

Analysis of Public Transportation Transaction Transjakarta

Prepared by Dyah Kusuma Ningrum

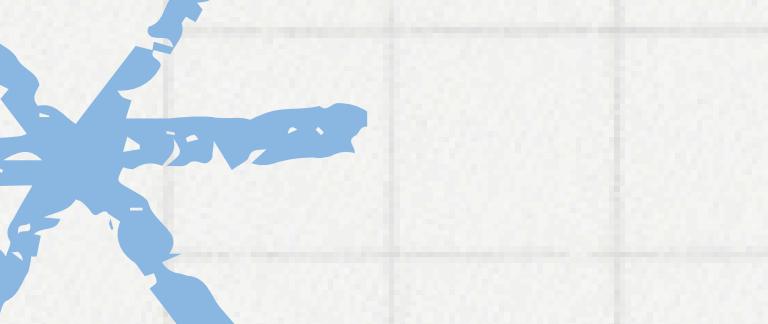


Overview

Dataset yang disediakan berisi catatan transaksi dari sistem transportasi Transjakarta. Dataset terdiri dari variabel-variabel berikut:

1. Transaction ID (transID), Fare Collection ID (payCardID), Corridor ID (corridorID), and Payment Amount (payAmount).
2. Demografi Pengguna/Penumpang, seperti payee's bank, name, sex, dan birth date.
3. Rincian Perjalanan Penumpang, seperti corridor name, direction, tap-in and tap-out stops, and times.

Tujuan dari analisis ini adalah untuk mengetahui pendapatan Transjakarta dengan menganalisis transaksi dari berbagai koridor.



Problem Statement

01.

Apa saja yang termasuk dalam top 10 koridor berdasarkan total pendapatan tertinggi?

02.

Bagaimana distribusi dari jumlah pembayaran yang ada?

03.

Bagaimana jumlah transaksi berdasarkan koridor untuk top 10 koridor berdasarkan total pendapatan tertinggi?



Data Analysis



Data Understanding

Dataset terkait terdiri dari 37.900 baris dan 22 kolom. Berikut adalah beberapa poin penting yang didapatkan dari overview data di atas:

Key columns:

1. transID: ID transaksi unik.
2. payCardID, payCardBank, payCardName: Rincian kartu pembayaran.
3. tapInStops, tapInStopsName, tapOutStops, tapOutStopsName: ID dan nama pemberhentian untuk naik dan turun bus.
4. tapInTime, tapOutTime: Waktu untuk naik dan turun bus.
5. payAmount: Jumlah yang dibayar untuk perjalanan.

Issue:

Terdapat missing values pada beberapa kolom seperti corridorName, corridorID, dan payAmount.

Data Summary:

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 37900 entries, 0 to 37899  
Data columns (total 22 items):  
 #   Column          Non-Null Count  Dtype     
---  --  
 0   transID         37900 non-null  int64    
 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  object    
 6   corridorID     37900 non-null  int64    
 7   corridorName    37900 non-null  object    
 8   direction        37900 non-null  object    
 9   tapInStops      37900 non-null  int64    
 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  float64  
 15  tapOutStops     37900 non-null  int64    
 16  tapOutStopsName 37900 non-null  object    
 17  tapOutStopsLat  37900 non-null  float64  
 18  tapOutStopsLon  37900 non-null  float64  
 19  stopEndSeq      37900 non-null  int64    
 20  tapOutTime      37900 non-null  float64  
 21  payAmount        37900 non-null  int64    
dtypes: int64(13), object(9)  
memory usage: 7.9 MB
```

Missing Values Per Column:

transID	0
payCardID	0
payCardBank	0
payCardName	0
payCardSex	0
payCardBirthDate	0
corridorID	1257
corridorName	1930
direction	0
tapInStops	1213
tapInStopsName	0
tapInStopsLat	0
tapInStopsLon	0
stopStartSeq	0
tapInTime	0
tapOutStops	2289
tapOutStopsName	1344
tapOutStopsLat	1344
tapOutStopsLon	1344
stopEndSeq	1344
tapOutTime	1344
payAmount	1007

Data Cleansing & Manipulation

