PROJECT 3 – DATA SCIENCE JOBS MARKET ANALYSIS

10.01.2024

Scope and Criteria

The purpose of this project is to design and implement an ETL pipeline that processes our data and stores it within a SQL database for future recalling and exploration.

Key Metrics:

Available positions, salaries, location, company rating

Resources:

Kaggle data set with 485 rows of data in 8 columns

Collaborators:

Chuck Bui

Amanuel Lebassi

Beau Massie

Christopher Turner

QUICK TAKE

- The "Cleaned Data Science Job Market & Salaries 2024" was chosen because it had many relevant key attributes that we were interested in, such as salary data, job titles, company names and ratings.
- We cleaned our database by handling null values, dropping unneeded information, reformatting values such as 'Date' to 'Days Listed, created new columns like Job Category from inside the 'Job Title' column as well as a 'Postings' to count the number of postings per state.
- We chose PostGreSQL for its ease of use for housing our database as well future exploration using SQL queries.
- We used the psycopg2 driver to automatically create our tables from within Python.
- A total number of 7 tables were created into PostGreSQL: a main database with all our cleaned and formatted data, then 6 other bite sized tables focused on each different job category as well as remote or non-remote.

LOAD THE DATA AND BEGIN CLEANING

[2]:
```
# import libraries
import csv
import pandas as pd
```

[3]:
```
file_path = 'data/clean_data.csv'
df = pd.read_csv(file_path)
df.head()
```

[3]:

	Job Title	Company Name	Location	Date	Job Link	Company Rating	Min Salary	Max Salary
0	Associate Stop Loss Underwriter	The Insurance Center\n2.7	Onalaska, WI	30d+	https://www.glassdoor.com/partner/jobListing.h...	2.7	57.0	84.0
1	Manager of Data Science	Nuvative, Inc.\n3.4	Wichita, KS	30d+	https://www.glassdoor.com/partner/jobListing.h...	3.4	106.0	157.0
2	Senior Data Product Manager	ProviderTrust\n4.2	Nashville, TN	11d	https://www.glassdoor.com/partner/jobListing.h...	4.2	105.0	141.0
3	Oncology Nurse Navigator	Inizio Engage\n3.6	Portland, OR	1d	https://www.glassdoor.com/partner/jobListing.h...	3.6	90.0	113.0
4	Head of Artificial Intelligence – Americas Region	Covestro\n3.6	Pittsburgh, PA	30d+	https://www.glassdoor.com/partner/jobListing.h...	3.6	89.0	148.0

[4]:
```
#checking if any non listed company rating docs have it in the compnay name
no_rating_df = df[df['Company Rating'].isnull()]
no_rating_df
```

[4]:

	Job Title	Company Name	Location	Date	Job Link	Company Rating	Min Salary	Max Salary
5	IT Manager	Western Welding Academy	Gillette, WY	26d	https://www.glassdoor.com/partner/jobListing.h...	NaN	NaN	NaN
6	Data Center Services Technician - IT, Crypto M...	Growler Mining	Tuscaloosa, AL	30d+	https://www.glassdoor.com/partner/jobListing.h...	NaN	NaN	NaN
7	Data Center Construction Coordinator	Applied Digital	Ellendale, ND	23d	https://www.glassdoor.com/partner/jobListing.h...	NaN	58.0	80.0
9	Principal Data Scientist	HelloGov	Tampa, FL	13d	https://www.glassdoor.com/partner/jobListing.h...	NaN	166.0	245.0
15	Sr. Digital Signal Processing Engineer	Cherish Health	Boston, MA	18d	https://www.glassdoor.com/partner/jobListing.h...	NaN	153.0	200.0
17	Senior Predictive Modeler	Sitewise Analytics	Dallas, TX	30d+	https://www.glassdoor.com/partner/jobListing.h...	NaN	NaN	NaN

[5]:
```
#removing the ratings from the company names
df['Company Name'] = df['Company Name'].str.split('\n').str[0]
df
```

[5]:

	Job Title	Company Name	Location	Date	Job Link	Company Rating	Min Salary	Max Salary
0	Associate Stop Loss Underwriter	The Insurance Center	Onalaska, WI	30d+	https://www.glassdoor.com/partner/jobListing.h...	2.7	57.0	84.0
1	Manager of Data Science	Nuvative, Inc.	Wichita, KS	30d+	https://www.glassdoor.com/partner/jobListing.h...	3.4	106.0	157.0
2	Senior Data Product Manager	ProviderTrust	Nashville, TN	11d	https://www.glassdoor.com/partner/jobListing.h...	4.2	105.0	141.0
3	Oncology Nurse Navigator	Inizio Engage	Portland, OR	1d	https://www.glassdoor.com/partner/jobListing.h...	3.6	90.0	113.0
4	Head of Artificial Intelligence – Americas Region	Covestro	Pittsburgh, PA	30d+	https://www.glassdoor.com/partner/jobListing.h...	3.6	89.0	148.0
...	...	...	...	...	...	...	...	...
480	Cloud Administrator	GM Financial	Arlington, TX	25d	https://www.glassdoor.com/partner/jobListing.h...	4.0	NaN	NaN
481	Robotics Engineer (AI)	Alpha Net Consulting	United States	4d	https://www.glassdoor.com/partner/jobListing.h...	NaN	NaN	NaN
482	Tchr of English- Newark School of Data Science...	Newark Board of Education	Newark, NJ	30d+	https://www.glassdoor.com/partner/jobListing.h...	3.3	62.0	107.0
483	Statistician	Sciome LLC	Research Triangle Park, NC	30d+	https://www.glassdoor.com/partner/jobListing.h...	NaN	NaN	NaN
484	Quantitative Analytics Manager - Data Modeling...	Freddie Mac	McLean, VA	5d	https://www.glassdoor.com/partner/jobListing.h...	3.6	140.0	210.0

485 rows × 8 columns

[6]:
```
# Dropping the Job Link column and creating a new DataFrame
df = df.drop('Job Link', axis=1)

# Display the cleaned DataFrame
df.head()
```

[6]:

	Job Title	Company Name	Location	Date	Company Rating	Min Salary	Max Salary
0	Associate Stop Loss Underwriter	The Insurance Center	Onalaska, WI	30d+	2.7	57.0	84.0
1	Manager of Data Science	Nuvative, Inc.	Wichita, KS	30d+	3.4	106.0	157.0
2	Senior Data Product Manager	ProviderTrust	Nashville, TN	11d	4.2	105.0	141.0
3	Oncology Nurse Navigator	Inizio Engage	Portland, OR	1d	3.6	90.0	113.0
4	Head of Artificial Intelligence – Americas Region	Covestro	Pittsburgh, PA	30d+	3.6	89.0	148.0

[7]:
```
# Remove 'd' and 'd+' from the days_listed column
df['Date'] = df['Date'].str.replace('d\+?', '', regex=True)
df['Date'] = df['Date'].str.replace('24h', '1', regex=False)
# Display the updated DataFrame
df.head()
```

CLEANING AND CREATING COLUMNS

[8]:
```
# Renaming the column
df.rename(columns={'Date': 'Days Listed'}, inplace=True)

# Display the updated DataFrame
df.head()
```

[8]:

	Job Title	Company Name	Location	Days Listed	Company Rating	Min Salary	Max Salary
0	Associate Stop Loss Underwriter	The Insurance Center	Onalaska, WI	30	2.7	57.0	84.0
1	Manager of Data Science	Nuvative, Inc.	Wichita, KS	30	3.4	106.0	157.0
2	Senior Data Product Manager	ProviderTrust	Nashville, TN	11	4.2	105.0	141.0
3	Oncology Nurse Navigator	Inizio Engage	Portland, OR	1	3.6	90.0	113.0
4	Head of Artificial Intelligence – Americas Region	Covestro	Pittsburgh, PA	30	3.6	89.0	148.0

[9]:
```
df = df[df['Max Salary'].notnull() & df['Min Salary'].notnull()]
df.head()
```

[9]:

	Job Title	Company Name	Location	Days Listed	Company Rating	Min Salary	Max Salary
0	Associate Stop Loss Underwriter	The Insurance Center	Onalaska, WI	30	2.7	57.0	84.0
1	Manager of Data Science	Nuvative, Inc.	Wichita, KS	30	3.4		
2	Senior Data Product Manager	ProviderTrust	Nashville, TN	11	4.2		
3	Oncology Nurse Navigator	Inizio Engage	Portland, OR	1	3.6		
4	Head of Artificial Intelligence – Americas Region	Covestro	Pittsburgh, PA	30	3.6		

10]:
```
# adding a average column by averaging the min and max
df['Average Salary'] = df[['Min Salary', 'Max Salary']].mean(axis=1)
df.head()
```

10]:

	Job Title	Company Name	Location	Days Listed	Company Rating	Min
0	Associate Stop Loss Underwriter	The Insurance Center	Onalaska, WI	30	2.7	
1	Manager of Data Science	Nuvative, Inc.	Wichita, KS	30	3.4	
2	Senior Data Product Manager	ProviderTrust	Nashville, TN	11	4.2	
3	Oncology Nurse Navigator	Inizio Engage	Portland, OR	1	3.6	
4	Head of Artificial Intelligence – Americas Region	Covestro	Pittsburgh, PA	30	3.6	

