# AI Forge: Predict bike rental demand - Project Proposal:

## ProjectTitle:A Regression Model for Predicting Bike Rental Demand

### 1. Problem Statement:

Predict daily/weekly/etc rental demand for shared vehicles (bikes, scooters, cars, etc.) using environmental and temporal factors (temperature, humidity, day type, holidays, seasonality, etc) and/or other task relevant features. The aim is to understand which factors most strongly influence demand and to support better resource allocation and planning.

### 2. Business Goal:

The primary goal is to develop a machine learning model that accurately predicts the bike rental demand with a significant accuracy. This model will empower the bike sharing business operations and delivery teams to proactively plan for peak and trough demand periods and allocate resources accordingly. The aim is to increase operational efficiency by 15% within the next fiscal year.

### 3.Data Source

We will use the "Rental Bike Sharing" dataset from Capital Bike Share. It's a well-structured, fictional dataset created by data scientists, making it ideal for this type of predictive modeling project.

* **Source Platform: Kaggle**
* **Full Citation:**
* Hadi Fanaee-T
* Laboratory of Artificial Intelligence and Decision Support (LIAAD), University of Porto  
  INESC Porto, Campus da FEUP  
  Rua Dr. Roberto Frias, 378  
  4200 - 465 Porto, Portugal
* Original Source: <http://capitalbikeshare.com/system-data>  
  Weather Information: <http://www.freemeteo.com>  
  Holiday Schedule: <http://dchr.dc.gov/page/holiday-schedule>

Kaggle. Retrieved October 2, 2025, from

https://www.kaggle.com/datasets/imakash3011/rental-bike-sharing

### 4.Tools & Technologies

* Programming Language: Python
* Core Libraries:
  + Data Manipulation & Analysis: Pandas, NumPy
  + Machine Learning: Scikit-learn
  + Data Visualization & Storytelling: Matplotlib, Seaborn
  + Development Environment: Our team will primarily use: Google Colab
* BITools: Power BI to create a final dashboard.

### 5.Project Workflow

The project will follow a structured data science lifecycle, visualized as follows:

Data Acquisition → Data Cleaning & Preprocessing → Exploratory Data Analysis (EDA) → Feature Engineering → Model Building & Training → Model Evaluation → Reporting & Visualization

1. Data Acquisition: Fetch the dataset from Kaggle using its API.
2. Preprocessing: Handle missing values (if any), encode categorical variables, and check for data inconsistencies.
3. EDA: Analyze features to understand their relationship with attrition using statistical summaries and visualizations.
4. Feature Engineering: Create new features from existing ones if necessary to improve model performance.
5. Modeling: Train several classification models (e.g., Logistic Regression, Random Forest, Gradient Boosting).
6. Evaluation: Assess model performance using metrics like Accuracy, Precision, Recall, and F1-Score. Select the best-performing model.
7. Visualization: Create an interactive dashboard in Tableau/PBI to present the key findings and predictions to stakeholders.

### 6.Data Extraction

Capital Bike Share - Rental Bike Sharing" dataset is acquired directly from the Kaggle repository. To ensure a professional and reproducible workflow, manually downloading the files is not done.

Instead, we will perform the following steps:

* **Automate the Process:** We will write a Python script that utilizes the official Kaggle API to connect to the source and download the dataset.
* **Ensure Reproducibility:** This scripted approach guarantees that the data extraction process is consistent and can be easily re-run by any team member or reviewer.
* **Prepare for Analysis:** The script will handle the unzipping of the downloaded files and load the data directly into a Pandas DataFrame, making it immediately available for the next phase of our project.
* Notebook: ***<TBD>***

### 7. Schema/Data Dictionary:

This data dictionary is created after inspecting the dataset.

Excel sheet: ***<TBD>.xlsx***