DIBIMBING

BUSINESS

IS THE PENALIZER WE HAVE CREATED APPROPRIATE FOR MEASURING THE ACCURACY OF THE CUSTOMER RETENTION MODEL?

CONTENT

01 background

02 objective

03 results and Discussion

04 bibliography

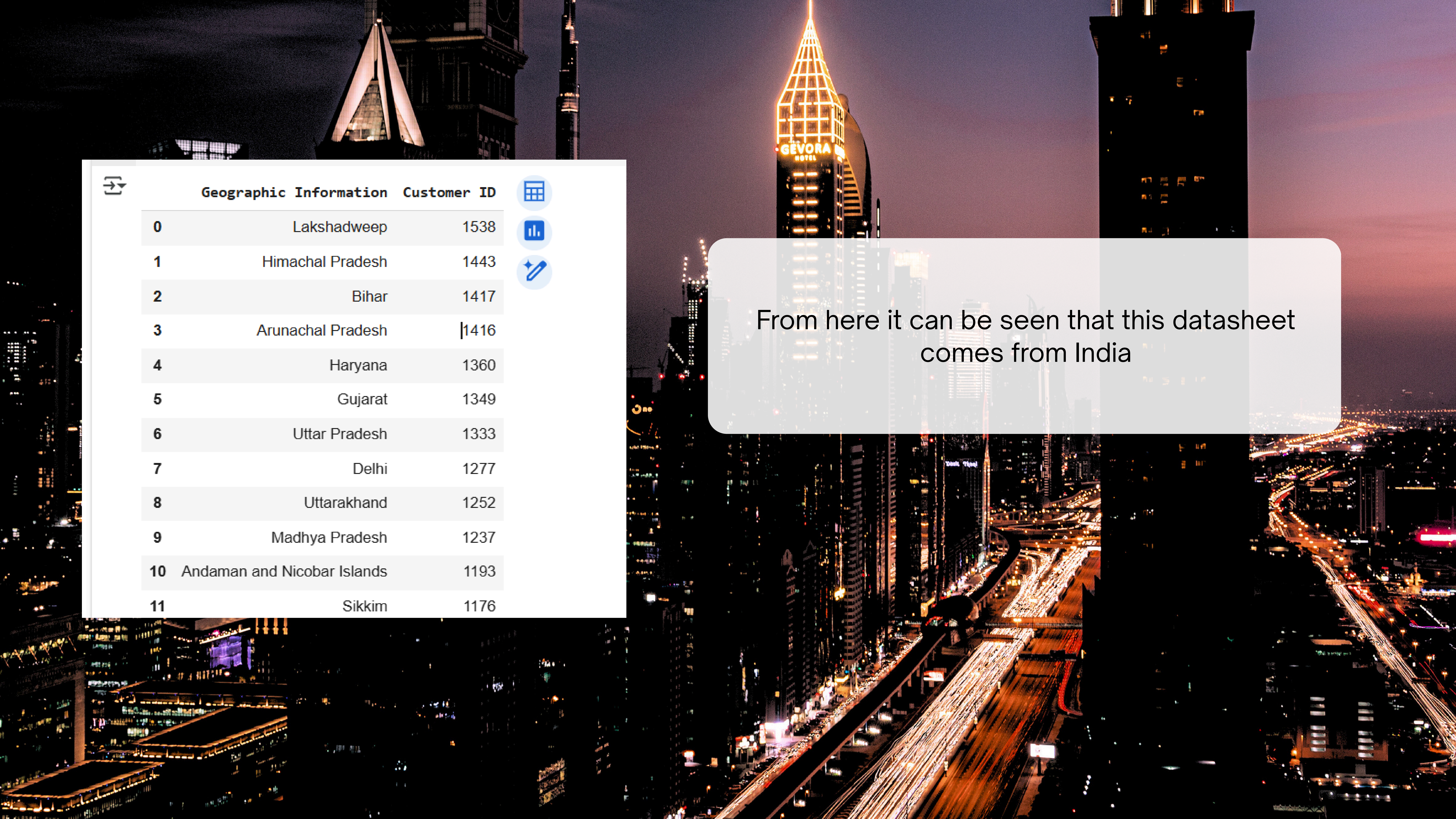
BACKGROUND

According to the Financial Services Authority (OJK), Indonesia currently has 150 insurance companies operating, one of which is the company I use.

According to Dalimunthe and Islami (2021), insurance companies must have two-way communication.

BACKGROUND

This dataset contains Customer ID, Age, Gender, Marital Status, Education Level, Geographic Information, Occupation, Income Level, Behavioral Data, Purchase History, Interactions with Customer Service, Insurance Products Owned, Coverage Amount, Premium Amount, Policy Type, Customer Preferences, Preferred Communication Channel, Preferred Contact Time, Preferred Language, and Segmentation Group.



	Geographic Information	Customer ID
0	Lakshadweep	1538
1	Himachal Pradesh	1443
2	Bihar	1417
3	Arunachal Pradesh	1416
4	Haryana	1360
5	Gujarat	1349
6	Uttar Pradesh	1333
7	Delhi	1277
8	Uttarakhand	1252
9	Madhya Pradesh	1237
10	Andaman and Nicobar Islands	1193
11	Sikkim	1176

