

# Project Proposal

staRstitions - Will Lieber, Wania Iftikhar Khan, AJ Tenser, Kami Akala

```
library(tidyverse)
library(tidymodels)
library(patchwork)
library(ggplot2)
install.packages("psych")
library(psych)
install.packages("readxl")
library(readxl)
library(dplyr)
```

## Introduction

## Data description

```
California_incarceration_rate_data <- read_excel("data/California incarceration rate data.xlsx")
```

```
ZIPCodeData2022_1_ <- read_excel("data/ZIPCodeData2022 (1).xlsx")
```

```
#renaming the columns
```

```
California_incarceration_rate_data <- California_incarceration_rate_data %>%
mutate(zip_code = `California ZIP codes`)
```

```
ZIPCodeData2022_1_ <- ZIPCodeData2022_1_ %>%
mutate(zip_code = `ZIP Code`) %>%
  mutate(zip_code = as.numeric(zip_code))
```

```
joined_data <- full_join(California_incarceration_rate_data, ZIPCodeData2022_1_, by = "zip_code")
```

```
#make sure that the zip codes are not repeated in either dataset
joined_data_clean <- joined_data %>%
  drop_na() %>%
  nrow() %>%
  print()
```

```
[1] 1594
```

```
joined_data %>%
  nrow()
```

```
[1] 1986
```

## Exploratory data analysis

...

## Analysis approach

...

## Data dictionary

The data dictionary can be found [here](#) [Update the link and remove this note!]