MBA 564 – Analytics Applications Across Business Functions

Module 8 Team Assignment – Final Project

G047 – Karoline Goncalves Melo Sa, Robert Howell, Jamie Landis, Seok Lee

Executive Summary

The work conducted concludes analyses of the University of Illinois at Urbana-Champaign's iMBA program. We formulated a satisfaction survey for the program students and analyzed social media sentiment for the program.

The survey measured satisfaction of five program characteristics and overall satisfaction by using Net Promoter Score (NPS). Additionally, seven demographic questions were asked to enable segmentation of the 97 respondents. Among other interesting findings, we discovered that overall student satisfaction is high (the median student is considered a "Promoter"); among male students, the satisfaction is lower and varies more; the program seems more appealing to those with dependents; and that generally students value curriculum diversity, tuition cost, and networking opportunities more than faculty excellence and ease of use of the technological platforms.

For social media sentiment, we analyzed mentions of the program within Reddit and Twitter using the Vader and sentiWordNet sentiment analysis. In addition, on Twitter, we also benchmarked the iMBA against other online MBAs. Overall, we found that the awareness of the program on social media is low (e.g. the volume of mentions is around 35 times in a month on Reddit). The sentiment of those mentions is mostly neutral 80%- 92%), followed by positive references to the program. Comparatively, the program sentiment is more favorable than competing online MBAs in general.

We recommend further analysis to be conducted in three areas. The first is to better understand differences in satisfaction among different demographics, such as gender and students with dependents. That will allow the program to offer a more tailored advising experience to address the gap between those populations. The second is to analyze why only 25% of students ranks the curriculum as "very satisfied" and make improvements to course offerings, given that is the main driver of overall satisfaction. Finally, additional investigation is needed to understand the root cause for the different sentiments about the program on social media.

To conclude, we recommend that the program further invest in its marketing and social media presence to increase brand awareness, such as hosting interactive sessions like Reddit AMA (ask me anything) to spark discussions about the benefits of the program. Moreover, our survey has shown that the program resonates more with students that have dependents. Specifically, social media campaigns should highlight the program flexibility for that segment.

Customer Satisfaction Survey Findings/Discussion

Survey Creation

To land upon a survey choice, the team brainstormed topics that were important to us. We agreed that the successful execution of the University of Illinois' iMBA program was critical to our graduation success as students. The iMBA is an accredited MBA program that democratizes learning opportunities for working professionals. It offers online courses and virtual engagement opportunities. Specifically, we looked at student satisfaction levels to assess if the iMBA program is delivering on student needs and expectations.

The team addressed how survey questions could be read and interpreted to ensure the questions were concise, easy-to-understand, not "double-barreled," provided valuable insight, and could be leveraged as leading indicators of student behavior. The questions avoided confusion and chaos through understanding the psychological response process of comprehension, retrieval, judgement, and selection (Willis, 2009).

- Comprehension: simple, offer "no response" options, have balanced scales, and avoids complex syntax.
- Retrieval: the distinctive questions occur at important student milestones, so easier to recall.
- Judgement: allow for a quick recall and assessment of the student's iMBA experiences.
- Selection: easy to translate judgement into a response with the simple scale and the open-ended question allows for better understanding of student feelings and attitudes.

Independent Variables

The five independent variables considered in the survey were the tuition cost, the faculty and staff excellence, the specializations and courses curriculum, the ease of use of the two main platforms used in the program (Compass2g and Coursera), and networking opportunities. These five key attributes are considered important to the success of the iMBA program and were rated on a consistent 7-point Likert scale as shown in the Table 1. The respondents indicated a degree of satisfaction or dissatisfaction.

Based on your experience, please rate your satisfaction with the iMBA program for each of the following:							
	Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Satisfied	Very satisfied
Tuition cost							
Faculty and staff excellence							
Courses and specializations curriculum							
Ease of use of Compass2g and Coursera							
Networking opportunities							

Table 1: iMBA Program Survey - Likert Scale

The tuition cost was chosen to consider the financial aspect and reflect the student perception of cost and return. The second variable evaluates the quality of the faculty and staff. The courses and specialization offerings variable assess the diversity of learning opportunities and how it aligns with the student's goals. The ease of use of the technological platforms rates the program resources and support. Finally, the networking variable addresses the opportunities promoted in the program to strengthen the student connections and the impact to their graduation success and career development.

Dependent Variable

The team landed on a dependent variable that would leverage a Net Promoter Score (NPS) (Qualtrics, 2020). We felt that NPS would be the most important indicator of the iMBA program's success: "*On a scale of 1-10, how likely are you to recommend the iMBA program to a friend or colleague?*"

0-6	Detractors (unhappy, flight-risk)
7-8	Passives (like, not love)
9-10	Promoters (love, will refer)

Table 2: NPS score

In addition, the team included an open-ended question "*How has your experience with the iMBA program been?*". NPS identifies student satisfaction levels, student loyalty, and how likely students are to recommend the iMBA program to others. NPS can also be used as a growth indicator. The NPS & open-ended questions can be used to predict student turnover rates and identify "at risk" students to flag for early intervention, and/or address gaps in the program.

Demographic Variables

The demographic information brings insights on improving satisfaction. The more we know the respondent, the better we can approach dissatisfied students to improve the quality of the program. The wage level, marital status, ethnicity, and employment conditions were disregarded because it could pose psychological discomfort.

Demographics	Category
Gender	▪Male ▪Female ▪Other ▪Prefer not to answer
Age	▪20-25 ▪25-30 ▪30-35 ▪35-40 ▪40-45 ▪45-50 ▪50+ ▪Prefer not to answer

Education level	▪Highschool ▪Bachelor ▪Master ▪PhD or higher ▪Prefer not to answer
Location	▪North America (US, Canada) ▪Central and South America ▪Europe ▪Africa ▪Asia ▪Oceania ▪Prefer not to answer
Area of prior study	▪Social science (Law, Economics, Geography, History, Politics, Psychology, Archeology, and more) related ▪ Liberal art (Literature, Philosophy, Sociology, Creative Arts, Communication, and more) related ▪STEM (Science, Tech, Engineering, Math) related ▪Business (ex: Marketing, Sales, corporate HR and more) related ▪IT and computer science related ▪Education related ▪Biological and Biomedical, Health profession related ▪Prefer not to answer
Dependents (except spouse)	▪Yes ▪No ▪Prefer not to answer
Currently employed?	▪Yes ▪No ▪Prefer not to answer

Table 3: Demographic Variables

Sampling Strategy

Our sampling strategy relied on social networking sites for the iMBA program and personal contacts from team members. Along with our collection methodology, we have crafted our questions to identify relevant information that allowed us to segment our data into usable and reliable information to address biases in the sampling strategy.

The main source of survey responses came from the Workplace social network provided by the iMBA program. Everyone with in the program has an account on this social networking platform. The platform does not mix with other programs in the University ensuring that we are only collecting data from our intended experimental unit. The network familiarity with these types of surveys provided incentive for the respondents to take the survey with good intent and supply honest, thoughtful feedback.

Statistical Results

The team collected 97 surveys in a week. The survey data was in exported in CSV format from Qualtrics and fed into R for statistical analysis. Regarding methodology for analysis, we reviewed the distributions of demographic variables among the respondents, ran t-tests and ANOVAs to see their impact to the NPS at 95% confidence.

