Providing targeted responsible foreign aid

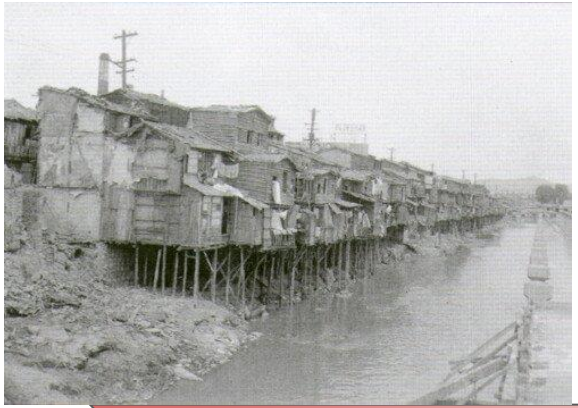
Group 5: Ron, Kushal, Jimin, Seunggil

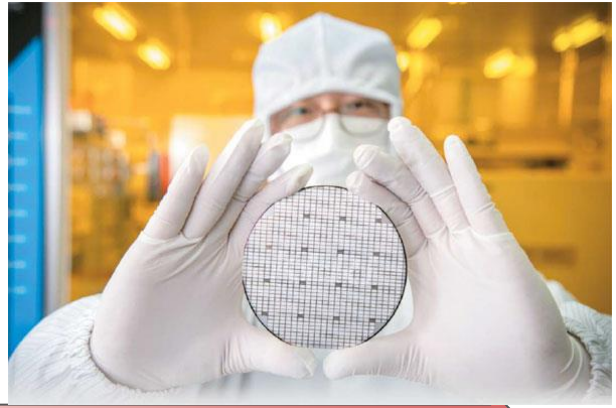
Contents

- **Project Background**
  - Talk about problems, current status, background
- **Data and EDA**
  - Data
    - Data collection, what type of data, data timeframe, countries focused
  - EDA
    - Foreign aid trends, GDP trends (line charts)
- **Methodology**
  - Model Building
- **Final results**
  - Final Charts/dashboard/word clouds/topic modeling
- **Conclusion/Future Work**

Why Foreign Aid?

Example of South Korea





→ Now, Korea became **Strong Alliance** of USA

→ Korea is Contributing to the **advancement of the world**

Skepticism about Foreign Aid

In 2020, US government had spent $51.05 billion US dollars in foreign aid

But, many mixed perspectives about effects of foreign aid exist

Many countries still remain as Least Developed Countries(LDC):

→ hasn't reduced poverty rate

<Question>

Provide Targeted, Responsible Foreign Aid !

→ Aid that has actual outcome

<Solution>

Focusing on TARGETED, RESPONSIBLE

Targeted: For which country does aid be helpful?

→ for which country aid should be given?

Responsible: The aid should helpful for the nations

→ In which way the recipient nation get beneficial results?

Various types of Foreign Aid

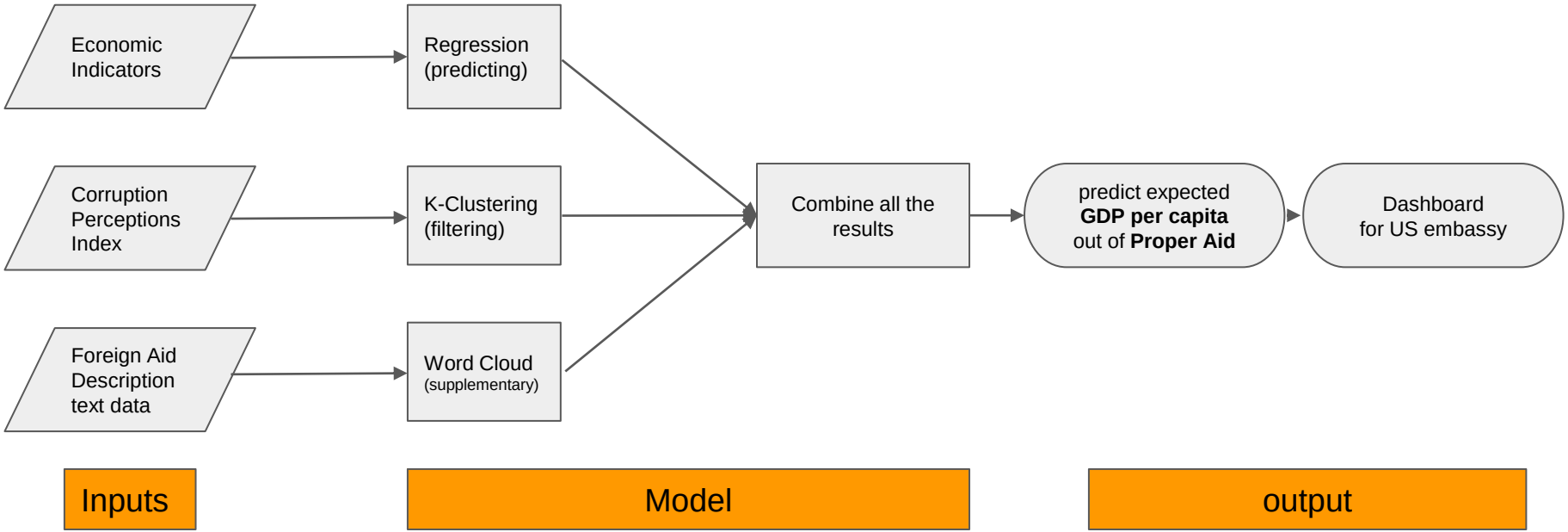
there are various kinds of foreign aid exist

