

Membership Classification Recommendation and Analysis from Customer Transaction History and Demographic



22-03-019-3

Stefani Angelika

Steven Alexander Liong

Pemilihan Variabel

✓ *"mean_purchase_price" : Rerata besar pengeluaran pada pembelian produk oleh customer setiap tahunnya.*

Dari analisis yang telah dilakukan, variabel yang digunakan sebagai fitur untuk mengklasifikasi kelas membership (Gold, Silver, dan Bronze) dari semua pelanggan Otago Prime yang melakukan transaksi di tahun 2018-2020 adalah "mean_purchase_price", yaitu rerata besar pengeluaran pada pembelian produk oleh customer setiap tahunnya.



Metode Analisis

1

Preprocessing Data

2

Feature Engineering

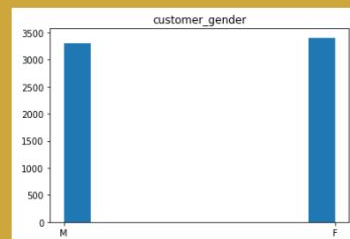
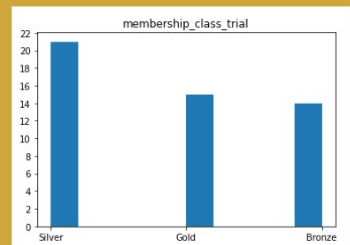
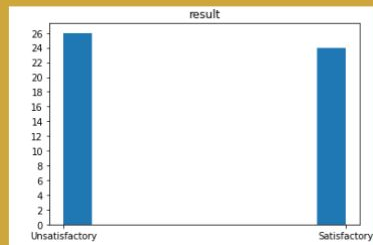
3

Modeling with
Support Vector
Machine

Preprocessing Data

◆ Exploratory Data Analysis

Exploratory Data Analysis (EDA) dilakukan dengan tujuan untuk mendapatkan suatu rangkuman yang mencakup karakteristik dari suatu dataset menggunakan bantuan statistika deskriptif serta visualisasi data.



Preprocessing Data

◆ Data Cleaning dan Data Manipulation

Untuk mendapatkan suatu dataset yang dapat diolah dengan baik, maka perlu dilakukannya data cleansing dan manipulasi data.

Langkah-langkah data cleansing :

1. penghilangan data duplikat
2. data pencilan
3. serta pengisian data yang kosong (NaN values).

Dalam data manipulation, dilakukan beberapa hal seperti :

1. pengubahan bentuk data
2. aggregasi
3. penggabungan data
4. pengelompokkan data -> data transaksi 2018, 2019, dan 2020

transaction_date_unix	transaction_date
1517715108000	2018-02-04 16:31:48+13:00
1540010091000	2018-10-20 17:34:51+13:00
1537926734000	2018-09-26 13:52:14+12:00

	transaction_id	transaction_date	transaction_region	purchase
customer_id				
257991748	17754	2018-05-27 08:39:51+12:00	Auckland	[{"id": 6, "qty": 1}]
257991748	78056	2018-10-16 06:32:23+13:00	Auckland	[{"id": 13, "qty": 4}]
257991748	18063	2018-10-06 23:15:59+13:00	Auckland	[{"id": 9, "qty": 1}]

	total_purchase_18	total_region_18	transaction_occurance_18
customer_id			
257991748	[{"id": 6, "qty": 1}, {"id": 13, "qty": 4}, {"id": 9, "qty": 1}]	[Auckland, Auckland, Auckland, Auckland]	[2018-05-27 08:39:51+12:00, 2018-10-16 06:32:23+13:00, 2018-10-06 23:15:59+13:00]
272805864	[{"id": 17, "qty": 4}, {"id": 16, "qty": 4}]	[Auckland, Auckland, Auckland, Auckland]	[2018-09-23 10:55:19+12:00, 2018-07-05 13:12:40+12:00]
452593661	[{"id": 1, "qty": 1}, {"id": 6, "qty": 1}]	[Auckland, Auckland, Auckland, Auckland]	[2018-01-29 17:21:04+13:00, 2018-11-24 22:03:20+13:00]

Data Kontinu

Untuk menentukan kolom dengan data kontinu mana saja yang berpengaruh secara signifikan dalam prediksi model, dilakukan beberapa langkah, yaitu:

1. Kruskal-Wallis Test -> Cek apakah suatu populasi data berpengaruh secara signifikan dengan populasi data lainnya
2. Logical Regression Test -> Tinjau Accuracy, Sensitivity, dan Specificity
3. Correlation Test -> Remove redundant feature

