

# The Queen's Birthday

## Abstract

This project explores trends in the google search query “queen’s birthday” using monthly data from the year 2004 to the present for New Zealand (NZ) and Great Britain (GB). First, we visualise the data, highlighting seasonal spikes in the number of queries, and noting some differences between the trends in GB and NZ. Second, we conduct a Bayesian time series analysis, fitting a Normal Dynamic Linear Model (NDLM) with seasonal and polynomial trend components for the NZ data. The upshot is a model which captures the seasonal trend very well and accurately forecasts a year in the future, with the caveat that the search query trend is likely to fundamentally change in the coming decade, necessitating re-modelling.

## Introduction

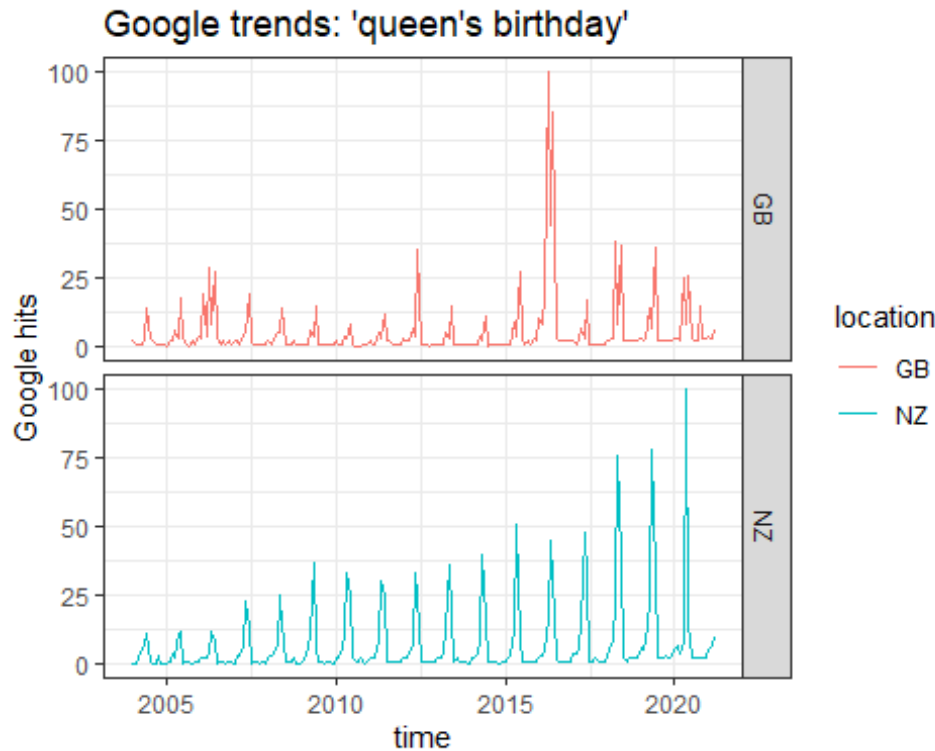
This project is using gtrends hits for “queen’s birthday”. The aims of the project are to:

1. Explore both the GB and NZ data visually and understand the key features.
2. Choose one of the two datasets and model it with a suitably constructed NDLM, keeping the final year’s worth of data held out.
3. Produce the filtering, smoothing and forecasting distributions for the model.
4. Analyse and constructively criticise the model’s effectiveness.

## Exploratory Data Analysis

While we do not specify *which* queen, the geo-locations of GB (Great Britain) and NZ (New Zealand), both of whose Head of State is Queen Elizabeth II, narrows down the likely candidates. To add some spice to the modelling problem, she has two “birthdays”:

- an actual date of birth (21st April)
- an Official Birthday (2nd Saturday of June)



As can be seen from the plot, there is an annual spike in hits across April, May and June in both the GB and NZ data. There are some characteristic differences however.

GB: has an exceptional spike during 2016. This corresponds to the QEII's 90th birthday. To a lesser extent this spike is also seen in 2006, the year of her 80th birthday. The GB data's peaks are also noticeably bimodal: outside of special anniversaries, there is a lesser peak in April (QEII's actual birthday) and a greater peak in June (her Official birthday).

NZ: does not demonstrate any exceptional spike, but instead a generally increasing trend in the seasonal spikes. Also noticeable is that the peaks are all unimodal and centred in May.

Some of these characteristics may be explained by different media coverage, events and celebrations in the two countries, e.g. Bank Holidays in the UK.

### In summary:

- Search query: "queen's birthday"
- Sampling rate: monthly aggregates
- Location: GB (Great Britain) and NZ (New Zealand)
- Characteristics:
  - evident seasonal trend across both locations peaked annually around May
  - GB: exceptional number of hits in 04/2016 and 06/2016 and bimodal peak structure.
  - NZ: generally increasing amplitude of seasonal peaks over time and unimodal peak structure.

