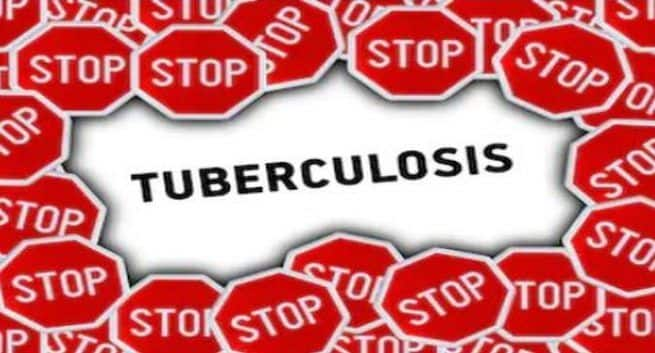
**Tuberculosis (TB)**



Tuberculosis (TB) is a disease caused by bacteria called Mycobacterium tuberculosis. The bacteria usually attack the lungs, but they can also damage other parts of the body.

TB spreads through the air when a person with TB of the lungs or throat coughs, sneezes, or talks. If you have been exposed, you should go to your doctor for tests. You are more likely to get TB if you have a weak immune system.

Symptoms of TB in the lungs may include

* A bad cough that lasts 3 weeks or longer
* Weight loss
* Loss of appetite
* Coughing up blood or mucus
* Weakness or fatigue
* Fever
* Night sweats

### TB Prevalence

### Prevalence refers to the number of cases of TB that are present in a particular population at a given time. Prevalence includes newly diagnosed people, plus people who were diagnosed in the past, and people who haven't even been diagnosed. Prevalence is usually, but not always given as a percentage of the population.

### The prevalence comes from surveys which are only produced for a limited number of countries.

### World Health Organisation (WHO) statistics

### All countries are asked to report their TB figures to the WHO. The [countries with TB](https://tbfacts.org/countries-with-tb/) page gives the notified figures for almost every country in the world.

### WHO uses the notified figures to help produce estimated total TB incidence statistics for each country. These and other statistics are the figures referred to on this page.

### TB related deaths

### TB is one of the top ten leading causes of death worldwide and the leading cause from a single infectious agent, ranking above HIV/AIDS.

In 2017 there were a total of 1,600,000 TB related deaths.

Also, an estimated 234,000 children died of TB in 2017 including children with HIV associated TB. People who have both [TB and HIV](https://tbfacts.org/tb-hiv/) when they die, are internationally classified as having died from HIV.

There is more about [deaths from TB](https://tbfacts.org/tb-deaths/) including the 234,000 mainly preventable deaths in children.

Dataset Information

The dataset given is about TB prevalence, all forms (per 100000 population per year) in different countries. Group countries based on how similar their situation has been year-by-year to understand the world situation regarding the tuberculosis disease.

Note:

The cluster information is given for reference.

Overview

Here, I have used clustering method to cluster countries into 3 groups.

* Countries with minimum TB Prevalence
* Countries with medium TB Prevalence
* Countries with maximum TB Prevalence

I have used KMeans clustering method to cluster all the countries into 3 groups. And then I have taken Decision tree (with PCA) as the classification model.