AI Boot Camp Project 2

Project 2

Team Members:

- Thomas Brown
- Robb Stenman
- Jed Murphy

Project Overview

Project Purpose / Description

Analyze a problem using machine learning (ML):

o Analyze Exchange Rates and the relationship to Gross Domestic Product Growth (GDP Growth).

Project Goals

Goal/Questions to be addressed

- Goals:

o Analyze Exchange Rates for 10 countries and access correlation to GDP Growth.
o Analyze all 11,762 Exchange Rates for all countries and analyze correlation to GDP Growth.
o The model demonstrates meaningful predictive power at least 75% classification accuracy or 0.80 R-squared.

- Questions:

o Are the exchange rates for a Country a good predictor of annual GDP Growth for that Country?
o How closely correlated are exchange rates and GDP Growth for a country?

The Data

Data Sources

GDP Growth Data from The Department of the Treasury and the Bureau of the Fiscal Service.

Exchange Rate Data from World Bank's Development Data Group.

GDP Growth Data - Clean and Consistent
- Removed regions (East Asia & Pacific) and categories (OECD Members) via inner join
- Used Melt function to convert Year columns to rows for exchange rate data matching

Exchange Rate Data – Inconsistent and Duplicative
- Country names in different cases (all caps, title case) and spelling variations
- Currency spelling/expression variations

Standardization Approach
- Used a dictionary to standardize names

Approach

Approach taken to achieve goals

The analysis was broken into steps with a Jupyter notebook for each step:
Step1 Build base data
Step2 Reformat data for analysis
Step3 Analyze data
Performance was measured and R-squared calculated.
Random Forest Regressor produced the best R-squared.

Performance Improvement after removal of: Ridge Regression, Linear Regression, Lasso Regression calculations, and Random Forest Regressor:

- Before Adjustment Duration: 0:00:05.307291
- After Adjustment Duration: 0:00:04.185599

- 1.2 Second Improvement

Project Milestones:

- Project ideation – Complete 5/27
- Data fetching – Complete 5/27
- Data exploration – Complete 5/30
- Data transformation – Complete 6/3
- Data analysis – Complete 6/6
- Testing – Complete 6/10
- Creating documentation – Complete 6/10
- Creating the presentation – Complete 6/10