KRUSKAL-WALLIS TEST

```
Statistics=29.973, p=0.000
total_transaction_18 : Different distributions (reject H0)
Statistics=35.770, p=0.000
purchase_price_18 : Different distributions (reject H0)
Statistics=32.682, p=0.000
total_quantity_18 : Different distributions (reject H0)
Statistics=35.770, p=0.000
mean_pur_price_18 : Different distributions (reject H0)
Statistics=30.057, p=0.000
spending_habit_day_18 : Different distributions (reject H0)
Statistics=35.770, p=0.000
user_duration_day_18 : Different distributions (reject H0)
Statistics=31.071, p=0.000
total_transaction_19 : Different distributions (reject H0)
Statistics=34.271, p=0.000
purchase_price_19 : Different distributions (reject H0)
Statistics=34.273, p=0.000
total_quantity_19 : Different distributions (reject H0)
Statistics=34.271, p=0.000
mean_pur_price_19 : Different distributions (reject H0)
Statistics=26.480, p=0.000
```

LOGISTIC REGRESSION TEST

	key_0	accuracy	sensitivity	specificity	mean_score
9	mean_pur_price_19	1.000000	0.555556	1.000000	0.851852
8	total_quantity_19	1.000000	0.555556	1.000000	0.851852
7	purchase_price_19	1.000000	0.555556	1.000000	0.851852
15	mean_pur_price_20	1.000000	0.511111	1.000000	0.837037
6	total_transaction_19	0.904762	0.483333	1.000000	0.796032
0	total_transaction_18	0.833333	0.433333	1.000000	0.755556
13	purchase_price_20	0.916667	0.447619	0.885714	0.750000
3	mean_pur_price_18	0.833333	0.388889	0.877778	0.700000
1	purchase_price_18	0.833333	0.388889	0.877778	0.700000
2	total_quantity_18	0.833333	0.388889	0.877778	0.700000
10	spending_habit_day_19	0.714286	0.361111	1.000000	0.691799
4	spending_habit_day_18	0.750000	0.366667	0.933333	0.683333
14	total_quantity_20	0.833333	0.400000	0.783333	0.672222
17	user_duration_day_20	0.583333	0.200000	0.638095	0.473810
5	user_duration_day_18	0.583333	0.200000	0.583333	0.455556
16	spending_habit_day_20	0.500000	0.133333	0.666667	0.433333
12	total_transaction_20	0.416667	0.066667	0.600000	0.361111
11	user_duration_day_19	0.333333	0.000000	0.583333	0.305556

CORRELATION TEST



Data Kategorikal

Dilakukan label encoding terlebih dahulu untuk menjadikan data ini ke dalam bentuk numerik, agar data categorical ini dapat dihitung dan dipertimbangkan dalam model.

Untuk memilih data categorical yang signifikan dalam pembentukan model prediksi target, dilakukan metode:

Chi Square Contingency Test (Independency Test) pada setiap kolomnya.

LABEL ENCODING

most_category_18	most_brand_18
Eggs	Auckland's Best
Breads	Auckland's Best
Eggs	Auckland's Best
Dairy Products	Auckland's Best
Eggs	Auckland's Best



most_category_18	most_brand_18
1	3
3	3
1	3
2	3
1	3

CHI SQUARE CONTINGENCY TEST

	dependent	p_value
most_category_18	no	0.992648
most_brand_18	no	1
most_region_18	no	0.973972
most_category_19	no	0.881179
most_brand_19	no	1
most_region_19	no	0.972551
most_category_20	no	0.983256
most_brand_20	no	1
most_region_20	no	0.973972

