

MELANIE DUONG, JOHNNY NGUYEN

DATA 400 PROJECT FINAL PRESENTATION



1. PROJECT OVERVIEW

A study of how data science job requirements shifted before and after COVID, with a focus on evolving skills, tools, and role expectations in the U.S. job market.

2. RESEARCH QUESTION

The guiding question: How have skills, tools, and role expectations for data science jobs evolved over time, especially during regression time?

3. DATASET

Description of the job postings dataset, its sources, key fields, and preprocessing steps.

4. ANALYSIS & RESULTS

Exploratory analysis, skill frequency comparison, clustering, and regression results.

5. INTERPRETATION & IMPLICATIONS

Key findings, limitations, and implications for data job seekers and educators.



PROJECT OVERVIEW

OBJECTIVE

Measure how Data Scientist job descriptions in the U.S. changed pre vs post COVID (skills + language).

DATA SOURCE

Jobspikr U.S. Data Scientist postings (2018–2019) + Google Search Jobs postings (2022–2025).

METHODOLOGY

Standardize column names, parse dates, and filter to Data Scientist titles. Utilize TF-IDF and KMeans, linear regression.

OUTCOME

Shows which skills/language increased overtime and whether post-COVID DS postings split into more distinct clusters/types than pre-COVID.

PROJECT OVERVIEW / PROJECT OVERVIEW



RESEARCH QUESTION



How have data science job skill requirements changed before and after COVID, and which skills show the most significant shifts over time?



SKILL TRENDS

Across job postings, how do skill mentions change between pre-COVID and post-COVID periods?

ROLE COMPARISON

How do skill requirements differ across data science roles before and after COVID?

EMERGING PROFILES

Can these changes be summarized into distinct post-COVID skill profiles?



DATA SUMMARY

Two Time Periods

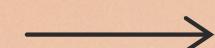
- Pre-COVID (2018–2019)
 - Jobspikr U.S. data science job postings
- Post-COVID (2022–2025)
 - Google Search job postings

Dataset Scope

- U.S.-based data science roles
- Job titles, descriptions, locations, and dates
- Combined structured metadata and text fields

	crawl_timestamp	url	job_title	category	company_name	city	state	country	inferred_city	inferred_state
0	2019-02-06 05:26:22 +0000	https://www.indeed.com/viewjob?jk=fd83355c2b23438c	Enterprise Data Scientist I	Accounting/Finance	Farmers Insurance Group	Woodland Hills	CA	USA	Woodland hills	California
1	2019-02-06 05:33:41 +0000	https://www.dice.com/jobs/detail/Data-Scientist-Luxoft-USA-Inc-Middletown-NJ-07748/newageny/802263	Data Scientist	NaN	Luxoft USA Inc	Middletown	NJ	USA	Middletown	New jersey
2	2019-02-06 05:33:35 +0000	https://www.dice.com/jobs/detail/Data-Scientist-Cincinnati-Bell-Technology-Solutions-New-York-NY-10001/10119424/464780	Data Scientist	NaN	Cincinnati Bell Technology Solutions	New York	NY	USA	New york	New york

Unnamed: 0	index	title	company_name	location	via	description	extensions
0	0	Data Analyst	Meta	Anywhere	via LinkedIn	In the intersection of compliance and analytics, we are seeking an analytical and data-savvy self-starter with a passion for... [15 hours ago, '101K-143K a year', 'Work from home', 'Full-time']	eyJqb2JfdGl0bGUIOiJEYXRhlEFuYWx5c3QiLCJodGlk2NpZCI6IihBR01
1	1	Data Analyst	ATC	United States	via LinkedIn	Job Title: Entry Level Business Analyst / Product Owner U.S. Citizens and those authorized to work in the U.S. are e... [12 hours ago, 'Full-time', 'Health insurance']	eyJqb2JfdGl0bGUIOiJEYXRhlEFuYWx5c3QiLCJodGlk2NpZCI6InJ4WUhw
2	2	Aeronautical Data Analyst	Garmin International, Inc.	Olathe, KS	via Indeed	Overview:\n\nWe are seeking a full-time...\nAeronautical Data Analyst in our Olathe, KS location. In this role, you ... [18 hours ago, 'Full-time']	eyJqb2JfdGl0bGUIOiBZXJvbmF1dGjYWwgRGF0YSBBbmFseXN0liwi+



