**Data Collection**

This project is based on the use case mention in the paper below: -

<https://projects.iq.harvard.edu/covid-pm/home>

According to this paper we are assuming there is 8% chance of people being affected by covid 19, when they are exposed to polluted area.

The size of the covid 19 particle varies from 65–125 nm, So I have collected the microclimate data having size pm2.5 for our use case.

I have collected the data from the following website:

<https://openaq.org/>

The following areas data have been collected: -

1. Alphington
2. Altona North 1
3. Altona North 2
4. Box Hill
5. Brighton
6. Brooklyn
7. Campbellfield
8. Coolaroo
9. Dallas
10. Footscray
11. Macleod
12. Melbourne CBD
13. MRI Campbellfield fire
14. Thomastown East

Data is gathered using the code in file *data\_collection.ipynb*

[https://deakin365.sharepoint.com/sites/Data2IntelligenceConsulting/Shared Documents/D2I (Melbourne City)/T2\_2020/Microclimate\_data/Pavan Work/data\_collection.ipynb](https://deakin365.sharepoint.com/sites/Data2IntelligenceConsulting/Shared%20Documents/D2I%20(Melbourne%20City)/T2_2020/Microclimate_data/Pavan%20Work/data_collection.ipynb)

After the data is collected in have Identified that some dates and the values are missing.

So, I have uploaded the data in the MSSQL and wrote the code to create the missing date.