

Get into groups of 3-4 and open the document that was shared through EdStem and on the DATA 6 website. Write your group name, members' names, and the area of study you are the most curious about.

In groups, answer the following questions. You will be given 25 minutes to work, and present afterwards!

1. What [Air District data](#) interests you for further exploration? There's plenty in the Public Data Center. Discuss your interests, then select one data type for the group.
2. How did your group select this data? Describe your group's decision-making process.
3. What are some compelling questions to investigate using this data? How would you approach exploring it? What insights are you hoping to gain about the world?

Example)

1.
 - a. Person A: Air quality measurement
 - b. Person B: how it was collected/ who is collecting the data
 - c. Person C: Open Burn Maps
 - Group's decision:
 - Emission break downs
2. We decided on emission breakdowns because ... this data seemed to be collected through ...
3. We saw decrease in emission overall, but there was one huge drop during ... We are curious if this has to do with off-road transportation or policies during that time that passed. We would approach using this data by Through this, we hope to understand what factors impact the emission break downs the most.

Kit-Kat : Group 1 - Claudia, Billy, Shreya

What we breathe:

1. The **"Air Quality Forecasts"** under What you Breathe interested our group because it is interesting to see up-to-date information on air quality conditions across the Bay

Area– also looking at pollutant levels for substances like ozone, particulate matter and nitrogen oxide.

2. In the forecast we can see that most of the bay has an index of 0-50 air quality meaning we are in the good range, we are able to measure the air quality by meteorologists using monitoring stations with sensors that detect elements within the air like nitrogen dioxide and sulfur dioxide along with satellite imagery and Performing PM2.5 and PM10 gravimetric analysis over 30 stations.
3. Immediate impacts of pollution on public health, during wildfires or high smog-days
 - Looking at the kinds of trends in data within specific times or different locations in the Bay Area (like how proximity to Bay impacts pollution) and we could identify areas that were disproportionately affected.

TooSleepy: Group 2 - Vinayak, Ricardo, Andy

1. Air Quality Index in Bay Area. The data is taken as air quality measurements of pollutants at the Air district's air monitoring stations. An AQI scale is provided serving as a categorical ordinal representation of quantitative data
2. We decided on Air Quality index because we could see the difference in air quality in the bay area over time. They took samples of air and tested how much of each pollutant (ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide, and fine and particulate matter) was in the air. They then compared these ratios to the federal standard and then quantified each rating with their scale, where 0-50 was good, 51-100 was moderate, 101-150 was unhealthy for some, 151-200 was unhealthy for most, 201-300 was very unhealthy, and over 301 was considered hazardous. They measured data for each pollutant daily in each place.
3. It would be interesting to study the relationship between air quality improvements with the enforcement of environmental laws or changes in the pollutant industry. We can also ask questions about how the economy, population growth, weather, and events or policies show how they influence air quality. A possible approach to tackle the data and its implications would be to stack it up to these variables and draw generalized conclusions

Group 3: Zipeng, Yaxin, Jerry

- Air quality in Santa Clara Valley (Silicon Valley)

- Reason: As a hub for high-tech companies, Silicon Valley has unique emission sources, such as data centers and electronics manufacturing.
- Reason: From a monthly dimension, it seems to have weaker air quality and more data than other areas.
- There's a spike of Ozone (O₃) pollution during June-October (Seasonal?)
- What are the specific areas or companies that contribute to the most pollution?
- Companies are recently promoting "Net zero emissions", are they really effective or improving air quality compared to the past.

- ☐ One thing that interests me is about the ozone measurement. In the website, it says "the line graph shows the hourly average concentrations for ozone, measured over the full day". For places such as Santa Clara Valley, there are separate line plots that describe the measurement of ozone in different sub-areas of the valley. The unit is parts per million and measurement ranges over either 1 hour or 8 hours. I notice that the two of the towns have similar ozone levels (Los Gatos and Gilroy). I wonder the reasons behind it other than the short distance between them.

Group 4 : Harry, Lucas, 3

- a. Person A: Air quality measurement
- b. Person B: Open Burn Maps
- c. Person C: Air quality measurement
- Group's decision:
 - Air Quality in Oakland, CA
- 4. We decided on Air Quality measurement because two of three group members chose this topic as the one that interested them the most. This data seemed to be collected and measured through data on various pollutants like particulate matter, ozone, and nitrogen dioxide at different locations across the region. To be specific, we decided to focus on Oakland, CA.
- 5. Oakland's air quality is good (0-50, according to the index). A lot of the most significant air pollutants come from oil-based companies (oil refineries, etc). Some questions we can use this data to investigate include "Which type of company releases

the most air pollutants (military-based, oil-refinery, etc) in an area?” “Is the air quality of an area closer to the ocean affected in any way by its location?” “Has air quality gotten worse over time? Are there any growth patterns we can notice over an X number of years? X number of months?” “At what degree do air pollutants noticeably affect air quality?”

6. Data investigation using this dataset would benefit communities and neighborhoods found in these areas by finding potential ways to reduce air pollution. The identification of companies who release the most pollutants would put a social focus on either finding a solution or getting rid of the company in the region in general.

Isabella's Section

Group 1

Sophia, Sherry, Ted, Caroline, Brody

1. What [Air District data](#) interests you for further exploration? There's plenty in the Public Data Center. Discuss your interests, then select one data type for the group.

