

Low Level Design

Us Pollution Analysis



Written By	Shubham Tembhurne
Document Version	0.3
Last Revised Date	14/05/25

DOCUMENT CONTROL

Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	14- May - 2025	Shubham temburne	Introduction and architecture defined
0.2	14 - May - 2025	Shubham tembhurne	Architecture & Architecture description appended and updated.

Contents

1.	Introduction.....	04
1.1	What is Low-Level Design Document?	04
1.2	Scope	04
1.3	Project Introduction.....	04
2.	Problem Statement.....	05
3.	Dataset Information.....	05
4.	Architecture	07
4.1	Architecture Description.....	09

1. Introduction

1.1 What is Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Us Pollution Analysis dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

1.3 Project Introduction

The goal of this project is to analyze the air pollution is the contamination of air due to the presence of substances in the atmosphere that are harmful to the health of humans and other living beings, or cause damage to the climate or to materials.

2. Problem Statement

Air pollution is the contamination of air due to the presence of substances in the atmosphere that are harmful to the health of humans and other living beings, or cause damage to the climate or to materials. This dataset provides you the pollution in the U.S. It contains daily data for the four major pollutants NO₂, O₃, SO₂ and CO each has 5 specific columns during 2006 and 2010

2. Dataset Information

State names: Lists of state names in US

State code: Code numbers belong to each state

County code: Unique code numbers specified for county

Site number: numbers given to the sites

Address: The Address of each locality where pollutant has been detected

Country names: List of county names

City names: Names of city

Date local: Day Month and the year of every local area when pollutant took place and the hierarchy

NO₂ units: Units represented by Parts Per Billion, NO₂ mean, NO₂ 1st MAX value, NO₂ 1st MAX hour, NO₂ AQI

O₃ units: Units represented by parts per million, O₃ mean, O₃ 1st MAX value, O₃ 1st MAX hour, O₃ AQI

SO₂ units: Units represented by Parts Per Billion, SO₂ mean, SO₂ 1st MAX value, SO₂ 1st MAX hour, SO₂ AQI

CO units: Units represented by parts per million, CO₂ mean, CO₂ 1st MAX value, CO₂ 1st MAX hour, CO₂ AQI

1. Architecture

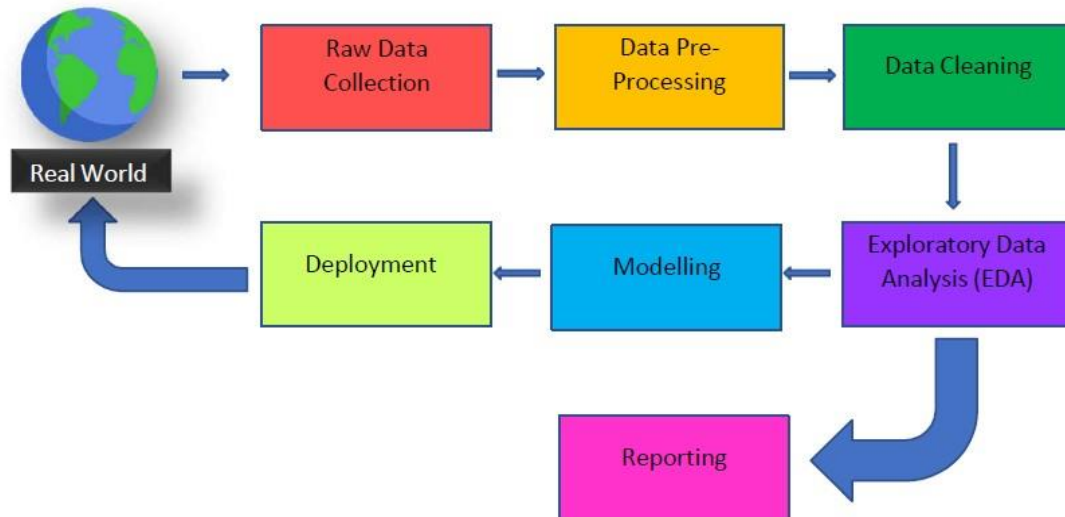


Tableau Server Architecture

Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software. Tableau Server architecture supports fast and flexible deployments.

