

**TASK**

**Exploratory Data Analysis on the Chocolate Imports Data Set**

**Introduction**

This data set consists of 5 types of entities: (a) the specification of an exporter in terms of name, (b) its assigned quantity exported, (c) Customs Value (d) date imported (e) the port received.

The first factor (a) corresponds to a quantity of confectionary imports into South Africa, under the Tarif Code 18062010, measured in kg. The second (b) relates the import value (measured in Rands). (c) refers the date the imports ere received by the Customs office, whilst (e) specifies which port of entry was used for the select imports.

**ANALYZE DATA**

**Categorical Data**

districtofficename, countryoforiginname,

**Ordinal**

transportcodedescription, worldregion

**Numerical Data**

Continuous

statisticalquantity, customsvalue, calendaryear

**Mixed Discrete Data**

yearmonth (mix of year and numerical month)

**Redundant Data**

tradetype, countryofdestination, countryofdestinationname, tariff,

tariffanddescription, statisticalunit, transportcode, districtofficecode,

countryoforigin

**Features that may contain Errors**

Errors may be linked to user capture error in categorical or numerical data.

WorldRegion would also benefit from correcting WorldRegion and split (between N.America

and S.America)

**Features that contain null\empty\blank entries**

transportcodedescription > countryoforigin contain missing values

transportcodedescription > countryoforigin incomplete dataset

**REDUNDANT DATA**

* all data, and codes reflects chocolate imports to South Africa (can be incorporated into heading)

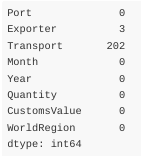
i.e.. Cocoa confectionary imports (18062010) to South Africa in KG

* all units measured in kilograms
* all transport codes have better descriptions (use that column)
* information already available in yearmonth
* district office code has better descriptions to use
* country of origin has descriptors available in separate column

**DATA CLEANING & MISSING DATA**

**MISSING DATA**

* Notice ‘Transport not specified” unknown variable, replace with NaN
* Count missing values per respective column



**Imputate Missing Data**

1) Transport Values (202 values missing)

* Ascertain Transport based on receiving office
* Method of transport can be derived from noting receiving port



**2) Exporter Values (3 values missing)**

* Conclude all port entries through Kopfontein relate to Botswana (only missing 3 values - all other samples relate Country 'Botswana' to port 'Kopfontein'
* Incorrect 'South Africa' values for exporter, change to 'Botswana' due to entry port INTO SA from Botswana

DATA STORIES AND VISUALIZATIONS

**Examine cases with single variable: Univariant Analysis**

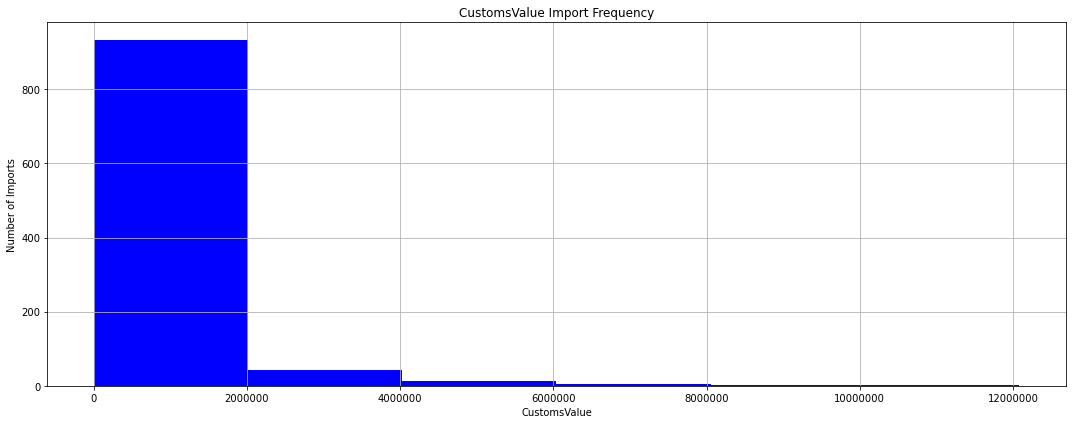
**Exporter and Transport Frequency Analysis**

