Software Design Specification

for

Model to Increase the volume of operations for a freight forwarding company

Version 3

Prepared by Adriana Ortiz and Marcel Socorro

CG Company

March 2, 2019

Table of Contents

Revision History

Name	**Date**	**Reason For Changes**	**Version**
Marcel Socorro	02/11/2019	Scope Adjustment	1
Adriana Ortiz	02/17/2019	Data inclusion	2
Adriana Ortiz Marcel Socorro	03/02/2019	Final Revision	3

1. Introduction

This documents purpose is to provide a high level design framework of the prescriptive analysis to find a strategy that will allow GC – a Freight Forwarder company to increase their volume of Operations.

It also provides specific information about the expected input, classes, and functions modeled to bring the expected outcome .

1.1 Goals

The project involves building a statistical model to allow the company to grow its operations volume by delivering a set of actions backed on the data analysis that will led them to reach the goal.

1.2 Statement of scope

Company lacks of measurement for their strategies and results, KPI's are very basic, the management is mostly supporting their decision making process in experience and feelings.
This project will help the management to support their decision making process with data, to solve this specific issue and hopefully scale the data analysis approach to other divisions or situations they might be having.

The project consist in the analysis of an internal and industry databases with relevant data from 2012 – 2016. Data is available in CVS (comma separated value) and will be stored in MySQL and coded with R.

Dataset	Description
Air export	GC Internal database
statsb2012	MIA transportation statistics
statsb2013	MIA transportation statistics
statsb2014	MIA transportation statistics
statsb2015	MIA transportation statistics
statsb2016	MIA transportation statistics

Table 1.

The analysis will focus on the parameters that are having the greatest impact for the company to continue growing their Air transportation service.

The expected outcome is a set of actions to be followed by the management to reach the goal of growing the volume of operations in 20% in the next year.

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:

-How to link statistically the behavior of the Market to 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.

1.3 Model context

The model will be placed in a Business context where CG is a Freight Forwarder that wants to grow their operation volume for an specific service.

The company has a limited customers portfolio, the main market they are servicing its South America, and the suppliers are the airlines flying from out of the US.

An example of the business dynamics is a customer that needs to ship cargo with a given chargeable weight (determined by weight and dimensions of the cargo) from point A to B.

The final price for the end customer its given by:

Selling rate (airline buying rate + company's markup) per Kg * Chargeable weight of the cargo

The hypothesis that needs to be proven its: If the company increases the volume of the operations (shipments), what will be the impact in the increase of the chargeable weight of the cargo to be transported?

.

1.4 Major constraints

- Limited access to industry data reports
- Management hesitant to expose its business data

2. Data design

For purposes of the analysis proposed for this project the following datasets and parameters will be used:

Air Export Dataset

Parameter
CodeDestinationAirport
Airline
AgentCode
Year
Month
Chargeable_weight

Table II

Statsb 2012 – 2016

Parameter
FREIGHT
CARRIER_NAME
ORIGIN
DEST
YEAR
MONTH

Table III

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.

2.1 Data sources

Air Export Dataset: 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. A query from the master database was obtained to develop this project.

Type: .cvs

Structure: Structured

Statsb 2012 – 2016: The Bureau of Transportation Statistics (BTS)

Type: .cvs

Structure: Structured

2.2 Internal data structure

The data to be evaluated for the purpose of this analysis, will be stored in MySQL and all the analysis are to be conducted with R.

2.3 Variable Description

2.3.1 Independent Variables

2.3.1.1 CodeDestinationAirport

1. *It represents the airports to which the cargo is sent*
2. *Data type: text*

2.3.1.2 Airline

1. *It represents the airline used to send the cargo*
2. *Data type: text*

2.3.1.3 Agent Code

3. *It represents the code used to identify the GCs customer*
4. *Data type: text*

