

LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE
ANALISIS PERSEBARAN AKTIVITAS PERSEWAAN PROPERTI DI
NEW YORK BERDASARKAN AIRBNB



HAMMAM FARIZ RAHARJO 123200062
FADHIL WICAKSONO NUR RAHMAN 123200160

PROGRAM STUDI INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
YOGYAKARTA
2022

HALAMAN PENGESAHAN

LAPORAN PROYEK AKHIR

Disusun oleh :

Hammam Fariz Raharjo 123200062

Fadhil Wicaksono Nur Rahman 123200160

Telah Diperiksa dan Disetujui oleh Asisten Praktikum Data Science

Pada Tanggal :

Asisten Praktikum

Asisten Praktikum

Rivano Ardiyan Taufiq Kurniawan

NIM.123190084

Ega Erinovian Megananda Putra

NIM.123190113

DAFTAR ISI

HALAMAN PENGESAHAN LAPORAN PROYEK AKHIR	3
DAFTAR ISI	4
PENDAHULUAN	5
METODE	6
Machine Learning Dengan Linear Regression	6
Data Analysis Menggunakan GGLOT2	6
Data Analysis Menggunakan GGLOT2 dan Shiny	7
HASIL DAN PEMBAHASAN	7
Analisis hosting area per kawasan	8
Machine Learning Dengan Linear Regression	12
KESIMPULAN	14

1. PENDAHULUAN

Dalam membuka sebuah usaha tentu diperlukan sebuah observasi untuk menentukan langkah dalam melakukan kegiatan bisnis. Dengan melakukan langkah observasi tentu akan membantu dalam penentuan keputusan. Penentuan keputusan yang tepat adalah sebuah jembatan menuju suksesnya sebuah bisnis yang akan dikembangkan. Dalam melakukan analisis maupun observasi tentu berdasarkan sebuah parameter-parameter yang menunjang bisnis. Dalam kasus ini, kami membahas tentang persebaran sebuah persewaan hotel maupun hostel di daerah New York City. New York adalah salah satu kota yang memiliki tingkat kunjungan turis yang tinggi dari tahun ke tahun. Dengan begitu, kesempatan untuk membuka usaha penginapan akan sangat menggiurkan mengingat potensi kota tersebut yang sangat besar.

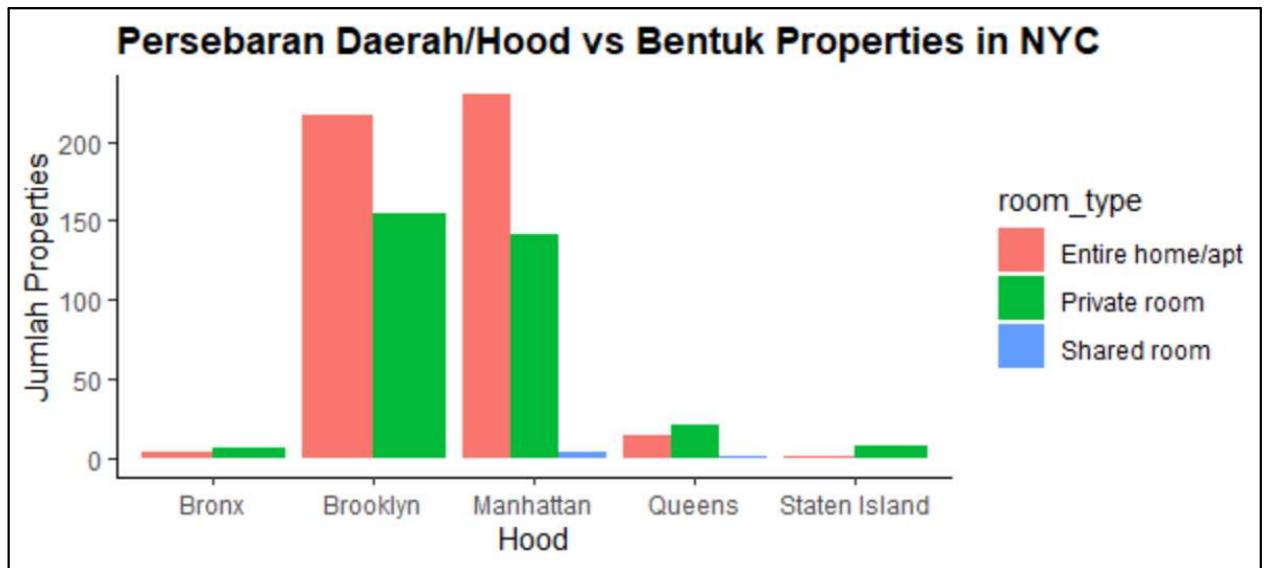
Di era saat ini, orang-orang melakukan reservasi tempat penginapan dengan begitu mudah, salah satunya menggunakan aplikasi Airbnb. Dengan aplikasi tersebut, para pengunjung akan mudah dalam melakukan pemesanan hotel atau tempat penginapan sedari jauh-jauh hari sebelum tiba di lokasi. Airbnb merupakan salah satu aplikasi pemesanan hotel dengan pengguna tertinggi di Amerika Serikat. Oleh karena itu, kami akan menggunakan dataset dari Airbnb untuk melakukan analisis persebaran hotel maupun hostel di Amerika Serikat terutama New York City. Dengan harapan dari sudut pandang bisnis dapat memperoleh pandangan dan potensi bisnis hotel di daerah tersebut. Tidak hanya itu tetapi juga dari sudut pandang pengunjung dapat melihat ringkasan dari persebaran tersebut yang nantinya dapat digunakan sebagai pertimbangan dalam pemesanan hotel maupun hostel.

