

Week 4 Progress Report

DATA 450 Capstone

Nick Nemkov

Week ending: 2/16/25

Time Log

Day	Time	# Hours	Task(s)	GH Commit(s)
Mon.				
Tue.				
Wed.				
Thu.				
Fri.				
Sat.				
Sun.	[1-5pm] ; [5:30- 7:30pm]	[4] ; [2]	[Load in datasets, check for NA values] ; [Begin separating data into categories.]	Link 1 ; Link 2

Total # of hours worked this week: [6]

What you accomplished (or attempted) this week:

I have loaded in the adult.data and stat_M2023_dl.xlsx files and began performing EDA on the dataframes. The adult.data file needed a manual input of a header. Direct paths to the files were used to read in the files to a Jupyter notebook file (salary_predictions.ipynb). For the EDA, I have checked the adult.data file for NA values. I have also began separating the education levels in adult_df (the adult.data dataframe) into 4 categories (Preschool, Primary, Secondary, and Higher). A scatter plot visualization will follow.

Any setbacks/roadblocks you experienced:

I need to further revise the questions for the proposal. I will be adding more visualizations and sub-questions to expand questions 1, 2, and 4. Due to this I will need to do more work next week for the first research question as I was working to expand the questions.

What you plan to work on next:

Once I divide adult_df into the 4 education categories, I will get into plotting the data based on Question 1. Throughout next week, I will finish Question 1 and begin working on Question 2.

Are you on track? If not, how will you get back on track?

I am not on track due to revisions needed to be done to the proposal. I will continue throughout Sunday to add to the proposal research questions. Next, week I will continue work on Question 1. I will try to put in extra work to complete both questions 1 and 2 for next week.