Predicting Asylum Application Outcomes in the European Union

Alexander Portland
Brown University
GitHub[1]

1. Introduction

In September 2023, EU+ countries received 108,000 asylum applications, the highest number since the refugee crisis of 2015-16[2]. Thus, it's important to understand what factors of an asylum application can affect one's odds of being accepted into a better life. Beyond that, any predictive model would be useful for examining systemic biases in national asylum adjudication methods, and allow for reflection on the lack of international convention on such processes[3]. The target variable for this project is the number of accepted asylum applications of people of a certain demographic applying from one country to another in a given quarter.

1.1 Previous Works

In work very similar to that conducted here, Carammia et al.[4] does an impressive job in predicting asylum applications, achieving a mean absolute error of just 81 on monthly predictions with a combination of data collection on causal factors and historical application records. This project focuses solely on the migr_asydcfstq[5] dataset, and thus will likely lose some predictive power in comparison by not considering external factors.

1.2 The Dataset

The dataset used in this project is the migr_asydcfstq[5] dataset containing "first instance decisions on applications by citizenship, age and sex" divided up by yearly quarters from 2008 up until the present. (Partial data from Q3 2023 was not considered in this study.)

Feature	**Description[6]**
Age	The age of the applicant
Sex	The sex of the applicant (Male, Female, or Unknown)
Citizenship	The applicant's country of citizenship
Geo (country of asylum)	The EU+ country where the applicant filed for asylum
Time Period	The yearly quarter of the application's filing
Decision	The applicant's final decision outcome

Table 1. Dataset features and descriptions.

This dataset is very large (with over 6 million rows)[5], and not Independent and Identically Distributed (IID) due to the timely nature of its data.

1.3 Feature Engineering

Three different kinds of positive outcomes are represented in the dataset. The "GENCONV" type was used to represent those granted asylum based on the 1951 Geneva Convention[7, 8], the "HUMSTAT" type represents those granted a humanitarian status[9], and "SUB_PROT" represents the subsidiary protections that some countries grant to those who don't qualify as official refugees, but would face serious harm if returned to their home country[10].

The rules for Geneva Convention status and subsidiary protection are controlled by the United Nations and European Union respectively[6, 8]. However, with no centralised rules on when humanitarian status should be granted, there are great discrepancies between how EU+ countries use the status.

Thus, I decided to merge these somewhat arbitrarily defined types into a single category to allow the models to focus on which features resulted in any positive decision. I similarly created a new feature for the total number of applications.

Engineered Feature	**Description**
Total Positive Decisions (*"total_pos"*)	All three of the positive outcomes ("GENCONV", "HUMSTAT" and "SUB_PROT") were merged into a single positive decision feature
Total Applications ("*total_apps"*)	The sum of all asylum applications within the quarter, regardless of outcome

Table 2. Engineered features and descriptions.

2. Exploratory Data Analysis

As it's important to understand the general migratory context that this data is a part of, I found it useful to look at how asylum application outcomes have changed over time. Figure 1 shows how application results have changed over time. We can clearly see the peak of the 2015-2016 refugee crisis, and how applications have been steadily increasing in recent years.

