



Winery Customer Analysis

Team 13: Belle Lin, Caroline Martina, Sarah Fauska, Ivan Huang, Weihan Weng

Executive Summary

Data Description

This dataset contains basic customer, sales, and marketing information (21 columns) from a winery between 2008-2010 related to 65,534 orders.

The Problem

The winery wants to boost sales by determining which customers to target and which marketing channels are the most effective for different customer segments based on their sales data.

The Proposed Solution

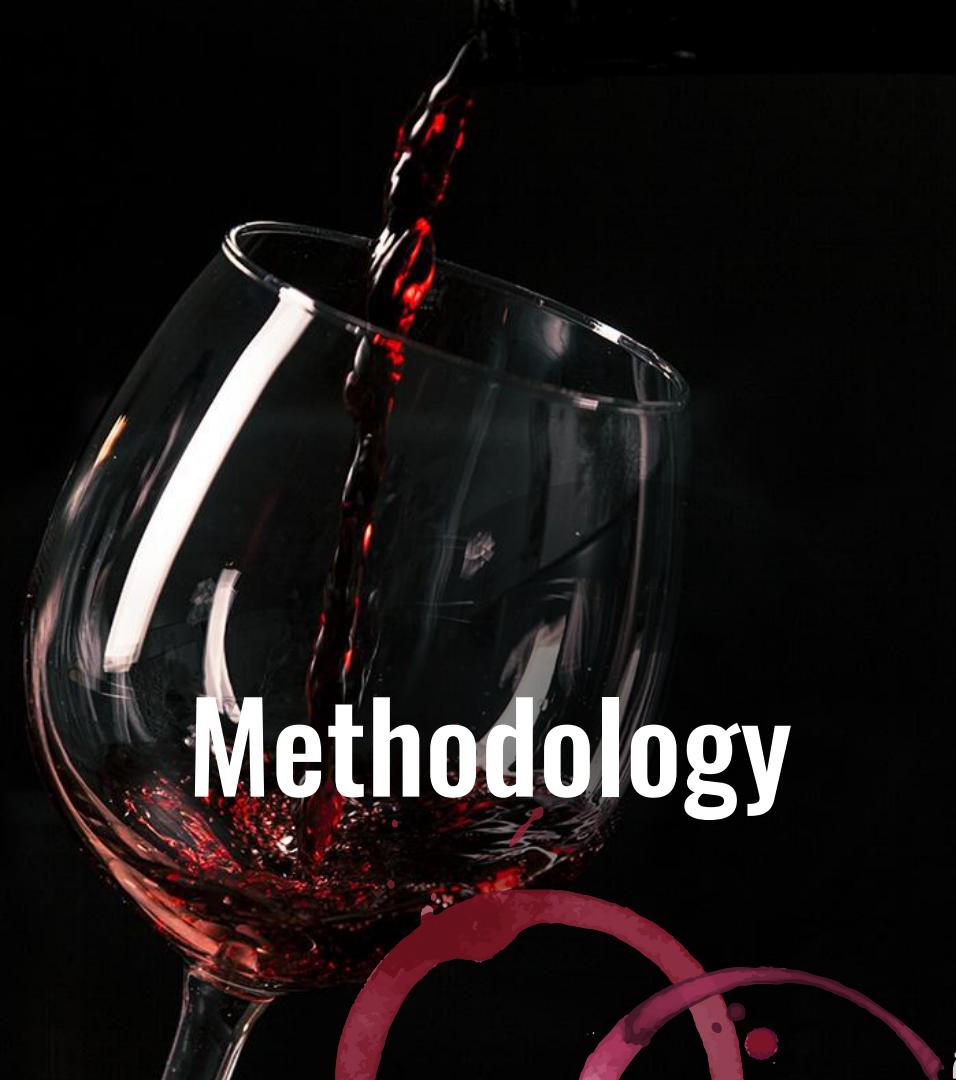
The winery can conduct a RFM (Recency, Frequency, Monetary) analysis to identify the high potential customer segments and strategic sales channel. In addition to this, the winery wants to evaluate and refine advertising strategies by analyzing the response patterns of each marketing channel to ensure the most effective promotional outreach.





