**PREDICTIVE MODEL OF STUDENT’S SUCCESS & CAREER COUNSELING USING DATA SCIENCE**

By

**Muhammad Qasim (19k-1612)**



**Report submitted for the**

**Degree of MS (Data Science)**

**National University of Computer & Emerging Sciences**

Karachi-Pakistan

**2019**

**PATEL HOSPITAL PATIENTS EXPLORATORY DATA ANALYSIS USING DATA SCIENCE**

**MS Course Project**

By

**Muhammad Qasim (19K-1612)**

Course Project Supervisor

**Dr. Muhammad Wasim**

**2019**

**National University of Computer & Emerging Sciences**

Karachi-Pakistan

Pakistan

**Table of Contents**

|  |  |  |  |
| --- | --- | --- | --- |
| **1.** | **Introduction** | | 1 |
|  | 1.1. | Problem Statement and Domain | 1 |
|  | 1.2. | List of Requirements | 1 |
|  | 1.3. | Introduction to Data Set | 1 |
|  | 1.4. | Overview of Statistical and Mathematical Methods for Data Science | 1 |
|  |  |  |  |
| **2.** | **Data Analytics** | | 4 |
|  | 2.1. | Age vs Disease | 4 |
|  | 2.2. | Gender vs Disease | 6 |
|  | 2.3. | Most common Disease in female | 7 |
|  | 2.4. | Most common Disease in male | 8 |
|  | 2.5. | In which year female having most common Disease | 9 |
|  | 2.6. | In which year male having most common Disease | 11 |
|  | 2.7. | Male vs Female disease pattern | 13 |
|  | 2.8. | Most common Disease in male and female | 14 |
|  | 2.9. | Which year is most crucial for female | 14 |
|  | 2.10 | Visualization of all results | 15 |
|  |  |  |  |
| **3.** | **Results and Conclusion** | | 16 |

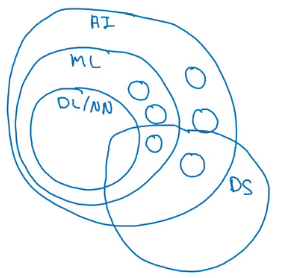
**1. Introduction**

Data Science is the area of study which involves extracting insights from vast amounts of data by the use of various scientific methods, algorithms, and processes. It helps you to discover hidden patterns from the raw data.

The term Data Science has emerged because of the evolution of mathematical statistics, data analysis, and big data.

Data Science is an interdisciplinary field that allows you to extract knowledge from structured or unstructured data. Data science enables you to translate a business problem into a research project and then translate it back into a practical solution. Key components of data science are given in figure 1.

Page1

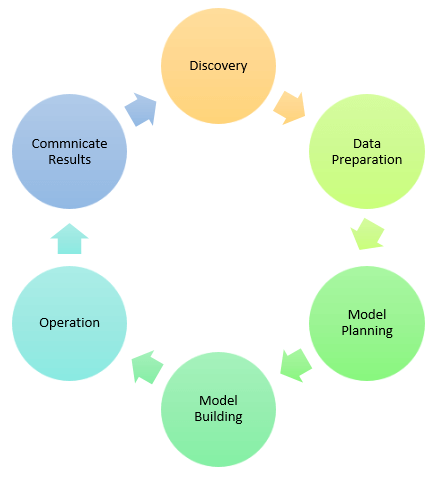


*Figure1. Key Components of Data Science*

## Data Science Process

Data science process is consists of six key modules  discovery, data preparation, model planning, model building, operation and communicate results. These processes are given in figure 2.

Page2



*Figure2. Data Science Processes Model*

### i. Discovery

Discovery step involves acquiring data from all the identified internal & external sources which helps you to answer the business question.

The data can be:

* Provided by Patel Hospital

### ii. Exploratory Analysis

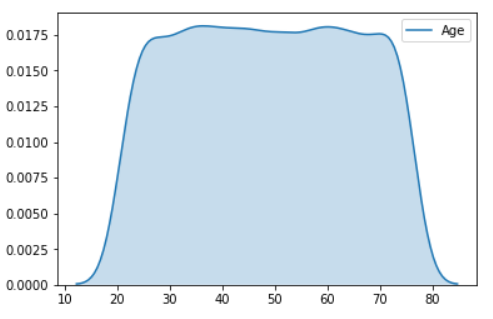
We have displayed best graph using matplot and seaborn visualization tools in python. We found some trends during Analysis. Now we can see each analysis on by one.

