

Project 1 Report

DATA MINING

CSE 572: Fall 2019

Submitted to:

Dr. Ayan Banerjee
Ira A. Fulton School of Engineering
Arizona State University

Submitted by(GROUP 5):

Sumanth Paranjape (sparanj2@asu.edu)
Sameena Hossain (shossai5@asu.edu)

October 11, 2019

Introduction:

We were assigned 6 tasks as part of Project 1 completion. These are:

- a) Extract 4 (one for each student) different types of time series features from only the CGM data cell array and CGM timestamp cell array
- b) For each time series explain why you chose such feature
- c) Show values of each of the features and argue that your intuition in step b is validated or disproved?
- d) Create a feature matrix where each row is a collection of features from each time series. SO if there are 75 time series and your feature length after concatenation of the 4 types of features is 17 then the feature matrix size will be 75 X 17
- e) Provide this feature matrix to PCA and derive the new feature matrix. Choose the top 5 features and plot them for each time series.
- f) For each feature in the top 5 argue why it is chosen as a top five feature in PCA?

We have divided the Project into two phases- Phase 1 and Phase 2. In phase 1, we pre-process the data by cleaning and organizing. In phase 2, we perform the tasks assigned in the project document.

Phase 1: Data Cleaning and Organization

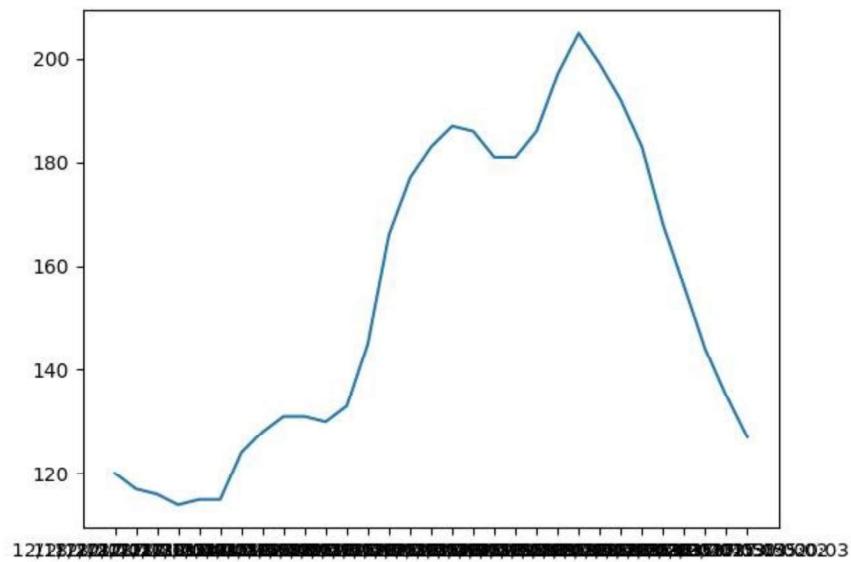
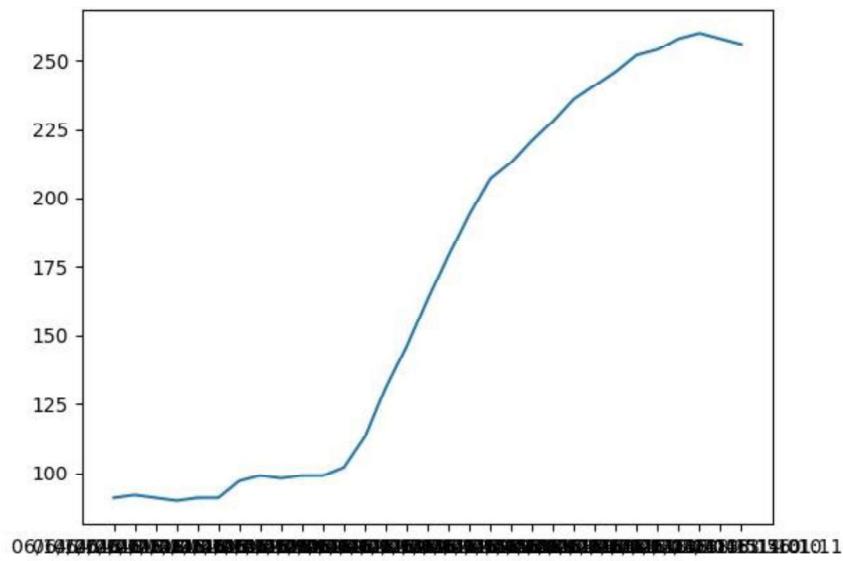
The CGM series data had missing values mostly in the last column for multiple cells, also for some cells on the in the first column. We wrote python code to clean the CGM series lunch data from patient wise files. We used forward fill and backward fill to fill those missing values. We also wrote python code to format the timestamps from the CGM time stamp files. The timestamps were converted into date and time format. We used python code to merge the CGM series and CGM time stamp data into one CSV file named ‘Patient 1 Merged Data’. The merged data was then ready to be used in performing the tasks in Phase 2. We use data from one patient (Patient 1) to perform the assigned task. Similar cleaning and organization can be performed for all 5 patients and subsequently the assigned tasks can be performed on those data as well.

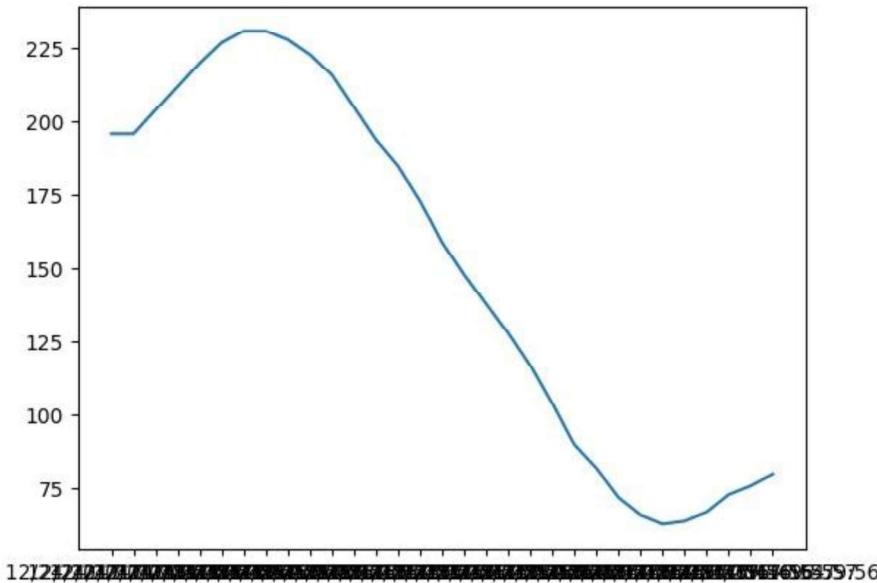
We provide a screen shot of the ‘Patient 1 Merged data’-

MERGED DATA FRAME																
0	1	2	3	4	5	6	62	63	64	65	66	67	68	69	70	71
0 06/14/2018 11:31:11	91.0	06/06/2018 12:33:59	216.0	06/05/2018 10:43:54	...	100.0	11/08/2017 11:37:03	99.0	11/06/2017 12:33:21	78.0
1 06/14/2018 11:30:10	92.0	06/06/2018 12:36:59	215.0	06/05/2018 10:48:54	...	100.0	11/08/2017 11:42:03	99.0	11/06/2017 12:36:21	77.0
2 06/14/2018 11:31:10	93.0	06/06/2018 12:34:59	232.0	06/05/2018 10:47:00	...	95.0	11/08/2017 11:47:06	100.0	11/06/2017 12:43:22	75.0
3 06/14/2018 11:46:11	90.0	06/06/2018 12:53:59	260.0	06/05/2018 10:58:53	...	90.0	11/08/2017 11:58:04	101.0	11/06/2017 12:53:22	76.0
4 06/14/2018 11:51:11	91.0	06/06/2018 12:53:59	261.0	06/05/2018 11:03:54	...	90.0	11/08/2017 11:57:03	105.0	11/06/2017 12:53:21	75.0
5 06/14/2018 11:56:10	91.0	06/06/2018 12:58:59	269.0	06/05/2018 11:08:54	...	92.0	11/08/2017 12:02:03	109.0	11/06/2017 12:58:21	75.0
6 06/14/2018 12:01:10	97.0	06/06/2018 13:03:59	275.0	06/05/2018 11:13:53	...	96.0	11/08/2017 12:07:03	116.0	11/06/2017 13:03:23	78.0
7 06/14/2018 12:02:10	99.0	06/06/2018 13:08:59	292.0	06/05/2018 11:18:53	...	100.0	11/08/2017 12:12:03	120.0	11/06/2017 13:12:22	79.0
8 06/14/2018 12:11:11	98.0	06/06/2018 13:14:59	294.0	06/05/2018 11:23:54	...	115.0	11/08/2017 12:17:04	128.0	11/06/2017 13:13:22	79.0
9 06/14/2018 12:16:11	99.0	06/06/2018 13:18:59	310.0	06/05/2018 11:28:54	...	126.0	11/08/2017 12:22:04	128.0	11/06/2017 13:18:23	78.0
10 06/14/2018 12:21:10	99.0	06/06/2018 13:23:59	324.0	06/05/2018 11:33:54	...	131.0	11/08/2017 12:27:03	126.0	11/06/2017 13:23:21	76.0
11 06/14/2018 12:26:10	100.0	06/06/2018 13:28:59	330.0	06/05/2018 11:38:54	...	132.0	11/08/2017 12:32:03	127.0	11/06/2017 13:32:22	74.0
12 06/14/2018 12:31:11	113.0	06/06/2018 13:34:00	345.0	06/05/2018 11:43:53	...	127.0	11/08/2017 12:37:04	125.0	11/06/2017 13:33:22	74.0
13 06/14/2018 12:36:11	131.0	06/06/2018 13:38:59	348.0	06/05/2018 11:48:54	...	121.0	11/08/2017 12:42:05	124.0	11/06/2017 13:38:23	75.0
14 06/14/2018 12:41:10	146.0	06/06/2018 13:43:59	349.0	06/05/2018 11:53:54	...	116.0	11/08/2017 12:47:05	123.0	11/06/2017 13:43:23	77.0
15 06/14/2018 12:46:10	146.0	06/06/2018 13:48:59	349.0	06/05/2018 11:58:54	...	116.0	11/08/2017 12:52:05	123.0	11/06/2017 13:52:23	77.0
16 06/14/2018 12:51:10	179.0	06/06/2018 13:54:00	340.0	06/05/2018 12:03:53	...	105.0	11/08/2017 12:57:06	125.0	11/06/2017 13:53:22	89.0
17 06/14/2018 12:56:11	194.0	06/06/2018 13:59:00	347.0	06/05/2018 12:08:54	...	99.0	11/08/2017 13:02:05	128.0	11/06/2017 13:58:22	161.0
18 06/14/2018 13:01:10	207.0	06/06/2018 14:03:59	347.0	06/05/2018 12:13:54	...	95.0	11/08/2017 13:07:05	133.0	11/06/2017 14:03:23	109.0
19 06/14/2018 13:06:10	221.0	06/06/2018 14:08:59	349.0	06/05/2018 12:18:54	...	99.0	11/08/2017 13:12:05	130.0	11/06/2017 14:12:23	107.0
20 06/14/2018 13:11:10	221.0	06/06/2018 14:14:00	343.0	06/05/2018 12:23:53	...	97.0	11/08/2017 13:17:06	132.0	11/06/2017 14:13:22	120.0
21 06/14/2018 13:16:11	228.0	06/06/2018 14:19:00	341.0	06/05/2018 12:28:53	...	96.0	11/08/2017 13:22:06	136.0	11/06/2017 14:18:22	122.0
22 06/14/2018 13:21:11	236.0	06/06/2018 14:23:59	335.0	06/05/2018 12:33:54	...	95.0	11/08/2017 13:27:05	127.0	11/06/2017 14:23:21	124.0
23 06/14/2018 13:26:10	240.0	06/06/2018 14:28:59	339.0	06/05/2018 12:38:54	...	96.0	11/08/2017 13:32:05	129.0	11/06/2017 14:32:23	126.0
24 06/14/2018 13:31:10	246.0	06/06/2018 14:33:59	346.0	06/05/2018 12:43:53	...	91.0	11/08/2017 13:37:06	120.0	11/06/2017 14:33:22	130.0
25 06/14/2018 13:36:10	252.0	06/06/2018 14:39:00	310.0	06/05/2018 12:48:53	...	87.0	11/08/2017 13:42:06	128.0	11/06/2017 14:38:22	145.0
26 06/14/2018 13:41:10	254.0	06/06/2018 14:44:00	303.0	06/05/2018 12:53:54	...	84.0	11/08/2017 13:47:05	126.0	11/06/2017 14:43:22	150.0
27 06/14/2018 13:46:10	259.0	06/06/2018 14:49:00	299.0	06/05/2018 12:58:54	...	79.0	11/08/2017 13:52:05	125.0	11/06/2017 14:52:22	150.0
28 06/14/2018 13:51:10	266.0	06/06/2018 14:53:59	260.0	06/05/2018 13:03:54	...	78.0	11/08/2017 13:57:05	111.0	11/06/2017 14:53:21	156.0
29 06/14/2018 13:56:10	258.0	06/06/2018 14:59:00	273.0	06/05/2018 13:08:53	...	78.0	11/08/2017 14:02:06	107.0	11/06/2017 14:58:22	160.0
30 06/14/2018 14:01:11	256.0	06/06/2018 14:04:00	260.0	06/05/2018 13:13:53	...	78.0	11/08/2017 14:07:06	106.0	11/06/2017 15:03:22	164.0