Support Vector Machine

PEMILIHAN VARIABEL UNTUK SVM

	accuracy	sensitivity	specificity	acc
[mean_purchase_price]	1	0.503922	1	1.00
[mean_total_qty]	0.972222	0.495425	1	0.96
[mean_mean_pur_price]	1	0.503922	1	1.00
[mean_total_transaction]	0.944444	0.46348	0.959559	0.92
[mean_purchase_price, mean_total_qty]	1	0.503922	1	1.00
[mean_purchase_price, mean_mean_pur_price]	1	0.503922	1	1.00
[mean_purchase_price, mean_total_transaction]	1	0.503922	1	1.00
[mean_total_qty, mean_mean_pur_price]	1	0.503922	1	1.00
[mean_total_qty, mean_total_transaction]	0.972222	0.495425	1	0.96
[mean_mean_pur_price, mean_total_transaction]	1	0.503922	1	1.00
[mean_purchase_price, mean_total_qty, mean_mean_pur_price]	1	0.503922	1	1.00
[mean_purchase_price, mean_total_qty, mean_total_transaction]	1	0.503922	1	1.00
[mean_purchase_price, mean_mean_pur_price, mean_total_transaction]	1	0.503922	1	1.00
[mean_total_qty, mean_mean_pur_price, mean_total_transaction]	1	0.503922	1	1.00
[mean_purchase_price, mean_total_qty, mean_mean_pur_price, mean_total_transaction]	1	0.503922	1	1.00

Metode Support Vector Machine (SVM) dipilih karena metode ini cocok untuk jumlah data yang cenderung kecil dalam melakukan klasifikasi suatu data. Terlebih, classifier ini dapat digunakan untuk multiclass classification yang cocok untuk membagi kelas membership menjadi 3, yaitu gold, silver, dan bronze.

I. Hasil apply model ke data "unsatisfactory"

customer_id	mean_purchase_price	mean_total_qty	mean_mean_pur_price	mean_total_transaction	membership_class_trial
997579682270	433.666667	37.333333	48.738889	9.333333	2
997458045673	345.000000	30.333333	41.377778	8.333333	1
996882350315	233.000000	22.333333	33.229798	7.000000	1
996783659998	255.666667	22.000000	40.014141	6.333333	1
996331197257	184.666667	16.666667	32.583333	6.000000	1

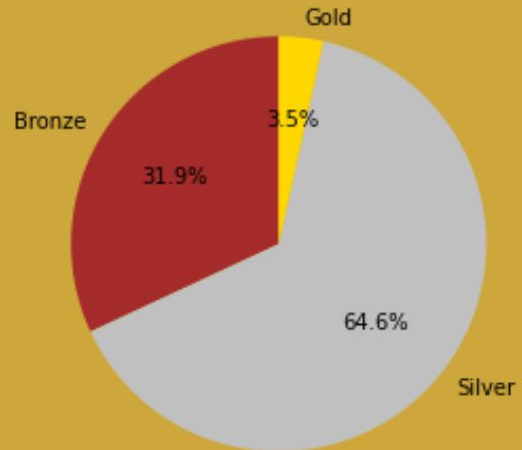
II. Hasil apply model ke customer_id lain selain data OPC

customer_id	mean_purchase_price	mean_total_qty	mean_mean_pur_price	mean_total_transaction	mean_spending_habit	membership_class_trial
257991748	280.000000	26.666667	42.555556	7.000000	54.466667	1
272805864	559.333333	48.666667	57.750000	8.666667	38.461905	2
452593661	363.666667	34.666667	37.686111	9.333333	38.055556	1
477858761	660.000000	52.333333	67.250000	9.000000	45.918519	2
571498795	322.666667	27.666667	49.358974	5.666667	7.629630	1
...
959434851573	593.883449	47.808171	67.072546	7.431747	50.889424	2
977390060838	538.550115	44.141505	48.331276	9.431747	57.766347	2
246059017174	767.763010	59.596449	55.262549	10.046494	39.761839	2
573291074774	707.096344	52.263116	58.329216	8.046494	46.172951	2
811709979790	788.096344	61.596449	57.111034	10.046494	36.284062	2

8801 rows x 6 columns

Membership Class Recommendation

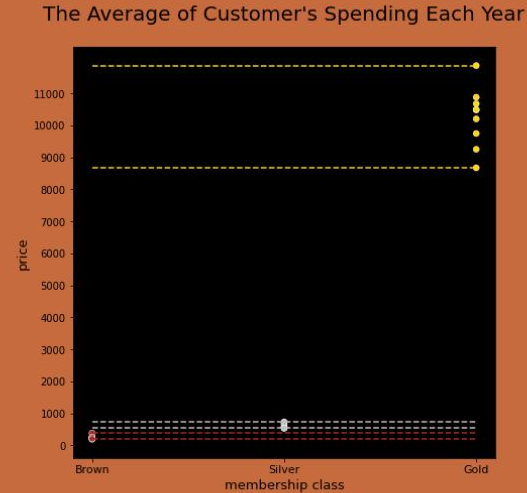
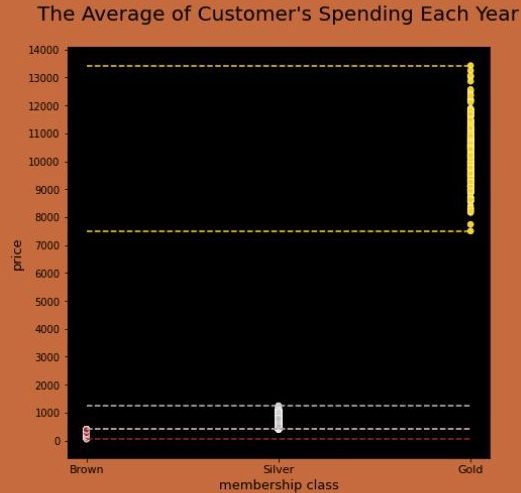
Membership Class



