Information Visualization lab

Second practical work

1 Introduction

We will continue working with the dataset with information about collisions in New York City, the clean version you used for the Lab1 project, but this time, we will work **only with year 2018**. In this practical work, we assume no cleaning is needed (though some data reduction and data augmentation may be required).

This time, our goal is to create an interactive visualization with multiple views that allow users to answer the following questions:

- Which weather condition and type of vehicle were present in the majority of accidents each month? And in the combination of all the months?

- In which area and at what hour did the majority of accidents each month happen? And in the combination of all the months?

- Which area presented the majority of taxi accidents during rainy days in June on Mondays at noon, 12am?

- Which day had more accidents during clear days in July in Manhattan?

We do not want repeated charts (Ex. having several charts for the first question, one per month, and one for the combination).

For "type of vehicle" we are going to specify the categories to use. For "weather conditions" we are going to provide you with the dataset to use.

You may add extra questions. Remember to create DataFrames with the specific columns that you want to use to avoid disconnection problems in the Colab environment.

The visualization must be a multi-view visualization. Think carefully about the design. Test and redesign. The final combination has to be visualized with Streamlit.

2 Data

Collisions dataset and NY map: From Lab1.
Weather dataset: Download *weather2018.csv*. Use the column *icon* as weather condition.

Type of vehicle (From column: vehicle type code 1):
TAXI = 'Taxi'
FIRE = 'Fire','FD tr','firet','fire','FIRE','fd tr','FD TR', and 'FIRET'
AMBULANCE = 'AMBUL','Ambulance','ambul','AMB','Ambul', 'AMBULANCE', and'AMBU'

3 Design and implementation

Before you start coding anything, you need to think about what visualizations will be provided. Note that the user needs to be able to answer the questions above with a single visualization, that will include multiple views.

Some views will contain several variables, so use visual cues wisely.

Like in the previous delivery, we need to see the design process, we want to understand the process you followed to reach the final visualization.

The final visualization must be designed with the tools from Streamlit.

4 Delivery instructions

You must include a step-by-step description of how to solve tasks.

The delivery must consist of a single ZIP file with a name that includes the authors, that contains the datasets (raw and clean), the Colab file(s) (ipnyb), the Python file that contains the streamlit code, and optional extra documents if required. The Colab file must be named after the names of the authors. Treat the Colab document as a report, including titles, boldfaces, etc., to make it easier to read.

The deadline for the delivery of this lab project is the **29th of December**.