PedalSense

Bike Demand Forecasting

Riding into Change

PedalSense is a predictive analytics company that helps bike-sharing startups make go-to-market and capacity decisions by forecasting daily demand using historical usage and weather data.

Our machine learning models support smarter launch decisions, better capacity planning, and improved rider experience across urban environments.

PedalSense

Main Goal

Build a regression model that forecasts daily bike rentals demand using **weather, seasonality, and calendar features** to support go-to-market and capacity planning decisions.

PedalSense

Data Selection and Preparation

We used a public Capital Bikeshare dataset from Washington D.C. covering historical data of 2 years for daily bike rentals.

The data includes **weather, seasonal and calendar** features and reflects real-world urban mobility patterns.

[Link to the Dataset]

Data Cleaning

Cleaned the data by removing irrelevant columns and ensuring data quality

Exploratory Data Analysis

Conducted EDA to understand demand patterns and validate feature relevance

Encoding and Scaling

Encoded categorical variables through “one-hot encoding” method and scaled numerical features through “StandardScaler”

Train–Test Split

Split the data into training and test sets for modelling (20/80)

credit: vecteezy.com

Focus Areas

PedalSense

1

Key Demand Drivers & Weather Sensitivity

Which weather and seasonal factors affects daily bike rental demand, and how sensitive is it to changes in temperature, humidity, and wind speed?

2

High vs Low Demand Periods

Can periods of high and low demand be identified and characterized based on predicted daily bike rentals?

3

Optimal Market Entry Timing

When is the optimal time of year to launch a bike-sharing service in order to maximize early adoption?

4

Capacity Planning & Usage Patterns

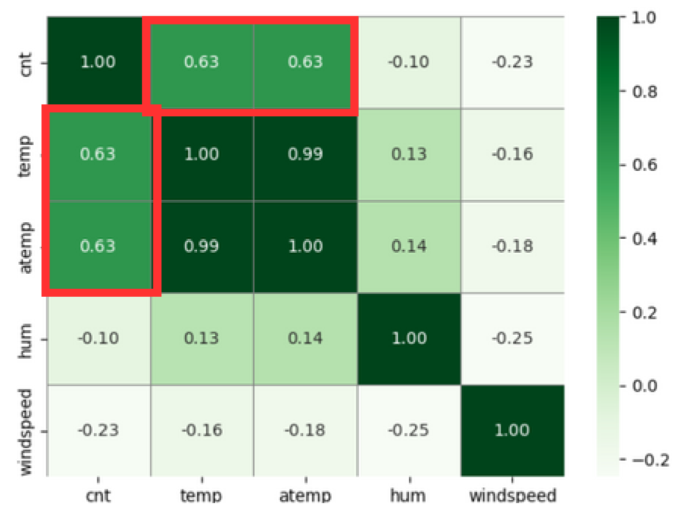
How many bikes should be deployed on a typical day, and how does demand differ between weekdays and weekends?

PedalSense

EDA Insights

Correlation Heatmap

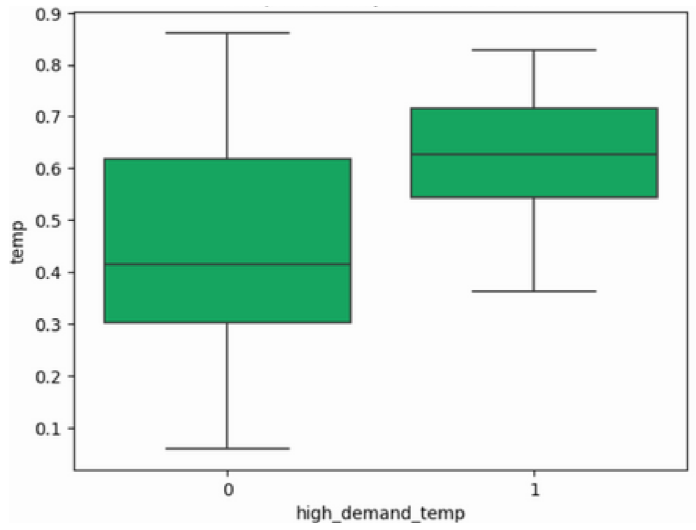
Correlation Between Weather Variables and Demand



Temperature → Strongest demand driver.