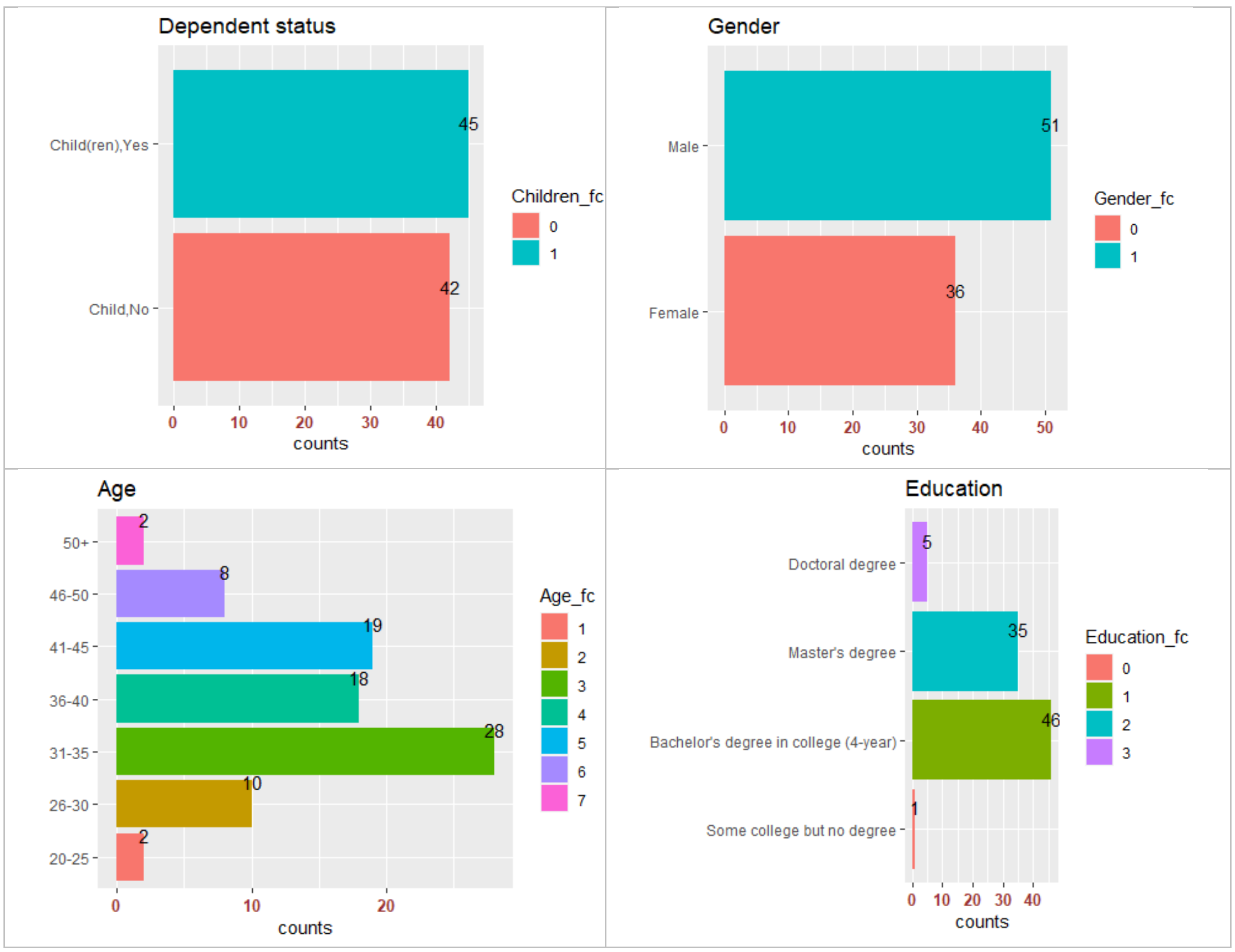
We ran a t-test for demographic variables with two options, such as dependent status and gender. On variables that had multiple segments, we analyzed “Area of prior study” using ANOVA, and

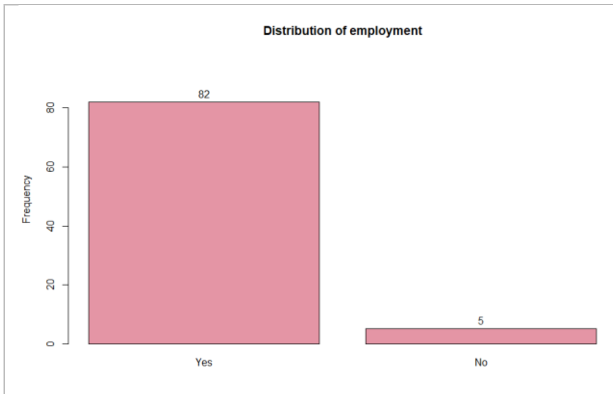
Education and Age together in a multiplicative ANOVA. Almost all respondents were employed and in North America, so we did not run statistical tests on those variables.

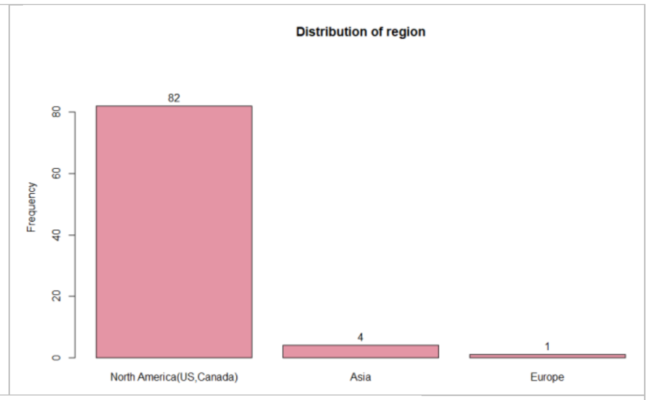
With program features (e.g. tuition cost), we ran a series of simple linear regressions to find significant explanatory variables. Last, we ran multi-variable linear regressions, starting with all program features, while examining variable p-values and model R-squared. The criteria for feature selection was the p-value via Backward Feature Elimination. We ended up with only three variables in the third and last model.

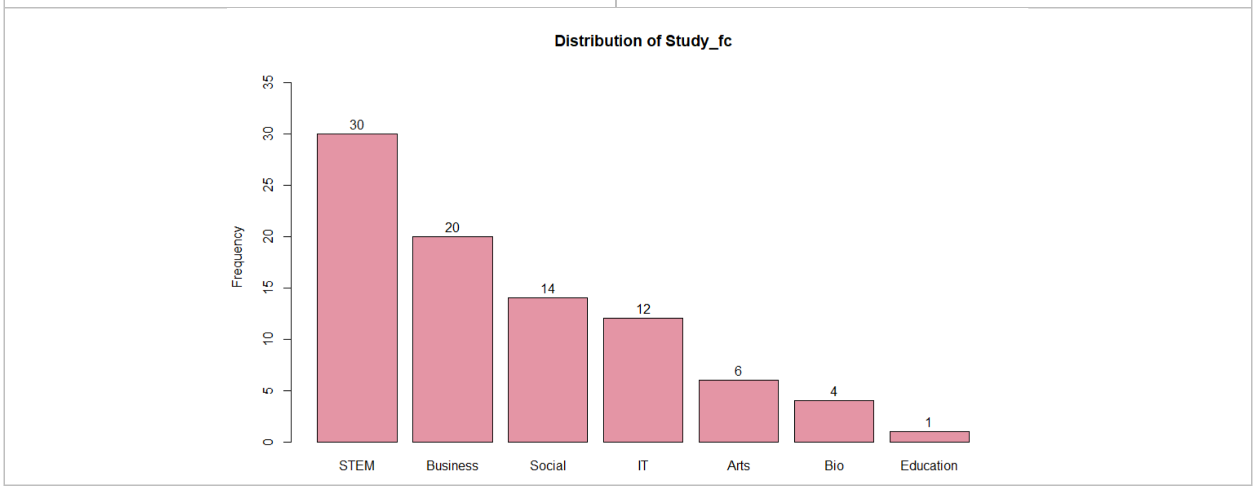
Demographic Variables Summary

Distributions



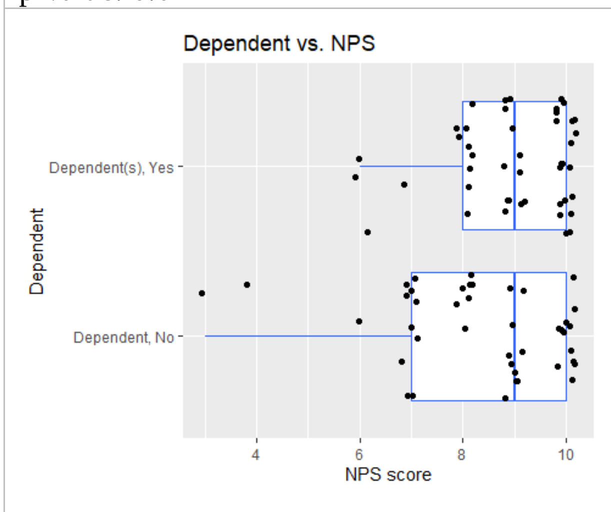
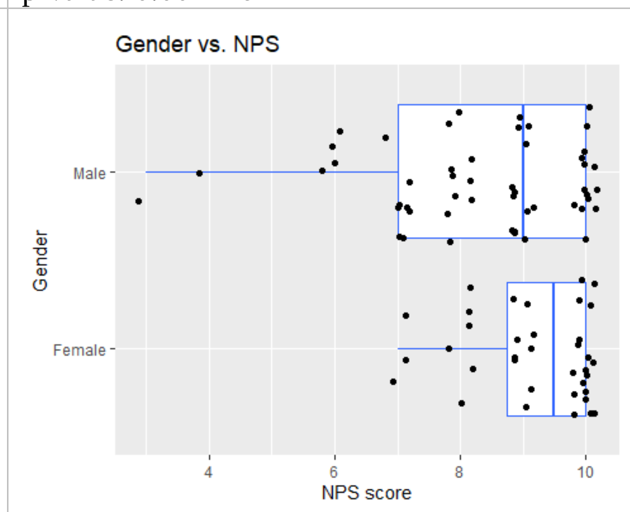






T-tests for Dependent Status and Gender

Statistical tests on dependent status and gender showed that both variables influenced NPS significantly at 95% confidence. The table below shows the t-test results and boxplot for them:

Dependent	**Gender**
p-value: 0.0221	p-value: 0.004126
	

ANOVA for Area of Prior Study

ANOVA for Area of prior study and NPS found that the Area of study did not influence NPS significantly at 95% confidence (p-value = 0.2911).

