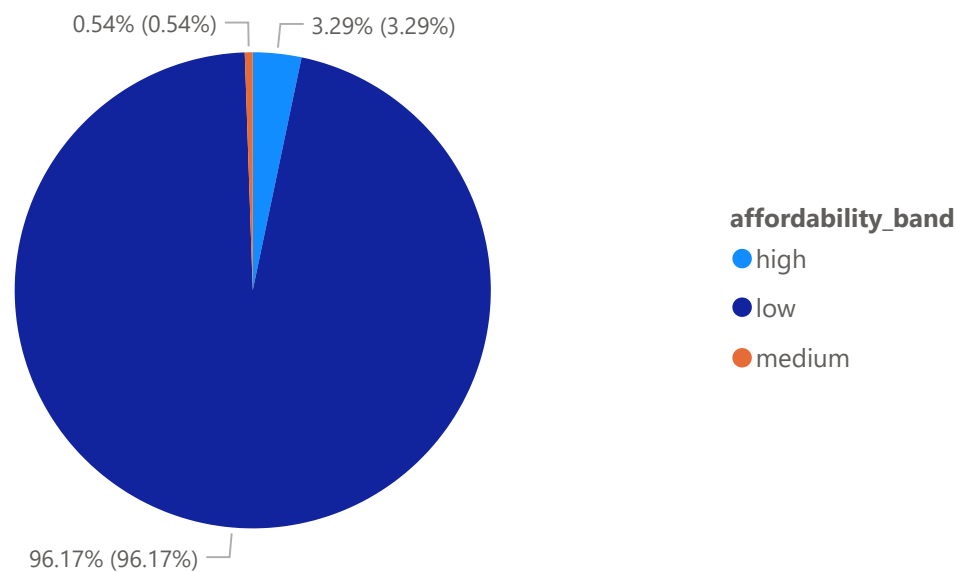


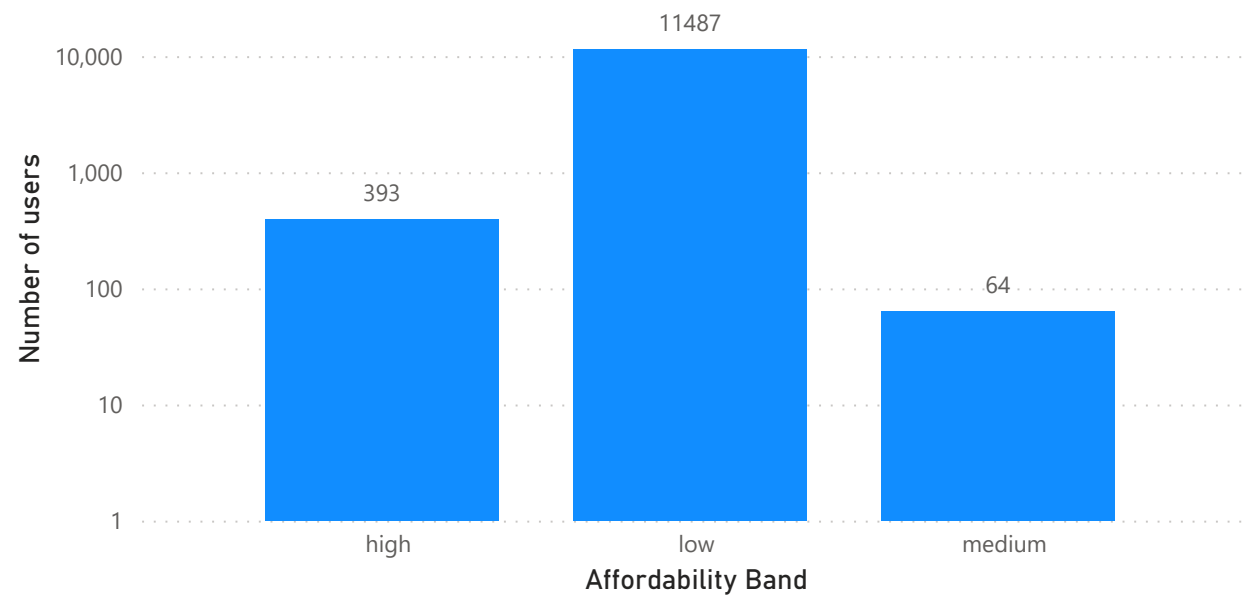
Understanding Affordability and Loan Suitability through Transaction Analysis and Visualisation

By Joseph Awonusi

Percent of users per affordability band



Number of Users per Affordability Band



Affordability Corridors

Low = Less than £0.00

Medium = £0.00 to £390.00

High = Above £390.00

Note

As date fields (transaction_date and created_on) mostly only include time, the corridor concept was applied over the entire dataset opposed to a representative average of monthly credit and debit transactions.