Patient 1 Merged Data

and some example plots from the data, the x axis represents the time stamps from each time series and the y axis represents the CGM reading-





Phase 2: Task Completion

In Phase 2, we performed the tasks assigned in the project document. The tasks and the steps taken in to complete each task is discussed as follows-

Task A: Feature Extraction

In this task we used multiple feature extraction methods such as Fast Fourier Transform, a set of Statistical Features (average, std) and mathematical calculation (percent) in order to select features that are significant.

- Feature Extraction Techniques-

Fast Fourier Transform-

The Fast Fourier Transform (FFT) is a mathematical method for transforming a function of time into a function of frequency. Sometimes it is described as transforming from the time domain to the frequency domain. It is very useful for analysis of time-dependent phenomena. It is an algorithm to sample the time domain signal and convert it into frequency domain sinusoidal signals to show where does the signal have highest amplitudes for given frequencies.

Here we calculate the variance of FFT, by using the built-in python FFT function. Finding the variance of FFT for the readings in each time series for a patient can provide important insight into patient's blood glucose readings by showing the peak values.

Statistical Features-

Statistical features give useful information about what is the maximum, minimum, mean, average and standard deviation of the data over the time. Choosing statistical features to analyze data is adopted by many researchers and research methods. Therefore, we decided to implement statistical feature extraction in order to be able extract significant feature from the merged Patient 1 data.

The computation of mean glucose values from CGM data is suggested as a descriptor for overall glycemic control. Computing of pre- and post-meal averages and their difference can serve as an indication of the overall effectiveness of pre-meal bolus timing and amount [1]. We make use of this reference and calculate the pre and most meal average for each time series for each patient. We intend to find significant information about the patient's glycemic control.

Mathematical Calculations-

Simple mathematical calculations can also help in extracting significant information from a time series data.

We used mathematical calculations to find the percent of times the patient was 'Hypoglycemic' and the percent of times the patient was 'Hyperglycemic'. We have used the ranges listed in [1] for determining the 'Hypoglycemic' and 'Hyperglycemic' states-

Hypoglycemic - below 70 mg/dL &

Hyperglycemic- above 180 mg/dL

We hypothesize that extracting these two features will help in identifying whether the patient's glucose levels are within the target range.

• Extracted Features-

We have extracted six features from the 'Patient 1 Merged Data', of which we hypothesize **five** of them to be of significance. They are as follows-

1. **Variance of FFT**
2. **Average of Pre and Post meal CGM readings**
3. **Hyperglycemic Percentage**
4. **Hypoglycemic Percentage**
5. **Standard Deviation**

The following shows a screenshot of the feature matrix

		Timestamp	Feature1	Feature2	Feature3	Feature4	Feature5	Feature6
0	06/14/2018	11:31:11 - 06/14/2018 14:01:11	1005339.0	101.795031	0.000000	45.161290	67.847874	167.612903
1	06/06/2018	12:33:59 - 06/06/2018 15:04:00	2870807.0	77.968944	0.000000	100.000000	41.416440	303.967742
2	06/05/2018	10:43:54 - 06/05/2018 13:13:53	1249650.0	92.378882	0.000000	51.612903	59.793157	192.935484
3	06/04/2018	11:13:49 - 06/04/2018 13:43:48	795109.0	76.770186	0.000000	41.935484	56.037393	150.935484
4	05/21/2018	11:38:45 - 05/21/2018 14:08:44	626259.0	29.478261	0.000000	0.000000	16.942211	142.838710
5	05/16/2018	12:04:40 - 05/16/2018 14:34:42	870645.0	92.031056	0.000000	48.387097	56.183678	158.677419
6	03/21/2018	11:34:04 - 03/21/2018 14:04:05	636710.0	50.198758	0.000000	16.129032	34.300351	140.258065
7	03/01/2018	11:14:02 - 03/01/2018 13:44:01	888197.0	78.099379	0.000000	38.709677	57.656361	151.741935
8	02/26/2018	11:13:47 - 02/26/2018 13:43:46	671963.0	57.198758	0.000000	19.354839	33.749398	144.451613
9	02/21/2018	11:19:40 - 02/21/2018 13:49:44	722144.0	33.173913	0.000000	22.580645	22.903271	152.612903
10	02/16/2018	12:40:01 - 02/16/2018 15:10:02	1003306.0	75.285714	0.000000	54.838710	51.703207	173.935484
11	02/15/2018	11:04:55 - 02/15/2018 13:34:56	981079.0	88.118012	0.000000	38.709677	67.922003	146.387097
12	02/12/2018	11:49:39 - 02/12/2018 14:19:41	774825.0	57.695652	0.000000	41.935484	37.412040	154.870968
13	02/06/2018	11:27:30 - 02/06/2018 13:57:28	741151.0	62.937888	0.000000	35.483871	40.966680	150.354839
14	12/29/2017	11:38:48 - 12/29/2017 14:08:47	317078.0	27.770186	0.000000	0.000000	13.766297	101.193548
15	12/27/2017	11:45:50 - 12/27/2017 14:15:50	727389.0	60.931677	0.000000	32.258065	46.043842	147.354839
16	12/26/2017	12:05:45 - 12/26/2017 14:35:44	976319.0	53.099379	0.000000	48.387097	38.221709	175.064516
17	12/24/2017	14:29:57 - 12/24/2017 16:59:56	785352.0	83.981366	12.903226	45.161290	62.439015	150.580645
18	12/23/2017	12:19:49 - 12/23/2017 14:49:50	518234.0	57.552795	0.000000	9.677419	40.905515	123.645161
19	12/19/2017	11:39:29 - 12/19/2017 14:09:29	732582.0	65.906832	0.000000	32.258065	38.011445	149.967742
20	12/18/2017	11:30:02 - 12/18/2017 14:00:03	748782.0	48.931677	0.000000	35.483871	30.375006	153.935484
21	12/15/2017	11:49:46 - 12/15/2017 14:19:47	251350.0	3.453416	0.000000	0.000000	6.239247	91.677419
22	12/13/2017	11:34:35 - 12/13/2017 14:04:36	766739.0	21.720497	0.000000	25.806452	27.377197	157.193548
23	12/12/2017	13:29:30 - 12/12/2017 15:59:29	391307.0	26.757764	3.225806	0.000000	26.459794	110.516129
24	11/27/2017	11:34:55 - 11/27/2017 14:04:54	810657.0	60.813665	0.000000	41.935484	38.147286	158.451613
25	11/22/2017	11:39:29 - 11/22/2017 14:09:29	332362.0	26.142857	0.000000	0.000000	18.135861	103.161290
26	11/20/2017	12:03:59 - 11/20/2017 14:33:59	298449.0	27.677019	3.225806	0.000000	19.797676	99.709677
27	11/17/2017	12:08:44 - 11/17/2017 14:38:45	598743.0	39.757764	0.000000	9.677419	20.466770	141.419355
28	11/14/2017	11:28:28 - 11/14/2017 13:58:29	637512.0	44.832298	0.000000	6.451613	29.054006	141.838710
29	11/11/2017	11:52:20 - 11/11/2017 14:22:21	634115.0	60.472050	0.000000	22.580645	32.955051	143.645161
30	11/09/2017	11:22:08 - 11/09/2017 13:52:08	306480.0	6.590062	0.000000	0.000000	15.323905	99.870968
31	11/08/2017	11:37:03 - 11/08/2017 14:07:06	439823.0	20.608696	0.000000	0.000000	10.930677	119.935484
32	11/06/2017	12:33:21 - 11/06/2017 15:03:22	355604.0	36.652174	0.000000	0.000000	31.598483	103.290323

Task B: Explain why these features were chosen

1. Variance of FFT-

FFT reduces the noise to reveal repeating patterns. For a small data set it is more apt to find the variance of FFT to get better insight into the data. This is the reason we chose this feature.

2. Average of Pre and Post meal CGM readings-

As discussed above, the average of post and pre meal CGM data is a descriptor for glycemic control. This feature can present a significant insight into the patients random glucose levels [1]. This is the reason we chose this feature.

