

The package used below will have to be explicitly imported

Package name: TSForecasting

All the packages mentioned below will be installed automatically installed in case they are missing :

- pandas
- numpy
- matplotlib
- dateutil
- re

The pandas package that is explicitly imported will be used to read the data supplied by you

```
In [1]: from TSForecasting.TSForecasting import TimeSeriesForecast
import pandas as pd
```

Using TensorFlow backend.

=====

To read the documentation of the class, just run the following code

```
In [2]: TimeSeriesForecast?
```

Now, we create an object of the TimeSeriesForecast class

This is object which has the model and methods to predict the imputed values

The object has a pretrained model on the data described below. The model is capable of performing imputations

```
In [3]: obj = TimeSeriesForecast()
```

```
Epoch 1/1
29188/29188 [=====] - 66s 2ms/step - loss: 0.0013
```

To read the documentation of any method:

1. Place the cursor inside the parenthesis of the function, hold down shift and press tab
2. Run the following code (substitute 'impute' with the name of the method)

```
In [4]: # obj.impute?
```

The following method describes the **CGM data** used to pre-train the model

```
In [5]: obj.dataDescribe()
```

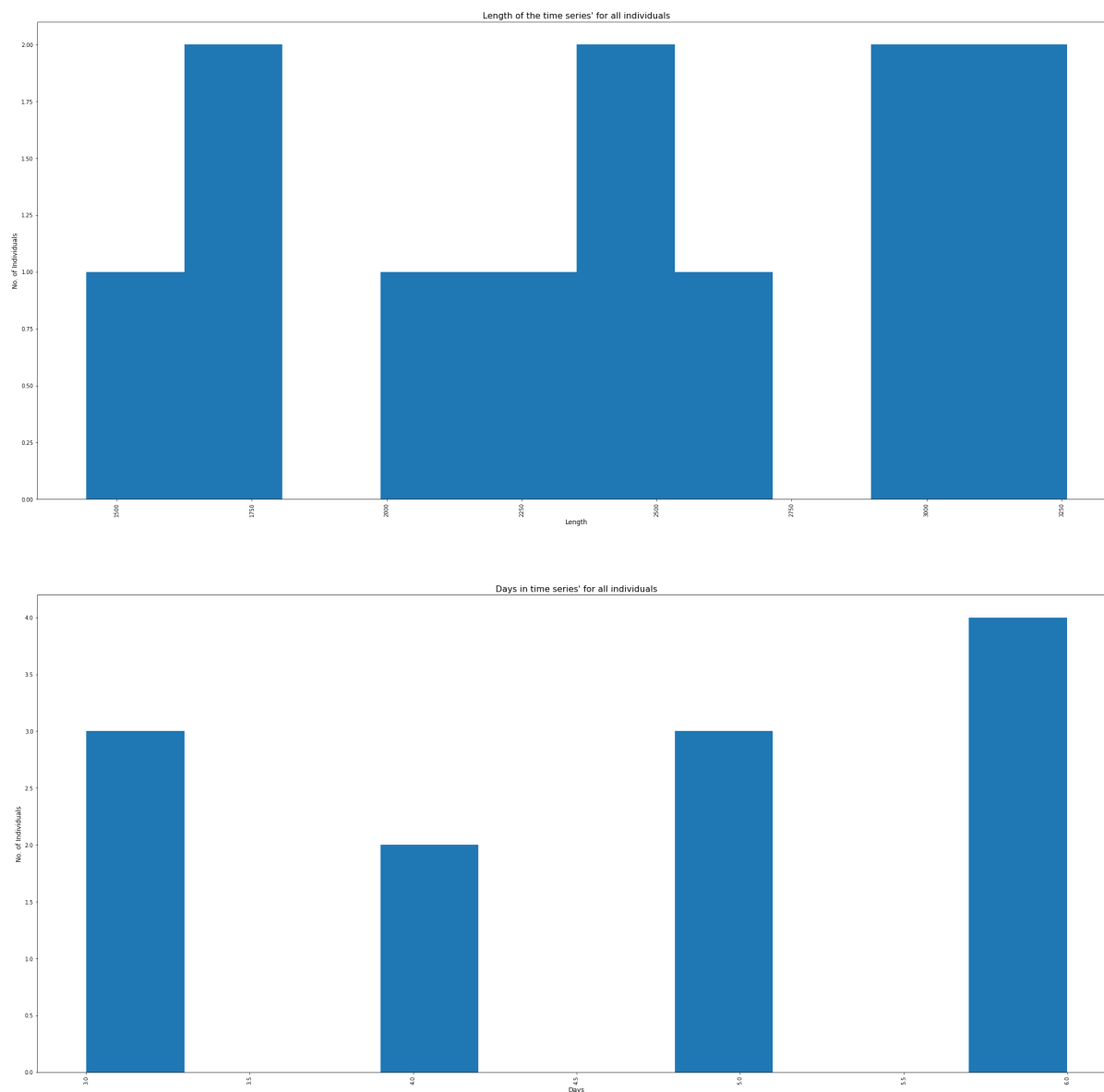
Here is a glimpse of the data:

	Display Time	subjectId	GlucoseValue
0	2016-06-25 09:14:00	ID01	108.000
1	2016-06-25 09:17:00	ID01	108.000
2	2016-06-25 09:20:00	ID01	108.000
3	2016-06-25 09:23:00	ID01	107.658
4	2016-06-25 09:26:00	ID01	107.496

	Length of readings	Max. Glucose Value	Min. Glucose Value	Days
count	12.000000	12.000000	12.000000	12
mean	2432.416667	256.847500	65.063333	5 days 01:34:15
std	605.555711	97.372823	24.159458	1 days 06:16:40.028065
min	1443.000000	170.390000	14.020000	3 days 00:06:00
25%	2015.250000	185.947500	58.397500	4 days 04:42:45
50%	2491.000000	230.130000	65.500000	5 days 04:30:00
75%	2989.000000	297.727500	79.762500	6 days 05:24:00
max	3260.000000	508.710000	95.400000	6 days 18:57:00

Here is the statistical analysis of the data:

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	MAGE Score	Days	Start	End
0	ID31	HEALTHY	1443	182.19	66.47	-1.0	3 days 00:06:00	2016-06-06 16:55:00	2016-06-09 17:01:00
1	ID30	HEALTHY	2472	187.20	59.40	-1.0	5 days 03:33:00	2016-12-21 07:29:00	2016-12-26 11:02:00
2	ID29	HEALTHY	1728	195.64	75.15	-1.0	3 days 14:21:00	2016-12-19 22:42:00	2016-12-23 13:03:00
3	ID23	PRE-DIABETIC	3260	326.83	64.53	-1.0	6 days 18:57:00	2016-10-11 15:34:00	2016-10-18 10:31:00
4	ID22	PRE-DIABETIC	2111	170.39	34.20	-1.0	4 days 09:30:00	2016-12-29 07:35:00	2017-01-02 17:05:00
5	ID21	PRE-DIABETIC	2686	214.20	66.60	-1.0	5 days 14:15:00	2016-11-22 07:26:00	2016-11-27 21:41:00
6	ID13	DIABETIC-2	1681	246.06	94.75	-1.0	3 days 12:00:00	2016-12-01 07:44:00	2016-12-04 19:44:00
7	ID12	DIABETIC-2	2980	271.80	93.60	-1.0	6 days 04:57:00	2016-08-02 08:19:00	2016-08-08 13:16:00
8	ID11	DIABETIC-2	2205	316.80	95.40	-1.0	4 days 14:12:00	2017-01-07 08:52:00	2017-01-11 23:04:00
9	ID03	DIABETIC-1	3097	508.71	61.25	-1.0	6 days 10:48:00	2016-05-26 06:09:00	2016-06-01 16:57:00
10	ID02	DIABETIC-1	3016	170.98	14.02	-1.0	6 days 06:45:00	2016-05-26 07:50:00	2016-06-01 14:35:00
11	ID01	DIABETIC-1	2510	291.37	55.39	-1.0	5 days 05:27:00	2016-06-25 09:14:00	2016-06-30 14:41:00

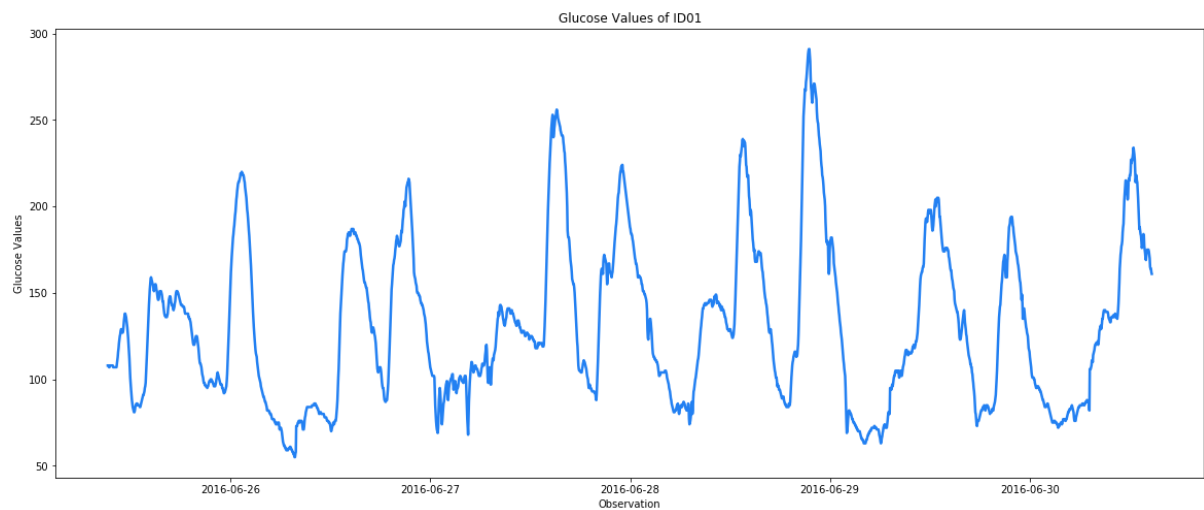


To see what a specific individual's glucose values look like, run the following line.

In between the quotes, pass the Subject ID of that individual

The different subject ID's can be found in the description table above

```
In [6]: obj.plotSpecific('ID01')
```



To test the pretrained model, write the path of the testing dataset below:

```
In [7]: testing_data = pd.read_csv("~/Desktop/NCSA_genomics/Python - notebooks/TSForecasting/Data/test/generated_test.txt", sep=",")
```

Now, we run the imputation model on the data you supplied above

The graph you see is the imputed glucose values

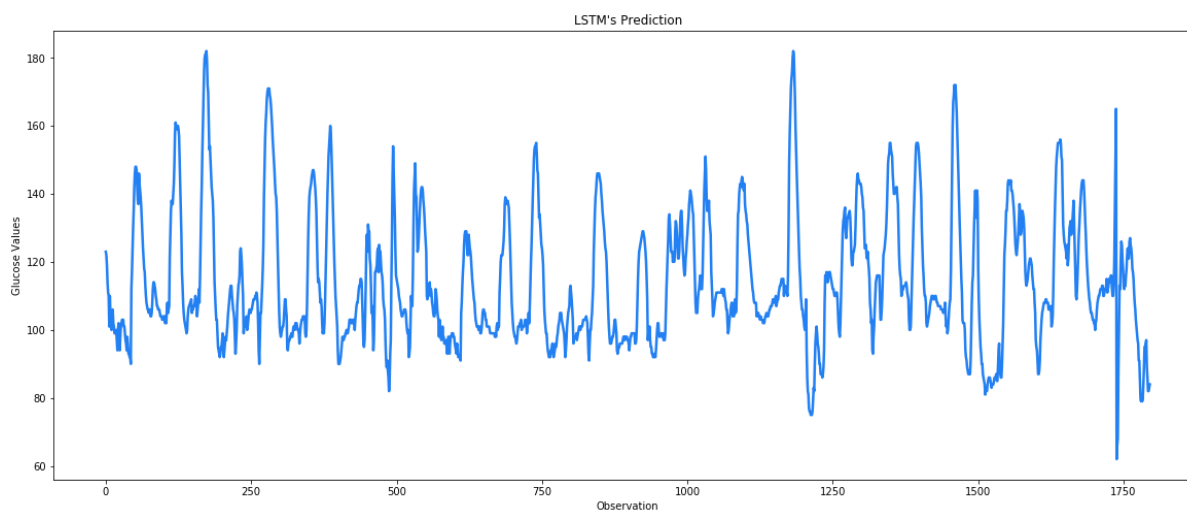
```
In [8]: obj.impute(testing_data)
```

Gap detected!

566/566 [=====] - 1s 1ms/step

Imputations performed!

File saved!



