**Capstone Project Ideas – Adarsh Srinivas**

**Mentor: Hobson Lane**

1. **Talking Data Mobile User Demographics**

**Objective and Project Overview:**

The objective is to build a model predicting users’ demographic characteristics based on various factors such as app usage, geolocation and mobile device properties. It’s an on-going Kaggle Competition with the deadline coinciding with the duration of my course. On account of some exposure in data wangling through one of my previous projects, I would really like to work on a project where in I can learn, understand, apply and use the machine learning methods to the best of my ability. Though Kaggle provides with cleaned, well-formatted data catered for analysis, there is some deal of data wrangling I might need to do in this dataset. Having said that, I would be delighted to start applying some cool techniques to see what insights are generated.

**Potential client:**

One of my friend's relative is a founder of a tech startup in Silicon Valley whose mission is to make mobile internet both affordable and available to areas where they are currently not available. They have partnered with a lot of Mobile network operators in Asia who work on providing specialized packages to millions of their customers. The idea behind this project to predict user behavior so as to cater marketing efforts to audiences and help improve their user experience is definitely something my friend's relative would love to know more about and would be benefitted by the findings as they try to look to solve some similar problems using data. I also hope to use this project and my portfolio to showcase and present my work in order to get a job in this field at his company as there could be potential openings related to Data science.

**The data** is available from:

<https://www.kaggle.com/c/talkingdata-mobile-user-demographics>

1. **Prediction of stock prices for some [around 10] of the top companies in the world**

**Objective and Project Overview:**

The goal is to identify the price at which the stock of a particular company will close on a give a day. Professional investors spend their entire careers in studying these stock values. Knowing that people who have very little or no knowledge about how the trend of the stock value will be in future, I aim to build a model that gives us future predictions of stock, thus letting people know when to buy, sell or hold stock for earning quickly. Investors are not able to predict a stock’s future performance due to various factors. The problem here is, what to use to get correct/accurately predicted stock values. Good machine learning algorithms could be used to try to predict the closing prices.

**The data** can be accessed through yahoo finance by building a web scraper to get data using yahoo’s API. Quandl can also be used to get similar data.

1. **Road Safety Clustering Analysis - Which state is safer to drive in USA?**

**Objective and Project Overview:**

The number and types of motor vehicle crash deaths differ widely among the 50 states and the District of Columbia. A state’s population has an obvious effect on the number of motor vehicle deaths. Fatality rates per capita and per vehicle miles traveled provide a way of examining motor vehicle deaths relative to the population and amount of driving. However, many factors can affect these rates, including types of vehicles driven, travel speeds, rates of licensure, state traffic laws, emergency care capabilities, weather, and topography. These factors can be considered and clustering can be done to identify the solution.

**The data** can be collected for different attributes from IIHS website - (http://www.iihs.org/iihs/topics/t/general-statistics/fatalityfacts/state-by-state-overview#Fatal-crash- totals) for year - 2014.

**Deliverables for the Capstone Project:**

1. Code for the project on GitHub.
2. A final paper explaining the problem, approach and findings
3. A slide deck or some other similar alternative