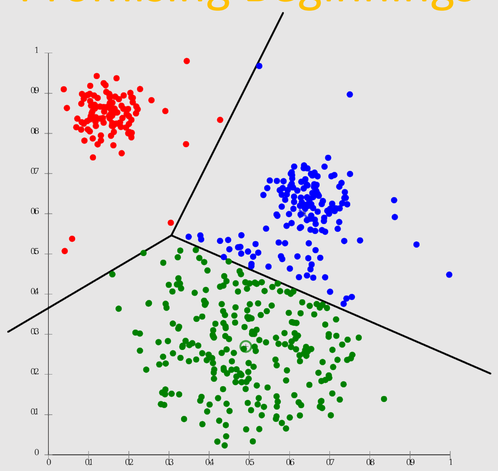
LEED Certification Prediction with K-Means Clustering Algorithm

AI for the Humanities
with Professors Chun and Elkins

By Jack Chase

Final Project
20 December 2018

Promising Beginnings

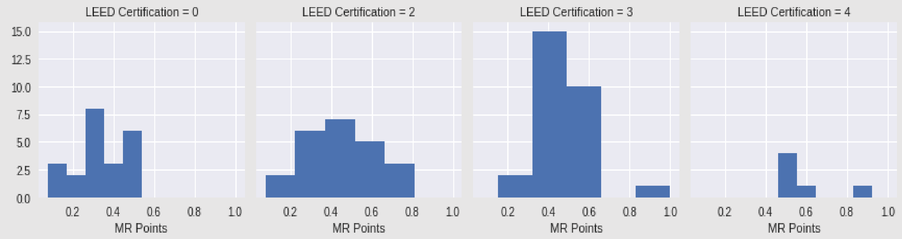


This project uses a K-Means Clustering algorithm. K-Means Clustering is a method of vector quantization, originally from signal processing that is popular for cluster analysis in data mining. K-Means Clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The example above has three clusters, and my project used four clusters, one for each LEED Certification. I coded my project in Python using Google's new online CoLaboratory (pictured below) and Jupyter Notebook.

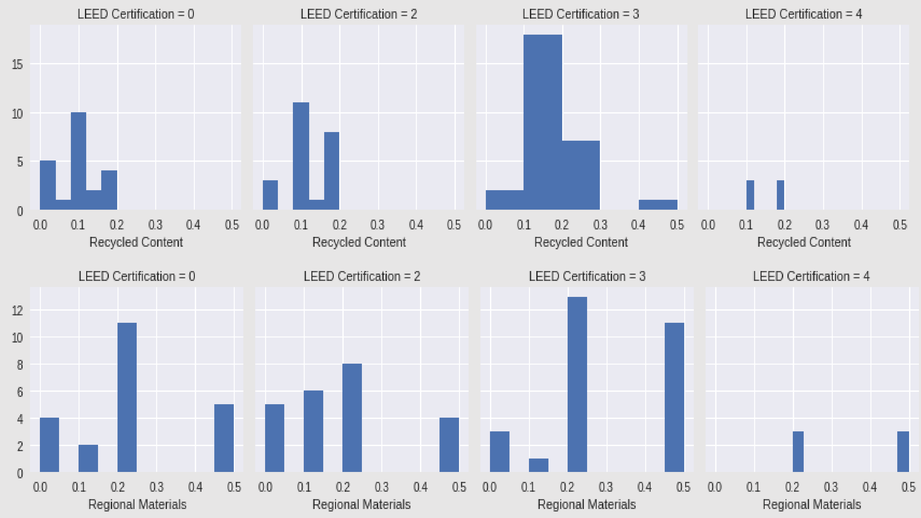
The goal of this presentation is to find the most accurate prediction agent of the LEED Certification, using the Materials and Resources sector of evaluation. When a project is evaluated for its LEED Certification there are six different sectors of evaluation, of which Materials and Resources is most rarely certified as perfect. In fact, the LEED evaluation above has never yet been given in the United States, and usually it is the Materials and Resources section that is responsible for the points lost. Thus, I was able to apply the K-Means Clustering algorithm to determine just how much the materials and resources play a part in the overall evaluation. The bar graphs below detail how a better percentage in the Materials and Resources sector lends the project to receive a better rating. However, these graphs are based on data alone, so let us apply the K-Means Clustering Algorithm.

