

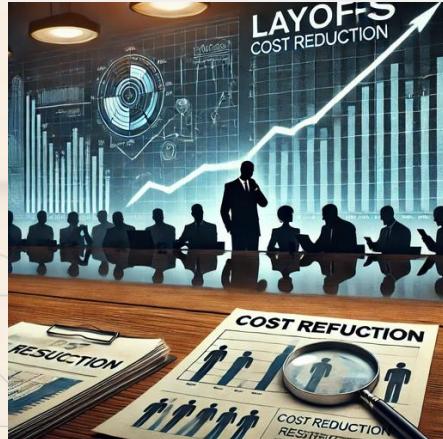
DATA ANALYSIS of -LAYOFFs-



**Anan Gamal Mohammed Mohammed
Dalia Abbas Ahmed Mohammed Shaltout
Merhan Hassan Mohammed Elsayaby
Zainab Kamal Kamel Maamoun**



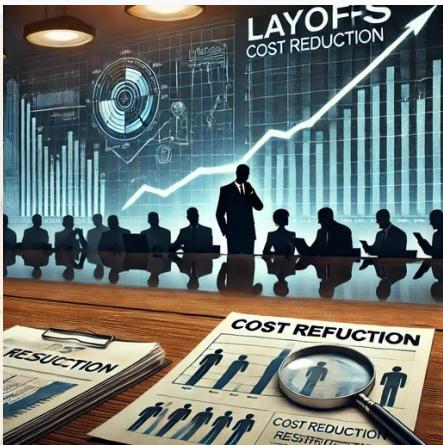
what is Lay-Off?



A layoff is the temporary or permanent termination of a worker's employment for reasons unrelated to the individual's performance on the job, also known as workforce reductions or employee downsizing.

Source: What a Layoff Means, Plus Statistics and an Example ([investopedia.com](https://www.investopedia.com/terms/l/layoff.asp))

Why Lay-Offs?



- Rapid change in technology.
- High international competition
- Changing customer needs
- Changing Regulations
- Regional downturns in economy

[Source: Tech sector layoffs explained: What you need to know \(techtarget.com\)](https://www.techtarget.com/itguru/explained/what-is-a-tech-sector-layoff)

Project Idea

The Problem



- Problem is
The Job loss all over the world by certain companies, industries also regions which affect economic stability.

The Solution



- Solution is
To visualize a Dash-Board allows stakeholders to understand patterns of layoffs within companies and regions.

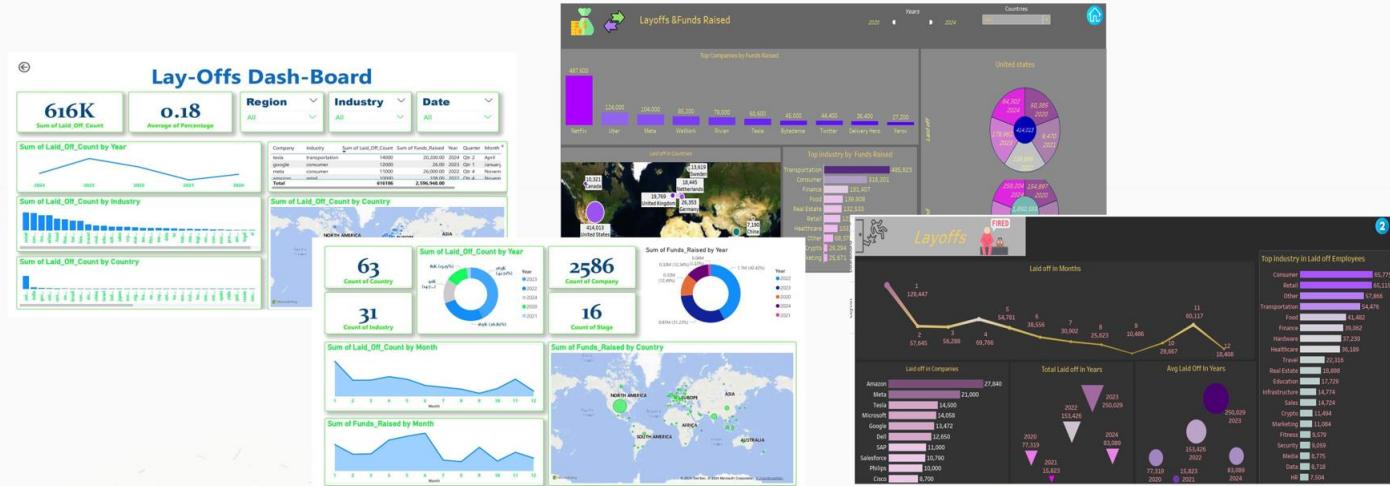
Unique Value Proposition



- Unique value proposition
is the ability to create Actionable insights in real time across the world.

PROJECT WIREFRAME

3
9
2



End Users & Features



The Primary Users

- Primary users' persona are:
- Human resources.
 - Economists.
 - Investors and Organizations.
 - Government Leaders.
 - Data Analysts.



