

# RevoBank's Customer Segment Analysis

Satryo Sunu Prabowo

[Colab file](#)

```
const events = [
  'dragenter',
  'dragleave',
  'dragover', // to allow drop
  'drop'
];
events.forEach(e => {
  fileDropZone.addEventListener(e, (ev) => {
    ev.preventDefault();
    if (ev.type === 'dragenter') {
    }
    if (ev.type === 'dragleave') {
    }
    if(ev.type === 'drop') {
      fileDropZone.classList.remove('solid-border');
      handleFiles(ev.dataTransfer.files)
        .then(values => values.map(tag => {
          tag.setAttribute('class', 'border rounded');
          fileDropZone.appendChild(tag)
        }));
    }
  });
});
```

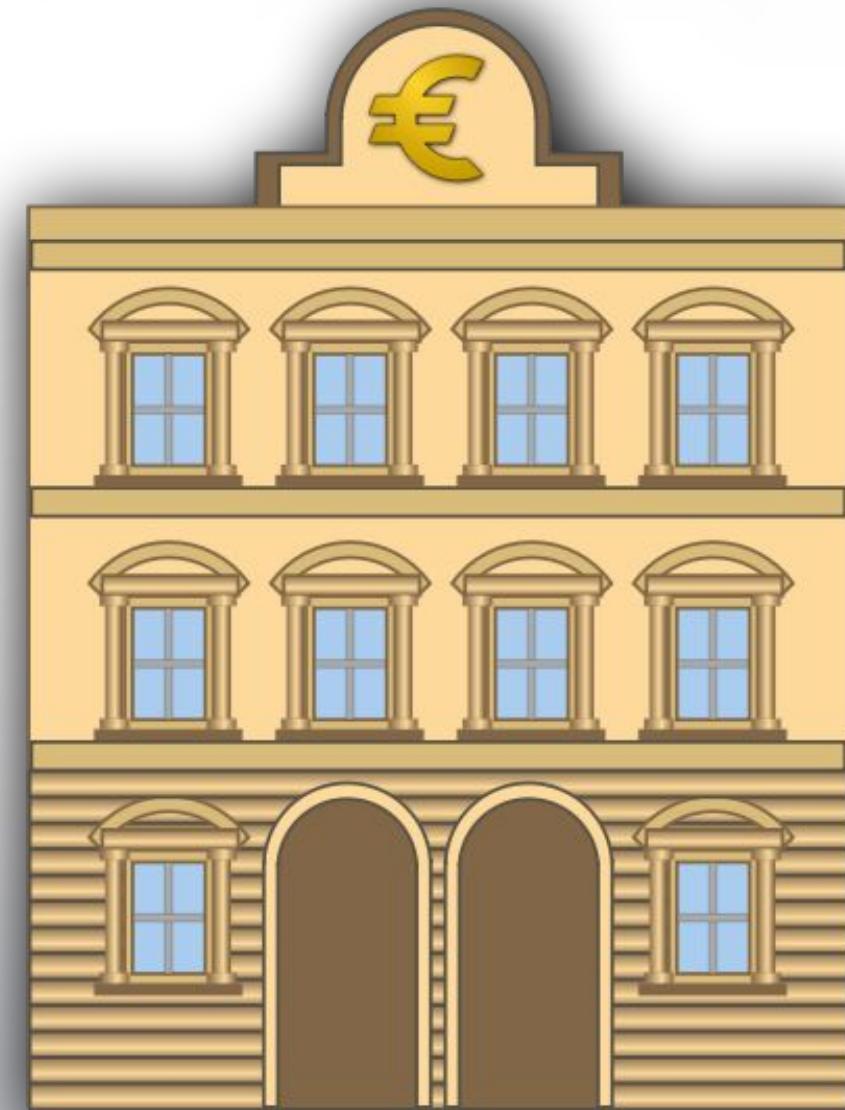
PROBLEMS TERMINAL

```
→ drag-and-drop git:(master) ✘ hs
Starting up http-server, serving .
Available on:
  http://127.0.0.1:8080
  http://10.2.0.73:8080
  Go to the server
  CMT+0200 (CEST)  GET /index.html
  CMT+0200 (CEST)  GET /style.css
  CMT+0200 (CEST)  GET /script.js
```

# Background

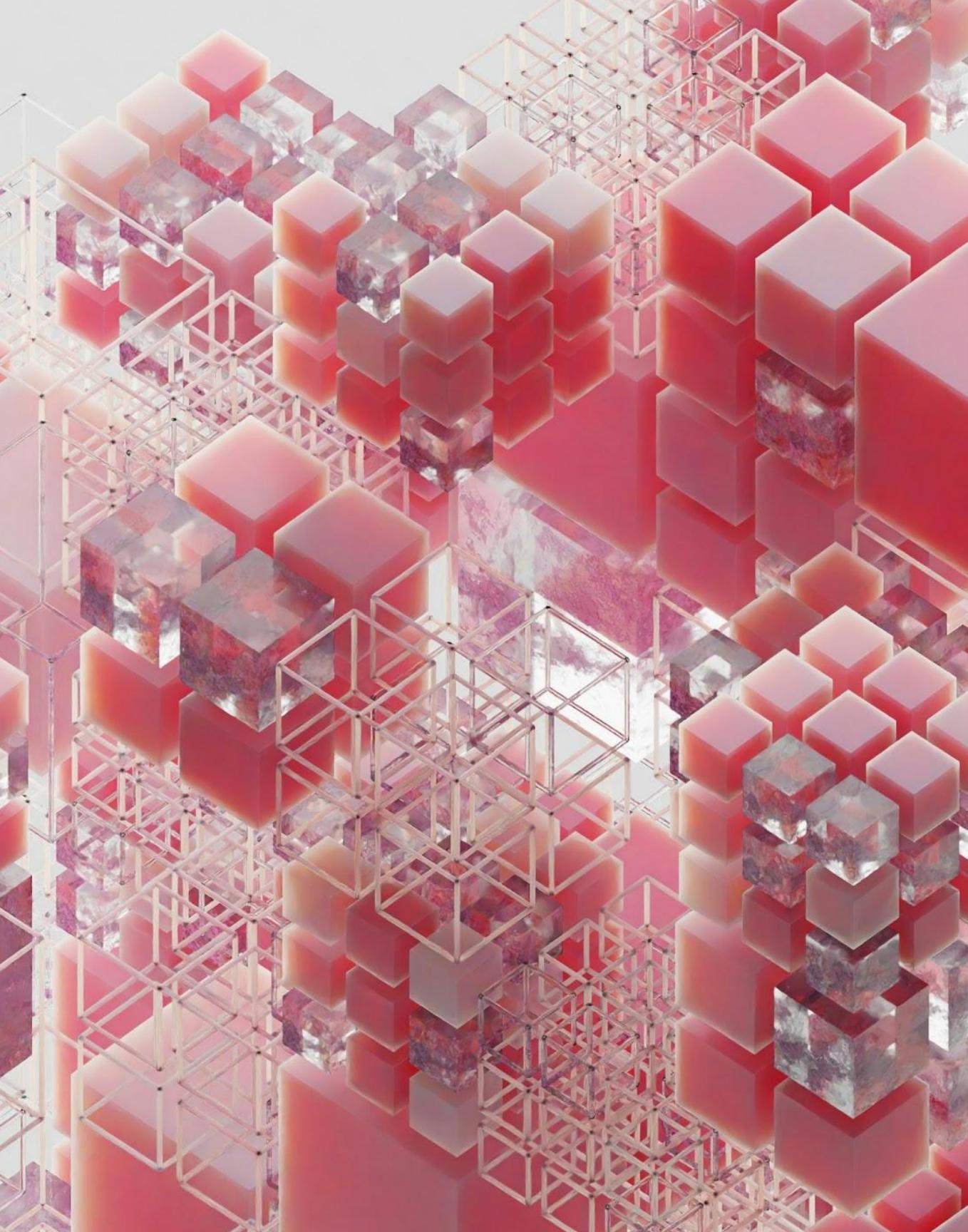
As a Data Analyst at RevoBank, a bank based in Europe, We are currently working with Performance Management (PM) to summarize sales performance over the last 3 years (36 months) to create user personas in order to encourage existing customers to use RevoBank credit card more.

This strategy aims for a 20% increase in credit card usage and a 10% improvement in customer engagement metrics within six months.



# Objectives

- Analyze sales data to identify which user demographics is contributing more on RevoBank credit card usage.
- Evaluate metrics such as average transaction value, total sales volume, and average frequency usage.
- Segment customers based on user demographics, purchasing behavior, and credit card frequency usage.



# Data Cleaning

Upon loading data set, here is the discovery of data that should be cleaned for analysis:



## Check for typo & unique values

There is 'XYZ' in 'account activity level' and 'F' in 'costumer\_value\_level', we will drop the values of 'XYZ' and 'F'.



## Change to correct datatype

Change 'account\_id' datatype from **integer** to **string**. Change 'birth\_date' datatype from **string** to **datetime**.



## Dealing with wrong & missing value

In column 'avg\_sales\_L36M' there is one wrong value of -1000 and 736 datas of null value.

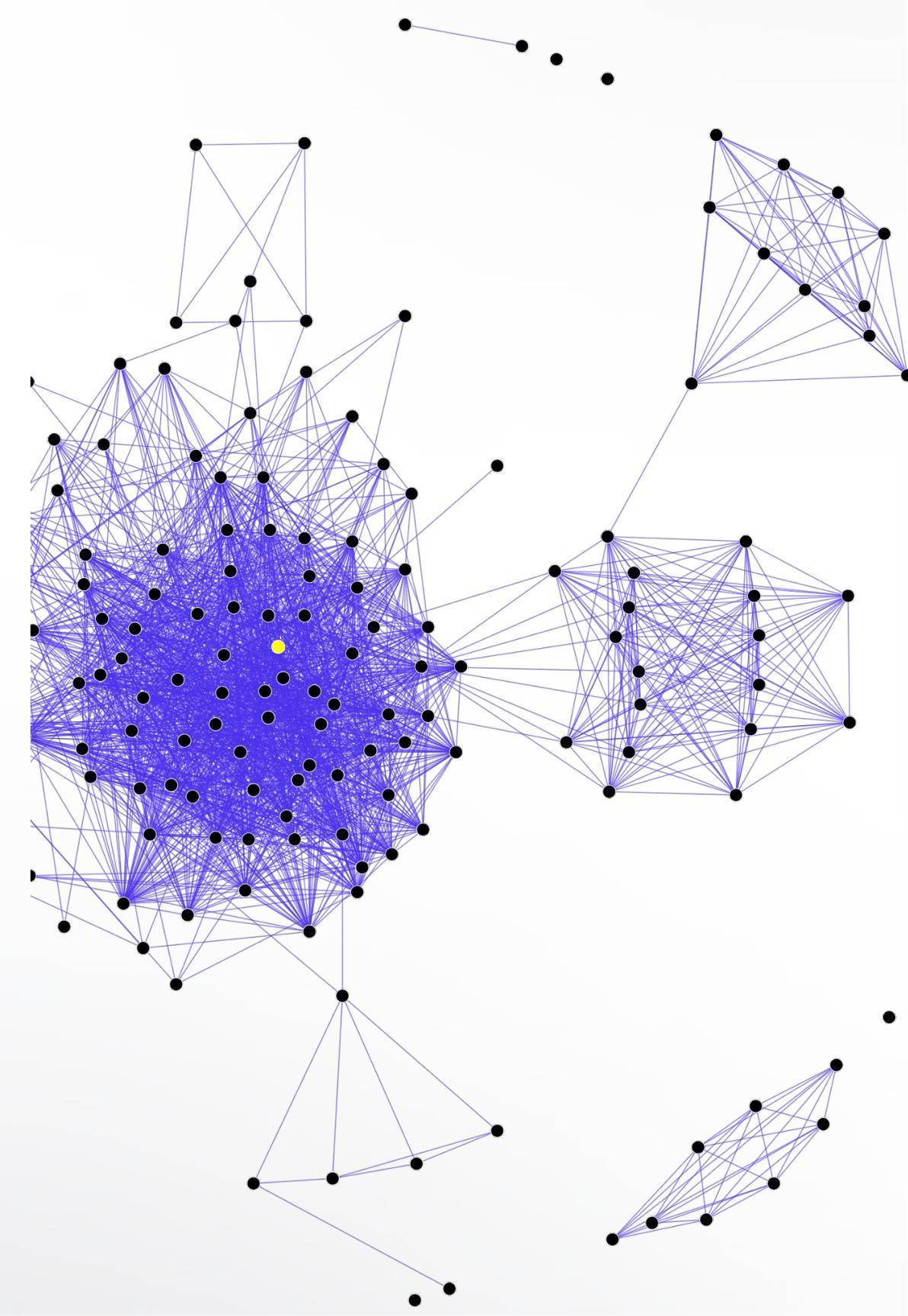
- We drop -1000 value since we don't know what data should we impute for the dataset, hence we drop one row of the data.
- We impute 736 datas of null value with 0. The reason is since count of sales of every null value is 0, basically means there are no sales made. Hence, we impute 0 for all null values in column 'avg\_sales\_L36M'



## Add new column

Added new column for 'age', derived from date of analysis (January 31, 2023) to birth date of users; and 'gender', derived from 'flag\_female' column which 1 is female, 0 is male.

