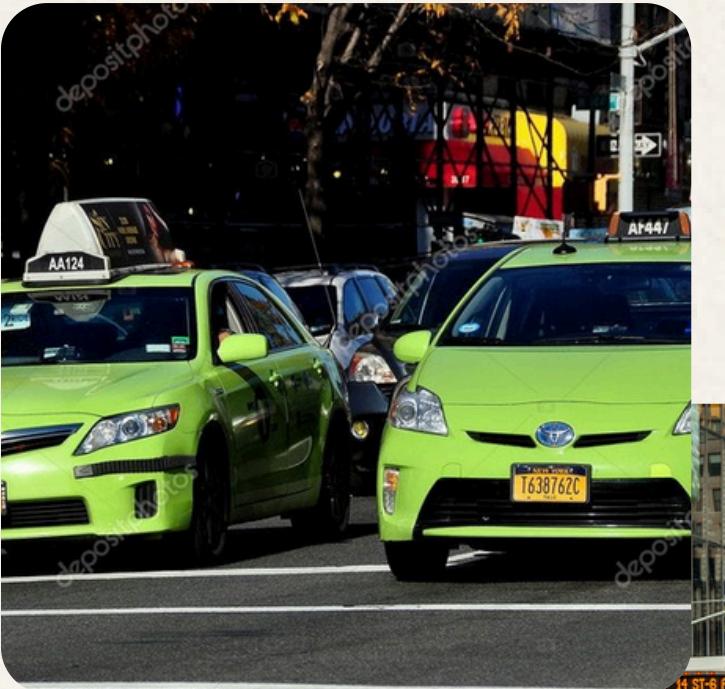




PURWADHIKA JCDS 0808

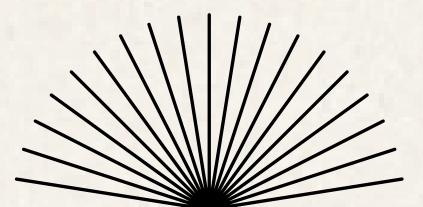


MENINGKATKAN KINERJA GREEN TAXI COMPANY

Penyusunan strategi berbasis data untuk meningkatkan layanan, efisiensi, dan kepuasan pelanggan

DISUSUN OLEH

JCDS 0808 - 007 - Yanfa Anandika



Bussiness Understanding

- Green Taxi diperkenalkan oleh NYC TLC pada tahun 2013
- Green Taxi bertujuan untuk menangani wilayah Manhattan Utara dan Luar Manhattan
- Kepemilikan Green Taxi bisa oleh individu atau perusahaan taxi maupun pengusaha lokal
- Regulasi, perizinan, lisensi di tangani oleh New York City Taxi & Limousine Commission (NYC TLC)
- Kendaraan berlisensi NYC TLC (Yellow Taxi, Green Taxi, For Hire Vehicle, Commuter Van, Paratransit Vehicle)



NYC[®]
**Taxi & Limousine
Commission**



Executive Summary

Latar Belakang

Green Taxi sebagai sarana transportasi warga atau wisatawan di kota New York

Stake Holders

Perusahaan Green Taxi

Permasalahan Utama

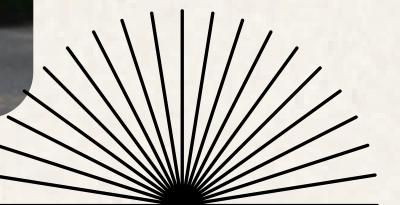
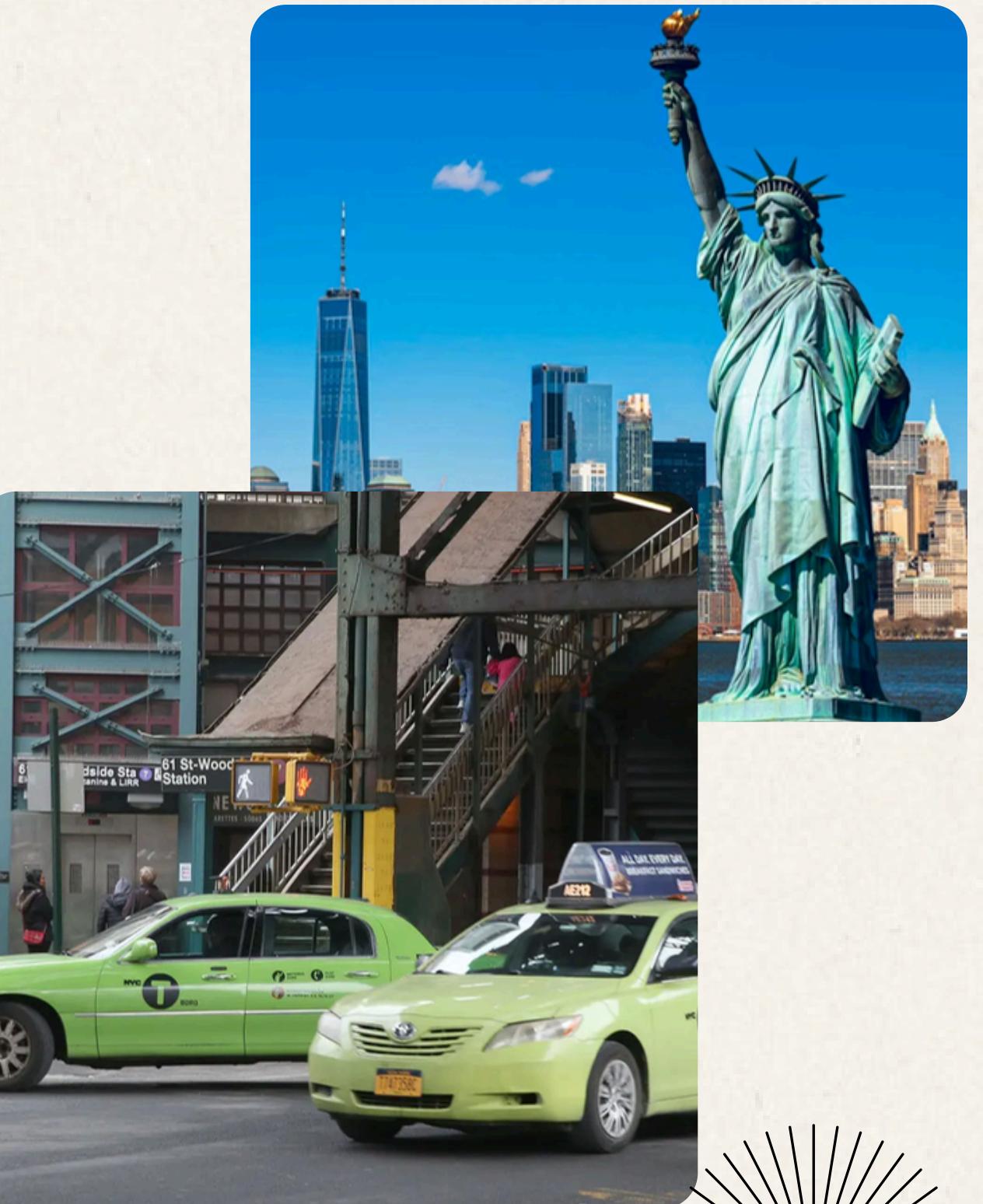
Persaingan dengan Yellow Taxi, Uber, Lyft

Rumusan Masalah

Meningkatkan kinerja dan kepuasan pelanggan Green Taxi

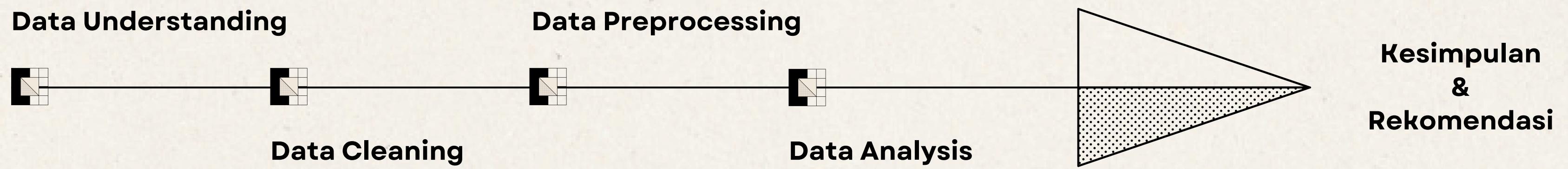
Strategi

- Mencari lokasi yang ramai permintaan
- Mencari waktu yang ramai permintaan
- Mengidentifikasi preferensi/perilaku pelanggan



Tahapan

Tahap tahap yang akan dilakukan untuk menyelesaikan permasalahan yang ada



Data Understanding

Data yang digunakan adalah data dari New York City Taxi & Limousine Commission (NYC TLC)

NYC TLC Trip Record.csv

Isi Data :

No	Kolom	Deskripsi
1	VendorID	Provider perangkat pencatatan perjalanan 1 = Creative Mobile Technologies,LLC. 2 = VeriFoneInc.
2	lpep_pickup_datetime	Waktu penjemputan penumpang
3	lpep_dropoff_datetime	Waktu penurunan penumpang
4	store_and_fwd_flag	Catatan perjalanan disimpan diperangkat sebelum dikirim ke server, karena kehilangan koneksi Y = Data disimpan diperangkat sebelum dikirim ke server N = Data langsung data dikirim
5	RatecodeID	Jenis tarif perjalanan 1 = Standard rate 2 = JFK Airport 3 = Newark 4 = Nassau or Westchester 5 = Negotiated fare 6 = Group ride
6	PULocationID	ID lokasi penjemputan penumpang berdasarkan zona geografis NYC TLC
7	DOLocationID	ID lokasi penurunan penumpang berdasarkan zona geografis NYC TLC
8	passenger_count	Jumlah penumpang dalam perjalanan
9	trip_distance	Jarak tempuh perjalanan
10	fare_amount	Tarif dasar berdasarkan waktu dan jarak