10/22/2024



The Key features

- Key features that address user needs is
- Layoffs trend Analysis
- Geographical Layoffs Map distribution.
- Average Layoffs Percentage.
- Data sharing and generating reports.



Features solving the problems

- Solve the problems of each end-user
The trend analysis, Geographic distribution Average And Reports will have positive Impact on users.

End Users & Features

Average Layoffs Percentage



Uncovering and acknowledge of the Layoffs
Average within companies and industries:

Impact investors, decision makers and leaders to invest more and expand or not.

Geographical Layoffs Map



- Map distribution illustrate Layoffs affected regions and require interventions
- Impact mostly government leaders and organizations to make support programs for layoffs.

Layoffs Trend Analysis

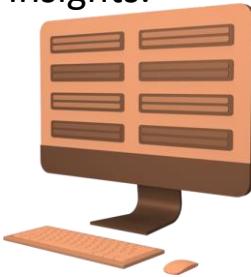


- The trend Analysis allows end users to identify economy changes and patterns
- Impact HR to create workforce strategy.

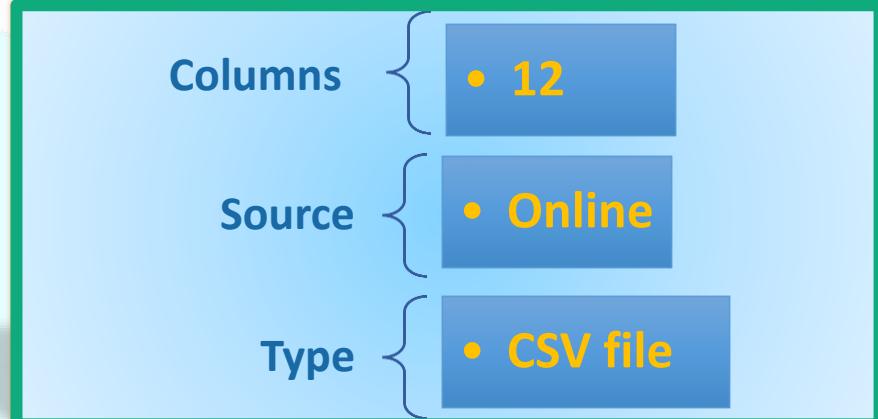
Data sharing



- Data Sharing also generating reports can be done easily
- Impact Data Analysts And decision makers by its informative insights.



DATA Structure before



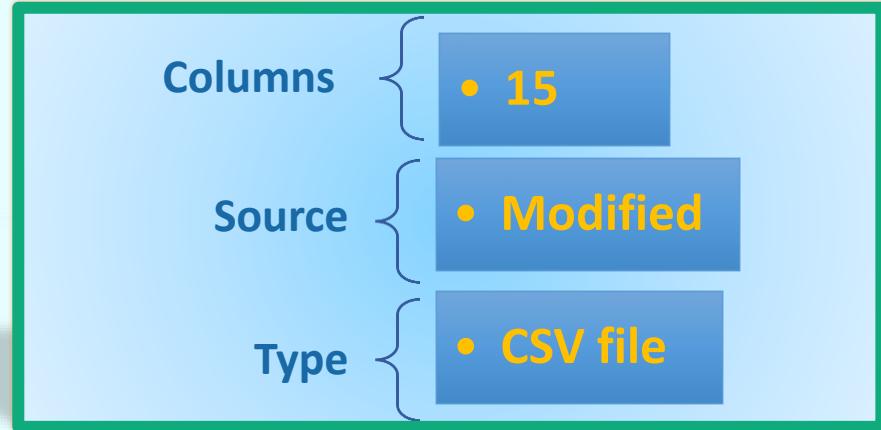
https://www.kaggle.com/datasets/theakhilb/layoffs-data-2022?utm_medium=social&utm_campaign=kaggle-dataset-share&utm_source=linkedin.

Columns Are:

- 1- Company. (where layoffs happened)
- 2- Location_HQ. (where layoff took place, allows analysis of economic study)
- 3- Industry. Describe company's operations)
- 4- Laid_off_count. (numbers of employees laid off)
- 5- Date.(time when layoffs occurred)
- 6- Source.(origin of data)
- 7- Funds_Raised.(amount of funds company raised)
- 8- Stage.(stage of layoffs)
- 9- Date_Added .(when it has been added)
- 10- Country. (where lay off took place, allows analysis of economic study)
- 11- Percentage. (percentage of total laid off employees)
- 12- List_of_employees_laid_off. (specific data)

	A	B	C	D	E	F	G	H	I	K	L	M	N
1	Company	Location_I	Industry	Laid_Off_Ct	Date	S	Country	Percentag	List_of_Employees_Laid_Off				
2	Oda	Oslo	Food	150	6/5/2024	https://te	691	Unknown	####	Norway	Unknown		
3	Pagaya	Tel Aviv	Finance	100	6/5/2024	https://w	2000	Post-IPO	####	Israel	0.2	Unknown	
4	Aleph Farr	Tel Aviv	Food	30	6/5/2024	https://w	119	Unknown	####	Israel	0.3	Unknown	
5	MoonPay	Dover	Crypto	30	6/5/2024	https://w	651	Unknown	####	United Sta	0.1	Unknown	
6	Yext	New York	Marketing		6/5/2024	https://w	117	Post-IPO	####	United Sta	0.12	Unknown	
7	Microsoft	Seattle	Other	1000	6/3/2024	https://w	1	Post-IPO	####	United States	Unknown	Unknown	
8	OrCam	Jerusalem	Healthcare	100	6/3/2024	https://w	86	Unknown	####	Israel	0.5	Unknown	
9	Google	SF Bay Are	Consumer	100	####	https://w	26	Post-IPO	####	United States	Unknown	Unknown	

