Final Project Report

for

MODEL TO INCREASE THE VOLUME *OF OPERATIONS FOR A FREIGHT FORWARDING COMPANY*

Version 1.0

Prepared by Adriana Ortiz & Marcel Socorro

St. Thomas University, CIS 545, Spring 2018

Date Last Modification 03-03-19
Date of Creation:02-28-19

Table of Contents

Revision History

Name	Date	Reason For Changes	Version

1. Introduction

A Freight Forwarder (CG Inc) wants to increase the Volume of Operations they do annually. Its main Service is to Export General Cargo to South America by Multimodal Transportation. The most profitable services of transportation for this Company is:

-Air Transportation

According to the Sales Team and the feedback they got from Customers, the Company has two main weaknesses for them to sell Air Transportation’s service:

-The selling rates for Air Transportation are too high which avoid the Company to sell more services (shipments). If the Company low their rates a 10%, Sales Team forecast an operations Increase of 20% for the following Year

-Pricing Department is not getting good rates with the Airlines.

According to Pricing Department Airlines ask for an Operations Increased of Chargeable weight of 15% to decrease the rates in at least a 10%.

The hypothesis that needs to be proven is that: The higher is the level of Operations then the higher should be the chargeable weight of the cargo to be transported.

The Company hesitates if using Big Data will help them to find the strategy that will lead them to reach their goal. This is why before they decide whether invest or not in Big Data, they would like to check the outcome in minor scale project.

1.1 Goals

-Which is the correct Strategy for Increasing Operations’ Volume in a 20%, for the following Year?

2. Theoretical Framework

In this project we will use some statistical concepts described as follow

Regression analysis can be used for two main purposes:

1. To describe a relation between two or more variables;

2. To predict the value of a variable (the predictand, sometimes called the dependent variable or response), based on one or more other variables (the predictors, sometimes called independent variables. (1)

Types of models A simple correlation or regression relates two variables only; a multiple correlation or regression relates several variables at the same time. Modelling and interpretations are much trickier in the multivariate case, because of the inter-relations between the variables. A linear relation models one variable as a linear function of one or several other variables. That is, a proportional change in the predictor results in a proportional change in the predictand or the modelled variable (1).

-The function ¨ts¨ is used to create time-series objects (2)

-A Pareto chart is a barplot where the categories are ordered in non-increasing order, and a line is also added to show the cumulative sum. (2)

-dplyr is the next iteration of plyr, focused on tools for working with data frames (hence the d in the name). It has three main goals:

Identify the most important data manipulation tools needed for data analysis and make them easy to use from R.

Provide blazing fast performance for in-memory data by writing key pieces in C++.

Use the same interface to work with data no matter where it's stored, whether in a data frame, a data table or database. (3)

-Quality Control Charts (qcc)

Create an object of class 'qcc' to perform statistical quality control. This object may then be used to plot Shewhart charts, drawing OC curves, computes capability indices, and more.(2)

-HeidiSQL is free software, and has the aim to be easy to learn. "Heidi" lets you see and edit data and structures from computers running one of the database systems MariaDB, MySQL, Microsoft SQL or PostgreSQL (4)

3. Prescriptive Data Analysis

Company’s Management is not driven by Prescriptive Analytics, KPI’s are very basic and decision-making process is based on feeling and experience. The analysis of this period with the prescriptive strategy could convince this Company’s Management to start weighing more these tools and Technics. As a first stage for this project the data to be used will be a 5 Years Analysis (2012-2016). CG Inc doesn’t want to take risks using current data with sensible information regarding the Company.

This is the reason why the main source will be a database with approximately 9932 rows with 14 columns.

The context of the Industry during the time frame analyzed will be compared together with Company´s data.

The scope is to analyze as much parameters as possible to start modeling a prescriptive outcome so that this model could be used as a starting point in a following stage by the Company with current data.

The essential requirements are:

-Operations statistical analysis for Air Transportation Service variables since 2012 up to 2016.

