

# In-Vehicle Coupon Recommendation

---

Dokumen  
Laporan Final  
Project

DEEP LEARNING 4.0



# LATAR BELAKANG MASALAH

Pemberian kupon merupakan salah satu cara untuk meningkatkan revenue dan mendapatkan pelanggan baru dalam sebuah bisnis. Berdasarkan pemberian kupon, 69% mereka akan merasa menjadi pelanggan yang pintar[2], 60% pelanggan akan mencoba produk baru[2], 54% melakukan pembelian impulsif[2], 88% pelanggan akan menggunakan kuponnya[3]. Hal tersebut menjadikan pemberian kupon salah satu metode untuk meningkatkan revenue dalam bisnis.

## **REFERENCE**

[1] The Pros & Cons Of Using Coupons For Your Business (investopedia.com)

[2] Valassis Research: Coupons and Discounts Appeal Far Beyond Savings - Valassis

[3] Coupon Statistics - Is Couponing Growing or Slowing? (spendmenot.com)

# PROBLEM STATEMENT

- Berdasarkan dari dataset yang dimiliki, hanya terdapat 56.84% pengemudi yang menerima kupon. Hal ini akan menghasilkan profit yang tidak optimal, karena jumlah pelanggan/pengemudi yang menerima kupon tidak sesuai ekspektasi awal/tidak memenuhi KPI campaign kupon tersebut.
- Langkah yang dapat dilakukan untuk menanggulangi hal tersebut ialah memprediksi tingkat *coupon acceptance* terhadap pelanggan/pengemudi.
- Membuat model yang dapat memprediksi dan menghasilkan rekomendasi (karakteristik pengemudi) yang tepat untuk menjadi *leads* sebagai penerima kupon



# EDA

## DESCRIPTIVE ANALYSIS

```

1 Dataset Info:
2 <class 'pandas.core.frame.DataFrame'>
3 RangeIndex: 12684 entries, 0 to 12683
4 Data columns (total 26 columns):
5 #   Column              Non-Null Count  Dtype
6 ---  -
7 0   destination         12684 non-null  object
8 1   passanger           12684 non-null  object
9 2   weather             12684 non-null  object
10 3   temperature         12684 non-null  int64
11 4   time                12684 non-null  object
12 5   coupon              12684 non-null  object
13 6   expiration          12684 non-null  object
14 7   gender              12684 non-null  object
15 8   age                 12684 non-null  object
16 9   maritalStatus       12684 non-null  object
17 10  has_children        12684 non-null  int64
18 11  education            12684 non-null  object
19 12  occupation           12684 non-null  object
20 13  income               12684 non-null  object
21 14  car                  108 non-null   object
22 15  Bar                  12577 non-null  object
23 16  CoffeeHouse         12467 non-null  object
24 17  CarryAway           12533 non-null  object
25 18  RestaurantLessThan20 12554 non-null  object
26 19  Restaurant20To50    12495 non-null  object
27 20  toCoupon_GEQ5min    12684 non-null  int64
28 21  toCoupon_GEQ15min   12684 non-null  int64
29 22  toCoupon_GEQ25min   12684 non-null  int64
30 23  direction_same       12684 non-null  int64
31 24  direction_opp        12684 non-null  int64
32 25  Y                    12684 non-null  int64
33 dtypes: int64(8), object(18)

```

```

36 Missing Values:
37 destination ..... 0
38 passanger ..... 0
39 weather ..... 0
40 temperature ..... 0
41 time ..... 0
42 coupon ..... 0
43 expiration ..... 0
44 gender ..... 0
45 age ..... 0
46 maritalStatus ..... 0
47 has_children ..... 0
48 education ..... 0
49 occupation ..... 0
50 income ..... 0
51 car ..... 12576
52 Bar ..... 107
53 CoffeeHouse ..... 217
54 CarryAway ..... 151
55 RestaurantLessThan20 ..... 130
56 Restaurant20To50 ..... 189
57 toCoupon_GEQ5min ..... 0
58 toCoupon_GEQ15min ..... 0
59 toCoupon_GEQ25min ..... 0
60 direction_same ..... 0
61 direction_opp ..... 0
62 Y ..... 0
63 dtype: int64

```

# EDA

## DESCRIPTIVE ANALYSIS

toCoupon_GEQ5min	
count	12684.0
mean	1.0
std	0.0
min	1.0
25%	1.0
50%	1.0
75%	1.0
max	1.0

Fitur toCoupon\_GEQ5min memiliki nilai yang unik. Nilai kuartil, mean hingga maksimum hanya menampilkan angka 1. Hal ini disebabkan oleh unique value yang dimiliki oleh kolom ini hanya 1.

