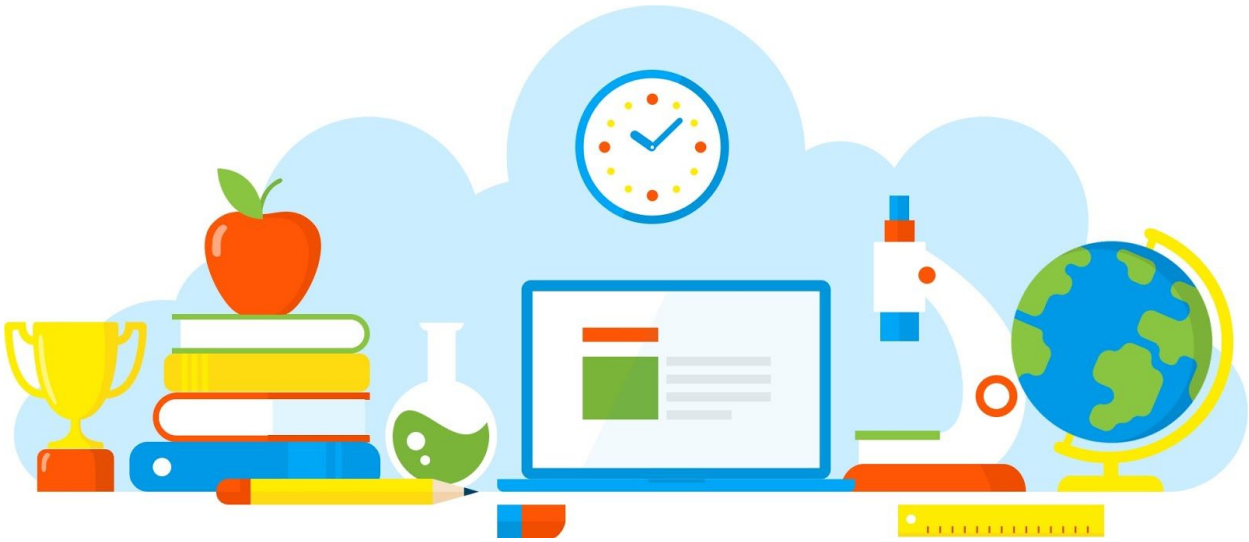

Data Science

Analysing Suicides in India

Aekansh Dixit
PES1201701808 (3C)

Introduction

I decided to analyse the Suicides In India from 2001 to 2012 and inferred some meaningful insights from the data gathered. The data set was downloaded from the internet and underwent cleaning in order to properly understand the analytics. There are both categorical data, as well as numeric data making it a challenge to properly interpret the same.

The Project Details

Dataset: Suicides In India (2001 - 2012)

Programming Language Used: R in RStudio

GitHub Link: Crazytics - Suicides In India

Total Rows: 238,000

Total Columns: 7

State	Year	Type_code	Type	Gender	Age_group	Total
A & N Islar	2001	Causes	Illness (Aid	Female	0-14	0
A & N Islar	2001	Causes	Bankruptc	Female	0-14	0
A & N Islar	2001	Causes	Cancellatic	Female	0-14	0
A & N Islar	2001	Causes	Physical Ab	Female	0-14	0
A & N Islar	2001	Causes	Dowry Dis	Female	0-14	0
A & N Islar	2001	Causes	Family Pro	Female	0-14	0
A & N Islar	2001	Causes	Ideologica	Female	0-14	0
A & N Islar	2001	Causes	Other Prol	Female	0-14	0
A & N Islar	2001	Causes	Property D	Female	0-14	0
A & N Islar	2001	Causes	Fall in Soci	Female	0-14	0
A & N Islar	2001	Causes	Illegitimat	Female	0-14	0
A & N Islar	2001	Causes	Failure in E	Female	0-14	0
A & N Islar	2001	Causes	Insanity/M	Female	0-14	0
A & N Islar	2001	Causes	Love Affai	Female	0-14	1
A & N Islar	2001	Causes	Profession	Female	0-14	0
A & N Islar	2001	Causes	Divorce	Female	0-14	0
A & N Islar	2001	Causes	Drug Abus	Female	0-14	0
A & N Islar	2001	Causes	Not having	Female	0-14	0
A & N Islar	2001	Causes	Causes No	Female	0-14	0
A & N Islar	2001	Causes	Unemploy	Female	0-14	0
A & N Islar	2001	Causes	Other Cau	Female	0-14	1
A & N Islar	2001	Causes	Poverty	Female	0-14	0
A & N Islar	2001	Causes	Death of D	Female	0-14	0
A & N Islar	2001	Causes	Cancer	Female	0-14	0
A & N Islar	2001	Causes	Suspected,	Female	0-14	0
A & N Islar	2001	Causes	Paralysis	Female	0-14	0
A & N Islar	2001	Causes	Property D	Male	0-14	0
A & N Islar	2001	Causes	Unemploy	Male	0-14	0
A & N Islar	2001	Causes	Poverty	Male	0-14	0
A & N Islar	2001	Causes	Family Pro	Male	0-14	0

The Dataset Downloaded in CSV

Gathering and Understanding the Dataset

The dataset can be downloaded from the given link above. Upon opening the **.csv** file, we realise the format of the dataset:

It seems to have 7 columns, namely:

1. **State** in which the data was recorded.
2. **Year** in which the data was recorded.
3. **Type_Code** is the type of data recorded, which can be broken into:
    a. **Causes:** The cause of the suicide.
    b. **Education Status:** The education status of the profile.
    c. **Means Adopted:** How did the profile commit suicide?
    d. **Professional Profile:** What was the professional status of the profile?
    e. **Social Status:** Was the profile married/divorced/Never Married, etc?
4. **Type** is the information of the categorical data presented in **3**.
5. **Gender** is the sex of the profile.
6. **Age Group** is the age group the profile falls under.
    a. **0-14**
    b. **15-29**
    c. **30-44**
    d. **45-59**
    e. **60+**
    f. **0-100+** is the special category for the profiles testing **Education Status** and **Social Status**.
7. **Total** is the number of profiles that suicided under that category.

Going through the dataset, it is clear that a lot of categorical data is clustered and that we need to clean the data in order for us to find some meaningful analysis. The only column that needs to be focused on is **Total**.

Questions to Answer

We'll be looking to answer 3 different questions using this dataset:

1) What is the trend of suicides committed by the females in India over the years?
2) Which age group commits the maximum number of suicides?
3) What is the major cause of suicides in students?

Getting Started

First Steps in R

As already guessed, the first step in R is to import the dataset. Thankfully, R makes it really easy to import CSV datasets. But before we do that, we need to install a few packages that we'll require later:

```
install.packages("dplyr")
install.packages("magrittr")
install.packages("lubridate")
install.packages("zoo")
install.packages("ggplot2")
install.packages("ggpubr")
install.packages("car")
```

Now that we have our packages ready, we're ready to begin. Let's go ahead and import our csv into a dataframe in R.

```
# Import Dataset
dataset <- read.csv("dataset.csv")
```

Now, let's select a few columns and see their listing in RStudio:

```
# Select State (Karnataka), Type, and Total Deaths
dataset %>%
  select(State,Type,Total) %>%
  filter(State == "Karnataka") %>%
  head()
```

This will give us the following output:

	State	Type	Total
1	Karnataka	Insanity/Mental Illness	11
2	Karnataka	Causes Not known	42
3	Karnataka	Property Dispute	0
4	Karnataka	Drug Abuse/Addiction	5
5	Karnataka	Cancer	0
6	Karnataka	Fall in Social Reputation	0