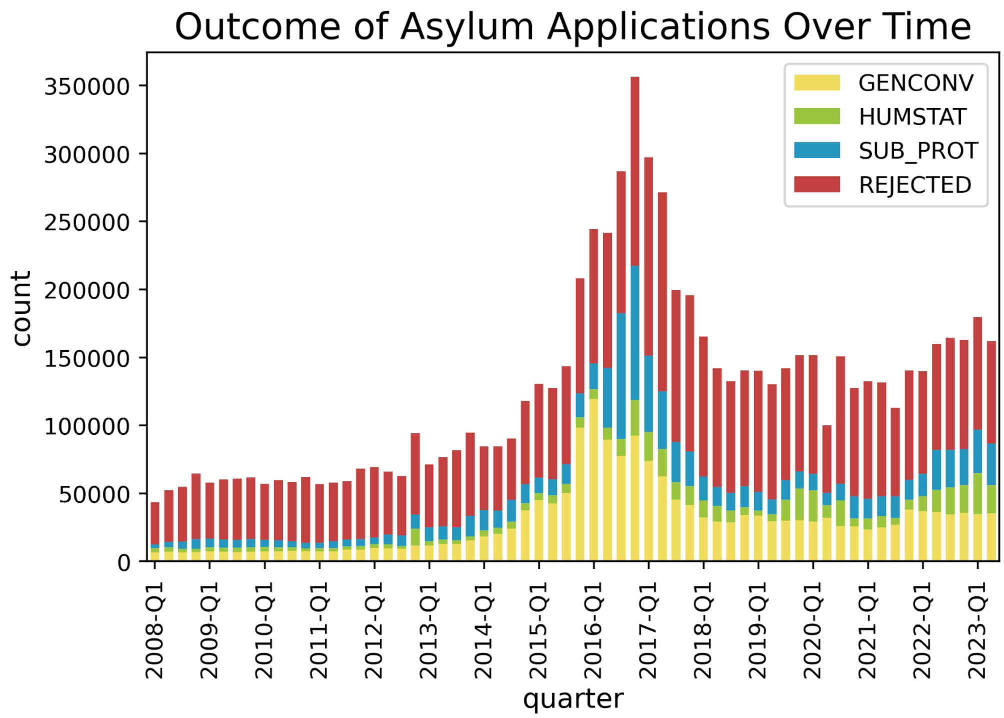

Fig 1. A decision-specific stacked bar chart showing the count of each asylum outcome by quarter since 2008.

Additionally, it's important to break down application outcomes based on demographics. The total outcome counts shown in Figure 2 highlight that the vast majority of those applying for asylum are between the ages of 18 and 34. By normalizing these counts with respect to the number of people in each age group, we can calculate acceptance rates (shown in Figure 3). These rates call attention to the fact that those aged 18-64 are much more likely to be denied asylum then their younger or older counterparts.

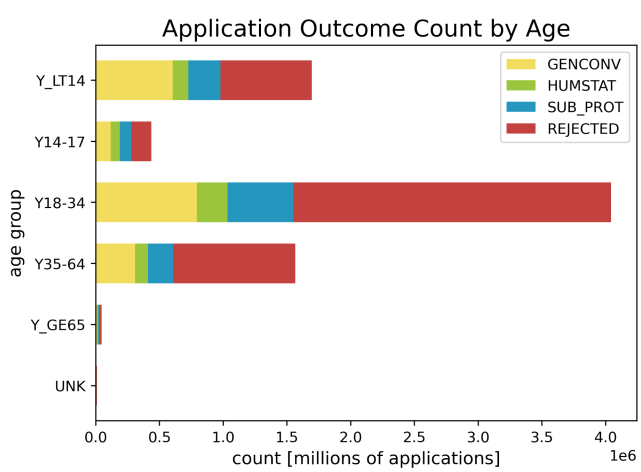

Fig 2. A stacked bar chart showing application outcome counts by age.

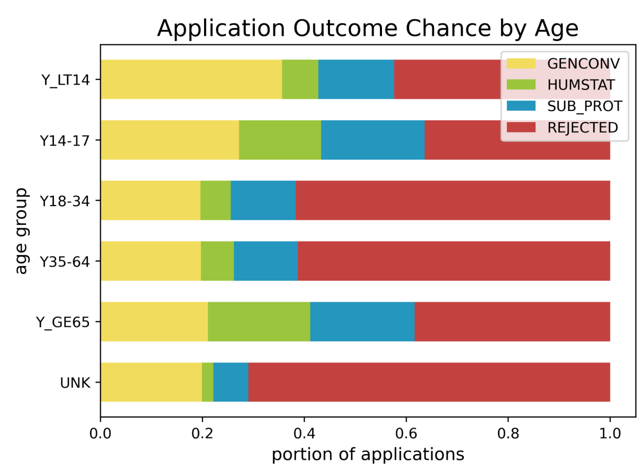

Fig 3. A stacked bar chart showing the relative portion of outcomes by age.

In addition to age demographics, and general global migratory trends, country of citizenship (see Figure 4) and country of asylum (see Figure 5) also appear to play an important role in the number of accepted applicants each quarter.

