



Lesson 5 Exercises

Stefanie Molin

April 18, 2017

Let's do some practice problems to challenge your understanding.

1. Query Vertica for spend by day for the last 5 days for 3 clients of your choice. Reshape the long dataframe into a wide one with each client as a column.
2. Read in a 2M row excerpt of the  catalog from the provided textfile using the `fread()` function from `data.table` (this is faster than base R and automatically detects options). The file will be read into a `data.table`. (a) Drop the `sqlid` column. (b) Rename the `id` column `external_id`. (c) Make the `name` and `external_id` columns keys. (d) Select the `name` and `external_id` of the most expensive item and least expensive item. Limit the name of the selection to 35 characters.
3. Using the  catalog you obtained in (2), (a) find number of products with extra data containing the word "promo". (b) Find the unique promo offers and display a few of them. You will need to use a regular expression to find the value in the `extra` field, then you will need to use `str_match()` to find that pattern, and use a function from the `apply` family to get the results of applying that function on all values of `extra`.

Hint: If you are having trouble with the regex, you can take a few entries of the `extra` column in the `data.table` and work on adapting a regex here: <http://regexr.com/>. Be sure to look at how `str_match()` works and pick an appropriate `apply` family member; depending on how you do this, you may need to change the type of the object you give the function from the `apply` family.