DATA Structure after



Columns Are:

- 1- Company type text.
- 2- Location_HQ type text.
- 3- Industry type text.
- 4- Laid_off_count type whole number.
- 5- Date type Date Date/time.
- 6- Source type text.
- 7- Funds_Raised type decimal number.
- 8- Stage type text.
- 9- Date_Added type Date/time.
- 10- Country type text.
- 11- Percentage type decimal number.
- 12- List_of_employees_laid_off type text.
- 13- Year type whole number.
- 14- Month type whole number.
- 15- Region type text.

	A	B	C	D	E	CSV Dataset of Layoffs										K	L	M	N	O	P
1	Company	Location_HQ	Industry	Laid_Off_Count	Date	Source	Funds_Raised	Stage	Date_Added	Country	Percentage	List_of_Employees_laid_off	Year	Month	Region						
2	oda	oslo	food	150	#####	https://te	691	Unknown	#####	norway	0	Unknown	2024	6	Europe						
3	pagaya	tel aviv	finance	100	#####	https://w	2000	Post-IPO	#####	israel	0.2	Unknown	2024	6	Middle East						
4	aleph farn	tel aviv	food	30	#####	https://w	119	Unknown	#####	israel	0.3	Unknown	2024	6	Middle East						
5	microsoft	denver	crypto	20	#####	https://w	651	Unknown	#####	United Sta	0.1	Unknown	2024	6	North America						

SQL

```
SELECT * FROM layoffs_data
LIMIT 10;

PRAGMA
table_info(layoffs_data); -- For
SQLite
-- OR
DESCRIBE layoffs_data; -- For
MySQL/PostgreSQL

SELECT company, date,
COUNT(*) AS row_count
FROM layoffs_data
GROUP BY company, date
HAVING COUNT(*) > 1;

SELECT
    SUM(CASE WHEN
num_laid_off IS NULL THEN 1
ELSE 0 END) AS
missing_num_laid_off,
    SUM(CASE WHEN date IS
NULL THEN 1 ELSE 0 END) AS
missing_date,
    SUM(CASE WHEN notes IS
NULL THEN 1 ELSE 0 END) AS
missing_notes
FROM layoffs_data;
```

1. Data Inspection

```
DELETE FROM layoffs_data
WHERE rowid NOT IN (
    SELECT MIN(rowid)
    FROM layoffs_data
    GROUP BY company, date
);
---
```

2. Remove Duplicate Records

Keep one occurrence
for each duplicate
entry based on
company and date:

```
SELECT
    COALESCE(num_laid_off,
'Unknown') AS num_laid_off,
    COALESCE(date, 'Unknown')
AS date,
    COALESCE(notes, 'Not
specified') AS notes
FROM layoffs_data;
---
```

3. Handle Missing Values (Without Deleting Nulls)

Use COALESCE to
provide fallback values
only in queries
without altering the
underlying data.

```
ALTER TABLE layoffs_data ADD  
COLUMN num_laid_off_clean  
INT;
```

```
UPDATE layoffs_data  
SET num_laid_off_clean =  
CAST(num_laid_off AS INT)  
WHERE num_laid_off IS NOT  
NULL;
```

4. Correct Data Types

Ensure data type correction without dropping columns:

Keep both old & new columns to avoid Data loss.

```
UPDATE layoffs_data  
SET company =  
LOWER(TRIM(company)),  
industry =  
LOWER(TRIM(industry)),  
location =  
LOWER(TRIM(location))  
WHERE company IS NOT NULL  
OR industry IS NOT NULL OR  
location IS NOT NULL;  
---
```

5. Standardize Text Fields

Handle text normalization while keeping NULL values intact

```
DELETE FROM layoffs_data  
WHERE num_laid_off < 0 OR  
percentage_laid_off > 100;  
---
```

6. Remove Outliers and Invalid Data

Avoid unnecessary deletions by keeping questionable records intact.

```
ALTER TABLE layoffs_data ADD
COLUMN year INT;
ALTER TABLE layoffs_data ADD
COLUMN month INT;

UPDATE layoffs_data
SET year = STRFTIME("%Y", date),
month = STRFTIME("%m", date)
WHERE date IS NOT NULL;

ALTER TABLE layoffs_data ADD
COLUMN region TEXT;

UPDATE layoffs_data
SET region = CASE
WHEN location LIKE '%USA%' OR location LIKE '%Canada%' THEN 'Americas'
WHEN location LIKE '%Germany%' OR location LIKE '%UK%' THEN 'EMEA'
WHEN location LIKE '%India%' OR location LIKE '%China%' THEN 'APAC'
ELSE 'Other'
END
WHERE location IS NOT NULL;
--
```

7. Data Enrichment
Populate new columns
while keeping NULL values
where appropriate.

```
SELECT *
FROM layoffs_data
WHERE num_laid_off IS NULL OR
date IS NULL;
```

```
SELECT company, date,
COUNT(*)
FROM layoffs_data
GROUP BY company, date
HAVING COUNT(*) > 1;
```

```
SELECT * FROM layoffs_data
LIMIT 10;
```

8. Quality Assurance
Check for missing data
without replacing or
deleting it.

Summary

Clean Data Ready for Visualization This workflow ensures:

1- Data Integrity:

No duplicates, outliers, or invalid data points.

