

JOB DATABASE FOR ANALYSTS' JOBS

The objective of this project is to create a database for job finders who are looking for any kind of analyst job roles, select from which job sites they will apply for that job. The database will consist of data from the most visited job sites like Indeed, Glassdoor, Monster etc., with columns Job Id, Job role, Job Description, Skills, Company, Location, Job site (link to the site), twitter info of the job post and company, twitter handles of the companies, most used hashtags, reviews, and ratings of the companies.

Using BeautifulSoup as well as Selenium in python, we scrape the necessary data required and the cleaning and munging of the data is done in Jupyter notebook. Later this cleaned data is converted into ‘.csv’ format and stored in SQL. By then storing this data in a SQL table we can perform the necessary operations such as Searching and Filtering the data.

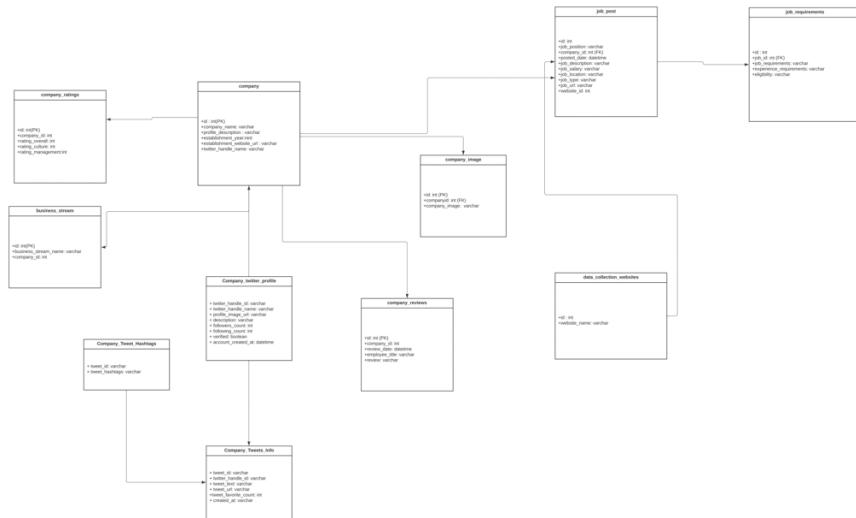


Figure 1: Class Diagram Diagram

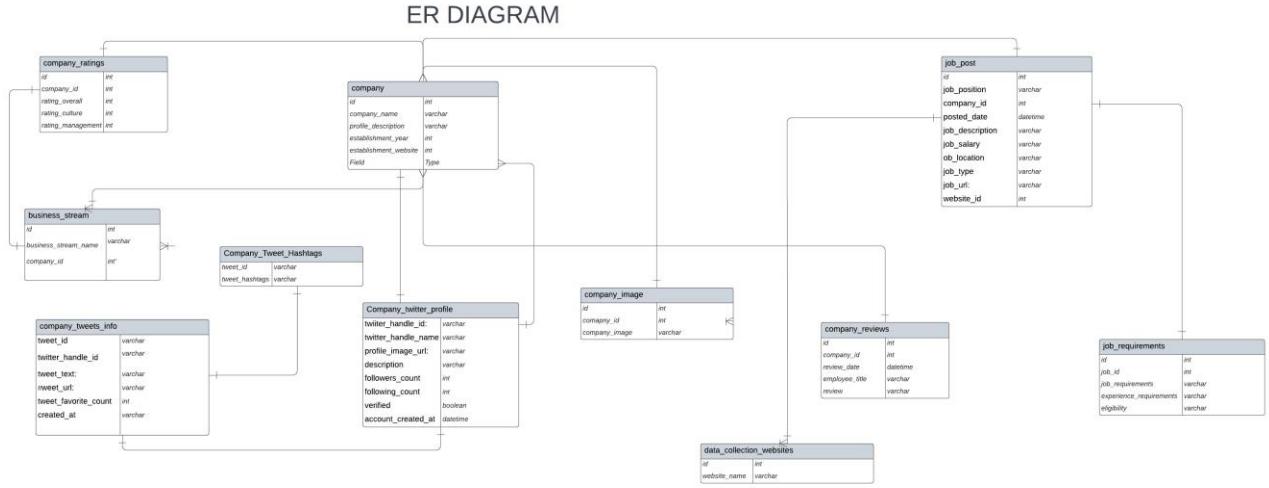


Figure 2: ER Diagram

Create Table commands for initial schema:

- CREATE TABLE company_ratings(id int AUTO_INCREMENT, company_id int, rating_overall int, rating_culture int, rating_management int, PRIMARY KEY(ID));

```

15
16 • select * from company_ratings;

```

00% 31:16

Result Grid Filter Rows: Search Edit: Export/Import: Fetch rows:

id company_id rating_overall rating_cultu... rating_managem...				
3	34	5	5	5
4	34	4	4	2
5	34	4	4	4
6	34	4	5	4
7	34	4	4	3

- CREATE TABLE business_stream(id int AUTO_INCREMENT, business_stream_name varchar(200), company_id int, PRIMARY KEY(id))
- CREATE TABLE company_image(id int AUTO_INCREMENT, company_id int, company_image varchar(1000), PRIMARY KEY(id));

```
15
16 • select * from business_stream;
```

00% 31:16

Result Grid Filter Rows: Search Edit: Export/Import:

id	business_stream_name	company_id
5	Venture Capital and Private Equity Principals	5
6	Retail Building Materials and Garden Equipment	2
7	Renewable Energy Semiconductor Manufacturi...	3
8	Information Services, Staffing and Recruiting, a...	4
9	IT Services and IT Consulting and Financial Ser...	4
10	Software Development	6

3. CREATE TABLE company_tweet_hashtags(tweet_id varchar(200),hashtags varchar(250));

```
14 • select * from company_tweet_hashtags;
```

00% 38:14

Result Grid Filter Rows: Search Export: Fetch rows:

tweet_id	hashtags
456741862833799168	MayDay
456741862833799168	TimeforChange
442265466942533632	InspireChange
286334698077188096	1
1591186428205359104	AI
1591135703223906304	VeteransDay

4. CREATE TABLE company_tweets_info(tweet_id varchar(200),twitter_handle_id varchar(200), tweet_text varchar(1000), tweet_urls varchar(1000), tweet_favorite_count int, created_at datetime);

```
15 • select * from company_tweets_info;
```

0% 35:15

Result Grid Filter Rows: Search Export: Fetch rows:

tweet_id	twitter_handle_id	tweet_text	tweet_urls	tweet_favorite_co...	created_at
529004023717112858	108203954	Here's Why Good Employees Quit http://t.co/0d9gMEIR	http://t.co/0d9gMEIR	17	2014-10-17 0
505372247557111808	108203954	Finding Happiness at Work. All by Yourself http://t.co/oVYJCHoHd	http://t.co/oVYJCHoHd	16	2014-09-01 0
48861289040256864	108203954	Pressure Will Turn You Into Either Dust Or A Diamond http://t.co/WSALFPqD	http://t.co/WSALFPqD	12	2014-07-14 0
482392983770517504	108203954	From shipping to 45 countries in 1995 to over 200 countries today - we've come a long way! http://t.co/dNPnTrjQS	http://t.co/dNPnTrjQS	11	2014-06-27 0
479542950704480256	108203954	Innovation at the next level Amazon Fire Phone it! Watch Video: http://t.co/5jXmrsdy7F4 Read Article... http://t.co/77jAmrXH1	http://t.co/5jXmrsdy7F4	9	2014-06-19 0

5. CREATE TABLE company_twitter_profile(twitter_handle_id varchar(200), twitter_handle_name varchar(100), profile_image_url varchar(1000), description varchar(350), followers_count int, following_count int, verified boolean, account_created_at datetime);

```
15 • select * from company_twitter_profile;
```

The screenshot shows the MySQL Workbench interface with a result grid. The table has columns: twitter_handle_name, twitter_handle_name, profile_image_url, description, followers_count, following_companies, verified, and account_created_at. The data includes entries for Netflix, FidelityJobs, AccentureJobsIN, and others.

twitter_handle_name	twitter_handle_name	profile_image_url	description	followers_count	following_companies	verified	account_created_at
16573941	netflix	http://pbs.twimg.com/profile_images/1235992718171467776/PaX2Bz1S_normal.jpg	The Lohanaissance is upon us!	21121308	2176	1	2008-10-03 04:16:17
16573941	netflix	http://pbs.twimg.com/profile_images/1235992718171467776/PaX2Bz1S_normal.jpg	The Lohanaissance is upon us!	21121329	2174	1	2008-10-03 04:16:17
395511307	FidelityJobs	http://pbs.twimg.com/profile_images/1491460562508730368/7x8AxTJ_normal.jpg	Fidelity provides financial expertise to help peop...	9064	835	1	2011-10-21 19:40:39
30839699	AccentureJobsIN	http://pbs.twimg.com/profile_images/1323091380227305472/dwEOIVS_normal.jpg	Find out more about Accenture in India, and abo...	63593	66	1	2009-04-13 09:56:19
10194682	Accenture	http://pbs.twimg.com/profile_images/1552786622936129536/JVw0DfVn_normal.jpg	Together, we deliver on the promise of technolo...	552614	1676	1	2007-11-12 21:34:00

6. CREATE TABLE company_reviews(id int AUTO_INCREMENT, company_id int, review_date datetime, employee_title varchar(200), review varchar(1000), PRIMARY KEY(id));

```
15  
16 • select * from company_reviews;
```

The screenshot shows the MySQL Workbench interface with a result grid. The table has columns: id, company_id, review_date, employee_title, and review. The data includes reviews from employees like Senior Software Engineer and Software Development Engineer.

id	company_id	review_date	employee_title	review
3	34	2019-04-17 00:00:00	Senior Software Engineer	Adopt Rust as C++ successor?
4	34	2019-04-15 00:00:00	NULL	Depends on the team
5	34	2019-04-10 00:00:00	Software Development Engineer	Great place to work
6	34	2019-04-04 00:00:00	Senior Software Engineer	Overall 8.0
7	34	2019-04-05 00:00:00	Senior Software Engineer	Good company to work

7. CREATE TABLE company(id int AUTO_INCREMENT, company_name varchar(100), profile_description varchar(5000), establishment_year int, establishment_website_url varchar(200), twitter_handle_name varchar(200), PRIMARY KEY(id));

```
15  
16 • select * from company;
```

The screenshot shows the MySQL Workbench interface with a result grid. The table has columns: id, company_name, profile_description, establishment_year, establishment_website_url, and twitter_handle_name. The data includes entries for Spencer Ogden, Parabolic Career, Selby Jennings, HireMatch, and Flexon Technologies Inc.