From here it can be seen that this datasheet comes from India

OBJECTIVE

01 Grouping customers based on purchasing behavior

02 Prioritize high-value customers

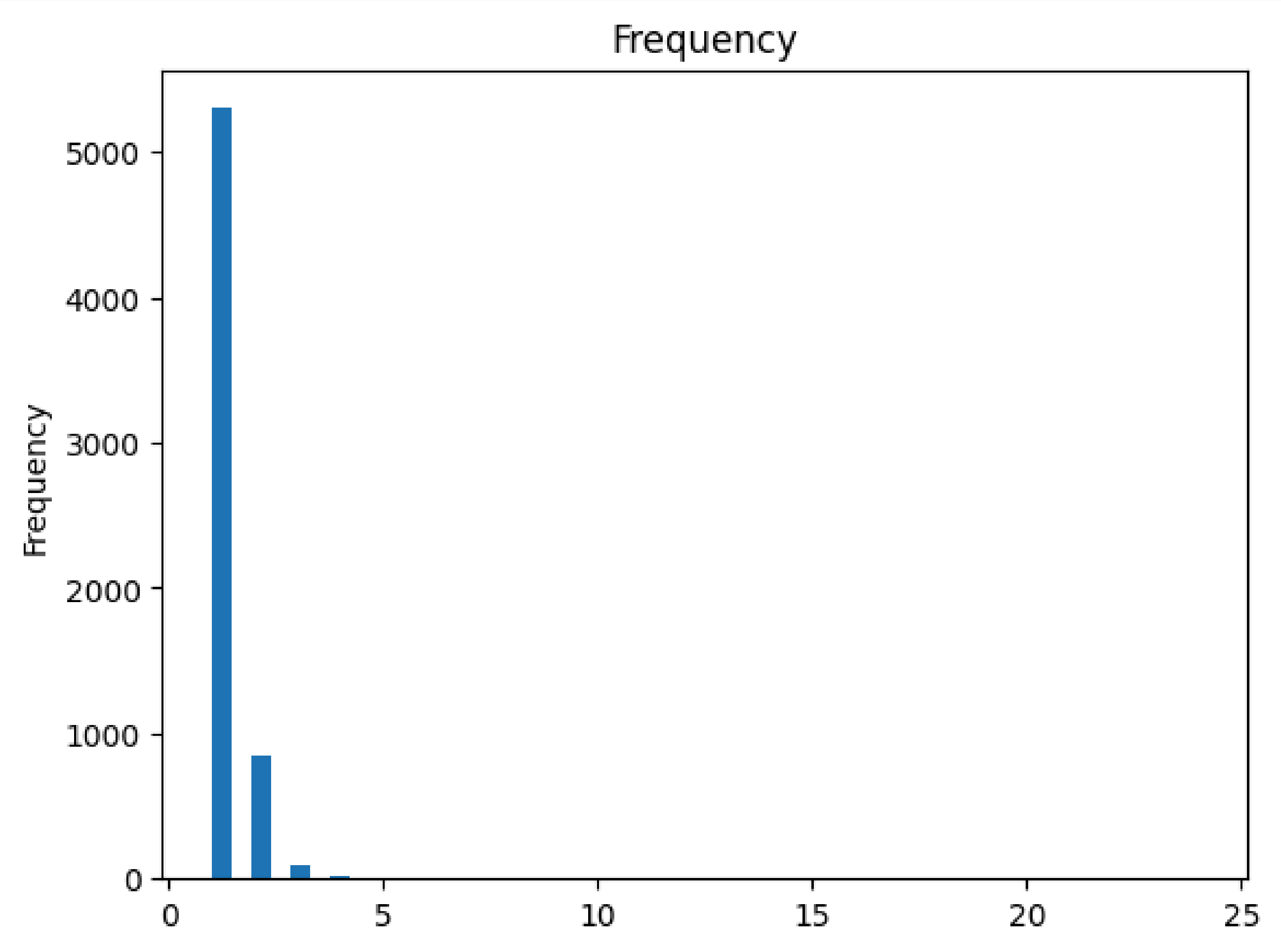
03 Designing more appropriate retention and acquisition strategies

