

Report Progress

Stage 0

Unison
Muhammad Ath Thaariq Amir Chandra

Sutan Faisal

Farah Nelia Satriani

Hafizh Fakhri Alfarisy

Salsabila Nur Yasmin

Unison



M Ath Thaariq A C

Data Analyst



Sutan Faisal

Data Scientist



Farah Melia S

Project Manager



Hafizh Fakhri A

Data Engineer



An aerial, high-angle view of a dense urban skyline, likely New York City, with numerous skyscrapers and buildings. The image is faded and serves as a background for the text.

Stage 0

**Extracting insights and create
solutions with potential business
or industrial impact.**

Dataset Exploration

Explorasi Dataset Awal: Calendar, Listings, dan Reviews

1. Pemahaman Umum Dataset

Dataset Airbnb yang tersedia terdiri dari tiga file utama:

- a. Calendar.csv (4 kolom)
Kolom: listing_id, date, available, price
Masalah: price memiliki 459.028 nilai hilang.
- b. Listings.csv (92 kolom)
Kolom utama: id, name, host_id, price, room_type, number_of_reviews, dll.
Masalah: Banyak nilai hilang dalam kolom seperti square_feet, security_deposit, cleaning_fee, weekly_price, dan monthly_price.
- c. Reviews.csv (6 kolom)
Kolom: listing_id, reviewer_id, reviewer_name, comments
Masalah: comments memiliki 18 nilai hilang.

Dataset Exploration

2. Struktur data

Setelah melakukan eksplorasi menggunakan Pandas di Python, berikut adalah insight awal:

- a. Calendar.csv memiliki kolom ``listing_id``, ``date``, ``available``, dan ``price`` dengan beberapa nilai kosong pada kolom ``price``.
- b. Listings.csv berisi berbagai informasi, termasuk jenis properti, lokasi, harga, dan kebijakan pembatalan.
- c. reviews.csv berisi informasi ulasan, termasuk ``listing_id``, ``reviewer_name``, dan ``comments``.

Dataset Exploration

3. Potensi Insight dari Dataset & Identifikasi

- a. Tren Harga Properti: Dari `calendar.csv`, kita bisa melihat fluktuasi harga harian dan apakah ada pola harga musiman.
- b. Korelasi Antara Fitur Listing: Misalnya, apakah `room_type`, `location`, atau `host_is_superhost` memengaruhi `price`?
- c. Popularitas Listing Berdasarkan Review: Dari `reviews.csv`, kita bisa melihat jumlah dan isi ulasan untuk menentukan properti dengan performa terbaik.

Data yang Hilang atau Perlu Dibersihkan

- a. Kolom `price` dalam `calendar.csv` memiliki nilai kosong jika properti tidak tersedia.
- b. Kolom `description` dalam `listing.csv` terkadang kosong atau sangat panjang.
- c. Kolom `reviews` dalam `reviews.csv` memiliki beberapa entri yang mengandung teks yang tidak terstruktur.

Dataset Exploration

Kesimpulan

1. Perlu pembersihan data: Penanganan nilai hilang di kolom harga (price, security_deposit, dll.).
2. Analisis harga: Bisa mengeksplorasi tren harga berdasarkan waktu dan lokasi.
3. Analisis review: Bisa menghubungkan jumlah review dengan harga atau ketersediaan listing.

Industry Problem Discussion

1. Identifikasi Industri yang Berkaitan

Dataset ini berasal dari platform Airbnb, yang beroperasi di sektor:

- Industri Perhotelan & Pariwisata
- Industri Properti & Sewa Jangka Pendek
- Industri Analitik Pasar Real Estate

Data ini mencakup informasi tentang harga sewa properti, ketersediaan kamar, ulasan pelanggan, serta karakteristik properti dan host. Maka dari itu, permasalahan bisnis yang dapat diselesaikan diantaranya adalah:

a. Optimalisasi Harga Sewa (Dynamic Pricing)

Masalah: Host kesulitan menentukan harga optimal untuk memaksimalkan pendapatan tanpa kehilangan peminat.

