R08

Predictive Analytics for Population Growth of Refugees in Asia

Umar Ibn Ali (1328435) | Md. Shaikot Hossen (1233903)
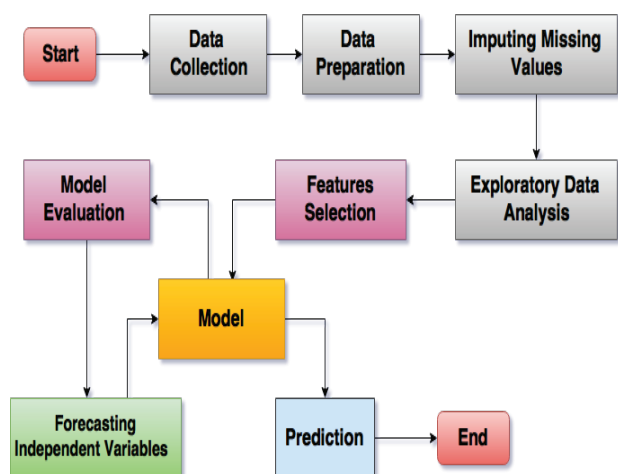
Supervised By
Madam Suriani Bt. Sulaiman

🔍 Problem Statement

The refugee population in Asia is on the rise. The countries that are receiving these refugees are not well equipped for the huge surplus in the number of refugees in the recent past. Thus, the refugees are unattended and the governments unprepared. Moreover, the sentiment of the citizens of these countries regarding refugees is not highlighted.
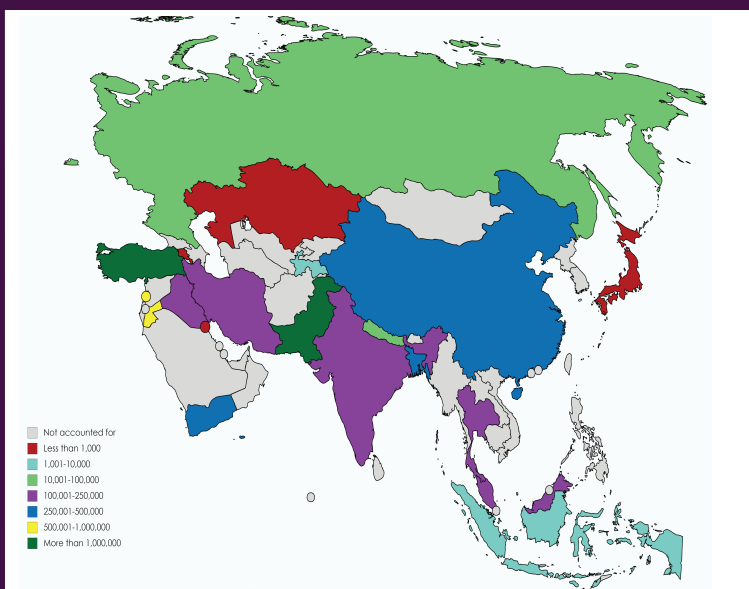
🔬 Project Objective

1) Predict the population of refugees in hosting Asian countries up to year 2022.

2) Analyze Tweets to evaluate the acceptability of Refugees in these Asian countries.

3) Visualization of model prediction and sentiment analysis results

Methodology Diagram

Refugees in Asia 2022

Not accounted for
Less than 1,000
1,001-10,000
10,001-100,000
100,001-250,000
250,001-500,000
500,001-1,000,000
More than 1,000,000

⊞ Datasets

Countries included: 20
Total records: 560

Methodology for inclusion:
1991-2016 refugee population mean over 2000
2007-2016 refugee population mean over 2000

➡ **Country Dataset:** 26 rows x 6 columns
The yearly data from these datasets are used as columns to get the dataset for each country for years 1991 to 2016.
  Datasets
  ~UNHCR Yearly Refugee Population Data from UNHCR
  ~GDP Per Capita from IMF
  ~Population Density from IMF
  ~Global Peace Index (GPI) from IEP
  ~No of Refugee Countries from UNHCR

➡ **Sentiment Analysis Dataset**
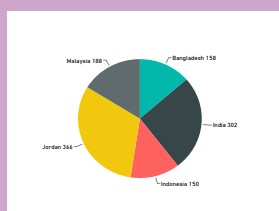 ~ Tweets about refugees from Tweetdeck for years 2014 to 2017

