

ADVANCED BIG DATA ANALYTICS - LAB Course

Di Rancang Oleh:

Lola Naomi Enzelin Mangunsong

[<https://github.com/LolaNaomi/Advance-Big-Data-Project-Case>]

TAXI SERVICES / NYC Green Taxi Siders Company

*Apa itu **Green Taxi** di New York City?*

Pada musim panas tahun 2013, Kota New York membuat program untuk mengurangi kesenjangan layanan taksi di wilayah yang jarang dikunjungi di NYC — **Harlem, Queens, Bronx, dan Brooklyn** (serta bagian utara Manhattan di atas East 96th Street dan West 110th Street). Program baru ini meluncurkan **Green Taxi** di NYC yang secara resmi dikenal sebagai Boro Taxis.

Namun, taksi hijau **boleh** masuk Manhattan untuk **menurunkan penumpang**, tetapi **tidak boleh mengambil penumpang** di jalan di daerah pusat Manhattan (di bawah batasan East 96th Street dan West 110th Street). Penumpang di Manhattan di luar batas ini bisa dijemput. Jadi, meskipun taksi hijau dapat melewati Manhattan atau menurunkan penumpang di sana, aturan utamanya adalah mereka **tidak diperbolehkan mengambil penumpang di pusat Manhattan** dari East 96th Street ke bawah.

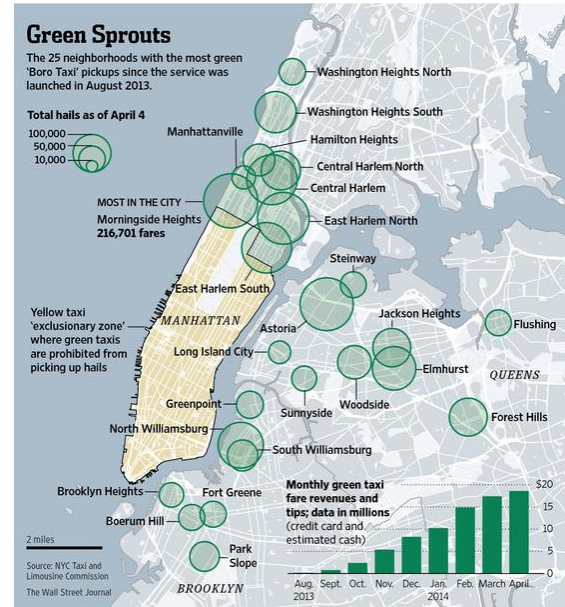


Fig 1. Wilayah Pickup Passenger Green Taxi |
Sumber: *Astoria Loves Those Green Taxis*

Dataset Source :

[TLC Trip Record Data - TLC](#) | [zone coordinates of NYC Yellow Taxi Trip Records](#) (Untuk Detil Lokasi dari ID Zona Pickup & Zona DropOff)

Link PBI : https://multimediasantara-my.sharepoint.com/:u:/g/personal/lola_naomi_student_umn_ac_id/EZREBXyvAKtNvIwlG2BgHRsBunLWHVSSmsyL978bPI9foQ?e=NSFQTK

Link SAS : <https://vfl-036.engage.sas.com/links/resources/report?uri=%2Freports%2Freports%2F04369c7b-dc67-4cce-af1e-de2db2291a44>

Deskripsi Dataset :

[Bulan Januari – Juli, Tahun 2024]

No	Attribut	Penjelasan
1	VendorID	ID dari vendor taksi yang menyediakan layanan. Biasanya memiliki nilai 1 (Creative Mobile Technologies, LLC) atau 2 (VeriFone Inc.) yang mengindikasikan vendor yang berbeda.
2	lpep_pickup_datetime	Tanggal dan waktu saat penumpang dijemput oleh Green Taxi.
3	lpep_dropoff_datetime	Tanggal dan waktu ketika penumpang diantar ke tujuan.

4	store_and_fwd_flag	Flag untuk menunjukkan apakah data perjalanan disimpan sementara sebelum dikirimkan ke server. Nilainya bisa 'Y' (Yes) atau 'N' (No).
5	RatecodeID	Kode tarif yang menggambarkan jenis perjalanan. Beberapa nilai yang mungkin termasuk: <ul style="list-style-type: none"> • 1: Standard Rate (Tarif biasa) • 2: Negotiated Fare (Tarif dinegosiasikan) • 3: Group Ride • 4: Rate untuk wilayah Westchester atau Nassau. • 5: Tarif tetap untuk perjalanan ke luar kota.
6	PULocationID	ID lokasi penjemputan. Ini adalah kode yang mengindikasikan zona di mana penjemputan dilakukan (sesuai dengan zona taksi di NYC).
7	DOLocationID	ID lokasi pengantaran. Kode ini menggambarkan zona pengantaran sesuai sistem zona taksi di NYC.
8	passenger_count	Jumlah penumpang dalam perjalanan taksi.
9	trip_distance	Jarak perjalanan dalam mil. Ini menggambarkan jarak dari penjemputan ke pengantaran.
10	fare_amount	Jumlah tarif utama yang dikenakan untuk perjalanan (tidak termasuk biaya tambahan seperti pajak atau tol).
11	extra	Biaya tambahan untuk perjalanan, misalnya biaya tambahan malam hari atau cuaca buruk.
12	mta_tax	Pajak yang dikenakan oleh Metropolitan Transportation Authority (MTA), umumnya bernilai tetap.
13	tip_amount	Jumlah tip yang diberikan oleh penumpang (jika pembayaran dilakukan menggunakan kartu kredit).
14	tolls_amount	Jumlah biaya tol yang dibebankan selama perjalanan (jika melalui jalan tol).
15	ehail_fee	ehail_fee
16	improvement_surcharge	Biaya perbaikan taksi yang dikenakan oleh NYC, umumnya tetap sebesar \$0.30 per perjalanan.
17	total_amount	Total biaya perjalanan yang harus dibayar penumpang, termasuk tarif dasar, biaya tambahan, pajak, tip, tol, dll
18	payment_type	Jenis pembayaran yang digunakan dalam perjalanan. Nilainya mungkin meliputi: <ul style="list-style-type: none"> • 1: Kartu kredit. • 2: Tunai. • 3: Tidak ada biaya (No charge). • 4: Dispute (Sengketa). • 5: Unknown (Tidak diketahui).
19	trip_type	Jenis perjalanan taksi. Nilainya bisa: <ul style="list-style-type: none"> • 1: Perjalanan normal (dengan penumpang yang dijemput di jalanan). • 2: Perjalanan yang dipesan sebelumnya melalui aplikasi.

20	congestion_surcharge	Biaya tambahan karena kemacetan. Umumnya dikenakan pada waktu-waktu tertentu di daerah Manhattan.
----	----------------------	---

[NUMBER 1]

1. Create the output of the business processes and company strategy that you apply according to the answer you gave in question number 2 UTS Theory.

JAWABAN:

1) Context Analysis

Untuk memulai analisis Business Intelligence pada data transportasi ini, pertama-tama dilakukan proses preprocessing menggunakan **Python** untuk memastikan kualitas dan kebersihan data. Berikut adalah beberapa tahapan preprocessing:

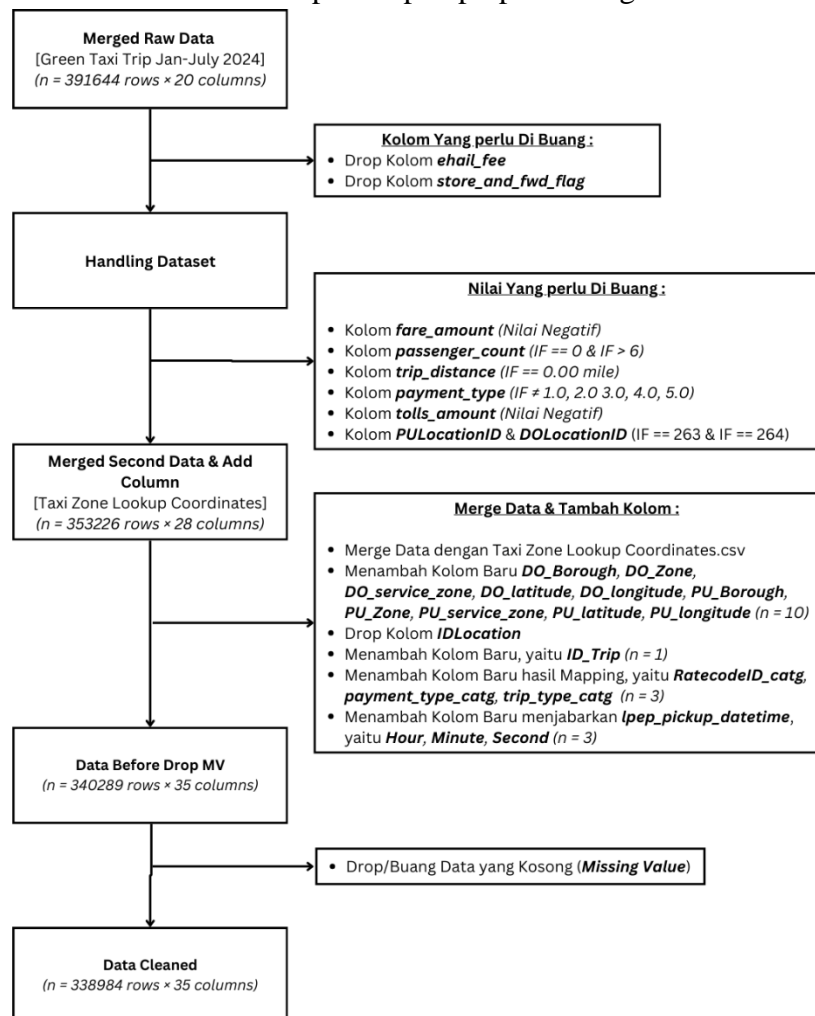


Fig 2. Proses Preprocessing Raw Data dengan Python

→ **Kolom yang Tidak Perlu :**

(1) **ehail_fee**

- **Alasan:** Kolom ini kosong, sehingga tidak memberikan informasi apapun untuk analisis.