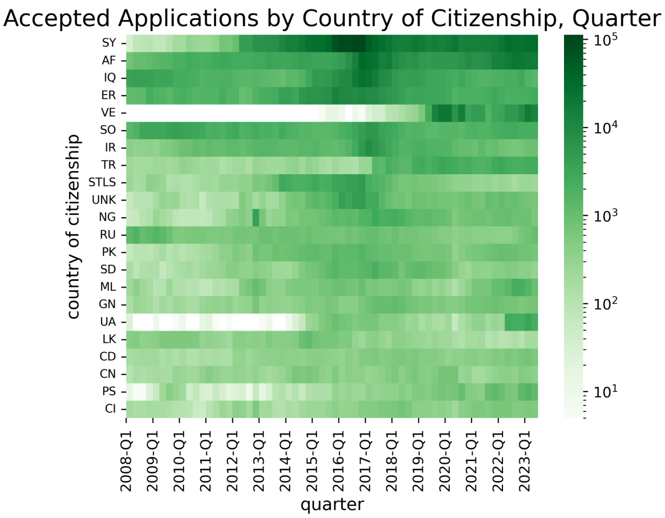

Fig 4. A heat map plotting the number of accepted applicants by their country of citizenship over time.

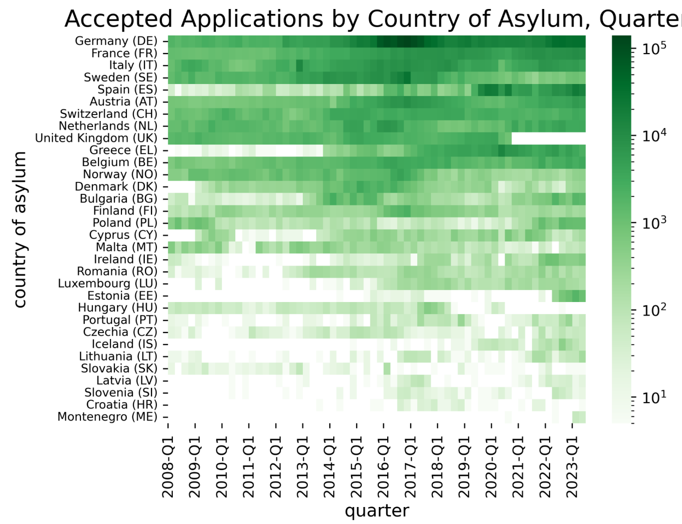

Fig 5. A heat map plotting the number of accepted applicants by their country of asylum over time.

3. Methodology

3.1 Splitting Strategy

Since our data is a time-series, the structure of sklearn's *TimeSeriesSplit* function was used as a model, ensuring we loop through different folds of the time-series, and leave a portion of the most recent data points aside as a test set. This ensures that there is no data leakage, and any models we develop can be accurately tested on data they haven't seen before.

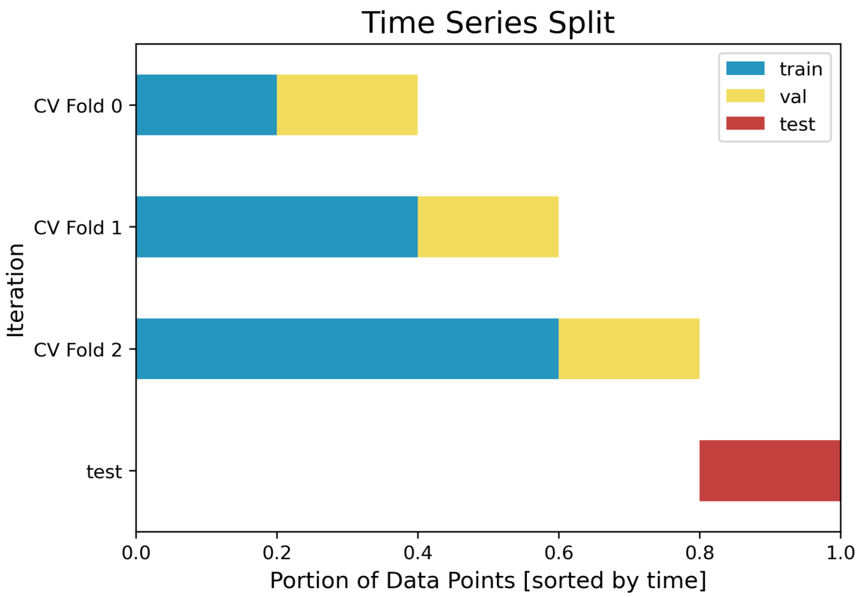

Fig 6. The *TimeSeriesSplit* strategy used in my project, with data points being sorted by time.

