SARB Features

January 24, 2021

```
[1]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt

     feature_set_sarb = pd.read_csv('data/sarb_features_data.csv').
      ↪drop(['unemployment rate'], axis=1).set_index('Date')
     target = pd.read_csv('data/sarb_target_data.csv').set_index('Date')
```

```
[2]: import warnings

     warnings.filterwarnings('ignore')
```

0.1 The full feature set

These feature were accessed from the South African Reserve Bank.

There are ***147 features in total****, these cover a significant portfion of the South African economy*

The data from 1922-01-01 to 2020-01-01 if it used for unemployment forecasting, deleting redudant observations is helpful

```
[3]: feature_set_sarb
```

```
[3]:             Final consumption expenditure by general government   \
     Date
     1922-01-01                                                NaN
     1922-02-01                                                NaN
     1922-03-01                                                NaN
     1922-04-01                                                NaN
     1922-05-01                                                NaN
     …                                                        …
     2019-09-01                                                NaN
     2019-10-01                                                NaN
     2019-11-01                                                NaN
     2019-12-01                                                NaN
     2020-01-01                                                NaN

                 Consolidated general government: Revenue  \
     Date
```

```
1922-01-01                                       NaN
1922-02-01                                       NaN
1922-03-01                                       NaN
1922-04-01                                       NaN
1922-05-01                                       NaN
…                                              …
2019-09-01                                       NaN
2019-10-01                                       NaN
2019-11-01                                       NaN
2019-12-01                                       NaN
2020-01-01                                       NaN

            Foreign liabilities: Total portfolio investment   \
Date
1922-01-01                                               NaN
1922-02-01                                               NaN
1922-03-01                                               NaN
1922-04-01                                               NaN
1922-05-01                                               NaN
…                                                      …
2019-09-01                                               NaN
2019-10-01                                               NaN
2019-11-01                                               NaN
2019-12-01                                               NaN
2020-01-01                                               NaN

            Foreign liabilities: Portfolio investment: Equity securities  \
Date
1922-01-01                                                NaN
1922-02-01                                                NaN
1922-03-01                                                NaN
1922-04-01                                                NaN
1922-05-01                                                NaN
…                                                       …
2019-09-01                                                NaN
2019-10-01                                                NaN
2019-11-01                                                NaN
2019-12-01                                                NaN
2020-01-01                                                NaN

            Domestic output: All groups   \
Date
1922-01-01                           NaN
1922-02-01                           NaN
1922-03-01                           NaN
1922-04-01                           NaN
1922-05-01                           NaN
```

```
…                                  …
2019-09-01                         114.3
2019-10-01                         114.6
2019-11-01                         114.3
2019-12-01                         114.5
2020-01-01                           NaN

            Final consumption expenditure by households: Total   \
Date
1922-01-01                                                NaN
1922-02-01                                                NaN
1922-03-01                                                NaN
1922-04-01                                                NaN
1922-05-01                                                NaN
…                                                       …
2019-09-01                                                NaN
2019-10-01                                                NaN
2019-11-01                                                NaN
2019-12-01                                                NaN
2020-01-01                                                NaN

            Gross fixed capital formation   \
Date
1922-01-01                             NaN
1922-02-01                             NaN
1922-03-01                             NaN
1922-04-01                             NaN
1922-05-01                             NaN
…                                    …
2019-09-01                             NaN
2019-10-01                             NaN
2019-11-01                             NaN
2019-12-01                             NaN
2020-01-01                             NaN

            SDDS - Financial derivative liabilities  \
Date
1922-01-01                                      NaN
1922-02-01                                      NaN
1922-03-01                                      NaN
1922-04-01                                      NaN
1922-05-01                                      NaN
…                                             …
2019-09-01                                      NaN
2019-10-01                                      NaN
2019-11-01                                      NaN
2019-12-01                                      NaN
```

```
2020-01-01                                      NaN

            Foreign liabilities: Portfolio investment: Debt securities  \
Date
1922-01-01                                                NaN
1922-02-01                                                NaN
1922-03-01                                                NaN
1922-04-01                                                NaN
1922-05-01                                                NaN
…                                                       …
