

NYC TLC

New York City Taxi and Limousine Commission

Data Analysis

Athaya Zahrani Irmansyah

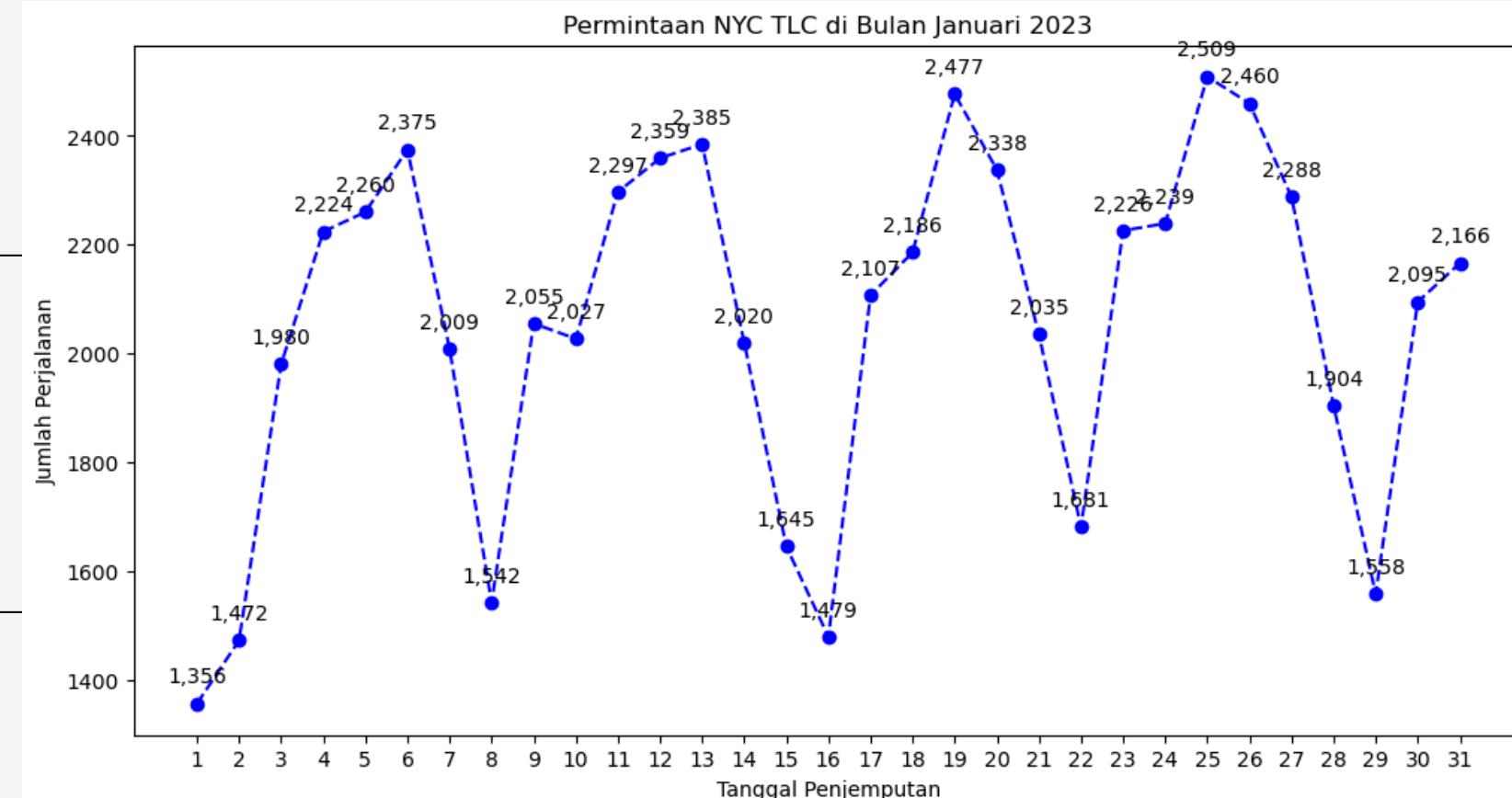
Thu, 8/8/2024



Executive Summary



Latar Belakang	Industri sewa kendaraan untuk mengatur dan memfasilitasi Taksi/Limosin untuk warga kota atau wisatawan di kota New York
Situasi	Layanan antar-jemput yang permintaannya sangat tinggi dan berfluktuasi sepanjang hari
Rumusan Masalah	Mengoptimalkan operasi dan meningkatkan profitabilitas dari dataset bulan Januari 2023
Questions	Bagaimana fluktuasi permintaan ? Bagaimana preferensi dan perilaku pelanggan ? Apa saja faktor-faktor yang mempengaruhi besaran tip ? Apa saja rekomendasi untuk meningkatkan pelayanan ?
Stakeholders	Pemerintah Kota New York Pengemudi Taksi dan Limosin Perusahaan Taksi dan Limosin Penumpang dan Masyarakat Umum



Data Summary



	kolom	data_type	null_value(%)	negative_value(%)	0_value(%)	duplicate	n_unique	sample_unique
0	VendorID	int64	0.000	0.000	0.000	0	2	[2, 1]
1	lpep_pickup_datetime	object	0.000	0.000	0.000	0	29448	[1/1/2023 0:26, 1/1/2023 0:51, 1/1/2023 0:35, ...]
2	lpep_dropoff_datetime	object	0.000	0.000	0.000	0	29624	[1/1/2023 0:37, 1/1/2023 0:57, 1/1/2023 0:41, ...]
3	store_and_fwd_flag	object	6.339	0.000	0.000	0	2	[N, Y, nan]
4	RatecodeID	float64	6.339	0.000	0.000	0	6	[1.0, 5.0, 4.0, 3.0, 2.0, 99.0, nan]
5	PULocationID	int64	0.000	0.000	0.000	0	226	[166, 24, 223, 41, 181, 255, 75, 66, 195, 83, ...]
6	DOLocationID	int64	0.000	0.000	0.000	0	249	[143, 43, 179, 238, 74, 262, 45, 75, 166, 140, ...]
7	passenger_count	float64	6.339	0.000	0.493	0	10	[1.0, 2.0, 4.0, 3.0, 6.0, 5.0, 0.0, 7.0, 9.0, ...]
8	trip_distance	float64	0.000	0.000	4.911	0	1870	[2.58, 1.81, 0.0, 1.3, 1.1, 2.78, 3.8, 1.88, 1...
9	fare_amount	float64	0.000	0.258	0.122	0	2553	[14.9, 10.7, 7.2, 6.5, 6.0, 17.7, 19.1, 14.2, ...]
10	extra	float64	0.000	0.117	59.703	0	16	[1.0, 0.5, 3.75, 0.0, 3.25, 6.0, 5.0, 2.75, 5....]
11	mta_tax	float64	0.000	0.242	8.471	0	6	[0.5, 1.5, 0.0, 1.0, -0.5, 2.75]
12	tip_amount	float64	0.000	0.013	42.251	0	1492	[4.03, 2.64, 1.94, 1.7, 0.0, 4.85, 1.0, 3.0, 3...
13	tolls_amount	float64	0.000	0.000	97.448	0	26	[0.0, 6.55, 4.0, 13.75, 9.0, 3.0, 11.75, 2.45, ...]
14	ehail_fee	float64	100.000	0.000	0.000	0	0	[nan]
15	improvement_surcharge	float64	0.000	0.255	0.273	0	5	[1.0, 0.0, 0.3, -1.0, -0.3]
16	total_amount	float64	0.000	0.262	0.101	0	4670	[24.18, 15.84, 11.64, 10.2, 8.0, 22.95, 29.2, ...]
17	payment_type	float64	6.339	0.000	0.000	0	5	[1.0, 2.0, 3.0, 4.0, 5.0, nan]
18	trip_type	float64	6.354	0.000	0.000	0	2	[1.0, 2.0, nan]
19	congestion_surcharge	float64	6.339	0.004	68.898	0	4	[2.75, 0.0, 2.5, -2.75, nan]

**68211 baris
20 kolom**

Supporting Data Summary



	kolom	data_type	null_value(%)	negative_value(%)	0_value(%)	duplicate	n_unique	sample_unique
0	OBJECTID	int64	0.0	0.0	0.0	0	263	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...]
1	Shape_Leng	float64	0.0	0.0	0.0	0	263	[0.116357453189, 0.43346966679, 0.084341105901...]
2	Shape_Area	float64	0.0	0.0	0.0	0	263	[0.0007823067885, 0.00486634037837, 0.00031441...]
3	zone	object	0.0	0.0	0.0	0	260	[Newark Airport, Jamaica Bay, Allerton/Pelham ...]
4	LocationID	int64	0.0	0.0	0.0	0	260	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...]
5	borough	object	0.0	0.0	0.0	0	6	[EWR, Queens, Bronx, Manhattan, Staten Island,...]
6	longitude	float64	0.0	0.0	0.0	0	263	[936681.7498053048, 1033536.0827942044, 102713...]
7	latitude	float64	0.0	0.0	0.0	0	263	[190522.13027768556, 161853.98229982704, 25473...]

no null, negative, 0, and duplicate values

Data Cleaning



Ubah tipe data *datetime*

```
1 lpep_pickup_datetime    68211 non-null   datetime64[ns]
2 lpep_dropoff_datetime   68211 non-null   datetime64[ns]
```

