Customer segmentation

Homework 1. The data - Bank 2
Group 6: Shumeiko Lyubov, Kozin Artem

The description for Task #1

What are the features in dataset? Are they significant?

The dataset consists of two dataframes:

- Clients,'s data and transactions data. Client's data have several features that consist of:user_id report employee_count_nm bankemplstatus customer_age.
- The transactions there are consist of user id, mcc code, currency ark, transaction amount and transaction data.

I analyzed most of the features in the dataset clients data as follows: user_id report   employee_count_nm   customer_age .But they showed insignificant values.
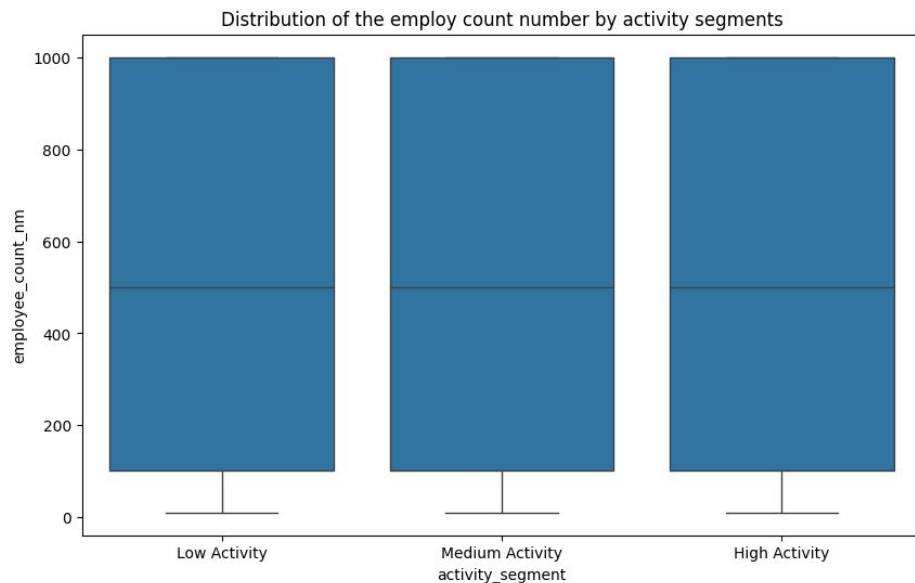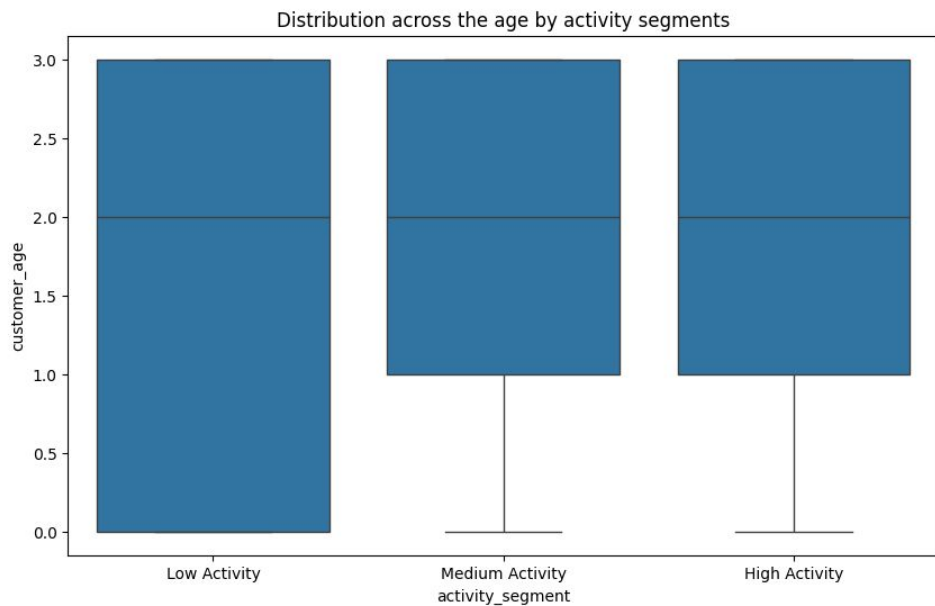
In dataframe transactions almost all features showed an extremely significant role (except currency rk),such as mcc_codetransaction_amt transaction dttm and user id.

User id helped me to identify the client transaction data, and transaction amount helped me to do SRF analysis, as well as the Recency Frequency and monetizing features. Also I used these features to identify the first first day of becoming a client of our company as a parameter of clustering as sames mcc.

activity_segment	count
Medium Activity	48515
Low Activity	30813
High Activity	16672

