FINAL SUMMARY REPORT

Day 1: Data Overview

Tasks Performed:

- *Loaded the Glassdoor job dataset*
- *Inspected shape, column names, and data types*
- *Checked for missing values and nulls*
- *Generated descriptive statistics of each column*

Output:

- *A complete structural summary of the dataset was generated.*
- *It revealed a well-sized dataset with thousands of job entries, key fields like Job Title, Company Name, Location, and job descriptions.*
- *Helped understand the distribution and completeness of data before proceeding with cleaning.*

Day 2: Data Cleaning

Tasks Performed:

- *Removed duplicate entries to avoid skewed insights*
- *Standardized text fields by converting them to lowercase and trimming spaces*
- *Filtered records to retain only data-relevant roles (keywords: data, AI, ML, etc.)*

Output:

- *A refined dataset named **cleaned_jobs.csv** containing only valid and relevant job postings.*
- *Clean, uniform text fields which improved downstream processing and skill extraction accuracy.*

Day 3: Top Job Titles

Tasks Performed:

- *Simplified job titles into broader categories (e.g., "Senior Data Analyst" → "Data Analyst")*
- *Counted occurrences of each role and generated visualizations*

Output:

- *Identified top job titles: **Data Analyst**, **Data Scientist**, **ML Engineer**, **Data Engineer***
- *Created:*
    - *A **bar chart** to show top 10 job roles by volume*
    - *A **pie chart** for proportional representation*

- *A **word cloud** depicting job title frequency visually*

These visuals clarified which roles are in demand across the job market.

Visualizations:

Figure 1- (A word cloud depicting job title frequency visually)

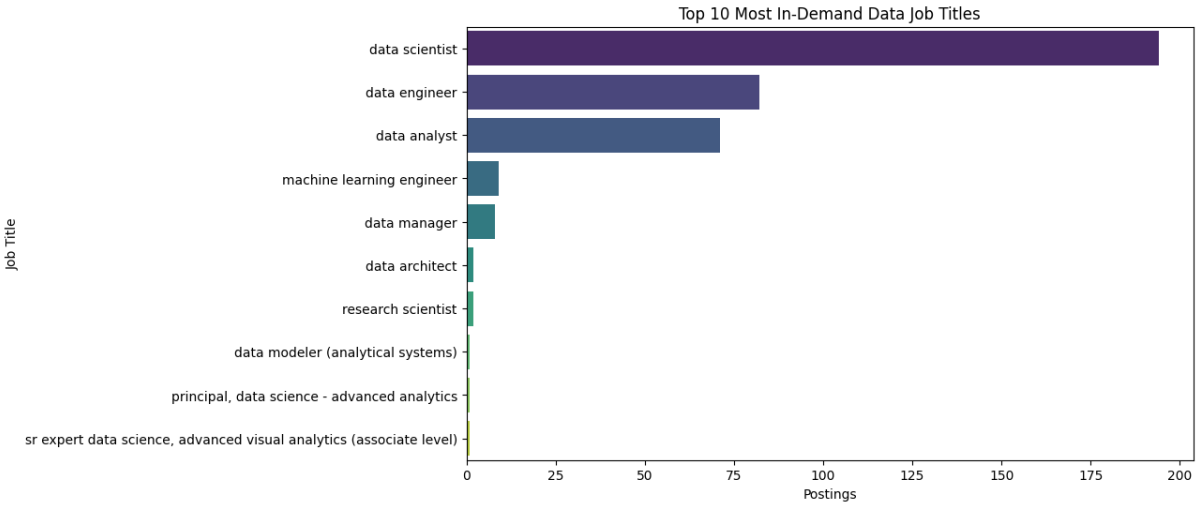


Figure 2 - (A bar chart to show top 10 job roles by volume)

