



Team Bear

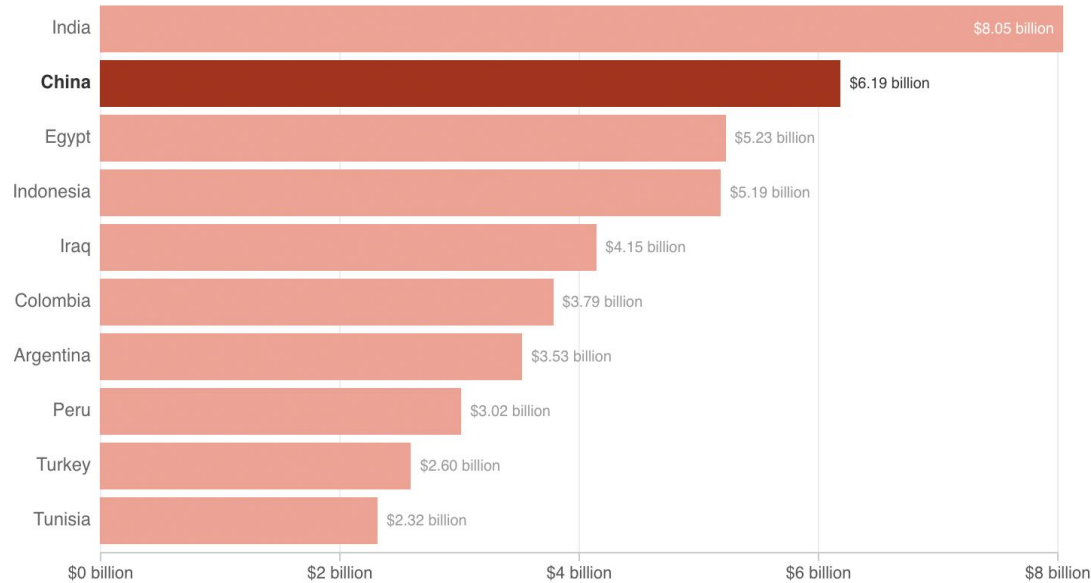
Qarik Corporate Project



Kashif Bari, Kevin Bombardier,
Attilio Castano, Bingjin Liu

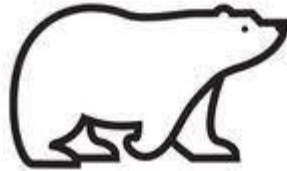


The World Bank

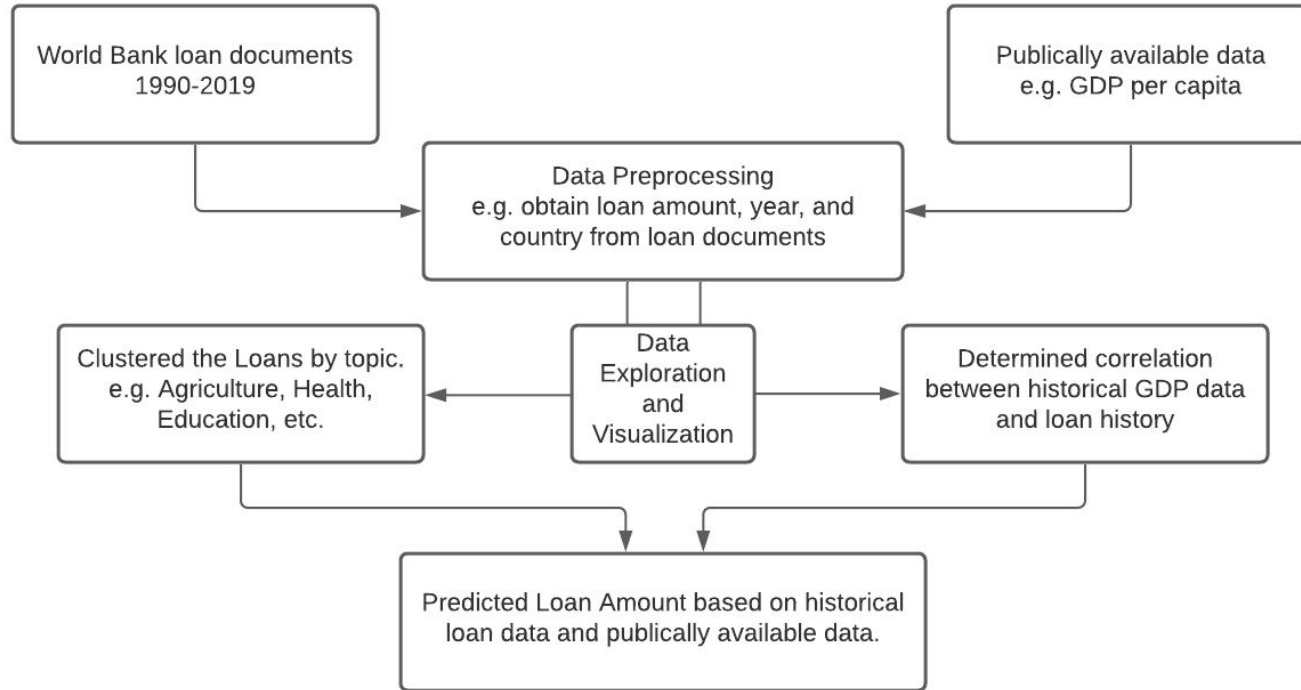


Loans given by the World Bank from 2016 to 2018

Our Goal



Project Execution Overview

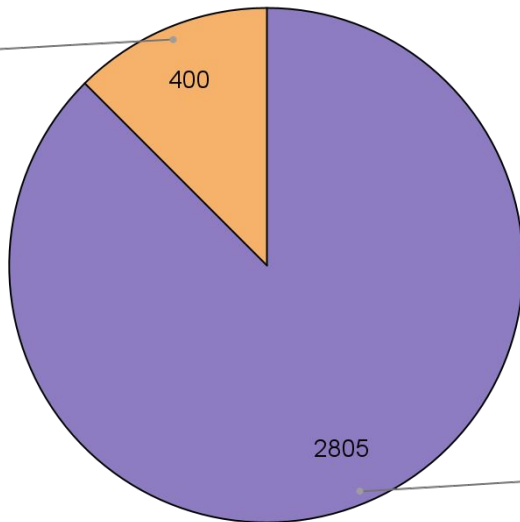


Loan Documents: from pdf to text

