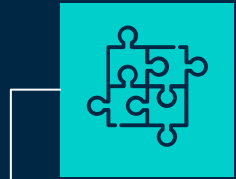


MARKETING ANALYTICS

The background is a dark navy blue. It is decorated with various geometric elements: small squares in teal, orange, and pink, some of which are solid and others are hollow outlines. Thin, light-colored vertical lines of varying lengths are scattered across the canvas, creating a modern, data-driven aesthetic.

Data Science Process



01

Business
Understanding



02

Data
Understanding



03

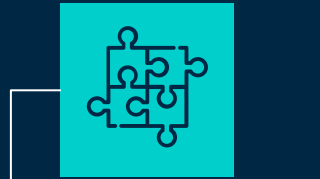
Data
Preparation

Data Science Process



04

Modelling and Task
Solving



05

Evaluation

Business Understanding

01

Business Understanding

Data Customer



- Customer profiles
- Product preferences
- Campaign successes/failures
- Channel performance



Goal

Mengevaluasi hasil penjualan dan strategi marketing saat ini



Menyusun usulan strategi marketing baru untuk meningkatkan penjualan kedepannya

Data Understanding

02

Deskripsi Dataset

Tampilan info setiap atribut

Data columns (total 22 columns):

#	Column	Non-Null	Count	Dtype
0	Tanggal Lahir	2240	non-null	object
1	Pergawean	2240	non-null	object
2	Status Kawin	2240	non-null	object
3	Penghasilan per Tahun	2216	non-null	object
4	Tanggungan Ortu	2240	non-null	int64
5	Tanggungan Anak	2240	non-null	int64
6	Mulai Berlangganan	2240	non-null	int64
7	Transaksi Terakhir	2240	non-null	int64
8	Sembako	2240	non-null	int64
9	Elektronik	2240	non-null	int64
10	Hobi	2240	non-null	int64
11	Pakaian	2240	non-null	int64
12	Buku	2240	non-null	int64
13	Kesehatan	2240	non-null	int64
14	Jml Transaksi Diskon	2240	non-null	int64
15	Jml Transaksi App	2240	non-null	int64
16	Jml Transaksi Web	2240	non-null	int64
17	Jml Transaksi Outlet	2240	non-null	int64
18	Jml Kunjungan App	2240	non-null	int64
19	Komplain	2240	non-null	object
20	Pulau	2240	non-null	object
21	Urutan Terima Penawaran	609	non-null	object

dtypes: int64(15), object(7)

memory usage: 385.1+ KB

Ada atribut yang memiliki **nama yang tidak general/umum** sehingga perlu diubah nama atributnya yaitu:

Pergawean (Bahasa Jawa) -> Pekerjaan (Bahasa Indonesia)

Terdapat beberapa atribut yang **tipe datanya belum sesuai** sehingga kami mencoba untuk mengubah tipe data atribut tersebut

- Tanggal Lahir (Object -> DateTime)
- Penghasilan per Tahun (Object -> Integer)

Deskripsi Dataset

Tampilan 5 data pertama

Tanggal Lahir	Pergawean	Status Kawin	Penghasilan per Tahun	Tanggungan Ortu	Tanggungan Anak	Mulai Berlangganan	Transaksi Terakhir	Sembako	Elektronik	...
5/29/1976	Kepala Divisi	Cerai	Rp66.476.000,-	0	1	-3140	99	372558	18877	...
10/5/1977	C-level	Menikah	Rp31.056.000,-	1	0	-3184	99	5558	10877	...
5/30/1976	Karyawan	Cerai	Rp46.310.000,-	1	0	-3234	99	185558	2877	...
6/26/1978	Karyawan	Menikah	Rp65.819.000,-	0	0	-3238	99	267558	38877	...
9/27/1969	Kepala Divisi	Menikah	Rp94.871.000,-	0	2	-3327	99	169558	24877	...

Buku	Kesehatan	Jml Transaksi Diskon	Jml Transaksi App	Jml Transaksi Web	Jml Transaksi Outlet	Jml Kunjungan App	Komplain	Pulau	Urutan Terima Penawaran
189861	218702	1	4	4	6	1	Tidak pernah	Jawa	5
861	37702	1	7	3	7	5	Tidak pernah	Sumatera	[1, 5]
2861	30702	1	3	2	5	2	Tidak pernah	NTT	NaN
861	702	1	1	0	2	7	Tidak pernah	Kalimantan	NaN
861	34702	2	3	1	2	7	Tidak pernah	Jawa	[2, 5]

Atribut "Urutan Terima Penawaran" :

- terdapat dalam bentuk Array dan perlu di-encode agar bisa dianalisis
- nilai NaN nya berarti bisa diartikan Customer belum menerima tawaran hingga tawaran terakhir

Deskripsi Dataset

Tampilan Deskriptif Statistik dari data

	Tanggungan Ortu	Tanggungan Anak	Mulai Berlangganan	Transaksi Terakhir
count	2240.000000	2240.000000	2240.000000	2240.000000
mean	0.444196	0.506250	-3014.582143	49.109375
std	0.538398	0.544538	202.122512	28.962453
min	0.000000	0.000000	-3360.000000	0.000000
25%	0.000000	0.000000	-3190.000000	24.000000
50%	0.000000	0.000000	-3016.500000	49.000000
75%	1.000000	1.000000	-2841.750000	74.000000
max	2.000000	2.000000	-2661.000000	99.000000

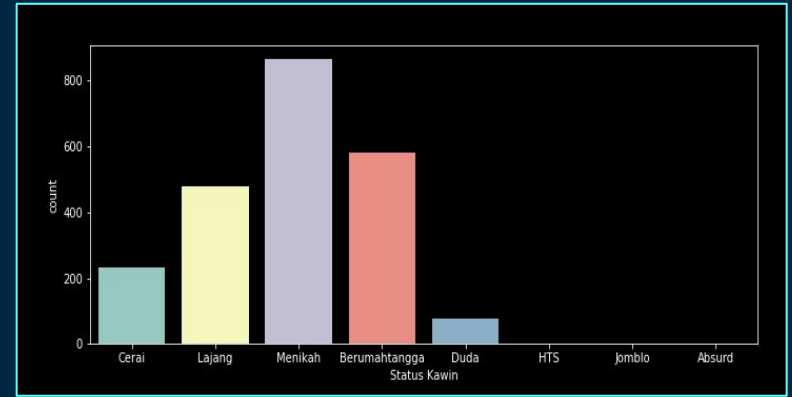
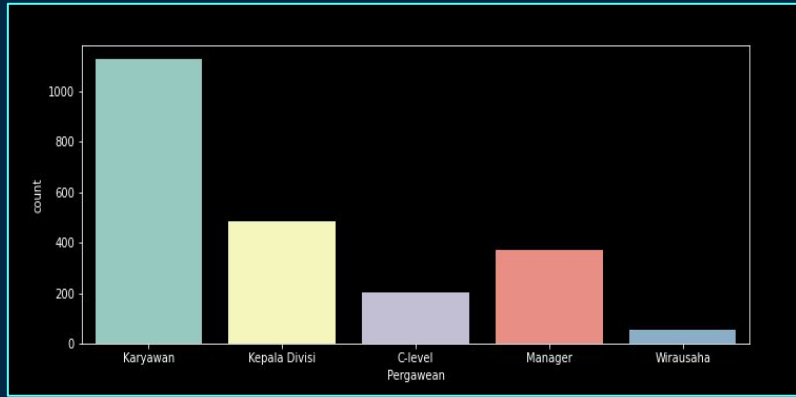
Mulai Berlangganan : memiliki nilai max yaitu -2661 dan nilai min yaitu -3360. Dari sini berarti dataset ini hanya berisi data customer lama (yang telah berlangganan setidaknya lebih dari 7 tahun)

	Tanggal Lahir	Pergawean	Status Kawin	Penghasilan per Tahun	Komplain
count	2240	2240	2240	2216	2240
unique	2106	5	8	1974	2
top	1/16/1976	Karyawan	Menikah	Rp7.500.000,-	Tidak pernah
freq	3	1127	864	12	2219

Komplain : terdapat 2 nilai unik dan modus yaitu Tidak pernah dengan frekuensi kemunculan 2219, berarti sekitar 99% customer tidak pernah melakukan komplain.

Exploratory Data Analysis

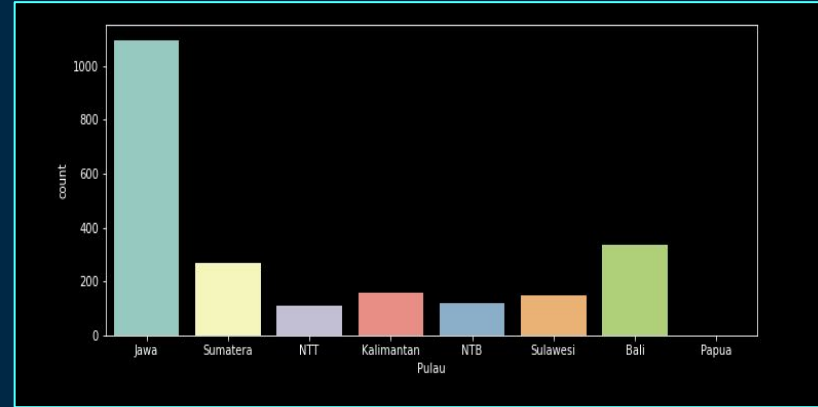
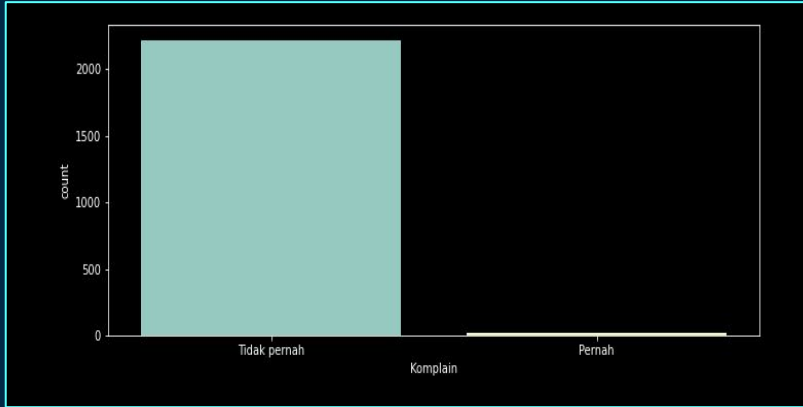
Countplot untuk Data Kategorik



- **Pergawean** : Terdapat 5 nilai unik dari fitur ini, namun sebenarnya Kepala Divisi, C-Level, Manager overlap dengan Karyawan
- **Status Kawin** : Terdapat 8 nilai unik dari fitur ini, namun ada beberapa atribut yang overlap seperti Cerai/Duda, Menikah/Berumah Tangga, Lajang/Jomblo, serta beberapa nilai unik yang tidak diketahui status kawinnya (HTS, Absurd)