(2) **store_and_fwd_flag**

- **Alasan:** Jika tidak ada prediksi yang melibatkan penyimpanan dan pengiriman, kolom ini mungkin tidak diperlukan. Biasanya hanya memiliki dua nilai, 'Y' dan 'N', yang tidak terlalu berpengaruh pada analisis.

→ **Nilai yang Perlu di Buang :**

(1) **passenger_count**

- **Alasan:** Jika passenger_count bernilai 0 (nol), karena tidak ada penumpang berarti perjalanan tersebut tidak relevan untuk analisis transportasi. Begitu juga jumlah penumpang yang > 6 karena tidak sesuai dengan aturan hukum.

(2) **trip_distance**

- **Alasan:** Jika trip_distance bernilai 0.00 mil, karena perjalanan tanpa jarak tidak akan memberikan data yang berguna.

(3) **fare_amount**

- **Alasan:** Jika fare_amount bernilai negatif atau sangat rendah, karena ini mungkin menunjukkan kesalahan dalam pencatatan data.

(4) **payment_type**

- **Alasan:** Jika nilai payment_type tidak termasuk dalam kategori yang valid (misalnya 1.0 hingga 5.0), ini bisa menandakan kesalahan data. ()

(5) **tolls_amount**

- **Alasan:** Jika tolls_amount bernilai negatif atau tidak logis, baris tersebut sebaiknya dihapus

(6) **tolls_amount**

- **Alasan:** PULocationID dan DOLocationID pada Dataset jika bernilai 263 atau 264 (Pada Dataset "taxi_zone_lookup_coordinates" value NaN).

→ **Merge Data :**

(1) **Merge pertama** : Menggabungkan df_cleaned dengan df_loc berdasarkan kolom PULocationID. Setelah merge, kolom dari df_loc diberi prefix PU_ untuk menunjukkan bahwa ini berasal dari PULocationID.

(2) **Merge kedua** : Setelah merge pertama, kita melakukan merge lagi menggunakan DOLocationID dengan prefix DO_ untuk kolom tambahan.

(3) **Drop kolom LocationID** : Setiap kali merge dilakukan, kolom LocationID yang duplikat dihapus setelah merge untuk menjaga kebersihan data.

→ **Kolom Tambahan :**

(1) Menambah Kolom Baru, yaitu **ID_Trip** ($n = 1$)

(2) Menambah Kolom Baru hasil Mapping, yaitu **RatecodeID_catg**, **payment_type_catg**, **trip_type_catg** ($n = 3$)

(3) Menambah Kolom Baru menjabarkan **lpep_pickup_datetime**, yaitu **Hour**, **Minute**, **Second** ($n = 3$)

2) Organization Analysis

Strategi berdasarkan SWOT-I Matrix:

(SO) Strategi – Kekuatan & Peluang:

- **Strategi:** Memanfaatkan kehadiran Green Taxi di area kurang terlayani untuk menganalisis titik jemput dan antar yang paling tinggi. Menggunakan data untuk meningkatkan aksesibilitas dengan fokus pada pola permintaan di wilayah tersebut dan menyusun strategi pemasaran yang lebih tepat sasaran.
- **Aksi:** Analisis zona pickup/drop-off dengan permintaan tertinggi di area focus (Heatmap & Grafik Trend)

Business Intelligence:

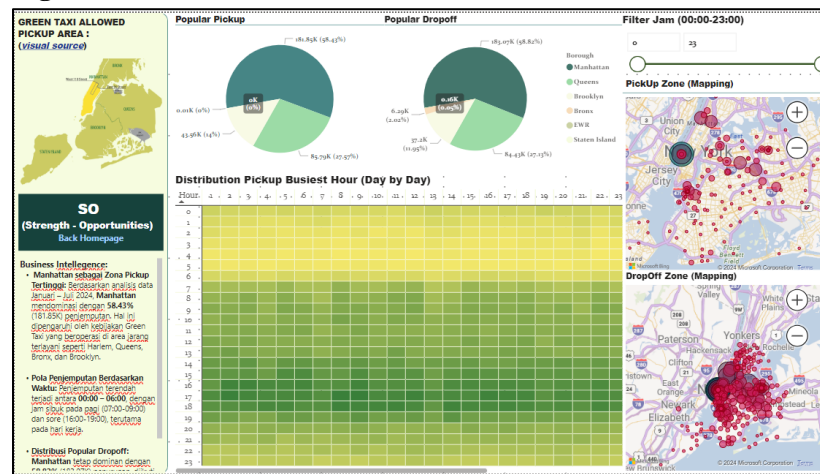


Fig 3. SO Strategy

Berdasarkan analisis data **NYC Green Taxi Siders Company** selama bulan Januari – Juli tahun 2024, **Zona PU (Pickup)** yang memiliki *demand Passenger Trips* (Pesanan Trip) tertinggi adalah **Borough Manhattan**. Berdasarkan kebijakan pemerintah NYC yang ditetapkan pada tahun 2013, layanan **Green Taxi** beroperasi di area seperti *Harlem, Queens, Bronx, Brooklyn*, serta bagian utara Manhattan di atas *East 96th Street* dan *West 110th Street* (fokus pada area yang jarang terlayani). Borough Manhattan menerima sekitar 65 juta pengunjung per tahun dan memiliki populasi sekitar $\pm 1,6$ juta penduduk, menjadikannya borough yang paling banyak dikunjungi Antonina (2024). Oleh karena itu, pemerintah menerapkan kebijakan tersebut.

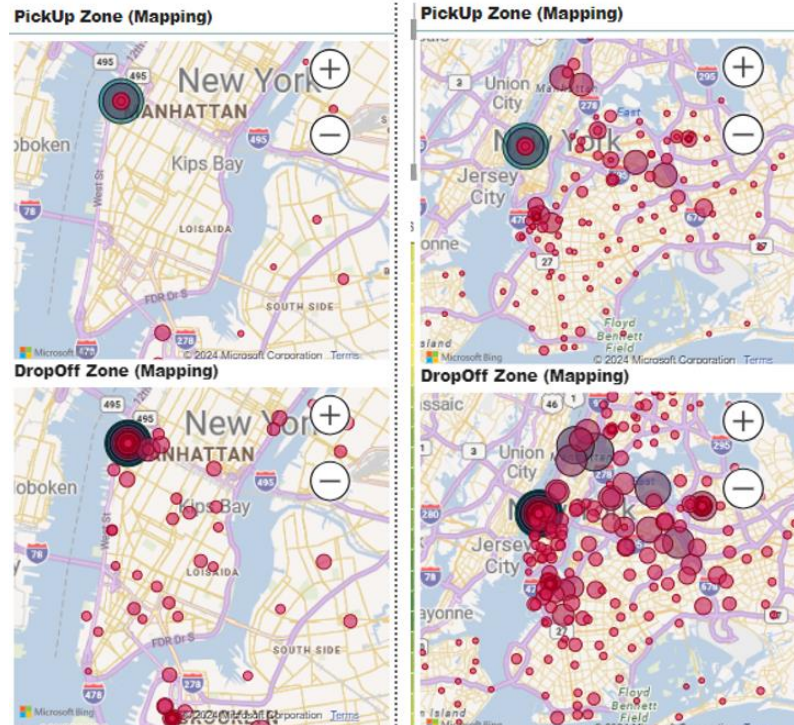


Fig 4. Zoom In Manhattan (Kiri) & Zoom Out NYC (Kanan)

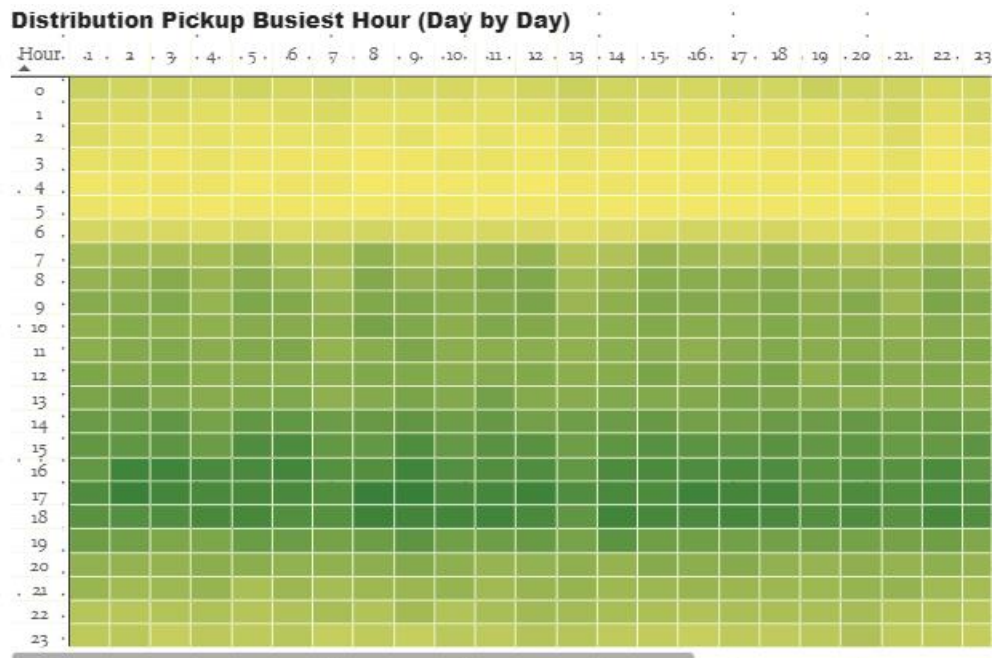


Fig 5. Busiest Hour (Jam Tingginya Demand Trip)

Pada Fig 5 menunjukkan analisis jam-jam dengan **demand trip** tertinggi berdasarkan visualisasi heatmap. Beberapa poin utama yang dapat disoroti adalah sebagai berikut:

1. **Periode Penjemputan Terendah** : Penjemputan terendah terjadi antara pukul **00:00 hingga 06:00**, ditunjukkan oleh warna yang lebih terang, mencerminkan frekuensi penjemputan yang rendah secara konsisten sepanjang minggu.