RESULTS AND DISCUSSION

Visualizing our frequency/recency matrix

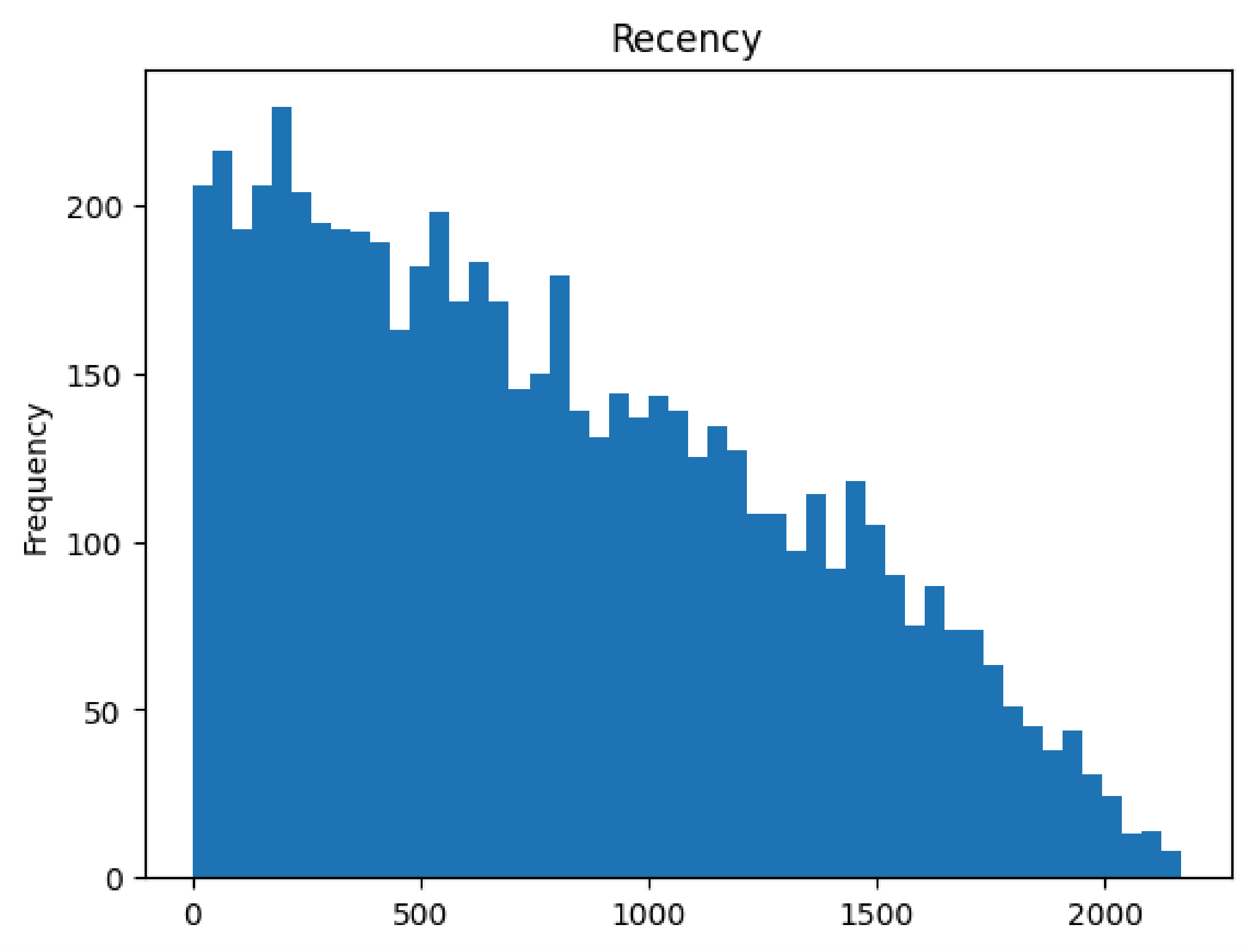
Customer transactions predictions

The Gamma-Gamma model and the independence assumption



If we look at the picture on the side, the average customer only shops once.

and if we look at the recency, most customers haven't shopped for too long, although there are also some who haven't shopped for a very long time.



```
[24] churn_rate = 1 - repeat_rate
     print("Churn Rate:", churn_rate)
```

