Term Project
World Heath Report Analysis of Tuberculosis

PARTICIPANT'S NAME: ARUN SHARMA (UB Person: 50206920)

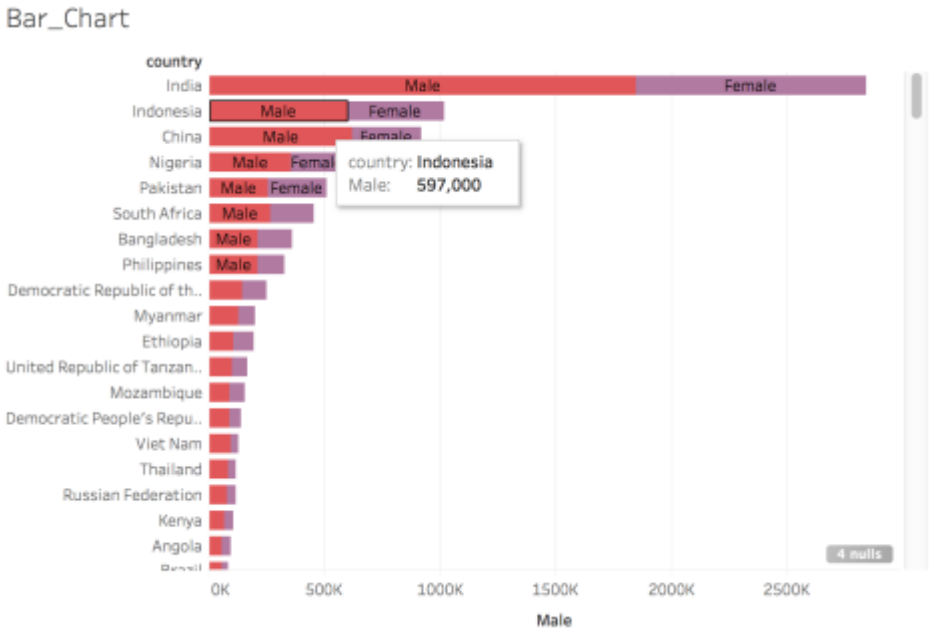
DATASET: Data provided by countries to WHO and estimates of TB burden generated by WHO for the *Global Tuberculosis Report* http://www.who.int/tb/country/data/download/en/
The dataset has around 3500 instance and 48 attributes
Dataset collected was of TB burden countries over the years.

Schema: Most of the schema columns is self-explanatory. Many of the column attributes of schema are not considered. The ones which are considered given below:
Death of TB, Deaths of HIV, Deaths with both TB and HIV, Incidents, Male, Female, Percentage, Detection Rate, Total Incidents and Ratio.

STORY:
Sheet 1: Gives an overall analysis of the fraction of the population affected by the user.

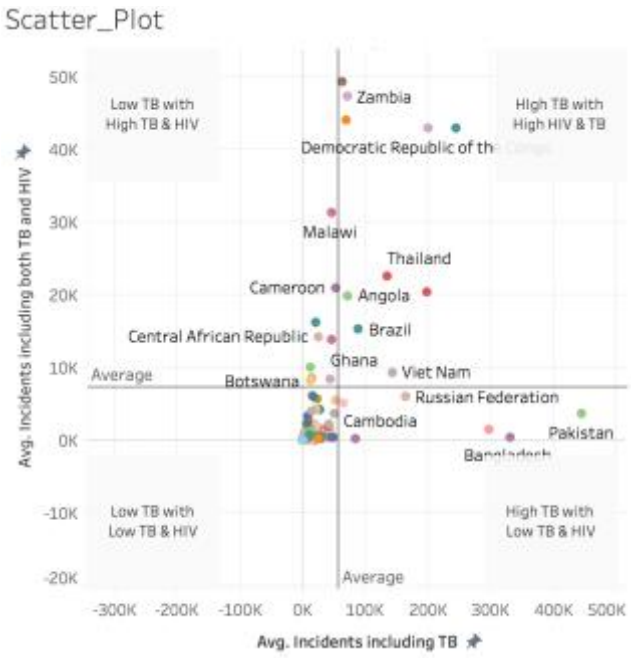
From the analysis, African countries such as South Africa, Namibia have the highest percentage of affected population followed by Asian countries like India, Pakistan, Bhutan etc. Sheet 2:

Giving gender based analysis of affected population all over the world. The bar chart gives country-wise exact number of male & female affected by TB.