2019-09-01                                                NaN
2019-10-01                                                NaN
2019-11-01                                                NaN
2019-12-01                                                NaN
2020-01-01                                                NaN

            Change in inventories   …  \
Date                                …
1922-01-01                     NaN  …
1922-02-01                     NaN  …
1922-03-01                     NaN  …
1922-04-01                     NaN  …
1922-05-01                     NaN  …
…                            …  …
2019-09-01                     NaN  …
2019-10-01                     NaN  …
2019-11-01                     NaN  …
2019-12-01                     NaN  …
2020-01-01                     NaN  …

            Physical volume of manufacturing production: Total   \
Date
1922-01-01                                                NaN
1922-02-01                                                NaN
1922-03-01                                                NaN
1922-04-01                                                NaN
1922-05-01                                                NaN
…                                                       …
2019-09-01                                               99.0
2019-10-01                                              101.5
2019-11-01                                              100.0
2019-12-01                                                NaN
2020-01-01                                                NaN

            Remuneration per worker in non-agricultural: Total   \
Date
1922-01-01                                                NaN
```

```
1922-02-01                                                NaN
1922-03-01                                                NaN
1922-04-01                                                NaN
1922-05-01                                                NaN
…                                                        …
2019-09-01                                                NaN
2019-10-01                                                NaN
2019-11-01                                                NaN
2019-12-01                                                NaN
2020-01-01                                                NaN

            Consolidated general government: Non-financial assets - Net  \
Date
1922-01-01                                                NaN
1922-02-01                                                NaN
1922-03-01                                                NaN
1922-04-01                                                NaN
1922-05-01                                                NaN
…                                                        …
2019-09-01                                                NaN
2019-10-01                                                NaN
2019-11-01                                                NaN
2019-12-01                                                NaN
2020-01-01                                                NaN

            Consolidated general government: Cash surplus / deficit  \
Date
1922-01-01                                                NaN
1922-02-01                                                NaN
1922-03-01                                                NaN
1922-04-01                                                NaN
1922-05-01                                                NaN
…                                                        …
2019-09-01                                                NaN
2019-10-01                                                NaN
2019-11-01                                                NaN
2019-12-01                                                NaN
2020-01-01                                                NaN

            CPI Headline   Gross domestic expenditure   \
Date
1922-01-01           0.6                         NaN
1922-02-01           0.6                         NaN
1922-03-01           0.6                         NaN
1922-04-01           0.6                         NaN
1922-05-01           0.6                         NaN
…                   …                           …
```

```
2019-09-01          113.4                          NaN
2019-10-01          113.4                          NaN
2019-11-01          113.5                          NaN
2019-12-01          113.8                          NaN
2020-01-01            NaN                          NaN

            Net cash-flow from operating activities   \
Date
1922-01-01                                       NaN
1922-02-01                                       NaN
1922-03-01                                       NaN
1922-04-01                                       NaN
1922-05-01                                       NaN
…                                              …
2019-09-01                                       NaN
2019-10-01                                       NaN
2019-11-01                                       NaN
2019-12-01                                       NaN
2020-01-01                                       NaN

            Non-agricultural employment: Total   \
Date
1922-01-01                                  NaN
1922-02-01                                  NaN
1922-03-01                                  NaN
1922-04-01                                  NaN
1922-05-01                                  NaN
…                                         …
2019-09-01                                  NaN
2019-10-01                                  NaN
2019-11-01                                  NaN
2019-12-01                                  NaN
2020-01-01                                  NaN

            Consolidated general government: Expense  Residual item
Date
1922-01-01                                       NaN             NaN
1922-02-01                                       NaN             NaN
1922-03-01                                       NaN             NaN
1922-04-01                                       NaN             NaN
1922-05-01                                       NaN             NaN
…                                              …             …
2019-09-01                                       NaN             NaN
2019-10-01                                       NaN             NaN
2019-11-01                                       NaN             NaN
2019-12-01                                       NaN             NaN
2020-01-01                                       NaN             NaN
```

```
[1432 rows x 147 columns]
```