# Exploratory Data Analysis



# Calculate sales data over 3 years

Since we have average data sales and total transactions of sales, we can combine and create new column called '**total\_sales**' to get the amount of total sales over past 3 years of each account.

	account_id	account_activity_level	customer_value_level	MOB	flag_female	avg_sales_L36M	cnt_sales_L36M	last_sales	month_since_last_sales	count_direct_promo_L12M	birth_date	client_age	gender	total_sales
0	100000006	X	A	91.0	0	25000.0	1	25000.0	26	5	1996-11-03	26	Male	25000.0
1	100000038	X	A	69.0	0	9000.0	5	10000.0	25	10	1971-04-24	51	Male	45000.0
2	100000382	X	A	123.0	1	17330.0	6	20000.0	18	6	1953-01-16	70	Female	103980.0
3	100000612	X	A	84.0	1	25830.0	6	25000.0	14	7	1997-06-07	25	Female	154980.0
4	100000757	X	A	40.0	0	25000.0	3	25000.0	17	5	1989-09-13	33	Male	75000.0
5	100000932	X	A	45.0	1	9000.0	4	10000.0	18	6	1971-11-05	51	Female	36000.0
6	100000996	X	A	105.0	1	10000.0	5	10000.0	19	6	1975-01-06	48	Female	50000.0
7	100001201	X	A	41.0	0	25000.0	2	25000.0	24	3	1973-09-04	49	Male	50000.0
8	100001304	X	A	126.0	1	17500.0	2	20000.0	17	6	1947-03-02	75	Female	35000.0
9	100001339	X	A	70.0	1	9800.0	4	10000.0	18	3	1951-06-14	71	Female	39200.0

We accumulate total sales of each account, we got total sales of **401,971,810.00 euro**

# How many percentage for users that has no sales?

```
const events = [
  'dragenter',
  'dragleave',
  'dragover', // to allow drop
  'drop'
];
events.forEach(e => {
  fileDropZone.addEventListener(e, (ev) => {
    ev.preventDefault();
    if (ev.type === 'dragenter') {
      if (ev.type === 'dragenter') {
        if (ev.type === 'dragleave') {
          if (ev.type === 'drop') {
            fileDropZone.classList.remove('solid-border');
            handleFiles(ev.dataTransfer.files)
              .then(values => values.map(tag => {
                tag.setAttribute('class', 'border rounded');
                fileDropZone.appendChild(tag)
              }));
          }
        }
      }
    }
  });
});
```

# Identify No Sales Percentage

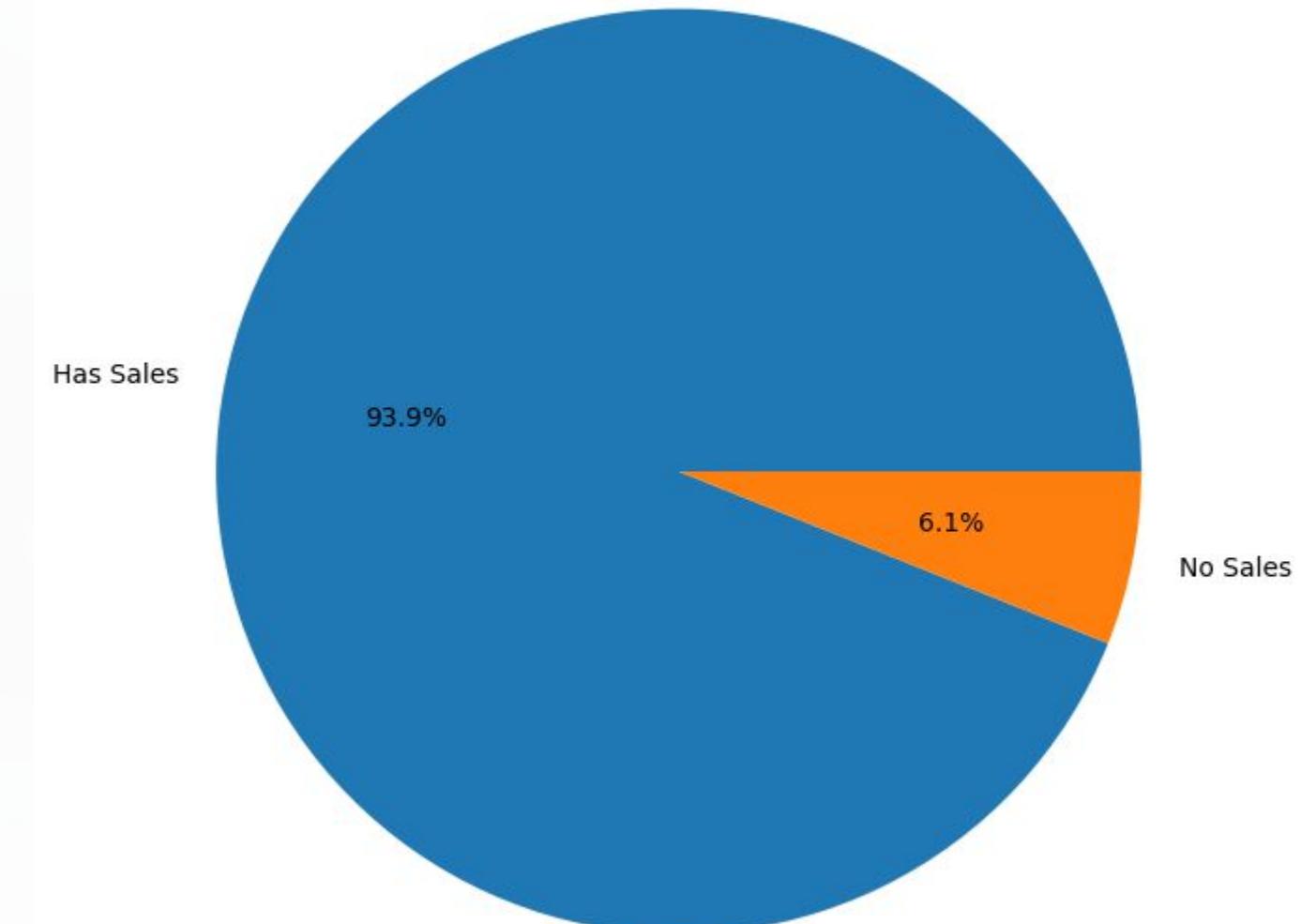
Comparison of Clients with Zero Sales vs Nonzero Sales

Identify the number of clients with no sales in the past 36 months.

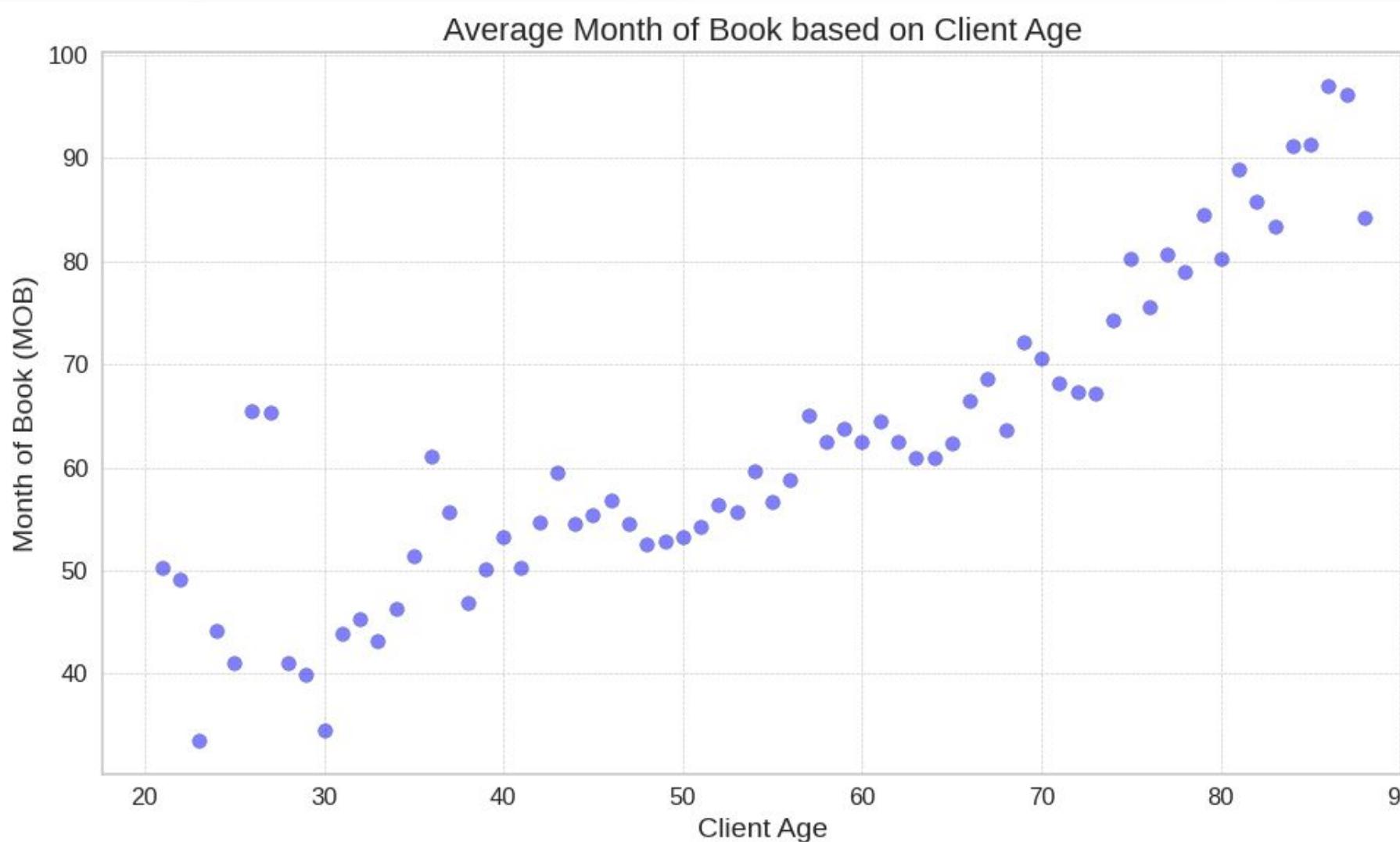
What percentage does this represent compared to the total number of clients?

There is **6.1%** clients that has no sales out of 12,337 entries.

That means around **750 users** has no transactions over the past 3 years on RevoBank. While this segment is relatively small, it highlights that there is a portion of clients who have not made any sales during the period under analysis.



# Compare user's age with average MOB



This scatter plot shows the data of client age to average Month of Book (MOB) of RevoBank. The scatter plot covers a wide range of client ages, from around 21 to 88 years old, providing a comprehensive view of the relationship across different age groups.

As we can see here that **older user** tend to have **longer MOB**, meaning older age **started using loan on early age**. There appears to be a **positive correlation** between **client age** and the **average month of book**. In general, as the client age increases, the average month of book also tends to increase.

