Bio Tech Market Segmentation Report

* By Sourav Nanda

**Fermi Estimation (Breakdown of Problem Statement)**

We are a team working under a Bio-Tech Startup going to launch its Home Check-Up Service with Online Booking offering the following initial services:

1. Full Body Check-Up with a Bio-Tech Device based on Blood Samples

2. Online Health Techs offering

i. Diabetes check-up device  
ii. Blood Pressure check-up device

iii. Vitamins deficiency check-up device

We have to analyse Medical Market in India with respect to the given problem statement using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use their product in terms of Geographic, Demographic, Psychographic, Behavioural.

**Data Sources (Data Collection)**

The quality of empirical data is critical for developing a valid segmentation solution. Organizations collect data to make appropriate evaluation of their existing consumer base, their market reputation and public opinion, market trends, customer preferences and competitors. Empirical data for segmentation studies can come from a range of sources:

* from survey studies
* from observations such as scanner data where purchases are recorded
* loyalty programs
* experimental studies

We have datasets from government sites, free data warehouses and Kaggle. Some of the sites we have collected our data from are noted below:

* <https://data.mendeley.com/>
* <https://www.ieee.org/>
* <https://www.kaggle.com/>

**Libraries Used (Python Modules)**

NumPy - for computational math

Pandas - for working on dataframes

Seaborn - for visualizing relations and statistics of data

matplotlib.pyplot - for visualizing relations and statistics of data

sklearn - for encoding categorical data, for unsupervised clustering to form segments

**Data Pre-processing (Feature Engineering)**

Diabetes Datasets Summery

Bangladesh\_UCI\_Diabetes\_Dataset

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 520 entries, 0 to 519

Data columns (total 17 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

# Logs -

* First, we extracted only useful columns: 'Age', 'Gender', 'Polyuria', 'Polydipsia', 'Polyphagia', 'partial paresis',' Alopecia', 'Obesity', 'class'.
* Then we changed all the object columns to numerical, such that we can perform segmentation on them
* Then we removed outliers present in ‘Age’ column.

0 Age 520 non-null int64

1 Gender 520 non-null object

2 Polyuria 520 non-null object

3 Polydipsia 520 non-null object

4 sudden weight loss 520 non-null object

5 weakness 520 non-null object

6 Polyphagia 520 non-null object

7 Genital thrush 520 non-null object

8 visual blurring 520 non-null object

9 Itching 520 non-null object

10 Irritability 520 non-null object

11 delayed healing 520 non-null object

12 partial paresis 520 non-null object

13 muscle stiffness 520 non-null object

14 Alopecia 520 non-null object

15 Obesity 520 non-null object

16 class 520 non-null object

dtypes: int64(1), object(16)

memory usage: 69.2+ KB

# ****India\_Diabetes\_Dataset\_2019****

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 952 entries, 0 to 951

# Logs -

* First, we removed Rows containing Null Values.
* Then we extracted only useful columns: 'Age', 'Gender', 'Family\_Diabetes',' highBP', 'Smoking',' Alcohol','Diabetic',’ BMI’, ’Stress’, ’BPLevel’, ’UriationFreq’.
* Then we changed all the object columns to numerical, such that we can perform segmentation on them
* We added new features like: obese\_meter
* Then we removed outliers present in ‘BMI’ column.

Data columns (total 18 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Age 952 non-null object

1 Gender 952 non-null object

2 Family\_Diabetes 952 non-null object

3 highBP 952 non-null object

4 PhysicallyActive 952 non-null object

5 BMI 948 non-null float64

6 Smoking 952 non-null object

7 Alcohol 952 non-null object

8 Sleep 952 non-null int64

9 SoundSleep 952 non-null int64

10 RegularMedicine 952 non-null object

11 JunkFood 952 non-null object

12 Stress 952 non-null object

13 BPLevel 952 non-null object

14 Pregancies 910 non-null float64

15 Pdiabetes 951 non-null object

16 UriationFreq 952 non-null object

17 Diabetic 951 non-null object

dtypes: float64(2), int64(2), object(14)

memory usage: 134.0+ KB

# ****North\_India\_Health\_Data\_500****

# Logs -

* First, we removed all unnecessary columns: 'Height(cms)','Skeleton Muscle(%age)','Body Age(yrs)','Waist(cms)','RM(Kcal)',
* Then we changed all the object columns to numerical, such that we can perform segmentation on them
* We added new features like: obese\_meter, hypertension\_meter, bloodsugar\_meter
* Then we removed outliers present in 'Birth Age(yrs)', 'Weight(kg)', 'Body Fat', 'Sugar PP' columns.

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 500 entries, 0 to 499

Data columns (total 16 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Birth Age(yrs) 500 non-null int64

1 Height(cms) 500 non-null float64

2 Gender 500 non-null int64

3 Weight(kg) 500 non-null float64

4 Body Fat 500 non-null float64

5 Visceral Fat 500 non-null float64

6 Skeleton Muscle(%age) 500 non-null float64

7 Body Age(yrs) 500 non-null int64

8 RM(Kcal) 500 non-null int64

9 BMI 500 non-null float64

10 Systolic BP 500 non-null int64

11 Diasttolic BP 500 non-null int64

12 Pulse 500 non-null int64

13 Suger Fasting 500 non-null int64

14 Sugar PP 500 non-null int64

15 Waist(cms) 500 non-null float64

dtypes: float64(7), int64(9)

memory usage: 62.6 KB

# ****Pima\_Diabetes\_Database****

<class 'pandas.core.frame.DataFrame'>

# Logs -

* First, we removed all unnecessary columns: 'Pregnancies','SkinThickness','DiabetesPedigreeFunction'.
* We added new features like: obese\_meter, hypertension\_meter, bloodsugar\_meter
* Then we removed outliers present in 'Glucose','BloodPressure','Insulin','BMI','Age' columns.

RangeIndex: 768 entries, 0 to 767

Data columns (total 9 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Pregnancies 768 non-null int64

1 Glucose 768 non-null int64

2 BloodPressure 768 non-null int64

3 SkinThickness 768 non-null int64

4 Insulin 768 non-null int64

5 BMI 768 non-null float64

6 DiabetesPedigreeFunction 768 non-null float64

7 Age 768 non-null int64

8 Outcome 768 non-null int64

dtypes: float64(2), int64(7)

memory usage: 54.1 KB

Hypertension Datasets Summery

# ****North\_India\_Health\_Data\_500****

We have already Preprocessed this dataset above, so we use the preprocessed version of this dataset.