0.2 Additional Feature or Target

Depending on the purpose the **unemployment rate** be considered as a *target*, if you want to forecast it. Or is another feature if you want to forecast something else

[4]:

```
target
```

[4]:

```
              unemployment rate
Date
1922-01-01                  NaN
1922-02-01                  NaN
1922-03-01                  NaN
1922-04-01                  NaN
1922-05-01                  NaN
…                         …
2019-09-01                 29.1
2019-10-01                  NaN
2019-11-01                  NaN
2019-12-01                 29.1
2020-01-01                  NaN

[1432 rows x 1 columns]
```

1 Important Note: Missing Data from mixing frequencies

1.0.1 Ensure to review SARB Feature List.pdf to see the frequency of each feature that was accessed fro the SARB (https://github.com/rudzanimulaudzi/sarb__feature__set/blob/main/SARB%20Feature%20

The missing data occurs because we are merging data that monthly and data that is quartely, hence all quartely data should be expected to have missing data. This is normal when dealing with multiple time series.

In this case the data is from 1922. Some data observations should be deleted according to match the target variable. This is will improve the percentage of missing data.

[7]:

```
#Here we visualize the frequency of missingness of each feature
feature = feature_set_sarb.isna().sum()/len(feature_set_sarb)

feature = feature.sort_values(ascending=False)
feature_df = pd.DataFrame(feature.index, columns=['Feature Name'])
feature_df['Missing Frequency'] = np.array(feature.values)
feature_df['Rank'] = feature_df['Missing Frequency'].rank(ascending=False)
feature_df
```

[7]:

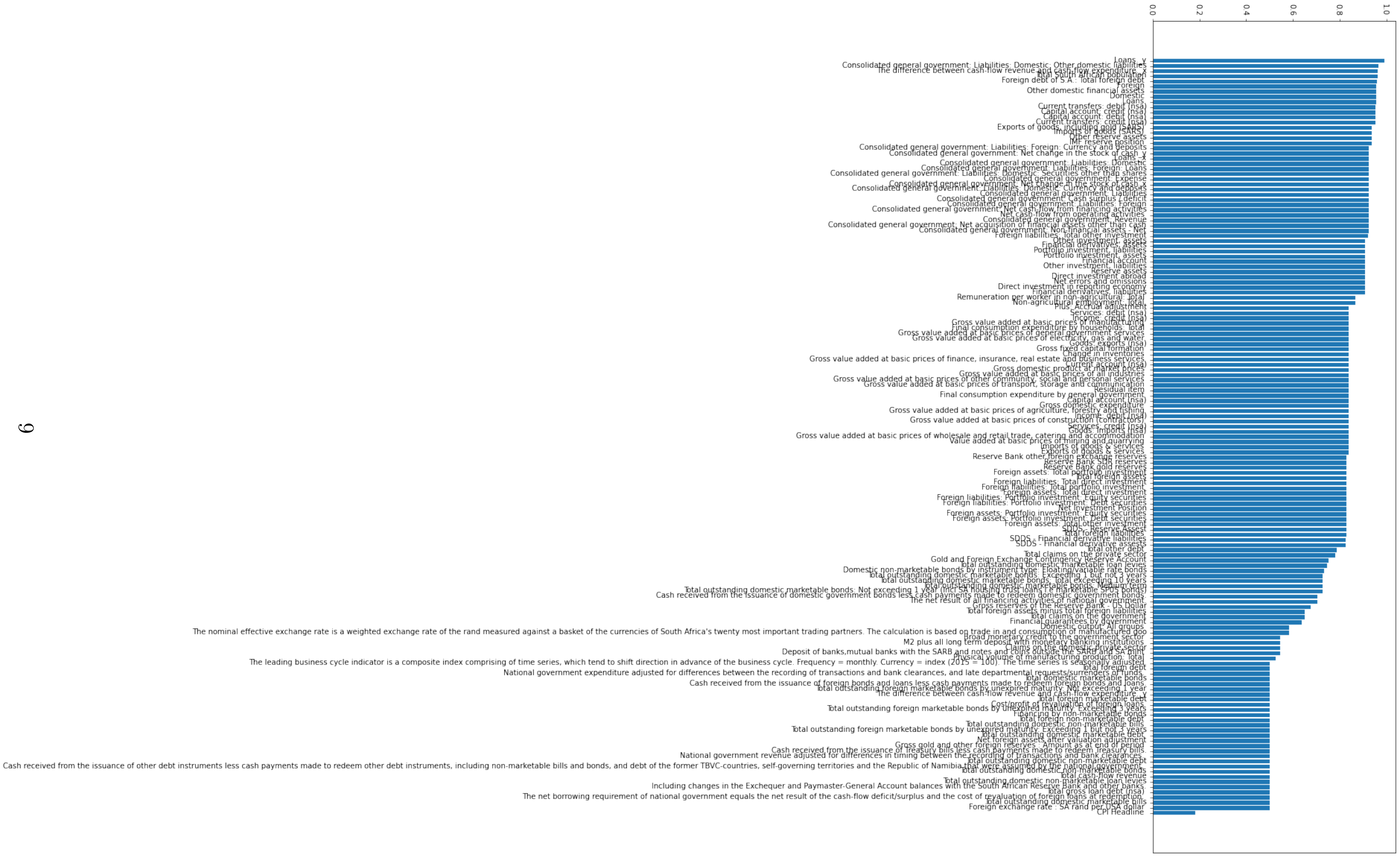
```
                                          Feature Name  Missing Frequency  \
0                                             Loans _y           0.986732
1    Consolidated general government: Liabilities: …           0.959497
2    The difference between cash-flow revenue and c…           0.958799
3                       Total South African population           0.958101
4             Foreign debt of S.A.: Total foreign debt           0.953212
..                                                 …                 …
142                        Total gross loan debt (nsa)           0.497207
143  The net borrowing requirement of national gove…           0.497207
144        Total outstanding domestic marketable bills           0.497207
145     Foreign exchange rate : SA rand per USA dollar           0.496508
146                                       CPI Headline           0.178771

      Rank
0      1.0
1      2.0
2      3.0
3      4.0
4      5.0
..     …
142  132.0
143  132.0
144  132.0
145  146.0
146  147.0

[147 rows x 3 columns]
```

[8]:

```python
plt.figure(figsize=(20,6))
plt.xticks(rotation=90)
plt.bar(feature_df['Feature Name'], feature_df['Missing Frequency'])
```

