High Level Design (HLD)

US Pollution Analysis

Revision Number: 1.0
Last date of revision: 14/05/2025

Shubham Tembhurne

Document Version Control

Date Issued	Version	Description	Author
14th May 2025	1.0	First Version of Complete HLD	Shubham Tembhurne

Contents

Abstract

This study analyzes air pollution trends in the United States from 2006 to 2010, focusing on key pollutants including Carbon Monoxide (CO), Ozone (O_3), Nitrogen Dioxide (NO_2), and Sulfur Dioxide (SO_2). The data reveals a general decline in the average and maximum concentrations of CO, NO_2, and SO_2 over the five-year period, suggesting notable improvements in air quality. CO and SO_2 levels peaked in 2006, while NO_2 reached its highest level in 2007. In contrast, O_3 exhibited a relatively constant average trend, with its peak concentration occurring in 2008. The Air Quality Index (AQI) values followed similar patterns, with the highest values recorded between 2006 and 2007, followed by a gradual decline. These trends reflect the effectiveness of air pollution control measures and regulatory efforts implemented during this period.

Overall, the analysis indicates a positive shift toward cleaner air and improved public health outcomes in the United States.

1 . Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2 . General Description

2.1 Product Perspective & Problem Statement

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

2.2 Tools used

Business Intelligence tools and libraries works such as NumPy, Pandas, Excel, Tableau, are used to build the whole framework.

3 . Design Details

3.1 Functional Architecture

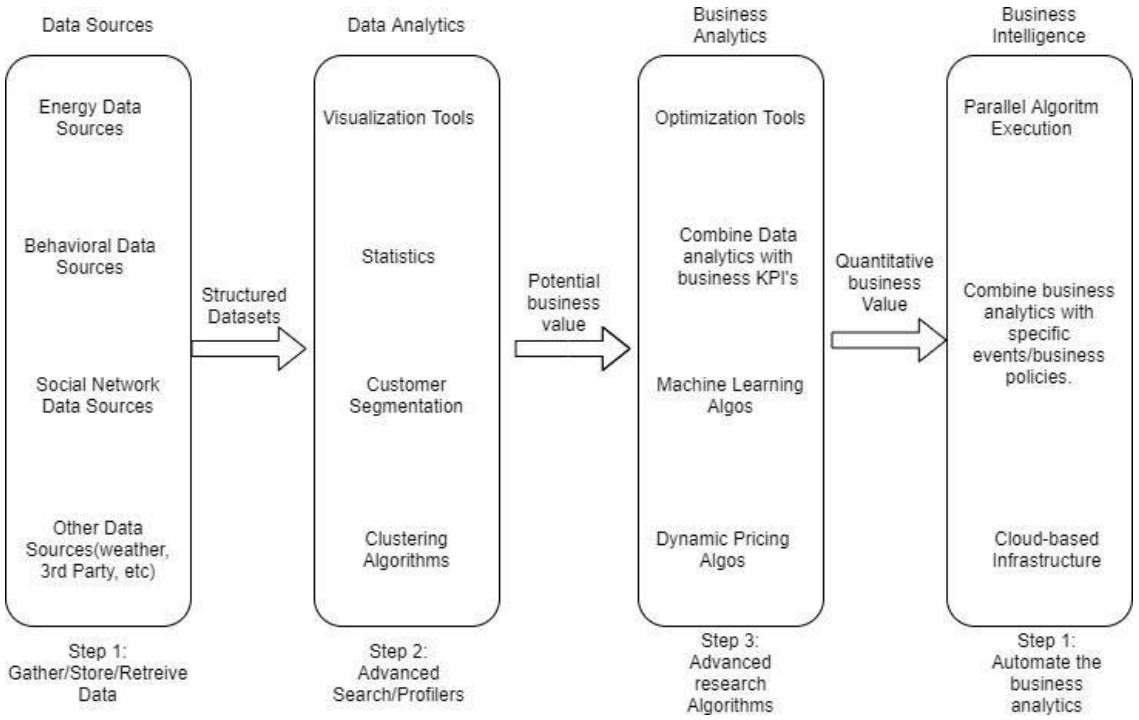
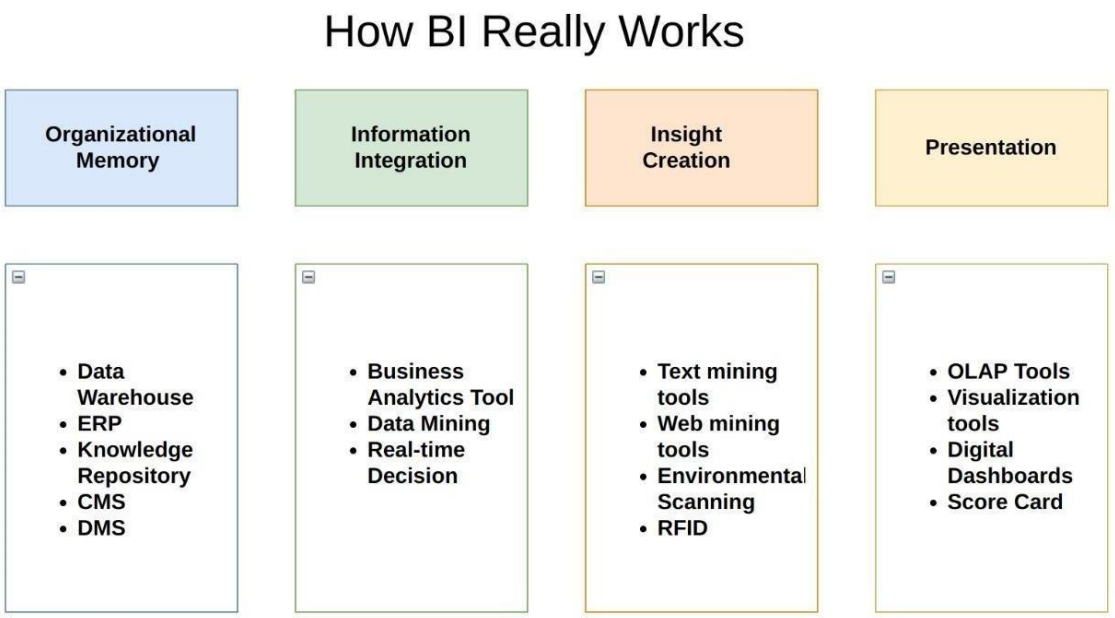