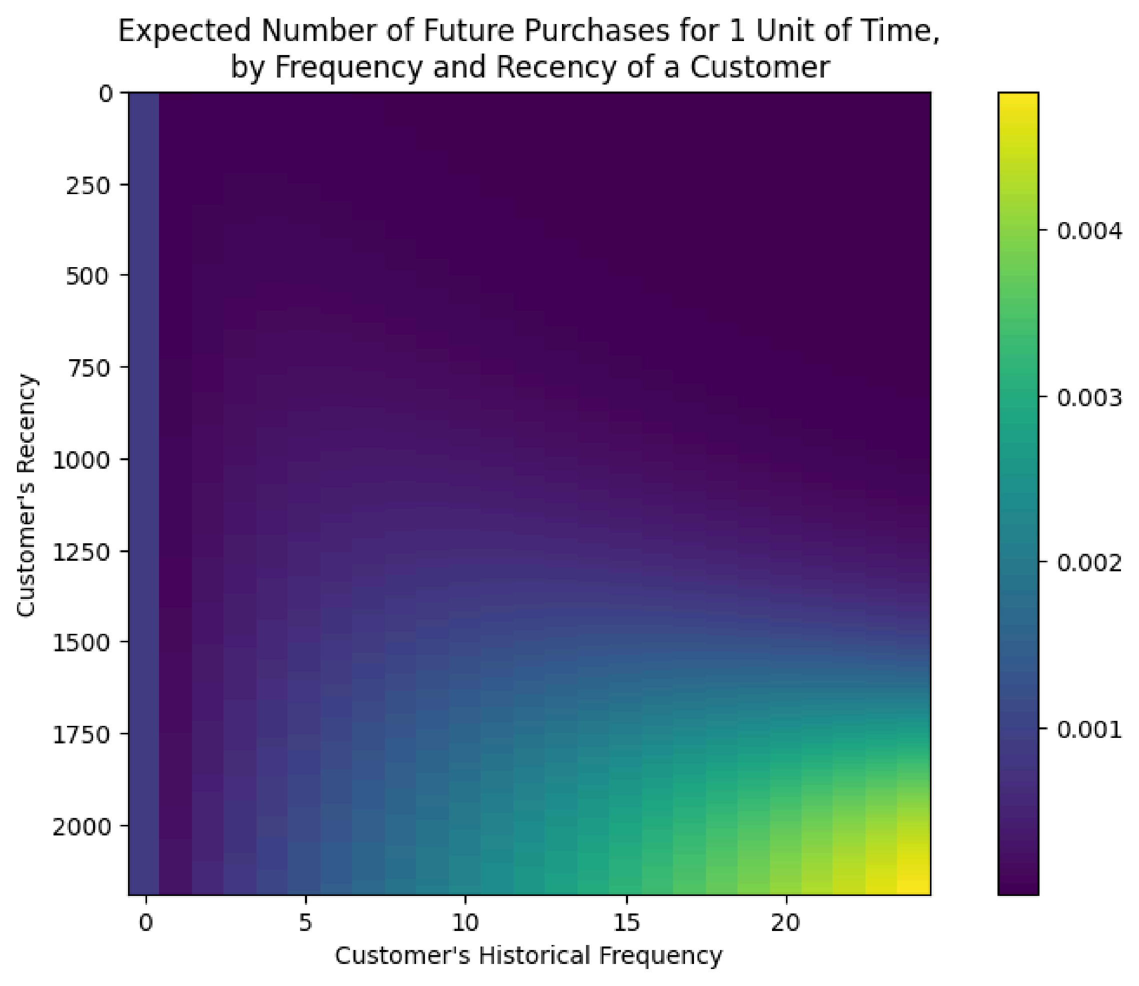
```
Churn Rate: 0.8004524310202001
```

high churn rate

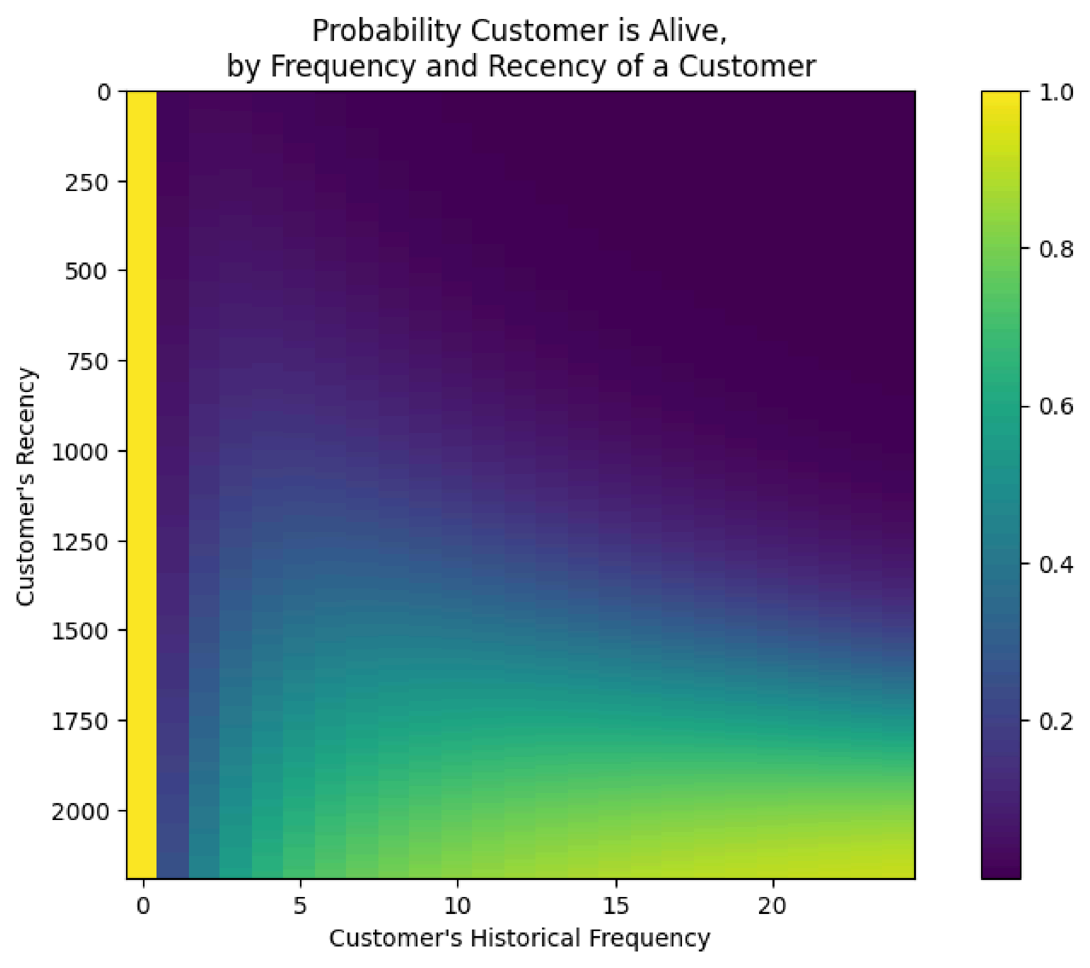
```
cltv_ready[['frequency', 'recency', 'T']].describe()
```

	frequency	recency	T
count	6257.000000	6257.000000	6257.000000
mean	1.177401	786.712003	1485.016941
std	0.531171	532.379488	513.273060
min	1.000000	1.000000	7.000000
25%	1.000000	330.000000	1143.000000
50%	1.000000	706.000000	1573.000000
75%	1.000000	1183.000000	1913.000000
max	24.000000	2167.000000	2188.000000

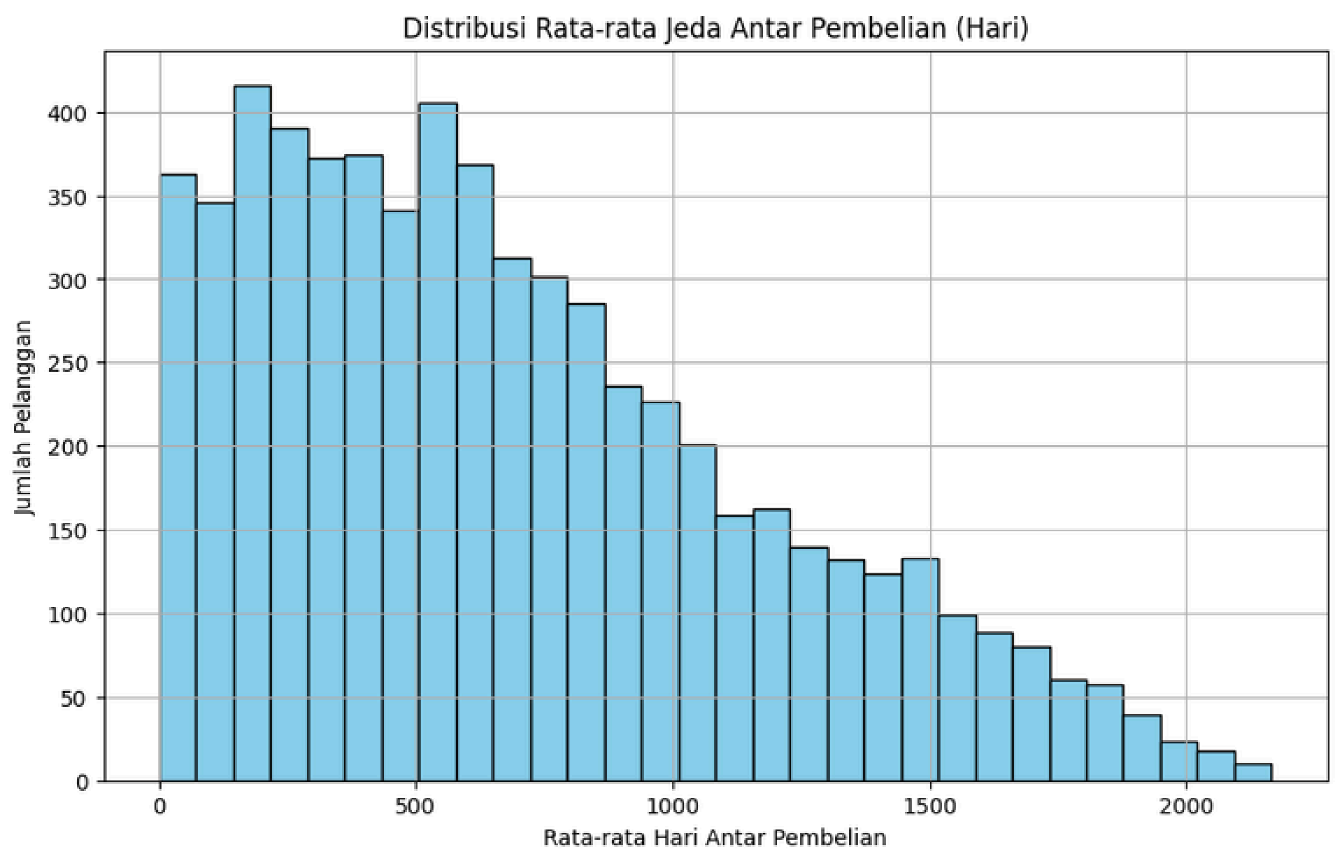
rencency below the T value



Customers with a high frequency of 10–24 purchases and high recency are those who have recently made purchases. These customers can be interpreted as still active and are very likely to make another purchase.