id	company_name	profile_description	establishment_y...	establishment_website_url	twitter_handle_na...
3	Spencer Ogden	The company specializes in providing skilled en...	2010	https://www.spencer-ogden.com/?source=googl...	SpenceOgden
4	Parabolic Career	Parabolic is a career coaching firm for financial...	2020	https://www.parabolicusa.com/	parabolic_llc
5	Selby Jennings	Selby Jennings is a leading specialist recruiter ...	2004	https://www.selbyjennings.com/	Selby_Jennings
6	HireMatch	Hirematch is a private company. The company c...	1905	http://www.hirematch.com	hire_match
7	Flexon Technologies Inc.	Flexon Technologies is a leading end-to-end tec...	2010	https://www.flexontechnologies.com/	FlexonTch

8. CREATE TABLE job_requirements(id int AUTO_INCREMENT, job_id int, job_requirements varchar(10000), experience_requirements varchar(1000), eligibility varchar(1000), PRIMARY KEY(id));

```

15
16 • select * from job_requirements;

```

00% 32:16

Result Grid Filter Rows: Search Edit: Export/Import:

id	job_id	job_requirements	experience_requireme...	eligibility
3	205	- Identify gaps in knowledge and overseeing inf... < 1 Year	< 1 Year	Everyone
4	206	- Rendering technical assistance to Database M... Job for Fresh Graduates	All Interested Candidates	
5	207	- Coordinating the work of subordinate employees; < 1 Year	Applicants must hold Armenian	
6	208	- Assist the Tavush Marz communities and com... 5 Years Job Exp.	Residence & citizenship - Tbilisi,	
7	209	- Manage and control the company's activities in... 3 Years Job Exp.	All Interested Candidates	

9. CREATE TABLE data_collection_websites(id int AUTO_INCREMENT, website_name varchar(100), PRIMARY KEY(id));

```

15
16 • select * from data_collection_websites;

```

00% 40:16

Result Grid Filter Rows: Search Edit:

id	website_name
1	LinkedIn
2	Glassdoor
3	Indeed
4	Monster
NULL	NULL

10. CREATE TABLE job_post(id int AUTO_INCREMENT, job_position varchar(200), company_id int, posted_date date, job_description text(50000), job_location varchar(200), job_salary varchar(300), job_type varchar(100), job_url varchar(1000), website_id int, PRIMARY KEY(id));

```

15
16 • select * from job_post;

```

00% 24:16

Result Grid Filter Rows: Search Edit: Export/Import:

job_position	company_id	posted_date	job_description	job_location	job_salary	job_type	job_url	website_id
Investment Banking Analyst and Associate I Ind...	3	2022-12-02	**Title: Investment Banking Analyst or Investme...	Denver, CO	NULL	Full-time	https://uk.linkedin.com/company/selby-jennings... 1	
Investment Banking Analyst and Associate I Ind...	4	2022-12-02	**Title: Investment Banking Analyst or Investme...	Denver, CO	NULL	Full-time	https://uk.linkedin.com/company/selby-jennings... 1	
Private Equity Analyst / Associate	5	2022-11-29	Selby Jennings has been engaged by a top-tier...	Austin, TX	NULL	Full-time	https://uk.linkedin.com/company/selby-jennings... 1	
Data Science Analyst (Remote)	6	2022-11-13	Position Purpose The Data Science Analyst i...	Atlanta, GA	NULL	Full-time	https://www.linkedin.com/company/the-home-de... 1	
Analyst/Associate	7	2022-11-30	I'm working with an industry-leading renewable...	New York, NY	NULL	Full-time	https://uk.linkedin.com/company/spencerrogden... 1	

Sources of data:

To obtain relevant and high-quality data we choose to scrape data from websites like LinkedIn, Twitter, Indeed, Monster, etc.

The data gathered from these websites is scraped using python libraries like BeautifulSoup and Selenium.

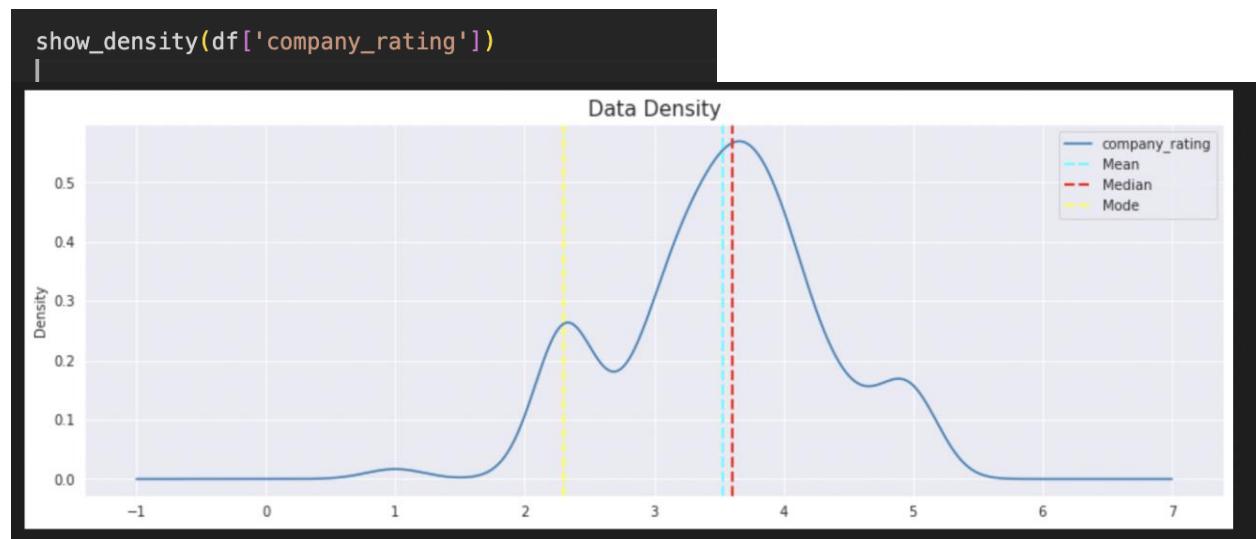
Also, data used in the database is collected and downloaded from data repositories and websites like ‘kaggle.com’.

https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Assignment_3/Glassdoor_WebScraping_BS.ipynb

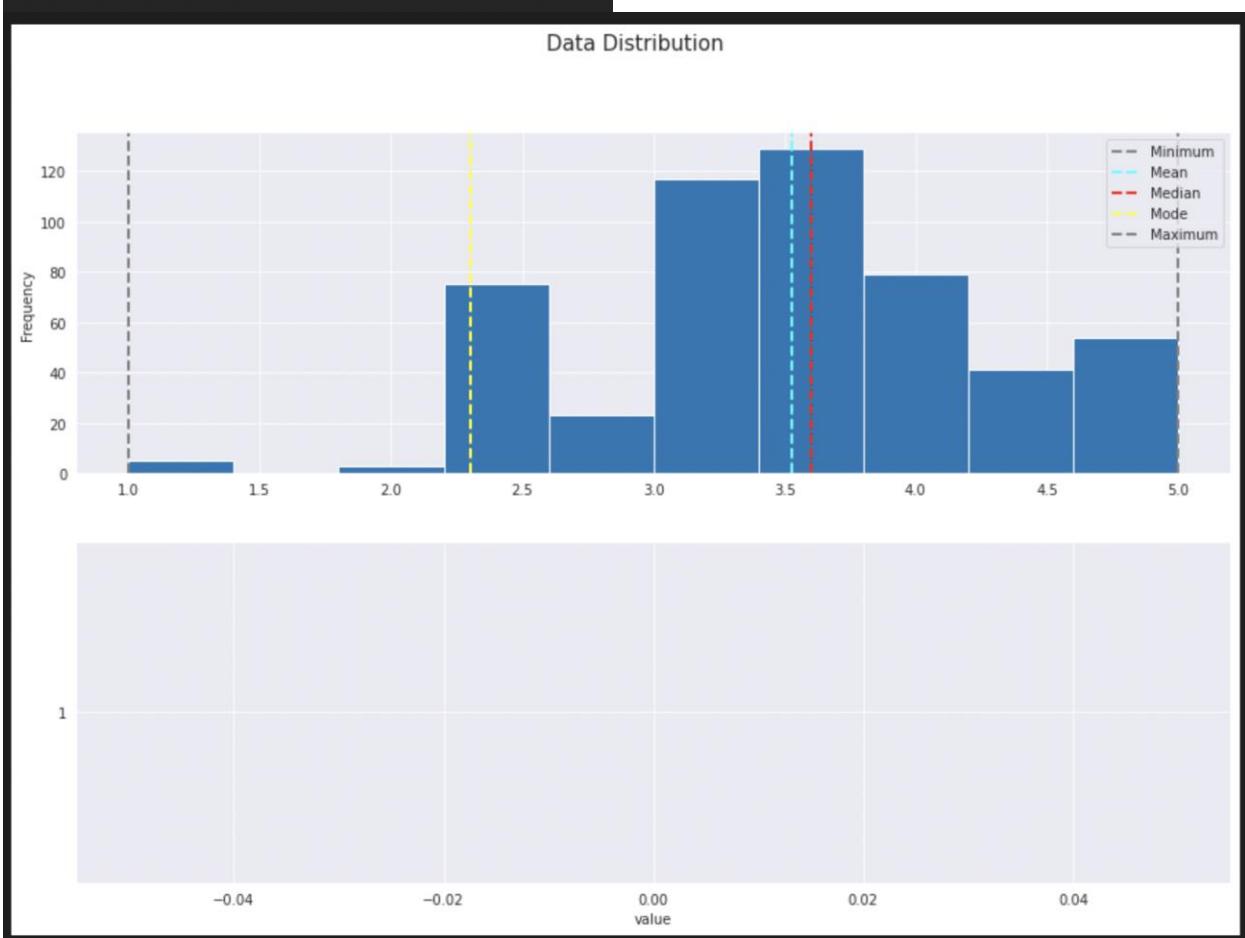
https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Assignment_3/LinkedIn_Scraping.ipynb

Data Visualization:

Further the data is visualized based on various factors like data-density and frequency:



```
show_distribution(df['company_rating'])
```

