## línea horizontal



Racism and Fatal Police Shooting in the US

25/12/2020 to 10/01/2021

**─**

Kapil Dadlani

Data Science

The Bridge

# General Vision

Individual project based on Fatal Police Shooting in the US focused on racism.

# Goals

The main goals are to use all the knowledge we have acquired to be able to get, clean, analyse and visualize data to be able to draw our conclusions and answer our hypothesis

# Specifications

In order to achieve the goals and make the delivery the most, all the requirements needed are specified:

## Software

Visual Studio Code

## Hardware

Processor = Minimum i5

RAM memory = 4GB (although 8GB onwards is recommended)

## Requirements

Programming language: Python 3.9.0

Libraries: Pandas, Numpy, Matplotlib, Seaborn, Plotly

# 

# Steps

**Pie chart representing the time needed for each part of the project:**



## Research the context

I researched if there were police brutality against non-white people in the US, specially fatal police shooting against black people. I also researched which other factor could influence in black people being police target

The hypothesis I have stated:

* Percentage of police shooting in the US is higher in black people.

## Get Data

To get my data I first searched various websites such as Statista, data.world, ourwoldindata, googledatasetsearch, google apis and Kaggle. After quite a few websites, I decided the best data set I found was the one from Kaggle.

## Data Wrangling

Within the data I downloaded, I had quite a few csv files which I had to explore and transform into dataframes using pandas library.

## Data Mining / Clean Data

I cleaned all the dataframes mentioned in the data wrangling step, individually. I changed column title formats so that all titles had the same format in all dataframes in order to be able to merge them and I deleted and added columns so that the dataframes were complete.

I repeated merging processes several times in order to get my final dataframe as I wanted it to be.

I checked for missing values such as NaNs and other non-numerical values that I didn´t need for my analysis. If there were any of those values, I deleted or replaced those rows. The decision of deleting or replacing these values, depended on the amount of missing values in comparison with the whole column (percentage rate). I also changed column types when needed.

## Visualization

Used all the libraries I had learnt to use. Plotly, Seaborn and Matplotlib were the ones I chose. Firstly, I showed different tendencies for each column in my dataset and then made my hypothesis visible in other various charts and graphs.

**Answering the project questions:**

1. **Was it possible to demonstrate the hypothesis? Why?**

I had all the information needed to answer my hypothesis with different visualizations. The hypothesis was correct:

We can conclude that there is a tendency for brutal police shooting in the US towards black male people

1. **What can you conclude about your data study?**

We can conclude that there is a tendency for brutal police shooting in the US towards black people. We can also say that brutal police shooting has been decreasing during the years

1. **What would you change If you need to do another EDA project?**

I would try to get more recent data (the ones I have are from 2015-2017).

1. **What did you learn doing this project?**

I learnt to use libraries I had not used before. I learnt to import other files (visualization.py, mining\_data\_tb.py, etc) into my main files so I can call the functions when needed. I learnt how to solve problems that appeared using google.

# Future Steps

As the course continues, I will be able to improve my project in few aspects.

For now, my thoughts for this project in future:

* Find more information related with this topic so I can analyse deeply racism in the US.
* Apply all the new concepts I will be acquiring in the course in this project.