Project #3 – Data Visualization Project

Exploratory to Explanatory

Project Description

It is time for the Data Visualization Project. Your task is to perform an Exploratory Data Analysis and turn it into an Explanatory Data Analysis! 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 analyzing & visualizing big data sets with code
- Take your Exploratory Data Analysis and turn it into an Explanatory Data Analysis
- Practise effective communication of the value of your analysis with storytelling

Project Requirements

The following requirements are **mandatory**:

- The data set should have many columns to enable you to do interesting analysis with group-by's, graphs, dissections, etc. It should preferably be large as well with hundreds or thousands of rows, but the number of columns is most important.
    - **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.
- Write comments about what you see in each plot.

- After the EDA, cherry pick the graphs that you want to put emphasis on.
- Use the principles of Aesthetic Plotting to create visually appealing graphs. Don't spend any time on this before you have your Data Storyline.

- 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, e.g. ratios.
- Use statistical tools in your analysis i.e.:
- Choose an Effective Visuals, less is more! Simple text, heatmap, line graph, bar chart - stacked, vertical, horizontal, scatterplots, boxplot, distribution plots.
- Visualize your findings!

Presentation

The presentation should take max 8 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 the Setup-Conflict-Resolution Framework or storytelling with data

Schedule

The presentations will take place… teachers announcement

Resources

Kaggle (for Datasets)

4 types of data analytics

Data to Viz

Good Luck!!