2. METODE

- *Machine Learning Dengan Linear Regression*
 1. Melakukan *data cleaning* dengan melakukan pembersihan duplikasi data dan data kosong.
 2. Melakukan *splitting* yakni data training dan data test
 3. Membuat model, linear regression

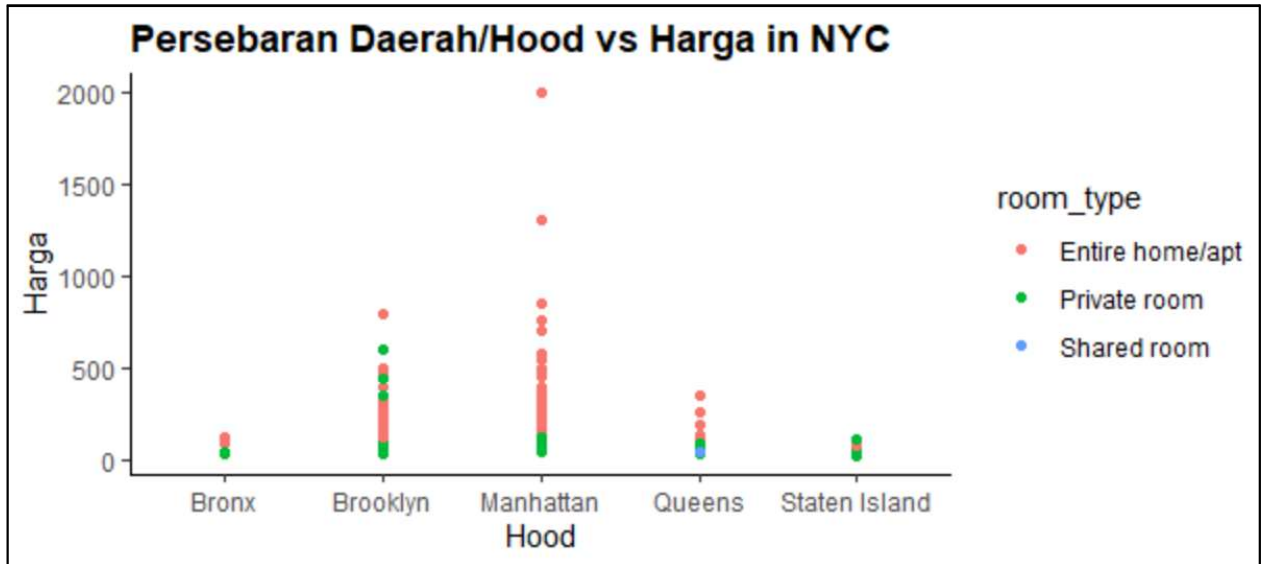
4. Melakukan training dengan model yang telah dibuat
 5. Melakukan fitting model ke data test
 6. Melakukan *prediction*
- *Data Analysis Menggunakan GGLOT2*
 1. Melakukan *subset* dari *dataset* yang digunakan agar data lebih mudah dibaca
 2. Membersihkan *subset* data dari data yang tidak masuk akal dengan parameter harga, yakni harga yang memiliki value 0
 3. Melakukan *grouping* dari *neighborhood* agar mudah dibaca
 4. Melakukan visualisasi data dengan ggplot2 untuk melihat perbandingan data properties dengan *hood*/daerah
 5. Melakukan visualisasi data dengan ggplot2 untuk melihat perbandingan daerah dengan harga
 6. Melakukan analisis mendalam dengan pemisahan data menjadi lebih detail
 7. Analisis daerah Brooklyn
 8. Analisis daerah Bronx
 9. Analisis daerah Manhattan
 10. Analisis daerah Queen
 11. Analisis daerah Staten Island
 12. Analisis trend harga sewa
 13. Analisis dan mengidentifikasi berapa lama pengunjung menyewa properti
 - *Data Analysis Menggunakan GGLOT2 dan Shiny*
 1. Melakukan pembuatan UI dengan fluidpage
 2. Melakukan penambahan sidebar panel dengan berbagai panel yang diperlukan
 3. Membuat variabel Server dengan variabel yang diperlukan
 4. Melakukan *view* terhadap hasil pengolahan data

