Eksplorasi Data Airbnb

# Introduction

Airbnb, Inc (Airbnb) adalah *marketplace* untuk memesan penginapan/*homestay* atau paket liburan, secara online melalui aplikasi. Pengguna cukup mengunduh aplikasi Airbnb melalui Google Play/Apple Store, kemudian pengguna akan mencari lokasi properti penginapan, berikut tanggal menginap. Aplikasi kemudian akan menampilkan daftar properti yang tersedia pada lokasi dan tanggal tersebut. Pengguna kemudian membayarkan *fee* kepada Airbnb, untuk diteruskan kepada pemilik properti. Dalam hal ini, perusahaan Airbnb tidak memiliki aset properti yang ditawarkan, namun Airbnb bertindak sebagai perantara (antara pemilik properti [host] dan pengguna). Airbnb berkantor pusat di San Fransisco, California, Amerika Serikat.

Dari properti yang tersedia di Airbnb, apakah faktor penentu harga properti di Airbnb?

# Methodology

## Data Understanding

SteveZheng (Steve Zheng), merupakan kontributor di Kaggle.com yang mengunggah dataset [“Airbnb price prediction”](https://www.kaggle.com/stevezhenghp/airbnb-price-prediction). Dataset merupakan data properti Airbnb yang ada di Amerika Serikat. Dataset terdiri dari **74111** baris data, dengan **29** fitur

Deskripsi fitur-fitur yang tersedia (menurut interpretasi penulis, dikarenakan tidak adanya dokumentasi resmi dari kontributor) dan yang digunakan, adalah sebagai berikut :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Feature | Description | Fitur digunakan? | Response/Predictor |
| 1 | id | Nomor ID properti | Tidak | - |
| 2 | log\_price | Harga sewa properti (dalam bentuk log) | Ya | Response |
| 3 | property\_type | Jenis properti (contoh : apartemen, rumah, dsb) | Ya | Predictor |
| 4 | room\_type | Jenis kamar yang disewakan. Contoh : 1) entire home/apt –> seluruh rumah/apartemen, 2) private room –> kamar pribadi yang disewakan, 3) Shared room –> kamar yang dibagi dengan pengguna lain) | Ya | Predictor |
| 5 | amenities | Kelengkapan lain dari properti | Tidak | - |
| 6 | accomodates | Jumlah kapasitas orang yang dapat menginap dalam properti yang disewakan | Ya | Predictor |
| 7 | bathrooms | Jumlah kamar mandi tersedia | Ya | Predictor |
| 8 | bed\_type | Jenis kasur yang disediakan : Real Bed, Futon, Pull-out Sofa, Couch atau Airbed | Ya | Predictor |
| 9 | cancellation\_policy | Persyaratan pembatalan | Ya | Predictor |
| 10 | cleaning\_fee | Ada atau tidaknya biaya pembersihan kamar | Ya | Predictor |
| 11 | city | Kota letak properti berada | Ya | Predictor |
| 12 | description | Deskripsi properti | Tidak | - |
| 13 | first\_review | Tanggal review pertama | Ya | Predictor |
| 14 | host\_has\_profile\_pic | Penanda apakah *host* menampilkan foto profil pada akun Airbnb | Tidak | - |
| 15 | host\_identity\_verified | Penanda apakah *host* telah diverifikasi oleh Airbnb | Tidak | - |
| 16 | host\_response\_rate | Kecepatan *host* dalam merespons setiap pertanyaan yang masuk | Tidak | - |
| 17 | host\_since | Tanggal *host* mulai memasukkan propertinya pada Airbnb | Tidak | - |
| 18 | instant\_bookable | Penanda apakah properti dapat langsung di book tanpa perlu adanya konfirmasi dari *host* | Ya | Predictor |
| 19 | last\_review | Tanggal review terakhir | Tidak | - |
| 20 | latitude | *Latitude* lokasi properti | Ya | Predictor |
| 21 | longitude | *Longitude* lokasi properti | Ya | Predictor |
| 22 | name | Nama properti | Tidak | - |
| 23 | neighbourhood | nama wilayah | Tidak | - |
| 24 | number\_of\_reviews | jumlah review yang diberikan | Tidak | - |
| 25 | review\_scores\_rating | Skor review | Ya | Predictor |
| 26 | thumbnail\_url | URL link gambar dari properti | Tidak | - |
| 27 | zipcode | Kodepos properti | Tidak | - |
| 28 | bedrooms | Banyak kamar tidur | Ya | Predictor |
| 29 | beds | Banyak tempat tidur yang disediakan | Ya | Predictor |

Adapun ringkasan dari nilai pada setiap feature adalah sebagai berikut :

library(tidyverse)

## -- Attaching packages -------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.2  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ----------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

airBNB<-read\_csv("train.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## id = col\_double(),  
## log\_price = col\_double(),  
## accommodates = col\_double(),  
## bathrooms = col\_double(),  
## cleaning\_fee = col\_logical(),  
## first\_review = col\_date(format = ""),  
## host\_has\_profile\_pic = col\_logical(),  
## host\_identity\_verified = col\_logical(),  
## host\_since = col\_date(format = ""),  
## instant\_bookable = col\_logical(),  
## last\_review = col\_date(format = ""),  
## latitude = col\_double(),  
## longitude = col\_double(),  
## number\_of\_reviews = col\_double(),  
## review\_scores\_rating = col\_double(),  
## bedrooms = col\_double(),  
## beds = col\_double()  
## )

## See spec(...) for full column specifications.

summary(airBNB)

## id log\_price property\_type room\_type   
## Min. : 344 Min. :0.000 Length:74111 Length:74111   
## 1st Qu.: 6261964 1st Qu.:4.317 Class :character Class :character   
## Median :12254147 Median :4.710 Mode :character Mode :character   
## Mean :11266617 Mean :4.782   
## 3rd Qu.:16402260 3rd Qu.:5.220   
## Max. :21230903 Max. :7.600   
##   
## amenities accommodates bathrooms bed\_type   
## Length:74111 Min. : 1.000 Min. :0.000 Length:74111   
## Class :character 1st Qu.: 2.000 1st Qu.:1.000 Class :character   
## Mode :character Median : 2.000 Median :1.000 Mode :character   
## Mean : 3.155 Mean :1.235   
## 3rd Qu.: 4.000 3rd Qu.:1.000   
## Max. :16.000 Max. :8.000   
## NA's :200   
## cancellation\_policy cleaning\_fee city description   
## Length:74111 Mode :logical Length:74111 Length:74111   
## Class :character FALSE:19708 Class :character Class :character   
## Mode :character TRUE :54403 Mode :character Mode :character   
##   
##   
##   
##   
## first\_review host\_has\_profile\_pic host\_identity\_verified  
## Min. :2008-11-17 Mode :logical Mode :logical   
## 1st Qu.:2015-06-28 FALSE:226 FALSE:24175   
## Median :2016-05-25 TRUE :73697 TRUE :49748   
## Mean :2016-01-14 NA's :188 NA's :188   
## 3rd Qu.:2017-01-02   
## Max. :2017-10-05   
## NA's :15864   
## host\_response\_rate host\_since instant\_bookable  
## Length:74111 Min. :2008-03-03 Mode :logical   
## Class :character 1st Qu.:2013-04-21 FALSE:54660   
## Mode :character Median :2014-09-28 TRUE :19451   
## Mean :2014-07-21   
## 3rd Qu.:2015-12-22   
## Max. :2017-10-04   
## NA's :188   
## last\_review latitude longitude name   
## Min. :2009-01-21 Min. :33.34 Min. :-122.51 Length:74111   
## 1st Qu.:2017-01-09 1st Qu.:34.13 1st Qu.:-118.34 Class :character   
## Median :2017-04-28 Median :40.66 Median : -77.00 Mode :character   
## Mean :2017-03-14 Mean :38.45 Mean : -92.40   
## 3rd Qu.:2017-09-08 3rd Qu.:40.75 3rd Qu.: -73.95   
## Max. :2017-10-05 Max. :42.39 Max. : -70.99   
## NA's :15827   
## neighbourhood number\_of\_reviews review\_scores\_rating  
## Length:74111 Min. : 0.0 Min. : 20.00   
## Class :character 1st Qu.: 1.0 1st Qu.: 92.00   
## Mode :character Median : 6.0 Median : 96.00   
## Mean : 20.9 Mean : 94.07   
## 3rd Qu.: 23.0 3rd Qu.:100.00   
## Max. :605.0 Max. :100.00   
## NA's :16722   
## thumbnail\_url zipcode bedrooms beds   
## Length:74111 Length:74111 Min. : 0.000 Min. : 0.000   
## Class :character Class :character 1st Qu.: 1.000 1st Qu.: 1.000   
## Mode :character Mode :character Median : 1.000 Median : 1.000   
## Mean : 1.266 Mean : 1.711   
## 3rd Qu.: 1.000 3rd Qu.: 2.000   
## Max. :10.000 Max. :18.000   
## NA's :91 NA's :131

