Assignment 3 - Report

Introduction:

For this assignment, I chose to research COVID-19 cases in England by using the data provided on the Government website[1]. My main focus for this assignment was on the number of new infections and determining which model best fits how the number of infections per day has changed between 30th March 2020 to 12th April 2023.

I started by performing some basic pre-processing and manipulation, which included turning the dataset into a data frame and then selecting the columns of data containing the number of cases each day. From this, I then produced several graphs to help fit the best model to the entire dataset and the data between 10th November 2021 and 13th January 2022 (650-715 days).

Analysis of the Data:

Number of COVID-19 Cases per day:

Figure 1 - Number of COVID-19 Cases per day in England.

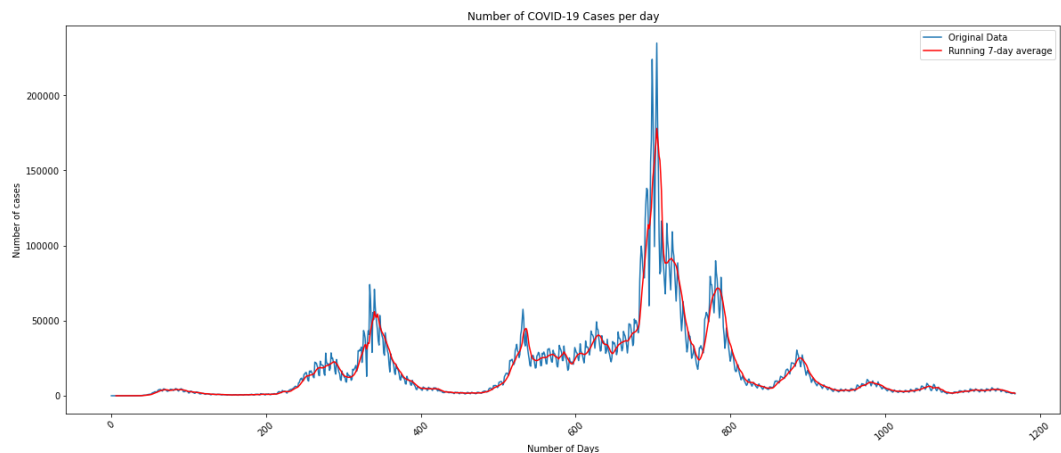Figure 1 shows the number of COVID-19 cases between 30th March 2020 and 12th April 2023. It is clear to see there are five peaks, with the highest peak of 220,000 cases at about 715 days from the start. The first of the peaks of 75,000 cases at 300 days saw a rapid decrease in cases, down to 3,000-6,000 cases per day, due to the implementation of a national lockdown. As the lockdown was eased off, the number of new cases began to grow from 525 days, until 700 days when the number of cases grew exponentially. This brought about new measures (e.g., face masks, COVID passes, vaccines, and testing) which helped to reduce the number of new cases.

Model Fitting:

I chose a quartic model to be the best model to fit the dataset. When comparing the cubic, quartic (figure 2) and quintic models, the cubic model was slightly underfitting the data whilst the quintic model is slightly overfitting, which can be seen in the tails of both models. The best fit parameters are: a = 109 ± 13, b = -2429 ± 269, c = 16119 ± 1790, d = -28463 ± 4398, e = 12828 ± 3171.

Figure 2 - Quartic Model Fitted to the Number of COVID-19 Cases per day in England.

Exponential Model Fitting:

For fitting an exponential model to the dataset (figure 3), I decided it would be best to fit the model to the small section of the data, where the growth of cases is the quickest. the section I chose was between 650 and 715 days as this was the lead-up to the highest peak in cases. the best fit parameters are: $x = 3.45e^4 ± 5.68e^3$, $y = 1.53e^{-1} ± 2.15e^{-2}$.
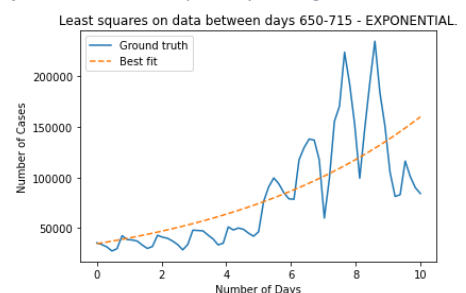
Figure 3 - Exponential Model Fitted to the Number of COVID-19 Cases per day in England between the 600th-715th days.

Summary:

Overall, the COVID-19 cases in England dataset shows it is very difficult to have a model which perfectly fits the data between 30th March 2020 to 12th April 2023, as there are many peaks, with different rates of increase and decrease. However, it is much easy and more sensible to fit a model to a specific section of the data e.g., the rise up to the highest peak between 650-715 days.
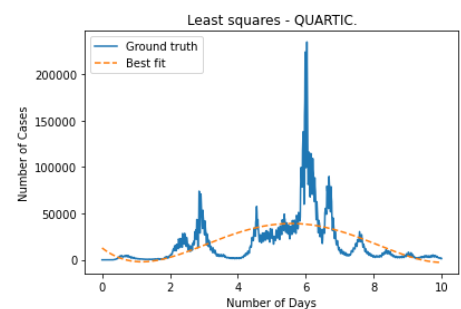
---

[1] Data Sourced from: https://coronavirus.data.gov.uk/details/cases?areaType=nation&areaName=England