

Kelompok FinPro-8 InsightSeekers

Risanto Darmawan

Asri Nur Azizah

Fransiska Angelina Widiyanti

Naufa Tasha Nabila



FINAL PROJECT STAGE 1

(Link jupyter notebook: https://github.com/rdarmawan48/FinPro8-InsightSeekers/blob/main/Homework_FinPro-8_InsightSeekers_Stage-1.ipynb)

Dataset

Dataset yang digunakan dalam analisis ini adalah kumpulan data Airbnb yang terdiri dari 20 file CSV. Dataset ini mencakup informasi mengenai penginapan Airbnb di akhir pekan (weekend) dan hari kerja (weekday) di 10 kota besar di Eropa. Dataset yang terdiri dari 20 file CSV kemudian dengan metode looping digabungkan menjadi 1 file dataset dengan penambahan fitur kota (nama kota), weekdays dan weekends. Kota-kota yang termasuk dalam analisis ini adalah Amsterdam, Athena, Barcelona, Berlin, Budapest, Lisbon, London, Paris, Roma, dan Wina. Setiap dataset berisi informasi penting seperti Unnamed:0, realSum, room_type, room_shared, room_private, person_capacity, host_is_superhost, multi, biz, cleanliness_rating, guest_satisfaction_overall, bedrooms, dist, metro_dist, attr_index, attr_index_norm, rest_index, rest_index_norm, lng, lat yang akan membantu dalam proses analisis yang bertujuan untuk optimalisasi harga sewa properti menggunakan model machine learning yang komprehensif untuk Airbnb yang dapat memaksimalkan pendapatan pemilik properti sekaligus meningkatkan kepuasan pelanggan melalui penentuan harga yang optimal dan relevan berdasarkan faktor-faktor seperti tipe kamar, lokasi, kebersihan, dan lainnya.

Data Quality

Data quality dalam data science mencakup berbagai aspek yang menentukan seberapa "baik" atau "layak" data tersebut untuk analisis berikut adalah beberapa hal yang relevan dalam konteks data quality:

- a. Accuracy : dataset Airbnb yang akan digunakan sudah mencerminkan informasi yang akurat mengenai penginapan di 10 kota besar di Eropa saat weekends dan weekdays.
- b. Completeness : Semua informasi yang diperlukan ada dan tercatat dalam dataset tersebut yang ditunjukkan dari nama-nama kolom yang ada. Tidak ada missing value (data yang hilang) atau atribut yang tidak lengkap.
- c. Consistency : Dataset telah diperiksa dan menunjukkan konsistensi dalam format dan nama kolom. Nama kolom yang terdapat pada dataset tersebut antara lain Unnamed:0, realSum, room_type, room_shared, room_private, person_capacity, host_is_superhost, multi, biz, cleanliness_rating, guest_satisfaction_overall, bedrooms, dist, metro_dist, attr_index, attr_index_norm, rest_index, rest_index_norm, lng, lat.
- d. Timeliness : dataset ini merupakan dataset histori yang diperoleh sebelum tahun 2024 sehingga dapat digunakan untuk membuat model machine learning yang bertujuan untuk optimalisasi harga sewa properti di waktu yang akan datang.
- e. Uniqueness : Dari dataset yang akan digunakan terdapat beberapa data nonunique terutama di kolom room_type, room_shared, room_private, person_capacity, host_is_superhost, multi, biz, cleanliness_rating, guest_satisfaction_overall, bedrooms. Hal ini dikarenakan pada kolom tersebut hanya terdapat beberapa pilihan, seperti TRUE FALSE, jenis-jenis ruang yang disediakan (private room, shared room, dan entire home/apt), angka 1 dan 0 di kolom multi dan biz, pemberian angka rating yang sama, dan jumlah tempat tidur. Sedangkan di kolom dist, metro_dist, attr_index merupakan data unique.
- f. Validity : Dari dataset yang akan digunakan terdapat beberapa kolom yang perlu dilakukan pengubahan tipe data seperti pada kolom room_type, room_shared, room_private, host_is_superhost, person_capacity, multi, biz, cleanliness_rating dan guest_satisfaction_overall serta drop kolom Unnamed:0.
- g. Reliability : dataset tersebut dapat diandalkan dan dapat digunakan dalam pengambilan keputusan. Keputusan yang diambil dari dataset tersebut oleh kelompok Data Science Finpro-8 InsightSeekers adalah optimalisasi harga sewa properti menggunakan model machine learning yang komprehensif untuk Airbnb. Tujuannya untuk memaksimalkan pendapatan pemilik properti sekaligus meningkatkan kepuasan pelanggan melalui penentuan harga yang optimal dan relevan berdasarkan faktor-faktor seperti tipe kamar, lokasi, kebersihan, dan lainnya
- h. Relevance : dataset yang diperoleh relevan dengan apa yang akan kami analisis, yaitu
 - Untuk menganalisis dampak karakteristik kamar terhadap harga.

- Untuk mengevaluasi peran lokasi (kota, dist, metro_dist) dalam menentukan harga Airbnb.
 - Untuk menilai bagaimana peringkat kebersihan dan kepuasan tamu memengaruhi harga properti.
 - Untuk membantu calon tamu menemukan akomodasi yang menawarkan nilai terbaik berdasarkan berbagai faktor.
 - Untuk memberikan wawasan yang akan membantu tuan rumah Airbnb menyesuaikan strategi harga untuk daya saing yang lebih baik.
 - Untuk mengidentifikasi faktor-faktor utama yang secara signifikan mempengaruhi harga properti Airbnb (realSum) di berbagai kota di Eropa.
- i. Accessibility : dataset tersebut dapat diakses dengan mudah diakses, terkait dengan format data yang mudah digunakan dan mendukung integrasi maupun analisis lebih lanjut.
 - j. Integrity : dataset tersebut tidak akan mengalami modifikasi atau kerusakan yang tidak diinginkan selama penyimpanan, pengolahan, atau transmisi sehingga analisis yang dilakukan terhadap dataset tersebut dapat akurat.

Data Cleaning

Data cleaning adalah proses penting dalam data science untuk memastikan bahwa data yang digunakan dalam analisis atau model adalah akurat, konsisten, dan bebas dari kesalahan. Berikut adalah langkah-langkah yang diambil dalam pembersihan data untuk dataset tersebut:

- a. Handling Missing Data : Tidak terdapat data yang hilang atau null dalam dataset yang ada.
- b. Handling Duplicate Data : Tidak terdapat baris atau kolom yang duplikat dalam dataset yang ada.
- c. Handling Inconsistent Data : Diperlukan penyelarasan nama kolom realSum. Kolom realSum merupakan kolom yang berisikan total biaya penginapan sehingga dapat diganti dengan nama 'price'.
- d. Kesalahan Pengetikan : Tidak terdapat kesalahan pengetikan, baik pada nama kolom maupun nama baris dalam dataset yang ada.
- e. Data Type Conversion : Terdapat beberapa kolom yang harus dilakukan perubahan tipe data, seperti
 - kolom room_type bisa dilakukan perubahan dari object (private room, shared room, entire home/apt) dilakukan encoding menjadi angka dan tipe data dilakukan perubahan dari object menjadi int64.