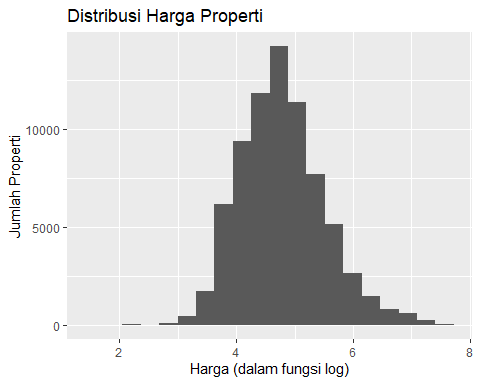
Dari summary diatas, diketahui bahwa terdapat beberapa feature yang memiliki nilai NA. Fitur review\_scores\_rating memiliki NA yang cukup tinggi. Hal ini diakibatkan karena properti tersebut belum memiliki rating. Diduga bahwa properti tersebut adalah properti baru. Oleh karena itu, dataset akan dibagi menjadi 2 *cluster*, yakni dataset untuk properti baru (belum pernah dipesan) dan properti lama (pernah dipesan). Adapun *instance* yang memiliki nilai NA untuk setiap cluster nya akan diabaikan dan tidak digunakan untuk proses pemodelan.

Dari data terdapat *instance* yang nilai log\_price nya adalah 0. Terkait hal ini, instance tersebut tidak akan dimasukkan kedalam proses modelling

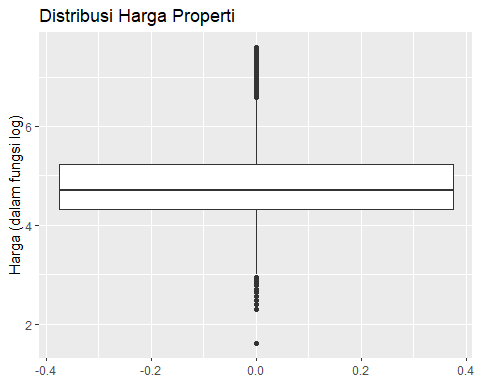
airBNB<-subset(airBNB,log\_price>0)

Secara garis besar, data yang digunakan memiliki sebaran harga sebagai berikut :

ggplot(airBNB,aes(log\_price))+geom\_histogram(bins=20)+labs(title = "Distribusi Harga Properti", x="Harga (dalam fungsi log)", y="Jumlah Properti")

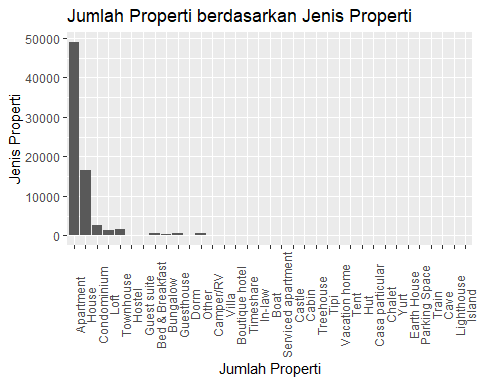


ggplot(airBNB,aes(x=NULL, y=log\_price))+geom\_boxplot()+labs(title = "Distribusi Harga Properti", y="Harga (dalam fungsi log)",x=NULL)

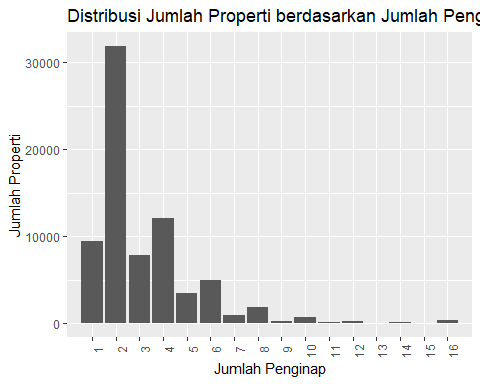
 Dari gambar terlihat bahwa harga mayoritas berada pada rentang 4.3 - 5.2

Jumlah properti berdasarkan jenis nya, dapat ditampilkan dalam grafik sebagai berikut :

ggplot(airBNB,aes(property\_type))+geom\_bar()+theme(axis.text.x = element\_text(angle = 90))+labs(title = "Jumlah Properti berdasarkan Jenis Properti", y="Jenis Properti",x="Jumlah Properti")+ scale\_x\_discrete(limits=c(unique(airBNB$property\_type)))

 Pada grafik diatas, terlihat bahwa dataset didominasi dengan properti berjenis Apartement (apartemen) dan House (rumah). Distribusi jumlah properti berdasarkan jumlah penginap dapat digambarkan sebagai berikut :

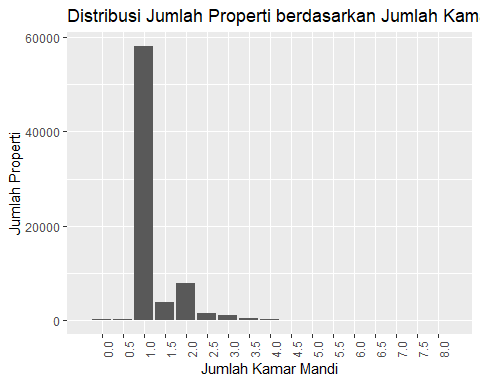
ggplot(airBNB,aes(accommodates))+geom\_bar()+theme(axis.text.x = element\_text(angle = 90))+labs(title = "Distribusi Jumlah Properti berdasarkan Jumlah Penginap yang bisa Ditampung", y="Jumlah Properti",x="Jumlah Penginap") + scale\_x\_discrete(limits=c(1:1:16))

 Dari grafik tersebut, mayoritas properti dapat menampung maksimal 2 penginap. Terdapat beberapa properti yang dapat menampung sampai 16 penginap.

Berdasarkan jumlah kamar mandi yang tersedia pada setiap properti, dapat digambarkan dalam grafik distribusi sebagai berikut :

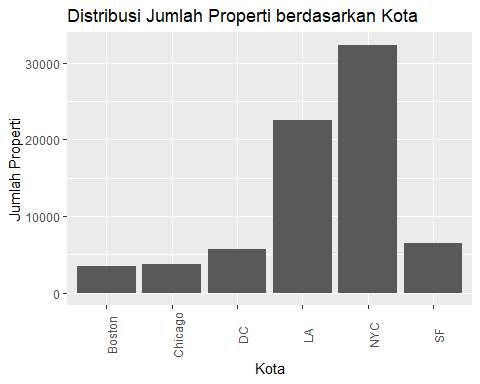
ggplot(airBNB,aes(bathrooms))+geom\_bar()+theme(axis.text.x = element\_text(angle = 90))+labs(title = "Distribusi Jumlah Properti berdasarkan Jumlah Kamar Mandi", y="Jumlah Properti",x="Jumlah Kamar Mandi") + scale\_x\_discrete(limits=seq(from = 0, to = 8, by = 0.5))

## Warning: Removed 200 rows containing non-finite values (stat\_count).

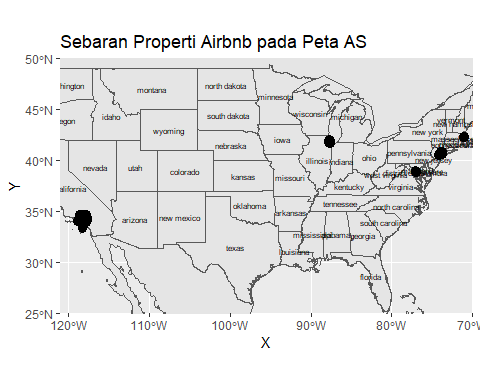
 Dari grafik dapat ditunjukkan bahwa minimal sebuah properti memiliki 1 kamar mandi. Beberapa properti memiliki kelipatan 0.5, ini artinya bahwa properti tersebut hanya memiliki kloset dan wastafel, namun tidak memiliki *bath tub* [“what does 1 1/2 bath means?”](https://www.homes.com/question/15848/what-does-1-1-2-bath-means/).

