James Lee

Jet.com Case Assignment

05/09/2017

# Contents

## Executive Summary

To measure which vendors are providing the best prices, we examine two different methods. Mean percentage deviation from average prices of products, and the same measure weighted by price * sales to better measure the magnitude of the impact the prices have on Jet's overall cost. Analysis shows that overall, **Jasmine's Shop offers prices that are 4.36% below the average of three vendors**. **However, once we take price*sales i.e. cost to Jet.com into consideration, Leo's bodega offers the best prices, offering 5% lower weighted prices than its competitors.**

To measure which products are most competitively priced, we use standard deviation at a SKU level. Products with the widest spreads are considered least competitively priced, and products with the smallest spreads are considered most competitively priced. We also look at scatterplots to visually scan any trends**. Electronics product low in sales are least competitively priced, and consumable products high in sales are most competitively priced.**

We also investigate the competitiveness of the vendors in each category by applying the same percentage deviation methods but aggregated at the category level. **Leo's Bodega offers the best prices in Consumables and Home while Jasmine's offers the best prices in Electronics.**

Upon further investigation of the scatterplots, we can see that each category has its own clusters of products that share similar characteristics. We investigate these clusters on a Sales level. W**e can see that Leo's competitive advantage comes from its low prices in high sales products across all three categories, but it isn't competitive in low sales products. This also explains why Leo's seem to offer the worst prices in unweighted average method.**

We can calculate the weighted average percentage deviations for each of these clusters. Below table is a summary of the analysis, and it tells a much different story than before.

| Category | Sales Cluster | Most Competitive Vendor | Least Competitive Vendor |
|---|---|---|---|
| **Electronics** | <=200 | Jasmine's | Leo's |
| **Electronics** | >200 | Leo's | Alex's |
| **Home** | <=350 | Alex's | Jasmine's |
| **Home** | >350 | Leo's | Jasmine's |
| **Consumables** | <500 | Alex's | Leo's |
| **Consumables** | 500-600 | Jasmine's | Leo's |
| **Consumables** | 1000-1100 | Alex's | Jasmine's |
| **Consumables** | >1100 | Leo's | Jasmine's |

**In summary, If Jet can pick and match vendors, it should select based on cluster analysis shown above. In a mutually exclusive scenario, Leo's should be chosen based on overall best pricing. However, Jet should be cautious on the fact that Leo's offers poor prices in other clusters.**

We also investigate K-means clustering method as a more mathematically sound approach in determining these sales clusters.

## Data Adjustments

1. These data points exhibit unusual characteristics that are not likely caused by variability. Therefore, they are removed from our dataset.
    1) 0 dollar prices
    2) (-) prices
    3) 99,999.00 prices
    4) Data points with (-) sales for certain questions

2. For questions where using sales and category mapping were appropriate, data points without a corresponding mapping on either category or the sales were removed.

## 1. Overall, How Are the Three Merchants Priced Relative to Each Other? Which Merchant Has the Best Pricing?

### a) Percentage Deviations from the Mean Price Approach

For each product and merchant, I looked at percentage deviation of its price from the product's mean price. I used R for the calculation because of its robust ability to do column and aggregation calculations. Code is attached in the appendix.

Then to look at the deviations on a merchant level, I took an average of all the percentage deviations by merchant.

$$\text{For each Merchant,}$$
$$\frac{1}{\#\ of\ Products}\sum_{i=1}^{\#\ of\ Products}\frac{Merchant's Price\ for\ i - Average\ of\ All\ the\ Merchants' Prices\ for\ Product\ i}{Average\ of\ All\ the\ Merchants' Prices\ for\ Product\ i}$$

Result:

| Merchant | Average Price Deviation | Rank |
|---|---|---|
| **Alex's Store** | -0.02410051 (-2.4%) | 2 |
| **Jasmine's Shop** | -0.04413444 (-4.4%) | 1 |
| **Leo's Bodega** | 0.06864556 (6.9%) | 3 |

**Result essentially tells us that on average, Jasmine's shop's prices are 4.36% below the product average price of all the vendors.** By this measure, Jasmine's Shop has the best consistent pricing across the board without taking Jet's sales into context.

**On the other hand, Leo's Bodega's prices are 6.84% above the average, and has the worst pricing in this context.**

### b) Weighted Percentage Deviations from the Mean Price Approach

However, above analysis is biased in a sense that we want to identify vendors that provide the cheapest prices in products that drive bigger portion of Jet's costs. So, it makes sense to weight these averages by how much impact each product has on Jet's costs, and calculate the weighted average. Below is an output once we take into consideration of the weights. I used Sales * Price i.e. total cost to Jet as weights.

| Merchant | Weighted Average Price Deviation | Rank |
|---|---|---|
| **Alex's Store** | 0.004503171 (.45%) | 3 |
| **Jasmine's Shop** | 0.071889696 (7.2%) | 2 |
| **Leo's Bodega** | -0.050365907 (-5.0%) | 1 |

**What the result is showing is that while Leo's Bodega offers the worst prices in pure average sense, it offers the cheapest prices in the products that have the most impact on Jet's overall cost.** In this sense, Leo's Bodega provides the best prices among all the vendors. **Also in contrast, Jasmine's Shop provides the worst pricing in this context. This can mean that Leo's selectively tries to make themselves more competitive in high demand products. It can also mean that Leo's infrastructure is better able to capitalize on economies of scales vs its peers.**

## 2. What Products Have the Most Competitive Pricing? Which Ones Have the Least Competitive Pricing?

### a) Standard Deviation Approach

I characterized competitively priced products as products having smaller standard deviation of prices. Intuition is that heavy competition and undercutting lead to prices that have a very small spread.

Bottom 3 in standard deviations, most competitively priced.