Multiplicative ANOVA for Education and Age

A multiplicative ANOVA was run to see whether education, age, and their interaction influence NPS. We found that neither variable nor their interaction affects NPS significantly at 95% confidence.

	p-value
education	0.1028128
age	0.6033702
education + age	0.1440985

Independent Variables (Program Features)

Individual Linear Regressions

After t-testing and running ANOVA on demographic variables, we ran linear regression for each independent variable (program features) to see what variables could explain NPS at a 95% confidence interval. The table below summarizes the results we got for each individual model:

Analysis	**p-value**	**R-squared**
Networking	5.18e-07	0.2552
Easiness	0.000353	0.1387
Faculty	3.1e-06	0.2246
Curriculum	7.84e-12	0.4215
Tuition	0.0007664	0.124

All variables had statistical significance on explaining the NPS when they were the single variable in a linear regression. With highest model R-squared, curriculum showed to be the variable that most explained the variance on NPS individually, followed by networking and faculty.

Multi-variable Linear Regressions

The previous individual regressions gave us insights into which explanatory variables could explain NPS. We then ran a multi-variable linear regression with all the independent variables above.

Model Fitting Trial #1

With all variables, the *faculty* and *easiness* showed not be significant for the model:

	p-value

Faculty	0.6449797
Curriculum	1.7038906×10^{-4} ***
Easiness	0.6466488
Tuition	0.0292951 *
Networking	0.0098461 **
R-squared	0.5067885

Model Fitting Trial #2

Re-running the model excluding *easiness*, since it has the highest p-value above (Backward Feature Elimination), the result showed the *faculty* was not significant to explain NPS in the 4 variables model.

	p-value
Faculty	0.6796889
Curriculum	1.4412179×10^{-5} ***
Tuition	0.0280843 *
Networking	0.0081088 **
R-squared	0.5054993

Model Fitting Trial #3

The final linear regression model excluding *faculty* is shown below:

	coefficient	**p-value**
Intercept	6.2158639	8.493906×10^{-29} ***
Curriculum	0.7869036	4.0144155×10^{-7} ***
Tuition	0.3239188	0.0140474 *
Networking	0.2537758	0.0067494 **
R-squared	0.5044638	

Findings

In terms of demographics of our sample, we found that most students came from a STEM background (~34%). If we add that to students from Information Technology, almost half of the respondents were from some STEM background, which is more than double the number of students coming from a Business background. Nevertheless, ANOVA on this variable did not show significant impact on NPS. Surprisingly, the level of education and the students age does not explain NPS significantly in our sample.

Additional analysis of how gender and dependent status of respondents impacted the NPS showed that both variables are related to significance variance of satisfaction, which can be a key indicator for program improvements to specific populations. For example, male students tend to be less satisfied with the program. For dependent status, satisfaction of students with no dependents is skewed to the left, possibly showing that the iMBA program flexibility might be less important to them.

When analyzing program features (independent variables) together, we found that the satisfaction with the specializations and course curriculum, tuition cost, and networking to be the most important drivers of overall satisfaction, when utilizing the variables that we chose. However, the final multi-variable regression model still only explains 50.4% (R-squared) of variance in NPS, showing that there is still room for further analysis.

Practical Recommendations

Based on the findings from our survey statistical analysis, we recommend the following:

- Further analysis is warranted to understand why male students are less satisfied with the program.
- Students with dependents showed on average to be more satisfied with the program. This could be used for targeted advertising to highlight the flexibility of the program.
- In our latest model, curriculum has the highest impact on satisfaction with very high confidence. Only 25% of students are "very satisfied" with the curriculum. Therefore, further analysis to understand how the program could improve the course offerings is recommended.
- Hold more student and alumni events to build stronger relationships and expand networking platforms. Evaluate the impact of the new GiesLink platform on networking satisfaction.
- Tuition cost satisfaction is already high (median score is "very satisfied"). Flexible financing programs could address satisfaction with outliers.

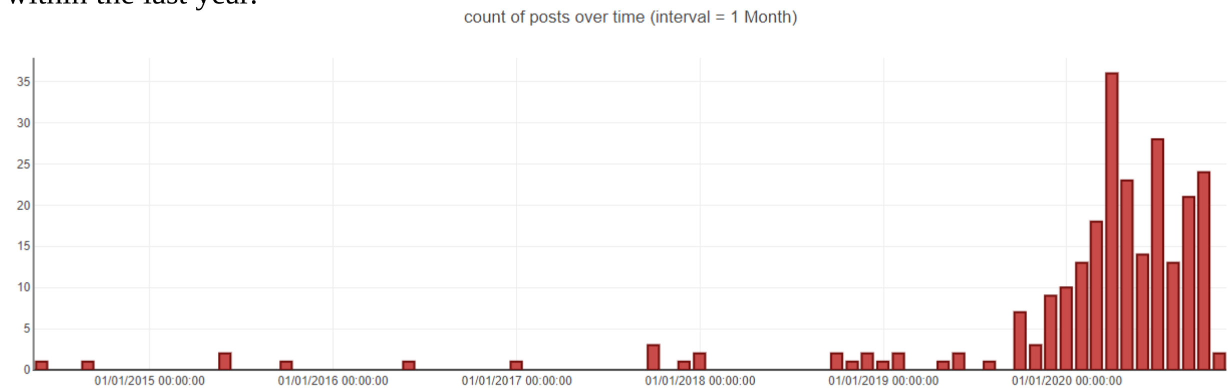
Social Data Text Analytics Findings/Discussion

To analyze the sentiment of the Illinois iMBA program further, the team collected additional social media data using the Social Media Intelligence and Learning Environment "SMILE" on the Social Media Macroscope (Yun et al., 2019). SMILE is an open source social media analytics environment. The team used two sentiment analysis algorithms. The first, Vader, is an open-sourced lexicon and rule-based social media sentiment analysis tool (Hutto, 2014). The second, sentiWordNet, is an opinion mining tool which pairs information retrieval and computational linguistics to get at the opinion a document expresses (Baccianella, 2010).

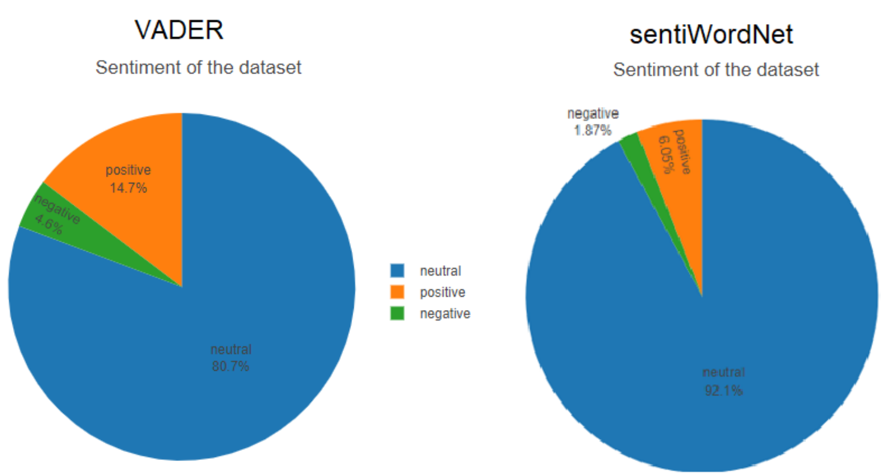
Reddit Mining Results and Findings

Due to a higher volume of mentions, the team chose to leverage the API for Reddit, which pulled 246 posts across the top 33 subreddits using the search teams and booleans ["Illinois" AND "iMBA"].

Despite the University of Illinois offering the iMBA program since 2015, Reddit posts across the top 33 subreddits regarding the Illinois iMBA program have only began increasing dramatically within the last year.



The sentiment of the dataset pulled by Vader shows that 80.7% of reddit posts made regarding the iMBA are neutral, 14.7% are positive, and 4.6% are negative. The sentiment analysis dataset pulled by sentiWordNet showed 92.1% neutral posts, 6.05% positive posts, and 1.87% negative posts.



At its peak, the Illinois iMBA was posted about 35 times over the span of a month on Reddit. Reddit is the world's third largest website with over 330 million active monthly users (Teams, 2020). Although awareness of the University of Illinois' iMBA program has grown since its inception in 2005, awareness of the program is still low based on Reddit engagement related to the program.

In addition, despite the high satisfaction ratings and high levels of survey takers identifying that they would recommend the program to others, the vast majority of reddit posts about the Illinois iMBA were neutral in nature. As identified through the survey results, the Illinois iMBA program offers many differentiating benefits such as low cost and high flexibility. These differentiators are not being spoken to in high volume within the online Reddit community, therefore, the team feels like insights from these neutral posts are just as critically important to understand as both the positive and neutral posts.

