

By : Fenny Anggraini

# MINI PROJECT

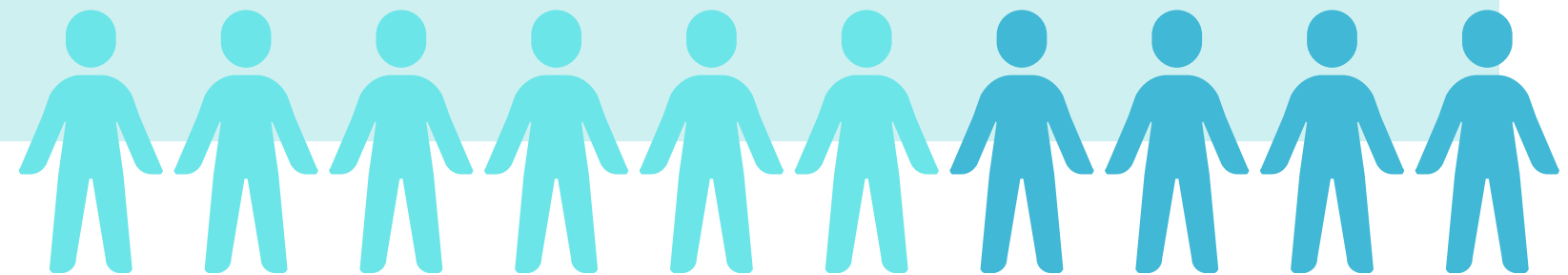
# DATA ENGINEER

CHICAGO TAXI TRIPS



# DATASET

Dataset Chicago Taxi Trips mencakup informasi lengkap tentang perjalanan taksi di Chicago, termasuk waktu mulai dan selesai perjalanan, durasi, jarak tempuh, lokasi penjemputan dan penurunan, biaya perjalanan (termasuk tarif, tips, dan tol), metode pembayaran, serta perusahaan taksi. Data ini digunakan untuk menganalisis pola perjalanan, mengoptimalkan operasional taksi, serta memahami tren transportasi dan perilaku konsumen di kota tersebut.



# DATASET OVERVIEW

- **unique\_key** : Kunci unik yang mengidentifikasi setiap perjalanan taksi. Membuat setiap entri perjalanan berbeda satu sama lain.
- **taxi\_id** : Identifikasi unik untuk taksi yang melakukan perjalanan. Melacak taksi mana yang melakukan perjalanan tertentu.
- **trip\_start\_timestamp** : Waktu mulai perjalanan, dibulatkan ke 15 menit terdekat. Menentukan kapan perjalanan dimulai.
- **trip\_end\_timestamp** : Waktu selesai perjalanan, dibulatkan ke 15 menit terdekat. Menentukan kapan perjalanan berakhir.
- **trip\_seconds** : Durasi perjalanan dalam detik. Mengukur lama waktu perjalanan berlangsung.
- **trip\_miles** : Jarak perjalanan dalam mil. Menunjukkan jarak yang ditempuh selama perjalanan.
- **pickup\_census\_tract** : Kode Census Tract untuk lokasi penjemputan, disembunyikan untuk privasi dalam beberapa perjalanan.
- **dropoff\_census\_tract** : Kode Census Tract untuk lokasi penurunan, disembunyikan untuk privasi dalam beberapa perjalanan.
- **pickup\_community\_area** : Kode Community Area tempat penjemputan terjadi.
- **dropoff\_community\_area** : Kode Community Area tempat penurunan terjadi.
- **fare** : Menunjukkan biaya awal yang dikenakan untuk perjalanan.
- **Tolls** : Biaya tol yang dibayar selama perjalanan.
- **Extras** : Biaya tambahan lain yang terkait dengan perjalanan (misalnya, biaya tambahan untuk layanan tertentu).
- **trip\_total** : Total biaya perjalanan, termasuk tarif dasar, tip, tol, dan biaya tambahan.
- **payment\_type** : Metode pembayaran yang digunakan (misalnya, tunai, kartu kredit).
- **Company** : Nama perusahaan taksi yang menyediakan layanan.
- **pickup\_latitude** : Latitude (garis lintang) lokasi penjemputan.
- **pickup\_longitude** : Longitude (garis bujur) lokasi penjemputan.
- **pickup\_location** : Lokasi pusat dari Census Tract atau Community Area untuk penjemputan (jika data lebih detail disembunyikan).
- **dropoff\_latitude** : Latitude (garis lintang) lokasi penurunan.
- **dropoff\_longitude** : Longitude (garis bujur) lokasi penurunan.
- **dropoff\_location** : Lokasi pusat dari Census Tract atau Community Area untuk penurunan (jika data lebih detail disembunyikan).