11	extra	Biaya tambahan (malam, atau jam sibuk) \$1 (jam sibuk 4pm - 8pm) \$1 (jam malam 8pm - 6am)
12	mta_tax	Pajak MTA (Metropolitan Transportation Authority) \$0.50
13	tip_amount	Jumlah tip yang diberikan penumpang dengan pembayaran credit card
14	tolls_amount	Biaya tol selama perjalanan
15	ehail_fee	Biaya layanan pemesanan online/aplikasi
16	improvement_surcharge	Biaya tambahan untuk pengembangan layanan \$0.30
17	total_amount	Total jumlah yang dibayar termasuk pajak dan biaya tambahan
18	payment_type	Metode pembayaran 1 = Credit card 2 = Cash 3 = No charge 4 = Dispute 5 = Unknown 6 = Voided trip
19	trip_type	Jenis perjalanan 1 = Street-hail 2 = Dispatch
20	congestion_surcharge	Biaya tambahan untuk kemacetan \$2.75 Green Taxi

Additional Data

Data tambahan yang digunakan adalah data dari New York City Taxi & Limousine Commission (NYC TLC)

`taxi_zone_lookup.csv`

Isi Data :

No	Kolom	Deskripsi
1	LocationID	ID wilayah
2	Borough	Nama wilayah
3	Zone	Nama zona
4	service_zone	Zona operasi Yellow Taxi / Green Taxi

Data Overview

Berikut adalah data utama dan tambahan yang telah di gabungkan

	column	unique_total	unique_sample
0	VendorID	2	[2, 1]
1	lpep_pickup_datetime	66575	[2023-01-01 00:26:10, 2023-01-01 00:51:03, 202...
2	lpep_dropoff_datetime	66519	[2023-01-01 00:37:11, 2023-01-01 00:57:49, 202...
3	store_and_fwd_flag	2	[N, Y, nan]
4	RatecodeID	6	[1.0, 5.0, 4.0, 3.0, 2.0, 99.0, nan]
5	PULocationID	226	[166, 24, 223, 41, 181, 255, 75, 66, 195, 83, ...]
6	DOLocationID	249	[143, 43, 179, 238, 74, 262, 45, 75, 166, 140, ...]
7	passenger_count	10	[1.0, 2.0, 4.0, 3.0, 6.0, 5.0, 0.0, 7.0, 9.0, ...]
8	trip_distance	1870	[2.58, 1.81, 0.0, 1.3, 1.1, 2.78, 3.8, 1.88, 1...
9	fare_amount	2553	[14.9, 10.7, 7.2, 6.5, 6.0, 17.7, 19.1, 14.2, ...]
10	extra	16	[1.0, 0.5, 3.75, 0.0, 3.25, 6.0, 5.0, 2.75, 5....]
11	mta_tax	6	[0.5, 1.5, 0.0, 1.0, -0.5, 2.75]
12	tip_amount	1492	[4.03, 2.64, 1.94, 1.7, 0.0, 4.85, 1.0, 3.0, 3...
13	tolls_amount	26	[0.0, 6.55, 4.0, 13.75, 9.0, 3.0, 11.75, 2.45, ...]
14	ehail_fee	0	[nan]
15	improvement_surcharge	5	[1.0, 0.0, 0.3, -1.0, -0.3]
16	total_amount	4670	[24.18, 15.84, 11.64, 10.2, 8.0, 22.95, 29.2, ...]
17	payment_type	5	[1.0, 2.0, 3.0, 4.0, 5.0, nan]
18	trip_type	2	[1.0, 2.0, nan]
19	congestion_surcharge	4	[2.75, 0.0, 2.5, -2.75, nan]
20	pu_borough	7	[Manhattan, Queens, Brooklyn, Bronx, Outside o...
21	pu_zone	224	[Morningside Heights, Bloomingdale, Steinway, ...]
22	do_borough	7	[Manhattan, Queens, Brooklyn, Outside of NYC, ...]
23	do_zone	247	[Lincoln Square West, Central Park, Old Astori...

Uji normalitas data menggunakan Shapiro-Wilk Test - 2000 Sampel

	kolom	pvalue	hasil
0	passenger_count	7.402673e-62	Tidak terdistribusi normal
1	trip_distance	1.215916e-72	Tidak terdistribusi normal
2	fare_amount	4.335857e-58	Tidak terdistribusi normal
3	extra	3.504858e-52	Tidak terdistribusi normal
4	mta_tax	2.809343e-55	Tidak terdistribusi normal
5	tip_amount	2.226392e-43	Tidak terdistribusi normal
6	tolls_amount	6.571658e-70	Tidak terdistribusi normal
7	improvement_surcharge	2.927246e-67	Tidak terdistribusi normal
8	total_amount	1.913345e-54	Tidak terdistribusi normal
9	congestion_surcharge	3.388059e-55	Tidak terdistribusi normal

Data Cleaning

VendorID	0.0%	VendorID	0
lpep_pickup_datetime	0.0%	lpep_pickup_datetime	0
lpep_dropoff_datetime	0.0%	lpep_dropoff_datetime	0
store_and_fwd_flag	6.34%	store_and_fwd_flag	4324
RatecodeID	6.34%	RatecodeID	4324
PULocationID	0.0%	PULocationID	0
DOLocationID	0.0%	DOLocationID	0
passenger_count	6.34%	passenger_count	4324
trip_distance	0.0%	trip_distance	0
fare_amount	0.0%	fare_amount	0
extra	0.0%	extra	0
mta_tax	0.0%	mta_tax	0
tip_amount	0.0%	tip_amount	0
tolls_amount	0.0%	tolls_amount	0
ehail_fee	100.0%	ehail_fee	68211
improvement_surcharge	0.0%	improvement_surcharge	0
total_amount	0.0%	total_amount	0
payment_type	6.34%	payment_type	4324
trip_type	6.35%	trip_type	4334
congestion_surcharge	6.34%	congestion_surcharge	4324
pu_borough	0.0%	pu_borough	0
pu_zone	0.0%	pu_zone	0
do_borough	0.0%	do_borough	0
do_zone	0.0%	do_zone	0

Menghapus kolom yang tidak digunakan dalam analisis

- ehail_fee
- VendorID
- store_and_fwd_flag

Cleaning kolom (RatecodeID)

- mengganti nilai 99 → 1
- berdasarkan pu_boroughnya
- mengisi missing value dengan median

Cleaning kolom (passenger_count)

- mengganti nilai 0 → 1
- mengisi missing value dengan median

Cleaning kolom (payment_type)

- mengisi missing value dengan nilai 5 yaitu tipe pembayaran Unknown

Cleaning kolom (trip_type)

- mengisi missing value dengan median

Cleaning kolom (congestion_surcharge)

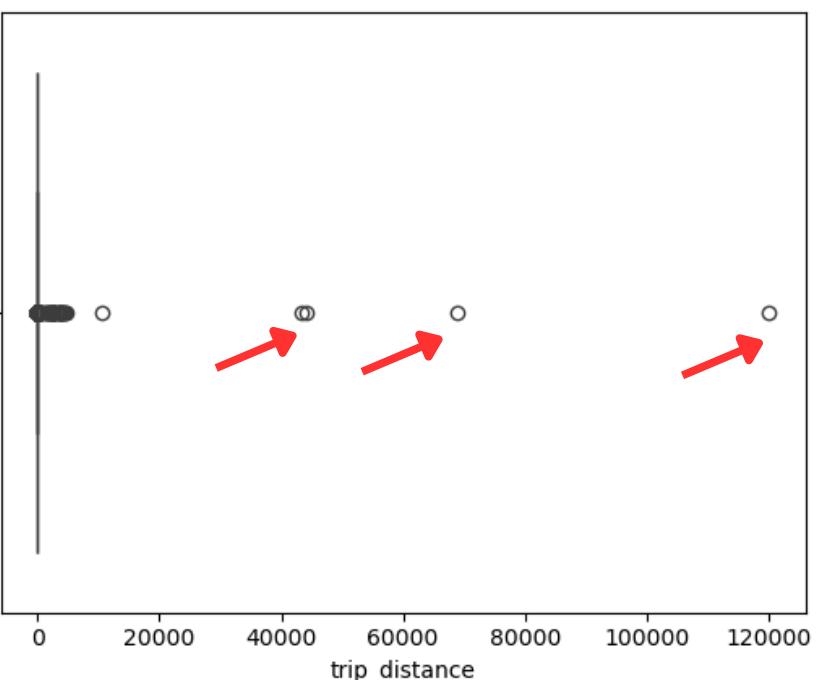
- menghapus nilai < 0
- mengisi missing value dengan median

Handling Outliers

Melakukan pengecekan outlier untuk dipertimbangkan dalam analisis

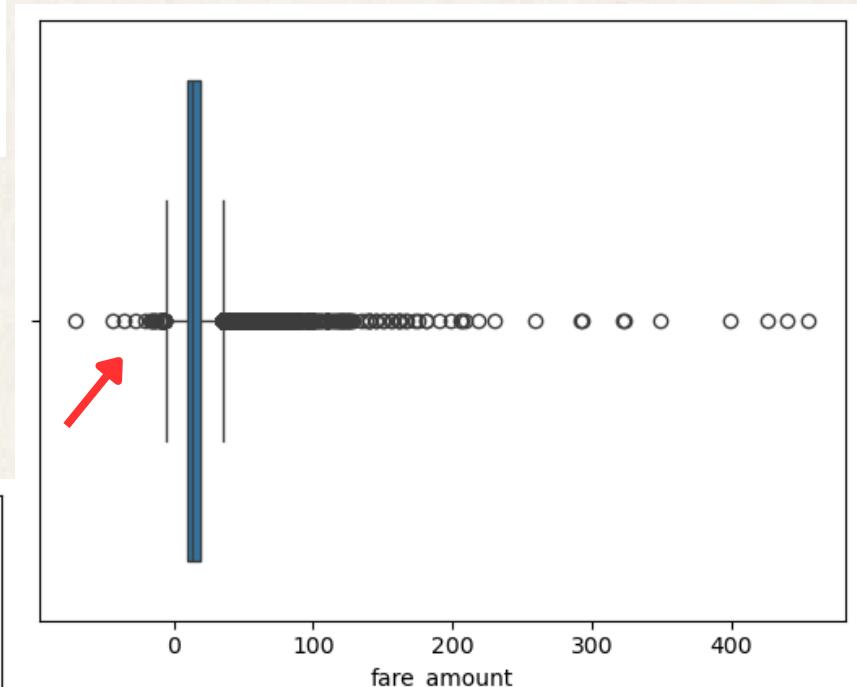
Outlier kolom (`trip_distance`)