-Statistical Prediction Analysis to prescribe an strategy to accomplish the goal.

Desirable requirements are:

-Compare statistically the behavior of the Market with CG’s result.

Future requirements are:

Design an Online Analytical Processing provided with information by a Datawarehouse using internal with structured data and external with structured/unstructured data.

CG works with a software based on a local server using a SQL Server as a RDBMS. This Database storage and process all the information related with Operations and Accounting.

For this project it will be used one Query with 14 columns of the Air Operations Table included in the Database. The query includes the total amount of Air shipments registered for 5 years.

Each row contains as input all the information related with one shipment. The analysis of the data selected allows to describe, diagnose, predict and prescribe the internal variables impacting in the Operations Level for Air Transportation Services.

3.1 Model context

The model will be placed in a Business context where CG is a Freight Forwarder that wants to grow their business operation.

They have a limited customers portfolio, the main market for their services is South America, and the services suppliers are the Airlines flying out the US.

Customer wants to ship a cargo with a certain chargeable weight (determined by weight and dimensions of the cargo) from point A to point B

The sale for an Air Transportation Service is made based on the selling rate (buying rate + Company's markup) offered to customers and determined by the multiplication of the selling rate per kilogram with the chargeable weight of the cargo.

One of the hypotheses that needs to be proven is that: The higher is the level of Operations then the higher should be the chargeable weight of the cargo to be transported.

Once approved selling rate, shipping Instructions for an air shipment are sent for the cargo previously quoted.

3.2 Major constraints

The level of Operations is determined for the quantity of shipments confirmed by Customers accepting the selling rates quoted.

The buying rate of CG is fixed by the Airlines according to the Quantity of Chargeable weight that CG as a customer can contribute for each specific lane (Airport A to Airport B)

3.3 Statistical Analysis

The 6 tables recorded in the Data Warehouse ¨Freight¨ are connected with R to do the Statistical Analysis.

This is the code used to connect R with the Warehouse in My SQL

```
#Connecting R with database freight in MySQL
```{r}
con=dbConnect(MySQL(),
              user='root', password ="",
dbname = 'freight', host ='127.0.0.1')
```
```

The first table to be analyzed is Air Export with 9932 columns and 6 columns as per below:

#	Name	Datatype
1	CodeDestinationAirport	VARCHAR
2	Airline	VARCHAR
3	AgentCode	VARCHAR
4	Year	INT
5	Month	INT
6	Chargeable_weight	FLOAT

Then we can do the queries from that table by putting the following code:

```
```{r}
result<-dbGetQuery(con, statement = "SELECT*FROM airExport")
```
```

3.3.1 Data behavior for Level of Operations per year

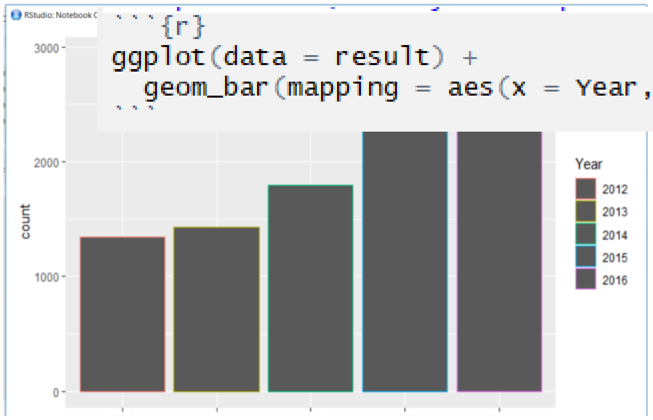
The dataset Air Export with information from within the company, gave us the insight to design the model.

CGs volume of operations show a sustained growth over the years. According to a series decomposition it could be observed the impact of Trend, Seasonality and a Random Variable

Graphic 1.

