

## Analysing the Relationship between the Human Development Index and Mortality Rates within countries.

The datasets were obtained from World Bank (World Bank, 2020) displaying the mortality rates of countries expressed as deaths per 1,000 individuals per year. The primary aim of this study was to apply three clustering algorithms; Affinity, K-Nearest Neighbours (KNN), and Agglomerative on mortality rates in 2019 and 1960 to observe countries with similar relationships, as illustrated in Figure 1.

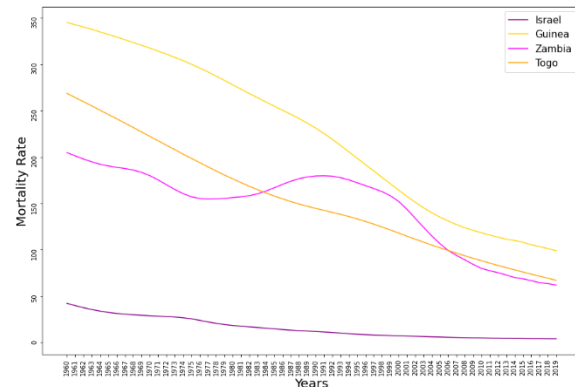


Figure 2 – Line graph of countries from each cluster using the Affinity algorithm.

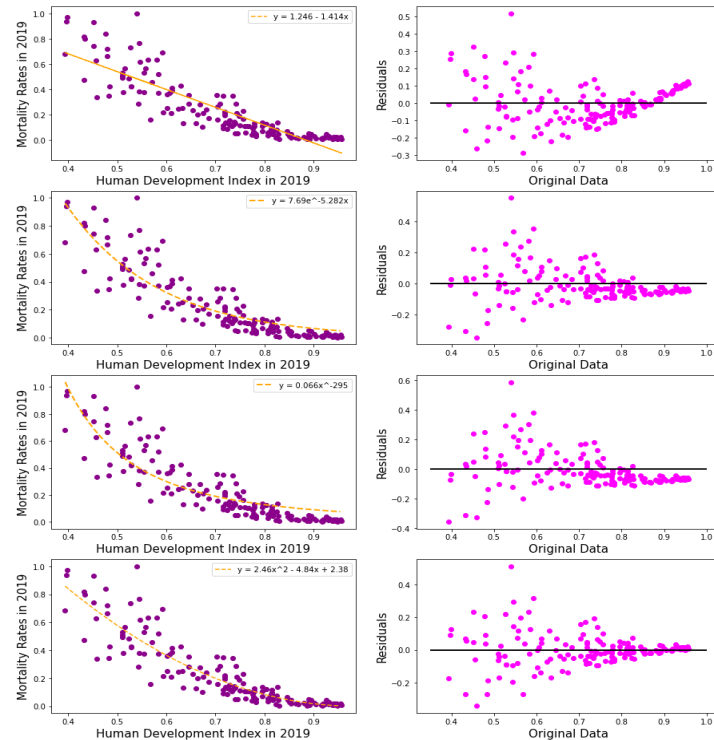


Figure 3 – Predictive models of A linear model B exponential model C Polynomial 1 model D Polynomial 2 model, and their respective residuals.

