



Group-6

DATA 225 DB Systems for Analytics

Project Report

Analyzing Layoffs in the Top 20 Affected States of the US

STUDENTS

Sheetal - 017526678
Aishwrya - 017557579
Lincy - 017414072
Dhruv - 017506541
Pranav - 016567083

Contents

1	Significance to the real world	1
2	Vital lessons learned from the Project	1
3	Innovation	1
4	Teamwork	2
4.1	Roles	2
5	Technical difficulty	2
6	Pair Programming	3
7	Scrum Practice	3
8	Usage of Grammarly	5
9	Performed substantial analysis using database techniques	6
10	Used a new database or data warehouse tool not covered in the HW or class	10
11	Used appropriate data modeling techniques	11
12	Used ETL tool	13
13	Demonstrated how Analytics support business decisions	14
14	Used RDBMS	16
15	Used Data warehouse	17
16	Includes DB Connectivity	17
17	Used NOSQL	18
	References	20

1 Significance to the real world

Analyzing layoffs in the top 20 affected states of the US is significant in order to understand how the economic dynamics are different from those of social and policy ones. It helps in providing insights into how the economic health of regions is impacted and it also indicates the area of growth or decline. By identifying such kinds of industries we are able to deal with the most significant challenge that is present. The policymakers will get a streamlined set of retraining programs or targeted support. In addition to this, understanding the social impact of layoffs is very important as it also includes the employment rates and the potential migration patterns which indeed helps allocating the resources which are effective enough to support the affected individuals and communities. This analysis informs strategic planning for businesses that anticipate the industry trends and also identify the opportunities that are essential for growth or diversification. Altogether, examining the layoffs in these 20 states offers valuable real-world data that not only informs the decisions of policymakers but also investors, job seekers, and businesses, who have more informed and effective responses to the challenges with respect to the economy and growth.

2 Vital lessons learned from the Project

In this whole project, we have learned about many valuable lessons which include industry vulnerabilities, policy effectiveness, and also the economic resilience which can be regional. This project has helped us to highlight certain industries that are more prone to layoffs due to multiple factors that have come up like automation and globalization. The study also highlights variations in regional economic resilience, providing information on things like worker adaptation and community responses. By gaining an understanding of the dynamics of corporate practices during layoffs and long-term employment trends, policymakers, businesses, and communities can better execute targeted interventions, promote economic resilience, and support workers in affected states.

3 Innovation

In order to structure the lessons that we have learned from this project innovation is an essential part of it to understand the effects of layoffs and address the various vulnerabilities in the industry, Innovative strategies must be developed to handle such problems. One of the examples of such strategies includes developing new technologies, business models, or training the workers. In addition to that there are various different approaches which are being researched to an enhance their effectiveness to support such workers and fostering economic residence such as Novel policy. The study also examines how companies use the technology in the creative way and calculate the moves to get around the challenges and capture that chances as the employment landscape changes. At the time of the layoffs, stakeholders can effectively manage any employment difficulties and Continue to move forward in

long term for economic growth by embracing the innovation in corporate tactics, policy making, and Community responses.

4 Teamwork

In this project teamwork is an integral part to achieve our goals. each member of the team contributes that own specialized skills and knowledge, fostering a collaborative environment where diverse perspectives are valued. through our effective communication and coordination, Rihanna's the collective expertise to conduct the rigorous data analysis, did I have meaningful insights, and develop various actionable strategies to address employment challenges in the affected States.

4.1 Roles

Data Extraction - Pranav and Dhruv

Data Cleaning using Python - Pranav and Dhruv

Data Modeling (ER diagrams) - Aishwarya, Sheetal, Lincy, Pranav, Dhruv

Data Warehouse - Aishwarya, Lincy

Design in MySQL - Aishwarya, Sheetal

Data Analysis- Aishwarya, Sheetal

Data Visualization - Aishwarya, Sheetal

Generation of Analysis Report - Lincy, Sheetal

Report Writing in Latex - Lincy, Sheetal, Aishwarya, Pranav.

5 Technical difficulty

- The primary issue that we encountered was that the data set was extremely cluttered and messy and very disorganized. It was very difficult to obtain up to date and trustworthy layoff data from the states that were impacted, especially because of the possible delays in reporting in the WARN act . Now to make sure that we have a complete and current data for analysis we are finding alternate sources such as respectable news organizations and research institutions that frequently release State specific layoff data.
- It can be difficult to integrate the data that is mixed from many different sources and formats. Due to which before analyzing be standardize the data formats as a part of a solution. To guarantee the correctness and the consistency across the data sets, which means this may require making manual tweaks or using data cleaning technologies, allowing for smooth integration for Reliable analysis.

- Sorting through massive volumes of data of employment laws from several States can be very difficult, especially given the different economic environments . To address this, we are using different tactics such as your categorizing States based on their economic sectors, utilizing various visualization tools for more detail insights, and doing comparison analysis to identify different patterns and and to acquire a thorough grasp Off a bigger economic picture. These various methodologies allows us to successfully navigate the Complexity of the data and extract important insights to inform our study.