car	
count	108
unique	5
top	Scooter and motorcycle
freq	22

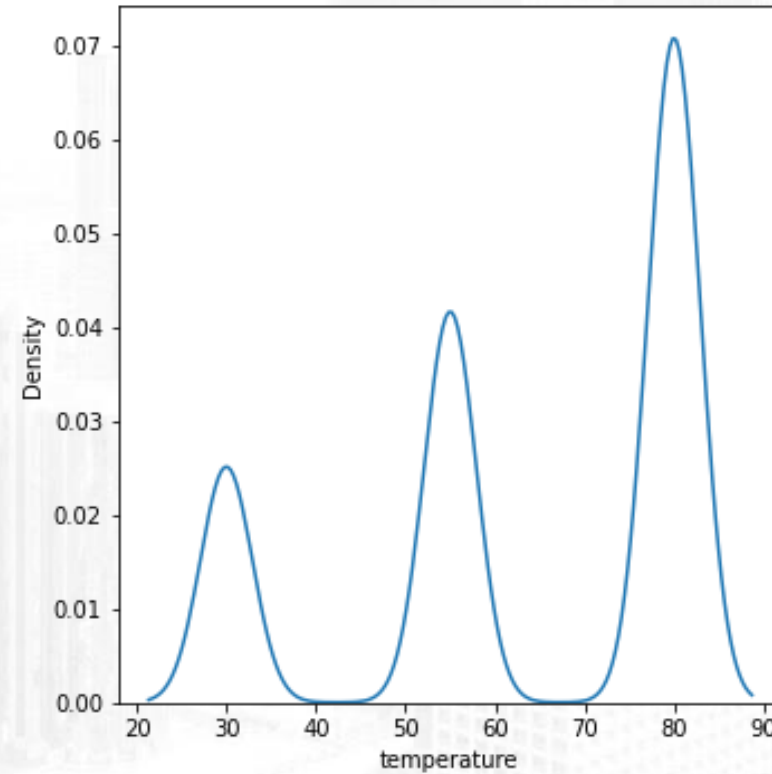
Berdasarkan informasi countnya, fitur car menjadi kandidat kuat untuk di-drop karena tingkat missing valuenya mencapai 99,14%

	direction_same	direction_opp
count	12684.000000	12684.000000
mean	0.214759	0.785241
std	0.410671	0.410671
min	0.000000	0.000000
25%	0.000000	1.000000
50%	0.000000	1.000000
75%	0.000000	1.000000
max	1.000000	1.000000

Nilai statistik direction\_opp dan direction\_same saling berseberangan, tetapi nilai stdnya sama. Hal ini dapat disebabkan karena value pada kedua kolom tersebut memiliki arti yang sama. Maka, salah satunya (direction\_opp) akan di-drop ketika pre-processing.