Business Questions

01. Which sales channel is the most effective for each consumer segment?
 02. What sales channel should we not target a consumer segment with?
 03. What are the purchasing behaviors of all our customers?
 04. Which customers are the best customers to target?
- 



Methodology

Cleaning and Summarizing Data

- Remove rows with N/A and Blank Data by using R Studio
- Break down data by sales/consumer segment/location/etc using Tableau & Excel

Advertising Response

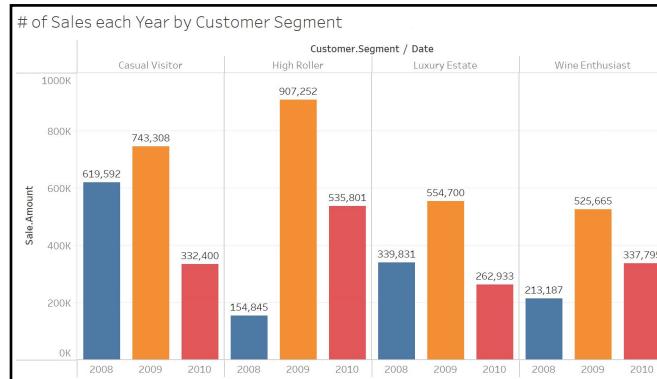
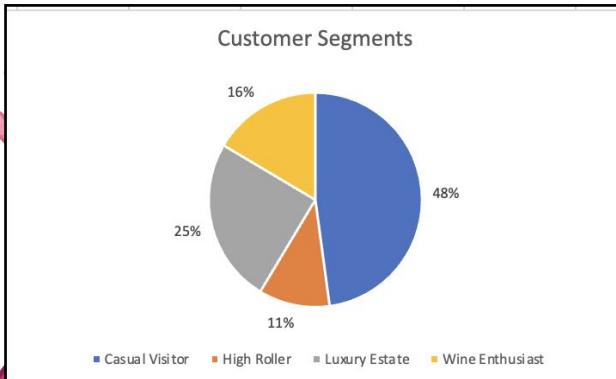
- Ran linear regression of each sales channel for each consumer segment

RFM Analysis

- Created Testing data using 2008 and 2009 data, training data using 2010 data
- Performed RFM Analysis using 2009 data
- Calculated response rate using 2010 data

Data Summary & Visualization

Basic Customer Segment Insights



Casual Visitors are the largest segment size (48%).

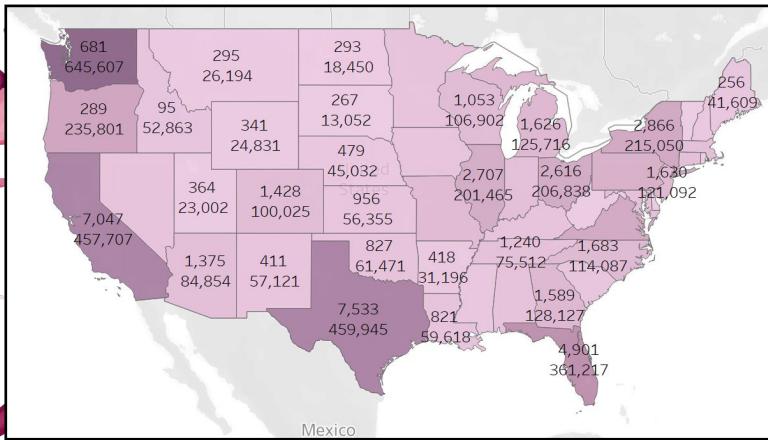
Wine Enthusiast make up 16% of our customer base but consistently spend the least.

Even though more customers were acquired in 2010, total sales revenue is going down.

High Rollers generate the most sales.