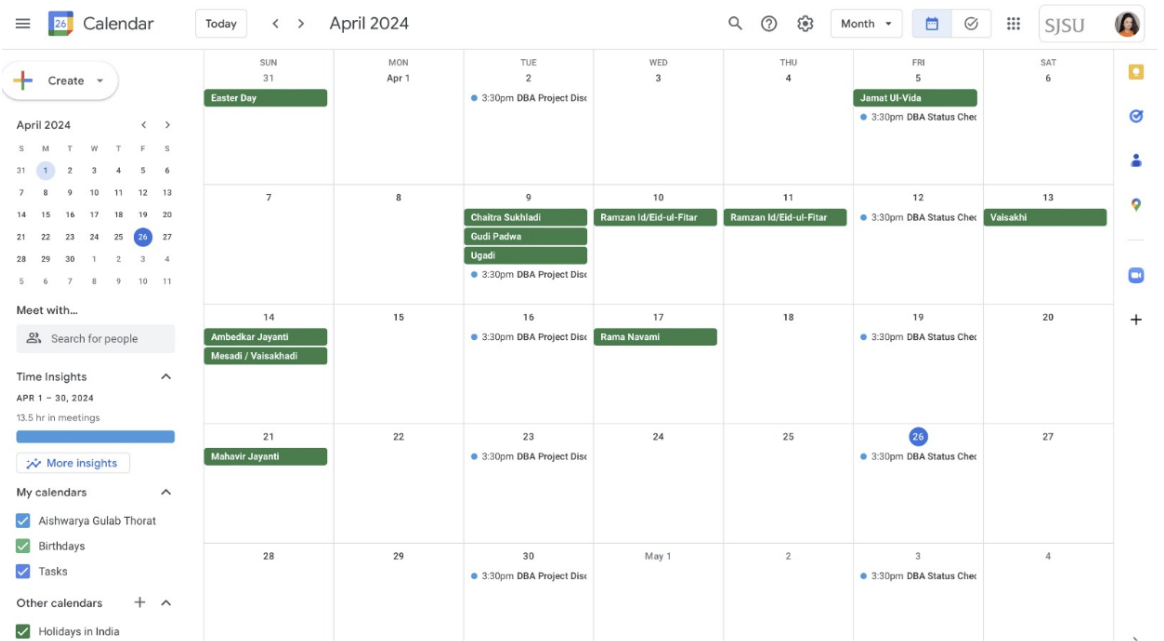
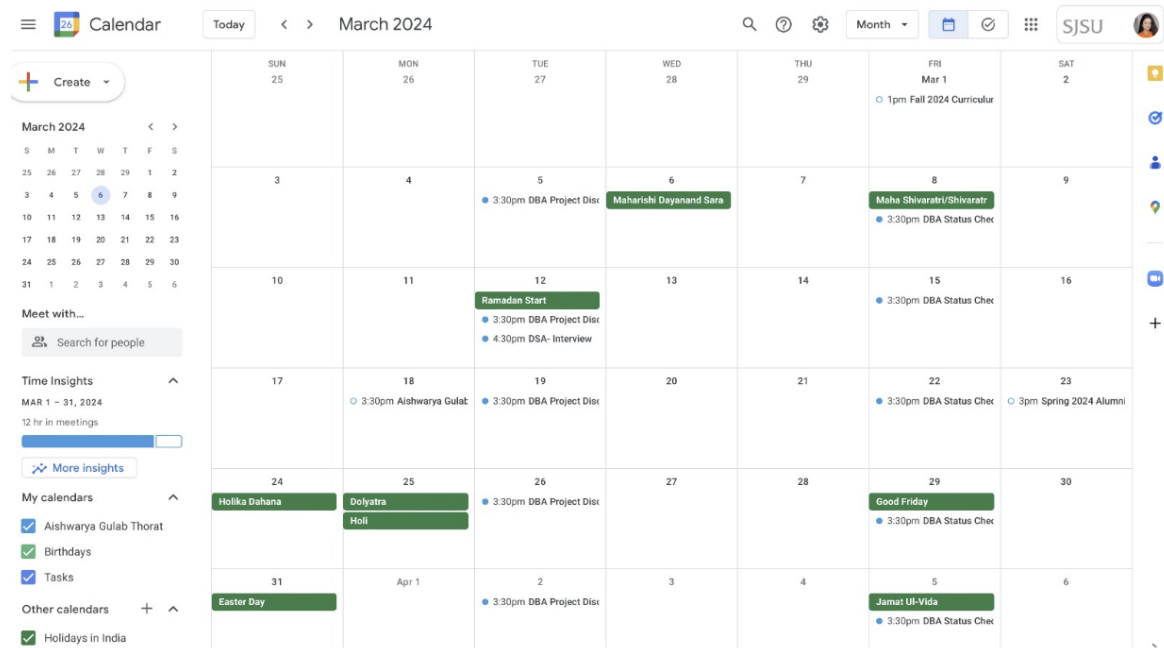
6 Pair Programming

In our project, we embraced pair programming to address the issue of dealing with a messy dataset. One team member assumed the role of the "driver," responsible for hands-on coding tasks like data cleaning and manipulation. Meanwhile, the other team member served as the "navigator," actively reviewing the code, suggesting improvements, and ensuring alignment with project objectives. This dynamic collaboration allowed us to efficiently tackle challenges, enhance code quality, and foster knowledge exchange within the team.

7 Scrum Practice

For this project, we've implemented Agile methodologies, specifically Scrum, utilizing one-week sprints to manage our project effectively. We've established regular sprint planning, weekly stand-up meetings, and sprint reviews to ensure continuous progress and adaptability to changing requirements. Our team maintains transparency and accountability through detailed meeting minutes, task boards, and other artifacts, which are regularly updated and submitted on Canvas for documentation and review.

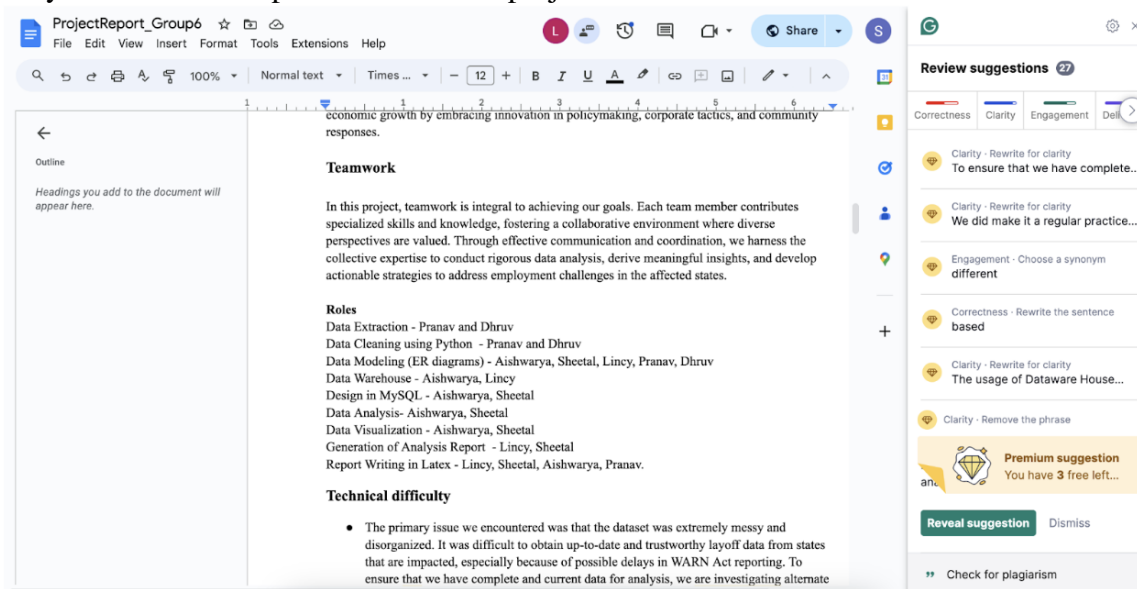
Data 225 Project Report



We did make it a regular practice to jot down the minutes of meeting and did share it to all the team members. One of the example is here: MOM of the DBA project meeting.

