

Airbnb_Dataset

PrabhTalwar

2024-03-02

```
library(readr)
library(visdat)
library(VIM)
```

```
## Loading required package: colorspace
```

```
## Warning: package 'colorspace' was built under R version 4.2.3
```

```
## Loading required package: grid
```

```
## VIM is ready to use.
```

```
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
```

```
##
```

```
## Attaching package: 'VIM'
```

```
## The following object is masked from 'package:datasets':
```

```
##
```

```
##      sleep
```

```
Airbnb_Dataset <- read_csv("Airbnb_Dataset.csv")
```

```
## Rows: 11349 Columns: 15
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr  (5): rent_type, city, rating, neighbourhood, room_type
```

```
## dbl (10): id, host_id, latitude, longitude, price, minimum_nights, number_of...
```

```
##
```

```
## i Use 'spec()' to retrieve the full column specification for this data.
```

```
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
head(Airbnb_Dataset)
```

```
## # A tibble: 6 x 15
```

```
##       id rent_type  city    rating host_id neighbourhood latitude longitude
```

```
##    <dbl> <chr>      <chr>    <chr>   <dbl> <chr>                <dbl>   <dbl>
```

```
## 1 1936740 Rental Unit Vancouver <NA>   5510458 Shaughnessy      49.3    -123.
```

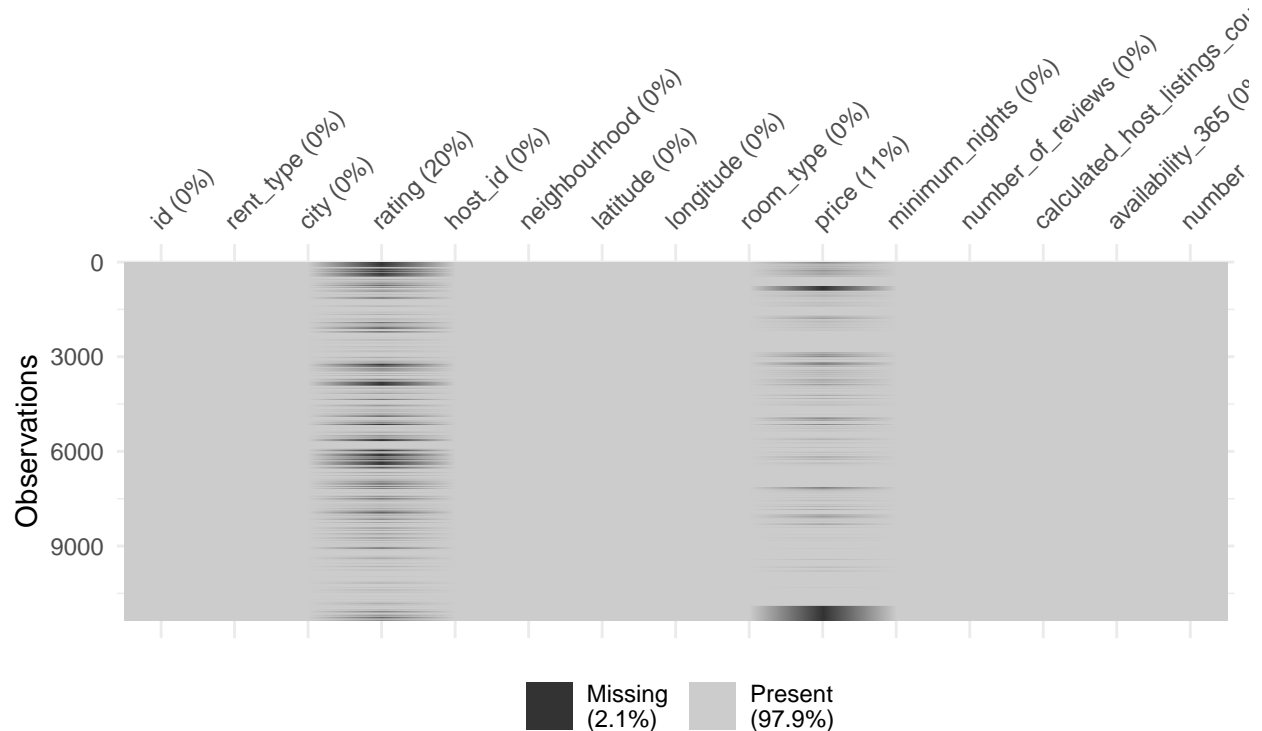
```
## 2 2876573 Rental Unit Vancouver <NA>      52926 Downtown      49.3      -123.
## 3 3902807 Townhouse   Vancouver <NA>      20209434 Kitsilano      49.3      -123.
## 4 4495075 Condo       Vancouver <NA>      9453431 West End        49.3      -123.
## 5 8017360 Rental Unit Vancouver <NA>      42335303 West End        49.3      -123.
## 6 8255521 Townhouse   Vancouver <NA>      38175858 West End        49.3      -123.
## # i 7 more variables: room_type <chr>, price <dbl>, minimum_nights <dbl>,
## #   number_of_reviews <dbl>, calculated_host_listings_count <dbl>,
## #   availability_365 <dbl>, number_of_reviews_ltm <dbl>
```

```
Airbnb_Dataset[sapply(Airbnb_Dataset, is.character)] <- lapply(
  Airbnb_Dataset[sapply(Airbnb_Dataset, is.character)], as.factor)

str(Airbnb_Dataset)
```

```
## spc_tbl_ [11,349 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ id : num [1:11349] 1936740 2876573 3902807 4495075 8017360 ...
## $ rent_type : Factor w/ 47 levels "Aparthotel","apartment",...: 34 34 43 16 34 4...
## $ city : Factor w/ 43 levels "Brentwood","Brentwood Bay",...: 41 41 41 41 4...
## $ rating : Factor w/ 114 levels "2.5","2.67","3",...: NA NA NA NA NA NA NA NA...
## $ host_id : num [1:11349] 5510458 52926 20209434 9453431 42335303 ...
## $ neighbourhood : Factor w/ 51 levels "Arbutus Ridge",...: 39 5 23 50 50 50 13 5 15 ...
## $ latitude : num [1:11349] 49.3 49.3 49.3 49.3 49.3 ...
## $ longitude : num [1:11349] -123 -123 -123 -123 -123 ...
## $ room_type : Factor w/ 4 levels "Entire home/apt",...: 1 1 1 1 1 1 1 1 1 ...
## $ price : num [1:11349] 100 163 395 115 NA 300 195 300 NA NA ...
## $ minimum_nights : num [1:11349] 30 30 30 30 30 30 30 30 30 ...
## $ number_of_reviews : num [1:11349] 0 0 0 0 0 0 0 0 0 ...
## $ calculated_host_listings_count: num [1:11349] 1 1 1 1 1 1 1 1 1 ...
## $ availability_365 : num [1:11349] 364 364 144 70 0 358 345 179 0 0 ...
## $ number_of_reviews_ltm : num [1:11349] 0 0 0 0 0 0 0 0 0 ...
## - attr(*, "spec")=
## .. cols(
## .. id = col_double(),
## .. rent_type = col_character(),
## .. city = col_character(),
## .. rating = col_character(),
## .. host_id = col_double(),
## .. neighbourhood = col_character(),
## .. latitude = col_double(),
## .. longitude = col_double(),
## .. room_type = col_character(),
## .. price = col_double(),
## .. minimum_nights = col_double(),
## .. number_of_reviews = col_double(),
## .. calculated_host_listings_count = col_double(),
## .. availability_365 = col_double(),
## .. number_of_reviews_ltm = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
vis_miss(Airbnb_Dataset)
```