- membatasi nilai maksimum di 100 mil
- karena diatas jarak tersebut sudah diluar state NYC



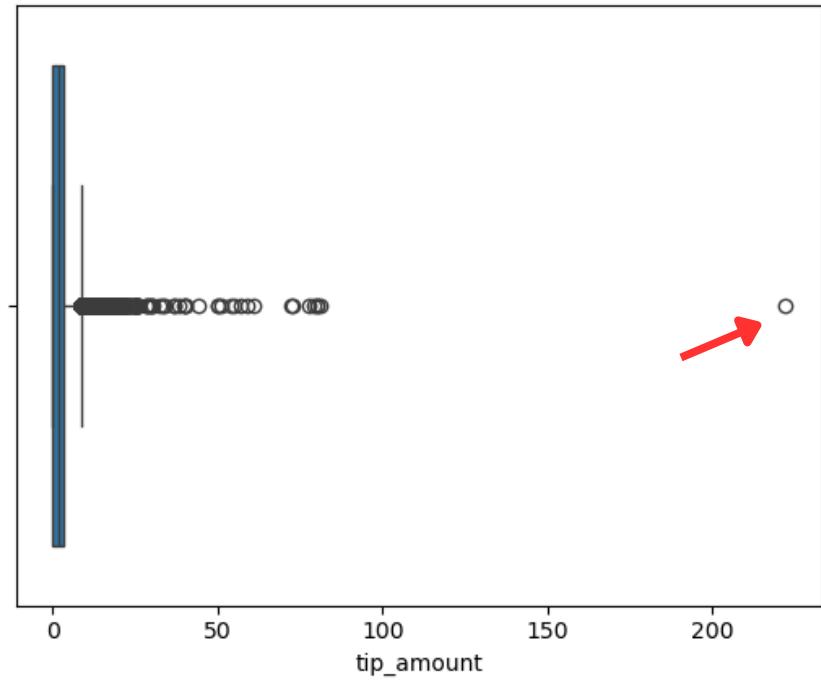
Outlier kolom (`fare_amount`)

- membatasi nilai terendah \$3
- sesuai dengan tarif minimum NYC TLC



Outlier kolom (`tip_amount`)

- membatasi nilai maksimum \$100



Data Preprocessing

Sebelum melakukan analisis lebih lanjut, dilakukan pengkondisian data sehingga sesuai dengan kebutuhan analisis

Mengubah nilai kolom sesuai artinya :

RatecodeID

- 1 = Standard rate
- 2 = JFK Airport
- 3 = Newark
- 4 = Nassau or Westchester
- 5 = Negotiated fare
- 6 = Group ride

payment_type

- 1 = Credit card
- 2 = Cash
- 3 = No charge
- 4 = Dispute
- 5 = Unknown
- 6 = Voided trip

trip_type

- 1 = Street-hail
- 2 = Dispatch

RatecodeID	payment_type	trip_type
1.0	1.0	1.0
1.0	1.0	1.0
1.0	1.0	1.0
1.0	1.0	1.0
1.0	1.0	1.0
...



RatecodeID	payment_type	trip_type
Standard Rate	Credit Card	Street-hail
Standard Rate	Credit Card	Street-hail
Standard Rate	Credit Card	Street-hail
Standard Rate	Credit Card	Street-hail
Standard Rate	Cash	Street-hail

Data Preprocessing

Membuat kolom (trip_duration)

- waktu dropoff - pickup
 - menghapus (trip_duration) = 0
 - menghapus (trip_duration) > 3 jam

Mengubah kolom tanggal menjadi datetime format

- (`lpep_pickup_datetime`) dan (`lpep_dropoff_datetime`) menjadi datetime
 - menghapus data selain tahun 2023 dan bulan Januari
 - membuat kolom (`pu_hour`)
 - membuat kolom (`pu_day`)
 - membuat kolom (`day_category`)
 - membuat kolom (`pu_date`)

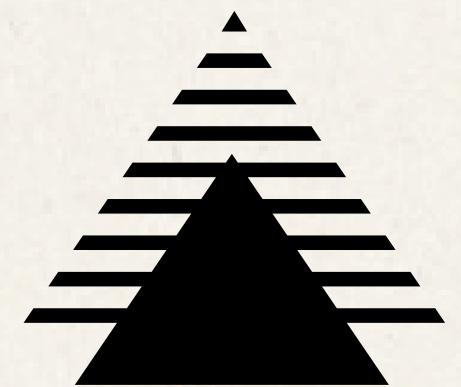
Membuat kolom total pembayaran yang valid (total_payment)

- menghitung ulang biaya biaya (fare_amount + extra + mta_tax + tip_amount + tolls_amount + improvement_surcharge + congestion_surcharge)
 - membandingkan dengan kolom (total_amount) yang tersedia dari data

Sebelum perhitungan ulang



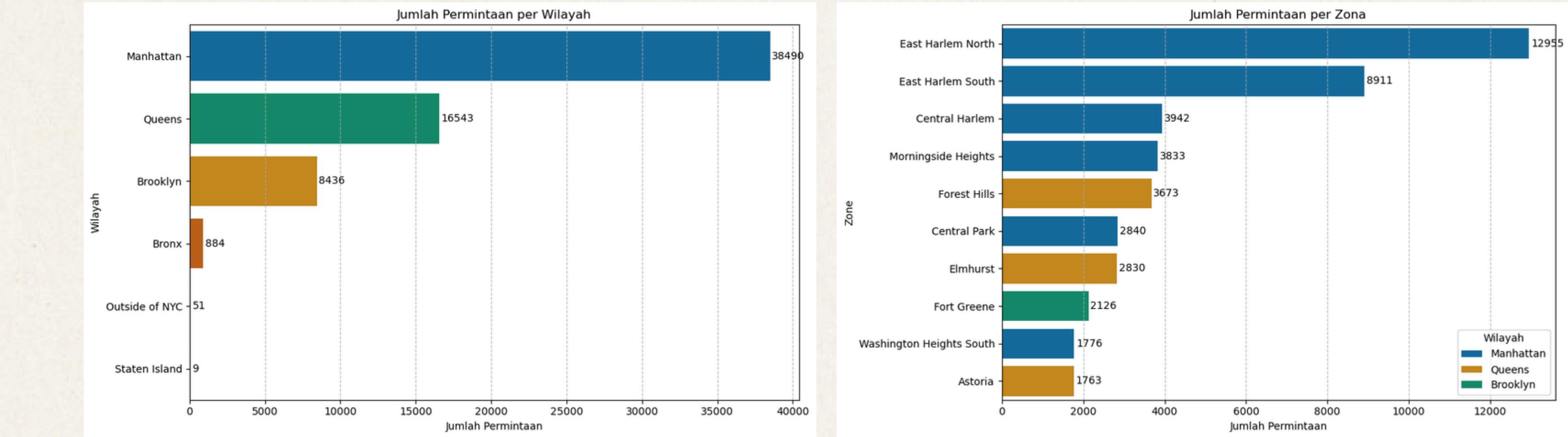
Data Analysis





**Mencari lokasi yang memiliki
permintaan tinggi**

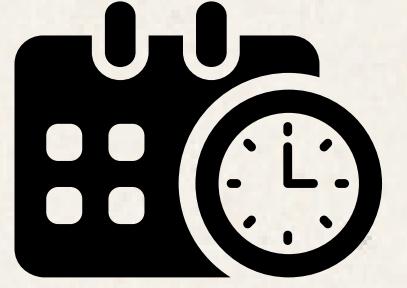
Mencari lokasi yang memiliki permintaan tinggi