11]:
```
df=df[df['Company Rating'].notnull()]
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 348 entries, 0 to 484
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Job Title       348 non-null    object
 1   Company Name    348 non-null    object
 2   Location        348 non-null    object
```

[12]:
```
# Create a 'remote' column by checking if 'Remote' is in either 'Location' or 'Job Title'
df['Remote'] = df['Location'].str.contains('Remote', case=False, regex=True) | df['Job Title'].str.contains('Remote', case=False, regex=True)
df
```

[12]:

	Job Title	Company Name	Location	Days Listed	Company Rating	Min Salary	Max Salary	Average Salary	Remote
0	Associate Stop Loss Underwriter	The Insurance Center	Onalaska, WI	30	2.7	57.0	84.0	70.5	False
1	Manager of Data Science	Nuvative, Inc.	Wichita, KS	30	3.4	106.0	157.0	131.5	False
2	Senior Data Product Manager	ProviderTrust	Nashville, TN	11	4.2	105.0	141.0	123.0	False
3	Oncology Nurse Navigator	Inizio Engage	Portland, OR	1	3.6	90.0	113.0	101.5	False
4	Head of Artificial Intelligence – Americas Region	Covestro	Pittsburgh, PA	30	3.6	89.0	148.0	118.5	False
...	...	...	...	...	...	...	...	...	...
474	Principal Data Analyst	Citizens	Irving, TX	1	3.6	110.0	160.0	135.0	False
476	Manager, Data Science Platform Engineering	Dave Inc.	Remote	30	4.0	169.0	271.0	220.0	True
479	Analytics Solution Owner	Network Coverage	Remote	30	4.4	75.0	125.0	100.0	True
482	Tchr of English- Newark School of Data Science...	Newark Board of Education	Newark, NJ	30	3.3	62.0	107.0	84.5	False
484	Quantitative Analytics Manager - Data Modeling...	Freddie Mac	McLean, VA	5	3.6	140.0	210.0	175.0	False

348 rows × 9 columns

CREATING 'JOB CATEGORY' AND 'COMPANY RATING'

[17]:
```
# Create a function to categorize jobs
def categorize_job(title):
    title = title.lower()
    if 'analyst' in title:
        return 'Analyst'
    elif 'analytics' in title:
        return 'Analyst'
    elif 'engineer' in title:
        return 'Engineer'
    elif 'scientist' in title:
        return 'Scientist'
    elif 'science' in title:
        return 'Scientist'
    elif 'manager' in title:
        return 'Manager'
    else:
        return 'Other'

# Apply the function to create a new column
clean_df['Job Category'] = clean_df['Job Title'].apply(categorize_job)
clean_df
```

[21]:
```
def categorize_rating(rating):
    if pd.isna(rating):
        return None
    elif 0 <= rating < 1.1:
        return '0-1'
    elif 1.1 <= rating < 2.1:
        return '1-2'
    elif 2.1 <= rating < 3.5:
        return '2-3'
    elif 3.5 <= rating <= 5:
        return '4-5'
    else:
        return None
# Apply the function to create a new column 'Rating Range'
clean_df.loc[:, 'Rating Range'] = clean_df['Company Rating'].apply(categorize_rating)
```

[22]:

	Job Title	Company Name	Days Listed	Company Rating	Min Salary	Max Salary	Average Salary	Remote	City	State	Job Category	Rating Range
0	Associate Stop Loss Underwriter	The Insurance Center	30	2.7	57.0	84.0	70.5	False	Onalaska	WI	Other	2-3
1	Manager of Data Science	Nuvative, Inc.	30	3.4	106.0	157.0	131.5	False	Wichita	KS	Scientist	2-3
2	Senior Data Product Manager	ProviderTrust	11	4.2	105.0	141.0	123.0	False	Nashville	TN	Manager	4-5
3	Oncology Nurse Navigator	Inizio Engage	1	3.6	90.0	113.0	101.5	False	Portland	OR	Other	4-5
4	Head of Artificial Intelligence – Americas Region	Covestro	30	3.6	89.0	148.0	118.5	False	Pittsburgh	PA	Other	4-5
...	...	...	...	...	...	...	...	...	...	...	...	...
474	Principal Data Analyst	Citizens	1	3.6	110.0	160.0	135.0	False	Irving	TX	Analyst	4-5
476	Manager, Data Science Platform Engineering	Dave Inc.	30	4.0	169.0	271.0	220.0	True	Remote	Remote	Engineer	4-5
479	Analytics Solution Owner	Network Coverage	30	4.4	75.0	125.0	100.0	True	Remote	Remote	Analyst	4-5
482	Tchr of English- Newark School of Data Science...	Newark Board of Education	30	3.3	62.0	107.0	84.5	False	Newark	NJ	Scientist	2-3
484	Quantitative Analytics Manager - Data Modeling...	Freddie Mac	5	3.6	140.0	210.0	175.0	False	McLean	VA	Analyst	4-5

