NYC Job Posting Data Analysis

Gabrielle Glasgow, Max Ross, Jason Fearnall

Overview

Introduction

Architecture Setup

Data Cleaning

Analysis Results

Questions/Comments

01 »

Introduction

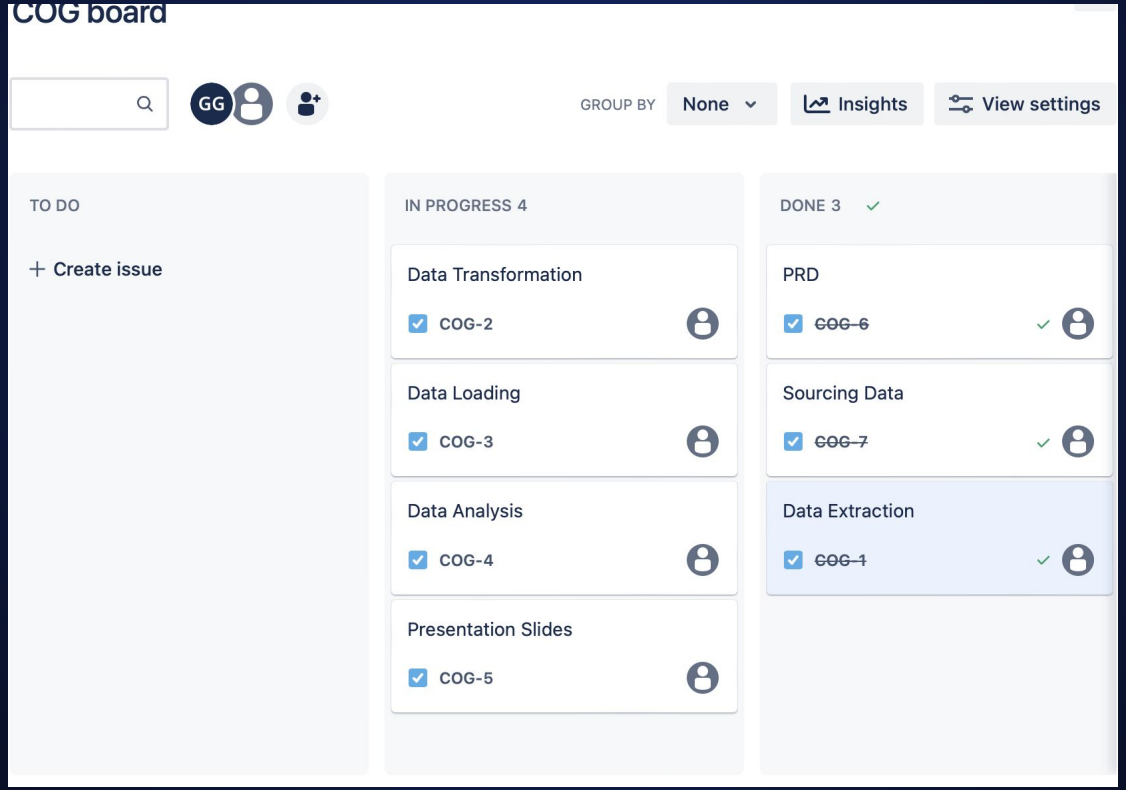
Our Company

Cognixia Gov Staffing is harnessing data to enhance recruitment strategies that will benefit both clients and candidates. Our Focus is on job postings in NYC, using data from the City of New York's official job site. This dataset provides a comprehensive view of the local job market.

Our Idea

By analyzing this data, we gain insights into recruitment complexities, enabling us to offer tailored solutions to our clients. Our goal is to connect the right candidates with the right opportunities in NYC driven by actionable insights from data analysis.

Project Setup/Timeline



02 »

Architecture Setup

IAM Roles and Policies

IAM Roles and Policies

Policy name	Type	Used as	Description
AmazonRedshift-Com...	Customer managed	Permissions policy (1)	-
AmazonRedshift-Com...	Customer managed	Permissions policy (1)	-
AWSGlueServiceRole-...	Customer managed	Permissions policy (1)	This policy will be used for Glue Crawler and Job execution. Please do NOT delete!
AWSLambdaBasicExec...	Customer managed	Permissions policy (1)	-
AWSLambdaBasicExec...	Customer managed	Permissions policy (1)	-
AWSLambdaVPCAcce...	Customer managed	Permissions policy (1)	-
AWSLambdaVPCAcce...	Customer managed	Permissions policy (1)	-
rds-proxy-policy-170...	Customer managed	Permissions policy (1)	-
rds-proxy-policy-170...	Customer managed	Permissions policy (1)	Allows RDS Proxy access to database connection credentials
rds-proxy-policy-170...	Customer managed	Permissions policy (1)	-

S3 Bucket

○	semistructuredata	US East (N. Virginia) us-east-1	Objects can be public	February 19, 2024, 11:46:34 (UTC-05:00)

- Added read and write permission
- Uploaded a csv file to the bucket
- S3 bucket served as our Lake

Name	Type	Last modified	Size	Storage class
data1.csv	csv	February 21, 2024, 14:54:22 (UTC-05:00)	6.2 MB	Standard
data2.csv	csv	February 21, 2024, 14:17:07 (UTC-05:00)	6.1 MB	Standard
data3.csv	csv	February 21, 2024, 14:18:07 (UTC-05:00)	6.1 MB	Standard
data4.csv	csv	February 21, 2024, 14:19:39 (UTC-05:00)	6.2 MB	Standard
data5.csv	csv	February 21, 2024, 14:21:06 (UTC-05:00)	6.1 MB	Standard
data6.csv	csv	February 21, 2024, 14:22:24 (UTC-05:00)	2.9 MB	Standard

