Data Analysis

Group ID: G04

Team Members: - Nachiket Khamar (21162102004)

- Devarsh Mistry (21162122002)

- Souhardya Sarkar (20162171029)

Guide/Mentor Name: Anoj Dixit

Institute of Computer Technology, Ganpat University

Problem Statement (Scope/Goals)

Problem Statement

"Efficiently Analyzing and Securing Diverse Data Sources.

Scope

- Facebook Ads
- Security Log Data
- General Data / Open Sources Data

Goals

- Segregation of Data
- Decision Making using Machine Learning with the data

Background

- With ever growing data, managing and analysis is becoming a tougher job.
- Motivation: Utilization of AI in various aspects of our life led us to believe that data analysis with most accurate data segregation and visuals are required.
- At ground level we want to just categorize and visualize data. Add-on features in future can include data prediction etc.

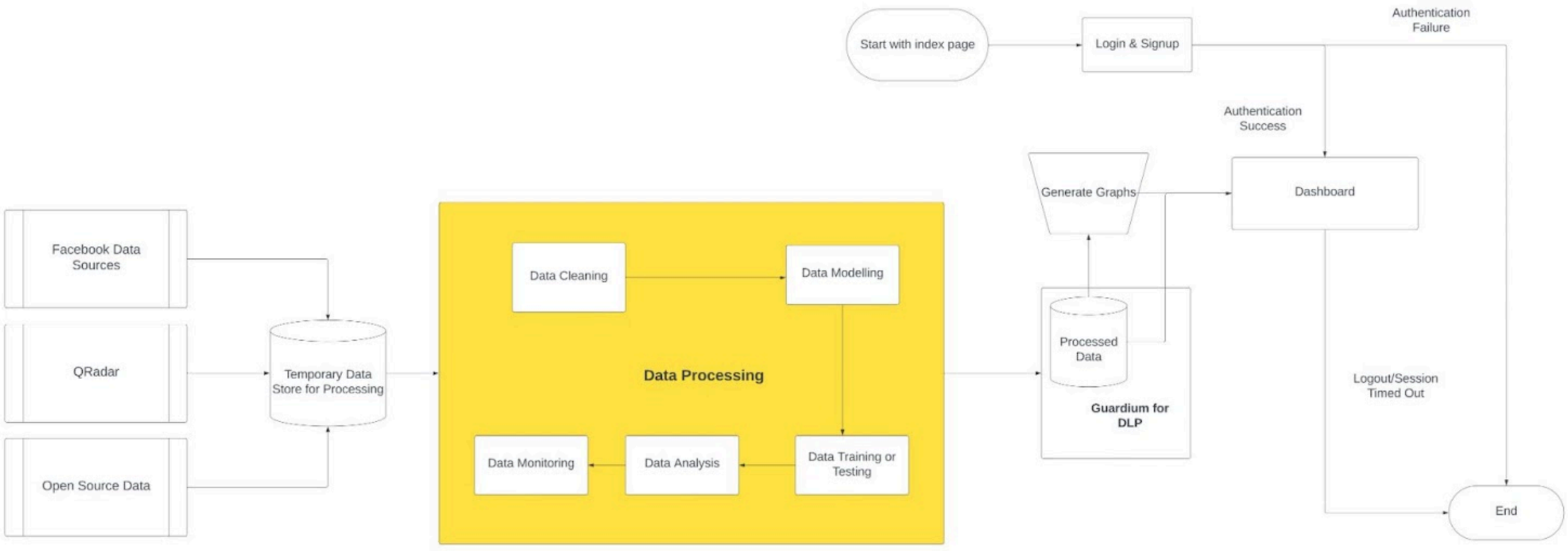
Feasibility Study

- Resources: Cloud Availability for deployment, Datasets for Model training
- Technology Stack: Python (Flask Framework)
- Data Security and Compliance: Data of users will be protected using DLP Protection softwares and QRadar for security and compliances
- Stakeholder buy-in: Constant support from Stakeholder and mentors