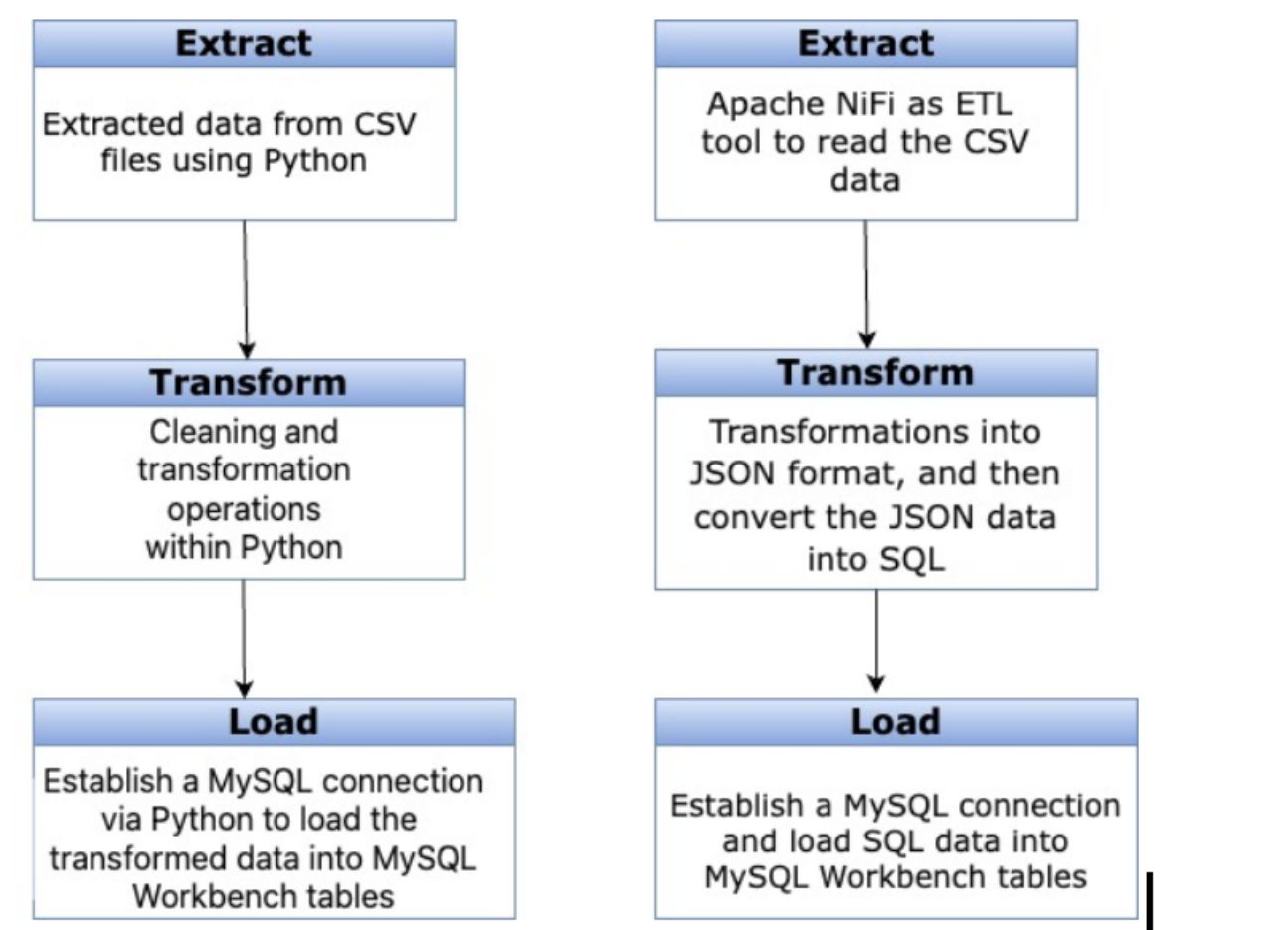
8 Usage of Grammarly

Using Grammarly in our project report was one of the best decisions that we took as it really helped us in making our content very much polished and error-free. The suggestions that Grammarly gave have helped us in correcting all of our spelling mistakes and grammar mistakes. We personally do feel that there is a huge improvement in the sentence structure that we are currently forming and it is maintaining a consistent writing style which is quite crucial in writing reports. There is a significant enhancement in the overall clarity and professionalism of our report which will definitely make it easier for the readers to understand our findings and recommendations. This assistance has saved a lot of time and effort in terms of proofreading and has allowed us to focus more on the content and analysis that we have performed for our project.



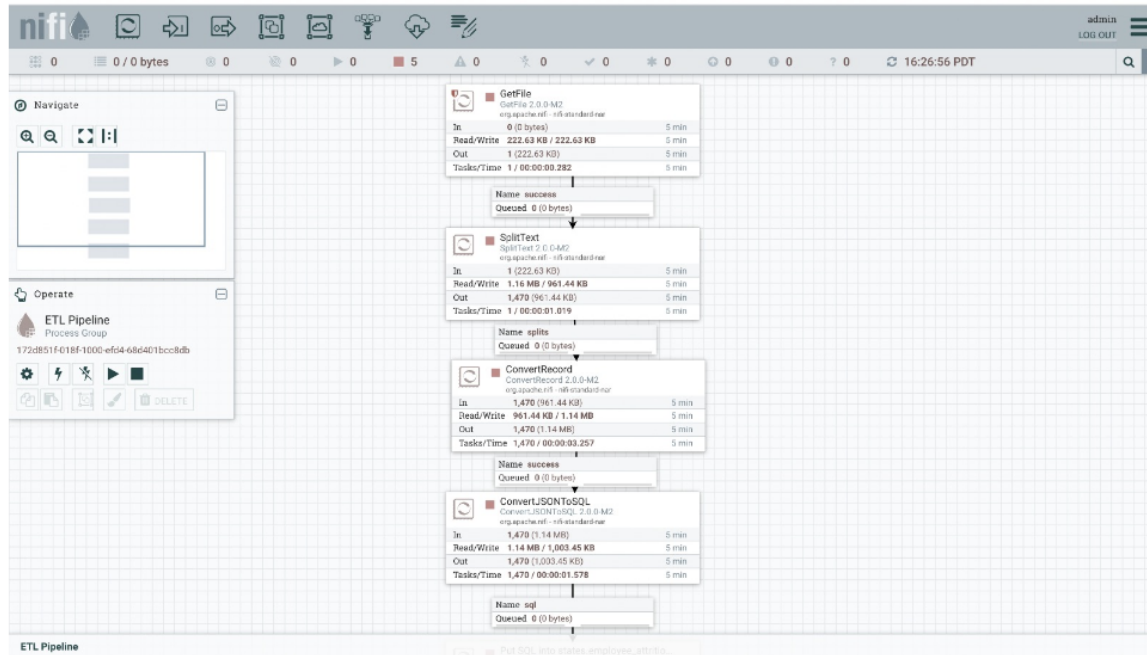
9 Performed substantial analysis using database techniques