# Is it true for Account Activity Level “X” is most active users?

```
ml style.css JS script.js

const events = [
  'dragenter',
  'dragleave',
  'dragover', // to allow drop
  'drop'
];
events.forEach(e => {
  fileDropZone.addEventListener(e, (ev) => {
    ev.preventDefault();
    if (ev.type === 'dragenter') {
      if (ev.type === 'dragleave') {
        if (ev.type === 'drop') {
          fileDropZone.classList.remove('solid-border');
          handleFiles(ev.dataTransfer.files)
            .then(values => values.map(tag => {
              tag.setAttribute('class', 'border rounded');
              fileDropZone.appendChild(tag)
            }));
        }
      }
    }
  });
});
```

PROBLEMS TERMINAL

→ drag-and-drop git:(master) ✘ hs

Starting up http-server, serving ./

Available on:

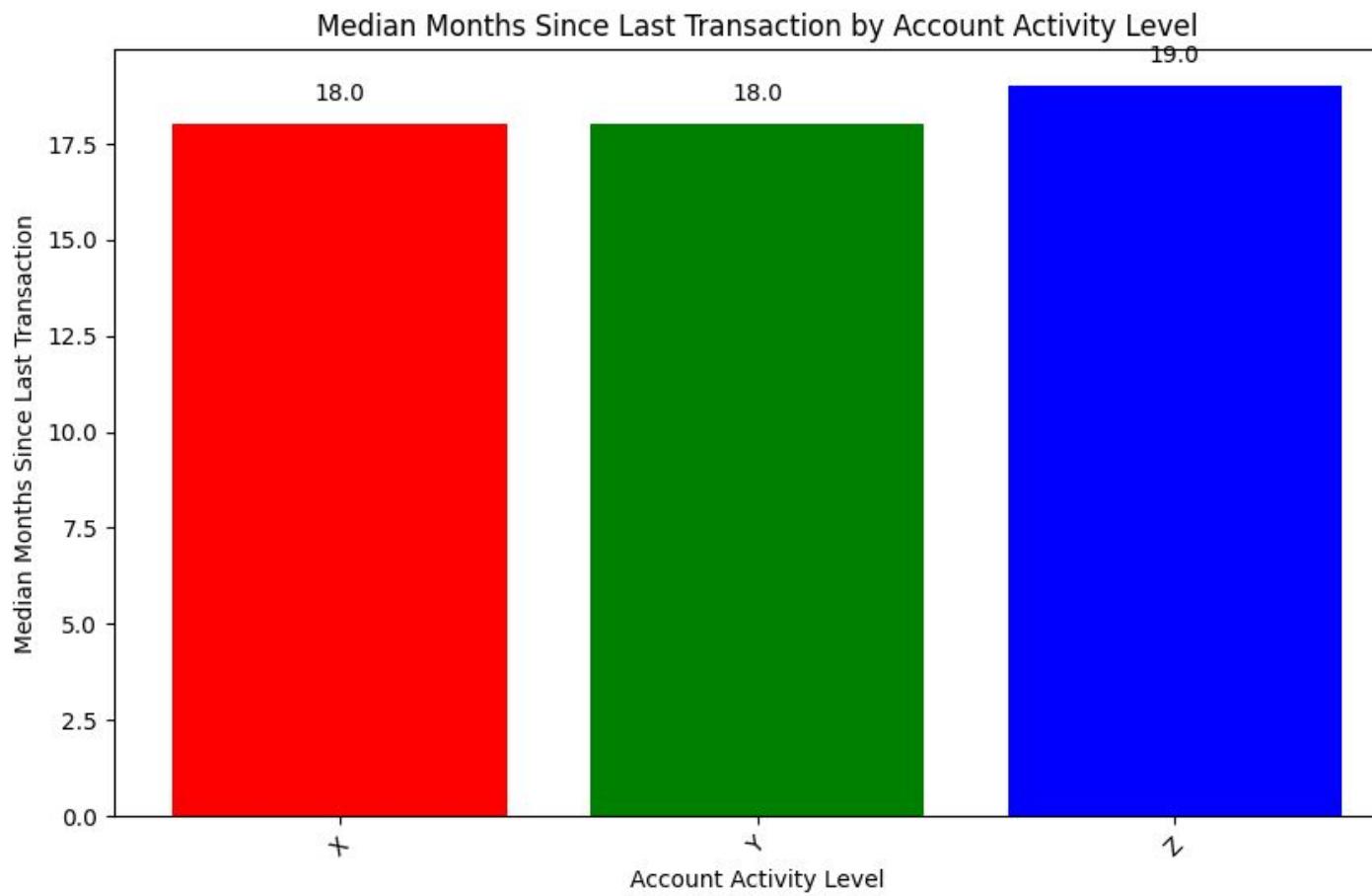
http://127.0.0.1:8080

http://10.2.0.73:8080

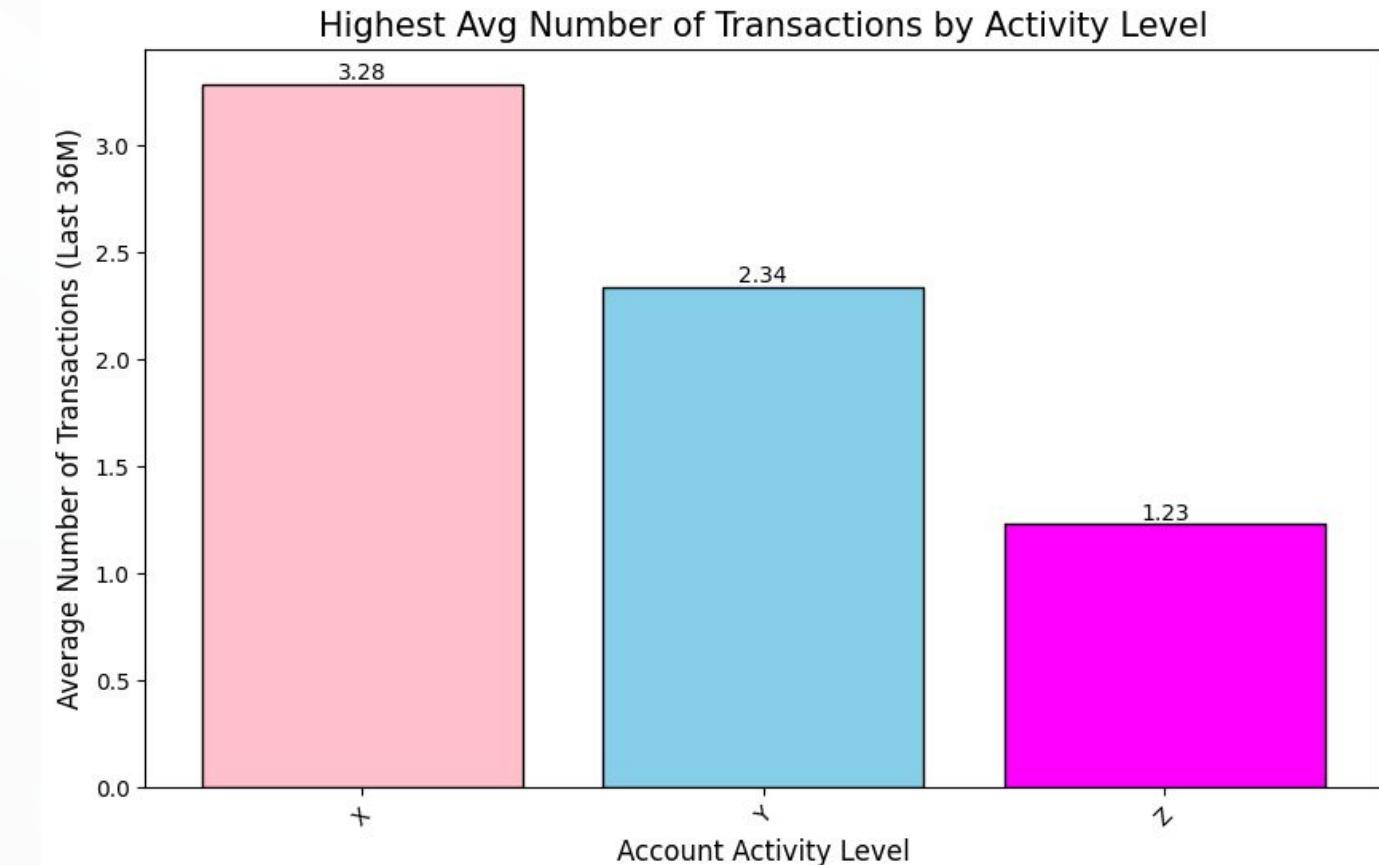
The server

GET /index.html (CEST) -> GET /style.css (CEST)

# Account Activity Level Transaction



While **X and Y** have the **same median value**, which is **18 months since last transaction**, **Z** has relatively have **longer time period** which is **19 months** since last transaction.



**X** have the **highest average number** of transactions, at **3.28 transactions** over the last 36 months. Followed with **Y** at **2.34 transactions** and **Z** only **1.23 transactions** over the last 36 months. This chart suggests that there is a clear correlation between account activity level and the number of transactions. Customers with higher activity levels tend to make more transactions.

# Which gender has more profit?

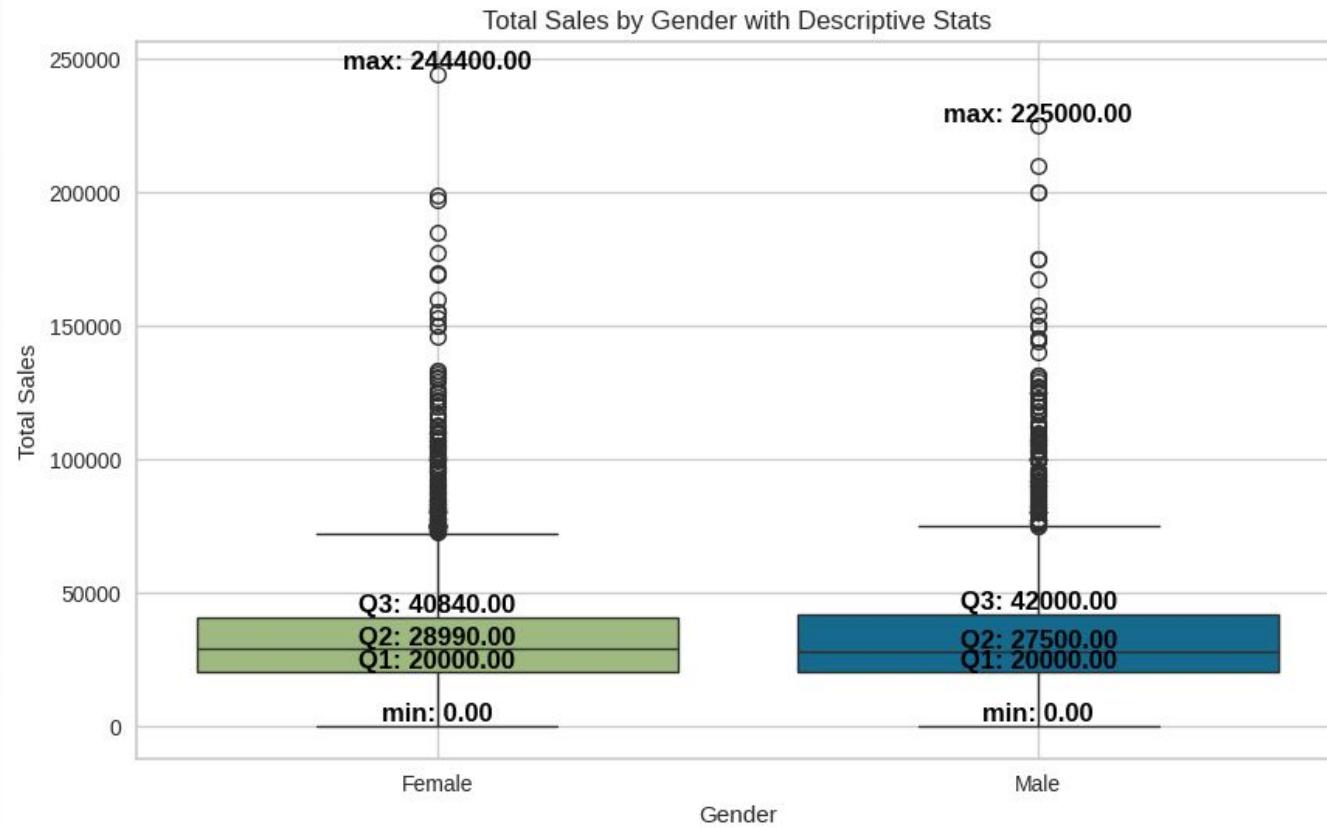
The screenshot shows a code editor window with a dark theme. At the top, there are tabs for 'style.css' and 'script.js'. The 'script.js' tab is active, showing the following code:

```
const events = [
  'dragenter',
  'dragleave',
  'dragover', // to allow drop
  'drop'
];
events.forEach(e => {
  fileDropZone.addEventListener(e, (ev) => {
    ev.preventDefault();
    if (ev.type === 'dragenter') {
    }
    if (ev.type === 'dragleave') {
    }
    if(ev.type === 'drop') {
      fileDropZone.classList.remove('solid-border');
      handleFiles(ev.dataTransfer.files)
        .then(values => values.map(tag => {
          tag.setAttribute('class', 'border rounded');
          fileDropZone.appendChild(tag)
        }));
    }
  });
});
```