# EDA

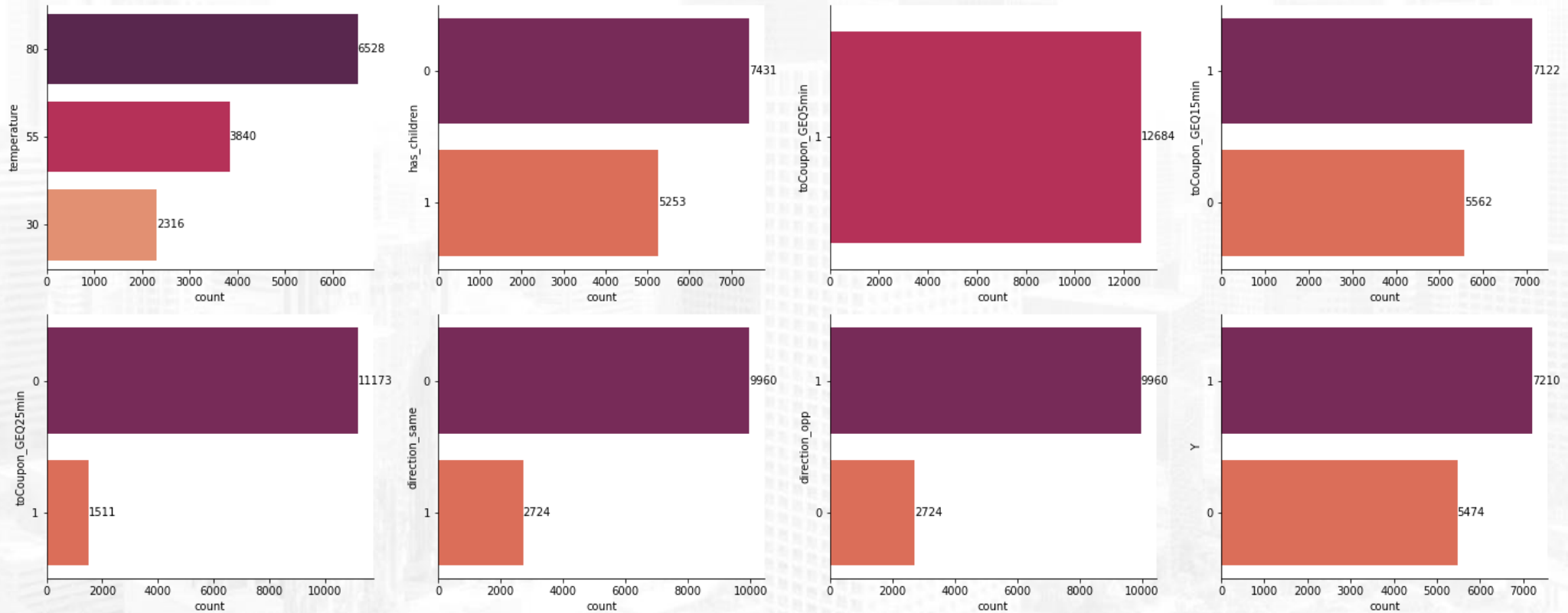
## UNIVARIATE ANALYSIS



fitur temperature memiliki distribusi trimodal, hal ini dikarenakan value pada fitur tersebut merupakan kategorikal sehingga distribusi data terkumpul pada 3 unique value tersebut.

# EDA

## UNIVARIATE ANALYSIS



- Fitur toCoupon\_GEQ5min hanya memiliki 1 unique value. Karena dirasa fitur ini tidak dapat memberikan insight, maka akan didrop pada tahap pre-processing.
- Fitur direction\_opp perlu didrop ketika data pre-processing karena fitur tersebut memiliki value yang artinya sama dengan fitur direction\_same, maka hanya memerlukan salah satunya saja.