3.2 Preprocessing

All categorical features (*country of citizenship*, *country of asylum* and *sex*) were preprocessed using sklearn's *OneHotEncoder*, which converts each of the potential values for those features into a new binary feature. Both *age*, and *time_period*, were ordinal features and were thus preprocessed using the *OrdinalEncoder*, to preserve the order of their categories. The remaining features were all numeric with no clear upper bound, so sklearn's *StandardScaler* was used to transform their mean to 0 and standard deviation to 1.

3.3 Autoregressive Features

To better capture the timely and potentially cyclical nature of asylum applications, autoregressive features were introduced for both *total applications* and *total positive outcomes*. Based on initial testing, a lag of 4 quarters had the most predictive power, so this is what I used throughout the process. With more time and resources, it would have been better to treat the feature lag as its own hyperparameter, but with my large dataset, this was unfortunately not feasible for the time-frame of this project.

3.4 Hyperparameter Tuning and Cross-Validation

The goal of hyperparameter tuning is to determine which parameters allow for training to produce the most effective model on a given dataset. For each parameter configuration, we loop through the folds described in 3.1, training a model on the *train* set and evaluating its performance on the *validation* set. After determining which parameters are most effective on the three folds, we can use those parameters to refit a model on the entire *training* set, and test its objective performance on the final *test* set to compare it with other models.

For each model, hyperparameters were tuned on an appropriate scale, and if any optimal parameters were on the bounds of initially tested values, models were retrained to explore the whole parameter space. RMSE was used as the evaluation metric throughout this process to maintain the units of the target variable.

	L2 Linear Regression	**L2 Linear Regression (w/ polynomial features)**	**Linear Support Vector Regression (SVR)**	**Random Forest Regressor**	**XGBoost**
Random States Tested	just one (deterministic models)			5 random states	
Parameters Tuned	*alpha* - coefficient of regularization (log scale)	*alpha* - coefficient of regularization (log scale) *degree* - degree of polynomial features (lin scale)	*C* - regularization parameter (log scale)	*max_depth* - maximum depth of the tree (log scale) *max_features* - the portion of features to use (lin scale)	*reg_alpha* - l1 regularization term (log scl) *reg_lambda* - l2 regularization term (log scl)
Optimal Parameters	*alpha* = 2848	*alpha* = 0.01; *degree* = 2	*C* = 1000	*max_depth* = 10; *max_features* = 1.0	reg_alpha = 1.0; reg_lambda = 0.1

Table 3. Hyperparameters tuned for each model

4. Results

4.1 Baselines

To contextualize model performance, 3 baselines were calculated. Vitally, a baseline model that always predicts that the same number of applications will be accepted as last year can achieve a RMSE value of 10.344 on the test set.

4.2 Model Performance

Each model was trained according to the pipeline described in 3, with multiple random states being tested for the non-deterministic models (see Figure 7). The best models of each type all outperformed baseline, with Ridge linear regression with polynomial features being the best by a significant margin. Figure 8 shows how our finalized models perform on a slice of the test set for a single demographic.