## Modelling

We will proceed to model, focusing just on the NZ data. We will use an NDLM model with:

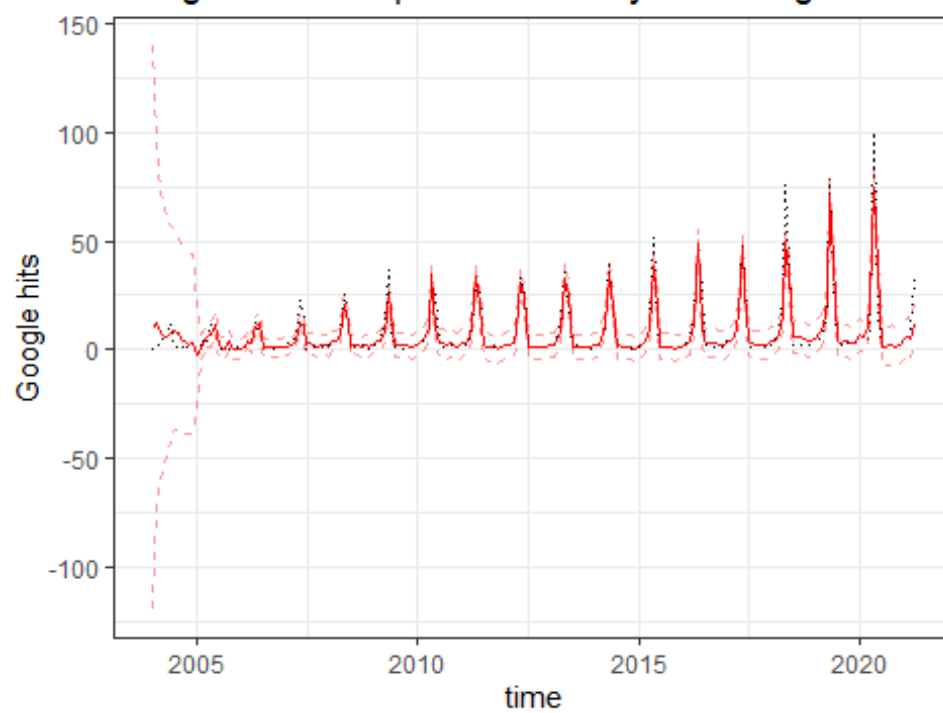
- a seasonal component of fundamental period 12 to capture the annual seasonality. To fit the shape of the within-season curve we will use 5 additional harmonic components.
- an order 2 polynomial component to capture the increase in the amplitude over time
- priors centred on a mean of 5 and a system covariance of 10 to approximately capture the default hit-level between peaks.

Based on the above, the model specs we use are:

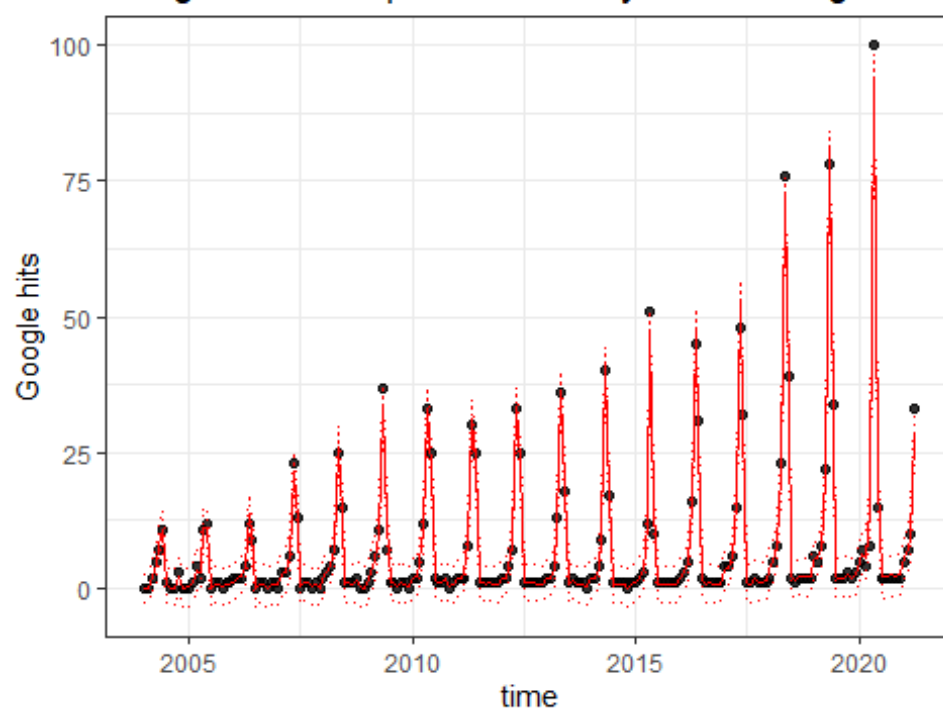
```
model_seasonal <- dlmModTrig(s=12, q=6, dV=10, dW=10)
model_trend <- dlmModPoly(order=2, dV=10, dW=rep(1,2), m0=c(5, 5))
model <- model_trend + model_seasonal
model$C0 <- 10*diag(13)
n0=1
S0=10
```

