Task:

Perform an EDA (exploratory data analysis) of the dataset. Prepare graphs of your data in plotly.

Build a forecasting model that will predict the future price of used (Używane) iPhones 11 (1 month into the future) and calculate several measures of model performance (e.g., MSE, MAE).

Bear in mind that there are more than one iPhone 11 model and specs. Try to invent an approach that will automatically assign offers to a particular iPhone 11 model based on available data (not only in the dataset).

Prepare a short report or use Jupyter notebook to **present your way of reasoning**.

Dataset:

The dataset consists of the offers that were put on OLX regarding a query “iphone 11” between 01-01-2021 and 24-02-2021. The dataset is as-is on the OLX website. Due to that the offers are mostly in polish, so use polish NLP models.

Data description:

- http – URL address of a particular offer;
- voivodeship – a (polish) voivodeship where the offer is given;
- scrap_time – timestamp when the offer was scraped;
- name – offer’s name;
- price – offer’s price;
- brand - OLX specific field; product’s brand;
- condition – OLX specific field; condition of the offered product;
- offer_from – OLX specific field; whether the seller is a company or private;
- type – OLX specific field; whether a product is working or not;
- description – offer’s main text;
- added_at – timestamp when the offer was added;
- views – how many views the webpage has at the moment of scrapping;
- user_since – timestamp when the seller has created an account.

Requirements:

- Python
- Whole process can be presented in Jupyter notebook

Other information:

Focus mostly on technicalities, but remember that what you produce should be later used by business. Do not try to create ‘the best’ model in terms of accuracy. Try to build a PoC. Remember to carefully comment and explain what you are implementing.

Any questions:

Send any questions to paulina.gniadzik@wesub.pl