2- Consistency:

Standardized text and accurate data types.

3- Enrichment:

Additional columns like year, month, and region for better analysis.

4- Scalability:

Structured, reusable SQL queries suitable for enterprise use.

N>B

Duplicates removed based on Company and Date.

Data types corrected (dates standardized, numeric values validated).

New analytical columns added (Year, Month, Region)



1.Import

2.Load

3.Handling missing values

```
7 # importing the necessary libraries
8 import pandas as pd
9 import numpy as np
10
11
12 # In[3]:
13
14
15 # loading the dataset
16 filePath = "C:\\\\Users\\\\AE-H\\\\Desktop\\\\archive\\\\layoffs_data.csv"
17 layoffsData = pd.read_csv(filePath)
18
19
20 # In[4]:
21
22
23 # handling missing values by filling with the median of the column in cases of numeric values and N/A in the case of categorical values
24 layoffsData['Laid_Off_Count'] = layoffsData['Laid_Off_Count'].fillna(layoffsData['Laid_Off_Count'].median())
25 layoffsData['Funds_Raised'] = layoffsData['Funds_Raised'].fillna(layoffsData['Funds_Raised'].median())
26 layoffsData['Percentage'] = layoffsData['Percentage'].fillna(layoffsData['Percentage'].median())
27 layoffsData['List_of_Employees_Laid_Off'] = layoffsData['List_of_Employees_Laid_Off'].replace('Unknown', 'N/A')
28 layoffsData['List_of_Employees_Laid_Off'] = layoffsData['List_of_Employees_Laid_Off'].fillna('N/A')
29
```



4.Convert Date type

5.Ensure numeric format

6.Remove duplications

```
34 # converting date values into datetime
35 layoffsData['Date'] = pd.to_datetime(layoffsData['Date'])
36 layoffsData['Date_Added'] = pd.to_datetime(layoffsData['Date_Added'])

37

38

39 # In[ ]:

40

41

42 # ensuring numeric columns are in numeric format
43 layoffsData['Laid_Off_Count'] = pd.to_numeric(layoffsData['Laid_Off_Count'])
44 layoffsData['Funds_Raised'] = pd.to_numeric(layoffsData['Funds_Raised'])
45 layoffsData['Percentage'] = pd.to_numeric(layoffsData['Percentage'])

46

47

48 # In[6]:

49

50

51 # removing Duplicates
52 layoffsData = layoffsData.drop_duplicates()

53
```

7. Ensure consistency of categorial values

8. Add new columns

9. Export CSV file

```
58 # ensuring consistency in categorical values
59 layoffsData['Industry'] = layoffsData['Industry'].str.strip().str.title()
60 layoffsData['Country'] = layoffsData['Country'].str.strip().str.title()
61 layoffsData['Location_HQ'] = layoffsData['Location_HQ'].str.strip().str.title()
62
63
64 # In[8]:
65
66
67 # adding a new 'Region' column
68 country_to_region = {
69     'United States': 'North America',
70     'Canada': 'North America',
71     'Israel': 'Middle East',
72     'Norway': 'Europe',
73     'India': 'Asia',
74     'Germany': 'Europe',
75     'United Kingdom': 'Europe',
76     'Australia': 'Oceania',
77 }
78 layoffsData['Region'] = layoffsData['Country'].map(country_to_region).fillna('Other')
79
80
81 # In[9]:
82
83
84 # exporting the cleaned data
85 layoffsData.to_csv('cleanedLayoffsData.csv', index=False)
86
```

Frameworks & Deliverables

Application used	Why
Zoom	Online meetings
SQL code	programming
Python code	programming
Plateau	Visualization
Power BI	Visualization
Power Point	Presentation