2.3.1.4 Qty Shipments

5. *Sum of the # of shipments*
6. *Data type: numeric*

2.3.2 Dependent Variables

2.3.2.1 Chargeable_weight

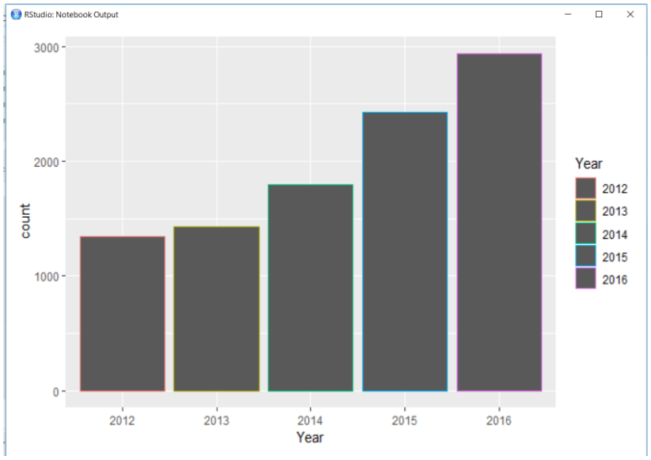
3. *Weight and dimension of the cargo being sent.*
4. *Data type: numeric*

2.4 Pre-design Analysis

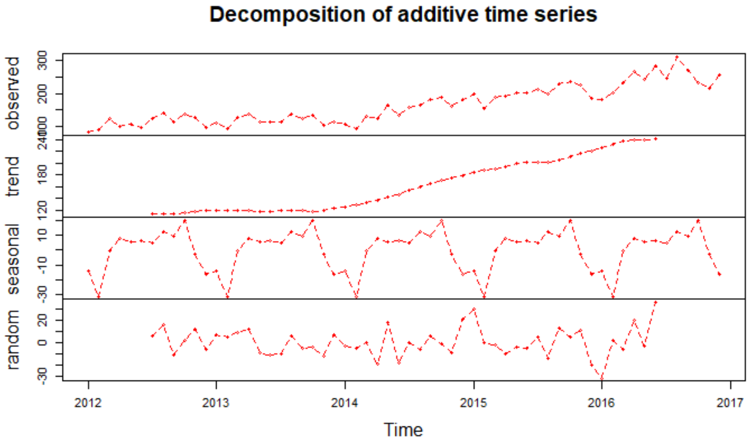
The dataset Air Export with information from within the company, gave us the insight to design the model.

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.



Graphic 2.

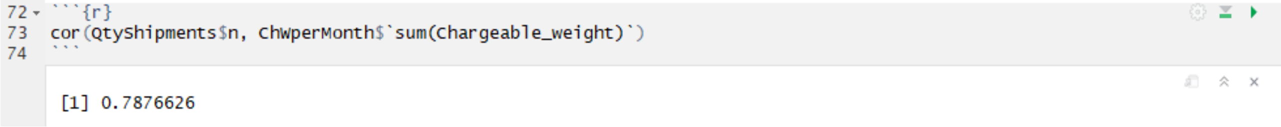


The hypothesis will be proved through the relation between the following variables:

-Qty Shipments per month

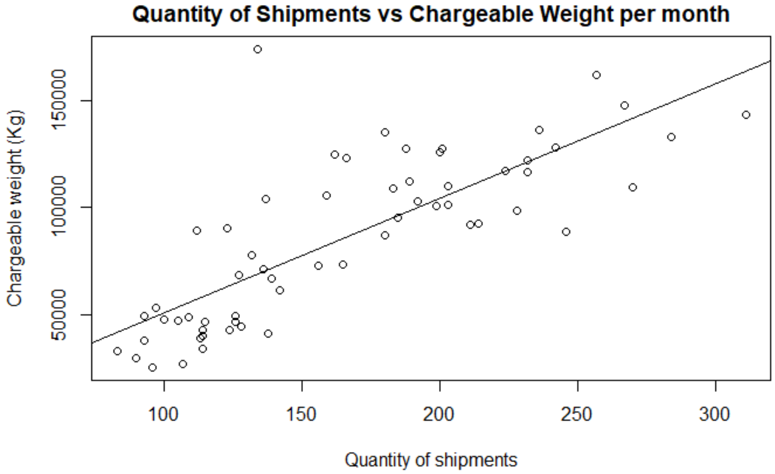
-Chargeable weight shipped by the Company per month

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



~~~
```{r}
cor(QtyShipments$n, ChWperMonth$`sum(Chargeable_weight)`)
```

[1] 0.7876626
~~~



