

MLDS 422 – Fall 2023
Homework 4
Due Thursday, 12/7/23 at 11:59pm

Answer the following in ONE of these languages: Rust / Scala / Julia

Question 1: Write a function [30 points]

- The function should:
 - Take in a string
 - Split the string on spaces
 - Remove a few punctuation marks (.,?!)
 - Make the text lowercase
 - Return the most common word in the string
- Input a test string ('I bought a sandwich with a side of chips!') into the function and print the output

Question 2: Analyze a data set [45 points]

- Read in the movie_reviews.csv file
 - Print the number of rows in the data set
- Filter the data to only contain “popular” movies that were released in theatres before 2010 (popular = movies with more audience reviews than the average number of audience reviews of all movies before 2010)
 - Print the number of rows in the output
- Using the filtered data, display the percent of movies that fall under each type of rating (R, PG-13, etc.)
 - Share an insight from the summary table
- Engineer a new feature from any of the existing columns
 - Create a summary table using the new feature
 - Share an insight from the summary table

Question 3: Explore an external library [15 points]

- Import a library that wasn't covered in class
- Explain how it works
- Show an example of the library in action

This final question should be completed within a word document or spreadsheet.

Question 4: Compare data science programming languages [10 points]

Fill in the following table:

Language	Summary	Strengths	Weaknesses
Python			
R			
Rust			
Scala			
Julia			