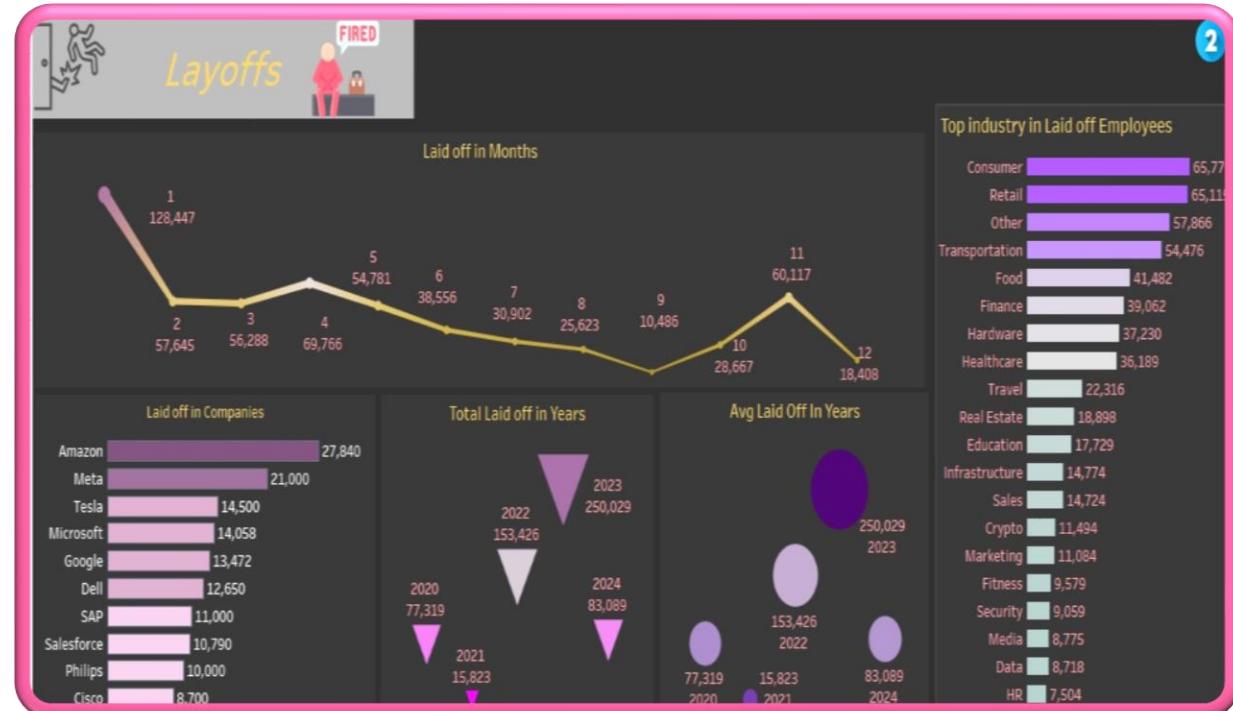
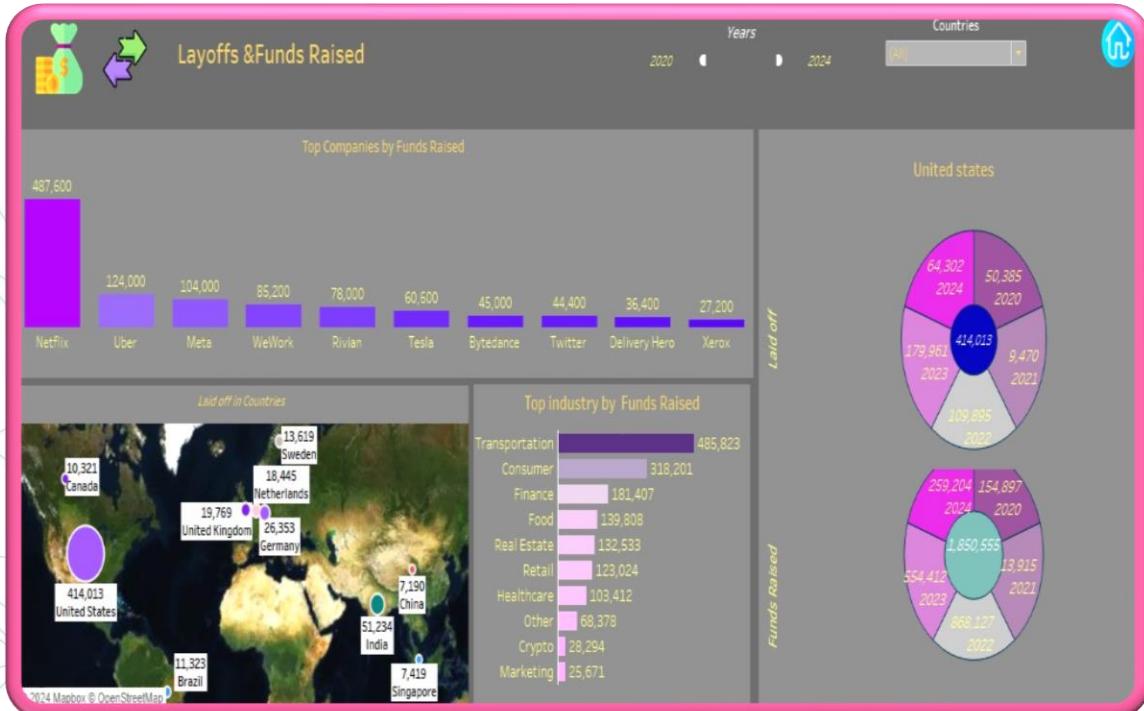


Deliverables:

- 1- CSV file of cleaned Data.
- 2- Dash-Board by Plateau.
- 3- Dash-Board by Power-BI.
- 4- Project Presentation.