DATA CAPSTONE

CLEANING STEPS

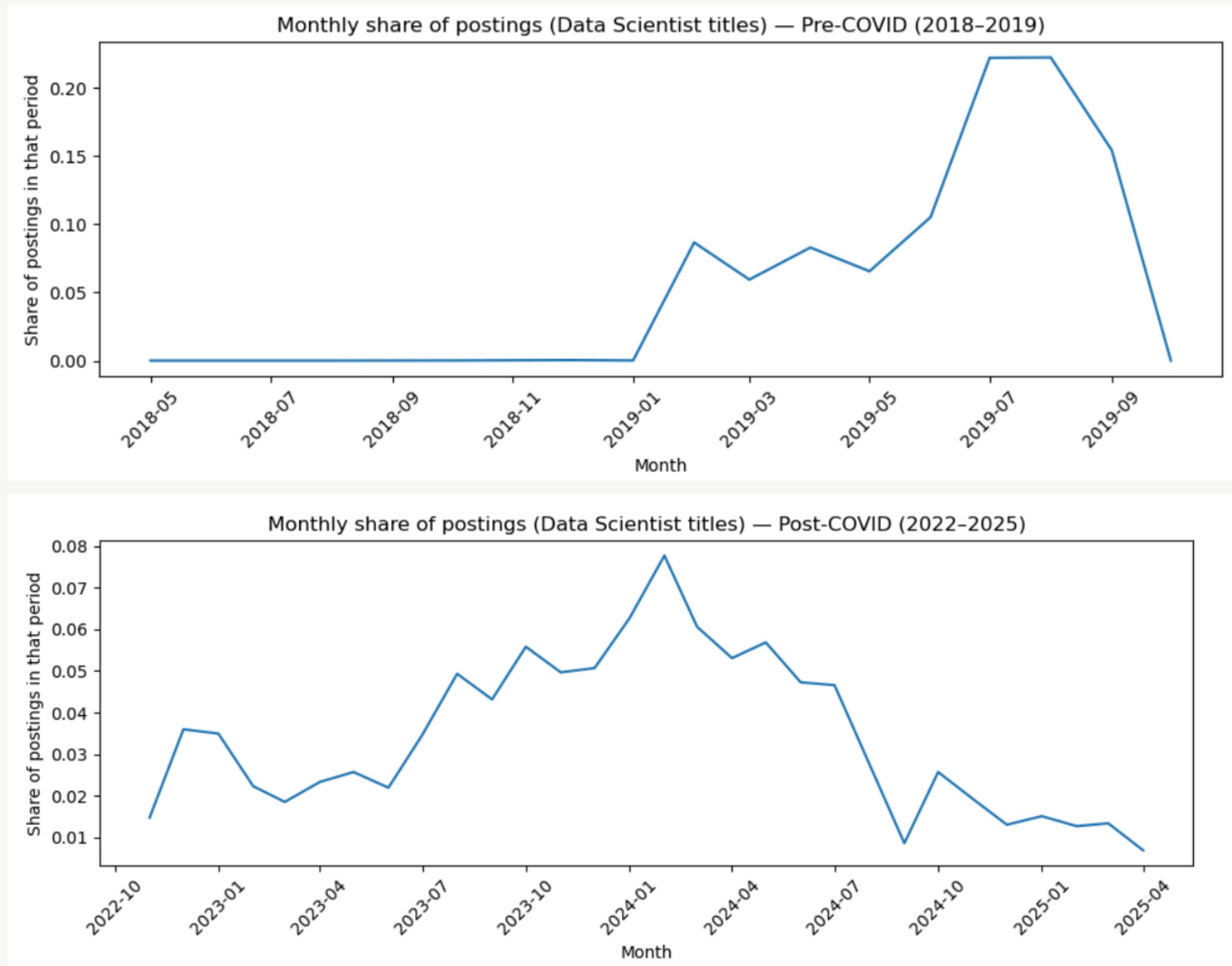
- Standardized datasets into one schema: date, title, description, company, location_raw, source
- Parsed dates + created year and Pre vs Post period labels (kept only 2018–2019 and 2022–2025)
- Cleaned text fields (trim whitespace, remove messy spacing) + dropped rows with missing dates
- Filtered to Data Scientist titles for apples-to-apples comparison
- Saved the final clean table as job_postings_clean.csv