[8]:

```
<BarContainer object of 147 artists>
```


0.0
0.2
0.4
0.6
0.8
1.0
Loans_y
Consolidated general government: Liabilities: Domestic: Other domestic liabilities
The difference between cash-flow revenue and cash-flow expenditure._x
Total South African population
Foreign debt of S.A.: Total foreign debt
Foreign
Other domestic financial assets
Domestic
Loans
Current transfers: debit (nsa)
Capital account: credit (nsa)
Capital account: debit (nsa)
Current transfers: credit (nsa)
Exports of goods, including gold (SARS)
Imports of goods (SARS)
Other reserve assets
IMF reserve position
Consolidated general government: Liabilities: Foreign: Currency and deposits
Consolidated general government: Net change in the stock of cash_y
Loans_x
Consolidated general government: Liabilities: Domestic
Consolidated general government: Liabilities: Foreign: Loans
Consolidated general government: Liabilities: Domestic: Securities other than shares
Consolidated general government: Expense
Consolidated general government: Net change in the stock of cash_x
Consolidated general government: Liabilities: Domestic: Currency and deposits
Consolidated general government: Liabilities
Consolidated general government: Cash surplus / deficit
Consolidated general government: Liabilities: Foreign
Consolidated general government: Net cash-flow from financing activities
Net cash-flow from operating activities
Consolidated general government: Revenue
Consolidated general government: Net acquisition of financial assets other than cash
Consolidated general government: Non-financial assets - Net
Foreign liabilities: Total other investment
Other investment, assets
Financial derivatives, assets
Portfolio investment, liabilities
Portfolio investment, assets
Financial account
Other investment, liabilities
Reserve assets
Direct investment abroad
Net errors and omissions
Direct investment in reporting economy
Financial derivatives, liabilities
Remuneration per worker in non-agricultural: Total
Non-agricultural employment: Total
Plus: Accrual adjustment
Services: debit (nsa)
Income: credit (nsa)
Gross value added at basic prices of manufacturing
Final consumption expenditure by households: Total
Gross value added at basic prices of general government services
Gross value added at basic prices of electricity, gas and water
Goods: exports (nsa)
Gross fixed capital formation
Change in inventories
Gross value added at basic prices of finance, insurance, real estate and business services
Current account (nsa)
Gross domestic product at market prices
Gross value added at basic prices of all industries
Gross value added at basic prices of other community, social and personal services
Gross value added at basic prices of transport, storage and communication
Residual item
Final consumption expenditure by general government
Capital account (nsa)
Gross domestic expenditure
Gross value added at basic prices of agriculture, forestry and fishing
Income: debit (nsa)
Gross value added at basic prices of construction (contractors)
Services: credit (nsa)
Goods: imports (nsa)
Gross value added at basic prices of wholesale and retail trade, catering and accommodation
Value added at basic prices of mining and quarrying
Imports of goods & services
Exports of goods & services
Reserve Bank other foreign exchange reserves
Reserve Bank SDR reserves
Reserve Bank gold reserves
Foreign assets: Total portfolio investment
Total foreign assets
Foreign liabilities: Total direct investment
Foreign liabilities: Total portfolio investment
Foreign assets: Total direct investment
Foreign liabilities: Portfolio investment: Equity securities
Foreign liabilities: Portfolio investment: Debt securities
Net Investment Position
Foreign assets: Portfolio investment: Equity securities
Foreign assets: Portfolio investment: Debt securities
Foreign assets: Total other investment
SDDS - Reserve Assest
Total foreign liabilities
SDDS - Financial derivative liabilities
SDDS - Financial derivative assests
Total other debt
Total claims on the private sector
Gold and Foreign Exchange Contingency Reserve Account
Total outstanding domestic marketable loan levies
Domestic non-marketable bonds by instrument type: Floating/variable rate bonds
Total outstanding domestic marketable bonds: Exceeding 1 but not 3 years
Total outstanding domestic marketable bonds: Total exceeding 10 years
Total outstanding domestic marketable bonds: Medium term
Total outstanding domestic marketable bonds: Not exceeding 1 year (incl SA housing trust loans i.e marketable SP05 bonds)
Cash received from the issuance of domestic government bonds less cash payments made to redeem domestic government bonds.
The net result of all financing activities of national government.
Gross reserves of the Reserve Bank - US Dollar
Total foreign assets minus total foreign liabilities
Total claims on the government
Financial guarantees by government
Domestic output: All groups
The nominal effective exchange rate is a weighted exchange rate of the rand measured against a basket of the currencies of South Africa's twenty most important trading partners. The calculation is based on trade in and consumption of manufactured goo
Broad monetary credit to the government sector
M2 plus all long term deposit with monetary banking institutions
Claims on the domestic private sector
Deposit of banks,mutual banks with the SARB and notes and coins outside the SARB and SA mint
Physical volume of manufacturing production: Total
The leading business cycle indicator is a composite index comprising of time series, which tend to shift direction in advance of the business cycle. Frequency = monthly. Currency = index (2015 = 100). The time series is seasonally adjusted.
Total foreign debt
National government expenditure adjusted for differences between the recording of transactions and bank clearances, and late departmental requests/surrenders of funds.
Total domestic marketable bonds
Cash received from the issuance of foreign bonds and loans less cash payments made to redeem foreign bonds and loans.
Total outstanding foreign marketable bonds by unexpired maturity: Not exceeding 1 year
The difference between cash-flow revenue and cash-flow expenditure._y
Total foreign marketable debt
Cost/profit of revaluation of foreign loans.
Total outstanding foreign marketable bonds by unexpired maturity: Exceeding 3 years
Financing by non-marketable bonds
Total foreign non-marketable debt
Total outstanding domestic non-marketable bills
Total outstanding foreign marketable bonds by unexpired maturity: Exceeding 1 but not 3 years
Total outstanding domestic marketable debt
Net foreign assets after valuation adjustment
Gross gold and other foreign reserves : Amount as at end of period
Cash received from the issuance of Treasury bills less cash payments made to redeem Treasury bills.
National government revenue adjusted for differences in timing between the recording of transactions and bank clearances.
Total outstanding domestic non-marketable debt
Cash received from the issuance of other debt instruments less cash payments made to redeem other debt instruments, including non-marketable bills and bonds, and debt of the former TBVC-countries, self-governing territories and the Republic of Namibia that were assumed by the national government.
Total outstanding domestic non-marketable bonds
Total cash-flow revenue
Total outstanding domestic non-marketable loan levies
Including changes in the Exchequer and Paymaster-General Account balances with the South African Reserve Bank and other banks.
Total gross loan debt (nsa)
The net borrowing requirement of national government equals the net result of the cash-flow deficit/surplus and the cost of revaluation of foreign loans at redemption.
Total outstanding domestic marketable bills
Foreign exchange rate : SA rand per USA dollar
CPI Headline