The trouble with the Materials and Resources evaluation is that it is a full score is only possible if all the construction materials are recycled and/or regional and that all construction waste must be "diverted" sustainably. With most modern ideology, especially the idea that newer is better, recycling the materials of whatever building previously inhabited the space or sourcing local materials is unappealing, especially to the financial sources of the project, which always want the new building to to stand out.

Humbling Conclusions

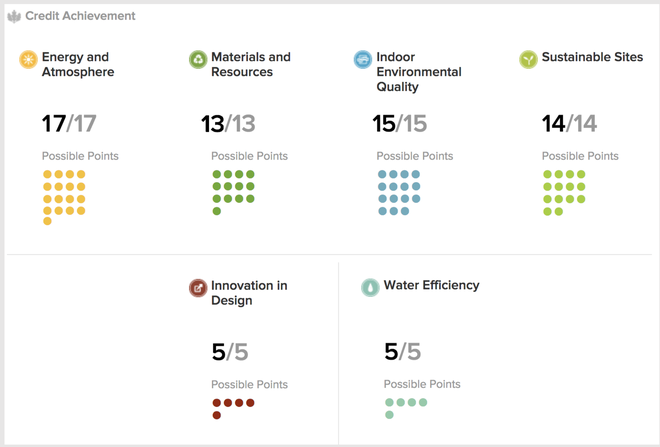


The graph above depicts the number of LEED Certifications (0 = Certified, 2 = Silver, … 4 = Platinum) by percentage evaluation of Materials and Resources points. I generated this graph and the following graphs using Python's matplotlib library, using the data that I collected from the U.S. Green Building Council's website.



As you can see in the graphs above, there is a strong correlation between stronger scores in the two categories depicted (Recycled Content and Regional Materials) and a higher LEED Certification. Any discrepancy could reasonably be explained by the lower overall chance of any project receiving a Platinum LEED Certification.

This is where my project took an unexpected turn. All the information I had gathered up to this point led me to believe that my hypothesis would be correct. Especially considering that Materials and Resources account for approximately 20% of the points able to be achieved for any given LEED Certification, when my K-Means Clustering Algorithm discovered that with all of the information I had collected, it came back with a result of only 24% clustering. Thus, despite that by far most of the energy that contributes to a building's carbon and pollution footprint is during its construction, LEED applies a weight to the Materials and Resources sector barely more than its points would suggest.



Thus, assuming that my K-Means Clustering algorithm is correct, I have first discovered that my hopes for this project were for naught. To accurately predict a rating using only the Materials and Resources sector is not possible. However, in discovering this, I have pointed out a major flaw in the weighting of the LEED Certification evaluations.

After manually collecting three-thousand points of data from all the LEED certifications available in the public domain, I compiled a dataset that includes LEED certifications from every state capital and several of the main cities in each state. This is where I became intimate with the dataset that I believed would be the major predicting factor in LEED certification. Out of the hundreds of cities that I examined, no project had a score higher than 92% in the Materials and Resources sector of evaluation without some of the ordinary factors, including building material reuse and regional resources, being completely and mysteriously absent from the data made available online. Thus, it my hypothesis that the Materials and Resources sector would be the determining sector.

Startling Insights

In familiarizing myself for many hours with the data that went into the K-Means Clustering algorithm, I always felt a sense of unease about the points-based system that LEED Certifications use. When I discovered that some buildings, labeled as "Corporate" had different measures for evaluation, that completely excluded the two categories that most strongly correlated themselves with a higher LEED Certification rating (and also an overall value for the environment), I was concerned. However, it seems that that concern ought to have been even more strong than it was initially. The points system, which somehow weighs equally whether a building collects recycling and whether it was built using regionally sourced materials, is more than flawed in the points within its category but also flawed in its weighting overall.

Thus, it seems that the certification depicted above is just a means for the people seeking it to half-bake a plan to get board approval and public funding. It serves the purpose of encouraging construction to be considerate of their impact on their environment, that much unquestionable. However, all the small, unacknowledged adobe houses in Santa Fe ought to receive Platinum LEED Certifications just as frequently as the 72,000 square foot behemoths that not only cost the tax payers *explicitly* but also *implicitly* in the negative externalities their construction cause, even while receiving LEED Gold.