

Supply Chain Data Analytics

Stan Brouwer¹, Liz Chan², Maaïke Lamberst³, Niek Schroor⁴

¹Vrije Universiteit,
²Master TSCM,
³Supply Chain Data analysis,
⁴Group 10,

Introduction

We analyze, forecast and interpret the [Superstore sales](#) provided by [Tableau](#) using different statistical and machine learning methods.

We describe our work in the PDF version. However, we would like to recommend reading our quarto manuscript *here* as it contains the **relevant** R code in the Article Notebook.

0.1 Data Pre-processing

The superstore data set we selected is of high quality. Thus we do the required data pre-processing, but included the hypothetical steps we would take were our data of lower quality to communicate our understanding of the data pre-processing process.

We took the following pre-processing steps:

- Improved column names by removing whitespaces
- Removed the Row_ID column as it can be inferred by it's index
- Removed all columns with a single unique value, as storing these would be [redundant](#)
- Ensured machine-readable date formats in yyyy-mm-dd as these usually differ per locale.
- Ensured proper decimal separators
- calculated the number of missing values (both NA and empty string “”) per column.

Source: [Article Notebook](#)

```
[1] "None of the columns contains missing values"
```

Source: [Article Notebook](#)

After these steps (and transposing the table for better document formatting), the data looks as follows:

Table 1: First 5 Rows of the Data (Transposed)

Order_ID	CA-2016-152156	CA-2016-152156	CA-2016-138688
Order_Date	2016-11-08	2016-11-08	2016-06-12
Ship_Date	2016-11-11	2016-11-11	2016-06-16
Ship_Mode	Second Class	Second Class	Second Class
Customer_ID	CG-12520	CG-12520	DV-13045
Customer_Name	Claire Gute	Claire Gute	Darrin Van Huff
Segment	Consumer	Consumer	Corporate
City	Henderson	Henderson	Los Angeles
State	Kentucky	Kentucky	California
Postal_Code	42420	42420	90036

Corresponding author: Stan Brouwer,

Region	South	South	West
Product	FUR-BO-10001798	FUR-CH-10000454	OFF-LA-10000240
Category	Furniture	Furniture	Office Supplies
Sub-Category	Bookcases	Chairs	Labels
Product	Flash Somerset Bookcase	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	Self-Adhesive Address Labels for Typewriters by Universal
Sales	261.96	731.94	14.62
Quantity	2	3	2
Discount	0	0	0
Profit	41.9136	219.5820	6.8714

Source: [Article Notebook](#)

There is some more processing to do, such as removing outliers. However, by doing so we impose our own assumptions on the data (possibly the outliers are actual sales?). We will visualize and qualitatively evaluate the data first, and then decide what other processing steps to take.

0.2 Section

This is a simple placeholder for the manuscript’s main document ([knuth84?](#)).

1 + 1

[1] 2

Source: [Article Notebook](#)

0.3 Introduction

Source: [Article Notebook](#)

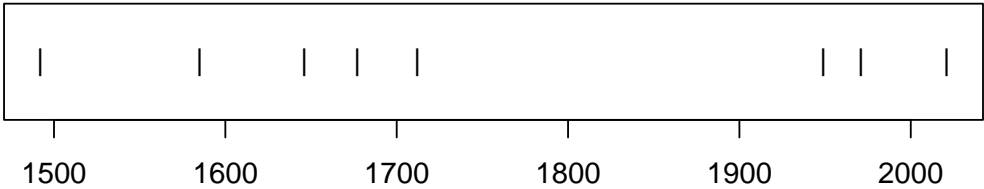


Figure 1: Timeline of recent earthquakes on La Palma

Source: [Article Notebook](#)

Source: [Article Notebook](#)

Based on data up to and including 1971, eruptions on La Palma happen every 79.8 years on average.

Studies of the magma systems feeding the volcano, such as Marrero et al. (2019), have proposed that there are two main magma reservoirs feeding the Cumbre Vieja volcano; one in the mantle (30-40km depth) which charges and in turn feeds a shallower crustal reservoir (10-20km depth).

Eight eruptions have been recorded since the late 1400s (Figure 1).

Data and methods are discussed in Section 0.4.

Let x denote the number of eruptions in a year. Then, x can be modeled by a Poisson distribution

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (1)$$

where λ is the rate of eruptions per year. Using Equation 1, the probability of an eruption in the next t years can be calculated.

Table 2: Recent historic eruptions on La Palma

Name	Year
Current	2021
Teneguía	1971
Nambroque	1949
El Charco	1712
Volcán San Antonio	1677
Volcán San Martin	1646
Tajuya near El Paso	1585
Montaña Quemada	1492

Table 2 summarises the eruptions recorded since the colonization of the islands by Europeans in the late 1400s.

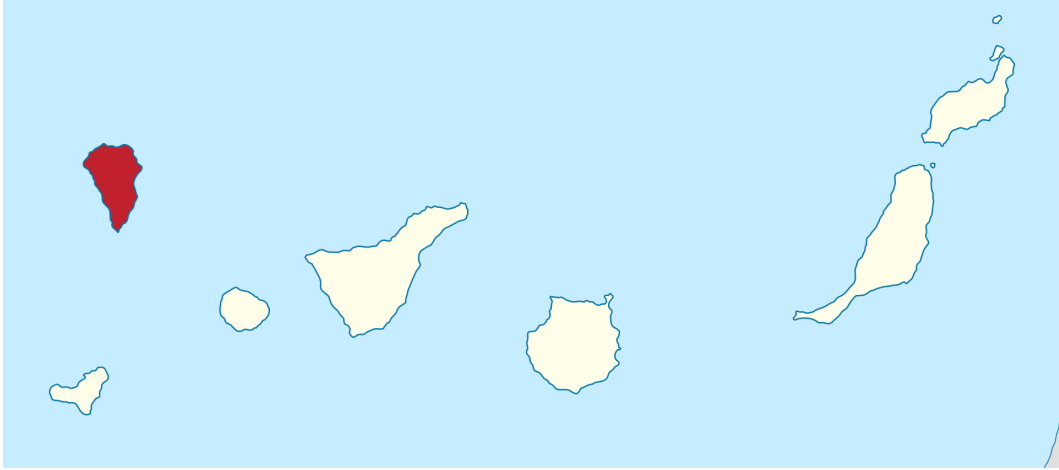


Figure 2: Map of La Palma

La Palma is one of the west most islands in the Volcanic Archipelago of the Canary Islands (Figure 2).

0.4 Data & Methods

0.5 Conclusion

References

Marrero, J., García, A., Berrocoso, M., Llinares, Á., Rodríguez-Losada, A., & Ortiz, R. (2019). Strategies for the development of volcanic hazard maps in monogenetic volcanic fields: The example of La Palma (Canary Islands). *Journal of Applied Volcanology*, 8. <https://doi.org/10.1186/s13617-019-0085-5>