Exploratory Data Analysis

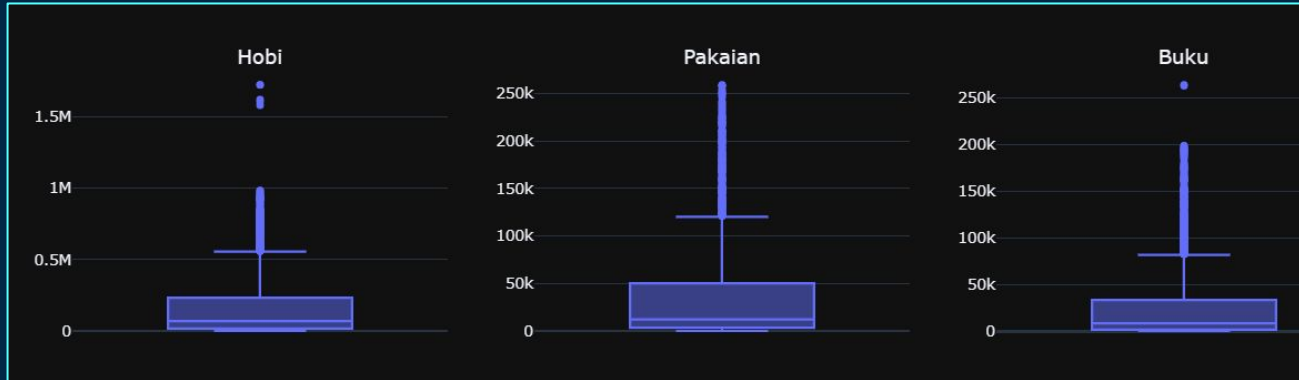
Countplot untuk Data Kategorik



- **Komplain** : Terlihat bahwa perbandingan jumlah Customer yang pernah komplain dan tidak pernah Komplain sangat jauh
- **Pulau** : Terlihat bahwa Customer tersebar ke 8 pulau, dengan Customer terbanyak terdapat di pulau Jawa, dan paling sedikit terdapat di pulau Papua.

Exploratory Data Analysis

Boxplot untuk Data Numerik



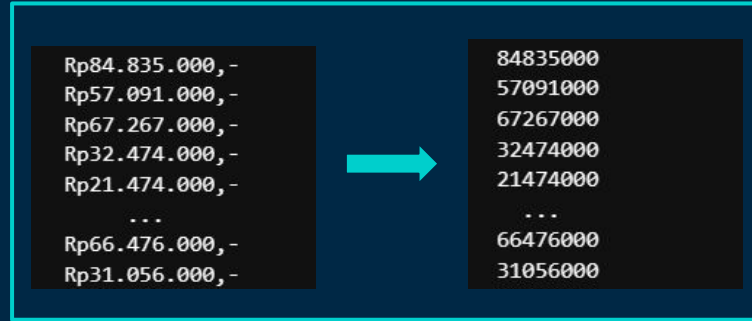
Terdapat outlier pada beberapa atribut, diantaranya pada atribut Sembako, Elektronik, Hobi, Pakaian, Buku, Kesehatan, Jml Transaksi Diskon, Jml Transaksi App, Jml Transaksi Web, dan Jml Kunjungan App.

Data Preparation

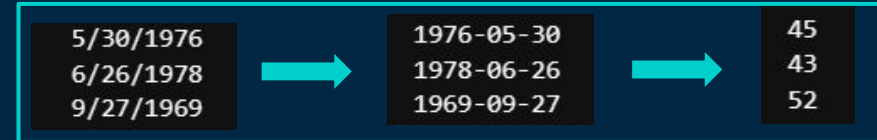
03

Data Cleaning

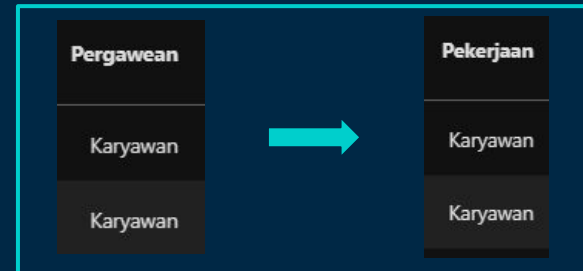
Penghasilan Per Tahun : menghilangkan tanda "Rp", ",", dan "." agar menghasilkan data yang hanya berupa angka. Kemudian mengubah tipe datanya dari Object menjadi Numeric



Tanggal Lahir : mengubah tipe data nya menjadi datetime dan membuat atribut baru dengan nama umur.



Pergawean : mengubah nama kolomnya menjadi "Pekerjaan" agar lebih general



Data Encoding

Status Kawin : Atribut ini memiliki 8 nilai unik, namun terdapat value yang dapat dijadikan kategori yang sama, seperti Cerai/Duda, Menikah/Berumah Tangga, Lajang/Jomblo, serta HTS/Absurd. Dilakukan encoding berdasarkan 4 kategori tersebut. Encoding menjadi kategori 1 : Cerai/Duda, 2 : Menikah/Berumah Tangga, 3 : Lajang/Jomblo, 4 : HTS/Absurd

```
dataset["Status Kawin"].unique()  
  
array(['Cerai', 'Lajang', 'Menikah', 'Berumahtangga', 'Duda', 'HTS',  
      'Jomblo', 'Absurd'], dtype=object)
```



```
dataset["Status Kawin"].unique()  
  
array([1, 3, 2, 4], dtype=int64)
```

Komplain : Sesuai deskripsi dataset, dilakukan encoding untuk mengubah valuenya menjadi integer 0 atau 1

```
dataset["Komplain"].unique()  
  
array(['Tidak pernah', 'Pernah'])
```



```
dataset["Komplain"].unique()  
  
array([0, 1], dtype=int64)
```

Data Encoding

Pekerjaan : Atribut ini memiliki 5 nilai unik, namun terdapat value yang dapat dijadikan kategori yang sama, seperti Karyawan/Manager/C-Level. Dilakukan encoding kategori untuk kategori Manager/C-Level/Kepala Divisi menjadi Karyawan.

```
dataset['Pekerjaan'].unique()  
  
array(['Karyawan', 'Kepala Divisi', 'C-level', 'Manager', 'Wirausaha'],  
      dtype=object)
```



```
dataset['Pekerjaan'].unique()  
  
array(['Karyawan', 'Wirausaha'], dtype=object)
```


Handle Missing Values

	Total	Percent
Urutan Terima Penawaran	1599	0.727480
Penghasilan per Tahun	24	0.010919

Missing value yang ada pada dataset kami hanya pada 2 fitur yaitu Urutan Terima Penawaran dan Penghasilan per Tahun. Missing value memiliki persentase yang cukup tinggi sehingga opsi kami pilih untuk handle missing value dengan melakukan replace nilai Null terhadap nilai median nya. Hal ini dilakukan agar tidak ada informasi penting yang hilang.

Untuk fitur Penghasilan per Tahun dilakukan replace nilai Null terhadap median nya, sedangkan Untuk fitur Urutan Terima Penawaran dilakukan replace nilai Null terhadap angka -1 untuk mengindikasikan bahwa customer tersebut belum menerima penawaran apapun.

Handle Outlier

Ada 11 fitur yang memiliki outlier. Karena jumlah outlier yang banyak dengan total hampir setengah dari banyak data, maka kami memilih opsi untuk mereplace outlier dengan nilai median agar tidak ada informasi yang hilang.

	Total	Percent	Range
Sembako	34	0.015469	1493000.0
Elektronik	240	0.109190	199000.0
Hobi	169	0.076888	1725000.0
Pakaian	220	0.100091	259000.0
Buku	234	0.106460	263000.0
Kesehatan	203	0.092357	362000.0
Jml Transaksi Diskon	85	0.038672	15.0
Jml Transaksi App	4	0.001820	27.0
Jml Transaksi Web	23	0.010464	28.0
Jml Kunjungan App	8	0.003640	20.0
Umur	3	0.001365	103.0

Data Encoding Atribut 'Urutan Terima Penawaran'

Seperti yang kita analisis sebelumnya pada Data Understanding, perlu dilakukan encode. Kami mengubah pun atribut ini menjadi 3 atribut berjenis berbeda, yaitu :

1. Mulai Terima pada Penawaran ke (Tawaran pertama yang diterima, inf jika belum ada tawaran yang diterima)
2. Terima Penawaran 0-5 (1 bila menerima tawaran ke i)
3. Jml Tawaran Diterima (total banyak tawaran yang diterima)

Urutan Terima Penawaran
5
[1, 5]
-1



Mulai Terima pada Penawaran ke	Terima Penawaran 0	Terima Penawaran 1	Terima Penawaran 2	Terima Penawaran 3	Terima Penawaran 4	Terima Penawaran 5	Jml Tawaran Diterima
5.0	0	0	0	0	0	1	1
1.0	0	1	0	0	0	1	2
inf	0	0	0	0	0	0	0

Modelling and Task Solving

04

Customer Segmentation [Task 1]

Buatlah cluster-cluster customer berdasarkan fitur yang cocok menurut Anda! Gunakan satu model clustering dan juga model RFM (simpan kalkulasi RFM sebagai fitur untuk soal selanjutnya). Apakah cluster yang dihasilkan dapat merepresentasikan customer berdasarkan suatu fakta



Behavioral Model

Primer, Sekunder,
Tersier



RFM Model

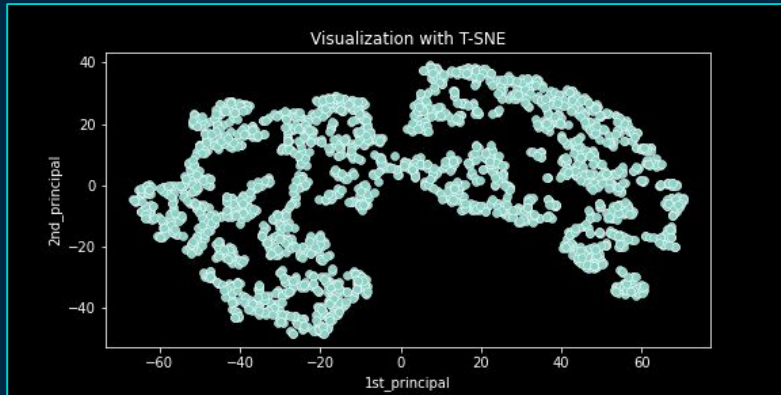
Recency, Frequency,
Monetary

Customer Segmentation with Behavioral Model

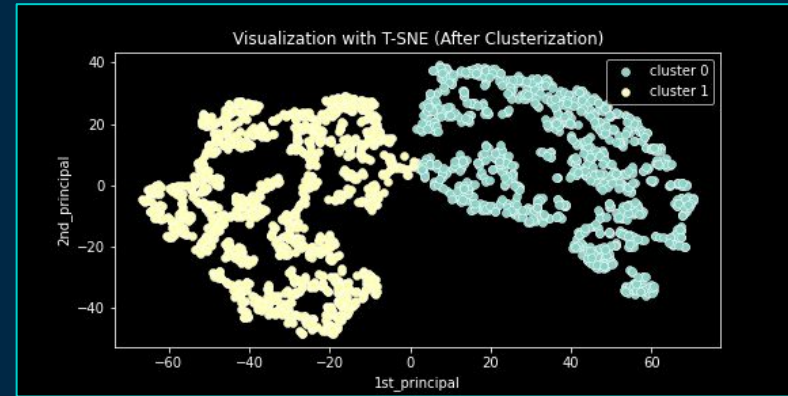
Define Feature for Behavior “Customer Needs of Product” :

