**Instructions:**

1. Please ﬁll in all the required information.
2. Avoid grammatical errors.

# EDA on the given dataset:

From the dataset we understand that

### The dataset contains weather information (Temperature, Humidity, Wind speed, Visibility, Dew point, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

* Most number of bikes are rented in the summer season and the lowest in the winter season.
* Over 96% of the bikes are rented on days that are considered as No Holiday.
* Most number of bikes are rented in the temperature range of 15 degrees to 30 degrees.
* The highest number of bike rentals have been done in the 18th hour, i.e. 6pm, and lowest in the 4th hour, i.e. 4am.
* Most of the bike rentals have been made when there is high visibility.

# Conclusion:

1. Results from ML models:

* Random Forest Regression is the best performing model with an r2 score of 0.6645.
* Lasso Regression (L1 regularization) is the worst performing model with an r2 score of 0.4264.
* Actual vs. Prediction visualization is done for all the 4 models.
* All 4 models have been explained with the help of SHAP library.
* Temperature and Hour are the two most important factors according to all the models.

2. Challenges faced:

* Removing Outliers.
* Encoding the categorical columns.
* Removing Multicollinearity from the dataset.
* Choosing Model explainability technique.

| Drive Link:-  https://colab.research.google.com/drive/103AzaUWbvdgEK68IwwZtTUIYL\_fATj56?usp=share\_link |
| --- |
|  |