MONASH University

FIT5147 – **Data Exploration Project**

Beneficiaries of the Health Scheme of the state Andhra Pradesh (NTR Vaidya Seva), India

VADDE PAVAN KUMAR

29772990

pvad0001@student.monash.edu

Introduction:

NTR Vaidya Seva (or Arogya Seva), which is the government scheme initiated by the Government of Andhra Pradesh, India. Under this scheme where lower-middle class and low-income citizens of Andhra Pradesh are getting benefited where they can obtain free medical support for many major diseases and ailments.

- Under this End-to-end cashless services for identified 1044 diseases under secondary and tertiary care through 400 Government and Corporate Network Hospitals.
- The beneficiaries The Scheme is intended to benefit 129.44 lakh families in all 13 districts of the state.

Problem Statement: The objective of this project is to explore scheme data and answer some interesting questions like.

Q1) which age groups of the population are highly getting benefited with this scheme and for what kind of diseases the treatment is being made often?
Q2) which particular district in the state is highly active in using the scheme, and to know whether it is reaching the rural people and whether it is equally being shared among the districts according to densities of the population?
Q3) which community is getting highly benefited from the scheme?

In order the obtain the answers for questions, where I am going to analyze around 4,80,000 records and other data set with 14 records.

```
1 data.shape
```

```
(479688, 24)
```

```
1 data1.shape
```

```
(14, 8)
```

We are using two data sets for analysis; one is Data set with health Treatment Data and the other population statistics of the state Andhra Pradesh, District wise as per 2011 census.

Data Wrangling and Data Cleaning:

- using Python for Wrangling and clean up the Data
- Data has been merged in to single excel sheet using the **Excel** where we have the different sheets across the years from 2013 to 2017.

```
1  data['SEX'].value_counts()
```

```
Male             260718
Female           178947
Male(Child)       25068
Female(Child)     14925
FEMALE               21
MALE                  9
```

- As we can see that there are **duplicate values** in this column. Male and MALE are not two different genders and as well as Female and FEMALE are not two different genders. We can replace the column names to fix the issue of duplicate data.
- We can also see that Male (Child) and Female (child), as we can replace them as the Boy for Male (Child) and Girl for Female (child) for easy purpose.

```
1 mappings = {'MALE' : 'Male', 'FEMALE' : 'Female', 'Male(Child)' : 'Boy', 'Female(Child)' : 'Girl'}
2 data['SEX'] = data['SEX'].replace(mappings)
3 data['SEX'].value_counts()
```

```
Male      260727
Female    178968
Boy        25068
Girl       14925
```

- Where I am going to drop couple of columns, where we are not going to use the data any such data for analysis, since there type and category of data is not appropriate.

```
1 data.columns
```

```
Index(['    ', 'AGE', 'SEX', 'CASTE_NAME', 'CATEGORY_CODE', 'CATEGORY_NAME',
       'SURGERY_CODE', 'SURGERY', 'VILLAGE', 'MANDAL_NAME', 'DISTRICT_NAME',
       'PREAUTH_DATE', 'PREAUTH_AMT', 'CLAIM_DATE', 'CLAIM_AMOUNT',
       'HOSP_NAME', 'HOSP_TYPE', 'HOSP_LOCATION', 'HOSP_DISTRICT',
       'SURGERY_DATE', 'DISCHARGE_DATE', 'Mortality Y / N', 'MORTALITY_DATE',
       'SRC_REGISTRATION'],
      dtype='object')
```

- From the above tabular data, will drop the SURGERY_DATE, DISCHARGE_DATE, MORTALITY_DATE, SRC_REGISTRATION
- Since we have no use where we can analyze based on time and date.

```
1
2 data = data.drop(["SURGERY_DATE", "DISCHARGE_DATE", "MORTALITY_DATE","SRC_REGISTRATION"], axis = 0)
3 data.shape
```

```
(479688, 20)
```

```
1 data['HOSP_TYPE'].value_counts()
```

```
C    369346
G    110342
```

```
1 mappings = {'C' : 'Corporate', 'G' : 'Government'}
2 data['HOSP_TYPE'] = data['HOSP_TYPE'].replace(mappings)
3 data['HOSP_TYPE'].value_counts()
```

```
Corporate     369346
Government    110342
```

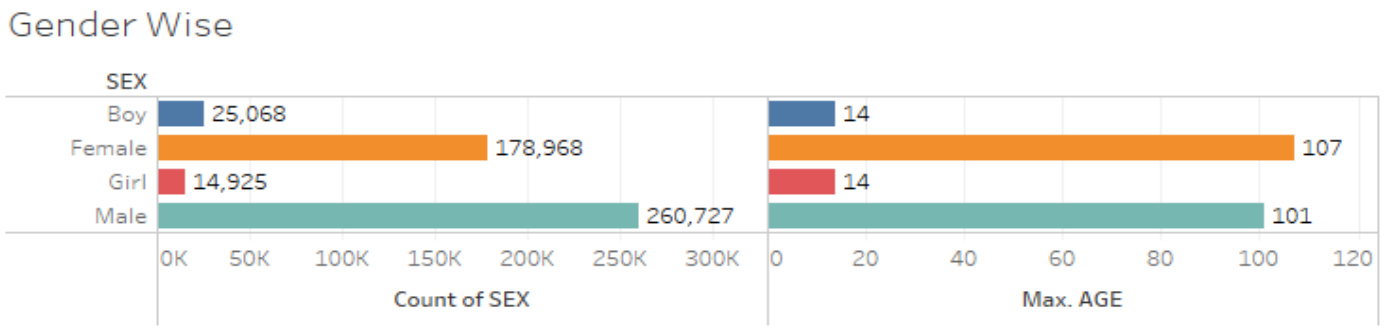
In this we have two different types of hospitals, like C and G, we are renaming it to the Corporate and Government hospitals.
It will be easy for us to understand and to read the data.

