**Datamart**

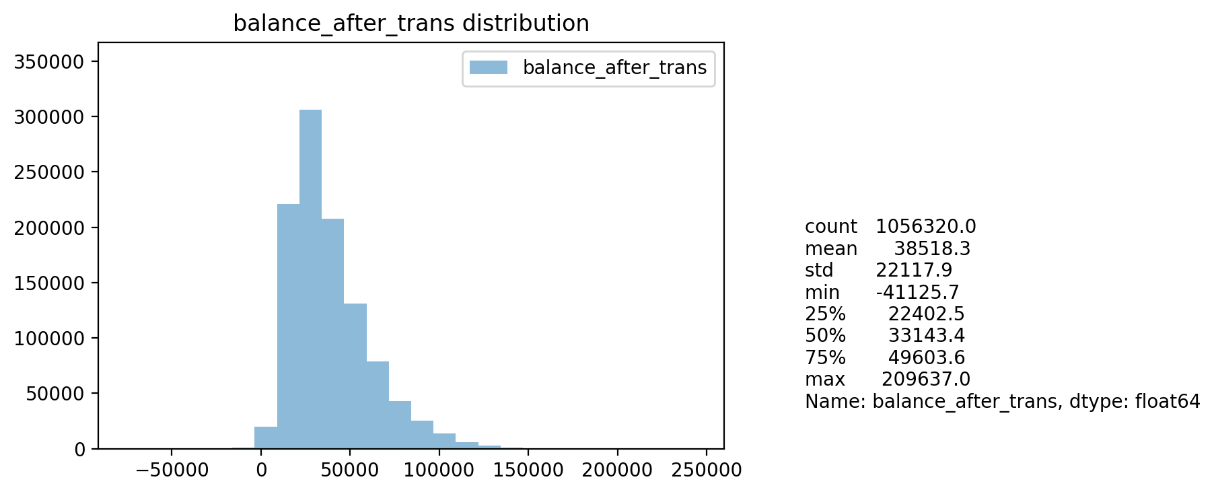
**Appendix 1 – Datamart table explanation**

**Transaction table**

There are 4,500 account ids in the transaction table, with total 1,056,320 transactions.

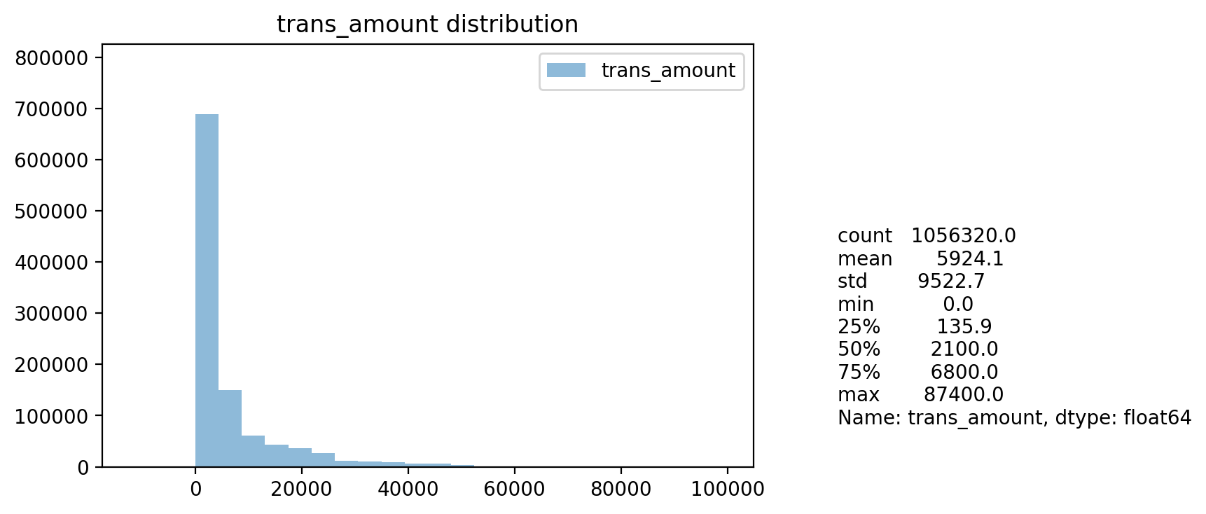
**Distribution of account balance after transaction**

Average balance after transaction is about 38k CZK. The distribution is also right-skewed, median balance is about 33k CZK and mode is falling around 25k CZK.