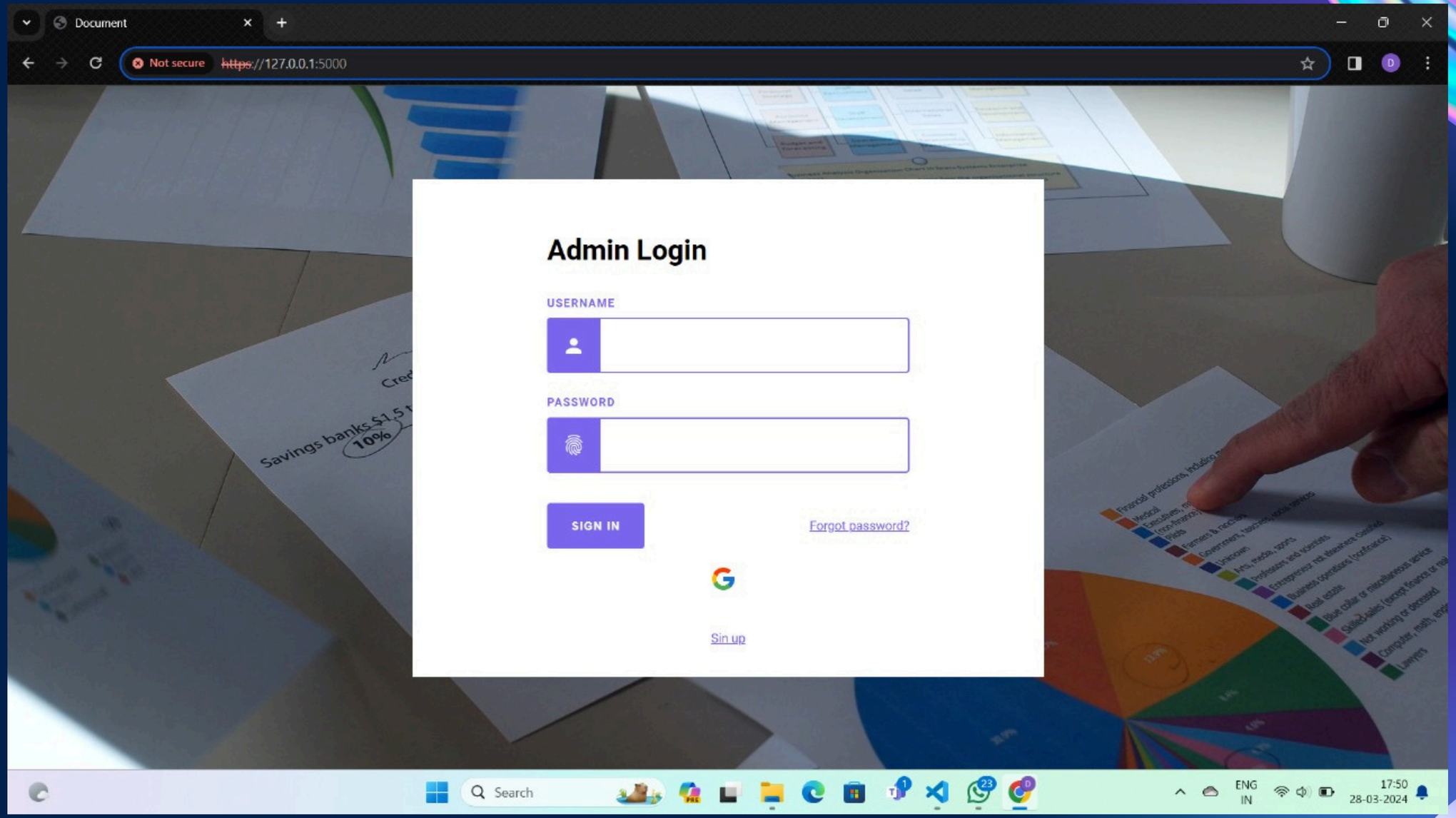
Data source

- Kaggle
- Google Analytics
- Github
- CodeLab