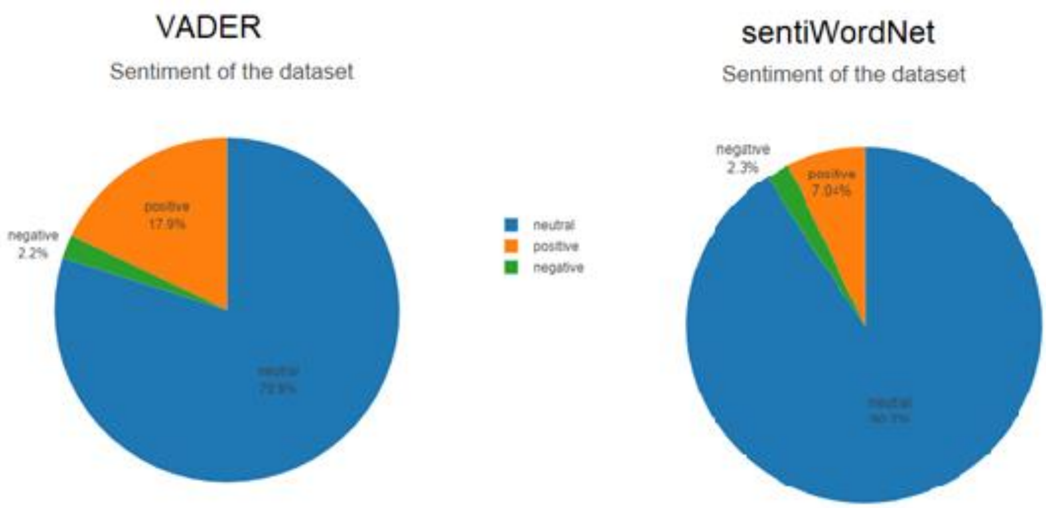
The team also notes that negative sentiment was low for both the Vader and sentiWordNet algorithms, 4.6% and 1.87% respectively. This is a positive indicator for continued growth of the program.

Twitter Mining Results and Findings

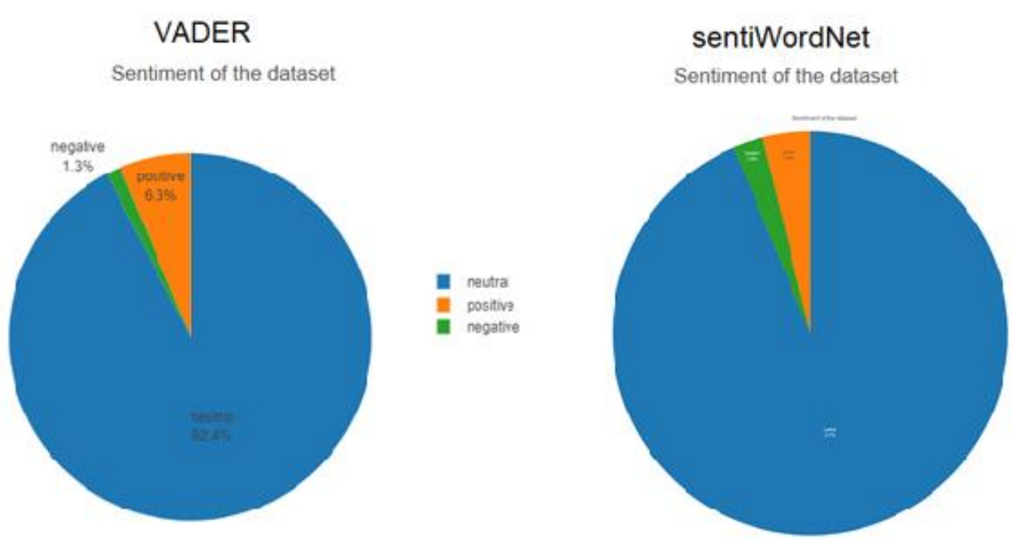
The term we chose to get a representation of the iMBA program was "Gies business" since the iMBA brand did not produce a sufficient number of findings through the analytic tool used on Twitter. Additionally, we looked at "Online MBA" to get a view of the aggregate analysis for online MBA programs and "HBS" to compare our results to another MBA program that is consistently ranked by various publications in the top 5 and renowned for its return on investment and opportunity.

We analyzed three different terms that we felt would give us insight into how our program, online MBAs, and Harvard's MBA compared on Twitter. The summary of results is shown below:

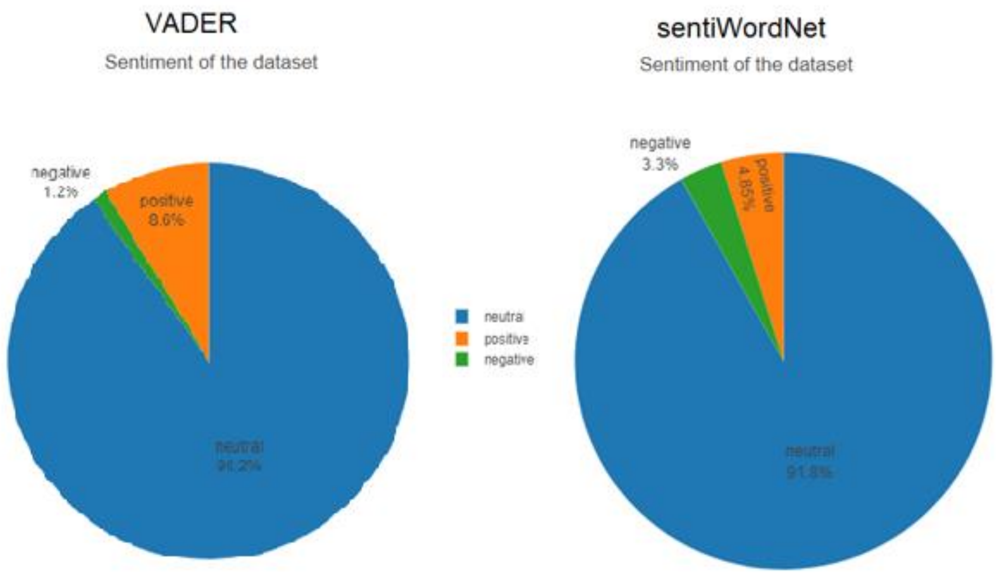
"GIES business"



"OnlineMBA"



“HBS”



The output result of VADER/sentiWordNet shown below show that the sentiment for Gies Business is 17.9% positive, just about 10% better than online MBA’s in total and the HBS sentiment.

While HBS costs of $73,440/year (HBS, 2020), more than 3 times that of the entire tuition for an iMBA through the University of Illinois, the overall positive sentiment for the Gies college of business is far greater. One would also expect that given these expectations and price there would be more positive sentiment. However, the Twitter sentiment was more favorable for Gies and online MBA programs. Also, Gies Business (iMBA) shows more positive sentiment than online MBAs generally.

Another insight that cannot be overlooked is the lack of presence of the branded term “iMBA” as well as the connection of iMBA to Gies business and the University of Illinois. As stated previously most tweets were from Gies itself. This seems to be a big miss from a marketing and branding standpoint as the analysis shows it is program designed for how MBA students today prefer to study and at a price that provides a positive return on investment for most people.

Additionally, the Twitter feeds produced in the sentimental analysis were mainly from the GIES business tweets. It may form a filter bubble (Fletcher, 2020) to view iMBA positively.

Practical Recommendations

To continue the growth and profitability of the Illinois iMBA program, the team has identified two critical recommendations relating to market sentiments and brand awareness.

1. Fully understand the market sentiments to determine the root causes for the positive, negative, and neutral sentiments. Key focus should be given to the neutral market sentiments and neutral sentiments are not indicative of brand loyalty and could ultimately result in a low net promoter score long term as well as positive sentiments to ensure they

are connected to value add items such as ability to network, quality of material, and post-graduate opportunities. One way to get at awareness and neutral sentiment understanding could be through leadership within the program hosting Reddit AMA (ask me anything) forum to spark discussion and debate about the program's many benefits and opportunities. It is recommended that these market sentiments also be understood for iMBA competitors in the industry for benchmarking purposes.

2. Fully understand the root cause of low brand and program awareness. Through understanding and acting on this knowledge, the Illinois iMBA program could reach a broader prospective student audience and better drive decisions when they are differentiating between MBA programs. We recommend that the program make a stronger connection from the branded iMBA to the University of Illinois and Gies Business school via increased social media presence and relevant hashtags, such as including #iMBA to each tweet regarding Gies.

Once the brand awareness, benefits, and reputation are known, market share and incremental sales will likely increase.

Conclusions

In conclusion, thorough analysis and assessment of the University of Illinois' iMBA program through satisfaction surveys of students and post-graduates as well as multiple social media sentiment analyses of key iMBA phrases has shown that the iMBA program is currently delivering on student needs and expectations. However, the team has determined that if the program wishes to continue profitable growth, there are some opportunities the University of Illinois should consider.

First a more nuanced analysis of why male students is seemingly less satisfied is needed to better market to and cater to that demographic. Second, we feel that there is an additional opportunity by making a concerted effort to highlight the satisfaction and flexibility for our students with dependents. Finally, the area we uncovered that would in our opinion have the broadest and most significant impact is to increase the marketing ties from the iMBA program to the overall University of Illinois. Doing this would help the program by taking advantage of the established brand and reputation. A few ways to do this would be by increasing alumni events for better networking results, evolve social media brand & presence, better leverage hashtags, and participate in more online forums.