High Level Statistics

All

Range of corridor: £1,243,773.03
Mean of corridor: £2,430.82
Median of corridor: £-1034.64
Standard deviation of corridor: £34,080.48

Low

Range of corridor: £7,0917.99
Mean of corridor: £-1,645.36
Median of corridor: £-1,084.93
Standard deviation of corridor: £2,063.02

Medium

Range of corridor: £327.33
Mean of corridor: £114.81
Median of corridor: £55.22
Standard deviation of corridor: £92.09

High

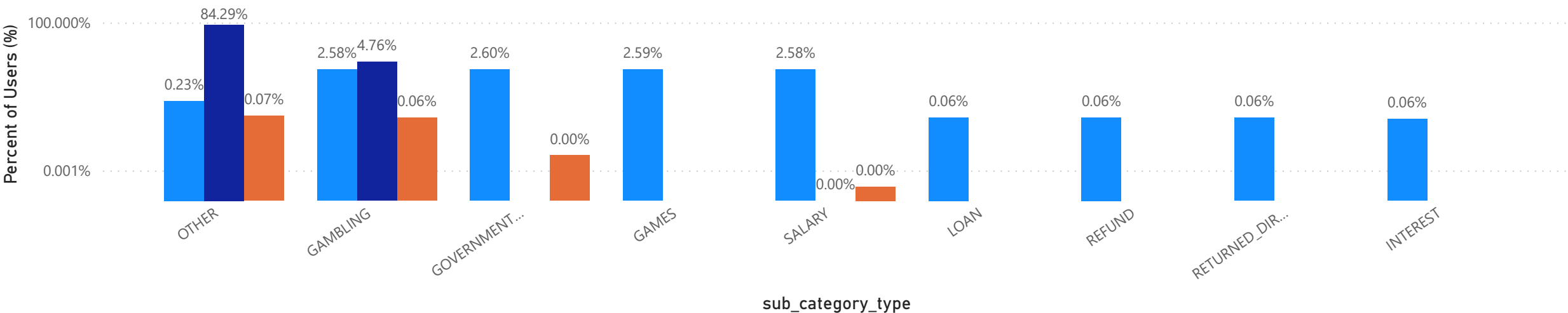
Range of corridor: £1,172,463.50
Mean of corridor: £121,950.80
Median of corridor: £139,015.00
Standard deviation of corridor: £143,012.83

Note: 'Corridor' is the difference between credit and debit transactions per user