- kolom `room_shared`, `room_private`, dan `host_is_superhost` bisa dilakukan perubahan isi kolom yang semula TRUE FALSE menjadi angka (1 dan 0) dengan mengubah tipe data boolean menjadi int64.
 - kolom `person_capacity` bisa dilakukan perubahan tipe data menjadi int64 karena isi dalam kolom tersebut merupakan angka bulat.
 - kolom `cleanliness_rating` dan `guest_satisfaction_overall` bisa dilakukan perubahan tipe data dari float64 menjadi int64 karena data yang terdapat dalam kolom tidak ada angka di belakang koma.
 - kolom `week` bisa dilakukan encoding dengan mengubah weekend dan weekday menjadi angka 1 dan 0. Selain itu, dilakukan perubahan tipe data dari object menjadi int64.
 - kolom `city` bisa dilakukan encoding dengan mengubah semua nama kota menjadi angka. Selain itu, dilakukan perubahan tipe data dari object menjadi int64.
 - kolom `unnamed:0` bisa dilakukan drop kolom karena tidak memiliki pengaruh terhadap fitur lain.
- f. Menangani Outliers : Outlier yang ada akan ditangani di data preprocessing menggunakan strategi yang sesuai untuk mengurangi dampak outliers terhadap proses analisis.
 - g. Normalization and Standardization : Fitur-fitur atau kolom yang memiliki sebaran data yang berbeda akan dilakukan penanganan di data preprocessing menggunakan strategi yang sesuai dengan sebaran data yang ada.
 - h. Menyelaraskan Label Kategori atau Nama : Semua label kategori dan nama dalam dataset yang ada telah selaras sehingga semua nama kolom dan data yang dimasukkan berada dalam format yang sama (huruf kecil) tanpa adanya singkatan.
 - i. Validasi Data : Dari dataset yang diperoleh sudah lengkap dan tidak memiliki nilai kosong (null atau NaN). Selain itu, terdapat tipe data yang perlu dilakukan pengubahan dan sudah dipastikan bahwa data yang ada adalah unique dan tidak ada duplikasi.

Data Integration

Pada proses menggabungkan data dari berbagai sumber untuk memberikan pandangan yang lebih komprehensif dan menyeluruh. dataset Airbnb yang diperoleh merupakan dataset yang memiliki format csv dan memiliki nama kolom yang sama sehingga dilakukan penggabungan data berdasarkan nama kolom yang ada dengan menggunakan 'concat'. Yang sebelumnya dilakukan data combine yang menggabungkan dua dataset, yang mewakili data untuk hari kerja dan akhir pekan dari suatu kota tertentu, menjadi satu DataFrame dengan kolom tambahan yang menunjukkan jenis minggu (hari kerja atau akhir pekan) dan nama kota.

EDA

- a. Pemahaman data : Untuk memahami dataset yang ada maka digunakan statistic deskriptif, seperti mean, median, modus, dan standar deviasi untuk memahami karakteristik dasar dari variabel yang ada.
- b. Pembersihan data : Karena tidak ada nilai yang hilang, maka tidak dilakukan handling missing data. Begitu juga dengan handling duplicated data tidak dilakukan karena tidak terdapat data yang duplikat.
- c. Visualisasi Data : Untuk mengetahui distribusi dari data yang ada, maka digunakan histogram dan boxplot untuk memahami distribusi variabel. Selain itu, juga menggunakan bar chart untuk melihat distribusi kategori, dan menggunakan scatter plot untuk melihat hubungan antara dua atau lebih variabel.
- d. Analisis Korelasi : Menggunakan matriks korelasi dengan heatmap untuk menilai kekuatan dan arah hubungan antara variabel numerik.

Feature Engineering

Terdapat fitur baru yang ditambahkan dari variabel yang ada, yaitu nama kota (city) dan week (weekday dan weekend) yang digunakan untuk meningkatkan pemodelan.

Feature

Feature yang akan digunakan adalah realSum, room_type, room_shared, room_private, person_capacity, host_is_superhost, multi, biz, cleanliness_rating, guest_satisfaction_overall, bedrooms, dist, metro_dist, attr_index, attr_index_norm, rest_index, rest_index_norm, lng, lat.

1. Descriptive Analysis

Hasil observasi dari `.info()` dan `.describe()` adalah sebagai berikut:

- a. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?
 - kolom `room_type` bisa dilakukan perubahan dari object (private room, shared room, entire home/apt) dilakukan encoding menjadi angka dan tipe data dilakukan perubahan dari object menjadi `int64`.
 - kolom `room_shared`, `room_private`, dan `host_is_superhost` bisa dilakukan perubahan isi kolom yang semula `TRUE FALSE` menjadi angka (1 dan 0) dengan mengubah tipe data boolean menjadi `int64`.
 - kolom `person_capacity` bisa dilakukan perubahan tipe data menjadi `int64` karena isi dalam kolom tersebut merupakan angka bulat.
 - kolom `cleanliness_rating` dan `guest_satisfaction_overall` bisa dilakukan perubahan tipe data dari `float64` menjadi `int64` karena data yang terdapat dalam kolom tidak ada angka di belakang koma.
 - kolom `week` bisa dilakukan encoding dengan mengubah weekend dan weekday menjadi angka 1 dan 0. Selain itu, dilakukan perubahan tipe data dari object menjadi `int64`.
 - kolom `city` bisa dilakukan encoding dengan mengubah semua nama kota menjadi angka. Selain itu, dilakukan perubahan tipe data dari object menjadi `int64`.
 - kolom `unnamed:0` bisa dilakukan drop kolom karena tidak memiliki pengaruh terhadap fitur lain.
- b. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?
Tidak ada kolom yang memiliki nilai kosong.
- c. Apakah ada kolom yang memiliki nilai summary agak aneh?(min/mean/median/max/unique/top/freq)
 - kolom `realSum` memiliki jarak yang cukup jauh antara nilai mean dan nilai max. Hal ini dapat berakibat adanya outlier yang akan mempengaruhi hasil analisis. Mungkin dapat dilakukan perubahan angka dalam kolom tersebut menjadi log untuk meminimalisir pengaruh outlier.
 - kolom `cleanliness_rating` memiliki jarak yang cukup dekat antara nilai mean dan nilai max. Hal ini mengindikasikan rating yang diberikan oleh customer paling banyak diberikan pada rentang nilai 8-10. Perlu dilakukan analisis lebih lanjut untuk tahu apakah bias atau tidak.

- kolom `guest_satisfaction_overall` memiliki jarak yang cukup dekat antara nilai mean dan nilai max. Hal ini mengindikasikan tingkat kepuasan yang diberikan oleh customer paling banyak diberikan pada nilai yang tinggi (87-100). Perlu dilakukan analisis lebih lanjut untuk tahu apakah bias atau tidak.
- kolom `dist` memiliki rentang jarak yang cukup besar dari 0.01 hingga 25.28. Setiap penginapan memiliki jarak yang berbeda dengan pusat kota dan hal ini menunjukkan bahwa terdapat penginapan yang sangat dekat dengan pusat kota dan sangat jauh dari pusat kota.
- kolom `metro_dist` memiliki rentang jarak yang cukup besar dari 0.002 hingga 14.27. Ini adalah jarak dari akomodasi yang ada di sekitar penginapan tersebut. Hal ini menunjukkan adanya penginapan yang jaraknya sangat dekat hingga terjauh dari transportasi publik.
- kolom `attr_index`, `attr_index_norm`, `rest_index`, dan `rest_index_norm` bisa dianggap sebagai outlier karena adanya rentang nilai mean dan max. Ini bisa menjadi properti yang berlokasi di area yang sangat dekat dengan banyak atraksi atau restoran.
- Nilai mean dan median yang cukup berbeda, terutama di kolom `attr_index_norm` dan `rest_index_norm`, menunjukkan distribusi yang cenderung condong ke bawah, dengan sebagian besar properti tidak memiliki akses maksimal.
- kolom `bedrooms` menunjukkan bahwa penginapan paling banyak menyediakan 1 dan 2 bedroom. Pada kolom tersebut juga terdapat angka 0 sebanyak 4481 yang artinya tidak tersedia bedroom. Hal ini dapat diindikasikan kemungkinan besar adalah properti tipe studio atau akomodasi bersama seperti hostel. Namun, ini bisa membingungkan jika diinterpretasikan secara literal, sehingga perlu dipastikan apakah memang ini definisi yang digunakan.