2.5 Tidying procedure

There was no normalization of the variables and no data was deleted

2.6 Database description

To conduct the statistical analysis the following databases were created:

- subsets of the observation per year

New Databases
Year 2012
Year 2013
Year 2014
Year 2015
Year 2016

Table IV

<Database(s) created as part of the application is(are) described. (Provide enough detail to recreate the database, like a full schema.)>

3. Model Architecture

The statistical model is going to use historical data collected from within the company, as well as industry related data to contrast with it.
As the company already uses MySQL, the data was obtained by querying the data base with the relevant parameters for this analysis, and continue to use MySQL to store the project's databases.
The model capabilities are going to be scalable and will offer opportunies to adapat to othes business requirements.
The data is going to be manipulated as required to make the statistical analysis, by creating subsets with R.

Expected input:

- Historical database from the company
- Industry database
- Company main concern to be answered by the model

Expected outcome

-A set of actions to be implemented by the company

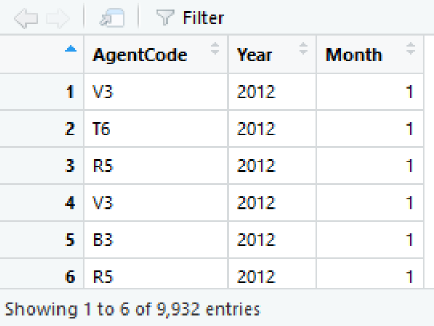
3.1 Type of Model

After analysis the distribution of each variables, a simple correlation or regression its going to be used to determine the relationship between target variables and understand the impact of any variation of them on the other.

3.2 Training set

The training set was constituted by, AgentCOde, Year, Month.

The size of the training set was 9,932 entries



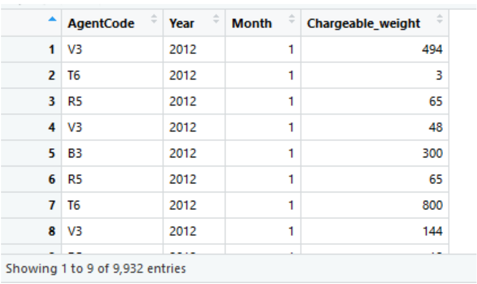
Filter

	AgentCode	Year	Month
1	V3	2012	1
2	T6	2012	1
3	R5	2012	1
4	V3	2012	1
5	B3	2012	1
6	R5	2012	1

Showing 1 to 6 of 9,932 entries

3.3 Testing set

The testing set is constituted by AgentCode, Year, Month, ChargeableWeight



	AgentCode	Year	Month	Chargeable_weight
1	V3	2012	1	494
2	T6	2012	1	3
3	R5	2012	1	65
4	V3	2012	1	48
5	B3	2012	1	300
6	R5	2012	1	65
7	T6	2012	1	800
8	V3	2012	1	144

Showing 1 to 9 of 9,932 entries

4. Approach

- identifying which are the top 20 clients for volume of operations within the company
- determining which are the airlines are getting most of the business with the company to identify opportunities to negotiate

4.1 Implementation Details

The implementation requires

- R
- Heidy MySQL
- XAMMP
- Office suite

4.2 System requirements

-Commercial machines can be used to run the model
-Wi-Fi connection available to connect with databases
- Cloud storage

5. Testing Strategy

5.1 Classes of tests

The test was conducted using the summarize and count function.

5.2 Expected response

The model should bring the count or Qty of shipments done per AgentCode,monthly and then yearly.

5.3 Performance bounds

A commercial machine should be able to run the test.

6. References

https://www.transtats.bts.gov/DL_SelectFields.asp

Appendix A: Glossary

Air Freight: the transportation of goods by aircraft

Chargeable weight: determined by weight and dimensions of the cargo

R: language and environment for statistical computing and graphics

Correlation coefficient: a number between −1 and +1 calculated so as to represent the linear dependence of two variables or sets of data

Heidy MySQL: is a free and open-source administration tool for MySQL and its forks, as well as Microsoft SQL Server and PostgreSQL.

XAMMP: free and open-source cross-platform web server solution stack package developed by Apache,

CSV: comma separated values

Appendix C: Issues List

- *T*he function predict with R, it not returning the expected value.