# ANALISIS 1

Hitung rata-rata, median, dan standar deviasi dari durasi perjalanan (trip\_seconds) untuk perjalanan yang dilakukan pada hari Senin dan Sabtu. Bandingkan hasil kedua hari tersebut.

Query ini menganalisis dataset perjalanan taksi di Chicago untuk membandingkan durasi perjalanan pada hari Senin dan Sabtu. Menggunakan Common Table Expression (CTE) bernama `trip\_data`, query ini mengekstrak hari dalam seminggu dari timestamp perjalanan dan durasi perjalanan dalam detik, hanya untuk hari-hari yang relevan (Senin dan Sabtu). Selanjutnya, query menghitung jumlah perjalanan, rata-rata durasi, median durasi, dan standar deviasi durasi untuk masing-masing hari, serta mengonversi angka hari ke nama hari ('Monday' dan 'Saturday'). Hasilnya dikelompokkan berdasarkan hari dan diurutkan, memberikan wawasan tentang pola perjalanan taksi di Chicago selama dua hari tersebut.



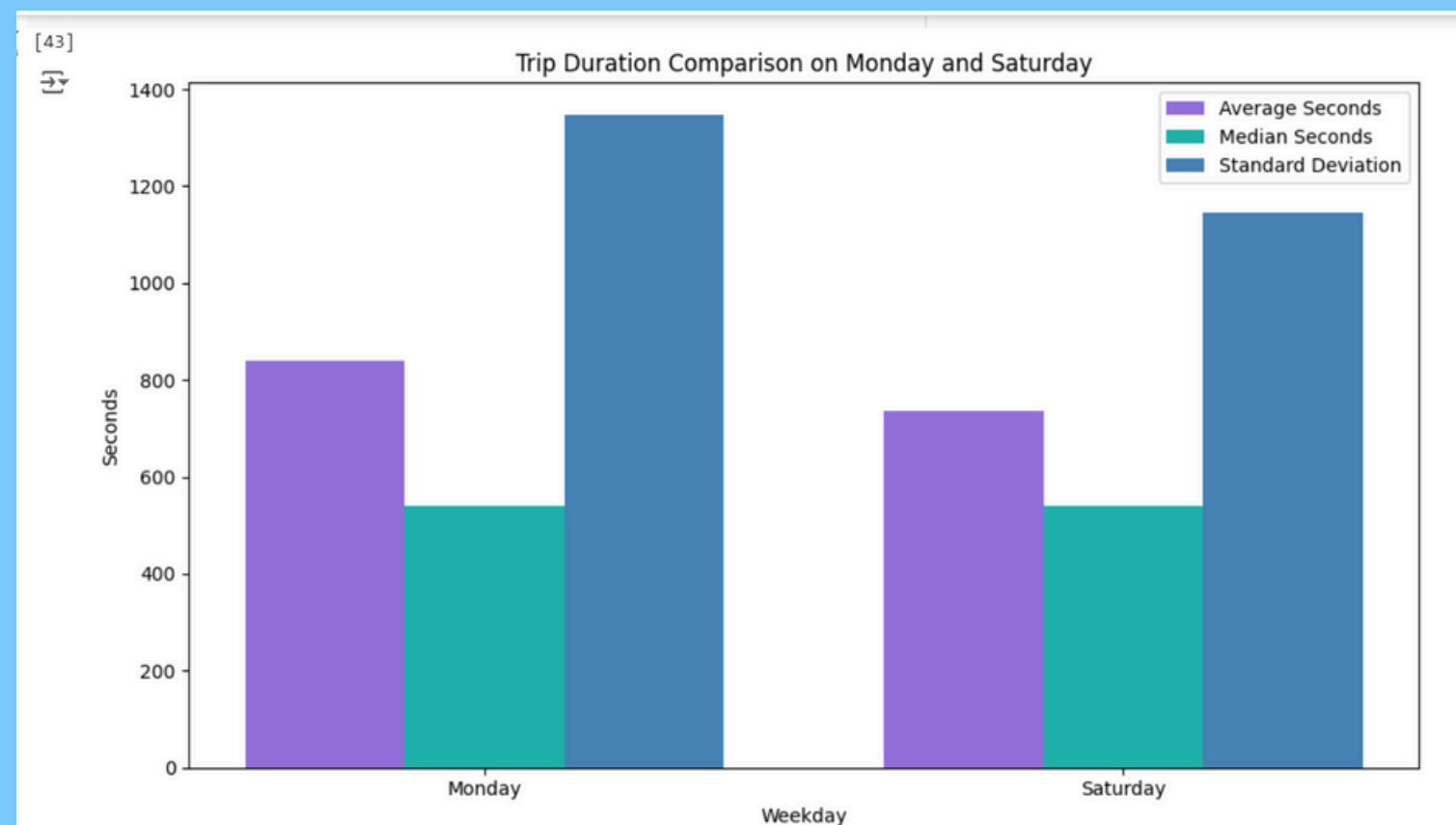
## PROGRAM

```
query = """
WITH trip_data AS (
SELECT
EXTRACT(DAYOFWEEK FROM trip_start_timestamp) AS day_of_week,
trip_seconds
FROM
`bigquery-public-data.chicago_taxi_trips.taxi_trips`
WHERE
EXTRACT(DAYOFWEEK FROM trip_start_timestamp) IN (2, 7) -- 2 = Monday, 7 = Saturday
)
SELECT
CASE
WHEN day_of_week = 2 THEN 'Monday'
WHEN day_of_week = 7 THEN 'Saturday'
END AS day_name,
COUNT(*) AS num_trips,
AVG(trip_seconds) AS avg_trip_seconds,
