Background:

An analytics/data consulting company has been noticing a trend of high performing employees leaving at higher rates. They want to understand what is driving high-performing employees to leave and to develop a model that helps predict an employee at risk of leaving. Currently there is no clearly defined way of defining high performers. Bob in IT has provided this file.

Objective:

Determine what is driving high-performing employees to leave and offer recommendations on how to stop this trend. Build a classification model that can be used in future to predict those at risk (probability) of leaving.

Deliverables:

- Analysis Presentation with Recommendations
- Model Equation;
- Notebook (.ipynb) with markdown documentation and appropriate headers showcasing your Exploratory data analysis, experiments
- Productionized Code, meaning a cleaned up ipynb that just has your final model object and the code required to run the pipeline of load, extract, transform, predict, evaluate (File containing model object; code to read new independent variables of same array as original, prediction, writing predictions to csv file)
- link to your git hub project repo. I will be doing git hub pull requests to review your code here

Mandatory:

- Apply what you've learned so far about exploratory data visualization (EDA, visualization), feature correlation, feature selection, preprocessing (scaling, typing, mapping, dropping), feature engineering (if applicable)
- Modelling experiments should be recorded and output as a dataframe to organize your thoughts model planning
  - from sklearn.model_selection import GridSearchCV
  - hint: if using grid search use mygridsearchobject.cv_results_ → dataframe
- Explore your model coefficients (data visualization) and give some interpretation why your best model was selected (might be challenging)
- Your own personal insight into other interesting trends or patterns you noticed during your EDA
- 3 user defined functions
- Create or use your designated project repo for this project:
  - README.md: make sure to fill this out to give a high level showcase about what the project is about, what domains does it cover from class, what you learned from your dataset/modelling, and what you learned from doing the project. This is to not only showcase your progress as you move along in the curriculum but also for anyone who visits (employers, etc) to see your abilities

Nice to haves:

- Python File (.py)
  - Your final code should just be your final model (no markdown or EDA here) and the functions that it needs to load in your data, preprocess, and run predictions
    - Hint: Putting your code into concrete user defined functions can help with this (not only for readability but also for testing later on)
  - should be able to run with this terminal command: python3 model.py

Audience:

Operations Manager; HR Director