- Manhattan memiliki permintaan terbanyak
- Diikuti oleh Queens dan Brooklyn

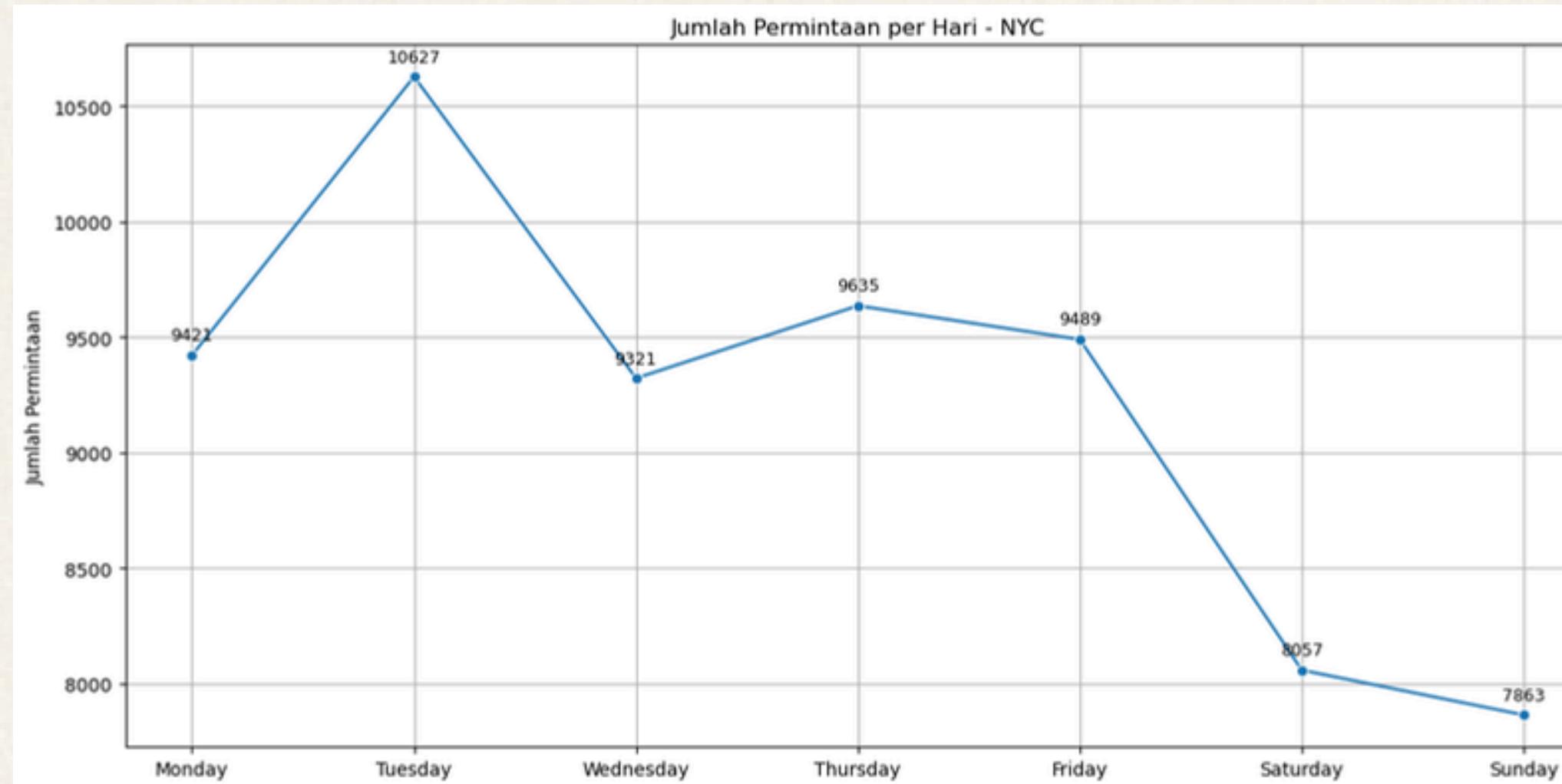
- Zona dengan permintaan terbanyak adalah zona di wilayah Manhattan

Wilayah serta zona tersebut akan menjadi pusat fokus utama untuk menyiapkan ketersediaan armada disekitar area tersebut

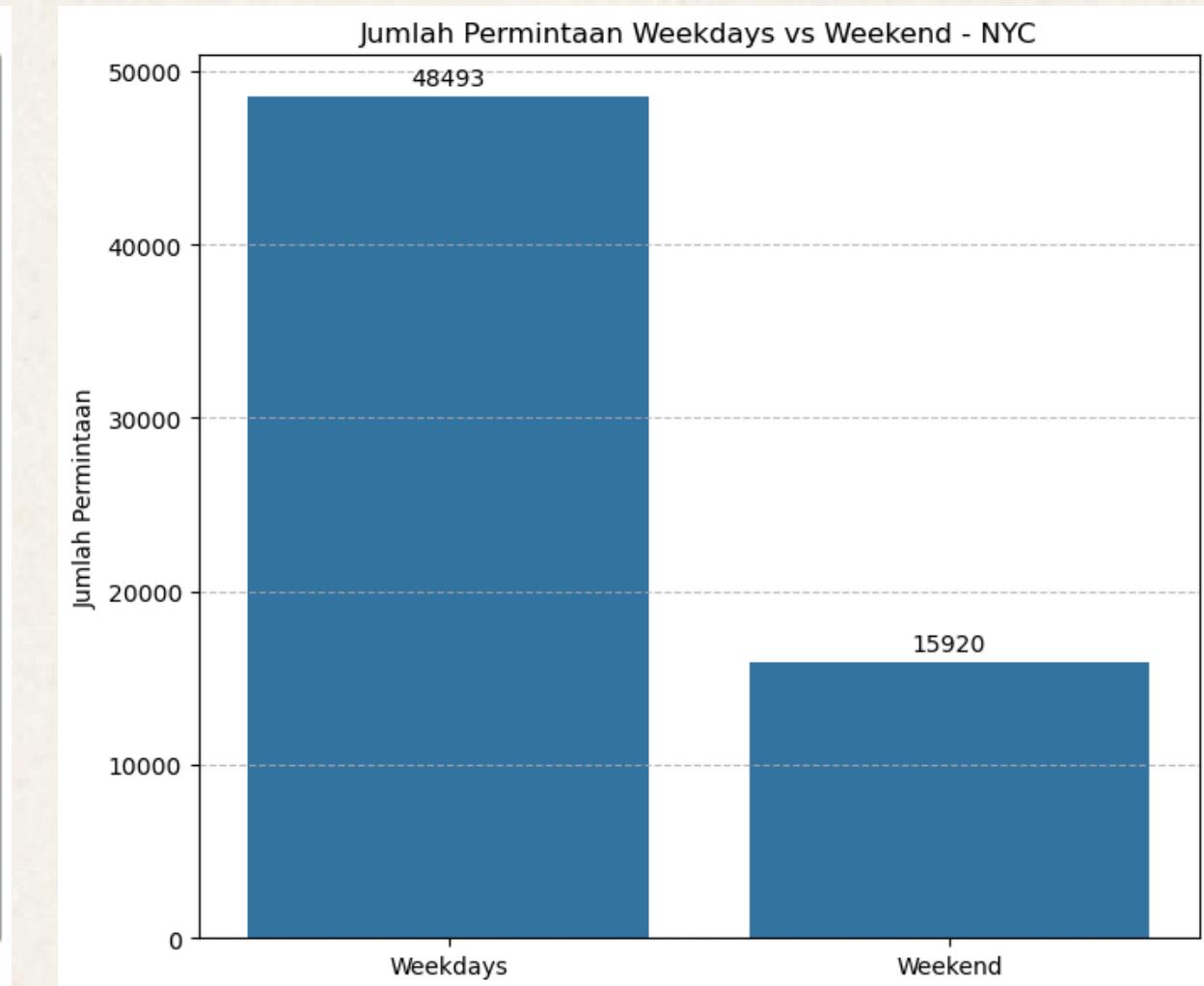


**Mencari waktu yang
memiliki permintaan tinggi**

1. Mencari waktu yang memiliki permintaan tinggi berdasarkan Hari

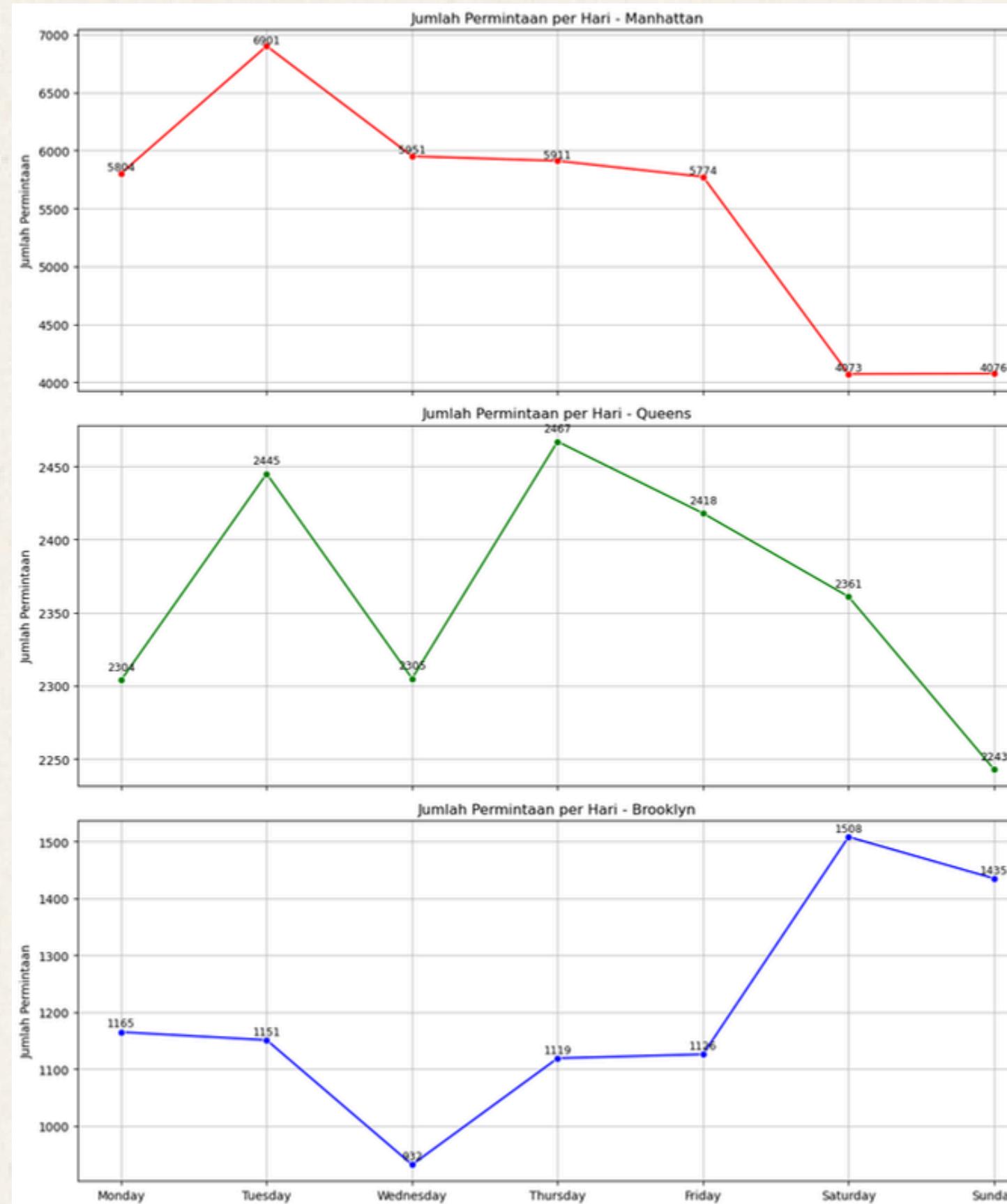


- Hari selasa puncak tertinggi permintaan
- Hari sabtu dan minggu cenderung mengalami penurunan permintaan