3. HASIL DAN PEMBAHASAN



Gambar 3.1 Persebaran *hood* dengan jenis properti

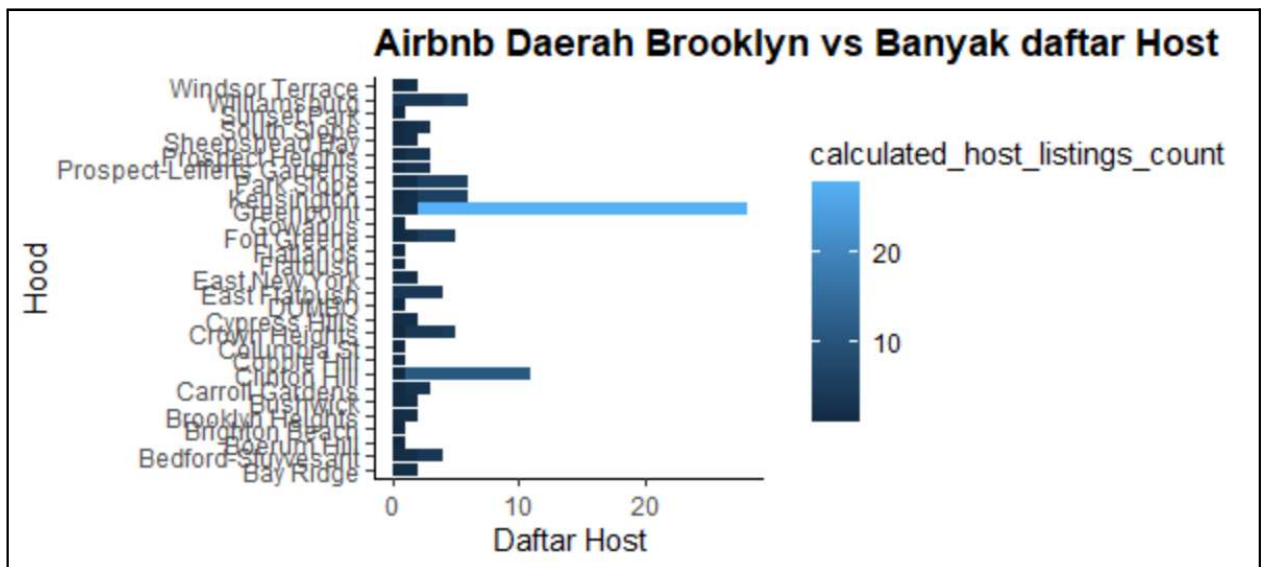
Dari grafik yang diperoleh, dapat dilihat bahwa Manhattan merupakan kawasan dengan paling banyak properti yang disewakan dan di Manhattan jenis properti yang paling banyak disewakan adalah apartemen.



