**Overview**

Building upon the previous capstone work, which focused on data aggregation, cleaning, processing, and visualization, this phase aims to develop machine learning models to predict hourly bike rental counts. The project will cover feature engineering, scaling, and regression modeling techniques to build an effective predictive model.

**Project Statement**

Develop an end-to-end machine learning pipeline to forecast hourly bike rentals using various regression techniques and performance evaluation methods

**Input dataset:** [Dataset](https://drive.google.com/drive/folders/1RFvzO7Eyidp4M9GCSO93izHXnaBFfQbF?usp=drive_link)

1. [Links to an external site.](https://drive.google.com/drive/folders/1RFvzO7Eyidp4M9GCSO93izHXnaBFfQbF?usp=drive_link)

**Steps to Perform**

**Task 1: Feature engineering (45 mins)**

* 1. Analyze the provided dataset and select relevant features
  2. Create new features such as:
     + Interaction features
  3. Encode categorical variables and handle missing values
  4. Scale the numerical features using StandardScaler
  5. Save the processed dataset as "bike\_rental\_features.csv"

**Task 2: Model building (75 mins)**

* 1. Implement various regression models including:
     + Linear Regression
     + Ridge Regression (L2 Regularization)
     + Lasso Regression (L1 Regularization)
     + Elastic Net Regression
  2. Perform hyperparameter tuning using GridSearchCV
  3. Evaluate model performance using:
     + Mean Absolute Error (MAE)
     + Mean Squared Error (MSE)
     + R-squared (R²)

**Task 3: Model building with polynomial features (45 mins)**

* 1. Create polynomial features for selected numerical columns
  2. Train models with polynomial features to capture non-linear relationships
  3. Compare results with linear models to assess improvements
  4. Save the best-performing model

**Task 4: Model evaluation and validation (45 mins)**

* 1. Perform cross-validation techniques to validate model performance (on both models- With Polynomial Features and without Polynomial Features)
  2. Assess models using test data
  3. Compare results across different regression models

**Task 5: Reporting and insights (30 mins)**

1.    Summarize findings and key takeaways from the analysis  
2.    Discuss feature importance and business implications  
3.    Provide recommendations for further improvements