Revenue2

May 26, 2020

1 Revenue Forecasting

2 Real world Problems for revenue forcasting:

Many entrepreneurs complain that building forecasts with any degree of accuracy takes a lot of time–time that could be spent selling rather than planning. But few investors will put money in your business if you're unable to provide a set of thoughtful forecasts.More important, proper financial forecasts will help you develop operational and staffing plans that will help make your business a success.

3 Introduction:

Company fetches the financial data from various clients as individual excel sheets. These sheet are then clubbed as zip file in empirical model sheet . Financial Company calculates regression model using stastical way.BUt we can solve this problem using machine learning. # Objective: The objective is to forecast sales of revenues from empirical sheet based on past sales.

[1]:
```
#Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from datetime import datetime
from pandas import Series
import warnings
warnings.filterwarnings("ignore")
plt.style.use('fivethirtyeight')
import itertools
import pickle
```

[2]:
```
import matplotlib
matplotlib.rcParams['axes.labelsize']=14
matplotlib.rcParams['xtick.labelsize']=12
matplotlib.rcParams['ytick.labelsize']=12
matplotlib.rcParams['text.color']='k'
```

[11]:
```
# extraction
```

[15]:
```
from zipfile import ZipFile

filename='Health_datasets.zip'
with ZipFile(filename,'r') as zip:
    zip.printdir()
    zip.extractall()
    print('Done')
```

```
File Name                                             Modified             Size
Medmine_datasets/CRY.xlsx                      2019-07-13 11:26:30      6074072
Medmine_datasets/decentral.xlsx                2019-07-13 11:53:38         7258
Done
```

4 Data Preprocessing

[3]:
```
#Read excel file
df=pd.read_excel('CRY.xlsx',sheet_name='Empirical Model')
df
```

[3]:

	Unnamed: 0	Empirical Model - Total US	Unnamed: 2	Unnamed: 3	\
0	NaN	Back	NaN	NaN	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	Fiscal year ends December 31st	NaN	NaN	
4	NaN	NaN	NaN	Monthly Sales	
..	…	…	…	…	
194	NaN	NaN	9	September	
195	NaN	NaN	NaN	FOURTH QUARTER	
196	NaN	NaN	10	October	
197	NaN	NaN	11	November	
198	NaN	NaN	12	December	

	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	\
0	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	Quarterly sales	Monthly Contribution	Reported Sales	Growth Rate	
..	…	…	…	…	
194	0.313191	NaN	NaN	NaN	
195	NaN	NaN	NaN	NaN	
196	0.34037	NaN	NaN	NaN	
197	0.312412	NaN	NaN	NaN	

```
198         0.347218                    NaN             NaN          NaN

                Unnamed: 8 Unnamed: 9  … Unnamed: 795  Unnamed: 796  \
0                      NaN        NaN  …           NaN           NaN
1                      NaN        NaN  …           NaN           NaN
2                      NaN        NaN  …           NaN           NaN
3                      NaN        NaN  …           NaN           NaN
4    % Sales Captured in DB  # of facs  …          9955          9956
..                     …        …  …  …           …             …
194                    NaN        NaN  …           NaN           NaN
195                    NaN        NaN  …           NaN           NaN
196                    NaN        NaN  …           NaN           NaN
197                    NaN        NaN  …           NaN           NaN
198                    NaN        NaN  …           NaN           NaN

    Unnamed: 797 Unnamed: 798 Unnamed: 799 Unnamed: 800 Unnamed: 801  \
0            NaN          NaN          NaN          NaN          NaN
1            NaN          NaN          NaN          NaN          NaN
2            NaN          NaN          NaN          NaN          NaN
3            NaN          NaN          NaN          NaN          NaN
4          10015        10018        10085        10101        10103
..           …            …            …            …            …
194          NaN          NaN          NaN          NaN          NaN
195          NaN          NaN          NaN          NaN          NaN
196          NaN          NaN          NaN          NaN          NaN
197          NaN          NaN          NaN          NaN          NaN
198          NaN          NaN          NaN          NaN          NaN

    Unnamed: 802 Unnamed: 803  Unnamed: 804
0            NaN          NaN           NaN
1            NaN          NaN           NaN
2            NaN          NaN           NaN
3            NaN          NaN           NaN
4          10364        10633  Total Result
..           …            …             …
194          NaN          NaN           NaN
195          NaN          NaN           NaN
196          NaN          NaN           NaN
197          NaN          NaN           NaN
198          NaN          NaN           NaN

[199 rows x 805 columns]
```