## UNIVARIATE ANALYSIS



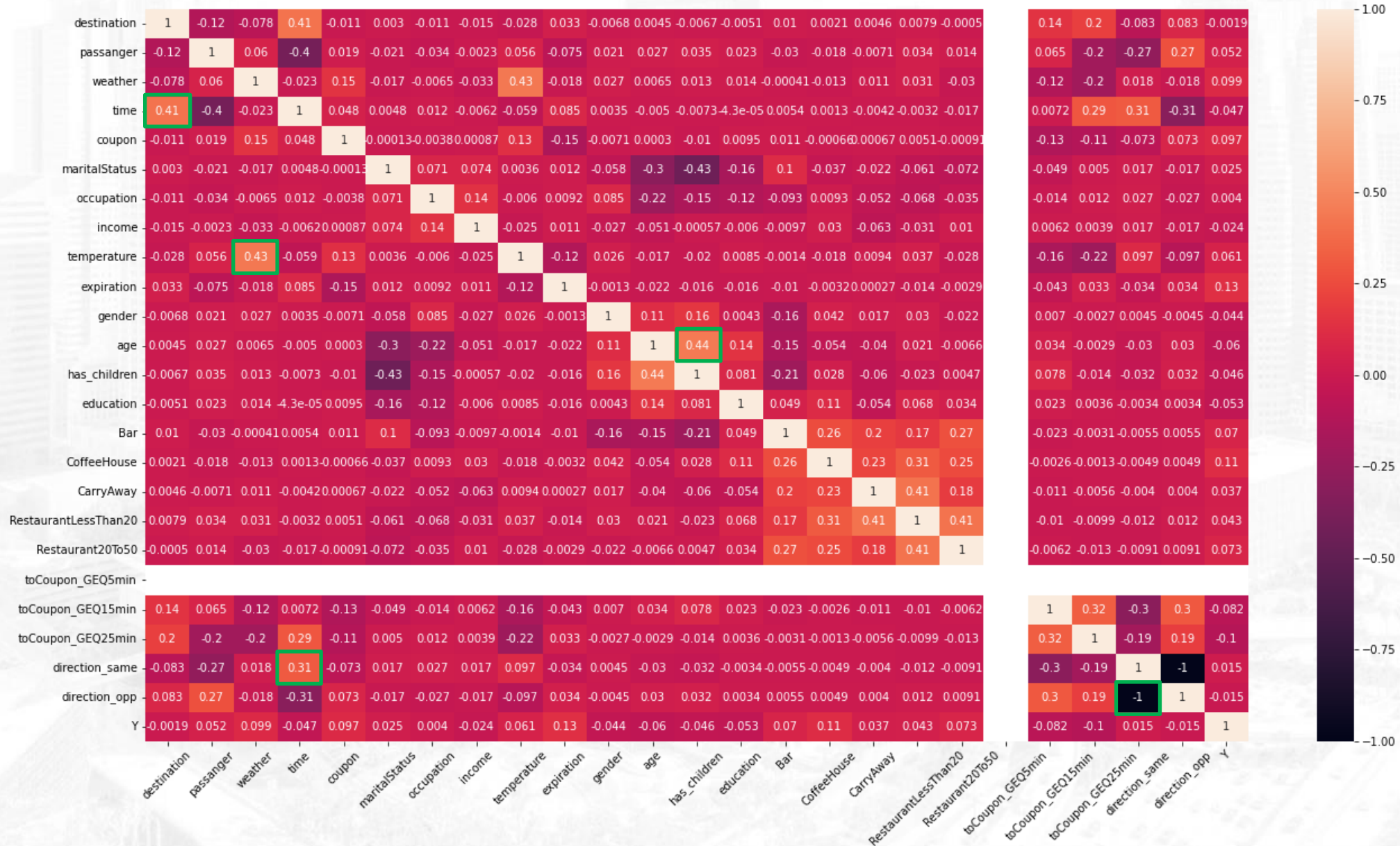


# EDA

## UNIVARIATE ANALYSIS

- Fitur occupation terlalu banyak kategorinya, perlu diringkas kategorinya pada saat data pre-processing. Pada fitur occupation didominasi oleh kategori unemployed.
- Fitur age dapat diringkas kategorinya pada saat data pre-processing.
- Fitur car terlalu banyak missing valuesnya, maka perlu didrop ketika data pre-processing.
- Fitur destination didominasi oleh 'No Urgent Place', fitur passenger didominasi oleh 'Alone', fitur weather didominasi oleh 'Sunny', fitur bar didominasi oleh 'never', fitur RestaurantLessThan20 didominasi oleh '1~3', fitur Restaurant20to50 didominasi oleh 'less1', fitur coupon didominasi oleh CoffeeHouse.

# MULTIVARIATE ANALYSIS



# EDA

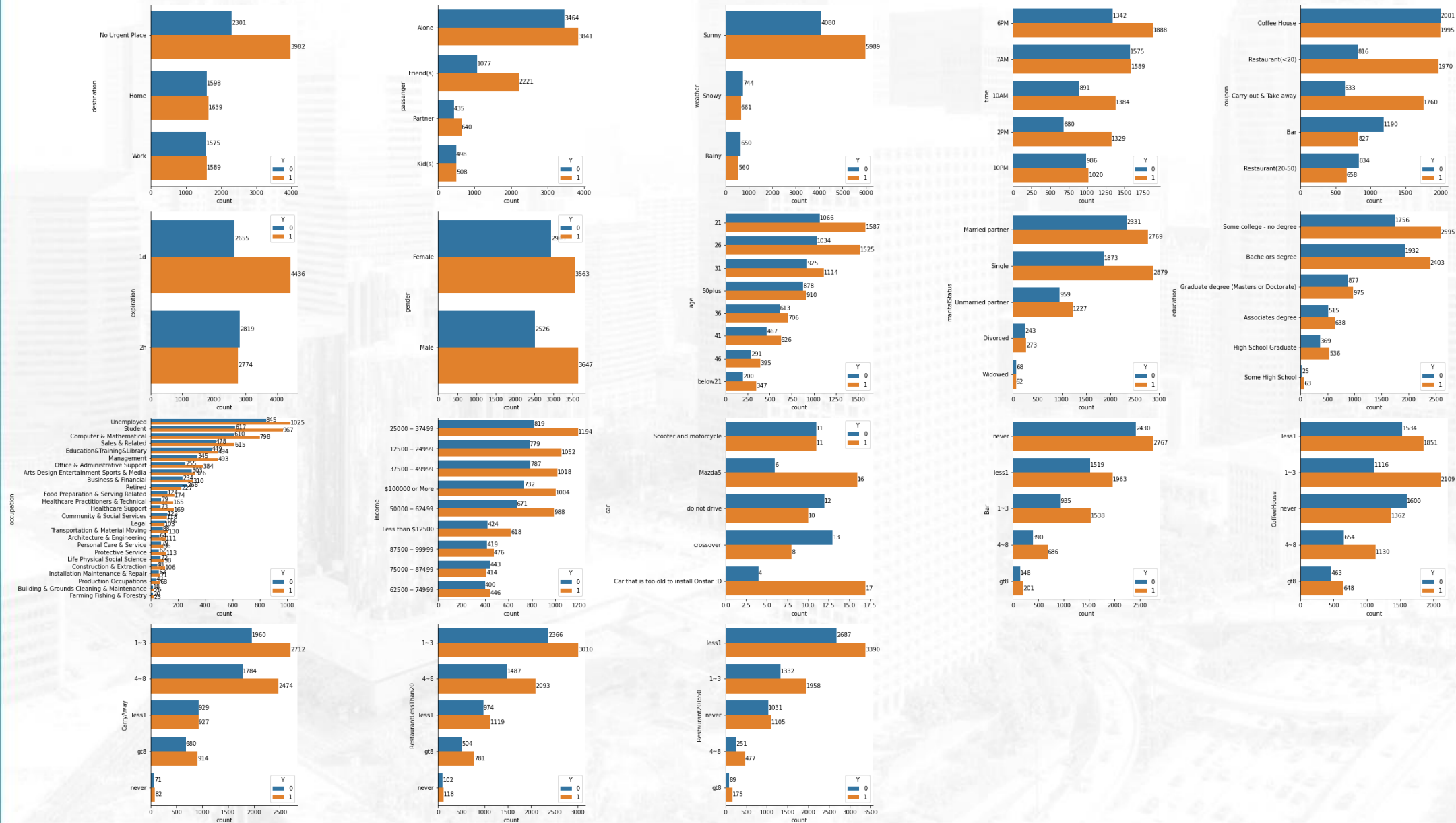
## MULTIVARIATE ANALYSIS

- Perhitungan korelasi diatas dilakukan terhadap seluruh kolom yang sebelumnya dilakukan label-encoding.
- Berdasarkan hasil perhtungan korelasi maka kolom/fitur yang relevan ialah:
 