2. **Perbedaan Pola Hari Kerja dan Akhir Pekan** : Pada akhir pekan, terutama pada **Jumat dan Sabtu**, penurunan frekuensi penjemputan tampak lebih lambat (sekitar pukul **02:00 hingga 03:00**), mengindikasikan aktivitas malam yang lebih panjang dibandingkan hari kerja.
3. **Jam Sibuk Pagi dan Sore** : Warna yang lebih gelap menunjukkan penjemputan meningkat selama jam sibuk pagi (**07:00 hingga 09:00**) dan sore (**16:00 hingga 19:00**) pada sebagian besar hari kerja.
4. **Distribusi Keseluruhan**: Pola penjemputan yang lebih tinggi konsisten terjadi selama jam sibuk hari kerja, meskipun pada akhir pekan terdapat pergeseran aktivitas ke malam hari.

Popular Pickup menunjukkan bahwa sebagian besar penjemputan terjadi di **Manhattan** dengan **58.43%** atau **181.85K** penjemputan. Wilayah **Queens** menyumbang **27.57%** dengan **85.79K** penjemputan, diikuti oleh **Brooklyn** dengan **14%** atau **43.56K** penjemputan. **Bronx** hanya menyumbang **0%** atau **0.01K** penjemputan. Untuk **Popular Dropoff**, **Manhattan** juga mendominasi dengan **58.82%** atau **183.07K** penurunan. **Queens** berada di urutan kedua dengan **27.13%** atau **84.43K** penurunan, diikuti oleh **Brooklyn** dengan **11.95%** atau **37.2K** penurunan. **Bronx** menyumbang **2.02%** dengan **6.29K** penurunan, sementara **EWB (Newark Airport)** mencatat **0.05%** atau **0.16K** penurunan.

(ST) Strategi – Kekuatan & Ancaman:

- **Strategi:** Menggunakan kemampuan prediksi untuk meningkatkan layanan pelanggan di area dengan permintaan rendah dengan menawarkan insentif untuk menarik penumpang. Mengoptimalkan operasional Green Taxi agar tetap efisien meskipun menghadapi persaingan dari Yellow Taxi.
- **Aksi:** Menganalisis rata-rata persentase tips dari total fare dan pajak yang dibayar penumpang untuk mengidentifikasi daerah dengan perilaku pelanggan yang royal.

Business Intelligence:

Berdasarkan hasil analisis, berikut adalah **rata-rata tips tertinggi** di berbagai borough:

- **Brooklyn:** Mencatat rata-rata tips tertinggi dengan persentase mencapai **1584%**.
- **Staten Island:** Rata-rata jarak tempuh adalah **7.6 mile** dengan rata-rata tips **\$1.8**.
- **Queens:** Rata-rata jarak tempuh **3.6 mile** dengan rata-rata tips **\$2.1**.
- **Brooklyn:** Rata-rata jarak tempuh **3.5 mile** dengan rata-rata tips **\$3.2**.
- **Manhattan:** Rata-rata jarak tempuh **2.4 mile** dengan rata-rata tips **\$2.4**.

Rekor tertinggi untuk jumlah tips diberikan di Brooklyn, mencapai **\$222**, sementara jarak tempuh rata-rata untuk trip adalah **3.45 mile**. Sebaliknya, untuk **Staten Island**, meskipun rute perjalanan paling jauh dengan rata-rata **7.6 mile**, rata-rata tips yang diberikan hanya **\$6**, menunjukkan adanya perbedaan perilaku dalam memberi tip di berbagai wilayah.

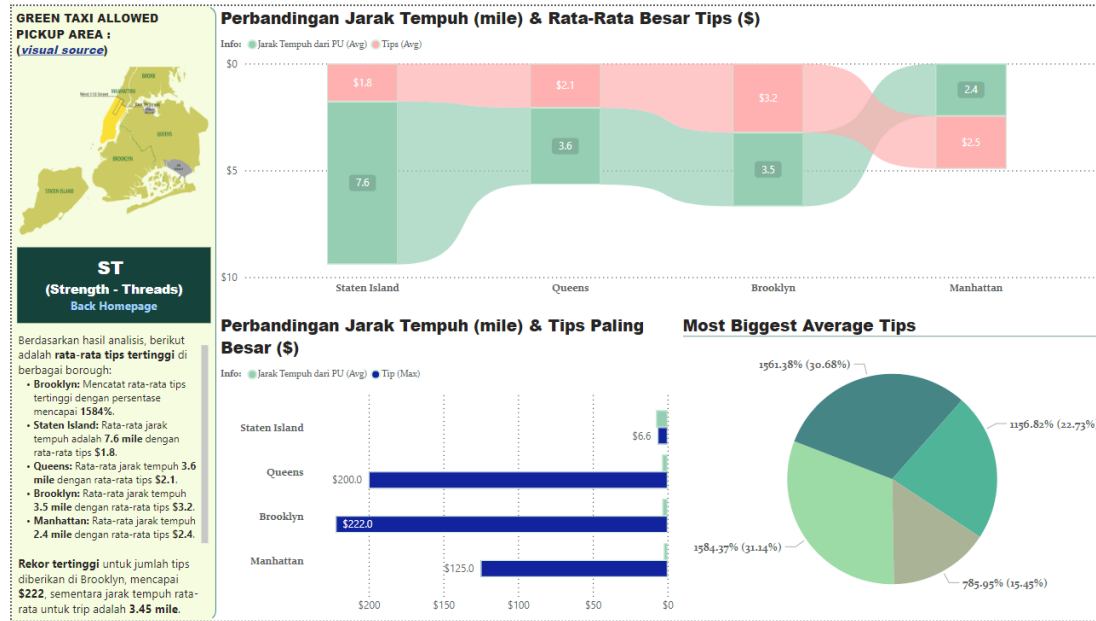


Fig 6. SO Strategy

(+) Manfaat Bagi Driver

- **Peluang Pendapatan yang Lebih Besar:**

Pengemudi memiliki potensi untuk mendapatkan tips yang lebih tinggi di borough seperti Brooklyn, yang mencatat rata-rata tips tertinggi.

- **Konsistensi Dalam Pemberian Tips:**

Penumpang di beberapa area menunjukkan perilaku memberi tip yang royal, memberikan keuntungan finansial yang lebih baik bagi pengemudi.

- **Analisis Data untuk Optimalisasi:**

Dengan adanya analisis perilaku penumpang, pengemudi dapat menyesuaikan layanan mereka untuk meningkatkan kepuasan pelanggan dan, pada gilirannya, meningkatkan tip.

- **Meningkatkan Loyalitas Pelanggan:**

Melalui program insentif dan layanan yang ditingkatkan, pengemudi dapat menciptakan hubungan yang lebih baik dengan penumpang, yang dapat mengarah pada peningkatan jumlah perjalanan berulang.

(WO) Strategi – Kelemahan & Peluang:

- **Strategi:** Mengumpulkan lebih banyak data tentang penumpang untuk memahami karakter dan perilaku mereka, yang dapat meningkatkan pelayanan dan pemasaran. Membangun kemitraan dengan layanan lain atau program pemerintah untuk memperoleh akses data tambahan yang dapat membantu meningkatkan layanan di area kurang terlayani.
- **Aksi:**
 1. Analisis perilaku penumpang (*jenis pembayaran, tip, rata-rata jarak tempuh*) untuk memaksimalkan layanan di zona yang paling menguntungkan.
 2. Analisis tipe pemesanan trip oleh penumpang (*Normal Trip, Pre-Book Trip*)

Business Intelligence

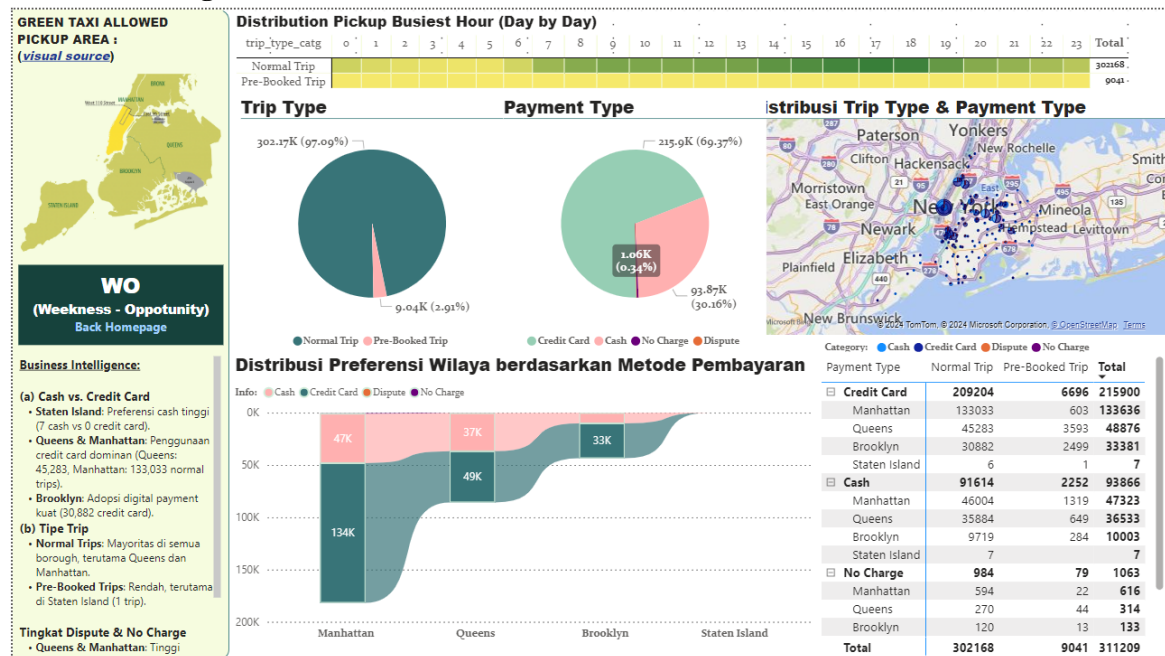


Fig 7. WO Strategy

(a) Cash vs. Credit Card :

