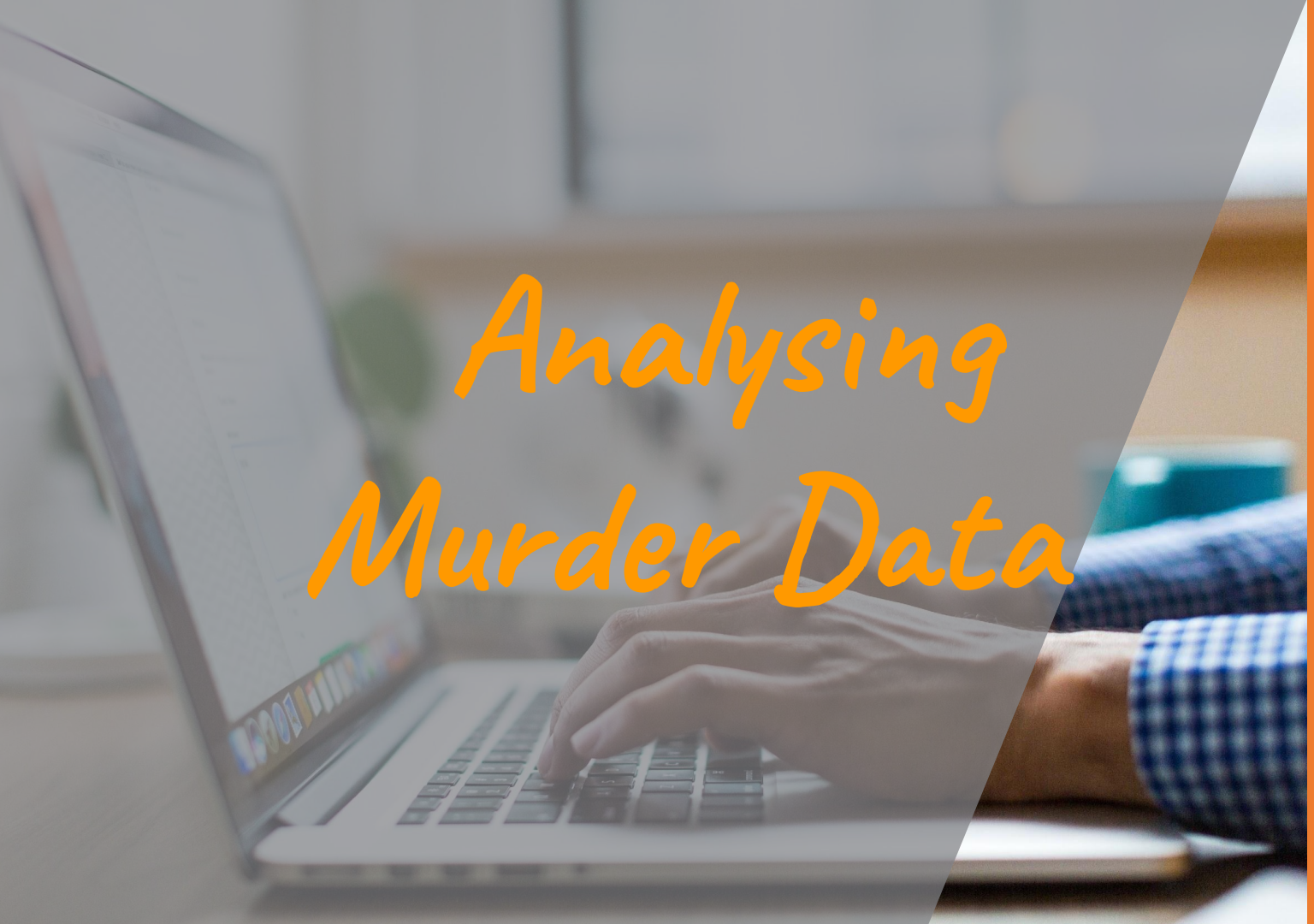

Analysing Murder Data

What were we doing in this project?

So, in this video, we are going to be analyzing the US murder data. This data contains the information about the murder incidents that happened in the United States from 1976 to 2017.

The dataset comes from an online project from the name http://www.murderdata.org/ and also this website gives one very simple algorithm for clustering the informations in the data provided based on the location, weapon used in the crime and the sex of the victim.

So, we will use this data as well as this algo. This algo has been implemented in SPSS but what we are going to do is implement it in the Python.

So, what is the Objective of this Mini Project?

Broadly in this project, we are going to do the following three things:

1. Some exploratory data analysis to make some hypothesis, understand the data and to answer those hypothesis.
2. Implement the murder data algorithms (SPSS one) in python.
3. Then, display the results on US map.

We already did in the last part...

1. EDA (Exploratory Data Analysis).
2. Implementation of the Algorithm.

In this video, we will plot the results..

1. Graph the Results...

To graph the results, the following steps needs to be followed

1. Get the US image.
2. Get the County names for the given top45 counties for largest number of unsolved cases.
3. Get their Latitude and Longitude (Various Counties).
4. Plot them on the image.

Get the US Image...

To get the US image, we will first need the information on the US min and max latitude and longitude. Once we have this, we can use the Export tool OPENSTREETMAP to get the US map image. Let us do that.

County Names

Let us get the county names. It is very simple. To do this what we need to do is just use the dictionary mapper that we created earlier.

Get the Latitude and Longitude of the counties..

To do that, I will use a data called geocode which has the geographic code of various locations. We don't need all of them. We just need top 45 of them. Now, to do that first we will have to match all the counties in the geo data and then pick up the corresponding latitude and longitude.

Fuzzywuzzy string matching for getting county names

To actually match the county names given in top 45 counties with maximum unsolved cases with the geo data that we have, we will use fuzzy matching which matches two string partially and give them a score.

Finally plot the scatter of Lat and long and image on same axis

Now, let us combine all of the above steps to actually create the desired map that we discussed about.

Thank you