Datasets

1. [Facebook dataset](https://www.kaggle.com/datasets/mrmorj/political-advertisements-from-facebook): ​​
2. [E-commerce dataset:](https://www.kaggle.com/datasets/lissetteg/ecommerce-dataset)
3. [Health Dataset(diabetes):](https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset)
4. [Energy consumption dataset Helsinki:](https://www.avoindata.fi/data/fi/dataset/helsingin-kaupungin-palvelukiinteistojen-energiankulutustietoja)
5. [Webserver logs:](https://www.kaggle.com/datasets/eliasdabbas/web-server-access-logs)
6. [Predicting the Future Transaction from Large and Imbalanced Banking Dataset:](https://www.kaggle.com/c/santander-customer-transaction-prediction)

Way of work:

1. Every group will have 10 mins presentation:
2. Every group (2 students) has to define between 2-3 research questions for at least 3 datasets
3. Data analysis tasks every sub-group should follow this protocol:
   * Descriptive statistics:
     + Building Summary
     + Calculate Central/Dispersion measures
       - Mean, STD, …
     + Get the distribution of the data (each column)
       - Define the distribution based on the plot
     + Analyze relationships between features
       - Correlation (Heatmap) and Pair plot
   * Exploratory Data Analysis (EDA) for every possible question:
     + Identify the variable and their types
     + Clean your data (error, remove duplicates, missing values, Outliers)
     + Transformation (Standardization, Normalization, encoding categorical to numerical)
     + Data Visualization (use the suitable visualization that you need)
   * Define your research questions/objectives
     + Perform Hypothesis testing
     + Interpret the results findings in the context of your research question or objective. Draw conclusions and make recommendations based on your analysis.
     + Communicate your results: Present your insights and conclusions in a clear and concise manner, using visualizations and descriptive statistics. Tailor your communication to your audience, whether it be technical or non-technical.