



## Optimalisasi Waktu Penjemputan Dan Lokasi Pada Data Histori Perjalanan NYC TLC Menggunakan Exploratory Data Analysis

Andy Hermawan<sup>1</sup>, Antonius Andriyanto<sup>2</sup>, Ryandri Alif Pratomo Putra<sup>3</sup>,  
William Armand Rahardjo<sup>4</sup>, Yogga Prastya Wijaya<sup>5</sup>

<sup>1</sup>Universitas Indraprasta PGRI Jakarta

<sup>2-5</sup>Purwadhika Digital School Jakarta

Email: [andy.hermawan@unindra.ac.id](mailto:andy.hermawan@unindra.ac.id)<sup>1</sup>, [antoniusyanuar99@gmail.com](mailto:antoniusyanuar99@gmail.com)<sup>2</sup>, [ryandrialif43@gmail.com](mailto:ryandrialif43@gmail.com)<sup>3</sup>,  
[williamarmandrahardjo@gmail.com](mailto:williamarmandrahardjo@gmail.com)<sup>4</sup>, [yoggapw@gmail.com](mailto:yoggapw@gmail.com)<sup>5</sup>

**Abstract.** This study analyzes the "NYC TLC Trip Record" dataset for the period January 1, 2023 to January 31, 2023 to understand taxi usage patterns in New York City. The objectives to be achieved in this analysis include: (1) Identify the days and times with the highest demand for taxi services, (2) Identify the boroughs with the highest demand for taxi services. We applied univariate analysis for this analysis. The results show that the day with the highest demand occurs on Tuesday for the densest time occurs in the vulnerable time of 3 pm to 6 pm. The boroughs with the highest taxi demand are Manhattan, Queens, and Brooklyn. This analysis provides the results for NYC TLC to develop a data-driven optimization strategy. This analysis not only helps in identifying demand hotspots but also provides insights for more efficient taxi scheduling and placement. With this analysis, it is expected that more effective pick-up time and location optimization strategies can be developed, thereby improving operational efficiency and customer satisfaction in taxi services in New York City.

**Keywords:** univariate analysis, NYC TLC, taxi usage patterns

**Abstrak.** Penelitian ini menganalisis dataset "NYC TLC Trip Record" untuk periode 1 Januari 2023 hingga 31 Januari 2023 guna memahami pola penggunaan taksi di New York City. Tujuan yang ingin dicapai pada analisis ini antara lain: (1) Mengidentifikasi hari dan waktu dengan permintaan tertinggi layanan taksi, (2) Mengetahui borough dengan permintaan tertinggi layanan taksi. Analisis ini kami menerapkan analisis univariat. Hasil menunjukkan bahwa hari dengan permintaan tertinggi terjadi di hari Selasa untuk waktu terpadatnya terjadi pada rentan waktu 3 sore hingga 6 sore. Untuk *borough* dengan permintaan taksi terbanyak yaitu Manhattan, Queens, dan Brooklyn. Analisa ini memberikan hasil analisis untuk NYC TLC mengembangkan strategi optimalisasi yang berbasis data. Analisis ini tidak hanya membantu dalam mengidentifikasi hotspot permintaan tetapi juga memberikan wawasan untuk penjadwalan dan penempatan taksi yang lebih efisien. Dengan analisis ini, diharapkan dapat dikembangkan strategi optimalisasi waktu dan lokasi penjemputan yang lebih efektif, sehingga dapat meningkatkan efisiensi operasional dan kepuasan pelanggan dalam layanan taksi di New York City.

**Kata kunci:** analisis univariat, NYC TLC, pola penggunaan taksi

### PENDAHULUAN

Layanan taksi di New York City, yang diatur oleh Taxi and Limousine Commission (TLC), merupakan bagian integral dari sistem transportasi kota. Dengan populasi dan jumlah wisatawan yang besar, permintaan akan layanan taksi sangat tinggi dan berfluktuasi sepanjang hari. Mengelola operasi taksi secara efektif adalah tantangan besar, terutama dalam hal mengoptimalkan waktu dan lokasi penjemputan untuk mengurangi waktu tunggu penumpang dan meningkatkan efisiensi armada.

Seiring dengan perkembangan teknologi dan meningkatnya ketersediaan data besar (*big data*), analisis data historis perjalanan taksi menjadi mungkin dan sangat bermanfaat. Data perjalanan yang direkam oleh TLC mencakup informasi detail tentang waktu dan lokasi

penjemputan serta pengantaran. Dengan menerapkan *Exploratory Data Analysis* (EDA) pada data ini, kita dapat mengeksplorasi pola-pola permintaan taksi dan memahami dinamika operasionalnya secara lebih mendalam.

