

LA Times Data Modeling Take Home Test

Thank you for your interest in a position with the Data Science team at the LA Times! We have huge volumes of traffic coming through our digital and mobile properties monthly, a large amount of curated editorial content being produced daily, and very active social channels and marketing campaigns. We are looking for candidates that are excited about working with these data, deriving insights, and building models to advance our business!

The purpose of this test is to give us a sense of a candidate's capabilities in a few different areas:

- Ability to dive into a data set and tease out insight
- Familiarity with machine learning algorithms and evaluating their performance
- Performing feature generation on a data set
- Coding ability and style
- Knowledge of modules and libraries used in machine learning

We ask that you complete your task in Python (preference for python 3.* but python 2.7 is okay) as this is one of the main languages we work with in-house. If you'd like to work in something else, please discuss with the person administering your test.

We also recommend that you do your work in a Jupyter notebook.

We expect the task to take at least a couple hours to complete. Thanks for your time and we look forward to seeing your solution.

Description

We have provided you with a file `posts.json.zip`. This file contains data on the content and performance of news posts on Facebook page over a one year time period. Each line of the file is a JSON document corresponding to a single post.

We ask that you develop a model to predict post performance for new posts a user may want to create for this account.

Feel free to use any standard modules or tools available, but please include a list of modules that you used in your solution to make it easy for someone on our end to run your code.

Evaluation

This is an opened ended problem and we do not have a single set of metrics we apply to evaluate your solution. Feel free to focus on aspects of the problem that demonstrate your

strengths! Some things we will look at:

- Coding style
- How did you define the problem you solved?
- How did you evaluate your solution?
- Was your model a good choice for the problem?
- What kind of features did you use?