DATA 400

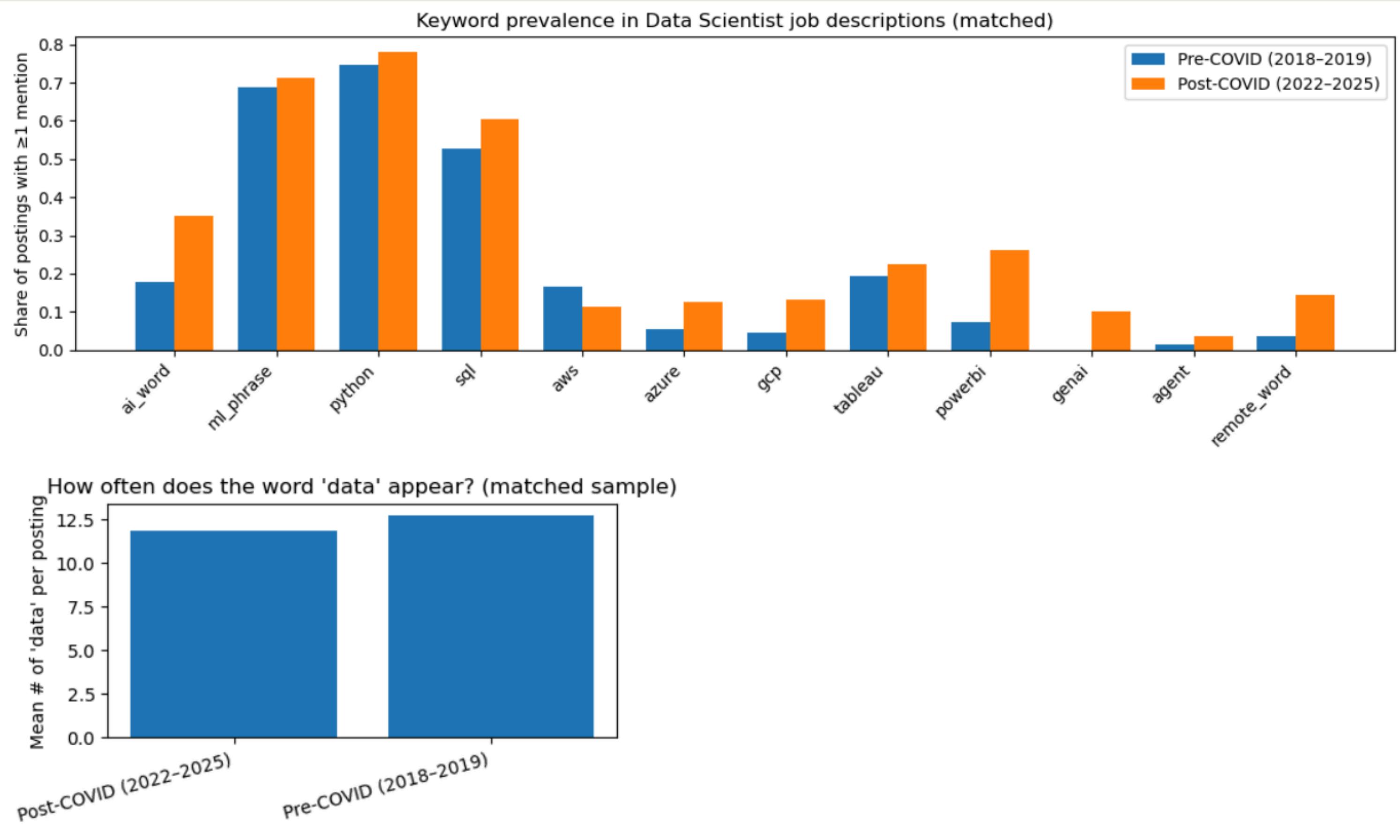
	source	period	date	year	title	company	location_raw	description
0	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Data Analyst III - Full Time - Days	Mercy	Springfield, MO	We're a Little Different Our mission is clear. We bring to life a healing ministry through our compassionate care an...
1	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Sr. Business Data Analyst	New American Funding	United States	Must be able to work on Pacific Standard Time Overview: Are you excited to analyze large datasets and help business ...
2	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Sr. Data Analyst	Medasource	United States	Sr Data Analyst _ Role Summary/Purpose ... The Senior Data Analyst role solves complex business problems, integratio...
3	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Data Analyst	Upwork	Anywhere	Apply at: https://datobite.com/ RESPONSIBILITIES... Managing master data, including creation, updates, and deletion....
4	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Data Analyst - Finance (Remote)	Drip	Anywhere	At Drip, we believe any ecommerce brand should have the tools to connect with their customers and compete online. Le...
5	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Data Analyst I	Neogen Corporation	Anywhere	It's fun to work in a company where people truly believe in what they are doing! Primary function is analysis, qualit...
6	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Data Analyst	Labcorp	Anywhere	Recognized as one of Fortune magazine's 2021 World's Most Admired Companies and named to Fast Company magazine's lis...
7	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Data Analyst / Analytics	ACS	United States	_Data Analyst - 100% Remote - Financial Services!_ We are searching for an experienced data analyst to join our team...
8	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	DATA ANALYST w/INFORMATICA EXPERIENCE	Business Advisory Solutions Group (BASG)	United States	Business Advisory Solutions Group (BASG) is seeking candidates for a Data Analyst / Business Analyst with Informatic...
9	gsearch	Post-COVID (2022–2025)	2022-11-04 03:40:11.151943	2022	Data Analyst II	EDWARD JONES	Bates City, MO	At Edward Jones, we help clients achieve their serious, long-term financial goals by understanding their needs and i...



