MAT 374 Project 1 **Due Oct. 16**

Objective

This project allows you to apply multiple linear regression techniques to a real dataset from Open Data Pennsylvania. You will identify a research question, fit an appropriate model, interpret the results, and summarize your findings in a short written report.

Project Requirements

1. **Choose a Real Dataset**
   - Find a dataset from Open Data Pennsylvania that contains at least one quantitative response variable (Y) and two or more predictor variables ($X_1, X_2, \ldots$) and at least 50 observations.
2. **Formulate a Research Question**
   - Clearly state what you aim to predict or explain. *For example: "How well do advertising spending and product price explain sales volume?"*
3. **Fit and Evaluate a Multiple Regression Model**
   - Fit a regression model.
   - Report model summary output (e.g., R^2, adjusted R^2, F-test, p-values).
   - Include relevant diagnostic plots.
4. **Interpret Your Results**
   - Interpret each estimated coefficient in context. Discuss which predictors are significant and what that implies.
   - Summarize the strength and direction of relationships.
5. **Make Predictions**
   - Use your fitted model to make a meaningful prediction for some new observations.
   - Report a confidence or prediction interval when appropriate.
6. **Write a Short Summary**
   - Research question and brief dataset description.
   - Variables used and model specification.
   - Key output with interpretation of parameters.
   - Include at least one example prediction.
   - Short reflection on model fit, limitations, and what you learned.

Submission Guidelines

Submit a report in html format that includes the link to the dataset, R code, output, and interpretations.

Rubrics

Ratings	Descriptions
0-5	Html report readability
0-5	R code reproducibility and dataset source
0-15	Proper model specification, interpretation of coefficients, and model output
0-10	Prediction and interval calculation
0-5	Clarity of conclusions and discussion
0-40	