We’ve developed two pipelines for loading various CSV files into MySQL Workbench. In the first approach, we utilize Python to ingest the CSV files, conduct cleaning and transformation operations within Python, and subsequently establish a MySQL connection via Python to load the transformed data into MySQL Workbench tables. Alternatively, in the second pipeline, we leverage Apache NiFi as our ETL tool to read the data, perform transformations into JSON format, and then convert the JSON data into SQL. This SQL data is then loaded into MySQL Workbench tables.



Data 225 Project Report

For the ETL tool, we have used Apache NiFi.



Once the data is in MySQL Workbench, we conduct various analyses, a few of which are listed below with the screenshots:

- Obtaining the total number of layoffs by states.

```
## Get Total number of layoffs by states
select s.state,sumViews(c.`Number of Workers`) as 'total_layoffs' from state_info s inner join company_info c
on c.Company=s.Company
group by s.state;
```

OUTPUT

state	total_layoffs	
Alabama	485608	
California	8018759	
Arizona	472937	
Texas	652335	
Ohio	1604367	
Florida	2519179	
Pennsylvania	1707482	
Michigan	1157543	
Kansas	569834	
New York	1338878	
Maryland	543950	
Georgia	1424520	
Washington	907892	
Virginia	695675	
New Jersey	1089643	
Missouri	586114	
Connecticut	229501	
Indiana	411112	
Result 2		

- Calculating the total number of layoffs by Closure/Layoff Statuses.

```
## Get Total number of layoffs by Closure/Layoff Statuses
select `Closure/Layoff`,sum(`Number of Workers`) as 'total_layoffs' from company_info
group by `Closure/Layoff`
having sum(`Number of Workers`)>0
order by total_layoffs desc;
```

OUTPUT

Closure/Layoff	total_layoffs	
Layoff	496963	
Closure	493889	
Layoff Temporary	430325	
Plant Closing	393842	
Layoff Permanent	390307	
Closure Permanent	254859	
Mass Layoff	250716	
Plant Closure	206746	
Plant Layoff	105520	
Closure Temporary	98004	
Temporary Plant Closi...	80737	
Temporary Plant Layoff	78221	
Facility closure	61631	
Closing	49545	
Layoff	43590	
Layoff Unknown at thi...	41795	
Possible Plant Closing	21732	
Plant Unit Closing	17112	
Result 3		

- Aggregating the total number of layoffs by city.

```
## Get Total number of layoffs by city
select s.city,sum(c.`Number of Workers`) as 'total_layoffs' from state_info s inner join company_info c
on c.Company=s.Company
group by s.city
having sum(c.`Number of Workers`)>0
order by sum(c.`Number of Workers`) desc;
```

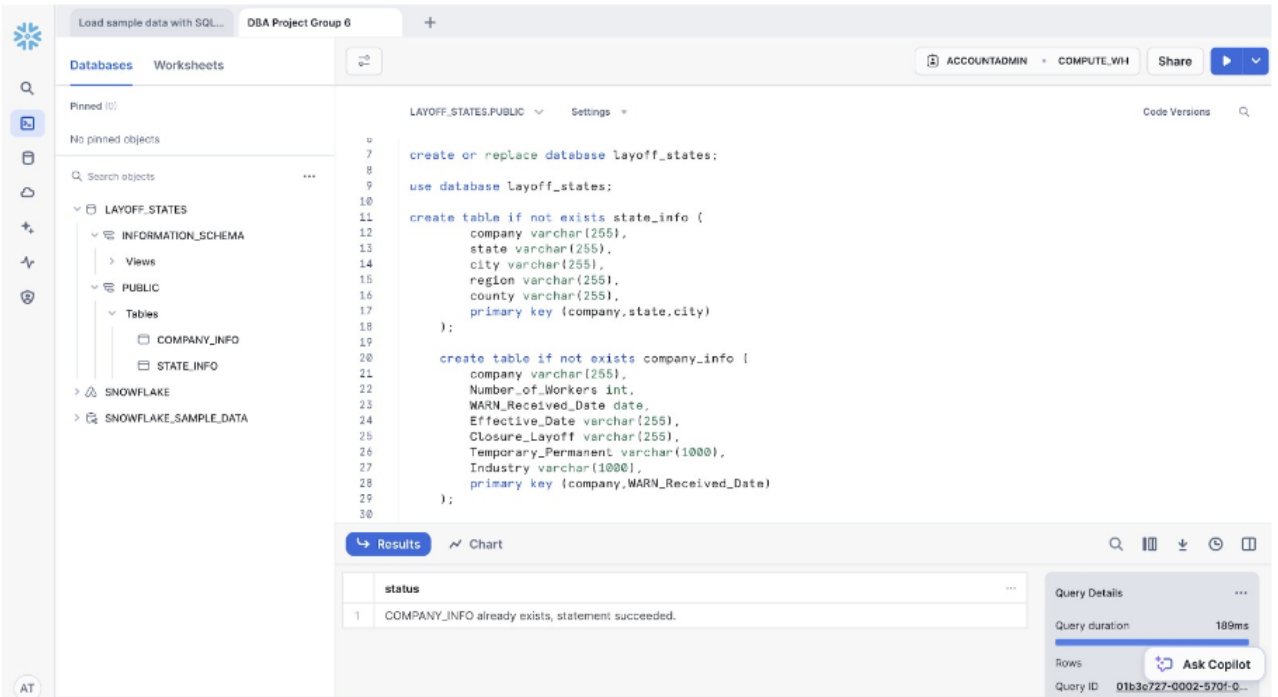
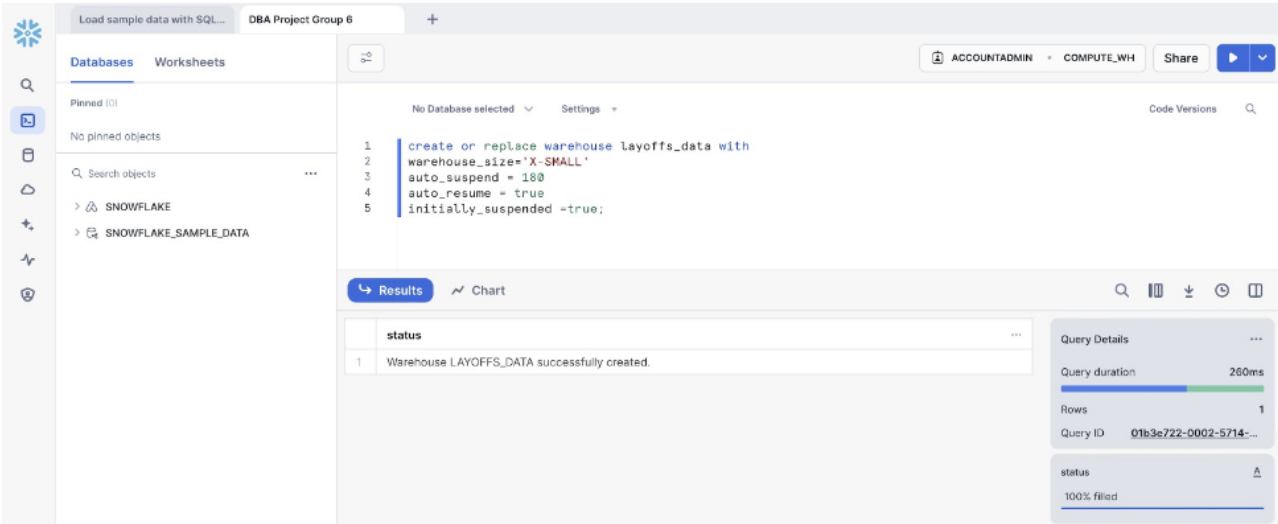
OUTPUT

