**Лабораторна робота № 1**

**Розвідувальний аналіз даних**

**Перед початком виконання**

1. Створити свій репозиторій на github, надати доступи для *onupmarchenko* та *janeplatonova*.
2. Зафіксувати середовище (environment) для виконання ваших робіт. (Див. практика 09.09-13.09, <https://pipenv.pypa.io/en/latest/index.html> ). Середовище фіксується в корені проєкту.
3. Для коду:
   1. створити окрему директорію з номером роботи (*LW1*)
   2. всі додаткові матеріали, котрі стосуються цієї роботи, мають бути розміщені в ній.
4. Використання jupyter notebook для цієї роботи можливе. Проте більше вітається організація коду в .py файлах. <https://ploomber.io/blog/nbs-myths/>  
   <https://www.reddit.com/r/Python/comments/spxim1/notebooks_suck_change_my_mind/>
5. Скачати дані відповідно до вашого варіанту.
6. Відповідний jupyter notebook із завданням розміщений в папці [Notebooks](https://drive.google.com/drive/folders/1M0-EuhGVhd7Xyf8sLhnEujhtCeX7P7RN?usp=sharing).

**Постановка задачі**

1. Ознайомитись з датасетом, визначити цільову змінну (target) та основну задачу прогнозної моделі. (2 бали)
2. Виконати етапи розвідувального аналізу, вказані у блокноті.   
   (4 бали)
3. Виконати творче завдання: спробувати знайти ознаки (змінні, “фічі”), які будуть найбільш корисними для передбачення target-змінної. Використовуйте будь-які доступні інструменти розвідувального аналізу (обчислення, зведені таблиці (pivot\_table, groupby), візуалізація, кореляційний аналіз).   
   (4 бали)

**Варіанти**

* Варіант 1. [Adult Income Prediction](https://www.kaggle.com/datasets/wenruliu/adult-income-dataset)
* Варіант 2. [Telecom Churn Prediction](https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset)
* Варіант 3. [Bank Marketing Dataset](https://www.kaggle.com/datasets/henriqueyamahata/bank-marketing)
* Варіант 4. [Online News Popularity](https://www.kaggle.com/datasets/btphan/online-news-popularity-dataset)