# ****HDHI Admission Dataset****

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 15757 entries, 0 to 15756

Data columns (total 56 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 SNO 15757 non-null int64

1 MRD No. 15757 non-null object

2 D.O.A 15757 non-null object

3 D.O.D 15757 non-null object

4 AGE 15757 non-null int64

# Logs -

* First we removed all unnecessary columns : 'SNO', 'MRD No.', 'D.O.A', 'D.O.D', 'TYPE OF ADMISSION-EMERGENCY/OPD', 'month year', 'DURATION OF STAY', 'duration of intensive unit stay', 'OUTCOME','SHOCK','PULMONARY EMBOLISM', 'CHEST INFECTION','UTI', 'NEURO CARDIOGENIC SYNCOPE', 'ORTHOSTATIC','INFECTIVE ENDOCARDITIS', 'DVT', 'AKI', 'CVA INFRACT', 'CVA BLEED', 'AF', 'VT', 'PSVT','ACS','SEVERE ANAEMIA', 'ANAEMIA','EF','CREATININE','CKD','STABLE ANGINA','ATYPICAL CHEST PAIN','CONGENITAL','UREA','TLC','PLATELETS'
* Then we removed all rows containing null values.
* We added new features like heart\_disease, by combining all columns of heart diseases.
* Then we removed outliers present in 'AGE','HB','GLUCOSE' columns.

5 GENDER 15757 non-null object

6 RURAL 15757 non-null object

7 TYPE OF ADMISSION-EMERGENCY/OPD 15757 non-null object

8 month year 15757 non-null object

9 DURATION OF STAY 15757 non-null int64

10 duration of intensive unit stay 15757 non-null int64

11 OUTCOME 15757 non-null object

12 SMOKING 15757 non-null int64

13 ALCOHOL 15757 non-null int64

14 DM 15757 non-null int64

15 HTN 15757 non-null int64

16 CAD 15757 non-null int64

17 PRIOR CMP 15757 non-null int64

18 CKD 15757 non-null int64

19 HB 15501 non-null float64

20 TLC 15467 non-null float64

21 PLATELETS 15463 non-null float64

22 GLUCOSE 14812 non-null float64

23 UREA 15513 non-null float64

24 CREATININE 15506 non-null float64

25 BNP 6676 non-null float64

26 RAISED CARDIAC ENZYMES 15757 non-null int64

27 EF 14158 non-null float64

28 SEVERE ANAEMIA 15757 non-null int64

29 ANAEMIA 15757 non-null int64

30 STABLE ANGINA 15757 non-null int64

31 ACS 15757 non-null int64

32 STEMI 15757 non-null int64

33 ATYPICAL CHEST PAIN 15757 non-null int64

34 HEART FAILURE 15757 non-null int64

35 HFREF 15757 non-null int64

36 HFNEF 15757 non-null int64

37 VALVULAR 15757 non-null int64

38 CHB 15757 non-null int64

39 SSS 15757 non-null int64

40 AKI 15757 non-null int64

41 CVA INFRACT 15757 non-null int64

42 CVA BLEED 15757 non-null int64

43 AF 15757 non-null int64

44 VT 15757 non-null int64

45 PSVT 15757 non-null int64

46 CONGENITAL 15757 non-null int64

47 UTI 15757 non-null int64

48 NEURO CARDIOGENIC SYNCOPE 15757 non-null int64

49 ORTHOSTATIC 15757 non-null int64

50 INFECTIVE ENDOCARDITIS 15757 non-null int64

51 DVT 15757 non-null int64

52 CARDIOGENIC SHOCK 15757 non-null int64

53 SHOCK 15757 non-null int64

54 PULMONARY EMBOLISM 15757 non-null int64

55 CHEST INFECTION 15757 non-null object

dtypes: float64(8), int64(39), object(9)

memory usage: 6.7+ MB

Cardiovascular Disease Datasets Summery

# ****CardioVascular Disease Dataset****

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1000 entries, 0 to 999

Data columns (total 14 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 patientid 1000 non-null int64

1 age 1000 non-null int64

2 gender 1000 non-null int64

3 chestpain 1000 non-null int64

4 restingBP 1000 non-null int64

5 serumcholestrol 1000 non-null int64

6 fastingbloodsugar 1000 non-null int64

# Logs -

* We removed all unnecessary columns: 'patientid','exerciseangia', 'oldpeak', 'slope', 'noofmajorvessels','chestpain'.
* The dataset is already clean so we didn’t require further pre-processing.

7 restingrelectro 1000 non-null int64

8 maxheartrate 1000 non-null int64

9 exerciseangia 1000 non-null int64

10 oldpeak 1000 non-null float64

11 slope 1000 non-null int64

12 noofmajorvessels 1000 non-null int64

13 target 1000 non-null int64

dtypes: float64(1), int64(13)

memory usage: 109.5 KB

**Data Analysis (Visualization)**

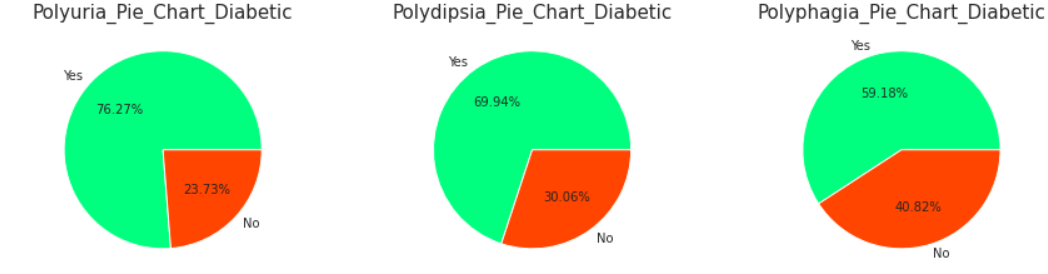
Diabetes Datasets Analysis

Bangladesh\_UCI\_Diabetes\_Dataset

We analysed the data after splitting the dataset into diabetic and non-diabetic patients. The following graphs are highlights of the analysis :

# 

The mean age for people suffering from diabetes is found to be : 48.58. From the box plot it is evident that the Age group suffering from diabetes is : 39 – 57 years.

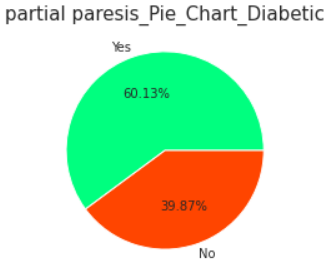


Its seen that among people suffering from Diabetes majority of them suffer from :

Polyuria (Excessive urination) : 76 %

Polydisia (Excess thirst): 70 %

Polyphagia (Excessive hunger): 59 %

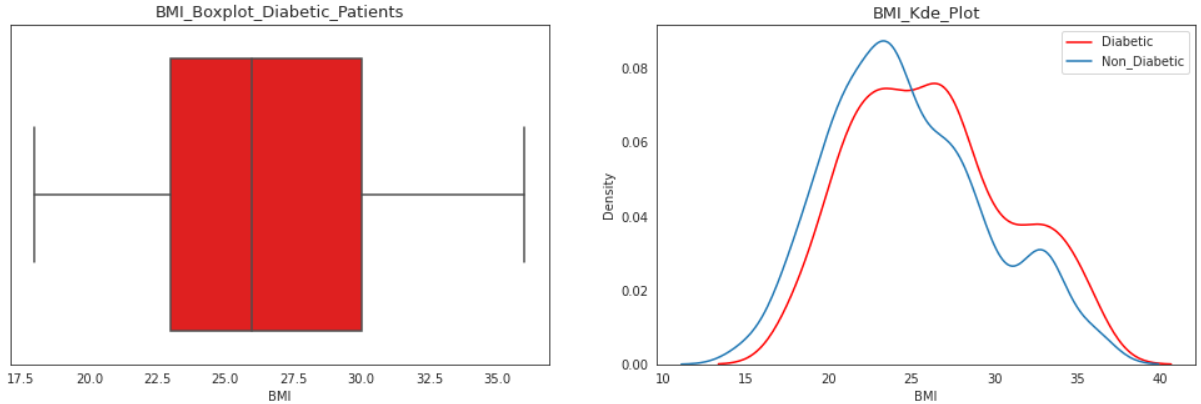


60% of Diabetic Patients are also seen to be suffering from Partial Paresis. (Incomplete Paralysis)