1. Passanger	8. Bar
2. Time	9. CoffeeHouse
3. Weather	10. Carry Away
4. Coupon	11. RestaurantLessThan20
5. Temperature	12. Restaurant20To50
6. Expiration	13. direction_same
7. Age	14. toCoupon_GEQ15min
8. Education	15. toCoupon_GEQ25min
- Adapun fitur yang dirasa memiliki korelasi terhadap target tetapi berdasarkan perhitungan korelasi nilainya sangat kecil, seperti destination, time, direction\_same. Pendugaan sejauh ini karena fitur tersebut tidak cocok dilakukan label-encoding karena tipe data pada fitur terebut bukanlah ordinal/2 unique values.
- Selain itu, fitur lainnya yang dirasa memiliki korelasi terhadap target tetapi berdasarkan perhitungan korelasi nilainya sangat kecil, seperti CarryAway dan RestaurantLessThan20. Pendugaan sejauh ini karena fitur tersebut masih memiliki missing values ketika dilakukan perhitungan korelasi, sehingga nilainya menjadi lebih kecil dari yang seharusnya.
- Fitur direction\_same memiliki nilai korelasi sempurna dengan direction\_opp, hal ini memperkuat tujuan aal untuk drop salah satunya agar tidak menimbulkan multicollinearity.



