**Crime in Los Angeles – Project 2 Data Visualization Dashboard**

Authors: António Pinto, Davide Farinati, Henrique Vaz, Philipp Metzger

1. **Dataset Description**

The main dataset in use in this project contains information on crimes that occurred in Los Angeles between 2010 and 2019. This dataset can be found on Kaggle.com at [1], where a CSV file with 2.12 million records is available for download.

The dataset contains 28 columns that contain information like dates, locations, crime descriptions and information about the crime victims, such as ethnicity, age and gender. The features used for the analysis and visualizations of this projects are: ‘DATE OCC’ (date of crime’s occurrence), ‘AREA NAME’ (the name of the area where it occurred), ‘Crm Cd Desc’ (a description of the occurred crime), ‘Vict Descent’ (the ethnicity of the victim), ‘LAT’, and ‘LON’ (the geographical location of the occurred crime).

Apart from this dataset, the authors also resorted to the Los Angeles Police Department’s website [2] in order to retrieve each area’s population and the square miles area of each area. Since for three areas of Los Angeles (‘Olympic’, ‘West LA’, and ‘Topanga’) this information was not available, the median of the respective variable was used for these three areas.

[Philipp please complete this last paragraph with the cases in which the median was used. Also, the next one should concern the ethnicities data that you used.]

1. **Visualizations and Interaction**
2. Inspiration

The content of the crime dataset used served as the source of inspiration for the visualizations created for this project.

1. KPIs / Big Numbers

In terms of visualizations, the first thing the reader will see are 3/4 numbers that will provide a very high-level perspective of data. Big numbers are a very common practice in dashboards and are used to give the viewer some context before diving into deeper analysis. The information is this visualization will tell where in Los Angeles is there a higher and lower crime incidence and compare the numbers of reported crimes between the selected year and the year before, so that the reader can have an idea of how things evolved.

1. Map of Los Angeles Crime in Selected Year - Choropleth / Map

This graphic displays a spatial heat map, using Los Angeles area, in order to show which areas had more crime incidence. The values in each area can be compared by means of a colour scale/intensity. This is a very handy graphic since it very concretely shows general areas of higher incidence, always considering a specific year selection.

1. Top 10 Most Common Crime Types – Bar Chart

An also very common practice in dashboards, since there is limited space, is to present top analysis using the most significant categories. In this case, from 142 described crime types, the most frequent 10 are show, for a given year. Furthermore, the user is given the possibility of passing the mouse over the bars to find the exact values of occurrences.

1. Line Chart

Describe it

1. Average Number of Reported Crimes per inhabitants per year for the respective area of Los Angeles - Bar Chart

This bar chart combines information on the population, the square miles area and the yearly average crime count of each area in one plot. The user is able to select an area of Los Angeles in the selector above, which results in the respective data to be visualized in the left bar of the two-bar plot. The right bar with the label “Total Los Angeles” serves as an object of comparison to give the user an insight on the magnitude of the crime rate in the respective area compared to Los Angeles as a whole. A two-bar plot was chosen for this visualization, because it is the most effective way to provide the described comparison to the viewer.

1. Victim’s Ethnicities Distribution for the respective area of Los Angeles - Donut Plot

This donut plot visualizes the ethnicity distribution of the crime victims of the area that is selected by the user using the same selector as above. For a better visual result only victim’s ethnicities that make up at least 1% of the occurred crimes are displayed as a separate category. The others are aggregated in the category “Other”. A donut plot was chosen, because it is able to transport the same amount of information as a pie chart, whilst using significantly less ink due to its whole in the middle.

1. Interactions

Besides graphics, the author also implemented two filtering options. The first one is a year slider, in which the user is allowed to select the year of interest to conduct his analysis. The year can be select by sliding within the given year range.

The second filtering option regards the neighborhood/area. Following the same logic, but in this case the user makes his selection using a drop-down menu with 21 different options. With this filter the user can perform comparisons between different areas and look for evidence or correlation between crime and area’s census data.(not sure if this is being too hard like implying causality)

1. **Discussion**
2. **References**

[1] https://www.kaggle.com/chaitanyakck/crime-data-from-2020-to-present/version/5?select=Crime\_Data\_from\_2010\_to\_2019.csv