BIO-TECH STARTUP LAUNCH

Feynn Labs Assignment

ABSTRACT

Analyze Medical Market in India with respect Bio-Tech Startup going to launch its Home Checkup Service with Online Booking offering for B.P. , Diabetes & Vitamins

Team Abhinav (E)

Contributors:-

- **Abhinav Kumar Singh**
- **Jasthi naveen**
- **Riya Pandey**
- **Yogeshwar Manerikar**

CONTENTS

Git link..... https://github.com/Abhi9vSingh/feynn_labs-projects

Introduction

With 462 million active internet users and 430 million active mobile internet users in India, the scope for e-health services and solutions can prove to be a game-changer in the patient care space. «Various programs are focused on making the entire spectrum of medical facilities available 24*7 through the web, mobile, SMS and Call center services.

The spectrum of these online interventions encompasses medical consultation, medical records, medicine supply management and Pan-India exchange of patient information. Some of the expected outcomes are delivery of better medical amenities in terms of equitable access, quality, affordability, lowering of disease burden, and efficient monitoring of health entitlements for citizens.

Benefits of Market Segmentation in healthcare

Market segmentation is a decision-making tool for the marketing manager in the crucial task of selecting a target market. In the Healthcare industry, a company willing to offer different services to different target audiences will lead to the sustainable growth of the company.

Problem Statement

You are a team working under a Bio-Tech Startup going to launch its Home Checkup Service with Online Booking offering the following initial services.

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

2. Online Health Techs offering

i. ***Diabetes*** checkup device,

ii. ***Blood Pressure*** checkup device,

iii. ***Vitamins*** deficiency checkup device

You have to ***Analyze 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, Behavioral.

Fermi Estimation

The problem statement says that Bio-Tech Startup going to launch its Home Checkup Service with Online Booking offering.

One Startup Company offering online health checkup. Bio-Tech Startup want to know what are the states in India where they can start their initial phase. **Give the insights based on the market segments Geographic, Demographic, Psychographic and Behavioral.**

We cover Geographic, demographic market segments

Startup providing

i. Diabetes checkup device,

ii. Blood Pressure checkup device,

iii. Vitamins deficiency checkup device

So we chose Vitamins deficiency as tar gate variable. Based on this we given feasible strategies.

.

Data Source

We collected the segment wise data. We collected from government of India website .

https://nhm.gov.in

https://data.gov.in

- Indian state wise vitamin deficiency

Deficiency of vitamins is given by state wise and the population count is given.

- Indian Population state wise
- Internet Subscriber in India

- Literacy Rate Across India

The data in this CSV file contains the data from the Govt. Of India website, regarding the literacy rate of the 35 states and union territories.

There are 3 key fields, literacy rate overall, literacy rate urban and literacy rate rural.

Data Pre-processing and Exploratory Data Analysis

Most EDA techniques are graphical in nature with a few quantitative techniques. The reason for the heavy reliance on graphics is that by its very nature the main role of EDA is to open-mindedly explore, and graphics gives the analysts unparalleled power to do so, enticing the data to reveal its structural secrets, and being always ready to gain some new, often unsuspected, insight into the data.

We have carried out extensive exploration on the available data to get as much understanding of the patterns as possible. Firstly we had to check for any missing values from the columns and also do a thorough check on if data values are consistent or not which was our main focus as analysis on inconsistent data is as good as garbage.

Data set represents the information of India, while other rows represent states, so we have to remove it as it is an outlier. So, we eliminate the extremes. selecting top 5 state from each dataset

Result is concluded on the bases of similarly.

Indian state wise vitamin deficiency

Data contains

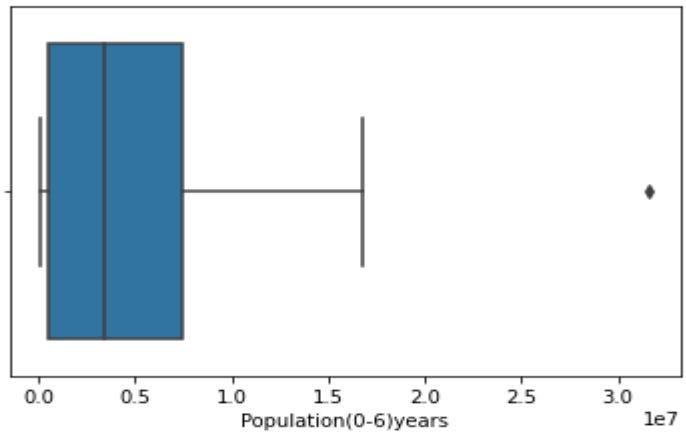
```
In [174]: data.head()
```

Out[174]:

	State	Population(0-6)years	VitA_deficit%	VitD_deficit%
1	Delhi	2016849	17.8	32.5
2	Haryana	3335537	26.1	27.6
3	Himachal Pradesh	793137	5.9	4.6
4	Jammu & Kashmir	1485803	8.7	22.9
5	Punjab	3171829	17.2	52.1

- One more outlier is present but we cannot drop that as it represents some state and if we remove that we will not get information about that state.

```
In [175]: sns.boxplot(x=data['Population(0-6)years'])
Out[175]: <AxesSubplot:xlabel='Population(0-6)years'>
```

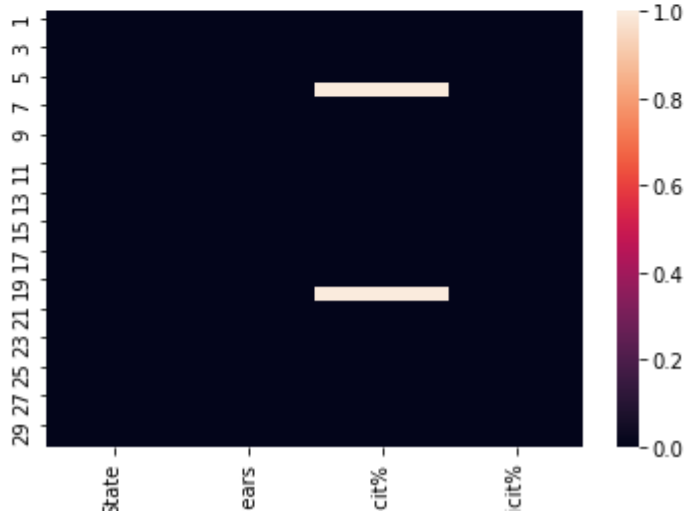


Null values

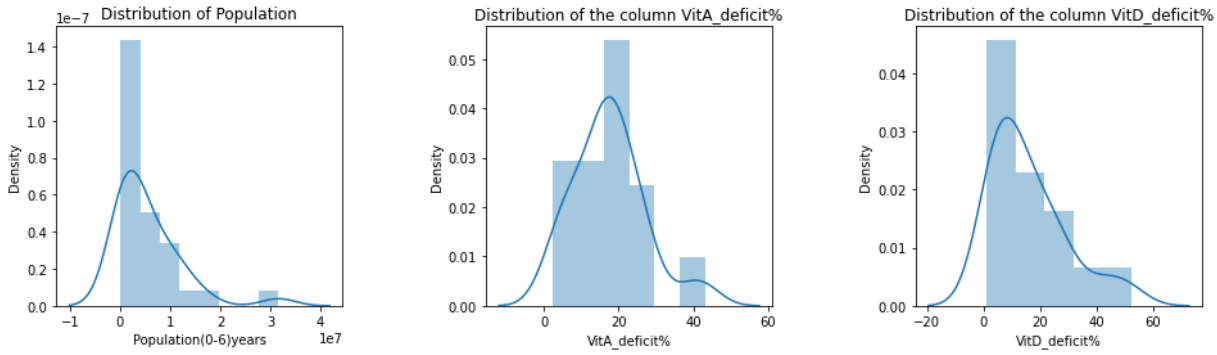
Checking for null Values

```
In [180]: data.isnull().sum()
Out[180]: State                  0
          Population(0-6)years   0
          VitA_deficit%          2
          VitD_deficit%          0
          dtype: int64

In [181]: sns.heatmap(data.isnull())
Out[181]: <AxesSubplot:>
```



- first lets relate the population of 0-6 years old in different states of India
- Now lets see the distribution of population(0-6 years), vitamin A,B deficiency



Adding columns in data repersenting total number of poulation having viatmin A and D deficiency respectively for each state

- Adding columns in data representing total number of populations having vitamin A and D deficiency respectively for each state

