

Dr. Kertcher

# Final Project

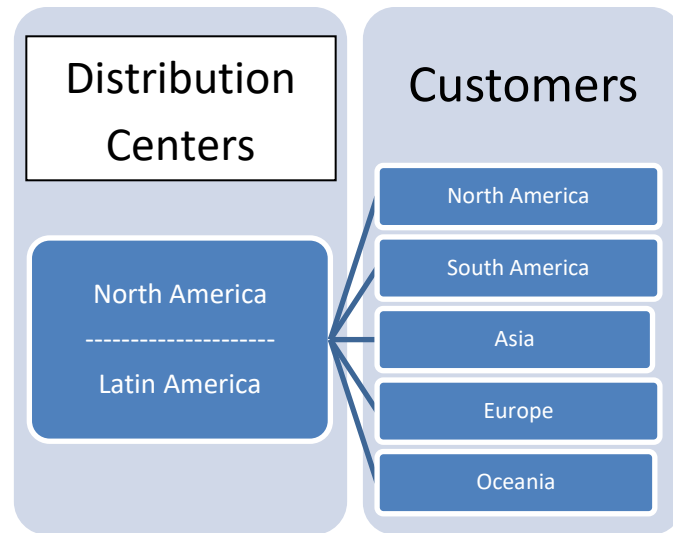
Onontsatsal Gantsog – 672164839



Spring 2017

## 1. Introduction

Nikko Company, based in Chicago, produces pieces for electronic devices. Their product is sold all over the world in five different continents. Two distribution centers are responsible for distributing their products to their customer segments; one located in North America and the other one in Latin America.



Ono and Roya, the managers of the company, have a plan in which the goal is to increase the profitability of the company. They are looking for the best way to increase profit. They are wondering which customer segment to focus on? Which distribution center works better for the company? Where and how they can invest for the best returns?

Nikko Company uses PROS software. PROS is a powerful Pricing Analytics Software, developed by PROS company, which can be used in almost all industries including retail, supply chain, healthcare, and insurance. Ono and Roya are using the data and information gathered from this software to make their decision about how to run the company for more profitability.

The objective of this research is that to help the managers to find which factors affect the profitability of the company.

## 2. Explore and Prepare Data

The data was gathered from PROS Pricing Analytics software in the manufacturer. This dataset originally contains 561 observations and 8 variables. Each observation of the data is aggregated value of a product grouping and a country, in other words net revenue of that observation is the total revenue of the product grouping in that country. In this section, variables, their descriptions and types are explained in detail. Along with the explanation, data exploring and cleaning is done in order to prepare them for the analysis;

- **Date:** A column is allocated to Date in this dataset; however, the date cannot be counted as a variable here because it is fixed. All the data is from 2016. All the dates are numeric and are 2016. This column has been removed from data before starting analysis.
- **Disty. Region:** This variable defines the distribution center location. The datatype is character and includes the name of the regions; Latin America with 110 observation and North America with 451 observations. Changed the name of the variable to “Center”. (Independent variable, Control variable)
- **Product Grouping:** Product Grouping is character type indicating the specific group of products (Independent variable). Forty six groups of products introduced in this variable. Changed the name of the variable into “Product”.
- **Cust. Geography:** Cust. Geography is representative of the country where the product sold in 2016. The data type is character. (Independent variable, Control variable). Customer country of only one of the observations is unknown and has value “NA” in the data. This sole observation is simply removed. Thirty one

countries buy their products. This variable name is changed to “Country” in the data frame.

- **Net Revenue:** is numeric, continuous data, and shows the net revenue of 2016. (Dependent variable). In cleaning data, all the commas are removed from the revenue. Exploring this variable, there are sixteen observations with negative revenue. Looking to the data more deeply, the quantity sold of above mentioned observations are also negative. Therefore, it is being known that the negative revenue is due to returns from the previous. So the decision is not to use them for the current year. Having negative quantity sold indicates that the product grouping items are returned from the previous year, so the negative revenue is not due to an investment and we can safely remove them from 2016 net revenue. Besides, there is no trend in these observations. Not they are related to a specific distributor, nor sell to a specific country. Moreover, the number of the negative revenue observation is only 2.8% of the observations and it is not significant. These 16 observations are removed from the dataset. The variable name is changed to “Revenue”.
- **Net Qty Sold:** Net Quantity Sold is the number of each group product sold in 2016. It is a numeric data and integer. (Dependent variable). As done with the revenue, commas are removed from these numeric data too. The issue with this data is that there is no specific measure for counting the product grouping. The products characteristics are different and they cannot be measured the same way. Among the entire this variable there are significantly different scales and measures. Some can be measured by feet and some by pound and/or other measures. In this regard,

this variable does not seem a good variable to be studied in the analysis. The name of the variable is changed to “Quantity”.

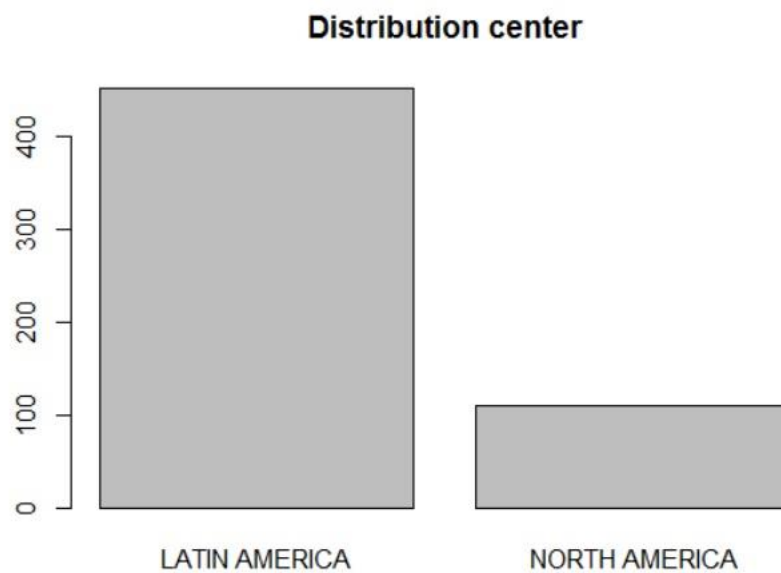
- **ASP Pricing:** Average Sell Price is a numeric, continuous variable defining the price set of group products in 2016. The variable name is changed to “Average Price”. (Independent variable)
- **Profitability Margin:** is a percentage showing the profitability of the group product in 2016 (Dependent variable). The variable name is changed to “ProfitMargin”.

### 3. Analyze data

#### 3.1. Distribution Center

From the following table it is visible that almost 80 percent of the observations are at Latin America distribution center and the rest are at North America distribution center.

