

Pola Perjalanan dan Perilaku Pengguna TransJakarta

Hani Handayani

Transjakarta

Transjakarta merupakan salah satu moda **transportasi publik** utama di Jakarta yang setiap harinya melayani jutaan penumpang. Sebagai sistem transportasi massal, Transjakarta berperan penting dalam mendukung mobilitas masyarakat, mengurangi kemacetan, dan menyediakan alternatif perjalanan yang lebih efisien.



Problem Statement

Tingginya volume penumpang transjakarta, maka diperlukan analisis mendalam untuk memahami **pola penggunaan rute, waktu perjalanan, karakteristik penumpang, serta metode pembayaran.** Hal ini penting agar Transjakarta dapat **mengoptimalkan operasional, mengatur kapasitas secara lebih tepat, serta meningkatkan pengalaman pengguna.**

Data Understanding

```

RangeIndex: 37900 entries, 0 to 37899
Data columns (total 22 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   transID                37900 non-null  object  
 1   payCardID              37900 non-null  int64   
 2   payCardBank            37900 non-null  object  
 3   payCardName            37900 non-null  object  
 4   payCardSex             37900 non-null  object  
 5   payCardBirthDate       37900 non-null  int64   
 6   corridorID             36643 non-null  object  
 7   corridorName           35970 non-null  object  
 8   direction              37900 non-null  float64  
 9   tapInStops             36687 non-null  object  
10   tapInStopsName         37900 non-null  object  
11   tapInStopsLat          37900 non-null  float64  
12   tapInStopsLon          37900 non-null  float64  
13   stopStartSeq           37900 non-null  int64   
14   tapInTime              37900 non-null  object  
15   tapOutStops            35611 non-null  object  
16   tapOutStopsName        36556 non-null  object  
17   tapOutStopsLat         36556 non-null  float64  
18   tapOutStopsLon         36556 non-null  float64  
19   stopEndSeq             36556 non-null  float64  
20   tapOutTime             36556 non-null  object  
21   payAmount              36893 non-null  float64  
dtypes: float64(7), int64(3), object(12)
memory usage: 6.4+ MB

```

Jumlah kolom pada dataset transjakarta adalah **37900** dan **22** Kolom.

Dari total keseluruhan dataset ada sekitar 10 kolom yang memiliki missing value dan tipe data yang tidak sesuai.

```

print(f"data duplikat :", df_tj.duplicated().sum())
data duplikat : 0

```


Data Cleaning

```
# mapping hanya untuk corridorID yang valid
corridor_map = (
    df_tj.dropna(subset=["corridorID", "corridorName"])
    .drop_duplicates("corridorID")
    .set_index("corridorID")["corridorName"]
    .to_dict()
)

# isi hanya kalau corridorID tidak NaN
df_tj["corridorName"] = df_tj.apply(
    lambda row: corridor_map.get(row["corridorID"], row["corridorName"])
    if pd.notna(row["corridorID"]) else row["corridorName"],
    axis=1
)

# mapping hanya untuk corridorName yang valid
corridor_name_map = (
    df_tj.dropna(subset=["corridorName", "corridorID"])
    .drop_duplicates("corridorName")
    .set_index("corridorName")["corridorID"]
    .to_dict()
)

# isi hanya kalau corridorName tidak NaN
df_tj["corridorID"] = df_tj.apply(
    lambda row: corridor_name_map.get(row["corridorName"], row["corridorID"])
    if pd.notna(row["corridorName"]) else row["corridorID"],
    axis=1
)
```

pada tahapan data cleaning dilakukan handling **missing value** dengan **fillna** dan **drop missing value**.

Data Transformation / Feature Engineering

```

Index: 32555 entries, 0 to 37899
Data columns (total 34 columns):
 #   Column              Non-Null Count  Dtype  
---  -
 0   transID              32555 non-null  object  
 1   payCardID            32555 non-null  int64   
 2   payCardBank          32555 non-null  object  
 3   payCardName          32555 non-null  object  
 4   payCardSex           32555 non-null  object  
 5   payCardBirthDate     32555 non-null  int64   
 6   corridorID           32555 non-null  object  
 7   corridorName         32555 non-null  object  
 8   direction            32555 non-null  float64  
 9   tapInStops           32555 non-null  object  
10   tapInStopsName       32555 non-null  object  
11   tapInStopsLat        32555 non-null  float64  
12   tapInStopsLon        32555 non-null  float64  
13   stopStartSeq         32555 non-null  int64   
14   tapInTime            32555 non-null  datetime64[ns]
15   tapOutStops          32555 non-null  object  
16   tapOutStopsName      32555 non-null  object  
17   tapOutStopsLat       32555 non-null  float64  
18   tapOutStopsLon       32555 non-null  float64  
19   stopEndSeq           32555 non-null  float64  
20   tapOutTime           32555 non-null  datetime64[ns]
21   payAmount            32555 non-null  float64  
22   age                  32555 non-null  int64   
23   dayName              32555 non-null  object  
24   isweekend            32555 non-null  int64   
25   date_tap_in          32555 non-null  object  
26   age_category         32555 non-null  object  
27   tapIn_hour           32555 non-null  object  
28   tapOut_hour          32555 non-null  object  
29   tapIn_category       32555 non-null  object  
30   tapout_category      32555 non-null  object  
31   duration_hour        32555 non-null  timedelta64[ns]
32   Distance             32555 non-null  float64  
33   duration_min         32555 non-null  float64  
dtypes: datetime64[ns](2), float64(9), int64(5), object(17), timedelta64[ns](1)
memory usage: 8.7+ MB

```

- buat kolom baru untuk age
- kolom baru untuk weekdays dan weekend atau day
- kolom durasi perjalanan
- kategori remaja/dewasa/lansia
- kategori waktu pagi/siang/sore/malam
- kategori hari
- kategori waktu

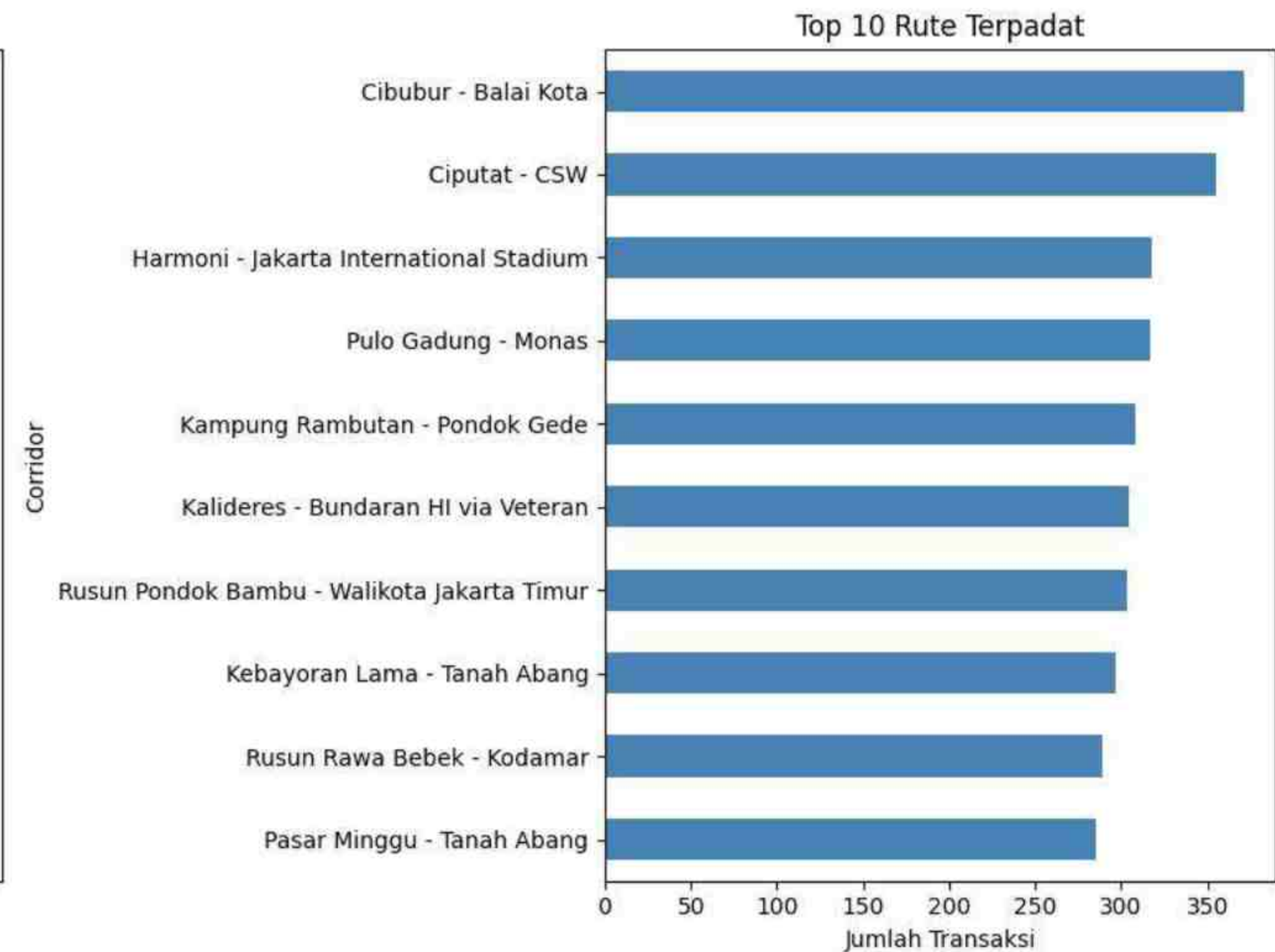
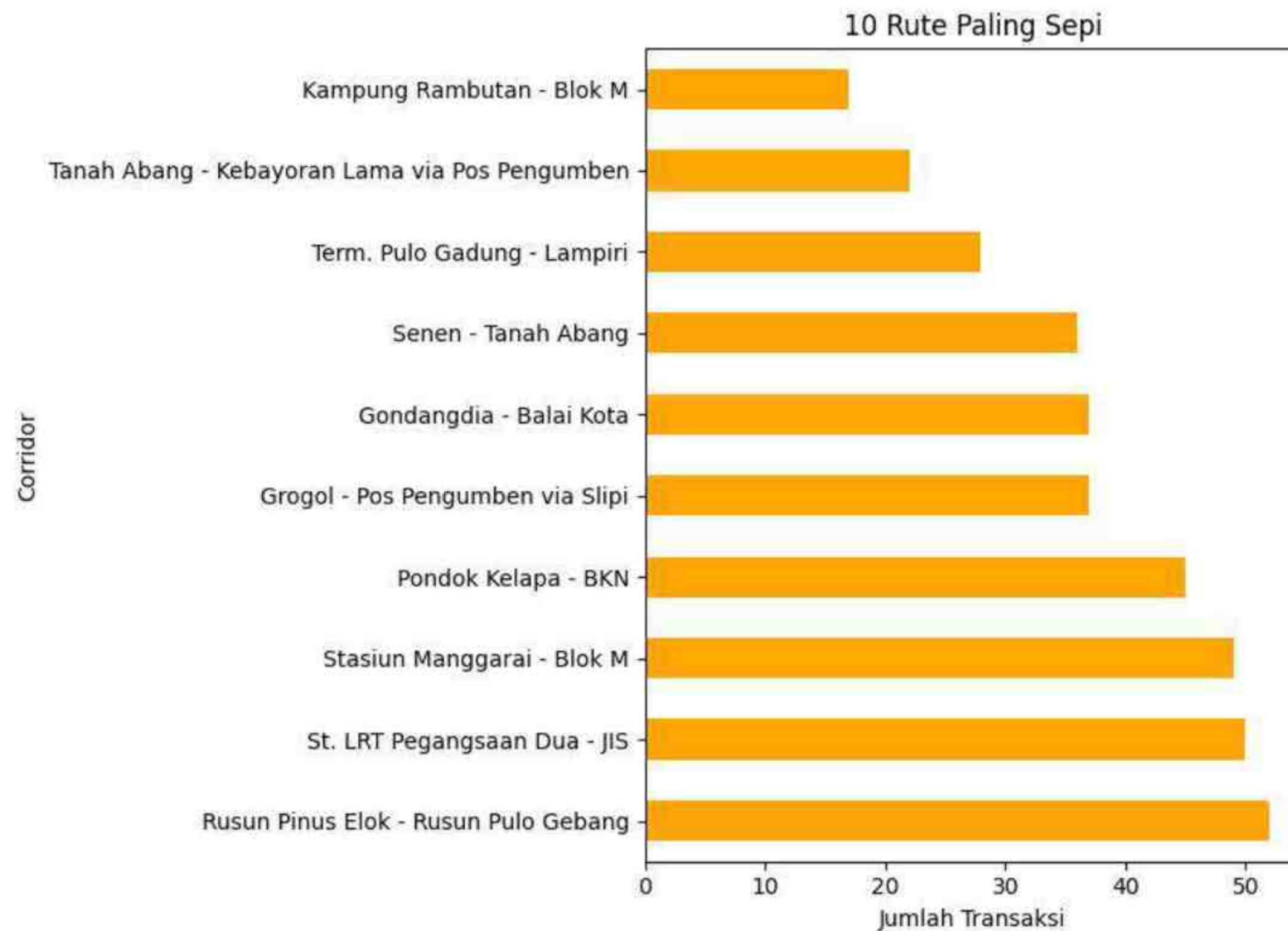
Exploratory Data Analysis (EDA)

