**Project Proposal: Analyzing Bike Sharing Demand**

**Issue of Interest:**

The issue of interest for this project is analyzing bike sharing demand in the context of ride-hailing apps such as Uber and Ola. Bike-sharing systems have become increasingly popular in urban areas as a sustainable mode of transportation, and this trend aligns with the goals of ride-hailing companies to provide convenient and eco-friendly travel options to their users.

**Importance:**

Analyzing bike-sharing demand is important for several reasons. Firstly, it can aid urban planners and policymakers in designing efficient transportation networks. Secondly, bike-sharing companies can use this analysis to optimize fleet management, station placement, and pricing strategies. Lastly, studying bike-sharing demand patterns can provide insights into commuting habits, environmental impact, and public health.

These are the questions I am planning to answer after my project is completed:

a) What are the key factors influencing bike-sharing demand?

b) How does bike sharing demand vary across different times of the day, days of the week, and seasons?

c) Can we predict future bike-sharing demand based on historical patterns and external factors?

d) How do weather or other external factors impact bike-sharing demand?

**Data Sources:**

To analyze bike sharing demand, I am using data from the Kaggle data source and the size of the data is nearly 1Gb

**Variables/Measures:**

The variables/measures that will be used in the analysis include:

1. datetime
2. season
3. holiday
4. workingday
5. weather
6. temp
7. atemp
8. humidity
9. windspeed
10. casual
11. registered
12. count

**Unit of Analysis:**

Data Information

**Features**

1. datetime - hourly date + timestamp
2. season - season of the year (1 = spring, 2 = summer, 3 = fall, 4 = winter)
3. holiday - whether the day is considered a holiday (1: holiday, 0: not a holiday)
4. workingday - whether the day is neither a weekend nor holiday (1: working day, 0: not a working day)
5. weather:
6. 1: Clear, Few clouds, Partly cloudy, Partly cloudy
7. 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
8. 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
9. 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
10. temp - temperature in Celsius
11. atemp - "feels like" temperature in Celsius
12. humidity - relative humidity
13. windspeed - wind speed
14. casual - number of non-registered user rentals initiated
15. registered - number of registered user rentals initiated
16. count - number of total rentals

**Techniques/Models:**

The below techniques are the models I am planning to use.

b) Regression analysis to determine the impact of factors like weather, time of day, and season on bike sharing demand.

c) Clustering to identify areas with similar demand patterns.

d) Predictive modeling using machine learning algorithms (e.g., Random Forest, Gradient Boosting) to forecast future bike sharing demand.

**Model Evaluation:**

The performance of the models will be evaluated using appropriate metrics based on the chosen techniques. For regression models, metrics like mean squared error (MSE) or R-squared can be used. For predictive models, metrics like root mean squared error (RMSE) or mean absolute error (MAE) can be employed. Cross-validation techniques can also be applied to assess the generalization capabilities of the models.

**Intended Outcomes:**

The primary outcome of this project is to gain a better understanding of the factors influencing bike-sharing demand and develop predictive analytics to forecast future demand. The results can inform decision-making for urban planners, bike sharing companies, and policymakers to optimize resource allocation, infrastructure planning, and improve overall transportation systems.