Springboard – Data Science Career Track Program
Capstone Project 2: Ideas
By Alan Siu 10/9/2022

Idea 1

(1) What is the business problem?

- How can we create a model that will accurately forecast how many customers will walk into three retail locations within a 95% confidence interval?
    - A retail location takes care of our Comcast customers by selling, upgrading, and downgrading our services (TV, Internet, Home Phone, Home Security, Mobile).
    - Comcast's hierarchy has 3 divisions (West, Central, and Northeast) that roll up to our corporate headquarters. Within each division, there are at least 5 different regions. Each region contains around 20 retail locations.
    - We want to forecast how many customers (door swings) walk into our locations on a monthly basis. We give our stores a goal on how many units (services) we sell to our customers every month based on door swings.
    - We currently have daily data and our goal is to forecast at least three months in advance. This will allow us to make any necessary changes to help our retail locations increased door swings when necessary.

(2) Who are the intended stakeholders, and why is this problem relevant to them?

- The intended stakeholders are the market managers and directors of retail locations. The problem is relevant to them because we need an accurate way of knowing how many door swings we will have at a site in a given month. Door swings are an indicator to predict how many sales will come from a location.

(3) Where are the datasets available?

- I have the data set available at work which shows past door swings and all information related to the retail locations. The data has about 15,000,000 rows and 235 columns. I know which columns we will need for this capstone project.

(4) What data science approaches do you anticipate you will use to model the business problem as a data science problem? (*)

- I anticipate using a supervised regression approach. I will have labeled datasets, and I want to predict numerical values.

- How do you anticipate that you will evaluate the performance of each of the data science approaches that you envision?

-
- I will evaluate the performance of this project by comparing our forecast to actual results. I would like to see the forecasted result be within a 95% confidence interval and hopefully get it to a 99% confidence interval.

(5) How do you anticipate that the intended clients will use the results of your CP2 to address the original business problem?

- My stakeholders will use this information to determine how we should set goals for each retail location. This solution will allow us to make better goal decisions and ensure that the front line has a consistent goal every month. The answer will also help with employee experience to retain the best talent. We want to make sure our employees are not given unobtainable goals. If we can forecast how many customers will come into our stores then we can give our sales reps at the stores a realistic goal. If the reps are happy with their goals then we can keep them employed.