Solusi:

- Analisis calendar.csv untuk tren harga dan ketersediaan.
- Perbandingan harga dengan kompetitor menggunakan listings.csv.
- Prediksi harga optimal dengan machine learning.

Industry Problem Discussion

b. Prediksi Tren Permintaan & Ketersediaan

Masalah: Host perlu memahami pola permintaan agar dapat menyesuaikan strategi pemasaran dan harga.

Solusi:

- Gunakan calendar.csv untuk mengidentifikasi musim dengan permintaan tinggi.
- Analisis jumlah ulasan di reviews.csv sebagai indikator okupansi.
- Identifikasi area dengan permintaan tinggi melalui listings.csv.

c. Analisis Kualitas Layanan & Kepuasan Pelanggan

Masalah: Host ingin meningkatkan pengalaman tamu untuk mendapatkan ulasan positif.

Solusi:

- Gunakan Sentiment Analysis pada reviews.csv untuk memahami feedback pelanggan.
- Analisis review_scores_rating dan host_response_time di listings.csv untuk menemukan faktor utama kepuasan pelanggan.

Industry Problem Discussion

d. Faktor Keberhasilan Listing

Masalah: Host perlu mengetahui faktor utama yang meningkatkan daya saing properti mereka.

Solusi:

- Analisis listings.csv untuk melihat pengaruh fasilitas, harga, dan kebijakan pembatalan.
- Identifikasi fitur yang paling disukai tamu melalui reviews.csv.

Kesimpulan:

- Dataset Airbnb ini relevan untuk pengoptimalan harga, prediksi permintaan properti, dan analisis kepuasan pelanggan.
- Pendekatan berbasis data dapat membantu host Airbnb meningkatkan strategi pemasaran dan pendapatan mereka.
- Studi akademik dan industri mendukung bahwa faktor-faktor seperti harga, lokasi, dan ulasan pelanggan memainkan peran penting dalam performa properti di Airbnb.

Formulasi Pertanyaan Analitis untuk Eksplorasi Dataset Airbnb

Berdasarkan eksplorasi dataset dan diskusi mengenai permasalahan industri, berikut adalah tiga pertanyaan analitis utama yang dapat mengarahkan langkah-langkah analisis lebih lanjut:

1. **Faktor yang Mempengaruhi Harga Sewa**
Masalah: Host sering kesulitan menentukan harga optimal.
Solusi:
 - Gunakan listings.csv untuk menganalisis variabel seperti lokasi, tipe properti, jumlah kamar, dan fasilitas.
 - Gunakan calendar.csv untuk melihat pola harga berdasarkan musim atau periode tertentu.
 - Terapkan regresi linear atau machine learning untuk memprediksi harga optimal.
2. **Tren Permintaan dan Ketersediaan Properti**
Masalah: Host ingin mengetahui periode dengan permintaan tertinggi untuk menyesuaikan harga dan strategi pemasaran.
Solusi:
 - Gunakan calendar.csv untuk menganalisis pola pemesanan berdasarkan waktu.
 - Gunakan reviews.csv sebagai indikator tingkat okupansi.
 - Identifikasi musim ramai dan sepi dengan analisis tren waktu (time series).

Formulasi Pertanyaan Analitis untuk Eksplorasi Dataset Airbnb

3. Pengaruh Sentimen Pelanggan terhadap Performa Listing

Masalah: Ulasan pelanggan sangat mempengaruhi keputusan calon penyewa.

Solusi:

- Gunakan Sentiment Analysis pada reviews.csv untuk memahami feedback pelanggan.
- Analisis hubungan antara sentimen, skor rating, dan jumlah pemesanan di listings.csv.
- Identifikasi aspek layanan yang perlu ditingkatkan untuk mendapatkan rating lebih tinggi.

Kesimpulan dan Langkah Selanjutnya

Analisis ini membantu host, investor, dan platform Airbnb dalam menentukan harga optimal, memahami tren permintaan, serta meningkatkan kualitas layanan berdasarkan ulasan pelanggan. Langkah berikutnya adalah menerapkan strategi berbasis data untuk meningkatkan daya saing properti.