[ ]:

[4]:
```
df=df.drop([0,1,2,3],axis=0)
df.head()
```

[4]:
```
   Unnamed: 0 Empirical Model - Total US  Unnamed: 2     Unnamed: 3  \
4         NaN                         NaN         NaN  Monthly Sales
5         NaN                        Q109  2009-01-01         239327
6         NaN                        Q109  2009-02-01         239655
7         NaN                        Q109  2009-03-01         338640
8         NaN                        Q209  2009-04-01         465898

        Unnamed: 4            Unnamed: 5      Unnamed: 6   Unnamed: 7  \
4  Quarterly sales  Monthly Contribution  Reported Sales  Growth Rate
5           817622              0.292711             NaN          NaN
6           817622              0.293112             NaN          NaN
7           817622              0.414177        22744000          NaN
8          1729262               0.26942             NaN          NaN

               Unnamed: 8 Unnamed: 9  … Unnamed: 795  Unnamed: 796  \
4  % Sales Captured in DB  # of facs  …          9955          9956
5                     NaN         -1  …           NaN           NaN
6                     NaN         -1  …           NaN           NaN
7               0.0359489         -1  …           NaN           NaN
8                     NaN         -1  …           NaN           NaN

  Unnamed: 797 Unnamed: 798 Unnamed: 799 Unnamed: 800 Unnamed: 801  \
4        10015        10018        10085        10101        10103
5          NaN          NaN          NaN          NaN          NaN
6          NaN          NaN          NaN          NaN          NaN
7          NaN          NaN          NaN          NaN          NaN
8          NaN          NaN          NaN          NaN          NaN

  Unnamed: 802 Unnamed: 803  Unnamed: 804
4        10364        10633  Total Result
5          NaN          NaN        239327
6          NaN          NaN        239655
7          NaN          NaN        338640
8          NaN          NaN        465898

[5 rows x 805 columns]
```

[5]:
```
df.head(10)
```

[5]:
```
   Unnamed: 0 Empirical Model - Total US  Unnamed: 2     Unnamed: 3  \
4         NaN                         NaN         NaN  Monthly Sales
5         NaN                        Q109  2009-01-01         239327
6         NaN                        Q109  2009-02-01         239655
7         NaN                        Q109  2009-03-01         338640
8         NaN                        Q209  2009-04-01         465898
9         NaN                        Q209  2009-05-01         531263
10        NaN                        Q209  2009-06-01         732101
```

```
11         NaN                      Q309  2009-07-01         772985
12         NaN                      Q309  2009-08-01         759454
13         NaN                      Q309  2009-09-01         886120

         Unnamed: 4            Unnamed: 5      Unnamed: 6   Unnamed: 7  \
4   Quarterly sales  Monthly Contribution  Reported Sales  Growth Rate
5            817622              0.292711             NaN          NaN
6            817622              0.293112             NaN          NaN
7            817622              0.414177        22744000          NaN
8           1729262               0.26942             NaN          NaN
9           1729262              0.307219             NaN          NaN
10          1729262               0.42336        23579000          NaN
11      2.41856e+06              0.319606             NaN          NaN
12      2.41856e+06              0.314011             NaN          NaN
13      2.41856e+06              0.366383        23941000          NaN

                Unnamed: 8 Unnamed: 9  … Unnamed: 795  Unnamed: 796  \
4   % Sales Captured in DB  # of facs  …         9955          9956
5                      NaN         -1  …          NaN           NaN
6                      NaN         -1  …          NaN           NaN
7                0.0359489         -1  …          NaN           NaN
8                      NaN         -1  …          NaN           NaN
9                      NaN         -1  …          NaN           NaN
10               0.0733391         -1  …          NaN           NaN
11                     NaN         -1  …          NaN           NaN
12                     NaN         -1  …          NaN           NaN
13                0.101022         -1  …          NaN           NaN

   Unnamed: 797 Unnamed: 798 Unnamed: 799 Unnamed: 800 Unnamed: 801  \
4         10015        10018        10085        10101        10103
5           NaN          NaN          NaN          NaN          NaN
6           NaN          NaN          NaN          NaN          NaN
7           NaN          NaN          NaN          NaN          NaN
8           NaN          NaN          NaN          NaN          NaN
9           NaN          NaN          NaN          NaN          NaN
10          NaN          NaN          NaN          NaN          NaN
11          NaN          NaN          NaN          NaN          NaN
12          NaN          NaN          NaN          NaN          NaN
13          NaN          NaN          NaN          NaN          NaN

   Unnamed: 802 Unnamed: 803  Unnamed: 804
4         10364        10633  Total Result
5           NaN          NaN        239327
6           NaN          NaN        239655
7           NaN          NaN        338640
8           NaN          NaN        465898
9           NaN          NaN        531263
```

```
10          NaN          NaN        732101
11          NaN          NaN        772985
12          NaN          NaN        759454
13          NaN          NaN        886120

[10 rows x 805 columns]
```

