**Picking the Best Google Office**

Abhinav Dulur

January 1st, 2019

**Objective**

Use machine learning to help newly employed Googlers pick the best office location.

**Introduction**

Business Problem: A recent graduate is offered a job by Google, but he/she must pick between various Google Office Branches. I will use my own personal experience as well as supervised machine learning to develop a decision tree to help people select which location they are suited for.

Target Audience: Recent graduates and anyone that is offered a job by Google.

**Data**

Foursquare location data will be used, as well as online data sets that contain data about all cities such as crime rate, cost of living, and average salary. Online data will be captured manually, and Foursquare location data will be captured using the Foursquare API.

From the Foursquare API, I will limit my search to 100 locations and 500 meters around all buildings, which is a very walkable distance. This data should be relatively clean and does not require too much data processing.

**Methodology**

I used a supervised machine learning technique, Decision Trees, to build my classifier. I am assuming that recent graduates will have similar tastes to myself (as I am also a recent graduate), which is why they will be able to use my Decision Tree.

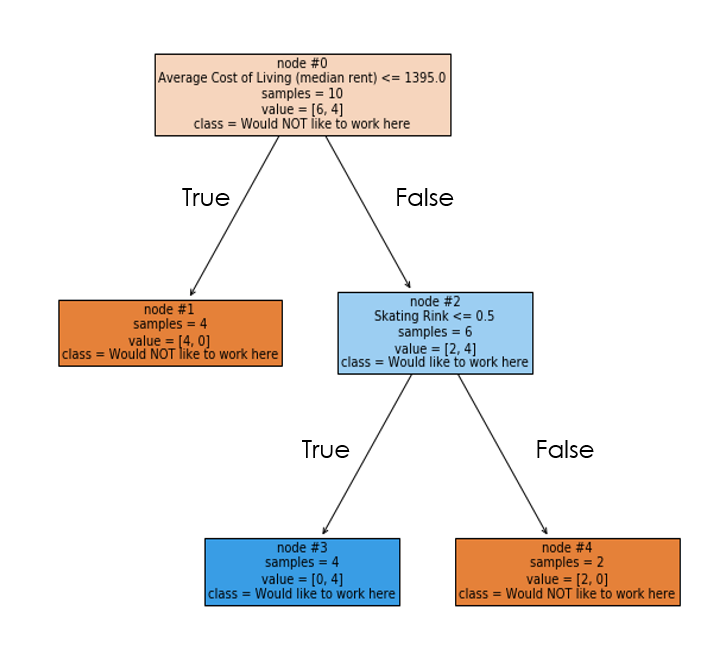
Due to the small sample size, I do not think it is prudent to do a train/test split. However, given a proper sample size (i.e. I visited many more google locations) then I would do the following:

Do a train/test split and calculate the accuracy of my model using a Confusion Matrix.

I opted to use a Decision Tree Classifier due to it’s simplicity and ease of use when explaining to individuals not familiar with machine learning or data science.

**Results**

Figure 1: Decision Tree



From the tree in Figure 1, we can see the first split occurs at the Average Cost of Living. So, if the Average Cost of Living is indeed <= 1395.0 (monthly rent) then I would not like to work there. If the location in question has a monthly rent of > 1395.0, then we can proceed to the next node.

Now, if the location does contain a Skating Rink, I would not like to work there. If the location does not, then I would like to work there. The main distinctive features that are shown is the median rent and the presence of a Skating Rink.

Since the decision tree does not always result in the same tree, and we are not running any statistics, I have run the classifier numerous times and have achieved the same result, so I will assume that this is the optimal tree.

**Discussion**

To reiterate, this tree could be made more robust with the inclusion of more data. However, to keep it realistic, I only included locations that I have personally visited. According to my tree, if someone my age and is a recent graduate is deciding between various Google office locations, then I would recommend them to look at the median rent as well as the presence of a Skating Rink.

I cannot be sure about the explanations behind these phenomena, but I can assume that I prefer high-rent areas due to the increased safety and quality of goods and services. I can also assume that I do not prefer Skating Rinks due to the crowd they typically attract (Skating Rinks are known to attract families and older individuals).

**Conclusion**

In conclusion, I have used personal experience, Foursquare location data, online sources, and supervised machine learning to create a Decision Tree Classifier to help recent graduates select a Google location to work for.

This classifier has concluded that to find the best location, one must look at the median rent and the presence of a Skating Rink.

In future work, to make the classifier more robust, I should visit more Google office locations, as well as add additional online data from various other online sources / API's.