

Analisis Kinerja Vendor Taksi di New York City

Sesilia Putri Subandi¹, Dearn Monica Manik², Pandra Insani Putra Azwar³,
Izza Lutfia⁴, Elok Fiola⁵, Syalaisha Andini Putriansyah⁶

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera

E-mail: ¹sesilia.122450012@student.itera.ac.id,

²dearni.122450075@student.itera.ac.id, ³pandra.121450137@student.itera.ac.id,

⁴izza.122450090@student.itera.ac.id, ⁵elok.122450051@student.itera.ac.id

⁶syalaisha.122450111@student.itera.ac.id

Abstrak

Transportasi memiliki peran penting dalam perkembangan suatu kota, terutama dalam meningkatkan mobilitas masyarakat dan mendukung pertumbuhan ekonomi. Di New York City, layanan taksi menjadi salah satu mode transportasi utama yang banyak digunakan. Tujuan dilakukannya penelitian ini untuk menganalisis jumlah transaksi, tren tarif, dan pola perjalanan per bulan berdasarkan Vendor ID. Metode yang digunakan dalam penelitian ini yaitu EDA (Exploratory Data Analysis) dan divisualisasikan dalam bentuk grafik. EDA dilakukan terhadap data yang sudah dilakukan preprocessing terlebih dahulu. Hasil EDA divisualisasikan menjadi tiga grafik berbeda yang mewakili jumlah transaksi, tren tarif, dan pola perjalanan per bulan. Jumlah transaksi paling tinggi dihasilkan oleh Vendor 2, diikuti oleh Vendor 1 dan Vendor 6. Tarif rata-rata tertinggi yaitu pada Vendor 2 seharga 20 USD pada November 2024, diikuti oleh Vendor 1 dengan harga sekitar 20 USD dan Vendor 6 dengan harga paling rendah mendekati 0 USD pada rentang waktu yang sama. Untuk pola perjalanan, Vendor 2 mengalami lonjakan di Desember 2008 hingga menyentuh angka 20 mil, kemudian kembali mengalami penurunan dan stabil di akhir tahun 2024 dengan jarak tempuh tertinggi di November 2024 sekitar 5 mil. Pada waktu yang sama di November 2024, Vendor 6 menyentuh rata-rata jarak tempuh 10 mil dan Vendor 1 memiliki jarak tempuh terendah sejauh 2.5 mil. Berdasarkan hasil, dapat ditarik kesimpulan bahwa Vendor 2 dari NYC TLC memiliki produktivitas paling tinggi yang tercatat berdasarkan jumlah transaksi dan jarak tempuh rata-rata.

Kata kunci: EDA, New York City, Taksi, Visualisasi

Abstract

Transportation plays an important role in the development of a city, especially in increasing people's mobility and supporting economic growth. In New York City, taxi services are one of the main modes of transportation that are widely used. The purpose of this study was to analyze the number of transactions, fare trends, and travel patterns per month based on Vendor ID. The method used in this study is EDA (Exploratory Data Analysis) and is visualized in the form of graphs. EDA is carried out on data that has been preprocessed first. The results of EDA are visualized into three different graphs representing the number of transactions, fare trends, and travel patterns per month. The highest number of transactions was generated by Vendor 2, followed by Vendor 1 and Vendor 6. The highest average fare was Vendor 2 for 20 USD in November 2024, followed by

Vendor 1 around 20 USD and Vendor 6 with the lowest price approaching 0 USD in the same time period. For the travel pattern, Vendor 2 experienced a spike in December 2008 to reach 20 miles, then decreased again and stabilized at the end of 2024 with the highest mileage in November 2024 around 5 miles. At the same time in November 2024, Vendor 6 touched an average mileage of 10 miles and Vendor 1 had the lowest mileage of 2.5 miles. Based on the results, it can be concluded that Vendor 2 from NYC TLC has the highest productivity recorded based on the number of transactions and average mileage.

Keywords: EDA, New York City, Taxi, Visualization

1. PENDAHULUAN

Transportasi umum berperan penting dalam pertumbuhan ekonomi kota dengan memfasilitasi mobilitas penduduk dan memperlancar distribusi barang serta jasa (Putri, 2022). Salah satu moda transportasi yang banyak digunakan adalah taksi. Taksi menjadi pilihan transportasi karena fleksibel dan mudah ditemukan terutama di kota metropolitan dengan aktivitas ekonomi yang tinggi seperti *New York*. Sebagai bagian dari sistem transportasi, taksi membantu menghubungkan berbagai wilayah dalam kota dan memberikan alternatif bagi masyarakat yang tidak memiliki kendaraan pribadi.