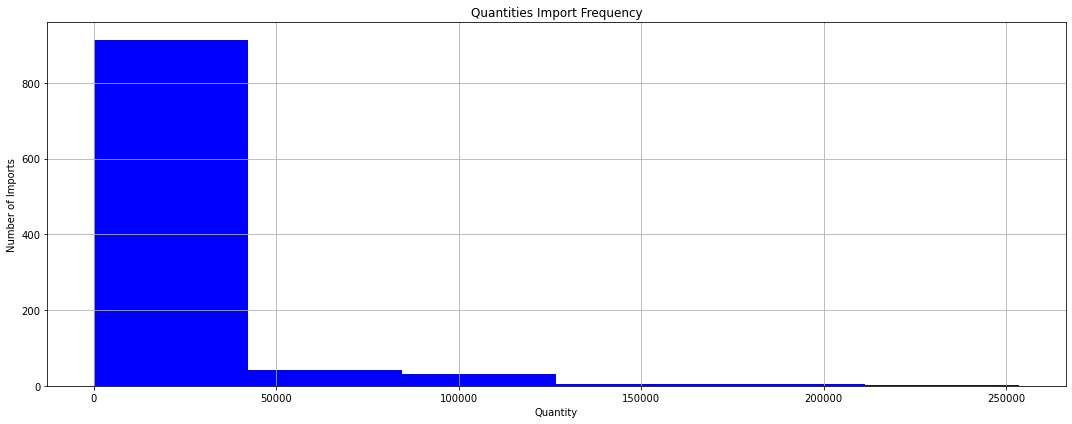
****

**Conclusion**

* Most chocolate imports are transported via Maritime Shipping (59.3%), followed by Air (25.7%) and Road (15.05)

**Mean CustomsValue and Quantity Imported**

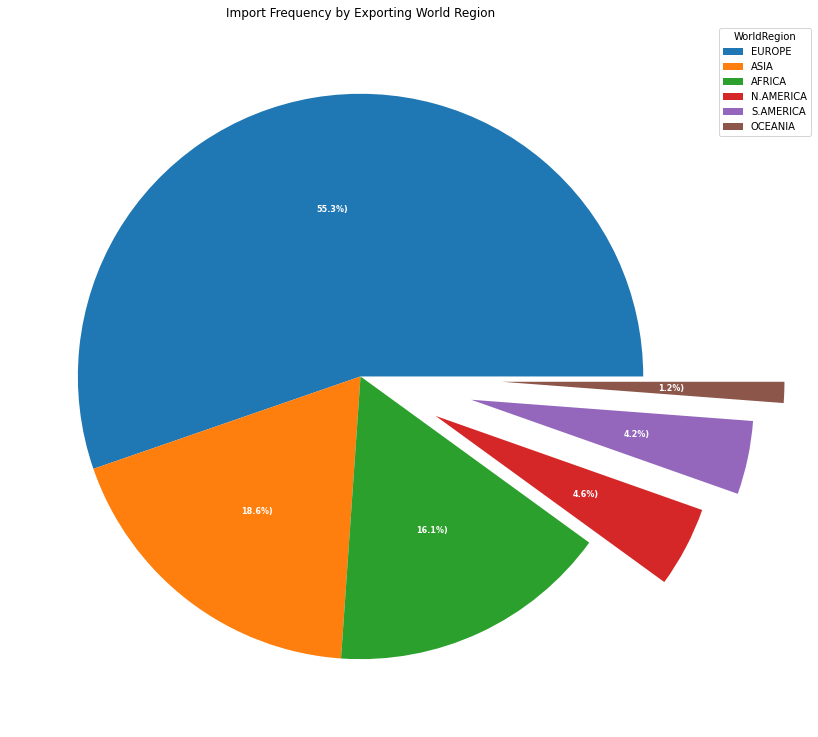
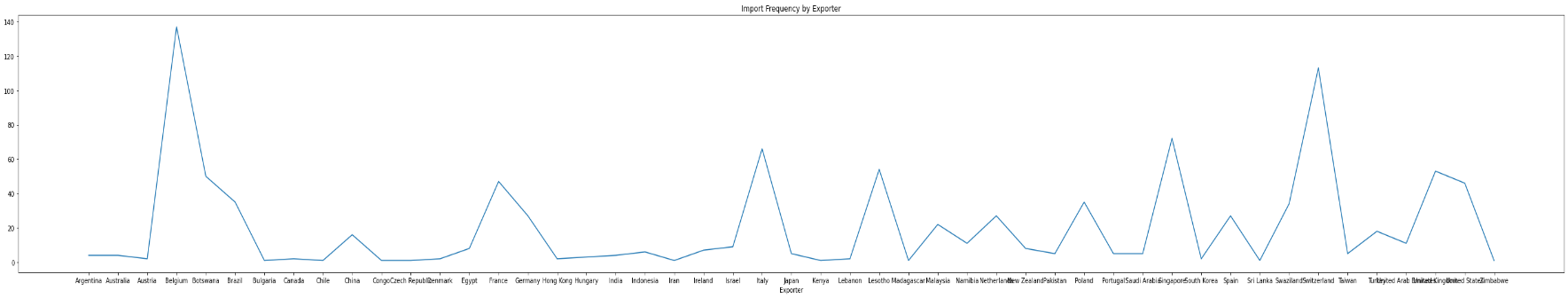