Penelitian terkait telah menunjukkan bahwa analisis data dapat membantu dalam mengidentifikasi pola permintaan transportasi. Studi sebelumnya telah menggunakan berbagai metode statistik dan pembelajaran mesin untuk memprediksi permintaan taksi dan mengoptimalkan rute. Namun, kebanyakan penelitian tersebut berfokus pada prediksi jangka pendek tanpa mempertimbangkan analisis mendalam mengenai pola waktu dan lokasi yang dapat memberikan wawasan untuk strategi penempatan taksi yang lebih efektif.

Kebaruan penelitian ini terletak pada pendekatannya yang komprehensif dalam menggunakan EDA untuk mengidentifikasi dan menganalisis pola spatio-temporal permintaan taksi di New York City. Dengan memahami distribusi penjemputan berdasarkan waktu (misalnya, jam, hari, dan musim) dan lokasi, penelitian ini bertujuan untuk mengembangkan strategi optimalisasi yang berbasis data. Analisis ini tidak hanya membantu dalam mengidentifikasi hotspot permintaan tetapi juga memberikan wawasan untuk penjadwalan dan penempatan taksi yang lebih efisien.

Urgensi penelitian ini semakin meningkat dengan semakin padatnya lalu lintas di New York City dan meningkatnya ekspektasi konsumen terhadap layanan taksi yang cepat dan andal. Optimalisasi operasi taksi tidak hanya akan menguntungkan operator taksi dalam hal pendapatan dan pemanfaatan sumber daya, tetapi juga meningkatkan pengalaman pelanggan dengan mengurangi waktu tunggu dan memperbaiki layanan.

Tujuan utama dari penelitian ini adalah untuk menerapkan EDA pada data histori perjalanan NYC TLC untuk mengungkap pola-pola penting dalam permintaan taksi. Dengan analisis ini, diharapkan dapat dikembangkan strategi optimalisasi waktu dan lokasi penjemputan yang lebih efektif, sehingga dapat meningkatkan efisiensi operasional dan kepuasan pelanggan dalam layanan taksi di New York City.

## **KAJIAN TEORITIS**

### **Grafik Garis**

Grafik garis yaitu grafik data berupa garis, yang diperoleh dari beberapa ruas garis dimana garis-garis tersebut menghubungkan titik-titik data. Grafik garis ini dapat digunakan apabila ingin memvisualisasikan perubahan dalam data seiring waktu atau urutan. Grafik garis sangat berguna untuk melihat tren, fluktuasi, serta perbandingan dalam data yang berbentuk *time-series*.

## **Diagram Batang**

Diagram batang adalah salah satu data visualisasi paling umum. Diagram batang adalah representasi grafis dari distribusi frekuensi yang menggunakan garis-garis vertikal atau horizontal untuk menggambarkan frekuensi masing-masing kelas. Diagram ini digunakan untuk membandingkan nilai-nilai kategori atau variabel yang berbeda. Terdapat dua jenis utama: diagram batang horizontal (*bar chart*) dan diagram batang vertikal (*column chart*).

## ***Heatmap***

*Heatmap* merupakan visualisasi data dua dimensi yang menggunakan skala warna untuk menunjukkan nilai dalam bentuk *grid*. Visualisasi melalui *heatmap* sering digunakan dalam menggambarkan korelasi antara dua variabel atau untuk melihat pola dalam data kompleks lebih terperinci.

## **METODE PENELITIAN**

Penelitian ini menggunakan pendekatan analisis univariat. Analisis univariat digunakan pada penelitian deskriptif dimana penelitian deskriptif memusatkan perhatian kepada pemecahan masalah-masalah aktual sebagaimana adanya pada saat penelitian dilaksanakan. Notoatmodjo (2018) menjelaskan bahwa analisis univariat bertujuan untuk menjelaskan atau mendeskripsikan karakteristik setiap variabel penelitian. Analisis univariat dilakukan menurut jenis data baik kategorik maupun numerik.

## HASIL DAN PEMBAHASAN

### Data Understanding

Berikut adalah penjelasan mengenai fitur-fitur yang terdapat dalam dataset ini. Dataset ini mencatat total 68.211 perjalanan, yang direpresentasikan dalam 20 kolom.

