

Homework 7 - Group

10/29/2019

Group Members

Melinda Ding (mjd4), Henry Creamer (hmc5), Caleigh Page (cap9), Aaryan Jadhav (aj52), Cole Morgan (cmm16)

Question 5

“group10.db” is the name of our data base. “flights” is our primary source, and “DaysOnZillow_Msa” is our secondary source. Our secondary data, DaysOnZillow_Msa, lists the average days for a listing on Zillow to come off the housing market in many cities.

```
library(RSQLite)
```

```
## Warning: package 'RSQLite' was built under R version 3.5.2
```

```
dcon <- dbConnect(SQLite(), dbname = "group10.db")
dbListTables(dcon)
```

```
## [1] "DaysOnZillow_Msa" "flights"
```

Primary Source

Query all flights out of IAH airport.

```
res <- dbSendQuery(conn = dcon, "
SELECT YEAR, MONTH, DAY_OF_MONTH, DAY_OF_WEEK, ORIGIN, DEST,
       CRS_DEP_TIME, DEP_TIME, CRS_ARR_TIME, ARR_TIME, DISTANCE
FROM flights
WHERE ORIGIN = 'IAH';")
IAHdata <- dbFetch(res, -1)
```

```
## Warning in result_fetch(res@ptr, n = n): Column `DEP_TIME`: mixed type,
## first seen values of type real, coercing other values of type string
```

```
## Warning in result_fetch(res@ptr, n = n): Column `ARR_TIME`: mixed type,
## first seen values of type real, coercing other values of type string
```

```
dbClearResult(res)
head(IAHdata)
```

```
##   YEAR MONTH DAY_OF_MONTH DAY_OF_WEEK ORIGIN DEST CRS_DEP_TIME DEP_TIME
## 1 2018     1           27           6   IAH  PHX          1420      1413
## 2 2018     1           27           6   IAH  LGA          1847      1837
## 3 2018     1           27           6   IAH  ATL           735       731
## 4 2018     1           27           6   IAH  MSY          1836      1829
## 5 2018     1           27           6   IAH  HNL           940      1039
## 6 2018     1           27           6   IAH  BWI          1200      1158
##   CRS_ARR_TIME ARR_TIME DISTANCE
## 1          1618    1559     1009
## 2          2318    2240     1416
## 3          1043    1031       689
## 4          1948    1936       305
## 5          1425    1455     3904
## 6          1602    1546     1235
```

Query all flights out of IAH Airport on christmas day.

```
# Get the flights
```

```
res <- dbSendQuery(conn = dcon, "  
SELECT YEAR, MONTH, DAY_OF_MONTH, DAY_OF_WEEK, ORIGIN, DEST,  
       CRS_DEP_TIME, DEP_TIME, CRS_ARR_TIME, ARR_TIME, DISTANCE  
FROM flights  
WHERE ORIGIN = 'IAH'  
AND MONTH = 12  
AND DAY_OF_MONTH = 25;")  
IAH_xmas <- dbFetch(res, -1)
```

```
## Warning in result_fetch(res@ptr, n = n): Column `DEP_TIME`: mixed type,  
## first seen values of type real, coercing other values of type string
```

```
## Warning in result_fetch(res@ptr, n = n): Column `ARR_TIME`: mixed type,  
## first seen values of type real, coercing other values of type string
```

```
dbClearResult(res)  
head(IAH_xmas)
```

```
##   YEAR MONTH DAY_OF_MONTH DAY_OF_WEEK ORIGIN DEST CRS_DEP_TIME DEP_TIME  
## 1 2018    12           25           2   IAH  MIA           1118      1206  
## 2 2018    12           25           2   IAH  DFW            816       810  
## 3 2018    12           25           2   IAH  DFW          1428      1422  
## 4 2018    12           25           2   IAH  CLT            527       524  
## 5 2018    12           25           2   IAH  LAX            655       651  
## 6 2018    12           25           2   IAH  CLT          1350      1401  
##   CRS_ARR_TIME ARR_TIME DISTANCE  
## 1           1446      1525       964  
## 2           935       929       224  
## 3          1550      1528       224  
## 4           851       827       912  
## 5           851       829      1379  
## 6          1716      1712       912
```