339 rows × 12 columns

[17]:

	Job Title	Company Name	Days Listed	Company Rating	Min Salary	Max Salary	Average Salary	Remote	City	State	Job Category
0	Associate Stop Loss Underwriter	The Insurance Center	30	2.7	57.0	84.0	70.5	False	Onalaska	WI	Other
1	Manager of Data Science	Nuvative, Inc.	30	3.4	106.0	157.0	131.5	False	Wichita	KS	Scientist
2	Senior Data Product Manager	ProviderTrust	11	4.2	105.0	141.0	123.0	False	Nashville	TN	Manager
3	Oncology Nurse Navigator	Inizio Engage	1	3.6	90.0	113.0	101.5	False	Portland	OR	Other
4	Head of Artificial Intelligence – Americas Region	Covestro	30	3.6	89.0	148.0	118.5	False	Pittsburgh	PA	Other
...	...	...	...	...	...	...	...	...	...	...	...
474	Principal Data Analyst	Citizens	1	3.6	110.0	160.0	135.0	False	Irving	TX	Analyst
476	Manager, Data Science Platform Engineering	Dave Inc.	30	4.0	169.0	271.0	220.0	True	Remote	None	Engineer
479	Analytics Solution Owner	Network Coverage	30	4.4	75.0	125.0	100.0	True	Remote	None	Analyst
482	Tchr of English- Newark School of Data Science...	Newark Board of Education	30	3.3	62.0	107.0	84.5	False	Newark	NJ	Scientist
484	Quantitative Analytics Manager - Data Modeling...	Freddie Mac	5	3.6	140.0	210.0	175.0	False	McLean	VA	Analyst

348 rows × 11 columns

CREATING 'POSTINGS' COLUMN AND OUR DATAFRAMES

```
[23]: # create a grouby for state locations - we'll use this for the size of our scatter bubbles
location_postings = clean_df.groupby('State').size().reset_index(name='Postings')
```

```
[24]: # merge locations_postings with our df to get postings column
clean_df = clean_df.merge(location_postings, on='State', how='left')
```

```
[25]: # check it
clean_df
```

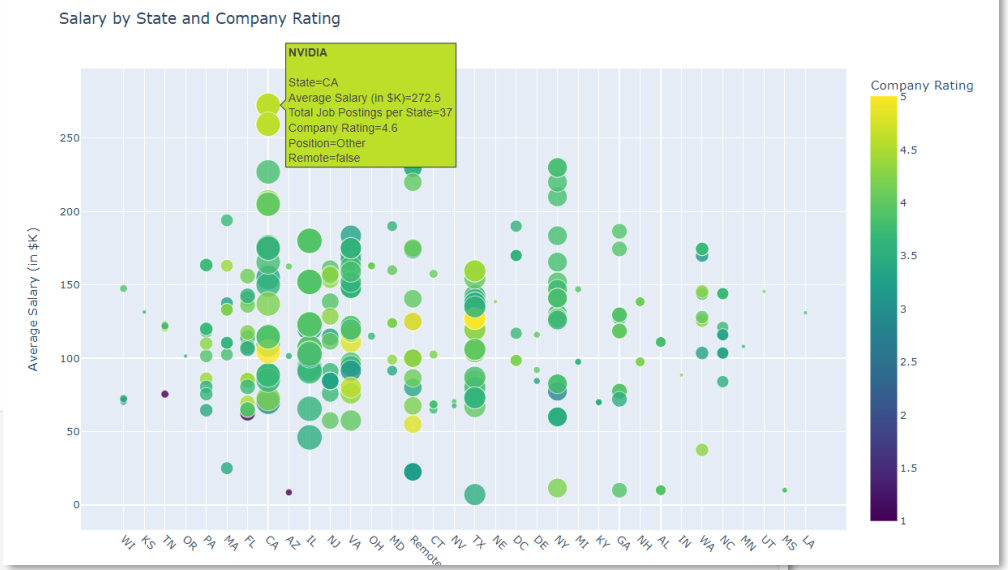
[25]:

	Job Title	Company Name	Days Listed	Company Rating	Min Salary	Max Salary	Average Salary	Remote	City	State	Job Category	Rating Range	Postings
0	Associate Stop Loss Underwriter	The Insurance Center	30	2.7	57.0	84.0	70.5	False	Onalaska	WI	Other	2-3	4
1	Manager of Data Science	Nuvative, Inc.	30	3.4	106.0	157.0	131.5	False	Wichita	KS	Scientist	2-3	1
2	Senior Data Product Manager	ProviderTrust	11	4.2	105.0	141.0	123.0	False	Nashville	TN	Manager	4-5	4
3	Oncology Nurse Navigator	Inizio Engage	1	3.6	90.0	113.0	101.5	False	Portland	OR	Other	4-5	1
4	Head of Artificial Intelligence – Americas Region	Covestro	30	3.6	89.0	148.0	118.5	False	Pittsburgh	PA	Other	4-5	11
...	...	...	...	...	...	...	...	...	...	...	...	...	...
334	Principal Data Analyst	Citizens	1	3.6	110.0	160.0	135.0	False	Irving	TX	Analyst	4-5	30
335	Manager, Data Science Platform Engineering	Dave Inc.	30	4.0	169.0	271.0	220.0	True	Remote	Remote	Engineer	4-5	21
336	Analytics Solution Owner	Network Coverage	30	4.4	75.0	125.0	100.0	True	Remote	Remote	Analyst	4-5	21
337	Tchr of English- Newark School of Data Science...	Newark Board of Education	30	3.3	62.0	107.0	84.5	False	Newark	NJ	Scientist	2-3	19
338	Quantitative Analytics Manager - Data Modeling...	Freddie Mac	5	3.6	140.0	210.0	175.0	False	McLean	VA	Analyst	4-5	28

339 rows × 13 columns

```
[28]: # create a remote df
remote_df = clean_df[clean_df['Remote']== True]

# create a non remote df
non_remote_df = clean_df[clean_df['Remote']== False]

# create df's for each position
sci_df = clean_df[clean_df['Job Category'] == 'Scientist']
eng_df = clean_df[clean_df['Job Category'] == 'Engineer']
mgt_df = clean_df[clean_df['Job Category'] == 'Manager']
ana_df = clean_df[clean_df['Job Category'] == 'Analyst']
other_df = clean_df[clean_df['Job Category'] == 'Other']
```

```
# plotly express scatter
fig = px.scatter(
    clean_df,
    x='State', y='Average Salary',
    size='Postings', color='Company Rating',
    hover_name='Company Name',
    hover_data=['Company Rating', 'Job Category', 'Remote'],
    title='Salary by State and Company Rating',
    labels={'State': 'State', 'Job Category': 'Position', 'Postings':'Total Job Postings per State','Average Salary': 'Average Salary (in $K)','Remote':'
    color_continuous_scale=px.colors.sequential.Viridis,
    height=700
)
# add interaction popup with a white border
fig.update_traces(marker=dict(opacity=.8,
                              line=dict(width=.7,
                                        color='white')))
fig.update_xaxes(tickangle=45)
fig.show()
```