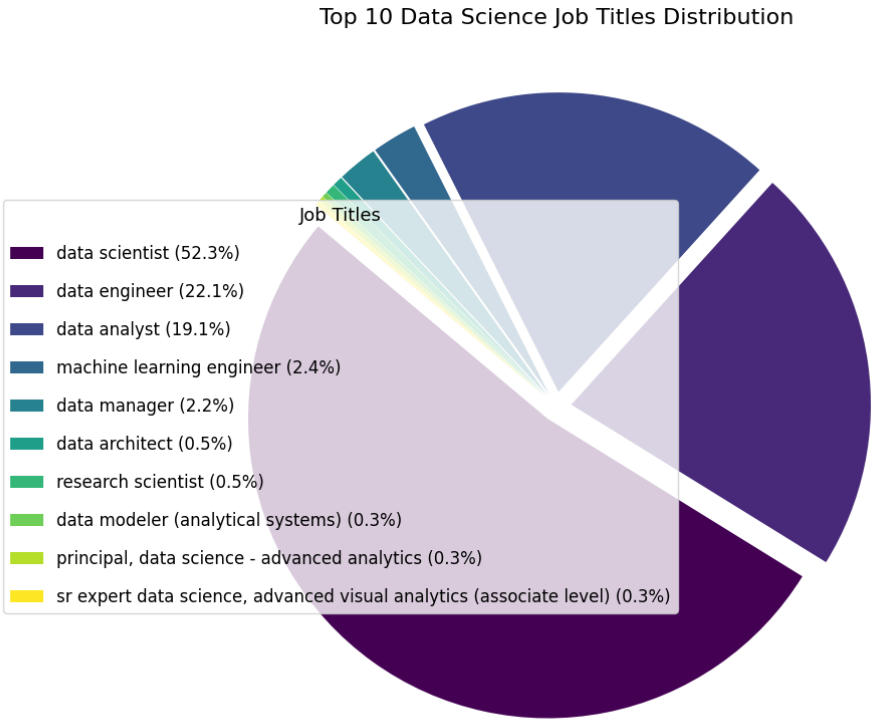


Figure 3 – (A pie chart for proportional representation of job titles)

Day 4: Top Hiring Companies

Tasks Performed:

- *Analysed job counts by company*
- *Segmented data for **global top 10 companies** and **India-specific (entry-level) hiring trends***

Output:

- *Visualizations revealed that companies(from dataset) like **massmutual**, , **pnnl**, **mitre**, and **novetta** led in hiring.*
- *Separate charts were generated:*
    - ***Global bar chart** of top 10 hiring companies*
    - ***India-entry-level specific view**, helping fresher candidates identify target employers*

Visualizations:

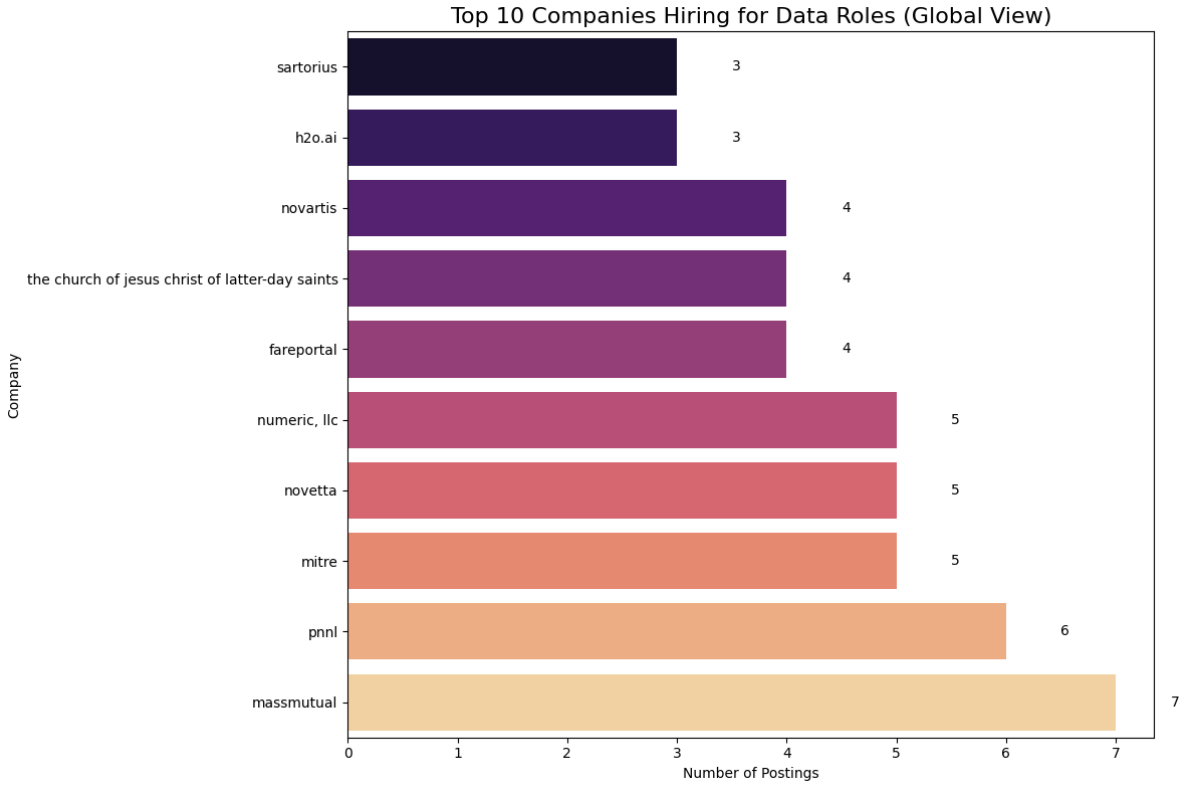
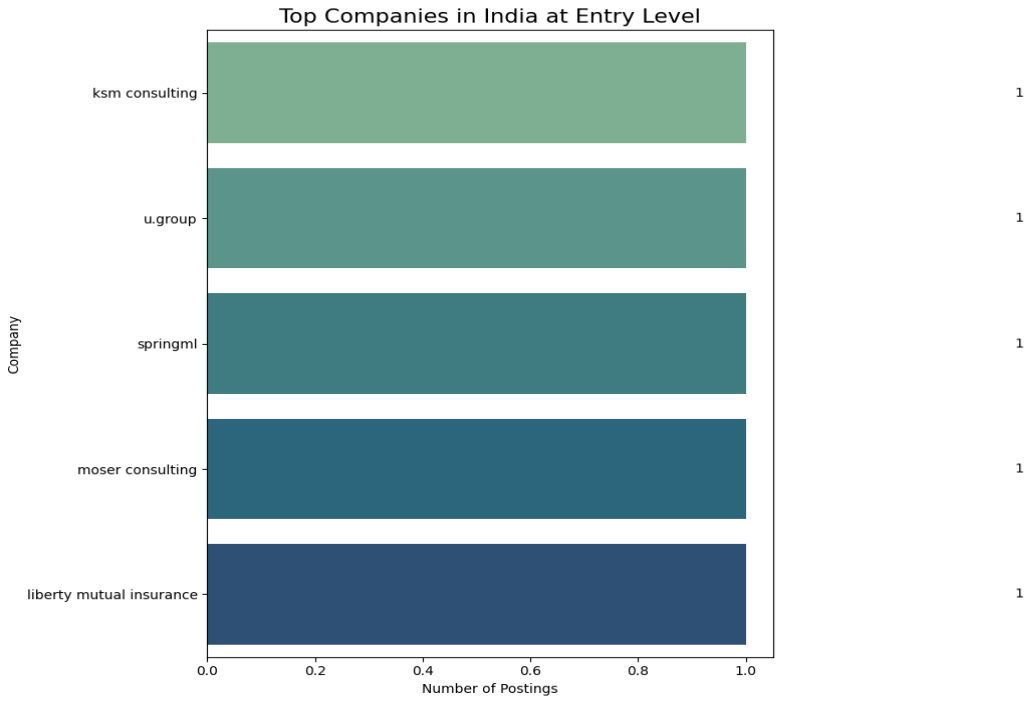


Figure 4 – (Global bar chart of top 10 hiring companies)



- Figure 5 – (***India-entry-level specific view****, helping fresher candidates identify target employers)*

Day 5: Location Insights

Tasks Performed:

- *Split the Location field into City and Country*
- *Identified top cities and countries hiring for data roles*
- *Visualized via:*
    - *Bar chart(top cities, top countries)*
    - ***Plotly choropleth map** for country-wise demand*
    - ***Folium interactive map** for top Indian cities*

Output:

- *India, USA, and Germany emerged as top hiring countries*
- *Cities like **New York**, **San Francisco**, and **Chicago** dominated hiring in the World.*
- *The maps provided a geographic understanding of job demand concentration*

Visualizations:

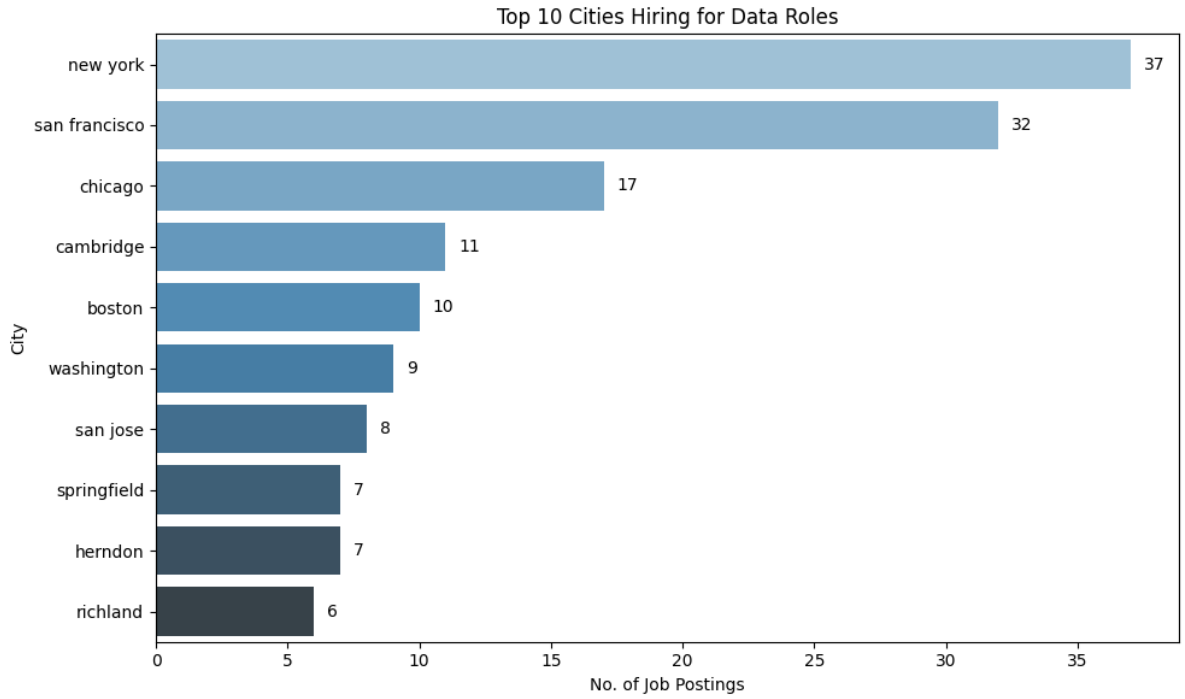


Figure 6 – (Top Cities Worldwide hiring for data roles)

File1-Plotly choropleth map for country-wise demand

File2-Folium interactive map for top Indian cities

Day 6: Skills Extraction

Tasks Performed:

- *Used NLP techniques and a custom keyword dictionary to extract **technical and soft skills** from job descriptions*
- *Aggregated frequency of each skill*
- *Visualized with:*
  - ***Word cloud** for extracted skills*
  - ***Bar chart** for top 20 skills*

Output:

- *Most common technical skills: **Python**, **Excel**, **SQL**, **AWS**, **Tableau**.*
- *Top soft skills: **Communication**, **Problem Solving**, **Teamwork.***
- *These outputs clarified what tools and abilities employers value most in data-related roles*

Visualizations:

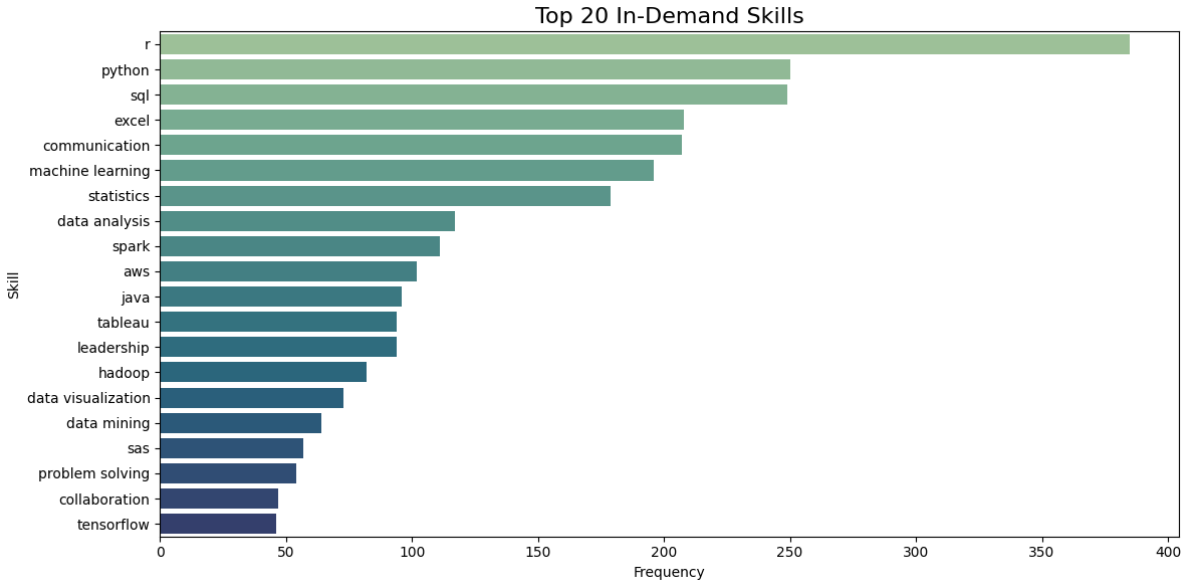


Figure 7 – (Bar chart for top 20 skills)

