

Project6__2019__EM

Elizabeth McGuckin

October 13, 2019

Worked with students Natalie McGuckin and Jayla Langford

Worked with TA Jun Kim

Question 1

```
#my bigDF = finalData  
finalData <- do.call(rbind, myresults)
```

Question 1a

```
finalRowName <- row.names(finalData)  
RownameDataframe <- as.data.frame(sapply(finalRowName, function(x) {strsplit(x, "/")}))  
finalData$country <- as.character(unlist(as.list(RownameDataframe[6,])))  
finalData$region <- as.character(unlist(as.list(RownameDataframe[7,])))  
finalData$city <- as.character(unlist(as.list(RownameDataframe[8,])))  
finalData$downloadeddate <- as.character(unlist(as.list(RownameDataframe[9,])))
```

Made new columns called country, region, city, and downloaded-date

Question 1b

```
rownames(finalData) <- NULL
```

Changed rownames

Question 2a

```
namesCol <- finalData$name  
  
AllCapitalLetter <- namesCol[namesCol == toupper(namesCol)]  
length(AllCapitalLetter)
```

```
## [1] 100918
```

[1] 100918

100918 entries of the name column are written in this dramatic way

```
AllContainNumbers <- namesCol[grepl("[[:digit:]]", namesCol)]  
length(AllContainNumbers)
```

```
## [1] 394404
```

[1] 394404

394404 entries of the name column are written in this dramatic way

```
AllContainSpecial <- namesCol[grepl('[[:punct:]]', namesCol)]  
length(AllContainSpecial)
```

```
## [1] 748559
```

[1] 748559

748559 entries of the name column are written in this dramatic way

Question 2b

```
reviewLos <- read.csv("/class/datamine/data/airbnb/united-states/ca/los-angeles/2019-07-08/visualisation")  
reviewsDate <- as.character(reviewLos$date)  
reviewsDate <- as.data.frame(sapply(reviewsDate, function(x) {strsplit(x, "-"))})  
years <- as.character(unlist(as.list(reviewsDate[1,])))  
table(years)
```

```
## years  
##   2009   2010   2011   2012   2013   2014   2015   2016   2017   2018   2019  
##    31    425   2152   5687  14729  35884  83780 172479 294319 476108 341559
```

number of reviews for properties in Los Angeles.

years

2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

31 425 2152 5687 14729 35884 83780 172479 294319 476108

2019

341559

Question 3a

```
American <- finalData[finalData$country == "united-states", ]  
tail(sort(tapply(American$price , American$city, mean)))
```

```
##      rhode-island      hawaii      austin      columbus      pacific-grove  
##      269.7520      317.5976      357.6869      376.9430      383.5779  
## twin-cities-msa  
##      558.2613
```

rhode-island hawaii austin columbus

269.7520 317.5976 357.6869 376.9430

pacific-grove twin-cities-msa

383.5779 558.2613

Twin city has the most expensive average AirBnB listing prices

Question 3b

```
twinCity <- American[American$city == "twin-cities-msa",]  
tail(sort(twinCity$price), n= 20)
```

```
## [1] 7000 7000 7150 7900 8000 9000 9400 10000 10000 10000 10000 10000 10000 10000  
## [15] 10000 10000 10000 10000 10000 15000
```

[1] 7000 7000 7150 7900 8000 9000 10000 10000 10000 10000

[12] 10000 10000 10000 10000 10000 10000 10000 10000 10000 15000

Rates for AirBnbs were really high for Super Bowl 2018 in Twin cities, therefore I think that these high rates are is mostly due to the Super Bowl in 2018