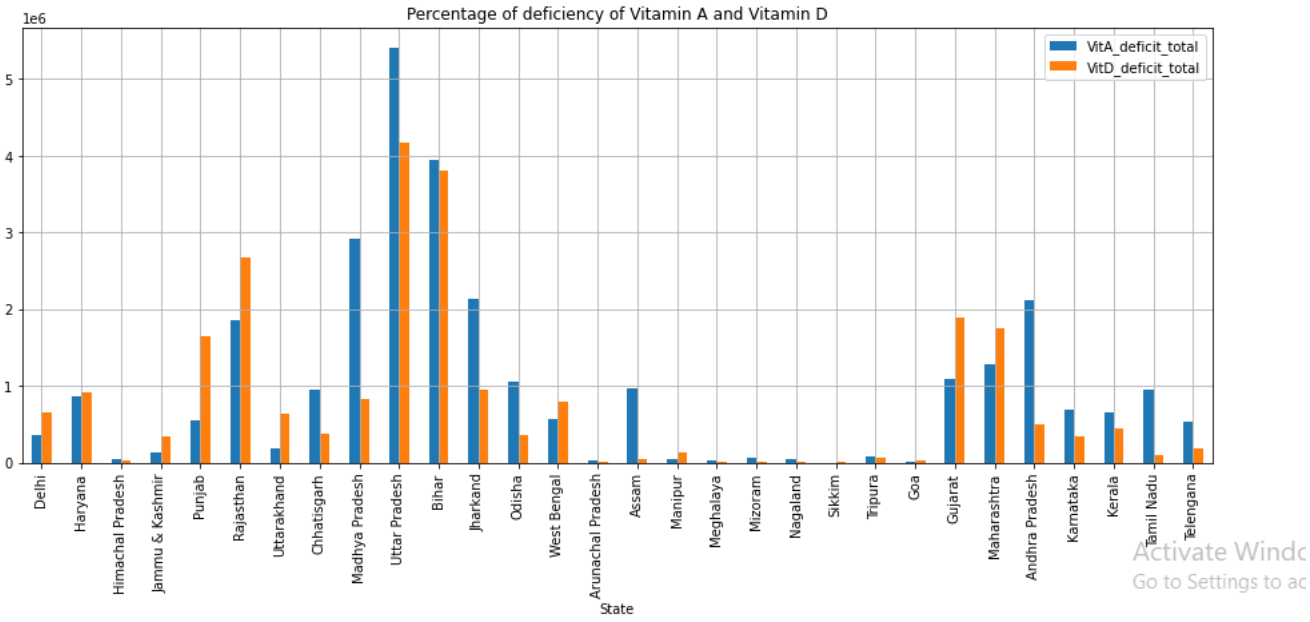
```
In [188]: data['VitA_deficit_total'] = data['VitA_deficit%'] * data['Population(0-6)years'] /100
          data['VitD_deficit_total'] = data['VitD_deficit%'] * data['Population(0-6)years'] /100

In [189]: data.head()
```

Out[189]:

	State	Population(0-6)years	VitA_deficit%	VitD_deficit%	VitA_deficit_total	VitD_deficit_total
1	Delhi	2016849	17.8	32.5	358999.122	655475.925
2	Haryana	3335537	26.1	27.6	870575.157	920608.212
3	Himachal Pradesh	793137	5.9	4.6	46795.083	36484.302
4	Jammu & Kashmir	1485803	8.7	22.9	129264.861	340248.887
5	Punjab	3171829	17.2	52.1	545554.588	1652522.909

