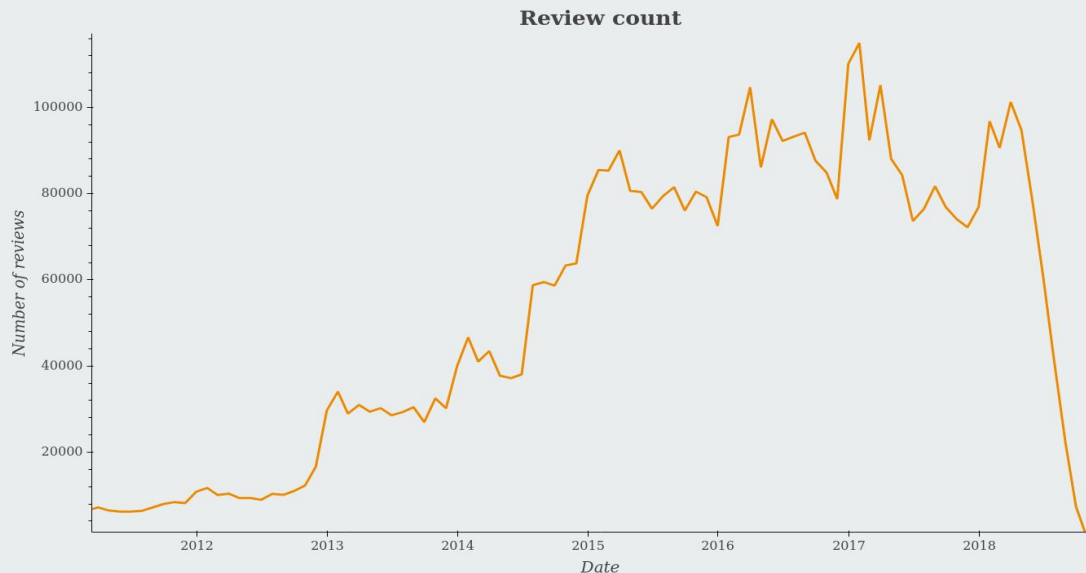# Impact of events and trends on Amazon

Jules Gottraux, Lucien Iseli, Florian Ravasi and Christina Mantonanaki

A quick view on the dataset

- 5 million reviews
- 2.7 million reviewers
- 5GB of text and metadata
- More than 90% between 2013 and end of 2018

Review count

From raw data to usable one

- NLP pipeline using nltk library on reviews' text

- Reviews' text : main axis of analysis

Here is how the pipeline works:

Amazon Customer

⭐⭐⭐⭐⭐ **yummy brownies and cookies**

May 30, 2018

Size: 18 Pieces | **Verified Purchase**

These brownies and cookies are very tasty. I would order them again. They also remained fresh for several days.

One person found this helpful

| Helpful | ⌄ Comment | Report abuse | Permalink |

1. Lower-case
2. Remove numbers and punctuation
3. Tokenization
4. Lemmatization
5. Remove stop-words

Amazon Customer

⭐⭐⭐⭐⭐ **yummy brownies and cookies**

May 30, 2018

Size: 18 Pieces | **Verified Purchase**

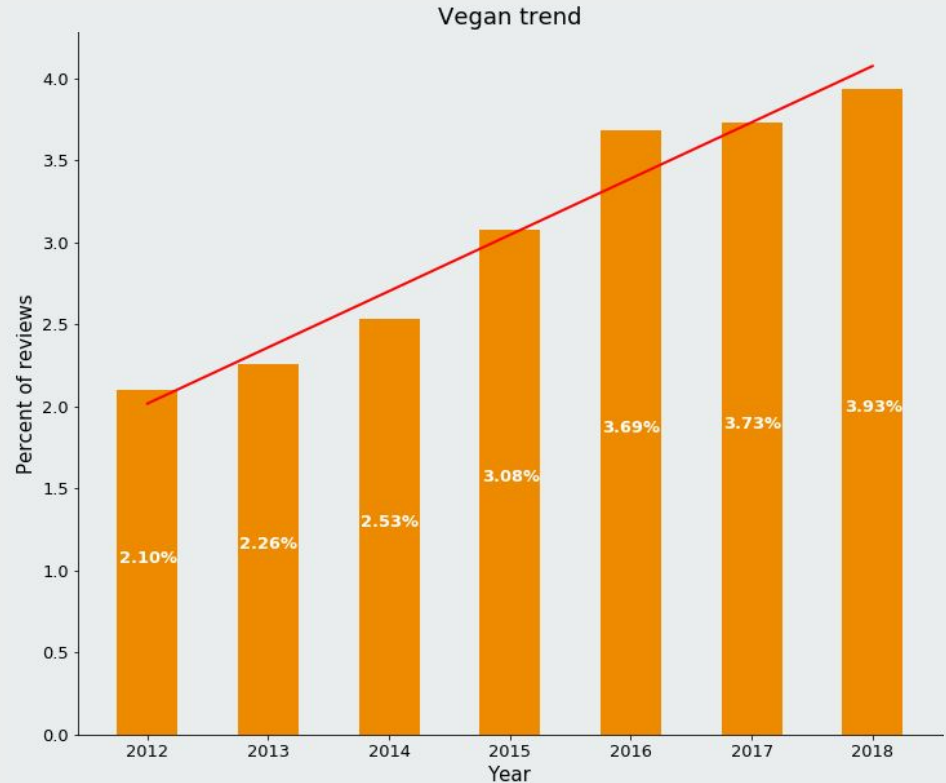**brownie cooky tasty would order also remain fresh several day**