**Tabel 1. Penjelasan Kolom pada Dataset**

### New York City TLC Trip Record Data Dictionary

Feature	Description
VendorID	A code indicating the LPEP provider that provided the record.  1 = Creative Mobile Technologies, LLC. 2 = VeriFone Inc.
lpep_pickup_datetime	The date and time when the meter was engaged.
lpep_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value.
Trip_distance	The elapsed trip distance in miles was reported by the taximeter.
PULocationID	TLC Taxi Zone in which the taximeter was engaged.
DOLocationID	TLC Taxi Zone in which the taximeter was disengaged.
RateCodeID	The final rate code is in effect at the end of the trip.  1 = Standard rate 2 =JFK 3 =Newark 4 =Nassau or Westchester 5 =Negotiated fare 6 =Group ride
Store_and_fwd_flag	This flag indicates whether the trip record was held in the vehicle memory before sending to the vendor, aka "store and forward," because the vehicle did not have a connection to the server.  Y = store and forward trip N = not a store and forward trip

Payment_type	A numeric code signifying how the passenger paid for the trip.  1 = Credit card 2 = Cash 3 = No charge 4 = Dispute 5 = Unknown 6 = Voided trip
Fare_amount	The time-and-distance fare is calculated by the meter. Extra Miscellaneous extras and surcharges. Currently, this only includes the \$0.50 and \$1 rush hour and overnight charges.
MTA_tax	\$0.50 MTA tax that is automatically triggered based on the metered rate in use.
Improvement_surcharge	\$0.30 improvement surcharge assessed on hailed trips at the flag drop. The improvement surcharge began being levied in 2015.
Tip_amount	This field is automatically populated for credit card tips. Cash tips are not included.
Tolls_amount	The total amount of all tolls paid in the trip.
Total_amount	The total amount charged to passengers. Does not include cash tips.
Trip_type	A code indicating whether the trip was a street hail or a dispatch that is automatically assigned based on the metered rate in use but can be altered by the driver.  1 = Street-hail 2 = Dispatch

## Data Cleaning

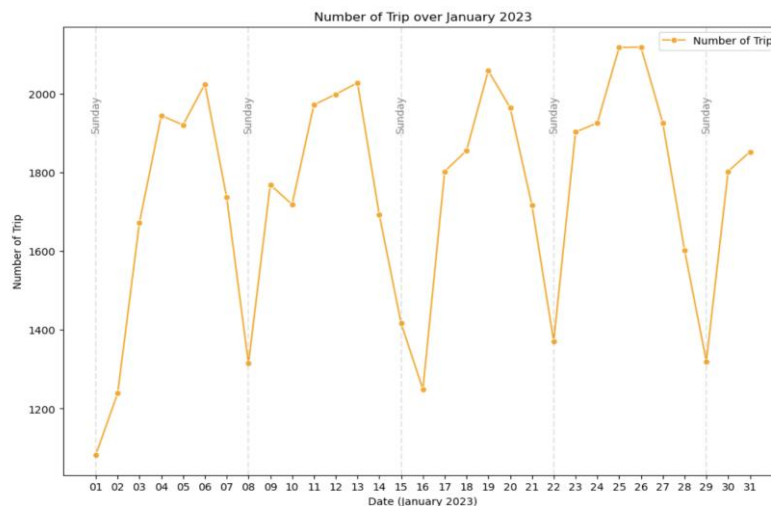
Tahapan pembersihan data yang dilakukan adalah sebagai berikut:

1. Penanganan Data yang Hilang: Tahap ini mencakup identifikasi dan penanganan nilai yang hilang dalam dataset. Hal ini dapat dilakukan dengan menggantikan nilai yang hilang dengan nilai rata-rata atau median dari kolom yang bersangkutan, atau dengan menggunakan metode lain seperti interpolasi.
2. Batasan Tipe Data: Tahap ini mencakup pemeriksaan dan perubahan tipe data untuk memastikan bahwa setiap variabel memiliki tipe data yang sesuai. Misalnya, memastikan variabel yang seharusnya bertipe numerik tidak memiliki tipe data string.

3. Penanganan Data yang Duplikat: Pada tahap ini, data yang memiliki nilai yang sama untuk setiap atributnya diidentifikasi dan dihapus dari dataset. Hal ini dilakukan untuk mencegah pengaruh data yang duplikat pada analisis.
4. Rentang Data yang Valid: Tahap ini melibatkan pemeriksaan apakah setiap nilai dalam dataset berada dalam rentang yang valid. Misalnya, memastikan bahwa umur seseorang tidak negatif atau bahwa tanggal lahir seseorang dalam format yang benar.
5. Validasi Silang Kolom: Tahap ini melibatkan pemeriksaan hubungan antara kolom-kolom dalam dataset untuk memastikan bahwa data yang ada konsisten dan valid. Misalnya, memastikan bahwa nilai total pengeluaran tidak melebihi nilai pendapatan yang tercatat.
6. Penanganan Pencilan: Tahap ini mencakup identifikasi dan penanganan nilai yang berbeda secara signifikan dari nilai-nilai lain dalam dataset. Hal ini dilakukan untuk mencegah nilai yang ekstrim ini mengganggu analisis yang dilakukan.