Hapus kolom ehail_fee

```
df.drop(columns=['ehail_fee'])
```

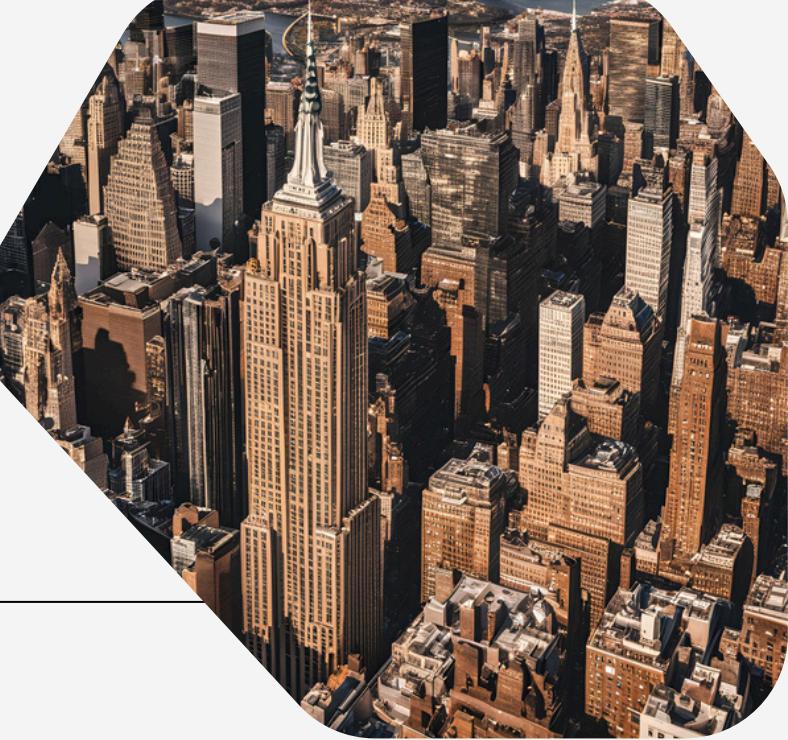
Zeros Value Handling

Kolom extra, mta_tax, tip_amount, tolls_amount, improvement_surcharge, dan congestion_surcharge **memungkinkan bernilai 0 -> Dibiarkan**

Kolom passenger_count, trip_distance, fare_amount, dan total_amount **tidak mungkin bernilai 0**

- Drop passenger_count = 0 dan trip_distance = 0
- Drop passenger_count = 0 dan total_amount = 0
- passenger_count = 0 -> isi dengan 1
- Kolom lainnya diseleksi dimana trip_distance, fare_amount, total_amount > 0

Missing Value Handling



Kolom store_and_fwd_flag

Isi menggunakan modusnya:

N = Bukan perjalanan *store and forward*

Kolom RatecodeID

Isi menggunakan mediannya:

1 = Tarif standar

Kolom passenger_count

Isi menggunakan mediannya:

1 = Jumlah penumpang sebanyak 1 orang, jumlah terbanyak, jumlah minimal

Kolom payment_type

tip_amount > 0 -> payment type = 1 = Cash

tip_amount = 0 -> payment_type = 5 = Pembayaran tidak diketahui

Kolom trip_type

Memiliki korelasi sangat kuat (0.916) dengan kolom RatecodeID (trip_type sangat dipengaruhi oleh RatecodeID)

RatecodeID 1.0 memiliki trip_type 1.0.

Missing value pada kolom trip_type yang memiliki nilai RatecodeID 1.0 bisa diisi 1.0.

Outliers Handling

- Tahun selain 2023 (2022 dan 2009)
- Bulan 12 (Desember) dan 2 (Februari)
- Ganti data RatecodeID kategori 99 dengan 1 (modusnya)
- Fare_amount di bawah \$3.00 (tetapan minimal)
- Jarak waktu yang bernilai < 1 menit (pembatalan trip)
- trip_distance < 0.2 miles dan fare_amount > \$3.00 (*initial charge* adalah sebesar \$3.00 dan akan bertambah sebesar \$0.7 setiap 0.2 miles)
- trip_distance > 180 miles (maksimal jarak per *shift*)
- Drop data pada kolom dengan outliers di bawah 5%



Mengubah Nilai Data



VendorID:

1 = 'Creative Mobile Technologies, LLC.'
2 = 'VeriFone Inc.'

store_and_fwd_flag:

Y = 'Store and Forward Trip'
N = 'Not Store and Forward Trip'

RatecodeID:

1 = 'Tarif Standar'
2 = 'JFK Airport'
3 = 'Newark'
4 = 'Nassau/Westchester'
5 = 'Biaya Negosiasi'
6 = 'Perjalanan Berkelompok'

trip_type:

1 = 'Street-hail'
2 = 'Dispatch'

payment_type:

1 = 'Kartu Kredit'
2 = 'Uang Tunai'
3 = 'Tanpa Biaya'
4 = 'Sengketa'
5 = 'Tidak Diketahui'
6 = 'Perjalanan yang Dibatalkan'

Menambah Kolom



10

- trip_duration (Durasi trip dalam menit)
- driving_speed (Kecepatan trip = Jarak / Waktu)
- trip_duration_range
- pickup_time_category
- day_category
- speed_range
- distance_range
- fare_amount_range
- tip_amount_range
- total_final