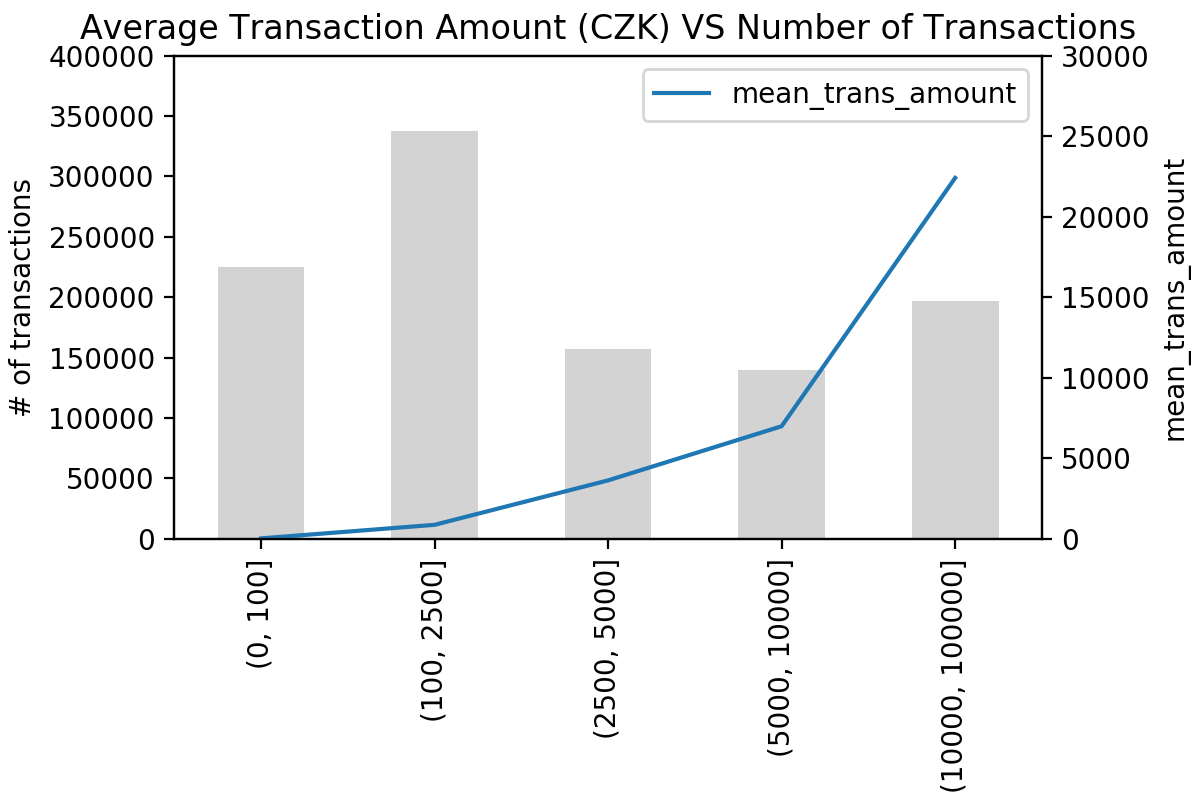


**Distribution of transaction**

The histogram of all transactions and summary statistics are shown below. The distribution is right-skewed, with average (whole timeframe, year 1993-1998) of 5.9k CZK. About 50% of total transactions was less than 2.1k CZK.