- We don’t have any missing values in the data
- We have completed Data Wrangling and Data Cleaning

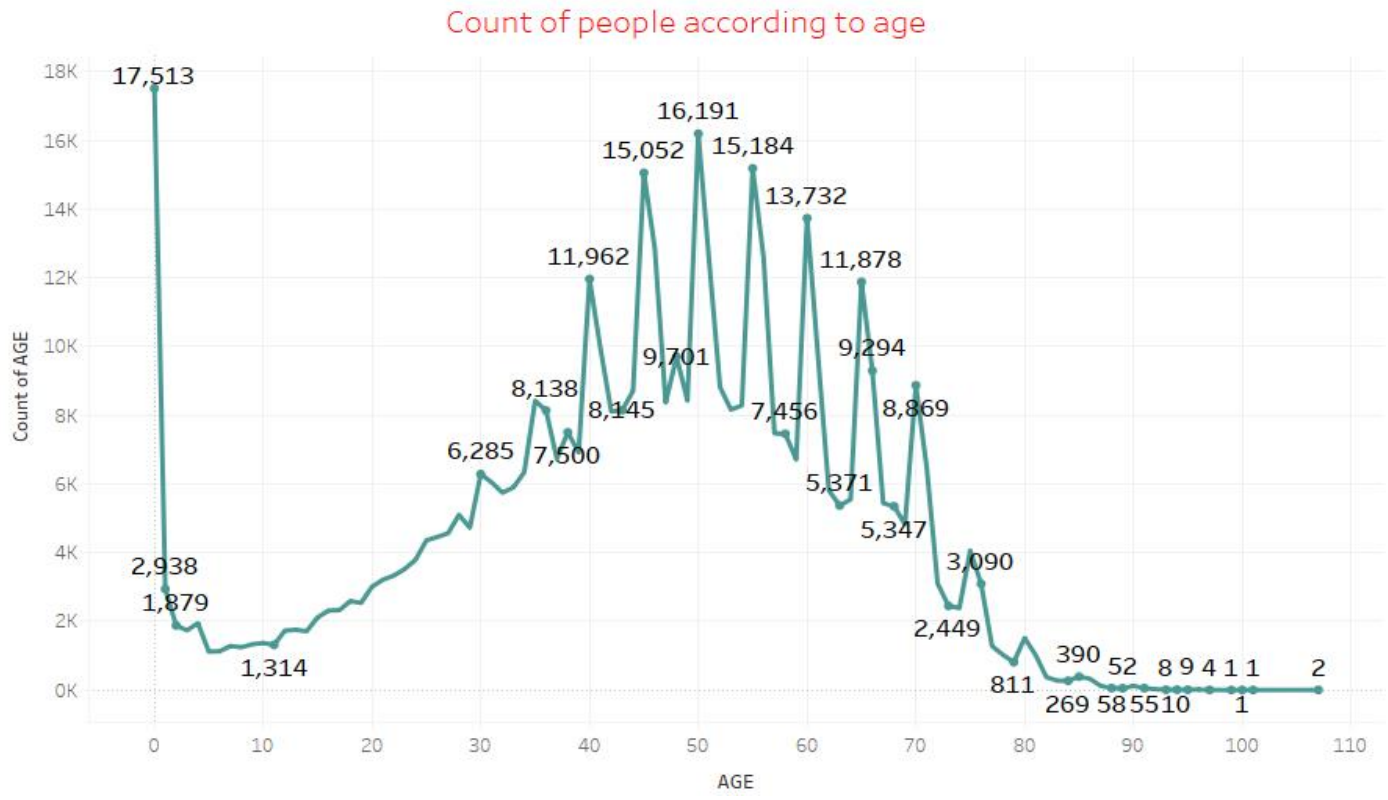
Data Checking: All the columns and attributes are in proper order and with the appropriate data types which doesn’t cause any problem in exploring the data.

Data Exploration:

1. **Age groups of the population are highly getting benefited with this scheme:**



- We can see that there more Males and Boys than the females
- Children are considered from age IN BETWEEN 0 -14 and rest are Elder people