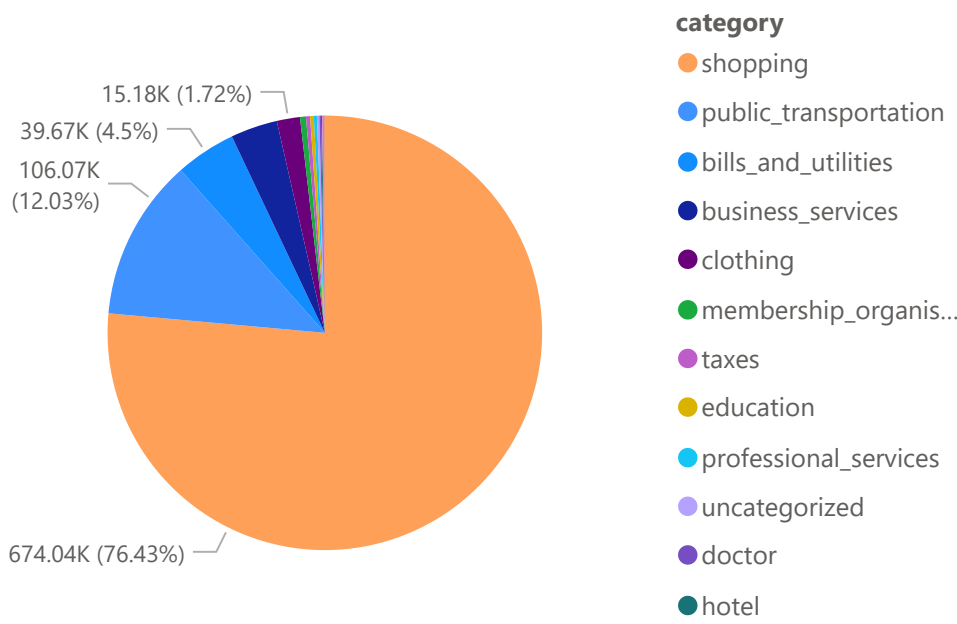
Overview of Data

Number of Users per Sub-Category

affordability_band high low medium



Number of Users per Category

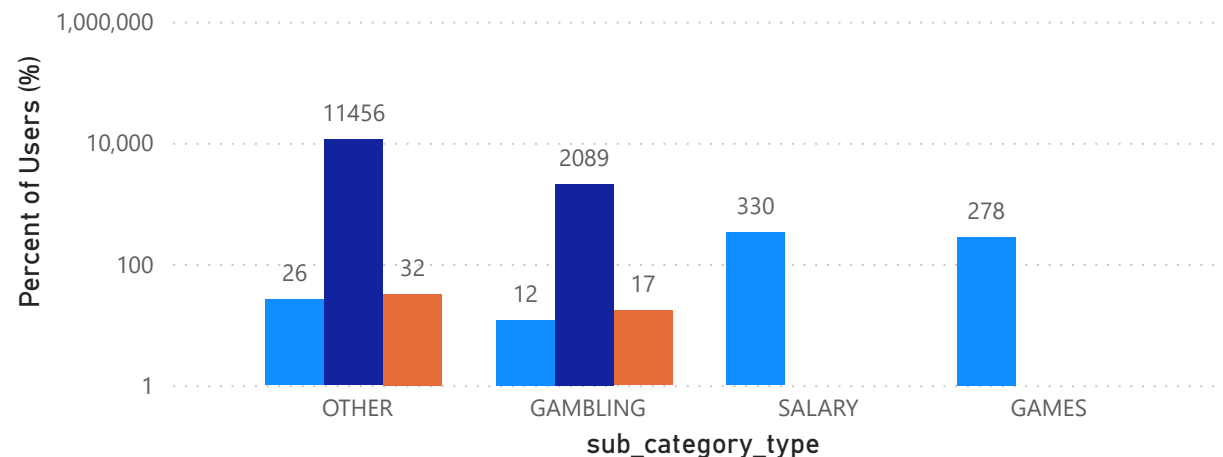


- The percentage of users per sub-category highlight 'Other' as a primary type of expence, of which is mainly made of shopping and public transportation.
- Of all transactional data 7.4% of records are related to gambling. The low affordability group has the highest proportion of records.
- Additionally, the only loans which are seen within the dataset belong to users within the high affordability band.

Overview of Data

Distinct Number of Users per Expense Sub-Category

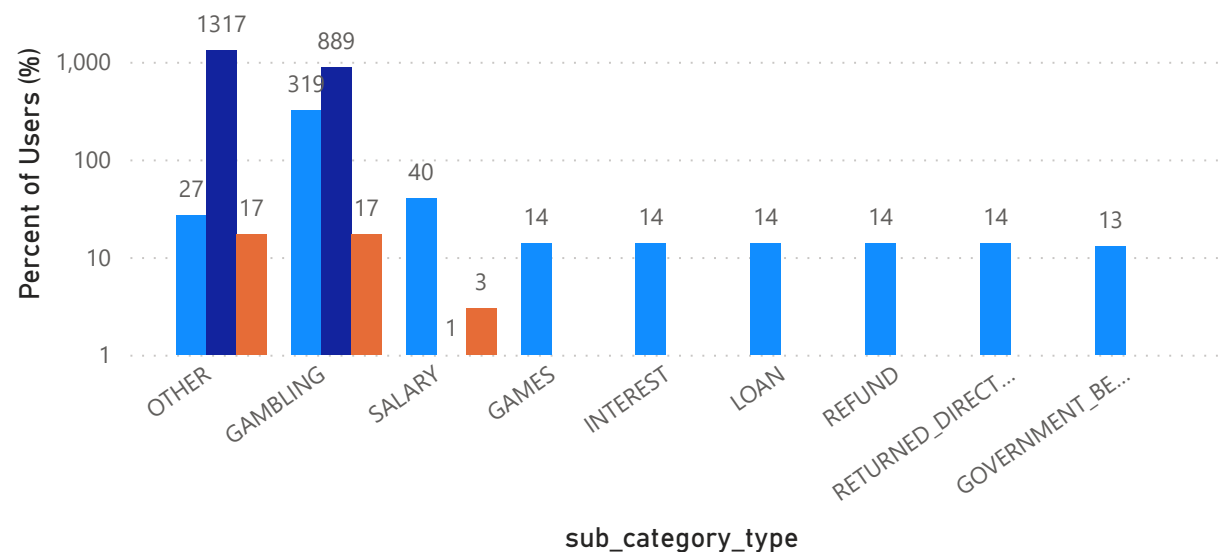
affordability_band ● high ● low ● medium



- When the 'category_type' is 'EXPENSE', the proportion of the high and medium affordability band involved in gambling drops significantly.
- Of the 11,487 low affordability users, 2,089 (18%) of them have been involved in gambling related expenses at least once. This is 3.0% and 2.7% for high and medium band respectfully.