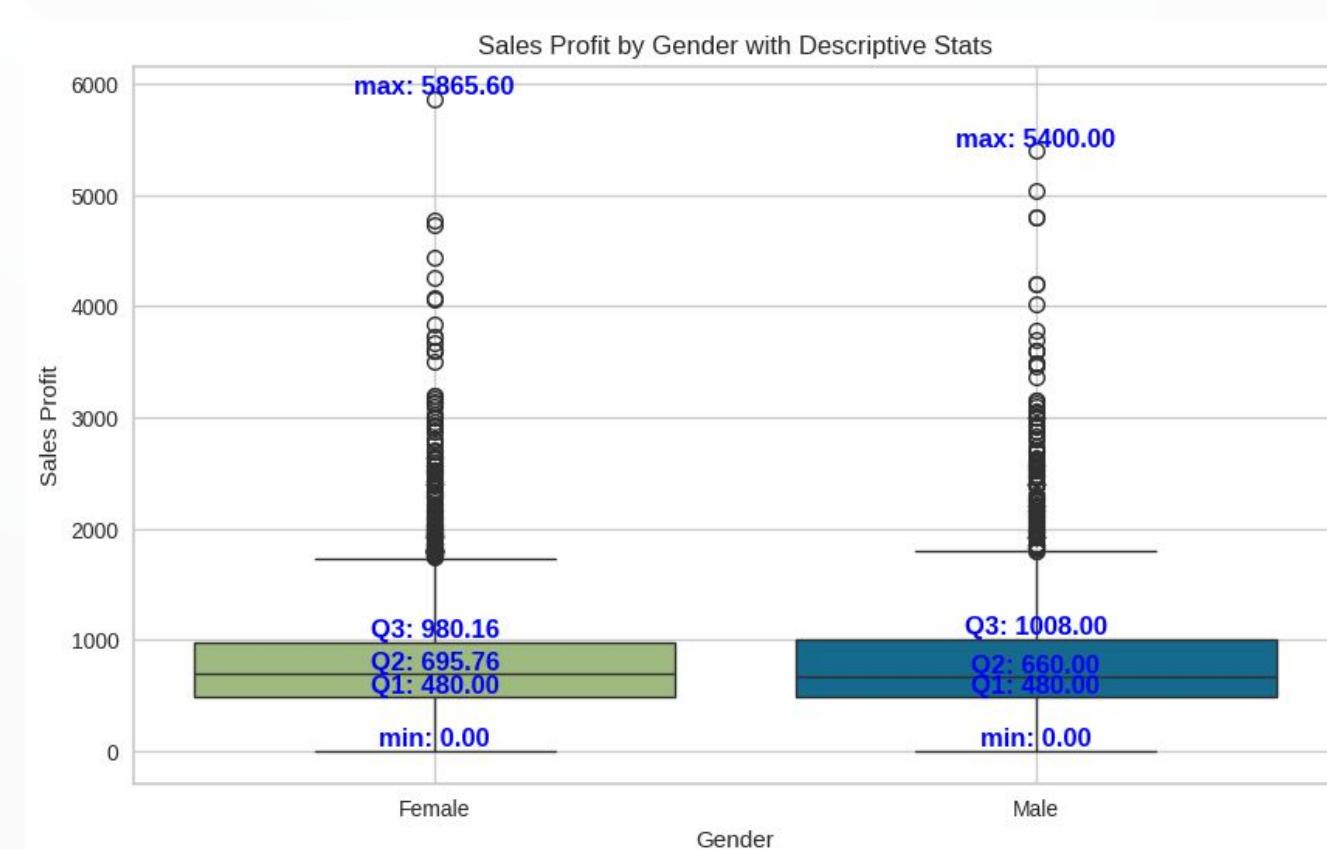
Below the code editor, there is a terminal window with the following output:

```
PROBLEMS TERMINAL
→ drag-and-drop git:(master) ✘ hs
Starting up http-server, serving .
Available on:
  http://127.0.0.1:8080
  http://10.2.0.73:8080
  Go to the server
  CMT+0200 (CEST)  GET /index.html
  CMT+0200 (CEST)  GET /style.css
  CMT+0200 (CEST)  GET /script.js
```

# Profitability by Gender



From the boxplot of total sales earned, **female** users has **more sales** compared to the **male** counterpart, contributing **54% of total sales** for RevoBank with **max sales of 2,444,000 EUR**. However, male users have more average sales compared to female.



Related to sales, in terms of profit, **female** users also have generated **more profit** compared to male users. Notice that the distribution of both sales and profit for female and male user is not normal as there are many outliers for users outside normal distribution.

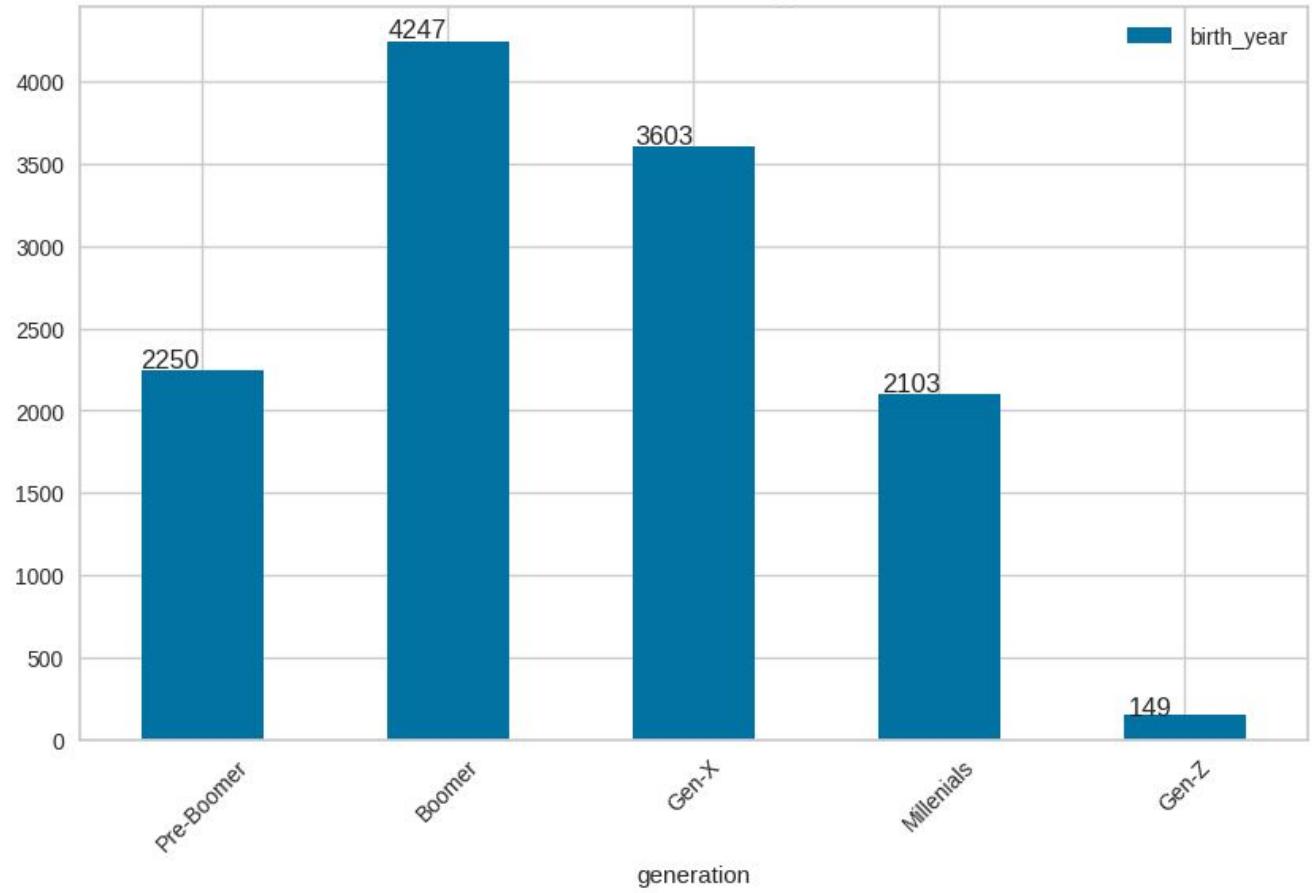
# Which age generation age group has more profit?

```
const events = [
  'dragenter',
  'dragleave',
  'dragover', // to allow drop
  'drop'
];
events.forEach(e => {
  fileDropZone.addEventListener(e, (ev) => {
    ev.preventDefault();
    if (ev.type === 'dragenter') {
    }
    if (ev.type === 'dragleave') {
    }
    if(ev.type === 'drop') {
      fileDropZone.classList.remove('solid-border');
      handleFiles(ev.dataTransfer.files)
        .then(values => values.map(tag => {
          tag.setAttribute('class', 'border rounded');
          fileDropZone.appendChild(tag)
        }));
    }
  });
});
```

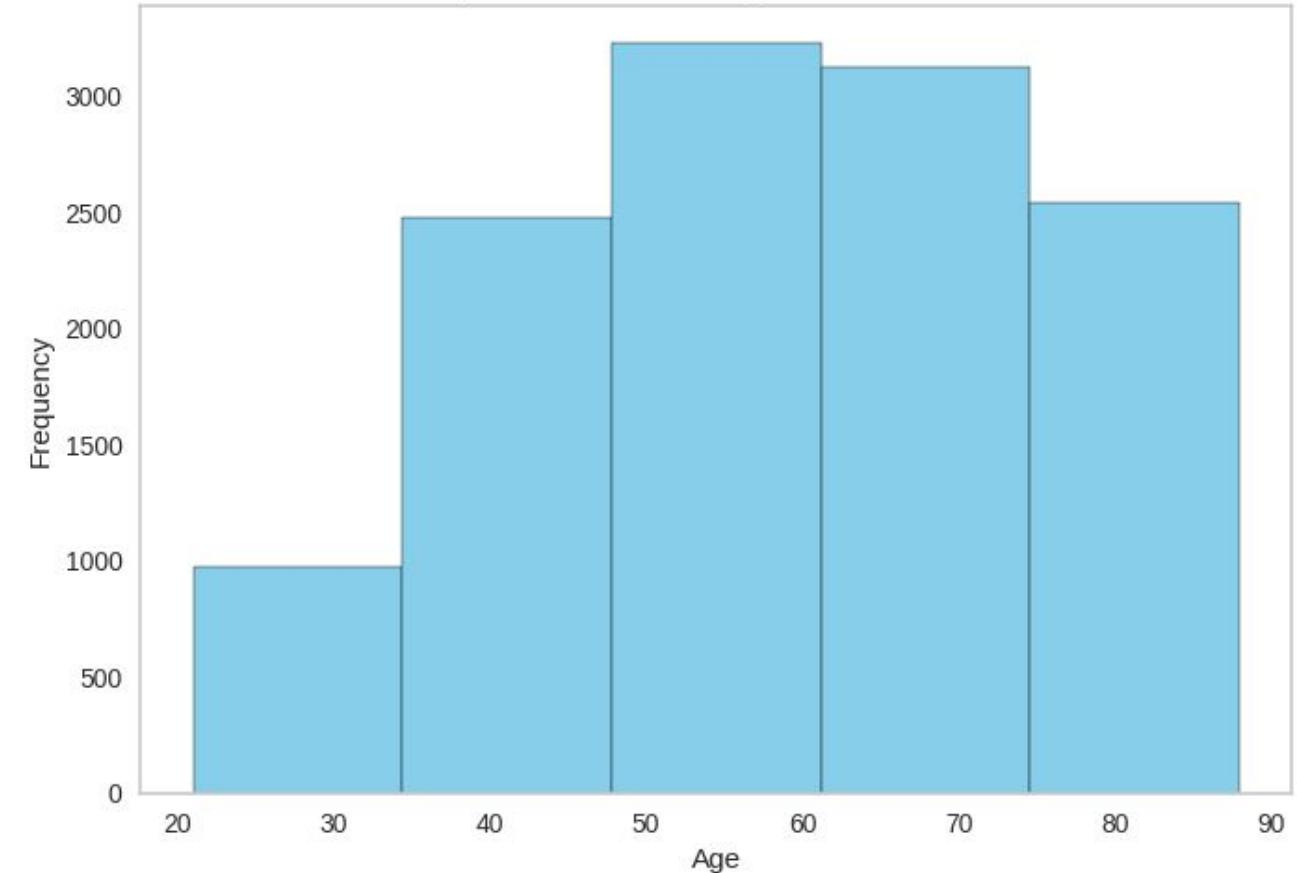
PROBLEMS TERMINAL

```
→ drag-and-drop git:(master) ✘ hs
Starting up http-server, serving .
Available on:
  http://127.0.0.1:8080
  http://10.2.0.73:8080
  Go to the server
  CMT+0200 (CEST)  GET /index.html
  CMT+0200 (CEST)  GET /style.css
  CMT+0200 (CEST)  GET /script.js
```