Algorithm: Multiple Linear Regression (MLR)
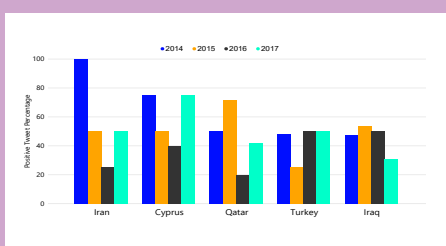**Language:** Python
Sentiment Analysis Tool: TextBlob

Modeling Tool: Scikit-learn
Model Evaluation: R squared, RMSE, MAE
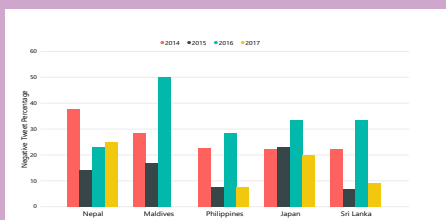**Environment:** Jupyter Notebook (IPython)

📈 Twitter Top 5

[Figure X1. Top 5 countries with the most number of tweets about refugees between 2014-2017]

[Figure X2: The yearly percentages for the top 5 countries with the highest rate of positive tweets based on the mean value.]

[Figure X3: The yearly percentages for the top 5 countries with the highest rate of negative tweets based on the mean value.]

📊 Prediction Results

Country	Actual Values	Predicted Values
Malaysia	92,287	91,934
China	317,260	308,638
Yemen	269,796	243,703
Tajikistan	2,731	2,297
Thailand	106,471	114,070

[Table X1: For year 2016, actual values were checked with predicted values to test the prediction results]
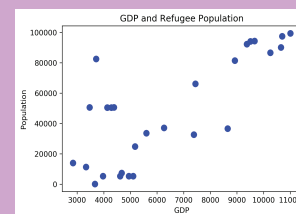
Country	2016	2018	2020	2022
Turkey	1,688,008	1,874,707	2,061,407	2,248,107
Pakistan	1,567,471	1,581,124	1,594,778	1,608,432
Lebanon	703,453	784,967	866,481	947,995
Jordan	603,137	664,238	725,339	786,440
Bangladesh	390,099	424,284	458,469	492,654

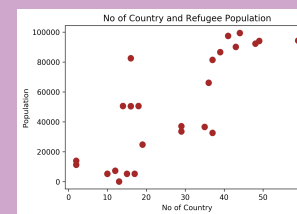[Table X2: Top 5 countries with the highest predicted population of refugees in 2022.]

Country	2016	2022	Percentage
Lebanon	703,453	947,995	34.76
Turkey	1,688,008	2,248,107	33.18
Jordan	603,137	786,440	30.39

[Table X3: Top 3 countries with the highest percentage increase in refugee population from 2016 to 2022.]
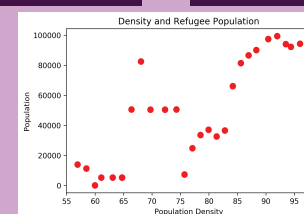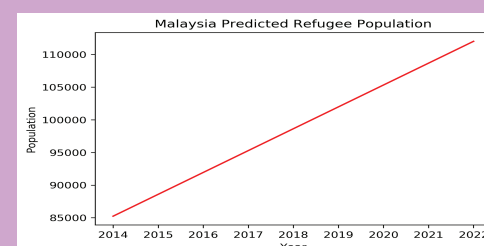
🥧 Results of Malaysia

[Figure X4:
Intercept: -15987 ; Coefficient: 10 ;
R-squared value: 0.60]

[Figure X5:
Intercept: 1005 ; Coefficient: 1811 ;
R-squared value: 0.65]

[Figure X6:
Intercept: -15987 ; Coefficient: 10 ;
R-squared value: 0.60]

[Figure X7: The variables GDP, No of Country, Density were forecasted and those values were used to build Malaysia's Model.
RMSE Score: 18867, MAE Score: 17423, R-squared value: 0.67

Predicted 2017 by the model: 95,281
Latest update July 26, 2017: 94,361 || 2022 prediction: 112,019]

💬 Discussion

The model predictions showed good results overall. There is a noticeable correlation of refugee population with percentage of positive tweets and number of tweets. The results of the prediction could be improved by monthly refugee population data from UNHCR as monthly values could detect the fluctuation in the refugee population better than yearly values. The increase of refugee population could be better predicted by adding the indicators used to make Global Peace Index dataset, such as "Number of refugees and displaced persons as percentage of population", "Military expenditure as a percentage of GDP" and "Political terror" among others. This data is not available to the public.

ICTSS   FYP SHOWCASE   FYP BCS