References

Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. (2010). *Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining*. Lrec. Vol. 10. No. 2010. 2010. Retrieved from http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf

Harvard Business School. (2020). *Annual Cost of Attendance*. HBS.edu. Retrieved from https://www.hbs.edu/mba/financial-aid/tuition-assistance/Pages/cost-of-attendance.aspx

Hutto, C.J. & Gilbert, E.E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media*

(ICWSM-14). Ann Arbor, MI, June 2014. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8109/8122

Fletcher, R. (2020). *The truth behind filter bubbles: Bursting some myths.* Reuter Institute. Retrieved from https://reutersinstitute.politics.ox.ac.uk/risj-review/truth-behind-filter-bubbles-bursting-some-myths

Mulikan, Hollace. (2017) Redefining the Term Networking at HBS.

Goncalves M. S., K. Howell, R., Landis, J. & Lee, S. T*eam Assignment Module 5*. (2020, Sep 15). University of Illinois. Course MBA 564, Applying Analytics Across Business Function.

Goncalves M. S., K; Howell, R.; Landis, J. & Lee, S. *Team Assignment Module 7*. (2020, Sep 29). University of Illinois. Course MBA564, Applying Analytics Across Business Function.

Team, F. (2020, Sep 25). *Reddit Statistics For 2020: Eye-Opening Usage & Traffic Data*. Retrieved from https://foundationinc.co/lab/reddit-statistics/

Qualtrics.com. (2020). *What is Net Promotor Score (NPS)? Definition & examples.* Retrieved from https://www.qualtrics.com/experience-management/customer/net-promoter-score/

Willis, G. (2009). *Cognitive aspects of survey methodology*. In: Lavrakas P, editor. Encyclopedia of Survey Research Methods. Vol 2. Thousand Oaks, CA: Sage Publications; 2009. pp. 103–106.

Yun, J. T., Vance, N., Wang, C., Marini, L., Troy, J., Donelson, C., Chin, C. L. & Henderson, M. D. (2019).*The Social Media Macroscope: A science gateway for research using social media data*. Future Generation Computer Systems. Retrieved from doi:10.1016/j.future.2019.10.029

Statistical Appendix

```
library(dplyr)
library(tidyverse) # Load the tidyverse package
#data read
imba_data<-read.csv("raw_data/iMBA experience_September 222020_1658.csv",stri
ngsAsFactors=FALSE)

imba_data<-na.omit(imba_data)
imba_data<-imba_data[-2:-1,]
imba_data<-imba_data[,18:(ncol(imba_data)-3)]


colnames(imba_data)<- c("tuition","faculty","curriculum",
                        "easiness","network","recommend_likely",
                        "nps","experience","gender",
                        "age","education","region",
                        "study","children","employment")


likert_list = c("very dissatisfied",
"dissatisfied",
"somewhat dissatisfied",
"neither satisfied nor dissatisfied",
"somewhat satisfied",
"satisfied",
"very satisfied")

edu_list = c(
"Some college but no degree",
"Bachelor's degree in college (4-year)",
"Master's degree" ,
"Doctoral degree" ,
"Professional degree (JD, MD)")

age_list = c("20-25",
             "26-30",
             "31-35",
             "36-40",
             "41-45",
             "46-50",
             "50+")


gender_list = c("Female","Male")

child_list =c("Child,No","Child(ren),Yes")

#data cleansing and re-coding
```

```r
imba_data1<-imba_data[- grep("Other|Prefer", imba_data$children),]


imba_data2<-imba_data1[- grep("Other|Prefer", imba_data1$gender),]


imba_data3<-imba_data2[- grep("Other|Prefer", imba_data2$employment),]


imba_data4<-imba_data3[- grep("Other|Prefer", imba_data3$age),]



imba_data5<-imba_data4[- grep("Other|Prefer", imba_data4$education),]



imba_data <- imba_data5[-which(imba_data5$nps == ""), ]



# re-coding for child
imba_data$Children_rc<-recode_factor(imba_data$children, No = 0,Yes = 1)
imba_data$Children_fc<-as.factor(imba_data$Children_rc)

# re-coding for gender
imba_data$Gender_rc<-recode_factor(imba_data$gender, Female = 0,Male = 1)
imba_data$Gender_fc<-as.factor(imba_data$Gender_rc)

# putting likert scale in lower case
for (i in 1:5)
{imba_data[,i]<-tolower(imba_data[,i])}




# education has NA, so omit again # but somehow running command not working
# but manual run for this code chunk works.
imba_data<-na.omit(imba_data)


# re-coding for age
imba_data$Age_rc<-recode(imba_data$age, "20-25"= 1,
                         "26-30"= 2,
                         "31-35"= 3,
                         "36-40"= 4,
                         "41-45"= 5,
                         "46-50"= 6,
                         "50+"   = 7)
```

```r
imba_data$Age_fc<-as.factor(imba_data$Age_rc)

# re-coding for easiness SPECIAL TREATMENT BECAUSE OF TYPO dissastisfied inst
ead of dissatisfied
imba_data$Easiness_rc<-recode(imba_data$easiness, "very dissastisfied" = -3,
                              "dissatisfied" = -2,
                              "somewhat dissatisfied" = -1,
                              "neither satisfied nor dissatisfied" = 0,
                              "somewhat satisfied" = 1,
                              "satisfied" = 2,
                              "very satisfied" = 3)

imba_data$Easiness_fc<-as.factor(imba_data$Easiness_rc)

# re-coding for faculty
imba_data$Faculty_rc<-recode(imba_data$faculty, "very dissatisfied" = -3,
                              "dissatisfied" = -2,
                              "somewhat dissatisfied" = -1,
                              "neither satisfied nor dissatisfied" = 0,
                              "somewhat satisfied" = 1,
                              "satisfied" = 2,
                              "very satisfied" = 3)

imba_data$Faculty_fc<-as.factor(imba_data$Faculty_rc)

imba_data$Tuition_rc<-recode(imba_data$tuition, "very dissatisfied" = -3,
                              "dissatisfied" = -2,
                              "somewhat dissatisfied" = -1,
                              "neither satisfied nor dissatisfied" = 0,
                              "somewhat satisfied" = 1,
                              "satisfied" = 2,
                              "very satisfied" = 3)


# re-coding for tuition
imba_data$Tuition_fc<-as.factor(imba_data$Tuition_rc)


# re-coding for education
imba_data$Education_rc<-recode(imba_data$education,
                               "Some college but no degree" = 0,
                               "Bachelor's degree in college (4-year)" =1,
                               "Master's degree" =2,
                               "Doctoral degree" = 3,
                               "Professional degree (JD, MD)" = 4)


imba_data$Education_fc<-as.factor(imba_data$Education_rc)

# re-coding for networking
```

```
imba_data$Networking_rc<-recode(imba_data$network, "very dissatisfied" = -3,
                              "dissatisfied" = -2,
                              "somewhat dissatisfied" = -1,
                              "neither satisfied nor dissatisfied" = 0,
                              "somewhat satisfied" = 1,
                              "satisfied" = 2,
                              "very satisfied" = 3)

imba_data$Networking_fc<-as.factor(imba_data$Networking_rc)

# re-coding for curriculum
imba_data$Curriculum_rc<-recode(imba_data$curriculum, "very dissatisfied" = -
3,
                              "dissatisfied" = -2,
                              "somewhat dissatisfied" = -1,
                              "neither satisfied nor dissatisfied" = 0,
                              "somewhat satisfied" = 1,
                              "satisfied" = 2,
                              "very satisfied" = 3)

imba_data$Curriculum_fc<-as.factor(imba_data$Curriculum_rc)

imba_data$nps<-as.numeric(imba_data$nps)


# run A/B testing by having children
x<-imba_data$nps[imba_data$Children_rc==1]
y<-imba_data$nps[imba_data$Children_rc==0]

# t test
t.test(x,y,paired=F)

##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 2.257, df = 73.881, p-value = 0.02696
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.08089881 1.30005358
## sample estimates:
## mean of x mean of y
##  9.000000  8.309524

# run A/B testing by gender (I think this should be paired)

rec_female<-imba_data$nps[imba_data$Gender_rc==0]
rec_male<-imba_data$nps[imba_data$Gender_rc==1]
```

```r
# t test
t.test(rec_female,rec_male,paired=F)

##
##  Welch Two Sample t-test
##
## data:  rec_female and rec_male
## t = 3.047, df = 83.864, p-value = 0.00309
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2962653 1.4096170
## sample estimates:
## mean of x mean of y
##  9.166667  8.313725

#run an ANOVA for recommendation and children
AOV1<- aov(nps~Children_rc,data=imba_data)

#summary
summary(AOV1)

##             Df Sum Sq Mean Sq F value Pr(>F)
## Children_rc  1  10.36  10.357    5.21  0.025 *
## Residuals   85 168.98   1.988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#run a multiplicative ANOVA for education and easiness
AOV2<-aov(nps~Education_rc*Age_rc,data= imba_data)
#summary
summary(AOV2)

##                     Df Sum Sq Mean Sq F value Pr(>F)
## Education_rc         1   5.53   5.535   2.721  0.103
## Age_rc               1   0.55   0.553   0.272  0.603
## Education_rc:Age_rc  1   4.42   4.423   2.174  0.144
## Residuals           83 168.82   2.034

# multiple linear regression

fit<- lm(nps~Education_rc,data= imba_data)
summary(fit)

##
## Call:
## lm(formula = nps ~ Education_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4619 -0.8668  0.1332  1.1332  1.5381
##
```

```
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.0569     0.4011  20.087   <2e-16 ***
## Education_rc   0.4050     0.2461   1.645    0.104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.43 on 85 degrees of freedom
## Multiple R-squared:  0.03086,    Adjusted R-squared:  0.01946
## F-statistic: 2.707 on 1 and 85 DF,  p-value: 0.1036

# education level is not related to recommendation

# linear regression by networking
fit_net<-lm(nps~Networking_rc, data=imba_data)
summary(fit_net)

##
## Call:
## lm(formula = nps ~ Networking_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2045 -0.7685  0.1888  0.7101  2.7955
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.24719    0.15655  52.680  < 2e-16 ***
## Networking_rc  0.52135    0.09845   5.296 9.16e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.26 on 85 degrees of freedom
## Multiple R-squared:  0.2481, Adjusted R-squared:  0.2392
## F-statistic: 28.04 on 1 and 85 DF,  p-value: 9.163e-07

# result shows networking is strongly related to recommendation
#plot(fit_net)

# what about easiness?
fit_easy<-lm(nps~Easiness_rc,data=imba_data)
summary(fit_easy)

##
## Call:
## lm(formula = nps ~ Easiness_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0117 -0.8297  0.1703  0.8067  2.9883
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   8.1025     0.2127  38.091  < 2e-16 ***
## Easiness_rc   0.3636     0.1003   3.624 0.000493 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.352 on 85 degrees of freedom
## Multiple R-squared:  0.1338, Adjusted R-squared:  0.1236 
## F-statistic: 13.13 on 1 and 85 DF,  p-value: 0.0004935
```

```
#plot(fit_easy)
# result shows easiness is somewhat related.

#faculty and curriculum

fit_fac_curr<-lm(nps~Faculty_rc+Curriculum_rc,data=imba_data)
summary(fit_fac_curr)
```

```
##
## Call:
## lm(formula = nps ~ Faculty_rc + Curriculum_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.9472 -0.5693  0.1506  0.7417  2.1506 
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)     6.3804     0.3611  17.672  < 2e-16 ***
## Faculty_rc      0.2801     0.1998   1.402    0.165    
## Curriculum_rc   0.9089     0.1630   5.576 2.93e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.104 on 84 degrees of freedom
## Multiple R-squared:  0.4293, Adjusted R-squared:  0.4157 
## F-statistic:  31.6 on 2 and 84 DF,  p-value: 5.868e-11
```

```
fit_fac<-lm(nps~Faculty_rc,data=imba_data)
summary(fit_fac)
```

```
##
## Call:
## lm(formula = nps ~ Faculty_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -4.7002 -0.6242  0.3758  0.4518  2.2998 
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.7762     0.4119  16.450  < 2e-16 ***
## Faculty_rc    0.9240     0.1898   4.869 5.11e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.284 on 85 degrees of freedom
## Multiple R-squared:  0.2181, Adjusted R-squared:  0.2089
## F-statistic: 23.71 on 1 and 85 DF,  p-value: 5.114e-06

fit_curr<-lm(nps~Curriculum_rc,data=imba_data)
summary(fit_curr)

##
## Call:
## lm(formula = nps ~ Curriculum_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8272 -0.7454  0.2137  0.7341  2.2546
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     6.7045     0.2788  24.043  < 2e-16 ***
## Curriculum_rc   1.0409     0.1338   7.781 1.56e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.11 on 85 degrees of freedom
## Multiple R-squared:  0.416,  Adjusted R-squared:  0.4091
## F-statistic: 60.54 on 1 and 85 DF,  p-value: 1.56e-11

# yes, the assumption is right!! and it's positive slope for faculty and curr
iculum

#What about tuition?
fit_tuition<-lm(nps~Tuition_rc,data=imba_data)
summary(fit_tuition)

##
## Call:
## lm(formula = nps ~ Tuition_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0298 -0.9016  0.5343  0.9702  2.0984
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.3374     0.4158  17.645  < 2e-16 ***
## Tuition_rc    0.5641     0.1652   3.414 0.000983 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.362 on 85 degrees of freedom
## Multiple R-squared:  0.1206, Adjusted R-squared:  0.1102
## F-statistic: 11.65 on 1 and 85 DF,  p-value: 0.0009834
```

```
# The best model search in multi linear regression
fit_best<-lm(nps~Faculty_rc+Curriculum_rc+Easiness_rc+Tuition_rc,data=imba_da
ta)
summary(fit_best)
```

```
##
## Call:
## lm(formula = nps ~ Faculty_rc + Curriculum_rc + Easiness_rc +
##     Tuition_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9830 -0.6005  0.0775  0.7601  2.1842
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.92877    0.40624  14.594  < 2e-16 ***
## Faculty_rc     0.14885    0.20809   0.715    0.476
## Curriculum_rc  0.84676    0.18190   4.655 1.23e-05 ***
## Easiness_rc    0.06496    0.09367   0.694    0.490
## Tuition_rc     0.31250    0.14240   2.194    0.031 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.082 on 82 degrees of freedom
## Multiple R-squared:  0.4642, Adjusted R-squared:  0.4381
## F-statistic: 17.76 on 4 and 82 DF,  p-value: 1.548e-10
```

```
# remove easiness
fit_best<-lm(nps~
               Curriculum_rc+Faculty_rc+Tuition_rc+Networking_rc,
             data=imba_data)

summary(fit_best)
```

```
##
## Call:
## lm(formula = nps ~ Curriculum_rc + Faculty_rc + Tuition_rc +
##     Networking_rc, data = imba_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8326 -0.5997  0.0921  0.6717  2.5097
##
```