CHECKING ON OUR DATA

```
Counts for each Rating Range
0-1      3
2-3     45
4-5    291
Name: Company Rating, dtype: int64

5 Highest paying companies for Company Name
NVIDIA      265.277778
Indeed      262.000000
Insurity    240.750000
Rokt        230.000000
Lime        229.500000
Name: Average Salary, dtype: float64

Top 5 rated companies
                        Company Rating  Average Salary  Postings
Company Name
Blackstone Group                   5.0           145.0      24.0
Openwork, LLC                      5.0           127.0      30.0
Penfield Search Partners           5.0           104.5      37.0
Vital Edge Solutions               5.0           125.0      21.0
Emergent Software                  4.8            55.0      21.0

5 Highest paying states sorted by State
CA    169.135135
OH    151.000000
DC    150.444444
VA    146.857143
UT    145.500000
Name: Average Salary, dtype: float64

Top 5 rated State
NE    4.300000
IN    4.300000
UT    4.200000
NH    4.133333
OH    4.000000
Name: Company Rating, dtype: float64
```

```python
# data exploring
# rating counts
rating_range_counts = clean_df.groupby('Rating Range')['Company Rating'].count()

#find top paying companies (may be skewed by number of jobs posted)
top_paying_companies = clean_df.groupby('Company Name')['Average Salary'].mean()
top_5_paying_companies = top_paying_companies.sort_values(ascending=False).head(5)

#Find top rated companies
top_rated_companies = clean_df.groupby('Company Name')[['Company Rating', 'Average Salary', 'Postings']].mean()
top_5_rated_companies = top_rated_companies.sort_values(by='Company Rating', ascending=False).head(5)

#top rated states
top_rated_states = df.groupby('State')['Company Rating'].mean().sort_values(ascending=False).head(5)

#top paying states could be useful to provide cost of living?
top_paying_states = df.groupby('State')['Average Salary'].mean()
print(f'Counts for each {rating_range_counts}')
print()
print(f'5 Highest paying companies for {top_5_paying_companies}'
print()
print('Top 5 rated companies')
print(top_5_rated_companies)
print()
print(f'5 Highest paying states sorted by {top_paying_states}')
print()
print(f'Top 5 rated {top_rated_states}')
```

```python
#category rating and salary  includes remote and onsite
average_rating_salary_by_category = clean_df.groupby('Job Category')[['Company Rating', 'Average Salary']].mean()
print(average_rating_salary_by_category)
```

```
              Company Rating  Average Salary
Job Category
Analyst             3.761176      115.700000
Engineer            3.833962      151.594340
Manager             3.900000      129.850000
Other               3.763077      109.284615
Scientist           3.858730      131.563492
```

```python
average_salary_remote = clean_df[clean_df['Remote'] == True]['Average Salary'].mean()
average_rating_remote = clean_df[clean_df['Remote'] == True]['Company Rating'].mean()
average_salary_non_remote = clean_df[clean_df['Remote'] == False]['Average Salary'].mean()
average_rating_non_remote = clean_df[clean_df['Remote'] == False]['Company Rating'].mean()
print(f"Average Salary for Remote Jobs: {average_salary_remote}")
print(f"Average Rating for Remote Jobs: {average_rating_remote}")
print(f"Average Salary for Non-Remote Jobs: {average_salary_non_remote}")
print(f"Average Rating for Non-Remote Jobs: {average_rating_non_remote}")
```

```
Average Salary for Remote Jobs: 119.6891891891892
Average Rating for Remote Jobs: 3.775675675675676
Average Salary for Non-Remote Jobs: 127.21688741721854
Average Rating for Non-Remote Jobs: 3.8178807947019866
```

```python
best_companies = clean_df[['Company Name', 'Average Salary', 'Company Rating', 'State', 'Postings']].sort_values(
    by=['Company Rating', 'Postings'],ascending=False)
best_companies.head()
```

	Company Name	Average Salary	Company Rating	State	Postings
282	Penfield Search Partners	104.5	5.0	CA	37
141	Openwork, LLC	127.0	5.0	TX	30
252	Openwork, LLC	127.0	5.0	TX	30
330	Openwork, LLC	127.0	5.0	TX	30
75	Blackstone Group	145.0	5.0	NY	24

CREATING TABLES IN POSTGRESQL

```
# run this in bash to install the psycopg2 database driver : pip install sqlalchemy psycopg2
# SQLAlchemy generates SQL statements and psycopg2 sends SQL statements to the database.
# engine = create_engine('postgresql://USERNAME:PASSWORD@localhost:5432/CREATED_DATABASE')
```

```
# job_data: this is the full data table - use psycopg2 to create tables in PostgreSQL
from sqlalchemy import create_engine

# Create an engine to connect to PostgreSQL
engine = create_engine('postgresql://postgres:postgres@localhost:5432/job_data')

data = clean_df

# Writing the data to a new table in the PostgreSQL database
data.to_sql('job_data', engine, if_exists='replace', index=False)

print("Data written to PostgreSQL successfully!")
```

```
Data written to PostgreSQL successfully!
```