Total customer based on generation

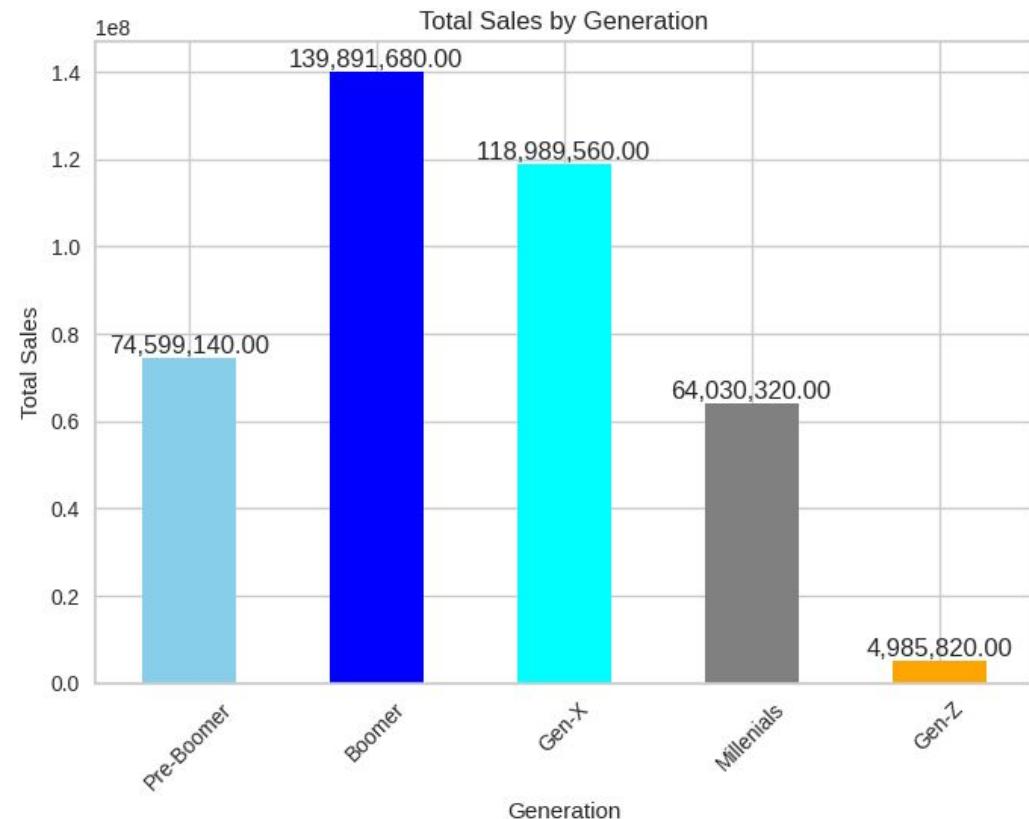
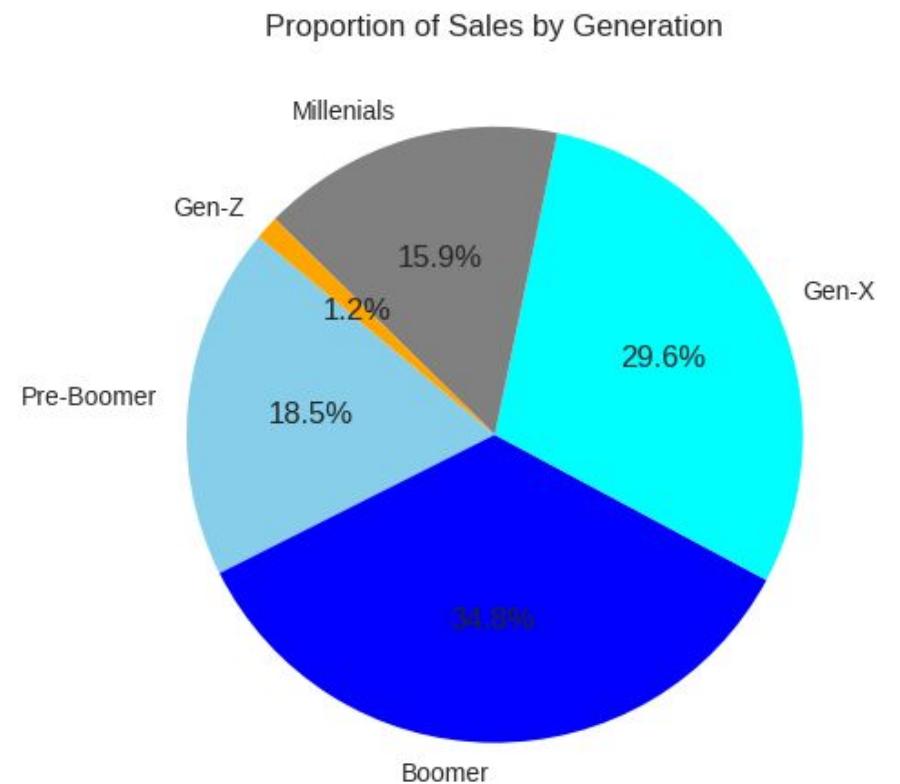


Age Distribution Throughout Generations



Based on the information provided that the **majority of users** are from the **Baby Boomer** generation (born between 1946-1964), and **the second largest segment** is **Generation X** (born between 1965-1980), with the age distribution showing a concentration in the **60s-70s** range.

This is consider unhealthy as time progress, later generations will take up much of the user based. We need to tailor our business needs to cater these later gens.



**Boomer** generation accounts for the **largest proportion** of sales at **34.8%**. Additionally, the bar chart reveals that the Boomer generation generated the **highest total sales** of **\$139,891,680.00**. This indicates that the Boomer demographic is currently the company's most significant revenue driver.

Furthermore, **Gen-X** generated the **second-highest total sales** of **\$118,989,560.00**. This highlights the need to cater to the preferences and needs of Gen-X consumers.

**Millennials** represent **15.9%** of the proportion of sales, indicating a **potential growth** opportunity for the company.

While **Gen-Z** currently accounts for a smaller proportion of sales at **1.2%**, their total sales of **\$4,985,820.00** suggest an emerging market that should not be overlooked. As this generation matures and gains purchasing power, the company should proactively position itself to attract and retain Gen-Z customers.

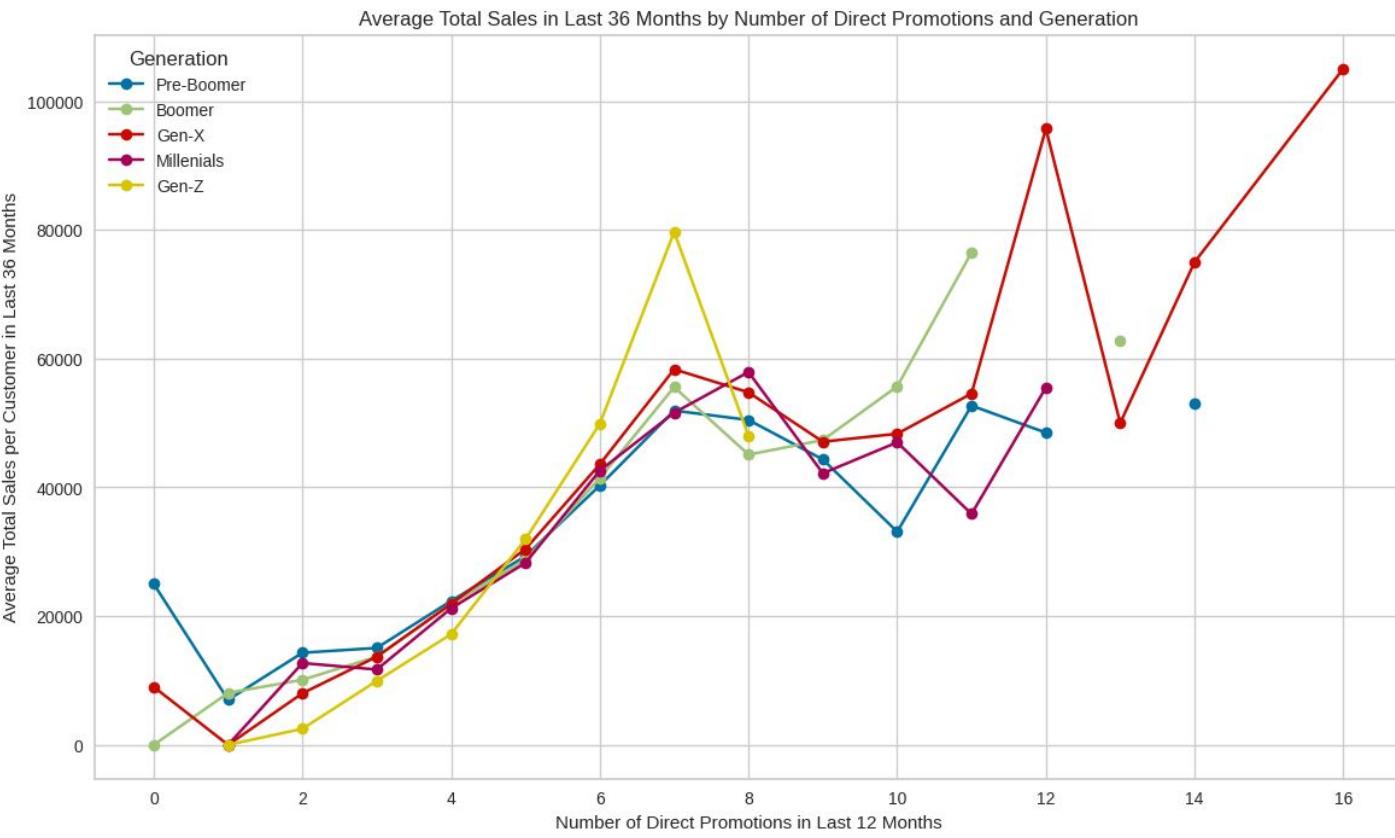
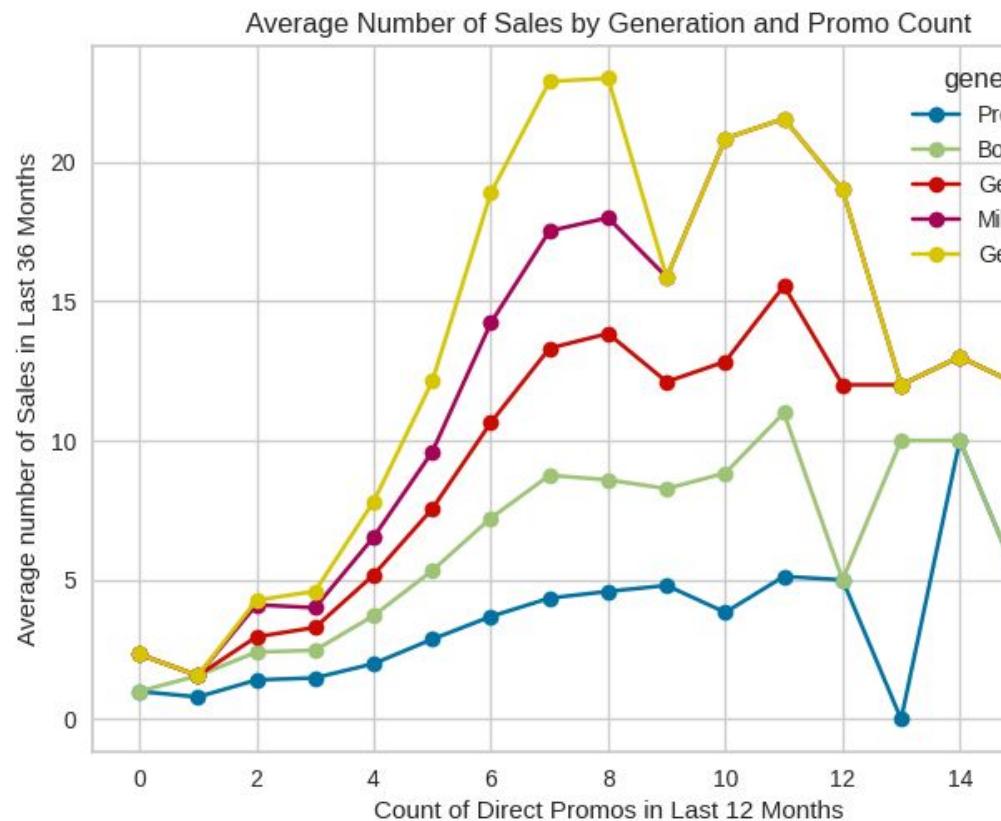


# Which age generation age group has more profit?

```
const events = [
  'dragenter',
  'dragleave',
  'dragover', // to allow drop
  'drop'
];
events.forEach(e => {
  fileDropZone.addEventListener(e, (ev) => {
    ev.preventDefault();
    if (ev.type === 'dragenter') {
    }
    if (ev.type === 'dragleave') {
    }
    if(ev.type === 'drop') {
      fileDropZone.classList.remove('solid-border');
      handleFiles(ev.dataTransfer.files)
        .then(values => values.map(tag => {
          tag.setAttribute('class', 'border rounded');
          fileDropZone.appendChild(tag)
        }));
    }
  });
});
```

PROBLEMS TERMINAL

```
→ drag-and-drop git:(master) ✘ hs
Starting up http-server, serving .
Available on:
  http://127.0.0.1:8080
  http://10.2.0.73:8080
  Go to the server
  CMT+0200 (CEST)  GET /index.html
  CMT+0200 (CEST)  GET /style.css
  CMT+0200 (CEST)  GET /script.js
```



The relationship between **direct promotions** and **sales** is relatively straightforward. There is a clear **upward trend** in both **average total sales** and **average number of sales** as the **number of direct promotions increases**. However, all generations have upwards sales up until **7 promotions given**.