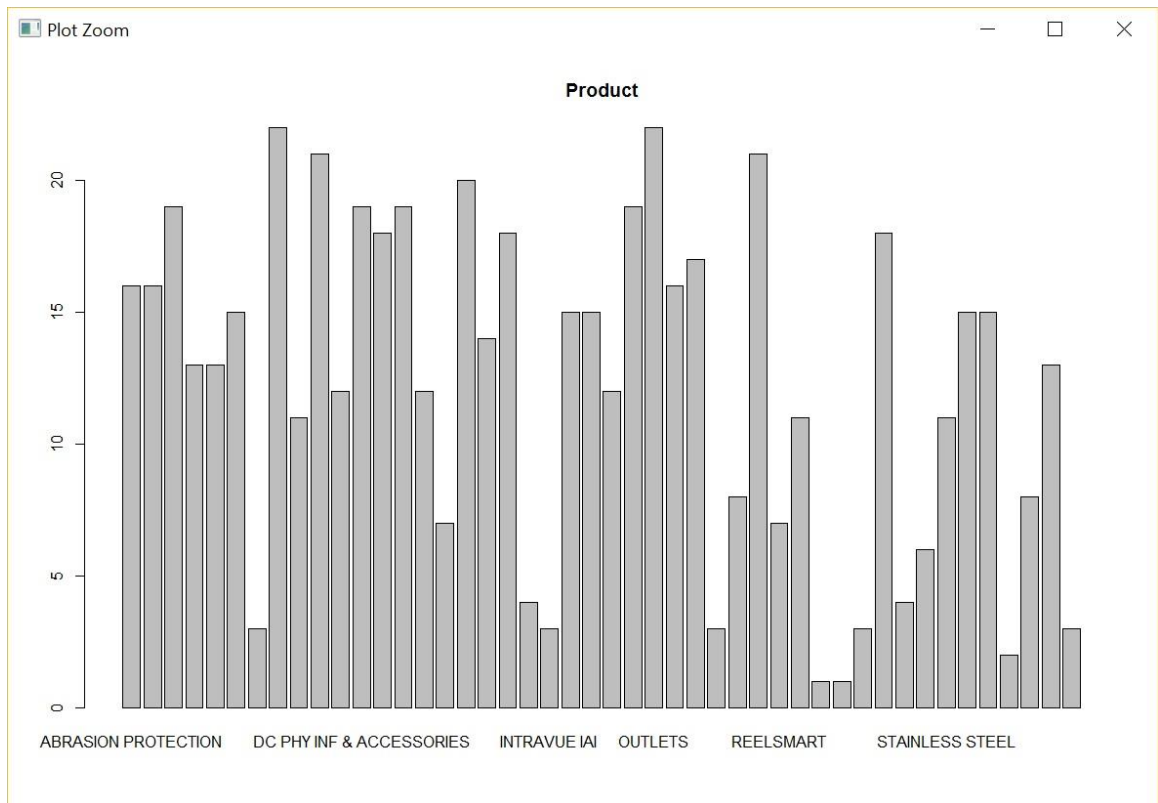
```
> table(pd$Disty..Region)
LATIN AMERICA NORTH AMERICA
          451          110
```



## 3.2. Product

```
> table(pd$Product.Grouping)
```

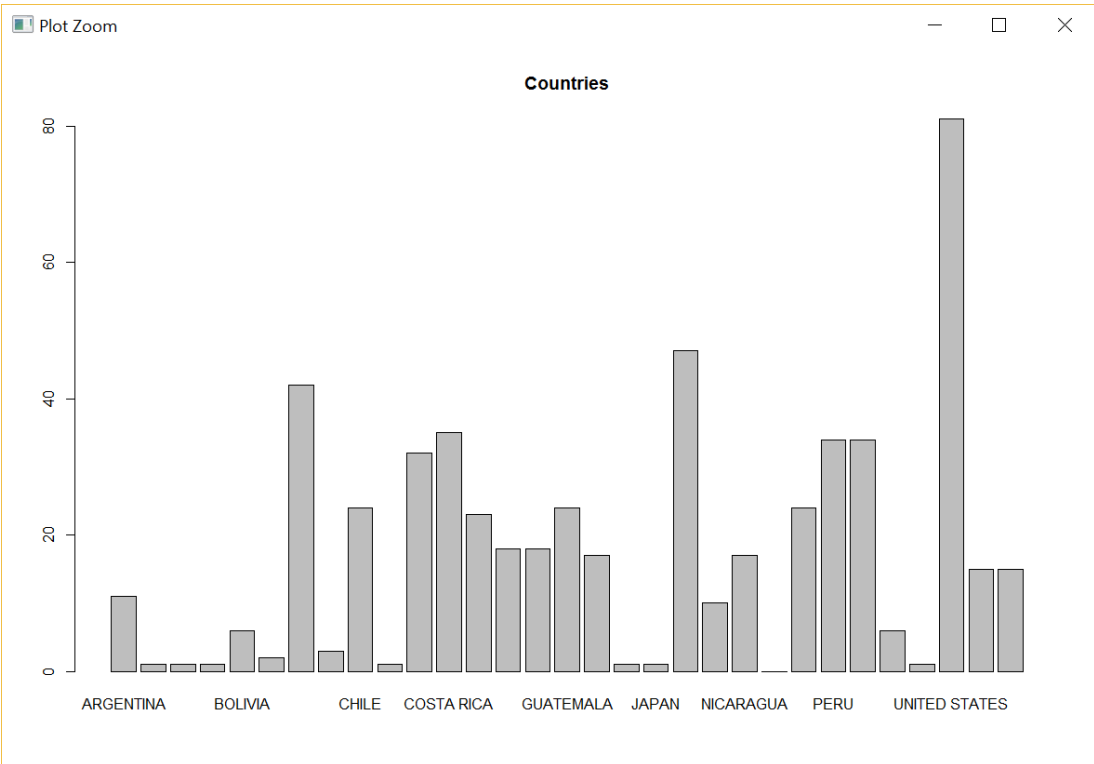
ABRASION PROTECTION	16	CABINETS AND THERMALS	16	CABLE - COPPER	19
CABLE - COPPER NETKEY	13	CABLE ACCESSORIES	13	COMPUTER PRINTABLES	15
COPPER CABLE IAI	3	COPPER CONNECTORS	22	COPPER CONNECTORS & PATCHCORDS IAI	11
COPPER PATCHCORDS	21	COPPER QUICKNET & HIGH SPEED	12	DC PHY INF & ACCESSORIES	19
FIBER CABLE	18	FIBER CONNECTIVITY	19	FIBER CONNECTORS	12
FIBER CONNECTORS & PATCHCORDS IAI	7	FIBER DISTRIBUTION	20	FIBER NETKEY	14
GROUNDING PRODUCTS	18	IMS NEW BUSINESSES	4	INTRAVUE IAI	3
LABELING SYSTEMS	15	LOOSE PIECE CABLE TIES	15	MISCELLANEOUS - OTHER	12
NETKEY COPPER CONNECTIVITY	19	OUTLETS	22	PATHWAYS	16
POWER CONNECTORS	17	PRE-CONFIGURED DATA CENTER	3	PRE-CONFIGURED IA	8
RACKS & CABLE MGMT	21	REELSMART	7	SAFETY PRODUCTS	11
SERVICES DC	1	SERVICES ENT	1	SERVICES IAI	3
SMARTZONE HARDWARE	18	SMARTZONE SERVICES	4	SMARTZONE SOFTWARE	6
STAINLESS STEEL	11	SURFACE RACEWAY	15	TERMINALS	15
TOOL SERVICE AND REPAIRS	2	WIREMARKERS	8	WIRING DUCT INCLUDES TOOLS	13
XMR TIES & TOOLS	3				



