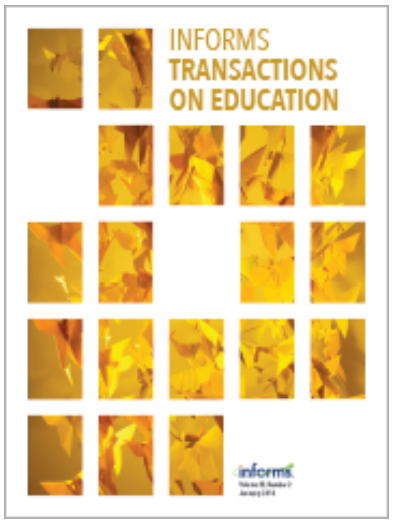

INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Case Article—Converting Zip Code Data into Distances: A Case Study for Teaching Business Analytics

Eric Huggins

To cite this article:
Eric Huggins (2019) Case Article—Converting Zip Code Data into Distances: A Case Study for Teaching Business Analytics. INFORMS Transactions on Education 19(2):105-107. https://doi.org/10.1287/ited.2018.0195

http://pubsonline.informs.org/journal/ited/

INFORMS TRANSACTIONS ON EDUCATION
Vol. 19, No. 2, January 2019, pp. 105–107
ISSN 1532-0545 (online)

Case Article

Converting Zip Code Data into Distances: A Case Study for Teaching Business Analytics

Eric Huggins[a]

[a] School of Business Administration, Fort Lewis College, Durango, Colorado 81301
Contact: huggins_e@fortlewis.edu, http://orcid.org/0000-0001-9724-4682 (EH)

Received: May 25, 2016
Revised: June 21, 2016; August 27, 2017
Accepted: August 31, 2017
Published Online in Articles in Advance: May 2, 2018

https://doi.org/10.1287/ited.2018.0195

Abstract. In this case study, we demonstrate how to convert a column of zip code data into approximate distances from a default zip code using only standard Microsoft (MS) Excel functions and a free zip code database available online. We illustrate one example of analytics that may be applied to these distances with real data. The Case Article discusses our experiences in teaching the case and highlights the positive aspects of it. In the Case Study, we develop a hypothetical scenario in order to convey the value of being able to convert standard zip code data into distances to the students reading the case. Finally, the Case Teaching Note provides guidance regarding classroom discussion as well as specific instructions for developing the MS Excel spreadsheet.

Supplemental Material: The student_zip_codes.zip is available at https://doi.org/10.1287/ited.2018.0195. The Case Teaching Note and final product example are available at https://www.informs.org/Publications/Subscribe/Access-Restricted-Materials.

Keywords: zip codes • microsoft excel • MIS • database • customer analytics • descriptive analytics • operations

1. Introduction

This case study evolved organically from the first offering of a new business analytics class in the business school at Fort Lewis College in Durango, Colorado. It was a great course—open to do whatever was interesting and we ended up creating from scratch a number of cases like this one. For this case, we made a request to a colleague in our office of institutional research to see if he had any recent requests for analytics to be done on our student body. He sent us the zip code information and asked if we could convert it into approximate mileages, which the class eventually did.

2. Literature Review

To the best of my knowledge, this is the first paper to pedagogically explain how to calculate distances based on zip codes using only Microsoft (MS) Excel. By no means do I suggest that I am the first person to *do* this or even *explain* how to do this, but I believe that I am the first author to explain how to do this in a pedagogical setting, i.e., a case study, using real-world data.

Finding the distance between any two zip codes is easy online. A quick Google search yields multiple websites such as https://www.freemaptools.com/, https://distancecheck.com/ and many others that will calculate the straight line and driving distances between any two zip codes. However, these websites do not explain how they are performing the calculation and further, they will only do one pair of zip codes at a time—the case study requires thousands of such calculations. There are commercially available products that will do the thousands of calculations as well as the analytics required for the case study (and more), but two big goals of this case study are to perform the analysis for *free* and to build the MS Excel tool from the ground up so that the students learn by doing, rather than just using prebuilt software. I have not been able to find free software online that will perform multiple calculations and more importantly, neither have my students—if there is an easy way, the students will usually find it!

A thorough (but nonexhaustive) review of the literature that uses zip codes to estimate distances yields a number of papers in the health services industry that examine how distance to a health care provider affects how patients use and experience said health care. Two papers that specifically consider how accurate these estimates are come from Phibbs and Luft (1995) and Berke and Shi (2009). The first paper concludes that straight line distance estimates correlate with actual patient travel times, with an overall correlation of 0.987. This particular statistic was a great find because I can now quote it to my students when

they question whether our straight line distance estimates are valid or not. The second paper discusses different ways to center where a zip code lies—a zip code after all is a two dimensional (three dimensional, really) space that we convert into a point to get latitude and longitude. Berke and Shi (2009) determine that the "population centroid" method is best, but all of their correlations are greater than 0.9. In both these papers, the authors use commercial software to calculate distances.

Similarly, LeBlanc et al. (2004) use zip code estimated distances in a large transportation optimization problem. Understanding that these measurements are underestimates, they multiply all distances by 1.14 as suggested by Simchi-Levi et al. (2000). Again, the authors of this study use commercial software to convert zip code pairs into distances. It is clear that being able to estimate distances between zip code pairs is a useful application.

