DASS Team 30
Sprint 2

Project Ideas

Study of existing work in the field

Introduction

Given the sensor data, we looked through work done on similar data by existing companies and for research done in the field. Based on this we came up with our final model, trying to incorporate some of the better features from the existing products while keeping in mind the feasibility of the feature.

Major Product in this category: Ambee

Ambee is a Bangalore based start up that sells individual air quality monitoring units to single consumers, while also providing data through an API on a subscription basis for researchers/developers.

The Ambee end-user product - Air Quality Tracker

Ambee sells individual units that monitor various pollutants in the immediate environment of the user. It has an interface on the unit itself to display the real time values, along with an app which allows visualization by providing the local heat map based on the user's sensor. It also provides the user a log of their past readings.

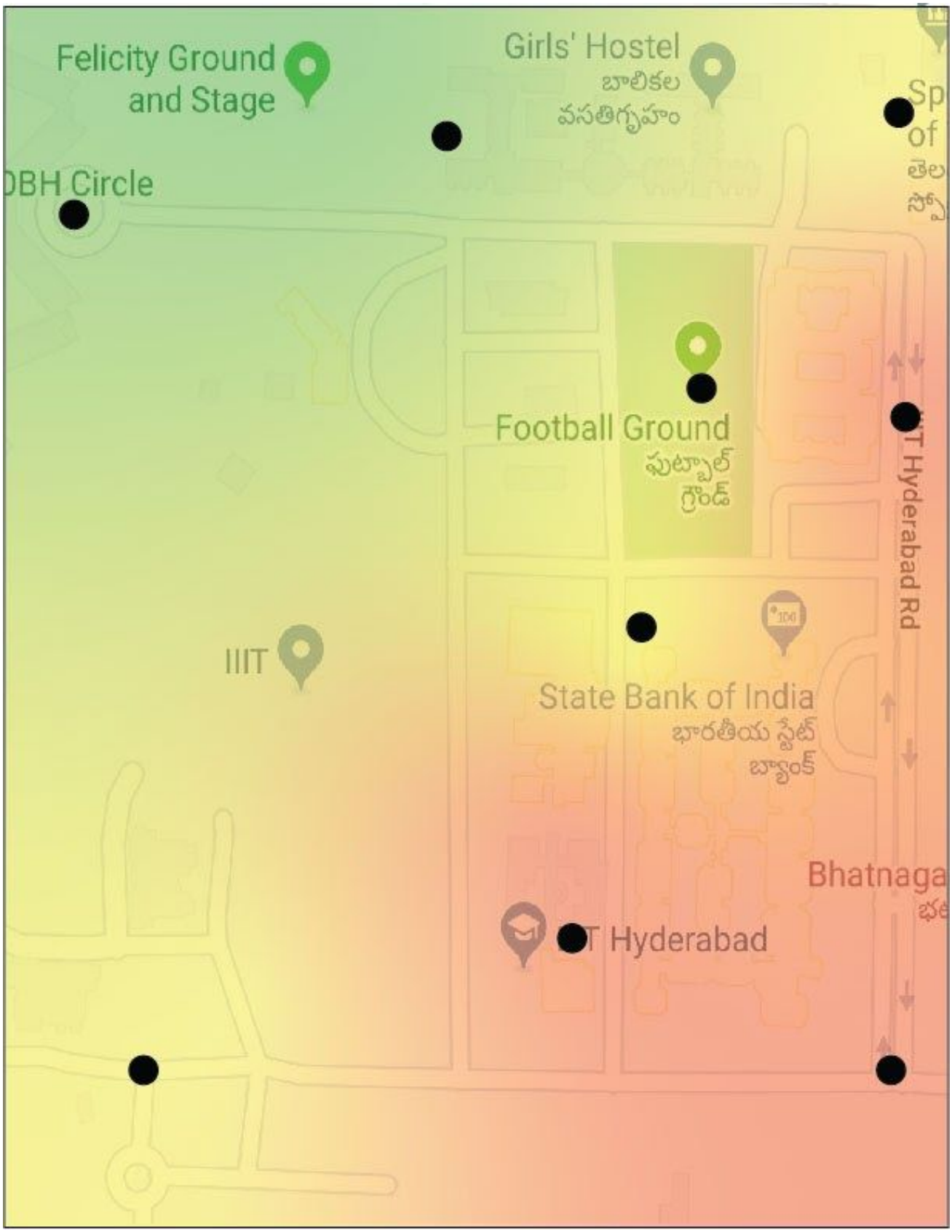
The Ambee API - raw data for analytics

The API interface provides raw data for use of subscriber, in a JSON format. The data provided is claimed to be realtime, with a 300 ms delay. It also provides a history for each place, which is a number of previous data points corresponding to that place.

Incorporating features of this product

We have integrated the heatmap feature of the individual sensor, and expanded it to show the combined readings of all the installed sensors on the campus, as well as the readings of each sensor. Having a separate interface for each sensor hinders the overall objective of visualizing the pollution data across the campus. Neither the API nor the individual sensors sold by Ambee provide any analysis on the data. Our system shall use data analytics for various purposes (expanded further).

On the next page, we have attached a sample image of what we are intending the heat map to look like. For this image, we have not used actual data, and it is only a conceptual representation. The color coding in this image is Red for Dangerously High values, Yellow for moderate and green for regions where the readings would be in the safe zone. These features are subject to change if needed.


Felicity Ground
and Stage
Girls' Hostel
బాలికల
వసతిగృహం
DBH Circle
Football Ground
ఫుట్బాల్
గ్రౌండ్
T Hyderabad Rd
IIIT
State Bank of India
భారతీయ స్టేట్
బ్యాంక్
Bhatnaga
T Hyderabad

Data Analytics

We went through the honors project report of Sai Krishna Charan as advised by the client and were interested in the following techniques, which we found relevant to our purpose