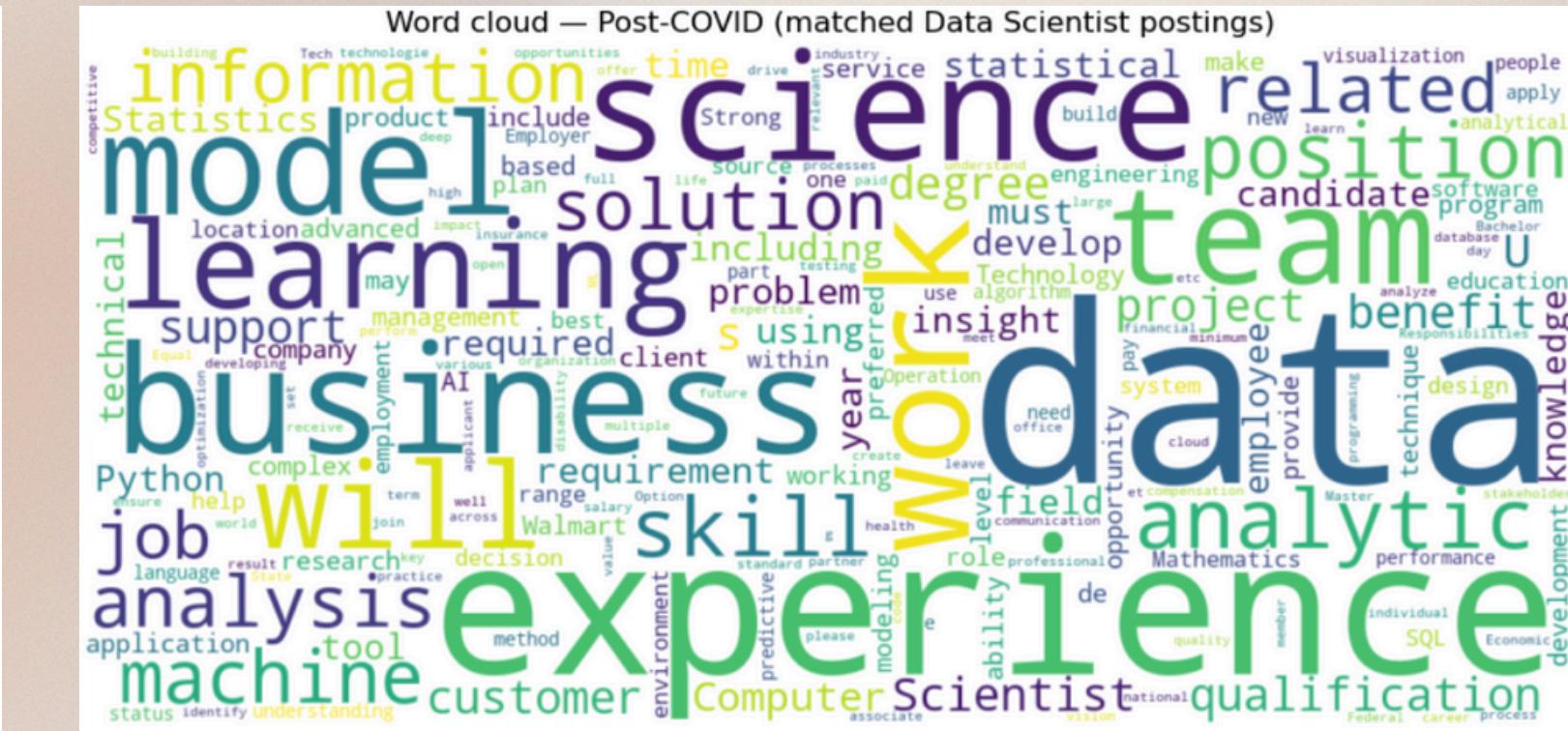
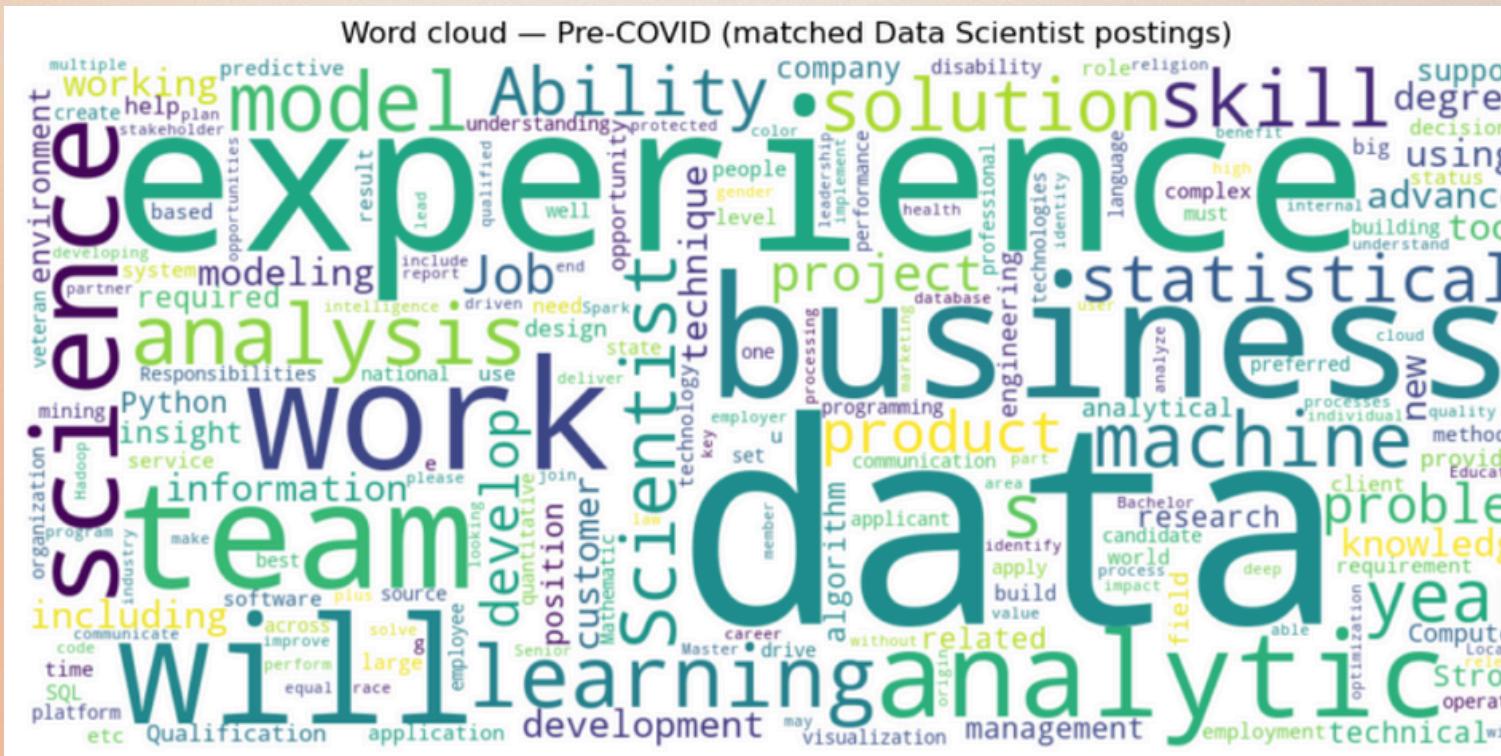
NORMALIZED MONTHLY POSTING PATTERNS (DATA SCIENTIST) – PRE VS POST COVID



