Inflation Rate in Botswana

A study about inflation in botswana and its predictive features

Lefhoko Thapelo

2022

# Introduction

This is a study about inflation in Botswana and whether it can be predicted. Inflation is the increase in the price of goods or decrease in the buying power of money over a period of time. Being able to predict inflation can be very useful to not only the central bank, but also potential investors, immigrants, and even individuals who might be keeping their funds in a savings account. In order to ensure that the market in a country remains stable, a central bank’s role is to keep it steady using its monetary policy instruments such as controlling interest rates for lending money to commercial banks. The Bank of Botswana has the objective of keeping Botswana’s inflation between 3% an 6%.

The data for this study was collected from the Central Bank of Botswana website. It consists of Botswana's daily inflation rates, exchange rates, and interest rates set by the central bank. In addition, South Africa's inflation rates data were also collected from the SA Stats website. This is because it is generally known that Botswana depends on South Africa for a lot of goods and services and there is a good chance that South Africa’s exchange rates influence Botswana’s exchange rates and inflation rates.

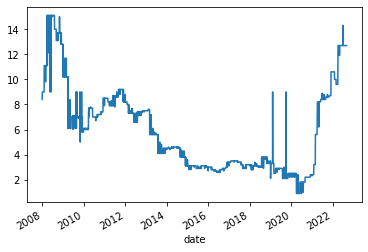
The goal of the project is to understand what factors affect inflation rate and to what degree they affect it. We also would like to see if we can predict inflation rate with good accuracy using various machine learning time series models.

# Data Wrangling

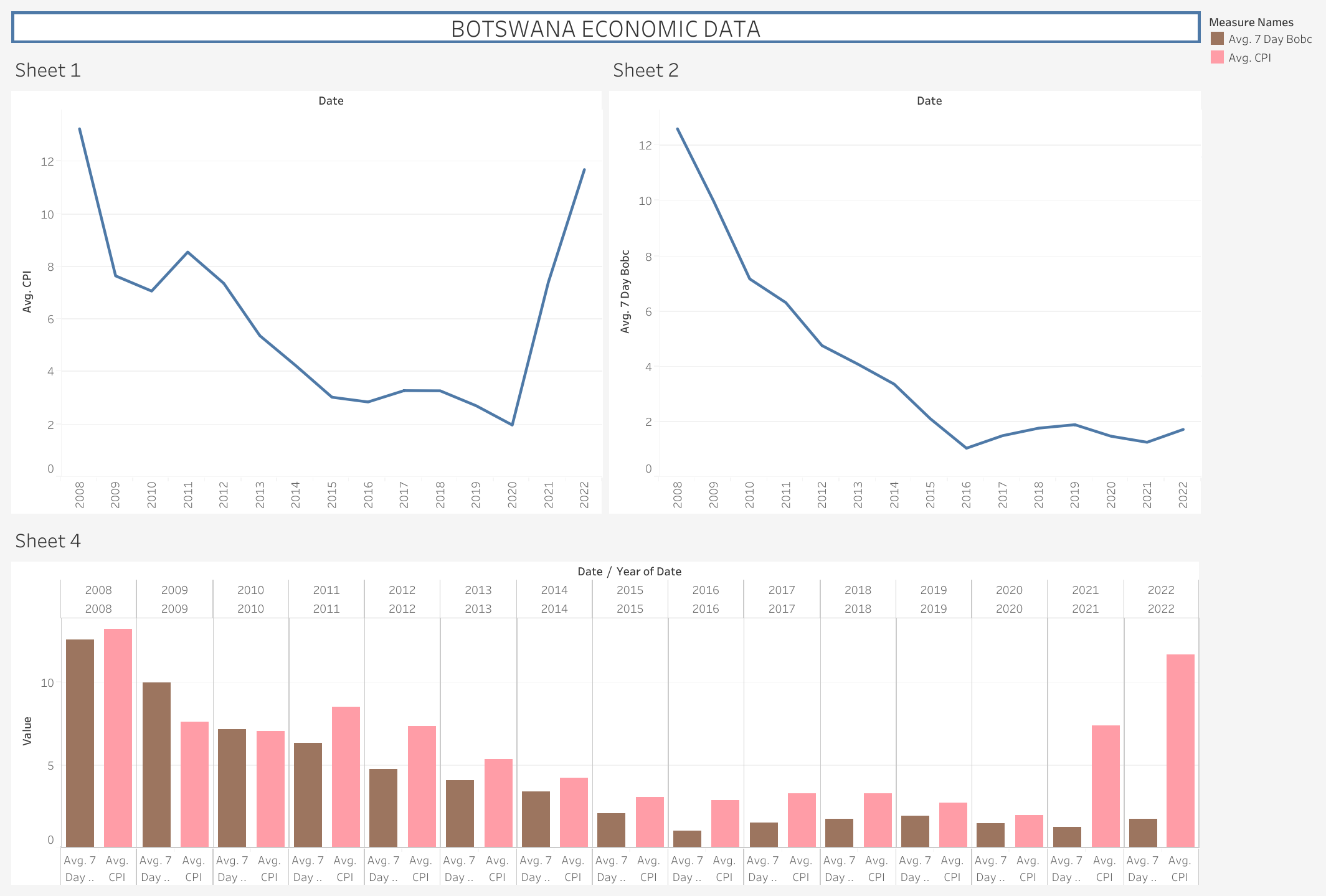
The data came in separate csv files for the inflation rate measures, the exchange rates, and the interest rates. All these files were merged into a single table or pandas DataFrame. The challenge was that parts of the data were recorded daily while other parts of the data were recorded monthly. So when the data was merged, it had a lot of missing values. For missing inflation rate values, a backward fill was done to ensure that inflation rates were consistent for each month by taking the value at the first entry of the month. Missing exchange rate values were all filled with the mean because this feature didn’t have any outliers. This means filling with the mean will not have much of an effect on the results from modelling.