## MULTIVARIATE ANALYSIS



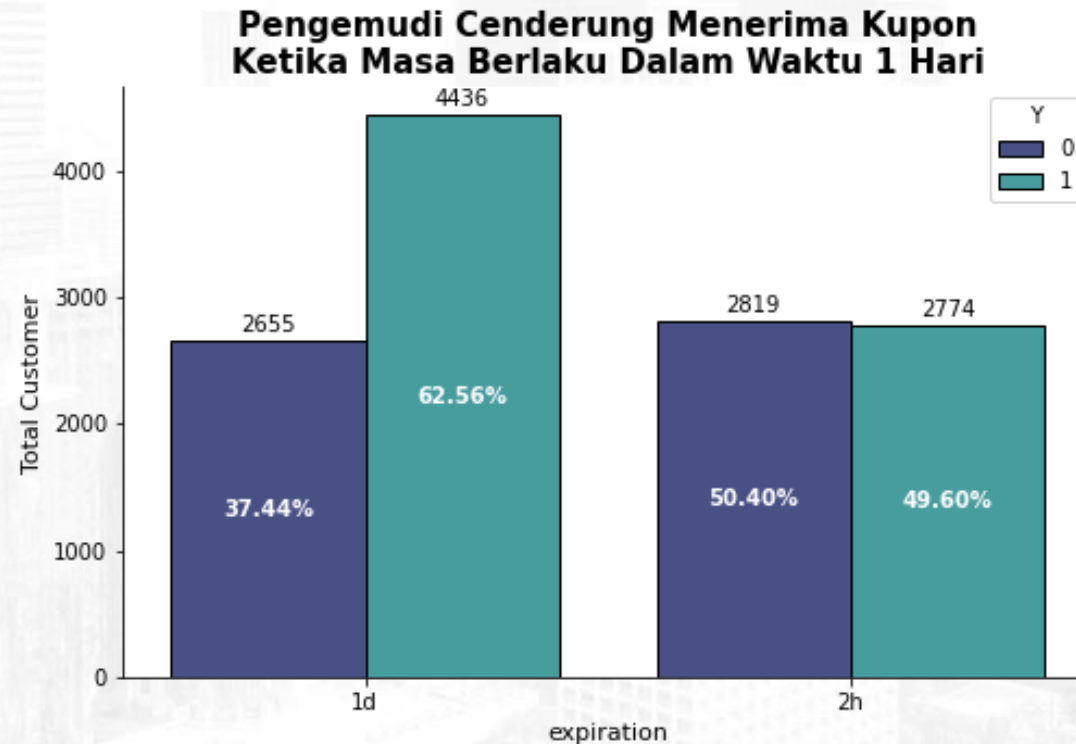


# EDA

## MULTIVARIATE ANALYSIS

- Pada fitur "**destination**" nampak bahwa pengendara yang paling banyak menerima kupon adalah pendedara dengan yang memiliki destinasi "**No Urgent Place**" sementara pada destinasi "**home**" dan "**work**" memiliki nilai setara lebih rendah .
- Pada fitur "**weather**" diketahui bahwa kupon paling banyak diterima oleh pengendara dalam kondisi cuaca "**sunny**" dibanding "**snowy**" dan "**rainy**"
- Diketahui "**time**" penerimaan kupon yang paling banyak adalah pada waktu "**10AM**", "**2PM**" dan "**6PM**" dibanding waktu lainnya.
- Diketahui pada fitur "**jenis kupon**" kupon yang paling banyak diterima oleh pengendara adalah jenis kupon "**RestaurantLessThan20**" dan "**carry out and take away**".
- pada fitur "**expiration**" kupon lebih banyak dipilih pada kategori expiration "**1d**" dibanding "**2h**".
- Pada fitur "**marital status**" penerima kupon didominasi oleh pengendara dengan status "**single**" dan "**marriage couple**".
- Pada tingkat "**education**" terlihat penerima paling banyak adalah dari status pendidikan "**college non-degre**" dan "**bachelor**".

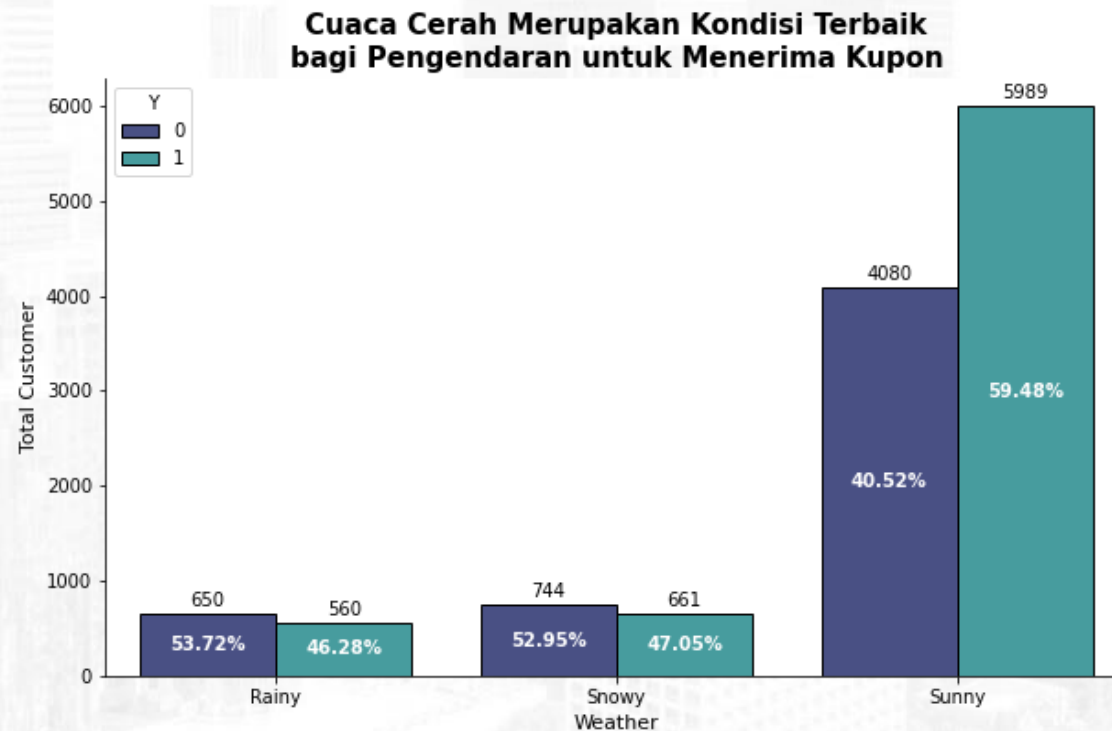
# BUSINESS INSIGHT (1)



Pelanggan/pengemudi yang menerima kupon dengan masa berlaku 1 hari memiliki rasio 62.56%, sedangkan pelanggan/pengemudi yang menerima kupon dengan masa berlaku 2 jam hanya 49.60%. Hal ini dapat disebabkan karena tidak semua pelanggan/pengemudi berpikir untuk menggunakan kupon sesegera mungkin karena ada pertimbangan aktivitas, lokasi berkendara, lokasi penggunaan kupon, dll.

Oleh karena itu, sangat disarankan kupon yang akan diberikan kepada pelanggan/pengemudi merupakan kupon yang masih memiliki masa berlaku 1 hari lamanya dan menghindari memberikan kupon yang masa berlakunya sebentar.