Figure 1: Functional Architecture of Business Intelligence



3.2 Optimization

Your data strategy drives performance

- Minimize the number of fields
- Minimize the number of records
- Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views

Reduce the marks (data points) in your view

- Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
- Remove unneeded dimensions from the detail shelf.
- Explore. Try displaying your data in different types of views.

Limit your filters by number and type

- Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.
- Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.
- Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete date filters.
- Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.
- Use parameters and action filters. These reduce the query load (and work across data sources).

Optimize and materialize your calculations

- Perform calculations in the database
- Reduce the number of nested calculations.
- Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.
    - LODs - Look at the number of unique dimension members in the calculation.
    - Table Calculations - the more marks in the view, the longer it will take to calculate.
- Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.

- Make groups with calculations. Like include filters, calculated groups load only `named members of the domain, whereas Tableau's group function loads the entire domain.
- Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings.
  Boolean>Int>Float>Date>Date Time>String

4 . KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the sales.

As and when, the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors

4.1 KPIs (Key Performance Indicators)

1. Average Pollutant by Month
2. Percentage distribution of AQI Categories.
3. Yearly Statistics for CO, NO2, SO2 and O3.
4. Highest Cities for CO, NO2, SO2, O3 Pollutants
5. Maximum CO, NO2, SO2 and O3 AQI by Year (2006-2010)
6. Maximum CO, NO2, SO2 and O3 AQI in each state
7. Highest States for CO, NO2, SO2 and O3 Pollutant

.