### 3.3.Country

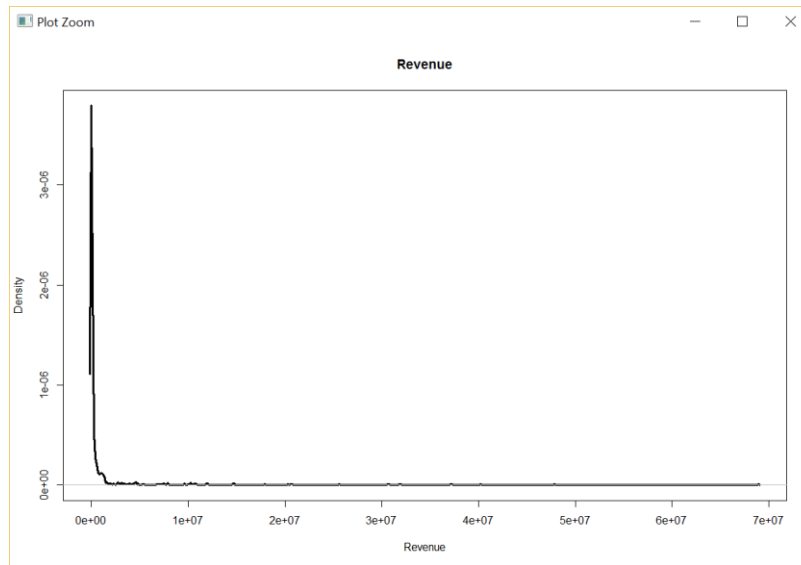
```
> table(pd3$Country)
```

ARGENTINA	AUSTRALIA	BAHAMAS	BARBADOS	BOLIVIA
11	1	1	1	6
BRAZIL	CANADA	CAYMAN ISLANDS	CHILE	CHINA
2	42	3	24	1
COLOMBIA	COSTA RICA	DOMINICAN REPUBLIC	ECUADOR	EL SALVADOR
32	35	23	18	18
GUATEMALA	HONDURAS	JAMAICA	JAPAN	MEXICO
24	17	1	1	47
NETHERLANDS ANTILLES	NICARAGUA	NOT AVAILABLE	PANAMA	PERU
10	17	0	24	34
PUERTO RICO	TRINIDAD & TOBAGO	UNITED ARAB EMIRATES	UNITED STATES	URUGUAY
34	6	1	81	15
VENEZUELA				
15				