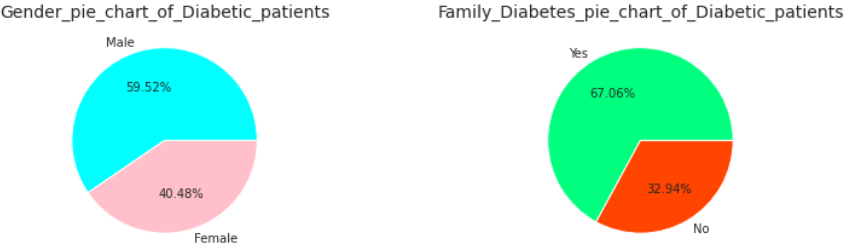
We have removed all those features which are irrelevant (and) or non-related with the disease & analysed only those features which can be medically examined.

# ****India\_Diabetes\_Dataset\_2019****

We analysed the data after splitting the dataset into diabetic and non-diabetic patients. The following graphs are highlights of the analysis :

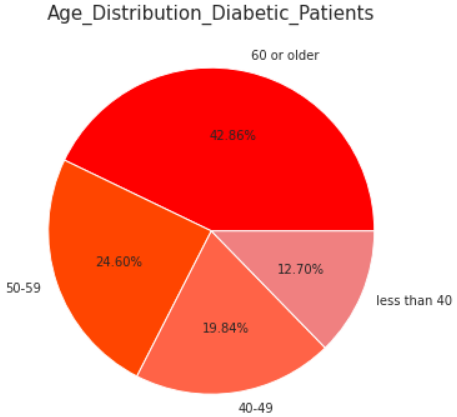


The mean BMI for people suffering from diabetes is found to be : 26.21. From the box plot it is evident that the BMI Range suffering from diabetes is : 23 – 30 , which says a lot of them are overweight.



Its seen that Male (60%) are more likely to suffer diabetes than females. And also if someone’s family has a diabetes background then they are 67% more likely to suffer from diabetes.

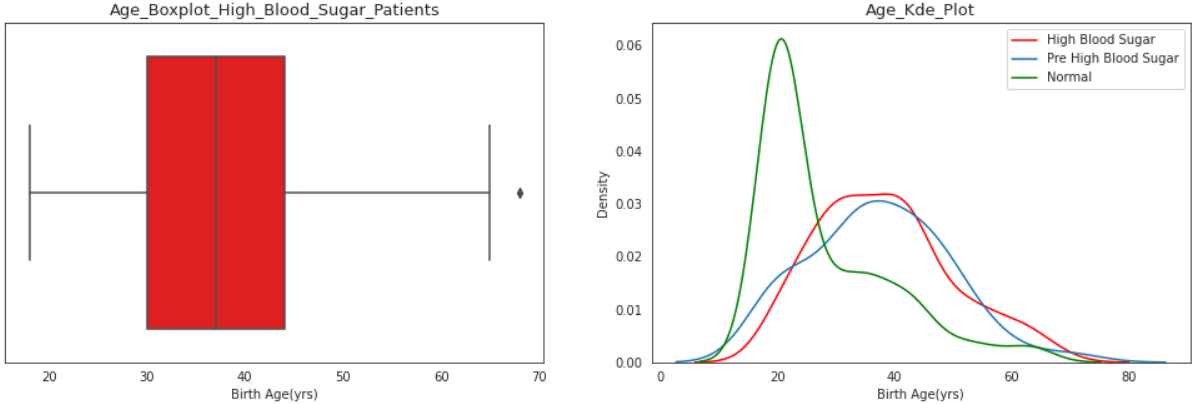
Age group 50-59 and 60+ are majority who suffer from diabetes in this dataset.



We have removed all those features which are irrelevant (and) or non-related with the disease & analysed only those features which can be medically examined.

**North\_India\_Health\_Data\_500**

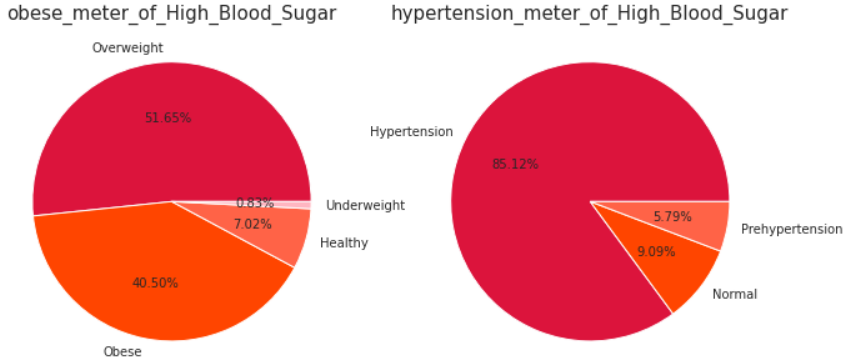
We analysed the data after splitting the dataset into Normal and Pre High Blood Sugar & High Blood Sugar patients. The following graphs are highlights of the analysis :



The mean age for people suffering from Pre High Blood Sugar is found to be : 37.25. From the box plot it is evident that the Age group suffering from Pre High Blood Sugar is : 28 – 45 years.

The mean age for people suffering from High Blood Sugar is found to be : 38.23. From the box plot it is evident that the Age group suffering from High Blood Sugar is : 30 – 44 years.

In this dataset the Age group suffering from diabetes is quite young compared to other datasets.



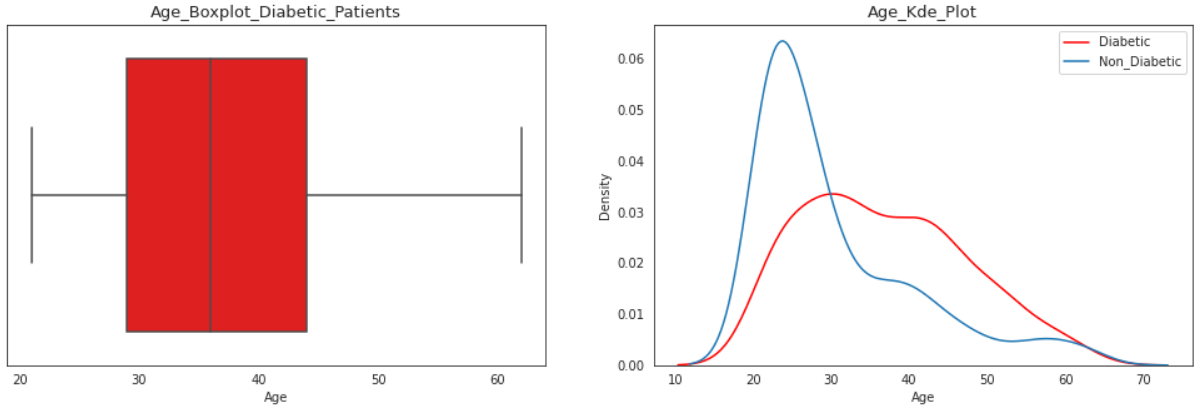
From above Pie charts its clear that:

* 92% of High Blood Sugar Patients are either Overweight or Obese
* 85 % of High Blood Sugar Patients suffer from Hypertension

We have removed all those features which are irrelevant (and) or non-related with the disease & analysed only those features which can be medically examined.