The following diagram shows Tableau Server's architecture:

Tableau Communication Flow

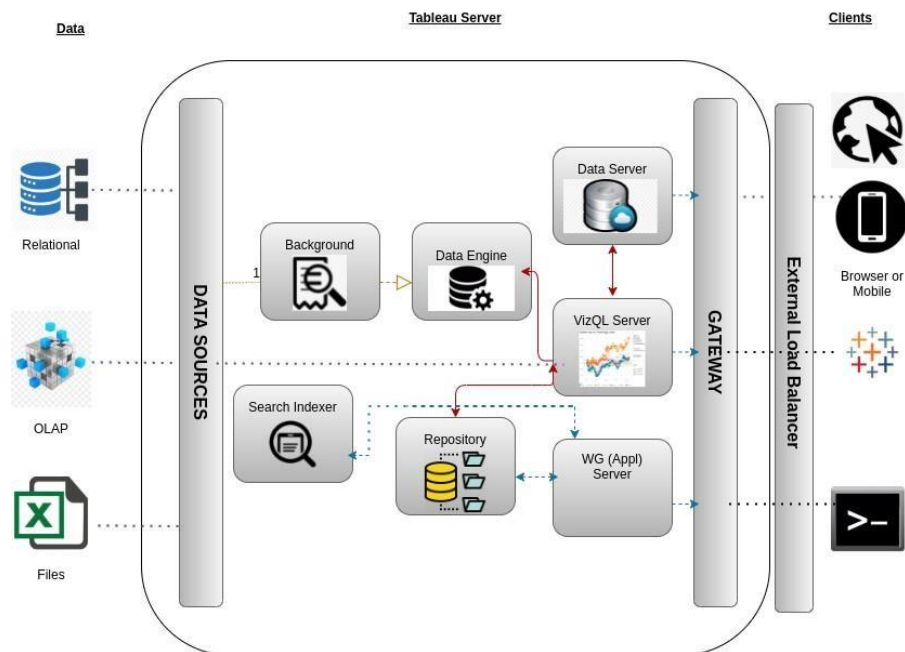


Tableau Server is internally managed by the multiple server processes.

1). Gateway/Load Balancer

It acts as an Entry gate to the Tableau Server and also balances the load to the Server if multiple Processes are configured.

2) Application Server: -

Application Server processes (wgserver.exe) handle browsing and permissions for the Tableau Server web and mobile interfaces. When a user opens a view in a client device, that user starts a session on Tableau Server. This means that an Application Server thread starts and checks the permissions for that user and that view.

3) Repository: -

Tableau Server Repository is a PostgreSQL database that stores server data. This data includes information about Tableau Server users, groups and group assignments, permissions, projects, data sources, and extract metadata and refresh information.

4) VIZQL Server: -

Once a view is opened, the client sends a request to the VizQL process (vizqlserver.exe). The VizQL process then sends queries directly to the data source, returning a result set that is rendered as images and presented to the user. Each VizQL Server has its own cache that can be shared across multiple users

5) Data Engine: -

It Stores data extracts and answers queries.

6) Backgrounder: -

The backgrounder Executes server tasks which includes refreshes scheduled extracts, tasks initiated from tabcmd and manages other background tasks.

7) Data Server: -

Data Server Manages connections to Tableau Server data sources

It also maintains metadata from Tableau Desktop, such as calculations, definitions, and groups.

3. Architecture Description

1. Raw Data Collection

The Dataset was taken from Kaggle Provided.

2. Data Pre-Processing

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data fed to the model to train.