5 . Deployment

Prioritizing data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today's most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

Tableau prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Tableau Server and Tableau Online leverage your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Tableau to match your requirements. Below is a comparison of the three types:

TYPE PROS CONS

Tableau Server - On Premises

- Full control of hardware and software
- Infrastructure and data remain behind your firewall
- Need dedicated administrators to manage hardware and software
- Additional infrastructure needed to access off-network (mobile, external)

Tableau Server - Public Cloud (IaaS)

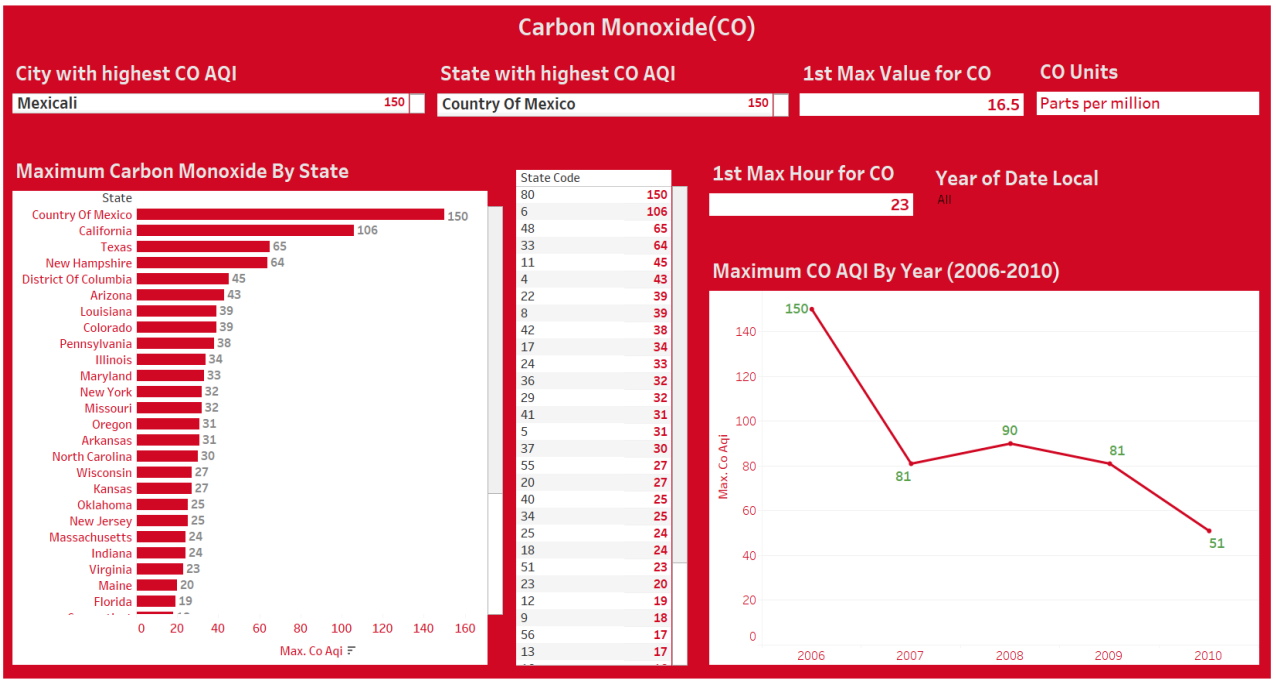
- Full control of software on managed hardware
- Puts infrastructure in same place as data (for migration to cloud)
- Flexibility to spin up/down hardware as needed
- Need dedicated administrators to manage software
- Additional infrastructure needed to access off-network (mobile, external)