- Disbursements(fund)
- obligations
- Bilateral Aid
- Military Aid
- Multilateral Aid
- Humanitarian Assistance
- …

Our Solution focuses on **Economic type of Aid**

→which might influences **GDP per capita**

Overview of our solution



datasets - source

1. World Bank Data
2. OECD data
3. Corruption Perception index
4. Text Analysis

Our Total Dataset :

World Development Indicators across countries

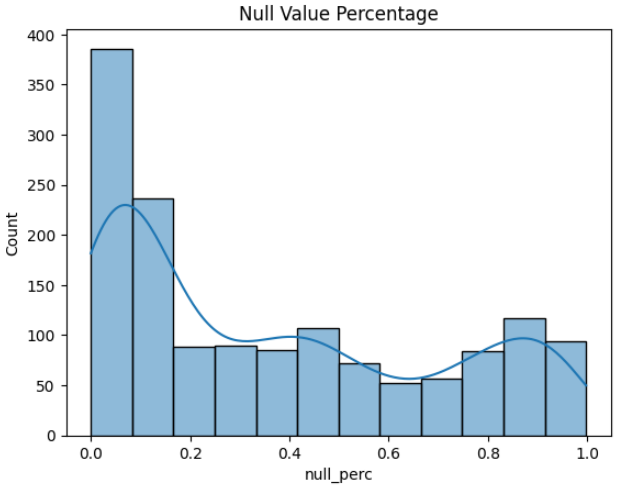
Countries & Indicators

Total Countries(34)

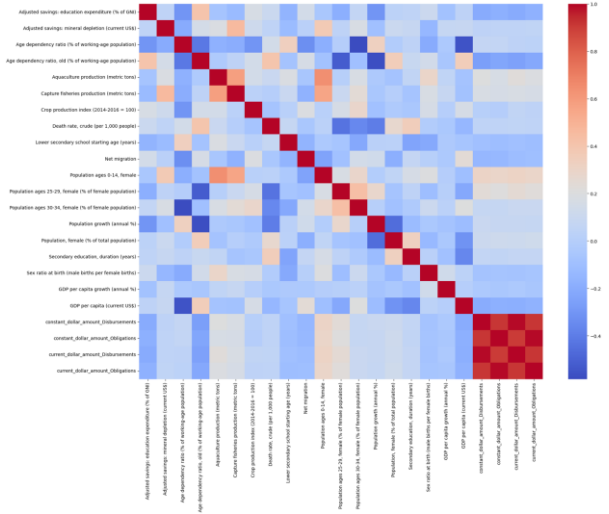
Latin America(24) + Southeast Asia(10)

Indicators (+ foreign aid)

