ABSTRACT OF THE THESIS

Time Series Analysis on U.S. Immigration Data
by
Elizabeth Kay Fabio
Master of Science in Big Data Analytics
San Diego State University, 2022

U.S. immigration data is analyzed using intervention event analysis and time series clustering. Most research is limited to immigration data in several major migration eras or only focus on immigrants with certain skills that were favored or disfavored by the U.S. This study conducts a comprehensive analysis of publicly available immigration data from 1820 to 2020. We examine impacts of political and economic events that have reshaped the U.S., compare their impacts on legal and undocumented immigration, and further investigate the significant migration patterns from sending countries. Upon understanding the trends and outliers in time series data, immigration flow can be better managed. With these goals in mind, U.S. immigration is assumed to follow a domestic politics and globalization approach. Domestic politics is best explained as a neutral arena for societal interest. Globalization is the international economic factors and social drivers that have a major influence on society.

Based on the intervention analysis, many events had a temporary negative impact. International economic events such as the Great Depression had a significant impact. Political events did have a significant effect depending on what the policy was directed at. For instance, some policies, including the Immigration Reform Act and Control Act of 1986 (IRCA), may have only affected border enforcement but not visa issuances and naturalization.

It is interesting to discover that the immigration-related policies, not only in the U.S. but also in sending countries, play a critical role in the flow of immigration. Hierarchical clustering algorithm was adopted to group these sending countries and to help better understand the dynamics of immigrants from different places. Both shape-based distance (SBD) and dynamic time warping (DTW) were used to measure the distance of time series data between countries because of their higher degree of accuracy when compared to other measures. As a result, two clusters were created for easier interpretation. After all the parameters were adjusted, the various clustering methods still resulted in similar prototypes. This study successfully provided in-depth analysis on political and economic events in time series data and used time series clustering to understand sending countries immigration history and behavior.