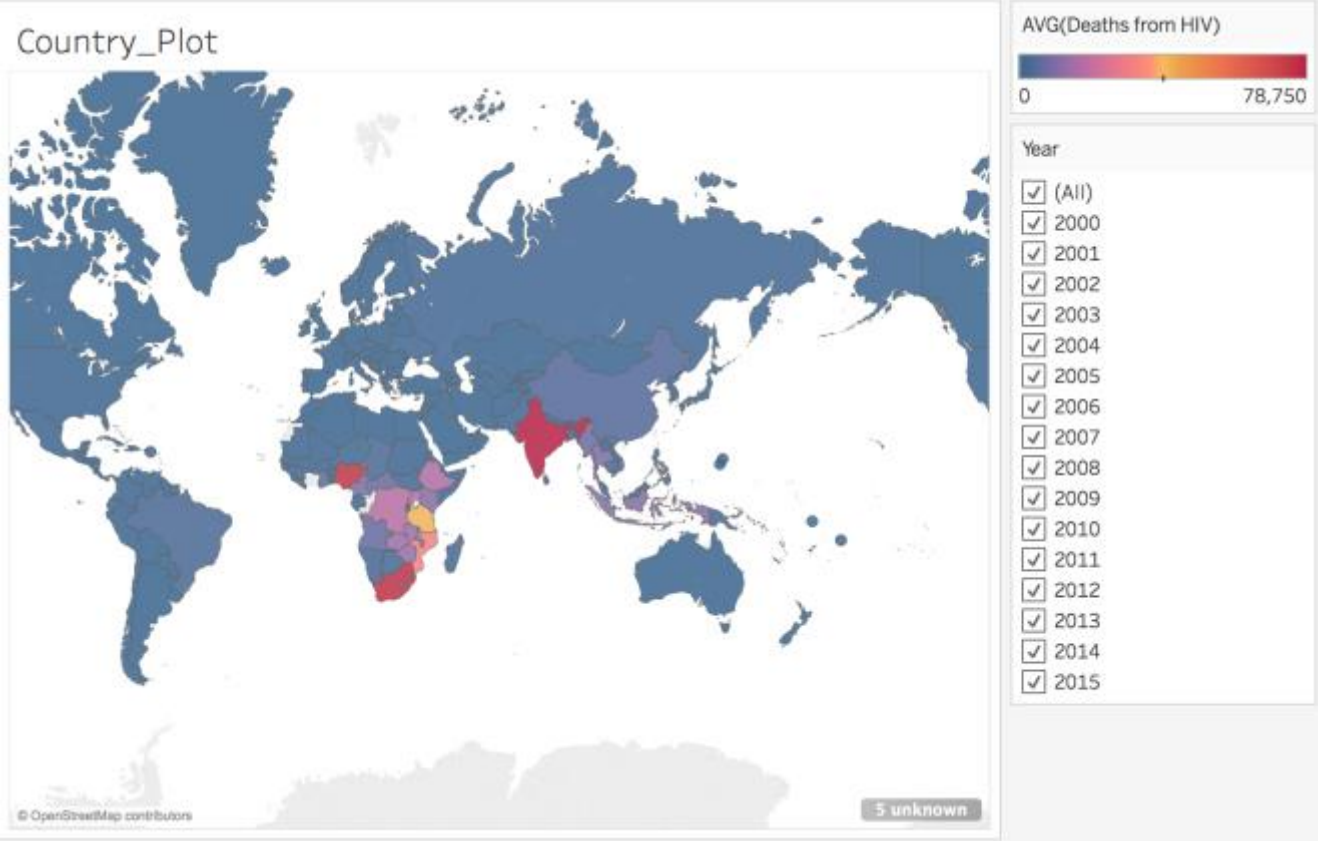
From the analysis, we conclude that males are most likely affected than females. This analysis may vary with actual male and female population present.

Sheet 3: Gives year-wise and country-wise correlation data of HIV and TB where X-axis shows average incidents of TB and Y axis gives average incidents including of TB and HIV.