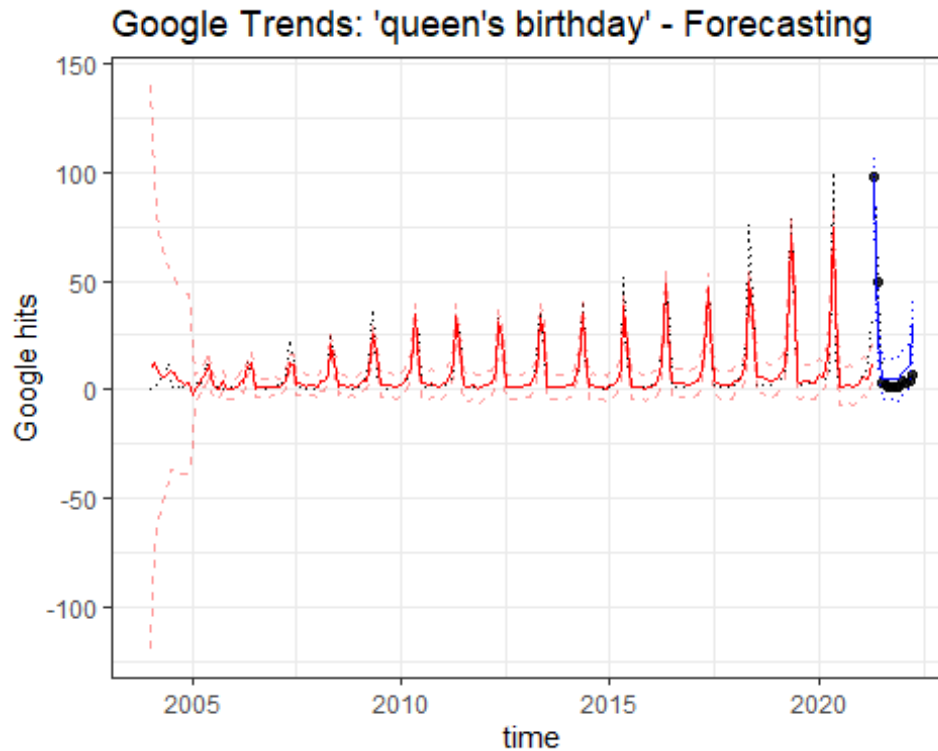
We assume unknown observational variances and unknown system covariance chosen via a discount factor. To choose this we use a grid search with the MSE criterion over the interval  $[0.7, 1]$ .

Google Trends: 'queen's birthday' - Filtering



Google Trends: 'queen's birthday' - Smoothing





## Discussion

### Performance

The forecasting performance (blue lines) on the held-out data is great! It accurately captures the increasing seasonal peak trend, and the quiet-month predictions are well within the confidence intervals. It is slightly over-estimating the number of hits for July 2021 - March 2022, but mostly well within the credible intervals. The last timepoint is April 2022, which, at the time of completing this project is still underway. We expect that the number of hits is incomplete at this time and suggest that the number is also weighted towards the end of the month, and thus our estimate may still prove to be accurate. The selected system covariance discount factor was 0.89.

### Caveats

This model is likely to perform well only for a set number of years going forward: it is unlikely to persist beyond the (finite) lifetime of the Queen as the next Head of State is most likely to be male. There is thus a reasonable chance that the trend will fundamentally alter or die out completely, necessitating a new modelling strategy.

One flaw with using an NDLM for count data is that count data which fluctuates near 0 is not well modelled under a Normal assumption. As such, our model would readily predict a negative number of hits (which is impossible in reality) and although that didn't turn out to be the case for our posterior trends, the credible intervals do include negative values and

are thus not well calibrated. To improve this within a Bayesian context we might need to consider other families of models using different conjugate pairs, perhaps with exponential link components.

## Improvements

Possible next steps to improve model:

- Include covariates such as special year indicators for particular birthdays such as 80th, 90th, 100th(!) and coronation anniversary years.
- Revisit the normality assumption and consider an assumption more appropriate for count data.