Distribution of Customer Age across Activity Segments

High Activity
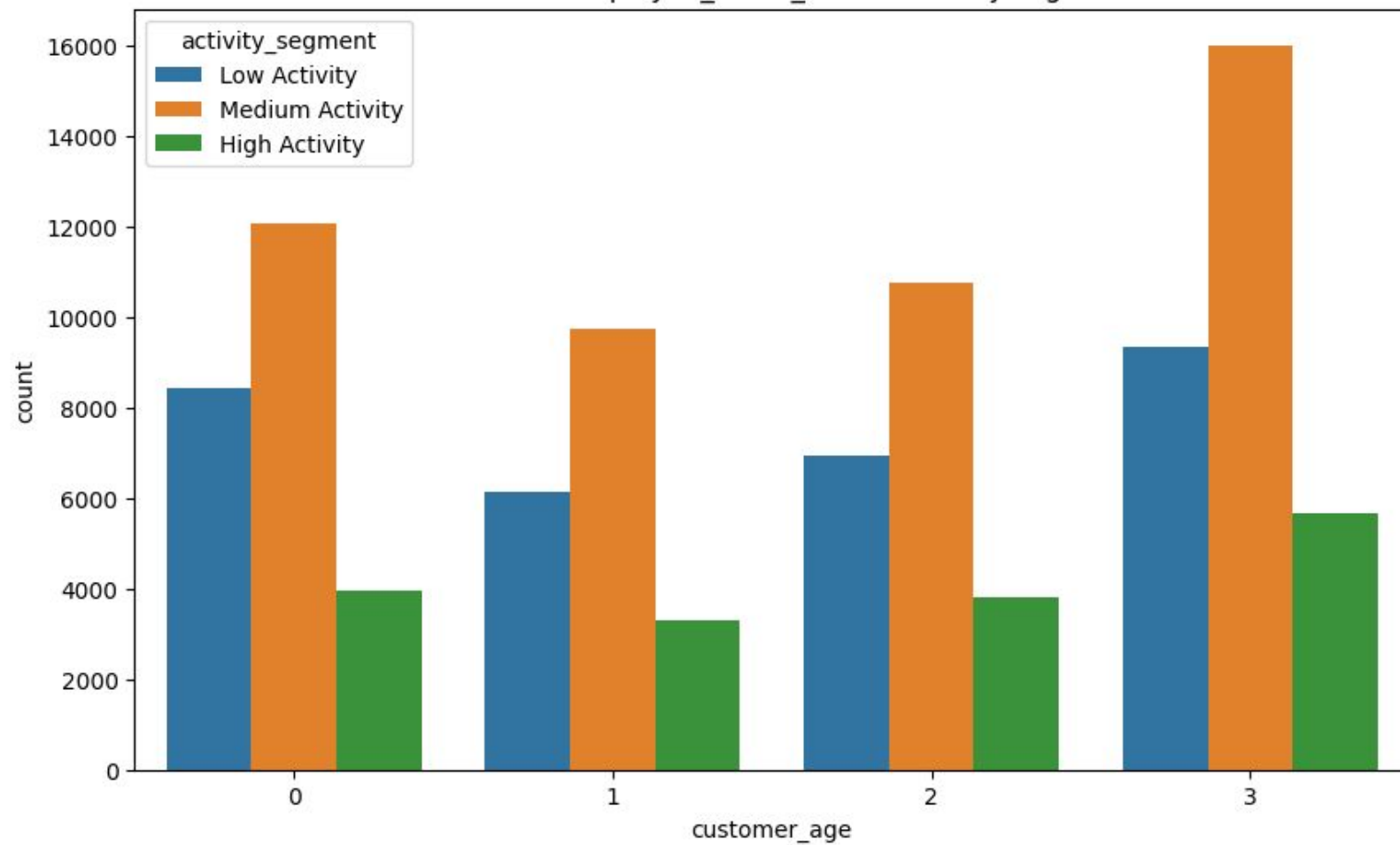
17.4%

Medium Activity

50.5%

32.1%
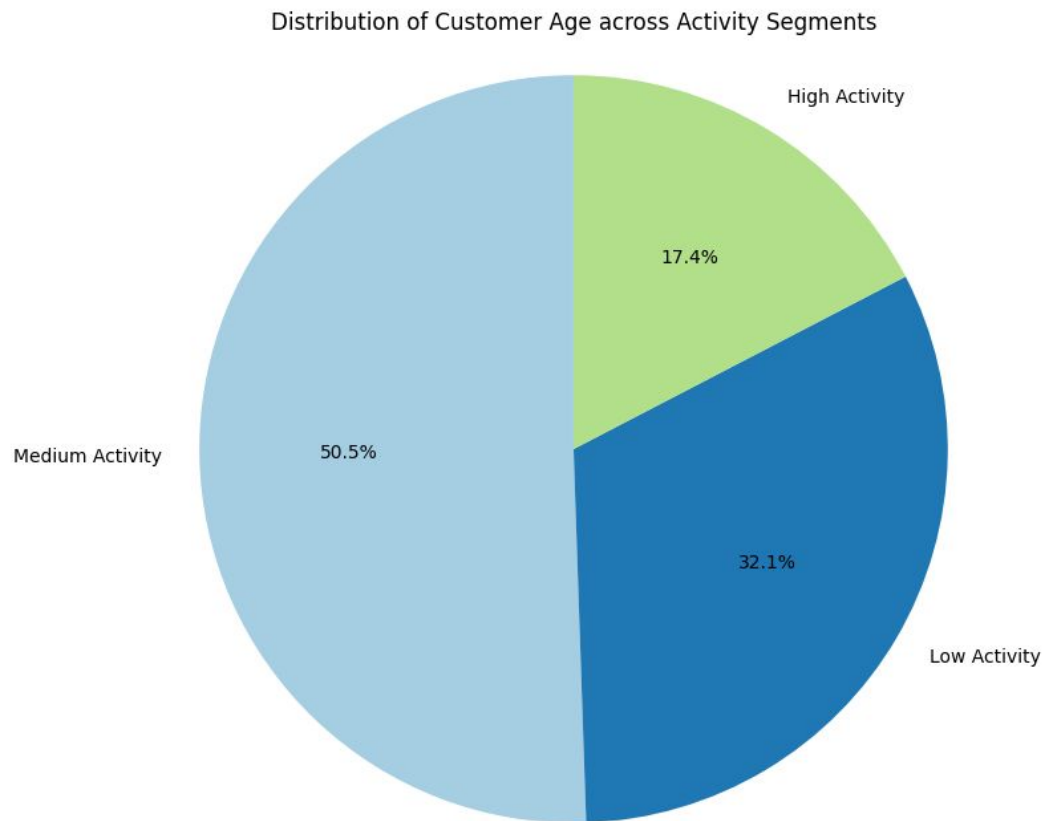
Low Activity

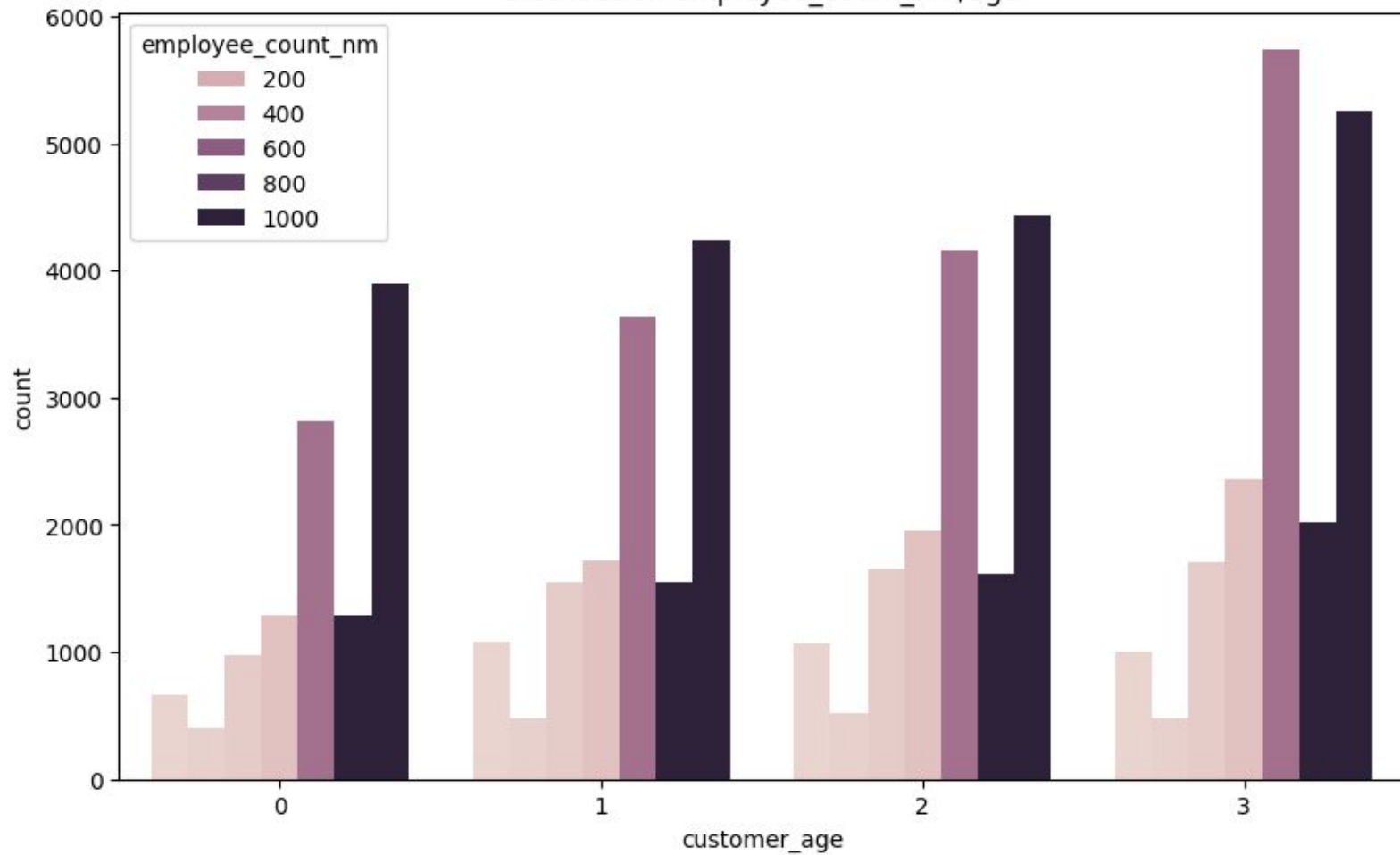# How our clients look like due to 'client's' data

Distribution employee_count_nm for activity segments
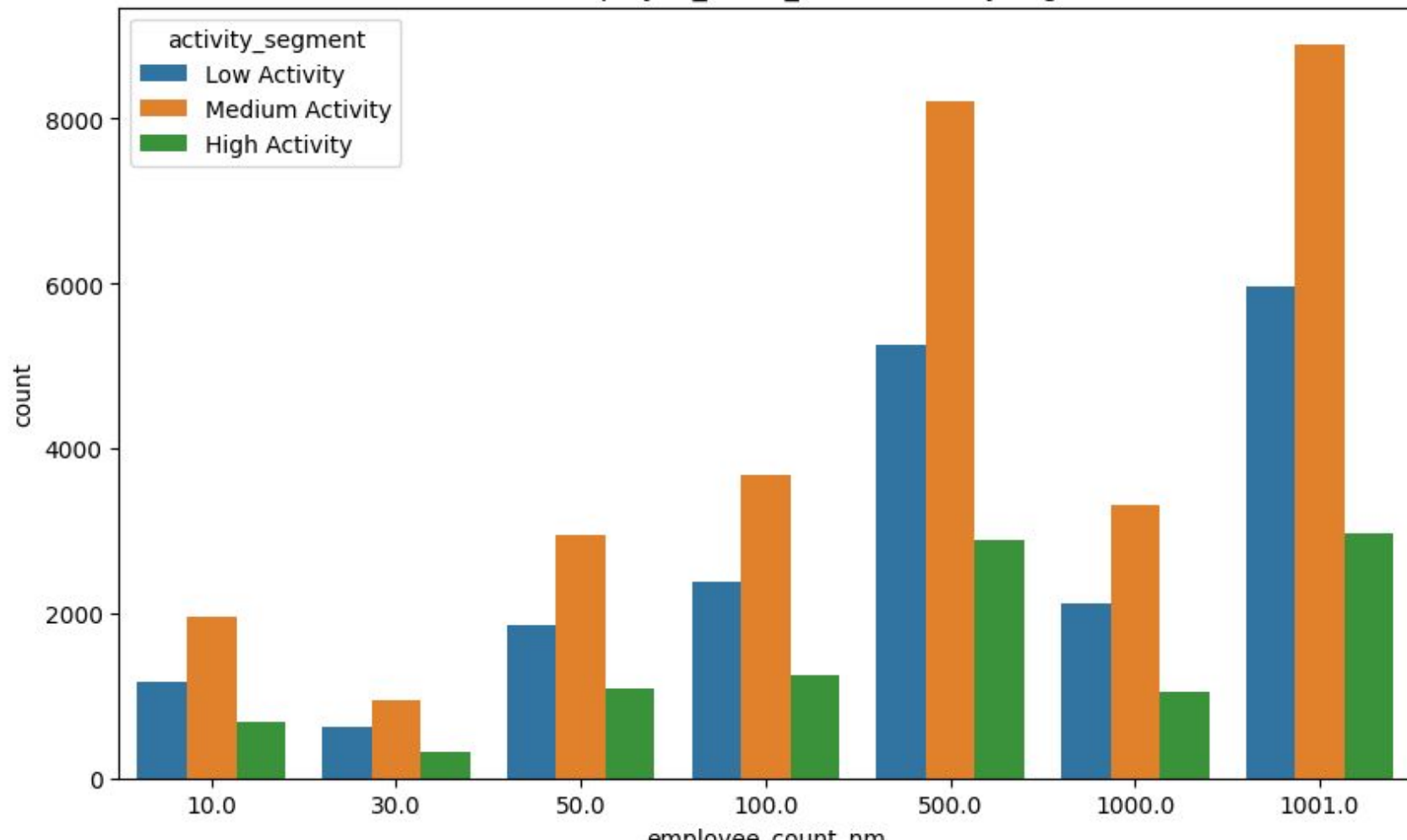
Distribution employee_count_nm for activity segments

Distribution employee_count_nm/age

Task #2

Cohort analysis.Analysis of customer behavior

How our clients move?

```
1 print(cohort_pivot)
```

period_number	0	1	2	3	4	5	6
cohort							
2021-10	484.0	475.0	475.0	473.0	466.0	471.0	443.0
2021-11	509.0	498.0	499.0	498.0	494.0		
2021-12	581.0	564.0	563.0	564.0	563.0		
2022-01	541.0	529.0	518.0	527.0	528.0		
2022-02	535.0	526.0	519.0	518.0	512.0		
2022-03	517.0	506.0	496.0	502.0	500.0		
2022-04	551.0	533.0	534.0	538.0	516.0		
2022-05	596.0	584.0	573.0	573.0	557.0		
2022-06	583.0	568.0	559.0	555.0	554.0		
2022-07	558.0	548.0	550.0	543.0	530.0		
2022-08	573.0	559.0	548.0	552.0	543.0		
2022-09	615.0	599.0	585.0	587.0	571.0		
2022-10	119.0	111.0	109.0	103.0	90.0		
2022-11	27.0	26.0	19.0	19.0	16.0		
2022-12	9.0	7.0	9.0	5.0	NaN		
2023-01	4.0	4.0	2.0	NaN	NaN		

```
1 print(retention_matrix)
```

period_number	0	1	2	3	4	5	6
cohort							
2021-10	1.0	0.983051	0.985321	0.965648	0.976090	0.971701	0.914195
2021-11	1.0	0.984200	0.960011	0.972036	0.972315	0.958753	0.821728
2021-12	1.0	0.946665	0.969773	0.971022	0.963902	0.944791	0.777792
2022-01	1.0	0.973264	0.972284	0.970185	0.962906	0.948768	0.808231
2022-02	1.0	0.973500	0.967524	0.969732	0.960249	0.932970	0.739153
2022-03	1.0	0.968586	0.963460	0.961751	0.946635	0.902340	0.707282
2022-04	1.0	0.973422	0.966369	0.959819	0.940799	0.915984	0.706512
2022-05	1.0	0.966994	0.959541	0.949012	0.939666	0.912930	0.684727
2022-06	1.0	0.965672	0.952363	0.954975	0.941169	0.901990	0.715796
2022-07	1.0	0.961028	0.959398	0.955388	0.936591	0.903133	0.674060
2022-08	1.0	0.962498	0.955505	0.950922	0.935488	0.869408	0.698782
2022-09	1.0	0.967318	0.956187	0.952043	0.910598	0.887981	0.714624
2022-10	1.0	0.909699	0.903567	0.818841	0.793200	0.639911	NaN
2022-11	1.0	0.886214	0.770241	0.634573	0.444201	NaN	NaN
2022-12	1.0	0.859504	0.768595	0.471074	NaN	NaN	NaN
2023-01	1.0	0.945652	0.543478	NaN	NaN	NaN	NaN
2023-02	1.0	0.906977	NaN	NaN	NaN	NaN	NaN
2023-03	1.0	NaN	NaN	NaN	NaN	NaN	NaN