Data Summary & Visualization

Highest Spending by State



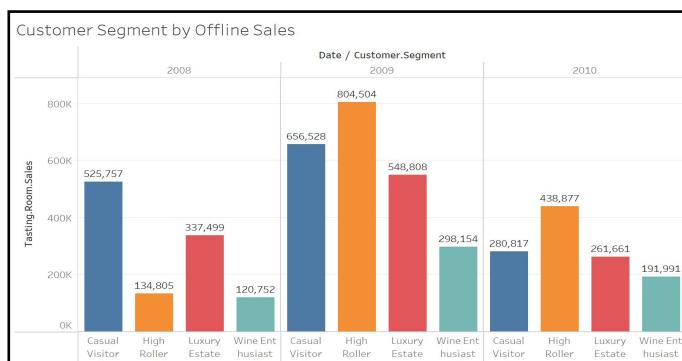
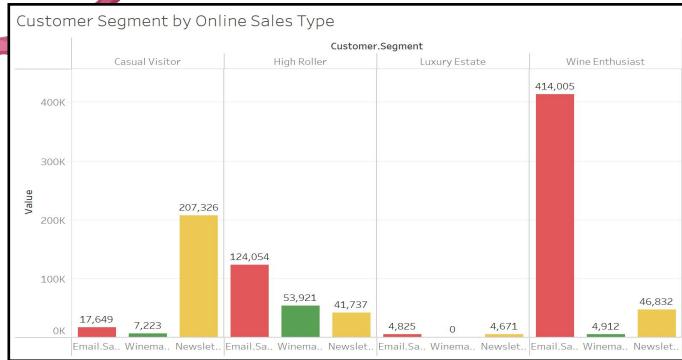
State	# of Customers	Total Sales Amount
WA	295	\$645,606.64
TX	2,515	\$459,944.89
CA	2389	\$457,706.58
FL	1661	\$361,216.8
OR	134	\$235,801.43

Washington (WA), Texas (TX), California (CA), Florida (FL), and Oregon (OR) are the 5 highest spending states for the winery

WA & OR have relatively fewer customers, but their total sales amounts are noteworthy, suggesting higher average spending per customer.

Data Summary & Visualization

Total Channel Sales by Customer Segment

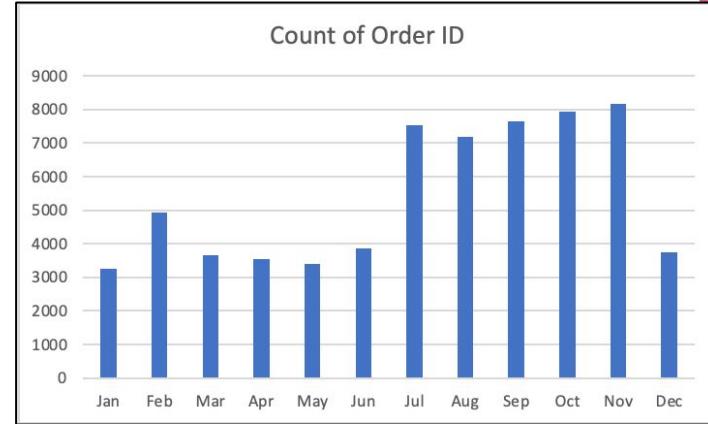
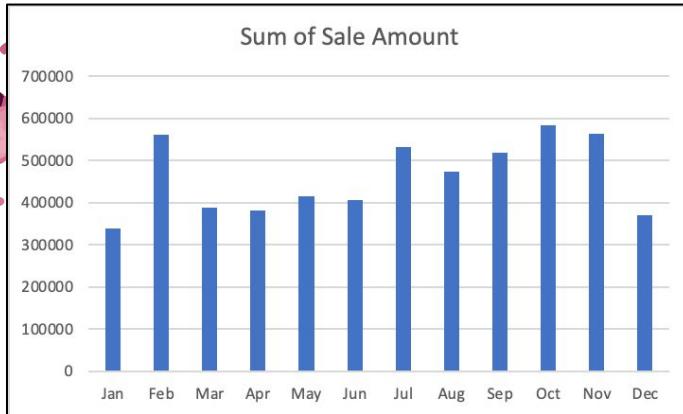


Customer Segment	Marketing Channel (Online/In Person)	
Casual Visitor	Newsletter	Tasting Room
High Roller	Email	
Luxury Estate	Email	
Wine Enthusiast	Email	

Assuming WineMaker Call started in 2010
The in-person Tasting Room is the highest generating sales channel

Data Summary & Visualization

Seasonal Sales



Total Sales Amount Peaks in February, despite low order counts.

People buy expensive wine for Valentines Day

Total Sales Amount and Sales Numbers are high in July - November, indicating wine bought for summer and holiday season.