Berdasarkan kota, distribusi properti dapat digambarkan sebagai berikut :

ggplot(airBNB,aes(city))+geom\_bar()+theme(axis.text.x = element\_text(angle = 90))+labs(title = "Distribusi Jumlah Properti berdasarkan Kota", y="Jumlah Properti",x="Kota")

 Data properti yang digunakan mayoritas terletak di New York City (sebanyak lebih dari 30.000 data), sedangkan Los Angeles menempati tempat kedua. Apabila lokasi properti digambarkan dalam peta AS, maka dapat digambarkan sebagai berikut :

library(mapdata)  
library(ggplot2)  
library(rnaturalearth)  
library(rnaturalearthdata)  
library(rgeos)  
library(maps)  
library(sf)  
library(tools)  
  
world <- ne\_countries(scale = "medium", returnclass = "sf")  
sites <- data.frame(longitude = airBNB$longitude,   
 latitude = airBNB$latitude  
 )  
states <- st\_as\_sf(map("state", plot = FALSE, fill = TRUE))  
states <- cbind(states, st\_coordinates(st\_centroid(states)))  
  
  
ggplot(data = world) +  
 geom\_sf() +  
 geom\_sf(data = states, fill = NA) +   
 geom\_text(data = states, aes(X, Y, label = ID), size = 2) +  
 geom\_point(data = sites, aes(x = longitude, y = latitude), size = 3,   
 shape = 16, fill = "red") +  
 coord\_sf(xlim = c(-121, -70), ylim = c(25, 50), expand = FALSE)+  
 labs(title = "Sebaran Properti Airbnb pada Peta AS")



## Metode Pemodelan

Metode yang digunakan untuk pemodelan adalah dengan menggunakan regresi linear. Dengan menggunakan software R, pemodelan ini dilakukan dengan memanggil fungsi lm(), dengan fitur price sebagai respon dan variabel lain sebagai prediktor. Nilai alpha yang digunakan untuk menguji hipotesis adalah 0.05.

Fungsi lm() di R akan otomatis mengubah fitur yang bertipe kategorikal menjadi *dummy variable*

# Result

## Cluster I : Properti Lama

Tahap pertama dilakukan *subsetting* terhadap dataset yang memiliki review\_scores\_rating lebih dari 0 (tidak NA). Data tersebut disimpan dalam variabel bernama subset\_airBNB\_lama

subset\_airBNB\_lama<-subset(airBNB,review\_scores\_rating>=0)

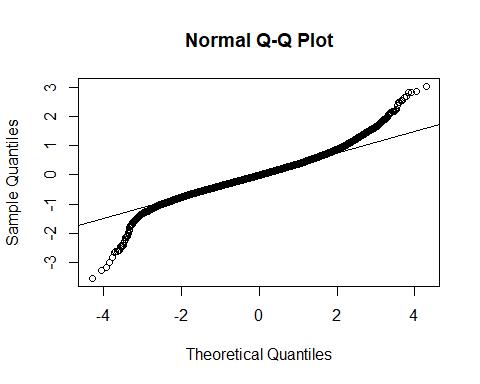
Pemodelan dibuat dengan memanggil fungsi lm() terhadap subset data properti lama

linearModel\_airBNB\_lama<-lm(log\_price ~ property\_type+room\_type+accommodates+bathrooms+  
 bed\_type+cancellation\_policy+cleaning\_fee+city+instant\_bookable+  
 latitude+longitude+review\_scores\_rating+bedrooms+beds, data=subset\_airBNB\_lama)  
summary(linearModel\_airBNB\_lama)

##   
## Call:  
## lm(formula = log\_price ~ property\_type + room\_type + accommodates +   
## bathrooms + bed\_type + cancellation\_policy + cleaning\_fee +   
## city + instant\_bookable + latitude + longitude + review\_scores\_rating +   
## bedrooms + beds, data = subset\_airBNB\_lama)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5482 -0.2615 -0.0133 0.2440 3.0319   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -7.724e+01 1.728e+00 -44.694 < 2e-16  
## property\_typeBed & Breakfast 1.509e-01 2.199e-02 6.861 6.89e-12  
## property\_typeBoat 2.494e-01 6.030e-02 4.136 3.54e-05  
## property\_typeBoutique hotel 2.184e-01 5.972e-02 3.657 0.000256  
## property\_typeBungalow -2.854e-02 2.360e-02 -1.209 0.226579  
## property\_typeCabin -1.227e-01 5.159e-02 -2.378 0.017421  
## property\_typeCamper/RV -3.461e-01 5.042e-02 -6.865 6.72e-12  
## property\_typeCastle 3.943e-01 1.133e-01 3.482 0.000498  
## property\_typeCave 2.787e-01 2.886e-01 0.965 0.334321  
## property\_typeChalet 7.359e-02 1.826e-01 0.403 0.686931  
## property\_typeCondominium 1.176e-01 9.567e-03 12.291 < 2e-16  
## property\_typeDorm -4.263e-01 4.082e-02 -10.443 < 2e-16  
## property\_typeEarth House 5.803e-02 2.356e-01 0.246 0.805471  
## property\_typeGuest suite -9.819e-02 4.114e-02 -2.387 0.017003  
## property\_typeGuesthouse -8.467e-02 2.051e-02 -4.127 3.68e-05  
## property\_typeHostel -5.363e-01 5.615e-02 -9.552 < 2e-16  
## property\_typeHouse -2.099e-02 4.782e-03 -4.389 1.14e-05  
## property\_typeHut -3.463e-01 1.543e-01 -2.244 0.024844  
## property\_typeIn-law -2.284e-01 5.167e-02 -4.421 9.84e-06  
## property\_typeIsland 9.392e-01 4.086e-01 2.299 0.021514  
## property\_typeLoft 1.851e-01 1.307e-02 14.160 < 2e-16  
## property\_typeOther 1.915e-02 2.011e-02 0.952 0.341141  
## property\_typeServiced apartment 2.038e-01 1.054e-01 1.933 0.053218  
## property\_typeTent -4.739e-01 1.093e-01 -4.336 1.45e-05  
## property\_typeTimeshare 5.492e-01 7.114e-02 7.720 1.18e-14  
## property\_typeTipi 3.746e-01 2.358e-01 1.588 0.112202  
## property\_typeTownhouse -5.774e-04 1.174e-02 -0.049 0.960770  
## property\_typeTrain 3.723e-01 2.887e-01 1.290 0.197160  
## property\_typeTreehouse 1.143e-01 1.667e-01 0.686 0.492932  
## property\_typeVacation home 3.613e-01 1.667e-01 2.168 0.030185  
## property\_typeVilla 2.072e-01 3.670e-02 5.648 1.64e-08  
## property\_typeYurt -1.325e-01 1.667e-01 -0.794 0.427015  
## room\_typePrivate room -5.990e-01 4.284e-03 -139.833 < 2e-16  
## room\_typeShared room -1.051e+00 1.198e-02 -87.722 < 2e-16  
## accommodates 7.780e-02 1.652e-03 47.085 < 2e-16  
## bathrooms 1.392e-01 3.899e-03 35.707 < 2e-16  
## bed\_typeCouch 7.821e-02 3.949e-02 1.981 0.047619  
## bed\_typeFuton -1.816e-02 2.773e-02 -0.655 0.512605  
## bed\_typePull-out Sofa 5.650e-02 2.879e-02 1.962 0.049744  
## bed\_typeReal Bed 6.205e-02 2.236e-02 2.776 0.005510  
## cancellation\_policymoderate 1.335e-02 4.920e-03 2.713 0.006678  
## cancellation\_policystrict 4.228e-02 4.602e-03 9.187 < 2e-16  
## cancellation\_policysuper\_strict\_30 3.711e-01 4.655e-02 7.971 1.60e-15  
## cancellation\_policysuper\_strict\_60 7.015e-01 1.294e-01 5.419 6.02e-08  
## cleaning\_feeTRUE 8.236e-04 4.515e-03 0.182 0.855268  
## cityChicago -1.761e+01 3.278e-01 -53.718 < 2e-16  
## cityDC -5.777e+00 1.503e-01 -38.431 < 2e-16  
## cityLA -4.830e+01 9.662e-01 -49.989 < 2e-16  
## cityNYC -2.770e+00 7.224e-02 -38.338 < 2e-16  
## citySF -5.276e+01 1.025e+00 -51.469 < 2e-16  
## instant\_bookableTRUE -4.493e-02 3.908e-03 -11.499 < 2e-16  
## latitude 1.598e-01 2.566e-02 6.227 4.78e-10  
## longitude -1.047e+00 1.971e-02 -53.103 < 2e-16  
## review\_scores\_rating 5.598e-03 2.210e-04 25.326 < 2e-16  
## bedrooms 1.438e-01 3.310e-03 43.455 < 2e-16  
## beds -3.900e-02 2.569e-03 -15.181 < 2e-16  
##   
## (Intercept) \*\*\*  
## property\_typeBed & Breakfast \*\*\*  
## property\_typeBoat \*\*\*  
## property\_typeBoutique hotel \*\*\*  
## property\_typeBungalow   
## property\_typeCabin \*   
## property\_typeCamper/RV \*\*\*  
## property\_typeCastle \*\*\*  
## property\_typeCave   
## property\_typeChalet   
## property\_typeCondominium \*\*\*  
## property\_typeDorm \*\*\*  
## property\_typeEarth House   
## property\_typeGuest suite \*   
## property\_typeGuesthouse \*\*\*  
## property\_typeHostel \*\*\*  
## property\_typeHouse \*\*\*  
## property\_typeHut \*   
## property\_typeIn-law \*\*\*  
## property\_typeIsland \*   
## property\_typeLoft \*\*\*  
## property\_typeOther   
## property\_typeServiced apartment .   
## property\_typeTent \*\*\*  
## property\_typeTimeshare \*\*\*  
## property\_typeTipi   
## property\_typeTownhouse   
## property\_typeTrain   
## property\_typeTreehouse   
## property\_typeVacation home \*   
## property\_typeVilla \*\*\*  
## property\_typeYurt   
## room\_typePrivate room \*\*\*  
## room\_typeShared room \*\*\*  
## accommodates \*\*\*  
## bathrooms \*\*\*  
## bed\_typeCouch \*   
## bed\_typeFuton   
## bed\_typePull-out Sofa \*   
## bed\_typeReal Bed \*\*   
## cancellation\_policymoderate \*\*   
## cancellation\_policystrict \*\*\*  
## cancellation\_policysuper\_strict\_30 \*\*\*  
## cancellation\_policysuper\_strict\_60 \*\*\*  
## cleaning\_feeTRUE   
## cityChicago \*\*\*  
## cityDC \*\*\*  
## cityLA \*\*\*  
## cityNYC \*\*\*  
## citySF \*\*\*  
## instant\_bookableTRUE \*\*\*  
## latitude \*\*\*  
## longitude \*\*\*  
## review\_scores\_rating \*\*\*  
## bedrooms \*\*\*  
## beds \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4081 on 57072 degrees of freedom  
## (260 observations deleted due to missingness)  
## Multiple R-squared: 0.6278, Adjusted R-squared: 0.6275   
## F-statistic: 1750 on 55 and 57072 DF, p-value: < 2.2e-16