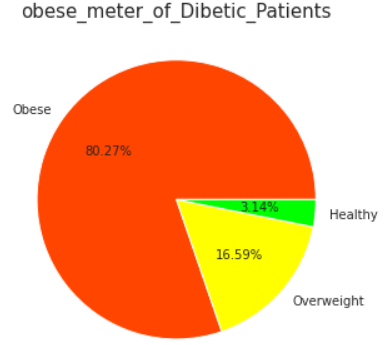
# ****Pima\_Diabetes\_Database****

We analysed the data after splitting the dataset into diabetic and non-diabetic patients. The following graphs are highlights of the analysis :



The mean age for people suffering from diabetes is found to be : 36.86. From the box plot it is evident that the Age group suffering from diabetes is : 29 – 44 years.

It represents the middle aged group (working group).



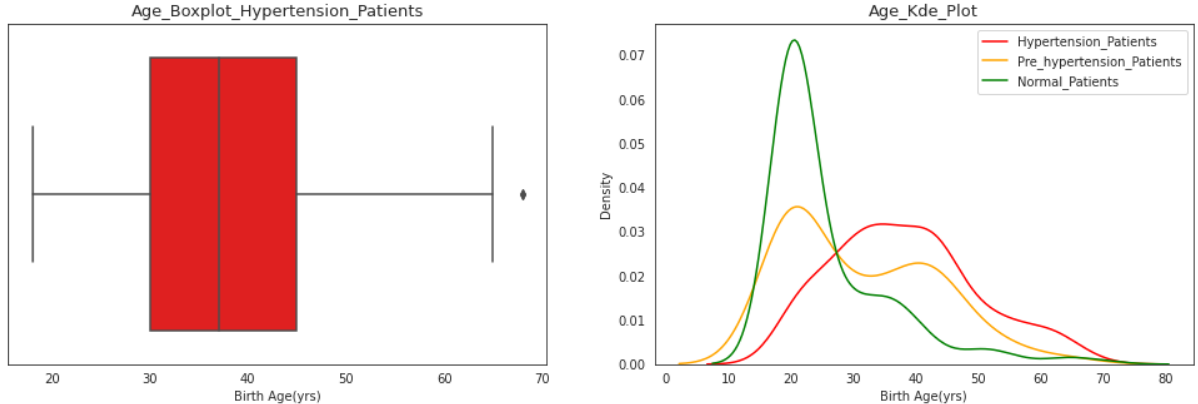
It’s Observed that 80% of diabetic patients are obese and 17% are overweight. So Diabetic Patients are also patients of obesity.

We have removed all those features which are irrelevant (and) or non-related with the disease & analysed only those features which can be medically examined.

Hypertension Datasets Analysis

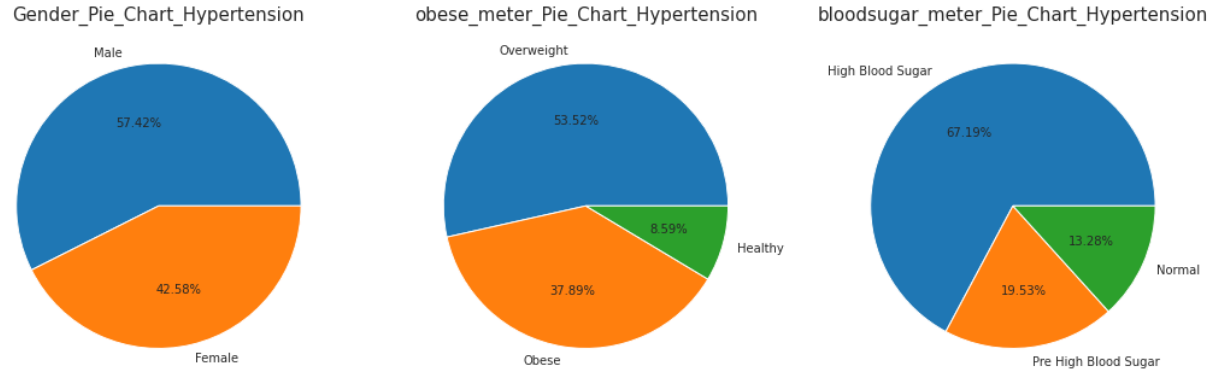
# ****North\_India\_Health\_Data\_500****

We analysed the data after splitting the dataset into Hypertension, Pre hypertension and normal patients. The following graphs are highlights of the analysis :



The mean age for people suffering from Pre Hypertension is found to be : 31.21. From the box plot it is evident that the Age group suffering from Pre Hypertension is : 21 – 42 years.

The mean age for people suffering from Hypertension is found to be : 38.28. From the box plot it is evident that the Age group suffering from Hypertension is : 30 – 45 years.



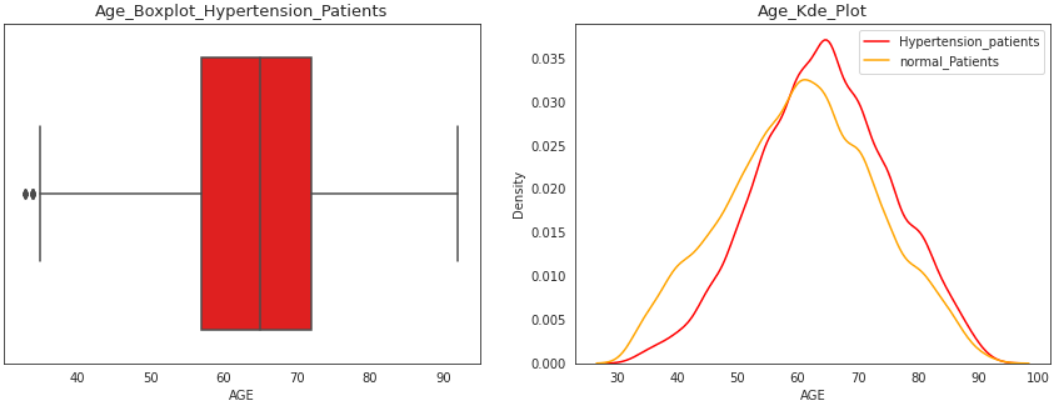
From above pie charts it’s clear that:

* Male (57 %) are more prone to Hypertension
* Overweight (54%) and Obese (38%) patients frequently suffer from Hypertension
* Patients with High Blood Sugar (67%) also suffer from Hypertension

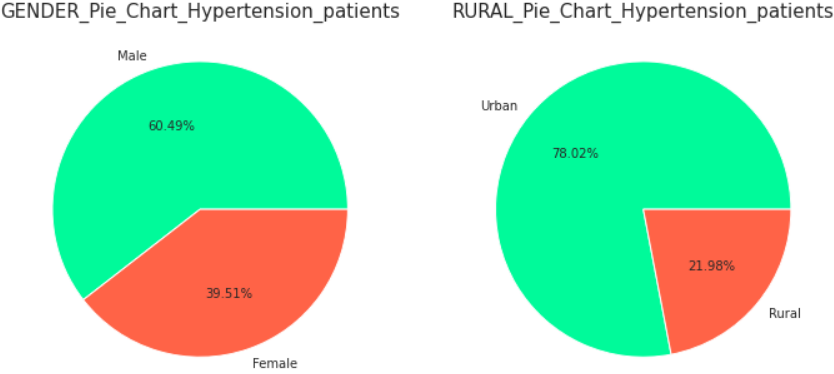
We have removed all those features which are irrelevant (and) or non-related with the disease & analysed only those features which can be medically examined.