Retention of clients throughout months

Monthly Cohorts: User Retention
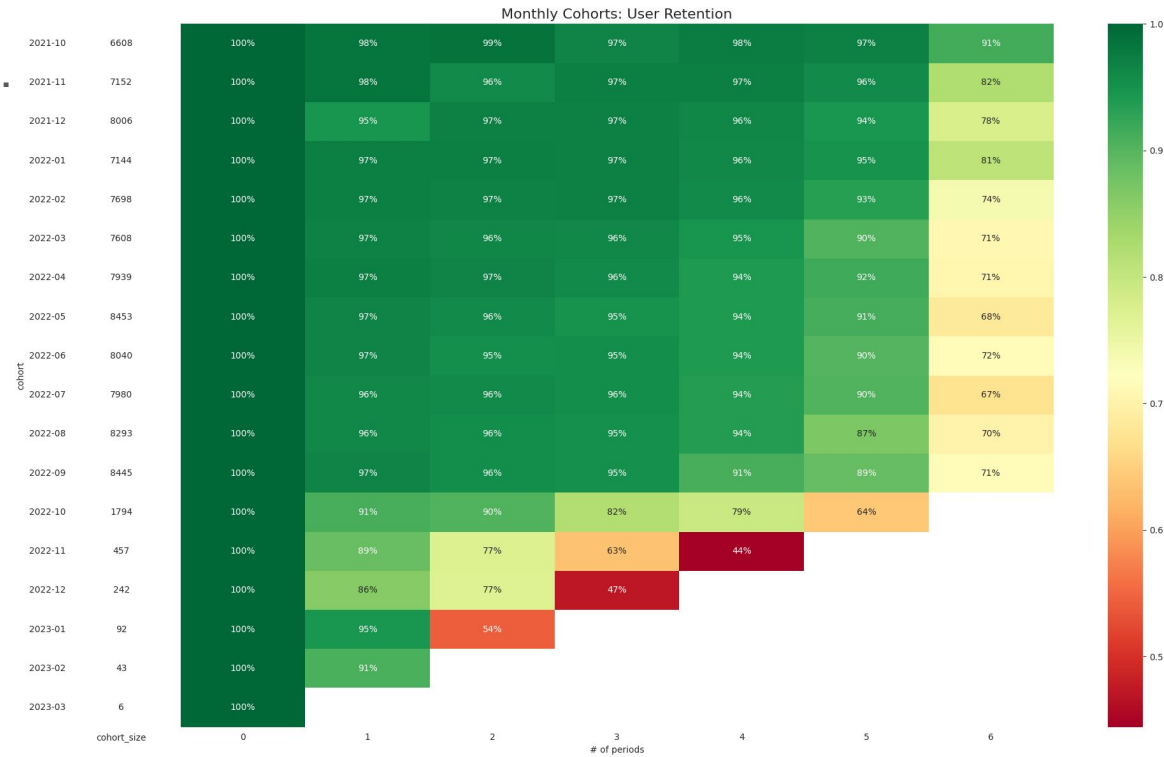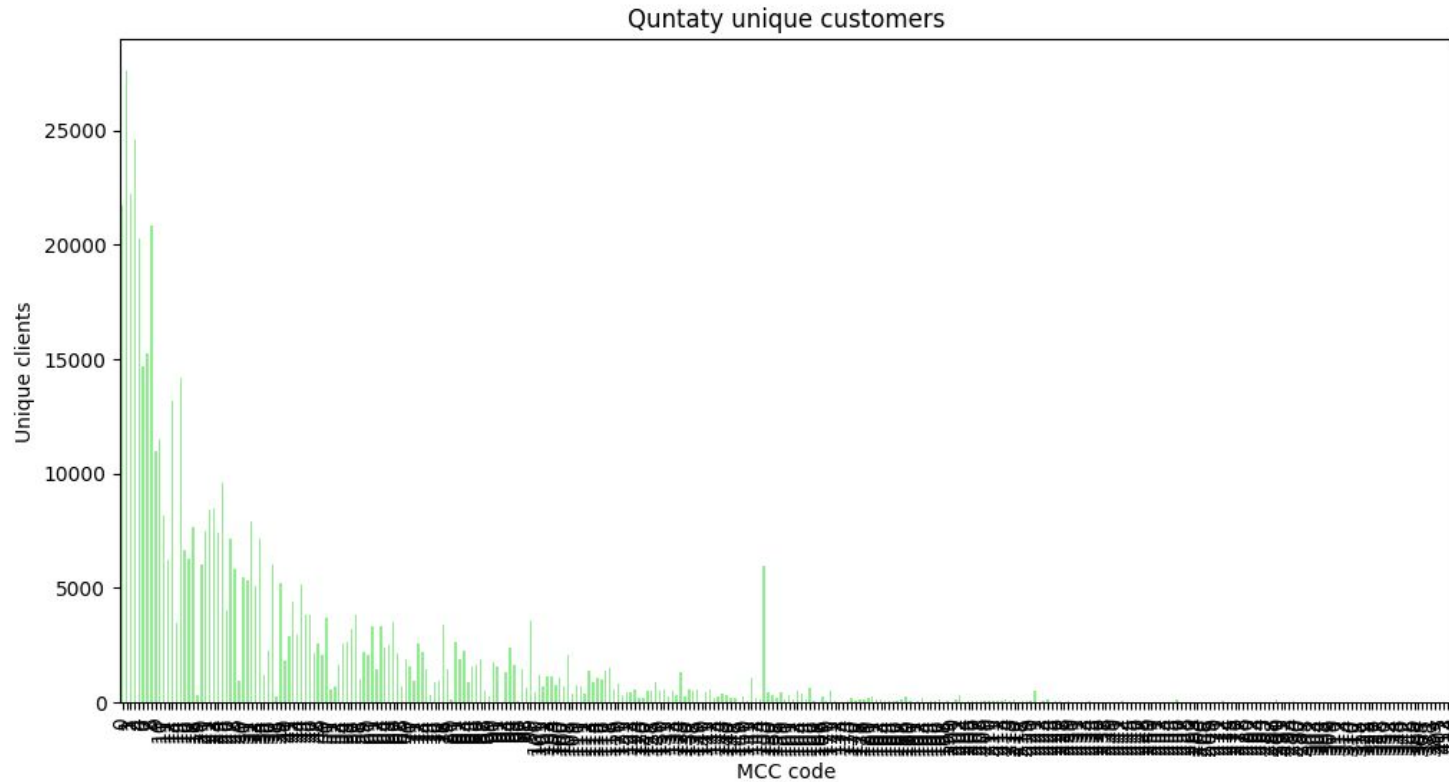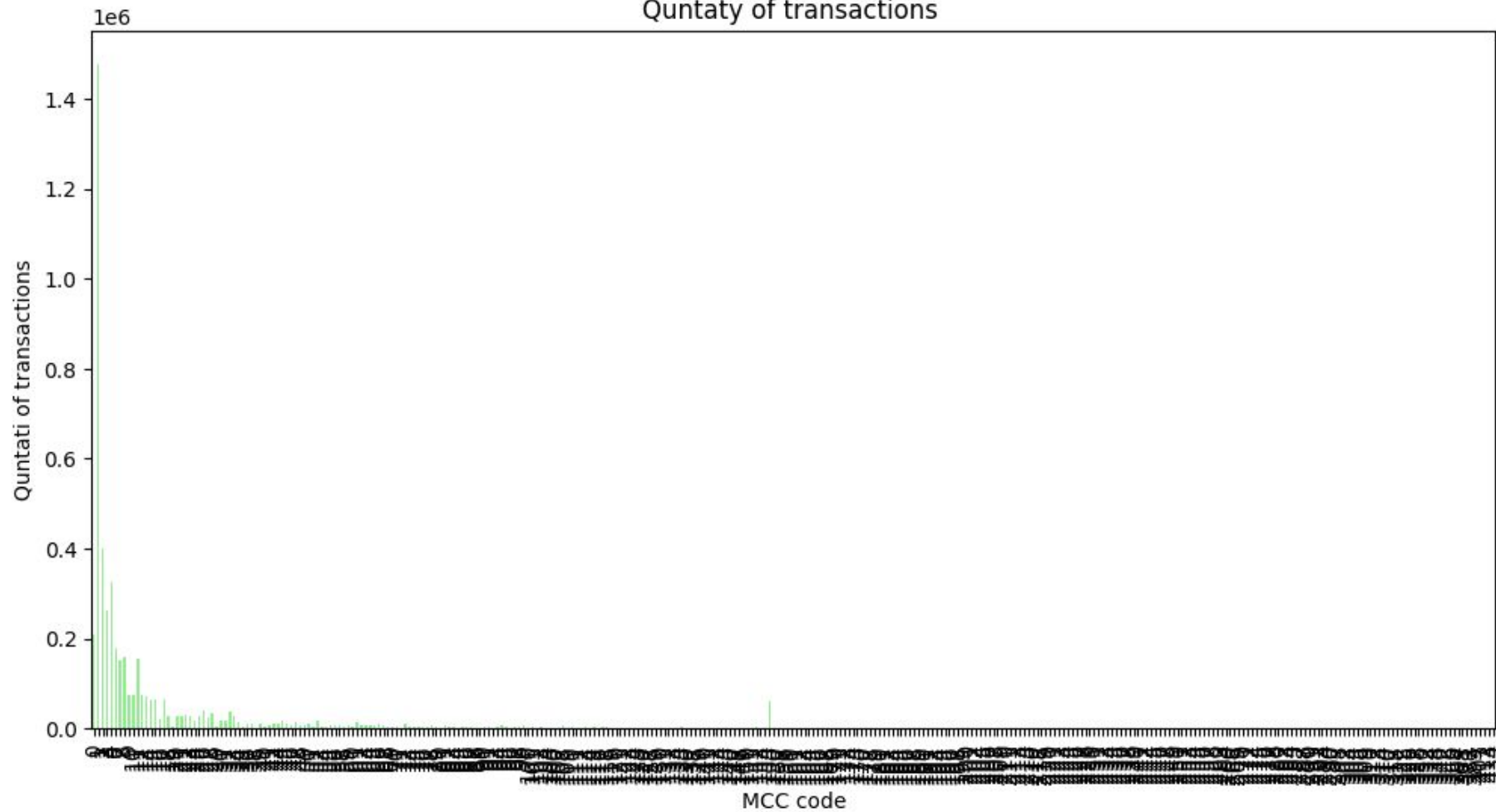
Why our clients shows worrible outflow of customers parameters of activity (especially at last months)

- what changed
- are there any dependencies of MCC
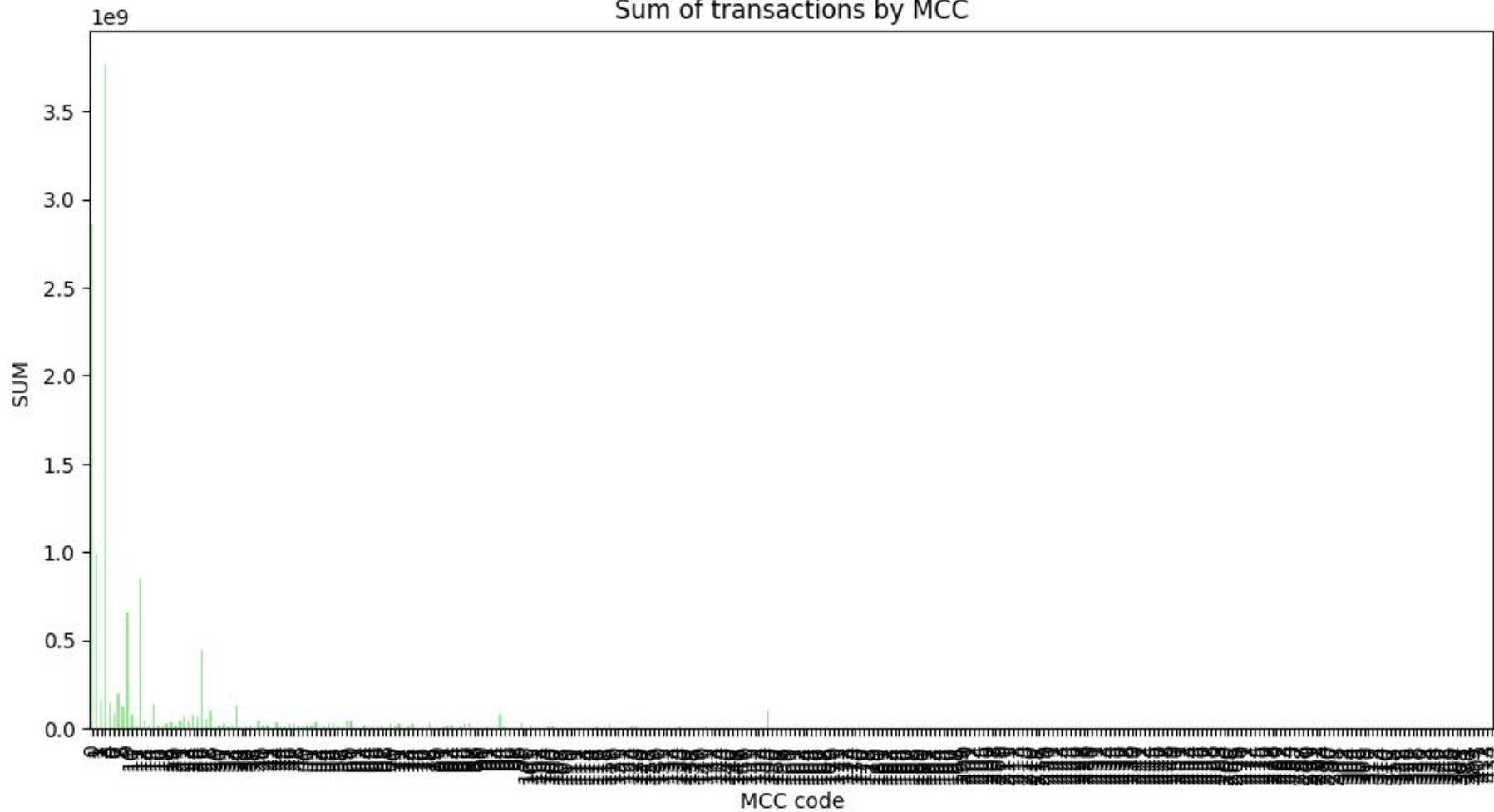- is it possible that the categories have also changed
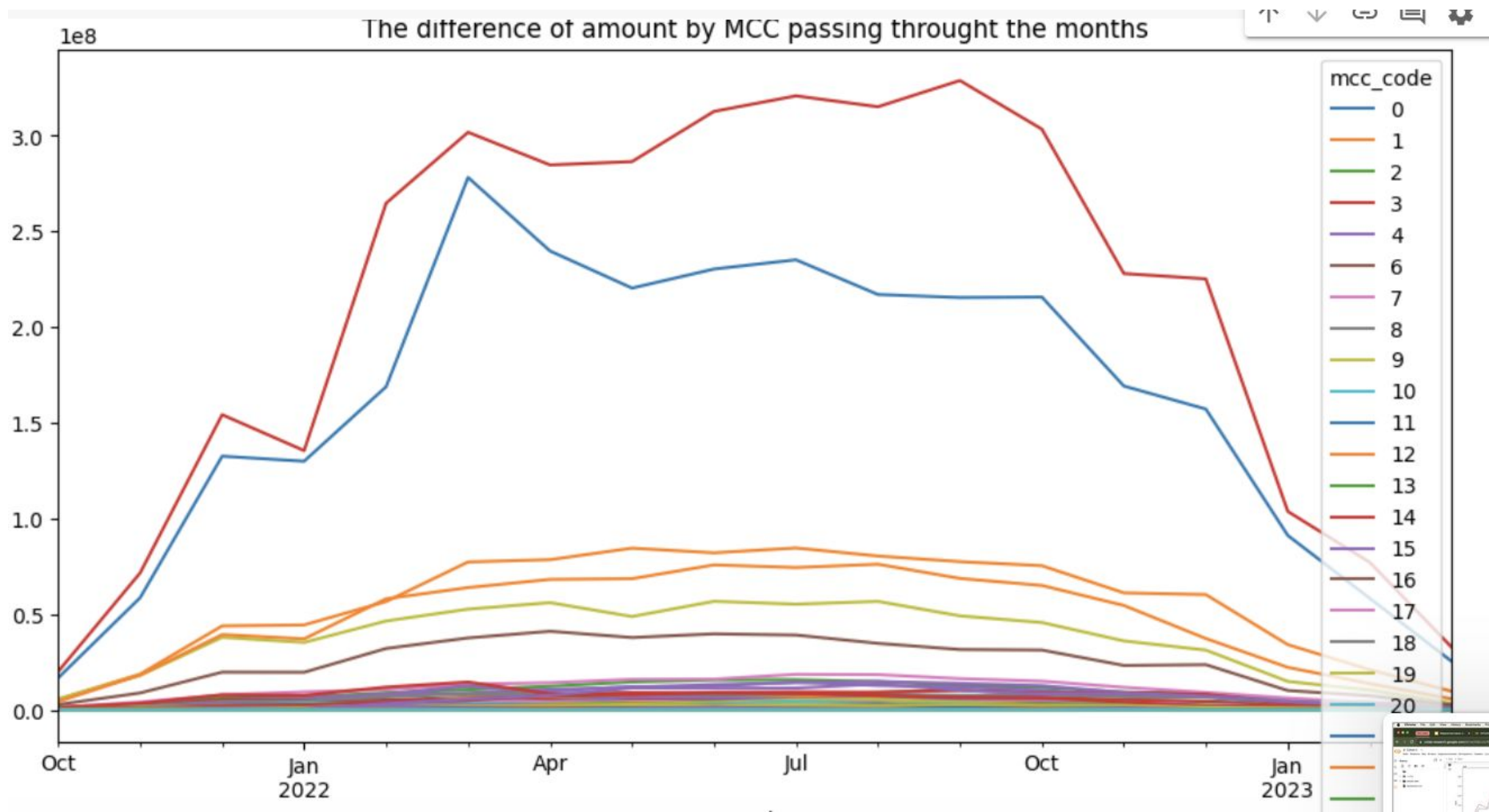
What is our customer behavior?