Dari hasil tersebut didapatkan nilai akurasi, yang diukur dengan R-squared sebesar 0.627 (62.7%). Adapun distribusi dari nilai residual dapat digambarkan sebagai berikut :

#ggqqplot(linearModel\_airBNB\_lama[["residuals"]])  
qqnorm(linearModel\_airBNB\_lama[["residuals"]])  
qqline(linearModel\_airBNB\_lama[["residuals"]])



#ggdensity(linearModel\_airBNB\_lama[["residuals"]])  
#shapiro.test(linearModel\_airBNB\_lama[["residuals"]])

Dari grafik tersebut terlihat bahwa residual dari pemodelan linear tidak terdistribusi normal. Model kemudian diperbaiki dengan menggunakan *feature selection*, dengan metode forward & backward (both), dengan sintaks sebagai berikut :

library(mlbench)  
library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

fs\_both<-stepAIC(linearModel\_airBNB\_lama,direction = "both")

## Start: AIC=-102345.5  
## log\_price ~ property\_type + room\_type + accommodates + bathrooms +   
## bed\_type + cancellation\_policy + cleaning\_fee + city + instant\_bookable +   
## latitude + longitude + review\_scores\_rating + bedrooms +   
## beds  
##   
## Df Sum of Sq RSS AIC  
## - cleaning\_fee 1 0.0 9505.0 -102347  
## <none> 9505.0 -102345  
## - bed\_type 4 5.0 9510.1 -102323  
## - latitude 1 6.5 9511.5 -102309  
## - instant\_bookable 1 22.0 9527.1 -102215  
## - cancellation\_policy 4 30.4 9535.4 -102171  
## - beds 1 38.4 9543.4 -102117  
## - review\_scores\_rating 1 106.8 9611.9 -101709  
## - property\_type 31 152.3 9657.3 -101499  
## - bathrooms 1 212.3 9717.4 -101085  
## - bedrooms 1 314.5 9819.5 -100488  
## - accommodates 1 369.2 9874.3 -100170  
## - longitude 1 469.6 9974.7 -99592  
## - city 5 1184.4 10689.4 -95647  
## - room\_type 2 3814.8 13319.8 -83073  
##   
## Step: AIC=-102347.4  
## log\_price ~ property\_type + room\_type + accommodates + bathrooms +   
## bed\_type + cancellation\_policy + city + instant\_bookable +   
## latitude + longitude + review\_scores\_rating + bedrooms +   
## beds  
##   
## Df Sum of Sq RSS AIC  
## <none> 9505.0 -102347  
## + cleaning\_fee 1 0.0 9505.0 -102345  
## - bed\_type 4 5.0 9510.1 -102325  
## - latitude 1 6.5 9511.5 -102311  
## - instant\_bookable 1 22.0 9527.1 -102217  
## - cancellation\_policy 4 31.2 9536.3 -102168  
## - beds 1 38.4 9543.5 -102119  
## - review\_scores\_rating 1 106.9 9612.0 -101710  
## - property\_type 31 152.3 9657.4 -101501  
## - bathrooms 1 212.3 9717.4 -101087  
## - bedrooms 1 314.5 9819.6 -100490  
## - accommodates 1 370.2 9875.2 -100167  
## - longitude 1 470.0 9975.1 -99592  
## - city 5 1185.2 10690.2 -95644  
## - room\_type 2 3875.4 13380.4 -82815

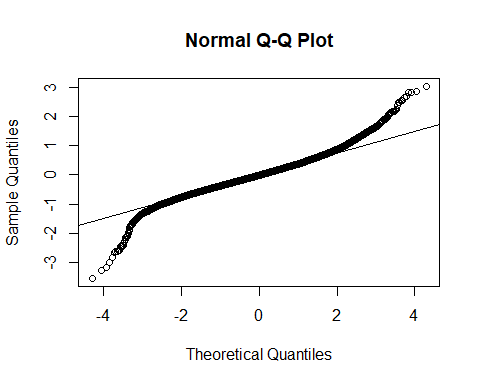
Dari proses feature selection, didapatkan hasil bahwa variabel cleaning\_fee dapat diabaikan dalam pemodelan. Fitur yang sudah diseleksi kemudian dimodelkan dengan regresi linear sebagai berikut :

linearModel\_airBNB\_lama\_fs<-lm(formula=fs\_both[["terms"]],data=subset\_airBNB\_lama)  
summary(linearModel\_airBNB\_lama\_fs)