Metrik Keberhasilan yang Digunakan

1. Tujuan Proyek
Menganalisis data listing Airbnb untuk memahami faktor yang memengaruhi harga sewa, tingkat okupansi, dan kepuasan pelanggan guna mendukung pengambilan keputusan bisnis.
2. Metrik Keberhasilan
 - a. Evaluasi Kualitas Data
 - Jumlah Data Hilang: <5% agar tetap valid untuk analisis.
 - Distribusi Variabel Kunci: Harga sewa, jumlah ulasan, dan skor ulasan harus memiliki distribusi yang wajar tanpa outlier ekstrem.
 - b. Wawasan Bisnis
 - Harga Sewa Rata-rata per Wilayah: Identifikasi area dengan harga tertinggi dan terendah.
 - Tingkat Okupansi: Rasio hari disewa vs. total hari dalam setahun untuk memahami pola permintaan.
 - Faktor yang Mempengaruhi Harga: Korelasi harga dengan lokasi, jumlah kamar, dan fasilitas.
 - Kepuasan Pelanggan: Analisis skor ulasan untuk menemukan faktor utama yang memengaruhi rating.

Metrik Keberhasilan yang Digunakan

c. Model Prediksi (Jika Diterapkan)

- Prediksi Harga (Dynamic Pricing):
 - Mean Absolute Error (MAE) < \$50 untuk akurasi yang baik.
 - R-squared (R^2) > 0.5 menunjukkan model mampu menjelaskan variabilitas harga.
- Sentimen Pelanggan:
 - Distribusi Sentimen (%) untuk melihat proporsi ulasan positif, netral, dan negatif.

3. Referensi dan Benchmarking

- Studi kasus Airbnb dan platform serupa untuk membandingkan metrik dengan tren industri
- Jurnal akademik tentang strategi harga dan okupansi dalam bisnis sewa jangka pendek.
- Polaritas Sentimen untuk memahami kecenderungan ulasan.

Assign Roles Airbnb Listing Analysis

Project Manager

Tugas:

- Menyusun timeline dan milestone proyek.
- Mengkoordinasikan komunikasi antar tim.
- Mengelola task dan progress menggunakan project management tools.

Tools:

Trello, Jira, atau Asana untuk manajemen proyek.
WhatsApp/Slack untuk komunikasi tim.

Data Engineer

Tugas:

- Mengumpulkan dan membersihkan dataset (listings.csv, calendar.csv, reviews.csv).
- Menyusun pipeline data untuk pemrosesan machine learning.
- Mengelola penyimpanan data (misalnya, dalam SQL atau cloud storage).

Tools:

Python (pandas, SQLAlchemy) untuk ETL.
Apache Airflow atau Luigi untuk pipeline data.
BigQuery atau PostgreSQL untuk penyimpanan.

Data Scientist

Tugas:

- Membangun dan melatih model prediksi harga sewa (Dynamic Pricing).
- Melakukan analisis sentimen pada ulasan pelanggan dengan NLP.
- Menentukan faktor yang paling berpengaruh dalam penentuan harga dan okupansi.

Tools:

Scikit-learn, TensorFlow, atau PyTorch untuk machine learning.
NLTK atau spaCy untuk analisis sentimen.
Jupyter Notebook untuk eksplorasi model.

Data Analyst

Tugas:

- Menganalisis tren harga dan okupansi berdasarkan data historis.
- Memberikan insight bisnis terkait faktor yang mempengaruhi harga.
- Membuat visualisasi data dan laporan untuk rekomendasi strategi.

Tools:

Tableau atau Power BI untuk visualisasi data.
SQL untuk kueri database.
Excel atau Google Sheets untuk analisis cepat.

The background of the slide is a repeating pattern of 3D cubes in various shades of teal and blue, creating a sense of depth and movement.

Raih Mimpi **#TanpaBatas**



Raih Mimpi **#TanpaBatas**



Thank you!

Raih Mimpi **#TanpaBatas**