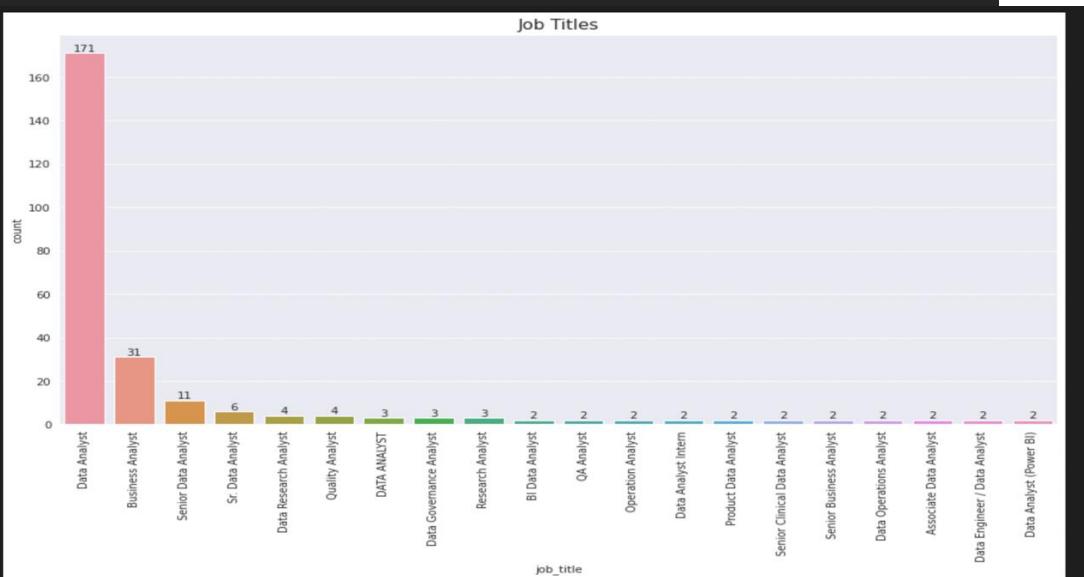


```

plt.figure(figsize = (15,8))
sns.countplot(x = 'job_title', data = df, order = df['job_title'].value_counts()[:20].index)
plt.title('Job Titles', fontsize = 15)
plt.xticks(rotation=90)

for index, value in enumerate(df['job_title'].value_counts()[:20].values):
    plt.text(index,value,str(value), ha = 'center', va= 'bottom')

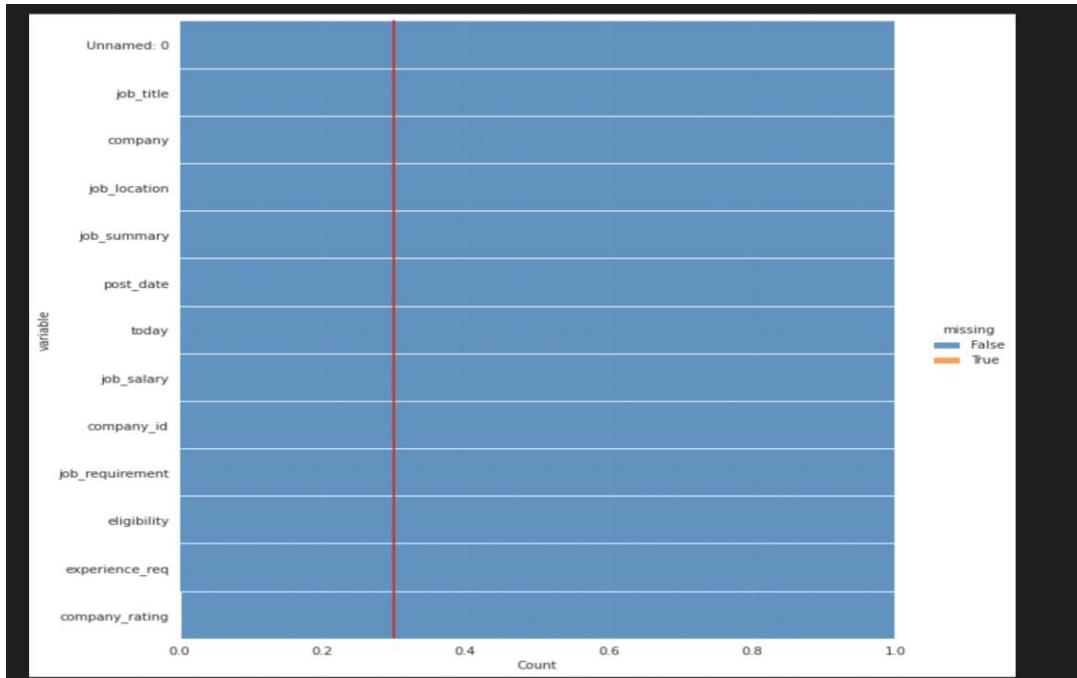
```



```

plt.figure(figsize = (10, 5))
sns.displot(
    data = df.isnull().melt(value_name = "missing"),
    y = "variable",
    hue = "missing",
    multiple = "fill",
    height = 9.25)
plt.axvline(0.3, color = "r")
plt.show()

```



Data Validation:

The data gathered is checked for its validity which includes three factors:

Completeness: refers to the extent to which an entity contains the information needed to describe a real-world object. The presence of null values, which are typically regarded as missing values, in tables in relational database systems can be used to determine how full a table is.

Consistency: The degree to which a set of semantic rules are violated such as a specific data type, an interval for a numerical column, or a set of values for a categorical column.

Accuracy: The correctness of the data and can be measured in two dimensions: syntactic and semantic. Semantic accuracy contrasts a value with its actual representation, while syntactic accuracy compares a value's representation with a domain of definition that corresponds.

Use Cases:

Based on the data gathered from various sources and the goal of the database we created UseCases and wrote SQL queries and Relational Algebra for the same.

Below are some of the examples:

1. **Use Case:** Select the company with most followers

Description: Companies with the maximum followers get listed

Actor: Follower

Steps:

Actor action: job seekers request the companies with the highest followers

System Responses: the companies with most followers get displayed and use case ends.

2. **Use Case:** Select the company with most likes in tweets

Description: Companies with the most likes in tweets get listed

Actor: Job Seeker

Steps:

Actor action: Job Seekers like the tweets posted by Companies

System Responses: The companies with most likes in tweets get listed and use case ends.

3. **Use Case:** The most active company in twitter

Description: the company that is the most active in twitter gets listed

Actor: employee

Steps:

Actor action: employees keep posting tweets and follow up regularly

System Responses: the company whose employees keep the company's page active on twitter is listed to be the most active and use case ends.

4. **Use Case:** Verify for a company on twitter

Description: A company verifies itself on twitter

Actor: employee

Pre-Condition: For a company to verify itself on twitter, it must first be registered.

Steps:

Actor action: employee provides a company to check if it's verified or not.

System Responses: The company that is verified on twitter gets displayed.

Post Condition: The Company gets verified

5. **Use Case:** List top 3 companies according to popular hashtags

Description: the first 3 companies with popular hashtags gets displayed

Actor: job seeker

Steps:

Actor action: job seeker asks to list the company's top 3 popular hashtags

System Responses: top 3 companies with hashtags get displayed.

6. **Use Case:** Salary for Data Analyst with a graduate degree

Description: the salary amount for a data analyst holding a graduate degree gets listed

Actor: Job Seeker

Precondition: For the salary to be listed, first the job seeker must hold a degree

Steps:

Actor action: the job seeker asks for the salary of a data analyst with a graduate degree

System Responses: data analyst's salary with a graduate degree is listed

7. **Use Case:** companies with ratings greater than 4

Description: those companies which have ratings greater than 4 are listed.

Actor: Job Seeker

Steps:

Actor action: the job seeker asks for the company's ratings greater than 4

System Responses: companies whose ratings are greater than 4 are listed.

8. **Use Case:** salaries of companies in Boston

Description: Salaries of those companies located in Boston are listed.

Actor: Job Seeker

Steps:

Actor action: the job seeker asks for the salaries of all the Boston-located companies.

System Responses: salaries of those companies located in Boston are listed.

9. Use Case: salary for Data Analyst

Description: Salaries of all data analysts are listed.

Actor: Job Seeker

Steps:

Actor action: the job seeker asks for the salaries of all the Data Analysts.

System Responses: salaries of all Data Analysts are listed.

10. Use Case: least paying job in California

Description: jobs in California which pay the least are listed.

Actor: Job Seeker

Steps:

Actor action: the job seeker asks for all the least paying jobs in California.

System Responses: All the jobs which pay the least in California are listed.

1. Use Case: companies with most job posting

Description: all the companies having the most job postings are listed

Actor: Job Seeker

Steps:

Actor action: the job seeker asks for all the companies having the maximum job postings.

System Responses: All the companies with maximum job postings are listed.

2. Use Case: List of all the internships

Description: all the internships are listed

Actor: Job Seeker

Steps:

Actor action: the job seeker asks for all the internships.

System Responses: All the internships are listed.

3. **Use Case:** average salary of data analyst

Description: Data Analyst's average salary is listed

Actor: Job Seeker

Steps:

Actor action: the job seeker asks for the average salary of Data Analyst.

System Responses: All the Data Analyst's average salary is listed.

4. **Use Case:** Average salary in Boston

Description: average salary in Boston is listed

Actor: Job Seeker

Steps:

Actor action: the job seeker asks all the average salary in Boston.

System Responses: All the average salary in Boston gets listed.

5. **Use Case:** companies with financial analyst jobs

Description: all the companies having financial analyst job roles get listed

Actor: Job Seeker

Steps:

Actor action: the job seeker asks all the companies with financial analyst roles.

System Responses: All the financial analyst job roles get listed.

6. **Use Case:** get the company twitter profile name for fidelity

Description: the twitter profile name for the company fidelity gets listed

Actor: Job Seeker

Steps:

Actor action: the job seeker gets the twitter profile name of the company fidelity

System Responses: the twitter profile name of fidelity gets listed

7.

8. **Use Case:** the company provides medical insurance for a particular job

Description: if there is medical insurance of a particular job by any company

Actor: Job Seeker

Steps:

Actor action: the job seeker checks if there is any medical insurance supported by a company for any job

System Responses: if any medical insurance is provided for a particular job for a company.

9. **Use Case:** the company sponsors international applicants

Description: if the company sponsors international applicants

Actor: Job Seeker

Steps:

Actor action: the job seeker checks if any company sponsors international applicants

System Responses: if any international applicants are sponsored by the company.

10. **Use Case:** review of a particular role for a particular company

Description: the review of a particular role for any given company gets displayed.

Actor: Job Seeker

Steps:

Actor action: the required job role of the given project gets displayed whenever the job seeker asks for it.

System Responses: the review of the given company gets listed.

11. **Use Case:** details of the remote data analyst roles

Description: the details of the data analyst roles in remote gets displayed.

Actor: Job Seeker

Steps:

Actor action: the data analyst roles' details in remote are displayed whenever the job seeker asks for it.

System Responses: the details of the job role get listed.

Data Normalization:

The data which is already cleaned, munged, and validated is further checked to see if it is in a normalized form and satisfies the conditions for 1NF, 2NF and 3NF.

If not, we made changes in those tables to convert them into a normalized format.

For the tables 'company', 'company_ratings' and 'company_twitter_profile' we can see that these tables are already in 2NF and there is no partial dependencies. Below are the screen shots for the said tables. This normalized data was achieved through proper cleaning and munging of data.

