

Final Project Checkpoint 1: Dataset Identification

Objective: Identify a dataset that can be used for multiple linear regression, explain its importance, and load the dataset into RStudio.

Task 1: Identify a Dataset (3 points)

- Choose a dataset that contains:
 - At least one **quantitative response variable** (dependent variable).
 - At least **three or more explanatory variables** (independent variables).
- The dataset should have enough specimens (observations) for meaningful analysis.
- Consider datasets from publicly available sources (e.g., Kaggle, UCI Machine Learning Repository, government databases) or simulated data if necessary.

Task 2: Justify the Dataset Selection (4 points)

- **Write a justification (150–250 words) explaining:**
 - Why this dataset is important or relevant in a real-world context.
 - Whether it is a real-world dataset (collected from actual observations) or a simulated dataset (generated artificially).
 - Any potential limitations of the dataset.

Task 3: Load and Verify the Dataset in RStudio (3 points)

1. Load the dataset into **RStudio**. If using a CSV file, use the following command:

```
dataset <- read.csv("your_dataset.csv")
```

2. Test that the dataset has been successfully loaded by running:

```
head(dataset)
nrow(dataset)
names(dataset)
```

Submit answers on [this Google Form](#)