# ****HDHI Admission Dataset****

We analysed the data after splitting the dataset into hypertension and normal patients. The following graphs are highlights of the analysis :



The mean age for people suffering from hypertension is found to be : 64.34. From the box plot it is evident that the Age group suffering from hypertension is : 57 – 72 years.



From the pie charts it’s clear that:

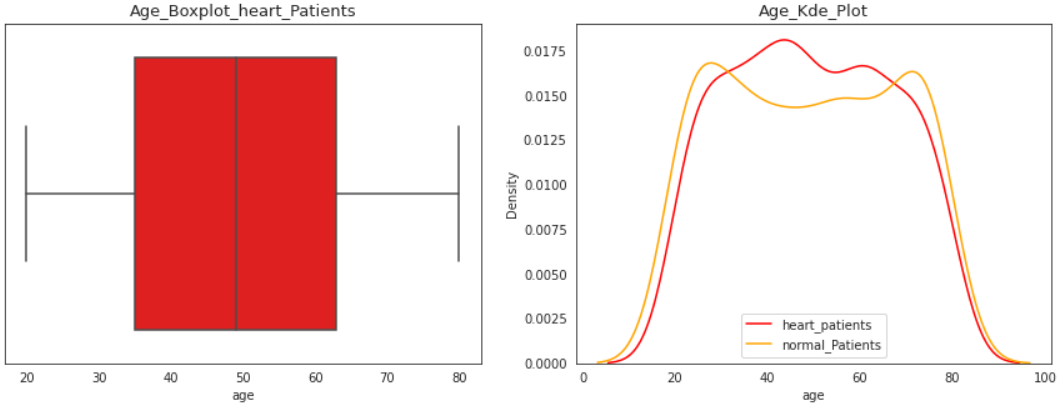
* Male (60%) are more prone to Hypertension
* People from Urban areas are 78% more likely to get Hypertension.

We have removed all those features which are irrelevant (and) or non-related with the disease & analysed only those features which can be medically examined.

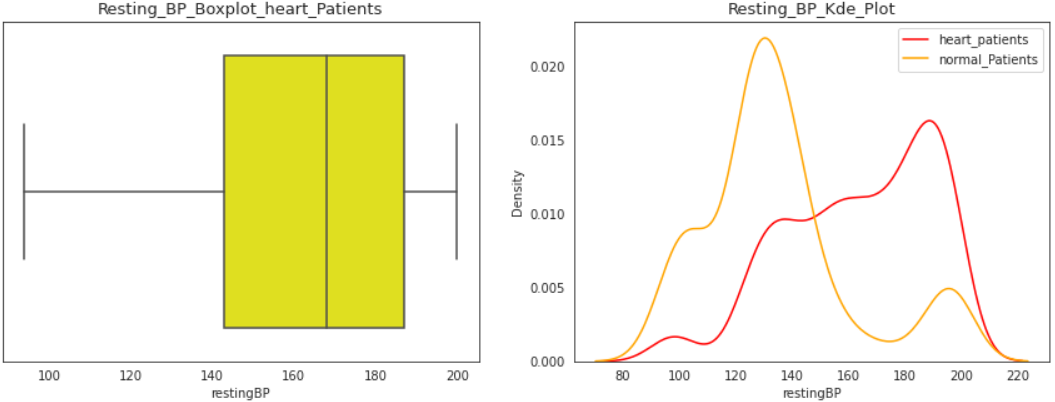
Cardiovascular Disease Datasets Analysis

# ****CardioVascular Disease Dataset****

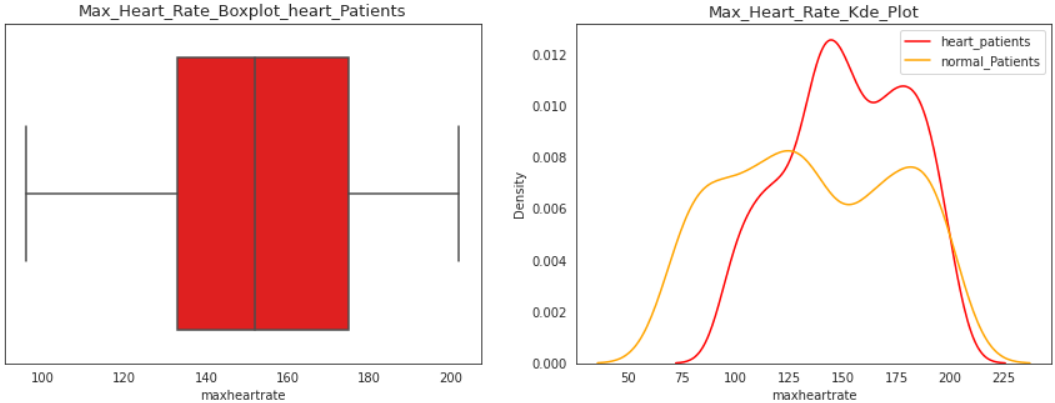
We analysed the data after splitting the dataset into Heart Patients and Normal Patients. The following graphs are highlights of the analysis :



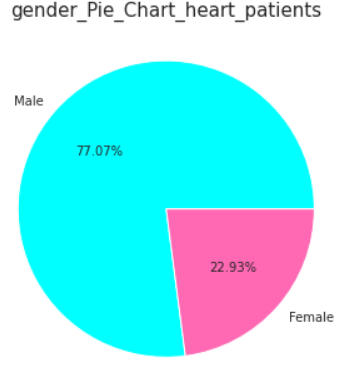
The mean age for people suffering from heart disease is found to be : 49.36. From the box plot it is evident that the Age group suffering from heart disease is : 35 – 63 years.



The mean Resting BP for people suffering from heart disease is found to be : 164.04. From the box plot it is evident that the Resting BP range suffering from heart disease is : 143 – 187 . This implies people suffering from heart diseases have high Resting BP.



The mean heart rate for people suffering from heart disease is found to be : 152.11. From the box plot it is evident that the heart rate range suffering from heart disease is : 133 – 175 . This implies people suffering from heart diseases tend to have high heart rate.

