

One of the key requirements of this role is that you're ready to be hands-on. The next step consists in completing 3 tasks. We expect you to finish them and submit them through Greenhouse. We will do our best to get back to you as soon as possible.

About Data

Set:

You will find 2 .csv files attached to this task. 1 of the files consist of courier's lifetime dependent features and other consist courier's weekly variant features. Features are renamed for confidentiality purposes and data dictionary will NOT be provided. However, in 2 different .csv files, same courier ID represents same courier.

Task 1: Exploratory Analysis and Data Munging

In this task, you are being expected to clean data, treat missing values, find out related features and finally label the data.

Every courier did not work every week. Thus, some of courier-week combinations' data are not provided. First, come up with a way to treat these missing values. Removing missing values are not suggested since provided data set is small and it will affect your predictive model's evaluation metric.

Create a report / dashboard and correlation matrix, in addition to results of your univariate and bivariate analysis and explain your findings.

Finally, label your data. If a specific courier's week 9, 10 and 11 data is not provided, we label this courier as "1" otherwise "0". After labeling, remove week 8(Yes including 8!), 9, 10 and 11 data to avoid bias in your next task. In addition, distribution of feature_3 is a hint how the data is generated.

Task 2: Create a Predictive Algorithm

In your second task, you were expected to create a logistic regression model that classifies your labels that done in the first task.

Create a logistic regression model by using Python or R. You are free to choose your algorithm and libraries / packages to use.

Finally, tune your hyper-parameters of your model by randomized search, grid search or any other search method and explain your reasoning for this choice.

Task 3: Evaluating the Model

Choose an evaluation metric fitting to your model and explain why you have chosen this and share your scores for this metric.

You need to send these 3 tasks before the date stated in the mail. If you don't send it by then we will assume you're not longer interested in the position. We believe both tasks are clear enough. However, if you have any technical question, please email them to Koray (koray@glovoapp.com).

Good
luck!