Quntaty of transactions

Sum of transactions by MCC

The difference of amount by MCC passing throught the months

mcc_code: 0, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
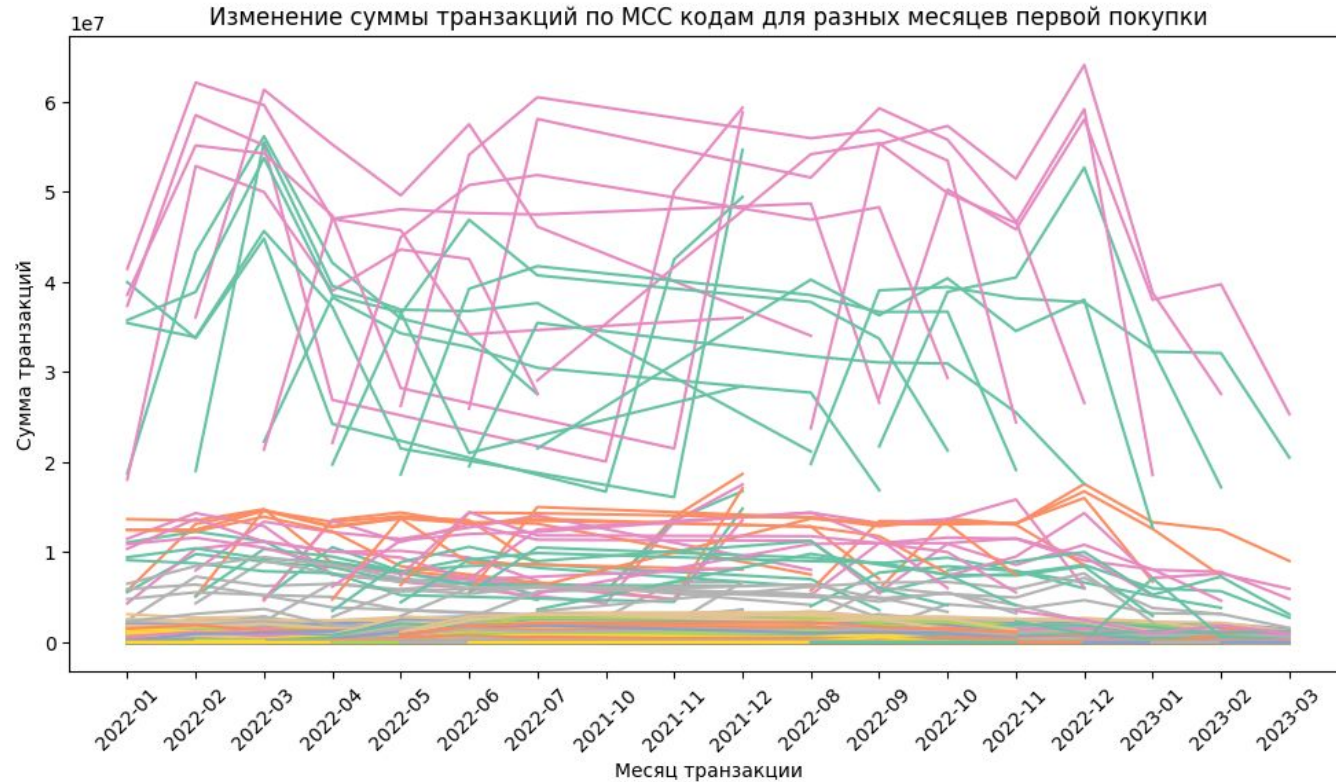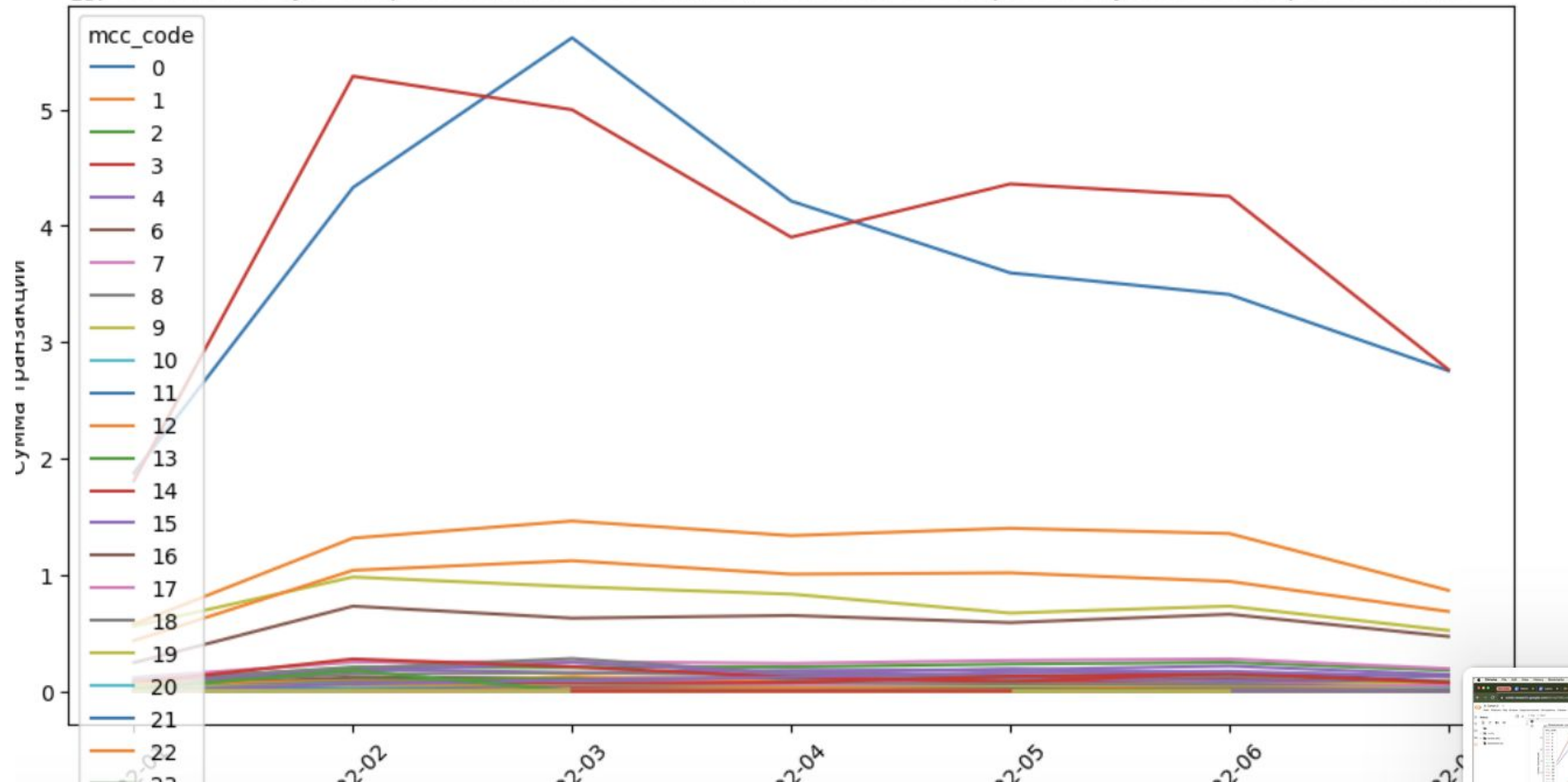
The theory of the changing product as cause.MCC feature