[6]:
```
df=df.drop(['Unnamed: 0'],axis=1)
```

[7]:
```
df.loc[4,'Unnamed: 2']='DateTime'
```

[8]:
```
df.head()
```

[8]:
```
  Empirical Model - Total US  Unnamed: 2     Unnamed: 3       Unnamed: 4  \
4                        NaN    DateTime  Monthly Sales  Quarterly sales
5                       Q109  2009-01-01         239327           817622
6                       Q109  2009-02-01         239655           817622
7                       Q109  2009-03-01         338640           817622
8                       Q209  2009-04-01         465898          1729262

             Unnamed: 5      Unnamed: 6   Unnamed: 7              Unnamed: 8  \
4  Monthly Contribution  Reported Sales  Growth Rate  % Sales Captured in DB
5              0.292711             NaN          NaN                     NaN
6              0.293112             NaN          NaN                     NaN
7              0.414177        22744000          NaN               0.0359489
8               0.26942             NaN          NaN                     NaN

  Unnamed: 9 Unnamed: 10  …  Unnamed: 795 Unnamed: 796 Unnamed: 797  \
4  # of facs   Sales/fac  …          9955         9956        10015
5         -1     -239327  …           NaN          NaN          NaN
6         -1     -239655  …           NaN          NaN          NaN
7         -1     -338640  …           NaN          NaN          NaN
8         -1     -465898  …           NaN          NaN          NaN

  Unnamed: 798 Unnamed: 799 Unnamed: 800 Unnamed: 801 Unnamed: 802  \
4        10018        10085        10101        10103        10364
5          NaN          NaN          NaN          NaN          NaN
6          NaN          NaN          NaN          NaN          NaN
7          NaN          NaN          NaN          NaN          NaN
8          NaN          NaN          NaN          NaN          NaN

  Unnamed: 803  Unnamed: 804
4        10633  Total Result
5          NaN        239327
6          NaN        239655
7          NaN        338640
8          NaN        465898
```

```
[5 rows x 804 columns]
```

[9]:
```
df=df.reset_index()
df.head(2)
```

[9]:
```
   index Empirical Model - Total US  Unnamed: 2     Unnamed: 3  \
0      4                        NaN    DateTime  Monthly Sales
1      5                       Q109  2009-01-01         239327

        Unnamed: 4            Unnamed: 5      Unnamed: 6   Unnamed: 7  \
0  Quarterly sales  Monthly Contribution  Reported Sales  Growth Rate
1           817622              0.292711             NaN          NaN

               Unnamed: 8 Unnamed: 9  … Unnamed: 795  Unnamed: 796  \
0  % Sales Captured in DB  # of facs  …         9955          9956
1                     NaN         -1  …          NaN           NaN

  Unnamed: 797 Unnamed: 798 Unnamed: 799 Unnamed: 800 Unnamed: 801  \
0        10015        10018        10085        10101        10103
1          NaN          NaN          NaN          NaN          NaN

  Unnamed: 802 Unnamed: 803  Unnamed: 804
0        10364        10633  Total Result
1          NaN          NaN        239327

[2 rows x 805 columns]
```