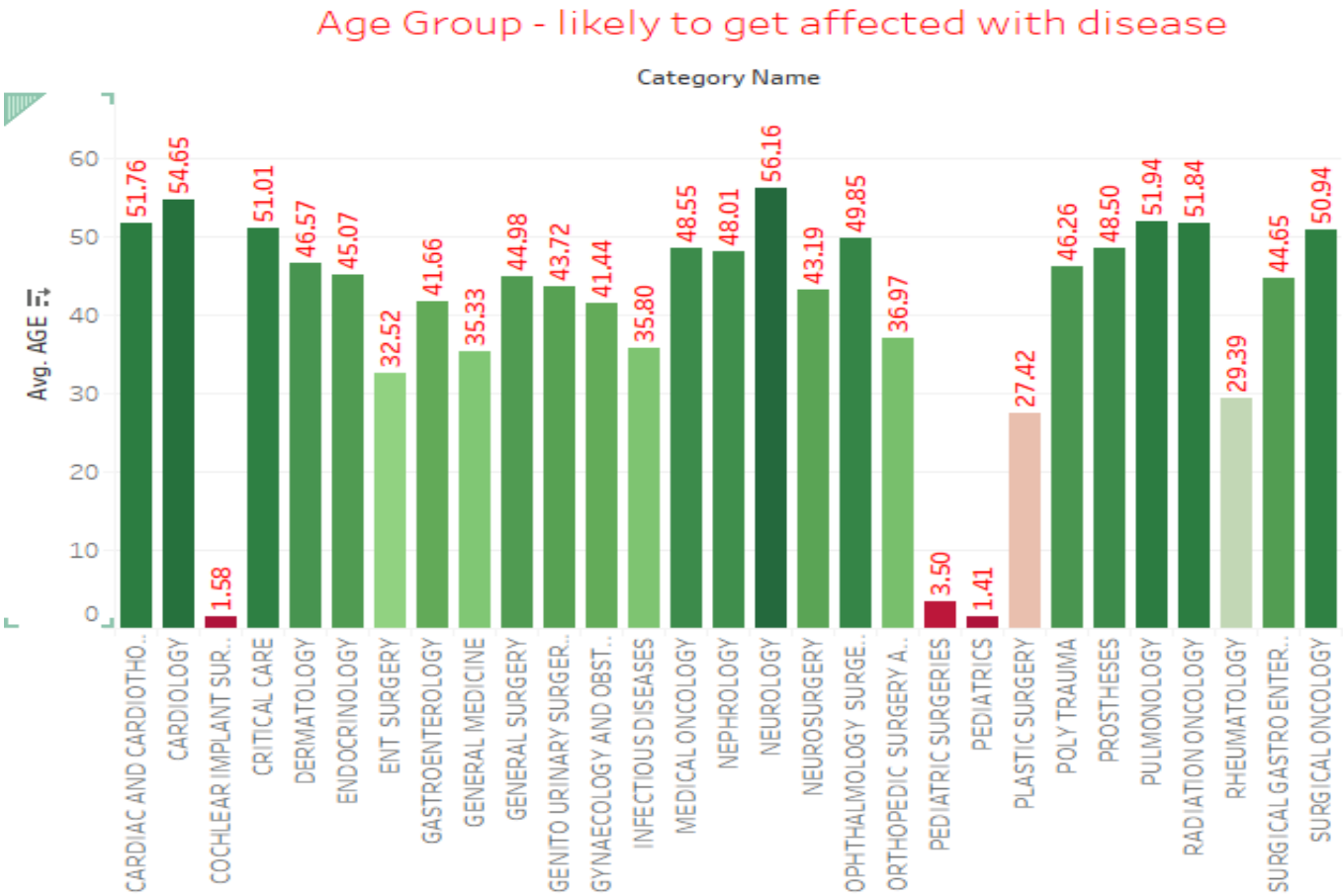


From the above plot we can clearly see that,

- The kids who were born recently are in the highest number getting benefited with the scheme.

- And then we can see there is rapid increase in usage of the scheme, and then within the age group of 45 to 60, there are many people getting benefited with the scheme.

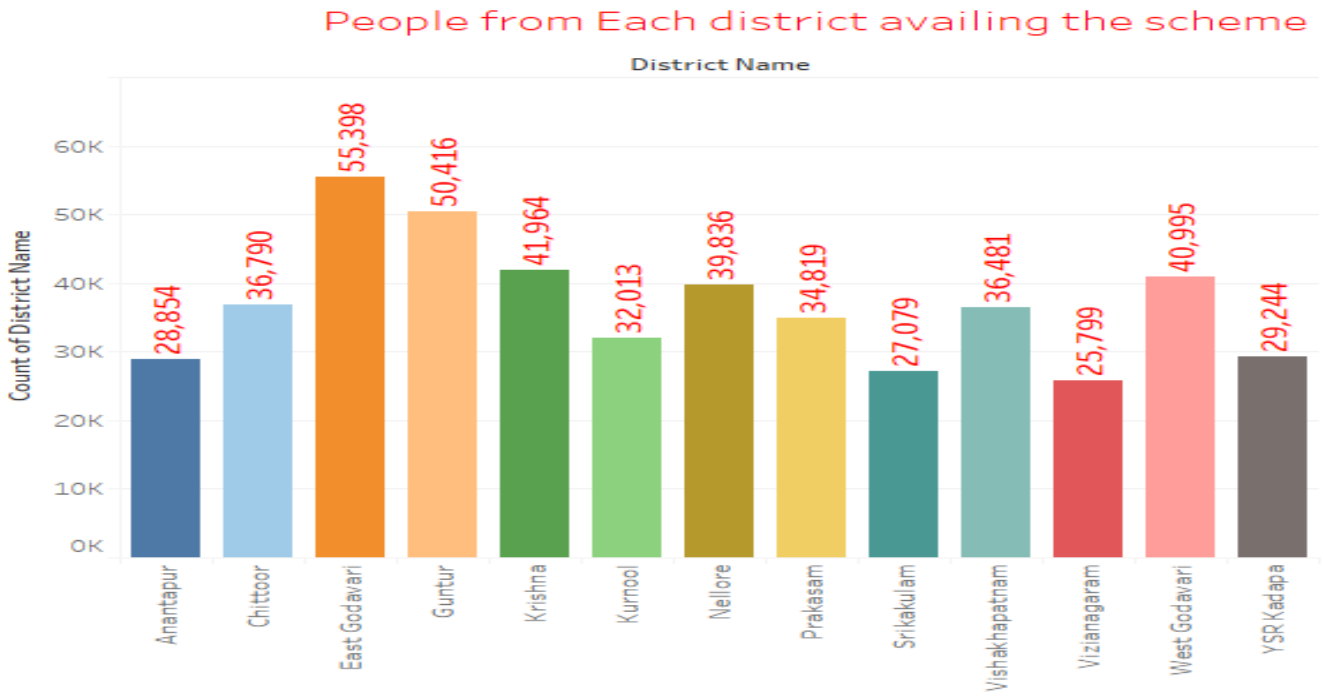
Average Age distribution plot people are getting the treatment