1478 -> 21

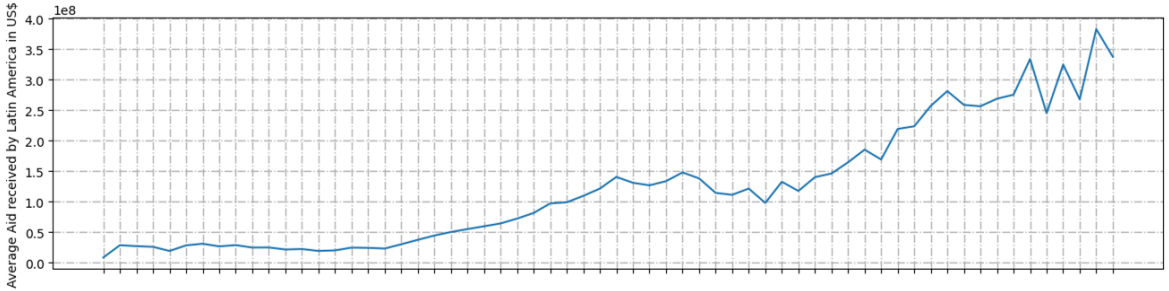


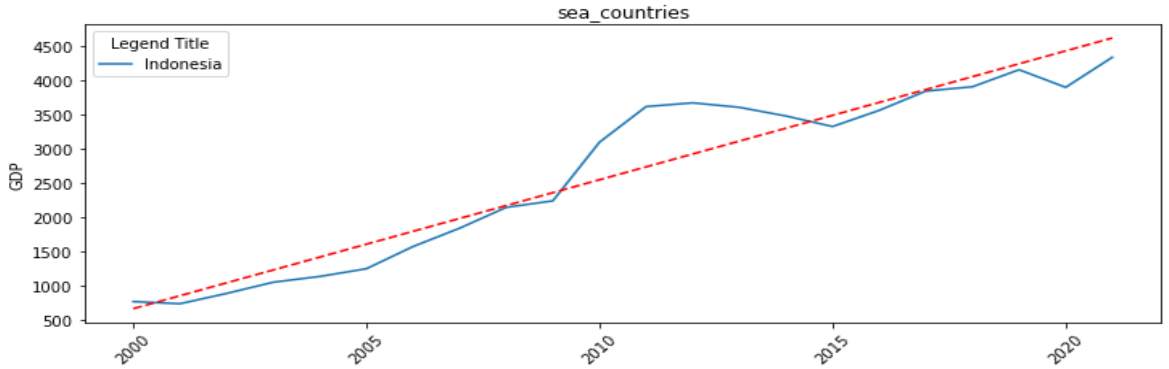
Percentage of null values of feature across all countries



Correlation of selected features

Trends & Correlation of Foreign Aid + GDP



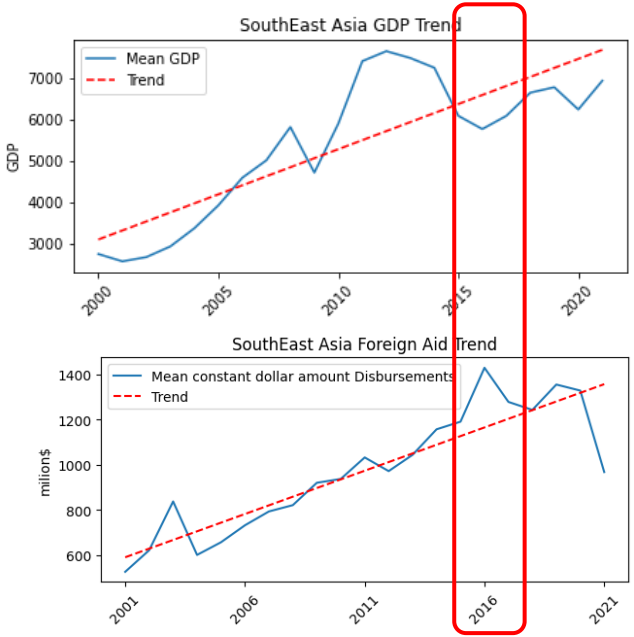


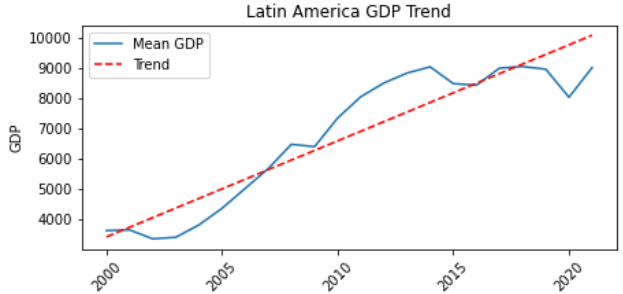
In Latin America and South East Asia countries,