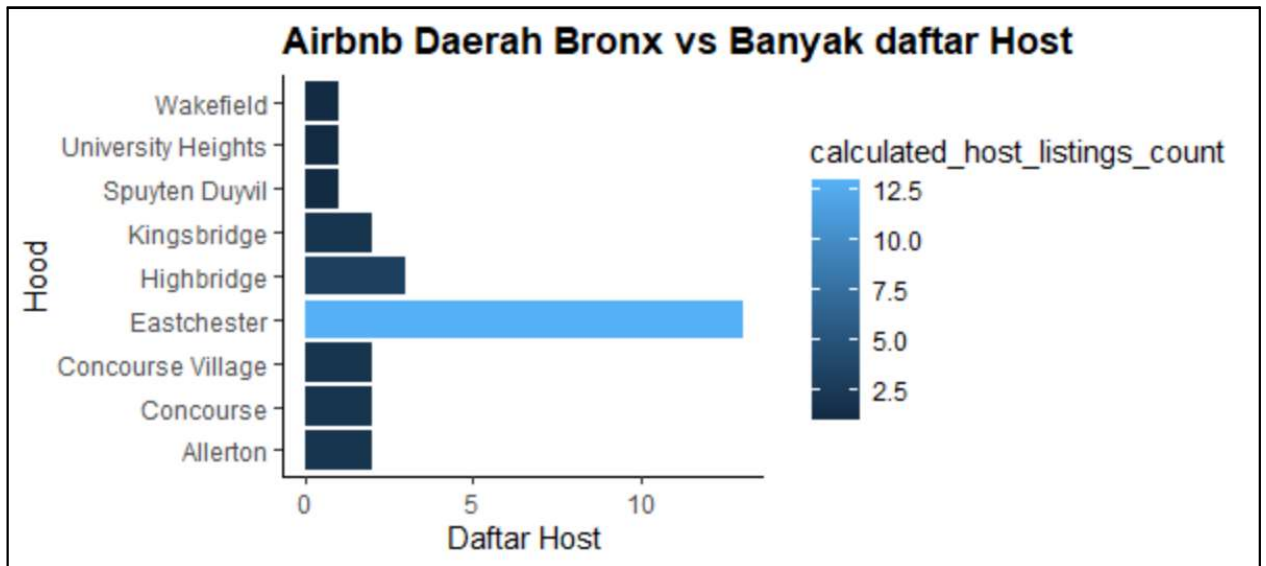
Gambar 3.2 Persebaran *hood* dengan harga properti

Dari grafik di atas dapat dilihat persebaran harga sewa properti di berbagai daerah di kawasan New York City. Untuk wilayah Brooklyn, Manhattan, Queen memiliki rata-rata harga yang lebih tinggi dibandingkan kawasan lain