This Process includes)

Handling Null/Missing Values

b) Handling Skewed Data

c) Outliers Detection and Removal

3. Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

a) Remove duplicate or irrelevant observations

b) Filter unwanted outliers

c) Renaming required attributes

4. Reporting

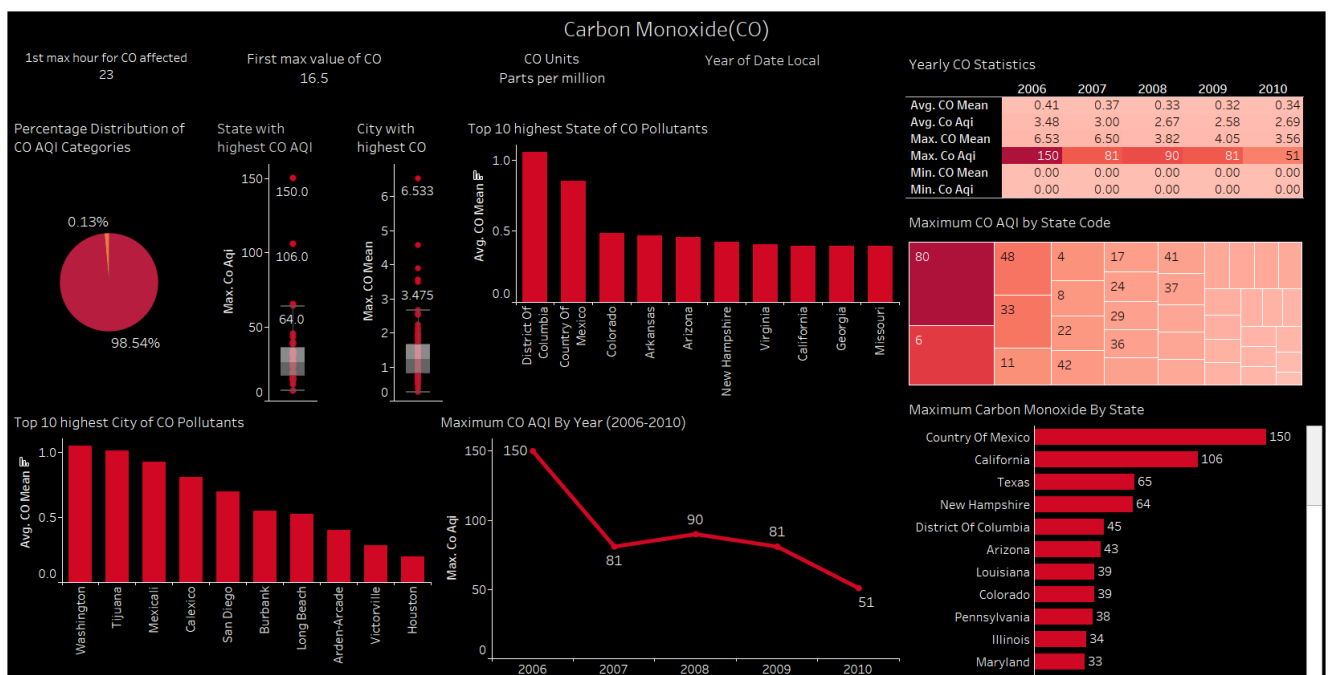
Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in the easy and self-explanatory report because your model will be used by many stakeholders who are not from a technical background. a) High-Level Design Document (HLD) b) Low-Level Design Document (LLD) c) Architecture d) Wireframe e) Detailed Project Report f) PowerPoint Presentation