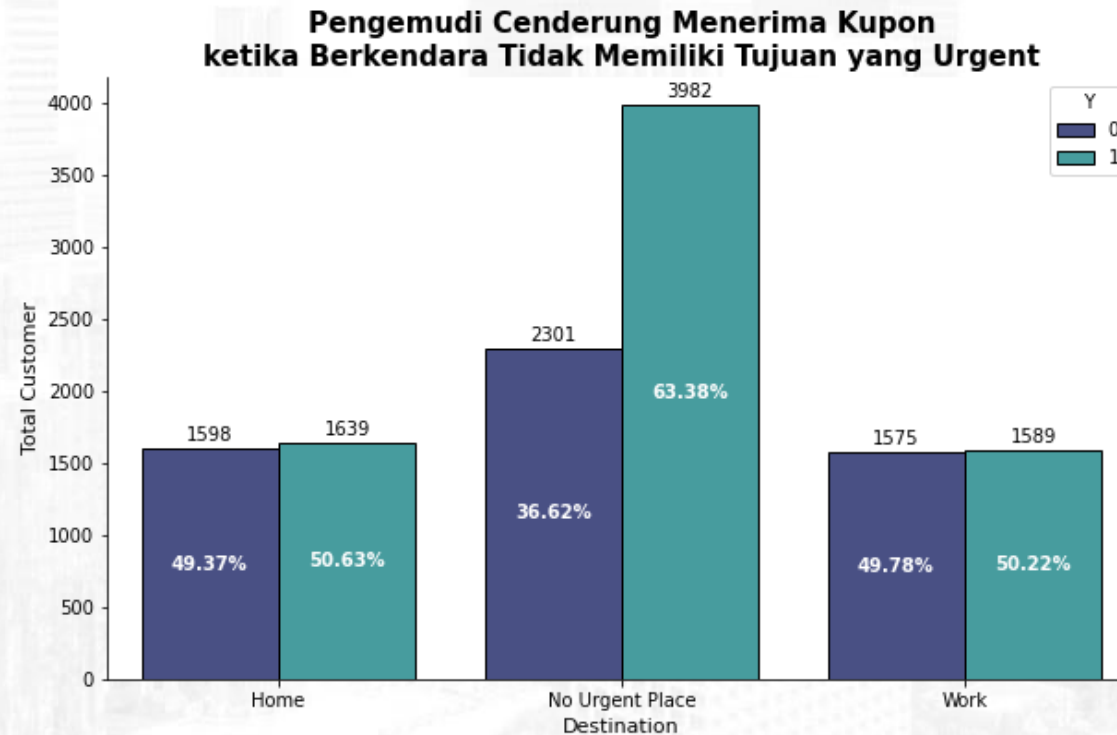
# BUSINESS INSIGHT (2)



Jumlah pelanggan/pengemudi yang menerima kupon jauh lebih banyak ketika cuaca cerah (10.097), selain itu rasio yang menerima kupon pun paling tinggi (59.48%) dibandingkan kondisi cuaca lainnya. Hal ini dapat disebabkan karena pelanggan/pengemudi tidak banyak yang keluar rumah/beraktivitas ketika cuaca hujan atau bersalju.

Oleh karena itu, sangat disarankan campaign pemberian kupon diadakan pada musim panas/semi. Hindari mengadakan campaign pemberian kupon pada musim hujan/salju.

# BUSINESS INSIGHT (3)

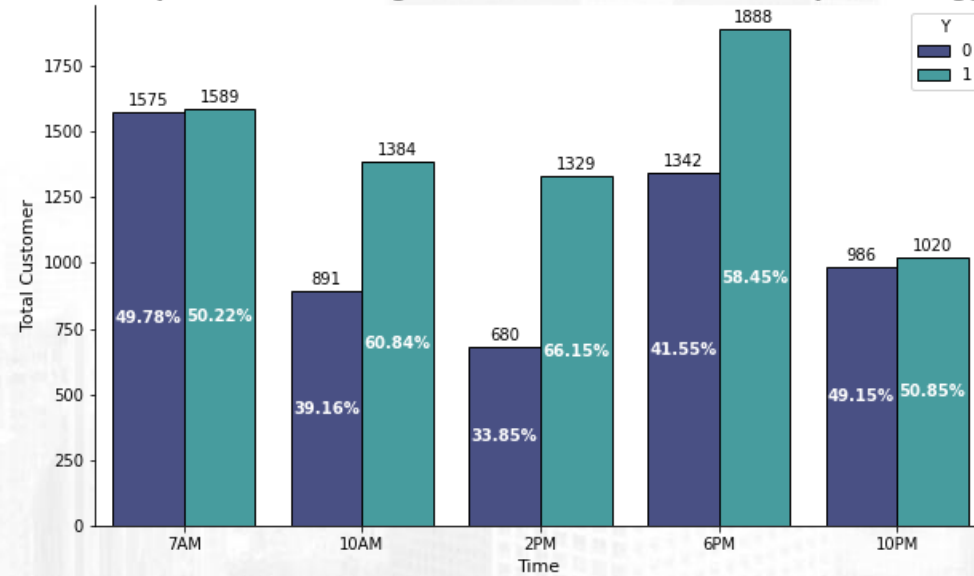


Jumlah pelanggan/pengemudi yang menerima kupon jauh lebih banyak ketika mereka sedang tidak memiliki tujuan tertentu ketika berkendara (3.982). Rasio penerimaan kuponnya pun paling tinggi (63.38%) dibanding yang lainnya. Hal ini disebabkan tidak adanya urgensi ketika berkendara artinya target pelanggan/pengemudi memiliki waktu kosong lebih untuk mengunjungi lokasi kupon untuk digunakan, maka dari itu tingkat penerimaannya lebih tinggi dibanding yang lain.