Lambda Function Overview

Function overview Info

Diagram | Template

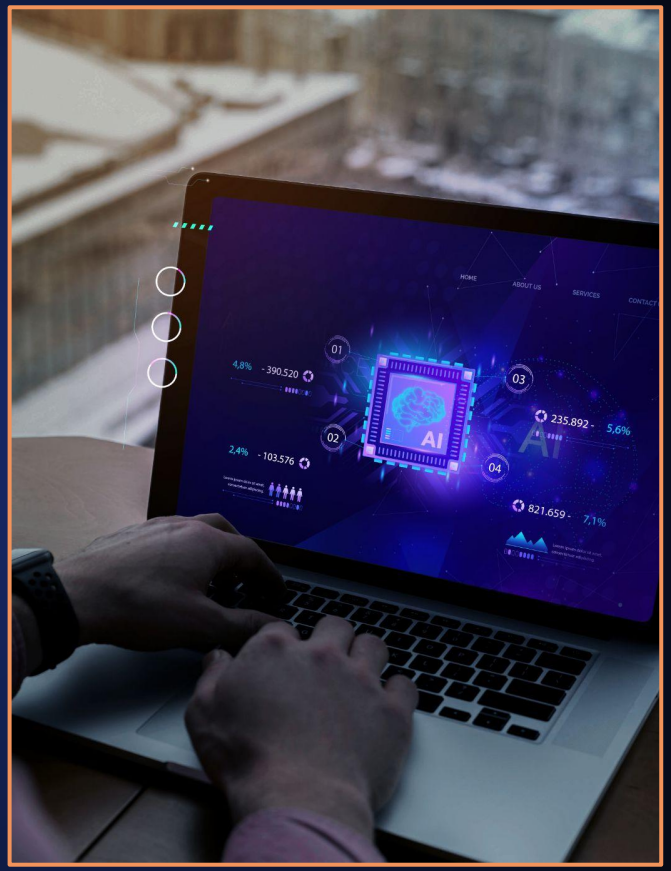
extractGovJson

Layers (1)

+ Add trigger

Lambda Function:

```python
import json
import boto3
import pandas as pd
import requests

s3 = boto3.client('s3')

def lambda_handler(event, context):
    bucket ='semistructuredata'
    url = 'https://data.cityofnewyork.us/resource/kpav-sd4t.json'

    params = {'$limit': 1000, '$offset':1000}
    response = requests.get(url,params=params)
    data = response.json()


    df = pd.DataFrame(data)
    df.drop_duplicates(inplace=True)


    csv_data = df.to_csv(index=False)
    fileName = 'data2.csv'
    uploadByteStream = bytes(csv_data.encode('utf-8'))
    s3.put_object(Bucket=bucket, Key=fileName, Body=uploadByteStream)
    print('Put Complete')
```



03 »

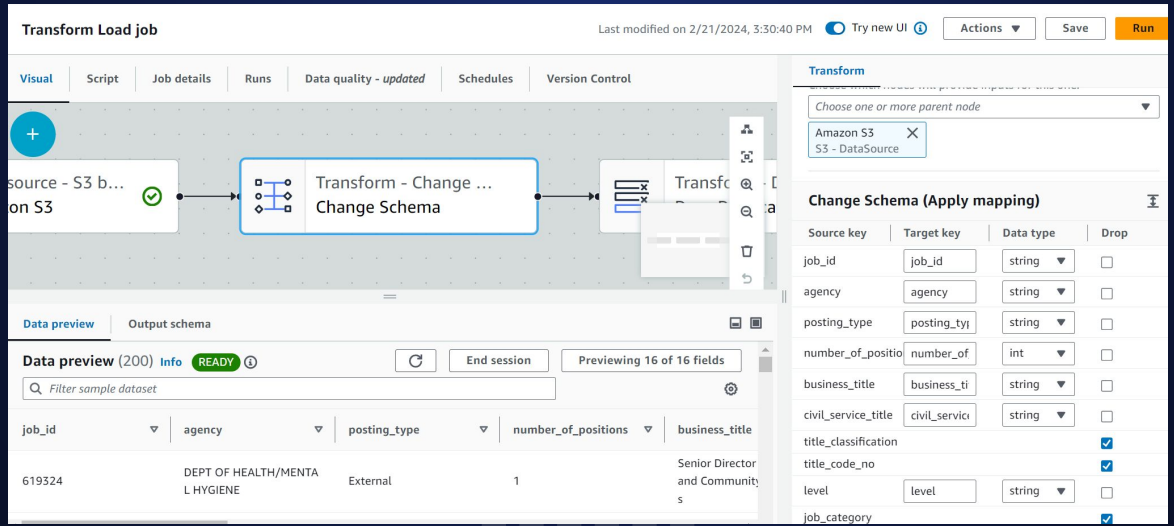
Data Transform & Load

Prior Configuration

- Uncleaned Data Stored in S3 Bucket
- Setup Data Warehouse (Redshift)
- IAM Role and permissions to access S3 and Redshift
- Glue Connector for Redshift DB connection
- Initial Exploratory Data Analysis to extract necessary features