APPROX_QUANTILES(trip_seconds, 2)[OFFSET(1)] AS median_trip_seconds,
STDDEV(trip_seconds) AS stddev_trip_seconds
FROM
trip_data
GROUP BY
day_of_week
ORDER BY
day_of_week;
"""

df1 = client.query(query).to_dataframe()
```

# ANALISIS 1

	day_name	num_trips	avg_trip_seconds	median_trip_seconds	stddev_trip_seconds
0	Monday	27315537	839.479784	540	1346.060208
1	Saturday	30678201	737.410552	540	1146.335888



Analisis hasil menunjukkan bahwa pada hari Sabtu terdapat 30.678.201 perjalanan taxi, lebih banyak dibandingkan dengan 27.315.537 perjalanan pada hari Senin, yang mencerminkan tingginya permintaan taxi untuk aktivitas sosial di akhir pekan. Rata-rata durasi perjalanan lebih singkat pada hari Sabtu (737,41 detik) dibandingkan dengan Senin (839,48 detik), mengindikasikan efisiensi perjalanan yang lebih baik pada hari tersebut. Meskipun median durasi perjalanan sama untuk kedua hari (540 detik), standar deviasi menunjukkan bahwa durasi perjalanan pada hari Senin lebih bervariasi (1.346,06 detik) dibandingkan dengan hari Sabtu (1.146,34 detik), yang mungkin disebabkan oleh kemacetan lalu lintas yang lebih tinggi dan ketidakpastian waktu perjalanan pada hari kerja. Secara keseluruhan, data ini menyoroti pola penggunaan taxi yang berbeda antara hari kerja dan akhir pekan, dengan dampak signifikan pada durasi dan permintaan perjalanan.