Temperature Separates Demand Levels

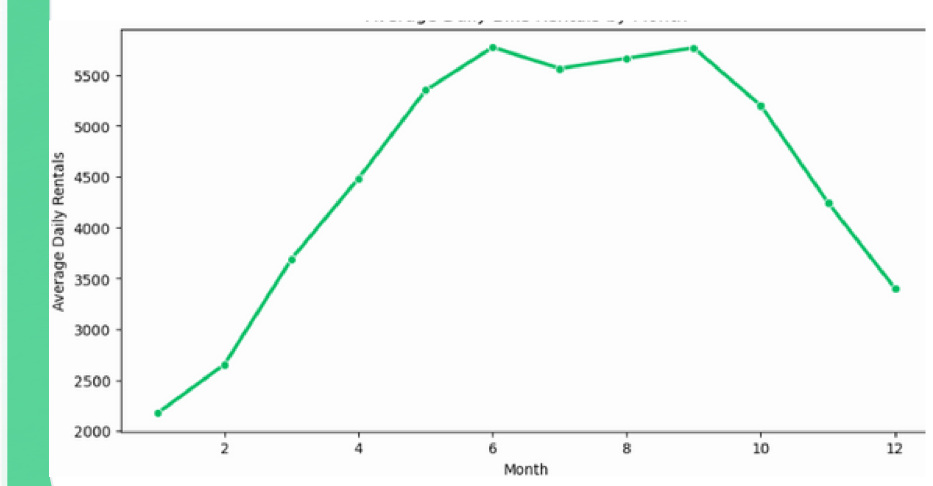
Temperature by Demand Level (High / Low)



high- and low-demand days → clear separation

Average Demand by Month

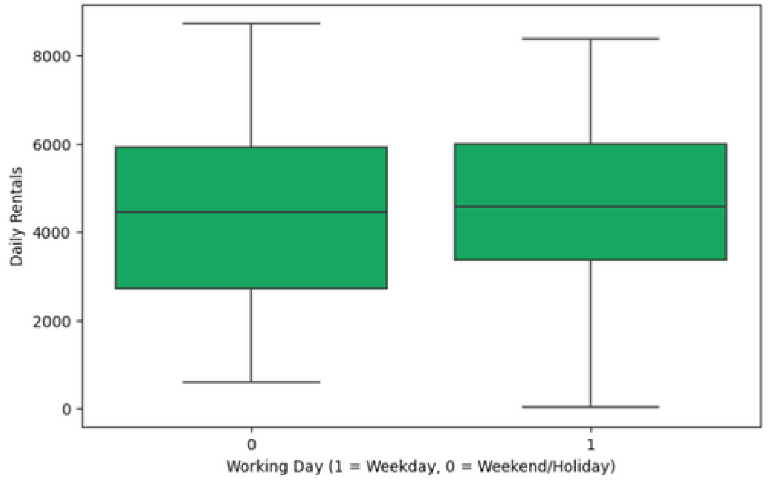
Average Daily Bike Rentals by Month



Demand → rises in spring and peaks in summer.

Weekday vs Weekend

Daily Bike Rentals: Weekdays vs Weekends



Weekday demand → more stable than weekends

RQ1 – What drives demand

- Demand is primarily driven by temperature and seasonality
- Other weather factors play a secondary role

RQ2 – Can demand levels be separated

- High and low-demand periods show clear and consistent differences
- Demand can be meaningfully classified

RQ3 – When to launch

- Demand increases sharply in spring
- Spring offers the best balance of growth and risk