In case you wish to train your own model, use the following code

Write the path to your file for the training and test set

Be careful of the following:

1. Enter the path of your file
2. The file should have only two columns in the following order with the exact names: Timestamp, GlucoseValue
3. The date of the timestamp should have the entire year (all 4 digits). Make sure the data is numeric (irrespective of the delimiter)

```
In [9]: training_data = pd.read_csv("~/Desktop/NCSA_genomics/Python - notebooks/  
TSForecasting/Data/CGM/CGM_Analyzer_Appended.csv")
```



```
In [10]: obj.dataDescribe(training_data)
```

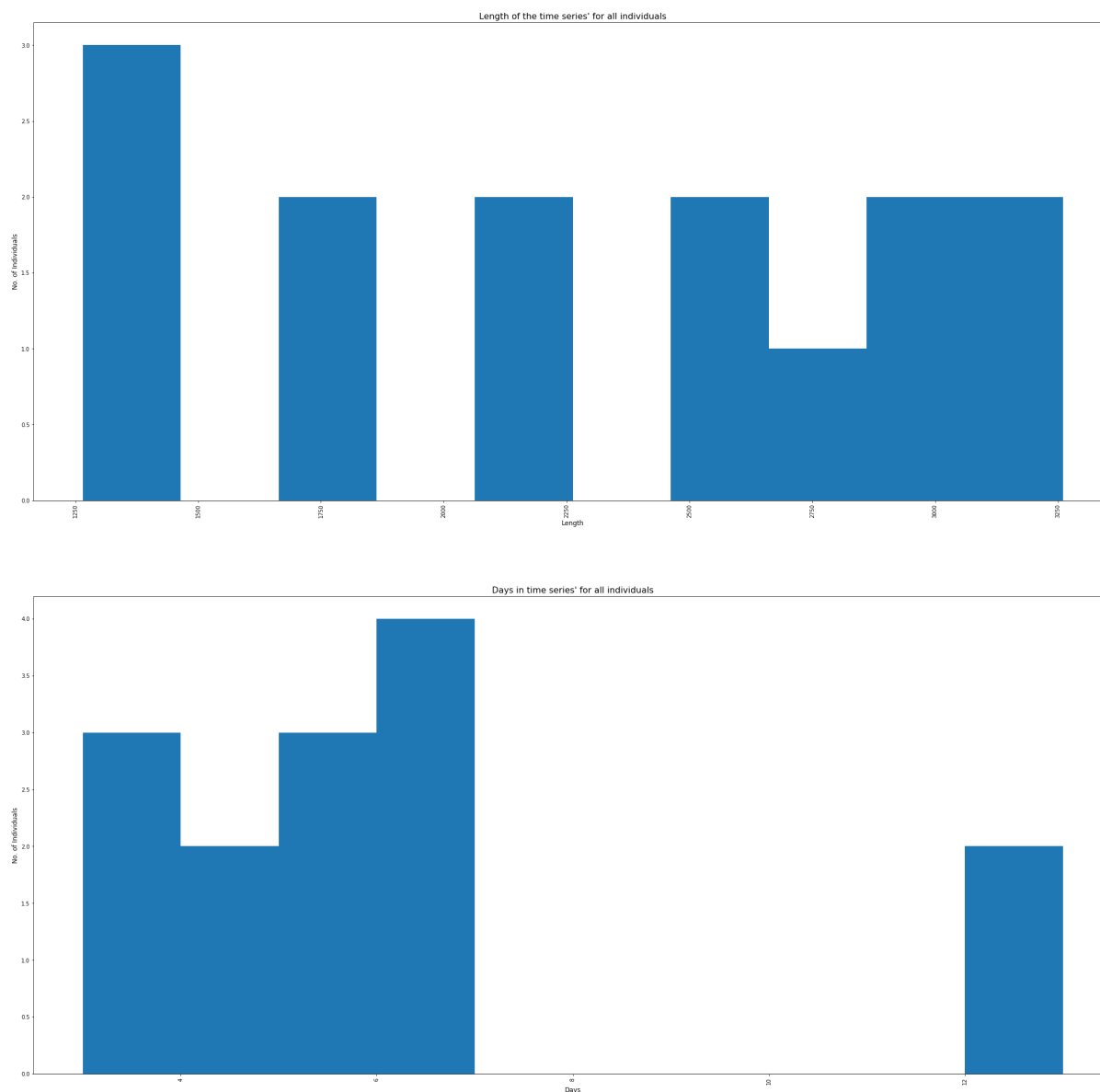
Here is a glimpse of the data:

	Display Time	subjectId	GlucoseValue
0	2016-06-25 09:14:00	ID01	108.000
1	2016-06-25 09:17:00	ID01	108.000
2	2016-06-25 09:20:00	ID01	108.000
3	2016-06-25 09:23:00	ID01	107.658
4	2016-06-25 09:26:00	ID01	107.496

	Length of readings	Max. Glucose Value	Min. Glucose Value	Days
count	14.000000	14.000000	14.000000	14
mean	2270.785714	246.283571	61.954286	6 days 07:28:30
std	692.305533	93.808260	23.631486	3 days 08:57:56.850301
min	1265.000000	163.800000	14.020000	3 days 00:06:00
25%	1692.750000	183.442500	49.097500	4 days 10:40:30
50%	2338.500000	208.100000	62.890000	5 days 09:51:00
75%	2906.500000	286.477500	73.012500	6 days 09:47:15
max	3260.000000	508.710000	95.400000	13 days 21:59:00

Here is the statistical analysis of the data:

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	MAGE Score	Days	Start	End
0	ID33	NAN	1265	163.80	39.60	-1.0	13 days 15:49:00	2019-02-21 14:59:00	2019-03-07 06:48:00
1	ID32	NAN	1337	202.00	47.00	-1.0	13 days 21:59:00	2018-08-01 12:00:00	2018-08-15 09:59:00
2	ID31	HEALTHY	1443	182.19	66.47	-1.0	3 days 00:06:00	2016-06-06 16:55:00	2016-06-09 17:01:00
3	ID30	HEALTHY	2472	187.20	59.40	-1.0	5 days 03:33:00	2016-12-21 07:29:00	2016-12-26 11:02:00
4	ID29	HEALTHY	1728	195.64	75.15	-1.0	3 days 14:21:00	2016-12-19 22:42:00	2016-12-23 13:03:00
5	ID23	PRE-DIABETIC	3260	326.83	64.53	-1.0	6 days 18:57:00	2016-10-11 15:34:00	2016-10-18 10:31:00
6	ID22	PRE-DIABETIC	2111	170.39	34.20	-1.0	4 days 09:30:00	2016-12-29 07:35:00	2017-01-02 17:05:00
7	ID21	PRE-DIABETIC	2686	214.20	66.60	-1.0	5 days 14:15:00	2016-11-22 07:26:00	2016-11-27 21:41:00
8	ID13	DIABETIC-2	1681	246.06	94.75	-1.0	3 days 12:00:00	2016-12-01 07:44:00	2016-12-04 19:44:00
9	ID12	DIABETIC-2	2980	271.80	93.60	-1.0	6 days 04:57:00	2016-08-02 08:19:00	2016-08-08 13:16:00
10	ID11	DIABETIC-2	2205	316.80	95.40	-1.0	4 days 14:12:00	2017-01-07 08:52:00	2017-01-11 23:04:00
11	ID03	DIABETIC-1	3097	508.71	61.25	-1.0	6 days 10:48:00	2016-05-26 06:09:00	2016-06-01 16:57:00
12	ID02	DIABETIC-1	3016	170.98	14.02	-1.0	6 days 06:45:00	2016-05-26 07:50:00	2016-06-01 14:35:00
13	ID01	DIABETIC-1	2510	291.37	55.39	-1.0	5 days 05:27:00	2016-06-25 09:14:00	2016-06-30 14:41:00