Page3

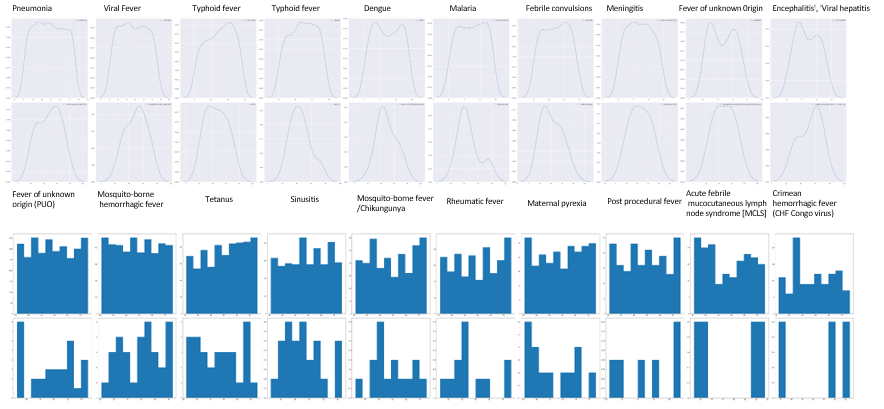
**2.1 Age vs Disease**

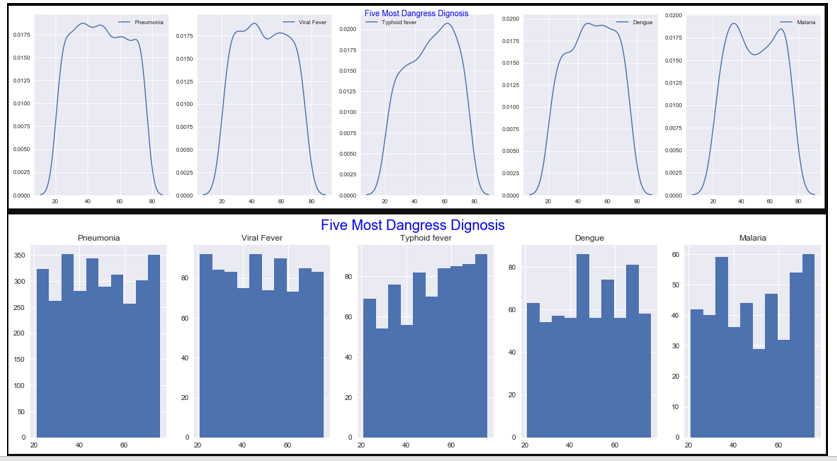
Patients age starting from 12 to 90

25 to 75 look like uniform distribution

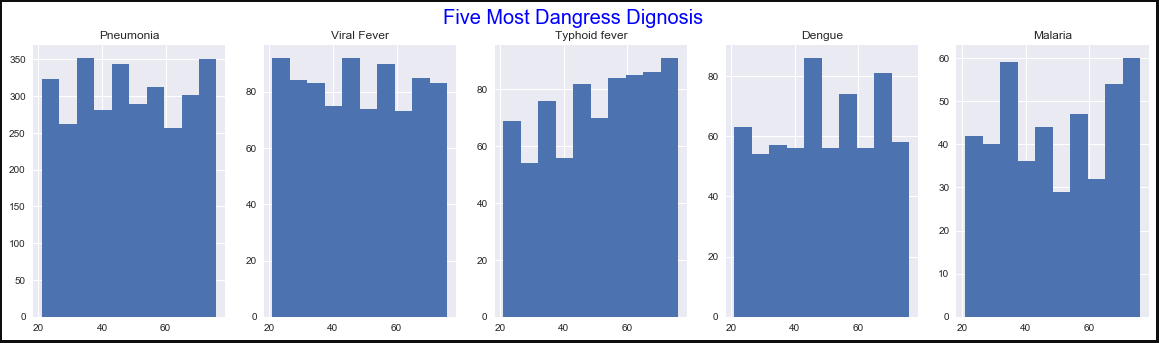


Page4





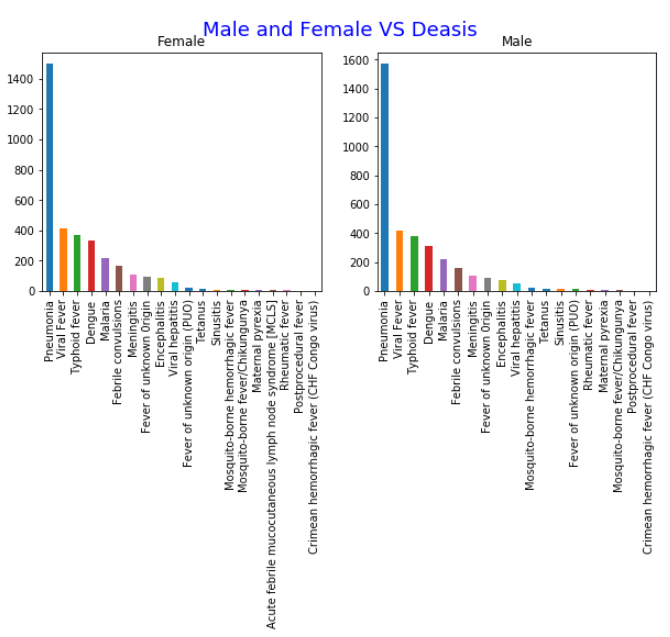
Page5



* Phenomena mots effected patients
  + 25 to 55 oldest people
  + 65 to 70 oldest people
* Viral Fever most effected patients
  + 20 to 38 or 50 to 55 oldest people
* Typhoid most effected patients
  + 55 to 60 oldest people oldest people

Page5