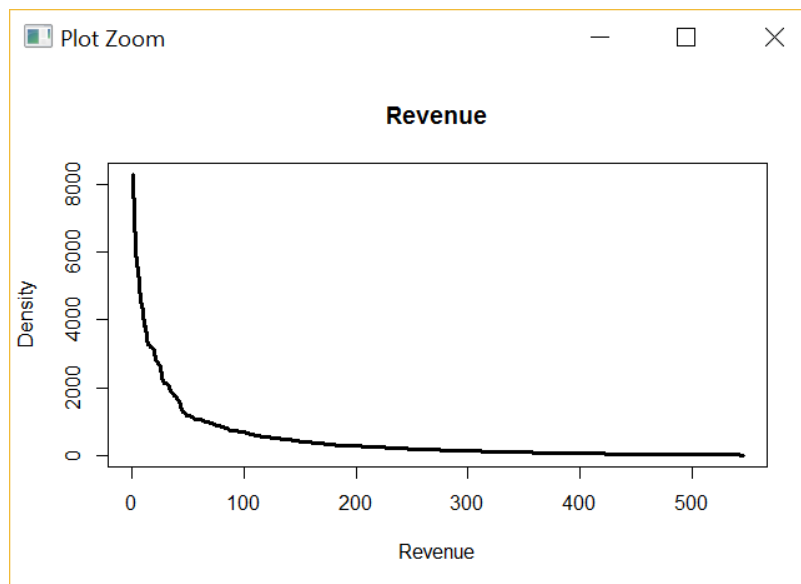


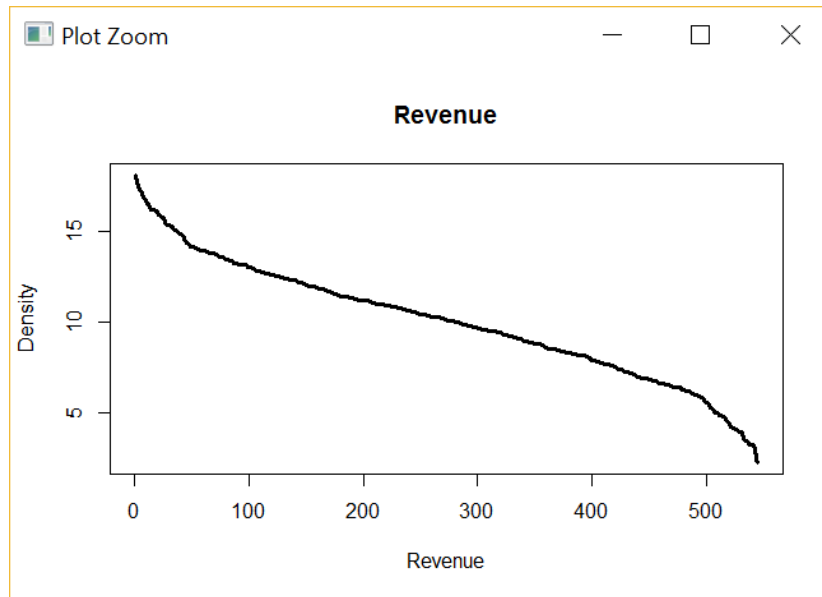


### 3.4.Revenue

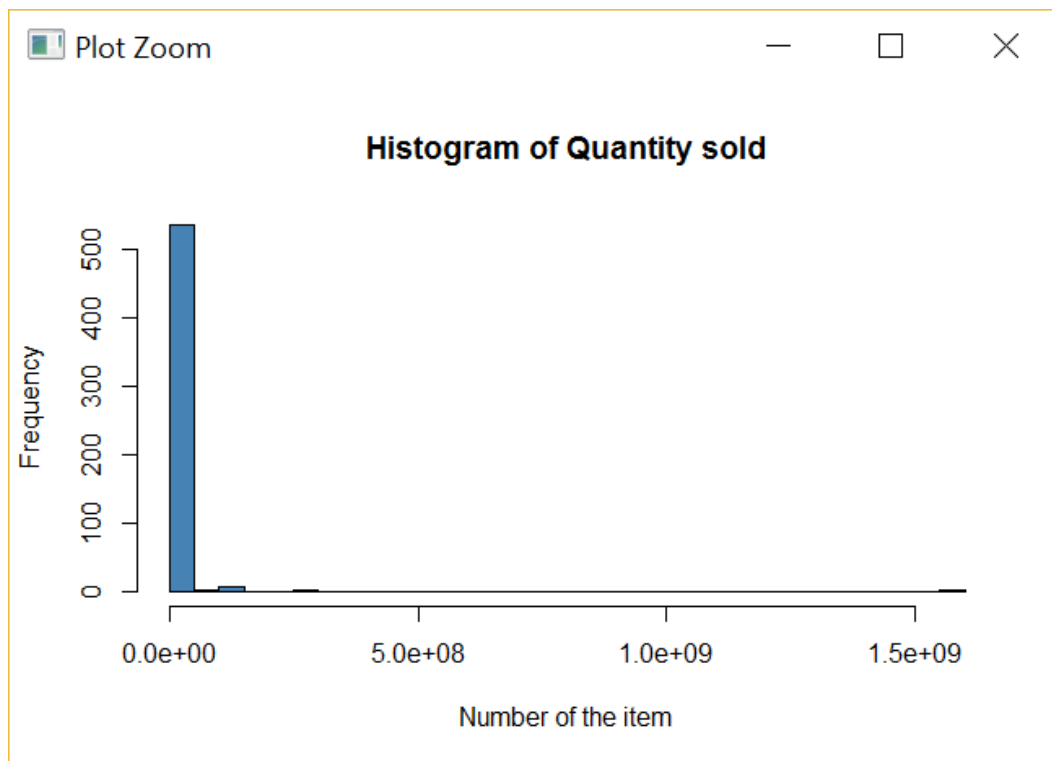


As depicted in the figure, Revenue is very skewed data. Followings are square root and log transformations respectively.



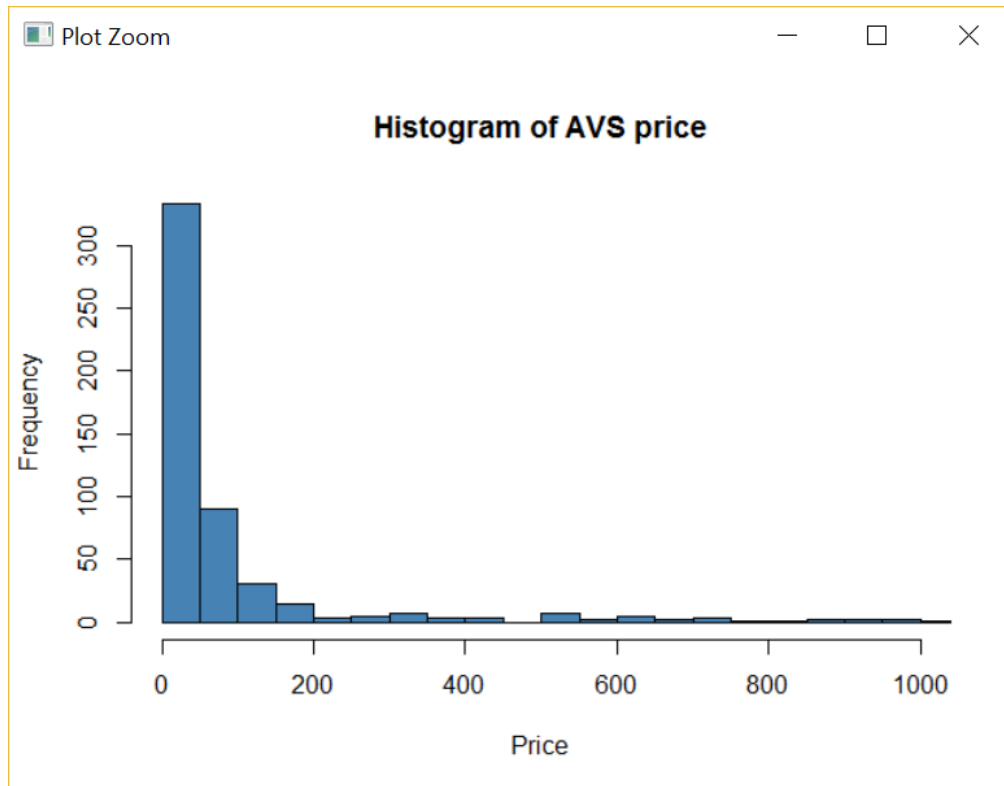


### 3.5.Quantity

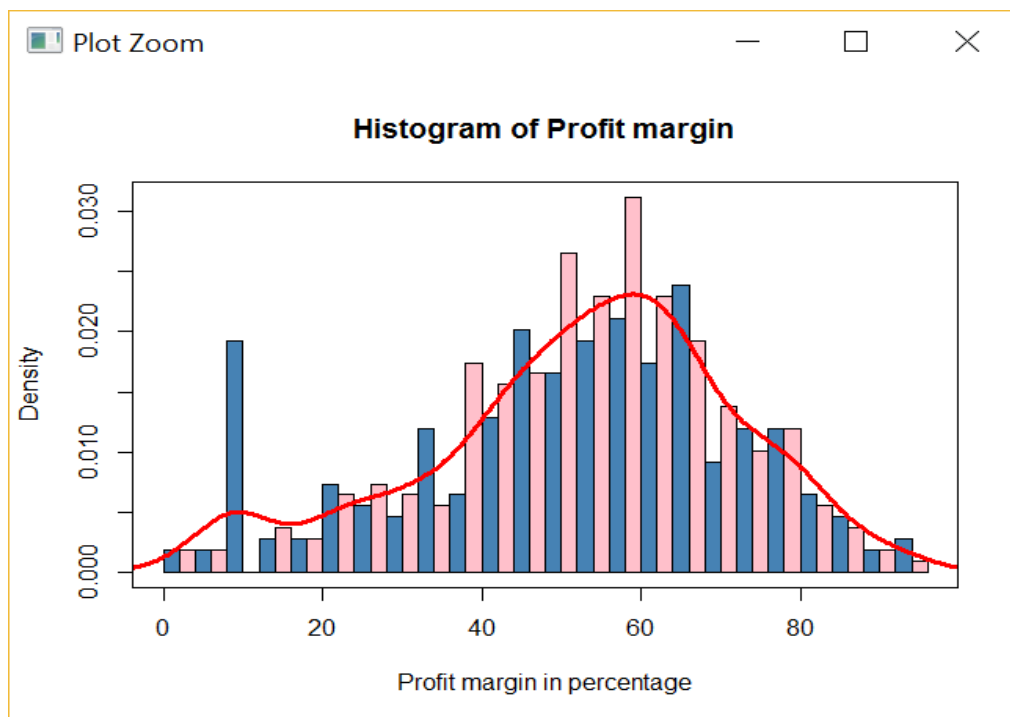


Similar to Revenue, Quantity is highly skewed data.

