Predictor table for y = transcount

|  |  |  |
| --- | --- | --- |
| Predictor | Expected sign of effect | Rationale |
| Revolving Indicator | -/+ | Customers who are in default often use credit cards less frequently (Neg) than actors who are able to pay their obligations in full each month (Pos). |
| Wealth tag | + | Those who are wealthier are more inclined to transact frequently. |
| Card type | + | Cardtype + Card tiers are frequently upgraded as users make more transactions. Higher Card Tier = Greater Assets + More Transactions. |
| Spend Category | -/+ | Spending on luxury items like airfare would be less common and required than shopping for groceries, driving a car, or buying personal items. |
| Transamount | + | More financial security and disposable money are indicated by higher value transactions, which can also lead to an increase in transaction volume. |
| ClientNum | None | It’s a unique ID. |
| Month | None | Even though month might have a seasonal effect on the number of transactions, Im assuming that it does not affect for this data. |

Predictor table for y = transamount

|  |  |  |
| --- | --- | --- |
| Predictor | Expected sign of effect | Rationale |
| Wealth tag | + | Transactions with people who are more financially secure or rich are also probably to be for bigger amounts. |
| Card type | + | This should have a favorable impact because card tiers are often upgraded when customers spend more money. Higher tier = more spending. |
| Revolving indicator | +/- | Delinquent and transactional consumers are expected to have opposing effects on transaction amounts because the revolving indicator is a measure of the customer's financial situation. |
| Spendcategory | +/- | As an example, flight is innately more expensive than groceries (expensive, positive impact) (affordable, negative impact). |
| Transcount | + | More transaction amounts overall are directly correlated with higher transaction counts. |
| Month | None | Month could have a seasonal impact on spendings. For example, In Christmas time, customers could use their credit card for shopping more. But I’m considering it doesn’t affect for this data because its just 5 months of data and difficult to extract patterns. |
| Client Number | None | Client number is an ID and doesn’t help much in estimating transamount. |

EDA- OLS may work for transcount. log(transamount) plot shows that an exponential model will be better for transamount.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, bar chart, histogram

Description automatically generated

Strong correlation between transcount and transamount.

**Models for transamount-**

**amount1.out -**This model does not obey the linear and Equality of variance assumptions. Hence we tried applying an exponential transformation on Y.

**amount2.out -** This model obeys all LINE assumptions except equality of variances. I have tried WLS and FGLS methods to reduce heteroscedasticity but in vain.

**LINE graphs for model 2.**

Graphical user interface, chart, histogram

Description automatically generated

Although Normal Q-Q plot fails on the edges, the histogram shows that the residuals are normal. We see the fanning in effect in the residuals vs. fitted graph and hence Equality of variances is violated. The D-W statistic is 1.830638, which is less than 2, indicating that there is evidence of positive autocorrelation in the residuals hence satisfying Independence assumption.

How do customers’ spending pattern vary by spending category, while controlling for other variables?

spendcategory: Compared to the reference category of "Personal", having the spend category of "Airlines", "Auto", "Business", or "Grocery" is associated with a positive effect on the log of transaction amount. Specifically, the estimated effect is 62.2%, 39.2%, 39.4%, and 61.0%, respectively, controlling for all other variables in the model.

What type of customers have the highest and lowest spending by card type? By what amount?

wealthtagEmergingAffluent:cardtypeGold have 11.3% higher in terms of transaction amount compared to the reference level of wealthtag :Mass market and cardType : Blue, holding all other variables constant. The lowest would be the reference level mentioned.

How do their spending patterns vary by wealth tag?

Customers tagged as "Affluent," "Emerging Affluent," and "High Net Worth" have estimated spending levels that are 1.798, 1.381, and 2.128 times higher, respectively, compared to the reference level of “Mass market”, holding other variables constant.

How do their spending patterns vary by revolving indicators?

The estimated coefficients for the three types of revolving indicators (Occasional Revolver, Revolver, and Transactor) are 0.780, 0.611, and 0.903, respectively, which means that holding all other variables constant, customers who are Occasional Revolvers, Revolvers, and Transactors are estimated to have 1.78, 1.84, and 2.47 times higher spending, respectively, compared to customers who are not revolvers.

Stragazer output-

Table

Description automatically generatedTable

Description automatically generated

**Models for transcount-**

**count1.out -**This model does not obey the linear and Equality of variance assumptions. Hence, we tried applying an exponential transformation on Y.

**count2.out -** This model obeys all LINE assumptions except equality of variances. I have tried WLS and FGLS methods to reduce heteroscedasticity but in vain.

How do customers’ spending pattern vary by spending category, while controlling for other variables?

Spending category "Auto" has an estimate of -0.1614, which means that when all other variables are held constant, spending on auto decreases by 16.14% compared to spending on airlines. Similarly, spending on "Business" and "Entertainment" decreases by 15.25% and 41.80%, respectively, compared to spending on airlines.

What type of customers have the highest and lowest transcount by card type? By what count?

The customer type "Emerging Affluent" with a "Platinum" card has the highest transcount, with an estimated coefficient of 0.1101. Compared to the reference category of "Mass Market" with a "Blue" card, this represents an increase in transcount of 11.01%.

How do their transaction counts vary by wealth tag?

Emerging Affluent category, High Net Worth customers have transaction counts that are 11.6% higher and Mass Market customers have transaction counts that are 28.8% lower.

How do their transactional counts vary by revolving indicators?

individuals with an OccasionalRevolver indicator tend to have 66.4% more transactions than those with a Delinquent indicator. individuals with a Revolver or Transactor indicator tend to have 50.4% and 81.9% more transactions, respectively, compared to those with a Delinquent indicator. These proportional differences hold after controlling for other predictors in the model.

Graphical user interface, chart, histogram

Description automatically generated