- menjawab permasalahan yang ada pada dataset transjakarta
- distribusi data
- hubungan antar data
- visualisasi data
- menemukan insight pada data



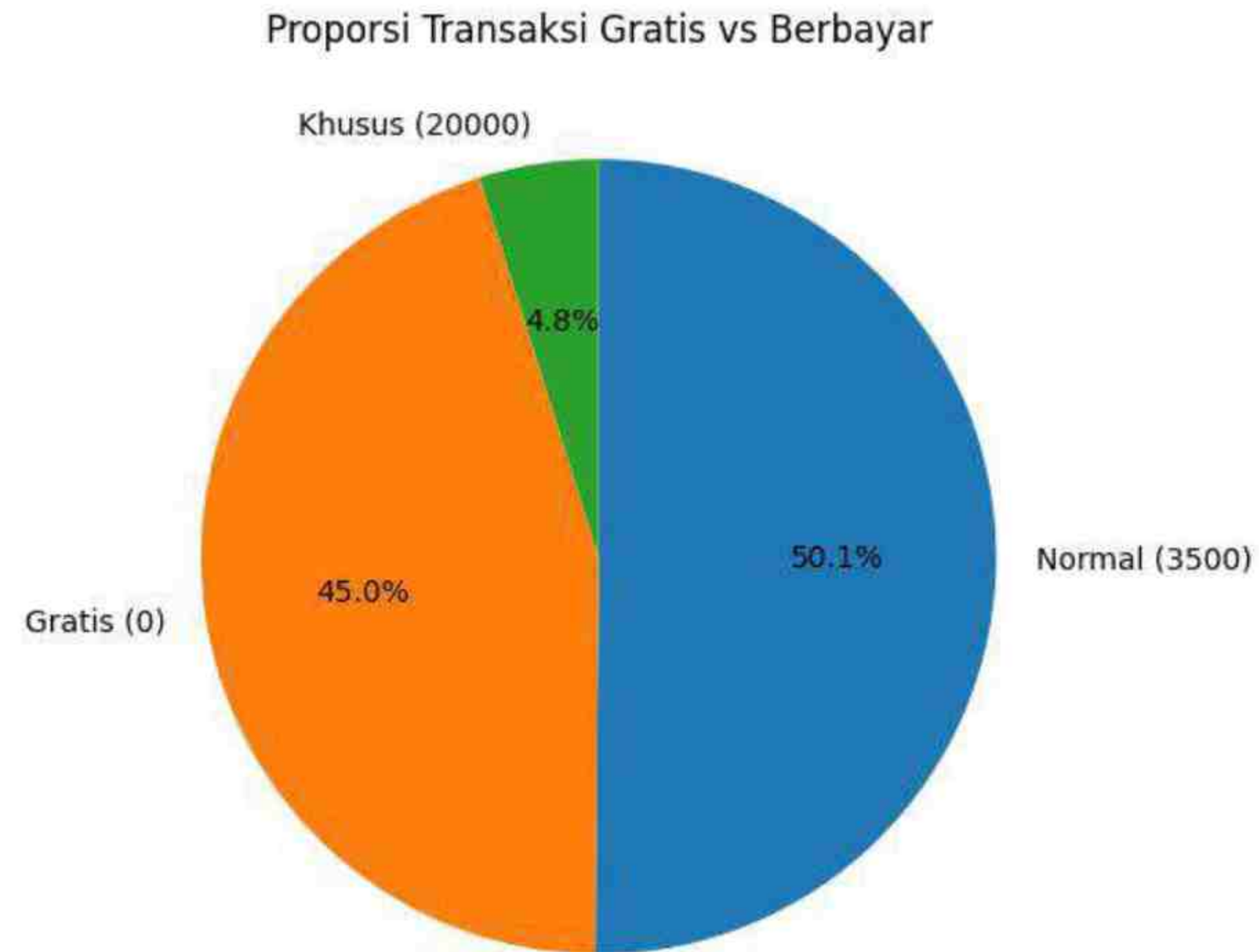
Exploratory Data Analysis (EDA)

1. Rute terpadat vs sepi



Exploratory Data Analysis (EDA)

