Final Report: Trends and Skill Analysis in Data Science Job Postings

- ✓ **Intern Name:** Vedant Waghmare
- ✓ **Internship Title:** AI & Data Science Intern – GennieSphere
- ✓ **Project Duration:** 10 Days
- ✓ **Tools Used:** Python, Pandas, NumPy, Seaborn, Matplotlib, spaCy (NLP), Google Colab

➢ Project Overview:

In this 10-day internship project, I explored trends in the data science job market using a real-world dataset of job postings collected between 2020 and 2023. My primary focus was to apply data science techniques — including data cleaning, natural language processing (NLP), and data visualization — to uncover insights about hiring patterns, in-demand skills, job roles, and geographical hiring trends.

This project helped me bridge the gap between academic concepts and real-world applications, simulating what data scientists and analysts do in talent analytics and market intelligence domains.

➢ Key Insights:

1. Most In-Demand Job Titles

- The most frequently occurring roles included Data Scientist, Data Analyst, and Machine Learning Engineer.
- Grouped similar variations (like “Sr Data Analyst” and “Analyst - Data”) using string matching and normalization.
- These titles dominated postings across regions, especially in the United States, India, and Europe.

2. Most Demanded Skills

- Top Technical Skills: Python, SQL, Spark, Hadoop, Azure, Excel
- Entry-Level Jobs: Emphasize SQL, Excel, Python
- Senior Roles: Demand cloud platforms, scalable tools, and ML frameworks
- This shows how technical depth and architectural thinking become increasingly important as one advances.

3. Regional Demand Trends

- Top hiring countries: United States, India, United Kingdom
- Remote/hybrid opportunities inferred through employee residence field
- Global demand shows data-driven decision making is a widespread shift

4. Company-Specific Skill Preferences

- Medium-sized companies showed the highest demand for diverse skills
- Large firms leaned toward deployment-oriented and big data technologies
- Small companies preferred generalists proficient in Python and SQL

5. Experience-Level Skill Evolution

- Entry-level jobs focus on SQL, Excel, Tableau
- Senior roles require cloud, scalable ML, and deployment tools
- Reflects a shift from task execution to system design with experience

➢ Visual Insights (with Graphs):

I used bar charts, WordClouds, and grouped plots to communicate skill trends, job titles, and company-wise comparisons.
(Visuals are available in the Jupyter Notebook and can be attached separately.)

➢ Final Deliverables:

1. Cleaned Dataset – cleaned_ds_jobs.csv
2. Jupyter Notebook – Documented and visualized
3. Final Report – This Word document
4. GitHub Repository – Optional

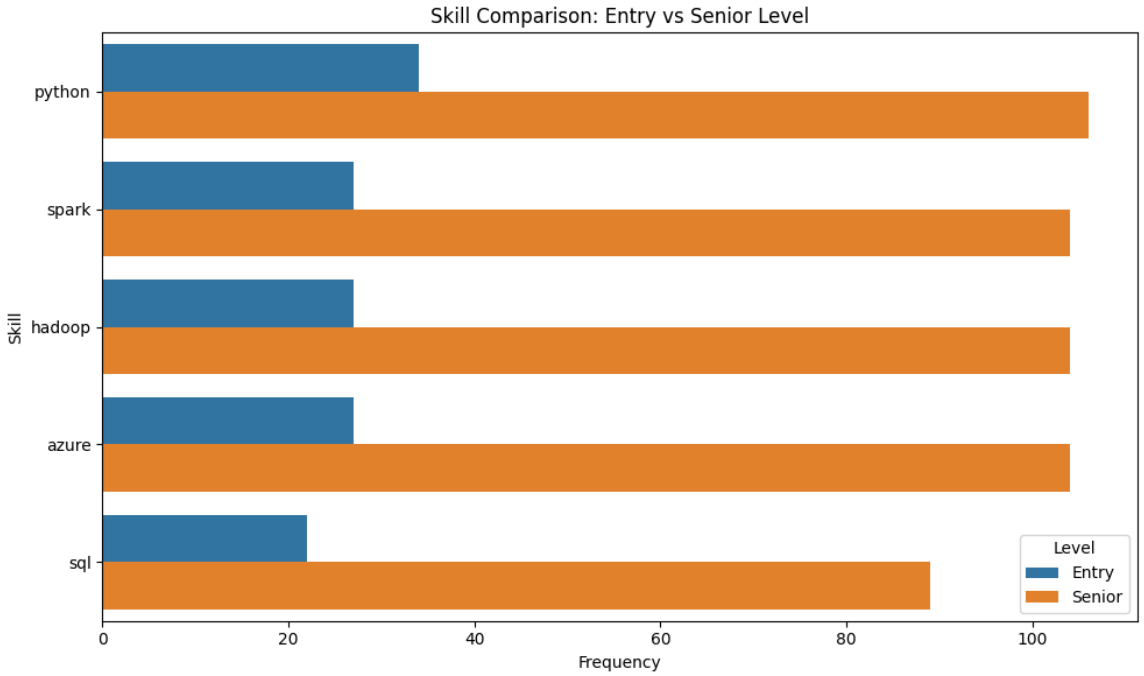
➢ What I Learned:

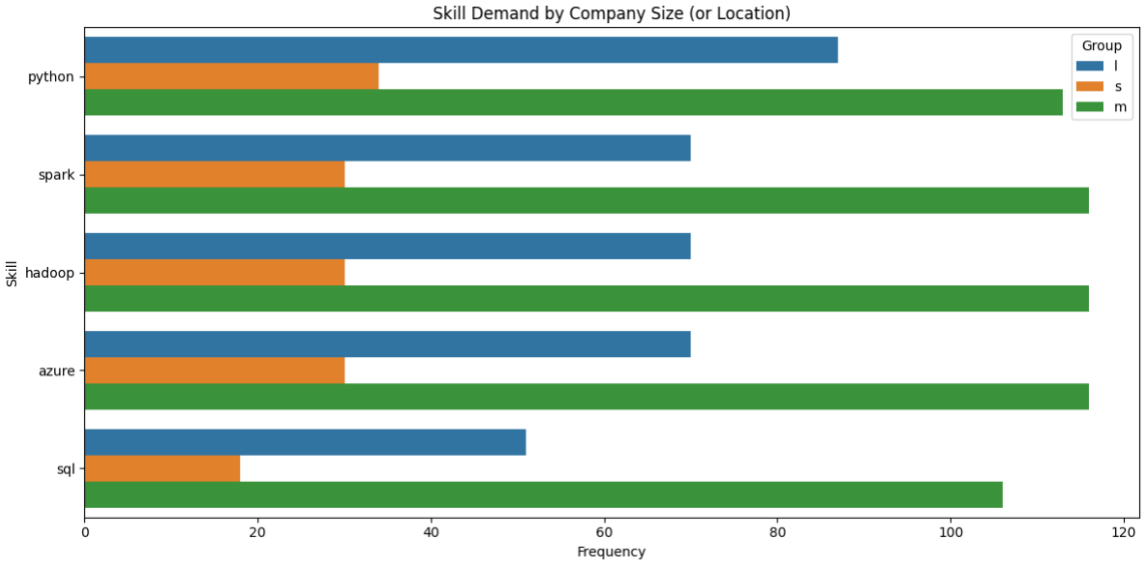
1. Hands-on experience with real-world data wrangling
2. NLP-based skill extraction using spaCy
3. Creating visualizations that tell stories from data
4. Structured reporting and insight communication
5. Better understanding of how data science informs hiring intelligence