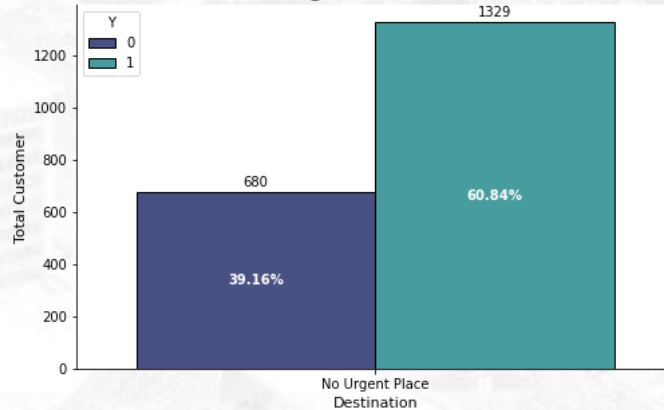


# BUSINESS INSIGHT (3)

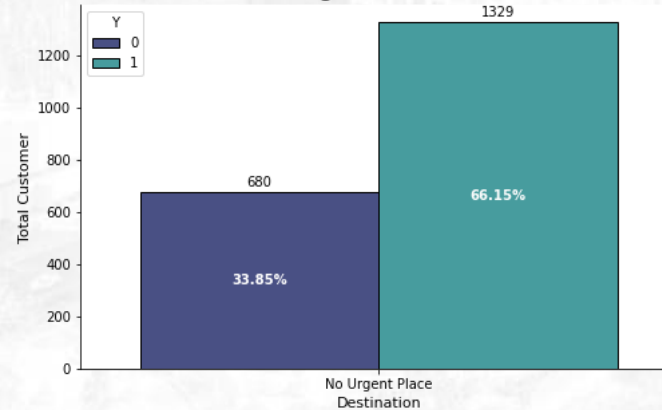
**Pukul 10 Pagi, 2 Siang dan 6 Sore**  
**Merupakan Waktu dengan Persentase Penerimaan Kupon Tertinggi**



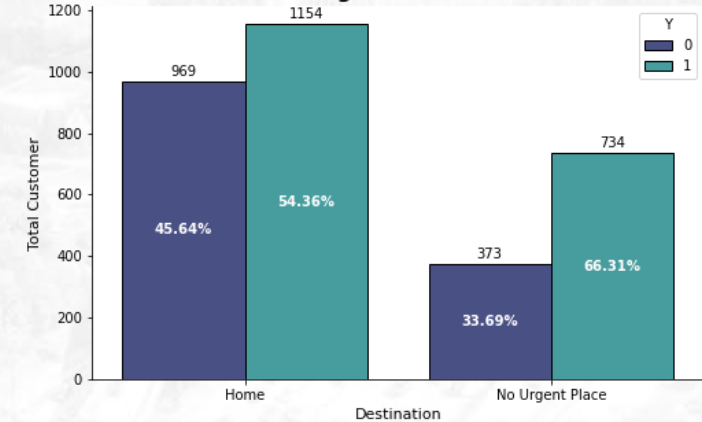
**Destinasi Pengemudi Pada Pukul 10AM**



**Destinasi Pengemudi Pada Pukul 2PM**



**Destinasi Pengemudi Pada Pukul 6PM**



# BUSINESS INSIGHT (3)

Jumlah pelanggan/pengemudi yang menerima kupon tinggi pada pukul 10AM (1.384), 2PM (1.329) dan 6PM (1.888). Setelah dianalisis lebih lanjut, ternyata hal ini disebabkan oleh destinasi pelanggan/pengemudi pada jam tersebut hanya 2, yakni tidak ada destinasi urgent dan rumah. Pada pukul 10AM dan 2PM pelanggan/pengemudi seluruhnya tidak memiliki destinasi urgent. Rasio penerimaan kuponnya-pun tinggi yakni 60.84% dan 66.15%. Lalu, pada pukul 6PM destinasi pelanggan/pengemudi terdiri atas tidak adanya destinasi urgent dan rumah. Meskipun begitu rasio penerimaan kuponnya tetap tinggi 54.36% dan 66.31%. Hal ini dapat disebabkan karena situasi pelanggan/pengemudi ketika berkendara pulang ke rumah sembari mencari makanan untuk di rumah.

Berdasarkan variabel destinasi dan waktu, maka sangat disarankan pemberian kupon dilakukan secara masif pada pukul 10AM, 2PM dan 6PM. Dikarenakan destinasi pelanggan/pengemudi pada pukul 6PM ada yang ke rumah, maka pemberian kupon CarryAway merupakan langkah yang tepat sasaran.