# ANALISIS 2



Temukan lima rute (dari community\_area awal ke community\_area tujuan) dengan jumlah perjalanan terbanyak pada tahun 2023.

Query ini menganalisis data perjalanan taksi di Chicago untuk tahun 2023 dengan tujuan untuk mengidentifikasi lima kombinasi area komunitas (pickup dan dropoff) yang paling banyak digunakan. Pertama, query memilih kolom pickup\_community\_area dan dropoff\_community\_area, serta menghitung jumlah perjalanan taksi (num\_trips) yang dilakukan dari masing-masing kombinasi area komunitas. Kondisi 'WHERE' memastikan bahwa kedua area komunitas tidak null dan hanya mempertimbangkan perjalanan yang terjadi pada tahun 2023. Data kemudian dikelompokkan berdasarkan kombinasi area komunitas yang diambil dan dijatuhkan, diurutkan berdasarkan jumlah perjalanan secara menurun. Akhirnya, query membatasi hasil hanya pada lima kombinasi teratas, memberikan wawasan tentang rute perjalanan taksi paling populer di Chicago pada tahun tersebut.

## PROGRAM

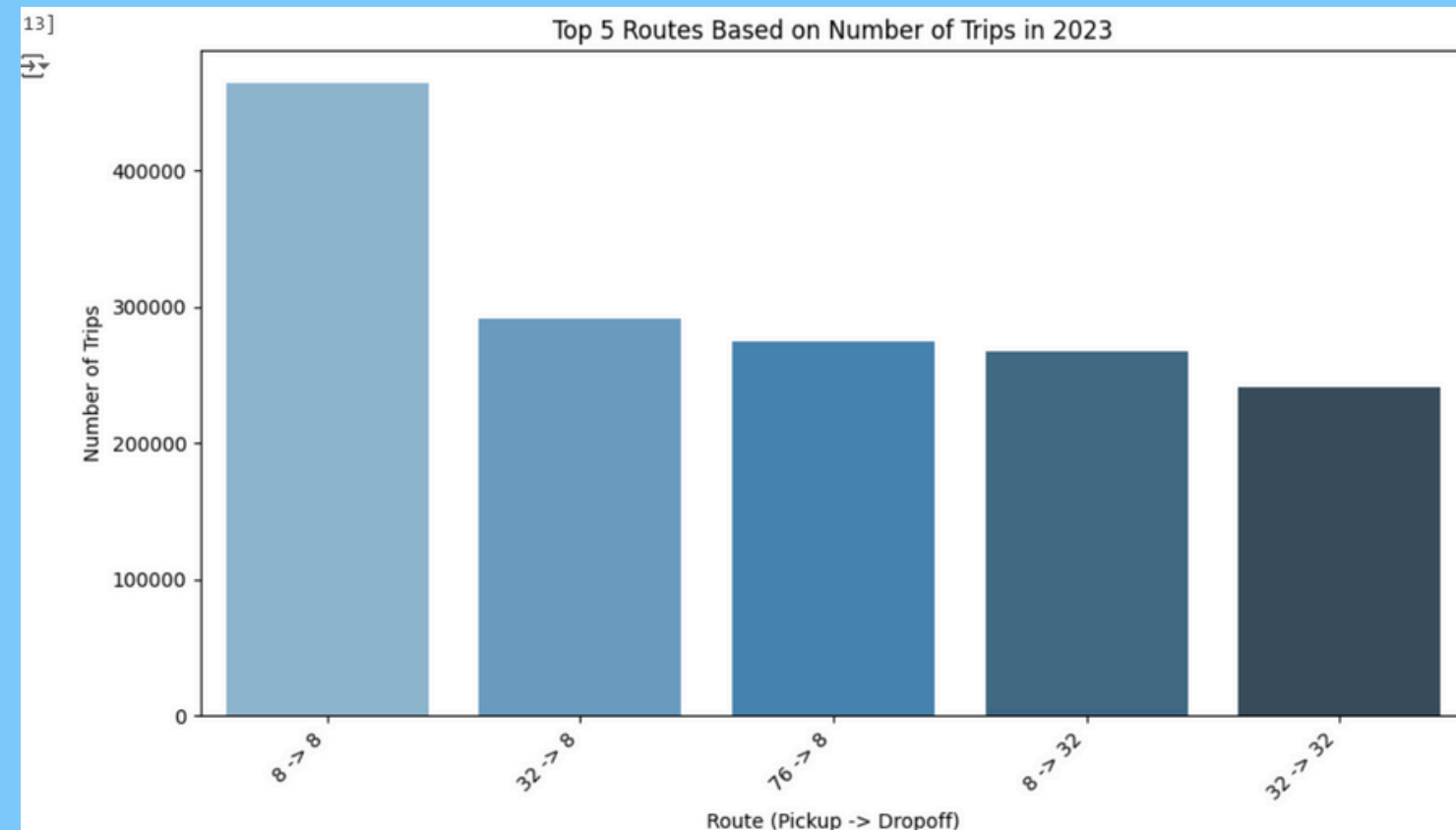
```
query = """
SELECT
    pickup_community_area,
    dropoff_community_area,
    COUNT(*) AS num_trips
FROM
    bigquery-public-data.chicago_taxi_trips.taxi_trips
WHERE
    pickup_community_area IS NOT NULL
    AND dropoff_community_area IS NOT NULL
    AND EXTRACT(year FROM trip_start_timestamp) = 2023
GROUP BY
    pickup_community_area,
    dropoff_community_area
ORDER BY
    num_trips DESC
LIMIT
    5;
"""
df2 = client.query(query).to_dataframe()
```