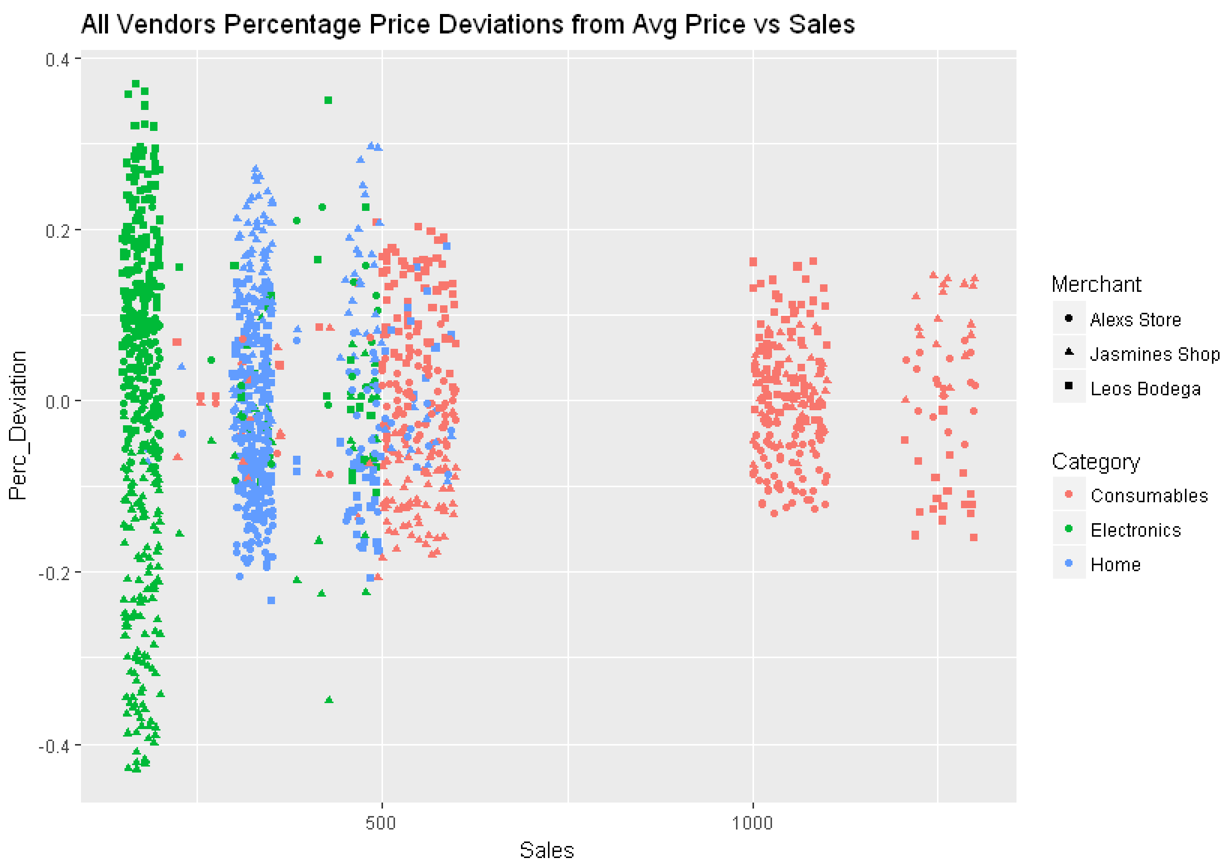
| JET SKU ID | Stdev |
|---|---|
| **J0149** | 0.03716455 |
| **J0412** | 0.06855854 |
| **J0283** | 0.10296265 |

Top 3 in standard deviations, least competitively priced.

| JET SKU ID | stdev |
|---|---|
| **J0126** | 1110.990414 |
| **J0127** | 350.021420 |
| **J0147** | 279.19833 |

While the output answers the question, to make a more applicable analysis, we should look at the data not on a product level but on a category level as shown in an image below.

### b) Scatterplot Analysis on Price Competition by Categories




**What the image is essentially telling us is that, the spreads of prices, i.e. competitiveness of pricing, depend on 2 factors – 1) Category 2) Sales. Consumables products and products that have high sales tend to have the most competition. Electronics, especially electronics with low sales, have the least competition. This may mean that the vendors try to undercut each other's prices more often in high demand products to try to capture higher market share.**

## 3. What Else Could We Learn from this Data?

### a) Competitive Advantage of Vendors in Different Categories

If the vendors are not mutually exclusive, Jet can potentially contract them by product categories. We apply the same weighted mean percentage deviation approach but do so on a category level. Result is shown below.

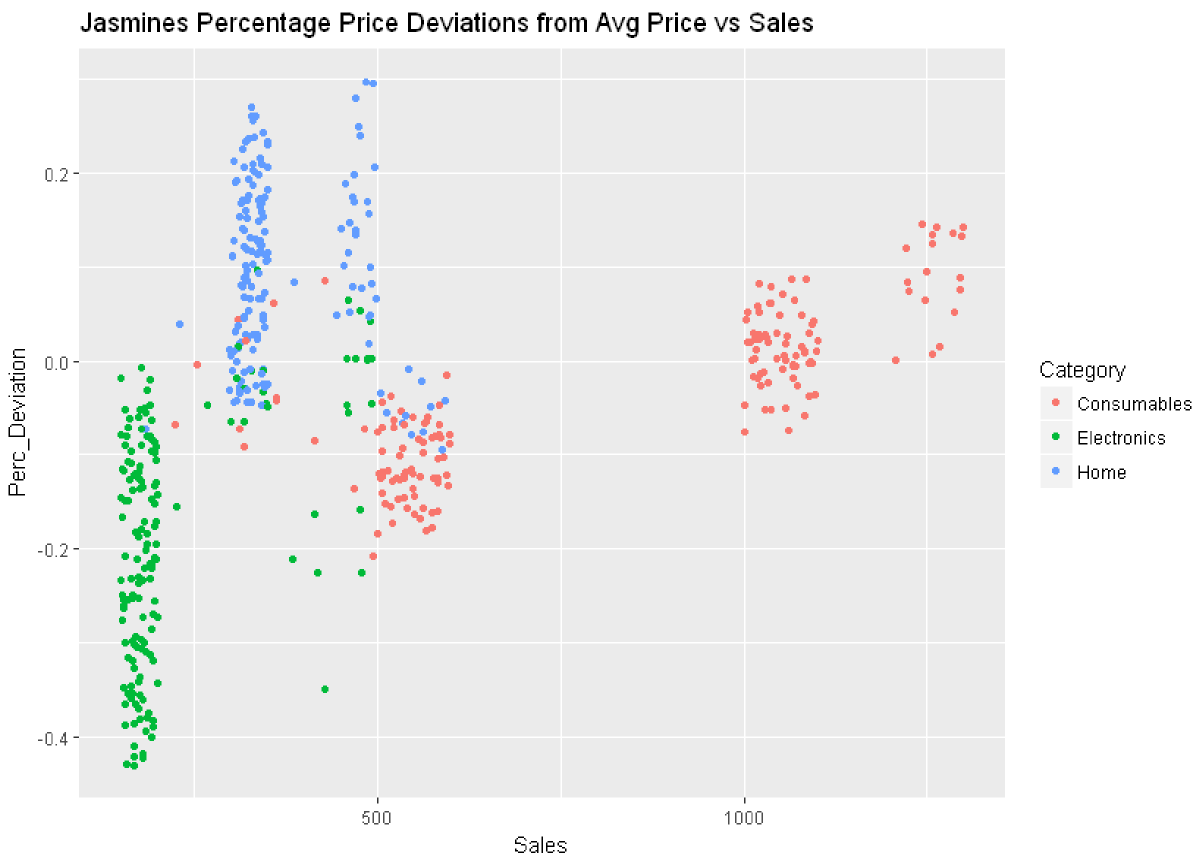
| Category | Merchant | Mean | Weighted Mean |
|---|---|---|---|
| **Consumables** | Alex's Store | -2.87% | .03% |
| **Consumables** | Jasmine's Shop | -3.4% | 4.9% |

