Population Migration Use Case

Team 58

Natalie Scavuzzo

Nino Cricco

Andrea Cuadros

Helen Haile

Rahel Mehreteab

Charles Pryor

Fiona Suliman

**Predicting Central American Migration**

Each year Central American countries, namely Honduras, Guatemala, Mexico, and El Salvador sees thousands of migrants fleeing the country. The migrants mainly attempt to immigrate to the United States however South America, the West Indies, and Europe are starting to see significant numbers of Central American Migrants. Factors that could influence migration are food insecurity, violent crime, politics, domestic violence, jobs, access to medical care, and poverty. This migration causes a massive immediate influx to countries that take them. Countries need to understand what exactly drives migration and how best to serve the immigrants as well as help the countries they are coming from retain their population.

**The Problem**

*What?*

How to help a country reverse its trend of migration.

*The approach?*

We will look at all of the factors that influence migration and see the influence on one’s decision to leave. As the numbers change, migrant numbers will change. We will see which features correlate with each other and assess how the change in the numbers raise or lower migration.

*Why does this problem matter?*

If we can assess which variables most influence population the world can respond by pouring adequate resources into these regions.

**Audience**

* World Health Organizations
* Local and Federal Governments
* Environmental Organizations
* Population Demographic Centers
* Sociological Research Institutions
* Economists

**Datasets**

* Preliminary datasets will be taken from the World Health Organization
* The U. S. State Department
* Pew Center for research and statistics

**Method**

We will clean all datasets then merge them and clean again. We will perform feature engineering on the X variables to see how each column relates to one another. Once the data is cleaned and put into a final pandas data frame we will check the assumptions.

Scadacity - homo and hetero

Multicollinearity - we will look at variance inflation factors and fix if numbers aren't within the metric.

We will assess if normalization, regularization, and standardization of the columns is needed.

We will then fit the model with a basic linear regression, then Polynomial Regression, Ridge, and Lasso and compare prediction accuracy and cross validate the model with one of the four suggested metrics preferable K-fold.

**Libraries & Packages**

* Pandas
* Numpy
* Math
* Seaborn
* Matplotlib
* Yellowbrick
* Scikit Learn
* Statsmodels
* Linear Regression
* Polynomial Regression
* Dash
* Flask
* Streamlit

**Interface**

A web application will be made through Dash, Flask, or Streamlit. It will give a brief overview of the problem, the ppt, visualizations bringing the data to life and an explanation of the interactive feature. This will allow someone who is trying to assess how much a population will change due to a variable to input different X values.

*Example:*

X1 - percent of population in poverty

X2 - number of incidents of domestic abuse

X3 - percent of population diagnosed as malnourished

X4 - number of jobs available

X5 - birth rate

X6 - crime rate

X7 - political approval rate

Y - migration amount

The application will have a section to fill in for all X values and once all entries are made, the site will show the predicted Y value(amount of people leaving).

**Timeline**

Week 5 - Find datasets, create Github

Week 6 - Clean datasets, start merging data sets

Week 7 - Start cleaning the amster data set, assess for problems then fix

Week 8 - Continued cleaning of data

Week 9 - Perform Exploratory Data Analysis(EDA) and Feature Engineering, start website

Week 10 - EDA and Feature Engineering continued, Website continued, ppt started, master annotated technical notebook starts to be assembled.

Week 11 - Model fitting & cross validation, Website, tech notebook cont’d, ppt.

Week 12 - Finalize website, ppt, tech notebook, clean github.

Week 13 - Present the project.