The code used to get the bar plot was

```{r}
ggplot(data = result) +
  geom_bar(mapping = aes(x = Year, colour = Year))
```

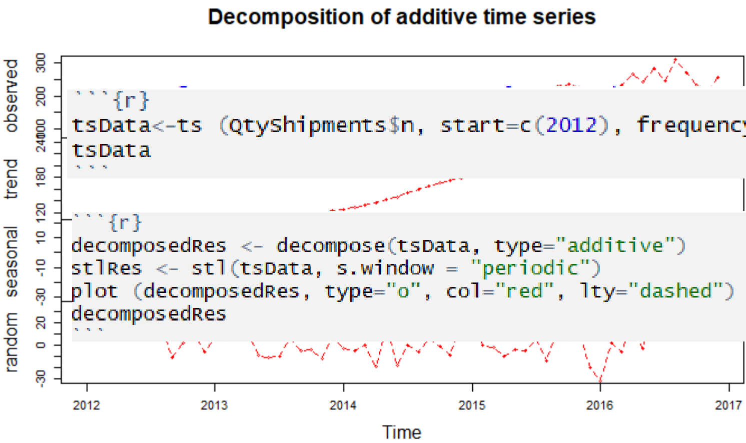


Graphic 2.

The code used to get the series decomposition was

```{r}
tsData<-ts (QtyShipments$n, start=c(2012), frequency=12)
tsData
```

```{r}
decomposedRes <- decompose(tsData, type="additive")
stlRes <- stl(tsData, s.window = "periodic")
plot (decomposedRes, type="o", col="red", lty="dashed")
decomposedRes
```



We can confirm that this is a Business with a continuous growing from 2012 to 2016. According to the series analysis there is remarkable trend, and seasonality with the highest Peak in October and the lowest Peak in February.

3.3.2 Grouping Quantity of shipments and Chargeable weight per month and Year

We have 60 observations per each month for the 5 years studied, and the code used was

Graphic 3

```{r}
byMonthYear <- group_by(result, Year, Month)
Summary <- summarise(byMonthYear,
  count = n(), sum(Chargeable_weight))
Summary
```

Year <int>	Month <int>	count <int>	sum(Chargeable_weight) <dbl>
2012	1	83	33243.00
2012	2	90	29642.48
2012	3	126	49501.47
2012	4	100	47767.00
2012	5	107	26873.00
2012	6	96	25520.06
2012	7	124	42907.32
2012	8	142	61485.00
2012	9	113	38794.00
2012	10	138	41080.00

1-10 of 60 rows Previous 1 2 3 4 5 6 Next

3.3.3 Finding the relation between Chargeable weight shipped per month and Quantity of Shipments per month

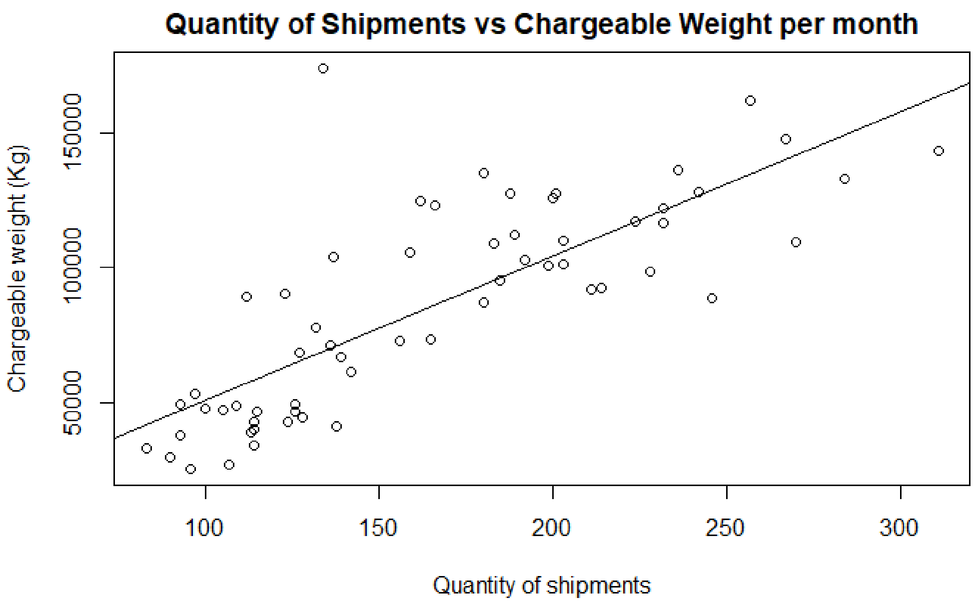
The function used to find the correlation was ¨cor¨

```{r}
cat("Correlation coefficient",cor(QtyShipments$n, ChWperMonth$`sum(Chargeable_weight)`))
```

```
Correlation coefficient 0.7876626
```

