

The word "ARUNDO" is written in a large, white, sans-serif font. It is centered against a blue background that features a network of white dots connected by thin white lines, resembling a molecular or data network structure.

# Arundo data science take home challenge

## May 2017

ARUNDO ANALYTICS  
HOUSTON | OSLO | PALO ALTO  
[WWW.ARUNDO.COM](http://WWW.ARUNDO.COM)



Arundo Analytics  
Houston | Oslo | Palo Alto  
www.arundo.com

## CHALLENGE DESCRIPTION AND GOALS

The data science take home challenge is intended to help us evaluate how you do on a task reflective of typical data science work at Arundo. The test is centered on building a predictive model and answering some basic data analysis questions, but the primary thing we are looking to understand is how well you can deliver something of value in a short time frame and how well you can communicate your results, process, and decisions that went into the work.

The goal of the take home challenge is to predict the number of engineering support requests that a set of field sites request on any given day. You are tasked with building a model that can predict the request count in the test data set and answering a few questions about the data.

This project is designed to be doable in 2 - 3 hours. We are not looking for a perfect prediction, but rather an understanding of how you approached the problem and the result you produced in a limited amount of time with limited data.

## DEADLINES

1. You should return your results and report **by midnight** on the day which it is given to you.
2. **Optional:** you have **1 week** to add anything else to your result that you think is interesting or would improve the result and report.

If you have questions, please send them to [roy.keyes@arundo.com](mailto:roy.keyes@arundo.com)

## TECHNICAL DETAILS

### FILES AND DATA

The data consists of two files:

1. Arundo\_take\_home\_challenge\_training\_set.csv
2. Arundo\_take\_home\_challenge\_test\_set.csv



Arundo Analytics  
Houston | Oslo | Palo Alto  
www.arundo.com

The data columns are:

- `date`: yyyy-mm-dd format
- `calendar_code`: 0 or 1 (a code describing certain calendar events)
- `request_count`: an integer (the number of support requests received on that date)
- `site_count`: an integer (the number of sites operating on that date)
- `max_temp`: a float (max temperature for that day in degrees Celsius)
- `min_temp`: a float (min temperature for that day in degrees Celsius)
- `precipitation`: a float (millimeters of precipitation on that date)
- `events`: a string (description of weather events on that date)

`request_count` is absent from the test set.

## RESULTS AND FORMATTING

Your end product should consist of:

- A CSV file with a single column of predicted request counts for the dates 2014-10-01 through 2014-10-31 corresponding to the provided test set. Please do *not* include a header line. Please title this file "predicted\_request\_counts.csv".
- A report explaining your result and the approach you took.

The error of the model is measured as RMSE (root mean squared error).

Additionally you should answer the following questions in your report:

1. What other data might help your forecast?
2. From the training data, which day of the week has the highest mean # of support requests?
3. From the training data, which day of the week has the lowest mean # of support requests?

We would prefer that you use Python for this challenge, but you may use any programming language that you are comfortable with (e.g. Python, R, Julia, etc). Please make informative comments in your code.

- Include your prediction code and the code you used to answer questions 2 and 3.
  - Include any additional insights or visualizations if you would like.
-



Arundo Analytics  
Houston | Oslo | Palo Alto  
[www.arundo.com](http://www.arundo.com)

A Jupyter notebook file is the preferred report format, but any readable format (HTML, text, markdown, PDF, etc) to is acceptable (in addition to the CSV file).

Return your CSV of predicted request counts and report using the link provided to you in the email.