Uber Price Prediction

For the final project, I will focus on a **regression task** to predict **ride prices at a given timestamp** using data from Uber and additional contextual data. The primary dataset contains information about Uber rides, including attributes such as timestamps, pickup and drop-off locations, and prices. To enhance the predictive power of the model, I will integrate **weather data** (temperature, precipitation, etc.) and national holidays as additional features, as these factors significantly influence ride demand and pricing. I also plan to split the rides by geographical area, allowing the model to capture location-specific patterns.

The project aims to identify key drivers of surge pricing and build predictive models that can accurately estimate the ride price at a given timestamp. This involves leveraging temporal patterns, such as peak and off-peak hours, day of the week, and seasonal trends, while incorporating external factors like adverse weather conditions and holiday periods, which are known to disrupt normal travel behavior.

To achieve this, I will use a range of models. For baseline comparisons, I will start with traditional regression models such as **Linear Regression** and **Ridge Regression** to capture the linear relationships between features and ride prices. Next, I will test some tree-based models, including **Gradient Boosting Models** like XGBoost and LightBoost, which are highly effective at capturing non-linear relationships and interactions between features. Additionally, I plan to experiment with a deep learning approach, such as a **Feedforward Neural Network**.

To evaluate the models, I will use regression metrics such as **Mean Absolute Error (MAE)**, **Root Mean Squared Error (RMSE)**, and **R-squared** to measure the accuracy and explanatory power of the predictions. I also plan to conduct a feature importance analysis to understand the most significant factors influencing ride prices. Finally, to optimize the models I use I plan to grid search hyper-parameters.

## Sources

* Uber prices – [Kaggle](https://www.kaggle.com/datasets/yasserh/uber-fares-dataset)
* Weather – [Meteomatics API](https://www.meteomatics.com/en/api/request/?ppc_keyword=&utm_term=&utm_campaign=Performance+Max+UK&utm_source=adwords&utm_medium=ppc&hsa_acc=5001518620&hsa_cam=21708623642&hsa_grp=&hsa_ad=&hsa_src=x&hsa_tgt=&hsa_kw=&hsa_mt=&hsa_net=adwords&hsa_ver=3&gad_source=1&gbraid=0AAAAADoNVHqM5_FLpjDa8WuIimFwlXuXb&gclid=EAIaIQobChMIpb_U1PGKigMVUmBHAR0qkxhqEAAYASAAEgLo3PD_BwE)