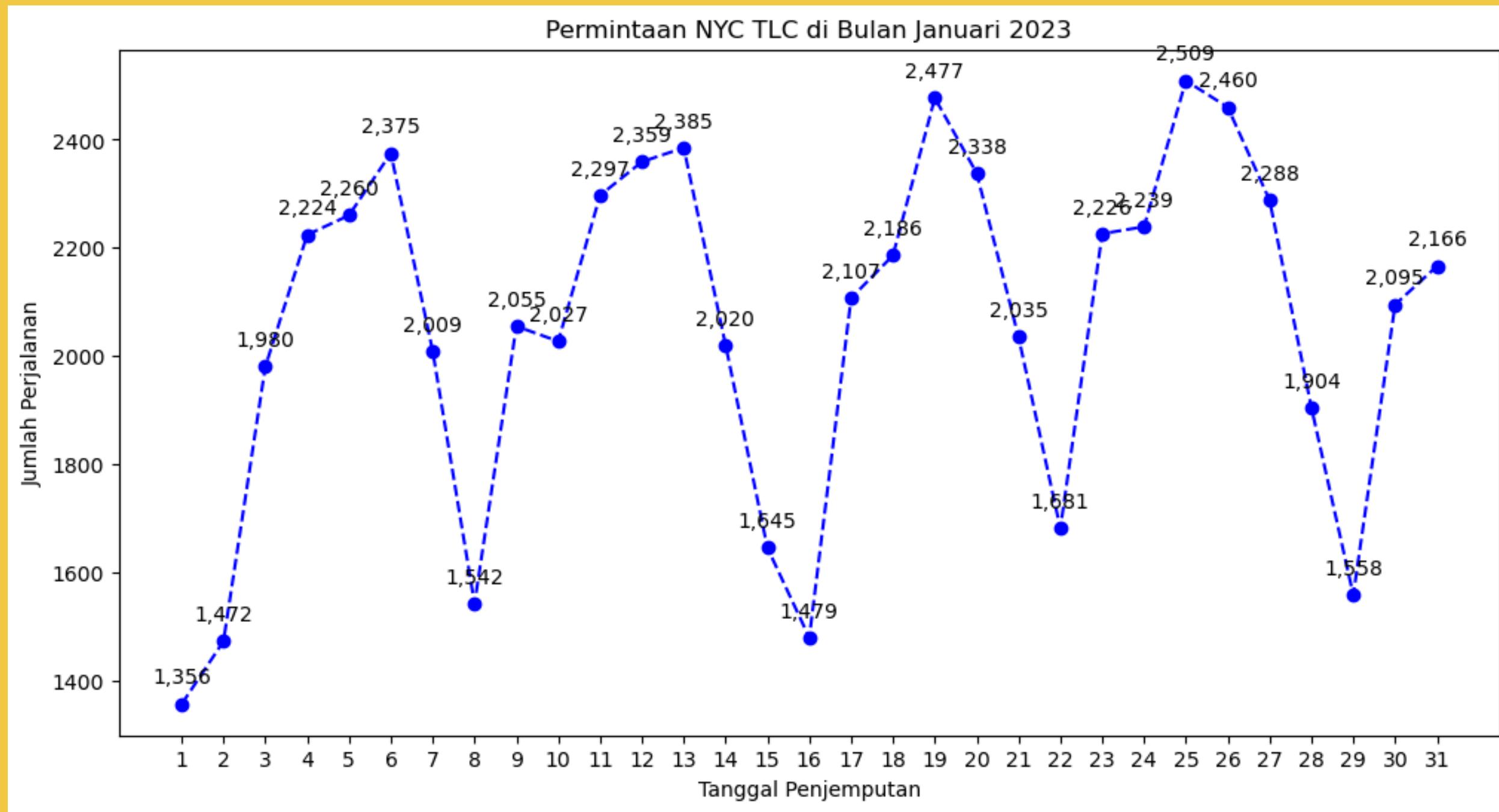
Data Analysis



**Kapan saja tanggal di bulan
Januari 2023 yang memiliki
permintaan tertinggi?**



Results

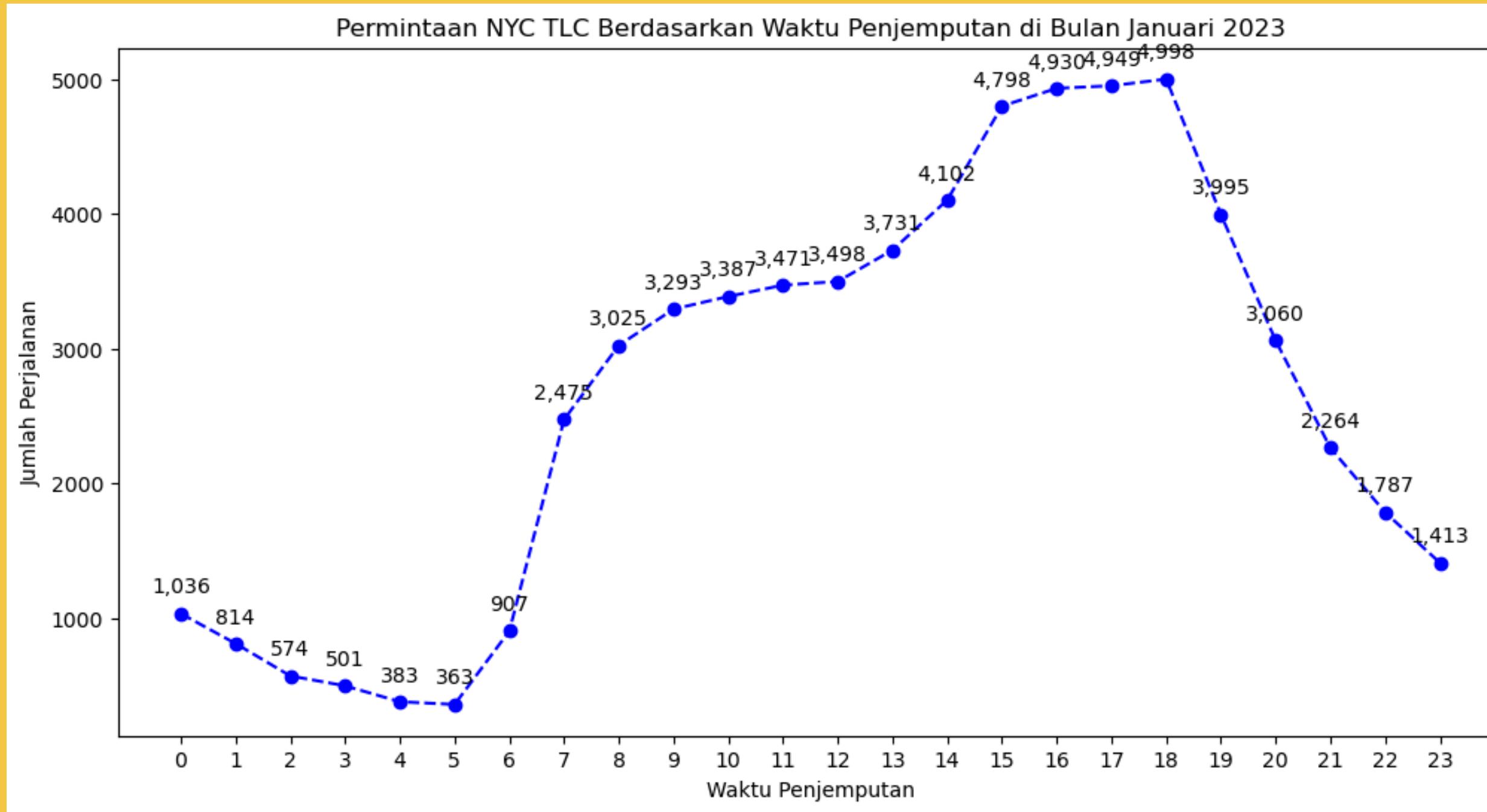


- **Tinggi:** hari terakhir weekday
- **Rendah:** Tahun baru dan hari libur
- Ada **pola berulang** setiap minggu

**Bagaimana variasi waktu
penjemputan (*pick up hour*)
yang memiliki permintaan
tertinggi?**

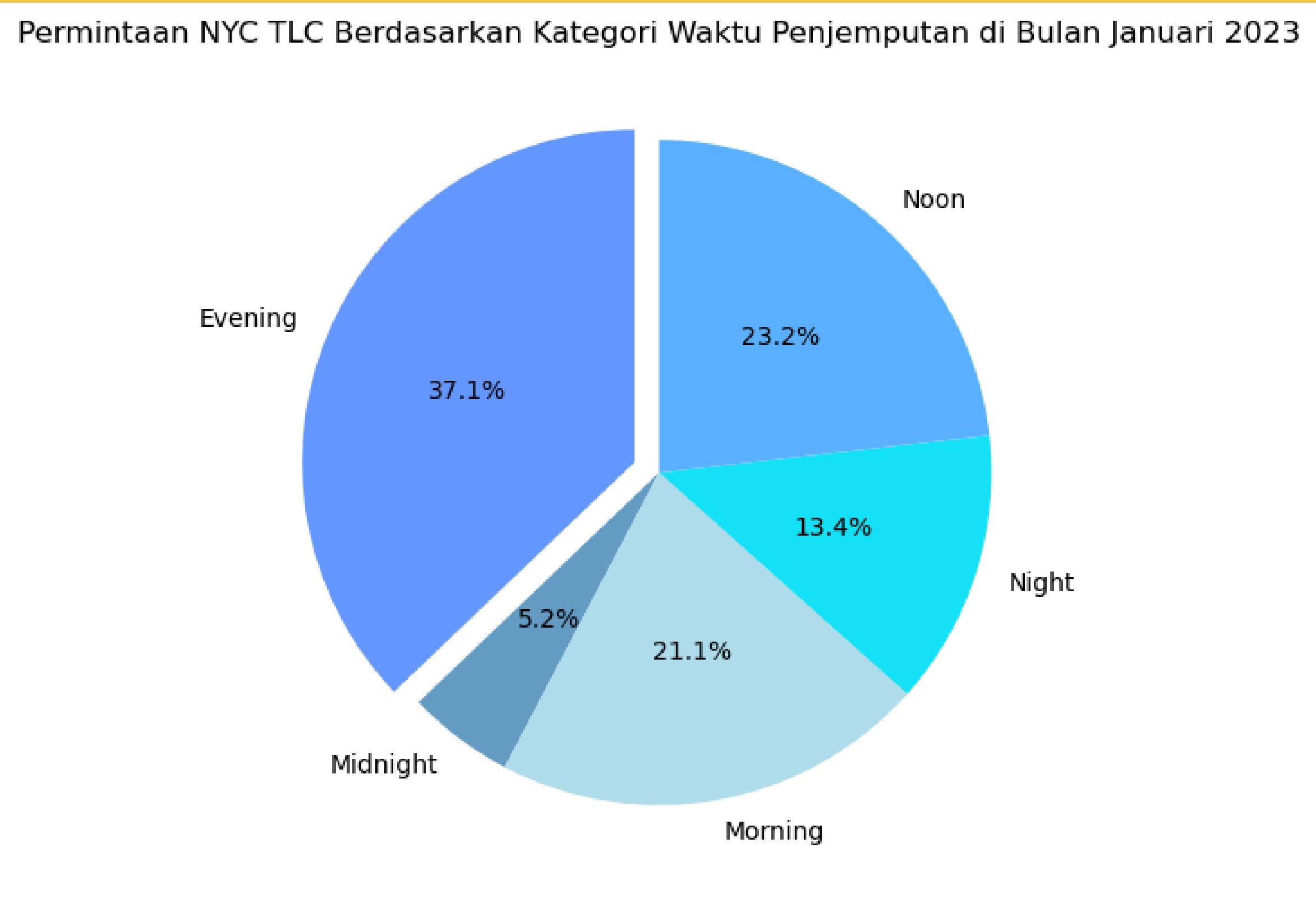


Berdasarkan waktu penjemputan



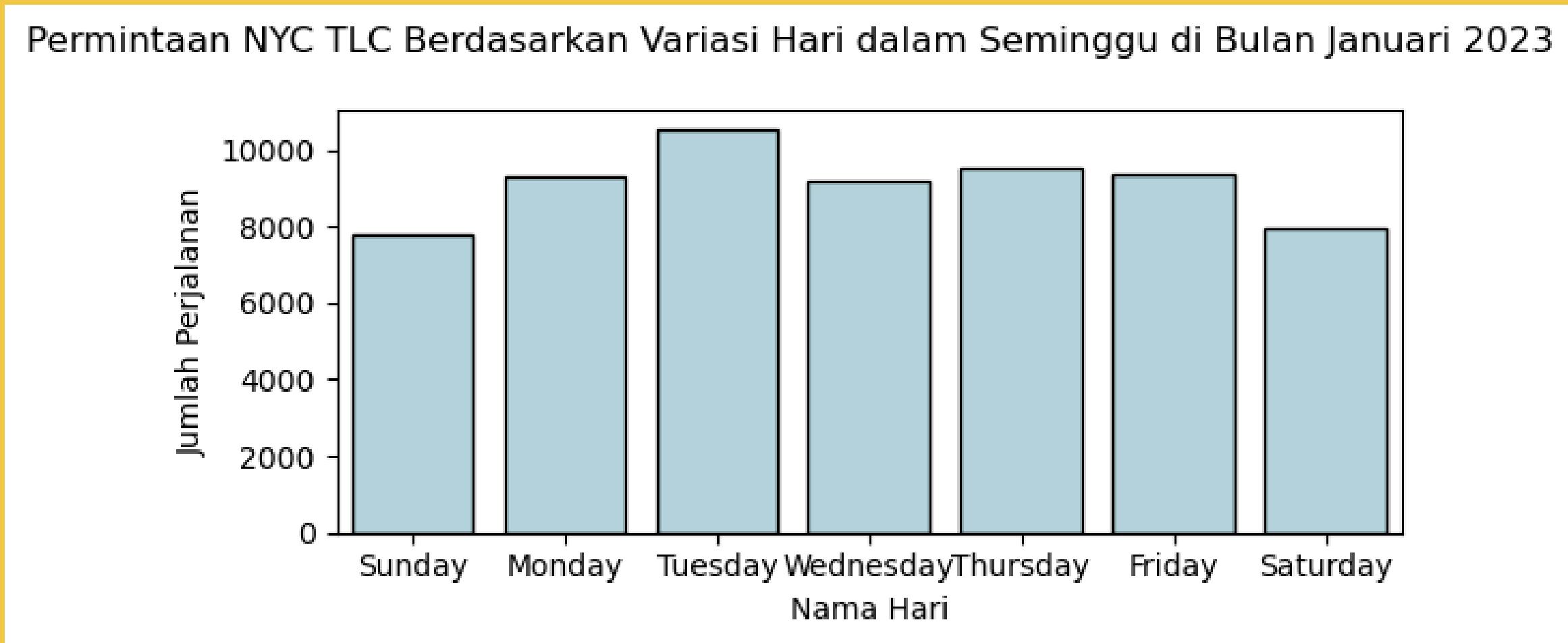
- **Naik:** 05:00 - 09:00 dan 13:00 - 15:00 (jam sibuk pagi dan sore)
- **Turun:** 18:00 - larut
- **Turun drastis:** 00.00 - subuh

