole to percentage of people who are selected for jury duty. This would be of interest to attorneys and jury consultants, as it would assist them in making the most strategic choices in voir dire.

3. Traffic accidents--Can we target which roads (and at which times of day on those roads) fatal accidents are most likely to occur? Some of this information may be obtained from NHTSA’s Fatality Analysis Reporting System.  Additionally, the Kansas DOT website, for example, appears to contain more detailed information as to road type, time of day trends and road conditions. (Their reports are in PDF form, however.) Florida has the FIRES system (Florida’s Integrated Report Exchange System) that provides similar information that must be searched, rather than simply displayed in tabular form. The challenge here will be finding the data in a suitable form. This topic would be of interest to law enforcement agencies, as they can make the best choices in deploying resources in those areas and at those times preemptively (i.e., putting speed traps or DUI checkpoints in those areas and shortly before those times when traffic fatalities are likely to occur in an attempt to encourage safe driving behavior) and to ensure that resources are available to quickly respond at those times. It would also be of interest to auto insurance companies of individuals that live in those areas in setting their rates.

4. Passenger Screening Algorithm on Kaggle—This project sponsored by the Department of Homeland Security attempts to predict which of 17 body zones is likely to contain a threat based on body scans using the High Definition-Advanced Imaging Technology (HD-AIT) system. The data required is provided therein. This topic is of interest to the Department of Homeland Security and the TSA in that it would improve the accuracy of threat detection systems and reduce the number of false positives.