```
Missing Values After Cleaning:  
transID          0  
payCardID        0  
payCardBank      0  
payCardName      0  
payCardSex        0  
payCardBirthDate 0  
corridorID        0  
corridorName      0  
direction         0  
tapInStops       1111  
tapInStopsName    0  
tapInStopsLat     0  
tapInStopsLon     0  
stopStartSeq      0  
tapInTime         0  
tapOutStops      2099  
tapOutStopsName   1226  
tapOutStopsLat    1226  
tapOutStopsLon    1226  
stopEndSeq        1226  
tapOutTime        1226  
payAmount          0  
dtype: int64
```

Key points:

1. Baris yang memiliki missing values pada kolom seperti corridorName, corridorID, dan payAmount telah dihapus. Dengan begitu, hanya data yang lengkap dan relevan yang tetap ada dalam dataset, sehingga analisis yang dilakukan lebih akurat.
2. Duplikasi transaksi berdasarkan transID telah dihapus, sehingga setiap transaksi hanya tercatat sekali dalam dataset.

Data Types & Formatting

```
Data Types After Formatting:  
transID          object  
payCardID       float64  
payCardBank     object  
payCardName     object  
payCardSex      object  
payCardBirthDate int64  
corridorID      object  
corridorName    object  
direction        int64  
tapInStops      object  
tapInStopsName  object  
tapInStopsLat   float64  
tapInStopsLon   float64  
stopStartSeq    int64  
tapInTime        datetime64[ns]  
tapOutStops     object  
tapOutStopsName object  
tapOutStopsLat  float64  
tapOutStopsLon  float64  
stopEndSeq      float64  
tapOutTime       datetime64[ns]  
payAmount        float64  
dtype: object
```

Key point:

1. Dari hasil tipe data setelah konversi, kita dapat memastikan bahwa kolom berisi informasi waktu telah ter-format dengan benar dan dapat digunakan untuk analisis lebih lanjut, seperti menghitung durasi perjalanan atau mengidentifikasi pola waktu.

Descriptive Analysis

Descriptive Statistics (Numerical):					tapOutStopsLat	tapOutStopsLon	stopEndSeq	\
count	3.488300e+04	34883.000000	34883.000000	34883.000000	33657.000000	33657.000000	33657.000000	
mean	4.189236e+17	1990.150790	0.500960	-6.214807	-6.214680	106.84150	21.318626	
min	6.040368e+10	1946.000000	0.000000	-6.394973	-6.394973	106.61473	1.000000	
25%	1.800480e+14	1982.000000	0.000000	-6.245990	-6.248460	106.80081	11.000000	
50%	3.507250e+15	1991.000000	1.000000	-6.214145	-6.214459	106.83489	18.000000	
75%	4.684820e+15	2001.000000	1.000000	-6.174903	-6.174360	106.88491	30.000000	
max	4.997690e+18	2012.000000	1.000000	-6.089429	-6.091746	107.02366	77.000000	
std	1.314201e+18	13.061228	0.500006	0.058490	0.059563	0.06172	13.894385	
					tapInTime	tapOutTime	payAmount	
count	34883.000000	34883.000000		34883		33657	34883.000000	
mean	106.841795	13.635754	2023-04-16 08:04:43.651062272		2023-04-16 09:15:02.920046336		2704.626896	
min	106.614730	0.000000	2023-04-01 06:22:00		2023-04-01 07:27:00		0.000000	
25%	106.801940	4.000000	2023-04-10 09:02:00		2023-04-10 10:21:00		0.000000	
50%	106.835300	10.000000	2023-04-16 17:15:00		2023-04-16 18:44:00		3500.000000	
75%	106.883630	19.000000	2023-04-22 05:09:00		2023-04-21 23:10:00		3500.000000	
max	107.023950	68.000000	2023-04-30 21:55:00		2023-04-30 23:23:00		20000.000000	
std	0.061047	12.301944		Nan		Nan	4221.111470	

Descriptive Analysis

Descriptive Statistics (Categorical):

	transID	payCardBank	payCardName	payCardSex	corridorID	\
count	34883	34883	34883	34883	34883	
unique	34883	6	1947	2	216	
top	EIIW227B8L34VB	dki	Suci	Wacana	F	1T
freq	1	17341	80	18579	391	

corridorName tapInStops tapInStopsName tapOutStops \

count	34883	33772	34883	32784
unique	211	2490	2520	2165
top	Cibubur - Balai Kota	P00170	Penjaringan	P00016
freq	391	225	232	282

tapOutStopsName

count	33657
unique	2183
top	BKN
freq	292

Descriptive Analysis

Payment Amount Summary:

```
count      34883.000000
mean       2704.626896
std        4221.111470
min        0.000000
25%        0.000000
50%        3500.000000
75%        3500.000000
max       20000.000000
```

Name: payAmount, dtype: float64

Top 5 Corridors by Transaction Count:

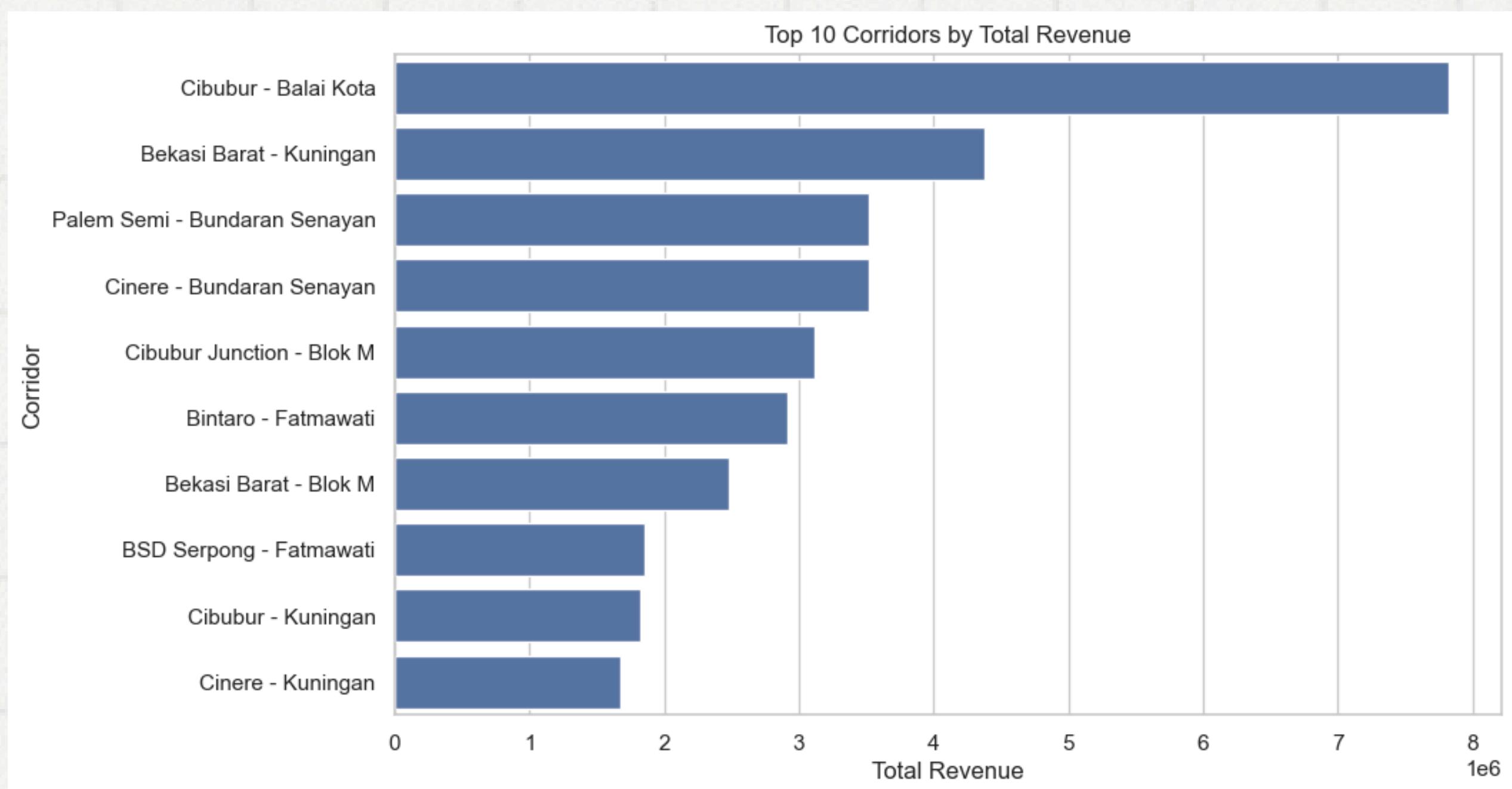
corridorName	count
Cibubur - Balai Kota	391
Ciputat - CSW	383
Pulo Gadung - Monas	339
Harmoni - Jakarta International Stadium	337
Kebayoran Lama - Tanah Abang	333

Name: count, dtype: int64

Data Visualization



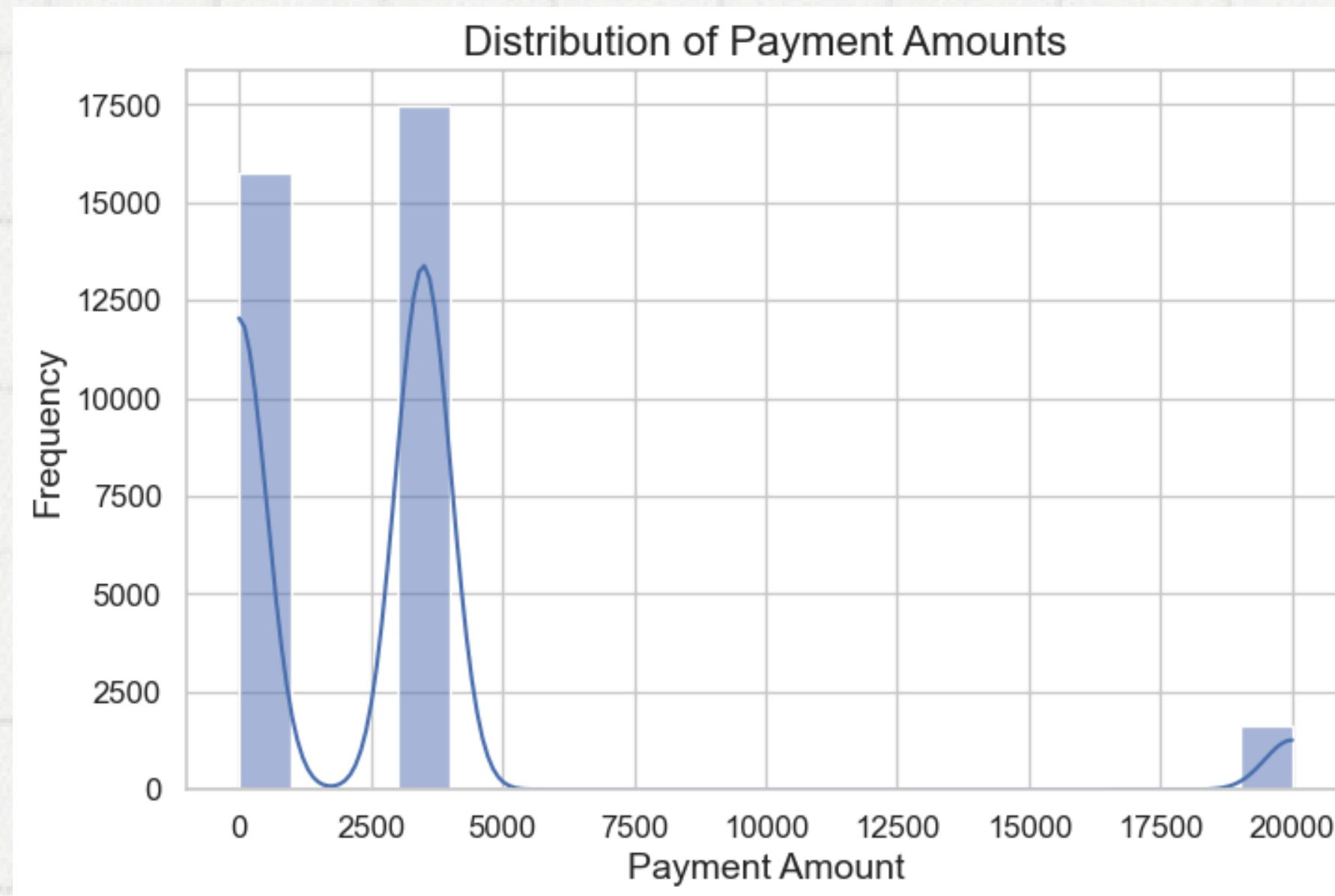
1. Top 10 Corridors by Total Revenue



Insight :

Top 10 koridor adalah seperti terlampir pada barplot di atas dengan total pendapat berkisar di antara 1,6 juta sampai dengan 7,8 juta dari masing-masing koridor.

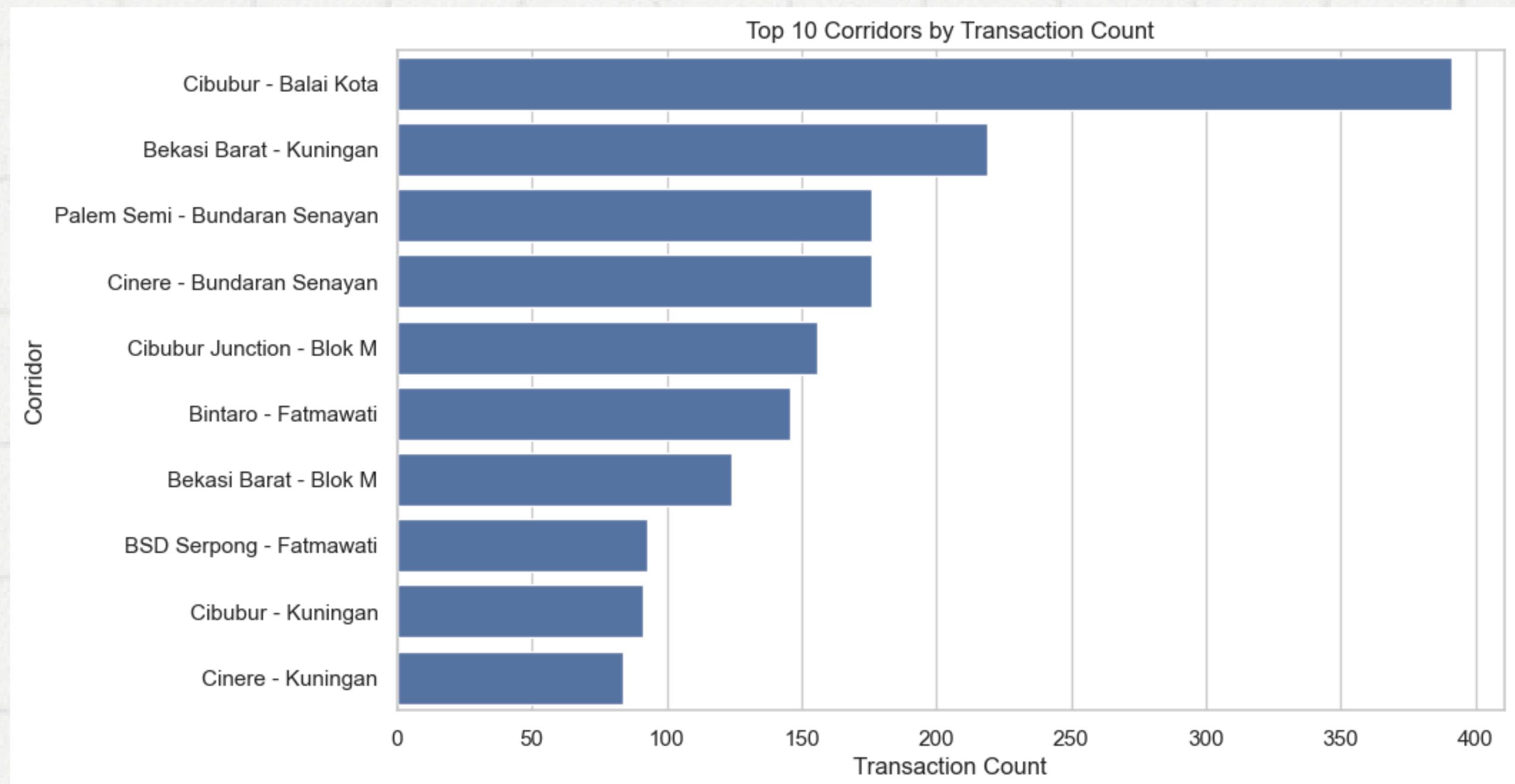
2. Distribution of Payment Amounts



Insight :

1. Berdasarkan hasil yang disajikan, tarif cenderung berbeda-beda, namun perbedaan tidak terlalu banyak karena hanya tersedia 2 jenis tarif.