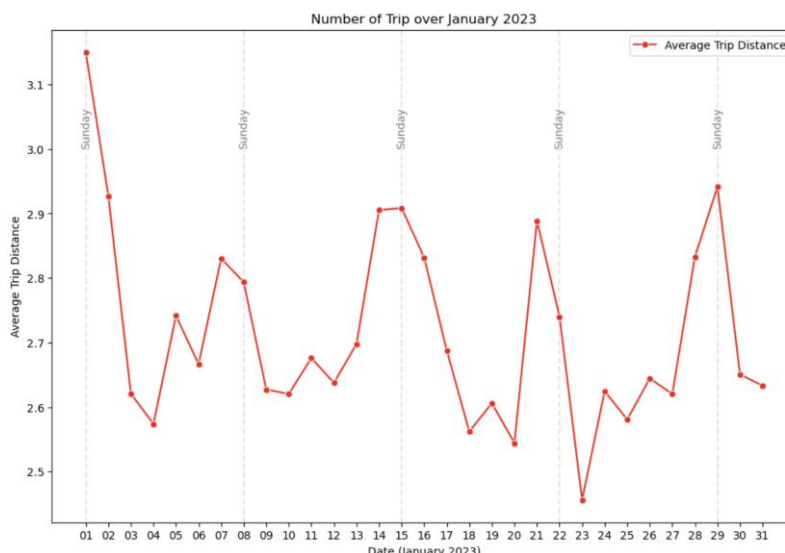
### Exploratory Data Analysis

Dalam bagian ini, kami akan melakukan analisis pada data yang telah melalui proses pembersihan untuk mengidentifikasi informasi atau pola penting. Kami akan menggunakan statistik deskriptif dan visualisasi grafis untuk mendapatkan pemahaman yang lebih baik tentang data.



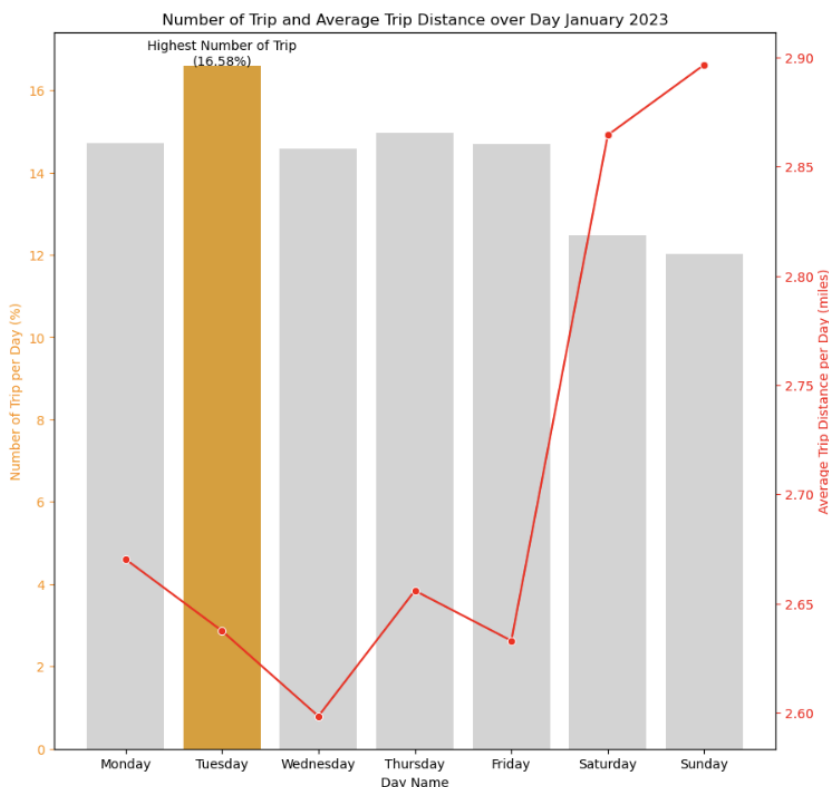
**Gambar 1. Grafik Jumlah Perjalanan Harian**

Grafik di atas menggambarkan tren jumlah perjalanan taksi selama Januari 2023, menunjukkan bahwa jumlahnya cenderung lebih tinggi selama hari kerja dibandingkan dengan akhir pekan. Pada tanggal 1 Januari, jumlah perjalanan terendah tercatat sebanyak 1.084, sementara pada tanggal 11 Januari mencatat jumlah perjalanan tertinggi sebanyak 1.974.



