**Ola Demand Forecasting: Summary Report**

This report summarizes the key findings from the data analysis and the performance evaluation of various machine learning models developed for predicting bike rental demand.

**Columns**

- Season (1: Winter, 2: Spring, 3: Summer, 4: Fall)
- Weather condition (1: Clear, 2: Misty, 3: Light Rain, 4: Heavy Rain)
- Temperature
- Humidity
- Windspeed
- Number of casual (non-registered) users
- Number of registered users
- Total number of users (casual + registered)

**Data Analysis Key Findings**

The analysis identified several significant factors driving hourly bike rental demand:

**Demand Patterns**

- **Temporal Trends:** Demand follows a strong bimodal pattern daily, peaking during the typical morning and evening commute hours. Seasonally, demand is highest in the summer (Season 3) and lowest in the winter (Season 1). Overall, there was a clear upward trend in total rentals from 2011 to 2012. Weekends generally exhibit slightly lower demand than weekdays.
- The total demand (count) is, as expected, highly correlated with both **registered users (0.97)** and **casual users (0.83)**.

**Environmental Factors**

There was very little correlation between the variables.

**Machine Learning Model Performance**

Five models were developed and assessed using K-Fold Cross-Validation for robust performance estimation. The primary evaluation metric was the Mean Absolute Error (MAE).

**Cross-Validation Results**

| Model | Mean MAE | Standard Deviation (Stability) |
|---|---|---|
| Lasso Regression | 50.6814 | 0.6321 (Lowest) |
| Linear Regression | 50.7273 | 0.6622 |
| Ridge Regression | 50.7273 | 0.6622 |
| SVR | 50.7008 | 0.6945 |
| Random Forest Regressor | 51.3403 | 0.7911 (Highest) |

The **Lasso Regression** model achieved the **best overall performance**, demonstrating the **lowest mean MAE (50.6814)** and the **highest stability** (lowest standard deviation) across the cross-validation folds.

**Identified Model Limitations and Future Steps**

While the models provide a baseline prediction, several limitations were identified that restrict current accuracy and generalization:

**1. Model Limitations**

- **Target being a sum of two variables:** Combining casual and registered users into a single count masks the distinct underlying demand patterns for these two key user segments if they were separate.
- **Data Span:** The two-year dataset (2011-2012) is insufficient for reliably capturing long-term trends or complex, multi-year cyclical patterns.

**2. Feature and Data Limitations**

- **Missing Features:** The models lack critical external factors such as:
  - Holidays/Special Events (e.g., local festivals, major sporting events).
  - Economic Conditions or local demographics.
- **Model Simplicity:** The standard regression models may not be fully capturing complex non-linear relationships present in the demand data, and they may struggle to accurately predict demand during extreme or unusual events.

**Recommended Next Steps**

1. **Enhance Feature Engineering:**
   - **Integrate External Data:** Incorporate specific **holiday calendars** and **local event schedules**.
   - **Improve Weather Data:** Source and include data on precipitation levels and other weather factors.
2. **Explore Advanced Modeling Techniques:**
   - Investigate and tune more powerful algorithms such as Gradient Boosting Machines (e.g., XGBoost) to better capture non-linearities and achieve higher predictive accuracy.
   - Consider time-series specific models to explicitly account for temporal dependencies.