### 3.6.Average Price

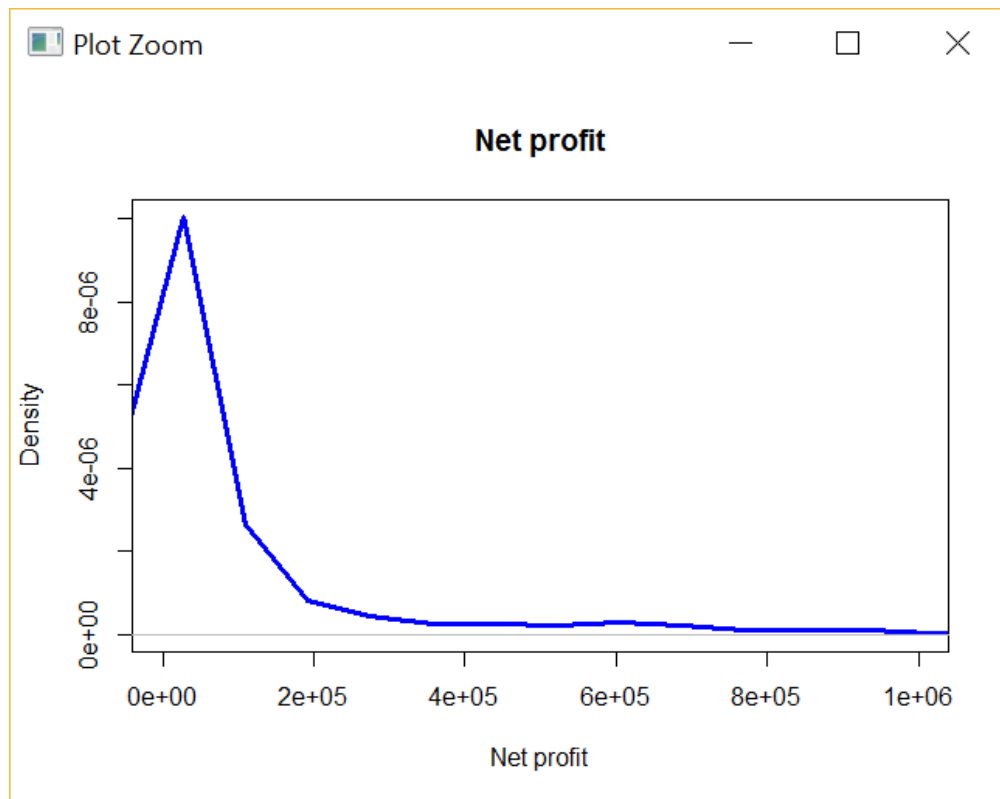


**3.7.Profit Margin:** Despite of the other variables, profit margin has normal distribution.



### 3.8. Net Profit

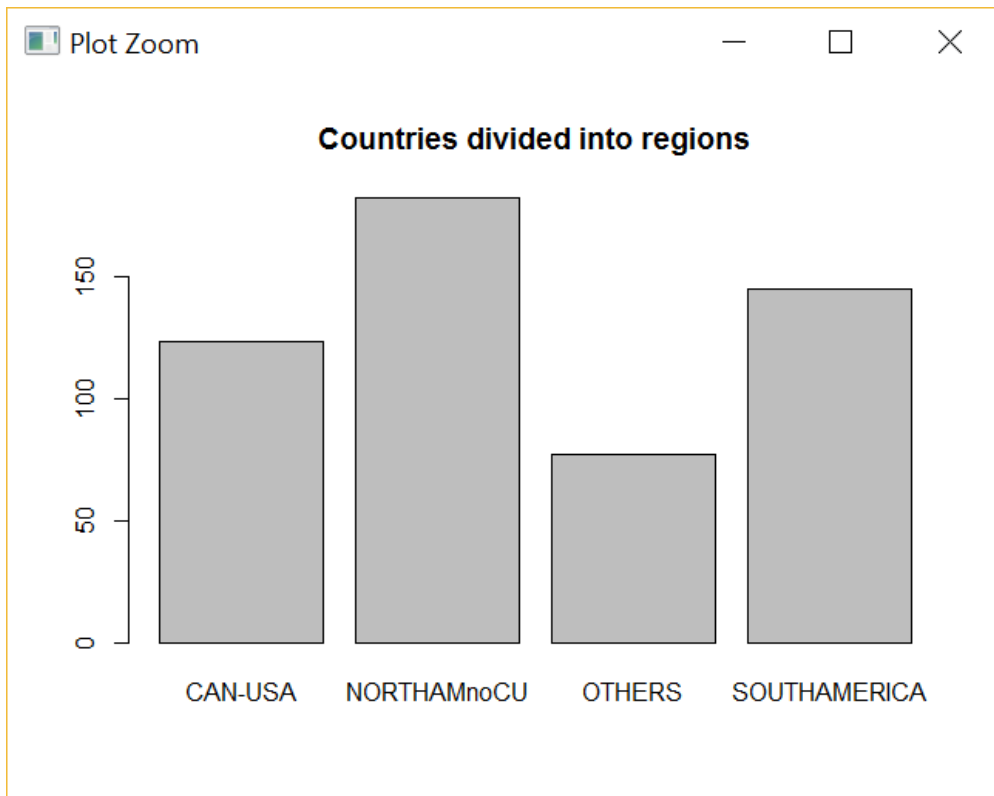
When exploring data and looking at their histogram; revenue, quantity sold, and pricing are all highly skewed. Profit margin is almost normally distributed. Furthermore, what really matters for the company managers is profitability. Revenue includes cost of goods sold and Profit margin is the percentage of profit, so another variable “NetProfit” is added to dataset which is numeric, continuous, and dependent variable. Net Profit is calculated by multiplying net revenue by profit margin.



### 3.9. Country Region

In order to be able to use this variable as a control variable, these countries are transformed as country regions so that we have only four country regions in which there are balanced numbers of transactions. Regions are as follow:

- ❖ United State and Canada (125 observations)
- ❖ All countries from North America other than Canada and the united states (186 observations)
- ❖ All countries of South America (151 observations)
- ❖ All Countries from Asia, Oceania, and Europe (79 observations)



#### 4. Research Question and Hypothesis

**Research topic** is to find out which factors affect the net profit of the company.

**First hypothesis** is that the customer segment and product grouping are independent. The reason with this hypothesis is to check that if certain customer segment is purchasing certain products.

