Florence Lee (fswlee)

SI 618: Project Proposal

Introduction

I have been working with Christopher Brooks, a Research Fellow at the University of Michigan (UM), looking at University of Michigan students' relationship with Massive Open Online Courses (MOOCs). We recently finished administering a survey to a random sampling of University of Michigan students. I want to use this dataset for my project.

The Data

The dataset contains 798 responses from University of Michigan students. The questions on the survey are almost all categorical variables, with some nominal (Yes/No), and others ordinal (likert scales, 1-5). Some of the main variables include whether students have enrolled in a MOOC before (Yes/No), reasons for enrolling in a MOOC (likert scale for every reason), whether UM should offer course credits for UM MOOCs (Yes/No), and how MOOCs might be funded (likert scale for every reason). These results will be combined with student demographic and background information such as age, gender, year, declared major, school, and GPA. Currently the survey dataset is in a csv file. I will need to code all the questions then create a dataset that combines the survey responses with student demographic information.

Exploratory Questions and Method

1. What are the differences in opinion between students who have taken a MOOC compared with those who have not taken a MOOC before?

To look at this, I want to create two factors, those who have taken a MOOC and those who have not taken a MOOC before. I want to look at all the opinion questions, such as whether UM should offer course credits for UM MOOCs, or whether UM MOOCs are a benefit to UM students, to see if there are differences between these two groups. To visualize this I would create histograms to show the different distribution of opinions between the two groups.

2. How do students think MOOCs should be funded?

For each hypothetical funding option, I want to look at the average opinion, then at the distribution of the responses to see how widely opinion is spread (between strongly agree and strongly disagree). To visualize this, I would create a histogram for each hypothetical funding option that illustrates the range of opinions.

3. Do students' opinion on the impact MOOCs have on their education differ depending on their education level?

To look at this, I need to create factors based on students' level of education (first to fifth year, graduate). I want to look at all the questions specifically related to students' education (ex. whether participation MOOCs has influenced academic study or whether students feel volunteering on UM MOOCs would be beneficial). For questions with a likert scale (ex. how likely would you be to enroll in a MOOC in the next six months?), I will create a histogram of the responses, split into facets based on level of education. For questions where the results are nominal categories, I will create a pie chart for each facet.

4. Do student opinions about MOOCs differ depending on their GPA?

First I want to use scatterplots to visualize the responses to different questions about student opinions on MOOC (ex. whether students think UM should give course credits for UM MOOCs, how UM MOOCs should be funded, and the level of perceived difficulty of MOOCs) by GPA. I will use smoothing to help determine if there are any trends that occur as GPA increases. These will be visualized in different scatterplots with the trend line.