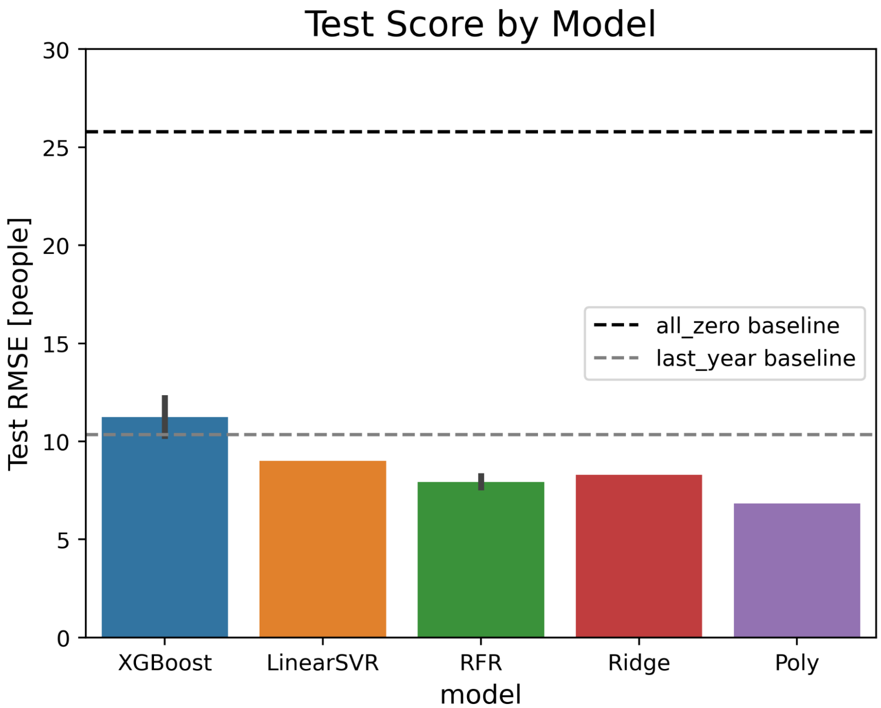

Fig 7. A bar chart showing test performance by model compared to baselines. Multiple XGBoost and RFR models were trained to test various random states.

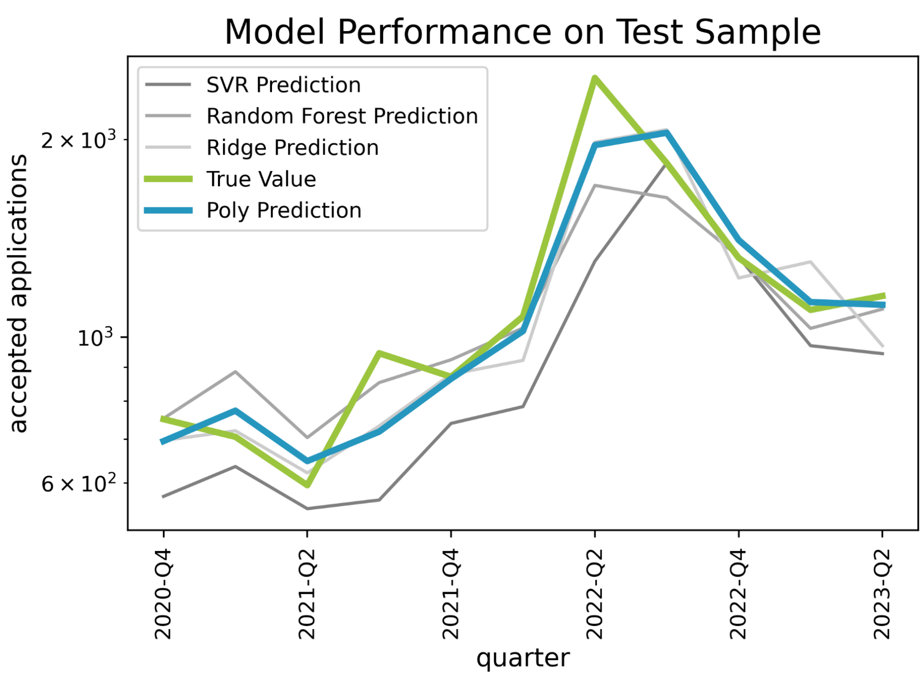

Fig 8. A line chart plotting model predictions on a subset of the test set (female applicants aged 18-34 applying to Germany with Syrian citizenship)

4.3 Model Interpretation - Global Feature Importance

In order to begin interpreting what makes these models effective, various metrics will be calculated to measure each feature's importance. Most simply, we can examine the weights of our best performing linear model (Ridge regression with polynomial features) to see which features it relies on most for its predictions. As can be seen in Figure 9, the lagged features are very important here, likely allowing the model to get a good understanding of the historical trends in migration that have led up until a certain point. From there, it appears that the model modifies its initial prediction based on the demographics of those applying (with certain citizenships and countries of asylum positively or negatively affecting the amount accepted).

The ease with which these modifications can be understood in a linear model is vital, as it allows us to examine trends during the asylum adjudication process that may be disproportionately affecting those of certain demographic groups. For example, why is an Albanian citizenship (*one_hot_citizen_AL*) so significantly detrimental to the number of refugees accepted in a quarter? These weights may highlight areas in which more research could be done to examine bias in adjudication.