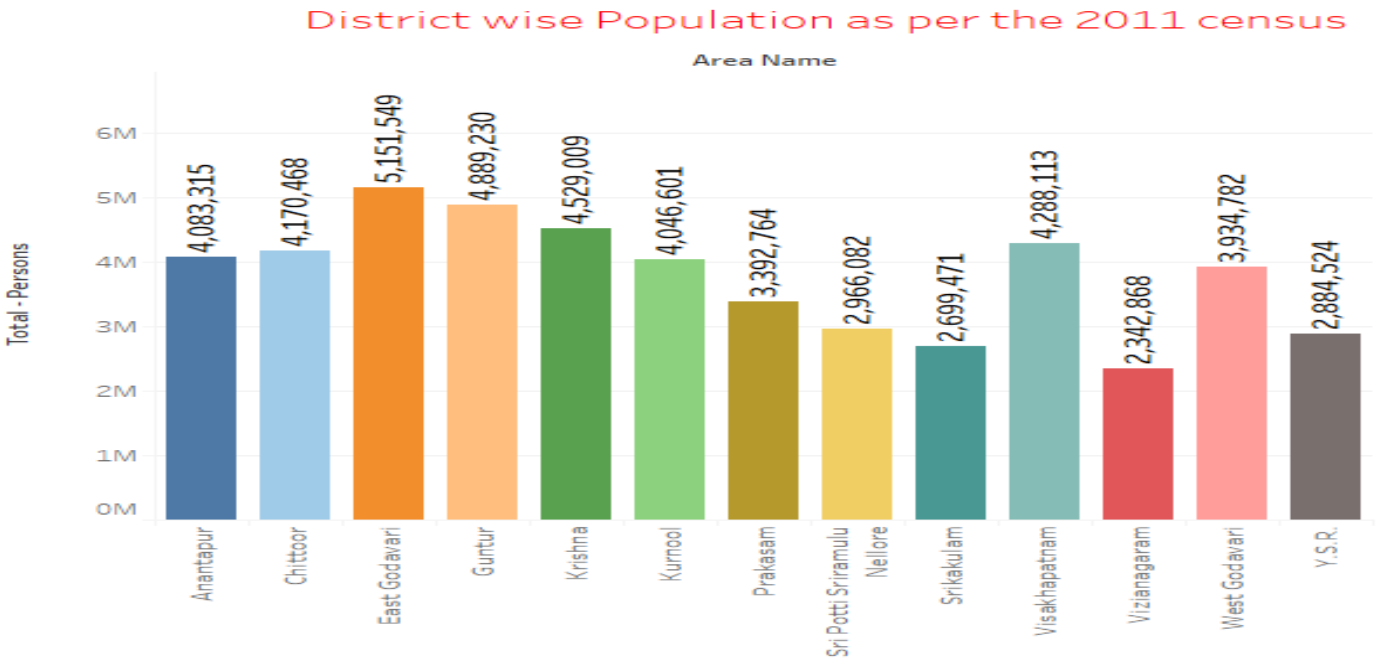


From the below graph,

- As we have clearly see from above graph , the kids who were recently born and **age groups between 40 to 60** are high among the people who are availing the scheme
- Children are undergoing through PEDIATRIC, PEDIATRIC SURGERIES (Usually done for Children with basic allergies) and COCHLEAR IMPLANT SURGERIES (which is performed to make sure prevent the child from hear loss)
- The people around the age 45 – 60 are undergoing the treatments under the scheme are
    1. CARDIOLOGY
    2. CARDIC AND CARDIOTHORACIC SURGERY
    3. CRITICAL CARE
    4. NEUROLOGY

DISTRICT WISE EXPLORING THE DATA:

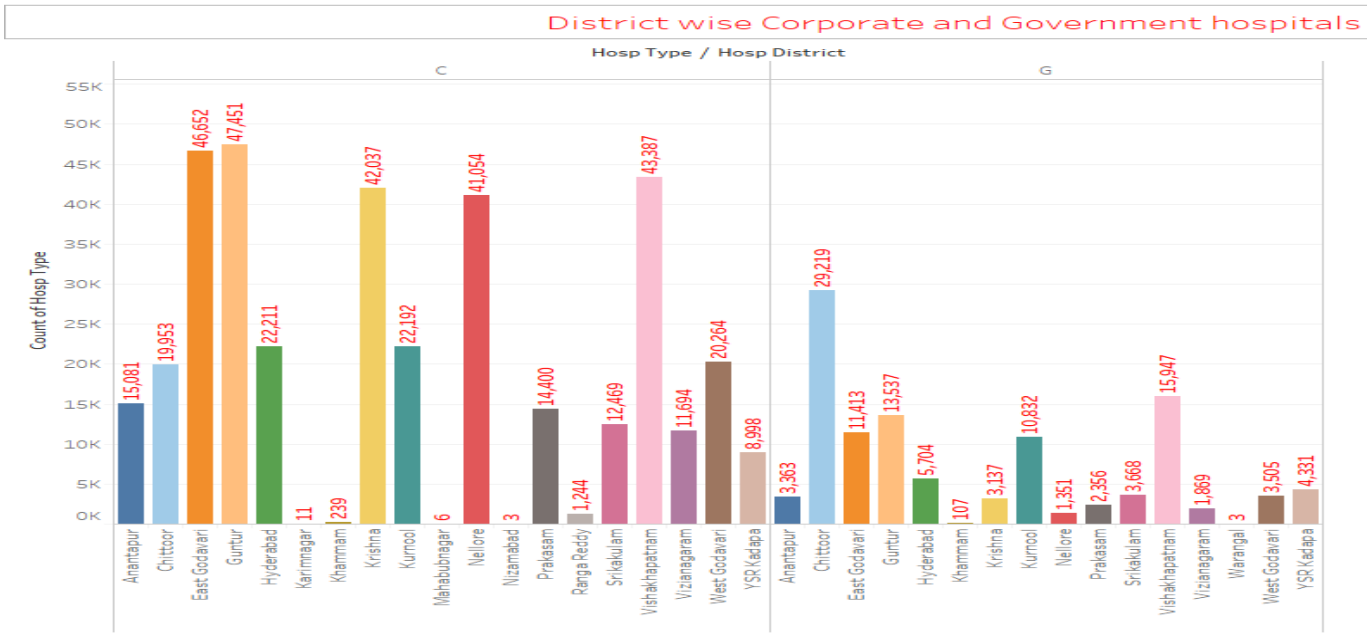




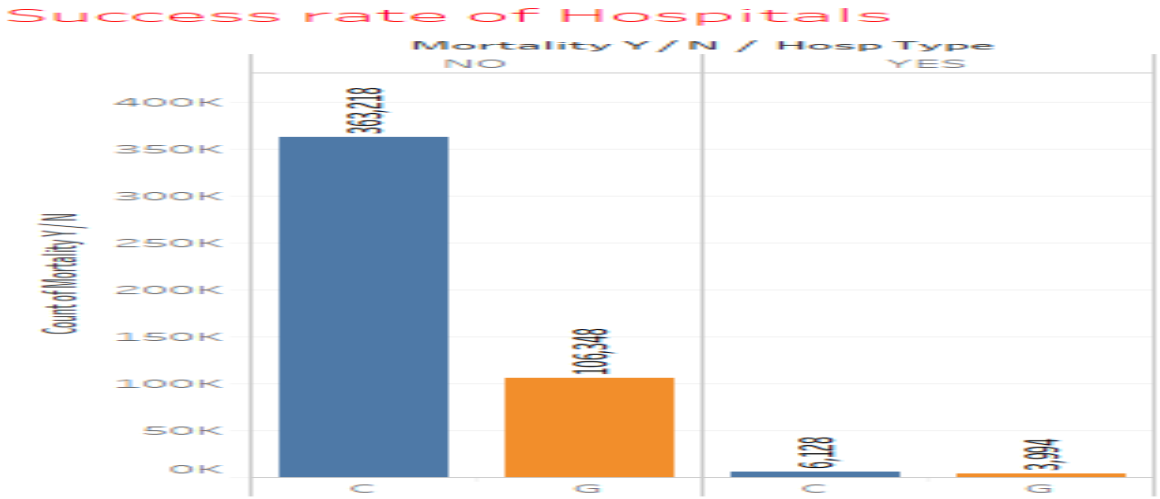
By looking at both the plots we can clearly say that East Godavari stands in first place according to the Population wise. In the likely manner they are utilizing the scheme very well

Analyzing the Hospital wise Data:

We are now exploring the different hospitals present in the state.



