

Fresh Geeks 2021 - Data Scientist Test

INTRODUCTION

Welcome to One Mount Data Scientist Recruit Challenge!

In this challenge, you are going to help Nexus Telecommunication Company understand and predict customer churn based on information about their account. The data consist of 19 predictors such as number of customer service calls, the area code, and usage duration collected during one month. The outcome is whether the customer churned (cancelled his plan) after that month.

We expect your deliverable to be a notebook file containing (1) code (preferably Python) and (2) answers to the TASKS section below.

Please contact Mr. Kien (kientt15@onemount.com) for any further questions.

Good luck,

One Mount Data Team

DATASET

The predictor names are self-explained, but the following notes might help:

- `account_length`: number of days from registration
- `number_vmail_messages`: number of voice mail messages
- `total_day_minutes`: total call duration in minutes made and received during day time
- `total_eve_charge`: total amount of money to be paid for calls during evening time
- `total_intl_calls`: total number of international calls

TASKS

- 1) Conduct exploratory data analysis with a focus on the relationship between the predictors and the outcome.
- 2) Fit a few models to the training set and tune them via resampling. What criteria should be used to evaluate the effectiveness of the models?
- 3) (Optional) Plot lift charts to compare models. If you wanted to find 80% of the churning customers, how many customers would have to be identified?