Demand prediction challenge

Background

One of the most important signals in Uber's system is demand. It's used in a variety of capacities, from calculating dynamic (surge) pricing to making decisions about the number of drivers needed to keep the system running smoothly. One of the ways we quantify demand is by tracking when users open the Uber app on their phones.

Your task is to predict future demand, based on historical data.

Dataset

Attached is a JSON object containing client logins from March 1st to May 1st 2012. All logins are from the Washington DC area and are each logged as a UTC timestamp.

Challenge

- Predict the number of logins (a proxy for demand) per hour from May 1 to May 15th 2012 UTC.

- Implement your solution as an API which accepts demand data and returns predictions.

- After you've built an API which returns the predictions for the specified date range, extend your API to accept a stream of demand data and continuously update the predictions.

Solution

- Return your May 1 to May 15th 2012 predictions in CSV format. The first column a ISO-formatted UTC timestamp of the start of the prediction period and the second column your prediction. For example, a few rows of predictions might look like:

  2012−05−01T00:00:00 ,33.02 ,
  2012−05−01T01:00:00 ,42.96 ,
  2012−05−01T02:00:00 ,12.03 ,

- You may use any programming language, but Python is preferred. If you're looking for a simple framework to build an API with, check out flask.

- Please include your code and a writeup explaining your methodology. Feel free to visualize or otherwise point out any interesting patterns you notice in the dataset - we're always excited about exploring our data in useful ways and give major bonus points to anyone for doing so.

- The target schedule for this project is ideally about a week; for a short-term coding challenge we're less interested in cutting edge research and more on solid approach and solid implementation.

Please let me know if you have any questions.