**Matt Allen**

**DS 700 Assignment 1**

**2/24/18**

**Exercise 1: Rent vs Buy model**

The variables in the model will eschew mortgage payment plus escrow versus rental. There are already lots of great examples of those like the one at the [New York Times](https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html). While this Rent or Buy model includes financial preparedness, the model’s primary focus is on whether you would enjoy owning a home versus renting an apartment. The variables and their descriptions are below.

**FICO700** is a variable set to 1 if your FICO score is greater than or equal to 700, or 0 if it is less than 700 or the score is unknown. To ensure the best possible mortgage interest rate, it would be advisable to get your FICO score above 700.

**DownPayment10** is a value indicating whether you have at least a 10% down payment ready to purchase a home at your target price. The value is 1 if you have the down payment and 0 if you do not.

**Lawn** is a ranking on a 1 to 5 scale on how much you enjoy mowing a lawn. 5 means lots of enjoyment, and 1 means very little enjoyment.

**Snow** is a ranking on a 1 to 5 scale on how much you enjoy shoveling snow. 5 means lots of enjoyment, and 1 means very little enjoyment.

**Garden** is a ranking on a 1 to 5 scale on how much you want your own garden. 5 means you really want a garden, and 1 means you could care less.

**Rejection** is a ranking on a 1 to 5 scale on how well you can handle rejection. In today’s seller’s market, you will face stiff competition from other buyers, so you will need to be able tolerate a lot of rejection and dashed hopes.

**LengthOfStay** is a ranking on a 1 to 5 scale on the likelihood that you will stay in the same place for at least five years. Five years is a likely minimum on how long it would take to make purchasing a home cost effective. A 5 in this case means that you are absolutely likely to stay at least five years, and a 1 means that you are very likely to move somewhere else within the next five years.

**Maintenance** is a ranking on a 1 to 5 scale on your willingness to complete maintenance tasks like plumbing or be smart enough to call a professional. If you would rather just be able to call the building maintenance person, then the ranking should close to a 1. If you are willing to fix things that break in your house, then the ranking should be close to 5.

**Homebody** is a ranking on a 1 to 5 scale of how you see yourself. Do you never want to leave the house? If so, you should rank yourself as a 5 on the homebody scale.

**NightLife** is a ranking on a 1 to 5 scale on how much you enjoy going out to restaurants and bars. Night life is much more likely to be available to you if you are renting an apartment downtown versus owning a home out in the suburbs. Someone who really likes the night life should rank themselves at 5.

The equation for rent of buy is below.

**FICO700\*DownPayment10\*(Lawn+Snow+Garden+Rejection+LengthOfStay+Maintenance+Homebody+NightLife)**

The range of this is equation is from 30 to 0. 30 indicates that you should absolutely buy a home. 0 indicates that you should definitely rent.

**Exercise 2 Video Reviews**

**Review these videos:**

**Interview with Dr. Sasi Pillay**

**Data Science: Where Are We Going?**

**Interview with Adam Hardy**

**Kenneth Cukier on Big Data**

**Tom Davenport - Analytics and Good Judgment**

**What are 10 key takeaways you identified from these videos? How are the takeaways important and relevant for data science and its effectiveness?**

Information Theory and Machine Learning are important to be able analyze more data.

Data Science is the future of business. We need more analytically trained people. People in non-analytical roles need to become more analytical to compete.

An organization should come together to make common definitions of data. This keeps analysis consistent across the organization.

Have common data infrastructure across the organization versus having analytic fiefdoms. Allows data analysis across multiple domains in the organization.

Think about the audience that you are presenting your analysis to, how it is relevant to them, and what kind of action they can take on it. Present technical work in a way that is easily understandable and useful.

Data has value and can be used to build businesses. The availability of a lot of data can lead to new insights.

There are many things that are being “datafied” now and will be in the future that had not been machine readable in the past. New forms of information will be discovered.

There are ethical dilemmas brought about by the new analysis. Just because you can predict that someone is likely to behave in a negative way, should you make an intervention?

When making business decisions, get multiple people involved and bring analytics to the decision process.

Understand the assumptions behind analytic models, and be ready to change the assumptions to accurately reflect reality.

**Exercise 3 Jimmy’s Best Burgers**

**What are the important elements for each of the steps below?**

**What information would you collect for each step?**

**Problem recognition and framing**

Jimmy’s Best Burgers has faced declining sales and revenues over the last year and a half. They have tried a few approaches already like adding vegetarian option. It has worked to be smart about its logistics. For example, the restaurant locations are near their distribution centers. With these approaches, they have still been losing money. Based on the description of the problem, it seems like they are doing the right things. It is possible that they need to change their target market, or shift dollars to the form of media that is most effective.

**Review of previous findings**

Find if the current market mix allocations are known. See if there were studies done to arrive at the current market mix. Discover if customer data is available.

**Modeling**

Based on the book *Keeping up with the Quants,* there are existing marketing mix optimization models that can be used to find optimal value in advertising.

**Data Collection**

Since the restaurants are franchised, sales data can be asked for from franchise owners. It may also be possible to have people giving surveys at locations and collecting data by observing the customers that come into the franchise locations.

**Data Analysis**

The data analysis can be done by me and my team. I am not sure what tools are available to us, and what the competency in marketing mix analysis is on my team. We may need to hire a consultant that has expertise in this area.

**Presentation and action**

The results can be reviewed by the data science team, the executive team, and the franchise owners to decide if the results make sense and on what actions to take next.

Davenport, Thomas H., and Jinho Kim. 2013. Keeping up with the quants: your guide to understanding and using analytics. Harvard Business Review Press, Boston, Massachusetts.

**Exercise 4 Article Analysis**

**Review the article, The Rise of Big Data. Discuss the opportunities and challenges, especially social and ethical challenges, related to big data as described in the article.**

Opportunities

Big data will give us the ability to quantify things that have not been quantified before. From the new data sources, we can gain new insights into phenomenon. Cars can be taught to drive to make transportation safer and more efficient. Machines can be taught to communicate through natural language. There is an opportunity for a different statistical emphasis. Instead of looking at small samples of data, we look at the population. We can think about correlations. Instead of why, there is an opportunity to discover associations. With big data, we have enough data to analyze more and more specific cases in data sets. There is the opportunity to make things into data that have never been measurable like the example of posteriors and floors. It also has become increasingly cheap to measure things that have never been measured before due to cost, which will allow more insights into the data.

Challenges

The large volumes of data have made the data source messier. The data may be less precise and accurate, but the volume of it may smooth away imprecision and inaccuracy. It is also possible that imprecision and inaccuracy can be systemic in the data, and lead to erroneous correlations. There is a danger of mistaking correlations for causations. Correlations can be beneficial for prediction as in the UPS truck maintenance example, but it can be over reaching as well. It gives states and corporations more power over individuals behavior. The state could interfere with free will deciding if someone is a terrorist based on correlations, when it is not always correct. People can put too much faith in data, and not fully understand its caveats. Things reduced to numbers can oversimplify a problem, and give a false sense of true understanding. It also begs the question of where do humans fit into a world of automation. What do all the people replaced by algorithms do?