**Conclusion**

* The highest and average Quantity of 45000kg of chocolate is imported monthly, contributing R2 000 000 on average
* Chocolate is rarely imported at higher volumes

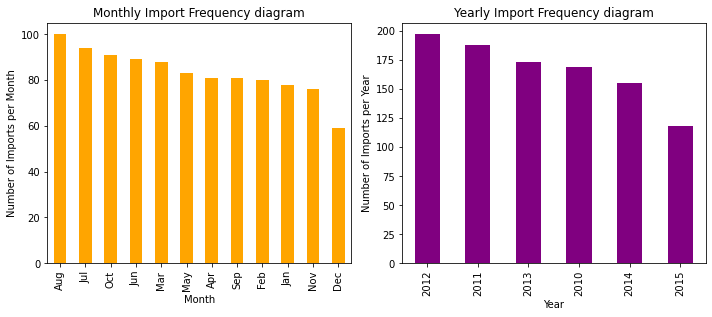
**Import Frequency by Locale**



**Conclusion**

* Europe, Asia and Africa contribute most to the SA import of chocolate The Americas contributing very little, with Oceania (Australia and New Zealand) contributing the least
* The largest importers by capital are Belgium, Switzerland, Singapore, Italy, Botswana, France and UK

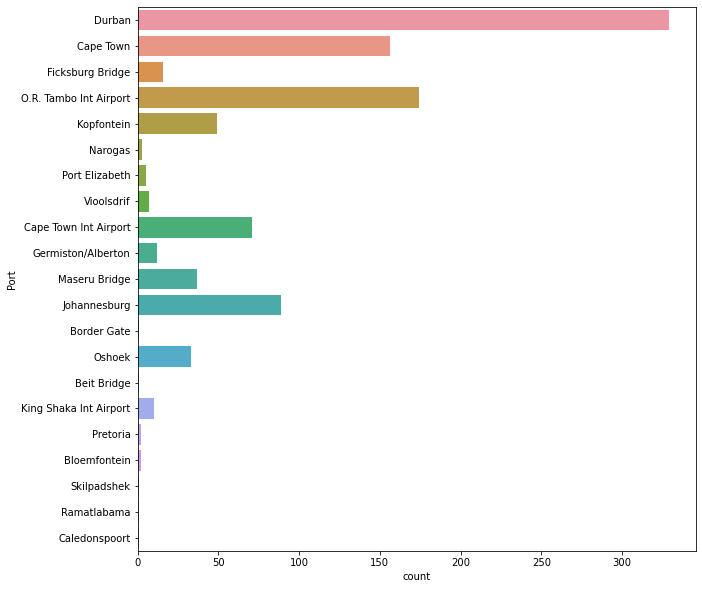
**Import Frequency Analysis**



**Conclusion**

* Winter Months (Mar-Aug) see the highest volumes of imports, with as much as a 40% difference in imports during summer
* 2011 and 2012(max) saw the largest import volumes, tapering off steeply (-25%) in 2015