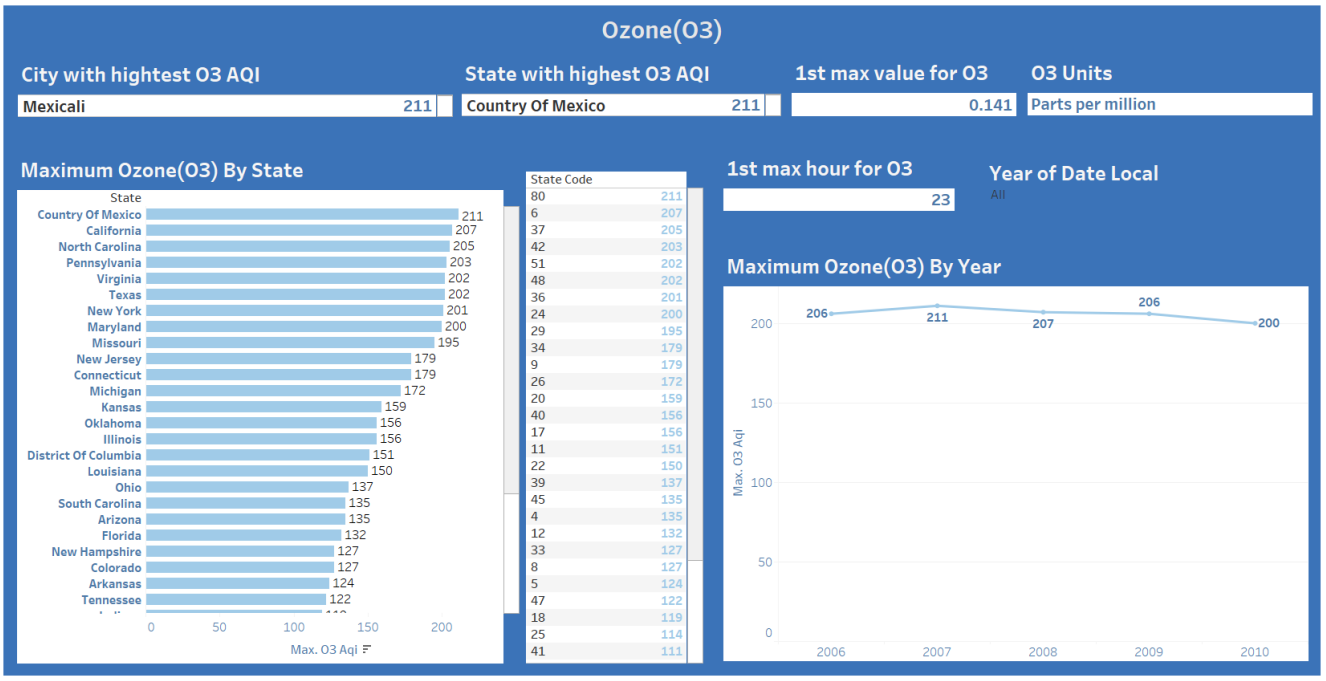
Tableau Online (SaaS)

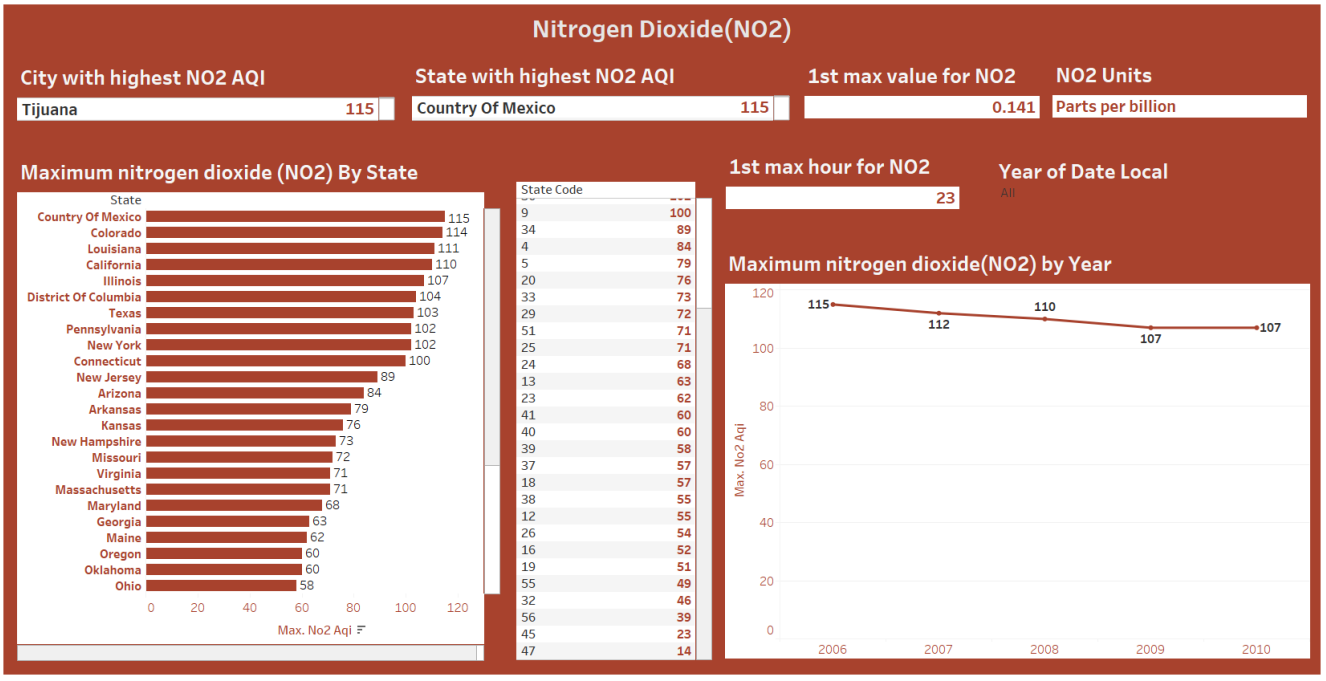
- Fully hosted solution (hardware, software upgrades)
- Fast to deploy
- Easy for external audience to access
- Single-site in multi-tenant environment
- Cubes are not supported
- No guest account access