Advertising Response Model

Coefficients	High Roller	Wine Enthusiast	Casual Visitor	Luxury Estate
Email Subscription	79.27	15.32	Not Significant	Not Significant
Newsletter Subscription	Not Significant	-18.83	-20.554	-9.72
Winemaker Calls	-78.27	-18.26	21.319	Not Significant

Email subscription is a great way to increase the amount of each purchase.

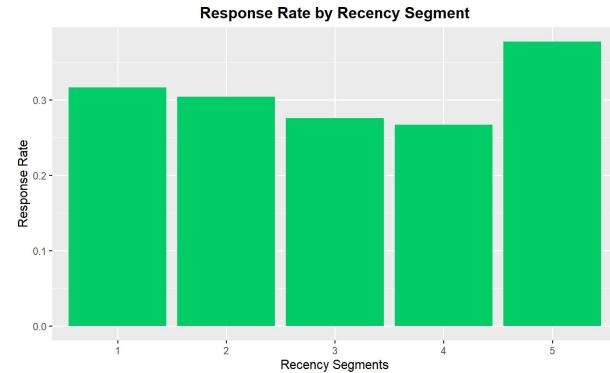
As for newsletter subscriptions, the subscription will cause customers to spend less on each purchase. The company can offer discount coupons to customers and encourage customers to purchase in-store.

For calling customers from winemakers, the company can focus more on calling casual visitors segments to earn more revenue from them on each purchase.

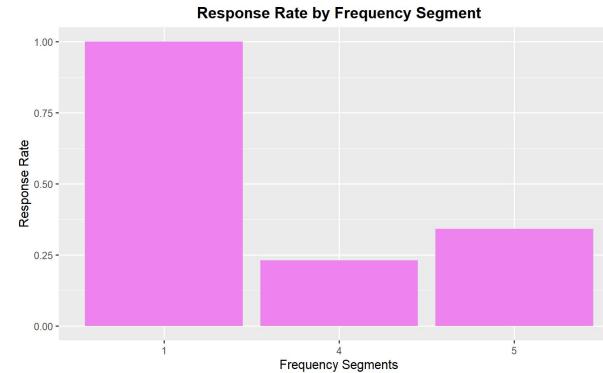


Key Findings

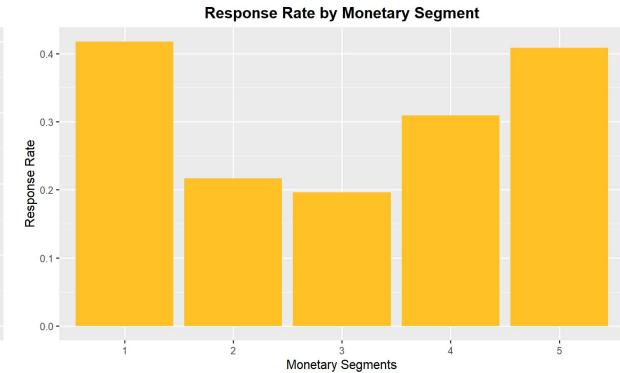
RFM Segments



Customers in 2008 and 2009 who purchase more recently are more likely to buy in the next year



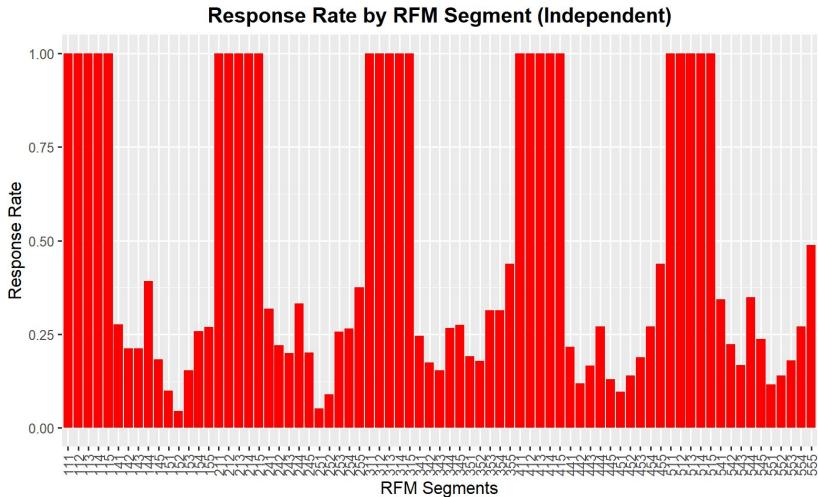
Customers who purchased less frequently in 2008 and 2009 are more likely to be a return buyer in 2010



Customers in 2008 and 2009 who purchased the lowest and highest total dollar amount and are more likely to buy in the next year.

