ZIP CODE CONVERSION

BA638\_Data Driven Decision Making & Optimization

Professor: Carol J. Cagle

Student’s Name : Hien Do

Khanh Dang

Ravi Bandi

# **Abstract**

Institute research department faces a challenge to define how far students are from campus (Atlanta or Macon) with a given standard data files including students’ ID, students’ states, and students’ zip codes. There are numerous business problems which are related to defining a student’s location. Firstly, the university cannot identify the proportion of in-state, out-state and international students. Secondly, a big alumnus event is being planned in Macon campus over the summer organizers wanted to invite all current students that called Macon home to the event. Thirdly onboarding a greater number of international students in the university. The stakeholders include students who are considered targeted customers, apart from available given information, if we can get more information from them such their ages, towns, cities, majors. Another stakeholder is the person or analytics department who must handle the zip code data file. The marketing department who will use the information to make a suitable campaign to approach more potential students, the professors and staff who will work closely with students later if they successfully enroll to university. The objective of the project is to identify an ideal location for all students to attend alumni events and formulating strategy, based on analysis to attract more outstate and international students.

# **Analysis scope:**

There are many software's available online to calculate the distances and time driving between two zip codes, but all of them are not free, moreover, these programs require time and training to use it efficiently. The research department only has the data with limited information of students, but they must analyze the data and utilize this information for better decision making. The university can control some basic information provided by students but cannot control the quality of input information. No additional budget or time can be allocated, so we can use tools like excel and math functions to fulfil the requests. Results generated from the analysis can be used for further business development such as in-person and online class programs, marketing events to promote the educational institution, or providing more additional service for students like a job fair.

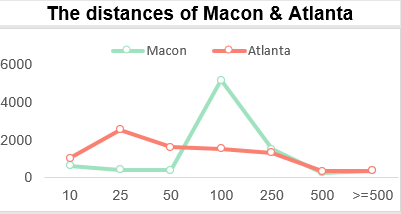
# **Data Process**

The original data**[1](#_Appendix" \t "_top)** includes students ID, state, zip code of 8,876 students. We need more information which is latitudes and longitudes to calculate the distances between two zip code. We downloaded the latitude and longitude for each zip code state in the US file in *federalgovernmentzipcodes.us* website, then we combine two files. After finding longitude and latitude for each zip code, there were 74 invalid values left, we manually search for each zip code on Google to figure out where the location. There are four rows that we cannot find any information, so we decided to remove them. There are 2 zip codes that students are from Texas and we can find its longitude and latitude as well. For 68 left rows, we found that there are exactly 20 zip codes coming from other countries outside the US such as India, Canada, Brazil, we labelled it as International students. The other 48 rows, these zip codes belong to many countries that means, many countries have the same first five digits postal code, so it would be better idea to collect more information such as the actual address of the student, or the city of each zip code. However, the number of invalid values accounts for a small percentage in the entire data, we decided to label it as “Unknown” now. The functions in excel are given in the appendix.

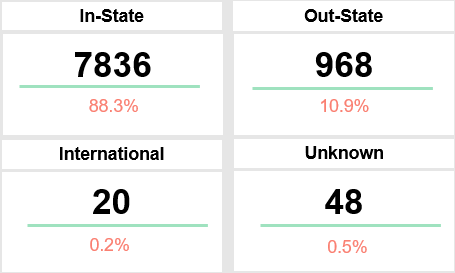
There are two methods to approach the distances straight line and great circle distance method. The distances between two coordinates (latitude and longitude) are calculated based on the formula called “great-circle distance”. The formula is given in the appendix. The distances model has been built in Excel with given data. With the model, we can calculate the distances of each student’s location automatically. Moreover, we also get more crucial outputs for further analyzing. For example, we calculate the number of in-state, out-state, international students.

# **Results**

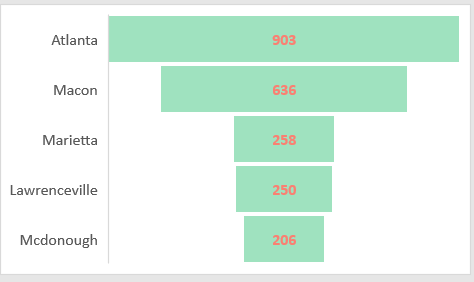
Results are taken from the Dashboard tab in the excel file:

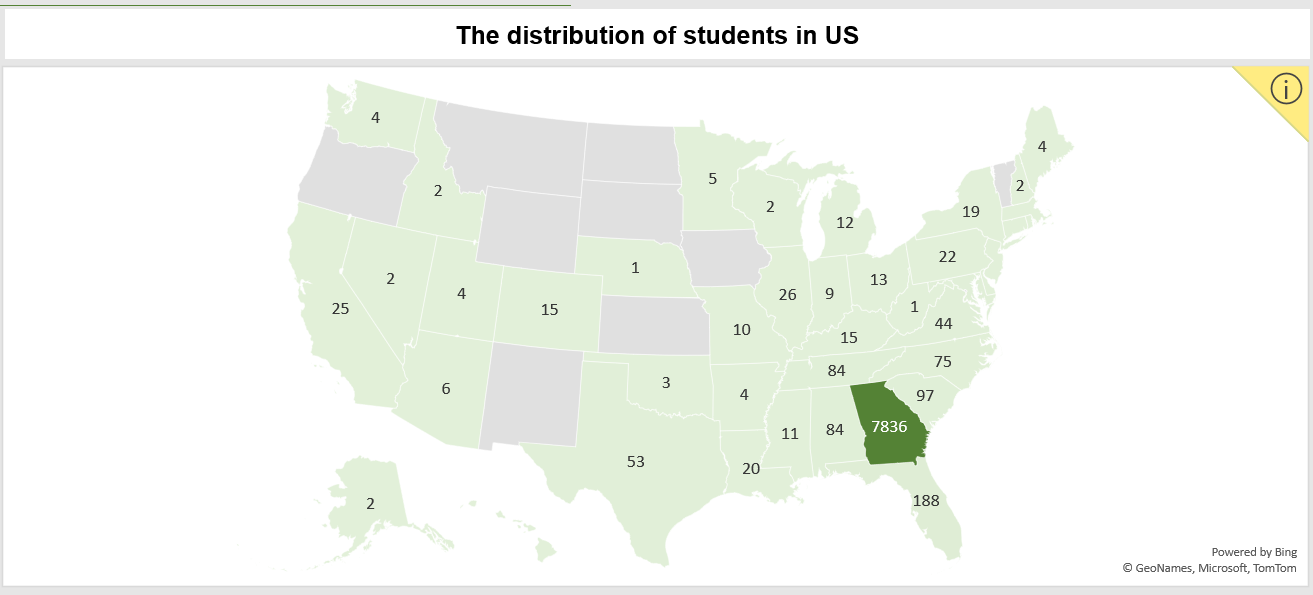


The graph shows the number of students who live within 10, 25, 50, 100, 250, 500 and greater than 500 miles from their location to two campuses Macon and Atlanta. As we can see from the table, most of our current students live far from Macon campus. More than 5,000 out of 8,000 students live within 50-100 miles. While the distribution of distances within the Atlanta campus seems to be nearly similar.



There are more than 88% of current students coming from GA which we label as in-state students, tuition fee for a resident student is much affordable than that of out-state or international students. However, the international students who are just 0.27% are considered to bring the most revenue to the university in comparison.





The top 5 cities that have the highest number of students are Atlanta, Macon, Marietta, Lawrenceville, McDonough. Whereas the top 5 states which have the largest number of students are Florida, South Carolina, Alabama, Tennessee, and North Carolina, these states are also the neighboring states of Georgia.

# **Recommendations**

The recommendations are based on the results we explored.

From specific students’ locations, the University can decide where they should hold the marketing campaign to attract more students effectively. For example:

* In-state students: They can hold the event in the top 5 cities in Atlanta, besides they can estimate the number of students living nearby and invite them to attend the events.
* Outstate students: Because the top 5 states are not far from GA, the university can assign someone to go high school directly to introduce about the university’s academic programs.
* International students: The university, especially the international department, should pay more attention to this group to increase revenue. For example, to connect with more global students, the university can collaborate with consultancies such as Shore light and also offer students feasible more scholarships.

Macon is our main campus, however, there are a high proportion of students live further from there which can be one of disadvantage that students will consider when they want to apply for a program. It can be the consequences of the shortage of housing around campus. The university should offer better housing packages or facilities to increase housing occupancy.