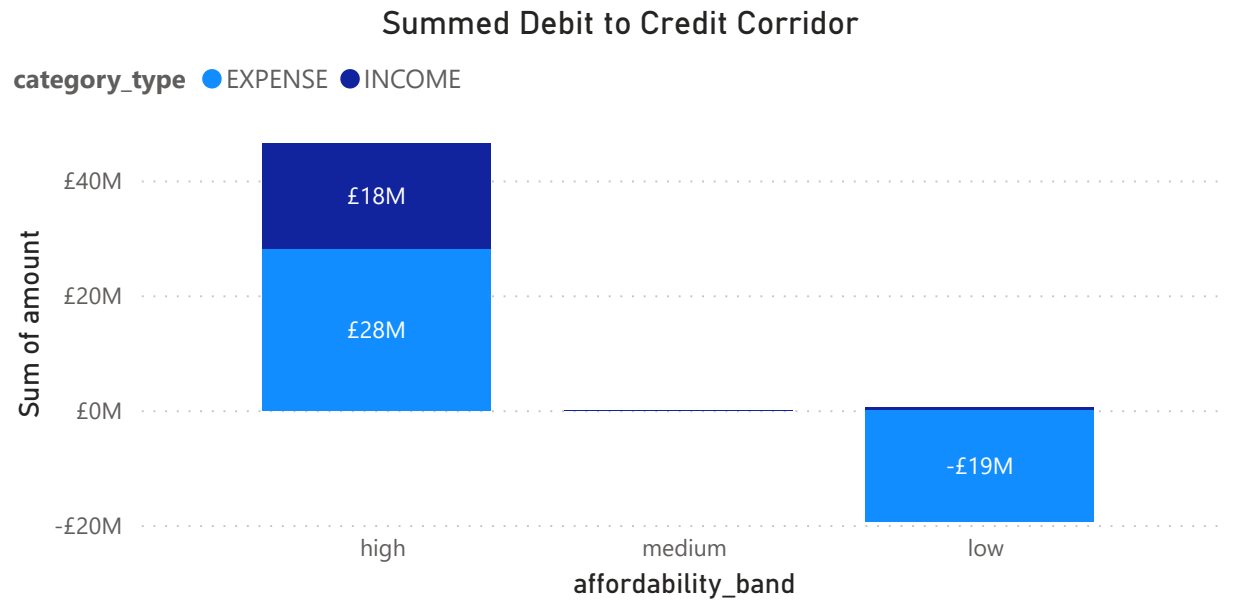
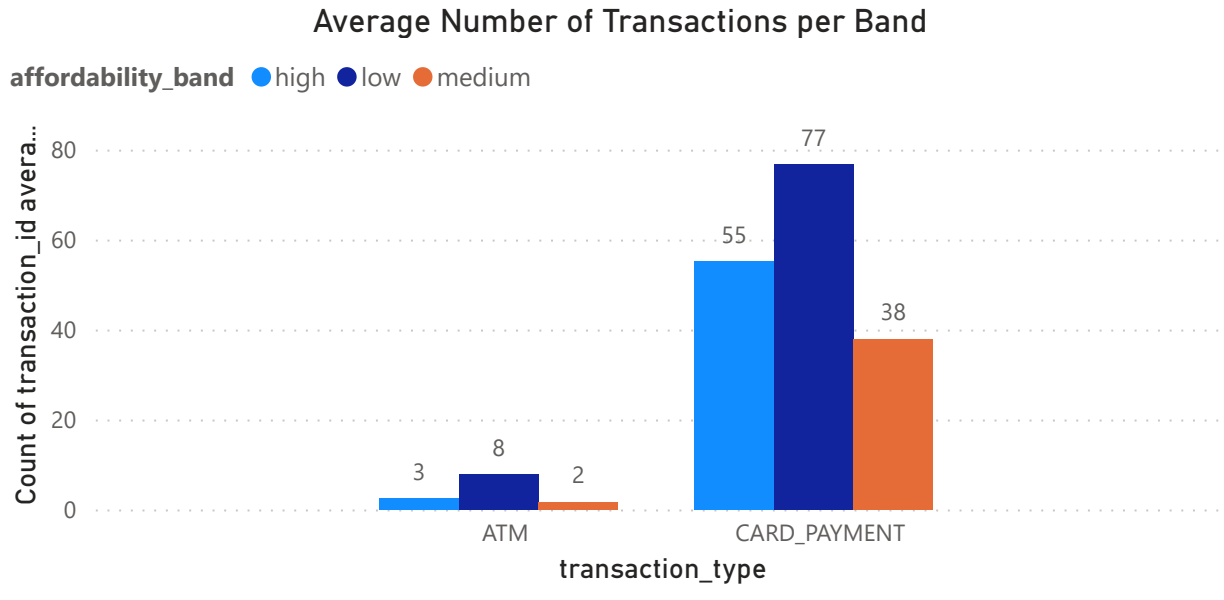
Distinct Number of Users per Income Sub-Category

affordability_band ● high ● low ● medium



- When the 'category_type' is 'INCOME', 'OTHER' remains the primary sub-category type (consists of 'clothing' and 'shopping' categories).
- A significant number of low (7%) and high (81%) affordability users are involved with gambling at least once with in this dataset.
- Only 40 of the high affordability users have evidence of a salary income within this dataset.
- The results indicate that income on the majority of case (outside of salary and government benifits) does not come from a reliable source.