Images vs. Native

Images

12.5%



- PyMuPDF for Native PDF's
- pyTesseract OCR for Scanned PDF's

Collect Publicly Available Data



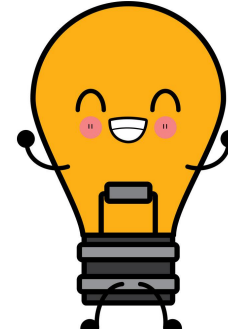
GDP



Stability (Absence
of Violence)



Literacy Rate



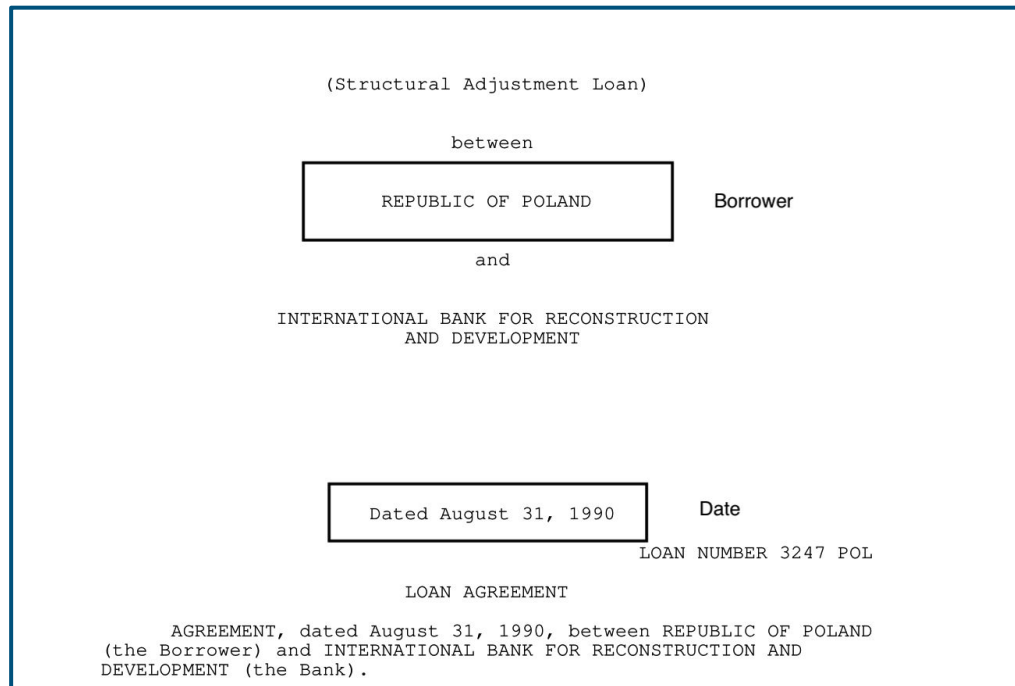
Electricity
Usage



Income
Inequality

Data Preprocessing

- Loan Amount
- Loan Date
- Borrower (Country)
- Project Name
- Project Description

