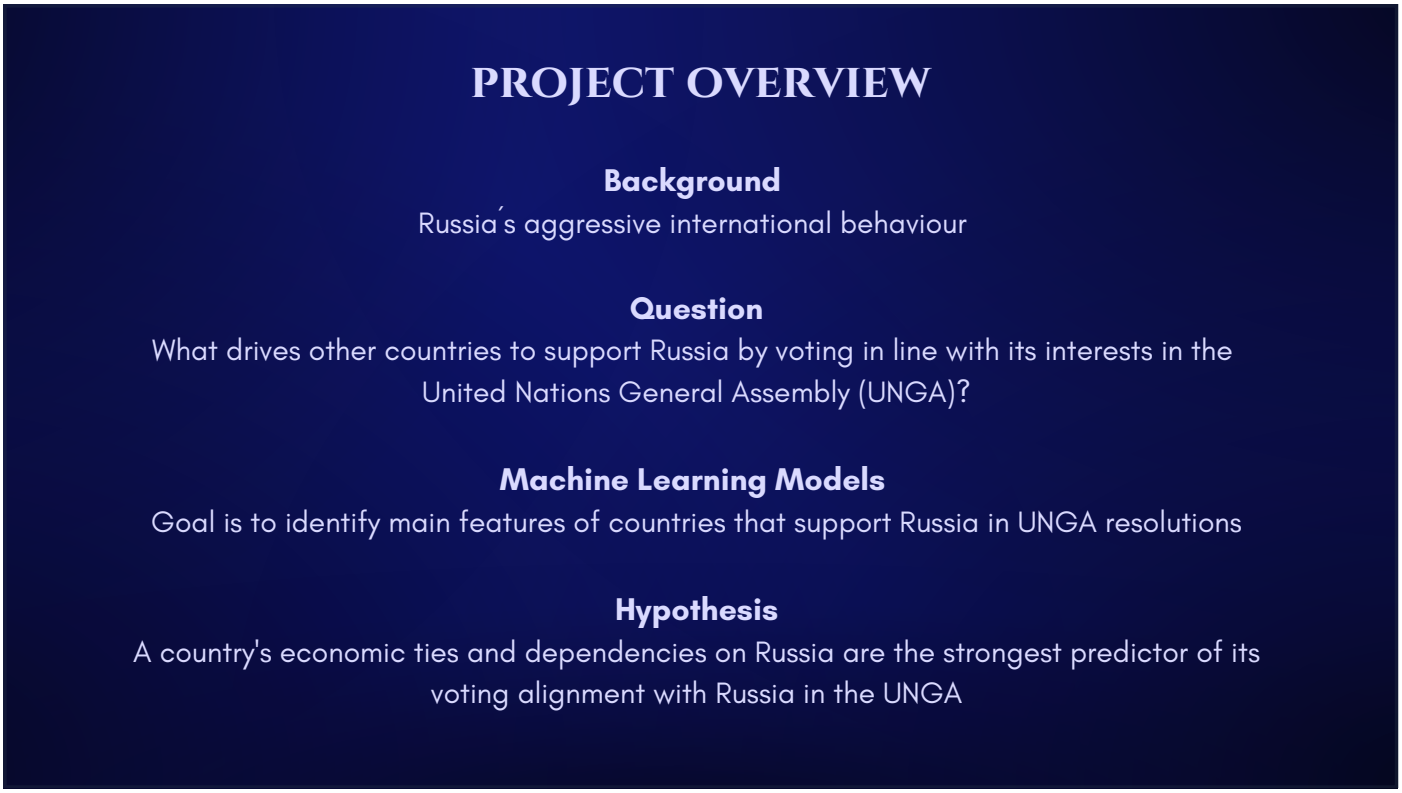

Welcome to the presentation of my data science project on the topic of ...




Background...
Question...
+ importance: if Russia can still gain allies despite its highly aggressive behaviour, that is very concerning. Therefore, understanding what motivates countries to support it is crucial.
ML Models
Hypothesis

# Data

**Dataset**

- 192 rows (based on UN members) and 12 columns
- Self-constructed from different sources (UN voting data, OECD, UNCTAD etc.)

