

Barbershop Customer Forecast



Peramalan Jumlah Konsumen dan Kunjungan Barber Shop

Data google business-transaction
from Inalyst

@michael sitanggang

BUSINESS UNDERSTANDING

Perbedaan bisnis barbershop dengan bisnis lain adalah adanya peluang bahwa pelanggan akan datang lagi atau tidak. Umumnya pelanggan yang pernah datang lebih dari 1 kali akan menjadi pelanggan tetap dan akan datang kembali. Sinkronisasi Data Transaksi dan Data Google Bisnis dapat dijadikan acuan pengadaaan promo dan jasa SEO. Hari weekday atau weekend juga bisa jadi berpengaruh terhadap banyaknya kedatangan customer.



Business Understanding ◦ ◦ ◦ ◦

Variabel data transaksi

Cukur_ke
Urutan
id_cust
Capster
Status
Haircut
Bayar
Tanggal
Hari
Total Payment

Variabel data google bisnis

Start Date, End Date, Overall rating, Total searches, Direct searches, Discovery searches, Total views, Search views, Maps views, Total actions, Website actions, Directions actions, Phone call actions.

GOALS | TARGET CUST

barbershop membutuhkan prediksi secara harian, berapa visitor yang akan datang pada beberapa hari yang akan datang.

AUG-2021



Exploratory Data Analysis



MAR-2022



Time Series Analysis



Data Understanding

Data Transaksi

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 2049 entries, 0 to 2048
```

```
Data columns (total 9 columns):
```

#	Column	Non-Null Count	Dtype
0	id_cust	2046 non-null	object
1	Status	2046 non-null	object
2	Cukur_ke	1993 non-null	float64
3	Capster	2046 non-null	object
4	Hari	2049 non-null	object
5	Tanggal	2049 non-null	object
6	Bayar	2049 non-null	object
7	Haircut	2049 non-null	int64
8	Total Payment	2049 non-null	int64

```
dtypes: float64(1), int64(2), object(6)
```

```
variasi label variabel Status sebanyak 2  
variasi label variabel Capster sebanyak 6  
variasi label variabel Hari sebanyak 7  
variasi label variabel Tanggal sebanyak 224  
variasi label variabel Bayar sebanyak 2
```

NB : nilai duplikat sebanyak 2

*persentase
nilai hilang*



*variasi label
var kategorik*



Cukur_ke	2.733041
Urutan	0.146413
id_cust	0.146413
Capster	0.146413
Status	0.146413
Haircut	0.000000
Bayar	0.000000
Tanggal	0.000000
Hari	0.000000
Total Payment	0.000000
dtype:	float64

Data Understanding

Data Google Bisnis

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18 entries, 1 to 18
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Start Date            18 non-null    datetime64[ns]
1   End Date              18 non-null    datetime64[ns]
2   Overall rating        18 non-null    int8
3   Total searches        18 non-null    int16
4   Direct searches       18 non-null    int16
5   Discovery searches    18 non-null    int16
6   Total views           18 non-null    int16
7   Search views          18 non-null    int16
8   Maps views            18 non-null    int16
9   Total actions         18 non-null    int16
10  Website actions       18 non-null    int8
11  Directions actions    18 non-null    int16
12  Phone call actions    18 non-null    int8
dtypes: datetime64[ns](2), int16(8), int8(3)
```

*persentase
nilai hilang*



Start Date	0.0
End Date	0.0
Overall rating	0.0
Total searches	0.0
Direct searches	0.0
Discovery searches	0.0
Total views	0.0
Search views	0.0
Maps views	0.0
Total actions	0.0
Website actions	0.0
Directions actions	0.0
Phone call actions	0.0
dtype:	float64

*tidak terdapat
var kategorik*



NB : tidak ditemukan nilai duplikat

Data Preparation

1. Imputasi nilai hilang variabel numerik dengan nilai median
2. Imputasi nilai hilang variabel kategorik dengan nilai modus

```
#membagi atribut numerik dan kategorik
num = df_tf.select_dtypes(exclude='object')
cat = df_tf.select_dtypes(include='object')

modetreat = cat.columns[cat.isna().any()].to_list()
mediantreat = num.columns[num.isna().any()].to_list()

for i in modetreat:
    cat.loc[cat.loc[:,i].isna(),i] = cat.loc[:,i].mode()[0] #var kategori diisi dengan modus
for i in mediantreat:
    num.loc[num.loc[:,i].isna(),i] = num.loc[:,i].median() #var numerik diisi dengan median

df_tf = pd.concat([num,cat],axis=1)
df_tf.isna().sum()
```

3. Hapus nilai duplikat `df_tf = df_tf.drop_duplicates()`

EDA

```
#Pelanggan dengan kunjungan > 1 kali  
cust_lama = df_tf["id_cust"].duplicated()  
df_custlama = df_tf[cust_lama]
```