1.0.2 Here I give an example of how to deal with missing data using Forward Fill approach. See https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.fillna.html

[14]:
```
#Data imputation strategy is foward fill but there many other options,
#choose based on your research needs and what gives higher accuracy
x_values_ffill = feature_set_sarb.fillna(method='ffill')
y_values_ffill = target.fillna(method='ffill')
```

[15]:
```
# Remove all data points before unemployment rate data is available.␣
 ↪Unemployment rate is my target variable.
valid_start = y_values_ffill.first_valid_index()
y_values_ffill = y_values_ffill[valid_start : ]
x_values_ffill = x_values_ffill[valid_start : ]
```

[16]:
```
#I fill with NA here to avoid any features that might be NA i.e. insurance
x_values_ffill = x_values_ffill.fillna(0)
x_values_ffill.isna().sum()
```

[16]:
```
Final consumption expenditure by general government               0
Consolidated general government: Revenue                           0
Foreign liabilities: Total portfolio investment                    0
Foreign liabilities: Portfolio investment: Equity securities       0
Domestic output: All groups                                        0
                                                                  ..
Gross domestic expenditure                                         0
Net cash-flow from operating activities                            0
Non-agricultural employment: Total                                 0
Consolidated general government: Expense                           0
Residual item                                                      0
Length: 147, dtype: int64
```