Key Findings

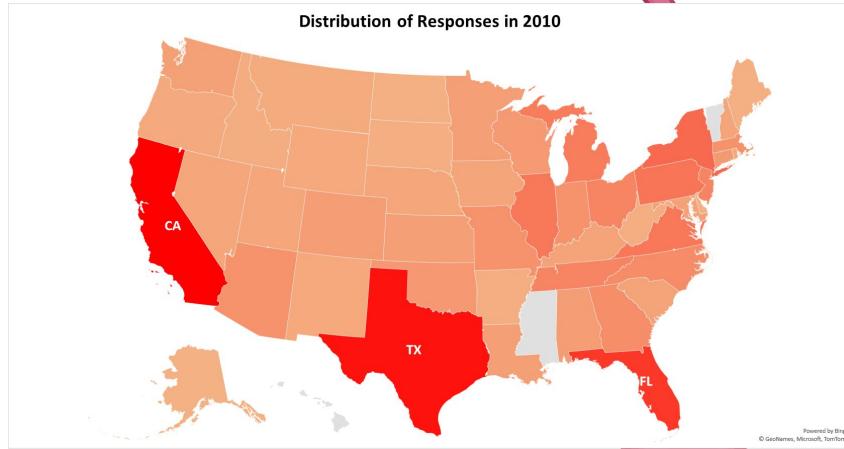
RFM Classifications



There are some segments are more likely to be a return buyer in the future

- 111-115
- 211-215
- 311-315
- 411-415
- 511-515

100% Response V.S. Location



The most loyal customers are from California, Texas and Florida

Implementation Recommendations



Target 5 highest spending States

- CA, TX, FL as Primary
- WA, OR as Secondary



Messaging Tactics

- Recent vs. non-recent buyers
- Holiday vs. non-holiday



Exclusive Emails

For High Rollers
and Wine
Enthusiasts



Different Wine Recommendations

For different
monetary segments



Get Wine Enthusiasts to Spend More

Email recommendations



Loyalty Programs/ Wine Club

For frequent purchasers

Possible Future Projects

Calculate Break-even RFM Response Rate

Only target consumers above the break even response rate (using cost data)