**Second hypothesis** is that the net profits from two distribution centers are equal.

**Third hypothesis** is that the net profits from four different country regions are the same.

**Regression analysis** runs to find out how much each independent variable affects the dependent variable.

#### 5. Methods

This section discusses how the hypotheses are tested and why each specific method is chosen for each hypothesis. The results of analysis are interpreted in this section too.

**Hypothesis1:** *The customer segment and product grouping are independent.*

A chi-square test is run to examine the independency between these two variables; product grouping and customer segment.

Pearson's Chi-squared test

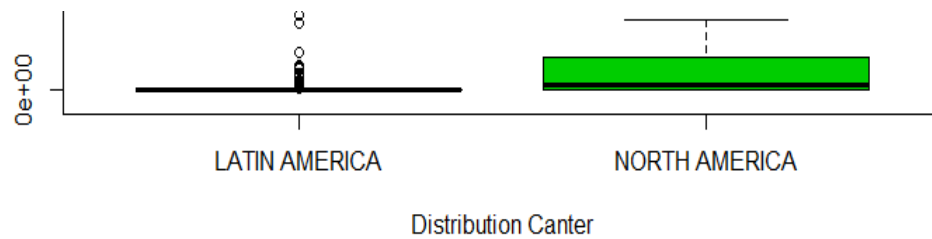
```
data: prodcregion
x-squared = 87.424, df = 135, p-value = 0.9995
```

This means it does not matter which specific product to be sent to which customer segment. There is no trend and all the customer segments need entire product groupings.

**Hypothesis2:** *net profits from two distribution centers are equal.*

t.test and wilcox test indicates that the difference between mean of net profit from two distribution centers is not zero. The result of net profit aggregation based on distribution centers shows the average of net profit for two centers in 2016 are as follow;

Distribution Center	Average net profit
Latin America	3023172.03
North America	97811.35



From figure: Net profit by centers, we see that mean of the centers are not equal. It is interesting that although most of the observations were at Latin America distribution center, the average Net profit is higher at North America distribution center. That means customers of Latin America are buying cheaper products than North America distribution center customers.

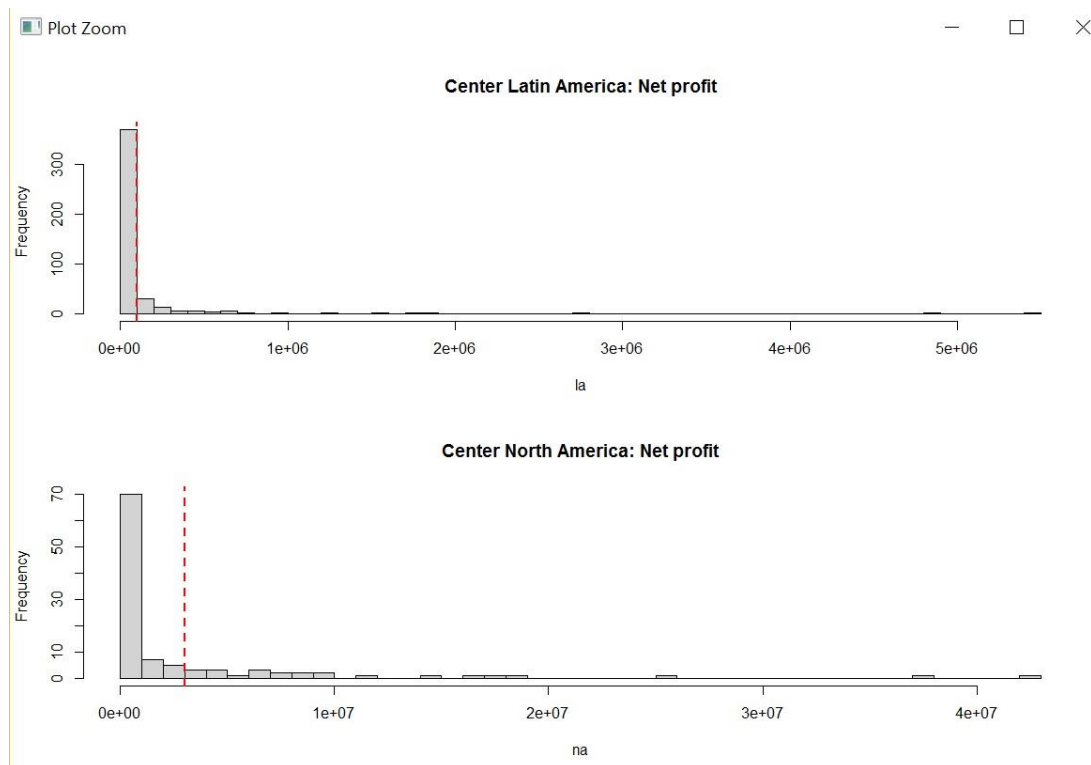
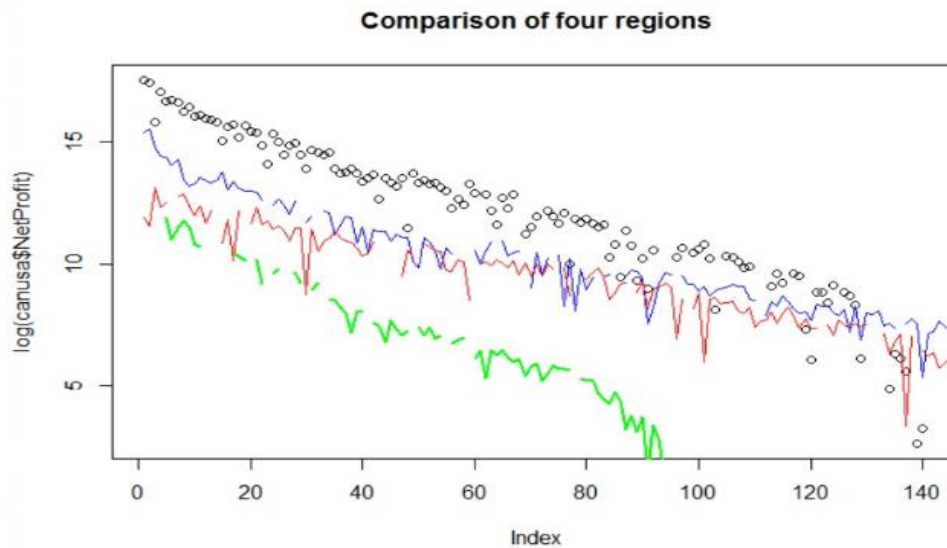


Figure: Net profit by centers

**Hypothesis3:** *The net profits from four different country regions are the same.*  
Using ANOVA, the mean of net profit is calculated for each individual country region. The result is as follow:

Country region	Average Net Profit
US and Canada	2626669.76
N. America except US & Canada	177891.68
South America	41673.95
Asia, Europe, Oceania	14444.32

Result shows that the net profit from US and Canada is the highest in 2016. Other countries from North America are ranked second almost the same as South America. Asia, Oceania, and Europe together have the lowest net profit for the company.