Berdasarkan kategori waktu penjemputan



- **Naik:** Pagi dan sore hari (pergi dan pulang beraktivitas)
- **Turun:** 18:00 - larut
- **Turun drastis:** 00.00 - subuh

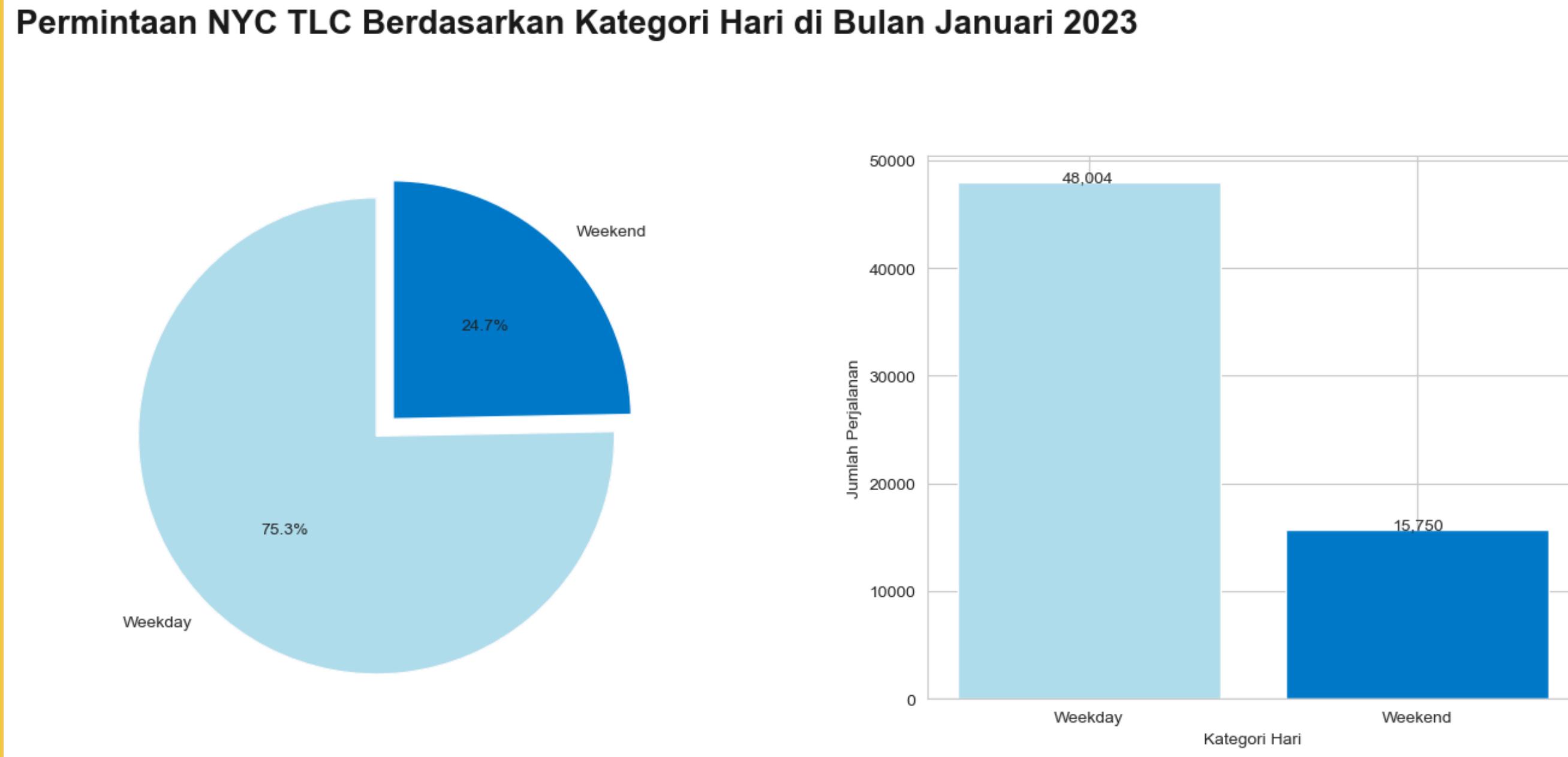
Berdasarkan hari dalam seminggu



Tinggi: Selasa
Rendah: Senin

Tetapi cenderung merata dengan rata-rata 9.107 perjalanan per hari

Berdasarkan kategori hari



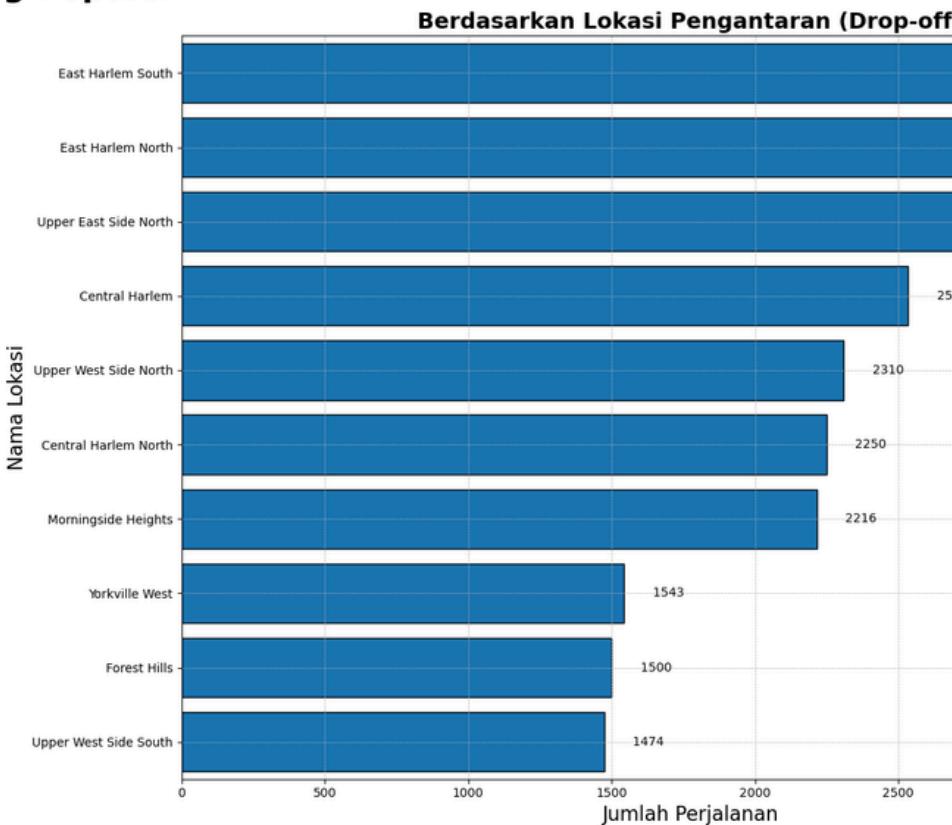
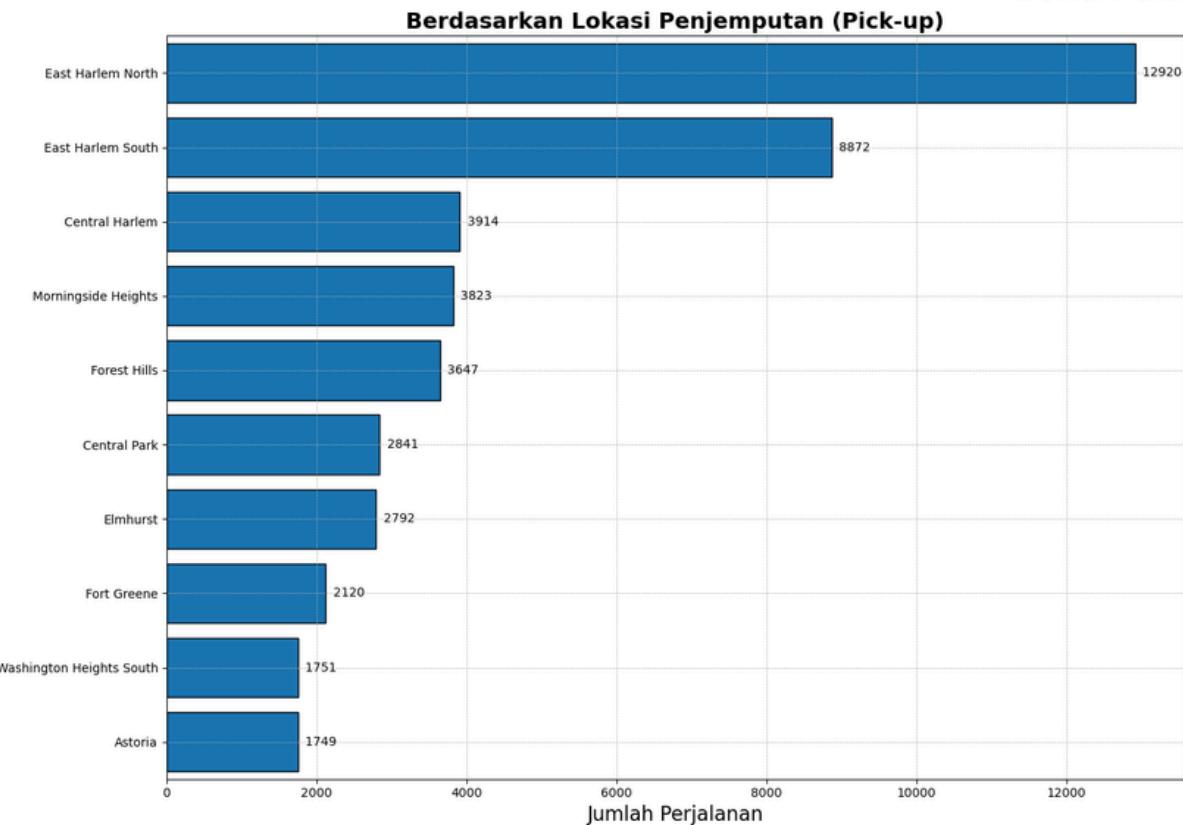
Hampir seperempat persentase
terjadi di kategori *weekend* (Sabtu-Minggu)