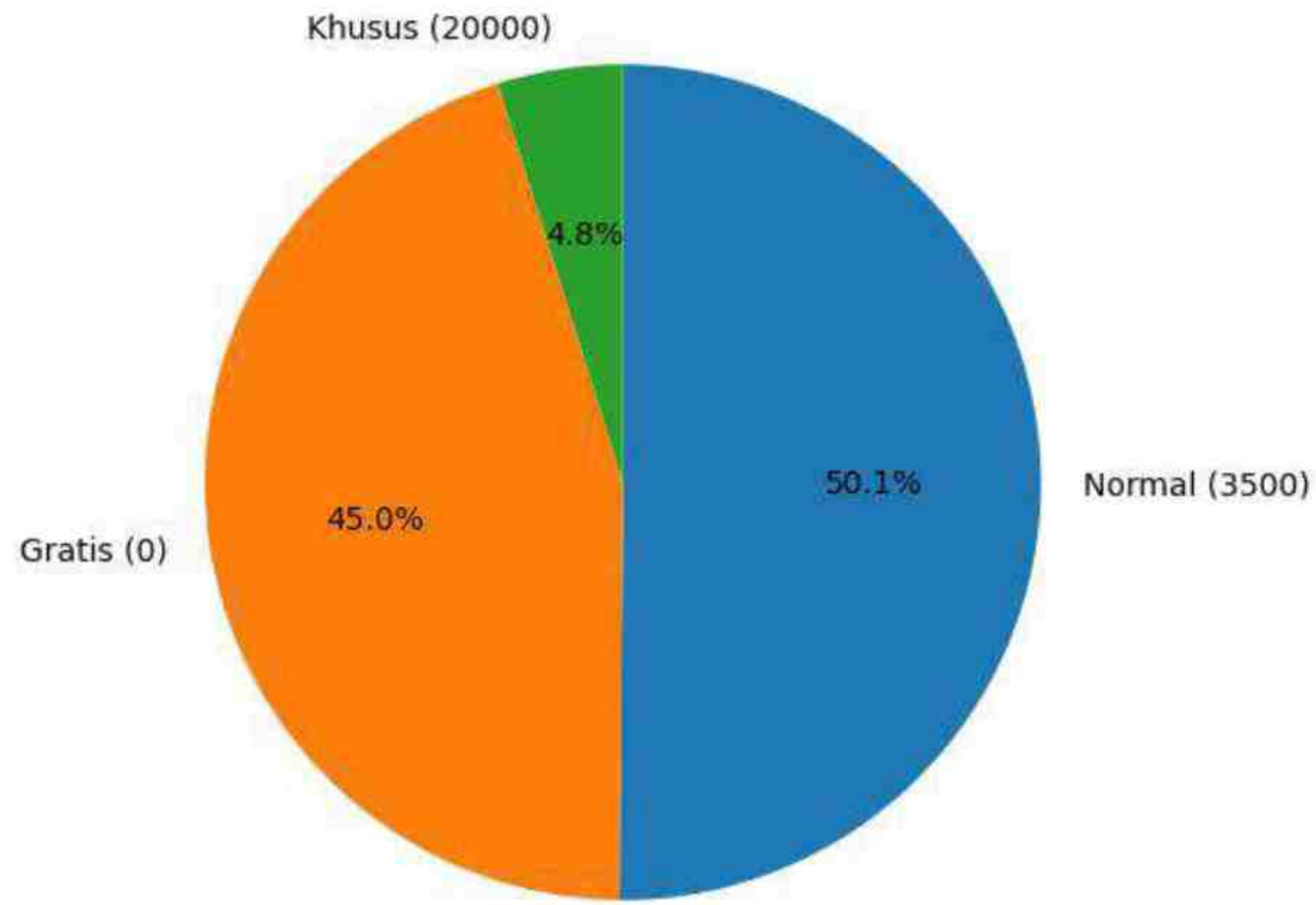
2. Proporsi Subsidi Gratis Vs Bayar



Exploratory Data Analysis (EDA)

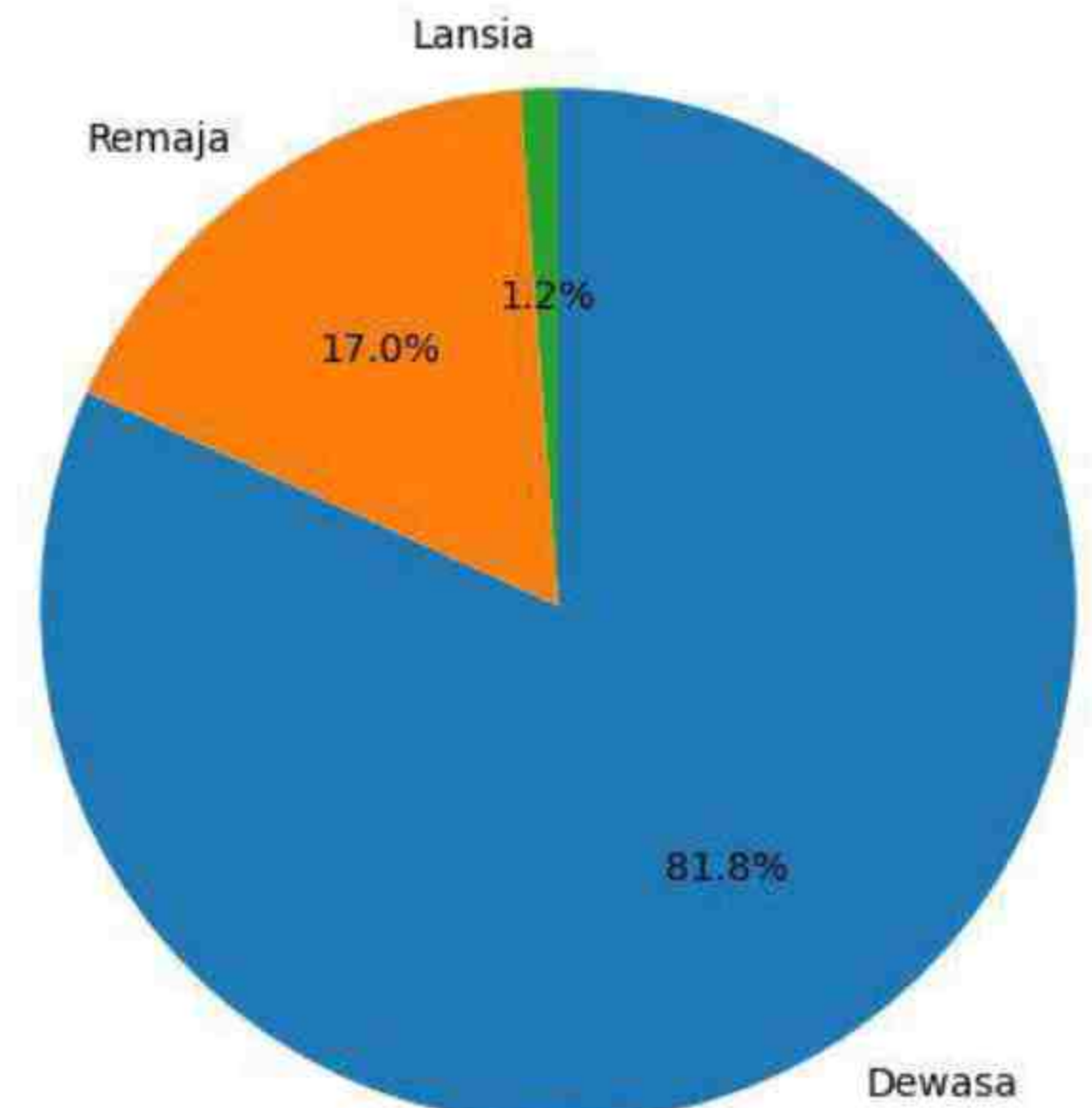
2. Proporsi Subsidi Gratis Vs Bayar

Proporsi Transaksi Gratis vs Berbayar



kenapa bisa ada transaksi free dan bayar?

Proporsi Kategori Usia Penerima Subsidi (Gratis)



Exploratory Data Analysis (EDA)

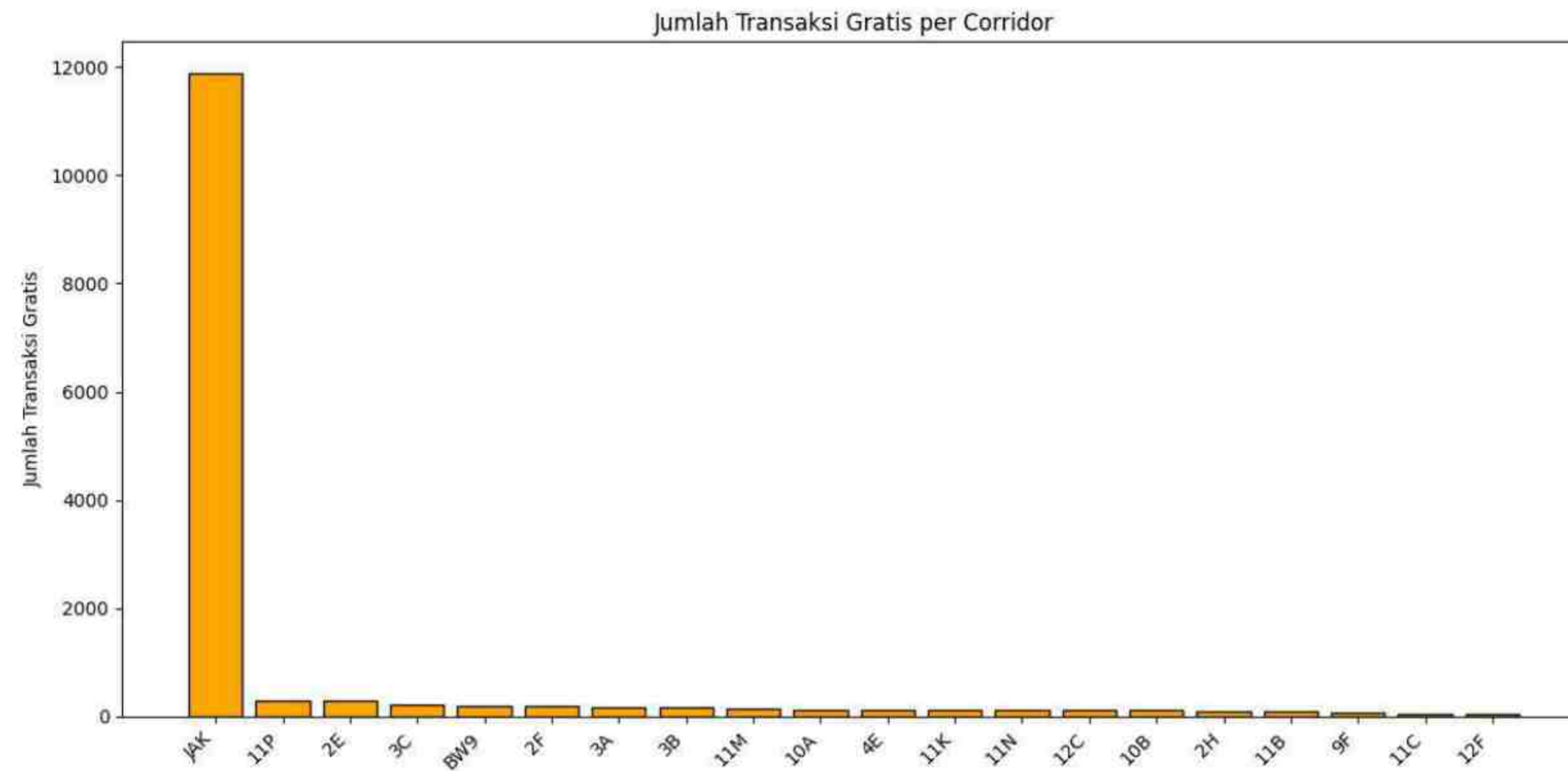
2. Proporsi Subsidi Gratis Vs Bayar

kenapa dewasa paling dominan mendapatkan subsidi, apakah benar semuanya karena pengguna bank DKI, atau PNS sehingga mendapatkan subsidi?

	payCardBank	bni	brizzi	dki	emoney	flazz	online
age_category							
Dewasa	643	890	6577	1946	889	1041	
Lansia	4	0	123	35	5	5	
Remaja	356	309	1040	506	124	163	

Exploratory Data Analysis (EDA)

3. Koridor Payamount Free



terdapat pola yang konsisten ketika data dilihat berdasarkan corridor ID, payAmount = 0, serta corridorName.