'company_table'

```
[ ] img_path = "/content/Screen Shot 2022-12-13 at 9.22.14 PM.png"
image = io.imread(img_path)
cv2.imshow(image)
```

Data output Messages Notifications

	id [PK] integer	company_name character varying (100)	profile_description character varying (5000)	establishment_year integer	establishment_website_url character varying (200)	twitter_handle_name character varying (200)
1	123	Amazon	Amazon.com Inc. is an ...	1994	https://www.amazon.com/	amazon
2	445	Google	Google LLC is an Americ...	1998	https://www.google.com/	Google

Figure 3: Normalization 1NF

Both the tables job posts and twitter hashtags tables had multi-valued columns and to attain 1NF the multivalued tables were separated.

To be in second normal form, a relation must be in first normal form and relation must not contain any partial dependency. A relation is in 2NF if it has No Partial Dependency, i.e., no non-prime attribute is dependent on any proper subset of any candidate key of the table.

For the tables 'company', 'company_ratings' and 'company_twitter_profile' we can see that these tables are already in 2NF and there is no partial dependencies. Below are the screen shots for the said tables. This normalized data was achieved through removing repeating groups.

Data output Messages Notifications

The screenshot shows a database interface with a toolbar at the top containing icons for new, open, save, delete, export, and search. Below the toolbar is a table with the following columns: id [PK] integer, company_name character varying (100), profile_description character varying (5000), establishment_year integer, establishment_website_url character varying (200), and twitter_handle_name character varying (200). The data consists of 10 rows:

	id [PK] integer	company_name character varying (100)	profile_description character varying (5000)	establishment_year integer	establishment_website_url character varying (200)	twitter_handle_name character varying (200)
1	123	Amazon	Amazon.com Inc. is an ...	1994	https://www.amazon.com/	amazon
2	445	Google	Google LLC is an Americ...	1998	https://www.google.com/	Google
3	724	Accenture	Accenture plc is an Irish...	1989	https://www.accenture.co...	Accenture
4	990	AdTech Holding	AdTech Holding establis...	1998	https://adtechholding.com/	Adtech_Systems
5	887	Maersk	An integrated transport ...	1996	https://www.maersk.com/	Maersk
6	444	Honeywell	Honeywell International ...	1906	https://www.honeywell.co...	honeywell
7	886	Financial Fabric	delivers trades, position...	2014	https://www.financialfabri...	FinancialFabric
8	109	Brickendon Consulting	We are transformation s...	2010	https://www.brickendon.c...	BrickendonIntl
9	665	PayPal	PayPal is the faster, saf...	1998	https://www.paypal.com/	Paypal
10	545	Northern Trust Corp.	is a leading provider of ...	1889	https://www.northerentrust....	NorthernTrust

Figure 4: Normalization 2NF

Company Rating Table

The screenshot shows a database interface with a toolbar at the top containing icons for new, open, save, delete, export, and search. Below the toolbar is a table with the following columns: id [PK] integer, company_id integer, rating_overall integer, rating_culture integer, and rating_management integer. The data consists of 10 rows:

	id [PK] integer	company_id integer	rating_overall integer	rating_culture integer	rating_management integer
1	109	34	5	3	5
2	123	34	4	5	5
3	444	34	5	4	5
4	445	34	5	4	3
5	545	34	5	5	5
6	665	34	4	3	5
7	724	34	4	4	5
8	886	34	5	4	5
9	887	34	5	5	3
10	990	34	5	5	4

Figure 5: Normalization 2NF

Company Twitter Profile Table

0)	twitter_handle_name character varying (100)	profile_image_url character varying (1000)	description character varying (350)	followers_count integer	following_count integer	verified boolean	account_created_at timestamp with time zone
1	Google	https://upload.wikimedi...	The worlds strongest n...	345	86	true	2012-10-20 12:12:12-04
2	Google	https://upload.wikimedi...	The worlds strongest n...	345	86	false	2012-10-20 12:12:12-04
3	Amazon	https://upload.wikimedi...	The worlds strongest n...	335	26	true	2012-10-10 12:11:12-04
4	Amazon	https://upload.wikimedi...	The worlds strongest n...	335	26	true	2012-10-10 12:11:12-04
5	Netflix	https://upload.wikimedi...	The worlds strongest n...	322	11	false	2012-10-11 12:09:12-04
6	Amazon	https://upload.wikimedi...	The worlds strongest n...	335	26	true	2012-10-10 12:11:12-04
7	Amazon	https://upload.wikimedi...	The worlds strongest n...	335	26	true	2012-10-10 12:11:12-04
8	Netflix	https://upload.wikimedi...	The worlds strongest n...	322	11	false	2012-10-11 12:09:12-04
9	Amazon	https://upload.wikimedi...	The worlds strongest n...	335	26	true	2012-10-10 12:11:12-04
10	Amazon	https://upload.wikimedi...	The worlds strongest n...	335	26	true	2012-10-10 12:11:12-04
11	Netflix	https://upload.wikimedi...	The worlds strongest n...	322	11	false	2012-10-11 12:09:12-04

Figure 6: Normalization 2NF

NORMALIZATION- 3NF In order to satisfy 3NF, a table ,must be satisfied with 1NF and 2NF, and must be free from transitive dependencies. Whenever some indirect relationship happens to cause functional dependency (FD), it is known as Transitive Dependency. Thus, if $A \rightarrow B$ and $B \rightarrow C$ are true, then $A \rightarrow C$ happens to be a transitive dependency. Thus, to achieve 3NF, one must eliminate the Transitive Dependency.

twitter_handle_name	followers_count	following_count
Google	345	86
Amazon	335	26
Amazon	335	26
Netflix	322	11
Amazon	335	26
Amazon	335	26
Netflix	322	11
Amazon	335	26
Amazon	335	26
Netflix	322	11

10 rows in set (0.00 sec)

Figure 7: Normalization 3NF

for the table `company_twitter_profile`, there is transitive dependency in the columns `twitter_handle_name`, `followers_count`, `following_count` and hence the table is split into 2 tables as above, in order to satisfy 3NF.

Data output Messages Notifications

	id [PK] integer	company_name character varying (100)	profile_description character varying (5000)	establishment_year integer	establishment_website_url character varying (200)	twitter_handle_name character varying (200)
1	123	Amazon	Amazon.com Inc. is an ...	1994	https://www.amazon.com/	amazon
2	445	Google	Google LLC is an Americ...	1998	https://www.google.com/	Google
3	724	Accenture	Accenture plc is an Irish...	1989	https://www.accenture.co...	Accenture
4	990	AdTech Holding	AdTech Holding establis...	1998	https://adtechholding.com/	Adtech_Systems
5	887	Maersk	An integrated transport ...	1996	https://www.maersk.com/	Maersk
6	444	Honeywell	Honeywell International ...	1906	https://www.honeywell.co...	honeywell
7	886	Financial Fabric	delivers trades, position...	2014	https://www.financialfabri...	FinancialFabric
8	109	Brickendon Consulting	We are transformation s...	2010	https://www.brickendon.c...	BrickendonIntl
9	665	PayPal	PayPal is the faster, saf...	1998	https://www.paypal.com/	Paypal
10	545	Northern Trust Corp.	is a leading provider of ...	1889	https://www.northerntrust....	NorthernTrust

Figure 8: Normalization 3NF

For the table "company" we can see that it is already in 3NF and there is no transitive dependencies. Above are the screen shots for the said tables. This normalized data was achieved through proper cleaning and munging of data.

Company Ratings Table

	id [PK] integer	company_id integer	rating_overall integer	rating_culture integer	rating_management integer	
1	109	34	5	3	5	
2	123	34	4	5	5	
3	444	34	5	4	5	
4	445	34	5	4	3	
5	545	34	5	5	5	
6	665	34	4	3	5	
7	724	34	4	4	5	
8	886	34	5	4	5	
9	887	34	5	5	3	
10	990	34	5	5	4	

For the tables 'company ratings' we can see that the table are already in 3NF and there is no Transitive dependencies.

OUTPUT OF VIEWS:

1 Creating and calling View for company ratings where the overall rating is more than 4.

```
In [ ]: CREATE VIEW company_ratings_above_4 AS
SELECT DISTINCT c.company_name,cr.rating_overall FROM company_ratings cr inner join company c on cr.company_id = c.id WHERE
cr.rating_overall > 4;

12
13
14 •  select * from company_ratings_above_4;
15
```

100% 39:14 |

Result Grid Filter Rows: Search Export:

company_name	rating_overall
Bloomberg	5
HireMatch	5
Spencer Ogden	5
Parabolic Career	5
Santander	5

Action Output | Time | Action | Response

2 Creating and calling View for job_posts_by_company where the company_id's count must be in descending order.

```
In [ ]: CREATE VIEW job_posts_by_company AS
select company_name, count(company_id) from company c inner join job_post jp on c.id = jp.company_id
group by company_name order by count(company_id) desc limit 1;

12
13
14 •  select * from job_posts_by_company;
15
```

100% 36:14 |

Result Grid Filter Rows: Search Export:

company_name	count(company_i...)
Accenture	32

Action Output | Time | Action | Response

3 Create and calling Views for job_intern where the job_position must have the word “intern” in it.

```
In [ ]: CREATE VIEW job_intern AS
select * from job_post where job_position like '%intern%';

12
13
14 •  select * from job_intern;
15
```

Result Grid Filter Rows: Search Export:

id	job_position	company_id	posted_date	job_description	job_location	job_salary	job_type	job_url	website_id
212	Intern - Business Analyst	57	2022-11-21	We believe that machine learning and other soft...	Remote	NULL	HULL	NULL	3
233	Data Analyst Intern	77	2022-12-01	Reorganizing data in a readable format. [? A...	Temporarily Remote in Chennai, Tamil Nadu	₹10,000 - ₹15,000 a month	HULL	HULL	3

job_intern 4

Action Output

Time	Action

Response Duration / Fetch

4 Create and calling Views for rating_management_google where the company name is “Google”.

```
In [ ]: CREATE VIEW rating_mgmt_Google AS
select Distinct c.company_name,cr.rating_management from company_ratings cr inner join company c on cr.company_id = c.id
where c.company_name = "Google";

12
13
14 •  select * from rating_mgmt_Google;
15
```

Result Grid Filter Rows: Search Export:

company_name rating_managem...
► Google NULL
► Google 4
► Google 5

rating_mgmt_Google 6

Action Output

Time	Action

Response Duration / Fetch

5 Create and calling Views for data_analyst_salary where the job_position must have the words “Data Analyst” and the job_salary should not be null.

```
In [ ]: CREATE VIEW data_analyst_salary AS
SELECT job_salary FROM job_post WHERE job_position LIKE '%Data Analyst%' AND job_salary IS NOT NULL;
```

The screenshot shows a database interface with the following details:

- Object Info** tab is selected.
- Session** tab is also visible.
- Tables:**
 - MealTable
 - MealPlanTable
 - RentTable
 - RoomTable
 - StudentTable
 - SysAdminTable
 - test
 - WardenQueryTable
- Columns:**
 - BuildingTable**: Int AI PK
 - buildingName**: varchar(45)
 - wardenAssigned**: varchar(45)
 - lBnkAvail**: int
 - zBhkAvail**: int
 - sharedAvail**: int
 - lBnkDef**: int
- Action Output** section shows two rows of logs:
 - 049 19:41:09 select tweet_text, tweet_favorite_count from company_tweets_info order by tweet_favorite_count desc limit 0, 10; 1 row(s) returned
 - 649 19:41:25 CREATE VIEW popular_tweets AS select tweet_text, tweet_favorite_count from company_tweets_info order by tweet_favorite_count desc limit 0, 10; 1 row(s) affected

6 Create and calling Views for financial_analyst_company where the job_position must have the words “Financial Analyst”.

```
In [ ]: CREATE VIEW financial_analyst_company AS
select c.company_name, jp.job_position from company c inner join job_post jp on c.id=jp.company_id
where job_position like '%Financial Analyst%';
```

The screenshot shows a database interface with the following details:

- Object Info** tab is selected.
- Session** tab is also visible.
- Tables:**
 - company
 - job_post
- Columns:**
 - financial_analyst_company**: (no columns listed)
- Action Output** section shows two rows of logs:
 - 12
 - 13
 - 14 • select * from financial_analyst_company;
 - 15

7 Create and calling Views for highest_follower_count ordered by followers_count in descending order.

```
In [ ]: CREATE VIEW highest_follower_count AS
select distinct twitter_handle_name,followers_count from company_twitter_profile pf order by followers_count desc limit 1;

12
13
14 • select * from highest_follower_count;
15

100% 37:14

Result Grid Filter Rows: Search Export:
|twitter_handle_na...|followers_cou...
> Google 27841687

highest_follower_count 8

Action Output Time Action Response Duration / Fetch
```

8 Create and Calling views for highest_fav_count order by tweet_favorite_count in descending order.

```
In [ ]: CREATE VIEW highest_fav_count AS
select twt.twitter_handle_id, twitter_handle_name, tweet_favorite_count from company_tweets_info twt
inner join company_twitter_profile pf on twt.twitter_handle_id = pf.twitter_handle_id
order by tweet_favorite_count desc limit 1;

12
13
14 • select * from highest_fav_count;
15

100% 32:14

Result Grid Filter Rows: Search Export:
|twitter_handle...|twitter_handle_na...|tweet_favorite_co...
> 16573941 netflix 99174

highest_fav_count 9

Action Output Time Action Response Duration / Fetch
```

9 Create and calling Views for job_salary_boston where the job_location must have the words "Boston" and the job_salary is not null.

```
In [ ]: CREATE VIEW job_salary_Boston AS
SELECT job_salary FROM job_post WHERE job_location LIKE '%Boston%' AND job_salary IS NOT NULL;
```

```
12
13
14 • select * from job_salary_Boston;
15
```

Result Grid		Filter Rows: <input type="text"/> Search	Export:	
job_salary				
► \$1L - \$1L (Glassdoor Est.)				
\$58T - \$86T (Glassdoor Est.)				
\$94T - \$1L (Glassdoor Est.)				
\$90T - \$1L (Employer Est.)				
\$70T - \$95T (Glassdoor Est.)				
job_salary_Boston				
Action Output	⌚			
	Time	Action	Response	Duration / Fetc

10 Create and calling Views for highest_tweet_count ordered by tweet_count descending limited to 1.

```
In [ ]: CREATE VIEW highest_tweet_count AS
select Distinct twt.twitter_handle_id , pf.twitter_handle_name, count(tweet_id) as tweet_count from
company_tweets_info twt inner join company_twitter_profile pf on twt.twitter_handle_id = pf.twitter_handle_id
group by twitter_handle_id, twitter_handle_name order by tweet_count desc limit 1;
```

```
12
13
14 • select * from highest_tweet_count;
15
16
```

Result Grid		Filter Rows: <input type="text"/> Search	Export:	
highest_tweet_count				
► twitter_handle_id twitter_handle_na... tweet_cou...				
108203954	amazoncareers	200		
Action Output	⌚			
	Time	Action	Response	Duration / Fetc

11 Create and calling Views for fidelity_twitter_verified where the twitter_handle_name must be “Fidelity Jobs”.

```
In [ ]: CREATE VIEW Fidelity_twitter_verified AS
select verified from company_twitter_profile
where twitter_handle_name = 'FidelityJobs';
```

```
12
13
14 • select * from Fidelity_twitter_verified;
15
16
```

Result Grid		Filter Rows: <input type="text"/> Search	Export:	
Fidelity_twitter_verified				
► verified				
1				
Action Output	⌚			
	Time	Action	Response	Duration / Fetc

12 Create and calling Views for accenture_twitter_handle where the company_name must have the phrase “Accenture”.

```
In [ ]: CREATE VIEW accenture_twitter_handle AS
    select pf.twitter_handle_name,c.company_name from company_twitter_profile pf inner join
    company c on pf.twitter_handle_name=c.twitter_handle_name
    where c.company_name LIKE '%Accenture%';
```

```
12
13
14 •   select * from accenture_twitter_handle;
15
16
100% 40:14
Result Grid Filter Rows: Search Export:
|twitter_handle_na... |company_name |
|> Accenture      |Accenture|
|                   |
|                   |
|                   |
| accenture_twitter_handle 13 |
Action Output Response Duration / Fetc
| Time | Action |
```

13 Create and calling Views for top3_company_popular_accounts and order by the count of hashtags in descending order limited to 1.

```
In [ ]: CREATE VIEW top3_company_popular_accounts AS
    select hashtags, count(hashtags), twitter_handle_name as popularity_count
    from company_tweet_hashtags ht
    inner join company_tweets_info twt on ht(tweet_id = twt(tweet_id)
    inner join company_twitter_profile pf on twt.twitter_handle_id = pf.twitter_handle_id
    group by hashtags, twitter_handle_name
    order by count(hashtags) desc limit 3;
```

```
12
13
14 •   select * from top3_company_popular_accounts;
15
16
100% 45:14
Result Grid Filter Rows: Search Export:
|hashtags |count(hashta... |popularity_co... |
|> FidelityAssociate 191 |FidelityJobs |
|TSxCapitalOne 134 |CapitalOne |
|vzcareers 100 |VerizonCareers |
|                   |
|                   |
|                   |
| top3_company_popular_accounts 14 |
Action Output Response Duration / Fetc
| Time | Action |
```

14 Create and calling Views for job_salary_MA where job_location like “%MA%” and job_salary must not be null.

```
In [ ]: CREATE VIEW job_salary_MA AS
SELECT job_salary from job_post where job_location LIKE '%MA%' AND job_salary IS NOT NULL;

12
13
14 • select * from job_salary_MA;
15
16
```

Result Grid Filter Rows: Search Export:

job_salary
\$1L - \$1L (Glassdoor Est.)
\$58T - \$86T (Glassdoor Est.)
\$94T - \$1L (Glassdoor Est.)
\$90T - \$1L (Employer Est.)
\$70T - \$95T (Glassdoor Est.)

Action Output Response Duration / Fetc

15 Create and calling Views for international_job_eligibility where the eligibility must not have the phrase “Citizen”.

```
In [ ]: CREATE VIEW international_job_eligibility AS
select b.eligibility from job_requirements b inner join job_post j on b.job_id = j.id where b.eligibility NOT LIKE '%citizens%';

12
13
14 • select * from international_job_eligibility;
15
16
```

Result Grid Filter Rows: Search Export:

eligibility
All capable people
University students
Everyone
All interested Candidates
All Interested Candidates

Action Output Response Duration / Fetc

16 Create and calling Views for data_analyst_full_time where the job_description must have the phrases “Data Analyst ” and job_type must be “Full Time”.

```
In [ ]: CREATE VIEW data_analyst_full_time AS
select * from job_post where job_description LIKE '%data analyst%' and job_type = 'Full-time';

12
13
14 •  select * from data_analyst_full_time;
15
16
```

100% 38:14

Result Grid Filter Rows: Search Export:

id job_position company_id posted_date job_description job_location job_salary job_type job_url								
10	Full-time Data Analyst / Entry level (Remote)	10	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
11	Full-time Data Analyst / Entry level (Remote)	11	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
12	Full-time Data Analyst / Entry level (Remote)	12	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
13	Full-time Data Analyst / Entry level (Remote)	13	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
14	Full-time Data Analyst / Entry level (Remote)	14	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...

data_analyst_full_time 18

Action Output

Time Action Response Duration / Fetch

17 Create and calling Views for data_analyst_remote where the job position must have the phrases “Data Analyst” and job position must be “Remote”.

```
In [ ]: CREATE VIEW data_analyst_remote AS
select * from job_post where job_position LIKE '%data analyst%' and job_position LIKE '%Remote%';

13
14 •  select * from data_analyst_remote;
15
16
```

100% 35:14

Result Grid Filter Rows: Search Export:

id job_position company_id posted_date job_description job_location job_salary job_type job_url								
10	Full-time Data Analyst / Entry level (Remote)	10	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
11	Full-time Data Analyst / Entry level (Remote)	11	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
12	Full-time Data Analyst / Entry level (Remote)	12	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
13	Full-time Data Analyst / Entry level (Remote)	13	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...
14	Full-time Data Analyst / Entry level (Remote)	14	2022-12-04	We are looking for a passionate Product Data A...	Austin, TX	NULL	Full-time	https://www.linkedin.com/company/hire-match?...