3. Discussion

I have taught this case twice now, both times with very good results. The first time the class had to develop everything from scratch—finding a good database, finding a reliable distance equation, and all sorts of Excel issues. The best part was that the students bought in to it immediately and I didn't have to spend a minute motivating them to get it done. They intrinsically believed that this was a useful project. A few of the best students took it even further, figuring out how put the information visually on a map of the United States in Excel for each county or state creating a choropleth map. To this day I do not really understand how they did it! The only downside to this experience was that it took up a lot of class time but I do not believe any of that time was wasted.

The second time I taught this case, again in my business analytics class, it had more structure (like what I have written in the Case section). The students still bought into it and it took quite a bit less class time, but perhaps the enthusiasm/excitement level was not the same because they were not creating it from scratch. When I teach this case for the third time next year, I plan to allow the students a little more time for discovery while still trying to keep the project on track.

The students in both classes were junior and senior undergraduate business majors. All of them had some previous Excel experience (having taken an MIS prerequisite) and some of them were flat out Excel experts. The course is an elective so the students had self-selected and I was lucky to have nearly all motivated, analytically-minded students. I believe that this case study is best-suited for an analytics, operations, or MIS class, with upper-level undergraduates. However, I feel that it could also be used in a broad spectrum of other courses as well.

Given the real data set supplied with the case, it could be used in almost any GIS class where students could analyze the data using MS Excel as well as using GIS-specific software like ArcGIS from Esri. Comparing the two techniques could lead to interesting and instructive results in my opinion. Similarly, the case could be reshaped to fit into an engineering or computer science class, perhaps with computer science students writing code to analyze the results using a mix of code and MS Excel functions to achieve the desired outcomes. I also think the case could be used effectively at the MBA level, leaving advanced students to their own devices to figure out how to do it and it might be even better if the students were responsible for finding/supplying their own zip code data and doing a more thorough analysis of the results.

For me, this case is a joy as it involves the best parts of teaching. I do not have to lecture too much, maybe half an hour, to explain what we are doing. As previously mentioned, the students (so far) have bought into the project immediately and cannot wait to get to the computer lab to get to work on it. I have the students work in teams of two or three and most of my time is spent helping them with Excel questions. I use the case as a major course assignment, worth approximately 10% of the students' course grade. Both times that I have taught the case, two great classroom discussions have come up: (1) Are straight line distances really appropriate? And (2) What should be done with the data that doesn't fit neatly into the spreadsheet? I comment on these discussions in the Case Teaching Note.

In general, the students are motivated to work on this case study and they readily grasp the big picture concept—converting the zip codes into distances and states and then doing some basic analytics; the biggest challenges are in the Excel details. As always (in my experience), working with relative and absolute cell referencing and the =vlookup function can be difficult as well as a few formatting issues. I spend most of my time in the computer lab with the students helping them troubleshoot in Excel. However, I feel this is a good investment of my time and the students' time since these are valuable skills to have.

In my opinion, there are three main benefits that my students earn by working through this case. First, I feel that they gain practical experience working with real data on a project that naturally has real world applications. Second, there is a point during the case (which I am sometimes lucky enough to witness) where each team realizes that they can do this for themselves, they do not need to buy expensive software (beyond Excel) to convert the data; further, some of the better teams recognize that they are capable of doing more with the data for free than some software can do. Third, this case helps the students realize that analytics is not a "one-size-fits-all" formula, but rather a philosophy

for approaching problems and then finding the best solution.

The Case has a threefold purpose: (1) To convert a large quantity of zip codes into states and distances (2) using only standard Microsoft Excel functions (3) for free. An online search will reveal a number of links to code in SQL or JavaScript that will perform the distance conversion as well, but these links require computer programming expertise that many students (my undergraduate business students in particular) do not have. To accomplish the work, students will need to have basic Microsoft Excel skills, including knowing how to use the =countif and =vlookup functions and understanding relative versus absolute referencing. I explain how to use these and provide specific Excel functions in the Case Teaching Note, leaving it up to the instructor how much to share in advance with the class.

Finally, what to do with the information once we have determined it? First, it is very easy to simply count up how many (or what percentage of) customers are from each state. This result may have obvious marketing or strategy implications. Second, we can determine how many (or what percentage of) customers are close to the organization, where "close" can be defined in whatever way best suits the organization, and how many are from far away. Clearly, the organization can utilize this information for better decision making. How to best use this information once we have determined it typically leads to a third terrific classroom discussion. My institution is located in a tourist destination; each time we have wrapped up this case study, the class has had a really good discussion about how to apply analytics to the tourism and hospitality field.

References

Berke EM, Shi X (2009) Computing travel time when the exact address is unknown: A comparison of point and polygon zip code approximation methods. *Internat. J. Health Geographics* 8(23), https://doi.org/10.1186/1476-072X-8-23.

LeBlanc LJ, Hill JA, Greenwell GW, Czesnat AO (2004) Nu-Kote's spreadsheet linear-programming models for optimizing transportation. *Interfaces* 34(2):139–146.

Phibbs CS, Luft HS (1995) Correlation of travel time on roads versus straight line distance. *Medical Care Res. Rev.* 52(4):532–542.

Simchi-Levi D, Kaminsky P, Simchi-Levi E (2000) *Designing and Managing the Supply Chain* (McGraw-Hill, Boston).