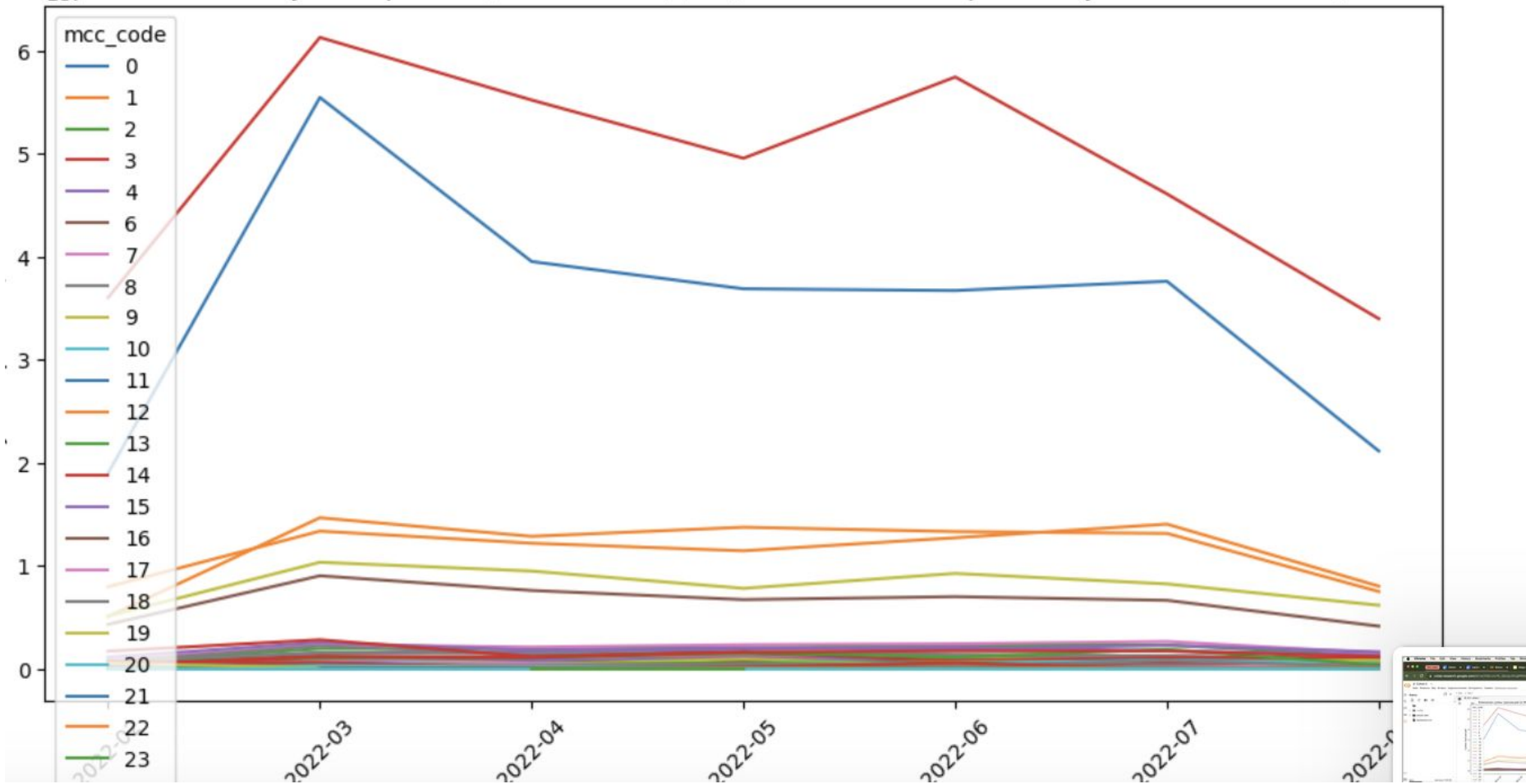

the color - the type of mcc

The behavior of different cluster based on the month of first purchase

Изменение суммы транзакций по МСС кодам для клиентов с первой покупкой в январе 2022 года

Изменение суммы транзакций по MCC кодам для клиентов с первой покупкой в 02 2022 года

mcc_code
- 0
- 1
- 2
- 3
- 4
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23

The description for Task #2

What can be said about the customers behavior?

1) We could see that our clients begin their way and after around 4-6 months, shows some decline in the activity.

2) Last months that we analyze, shows some more dependency in the decline, which is the risk parameter

3) Based on my analysis, the majority of activities over the years are concentrated around Product 3 and Product 11, which account for a significant portion of the total engagement.

4) Additionally, nearly 99% of all activities and associated revenue fall within the MSc range of 0 to 200, with only one exception. Notably, 60% of all activity occurs within the more focused range of 1 to 40, highlighting a concentrated area of interest or investment within this narrower range.

5) Only two MCC transactions used by our customers most significant ones Eating places and Restaurants and Hobby, Toy, and Game Shops

RFM analysis и Cluster analysis

Task #3

	count
segment	
Promising	191173
New Customers	196460
Need Attention	216750
Can't Lose	350490
About to Sleep	414114
Champions	480503
At Risk	767789
Potential Loyalists	785994
Loyal Customers	818085
Hibernating	824497

dtype: int64

Distribution of Segments
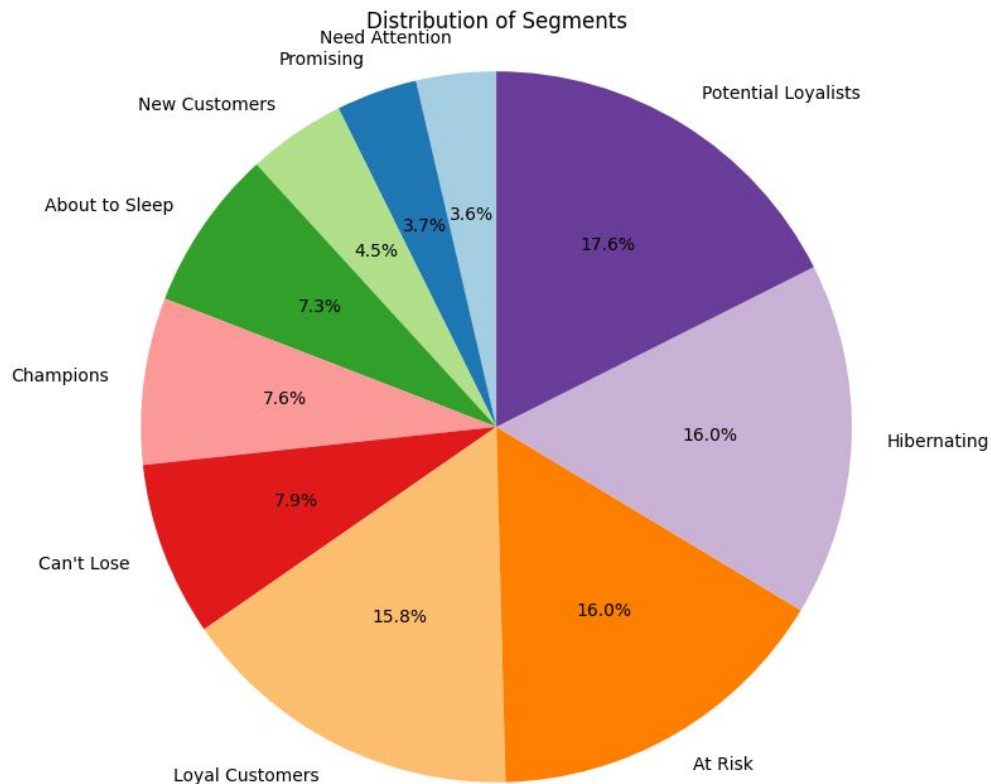
Distribution of recency
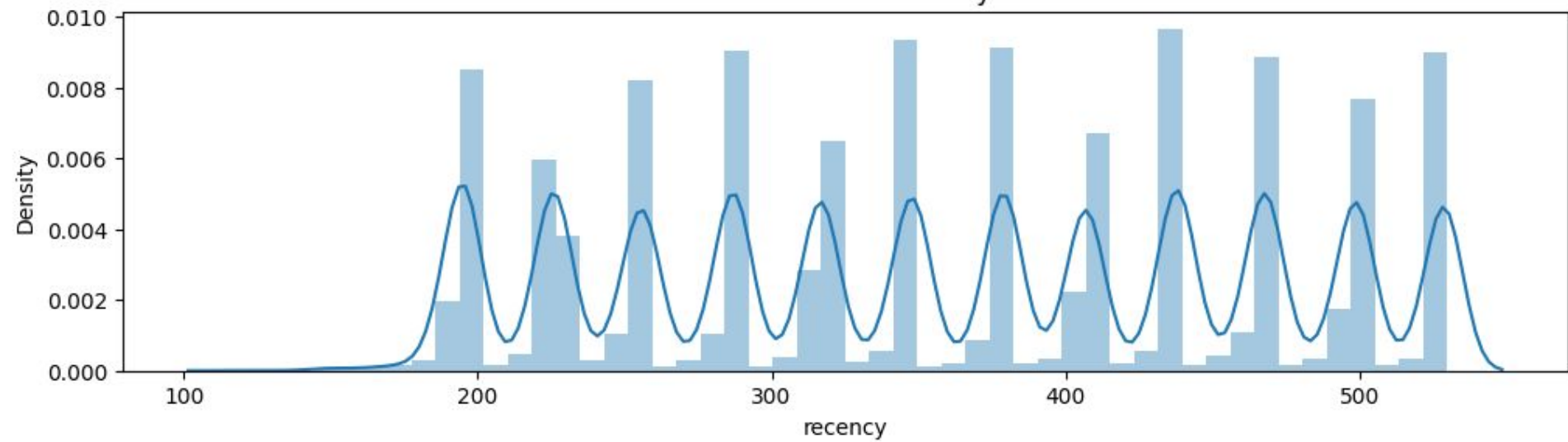
Distribution of frequency