**Gambar 2. Grafik Rata - Rata Jarak Perjalanan Harian**

Grafik tersebut menampilkan tren rata-rata jarak perjalanan taksi selama Januari 2023, yang menunjukkan bahwa rata-rata jarak perjalanan cenderung lebih tinggi selama akhir pekan dibandingkan dengan hari kerja. Pada tanggal 1 Januari, rata-rata jarak perjalanan terendah mencapai 3,29 mil, sementara pada tanggal 11 Januari mencapai puncak tertinggi dengan rata-rata 2,79 mil.



**Gambar 3. Grafik Rata - Rata Jarak Perjalanan dan Jumlah Perjalanan Per Jenis Hari**

Grafik tersebut mengungkap pola kebiasaan penumpang berdasarkan hari dalam seminggu. Hari Selasa menonjol sebagai hari dengan jumlah proporsi perjalanan terbanyak dibandingkan hari-hari lainnya yaitu sebanyak 8974 perjalanan (16,58%), sedangkan Hari Minggu menonjol sebagai hari dengan rata-rata perjalanan terjauh yaitu sejauh 2,89 mil.



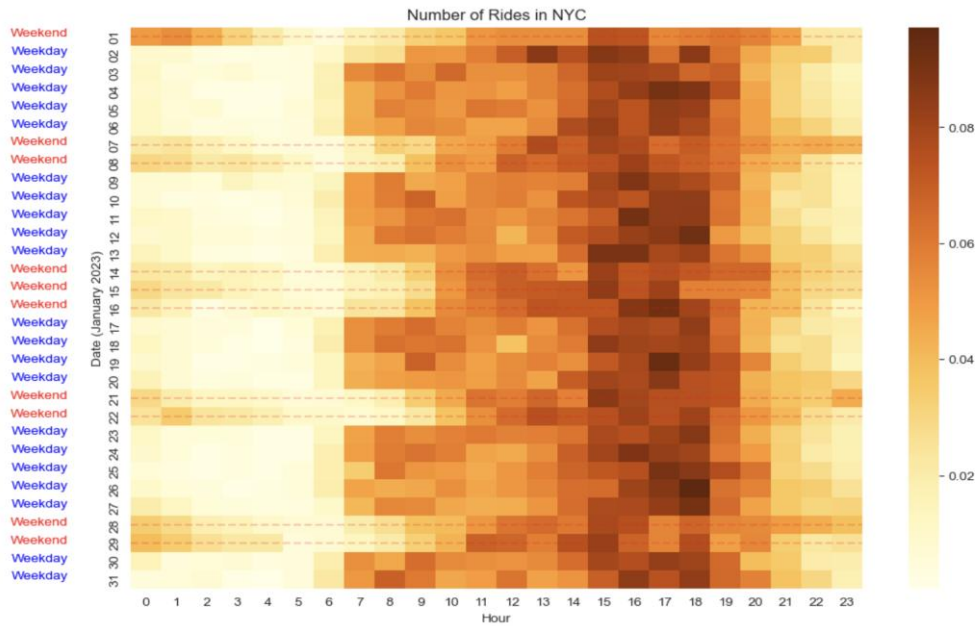
**Gambar 4. Grafik Jumlah Perjalanan Per Jam**

Grafik tersebut menggambarkan pola perjalanan berdasarkan jam dalam sehari, dibedakan antara hari kerja dan hari libur. Pada hari kerja, terlihat lonjakan signifikan pada jumlah perjalanan mulai pukul 08.00 yang mencapai puncaknya pada pukul 18.00. Sementara itu, pada hari libur, lonjakan jumlah perjalanan berada pada pukul 15.00 namun tidak sebesar lonjakan pada hari kerja.