1. Primer : Sembako, Pakaian, Kesehatan
2. Sekunder : Buku, Elektronik
3. Tersier : Hobi

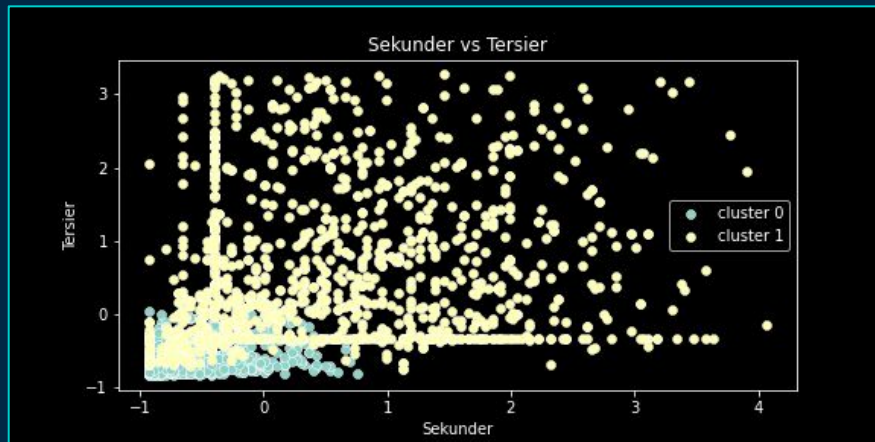
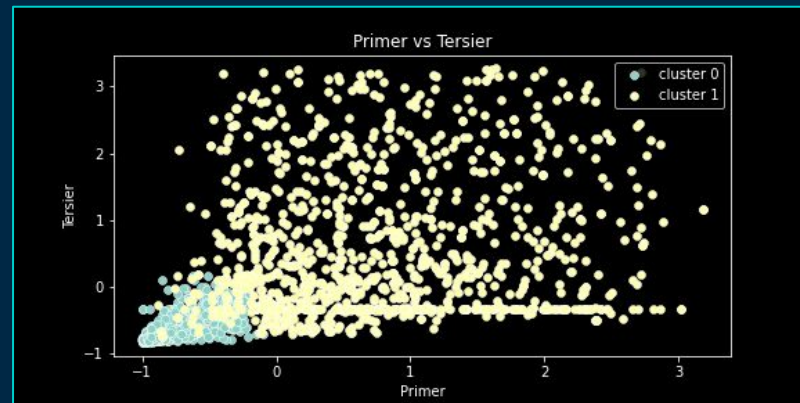
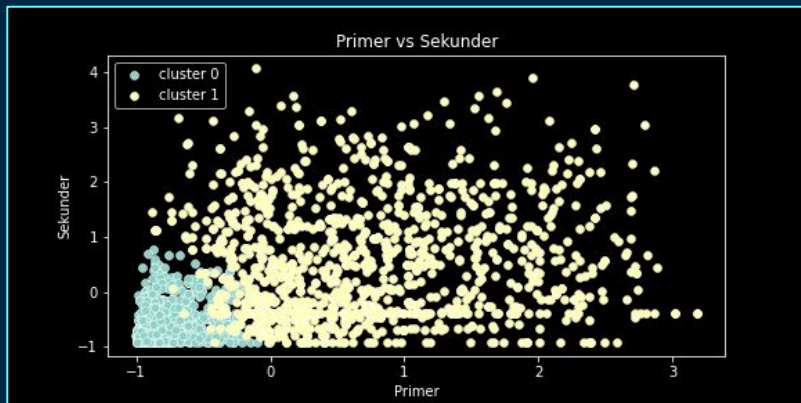
T-SNE Visualization



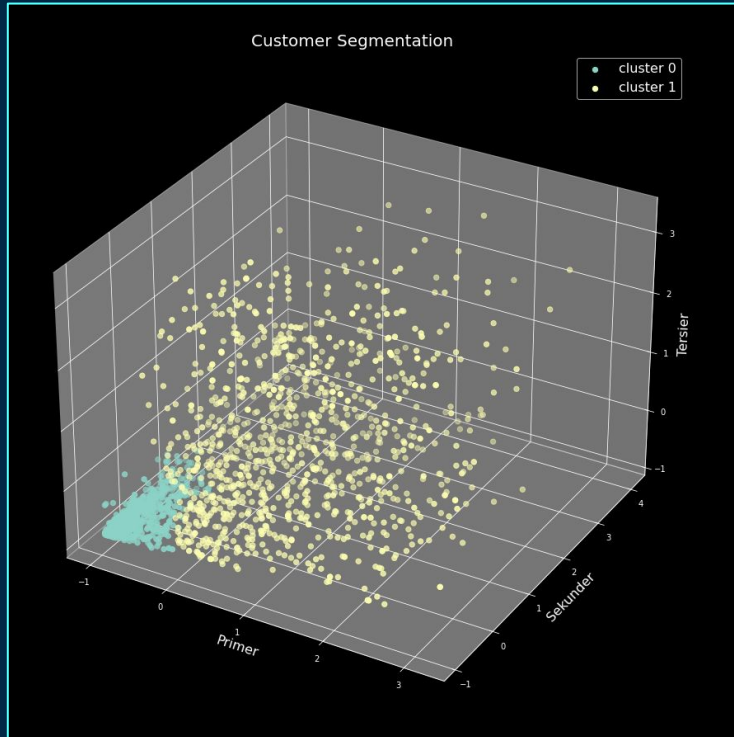
K-Means
Clustering
(K=2)



2D Visualization



3D Visualization



Cluster 1 (Dependent Customer) : High Primer, High Sekunder dan High Tersier.



Cluster 0 (Non-Dependent Customer) : Less Primer, Less Sekunder dan Less Tersier.

Usulan Strategi Marketing

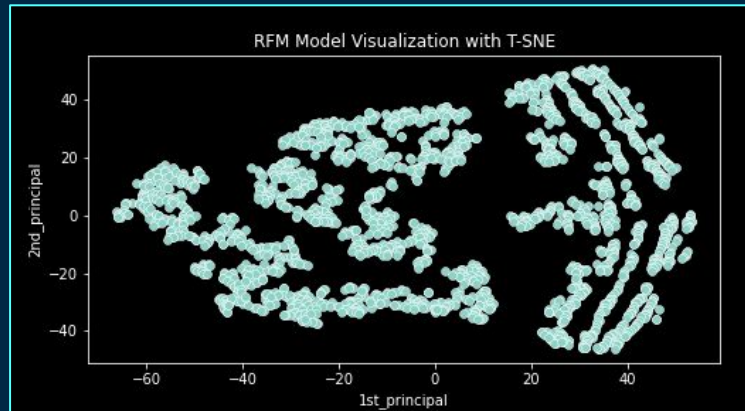
- Untuk Customer pada Cluster 1 (Dependent Customer) perlu adanya tambahan pemberian penawaran yang berdasarkan tipe kebutuhan yang belum diminatinya, misalnya jika ia sudah minat kebutuhan primer, maka kita bisa memberikan offer di kebutuhan sekunder atau tersier, dengan meningkatkan ketergantungan customer kita akan dapat meningkatkan penjualan.
- Untuk Customer pada Cluster 0 (Non-Dependent Customer), jumlahnya bisa diminimalkan dengan lebih meningkatkan branding perusahaan serta memberikan harga yang lebih menarik dibandingkan kompetitor, dengan ini akan menarik customer-customer ini untuk bergantung ke perusahaan kita dalam aspek kebutuhan.

Customer Segmentation with RFM Model

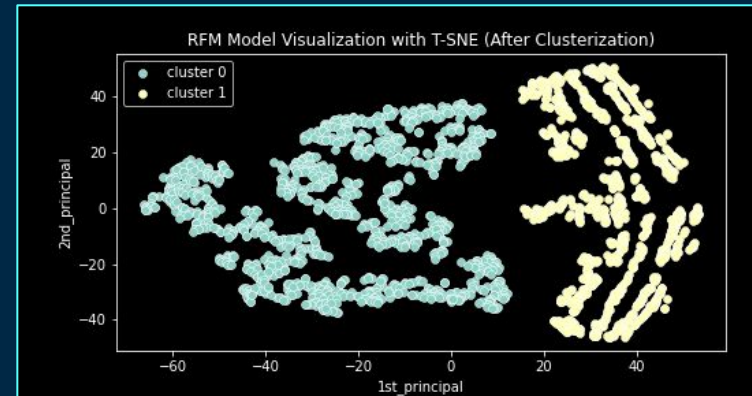
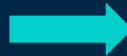
Define feature for RFM Model : (Anitha et al, 2019)

- Recency** : It refers to the number of days before the reference date when a customer made the last purchase. Lesser the value of recency, higher is the customer visit to a store. (Transaksi Terakhir)
- Frequency** : It is the period between two subsequent purchases of a customer. Higher the value of Frequency, more is the customer visit to the company. (Jml Transaksi App, Jml Transaksi Web, Jml Transaksi Outlet)
- Monetary** : This refers to the amount of money spent by a customer during a specific period of time. Higher the value, more is the profit generated to the company. (Sembako, Elektronik, Pakaian, Hobi, Buku, Kesehatan)

