Blue Owl Simple Challenge

Data Science - Insurance

Hi! Welcome to the first step of the Blue Owl interview process - the simple challenge. Success in the simple challenge leads to the final two steps of the interview process:

1. Informal chat with Blue Owl founders
2. Full technical challenge

For the simple challenge, use “train.csv” to predict the *outcome* variable using a GLM with a log-link and Poisson distribution. The category named “categorical” is a categorical column and the column named “numeric” is a numeric column. Both should be used as independent variables.

Requirements:

1. All code must be written in Python and must be in a Jupyter notebook
2. The first cell in the notebook must include:
    a. Your last name (please don’t include any other identifying information)
    b. The date
3. You must output the GLM’s parameter estimates.
4. Your code must be able to predict all five observations in the “test.csv” dataset. The last cell in the notebook must output the first five predicted values of the *outcome* variable for test.csv.
5. A key point of evaluation is how well written the code is. Please write the code as if you are writing it for a production setting. No need to wrap the code in a service or write Dockerfiles. Just ensure the code you write is not hacked together.

If you are spending more than an hour on this simple challenge because there are so many things you want to demonstrate, you are spending too much time on it. If you are spending more than an hour on it because you don’t know where to start, please be warned that the full technical challenge will be considerably more difficult.

When you are finished, please download a copy of your notebook in the following three formats: ipynb, html, pdf. Email these files to rachelle.valk@blueowl.xyz with the subject “Data Science - Insurance prescreen”. In the email to Rachelle, please include your full name and a few one hour blocks of time you have as options for the next step of the interview process. Best of luck!