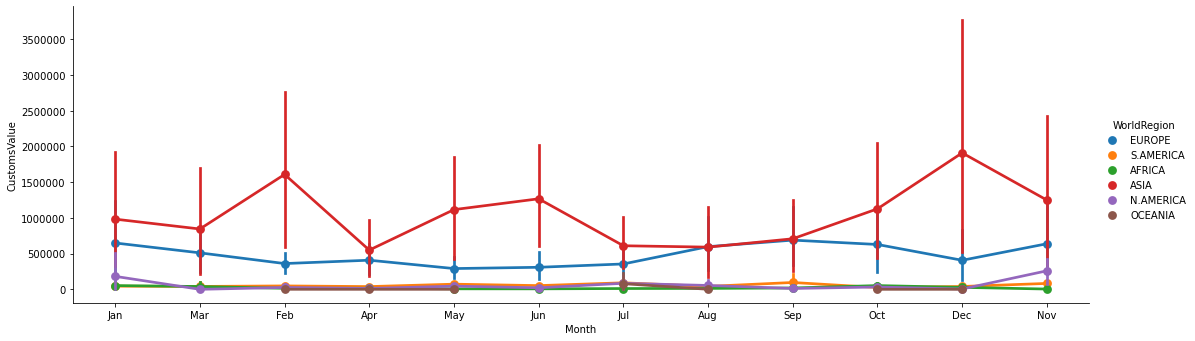
**Port Use Frequency Analysis**



**Conclusion**

* The most popular ports for import by shipping are based in Durban and Cape town
* OR Tambo Int Airport managing larger quantities than that of Cape Town Harbour

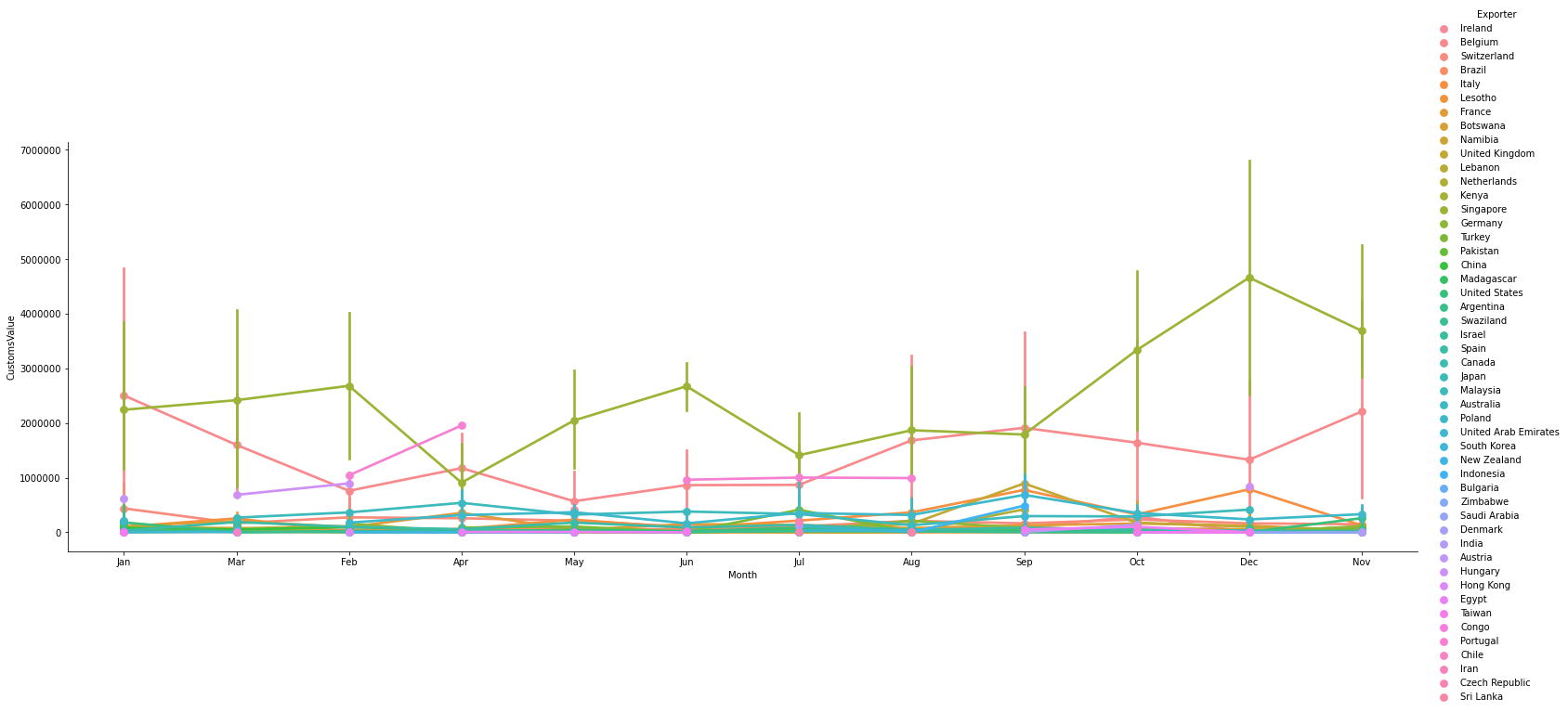
**CustomsValue Contribution by world region**