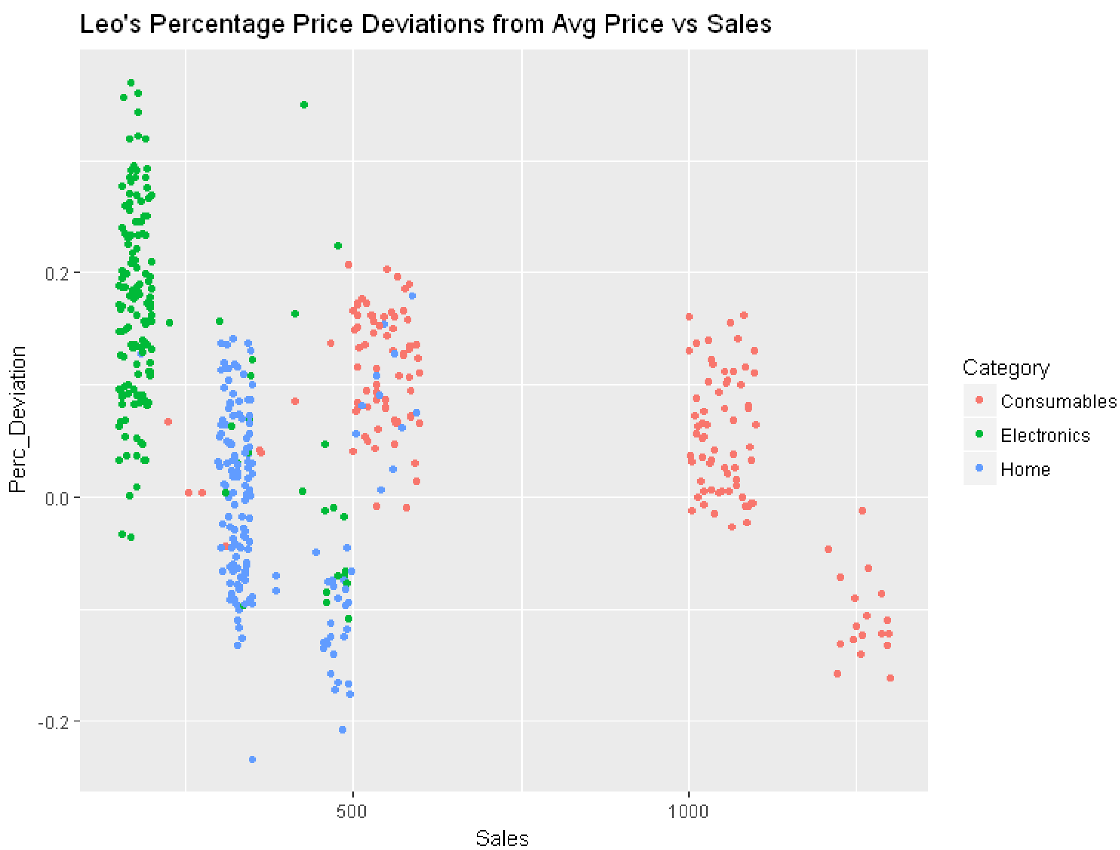
| | | | |
|---|---|---|---|
| **Consumables** | Leo's Bodega | 6.3% | -3.4% |
| **Electronic** | Alex's Store | 4.6% | 9.9% |
| **Electronic** | Jasmine's Shop | -19.0% | -4.8% |
| **Electronic** | Leo's Bodega | 14.9% | .3% |
| **Home** | Alex's Store | -9.1% | -4.5% |
| **Home** | Jasmine's Shop | 9.4% | 16.4% |
| **Home** | Leo's Bodega | -.73% | -11.6% |

**When taking Jet's sales into consideration, Leo's Bodega has the most competitive pricing in consumables and Home while Jasmine's has the most competitive pricing in Electronics**

## b) Investigating Clusters of Products Based on Sales

While above analysis reveals interesting summary on vendors' competitiveness on a category level, when plotting price deviations vs sales, we can see that **there are clusters of products within categories that exhibit different characteristics. Plots shown below also complement analysis done in part a). You can clearly see each vendor's competitive advantages in different categories.**




Leo's Percentage Price Deviations from Avg Price vs Sales
Perc_Deviation
0.2
0.0
-0.2
500
1000
Sales
Category
Consumables
Electronics
Home


Alex's Store Percentage Price Deviations from Avg Price vs Sales
Perc_Deviation
0.2
0.1
0.0
-0.1
-0.2
500
1000
Sales
Category
Consumables
Electronics
Home

As we can see that within Consumables category, there are approximately three different clearly identifiable clusters each having different spreads of price deviations from the mean.

We can do the same weighted average deviation calculation as before but broken down by these visually apparent clusters. Results are summarized below.

| Category | Sales Cluster | Most Competitive Vendor | Least Competitive Vendor |
|---|---|---|---|
| **Electronics** | <200 | Jasmine's by wide margin | Leo's by wide margin |
| **Electronics** | Others | Leo's | Alex's by wide margin |
| **Home** | <350 | Alex's store by a wide margn | Jasmine's by wide margin |
| **Home** | >350 | Leo's by a wide margin | Jasmine's by a wide margin |
| **Consumables** | <500 | Alex's on a weighted basis. Jasmine's on pure price basis | Leo's by wide margin |
| **Consumables** | 500-600 | Jasmine's Shop by a wide margin | Leo's by wide margin |
| **Consumables** | 1000-1100 | Alex's Store | Jasmine's Shop on a weighted basis. Leo's on a pure % price basis |
| **Consumables** | >1100 | Leo's Bodega | Jasmine's shop by wide margin |