Sisanya terjadi di kategori *weekday* (Senin-Jumat)

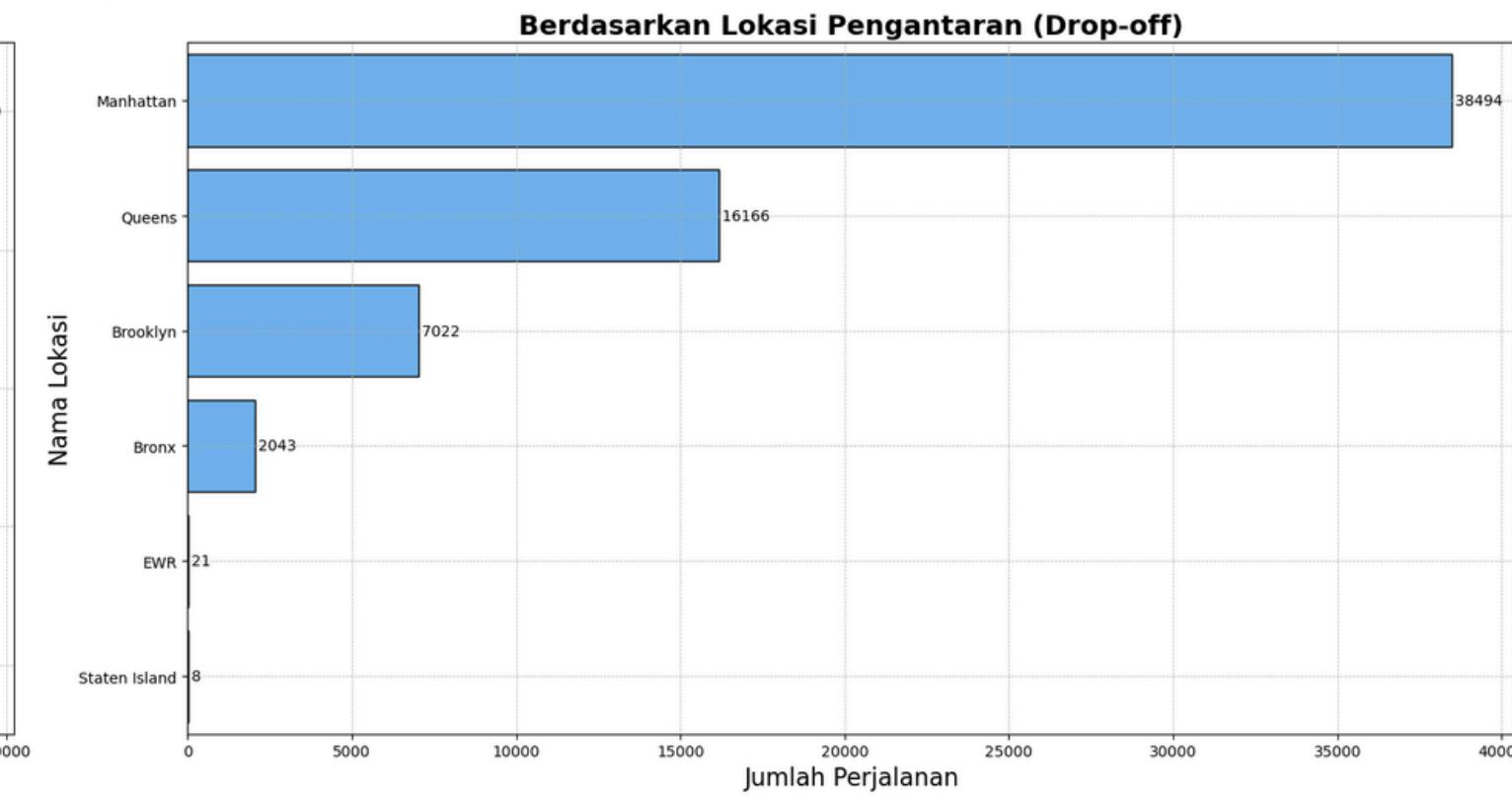
**Dimana saja *zona* dan
borough penjemputan (*pick-*
up) dan pengantaran (*drop-*
off) yang paling populer?**



Zona Paling Populer



Borough Paling Populer

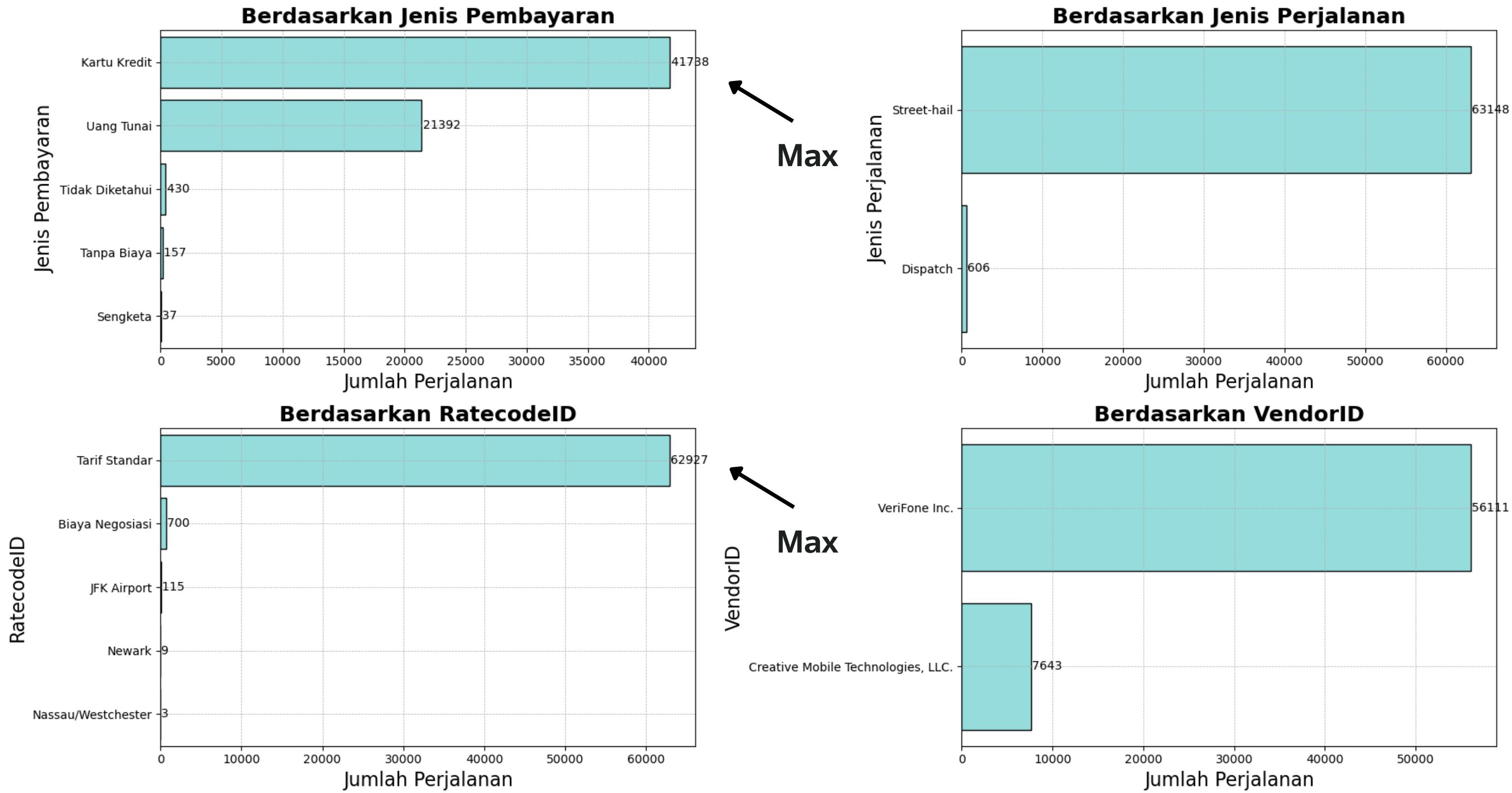


- **Zona Harlem (East Harlem North, South, dan Central)** menjadi zona dengan jumlah perjalanan (penjemputan & pengantaran) terbanyak
- **Borough Manhattan** (yang berisi zona Harlem) memiliki jumlah perjalanan terbanyak