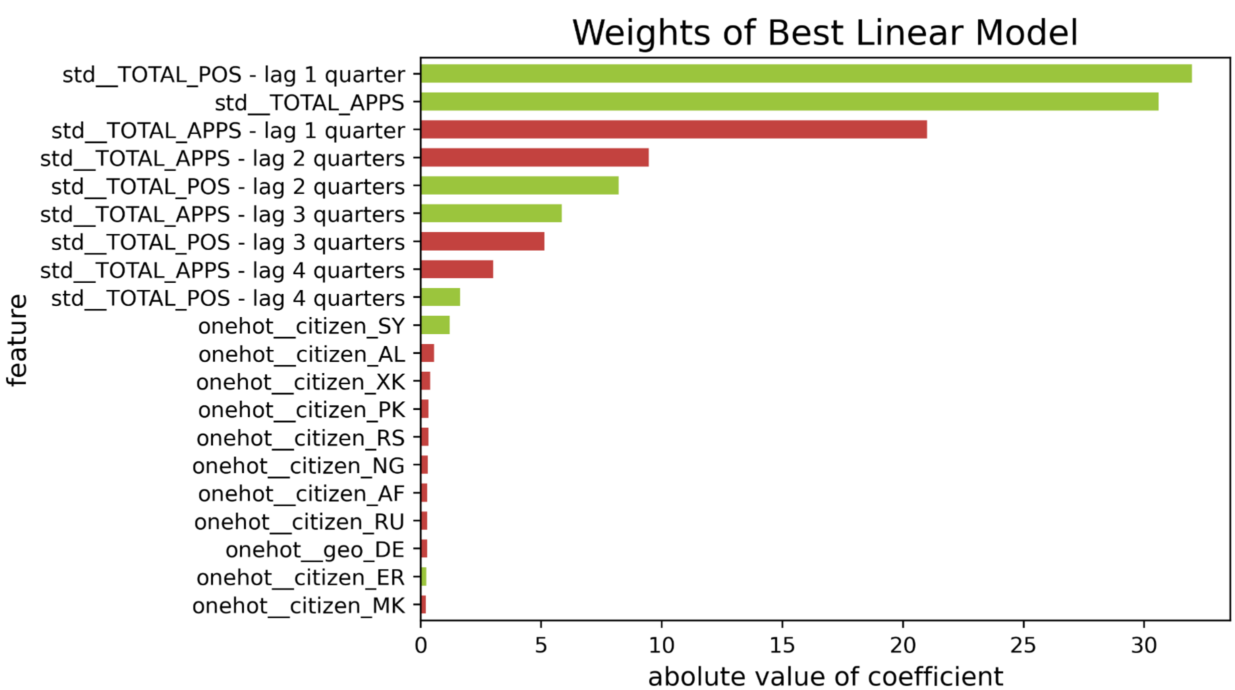

Fig 9. A bar chart showing the abs coefficients for each feature within the best linear regression model. Green indicates a positive coefficient while red represents negative.

This is corroborated by the metrics of the best performing XGBoost model (see Figures 10 & 11), in which the autocorrelated features similarly contribute heavily to model performance. Additionally, it seems that the XGBoost model pays more consideration to features such as *sex* and *time_period*, which rank highly in XGB gain despite having no significant impact on the best linear models.