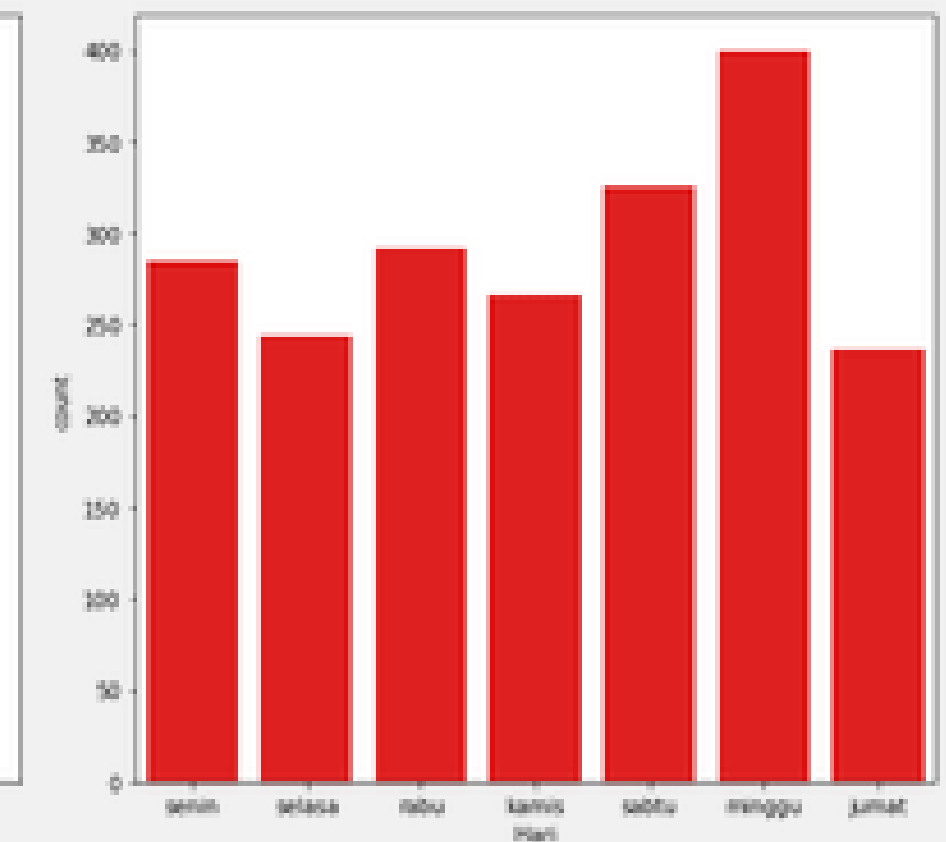
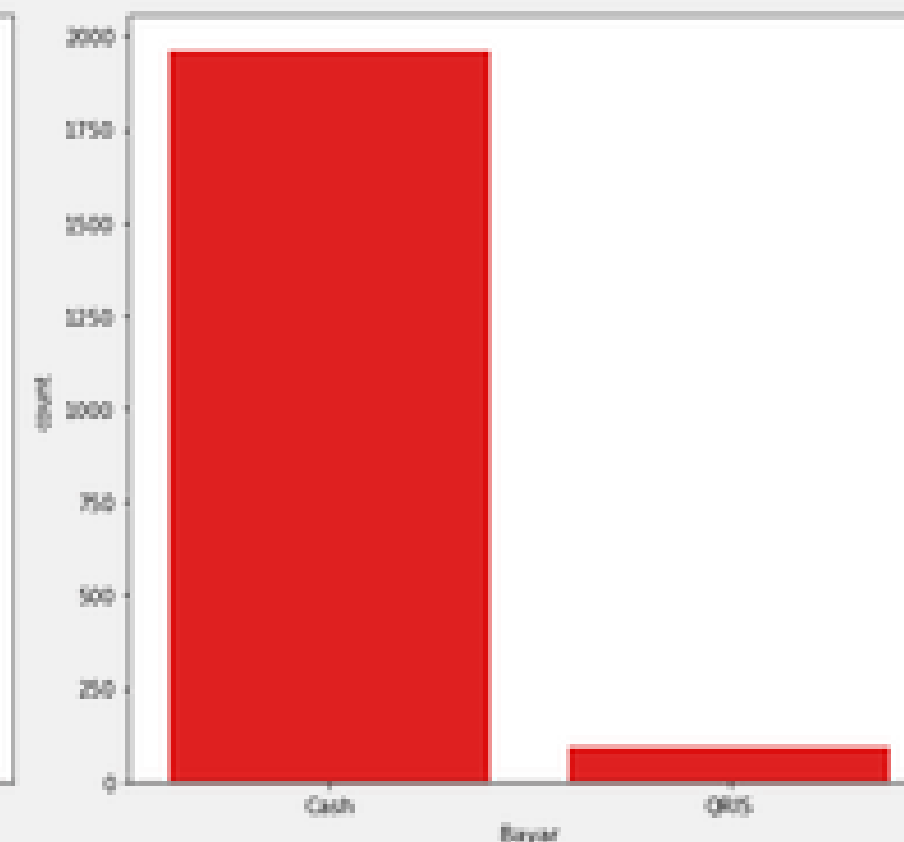
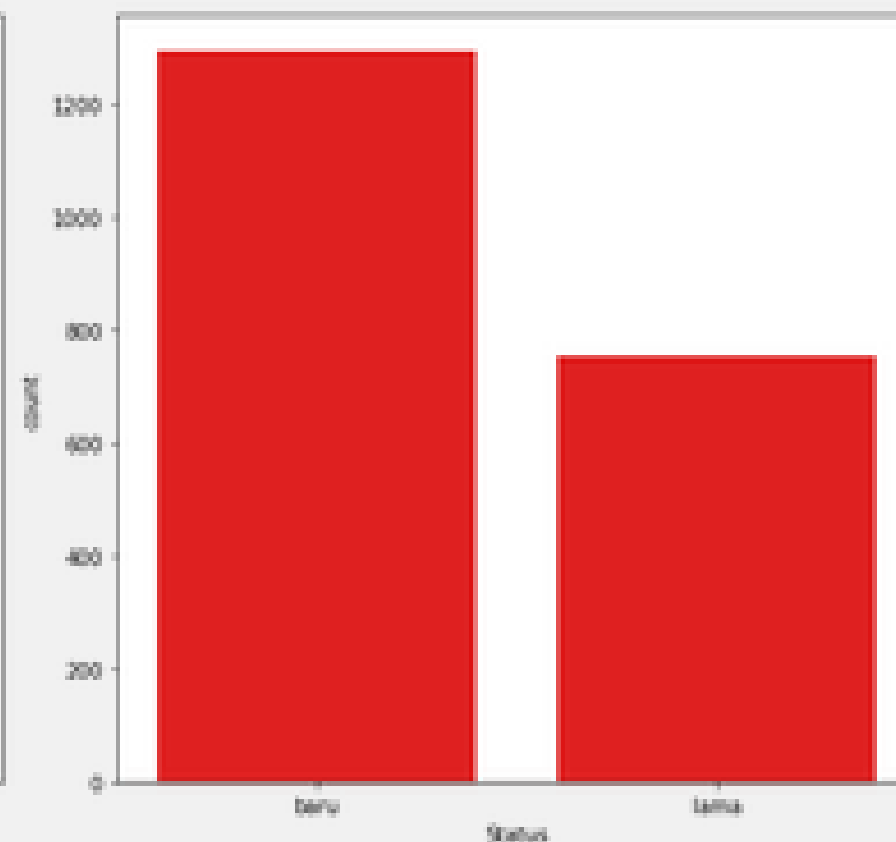
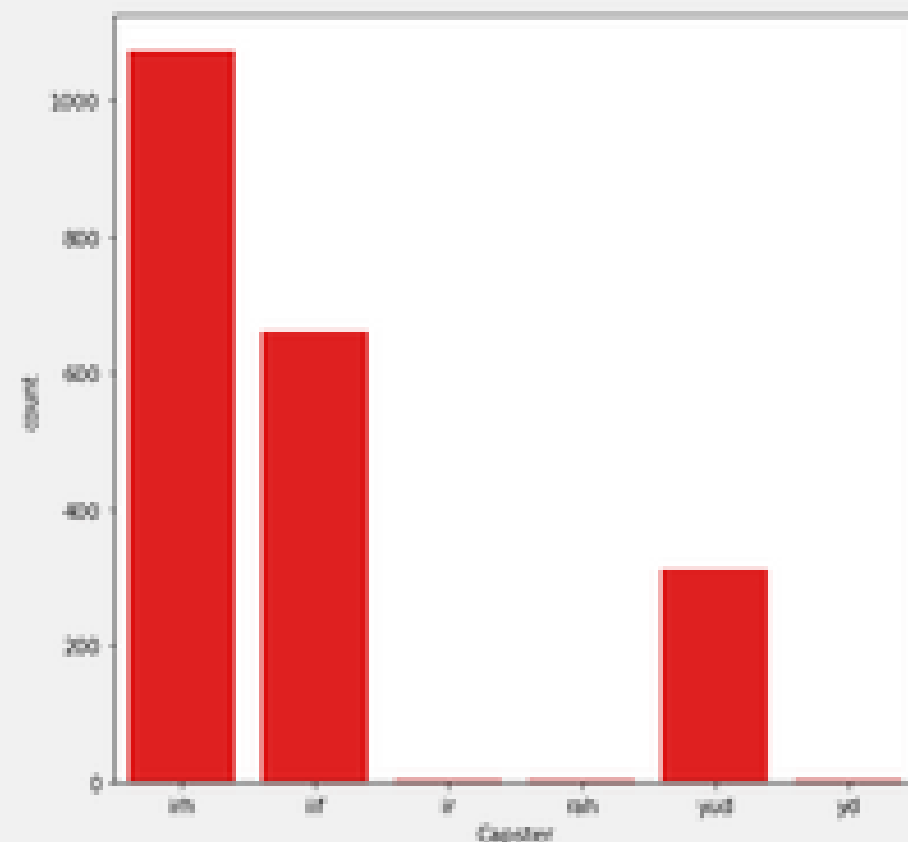
```
#Jumlah Pelanggan Tetap  
df_custlama['id_cust'].nunique()
```