1. Di Staten Island, terdapat lebih banyak penggunaan **cash** (7) dibandingkan **credit card** (7 total) untuk perjalanan normal, menunjukkan preferensi penumpang yang lebih cenderung menggunakan uang tunai.
2. Di Queens dan Manhattan, penggunaan **credit card** jauh lebih tinggi, dengan Queens mencatat 45,283 perjalanan normal dan Manhattan 133,033, **menunjukkan bahwa penumpang di daerah ini lebih nyaman menggunakan pembayaran non-tunai.**
3. Brooklyn juga menunjukkan preferensi kuat terhadap credit card untuk perjalanan normal (30,882), menunjukkan adopsi pembayaran digital yang tinggi di daerah ini

(b) Tipe Trip :

Normal Trip vs. Pre-Booked Trip:

1. Sebagian besar perjalanan adalah normal trips di semua borough, dengan Queens dan Manhattan memiliki jumlah yang sangat tinggi, masing-masing 45,283 dan 133,033 untuk pembayaran dengan credit card.
2. Pre-booked trips di Queens dan Manhattan menunjukkan angka yang lebih rendah (649 dan 1,319), yang mengindikasikan bahwa penumpang lebih memilih untuk memesan perjalanan mereka pada saat itu daripada merencanakan sebelumnya.
3. Staten Island memiliki jumlah perjalanan pre-booked yang paling rendah, dengan hanya 1 perjalanan menggunakan credit card.

Tingkat Dispute dan No Charge :

- Di Queens dan Manhattan, terdapat sejumlah **dispute** dan **no charge** untuk perjalanan normal dan pre-booked. Ini menunjukkan potensi masalah dalam pelayanan yang mungkin perlu ditangani untuk meningkatkan kepuasan pelanggan.

- Khususnya, Manhattan mencatat **268 disputes** untuk perjalanan normal dan **594 no charge**, yang menandakan adanya masalah yang perlu diidentifikasi dan diperbaiki untuk menjaga reputasi layanan.

Strategi untuk meningkatkan layanan taksi mencakup **pengumpulan data** tentang *preferensi pembayaran, jenis trip, dan kondisi perjalanan* untuk memahami **perilaku penumpang**.

Manfaat strategi bagi **perusahaan**:

- **Meningkatkan layanan:** Penyesuaian berdasarkan preferensi penumpang di tiap area.
- **Efisiensi operasional:** Optimalkan alokasi kendaraan (Sehingga Driver yang tidak Cashless lebih diutarakan di daerah yang mayoritas pembayaran cash).
- **Mitigasi masalah:** Perbaiki di area dengan tingkat dispute tinggi.

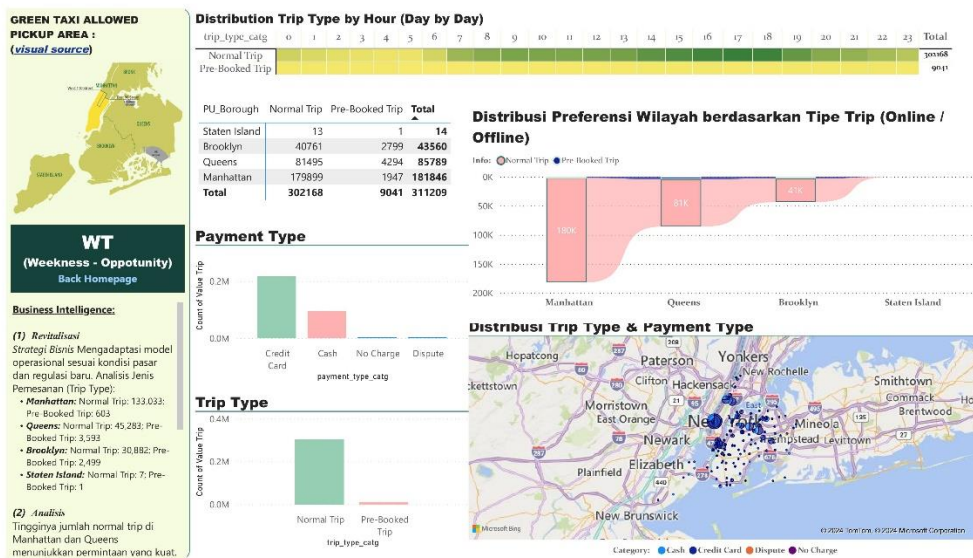
Bagi **pengemudi**:

- **Pendapatan maksimal:** Fokus di area dengan tip dan pembayaran tinggi.
- **Pilihan area sesuai preferensi:** Mengurangi masalah kembalian di zona cash tinggi.
- **Mengurangi waktu kosong:** Optimalkan rute sesuai permintaan tertinggi.
- **Kepuasan pelanggan:** Layanan lebih personal, meningkatkan loyalitas.

(WT) Strategi – Kelemahan & Ancaman:

- **Strategi:** Merevitalisasi strategi bisnis dengan mengubah model operasional agar sesuai dengan kondisi pasar dan regulasi baru. Mengembangkan program loyalitas untuk mempertahankan pelanggan dan meningkatkan permintaan di area dengan persaingan tinggi.
- **Aksi:** Mengidentifikasi jenis pemesanan (*online/offline*) untuk memahami dinamika pasar: banyaknya pemesanan online menunjukkan aplikasi yang baik (+) dan kemungkinan kurangnya armada di daerah tersebut (-), sedangkan banyaknya pemesanan offline menunjukkan banyaknya armada yang beroperasi.

Business Intelligence



(1) Revitalisasi Strategi Bisnis

Mengadaptasi model operasional sesuai kondisi pasar dan regulasi baru.

Analisis Jenis Pemesanan (Trip Type):

- **Manhattan:** Normal Trip: 133,033; Pre-Booked Trip: 603
- **Queens:** Normal Trip: 45,283; Pre-Booked Trip: 3,593
- **Brooklyn:** Normal Trip: 30,882; Pre-Booked Trip: 2,499
- **Staten Island:** Normal Trip: 7; Pre-Booked Trip: 1

(2) Analisis

Tingginya jumlah normal trip di Manhattan dan Queens menunjukkan permintaan yang kuat. Rendahnya pre-booked trips, terutama di Staten Island, menunjukkan potensi pasar yang belum dimanfaatkan.

(3) Rekomendasi

- Optimalisasi Armada: Fokuskan armada di Manhattan dan Queens, mengingat permintaan tinggi untuk normal trip.
- Tingkatkan Kesadaran Pre-Booked Trip: Luncurkan kampanye pemasaran untuk layanan pre-booking, khususnya di area dengan banyak normal trip.
- Analisis Pemesanan: Pantau frekuensi pemesanan online dan offline untuk mengidentifikasi area yang memerlukan peningkatan layanan.

[NUMBER 2]

2. Create the output of the Analytics process that you apply according to the answer you gave in question number 2 UTS Theory.

JAWABAN:

1) Menambah Kolom Baru PU_LOCATION & DO_LOCATION

Kelemahan & Strategi: Data yang ada pada dataset TLC Trip Record menunjukkan bahwa titik koordinat untuk lokasi penjemputan (PickUp) dan pengantaran (DropOff) tidak akurat atau tidak sesuai dengan lokasi yang sebenarnya. Hal ini dapat mengganggu analisis permintaan dan pemetaan pola perjalanan.

1 PU_LOCATION = 'NYC_Green_Taxi_2024_Cleaned(M1-M7)'[PU_Zone]&', "&'NYC_Green_Taxi_2024_Cleaned(M1-M7)'[PU_Borough]&', "&'NEW YORK"	
DO_LOCATION	PU_LOCATION
East Harlem South, Manhattan, NEW YORK	East Harlem North, Manhattan, NEW YORK
East Harlem South, Manhattan, NEW YORK	East Harlem North, Manhattan, NEW YORK
East Harlem South, Manhattan, NEW YORK	East Harlem North, Manhattan, NEW YORK
East Harlem South, Manhattan, NEW YORK	East Harlem North, Manhattan, NEW YORK
East Harlem South, Manhattan, NEW YORK	East Harlem North, Manhattan, NEW YORK

2) Mengubah Format Data Menjadi Currency

trip_distance	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount
1.25	\$7.20	\$0	\$0.50	\$0.0	\$0	\$1	\$8.7
1.37	\$12.80	\$0	\$0.50	\$0.0	\$0	\$1	\$14.3
0.71	\$7.20	\$0	\$0.50	\$0.0	\$0	\$1	\$8.7
0.83	\$6.50	\$0	\$0.50	\$0.0	\$0	\$1	\$8
1.9	\$12.80	\$0	\$0.50	\$0.0	\$0	\$1	\$14.3
1.8	\$12.80	\$0	\$0.50	\$0.0	\$0	\$1	\$14.3
1.78	\$12.10	\$0	\$0.50	\$0.0	\$0	\$1	\$13.6
0.8	\$5.80	\$0	\$0.50	\$0.0	\$0	\$1	\$7.3
1.07	\$12.10	\$0	\$0.50	\$0.0	\$0	\$1	\$13.6
1.22	\$7.20	\$0	\$0.50	\$0.0	\$0	\$1	\$8.7
1.06	\$7.20	\$0	\$0.50	\$0.0	\$0	\$1	\$8.7
0.84	\$5.80	\$0	\$0.50	\$0.0	\$0	\$1	\$7.3
0.95	\$7.90	\$0	\$0.50	\$0.0	\$0	\$1	\$9.4
1.2	\$7.90	\$0	\$0.50	\$0.0	\$0	\$1	\$9.4
0.94	\$7.90	\$0	\$0.50	\$0.0	\$0	\$1	\$9.4
1.41	\$7.90	\$0	\$0.50	\$0.0	\$0	\$1	\$9.4

3) Menambah Kolom Presentas Tip

Mengetahui behaviour customer berdasarkan persentase tip dari total fare yang dia bayar.

```

1 tip_percentage =
2 IF(
3   ('NYC_Green_Taxi_2024_Cleaned(M1-M7)'[fare_amount] + 'NYC_Green_Taxi_2024_Cleaned(M1-M7)'[tolls_amount] + 'NYC_Green_Taxi_2024_Cleaned(M1-M7)'
4    [tip_amount]) > 0,
5   ('NYC_Green_Taxi_2024_Cleaned(M1-M7)'[tip_amount] / ('NYC_Green_Taxi_2024_Cleaned(M1-M7)'[fare_amount] + 'NYC_Green_Taxi_2024_Cleaned(M1-M7)'
6    [tolls_amount] + 'NYC_Green_Taxi_2024_Cleaned(M1-M7)'
7    [tip_amount])) * 100

```

tip_percentage
274285.71%
228571.43%
186554.62%
178095.24%

Mengetahui behaviour customer berdasarkan persentase tip dari total fare yang dia bayar. Mengapa bisa menyetun >100%? Karena beberapa behaviour pelanggan sangat royal dan memberikan tips berkali – kali lipat dari total fare amountnya.