Return on Investment for Sales Channels

Using cost data to see which channels generate the highest returns



Propensity Scoring

To target certain consumers with specific marketing tactics



Purchase Propensity

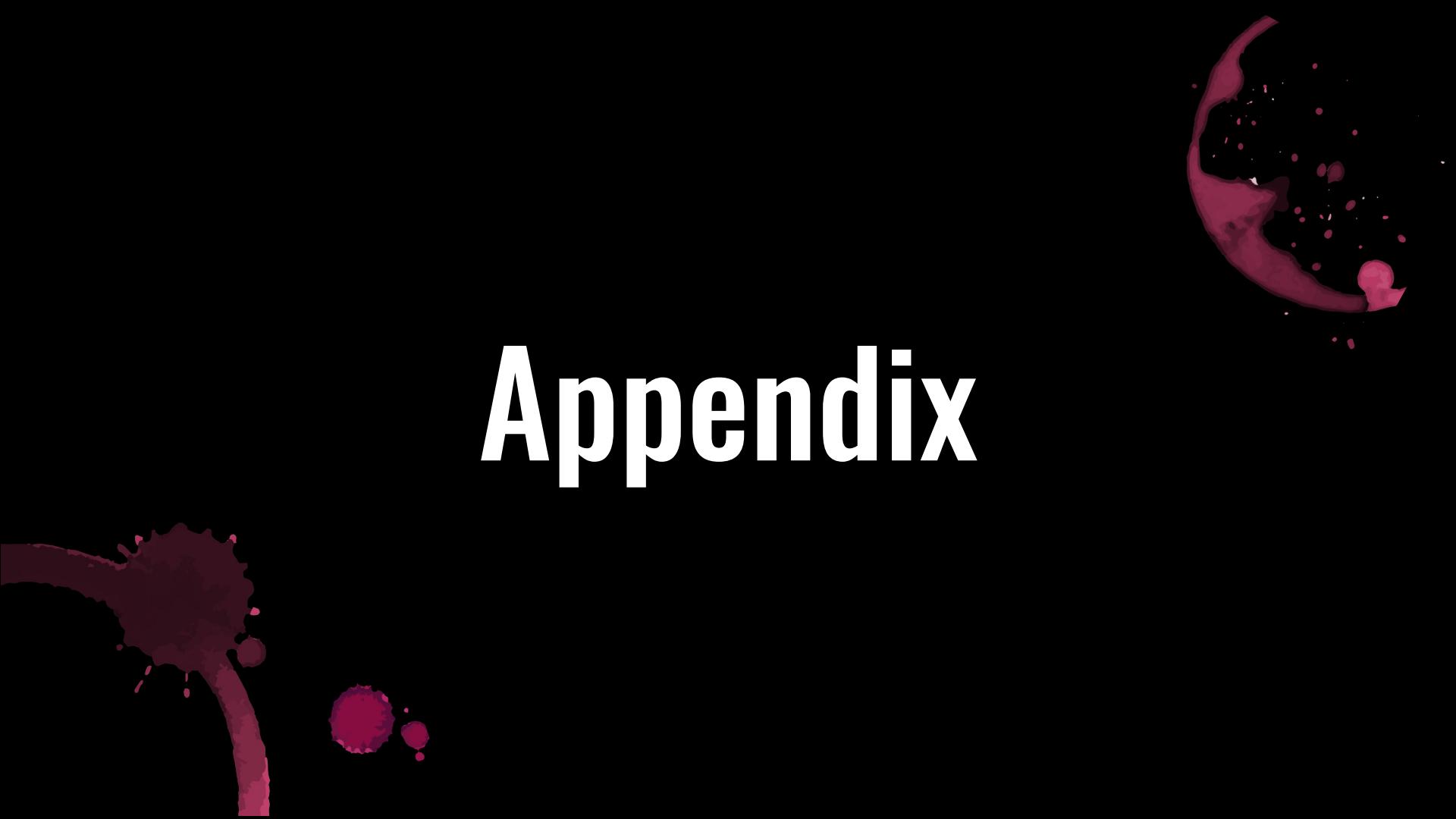
To determine which customers are most likely to buy





Thank You!

Appendix





Advertising Response Model - High Roller

```
##  
## Call:  
## lm(formula = Sale.Amount ~ Email.Subscr + Newsletter.Subscr +  
##     Winemaker.call, data = High_Roller_data_sep)  
##  
## Residuals:  
##      Min    1Q Median    3Q   Max  
## -245.2 -124.1  -96.1  -65.9 9280.4  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 164.116    9.408 17.444 < 2e-16 ***  
## Email.Subscr 79.279   17.075  4.643 3.48e-06 ***  
## Newsletter.Subscr -17.988   17.926 -1.003   0.316  
## Winemaker.call -78.279   16.743 -4.675 2.98e-06 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 513.1 on 9859 degrees of freedom  
## Multiple R-squared:  0.00377,    Adjusted R-squared:  0.003466  
## F-statistic: 12.44 on 3 and 9859 DF,  p-value: 4.1e-08
```





Advertising Response Model - Wine Enthusiast

```
##  
## Call:  
## lm(formula = Sale.Amount ~ Email.Subscr + Newsletter.Subscr +  
##   Winemaker.call, data = Wine_Enthusiast_data_sep)  
##  
## Residuals:  
##    Min      1Q Median      3Q     Max  
## -116.7   -60.7   -41.7   -11.7  6205.9  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  99.448    6.218   15.994 <2e-16 ***  
## Email.Subscr 15.323    8.702   1.761   0.0783 .  
## Newsletter.Subscr -18.834   8.763  -2.149   0.0316 *  
## Winemaker.call -18.266   7.800  -2.342   0.0192 *  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 294 on 12340 degrees of freedom  
## Multiple R-squared:  0.001992,  Adjusted R-squared:  0.00175  
## F-statistic: 8.211 on 3 and 12340 DF,  p-value: 1.862e-05
```