##   
## Call:  
## lm(formula = fs\_both[["terms"]], data = subset\_airBNB\_lama)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5487 -0.2615 -0.0133 0.2441 3.0314   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -7.724e+01 1.728e+00 -44.697 < 2e-16  
## property\_typeBed & Breakfast 1.508e-01 2.198e-02 6.859 7.00e-12  
## property\_typeBoat 2.492e-01 6.029e-02 4.133 3.58e-05  
## property\_typeBoutique hotel 2.183e-01 5.972e-02 3.655 0.000257  
## property\_typeBungalow -2.848e-02 2.360e-02 -1.207 0.227444  
## property\_typeCabin -1.227e-01 5.159e-02 -2.379 0.017349  
## property\_typeCamper/RV -3.462e-01 5.042e-02 -6.866 6.65e-12  
## property\_typeCastle 3.944e-01 1.133e-01 3.483 0.000497  
## property\_typeCave 2.784e-01 2.886e-01 0.965 0.334773  
## property\_typeChalet 7.372e-02 1.826e-01 0.404 0.686430  
## property\_typeCondominium 1.176e-01 9.567e-03 12.293 < 2e-16  
## property\_typeDorm -4.265e-01 4.081e-02 -10.450 < 2e-16  
## property\_typeEarth House 5.813e-02 2.356e-01 0.247 0.805152  
## property\_typeGuest suite -9.817e-02 4.114e-02 -2.386 0.017022  
## property\_typeGuesthouse -8.463e-02 2.051e-02 -4.126 3.70e-05  
## property\_typeHostel -5.367e-01 5.611e-02 -9.564 < 2e-16  
## property\_typeHouse -2.098e-02 4.781e-03 -4.387 1.15e-05  
## property\_typeHut -3.464e-01 1.543e-01 -2.244 0.024824  
## property\_typeIn-law -2.284e-01 5.167e-02 -4.421 9.86e-06  
## property\_typeIsland 9.393e-01 4.086e-01 2.299 0.021500  
## property\_typeLoft 1.851e-01 1.307e-02 14.160 < 2e-16  
## property\_typeOther 1.908e-02 2.011e-02 0.949 0.342646  
## property\_typeServiced apartment 2.039e-01 1.054e-01 1.934 0.053151  
## property\_typeTent -4.742e-01 1.093e-01 -4.339 1.43e-05  
## property\_typeTimeshare 5.490e-01 7.113e-02 7.718 1.20e-14  
## property\_typeTipi 3.746e-01 2.358e-01 1.589 0.112132  
## property\_typeTownhouse -5.279e-04 1.174e-02 -0.045 0.964124  
## property\_typeTrain 3.720e-01 2.886e-01 1.289 0.197476  
## property\_typeTreehouse 1.141e-01 1.667e-01 0.685 0.493435  
## property\_typeVacation home 3.610e-01 1.667e-01 2.166 0.030302  
## property\_typeVilla 2.073e-01 3.670e-02 5.648 1.63e-08  
## property\_typeYurt -1.324e-01 1.667e-01 -0.794 0.427149  
## room\_typePrivate room -5.991e-01 4.257e-03 -140.734 < 2e-16  
## room\_typeShared room -1.051e+00 1.194e-02 -88.004 < 2e-16  
## accommodates 7.782e-02 1.651e-03 47.145 < 2e-16  
## bathrooms 1.392e-01 3.898e-03 35.708 < 2e-16  
## bed\_typeCouch 7.825e-02 3.949e-02 1.982 0.047506  
## bed\_typeFuton -1.816e-02 2.773e-02 -0.655 0.512506  
## bed\_typePull-out Sofa 5.653e-02 2.879e-02 1.963 0.049624  
## bed\_typeReal Bed 6.205e-02 2.236e-02 2.776 0.005512  
## cancellation\_policymoderate 1.351e-02 4.836e-03 2.794 0.005214  
## cancellation\_policystrict 4.248e-02 4.468e-03 9.506 < 2e-16  
## cancellation\_policysuper\_strict\_30 3.712e-01 4.655e-02 7.975 1.55e-15  
## cancellation\_policysuper\_strict\_60 7.016e-01 1.294e-01 5.420 5.98e-08  
## cityChicago -1.761e+01 3.277e-01 -53.743 < 2e-16  
## cityDC -5.778e+00 1.503e-01 -38.457 < 2e-16  
## cityLA -4.831e+01 9.658e-01 -50.016 < 2e-16  
## cityNYC -2.770e+00 7.221e-02 -38.363 < 2e-16  
## citySF -5.277e+01 1.025e+00 -51.493 < 2e-16  
## instant\_bookableTRUE -4.492e-02 3.907e-03 -11.498 < 2e-16  
## latitude 1.597e-01 2.566e-02 6.225 4.85e-10  
## longitude -1.047e+00 1.971e-02 -53.125 < 2e-16  
## review\_scores\_rating 5.599e-03 2.209e-04 25.340 < 2e-16  
## bedrooms 1.438e-01 3.310e-03 43.458 < 2e-16  
## beds -3.900e-02 2.568e-03 -15.188 < 2e-16  
##   
## (Intercept) \*\*\*  
## property\_typeBed & Breakfast \*\*\*  
## property\_typeBoat \*\*\*  
## property\_typeBoutique hotel \*\*\*  
## property\_typeBungalow   
## property\_typeCabin \*   
## property\_typeCamper/RV \*\*\*  
## property\_typeCastle \*\*\*  
## property\_typeCave   
## property\_typeChalet   
## property\_typeCondominium \*\*\*  
## property\_typeDorm \*\*\*  
## property\_typeEarth House   
## property\_typeGuest suite \*   
## property\_typeGuesthouse \*\*\*  
## property\_typeHostel \*\*\*  
## property\_typeHouse \*\*\*  
## property\_typeHut \*   
## property\_typeIn-law \*\*\*  
## property\_typeIsland \*   
## property\_typeLoft \*\*\*  
## property\_typeOther   
## property\_typeServiced apartment .   
## property\_typeTent \*\*\*  
## property\_typeTimeshare \*\*\*  
## property\_typeTipi   
## property\_typeTownhouse   
## property\_typeTrain   
## property\_typeTreehouse   
## property\_typeVacation home \*   
## property\_typeVilla \*\*\*  
## property\_typeYurt   
## room\_typePrivate room \*\*\*  
## room\_typeShared room \*\*\*  
## accommodates \*\*\*  
## bathrooms \*\*\*  
## bed\_typeCouch \*   
## bed\_typeFuton   
## bed\_typePull-out Sofa \*   
## bed\_typeReal Bed \*\*   
## cancellation\_policymoderate \*\*   
## cancellation\_policystrict \*\*\*  
## cancellation\_policysuper\_strict\_30 \*\*\*  
## cancellation\_policysuper\_strict\_60 \*\*\*  
## cityChicago \*\*\*  
## cityDC \*\*\*  
## cityLA \*\*\*  
## cityNYC \*\*\*  
## citySF \*\*\*  
## instant\_bookableTRUE \*\*\*  
## latitude \*\*\*  
## longitude \*\*\*  
## review\_scores\_rating \*\*\*  
## bedrooms \*\*\*  
## beds \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4081 on 57073 degrees of freedom  
## (260 observations deleted due to missingness)  
## Multiple R-squared: 0.6278, Adjusted R-squared: 0.6275   
## F-statistic: 1783 on 54 and 57073 DF, p-value: < 2.2e-16

Namun demikian, akurasi model tetap tidak berubah, yakni berada pada 62.7%. Adapun residual dari model tidak terdistribusi normal.

qqnorm(linearModel\_airBNB\_lama\_fs[["residuals"]])  
qqline(linearModel\_airBNB\_lama\_fs[["residuals"]])

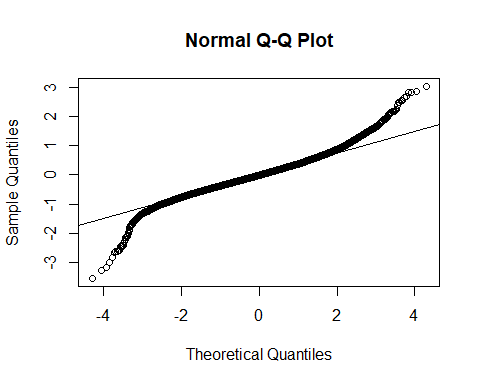
 Secara teori, apabila variabel independen bernilai 0, maka variabel dependen (log\_price) akan bernilai 0. Sehingga, pemodelan dapat dilakukan dengan mengabaikan *intercept*.

linearModel\_airBNB\_lama\_nointercept<-lm(log\_price ~ 0+property\_type+room\_type+accommodates+bathrooms+  
 bed\_type+cancellation\_policy+city+instant\_bookable+  
 latitude+longitude+review\_scores\_rating+bedrooms+beds, data=subset\_airBNB\_lama)  
summary(linearModel\_airBNB\_lama\_nointercept)