Plateau Dash-board

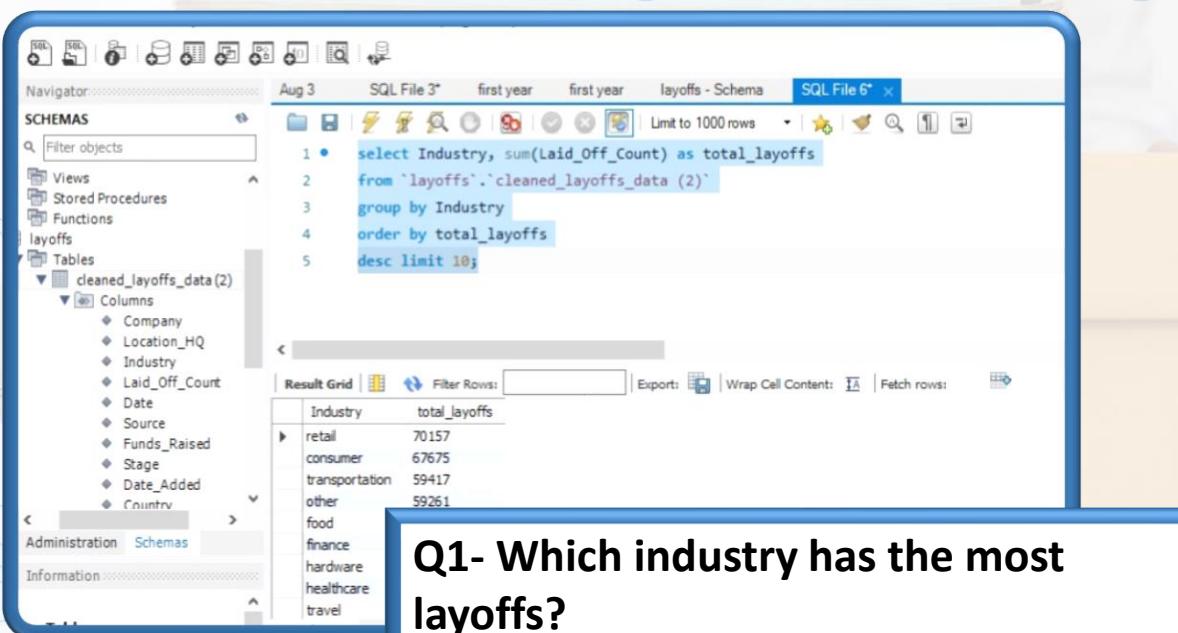


Power BI Dash-board



MySQL

Exploratory Data Analysis

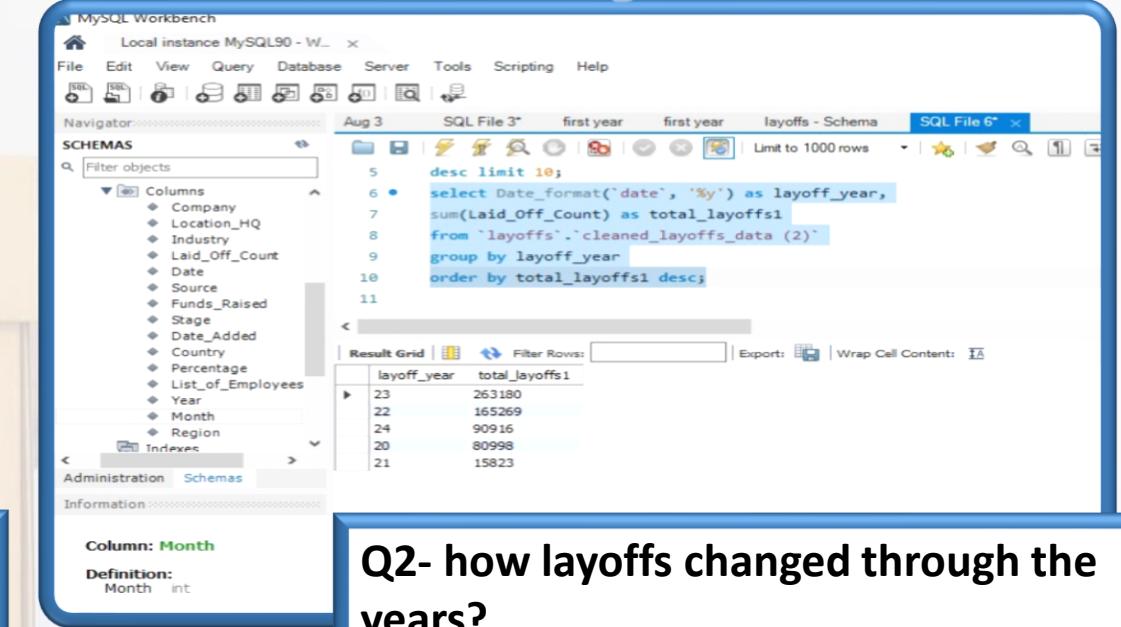


```
1 • select Industry, sum(Laid_Off_Count) as total_layoffs
2   from `layoffs`.`cleaned_layoffs_data` (2)
3   group by Industry
4   order by total_layoffs
5   desc limit 10;
```

Industry	total_layoffs
retail	70157
consumer	67675
transportation	59417
other	59261
food	
finance	
hardware	
healthcare	
travel	

Q1- Which industry has the most layoffs?

