Replication Write Up

Lucy Kinnear

April 2022

1 Summary

The project I chose to replicate was attempting to create a model that could predict patterns and trends in Google search terms for “depression” by modelling on past trend information.

1.1 Data

Original data involved search data across the US from Google Trends, and as part of the additive part of the project I looked at Australia and Norway’s Google Trends data.

1.2 Motivation

The purpose of Big Data utilisation is shifting from monitoring to forecasting. Being able to predict trend, seasonality, and cycles of interests is valuable to both businesses and social stakeholders. Google Trends is an excellent tool to observe information seeking in real time, offering instant reflection of needs, wants, and interests of users.

1.3 Why Do We Care?

Forecasting data, if reliable, could inform timing of advertising for products, campaigns, or social intervention programmes. For example, if there is a peak in search trends in January for “depression” this could inform mental health practitioner companies (e.g. “Better Help”) or charities like Samaritans to boost their advertisement spending then or to start a new initiative at that point in the year.

2 The Replication

The data itself was very simple, therefore required no cleaning. They first created a time series from the data, then split the data set 80/20. They then trained and tested the split data set and used ARIMA to predict. I have included code and results below.

2.1 Code

```
# changing the dataset into a timeseries
depressionUS.timeseries <- ts(depressionUS,start = c(2004,1), frequency = 12)
depressionUS.timeseries

## 80/20 split data train test
h2US <- 38L
trainUS <- head(depressionUS.timeseries, round(length(depressionUS.timeseries) - h2US))
testUS <- tail(depressionUS.timeseries, h2US)

trainDataUS <- trainUS
testDataUS <- testUS
# arima model
arimaModUS <- auto.arima(trainDataUS, stepwise=FALSE, approximation=FALSE)
arimaMod.FrUS <-forecast(arimaModUS,h=38)
# plotting
plot(arimaMod.FrUS, main ="US Forecast")
lines(testDataUS, col="red")
legend("topleft",lty=1,bty = "n",col=c("red","blue"),c("testData","ARIMAPred"))
# checking
accuracy(arimaMod.FrUS, testDataUS)
```

2.2 Results

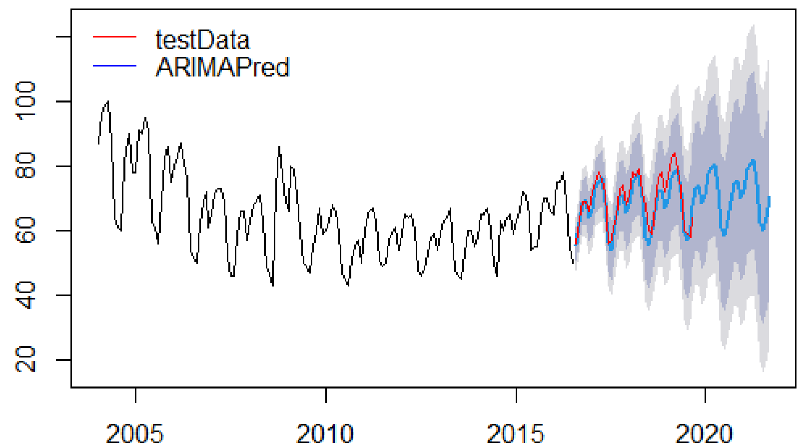
Let's look at the trained forecast versus the untrained forecast on the next page.