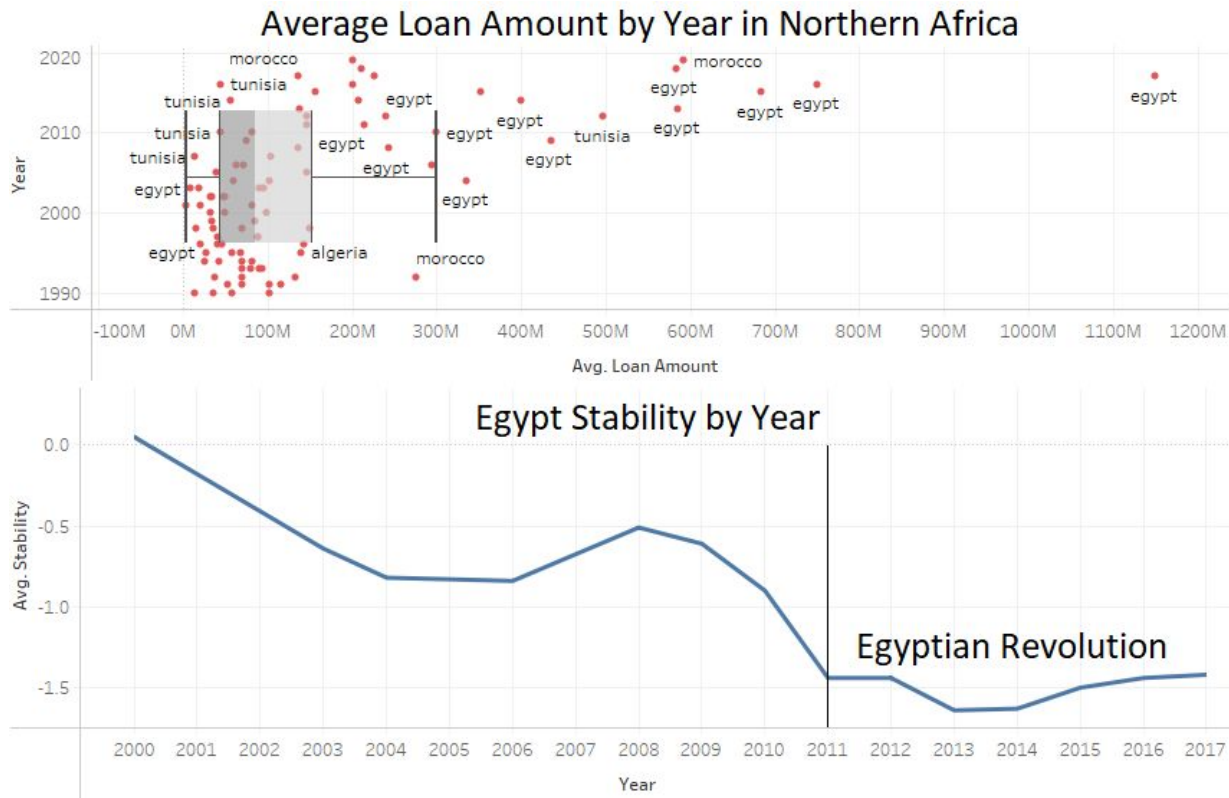


Average Loan Amount by Year and Region



The plot of average of Loan Amount for Year broken down by Region. The marks are labeled by Country.

Comparison of Loan Amount and Stability



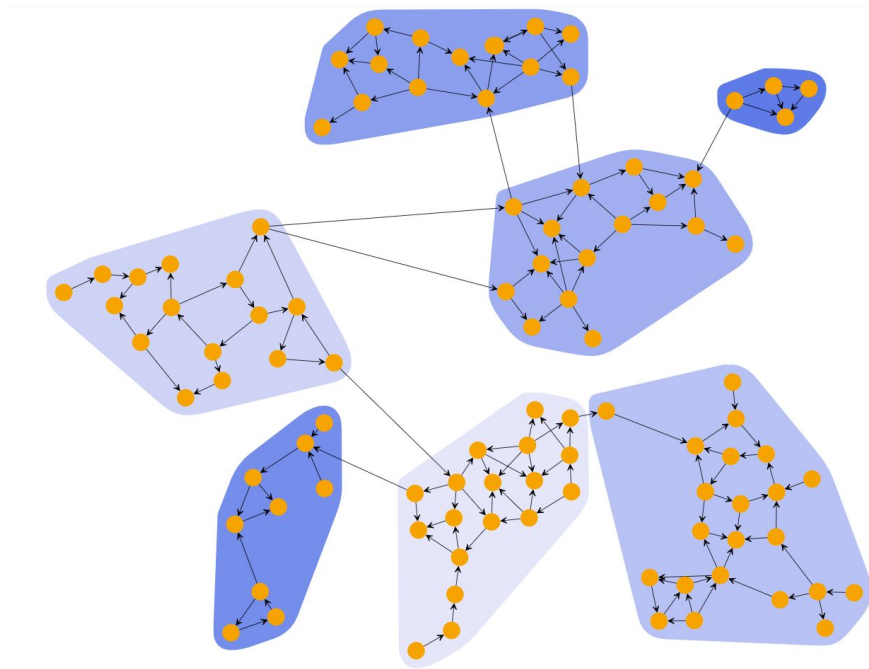
Two Approaches to Topic Analysis Clustering