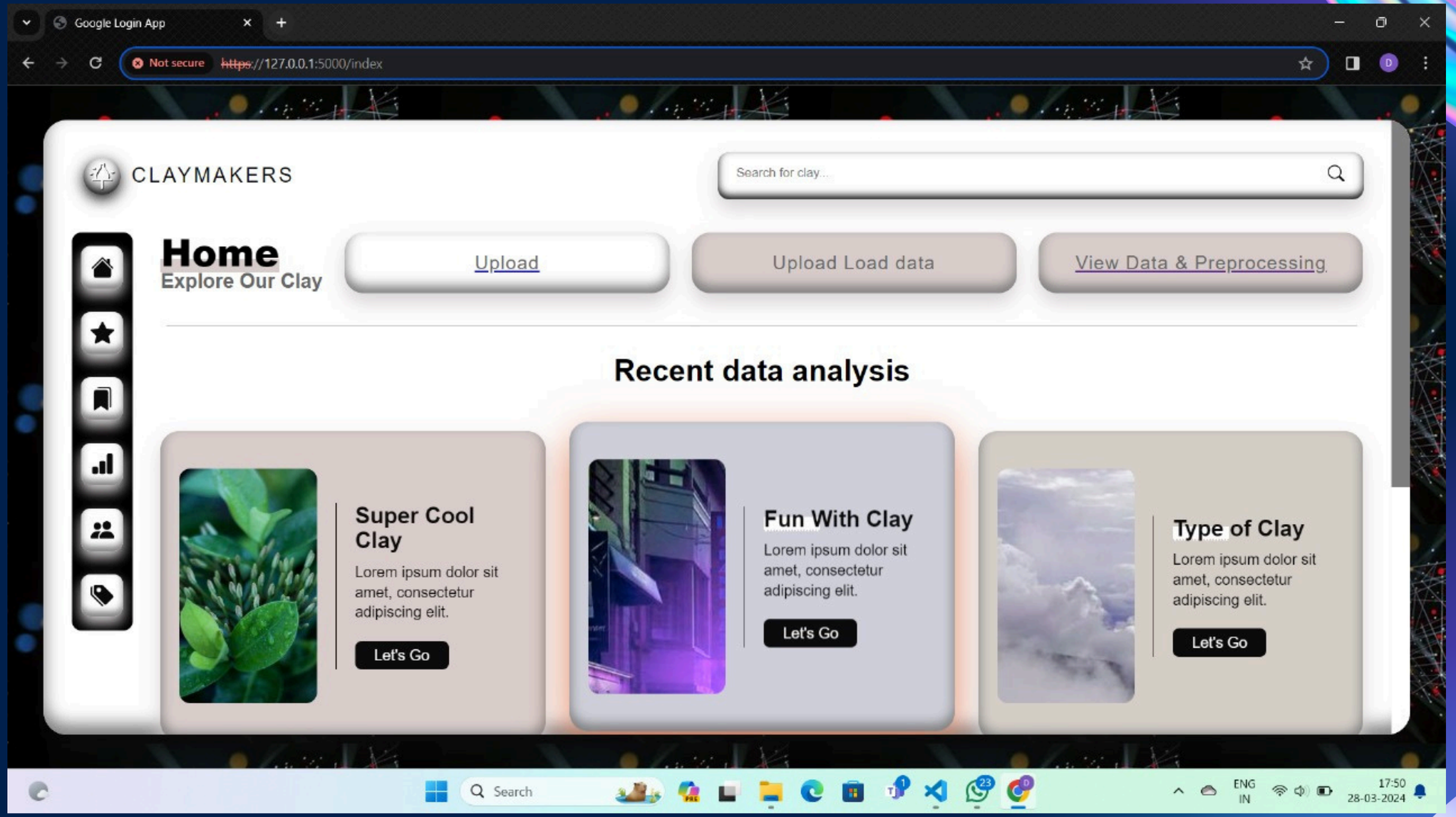
Project Flow



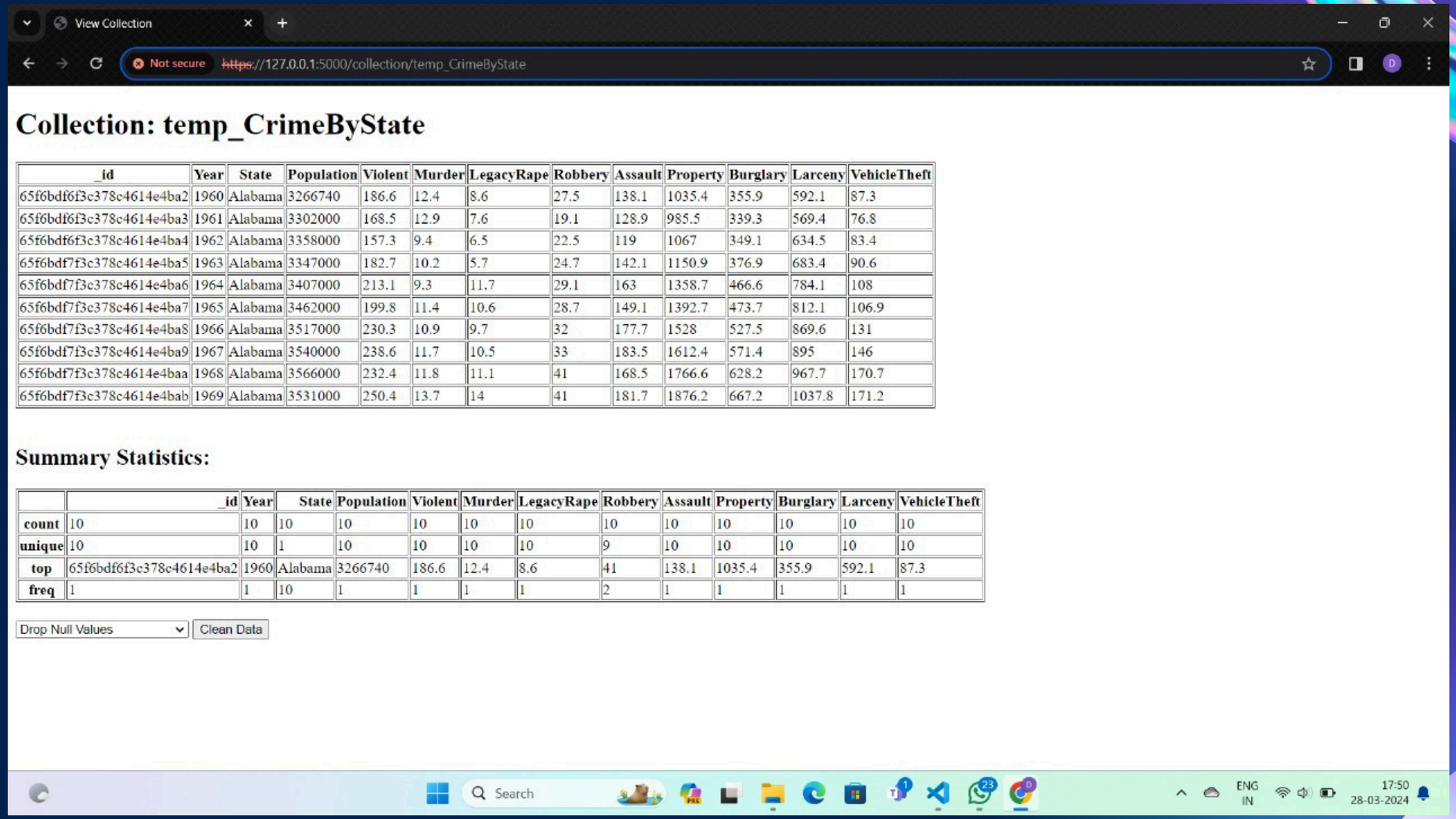
UI


Document
Not secure
https://127.0.0.1:5000
Admin Login
USERNAME
PASSWORD
SIGN IN
Forgot password?
Sin up
Search
ENG
IN
17:50
28-03-2024