3 Adding to the Project

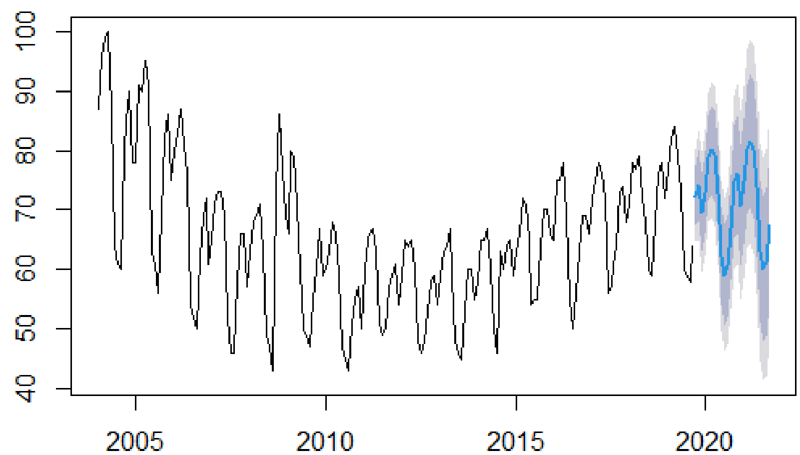
3.1 Norway and Australia

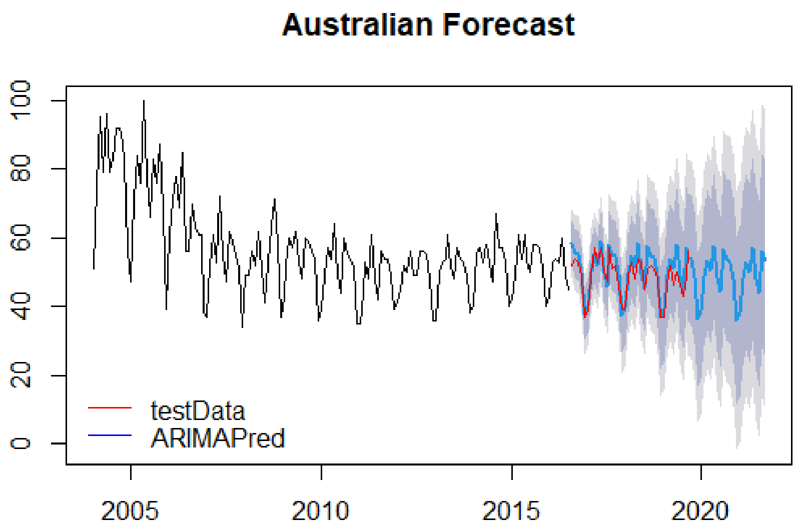
Using the same methods as the original, I investigated Australia and Norway. We saw distinct seasonality in the US dataset, consistent with seasonal depression or low mood as the weather turns colder and the days become shorter. Australia is a sunny country, Norway is a cloudy and cold country, I wanted to see if seasonality was impacted differently because of the weather situation. There was a clear seasonality to the data across US and Australia, but less pronounced in Norway. This may reflect that Australia and US have more significant change in weather in comparison to Norway. Trend wise, Norway is increasing its search of "depression" but US and Australia's trends are declining.

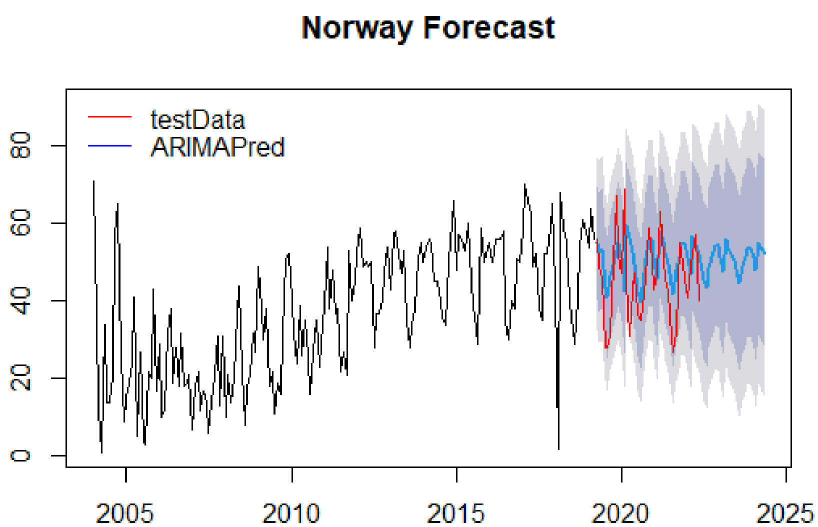
US Forecast



non-trained US Forecast



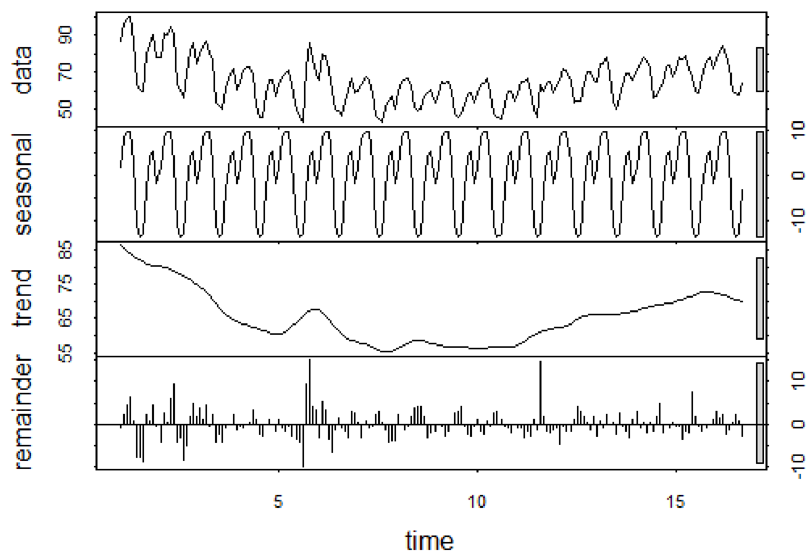

Australian Forecast
testData
ARIMAPred
0
20
40
60
80
100
2005
2010
2015
2020


Norway Forecast
testData
ARIMAPred
0
20
40
60
80
2005
2010
2015
2020
2025

3.2 Decomposing the Time Series

My final contribution was decomposing the US and Australian time series data. I chose these two countries because the trend and seasonality in Norway was less clear.

Indiviual Components of US Time Series



Indiviual Components of Australian Time Series