**Target**

- **Pro-Russian Voting Index:** continuous values between 0 and 1
  35 UNGA resolutions crucial for Russian foreign policy (Ukraine, Georgia)
  Timeframe 2008-2023, coding: support = 1, against = 0, abstentions/absences = 0,5,
  Index per country based on arithmetic mean (sum values/ number resolutions)

**Features** (for each UN member)

- **Economic**: Mean Russian Aid Amount per Year ($ MM), Bilateral Investment Treaty with Russia (0/1), Export and Import Partner Share, GDP per capita ($)
- **Other**: Regime Type (democratic/authoritarian), Distance to Moscow in km, Comecon Membership (0/1), Defense Cooperation Agreement with Russia (0/1), Membership in Organisation with Russia (0/1)

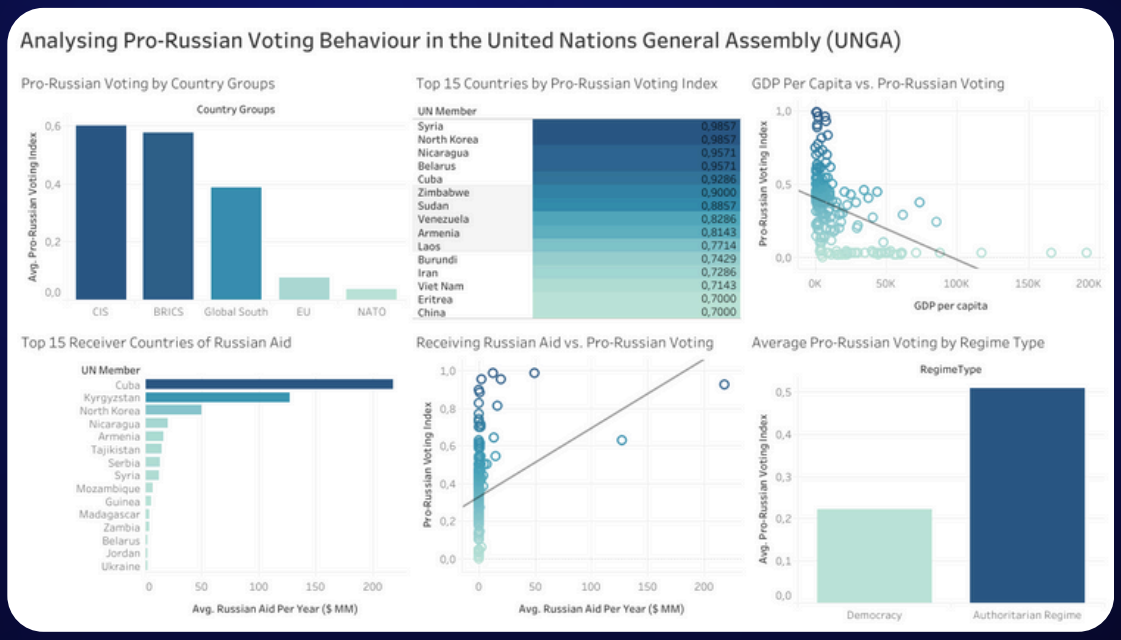
Dataset ...
Target...
Creating index: coded votes in support of Russia´s position as 1, against Russia´s interests as 0,
abstentions/absences = 0,5
For each country I calculated the arithmetic mean by taking the sum of all values dividing it  by number of resolutions

Comecon = Council für Mutual Economic Assistance

Org membership in BRICS, CIS, CSTO, EAEU, SCO

(mean values are based on target variable 2008-2023)

# Exploratory Data Analysis



# Data preparation

**Problem**

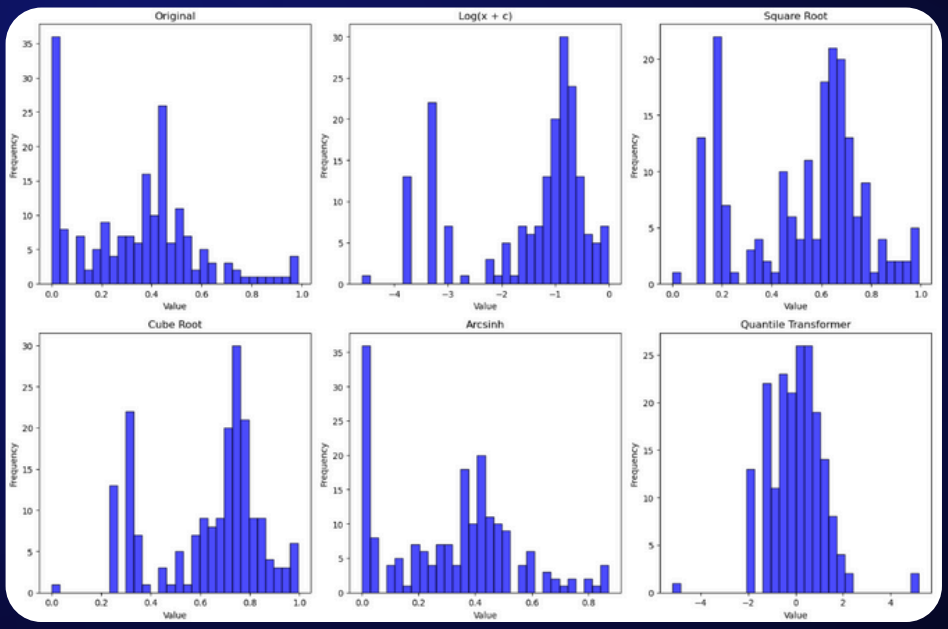
- Target Pro-Russian Voting Index not normally distributed