Berdasarkan hasil klasifikasi dari model Support Vector Machine (SVM) yang telah dibentuk menggunakan fitur “mean_purchase_price”, yaitu rerata besar pengeluaran pada pembelian produk oleh customer tiap tahunnya, didapatkan proporsi keseluruhan membership class dari 8851 customer_id (termasuk customer_id pada dataset OPC)

Kelas membership dengan proporsi terbesar adalah kelas Silver dengan jumlah 64.6% dari keseluruhan data, lalu kelas Bronze dengan 31.9%, dan terakhir kelas Gold dengan 3.5%.

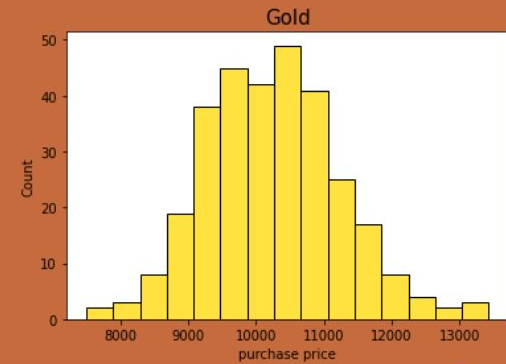
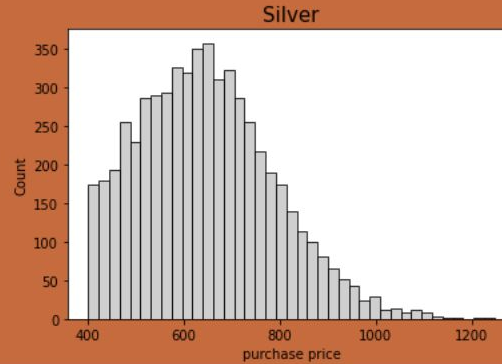
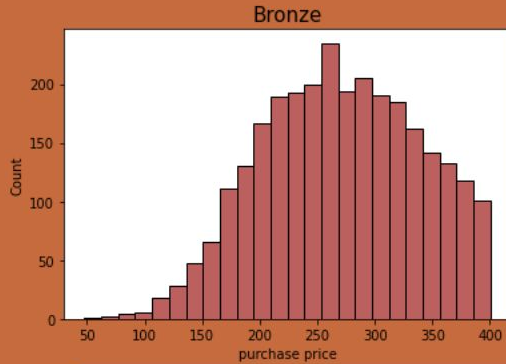
Membership Class Recommendation



Dari gambar di kiri, model yang dibentuk mengklasifikasikan kelas membership dari customer berdasarkan rata besarnya pengeluaran pada pembelian produk oleh customer tiap tahunnya. Model ini menyengmentasikan customer menjadi tepat 3 kelas, sama seperti yang terlihat pada training data, yaitu pada gambar di kanan.

Berdasarkan kedua gambar ini, dapat dilihat bahwa model mencerminkan hal yang sama seperti pada training data dan untuk lebih jelasnya dapat dilihat dengan lebih detil persebaran rata besar pengeluaran pada pembelian produk oleh customer tiap tahunnya pada setiap kelasnya

Membership Class Recommendation



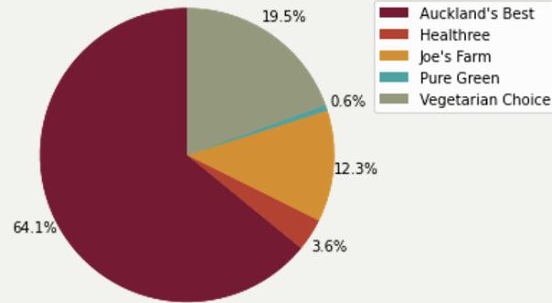
Dari ketiga gambar diatas, dapat dilihat distribusi data dari setiap kelas yang disegmentasikan oleh model, yaitu kelas Bronze, Silver, dan Gold.