There are peaks and valleys in the average sales figures after 7 promotions, indicating that these groups may be more selective or discerning in their response to promotional offers. Nevertheless, the sales is quite stagnant across generations. **Gen-X** has sales with **max 16 promotions given**, indicating that increasing the frequency of direct promotions could be an effective way to drive higher sales.

While **Gen-Z** contributes low sales compared other generations, Gen-Z has **second highest transaction made**, after Pre-boomer.

We can leverage this behavior to target younger audience, develop tailored promotional offers, that resonate with the unique characteristics of this group.



# Business Recommendation

## Tailor Younger Audience

- Create a **marketing plan** to cater the young generations. Conduct **market research** to identify young adults' spending habits, preferred brands, and financial goals.
- Millennials and Gen Z are digital natives, so it's crucial to have a **strong online and mobile presence**. Optimize the credit card application process for seamless digital onboarding, and create a user-friendly mobile app with features like spend tracking, reward redemption, and payment scheduling.

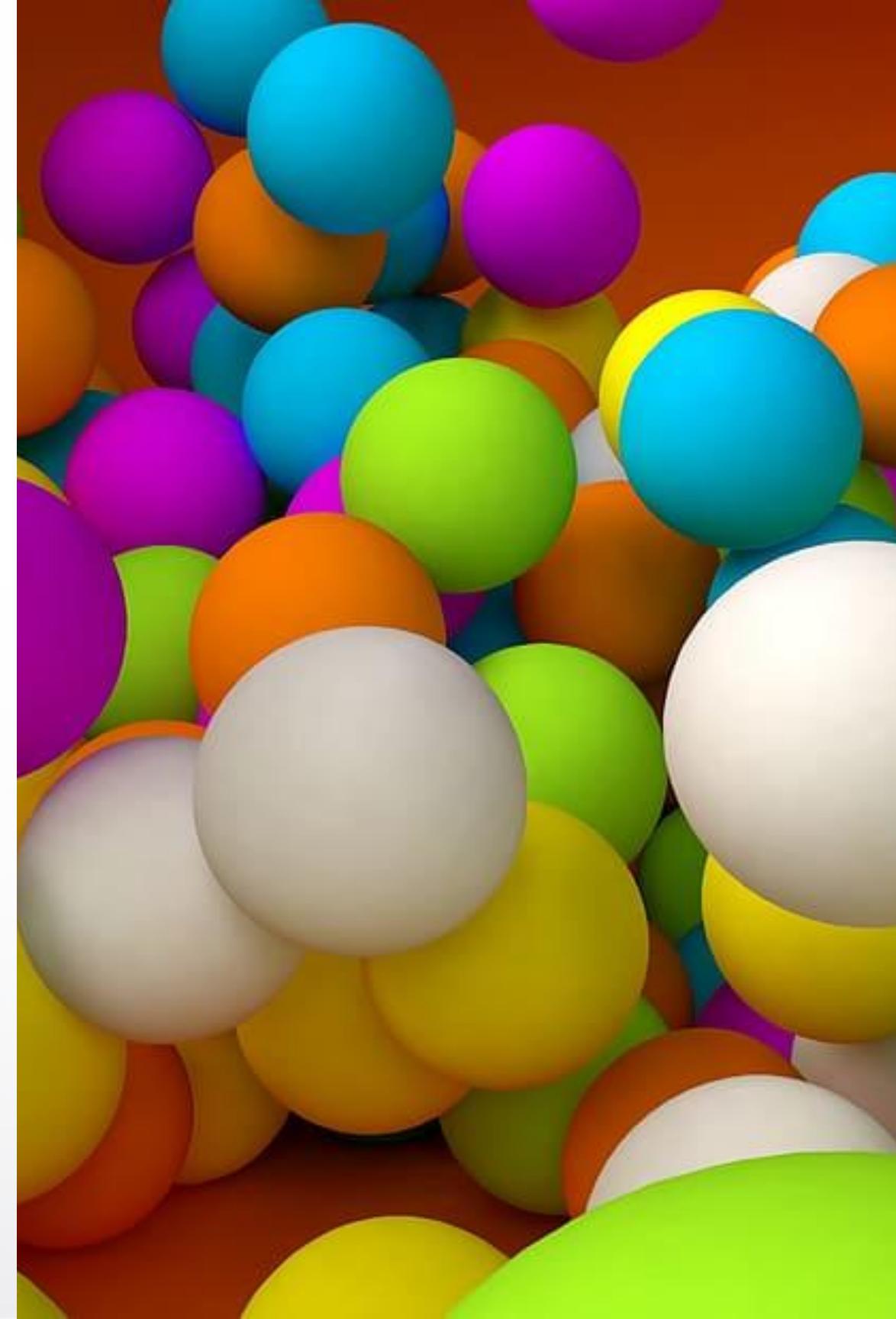
## Revamp Promotions and Rewards

- **Rewards and Cashback Programs:** Offer **attractive rewards and cashback programs** tailored to the interests and spending habits of younger consumers. Partner with popular brands, streaming services, and e-commerce platforms to offer rewards that resonate with this audience.
- Offer **introductory APR promotions** or waived annual fees for the first year to incentivize sign-ups. Provide bonus rewards or cashback for reaching specific spending thresholds within the first few months to encourage immediate usage.

## Upsell Card Benefits

- Incorporate **gamification** elements, such as **point-based challenges** or **virtual badges**, to make credit card usage more engaging. Integrate with social media platforms to allow users to share their rewards progress and achievements with their peers.
- **Focus on premium features:** Upsell the value of additional benefits like travel insurance, extended warranties, airport lounge access, or purchase protection.

# Segmenting & Clustering



# RFM Segmentation

By analyzing these three aspects of customer behavior, we can group customers with similar traits and tailor marketing strategies accordingly. RFM itself stands for **Recency, Frequency, and Monetary**.



## Recency

How **recently** a customer has conducted a transaction with the bank. We will consider this with **months since last transaction**.



## Frequency

How **frequent** a customer has conducted a transaction with the bank. This can be counted with **number of transactions**.



## Monetary

The total **amount of money** a customer brings to the bank. Variable of the analysis for monetary is **total sales in last 36 months**.

# What to consider using RFM

Considering using RFM for segmenting the customer persona, essentially we are looking for a **best segment** for maximizing the value of our business. The criteria has the best for RFM segmentation as follows:

- Customer that has **low recency**, meaning that has low number of months since last transactions.
- Customer that has **high frequency**, meaning that customer has high number of total transactions.
- Customer that **high monetary**, meaning that customer has high amount of sales spent.

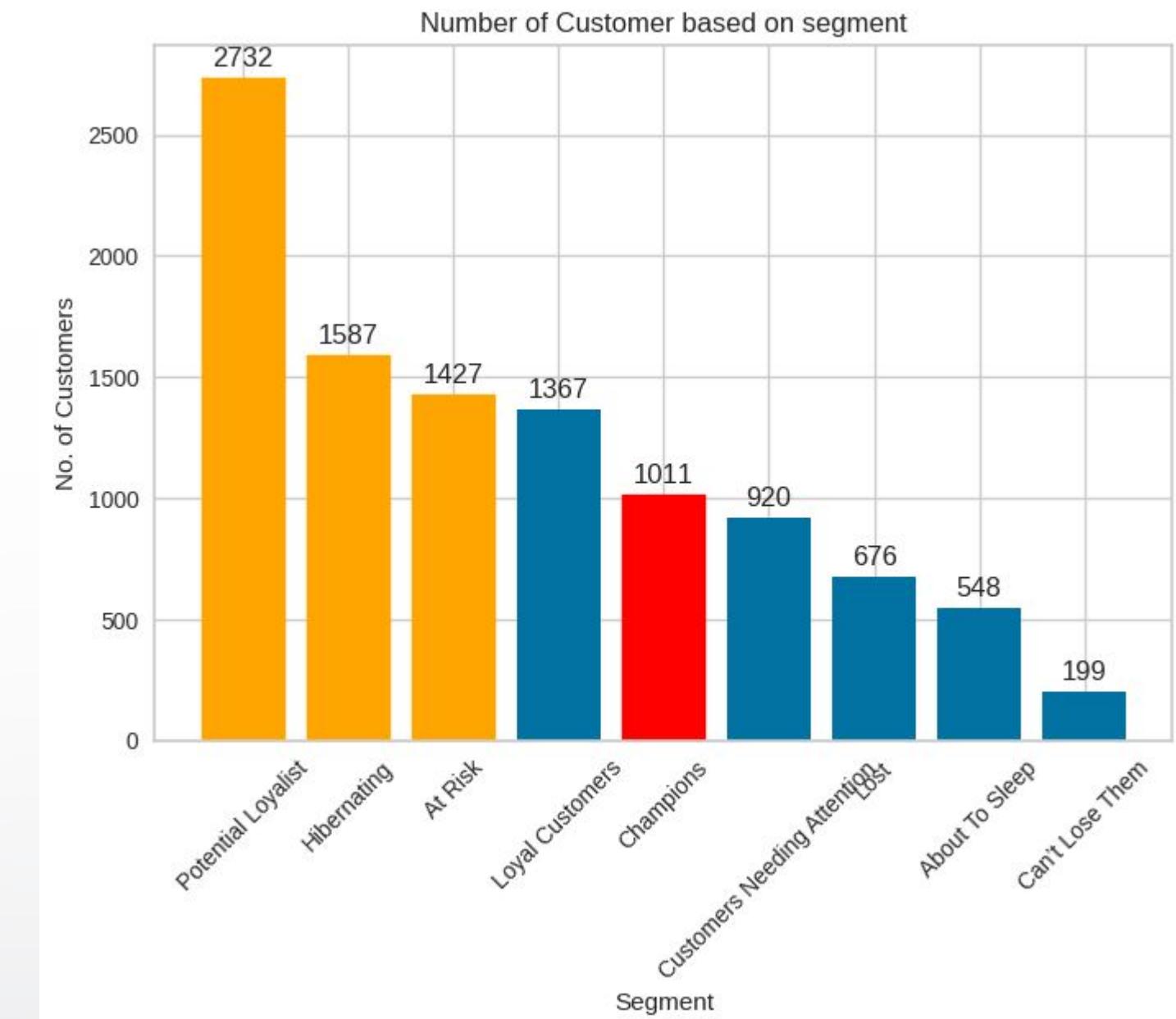


# RFM Scoring

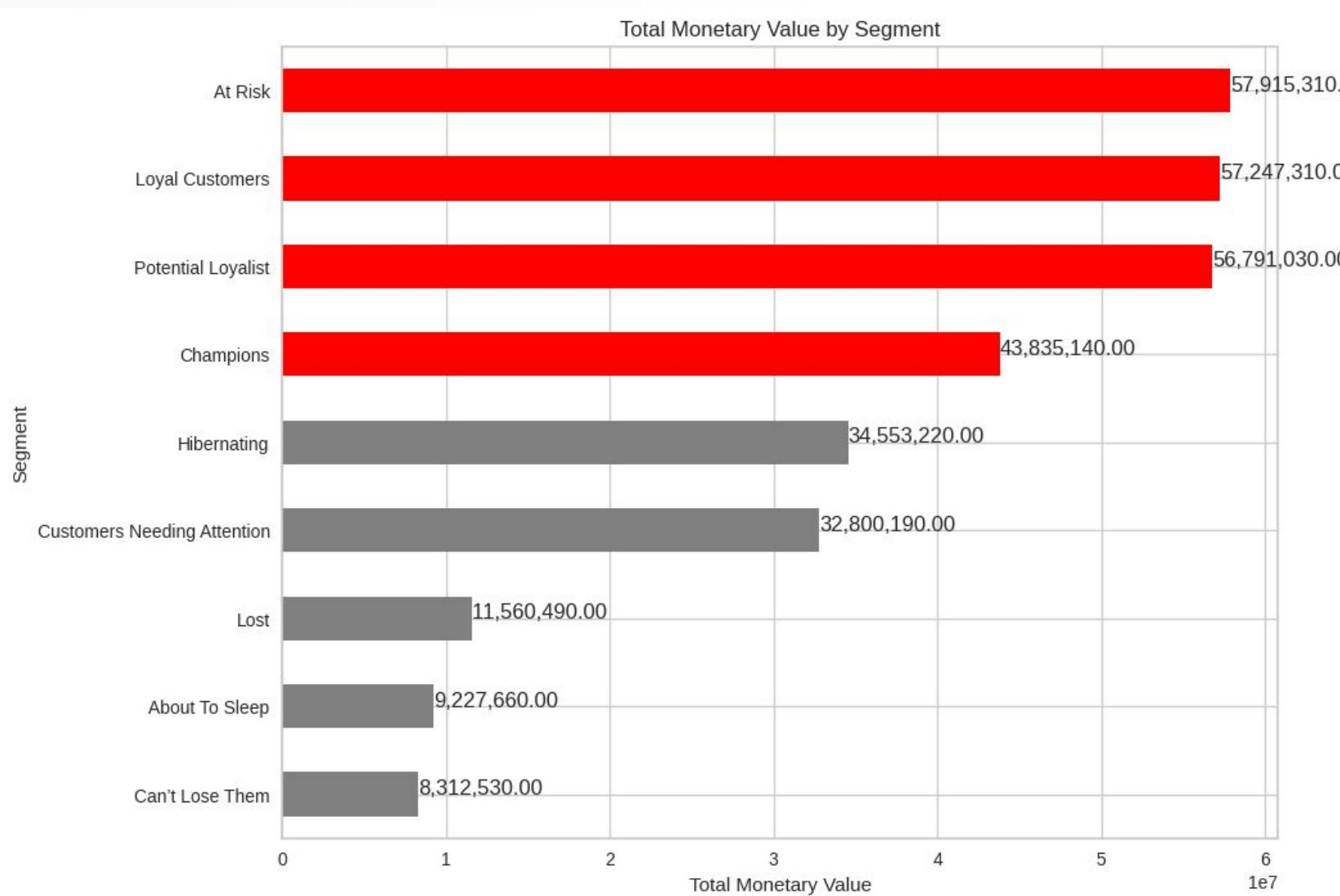
Score	1 (Worst)	2	3	4	5 (Best)
Recency	21-28	19-21	17-19	16-17	3-16
Frequency			1-2	2-4	4-12
Monetary	4,000-20,000	20,000-25,000	25,000-30,000	30,000-40,000	40,000-58,030