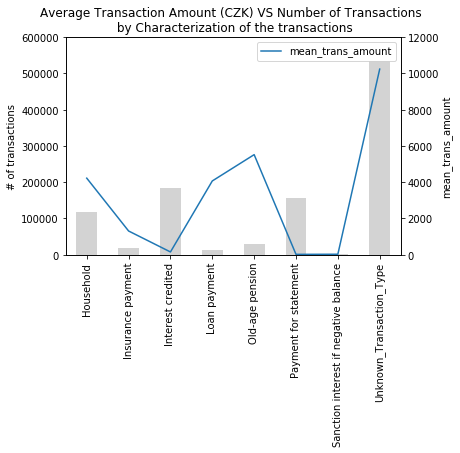
Based on the rough summary above, we performed discretizing into 5 groups, it shows that most transaction amount between 100-2500 CZK has the most occurrence in the dataset.



|  |  |  |
| --- | --- | --- |
| Transaction amount (range) | Average amount | # of transactions |
| (0, 100] | 32 | 225,242 |
| (100, 2500] | 873 | 337,789 |
| (2500, 5000] | 3,631 | 156,765 |
| (5000, 10000] | 6,992 | 139,705 |
| (10000, 100000] | 22,406 | 196,805 |

**Distribution of average transaction by characterization of transaction**

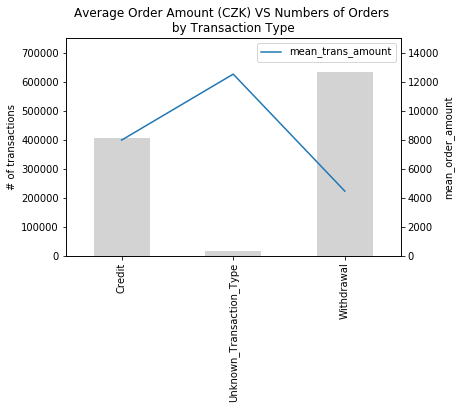
Most transactions are falling in ‘Unknown type’ group, this group also has the highest average transaction amount. It would be more useful for the bank to find out and try to identify this, to leverage their data usage.



|  |  |  |
| --- | --- | --- |
| Characterize of Transaction | Average amount | # of transactions |
| Household | 4,222 | 118,065 |
| Insurance payment | 1,307 | 18,500 |
| Interest credited | 150 | 183,114 |
| Loan payment | 4,069 | 13,580 |
| Old-age pension | 5,520 | 30,338 |
| Payment for statement | 17 | 155,832 |
| Sanction interest if negative balance | 24 | 1,577 |
| Unknown Transaction Type | 10,241 | 535,314 |

**Distribution of average transaction by transaction type**

More than 60% of transactions are ‘Withdrawal’, with the average amount of 4.5k CZK. However, a small number of transactions were not categorized and labelled as ‘Unknown’. This group has the highest average transaction amount and again, we would suggest the bank to have a closer look and identify this transaction type.