[17]:
```
x_values_ffill
```

[17]:
```
            Final consumption expenditure by general government   \
Date
1970-03-01                                           142014.0
1970-04-01                                           142014.0
1970-05-01                                           142014.0
1970-06-01                                           142014.0
1970-07-01                                           142014.0
…                                                       …
2019-09-01                                           653236.0
2019-10-01                                           653236.0
```

```
2019-11-01                                           653236.0
2019-12-01                                           653236.0
2020-01-01                                           653236.0

            Consolidated general government: Revenue  \
Date
1970-03-01                                       0.0
1970-04-01                                       0.0
1970-05-01                                       0.0
1970-06-01                                       0.0
1970-07-01                                       0.0
…                                              …
2019-09-01                                  462964.0
2019-10-01                                  462964.0
2019-11-01                                  462964.0
2019-12-01                                  462964.0
2020-01-01                                  462964.0

            Foreign liabilities: Total portfolio investment   \
Date
1970-03-01                                               2.0
1970-04-01                                               2.0
1970-05-01                                               2.0
1970-06-01                                               2.0
1970-07-01                                               2.0
…                                                      …
2019-09-01                                            3313.0
2019-10-01                                            3313.0
2019-11-01                                            3313.0
2019-12-01                                            3313.0
2020-01-01                                            3313.0

            Foreign liabilities: Portfolio investment: Equity securities  \
Date
1970-03-01                                                2.0
1970-04-01                                                2.0
1970-05-01                                                2.0
1970-06-01                                                2.0
1970-07-01                                                2.0
…                                                       …
2019-09-01                                             2036.0
2019-10-01                                             2036.0
2019-11-01                                             2036.0
2019-12-01                                             2036.0
2020-01-01                                             2036.0

            Domestic output: All groups   \
Date
1970-03-01                           4.1
1970-04-01                           4.2
1970-05-01                           4.2
1970-06-01                           4.2
1970-07-01                           4.2
…                                  …
2019-09-01                         114.3
2019-10-01                         114.6
2019-11-01                         114.3
2019-12-01                         114.5
2020-01-01                         114.5

            Final consumption expenditure by households: Total   \
Date
1970-03-01                                           459049.0
1970-04-01                                           459049.0
1970-05-01                                           459049.0
1970-06-01                                           459049.0
1970-07-01                                           459049.0
…                                                       …
2019-09-01                                          1961051.0
2019-10-01                                          1961051.0
2019-11-01                                          1961051.0
2019-12-01                                          1961051.0
2020-01-01                                          1961051.0

            Gross fixed capital formation   \
Date
1970-03-01                        176103.0
1970-04-01                        176103.0
1970-05-01                        176103.0
1970-06-01                        176103.0
1970-07-01                        176103.0
…                                    …
2019-09-01                        613640.0
2019-10-01                        613640.0
2019-11-01                        613640.0
2019-12-01                        613640.0
2020-01-01                        613640.0

            SDDS - Financial derivative liabilities  \
Date
1970-03-01                                      0.0
1970-04-01                                      0.0
1970-05-01                                      0.0
1970-06-01                                      0.0
```

```
1970-07-01                                      0.0
…                                             …
2019-09-01                                    118.0
2019-10-01                                    118.0
2019-11-01                                    118.0
2019-12-01                                    118.0
2020-01-01                                    118.0

            Foreign liabilities: Portfolio investment: Debt securities  \
Date
1970-03-01                                                      0.0
1970-04-01                                                      0.0
1970-05-01                                                      0.0
1970-06-01                                                      0.0
1970-07-01                                                      0.0
…                                                       …
2019-09-01                                                   1277.0
2019-10-01                                                   1277.0
2019-11-01                                                   1277.0
2019-12-01                                                   1277.0
2020-01-01                                                   1277.0

            Change in inventories   …  \
Date                                …
1970-03-01                 18617.0  …
1970-04-01                 18617.0  …
1970-05-01                 18617.0  …
1970-06-01                 18617.0  …
1970-07-01                 18617.0  …
…                            …  …
2019-09-01                 -9526.0  …
2019-10-01                 -9526.0  …
2019-11-01                 -9526.0  …
2019-12-01                 -9526.0  …
2020-01-01                 -9526.0  …

            Physical volume of manufacturing production: Total   \
Date
1970-03-01                                                 38.3
1970-04-01                                                 41.5