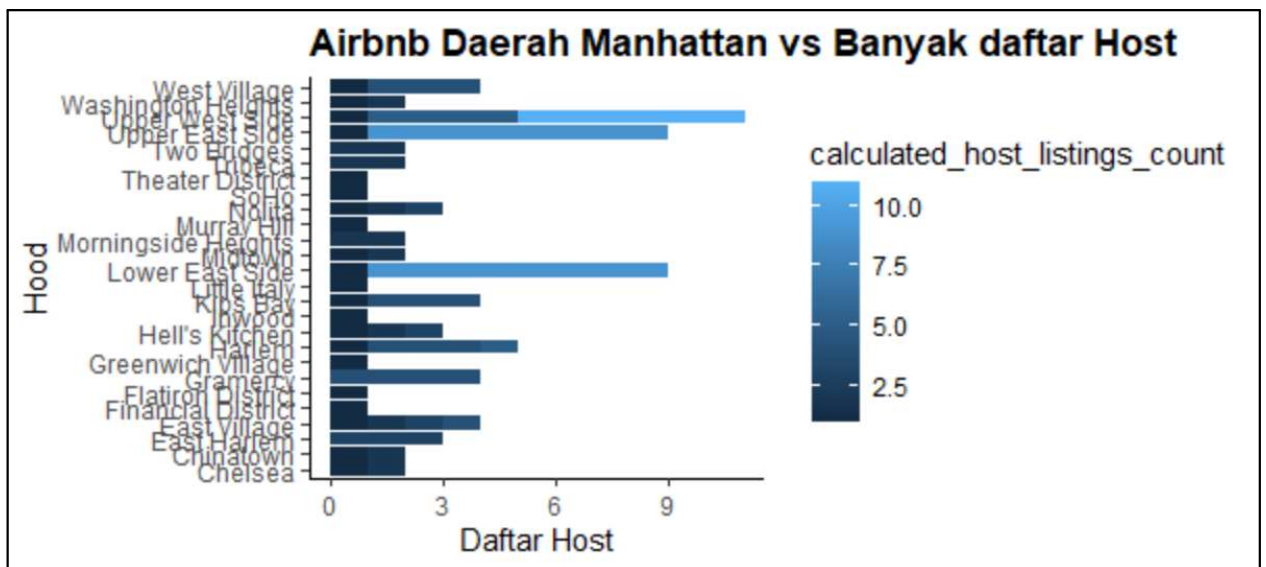
- *Analisis hosting area per kawasan*



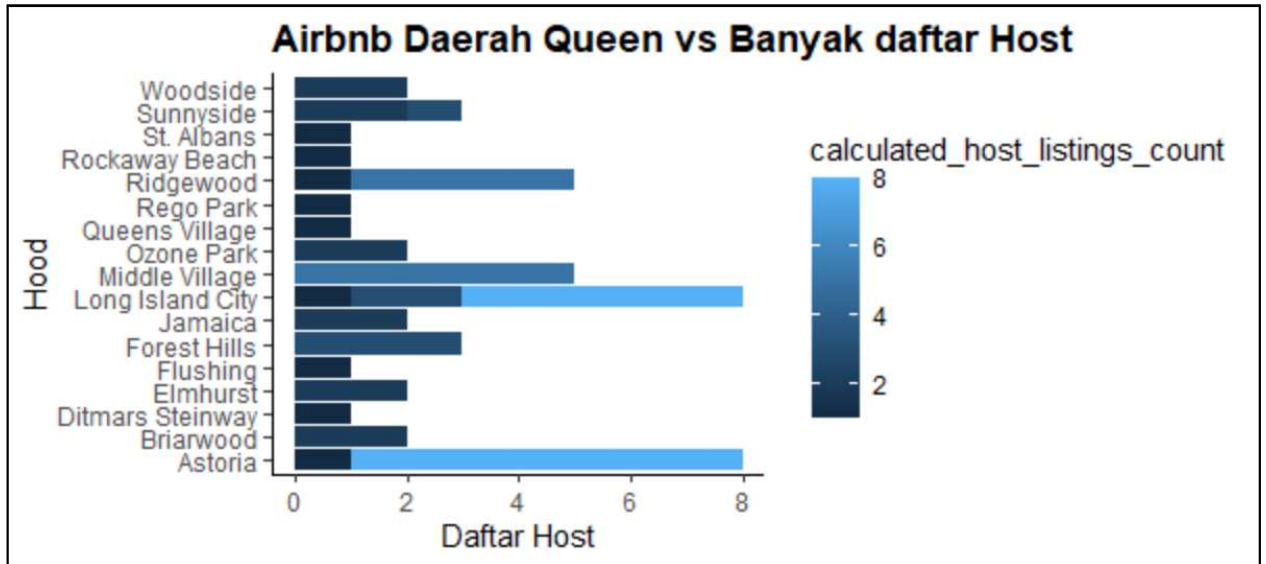
Gambar 3.3 Persebaran jumlah persewaan properti di daerah Brooklyn