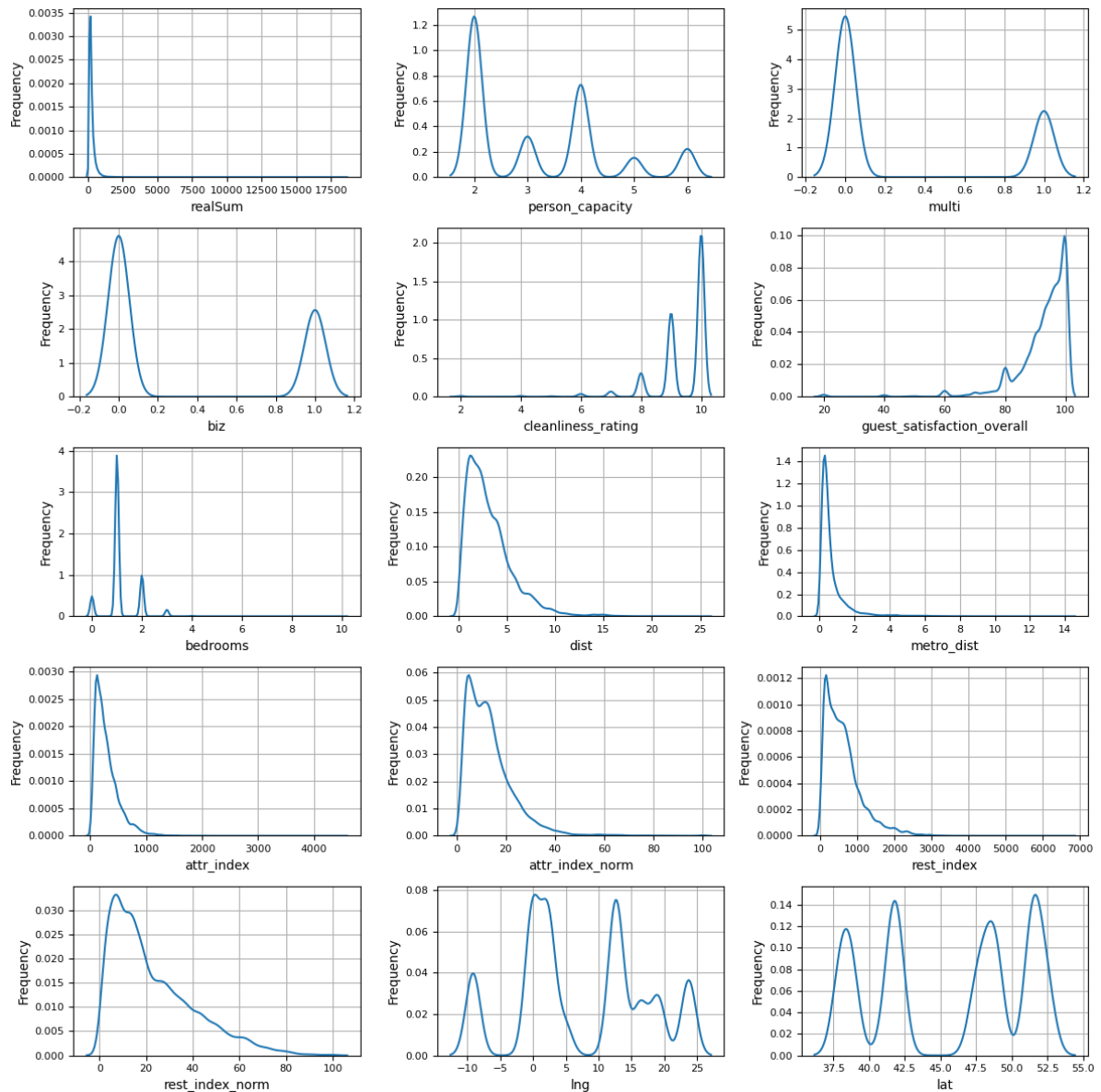
* Untuk masing-masing jenis observasi, tuliskan juga jika tidak ada masalah, misal untuk A:

“Semua tipe data sudah sesuai

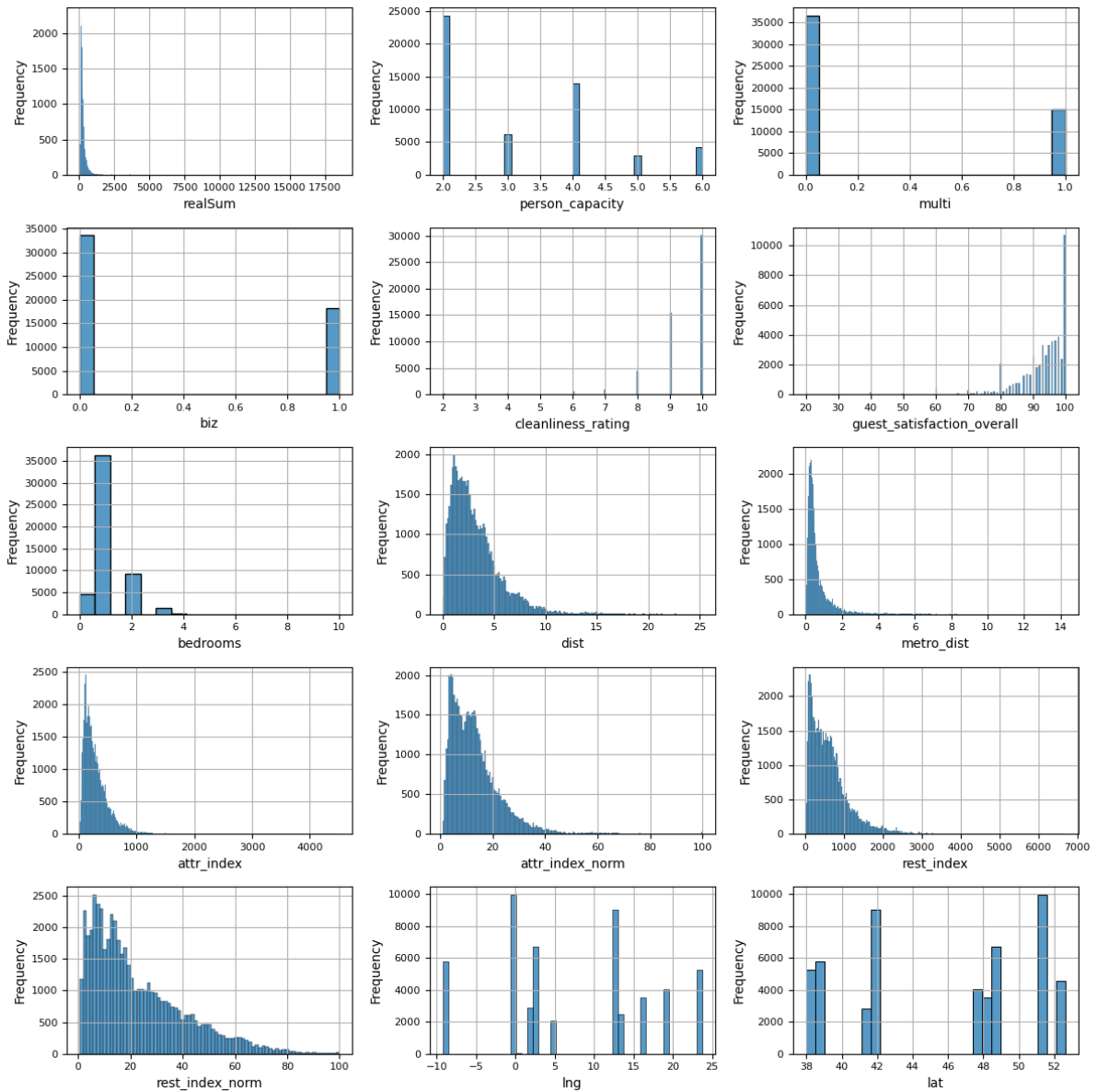
2. Univariate Analysis

Gunakan visualisasi untuk melihat distribusi masing-masing kolom (feature maupun target). Tuliskan hasil observasinya, misalnya jika ada suatu kolom yang distribusinya menarik (misal skewed, bimodal, ada outlier, ada nilai yang mendominasi, kategorinya terlalu banyak, dsb). Jelaskan juga apa yang harus di-follow up saat data pre-processing