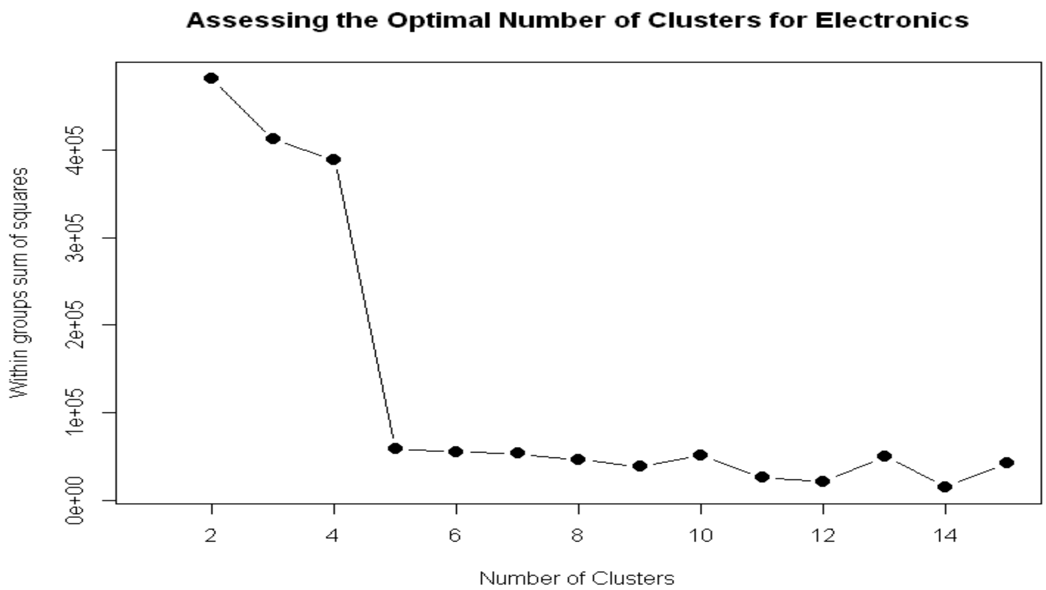
Key take away is that on an overall basis, Leo's showed strong pricing weighted by price*sales. **If Jet can pick and match vendors selectively, it should select based on clusters shown above. In a mutually exclusive scenario, Leo's should be chosen based on overall best pricing. However, Jet should be cautious on the fact that Leo's offers poor prices in other clusters. Choosing Leo's may potentially lead to unhappy customers who are looking to purchase niche products.**

## c) Using K-Means Clustering for More Mathematically Sound Determination of Clusters

A problem with the analysis above is that the clusters are chosen arbitrarily based on a visual scan. While on a larger scale, clusters are very apparent, there are also products that are not very clear which group it should fall into. For this reason, we can apply a k-clustering algorithm to come down to a more mathematically sound grouping.

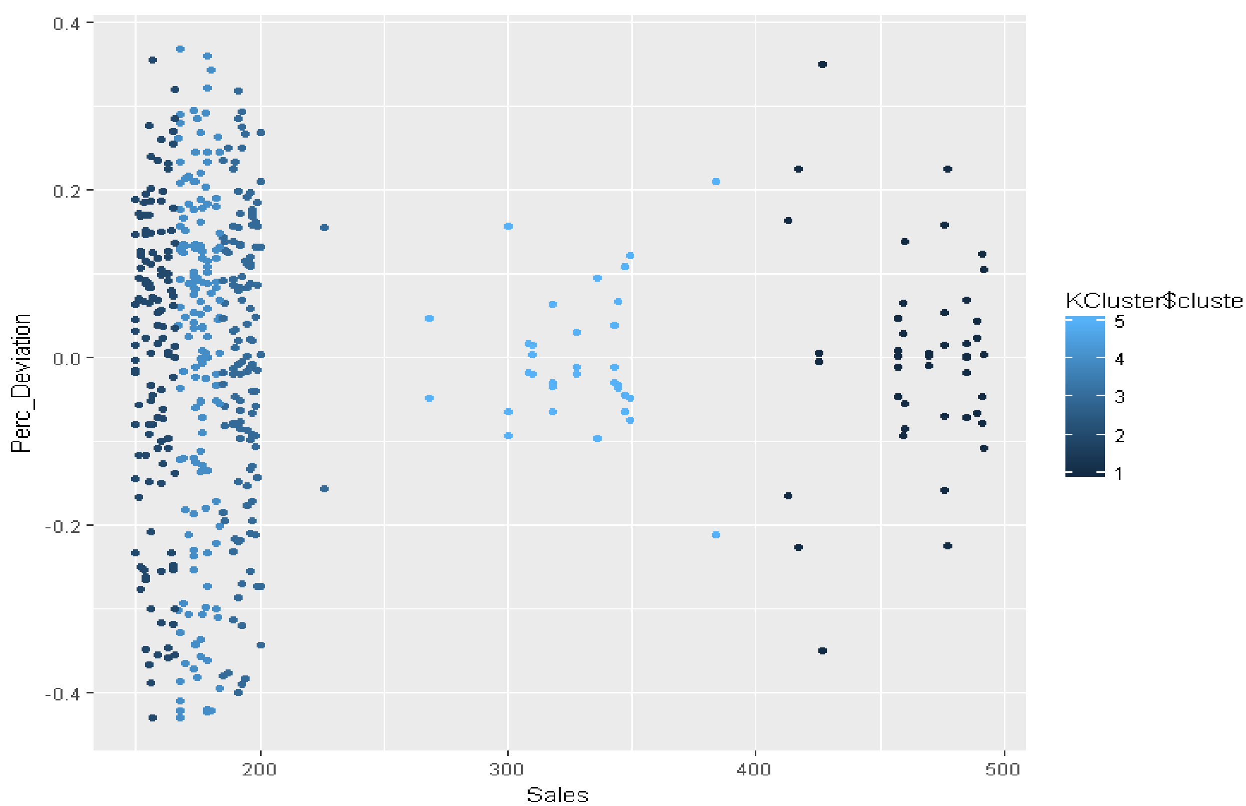
K-means algorithm will choose clusters in such a manner that it will minimize the sum of squares. I.e. it will choose number of clusters that will minimize the distance between the clusters and the data points that belong in the clusters. We apply the method at a sales level to determine mathematically sound clusters.

First, we investigate the right number of clusters by plotting sum of squares for different number of clusters for Electronics category. We want to minimize the sum of squares as much as possible. In that sense, we will be minimizing the distance between data points from the clusters. This indicates that the model is a tighter fit.




Surprisingly, clusters of 2 and 3, which was used originally, lead to very poor sum of squares. We can see that 14 clusters are optimal to reduce the within-clusters sum of squares as much as possible. However, a cluster number of 5 should be good enough without overfitting the data too much.