city	total_layoffs
Los Angeles	322264
Atlanta	313515
New York	290109
San Francisco	282551
Orlando	229953
Miami	219081
San Diego	218450
Tampa	217473
Jacksonville	202845
Wichita	198586
none	190234
Seattle	185180
Philadelphia	177007
Columbus	169916
San Jose	169266
St. Louis	161547
Irvine	143767
Newark	136086
Result 4	

These analyses allow us to gain insights into the distribution of layoffs across different geographical regions and closure statuses, enabling informed decision-making and strategic planning.

10 Used a new database or data warehouse tool not covered in the HW or class

In addition to the pipelines mentioned above, we've established a data warehouse database in **Snowflake** and configured tables within it. So in this, we have loaded the transformed data into this Snowflake database, which enables the reporting capabilities. The screenshots below show the visual confirmation of our setup and our data-loading processes.



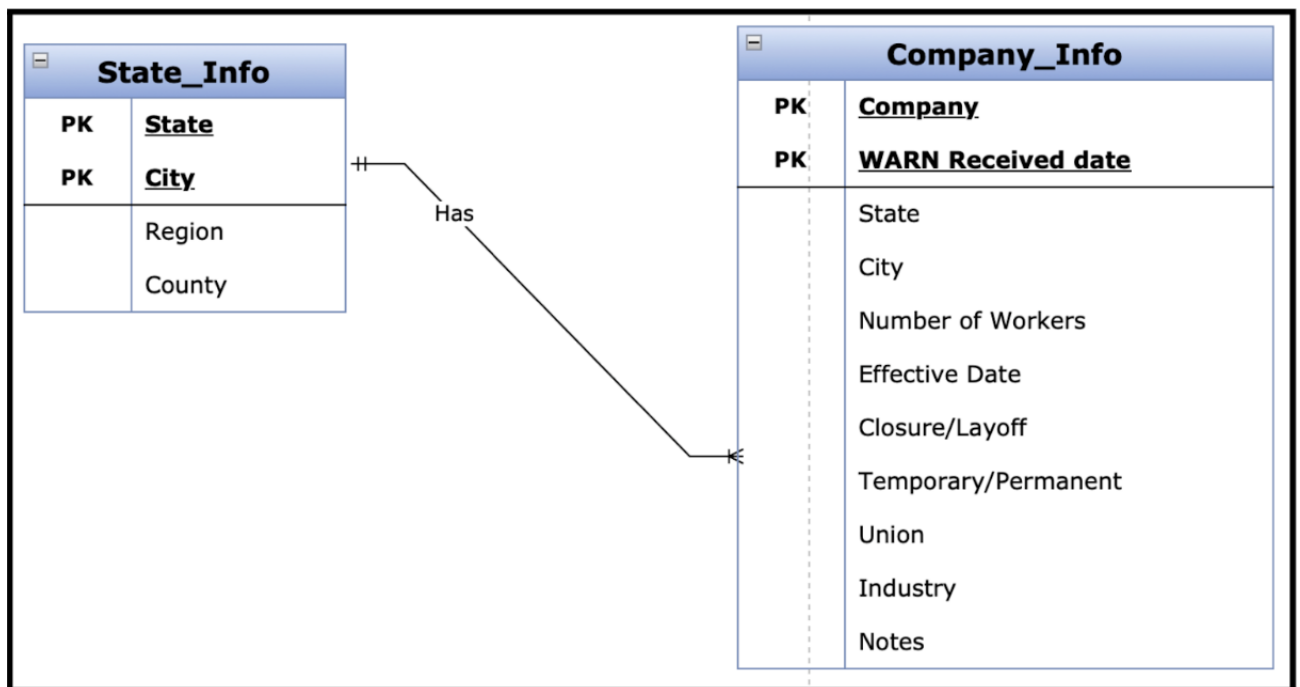
11 Used appropriate data modeling techniques

In this project, we have used to data modeling techniques to understand the different patterns of