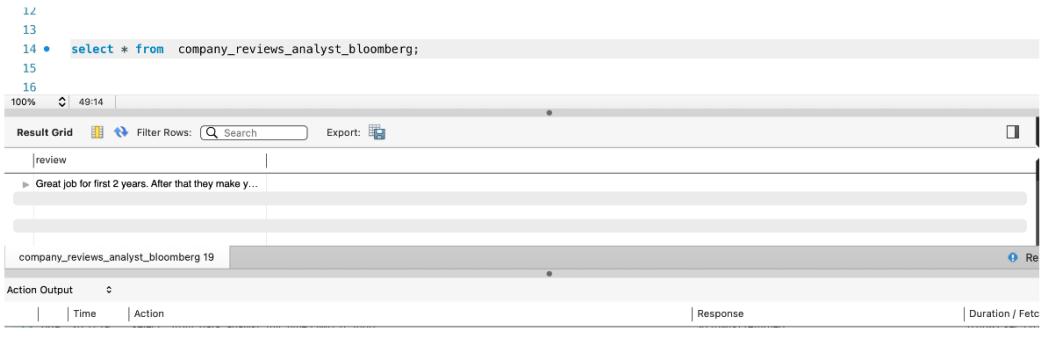
data_analyst_remote 18

Action Output

Time Action Response Duration / Fetch

18 Create and calling Views for company_reviews_analyst_bloomberg where the employee_title must be “analyst” and company name “Bloomerg”.

```
In [ ]: CREATE VIEW company_reviews_analyst_bloomberg AS
select review from company_reviews inner join company c on company_reviews.company_id = c.id
Where employee_title LIKE '%analyst%' and company_name LIKE '%Bloomberg%';
```



```
12
13
14 •  select * from company_reviews_analyst_bloomberg;
15
16
```

100% 49:14

Result Grid Filter Rows: Search Export:

review
» Great job for first 2 years. After that they make y...

company_reviews_analyst_bloomberg 19

Action Output 33: Time Action Response Duration / Fetc

19 Create and calling Views for job_2_or_more_experience where the experience requirements is Not Null and job_post is “Data Analyst”.

```
In [ ]: CREATE VIEW job_2_or_more_experience AS
SELECT job_position,posted_date,job_description,job_location,job_salary,job_url,job_type,job_requirements,experience_requirements
FROM job_post jp LEFT JOIN job_requirements jr ON jp.id = jr.id WHERE jr.experience_requirements NOT LIKE '%No Experience%' AND
```



```
12
13
14 •  select * from company_reviews_analyst_bloomberg;
15
16
```

100% 33:14

Result Grid Filter Rows: Search Export:

review
» Great job for first 2 years. After that they make you want to leave

company_reviews_analyst_bloomberg 20

Action Output 33: Time Action Response Duration / Fetc

20 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: CREATE VIEW job_requirements_analyst AS
SELECT job_position,job_requirements FROM job_requirements jr INNER JOIN job_post jp ON jp.id = jr.job_id WHERE jp.job_position LIKE '%Data Analyst%';
12
13
14 • select * from job_requirements_analyst;
15
16
```

Result Grid Filter Rows: Search Export:

job_position	job_requirements
Senior Data Analyst	- Supervises financial management and admins...
Data Analyst	- Working with the Country Director to provide e...
Data analyst	- Identify gaps in knowledge and overseeing inf...
Operation Analyst	- Rendering technical assistance to Database M...
KYC System Analyst	- Coordinating the work of subordinate employees;

Action Output | Time | Action | Response | Duration / Fetch

21 Create and calling Views for job_post_freshers_MA where the job_requirements is No requirements and the job location must be Massachusetts and job location must have the word MA.

```
In [ ]: CREATE VIEW job_post_freshers_MA AS
SELECT job_position,posted_date,job_description,job_location,job_salary,job_url,job_type,job_requirements,experience_requirements
FROM job_post jp INNER JOIN job_requirements jr ON jp.id = jr.job_id
WHERE jr.experience_requirements LIKE '%No Experience%' AND jp.job_location LIKE '%Massachusetts%' OR jp.job_location LIKE '%MA%';
12
13
14 • select * from job_post_freshers_MA;
15
16
```

Result Grid Filter Rows: Search Export:

job_position	posted_date	job_description	job_location	job_salary	job_url	job_type	job_requirements	experience_re
Data Analyst	2022-12-05	The Business Intelligence Advisor supports in th...	Pune, Maharashtra	NULL	NULL	NULL	Consultant will develop a clear and thorough	1 Year Job Exp
QA Analyst	2022-11-24	Our cutting-edge cloud platform uses AI and ma...	Pune, Maharashtra	NULL	NULL	NULL	- Monitor the overall programs performance and...	1 Year Job Exp
Data Science Platform Support Analyst	2022-12-03	Use insights based on data to inform decisions....	Pune, Maharashtra	NULL	NULL	NULL	Primary responsibilities include: recruitment and	1 Year Job Exp
Data Analyst - BPO / BPM	2022-11-25	2+ years experience as Data Analyst in busines...	Pune, Maharashtra	NULL	NULL	NULL	As a project assistant you will be responsible	2 Years Job Ex

Action Output | Time | Action | Response | Duration / Fetch

22 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: CREATE VIEW company_stream_information AS  
SELECT company_name, business_stream_name from company c INNER JOIN business_stream bs on c.id = bs.company_id  
WHERE bs.business_stream_name LIKE '%information%';
```

company_name	business_stream_name
Parabolic Career	Information Services, Staffing and Recruiting, a...
Diverse Lynx	Technology, Information and Internet
Diverse Lynx	Technology, Information and Internet
Diverse Lynx	Technology, Information and Internet
Diverse Lynx	Technology, Information and Internet
	company_stream_information 24

23 Create and calling Views for job_2_or_more_experience where the experience requirements is Not Null and job_post is “Data Analyst”.

```
In [ ]: CREATE VIEW job_eligibility_req_data_analyst AS  
SELECT jr.job_requirements, jr.experience_requirements FROM job_requirements jr  
INNER JOIN job_post jp ON jr.job_id = jp.id WHERE jp.job_position LIKE '%data analyst%';
```

job_requirements	experience_requirements
- Identify gaps in knowledge and overseeing inf...	< 1 Year
- Assist the Tavush Marz communities and com...	5 Years Job Exp.
- Network monitoring and administration;	4 Years Job Exp.
Graphic Designer will be responsible for every	1 Year Job Exp.
- Responsible for managing office administratio...	1 Year Job Exp.
	job_eligibility_req_data_analyst 25

24 Create and calling Views for company_reviews_analyst_bloomerg where the employee_title must be “analyst” and company name “Bloomerg”.

```
In [ ]: CREATE VIEW company_estbl_year AS
SELECT company_name, followers_count FROM company c INNER JOIN company_twitter_profile ct ON c.twitter_handle_name = ct.twitter_handle_name
WHERE establishment_year <= 2008;
```

12
13
14 • select * from company_estbl_year;
15
16

Result Grid Filter Rows: Search Export:

company_name	followers_cou...
Google	27841687
Accenture	551832

Action Output Response Duration / Fetc

Time Action

25 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: CREATE VIEW company_stream_information AS
SELECT company_name, business_stream_name FROM company c INNER JOIN business_stream bs ON c.id = bs.company_id
WHERE bs.business_stream_name LIKE '%information%';
```

```
12  
13  
14 • select * from company_stream_information;  
15  
16
```

Result Grid Filter Rows: Search Export:

company_name	business_stream_name
Parabolic Career	Information Services, Staffing and Recruiting, a...
Diverse Lynx	Technology, Information and Internet
Diverse Lynx	Technology, Information and Internet
Diverse Lynx	Technology, Information and Internet
Diverse Lynx	Technology, Information and Internet

Action Output Response Duration / Fetc

Time Action

26 Create and calling Views for company_reviews_analyst_bloomerg where the employee_title must be “analyst” and company name “Bloomerg”.

```
In [ ]: CREATE VIEW job_from_monster AS
SELECT job_position,posted_date,job_description,job_location,job_salary,job_type,job_url,website_name FROM job_post jp LEFT JOIN
12
13
14 • select * from job_from_monster;
15
16
100% 32:14
Result Grid Filter Rows: Search Export:
job_position | posted_date | job_description | job_location | job_salary | job_type | job_url | website_name |
Investment Banking Analyst and Associate I Ind... 2022-12-02 **Title: Investment Banking Analyst or Investme... Denver, CO Full-time https://uk.linkedin.com/company/selby-jennings... Monster
job_from_monster 28
Action Output
| Time | Action | Response | Duration / Fetc
```

27 Create and calling Views for international_job_eligibility where the eligibility must not have the phrase “Citizen”.

```
In [ ]: CREATE VIEW company_review_Parabolic AS
SELECT review FROM company_reviews cr INNER JOIN company c ON cr.company_id = c.id
WHERE company_name = 'Parabolic Career';
12
13
14 • select * from company_review_Parabolic;
15
16
100% 40:14
Result Grid Filter Rows: Search Export:
review |
Great Culture
New York Life - Hard Work Pays Off
Employee Friendly Company
Great to get ahead in the finance industry
Good Company!
company_review_Parabolic 29
Action Output
| Time | Action | Response | Duration / Fetc
```

28 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: 
    on,job_location,job_salary,job_url,job_type,job_requirements,experience_requirements,eligibility FROM job_post jp LEFT JOIN job_re
    > Experience%' AND jr.experience_requirements NOT LIKE '%4 Year%';

12
13
14 • select * from job_with_1to3_exp;
15
16
```

Result Grid Filter Rows: Search Export:

job_position	posted_date	job_description	job_location	job_salary	job_url	job_type	job_requirements
Senior Data Analyst	2022-11-30	Work on the development of the PropellerAds pr...	Bengaluru, Karnataka	₹80,045 a month	NULL	NULL	- Supervises financial ma...
Data Analyst	2022-11-28	Analyze data and develop machine learning mo...	Chennai, Tamil Nadu	NULL	NULL	NULL	- Working with the Count...
Data analyst	2022-11-28	Join a company that is transforming from a trad...	Bengaluru, Karnataka	NULL	NULL	NULL	- Identify gaps in knowled...
Operation Analyst	NULL	It combines traditional big data and machine le...	Kolkata, West Bengal	₹12,000 - ₹15,000 a month	NULL	NULL	- Rendering technical ass...
KYC System Analyst	2022-11-28	Focused on three practice areas of digital, data...	Remote in Hyderabad, Telangana	NULL	NULL	NULL	- Coordinating the work o...

Action Output | Time | Action | Response | Duration / Fetc

29 Create and calling Views for job_2_or_more_experience where the experience requirements is Not Null and job_post is “Data Analyst”.

```
In [ ]: 
CREATE VIEW job_with_more_3exp AS
SELECT job_position,posted_date,job_description,job_location,job_salary,job_url,job_type,job_requirements,experience_requirements,
WHERE jr.experience_requirements NOT LIKE '%No Experience%' AND jr.experience_requirements NOT LIKE '%Fresh%' AND jr.experience_>

12
13
14 • select * from job_with_more_3exp;
15
16
```

Result Grid Filter Rows: Search Export:

job_position	posted_date	job_description	job_location	job_salary	job_url	job_type	job_requirements
Data Analyst 2-B	2022-12-01	For example, using machine learning to determi...	Bengaluru, Karnataka+1 location	NULL	NULL	NULL	- Assist the Ta...
Analyst, Risk Analytics (PPNR/Machine Learning)	2022-11-24	Develops understanding of time series analysis...	Bengaluru, Karnataka	NULL	NULL	NULL	- Manage and...
Data Analyst	2022-11-16	Solutions to the top Financial Institutions in Ind...	Gurgaon, Haryana	NULL	NULL	NULL	- Network mor...
Intern - Business Analyst	2022-11-21	We believe that machine learning and other soft...	Remote	NULL	NULL	NULL	- Maintain CH...
Data Analyst Intern	2022-12-01	Reorganizing data in a readable format. A... Temporarily Remote in Chennai, Tamil Nadu	₹10,000 - ₹15,000 a month	NULL	NULL	NULL	- To welcome ...