RQ4 – How to operate

- Weekdays are more predictable
- Weekends require flexible rebalancing

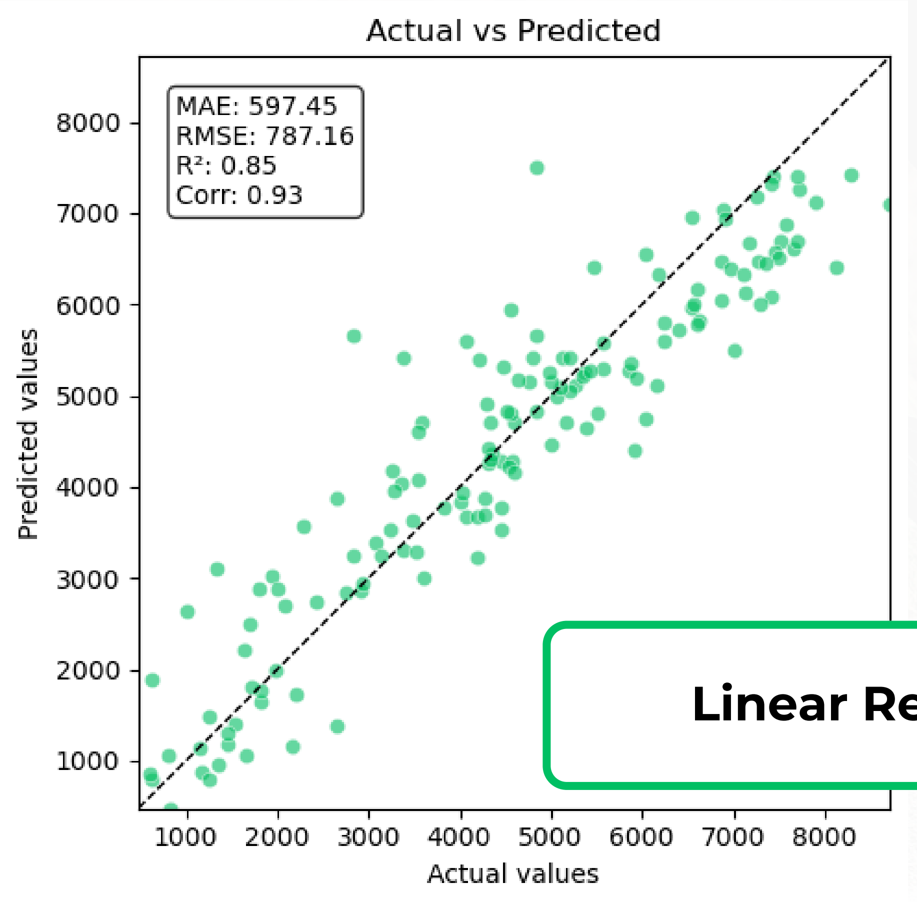
Feature Engineering and Selection

PedalSense

Decisions	*Challenges*
High correlation between temperature and feels-like temperature Risk of redundancy and multicollinearity. **Target variable “count” was composed of 2 other columns** Risk of target information leakage. **20/80 train–test split** Used a 20/80 train–test split to evaluate generalization performance.	**Categorical and numerical features were not separated as 2 separate datasets** Encoding was applied without isolating feature types which resulted in suboptimal feature representation and lower model performance. **Year-based split** As we had 2 years in our dataset tried to train on Year X to predict Year Y. This split reduced training data significantly and model performance was unstable due to limited sample size.