Distribution of monetary
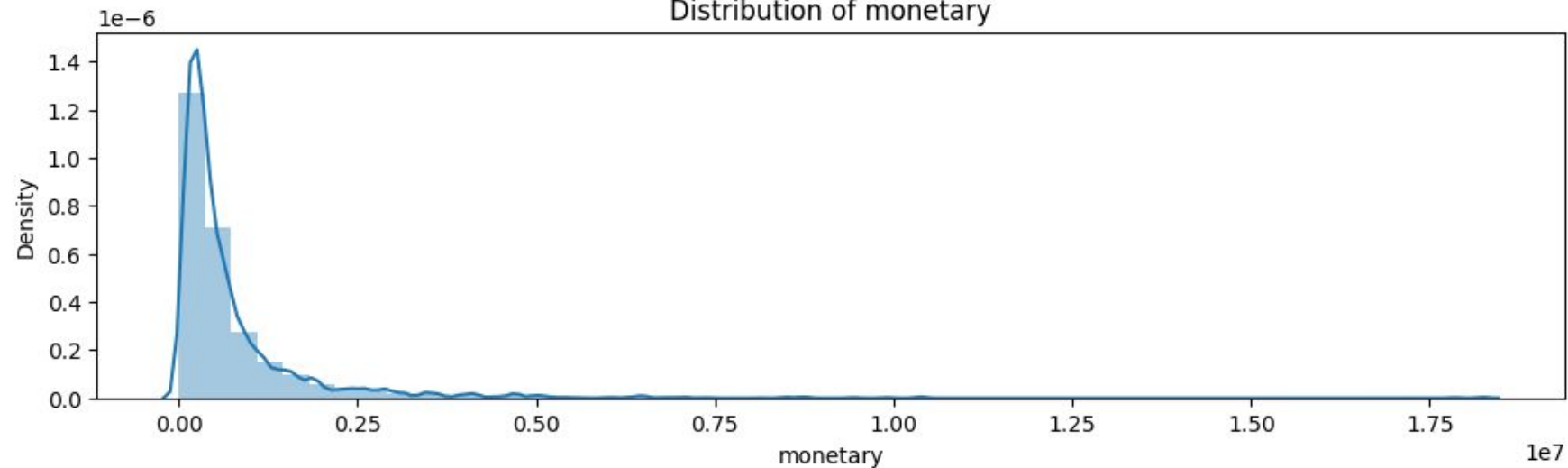
Distribution of recency
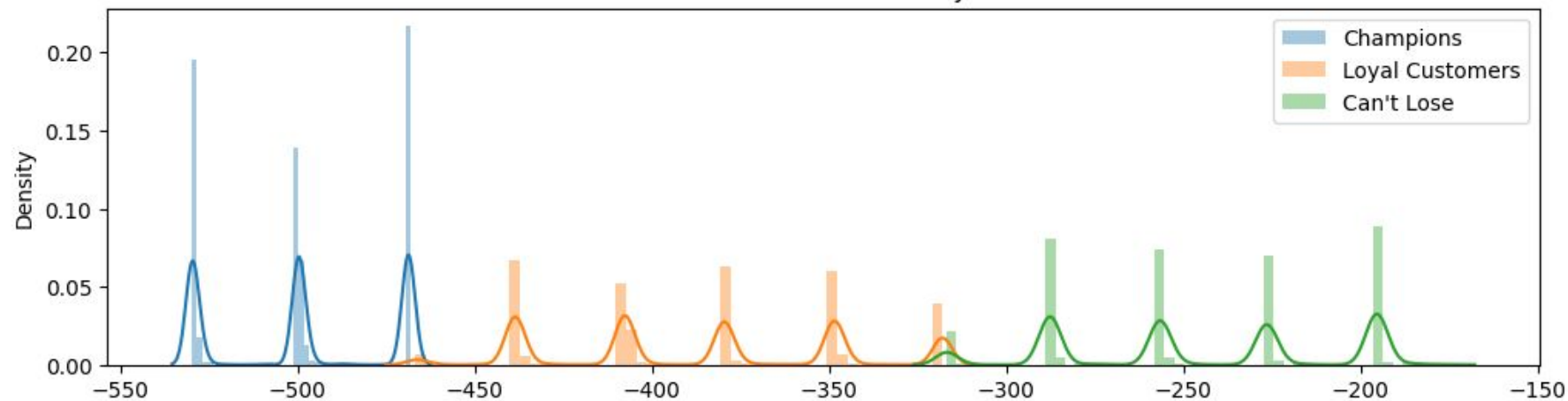
Distribution of frequency
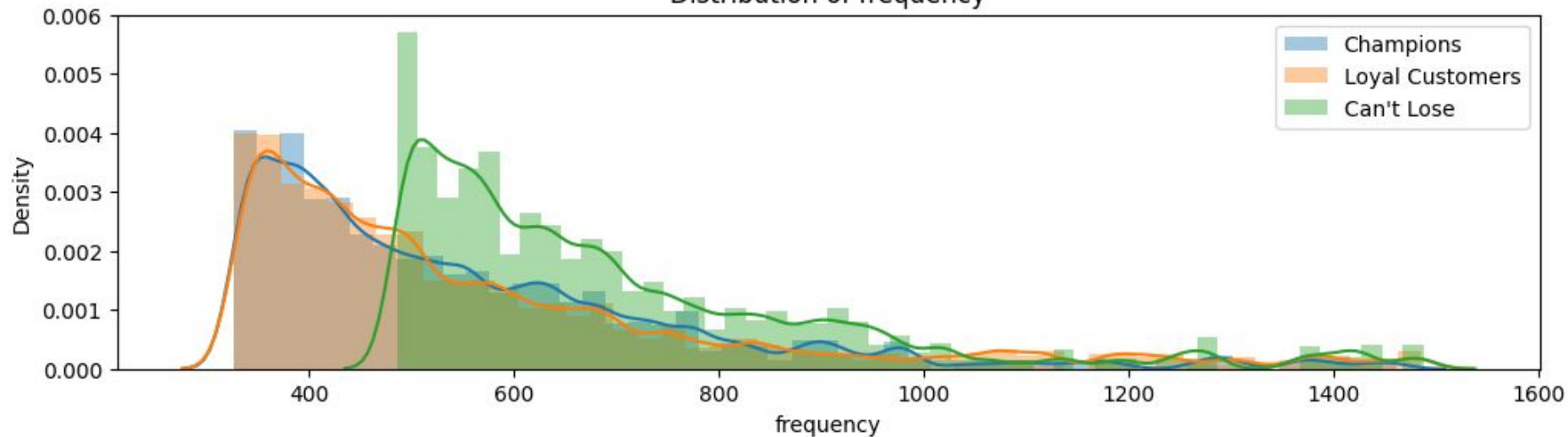
Distribution of monetary
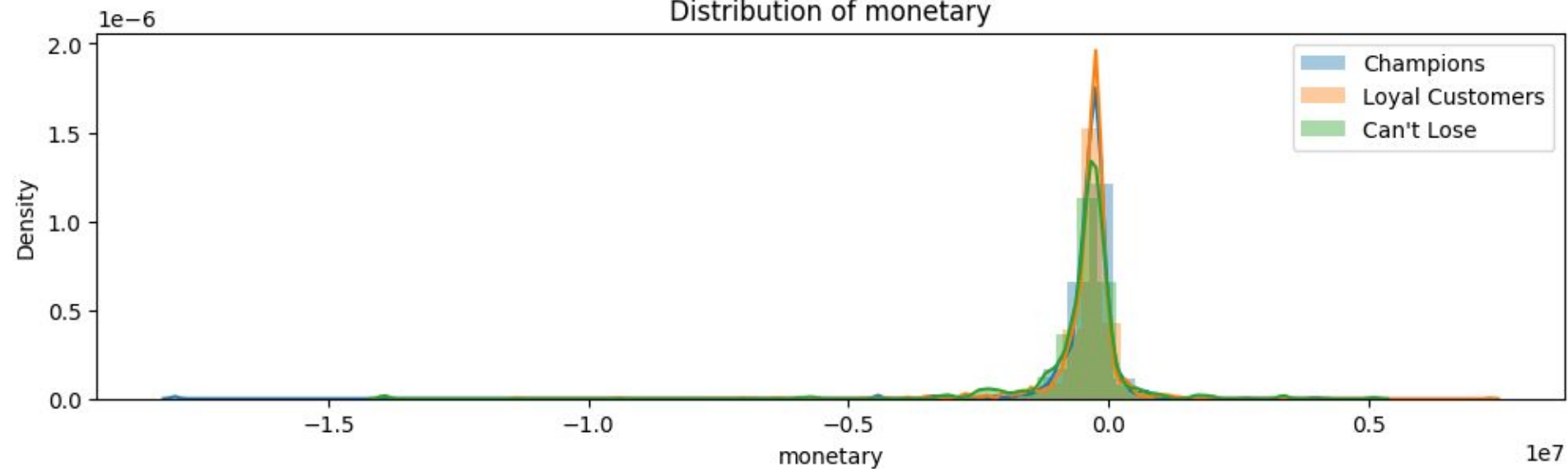
RFM Segments of Customers (count)

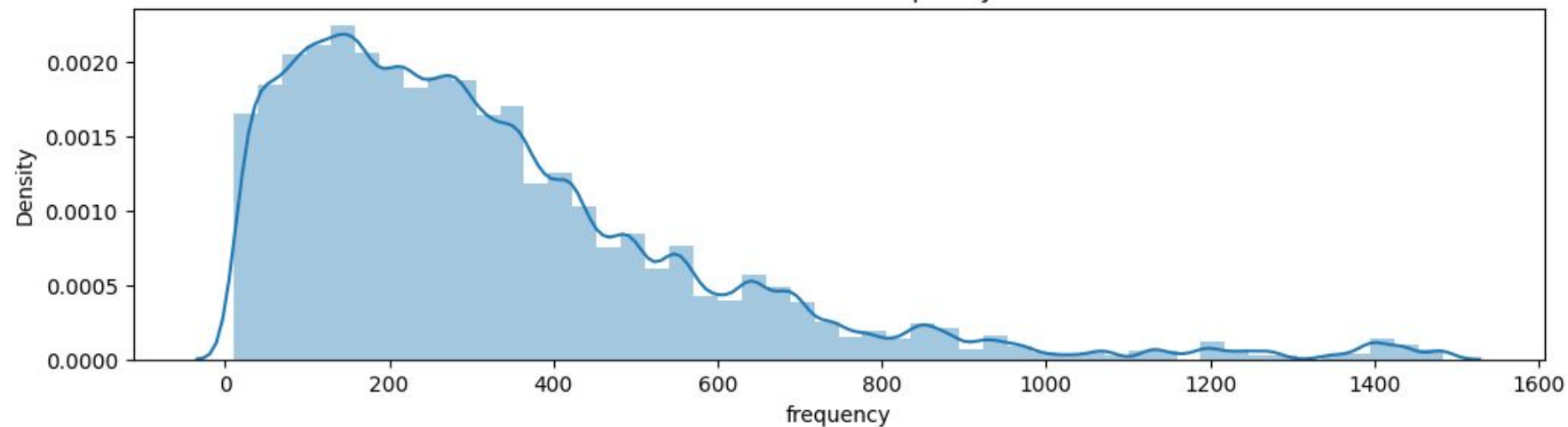|		
Promising (8.5%)	At Risk (13.6%)	Hibernating (13.7%)
Potential Loyalists (7.4%)	Need Attention (10.6%)	Can't Lose (13.6%)
New Customers (6.3%)		
Champions (6.2%)	Loyal Customers (9.4%)	About to Sleep (10.6%)
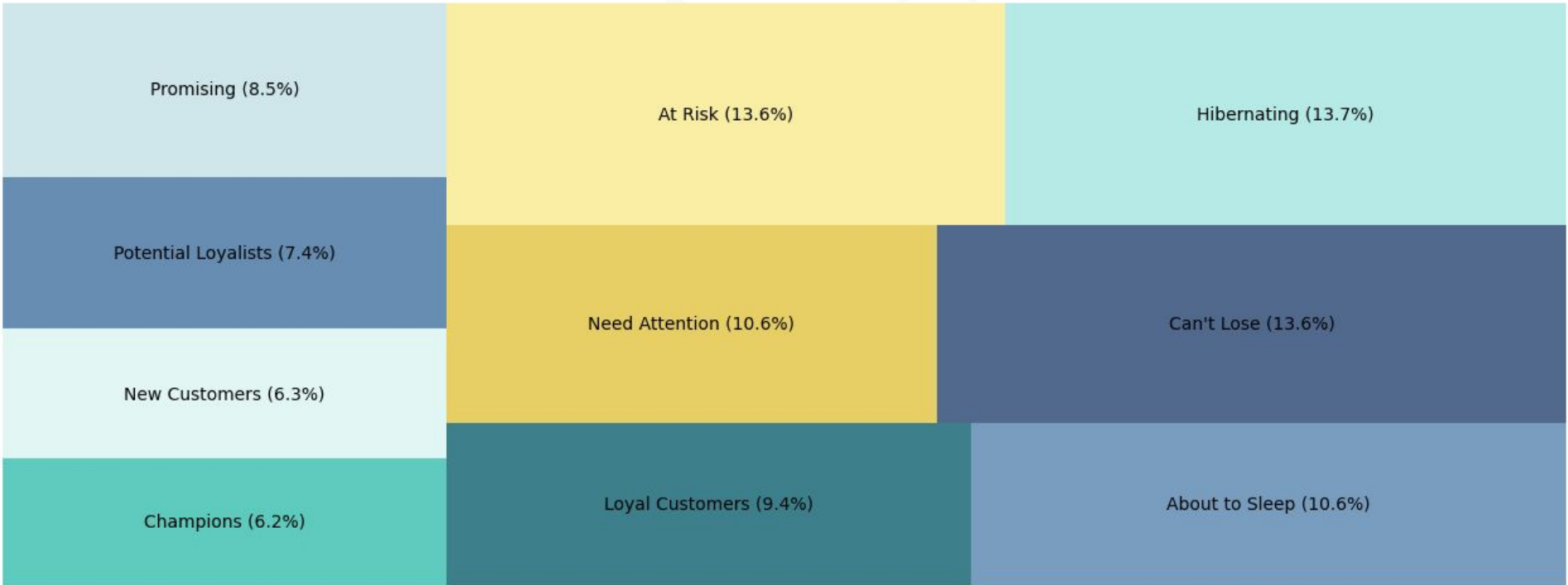
RFM Segments of Customers (monetary)

New Customers (2.2%)

Need Attention (2.6%)

Can't Lose (16.3%)

Hibernating (7.7%)

About to Sleep (3.8%)

At Risk (13.0%)

Champions (14.0%)

Loyal Customers (25.9%)

Potential Loyalists (12.7%)