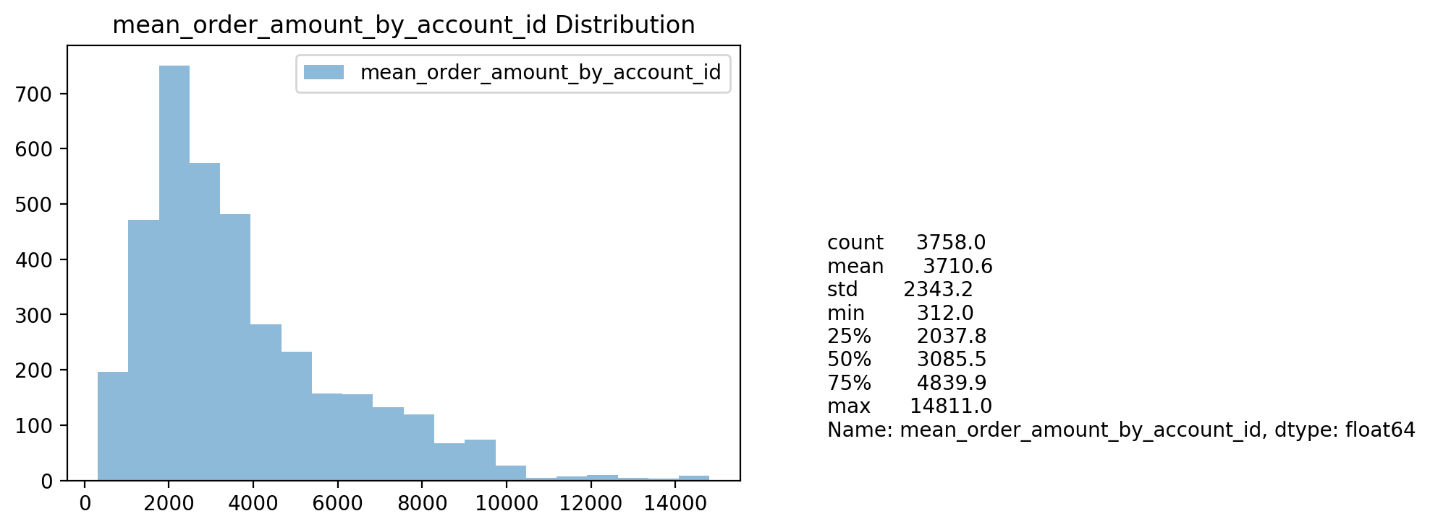
|  |  |  |
| --- | --- | --- |
| Transaction type | Average amount | # of transactions |
| Credit | 7,967 | 405,083 |
| Unknown Transaction Type | 12,517 | 16,666 |
| Withdrawal | 4,447 | 634,571 |

**Permanent Order table**

There are 3,758 unique account ids in the order table, with total of 6,471 records

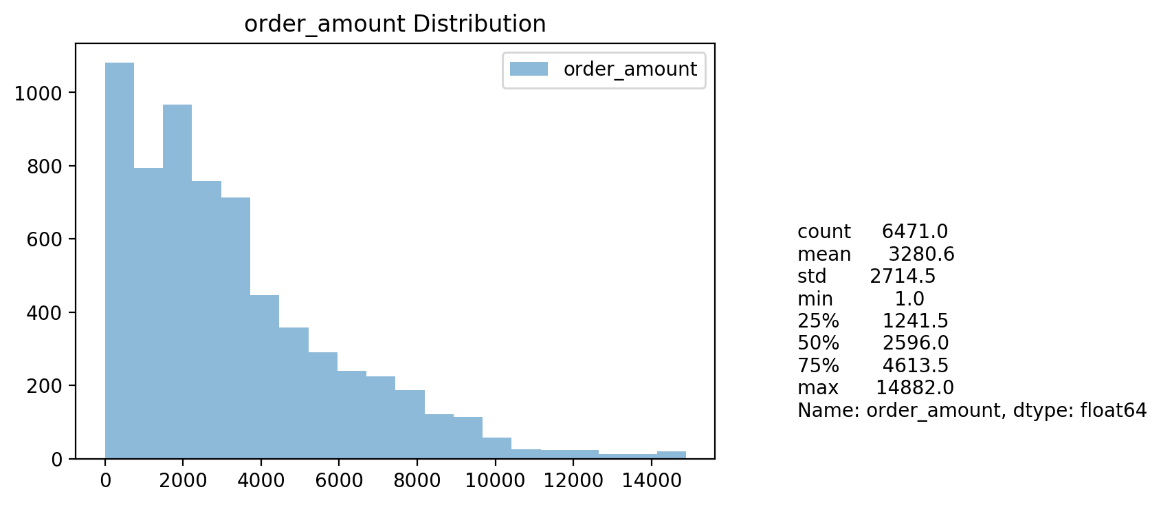
**Distribution of order amount by account id**

Average amount per order for each account id is about 3.7k CZK. The distribution is right-skewed, with median about 3.1k CZK and mode is falling around 2.5k CZK.