- Hari hari kerja (Weekdays) lebih ramai dibanding akhir pekan (Weekend)

1. Mencari waktu yang memiliki permintaan tinggi berdasarkan Hari



Wilayah Manhattan

Permintaan yang tinggi di hari Selasa
Permintaan yang rendah di hari Sabtu - Minggu

Wilayah Queens

Permintaan yang tinggi di hari Selasa dan Kamis
Permintaan yang rendah di hari Minggu

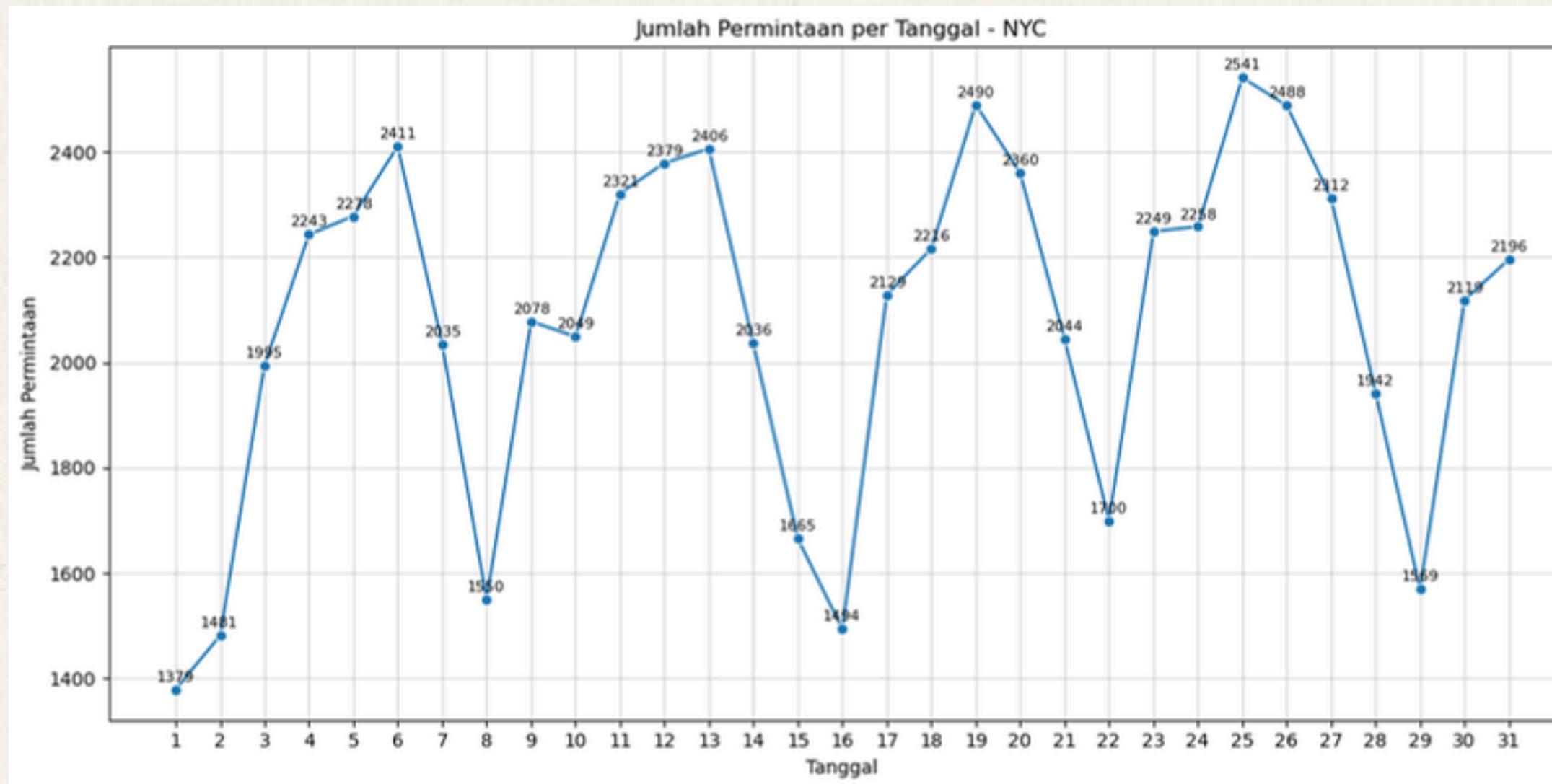
Wilayah Brooklyn

Permintaan yang tinggi di hari Sabtu dan Minggu
Permintaan yang rendah di hari Rabu

Strategi :

- Memfokuskan ketersediaan armada pada hari Weekdays untuk wilayah **Manhattan dan Queens**
- Untuk wilayah **Brooklyn** kita fokuskan ketersediaan armada pada hari Weekend dengan mengalihkan beberapa armada Weekend dari wilayah lain

2. Mencari waktu yang memiliki permintaan tinggi berdasarkan Tanggal



Calendar for January 2023 (United States)

Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

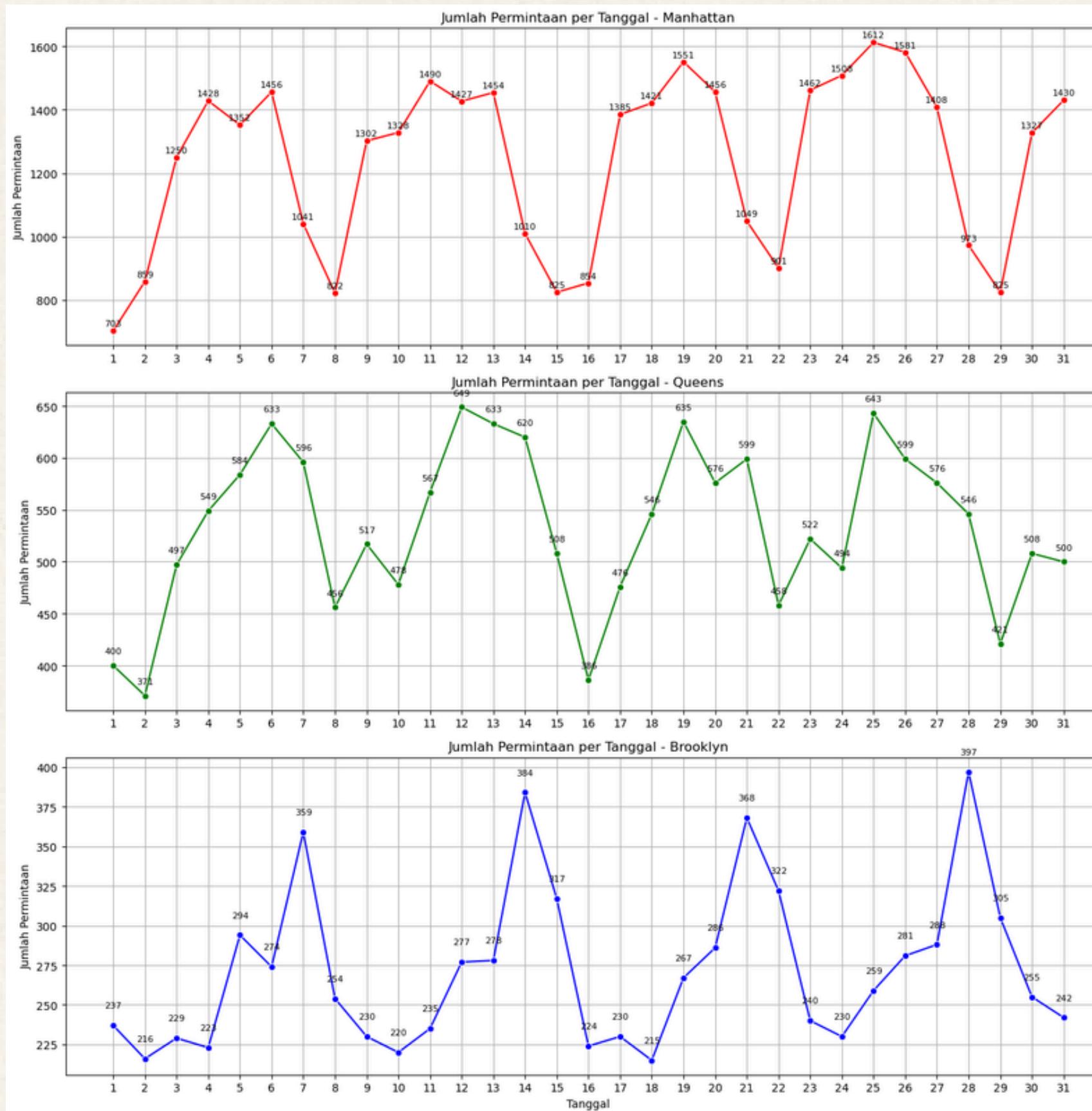
Phases of the Moon: 6:○ 14:○ 21:● 28:○

Holidays and Observances: 1: New Year's Day, 2: 'New Year's Day' day off, 16: Martin Luther King Jr. Day

- Tanggal 6, 13, 19, 25 puncak permintaan tertinggi (Weekdays)
- Tanggal 1 dan 2 → perayaan Tahun Baru
- Tanggal 16 → Martin Luther King Jr.