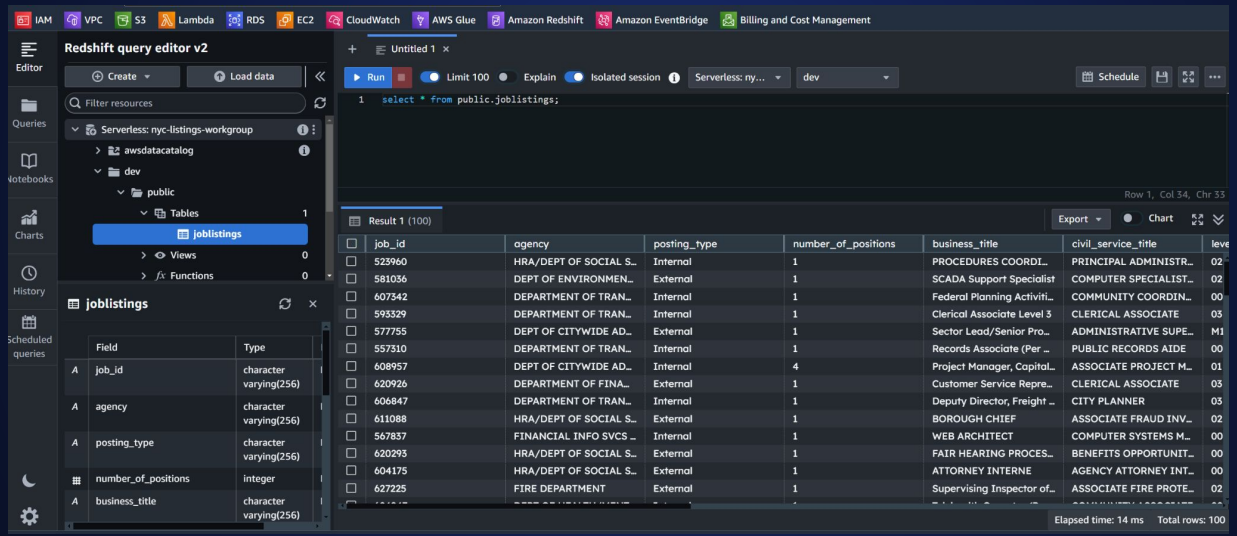
AWS GLUE

- Used AWS Glue for ETL Jobs

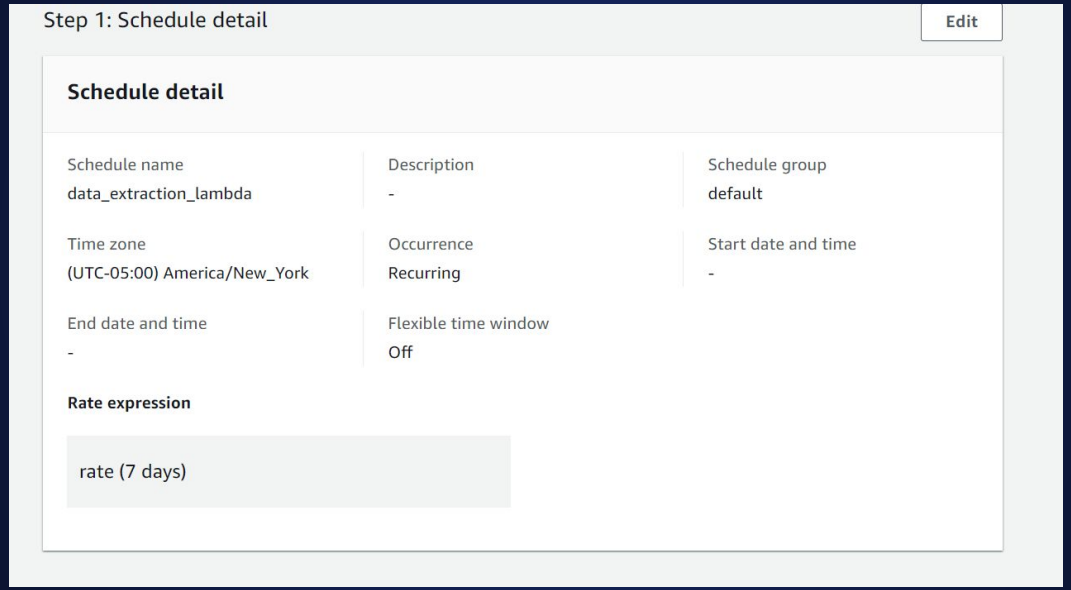


AWS RedShift

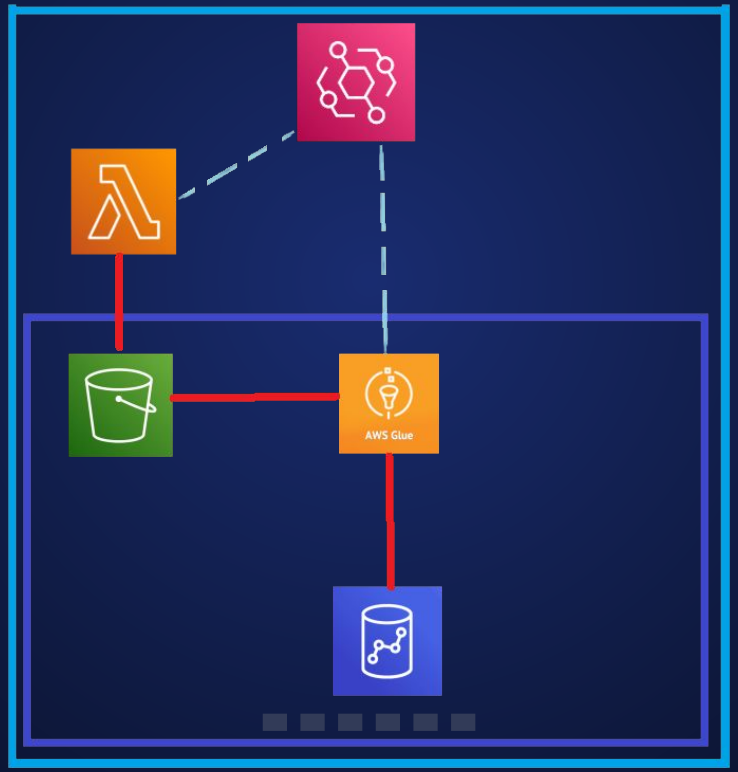
- Served as our Warehouse



EventBridge Scheduler



Pipeline Architecture



04 »

Data Cleaning

UNCLEANED

```
['Job ID',
 'Agency',
 'Posting Type',
 '# Of Positions',
 'Business Title',
 'Civil Service Title',
 'Title Classification',
 'Title Code No',
 'Level',
 'Job Category',
 'FT/PT',
 'Career Level',
 'Salary Range From',
 'Salary Range To',
 'Salary Frequency',
 'Work Location',
 'Division/Work Unit',
 'Job Description',
 'Min Qualifications',
 'Preferred Skills',
 'Additional Information',
 'To Apply',
 'Hours/Shift',
 'Work Location 1',
 'Recruitment Contact',
 'Residency Requirement',
 'Posting Date',
 'Post Until',
 'Posting Updated',
 'Process Date']
```

30 columns

VS.