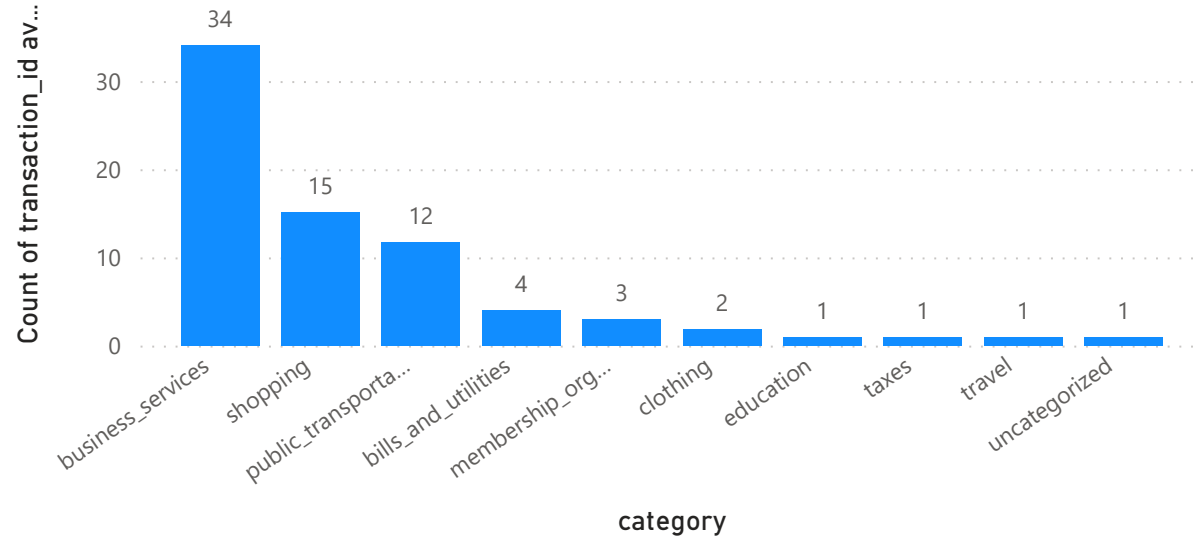
Overview of Data



- The clustered bar plot shows that low affordability users make more expense related transactions than medium and high affordability users on average (median).
- The summed debit to credit corridor lies far to the negative for low affordability users. The weight of expenses vs. income would indicate difficulty of paying back any kind of loan for users within this band.
- Medium