352

>>> jumlah konsumen dengan kedatangan lebih dari 1 kali sebanyak 352 orang



insight pada halaman selanjutnya



insight

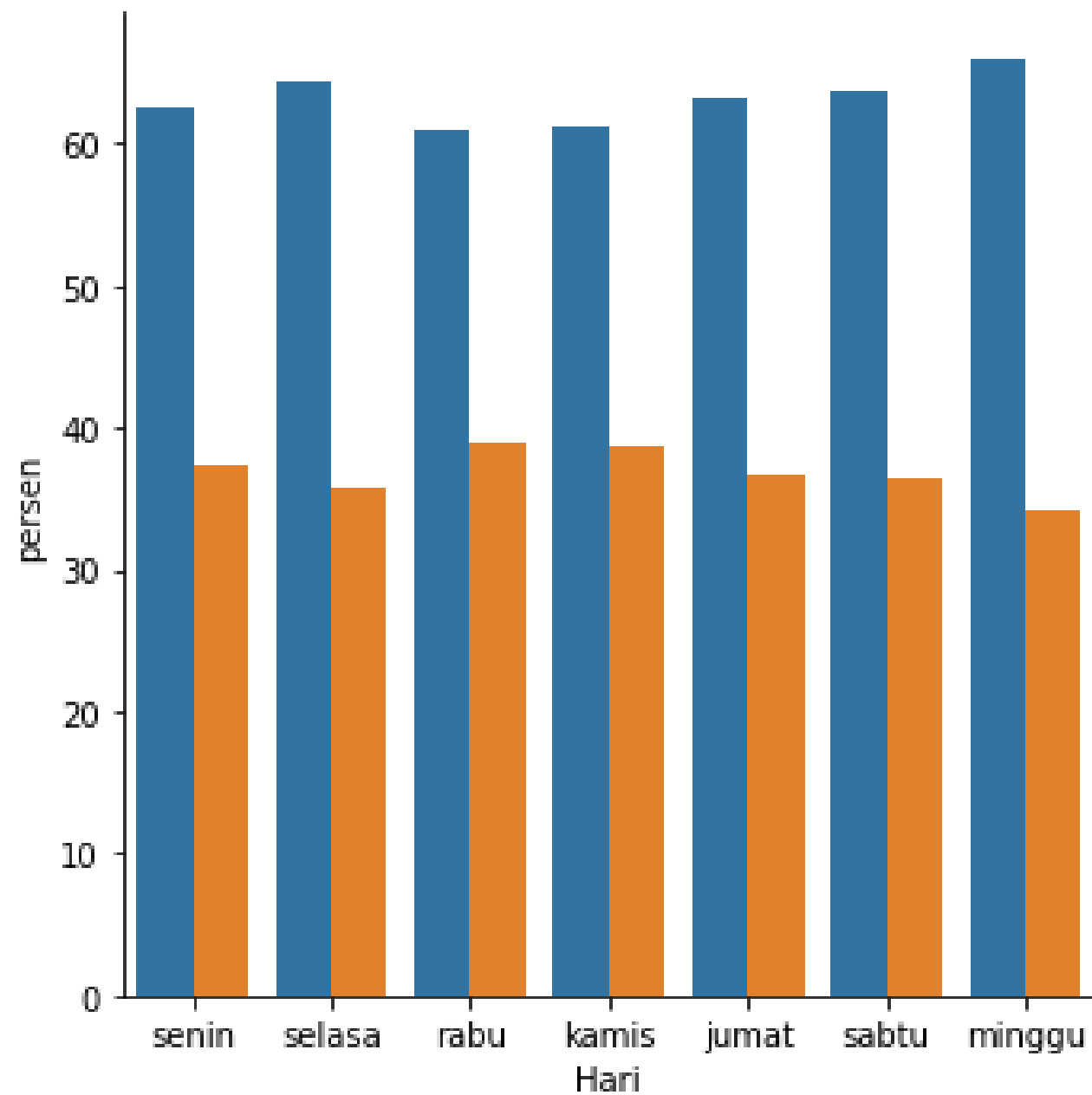
- 1.terdapat label terlihat imbalance dalam variabel capster maka akan dilakukan pemeriksaan.
- 2.terlihat pengujung baru lebih banyak dibandingkan pengujung lama.
- 3.terlihat pembayaran cash lebih banyak digunakan dibandingkan Qris.
- 4.asumsi "hari weekday atau weekend juga bisa jadi berpengaruh terhadap banyaknya kedatangan customer" diterima.

```
#cek variabel capster yang terlihat kosong dalam data
names = ['ir', 'rah', 'yd']
df_tf[df_tf.Capster.isin(names)]
```

	Cukur_ke	Haircut	Total Payment	id_cust	Status	Capster	Hari	Tanggal	Bayar
1089	2.0	35000	35000	25506597Rooma	lama	ir	rabu	2022-01-19	Cash
1181	2.0	35000	35000	25825025Rihru	lama	rah	jumat	2022-01-28	Cash
1602	3.0	35000	35000	2561701Yusuf	lama	yd	jumat	2022-03-04	Cash



3 label yang terlihat imbalance hanya memiliki masing-masing 1 data.