3. Hyperglycemic Percentage

The percentage of times when the patient was hyperglycemic can provide a significant insight into the patient's overall health. This information can be used to notify the patient if they are at the risk of being extremely hyperglycemic and should seek medical attention. This is the reason we chose this feature.

4. Hypoglycemic Percentage

The percentage of times when the patient was hypoglycemic can also provide significant insight into the patient's overall health. This information can be used to notify the patient if they are at the risk of being extremely hypoglycemic and should eat something or to seek medical attention. This is the reason we chose this feature.

5. Standard Deviation

The standard deviation of the blood glucose level in each time series can provide significant insight into the patient's blood sugar stability. This is the reason we chose this feature.

Task C: Validation for the chosen features

1. Variance of FFT

As we hypothesized the variance of FFT shows significant information about the patient's glucose level in each of the time series data. The following two screen shots show the merged data and the variance of FFT. As can be seen, variance of FFT indicates the fluctuations in the CGM data.

	MERGED DATA FRAME																	
	0	1	2	3	4	...	61	62	63	64	65							
0	06/14/2018 11:31:11	91.0	06/06/2018 12:33:59	216.0	06/05/2018 10:43:54	...	100.0	11/08/2017 11:37:03	99.0	11/06/2017 12:33:21	78.0							
1	06/14/2018 11:36:10	92.0	06/06/2018 12:38:50	216.0	06/05/2018 10:48:54	...	100.0	11/08/2017 11:42:03	99.0	11/06/2017 12:38:21	77.0							
2	06/14/2018 11:41:10	91.0	06/06/2018 12:44:00	232.0	06/05/2018 10:53:53	...	95.0	11/08/2017 11:47:04	99.0	11/06/2017 12:43:22	75.0							
3	06/14/2018 11:46:11	98.0	06/06/2018 12:49:00	245.0	06/05/2018 10:58:53	...	91.0	11/08/2017 11:52:04	101.0	11/06/2017 12:48:22	74.0							
4	06/14/2018 11:51:11	91.0	06/06/2018 12:53:59	261.0	06/05/2018 11:03:54	...	89.0	11/08/2017 11:57:03	105.0	11/06/2017 12:53:21	75.0							
5	06/14/2018 11:56:10	91.0	06/06/2018 12:58:59	269.0	06/05/2018 11:08:54	...	92.0	11/08/2017 12:02:03	109.0	11/06/2017 12:58:21	75.0							
6	06/14/2018 12:01:10	97.0	06/06/2018 13:03:59	275.0	06/05/2018 11:13:53	...	96.0	11/08/2017 12:07:03	116.0	11/06/2017 13:03:22	78.0							
7	06/14/2018 12:06:10	99.0	06/06/2018 13:09:00	284.0	06/05/2018 11:18:53	...	103.0	11/08/2017 12:12:04	124.0	11/06/2017 13:08:22	79.0							
8	06/14/2018 12:11:11	98.0	06/06/2018 13:14:00	294.0	06/05/2018 11:23:54	...	115.0	11/08/2017 12:17:04	128.0	11/06/2017 13:13:22	79.0							
9	06/14/2018 12:16:11	99.0	06/06/2018 13:18:59	310.0	06/05/2018 11:28:54	...	126.0	11/08/2017 12:22:04	128.0	11/06/2017 13:18:21	78.0							
10	06/14/2018 12:21:10	99.0	06/06/2018 13:23:59	324.0	06/05/2018 11:33:54	...	131.0	11/08/2017 12:27:03	126.0	11/06/2017 13:23:21	76.0							
11	06/14/2018 12:26:10	102.0	06/06/2018 13:29:00	335.0	06/05/2018 11:38:53	...	132.0	11/08/2017 12:32:04	126.0	11/06/2017 13:28:22	74.0							
12	06/14/2018 12:31:11	113.0	06/06/2018 13:34:00	345.0	06/05/2018 11:43:53	...	127.0	11/08/2017 12:37:04	125.0	11/06/2017 13:33:22	74.0							
13	06/14/2018 12:36:11	131.0	06/06/2018 13:38:59	348.0	06/05/2018 11:48:54	...	121.0	11/08/2017 12:42:05	124.0	11/06/2017 13:38:21	75.0							
14	06/14/2018 12:41:10	146.0	06/06/2018 13:43:59	349.0	06/05/2018 11:53:54	...	116.0	11/08/2017 12:47:05	123.0	11/06/2017 13:43:21	77.0							
15	06/14/2018 12:46:10	161.0	06/06/2018 13:48:50	351.0	06/05/2018 11:58:53	...	110.0	11/08/2017 12:52:06	123.0	11/06/2017 13:48:22	R1.0							
16	06/14/2018 12:51:10	179.0	06/06/2018 13:54:00	349.0	06/05/2018 12:03:53	...	105.0	11/08/2017 12:57:06	125.0	11/06/2017 13:53:22	89.0							
17	06/14/2018 12:56:11	194.0	06/06/2018 13:59:00	347.0	06/05/2018 12:08:54	...	99.0	11/08/2017 13:02:05	128.0	11/06/2017 13:58:22	101.0							
18	06/14/2018 13:01:11	207.0	06/06/2018 14:03:59	347.0	06/05/2018 12:13:54	...	95.0	11/08/2017 13:07:05	133.0	11/06/2017 14:03:21	109.0							
19	06/14/2018 13:06:10	213.0	06/06/2018 14:08:59	345.0	06/05/2018 12:18:54	...	96.0	11/08/2017 13:12:05	133.0	11/06/2017 14:08:21	115.0							
20	06/14/2018 13:11:10	221.0	06/06/2018 14:14:00	343.0	06/05/2018 12:23:53	...	97.0	11/08/2017 13:17:06	132.0	11/06/2017 14:13:22	120.0							
21	06/14/2018 13:16:11	228.0	06/06/2018 14:19:00	341.0	06/05/2018 12:28:53	...	96.0	11/08/2017 13:22:06	130.0	11/06/2017 14:18:22	122.0							
22	06/14/2018 13:21:11	236.0	06/06/2018 14:23:59	335.0	06/05/2018 12:33:54	...	95.0	11/08/2017 13:27:05	127.0	11/06/2017 14:23:21	124.0							
23	06/14/2018 13:26:10	241.0	06/06/2018 14:28:59	325.0	06/05/2018 12:38:54	...	94.0	11/08/2017 13:32:05	127.0	11/06/2017 14:28:21	131.0							
24	06/14/2018 13:31:10	246.0	06/06/2018 14:33:59	316.0	06/05/2018 12:43:53	...	91.0	11/08/2017 13:37:06	128.0	11/06/2017 14:33:22	139.0							
25	06/14/2018 13:36:10	252.0	06/06/2018 14:39:00	310.0	06/05/2018 12:48:53	...	87.0	11/08/2017 13:42:06	128.0	11/06/2017 14:38:22	145.0							
26	06/14/2018 13:41:11	254.0	06/06/2018 14:44:00	303.0	06/05/2018 12:53:54	...	84.0	11/08/2017 13:47:05	126.0	11/06/2017 14:43:22	150.0							
27	06/14/2018 13:46:11	258.0	06/06/2018 14:48:59	292.0	06/05/2018 12:58:54	...	79.0	11/08/2017 13:52:05	122.0	11/06/2017 14:48:21	152.0							
28	06/14/2018 13:51:10	260.0	06/06/2018 14:53:59	283.0	06/05/2018 13:03:54	...	78.0	11/08/2017 13:57:05	111.0	11/06/2017 14:53:21	156.0							
29	06/14/2018 13:56:10	258.0	06/06/2018 14:59:00	273.0	06/05/2018 13:08:53	...	78.0	11/08/2017 14:02:06	107.0	11/06/2017 14:58:22	160.0							
30	06/14/2018 14:01:11	256.0	06/06/2018 15:04:00	260.0	06/05/2018 13:13:53	...	78.0	11/08/2017 14:07:06	106.0	11/06/2017 15:03:22	164.0							

Patient 1 Merged Data

					Timestamp	Feature1
0	06/14/2018	11:31:11	-	06/14/2018	14:01:11	1005339.0
1	06/06/2018	12:33:59	-	06/06/2018	15:04:00	2870807.0
2	06/05/2018	10:43:54	-	06/05/2018	13:13:53	1249650.0
3	06/04/2018	11:13:49	-	06/04/2018	13:43:48	795109.0
4	05/21/2018	11:38:45	-	05/21/2018	14:08:44	626259.0
5	05/16/2018	12:04:40	-	05/16/2018	14:34:42	870645.0
6	03/21/2018	11:34:04	-	03/21/2018	14:04:05	636710.0
7	03/01/2018	11:14:02	-	03/01/2018	13:44:01	808197.0
8	02/26/2018	11:13:47	-	02/26/2018	13:43:46	671963.0
9	02/21/2018	11:19:40	-	02/21/2018	13:49:44	722144.0
10	02/16/2018	12:40:01	-	02/16/2018	15:10:02	1003306.0
11	02/15/2018	11:04:55	-	02/15/2018	13:34:56	801079.0
12	02/12/2018	11:49:39	-	02/12/2018	14:19:41	774825.0
13	02/06/2018	11:27:30	-	02/06/2018	13:57:28	741151.0
14	12/29/2017	11:38:48	-	12/29/2017	14:08:47	317078.0
15	12/27/2017	11:45:50	-	12/27/2017	14:15:50	727389.0
16	12/26/2017	12:05:45	-	12/26/2017	14:35:44	976319.0
17	12/24/2017	14:29:57	-	12/24/2017	16:59:56	785352.0
18	12/23/2017	12:19:49	-	12/23/2017	14:49:50	518234.0
19	12/19/2017	11:39:29	-	12/19/2017	14:09:29	732582.0
20	12/18/2017	11:30:02	-	12/18/2017	14:00:03	748782.0
21	12/15/2017	11:49:46	-	12/15/2017	14:19:47	251350.0
22	12/13/2017	11:34:35	-	12/13/2017	14:04:36	766739.0
23	12/12/2017	13:29:30	-	12/12/2017	15:59:29	391307.0
24	11/27/2017	11:34:55	-	11/27/2017	14:04:54	810657.0
25	11/22/2017	11:39:29	-	11/22/2017	14:09:29	332362.0
26	11/20/2017	12:03:59	-	11/20/2017	14:33:59	298449.0
27	11/17/2017	12:08:44	-	11/17/2017	14:38:45	598743.0
28	11/14/2017	11:28:28	-	11/14/2017	13:58:29	637512.0
29	11/11/2017	11:52:20	-	11/11/2017	14:22:21	634115.0
30	11/09/2017	11:22:08	-	11/09/2017	13:52:08	306480.0
31	11/08/2017	11:37:03	-	11/08/2017	14:07:06	439823.0
32	11/06/2017	12:33:21	-	11/06/2017	15:03:22	355604.0

Variance of Fast Fourier Transform of Patient 1 CGM Data

2. Average of Pre and Post meal CGM readings-

As we hypothesized, the average of post and pre meal CGM data gives an insight into the patient's random glucose levels. The following screen shot shows the Average Pre and Post Meal CGM data for Patient 1. It provides significant information about the time when the patient had higher average Pre and Post meal glucose level and when it was lesser.

