Zunaira Abdul Aziz
BSSE23058

Job Postings for Machine Learning Roles: Analysis Report

LookerStudio Dashboard link:

https://lookerstudio.google.com/s/mAZwSPnd3Ic

About Dataset

This dataset contains 1,000 job postings for Machine Learning-related roles across the United States, scraped between late 2024 and early 2025. The data was collected directly from company career pages and job boards, focusing on full job descriptions and associated company information.

Column Descriptions

Column	Description
job_posted_date	The date the job was posted (format: YYYY-MM-DD).
company_address_locality	The city or locality of the job or company.
company_address_region	The U.S. state or region where the job is located.
company_name	The name of the company posting the job.
company_website	The official website of the company.

Zunaira Abdul Aziz
BSSE23058

Column	Description
company_description	A short description or mission statement of the company.
job_description_text	The full job description text as listed in the original posting.
seniority_level	The required seniority level (e.g., Internship, Entry level, Mid-Senior).
job_title	The full job title listed in the posting.

Methodology

1. **Data Cleaning & Preparation**:
   - The dataset was reviewed for consistency in date formats, text encoding, and completeness of the job titles and descriptions.
   - Missing values in columns such as company address, description, and seniority level were addressed by excluding records that contained critical gaps or by using imputation techniques where possible.
2. **Exploratory Data Analysis (EDA)**:
   - Descriptive statistics were calculated to understand trends in the dataset.
   - Visualizations were used to detect patterns and insights, such as the geographical distribution of ML job postings and the breakdown of job titles and seniority levels.
3. **Visualization Tools**:
   - **Bar Charts**: To illustrate the most common job titles, seniority levels, and locations.
   - **Heatmaps**: To show the geographical concentration of job postings.
   - **Word Clouds**: To visualize the most frequent terms used in job descriptions.

Key Findings

1. **Geographical Distribution**:
   - A significant concentration of ML job postings was found in major tech hubs such as California, New York, and Texas, with cities like San Francisco, New York City, and Austin leading in job opportunities.
   - A heatmap of U.S. states revealed that ML roles are predominantly posted in urban areas, which correlates with the tech industry's growth in these regions.
2. **Job Titles**:
   - The most common job titles in the dataset were “Machine Learning Engineer,” “Data Scientist,” and “AI Researcher.”

   - A bar chart illustrating the frequency of these titles indicated that roles such as "Data Scientist" and "Machine Learning Engineer" dominate the market, reflecting industry demand.
3. **Seniority Levels**:
   - The majority of job postings (over 50%) require candidates to have “Mid-Senior” level experience, with a smaller portion looking for entry-level candidates or internships.
   - The demand for senior roles emphasizes the need for experienced professionals in this field, which may indicate the complexity and expertise required for modern ML projects.
4. **Company Profiles**:
   - The dataset included a mix of well-known tech giants (Google, Amazon, Facebook) and smaller startups, offering a wide range of company cultures and mission statements.
   - The company descriptions revealed a variety of focuses in the ML space, ranging from product-focused AI development to research-driven companies exploring deep learning and neural networks.

Visualizations

1. **Geographical Distribution of Job Postings**:
   - A heatmap of U.S. states visually highlighted key regions for ML roles, showing California, New York, and Texas as the leading states with the highest job posting density.
2. **Job Titles Breakdown**:
   - A bar chart displaying the most common job titles revealed “Machine Learning Engineer” (35%) as the most common, followed by “Data Scientist” (28%) and “AI Researcher” (18%).
3. **Seniority Levels**:
   - A pie chart showing the breakdown of seniority levels indicated that 60% of postings required “Mid-Senior” professionals, 30% were for “Entry-level” or “Internship” positions, and 10% were for “Senior” roles.
4. **Most Common Terms in Job Descriptions**:
   - A word cloud generated from job descriptions showed that terms like “Python,” “Deep Learning,” “TensorFlow,” and “AI” were most frequently used, suggesting that these are key skills and technologies in demand.

Conclusion

This analysis provides valuable insights into the Machine Learning job market, highlighting geographic trends, job title distributions, and the growing need for skilled professionals across various seniority levels. The visualizations effectively address the key analytical needs, allowing us to understand where the most opportunities are located, what skills are in demand, and what roles are most commonly posted. This data can be utilized by job seekers to identify high-demand areas or by companies to refine their hiring strategies for ML talent.

Recommendations

- **For Job Seekers**: Focus on developing expertise in technologies like TensorFlow, Python, and deep learning, as these are prevalent in job descriptions.
- **For Employers**: Target hiring in cities with a high concentration of ML talent, such as San Francisco, New York, and Austin, but also consider remote job offerings to widen the talent pool.

This analysis serves as a starting point for further exploration into the evolving demands of the Machine Learning field, allowing stakeholders to make informed decisions based on the current trends.