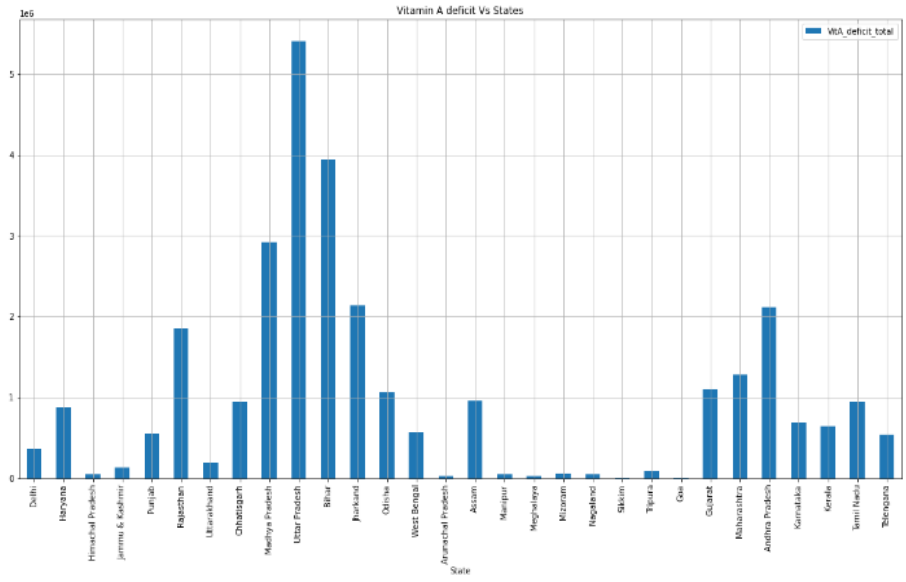
Comparison of Vitamin A and D deficiency in different states



- With this We can conclude in most states vitamin A Deficiency is more common, So our medical service should focus more vitamin A rich medicines.

```
In [214…  fig, ax= plt.subplots(figsize=(20,10))
          data.plot(x= 'State', y ="VitA_deficit_total", kind = 'bar',
                  title ='Vitamin A deficit Vs States', grid=True,ax=ax)
```

```
Out[214…  <AxesSubplot:title={'center':'Vitamin A deficit Vs States'}, xlabel='State'>
```



In [223…

```
#Distribution of Vitamin D deficiency over different states
```

In [224…

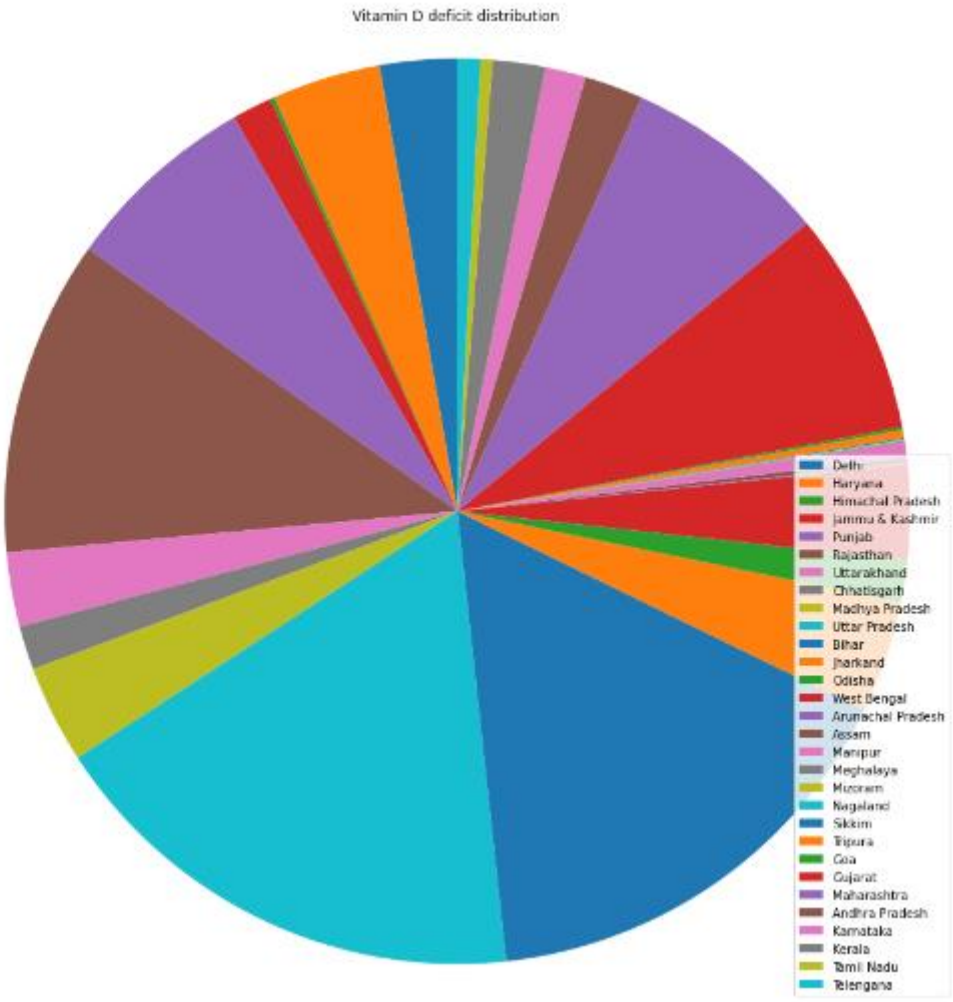
```
pie, ax = plt.subplots(figsize=[15,15])
patches, texts = plt.pie(data['VitD_deficit_total'], startangle=90)
plt.title("Vitamin D deficit distribution")
plt.legend(patches, labels=data['State'], loc="lower right")
plt.axis('equal')
```

```
<ipython-input-224-2bf67607ef90>:4: UserWarning:

You have mixed positional and keyword arguments, some input may be discarded.
```

