Data

2024-01-04

# Project’s Data

## Data Source

* YÖK ATLAS database, accessed through the “thestats” package in R.
* Focuses on Industrial Engineering departments at universities in Ankara.
* Includes data on student preferences (1st to 9th choices) and placements (2018-2020).

[Click](https://cran.r-project.org/web/packages/thestats/thestats.pdf) for more information about “thestats” package.

## Data Cleaning

* Woodworking industrial engineering department is removed from our data.
* Preferences that convey the same meaning but are written differently (i.e., ’ (English)’ ‘(English)’) are not included.
* Preferences of students who did not enter with a full scholarship are not included.
* Missing values are handled (NA).
* Columns are renamed for clarity.

**A breakdown of the code:**

**1. Loading Libraries:**

* library(tidyverse): Loads the tidyversepackage for efficient data manipulation.
* library(thestats): Loads the thestats package to access the YÖK ATLAS database.

**2. Loading Data:**

* data <- list\_score(...): Retrieves specific data from the YÖK ATLAS database using the list\_score function:
  + Filters for Industrial Engineering departments in Ankara.
  + Selects variables related to student choices and placements.
  + Employs English language for variable names (lang = "en").

**3. Filtering Data:**

* our\_data <- subset(data, department != "Woodworking Industrial Engineering"): Excludes the “Woodworking Industrial Engineering” department from the dataset.

**4. Handling Missing Values:**

* our\_data <- na.omit(our\_data): Removes any rows with missing values (NA).

**5. Renaming Columns:**

* colnames(our\_data) <- c(...): Assigns more descriptive names to the columns for clarity.

*Now we have a clean dataset called our\_data ready for further analysis to explore our project’s objectives*

You can reach our cleaned data through this link : [Our data](https://github.com/emu-hacettepe-analytics/emu430-fall2023-team-emutrend_explorers/blob/master/ourdata.rda)

## Project Objectives

#### Analyze demand and placement trends:

* Compare demand for Industrial Engineering programs across universities.
* Assess placement success rates in these programs.
* Identify trends in student interest over time.

#### Evaluate competition and differences:

* Analyze competition between universities for Industrial Engineering students.
* Examine differences between programs at different universities.
* Analyze education quality: Compare perceived quality of education based on student preferences.

#### Guide student choices:

* Help future students make informed decisions about which university to choose for Industrial Engineering, considering others’ placement success and preferences.

### Key Data Elements

1. ID’s of each university
2. Year
3. Type of university based on State or Private
4. Program Code
5. Faculty (Engineering Faculty)
6. Department (Industrial Engineering)
7. University Branch (Ankara-based universities)
8. Columns 8-16: Number of times chosen as 1st, 2nd, …, 9th preference
9. Columns 17-25: Number of students placed in order of choice

## Project Potential

* Insight into student preferences and university competition for Industrial Engineering.
* Understanding of factors influencing student choices and placement success.
* Guidance for future students in selecting universities in Industrial Engineering programs in Ankara.
* Potential for further analysis on educational quality, student quality, and social facilities.

## Final Dataset

library(thestats)

## Zorunlu paket yükleniyor: dplyr

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.0  
## ✔ readr 2.1.4

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(gridExtra)

##   
## Attaching package: 'gridExtra'  
##   
## The following object is masked from 'package:dplyr':  
##   
## combine

data <- list\_score(region\_names = "all", city\_names = "Ankara",  
 university\_names = "all", department\_names= "Industrial Engineering",  
 lang = "en", var\_ids=c("X141", "X142", "X143", "X144", "X145", "X146","X147", "X148", "X149",   
 "X151", "X152", "X153", "X154", "X155", "X156", "X157", "X158", "X159"))  
selected\_university <- c("Industrial Engineering (English) (Scholarship)","Industrial Engineering (English)","Industrial Engineering(English)",  
"Industrial Engineering (English) (Scholarship)","   
Industrial Engineering(English)(Scholarship)","Industrial Engineering (Scholarship)","   
Industrial Engineering (English) (Scholarship)","Industrial Engineering(English)(Scholarship)","Industrial Engineering (English) (Scholarship)",  
"Industrial Engineering(English)(Scholarship)","Industrial Engineering (English)","Industrial Engineering(English)","Industrial Engineering (Scholarship)",  
"Industrial Engineering (Scholarship)","Industrial Engineering (English) (Scholarship)","Industrial Engineering (English) (Scholarship)")  
  