**Bagaimana preferensi dan
prilaku pelanggan ketika
menggunakan NYC TLC?**



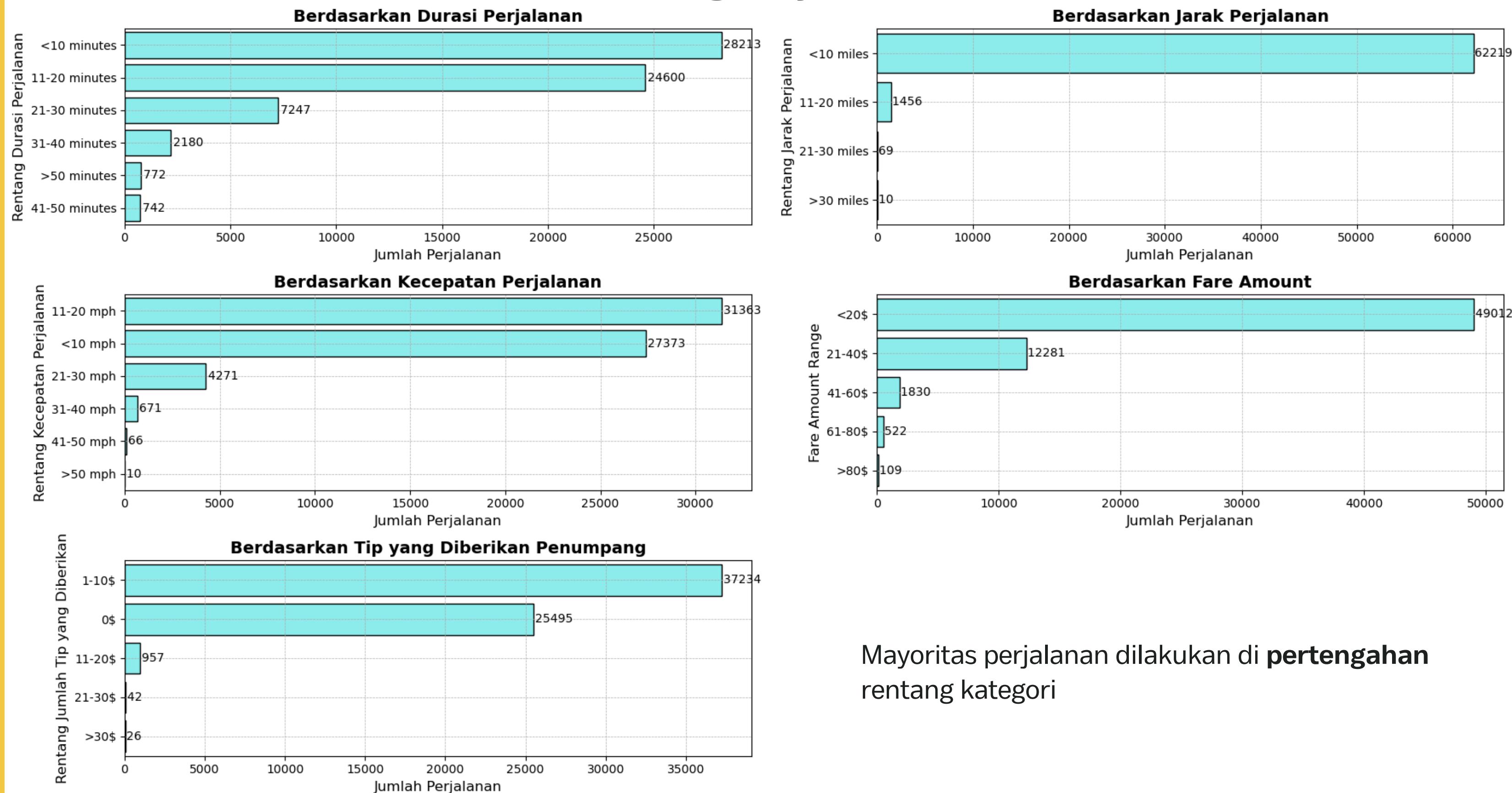
Preferensi dan Prilaku Pelanggan



**Bagaimana *variasi range*
perjalanan yang dilakukan?**



Variasi Rentang Perjalanan NYC TLC

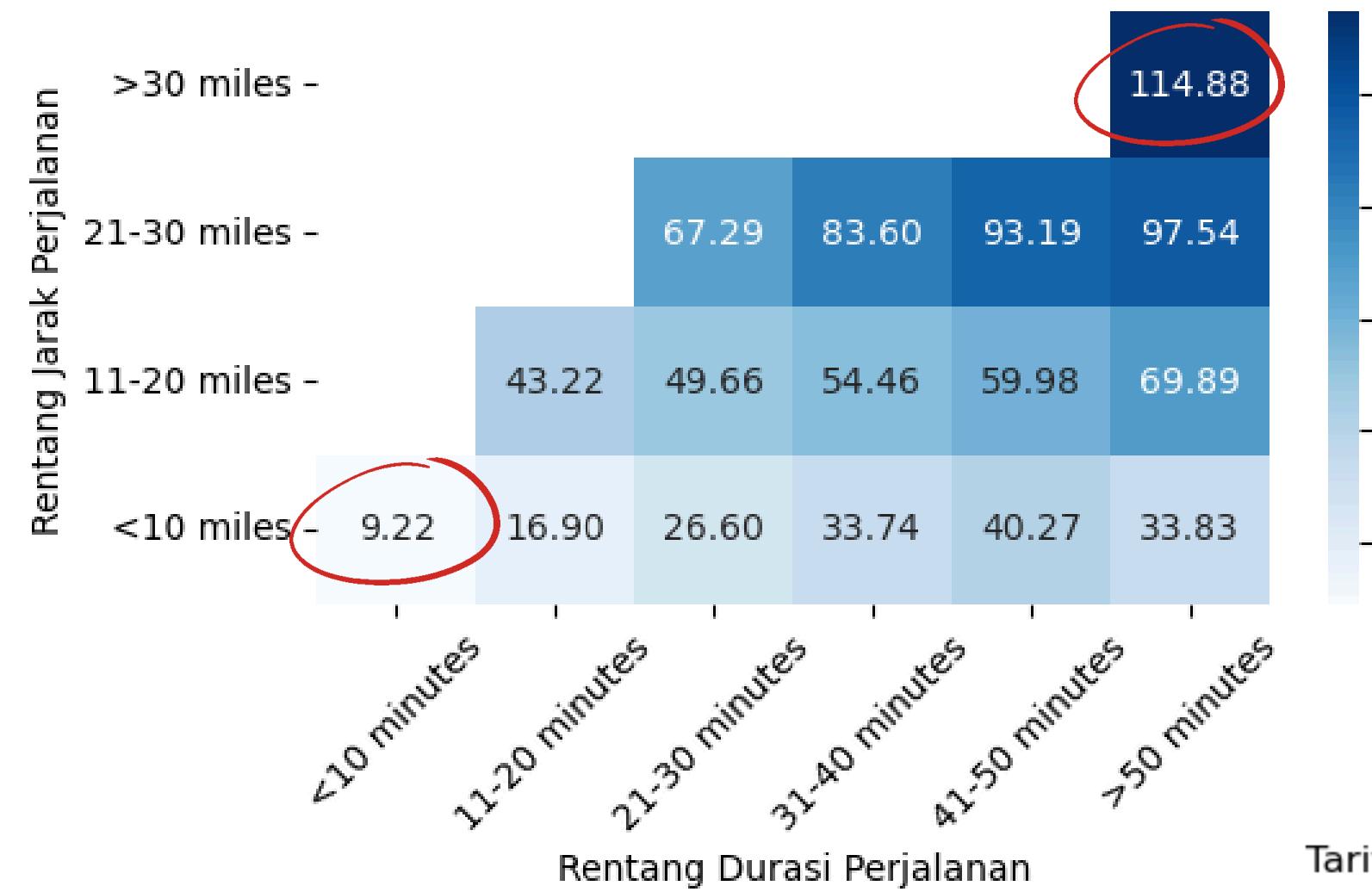


Mayoritas perjalanan dilakukan di **pertengahan** rentang kategori

**Bagaimana pengaruh jarak
perjalanan dan durasi
perjalanan terhadap *fare
amount* dan *total amount*?**

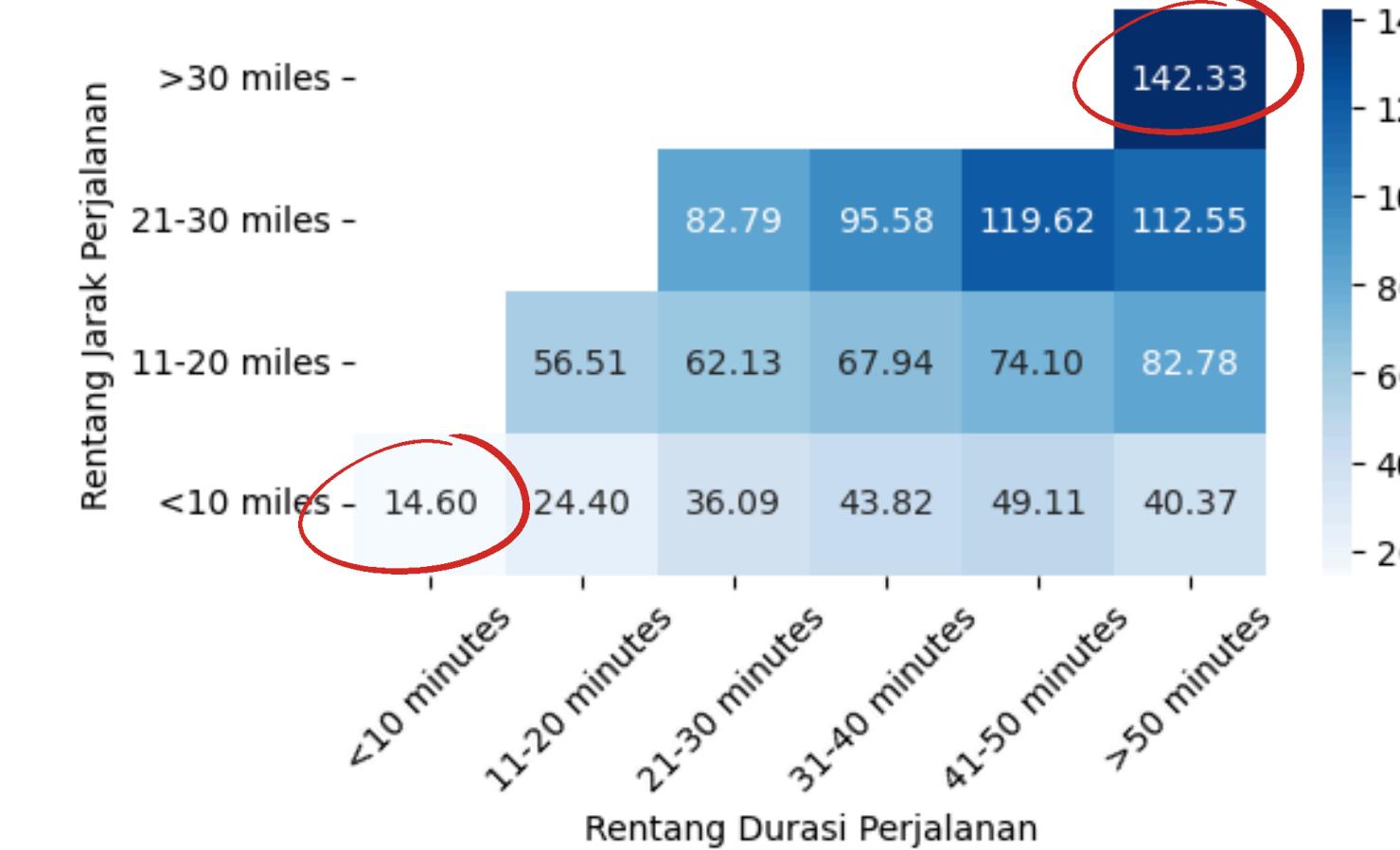


Rata-rata Fare Amount Berdasarkan Jarak dan Durasi Perjalanan pada Bulan Januari 2023