Out[224…

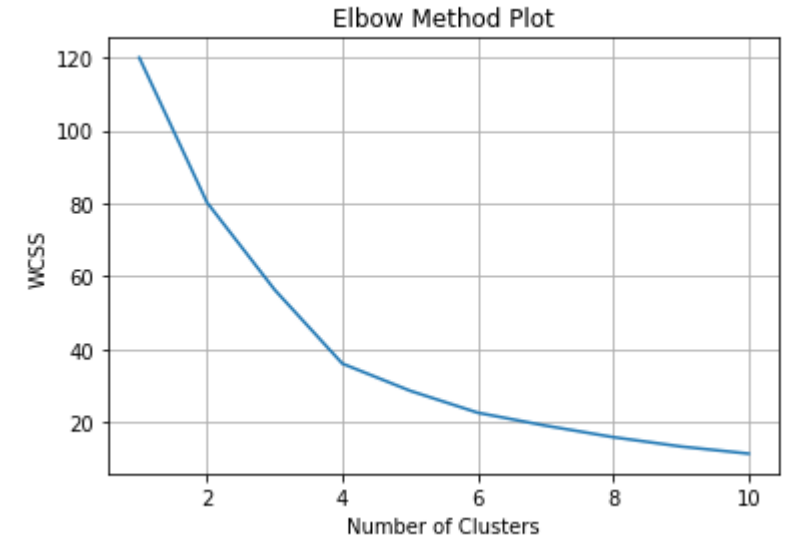
```
(-1.1069749665484279,
 1.1004165173396536,
 -1.1148496795770613,
 1.1007071305043223)
```



According to these to pie charts Gujarat, Rajasthan, Bihar and Uttar Pradesh are prominent areas to implement our business model as we can see there is more scope of gaining customers for Vitamin's deficiency checkup.

Clustering similar states together to operate our business similarly in similar states

Using ML model




```
kmeans = KMeans(n_clusters= 5)
label = kmeans.fit_predict(train_std)
print(label)
```

```
[0 0 4 4 0 2 0 1 1 3 3 1 1 4 4 1 0 4 1 4 4 1 4 2 2 1 4 4 4 1]
```

```
data['Cluster']=label
data.head()
```

	State	Population(0-6)years	VitA_deficit%	VitD_deficit%	VitA_deficit_total	VitD_deficit_total	Cluster
1	Delhi	2016849	17.8	32.5	358999.122	655475.925	0
2	Haryana	3335537	26.1	27.6	870575.157	920608.212	0
3	Himachal Pradesh	793137	5.9	4.6	46795.083	36484.302	4
4	Jammu & Kashmir	1485803	8.7	22.9	129264.861	340248.887	4
5	Punjab	3171829	17.2	52.1	545554.588	1652522.909	0

```
States in Cluster1 are  ['Delhi' 'Haryana' 'Punjab' 'Uttarakhand' 'Manipur']
States in Cluster2 are  ['Chhatisgarh' 'Madhya Pradesh' 'Jharkand' 'Odisha' 'Assam' 'Mizoram'
 'Tripura' 'Andhra Pradesh' 'Telengana']
States in Cluster3 are  ['Rajasthan' 'Gujarat' 'Maharashtra']
States in Cluster4 are  ['Uttar Pradesh' 'Bihar']
States in Cluster5 are  ['Himachal Pradesh' 'Jammu & Kashmir' 'West Bengal' 'Arunachal Pradesh'
 'Meghalaya' 'Nagaland' 'Sikkim' 'Goa' 'Karnataka' 'Kerala' 'Tamil Nadu']
```