1. Uni Variate analysis via parabolistic Histograms to see trends of data

2. Multivariate analysis via

   a. Parallel coordinates to display correlation between parameters.

   b. Stacked Bar representation shows relationship between attributes

After some consideration, we decided on the following ways of representing data:

Parabolic Histograms with stacking

For each sensor, we decided to have a histogram for the last 24 hours of data. Parabolic Histograms help in identifying the general trend of the that particular parameter over time. On top of that, by allowing stacking (or overlapping) histograms, it becomes very easy to visualize the correlation between the different parameters. While a correlation matrix can give a more accurate correlation, it is not at all intuitive to a user without an accurate understanding of what it actually is, and since our targeted user base need not have that, we decided to instead use parabolic histograms, which are much more intuitive.

Live Line Charts

Along with the histograms, we will also be presenting the data in line graphs which will be updated in real time as we receive the data from the sensors. We will deal with missing or erroneous data by using interpolation to ensure coherency of the data.

Statistical Analysis

Along with histograms, we will also be providing basic statistics of each parameter, including hourly, 8-hourly and daily averages. Along with that, we are looking at coming up with a way to combine all these parameters to create an all round “Health Index”, and do

the same analysis on this combined parameter (along with the histograms mentioned above).

We also will provide statistics in fixed time slots (eg. 10 AM to 12 PM, 12 PM to 2 PM and so on). This is done primarily to help with open air event scheduling, such sports matches/competitions or cultural events at the amphitheater. We came up with this after observing recent news headlines of major sports events being affected by bad weather conditions - India vs Sri Lanka Test in New Delhi, where play was halted due to bowlers having trouble breathing in polluted air, players being forced to withdraw from the Australian Open Tennis Championship because of excessive smoke, bad light regularly ending Test Cricket days prematurely to name a few.

Conclusion

We came with two methods to visualize the data:

1. A real time heat map of the campus to show current air conditions on campus.
2. Histograms to allow visualization of the correlation and time-dependent trends of the various parameters.

Apart from this, we use basic statistics such as averages on all the parameters to help with more specific tasks such as event scheduling.