Advertising Response Model - Casual Visitor

```
##  
## Call:  
## lm(formula = Sale.Amount ~ Email.Subscr + Newsletter.Subscr +  
##   Winemaker.call, data = Casual_Visitor_data_sep)  
##  
## Residuals:  
##    Min      1Q Median      3Q     Max  
## -89.3   -41.5  -25.7   -1.5  7264.6  
##  
## Coefficients:  
##                               Estimate Std. Error t value Pr(>|t|)  
## (Intercept)           66.525     1.632  40.761 < 2e-16 ***  
## Email.Subscr        -4.307     4.881  -0.883   0.378  
## Newsletter.Subscr -20.554     3.521  -5.837 5.37e-09 ***  
## Winemaker.call       21.319     5.163   4.129 3.65e-05 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 222.9 on 27015 degrees of freedom  
## Multiple R-squared:  0.001564,  Adjusted R-squared:  0.001453  
## F-statistic: 14.1 on 3 and 27015 DF,  p-value: 3.508e-09
```





Advertising Response Model - Luxury Estate

```
##  
## Call:  
## lm(formula = Sale.Amount ~ Email.Subscr + Newsletter.Subscr +  
##     Winemaker.call, data = Luxury_Estate_data_sep)  
##  
## Residuals:  
##      Min      1Q Median      3Q     Max  
## -85.0   -51.5  -34.8   -4.8 7983.5  
##  
## Coefficients:  
##                               Estimate Std. Error t value Pr(>|t|)  
## (Intercept)            75.756     2.379  31.846 <2e-16 ***  
## Email.Subscr          2.952      8.208   0.360    0.719  
## Newsletter.Subscr   -9.728      5.407  -1.799    0.072 .  
## Winemaker.call        4.518      8.685   0.520    0.603  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 252.7 on 15621 degrees of freedom  
## Multiple R-squared:  0.0002075, Adjusted R-squared:  1.547e-05  
## F-statistic: 1.081 on 3 and 15621 DF, p-value: 0.3559
```





RFM Model - Response Segments

R	Response		Row Total
	0	1	
1	2841	1316	4157
	0.683	0.317	0.199
2	2906	1268	4174
	0.696	0.304	0.200
3	2989	1139	4128
	0.724	0.276	0.197
4	3058	1114	4172
	0.733	0.267	0.199
5	2671	1617	4288
	0.623	0.377	0.205
Column Total	14465	6454	20919





RFM Model - Frequency Segments

F	Response		Row Total
	0	1	
1	0	1159	1159
	0.000	1.000	0.055
4	10179	3073	13252
	0.768	0.232	0.633
5	4286	2222	6508
	0.659	0.341	0.311
Column Total	14465	6454	20919