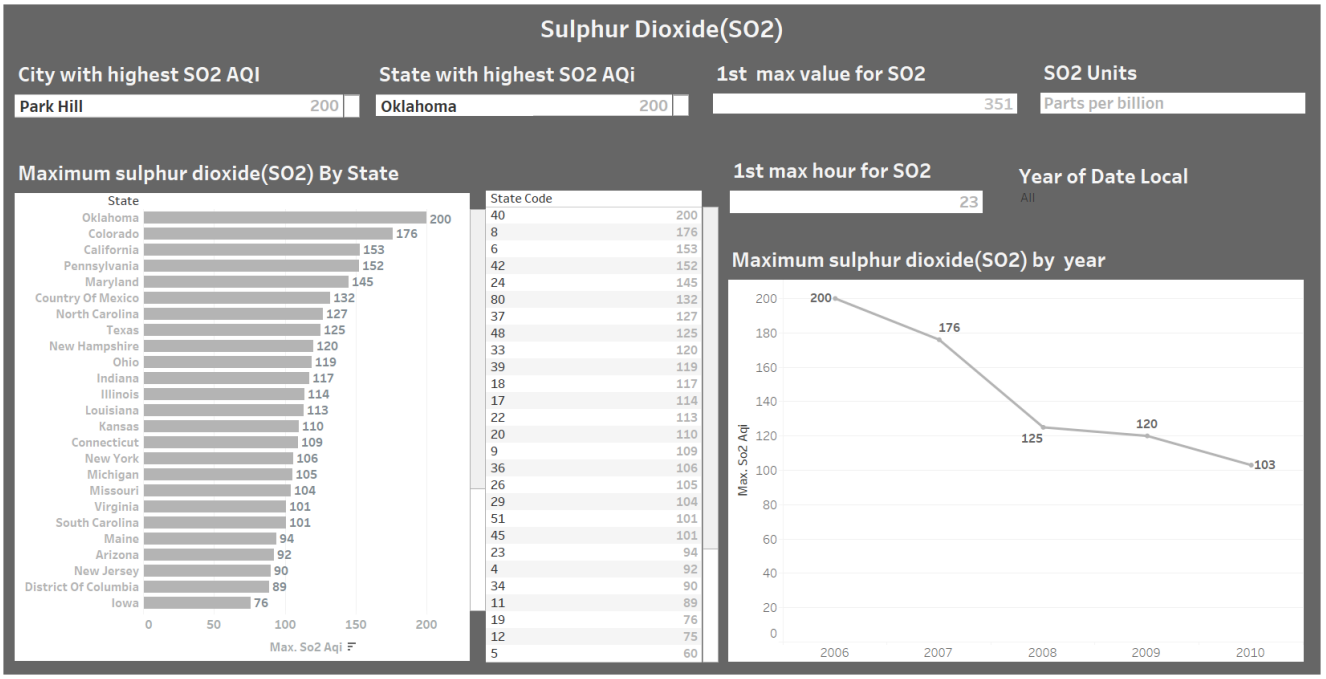
High Level Design (HLD)

