

ICA 1

Derien Weatherspoon

2023-01-18

```
order_details <- read.csv("order_details.csv")
orders <- read.csv("orders.csv")
territories <- read.csv("territories.csv")
regions <- read.csv("regions.csv")
employee_territories <- read.csv("employee_territories.csv")
employees <- read.csv("employees.csv")
customers <- read.csv("customers.csv")
shippers <- read.csv("shippers.csv")
suppliers <- read.csv("suppliers.csv")
products <- read.csv("products.csv")
categories <- read.csv("categories.csv")
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library/
##   dlopen(/Library/Frameworks/R.framework/Resources/modules//R_X11.so, 0x0006): Library not loaded: /
##   Referenced from: <51CE3E53-B0D5-3EC3-9C0E-347A62D97B61> /Library/Frameworks/R.framework/Versions/4
##   Reason: tried: '/opt/X11/lib/libSM.6.dylib' (no such file), '/System/Volumes/Preboot/Cryptexes/OS/
```

```
## Could not load tcltk. Will use slower R code instead.
```

```
## Loading required package: RSQLite
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##   filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##   intersect, setdiff, setequal, union
```

#Question 1: Join together the regions and territories datasets, leaving out duplicate columns.

```
Q1 <- sqldf("SELECT *
            FROM regions
            INNER JOIN territories")
```

```
head(Q1)
```

	regionID	regionDescription	territoryID	territoryDescription	regionID
## 1	1	Eastern	1581	Westboro	1
## 2	1	Eastern	1730	Bedford	1
## 3	1	Eastern	1833	Georgetow	1
## 4	1	Eastern	2116	Boston	1
## 5	1	Eastern	2139	Cambridge	1
## 6	1	Eastern	2184	Braintree	1

#Question 2: In the orders_detail data, calculate the average unitPrice for each productid. #summary

```
Q2 <- sqldf("SELECT avg(unitPrice), productId
            FROM order_details
            GROUP BY productId")
```

```
head(Q2)
```

	avg(unitPrice)	productID
## 1	17.14737	1
## 2	17.87727	2
## 3	9.50000	3
## 4	20.68000	4
## 5	19.61000	5
## 6	24.16667	6

#Question 3: Join together customers to orders, and, grouping by customer, #calculate the average freight cost

#for each, so that the result has average freight cost and customer name.

```
Q3 <- sqldf("SELECT avg(freight), customers
            #FROM orders
            #INNER JOIN customers
            #GROUP BY customers")
#couldn't get to work
Q3 <- sqldf("SELECT avg(freight), customers.customerID
            FROM customers
            INNER JOIN orders
            ON customers.customerID = orders.customerID
            GROUP BY customers.customerID")
```

```
Q3
```

	avg(freight)	customerID
## 1	37.59667	ALFKI
## 2	24.355000	ANATR
## 3	38.360000	ANTON
## 4	36.303846	AROUT
## 5	86.640000	BERGS
## 6	24.037143	BLAUS
## 7	56.696364	BLONP
## 8	63.723333	BOLID