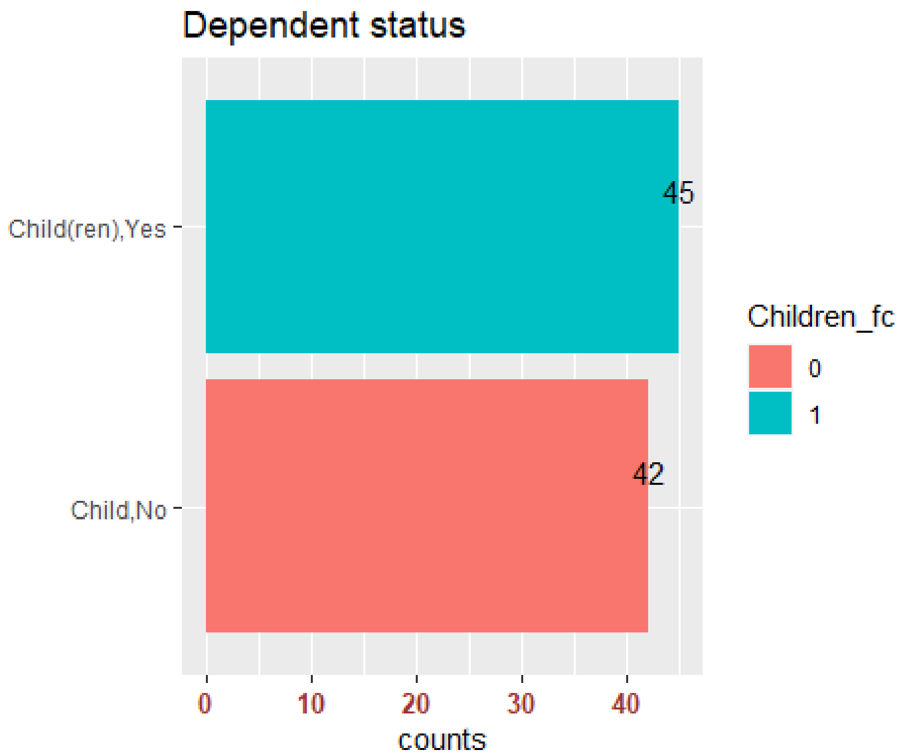
```
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)    
## (Intercept)     6.15294    0.39726  15.488  < 2e-16 ***
## Curriculum_rc   0.75452    0.16353   4.614 1.44e-05 ***
## Faculty_rc      0.08274    0.19967   0.414  0.67969    
## Tuition_rc      0.30591    0.13683   2.236  0.02808 *  
## Networking_rc   0.25019    0.09220   2.714  0.00811 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.04 on 82 degrees of freedom
## Multiple R-squared:  0.5055, Adjusted R-squared:  0.4814 
## F-statistic: 20.96 on 4 and 82 DF,  p-value: 6.278e-12
```

```
#let's remove faculty
fit_best<-lm(nps~Curriculum_rc+Tuition_rc+Networking_rc,data=imba_data)
summary(fit_best)
```