4) Filter Keseluruhan

payment_type_catg ^ ^

is Cash, Credit Card, D... ^ ^

Filter type ^

Basic filtering ^

Search

☒ Select all

☒ Cash 93866

☒ Credit Card 215900

☒ Dispute 380

☒ No Charge 1063

☐ Unknown 9

☐ Require single selection

1) Descriptive Analytics

Dashboard Descriptive Analytics memberikan wawasan komprehensif tentang operasi taksi dengan fitur-fitur penting seperti filter berdasarkan bulan dan jam, serta statistik utama yang mencakup rata-rata tarif, pajak, total perjalanan, jumlah penumpang, dan jarak tempuh. Informasi mengenai tipe trip (online atau offline) dan metode pembayaran (kartu kredit, tunai, tidak dikenakan biaya, atau sengketa) juga ditampilkan. Selain itu, peta distribusi permintaan untuk lokasi pickup dan drop-off serta grafik distribusi jam sibuk membantu mengidentifikasi pola permintaan. Analisis ini mendukung pengambilan keputusan strategis untuk meningkatkan efisiensi operasional dan pelayanan kepada pelanggan. Data ini mencakup periode Januari 2024 hingga Juli 2024.

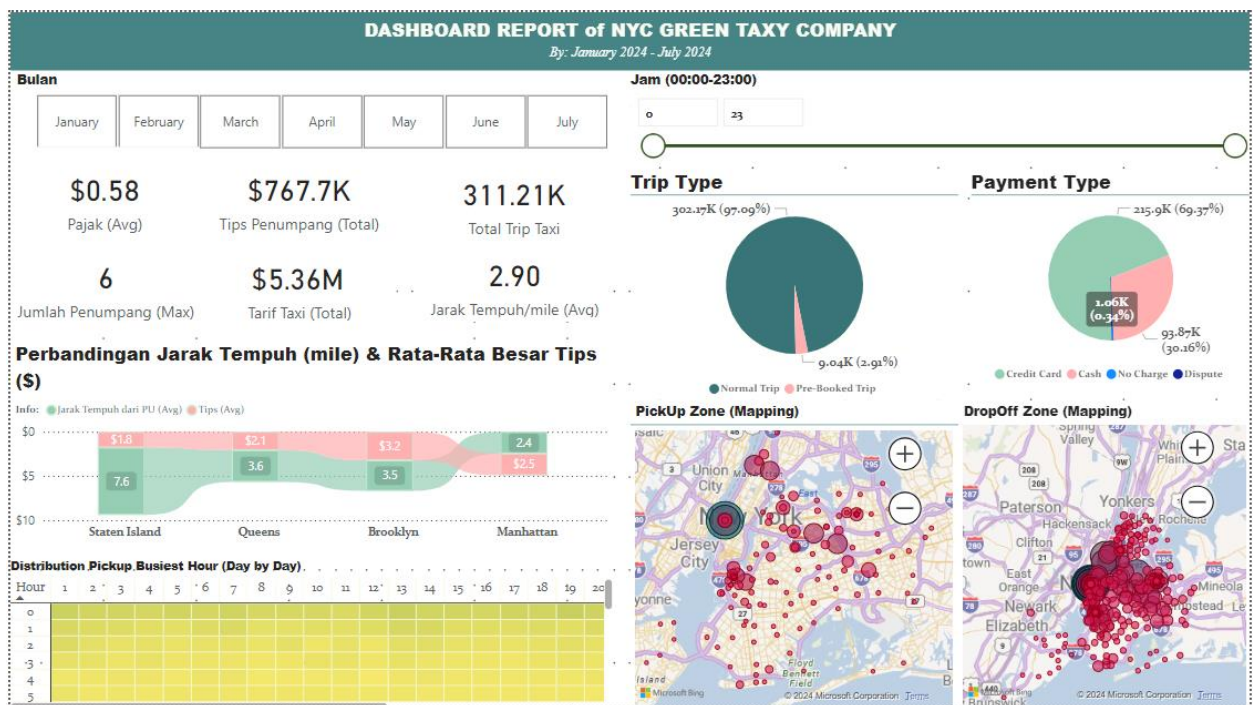


Fig 8. Dashboard Descriptive

2) Predictive Analytics

Data ini mencakup periode Januari 2024 hingga Juli 2024. Pada dashboard predictive, terdapat prediksi rata-rata pajak, yang kemungkinan besar akan mengalami penurunan di masa depan, meskipun diperkirakan akan meningkat kembali pada pertengahan Agustus. Rata-rata tarif juga diprediksi akan melonjak pesat pada awal Januari, seiring dengan tingginya permintaan di awal bulan baru, setelah mengalami penurunan drastis di akhir Juli. Selanjutnya, ada prediksi untuk permintaan perjalanan taksi, di mana setelah akhir Juli, diperkirakan permintaan akan menurun, tetapi akan kembali meningkat pada awal Agustus, meskipun belum mencapai lonjakan signifikan seperti bulan-bulan sebelum Juli. Rata-rata tips dari penumpang diprediksi akan tetap konstan ke depannya, yang menunjukkan perlunya pengembangan strategi untuk **NYC Green Taxi Siders Company by Lola**.

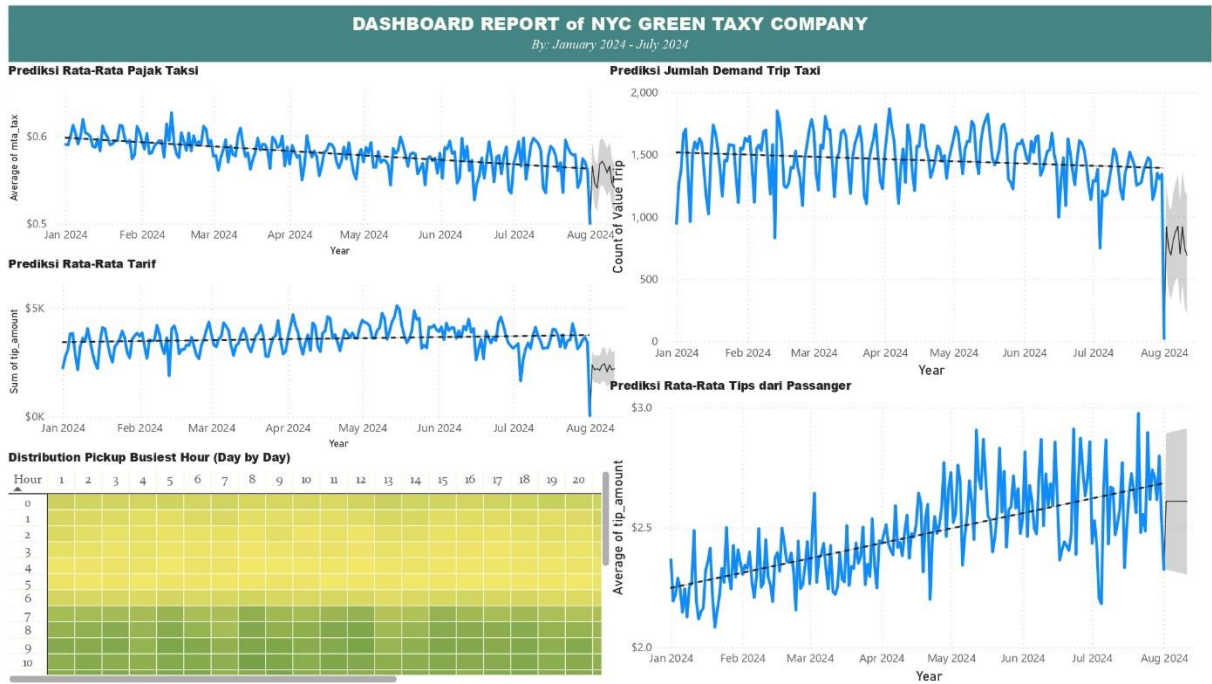


Fig 9. Dashboard Predictives

[NUMBER 3]

3. Create an interactive visualization as Business Reporting homepage that reflects the analytics application you deploy, illustrating all aspects of your answer at UTS theory number 2.