Meskipun memiliki berbagai keunggulan, industri taksi juga menghadapi berbagai tantangan seperti ketimpangan antara permintaan dan ketersediaan layanan, tarif yang bervariasi, serta persaingan ketat antar vendor taksi. Ketidakseimbangan antara jumlah taksi dan permintaan ke berbagai wilayah dapat menyebabkan kesulitan bagi penumpang dalam mendapatkan layanan, terutama ke daerah dengan kepadatan penduduk yang tinggi atau pada jam sibuk.

Di *New York City*, layanan taksi dikelola oleh *NYC Taxi & Limousine Commission (TLC)*, yang bertanggung jawab dalam mengatur operasional serta memastikan standar pelayanan yang sesuai. Untuk memahami dan mengevaluasi kinerja layanan taksi di NYC, diperlukan analisis kinerja berbagai vendor taksi di *New York City* berdasarkan data perjalanan dalam skala besar. Dengan pendekatan berbasis big data, studi ini diharapkan dapat memberikan wawasan mengenai efektivitas layanan taksi di NYC guna meningkatkan kualitas layanan transportasi perkotaan.

Tujuan dilaksanakannya penelitian ini yaitu:

- Menganalisis jumlah transaksi layanan taksi per bulan berdasarkan Vendor ID.
- Menganalisis tren tarif rata-rata per bulan layanan taksi di NYC berdasarkan Vendor ID.
- Menganalisis pola perjalanan taksi berdasarkan jarak tempuh rata-ratanya.
- Menganalisis Vendor dengan produktivitas paling baik berdasarkan jumlah transaksi dan jarak tempuhnya.

2. METODE

2.1 Data

Penelitian ini menggunakan dataset yang berasal dari *NYC Taxi & Limousine Commission (TLC)*, lembaga yang secara resmi mengelola data perjalanan taksi. Dataset yang dianalisis dalam studi kasus ini berfokus pada *Yellow Taxi Trip Data*, yang mencakup perjalanan taksi

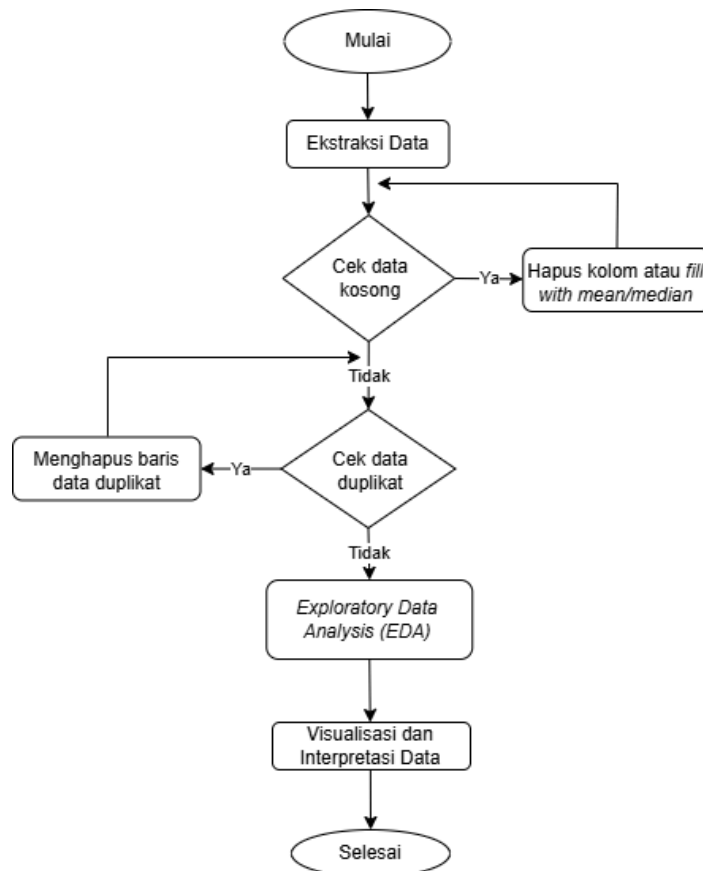
konvensional *Yellow Medallion Taxicabs*. Salah satu informasi penting dalam dataset ini adalah VendorID, yang mencatat identitas penyedia layanan pembayaran serta perusahaan teknologi yang menangani transaksi tarif taksi.