# ANALISIS 2



	pickup_community_area	dropoff_community_area	num_trips
0	8	8	464844
1	32	8	291722
2	76	8	274747
3	8	32	267673
4	32	32	241596



Hasil analisis menunjukkan bahwa kombinasi area komunitas 8 ke area komunitas 8 adalah yang paling banyak digunakan dengan total 464.844 perjalanan taksi, mencerminkan aktivitas tinggi di area tersebut, mungkin karena adanya pusat kegiatan atau fasilitas penting. Area komunitas 32 ke area komunitas 8 menempati posisi kedua dengan 291.722 perjalanan, menunjukkan konektivitas yang signifikan antara keduanya. Selanjutnya, perjalanan dari area komunitas 76 ke area komunitas 8 mencatat 274.747 perjalanan, menyoroti pentingnya area 8 sebagai tujuan utama. Kombinasi perjalanan dari area komunitas 8 ke area komunitas 32 (267.673 perjalanan) dan dari area komunitas 32 ke area komunitas 32 (241.596 perjalanan) juga menunjukkan tingginya permintaan di kedua komunitas tersebut. Secara keseluruhan, hasil ini mencerminkan pola perjalanan yang kuat di sekitar area komunitas 8 dan 32, menandakan bahwa kedua area tersebut mungkin merupakan pusat mobilitas yang penting dalam jaringan transportasi taksi di Chicago.

# ANALISIS 3



Bandingkan biaya rata-rata perjalanan taksi (fare, tips, dan taxes) berdasarkan metode pembayaran pada tahun 2019.

Query ini menganalisis data perjalanan taksi di Chicago untuk tahun 2019 dengan fokus pada jenis pembayaran yang digunakan oleh penumpang. Query ini memilih kolom `payment_type` dan menghitung rata-rata tarif perjalanan (`average_fare`), rata-rata tips (`average_tips`), dan rata-rata tol (`average_tolls`) untuk masing-masing jenis pembayaran. Kondisi 'WHERE' memastikan bahwa hanya perjalanan dari tahun 2019 yang diperhitungkan. Data kemudian dikelompokkan berdasarkan `payment_type` untuk menghitung rata-rata secara terpisah untuk setiap jenis pembayaran yang digunakan. Hasil akhir diurutkan berdasarkan `average_fare` secara menurun, sehingga jenis pembayaran dengan tarif rata-rata tertinggi akan muncul terlebih dahulu. Dengan demikian, query ini memberikan wawasan tentang bagaimana tarif, tips, dan tol bervariasi berdasarkan metode pembayaran yang dipilih oleh penumpang taksi.