**Solution**

- Quantile Transformer most normal distribution

**Trade-Off**

- Predictions can be back-transformed but coefficients not interpretable in original scale anymore



# Feature Selection & Engineering

**Low Correlation with Target < 0.1**

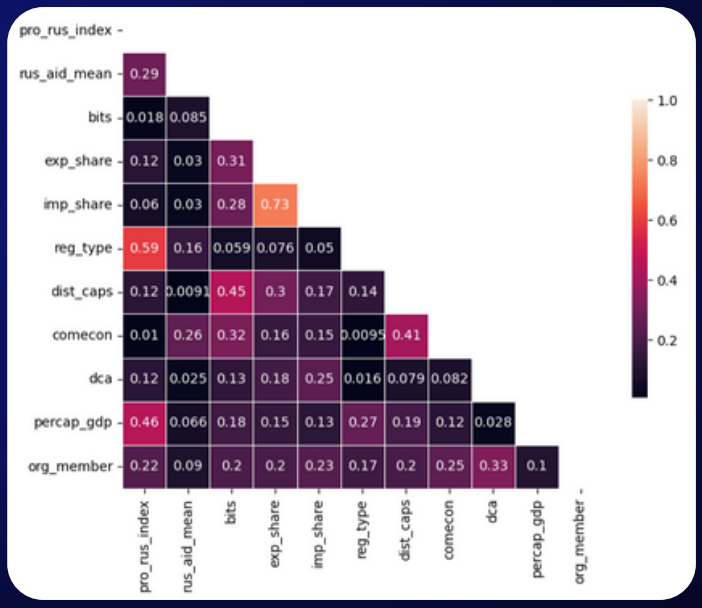
- Bilateral Investment Treaty
- Comecon Membership
- Import Partner Share

**High Feature Correlation**

- Import and Export Partner Share
- Already dropped Import

**Feature Engineering**

- Normalization of features for distance-based ML models due to big difference in values (GDP per capita, distance to Moscow etc.)



## Overview Of model Performance

**Distance-Based Models** with normalized features and transformed target

| Model | $R^2$ | Cross-Validation $R^2$ |
|---|---|---|
| **KNN** | 0.50 | 0.42 |
| **Linear Regression** | 0.32 | 0.28 |

- Performance drop after cross-validation indicates that the models were not generalising well before due to the quite small dataset with features that vary a lot for every country
- Selected random state 42 performed better than the average of multiple train/test splits

Applied cross-validation + hyperparameter tuning in the form of grid-search for all models