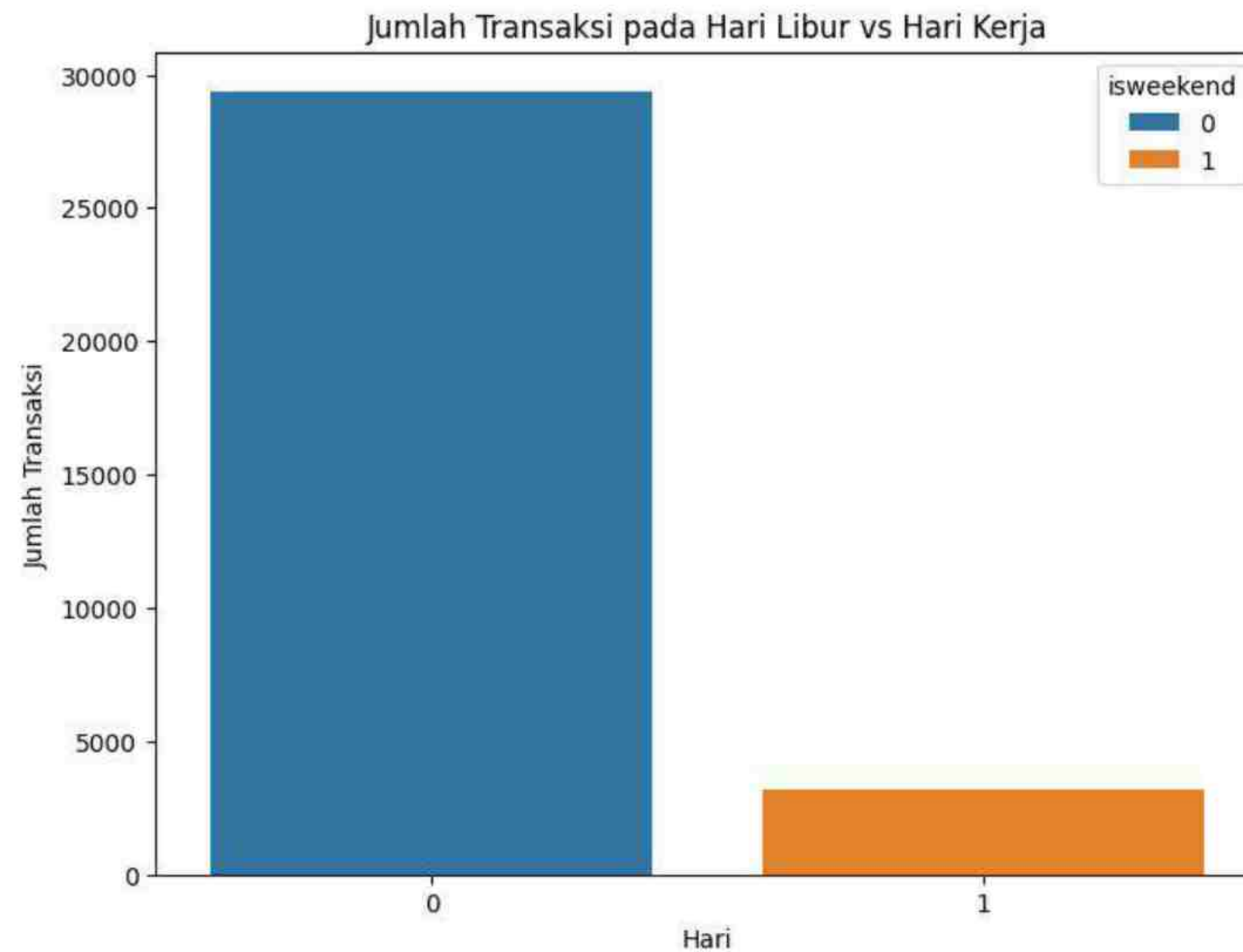
Corridor ID '**JAK**' merupakan layanan mikrotrans yang memang digratiskan.

Corridor ID '**BW9**' adalah bus pariwisata dengan rute Monas – Pantai Indah Kapuk, yang juga termasuk kategori bus gratis.

Sementara itu, corridor ID yang berawalan angka menunjukkan rute menuju **rumah susun (Rusun)**. Berdasarkan informasi dari Transjakarta, rute ini diasumsikan mendapat subsidi, karena ditujukan bagi para penghuni Rusun.

Exploratory Data Analysis (EDA)

4. Weekend vs Weekdays



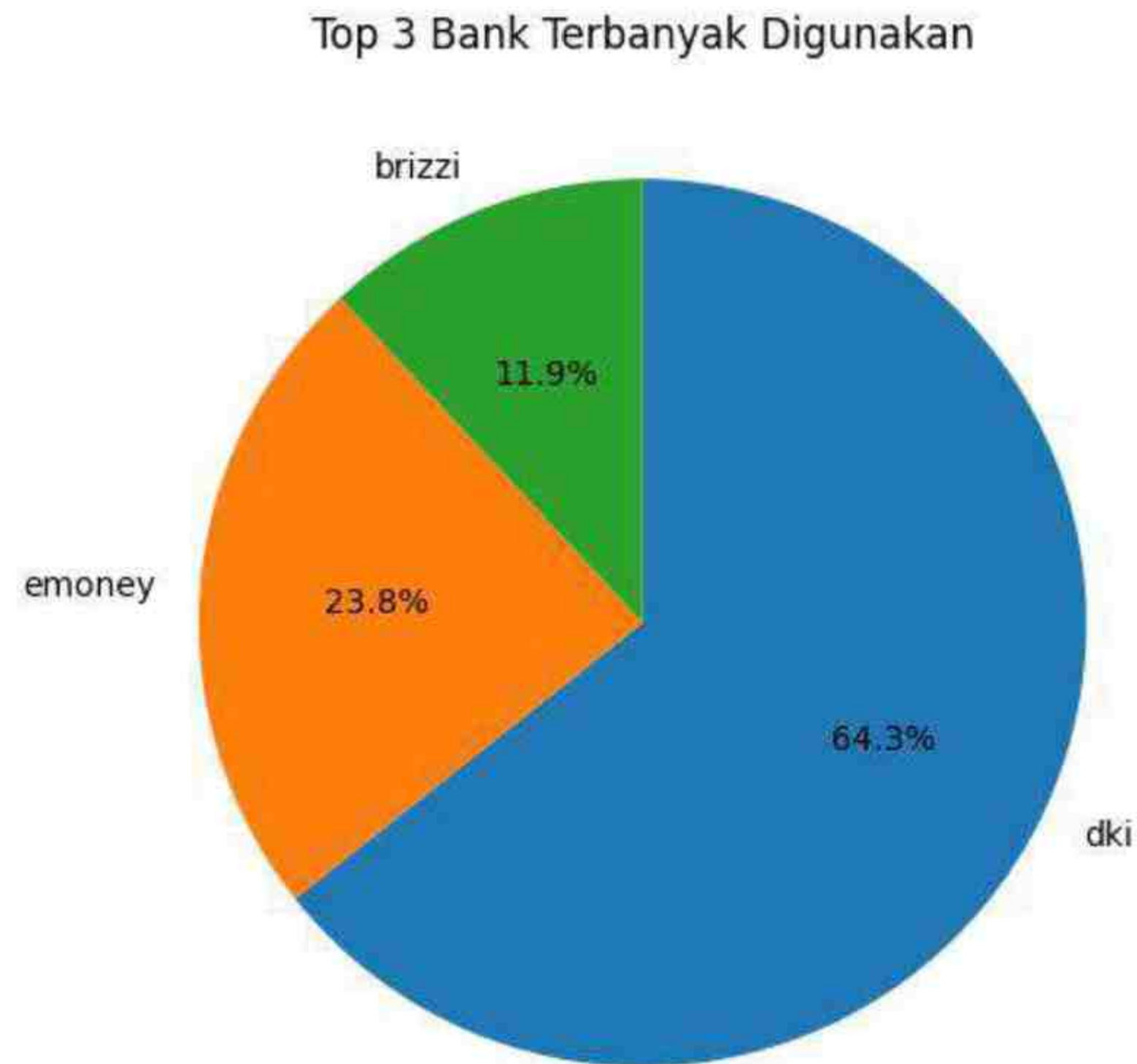
Notes:

0 → Weekdays

1 → Weekend

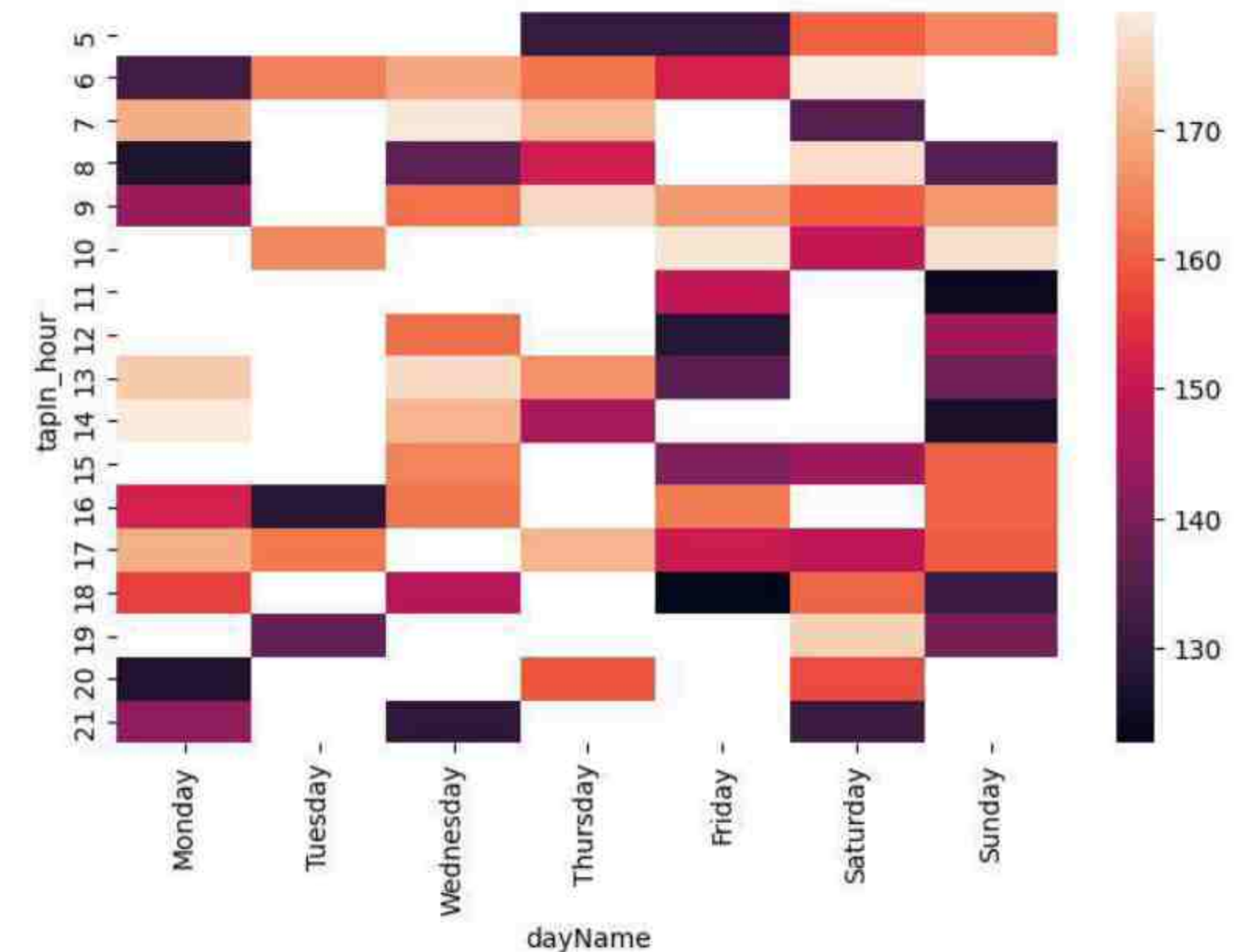
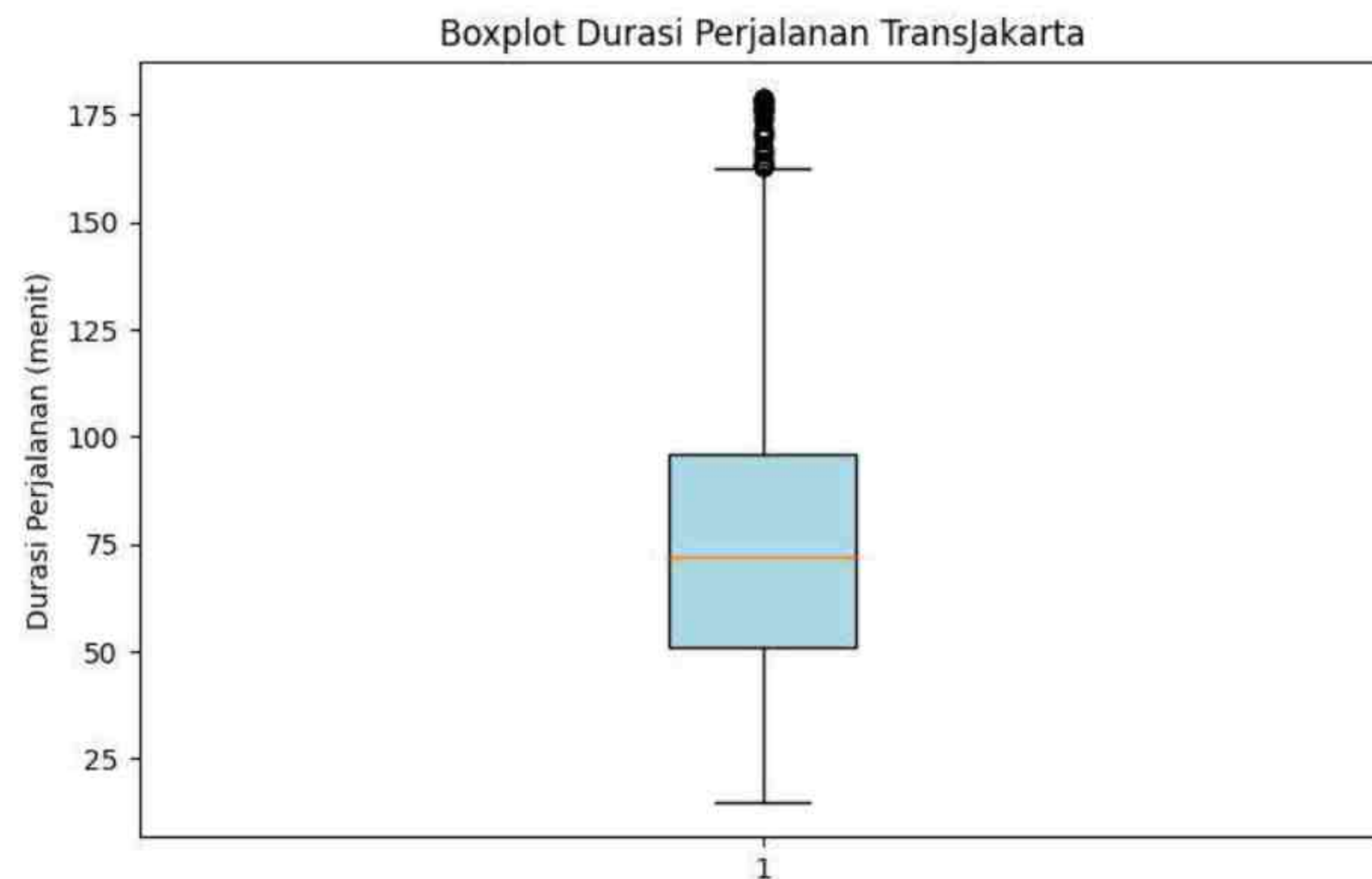
Exploratory Data Analysis (EDA)

4. PayCard yang paling banyak digunakan



Exploratory Data Analysis (EDA)