Query the number of flights per month out of IAH.

```
res <- dbSendQuery(conn = dcon, "  
SELECT MONTH, count(*) as NUM_FLIGHTS  
FROM flights  
WHERE ORIGIN = 'IAH'  
GROUP BY MONTH;")  
monthly_count <- dbFetch(res, -1)  
dbClearResult(res)  
monthly_count
```

```
##   MONTH NUM_FLIGHTS  
## 1      1      14058  
## 2      2      12974  
## 3      3      14738  
## 4      4      14076  
## 5      5      14407  
## 6      6      14735  
## 7      7      15060  
## 8      8      15279  
## 9      9      14210
```

```
## 10      10      14890
## 11      11      14575
## 12      12      14804
```

Secondary Data

Find data for all Texas cities in 2018

```
res <- dbSendQuery(conn = dcon, "
SELECT RegionName, Jan_18, Feb_18, March_18, April_18, May_18, June_18, July_18, Aug_18, Sept_18, Oct_18,
FROM DaysOnZillow_Msa
WHERE RegionName LIKE '%TX%';")
allTexasListings <- dbFetch(res, -1)
```

```
## Warning in result_fetch(res@ptr, n = n): Column `Jan_18`: mixed type, first
## seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `Feb_18`: mixed type, first
## seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `March_18`: mixed type,
## first seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `April_18`: mixed type,
## first seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `May_18`: mixed type, first
## seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `June_18`: mixed type,
## first seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `July_18`: mixed type,
## first seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `Aug_18`: mixed type, first
## seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `Sept_18`: mixed type,
## first seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `Oct_18`: mixed type, first
## seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `Nov_18`: mixed type, first
## seen values of type real, coercing other values of type string
## Warning in result_fetch(res@ptr, n = n): Column `Dec_18`: mixed type, first
## seen values of type real, coercing other values of type string
```

```
dbClearResult(res)
allTexasListings
```

```
##           RegionName Jan_18 Feb_18 March_18 April_18 May_18 June_18
## 1 Dallas-Fort Worth, TX  69.0  64.0    53.0    48.0   48.0   52.0
## 2 Houston, TX          90.0  93.0    74.0    63.0   63.0   65.0
## 3 San Antonio, TX      84.0  80.0    64.0    61.0   58.0   59.0
## 4 Austin, TX           83.0  80.0    56.0    50.0   51.0   53.0
## 5 El Paso, TX          98.0 113.0    84.0    90.0   99.0   86.0
## 6 McAllen, TX           0.0   0.0     0.0   123.5  107.0  104.0
## 7 Corpus Christi, TX  146.0 125.0   105.0   119.0   83.0   80.0
```

