Lesson 2 Solutions

Stefanie Molin

April 13, 2017

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

1. Create a CSV with the partner names of 5 accounts from your pod (named exactly how they would appear in the database) and the name of the AS on the account. Then read from the CSV into a dataframe. Run a simple query (not a dynamic one) for all accounts managed by the AS's in your pod, the partner ID, and the partner name. Merge the two data frames (inner join). Note that depending on how you define the column names you may have to use the `by` arguments to `merge()`.

```
# read in the CSV we created and take a look at it
(csvdf <- read.csv("sample_file.csv", stringsAsFactors = FALSE))
```

```
##          partner       AS
## 1 [redacted]
## 2
## 3
## 4
## 5
```

```
# capitalize partner names so they are like the database
csvdf$partner <- toupper(csvdf$partner)

# query Vertica for your pod's book of business (username/password already defined)
# QueryVertica() function already sourced
query <- "
SELECT
    [redacted]
FROM
    [redacted]
WHERE
    account_strategist_employee_id IN (
        SELECT
            employee_id
        FROM
            [redacted]
        WHERE
            full_name IN ('[redacted]', '[redacted]', '[redacted]')
        GROUP BY
            employee_id)
GROUP BY
    [redacted]
"

verticadf <- QueryVertica(username, query, password)

# inner join and map the columns
(pod <- merge(csvdf, verticadf, by.x = "[redacted]", by.y = "[redacted]"))
```

```
##          partner       AS merchant_id
## 1 [redacted]
## 2
## 3
## 4
## 5
```

2. Create another dataframe with 2 columns: the pod leader and the work level of the account for the accounts found in (1). Use `cbind()` to combine them into 1 dataframe. Use `rbind()` to add an additional row of your choosing to the dataframe.

```
# create podDetails dataframe
podDetails <- data.frame(pod_leader = "[redacted]",
                         work_level = c("Low", "Med", "High", "Low", "Med"),
                         stringsAsFactors = FALSE)

# add podDetails to the right of pod
(pod <- cbind(pod, podDetails))
```

```
##          partner       AS merchant_id pod_leader work_level
## 1 [redacted]                                          Low
## 2                                                     Med
## 3                                                    High
## 4                                                     Low
## 5                                                     Med
```

```
# make a new row
newRow <- data.frame(partner = '[redacted]', AS = '[redacted]', merchant_id = [redacted],
                     pod_leader = [redacted], work_level = "High",
                     stringsAsFactors = FALSE)

# add new row to the bottom of pod
(pod <- rbind(pod, newRow))
```

```
##          partner       AS merchant_id pod_leader work_level
## 1 [redacted]                                          Low
## 2                                                     Med
## 3                                                    High
## 4                                                     Low
## 5                                                     Med
## 6                                                    High
```

3. Adapt your query from (1) to query for only one AS but let the AS be specified at the function call rather than in the SQL itself. Write a function that takes care of the whole process.

```
#' @description Query Vertica for the book of business of given AS
#'
#' @param AS_name AS name
#' @param username Vertica login
#' @param password Vertica password to access the database
#'
#' @return dataframe of partner name and IDs
#'
#' @note Here for the purposes of this exercise the function has
#' default values for username meaning they don't have to be in
#' the call to QueryVertica(). (username/password are predefined).
#'

getBookOfBusiness <- function(AS_name, username = "s.molin", password){
  query <- "
  SELECT
   [redacted]
  FROM
    [redacted]
  WHERE
    account_strategist_employee_id IN (
        SELECT
            employee_id
        FROM
            [redacted]
        WHERE
            full_name = '%s'
        GROUP BY
            employee_id)
  GROUP BY
   [redacted]
  "

  # QueryVertica() function already sourced (username/password already defined)
  df <- QueryVertica(username, sprintf(query, toupper(AS_name)), password)

  # return the dataframe of results
  return(df)
}

head(getBookOfBusiness("[redacted]", password = password))
```

```
##   merchant_id       merchant_name
## 1 [redacted]
## 2
## 3
## 4
## 5
## 6
```