				Timestamp	Feature2
0	06/14/2018	11:31:11	-	06/14/2018 14:01:11	101.795031
1	06/06/2018	12:33:59	-	06/06/2018 15:04:00	77.968944
2	06/05/2018	10:43:54	-	06/05/2018 13:13:53	92.378882
3	06/04/2018	11:13:49	-	06/04/2018 13:43:48	76.770186
4	05/21/2018	11:38:45	-	05/21/2018 14:08:44	29.478261
5	05/16/2018	12:04:40	-	05/16/2018 14:34:42	92.031056
6	03/21/2018	11:34:04	-	03/21/2018 14:04:05	50.198758
7	03/01/2018	11:14:02	-	03/01/2018 13:44:01	78.099379
8	02/26/2018	11:13:47	-	02/26/2018 13:43:46	57.198758
9	02/21/2018	11:19:40	-	02/21/2018 13:49:44	33.173913
10	02/16/2018	12:40:01	-	02/16/2018 15:10:02	75.285714
11	02/15/2018	11:04:55	-	02/15/2018 13:34:56	88.118012
12	02/12/2018	11:49:39	-	02/12/2018 14:19:41	57.695652
13	02/06/2018	11:27:30	-	02/06/2018 13:57:28	62.937888
14	12/29/2017	11:38:48	-	12/29/2017 14:08:47	27.770186
15	12/27/2017	11:45:50	-	12/27/2017 14:15:50	60.931677
16	12/26/2017	12:05:45	-	12/26/2017 14:35:44	53.099379
17	12/24/2017	14:29:57	-	12/24/2017 16:59:56	83.981366
18	12/23/2017	12:19:49	-	12/23/2017 14:49:50	57.552795
19	12/19/2017	11:39:29	-	12/19/2017 14:09:29	65.906832
20	12/18/2017	11:30:02	-	12/18/2017 14:00:03	48.931677
21	12/15/2017	11:49:46	-	12/15/2017 14:19:47	3.453416
22	12/13/2017	11:34:35	-	12/13/2017 14:04:36	21.720497
23	12/12/2017	13:29:30	-	12/12/2017 15:59:29	26.757764
24	11/27/2017	11:34:55	-	11/27/2017 14:04:54	60.813665
25	11/22/2017	11:39:29	-	11/22/2017 14:09:29	26.142857
26	11/20/2017	12:03:59	-	11/20/2017 14:33:59	27.677019
27	11/17/2017	12:08:44	-	11/17/2017 14:38:45	39.757764
28	11/14/2017	11:28:28	-	11/14/2017 13:58:29	44.832298
29	11/11/2017	11:52:20	-	11/11/2017 14:22:21	60.472050
30	11/09/2017	11:22:08	-	11/09/2017 13:52:08	6.590062
31	11/08/2017	11:37:03	-	11/08/2017 14:07:06	20.608696
32	11/06/2017	12:33:21	-	11/06/2017 15:03:22	36.652174

Average Pre and Post Meal CGM data

3. Hyperglycemic Percentage

As we hypothesized, the percentage of times when the patient was hyperglycemic can provide a significant insight into the patient's overall health. The following screenshot shows Percentage of times Patient 1 was hyperglycemic. We can see that in general the patient was hyperglycemic only a few of the times.

			Timestamp	Feature3
0	06/14/2018	11:31:11	- 06/14/2018 14:01:11	0.000000
1	06/06/2018	12:33:59	- 06/06/2018 15:04:00	0.000000
2	06/05/2018	10:43:54	- 06/05/2018 13:13:53	0.000000
3	06/04/2018	11:13:49	- 06/04/2018 13:43:48	0.000000
4	05/21/2018	11:38:45	- 05/21/2018 14:08:44	0.000000
5	05/16/2018	12:04:40	- 05/16/2018 14:34:42	0.000000
6	03/21/2018	11:34:04	- 03/21/2018 14:04:05	0.000000
7	03/01/2018	11:14:02	- 03/01/2018 13:44:01	0.000000
8	02/26/2018	11:13:47	- 02/26/2018 13:43:46	0.000000
9	02/21/2018	11:19:40	- 02/21/2018 13:49:44	0.000000
10	02/16/2018	12:40:01	- 02/16/2018 15:10:02	0.000000
11	02/15/2018	11:04:55	- 02/15/2018 13:34:56	0.000000
12	02/12/2018	11:49:39	- 02/12/2018 14:19:41	0.000000
13	02/06/2018	11:27:30	- 02/06/2018 13:57:28	0.000000
14	12/29/2017	11:38:48	- 12/29/2017 14:08:47	0.000000
15	12/27/2017	11:45:50	- 12/27/2017 14:15:50	0.000000
16	12/26/2017	12:05:45	- 12/26/2017 14:35:44	0.000000
17	12/24/2017	14:29:57	- 12/24/2017 16:59:56	12.903226
18	12/23/2017	12:19:49	- 12/23/2017 14:49:50	0.000000
19	12/19/2017	11:39:29	- 12/19/2017 14:09:29	0.000000
20	12/18/2017	11:30:02	- 12/18/2017 14:00:03	0.000000
21	12/15/2017	11:49:46	- 12/15/2017 14:19:47	0.000000
22	12/13/2017	11:34:35	- 12/13/2017 14:04:36	0.000000
23	12/12/2017	13:29:30	- 12/12/2017 15:59:29	3.225806
24	11/27/2017	11:34:55	- 11/27/2017 14:04:54	0.000000
25	11/22/2017	11:39:29	- 11/22/2017 14:09:29	0.000000
26	11/20/2017	12:03:59	- 11/20/2017 14:33:59	3.225806
27	11/17/2017	12:08:44	- 11/17/2017 14:38:45	0.000000
28	11/14/2017	11:28:28	- 11/14/2017 13:58:29	0.000000
29	11/11/2017	11:52:20	- 11/11/2017 14:22:21	0.000000
30	11/09/2017	11:22:08	- 11/09/2017 13:52:08	0.000000
31	11/08/2017	11:37:03	- 11/08/2017 14:07:06	0.000000
32	11/06/2017	12:33:21	- 11/06/2017 15:03:22	0.000000

Hyperglycemic Percentage for Patient 1 CGM Data

4. Hypoglycemic Percentage

As we hypothesized, the percentage of times when the patient was hypoglycemic can provide a significant insight into the patient's overall health as the following screenshot shows. We can see that the patient was hypoglycemic in higher percentage of times than he was hyperglycemic.

				Timestamp	Feature4
0	06/14/2018	11:31:11	-	06/14/2018 14:01:11	45.161290
1	06/06/2018	12:33:59	-	06/06/2018 15:04:00	100.000000
2	06/05/2018	10:43:54	-	06/05/2018 13:13:53	51.612903
3	06/04/2018	11:13:49	-	06/04/2018 13:43:48	41.935484
4	05/21/2018	11:38:45	-	05/21/2018 14:08:44	0.000000
5	05/16/2018	12:04:40	-	05/16/2018 14:34:42	48.387097
6	03/21/2018	11:34:04	-	03/21/2018 14:04:05	16.129032
7	03/01/2018	11:14:02	-	03/01/2018 13:44:01	38.709677
8	02/26/2018	11:13:47	-	02/26/2018 13:43:46	19.354839
9	02/21/2018	11:19:40	-	02/21/2018 13:49:44	22.580645
10	02/16/2018	12:40:01	-	02/16/2018 15:10:02	54.838710
11	02/15/2018	11:04:55	-	02/15/2018 13:34:56	38.709677
12	02/12/2018	11:49:39	-	02/12/2018 14:19:41	41.935484
13	02/06/2018	11:27:30	-	02/06/2018 13:57:28	35.483871
14	12/29/2017	11:38:48	-	12/29/2017 14:08:47	0.000000
15	12/27/2017	11:45:50	-	12/27/2017 14:15:50	32.258065
16	12/26/2017	12:05:45	-	12/26/2017 14:35:44	48.387097
17	12/24/2017	14:29:57	-	12/24/2017 16:59:56	45.161290
18	12/23/2017	12:19:49	-	12/23/2017 14:49:50	9.677419
19	12/19/2017	11:39:29	-	12/19/2017 14:09:29	32.258065
20	12/18/2017	11:30:02	-	12/18/2017 14:00:03	35.483871
21	12/15/2017	11:49:46	-	12/15/2017 14:19:47	0.000000
22	12/13/2017	11:34:35	-	12/13/2017 14:04:36	25.806452
23	12/12/2017	13:29:30	-	12/12/2017 15:59:29	0.000000
24	11/27/2017	11:34:55	-	11/27/2017 14:04:54	41.935484
25	11/22/2017	11:39:29	-	11/22/2017 14:09:29	0.000000
26	11/20/2017	12:03:59	-	11/20/2017 14:33:59	0.000000
27	11/17/2017	12:08:44	-	11/17/2017 14:38:45	9.677419
28	11/14/2017	11:28:28	-	11/14/2017 13:58:29	6.451613
29	11/11/2017	11:52:20	-	11/11/2017 14:22:21	22.580645
30	11/09/2017	11:22:08	-	11/09/2017 13:52:08	0.000000
31	11/08/2017	11:37:03	-	11/08/2017 14:07:06	0.000000
32	11/06/2017	12:33:21	-	11/06/2017 15:03:22	0.000000

Percentage Hypoglycemic for Patient 1 CGM Data

5. Standard Deviation

As we hypothesized, the standard deviation of the blood glucose level provides important information about patient's blood sugar stability. We can see the data varies significantly except for one specific time series.