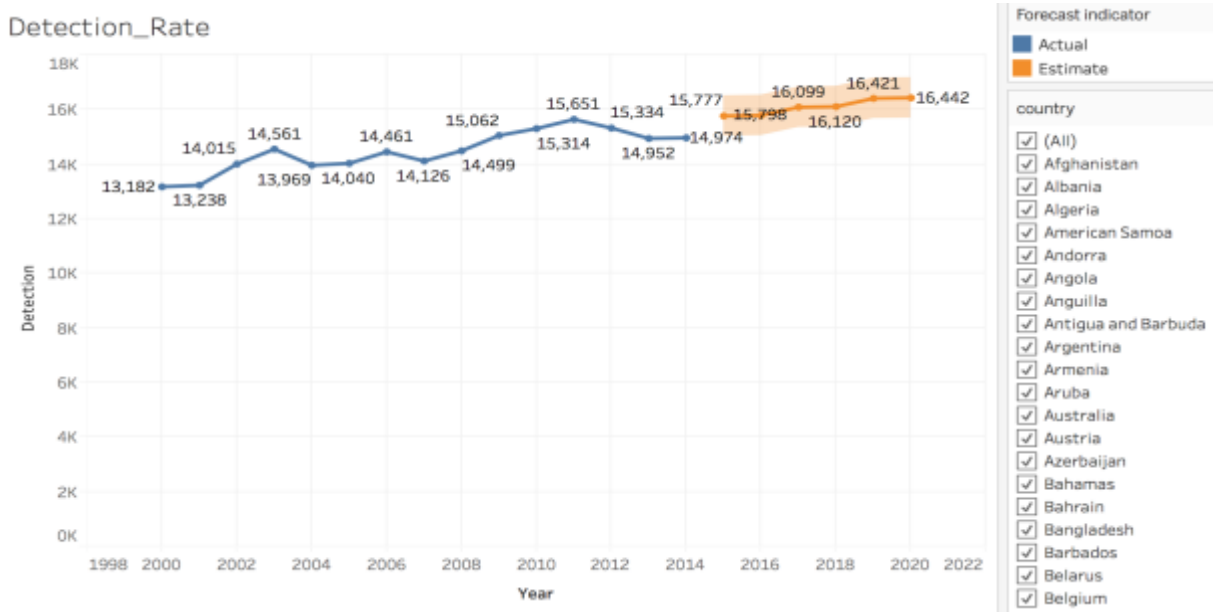


High TB with High HIV and TB: Countries which have high TB will most likely have incidents with HIV.
Low TB with High TB & HIV: Countries which have low TB incidents but has high TB and HIV incidents. This means that mostly those incidents are independent to TB.
Low TB with Low TB & HIV: Countries which are not affected with either of them.
Low TB with high TB and HIV: Countries which are less affected with TB but those cases are mostly TB and HIV.

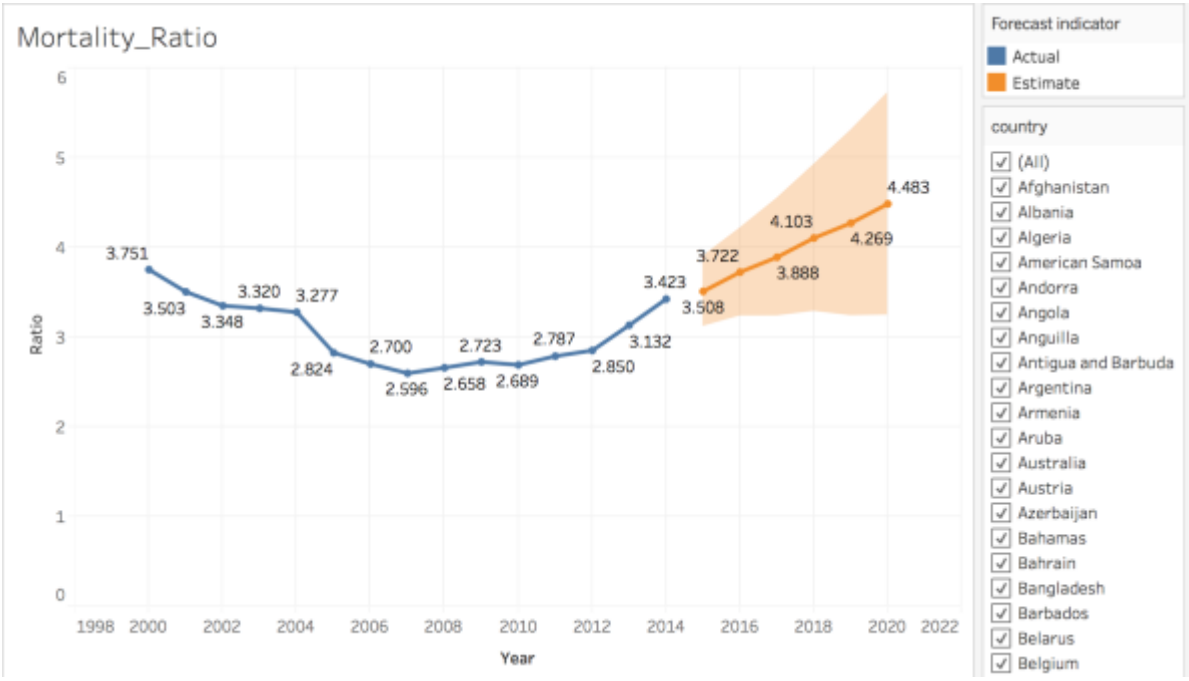
Sheet 4: Gives an overall year-wise analysis on the world map of average mortality rate of with HIV throughout the world along with the information of TB cases.



Sheet 5: Gives the prediction of number of cases which are detected as TB. Year-wise prediction analysis of data was done for next 3 years.



Sheet 6: Gives the prediction of year-wise mortality ratio of total death due to TB over total deaths due to HIV. The prediction indicates that the ratio increases year-wise since the number of deaths due to TB will be greater than deaths due to HIV.



Data Product: The goal is to study the healthcare data and perform statistical and predictive analysis giving a study Tuberculosis and its correlation with HIV over the years in different countries.

We have given country wise and year wise statistics and by these graphs we found many conclusion about what is the status of TB disease around the word and how does this affect the overall population through percentage and gender.

Also, gives an intuition that people who are affected by TB are most likely affected by HIV in some of the countries.

The world map shows the regions which are affected the most and finally the predictive analysis of some complex measures such as detection rate over next 3 years and the ratio of people who are affected by Tb to people who are affected by HIV.

Finally, we conclude that in some countries especially in Africa, the problems still persist because of lack of treatment facilities or awareness which should be resolved asap.

Channel: The product is channeled through web application on Tableau public with link below: https://public.tableau.com/profile/publish/Term_Project/Story#!/publish-confirm