## **CAPSTONE PROJECT PROPOSAL - Bad Teeth, Sugar & Government Health Spending**

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

A: Goal is to find relationship between oral health of a country with national parameters such as GDP, adult literacy rate, per capita health expenditure and sugar consumption per person.

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 client can be

* Global and national oral health monitoring organizations – Analysis can help them improve oral health of the population by taking corrective actions on the influencing factors.
* NGO’s – To create awareness campaigns to improve oral health conditions.
* Companies involved in dental products – To know marketable areas for their products to conduct advertising campaigns.

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 one of the Kaggle data sets. The data sets can be downloaded from this site - <https://www.kaggle.com/angelmm/healthteethsugar>.

The data sets consists of 5 comma separated value files –

1. adultliteracy.csv
2. badteeth.csv
3. gdp.csv
4. healthexpend.csv
5. sugar\_consumption.csv

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: Goal of the project will be to build a regression model to predict the relation between predictor variable – oral health of population and response variables – GDP, adult literacy rate, per capita health expenditure and sugar consumption per person.

Drawbacks of the dataset that can hamper with the model

* The bad teeth data set only has entries for 2004.
* Lots of missing values, NA’s present in data set. Might not be able to fill the missing values correctly due to poor data quality.

Q: What are your deliverables?

A: The following are the intended deliverables of the project

* Code Project in R and Python scripts
* 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