Maka kita fokuskan armada di tanggal selain libur

2. Mencari waktu yang memiliki permintaan tinggi berdasarkan Tanggal



Calendar for January 2023 (United States)						
Sun	Mon	Tue	Wednesday	Thu	Fri	Sat
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

Phases of the Moon: 6:○ 14:○ 21:● 28:○
Holidays and Observances: 1: New Year's Day, 2: 'New Year's Day' day off, 16: Martin Luther King Jr. Day

Wilayah Manhattan & Queens

Permintaan yang tinggi di tanggal 6, 12, 19, 25
Permintaan yang rendah di tanggal 8, 16, 22, 29

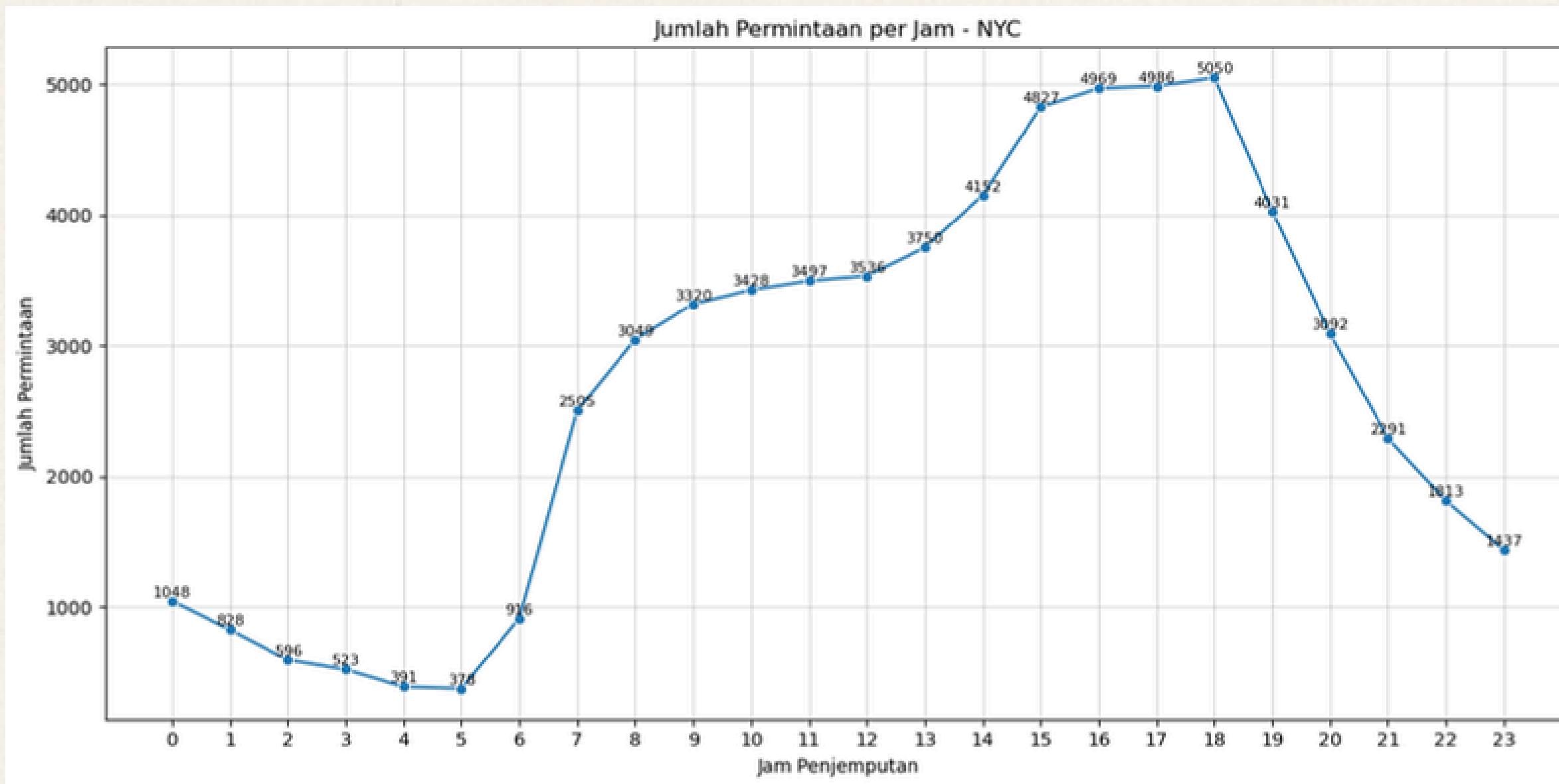
Wilayah Brooklyn

Permintaan yang tinggi di tanggal 7, 14, 21, 28
Permintaan yang rendah di tanggal 2, 10, 18, 24, 31

Strategi:

- Memfokuskan ketersediaan armada pada tanggal selain libur atau hari kerja (Manhattan & Queens)
- Untuk wilayah Brooklyn kita fokuskan ketersediaan armada pada akhir pekan mengalihkan beberapa armada dari wilayah lain

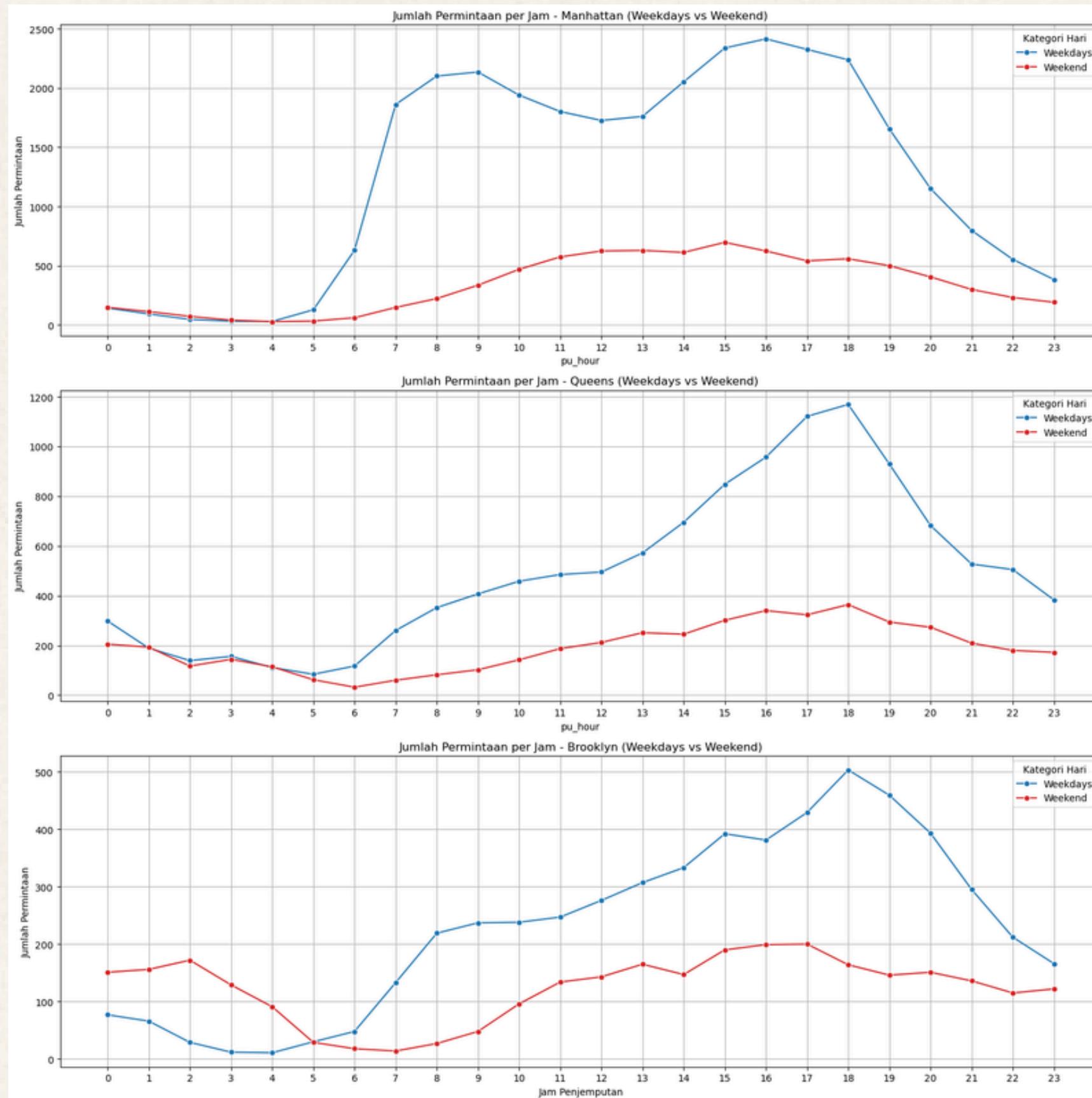
3. Mencari waktu yang memiliki permintaan tinggi berdasarkan Jam



- Permintaan **tertinggi** pada pukul 15:00 - 18:00
- Permintaan **yang mulai tinggi** pada pukul 07:00 - 09:00
- **Penurunan** permintaan terjadi pada pukul 19:00 - 05:00

Maka kita fokuskan armada pada jam - jam sibuk

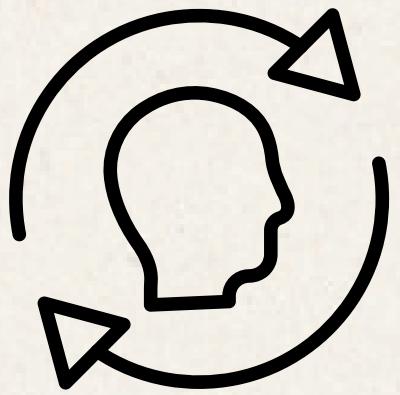
3. Mencari waktu yang memiliki permintaan tinggi berdasarkan Jam



- Permintaan tertinggi cenderung sama untuk ketiga wilayah
- Pukul 18:00 menjadi puncak permintaan tertinggi
- Pada wilayah **Manhattan** pukul 7:00 sudah terjadi lonjakan permintaan