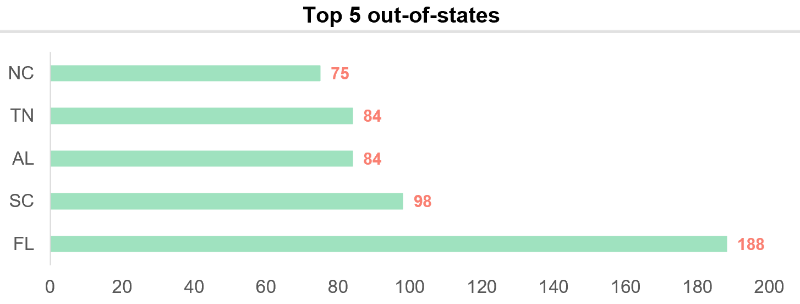
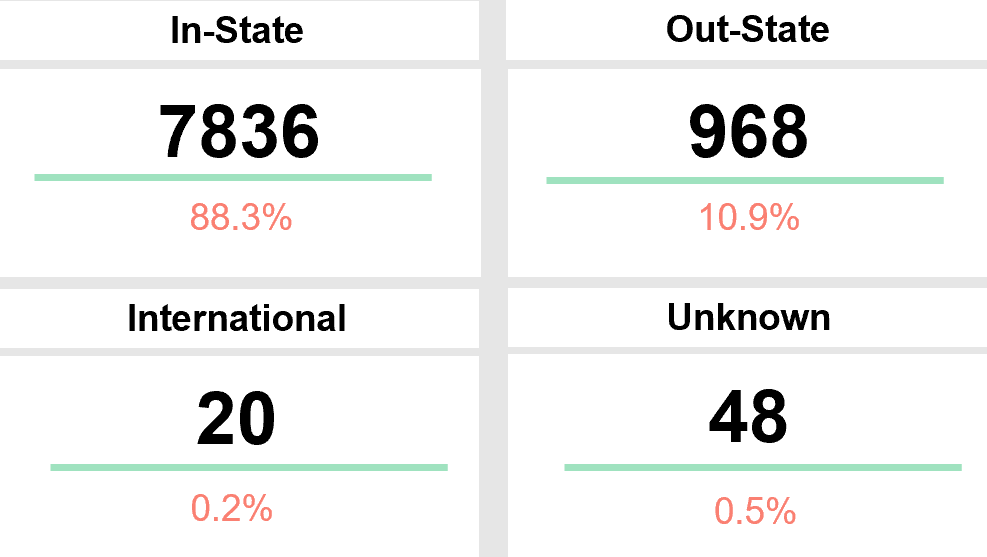
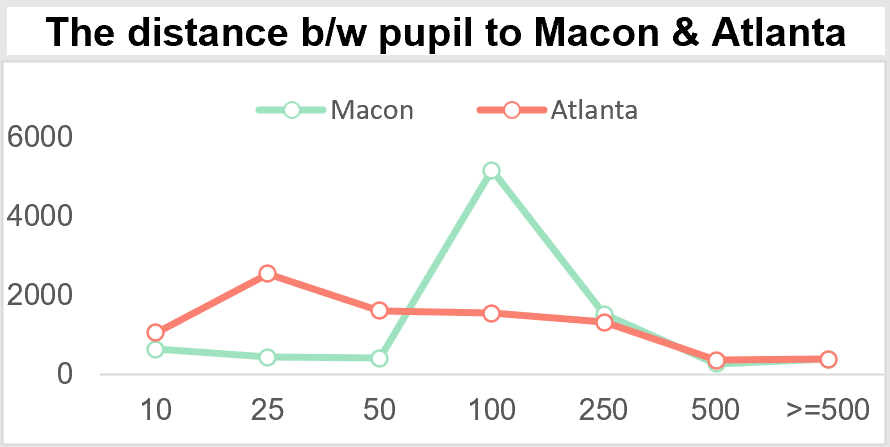
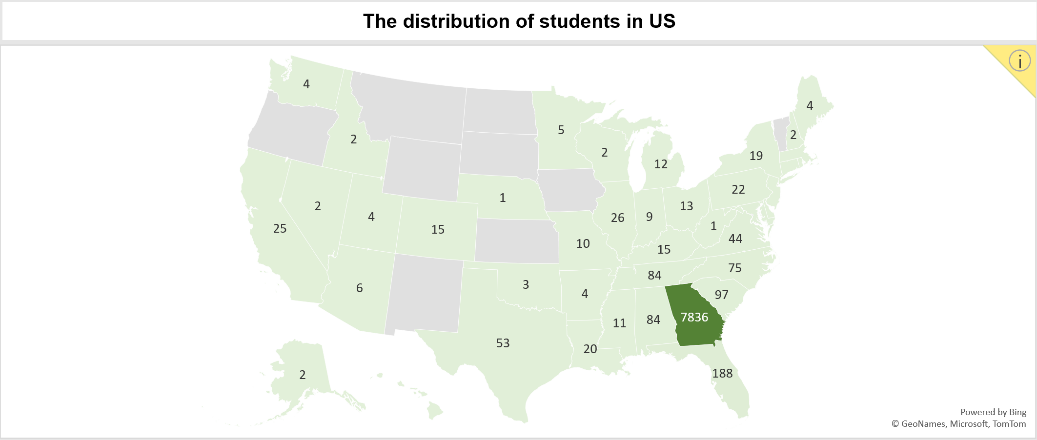
Overall Assessment

This analysis is not only used for the academic institution but also can be used by any organization, especially retailing industries, where the cost of transportation accounts a lot in the entire total cost. For example, a retail store wants to make efficiency in its distribution channel, they should analyze their customers’ location.

# **Appendix**

1. Data Dictionary

|  |  |  |
| --- | --- | --- |
| **Field name** | **Description** | **Data Type** |
| System Student ID | Student Identification number | Number |
| State | State code | Char |
| Student Zip Code | Postal code of the student location | Varchar |
| Zip Code Standard | Cleaned postal codes of student location | Varchar |
| Latitude | Geocode of location | Number |
| Longitude | Geocode of location | Number |
| Distance from Macon | Distance from student’s location to Macon | Number |
| Distance from Atlanta | Distance from student’s location to Atlanta | Number |
| Country | Country of student’s location | Char |
| City | City of student’s location | Char |

1. Distance in Miles= ACOS[(sin(Lat\_place\_1\*PI()/180)\*sin(Lat\_place\_2\*PI()/180)+cos(Lat\_place\_1\*PI()/180)\*cos(Lat\_place\_2\*PI()/180)\*cos(Lon\_place\_2\*PI()/180-Lon\_place\_1\*PI()/180)) ] \*3443.8985\*1.15077945
2. <https://blog.batchgeo.com/manipulating-coordinates-in-excel/>
3. 
4. 
5. 
6. 
7. 