## **CAPSTONE PROJECT PROPOSAL - Bay Area Bike Share**

## Q: What is the problem you want to solve?

A: Answers to lots of questions concerning the stakeholders – bike users, civic transport planners, advertising companies, etc. can be searched within this data set. Some of them can be –

* Finding the difference and similarities morning commute and evening rush period.
* Profile and usage of users in non-peak hours.
* Mostly frequently used start and end stations.
* Influence of weather on the usage of bikes.
* Patterns in memberships.

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

A: The possible clients can be -

* Civic city planners to gauze the popularity of eco friendly mass commutation transport and plan public transport system.
* Advertising agencies to market their products at popular stations and other locations where it can attract maximum coverage for their products.
* Bikes providers to determine demand response for bikes based on weather, time and date variables.
* Users of the bike share transport program.

Q: What data are you going to use for this? How will you acquire this data?

A: The dataset for this problem is borrowed from the Kaggle data sets. The data sets can be downloaded from this site - <https://www.kaggle.com/benhamner/sf-bay-area-bike-share>.

The data sets consists of 4 comma separated value files –

1. station.csv - Each row represents a station where users can pickup or return bikes.
2. status.csv - Each row represents the number of bikes and docks available for given station and minute.
3. trip.csv - Each row represents an individual bike trip.
4. weather.csv - Each row represents the weather for a specific day and zip code in the bay area.

These files are to be analysed to build the predictive model.

In brief, outline your approach to solving this problem (knowing that this might change later).

A: The models to be developed will be determined on the questions that we decide to ask on the dataset. Broadly following models can be thought about.

* Regression models – To predict a response variable based on as set of predictor variables.
* Time Series Forecasting – To determine the demand response curve for the bikes.
* Classification – For membership subscriptions.

Possible challenges

* A huge data set and will involve substantial time in data munging to get meaningful data to begin analyse on. Mismatched formats in data, time weather formats will require an approach to represent the data in a way it can be used for modelling.

Q: What are your deliverables?

A: The following are the intended deliverables of the project

* Code Project in R
* Project Document clearly outlining
* The problem statement being solved.
* Approaches used to model the problem
* Explanation of data sets and predictive models used in modelling
* Results of the predictive model
* Conclusion and further recommendations
* Business study case presentation for prospective customers
* R Shiny web application