UI


Google Login App
Not secure
https://127.0.0.1:5000/index
CLAYMAKERS
Search for clay...
Home
Explore Our Clay
Upload
Upload Load data
View Data & Preprocessing
Recent data analysis
Super Cool Clay
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Let's Go
Fun With Clay
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Let's Go
Type of Clay
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Let's Go
Search
ENG
IN
17:50
28-03-2024

UI



Collection: temp_CrimeByState

_id	Year	State	Population	Violent	Murder	LegacyRape	Robbery	Assault	Property	Burglary	Larceny	VehicleTheft
65f6bdf6f3c378c4614e4ba2	1960	Alabama	3266740	186.6	12.4	8.6	27.5	138.1	1035.4	355.9	592.1	87.3
65f6bdf6f3c378c4614e4ba3	1961	Alabama	3302000	168.5	12.9	7.6	19.1	128.9	985.5	339.3	569.4	76.8
65f6bdf6f3c378c4614e4ba4	1962	Alabama	3358000	157.3	9.4	6.5	22.5	119	1067	349.1	634.5	83.4
65f6bdf7f3c378c4614e4ba5	1963	Alabama	3347000	182.7	10.2	5.7	24.7	142.1	1150.9	376.9	683.4	90.6
65f6bdf7f3c378c4614e4ba6	1964	Alabama	3407000	213.1	9.3	11.7	29.1	163	1358.7	466.6	784.1	108
65f6bdf7f3c378c4614e4ba7	1965	Alabama	3462000	199.8	11.4	10.6	28.7	149.1	1392.7	473.7	812.1	106.9
65f6bdf7f3c378c4614e4ba8	1966	Alabama	3517000	230.3	10.9	9.7	32	177.7	1528	527.5	869.6	131
65f6bdf7f3c378c4614e4ba9	1967	Alabama	3540000	238.6	11.7	10.5	33	183.5	1612.4	571.4	895	146
65f6bdf7f3c378c4614e4baa	1968	Alabama	3566000	232.4	11.8	11.1	41	168.5	1766.6	628.2	967.7	170.7
65f6bdf7f3c378c4614e4bab	1969	Alabama	3531000	250.4	13.7	14	41	181.7	1876.2	667.2	1037.8	171.2

Summary Statistics:

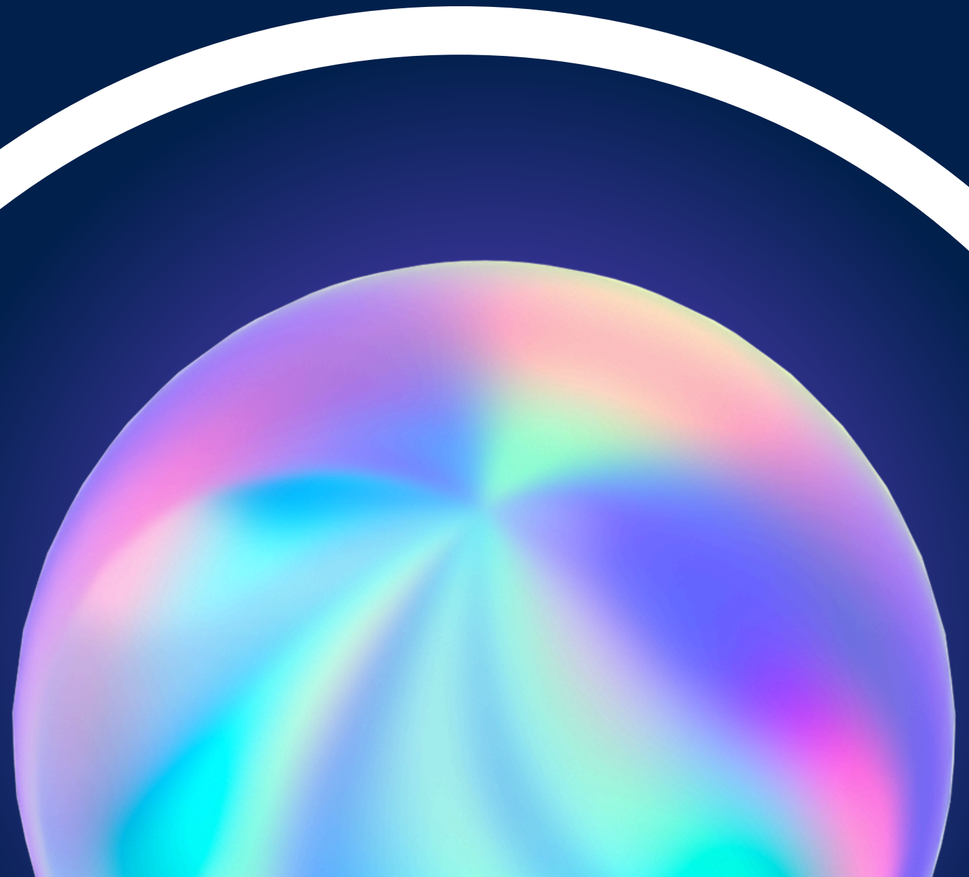
	_id	Year	State	Population	Violent	Murder	LegacyRape	Robbery	Assault	Property	Burglary	Larceny	VehicleTheft
count	10	10	10	10	10	10	10	10	10	10	10	10	10
unique	10	10	1	10	10	10	10	9	10	10	10	10	10
top	65f6bdf6f3c378c4614e4ba2	1960	Alabama	3266740	186.6	12.4	8.6	41	138.1	1035.4	355.9	592.1	87.3
freq	1	1	10	1	1	1	1	2	1	1	1	1	1