2.2 Diagram Alir

Proses penelitian ini terbagi menjadi dua tahap utama: ekstraksi dan pengolahan data (ETL), serta analisis eksploratif data (EDA). Pada tahap pertama, data dari *Yellow Taxi Trip Data* yang tersimpan di Amazon AWS S3 diambil dan dibersihkan untuk keperluan analisis. Proses ini meliputi pengumpulan data mentah dan pembersihan nilai yang hilang atau tidak valid, serta penyesuaian format agar lebih terstruktur. Setelah itu, data dimuat ke dalam sistem penyimpanan untuk mempermudah analisis selanjutnya.

Setelah data siap, tahap analisis eksploratif (EDA) dilakukan untuk mengidentifikasi pola dan karakteristik melalui statistik dan visualisasi. Langkah ini mencakup identifikasi distribusi data, deteksi anomali, dan analisis berbagai indikator. Hasil dari analisis ini kemudian divisualisasikan dalam grafik dan diagram, yang bertujuan untuk memudahkan pemahaman serta memberikan wawasan yang lebih dalam mengenai data yang telah dianalisis.

Tahapan atau proses yang dilakukan pada penelitian ini dapat dilihat pada Gambar 1. berikut:



Gambar 1. Diagram alir penelitian

2.3 Ekstraksi Data

Metode Ekstraksi data merupakan teknik yang digunakan untuk mengambil, mengubah, mengolah data dari berbagai sumber agar siap untuk dianalisis. Ekstraksi data biasanya melibatkan proses ETL (*Extract, Transform, Load*). Menurut Kimball & Caserta (2011), ETL merupakan pendekatan yang digunakan untuk mengumpulkan data dari berbagai sumber, membersihkan serta menyusun ulang formatnya agar sesuai dengan kebutuhan analisis, lalu menyimpannya dalam lingkungan yang lebih terstruktur seperti data *warehouse* (Kimball dan Caserta, 2011).

2.4 EDA (*Exploratory Data Analysis*)

Exploratory Data Analysis (EDA) adalah proses eksplorasi awal terhadap data untuk memahami pola, distribusi, dan karakteristik sebelum dilakukan analisis lebih lanjut. EDA bertujuan untuk menemukan wawasan penting, mendeteksi anomali, serta memastikan kualitas data agar siap digunakan dalam pemodelan atau pengambilan keputusan.

Metode EDA yang digunakan dalam penelitian ini mengadopsi pendekatan analisis univariat dalam studi deskriptif, yang berfokus pada pemecahan permasalahan aktual untuk membantu mengidentifikasi pola tersembunyi dan memastikan bahwa data siap untuk tahap analisis lebih lanjut seperti pemodelan statistik atau *machine learning*.

2.4.1 Karakteristik Data

Merupakan tahap awal dalam eksplorasi data yang bertujuan untuk memahami struktur dataset. Karakteristik data mencakup tipe data (numerik, kategorikal, waktu, teks), ukuran dataset, distribusi, kualitas, hubungan antar variabel, dan keunikan data.

2.4.2 Transformasi Data

Transformasi data merupakan proses penting dalam analisis data yang bertujuan untuk meningkatkan kualitas analisis dengan memastikan variabel yang digunakan memiliki format yang sesuai. Proses ini mencakup berbagai teknik, seperti mengatasi nilai yang hilang dengan mengisi atau menghapus data yang tidak lengkap agar dataset tetap konsisten. Selain itu, transformasi data juga melibatkan penghapusan outlier untuk mencegah pengaruh nilai ekstrim yang dapat mengganggu hasil analisis. Standarisasi format waktu juga diperlukan untuk memastikan konsistensi data temporal, sehingga analisis dapat dilakukan dengan lebih akurat. Selain itu, deteksi dan penghapusan data duplikat dilakukan untuk menghindari bias dalam analisis. Dengan menerapkan langkah-langkah ini, transformasi data memastikan bahwa dataset siap digunakan dalam analisis secara efektif dan akurat.