One person found this helpful

| Helpful | ⌄ Comment | Report abuse | Permalink |

Trends throughout the years

- Proportion of reviews with trend related words
- Linear regression and hypothesis testing for slope
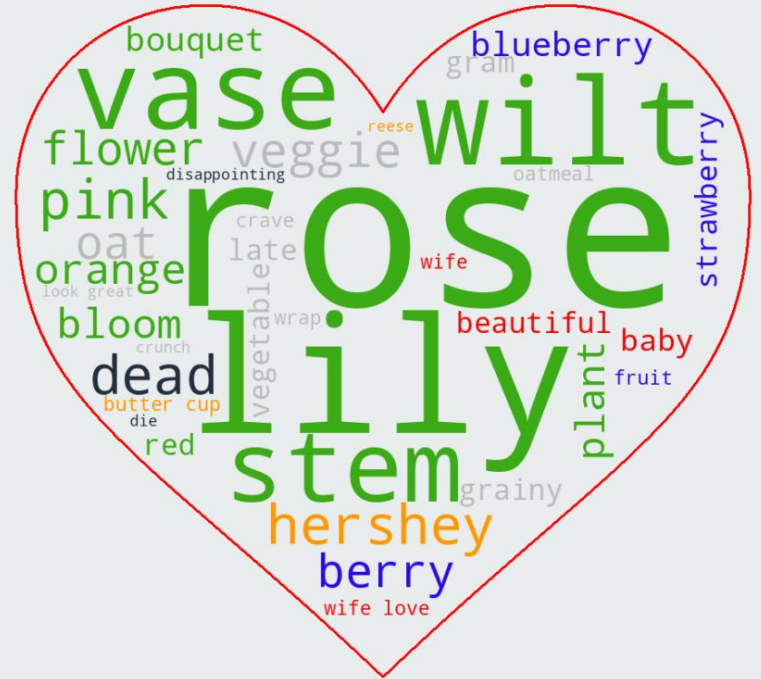- P-value essentially 0

Vegan trend

Event characterization

The evolution of word frequencies allow us to measure how much a word characterises an event.

Several themes present:

- Flowers in green
- Love and family in red
- Sweets and chocolate in orange
- Fruits in blue
- (Flowers's) death in black
- Unrelated in gray
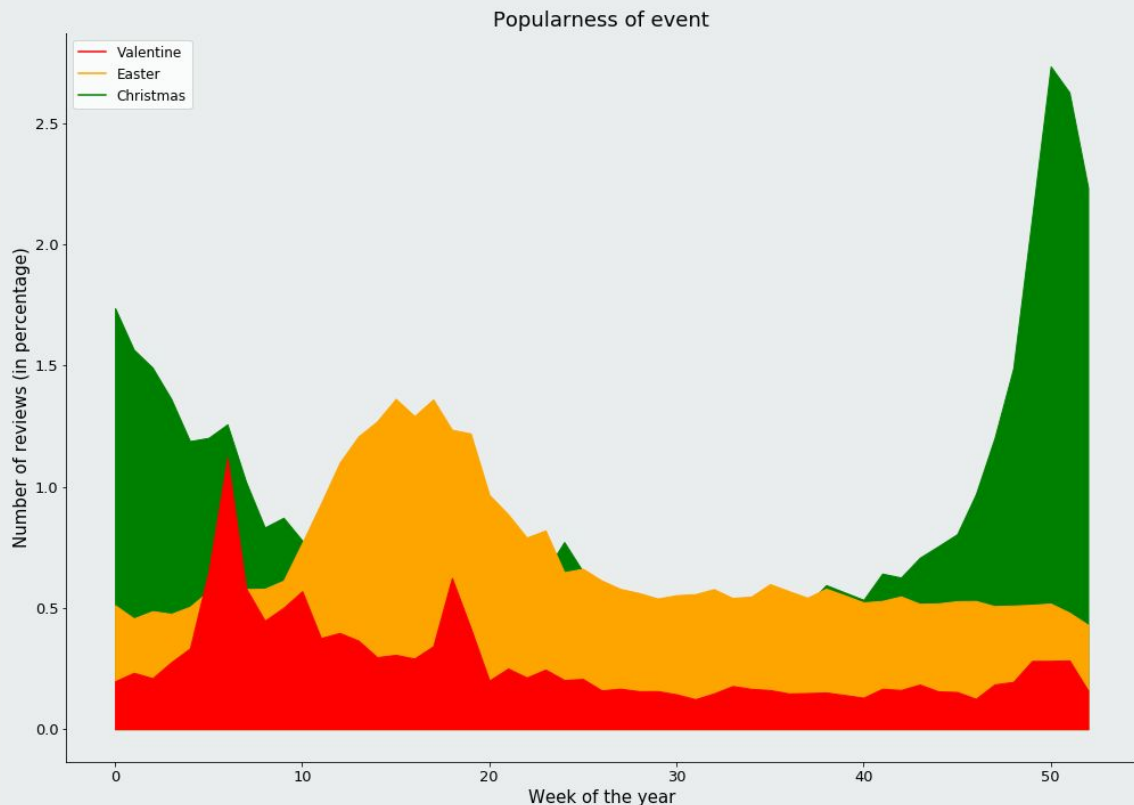
Structural bayesian time series

1. Choose one product that you suspect the event has influence on
2. Predict what the product would have been without the impact of the event

Need to choose correlated products, otherwise it does not make sense.

Prediction of nut gift basket
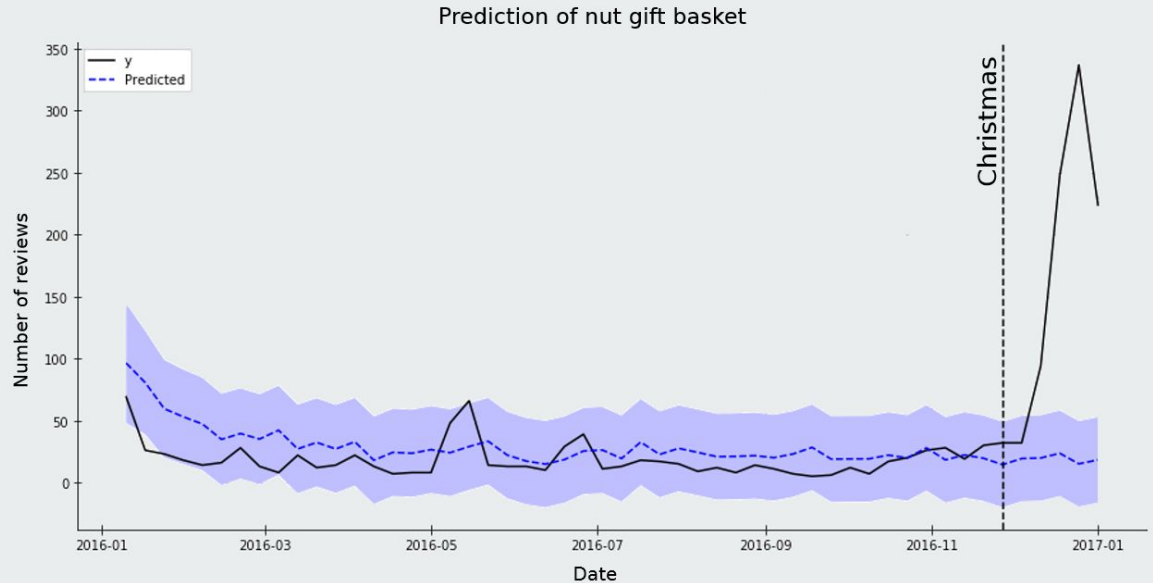
Events comparison

- Measure presence of event using its words characterization
- Christmas is clearly the most popular event

Popularness of event

Thank you!