**2.2 Gender vs Disease**



* Most Common disease in males (3,480)
  + Pneumonia (1,572)
  + Viral Fever (418)
  + Typhoid Fever (381)
* Most Common disease in Females (3,427)
  + Pneumonia (1500)
  + Viral Fever (413)
  + Typhoid Fever (372)

Most Common disease in males (3,480)

Pneumonia (1,572)

Viral Fever (418)

Typhoid Fever (381)

Most Common disease in Females (3,427)

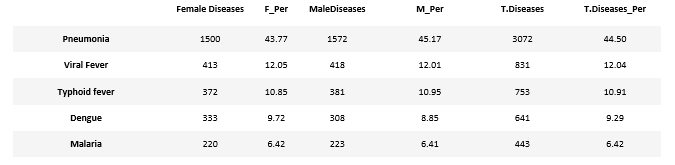
Pneumonia (1500)

Viral Fever (413)

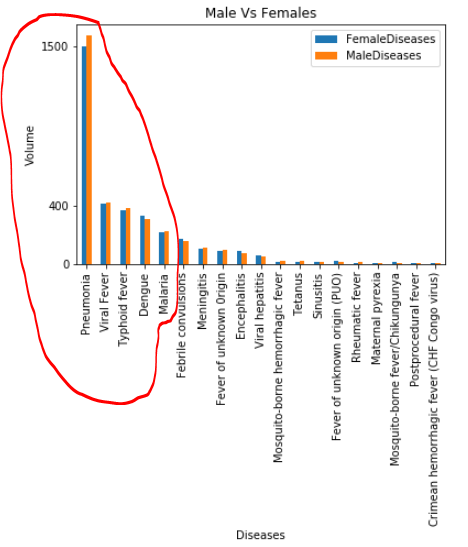
Typhoid Fever (372)

Page6

Below table show Top Five Disease:

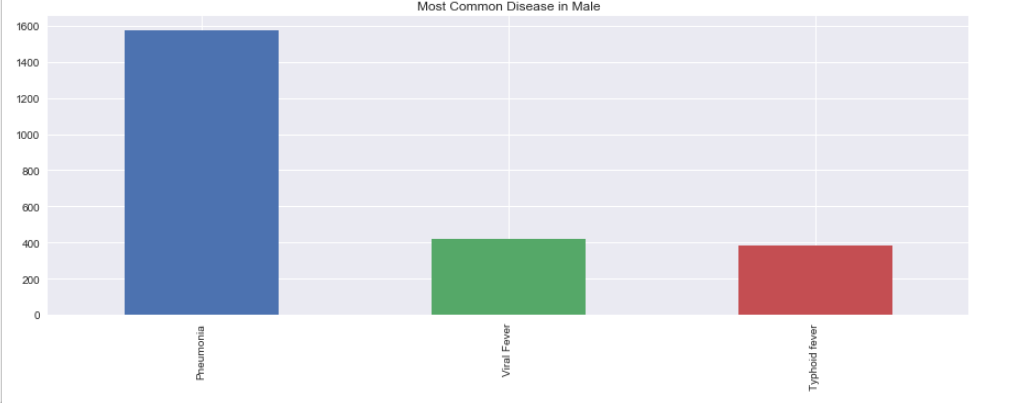


1. Pneumonia
   1. Effect male more than Females
   2. Most top disease effect to both
2. Viral Fever
   1. Effect females more than Males
3. Typhoid Fever
   1. Effect males more than Females
   2. Pneumonia
      1. Effect male more than Females
      2. These diseases are effected to both
   3. Viral Fever
      1. Effect females more than Males
   4. Typhoid Fever
      1. Effect males more than Females