## 9	79.874706	BONAP
## 10	56.710714	BOTTM
## 11	28.131000	BSBEV
## 12	12.126667	CACTU
## 13	3.250000	CENTC
## 14	45.905000	CHOPS
## 15	37.564000	COMMI
## 16	17.873333	CONSH
## 17	51.006667	DRACD
## 18	15.925000	DUMON
## 19	104.042500	EASTC
## 20	206.846333	ERNSH
## 21	33.250000	FAMIA
## 22	127.588000	FOLIG
## 23	88.320000	FOLKO
## 24	93.562667	FRANK
## 25	57.140000	FRANR
## 26	12.521667	FRANS
## 27	34.833750	FURIB
## 28	7.596000	GALED
## 29	56.827000	GODOS
## 30	35.820000	GOURL
## 31	98.873636	GREAL
## 32	33.900000	GROSR
## 33	51.769286	HANAR
## 34	69.953333	HILAA
## 35	41.416000	HUNGC
## 36	145.012632	HUNGO
## 37	36.365000	ISLAT
## 38	58.120000	KOENE
## 39	21.872500	LACOR
## 40	45.415714	LAMAI
## 41	3.306667	LAUGB
## 42	9.700000	LAZYK
## 43	67.802000	LEHMS
## 44	50.527500	LETSS
## 45	52.457857	LILAS
## 46	56.150833	LINOD
## 47	9.982500	LONEP
## 48	46.975000	MAGAA
## 49	65.558571	MAISD
## 50	107.247692	MEREP
## 51	64.408000	MORGK
## 52	12.530000	NORTS
## 53	61.328000	OCEAN
## 54	98.353000	OLDWO
## 55	86.274000	OTTIK
## 56	46.326667	PERIC
## 57	118.611000	PICCO
## 58	72.972000	PRINI
## 59	36.394444	QUEDE
## 60	152.515385	QUEEN
## 61	200.201071	QUICK
## 62	43.836000	RANCH

```
## 63 118.567222 RATTC
## 64 26.630000 REGGC
## 65 57.540909 RICAR
## 66 100.129000 RICSU
## 67 12.894000 ROMEY
## 68 45.916667 SANTG
## 69 215.603226 SAVEA
## 70 101.534444 SEVES
## 71 64.121429 SIMOB
## 72 27.070000 SPECB
## 73 62.074444 SPLIR
## 74 68.435833 SUPRD
## 75 65.522500 THEBI
## 76 43.320000 THECR
## 77 20.995000 TOMSP
## 78 47.563000 TORTU
## 79 45.760000 TRADH
## 80 23.336667 TRAIH
## 81 86.121818 VAFBE
## 82 49.325000 VICTE
## 83 11.682000 VINET
## 84 43.287000 WANDK
## 85 54.832000 WARTH
## 86 21.634444 WELLI
## 87 96.647143 WHITC
## 88 12.630000 WILMK
## 89 25.105714 WOLZA
```

#Question 4: Join together suppliers and shippers without a key (a Cartesian join)

```
Q4 <- sqldf("SELECT *
             FROM suppliers
             INNER JOIN shippers")
head(Q4)
```

```
## SupplierID      CompanyName      ContactName      ContactTitle
## 1          1      Exotic Liquids Charlotte Cooper Purchasing Manager
## 2          1      Exotic Liquids Charlotte Cooper Purchasing Manager
## 3          1      Exotic Liquids Charlotte Cooper Purchasing Manager
## 4          2 New Orleans Cajun Delights Shelley Burke Order Administrator
## 5          2 New Orleans Cajun Delights Shelley Burke Order Administrator
## 6          2 New Orleans Cajun Delights Shelley Burke Order Administrator
##           Address      City Region PostalCode Country      Phone Fax
## 1 49 Gilbert St.      London  NULL      EC1 4SD      UK (171) 555-2222 NULL
## 2 49 Gilbert St.      London  NULL      EC1 4SD      UK (171) 555-2222 NULL
## 3 49 Gilbert St.      London  NULL      EC1 4SD      UK (171) 555-2222 NULL
## 4 P.O. Box 78934 New Orleans LA      70117      USA (100) 555-4822 NULL
## 5 P.O. Box 78934 New Orleans LA      70117      USA (100) 555-4822 NULL
## 6 P.O. Box 78934 New Orleans LA      70117      USA (100) 555-4822 NULL
##           HomePage shipperID      companyName      phone
## 1          NULL          1 Speedy Express (503) 555-9831
## 2          NULL          2 United Package (503) 555-3199
## 3          NULL          3 Federal Shipping (503) 555-9931
## 4 #CAJUN.HTM#          1 Speedy Express (503) 555-9831
## 5 #CAJUN.HTM#          2 United Package (503) 555-3199
```

```
## 6 #CAJUN.HTM#          3 Federal Shipping (503) 555-9931
```

```
#Question 5: Join together employees to customers, with the result having  
#customer name, country, and employee first and last name. (2 joins, customerId, employeeId)  
Q5 <- sqldf("SELECT customers.contactName, customers.country, employees.firstName, employees.lastName  
            FROM customers  
            INNER JOIN employees  
            INNER JOIN orders  
            ON customers.customerID = orders.customerID  
               AND employees.employeeID = orders.employeeID")  
  
head(Q5)
```

```
##   contactName country firstName lastName  
## 1 Maria Anders Germany      Nancy  Davolio  
## 2 Maria Anders Germany      Nancy  Davolio  
## 3 Maria Anders Germany      Janet Leverling  
## 4 Maria Anders Germany Margaret  Peacock  
## 5 Maria Anders Germany Margaret  Peacock  
## 6 Maria Anders Germany Michael   Suyama
```