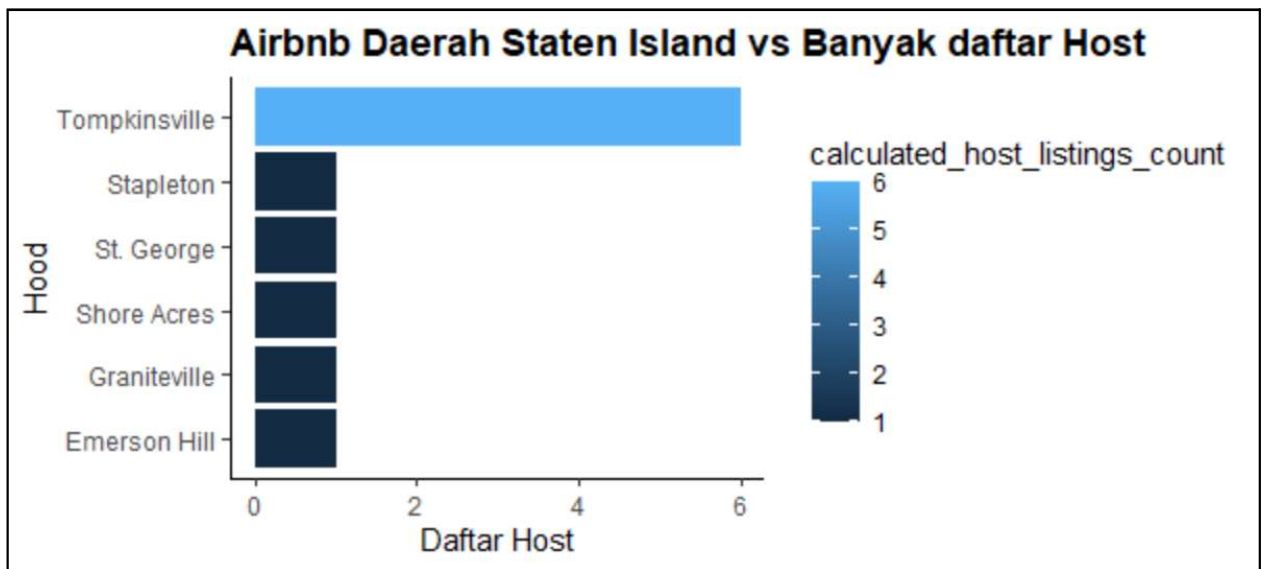
Gambar 3.4 Persebaran jumlah persewaan properti di daerah Bronx



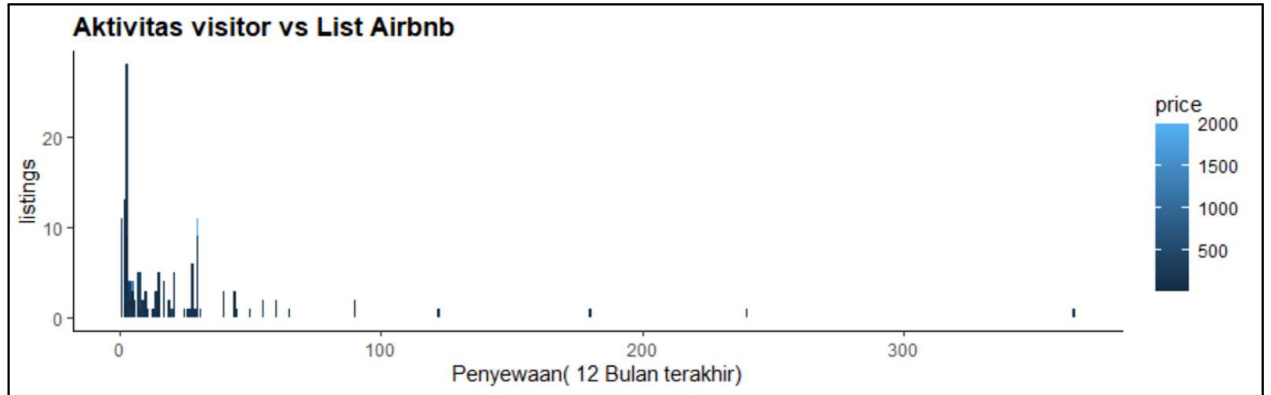
Gambar 3.5 Persebaran jumlah persewaan properti di daerah Manhattan