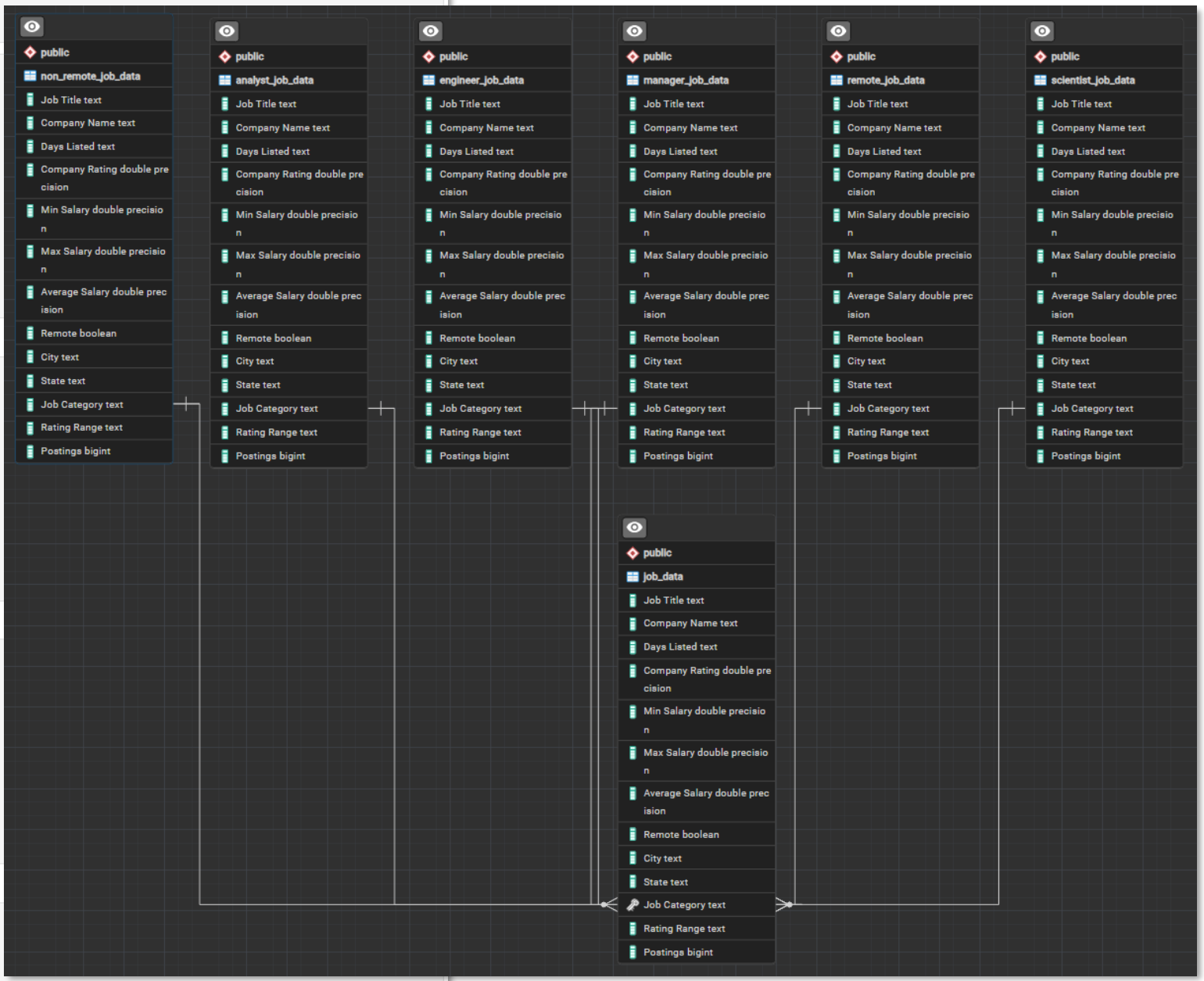
```
# non_remote_job_data- use psycopg2 to create tables in PostgreSQL

# Create an engine to connect to PostgreSQL
engine = create_engine('postgresql://postgres:postgres@localhost:5432/job_data')

data = non_remote_df

# Writing the data to a new table in the PostgreSQL database
data.to_sql('non_remote_job_data', engine, if_exists='replace', index=False)

print("Data written to PostgreSQL successfully!")
```

```
Data written to PostgreSQL successfully!
```

```
# remote_job_data - use psycopg2 to create tables in PostgreSQL

# Create an engine to connect to PostgreSQL
engine = create_engine('postgresql://postgres:postgres@localhost:5432/job_data')

data = remote_df

# Writing the data to a new table in the PostgreSQL database
data.to_sql('remote_job_data', engine, if_exists='replace', index=False)

print("Data written to PostgreSQL successfully!")
```

```
Data written to PostgreSQL successfully!
```

```
# scientist_job_data - use psycopg2 to create tables in PostgreSQL

# Create an engine to connect to PostgreSQL
engine = create_engine('postgresql://postgres:postgres@localhost:5432/job_data')
```



CHECKING TABLES IN POSTGRESQL

Query Query History

```sql
SELECT * FROM job_data;

SELECT * FROM analyst_job_data;

SELECT * FROM engineer_job_data;

SELECT * FROM manager_job_data;

SELECT * FROM scientist_job_data;

SELECT * FROM non_remote_job_data;

SELECT * FROM remote_job_data;
```