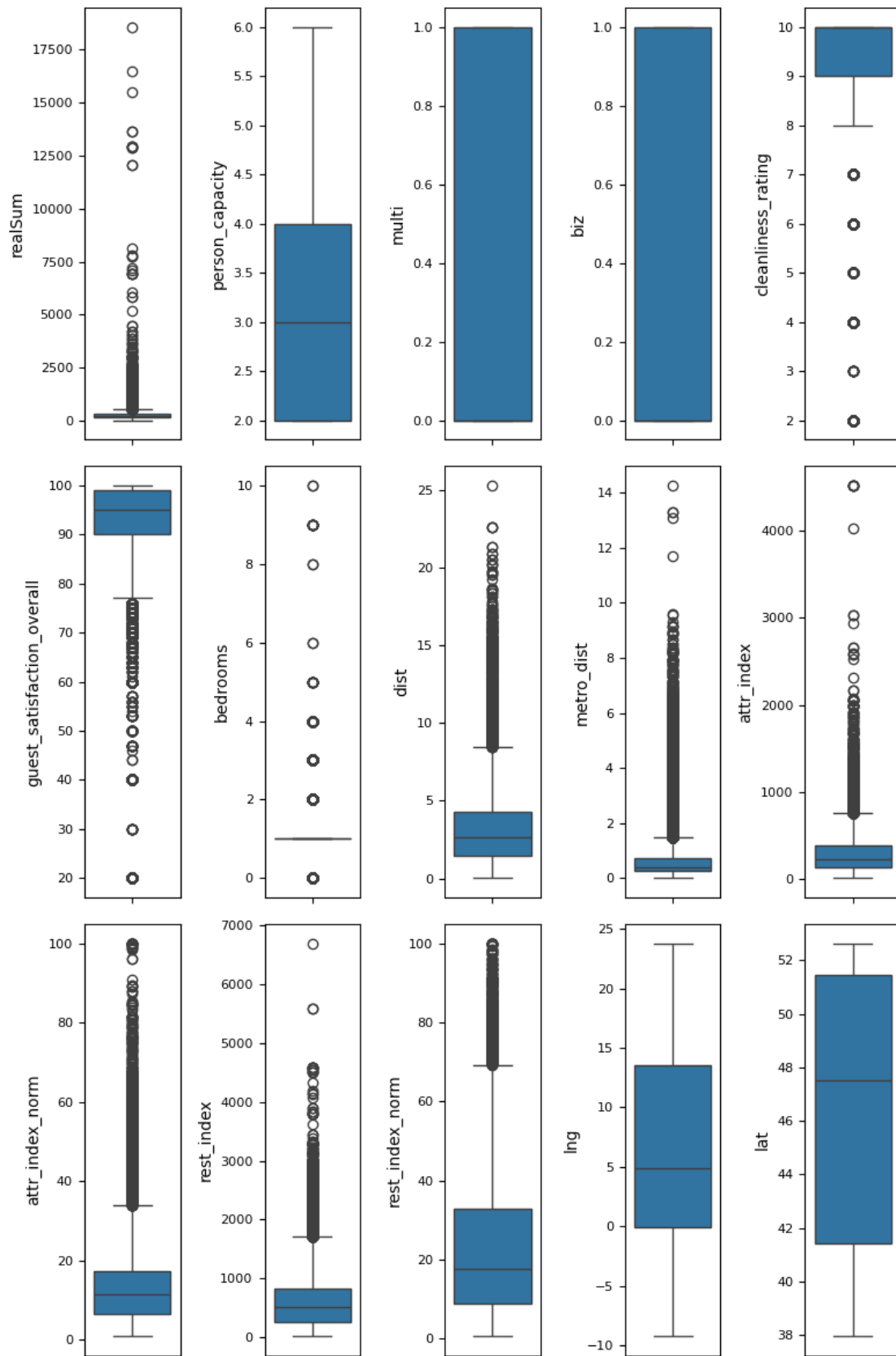
Kde Plot



Histogram

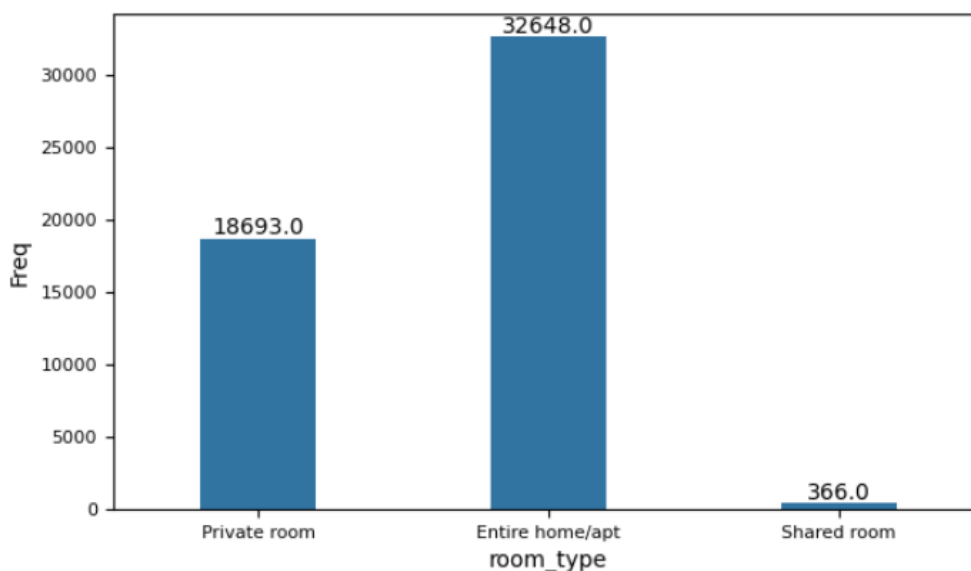


Box Plot

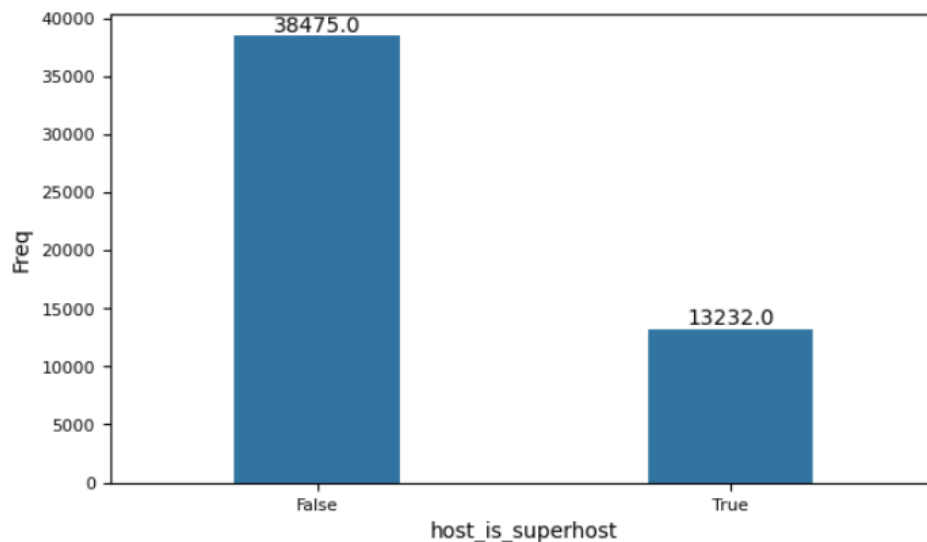


Hasil Observasi :