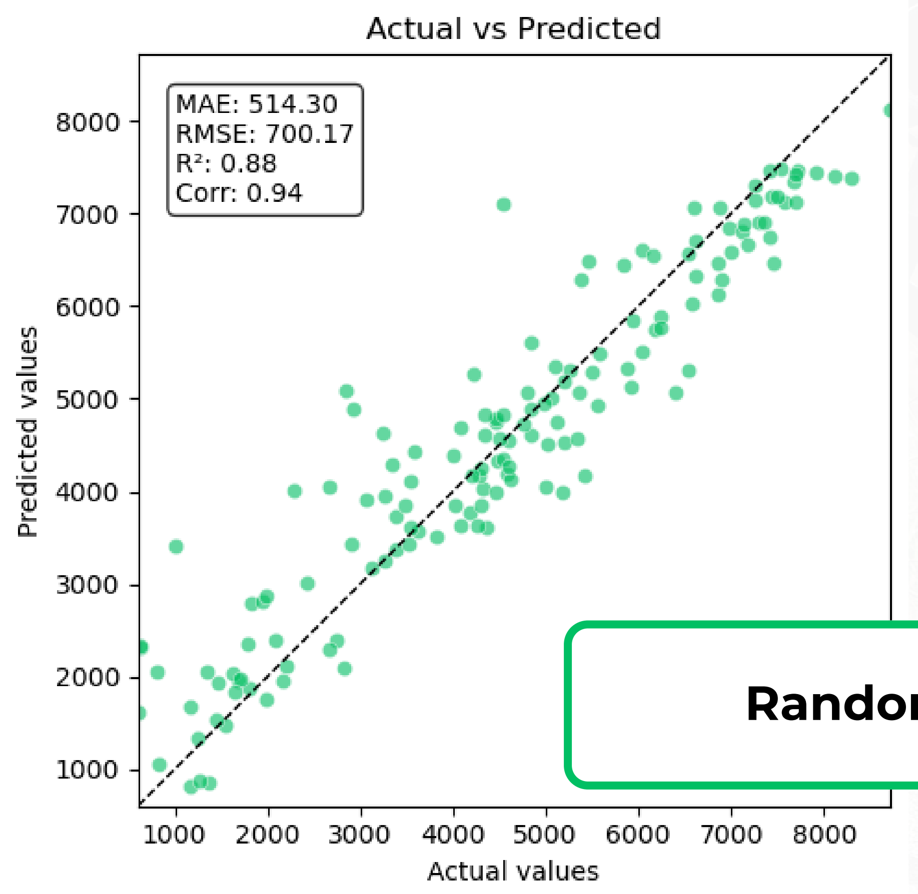
Models Evaluation

Metric: Test-set R^2 score



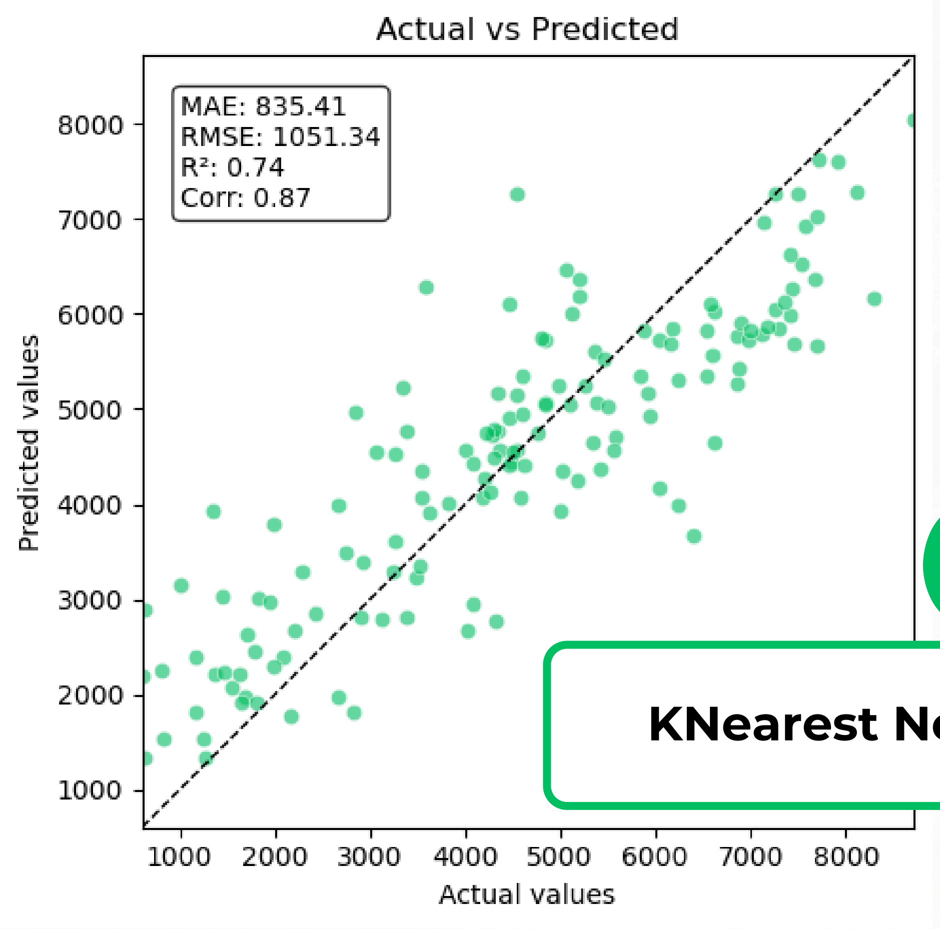
0.85

Linear Regression