Drop Null Values | Clean Data

Tools and Technologies in Use

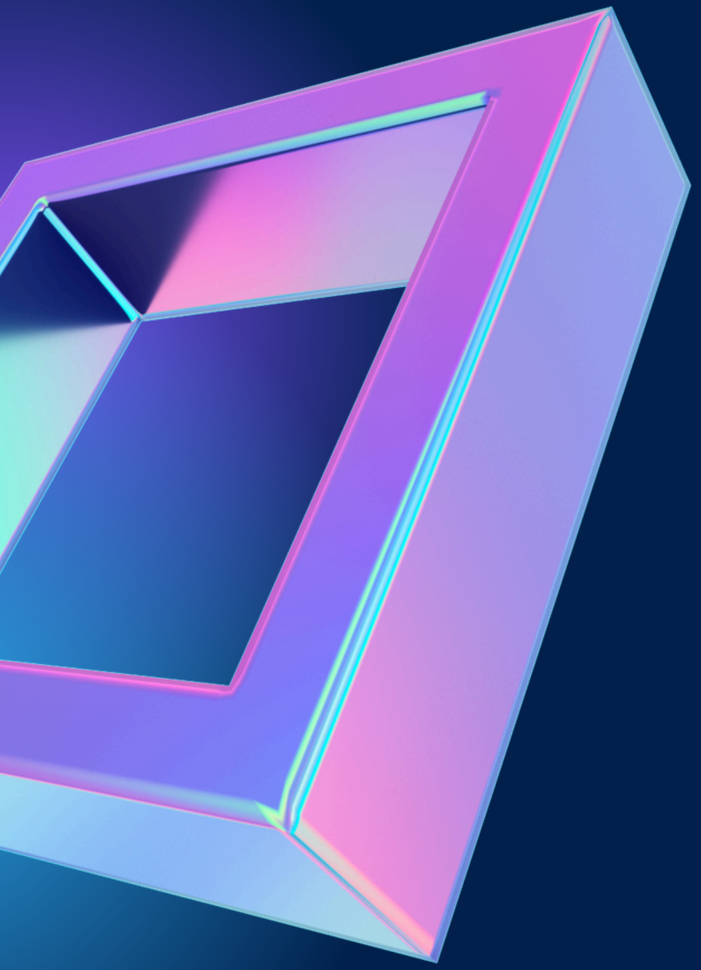
- CI/CD Pipeline
- Integration of Modules and Security Measures
- Custom Algorithms

Approach



Tools

- 2FA, Guardium for DLP, QRadar
- IBM SPSS Modeller
- Python & Flask
- Google Cloud Console
- UI/UX



Thank You