- Distribusi harga (realSum), jarak terhadap pusat kota (dist), jarak terhadap metro (metro_dist), penilaian daya tarik pasar (attr_index), penilaian daya tarik pasar yang sudah dinormalisasi (attr_index_norm), penilaian berdasarkan keberadaan restaurant (rest_index), dan penilaian berdasarkan keberadaan restaurant yang sudah di normalisasi (rest_index_norm) memiliki distribusi yang berbentuk *right skewed*. Kondisi tersebut menunjukkan bahwa perlu adanya proses normalisasi terhadap feature-feature tersebut dengan menggunakan log transformation.
- Feature harga (realSum) memiliki distribusi yang berbentuk *right skewed* dengan *long tailed* yang berarti mayoritas harga berada di kisaran 0-2.500 EUR dan terdapat nilai outlier yang jauh lebih besar dibandingkan harga lainnya. Feature ini perlu dilakukan log transformasi.
- Rating kebersihan cenderung memiliki nilai yang tinggi yaitu antara nilai 8 sampai dengan 10. Hanya ada beberapa tempat yang memiliki nilai yang rendah. Rating 10 adalah yang nilai rating paling banyak diberikan.
- Distribusi tingkat kepuasan berbentuk left-skewed yang menunjukkan bahwa tingkat kepuasan cenderung memiliki nilai yang besar yaitu berkisar antara 80-100.
- Terdapat outliers pada feature realSum, cleanliness_rating, guest_satisfaction_overall, bedrooms, dist, metro_dist, attr_index, attr_index_norm, rest_index, dan, rest_index_norm.

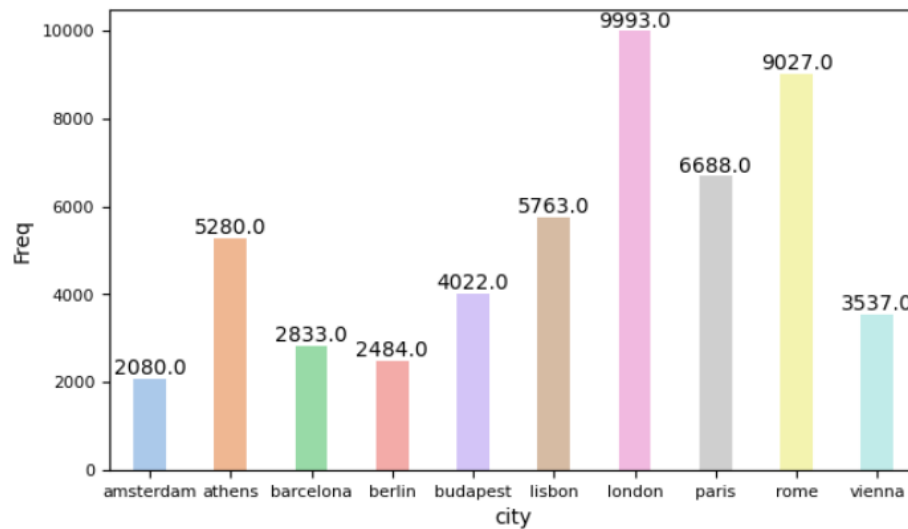


- Terdapat 3 tipe kamar yang disewakan di Airbnb. Tipe kamar yang paling banyak disewakan oleh Airbnb di Eropa adalah Entire home/apartment yaitu sekitar 32.648 tempat. Sedangkan, tipe kamar yang paling sedikit adalah shared room yaitu sebanyak 366 tempat saja. Shared room memiliki jumlah yang jauh lebih sedikit jika dibandingkan dengan kedua tipe kamar lainnya.



Gambar Apa

- Sebagian besar host atau pemilik hunian bukan merupakan superhost.



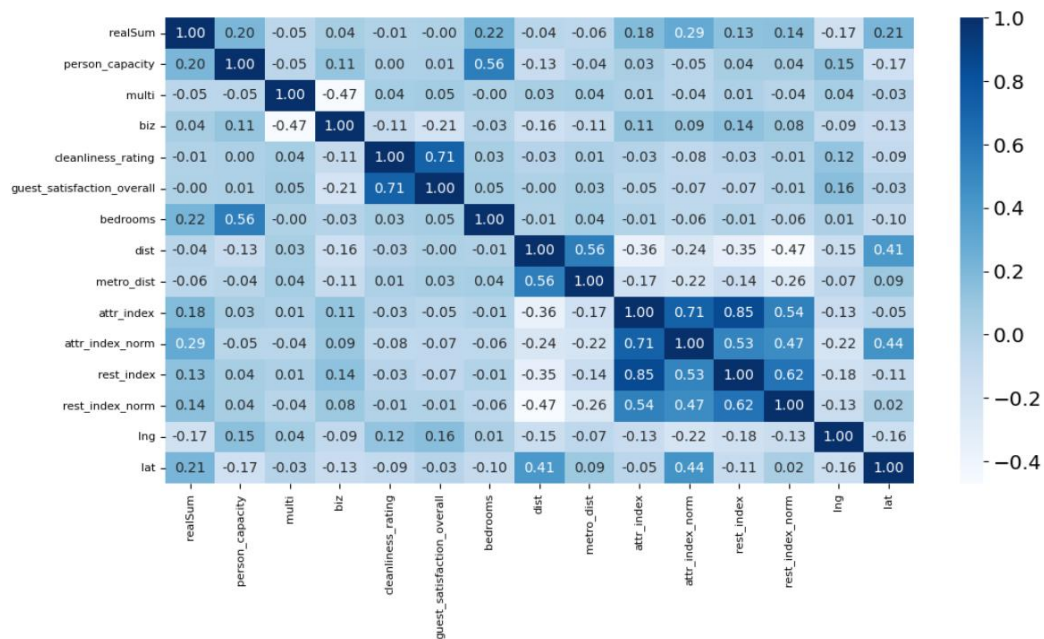
- Kota London memiliki jumlah hunian yang disewakan yang paling banyak yaitu sebanyak 9993 tempat. Sedangkan, Amsterdam memiliki jumlah yang paling sedikit yaitu sebanyak 2080 tempat.

