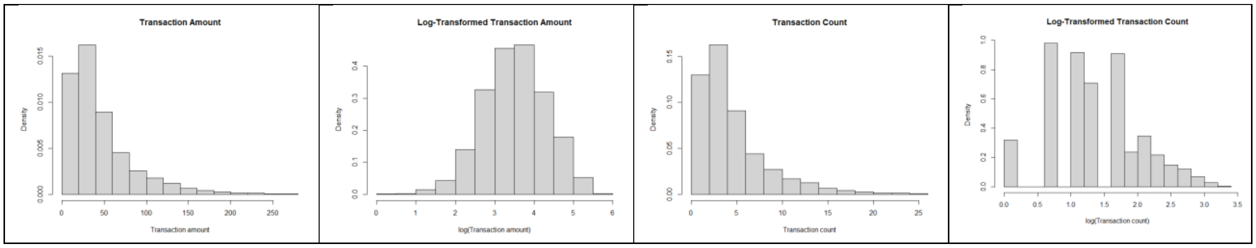
Credit Card Transactions

Predictor table:

Predictor	Effect	Rationale
DV: TransAmount		
Month	+/-	Since the transaction amount may vary based on factors such as seasonality, holidays, and promotions, the month variable could have a positive or negative effect on predicting transaction amount.
WealthTag	+	High net worth individuals may make larger purchases or spend more frequently on luxury items, while mass-market customers may be more price-sensitive and focused on necessities.
CardType	+	Platinum cardholders may have higher spending limits or more rewards points, which could incentivize them to spend more.
RevolvingIndicator	+	Customers who carry balances from month to month (revolvers) may be more likely to incur interest charges and fees, which could limit their spending capacity (must be ordered in ascending order).
SpendCategory	+/-	Personal vs. business. May have different pattern and magnitude of spend
TransCount	+/-	It is possible that customers who buy more frequently spend more, but it is also possible that customers who buy less frequently spend more per purchase.
Excluded: ClientNum		
DV: TransCount		
Month	+/-	Since the number of transactions may vary based on factors such as seasonality, holidays, and promotions, the month variable could have a positive or negative effect on predicting transaction count.
WealthTag	+/-	High net worth individuals may be more likely to make larger purchases or spend more frequently on luxury items, while mass-market customers may be more price-sensitive and focused on necessities.
CardType	+/-	Platinum cardholders may have higher spending limits or more rewards points, which could incentivize them to spend more.
RevolvingIndicator	+/-	Customers who carry balances from month to month (revolvers) may be more likely to incur interest charges and fees, which could limit their spending capacity
SpendCategory	+/-	Spending behavior may vary based on the category of spend (e.g. personal vs. business).
Excluded: ClientNum, TransAmount		

Exploratory Data Analysis

Histograms of DV:
The distributions of transaction amount and transaction counts are right-skewed, and hence, OLS regression will not be suitable. The distributions of log-transformed transaction amount and count are close to normal, and therefore, more suited for OLS regression.



Correlations:
Almost every predictor is a factor variable; the only two continuous variables are TransCount and TransAmount. However the correlation between them is 0.97, i.e., they are almost the same variable. So we can't use TransCount as a predictor of TransAmount.

cor(d$TransCount, d$TransAmount, use="complete.obs")
0.9720903

Regression Analysis

```
Count_Model <- lm(log(TransCount) ~ Month + ClientNum + WealthTag + CardType +
    RevolvingIndicator + SpendCategory, data = d)
Amount_Model <- lm(log(TransAmount) ~ Month + ClientNum + WealthTag + CardType +
    RevolvingIndicator + SpendCategory, data = d)
```

	Dependent variable:	
	log(TransCount) (1)	log(TransAmount) (2)
Month1302	0.020* (0.010)	0.014 (0.011)
Month1303	0.076*** (0.010)	0.061*** (0.011)
Month1304	0.006 (0.010)	-0.008 (0.011)
Month1305	0.018* (0.010)	0.007 (0.011)
Month1306	0.084*** (0.010)	0.090*** (0.011)
WealthTagAffluent	0.662*** (0.009)	0.594*** (0.012)
WealthTagEmergingAffluent	0.384*** (0.008)	0.359*** (0.010)
WealthTagHighNetWorth	0.864*** (0.009)	0.768*** (0.013)
CardTypeGold	0.402*** (0.007)	0.370*** (0.009)
CardTypePlatinum	0.683*** (0.007)	0.606*** (0.011)
RevolvingIndicatorDelinquent	-0.970*** (0.008)	-0.865*** (0.014)
RevolvingIndicatorOccasionalRevolver	-0.176*** (0.008)	-0.137*** (0.009)
RevolvingIndicatorRevolver	-0.363*** (0.008)	-0.308*** (0.010)
SpendCategoryAirlines	0.009 (0.010)	0.010 (0.011)
SpendCategoryAuto	-0.288*** (0.010)	-0.236*** (0.011)
SpendCategoryBusiness	-0.277*** (0.010)	-0.235*** (0.011)
SpendCategoryEntertainment	-0.700*** (0.010)	-0.608*** (0.013)
SpendCategoryPersonal	-0.691*** (0.010)	-0.608*** (0.013)
log(TransCount)		0.283*** (0.011)
Constant	1.208*** (0.013)	2.944*** (0.019)
Observations	8,052	8,048
R2	0.847	0.880
Adjusted R2	0.847	0.880
Residual Std. Error	0.269 (df = 8033)	0.276 (df = 8028)
F Statistic	2,477.596*** (df = 18; 8033)	3,110.060*** (df = 19; 8028)
Note:		*p<0.1; **p<0.05; ***p<0.01

Note: We could include interactions in these models. However, interpreting interactions of factor variables would be quite challenging.