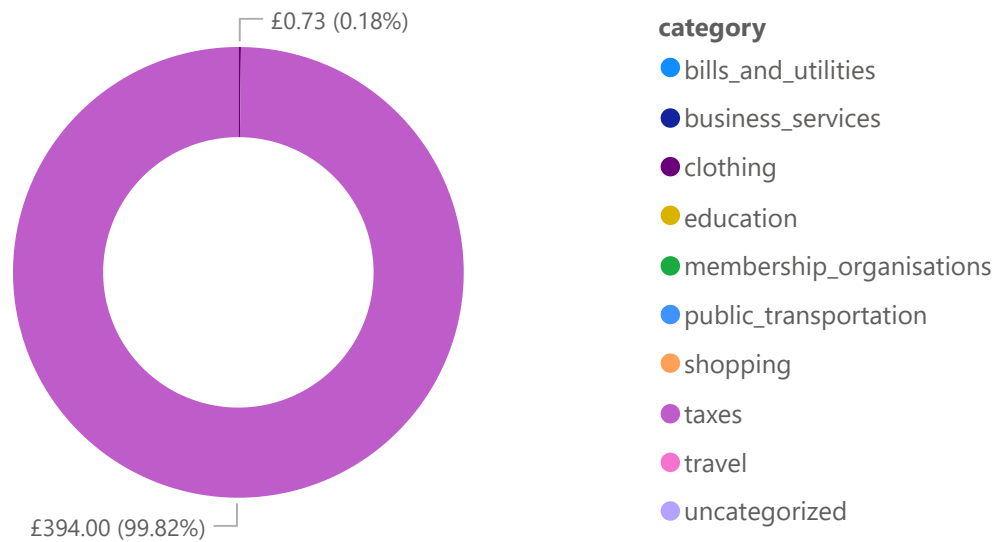
Medium Affordability Band

Frequency of transactions per Category

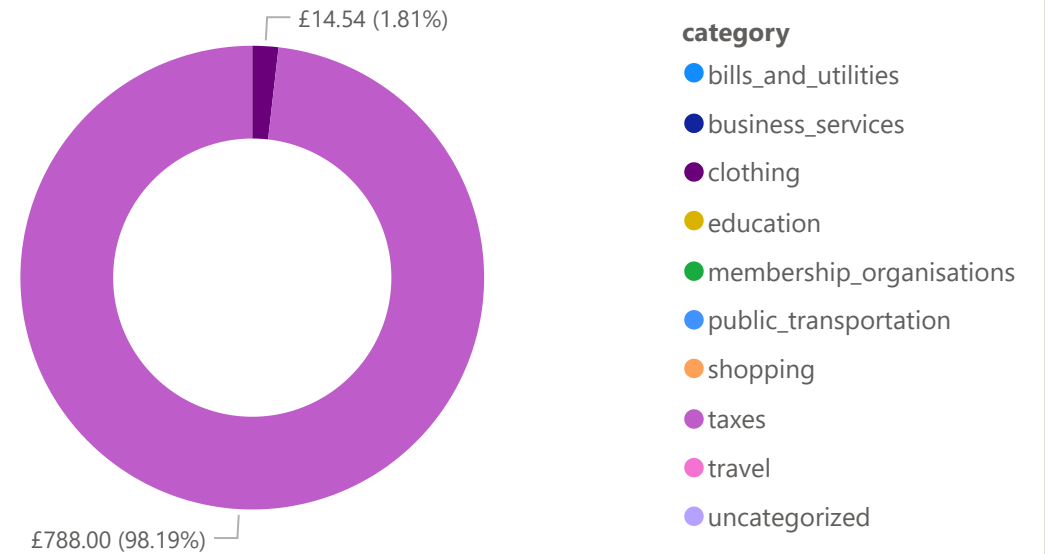


- For the medium affordability band, 'business services' is the primary category for expense frequency. However, the doughnut chart shows that the majority of expenses is committed towards taxes (85%)
- Note, the median is used as an average to mitigate outlier influence.
- When looking at the sum amount of transactions, taxes also outweigh every other category