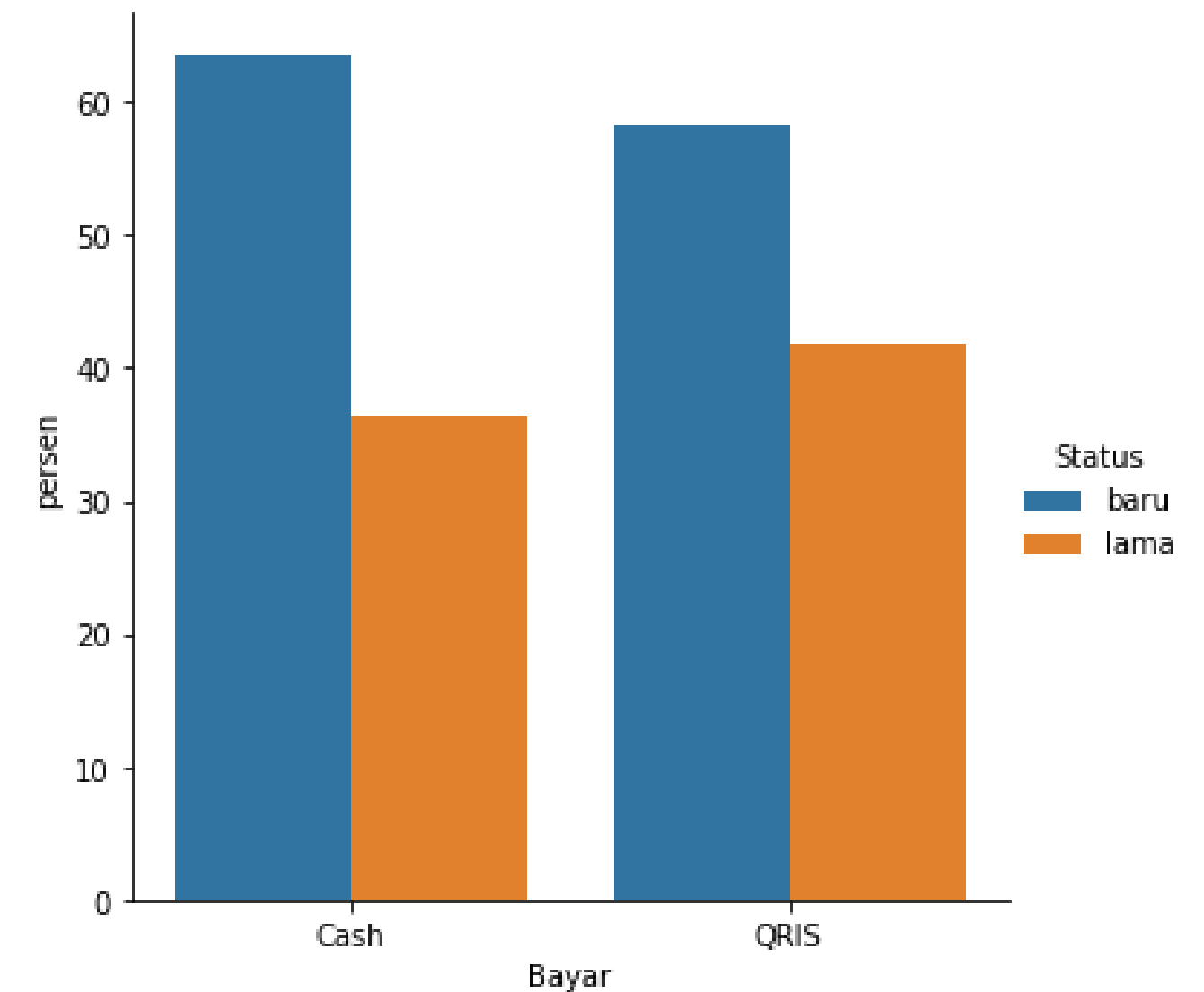


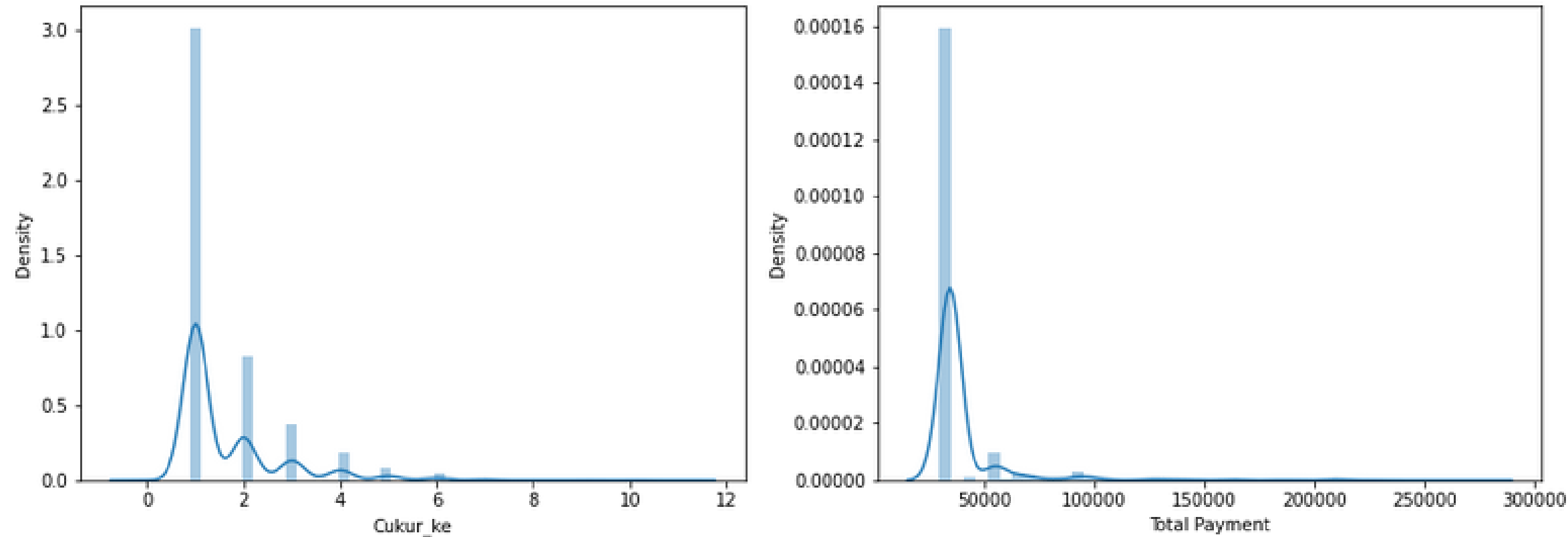
insight

- * Pengunjung baru relatif meningkat mendekati hari weekend
- * Pengunjung lama relatif menurun mendekati hari weekend

insight

- * Pengunjung baru relatif melakukan pembayaran Cash
- * Pengunjung lama relatif melakukan pembayaran Qris





- Pengunjung lama tidak bisa langsung disebut sebagai pelanggan tetap,
- Pengunjung baru dan lama relatif hanya melakukan `hair cut + pomade` [Total Payment] dan tidak terlihat permintaan terhadap fasilitas lainnya.