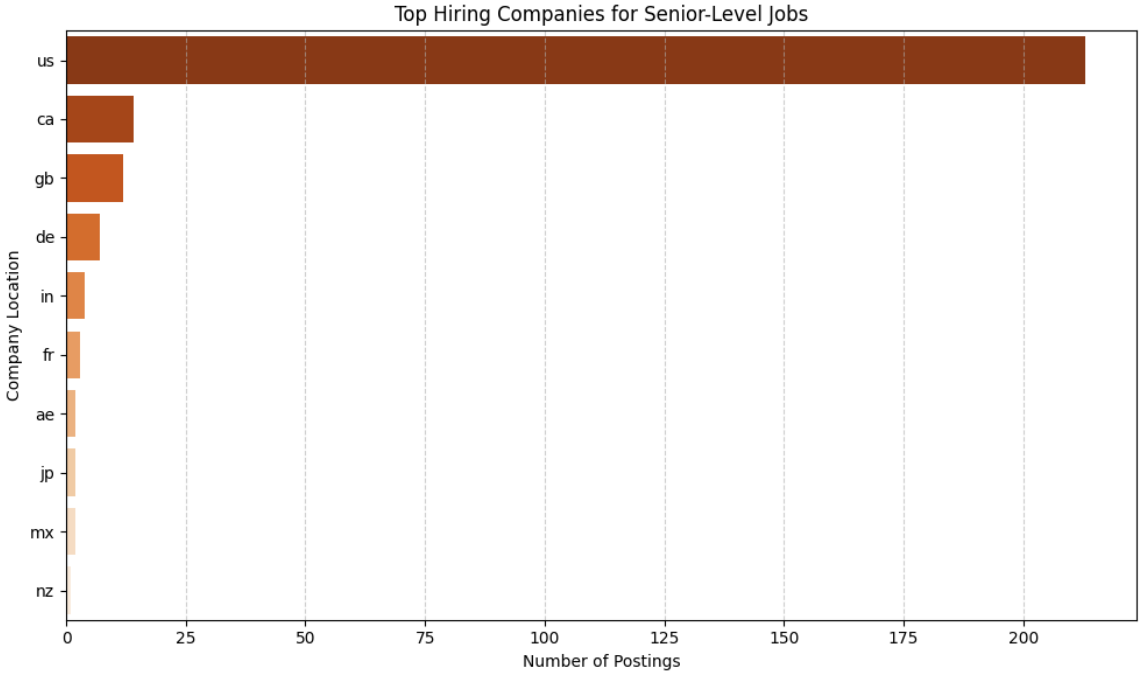
➢ Final Note:

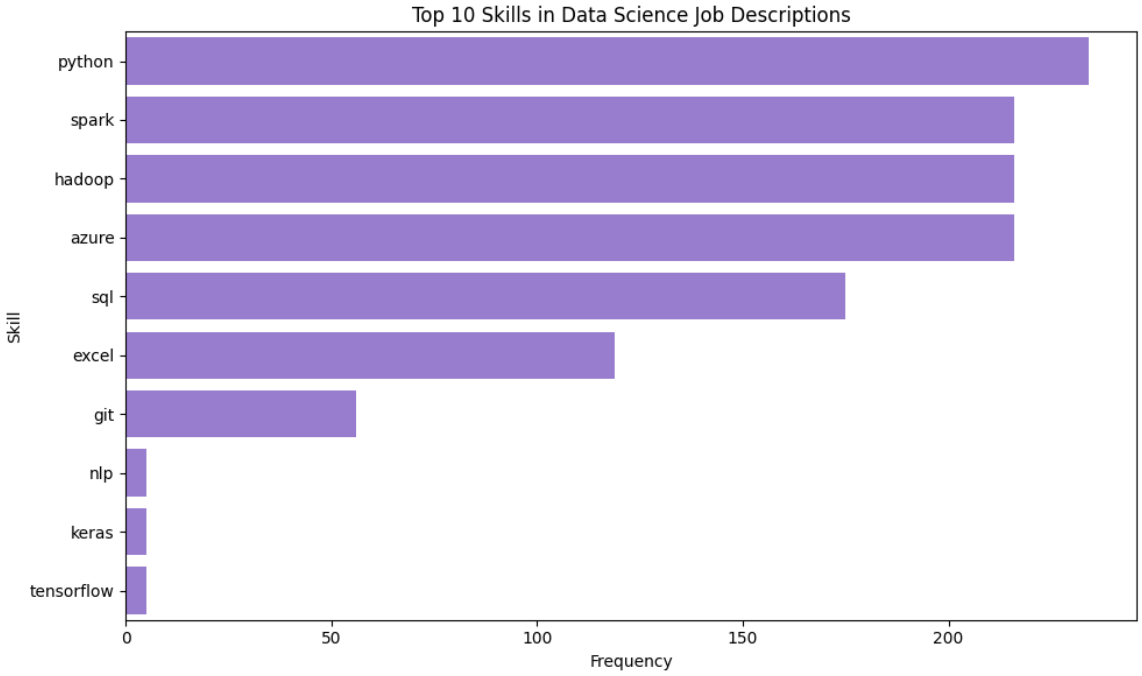
Thank you for the opportunity to work on this meaningful and insightful project. It has helped me grow not only in my technical skills but also in presenting analytical findings to non-technical audiences.

➢ Visualizations:






Top Hiring Companies for Senior-Level Jobs
Company Location
us
ca
gb
de
in
fr
ae
jp
mx
nz
0
25
50
75
100
125
150
175
200
Number of Postings


Top 10 Skills in Data Science Job Descriptions
Skill
python
spark
hadoop
azure
sql
excel
git
nlp
keras
tensorflow
0
50
100
150
200
Frequency