				Timestamp	Feature5
0	06/14/2018	11:31:11	-	06/14/2018 14:01:11	67.847874
1	06/06/2018	12:33:59	-	06/06/2018 15:04:00	41.416440
2	06/05/2018	10:43:54	-	06/05/2018 13:13:53	59.793157
3	06/04/2018	11:13:49	-	06/04/2018 13:43:48	56.037393
4	05/21/2018	11:38:45	-	05/21/2018 14:08:44	16.942211
5	05/16/2018	12:04:40	-	05/16/2018 14:34:42	56.183678
6	03/21/2018	11:34:04	-	03/21/2018 14:04:05	34.300351
7	03/01/2018	11:14:02	-	03/01/2018 13:44:01	57.656361
8	02/26/2018	11:13:47	-	02/26/2018 13:43:46	33.749398
9	02/21/2018	11:19:40	-	02/21/2018 13:49:44	22.903271
10	02/16/2018	12:40:01	-	02/16/2018 15:10:02	51.703207
11	02/15/2018	11:04:55	-	02/15/2018 13:34:56	67.922003
12	02/12/2018	11:49:39	-	02/12/2018 14:19:41	37.412040
13	02/06/2018	11:27:30	-	02/06/2018 13:57:28	40.960680
14	12/29/2017	11:38:48	-	12/29/2017 14:08:47	13.766297
15	12/27/2017	11:45:50	-	12/27/2017 14:15:50	46.043842
16	12/26/2017	12:05:45	-	12/26/2017 14:35:44	38.221709
17	12/24/2017	14:29:57	-	12/24/2017 16:59:56	62.439015
18	12/23/2017	12:19:49	-	12/23/2017 14:49:50	40.905515
19	12/19/2017	11:39:29	-	12/19/2017 14:09:29	38.011445
20	12/18/2017	11:30:02	-	12/18/2017 14:00:03	30.375006
21	12/15/2017	11:49:46	-	12/15/2017 14:19:47	6.239247
22	12/13/2017	11:34:35	-	12/13/2017 14:04:36	27.377197
23	12/12/2017	13:29:30	-	12/12/2017 15:59:29	26.459794
24	11/27/2017	11:34:55	-	11/27/2017 14:04:54	38.147286
25	11/22/2017	11:39:29	-	11/22/2017 14:09:29	18.135861
26	11/20/2017	12:03:59	-	11/20/2017 14:33:59	19.797676
27	11/17/2017	12:08:44	-	11/17/2017 14:38:45	20.466770
28	11/14/2017	11:28:28	-	11/14/2017 13:58:29	29.054006
29	11/11/2017	11:52:20	-	11/11/2017 14:22:21	32.955051
30	11/09/2017	11:22:08	-	11/09/2017 13:52:08	15.323905
31	11/08/2017	11:37:03	-	11/08/2017 14:07:06	10.930677
32	11/06/2017	12:33:21	-	11/06/2017 15:03:22	31.598483

Standard Deviation of Patient 1 CGM Data

Task D: Feature Matrix Creation

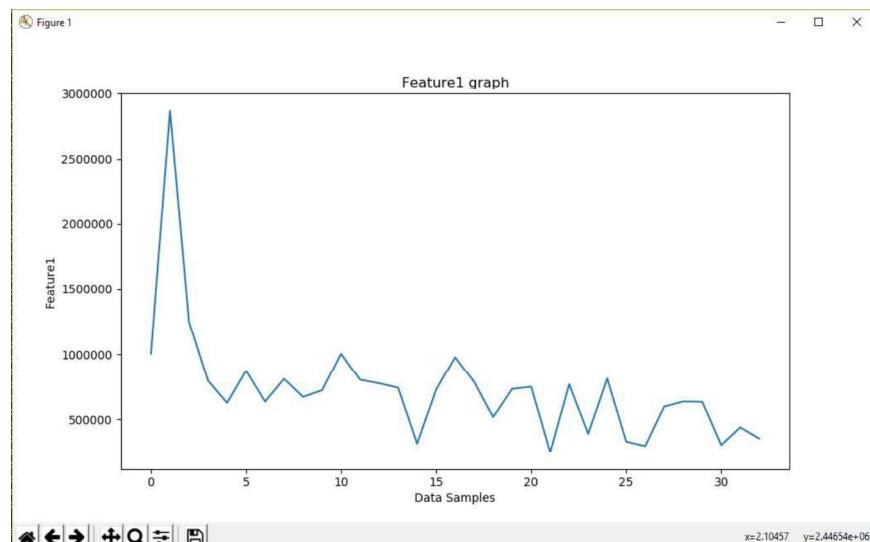
As mentioned in the earlier section, we extracted six features from our Patient 1 Merged CGM Data. So, the size of the feature matrix was 33 X 7 including the time stamps. The following is the screenshot of the Feature Matrix-

		Timestamp	Feature1	Feature2	Feature3	Feature4	Feature5	Feature6
0	06/14/2018 11:31:11	- 06/14/2018 14:01:11	1005339.0	101.795031	0.000000	45.161290	67.847874	167.612903
1	06/06/2018 12:33:59	- 06/06/2018 15:04:00	2870807.0	77.968944	0.000000	100.000000	41.416440	303.967742
2	06/05/2018 10:43:54	- 06/05/2018 13:13:53	1249650.0	92.378882	0.000000	51.612903	59.793157	192.935484
3	06/04/2018 11:13:49	- 06/04/2018 13:43:48	795109.0	76.770186	0.000000	41.935484	56.037393	150.935484
4	05/21/2018 11:38:45	- 05/21/2018 14:08:44	626259.0	29.478261	0.000000	0.000000	16.942211	142.838710
5	05/16/2018 12:04:40	- 05/16/2018 14:34:42	870645.0	92.031056	0.000000	48.387097	56.183678	158.677419
6	03/21/2018 11:34:04	- 03/21/2018 14:04:05	636710.0	50.198758	0.000000	16.129032	34.300351	140.258065
7	03/01/2018 11:14:02	- 03/01/2018 13:44:01	808197.0	78.099379	0.000000	38.709677	57.656361	151.741935
8	02/26/2018 11:13:47	- 02/26/2018 13:43:46	671963.0	57.198758	0.000000	19.354839	33.749398	144.451613
9	02/21/2018 11:19:40	- 02/21/2018 13:49:44	722144.0	33.173913	0.000000	22.580645	22.903271	152.612903
10	02/16/2018 12:40:01	- 02/16/2018 15:10:02	1003306.0	75.285714	0.000000	54.838710	51.703207	173.935484
11	02/15/2018 11:04:55	- 02/15/2018 13:34:56	801079.0	88.118012	0.000000	38.709677	67.022003	146.387097
12	02/12/2018 11:49:39	- 02/12/2018 14:19:41	774825.0	57.695652	0.000000	41.935484	37.412040	154.870968
13	02/06/2018 11:27:30	- 02/06/2018 13:57:28	741151.0	62.937888	0.000000	35.483871	40.960680	150.354839
14	12/29/2017 11:38:48	- 12/29/2017 14:08:47	317078.0	27.770186	0.000000	0.000000	13.766297	101.193548
15	12/27/2017 11:45:50	- 12/27/2017 14:15:50	727389.0	60.931677	0.000000	32.258065	46.043842	147.354839
16	12/26/2017 12:05:45	- 12/26/2017 14:35:44	976319.0	53.099379	0.000000	48.387097	38.221709	175.064516
17	12/24/2017 14:29:57	- 12/24/2017 16:59:56	785352.0	83.981366	12.903226	45.161290	62.439015	150.580645
18	12/23/2017 12:19:49	- 12/23/2017 14:49:50	518234.0	57.552795	0.000000	9.677419	40.905515	123.645161
19	12/19/2017 11:39:29	- 12/19/2017 14:09:29	732582.0	65.906832	0.000000	32.258065	38.011445	149.967742
20	12/18/2017 11:30:02	- 12/18/2017 14:00:03	748782.0	48.931677	0.000000	35.483871	30.375006	153.935484
21	12/15/2017 11:49:46	- 12/15/2017 14:19:47	251350.0	3.453416	0.000000	0.000000	6.239247	91.677419
22	12/13/2017 11:34:35	- 12/13/2017 14:04:36	7666739.0	21.720497	0.000000	25.806452	27.377197	157.193548
23	12/12/2017 13:29:30	- 12/12/2017 15:59:29	391307.0	26.757764	3.225806	0.000000	26.459794	110.516129
24	11/27/2017 11:34:55	- 11/27/2017 14:04:54	810657.0	60.813665	0.000000	41.935484	38.147286	158.451613
25	11/22/2017 11:39:29	- 11/22/2017 14:09:29	332362.0	26.142857	0.000000	0.000000	18.135861	103.161290
26	11/20/2017 12:03:59	- 11/20/2017 14:33:59	298449.0	27.677019	3.225806	0.000000	19.797676	99.709677
27	11/17/2017 12:08:44	- 11/17/2017 14:38:45	598743.0	39.757764	0.000000	9.677419	20.466770	141.419355
28	11/14/2017 11:28:28	- 11/14/2017 13:58:29	637512.0	44.832298	0.000000	6.451613	29.054006	141.838710
29	11/11/2017 11:52:20	- 11/11/2017 14:22:21	634115.0	60.472050	0.000000	22.580645	32.955051	143.645161
30	11/09/2017 11:22:08	- 11/09/2017 13:52:08	306480.0	6.590062	0.000000	0.000000	15.323905	99.870968
31	11/08/2017 11:37:03	- 11/08/2017 14:07:06	439823.0	20.608696	0.000000	0.000000	10.930677	119.935484
32	11/06/2017 12:33:21	- 11/06/2017 15:03:22	355604.0	36.652174	0.000000	0.000000	31.598483	103.290323