SKILL KEYWORD PREVALENCE IN DATA SCIENTIST POSTINGS – PRE VS POST COVID



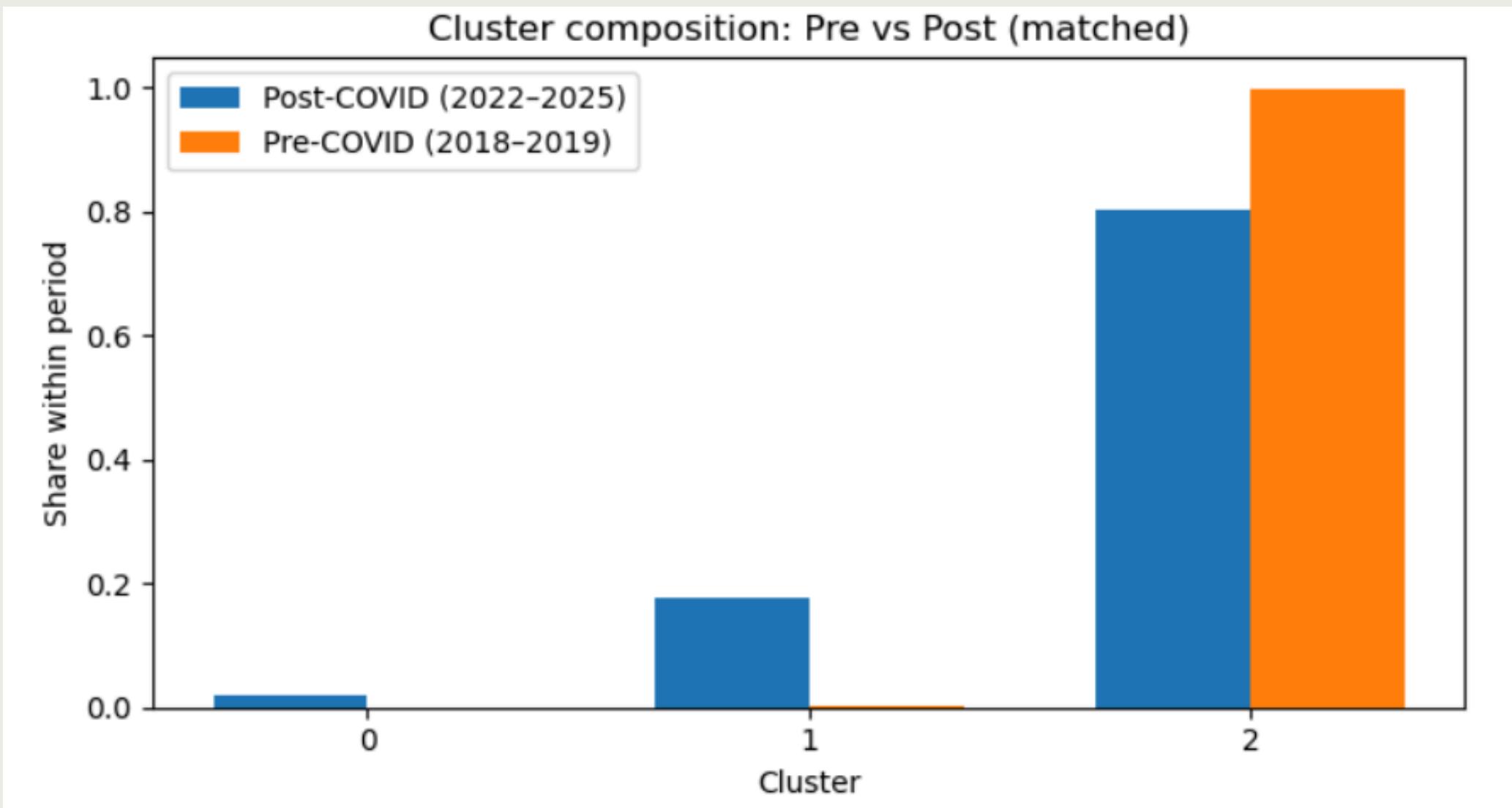
JOB DESCRIPTION LANGUAGE SNAPSHOT – PRE VS POST COVID



- Both periods share the same core DS vocabulary, suggesting baseline expectations stayed consistent
 - Post-COVID descriptions lean more business/team/role-structure oriented.
 - The keyword prevalences are the evidence for specific skill shifts.