We are looking for those 3 criterias (low recency, high frequency, and high monetary), that is **Champions** segment. The bar chart below provides valuable insights into the distribution of customers across different segments.

- While the **potential loyalist** have the highest number of customers of 2732, it's worth noting that the **champions**, with 1011 customers, is not among the top three.
- By not prioritizing the champions segment, the bank might be **missing out on significant revenue and profitability**. Champions tend to be brand advocates and generate more referrals compared to other segments.

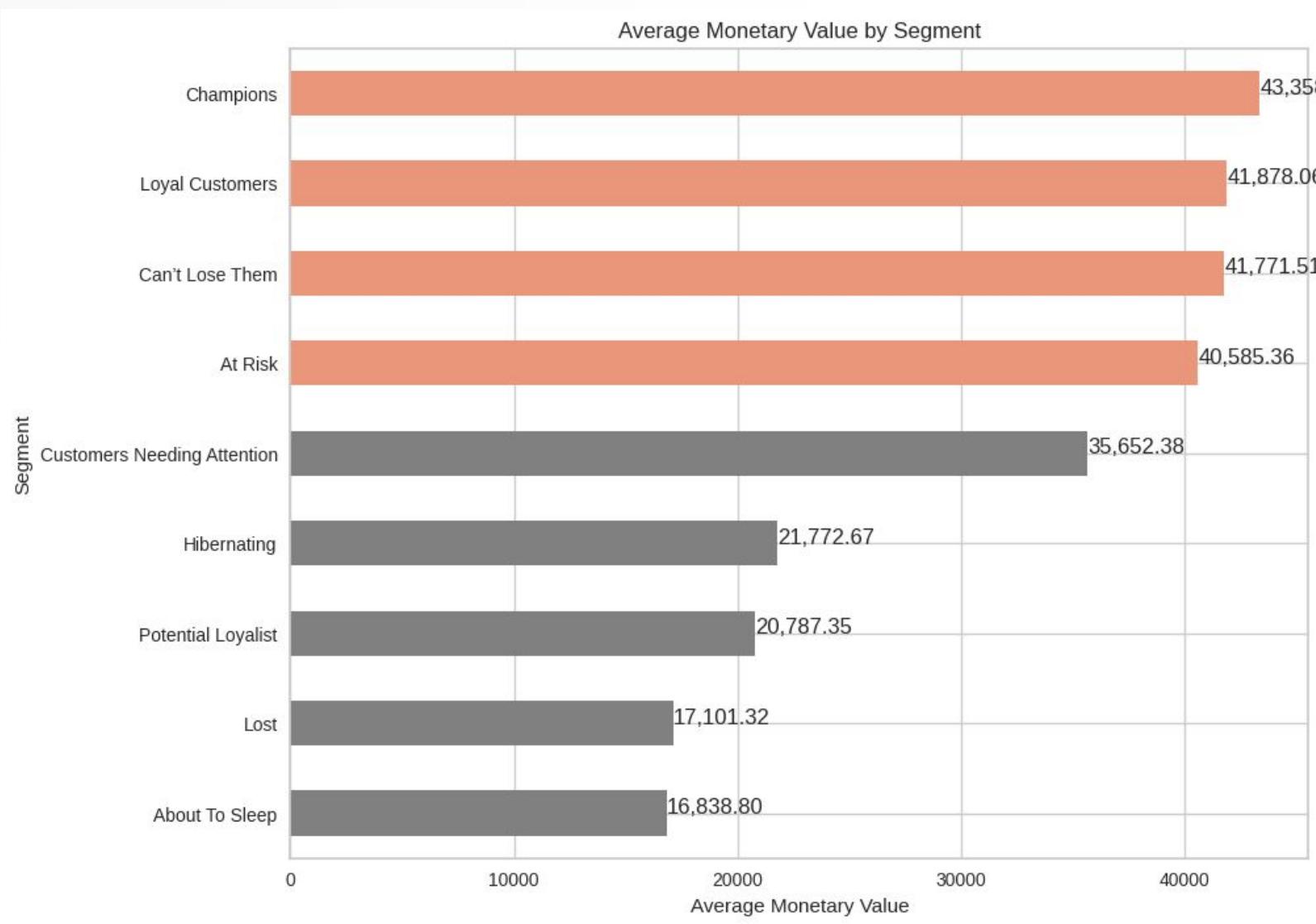


# Total Monetary Value per Segment



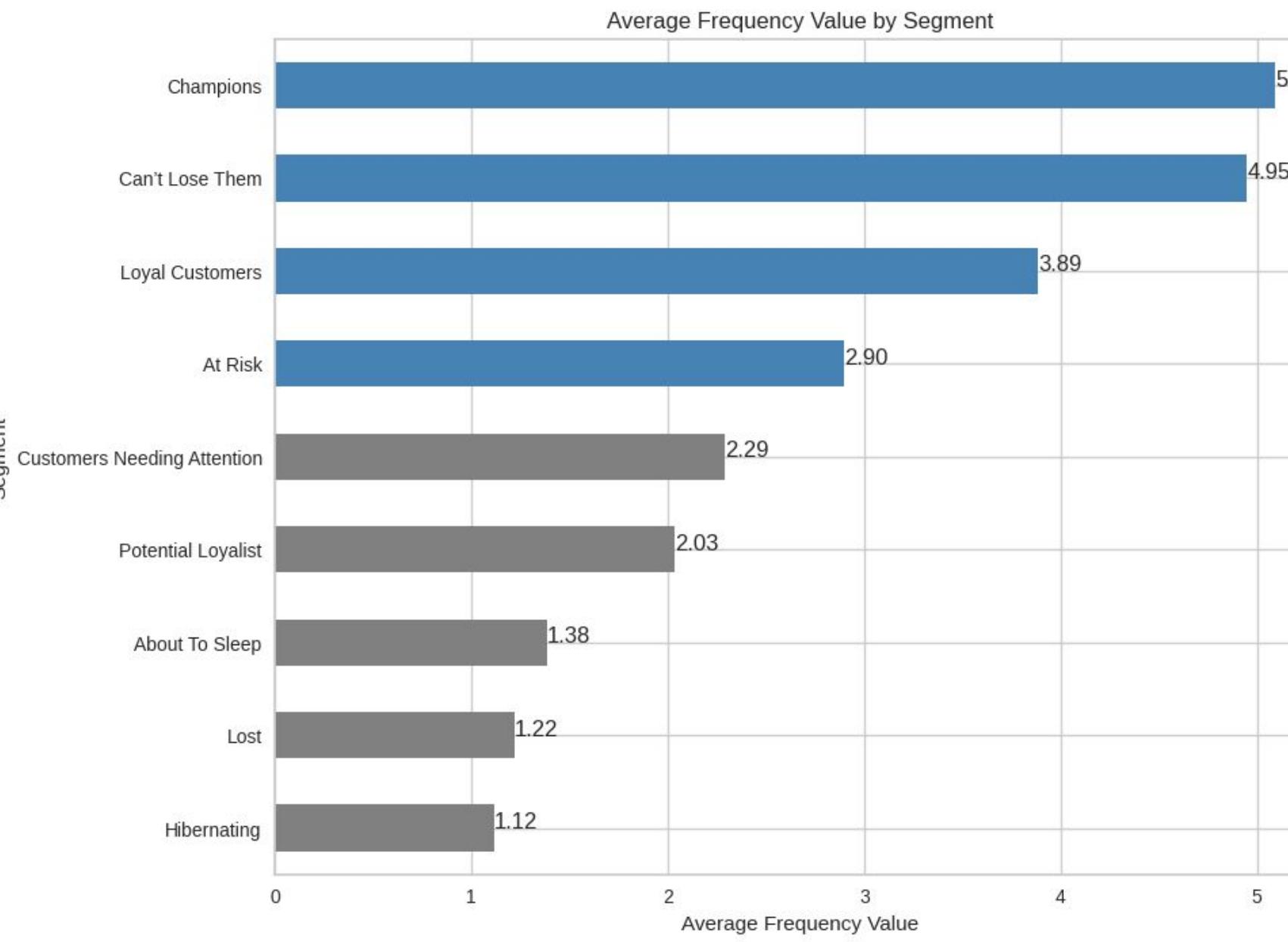
- Even though "**At Risk**" customers contribute a highest amount of money to the bank (57,915,310), they are at risk of churning.
- Segments like **Potential Loyalists** (56,791,030.00) and **Loyal Customers** (\$57,247,310.00) have a high total monetary value, indicating potential for converting them into even more engaged customers.
- Interestingly, **Champions**, the segment typically representing high-value and engaged customers, are not among the top spenders. They contribute 43,835,140.00, ranking fourth in total monetary value.

# Average Monetary Value per Segment



- **Champions** are the top spenders, with an average monthly value of 43,358.20. Even though the total value is not high, average value per transaction is still highest.
- “**Can't Lose Them**” has an average monthly value of 41,771.51, ranking third. While the **segment size** (191) is the **smallest**, their **average transaction value is significant**. This indicates a group of customers with substantial spending power, even though they transact infrequently.
- **Loyal Customer** maintains in second highest average monetary, **At Risk** is considerably high but again this value has potential to attrition.

# Average Frequency Value per Segment



- Despite not being the most frequent transactors (average frequency value of 4.95), **Can't Lose Them** customers still rank high. This suggests that while their transaction frequency may be lower, they are still relatively engaged with the bank.
- While the **At Risk** has a high average monetary value (as discussed previously), their average frequency is somewhat lower. This indicates a high level of churn as previously stated in the monetary analysis.
- Loyal customers** also show a healthy average frequency of 3.89, indicating consistent interaction with the bank. They might be a good target group for retention efforts through loyalty programs or personalized communications.