## PROGRAM

```
query = """
SELECT
    payment_type,
    AVG(fare) AS average_fare,
    AVG(tips) AS average_tips,
    AVG(tolls) AS average_tolls
FROM
    bigquery-public-data.chicago_taxi_trips.taxi_trips
WHERE
    EXTRACT(year FROM trip_start_timestamp) = 2019
GROUP BY
    payment_type
ORDER BY
    average_fare DESC;
"""

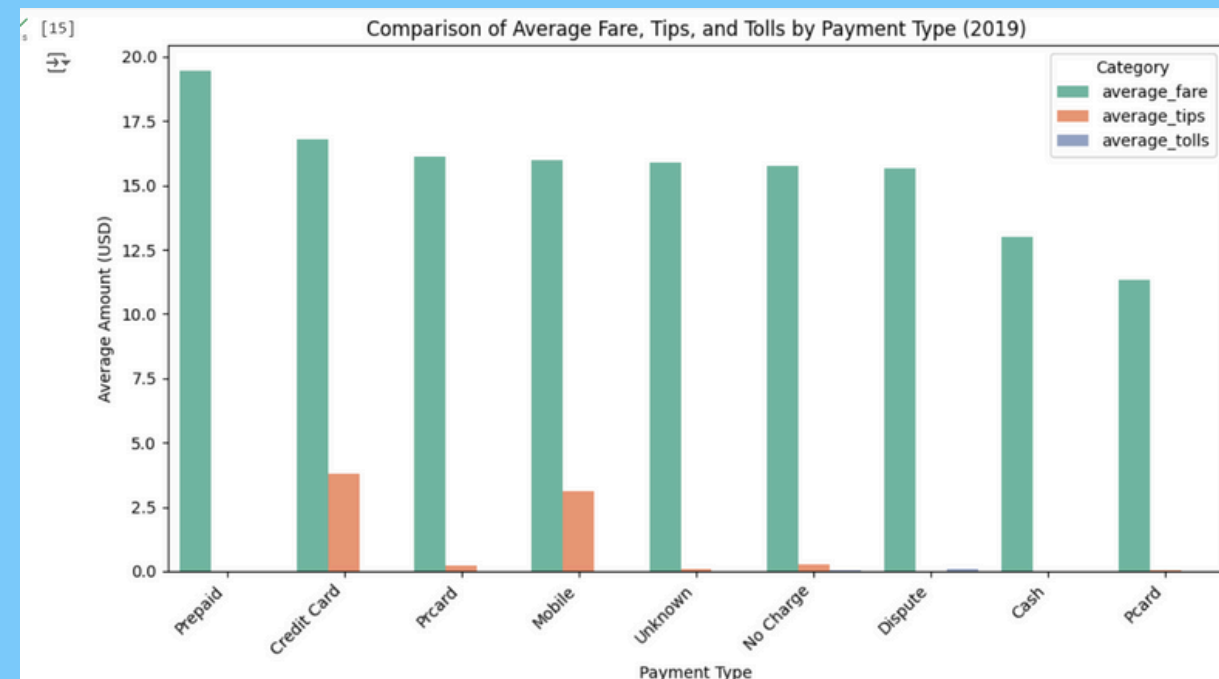
df3 = client.query(query).to_dataframe()
```



# ANALISIS 3



	payment_type	average_fare	average_tips	average_tolls
0	Prepaid	19.461415	0.000000	0.000000
1	Credit Card	16.814138	3.774545	0.002267
2	Prcard	16.130003	0.203528	0.001529
3	Mobile	15.973174	3.112698	0.000050
4	Unknown	15.872380	0.082129	0.000147
5	No Charge	15.753510	0.249532	0.019142
6	Dispute	15.655840	0.001446	0.078991
7	Cash	12.978306	0.002387	0.001791
8	Pcard	11.327471	0.038023	0.000000



Tabel tersebut menunjukkan rata-rata tarif (average\_fare), tips (average\_tips), dan tol (average\_tolls) untuk setiap jenis pembayaran (payment\_type) berdasarkan data perjalanan taksi di Chicago pada tahun 2019.