Untuk kelas bronze, nilai purchase price berada pada selang 47 sampai 400 satuan harga dan data berpusat pada nilai 200 sampai 300 satuan harga. untuk kelas Silver, nilai purchase price berada pada selang 401 sampai 1248 satuan harga dan data berpusat pada nilai 500 sampai 750 satuan harga. untuk kelas Gold, nilai purchase price berada pada selang 7502 sampai 13428 satuan harga dan data berpusat pada nilai 9000 sampai 11000 satuan harga.

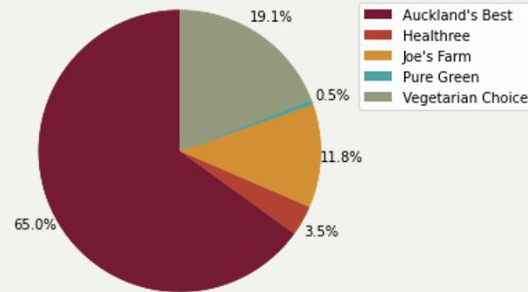
Setiap kelas tersegmentasi dengan baik berdasarkan rerata besarnya pengeluaran pada pembelian produk oleh customer tiap tahunnya. Tidak ada kelas yang saling beririsan dengan kelas lain, yang menandakan hasil prediksi dari model ini sangat baik dan sesuai dengan tujuan penelitian yang ada.

Brand Analysis

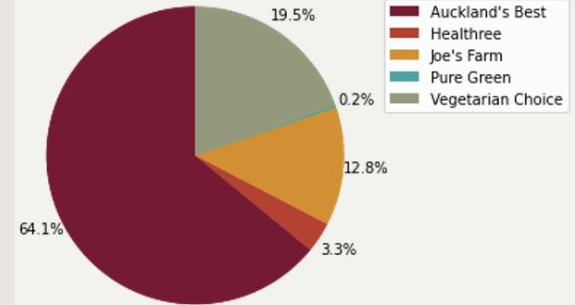
2018 Customer's Favorite Brand



2019 Customer's Favorite Brand



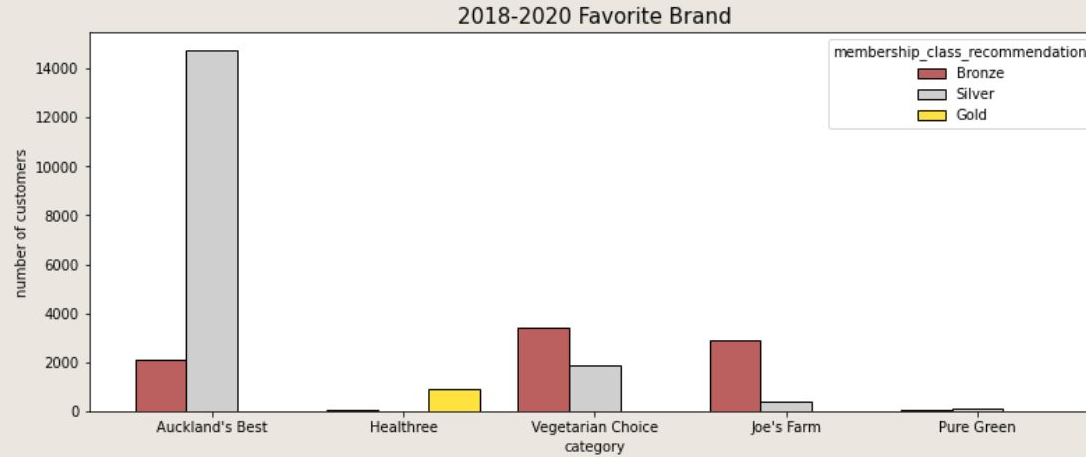
2020 Customer's Favorite Brand



Selanjutnya, analisis dilakukan pada brand produk favorit dalam transaksi tahun 2018-2020. Penentuan brand favorit didapat dengan cara menghitung brand yang produknya paling banyak yang dibeli oleh tiap customer setiap tahunnya. Dengan demikian, hasil dari gambar 3.5.1 merupakan visualisasi proporsi dari banyaknya customer dengan brand yang sering dibelinya masing-masing. Sebagai contoh, angka 64.1% dari gambar 3.5 mengartikan terdapat 64.1% customer dari jumlah customer yang memiliki “Auckland’s Best” sebagai brand unggulannya atau dengan kata lain, terdapat 64.1% customer yang memiliki brand “Auckland’s Best” sebagai brand yang produknya paling banyak dibeli pada tahun 2018.