##   
## Call:  
## lm(formula = log\_price ~ 0 + property\_type + room\_type + accommodates +   
## bathrooms + bed\_type + cancellation\_policy + city + instant\_bookable +   
## latitude + longitude + review\_scores\_rating + bedrooms +   
## beds, data = subset\_airBNB\_lama)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5487 -0.2615 -0.0133 0.2441 3.0314   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## property\_typeApartment -7.724e+01 1.728e+00 -44.697 < 2e-16  
## property\_typeBed & Breakfast -7.709e+01 1.728e+00 -44.610 < 2e-16  
## property\_typeBoat -7.699e+01 1.728e+00 -44.553 < 2e-16  
## property\_typeBoutique hotel -7.702e+01 1.729e+00 -44.544 < 2e-16  
## property\_typeBungalow -7.727e+01 1.728e+00 -44.711 < 2e-16  
## property\_typeCabin -7.736e+01 1.730e+00 -44.721 < 2e-16  
## property\_typeCamper/RV -7.759e+01 1.731e+00 -44.831 < 2e-16  
## property\_typeCastle -7.685e+01 1.733e+00 -44.352 < 2e-16  
## property\_typeCave -7.696e+01 1.752e+00 -43.928 < 2e-16  
## property\_typeChalet -7.717e+01 1.736e+00 -44.452 < 2e-16  
## property\_typeCondominium -7.712e+01 1.728e+00 -44.634 < 2e-16  
## property\_typeDorm -7.767e+01 1.728e+00 -44.947 < 2e-16  
## property\_typeEarth House -7.718e+01 1.743e+00 -44.274 < 2e-16  
## property\_typeGuest suite -7.734e+01 1.728e+00 -44.755 < 2e-16  
## property\_typeGuesthouse -7.733e+01 1.729e+00 -44.730 < 2e-16  
## property\_typeHostel -7.778e+01 1.728e+00 -45.006 < 2e-16  
## property\_typeHouse -7.726e+01 1.728e+00 -44.720 < 2e-16  
## property\_typeHut -7.759e+01 1.735e+00 -44.727 < 2e-16  
## property\_typeIn-law -7.747e+01 1.729e+00 -44.816 < 2e-16  
## property\_typeIsland -7.630e+01 1.765e+00 -43.231 < 2e-16  
## property\_typeLoft -7.706e+01 1.728e+00 -44.601 < 2e-16  
## property\_typeOther -7.722e+01 1.728e+00 -44.684 < 2e-16  
## property\_typeServiced apartment -7.704e+01 1.731e+00 -44.505 < 2e-16  
## property\_typeTent -7.772e+01 1.733e+00 -44.843 < 2e-16  
## property\_typeTimeshare -7.669e+01 1.730e+00 -44.338 < 2e-16  
## property\_typeTipi -7.687e+01 1.747e+00 -43.995 < 2e-16  
## property\_typeTownhouse -7.724e+01 1.728e+00 -44.707 < 2e-16  
## property\_typeTrain -7.687e+01 1.756e+00 -43.785 < 2e-16  
## property\_typeTreehouse -7.713e+01 1.736e+00 -44.434 < 2e-16  
## property\_typeVacation home -7.688e+01 1.735e+00 -44.305 < 2e-16  
## property\_typeVilla -7.703e+01 1.726e+00 -44.634 < 2e-16  
## property\_typeYurt -7.737e+01 1.739e+00 -44.482 < 2e-16  
## room\_typePrivate room -5.991e-01 4.257e-03 -140.734 < 2e-16  
## room\_typeShared room -1.051e+00 1.194e-02 -88.004 < 2e-16  
## accommodates 7.782e-02 1.651e-03 47.145 < 2e-16  
## bathrooms 1.392e-01 3.898e-03 35.708 < 2e-16  
## bed\_typeCouch 7.825e-02 3.949e-02 1.982 0.04751  
## bed\_typeFuton -1.816e-02 2.773e-02 -0.655 0.51251  
## bed\_typePull-out Sofa 5.653e-02 2.879e-02 1.963 0.04962  
## bed\_typeReal Bed 6.205e-02 2.236e-02 2.776 0.00551  
## cancellation\_policymoderate 1.351e-02 4.836e-03 2.794 0.00521  
## cancellation\_policystrict 4.248e-02 4.468e-03 9.506 < 2e-16  
## cancellation\_policysuper\_strict\_30 3.712e-01 4.655e-02 7.975 1.55e-15  
## cancellation\_policysuper\_strict\_60 7.016e-01 1.294e-01 5.420 5.98e-08  
## cityChicago -1.761e+01 3.277e-01 -53.743 < 2e-16  
## cityDC -5.778e+00 1.503e-01 -38.457 < 2e-16  
## cityLA -4.831e+01 9.658e-01 -50.016 < 2e-16  
## cityNYC -2.770e+00 7.221e-02 -38.363 < 2e-16  
## citySF -5.277e+01 1.025e+00 -51.493 < 2e-16  
## instant\_bookableTRUE -4.492e-02 3.907e-03 -11.498 < 2e-16  
## latitude 1.597e-01 2.566e-02 6.225 4.85e-10  
## longitude -1.047e+00 1.971e-02 -53.125 < 2e-16  
## review\_scores\_rating 5.599e-03 2.209e-04 25.340 < 2e-16  
## bedrooms 1.438e-01 3.310e-03 43.458 < 2e-16  
## beds -3.900e-02 2.568e-03 -15.188 < 2e-16  
##   
## property\_typeApartment \*\*\*  
## property\_typeBed & Breakfast \*\*\*  
## property\_typeBoat \*\*\*  
## property\_typeBoutique hotel \*\*\*  
## property\_typeBungalow \*\*\*  
## property\_typeCabin \*\*\*  
## property\_typeCamper/RV \*\*\*  
## property\_typeCastle \*\*\*  
## property\_typeCave \*\*\*  
## property\_typeChalet \*\*\*  
## property\_typeCondominium \*\*\*  
## property\_typeDorm \*\*\*  
## property\_typeEarth House \*\*\*  
## property\_typeGuest suite \*\*\*  
## property\_typeGuesthouse \*\*\*  
## property\_typeHostel \*\*\*  
## property\_typeHouse \*\*\*  
## property\_typeHut \*\*\*  
## property\_typeIn-law \*\*\*  
## property\_typeIsland \*\*\*  
## property\_typeLoft \*\*\*  
## property\_typeOther \*\*\*  
## property\_typeServiced apartment \*\*\*  
## property\_typeTent \*\*\*  
## property\_typeTimeshare \*\*\*  
## property\_typeTipi \*\*\*  
## property\_typeTownhouse \*\*\*  
## property\_typeTrain \*\*\*  
## property\_typeTreehouse \*\*\*  
## property\_typeVacation home \*\*\*  
## property\_typeVilla \*\*\*  
## property\_typeYurt \*\*\*  
## room\_typePrivate room \*\*\*  
## room\_typeShared room \*\*\*  
## accommodates \*\*\*  
## bathrooms \*\*\*  
## bed\_typeCouch \*   
## bed\_typeFuton   
## bed\_typePull-out Sofa \*   
## bed\_typeReal Bed \*\*   
## cancellation\_policymoderate \*\*   
## cancellation\_policystrict \*\*\*  
## cancellation\_policysuper\_strict\_30 \*\*\*  
## cancellation\_policysuper\_strict\_60 \*\*\*  
## cityChicago \*\*\*  
## cityDC \*\*\*  
## cityLA \*\*\*  
## cityNYC \*\*\*  
## citySF \*\*\*  
## instant\_bookableTRUE \*\*\*  
## latitude \*\*\*  
## longitude \*\*\*  
## review\_scores\_rating \*\*\*  
## bedrooms \*\*\*  
## beds \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4081 on 57073 degrees of freedom  
## (260 observations deleted due to missingness)  
## Multiple R-squared: 0.9928, Adjusted R-squared: 0.9928   
## F-statistic: 1.425e+05 on 55 and 57073 DF, p-value: < 2.2e-16

Hasilnya, terjadi peningkatan akurasi dari 62.7% menjadi 99.2%. Namun demikian, residual tetap tidak terdistribusi linear

qqnorm(linearModel\_airBNB\_lama\_nointercept[["residuals"]])  
qqline(linearModel\_airBNB\_lama\_nointercept[["residuals"]])



## Cluster 2 : Properti Baru

Pada tahap ini dilakukan pemodelan terhadap properti baru, yakni properti yang belum memiliki rating. Langkah awal adalah dengan membuat *subset* atas data properti yang belum memiliki rating

subset\_airBNB\_baru<-subset(airBNB,is.na(review\_scores\_rating),)

Selanjutnya dilakukan pemodelan dengan menggunakan metode regresi linear lm(), dengan mengabaikan *intercept*

linearModel\_airBNB\_baru<-lm(log\_price ~ 0+property\_type+room\_type+accommodates+bathrooms+  
 bed\_type+cancellation\_policy+cleaning\_fee+city+instant\_bookable+  
 latitude+longitude+bedrooms+beds, data=subset\_airBNB\_baru)  
summary(linearModel\_airBNB\_baru)