- Prepaid memiliki rata-rata tarif tertinggi sebesar \$19,46, tetapi tidak ada tips dan tol yang dilaporkan.
- Credit Card dan Mobile menunjukkan tarif yang kompetitif dengan rata-rata tips yang signifikan, masing-masing sebesar \$3,77 dan \$3,11, serta biaya tol yang minimal.
- Prcard memiliki tarif dan tips yang lebih rendah, tetapi tetap mencerminkan pola pemberian tips.
- Dispute menunjukkan rata-rata tol yang cukup tinggi dibandingkan dengan jenis pembayaran lainnya, dengan \$0,08, mungkin menunjukkan ketidakpuasan pelanggan.
- Cash memiliki rata-rata tarif, tips, dan tol yang lebih rendah dibandingkan metode pembayaran non-tunai, menunjukkan bahwa penumpang yang membayar dengan tunai cenderung memberi sedikit atau tidak memberi tips sama sekali.

# REKOMENDASI BISNIS

- Promosi dan Penawaran Khusus untuk Pengguna Taksi di Akhir Pekan: Mengingat tingginya permintaan perjalanan taksi pada hari Sabtu, perusahaan dapat mempertimbangkan untuk menawarkan promosi khusus, seperti diskon tarif atau program loyalitas, untuk menarik lebih banyak pelanggan di akhir pekan.
- Optimalisasi Layanan di Area Komunitas Strategis: Dengan data yang menunjukkan tingginya permintaan perjalanan antara area komunitas 8 dan 32, perusahaan harus meningkatkan frekuensi dan ketersediaan taksi di area ini. Ini dapat mencakup penempatan taksi lebih banyak di lokasi-lokasi strategis atau penawaran khusus untuk perjalanan antara kedua area tersebut.
- Peningkatan Fasilitas Pembayaran: Meningkatkan pengalaman pembayaran untuk pengguna kartu kredit dan mobile dapat berpotensi meningkatkan rata-rata tarif dan tips. Perusahaan dapat mempertimbangkan untuk memperkenalkan fitur baru dalam aplikasi pembayaran yang memberikan insentif bagi pelanggan untuk memberi tips.
- Pendidikan Pelanggan tentang Pemberian Tips: Mengingat rendahnya tingkat tips untuk pembayaran tunai, perusahaan bisa menjalankan kampanye edukasi untuk meningkatkan kesadaran pelanggan tentang pentingnya memberi tips, yang dapat membantu meningkatkan pendapatan pengemudi.

# KESIMPULAN

Analisis data perjalanan taksi di Chicago memberikan wawasan berharga tentang pola penggunaan taksi berdasarkan waktu dan lokasi. Terdapat perbedaan yang signifikan dalam permintaan dan durasi perjalanan antara hari kerja dan akhir pekan, dengan peningkatan penggunaan taksi yang terlihat pada hari Sabtu. Area komunitas 8 dan 32 menunjukkan pola perjalanan yang kuat, yang menandakan pentingnya area tersebut sebagai pusat mobilitas. Selain itu, variasi dalam tarif dan tips berdasarkan metode pembayaran menunjukkan peluang untuk meningkatkan pengalaman pelanggan dan pendapatan. Dengan menerapkan rekomendasi bisnis yang ditawarkan, perusahaan taksi dapat lebih baik memenuhi kebutuhan pelanggan, meningkatkan kepuasan, dan pada akhirnya meningkatkan pendapatan mereka.