**Conclusion**

* In terms of Value contribution from CustomsValues, we note that seasonal February and December sees large import of chocolate from Asian countries - presumably to maintain demand for festivities (Valentine’s Day and Christmas)
* Note the large gauge to indicate the volatility during the times of trade as price vary largely from a high of R3.5m imposed, to a low of R75 000. This may be due to larger competitive markets competing for chocolate at the same time
* Interestingly, colder months also see a spike in demand and CustomsValue by Asian countries

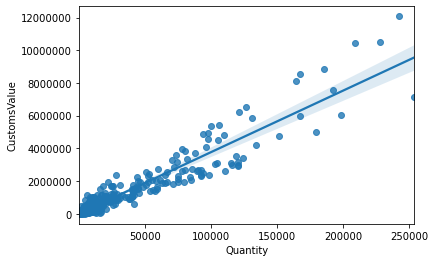
**CustomsValue Contribution by Importer**



**Conclusion**

* By country alone, we see Singapore is by far the single largest contributor to our Chocolate Customs Market, outstripping the world-renowned chocolate manufacture Belgium by as much as 4x fold.
* Notice the peaks and troughs match the larger Asian contribution, in terms of shape and scale - supporting the conclusion that Singapore is SA's largest importer of chocolates

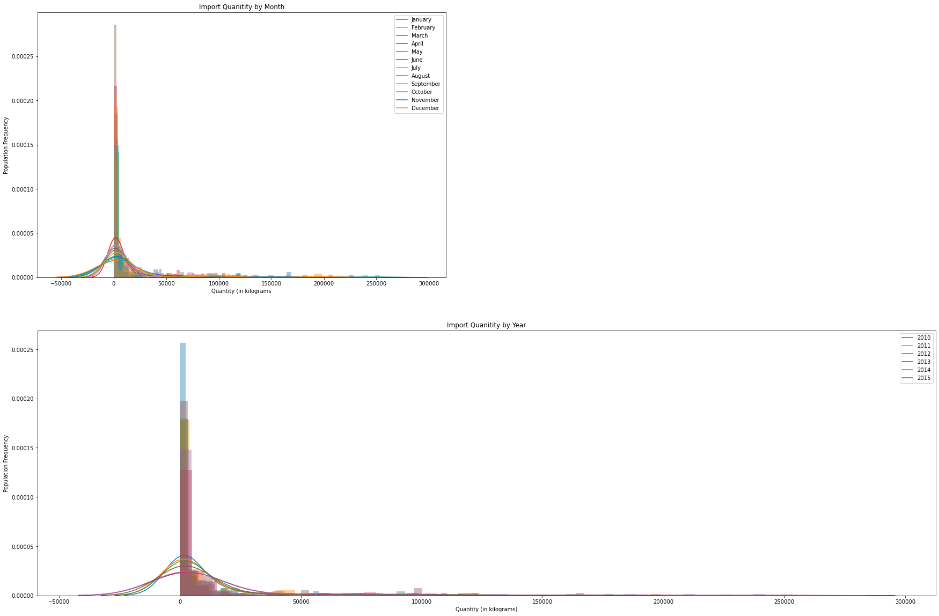
**CustomsValue Related to Quantity Imported**



**Conclusion**

* Customsvalue remains largely stable in lower quantities imported
* Although linear pricing trend is noted, we see a more upward weighted distribution of chocolates toward larger quantities imported.
* We can assume that larger quantities imported are subjected to larger global competition

**Quantity Imported Related to Relative Timeframe Imported**



**Conclusion**

* On Average, we note that April is set to see the largest number of for small Quantities imported
* It is noted that most imports of small volume
* 2010 sees the market and subsequent imports demand larger imports of chocolates
* 2013 and 2014 saw market imports drop 45% in quantity

**THIS REPORT WAS COMPILED BY : Lèan Gouws**