Action Output | Time | Action | Response | Duration / Fetc

30 Create and calling Views for job_2_or_more_experience where the experience requirements is Not Null and job_post is “Data Analyst”.

```
In [ ]: CREATE VIEW company_review_google AS
SELECT review from company_reviews cr INNER JOIN company c ON cr.id = c.id
WHERE company_name = 'Google';

13
14 • select * from company_review_google;
15
16
```

100% 38:14 |

Result Grid Filter Rows: Search Export: |

review
How The Best Can Be Better

company_review_google 32 | Re

Action Output | Time | Action | Response | Duration / Fetc



31 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: CREATE VIEW twitter_handle_for_IT_stream AS
SELECT c.twitter_handle_name, business_stream_name from company c LEFT JOIN business_stream b ON c.id = b.company_id
LEFT JOIN company_twitter_profile ct ON c.twitter_handle_name = ct.twitter_handle_name
WHERE business_stream_name LIKE '%IT%';

12
13
14 • select * from twitter_handle_for_IT_stream;
15
16
```

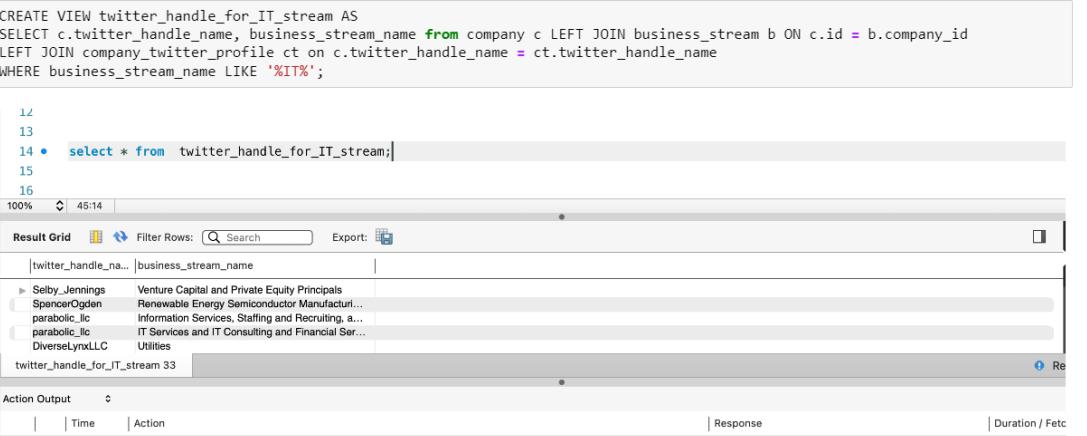
100% 45:14 |

Result Grid Filter Rows: Search Export: |

twitter_handle_name	business_stream_name
Selby_Jennings	Venture Capital and Private Equity Principals
SpencerOgden	Renewable Energy Semiconductor Manufactur...
parabolic_llc	Information Services, Staffing and Recruiting, a...
parabolic_llc	IT Services and IT Consulting and Financial Ser...
DiverseLynxLLC	Utilities

twitter_handle_for_IT_stream 33 | Re

Action Output | Time | Action | Response | Duration / Fetc



32 Create and calling Views for company_reviews_analyst_bloomerg where the employee_title must be “analyst” and company name “Bloomerg”.

```
In [ ]: CREATE VIEW rating_culture_Amazon AS
SELECT avg(rating_culture), company_name FROM company_ratings cr LEFT JOIN company c ON cr.company_id = c.id
WHERE company_name = 'Amazon';

12
13
14 • select * from rating_culture_Amazon;
15
16
```

Result Grid Filter Rows: Search Export:

avg(rating_culture)	company_name
3.5357	Amazon

rating_culture_Amazon 34

Action Output Response Duration / Fetch

33 Create and calling Views for job_2_or_more_experience where the experience requirements is Not Null and job_post is “Data Analyst”.

```
In [ ]: CREATE VIEW job_1yr_remote AS
SELECT job_position,posted_date,job_description,job_location,job_salary,job_url,job_type,job_requirements,experience_requirements
WHERE jr.experience_requirements LIKE '%1 Year' AND jp.job_location LIKE '%Remote%';
12
13
14 • select * from job_1yr_remote;
15
16
```

Result Grid Filter Rows: Search Export:

job_position	posted_date	job_description	job_location	job_salary	job_url	job_type	job_requirements	experience_requirements
KYC System Analyst	2022-11-28	Focused on three practice areas of digital, data...	Remote in Hyderabad, Telangana	HULL	HULL	HULL	- Coordinating the work of subordinate employees; < 1 Ye...	< 1 Year
Technical Business Analyst	2022-11-22	A bit of machine learning, such as regression an...	Remote in Bengaluru, Karnataka	₹40,000 a month	HULL	HULL	- Lead the effort to identify targets of opportunity...	< 1 Year

job_1yr_remote 35

Action Output Response Duration / Fetch

34 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: CREATE VIEW company_review_BOE_after_dec2018 AS
SELECT review from company_reviews cr LEFT JOIN company c on cr.company_id = c.id WHERE company_name = "Bank of England" AND cr.
12
13
14 •   select * from company_review_BOE_after_dec2018;
15
16
100% 49:14
Result Grid Filter Rows: Search Export:
|review|
▶ Nice place to work
Overall great place to work
Great people that care!
Loan processor
Professional Management
company_review_BOE_after_dec2018 36
Action Output
| Time | Action | Response | Duration / Fetc

```

35 Create and calling Views for company_reviews_analyst_bloomberg where the employee_title must be “analyst” and company name “Bloomberg”.

```
In [ ]: CREATE VIEW google_twitter_handle AS
select pf.twitter_handle_name,c.company_name from company_twitter_profile pf inner join
company c on pf.twitter_handle_name=c.twitter_handle_name
where c.company_name LIKE '%Google%';
12
13
14 •   select * from google_twitter_handle;
15
16
100% 37:14
Result Grid Filter Rows: Search Export:
|twitter_handle_na... |company_name |
▶ Google Google
google_twitter_handle 37
Action Output
| Time | Action | Response | Duration / Fetc

```

36 Create and calling Views for job_2_or_more_experience where the experience requirements is Not Null and job_post is “Data Analyst”.

```
In [ ]: CREATE VIEW tweets_posted_okta AS
select distinct tw(tweet_id, pf.twitter_handle_name, tw(tweet_text, tw.created_at as tweet_post_date
from company_tweets_info tw inner join company_twitter_profile pf on tw.twitter_handle_id = pf.twitter_handle_id
where tw.created_at > now() - interval 30 day and pf.twitter_handle_name = 'okta';

12
13
14 • select * from tweets_posted_okta;
15
16
```

Result Grid Filter Rows: Search Export:

tweet_id	twitter_handle_name	tweet_text	tweet_post_date
1600891712754176005	okta	Every year, @JINC helps over 63,500 kids in th...	2022-12-08 16:35:02
1600769653579587584	okta	At our #Oktae22 online event, key insights into...	2022-12-08 08:30:00
1600657719752785920	okta	@Hox_connor H @Hox_connor , Each compan...	2022-12-08 01:05:13
1600642881810833411	okta	The best of #Oktae22 is coming to YOU via we...	2022-12-08 00:06:16
160061656861249537	okta	Digital knowledge has become an imperative ski...	2022-12-07 22:21:42

Action Output | Response | Duration / Fetch

37 Create and calling Views for company_reviews_analyst_bloomerg where the employee_title must be “analyst” and company name “Bloomerg”.

```
In [ ]: CREATE VIEW fidelity_account_created_at AS
select twitter_handle_name, account_created_at from company_twitter_profile where twitter_handle_name = 'FidelityJobs';

12
13
14 • select * from fidelity_account_created_at;
15
16
```

Result Grid Filter Rows: Search Export:

twitter_handle_name	account_created_at
FidelityJobs	2011-10-21 19:40:39

fidelity_account_created_at 39

Action Output | Response | Duration / Fetch

38 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: CREATE VIEW popular_hashtags AS
select hashtags, count(hashtags) as popularity_count
from company_tweet_hashtags
group by hashtags
order by COUNT(hashtags) DESC;
```

hashtag	popularity_count
FidelityAssociate	191
TSCapitalOne	134
vzcareers	100
Oktane22	92
Lakehouse	69

39 Create and calling Views for company_reviews_analyst_bloomerg where the employee_title must be “analyst” and company name “Bloomerg”.

```
In [ ]: CREATE VIEW popular_tweets AS
select tweet_text, tweet_favorite_count from company_tweets_info
order by tweet_favorite_count desc;
```

tweet_text	tweet_favorite_count
Obsessed with Rosé and Lisa recreating this sc... 99174	
That's a wrap on Heartstopper Season 2 🍷 81843	
Wednesday now holds the record for most hour... 81670	
The game is back on! Alice in Borderland Seas... 73651	
We unleashed Thing on the streets of New York... 69411	

40 Create and calling Views for job_requirements_analyst where the job_position must have the word “Data Analyst”.

```
In [ ]: CREATE VIEW popular_tweets AS
select tweet_text, tweet_favorite_count from company_tweets_info
order by tweet_favorite_count desc;

12
13
14 • select * from popular_tweets;
15
16
```

100% 31:14

Result Grid Filter Rows: Search Export: Fetch rows: |

tweet_text	tweet_favorite_co...
Obsessed with Rosé and Lisa recreating this sc...	99174
That's a wrap on Heartstopper Season 2 🎉...	81843
Wednesday now holds the record for most hour...	81670
The game is back on! Alice in Borderland Seas...	73651
We unleashed Thing on the streets of New York... 69411	

popular_tweets 41

Action Output | Time | Action | Response | Duration / Fetc

MAJOR USE CASES:

1. Top 2 companies with highest job postings

```
select company_name, count(company_id) from company c inner join job_post jp on c.id = jp.company_id
group by company_name order by count(company_id) desc limit 2;

--  
13
14 • select company_name, count(company_id) from company c inner join job_post jp on c.id = jp.company_id
15 group by company_name order by count(company_id) desc limit 2;
16
```

100% 63:15

Result Grid Filter Rows: Search Export: Fetch rows: |

company_name	count(company_i...
Accenture	32
Deloitte	23

2. Management rating of Google

```
select Distinct c.company_name,cr.rating_management from company_ratings cr inner join company c on cr.company_id = c.id
where c.company_name = "Google";

--  
13
14 • select Distinct c.company_name,cr.rating_management from company_ratings cr inner join company c on cr.company_id = c.id
15 where c.company_name = "Google";
16
```

100% 33:15

Result Grid Filter Rows: Search Export: |

company_name	rating_managem...
Google	NULL
Google	4
Google	5

3. Companies offering Financial Analyst Post