Use the following function to train the model on the data you just supplied

```
In [11]: obj.train(training_data)
```

```
Train shape: (31790, 1, 1)
```

```
Epoch 1/1
```

```
31790/31790 [=====] - 71s 2ms/step - loss: 3.5  
978e-04
```

To test the model you just trained, write the path of the testing dataset below:

```
In [12]: testing_data = pd.read_csv("~/Desktop/NCSA_genomics/Python - notebooks/TSForecasting/Data/test/generated_test.txt", sep=",")
```

Now, we run the imputation model on the data you supplied above

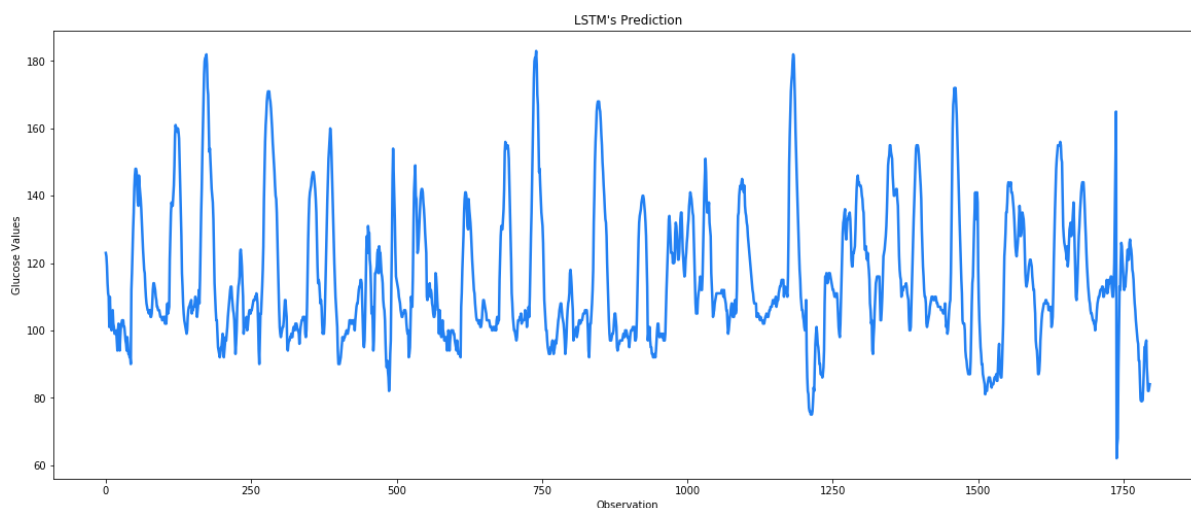
```
In [13]: obj.impute(testing_data)
```

Gap detected!

566/566 [=====] - 1s 918us/step

Imputations performed!

File saved!



Now, lets training the model on the HALL Dataset

```
In [14]: training_data = pd.read_csv("~/Desktop/NCSA_genomics/Python - notebooks/  
TSForecasting/Data/Hall/Hall_data.csv")
```

The following method describes the **raw HALL data** before processing

The data needs to be processed before we use it to train our imputation model

```
In [15]: obj.rawData()
```

Here is a glimpse of the data:

	Display Time	GlucoseValue	subjectId
0	2014-02-03 03:42:12	93	1636-69-001
1	2014-02-03 03:47:12	93	1636-69-001
2	2014-02-03 03:52:12	93	1636-69-001
3	2014-02-03 03:57:12	95	1636-69-001
4	2014-02-03 04:02:12	96	1636-69-001

	Length of readings	Gapsize	Missing Values	Percent of missing values	Days
count	57.000000	57	57.000000	57.000000	57
mean	1849.578947	21 days 22:01:48.929824	27.473684	1.427018	29 days 02:20:02.333333
std	104.808149	73 days 09:22:16.337099	32.057485	1.504522	74 days 06:47:55.221313
min	1584.000000	0 days 00:10:00	1.000000	0.060000	5 days 16:14:32
25%	1796.000000	0 days 01:15:00	8.000000	0.440000	6 days 07:39:25
50%	1822.000000	0 days 04:25:00	20.000000	1.140000	6 days 23:39:38
75%	1877.000000	0 days 19:09:56	31.000000	1.690000	8 days 07:49:13
max	2361.000000	416 days 17:21:34	176.000000	7.450000	423 days 11:25:54

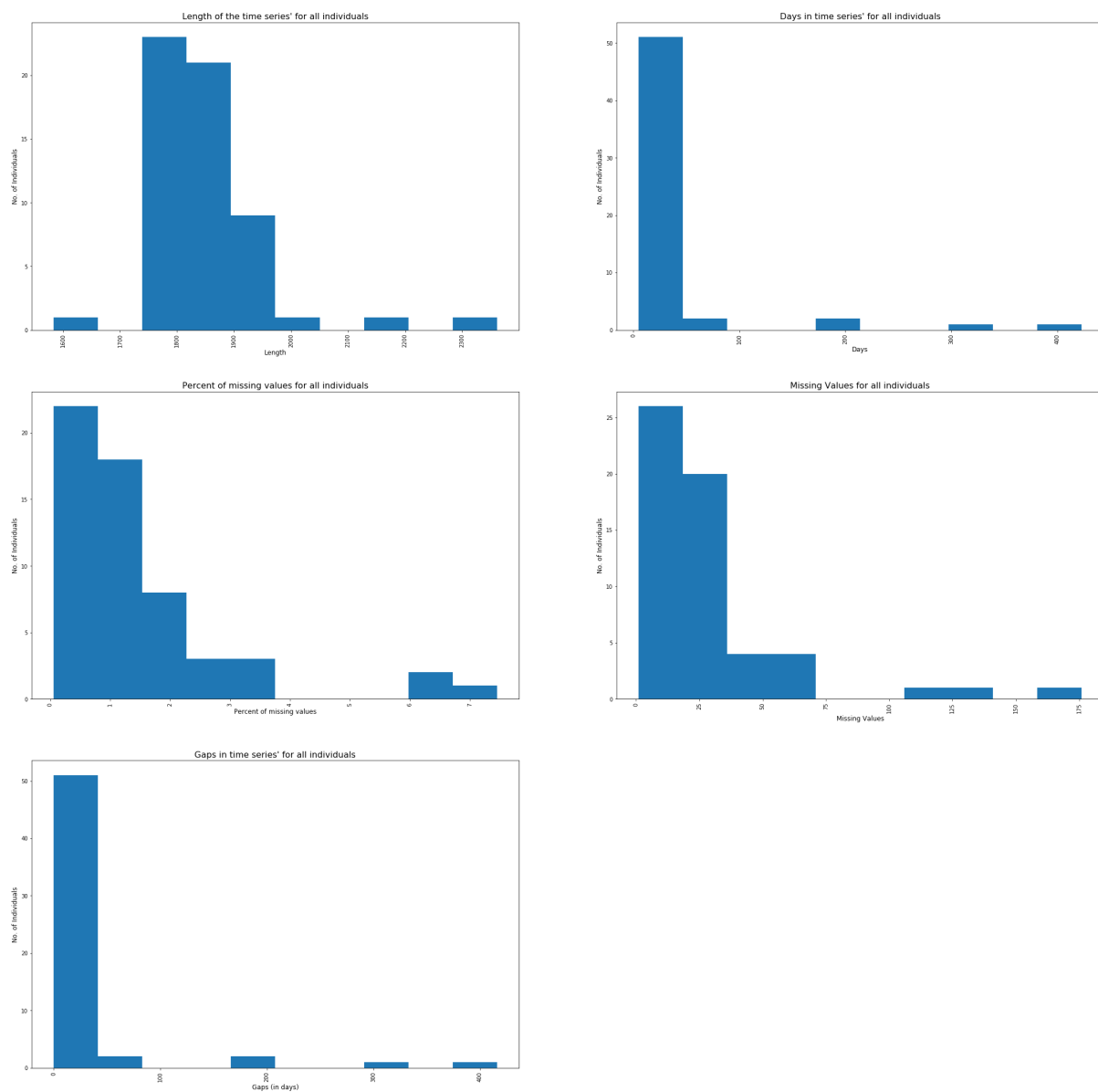
Here is the statistical analysis of the data:

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
0	2133-041	non-diabetic	1858	99	100	0 days 20:59:57	17	0.91	8 days 02:09:26	2017-07-03 18:32:06
1	2133-040	non-diabetic	1792	99	100	0 days 01:19:59	34	1.90	6 days 10:44:37	2017-06-28 13:34:52
2	2133-039	diabetic	2013	99	100	0 days 11:44:58	130	6.46	9 days 01:34:20	2017-06-05 12:23:22
3	2133-037	non-diabetic	1785	99	100	0 days 01:14:59	19	1.06	6 days 07:39:25	2017-06-05 14:11:14
4	2133-036	pre-diabetic	1954	99	100	0 days 06:49:59	119	6.09	8 days 15:04:26	2017-06-01 15:26:53
5	2133-035	pre-diabetic	1830	99	100	0 days 06:50:00	57	3.11	7 days 05:49:42	2017-06-01 17:17:00
6	2133-033	non-diabetic	1892	99	100	0 days 07:29:59	26	1.37	7 days 07:04:31	2017-05-19 12:58:22
7	2133-032	non-diabetic	1781	99	100	0 days 00:10:00	3	0.17	6 days 04:34:25	2017-05-19 13:31:19
8	2133-030	non-diabetic	1818	99	100	0 days 09:24:59	7	0.39	6 days 18:44:36	2017-05-03 14:13:09
9	2133-028	non-diabetic	1850	99	100	0 days 11:09:58	3	0.16	7 days 02:04:30	2017-05-10 00:03:32
10	2133-027	pre-diabetic	1936	99	100	0 days 16:04:57	29	1.50	8 days 02:13:55	2017-04-24 14:49:24
11	2133-026	non-diabetic	1782	99	100	0 days 00:10:00	3	0.17	6 days 04:39:35	2017-04-19 14:10:22
12	2133-025	non-diabetic	1865	99	100	0 days 01:55:05	10	0.54	6 days 18:09:38	2017-04-19 13:34:14
13	2133-024	pre-diabetic	1821	99	100	0 days 02:30:00	8	0.44	6 days 13:09:23	2017-04-17 14:14:20
14	2133-023	non-diabetic	1838	Low	100	0 days 19:09:56	13	0.71	7 days 10:09:25	2017-04-17 14:49:45
15	2133-022	non-diabetic	1814	Low	100	0 days 03:09:59	16	0.88	6 days 12:34:33	2017-03-22 13:34:24

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
16	2133-021	pre-diabetic	1797	99	100	0 days 01:19:58	7	0.39	6 days 07:39:27	2017-03-17 13:34:51
17	2133-020	non-diabetic	1826	99	100	0 days 01:15:00	5	0.27	6 days 10:04:43	2017-03-17 12:56:49
18	2133-019	pre-diabetic	1801	99	100	2 days 19:09:47	20	1.11	9 days 04:34:17	2017-03-15 16:37:47
19	2133-018	diabetic	1775	99	100	0 days 00:15:00	8	0.45	6 days 04:39:35	2017-03-14 13:30:04
20	2133-017	pre-diabetic	1799	99	100	0 days 01:14:59	21	1.17	6 days 10:04:23	2017-03-13 12:22:38
21	2133-015	pre-diabetic	1835	99	100	0 days 02:30:00	39	2.13	6 days 17:34:22	2017-01-30 13:33:25
22	2133-013	non-diabetic	1960	Low	100	0 days 07:24:58	68	3.47	8 days 07:49:13	2017-01-11 15:25:11
23	2133-012	non-diabetic	1935	99	100	5 days 03:14:36	13	0.67	12 days 10:14:03	2017-01-11 14:49:44
24	2133-011	non-diabetic	1933	Low	100	0 days 18:34:56	29	1.50	9 days 05:54:03	2017-01-10 15:29:05
25	2133-010	non-diabetic	1832	99	100	0 days 03:09:59	31	1.69	6 days 17:29:32	2016-11-21 15:28:45
26	2133-009	non-diabetic	1781	99	100	0 days 00:10:00	3	0.17	6 days 04:34:44	2016-11-21 13:33:38
27	2133-008	non-diabetic	1805	99	100	0 days 01:15:00	2	0.11	6 days 07:34:31	2016-11-22 00:03:05
28	2133-007	non-diabetic	1877	99	100	0 days 21:39:56	51	2.72	7 days 23:14:19	2016-11-01 13:30:28
29	2133-006	non-diabetic	1777	99	100	0 days 00:10:01	6	0.34	6 days 04:29:35	2016-11-01 00:00:51
30	2133-004	diabetic	1776	99	100	0 days 00:15:00	6	0.34	6 days 04:29:28	2016-09-21 00:04:11
31	2133-003	non-diabetic	1805	99	100	0 days 00:40:00	3	0.17	6 days 07:04:28	2016-09-02 17:51:46