The dataset shows the most affected industry sectors and
The answer is retail of total of 70157 layoffs



```
5
6 • select Date_format(`date`, '%y') as layoff_year,
7   sum(Laid_Off_Count) as total_layoffs1
8   from `layoffs`.`cleaned_layoffs_data` (2)
9   group by layoff_year
10  order by total_layoffs1 desc;
```

layoff_year	total_layoffs1
23	26180
22	165269
24	90916
20	80998
21	15823

Q2- how layoffs changed through the years?

Trends query shows the most affected year and
The answer is 2023

MySQL Exploratory Data Analysis

```

    9 group by layoff_year
    10 order by total_layoffs1 desc;
    11 • select Region,
    12 sum(Laid_Off_Count) as total_layoffs2
    13 from `layoffs`.`cleaned_layoffs_data` (2)
    14 group by Region
    15 order by total_layoffs2 desc;
  
```

Region	total_layoffs2
North America	424334
Other	84002
Asia	51234
Europe	46727
Middle East	6266
Ocean	

Q3- Which region has the most layoffs?

The dataset shows the most affected region and

The answer is North America

```

    14 group by Region
    15 order by total_layoffs2 desc;
    16 • select Company,
    17 sum(Laid_Off_Count) as total_layoffs3
    18 from `layoffs`.`cleaned_layoffs_data` (2)
    19 group by Company
    20 order by total_layoffs3 desc limit 10;
  
```

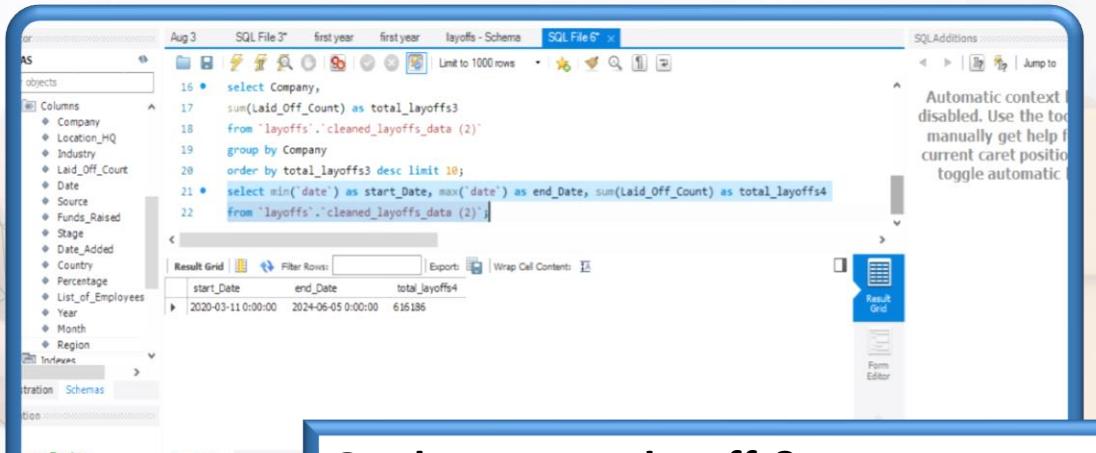
Company	total_layoffs3
amazon	27840
meta	21000
tesla	14500
microsoft	14058

Q4- Which companies is the most affected?

The dataset shows the most affected companies and

The answer is Amazon with total '27840' layoffs.

MySQL Exploratory Data Analysis



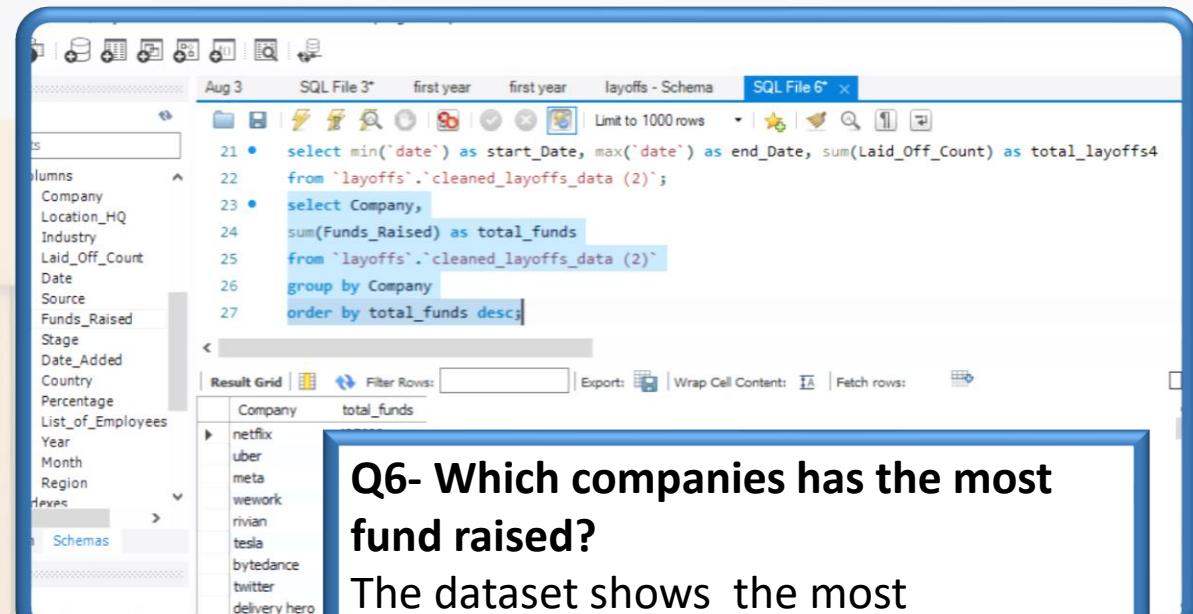
```

16 • select Company,
17   sum(Laid_Off_Count) as total_layoffs3
18   from `layoffs`.`cleaned_layoffs_data` (2)
19   group by Company
20   order by total_layoffs3 desc limit 10;
21 • select min(`date`) as start_Date, max(`date`) as end_Date, sum(Laid_Off_Count) as total_layoffs4
22   from `layoffs`.`cleaned_layoffs_data` (2);
  