1970-05-01                                                 39.8
1970-06-01                                                 40.6
1970-07-01                                                 41.7
…                                                       …
2019-09-01                                                 99.0
2019-10-01                                                101.5
2019-11-01                                                100.0
```

```
2019-12-01                                              100.0
2020-01-01                                              100.0

            Remuneration per worker in non-agricultural: Total   \
Date                                                            
1970-03-01                                                0.9
1970-04-01                                                0.9
1970-05-01                                                0.9
1970-06-01                                                0.9
1970-07-01                                                0.9
…                                                       …
2019-09-01                                              173.4
2019-10-01                                              173.4
2019-11-01                                              173.4
2019-12-01                                              173.4
2020-01-01                                              173.4

            Consolidated general government: Non-financial assets - Net  \
Date                                                                     
1970-03-01                                                0.0
1970-04-01                                                0.0
1970-05-01                                                0.0
1970-06-01                                                0.0
1970-07-01                                                0.0
…                                                       …
2019-09-01                                           -26886.0
2019-10-01                                           -26886.0
2019-11-01                                           -26886.0
2019-12-01                                           -26886.0
2020-01-01                                           -26886.0

            Consolidated general government: Cash surplus / deficit  \
Date                                                                 
1970-03-01                                                0.0
1970-04-01                                                0.0
1970-05-01                                                0.0
1970-06-01                                                0.0
1970-07-01                                                0.0
…                                                       …
2019-09-01                                           -82087.0
2019-10-01                                           -82087.0
2019-11-01                                           -82087.0
2019-12-01                                           -82087.0
2020-01-01                                           -82087.0

            CPI Headline   Gross domestic expenditure   \
Date                                                      
```

```
1970-03-01            1.6                     965734.0
1970-04-01            1.6                     965734.0
1970-05-01            1.6                     965734.0
1970-06-01            1.6                     965734.0
1970-07-01            1.6                     965734.0
…                   …                          …
2019-09-01          113.4                    3222339.0
2019-10-01          113.4                    3222339.0
2019-11-01          113.5                    3222339.0
2019-12-01          113.8                    3222339.0
2020-01-01          113.8                    3222339.0

            Net cash-flow from operating activities   \
Date
1970-03-01                                       0.0
1970-04-01                                       0.0
1970-05-01                                       0.0
1970-06-01                                       0.0
1970-07-01                                       0.0
…                                              …
2019-09-01                                  -55201.0
2019-10-01                                  -55201.0
2019-11-01                                  -55201.0
2019-12-01                                  -55201.0
2020-01-01                                  -55201.0

            Non-agricultural employment: Total   \
Date
1970-03-01                                 49.2
1970-04-01                                 49.2
1970-05-01                                 49.2
1970-06-01                                 49.2
1970-07-01                                 49.2
…                                         …
2019-09-01                                106.8
2019-10-01                                106.8
2019-11-01                                106.8
2019-12-01                                106.8
2020-01-01                                106.8

            Consolidated general government: Expense  Residual item
Date
1970-03-01                                       0.0        169951.0
1970-04-01                                       0.0        169951.0
1970-05-01                                       0.0        169951.0
1970-06-01                                       0.0        169951.0
1970-07-01                                       0.0        169951.0
```

```
…                                                    …               …
2019-09-01                                    518165.0          3938.0
2019-10-01                                    518165.0          3938.0
2019-11-01                                    518165.0          3938.0
2019-12-01                                    518165.0          3938.0
2020-01-01                                    518165.0          3938.0

[794 rows x 147 columns]
```

2 Options for dealing with mixed frequencies and high dimensional data from SARB

1. Only use data with the same data frequency i.e. delete monthy data and only use quartely because unemployment is quartely OR delete all quarterly and use monthly but using sampling techniques / imputation to fill missing data
2. Use sampling techniques, upsample all monthly data to be quartely so that all your data is now in the same frequency
3. Use data imputation techniques, I gave an example above. You must use the same one for your features and target
4. Remove low variance and duplicate feature using statistical scores

These four should you leave you with the right data to model either the South African unemployment rate (which I suggest you use) or any other macroeconomic variable in the data.

Also remember your features and target are in different dataframes. Use pd.concat to merge these two i.e pd.concat([x_values_ffill, y_values_ffill], axis=1)

[ ]: