

# Data

## Data Source:

The dataset, sourced from TUIK, covers convicts received into prison across Turkey between 1990 and 2020. It includes information the educational status of the convicts. We chose this dataset because of its potential to unveil the connection between educational background and criminal behavior.

Our objectives include:

We will understand the relationship between education and crime. By analyzing the data set, we will gain insights into how educational attainment influences incarceration. We will develop proficiency in data manipulation and analysis using R. We will learn how to import, clean, transform, and visualize data using R's built-in functions and libraries. We enhance critical thinking and problem-solving skills. We will apply their knowledge of R programming to draw meaningful conclusions from the data.

Our preliminary plans include:

We will import the data set into R and perform any necessary cleaning and data preparation steps. We will visualize the data using various plots and charts to gain an initial understanding of the distribution of variables and the relationships between them. We will interpret the results of the analysis and communicate the findings in a clear and concise manner. We have two main data sources. One of them comes from <https://biruni.tuik.gov.tr/> web site. Thanks to this site we can contact a data base which take necessary command according to data base's variable than turns a file which is excel format. Other one is an excel file which include information about crime types and number of convicts according to crime type. After some arrangement (cleaning, combining etc.) we collected our data in a excel file.

## How Import Data:

We began the analysis by downloading our data from Tuik. We downloaded it as Excel. Later, we made a few adjustments in Excel because our data was very complicated. We have edited it in a way that is more understandable and easy to analyze. For example, we converted "Years" into rows and "Education Status" into columns. It also allowed us to see the missing data more clearly. And it turned into a way to analyze easily.

We began the analysis by downloading our data from Tuik. We installed the "readxl" package. Then we imported the data into R using the "read\_excel" function. We converted the data frame to process our data. Our codes are as follows:

```
library(readxl)
data_criminals <- read_excel("data_criminals.xlsx")
```

Then, we store our final dataset in .RData format. Our codes are as follows:

```
save(data_criminals, file = "data_criminals.RData")  
  
load("data_criminals.RData")  
  
print(data_criminals)
```

[Explicit Link to Our .RData File](#)

## Exploratory Data Analysis:

Our data set consists of 31 rows and 9 columns. Each line represents the years between 1990 and 2020. The columns give information about their educational status.

Educational status are listed as follows:

Illiterate, Literate but not Graduated from a School, Primary School, Primary Education, Junior High School and Vocational School at Junior High School Level, High School and Vocational School at High School Level, Higher Education and Unknown.

There is no data between 1990-2002 in the “Primary Education” column. Likewise, there is no data between 1990-2020 in the “Unknown” column.