Data Output Messages Notifications

	Job Title text	Company Name text	Days Listed text	Company Rating double precision	Min Salary double precision	Max Salary double precision	Average Salary double precision	Remote boolean	City text	State text	Job Category text	Rating Range text	Postings bigint
1	Associate Stop Loss Underwriter	The Insurance Center	30	2.7	57	84	70.5	false	Onalaska	WI	Other	2-3	4
2	Manager of Data Science	Nuvative, Inc.	30	3.4	106	157	131.5	false	Wichita	KS	Scientist	2-3	1
3	Senior Data Product Manager	ProviderTrust	11	4.2	105	141	123	false	Nashville	TN	Manager	4-5	4
4	Oncology Nurse Navigator	Inizio Engage	1	3.6	90	113	101.5	false	Portland	OR	Other	4-5	1
5	Head of Artificial Intelligence – Americas Region	Covestro	30	3.6	89	148	118.5	false	Pittsburgh	PA	Other	4-5	11
6	Senior AIOps Engineer	Health Data Analytics Institute	30	4.4	151	175	163	false	Dedham	MA	Engineer	4-5	10
7	Training Department Supervisor	Esquire Law Services	27	1	60	65	62.5	false	Palm Beach Gardens	FL	Other	0-1	15
8	Senior Data Scientist	Idea Financial	30	4.1	112	169	140.5	false	Miami	FL	Scientist	4-5	15
9	Machine Learning Engineer	Digital Force Technologies	30	3.3	140	170	155	false	San Diego	CA	Engineer	2-3	37
10	IT Site Engineer (labs)	Oxford Global Resources	6	3.4	50	0	25	false	Lexington	MA	Engineer	2-3	10
11	RN Clinical Educator	Inizio Engage	1	3.6	90	113	101.5	false	Phoenix	AZ	Other	4-5	3
12	Junior Game Mathematician	Incredible Technologies	30	3.5	37	55	46	false	Vernon Hills	IL	Other	4-5	40
13	Senior Manager Advanced Analytics (US)	TD Bank	30	3.9	111	166	138.5	false	Mount Laurel	NJ	Analyst	4-5	19
14	AI/ML Subject Matter Expert	Thomas & Herbert Consulting LLC	30	2.4	165	175	170	false	Springfield	VA	Other	2-3	28
15	Lead AI Developer	SeeScan Inc.	30	3.6	125	175	150	false	San Diego	CA	Other	4-5	37
16	Data Scientist	Commonwealth Health Insurance Connector Auth...	30	3.8	100	105	102.5	false	Boston	MA	Scientist	4-5	10
17	Data Scientist - Manufacturing Analytics	Pilot Chemical Co.	30	3.7	95	135	115	false	West Chester	OH	Analyst	4-5	4
18	Project Manager/ Senior Data Analyst	HumanTouch LLC	6	3.4	75	108	91.5	false	Patuxent River	MD	Analyst	2-3	7
19	Data Science Technical Fellow	Indeed	30	4.1	214	310	262	true	Remote	Remote	Scientist	4-5	21
20	Artificial Intelligence (AI) Cybersecurity Architect	The Travelers Companies, Inc.	2	4.1	119	196	157.5	false	Hartford	CT	Other	4-5	5
21	Director - Systems and Technology - Data, Artificial Intelligence and Portfolio Delivery	SoCalGas	30	4.4	165	248	206.5	false	Los Angeles	CA	Other	4-5	37
22	AI Integration Specialist	De Castroverde Law Group	30	3.9	54	87	70.5	false	Las Vegas	NV	Other	4-5	2
23	Data Analyst (Onsite)	The Greentree Group	30	4.5	55	84	69.5	false	Tallahassee	FL	Analyst	4-5	15
24	Digital Marketing Specialist (Paid Media Focus)	Gentle Monster	30	2.9	65	75	70	false	Anaheim	CA	Other	2-3	37
25	Senior Data Scientist	Navy Federal Credit Union	2	4.1	98	174	136	false	Pensacola	FL	Scientist	4-5	15
26	Statistician	Gladney Center for Adoption	4	4	52	82	67	false	Fort Worth	TX	Other	4-5	30
27	Senior Research Scientist - Adversarial Machine Learning	Carnegie Mellon University	30	4.4	61	111	86	false	Pittsburgh	PA	Scientist	4-5	11
28	Principal Data Scientist - Time Series	Navy Federal Credit Union	30	4.1	112	200	156	false	Pensacola	FL	Scientist	4-5	15
29	NC3 Systems Engineer - Omaha	Johns Hopkins Applied Physics Laboratory (APL)	30	4.3	117	160	138.5	false	Offutt A F B	NE	Engineer	4-5	1
30	Computational Engineer (Level II or Senior)	St. Jude Children's Research Hospital	30	4.5	86	155	120.5	false	Memphis	TN	Engineer	4-5	4
31	Employee Relations Manager	Esquire Law Services	30	1	73	78	75.5	false	Nashville	TN	Manager	0-1	4