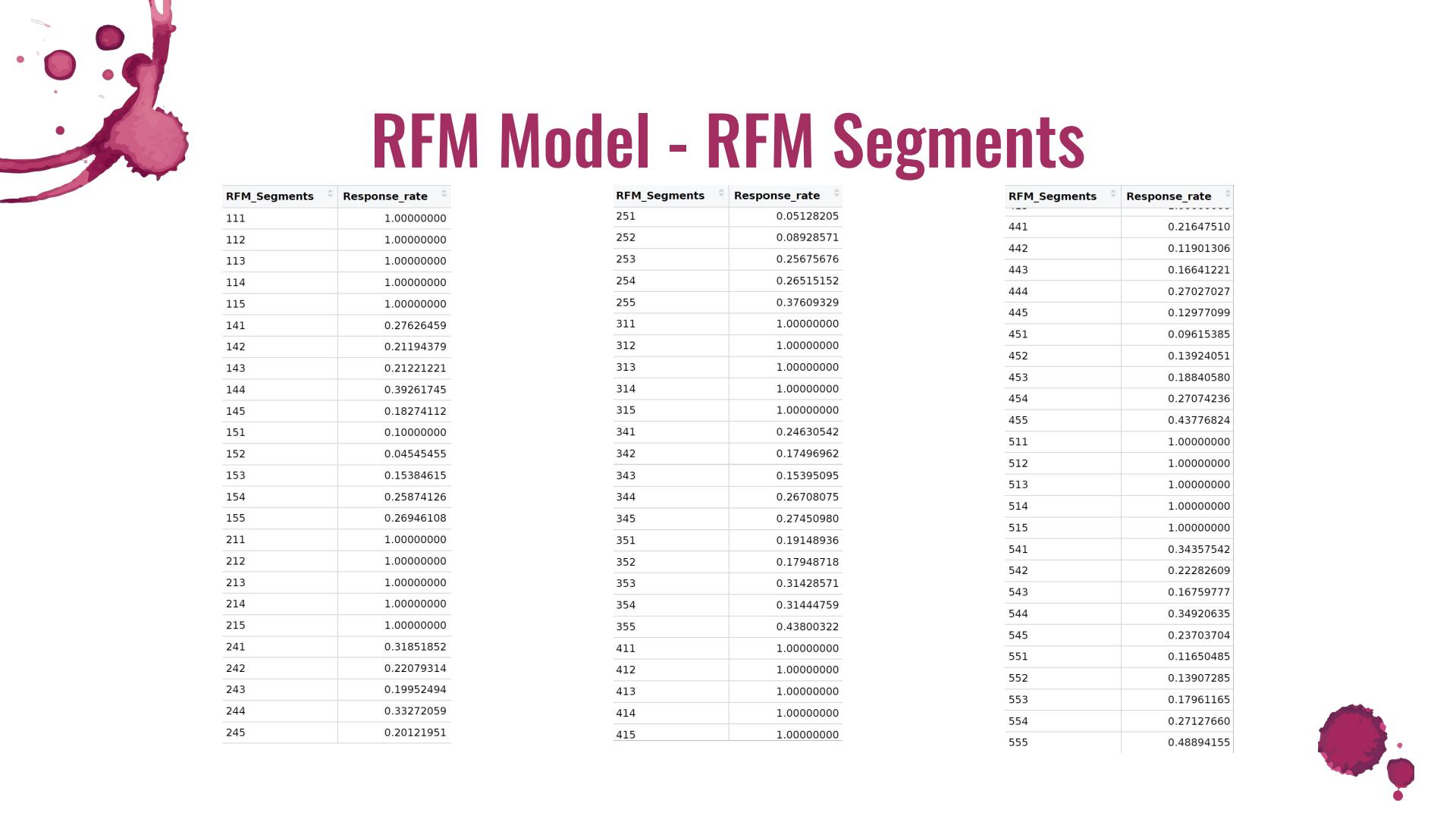




RFM Model - Monetary Segments

M	Response		Row Total	
	0	1		
1	53	9	62	
	0.855	0.145	0.003	
2	5861	341	6202	
	0.945	0.055	0.296	
3	3083	1848	4931	
	0.625	0.375	0.236	
4	3148	2392	5540	
	0.568	0.432	0.265	
5	2320	1864	4184	
	0.554	0.446	0.200	
Column Total	14465	6454	20919	





RFM Model - RFM Segments

RFM_Segments	Response_rate
111	1.00000000
112	1.00000000
113	1.00000000
114	1.00000000
115	1.00000000
141	0.27626459
142	0.21194379
143	0.21221221
144	0.39261745
145	0.18274112
151	0.10000000
152	0.04545455
153	0.15384615
154	0.25874126
155	0.26946108
211	1.00000000
212	1.00000000
213	1.00000000
214	1.00000000
215	1.00000000
241	0.31851852
242	0.22079314
243	0.19952494
244	0.33272059
245	0.20121951

RFM_Segments	Response_rate
251	0.05128205
252	0.08928571
253	0.25675676
254	0.26515152
255	0.37609329
311	1.00000000
312	1.00000000
313	1.00000000
314	1.00000000
315	1.00000000
341	0.24630542
342	0.17496962
343	0.15395095
344	0.26708075
345	0.27450980
351	0.19148936
352	0.17948718
353	0.31428571
354	0.31444759
355	0.43800322
411	1.00000000
412	1.00000000
413	1.00000000
414	1.00000000
415	1.00000000

RFM_Segments	Response_rate
441	0.21647510
442	0.11901306
443	0.16641221
444	0.27027027
445	0.12977099
451	0.09615385
452	0.13924051
453	0.18840580
454	0.27074236
455	0.43776824
511	1.00000000
512	1.00000000
513	1.00000000
514	1.00000000
515	1.00000000
541	0.34357542
542	0.22282609
543	0.16759777
544	0.34920635
545	0.23703704
551	0.11650485
552	0.13907285
553	0.17961165
554	0.27127660
555	0.48894155