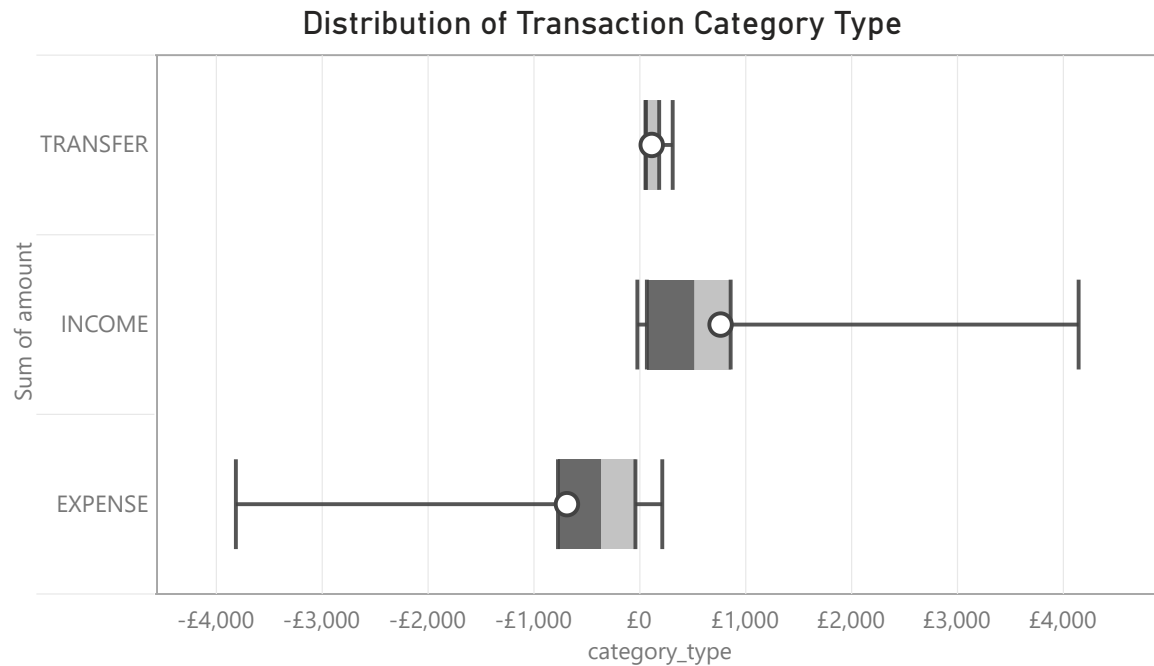
Average of amount by category



Sum of amount by category



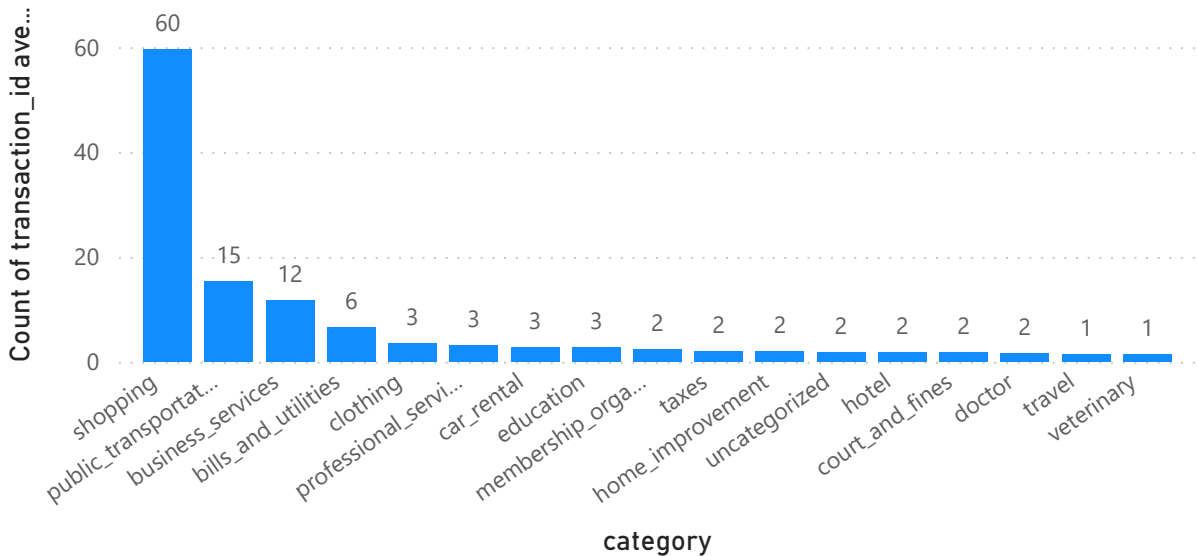
Medium Affordability Band



- Typically, the income amounts are not too much greater than the expenses seen within the dataset for medium affordability users.
- Income and expense has their mean lie towards the upper and lower quartile respectively indicating a skew within the dataset.
- It may be possible for the medium affordability band users to be able to make loan repayments, however caution is advised on a per user basis to ensure they make enough to repay the loan.

