# Employee Data Analysis Project

This project analyzes employee data from CSV files using SQLite and Python. It executes various SQL queries to extract meaningful insights and measures performance based on different file sizes (1 MB, 10 MB, 100 MB).

## Database Schema

The database `employee\_data` table has the following schema:

|  |  |
| --- | --- |
| Column Name | Data Type |
| PersonID | int(11) |
| PersonName | string(100) |
| SchoolID | string(20) |
| SchoolName | string(100) |
| SchoolCampus | string(50) |
| DepartmentName | string(100) |
| DepartmentID | string(20) |
| BirthDate | date |
| StillWorking | bool |
| JobID | string(20) |
| JobTitle | string(100) |
| Earnings | float(15, 2) |
| EarningsYear | int(4) |

## Executed Queries

1. Query 1:   
  
**SELECT PersonName FROM employee\_data WHERE BirthDate < '1975-01-01' AND Earnings > 130000;**

Execution Time:

1 MB File: 0.0788 seconds

10 MB File: 0.0430 seconds

100 MB File: 0.0446 seconds

A graph with a line

Description automatically generated

2. Query 2:  
  
**SELECT PersonName, SchoolName FROM employee\_data WHERE Earnings > 400000 AND StillWorking = 0;**

Execution Time:

1 MB File: 0.0145 seconds

10 MB File: 0.0139 seconds

100 MB File: 0.0139 seconds

A graph with a line going up

Description automatically generated

3. Query 3:   
  
**SELECT PersonName FROM employee\_data WHERE SchoolName = 'University of Texas' AND JobTitle = 'Lecturer' AND StillWorking = 0;**

Execution Time:

1 MB File: 0.0120 seconds

10 MB File: 0.0120 seconds

100 MB File: 0.0131 seconds

A graph with a line

Description automatically generated

4. Query 4:   
  
**SELECT SchoolName, SchoolCampus FROM employee\_data WHERE StillWorking = 1 GROUP BY SchoolName, SchoolCampus ORDER BY COUNT(\*) DESC LIMIT 1;**

Execution Time:

1 MB File: 0.0190 seconds

10 MB File: 0.0196 seconds

100 MB File: 0.0195 seconds

A graph with a line

Description automatically generated

5. Query 5:   
  
**SELECT PersonName, JobTitle, DepartmentName, SchoolName, Earnings FROM employee\_data WHERE PersonName = 'Suraj Basavaraj Rajolad';**

Execution Time:

1 MB File: 0.0108 seconds

10 MB File: 0.0103 seconds

100 MB File: 0.0108 seconds

A graph with blue lines

Description automatically generated

6. Query 6:   
  
**SELECT DepartmentName FROM employee\_data GROUP BY DepartmentName ORDER BY AVG(Earnings) DESC LIMIT 1;**

Execution Time:

1 MB File: 0.0455 seconds

10 MB File: 0.0462 seconds

100 MB File: 0.0434 seconds

## A graph with a line going up Description automatically generated

## Files

main.py: The main script to execute queries and measure execution time.

plot.py: A script to generate plots comparing query execution times across different file sizes.

datasets/: Directory containing the CSV files (1 MB, 10 MB, and 100 MB).

## Usage

Run the main.py file to execute all queries and measure the execution time for each file size. The results will be printed in the terminal, and graphs will be generated to visualize the execution times.