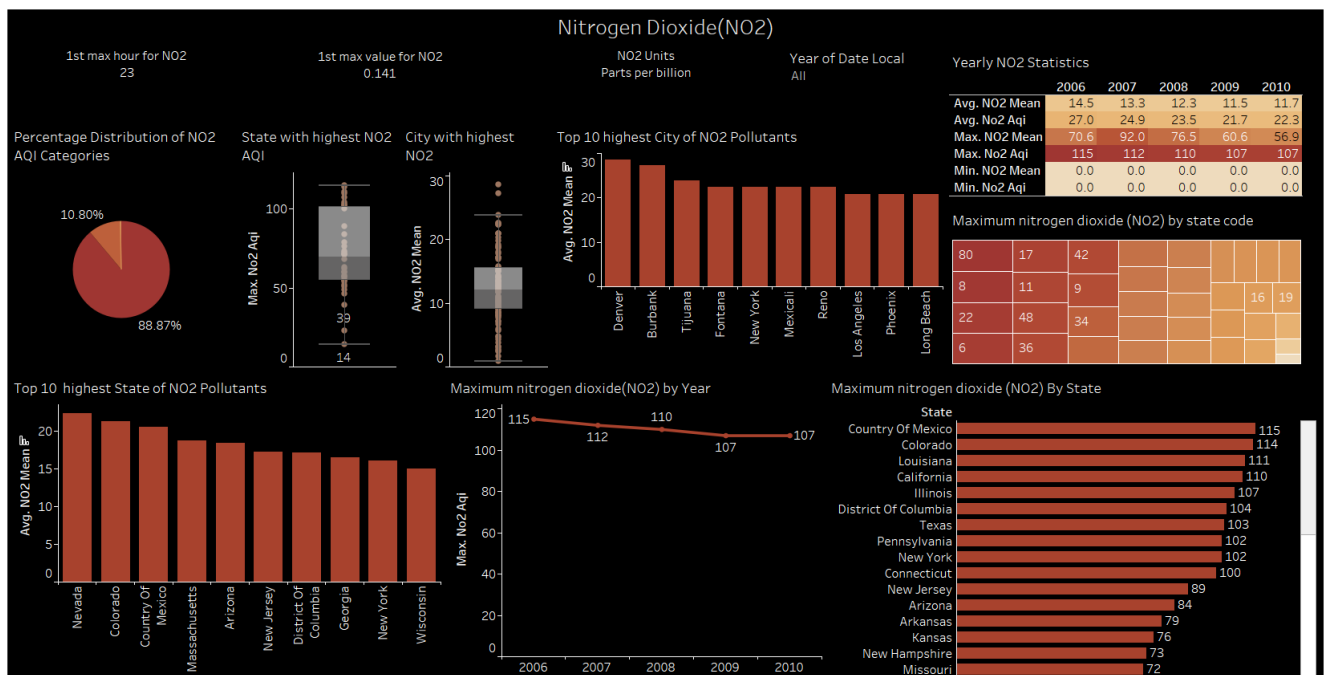
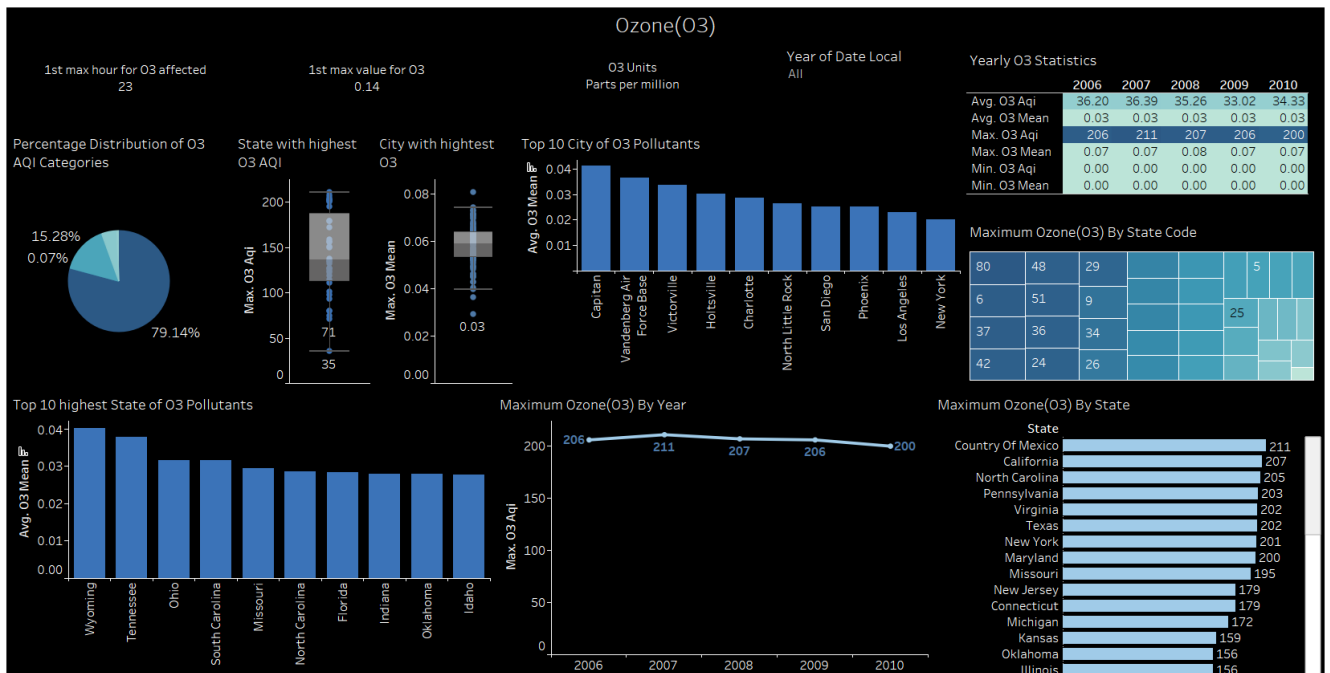
5. Modelling Data

