Alec Brooks

Eriko Sakamura

DATA 101

26 October 2023

College Rankings: A Personal View

In this lab we are creating a college ranking using personally relevant rankings. The stakeholder we are creating this ranking for is a first generation female student who is interested in a coed diverse student body with cheaper tuition and high retention rate. While creating our weights I organized our four variables into two categories, “required”, and “preferential”. My reasoning for doing this is that some of our metrics seemed more valuable to overall academic success. Namely, retention rate and cost. Whereas, the degree of multiracial first generation students is important but not as important to the overall success of our stakeholder. Although our weights are relatively even I weighed our required two NET_PRICE / RET_FT4 at 0.3 and D150_4_2MOR / FIRST_GEN at 0.2.

Initial Rankings

1 : University of Phoenix-Arizona	2 : University of Arizona
3 : Stanford University	4 : Apex School of Theology
5 : Huntsville Bible College	6 : University of Florida
7 : Talmudical Seminary of Bobov	8 : Uta Mesivta of Kiryas Joel
9 : California State University-Los Angeles	10 : University of Michigan-Ann Arbor
11 : University of Hawaii at Manoa	12 : California State University-Dominguez Hills
13 : California State University-San Bernardino	14 : Princeton University
15 : University of Nevada-Las Vegas	16 : The University of Texas Rio Grande Valley
17 : CUNY Bernard M Baruch College	18 : Inter American University of Puerto Rico-Barranquitas
19 : California State University-Stanislaus	20 : CUNY Lehman College
21 : Rutgers University-New Brunswick	22 : Texas Southmost College
23 : California State University-Fresno	24 : University of Arkansas-Fort Smith
25 : CUNY City College	

Face Validity

Looking at our initial rankings we can see that the University of Phoenix-Arizona is at the top of our list. Taking a closer look we can see that Southwest states namely, Arizona, Texas, and Nevada along with parts of southern California are common on our list as well. To verify the validity of our list we can look at regional metrics and demographics. Although our list has a relatively diverse range of tuition costs ("Average Cost of College by State, EducationData.org"). This seems to be a trade off with graduation retention rates ("Graduation Rates, higheredinfo.org"). Considering we are weighting both cost and retention rate the same it seems reasonable we would find a trend accounting for both.

Comparing Variable Sensitivity

In the pursuit of validating our model we must consider the sensitivity our model has to different variables. We first run our model while eliminating a variable to see how impactful it is on the result of our model. We will do this by examining the mean of our model after excluding the variables FIRST_GEN and then again excluding $NET_PRICE. After the elimination of $FIRST_GEN we see the mean of our new model compared to our old is 634.2598. Whereas when we compare the $NET_PRICE excluded model we see the mean is 581.6156. This indicates that both of our variables are impactful, while the mean of the excluded $FIRST_GEN variable is a bit more impactful than the $NET_PRICE variant.

Model Adjustment	**Mean**
Exclude $FIRST_GEN	634.2598
Exclude $NET_PRICE.	581.6156

Cited Works

Hanson, Melanie. “Average Cost of College by State” EducationData.org, April 26, 2023, https://educationdata.org/average-cost-of-college-by-state

“Graduation Rates” higheredinfo.org, 2018, http://www.higheredinfo.org/dbrowser/index.php?submeasure=27&year=2018&level=nation&mode=graph&state=

Lab Code Block

```
library(readr)
D1 <- read_csv("CollegeDataSmall.csv")

D1<- D1[which(D1$WOMENONLY==0),]

## Weights
#First Generation students
FIRST_GEN_WEIGHT = 0.2
#Retention rate
RET_FT4_WEIGHT = 0.3
#The cost of the school
NET_PRICE_WEIGHT = 0.3
#Proportion of multiracial students
D150_4_2MOR_WEIGHT = 0.2

#model 1
D1$Rank1 <- (FIRST_GEN_WEIGHT)*D1$FIRST_GEN + (RET_FT4_WEIGHT)*D1$RET_FT4_POOLED +
(D150_4_2MOR_WEIGHT)*D1$D150_4_2MOR - (NET_PRICE_WEIGHT)*D1$NET_PRICE
D1R<-D1[order(-D1$Rank1),]
D1R$INSTNM[1:25]

Model1_Index <- D1R[,1]

#model 2 - without $FIRST_GEN
D1$Rank2 <- (RET_FT4_WEIGHT)*D1$RET_FT4_POOLED + (D150_4_2MOR_WEIGHT)*D1$D150_4_2MOR -
(NET_PRICE_WEIGHT)*D1$NET_PRICE
D1R<-D1[order(-D1$Rank2),]
D1R$INSTNM[1:25]

Model2_Index <- D1R[,1]

#model 3 - without $NET_PRICE
D1$Rank3 <- (FIRST_GEN_WEIGHT)*D1$FIRST_GEN + (RET_FT4_WEIGHT)*D1$RET_FT4_POOLED +
(D150_4_2MOR_WEIGHT)*D1$D150_4_2MOR
D1R<-D1[order(-D1$Rank3),]
D1R$INSTNM[1:25]

Model3_Index <- D1R[,1]

#compare mod 1 to mod 2
D1$Dif1_3 <- abs(Model1_Index - Model2_Index)
mean(unlist(D1$Dif1_3))

#compare mod 1 to mod 3
D1$Dif1_4 <- abs(Model1_Index - Model3_Index)
mean(unlist(D1$Dif1_4))
```