  
  
  
our\_data <-na.omit(data)  
  
colnames(our\_data) <- c("ID", "Year","Type","Program Code","University","Faculty","Department","choice\_1st","choice\_2nd",  
 "choice\_3rd","choice\_4th","choice\_5th","choice\_6th",  
 "choice\_7th","choice\_8th","choice\_9th",  
 "placed\_1st", "placed\_2nd", "placed\_3rd",  
 "placed\_4th","placed\_5th","placed\_6th",  
 "placed\_7th","placed\_8th","placed\_9th")  
  
our\_data <- our\_data %>% filter (Department %in% selected\_university)  
print(head(our\_data))

## ID Year Type Program Code University Faculty  
## 1 1041 2020 State 104112108 Gazi University Engineering faculty  
## 2 1041 2019 State 104112108 Gazi University Engineering faculty  
## 3 1041 2018 State 104112108 Gazi University Engineering faculty  
## 4 1048 2020 State 104810529 Hacettepe University Engineering faculty  
## 5 1048 2019 State 104810529 Hacettepe University Engineering faculty  
## 6 1048 2018 State 104810529 Hacettepe University Engineering faculty  
## Department choice\_1st choice\_2nd choice\_3rd choice\_4th  
## 1 Industrial Engineering (English) 57 103 108 102  
## 2 Industrial Engineering(English) 61 82 75 94  
## 3 Industrial Engineering(English) 60 93 116 144  
## 4 Industrial Engineering (English) 176 201 208 154  
## 5 Industrial Engineering(English) 126 115 147 122  
## 6 Industrial Engineering(English) 128 142 212 170  
## choice\_5th choice\_6th choice\_7th choice\_8th choice\_9th placed\_1st placed\_2nd  
## 1 96 105 109 69 74 0 2  
## 2 99 72 101 81 77 1 5  
## 3 142 132 129 117 118 0 0  
## 4 138 162 120 121 110 7 8  
## 5 131 132 115 92 89 3 6  
## 6 175 177 149 153 139 1 0  
## placed\_3rd placed\_4th placed\_5th placed\_6th placed\_7th placed\_8th placed\_9th  
## 1 2 4 3 6 1 2 5  
## 2 0 4 7 2 4 2 2  
## 3 3 3 1 3 1 2 0  
## 4 9 4 8 15 5 3 4  
## 5 7 10 11 8 4 2 1  
## 6 4 4 4 3 6 5 7

# EDA Analysis

In this project, number of choices of Industrial Engineering departments of universities in Ankara and the number of students placed in order of choice between 2018-2020 are provided.

In our dataset, we use industrial engineering data. Therefore, in the “Faculty” column, we have “Engineering Faculty,” and in the “Department” section, we specifically denote “Industrial Engineering.” The university branch is designated as Ankara-based universities in the dataset. The columns from 8 to 16 provide information on how many times the Industrial Engineering department at that university was chosen as the 1st, 2nd, 3rd, and so on, up to the 9th preference. The columns from 17 to 25 give the number of students placed in order of choice.

Through this data set, we will be able to determine which university’s Industrial Engineering department is more preferred, and in conjunction with this parameter, we can understand which university is more favorable for students. Building upon these preferences, analyses can be conducted on the educational quality, student quality, and social facilities of the universities. Additionally, based on the placement preferences of admitted students, an analysis can be performed to understand the students’ preferences for a particular university and assess their level of interest.

In conclusion, this data set fundamentally provides access to a wide range of analytics, allowing a thorough assessment of universities based on the preferences of students majoring in Industrial Engineering. This makes it easier for us to comprehend the educational environment more deeply and to make judgments.