COSC 410 S24 Lab 4: Using data to answer concrete questions

This Lab is due by 11PM EST on the day of the lab (Feb 26/27 depending on your lab). You should work in groups of two or three for this lab.

Introduction

The purpose of this lab, and the associated HW3, is to strengthen your ability to find answers to concrete questions from a given dataset. This is an important skill as you start thinking about your final projects. By completing the lab and the HW, you will demonstrate that you understand:

1. Given a broad question and dataset, operationalize the question in terms of the data you have available.
2. Fit a linear regression model and interpret the learned parameters.
3. Compute the goodness of a linear regression model and use this to reason about the usefulness of features.
4. Given a question and a learned model, come up with an answer to the question (or justify why it is not possible to come up with a clear answer)
5. Given some code, explain what is the task the code is trying solve, why, and how it solves it.

Structure

This lab has two parts:

1. Work through an ipynb notebook to answer the following question: Are women given worse car rates than men?
2. Look through the requirements for HW3 and brainstorm what you might include in each of the sections.

Provided Files

- Lab4.ipynb and car_sales_data.csv for the lab
- HW3.ipynb and energy_efficiency_data.csv for HW3

Part 1

Work through Lab4.ipynb with your partner. Then submit `Lab4.ipynb` with the cell outputs saved to Gradescope/Moodle for your lab section.

Part 2

Make sure to leave yourself at least 15-20 minutes to work on this part

Homework 1 requires you to generalize the approach you adopted in `Lab4.ipynb` to the generate a proposal for a construction company. The starter file for this homework `HW3.ipynb` is deliberately sparse. We hope that working with this more open ended format will help develop the skills you need for the final project.

Brainstorm how you might approach this homework with your partner.