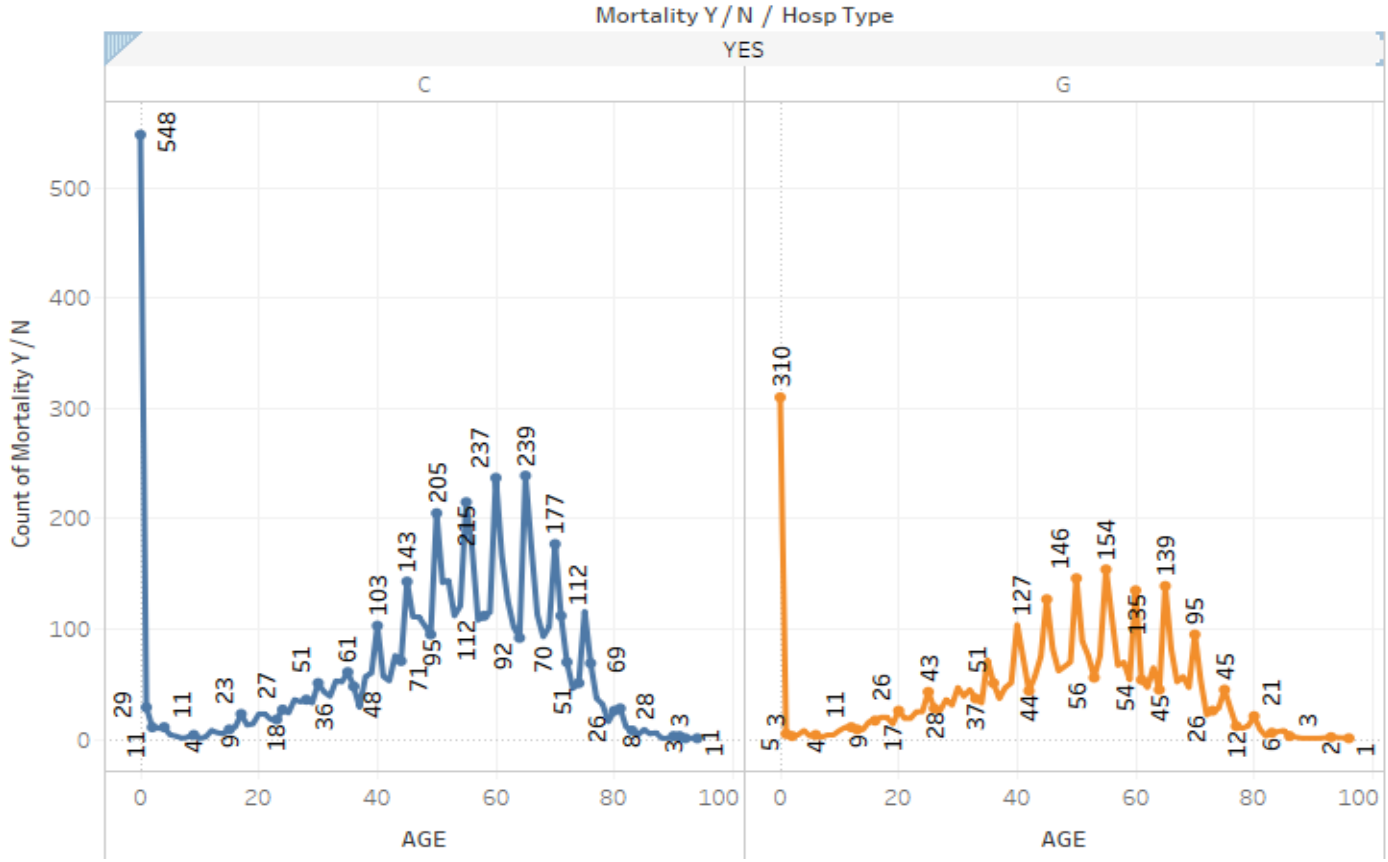
- Where we can clearly observe that there are less number of Government Hospital cases recorded in each district than the Private Hospital cases.
- From this we can clearly analyze that people aren't trusting equally as they are trusting the Private Hospitals, also can be interpreted that Government hospitals doesn't have enough resources.
- Only place where Chittoor is the district has higher cases in Government hospitals.



The above plot explains us that, the mortality rate is less in both government and Private Hospitals.

People who are using this scheme mostly there is success rate with the treatment

Failure rate, Deaths in the hospitals under the Scheme



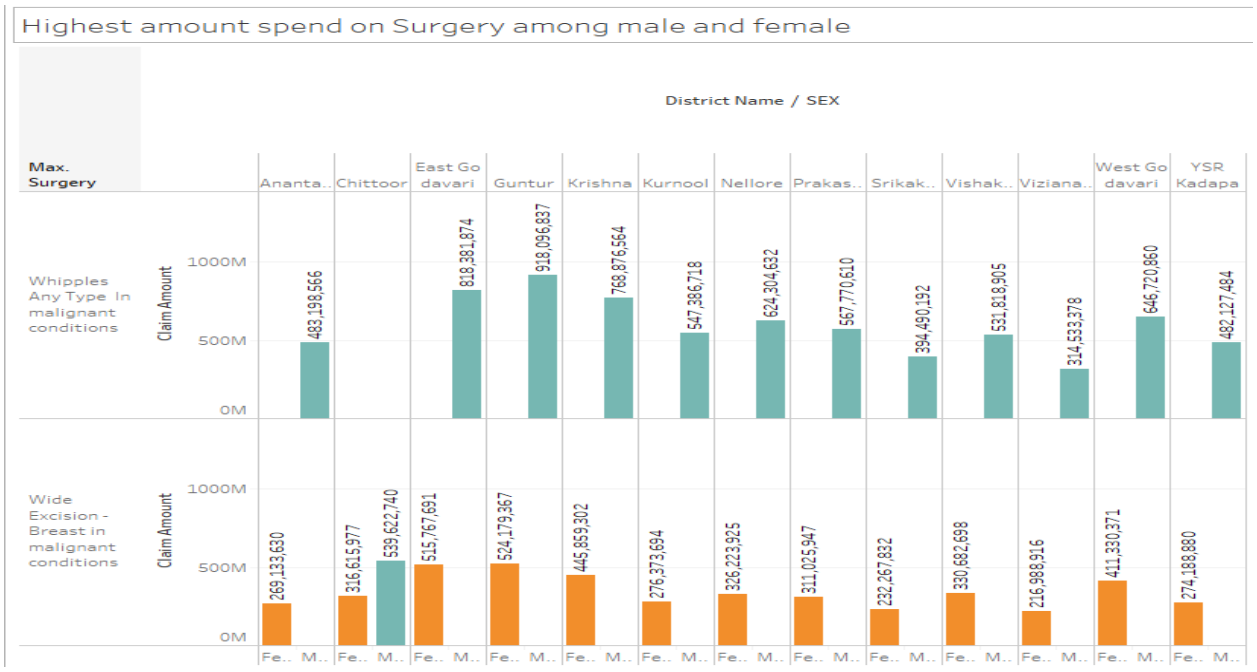
We can see that, it is so sad that there is lot more child death rate when compared to the other Age groups.