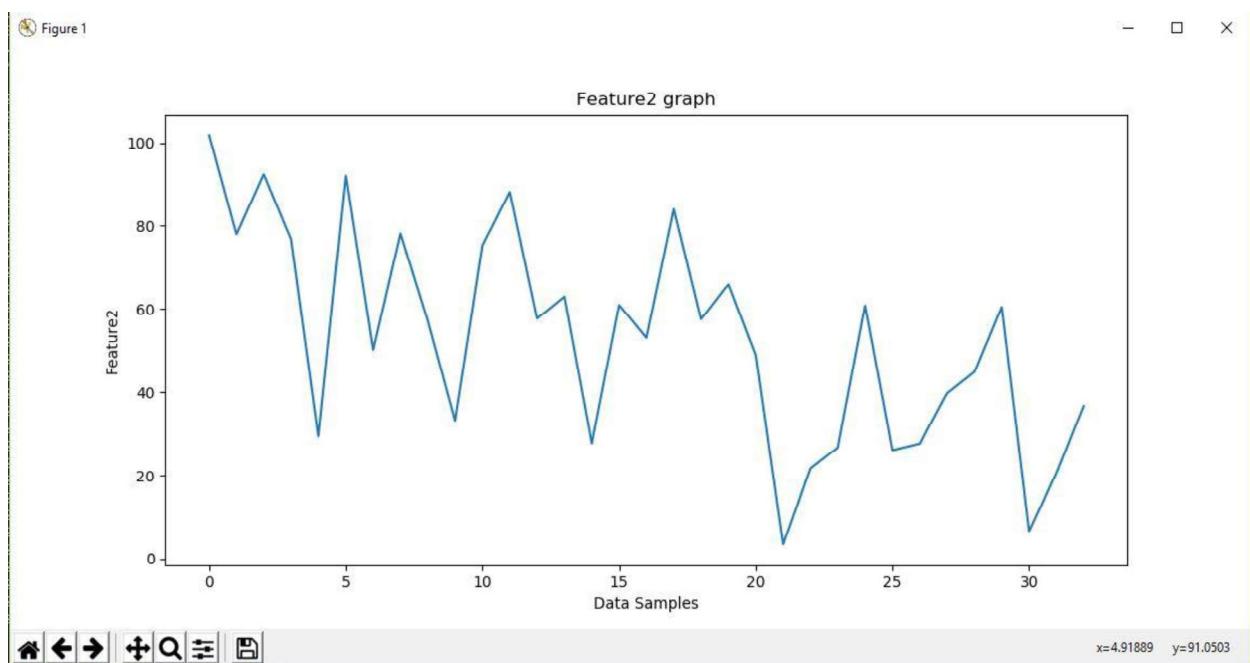
33 X 7 Feature Matrix

We also provide the plots from different features-

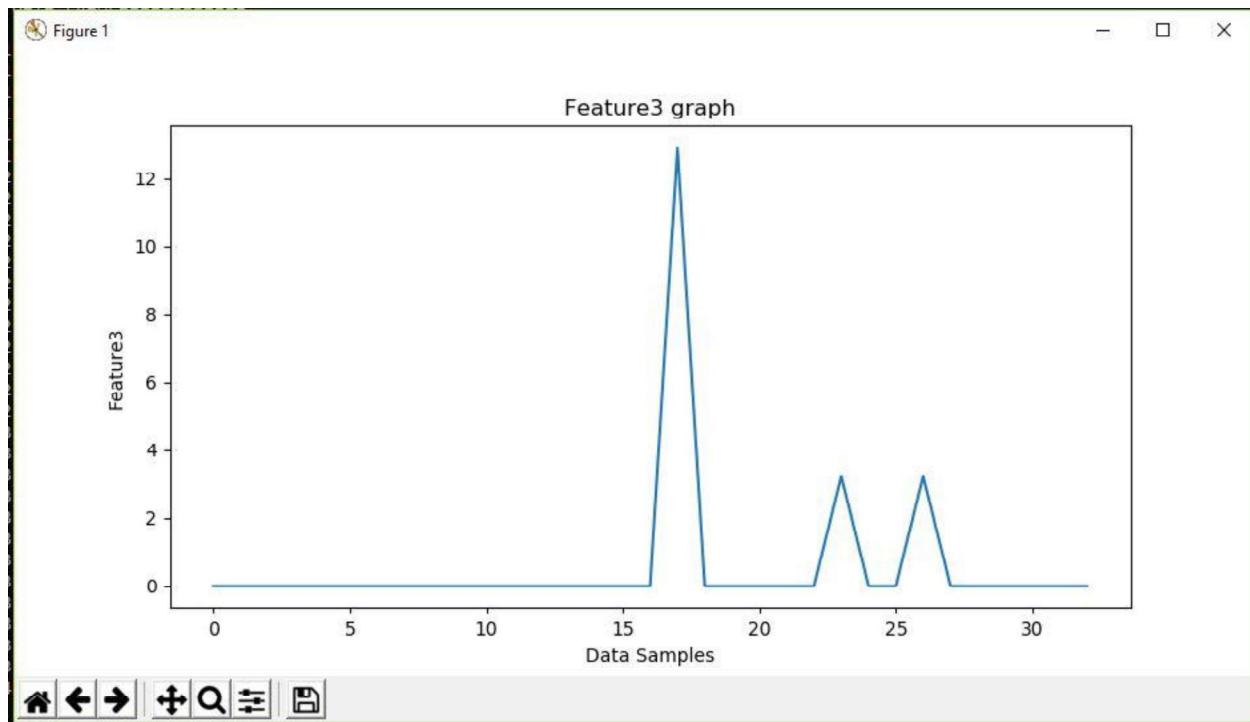
1. Variance of FFT



2. Average of Pre and Post meal CGM readings

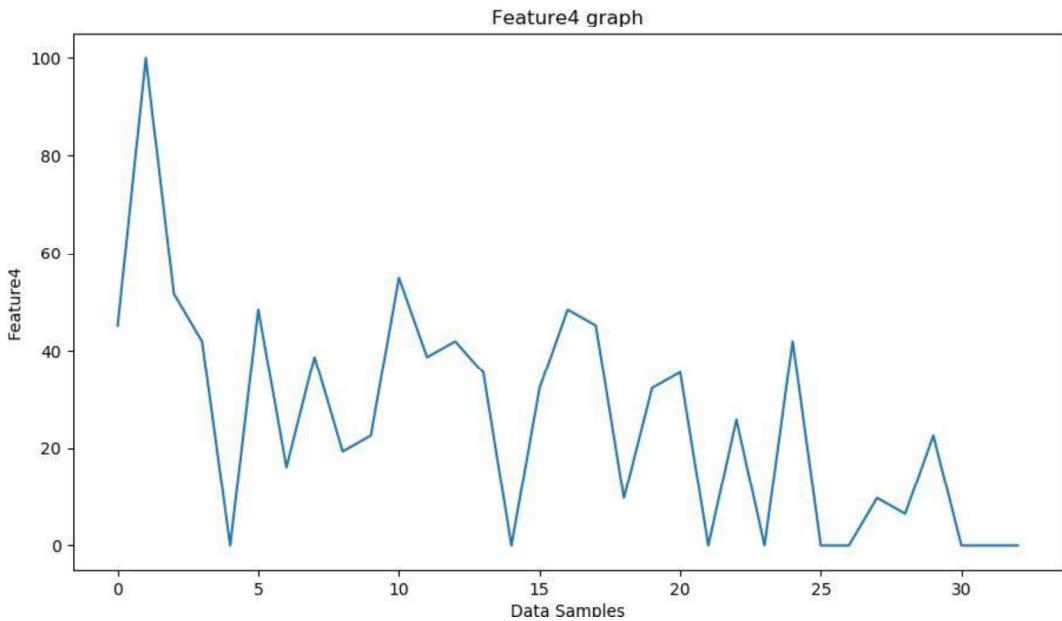


3. Hyperglycemic Percentage

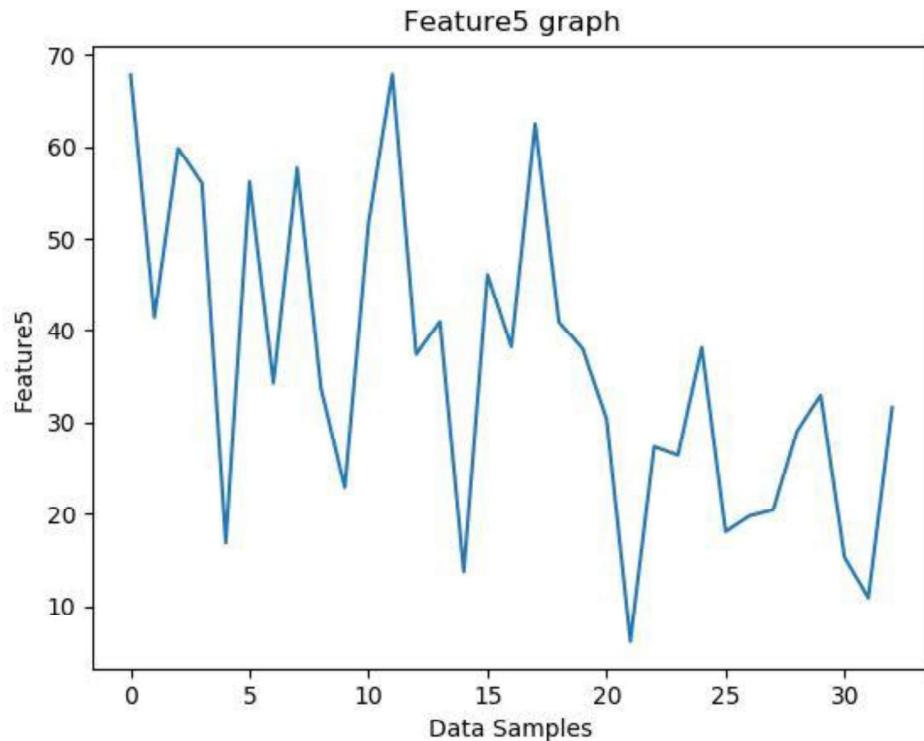


4. Hypoglycemic Percentage

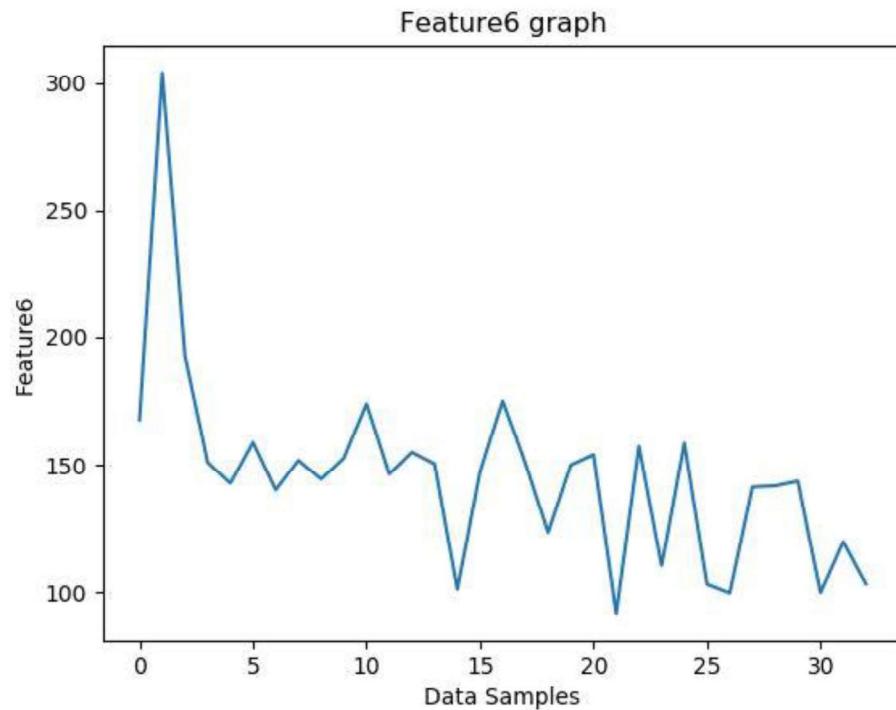
Figure 1



5. Standard Deviation



6. Mean



Task E: Feature Matrix to PCA

PCA-

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors (each being a linear combination of the variables and containing n observations) are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables.