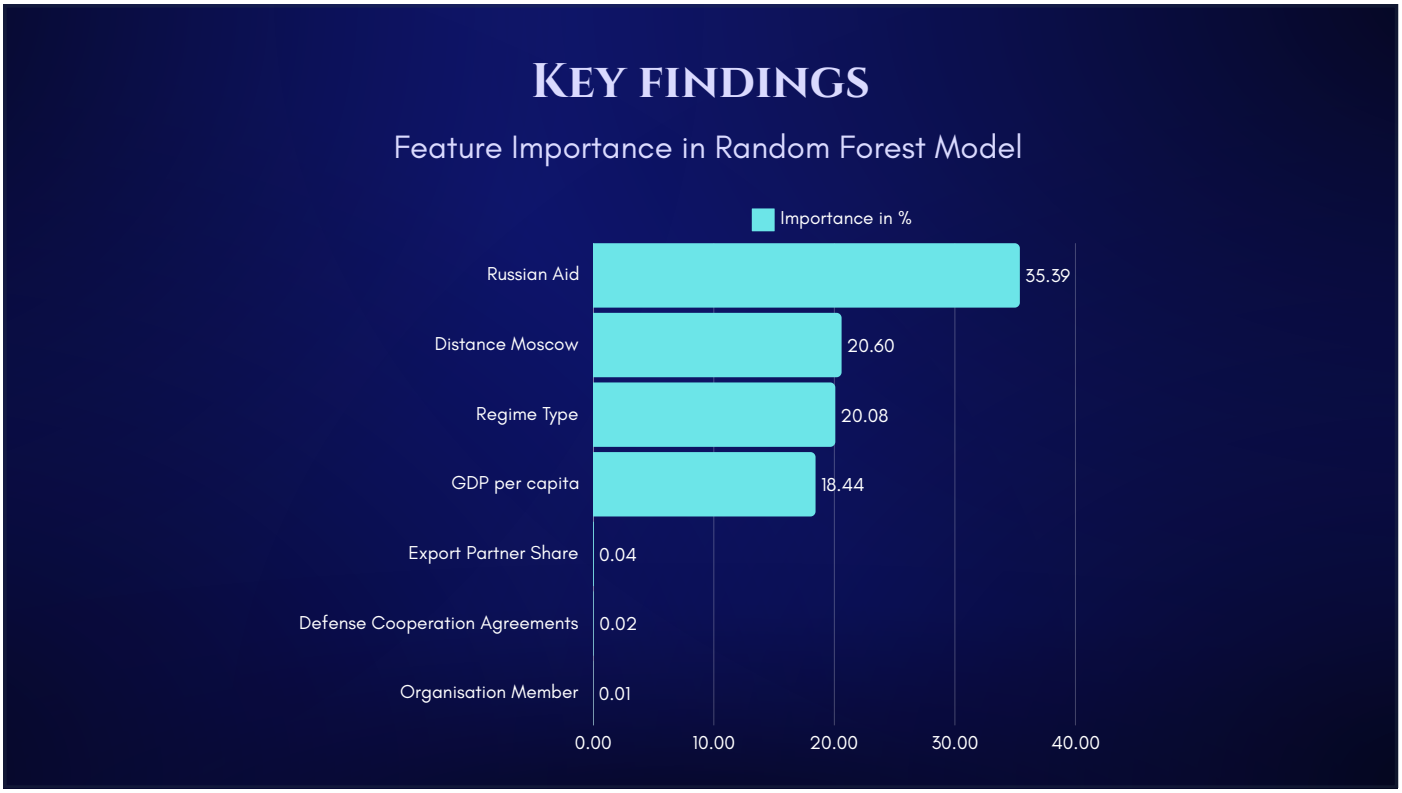
## Overview Of model Performance

**Decision Tree and Ensemble Methods**

| Model | $R^2$ | Cross-Validation $R^2$ |
|---|---|---|
| **Decision Tree** | 0.65 | 0.36 |
| **Random Forest** | 0.61 | **0.56** |
| **Ada Boost** | 0.54 | 0.53 |
| **Gradient Boost** | 0.46 | 0.51 |

Since performance drop legitimate – best model Random Forest with $R^2$ of 0.56

Since performance drop legitimate - best model Random Forest with $R^2$ of 0.56
or in other words - its a feature, not a bug




... % importance in predicting the target variable

# Limitations & Future work

**Explaining vs. Predicting**

- Focus rather on *explaining features* that lead to outcome than training a model to be able to predict the outcome itself
- Models do not have much application in making actual predictions because no new "rows" expected (new countries joining UN), only if country characteristics change, do the predictions change

**Model Performance**

- $R^2$ of 0.56 indicates room for improvement - missing features?
- Future work could include additional features which matter for predicting Pro-Russian voting (requires theoretical research before model building)

**Methods**

- Trying more advanced models?
- Using Time-series data?

# Conclusion

**Four Important Features in Pro-Russian Voting Identified**

- **Russian Aid**: Countries that receive more development assistance from Russia, on average have a higher Pro-Russian Voting Index
- **Distance to Moscow**: Russia on average receives the most voting support from countries that are located further away
- **Regime Type**: Authoritarian countries on average have a higher voting alignment with Russia than democracies
- **GDP per Capita**: Countries with a lower GDP per capita on average tend to support Russia more in the UNGA than countries with a higher GDP per capita

**Confirmation of Hypothesis**

- Import and Export Partner Share & Bilateral Investment Treaties not relevant (economic ties)
- Receipt of Russian Aid most important feature & GDP per capita important (dependencies)
- Economic dependencies on Russia are the strongest predictor for voting alignment with Russia in the UNGA

Despite the just mentioned limitations...

far away countries better allies - probably because Russia has destroyed the relationship to its closest neighbours and Europe


Thank You
Buying Friends?
Modelling Pro-Russian Voting Behaviour
in the United Nations General Assembly
Data Science Project by Maria