Project #3 – Data Visualization Project
Big Data Analysis

Project Description

It is time for the Data Visualization Project. You will work with a large data set of your own choice and analyse the heck out of it. This project is a perfect opportunity to enrich your code portfolio with a proper Data Analysis & Visualization Project.

Project Goals

- Grow autonomy in efficiently analysing & visualizing big data sets with code
- Practise effective communication of the value of your analysis with storytelling

Project Requirements

The following requirements are **mandatory**:

- The data set must have a minimum of 10 000 rows and 10 columns
  - Present your chosen data to us before finalizing your choice
- Continuously use Trello Board for project management
- Conduct an Exploratory Data Analysis (**EDA**) that results in:
  - A descriptive analysis
  - A diagnostic analysis (building upon the descriptive analysis)
- Put the project on your GitHub

Technical Requirements

- Put the data into (one or more) Pandas DataFrame(s)
- Clean your DataFrame(s)
- Slice the DataFrame(s) in interesting ways
- Reshape the data (using groupby + unstack/pivot_table etc.) and perform agg. operations
- Create your own features/columns
- Use statistical tools in your analysis i.e.:
    - Understand what data you are working with through classifying it
    - Correlations, standard deviations etc.
    - Boxplot / handling outliers
    - Distribution types
    - Etc…
- Visualize your findings!

Presentation

The presentation should take max 5 minutes

The slides should include the following (not necessarily in this order):

- Clear description of your data set
    - Visualize!
- Clear description of the question(s) you were exploring
- Clear communication of your main finding(s)
    - Visualize!
- Learnings / highlights/Improvements

- Use Storytelling and e.g. the 5 stages of a story and the other tips we discussed for maximal impact

Schedule

The presentations will take place... teachers announcement

Resources

[Kaggle (for Datasets)](#)

[4 types of data analytics](#)

[Data to Viz](#)

Gustavs slides!

Good Luck!!