We're choosing the air quality index. What we found interesting is that the air quality index shows the air quality history. Additionally, we found it interesting that some regions have worse air quality than others. Northern Zone and Coast/Central Bay have the better air quality in the past year whereas the Eastern Zone, South Central Bay, and Santa Clara Valley have a little worse air quality.

2. How did your group select this data? Describe your group's decision-making process.

We chose to analyze AQI (air quality index) over time because it gives an easily interpretable number that quantifies how healthy it is to breathe the air in a given area. Additionally, it shows how the air quality changes over the course for the year, and can be used to assume which areas will have different air qualities and when.

3. What are some compelling questions to investigate using this data? How would you approach exploring it? What insights are you hoping to gain about the world?
- Which areas are consistently more hazardous to live in?
 - How bad does the quality of air need to be to be hazardous? For how long
 - can humans survive in each air quality?
 - What would explain the spikes of unhealthy air in some locations? (Wildfires, wind currents, pollution, ect.)
 - Year to year, is the air quality becoming worse overall? Not just in certain areas?
 -

Group **Orange**

Tahseen Haque, Philipp Sommer, Ethan Shen

1.

- a. Ethan shen: emission reduction
- b. Tahseen Haque: Flare Data
- c. Philipp: Air pollution data

Group's decision: Flare data

- Flare data break downs

2. We decided on flare data because we don't know much about the topic and it sounds interesting. Tahseen proposed it and we as a group agreed. This data seemed to be collected through direct vent gas flow measurements in refineries. Chevron is the main company that emits these pollutants.

3. Q: What are some compelling questions to investigate using this data? How would you approach exploring it? What insights are you hoping to gain about the world?

- How does pollution vary between refineries?
- Are there seasonal patterns (global)
- Is there a trend to increasing/decreasing pollution over the years?
- What are the main pollutants in refineries?

Exploration approach:

1. Extracting the data, either manually or by downloading it
2. Optimizing/cleaning the data in
3. Visualizing the results
4. Building hypotheses/interpretations based on the results

Insights:

- Are there any interesting patterns that you could apply on a broader scale

Group: 3

Anne Zhang, Alp Senay, Jesse Cruz

1.

- d.** Person A: Air quality measurement
 - e.** Person B: how it was collected/ who is collecting the data
 - f.** Person C: Open Burn Maps
- Group's decision:
 - Open burn maps
- 7. We decided on emission breakdowns because fire control is important in california.
This data seemed to be collected through
- 8.**

Group 4: Annie Pineda, Junyi Qian

- a. Annie Pineda: Emission Reduction Credit Bank Status
 - b. Junyi Qian: Air Quality Forecast Map
- Group's decision:
 - Flare Data
- 9. We decided on the Flare data because there's a lot of data that comes in and it's interesting to look at the differences for each month. For example, there's daily data for different emission types. It's also something that we don't know much about. The data is visualized monthly by emission type and company. This data seemed to be collected through samples that came from direct measurements during the release of emissions.
- 10. There is still a lot of missing data here, like for Chevron Richmond, the data in D&R remained zero for several years. And different refineries report different categories of data.

Group Dylan

In groups, answer the following questions. You will be given 25 minutes to work, and present afterwards!

1. What [Air District data](#) interests you for further exploration? There's plenty in the Public Data Center. Discuss your interests, then select one data type for the group.

We would like to further explore data from the "Who we regulate" dataset as we want to further understand which companies are producing the most pollution in our area and how they are regulated while still allowing for proficient operation. We would also like to see if the regulations are centered on plants, companies, or a mix of both. The visualizations such as the air quality index map and we're really interested in seeing the extreme values that have occurred since the collection of this data has taken place, like when was the air quality at its all time worse and what caused that. We'd also like to see a map of where the companies being regulated are to see if we can draw any conclusions on where a more industrial area is.

2. How did your group select this data? Describe your group's decision-making process. Given a choice of empirical statistics and qualitative information, we opted for qualitative information to better understand who they chose to regulate. Our group

also was interested in public policy, thus the agency's choice on who to regulate was a public policy decision we wished to study further.

3. What are some compelling questions to investigate using this data? How would you approach exploring it? What insights are you hoping to gain about the world?
 - When issuing public notices, how can we make these notices more accessible on a community level and make communities more aware of projects implemented by the Air District?
 - Why do we regulate certain companies, but not others?
 - What is the metric used to regulate companies, and what metrics do we use?
 - What are the metrics we use when analyzing air quality? For instance, for traffic accidents data, some countries may calculate incidence of injuries per 100000 commuters, others may calculate death per length of road traveled? How do different countries measure "air quality"? Is there an international standard for better global comparison?

Group

Group Huanbo Meng

What [Air District data](#) interests you for further exploration? There's plenty in the Public Data Center. Discuss your interests, then select one data type for the group.

I personally lean towards the topic of data statistics on particle sizes in the air, such as the commonly mentioned PM2.5 and PM5.0. I believe relevant data can be found on our project database website.

How did your group select this data? Describe your group's decision-making process.

First of all, this data has been accumulated over a long period and is not a sudden or rare data type. I believe the feedback regarding its reliability is relatively clear.

What are some compelling questions to investigate using this data? How would you approach exploring it? What insights are you hoping to gain about the world?

The most compelling aspect is certainly the issues related to respiratory health. People hope to use this reliable data to identify potential problems, such as analyzing dust particle sizes by region to assess the air safety of a community, formulating air optimization policies for specific areas, or taking targeted measures.