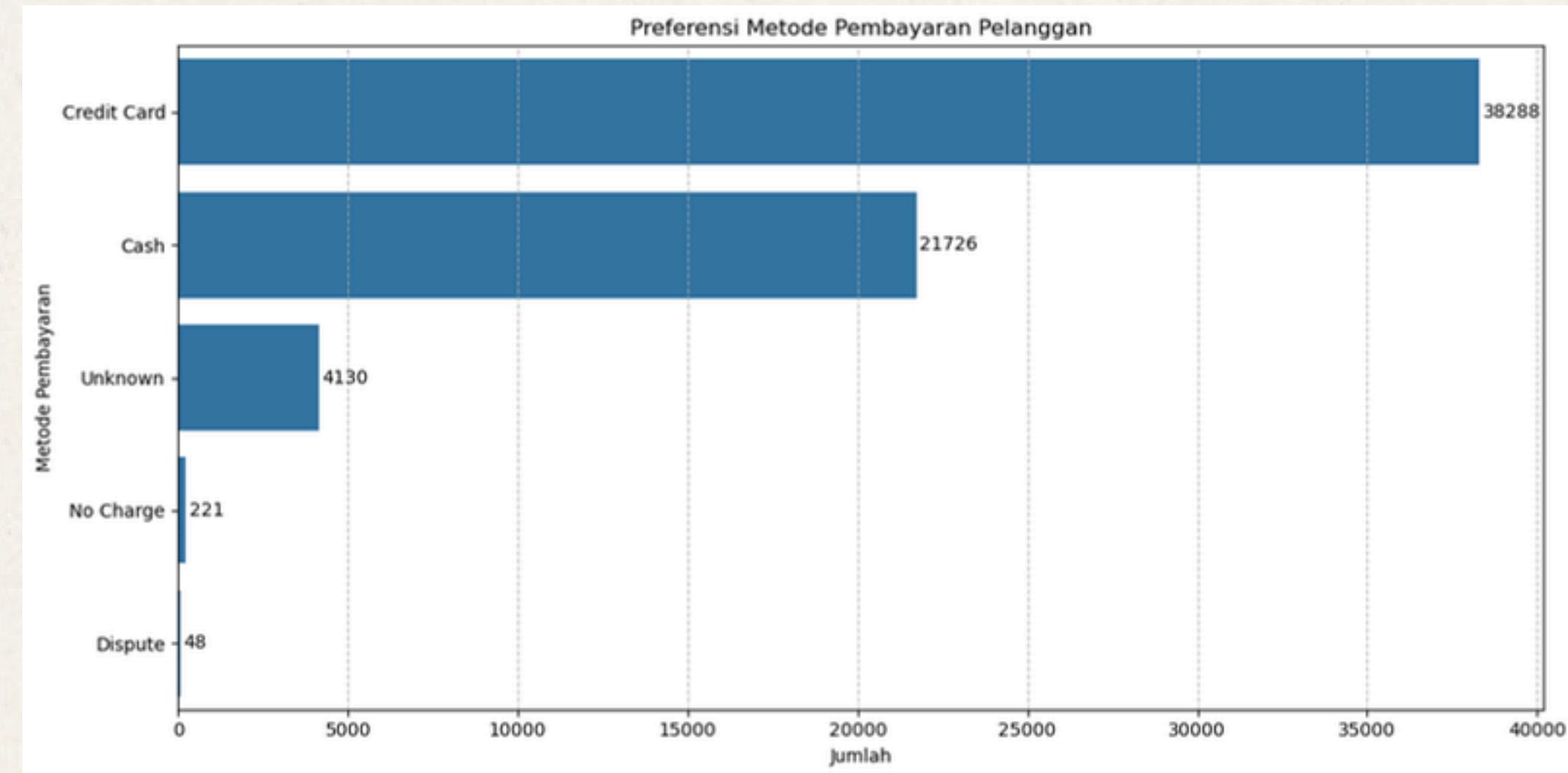
Strategi:

- Memfokuskan ketersediaan armada pada jam-jam sibuk



**Bagaimana preferensi / perilaku
pelanggan Green Taxi**

Jenis pembayaran yang sering digunakan pelanggan

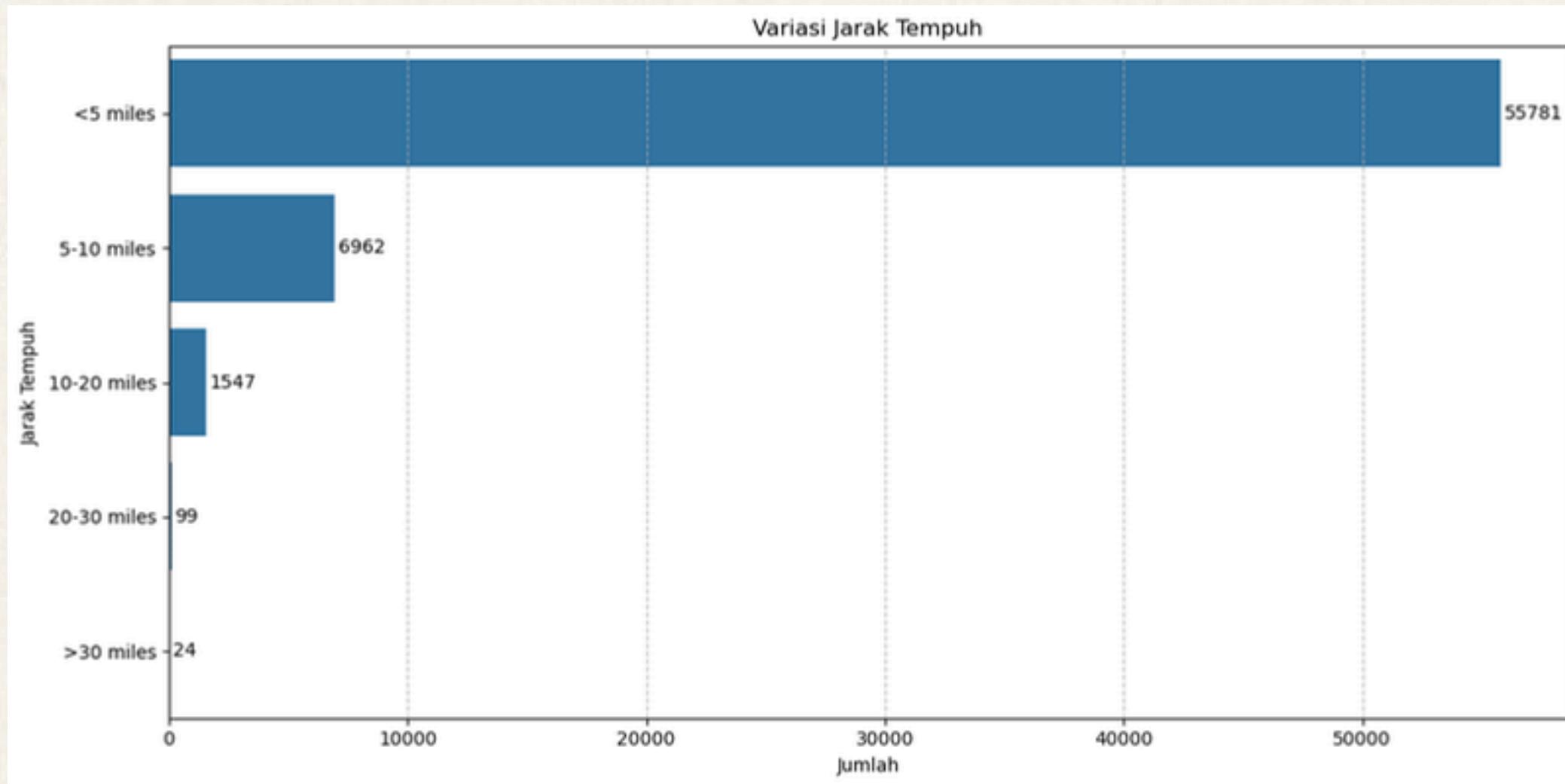


- Banyak pelanggan membayar dengan kartu kredit
- Beberapa pelanggan masih membayar dengan uang tunai

Strategi:

- Memastikan sistem pembayaran elektronik selalu tersedia, cepat, berjalan dengan baik
- Melakukan integrasi dari e-wallet lain