2.4.3 Visualisasi

Penelitian ini memanfaatkan visualisasi data untuk menyajikan hasil analisis dalam bentuk diagram atau grafik, sehingga memudahkan pemahaman terhadap pola, distribusi, dan hubungan antar variabel dalam dataset. Teknik visualisasi yang digunakan mencakup Heatmap, Histogram, dan Scatterplot, yang membantu mengidentifikasi pola dalam data secara lebih jelas dan informatif.

3. HASIL DAN PEMBAHASAN

3.1 Jumlah Transaksi

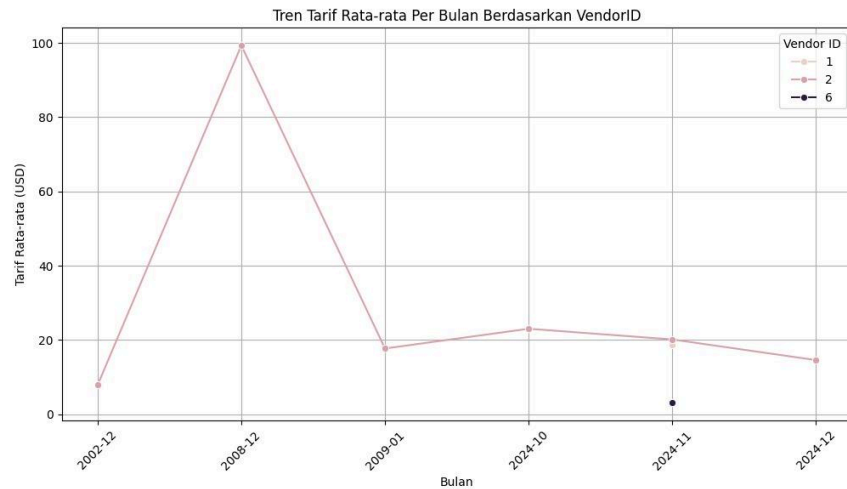


Gambar 2. Jumlah transaksi per bulan berdasarkan Vendor ID

Sebuah grafik jumlah transaksi per bulan berdasarkan Vendor ID menunjukkan beberapa anomali dalam data. Sebelum 2024, tidak ada transaksi yang tercatat, hal ini mungkin disebabkan oleh kesalahan tanggal atau data yang tidak valid. Vendor ID 1 dan 2 mengalami peningkatan besar dalam jumlah transaksi pada November 2024, sementara VendorID 6 mengalami penurunan yang signifikan dalam jumlah transaksi. Namun, jumlah transaksi turun drastis hingga nol pada Desember 2024 setelah mencapai puncaknya di bulan November. Pola ini menunjukkan bahwa tren perjalanan tidak konsisten, yang memerlukan penelitian lebih lanjut.

Perubahan dalam kebijakan operasional, peningkatan permintaan perjalanan karena musim liburan, atau kemungkinan masalah dengan sistem pencatatan data dapat menjadi beberapa penyebab fenomena ini. Ada kesalahan input, entitas bisnis baru, atau perubahan dalam sistem pengelolaan vendor dapat ditunjukkan oleh kehadiran VendorID 6 dengan jumlah transaksi yang lebih rendah daripada dua vendor utama. Oleh karena itu, untuk mendapatkan pemahaman yang lebih baik tentang kinerja vendor taksi di New York City, analisis lebih lanjut diperlukan untuk memastikan keakuratan timestamp, validitas VendorID, dan elemen luar yang dapat mempengaruhi tren perjalanan ini.

3.2 Tren Tarif Rata-rata Taksi NYC



Gambar 3. Tren tarif rata-rata per bulan berdasarkan vendor ID pada layanan taksi NYC

Gambar 3 menunjukkan tren tarif rata-rata per bulan berdasarkan Vendor ID untuk layanan taksi di NYC. Terdapat beberapa anomali dalam data, terutama pada skala waktu yang mencakup tahun 2002 dan 2008, yang kemungkinan besar merupakan kesalahan dalam pencatatan tanggal. Selain itu, terdapat lonjakan tarif yang signifikan pada Desember 2008, dengan rata-rata lebih dari 100 USD, jauh lebih tinggi dibandingkan bulan lainnya, sehingga perlu diperiksa apakah ada kesalahan dalam data atau faktor eksternal yang menyebabkan lonjakan ini. Sementara itu, tren tarif pada tahun 2024 menunjukkan pola yang lebih stabil dengan sedikit penurunan dari Oktober hingga Desember. Vendor dengan ID 1 dan 2 memiliki tren serupa, sedangkan Vendor 6 hanya muncul pada bulan November dengan tarif yang lebih rendah dibandingkan vendor lainnya. Hal ini mengindikasikan perlunya pembersihan dan validasi data sebelum menarik kesimpulan lebih lanjut mengenai pola tarif taksi di NYC.