[ ]:
```
```

[10]:
```
df.columns=df.iloc[0]
```

[11]:
```
df.head()
```

[11]:
```
0  4   NaN    DateTime  Monthly Sales  Quarterly sales  Monthly Contribution  \
0  4   NaN    DateTime  Monthly Sales  Quarterly sales  Monthly Contribution
1  5  Q109  2009-01-01         239327           817622              0.292711
2  6  Q109  2009-02-01         239655           817622              0.293112
3  7  Q109  2009-03-01         338640           817622              0.414177
4  8  Q209  2009-04-01         465898          1729262               0.26942

0  Reported Sales  Growth Rate  % Sales Captured in DB  # of facs  …  9955  \
0  Reported Sales  Growth Rate  % Sales Captured in DB  # of facs  …  9955
1             NaN          NaN                     NaN         -1  …   NaN
2             NaN          NaN                     NaN         -1  …   NaN
3        22744000          NaN               0.0359489         -1  …   NaN
4             NaN          NaN                     NaN         -1  …   NaN
```

```
0  9956  10015  10018  10085  10101  10103  10364  10633  Total Result
0  9956  10015  10018  10085  10101  10103  10364  10633  Total Result
1   NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN        239327
2   NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN        239655
3   NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN        338640
4   NaN    NaN    NaN    NaN    NaN    NaN    NaN    NaN        465898

[5 rows x 805 columns]
```

[12]:
```
df=df.reindex(df.index.drop(0))
```

[13]:
```
df=df.drop([4],axis=1)
df.head()
```

[13]:
```
0   NaN    DateTime Monthly Sales Quarterly sales Monthly Contribution  \
1  Q109  2009-01-01        239327          817622             0.292711
2  Q109  2009-02-01        239655          817622             0.293112
3  Q109  2009-03-01        338640          817622             0.414177
4  Q209  2009-04-01        465898         1729262              0.26942
5  Q209  2009-05-01        531263         1729262             0.307219

0 Reported Sales Growth Rate % Sales Captured in DB # of facs Sales/fac  …  \
1                        NaN                  NaN        -1   -239327  …
2                        NaN                  NaN        -1   -239655  …
3                   22744000                  NaN        -1   -338640  …
4                        NaN                  NaN        -1   -465898  …
5                        NaN                  NaN        -1   -531263  …

0  9955 9956 10015 10018 10085 10101 10103 10364 10633 Total Result
1   NaN  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN        239327
2   NaN  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN        239655
3   NaN  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN        338640
4   NaN  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN        465898
5   NaN  NaN   NaN   NaN   NaN   NaN   NaN   NaN   NaN        531263

[5 rows x 804 columns]
```

[14]:
```
final_df=df.iloc[:,1:10]
```

[15]:
```
final_df.head()
```

[15]:
```
0    DateTime Monthly Sales Quarterly sales Monthly Contribution  \
1  2009-01-01        239327          817622             0.292711
2  2009-02-01        239655          817622             0.293112
3  2009-03-01        338640          817622             0.414177
4  2009-04-01        465898         1729262              0.26942
```

```
5  2009-05-01        531263         1729262             0.307219

0 Reported Sales Growth Rate % Sales Captured in DB # of facs Sales/fac
1            NaN         NaN                    NaN        -1   -239327
2            NaN         NaN                    NaN        -1   -239655
3       22744000         NaN              0.0359489        -1   -338640
4            NaN         NaN                    NaN        -1   -465898
5            NaN         NaN                    NaN        -1   -531263
```

[16]:
```
monthly_df=final_df.iloc[:,:2]
```

[17]:
```
monthly_df.head()
```

[17]:
```
0    DateTime Monthly Sales
1  2009-01-01        239327
2  2009-02-01        239655
3  2009-03-01        338640
4  2009-04-01        465898
5  2009-05-01        531263
```

[18]:
```
monthly_df['DateTime']=pd.to_datetime(monthly_df['DateTime'])
```

[19]:
```
monthly_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 194 entries, 1 to 194
Data columns (total 2 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   DateTime       130 non-null    datetime64[ns]
 1   Monthly Sales  142 non-null    object
dtypes: datetime64[ns](1), object(1)
memory usage: 4.5+ KB
```

