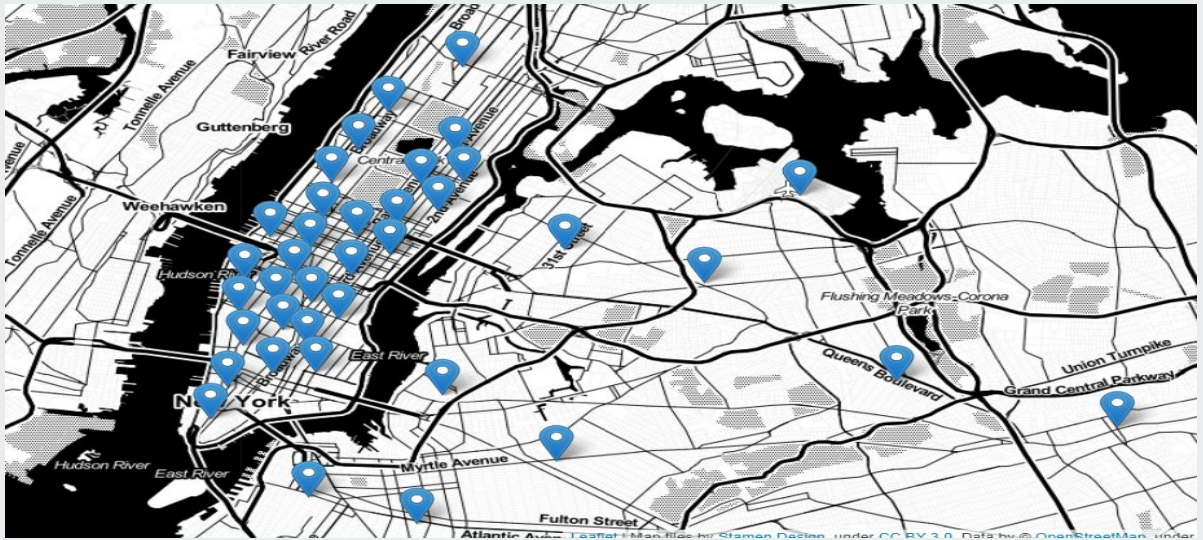
Taxi Demand Prediction



By Eddie Amaitum

Demand in taxi/car share industry

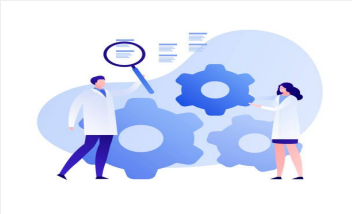
Uber

Drivers (Supply) Vs Customers(Demand)

Companies need to forecast demand

Solution: Predict taxi demand per hour by location

Operations team can adjust the distribution of drivers



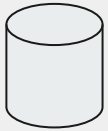
\+ $ 3 for pick-ups around the arena

Operations team

Impact

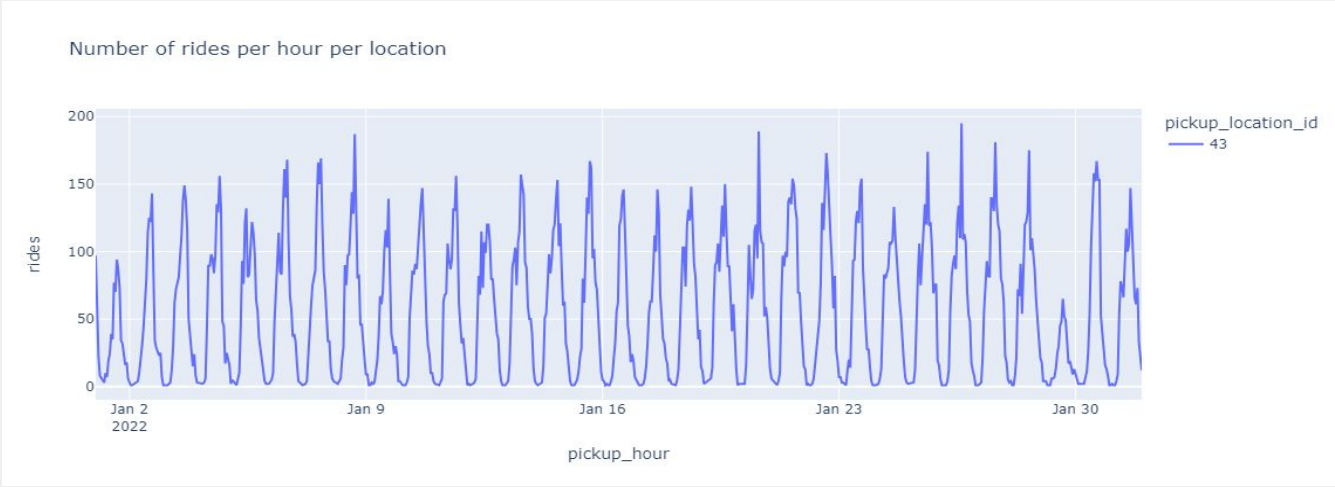
- Predictive models shown to cut wait times by up to 20%
- Improved efficiency in driver deployment hence companies generate more revenue
- Increased customer satisfaction

Dataset

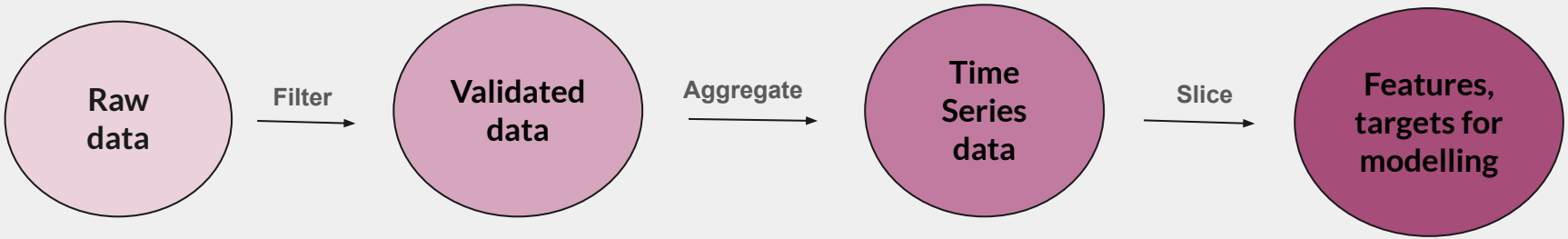


- NYC Taxi & Limousine Commission (TLC) Trip Records
- The data is relatively clean
- Feature engineering needed

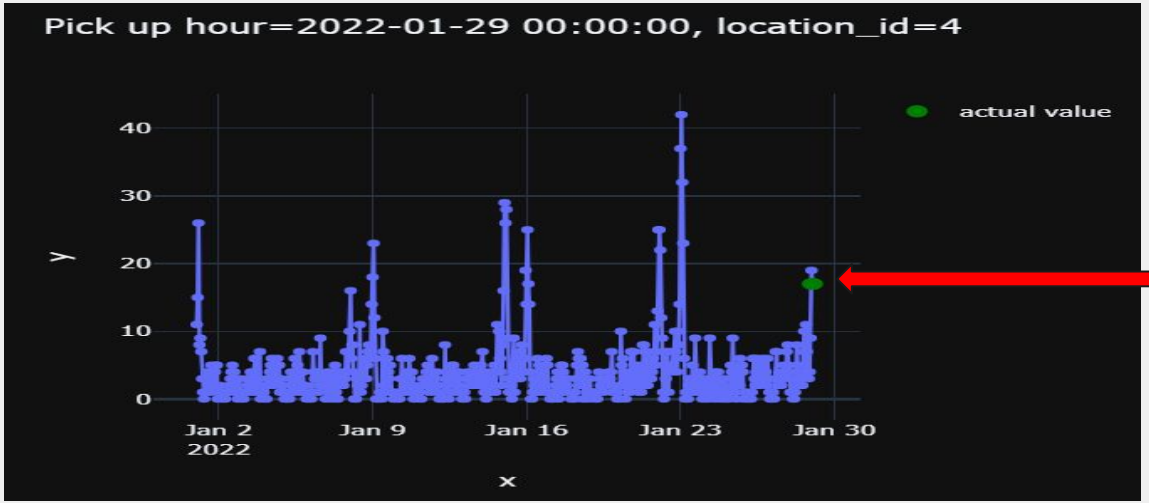
Sample Time Series Data



Data transformation into (features, targets)

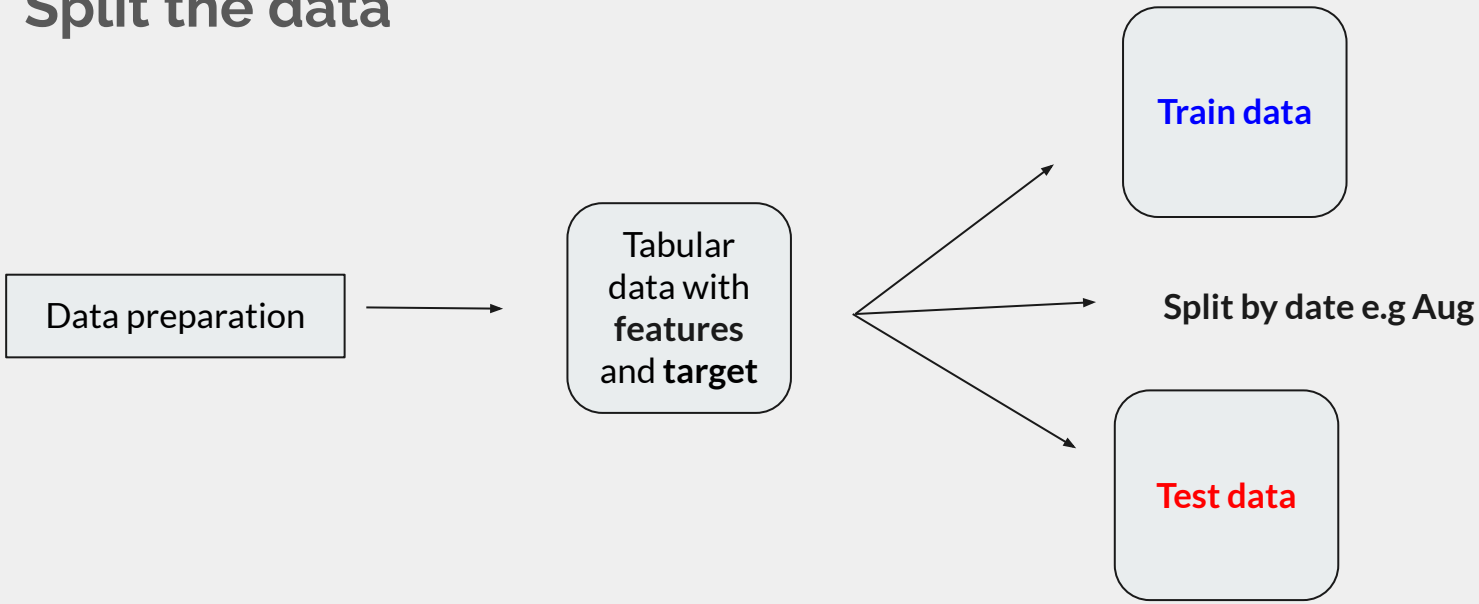


Sample Features & Target Variables

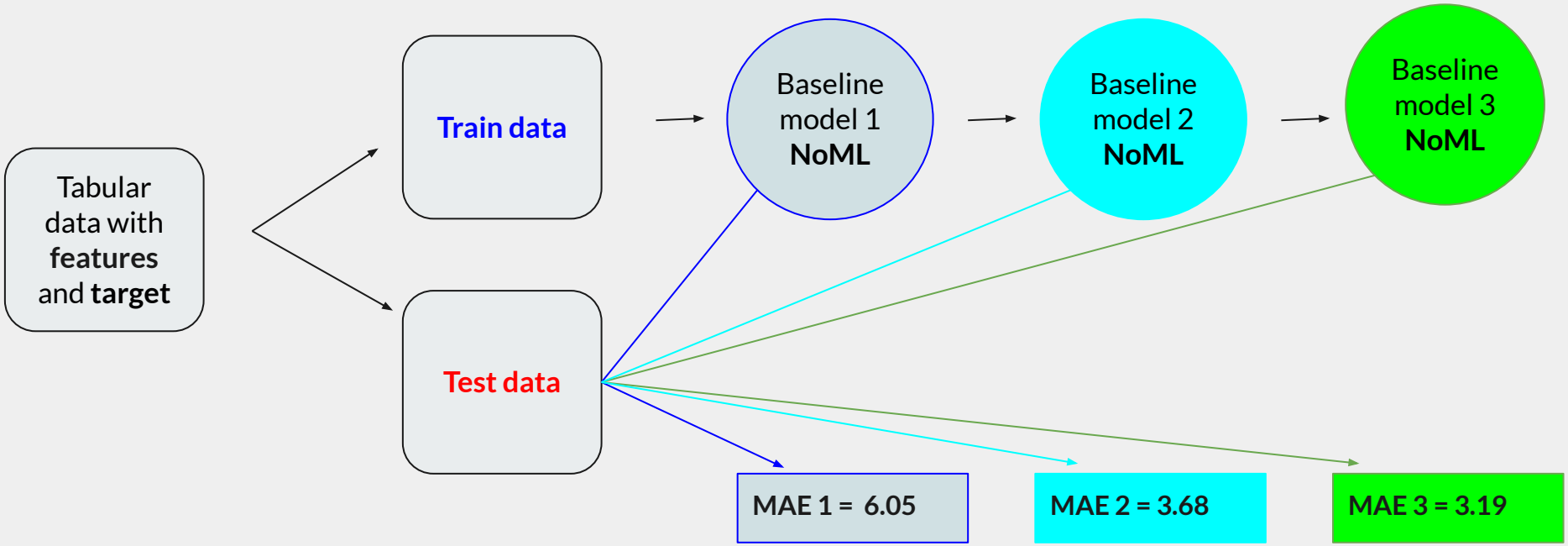


Target

Split the data



Baseline models

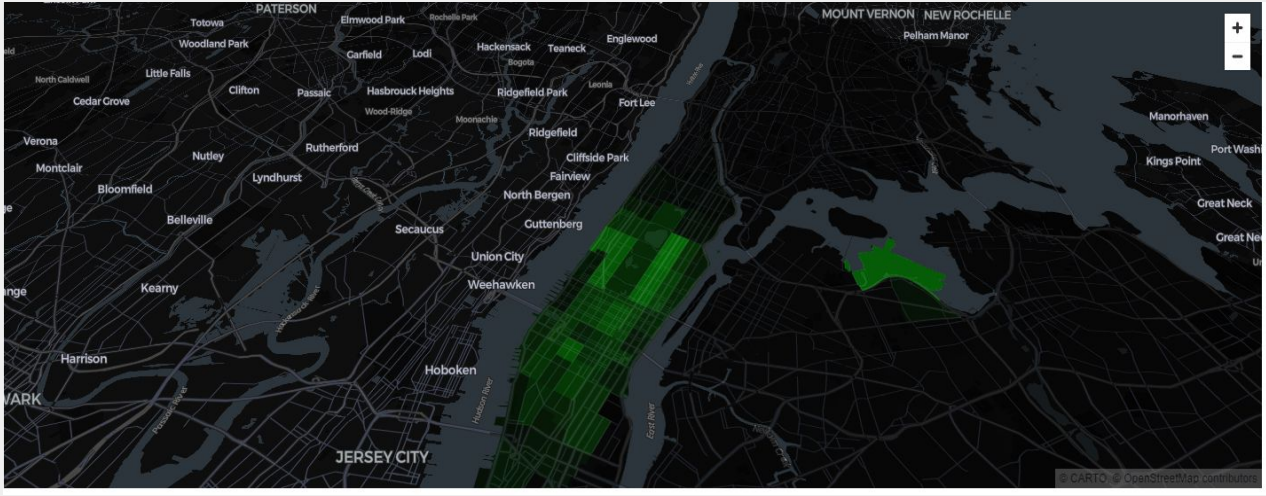


Model comparison

Model	Mean Absolute Error (MAE)	Notes
Ad Hoc model 1	6.05	Baseline model
Ad Hoc model 2	3.68	Baseline model
Ad Hoc model 3	3.19	Baseline model
XGBoost	2.70	Models improved
Lightgbm	2.57	Models improved
Lightgbm + feature engineering	2.59	Added average rides per month
Lightgbm + hyperparameter tuning	2.54(num_leaves,min_child_samples,etc)	Best model for production

- ➔ Further improve model performance by adding more features
- ➔ Build pipelines to automate processes
- ➔ Complete model operationalization

Application dashboard



Eddie Amaitum
Data Scientist

GitHub

THANK YOU