CLEANED

```
['Job_id',
 'Agency',
 'Posting_type',
 'Num_pos',
 'Business_title',
 'Civil_title',
 'Level',
 'Category',
 'Ft_Pt',
 'Career_level',
 'Yearly_salary',
 'Location',
 'Division',
 'Post_Date']
```

14 columns

Cleaning Code

Salary Range From	Salary Range To	Salary Frequency
399.68	399.68	Daily
74650.00	109409.00	Annual

Annual Base Salary
83932.8
74650.0

```python
#Rename FT/PT columns for easier calculations
df =df.rename(columns={'Full-Time/Part-Time indicator': 'FT/PT'})

#Create new column to hold Base Salary Calculations
df['Base Salary'] = 0

#Calculate Daily salary rate to yearly, and copy Annual as yearly
df.loc[df['Salary Frequency'] == 'Daily', 'Base Salary'] = round(df['Salary Range From'].mul(210), 2)
df.loc[df['Salary Frequency'] == 'Annual', 'Base Salary'] = round(df['Salary Range From'], 2)

#calculate Base Salary for Hourly Employees dependent on their FT/PT Status
df['Base Salary'] = np.where((df['FT/PT'] == 'P') & (df['Salary Frequency'] == 'Hourly') , round(df['Salary Range From'].mul(20*52), 2), df['Base Salary'])
df['Base Salary'] = np.where((df['FT/PT'] == 'F') & (df['Salary Frequency'] == 'Hourly') , round(df['Salary Range From'].mul(40*52), 2), df['Base Salary'])

#move Base Salary column to desired location in table
column_to_move = df.pop("Base Salary")
df.insert(12, "Annual Base Salary", column_to_move)
```

Python

```python
#Drop unnecessary columns
df = df.drop(['Title Classification', 'Title Code No', 'Salary Range To', 'Salary Range From', 'Salary Frequency', 'Hours/Shift', 'Post Until', 'Posting Updated', 'Process Date'], axis=1)
df = df.drop(['Preferred Skills', 'Additional Information', 'To Apply', 'Recruitment Contact', 'Work Location 1', 'Job Description', 'Minimum Qual Requirements', 'Residency Requirement'], a
```

Python

```python
#rename columns for query managability
df =df.rename(columns={'FT/PT': 'Ft_Pt', 'Job ID' : 'Job_id', '# Of Positions' : 'Num_pos', 'Business Title': 'Business_title', 'Civil Service Title' : 'Civil_title'})
df =df.rename(columns={'Job Category' : 'Category', 'Career Level' : 'Career_level', 'Posting Type': 'Posting_type', 'Annual Base Salary' : 'Yearly_salary'})
df =df.rename(columns={'Work Location' : 'Location', 'Division/Work Unit' : 'Division', 'Posting Date' : 'Post_Date'})
```

Python

Schema

Job_NYC_Postings		
PK	Job ID	INT
	Agency	VARCHAR(30)
	Posting Type	VARCHAR(25)
	# of Positions	INT
	Business Title	VARCHAR(35)
	Civil Service Title	VARCHAR(50)
	Level	VARCHAR(40)
	Job Category	VARCHAR(1)
	FT/PT	VARCHAR(35)
	Career Level	VARCHAR(25)
	Annual Base Salary	DOUBLE
	Work Location	VARCHAR(55)
	Division/Work Unit	VARCHAR(35)
	Posting Date	DATE

CLEANED DATAFRAME

	Job_id	Agency	Posting_type	Num_pos	Business_title	Civil_title	Level	Category	Ft_Pt	Career_level	Yearly_salary	Location	Division	Post_Date
0	589025	DEPARTMENT OF TRANSPORTATION	External	2	Supervisor Highway Repairer	SUPERVISOR HIGHWAY REPAIRER	00	Building Operations & Maintenance	F	Experienced (non-manager)	83932.8	55 Water St Ny Ny	Preventive Maintenance	06/14/2023
1	552763	DEPT OF ENVIRONMENT PROTECTION	Internal	1	Asset Management Program Planner	ASSOCIATE PROJECT MANAGER	02	Engineering, Architecture, & Planning	F	Experienced (non-manager)	74650.0	96-05 Horace Harding Expway	ASSET MANAGEMENT PLANNING	09/29/2022
2	595336	HRA/DEPT OF SOCIAL SERVICES	External	7	CASE MANAGEMENT SUPERVISOR	ASSOCIATE BENEFITS OPPORTUNITY	01	Administration & Human Resources Social Services	F	Experienced (non-manager)	56677.0	95 Evergreen Ave Brooklyn	FIA Operations-NM	07/27/2023
3	597592	DEPT OF ENVIRONMENT PROTECTION	Internal	1	ASSISTANT PROJECT MANAGER	PROJECT MANAGER INTERN#	00	Engineering, Architecture, & Planning Public S...	F	Experienced (non-manager)	56313.0	465 Columbus Ave. Valhalla, Ny	UPSTATE CAPITAL CONSTR. - IFA	08/17/2023
4	607064	DEPARTMENT OF CITY PLANNING	Internal	1	Database Administrator	COMPUTER SPECIALIST (SOFTWARE)	03	Technology, Data & Innovation	F	Experienced (non-manager)	115854.0	120 Broadway, New York, NY	ITD-Application Service Mgmt	09/25/2023
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5476	618573	BRONX DISTRICT ATTORNEY	External	1	Supervisor of Onboarding	ADMINISTRATIVE MANAGER	M2	Administration & Human Resources Communication...	F	Manager	95000.0	198 E161St Street Bronx N.Y.	Support Staff	12/04/2023
5477	620496	DEPT OF HEALTH/MENTAL HYGIENE	External	1	First Line Supervisor, Bureau of Hepatitis, HI...	SUPERVISING PUBLIC HEALTH ADVI	00	Constituent Services & Community Programs Heal...	F	Experienced (non-manager)	59301.0	42-09 28th Street	STI Programs	12/19/2023
5478	623202	OFFICE OF THE COMPTROLLER	Internal	1	Research Support Intern for the Chief Investme...	SUMMER COLLEGE INTERN	00	Finance, Accounting, & Procurement Policy, Res...	P	Student	16120.0	1 Centre St., N.Y.	Asset Management	01/12/2024

05 »

Analysis

Using: Pandas, Numpy, Sqlite3, matplotlib and seaborn

Top Paying NYC Jobs

```
query = 'SELECT Agency, Business_title, Civil_title, Yearly_salary FROM NYC_job_postings WHERE Yearly_salary >= 100000 ORDER BY Yearly_salary DESC'
pd.read_sql_query(query, engine).drop_duplicates()
```