[20]:
```
monthly_df.set_index('DateTime',inplace=True)
```

monthly_df

[21]:
```
len(monthly_df)
```

[21]:
```
194
```

[22]:
```
monthly_df.tail(77)
```

[22]:
```
0                              Monthly Sales
DateTime
2018-10-01 00:00:00.000000000     1.58714e+06
```

```
NaT                                          NaN
NaT                                          NaN
NaT                                          NaN
NaT                                          NaN
…                                            …
1970-01-01 00:00:00.000000009       September
NaT                                FOURTH QUARTER
1970-01-01 00:00:00.000000010         October
1970-01-01 00:00:00.000000011       November
1970-01-01 00:00:00.000000012        December

[77 rows x 1 columns]
```

[23]:
```
monthly_df.isnull().sum()
```

[23]:
```
0
Monthly Sales    52
dtype: int64
```

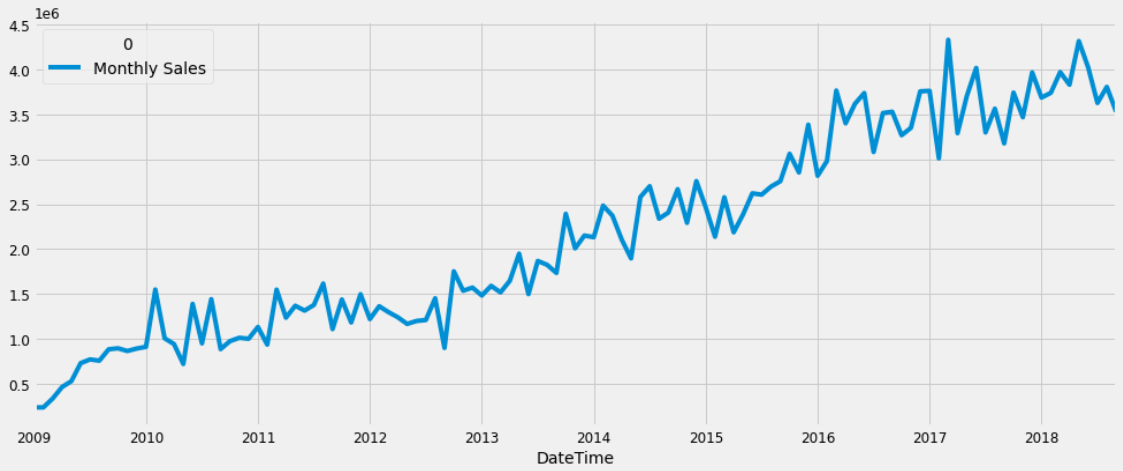
[46]:
```
indexedDataset=monthly_df[:117]
```

[47]:
```
indexedDataset.tail(10)
```

[47]:
```
0          Monthly Sales
DateTime
2017-12-01   3.96596e+06
2018-01-01   3.68591e+06
2018-02-01   3.73894e+06
2018-03-01   3.96962e+06
2018-04-01   3.82884e+06
2018-05-01   4.31492e+06
2018-06-01   4.02291e+06
2018-07-01   3.62582e+06
2018-08-01   3.80472e+06
2018-09-01    3.5292e+06
```