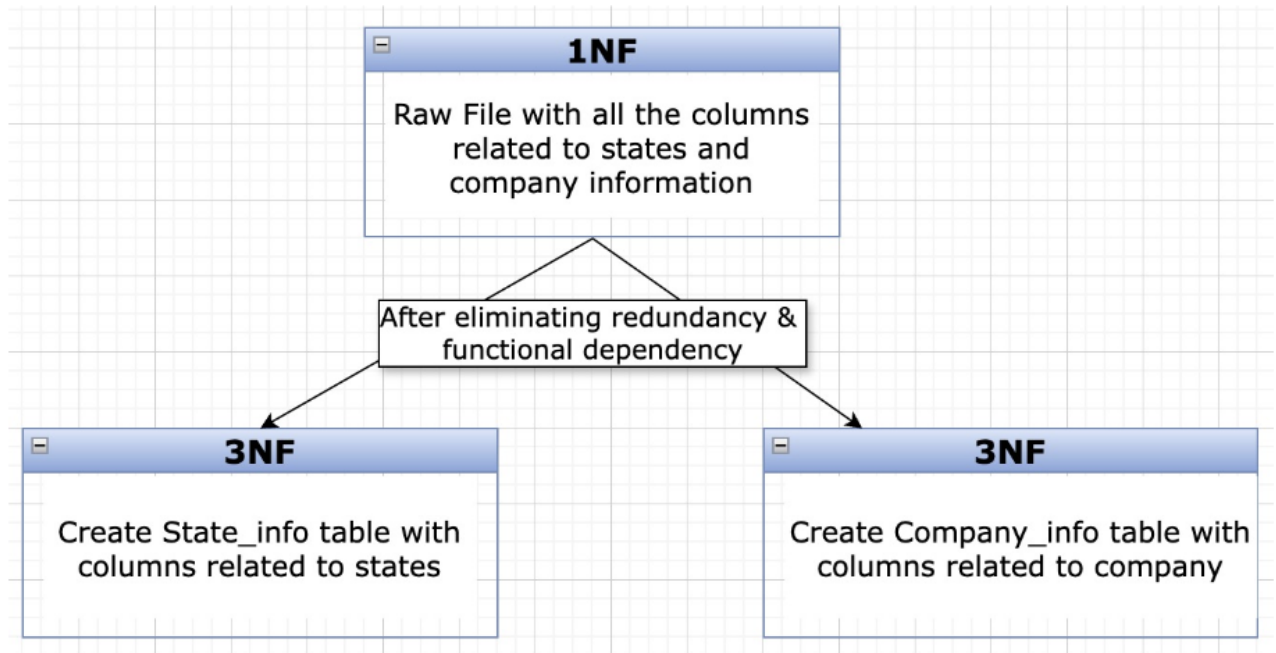
layoffs across the most affected states in the United States.

1. **Entity relationship modeling** we use this technique to identify and Define the entities such as the state industry companies and the relationships such as layoff affecting employee rates Etc within the given data set by using this ER relationship diagrams. It helps us to visualize the relationships and to make it easier to understand the complex data structures.

ER Diagram



2. **Normalization** We have used normalization and we applied it to ensure that the database is organized efficiently which reduces the redundancy issues and improves the data integrity. The normalization techniques that we have applied are First Normal Form (1NF), Second Normal Form (2NF), and Third Normal Form (3NF), to ensure data integrity and optimize database performance.



3. **Dimensional modeling** by using this technique we organize the data into different dimensions and facts such as the number of layoffs unemployment rate Etc this basically creates a data warehouse which is optimized for analytical queries which enables us to analyze the layoffs from different perspectives easily. We've established a data warehouse database in **Snowflake** and configured tables within it.
4. **Data Modelling Tool:** We have leveraged the usage of the MySQL database to perform various data modeling activities.