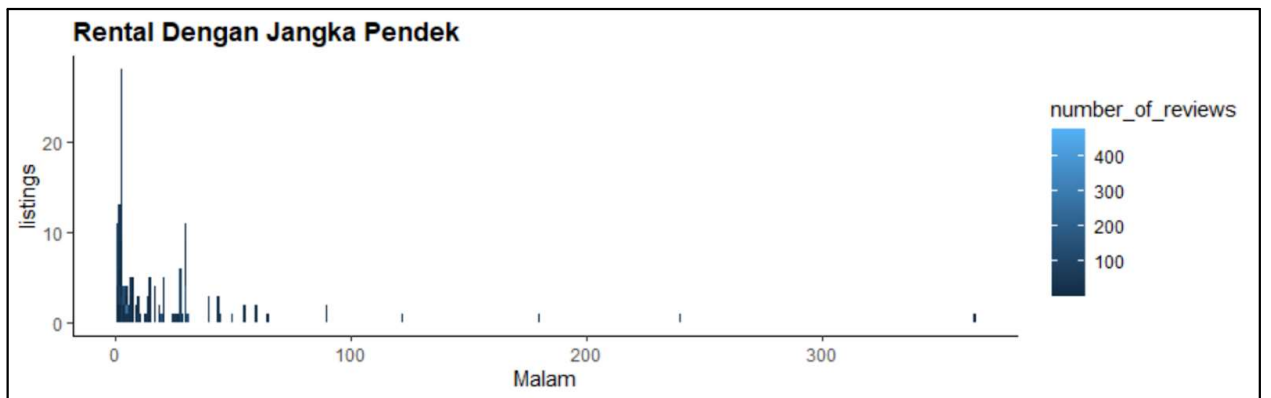
Gambar 3.6 Persebaran jumlah persewaan properti di daerah Queen



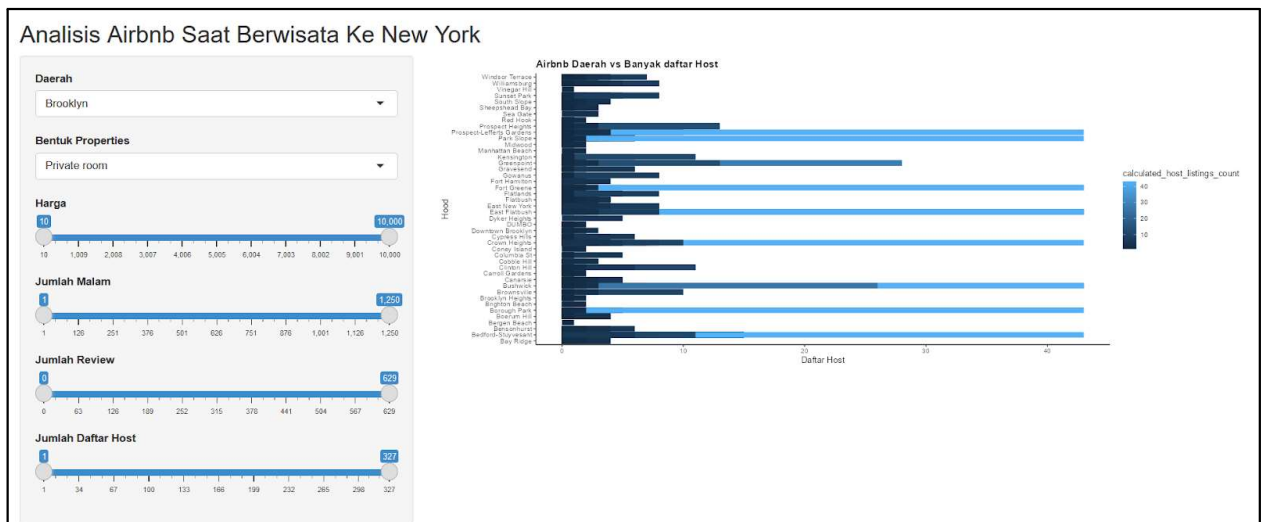
Gambar 3.7 Persebaran jumlah persewaan properti di daerah Staten Island