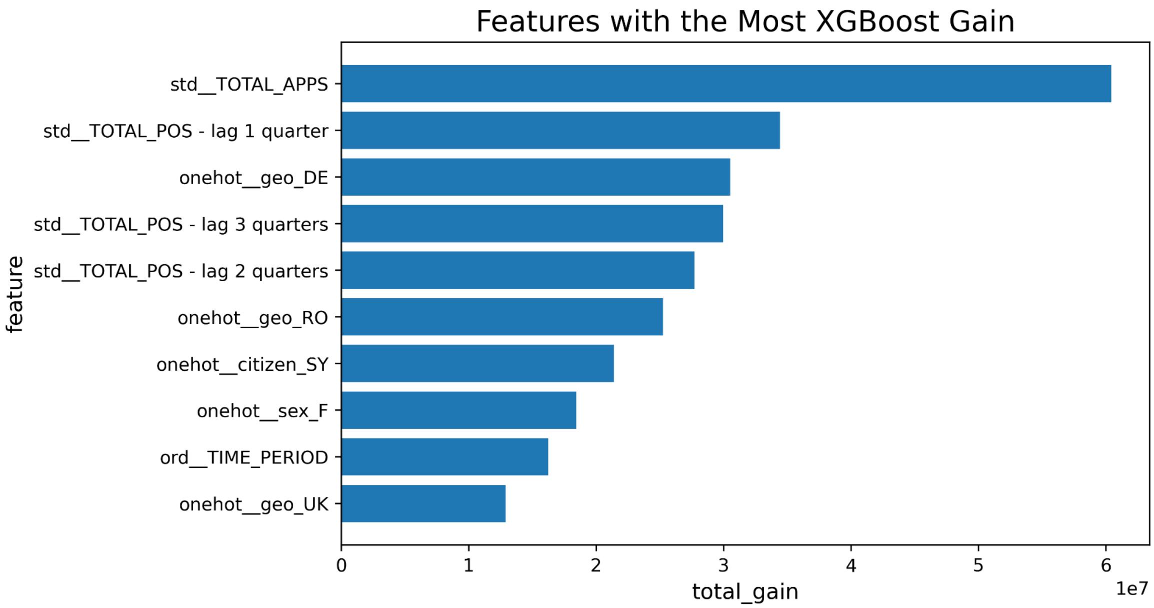

Fig 10. A bar chart showing the gain of each feature in the best performing XGBoost model.

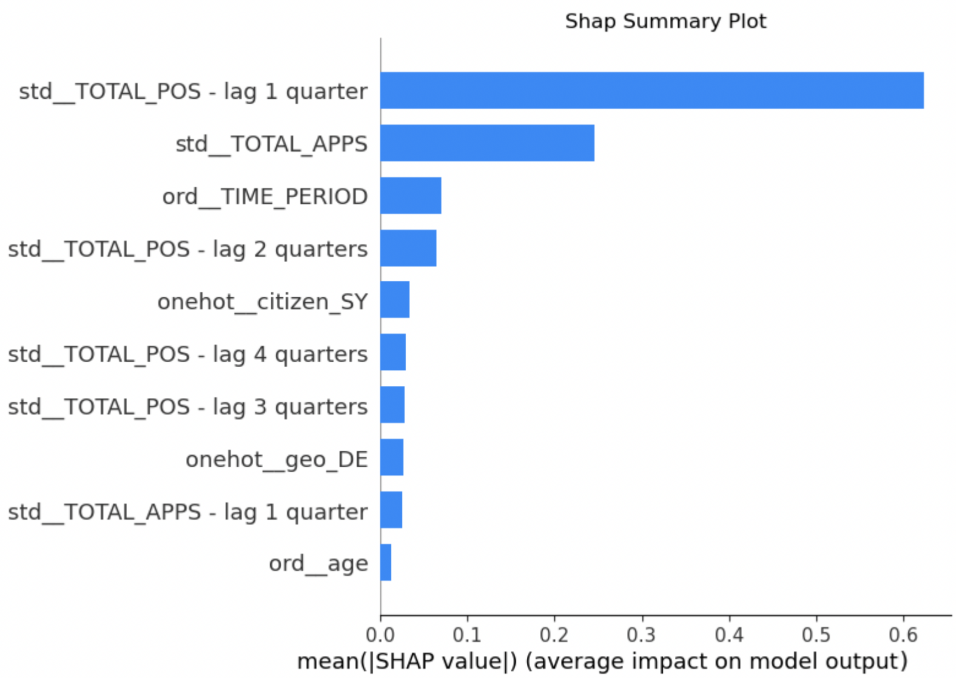

Fig 11. A bar chart showing the SHAP values of each feature (based on the best XGBoost model).

4.4 Model Interpretation - Local Feature Importance

From the SHAP values calculated above, we can create force plots (see Figures 12a-c) for three different datapoints from the test set to see how the model arrives at its predictions. This again reinforces the notion that the vast majority of the XGBoost model's consideration is typically placed on historical data and lagged features, with demographic data being used to tune an initial estimation based on specific features of the applicant pool.