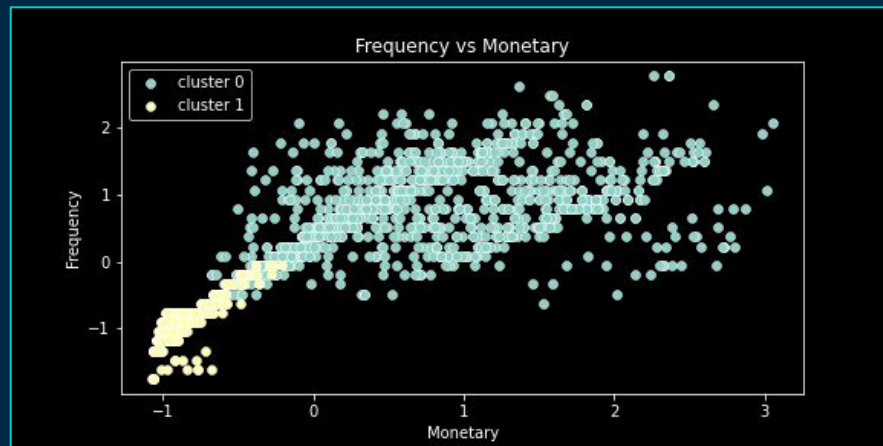
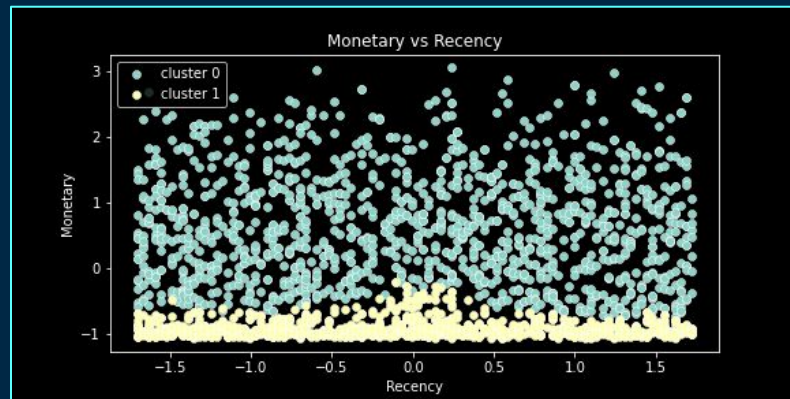
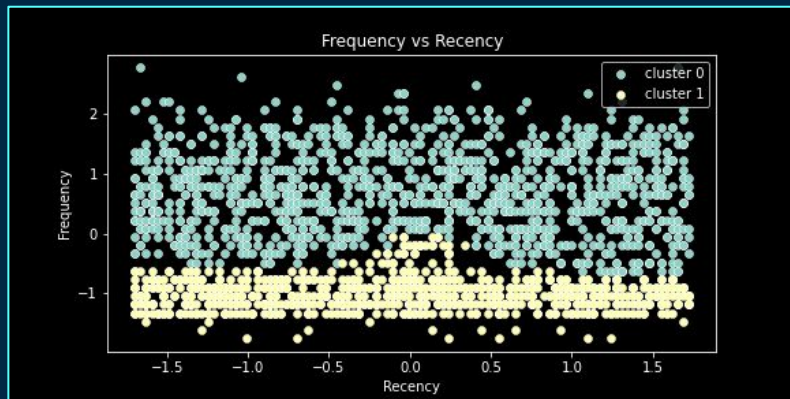
T-SNE Visualization



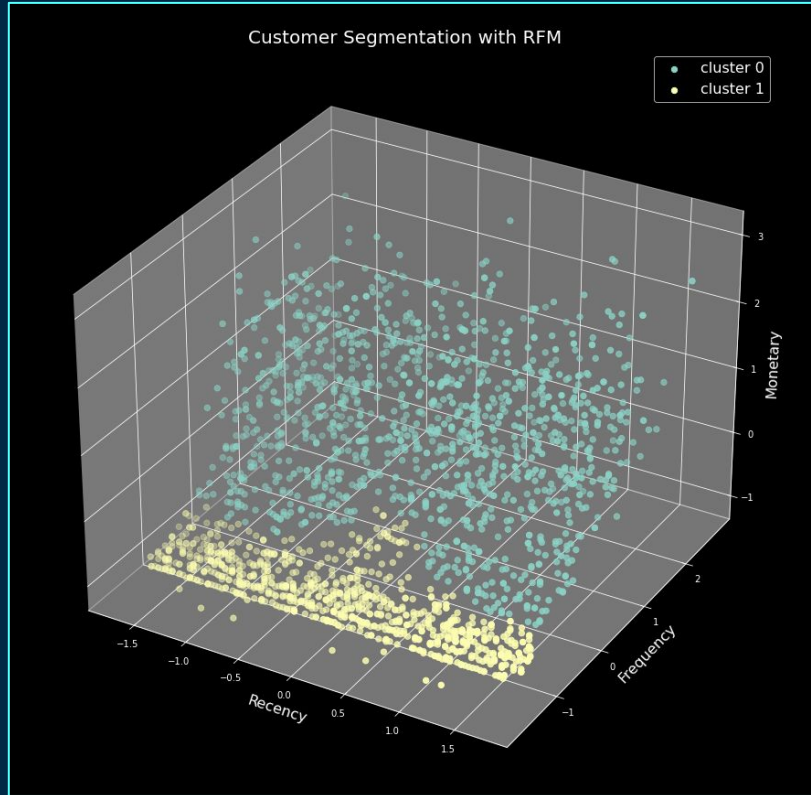
Agglomerative
Clustering
(Single Link)



2D Visualization



3D Visualization



Cluster 0 (Potential Customer) : High Frequency, High Monetary and Spread Evenly Recency



Cluster 1 (Unreachable Customer) : Less Frequency, Less Monetary, Spread Evenly Recency

Usulan Strategi Marketing

- Untuk Customer pada Cluster 0 (Potential Customer) , kita bisa memberikan small offer kepada customer yg recencynya tinggi agar tetap membeli di perusahaan kita.
- Untuk Customer pada Cluster 1 (Unreachable Customer), jumlahnya bisa diminimalkan dengan lebih menjangkau kembali Cluster dengan tipe ini, misalnya dengan big offer untuk Customer lama yang melakukan pembelian kembali.



Customer Offer Acceptance Analysis [Task 2]

Bagaimana kecenderungan customer menerima suatu tawaran lebih dari sekali dan kira-kira butuh berapa kali penawaran hingga customer dapat menerima tawaran?

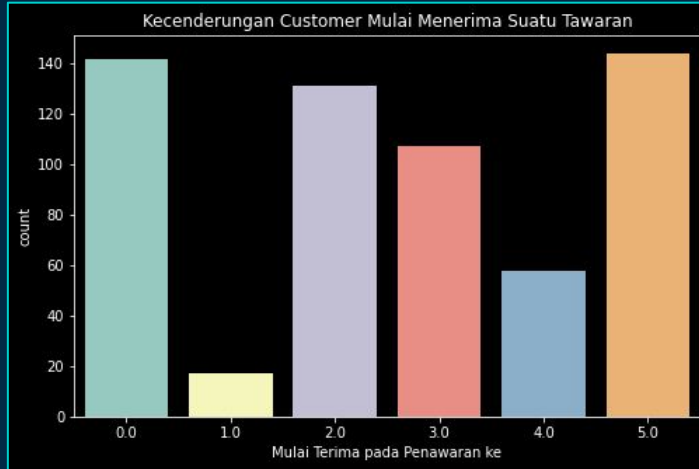


Customer Offer Acceptance Analysis [Task 2]



- kecenderungan Customer dalam menerima suatu tawaran menurun secara geometris
- Kebanyakan Customer hanya menerima tawaran sekali,
- hanya sedikit yang menerima 5 tawaran
- tidak ada yg menerima hingga 6 tawaran (seluruh tawaran)

Customer Offer Acceptance Analysis [Task 2]



	Count	Percentage
Mulai Terima pada Penawaran ke		
0.0	142	23.706177
1.0	17	2.838063
2.0	131	21.869783
3.0	107	17.863105
4.0	58	9.682805
5.0	144	24.040067

- Walau sudah cukup banyak customer yang menerima suatu tawaran pada tawaran pertama, tapi Customer cenderung mulai menerima suatu tawaran pada tawaran terakhir, sehingga butuh 5 tawaran sebelumnya untuk seorang Customer mau menerima tawaran yang diberikan.
- Sekitar 23% customer yang mau menerima tawaran di tawaran pertama (dari seluruh customer yang menerima suatu tawaran)

Hasil Analisis dan Usulan Strategi Marketing

- Ada kemungkinan customer tidak tertarik pada tawaran-tawaran yang diberikan terutama pada tawaran pertama yang diberikan.
- Usulan strategi marketingnya agar pada tawaran pertama bisa diberikan big offer (tawaran yg menarik), dengan memberikan tawaran di awal yang menarik, akan meningkatkan peluang suatu tawaran untuk diterima lebih dari sekali, karena yang menyebabkan tidak banyak tawaran yang diterima lebih dari sekali adalah kecenderungan seorang customer untuk menerima tawaran di tawaran pertama yang masih kurang (hanya 23%)

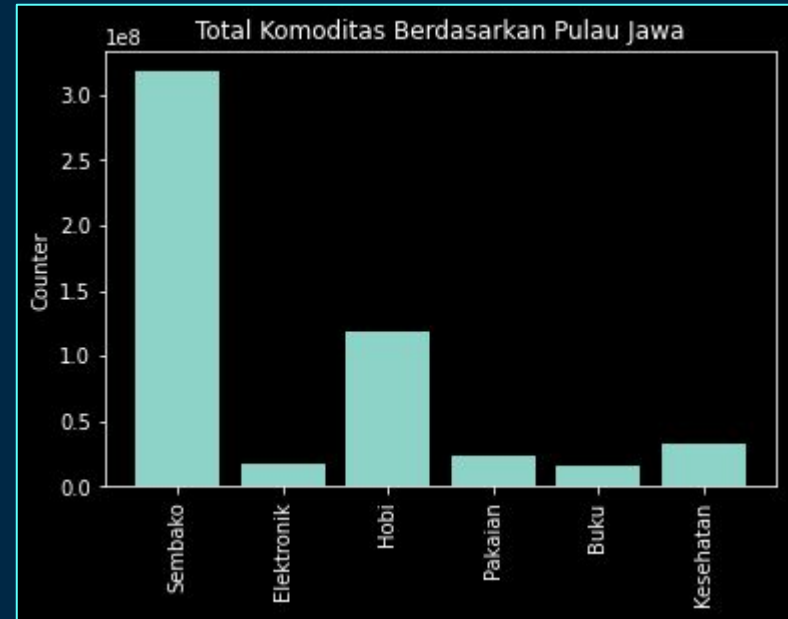


Profitable Comodity Analysis [Task 3]

Apa saja jenis komoditas yang diminati oleh customer dari masing-masing pulau dan pulau mana yang memberikan keuntungan terbesar pada setiap komoditas?