##   
## Call:  
## lm(formula = log\_price ~ 0 + property\_type + room\_type + accommodates +   
## bathrooms + bed\_type + cancellation\_policy + cleaning\_fee +   
## city + instant\_bookable + latitude + longitude + bedrooms +   
## beds, data = subset\_airBNB\_baru)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7495 -0.3809 -0.0483 0.3153 3.7819   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## property\_typeApartment -67.009075 4.105261 -16.323 < 2e-16  
## property\_typeBed & Breakfast -66.709590 4.104116 -16.254 < 2e-16  
## property\_typeBoat -66.618446 4.103777 -16.233 < 2e-16  
## property\_typeBoutique hotel -66.132188 4.108240 -16.097 < 2e-16  
## property\_typeBungalow -66.972227 4.106066 -16.311 < 2e-16  
## property\_typeCabin -66.921083 4.117325 -16.254 < 2e-16  
## property\_typeCamper/RV -66.951557 4.109373 -16.292 < 2e-16  
## property\_typeCasa particular -66.764163 4.151915 -16.080 < 2e-16  
## property\_typeChalet -67.050057 4.150821 -16.153 < 2e-16  
## property\_typeCondominium -66.807812 4.104776 -16.276 < 2e-16  
## property\_typeDorm -67.594231 4.105383 -16.465 < 2e-16  
## property\_typeEarth House -66.194194 4.163248 -15.900 < 2e-16  
## property\_typeGuest suite -67.015325 4.104324 -16.328 < 2e-16  
## property\_typeGuesthouse -67.080089 4.107116 -16.333 < 2e-16  
## property\_typeHostel -67.592273 4.106715 -16.459 < 2e-16  
## property\_typeHouse -66.962014 4.103850 -16.317 < 2e-16  
## property\_typeHut -68.881188 4.154368 -16.580 < 2e-16  
## property\_typeIn-law -67.630873 4.110601 -16.453 < 2e-16  
## property\_typeLighthouse -66.738536 4.133467 -16.146 < 2e-16  
## property\_typeLoft -66.810557 4.104157 -16.279 < 2e-16  
## property\_typeOther -66.688140 4.105570 -16.243 < 2e-16  
## property\_typeServiced apartment -66.445132 4.111394 -16.161 < 2e-16  
## property\_typeTent -67.040563 4.117250 -16.283 < 2e-16  
## property\_typeTimeshare -66.533735 4.107033 -16.200 < 2e-16  
## property\_typeTownhouse -67.002711 4.104248 -16.325 < 2e-16  
## property\_typeTreehouse -68.393739 4.146087 -16.496 < 2e-16  
## property\_typeVacation home -66.827864 4.105510 -16.278 < 2e-16  
## property\_typeVilla -66.660837 4.099826 -16.259 < 2e-16  
## property\_typeYurt -66.797503 4.123374 -16.200 < 2e-16  
## room\_typePrivate room -0.621579 0.011314 -54.941 < 2e-16  
## room\_typeShared room -0.972031 0.025178 -38.606 < 2e-16  
## accommodates 0.102521 0.003810 26.910 < 2e-16  
## bathrooms 0.138992 0.009833 14.136 < 2e-16  
## bed\_typeCouch 0.139193 0.077430 1.798 0.07225  
## bed\_typeFuton 0.042929 0.070957 0.605 0.54519  
## bed\_typePull-out Sofa 0.085017 0.080935 1.050 0.29354  
## bed\_typeReal Bed 0.021310 0.051882 0.411 0.68127  
## cancellation\_policymoderate -0.065886 0.013628 -4.835 1.35e-06  
## cancellation\_policystrict -0.030926 0.011564 -2.674 0.00749  
## cancellation\_policysuper\_strict\_30 0.135686 0.110542 1.227 0.21967  
## cancellation\_policysuper\_strict\_60 0.536809 0.226969 2.365 0.01804  
## cleaning\_feeTRUE -0.086947 0.010109 -8.601 < 2e-16  
## cityChicago -15.735659 0.807374 -19.490 < 2e-16  
## cityDC -4.783601 0.380678 -12.566 < 2e-16  
## cityLA -43.557213 2.406597 -18.099 < 2e-16  
## cityNYC -2.494495 0.183284 -13.610 < 2e-16  
## citySF -47.318226 2.537223 -18.650 < 2e-16  
## instant\_bookableTRUE -0.049741 0.010818 -4.598 4.30e-06  
## latitude 0.113642 0.063075 1.802 0.07161  
## longitude -0.940232 0.048442 -19.409 < 2e-16  
## bedrooms 0.117725 0.008727 13.490 < 2e-16  
## beds -0.045007 0.006726 -6.691 2.28e-11  
##   
## property\_typeApartment \*\*\*  
## property\_typeBed & Breakfast \*\*\*  
## property\_typeBoat \*\*\*  
## property\_typeBoutique hotel \*\*\*  
## property\_typeBungalow \*\*\*  
## property\_typeCabin \*\*\*  
## property\_typeCamper/RV \*\*\*  
## property\_typeCasa particular \*\*\*  
## property\_typeChalet \*\*\*  
## property\_typeCondominium \*\*\*  
## property\_typeDorm \*\*\*  
## property\_typeEarth House \*\*\*  
## property\_typeGuest suite \*\*\*  
## property\_typeGuesthouse \*\*\*  
## property\_typeHostel \*\*\*  
## property\_typeHouse \*\*\*  
## property\_typeHut \*\*\*  
## property\_typeIn-law \*\*\*  
## property\_typeLighthouse \*\*\*  
## property\_typeLoft \*\*\*  
## property\_typeOther \*\*\*  
## property\_typeServiced apartment \*\*\*  
## property\_typeTent \*\*\*  
## property\_typeTimeshare \*\*\*  
## property\_typeTownhouse \*\*\*  
## property\_typeTreehouse \*\*\*  
## property\_typeVacation home \*\*\*  
## property\_typeVilla \*\*\*  
## property\_typeYurt \*\*\*  
## room\_typePrivate room \*\*\*  
## room\_typeShared room \*\*\*  
## accommodates \*\*\*  
## bathrooms \*\*\*  
## bed\_typeCouch .   
## bed\_typeFuton   
## bed\_typePull-out Sofa   
## bed\_typeReal Bed   
## cancellation\_policymoderate \*\*\*  
## cancellation\_policystrict \*\*   
## cancellation\_policysuper\_strict\_30   
## cancellation\_policysuper\_strict\_60 \*   
## cleaning\_feeTRUE \*\*\*  
## cityChicago \*\*\*  
## cityDC \*\*\*  
## cityLA \*\*\*  
## cityNYC \*\*\*  
## citySF \*\*\*  
## instant\_bookableTRUE \*\*\*  
## latitude .   
## longitude \*\*\*  
## bedrooms \*\*\*  
## beds \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5908 on 16585 degrees of freedom  
## (85 observations deleted due to missingness)  
## Multiple R-squared: 0.9859, Adjusted R-squared: 0.9859   
## F-statistic: 2.232e+04 on 52 and 16585 DF, p-value: < 2.2e-16

Dari proses pemodelan diatas, didapatkan akurasi model sebesar 0.985 (98.5%). Bila dilakukan proses *feature selection* didapatkan hasil sebagai berikut :

fs\_both\_baru<-stepAIC(linearModel\_airBNB\_baru,direction = "both")

## Start: AIC=-17460.98  
## log\_price ~ 0 + property\_type + room\_type + accommodates + bathrooms +   
## bed\_type + cancellation\_policy + cleaning\_fee + city + instant\_bookable +   
## latitude + longitude + bedrooms + beds  
##   
## Df Sum of Sq RSS AIC  
## - bed\_type 4 1.86 5790.2 -17464  
## <none> 5788.4 -17461  
## - latitude 1 1.13 5789.5 -17460  
## - instant\_bookable 1 7.38 5795.8 -17442  
## - cancellation\_policy 4 11.50 5799.9 -17436  
## - beds 1 15.63 5804.0 -17418  
## - cleaning\_fee 1 25.82 5814.2 -17389  
## - bedrooms 1 63.52 5851.9 -17281  
## - bathrooms 1 69.74 5858.1 -17264  
## - longitude 1 131.48 5919.9 -17089  
## - property\_type 29 203.46 5991.8 -16944  
## - accommodates 1 252.74 6041.1 -16752  
## - city 5 593.57 6381.9 -15847  
## - room\_type 2 1248.09 7036.5 -14216  
##   
## Step: AIC=-17463.63  
## log\_price ~ property\_type + room\_type + accommodates + bathrooms +   
## cancellation\_policy + cleaning\_fee + city + instant\_bookable +   
## latitude + longitude + bedrooms + beds - 1  
##   
## Df Sum of Sq RSS AIC  
## <none> 5790.2 -17464  
## - latitude 1 1.18 5791.4 -17462  
## + bed\_type 4 1.86 5788.4 -17461  
## - instant\_bookable 1 7.58 5797.8 -17444  
## - cancellation\_policy 4 11.52 5801.8 -17439  
## - beds 1 15.81 5806.0 -17420  
## - cleaning\_fee 1 25.69 5815.9 -17392  
## - bedrooms 1 64.04 5854.3 -17283  
## - bathrooms 1 69.34 5859.6 -17268  
## - longitude 1 131.97 5922.2 -17091  
## - property\_type 29 206.12 5996.3 -16940  
## - accommodates 1 252.20 6042.4 -16756  
## - city 5 595.44 6385.7 -15845  
## - room\_type 2 1255.35 7045.6 -14203

Dari hasil *feature selection*, didapatkan hasil bahwa model akan mengabaikan fitur bed\_type. Adapun hasil regresi linear yang didapatkan sebagai berikut :

summary(fs\_both\_baru)