Semakin jauh dan lama perjalanan,
maka **semakin besar fare amount**
dan total amount yang dikeluarkan
oleh penumpang, dan sebaliknya

Tarif Total Rata-rata Berdasarkan Jarak dan Durasi Perjalanan pada Bulan Januari 2023

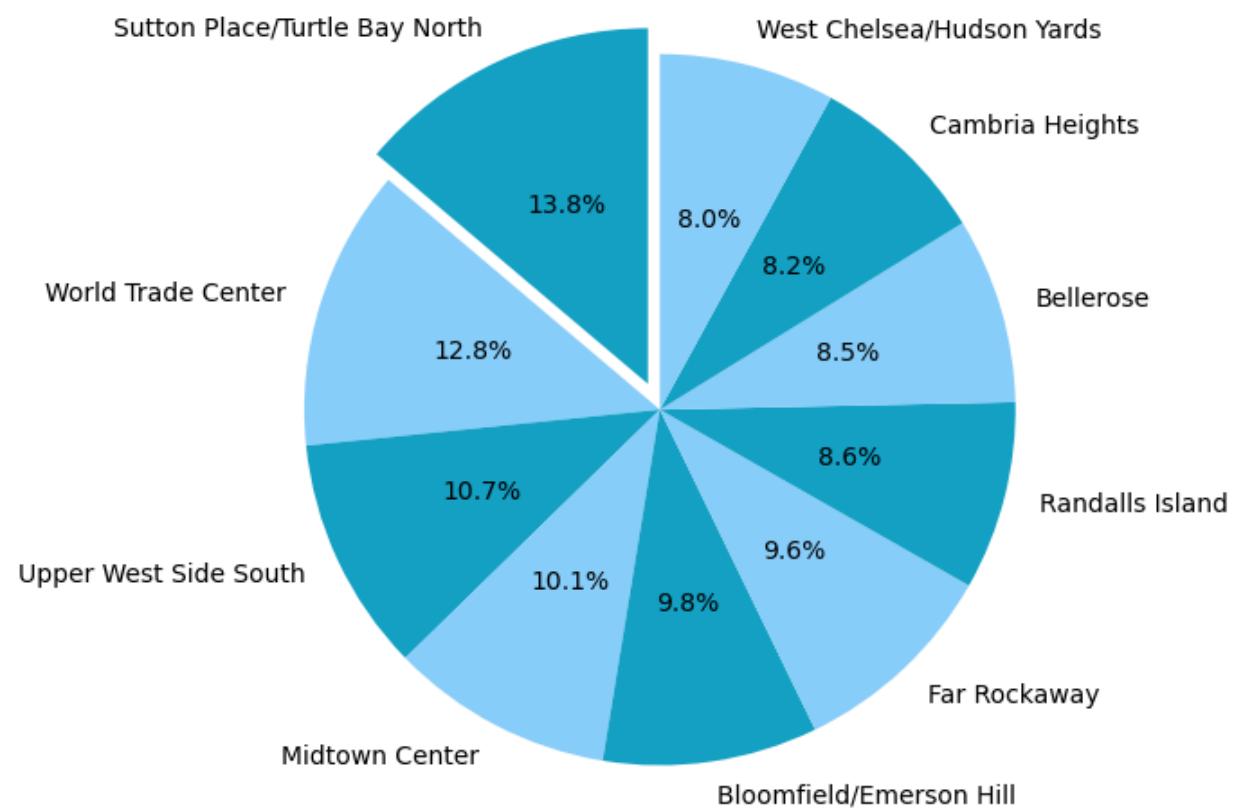


Apakah ada hubungan antara
zona dan ***borough*** dengan
total amount yang
dikeluarkan oleh penumpang?

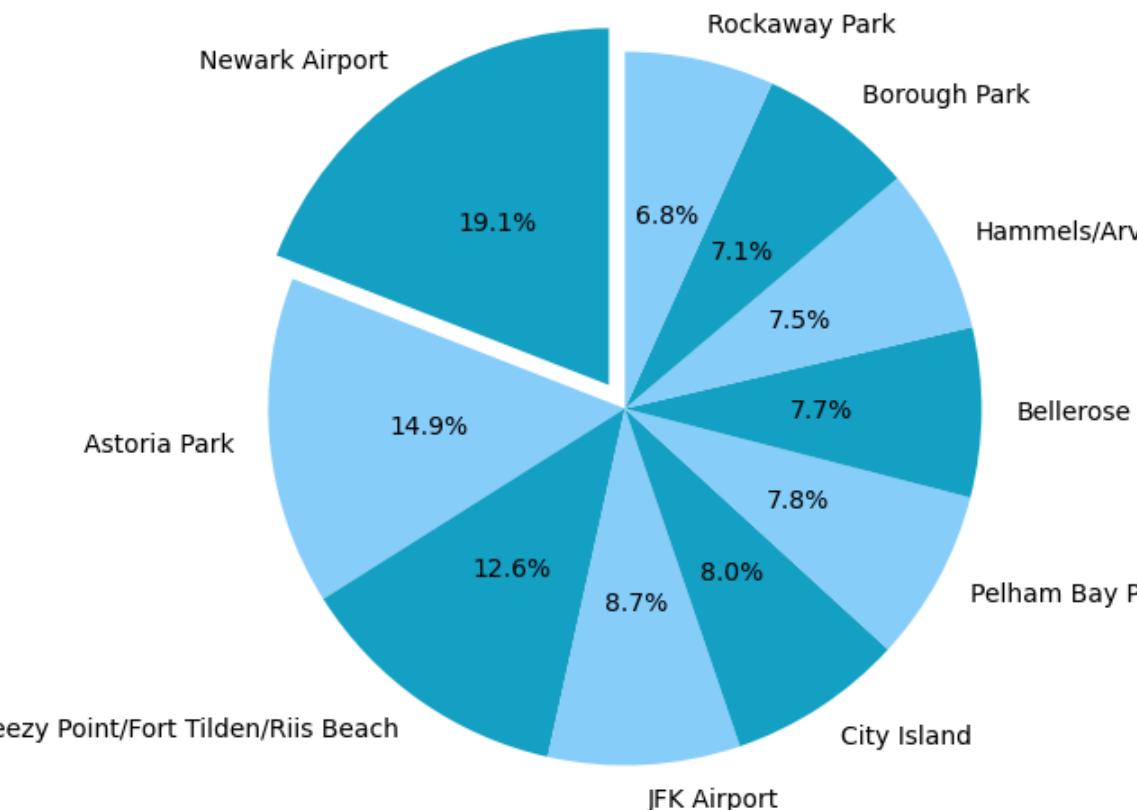


Persentase Rata-rata Total Biaya yang Dikeluarkan Penumpang

Berdasarkan Zona Lokasi Penjemputan (Pick-up)



Berdasarkan Zona Lokasi Pengantaran (Drop-off)

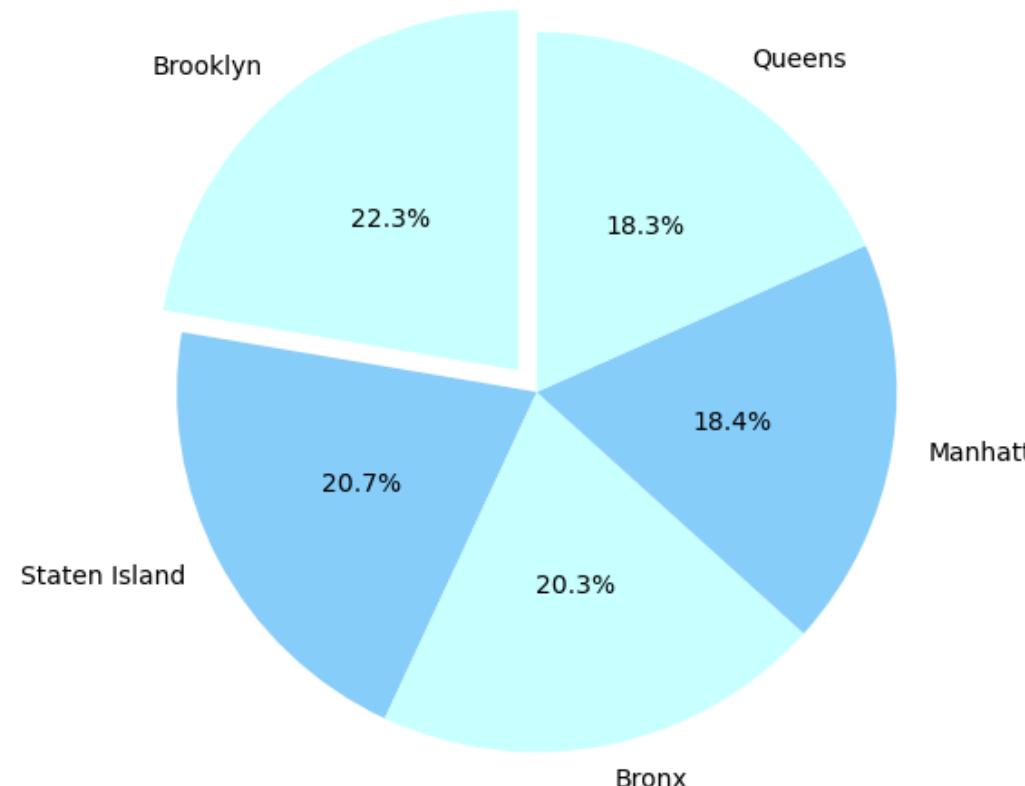


Zona Harlem yang menjadi zona dengan jumlah perjalanan terbanyak tidak masuk ke 10 zona dengan total biaya yang dikeluarkan paling banyak

