

HIREME

1. Problem Statement

The problem I am trying to solve is something that affects me and my classmates. I have applied to over 500 internships and have yet to receive an interview. By making this project, I aim to get the information about the market and will try to understand the landscape:

- 1) Which skills (Python, SQL, Machine Learning, Cloud, etc.) are most in demand
- 2) Which roles (Data Scientist, Software Engineer, ML Engineer, DevOps) are trending
- 3) How salary ranges differ across roles and locations
- 4) What regional patterns exist (e.g., West Coast vs East Coast)
- 5) How job descriptions differ in required experience & technologies

2. Data Collection

I will scrape from job portals with static HTML structures, such as RemoteOK, WeWorkRemotely, AllJobsNet, Himalayas.app, and Indeed. For each job posting, I will collect details like Job title, Company name, Location (remote/hybrid/onsite + region), Salary (if available), Skills/tech stack mentioned in the description, Job type (full-time, contract), Date posted, and Job URL.

3. Analysis

- Skill frequency analysis: Identify the most in-demand technologies (e.g., Python, SQL, AWS, Java, React).
- Job title analysis: Determine which roles appear most often (Software Engineer, Data Scientist, ML Engineer, DevOps Engineer).
- Salary analysis: Compare salary distributions across roles, skills, and job types (remote vs onsite).
- Geographic patterns: Understand how opportunities and compensation vary by region.
- Skill correlation analysis: Discover which skills commonly occur together.

4. Visualizations

Using Matplotlib or other Python visualization libraries, I will create:

- Bar charts showing the most frequent skills and top job roles
- Box plots comparing salary ranges across job titles or skills
- Scatter plots relating salary to skills or job categories
- Heatmaps illustrating correlations between key technical skills
- Word clouds highlighting common job description terms.