So, we provide the feature matrix obtained in Task D to PCA in order to derive a new feature matrix. The new feature matrix derived is as follows-

	FEATURE MATRIX AFTER DIMENSIONALITY REDUCTION					Timestamp
	PC1	PC2	PC3	PC4	PC5	
0	2.644232	0.072727	-1.188127	-0.287959	0.046731	06/14/2018 11:31:11 - 06/14/2018 14:01:11
1	4.511511	-2.185749	3.090703	-0.404788	0.101515	06/06/2018 12:33:59 - 06/06/2018 15:04:00
2	2.685497	-0.307043	-0.407349	-0.250860	0.032656	06/05/2018 10:43:54 - 06/05/2018 13:13:53
3	1.491420	-0.011836	-0.783163	0.178248	0.119952	06/04/2018 11:13:49 - 06/04/2018 13:43:48
4	-1.784894	-0.322075	0.346232	-0.462879	-0.012850	05/21/2018 11:38:45 - 05/21/2018 14:08:44
5	2.033988	-0.047590	-0.844892	0.086978	-0.283109	05/16/2018 12:04:40 - 05/16/2018 14:34:42
6	-0.458403	-0.119155	-0.266988	-0.180960	0.039834	03/21/2018 11:34:04 - 03/21/2018 14:04:05
7	1.505841	0.028443	-0.870826	0.040065	0.186838	03/01/2018 11:14:02 - 03/01/2018 13:44:01
8	-0.221238	-0.146277	-0.273572	-0.216368	-0.172588	02/26/2018 11:13:47 - 02/26/2018 13:43:46
9	-0.913327	-0.456113	0.514007	0.144586	-0.000646	02/21/2018 11:19:40 - 02/21/2018 13:49:44
10	1.839057	-0.325187	-0.203618	0.372009	-0.018916	02/16/2018 12:40:01 - 02/16/2018 15:10:02
11	2.010976	0.215802	-1.329831	-0.065399	0.349199	02/15/2018 11:04:55 - 02/15/2018 13:34:56
12	0.523047	-0.330397	0.018926	0.421960	-0.169122	02/12/2018 11:49:39 - 02/12/2018 14:19:41
13	0.553859	-0.189280	-0.285448	0.180821	-0.121312	02/06/2018 11:27:30 - 02/06/2018 13:57:28
14	-2.226548	-0.138015	0.049071	-0.079303	-0.265352	12/29/2017 11:38:48 - 12/29/2017 14:08:47
15	0.576255	-0.095701	-0.457263	0.132173	0.161171	12/27/2017 11:45:50 - 12/27/2017 14:15:50
16	0.804898	-0.536813	0.413222	0.466430	0.036600	12/26/2017 12:05:45 - 12/26/2017 14:35:44
17	2.309712	5.085918	1.400970	0.047529	-0.057318	12/24/2017 14:29:57 - 12/24/2017 16:59:56
18	-0.378430	0.136755	-0.805514	-0.342576	0.114392	12/23/2017 12:19:49 - 12/23/2017 14:49:50
19	0.443616	-0.187705	-0.296163	0.031961	-0.297787	12/19/2017 11:39:29 - 12/19/2017 14:09:29
20	-0.042157	-0.398277	0.244915	0.341763	-0.193439	12/18/2017 11:30:02 - 12/18/2017 14:00:03
21	-3.016747	-0.293519	0.552074	0.325117	0.017037	12/15/2017 11:49:46 - 12/15/2017 14:19:47
22	-0.895337	-0.500248	0.660031	0.376174	0.477897	12/13/2017 11:34:35 - 12/13/2017 14:04:36
23	-1.689640	1.222740	0.413014	-0.158464	0.270749	12/12/2017 13:29:30 - 12/12/2017 15:59:29
24	0.645130	-0.334363	-0.002203	0.336811	-0.200230	11/27/2017 11:34:55 - 11/27/2017 14:04:54
25	-2.113990	-0.097124	-0.030910	-0.061782	-0.033948	11/22/2017 11:39:29 - 11/22/2017 14:09:29
26	-1.963432	1.206520	0.465728	-0.080016	-0.078770	11/20/2017 12:03:59 - 11/20/2017 14:33:59
27	-1.273102	-0.277459	0.164174	-0.242365	-0.229153	11/17/2017 12:08:44 - 11/17/2017 14:38:45
28	-0.947738	-0.146088	-0.147026	-0.444800	0.043324	11/14/2017 11:28:28 - 11/14/2017 13:58:29
29	-0.141529	-0.137069	-0.313774	-0.112138	-0.337656	11/11/2017 11:52:20 - 11/11/2017 14:22:21
30	-2.625376	-0.198580	0.316605	0.240530	0.341833	11/09/2017 11:22:08 - 11/09/2017 13:52:08
31	-2.334490	-0.299980	0.405621	-0.131198	-0.131966	11/08/2017 11:37:03 - 11/08/2017 14:07:06
32	-1.472662	0.112736	-0.548625	-0.201299	0.264435	11/06/2017 12:33:21 - 11/06/2017 15:03:22
-----END OF PRINCIPAL COMPONENT ANALYSIS-----						

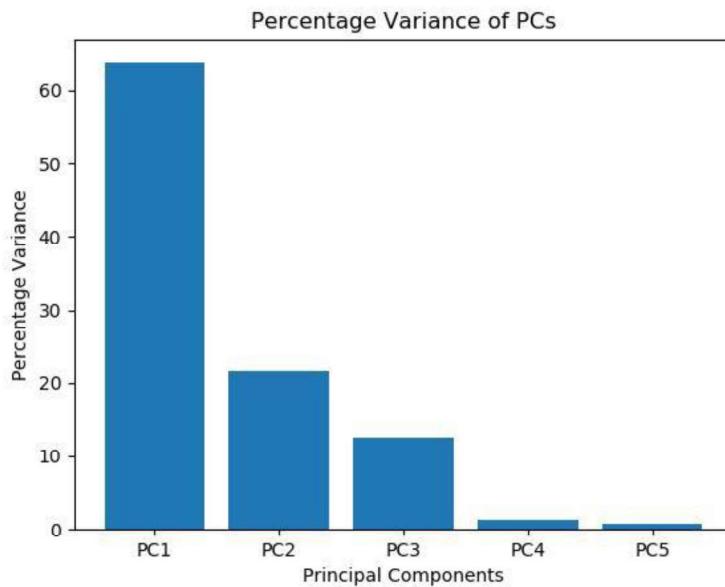
Feature Matrix derived from PCA

We take these top 5 features and plot them for each time series-

Task F: Top 5 Features in PCA

Principle Component Analysis (PCA) is used to emphasize variation and bring out strong patterns in a dataset. It is used to make data easy to explore. It makes the comparison among data values easier.

Following are the Top 5 Features derived by PCA. We provide the plots for those features and argument as to why those were chosen by PCA-



The mapping of the features to PCA features is as follows- (screenshot provided)

	θ	1
0	PC0	Feature4
1	PC1	Feature3
2	PC2	Feature1
3	PC3	Feature4
4	PC4	Feature5

- PC1- Percentage of Hypoglycemic readings

Since PCA finds the eigen vectors of a covariance matrix with the highest eigenvalues, hence it chooses the percentage of hypoglycemic readings to be the top feature. As seen in the earlier section, Percentage of Hypoglycemic readings has the most varied data.

- PC2- Percentage of Hyperglycemic readings

In the percentage of hyperglycemic values, there were some significant diverse values, hence that is chosen by the PCA to be top feature as well.

- PC3- Variance of FFT

Variance of FFT also had some values that were significantly diverse than the rest, hence this is also chosen to be a top feature by PCA.

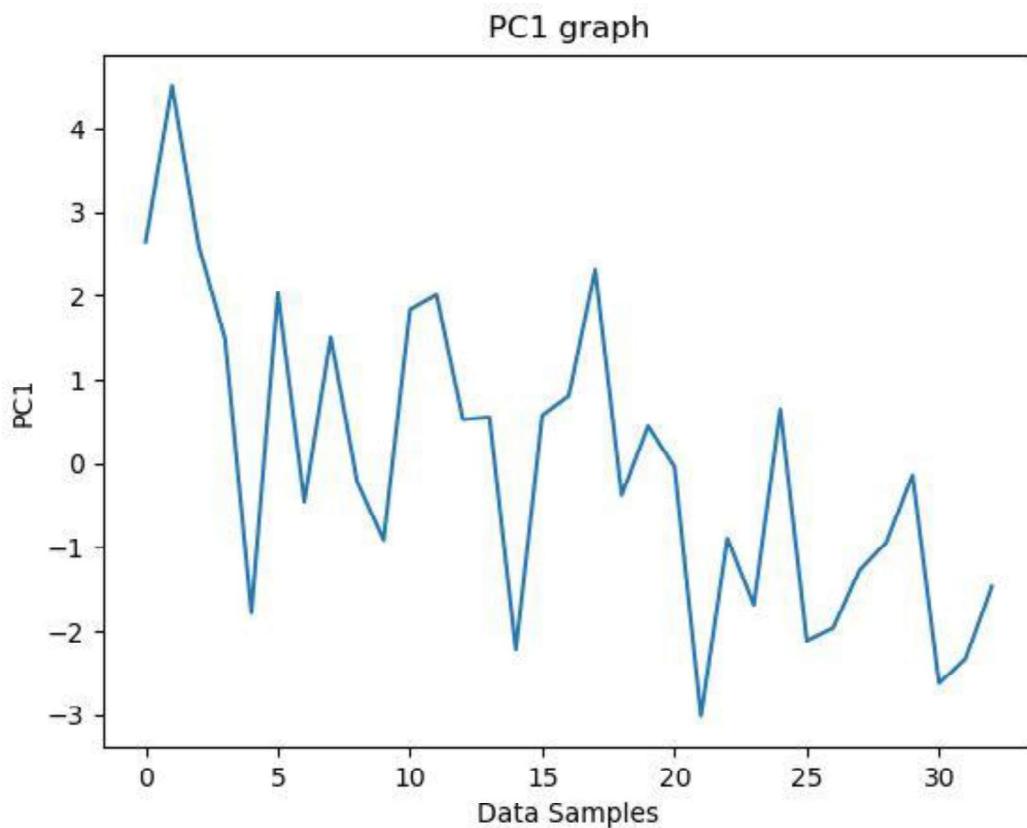
- PC4- Percentage of Hyperglycemic readings

Percentage of hyperglycemic can also be less significant if the distribution was regular. Hence the percentage of variance is lower, but also chosen to be one of the top 5 features.