Latent Dirichlet Allocation (LDA)

- Generative model
- 7 topics/clusters
- Evaluation of model
 - Topic Coherence: 0.586
 - Perplexity: -6.718

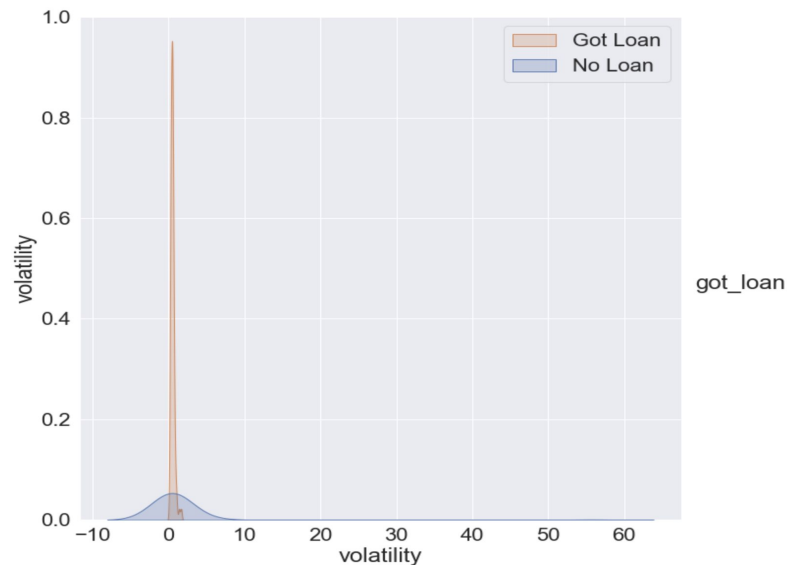
GloVe & KMeans

- GloVe: Transforms words into vectors
- KMeans: Clustering technique for vectors
- 11 topics/clusters
- Evaluation of model
 - Silhouette score: 0.112