There are lot of children who are losing their lives within a year suffering from various diseases, where we already know due to pediatrics issue.

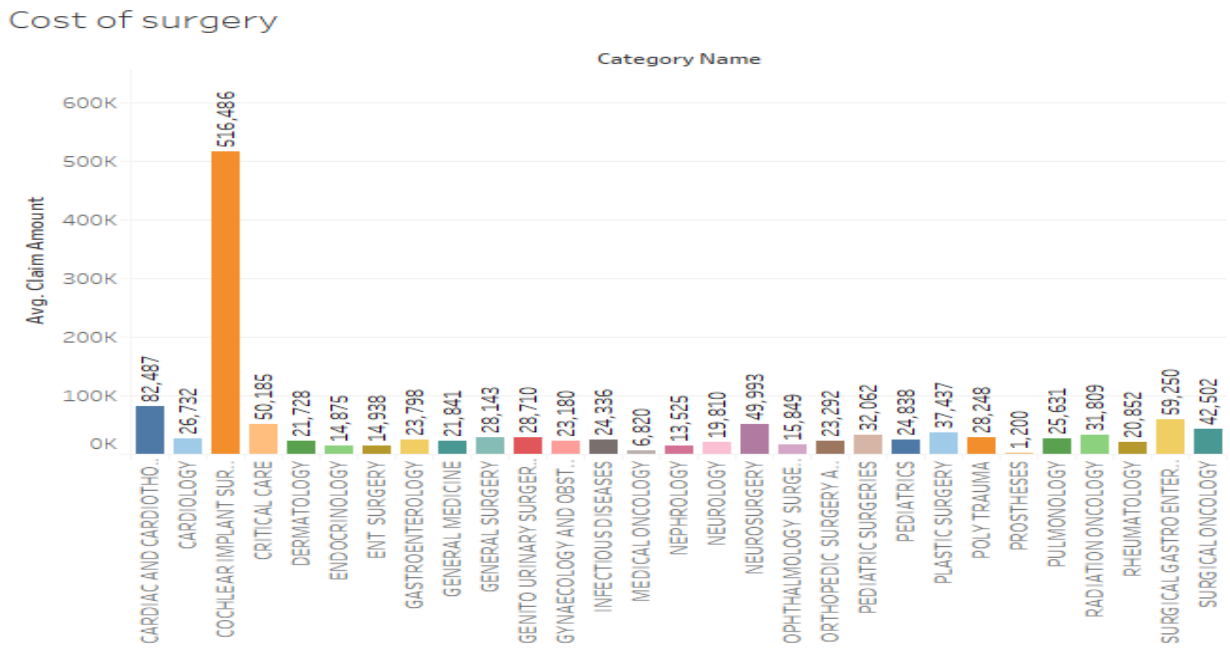
Then after that people between the age of 45 – 65 are likely loosing there life while being treated.

We have already analyzed the disease with which people are suffering.

Major disease getting treated based on Gender:

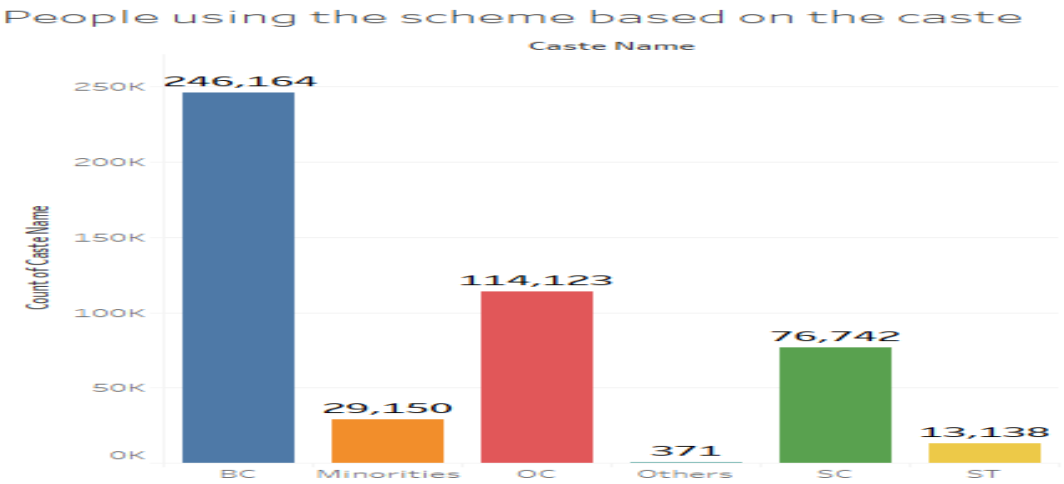


We can see that most surgery happened for male is whipples, for the females Breast malignant functions.



