SQL- LAYOFF PROJECT

By Mukesh Bhatt
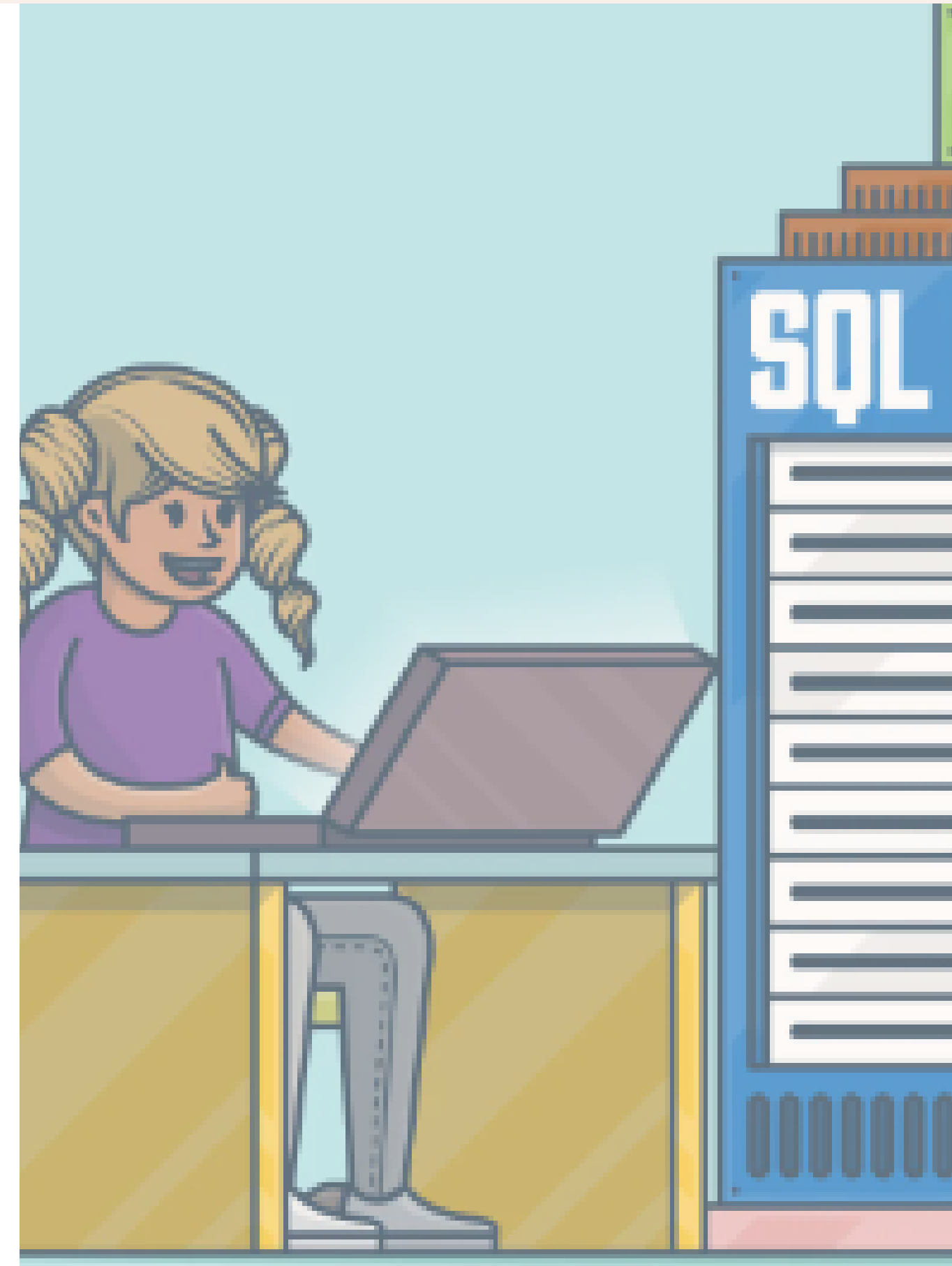
PROJECT GOAL

I am an aspiring Data Analyst working with SQL on a layoff dataset project. The goal was to clean the data by:

1.   Removing duplicates and creating a new table to preserve raw data.
2.   Standardizing and fixing data errors.
3.   Handling null values appropriately.
4.   Removing unnecessary columns and rows.