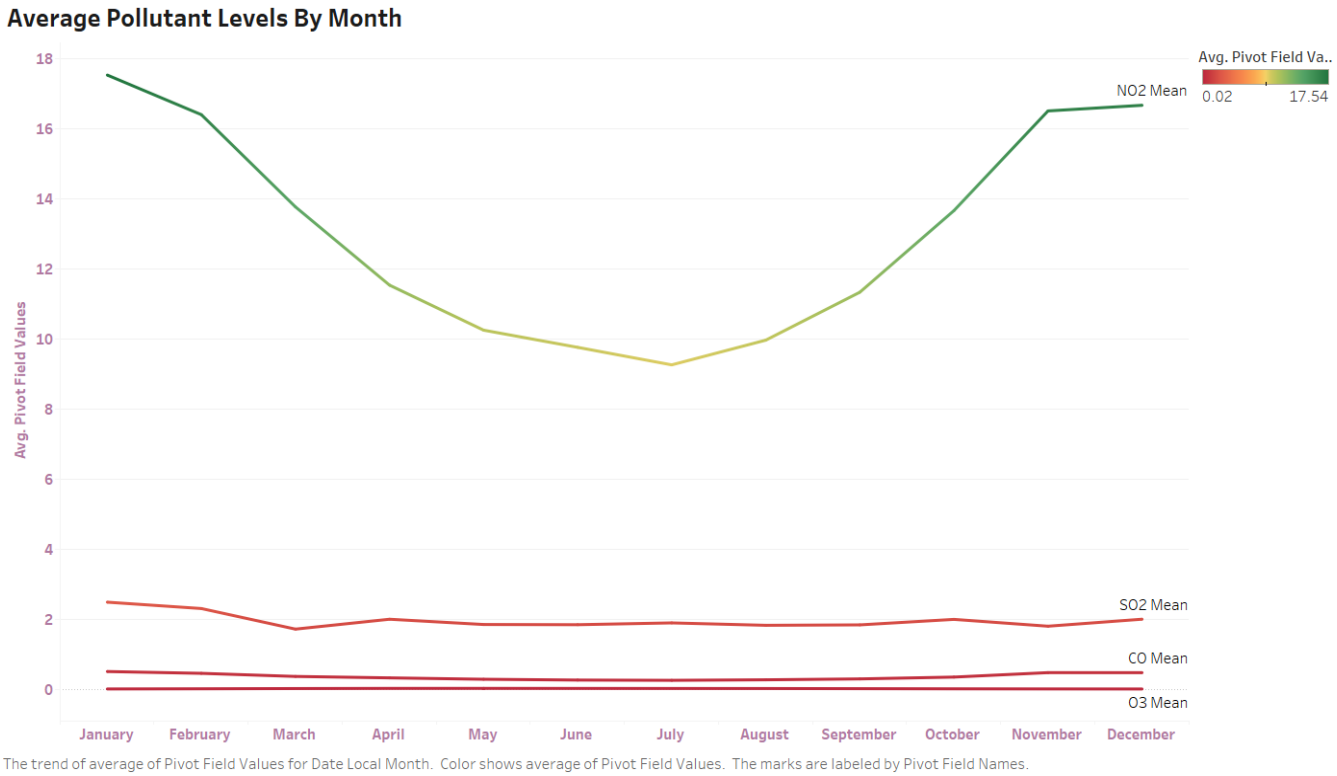
Depending on your organizational roles and responsibilities, Tableau Server should be installed by a systems administrator and the designated Tableau Server Administrator in coordination with the appropriate IT roles. For Tableau Online, you will integrate with your existing technology and configure the site settings. The Data & Analytics Survey, completed by business teams, identifies and prioritizes data use cases, audience size, and users. You will use the information collected in both surveys to plan your deployment strategy, including sizing, installation, and configuration of your Tableau Server or integration and configuration of Tableau Online. In addition to installing Tableau Server or configuring Tableau Online, administrators will also need to plan for the client software installation of Tableau Prep Builder, Tableau Desktop, Tableau Mobile, and Tableau Bridge for Tableau Online where applicable.