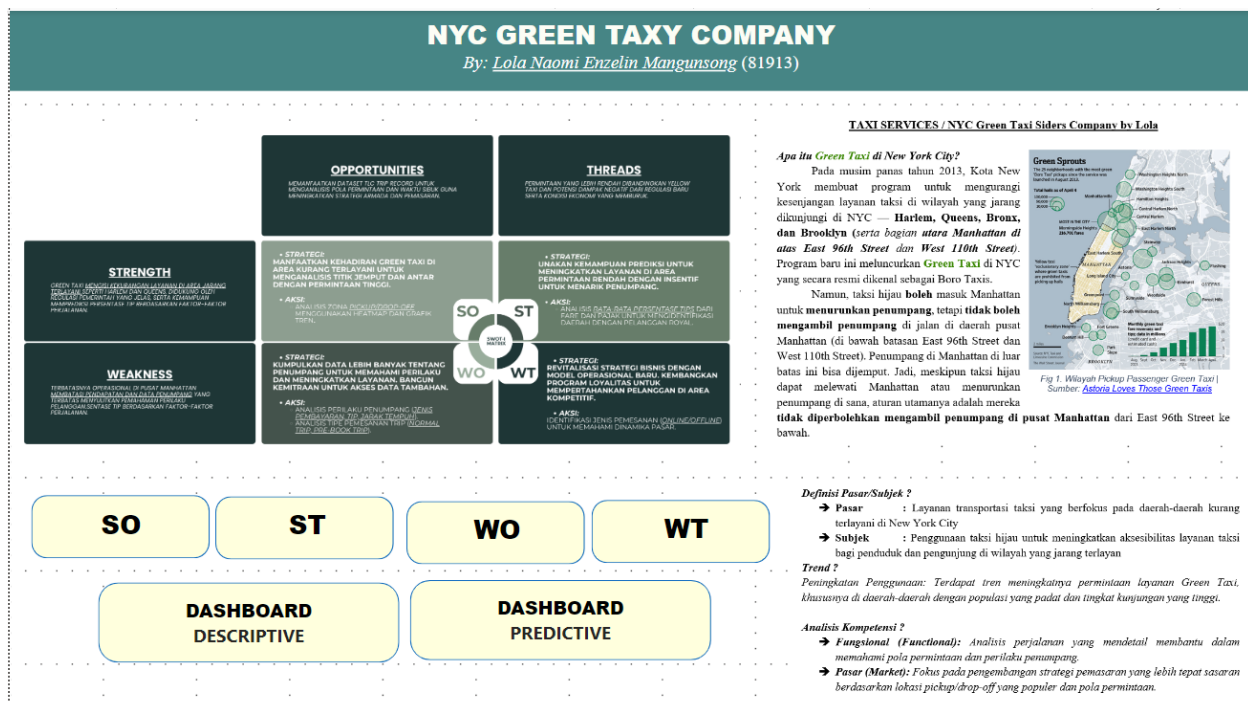
JAWABAN:

(Home Page)



Pada Home page terdapat 2 button yang dapat menavigasi ke page lainnya.

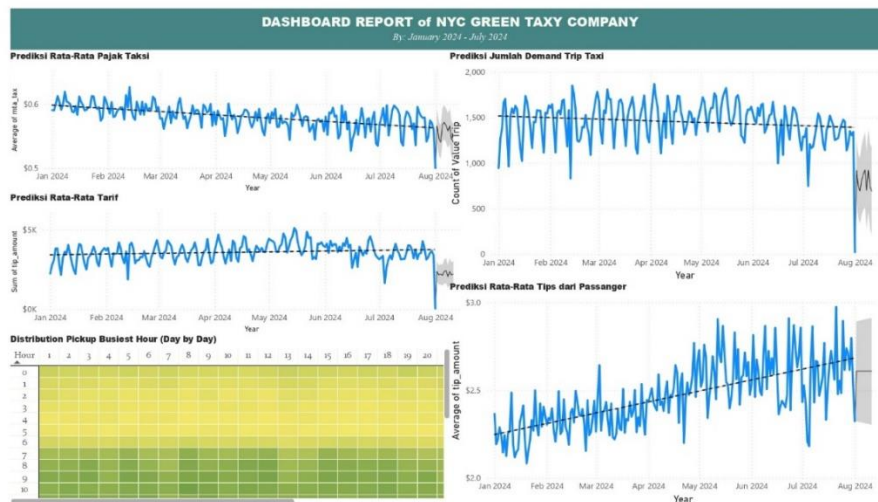
1. Button “About Company”
2. Button “Dashboard”



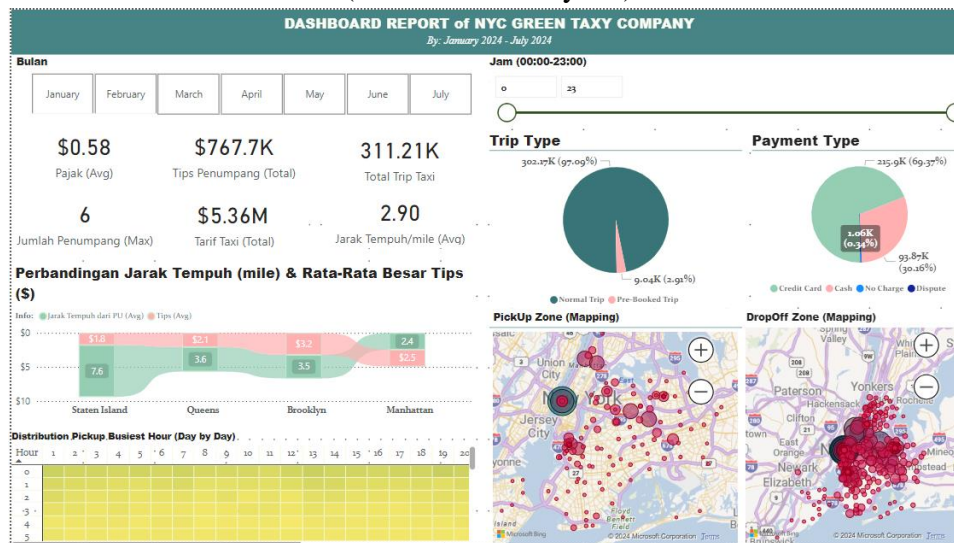
Pada saat klik “About Company”, akan muncul deskripsi perusahaan dan SWOT-I Matrix yang merupakan rancangan strategi (Business Strategy). Terdapat 5 button yang akan menavigasi anda ke page lainnya :

1. Button “Dashboard” -> Kembali ke Dashboard
2. Button “SO” -> Halaman SO
3. Button “ST” -> Halaman ST
4. Button “WO” -> Halaman WO
5. Button “WT” -> Halaman WT

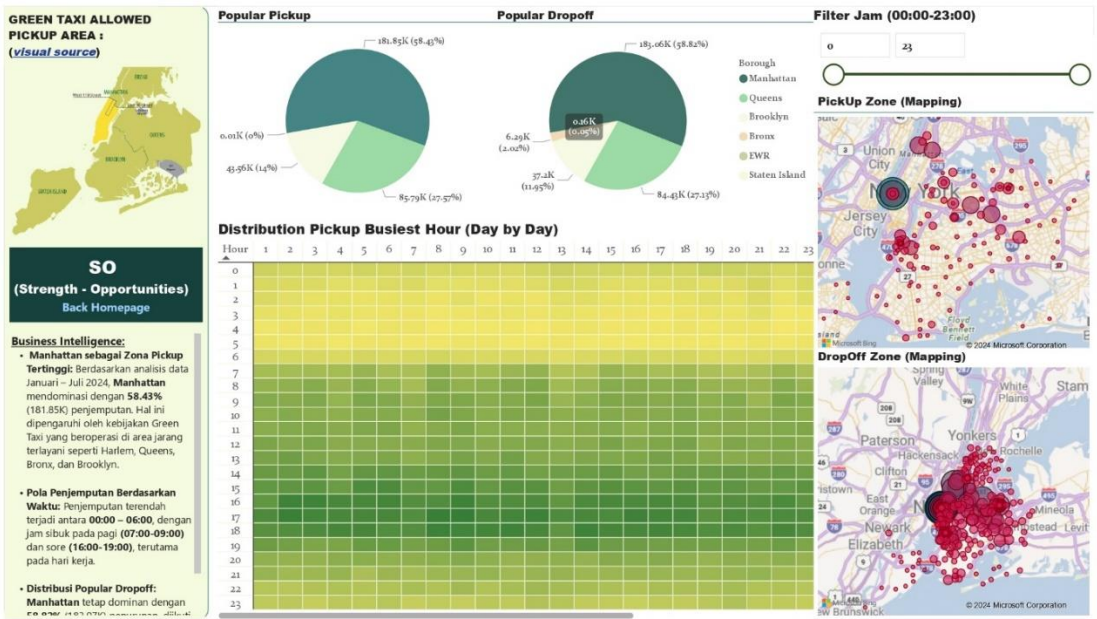
(Dashboard Predictive)



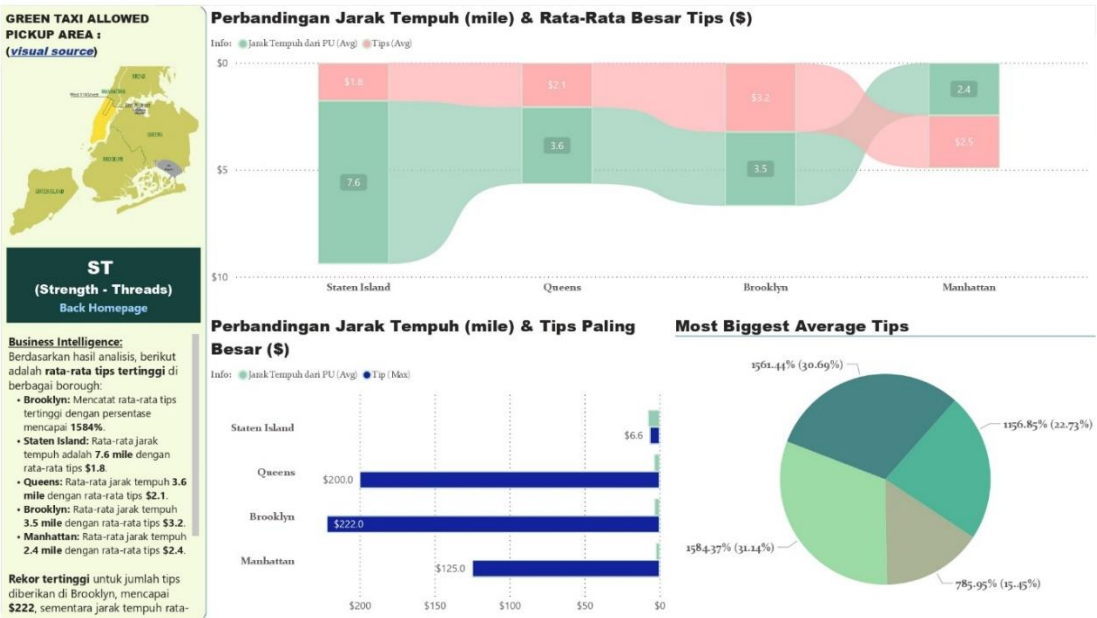
(Dashboard Analytics)



(SO Strategy Page)



(ST Strategy Page)



(WO Strategy Page)

GREEN TAXI ALLOWED
PICKUP AREA :
(visual source)



WO
(Weekness - Opportunity)
Back Homepage

Business Intelligence:

(a) Cash vs. Credit Card

- **Staten Island:** Preferensi cash tinggi (7 cash vs 0 credit card).
- **Queens & Manhattan:** Penggunaan credit card dominan (Queens: 45,283; Manhattan: 133,033 normal trips).
- **Brooklyn:** Adopsi digital payment kuat (30,882 credit card).

