Explanatory Visualization with MOOC Dataset

Qi Zhao

April 4, 2018

1 Links

This is my **final version** and my **original version** of my Tableau story. You may also check out my **markdown file** to see how I did feature engineering from the original **Kaggle dataset**.

2 Summary

Through feature engineering I got features of 70,000 ID's who enrolled in an online course between 2013 and 2014. They are either labeled "finished the course" or "dropped the course".
One thing we need to notice is that each student may enroll in multiple courses at the same time and therefore get multiple enrollment ID. My findings are about the difference between those who dropped the course and those who didn't.

3 Design

• I chose this dataset because I am attending the Udacity Nanodegree. I feel it would be meaningful to do some analysis about students' learning pattern.
• I mainly used bar chart because after tried out all the fancy plots (pie charts, line charts, bubble charts, etc) I still feel the bar chart is the best way for me to explain my findings considering my dataset is already complicated.
• The first dashboard I simply listed all the huge difference between the "drop" group and "finish" group.
• The second dashboard showed the similarity and difference of the distributions of some important features in two different semesters.
• The third and fourth dashboard was about the impact of the starting time. Did the student started in the first week or second week of the class? Which weekday did the students started their study?
• I chose "hours" as the unit for efficient study time and I chose "days" as the unit for the length of the whole study period. In the first version, I forgot to point it out and people were confused. I add the unit in my axis label in the final version.
• In the first project I put too many information in one dashboard and it was hard to read. In the last version, I separated the dashboards and made it less crowded.

4 Feedback

I got many feed backs in one day and I really appreciate those who helped me! Here are the valuable feedbacks.

"Looks good to me, I like the added notes for clarity. Spotted a small typo, I think it should be 'Quitter'. Also, the only graph that was a bit difficult to understand was the 'Average Processes Period', not sure what the units are for that one? I'm assuming days but would be helpful to see it in the label." - Aamir Raza

"I liked that you split the "impact of course size" on the first slide into separate columns instead of overlapping them. The x-axis of the "impact of study time" seems cut off. I like the analysis

about the clustering of dates and the semesters in the second slide. I wasn't clear on what "effective time" and "process period" meant in the second slide. I wasn't sure what the 0-2.0 scale on the "which weekday do people start to study" meant (I think it's days, but then hours are used in the last visual, so I was a little confused...a label might help). If I understood that visualization, that is a very interesting finding that the day you started also seems to correspond to the day on which you spent most time studying." - Fred

"I like the update, spreading it out definitely helped. Still don't know what the avg process period unit is, is it days? Same for the last slide, I think it would be nice to know the units of the Avg Effective Time (hrs) within the chart y label. What fred mentioned earlier, the x-axis for the 'The impact of study time' chart still shows cutoff for me in Firefox, not sure if that's just Tableau formatting issues or what. Nice choice on colours, they're pleasing to the eye"- Aamir Raza

"I like the organization. Lots of info, but not crowded or too busy" - roywangf

5 Resources

This is my **Github** that contains all my files including the original and cleaned dataset along with my R code and Tableau workbook.