

Revenue1

May 25, 2020

1 Revenue Forecasting

2 Real world Problems for revenue forecasting:

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.

```
[6]: #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
```

```
[8]: 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

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

```
[17]: 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]

```
[ ]:
```

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

```

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

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

```

[21]: Empirical Model - Total US Unnamed: 2 Unnamed: 3 Unnamed: 4 \
4      NaN NaN NaN 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
9      Q209 2009-05-01 531263 1729262
10     Q209 2009-06-01 732101 1729262

```

11	Q309	2009-07-01	772985	2.41856e+06
12	Q309	2009-08-01	759454	2.41856e+06
13	Q309	2009-09-01	886120	2.41856e+06

	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
9	0.307219	NaN	NaN	NaN
10	0.42336	23579000	NaN	0.0733391
11	0.319606	NaN	NaN	NaN
12	0.314011	NaN	NaN	NaN
13	0.366383	23941000	NaN	0.101022

	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
9	-1	-531263	...	NaN	NaN	NaN
10	-1	-732101	...	NaN	NaN	NaN
11	-1	-772985	...	NaN	NaN	NaN
12	-1	-759454	...	NaN	NaN	NaN
13	-1	-886120	...	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
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: 803	Unnamed: 804
4	10633	Total Result
5	NaN	239327
6	NaN	239655
7	NaN	338640
8	NaN	465898
9	NaN	531263

```

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

```

[10 rows x 804 columns]

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

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

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

```

[23]:  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]

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

```
[24]:      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]

[]:

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

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

```
[29]: 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]

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

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

```
[36]: 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 NaN NaN -1 -239327 ...
2 NaN NaN NaN NaN -1 -239655 ...
3 22744000 NaN 0.0359489 -1 -338640 ...
4 NaN NaN NaN NaN -1 -465898 ...
5 NaN NaN NaN NaN -1 -531263 ...

```

```

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

```

[5 rows x 805 columns]

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

```
[37]: 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
.. .. .. .. ..
```



```

190          9      September      0.313191      NaN
191         NaN FOURTH QUARTER      NaN      NaN
192         10      October        0.34037      NaN
193         11      November       0.312412      NaN
194         12      December       0.347218      NaN

```

```

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
..         ...      ...      ...      ...
190         NaN      NaN      NaN      NaN      NaN
191         NaN      NaN      NaN      NaN      NaN
192         NaN      NaN      NaN      NaN      NaN
193         NaN      NaN      NaN      NaN      NaN
194         NaN      NaN      NaN      NaN      NaN

```

[194 rows x 9 columns]

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

```

[38]: 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

```

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

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

```

[49]: 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

```

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

```
[52]: 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
```

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

```
[ ]:
```