# Exploratory Data Analysis

The only feature which seemed to have a relatively high correlation with CPI was interest rate for Bank of Botswana Certificate (BOBC) 7 Days. The Bank of Botswana uses this instrument to lend money to commercial banks and control the circulation of currency. The pearson correlation value between CPI and interest on BOBC 7 days was 0.599. This is counterintuitive because the expectation is that when interest rates go up, they should negatively impact CPI. One possible explanation for this is that the central bank anticipates rises in inflation rates and responds by increasing interest rates in order to moderate the inflation between 3% and 6% as their objective states.



The figure above shows the average CPI from 2008 to 2022. It peaks around 2008 and 2022 likely due to the 2008 global recession and the 2022 war between Russia and Ukraine. Covid 19 may have also influenced the spike in inflation in 2022.



Above is a screenshot of a tableau dashboard that shows the general trend of CPI and BOBC 7 days interest rate from 2008 to 2022.

# Data Pre-processing

There was hardly any feature engineering done on the dataset because all features were already numerical. The only work that was done was to convert the date field into a numeric field showing the number of days since the latest date of September 2, 2022. So the record for September 2, 2022 had zero in its date field. There was no need for dimensionality reduction since the features were very few.

Some basic linear regression models were used as a baseline. These were sklearn’s LinearRegression, and Ridge models. To do this, the data was split into lagged values and future values. This was done because future data cannot be used for predictions. Doing this can result in testing results not being representative of a model’s actual effectiveness or accuracy. Sklearn’s time series split was used to split the data.

The linear regression models performed very well with R2 scores of approximately one and MSE/MAE values less than two. This may have been because the predicted portion did not include the events that caused the spikes in 2008 and 2022.

# Modelling

In the modelling notebook, various regressor models were used to predict the inflation data. These included the decision tree regressor, random forest regressor, gradient boosting regressor, Bayesian ridge regressor, and time series prediction using StatsForecasts AutoARIMA and ETS models.

The performance of these models was not consistent. The models must have been able to predict some parts of the data than other parts. This makes it difficult to show average values since the values are so spread. Instead the values for each of the cross-validation folds will be displayed below:

Decision Tree Regressor:

Fold 1

R2 Score: -0.9820142471552693

Mean Squared Error: 0.4041878980891719

Root Mean Squared Error: 0.6357577353750183

Mean Absolute Error: 0.514171974522293

Fold 2

R2 Score: 0.5824543760116561

Mean Squared Error: 0.02979299363057424

Root Mean Squared Error: 0.17260647041920021

Mean Absolute Error: 0.14410828025477934

Fold 3

R2 Score: -0.9803901525821004

Mean Squared Error: 4.012531847133747

Root Mean Squared Error: 2.0031305117574707

Mean Absolute Error: 1.4664012738853482

Fold 4

R2 Score: -0.11593969595918852

Mean Squared Error: 0.6290923566878976

Root Mean Squared Error: 0.7931534256925943

Mean Absolute Error: 0.6218152866242038

Fold 5

R2 Score: 0.6747499065485201

Mean Squared Error: 4.747643312101891

Root Mean Squared Error: 2.178908743408473

Mean Absolute Error: 1.4165605095541358

Random Forest Regressor:

Fold 1

R2 Score: -0.7238677334708676

Mean Squared Error: 0.3515446353503183

Root Mean Squared Error: 0.5929119962948282

Mean Absolute Error: 0.4685111464968153

Fold 2

R2 Score: 0.695375584170054

Mean Squared Error: 0.021735764331210754

Root Mean Squared Error: 0.14743054070039477

Mean Absolute Error: 0.11782165605095653

Fold 3

R2 Score: 0.6015637257593847

Mean Squared Error: 0.8072844824840676

Root Mean Squared Error: 0.898490112624545

Mean Absolute Error: 0.6581640127388484

Fold 4

R2 Score: 0.16169983246779707

Mean Squared Error: 0.4725777117834328

Root Mean Squared Error: 0.687442878924084

Mean Absolute Error: 0.5116640127388499

Fold 5

R2 Score: 0.7325777901782987

Mean Squared Error: 3.903535437898063

Root Mean Squared Error: 1.9757366823284075

Mean Absolute Error: 1.2856544585987206

Gradient Boosting Regressor

Fold 0

R2 Score: -0.08055998537263842

Mean Squared Error: 0.22035627134057553

Root Mean Squared Error: 0.4694212088738381

Mean Absolute Error: 0.3673285472139096

Fold 1

R2 Score: 0.7936745873998956

Mean Squared Error: 0.014721868342684657

Root Mean Squared Error: 0.12133370653979321

Mean Absolute Error: 0.08267909095044809

Fold 2

R2 Score: 0.6596489424683245

Mean Squared Error: 0.6895961665790409

Root Mean Squared Error: 0.8304192715604816

Mean Absolute Error: 0.7085774212051785

Fold 3

R2 Score: 0.5966108964981174

Mean Squared Error: 0.2274038666274833

Root Mean Squared Error: 0.47686881490351546

Mean Absolute Error: 0.38547030408443717

Fold 4

R2 Score: 0.7928261933513724

Mean Squared Error: 3.0240954803131403

Root Mean Squared Error: 1.7389926625242387

Mean Absolute Error: 1.0533278884982948

Bayesian Ridge:

Fold 1

R2 Score: 0.9999999999999984

Mean Squared Error: 3.162119348315234e-16

Root Mean Squared Error: 1.7782348968331584e-08

Mean Absolute Error: 1.5682484269441835e-08

Fold 2

R2 Score: 1.0

Mean Squared Error: 1.6012674378371194e-19

Root Mean Squared Error: 4.0015839836708653e-10

Mean Absolute Error: 3.1474709604375455e-10

Fold 3

R2 Score: 1.0

Mean Squared Error: 3.587366386330994e-20

Root Mean Squared Error: 1.894034420577143e-10

Mean Absolute Error: 1.754642623695188e-10

Fold 4

R2 Score: 1.0

Mean Squared Error: 8.104309694218943e-22

Root Mean Squared Error: 2.846806929564937e-11

Mean Absolute Error: 2.525326020484001e-11

Fold 5

R2 Score: 1.0

Mean Squared Error: 1.1643686698463575e-21

Root Mean Squared Error: 3.4122846743001345e-11

Mean Absolute Error: 2.6537306499467925e-11

StatsForecast Models (AutoARIMA/ETS)

AUTO ARIMA

Fold 1

R2 Score: -1987.5466768126776

Mean Squared Error: 171.7928541239923

Root Mean Squared Error: 13.106977306915287

Mean Absolute Error: 12.939298834609982

ETS

Fold 1

R2 Score: -1075.1931698043077

Mean Squared Error: 92.9735763234711

Root Mean Squared Error: 9.642280659857972

Mean Absolute Error: 9.637799809265136

AUTO ARIMA

Fold 2

R2 Score: -323.34166301753623

Mean Squared Error: 150.35455875204494

Root Mean Squared Error: 12.261914970837342

Mean Absolute Error: 12.172532306289675

ETS

Fold 2

R2 Score: -165.03066564116605

Mean Squared Error: 76.96657666343694

Root Mean Squared Error: 8.773059709328152

Mean Absolute Error: 8.746599809265138

AUTO ARIMA

Fold 3

R2 Score: -64.19617989099494

Mean Squared Error: 77.42014807104579

Root Mean Squared Error: 8.79887197719377

Mean Absolute Error: 8.712984803390503

ETS

Fold 3

R2 Score: -23.23007601685278

Mean Squared Error: 28.773097996444744

Root Mean Squared Error: 5.364056114214759

Mean Absolute Error: 5.252199809265138

AUTO ARIMA

Fold 4

R2 Score: -147.4979649226479

Mean Squared Error: 83.14795466613889

Root Mean Squared Error: 9.11855003090617

Mean Absolute Error: 8.731815352249145

ETS

Fold 4

R2 Score: -50.52382478702004

Mean Squared Error: 28.849557971038855

Root Mean Squared Error: 5.371178452727004

Mean Absolute Error: 5.318799809265136

AUTO ARIMA

Fold 5

R2 Score: -0.29079228107420385

Mean Squared Error: 19.519581599884177

Root Mean Squared Error: 4.418097056412883

Mean Absolute Error: 3.9002001693725585

ETS

Fold 5

R2 Score: -0.29079228107420385

Mean Squared Error: 19.519581599884177

Root Mean Squared Error: 4.418097056412883

Mean Absolute Error: 3.9002001693725585

# Conclusion

Inflation is a difficult quantity to model because there are many economic factors that can influence it. Also, the central bank should be considered because it takes actions based on its assessment of the economic landscape.

The best performing models were the linear regression models. The performance is better than expected. The accuracy may have dropped significantly if the predictions were made to cover a larger time frame. Nevertheless, this means the linear regression model does perform quite well in predicting inflation.