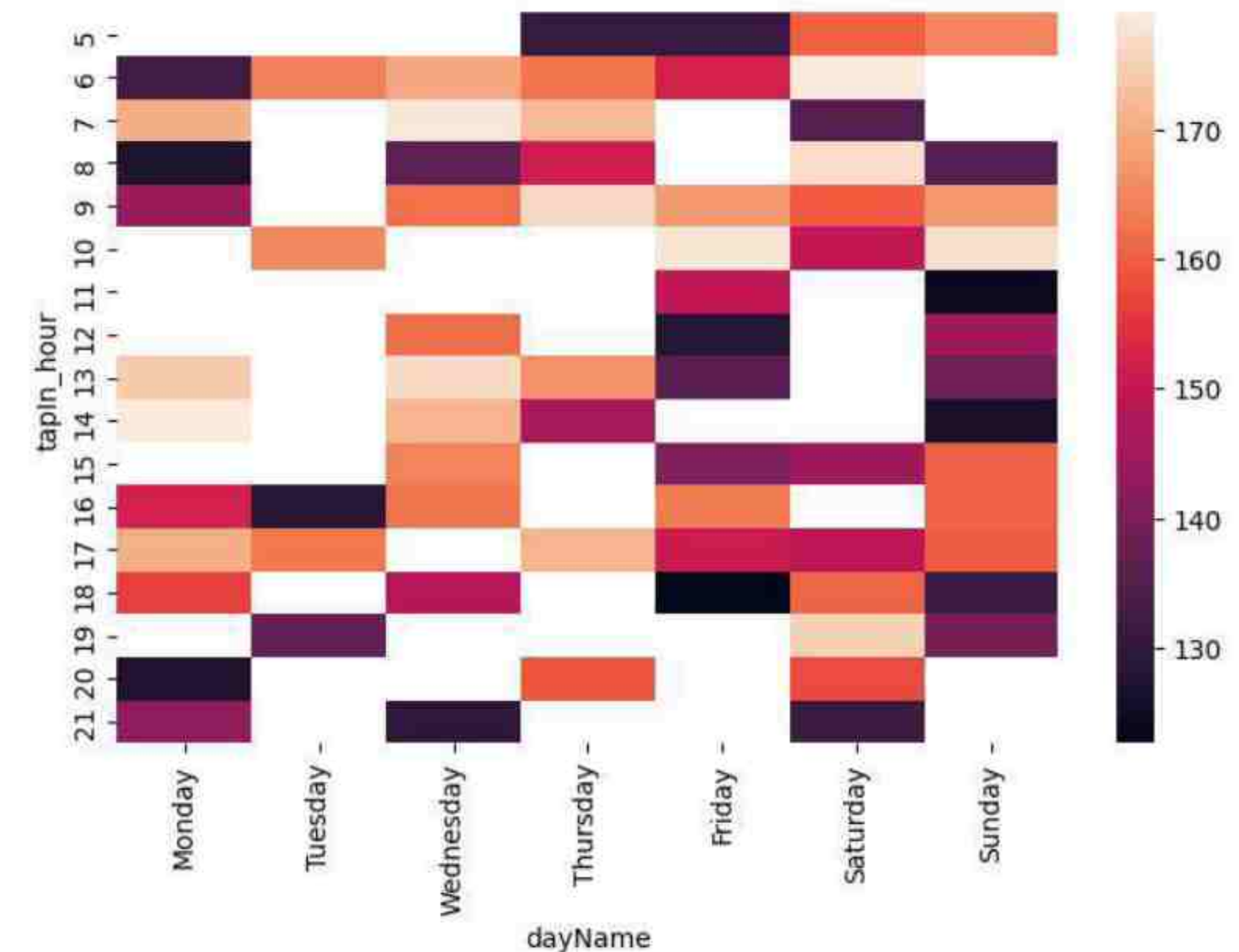
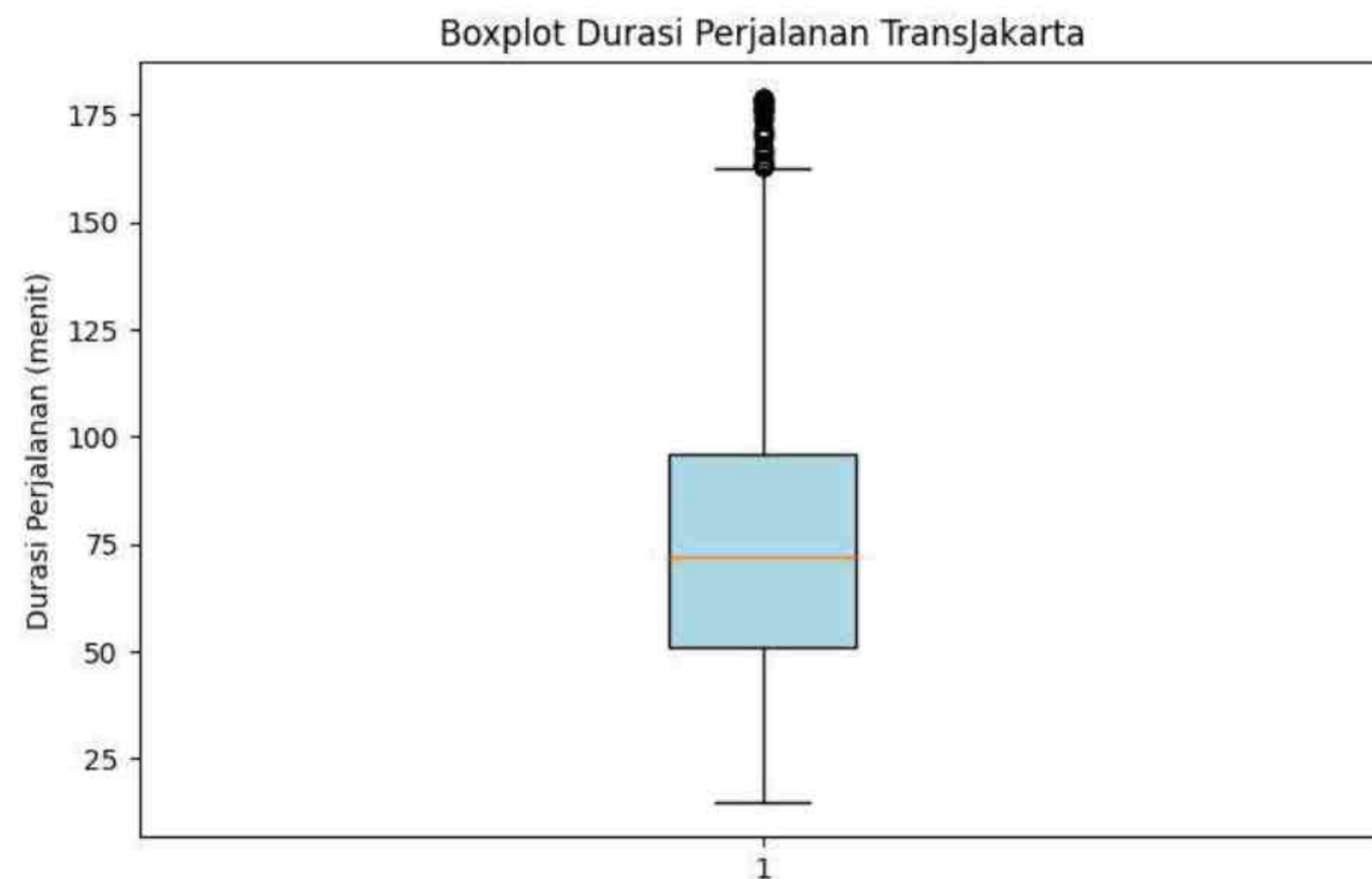
Distribusi Durasi Perjalanan



- Rata-rata durasi perjalanan yang sangat panjang ini paling banyak terjadi pada jam 14 sampai 17 (sore hari) terutama di hari kerja (Senin sampai Jumat).
- Warna terang pada heatmap mengindikasikan bahwa pengguna yang memulai perjalanan pada jam dan hari tersebut mengalami durasi perjalanan yang jauh lebih lama dari biasanya.
- ada potensi kemacetan atau gangguan lalu lintas pada waktu dan hari tersebut, namun dengan keterbatasan data, penyebab pasti dari durasi perjalanan yang lama tidak dapat diidentifikasi secara langsung.

Exploratory Data Analysis (EDA)

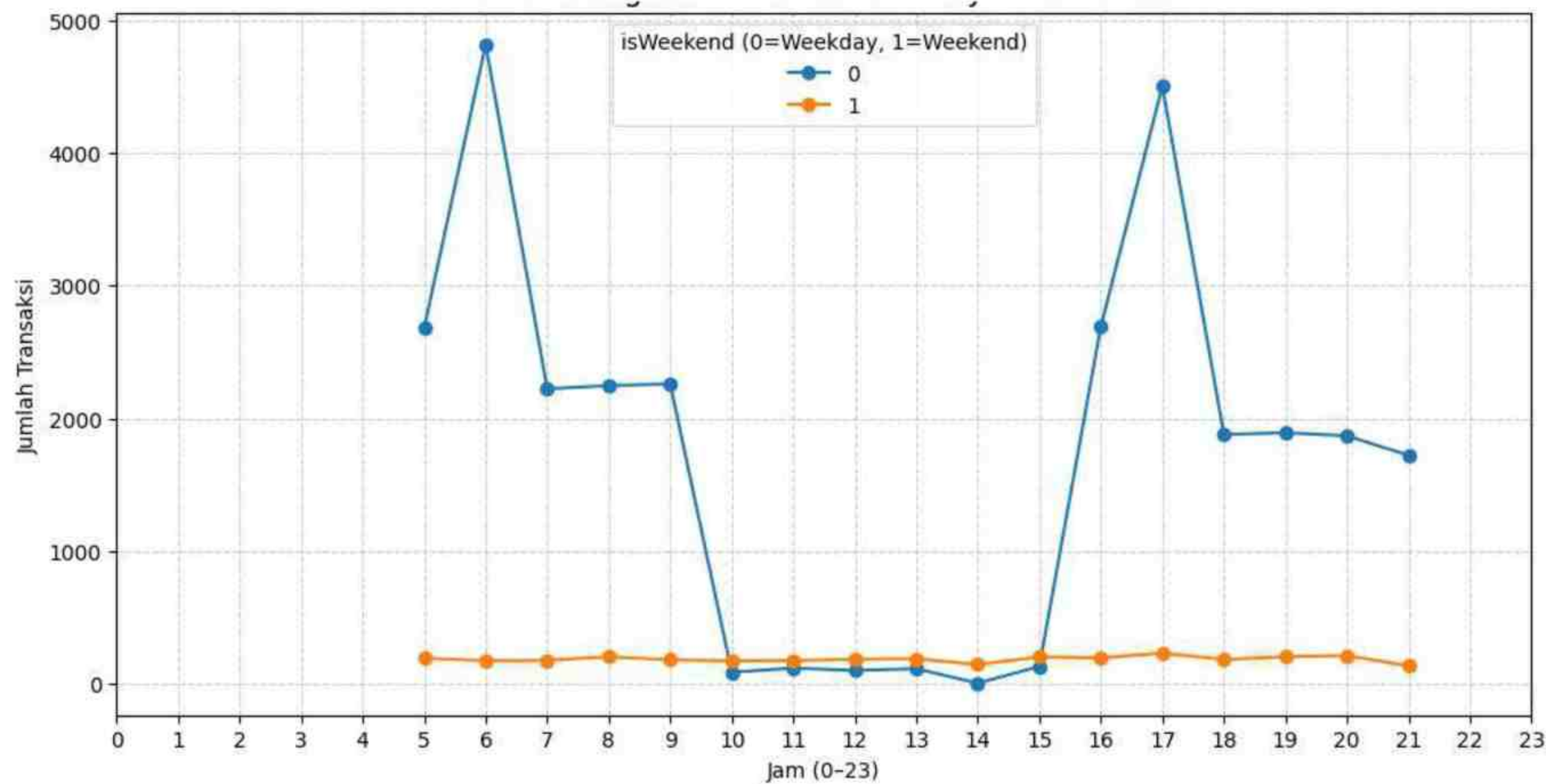
Distribusi Durasi Perjalanan



- Rata-rata durasi perjalanan yang sangat panjang ini paling banyak terjadi pada jam 14 sampai 17 (sore hari) terutama di hari kerja (Senin sampai Jumat).
- Warna terang pada heatmap mengindikasikan bahwa pengguna yang memulai perjalanan pada jam dan hari tersebut mengalami durasi perjalanan yang jauh lebih lama dari biasanya.
- ada potensi kemacetan atau gangguan lalu lintas pada waktu dan hari tersebut, namun dengan keterbatasan data, penyebab pasti dari durasi perjalanan yang lama tidak dapat diidentifikasi secara langsung.