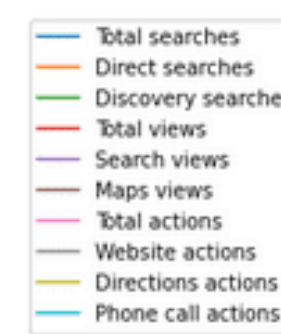
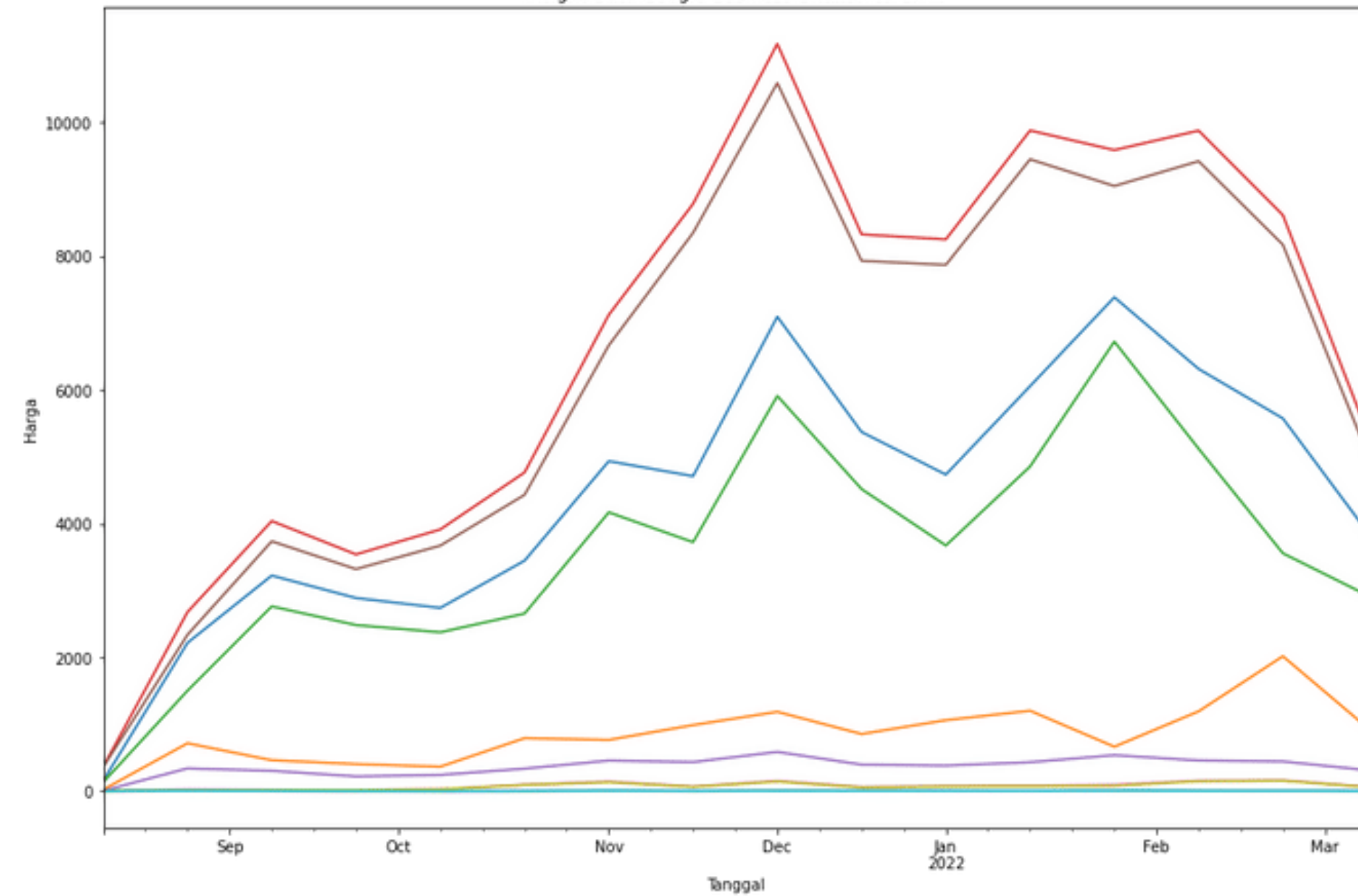
```
df_x = df_tf.groupby(['id_cust', 'Tanggal'])['Cukur_ke'].mean().reset_index()
diff = []

for i in df_x['id_cust'].unique():
    x = df_x[df_x.id_cust==i].reset_index()
    n = len(x)
    for i in range(1,n):
        diff.append(diff_dates(x.iloc[i-1,2],x.iloc[i,2]))

print('Rata-rata jarak hari pengunjung lama pangkas kembali {} hari'.format(round(sum(diff) / len(diff),0)))
```