Modelling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

6. Deployment

We created a Tableau Dashboard





Sulphur Dioxide(SO2)

1st max hour for SO2
23

1st max value for SO2
351

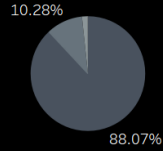
SO2 Units
Parts per billion

Year of Date Local
All

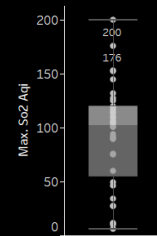
Yearly SO2 Statistics

	2006	2007	2008	2009	2010
Avg. SO2 Mean	2.6	2.3	1.9	1.6	1.2
Avg. So2 Aqi	5.0	4.5	3.7	2.8	1.9
Max. SO2 Mean	321.6	32.0	31.0	33.0	25.9
Max. So2 Aqi	200	176	125	120	103
Min. SO2 Mean	0.0	0.0	0.0	0.0	-0.5
Min. So2 Aqi	0.0	0.0	0.0	0.0	0.0

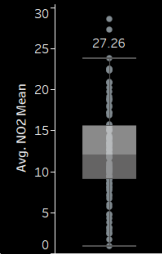
Percentage Distribution of SO2 Aqi
Categories



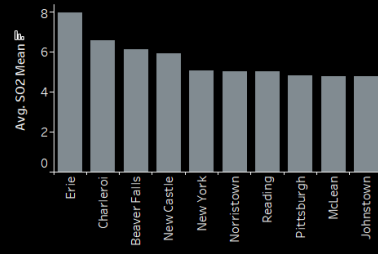
State with highest
SO2 Aqi



City with highest SO2



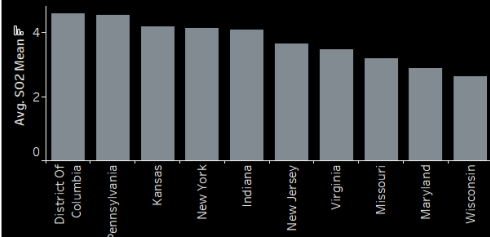
Top 10 highest City of SO2 Pollutants



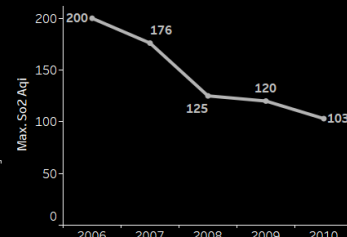
Maximum sulphur dioxide (SO2) by state code

40	24	33	22	29	45	51	23
8	80	39	20				
6	37	18	9	4			5
42	48	17	36	34			
			26	11			

Top 10 highest State of SO2 Pollutants



Maximum sulphur dioxide(SO2) by year



Maximum sulphur dioxide(SO2) By State