**Gambar 5. Grafik Rata - Rata Jarak Perjalanan Per Jam**

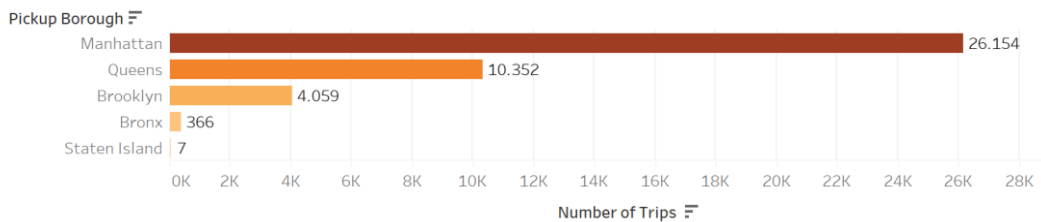
Grafik tersebut memperlihatkan pola rata-rata jarak perjalanan berdasarkan jam dalam sehari, dengan perbedaan antara hari kerja dan hari libur. Pada keduanya, terlihat lonjakan yang signifikan pada rata-rata jarak perjalanan mulai pukul 05.00, namun lonjakan ini lebih tinggi pada hari libur daripada pada hari kerja.



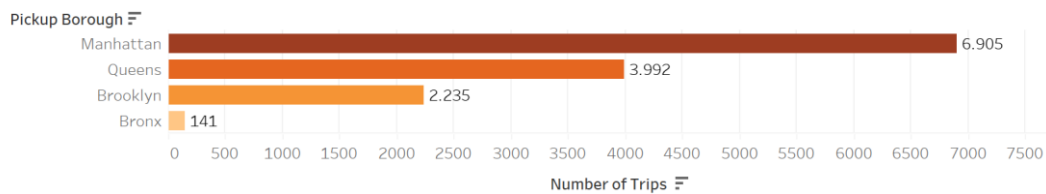
**Gambar 6. Grafik Kepadatan Jumlah Perjalanan Per Jam dan Per Hari**

Grafik tersebut menggambarkan pola kebiasaan penumpang berdasarkan hari dan jam. Pada hari kerja (Senin-Jumat), terjadi peningkatan jumlah perjalanan taksi mulai dari jam 7 pagi hingga pukul 20 malam. Sementara pada hari libur (Sabtu-Minggu), lonjakan jumlah perjalanan taksi terjadi mulai dari jam 10 pagi hingga tengah malam.

Frequency of Pickup Zone Per Borough - Weekdays

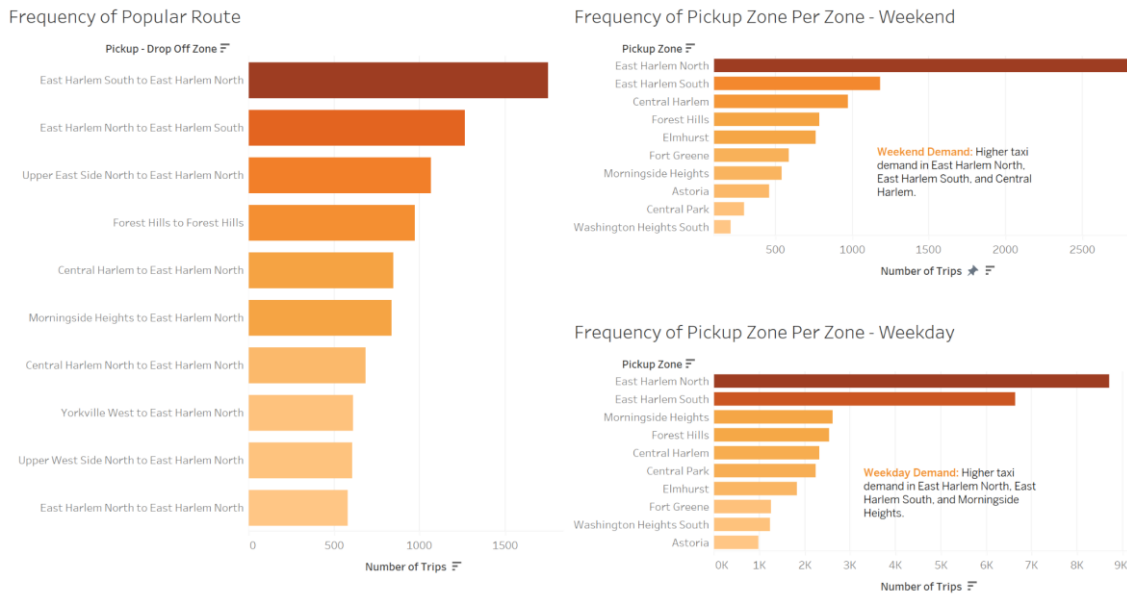


Frequency of Pickup Zone Per Borough - Weekend



**Gambar 7. Grafik Jumlah Perjalanan Per Wilayah**

Grafik tersebut menunjukkan frekuensi jumlah perjalanan per wilayah di New York City (NYC) pada hari kerja dan hari libur. Wilayah penjemputan tertinggi pada hari kerja adalah Manhattan, Queens, Brooklyn, Bronx, dan Staten Island secara berurutan. Sedangkan pada hari libur, wilayah penjemputan tertinggi adalah Manhattan, Queens, Brooklyn, dan Bronx secara berurutan.