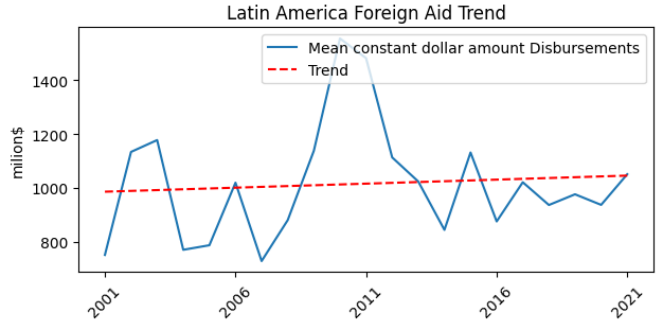
As average aid increases, GDP is also increased.

- There is Correlation between GDP and Foreign Aid

Current status of Foreign Aid and GDP







currently only historical data is considered for budgeting next year

Possible solution : predictive modeling for budgeting next year

Modelling

Regression model for predicting:

Target outcome

- GDP per capita (current US $)

Explanatory variables

- Adjusted savings education expenditure of GNI
- Adjusted savings mineral depletion current US
- Age dependency ratio of working age population
- Age dependency ratio old of working age population
- Aquaculture production metric tons
- Capture fisheries production metric tons
- Death rate crude per 1000 people
- Lower secondary school starting age years
- Net migration
- Population ages 25-29 female offe male population
- Population female of total population
- Secondary education duration years
- Sex ratio at birth male births per female births
- GDP per capita growth annual
- constant_dollar_amount_Obligations
- current_dollar_amount_Obligations

Baseline Model

- To build hypothesis and assumptions about underlying dataset
- Linear regression model
- Results:
  - Most features are significant
  - Reported error (adj R2) = 0.70
  - RMSE = $ 4854.85
  - F-statistics p-value < 2.2e-16

```
Residuals:
     Min       1Q   Median       3Q      Max
-10808.0  -3145.1   -159.7   2582.2  29119.7

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              2.931e+05  2.415e+04  12.138  < 2e-16 ***
AdjustedsavingseducationexpenditureofGNI -1.426e+03  1.159e+02 -12.306  < 2e-16 ***
AdjustedsavingsmineraldepletioncurrentUS  4.243e-07  8.335e-08   5.091 4.80e-07 ***
Agedependencyratioofworkingagepopulation -4.863e+02  2.594e+01 -18.750  < 2e-16 ***
Agedependencyratiooldofworkingagepopulation 8.925e+02  8.609e+01  10.368  < 2e-16 ***
Aquacultureproductionmetrictons          -2.594e-04  1.446e-04  -1.794 0.073298 .
Capturefisheriesproductionmetrictons     -1.648e-03  1.752e-04  -9.403  < 2e-16 ***
Deathratecrudeper1000people              -3.538e+02  1.762e+02  -2.008 0.045100 *
Lowersecondaryschoolstartingageyears      2.053e+03  5.447e+02   3.769 0.000181 ***
Netmigration                             -8.995e-03  2.830e-03  -3.179 0.001557 **
Populationages2529femaleoffemalepopulation -1.339e+03  3.940e+02  -3.398 0.000724 ***
Populationfemaleoftotalpopulation        -4.526e+03  3.017e+02 -14.999  < 2e-16 ***
Secondaryeducationdurationyears          -2.557e+03  3.548e+02  -7.208 1.75e-12 ***
Sexratioatbirthmalebirthsperfemalebirths -2.947e+04  1.454e+04  -2.027 0.043066 *
GDPpercapitagrowthannual                 -8.124e+01  4.595e+01  -1.768 0.077565 .
constant_dollar_amount_Obligations        2.097e-05  1.031e-05   2.033 0.042494 *
current_dollar_amount_Obligations        -2.299e-05  1.201e-05  -1.914 0.056052 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4925 on 588 degrees of freedom
Multiple R-squared:  0.7095,    Adjusted R-squared:  0.7016
F-statistic: 89.76 on 16 and 588 DF,  p-value: < 2.2e-16
```

Conclusion:

- GDP can be modelled as a function of multiple World Indicators, including Foreign Aid
- Regression results are promising yet due to high error, needs some non-linear approach