✓ 0.0s

	Agency	Business_title	Civil_title	Yearly_salary
0	DEPARTMENT OF CORRECTION	Deputy Commissioner, Training & Development	DEPUTY COMMISSIONER (DOC)	225000.00
1	DEPT OF ENVIRONMENT PROTECTION	Deputy Commissioner, Environmental Compliance	DEPUTY COMMISSIONER (DEP)	225000.00
3	DEPT OF ENVIRONMENT PROTECTION	Deputy Commissioner, Bureau of Customer Services	DEPUTY COMMISSIONER (DEP)	225000.00
6	BRONX DISTRICT ATTORNEY	Chief Human Resources Officer	ADMINISTRATIVE LABOR RELATIONS	207500.00
7	BRONX DISTRICT ATTORNEY	Bureau Chief of Recruitment	ASSISTANT DISTRICT ATTORNEY (S	207500.00
10	HOUSING PRESERVATION & DVLPMNT	Deputy Commissioner, Office of Asset and Prope...	DEPUTY COMMISSIONER (HOUSING P	207000.00
12	HOUSING PRESERVATION & DVLPMNT	Deputy Commissioner, Office of Policy and Stra...	DEPUTY COMMISSIONER (HOUSING P	207000.00
14	DEPARTMENT OF CORRECTION	Assistant Commissioner, Facility Operations	FACILITY SUPERVISOR (CORRECTIO	201587.00
16	DEPARTMENT OF CORRECTION	Associate Commissioner - Legal Matters	ASSOCIATE COMMISSIONER (DOC)	200000.00
17	DEPARTMENT OF CORRECTION	Deputy Commissioner, General Counsel	EXECUTIVE AGENCY COUNSEL	200000.00
22	DEPARTMENT OF CORRECTION	Associate Commissioner, Operations	ASSOCIATE COMMISSIONER (DOC)	200000.00
24	DEPARTMENT OF CORRECTION	Chief Surgeon, Health Management Division	CITY MEDICAL DIRECTOR	195000.00
25	DEPARTMENT OF CORRECTION	Deputy Commissioner, Chief Information Officer	DEPUTY COMMISSIONER (DOC)	195000.00
28	HRA/DEPT OF SOCIAL SERVICES	DEPUTY DIRECTOR, STRATEGY AND OPERATIONS	EXECUTIVE DEPUTY ADMINISTRATOR	190753.00
30	BRONX DISTRICT ATTORNEY	Bureau Chief - Chief of Crime Strategies Unit	ASSISTANT DISTRICT ATTORNEY (S	190000.00
32	LAW DEPARTMENT	Deputy Chief, Labor and Employment Law Division	ASSISTANT CORPORATION COUNSEL	186827.00
34	OFFICE OF MANAGEMENT & BUDGET	Associate Director FEDERAL AND STATE GRANTS M...	BUDGET ANALYST (OMB)-MANAGERIA	185886.00
36	DEPARTMENT OF CORRECTION	Associate Commissioner, Enterprise Development...	ASSOCIATE COMMISSIONER (DOC)	180000.00
37	DEPARTMENT OF TRANSPORTATION	Associate Deputy Commissioner	DEPUTY COMMISSIONER (DOT)	180000.00
40	DEPT OF ENVIRONMENT PROTECTION	Assistant Commissioner for Strategy and Innova...	ASSISTANT COMMISSIONER (DEP)	175000.00
41	DEPT OF ENVIRONMENT PROTECTION	Assistant Commissioner, OD&HR	ASSISTANT COMMISSIONER (DEP)	175000.00
43	DEPARTMENT OF CORRECTION	Associate Commissioner - Legal Matters	ASSOCIATE COMMISSIONER (DOC)	175000.00

Top Paying Entry Level Jobs

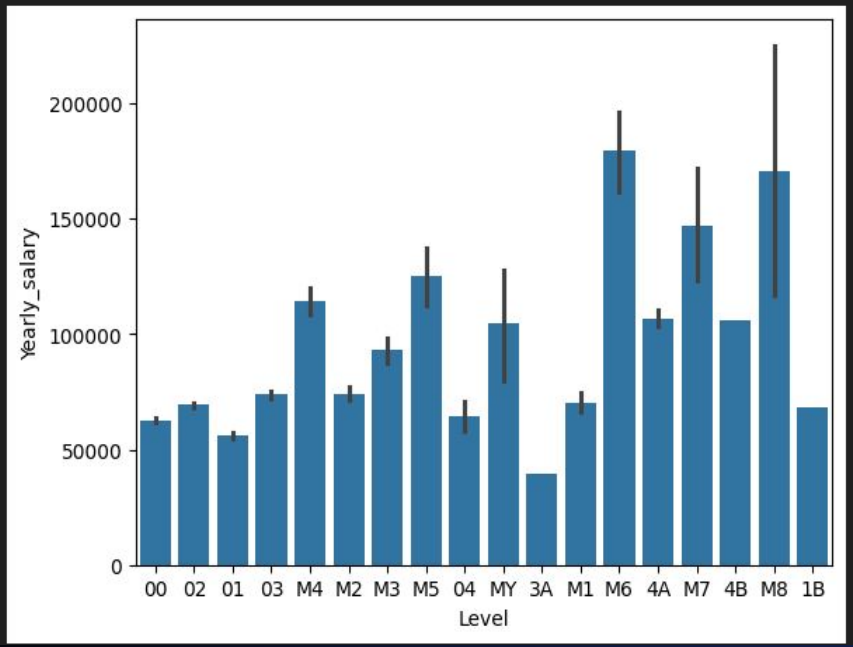
```
query = """SELECT Agency, Business_title, Civil_title, Yearly_salary
        FROM NYC_job_postings
        WHERE Career_level = 'Entry-Level'
        AND Ft_Pt = 'F'
        ORDER BY Yearly_salary DESC"""

pd.read_sql_query(query, engine).drop_duplicates()
```