Jenis Komoditas yang Diminati oleh Customer dari masing-masing Pulau



Jenis Komoditas yang Diminati oleh Customer dari masing-masing Pulau



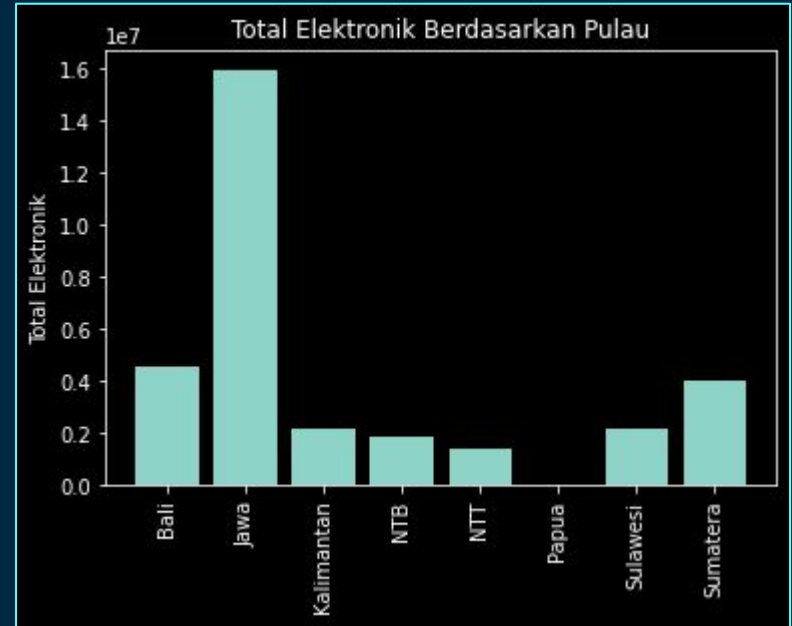
Jenis Komoditas yang Diminati oleh Customer dari masing-masing Pulau



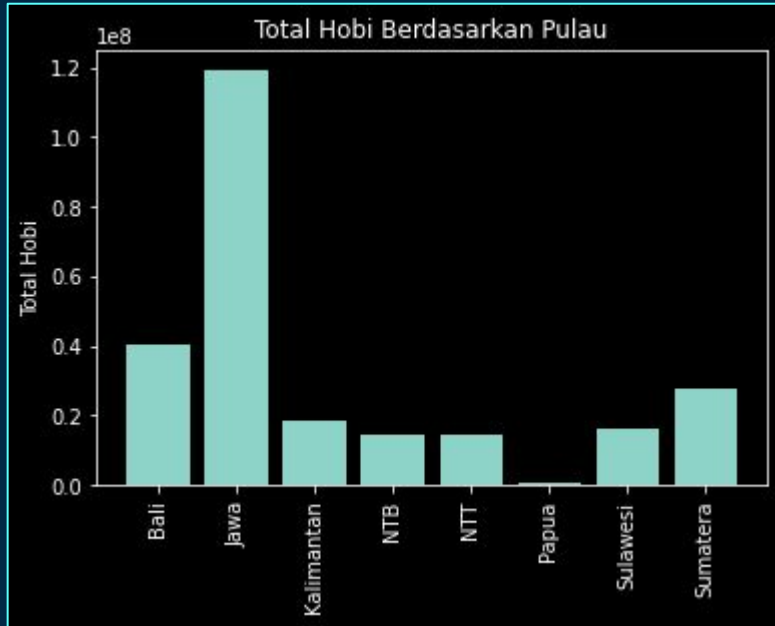
Jenis Komoditas yang Diminati oleh Customer dari masing-masing Pulau



Pulau yang memberikan keuntungan terbesar pada setiap Komoditas



Pulau yang memberikan keuntungan terbesar pada setiap Komoditas



Pulau yang memberikan keuntungan terbesar pada setiap Komoditas



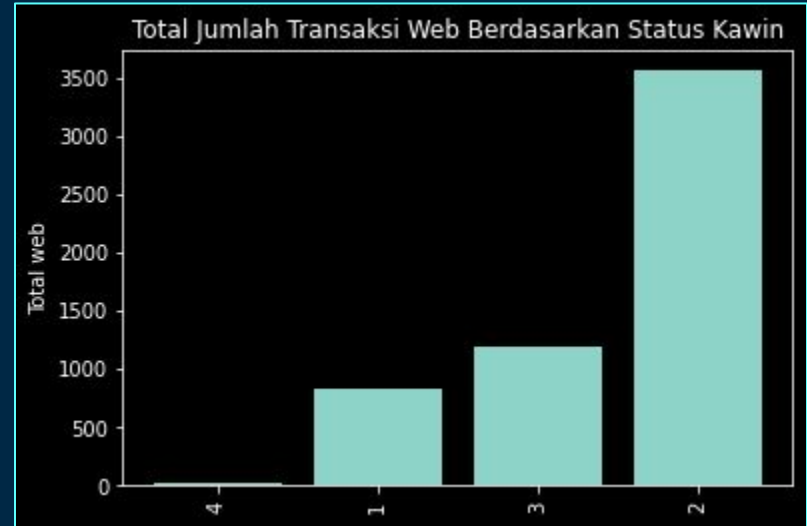
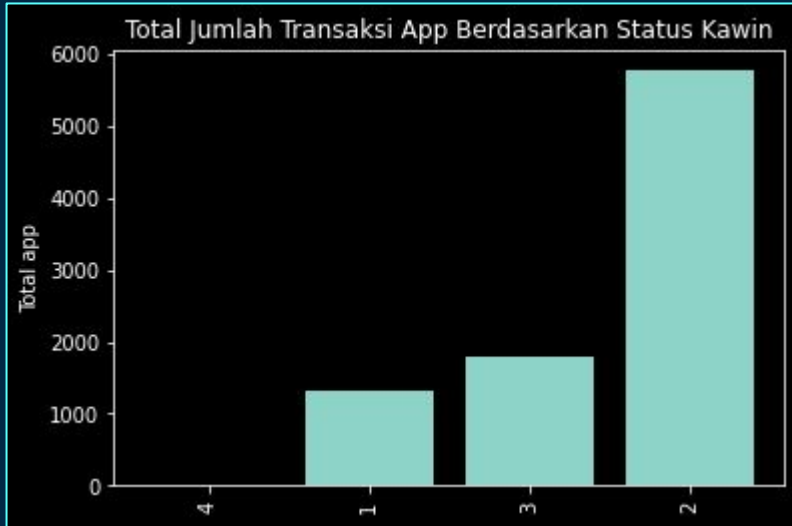
Hasil Analisis dan Usulan Strategi Marketing

- Berdasarkan analisis Komoditas yang paling diminati di setiap Pulau maka Sembako yang paling banyak diminati kemudian peringkat 2 paling diminati adalah hobi.
- Maka strategi marketing yang bisa dilakukan adalah meningkatkan kualitas dan variasi terhadap komoditas lain seperti elektronik, buku, pakaian, dan kesehatan. Lalu bisa juga dengan meningkatkan penawaran dengan memberikan promo menarik pada komoditas tersebut.
- Peningkatan promosi di luar pulau jawa agar penjualan lebih merata di seluruh pulau

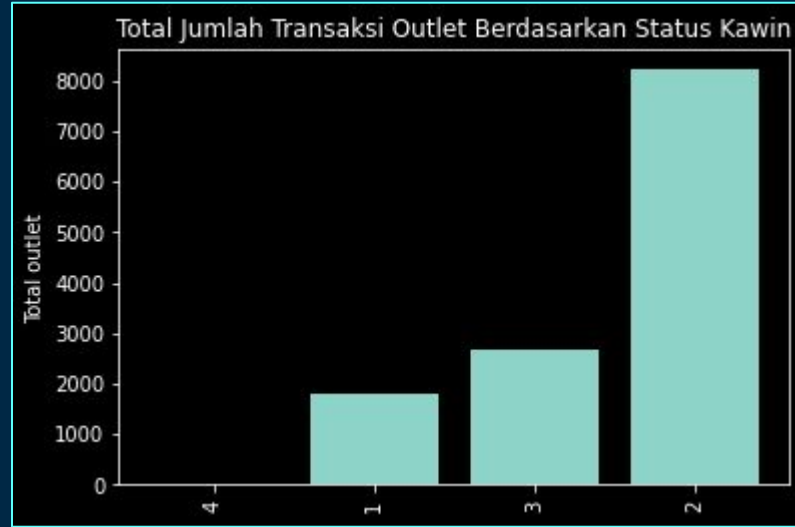
Platform Preference vs Customer Demographic Analysis [Task 4]

Bagaimana hubungan jenis platform yang sering customer pakai jika dibandingkan dengan demografi (umur, pekerjaan, status kawin)?

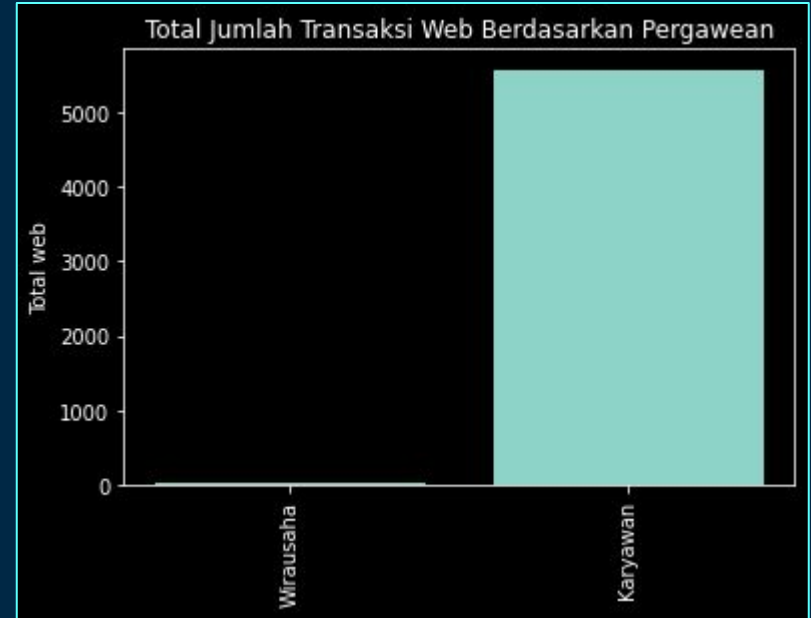
Jenis Platform yang Digunakan Berdasarkan Status Kawin



Jenis Platform yang Digunakan Berdasarkan Status Kawin



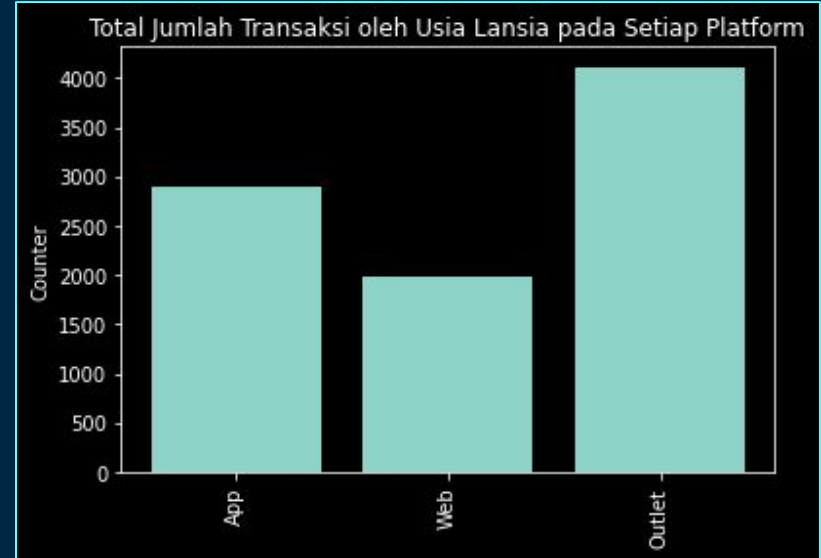
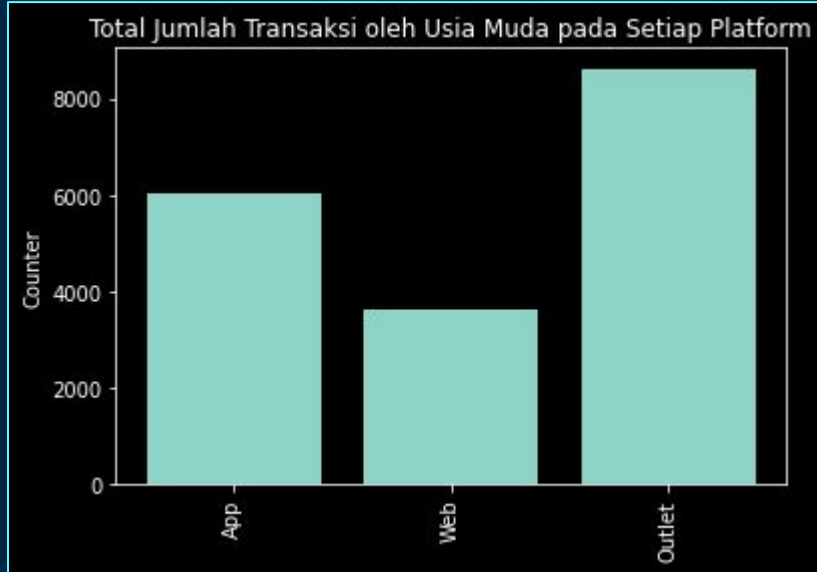
Jenis Platform yang Digunakan Berdasarkan Jenis Pekerjaan