This process ensured the dataset was ready for analysis.

```sql
CREATE TABLE `layoffs_staging2` (
    `company` TEXT,
    `location` TEXT,
    `industry` TEXT,
    `total_laid_off` INT DEFAULT NULL,
    `percentage_laid_off` TEXT,
    `date` TEXT,
    `stage` TEXT,
    `country` TEXT,
    `funds_raised_millions` INT DEFAULT NULL,
    `row_num` INT
)  ENGINE=INNODB DEFAULT CHARSET=UTF8MB4 COLLATE = UTF8MB4_0900_AI_CI;


SELECT
    *
FROM
    layoffs_staging2;


insert into layoffs_staging2
select * ,
row_number() over(
partition by company , location, industry , total_laid_off, percentage_laid_off, `date` ,stage, country
, funds_raised_millions  ) as row_num
from layoffs_staging ;
```

CHECK FOR DUPLICATES AND REMOVE

```sql
(select * ,
row_number() over(
partition by company , location, industry ,
total_laid_off, percentage_laid_off, `date` ,
stage, country
, funds_raised_millions  ) as row_num
from layoffs_staging)


select * from duplicate_cte
where row_num > 1;
```


| Result Grid | | | Filter Rows: | | | Export: | | | Wrap Cell Content: | | |

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions	row_num
Casper	New York City	Retail	NULL	NULL	9/14/2021	Post-IPO	United States	339	2
Cazoo	London	Transportation	750	0.15	6/7/2022	Post-IPO	United Kingdom	2000	2
Hibob	Tel Aviv	HR	70	0.3	3/30/2020	Series A	Israel	45	2
Wildlife Studios	Sao Paulo	Consumer	300	0.2	11/28/2022	Unknown	Brazil	260	2
Yahoo	SF Bay Area	Consumer	1600	0.2	2/9/2023	Acquired	United States	6	2

