Daniel Jurado – Thinkful Final Capstone Proposal

Data Science Flex

**What is the problem you are attempting to solve?**

When selecting the first capstone project I wanted to find a business-related dataset that I would be able to use for all three capstones. The logic behind this was to simulate working experience as if I were working for a company where I delivered various projects. Luckily, I came across a 30GB dataset for a mobile, music streaming app called KKBOX.

For my first project, I took a month break to learn Apache Spark so I can effectively predict customer churn. Dealing with so much data I took this as a great opportunity to learn about distributed computing. My second project built off my first as I ran a customer segmentation analysis. By understanding my customers better I was able to understand what makes each customer different from one to another, and what behaviors lead to them churning.

For my last project, I want to bring this all full circle by calculating customer lifetime value and it’s effect on customer segmentation.

**How is your solution valuable?**

The problem with a churn prediction model is that the predictions are use-less unless you know what you will actually be losing in terms of value. Although I was able to create several models for churn, I will need to calculate Customer Lifetime Value in order to deliver on this business goal.

**What is your data source and how will you access it?**

The datasource was acquired on Kaggle and is currently in a database I create on Google Bigquery.

**What techniques from the course do you anticipate using?**

Everything I have learned in this course will be needed for the completion of this project. Calculating CLV isn’t so straight forward and thus multiple scenarios will need to be created to cover all possible assumptions. The results will then need to be taken into consideration with my first two projects to see if this information is an overall value add.

**What do you anticipate to be the biggest challenge you’ll face?**

Learning CLV in a timely manner to deliver this project on time.