3. Transaction Count by Corridor for the Top 10 Corridors



Insight :

Total transaksi dari top 10 koridor yang sudah ditampilkan adalah berkisar antara 84 sampai dengan 391 transaksi dari masing-masing koridor.

Conclusion

High-Revenue Corridors

Koridor dengan total pendapatan tertinggi menunjukkan banyak peminat/banyak digunakan penumpang dan menjadi penyumbang utama pendapatan untuk sistem transportasi. Top 10 koridor dengan pendapatan tertinggi adalah:

corridorName	total_revenue	average_revenue
Cibubur - Balai Kota	7820000.0	20000.0
Bekasi Barat - Kuningan	4380000.0	20000.0
Palem Semi - Bundaran Senayan	3520000.0	20000.0
Cinere - Bundaran Senayan	3520000.0	20000.0
Cibubur Junction - Blok M	3120000.0	20000.0
Bintaro - Fatmawati	2920000.0	20000.0
Bekasi Barat - Blok M	2480000.0	20000.0
BSD Serpong - Fatmawati	1860000.0	20000.0
Cibubur - Kuningan	1820000.0	20000.0
Cinere - Kuningan	1680000.0	20000.0

Conclusion

Distribution of Payment Amounts

Dengan melihat payAmount, kita dapat mengetahui tarif yang dibayar di setiap koridor. Apabila tarif konsisten, berarti harganya standar. Apabila tarif berbeda-beda, kemungkinan karena jarak tempuh, jenis bus/armada, atau jenis tarif yang dipilih juga berbeda

Berdasarkan hasil yang disajikan, tarif cenderung berbeda-beda, namun perbedaan tidak terlalu banyak karena hanya tersedia 2 jenis tarif.

Conclusion

Total Transaction of High-Revenue Corridors

Koridor dengan jumlah transaksi tinggi menunjukkan rute yang sering dipakai. Biasanya, ini rute utama yang banyak dilalui penumpang atau area yang ramai. Berikut adalah jumlah transaksi dari top 10 koridor yang telah disebutkan:

corridorName	transaction_count
Cibubur - Balai Kota	391
Bekasi Barat - Kuningan	219
Palem Semi - Bundaran Senayan	176
Cinere - Bundaran Senayan	176
Cibubur Junction - Blok M	156
Bintaro - Fatmawati	146
Bekasi Barat - Blok M	124
BSD Serpong - Fatmawati	93
Cibubur - Kuningan	91
Cinere - Kuningan	84



Recommendation

1. Resource Allocation

Koridor dengan pendapatan tinggi dan volume tinggi harus diprioritaskan untuk alokasi sumber daya. Menambahkan lebih banyak bus/armada atau meningkatkan frekuensi layanan pada rute-rute ini dapat meningkatkan kepuasan penumpang dan mengurangi kepadatan.

2. Targeted Marketing and Promotions

Perlu dipertimbangkan strategi promosi untuk koridor dengan jumlah transaksi atau pendapatan rata-rata yang lebih rendah. Tarif yang lebih rendah selama jam-jam sepi dapat meningkatkan jumlah penumpang pada rute yang jarang digunakan.

3. Collaboration with External Parties

Meningkatkan kolaborasi dengan pihak eksternal (misal: sponsor) untuk koridor-koridor dengan volume tinggi, sehingga dapat menghasilkan pendapatan lain yang bisa digunakan untuk pembangunan koridor serta bus/armada transportasi itu sendiri.



Thank You!