Rata-rata jarak hari pengunjung lama pangkas kembali 50.0 hari

Insight Data Google Business 1 tahun terakhir



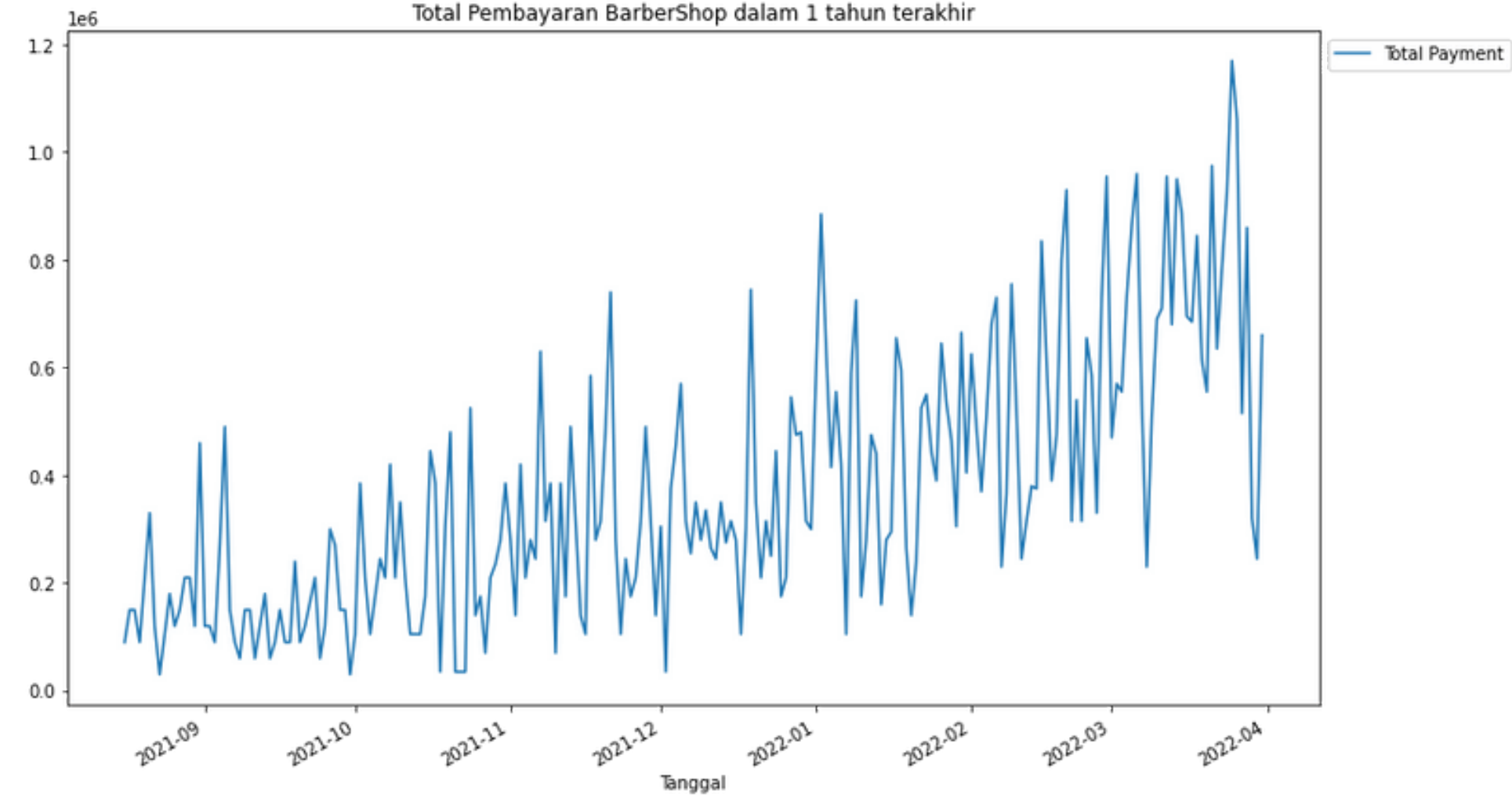
insight

- Variabel Total searches, Discovery searches, Total views, Maps views memiliki pola trend yang sama dalam 1 tahun terakhir
- Variabel Direct searches, Search views, Total actions, Website actions, Directions actions, Phone call actions memiliki pola trend yang sama dalam 1 tahun terakhir