Selanjutnya dari ketiga gambar diatas untuk tahun 2018 – 2020, proporsi brand favorit memiliki trend yang sama, yaitu “Auckland’s Best” sebagai brand paling favorit, diikuti dengan “Vegetarian Choice”, kemudian “Joe’s Farm”, lalu “Healththree”, dan yang terakhir adalah “Pure Green”.

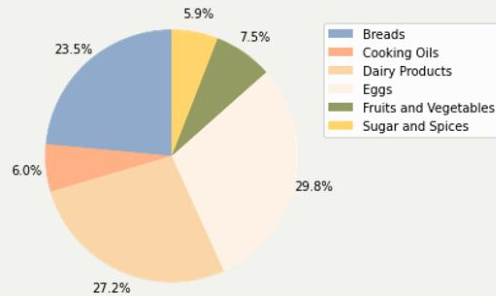
Brand Analysis



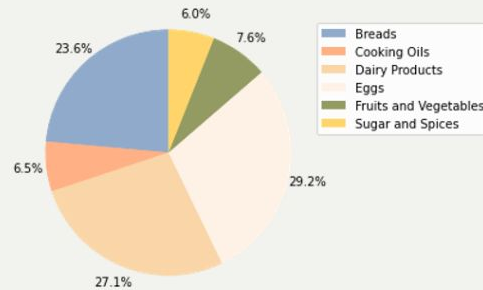
Customer dengan rekomendasi kelas Silver sebagian besar memiliki “Auckland’s Best” sebagai brand unggulannya. Hal ini merupakan hal yang baik bagi brand “Auckland’s Best” dikarenakan mayoritas customer berasal dari rekomendasi kelas Silver sehingga dengan memiliki banyak customer dengan rekomendasi kelas Silver sebagai pelanggan, maka penjualan dari brand tersebut dapat dipastikan baik untuk cakupan Otago Prime. Sedangkan brand “Pure Green” perlu memerhatikan kembali kondisi penjualan seperti strategi marketing, penjualan, dan kualitas produk untuk keluar dari posisi terendah brand favorit customer di tiap tahunnya.

