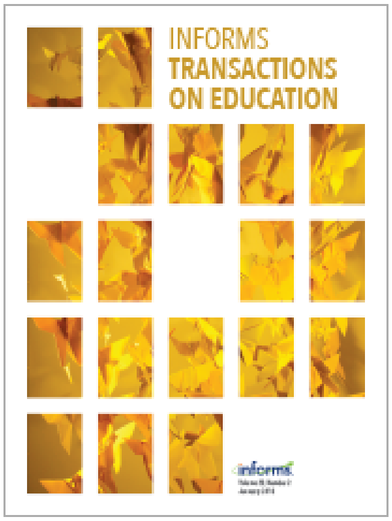

INFORMS Transactions on Education

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

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

Eric Huggins

To cite this article:
Eric Huggins (2019) Case—Converting Zip Code Data into Distances: A Case Study for Teaching Business Analytics. INFORMS Transactions on Education 19(2):108-109. https://doi.org/10.1287/ited.2018.0196

INFORMS TRANSACTIONS ON EDUCATION
Vol. 19, No. 2, January 2018, pp. 108–109
ISSN 1532-0545 (online)
http://pubsonline.informs.org/journal/ited/

Case

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.2018.0196

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

"We need to know where our students are coming from!" thought James to himself as he tried to get a handle on all the requests for information that had recently been filling up his email inbox. James works as a data analyst for the office of institutional research at a small college in southwest Colorado. In the past couple of weeks, he had received more requests regarding information about where students come from than he could count on one hand; as usual, all of the requests were urgent and some of them were from very high up and could not be ignored.

Luckily, all the requests shared a common theme so James felt that if he could solve one of them, he would effectively solve them all. But how? At his disposal, James had the database of all students who attended the college which included the standard fields such as the student's name, age, hometown, zip code, etc. He also had the newest version of Microsoft Excel on a recently refreshed computer on his desk. But he was not sure how to get started, so as he often did, James decided to bounce some ideas off of his colleague Maria.

"What's new in the frantic world of data analytics?" asked Maria, sensing that her colleague had an array of urgent requests as he typically did.

"Good morning to you, too" was James's customary reply. "But since you asked, I've got half a dozen emails wanting to know information about the location demographics of our 3,600 students. The requests are all a little different, but I think if I can figure out how to answer one of them the others will all fall into place." James went on to explain that the two most important requests—based on urgency and who had sent them—were as follows:

1. The first request was quite ambiguous. The provost's office wanted to know how many students at the college were "local," how many were "regional," and how many were from "far away." The second half of this request was more concrete: What percentage of students were from each state, specifically which states made up the bulk of enrolment at the college?

2. The second request was simply to know which students were from "the Denver area." A big alumni event was being planned in Denver over the summer and the person in charge of it wanted to invite all current students that called Denver home to the event.

Maria thought about each of these for a few moments, deducing why each question was trickier than it might seem at first glance. But rather than letting James off easy, she decided to needle him just a little bit more: "Number two is easy, isn't it?" she asked. "You have the student database, just look up which students list Denver as their hometown, right?"

"No it's not easy!" was James's exasperated reply. "Denver is just the city. Like most cities, it has dozens of suburbs and towns that are included in 'the Denver area' but I have no clear definition of what that is, and there are too many suburbs and towns in the Denver area to run an exhaustive list on. What I need is a quantifiable definition of what the Denver area is."

"OK, I get your point" conceded Maria. "I suppose number one isn't simple, either. But just for fun, why don't you just use specialized software for this? As you know, there are a number of commercially available software programs that are specifically designed to answer these types of questions both numerically and with impressive graphics and data visualizations."

"*As you know*," replied James with a smile, "our office does not have the budget to purchase these programs and even if we did submit a successful budget request, it would take months to be approved. Not to mention that those programs require time and training to use effectively. Maybe I could figure it out but it would take longer than I have. I need to solve these problems now with the information and tools that I have."

At this point, Maria and James put their heads together and continued to think about the problem. Maria asked James to reiterate what information he did have from the student database and to identify what the root of the problem was—James needed to quantify abstract terms like "the Denver area" and "local" and so on. Eventually, they honed in on the answer: *distance*. "The Denver area" could be defined, at least approximately, as anywhere within 25 miles of the city center, and if the person who put in the request did not concur with the 25 mile definition, they could use any distance they wanted. Further, they could define "local" and "regional" similarly, within 50 and 250 miles, respectively. If they could calculate the distances between the students' hometowns and specific locations, the solutions to all of James's requests would be in sight!

But how? Again, working together, James and Maria thought about the information that they did have in the student database. They had the address of each student, but finding the distance for each of more than 3,600 addresses to a specific, default address was daunting. It could be accomplished (for free) by using a mapping website and comparing each address to a specific address, but this would take several grueling days to complete and would have to be redone every time the specific address changed. They needed an automated system that would work instantly, and finally they arrived at the answer: *zip codes*.

"Pythagoras!" exclaimed James. "We also have the zip codes for each student. I have found an Excel spreadsheet online that lists the latitude and longitude (and state) for every zip code. We compare the latitude and longitude for each student to the latitude and longitude of the specific location we want, find the hypotenuse of the triangle and that is the approximate distance we need. Sure, it's not exact, but it's a very close proxy to what we want."

"The last time I checked," chided Maria, "the Earth isn't flat, so the Pythagorean Theorem won't work. But I bet there's a formula that accounts for the curvature of the Earth, let me check online.... Yep, here it is, and it's written in MS Excel format to boot!"

Assignment

Your assignment is to create the spreadsheet that James needs to answer requests 1 and 2 from the previous page. You will need to find the database online that provides the latitude, longitude and state for each zip code in the United States. (Hint: federalgovernmentzipcodes.us.) You will also need to find the MS Excel formula that calculates the distance between two locations based on their latitudes and longitudes. (Hint: Google the previous sentence.)

As a suggestion, create a spreadsheet with three sheets (or tabs). Make one sheet the Dashboard, where you will display all of your final results. You will want this sheet to look sharp so use whatever Excel formatting tricks you like to make your final results jump off the page. Make the second sheet the Data page, where you will insert the 3,647 zip codes from the included data file in one column and list the distances from the default location and the state in two other columns. Finally, the third sheet will be huge, including the entire database you found online along with a new column of the distance calculations between each zip code and the default location.

Request 1 Specifics

The default location is Durango, Colorado, with zip code **81301**. The provost's office has requested to know what percentage of students live within 10, 25, 50, 100, 250, and 500 miles of this location (and what percentage are from more than 500 miles away), both relative and cumulative frequency. Second, what percentage of students come from each state? (Spoiler Alert: Colorado is more than half. But what other states have significant enrolment and are there any states with no students?)

Request 2 Specifics

The default location is downtown Denver, Colorado, with zip code **80202**. Since the names of the students are not included in the data file, simply count how many students are from within a 25 mile radius of the city center. (If the names were available, it would be straightforward to sort the data and create a list of students to invite to the alumni event.)

Further Consideration

The previously mentioned hypothetical scenario is loosely based on a real situation. The value of knowing student location demographics to an academic institution should be apparent. How would this apply to many businesses? What kinds of businesses and organizations would have lists of where their customers/clients are from, i.e., their zip codes? How could these businesses and organizations use this information, converted into distances, to improve communication with their clients, marketing to their customers, distribution of products, etc.?