## 8	Killeen, TX	100.0	91.5	96.0	78.0	67.0	69.0
## 9	Beaumont, TX	83.0	95.0	76.0	68.0	61.0	66.0
## 10	Lubbock, TX	86.0	102.0	76.0	65.5	69.5	66.0
## 11	Amarillo, TX	103.5	111.5	103.5	78.0	76.0	74.0
## 12	College Station, TX	82.0	87.0	67.0	67.0	65.0	69.0
## 13	Longview, TX	158.5	145.0	112.0	96.5	139.0	113.0
## 14	Tyler, TX	106.0	96.5	91.0	74.0	70.0	70.0
## 15	Abilene, TX	72.5	93.0	77.0	73.0	64.0	67.0
## 16	Wichita Falls, TX	98.0	110.0	65.0	68.0	59.0	62.0
## 17	Midland, TX	83.0	77.0	55.5	65.5	56.0	58.0
## 18	Odessa, TX	82.0	91.5	78.5	81.0	65.0	59.0
## 19	Sherman, TX	0.0	0.0	0.0	0.0	0.0	0.0
## 20	San Angelo, TX	92.0	101.0	69.0	76.0	64.0	61.5
## 21	Victoria, TX	104.0	110.5	95.0	67.0	73.5	67.0
## 22	Lufkin, TX	0.0	135.5	109.5	73.0	106.5	72.5
## 23	Athens, TX	112.0	130.0	142.0	110.0	63.0	66.0
## 24	Marshall, TX	0.0	0.0	0.0	0.0	0.0	98.0
## 25	Palestine, TX	0.0	0.0	0.0	0.0	0.0	0.0
## 26	Paris, TX	0.0	0.0	0.0	0.0	0.0	0.0
## 27	Corsicana, TX	0.0	0.0	0.0	0.0	0.0	0.0
## 28	Gainesville, TX	0.0	0.0	0.0	67.0	60.0	54.0
## 29	Big Spring, TX	0.0	0.0	0.0	0.0	0.0	0.0
## 30	Brenham, TX	0.0	0.0	0.0	52.0	69.5	65.0
##	July_18	Aug_18	Sept_18	Oct_18	Nov_18	Dec_18	
## 1	52.0	55.0	60.0	64.0	66.0	73.0	
## 2	68.0	69.0	76.0	82.0	90.0	89.0	
## 3	60.0	61.0	66.0	70.0	78.0	77.0	
## 4	57.0	62.0	64.0	71.0	80.0	77.0	
## 5	96.5	94.0	89.5	89.0	103.0	93.5	
## 6	99.0	102.0	103.0	119.5	109.0	119.0	
## 7	88.0	89.0	89.5	103.5	95.0	98.0	
## 8	67.0	63.0	72.0	76.0	77.0	80.0	
## 9	63.0	70.0	63.0	76.0	81.0	85.0	
## 10	64.0	64.0	70.0	75.0	83.0	98.0	
## 11	76.0	64.0	82.0	82.0	83.0	90.5	
## 12	69.0	69.0	77.0	104.0	102.0	105.0	
## 13	91.0	87.0	88.0	115.5	120.0	124.0	
## 14	68.5	69.0	84.0	86.0	80.0	91.0	
## 15	62.0	68.5	76.0	77.0	76.5	76.0	
## 16	56.0	54.5	77.0	74.0	70.0	70.0	
## 17	53.0	56.0	58.0	65.0	70.0	79.0	
## 18	61.5	67.0	62.0	71.0	55.5	64.0	
## 19	0.0	0.0	0.0	70.0	68.0	66.0	
## 20	64.0	64.0	70.0	67.0	73.0	97.0	
## 21	75.0	67.5	68.0	90.0	69.0	75.0	
## 22	70.0	89.5	81.0	99.0	144.0	106.5	
## 23	75.0	79.5	92.0	93.0	103.0	105.0	
## 24	97.0	110.0	122.0	101.0	116.0	103.0	
## 25	0.0	0.0	0.0	0.0	0.0	0.0	
## 26	0.0	0.0	0.0	0.0	0.0	0.0	
## 27	62.0	61.0	65.0	64.0	60.0	75.0	
## 28	67.0	62.5	92.0	95.0	124.0	92.0	
## 29	0.0	0.0	0.0	0.0	83.0	121.5	
## 30	75.0	95.0	87.0	81.5	67.5	67.0	

```

#Find all cities where days is less than 100 in December of 2018
res <- dbSendQuery(conn = dcon, "
SELECT RegionName
FROM DaysOnZillow_Msa
WHERE Dec_18 < 100;")
all100dayslessdf <- dbFetch(res, -1)
dbClearResult(res)
head(all100dayslessdf)

```

```

##                RegionName
## 1                United States
## 2 Los Angeles-Long Beach-Anaheim, CA
## 3                Dallas-Fort Worth, TX
## 4                Philadelphia, PA
## 5                Houston, TX
## 6                Washington, DC

```