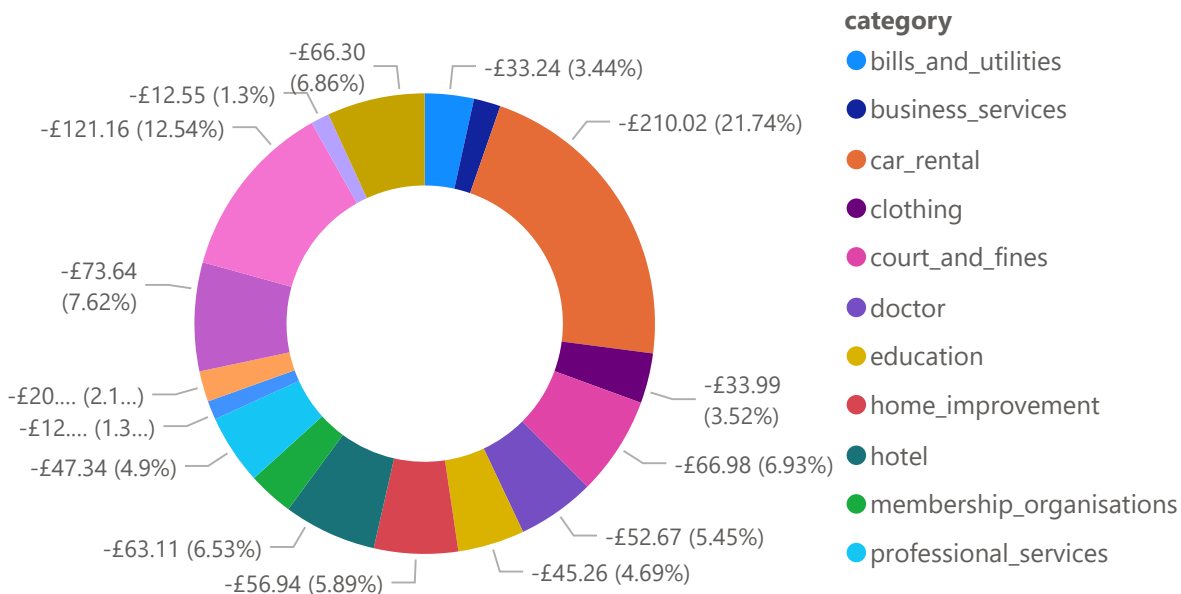
Low Affordability Band

Frequency of transactions per Category

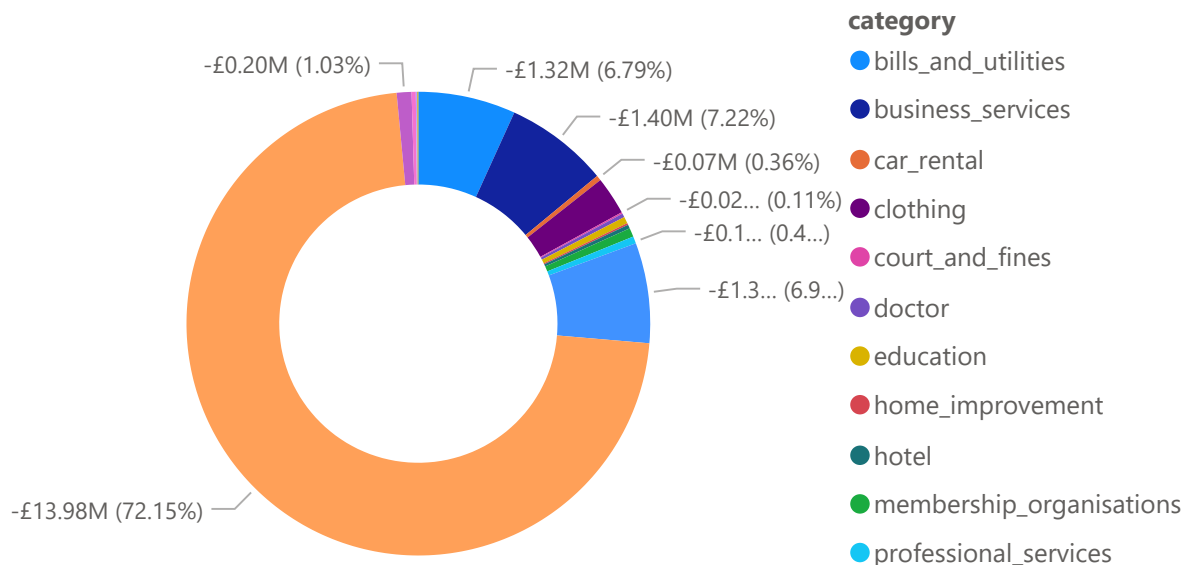


- Low affordability has a high number of expense related transactions found in the 'shopping' category.
- Although the average cost of shopping per transaction is significantly lower than car rental, travel and taxes, it's sum total heavily outweighs all other categories.
- Low affordability users have a greater diversity of expense categories than high and medium affordability groups.

Average of amount by category

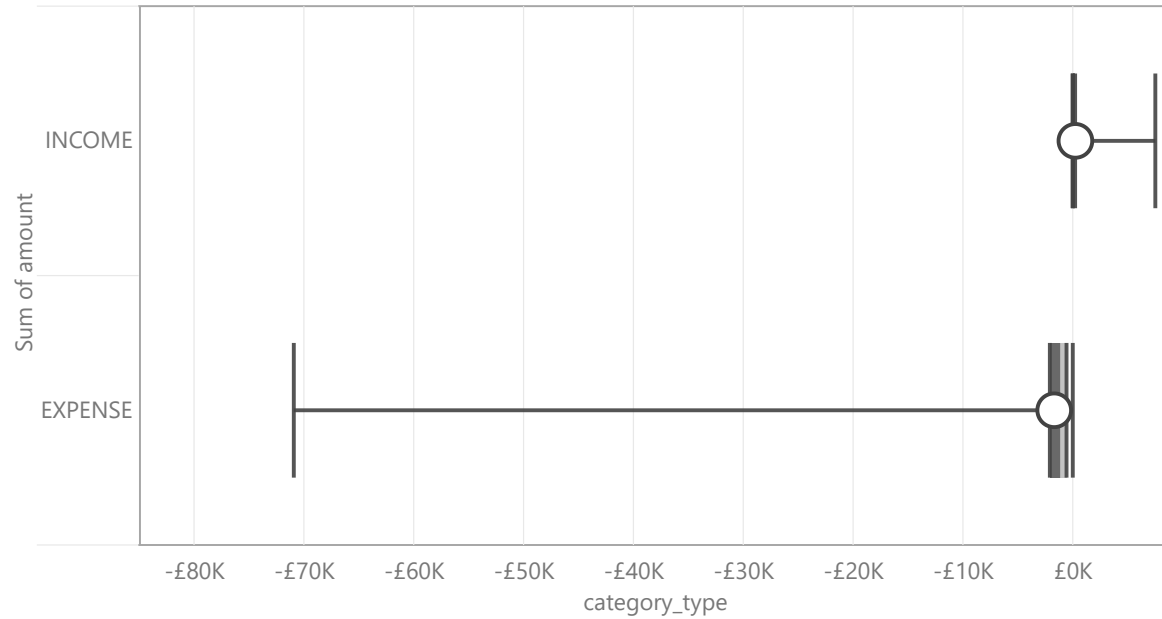


Sum of amount by category



Low Affordability Band

Distribution of Transaction Category Type



- The low affordability band has a very narrow margin for income related transactions.
- Expense related transactions have a narrower margin than the medium affordability does.
- Typically, it's ill-advised to provide someone within the low affordability band with a loan as the likelihood of repayment is very low.