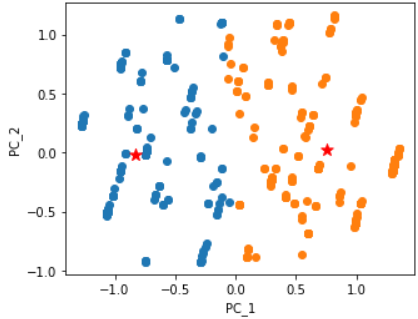


From the pie chart it’s clear that there are more Male (77%) heart patients in the dataset.

**Segment Extraction (K-Means Clustering)**

Diabetes Datasets Segmentation

Bangladesh\_UCI\_Diabetes\_Dataset



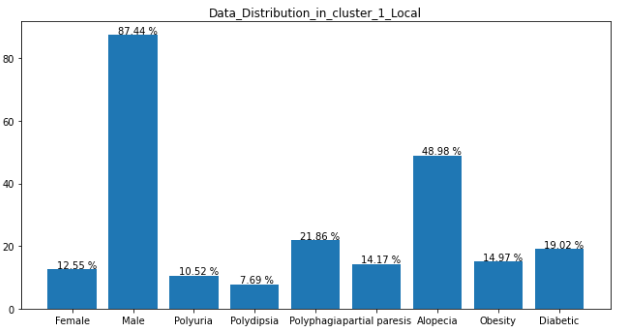
It is a scatter plot of Principal\_component\_1 Vs Principal\_component\_2 of the dataset.

Number of Clusters = 2

*Cluster\_1 : (Mostly Non-Diabetic Patients)*

It contains 14.87% of all diabetic patients in the whole dataset. Means it is the cluster of majorly non-diabetic patients.

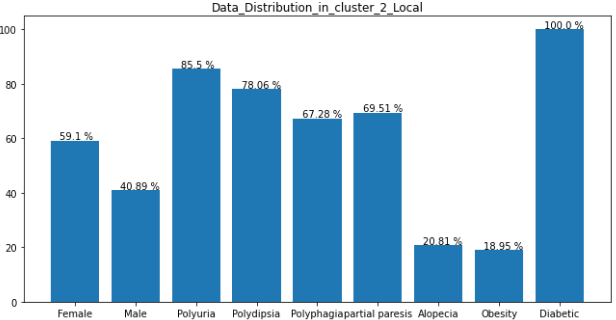
The Age distribution is : mean Age = 45.97, target range = 37 - 54



*Cluster\_2 : (Diabetic Patients)*

It contains 85.12% of all diabetic patients in the whole dataset. Means it is the cluster of Diabetic patients.

The Age distribution is : mean Age = 49.42, target range = 40 - 57



# ****India\_Diabetes\_Dataset\_2019****

# 

It is a scatter plot of Principal\_component\_1 Vs Principal\_component\_2 of the dataset.

Number of Clusters = 2

*Cluster\_1 : (Diabetic Patients)*

It contains 94.04% of all diabetic patients in the whole dataset. Means it is the cluster of Diabetic patients.

Age distribution : 43% are 60+ and 29% are between 50 – 59, take up the majority.

BP Level distribution : 60% of the people in cluster suffer from High Blood Pressure.

# 

*Cluster\_2 : (Non-Diabetic Patients)*

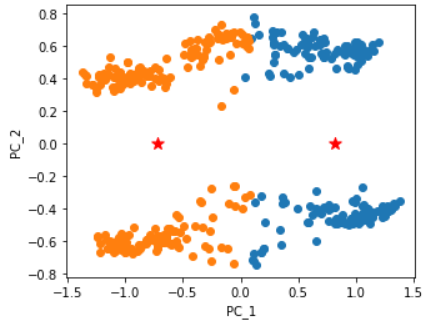
It contains 5.95% of all diabetic patients in the whole dataset. Means it is the cluster of Non-Diabetic patients.

Age distribution : 73.21% are less than 40 years old, take up the majority.

BP Level distribution : 91% of the people in cluster suffer from Pre High Blood Pressure.

# 

**North\_India\_Health\_Data\_500**

****

It is a scatter plot of Principal\_component\_1 Vs Principal\_component\_2 of the dataset.

Number of Clusters = 2

*Cluster\_1 : (Healthy Patients)*

It contains 85% of all Normal Blood Sugar Patients & 97.64% of all Normal Blood Pressure Patients in the whole dataset. Means it is the cluster of Healthy patients.

Age distribution :

* Mean = 27
* Target range = 20 - 33

Pulse distribution :

* Mean = 82.26
* Target range = 73 – 90

Weight : 71% of the population has Normal Weight.

*Cluster\_2 : (Diabetes & Hypertension Patients)*

It contains 95.87% of all High Blood Sugar Patients & 94.53% of all High Blood Pressure Patients in the whole dataset. Means it is the cluster of Healthy patients.

Age distribution :

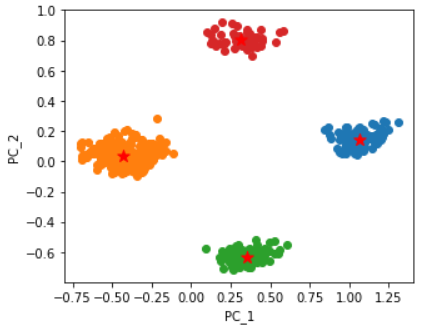
* Mean = 38.33
* Target range = 30 - 45

Pulse distribution :

* Mean = 146.21
* Target range = 87 - 99

Weight : 55.46% of the population are Overweight, & 39.84% of the population are Obese.

# ****Pima\_Diabetes\_Database****



It is a scatter plot of Principal\_component\_1 Vs Principal\_component\_2 of the dataset.

Number of Clusters = 4

*Cluster\_1 : (Diabetic Patients)*

It contains 48% of all Diabetic Patients of the hole dataset, but the cluster is a pure set of only diabetic patients. Means it is the cluster of Diabetic patients.

Age distribution :

* Mean = 38.31
* Target range = 29 - 47

Insulin distribution :

* Mean = 131.16
* Target range = 0 – 210

Glucose distribution :

* Mean = 167.50
* Target range = 152 - 181

Weight : 85.32% of the population are Obese.

*Cluster\_2 : (Non - Diabetic Patients)*

It contains 0% of all Diabetic Patients of the hole dataset, and the cluster is a pure set of only non - diabetic patients. Means it is the cluster of Non - Diabetic patients.

Age distribution :

* Mean = 29.80
* Target range = 22 - 34

Insulin distribution :

* Mean = 62.20
* Target range = 0 - 101

Glucose distribution :

* Mean = 104.20
* Target range = 91 - 118

Weight : 51.5% of the population are Obese & 27.5% are Overweight.

*Cluster\_3 : (Diabetic Patients)*

It contains 51.12% of all Diabetic Patients of the hole dataset, but the cluster is a pure set of only diabetic patients. Means it is the cluster of Diabetic patients.

Age distribution :

* Mean = 35.47
* Target range = 28 - 43

Insulin distribution :

* Mean = 68.42
* Target range = 0 - 140

Glucose distribution :

* Mean = 115.97
* Target range = 106 - 128

Weight : 75.43% of the population are Obese & 22.8% are Overweight.