So, we choose clusters of 5 and compare our vendor's competitiveness within each cluster. Below is a plot of the original scatterplot but colored by the clusters the data points belong to.



Below is a summary of the K-cluster analysis and which vendors were most competitive in each cluster. Competitiveness is measured once again by the weighted average method.

| Category | Cluster Center | Most Competitive | Least Competitive |
|---|---|---|---|
| **Electronics** | 158.844 | Jasmine's | Leo's |
| **Electronics** | 175 | Jasmine's | Leo's |
| **Electronics** | 193.488 | Jasmine's | Leo's |
| **Electronics** | 327.79 | Leo's | Alex's |
| **Electronics** | 465.35 | Leo's | Alex's |

In this case, while our K-means algorithm identified 3 more clusters**, it didn't change our analysis that much**. Leo's is still the most competitive in pricing in higher Sales Electronics products while Jasmine's is still the most competitive in lower sales Electronics products. Our analysis of clusters in other categories basically tell the same story, **but it is reassuring that our original model is not too far off from a mathematically sound model.** Therefore, I will not further explain my analysis, but attach the rest of the analysis in the appendix.

# Appendix

## a) R Code Used for the Case

```
library(readxl)
library(plyr)
library(ggplot2)

##Import Combined Data
Merchant_Data <- read_excel("C:/Users/James/Desktop/Programming/R
Environment/Analytics_Interview_Case_-_With_Sales_Data.xlsx",
                sheet = "Merchant Data")

#Get mean price for each JET SKU ID
SKU_ID_Avg <- setNames(aggregate(Merchant_Data[, 4],Merchant_Data[,3],mean ),c("Jet SKU ID",
"Average_Price"))

#Combine average prices into the original Table
Merchant_Data_With_Avg <- (merge(Merchant_Data,SKU_ID_Avg, by = 'Jet SKU ID'))

#Compute % deviations of prices from the means
Merchant_Data_With_Avg_Ndeviations <- within(Merchant_Data_With_Avg,
                              Perc_Deviation <-
                                (Price - Average_Price)/Average_Price)
#ANSWER TO #1
##Calculate Average deviations by Merchant
Merch_Perc_Deviation_Avg <-
setNames(aggregate(Merchant_Data_With_Avg_Ndeviations$Perc_Deviation,
                          by= list(Merchant = Merchant_Data_With_Avg_Ndeviations$Merchant),
                          mean),c("Merchant", "Average_Price_Deviation"))
```

```
#1.b Weighted Average by Average Price * Sales
#Calculate Weights
Merchant_Data_With_Avg_Ndeviations <- within(Merchant_Data_With_Avg_Ndeviations,
                                Weight <- Price * Sales)

#Perform weighted average analysis
ddply(na.omit(Merchant_Data_With_Avg_Ndeviations), .(Merchant), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

#ANsWER TO #2
#2. Calculate Standard Deviations on Product Level
 stdev<-
ddply(na.omit(Merchant_Data_With_Avg_Ndeviations), .(Merchant_Data_With_Avg_Ndeviations$`Jet
SKU ID`), summarize,
       stdev = sd(Price))

#3 Calculate Vendor Competitiveness by Category
ddply(na.omit(Merchant_Data_With_Avg_Ndeviations), c(.(Category),.(Merchant)), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

#Plotting Price Deviations vs Sales
Leos_Bodega <-
Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Merchant == 'Leos
Bodega',]
Alex_Store <-
Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Merchant == 'Alexs
Store',]
Jasmines_Shop <-
Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Merchant == 'Jasmines
Shop',]

#Plot overall
qplot(Sales, Perc_Deviation
     ,colour = Category, shape = Merchant,
     main = "All Vendors Percentage Price Deviations from Avg Price vs Sales",
     data = Merchant_Data_With_Avg_Ndeviations
     )

#Plot Jasmine's
qplot(Sales, Perc_Deviation
     ,colour = Category,
     main = "Jasmine's Percentage Price Deviations from Avg Price vs Sales",
     data = Jasmines_Shop)

#Plot Alex's
qplot(Sales, Perc_Deviation
```

```
     ,colour = Category,
     main = "Alex's Store Percentage Price Deviations from Avg Price vs Sales",
     data = Alex_Store)

#Plot Leo's
qplot(Sales, Perc_Deviation
     ,colour = Category,
     main = "Leo's Percentage Price Deviations from Avg Price vs Sales",
     data = Leos_Bodega)

#Calculate Vendor Competitiveness Amongst Clusters

ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
<200 &
             Merchant_Data_With_Avg_Ndeviations$Category=='Electronics',]), .(Merchant), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
>=200 &

Merchant_Data_With_Avg_Ndeviations$Category=='Electronics',]), .(Merchant), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
<350 &
                                        Merchant_Data_With_Avg_Ndeviations$Category=='Home',]), .(Merchant),
summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
>=350 &
                                        Merchant_Data_With_Avg_Ndeviations$Category=='Home',]), .(Merchant),
summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))


ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
<500 &

Merchant_Data_With_Avg_Ndeviations$Category=='Consumables',]), .(Merchant), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))
```

