# Uber Data Analysis

## Link:

<https://www.kaggle.com/datasets/bhanupratapbiswas/uber-data-analysis>

## Opis:

Databáza obsahuje informácie o jazdách Uber za rok 2016 (ich začiatok, koniec, kategória a miesto, kde sa jazda začína).

## Kod Link:

<https://www.kaggle.com/code/dariagrinenko/uber-data-analysis-and-prediction-of-miles/edit>

## Opis kodu:

Tento kód analyzuje databázu a zobrazuje informácie vo forme grafickej podoby s pomocou knižníc:

* numpy (matematické výpočty)
* pandas (práca s tabuľkovými dátami)
* matplotlib (grafy a vizualizácie)
* scikit-learn (strojové učenie)
* xgboost (gradientný boosting pre predikcie)
* ightgbm (rýchly gradientný boosting pre veľké dáta)
* catboost (gradientný boosting, dobre zvláda kategóriálne dáta)

\*Boosting je technika v strojovom učení, kde sa kombinuje viacero slabých modelov do jedného silného modelu.

O skúške programu:  
Aby som ho mohla vyskúšať, najprv som musela vytvoriť **venv** (ako som pochopila, ide o virtuálne prostredie). Potom som obnovila **pip** na verziu 25.2 (aby nové knižnice vedeli byť stiahnuté) a takisto som potrebovala zmeniť verziu Pythonu na staršiu, ktorá podporuje všetky knižnice. Takisto som v niektorých miestach musela zmeniť cestu súborov, aby sa kód vedel napojiť na databázu. Na zistenie informácií o tom, ako spraviť, aby kód začal fungovať, som použila ChatGPT, ktorý mi taktiež pomohol so slovenskou diakritikou.

+ screenshoty, že kód funguje…