##   
## Call:  
## lm(formula = log\_price ~ property\_type + room\_type + accommodates +   
## bathrooms + cancellation\_policy + cleaning\_fee + city + instant\_bookable +   
## latitude + longitude + bedrooms + beds - 1, data = subset\_airBNB\_baru)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7498 -0.3812 -0.0482 0.3153 3.7707   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## property\_typeApartment -67.192068 4.102791 -16.377 < 2e-16  
## property\_typeBed & Breakfast -66.893909 4.101621 -16.309 < 2e-16  
## property\_typeBoat -66.801233 4.101305 -16.288 < 2e-16  
## property\_typeBoutique hotel -66.314729 4.105762 -16.152 < 2e-16  
## property\_typeBungalow -67.154280 4.103643 -16.365 < 2e-16  
## property\_typeCabin -67.106583 4.114821 -16.309 < 2e-16  
## property\_typeCamper/RV -67.130433 4.106980 -16.345 < 2e-16  
## property\_typeCasa particular -66.947795 4.149463 -16.134 < 2e-16  
## property\_typeChalet -67.234267 4.148356 -16.207 < 2e-16  
## property\_typeCondominium -66.989989 4.102315 -16.330 < 2e-16  
## property\_typeDorm -67.782121 4.102855 -16.521 < 2e-16  
## property\_typeEarth House -66.259979 4.161505 -15.922 < 2e-16  
## property\_typeGuest suite -67.198485 4.101838 -16.383 < 2e-16  
## property\_typeGuesthouse -67.263723 4.104638 -16.387 < 2e-16  
## property\_typeHostel -67.781707 4.104151 -16.515 < 2e-16  
## property\_typeHouse -67.145275 4.101370 -16.371 < 2e-16  
## property\_typeHut -69.074599 4.151790 -16.637 < 2e-16  
## property\_typeIn-law -67.815622 4.108103 -16.508 < 2e-16  
## property\_typeLighthouse -66.921317 4.131028 -16.200 < 2e-16  
## property\_typeLoft -66.991124 4.101707 -16.332 < 2e-16  
## property\_typeOther -66.865555 4.103152 -16.296 < 2e-16  
## property\_typeServiced apartment -66.628597 4.108917 -16.216 < 2e-16  
## property\_typeTent -67.232179 4.115023 -16.338 < 2e-16  
## property\_typeTimeshare -66.716567 4.104557 -16.254 < 2e-16  
## property\_typeTownhouse -67.185603 4.101778 -16.380 < 2e-16  
## property\_typeTreehouse -68.577796 4.143623 -16.550 < 2e-16  
## property\_typeVacation home -67.010103 4.103052 -16.332 < 2e-16  
## property\_typeVilla -66.843723 4.097347 -16.314 < 2e-16  
## property\_typeYurt -66.981741 4.120885 -16.254 < 2e-16  
## room\_typePrivate room -0.621521 0.011307 -54.968 < 2e-16  
## room\_typeShared room -0.959926 0.024495 -39.188 < 2e-16  
## accommodates 0.102386 0.003809 26.880 < 2e-16  
## bathrooms 0.138566 0.009831 14.094 < 2e-16  
## cancellation\_policymoderate -0.065816 0.013628 -4.829 1.38e-06  
## cancellation\_policystrict -0.031352 0.011561 -2.712 0.0067  
## cancellation\_policysuper\_strict\_30 0.134799 0.110545 1.219 0.2227  
## cancellation\_policysuper\_strict\_60 0.537642 0.226978 2.369 0.0179  
## cleaning\_feeTRUE -0.086723 0.010108 -8.580 < 2e-16  
## cityChicago -15.762004 0.807294 -19.525 < 2e-16  
## cityDC -4.785127 0.380640 -12.571 < 2e-16  
## cityLA -43.614214 2.406441 -18.124 < 2e-16  
## cityNYC -2.495348 0.183270 -13.616 < 2e-16  
## citySF -47.389762 2.537026 -18.679 < 2e-16  
## instant\_bookableTRUE -0.050359 0.010806 -4.660 3.18e-06  
## latitude 0.115812 0.063048 1.837 0.0662  
## longitude -0.941829 0.048437 -19.444 < 2e-16  
## bedrooms 0.118175 0.008724 13.545 < 2e-16  
## beds -0.045262 0.006725 -6.731 1.74e-11  
##   
## property\_typeApartment \*\*\*  
## property\_typeBed & Breakfast \*\*\*  
## property\_typeBoat \*\*\*  
## property\_typeBoutique hotel \*\*\*  
## property\_typeBungalow \*\*\*  
## property\_typeCabin \*\*\*  
## property\_typeCamper/RV \*\*\*  
## property\_typeCasa particular \*\*\*  
## property\_typeChalet \*\*\*  
## property\_typeCondominium \*\*\*  
## property\_typeDorm \*\*\*  
## property\_typeEarth House \*\*\*  
## property\_typeGuest suite \*\*\*  
## property\_typeGuesthouse \*\*\*  
## property\_typeHostel \*\*\*  
## property\_typeHouse \*\*\*  
## property\_typeHut \*\*\*  
## property\_typeIn-law \*\*\*  
## property\_typeLighthouse \*\*\*  
## property\_typeLoft \*\*\*  
## property\_typeOther \*\*\*  
## property\_typeServiced apartment \*\*\*  
## property\_typeTent \*\*\*  
## property\_typeTimeshare \*\*\*  
## property\_typeTownhouse \*\*\*  
## property\_typeTreehouse \*\*\*  
## property\_typeVacation home \*\*\*  
## property\_typeVilla \*\*\*  
## property\_typeYurt \*\*\*  
## room\_typePrivate room \*\*\*  
## room\_typeShared room \*\*\*  
## accommodates \*\*\*  
## bathrooms \*\*\*  
## cancellation\_policymoderate \*\*\*  
## cancellation\_policystrict \*\*   
## cancellation\_policysuper\_strict\_30   
## cancellation\_policysuper\_strict\_60 \*   
## cleaning\_feeTRUE \*\*\*  
## cityChicago \*\*\*  
## cityDC \*\*\*  
## cityLA \*\*\*  
## cityNYC \*\*\*  
## citySF \*\*\*  
## instant\_bookableTRUE \*\*\*  
## latitude .   
## longitude \*\*\*  
## bedrooms \*\*\*  
## beds \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5908 on 16589 degrees of freedom  
## (85 observations deleted due to missingness)  
## Multiple R-squared: 0.9859, Adjusted R-squared: 0.9859   
## F-statistic: 2.418e+04 on 48 and 16589 DF, p-value: < 2.2e-16

Dari hasil regresi linear dengan fitur yang sudah dilakukan seleksi, akan menghasilkan nilai akurasi yang tidak berubah, yakni sebesar 98.5%

# Discussion

Dari hasil pemodelan kedua subset data, yakni data properti Airbnb baru dan *existing*, didapatkan hasil bahwa fitur berikut mempengaruhi harga sewa properti *existing* :

* Tipe properti (contoh : Villa, Rumah, Kondominium, dsb)
* Tipe kamar (contoh : kamar pribadi, shared room, dsb)
* Jumlah tamu yang dapat ditampung dalam properti. Semakin banyak jumlah tamu yang dapat di tampung, semakin mahal harga sewanya
* Jumlah kamar mandi yang tersedia
* Jenis *bed* yang disediakan. Properti yang memiliki bed berjenis *futon* akan semakin murah harganya
* *Cancellation Policy*
* Kota letak properti. Kota Boston memiliki rata-rata harga properti tertinggi daripada keseluruhan data
* Ketersediaan fitur *instant\_bookable*. Properti yang tidak memiliki fitur tersebut akan semakin murah
* *Latitude* & *longitude* properti. Semakin properti terletak di daerah timur & utara AS, semakin mahal harga sewanya.
* Semakin tinggi nilai review, semakin mahal harga propertinya.
* Jumlah kamar tidur melambangkan semakin banyaknya tamu yang dapat ditampung dalam properti. Sehingga fitur bedrooms mempengaruhi harga properti
* Jumlah tempat tidur yang disediakan akan mempengaruhi harga properti. Semakin banyak jumlah tempat tidur, semakin rendah harganya

Untuk properti baru, fitur yang mempengaruhi, antara lain :

* Tipe properti
* Jenis kamar (private/share)
* Jumlah tamu yang dapat ditampung (semakin banyak semakin tinggi harga sewa properti)
* Jumlah kamar mandi tersedia
* *Cancellation Policy*
* *Cleaning Fee*
* Kota letak properti
* Fitur *instant bookable* (apabila ada, semakin mengurangi harga sewa)
* *Latitude* & *longitude* properti. Semakin properti terletak di daerah timur & utara AS, semakin mahal harga sewanya.
* Jumlah ruang tidur dan tempat tidur mempengaruhi harga properti