**Capstone Project Proposal**

Jorge Moscat

1. **What is the problem you want to solve?**

I want to understand how can energy consumption data from smart meters installed in homes can be use to model the routines and behavior of customers.

1. **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?**

The main client of this kind of analysis will be Utilities companies. With the roll-out of smart meters to households across the country, Utilities now have the potential to capture and store (almost real-time) how much energy are their customers using at any point in time. In the past with analog meters, your electricity company had to send workers to your home to physically check how much you had consumed in a given period. Now, with this vast amount of smart meter data, Utilities companies can model how their customers behave, predict the demand based on historical data, identify when customers are away from home etc…

1. **What data are you going to use for this? How will you acquire this data?**

I have found a large dataset of 7.5Gbs of smart meter data that a Spanish utilities company published for a Datathon. It seems that this dataset contains 1-hour internal electricity readings from more than 100k customers across Spain for a 2 year period.

Url: [www.endesaenergychallenges.com/datathon/](http://www.endesaenergychallenges.com/datathon/)

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

1 – Initial exploration of the dataset

2 – Data clean up and selection of a subset of the full dataset

3 – Online research of what are the most popular kind of analysis performed on smartmeter data

4 – Select 3-4 analysis to be perform

5 – Code and test hypothesis, re-evaluate goals if necessary

6 – Deliver the project

1. **What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

A – Code file with all the analysis performed in R

B – R-markdown file with the clean code and full detailed report

C – PowerPoint presentation with a summary of the key insights