**Gambar 8. Grafik Jumlah Perjalanan Per Zona**

Grafik tersebut menunjukkan frekuensi jumlah perjalanan per zona di New York City (NYC) pada hari kerja dan hari libur. Zona penjemputan tertinggi pada hari kerja adalah East Harlem North, East Harlem South, Central Harlem, Forest Hills, dan Elmhurst secara beruntun. Sedangkan pada hari libur, zona penjemputan tertinggi adalah East Harlem North, East Harlem South, Morningside Heights, Forest Hills, dan Central Harlem secara beruntun. Rute perjalanan terpopuler adalah perjalanan dari East Harlem South ke East Harlem North.

## KESIMPULAN DAN SARAN

Kesimpulan setelah dilakukan analisa pada dataset NYC TLC Trip Record adalah penumpang taksi NYC memiliki pola mingguan sepanjang 01 Januari 2023 hingga 31 Januari 2023 yaitu hari Selasa merupakan hari dengan jumlah pemesanan taksi terbanyak sedangkan untuk waktu terpadatnya berada di rentan waktu pukul 3 sore hingga pukul 6 sore. Selain itu untuk rata-rata bepergian terjauh terjadi pada puncaknya di hari Minggu di rentan waktu pukul 2 subuh hingga pukul 5 pagi. Pada hari kerja, pemesanan taksi cenderung lebih banyak pada pukul 6 pagi hingga pukul 8 malam dan pada hari libur, pemesanan taksi lebih banyak pada pukul 10 pagi hingga pukul 3 subuh.

Selain itu dari analisa juga didapatkan bahwa terdapat 3 borough dengan permintaan taksi terbanyak yaitu Manhattan, Queens, dan Brooklyn dimana untuk zona dengan pickup terbanyak pada hari libur berada di zona East Harlem North, East Harlem South, Central Harlem, Forest Hills, dan Elmhurst sedangkan zona dengan drop off terbanyak pada hari libur adalah Central Harlem, East Harlem North, Central Harlem North, Upper East Side North, dan East Harlem South.

Pada hari kerja, zona East Harlem North, East Harlem South, Morningside Heights, Forest Hills, dan Central Harlem adalah zona dengan pickup terbanyak, sedangkan zona East Harlem South, East Harlem North, Upper East Side North, Central Harlem, dan Upper West Side North adalah zona dengan drop off terbanyak. Zona East Harlem North menjadi lokasi pickup terpadat pada 10 rute perjalanan terpopuler baik pada hari kerja maupun akhir pekan. Secara proporsi, penumpang lebih cenderung melakukan perjalanan lebih jauh pada hari libur dibandingkan hari kerja. Hasil analisis ini memberikan wawasan penting bagi NYC TLC dalam merencanakan dan mengelola layanan taksi di New York City.

Dari kesimpulan yang didapatkan maka rekomendasi yang dapat diberikan adalah mengoptimalkan layanan pada hari Selasa dengan menambah jumlah armada atau mengatur penjadwalan yang lebih efisien, meningkatkan keamanan penumpang dan supir untuk perjalanan panjang di malam hari dengan menambahkan alat pelacak di setiap taksi, meningkatkan layanan di borough dengan permintaan yang tinggi seperti di Brooklyn, Manhattan, Queens dan zona East Harlem North dikarenakan zona tersebut merupakan titik pickup terbanyak, meningkatkan layanan di jam tertentu seperti saat peak hour untuk permintaan yang tinggi terhadap pelayanan taksi.

Untuk keterbatasan penelitian pada analisis ini adalah dataset NYC TLC yang digunakan hanya mencakup satu bulan sehingga terdapat keterbatasan waktu sehingga pola yang ditemukan mungkin tidak sepenuhnya mewakili tren jangka panjang yang bisa terjadi sepanjang tahun. Oleh karena itu rekomendasi yang dapat diberikan untuk penelitian yang akan datang adalah menggunakan data dengan kurun waktu setahun sehingga dapat mewakili tren sepanjang tahun.

