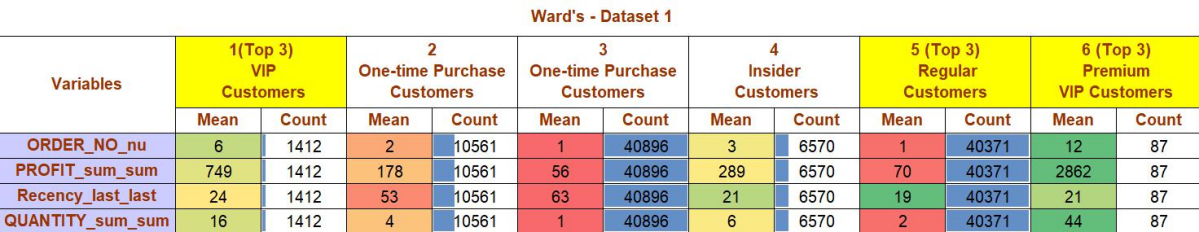
**Data Analysis**

1. **Identify Segments**

We chose Segment 1 (VIP Customers) and Segment 5 (Regular Customers) to conduct the analysis (see Figure 1.). The reasons were: (1) Their sample sizes were big enough (> 100); (2) They had little similarities with each other: VIP Customers was a small segment with more profitable users, while Regular Customers was a much larger segment with less profitable users.

***Figure 1. Final Six Segments and Top 3 Segments of Project 2***



1. **Choosing Variables**

Dependent Variable: PROFIT\_sum\_sum

Independent Variable: REVENUE\_sum\_sum, ORDER\_NO\_nu, Recency\_last\_last, QUANTITY\_sum\_sum, PAY\_METHOD\_last

1. **Dummy Coding Variable: PAY\_METHOD\_last**

AX - American Express, MC - MasterCard, VI - VISA, OT - Other (We recoded other credit cards into a new variable “OT”). After we finished dummy coding the ‘PAY\_METHOD\_last’ variable into 3 variables (Number of variables = Number of categories 4 - 1) and used ‘PROFIT\_sum\_sum’ to ran regression analysis with them, we found that only ‘AX’ variable was significant (see Figure 2. in Appendices).

1. **Original Regression Model and Hypothesis**

Original Regression Model：

**PROFIT\_sum\_sum = β0 + β1·REVENUE\_sum\_sum + β2·ORDER\_NO\_nu +**

**β3·****Recency\_last\_last + β4·QUANTITY\_sum\_sum + β5·AX + ε;**

Null and Alternative Hypothesis:

H0: β1 = β2 = β3 = β4 = β5 = 0 (No relation with PROFIT\_sum\_sum, PROFIT\_sum\_sum = β0 + ε)

H1: At least one βi is not equal to zero or not all βi’s are equal to zero.

1. **Calibration and Run Regression Model in Segment 1 (VIP Customers) and Segment 5 (Regular Customers)**
2. Set filter to choose 60% of Segment 1’s and Segment 5’s data (two steps, selected 60% of all cases and then selected cases if ‘(filter1\_calibration=1)&(K\_meanswards6\_1=1) / (filter1\_calibration=1)&(K\_meanswards6\_1=5)’;
3. Ran Regression model and removed insignificant independent variables. ANOVA Sig. = .000 (see Figure 3. and Figure 4 in Appendices), very strong to reject null hypothesis. So either β1, β2, β3, β4, β5 ≠ 0, and independent variable have relations with dependent variables;
4. Adjusted R squares were 93.30% and 92.30% (see Figure 5. and Figure 6. in Appendices) and ‘REVENUE\_sum\_sum’ of Segment 1 contribution was very high, we removed it and found ‘ORDER\_NO\_nu’ and ‘Recency\_last\_last’ were not significant (See Figure 7. in Appendices) in Segment 1. Eventually we removed ‘ORDER\_NO\_nu’ and ‘Recency\_last\_last’ from Segment 1’s regression model and kept all independent variables from Segment 5’s regression model (all independent variables were significant, even after we removed ‘REVENUE\_sum\_sum’ and ran the regression analysis);
5. For Segment 1’s regression mode1, ‘REVENUE\_sum\_sum’ had a stronger effect on the dependent variable; for Segment 5’s regression model, ‘REVENUE\_sum\_sum’ and ‘’AX had stronger effects on the dependent variable.
6. **Regression Assumptions and Issues (Residuals & Multicollinearity & Outliers)**
7. Both two segments’ Residual Mean were 0.000 (see Figure 8. and Figure 9. in Appendices);
8. Independent variables’ VIF from two segments were all less than four (see Figure 10. and Figure 11. in Appendices);
9. Checked LEV Values (see Figure 12. in Appendices), removed data that >.05 (outliers) in Segment 1; checked data that was .02 < LEV < .05 and ran regression analysis with and without these data, we did not have to remove these data since no significant changes occurred in Segment 1. Segment 5 had no outliers (see Figure 13. in Appendices).
10. **Finalized all independent variables and Ran Final Regression Analysis**

Segment 1 Final Regression Model: (see Figure 14. in Appendices)

**Profit = 38.52 + 0.48·REVENUE\_sum\_sum + 1.86·QUANTITY\_sum\_sum + 17.39·AX**

Segment 5 Final Regression Model: (see Figure 15. in Appendices)

**Profit = 0.92 + 0.49·REVENUE\_sum\_sum + 1.70·ORDER\_NO\_nu - 0.20·Recency\_last\_last + 3.44·QUANTITY\_sum\_sum + 0.58·AX**

1. **Validation and Prepared for further managerial implications**
2. Set filter to choose 40% of Segment 1’s and Segment 5’s data (two steps, selected 40% of all cases and then selected cases if (filter1\_calibration=0)&(K\_meanswards6\_1=1)/ / (filter1\_calibration=0)&(K\_meanswards6\_1=5);
3. Computed variable using final regression model and compared with original profit variable in data view for Segment 1 and Segment 5 (see Figure 16. and Figure 17. in Appendices);
4. For Segment 1 and Segment 5, used original profit and predicted profit to compute SSE and SSR in Excel, and computed the R Square values (SST = SSE + SSR, R Square = SSR / SST) to see if the two models were good fit (see Figure 18. in Appendices).
5. **Summary of General Managerial Implications for Segment 1 and Segment 5**
6. For VIP Customers, every dollar increases in revenue, the profit will increase 0.48 dollars; every unit increases in quantity, the profit will increase 1.86 dollars; customers using American Express credit card as a payment method will increase 17.39 dollars profit, compared to other credit cards users. Overall, 93.30% of the profit could be explained by our regression model.
7. For Regular Customers, every dollar increases in revenue, the profit will increase 0.49 dollars; every unit increases in purchase order, the profit will increase 1.70 dollars; every month increases in recency, the profit will decrease 0.20 dollars; every unit increases in quantity, the profit will increase 3.44 dollars; customers using American Express credit cards as a payment method will increase 0.58 dollars profit, compared to other credit cards users. Overall, 92.30% of the profit could be explained by our regression model.