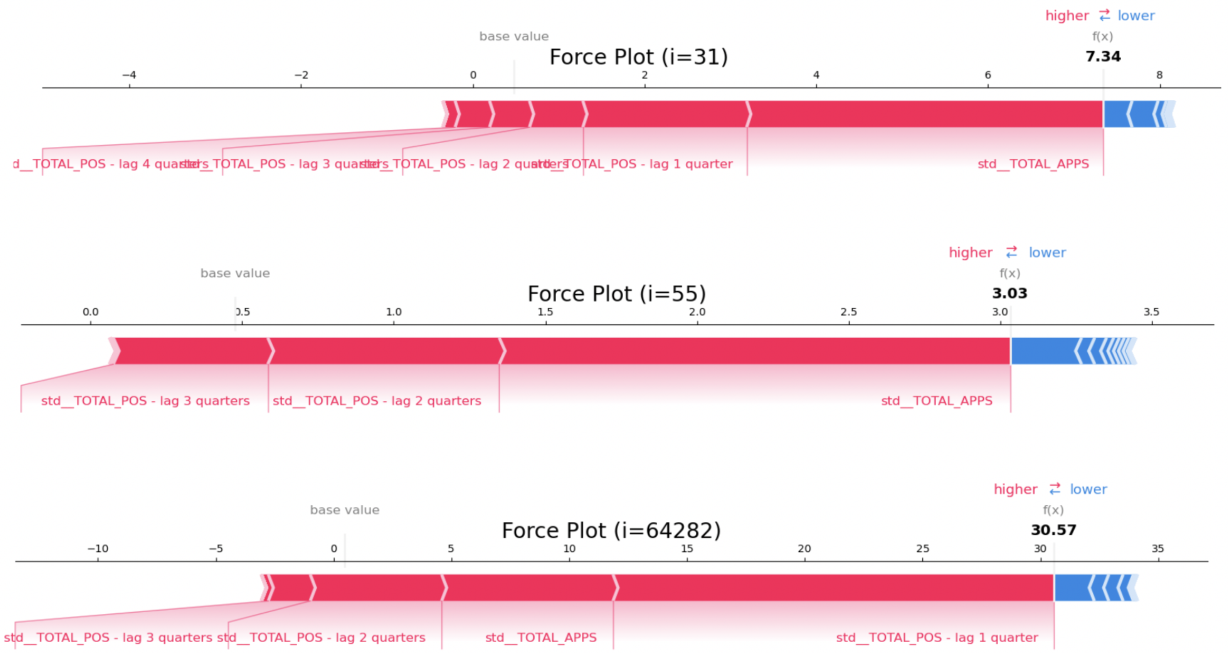

Fig 12a-c. Force plots showing the local feature importance for three example datapoints. Their true values are 10, 5, and 25 respectively.

5. Outlook

Given the large dataset used, computing power restraints had an impact on what could be done in the project's timeframe. Given more time or computing power, areas for improvement include:

- **Improvements in splitting methodology** such as doing hyperparameter optimization for each fold of the split data.
- Treating autoregressive **lag as a full hyperparameter** to further optimize historical data use.
- **Applying more complicated models** (like SVR with an rbf kernel) or attempting to more finely explore hyperparameters and random states.
- **Continued feature engineering** to encode country locations as continuous features (e.g. coordinates), allowing the model to potentially pick up on geographic trends.

6. References

[1] *migr: Predicting Asylum Application Outcomes in the European Union*, Alexander Portland, 2023. GitHub Repository. Available: https://github.com/AlexanderPortland/migr

[2] European Union Agency for Asylum, "Latest Asylum Trends." EUAA.europa.eu. https://euaa.europa.eu/latest-asylum-trends-asylum (accessed Oct 20, 2023)

[3] P. Katsikouli et al., "Machine Learning and Asylum Adjudications: From Analysis of Variations to Outcome Predictions," *Mobile*, no. 3, Feb. 2023.

[4] M. Carammia, S. M. Iacus and T. Wilkin, "Forecasting Asylum-Related Migration Flows with Machine Learning and Data at Scale," *Scientific Reports*, Jan. 2022.

[5] *First instance decisions on applications by citizenship, age and sex - quarterly data*, Eurostat, Dec. 2023. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/migr_asydcfstq/

[6] Eurostat, "Decisions on applications and resettlement: Reference Metadata in Euro SDMX Metadata Structure (ESMS)." https://ec.europa.eu/eurostat/cache/metadata/en/migr_asydec_esms.htm (accessed Sept. 2023)

[7] United Nations High Commissioner for Refugees, "About the UNHCR: The 1951 Refugee Convention." https://www.unhcr.org/us/about-unhcr/who-we-are/1951-refugee-convention (accessed Oct. 2023)

[8] United Nations Special Conference. *"Convention Relating to the Status of Refugees."* July 1951.

[9] J. Vassiliou, "What is the difference between refugee status and humanitarian protection?" Free Movement. https://freemovement.org.uk/difference-between-refugee-status/ (accessed Oct 2023)

[10] European Commission, "subsidiary protection." home-affiars.ec.europa.net. https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/subsidiary-protection_en (accessed Oct 2023)