Exploratory Data Analysis (EDA)

5. Peak Hours



weekday → Lonjakan tajam di jam **6–7 pagi** → kemungkinan besar orang berangkat kerja/sekolah.

Lonjakan lagi di jam **17–18 sore** → jam pulang kantor/sekolah.

weekend → Tidak ada lonjakan signifikan di jam tertentu.

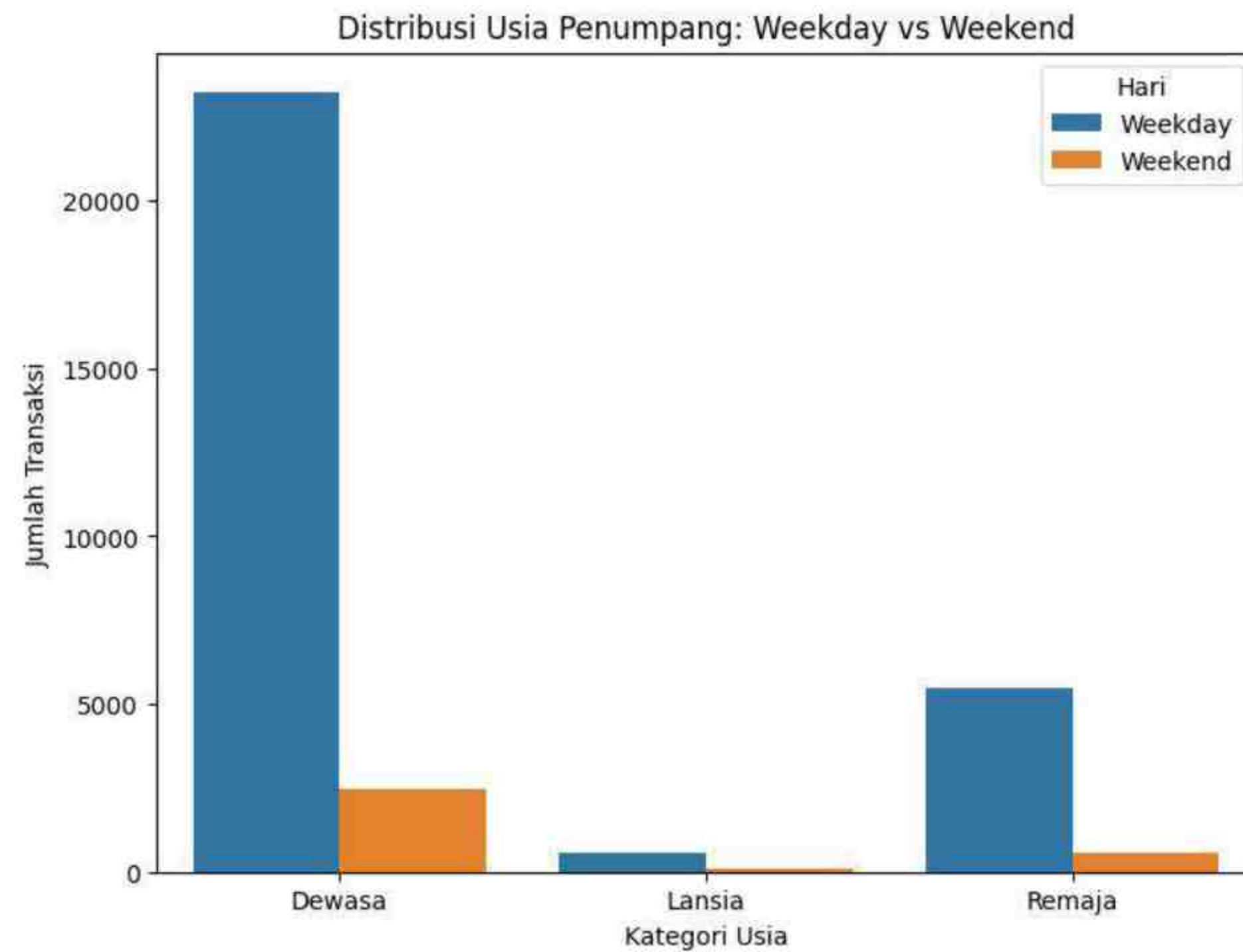
Exploratory Data Analysis (EDA)

6. Proporsi Gender Pengguna Transjakarta



Uji Inferensial

1. Apakah distribusi usia penumpang berbeda antara weekday & weekend?



Hasil Uji Chi-Square Usia × Hari

Chi-Square = 2.4687191367892414

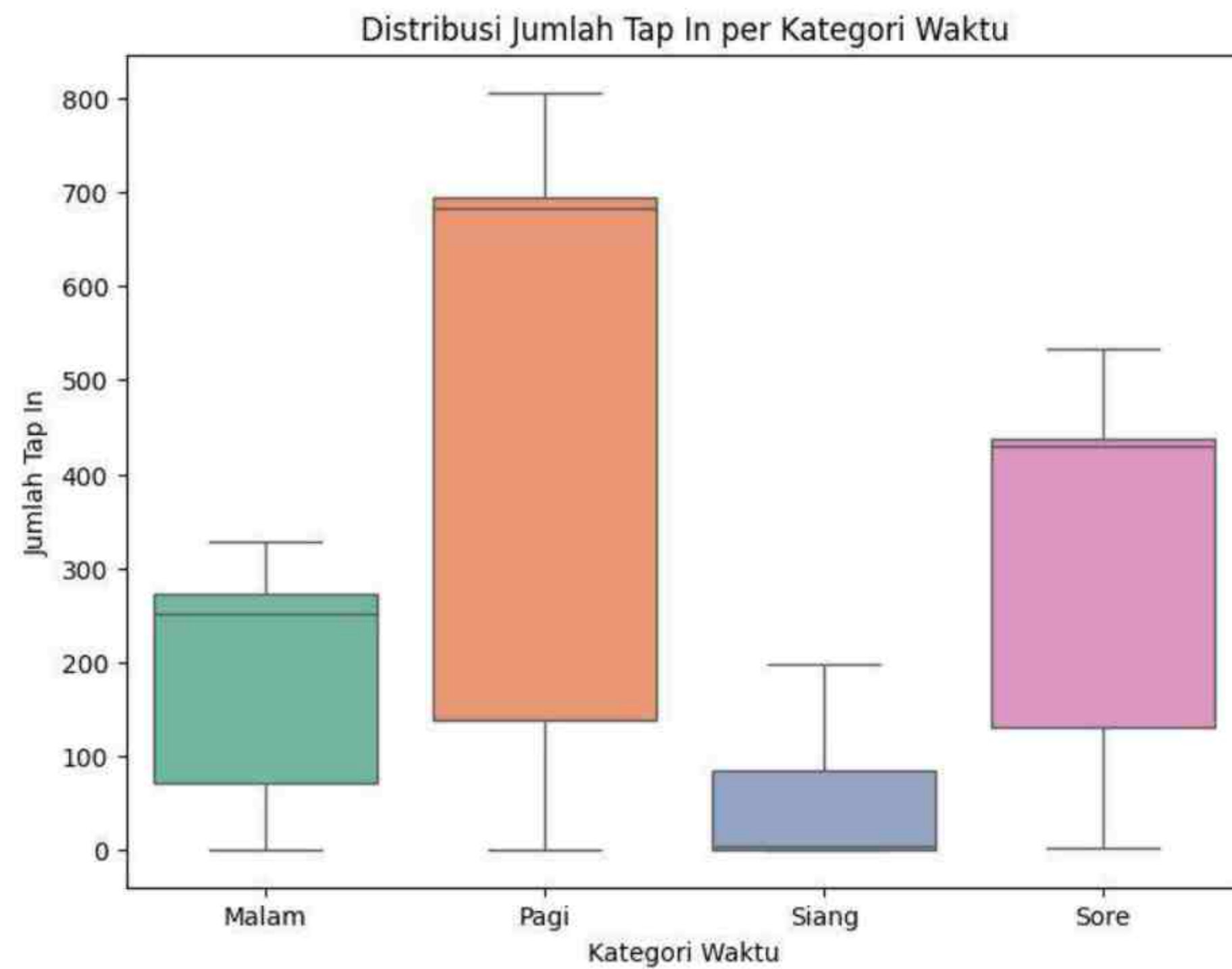
p-value = 0.29102108179767255

Degrees of freedom = 2

Tidak ada perbedaan signifikan distribusi usia antara weekday dan weekend.

Uji Inferensial

2. apakah ada perbedaan signifikan rata2 jumlah tapin disemua koridor (pagi/siang/sore/malam?)