Hence our business strategy should be:

- **Similar for states ['Uttar Pradesh' 'Bihar']**
- **Similar for states ['Himachal Pradesh' 'Jammu & Kashmir' 'West Bengal' 'Arunachal Pradesh' 'Meghalaya' 'Nagaland' 'Sikkim' 'Goa' 'Karnataka' 'Kerala' 'Tamil Nadu']**
- **Similar for states ['Delhi' 'Haryana' 'Punjab' 'Uttarakhand' 'Manipur']**
- **Similar for states ['Rajasthan' 'Gujarat' 'Maharashtra']**

Indian Population state wise

```
In [2]: data= pd.read_csv('state_wise_population__2019.csv')
        data.head(8)
```

Out[2]:

	State	population_total	population_male	population_female
0	1	JHARKHAND	4278259	4278259
1	2	MANIPUR	264986	128931
2	3	GUJARAT	17835049	9541688
3	4	WEST BENGAL	18063509	9357777
4	5	TAMIL NADU	13879395	6957261
5	6	NCT OF DELHI	13481997	7201322
6	7	ASSAM	1391154	720132
7	8	MADHYA PRADESH	11023091	5761143

Add male and female as total population

In [4]:
```
data['total']=data['population_male']+data['population_female']
```

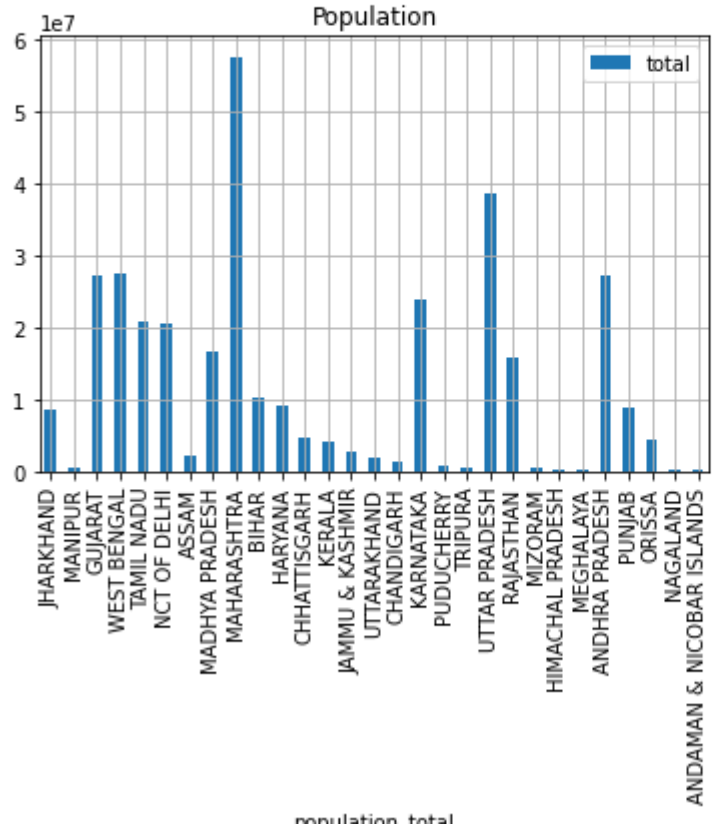
In [5]:
```
data
```

Out[5]:

	State		population_total	population_male	population_female	total
0	1	JHARKHAND	4278259	4278259	8556518	
1	2	MANIPUR	264986	128931	393917	
2	3	GUJARAT	17835049	9541688	27376737	
3	4	WEST BENGAL	18063509	9357777	27421286	
4	5	TAMIL NADU	13879395	6957261	20836656	
5	6	NCT OF DELHI	13481997	7201322	20683319	
6	7	ASSAM	1391154	720132	2111286	
7	8	MADHYA PRADESH	11023091	5761143	16784234	
8	9	MAHARASHTRA	37722136	19961736	57683872	
9	10	BIHAR	6714516	3558722	10273238	
10	11	HARYANA	5962223	3181968	9134191	

```
]: data.plot(x= 'population_total', y ='total', kind = 'bar',
             title ='Population', grid=True)
```

```
]: <AxesSubplot:title={'center':'Population'}, xlabel='population_total'>
```



So, there are many states which can be selected for our start up to launch their services in solely based on population count.