#KNN-IMPUTATION

```
Airbnb <- kNN(Airbnb_Dataset, variable = c("rating", "price"), k = 7)

summary(Airbnb)
```

```
##           id           rent_type           city
## Min.   :1.591e+03   Home       :3471   Vancouver   :6681
## 1st Qu.:3.220e+07   Rental Unit:2437   Victoria     :3067
## Median :5.418e+07   Condo       :2057   Sooke        : 366
## Mean   :4.133e+17   Guest suite:1646   Salt Spring Island: 319
## 3rd Qu.:8.582e+17   Guesthouse  : 314   North Saanich : 134
## Max.   :1.053e+18   Cabin       : 199   Port Renfrew  : 120
##           (Other)   :1225   (Other)      : 662
##           rating      host_id      neighbourhood      latitude
## 5           :2732   Min.       : 1748   Downtown     :2413   Min.       :48.33
## New         :1269   1st Qu.: 25125836   Saanich       : 842   1st Qu.:48.46
## 4.92        : 311   Median :104426256   West End      : 512   Median :49.24
## 4.88        : 282   Mean   :177629778   Kitsilano     : 487   Mean   :48.95
## 4.94        : 271   3rd Qu.:308765087   Mount Pleasant: 438   3rd Qu.:49.27
## 4.93        : 264   Max.    :552034232   Langford      : 420   Max.    :49.29
## (Other):6220           (Other)      :6237
##           longitude      room_type      price      minimum_nights
## Min.     :-124.4   Entire home/apt:9450   Min.       : 14.0   Min.       : 1.00
## 1st Qu.  :-123.4   Hotel room       : 13   1st Qu.: 104.0   1st Qu.: 1.00
## Median   :-123.1   Private room    :1850   Median : 145.0   Median : 2.00
```

```
## Mean      :-123.3   Shared room      : 36   Mean      : 197.5   Mean      : 12.88
## 3rd Qu.: -123.1                                3rd Qu.: 210.0   3rd Qu.: 30.00
## Max.      :-123.0                                Max.      :28650.0   Max.      :700.00
##
## number_of_reviews calculated_host_listings_count availability_365
## Min.      : 0.00   Min.      : 1.000                                Min.      : 0
## 1st Qu.: 2.00   1st Qu.: 1.000                                1st Qu.: 8
## Median : 15.00   Median : 1.000                                Median :127
## Mean      : 49.08   Mean      : 7.133                                Mean      :148
## 3rd Qu.: 60.00   3rd Qu.: 4.000                                3rd Qu.:267
## Max.      :1006.00   Max.      :134.000                                Max.      :365
##
## number_of_reviews_ltm rating_imp      price_imp
## Min.      : 0.00      Mode :logical   Mode :logical
## 1st Qu.: 0.00      FALSE:9048   FALSE:10078
## Median : 4.00      TRUE :2301    TRUE :1271
## Mean      : 14.83
## 3rd Qu.: 22.00
## Max.      :182.00
##
```

```
?write.csv
```

```
## starting httpd help server ... done
```

```
getwd()
```

```
## [1] "C:/Users/prabh/OneDrive/Documents"
```

```
write.csv(Airbnb, file = "Airbnb.csv")
```