Sample Description

The RFM analysis in this project is based on **3900+ rows** of transactions with **256 customers** from October 2006 to October 2018.

For each customer, the average number of transactions is about **15.28** times, and the average money spent is about **$3783.51** with **$247.6** for each transaction.

The dataset has a gender distribution that is more approach to the **male customers (209)** than the **female customers (47)**. But gender has a small impact on the average number of transactions and the money spent.

Gender Distribution

Spend per Trasaction

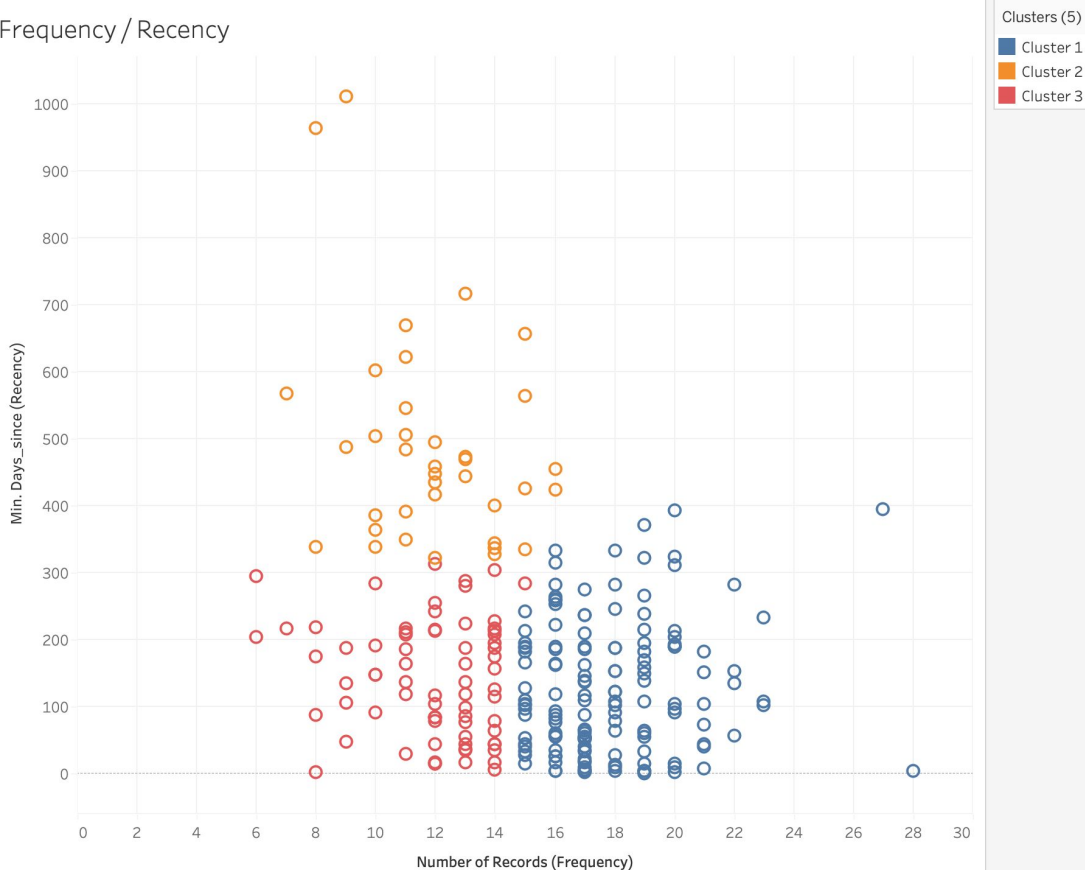Avg Transactions per Customer

Customer Sales

Total Sales
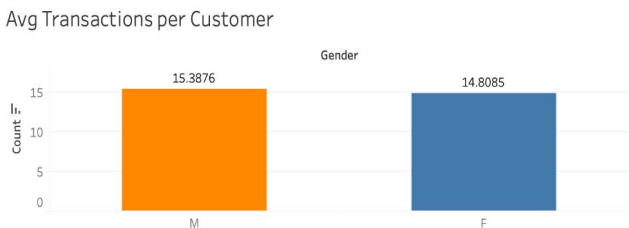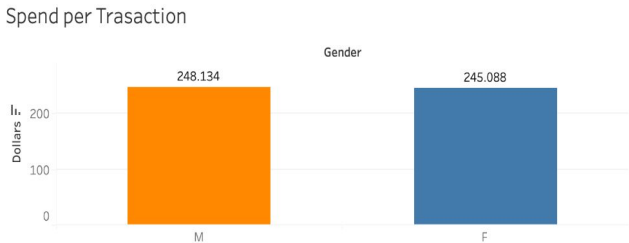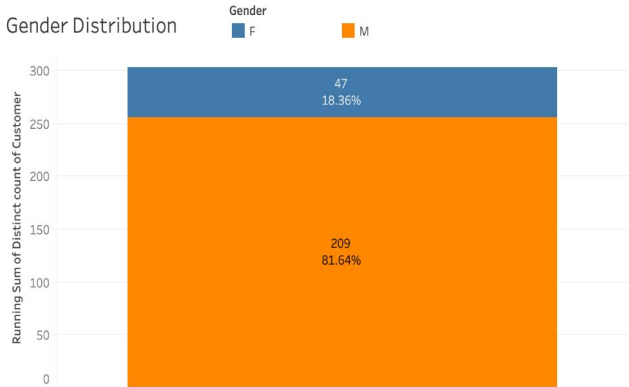
Segmentation/Clustering Solution

We finally decided to use **Frequency and Recency** to cluster customers.

In this final solution, the 256 customers have been grouped into 3 different clusters

Segmentation/Clustering - Clusters Description

Cluster1: Champions
85% of customers are male, purchase every 8 months and spend **$249** each time, the last purchase time was **124** days ago.

Cluster2: About to sleep
89% of customers are male, purchase every 11 months and spend **$244** each time, the last purchase time was **488** days ago.