Product Category Analysis

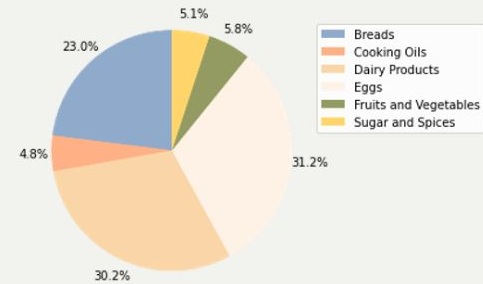
2018 Customer's Favorite Category



2019 Customer's Favorite Category



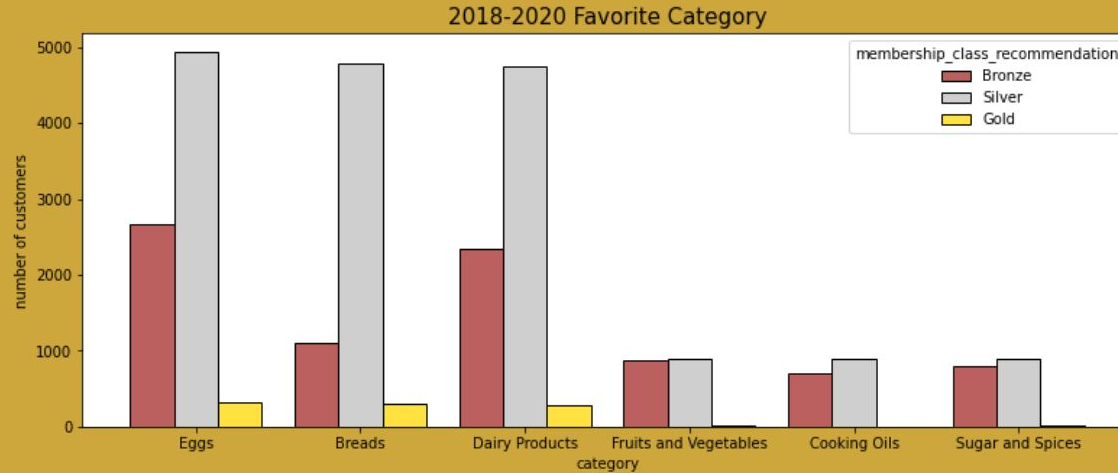
2020 Customer's Favorite Category



Analisis selanjutnya dilakukan pada kategori produk favorit dalam transaksi tahun 2018-2020. Kategori yang paling sering dibeli didapat dengan cara yang sama seperti pada brand favorit, yaitu dengan menghitung kategori yang produknya paling banyak yang dibeli oleh tiap customer setiap tahunnya. Dengan demikian, hasil dari gambar 3.7.1 merupakan visualisasi proporsi dari banyaknya customer dengan kategori yang paling banyak dibeli oleh masing-masing customer.

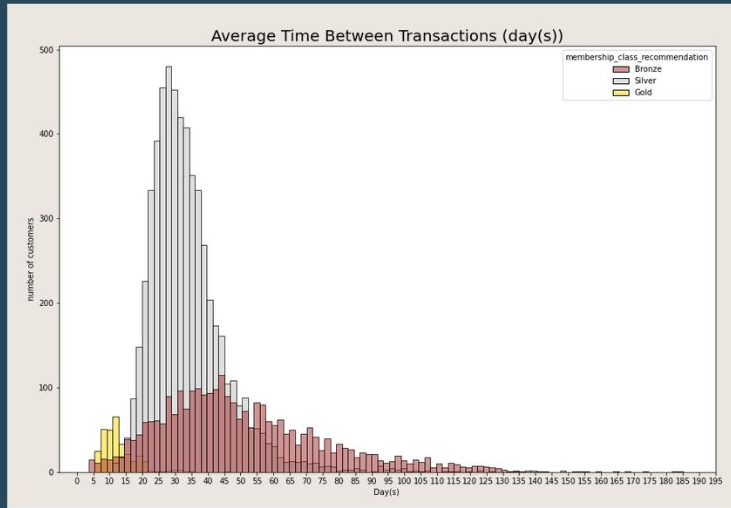
Untuk tahun 2018 – 2020, proporsi kategori paling favorit memiliki trend yang sama, yaitu “Eggs” sebagai kategori paling favorit, diikuti dengan “Dairy Products”, kemudian “Breads”, lalu “Fruits and Vegetables”, dan “Cooking Oils” dan “Sugar and Spice” yang bertukar posisi di posisi terakhir.

Product Category Analysis



customer dengan rekomendasi kelas membership Bronze, Silver, dan Gold sebagian besar memiliki “Eggs”, “Dairy Products”, dan “Breads” sebagai kategori produk unggulannya. Oleh karena itu, penulis merekomendasikan produsen setiap brand untuk menjaga ketersediaan produk yang telah disebutkan untuk bersaing dalam pasar.

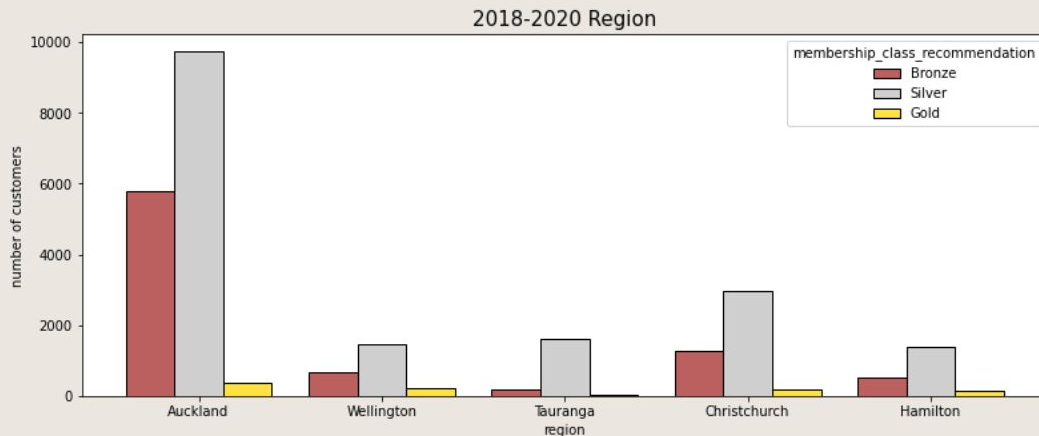
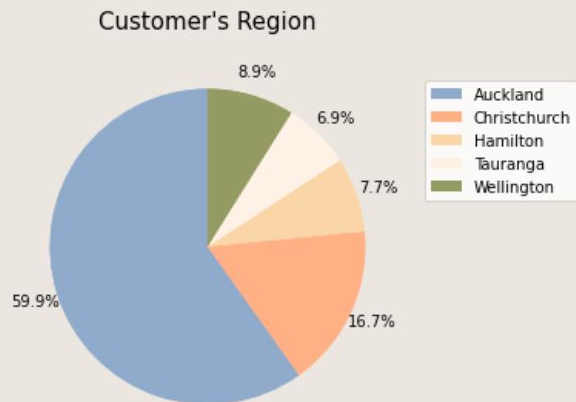
Transaction Time Analysis



