New Attempt

- Due Jul 23, 2023 by 11:59pm
- Points 100
- Submitting a file upload
- File Types html, rmd, and ipynb
- Available after Jun 15, 2023 at 12am

A7 Project Assignment

Instructions

Objectives: Use the skills you have learned in this class to analyze and model a new dataset.

Consider that you would be in a critical meeting with stakeholders and reading from this Rmd to explain your exploration, modeling and findings. Your code as well as notes in the Rmd are crucial to convey your analysis.

Python: You may also complete this assignment using Google Colab and coding with Pandas, scikit-learn if you choose. You must submit your ipynb and html. The requirements of the assignment are still the same.

Input files:

- **census.csv (https://utah.instructure.com/courses/882385/files/147060830?wrap=1) (https://utah.instructure.com/courses/882385/files/147060830/download?download_frd=1)**

Data Source: The Census

Alas you have very little information about this dataset. The target variable is a column named y. Your goal is to accurately predict income based on the data provided so that our company can market financial products to these individuals. The rest you must figure out yourself.

Task I - EDA (25%)

Perform exploratory data analysis on the dataset. Add text to note items from code blocks you have observed.

Task II Data Preparation (15%):

Prepare your data for modeling.

Task III - Model Building (30%)

Build a variety of models and keep them to show your efforts to achieve good performance.
Your stakeholders prefer interpretability over performance. Consider this as you choose your models.
How many models are sufficient? Enough to show that you found underfitting, overfitting and a good balance between the two. Your stakeholders still want a good model however.

Task IV - Reflections (30%)

Write up your summary findings in a paragraph to conclude your analysis.

Render A7_LastName_FirstName.Rmd to a HTML file and submit.

(Please do not compress these two files into one zip file. Submit them as separate files. Be sure that the HTML output file contains assignment title, author name – you, the file creation date, and table of content including section numbers.)

A7 - Rubric

Criteria	Ratings				Pts
Task 1: EDA	**25 pts** **Full Marks**	**19 pts** **Mostly Met** Most work was done in this section was of high quality and included items we have seen throughout the course.	**13 pts** **Partially Met** While work was done in this section it is either incorrect, or insufficient.	**0 pts** **No Marks** This section was either missing or extremely short and insufficient.	25 pts
Task 2: Data Preparation	**15 pts** **Full Marks**	**12 pts** **Mostly Met** Most work was done in this section was of high quality and included items we have seen throughout the course.	**8 pts** **Partially Met** While work was done in this section it is either incorrect, or insufficient.	**0 pts** **No Marks** This section was either missing or extremely short and insufficient.	15 pts
Task 3: Model Building	**30 pts** **Full Marks**	**23 pts** **Mostly Met** Most work was done in this section was of high quality and included items we have seen throughout the course.	**15 pts** **Partially Met** While work was done in this section it is either incorrect, or insufficient.	**0 pts** **No Marks** This section was either missing or extremely short and insufficient.	30 pts
Task 4: Reflections	**30 pts** **Full Marks**	**23 pts** **Mostly Met** Most work was done in this section was of high quality and included items we have seen throughout the course.	**15 pts** **Partially Met** While work was done in this section it is either incorrect, or insufficient.	**0 pts** **No Marks** This section was either missing or extremely short and insufficient.	30 pts
					Total Points: 100