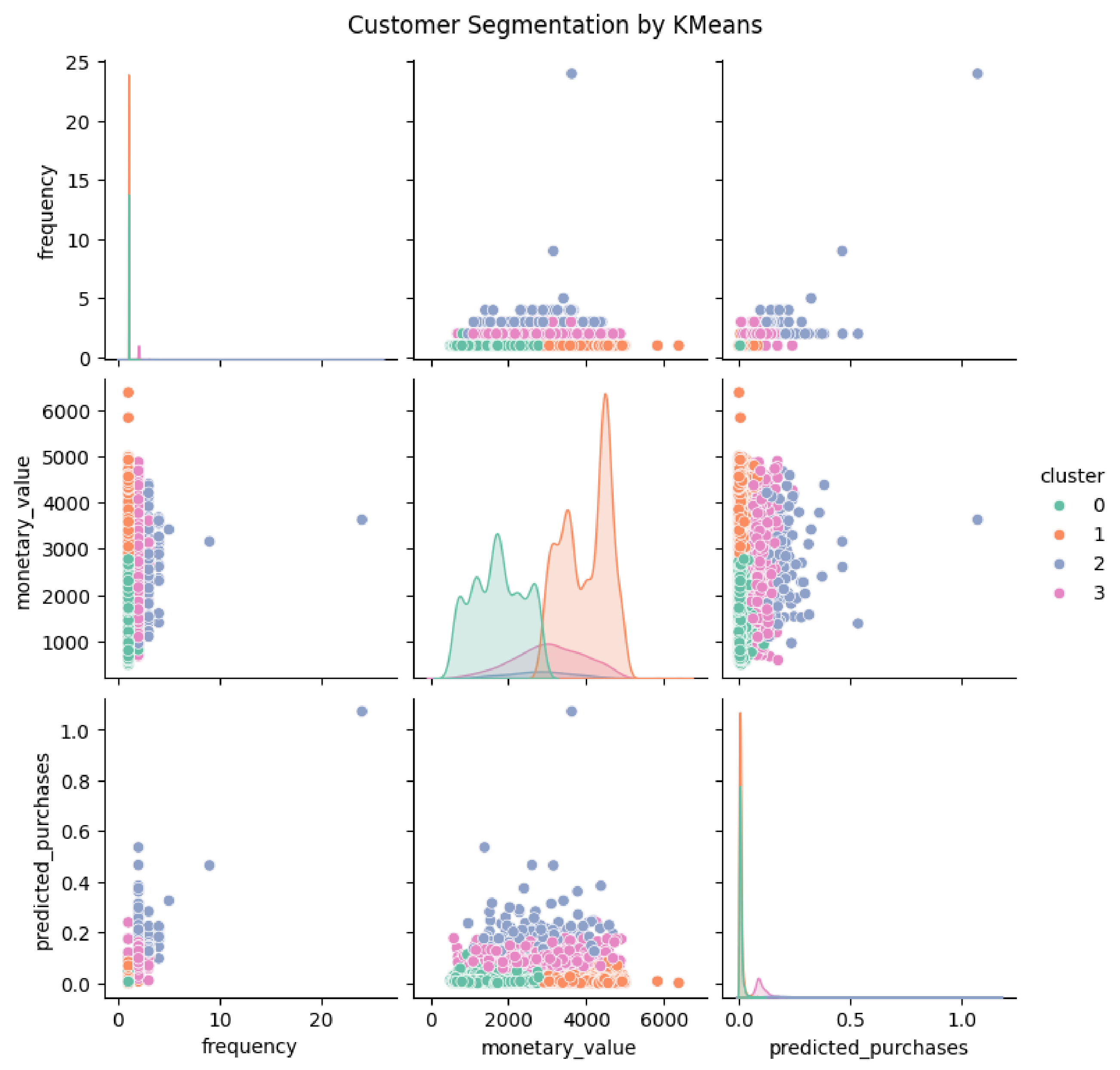
Most of the customers are in the dark purple area on the top left which means low frequency and high recency, meaning customers rarely buy and have not bought for a long time, but there are still some customers in the bottom right with the yellow-green color, these are loyal customers.