Cluster3: Recent Customers
78% of customers are male, purchase every 11 months and spend **$246** each time, the last purchase time was **140** days ago.

Frequency / Recency

Clusters (5)
- Cluster 1
- Cluster 2
- Cluster 3

Regression- Predict the Future Trend

The **Total Size of Spent** for each customer can be predicted by our **Clusters** solution and **Gender** :

We will choose **Champions (Cluster 1)**, especially male customers, as the "key customers" to restrict the promotion for free trials.

Reasons:
Cluster 1(Intercept) : 4434.04 is the **base total money spent**. If the customer belongs to Champions (cluster 1) and is female, the predicted total money spent will be 4434.04.

Cluster 2 : -1529.59 represents a **decrease** in total money spent for the customer who is About to Sleep (cluster 2) with respect to customers who is Champions (cluster 1).

Cluster 3 : -1509.54 represents a **decrease** in total money spent for customers is Recent Customer (cluster 3) with respect to customers is Champions (cluster 1).

Gender : 1.23 represents an **increase** in total money spent for male with respect to female
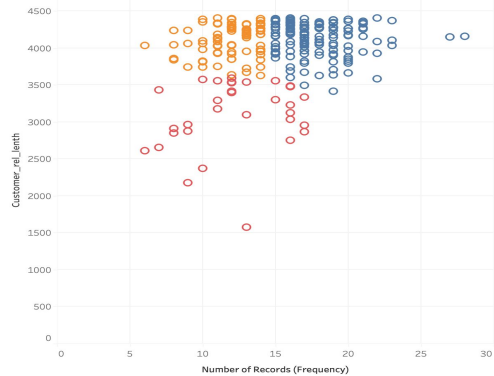
Conclusion

After multiple testing and predicting, we suggest choosing :