Together with correlation coefficient the plot for both variables was obtained to confirm that these 2 variables are dependent.

Graphic 4



The Correlation coefficient of 0.79 shows a high correlation between the chargeable weight and the quantity of shipments, and together with a line that fits in the model of Linear Regression, we could confirm that Quantity of Shipments per month and Chargeable weight per month are dependents. Chargeable weight could be estimated through a Linear Regression Model by Quantity of Shipments

This equation was obtained through the following code:

````
```{r}
relation <- lm(ChWperMonth$`sum(Chargeable_weight)` ~ QtyShipments$n,data=result)
relation
```
````

This is the result

```
Call:
lm(formula = ChWperMonth$`sum(Chargeable_weight)` ~ QtyShipments$n,
    data = result)

Coefficients:
   (Intercept)  QtyShipments$n
       -2544.1           533.8
```

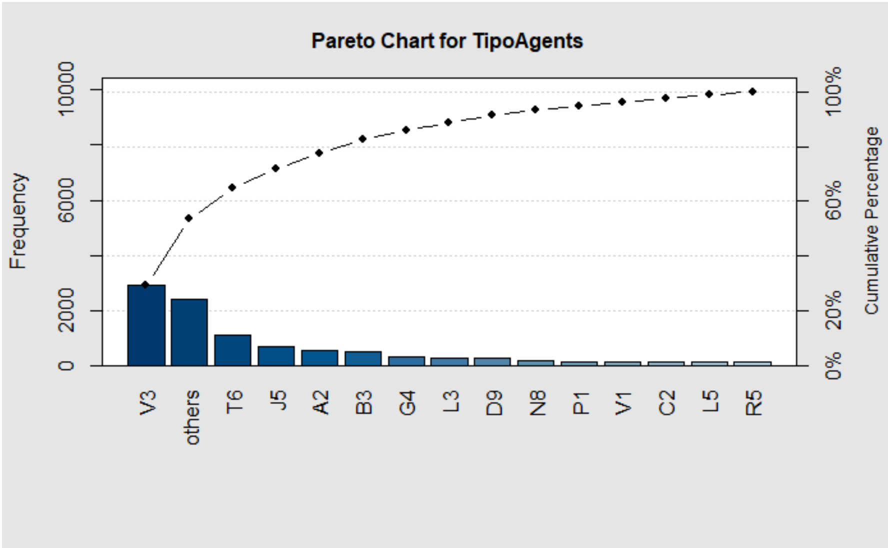
```
Chargeable weight per month= Quantity of Shipments*533.8 - 2544.1
```

3.3.4 Estimating Chargeable weight per Month through Quantity of Shipments

If the Company could low the selling rates in a 10%, Sales could get 20% more of shipments. Let's calculate the average of Quantity of shipments for 2016 to simulate a 20% increase of operations for the following Year