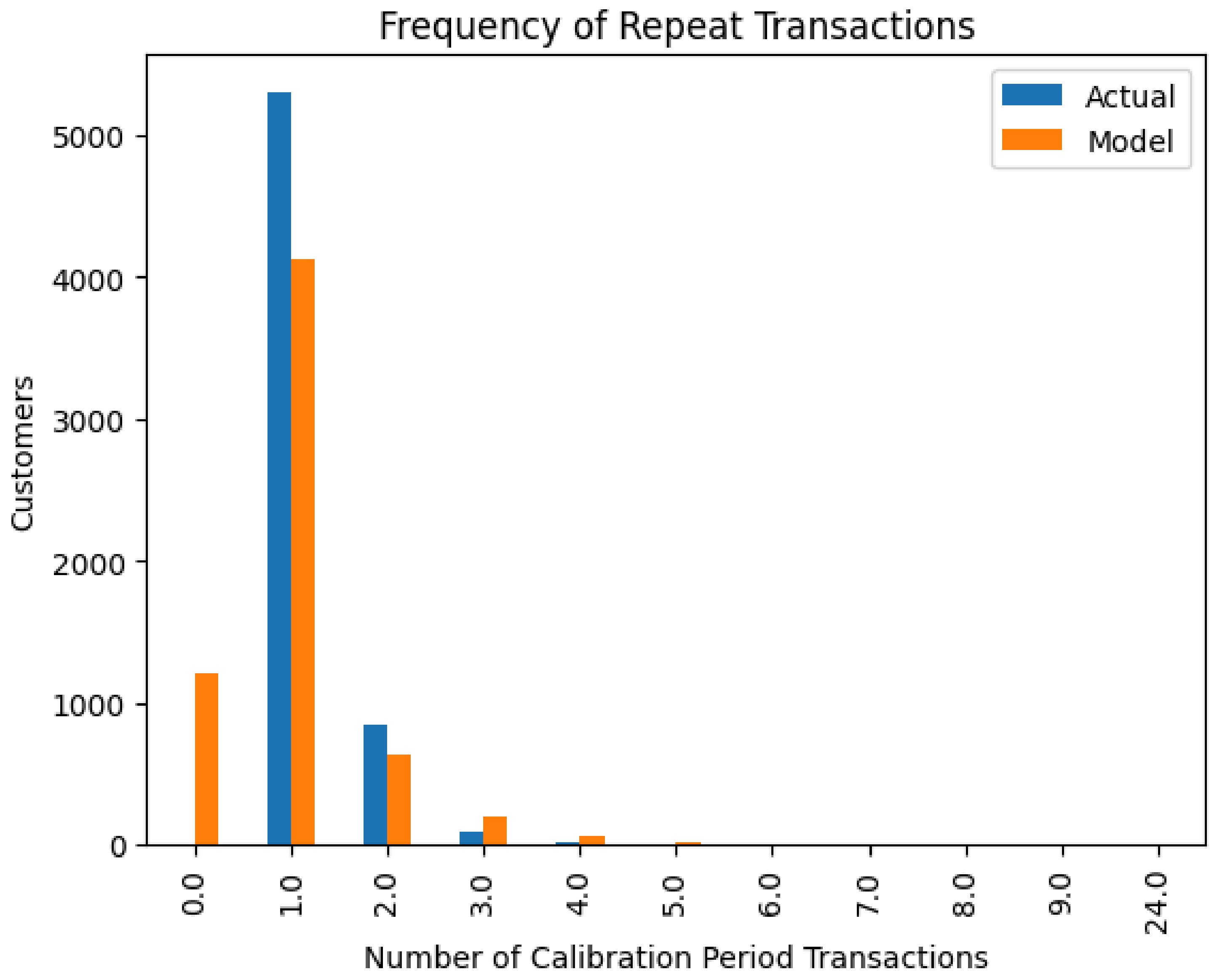


0-600 customers at most after that it drops significantly

using K-Means for clustering

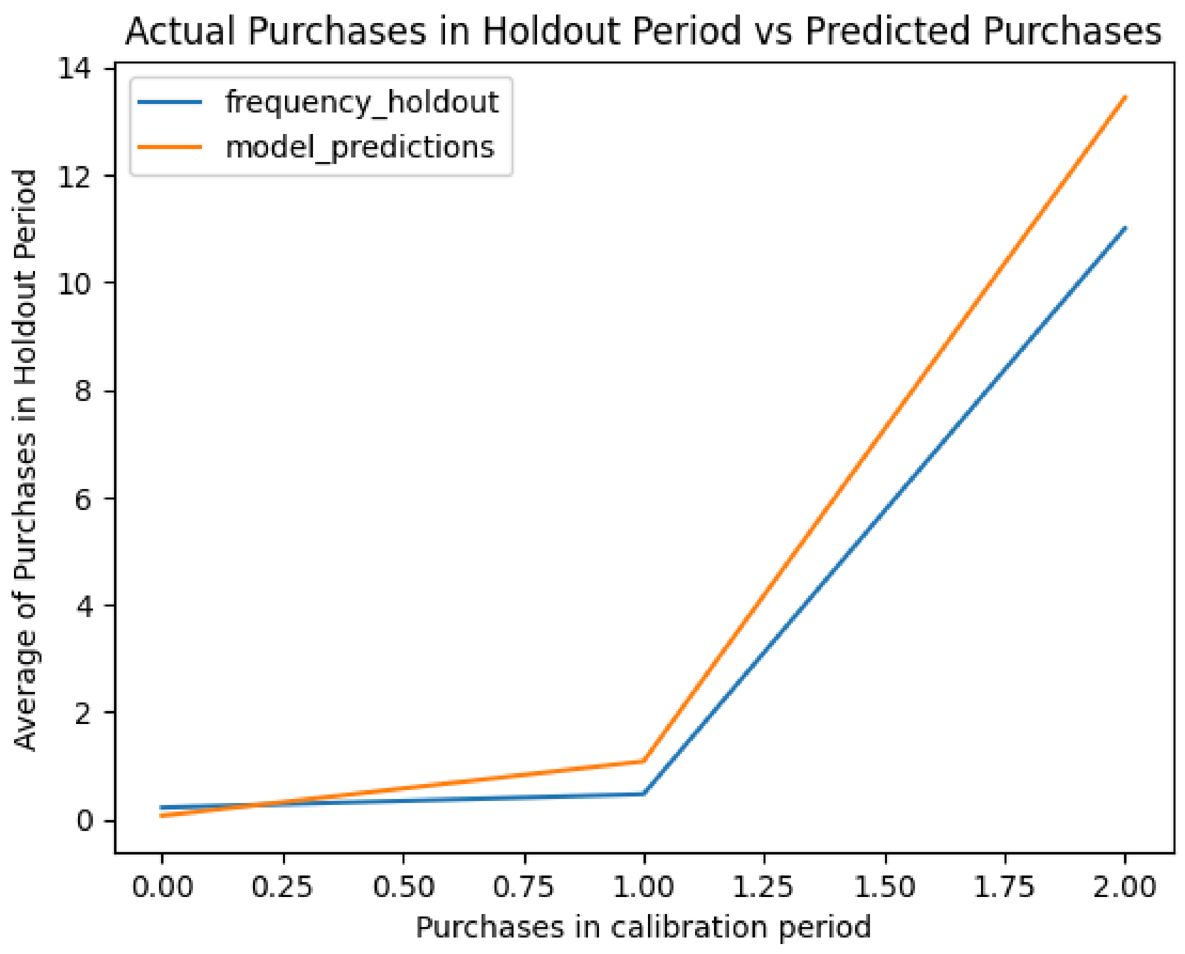
Cluster	Ciri-ciri Utama
0	Loyal, sering belanja, nilai tinggi
1	Sekali belanja besar, jarang kembali
2	Stabil, pembeli reguler
3	Tidak aktif, nilai kecil





Based on the figure, the model's predictions are not far off from the actual data, especially for low frequencies (1–2 transactions).