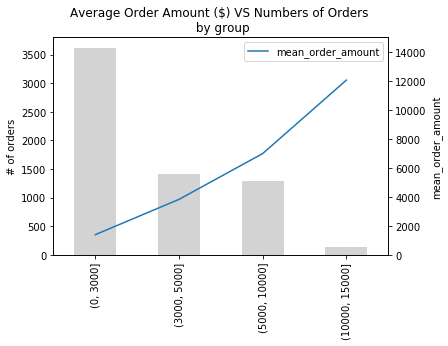


**Distribution of overall order amount**

Average amount per order is about 3.2k CZK. The distribution is right-skewed, with median about 2.6k CZK. It is also interesting to note that there are some order which has the order amount as low as 1 CZK in the database.



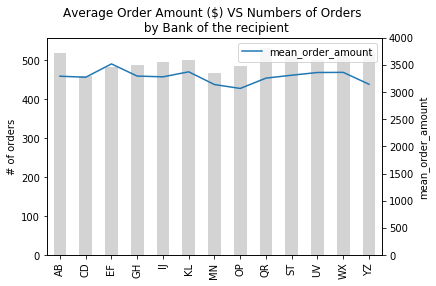
Based on the rough summary above, we performed discretizing into 4 groups, it shows that most transaction amount below 3k CZK has the most occurrence in the dataset.



|  |  |  |
| --- | --- | --- |
| Order amount (range) | Average amount | # of Orders |
| (0, 3000] | 1,398 | 3,618 |
| (3000, 5000] | 3,833 | 1,416 |
| (5000, 10000] | 6,992 | 1,300 |
| (10000, 15000] | 12,062 | 137 |

**# Order amount and # Orders for each Recipient Bank**

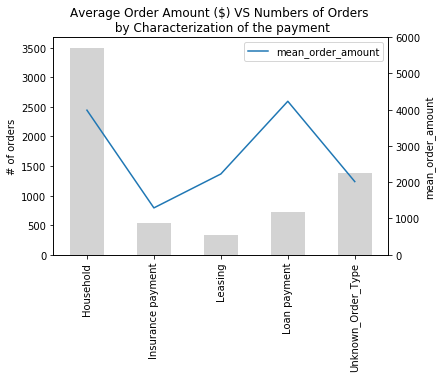
Average amount per order is about 3k CZK. All banks are equally distributed for both number of orders and the average amount.



|  |  |  |
| --- | --- | --- |
| Order Recipient Bank | Average amount | # of Orders |
| AB | 3,290 | 519 |
| CD | 3,271 | 458 |
| EF | 3,516 | 483 |
| GH | 3,292 | 487 |
| IJ | 3,279 | 496 |
| KL | 3,371 | 500 |
| MN | 3,136 | 466 |
| OP | 3,065 | 485 |
| QR | 3,255 | 531 |
| ST | 3,309 | 511 |
| UV | 3,358 | 499 |
| WX | 3,361 | 515 |
| YZ | 3,142 | 521 |

**Distribution of average order amount by characterization of payment**

Most transactions are falling in ‘Unknown type’ group, this group also has the highest average transaction amount. It would be more useful for the bank to find out and try to identify this, to leverage their data usage.



|  |  |  |
| --- | --- | --- |
| Characterize of Order | Average amount | # of Orders |
| Household | 3,988 | 3,502 |
| Insurance payment | 1,291 | 532 |
| Leasing | 2,227 | 341 |
| Loan payment | 4,233 | 717 |
| Unknown Order Type | 2,017 | 1,379 |

**Basetable Analysis**

‘trans\_ratio\_98\_97’ column derives from the ratio of transaction growth in the past year. After sorted this column descending, it reveals the top 10% customers with the most transaction growth in the past year and also the bottom 10% which had the least transaction growth. With this information, the bank can take appropriate action on this.

The data was exported into two csv files; bottom10percent\_customer.csv, top10percent\_customer.csv.