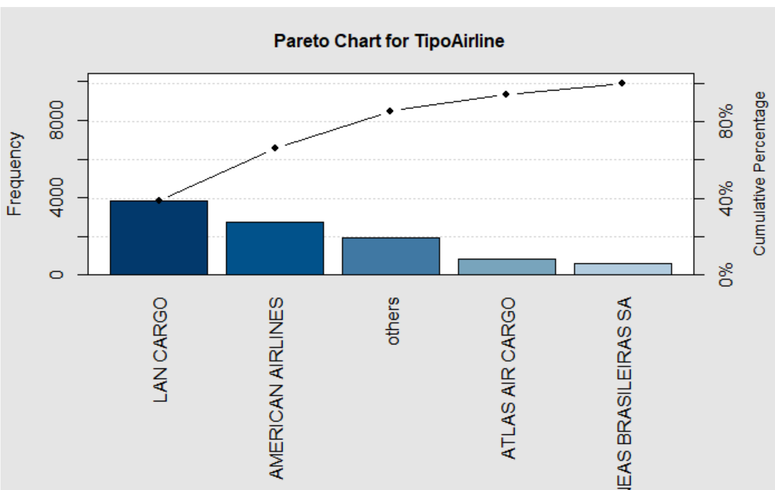
```
```{r}
year2016<-subset(result,Year=="2016")
byMonthYear2016 <- group_by(year2016, Year, Month)
QtyShipments2016<-count(byMonthYear2016)
AveQtyShipments2016<-mean(QtyShipments2016$n)
cat("Average Shipments per month 2016:",AveQtyShipments2016)
```
```

```
Average Shipments per month 2016: 244.8333
```

Then let´s estimate an increase of the Quantity of shipments per month through the average according to the Forecast of Sales assuming that Pricing can get a reduction of 10% on the buying rates

```
```{r}
ShipmentsIncreased<-mean(QtyShipments2016$n)*1.2
cat("Shipments per month increased in a 20%:",ShipmentsIncreased)
```
```

```
Shipments per month increased in a 20%: 293.8
```

To predict the increasing of chargeable weight, let's use the average of shipment´s quantity per month for last Year with the 20% increase previously estimated.

-Predict function: predict(relation, given Quantity of shipments)

```
```{r}
shipmentsInc<-data.frame(QtyShipments$n==294)
estChgwMonth <- predict(relation, shipmentsInc)
estChgwMonth
```
```

Simulating in the linear equation for 294 shipments, the chargeable weight estimated is 154393.1 Kg.

If the average for Chargeable weight per month estimated for 2016 was 119211.6 Kg, and the Chargeable weight estimated for the new level of Operations (through the linear regression) was 154393.1 we could compare both values to check the difference between them.

```
```{r}
difChgw<-(154393.1/AveChWperMonth2016-1)*100
cat("Increase of the Average of Chargeable weight per month:",difChgw,"%")
```
```

```
Increase of the Average of Chargeable weight per month: 29.51184 %
```

The difference is higher than 20%, asked for the Airlines to concede a 10% discount on their rates. Then the recommendation to the Company is to get a commitment with the Airlines confirming them that they could bring to the table at least 20% more of Cargo (Chargeable weight).

3.3.5 Focusing on the main targets.

It will be a waste of time and energy negotiating with all the Airlines for all destinations, as well as providing reductions on selling rates for all customers. Using Paretto Techniques for each of these 3 variables will allow to apply the strategy found on the core of the business of the Company

In 3 steps the same process will be applied for each of these 3 variables

1-Grouping by Agents (Customers)

```
```{r}
byAirport <- group_by(result, CodeDestinationAirport)
destination <- summarise(byAirport,
  "Shipments"= n(), "Chargeable"=sum(Chargeable_weight))
sortdestination <- destination[order(-destination$Shipments),]
sortdestination
```
```