Promising (1.7%)
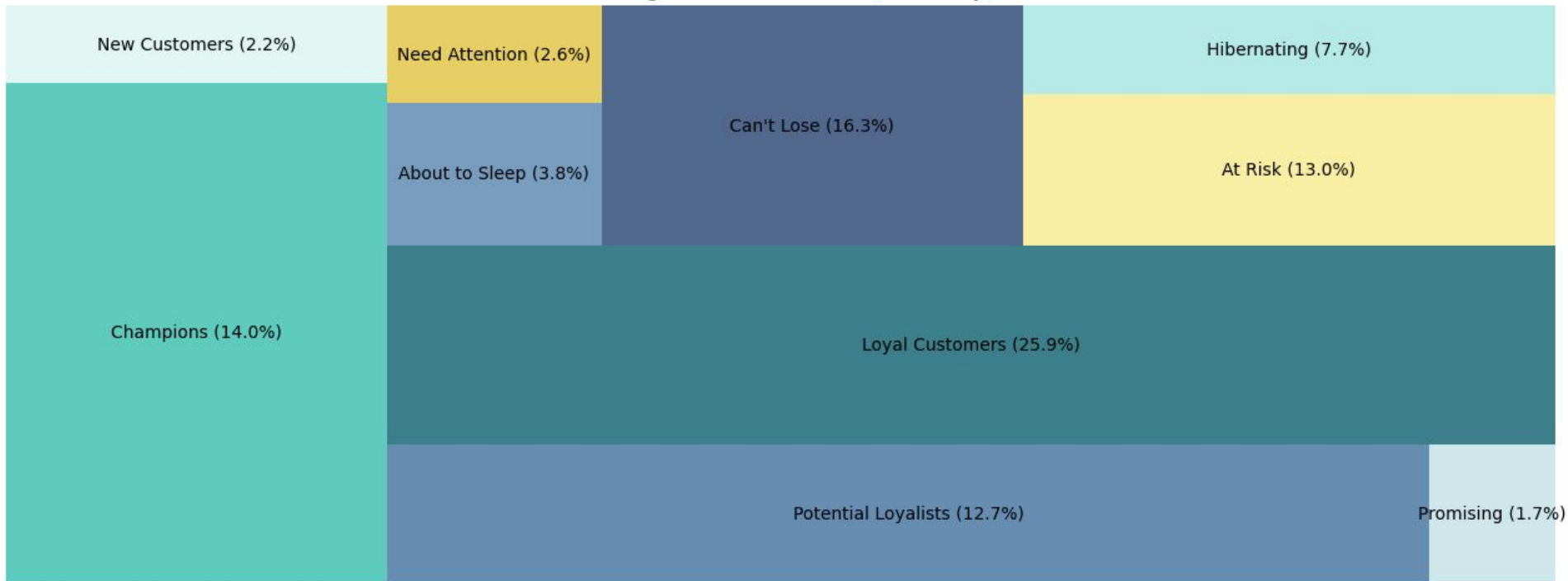
Choose the optimal number of clusters (the task)

Average Recency for Each Cluster

Average Frequency for Each Cluster

Average Monetary Value for Each Cluster
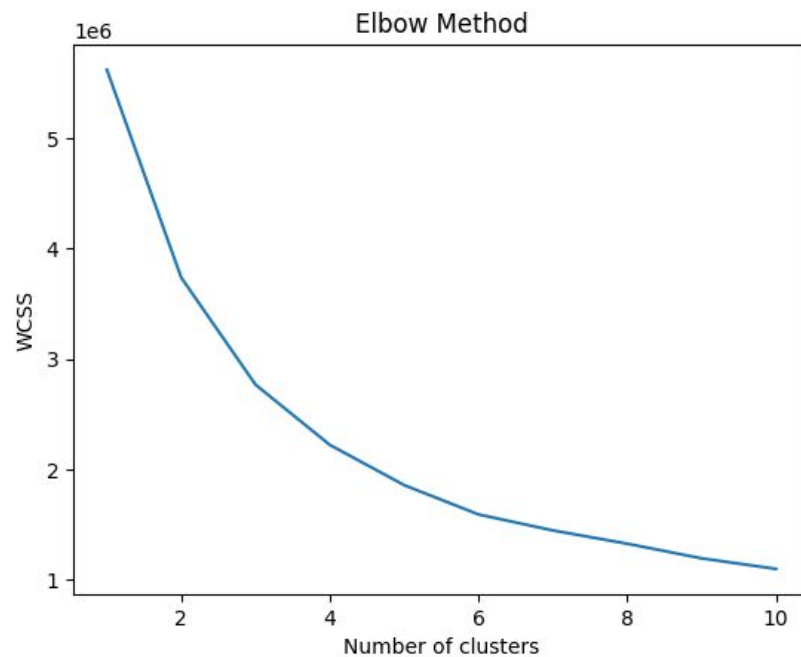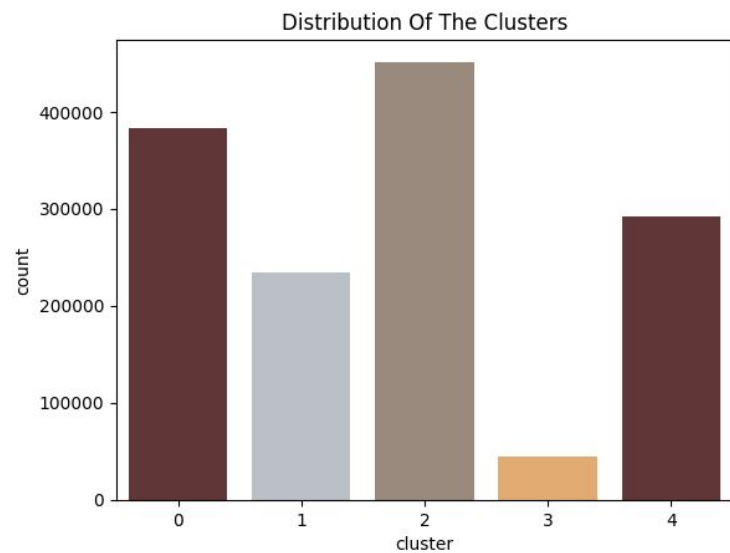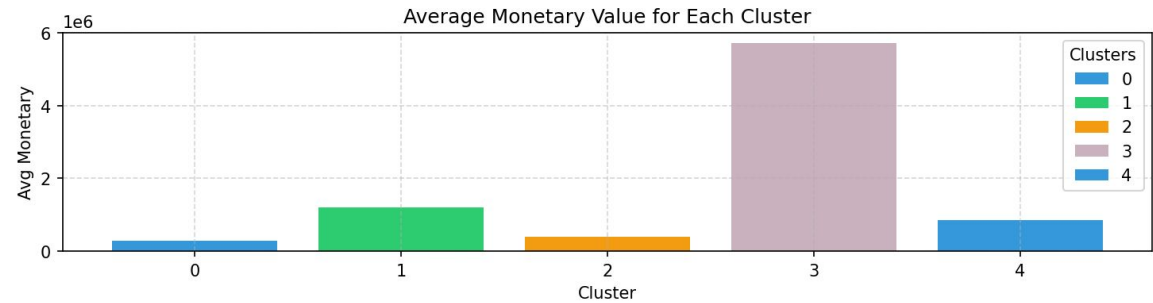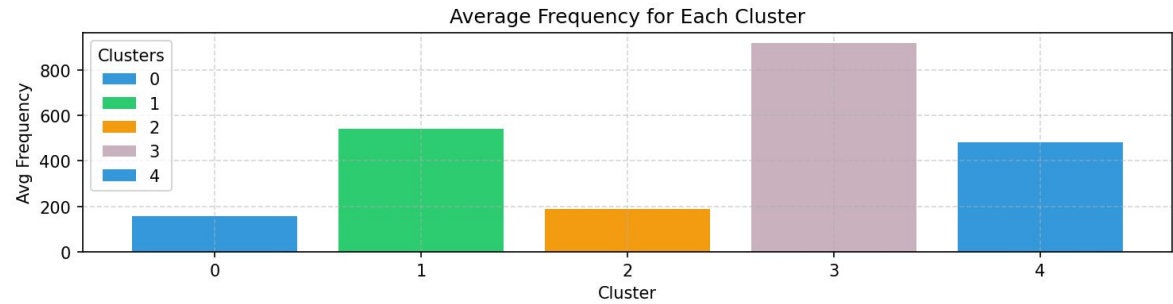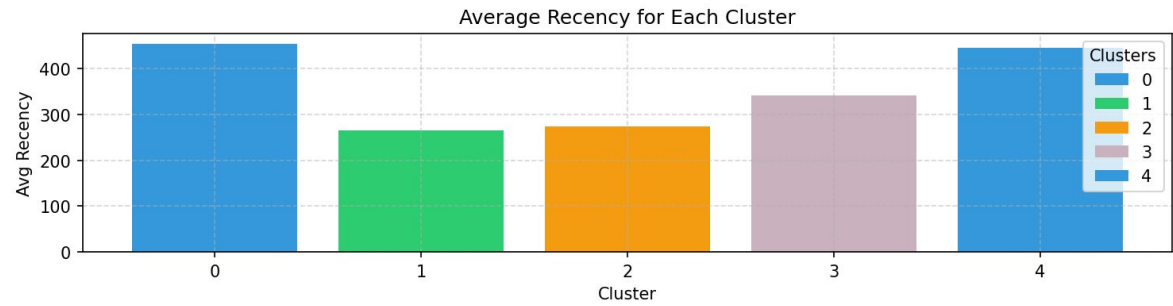
Trying the various combinations:explanation the way of the most appropriate one

1) Previously I identified various combinations of the features, firstly identified features from the client's data and it shows that there are non-significant dependency on their report and other variable. There are no 0 dependency in how large is company
2) Almost no dependency how long does it use it our product. Except the statement that the **lowest activity** mostly shown by the really **old customers**.
3) The clustering by **RFM shows** us really good numbers but as it seems it's not so easy to explain what exact parameters (except for recency and frequency of course) the way how we can define our customer and how we can influence on them)
4) I experimented with clustering based solely on the **first day of using our product** in combination with the **MSC metric**, but this approach did not yield meaningful clusters. It helps me to understand the customer activity but to commonly .However, this analysis revealed significant potential to explore beyond traditional clustering methods.
5) Specifically, it suggests that incorporating MSC **usage patterns** could provide more valuable insights and improve the segmentation approach.

The description for Task #4

Describe the customers in each group. Which significant characteristics can be underlined?

Here we provide to systems of clusters **closely** connected to each other

1) Classical RFM segmentation (2)

   Champions and Loyal Customers

   At Risk, **Can't Lose**, and **Hibernating**

   Promising, **Potential Loyalists**, and **New Customers**

2) The predicted clusters based on rfm analysis (1 - firstly described)

The predicted segmentation

Cluster 3 is the most active and valuable customers who spend the most money and buy the most often.

Cluster 0 and 4 are the customers with the lowest activity who have not made purchases for a long time, but at the same time differ in frequency and monetary value.

Cluster 1 and 2 are the middle segments where customers make purchases relatively recently and with moderate frequency, but differ in spending levels.

The characteristics of predicted segmentation

Cluster 0: Recency (Statute of limitations): High Recency value (over **400**). This means that customers from this cluster have not made purchases for a long time. Frequency (Frequency): Low frequency (less than **200**). Customers make rare purchases. Monetary (Monetary value): Low monetary value. Customers spend relatively little money. Characteristic: These may be inactive or "sleeping" customers who have not made purchases for a long time and do so rarely, with small amounts.

2. **Cluster 1: Recency**: The average Recency value (about **250**). Customers have not made purchases recently, but these are not the oldest customers either. Frequency: Average frequency (about **400).** Customers make purchases quite often. Monetary: The average monetary value. Customers spend moderate amounts. Characteristic: These can be active customers with an average frequency and moderate expenses.

3. **Cluster 2: Recency**: The average Recency value (about **300**). Customers have been shopping not so long ago, but some time has passed. Frequency: Low frequency (less than **200**). Customers don't shop very often. Monetary: Low monetary value. Customers spend little money. Characteristic: These may be customers who buy rarely and spend relatively small amounts, perhaps only during certain periods.

4. **Cluster 3: Recency**: The average Recency value (about **300**), which indicates active customers who have made purchases recently. Frequency: Very high frequency (about **800**). These are customers who make purchases very often. Monetary: Very high monetary value (about 5-6 million). Customers spend large amounts. Characteristic: These are probably the most valuable customers who often make purchases and spend significant amounts. This may be a key segment for the business.

5. **Cluster 4: Recency:** High Recency value (about **400**). Customers have not made purchases for a long time. Frequency: Low frequency (about **300**). Customers don't shop often. Monetary: Low monetary value. Customers spend relatively small amounts. Characteristic: These may be customers who are not active, but at the same time can make periodic purchases with moderate amounts.