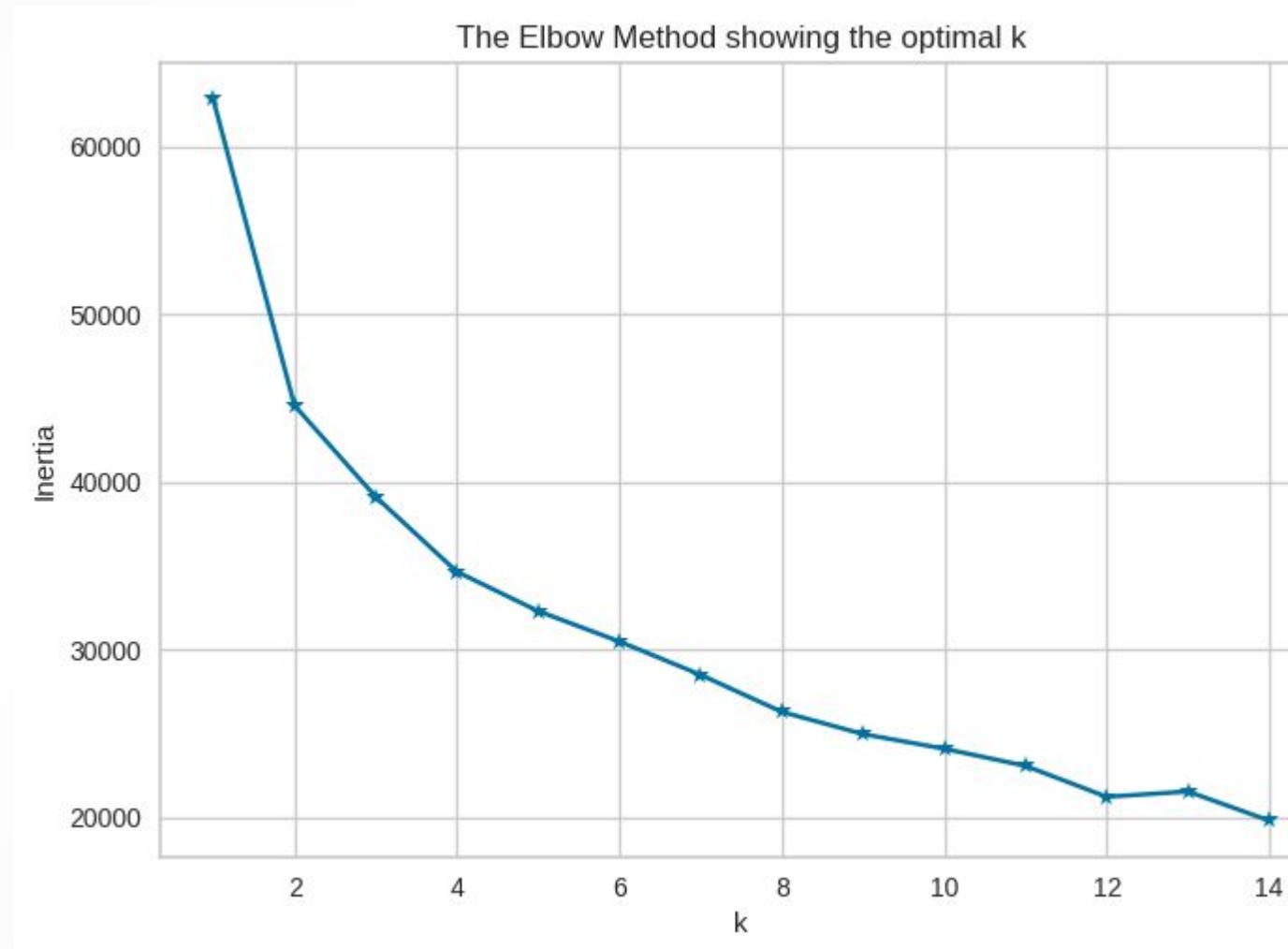
# K-means Clustering

Aside from RFM segmentation, K-means clustering also a valid method to determine customer behavior. K-means is able to overcome some of the limitation of RFM by automatically grouping customers based on various features (can include RFM factors and more). Some of the parameters chosen to determine the cluster of the customers of RevoBank as follow:

- Client age
- MOB
- Average sales
- Number of sales (transactions)
- Month since last sales
- Number of promotion received

# Clustering Analysis

There are 2 methods that allows for determining number of clusters (k), Elbow Method and Silhouette Method. First, let see the elbow method looks like.



Elbow point is used as a **cutoff point** in mathematical optimization to decide at which point the diminishing returns are no longer worth the additional cost. Since it's quite hard to determine the break point, let use Silhouette Method

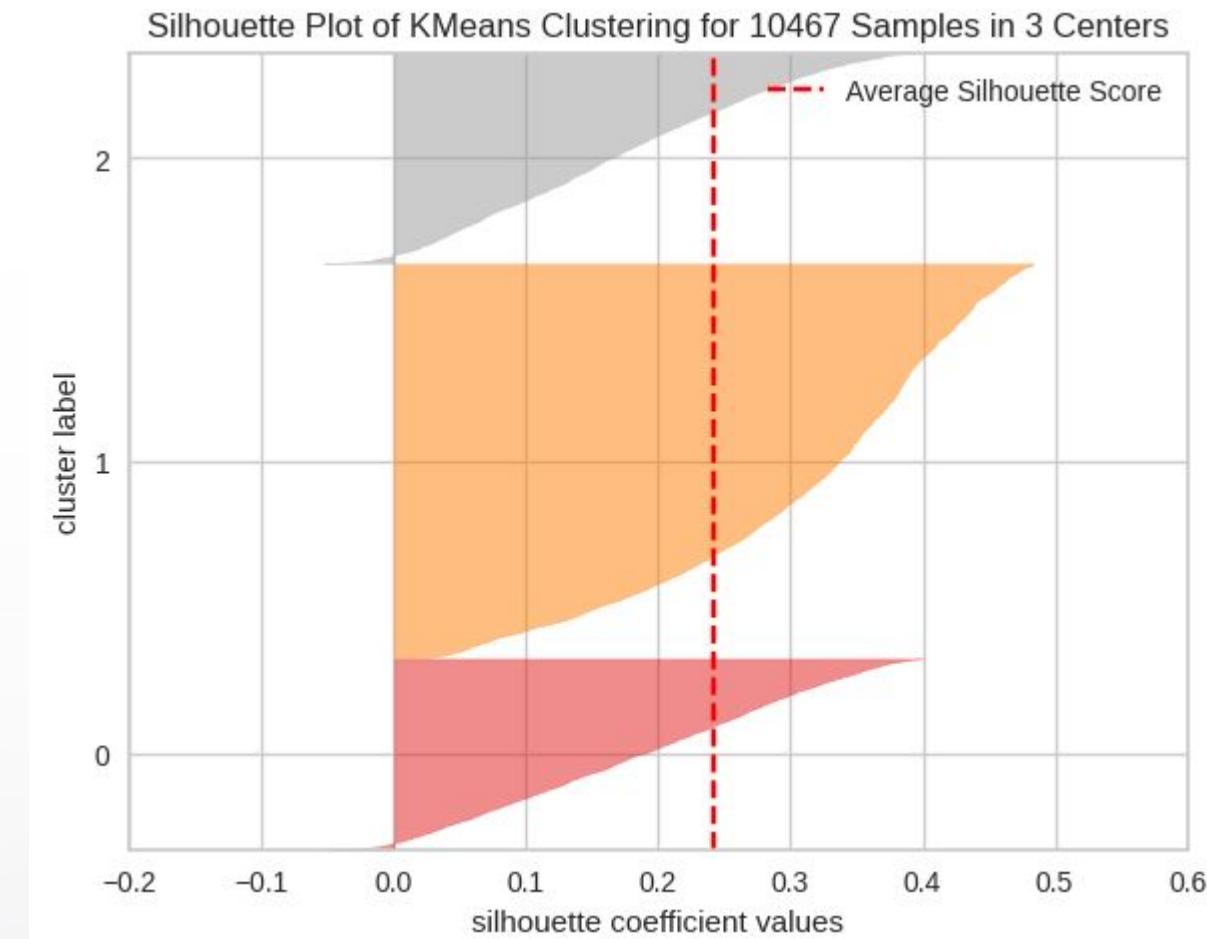


# Clustering Analysis

Silhouette is better since I'm able to see the average silhouette score and see which number of clusters that has least negative silhouette score.

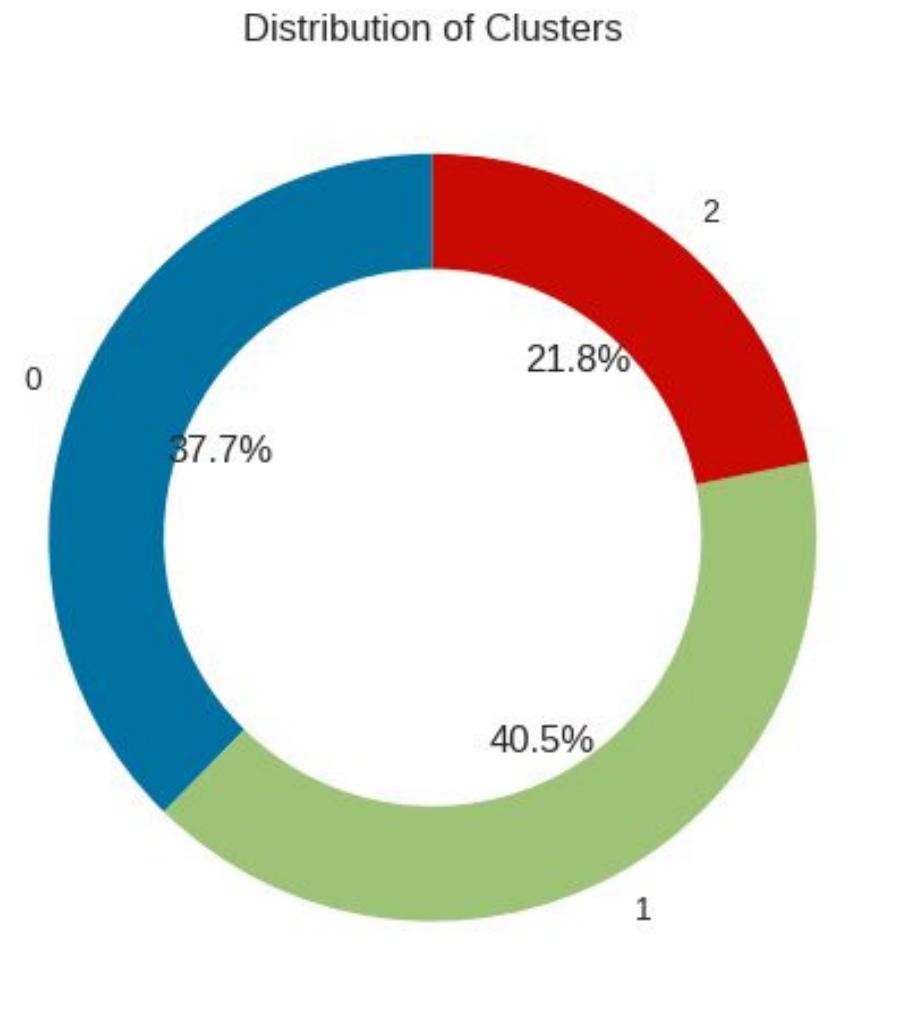


Of all the number of cluster shown, I picked **k= 3** as this has high average silhouette score and least negative silhouette coefficient values.



# K-means Result

K-means analysis interprets 3 clusters as a results. 40.5% of customer base is Cluster 1, which called **Loyal Mature Customers**, 37.7% is Cluster 0, which called **Sporadic Premium Patrons**, and 21.8% is Cluster 2 called **High-Potential Occasional Buyers**.



## Cluster 0: Sporadic Premium Patrons

This cluster represents younger customers (average age 54.69 years) with a relatively short tenure (40.08 months on average). They have received fewer promotions (5.09 on average) but are frequent users (average 17.44 months since last sales). Despite having high average sales of €19,848.78, they have a low number of transactions (average 1.5).

## Cluster 1: Loyal Mature Customers

Largest cluster, consisting of older customers (average age 63.9 years) with the longest tenure (89.2 months on average). They have received the highest number of promotions (5.63 on average) and are frequent users (average 17.39 months since last sales). However, their average sales amount is lower at €8,705.68, but they have a high number of transactions (4.11 on average).

## Cluster 2: High-Potential Occasional Buyers

Smallest cluster, with customers having an average age of 56.15 years and a medium tenure of 50.03 months. They have received the fewest promotions (4.12 on average) and are occasional users (average 23.2 months since last sales). However, their average sales amount of \$18,161.38 is high, but they have a low number of transactions (1.42 on average).

# Recommendations





## Develop plan to suits younger age

- Based on **Cluster 0** and previous analysis, consider developing products or services that appeal to **younger customers**, potentially focusing on digital banking solutions or investment options geared towards their financial goals.
- Personalize communication and also marketing plans to highlight benefits relevant to younger demographics. Focusing on younger potential **customer acquisition** will make the business more sustainable.

## Enhance Customer Engagement and Loyalty Initiatives

- Design loyalty programs and retention strategies for **Loyal Mature Customers** to maintain long-standing relationships.
- Introduce customer nurturing programs to transition **High-Potential Occasional Buyers** into more frequent and loyal customers, such as **educational content** (webinars, articles) focused on topics relevant to mature customers, such as wealth management strategies, or fraud prevention tips.

## Expand Product and Service Offerings

- Introduce specialized products and services aligned with different customer segments (e.g., wealth management).
- **Cross-sell** complementary products and services to deepen customer relationships and increase share of wallet.
- Analyze their **purchase history** and **recommend relevant products or services**. This could be done through email marketing or targeted ads on the RevoBank mobile app.

# Thank you

Let's discuss the topic further!



[Satryo Sunu Prabowo](#)