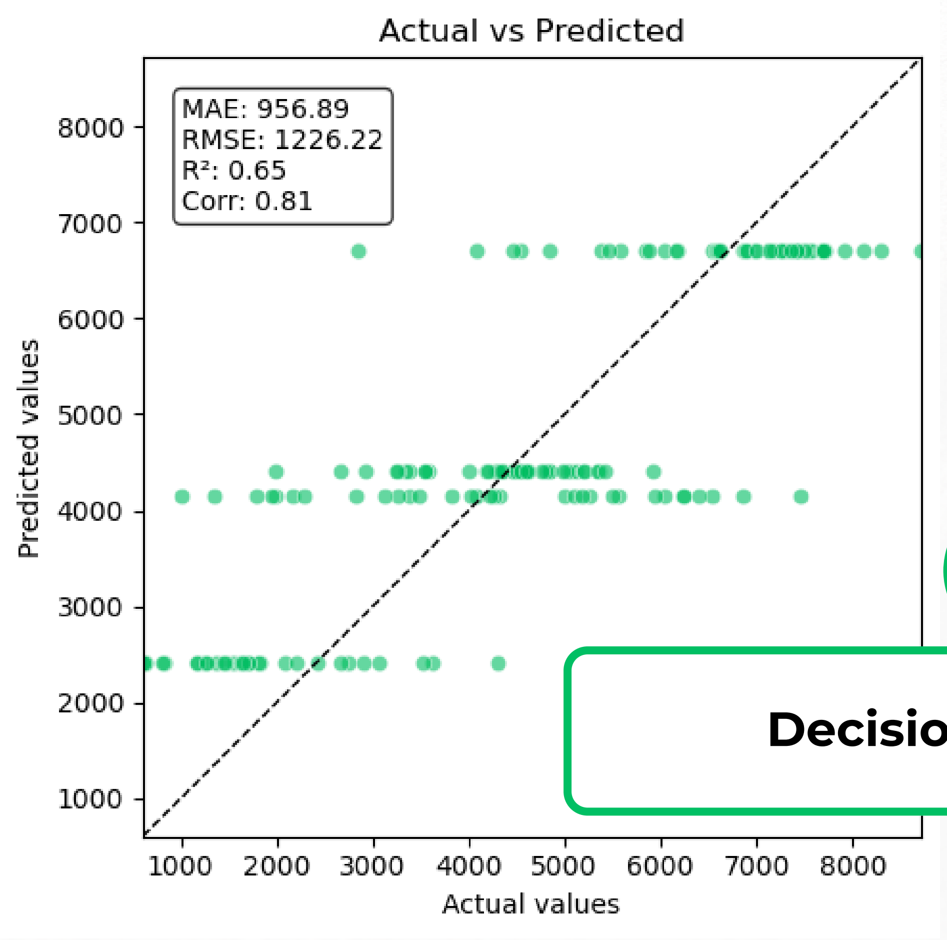
0.88

Random Forest



0.74

KNearest Neighbours



0.65

Decision Tree

PedalSense

Hyperparameter Tuning & Model Optimization