Gambar 3.8 Aktivitas *Visitor* dalam penyewaan per malam



Gambar 3.9 Banyak *visitor* dengan penyewaan dalam jangka pendek



Gambar 3.10 Implementasi Shiny dan ggplot2 dalam dataset

- *Machine Learning Dengan Linear Regression*

1. *Data Cleaning*

```
df_airbnb <- df_airbnb %>%  
  distinct() %>%  
  clean_names() %>%  
  drop_na()  
  
airbnb_hood <- subset(airbnb_hood, price > 0)  
airbnb_property <- subset(airbnb_property, price > 0)  
df_airbnb_clean <- subset(df_airbnb, price > 0)
```

Listing Code 3.1

→ Digunakan untuk membersihkan data kosong yang dapat mempengaruhi performa model algoritma *machine learning*, seperti data kosong dan duplikasi data.

2. *Data Splitting*

```
set.seed(123)  
  
df_airbnb_split <- initial_split(df_airbnb_clean, prop =  
0.8, strata = price)  
  
df_airbnb_train <- training(df_airbnb_split)  
df_airbnb_test <- testing(df_airbnb_split)
```

Listing Code 3.2

→ *splitting data* antara *data training* dan *data test*

3. *Recipe*

```

airbnb_recipe <- df_airbnb_train %>%

  recipe() %>%

  update_role(price, new_role = "outcome") %>%

  update_role(number_of_reviews, minimum_nights, new_role =
"predictor") %>%

  update_role(host_id, new_role = "id") %>%

  step_corr(all_predictors(), threshold = 0.9)

```

Listing Code 3.3

4. Modeling

```

airbnb_model <- linear_reg() %>%

  set_engine("lm") %>%

  set_mode("regression")

```

Listing Code 3.4

→ Membuat model algoritma yakni Linear Regression yang nantinya akan diimplementasikan ke dalam *dataset* yang telah dibersihkan.

5. Training

```

airbnb_workflow <- workflow() %>%

  add_recipe(airbnb_recipe) %>%

  add_model(airbnb_model)

```

Listing Code 3.5

→ Melakukan *training model* dengan mengimplementasikan model algoritma ke dalam *data training*

6. Fitting

```

airbnb_fit <- airbnb_workflow %>%

```

```
fit(data = df_airbnb_train)

summary(airbnb_fit)
```

Listing Code 3.6

→ Mengimplementasi model algoritma ke dalam *data test* untuk mengetahui prediksi harga *real*

7. Prediction

```
airbnb_pred <- airbnb_fit %>%
  predict(df_airbnb_test)

bind_cols(df_airbnb_test, airbnb_pred)

airbnb_pred
```

Listing Code 3.7

→ Menampilkan data hasil prediksi

4. KESIMPULAN

Dari hasil analisis yang dilakukan diperoleh beberapa grafik yang menampilkan kondisi yang tercatat dalam file dataset. Dari kondisi tersebut kita dapat mengambil hasil analisis dan observasi. Dari grafik yang ada, secara garis besar dapat diambil kesimpulan bahwa Manhattan adalah kawasan di New York dengan aktivitas pengunjung paling padat dan juga dengan jumlah properti paling banyak diantara daerah lain di New York. Sedangkan daerah dengan jumlah properti yang sedikit adalah Staten Island dengan aktivitas pengunjung yang terendah dibanding dengan lainnya.

Di sisi lain, jenis properti yang paling banyak disewakan di daerah New York City adalah properti berjenis rumah atau apartemen dengan range harga sampai dengan 10000 dollars. Dari hasil analisis tersebut, banyak pengunjung lebih memilih untuk menyewa seluruh rumah atau apartemen untuk berkunjung di New York City.