- ◇ USA and Canada
- North America without Canada and the US
- South America
- Others (Asia, Oceania, Europe)

Figure: Plot for comparison of four regions

Because Net Profit variable is highly skewed, in order to compare the Net Profit of the regions in a meaningful visualization, transformation of log is used on the data. With this transformation Net Profit is displayed as an almost straight downward slope line.

Pair name	Diff	Lwr	Upr	P adj
NA - CanUSA	-2448778.07	-3385197	-1512358.9	0.00
Others- CanUSA	-2612225.44	-3778030	-1446421.0	0.00
SA-CanUSA	-2584995.80	-3568417	-1601574.9	0.00
Others-NA	-163447.36	-1254077	927182.6	0.98
SA-NA	-136217.73	-1029241	756805.6	0.98
SA-Others	27229.64	-1104013	1158472.2	1.00

Table: ANOVA – TukeyHSD test



From table ANOVA-TukeyHSD test too, it is visible that there is most difference between the country regions of Others and USA and Canada.

### Correlation

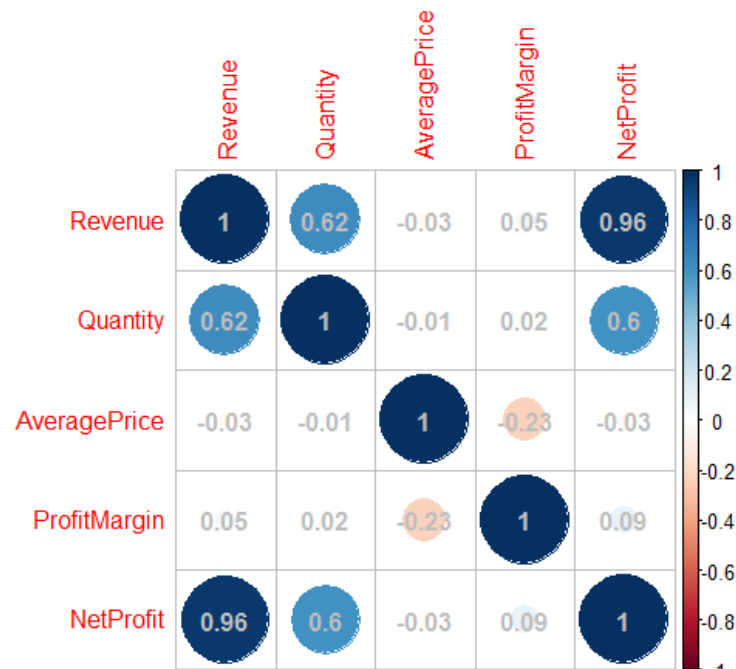


Figure: Correlation of Revenue, Quantity, Average Price, Profit Margin and Net Profit.

Since Net Profit is created from Revenue and Profit Margin, there will be correlation among those variables. But, from the figure, it is visible that Revenue is highly correlated to Net Profit while Profit Margin's correlation value is 0.09.

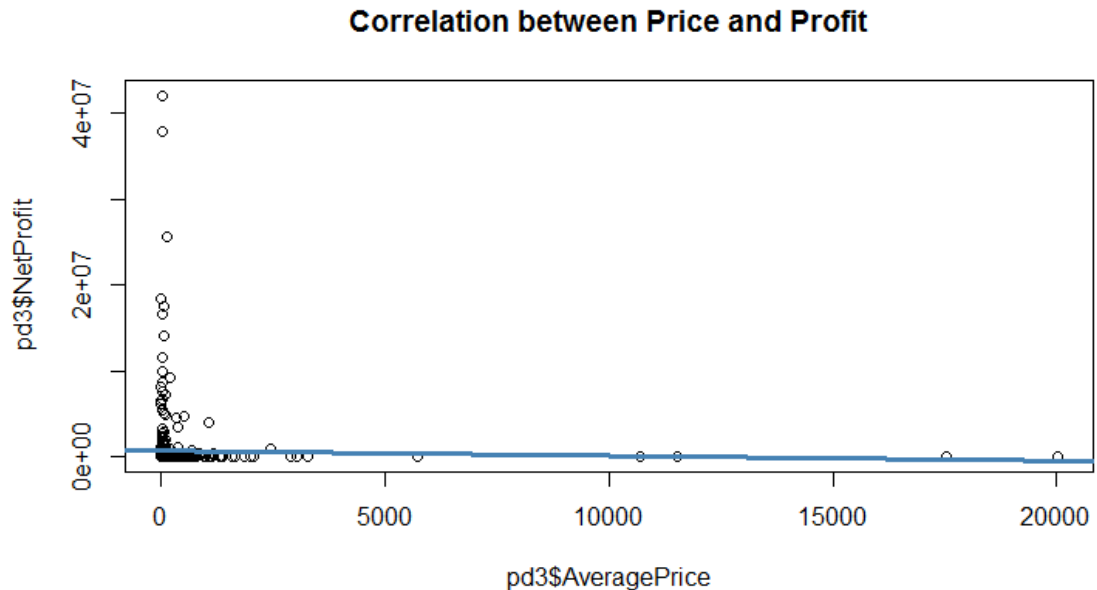
Another interesting correlation is Quantity and Revenue with value of 0.62. Although it is logical that higher quantity higher revenue, it could mean that there are not many high value products. Because Quantity has correlation with Revenue, it also has correlation with Net Profit with value of 0.6.

There is negative correlation between Average Price and Profit Margin with value of -0.23. It means lower the average price for the product higher the profit margin.

	Revenue	Quantity	AveragePrice	ProfitMargin	NetProfit
Revenue	1.00	0.62	-0.03	0.05	0.96
Quantity	0.62	1.00	-0.01	0.02	0.60
AveragePrice	-0.03	-0.01	1.00	-0.23	-0.03
ProfitMargin	0.05	0.02	-0.23	1.00	0.09
NetProfit	0.96	0.60	-0.03	0.09	1.00

**Regression Analysis:** Regression shows how much dependent variable is affected by each correlated independent variable.

The following figure shows that the net profit has not affected by price change.



Correlation between price and profit is 0.03 which is closer to zero rather than 1. The result of this correlation is consistent with the coefficient of average price in regression model.

Intercept is the dependent variable where all the independent variables are zero. Each coefficient is representative of the amount of change in dependent variable when the independent variable changes one unit. Positive coefficient shows that dependent variable and independent variable change the same direction. For example, in this regression; when quantity sold increases the net profit also increases. The negative coefficient for price indicates that when the price goes up the profit would decrease.