This indicates that the model is quite capable of representing the data, with some underpredictions occurring when transactions exceed three.



The model's trends are similar to the actual data, although the predictions are slightly higher.

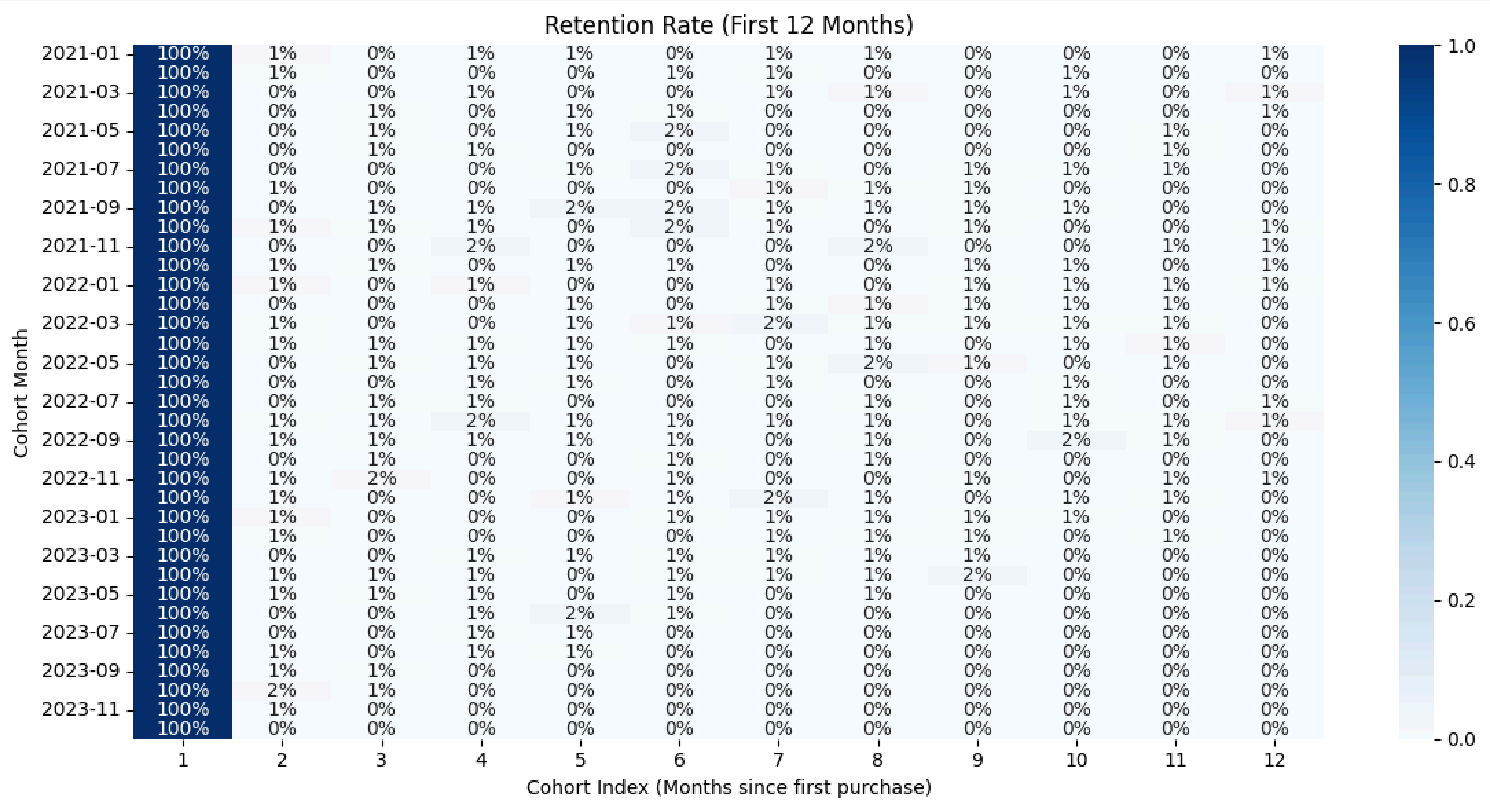
This means the model is quite good at capturing purchasing behavior patterns.

customer_id	frequency	recency	T	monetary_value	prob_alive	predicted_purchases
87858	2.0	108.0	324.0	2400.500000	0.358610	0.375091
35063	2.0	132.0	335.0	4376.500000	0.364952	0.384747
100000	9.0	1734.0	1843.0	3154.444444	0.766319	0.465205
37975	2.0	192.0	308.0	2605.500000	0.405410	0.466731
68541	2.0	190.0	268.0	1385.500000	0.423114	0.536398
1	24.0	2134.0	2165.0	3626.291667	0.910900	1.071894

If we look at customer probability histories, these are the customers predicted to make the most purchases in the next 100 times.

customer_id	frequency	recency	T	monetary_value	prob_alive	predicted_purchases
13360	4.0	2025.0	2157.0	2362.500000	0.600631	0.175312
62716	4.0	1697.0	1783.0	2611.750000	0.616420	0.211102
66522	4.0	1694.0	1731.0	2959.250000	0.636601	0.217847
37770	4.0	2127.0	2166.0	3118.000000	0.637043	0.175456
100000	9.0	1734.0	1843.0	3154.444444	0.766319	0.465205
1	24.0	2134.0	2165.0	3626.291667	0.910900	1.071894

These are customers who are MOST likely still active, so the chance of buying again is high.