Jarak penggunaan taxi oleh pelanggan

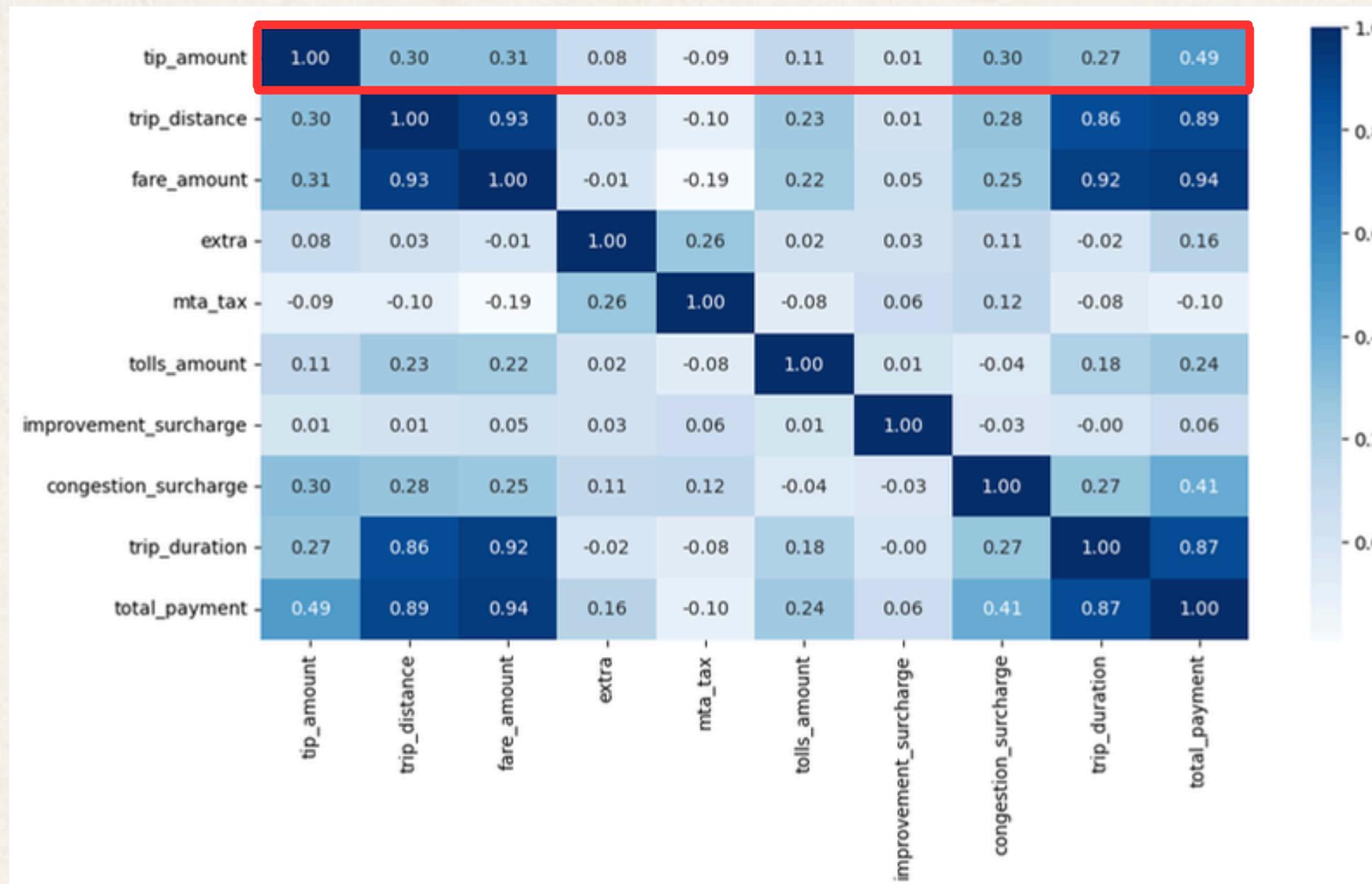


- Lebih banyak pelanggan menggunakan taxi untuk mobilitas < 5 mil
- Sangat jarang pelanggan menggunakan taxi > 10 mil

Strategi:

- Mencoba penerapan tarif flat untuk jarak tertentu < 5 mil
- Untuk menarik dan mempertahankan pelanggan setia Green Taxi

Bagaimana hubungan antar jumlah tip yang diberikan pelanggan ?



Nilai Korelasi	Tingkat Korelasi
0 - 0.19	Sangat Lemah
0.2 - 0.39	Lemah
0.4 - 0.59	Sedang
0.6 - 0.79	Kuat
0.8 - 1	Sangat Kuat

- Korelasi tertinggi (tip_amount) adalah dengan (total_payment) dengan tingkat korelasi **Sedang**
- Semakin besar (total_payment) biasanya tip yang diberikan semakin besar
- Korelasi **Lemah** (tip_amount) dengan (trip_distance) dan (congestion_surcharge)

Kesimpulan :

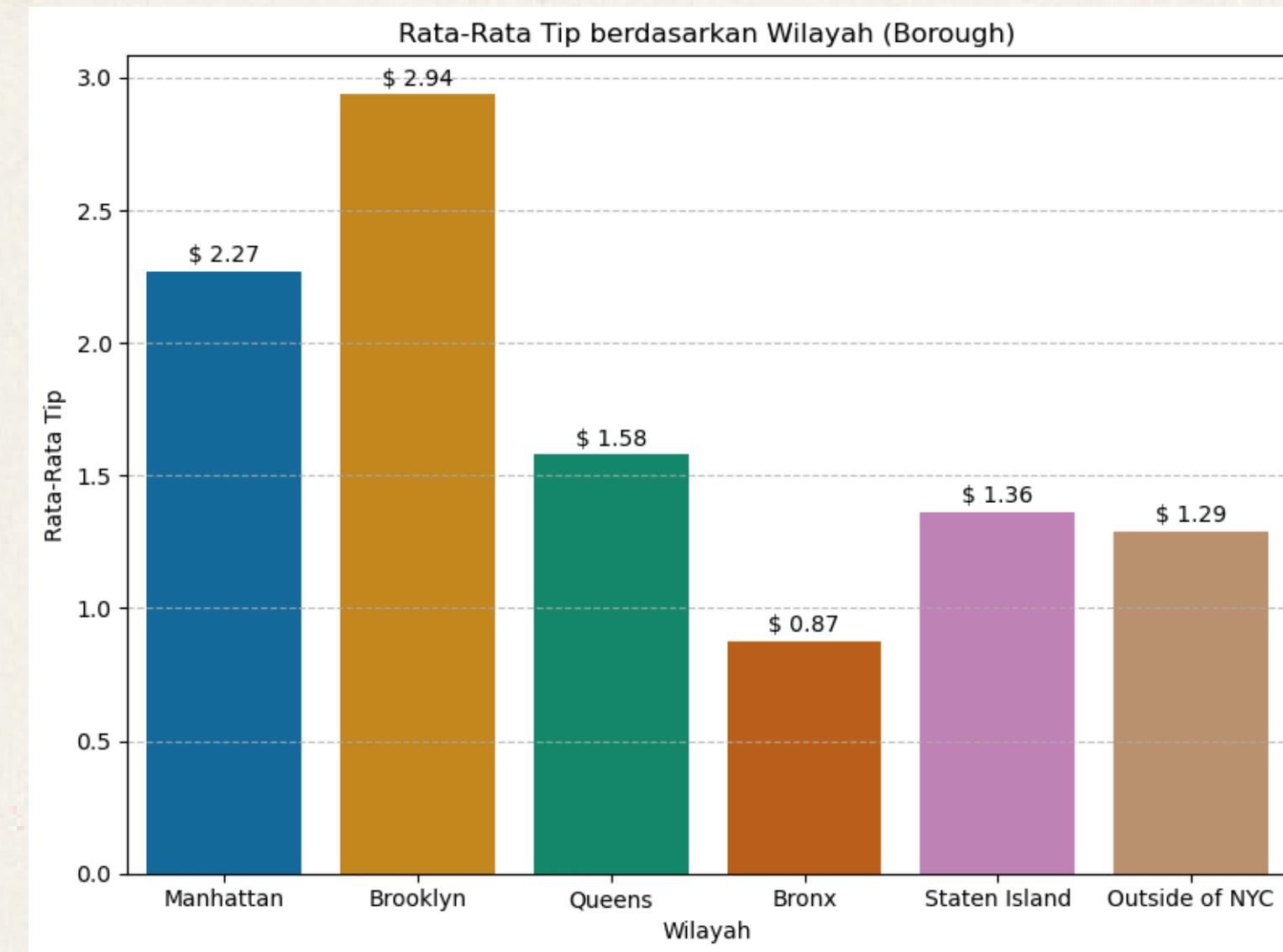
- Jumlah tip yang diberikan pelanggan bisa menjadi indikasi kepuasan pelanggan
- Namun data yang ada tidak dapat menjelaskan dengan jelas
- Hanya oleh (total_payment) (trip_distance) dan (trip_duration)
- Ada faktor lain yang mempengaruhi seperti : kenyamanan, keramahan driver dan kondisi kendaraan

Strategi :

- Membuat rating penilaian oleh pelanggan
- Membuat survei kepuasan pelanggan

Bagaimana jumlah tip yang diberikan pelanggan berdasarkan wilayah ?

1. Uji Statistik Kruskal Wallis
 - Berhasil membuktikan adanya perbedaan rata rata tip antar wilayah
2. Uji Statistik Mann-Whitney U
 - Berhasil membuktikan bahwa rata rata tip wilayah Brooklyn lebih besar dari Manhattan



- Wilayah dengan rata rata tip tertinggi → Brooklyn
- Brooklyn dan Manhattan rata rata tip > \$2

Strategi:

- Tetap tingkatkan pelayanan disemua wilayah
- Walaupun rata rata tip Manhattan lebih kecil dari Brooklyn, manfaatkan volume permintaan yang tinggi dan tidak hanya berfokus pada tip

Kesimpulan

No	Kesimpulan	Rekomendasi
1	Wilayah dengan permintaan tinggi Manhattan, Queens, Brooklyn	Mengoptimalkan armada di Manhattan terutama pada jam sibuk, alihkan zona rendah ke zona tinggi permintaan
2	Zona dengan permintaan tertinggi berada pada wilayah Manhattan	
3	Manhattan & Queens lebih ramai saat Weekdays Brooklyn lebih ramai saat Weekend	Mengalihkan fokus beberapa armada ke Brooklyn pada akhir pekan
4	Penurunan permintaan pada tanggal merah/libur	Mengurangi jumlah armada yang aktif di hari libur untuk efisiensi operasional
5	Jam sibuk Pagi dan Sore mendominasi permintaan	Menyediakan armada lebih responsif pada jam sibuk agar supply selalu tersedia
6	Kartu kredit lebih banyak digunakan	<ul style="list-style-type: none">• Memastikan fungsi pembayaran elektronik berjalan baik• Kerja sama dengan penyedia pembayaran elektronik
7	Penggunaan taxi lebih banyak untuk jarak <5 mil	<ul style="list-style-type: none">• Menerapkan tarif flat perjalanan <5 mil• Perkuat layanan area Central Business District yaitu Manhattan
8	Besaran tip sebagai indikasi kepuasan pelanggan kurang menjelaskan	Membuat survei kepuasan pelanggan untuk evaluasi kinerja
9	Adanya faktor lain yang mempengaruhi besaran tip	Meningkatkan keramahan driver, menjaga kebersihan kendaraan
10	Setiap wilayah memiliki besaran tip yang berbeda	Mengoptimalkan pelayanan disetiap wilayah, jangan terfokus dengan tip, tetapi manfaatkan volume tinggi

Terima Kasih!

Saya sangat mengharapkan saran, masukan, maupun pertanyaan dari anda untuk solusi yang lebih baik.

E-mail yanfaanandika21@gmail.com

LinkedIn linkedin.com/in/yanfa-anandika

GitHub github.com/yanfa121