From the above plot we can see COCHLEAR IMPLANT SURGERY is costing higher than any other.

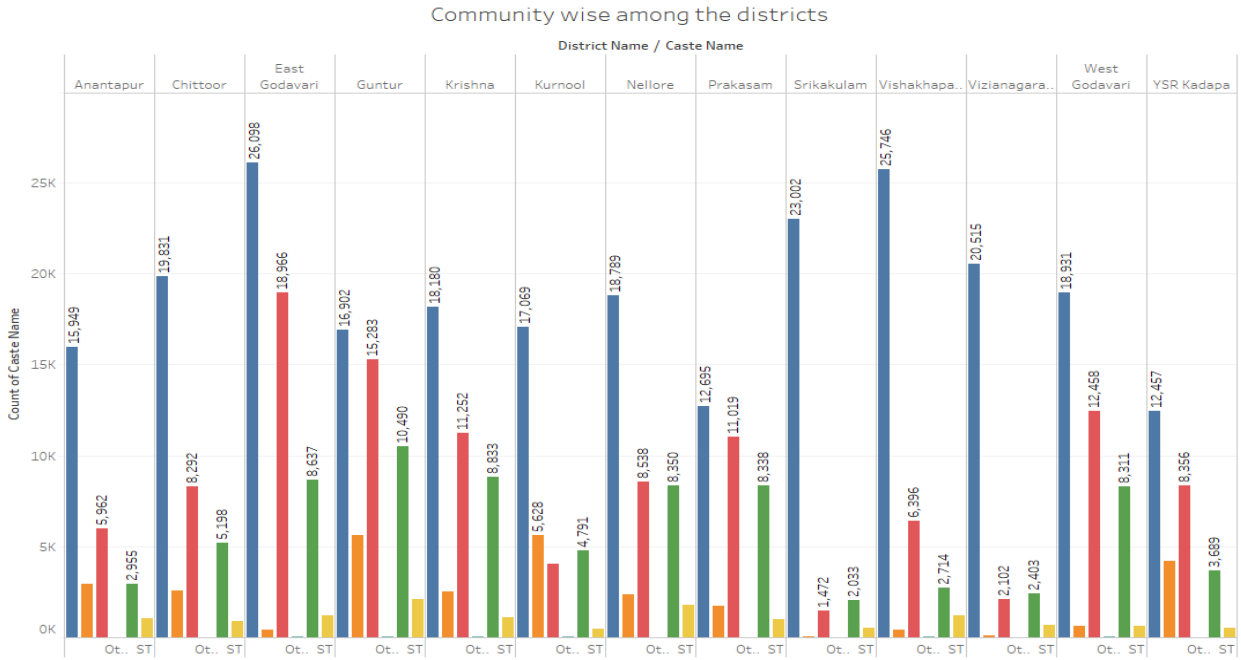
Community wise Data Analysis:



❖ Caste System in the state of Andhra Pradesh, are :

Backward Class (B.C), Minorities, Open Caste(OC), Scheduled Tribes(S.C), Scheduled Castes (S.T)

From the above plot we can clearly see that, BC are the people who are majorly availing the scheme.



From the above plot we can clearly see that, Backward class people are availing the scheme mostly, in all the districts.

Conclusion:

After exploring the data for the health scheme, I can conclude that,

- Mostly Men are among people who are actively using the scheme.
- Among the districts we can see that West Godavari is the District, where many people are using the scheme rather compared to other districts.
- When we are checking the hospitals, we can clearly see that they are lot more Private Hospitals rather than the Government Hospitals, which are active in the scheme.
- This clearly states that there were no proper resources for all kind of treatments.
- Corporate or Private Hospitals are earning by using this scheme.
- We can clearly there is equal amount of people they are accessing the scheme, when compared to the densities of the population.
- Interesting Facts we got under this analysis is death rate when the people are using this scheme.
- We found there is lot more success rate in both Government and Private hospitals when dealing with the health of the people under this scheme.
- On major analysis is we have seen that the new born infants are the once where the death rate is more.
- According to the community wise, Backward Class are the people who are actively using this scheme.
- This clearly a sign that where the scheme is being used properly, since the scheme is designed for backward people and lower class people with less income.
- That shows the conclusive analysis in different areas precisely.

- Where I can conclude by saying that, exploring the data considered has answered all the questions, and in fact also able to find out some interesting things, while exploring the data .

Reflection:

After Analyzing the data where I can clearly say that, which helped me in understanding actually how much Government is spending on this Scheme.

Actually tried of getting Budget information, actually to know how much Government is allocating every year and how much is being used properly and reaching the Back Ward class people as per the scheme. That will be collected and analyzed.

References to the Data Sets:

Major Data Set NTR Vaidya Seva in the state of Andhra Pradesh

Data Set 1 : http://www.ntrvaidyaseva.ap.gov.in/web/guest/explore_data

We will get data sets year wise data , where I have collected the data year wise and then merged into single data set for exploring the data.

Another Data Set in order to get the District wise population according to 2011 census

Data Set 2 : https://www.census2011.co.in/census/state/andhra+pradesh.html

Another caste based information gathered from

Data Set 3: https://en.wikipedia.org/wiki/Caste_system_in_India

References:

- Wikipedia for caste system in India and in state of Andhra Pradesh
- Stack Overflow
- Data.gov.in
- India 2011 census data
- Ap.gov.in for official data set of different years