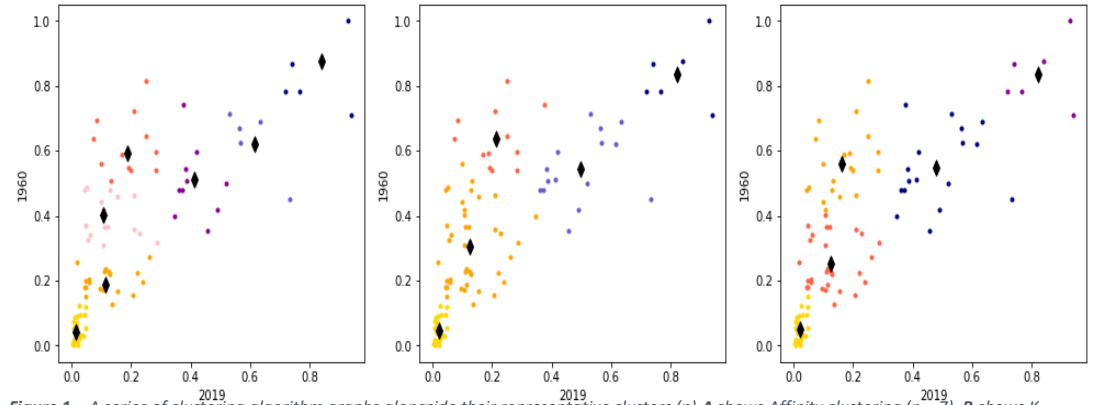


Figure 1 – A series of clustering algorithm graphs alongside their representative clusters (n) A shows Affinity clustering (n = 7), B shows K-Nearest Neighbours clustering (n = 5), and C shows Agglomerative clustering (n = 5).

Further visual analysis of the affinity clustering was observed in Figure 2. It can be shown that countries with the least change on mortality rates in 1960 and 2019 were successfully clustered into separate groups. Whereas clustered countries with the greatest change in mortality rate from 1960 to 2019 overlapped. Interestingly, both agglomerative and KNN algorithms clustered these countries into the same cluster. Thus, affinity clustering algorithm was unable to 100% discriminate between countries.

Additionally, the study analysed the relationship between Human Development Index's (HDI) of countries and their respective mortality rates in 2019. The HDI provides a single index measured on life expectancy, average years of schooling for adults, and gross national income per capita. Other studies have observed HDI can predict mortality rates for individual countries (Lee, Park, et al., 1997). An unpaired students t-test was applied to the clustering labels to observe whether the mortality rate reflects the HDI in 2019. The results have shown failure to reject the null hypothesis, indicating the distributions are the same ( $stat=-0.864$ ,  $p=0.388$ ).

The study applied several predictive models using interpolation techniques. As shown in Figure 3, four interpolation equations were applied to the data and their residuals were plotted for visualisation. From the graphs the polynomial equation  $y = 2.46x^2 - 4.84x + 2.38$  was the most appropriate because the residuals are distributed uniformly around  $x=0$  (Sperandei, 2014).

Furthermore, the study applied this predictive model to two further years; 2000 and 2017 to analyse the residuals to see if it remained to be a best fit. As shown in Figure 4 the equation conforms to 2017 however, is shifted far right in 2000. Additionally, the fitted response values vs the residuals of the line, clearly show the variance of the residuals increases with response variable magnitude. Thus, the model does not respect homoscedasticity. In turn, transformation may be required to improve the model quality. Future research aiming to explore this relationship, could incorporate the other parameters used to determine HDI to perform feature selection of these variables improving the quality of the model.

References:

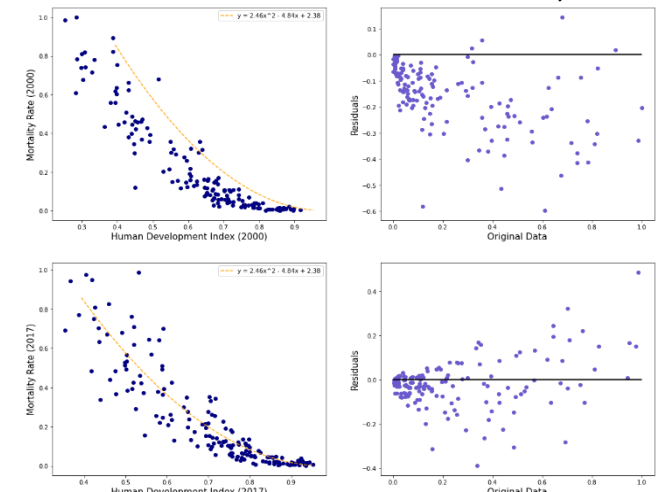


Figure 4 – Predictive models applied to A year 2000 and B year 2017 and their respective residuals.

Lee, K. S., Park, S. C., Khoshnood, B., Hsieh, H. L., & Mittendorf, R. (1997). Human development index as a predictor of infant and maternal mortality rates. *The Journal of pediatrics*, 131(3), 430–433.

Sperandei S. (2014). Understanding logistic regression analysis. *Biochemia medica*, 24(1), 12–18.