STANDARDIZE DATA AND FIX ERRORS

```sql
SELECT DISTINCT
    country, TRIM(TRAILING '.' FROM country)
FROM
    layoffs_staging2;


UPDATE layoffs_staging2
SET
    country = TRIM(TRAILING '.' FROM country)
WHERE
    country LIKE 'united%states%';


-- text to date format :
alter table `layoffs_staging2`
modify column `date` date;


SELECT
    `date`
FROM
    layoffs_staging2;
```
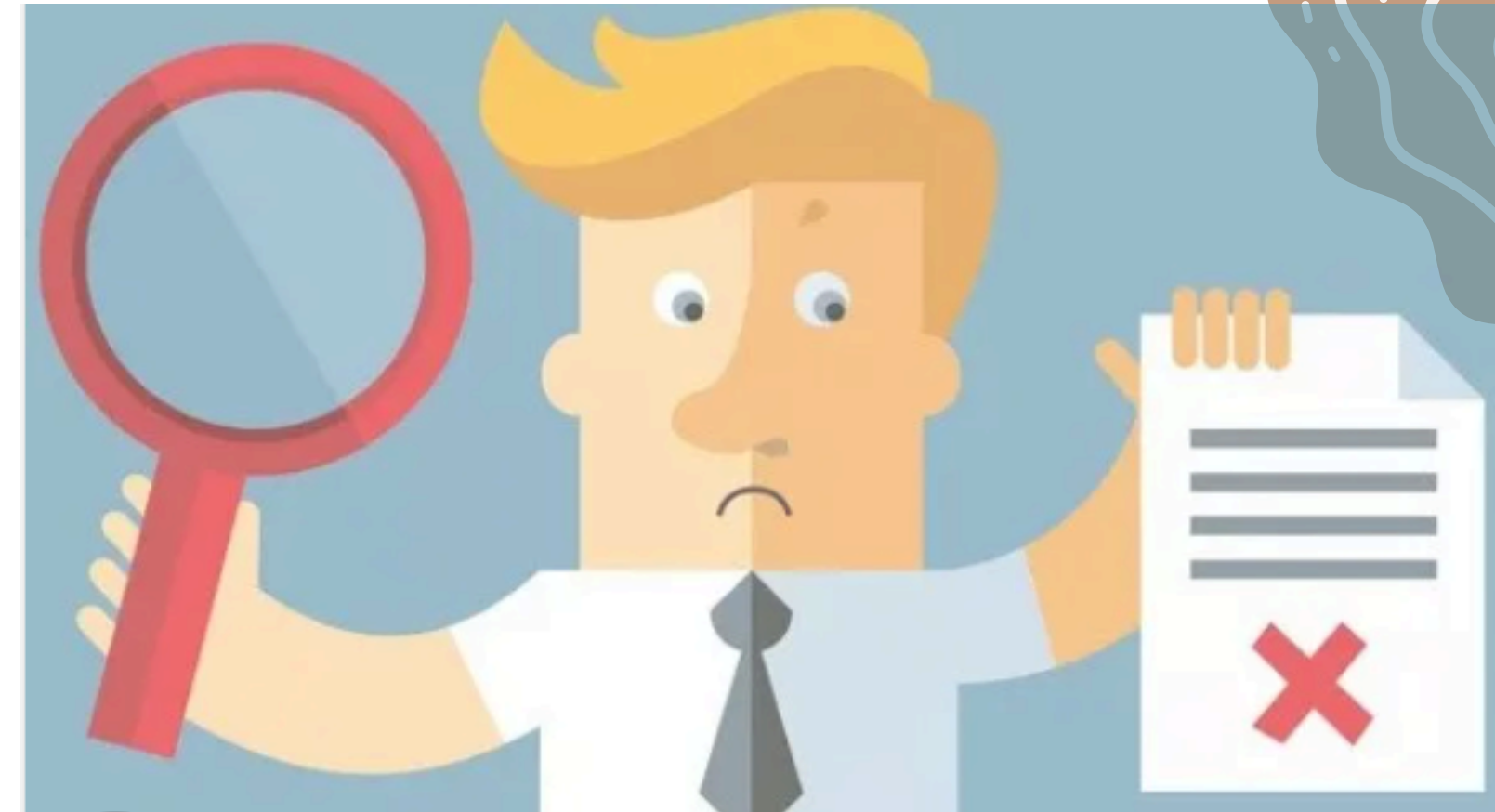
```sql
UPDATE layoffs_staging2
SET
    company = TRIM(company);


UPDATE layoffs_staging2
SET
    industry = 'Crypto '
WHERE
    industry LIKE ' Crypto%';


SELECT
    industry
FROM
    layoffs_staging2;


SELECT DISTINCT
    country, TRIM(TRAILING '.' FROM country)
FROM
    layoffs_staging2;
```

company	industry	date
Included Health	Healthcare	2022-07-25
&Open	Marketing	2022-11-17
#Paid	Marketing	2023-01-27
100 Thieves	Consumer	2022-07-13
10X Genomics	Healthcare	2022-08-04
1stdibs	Retail	2020-04-02
2TM	Crypto	2022-06-01
2TM	Crypto	2022-09-01
2U	Education	2022-07-28
54gene	Healthcare	2022-08-29
5B Solar	Energy	2022-06-03
6sense	Sales	2022-10-12
80 Acres Farms	Food	2023-01-18

LOOK AT NULL VALUES AND FIX

```sql
select *
 from layoffs_staging2
where total_laid_off is null and
percentage_laid_off is null;


update layoffs_staging2
set industry = 'Travel'
where company like 'Airbnb%';


select * from layoffs_staging2
where company = 'airbnb';


delete
 from layoffs_staging2
where total_laid_off is null and
percentage_laid_off is null;


select * from layoffs_staging2;
alter table  layoffs_staging2
drop  row_num ;
```

	company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_
•									

ANALYSIS

THE DATASET CONTAINS 2,361 RECORDS WITH 9 COLUMNS, INCLUDING COMPANY NAMES, LOCATIONS, INDUSTRY, LAYOFFS DATA, AND FUNDING INFORMATION. WHERE WE CLEAN THE DATA , WORK IN NULL AND BLANK VALUE AND REPRESENT DATA TO PERFORM EDA .

THANK YOU