3. Multivariate Analysis

Lakukan multivariate analysis (seperti correlation heatmap dan category plots, sesuai yang diajarkan di kelas). Tuliskan hasil observasinya. Tuliskan juga jika memang tidak ada feature yang saling berkorelasi

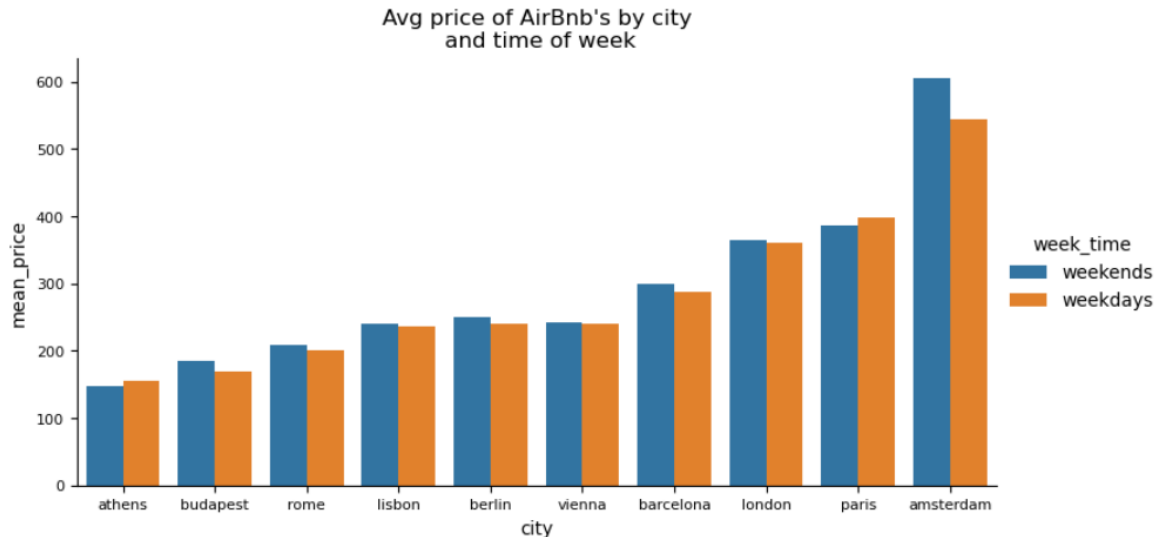
Hasil Observasi:

- Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?

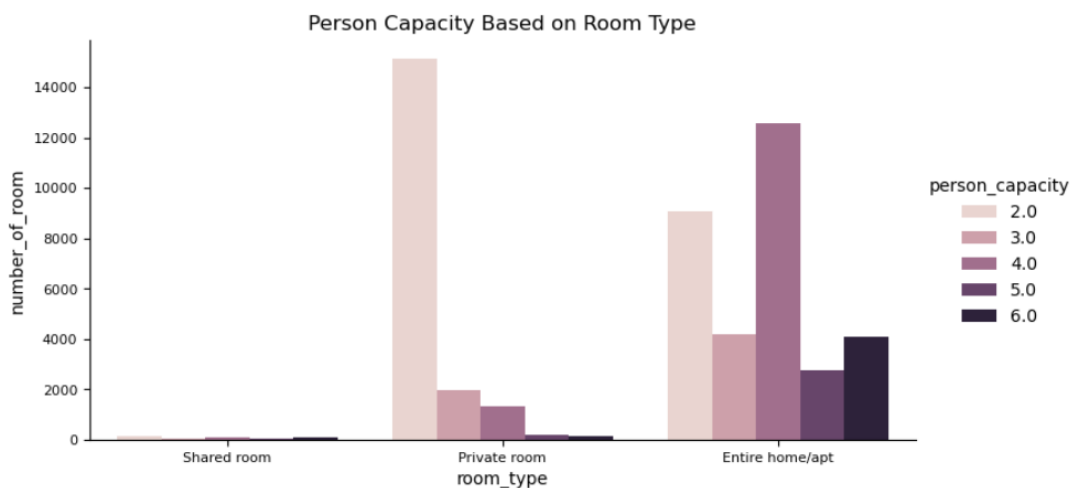


- Setiap feature memiliki korelasi dengan target yang berbeda-beda. Feature kapasitas orang, business indicator, jumlah kamar, attr_index, attr_index_norm, penilaian berdasarkan ketersediaan restoran, dan koordinat garis lintang memiliki korelasi positif terhadap target. Korelasi positif menunjukkan bahwa apabila feature meningkat maka target juga ikut meningkat.
 - Feature yang memiliki korelasi yang paling kuat dengan target adalah feature attr_index_norm atau penilaian berdasarkan daya Tarik lokasi dan daerah sekitar dengan nilai korelasinya adalah 0,29.
 - Feature yang memiliki hubungan yang paling lemah terhadap target adalah feature rating kebersihan dengan nilai korelasinya -0,01.
 - Feature guest_satisfaction_overall tidak memiliki hubungan atau korelasi dengan target.
- Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?

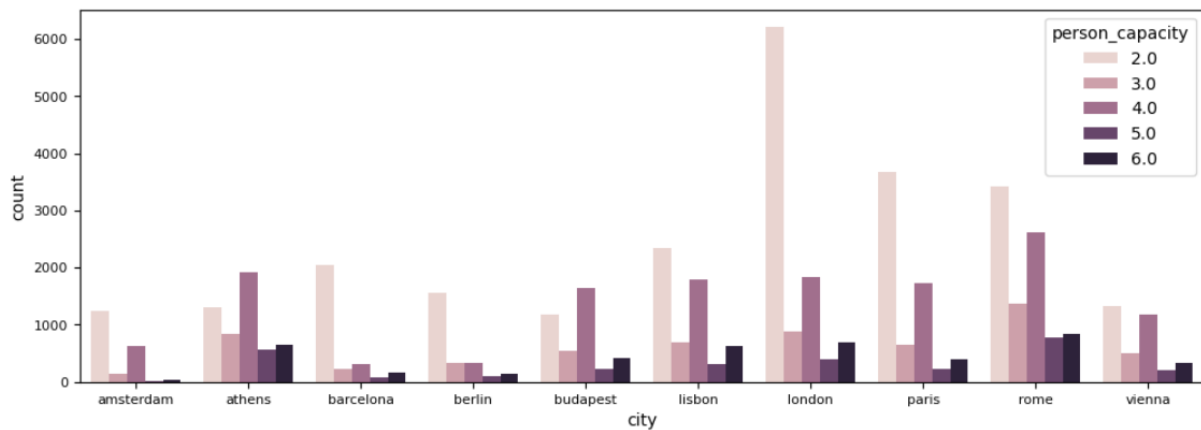
- Feature yang memiliki korelasi yang cukup kuat adalah feature `attr_index` dengan `attr_index_norm`, `attr_index` dengan `rest_index`, `attr_index` dengan `rest_index_norm`, `attr_index_norm` dengan `rest_index`, `attr_index_norm` dengan `rest_index_norm`, dan `rest_index` dengan `rest_index_norm`.



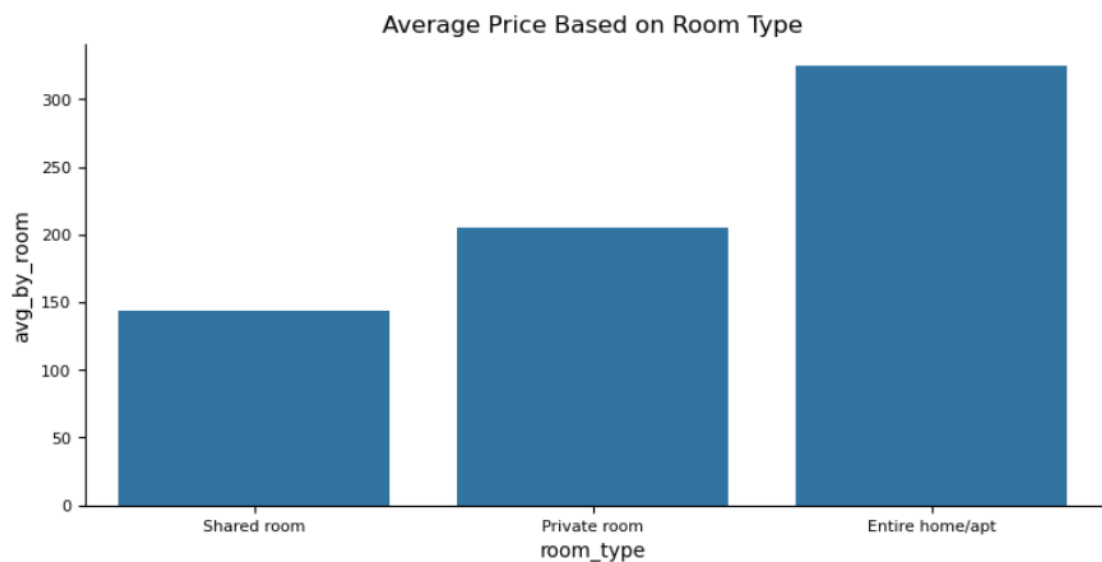
- Rerata harga sewa paling tinggi berada di kota amsterdam dan yang paling rendah adalah kota athens
- Perbandingan antara rata-rata harga sewa pada saat weekend dan weekday tidak terpaut terlalu jauh, umumnya harga sewa pada saat weekend lebih besar dibandingkan ketika weekday. Namun terdapat satu kota yang rata-rata harga sewa pada saat weekday lebih besar dibandingkan pada saat weekend yaitu kota athens