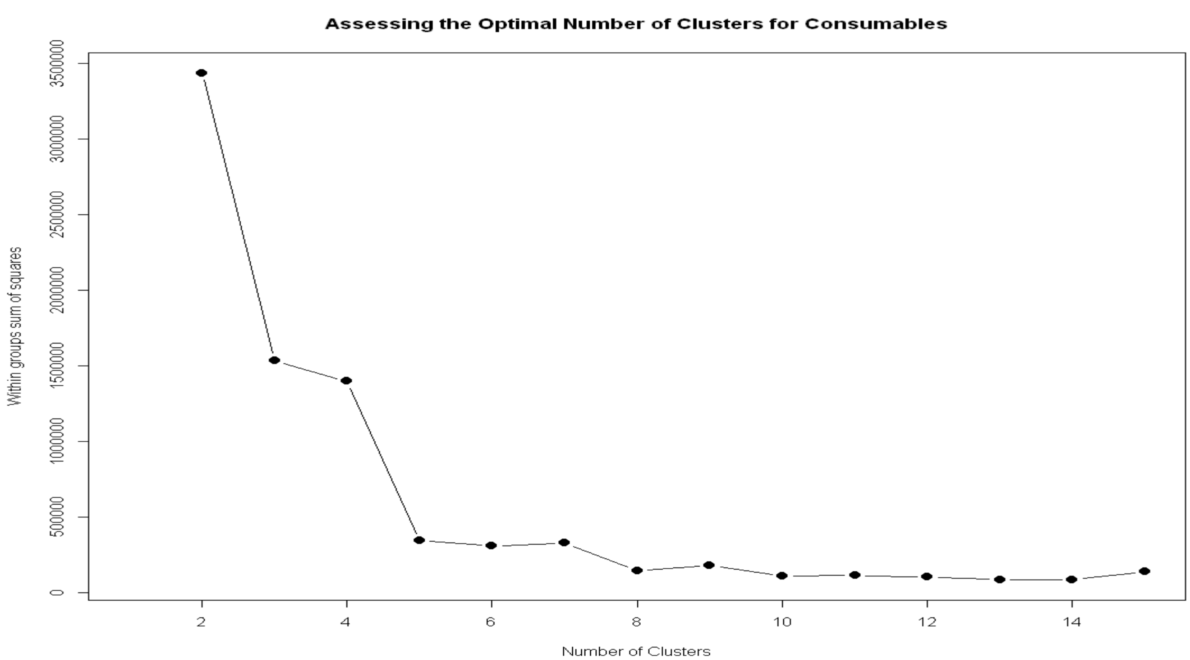
```
ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
>=500 &
                                  Merchant_Data_With_Avg_Ndeviations$Sales <=600 &

Merchant_Data_With_Avg_Ndeviations$Category=='Consumables',]), .(Merchant), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))


ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
>=1000 &
                                  Merchant_Data_With_Avg_Ndeviations$Sales <=1100 &

Merchant_Data_With_Avg_Ndeviations$Category=='Consumables',]), .(Merchant), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

ddply(na.omit(Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Sales
>1100 &

Merchant_Data_With_Avg_Ndeviations$Category=='Consumables',]), .(Merchant), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

##kmeans Calculation

#Isolate Electronics Data
ElectronicsData <-
Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Category=='Electronics
',]

#Identify Appropriate Number of Clusters
mydata <-ElectronicsData
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(mydata[,c(6,8)],
                              centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
    ylab="Within groups sum of squares",
    main="Assessing the Optimal Number of Clusters for Electronics",
    pch=20, cex=2)

#Group data into best number of clusters
KCluster <- kmeans( ElectronicsData[,c(6)]
                , 5, nstart = 20)

#Plot the data with clusters
qplot(Sales, Perc_Deviation
     , color = KCluster$cluster,
```

```
     data =ElectronicsData
)

ElectronicsData$Cluster <- KCluster$cluster
ddply(na.omit(ElectronicsData), c(.(Cluster),.(Merchant)), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))


##Consumables##
#Isolate Consumables Data
ConsumablesData <-
Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Category=='Consumab
les',]

#Identify Appropriate Number of Clusters
mydata <-ConsumablesData
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(mydata[,c(6)],
                                     centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
    ylab="Within groups sum of squares",
    main="Assessing the Optimal Number of Clusters for Consumables",
    pch=20, cex=2)

#Group data into best number of clusters
KCluster <- kmeans( ConsumablesData[,c(6)]
                    , 5, nstart = 20)

#Plot the data with clusters
qplot(Sales, Perc_Deviation
     , color = KCluster$cluster,
     data =ConsumablesData
)


#Calculate Vendors Competitiveness in Each Clusters
ConsumablesData$Cluster <- KCluster$cluster
ddply(na.omit(ConsumablesData), c(.(Cluster),.(Merchant)), summarize,
     mean = mean(Perc_Deviation),
     wmean = weighted.mean(Perc_Deviation, w=Weight))

##HOME##
#Isolate Home Data
HomeData <-
Merchant_Data_With_Avg_Ndeviations[Merchant_Data_With_Avg_Ndeviations$Category=='Home',]

#Identify Appropriate Number of Clusters
```

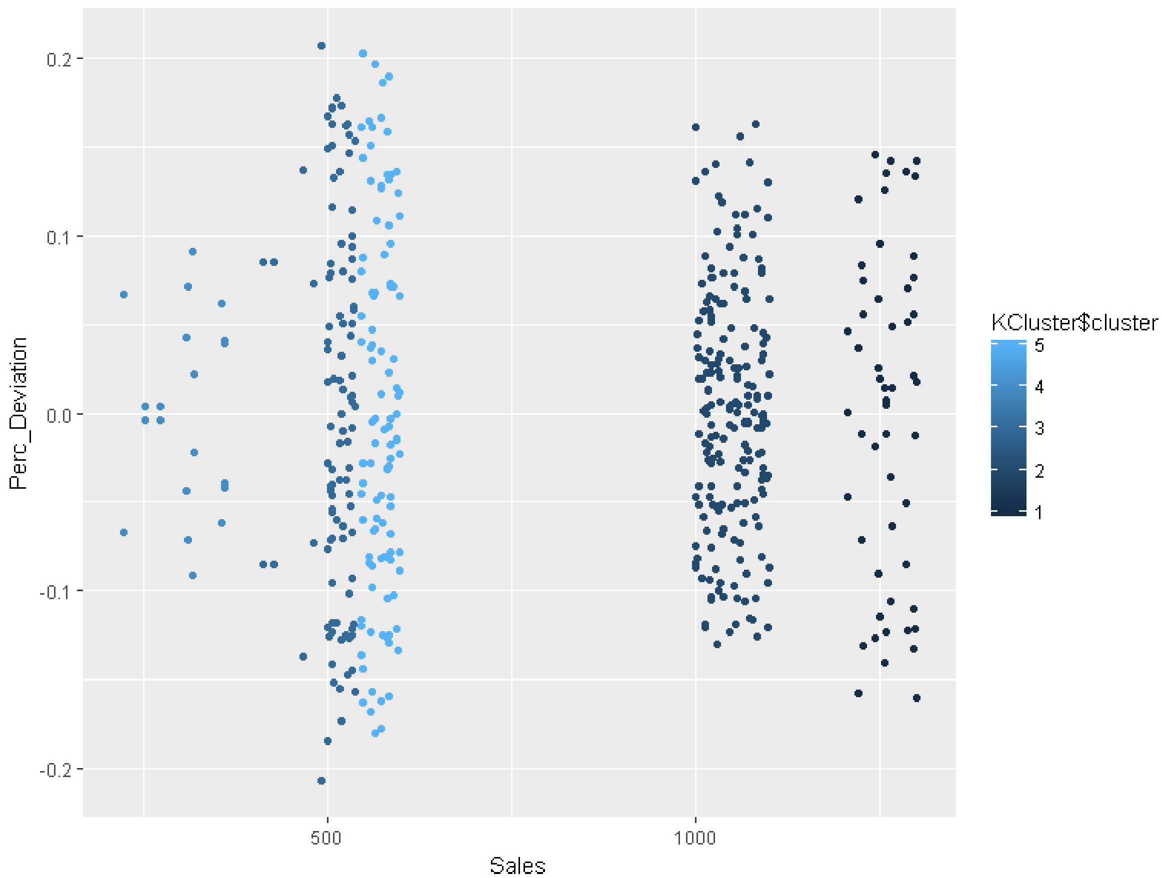
```
mydata <-HomeData
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(mydata[,c(6)],
                                   centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
     ylab="Within groups sum of squares",
     main="Assessing the Optimal Number of Clusters for Home",
     pch=20, cex=2)

#Group data into best number of clusters
KCluster <- kmeans( HomeData[,c(6,8)]
                   , 2, nstart = 20)

#Plot the data with clusters
qplot(Sales, Perc_Deviation
      , color = KCluster$cluster,
      data =HomeData
)

#Calculate Vendors Competitiveness in Each Clusters
HomeData$Cluster <- KCluster$cluster
ddply(na.omit(HomeData), c(.(Cluster),.(Merchant)), summarize,
      mean = mean(Perc_Deviation),
      wmean = weighted.mean(Perc_Deviation, w=Weight))
```