From here, it can be seen that the average customer will return within 12 months is 1-2%. This is normal because this company offers insurance services.

```
              frequency  recency       T  monetary_value  prob_alive  \
customer_id
1                  24.0   2134.0  2165.0     3626.291667    0.910900
598                 1.0   1040.0  1202.0     3558.000000    0.217734
837                 1.0    530.0   817.0     4863.000000    0.183983
839                 1.0    997.0  1585.0     2211.000000    0.139103
858                 1.0     85.0  1219.0     3856.000000    0.058809

              predicted_purchases
customer_id
1                        1.071894
598                      0.013067
837                      0.014978
839                      0.007697
858                      0.001905


Customers with atleast 1 repeat purchase
 6257
```

This means that out of all 6257 customers, there are at least 1 valid repeat purchase to calculate the lifetime value.

```
summary[['monetary_value', 'frequency']].corr()
```

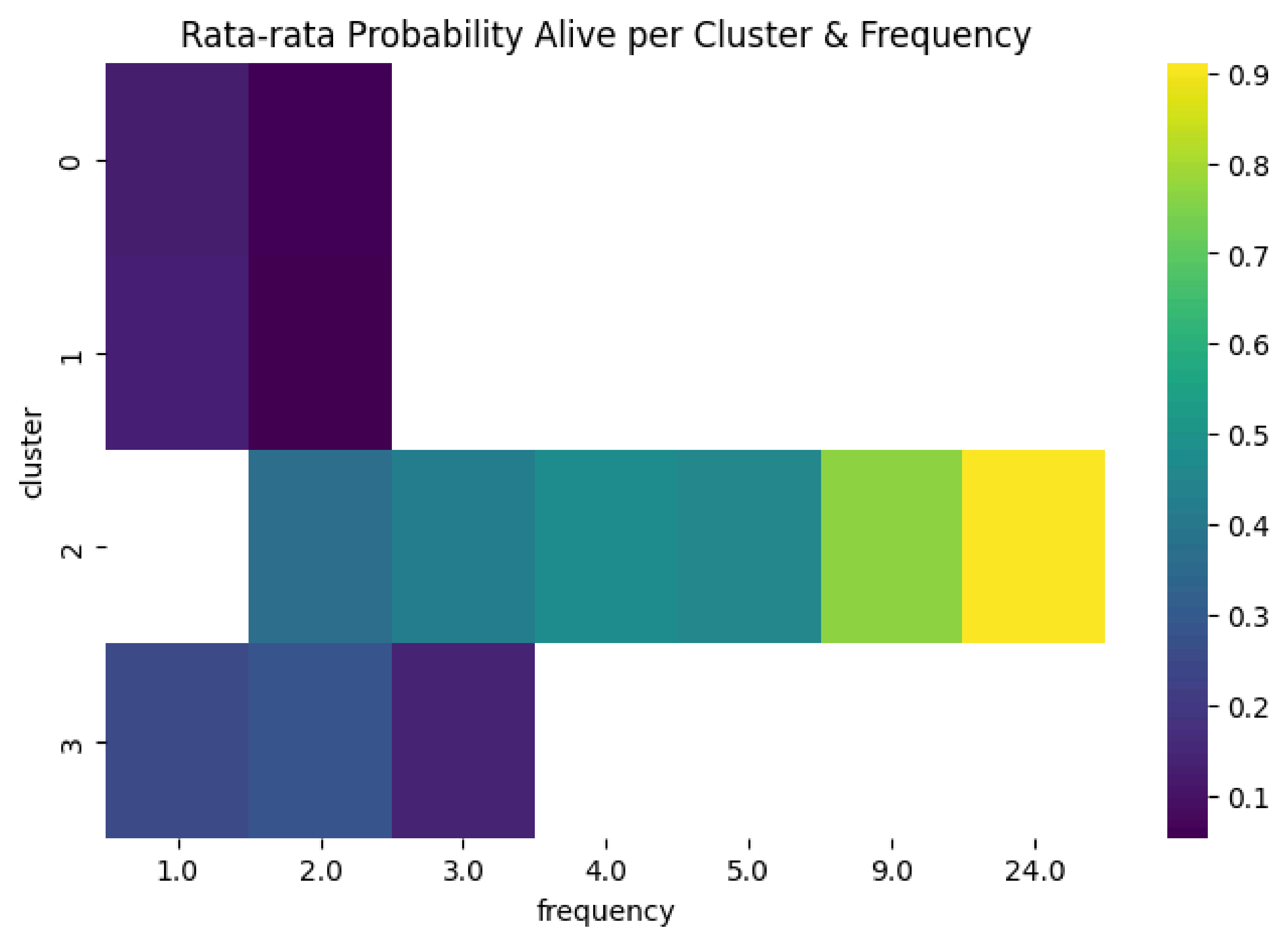
	monetary_value	frequency
monetary_value	1.000000	-0.006971
frequency	-0.006971	1.000000

Purchase frequency does NOT correlate with average purchase value, so it is permissible to use Gamma-Gamma to estimate monetary value.

	prob_alive
cluster	
0	0.130506
1	0.130766
2	0.416489
3	0.288037

Cluster two has a high chance of surviving in this company.

if from the frequency cluster two is also very high



```
customer_id
97970     1078.847051
98084     1807.197834
98121     3793.800985
98276     4733.267937
98404     3230.120813
98418     1035.568237
98745     1830.420613
98860     4824.047889
99267      864.564140
100000    3174.353827
dtype: float64
```

Customer 97970 is predicted to spend an average of 1,078 per transaction in the future.

```
Expected conditional average profit: 3189.301664306418, Average profit: 3023.1962925967364
```

The average predicted profit per trade is slightly higher than the historical average. This is normal, as the model corrects to the mean.

```
customer_id
1       5484.686521
598      318.107206
837      396.137543
839      117.932796
858       92.790864
994      128.288824
1023     487.693135
1039     124.718697
1099     205.684327
1224      65.978930
Name: clv, dtype: float64
```

Customer ID 1 is predicted to generate 5,484 revenue in the next 12 months (based on shopping frequency & average value).

Strategi dan Rekomendasi Bisnis

1. Focus on customers who have purchased between 0-1000 days by providing attractive offers and discounts on products to increase their chances.
2. Re-attract customers who have not purchased for a long time, for example, by providing discount vouchers for customers who have not purchased for a long time.
3. create a loyalty program to increase customer loyalty to this company.

bibliography

Dalimunthe M.H dan Aslami N.2021.Perencanaan dan Strategi Pemasaran Asuransi.VISA: Journal of Vision and Ideas 1(1): 54-67[Publish]

https://ojk.go.id/id/kanal/iknb/data-dan-statistik/asuransi/Documents/Pages/Statistik-Perasuransian-2022/Statistik%20Perasuransian%202022.pdf

Dashboard interaktif klik disini