- Cluster Solution with Frequency and Recency

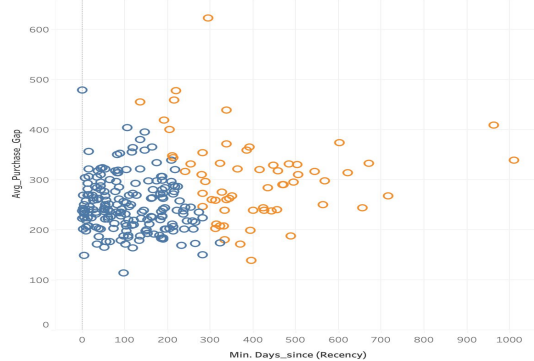- Champions (Cluster 1), especially male customers, as the key customers to restrict promotion

Thanks!
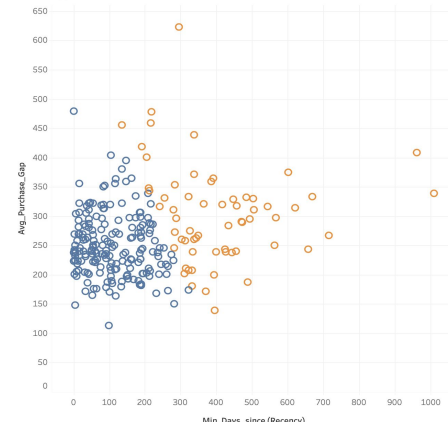
Appendix - Other Cluster Solutions
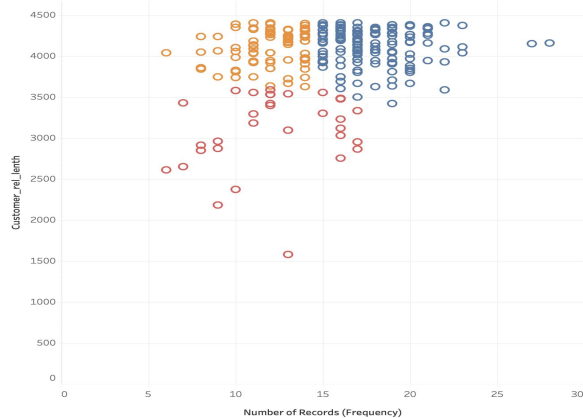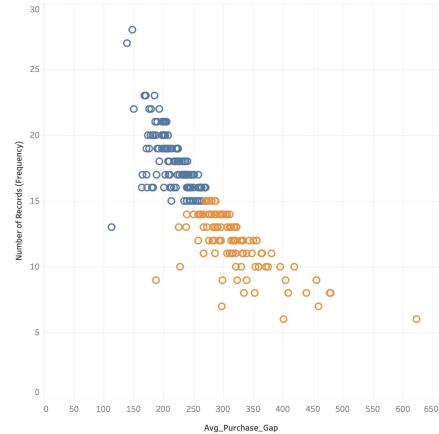
Appendix - Regression
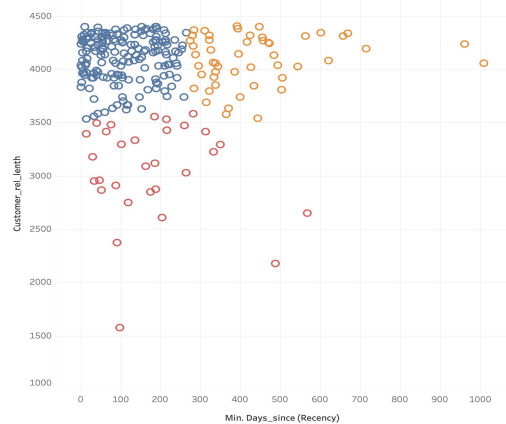
Best Regression Result with Total Money Spent:

Regression Statistics	
Multiple R	0.76664536
R Square	0.58774511
Adjusted R Square	0.58283731
Standard Error	633.671548
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	3	144262071	48087357	119.757437	3.1667E-48
Residual	252	101187987	401539.631		
Total	255	245450058			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4434.03828	102.635481	43.2018072	1.689E-118	4231.90566	4636.17089	4231.90566	4636.17089
Gender_M	1.22912604	103.878623	0.01183233	0.99056875	-203.351757	205.81001	-203.351757	205.81001
Frequency/Recency_Cluster 2	-1529.59398	116.714418	-13.1054416	2.8494E-30	-1759.45397	-1299.734	-1759.45397	-1299.734
Frequency/Recency_Cluster 3	-1509.54396	91.9418608	-16.4184621	1.0916E-41	-1690.61631	-1328.4716	-1690.61631	-1328.4716

Appendix - Regression

Best Regression Result with Average Money Spent:

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.07530976								
R Square	0.00567156								
Adjusted R Square	-0.00218874								
Standard Error	22.8829355								
Observations	256								
ANOVA									
	df	*SS*	*MS*	*F*	*Significance F*				
Regression	2	755.642947	377.821474	0.72154457	0.48699835				
Residual	253	132478.07	523.628736						
Total	255	133233.713							
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*	
Intercept	248.893307	1.89380522	131.424977	7.472E-235	245.163676	252.622939	245.163676	252.622939	
Frequency/Recency_Cluster 2	-4.49479073	4.21172485	-1.06720902	0.28689475	-12.7892976	3.79971619	-12.7892976	3.79971619	
Frequency/Recency_Cluster 3	-2.6557774	3.28016685	-0.80964705	0.41890388	-9.11568816	3.80413335	-9.11568816	3.80413335	

Appendix - Regression

Other Regression Result with Total Money Spent:

SUMMARY OUTPUT F_Cust_Length

Regression Statistics

Multiple R	0.74609728
R Square	0.55666114
Adjusted R Sq	0.5513833
Standard Error	657.126826
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	3	136632510	45544170	105.471324	2.9295E-44
Residual	252	108817548	431815.666		
Total	255	245450058			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4414.57745	105.691416	41.7685525	2.947E-115	4206.42641	4622.72849	4206.42641	4622.72849
Gender_M	26.6916848	106.695203	0.25016762	0.80266155	-183.43623	236.819603	-183.43623	236.819603
Frequency/Cu	-1501.198	91.635783	-16.382225	1.4567E-41	-1681.6676	-1320.7284	-1681.6676	-1320.7284
Frequency/Cu	-1376.0361	128.589326	-10.701013	2.8258E-22	-1629.2828	-1122.7894	-1629.2828	-1122.7894

SUMMARY OUTPUT F_Gap

Regression Statistics

Multiple R	0.76612889
R Square	0.58695347
Adjusted R Sq	0.58368828
Standard Error	633.024905
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	2	144067764	72033882	179.760897	2.6538E-49
Residual	253	101382294	400720.53		
Total	255	245450058			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4498.9931	103.196595	43.5963327	1.189E-119	4295.75929	4702.22691	4295.75929	4702.22691
Gender_M	-50.68869	102.977363	-0.4922314	0.62298263	-253.49074	152.113363	-253.49074	152.113363
Frequency/Av	-1513.7658	80.2171519	-18.870849	3.5167E-50	-1671.7442	-1355.7873	-1671.7442	-1355.7873

SUMMARY OUTPUT R_Gap

Regression Statistics

Multiple R	0.43277581
R Square	0.1872949
Adjusted R Sq	0.18087036
Standard Error	887.948546
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	2	45971544.9	22985772.4	29.1530167	4.0413E-12
Residual	253	199478513	788452.621		
Total	255	245450058			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3775.27656	130.965019	28.8266027	6.6383E-82	3517.35604	4033.19708	3517.35604	4033.19708
Gender_M	300.67758	144.115868	2.08635999	0.03794765	16.8579805	584.49718	16.8579805	584.49718
Recency/Avg	-979.57119	130.237834	-7.5214027	9.4451E-13	-1236.0596	-723.08278	-1236.0596	-723.08278

