

# Supply Chain Data Analytics

Stan Brouwer<sup>1</sup>, Liz Chan<sup>2</sup>, Maaïke Lamberst<sup>3</sup>, Niek Schroor<sup>4</sup>

<sup>1</sup>Vrije Universiteit,

<sup>2</sup>Master TSCM,

<sup>3</sup>Supply Chain Data analysis,

<sup>4</sup>Group 10,

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.

```
[1] "No missing values"
```

Source: [Article Notebook](#)

We also ran some descriptive statistics to check unlikely or impossible values, outliers, means, etc.

```
# A tibble: 5 x 19
  Order_ID Order_Date Ship_Date Ship_Mode Customer_ID Customer_Name Segment
  <chr>    <date>    <date>    <chr>    <chr>    <chr>    <chr>
1 CA-2016-152~ 2016-11-08 2016-11-11 Second C~ CG-12520  Claire Gute  Consum~
2 CA-2016-152~ 2016-11-08 2016-11-11 Second C~ CG-12520  Claire Gute  Consum~
3 CA-2016-138~ 2016-06-12 2016-06-16 Second C~ DV-13045  Darrin Van H~ Corpor~
4 US-2015-108~ 2015-10-11 2015-10-18 Standard~ S0-20335  Sean O'Donne~ Consum~
5 US-2015-108~ 2015-10-11 2015-10-18 Standard~ S0-20335  Sean O'Donne~ Consum~
# i 12 more variables: City <chr>, State <chr>, Postal_Code <dbl>,
#   Region <chr>, Product_ID <chr>, Category <chr>, `Sub-Category` <chr>,
#   Product_Name <chr>, Sales <dbl>, Quantity <dbl>, Discount <dbl>,
#   Profit <dbl>
```

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

---

Corresponding author: Stan Brouwer,

sales?). We will visualize and qualitatively evaluate the data first, and then decide what other processing steps to take.

Table 1: First 5 Rows of the Data

OrderID	ShipDate	ShipTime	ShipClass	CustomerID	CustomerName	CustomerAddress	CustomerCity	CustomerState	CustomerPostalCode	Region	ProductID	ProductCategory	ProductSubcategory	ProductDescription	ProductPrice	ProductQuantity	ProductDiscount	ProductProfit
CA-2016-11-152158	2016-11-11	11:00	Second Class	CG-1252	CG- Claire Guter	Consolidated	Henrieville	NY	12438	South	HURFURNITURE	BO-10001798	Bookcase	Somerset Collection Bookcase	261.26	100	0.00	41.9136
CA-2016-11-152158	2016-11-11	11:00	Second Class	CG-1252	CG- Claire Guter	Consolidated	Henrieville	NY	12438	South	HURFURNITURE	CH-10000454	Chair	Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	731.34	100	0.00	219.5820
CA-2016-06-138682	2016-06-06	06:16	Second Class	LV-1304	LV- Darrick Van Huff	Corporate	Los Angeles	CA	90035	West	OFFICE	LA-10000246	Label	Self-Adhesive Address Labels for Type-writers by Universal	14.62	200	0.00	6.8714
US-2015-10-108966	2015-10-10	10:18	Standard Class	SO-20335	SO- Sean O'Donnell	Consolidated	Fort Lauderdale	FL	33315	South	HURFURNITURE	TA-10000577	Table	Bretford CR4500 Series Slim Rectangular Table	957.57	75	0.45	-383.0310
US-2015-10-108966	2015-10-10	10:18	Standard Class	SO-20335	SO- Sean O'Donnell	Consolidated	Fort Lauderdale	FL	33315	South	OFFICE	ST-10000766	Storage	Edison Fold 'N Roll Cart System	22.32	80	0.20	2.5164

Source: [Article Notebook](#)

## 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](#)

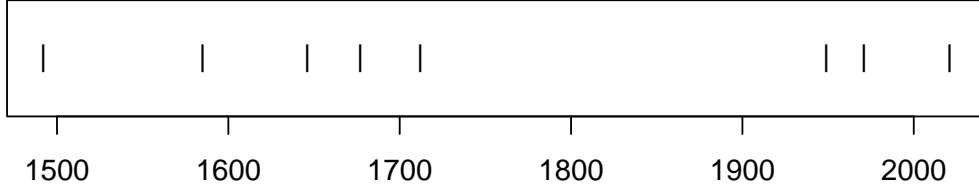


Figure 1: Timeline of recent earthquakes on La Palma

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  $\lambda$  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.

