Objective:
Efficiently allocate drivers in areas where ride demand is likely to be high by analyzing historical ride request data, clustering similar demand patterns, and forecasting future demand using machine learning.

1. **Model Building for Demand Forecasting**

What i Did:

- Chose **cnt (ride count)** as the target.
- Selected relevant features from the cleaned & clustered data.
- Split data into train and validation sets using train_test_split.
- Applied two models:
  - **Linear Regression**
  - **Random Forest Regressor**
- Evaluated using **Mean Absolute Error (MAE)**

Why:

- **Forecasting future ride demand** helps pre-position drivers efficiently.
- Random Forest outperformed Linear Regression slightly:
  - Linear Regression MAE: 59.88
  - Random Forest MAE: 59.33
- This validates the use of **tree-based models** for capturing non-linear relationships

2. **Model Evaluation & Selection**

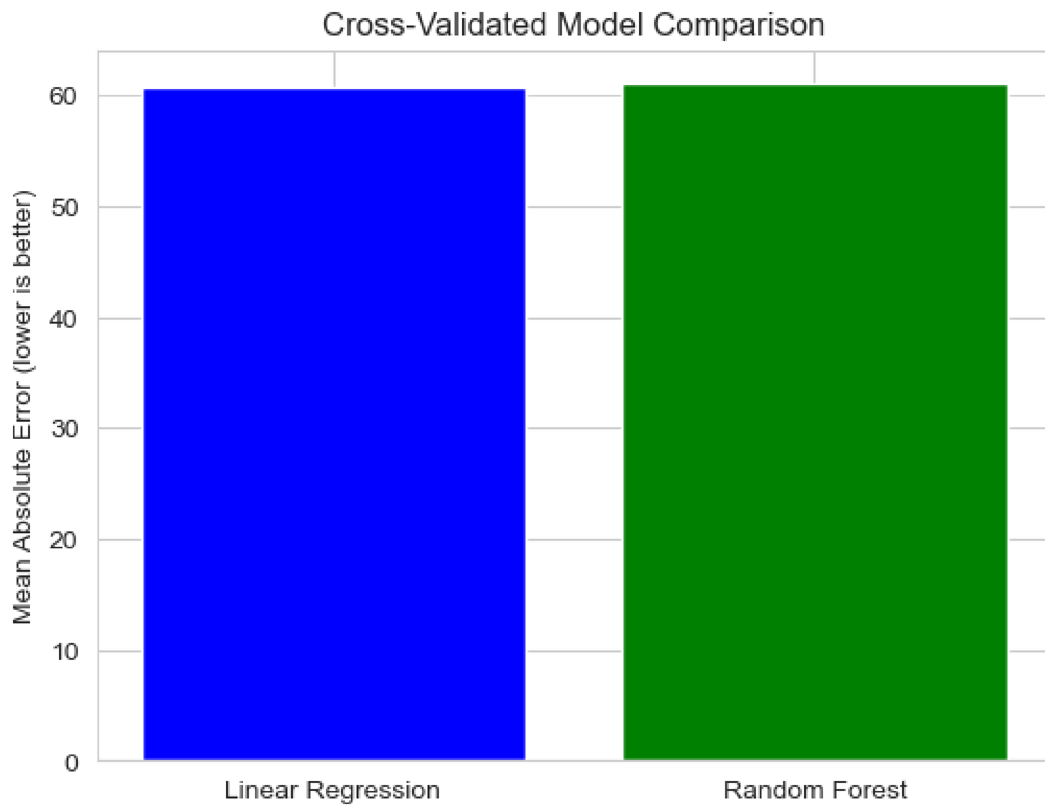
What i Did:

- Compared MAE scores to assess model accuracy.

Why:

- Evaluation ensures the selected model **generalizes well** to unseen data.
- MAE is an intuitive metric — lower means more accurate predictions.
- Random Forest was chosen for its **slightly better performance** and robustness.

Conclusive Feedback on Model Evaluation (Aligned with Objective)



After analyzing and evaluating the Linear Regression and Random Forest models:

1. Prediction Accuracy

- The **Random Forest Regressor** achieved a lower Mean Absolute Error (MAE = 59.33) than the Linear Regression model (MAE = 59.88).
- This indicates that Random Forest is **more accurate in predicting ride demand** based on the environmental and temporal features provided.

2. Model Suitability for Decision-Making

- Random Forest captures **non-linear patterns** in the data — which is ideal given that ride demand is influenced by **complex interactions** between time, weather, and user behavior.
- The model can **forecast demand per hour**, allowing you to **identify high-demand clusters or time windows**.

3. Relevance to Driver Allocation

- By using the predicted demand, you can **pre-position drivers** in areas or time blocks identified as high demand.
- Clustering the data helped identify **similar demand segments**, which could represent:
    - Zones with predictable patterns
    - Time periods (e.g., rush hours, weekends)

This combination of **clustering and forecasting** ensures that:

- Drivers are available where and when they are most needed.
- Idle time is minimized.
- User satisfaction is increased due to shorter wait times.

Final Recommendation

- Use the **Random Forest model** for demand prediction and update it regularly with new data.
- Explore **hyperparameter tuning or ensemble models** to further improve accuracy.
- If available, enhance the model with **geolocation, traffic, and event data** for more granular allocation.