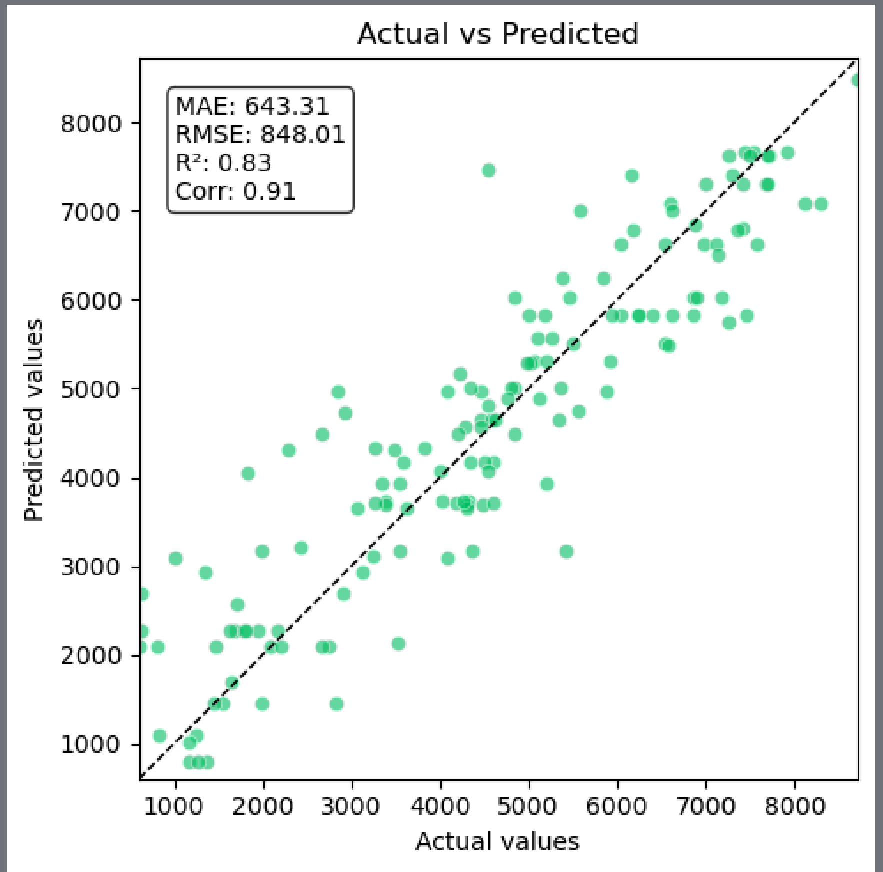
Grid Search

Decision Tree

Test R2 score:

0.65

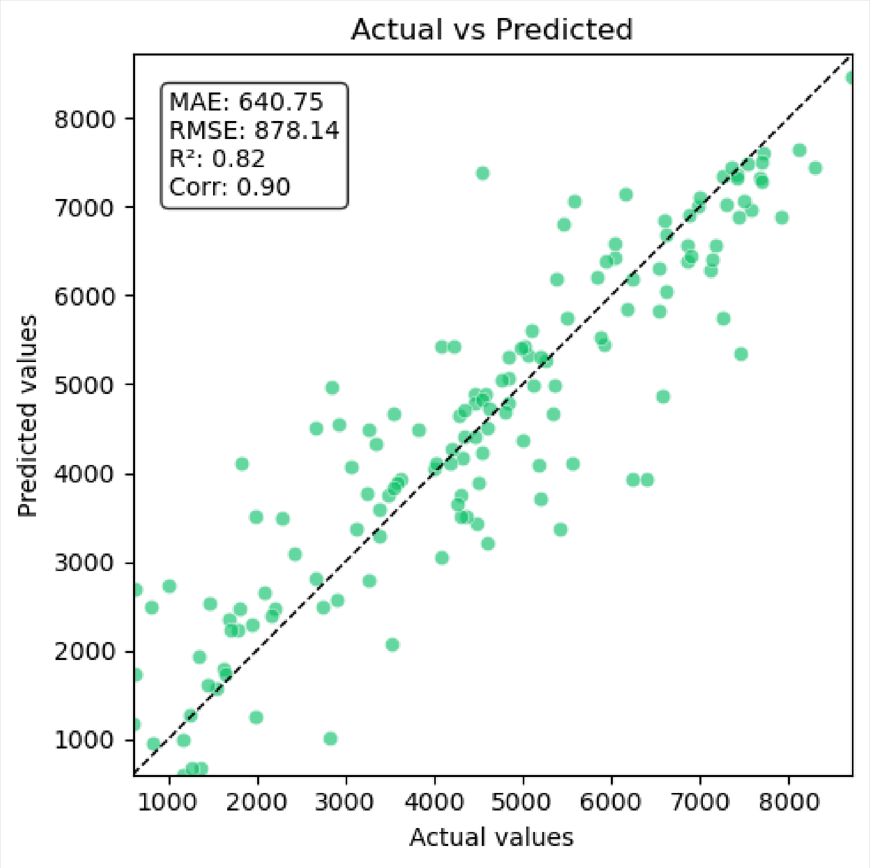
0.83



Gradient Boosting

Test R2 score:

0.82



Key Findings and Insights

PedalSense

Temperature

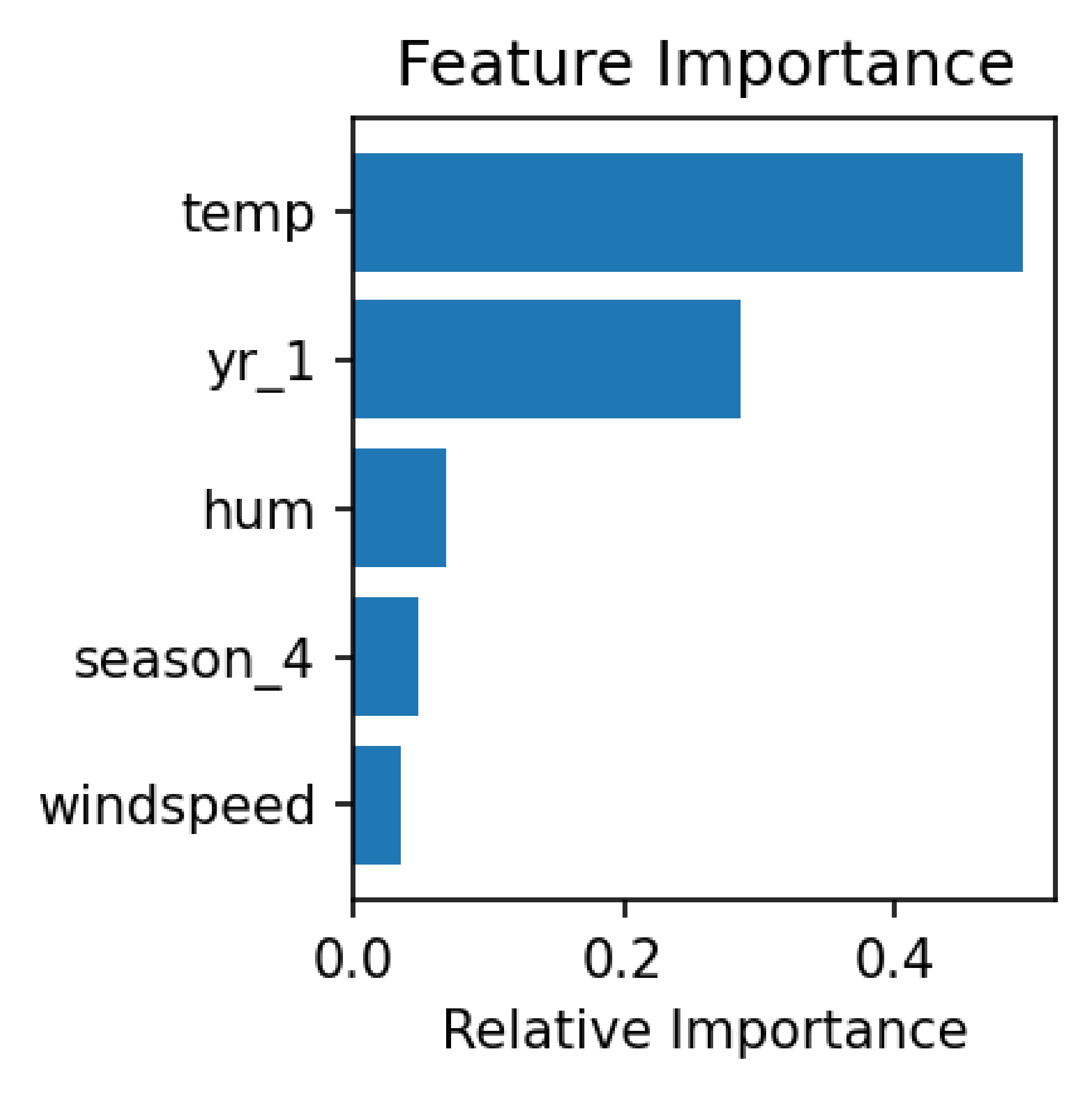
- Most important driver of bike demand

Weather variables

- Humidity, wind, weather situation influence demand, but less than temperature

Calendar features

- Season and year are driver indicator of demand



Real-World Application and Impact

PedalSense

Future Work and Improvements

PedalSense

More data on costs

With pricing and revenue data, additional engineered features could include:

- Revenue per rental
- Price elasticity indicators
- Profit margins
- Season-based pricing

More data on location

With location-level data, new engineered features could include:

- Identification of high-demand hotspots by station or area
- Optimized bike allocation strategies per zone

Thank You

Jim Hopper Team

Thank you for exploring this journey with us.

Davy Goupil
Ghazal Hassanzadeh
Helena Sospedra
Rui Parreira