2- Finding the vital customers and grouping in "others"" the Trivial customers according to the Quantity of shipments ordered by customer

```
```{r}
trivialsAgents<-subset(agent,Shipments<120)
mainAgents<-subset(agent,Shipments>=120)
otherAgents<-data.frame("AgentCode"="others","Shipments"=sum(trivialsAgents$Shipments),"Chargeab
le"=sum(trivialsAgents$Chargeable))
destParetoAgents<-rbind(mainAgents,otherAgents)
destParetoAgents$frequency<-prop.table(destParetoAgents$Shipments)
destParetoAgents$cumFrequency<-cumsum(destParetoAgents$frequency)
sortAgent <- destParetoAgents[order(-destParetoAgents$frequency),]
sortAgent
```
```

3-Graphiquing Paretto chart

```
```{r}
TipoAgents<-destParetoAgents$Shipments
names(TipoAgents)<-(destParetoAgents$AgentCode)
TipoAgents
pareto.chart(TipoAgents,cumperc=seq(0,100,by=20,ylim=80))
```
```

Graphic 5

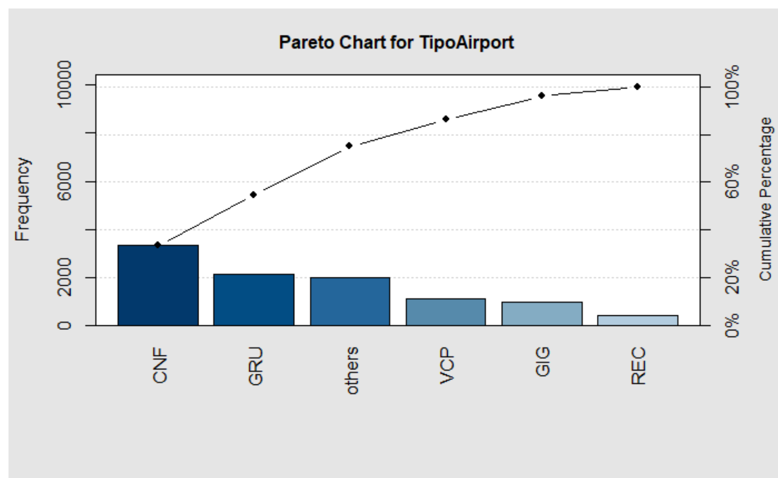


Same 3 steps are followed for the variables Agent Variable (Customers) and Airlines

Graphic 6



Graphic 7



4. Conclusions

- For the past 5 Years the Quantity of shipments per month were observed with a marked Trend to grow and a Seasonality with the highest Peak in October and the lowest Peak in February.

- There is a high a relation coefficient between Quantity of Shipments per month and Chargeable weight per Month.

- Chargeable weigh per month can be estimated through a Linear Regression Model using Quantity of shipments as independent variable.

- It was proved statistically the following strategy:

  - -Pricing can get a commitment with the Airlines by promising an increase of chargeable weight in at least 15% if the Airlines low the buying rate in at least 10%.

  - -Lowing the buying rate with Airlines will allow Sales to low the Selling rates in a 10% and then to increase the Level of Operations in at least 20% more shipments per month.

- Using Paretto Techniques for each of these 3 variables will allow to apply the strategy found on the core of the business of the Company

- The main destinations to negotiate discounts with the Airlines are:

  - CNF---GRU---VCP---REC---GIG

- The quantity of customers to apply the discount on selling rates are 17 customers in a total of 170

- The main airlines to negotiate discounts are:

  - American Airlines

  - Lan Cargo

  - Atlas Cargo

  - ABSA

- According to the correlation analysis of the data obtained from Bureau of Transportation Statistics and CG´s data there is not a good correlation coefficient. Chargeable weight per month of CG doesn't have a dependent relation with chargeable weight shipped out from Miami International Airport.

5. Bibliography

1- Tutorial: An example of statistical data analysis using the R environment for statistical computing D G Rossiter Version 1.4; May 6, 2017

2- https://www.rdocumentation.org/packages/stats/versions/3.5.2/topics

3- https://www.r-project.org/nosvn/pandoc/dplyr.html

4- https://www.heidisql.com/