Advanced Models

KNN

- Non-parametric memory based model
- East to understand and implement

RMSE = $ 8464.376 ▲

Conclusion
- Error is high
- Prediction will be $8464.37 off from actual

Decision Tree

- Non-linear model
- Easy to understand and implement
- Better explainability

RMSE = $ 2404.05 ▼

Conclusion
- Error is high
- Prediction off by $2404.05 from actual

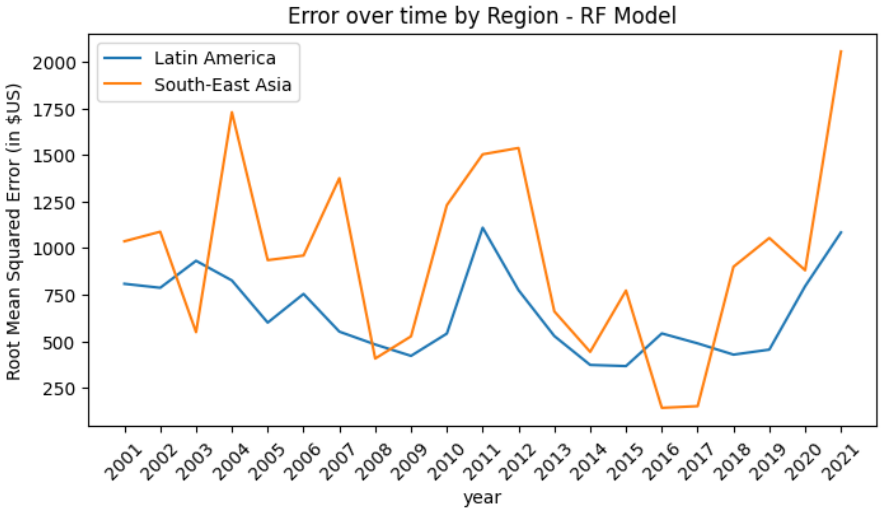
Random Forest

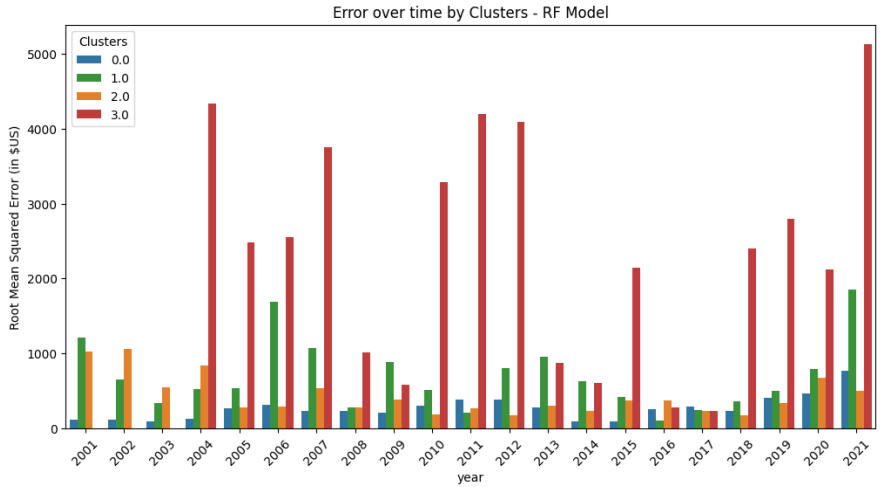
- Combination of multiple decision trees
- Captures non-linear relationship within features

RMSE = $ 1120.86 ▼

Conclusion
- Low error
- Only off by $1120.86 from actual

Model drawbacks





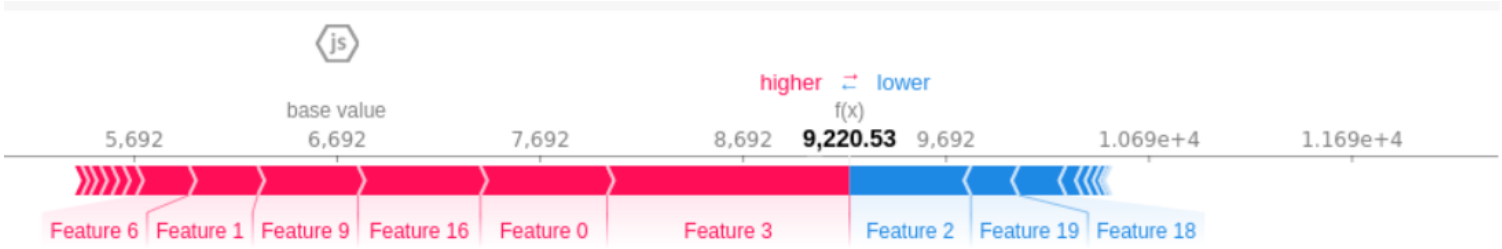

```
Cluster 0.0:  ['Cambodia' 'Guatemala' 'Haiti' 'Nicaragua']
Cluster 1.0:  ['Costa Rica' 'Cuba' 'Malaysia']
Cluster 2.0:  ['Colombia' 'Ecuador' 'El Salvador' 'Indonesia' 'Mexico' 'Philippines'
 'Thailand' 'Vietnam']
Cluster 3.0:  ['Singapore']
```