(b) Tipe Trip

- **Normal Trips:** Mayoritas di semua borough, terutama Queens dan Manhattan.
- **Pre-Booked Trips:** Rendah, terutama di Staten Island (1 trip).

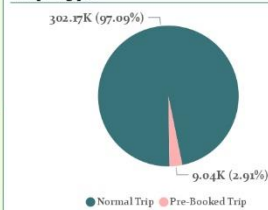
Tingkat Dispute & No Charge

- **Queens & Manhattan:** Tinggi

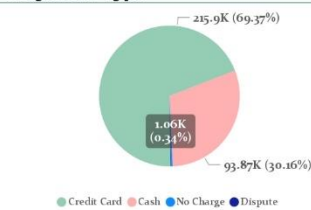
Distribution Payment Type by Hour (Day by Day)

payment_type_catg	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Total
Cash																									93866
Credit Card																									215900
Dispute																									380
No Charge																									1063

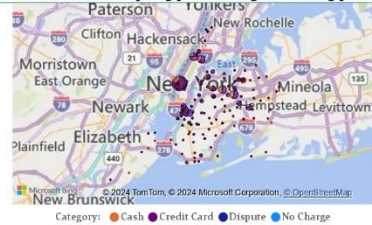
Trip Type



Payment Type

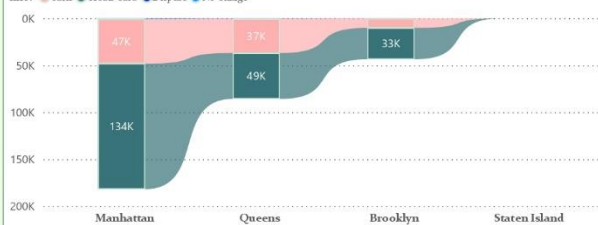


Distribusi Trip Type & Payment Type



Distribusi Preferensi Wilayah berdasarkan Metode Pembayaran

Info: Cash Credit Card Dispute No Charge



Payment Type	Normal Trip	Pre-Booked Trip	Total
Credit Card	209204	6696	215900
Manhattan	133033	603	133636
Queens	45283	3593	48876
Brooklyn	30882	2499	33381
Staten Island	6	1	7
Cash	91614	2252	93866
Manhattan	46004	1319	47323
Queens	35884	649	36533
Brooklyn	9719	284	10003
Staten Island	7		7
No Charge	984	79	1063
Manhattan	594	22	616
Queens	770	44	814
Total	302168	9041	311209

(WT Strategy Page)

GREEN TAXI ALLOWED
PICKUP AREA :
(visual source)



WT
(Weekness - Opportunity)
Back Homepage

Business Intelligence:

(1) Revitalisasi

Strategi Bisnis Mengadaptasi model operasional sesuai kondisi pasar dan regulasi baru. Analisis Jenis Pemesanan (Trip Type):

- **Manhattan:** Normal Trip: 133,033; Pre-Booked Trip: 603
- **Queens:** Normal Trip: 45,283; Pre-Booked Trip: 3,593
- **Brooklyn:** Normal Trip: 30,882; Pre-Booked Trip: 2,499
- **Staten Island:** Normal Trip: 7; Pre-Booked Trip: 1

(2) Analisis

Tingginya jumlah normal trip di Manhattan dan Queens menunjukkan permintaan yang kuat.

Distribution Trip Type by Hour (Day by Day)

trip_type_catg	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Total
Normal Trip																									302168
Pre-Booked Trip																									9041

PU_Borough	Normal Trip	Pre-Booked Trip	Total
Staten Island	13	1	14
Brooklyn	40761	2799	43560
Queens	81495	4294	85789
Manhattan	179899	1947	181846
Total	302168	9041	311209

Payment Type

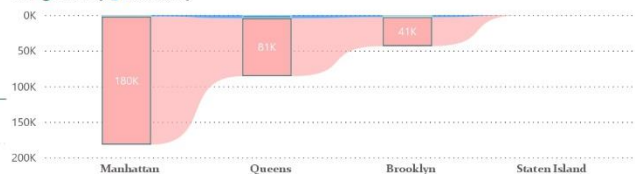


Trip Type



Distribusi Preferensi Wilayah berdasarkan Tipe Trip (Online / Offline)

Info: Normal Trip Pre-Booked Trip



Distribusi Trip Type & Payment Type



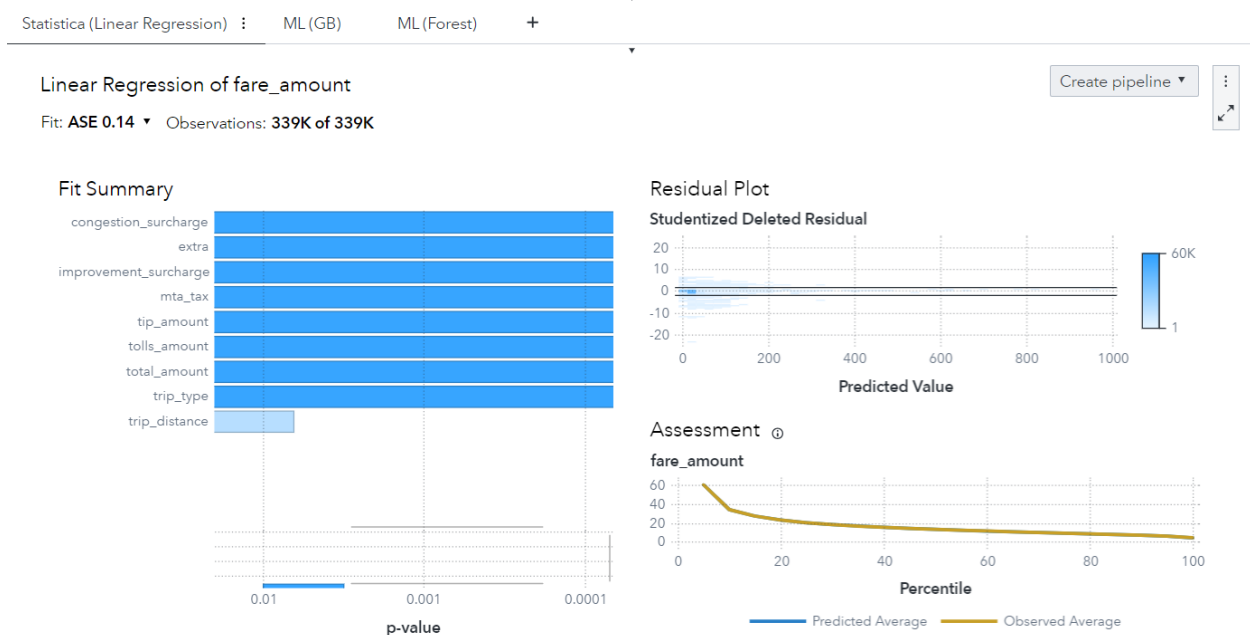
NUMBER 4

4. Create an analytical output which is a derivative of the Business Intelligence homepage that was made, according to the type of analytical process you did in answer to UTS theory number 3.

JAWABAN:

Statistical Methods

(1) Linear Regression

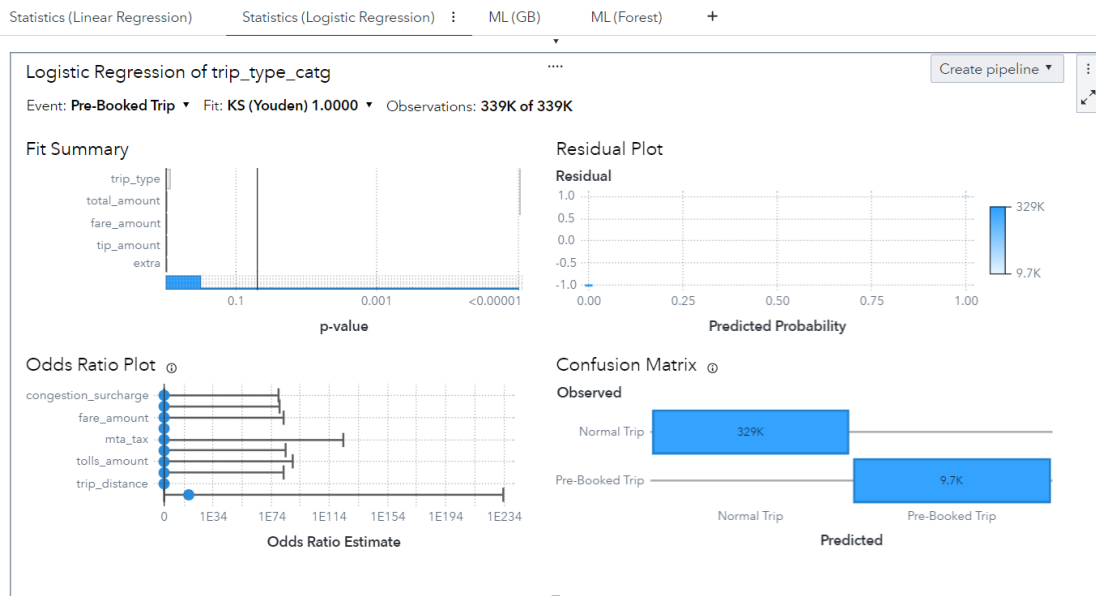


Hasil Menunjukkan:

Measures	
Adjusted R-Squared	0.0993
ASE	0.14
Mean Square Error	0.14
R-Squared	0.9993

Hasil analisis regresi linear menunjukkan bahwa tarif taksi (*fare_amount*) sangat dipengaruhi oleh variabel seperti Congestion Surcharge, tip amount, dan trip distance, dengan nilai R-Squared yang sangat tinggi (0.9993) menandakan hampir semua variasi tarif dapat dijelaskan oleh variabel tersebut. Namun, nilai Adjusted R-Squared yang rendah (0.0993) mengindikasikan bahwa model mungkin terlalu kompleks, sehingga kurang akurat untuk prediksi di luar dataset yang digunakan. Nilai Mean Square Error (MSE) dan Average Squared Error (ASE) masing-masing sebesar 0.14, menunjukkan bahwa meskipun model ini memiliki akurasi tinggi dalam menjelaskan variasi data, ada potensi untuk generalisasi yang kurang baik.