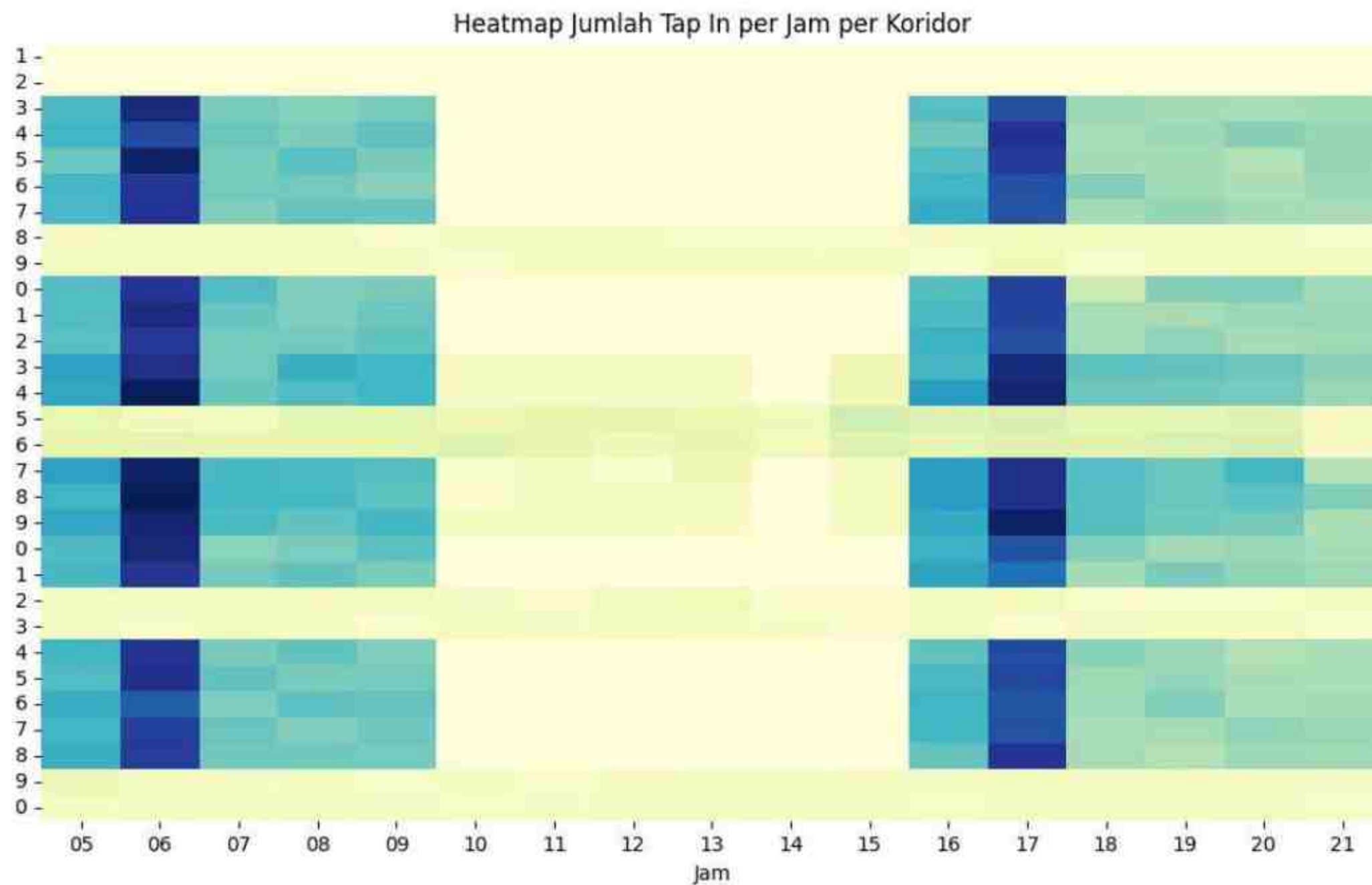
ANOVA & Kruskal-Wallis → ada perbedaan signifikan rata-rata/median jumlah tap-in antar kategori waktu.

Post-hoc Tukey →

- jumlah tap in penumpang paling tinggi terjadi pada pagi hari, diikuti oleh sore hari.
- Hal ini menggambarkan bahwa jam sibuk utama pengguna Transjakarta adalah pada waktu berangkat kerja/sekolah di pagi hari serta saat pulang di sore hari. Sementara itu,** pada siang dan malam hari, aktivitas penumpang relatif lebih rendah.

Uji Inferensial

3. apakah ada perbedaan rata-rata penumpang di kedua jam sibuk (6 pagi) dan (17 sore)?



T-statistic: 0.11553675745737003

p-value: 0.9084404899555163

Gagal tolak H_0 → Tidak ada perbedaan signifikan rata-rata jumlah penumpang di pagi dan sore

Man whitney :

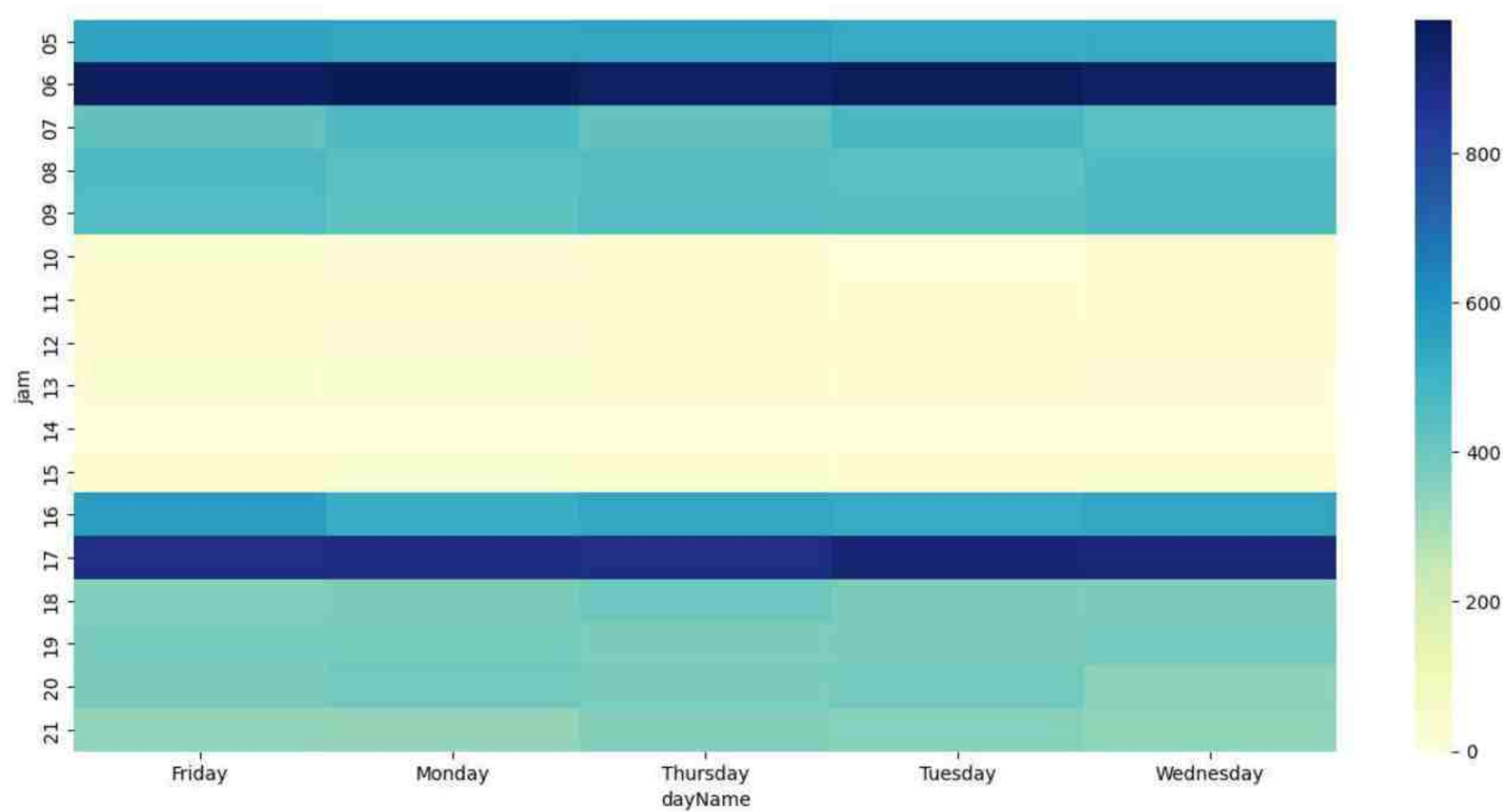
U-statistic: 471.0

p-value: 0.30310730226427895

Gagal tolak H_0 → Tidak ada perbedaan signifikan rata-rata jumlah penumpang di pagi dan sore

Uji Inferensial

4. hari yang paling sibuk



Hasil ANOVA

F-statistic: 0.00022927143543323238

p-value: 0.9999998922755303

Gagal tolak H_0 → Tidak ada perbedaan signifikan rata-rata jumlah penumpang antara hari-hari di weekdays

Hasil Kruskal-Wallis

H-statistic: 0.05496045218532316

p-value: 0.9996292651708816

Gagal tolak H_0 → Tidak ada perbedaan signifikan rata-rata jumlah penumpang antara hari-hari di weekdays

Kesimpulan Analisis



1. Profil Penumpang

- Mayoritas: dewasa produktif, rata-rata usia 33 tahun
- Metode pembayaran dominan: Bank DKI

2. Pola Penggunaan

- Lebih banyak di weekdays daripada weekend
- Rute padat: Cibubur–Balaikota & Ciputat
- Rute sepi: Kampung Rambutan–Blok M (± 17 /hari)

3. Subsidi & Koridor Gratis

- Gratis vs membayar hampir seimbang
- Penerima subsidi dominan dewasa; lansia $\pm 1\%$
- Koridor khusus: JAK (mikrotrans), BW9 (bus wisata), rusun

4. Pola Jam Sibuk

- Puncak: pagi (06–09) → signifikan lebih tinggi
- Sepi: siang (11–14) → signifikan lebih rendah
- Sore & malam relatif serupa

Rekomendasi Bisnis



1. Armada & Jadwal

- Tambah armada di jam sibuk pagi (06.00) & sore (17.00) pada weekdays.
- Evaluasi rute sepi → lakukan penyesuaian jadwal.

2. Promosi & Tarif

- Terapkan strategi promosi / tarif khusus di weekend untuk menarik penumpang.

3. Gender-Based Space

- Komposisi penumpang: 53% perempuan, 46% laki-laki.
- Optimalkan pembagian ruang 50:50 untuk kenyamanan & keamanan gender.

4. Alokasi Koridor Prioritas

- Fokus pada koridor dengan aktivitas tinggi di jam sibuk sebagai prioritas utama alokasi sumber daya.

Thanks



Hani Handayani