## b) Rest of K-Clustering Analysis

Consumables

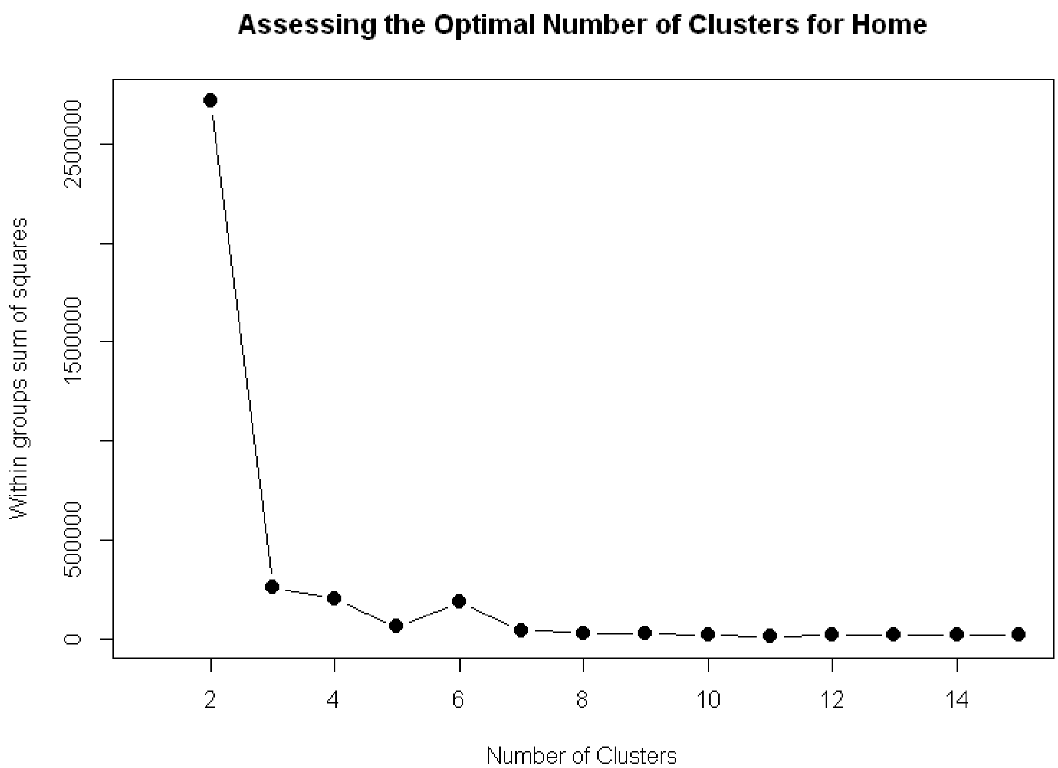


We therefore use 5 clusters.

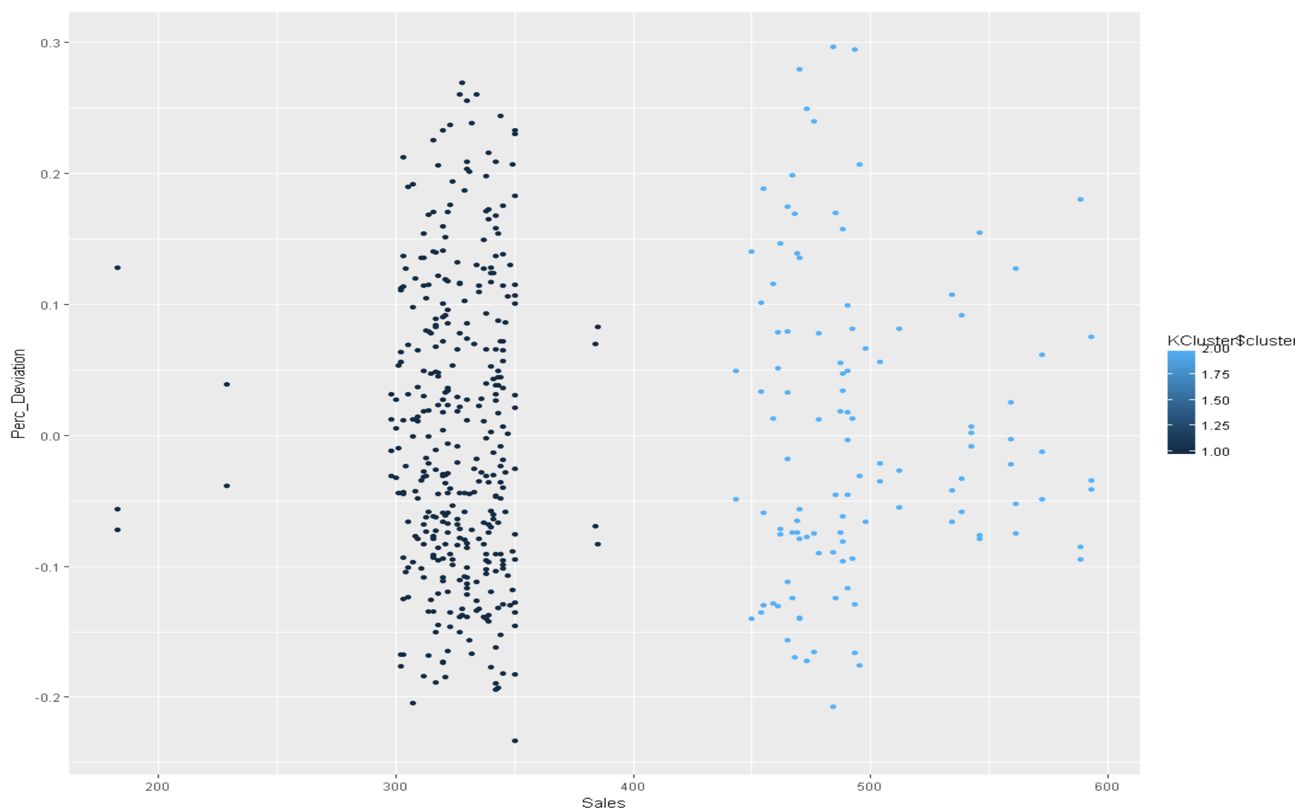



```
Cluster      Merchant        mean       wmean
1         1   Alexs Store  0.015364109  0.01678657
2         1 Jasmines Shop  0.090936427  0.09968869
3         1   Leos Bodega -0.106300536 -0.11291788
4         2   Alexs Store -0.070155960 -0.03923642
5         2 Jasmines Shop  0.011341841  0.02334427
6         2   Leos Bodega  0.057734797  0.01861944
7         3   Alexs Store -0.003312025  0.03560698
8         3 Jasmines Shop -0.107237437 -0.11685006
9         3   Leos Bodega  0.116760133  0.08726077
10        4   Alexs Store  0.014899294 -0.01916158
11        4 Jasmines Shop -0.020814357  0.02162216
12        4   Leos Bodega  0.018805457 -0.01663517
13        5   Alexs Store -0.005411566  0.01291529
14        5 Jasmines Shop -0.108529192 -0.10507161
15        5   Leos Bodega  0.113640116  0.10018578
```