[48]:
```
#nana
indexedDataset=indexedDataset.fillna(method='ffill')
```

[49]:
```
indexedDataset.plot(figsize=(15,6))
plt.show()
```



5 Stationarity test

[50]:
```
from statsmodels.tsa.stattools import adfuller
```

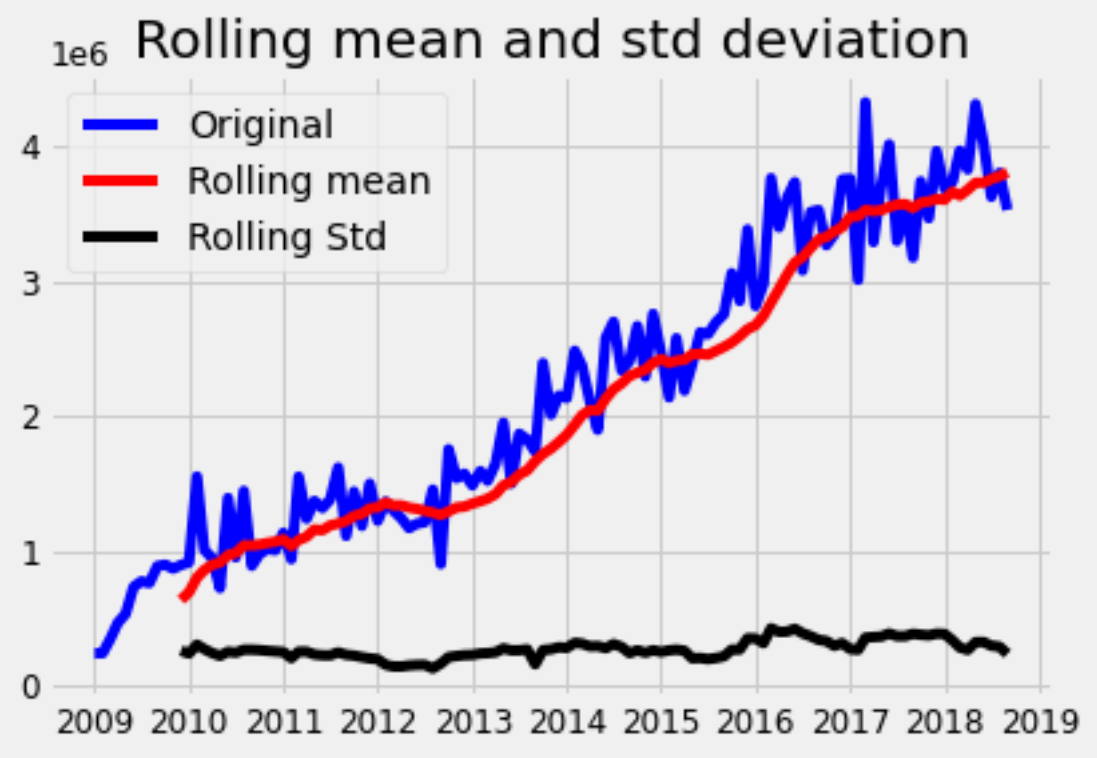
[51]:
```
def test_stationarity(df):
    movingAverage=df.rolling(window=12).mean()
    movingSTD=df.rolling(window=12).std()
    #plotting
    orin=plt.plot(df,color='blue',label='Original')
    mean=plt.plot(movingAverage,color='red',label='Rolling mean')
    std=plt.plot(movingSTD,color='black',label='Rolling Std')
    plt.legend(loc='best')
    plt.title('Rolling mean and std deviation')
    plt.show(block=False)

    # ADF
    dftest=adfuller(df['Monthly Sales'],autolag='AIC')
    print(dftest)
    dfoutput=pd.Series(dftest[0:4],index=['Test␣
 ↪Statistic','p-value','#LagsUsed',
                                          'Number of Observation Used'])
    for key,value in dftest[4].items():
        dfoutput['Critical (%s)'%key]=value
    print(dfoutput)
```

[52]:
```
type(indexedDataset)
```

[52]:
```
pandas.core.frame.DataFrame
```

[53]:
```
test_stationarity(indexedDataset)
```




```
(-0.2595805213451949, 0.9310370589625145, 11, 105, {'1%': -3.4942202045135513,
'5%': -2.889485291005291, '10%': -2.5816762131519275}, 2874.515411115559)
Test Statistic                  -0.259581
p-value                          0.931037
#LagsUsed                       11.000000
Number of Observation Used     105.000000
Critical (1%)                   -3.494220
Critical (5%)                   -2.889485
Critical (10%)                  -2.581676
dtype: float64
```

[54]:
```
# 0.05 dea
```

[56]:
```
"1. Model to apply time series 2. Pattern and plot pattern  3. Forescast our␣
↪result and plot "
```

[56]:
```
'1. Model to apply time series 2. Pattern and plot pattern  3. Forescast our
result and plot '
```

[ ]:
```

```