Jenis Platform yang Digunakan Berdasarkan Jenis Pekerjaan



Jenis Platform yang Digunakan Berdasarkan Usia



Hasil Analisis dan Usulan Strategi Marketing

Berdasarkan analisis penggunaan platform berdasarkan usia maka strategi marketing yang dapat diterapkan adalah

- meningkatkan kualitas pelayanan di outlet karena baik usia muda maupun dewasa paling banyak transaksi melalui outlet
- Selanjutnya perlu juga dilakukan peningkatan kualitas web dan app karena transaksi melalui web ternyata yang paling sedikit. Hal ini bisa jadi karena ada masalah pada fitur yang ada di web dan app. Sehingga kami merekomendasikan untuk dilakukan analisis lebih lanjut terhadap fitur yang ada di web dan app.

Hasil Analisis dan Usulan Strategi Marketing

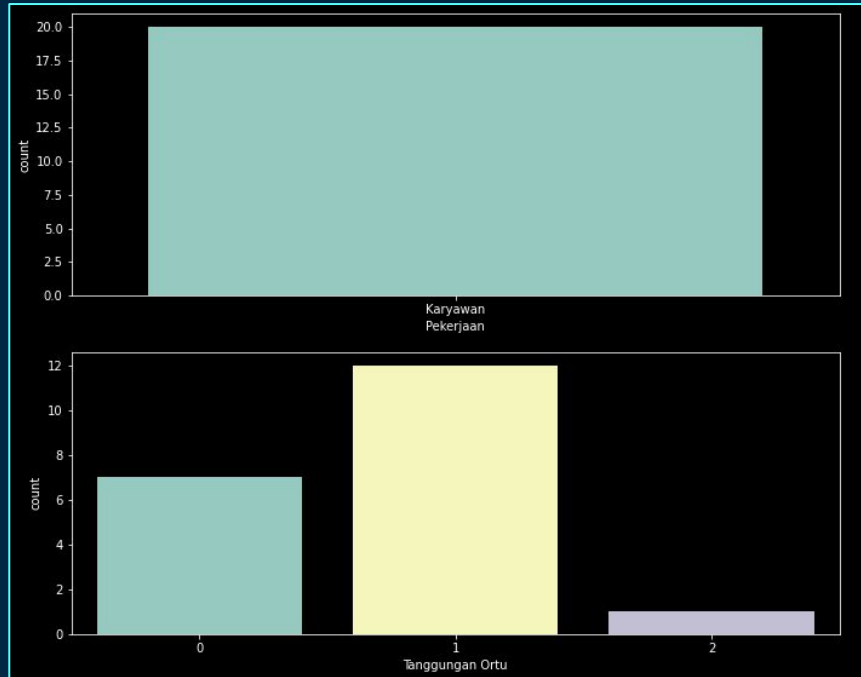
- Berdasarkan analisis penggunaan platform berdasarkan jenis pekerjaan maka strategi marketing yang dapat diterapkan adalah meningkatkan promosi agar semua kalangan bisa dijangkau oleh penjual. Hal ini karena ternyata kalangan karyawan saja yang lebih sering melakukan transaksi melalui berbagai platform
- Berdasarkan analisis penggunaan platform berdasarkan status kawin maka strategi marketing yang dapat diterapkan adalah menambahkan variasi produk yang dijual agar semua kalangan dari berbagai status tertarik untuk melakukan transaksi pada perusahaan marketing ini. Hal ini karena terlihat bahwa yang sering melakukan transaksi adalah yang statusnya menikah dan berumahtangga. Hal ini bisa saja karena produk yang dijual memang lebih menawarkan produk yang dibutuhkan oleh orang yang berumahtangga.

Customer Complaint Analysis and Prediction Model [Task 5]

Informasi apa yang bisa kalian gali dari pelanggan yang melakukan komplain? Apakah komplain dapat diprediksi?



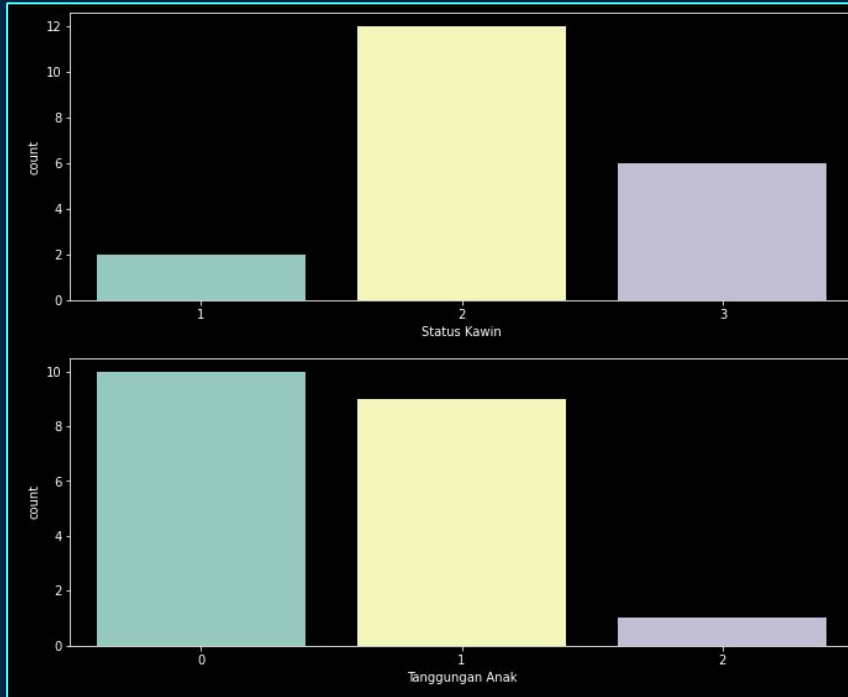
Informasi yang didapatkan dari pelanggan yang melakukan komplain [Task 5]



Customer yang paling banyak melakukan komplain, yaitu:

- Dari 20 customer yang melakukan komplain, semuanya merupakan **Karyawan**
- Customer yang memiliki jumlah **tanggungjan orang tua** sebanyak 1 orang, yaitu sebanyak 12 customer

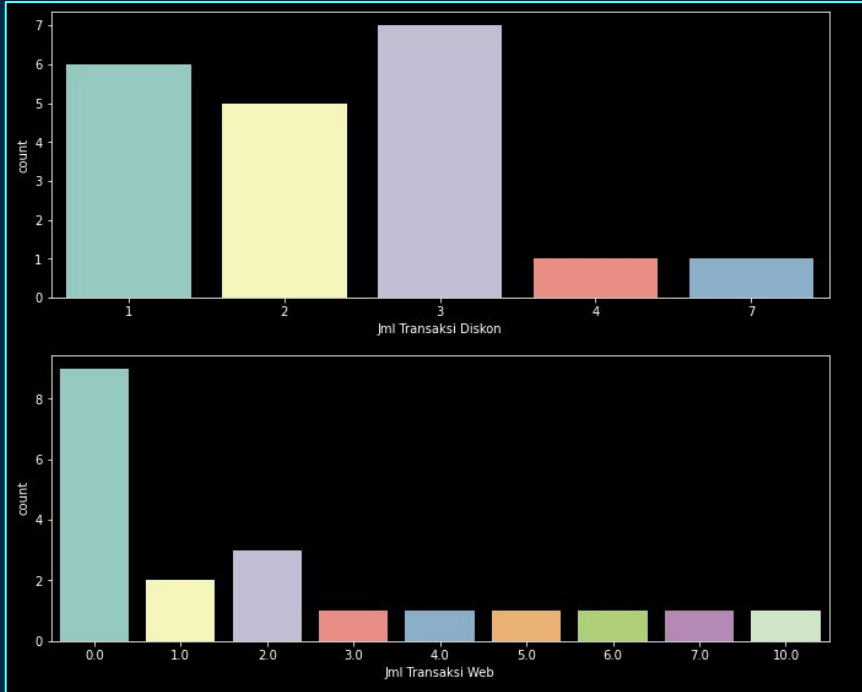
Informasi yang didapatkan dari pelanggan yang melakukan komplain [Task 5]



Customer yang paling banyak melakukan komplain, yaitu:

- Customer yang sudah menikah/berumah tangga, yaitu sebanyak 12 customer
- Customer yang tidak memiliki tanggungan anak, yaitu sebanyak 10 customer

Informasi yang didapatkan dari pelanggan yang melakukan komplain [Task 5]

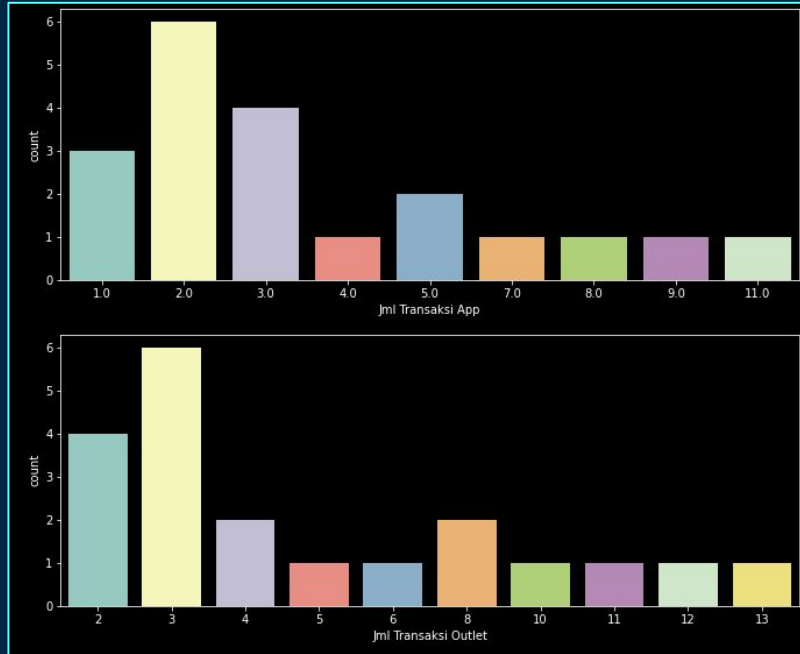


Customer yang paling banyak melakukan komplain, yaitu:

- Customer yang melakukan transaksi dengan diskon sejumlah 3 kali, yaitu sebanyak 7 customer
- Customer yang tidak pernah melakukan transaksi melalui web, yaitu sebanyak 9 customer

Customer yang melakukan komplain cenderung merupakan customer yang baru melakukan sedikit transaksi dengan diskon maupun transaksi melalui Web

Informasi yang didapatkan dari pelanggan yang melakukan komplain [Task 5]

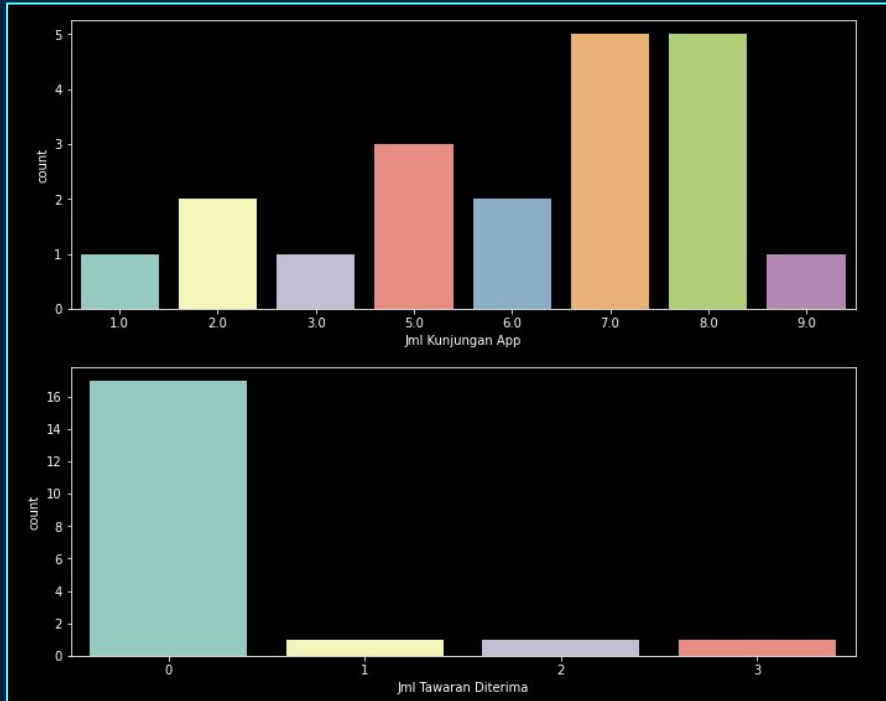


Customer yang paling banyak melakukan komplain, yaitu:

- Customer yang melakukan transaksi melalui app sejumlah 2 kali, yaitu sebanyak 6 customer
- Customer yang melakukan transaksi melalui outlet sejumlah 3 kali, yaitu sebanyak 6 customer

Customer yang melakukan komplain cenderung merupakan customer yang baru melakukan sedikit transaksi baik melalui App maupun Outlet

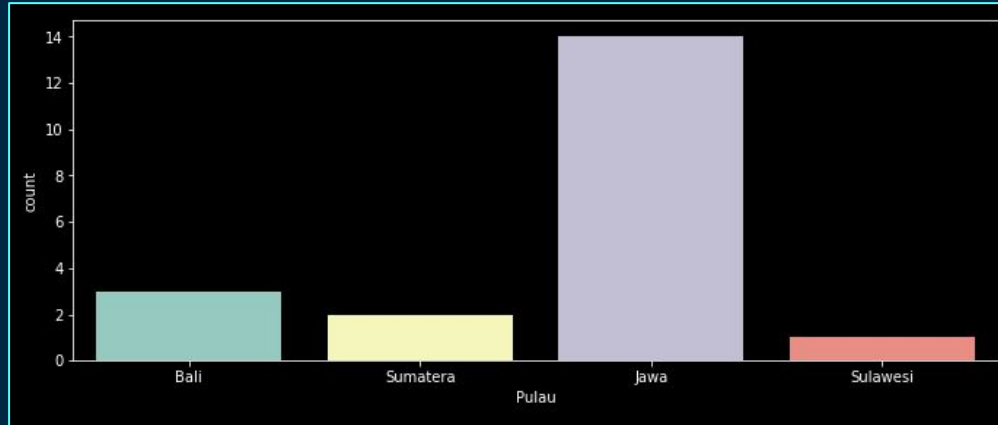
Informasi yang didapatkan dari pelanggan yang melakukan komplain [Task 5]



Customer yang paling banyak melakukan komplain, yaitu:

- Customer yang mengunjungi app sejumlah 7 atau 8 kali, yaitu sebanyak 5 customer
- Customer yang tidak pernah menerima tawaran, yaitu sebanyak 17 customer

Informasi yang didapatkan dari pelanggan yang melakukan komplain [Task 5]



Customer yang paling banyak melakukan komplain, yaitu:

- Customer yang berdomisili di Pulau Jawa, yaitu sebanyak 14 customer

Apakah komplain dapat diprediksi? [Task 5]

Jumlah customer yang pernah melakukan komplain sebanyak 20 customer, sedangkan jumlah customer yang tidak pernah melakukan komplain sebanyak 2178 customer.

Perbandingan antara jumlah customer yang pernah melakukan komplain dan yang tidak pernah melakukan komplain tersebut cukup besar. Hal ini dapat mengindikasikan adanya *imbalance* pada data.

Oleh karena itu, untuk memprediksi atribut 'Komplain' perlu untuk dilakukan *re-sample* pada *training set* sebagai teknik untuk melakukan *imbalance classification*, salah satunya yaitu dengan *teknik oversampling*:

- Random Oversampling
- SMOTE
- Borderline SMOTE
- Borderline-SMOTE SVM
- ADASYN

Discount vs Offer Acceptance Analysis

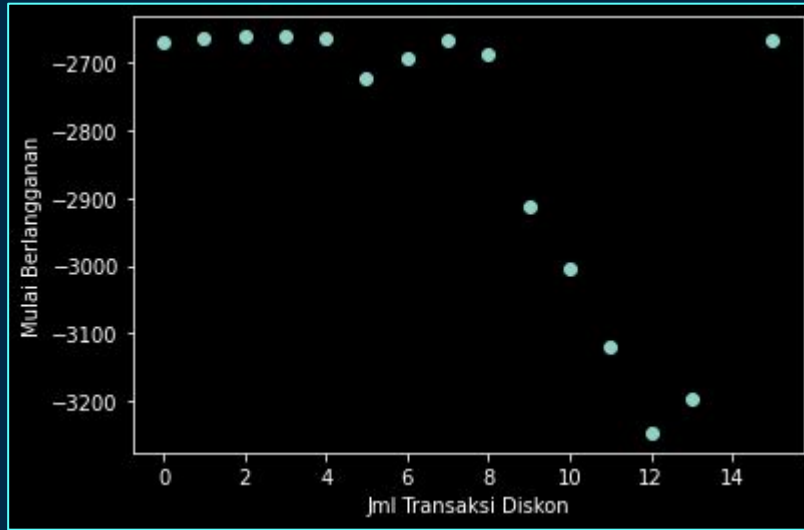
[Task 6]

Seberapa sering discount diberikan ke customer?
Adakah hubungan antara discount dengan offer acceptance?



Seberapa sering discount diberikan ke customer? [Task 6]

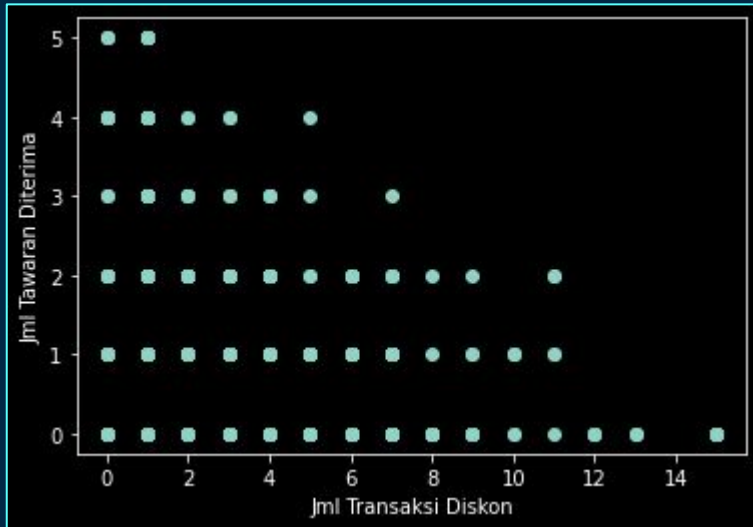
Analisis dilakukan dengan melihat data jumlah hari customer berlangganan yang paling minimal berdasarkan masing-masing jumlah transaksi diskon yang dilakukan customer.



Diketahui bahwa customer pada dataset ini sudah berlangganan selama setidaknya 7 tahun, sementara jumlah maksimal transaksi diskon yang dilakukan customer adalah sebanyak 15 transaksi.

Dapat disimpulkan bahwa diskon yang diberikan ke customer cukup jarang.

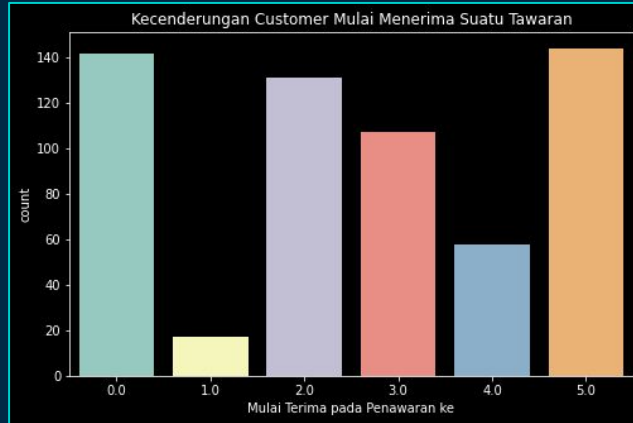
Adakah hubungan antara discount dengan offer acceptance? [Task 6]



Diskon merupakan salah satu penawaran yang umum ditawarkan kepada customer pada suatu praktik bisnis. Namun, pada kasus ini tidak dipastikan bahwa penawaran yang diberikan hanya sebatas diskon saja, sehingga dapat pula diberikan jenis penawaran lainnya. Pada visualisasi tersebut digambarkan bahwa terdapat customer yang jumlah transaksi diskonnya 0, namun menerima sejumlah tawaran tertentu (yang bisa saja bukan jenis penawaran diskon).

Dari visualisasi tersebut, didapatkan pula informasi bahwa terdapat customer yang tidak menerima tawaran namun melakukan sejumlah transaksi diskon. Dengan begitu, kami menyimpulkan bahwa tawaran dapat diberikan hanya pada periode waktu tertentu, sedangkan transaksi diskon dihitung sejak customer mulai berlangganan.