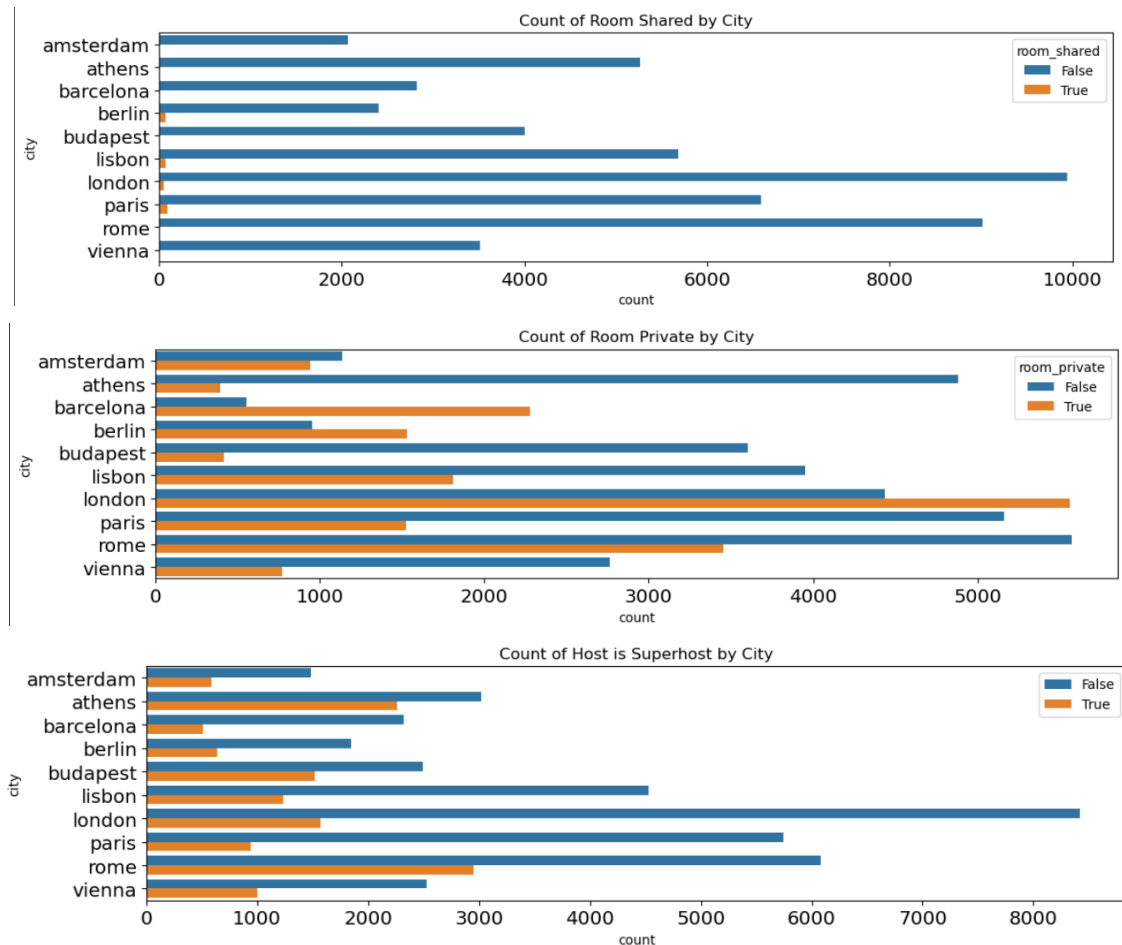
- Hunian sewa yang paling banyak disewakan berupa private room dengan kapasitas 2 orang.



- person capacity yang berisi 2 paling banyak terdapat di kota London dan paling sedikit terdapat di kota Budapest
- person capacity yang berisi 3 paling banyak terdapat di kota Roma dan paling sedikit terdapat di kota Amsterdam
- person capacity yang berisi 4 paling banyak terdapat di kota Roma dan paling sedikit terdapat di kota Berlin
- person capacity yang berisi 5 paling banyak terdapat di kota Roma dan paling sedikit terdapat di kota Amsterdam
- person capacity yang berisi 6 paling banyak terdapat di kota Roma dan paling banyak terdapat di kota Amsterdam



- Rata-rata berdasarkan tipe kamar yang paling tinggi adalah sewa hunian seluruh rumah atau apartemen.



- Penginapan yang terdapat di beberapa kota yang ada, seperti di Amsterdam, Athena, Barcelona, Berlin, Budapest, Lisbon, London, Paris, Rome, dan Vienna banyak yang tidak menyediakan shared room. Meskipun beberapa penginapan yang terdapat di kota Berlin, Lisbon, London, dan Paris menyediakan shared room tapi tidak terlalu banyak. Kota yang paling banyak menyediakan shared room adalah Paris.
- Penginapan yang terdapat di beberapa kota yang ada, seperti di Amsterdam, Athena, Budapest, Lisbon, Paris, Rome, dan Vienna banyak yang tidak menyediakan private room dibandingkan dengan yang menyediakan. Khusus di kota Athena lebih banyak penginapan yang tidak menyediakan private room. Sedangkan di kota Barcelona, Berlin, London terdapat lebih banyak penginapan yang menyediakan private room dibandingkan dengan yang tidak, khususnya di kota Barcelona lebih banyak penginapan yang menyediakan private room.
- Terdapat lebih banyak penginapan di semua kota yang ada yang tidak memberikan pengalaman menginap yang menyenangkan bagi tamu dibandingkan dengan jumlah penginapan yang terdapat di seluruh kota yang ada yang memberikan pengalaman yang menyenangkan bagi tamu.

4. Business Insight

Berdasarkan hasil exploratory data analysis (EDA) yang telah kami lakukan, berikut beberapa insight dan business recommendation yang bisa kami berikan:

Business Insight

- Diketahui properti dengan tipe *entire home/apartment* memiliki permintaan atau minat tertinggi di hampir semua kota. Tak hanya itu, properti ini menjadi properti dengan harga tertinggi dibanding tipe properti lainnya. Hal ini mengindikasikan bahwa rata-rata pengunjung rela membayar lebih mahal untuk properti *entire home/apartment* karena fasilitas yang diberikan seperti area yang lebih luas, kamar yang lebih banyak, fasilitas yang lebih lengkap, pengalaman, hingga privasi dan kenyamanan yang lebih besar;
- Jika melihat tipe dan kapasitas properti, diketahui properti tipe *private room* banyak disewa oleh 2 orang. Besar kemungkinan mereka merupakan sebagai pasangan atau travellers yang senang berpergian. Selain itu, properti dengan tipe *entire home/apartement* rata-rata banyak dipesan oleh rombongan dengan jumlah anggota 3-4 orang, terdapat kemungkinan bahwa rombongan ini merupakan keluarga kecil atau grup pertemanan. Sedangkan itu, rombongan orang banyak (+5 orang) tidak terlalu menunjukkan intensitas pemesanan yang masif, mengindikasikan adanya keterbatasan peluang di kota-kota tertentu;
- Beberapa kota seperti Athena, Budapest, Roma, hingga Lisbon memiliki rata-rata harga sewa yang lebih rendah dibanding dengan kota-kota lainnya, terutama untuk tipe kamar *shared room*. Hal ini dapat mengindikasikan bahwa kota-kota ini lebih banyak dikunjungi oleh wisatawan dengan anggaran yang lebih rendah dan menginginkan akomodasi yang lebih terjangkau. Hal ini berbeda dengan kota-kota besar lainnya yang memiliki harga sewa yang lebih tinggi dan variasi properti yang lebih beragam. Hal ini dapat terjadi karena beberapa faktor seperti daya tarik wisata, pusat kuliner, reputasi kota, infrastruktur, regulasi pemerintah, hingga biaya hidup yang tinggi di kota-kota tersebut.

