DATA 325 Project 1: Admissions Consulting

Each year, The College of Wooster buys information on over 100,000 U.S. high school students, through The College Board (which runs the SAT and ACT tests) and other organizations. Wooster’s Admissions office reaches out in various ways to these prospects, and a small percentage eventually apply to Wooster.

**Your task is to analyze the data, *finding actionable insights*, that would help the Admissions Staff to refine their name-buying practices and associated outreach**. Location is likely important, and you may find other criteria that are also relevant. There are about 3.6 million U.S. high school graduates each year, so Wooster purchases information on only a small fraction of the available pool.

Your deliverable is a ***seven-minute non-technical talk with slides***, appropriate for Admissions leaders, to be followed by a brief Q&A. You should also prepare a short report (2-3 pages) summarizing your findings with bullet points of key information, possibly one or two data visualizations, etc. There will be a space to submit your presentation on Moodle in advance of your presentation. Presentations will take place during class on February 15th and February 17th. **Everyone must submit *all* their files online in Moodle by February 14th at 11:59PM.**

Information about the data:

Anonymized data is provided on 263,139 domestic prospects for the College’s classes of 2021 and 2022. In order, the data attributes are as follows:

* ID number (sequential positive integers)
* Class of 2022 prospect (binary, as opposed to class of 2021 prospect)
* Self-identification as male
* Underrepresented minority status (Black/African-American and/or Latino/Latina)
* Home zip code (five-digit number)
* Home state (two-letter postal abbreviation, this is the only non-numeric column of data)
* Whether the student appeared on lists of students scoring 28-36 or 23027 on the ACT test (Note: not all students with ACT scores in those ranges appear on these lists.)
* Sophomore year in high school when first contacted (as opposed to junior year)
* Code number for the student’s first-choice major interest, if any (see list of major codes)

Using US Census Bureau data, the following are provided about the student’s home zip code area:

* Population density (person per square mile)
* Percentage of Black/African-American residents
* Percentage of Latino/Latina residents
* Percentage of local students attending private high schools
* Percentage of the adult population holding at least a bachelor’s (4-year college) degree
* Percentage of the adult population holding an advanced degree (e.g., M.A., Ph.D., M.D)
* Median income for households, families, and families with children under 18 at home
* Percentage by economic class (Lower, Lower-Middle, Middle, Upper-Middle, and Upper)
* CBSA number (a census reference to a city/town or local area, see provided list)
* Whether the CBSA is in a metropolitan area, as oppose to a micropolitan (more rural) area
* CSA number (Combined Statistical Area, for larger metro regions, see provided list), if applicable
* Distance to Wooster, in miles
* Latitude (north) and Longitude (west) coordinates of zip code (maybe useful for any mapping)

The associated dependent variables are as follows:

* Whether the student made an inquiry for more information about Wooster
* Whether the student eventually applied for admission to Wooster (our key target variable)