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
32	2133-002	non-diabetic	1806	99	100	0 days 01:15:02	15	0.83	6 days 10:44:31	2016-09-09 16:02:22
33	2133-001	non-diabetic	1814	99	100	0 days 15:29:57	11	0.61	7 days 00:59:29	2016-08-03 00:00:14
34	1636-70-1010	pre-diabetic	1820	99	100	81 days 21:34:32	26	1.43	88 days 21:13:56	2016-03-02 15:29:17
35	1636-70-1008	non-diabetic	1760	99	100	0 days 00:15:01	20	1.14	6 days 04:34:32	2015-12-07 13:34:23
36	1636-70-1005	pre-diabetic	1846	99	100	1 days 01:19:56	23	1.25	8 days 03:24:24	2016-04-06 14:13:52
37	1636-70-1003	non-diabetic	1895	99	100	0 days 04:24:59	39	2.06	7 days 07:39:28	2015-12-01 00:03:03
38	1636-70-1002	non-diabetic	1794	99	100	0 days 01:20:00	11	0.61	6 days 07:34:30	2015-11-12 00:02:25
39	1636-69-123	non-diabetic	1905	99	100	1 days 03:49:54	16	0.84	8 days 01:34:20	2016-02-17 00:01:50
40	1636-69-114	pre-diabetic	1796	99	100	0 days 00:40:00	9	0.50	6 days 06:59:31	2015-10-13 00:03:47
41	1636-69-111	non-diabetic	1868	Low	100	14 days 05:39:11	22	1.18	21 days 00:18:50	2015-09-15 03:41:33
42	1636-69-107	non-diabetic	1897	99	100	0 days 06:49:59	45	2.37	7 days 10:09:27	2015-09-24 13:33:14
43	1636-69-104	non-diabetic	2361	99	100	172 days 20:31:01	176	7.45	201 days 18:23:29	2015-08-05 14:46:39
44	1636-69-100	non-diabetic	1880	99	100	168 days 03:11:24	35	1.86	175 days 02:10:50	2015-09-15 00:03:15
45	1636-69-091	diabetic	1803	99	100	0 days 01:15:00	23	1.28	6 days 10:09:39	2015-11-04 14:48:20
46	1636-69-090	pre-diabetic	1863	99	100	0 days 04:25:00	22	1.18	6 days 23:39:38	2016-02-10 00:02:40
47	1636-69-069	non-diabetic	1897	99	100	14 days 14:08:43	59	3.11	22 days 15:48:01	2015-07-20 13:33:11

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
48	1636-69-064	non-diabetic	1584	99	100	0 days 01:55:01	23	1.45	5 days 16:14:32	2015-03-26 14:49:42
49	1636-69-060	non-diabetic	1821	99	100	0 days 02:30:00	30	1.65	6 days 14:59:31	2015-11-17 15:25:53
50	1636-69-053	non-diabetic	1867	99	100	292 days 03:12:36	30	1.61	299 days 01:32:14	2015-05-07 18:34:18
51	1636-69-048	non-diabetic	1779	99	100	0 days 00:20:00	3	0.17	6 days 04:34:29	2015-11-26 00:36:15
52	1636-69-035	non-diabetic	2180	99	100	0 days 12:24:58	58	2.66	11 days 01:04:20	2016-01-27 00:02:25
53	1636-69-032	pre-diabetic	1783	99	100	0 days 00:10:00	1	0.06	6 days 04:34:32	2016-01-13 12:58:17
54	1636-69-028	non-diabetic	1822	99	100	68 days 12:25:47	24	1.32	75 days 07:45:22	2015-04-02 13:33:15
55	1636-69-026	pre-diabetic	1796	99	100	0 days 08:39:59	32	1.78	7 days 01:29:34	2015-11-24 00:37:20
56	1636-69-001	diabetic	1846	99	100	416 days 17:21:34	7	0.38	423 days 11:25:54	2014-02-03 03:42:12



The following method describes the data after processing

The imputation model has been trained on this processed data

We use the following processed (cleaned) data for training our model

This minimizes erroneous data being fed to the model

```
In [16]: obj.processedData()
```

Here is a glimpse of the data:

	Display Time	GlucoseValue	subjectId
0	2014-02-03 03:42:12	93	1636-69-001
1	2014-02-03 03:47:12	93	1636-69-001
2	2014-02-03 03:52:12	93	1636-69-001
3	2014-02-03 03:57:12	95	1636-69-001
4	2014-02-03 04:02:12	96	1636-69-001

	Length of readings	Gapsize	Missing Values	Percent of missing values	Days
count	60.000000	60	60.000000	60.000000	60
mean	1727.900000	0 days 05:49:09.116666	25.200000	1.490000	6 days 15:30:15.116666
std	288.173835	0 days 06:52:22.017395	26.647606	1.688264	1 days 08:38:16.086564
min	580.000000	0 days 00:10:00	1.000000	0.060000	2 days 01:54:50
25%	1776.750000	0 days 01:14:59.750000	7.000000	0.447500	6 days 04:38:22.250000
50%	1804.500000	0 days 02:49:59.500000	18.000000	1.110000	6 days 10:44:34
75%	1847.000000	0 days 07:31:13.750000	30.250000	1.690000	7 days 03:00:48
max	2180.000000	1 days 03:49:54	130.000000	7.740000	11 days 01:04:20

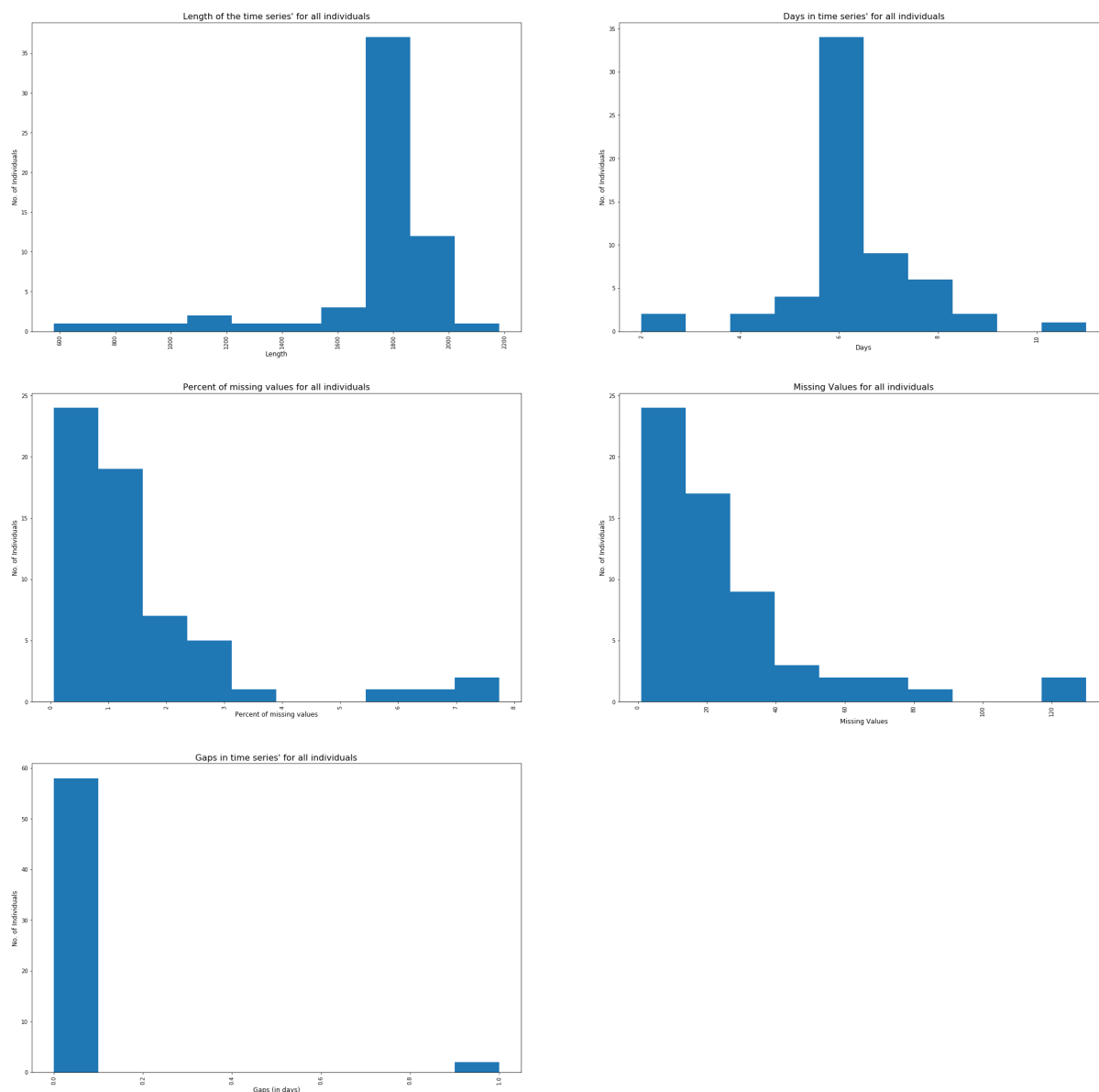
Here is the statistical analysis of the data:

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
0	2133-041	non-diabetic	1858	204	51	0 days 20:59:57	17	0.91	8 days 02:09:26	2017-07-03 18:32:06
1	2133-040	non-diabetic	1792	171	59	0 days 01:19:59	34	1.90	6 days 10:44:37	2017-06-28 13:34:52
2	2133-039	diabetic	2013	204	50	0 days 11:44:58	130	6.46	9 days 01:34:20	2017-06-05 12:23:22
3	2133-037	non-diabetic	1785	153	48	0 days 01:14:59	19	1.06	6 days 07:39:25	2017-06-05 14:11:14
4	2133-036	pre-diabetic	1954	214	58	0 days 06:49:59	119	6.09	8 days 15:04:26	2017-06-01 15:26:53
5	2133-035	pre-diabetic	1830	190	47	0 days 06:50:00	57	3.11	7 days 05:49:42	2017-06-01 17:17:00
6	2133-033	non-diabetic	1892	161	49	0 days 07:29:59	26	1.37	7 days 07:04:31	2017-05-19 12:58:22
7	2133-032	non-diabetic	1781	147	69	0 days 00:10:00	3	0.17	6 days 04:34:25	2017-05-19 13:31:19
8	2133-030	non-diabetic	1818	147	41	0 days 09:24:59	7	0.39	6 days 18:44:36	2017-05-03 14:13:09
9	2133-028	non-diabetic	1850	111	45	0 days 11:09:58	3	0.16	7 days 02:04:30	2017-05-10 00:03:32
10	2133-027	pre-diabetic	1936	155	60	0 days 16:04:57	29	1.50	8 days 02:13:55	2017-04-24 14:49:24
11	2133-026	non-diabetic	1782	166	45	0 days 00:10:00	3	0.17	6 days 04:39:35	2017-04-19 14:10:22
12	2133-025	non-diabetic	1865	175	45	0 days 01:55:05	10	0.54	6 days 18:09:38	2017-04-19 13:34:14
13	2133-024	pre-diabetic	1821	180	41	0 days 02:30:00	8	0.44	6 days 13:09:23	2017-04-17 14:14:20
14	2133-023	non-diabetic	1838	142	44	0 days 19:09:56	13	0.71	7 days 10:09:25	2017-04-17 14:49:45
15	2133-022	non-diabetic	1814	204	40	0 days 03:09:59	16	0.88	6 days 12:34:33	2017-03-22 13:34:24

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
16	2133-021	pre-diabetic	1797	236	62	0 days 01:19:58	7	0.39	6 days 07:39:27	2017-03-17 13:34:51
17	2133-020	non-diabetic	1826	201	40	0 days 01:15:00	5	0.27	6 days 10:04:43	2017-03-17 12:56:49
18	2133-019-2	pre-diabetic	580	168	63	0 days 01:19:59	4	0.69	2 days 01:54:50	2017-03-15 16:37:47
19	2133-019	pre-diabetic	1221	192	53	0 days 00:20:01	15	1.23	4 days 07:29:40	2017-03-20 13:42:24
20	2133-018	diabetic	1775	303	73	0 days 00:15:00	8	0.45	6 days 04:39:35	2017-03-14 13:30:04
21	2133-017	pre-diabetic	1799	182	68	0 days 01:14:59	21	1.17	6 days 10:04:23	2017-03-13 12:22:38
22	2133-015	pre-diabetic	1835	213	58	0 days 02:30:00	39	2.13	6 days 17:34:22	2017-01-30 13:33:25
23	2133-013	non-diabetic	1960	206	41	0 days 07:24:58	68	3.47	8 days 07:49:13	2017-01-11 15:25:11
24	2133-012	non-diabetic	1809	148	49	0 days 03:45:00	12	0.66	6 days 20:34:29	2017-01-11 14:49:44
25	2133-011	non-diabetic	1933	204	47	0 days 18:34:56	29	1.50	9 days 05:54:03	2017-01-10 15:29:05
26	2133-010	non-diabetic	1832	143	62	0 days 03:09:59	31	1.69	6 days 17:29:32	2016-11-21 15:28:45
27	2133-009	non-diabetic	1781	183	59	0 days 00:10:00	3	0.17	6 days 04:34:44	2016-11-21 13:33:38
28	2133-008	non-diabetic	1805	137	49	0 days 01:15:00	2	0.11	6 days 07:34:31	2016-11-22 00:03:05
29	2133-007	non-diabetic	1877	207	65	0 days 21:39:56	51	2.72	7 days 23:14:19	2016-11-01 13:30:28
30	2133-006	non-diabetic	1777	148	54	0 days 00:10:01	6	0.34	6 days 04:29:35	2016-11-01 00:00:51
31	2133-004	diabetic	1776	246	61	0 days 00:15:00	6	0.34	6 days 04:29:28	2016-09-21 00:04:11