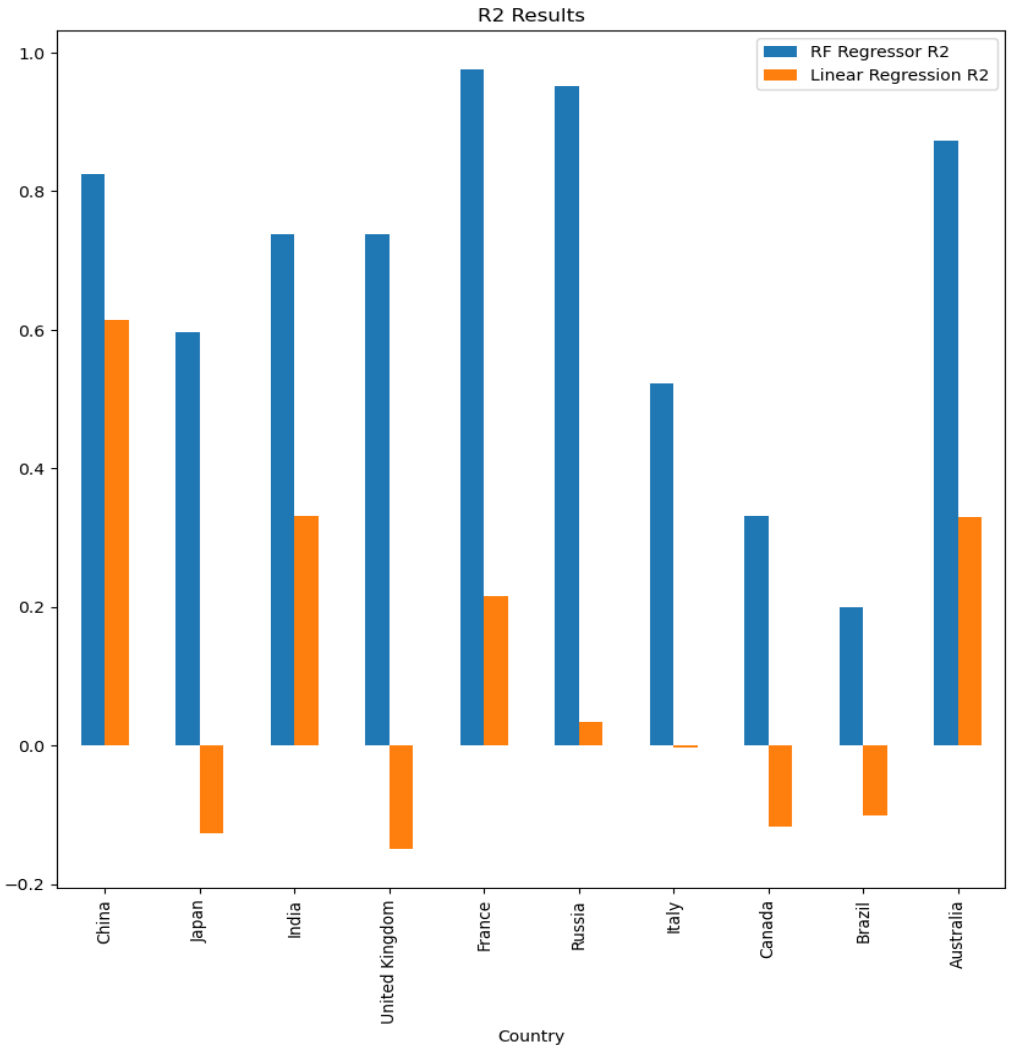
Result/Conclusion

Total of R2 Results using RF Regressor for all Countries taken together was 0.8799472692666176.

Country	Ridge MSE	Lasso Regression MSE	Linear Regression R2	Linear Regression MSE	RF Regressor R2
China	4.159085752	6.123385331	0.61371493	4.12010746	0.824533866
Japan	8.848326248	8.832380251	-0.127306103	8.847658167	0.596121356
India	7.708774877	10.53370344	0.330875249	7.679940508	0.737287742
United Kingdom	15.28339513	15.0632336	-0.149290932	15.32888591	0.737287742
France	4.191556172	5.357184351	0.216185394	4.177402261	0.975713065
Russia	19.6133965	18.64163694	0.033389697	19.6250163	0.95110876
Italy	10.77178642	10.72117766	-0.003361944	10.79370107	0.523217691
Canada	7.123663353	7.236712346	-0.117751931	7.123913001	0.331864356
Brazil	8.047473441	8.244038202	-0.10140481	8.109314673	0.199654724
Australia	0.949782473	1.446061229	0.329694694	0.943456912	0.873330075
Total for all Countries *					**0.879947269**

* " Random Forest Regressor" was run over all 11,762 US Treasury Exchange Rates

Result/Conclusion



Summary

Question 1: Are the exchange rates for a Country a good predictor of annual GDP Growth for that country?

R-squared is used as a measure of fit, or accuracy of the model, but what it actually tells you is about variance.
Our overall R-squared below indicates the Random Forest Regressor model is a good fit.
Training (R2) Score: 0.9764628099591433
Testing (R2) Score: 0.8799472692666176

Question 2: How closely correlated are exchange rates and GDP Growth for a country?

Pearson Correlation: -0.037599032386614 48 (weak)
Exchange Rates and GDP Growth are negatively correlated. As Exchange Rates go up, GDP Growth goes down.
Pearson is a liner correlation measure, Random Forest Feature Importance should be use for this measure.

Random Forest Feature importance refers to techniques that calculate a score for all the input features for a given model.
The scores represent the “importance” of each feature.
A higher score means that the specific feature will have a larger effect on the model that is being used to predict a certain variable.
Random Forest Feature Importance measurement:
Exchange Rate 0.391174
Effective Date 0.299577
Country Code 0.309249
Exchange Rate is the most "important" and has the larger effect on the model to predict GDP Growth.

Future Considerations

Additional research and data remediation for the issue below:

- In the data, the main issue was that some countries changed currencies during the period covered by the data. In this situation, the old and new currencies were both listed.
- Predict next year’s GDP Growth for a country, given the estimated Exchange Rate for that country.