Page7

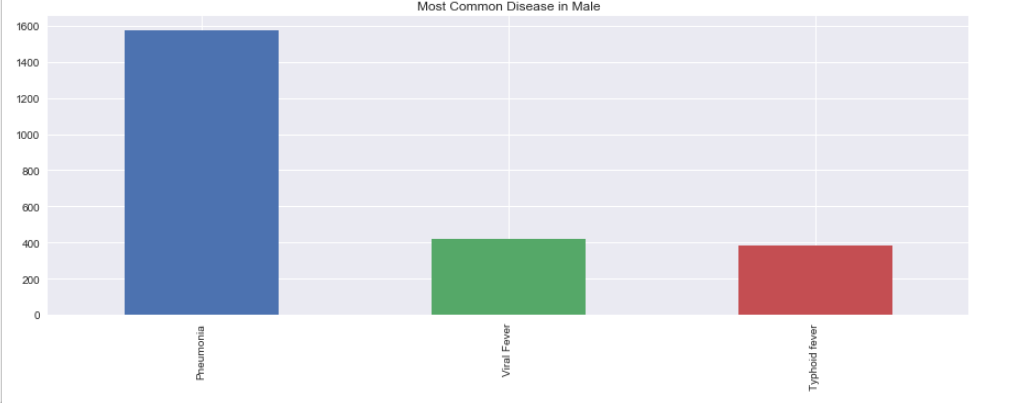
1. **Most common Disease in Females**

Pneumonia 1500

Viral Fever 413

Typhoid fever 372

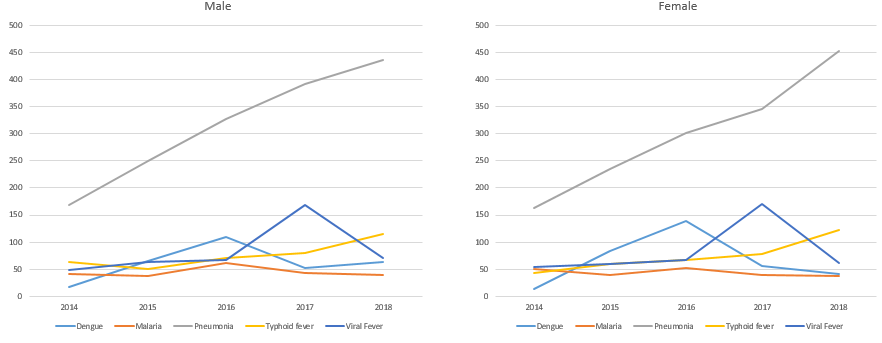
1. **Most common Disease in Male**



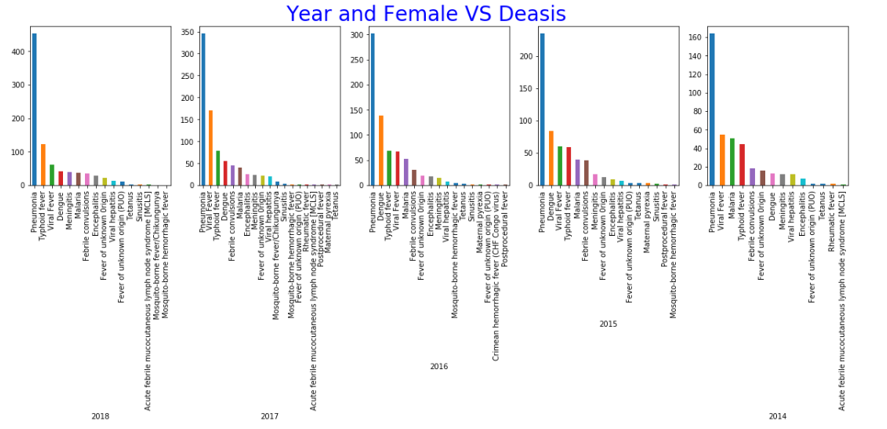
* Pneumonia 1,572
* Viral Fever 418
* Typhoid fever 381

Page8

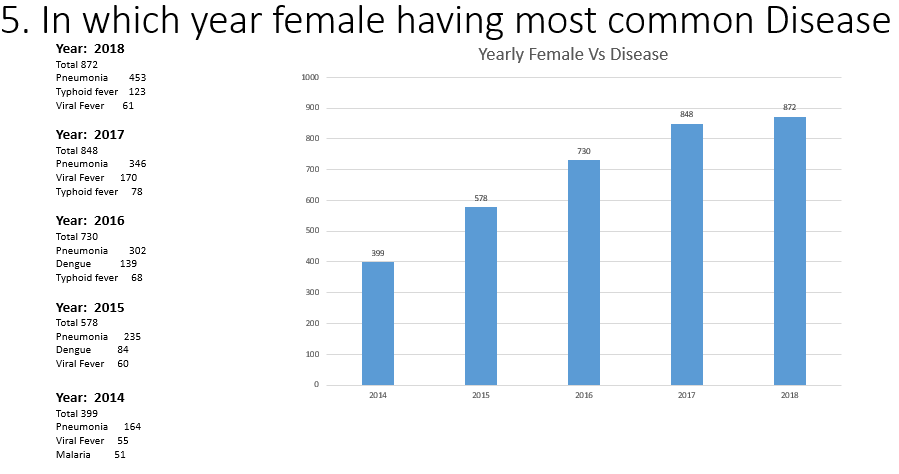
### Yearly Male VS female having most common Disease