(2) Logistic Regression



Hasil Menunjukkan:

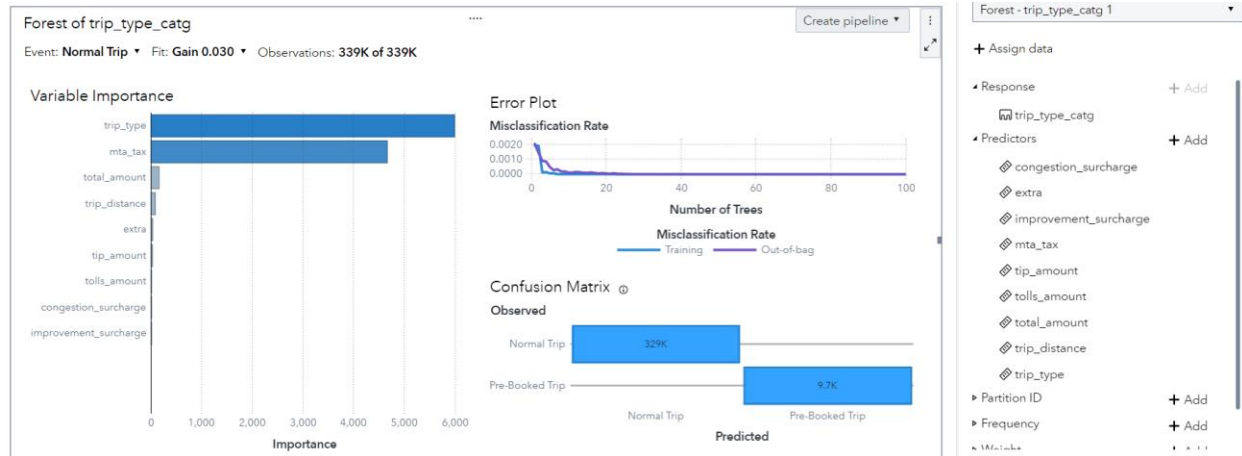
Measures	
Missclassification	0.0000
KS (Youden)	1.000
F1-Score	1.000
FDR (False Discovery Rate)	0.000
FPR (False Positive Rate)	0.000
Gain	1.000
C Statistic	1.000

Model regresi logistik menunjukkan ketepatan tinggi dalam mengklasifikasikan kategori pemesanan (*trip_type_category*) dengan nilai misclassification rate sebesar 0.0000, yang menandakan tidak ada kesalahan klasifikasi. KS statistic dan C Statistic masing-masing mencapai 1.000, menunjukkan kemampuan prediksi yang sangat baik. Selain itu, F1-Score juga mencapai 1.000, mencerminkan keseimbangan sempurna antara presisi dan recall. Nilai False Discovery Rate (FDR) dan False Positive Rate (FPR) yang nol menunjukkan keandalan tinggi model ini dalam situasi nyata, tanpa menghasilkan kesalahan positif.

Machine Learnings

(1) Forest

Sama dengan logistic regression, nilai target untuk Forest adalah **trip_type_category**, yang bertujuan untuk menganalisis perilaku penumpang dalam memesan secara online atau melalui aplikasi. Variabel kontinu yang digunakan sama dengan linear regression, yaitu: **congestion_surcharge**, **extra**, **improvement_surcharge**, **mta_tax**, **tip_amount**, **tolls_amount**, **total_amount**, **trip_distance**, dan **trip_type**.



Hasil Menunjukkan:

Measures	
Missclassification	0.0000
KS (Youden)	1.000
F1-Score	1.000
FDR (False Discovery Rate)	0.000
FPR (False Positive Rate)	0.000
Gain	0.030
C Statistic	1.000

Kesimpulan dari hasil analisis Forest menunjukkan bahwa model ini sangat efektif dalam mengklasifikasikan jenis pemesanan taksi (trip_type_category) berdasarkan variabel kontinu yang digunakan. Nilai **misclassification rate sebesar 0.0000** menunjukkan bahwa tidak ada kesalahan klasifikasi yang terjadi, yang mengindikasikan ketepatan model dalam mengidentifikasi tipe pemesanan. Nilai **KS statistic dan C Statistic masing-masing 1.000** menandakan bahwa model memiliki kemampuan prediksi yang sangat baik, dengan pemisahan yang optimal antara kategori pemesanan.

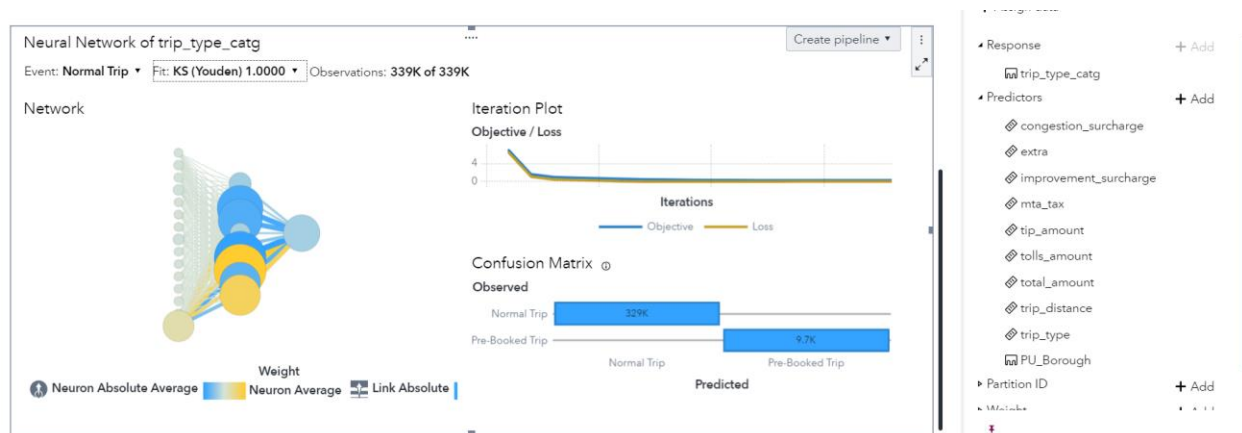
Selain itu, nilai **F1-Score sebesar 1.000** menunjukkan keseimbangan sempurna antara presisi dan recall, sementara nilai **False Discovery Rate (FDR) dan False Positive Rate (FPR) yang nol** menegaskan bahwa model tidak menghasilkan kesalahan positif, sehingga prediksi model sangat akurat. Meskipun nilai **Gain hanya sebesar 0.030**, keseluruhan metrik lainnya

menunjukkan bahwa Forest adalah model yang andal dalam menganalisis perilaku pemesanan penumpang baik secara online maupun melalui aplikasi.

(2) Neural Network

Sama dengan Forest, nilai target untuk Forest adalah **trip_type_category**, yang bertujuan untuk menganalisis perilaku penumpang dalam memesan secara online atau melalui aplikasi. Variabel kontinu yang digunakan sedikit berbeda dengan penambahan PU_Borough, agar semakin mengenail behaviour passanger dari dari segi geografis, yaitu: PU_Borough

, Pcongestion_surcharge, extra, improvement_surcharge, mta_tax, tip_amount, tolls_amount, total_amount, trip_distance, dan trip_type



Hasil Menunjukkan:

Measures	
Missclassification	0.000
KS (Youden)	1.000
F1-Score	1.000
FDR (False Discovery Rate)	0.000
Tau	0.056
FPR (False Positive Rate)	0.000
Gain	0.030
C Statistic	1.000

Hasil analisis *Neural Network* dalam mengklasifikasikan jenis pemesanan taksi (*trip_type_category*) menunjukkan akurasi yang sangat tinggi dengan *misclassification rate* sebesar 0.000, yang menandakan ketepatan sempurna dalam mengidentifikasi tipe pemesanan. Beberapa poin utama dari hasil evaluasi menunjukkan bahwa KS (Youden) dan C Statistic masing-masing mencapai 1.000, menunjukkan pemisahan optimal antara kelas positif dan negatif. F1-Score sebesar 1.000 juga menandakan keseimbangan sempurna antara presisi dan *recall*, memastikan bahwa model mampu mendeteksi semua kasus yang relevan tanpa kesalahan

klasifikasi. Nilai FDR dan FPR yang sama-sama nol menunjukkan bahwa model ini tidak menghasilkan kesalahan positif, sehingga prediksi yang dihasilkan sangat andal. Meski nilai Tau sebesar 0.056 dan Gain sebesar 0.030 hanya menambahkan informasi terbatas, metrik keseluruhan tetap kuat. Secara keseluruhan, *Neural Network* terbukti sangat andal dalam memprediksi perilaku pemesanan penumpang, terutama dengan variabel tambahan seperti PU_Borough sebagai informasi geografis, menjadikannya pilihan efektif untuk aplikasi perusahaan.