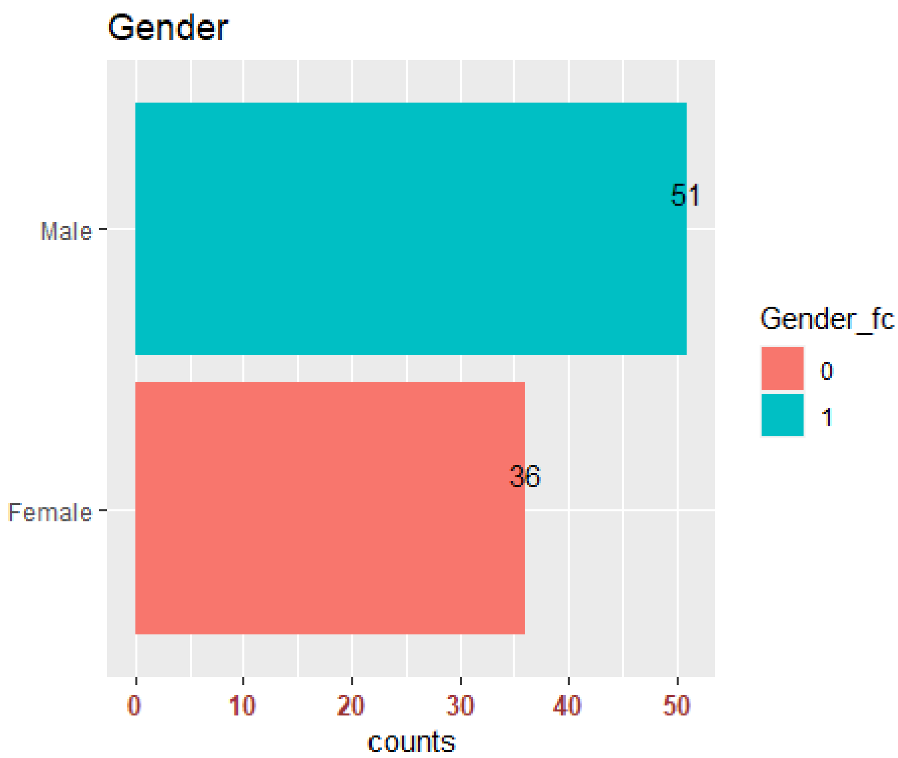
```
## 
## Call:
## lm(formula = nps ~ Curriculum_rc + Tuition_rc + Networking_rc, 
##     data = imba_data)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.8021 -0.6303  0.0956  0.7182  2.5330 
## 
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)    
## (Intercept)     6.21586    0.36525  17.018  < 2e-16 ***
## Curriculum_rc   0.78690    0.14293   5.506 4.01e-07 ***
## Tuition_rc      0.32392    0.12910   2.509  0.01405 *  
## Networking_rc   0.25378    0.09133   2.779  0.00675 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.035 on 83 degrees of freedom
## Multiple R-squared:  0.5045, Adjusted R-squared:  0.4866 
## F-statistic: 28.17 on 3 and 83 DF,  p-value: 1.168e-12
```