RESULTS SNAPSHOT: CLUSTERS AND POST-PRE SKILL SHIFTS

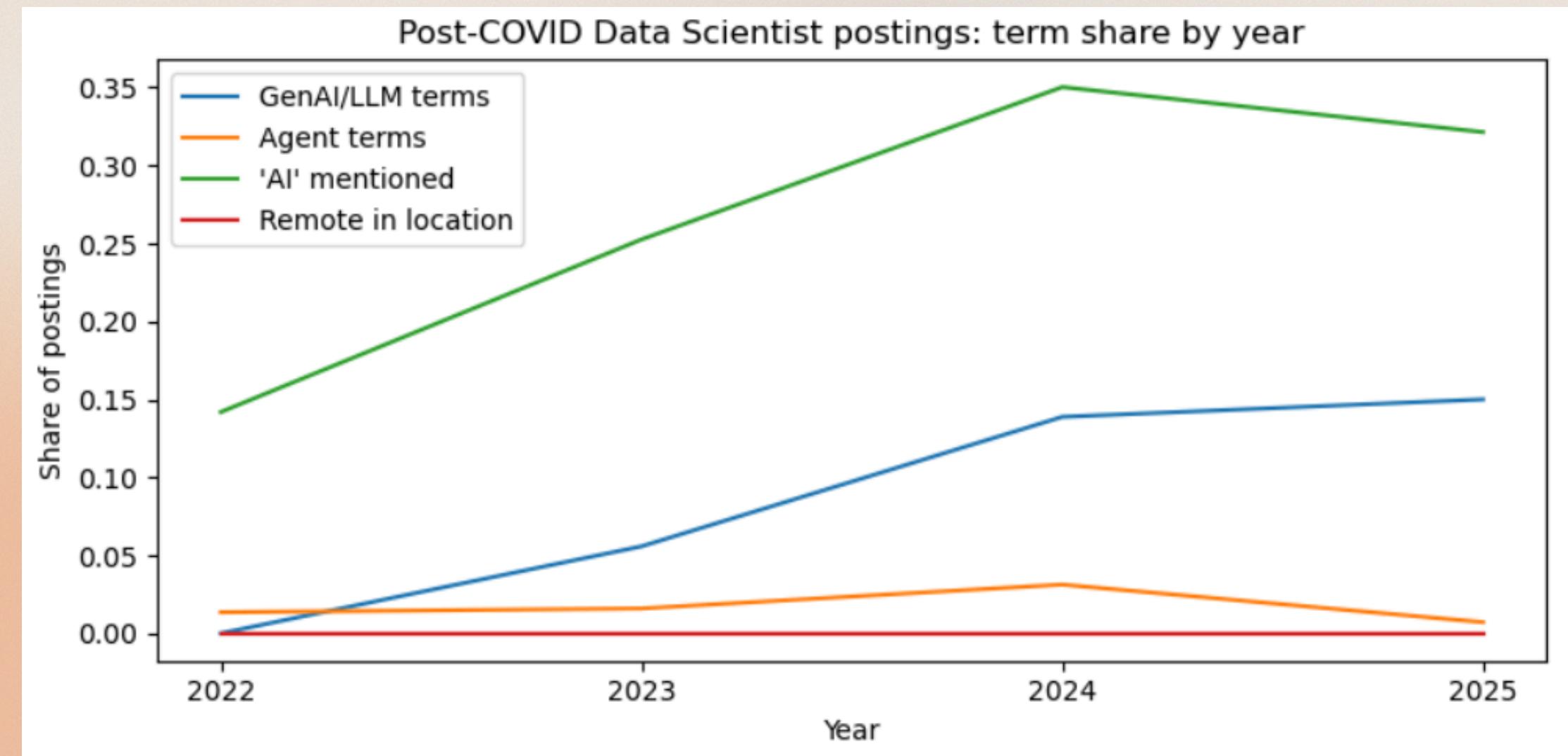


period	cluster	n	share_within_period
0	Post-COVID (2022–2025)	0	8
1	Post-COVID (2022–2025)	1	71
2	Post-COVID (2022–2025)	2	322
3	Pre-COVID (2018–2019)	1	1
4	Pre-COVID (2018–2019)	2	399

outcome	estimated_Post_minus_Pre
4 powerbi_any	0.189526
1 ai_word_any	0.172070
2 genai_any	0.099751
3 gcp_any	0.087282
6 azure_any	0.072319
5 aws_any	-0.049875
0 data_word	-0.890274



GROWTH OF AI & GENAI MENTIONS IN DS JOB DESCRIPTIONS



- GenAI/LLM mentions rise sharply from 2022 → 2024, then level off in 2025.
- “AI” mentions increase a lot overall.
- Agent terms stay low across all years.
- “Remote” in the location field is basically flat/near zero here.



ETHICAL, LEGAL, & SOCIETAL IMPLICATIONS

- Bias and signaling effects in job description language
- Uneven access to emerging skills (e.g., AI and GenAI tools)
- Limitations of scraped job posting data
- Risk of inflating “required” skill expectations
- Responsible use of publicly available hiring data

IMPLICATIONS FOR STAKEHOLDERS

- Guidance for students and early-career job seekers
- Curriculum alignment for data science education
- Insight for career services and advising
- Benchmarking trends for academic programs
- Transparency signals for employers



CONCLUSIONS

- Job postings reflect real-time changes in labor market expectations
- Core data skills remained stable before and after COVID
- AI and GenAI language increased substantially post-COVID
- Job descriptions show greater specialization and diversity
- Results highlight how external shocks reshape technical roles