*Cluster\_4 : (Non - Diabetic Patients)*

It contains 0% of all Diabetic Patients of the hole dataset, and the cluster is a pure set of only non - diabetic patients. Means it is the cluster of Non - Diabetic patients.

Age distribution :

* Mean = 35.07
* Target range = 24 - 41

Insulin distribution :

* Mean = 126.94
* Target range = 0 - 193

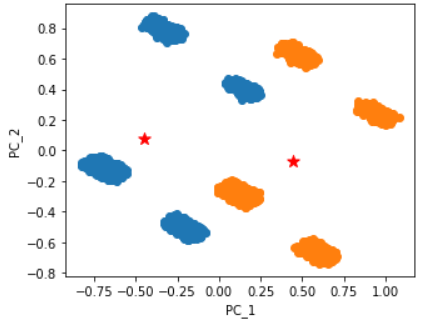
Glucose distribution :

* Mean = 154.25
* Target range = 144 - 158

Weight : 61.11% of the population are Obese & 29.62% are Overweight.

Hypertension Datasets Analysis

# ****HDHI Admission Dataset****



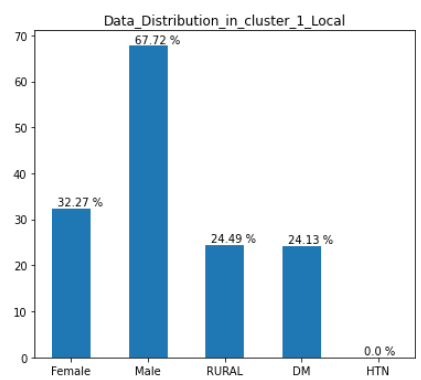
It is a scatter plot of Principal\_component\_1 Vs Principal\_component\_2 of the dataset.

Number of Clusters = 2

*Cluster\_1 : (Healthy Patients)*

It contains 0% of all Hypertension patients in the whole dataset. Means it is the cluster of healthy patients.

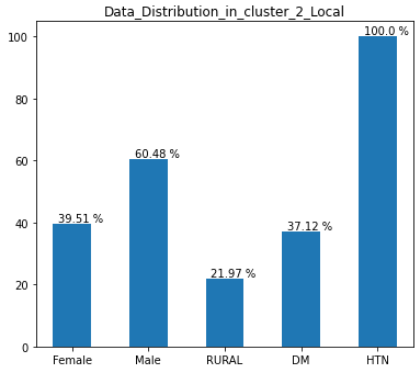
The Age distribution is : mean Age = 60.60, target range = 52 – 70



*Cluster\_2 : (Hypertension Patients)*

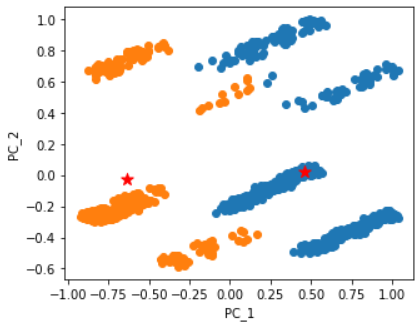
It contains 100% of all Hypertension patients in the whole dataset. Means it is the cluster of Hypertension patients.

The Age distribution is : mean Age = 64.34, target range = 57 - 72



Cardiovascular Disease Datasets Analysis

# ****CardioVascular Disease Dataset****



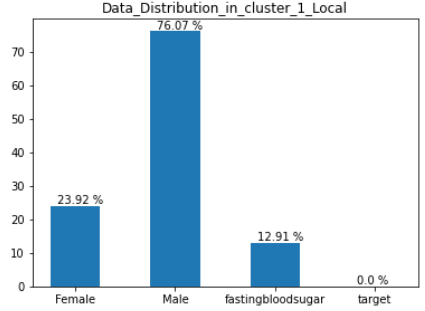
It is a scatter plot of Principal\_component\_1 Vs Principal\_component\_2 of the dataset.

Number of Clusters = 2

*Cluster\_1 : (Healthy Patients)*

It contains 0% of all Heart patients in the whole dataset. Means it is the cluster of healthy patients.

The Age distribution is :



* mean = 49.46
* target range = 35 – 63

The Resting BP distribution is :

* mean = 164.06
* target range = 143 - 187

The Serum Cholestrol distribution is :

* mean = 333.84
* target range = 241 - 456

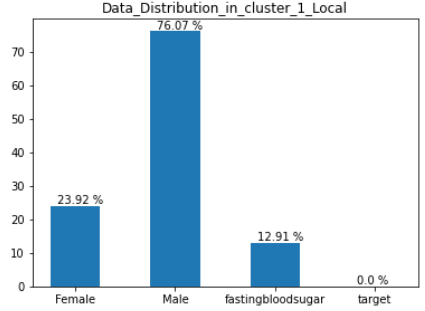
The Max Heart Rate distribution is :

* mean = 152.16
* target range = 133 - 175

*Cluster\_2 : (Heart Patients)*

It contains 100% of all Heart patients in the whole dataset. Means it is the cluster of heart patients.

The Age distribution is :



* mean = 48.93
* target range = 32 - 66

The Resting BP distribution is :

* mean = 134.59
* target range = 122 - 142

The Serum Cholestrol distribution is :

* mean = 280.26
* target range = 230 - 345

The Max Heart Rate distribution is :

* mean = 136.16
* target range = 103 - 171

**Profiling & Describing Potential Segments**

Diabetes

* Diabetic Patients tend to suffer from other diseases like : Partial Paresis , Polyuria , Polydisia & Polyphagia.
* The Mean age across all datasets for people suffering from diabetes came out to be : 41.22 years. This indicates middle-aged and older people are more prone to diabetes.
* If a family has Diabetes roots, then the person is more likely to have diabetes.
* Male are more prone to diabetes than Females
* Large number of Overweight & Obese people suffer from Diabetes
* People suffering from Hypertension are also found to be patient of Diabetes.

Hypertension

* The Mean age across all datasets for people suffering from hypertension came out to be : 51.31 years. This indicates older people are more prone to hypertension.
* Male are more prone to hypertension than Females
* Large number of Overweight & Obese people suffer from hypertension.
* People suffering from High Blood Sugar are also found to be patient of hypertension.
* People living in urban areas are more likely to have Hypertension.

Heart Disease

* The Mean age across all datasets for people suffering from heart disease came out to be : 49.36 years. This indicates middle aged people are more prone to heart disease.
* Male are more prone to heart disease than Females
* Heart patients are observed to have high resting BP

**Selection Of Target Segment**

Diabetes

* Target Age : 35 – 60, mostly middle aged and older people.
* Target Sex : Male (dominant), Female (less dominant)
* People Suffering from Partial Paresis , Polyuria , Polydisia & Polyphagia.
* High Blood Pressure patients (systolic: 140 mm Hg or higher, diastolic: 90 mm Hg or higher)
* People with diabetic family background
* BMI > 25 (Overweight and Obese People)
* Blood Sugar Level 126 mg/dL or above.