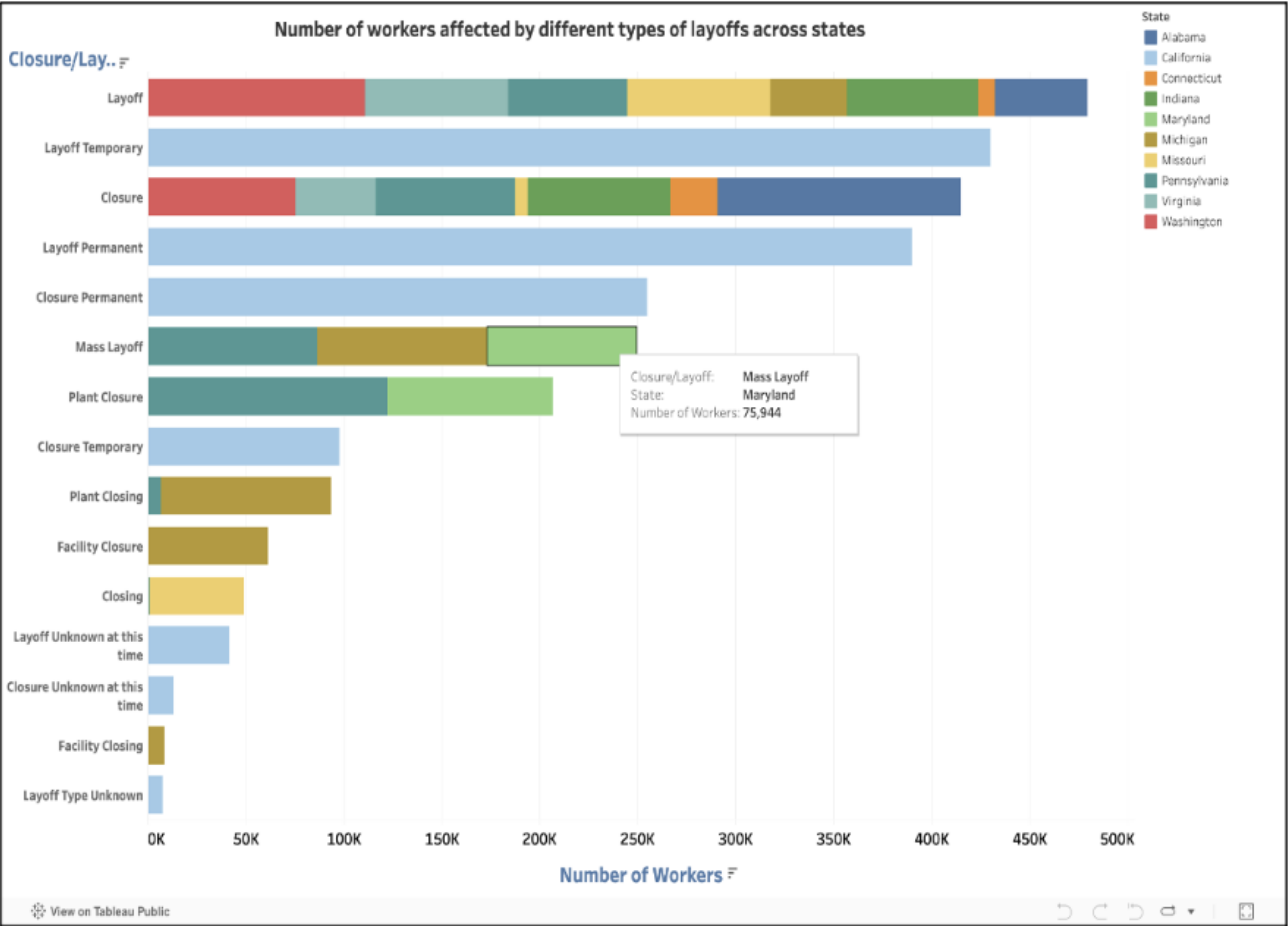
12 Used ETL tool

We have used Apache NiFi as the ETL tool which has played a very crucial role in our project workflow. We have used this to read the data, perform transformations into JSON format, and then convert the JSON data into SQL. This SQL data is then loaded into MySQL Workbench tables. This helped in standardizing the data which ensured the data was consistent and accurate.

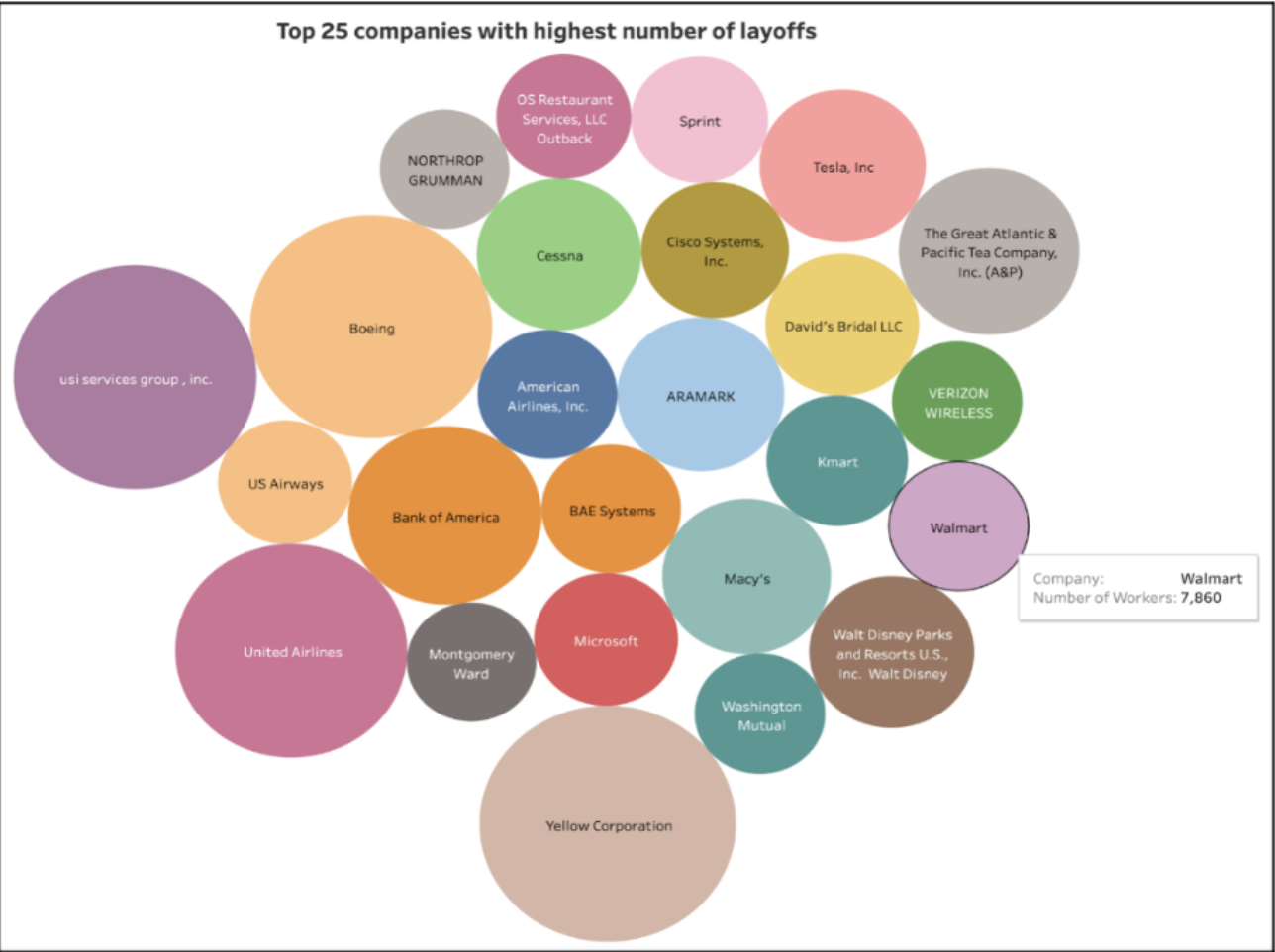
We have also utilized Python as an ETL Tool to ingest the CSV files, post which we also performed the cleaning and transformation operations within Python. Further, we established a MySQL connection via Python in order to load the transformed data into MySQL Workbench tables.

1000

• • • • •



The graph breaks down layoffs into different categories, such as closures (plant, facility), temporary and permanent layoffs, and mass layoffs. By analyzing the distribution of these categories across states, you can identify which types of layoffs are most prevalent in different regions.



This shows the top 25 companies with the highest number of layoffs. It lists companies like Yellow Corporation, USI service groups, US Airways, United Airlines, Macy's, and Bank of America as the top listed companies with highest layoffs.

14 Used RDBMS

We used RDBMS - **MYSQL** for this particular project which played a crucial role in the structuring of the data which would then be defined into tables and relationships. The SQL queries also helped us in analyzing the data based on the different states Industries and other dimensions it also helped us in understanding and performing complex joins and aggregations for detailed analysis.

15 Used Data warehouse

We've created a data warehouse in **Snowflake** and configured tables within it. So in this, we have loaded the transformed data into this Snowflake database. There is continuous improvement through the usage of this and the performance monitoring and decision making is also enhanced.