Home




We employ 2 clusters



| | Cluster | Merchant | mean | wmean |
|---|---|---|---|---|
| 1 | 1 | Alexs Store | -0.108177335 | -0.09695954 |
| 2 | 1 | Jasmines Shop | 0.096442762 | 0.18544336 |
| 3 | 1 | Leos Bodega | 0.007361416 | **-0.10741361** |
| 4 | 2 | Alexs Store | -0.037897523 | -0.03152964 |
| 5 | 2 | Jasmines Shop | 0.088262710 | 0.15232820 |
| 6 | 2 | Leos Bodega | -0.055136144 | **-0.12070079** |

Using this method, Leo's comes ahead in weighted mean for both clusters.

## c) R Outputs of Visual Cluster Analysis

Cluster Analysis

Electronics & Sales <=200

| | Merchant | mean | wmean |
|---|---|---|---|
| 1 | Alex's Store | 0.04773944 | 0.08913849 |
| 2 | Jasmine's Shop | -0.21872862 | -0.19314980 |
| 3 | Leo's Bodega | 0.16970355 | 0.12741897 |

Electronics & Others

| | Merchant | mean | wmean |
|---|---|---|---|
| 1 | Alex's Store | 0.03687443 | 0.104387146 |
| 2 | Jasmine's Shop | -0.07303046 | -0.006173171 |
| 3 | Leo's Bodega | 0.04922283 | -0.061484714 |

Home & Sales <=350

| | Merchant | mean | wmean |
|---|---|---|---|
| 1 | Alex's Store | -0.10864753 | -0.10968588 |
| 2 | Jasmine's Shop | 0.09426318 | 0.13823929 |
| 3 | Leo's Bodega | 0.01069833 | -0.01692674 |

Home & Sales >350

| | Merchant | mean | wmean |
|---|---|---|---|
| 1 | Alex's Store | -0.04649026 | -0.03044443 |
| 2 | Jasmine's Shop | 0.09495645 | 0.16872884 |
| 3 | Leo's Bodega | -0.05383899 | -0.14005252 |

Consumables & Sales <500

```
     Merchant        mean        wmean
1   Alex's Store  0.008950697 -0.0174932535
2 Jasmine's Shop -0.043163331 -0.0001484542
3   Leo's Bodega  0.060181306  0.0281218087
```

Consumables & Sales >=500 & Sales <=600

```
     Merchant        mean      wmean
1   Alex's Store -0.004329977  0.02963441
2 Jasmine's Shop -0.109711063 -0.11391549
3   Leo's Bodega  0.113911786  0.08945818
```

Consumables & Sales >=1000 & Sales <=1100

```
     Merchant       mean      wmean
1   Alex's Store -0.07015596 -0.03923642
2 Jasmine's Shop  0.01134184  0.02334427
3   Leo's Bodega  0.05773480  0.01861944
```

Consumables & Sales >1100

```
     Merchant       mean      wmean
1   Alex's Store  0.01536411  0.01678657
2 Jasmine's Shop  0.09093643  0.09968869
3   Leo's Bodega -0.10630054 -0.11291788
```

## d) Sales * Price Approach to Answering #1

This is a much more simplistic approach and makes a lot of intuitive sense. However, some vendors sell less number of products than others. This bias the approach in favoring vendors that have low number of product offerings. So we can either 1) use another method to give a more complete picture which is what I did or 2) Omit all products that only have 2 sellers which are roughly 50 products out of 500.

I included my original analysis on this approach before I realized the flaw in the approach due to this factor. To get the total cost to Jet, multiply sales * price and aggregate the measure onto Merchant level. Vendor with the lowest cost, i.e. Leo's, offers the best prices.

| Merchant | Total Cost | Rank |
|---|---|---|
| Leo's Bodega | $(21,682,565.82) | 1 |
| Alex's Store | $(21,821,501.63) | 2 |
| Jasmine's Shop | $(25,585,652.53) | 3 |

**Within this context, Leo's has the best pricing, and Jasmine's has the worst.**

*By category,*

| *Merchant* | *Consumables Total Cost* | *Rank* |
|---|---|---|
| *Leo's Bodega* | *$(12,082,647.59)* | *1* |
| *Alex's Store* | *$(12,486,838.45)* | *2* |
| *Jasmine's Shop* | *$(13,287,615.96)* | *3* |

| *Merchant* | *Home Total Cost* | *Rank* |
|---|---|---|
| *Leo's Bodega* | *$(5,975,237.21)* | *2* |
| *Alex's Store* | *$(5,779,851.32)* | *1* |
| *Jasmine's Shop* | *$(8,403,722.22)* | *3* |

| *Merchant* | *Electronics Total Cost* | *Rank* |
|---|---|---|
| *Leo's Bodega* | *$(3,624,681.02)* | *2* |
| *Alex's Store* | *$(3,554,811.86)* | *1* |
| *Jasmine's Shop* | *$(3,894,314.35)* | *3* |