insight

Income BarberShope relatif meningkat setiap bulannya dalam 1 tahun terakhir

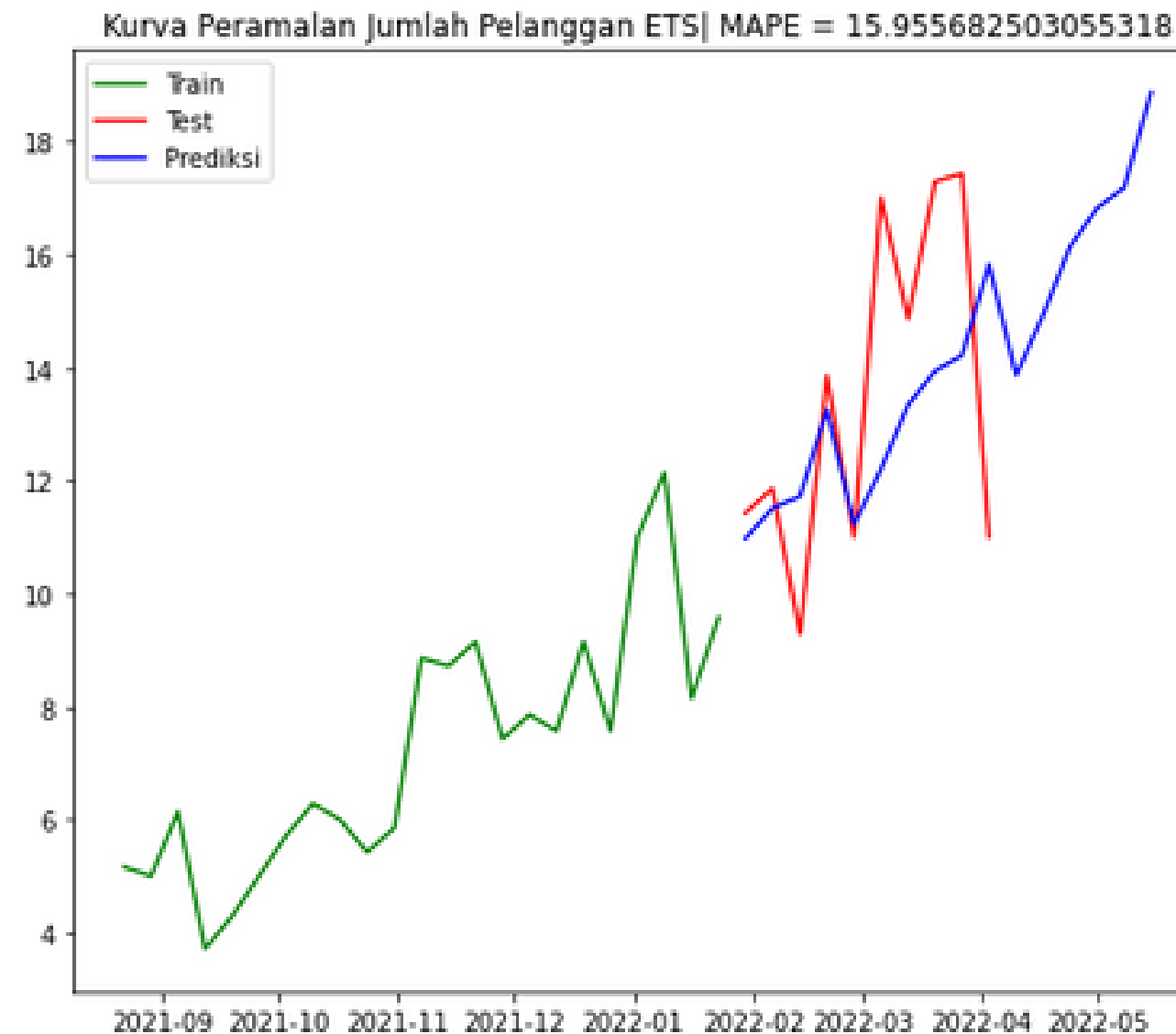
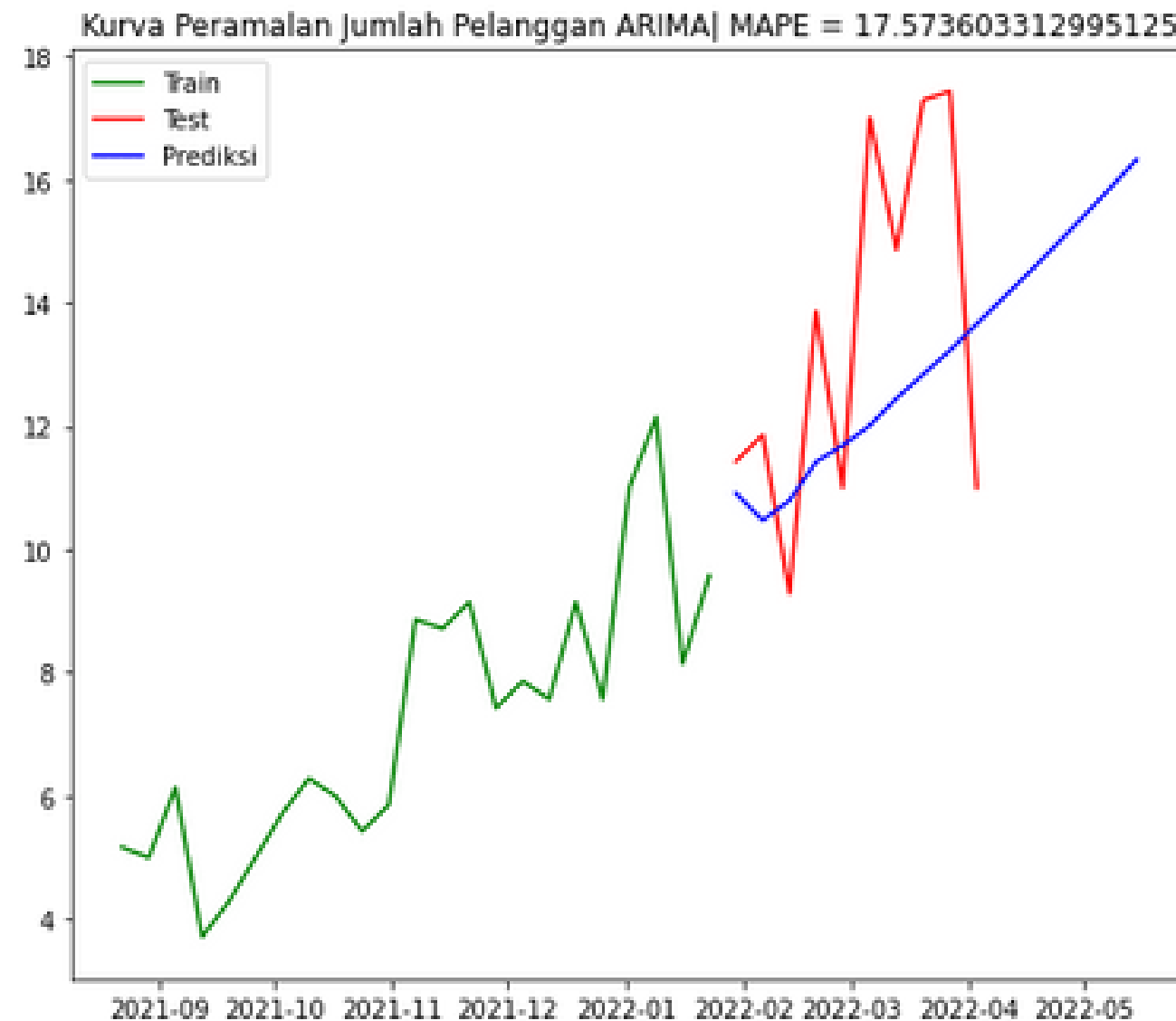
Total Pembayaran BarberShop dalam 1 tahun terakhir