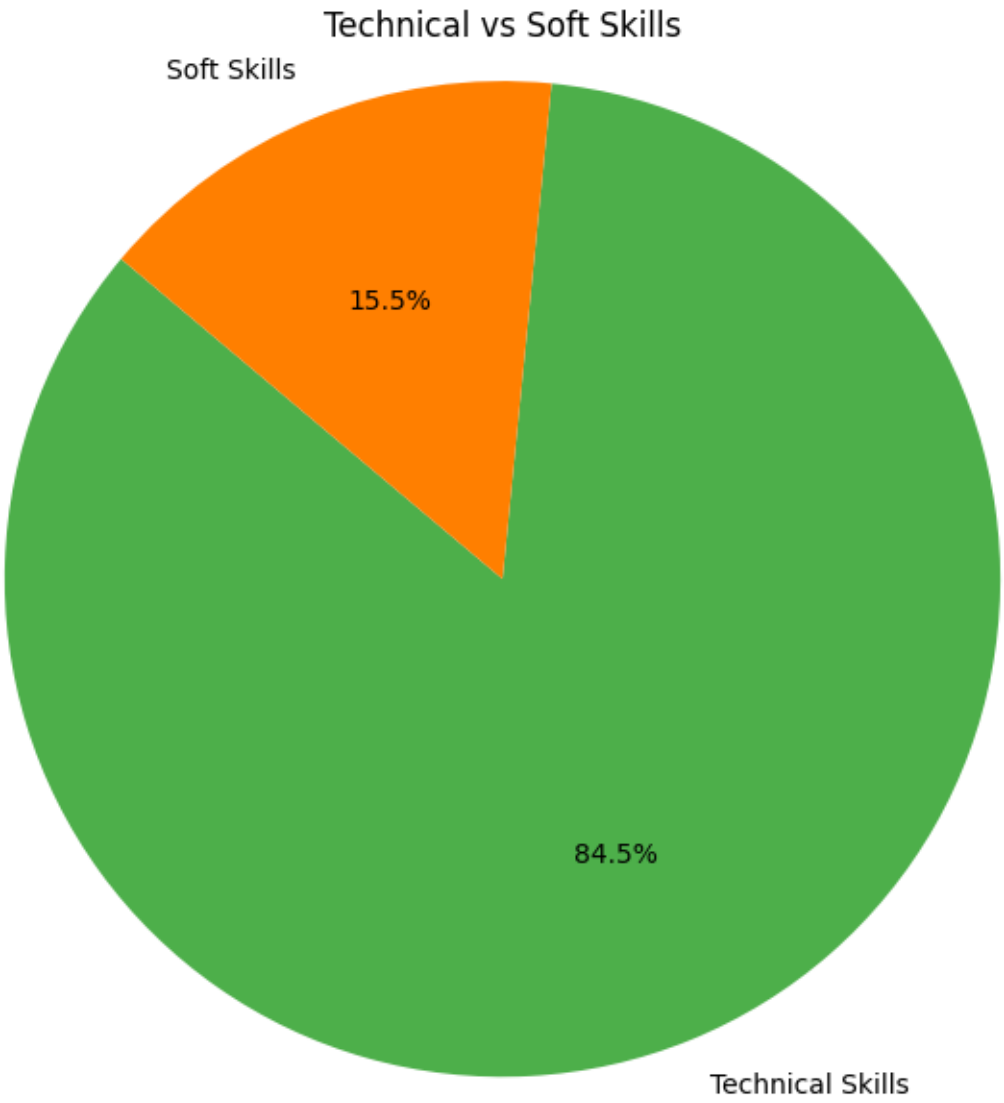


Figure 8 – (Pie Chart for comparison between Technical and Soft skills)

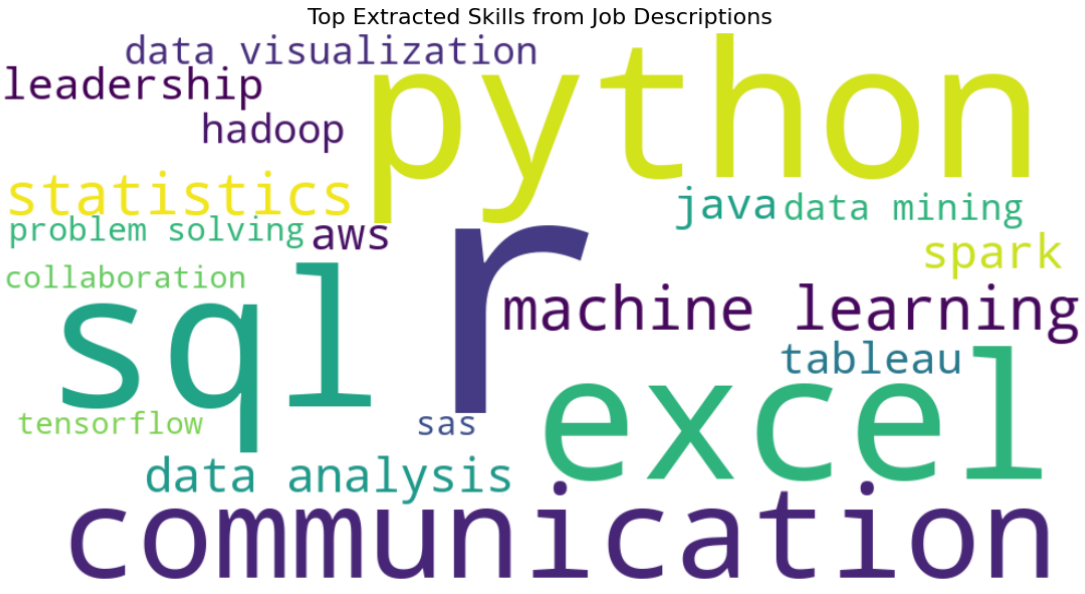


Figure 9 – (Word cloud for extracted skills)

Day 7: Skills Across Job Levels

Tasks Performed:

- *Job titles were categorized into **Entry**, **Mid**, and **Senior** levels*
- *Skill frequency was calculated for each level*
- *Heatmap and grouped bar chart were generated for comparison*

Output:

- ***Python** and **SQL** were in demand at all levels*
- ***Excel** was dominant in entry-level roles*
- ***AWS**, **Spark**, and **Deep Learning** were more emphasized in **Mid** and **Senior** positions*
- *The visualizations helped distinguish skill progression by experience for **excel.***

Visualizations:

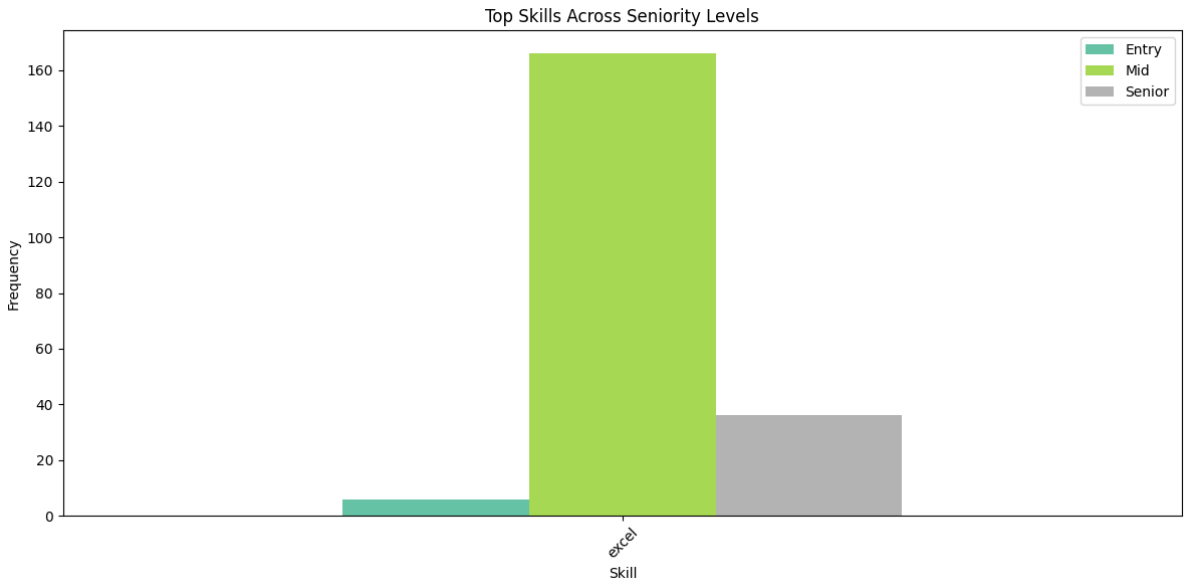


Figure 10 – (Comparison bar for skills showing excel is dominant in entry-level roles)

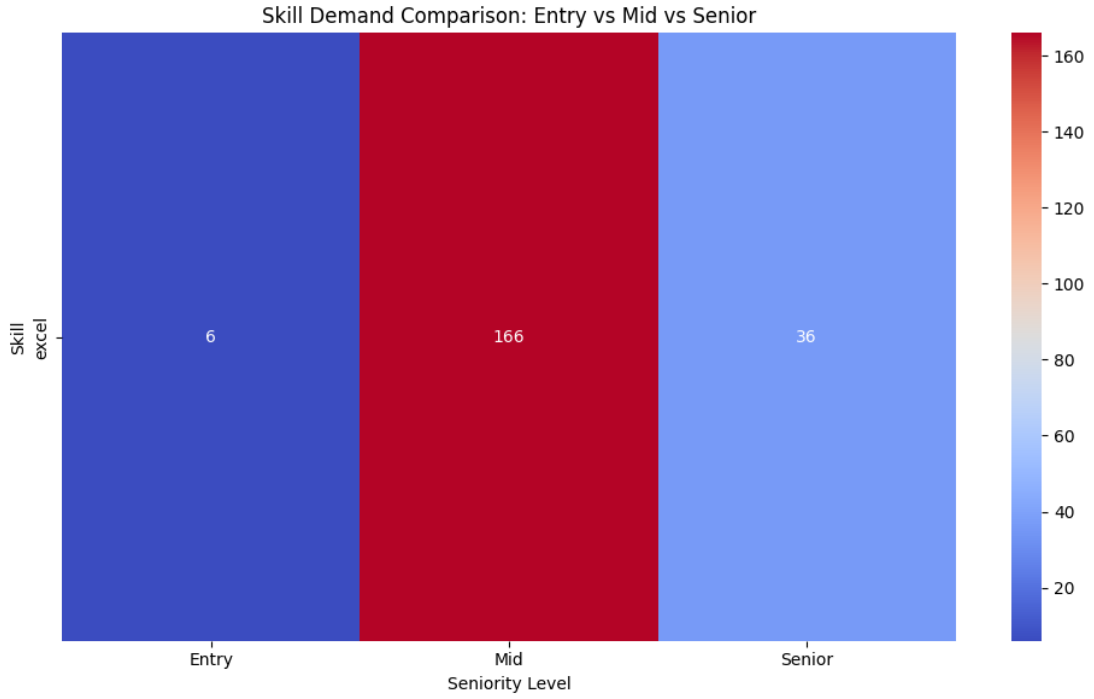


Figure 11 – (Heatmap for comparison of skills with respect to seniority wise demand roles)

Day 8: Skill Demand by Company

Tasks Performed:

- *Selected top 5 companies from earlier analysis*

- *Compared how often skills like **Python**, **R**, **Excel**, **AWS**, and **Spark** were mentioned in job descriptions*
- *Plotted a grouped bar chart for side-by-side comparison*

Output:

- ***Python** was universally required across all top companies*
- ***AWS** and **Spark** were more commonly required by **tech-intensive firms***
- ***R** appeared primarily in research-heavy roles*
- *Helped understand how different companies prioritize different tech stacks*

Visualizations:

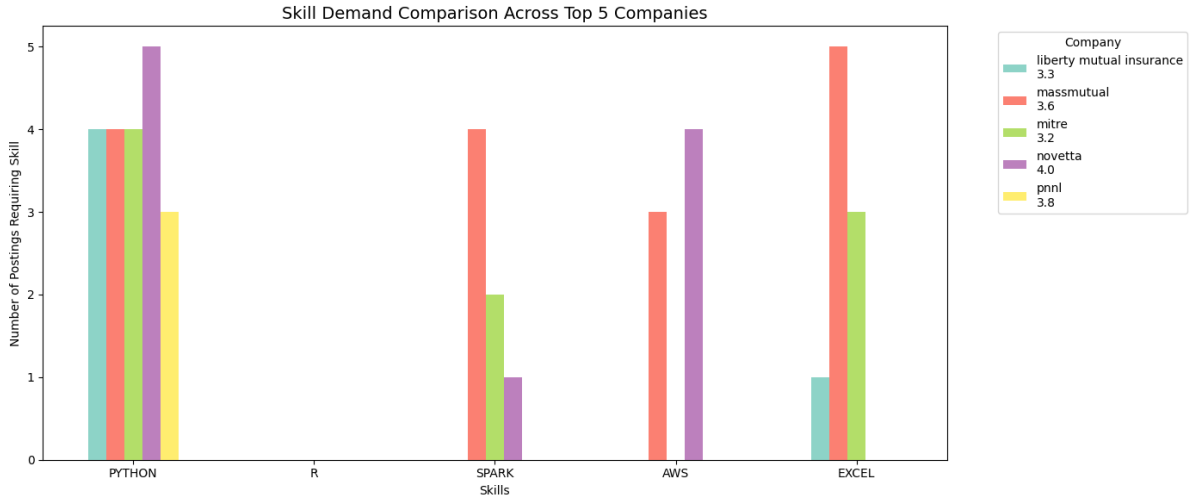


Figure 12 – (A grouped bar chart for side-by-side comparison of skills across top companies)

Day 9: Final Insights & Summary Report

Objective

The primary objective of this project was to analyze global job postings related to data and artificial intelligence fields to extract meaningful insights regarding:

- *In-demand skills*
- *High-growth job titles*
- *Regional hiring patterns*

---

1. Most In-Demand Skills

Skill	Demand Level
Python	Very High
Excel	High
SQL	Moderate
AWS	Growing Demand
Spark	Niche but Rising
R	Academic/Research

Insight:
Python and Excel are the most sought-after skills across data roles. Meanwhile, cloud technologies like AWS and Spark are rapidly gaining popularity, especially in mid to senior roles.

---

2. High-Growth Job Titles

Job Title	Trend
Data Analyst	Most Common Overall
Data Scientist	High Global Demand
Machine Learning Engineer	Increasing Rapidly
BI Analyst	Often Clubbed with Analyst Roles
Data Engineer	Niche but Critical

Insight:
Entry-level positions are dominated by analyst roles, but there is a noticeable increase in demand for engineers in AI and ML domains.

---

3. Regional Hiring Hubs

Region	Hiring Trend
India	Highest in Asia
United States	Strong Demand
UK & Germany	Moderate Growth
Remote Jobs	Increasing Trend

Insight:
India stands out as a major hiring hub, particularly for entry and mid-level roles. Remote opportunities are also gaining traction globally.

Recommendations

- *Strengthen core skills in **Python, SQL, and Excel**.*
- *Build expertise in **AWS and Spark** for domain specialization.*
- *Target **Data Analyst** or **Intern** roles to enter the job market.*
- *Monitor hiring patterns by **company and region** for targeted applications.*
- *Consider adding **date/time dimensions** in data to track **temporal hiring trends** in future iterations.*

All daily tasks (1-9) completed successfully.