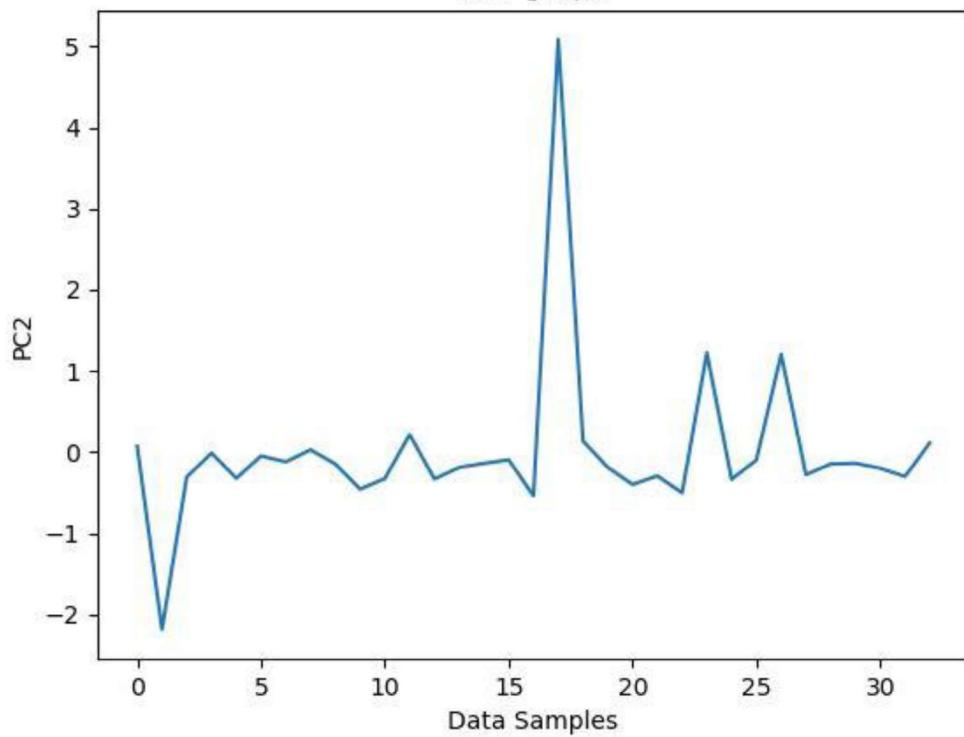
- PC5- Standard Deviation of CGM data

Standard deviation of CGM data provides lowest percentage variance of data among the top features.

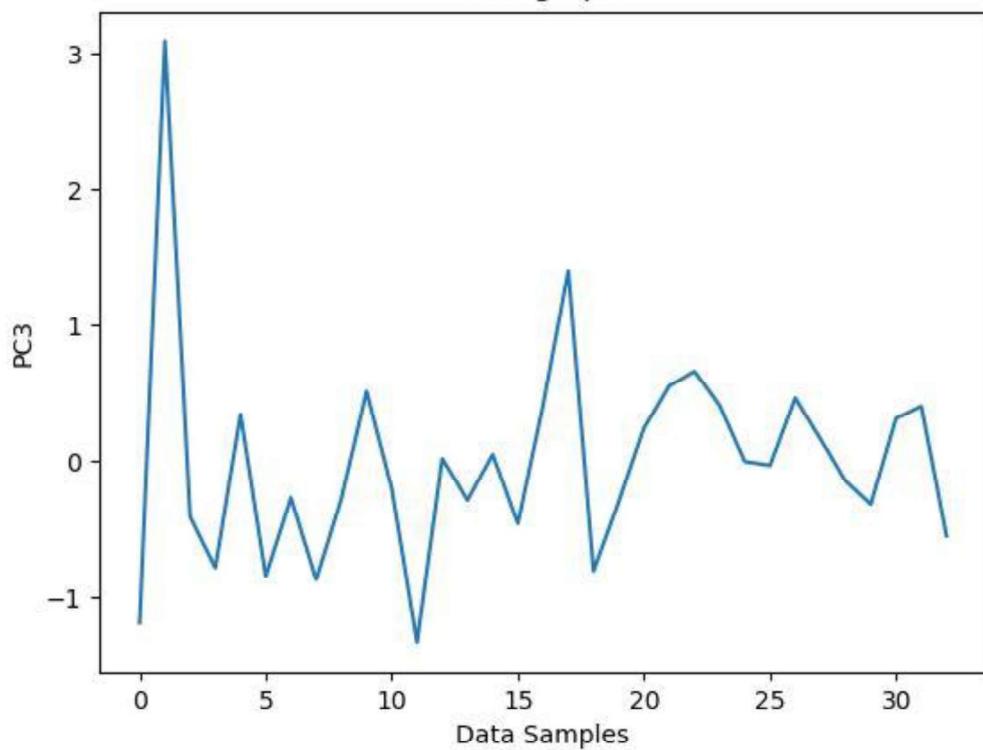
The following are the PCA feature plots-



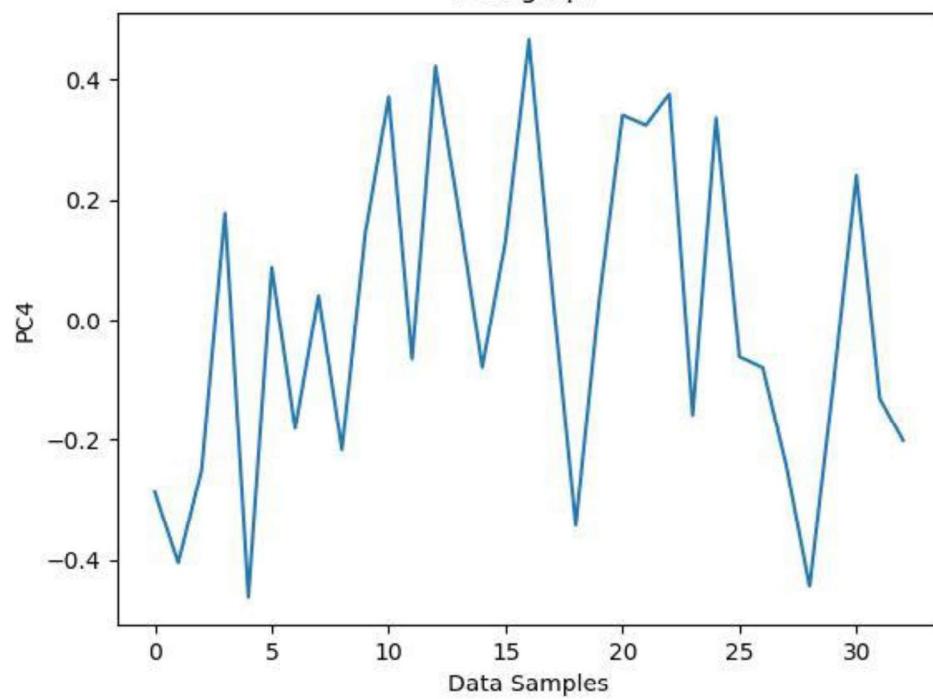
PC2 graph



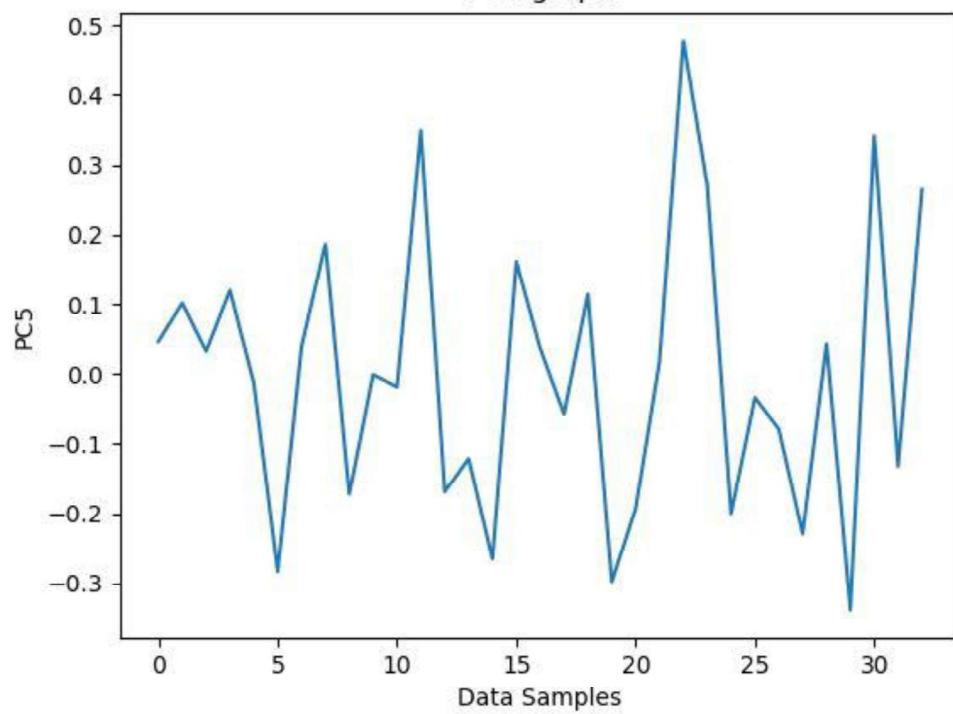
PC3 graph



PC4 graph

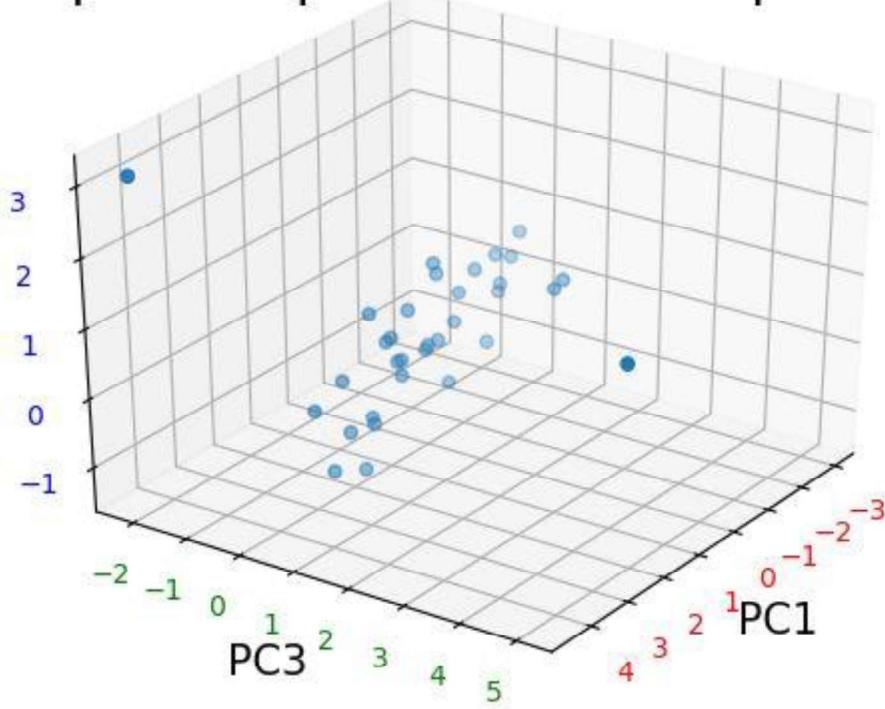


PC5 graph



The top 3 PCA features comparison is shown in following graph-

Top 3 components for comparing



References:

1. Clarke, W., & Kovatchev, B. (2009). Statistical tools to analyze continuous glucose monitor data. *Diabetes technology & therapeutics*, 11 Suppl 1(Suppl 1), S45–S54. doi:10.1089/dia.2008.0138