*Group Name*: Population Profits

*Group Members*: Matthew Gerloff, Michelle Chen, Natalie Ramdeo, Edward Reyes-Pinto

**Executive Summary**

**Key Objectives**

Since our group began this project, we had the goal of providing predictive analysis that would reveal relevant truths in the business sector. We felt we were in an excellent position to create these predictions with our U.S. Population dataset. At this point we faced our first major hurdle, which was deciding which of our variables could we realistically create predictions for, and would the subsequent analysis be as meaningful as we had hoped. After successfully creating a predictive decision tree model related to our Migrants variable, our main objective became providing answers toward how Migration will affect the future of U.S. business.

**Findings**

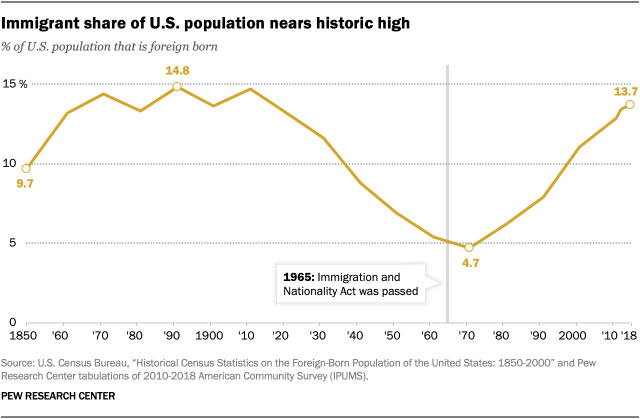
We found that in order to predict the future of Migration, we had to look toward the past. Our data revealed Migration levels had sustained periods of growth followed by steep declines in recent years. With additional research, it became clear this dramatic movement was caused in no small part by U.S. Legislation. Two pieces of legislation stood out to us the most. In 1965, when Migration levels were hovering at very low levels, The Immigration and Nationality Act was passed. This eliminated the National Origins quota system in place mainly due to large scale wars. Since that year, the number of Immigrants living in the U.S. has more than quadrupled, due to these relaxed regulations and continued population growth.When looking at our data, we noticed a sort of “peak’ in the Migration numbers between 1995 and 2000, and sure enough, we found more legislation at the root of this. The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 imposed stricter eligibility requirements and created barriers for immigrants to access public assistance programs. Through research and analysis of our data, we concluded that when the U.S. government isn’t happy with current Migration numbers, it has the power to enact change and that change seems to always have a massive impact.

**Visualizations**

Our visualizations were created in Tableau. We wanted to see trends and patterns that occurred throughout the years and so we had it as a constant x-axis variable.



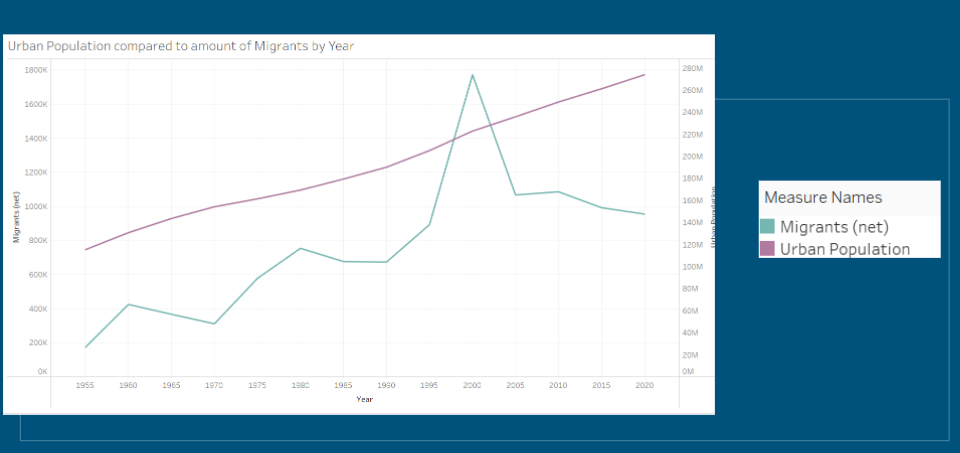
Our first visualization showcased the Migrant status in the US from 1955-2020. Through this visualization, we can see that the migrant status was consistently over the average after 1990 until 2020, where it had decreased.



Our second visualization was sourced from the Pew Research Center and illustrated how effective and impactful legislative policies were for migration in the US. This visual shows the impact of the Immigration and Nationality Act which has quadrupled the number of immigrants since 1965.



The third visualization is a clearer image of the first visualization. The steep decline in Migration numbers after the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 is clearly visible here.



Our final visualization was created in order to compare the trends of the US Urban population with the number of Migrants coming in each year. Highlighted here is the steady rise of Urban Population not being effected by high or low Migration years.

**Machine Learning**

In order to create our Machine Learning model, additional columns were added to our dataset. The Migrants column in itself was not eligible for a Decision Tree model, so we looked to determine if it was above or below an average that we could set. Since population numbers are known to always increase, we instead found the average by comparing Migrant numbers as a percentage of the population for a given year. Finally, in order to make the data readable for the Decision Tree predictor, we created the Migrant Status Encoded column that translated “High Migrants” to a 1 and “Low Migrants” to a 0. With the data adjusted in this way on Excel, it was uploaded to KNIME. From there, it was partitioned with a 70% training set with the remaining 30% used as testing data. The rest of our KNIME workflow was simply the Decision Tree Learner, the Decision Tree Predictor, and the Scorer to view results.

**Scores**

Our testing set only consisted of 6 fields, which is admittedly too small to be considered with full confidence. However, our Scoring results were that all rows were correctly classified, with a 100% accuracy rating.

**Conclusions**

With the knowledge of our smaller testing set in mind, we’ll prioritize a larger amount of rows in future datasets. It became clear that the more data that is present in terms of rows, the more meaning the dataset has. We were still pleased to complete a working model, and came away with some conclusions as we had hoped toward what Migration says about the future of business. These can be found in the Insights & Analysis slides toward the end of our final presentation. Most of all we learned effective ways to think critically about a dataset, and got valuable experience in applying what we learned in Tableau and KNIME to something we could create meaning for.