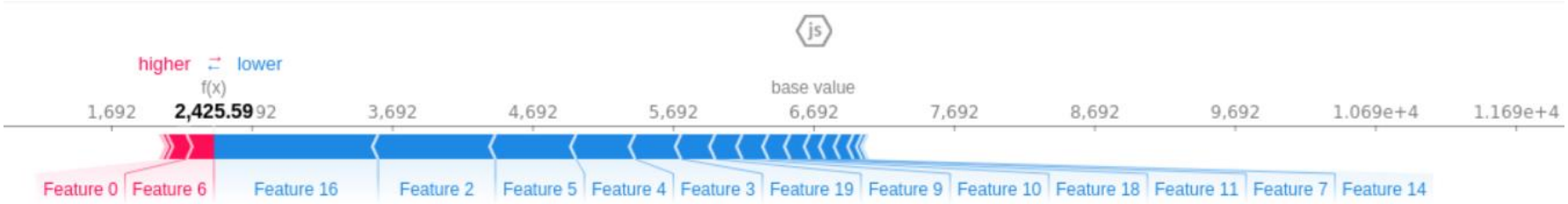
- Not performing well on recent data, error peaks for latest year metrics

- Cluster 2 performing very well, will be reliable for providing targeted aid
- Cluster 3 (Singapore) has higher RMSE over time

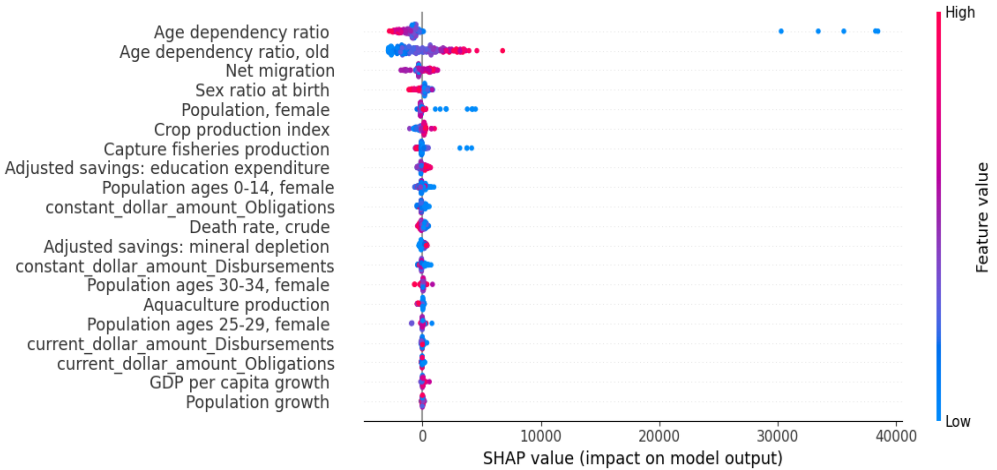
Model Interpretability

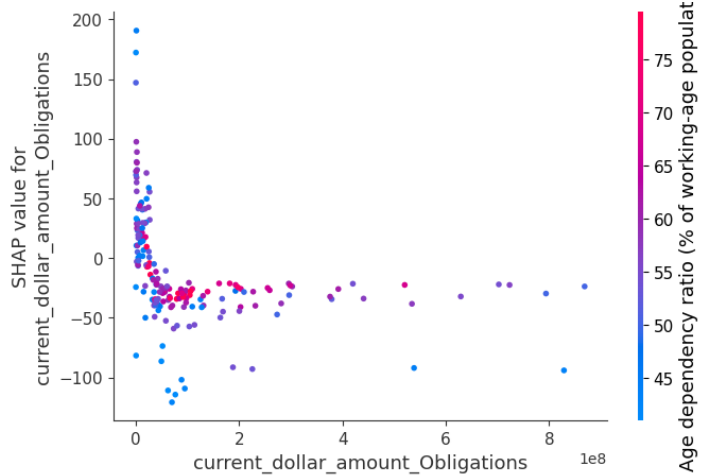
- Explaining model predictions using SHAP (**SHapley Additive exPlanations**) Values
- Two instances from dataset and how each feature is drawing the prediction towards or away from the expected value ($ 6692)





Model Interpretability - contd.