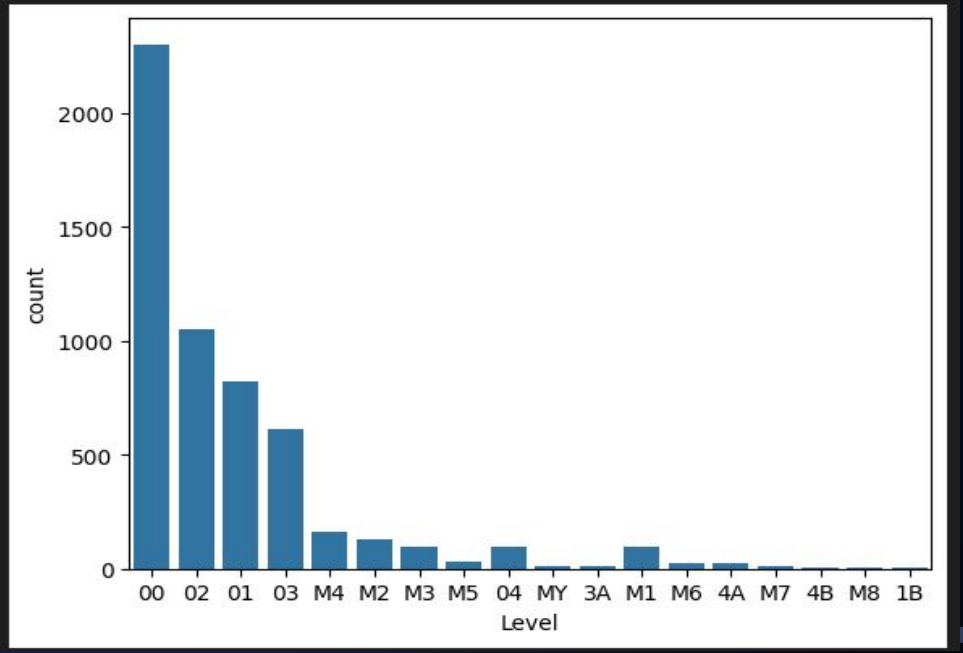
✓ 0.0s

	Agency	Business_title	Civil_title	Yearly_salary
0	DEPT. OF HOMELESS SERVICES	Sheet Metal Worker	SHEET METAL WORKER	120473.6
2	DEPT OF ENVIRONMENT PROTECTION	Plumber	PLUMBER	118268.8
4	NYC HOUSING AUTHORITY	PLASTERER	PLASTERER	100401.6
20	NYC HOUSING AUTHORITY	PLASTERER (HOUSING AUTH-40 HRS/WK)	PLASTERER	93225.6
22	DEPT OF HEALTH/MENTAL HYGIENE	CIR Statistician, Bureau of Immunization Stat...	CITY RESEARCH SCIENTIST	92301.0
24	DISTRICT ATTORNEY RICHMOND COU	Assistant District Attorney	ASSISTANT DISTRICT ATTORNEY (S	85000.0
26	DEPARTMENT OF TRANSPORTATION	Security Analyst	CYBER SECURITY ANALYST	78795.0
28	DEPT OF ENVIRONMENT PROTECTION	Machinist	MACHINIST	77542.4
32	LAW DEPARTMENT	Assistant Corporation Counsel - Administrative...	ASSISTANT CORPORATION COUNSEL	75677.0
34	BRONX DISTRICT ATTORNEY	Assistant District Attorney Fall 2024	ASSISTANT DISTRICT ATTORNEY (S	73579.0
36	DEPT OF ENVIRONMENT PROTECTION	Machinist's Helper	MACHINIST'S HELPER	73236.8
37	DEPT OF ENVIRONMENT PROTECTION	MACHINIST HELPER	MACHINIST'S HELPER	73236.8
42	DEPT OF HEALTH/MENTAL HYGIENE	Budget and Grant Analyst	ASSOCIATE STAFF ANALYST	70611.0
44	DEPT OF HEALTH/MENTAL HYGIENE	Data Analyst, Bureau of Hepatitis, HIV, and STI	CITY RESEARCH SCIENTIST	70087.0
46	OFFICE OF THE COMPTROLLER	Staff Auditor	ACCOUNTANT	70000.0
47	OFFICE OF THE COMPTROLLER	Staff Auditor	MANAGEMENT AUDITOR	70000.0
48	HUMAN RIGHTS COMMISSION	AGENCY ATTORNEY I	AGENCY ATTORNEY	69090.0
50	DEPARTMENT OF CORRECTION	Agency Attorney Interne	AGENCY ATTORNEY INTERNE	68183.0

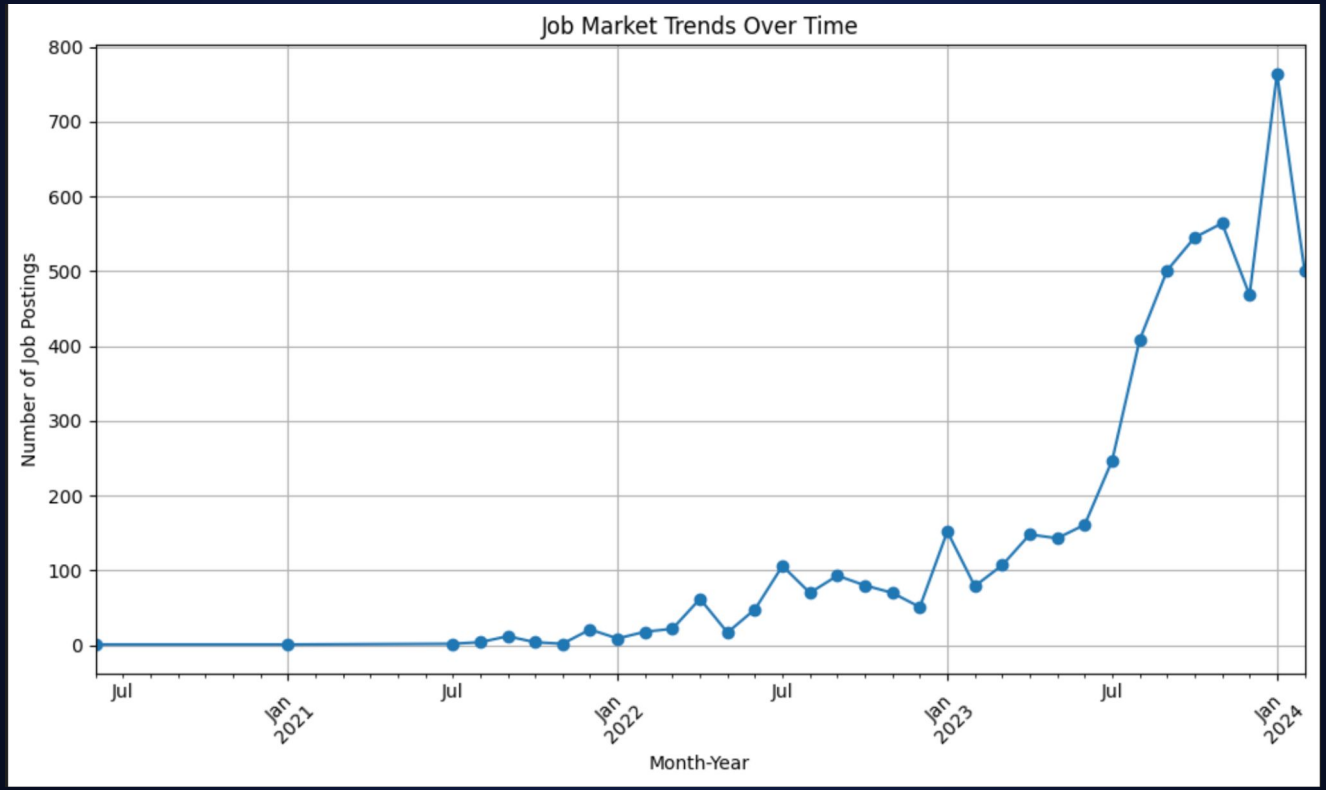
Correlations between Job Salary and number of jobs available.