```

start_Date	end_Date	total_layoffs4
2020-03-11 00:00:00	2024-06-05 00:00:00	616186

Q5- how many layoffs?

The dataset shows the total layoffs during a specific period and The answer is 616186 between 11/3/2020 till 5/6/2024.



```

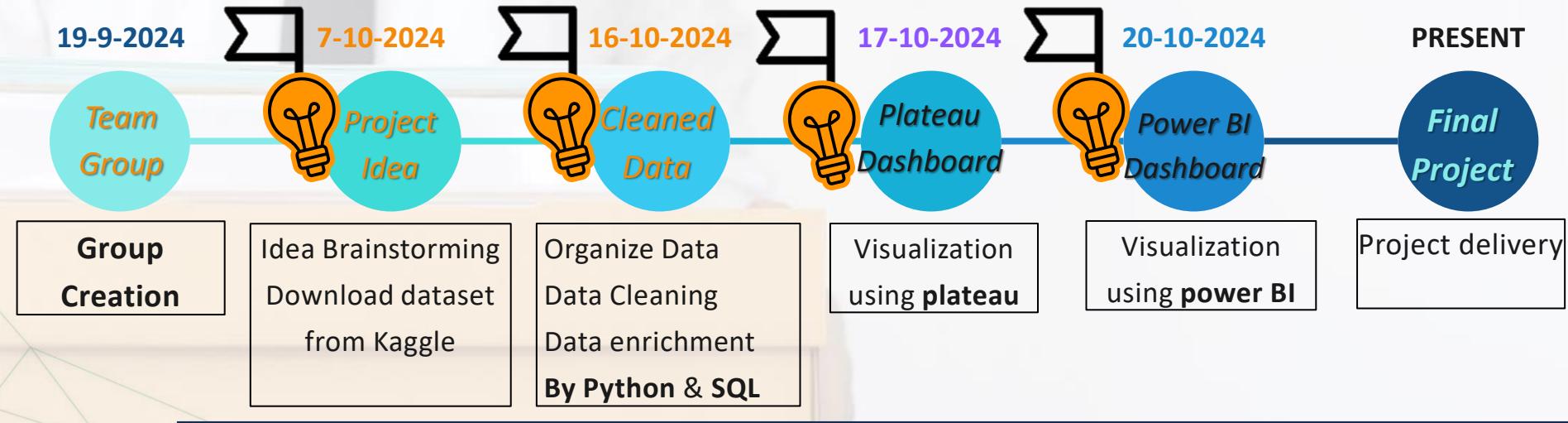
21 • select min(`date`) as start_Date, max(`date`) as end_Date, sum(Laid_Off_Count) as total_layoffs4
22   from `layoffs`.`cleaned_layoffs_data` (2);
23 • select Company,
24   sum(Funds_Raised) as total_funds
25   from `layoffs`.`cleaned_layoffs_data` (2)
26   group by Company
27   order by total_funds desc;
  
```

Company	total_funds
netflix	
uber	
meta	
wework	
rivian	
tesla	
bytedance	
twitter	
delivery hero	

Q6- Which companies has the most fund raised?

The dataset shows the most companies has fund raised and The answer is Netflix with total '487600' of fund raised.

TIMELINE FOR DELIVERABLES



Key responsibilities	Team member
SQL code (data preparation: cleaning & enrichment)	Zainab Kamal
Python code (data preparation: cleaning & enrichment)	Anan Gamal
Plateau Dashboard	Merhan Hassan
Power BI Dashboard, SQL EDA, Presentation	Dalia Abbas



THANK YOU

**Anan Gamal Mohammed Mohammed
Dalia Abbas Ahmed Mohammed Shaltout
Merhan Hassan Mohammed Elsayaby
Zainab Kamal Kamel Maamoun**