Gambar diatas menggambarkan distribusi data rerata waktu antar transaksi dalam satuan hari dari setiap kelas membership. Dari gambar tersebut, customer dengan rekomendasi kelas Gold memiliki mayoritas rerata waktu antar transaksi di antara 10-15 hari. Selanjutnya, customer dengan rekomendasi kelas Silver memiliki mayoritas rerata waktu transaksi di antara 25-40 hari. Selanjutnya yang terakhir, customer dengan rekomendasi kelas Bronze memiliki mayoritas rerata waktu transaksi di antara 35-50 hari. Dari hasil ini, didapatkan bahwa customer dengan rekomendasi kelas Gold memiliki selang waktu terjadinya suatu transaksi dengan transaksi lainnya yang paling dekat atau cepat, sehingga kelas membership Gold ini memiliki frekuensi terjadinya transaksi paling banyak dibandingkan dengan kelas lainnya. Kemudian, rerata waktu antar pembelian diikuti oleh kelas Silver dan kelas Bronze dengan nilai yang paling lama.

Oleh karena itu, dapat dikatakan bahwa customer dengan rekomendasi membership kelas Gold memiliki kebiasaan membeli 10 -15 hari sekali (1 - 2 minggu sekali) dan merupakan kelas dengan frekuensi pembelian paling tinggi (paling sering melakukan transaksi). Kemudian, disusul kelas Silver yang memiliki kebiasaan membeli 25 - 40 hari sekali (3 - 5 minggu sekali) sehingga memiliki frekuensi pembelian menengah dan kelas Bronze yang memiliki kebiasaan membeli 35 - 50 hari sekali (5 - 7 minggu sekali) sehingga memiliki frekuensi pembelian paling rendah.

Transaction Region Analysis



Didapat bahwa mayoritas customer melakukan transaksi di wilayah Auckland. Selanjutnya, berdasarkan Gambar 3.11, sebagian besar customer dengan rekomendasi membership kelas Silver dan Bronze melakukan transaksi di Auckland. Sementara itu, customer dengan rekomendasi membership kelas Gold melakukan transaksi di wilayah yang tersebar cukup merata di Auckland, Wellington, Christchurch, dan Hamilton.

KESIMPULAN

Variabel Prediktor

“mean_purchase_price”

yaitu rerata besar pengeluaran pada pembelian produk oleh customer setiap tahunnya.

**8851
customer**

Rekomendasi Kelas Membership

Bronze

**2823 cust
31.9%**

Silver

**5718 cust
64.6%**

Gold

**310 cust
3.5%**

KESIMPULAN

Rekomendasi bagi Otago Prime

Lebih memerhatikan pelanggan yang berada di rekomendasi kelas Silver dan Gold dengan memberi sejumlah promo khusus atau point bagi member

Gold

Rerata besar pengeluaran pada pembelian produk oleh customer setiap tahunnya yang terbesar & jumlah frekuensi pembelian terbanyak

Silver

Jumlah customer paling banyak memiliki jumlah pembelian menengah serta memiliki andil yang cukup besar dalam meningkatkan penjualan

Menjalin hubungan yang baik dengan brand favorit, yaitu "Auckland's Best" dan "Vegetarian Choice", karena kedua brand ini berperan besar dalam menjangkau customer Silver dan Gold.

Mempertimbangkan produk kategori Eggs, dan Dairy Products dalam mengatur stok dan distribusi barang, karena kedua produk ini kategori produk yang terjual paling banyak.

Untuk pengiriman produk, Otago Prime perlu memusatkan pengiriman di region Auckland karena banyak sekali customer yang melakukan transaksi di region ini.

Terima Kasih
