Project: University Entrance Exam Analysis

Dol_R

8/24/2020

Contents

Introduction and Key Takeaways

In Turkey, every year millions of students take the university entrance exam. After the announcement of results, participants list their university and department choices and they are placed according to their ranks.

This analysis focuses on universities and departments popularity over the years. Hopefully, it would help future participants in their decision making process.

Key Takeaways:

- We analyzed University Exam Entrance data
- We explored the popularity trends of the universities and departments over the years
- We compared top selected universities and departments

Summary of the Data and Explanations

Using **University Exam** data from *Hacettepe University's Website*, we obtained university results of years 2016-2020. Each year is on a separate Excel spreadsheet. Since every year some departments are opened and some are shut down, datasets will have different number of rows. There are 9 variables and more than 10000 rows for each dataset.

university: Name of the university **city**: University's location **department**: Name of the department **type**: Type of the exam **quota**: Maximum number of participants to be accepted, upper bound for accepted_number **accepted_number**: Accepted number of participants to the selected university's selected department **lowest_score**: Lowest score of the accepted participant's scores **highest_score**: Highest score of the accepted participant's scores **lowest_ranking**: Lowest ranking of the accepted participant's scores, last accepted person's ranking

Objectives:

- Cleaning and manipulation of datasets
- Exploration of the popularity trends of the universities and departments
- Comparison by visualization
- Finding top selected universities and departments

Data Preprocessing

In order to ease the reading process, functions `mani97` and `mani86` are created. These functions remove the columns related to the percentage of change regarding the previous year. They also rename the columns and remove the last empty rows. The raw data for some years have an additional column at the end, therefore two functions were needed.

```
knitr::opts_chunk$set(echo = TRUE) # applies to all chunks
mani97 <- function(data){
  newdata <- data %>%
  rename(
    number = c(1),
    difference = c(2),
    university = c(3),
    city = c(4),
    department = c(5),
    type = c(6),
    quota = c(7),
    difference_quota = c(8),
    accepted_number = c(9),
    lowest_score = c(10),
    highest_score = c(11),
    lowest_ranking = c(12),
    difference_ranking = c(13)
  )  %>% select(-number, -difference, -difference_quota, -difference_ranking) %>% slice_head(n=nrow(data
  return(newdata)
}
mani86 <- function(data){
  newdata <- data %>%
  rename(
    number = c(1),
    difference = c(2),
    university = c(3),
    city = c(4),
    department = c(5),
    type = c(6),
    quota = c(7),
    difference_quota = c(8),
    accepted_number = c(9),
```

```
    lowest_score = c(10),
    highest_score = c(11),
    lowest_ranking = c(12)
  )  %>% select(-number, -difference, -difference_quota) %>% slice_head(n=nrow(data)-9)
  return(newdata)
}
```

After loading the required libraries, we read the .xlsx file from our project repository.

```
library(tidyverse)
library(lubridate)
library(tinytex)
library(readxl)
library(tidyr)
library(httr)
url<-'https://github.com/pjournal/boun01g-dol-r/blob/gh-pages/uni_exam_project/uni_exam.xlsx?raw=true'
GET(url, write_disk(tf <- tempfile(fileext = ".xlsx")))
```

```
raw_df20 <- read_excel(tf, sheet="20", skip=21)
raw_df19 <- read_excel(tf, sheet="19", skip=21)
raw_df18 <- read_excel(tf, sheet="18", skip=21)
raw_df17 <- read_excel(tf, sheet="17", skip=21)
raw_df16 <- read_excel(tf, sheet="16", skip=21)
file.remove(tf)
```

Loading the Libraries and Datasets

We have 5 datasets, each has 9 columns. All have different number of rows. For example there are 10617 rows in dataset for the year 2020.

```
data2020 <- mani97(raw_df20)
data2020 <- mani97(raw_df20)
data2019 <- mani97(raw_df19)
data2018 <- mani86(raw_df18)
data2017 <- mani97(raw_df17)
data2016 <- mani97(raw_df16)

data2020 %>% summarise(exam20=n()) %>% mutate(data2019 %>% summarise(exam19=n())) %>% mutate(data2018 %
```

```
## # A tibble: 1 x 5
##   exam20 exam19 exam18 exam17 exam16
##    <int>  <int>  <int>  <int>  <int>
## 1  10617  11402  11958  11484  10657
```

As an example, let's observe dataset for the year 2020. Four of the variables are categorical and the rest is numerical.

```
data2020 %>% arrange(desc(highest_score)) %>% glimpse()
```

```
## Rows: 10,617
```

```
## Columns: 9
## $ university      <chr> "KOÇ ÜNİVERSİTESİ", "BOĞAZİÇİ ÜNİVERSİTESİ", "BOĞAZ...
## $ city            <chr> "İSTANBUL", "İSTANBUL", "İSTANBUL", "ANKARA", "İSTA...
## $ department      <chr> "Tıp Fakültesi (İngilizce) (Burslu)", "Elektrik-Ele...
## $ type            <chr> "SAY", "SAY", "SAY", "SAY", "SAY", "SAY", "SAY", "S...
## $ quota           <dbl> 14, 82, 88, 45, 8, 10, 15, 45, 20, 72, 15, 14, 103,...
## $ accepted_number <dbl> 14, 82, 88, 45, 8, 10, 15, 45, 20, 72, 15, 14, 103,...
## $ lowest_score    <dbl> 553.8035, 542.2315, 546.3472, 541.4214, 525.6764, 5...
## $ highest_score   <dbl> 571.4231, 567.0095, 564.3493, 562.3470, 561.2876, 5...
## $ lowest_ranking  <dbl> NA, 905, 486, 996, 3940, NA, 579, 797, 204, 1440, 6...
```

Exploratory Data Analysis

Top Selected Universities

Top Selected Departments

Top Cities

More...

Conclusion

References

Hacettepe University's Website.