3.3 Pola Perjalanan Taksi Berdasarkan Jarak Tempuh Rata-rata



Gambar 4. Pola perjalanan jarak tempuh rata-rata per bulan berdasarkan vendor ID

Gambar 4 menunjukkan pola perjalanan taksi berdasarkan jarak tempuh rata-rata (dalam mil) per bulan untuk berbagai Vendor ID. Terdapat lonjakan signifikan pada salah satu bulan di sekitar tahun 2008, di mana jarak tempuh rata-rata mencapai lebih dari 20 mil, yang kemudian diikuti oleh penurunan drastis pada bulan berikutnya. Setelah penurunan ini, jarak tempuh rata-rata tampak lebih stabil dengan fluktuasi kecil. Selain itu, terdapat satu titik data yang tampaknya berasal dari Vendor ID 6, yang memiliki nilai tertinggi dibandingkan data lainnya di sekitar bulan November 2024. Fluktuasi yang signifikan ini bisa disebabkan oleh perubahan dalam pola perjalanan, perubahan kebijakan tarif, atau faktor eksternal seperti kondisi ekonomi dan regulasi transportasi.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan dari pengolahan hingga visualisasi data yang sudah dilakukan, didapatkan kesimpulan bahwa:

- Tren perjalanan taksi berdasarkan jumlah transaksi per bulan untuk tiap vendor tidak menghasilkan suatu pola tertentu yang menandakan bahwa tren ini tidak konsisten. Vendor 2 menunjukkan lonjakan tren yang signifikan pada November 2024, namun kembali mengalami penurunan setelahnya, sementara untuk Vendor 1 dan 6 tidak menunjukkan pola.
- Tren tarif rata-rata taksi di NYC menunjukkan pola yang konstan pada harga kurang lebih 20 USD pada akhir tahun 2024 untuk Vendor 2, namun sebelumnya Vendor 2 mengalami lonjakan tarif pada Desember 2008 hingga angka 100 USD. Sementara untuk vendor lain, tidak menunjukkan pola yang menentu.
- Pola perjalanan layanan taksi Vendor 1 memiliki jarak perjalanan paling dekat pada nilai 2.5 mil, Vendor 2 di angka 5 mil dan Vendor 6 dengan

rata-rata jarak terjauh 10 mil. Vendor 2 mengalami lonjakan pola pada Desember 2008 di angka 20 mil.

- Vendor 2 memiliki produktivitas paling baik yang tercatat di dataset NYC TLC berdasarkan jumlah transaksi dan jarak tempuh rata-ratanya per bulan.

DAFTAR PUSTAKA

Putri, N. I. P. (2022, Juni 1). Pengaruh transportasi umum terhadap perkembangan

antar wilayah administrasi di Jakarta. Universitas Islam Negeri Sunan

Ampel. Retrieved from <https://journal.itltrisakti.ac.id/index.php/jmtbt>

Adam, A. (2022, Agustus 22). *Data Preprocessing: Pengertian, Manfaat, dan*

Tahapan Kerjanya. Accurate Online. Retrieved February 28, 2025, from

<https://accurate.id/teknologi/data-preprocessing/>

dibimbing.id - 4 Tahap Preprocessing Data, Beserta Penjelasan & Studi Kasus.

(2024, February 24). Dibimbing.id. Retrieved February 28, 2025, from

[https://dibimbing.id/blog/detail/mengenal-apa-itu-tahap-preprocessing-dat](https://dibimbing.id/blog/detail/mengenal-apa-itu-tahap-preprocessing-data)

a

Kimball, R., & Caserta, J. (2011). *The Data Warehouse ETL Toolkit: Practical*

Techniques for Extracting, Cleaning, Conforming, and Delivering Data.

Wiley.

Trivusi. (2022, September 16). *Pengertian dan Teknik Data Preprocessing dalam*

Data Mining. Trivusi. Retrieved February 28, 2025, from

<https://www.trivusi.web.id/2022/09/data-preprocessing.html>