SUMMARY OUTPUT R_Cust_Length

Regression Statistics

Multiple R	0.38283652
R Square	0.1465638
Adjusted R Sq	0.13640385
Standard Error	911.731224
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	3	35974094.3	11991364.8	14.4256356	1.0571E-08
Residual	252	209475964	831253.824		
Total	255	245450058			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3799.46942	135.667043	28.0058395	2.4245E-79	3532.28371	4066.65513	3532.28371	4066.65513
Gender_M	288.411967	148.188041	1.9462567	0.05273536	-3.4328748	580.25681	-3.4328748	580.25681
Recency/Cust	-780.89982	146.05292	-5.3466909	2.0079E-07	-1068.5397	-493.25993	-1068.5397	-493.25993
Recency/Cust	-817.91273	180.284447	-4.5367903	8.8442E-06	-1172.9689	-462.8565	-1172.9689	-462.8565

Appendix - Regression

Other Regression Result with Average Money Spent:

SUMMARY OUTPUT F_Cust_Length

Regression Statistics	
Multiple R	0.1053003
R Square	0.01108815
Adjusted R Sq	0.00327067
Standard Erro	22.8205231
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	2	1477.31584	738.657918	1.41837859	0.24402472
Residual	253	131756.397	520.776274		
Total	255	133233.713			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	248.912625	1.91505576	129.976699	1.184E-233	245.141143	252.684107	245.141143	252.684107
Frequency/Cu	-4.8870679	3.16518068	-1.5440091	0.12383599	-11.120527	1.34639087	-11.120527	1.34639087
Frequency/Cu	1.11297755	4.46561274	0.24923289	0.80338286	-7.6815323	9.90748742	-7.6815323	9.90748742

SUMMARY OUTPUT R_Cust_Length

Regression Statistics	
Multiple R	0.14394616
R Square	0.0207205
Adjusted R Sq	0.01297916
Standard Erro	22.7091112
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	2	2760.66884	1380.33442	2.6766035	0.07074456
Residual	253	130473.044	515.703733		
Total	255	133233.713			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	248.808303	1.71664745	144.938498	1.82E-245	245.427564	252.189043	245.427564	252.189043
Recency/Cust	-7.787606	3.6136846	-2.155032	0.03210315	-14.904341	-0.6708704	-14.904341	-0.6708704
Recency/Cust	1.95833055	4.4874272	0.43640386	0.66291592	-6.8791404	10.7958015	-6.8791404	10.7958015

SUMMARY OUTPUT F_Gap

Regression Statistics	
Multiple R	0.07086346
R Square	0.00502163
Adjusted R Sq	0.00110439
Standard Erro	22.8453085
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	1	669.050461	669.050461	1.2819315	0.25860793
Residual	254	132564.663	521.908121		
Total	255	133233.713			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	248.934855	1.9171357	129.847279	3.017E-234	245.159349	252.710362	245.159349	252.710362
Frequency/Av	-3.252766	2.87289936	-1.1322241	0.25860793	-8.9105033	2.40497128	-8.9105033	2.40497128

SUMMARY OUTPUT R_Gap

Regression Statistics	
Multiple R	0.08833244
R Square	0.00780262
Adjusted R Sq	0.00389633
Standard Erro	22.8133596
Observations	256

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1039.57192	1039.57192	1.99745061	0.15878833
Residual	254	132194.141	520.449375		
Total	255	133233.713			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	248.625564	1.63790376	151.794977	2.96E-251	245.399962	251.851166	245.399962	251.851166
Recency/Avg_	-4.7038184	3.32822378	-1.4133119	0.15878833	-11.258248	1.85061094	-11.258248	1.85061094