Factors Influencing Bike Ride Demand

Several factors can significantly affect the demand for bike rides in a city or ride-hailing platform like Ola:

1. **Time-Based Features**
    - **Hour of the day**: Demand typically peaks during rush hours (morning/evening).
    - **Day of the week**: Weekdays might show work-related demand, weekends might reflect leisure rides.
    - **Season**: Demand may vary by weather season — higher in summer and lower in monsoon or cold seasons.
2. **Weather Conditions**
    - **Temperature**: Extremely high or low temperatures may reduce ride frequency.
    - **Humidity**: High humidity might discourage rides due to discomfort.
    - **Windspeed**: High winds might deter users from choosing bike rides.
    - **Weather situation**: Rain, fog, or stormy weather reduces visibility and comfort, often lowering demand.
3. **User Behavior & Events**
    - Although not in my dataset, **events**, **holidays**, and **traffic congestion** also affect demand.

Machine Learning Model Used for Forecasting

Data Preparation Steps:

- **Missing value treatment**: I handled missing values via mean imputation to retain data volume.
- **Feature scaling**: StandardScaler was used to normalize features for better model performance.
- **Feature selection**: I selected numerical and categorical features that impact demand — such as hour, season, temperature, etc.
- **Clustering**: KMeans was optionally used to group similar demand patterns.

Models Built:

I implemented and evaluated two models:

1. **Linear Regression**
    - Assumes a linear relationship between input features and ride demand.
    - Fast, interpretable, but less accurate for complex non-linear patterns.

2. **Random Forest Regressor**
   - An ensemble of decision trees capable of capturing complex interactions.
   - Provides better accuracy and robustness to outliers and noise.

Model Evaluation

Metrics Used:

- **Mean Absolute Error (MAE)** was the primary evaluation metric.

Model	**MAE (lower is better)**
Linear Regression	59.88
Random Forest	59.33 ☑ (Best)

☑**Insights:**

- The **Random Forest Regressor outperformed Linear Regression**, indicating that demand is influenced by non-linear relationships between factors like time and weather.
- The lower MAE suggests that Random Forest can predict ride demand with **better accuracy**, enabling more informed decisions on driver placement.

Limitations of the Model

Despite good performance, the model has some limitations:

1. **Lack of Spatial Data**:
   - The dataset does not include **location or zone information**, which is crucial for driver allocation in real-world use.
2. **Missing Real-time or Contextual Features**:
   - No data on events, holidays, traffic, or actual user bookings.
   - These would improve demand prediction accuracy.
3. **Overfitting Risk**:
   - Complex models like Random Forest may overfit the training data if not carefully tuned.
   - I used a basic version; **hyperparameter tuning** could help.
4. **Static Data**:
   - If the dataset is old or limited in size, it may not generalize to current demand trends.

Conclusion

This project developed a model that forecasts bike ride demand based on historical time and weather data. The **Random Forest model** proved to be the best choice for accurate predictions. These predictions can be used to:

- Efficiently **allocate drivers** in high-demand zones.
- Reduce customer wait times.
- Improve platform efficiency and customer satisfaction.

For real-world deployment, incorporating **geolocation**, **real-time updates**, and **external event data** would further boost performance and relevance.