Business Recommendation:

- Mengingat adanya permintaan tinggi untuk tipe properti *entire home/apartment*, maka baik host maupun Airbnb dapat melakukan strategi tertentu guna meningkatkan daya tarik dan ketersediaan jenis properti ini. Mulai dari menonjolkan keunggulan fasilitas yang lebih lengkap dan memadai, aksesibilitas, dan ragam penawaran lainnya melalui dokumentasi yang jelas, lengkap, dan menarik. Jika dibutuhkan, baik Airbnb atau host dapat menetapkan

dynamic pricing harga sewa properti dan juga promo tergantung seberapa tinggi-rendahnya permintaan di setiap kota. Hal ini dilakukan guna menarik pengunjung sekaligus meningkatkan revenue pemilik properti. Cara lain yang dapat ditempuh bisa juga dengan memberikan panduan harga yang pasti guna memastikan bahwa tarif yang dipatok tetap kompetitif di antara kompetitor lainnya;

- Guna memudahkan pengunjung dalam mendapatkan properti yang sesuai dengan kebutuhan, kami merekomendasikan untuk membangun sebuah model *machine learning* berbasis generative AI yang dapat mengoptimalkan hasil pencarian properti berdasarkan kebutuhan anggaran pengunjung, ragam lokasi wisata terdekat, kapasitas properti, preferensi pelanggan, hingga aksesibilitas dengan pusat kota dan wisata. Tak hanya itu, model ini kiranya juga dapat merekomendasikan paket bundling yang dinamis antara harga sewa properti dengan paket wisata sehingga pengunjung bisa mendapatkan paket liburan yang lebih terjangkau dan memudahkan dalam merencanakan perjalanan yang lebih terstruktur.
- Berdasarkan hasil analisis yang menunjukkan bahwa properti yang dapat menampung dua orang (yang kemungkinan besar merupakan pasangan atau traveller) merupakan properti yang paling banyak dipesan, maka Airbnb dapat memberikan penawaran atau diskon tertentu bagi tamu yang memenuhi kualifikasi tersebut. Untuk tamu keluarga kecil atau grup pertemanan, bisa pula diberikan rekomendasi properti yang cocok untuk kebutuhan dan tipe tamu ini mulai dari fasilitas ramah keluarga, ruang bermain, area yang luas, kamar yang lebih banyak, dan lain-lain. Meskipun ada keterbatasan permintaan untuk grup besar, *Airbnb* bisa mengevaluasi kota-kota yang potensial untuk mengembangkan properti yang cocok bagi rombongan (5+ orang). Kota wisata dengan daya tarik keluarga besar atau acara khusus seperti pernikahan, konferensi, atau festival dapat memiliki potensi akomodasi besar. Jika ada potensi ini, *Airbnb* dapat membantu host untuk menyesuaikan fasilitas dan memperbaiki properti agar dapat menampung lebih banyak tamu.
- Bagi kota-kota dengan harga rata-rata sewa yang lebih rendah, ada baiknya bagi host maupun Airbnb untuk lebih fokus pada segmen wisatawan dengan budget atau biaya rendah disertai dengan personalisasi kebutuhan pengunjung. Hal ini dapat meliputi akomodasi-akomodasi yang terjangkau dan *worth-to-try* agar pengunjung berkenan datang dan menginap. Diperlukan pula rekomendasi khusus bagi para wisatawan yang mencari akomodasi terjangkau agar mereka mau untuk menginap lebih lama. Untuk beberapa kasus, host juga bisa menyediakan opsi *shared room* dengan harga yg lebih rendah dengan fasilitas standar. Kemungkinan besar hal ini akan cocok bagi kalangan backpacker, solo

traveller, dan lain-lain. Bagi kota-kota besar dengan harga yang lebih tinggi, bisa menekankan promosi dan daya tarik eksklusif yang ditawarkan oleh kota tersebut.

5. Git

Kolaborasi pengerjaan bersama digunakan Git dan GitHub, terutama untuk mengelola versi atau history dari file, berikut adalah langkah-langkah serta penjelasan lengkapnya:

a. Membuat Repository Git

Inisialisasi Repository: ini adalah pertama kali menggunakan Git di folder proyek, buka command prompt di direktori proyek dan jalankan perintah berikut:

```
:>git init
```

Perintah ini akan membuat repository Git lokal di direktori tersebut, sehingga dapat mulai melakukan version control di dalamnya.

b. Buat Repository di GitHub:

- Masuk ke akun GitHub dan klik tombol New Repository.
- Isi nama repository FinPro-8,
- Kemudian buat repository menjadi Public (terbuka untuk umum) Klik Create Repository.

c. Hubungkan Repository Lokal dengan Repository di GitHub:

Salin URL dari repository GitHub yang baru dibuat.

Di terminal, hubungkan repository lokal dengan repository di GitHub menggunakan perintah berikut:

```
:>git remote add origin https://github.com/rdarmawan48/FinPro8-InsightSeekers
```

d. Upload File Proyek (Notebook dan File Lainnya) ke Repository

- Tambahkan File ke Repository:
- Pindahkan file notebook (.ipynb), file .py, atau file lainnya ke dalam folder proyek. Contohnya, file airbnb_analysis.ipynb.
- Tambahkan file tersebut ke staging area dengan perintah:

```
:>git add airbnb_analysis.ipynb
```
- Jika ingin menambahkan semua file sekaligus:

```
:>git add
```

e. Commit Perubahan: Buat pesan commit yang menggambarkan perubahan yang dilakukan, misalnya:

```
:>git commit -m "Add FinPro8 homework files"
```

- f.** Push ke Repository GitHub: Kirim file dan commit ke repository GitHub dengan perintah berikut:

```
:>git push -u origin main
```

- g.** Buat dan Update File README.md

Buat File README.md: Untuk memberikan informasi tentang proyek, membuat file README.md di root folder proyek.

- h.** Konten README.md:

FinPro8-InsightSeekers

Kelompok 8 Data Science InsightSeekers

Airbnb Europe EDA Project

Laporan Homework stage 1 ini berisi penjelasan , Exploratory Data Analysis (EDA) Final Project Airbnb Prices in European Cities sampai dengan Business Insight

Dataset

Data Quality

Data Cleaning

Data Integration

EDA

Feature Engineering

1. Descriptive Analysis
2. Univariate Analysis
3. Multivariate Analysis
4. Business Insight
5. Git

Getting Started

Clone this repository.

Open Homework_FinPro-8_InsightSeekers-Stage_1.ipynb notebook.

- i.** Commit dan Push README.md: Setelah membuat atau memperbarui README.md, dilakukan commit dan push ke GitHub:

```
:>git add README.md
```

```
:>git commit -m "Add project summary and insights to README.md"
```

```
:>git push origin main
```

- j.** Repository di GitHub setelah status public digunakan team sebagai kolaborator