7 Appendix #2 (survey visualization)

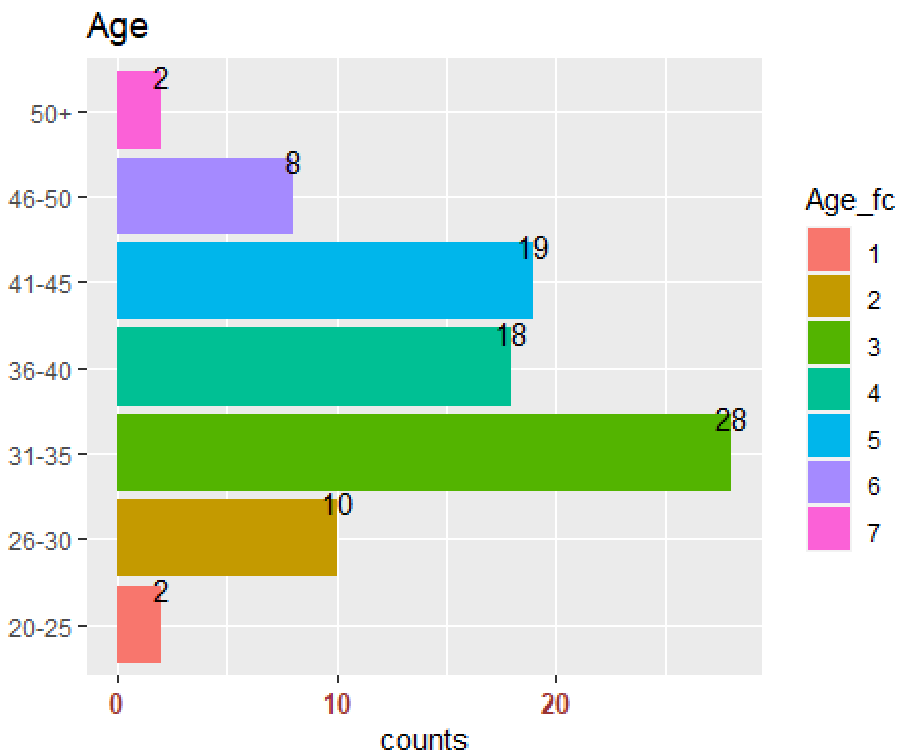
—–	No	Yes
Children	0	1



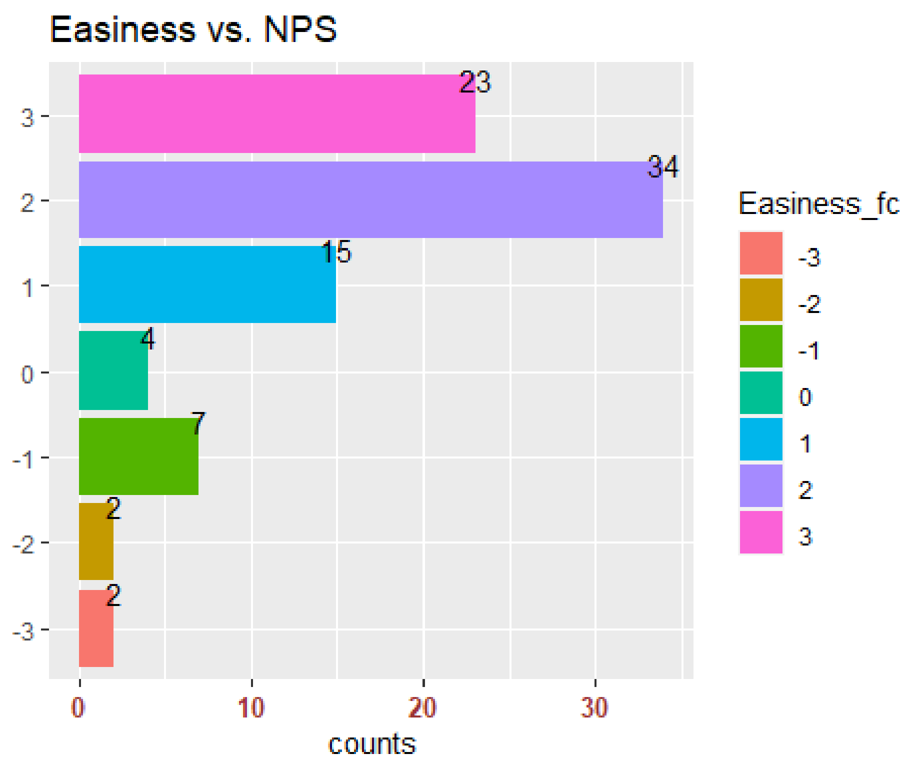
—–	Female	male
Gender	0	1


Gender
Male
Female
51
36
0
10
20
30
40
50
counts
Gender_fc
0
1

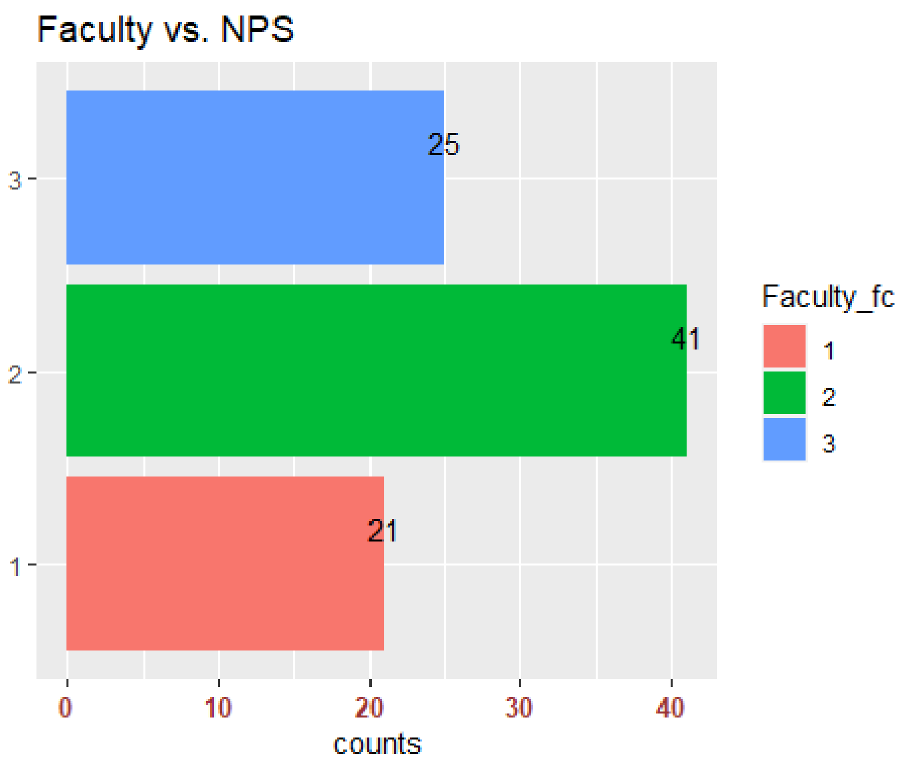
—–	20-25	26-30	31-35	36-40	41-45	46-50	50+
age	1	2	3	4	5	6	7



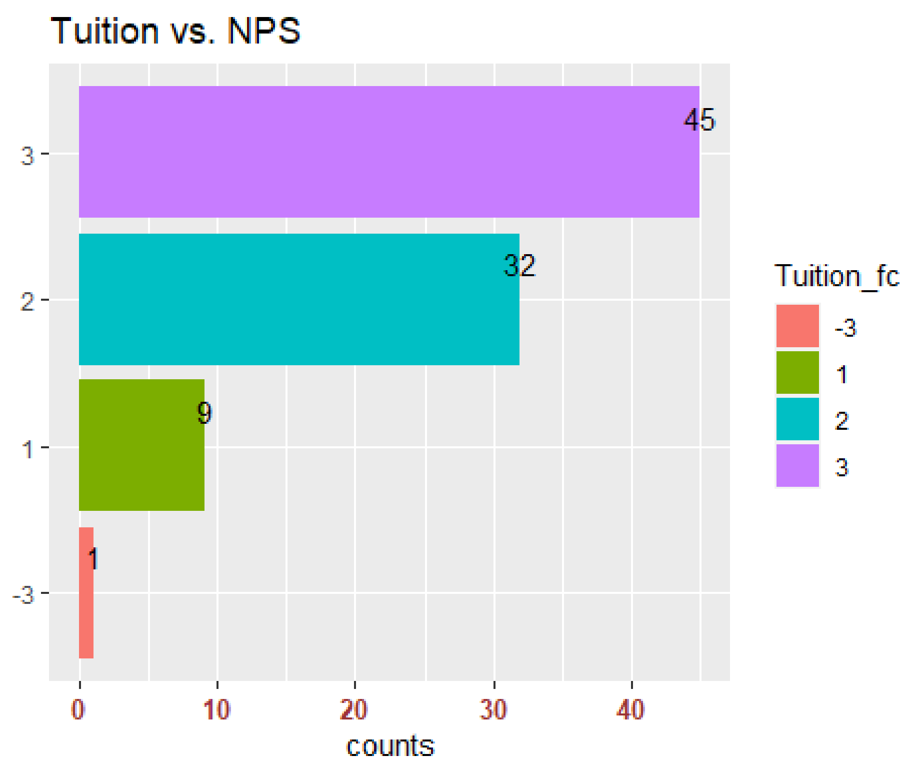
—–	Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Satisfied	Very satisfied
Easy of usage	-3	-2	-1	0	1	2	3



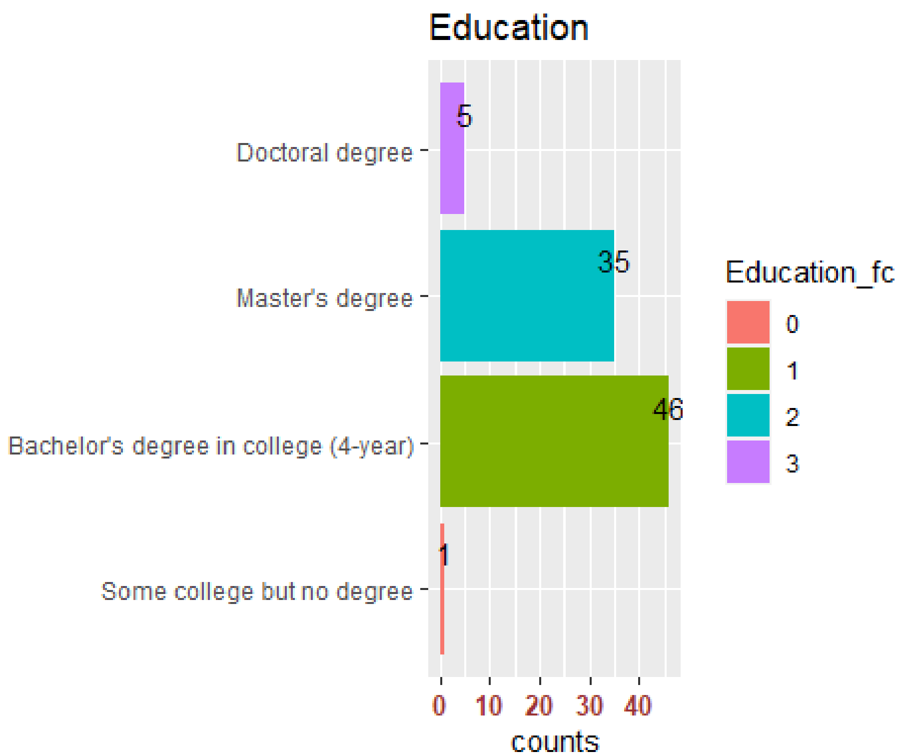
—–	Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Satisfied	Very satisfied
Faculty and staff excellence	-3	-2	-1	0	1	2	3



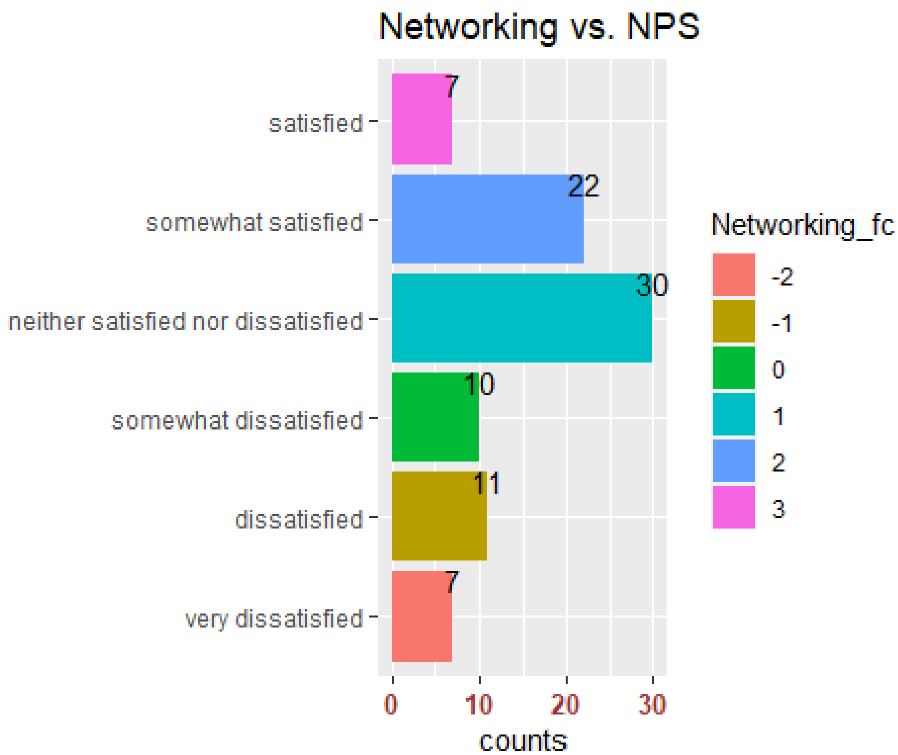
—–	Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Satisfied	Very satisfied
Tuition	-3	-2	-1	0	1	2	3



—–	Some college but no degree	Bachelor’s degree in college (4-year)	Master’s degree	Doctoral degree	Professional degree (JD, MD)	Other/Prefer not to share
Education level	0	1	2	3	4	5



—–	Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Satisfied	Very satisfied
Networking	-3	-2	-1	0	1	2	3



—–	Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Satisfied	Very satisfied
Curriculum	-3	-2	-1	0	1	2	3