**Accomplished So far:**

* Data collection

1. Air Quality Data: We have collected Fine Particulate Matter (PM2.5) data from Bay Area Air Quality Management site using a python script that does some pre-processing.
2. Purple Air Data: Live data capture API to do prediction on number of illness cases that can be expected during any given week.
3. Air borne diseases data:
4. All deaths related with Air Borne diseases
5. Asthma1
6. Asthma2
7. Asthma3
8. Asthma4: We have collected this data set from tracking california site writing python script
9. ILI (influenza-like-illness) emergency visits — a dataset representing emergency room visits from Flu like sickness.

* Data preprocessing and understanding the data: Each type of data set is being analyzed individually to understand the signature of data and come up with a pre-processing technique.Then we are trying to find correlation between air quality data and air borne diseases data.
* Initial data visualization: Pair plot, Correlation plot, swarm plot, histogram on various period data

**Findings and Results so far:**

* The PM2.5 – fine particle index is one of the key parameters that contribute towards the air quality index.
* Effects of air quality on various air borne disease cases vary by season and location. This is seen by simply looking at ILI cases during flu season vs. those during other parts of the year.
* Asthma cases seem to be more directly related to the air quality being breathed by the subjects.

**Difficulties:**

* Time constraint is a major difficulty. All 4 of us are full time employees.
* Finding up to date data for all of the different datasets as we are collecting data from different sources.

**Remaining:**

* Model design (11/14)
* Model evaluation and tweaking (11/21)
* Holistic results generation regarding effects of PM 2.5 AQI on various air borne illnesses. (11/28)
* Problem statement refinement depending on the relationship between P2.5 AQI and various air-borne illness data. Possibly extending the problem statement to cover more ground if needed. (Parallely)
* Model deployment
* Project report