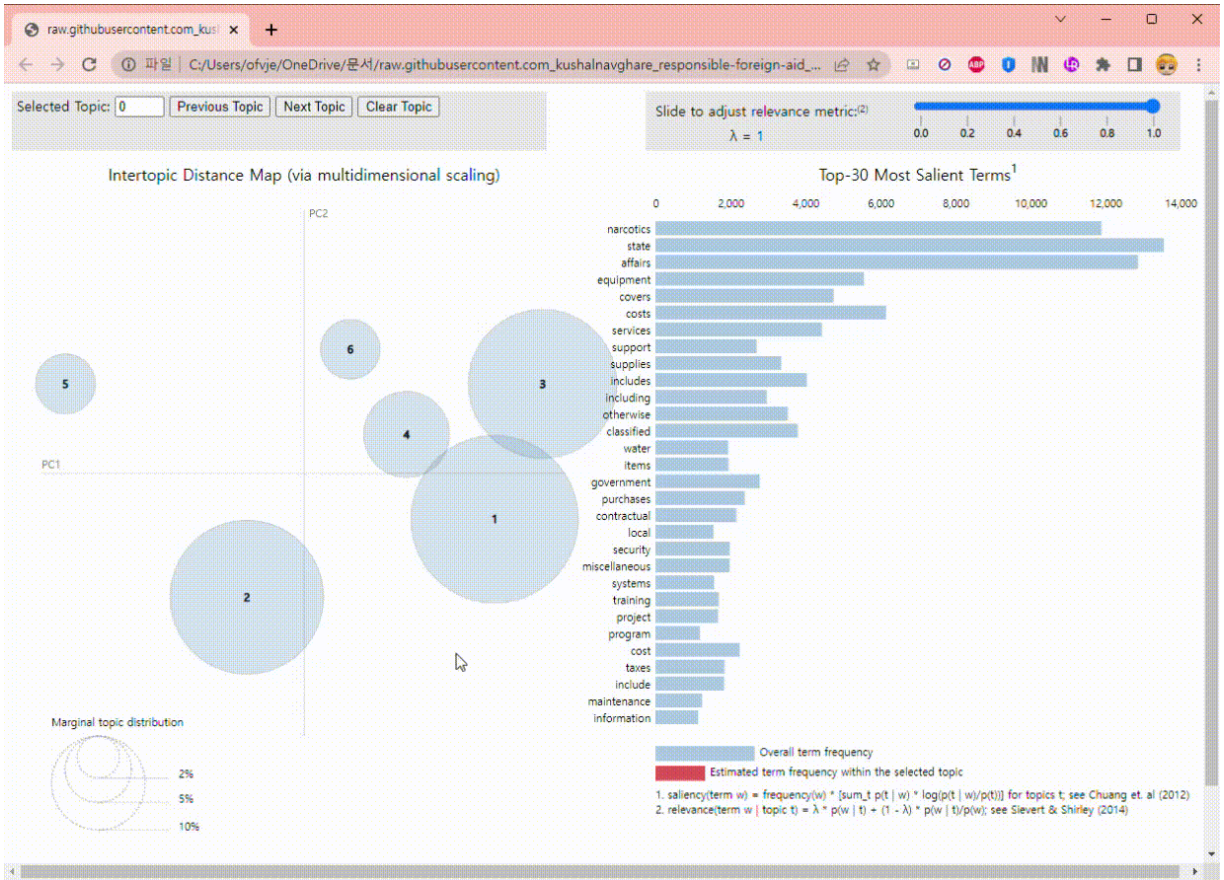




- Overall high Age dependency ratio of working age population will draw model predictions away from expected value ($6692.27) - **GDP will increase with high ratio of working population overall**
- Increase in Crop production index will have higher impact on model prediction - **Agriculture production will improve the economic condition of countries**

Topic modeling:

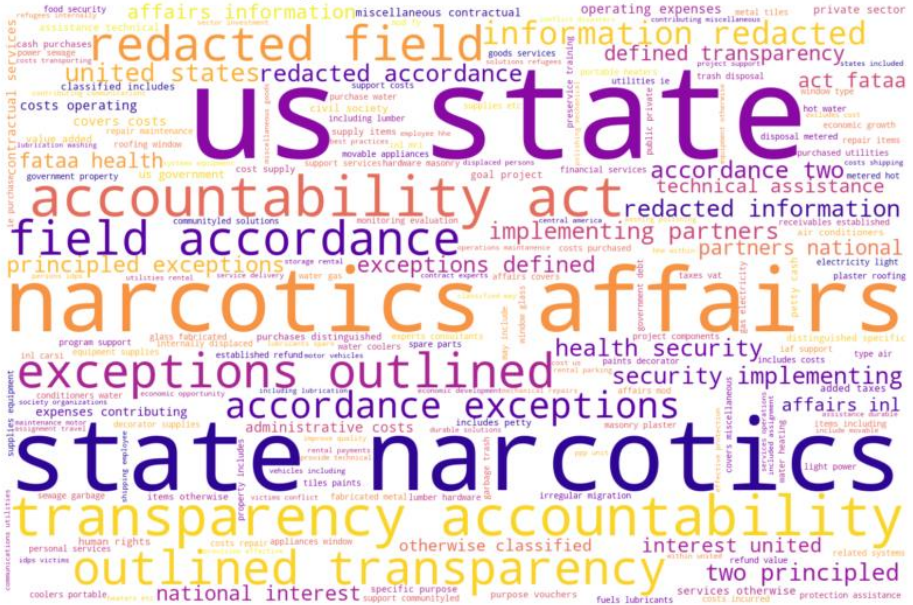
Clustering Activities of countries based on similarities



Final Results: SEA & LA

Bi-grams from South East Asia Activity

Bi-grams from Latin America Activity



Final Results : Dashboard

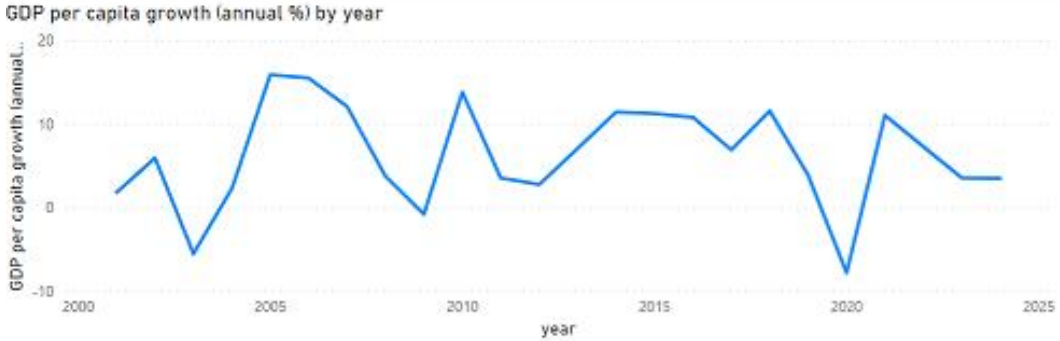
year

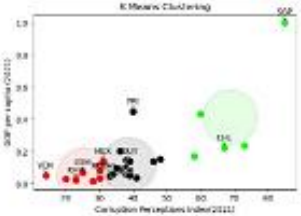
All

country_name

Dominican Republic

year	country_name	constant_dollar_amount_Obligations	constant_dollar_amount_Disbursements	current_dollar_amount_Disbursements	current_dollar_amount_Obligations	GDP per capita growth (annual %)	actual	prec
2021	Dominican Republic	51M	71M	72M	52M	11.08	8,476.75	7,9
2023	Dominican Republic	69M	64M	56M	60M	3.61	5,687.63	5,7
2024	Dominican Republic	69M	64M	56M	60M	3.57	5,739.84	5,7





Future work.

- Domain Understanding
  - Involvement of US Embassy for domain knowledge to better understand the process and identify areas where this solution can fit
  - Come up with a survey monkey questionnaire where the different users at the U.S embassy can review any model changes or suggestions.
  - Continuation of topic modeling:
    - Creating policies for different countries.
- Process
  - Improvements to the dashboard according to US Embassy requirements
  - Use best analytical practices to avoid discrepancies within data and modeling part