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
32	2133-003	non-diabetic	1805	190	62	0 days 00:40:00	3	0.17	6 days 07:04:28	2016-09-02 17:51:46
33	2133-002	non-diabetic	1806	163	60	0 days 01:15:02	15	0.83	6 days 10:44:31	2016-09-09 16:02:22
34	2133-001	non-diabetic	1814	186	47	0 days 15:29:57	11	0.61	7 days 00:59:29	2016-08-03 00:00:14
35	1636-70-1010	pre-diabetic	1393	186	75	0 days 04:34:58	15	1.08	5 days 02:04:34	2016-05-25 10:38:39
36	1636-70-1008	non-diabetic	1760	223	51	0 days 00:15:01	20	1.14	6 days 04:34:32	2015-12-07 13:34:23
37	1636-70-1005	pre-diabetic	1846	225	52	1 days 01:19:56	23	1.25	8 days 03:24:24	2016-04-06 14:13:52
38	1636-70-1003	non-diabetic	1895	146	46	0 days 04:24:59	39	2.06	7 days 07:39:28	2015-12-01 00:03:03
39	1636-70-1002	non-diabetic	1794	168	57	0 days 01:20:00	11	0.61	6 days 07:34:30	2015-11-12 00:02:25
40	1636-69-123	non-diabetic	1905	138	49	1 days 03:49:54	16	0.84	8 days 01:34:20	2016-02-17 00:01:50
41	1636-69-114	pre-diabetic	1796	163	76	0 days 00:40:00	9	0.50	6 days 06:59:31	2015-10-13 00:03:47
42	1636-69-111	non-diabetic	1720	135	48	0 days 02:00:00	21	1.22	6 days 06:24:41	2015-09-15 03:41:33
43	1636-69-107	non-diabetic	1897	195	66	0 days 06:49:59	45	2.37	7 days 10:09:27	2015-09-24 13:33:14
44	1636-69-104-2	non-diabetic	1153	140	46	0 days 11:44:58	86	7.46	6 days 08:49:31	2016-02-17 00:20:37
45	1636-69-104	non-diabetic	969	152	61	0 days 06:49:58	75	7.74	5 days 14:59:34	2015-08-05 14:46:39
46	1636-69-100	non-diabetic	1716	153	66	0 days 02:29:59	29	1.69	6 days 08:49:28	2015-09-15 00:03:15
47	1636-69-091	diabetic	1803	163	70	0 days 01:15:00	23	1.28	6 days 10:09:39	2015-11-04 14:48:20

	Subject ID	Status	Length of readings	Max. Glucose Value	Min. Glucose Value	Gapsizes	Missing Values	Percent of missing values	Days	Start
48	1636-69-090	pre-diabetic	1863	195	54	0 days 04:25:00	22	1.18	6 days 23:39:38	2016-02-10 00:02:40
49	1636-69-069	non-diabetic	1591	187	69	0 days 07:34:58	48	3.02	6 days 20:39:24	2015-07-20 13:33:11
50	1636-69-064	non-diabetic	1584	218	43	0 days 01:55:01	23	1.45	5 days 16:14:32	2015-03-26 14:49:42
51	1636-69-060	non-diabetic	1821	191	91	0 days 02:30:00	30	1.65	6 days 14:59:31	2015-11-17 15:25:53
52	1636-69-053	non-diabetic	1804	183	43	0 days 03:45:00	26	1.44	6 days 16:49:39	2015-05-07 18:34:18
53	1636-69-048	non-diabetic	1779	144	63	0 days 00:20:00	3	0.17	6 days 04:34:29	2015-11-26 00:36:15
54	1636-69-035	non-diabetic	2180	234	69	0 days 12:24:58	58	2.66	11 days 01:04:20	2016-01-27 00:02:25
55	1636-69-032	pre-diabetic	1783	191	67	0 days 00:10:00	1	0.06	6 days 04:34:32	2016-01-13 12:58:17
56	1636-69-028	non-diabetic	1543	168	75	0 days 04:59:59	21	1.36	5 days 19:59:40	2015-04-02 13:33:15
57	1636-69-026	pre-diabetic	1796	182	62	0 days 08:39:59	32	1.78	7 days 01:29:34	2015-11-24 00:37:20
58	1636-69-001-2	diabetic	1092	250	64	0 days 04:19:59	5	0.46	4 days 03:14:35	2015-03-29 11:53:31
59	1636-69-001	diabetic	754	225	68	0 days 00:10:00	1	0.13	2 days 14:49:45	2014-02-03 03:42:12



```
In [17]: obj.train(training_data)
```

Train shape: (103673, 1, 1)

Epoch 1/1

103673/103673 [=====] - 249s 2ms/step - loss: 4.4758e-04

To test the model we trained on the HALL dataset, write the path of the testing dataset below:

```
In [18]: testing_data = pd.read_csv("~/Desktop/NCSA_genomics/Python - notebooks/T  
SForecasting/Data/test/generated_test.txt", sep=",")
```

Now, we run the imputation model on the data you supplied above

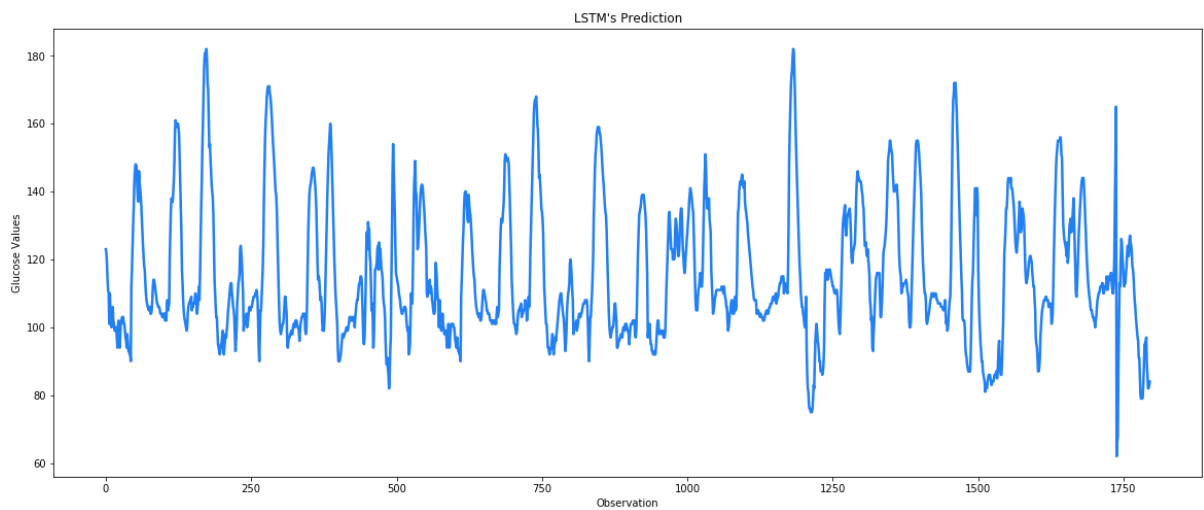
```
In [19]: obj.impute(testing_data)
```

Gap detected!

566/566 [=====] - 1s 898us/step

Imputations performed!

File saved!



```
In [ ]:
```