## **Take Home Assignment**

Congratulations, you have made it to the take home assignment portion of our interview process! This assignment should take an hour or two to complete. Thank you for your time in interviewing with us and please reach out if you have any questions about the assignment.

### **Dataset**

The training and testing datasets are included in this folder.

In this assignment you will be evaluating UCI’s Census Income dataset. The classification task is to determine whether a person makes over 50K a year. Please examine the prediction using a minimum of 2 machine learning model endpoints and return a brief analysis of the best result. The feature space is as follows, and contains a mixture of numerical and categorical inputs.

1. age: continuous.
2. workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
3. fnlwgt: continuous.
4. education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
5. education-num: continuous.
6. marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
7. occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
8. relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
9. race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
10. sex: Female, Male.
11. capital-gain: continuous.
12. capital-loss: continuous.
13. hours-per-week: continuous.
14. native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

Tasks

1. Use any two ML models to predict whether the income will be >50k or <=50k for each adult in the test dataset. Please include all the procedures and code as well as the results.
   1. During the data preprocessing stage of the workflow, please include at least one data visualization of your feature space.
2. Please set up an endpoint API for predicting the result from your best ML model. Your endpoint should take the input data in the test dataset (single row of data or a batch) and return the predictions. Please include your code and steps to use them.