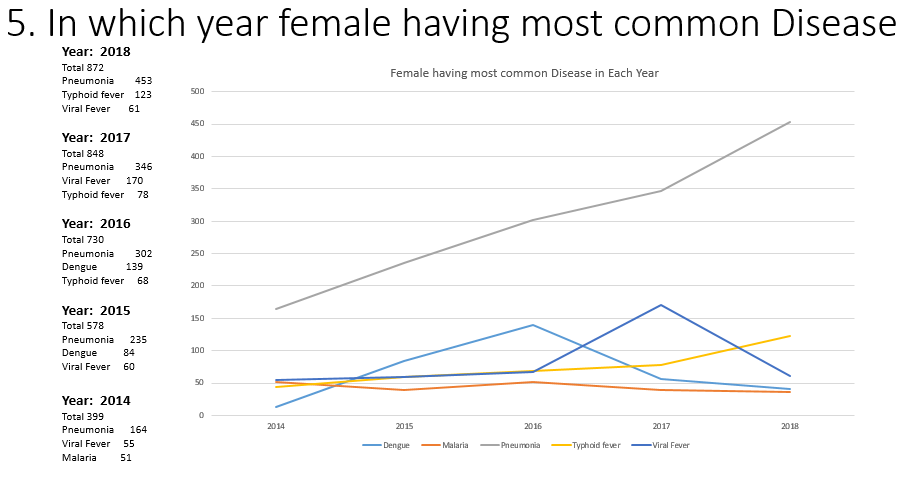


**5. In which year female having most common Disease**

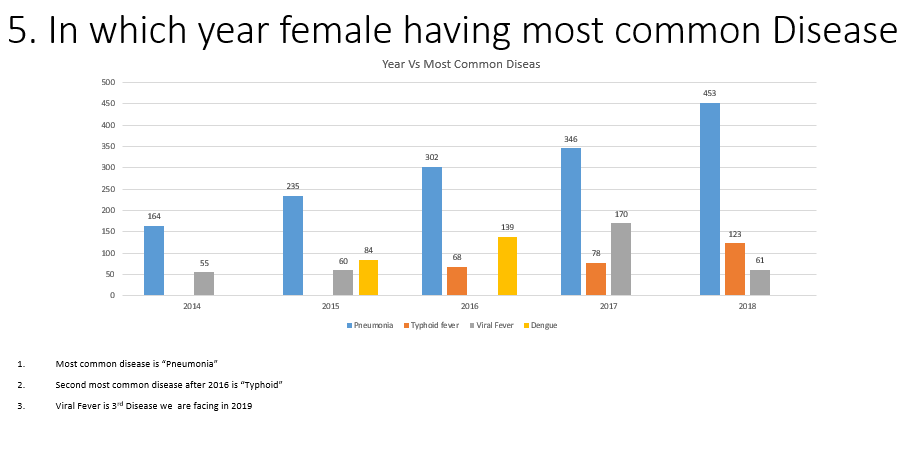


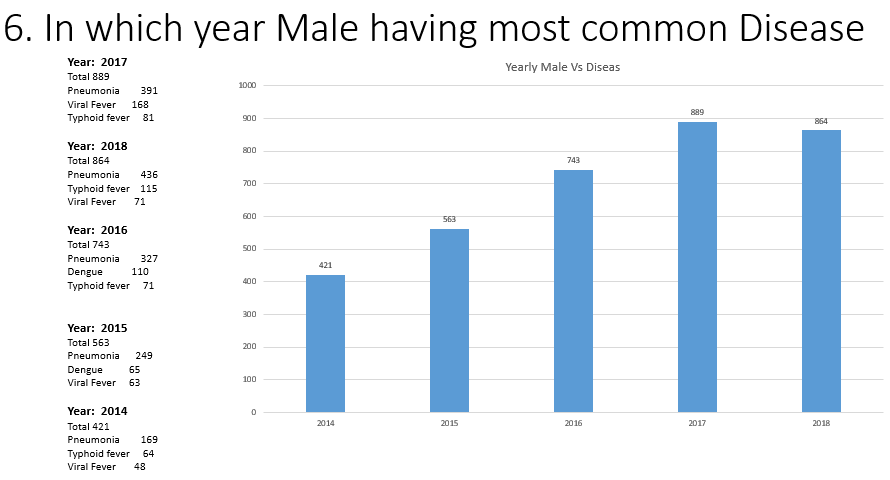
Page9



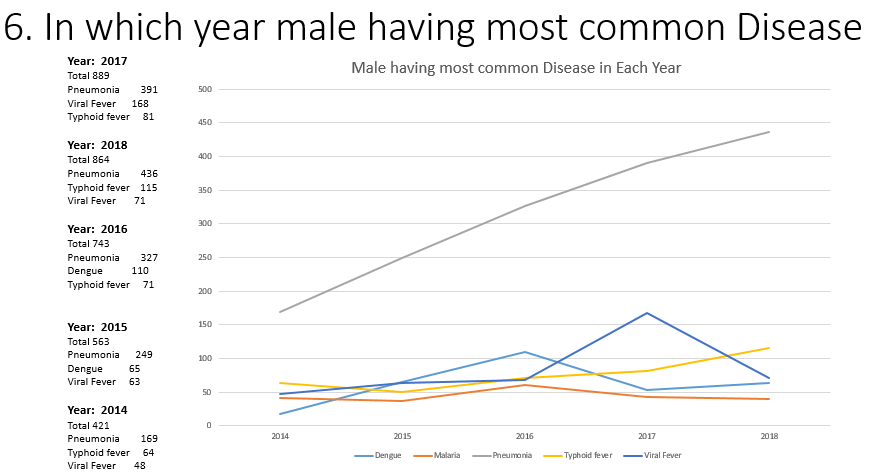


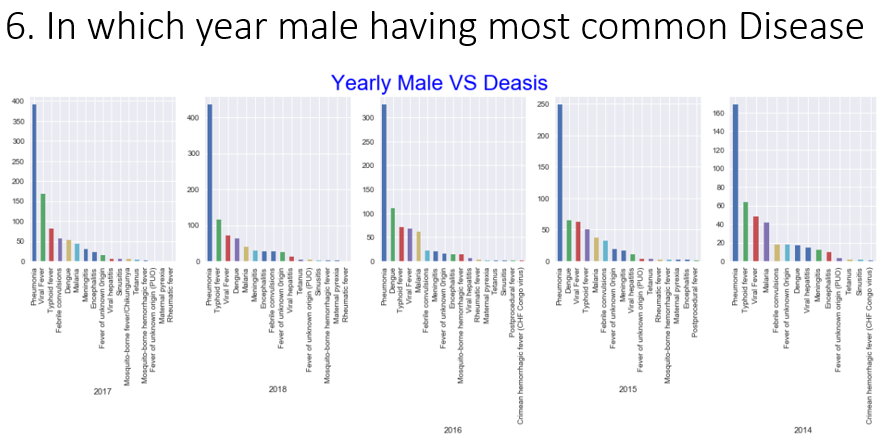
Page10



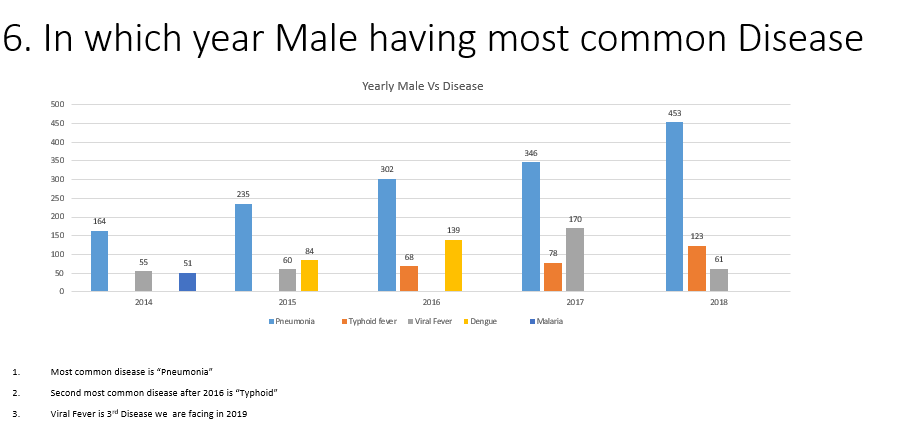


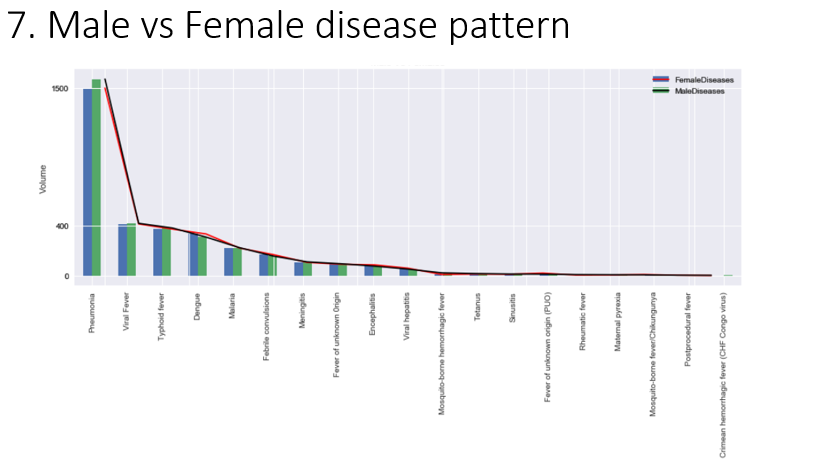
Page11



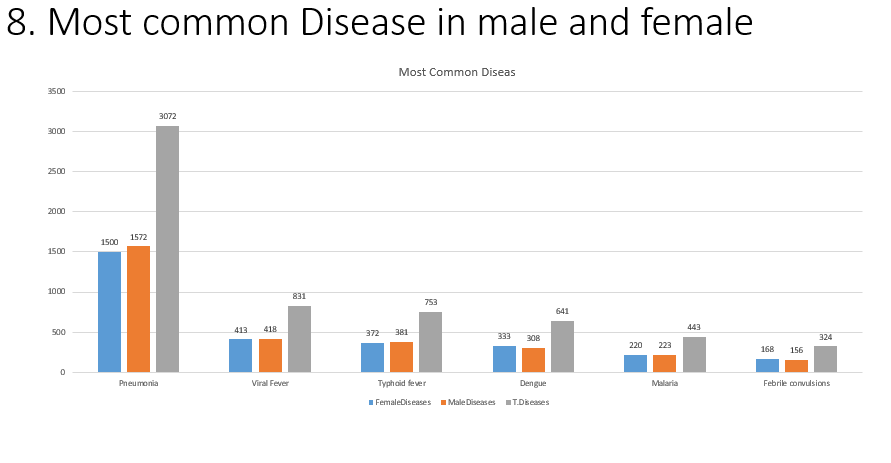


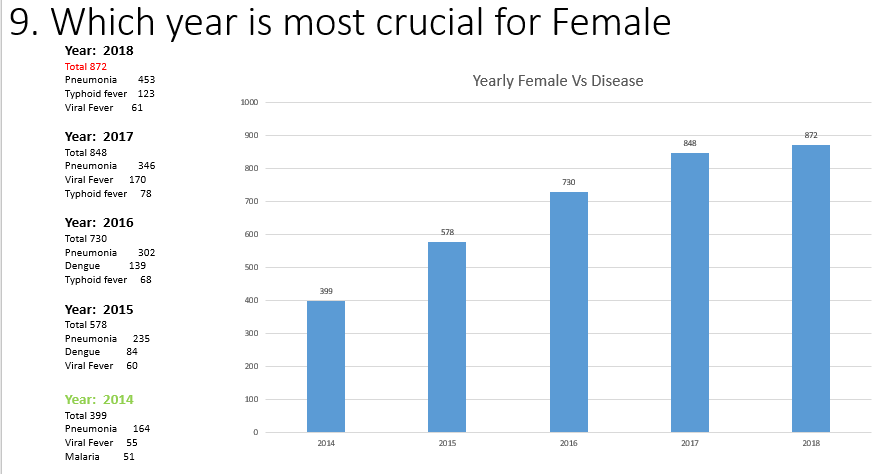
Page12



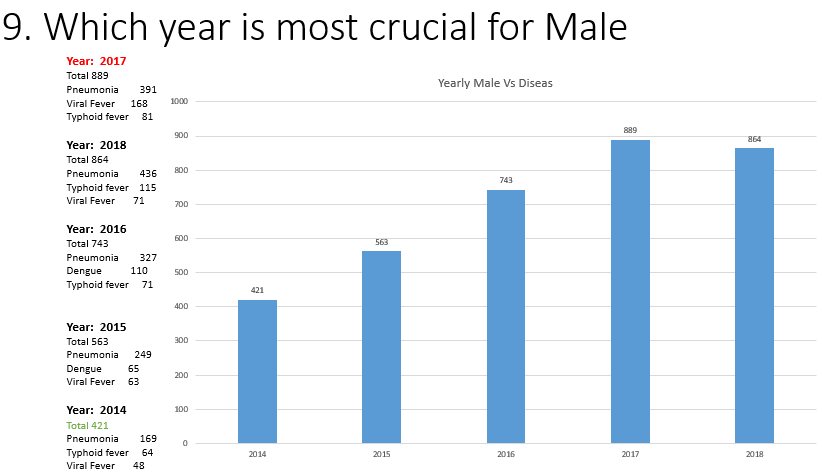


Page13





Page14



**10. Visualization of all results**

1. Seaborn, matplotlib and pandas plot
2. Bar chart, line chart
3. Subplots

Page15

### iii. Model Planning

using Machine learning technique to draw the relation between input variables. Planning for a model is performed by using different statistical formulas and visualization tools. SQL analysis services, ML, and DL are some of the tools used for this purpose.

### iv. Model Building

In this step, the actual model building process starts. Here, Data scientist distributes datasets for training and testing. Techniques like association, classification, and clustering are applied to the training data set. The model once prepared is tested against the "testing" dataset.

### v. Operationalize

In this stage, you deliver the final baselined model with reports, code, and technical documents. Model is deployed into a real-time production environment after thorough testing.

### vi. Communicate Results

In this stage, the key findings are communicated to all stakeholders. This helps you to decide if the results of the project are a success or a failure based on the inputs from the model.

Page16