

## General info

[Dataset link](#)

**Submission Form** colab-notebook.

## Part 1: EDA (8 max)

### Evaluation criteria

1. Univariate analysis (consider features separately, their distribution, descriptive statistics, anomalies, omissions, etc.) - 4 points;
2. Multivariate analysis (consider features in pairs, try to introduce new features to consider their relationship with other features, etc.) - 4 points;

## Part 2: DP (6 max)

Prepare data to dimensionality reduction step:

- Fill missing values if needed, explain the selected strategy - 2 points
- Deal with outliers if needed, explain the selected strategy - 2 points
- Prepare categorical data if needed - 2 points

## Part 3: DR (10 max)

### Overview

Use the same dataset you've used during this module. It would be nice if you save the result from the data preparation step and use it in the current task.

The main goal here is to use different dimensionality reduction techniques and show results

### Evaluation criteria

1. at least 3 different dimensionality reductions techniques (SVD, t-SNE, PCA, NMF, UMAP) used - 4 points
2. add visualizations for 3 different dimensionality reductions techniques - 3 points
3. calculate explained variance for 2 different dimensionality reductions techniques - 1 point
4. select the best number of components for at least 1 dimensionality reduction technique - 2 point

## General criteria for whole work (12 max)

1. The presence of informative visualizations (the presence of headings, axis labels, legends, etc.) - 3 points;

2. Explanations and comments regarding data actions performed - 3 points;
3. Active code reuse (helper methods, classes, etc.) -3 points.
4. The presence of a clear structure in the work (introduction, separate sections on the issues under investigation, conclusion, further steps to take) - 3 points;

## Useful links

<https://matplotlib.org/> (Links to an external site.);

<https://seaborn.pydata.org/> (Links to an external site.);

<https://plot.ly/python/> (Links to an external site.);

<https://bokeh.pydata.org/en/latest/docs/gallery.html#gallery> (Links to an external site.);

<https://python-graph-gallery.com/> (Links to an external site.);

<https://www.datacamp.com/courses/introduction-to-data-visualization-with-python> (Links to an external site.);

<https://www.datacamp.com/courses/improving-your-data-visualizations-in-python> (Links to an external site.);

<https://campus.datacamp.com/courses/data-visualization-with-seaborn> (Links to an external site.);

<https://www.dataquest.io/blog/python-data-visualization-libraries/> (Links to an external site.);

<https://www.dataquest.io/course/exploratory-data-visualization/> (Links to an external site.);

<https://dev.to/sakim/python-cheat-sheet-for-beginners-and-experts-37ok> (Links to an external site.);

<http://www.machinelearning.ru/wiki/index.php?title=Crisp-dm> (Links to an external site.).