Interpretation

1. How do customers' spending patterns vary by spending category, while controlling for other variables?
Compared to grocery, customers have 0.9% more transactions on airlines and spend 1.2% more, 28.8% fewer transactions on auto and spend 31.4% less, 27.7% fewer transactions on business purchases and spend 31.4% less, 70% fewer transactions on entertainment and spend 80.6% less, and 69.1% fewer transactions on personal purchases and spend 80.3% less. The effects indicated in red are unexpected.

2. What type of customers have the highest and lowest spending by card type? By what amount?
Compared to customers with blue card, customers with gold card have 40.2% more transactions and spend 48.4% more, and customers with platinum card have 68.3% more transactions and spend 79.9% more.

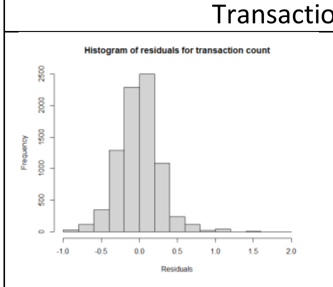
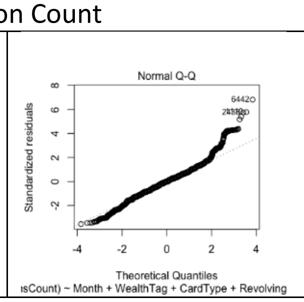
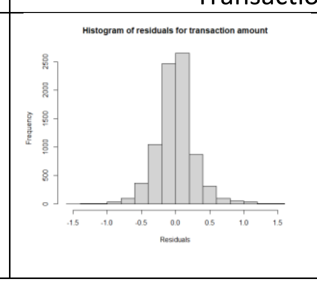
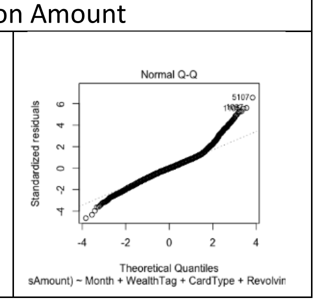
3. How do their spending patterns vary by wealth tag?
Compared to mass market customers, emerging affluent customers have 38.4% more transactions and spend 46.8% more, affluent customers have 66.2% more transactions and spend 78.1% more, and high net worth customers have 86.4% more transactions and spend 101.3% more.

4. How do their spending patterns vary by revolving indicators?
Compared to transactors, occasional revolvers have 17.6% fewer transactions and spend 18.7% less, revolvers have 36.3% fewer transactions and spend 41.1% less, and delinquents have 97% fewer transactions and spend 114% less.

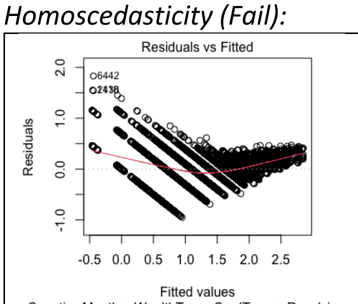
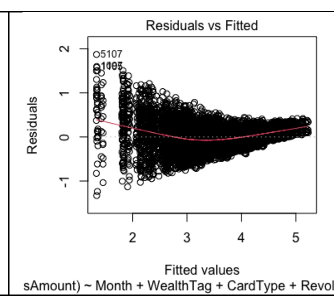
Assumptions

Linearity: Cannot be tested because all of our predictors are categorical variables.

Normality (Pass): There is some deviations from normality for residuals from both models at high quantiles, but they look fairly normal for low-mid range quantiles.

Transaction Count		Transaction Amount	
			

Homoscedasticity (Fail):

	Transaction Count: Residual plot shows that residuals are heteroskedastic. Bartlett's test also shows p<2.2e-16		Transaction Amount: Residual plot shows that residuals are heteroskedastic. Bartlett's test also shows p<2.2e-16

Multicollinearity (Pass): VIF tests shows that all independent variables in both count and amount models have GVIF^(1/(2*Df)) values less than 5, indicating no significant multicollinearity.

vif(Count_Model)				vif(Amount_Model)			
	GVIF	Df	GVIF^(1/(2*Df))		GVIF	Df	GVIF^(1/(2*Df))
Month	1.004007	5	1.000400	Month	1.004026	5	1.000402
WealthTag	1.002456	3	1.000409	WealthTag	1.002450	3	1.000408
CardType	1.001240	2	1.000310	CardType	1.001218	2	1.000304
RevolvingIndicator	1.001769	3	1.000295	RevolvingIndicator	1.001790	3	1.000298
SpendCategory	1.004598	5	1.000459	SpendCategory	1.004611	5	1.000460

*Independence (Pass):*Durbin-Watson test shows residuals in both count and amount models have DW statistic in the [1.5-2.5] range, indicating no severe violation of the independence assumption.

dwtest(Count_Model	dwtest(Amount_Model)
DW = 1.6434, p-value < 2.2e-16	DW = 1.6924, p-value < 2.2e-16