Ozone(O3)
City with hightest O3 AQI
Mexicali 211
State with highest O3 AQI
Country Of Mexico 211
1st max value for O3
0.141
O3 Units
Parts per million
Maximum Ozone(O3) By State
State
Country Of Mexico 211
California 207
North Carolina 205
Pennsylvania 203
Virginia 202
Texas 202
New York 201
Maryland 200
Missouri 195
New Jersey 179
Connecticut 179
Michigan 172
Kansas 159
Oklahoma 156
Illinois 156
District Of Columbia 151
Louisiana 150
Ohio 137
South Carolina 135
Arizona 135
Florida 132
New Hampshire 127
Colorado 127
Arkansas 124
Tennessee 122
0 50 100 150 200
Max. O3 Aqi
State Code
80 211
6 207
37 205
42 203
51 202
48 202
36 201
24 200
29 195
34 179
9 179
26 172
20 159
40 156
17 156
11 151
22 150
39 137
45 135
4 135
12 132
33 127
8 127
5 124
47 122
18 119
25 114
41 111
1st max hour for O3
23
Year of Date Local
All
Maximum Ozone(O3) By Year
206 211 207 206 200
200 150 100 50 0
Max. O3 Aqi
2006 2007 2008 2009 2010


Nitrogen Dioxide(NO2)
City with highest NO2 AQI
Tijuana 115
State with highest NO2 AQI
Country Of Mexico 115
1st max value for NO2
0.141
NO2 Units
Parts per billion
Maximum nitrogen dioxide (NO2) By State
State
Country Of Mexico 115
Colorado 114
Louisiana 111
California 110
Illinois 107
District Of Columbia 104
Texas 103
Pennsylvania 102
New York 102
Connecticut 100
New Jersey 89
Arizona 84
Arkansas 79
Kansas 76
New Hampshire 73
Missouri 72
Virginia 71
Massachusetts 71
Maryland 68
Georgia 63
Maine 62
Oregon 60
Oklahoma 60
Ohio 58
0 20 40 60 80 100 120
Max. No2 Aqi
State Code
9 100
34 89
4 84
5 79
20 76
33 73
29 72
51 71
25 71
24 68
13 63
23 62
41 60
40 60
39 58
37 57
18 57
38 55
12 55
26 54
16 52
19 51
55 49
32 46
56 39
45 23
47 14
1st max hour for NO2
23
Year of Date Local
All
Maximum nitrogen dioxide(NO2) by Year
115 112 110 107 107
120 100 80 60 40 20 0
Max. No2 Aqi
2006 2007 2008 2009 2010


Sulphur Dioxide(SO2)
City with highest SO2 AQI
Park Hill 200
State with highest SO2 AQi
Oklahoma 200
1st max value for SO2
351
SO2 Units
Parts per billion
Maximum sulphur dioxide(SO2) By State
State
Oklahoma 200
Colorado 176
California 153
Pennsylvania 152
Maryland 145
Country Of Mexico 132
North Carolina 127
Texas 125
New Hampshire 120
Ohio 119
Indiana 117
Illinois 114
Louisiana 113
Kansas 110
Connecticut 109
New York 106
Michigan 105
Missouri 104
Virginia 101
South Carolina 101
Maine 94
Arizona 92
New Jersey 90
District Of Columbia 89
Iowa 76
0 50 100 150 200
Max. So2 Aqi
State Code
40 200
8 176
6 153
42 152
24 145
80 132
37 127
48 125
33 120
39 119
18 117
17 114
22 113
20 110
9 109
36 106
26 105
29 104
51 101
45 101
23 94
4 92
34 90
11 89
19 76
12 75
5 60
1st max hour for SO2
23
Year of Date Local
All
Maximum sulphur dioxide(SO2) by year
Max. So2 Aqi
200 180 160 140 120 100 80 60 40 20 0
200
176
125
120
103
2006 2007 2008 2009 2010


Average Pollutant Levels By Month
Avg. Pivot Field Va..
0.02
17.54
Avg. Pivot Field Values
18 16 14 12 10 8 6 4 2 0
NO2 Mean
SO2 Mean
CO Mean
O3 Mean
January February March April May June July August September October November December
The trend of average of Pivot Field Values for Date Local Month. Color shows average of Pivot Field Values. The marks are labeled by Pivot Field Names.