Job Level in relation to Salary vs. Number of Jobs available for that level

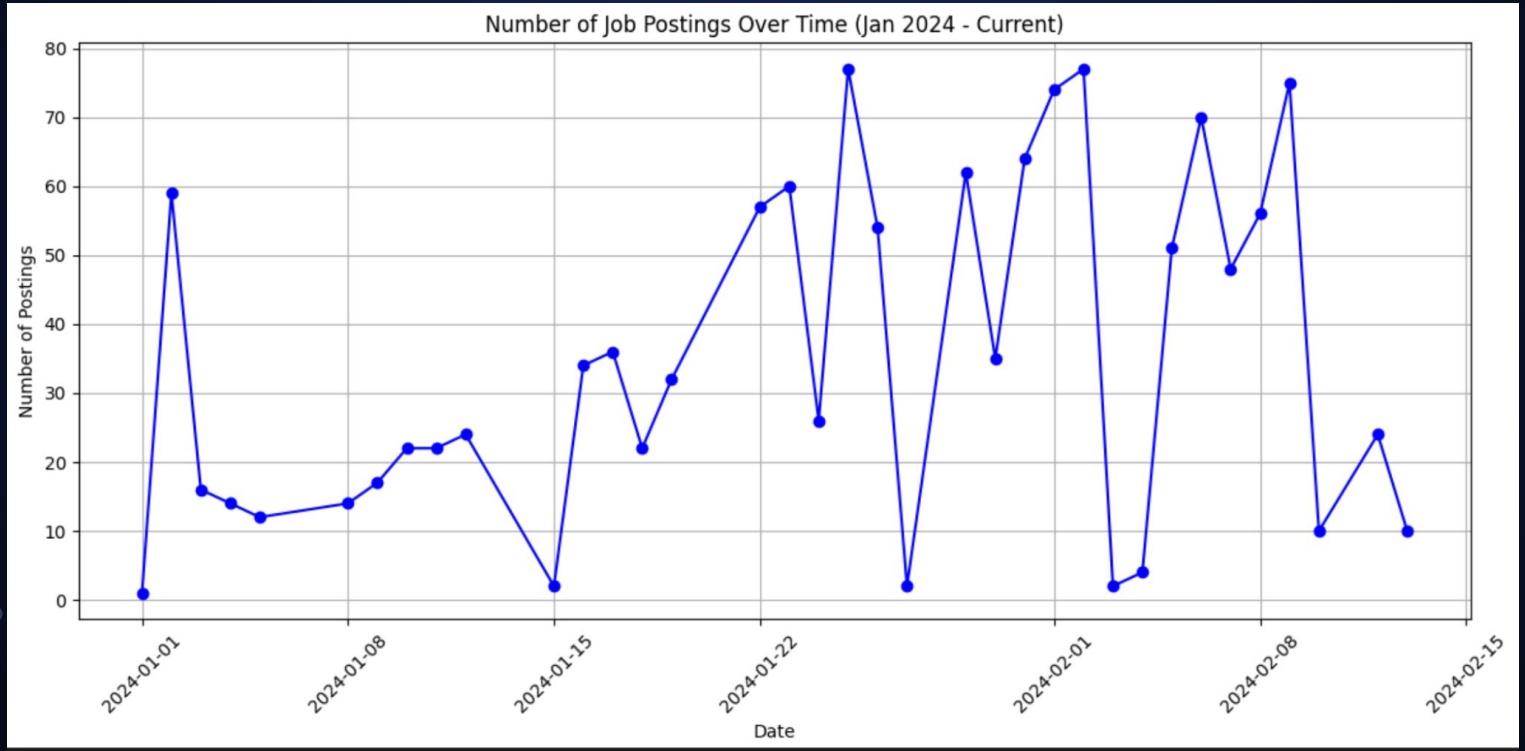




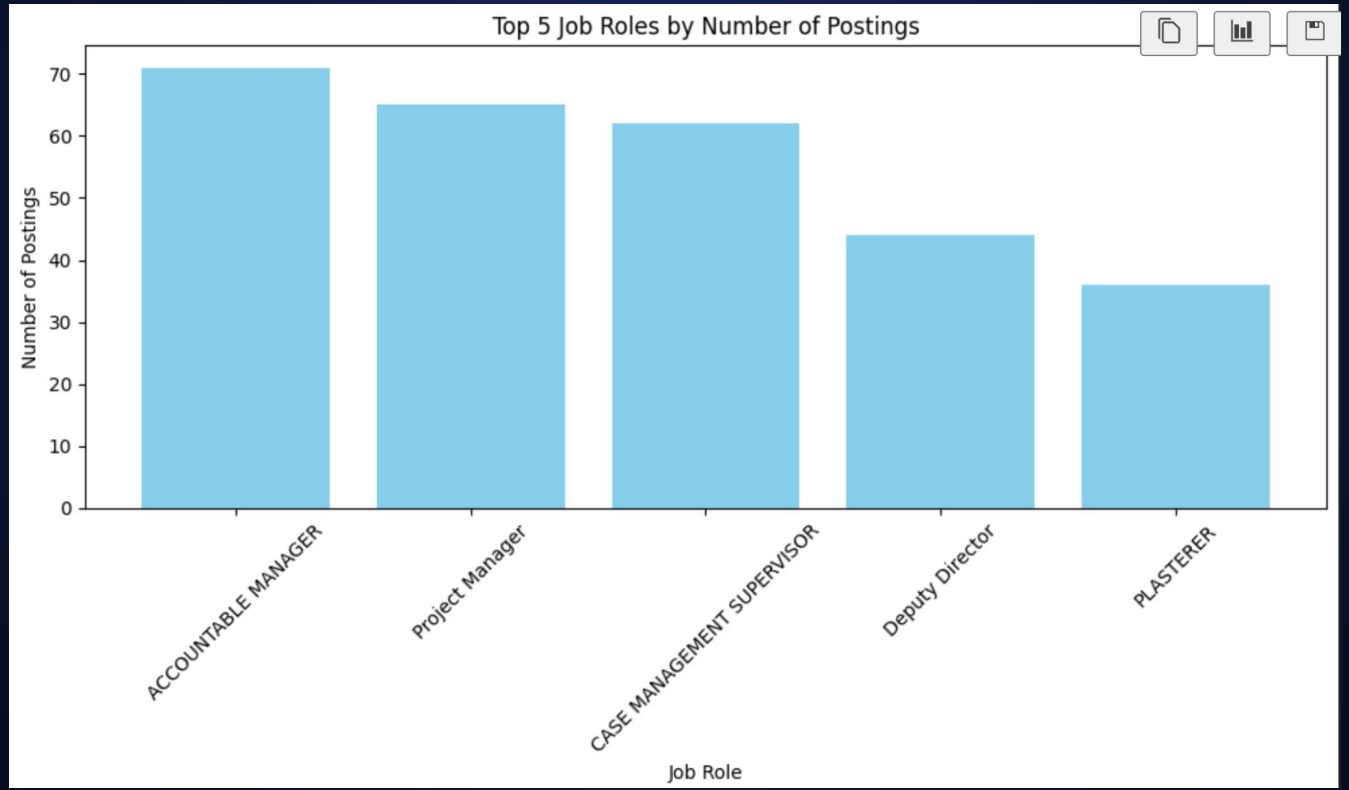
Job Market Trend



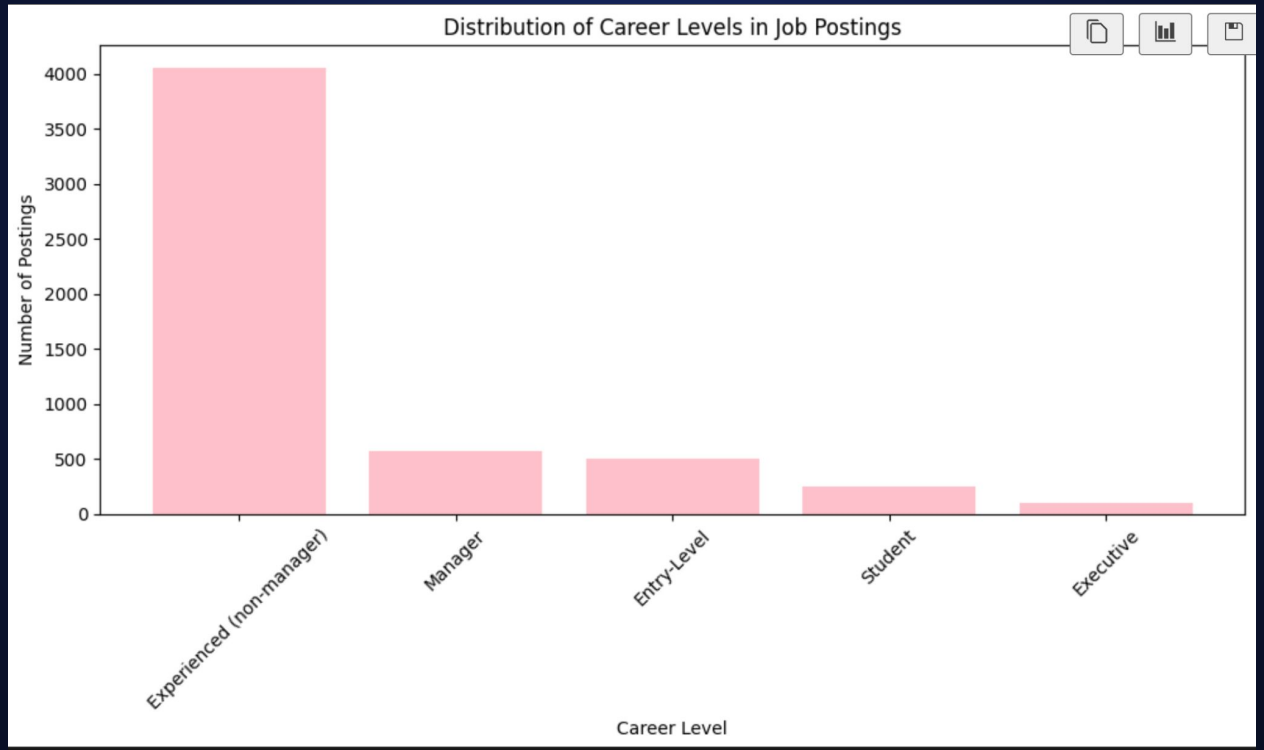
Job Market Trend Jan ‘24 – current



Top 5 Roles by Posting



Career Level Job Postings



Challenges and Resolutions

- Associate the route table to endpoint
- String lengths too great for table transfer
- Creating a table with Glue
- Running out of Lambda memory in free tier

Takeaways:

- Planning - ensuring all aspects of the project are clearly defined
- Organization and keeping code from expanding into too many areas
- Visual ETL jobs make your job easier but come with a price
- Scheduling Lambda's or any service to run is very easy with Eventbridge Scheduler (and free for default scheduling)

Key ideas in market opportunity

Stay ahead of trends

Monitor changes in the marketplace, such as shifts in job requirements to stay competitive and capitalize on new opportunities and trends

Innovative way to use the Data as a Job Seeker

Utilize the data to see what qualifications are needed to land your potential role and upskill from there

Seasonal spikes

Identify times of year when demand for employees may spike. This can be based on quarterly returns, leaving companies with more funds for hiring and potential growth

» Thanks! »

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**