Most likely more business will be generated from states like:

Rajasthan Uttar Pradesh Bihar West Bengal Madhya Pradesh Maharashtra Andhra Pradesh

Note: - These are states with population greater than 10 million and this does not visualize whole scenario it is just a speculation based on Total population count of the above given states.

Literacy Rate Across India

data

```
In [3]: df.head() # Let's take a look at the data.
```

Out[3]:

	Category	Country/ States/ Union Territories Name	Literacy Rate (Persons) - Total - 2001	Literacy Rate (Persons) - Total - 2011	Literacy Rate (Persons) - Rural - 2001	Literacy Rate (Persons) - Rural - 2011	Literacy Rate (Persons) - Urban - 2001	Literacy Rate (Persons) - Urban - 2011
0	Country	INDIA	64.8	73.0	58.7	67.8	79.9	84.1
1	State	Andhra Pradesh	60.5	67.0	54.5	60.4	76.1	80.1
2	State	Arunachal Pradesh	54.3	65.4	47.8	59.9	78.3	82.9
3	State	Assam	63.3	72.2	59.7	69.3	85.3	88.5
4	State	Bihar	47.0	61.8	43.9	59.8	71.9	76.9

- We have data for two years 2011 and 2001 which have a difference of a decade between them. We can generate new attribute to see the percentage change in literacy rate over the decade

```
In [57]: #sort the values of Urban-2011 column and take upper 5 rows
         lowest_u_2011=df.sort_values('Urban - 2011',na_position='first').head()
```

```
In [58]: lowest_u_2011
```

Out[58]:

	Category	States/ Union Territories	Total - 2001	Total - 2011	Rural - 2001	Rural - 2011	Urban - 2001	Urban - 2011	Total - Per. Change	Rural - Per. Change
26	State	Uttar Pradesh	56.3	67.7	52.5	65.5	69.8	75.1	0.202487	0.23090
4	State	Bihar	47.0	61.8	43.9	59.8	71.9	76.9	0.314894	0.33829
10	State	Jammu & Kashmir	55.5	67.2	49.8	63.2	71.9	77.1	0.210811	0.24144
22	State	Rajasthan	60.4	66.1	55.3	61.4	76.2	79.7	0.094371	0.10099
1	State	Andhra Pradesh	60.5	67.0	54.5	60.4	76.1	80.1	0.107438	0.09752

```
In [28]: highest_2011
```

Out[28]:

	Category	States/ Union Territories	Total - 2001	Total - 2011	Rural - 2001	Rural - 2011	Urban - 2001	Urban - 2011	Total - Per. Change	Rural - Per. Change	Urban - Per. Change
25	State	Tripura	73.2	87.2	69.7	84.9	89.2	93.5	0.191257	0.207650	0.058743
6	State	Goa	82.0	88.7	79.7	86.6	84.4	90.0	0.081707	0.084146	0.068293
18	State	Mizoram	88.8	91.3	81.3	84.1	96.1	97.6	0.028153	0.031532	0.016892
33	Union Territory	Lakshadweep	86.7	91.8	85.0	91.6	88.6	91.9	0.058824	0.076125	0.038062
13	State	Kerala	90.9	94.0	90.0	93.0	93.2	95.1	0.034103	0.033003	0.020902

So, the highest literacy score in urban area Goa, Mizoram, Kerala. Having education helps easy because of their awareness on health, importance of medical checkups and minimum knowledge on using online services

Conclusion:

It's very difficult to provide a particular, business numbers in terms of revenue, cost and profit with the limited data available here. there are direct and indirect factors influencing the cost and business model.

- **The business model should also use mobile booking so that even lay man can use mobile to book home checkup with ease.**

Implementing our model in Huge Population areas makes easy for business in every possible way, it makes us reach to our customer fast when compare to other places. By providing good services with minimum cost, we can take over major share in the current business

- **By identifying particular problems in certain areas will help us setting up and providing that checkup services only, this helps us a lot of money**