16 Includes DB Connectivity

We utilize Python to ingest the CSV files, conduct cleaning and transformation operations within Python, and subsequently **establish a MySQL connection via Python** to load the transformed data into MySQL Workbench tables. The attached screenshot below shows how the connection is established.

Connect to MYSQL Server, Create Tables and Upload Data into Tables

```
import mysql.connector
import csv
import getpass

try:
    password = getpass.getpass("Enter MySQL password: ")
    connection = mysql.connector.connect(
        host='localhost',
        user=input('Enter Username: '),
        password=password
    )
    print("Connected to MySQL server successfully!")

    cursor = connection.cursor()

    # Use newly created database
    cursor.execute("use States")

    cursor.execute("""
create table if not exists state_info (
    company varchar(255),
    state varchar(255),
    city varchar(255),
    region varchar(255),
    county varchar(255),
    primary key (company,state,city)
)
""")
    print("state_info table created successfully!")

    cursor.execute("""
create table if not exists company_info (
    company varchar(255),
    Number_of_Workers int,
    WARN_Received_Date date,
    Effective_Date varchar(255),
    Closure_Layoff varchar(255),
    Temporary_Permanent varchar(1000),
    Industry varchar(1000),
    primary key (company,WARN_Received_Date)
)
""")
    print("company_info table created successfully!")

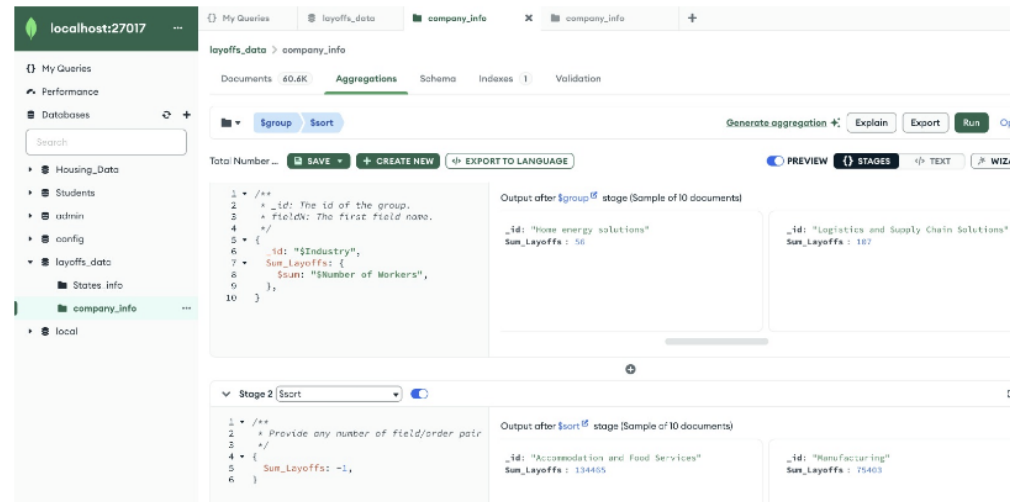
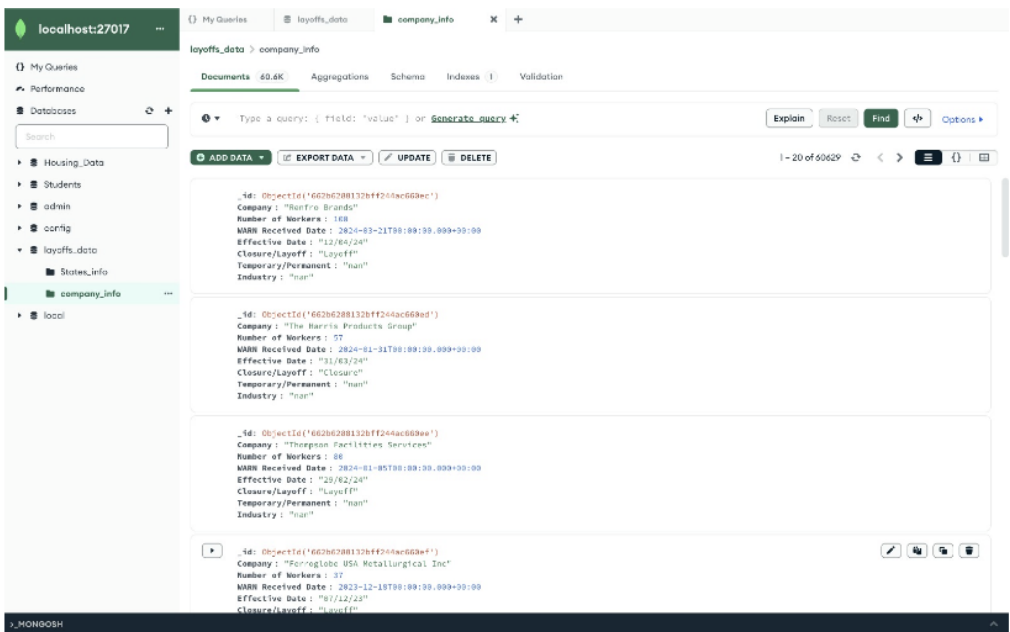
    connection.commit()
except mysql.connector.Error as error:
    print("Failed to connect to MySQL server:", error)
finally:
    if 'cursor' in locals():
        cursor.close()
    if 'connection' in locals() and connection is connected():
```

17 Used NOSQL

We have leveraged the usage of MongoDB for the NoSQL part as NoSQL databases are very important, particularly for managing a different data types which gives us scalability. NoSQL database provides us flexibility in storing such data. We use NoSQL in our project which helps us understand the context of the layoffs in the US. We have performed aggregations in the MongoDB compass to

Data 225 Project Report

perform essential analysis.



References

- [1] <https://www.cnn.com/2024/01/20/as-companies-lay-off-even-more-workers-they-could-be-making-a-mistake.html>
- [2] <https://layoffs.fyi/category/analysis/>
- [3] <https://www.wartracker.com/>
- [4] <https://layoffdata.com/data/>
- [5] <https://app.grammarly.com/>
- [6] <https://openai.com/blog/chatgpt>
- [7] <https://www.cloudflare.com/learning/security/api/what-is-api-call/>
- [8] <https://martinfowler.com/articles/on-pair-programming.html>
- [9] <https://www.computerworld.com/article/1617283/tech-layoffs-in-2023-a-timeline.html>