```
select c.company_name, jp.job_position from company c inner join job_post jp on c.id=jp.company_id  
where job_position like '%Financial Analyst%';
```

```
--  
14  
15 • select c.company_name, jp.job_position from company c inner join job_post jp on c.id=jp.company_id  
16 where job_position like '%Financial Analyst%';  
17
```

100% 47:16

Result Grid Filter Rows: Search Export:

company_name	job_position
Fidelity Investments	Financial Analyst
Fidelity Investments	Senior Financial Analyst
Deloitte	Financial Analyst

4. Top company with Highest Tweets Count

```
select Distinct twt.twitter_handle_id , pf.twitter_handle_name, count(tweet_id) as tweet_count from  
company_tweets_info twt inner join company_twitter_profile pf on twt.twitter_handle_id = pf.twitter_handle_id  
group by twitter_handle_id, twitter_handle_name order by tweet_count desc limit 1;
```

```
--  
15 • select Distinct twt.twitter_handle_id , pf.twitter_handle_name, count(tweet_id) as tweet_count from  
16 company_tweets_info twt inner join company_twitter_profile pf on twt.twitter_handle_id = pf.twitter_handle_id  
17 group by twitter_handle_id, twitter_handle_name order by tweet_count desc limit 1;  
18
```

100% 83:17

Result Grid Filter Rows: Search Export: Fetch rows:

twitter_handle_id	twitter_handle_name	tweet_count
108203954	amazoncareers	200

5. Twitter handle name of Accenture

```
select pf.twitter_handle_name,c.company_name from company_twitter_profile pf inner join
company c on pf.twitter_handle_name=c.twitter_handle_name
where c.company_name LIKE '%Accenture%';
```

```
--  
14 •  select pf.twitter_handle_name,c.company_name from company_twitter_profile pf inner join
15      company c on pf.twitter_handle_name=c.twitter_handle_name
16      where c.company_name LIKE '%Accenture%';
17  
100%  41:16 |  
  
Result Grid  Filter Rows: (Search) Export:  
|twitter_handle_na... |company_name |  
▶ Accenture    Accenture
```

6. Display details of Data Analyst Full-Time Roles

```
select * from job_post where job_description LIKE '%data analyst%'
and job_type = 'Full-time';
```

```
--  
13
14 •  select * from job_post where job_description LIKE '%data analyst%' and job_type = 'Full-time';
15
16  
100%  95:14 |  
  
Result Grid  Filter Rows: (Search) Edit: Export/Import:   
|id |job_position |company_id |posted_date|job_description |job_location |job_salary |job_type |job_url |
| 10 |Full-time Data Analyst / Entry level (Remote) | 10 |2022-12-04 |We are looking for a passionate Product Data A... | Austin, TX | NULL | Full-time | https://www.linkedin.com/comp...
| 11 |Full-time Data Analyst / Entry level (Remote) | 11 |2022-12-04 |We are looking for a passionate Product Data A... | Austin, TX | NULL | Full-time | https://www.linkedin.com/comp...
| 12 |Full-time Data Analyst / Entry level (Remote) | 12 |2022-12-04 |We are looking for a passionate Product Data A... | Austin, TX | NULL | Full-time | https://www.linkedin.com/comp...
| 13 |Full-time Data Analyst / Entry level (Remote) | 13 |2022-12-04 |We are looking for a passionate Product Data A... | Austin, TX | NULL | Full-time | https://www.linkedin.com/comp...
| 14 |Full-time Data Analyst / Entry level (Remote) | 14 |2022-12-04 |We are looking for a passionate Product Data A... | Austin, TX | NULL | Full-time | https://www.linkedin.com/comp...
```

7. Company Reviews of Bloomberg

```
select review from company_reviews inner join company c on company_reviews.company_id = c.id
Where employee_title LIKE '%analyst%' and company_name LIKE '%Bloomberg%';
```

```
--  
13
14 •  select review from company_reviews inner join company c on company_reviews.company_id = c.id
15      Where employee_title LIKE '%analyst%' and company_name LIKE '%Bloomberg%';
16  
100%  76:15 |  
  
Result Grid  Filter Rows: (Search) Export:  
|review |  
▶ Great job for first 2 years. After that they make you want to leave
```

8. Job Requirements of Analyst posts

```
SELECT job_position,job_requirements FROM job_requirements jr INNER JOIN
job_post jp ON jp.id = jr.job_id WHERE jp.job_position LIKE '%analyst%';
```

```
--  
13  
14 •  SELECT job_position,job_requirements FROM job_requirements jr INNER JOIN
15   job_post jp ON jp.id = jr.job_id WHERE jp.job_position LIKE '%analyst%';
16
```

100% ◇ 1:15

Result Grid Filter Rows: Search Export:

job_position	job_requirements
► Senior Data Analyst	- Supervises financial management and adminis...
Data Analyst	- Working with the Country Director to provide e...
Data analyst	- Identify gaps in knowledge and overseeing inf...
Operation Analyst	- Rendering technical assistance to Database M...
KYC System Analyst	- Coordinating the work of subordinate employees;

9. Fresher job posts at MA Locations

```
SELECT job_position,posted_date,job_description,job_location,job_salary,job_url,job_type,job_requirements,experience_requirements
FROM job_post jp INNER JOIN job_requirements jr ON jp.id = jr.job_id
WHERE jr.experience_requirements LIKE '%No Experience%' AND jp.job_location LIKE jp.job_location LIKE '%MA%';
```

```
13
14 •  SELECT job_position,posted_date,job_description,job_location,job_salary,job_url,job_type,job_requirements,experience_requirements,eligibility
15   FROM job_post jp INNER JOIN job_requirements jr ON jp.id = jr.job_id
16   WHERE jr.experience_requirements LIKE '%No Experience%' AND jp.job_location LIKE
17   '%Massachusetts%' or jp.job_location LIKE '%MA%';
18
19
```

100% ◇ 1:18

Result Grid Filter Rows: Search Export:

job_position	posted_date	job_description	job_location	job_salary	job_url	job_type	job_requirements
► Data Analyst	2023-12-05	The Business Intelligence Advisor supports in th...	Pune, Maharashtra	NULL	NULL	NULL	Consultant will develop a clear and thorough...
QA Analyst	2022-11-24	Our cutting-edge cloud platform uses AI and ma...	Pune, Maharashtra	NULL	NULL	NULL	- Monitor the overall programs performance
Data Science Platform Support Analyst	2022-12-03	Use insights based on data to inform decisions....	Pune, Maharashtra	NULL	NULL	NULL	Primary responsibilities include: recruitment
Data Analyst - BPO / BPM	2022-11-25	2+ years experience as Data Analyst in business...	Pune, Maharashtra	NULL	NULL	NULL	As a project assistant you will be responsible

10. Twitter details of companies established after 2008

The screenshot shows a MySQL database interface with a query editor at the top containing the following SQL code:

```

SELECT company_name, followers_count FROM company c INNER JOIN company_twitter_profile ct ON c.twitter_handle_name = ct.twitter_handle_name
WHERE establishment_year <= 2008;

```

The results grid below displays two rows of data:

company_name	followers_count
Google	27841687
Accenture	551832

STEPS FOLLOWED FOR CHOOSING DATASETS:

By Scraping data from Twitter, we got all the details of company twitter handles like followers, handle id, handle name, description, verification status etc

Steps followed by us were :

- Collected all the API key required to connect to twitter
 - Connected to twitter using "tweepy"
 - Extracted all the tweets
 - Read twitter profile name of companies
 - Scrapped the data related to twitter profile in a list
 - Extracted the company tweets
 - Created a Dataframe of the extracted data
 - Appended Data Related to Company Info on Twitter From the tweets to data frame
 - Data appended Twitter handle id, handle name, profile image, description, followers, following, verification status, account date, hashtags, separating the hashtag list
- Connection to mysql using sql-alchemy and pushing data to twitter data in mysql

Script List:

https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Twitter_Bot_Assignment2/Twitter_Bot_Jobs_Database_v1.ipynb

Web Scrapping:

We collect the essential data using Beautiful Soup and Selenium in Python and saved it as a DataFrame from the most popular job sites including LinkedIn, and Glassdoor, it will be included in the database, with columns for Job Id, Job role, Job Description, Skills, Company, Location, and Job site. Later, the duplicate and null data in this CSV file are removed. The essential procedures, such as searching and filtering the data, can subsequently be carried out by storing this data in a SQL table.

Steps followed :

- Extracted data from LinkedIn site

https://www.linkedin.com/jobs/search?keywords=Analyst&location=United%20States&&geoId=103644278&trk=public_jobs_jobs-search-bar_search-submit&position=1&pageNum=0

- Extracted Job List
- Filtering The data to get only analyst roles
- Cleaning the data required
- Performed visualization
- Connection to my sql using sql alchemy and pushing data to twitter data in my sql. We Selected the appropriate data source to acquire accurate and relevant data. The majority of the information for this assignment was gathered through job websites.
- We audited, cleansed, and thoroughly confirmed the raw data that was scraped. The raw data had to be downloaded and reformatted for this.

Script Link:

https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Assignment_3/LinkedIn_Scrapping.ipynb

Glassdoor Webscrapping Using BeautifulSoup:

The data from ‘glassdoor.com’ is scraped using BeautifulSoup.

- The data collected from scraping is cleaned using python script and stored in lists.
- Cleaned data is audited and checked for accuracy before going further.
- The data stored in lists are stored in a DataFrame using the ‘pandas’ library.
- Using the ‘sqlalchemy’ library these dataframes are added to the connected database.
- Also, DataVisualisation for various columns like ‘job_position’ and ‘company_name’ and their count is done using the ‘plotly’ library.

Script Link:

https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Assignment_3/Glassdoor_WebScraping_BS.ipynb

Glassdoor Rating Review Scraping:

- The ratings and reviews are collected from a data repository.
- Collected data is visually represented using the ‘matplotlib’ library.

Script Link:

https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Assignment_3/Glassdoor_Rating_Review.ipynb

Job_data_Data_Repository

- The data downloaded from data repositories is read into the notebook using pandas.
- The data collected is cleaned using python script and stored in lists.
- Cleaned data is audited and checked for accuracy before going further.
- The data stored in lists are stored in a DataFrame using the ‘pandas’ library.
- Using the ‘sqlalchemy’ library these dataframes are added to the connected database.

Script Link:

https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Assignment_3/Job_data_Data_Repository.ipynb

https://github.com/aiskunks/Jobs_Database/blob/main/Analyst_Roles/Assignment_3/Jobs_Database_Data_Repository_1.ipynb

CONCLUSION:

Web scraping in linkedin, twitter and glassdoor was done, and the data which was taken from the data repository was cleaned and used in the database. Several create and insert queries were executed for the database chosen. All python codes were successfully implemented and tested without any errors. An ER diagram was created for the project and SQL queries were written & implemented for different use cases and views were created.