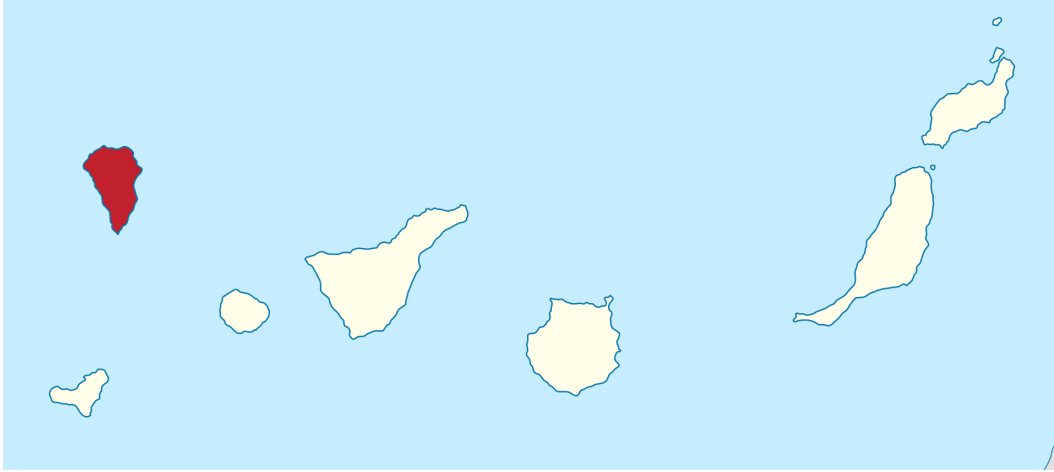


Figure 2: Map of La Palma

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

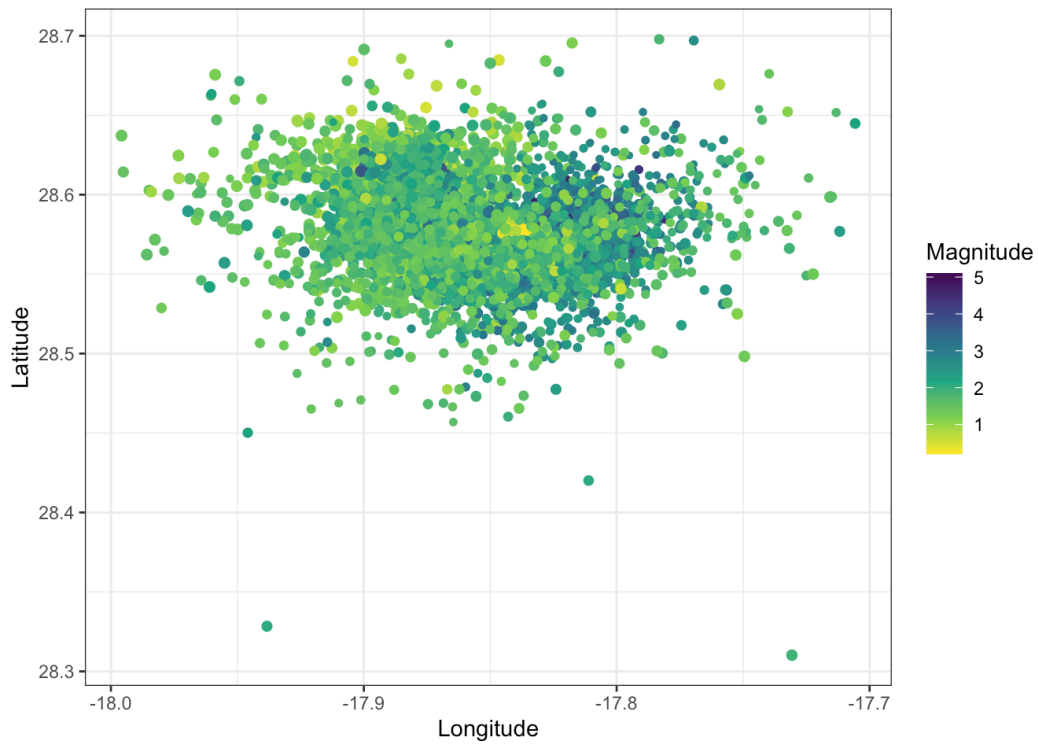


Figure 3: Locations of earthquakes on La Palma since 2017

Source: [Explore Earthquakes](#)

Figure 3 shows the location of recent Earthquakes on La Palma.

**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>