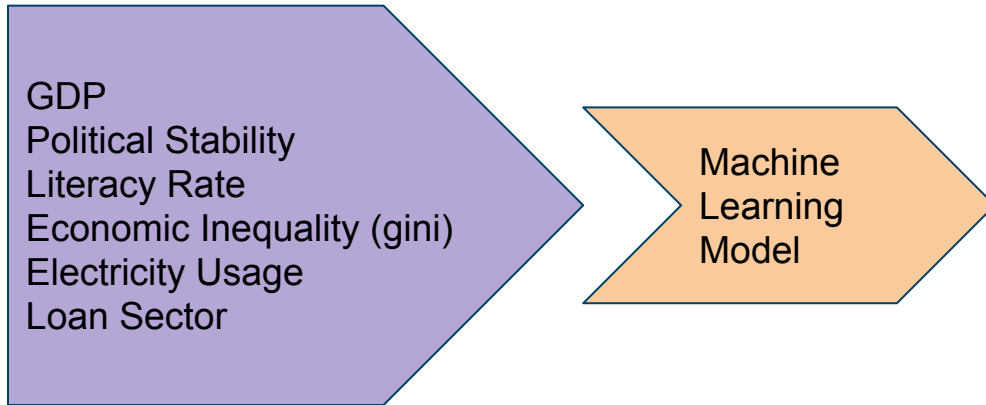


Predicting if a country got a loan from GDP Analysis

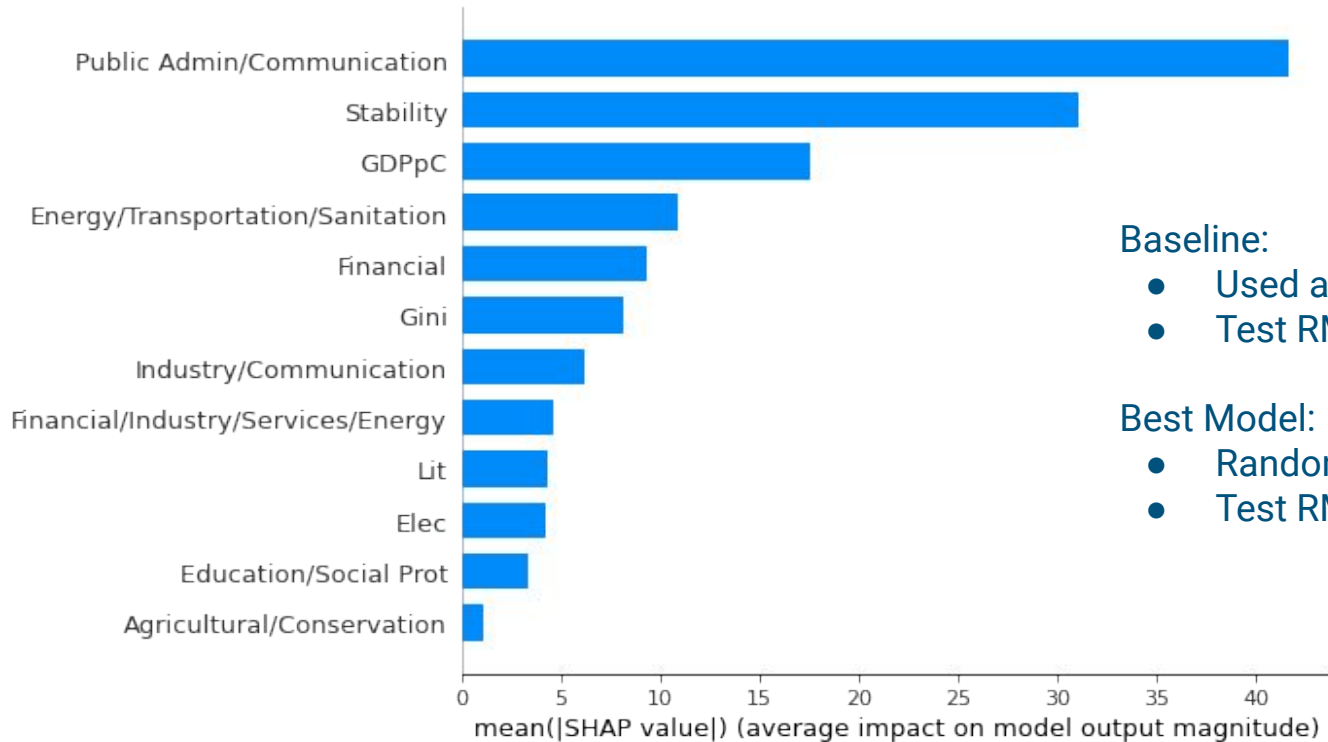
- Quantified correlation between
 - Historical GDP data
 - GDP Volatility = Normalized RMSE
 - Loan history
- Techniques
 - Linear Regression for feature eng.
 - Random Forest for classification.
- Evaluation of model
 - ROC AUC score ~0.75
 - Accuracy ~70%



Predicting Loan Amount



Ranking of features by mean SHAP values



Baseline:

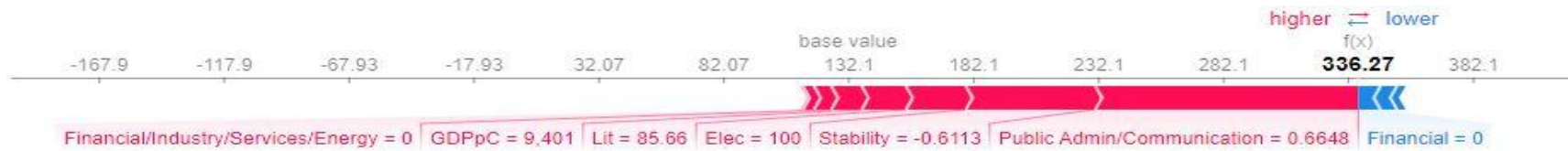
- Used average loan amount
- Test RMSE: \$164.55M

Best Model:

- Random Forest
- Test RMSE: \$121.67M

Conclusions

- Low GDP volatility is correlated to more loans.
- Political instability is not a barrier to apply for loans.
- Sector of the loan is correlated with the magnitude of the loan, as shown below.



For more details: <https://github.com/DoctorBomb/Qarik-Team-Bear>

Future Work

- Explore in more detail sector-specific trends.
- Obtain more data relevant to predicting loan amounts.

*thank
you*