Customer Offer Acceptance Prediction Model [Task 7]



Multi Class Classification (Hota et al, 2018)

Menggunakan K-Nearest Neighbor, dioptimisasi dengan Grid CV untuk mencari hyperparameter tuning yang tepat

Imbalanced Classification

Oversampling : Borderline
SMOTE

Sembako Nominal Transaction Prediction Model [Task 8]

Menggunakan 3 model, yaitu :

- Linear Regression
- Ridge Regression
- Lasso Regression

Model yang tepat dilihat berdasarkan nilai R^2 , MSE, RMSE, dan MAE.

Berdasarkan fitur yang digunakan, apakah terdapat data leakage? [Task 8]

Fitur yang digunakan pada task ini yaitu, Penghasilan per Tahun, Tanggungan Ortu, Tanggungan Anak, Mulai Berlangganan, Transaksi Terakhir, Sembako, Elektronik, Hobi, Pakaian, Buku, Kesehatan, Jml Transaksi Diskon, Jml Transaksi App, Jml Transaksi Web, Jml Transaksi Outlet, Jml Kunjungan App.

Berdasarkan fitur yang digunakan, terdapat data leakage yaitu Penghasilan pertahun, Tanggungan Ortu, dan Tanggungan anak, karena merupakan data pribadi yang seharusnya tidak diekspos.

Evaluation

05

Evaluasi Model Prediksi

Customer Complaint Analysis and Prediction Model [Task 5]

- Random Oversampling

```
scores1 = cross_val_score(pipeline, X, y, scoring='f1_micro', cv=cv, n_jobs=-1)
score1 = mean(scores1)
scores2 = cross_val_score(pipeline, X, y, scoring='roc_auc', cv=cv, n_jobs=-1)
score2 = mean(scores2)
print('F-measure: %.3f' % score1)
print('Mean ROC AUC: %.3f' % score2)
```

```
F-measure: 0.984
Mean ROC AUC: 0.554
```

Evaluasi Model Prediksi

Customer Complaint Analysis and Prediction Model [Task 5]

- Borderline SMOTE

```
scores1 = cross_val_score(pipeline, X, y, scoring='f1_micro', cv=cv, n_jobs=-1)
score1 = mean(scores1)
scores2 = cross_val_score(pipeline, X, y, scoring='roc_auc', cv=cv, n_jobs=-1)
score2 = mean(scores2)
print('F-measure: %.3f' % score1)
print('Mean ROC AUC: %.3f' % score2)
```

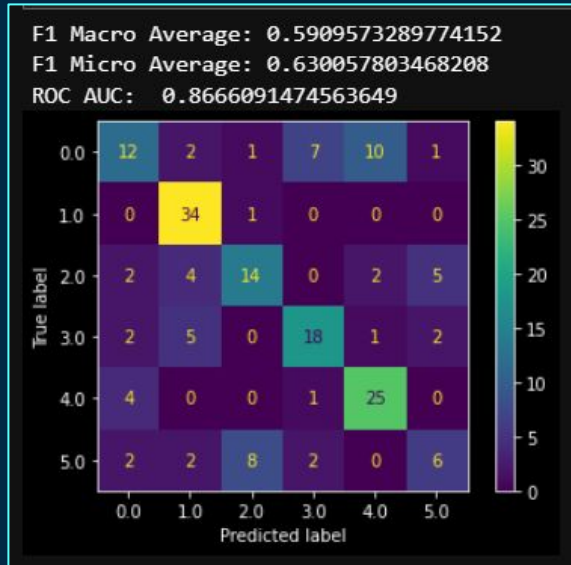
```
F-measure: 0.980
Mean ROC AUC: 0.560
```

Evaluasi Model Prediksi

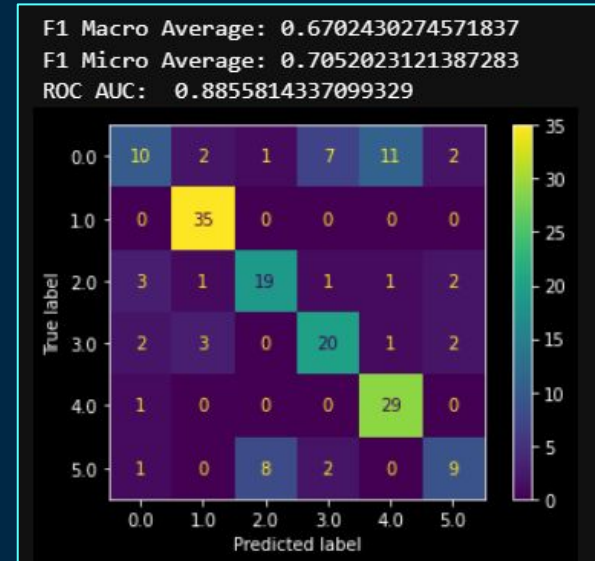
Customer Offer Acceptance Prediction Model [Task 7]

(Agrawal et al, 2015)

KNN (tanpa hyperparameter tuning)



KNN (dengan Grid CV untuk hyperparameter tuning)



Evaluasi Model Prediksi

Sembako Nominal Transaction Prediction Model [Task 8]

- Linear Regression memiliki nilai $MSE = RMSE = MAE = 0$ dan $R^2 = 1$

```
MSE: 0.00  
MAE: 0.00  
RMSE: 0.00  
R^2: 1.00
```

Berdasarkan hasil evaluasi terhadap model Linear Regression yang dibuat, akan mungkin terjadi overfitting, oleh karena itu dilakukan regularisasi dengan Ridge Regression dan Lasso untuk mencegah overfit tersebut.

Evaluasi Model Prediksi

Sembako Nominal Transaction Prediction Model [Task 8]

- Lasso Regression :

```
a : 1e-06 R2 : 0.9999999818928075 MSE : 1292.028767568812 RMSE : 35.944801676581996 MAE : 28.034515275637474
a : 1e-05 R2 : 0.9999999818928573 MSE : 1292.0244013998129 RMSE : 35.9447409421714 MAE : 28.034465953043156
a : 0.0001 R2 : 0.9999999818933565 MSE : 1291.9807405702509 RMSE : 35.944133604390174 MAE : 28.03397272730821
a : 0.001 R2 : 0.9999999818983749 MSE : 1291.5422711904907 RMSE : 35.938033769121134 MAE : 28.029020090590173
a : 0.01 R2 : 0.9999999819488035 MSE : 1287.14277704229 RMSE : 35.87677211013123 MAE : 27.97925042398318
a : 0.1 R2 : 0.9999999824518895 MSE : 1243.013710352937 RMSE : 35.25639956593607 MAE : 27.47244754991076
a : 0.5 R2 : 0.9999999844959392 MSE : 1066.1026292620663 RMSE : 32.65122707130723 MAE : 25.294177190120998
a : 1 R2 : 0.9999999868130351 MSE : 879.8030619926221 RMSE : 29.66147437321048 MAE : 22.898749256643793
a : 5 R2 : 0.9999999915249227 MSE : 587.991288618891 RMSE : 24.248531679647964 MAE : 18.52763411327579
a : 10 R2 : 0.9999999936164896 MSE : 510.54546293573515 RMSE : 22.595253106255196 MAE : 17.362448367184456
a : 20 R2 : 0.9999999874397304 MSE : 1058.5817321792313 RMSE : 32.53585302676466 MAE : 27.04330382324071
```

Evaluasi Model Prediksi

Sembako Nominal Transaction Prediction Model [Task 8]

- Ridge Regression :

```
a : 1e-06 R2 : 1.0 MSE : 6.365143614764122e-08 RMSE : 0.0002522923624441319 MAE : 0.00015493021311351296
a : 1e-05 R2 : 0.9999999999999999 MSE : 6.3651437146856285e-06 RMSE : 0.0025229236442440403 MAE : 0.0015493019781195362
a : 0.0001 R2 : 0.9999999999999923 MSE : 0.0006365141867732892 RMSE : 0.025229232782098016 MAE : 0.01549301739767984
a : 0.001 R2 : 0.9999999999992385 MSE : 0.06365121179724245 RMSE : 0.25229191781989857 MAE : 0.15492993638124034
a : 0.01 R2 : 0.9999999999238515 MSE : 6.36491397818324 RMSE : 2.522878114016458 MAE : 1.549275584669632
a : 0.1 R2 : 0.9999999923875835 MSE : 636.284248424296 RMSE : 25.224675387887473 MAE : 15.490378285504553
a : 0.5 R2 : 0.99999809959847 MSE : 15884.12205952054 RMSE : 126.03222627376118 MAE : 77.3991040866336
a : 1 R2 : 0.9999992411872359 MSE : 63421.864729081164 RMSE : 251.83698046371418 MAE : 154.66645684594872
a : 5 R2 : 0.9999812958404561 MSE : 1562915.3776221653 RMSE : 1250.1661400078651 MAE : 768.1115795004996
a : 10 R2 : 0.999926480885732 MSE : 6141368.294086609 RMSE : 2478.178422569006 MAE : 1523.3893645618862
a : 20 R2 : 0.9997158810803186 MSE : 23719479.686711162 RMSE : 4870.2648476968025 MAE : 2997.0549399180177
```

Pada model Ridge Regression, nilai MAE, MSE, RMSE lebih kecil dibandingkan pada model Lasso Regression. Dan nilai R^2 di model Ridge lebih besar dibandingkan model Lasso. Oleh karena itu, model yang lebih baik untuk dataset ini adalah model yang dibuat dengan Ridge Regression.

Rangkuman Usulan Strategi Marketing

- Pemberian promo/tawaran bergantung dengan cluster customer (sesuai dengan behavioral model atau rfm model)
- Pemberian tawaran yang menarik pada tawaran pertama yang diberikan ke seorang customer
- Peningkatan promosi di luar pulau jawa
- Peningkatan promosi untuk produk selain Sembako dan Hobi
- Meningkatkan pelayanan di outlet
- Menganalisis apakah terdapat masalah pada fitur di web dan app
- Peningkatan promosi untuk sektor pekerjaan selain karyawan
- Peningkatan variasi produk agar tidak hanya kebutuhan pokok untuk rumah tangga saja

Daftar Pustaka

Anitha, P., & Patil, M. M. (2019). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University - Computer and Information Sciences.
<https://doi.org/10.1016/j.jksuci.2019.12.011>

Hota, S., & Pathak, S. (2018). KNN classifier based approach for multi-class sentiment analysis of twitter data. International Journal of Engineering & Technology, 7(3), 1372.
<https://doi.org/10.14419/ijet.v7i3.12656>

Agrawal, A., Viktor, H.L., & Pakuet, E. (2015). 7th International Conference on Knowledge Discovery and Information Retrieval, Volume 1: KDIR, pages 226-234 ISBN: 978-989-758-158-8
<https://www.scitepress.org/papers/2015/55955/55955.pdf>

The background is a dark blue gradient. It is decorated with several vertical white lines of varying lengths. Scattered throughout are small squares in teal, pink, and orange. Some squares are solid, while others are outlined.

THANK YOU

CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), and infographics & images by [Freepik](#)
Please keep this slide for attribution