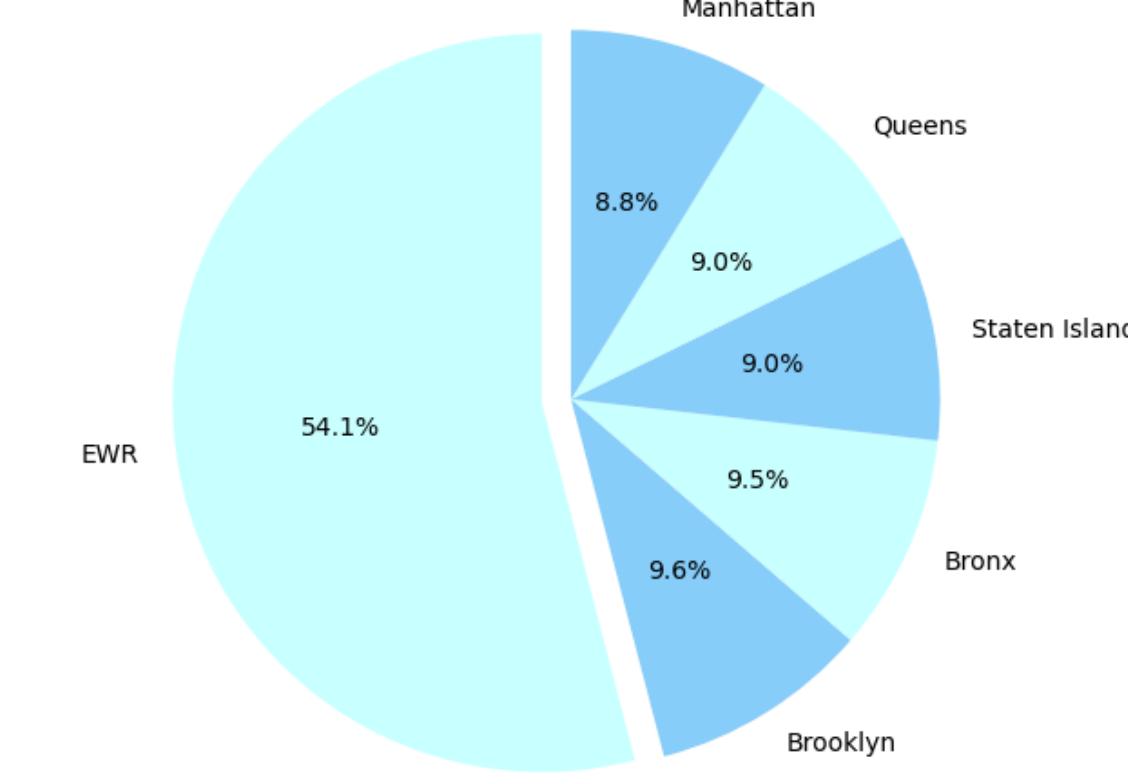
Persentase Rata-rata Total Biaya yang Dikeluarkan Penumpang

Borough Manhattan yang menjadi borough dengan jumlah perjalanan terbanyak tidak masuk ke borough dengan total biaya yang dikeluarkan paling banyak

Berdasarkan Borough Lokasi Penjemputan (Pick-up)

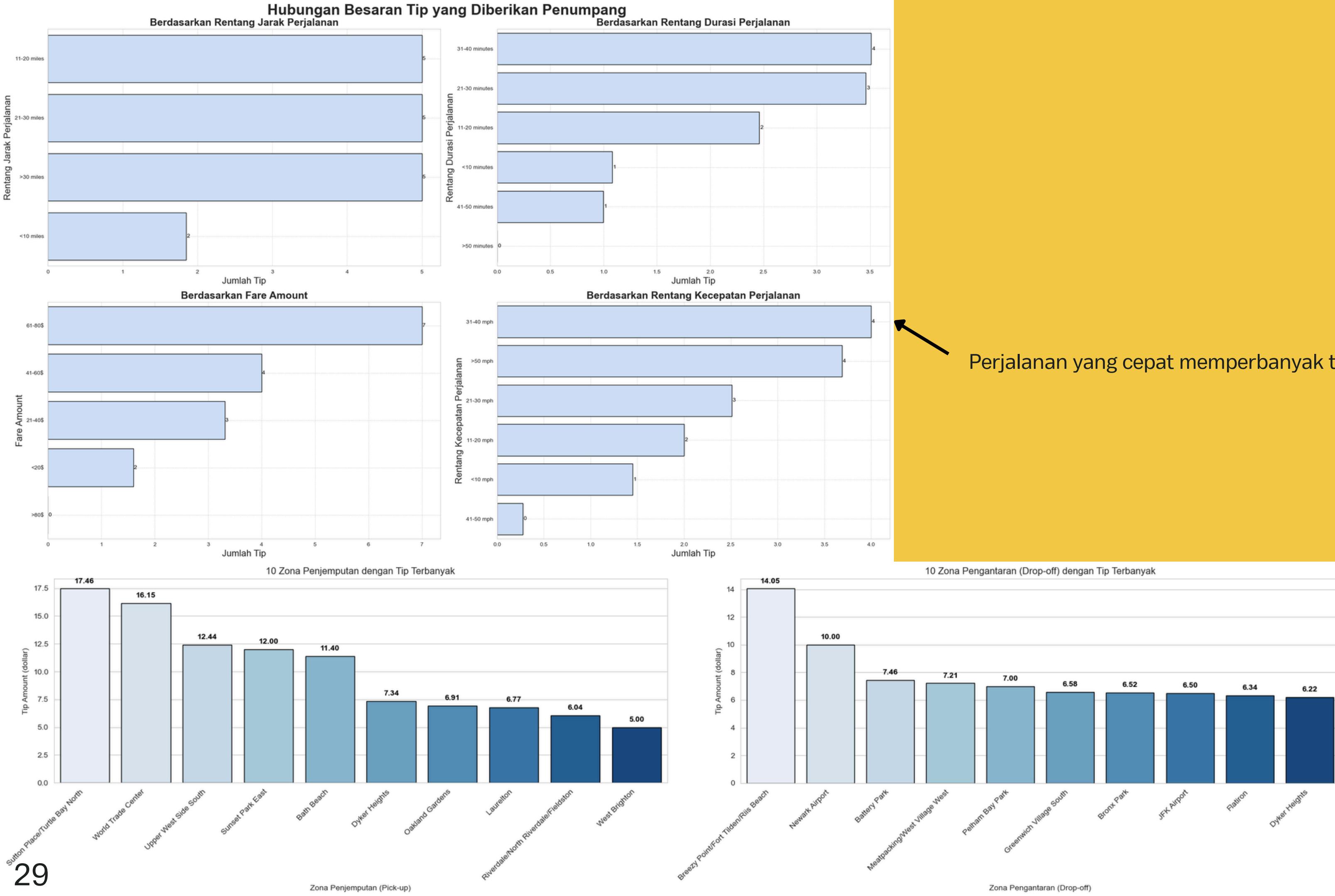


Berdasarkan Borough Lokasi Pengantaran (Drop-off)



Apa saja faktor yang mempengaruhi besaran tip yang diberikan oleh penumpang?





Kesimpulan

- Permintaan **tinggi** pada akhir **weekdays** dan **rendah** pada **weekend** dan **hari libur**, terutama di **jam-jam berangkat dan pulang beraktivitas**.
- **Zona Harlem di borough Manhattan** memiliki jumlah perjalanan (penjemputan & pengantaran) terbanyak.
- Pelanggan mayoritas membayar menggunakan **kartu kredit** dan melakukan perjalanan **street-hail** dengan **tarif standar (perjalanan dalam kota)**.
- Perjalanan mayoritas **kurang dari 10 mil** dan **kurang dari 10 menit**.
- Jumlah total yang perlu dibayar oleh penumpang meningkat sejalan dengan meningkatnya jarak dan durasi perjalanan.
- Lokasi zona penjemputan dan pengantaran mempengaruhi total biaya berdasarkan kategori dan karakteristik lokasi.
- Perbedaan jarak, durasi, kecepatan, jumlah pembayaran, dan lokasi penumpang mempengaruhi besaran tip.



Saran

- Tingkatkan jumlah TLC di waktu sibuk serta zona permintaan yang tinggi untuk mengurangi waktu tunggu pelanggan.
 - Berikan sejumlah diskon dan kurangi jumlah taksi di waktu sepi serta zona permintaan yang rendah untuk menarik perhatian pelanggan.
 - Berikan intensif ke pengebudi yang beroperasi di area permintaan yang tinggi.
-
- Memberi tarif lebih tinggi di waktu sibuk dan di zona dengan permintaan tinggi, dan sebaliknya.
 - Menawarkan paket tarif untuk pelanggan setia.
 - Berikan diskon untuk perjalanan jarak jauh dan memesan melalui *dispatch*.



Tableau



DATA ANALYSIS OF NEW YORK CITY TAXIS AND LIMOUSINES COMMISSION (NYC TLC) TRIP RECORD

NYC TLC is a company/agency under the government of New York City in the United States which includes the vehicle rental industry, regulates, and facilitates Taxis/Limousines that can be used by city residents or tourists in New York City

Demand Analysis based on Time

Pick-up and Drop-off Demand Analysis

Behaviour and Preferences Demand Analysis

Range Demand Analysis

Price Analysis

Heat Map Analysis

by Athaya Zahrani Irmansyah



Thank You

New York City Taxi and Limousine Commission
(NYC TLC) Data Analysis



Athaya Zahrani Irmansyah

Wed, 8/7/2024