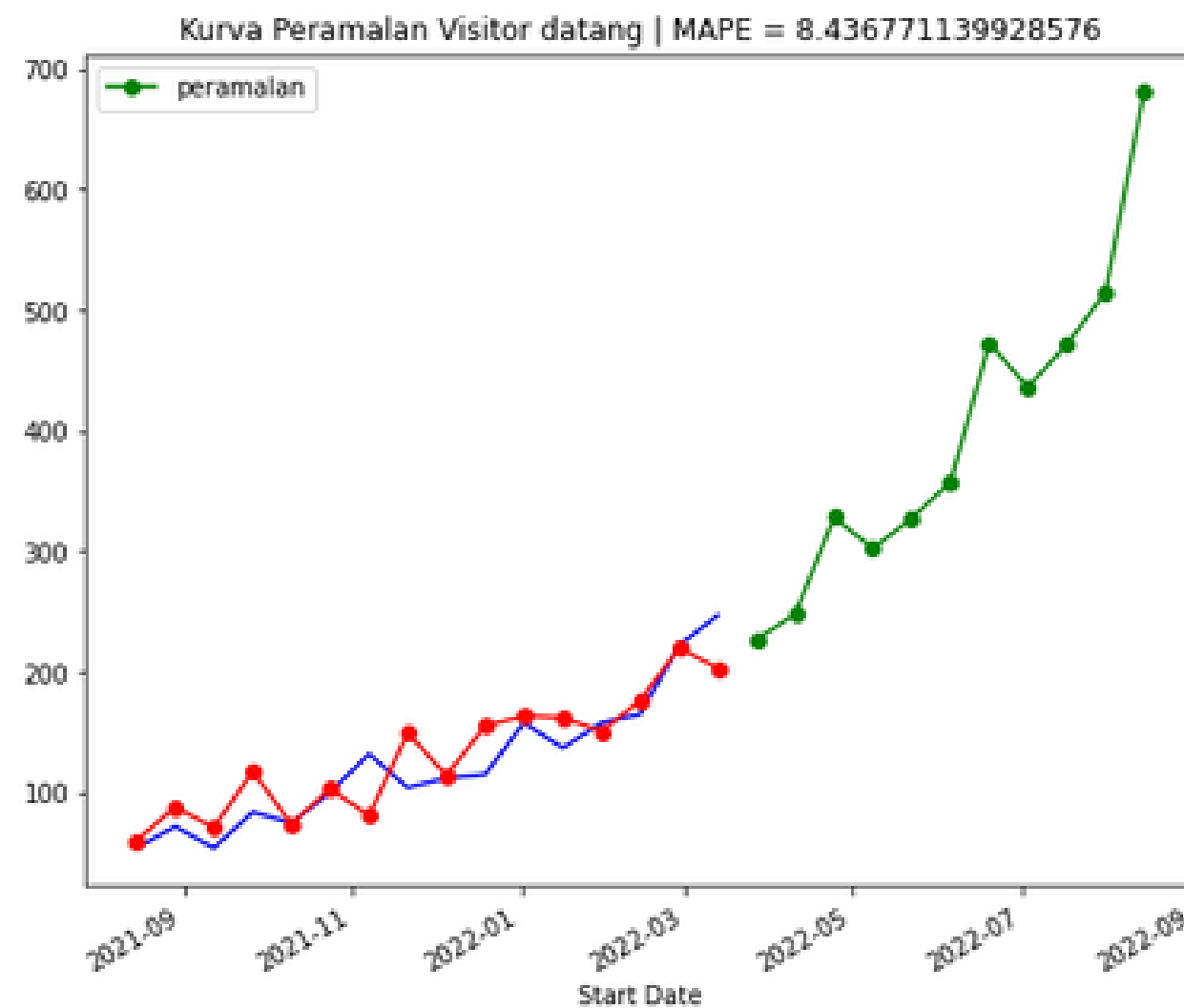
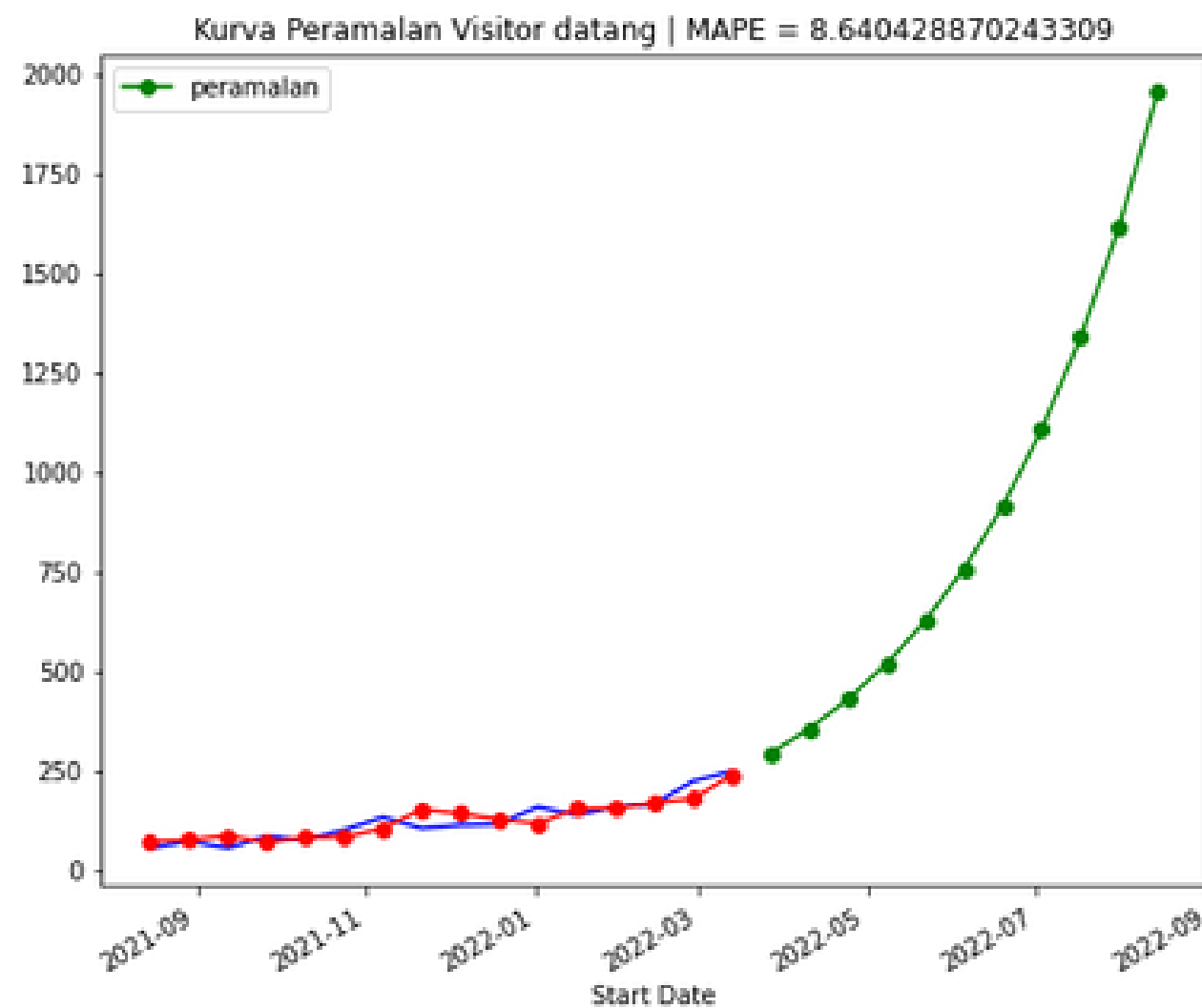
Modelling Time Series

sesuai dengan pola trend dan musiman kedua data maka;

1. Data Transaksi membandingkan model Arima dan ETS
2. Data Google Bisnis menggunakan model eksponensial smoothing



MAPE model Eksponensial Smoothing < model ARIMA untuk kedatangan pelanggan berdasarkan data transaksi



MAPE model Eksponensial Smoothing lebih kecil setelah dilakukan tuning parameter, untuk jumlah visitor BarberShop berdasarkan data google bisnis

fix model = MAPE terkecil =

Data Transaksi = Eksponensial Smoothing parameter periode musiman(6) dan trend multiplicative ; sesasonal additive

Data Google Bisnis = Eksponensial Smoothing parameter periode musiman(4) dan trend multiplicative ; sesasonal multiplicative

KESIMPULAN

1. Asumsi "hari weekday atau weekend juga bisa jadi berpengaruh terhadap banyaknya kedatangan customer" diterima dan didominasi pengunjung baru / customer baru,
2. PELUANG peningkatan customer baru relatif besar dan dipengaruhi optimalisasi jasa SEO pada google bisnis,
3. Memberikan paket layanan untuk meningkatkan ketertarikan konsumen terhadap fasilitas lainnya >>> [Lakukan penelitian bersifat kuisioner untuk pelanggan],
4. Memberi potongan harga pada pembayaran QRIS untuk digitalisasi data dan kepentingan analisis >>> [Opsi untuk membuat paket layanan kedepannya],
5. Meningkatkan optimalisasi jasa SEO untuk peluang ekspansi bisnis yang lebih besar.

@michael sitanggang