## DAFTAR REFERENSI

- Ascarya Solution. (2023, March 11). Analisis Univariat dan Bivariat, Cara Olah Datanya. Ascarya Solution. Retrieved June 20, 2024, from <https://ascarya.or.id/analisis-univariat-dan-bivariat/>
- Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), 131.
- Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python* (2nd ed.). California: O'Reilly Media.
- Cahyaningrum, W., & Gunawan. (2023). Hubungan Antara Intensitas Penggunaan Gawai dengan Perkembangan Literasi Awal Pada Anak Prasekolah Usia 4-6 Tahun di TK A isyiyah Kedungharjo Mantingan. *Jurnal Terapi Wicara dan Bahasa*, Volume 1 Nomor 2 Tahun 2023, 228-239. Retrieved from <https://jtwb.org/index.php/jtwb/article/view/60>

- City of New York. (n.d.). NYC Taxi Zones. Retrieved June 20, 2024, from <https://data.cityofnewyork.us/Transportation/NYC-Taxi-Zones/d3c5-ddgc>
- City of New York. (n.d.). Taxi & Limousine Commission. Retrieved June 20, 2024, from <https://www.nyc.gov/site/tlc/index.page>
- Gu, Z. (2022). Complex heatmap visualization. *Imeta*, 1(3), e43.
- Komorowski, M., Marshall, D. C., Saliccioli, J. D., & Crutain, Y. (September 2016). Exploratory Data Analysis. In M. Komorowski, D. C. Marshall, J. D. Saliccioli, & Y. Crutain (Eds.), *Secondary Analysis of Electronic Health Records* (pp. 185-203). doi:10.1007/978-3-319-43742-2\_15. License: CC BY-NC 4.0. Imperial College London.
- Kosmos, P. (2024, January 15). Visualisasi Data: Pengertian, Fungsi, dan contohnya. Telkom University. Retrieved June 20, 2024, from <https://dac.telkomuniversity.ac.id/visualisasi-data-pengertian-fungsi-dan-contohnya/>
- Lembaga Layanan Pendidikan Tinggi Wilayah V Yogyakarta. (2021, August 23). 9 Jenis Chart Yang Dapat Anda Gunakan Untuk Visualisasi Data Dalam Presentasi Anda. Artikel. Retrieved June 20, 2024, from <https://lldikti5.kemdikbud.go.id/home/detailpost/9-jenis-chart-yang-dapat-anda-gunakan-untuk-visualisasi-data-dalam-presentasi-anda/>
- Lizana, H. I., & Ridho, F. (2021). Implementasi dan Evaluasi Visualisasi Data Interaktif pada Publikasi Laporan Bulanan Data Sosial Ekonomi Indonesia. *Seminar Nasional Official Statistics*, 2021(1), 947-957. doi:10.34123/semnasoffstat.v2021i1.795
- Mujilawati, S. (2021). VISUALISASI DATA HASIL KLASIFIKASI NAÏVE BAYES DENGAN MATPLOTLIB PADA PYTHON. *Prosiding SNST ke-11 Tahun 2021*, Volume 1 No. 1(E. 31). Retrieved from [https://publikasiilmiah.unwahas.ac.id/PROSIDING\\_SNST\\_FT/article/view/5164](https://publikasiilmiah.unwahas.ac.id/PROSIDING_SNST_FT/article/view/5164)
- Nasution, L. M. (2017). Statistik deskriptif. *Hikmah*, 14(1), 49-55.
- Perer, A., & Shneiderman, B. (2008). Integrating Statistics and Visualization: Case Studies of Gaining Clarity during Exploratory Data Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 265-274). Florence, Italy: ACM.
- Rao, A. S., Vardhan, B. V., & Shaik, H. (2021). Role of Exploratory Data Analysis in Data Science. In *Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES 2021)* (pp. 1457-1461). doi: 10.1109/ICCES51350.2021.9488986.
- Sagala, N. T., & Aryatama, F. Y. (2022). Exploratory Data Analysis (EDA): A Study of Olympic Medallist. *Sistemasi: Jurnal Sistem Informasi*, 11(3), 578-587.
- Sahoo, K., Samal, A., Pramanik, J., & Pani, S. (2019). Exploratory Data Analysis using Python. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4727-4735. doi: 10.35940/ijitee.L3591.1081219.
- Soendari, T. (2012). *Metode Penelitian Deskriptif*. Bandung, UPI.

Tukey, J. W. (1977). *Exploratory Data Analysis (Vol. 2)*. Boston: Addison-Wesley.

Velleman, P. F., & Hoaglin, D. C. (2014). *Applications, Basics, and Computing of Exploratory Data Analysis*. Washington: The Internet-First University Press.

Waskom, M. (2022). *Seaborn: statistical data visualization*. Retrieved June 20, 2024, from <https://seaborn.pydata.org/>