Hypertension

* Target Age : 55 or Above, mostly older people.
* Target Sex : Male (dominant), Female (less dominant)
* High Blood Pressure patients (systolic: 140 mm Hg or higher, diastolic: 90 mm Hg or higher)
* People Belonging to URBAN areas

Heart Disease

* Target Age : 32 - 66, mostly middle aged and older people.
* Target Age : 55 or Above, mostly older people.
* Cholestrol range = 230 – 345

**Finding Out Target Market Region (Fermi Estimation)**

Diabetes

Top Highest Diabetes States in India:

* Tamil Nadu : With 13% of urban and 3.5% of rural population is suffering from diabetes.
* Punjab : 4.6% of Population suffering from diabetes.
* Karnataka: 7.5% of Population suffering from diabetes.
* Kerela : The most cases of diabetes in Kerala fall under 45 to 69 years category. That counts to 19.4% of people suffering from Diabetes.
* Gujarat :  That accounts for 7.1% of diabetic cases in Gujarat for India.

Considering The Population, the most amount of diabetic patients reside in:

* Tamil Nadu (5.8 million patients)
* Karnataka (4.8million)
* Kerela (6.7 million patients)
* Gujarat (4.4 million)

The Internet User Base in :

* Tamil Nadu (93% in urban and 40% in rural) = 4.6 million Potential Customers
* Karnataka (9.3%) = 0.4 million Potential Customers
* Kerela (54%) = 3.6 million Potential Customers
* Gujarat (10%) = 0.4 million Potential Customers

Thus After all considerations the final Target market regions are :

* **Tamil Nadu** (93% in urban and 40% in rural) = 4.6 million Potential Customers
* **Kerela** (54%) = 3.6 million Potential Customers

Hypertension

Top Highest Hypertension States in India:

* Maharastra : 25.1% of Population suffering from hypertension.
* Andhra Pradesh : 13.3% of Population suffering from hypertension.
* Odisha: 9% of Population suffering from hypertension..
* Chhattisgarh: 8.4% of Population suffering from hypertension.
* Gujarat :  6.7% of Population suffering from hypertension.

Considering The Population, the most amount of hypertension patients reside in:

* Maharastra : 28 million patients
* Andhra Pradesh : 6.5 million patients
* Odisha: 3.9 million patients
* Gujarat :  0.4.2 million patients

The Internet User Base in :

* Maharastra (61%) : 17 million Potential Customers
* Andhra Pradesh (31%) : 1.2 million Potential Customers
* Odisha: (31%) : 2 million Potential Customers
* Gujarat (10%):  0.4 million Potential Customers

Thus After all considerations the final Target market region is :

* **Maharastra** (61%) : 17 million Potential Customers

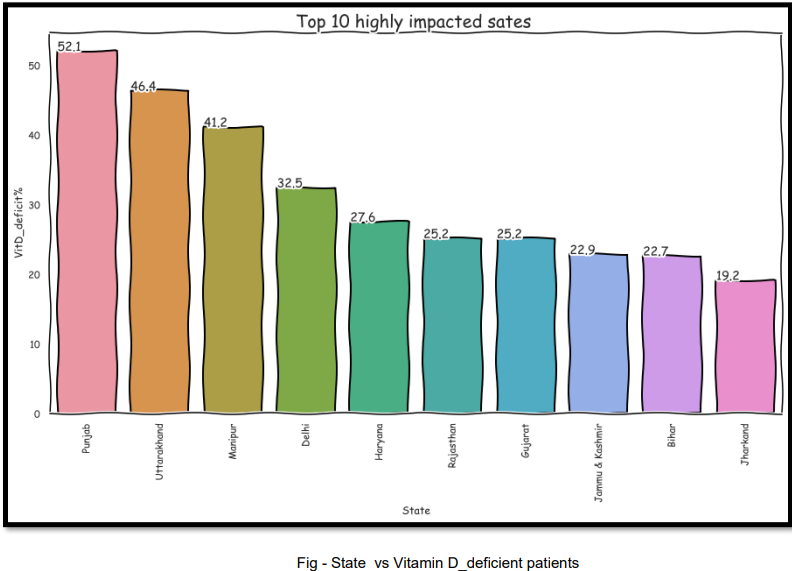
Heart Disease

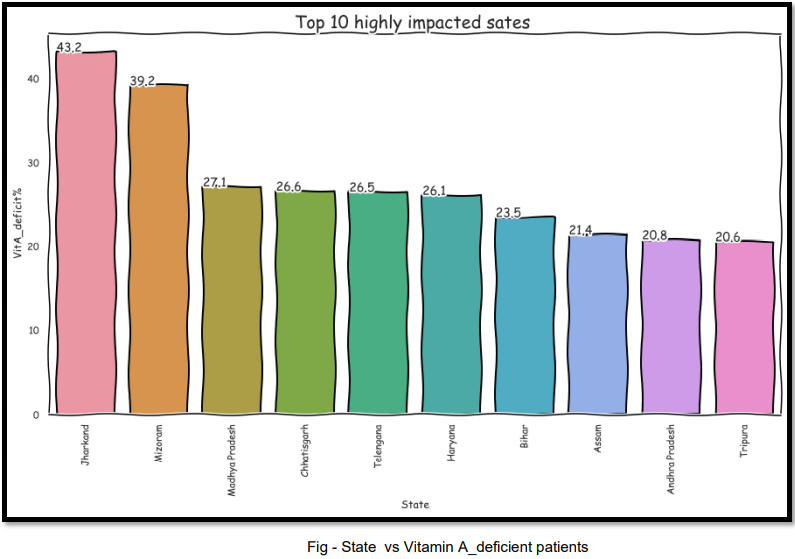
Top Highest Heart Disease States in India:

* Tamil Nadu
* Punjab
* Kerela

Since we already have our target market of Diabetes in **Kerela** and **Tamil Nadu** , setting it for Heart Disease is economical and beneficial.

Vitamin Deficiency





* From above charts it is evident that Vitamin Deficiency predominantly exists in North India.
* So Potential Target Market can Include:
  + Madhya Pradesh
  + Punjab
  + Jharkhand
  + Chhatisgarh

Because these states have huge population and a lot of potential consumer.

Full Body Health Checkup

It can be included in above all suggested target markets as it is a vital part of online health service offering and can be integrated into all of the above mentioned diseases.

**Customizing the Marketing Mix**

Product

1. Full Body Check-Up with a Bio-Tech Device based on Blood Samples

2. Online Health Techs offering

i. Diabetes check-up device  
ii. Blood Pressure check-up device

iii. Vitamins deficiency check-up device

Target Market

Diabetes & Heart Disease: Kerela, Tamil Nadu

Hypertension : Maharastra

Vitamin Deficiency : Any two of Madhya Pradesh , Punjab , Jharkhand , Chhatisgarh

Pricing Model

In current market full body checkup costs : Rs 1200

Daibetes , Hypertension Checkup device costs : Rs 500 – Rs 1500

Depending the quality of device and service provided by the Start Up, they can charge accordingly.

Promotion

Promotions can be done using :

- Medical shop branding

- Bus advertising

- Cab advertising

- YouTube and TV ads.

**GitHub Link**

<https://github.com/ruck-45/Bio_TEch_Market_Segmentation_India/tree/master>