The Classical RFM

Champions (14.0% $ - . **Champions (6.2%))**

- **Characteristics**: These are the best customers. They have made recent, frequent purchases, and contribute significantly to the overall revenue.
- **Monetary Significance**: This group contributes a **high proportion (14%)** of the total revenue, making them critical to the business's financial health.
- **Strategy**: **Retain and engage** these customers with exclusive rewards, early access to products, or loyalty perks to keep them engaged and spending.

Loyal Customers (25.9%$ -(9.4%))

- **Characteristics**: Customers who regularly buy from the company and have established a strong relationship with the brand.
- **Monetary Significance**: This is the **largest segment by revenue (25.9%)**, meaning these customers are crucial to maintaining stable revenue streams.
- **Strategy**: **Maintain their loyalty** through consistent engagement, personalized offers, and loyalty programs to maximize lifetime value.

Can't Lose (16.3%$-14)

- **Characteristics**: Previously high-value customers who have become inactive. Their purchasing frequency has dropped, but they once contributed significantly.
- **Monetary Significance**: This segment contributes a **substantial 16.3% of total revenue** despite being at risk of churn.
- **Strategy**: Focus on **win-back campaigns**, targeted incentives, or premium offers to revive their interest and prevent revenue loss.

At Risk (13.0%$-14)

- **Characteristics**: Customers who haven't made a recent purchase and are at high risk of leaving altogether.
- **Monetary Significance**: This group still contributes **13% of total revenue**, so re-engagement is important to avoid further decline.
- **Strategy**: **Reactivation campaigns**, personalized communication, or special discounts can help re-engage these customers and prevent churn.

Potential Loyalists (12.7%)

- **Characteristics**: These customers have made a few purchases and show signs of increasing engagement, potentially becoming loyal.
- **Monetary Significance**: This segment is responsible for **12.7% of revenue**, suggesting strong potential for growth.
- **Strategy**: **Encourage loyalty** through personalized engagement, upsell opportunities, and incentives that increase their purchasing frequency

Hibernating (7.7%-13.7) -

- **Characteristics**: Customers who haven't purchased in a long time, with very low activity levels.
- **Monetary Significance**: Contributing **7.7% of total revenue**, this group represents a smaller portion but still has potential to re-engage.
- **Strategy**: **Reactivation efforts**, such as seasonal promotions or limited-time offers, could bring some of these customers back to making purchases.

About to Sleep (3.8%$- 10)

- **Characteristics**: Customers who have not made a purchase recently and are close to becoming inactive.
- **Monetary Significance**: They account for **3.8% of revenue**, indicating a need for timely re-engagement.
- **Strategy**: **Prevent them from becoming inactive** by offering time-sensitive discounts or reactivation campaigns.

Need Attention (2.6%$-10.6)

- **Characteristics**: Customers who have not been very active lately but were once engaged. They have not made recent purchases and need to be re-engaged.
- **Monetary Significance**: They contribute **2.6% of revenue**, indicating their purchasing activity has significantly dropped.
- **Strategy**: Use **personalized communication** to win back their attention through special offers, reminders, or loyalty incentives.

New Customers (2.2%$- 6.3)

- **Characteristics**: These customers have recently made their first purchase, showing early signs of engagement.
- **Monetary Significance**: While they only account for **2.2% of revenue**, they represent future growth potential.
- **Strategy**: **Onboarding and follow-up** strategies are critical here, such as offering tailored promotions or product recommendations to encourage repeat purchases.

10. Promising (1.7%$ - 8.5)

- **Characteristics**: Customers who have started engaging but are not yet frequent purchasers. They show potential but need nurturing.
- **Monetary Significance**: This segment contributes the least to revenue at **1.7%**, but shows room for growth.
- **Strategy**: **Nurture** this group by offering personalized product recommendations, upsell opportunities, and engagement strategies to increase their spending and loyalty.

Task #5

Choose the KPI ("metrics of success") for each segment

Kpi for Classical RFM

Champions (14.0% of revenue)

- **KPI: Customer Lifetime Value (CLV)**
    - **Why:** Champions are the most loyal and high-spending customers. Maximizing their lifetime value is crucial to maintaining long-term profitability. Because their **loyalty needs to be supported**. we have to fix their level. Moreover, **Clv it's a long term** - to have this kind if clients not so important to have today as such it's important to have for as long as it could
    - **Goal: Increase CLV** by focusing on repeat purchases, upselling, cross-selling, and personalized offers. Keeping these customers loyal ensures sustained revenue.

. Loyal Customers (25.9% of revenue)

- **KPI: Purchase Frequency**
    - **Why:** Loyal customers contribute significantly to revenue, but increasing how often they make purchases can drive even more growth.
    - **Goal: Increase purchase frequency** through loyalty programs, personalized promotions, and consistent engagement to encourage repeat buying.

Can't Lose (16.3% of revenue)

- **KPI: Churn Rate**
  - **Why:** This segment includes customers who were previously high-value but are now at risk of leaving. Reducing churn rate in this group is critical to preventing revenue loss.
  - **Goal: Decrease churn rate** by implementing win-back campaigns, offering personalized incentives, and addressing reasons for disengagement.

4. At Risk (13.0% of revenue)

- **KPI: Reactivation Rate**
  - **Why:** These customers haven't made a recent purchase and are on the verge of becoming inactive. The focus is on bringing them back.
  - **Goal: Increase reactivation rate** through targeted marketing, personalized offers, and re-engagement campaigns designed to reignite their interest in the brand.

Potential Loyalists (12.7% of revenue)

- **KPI: Average Order Value (AOV)**
  - **Why:** These customers have potential to become loyal but need to increase their purchase value. Growing their average order value will lead to higher revenue from this segment.
  - **Goal: Increase AOV** by promoting premium products, offering bundles, or cross-sell opportunities to encourage higher spending on each purchase.

Hibernating (7.7% of revenue)

- **KPI: Reactivation Rate**
  - **Why:** These customers haven't made a purchase in a long time and are nearly inactive. The goal is to bring them back into the buying cycle.
  - **Goal: Increase reactivation rate** by offering seasonal promotions, exclusive discounts, or personalized re-engagement strategies to get them to purchase again.
  - **About to Sleep (3.8% of revenue)**
- **KPI: Purchase Frequency**
  - **Why:** These customers are at risk of becoming inactive and need to be encouraged to make more frequent purchases before they churn.
  - **Goal: Increase purchase frequency** by offering limited-time promotions, personalized reminders, or loyalty program incentives that encourage more frequent engagement.

Need Attention (2.6% of revenue)

- **KPI: Engagement Rate**
  - **Why:** This segment consists of customers who were once active but need re-engagement. Tracking how often they interact with communications (emails, ads, etc.) can indicate progress toward reactivation.
  - **Goal: Increase engagement rate** through targeted marketing campaigns, personalized messages, and exclusive offers that draw their attention back to the brand.

New Customers (2.2% of revenue)

- **KPI: Repeat Purchase Rate**
  - **Why:** New customers need to be encouraged to make repeat purchases to increase their lifetime value and help them transition to more engaged segments.
  - **Goal: Increase repeat purchase rate** by offering follow-up deals, incentives for the next purchase, and strong onboarding campaigns that ensure customer satisfaction.

Promising (1.7% of revenue)

- **KPI: Conversion Rate**
  - **Why:** These customers show potential but have not fully committed to purchasing regularly. Conversion rate will measure how well they can be moved to loyal customers.
  - **Goal: Increase conversion rate** by nurturing them with personalized offers, product recommendations, and reminders based on their browsing or purchase history.

KPI for predicted clusters

Cluster 0 – Reactivation Rate: Focus on re-engaging inactive customers.

Cluster 1 – Purchase Frequency: Encourage these moderately engaged customers to buy even more frequently.

Cluster 2 – Average Order Value (AOV): Increase spending per transaction to make each purchase more valuable.

Cluster 3 – Customer Lifetime Value (CLV): Nurture high-frequency, high-value customers to maximize their long-term contribution.

Cluster 4 – Churn Rate: Prevent churn by keeping these moderate-value customers engaged before they become inactive.

Task #6

Generate/Test 3-5 hypothesis of possible marketing initiatives

Hypothesis №1 - Increasing **C**ustomer **L**ifetime **V**alue (CLV) for Champions

- **Idea**: Offer Exclusive VIP membership perks (think early access to products) that are personalized.

- How to test?
  - Use of data from the Champions segment
  - Statistical test: paired t-test (if normally distributed)
  - Or a Wilcoxon signed-rank test

- **Hypothesis**: Offering personalized product recommendations will increase the **CLV** of Champions by 10%

- Decision to analyze the CLV before and after the hypothetical campaigns using Champion segment, and test the impact on CLV.

- Results?
  - Wilcoxon: 0.0, null hypothesis rejected
  - Maybe 10% increase was a safe bet

Hypothesis №2 - Reducing Churn for At-Risk(ers)

Idea: Execute targeted win-back campaigns offering

personalized incentives (e.g., discounts, special offers)

based on their past behavior

Hypothesis: Offering a targeted win-back campaign

will reduce churn by 15% in At-Risk customers.

How to test?

- ○ Use of data from the At-Risk segment

- ○ Statistical test: paired t-test (if normally distributed)

- ○ Shapiro-Wilk Test

Results?

- ○ Shapiro-Wilk Test p-value (normality test):

  1.0

- ○ Paired t-test p-value: 0.0

Hypothesis №3 - 20% Increase in Purchase Frequency in Loyal Customers

Idea: Tiered Loyalty Program including bonuses, gifts/giveaways and discounts

Hypothesis: Loyal customers have higher Purchase Frequency than Potential Loyalists.

How to test?

- Use of data from the Champions segment
- Statistical test: Shapiro-Wilk test, t-test (if normally distributed)
- Mann-Whitney U

Results?

- T-statistic: -198.36503606633838
- P-value: 0.0
- Wilcoxon test: p-value = 0.0

Hypothesis №4 - Increasing Average Order Value (AOV) of Potential Loyalists

- **Idea**: Offering product bundles to Potential loyalists to get them interested in buying more (Neural dopamine-based encoding)

- **How to test?**
  - Shapiro-Wilk, paired t-test
  - Wilcoxon signed-rank test

- **Hypothesis**: Offering product bundles will increase the Average Order Value (AOV) of Potential Loyalists by 15%.

- **Results:**
  - T-statistic: -67.3591566830833
  - P-value: 0.0
  - Wilcoxon test: p-value = 0.0

Task #7

Marketing strategy.The potential results of suggested strategy

Strategy of Segment 1: Champions (14% of revenue)

Offer
- VIP Banking services
- Cheaper loans, higher % savings, priority customer support

Implement
- Personalized recommendations such as exclusive investment offers

Focus
- On cross-selling premium offerings
- Such as investment products, wealth management

Strategy of Segment 2: Loyal Customers (25.9% of revenue)

Introduce a
- Tiered loyalty program

That offers
- Escalating rewards (e.g., Bronze, Silver, Gold) for increased purchase frequency.

Would finally offer
- Reduced fees or special loan offers

Use
- personalized promotions and reminders to drive repeat purchases.
- Could manifest in discounts based on what the individual purchases

Strategy of Segment 4: Potential Loyalists (12.7% of revenue)

Encourage
- Higher spending **by** offering bundled products
  - For example, offering a discount on a personal loan when customer opens a credit card.

Cross-sell
- Opportunities based on their previous purchases.
  - Such as Higher % on Saving accounts

Evaluating the strategy

Segment 1 Strategy:

- Strengths: Capitalization on existing loyalty and willingness to buy. More likely to respond positively to new offers.
- Weaknesses: Champions may already be using many products/services of bank.
  - It's important to not over-saturate the offerings.
  - There is also risk of neglecting other segments if focus on Champions is too big

Segment 2 Strategy:

- Strengths: More likely to respond positively since they are already engaged. A tiered program can give a sense of progress and can motivate to be involved more.
- Weaknesses: If the rewards are not meaningful enough, segment may not react well enough.
  - At the same time though, profitability should not be harmed by overspending on rewards.

Evaluating the strategy

Strategy 3:

- Strengths: Potential loyalists segment has potential for growth. Cross selling can lead to gains
  - Bundling is good as it simplifies decision-making process
- Weaknesses: Some customers may be hesitating when presented with adoption of a new product

Thank you!