With Center, Quantity, AveragePrice, CountryRegion and Product variables as independent variables the regression of highest Adjusted R-squared was 0.4433 without data transformation.

```

> Reg <-lm(NetProfit ~ Center+Quantity+AveragePrice+CountryRegion+Product, data = pd3)
> summary(Reg)

Call:
lm(formula = NetProfit ~ Center + Quantity + AveragePrice + CountryRegion +
    Product, data = pd3)

Residuals:
    Min       1Q   Median       3Q      Max
-3339458  -802135   -6053    374118   32452563

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.059e+06  7.316e+05   1.448  0.14828
CenterNORTH AMERICA  1.962e+06  3.874e+05   5.066 5.83e-07 ***
Quantity          2.595e-02  1.590e-03  16.322 < 2e-16 ***
AveragePrice       4.308e+01  1.406e+02   0.306  0.75942
CountryRegionNORTHAMnoCU -9.367e+05  3.868e+05  -2.422  0.01582 *
CountryRegionOTHERS  -1.409e+06  4.299e+05  -3.277  0.00112 **
CountryRegionSOUTHAMERICA -1.076e+06  4.053e+05  -2.655  0.00819 **

ProductCOPPER CONNECTORS      2.072e+06  8.549e+05   2.424  0.01573 *
ProductCOPPER CONNECTORS & PATCHCORDS IAI -6.410e+05  1.015e+06  -0.632  0.52794
ProductCOPPER PATCHCORDS      4.062e+05  8.550e+05   0.475  0.63494
ProductCOPPER QUICKNET & HIGH SPEED -2.540e+03  1.050e+06  -0.002  0.99807
ProductDC PHY INF & ACCESSORIES -1.960e+05  8.733e+05  -0.224  0.82256

-----
ProductWIRING DUCT INCLUDES TOOLS      1.055e+06  9.649e+05   1.093  0.27495
ProductXMR TIES & TOOLS      -2.519e+06  1.567e+06  -1.607  0.10872
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2450000 on 475 degrees of freedom
(18 observations deleted due to missingness)
Multiple R-squared:  0.4973,    Adjusted R-squared:  0.4433
F-statistic: 9.214 on 51 and 475 DF,  p-value: < 2.2e-16

```

Figure: Initial regression with Adjusted R-squared: 0.4433

T value defines the p value, and p value shows the probability of randomness. Here p value is small and we reject the null. Null is that this relationship occurred by chance. So we reject the null, therefore, this regression for quantity and distribution center did not happen by chance.

But considering NetProfit, Quantity and AveragePrice variables are highly skewed Regression is run again with data transformation of square root (sqrt function). This time, Adjusted R-squared is improved as 0.5641. Refer to the figure: Final regression

R-squared is representative of covariance and shows that 56% of the variations in dependent variable can be explained by this regression. However, adjusted R-squared is more reliable because it takes into account the sample size. Adjusted R-squared is equal to 0.5641. This means that more than half of movement in net profit is due to changes in quantity sold, distribution center, and price.

```

> reglog <-lm(sqrt(NetProfit) ~ sqrt(Quantity) + sqrt(AveragePrice)+Center+CountryRegion+Product, data = pd3)
> summary(reglog)

Call:
lm(formula = sqrt(NetProfit) ~ sqrt(Quantity) + sqrt(AveragePrice) +
    Center + CountryRegion + Product, data = pd3)

Residuals:
    Min       1Q   Median       3Q      Max
-1079.4   -221.3    -38.0    112.8   4129.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.772e+02  1.484e+02   3.215  0.00139 **
sqrt(Quantity)  1.693e-01  1.081e-02  15.662 < 2e-16 ***
sqrt(AveragePrice) 1.772e+00  3.215e+00   0.551  0.58172
CenterNORTH AMERICA  5.329e+02  7.875e+01   6.767  3.89e-11 ***
CountryRegionNORTHAMnOCU -3.146e+02  7.830e+01  -4.019  6.81e-05 ***
CountryRegionOTHERS -5.141e+02  8.726e+01  -5.892  7.25e-09 ***
CountryRegionSOUTHAMERICA -3.725e+02  8.212e+01  -4.536  7.27e-06 ***
ProductCABINETS AND THERMALS  5.708e+01  1.936e+02   0.295  0.76826
ProductCABLE - COPPER  7.537e+01  1.796e+02   0.420  0.67489
ProductCABLE - COPPER NETKEY  9.511e+01  1.906e+02   0.499  0.61800
ProductCABLE ACCESSORIES -9.675e+01  1.914e+02  -0.506  0.61343
ProductCOMPUTER PRINTABLES -2.319e+02  1.877e+02  -1.235  0.21744
ProductCOPPER CABLE IAI -5.889e+02  3.163e+02  -1.862  0.06319 .
ProductCOPPER CONNECTORS  4.888e+02  1.727e+02   2.831  0.00484 **
↓
ProductSERVICES IAI -7.788e+02  4.783e+02  -1.628  0.10415
ProductSMARTZONE HARDWARE -8.521e+01  1.802e+02  -0.473  0.63649
ProductSMARTZONE SERVICES -4.324e+02  3.208e+02  -1.348  0.17830
ProductSMARTZONE SOFTWARE -3.568e+02  2.649e+02  -1.347  0.17876
ProductSTAINLESS STEEL -1.441e+02  2.090e+02  -0.690  0.49081
ProductSURFACE RACEWAY  5.237e+00  1.880e+02   0.028  0.97779
ProductTERMINALS -1.682e+02  1.845e+02  -0.912  0.36244
ProductTOOL SERVICE AND REPAIRS -1.048e+03  3.834e+02  -2.733  0.00652 **
ProductWIREMARKERS -2.690e+02  2.197e+02  -1.225  0.22137
ProductWIRING DUCT INCLUDES TOOLS 1.542e+02  1.948e+02   0.791  0.42925
ProductXMR TIES & TOOLS -1.020e+03  3.191e+02  -3.196  0.00149 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 493.7 on 475 degrees of freedom
(18 observations deleted due to missingness)
Multiple R-squared:  0.6063,    Adjusted R-squared:  0.5641
F-statistic: 14.35 on 51 and 475 DF, p-value: < 2.2e-16

```

Figure: Final regression with Adjusted R-squared: 0.5641

## 6. Conclusion

The company should make a unique decision for all its products regardless of the country they are sold. These two variables are independent.

Despite the fact that the number of transactions of Latin America was the majority of the transactions, the average net profit is higher at North America distribution center. The company should be aware more transactions do not mean more profit.

The company should focus on North America distribution center as well as customers from Canada and US who are the most profitable elements for the company. The company has a good market share in Canada and US. It is better off not losing that to the competitors. On the other hand, regarding promotion budget, the company should spend more in Asia, Europe, and Oceania and make this customer segment more aware of its products. Therefore, the company will improve net profit out of those regions as well.

The company should improve their capacity and production so that they can increase quantity sold. They should not increase the price because they would lose their customer and as a result their net profit would decrease.