LinkedIn
Job Postings Analysis

Github

Introduction

Purpose

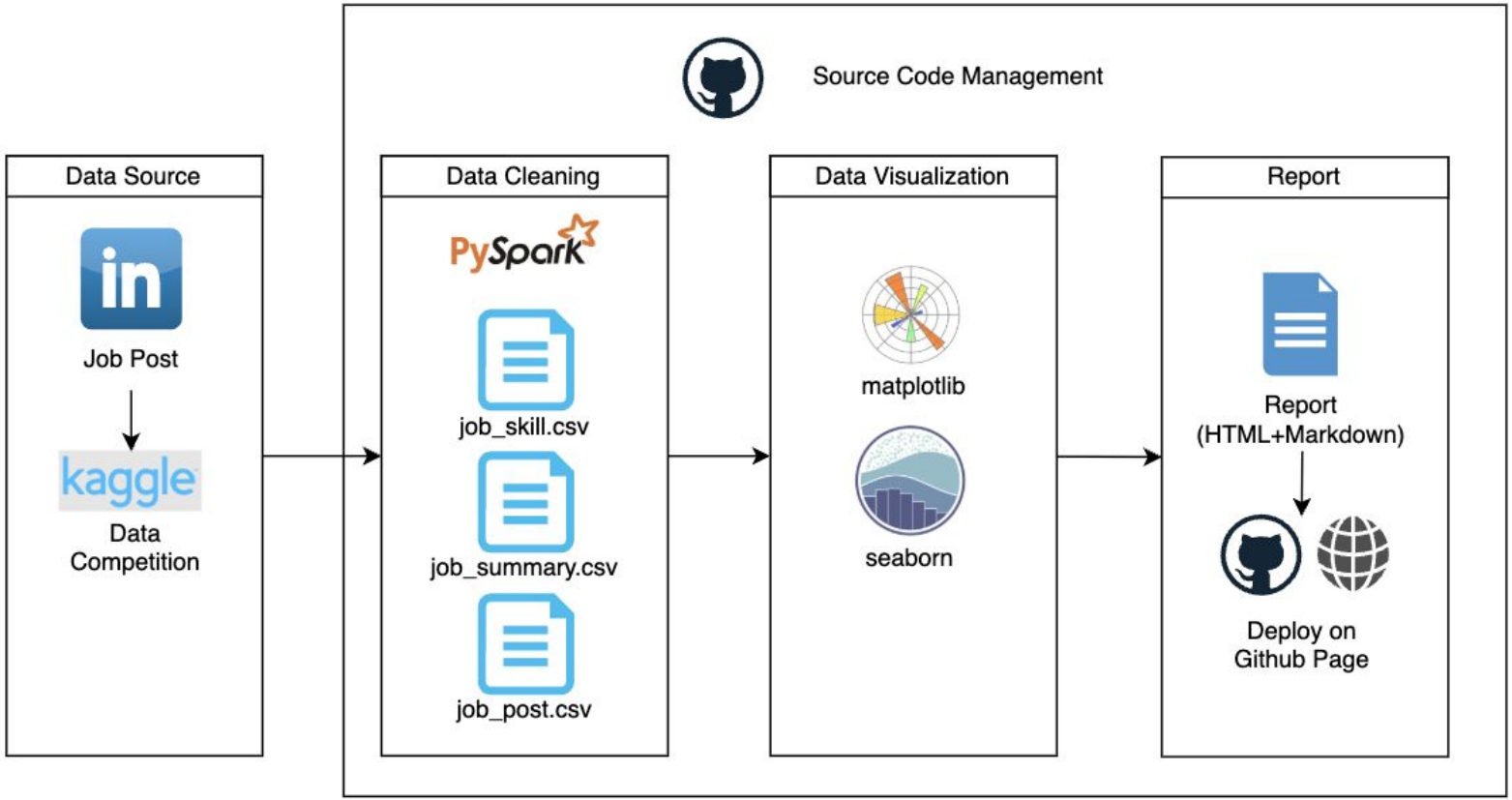
Personal Career Insights

Discover which skills employers are actively seeking in the current job market and identify cities with the most career opportunities.

Portfolio Demonstration

Showcase advanced big data processing techniques using PySpark to analyze massive datasets efficiently and effectively.

Dataflow


Source Code Management
Data Source
Job Post
kaggle
Data Competition
Data Cleaning
PySpark
job_skill.csv
job_summary.csv
job_post.csv
Data Visualization
matplotlib
seaborn
Report
Report (HTML+Markdown)
Deploy on Github Page

Dataset Overview

Data source: Kaggle's LinkedIn Jobs and Skills 2024 dataset, totaling approximately 6GB of raw job market intelligence.

linkedin_job_postings.csv

396 MB

1.3M rows of job posting data including titles, locations, and company information

job_skills.csv

642 MB

1.3M rows mapping specific skills requirements to individual job postings

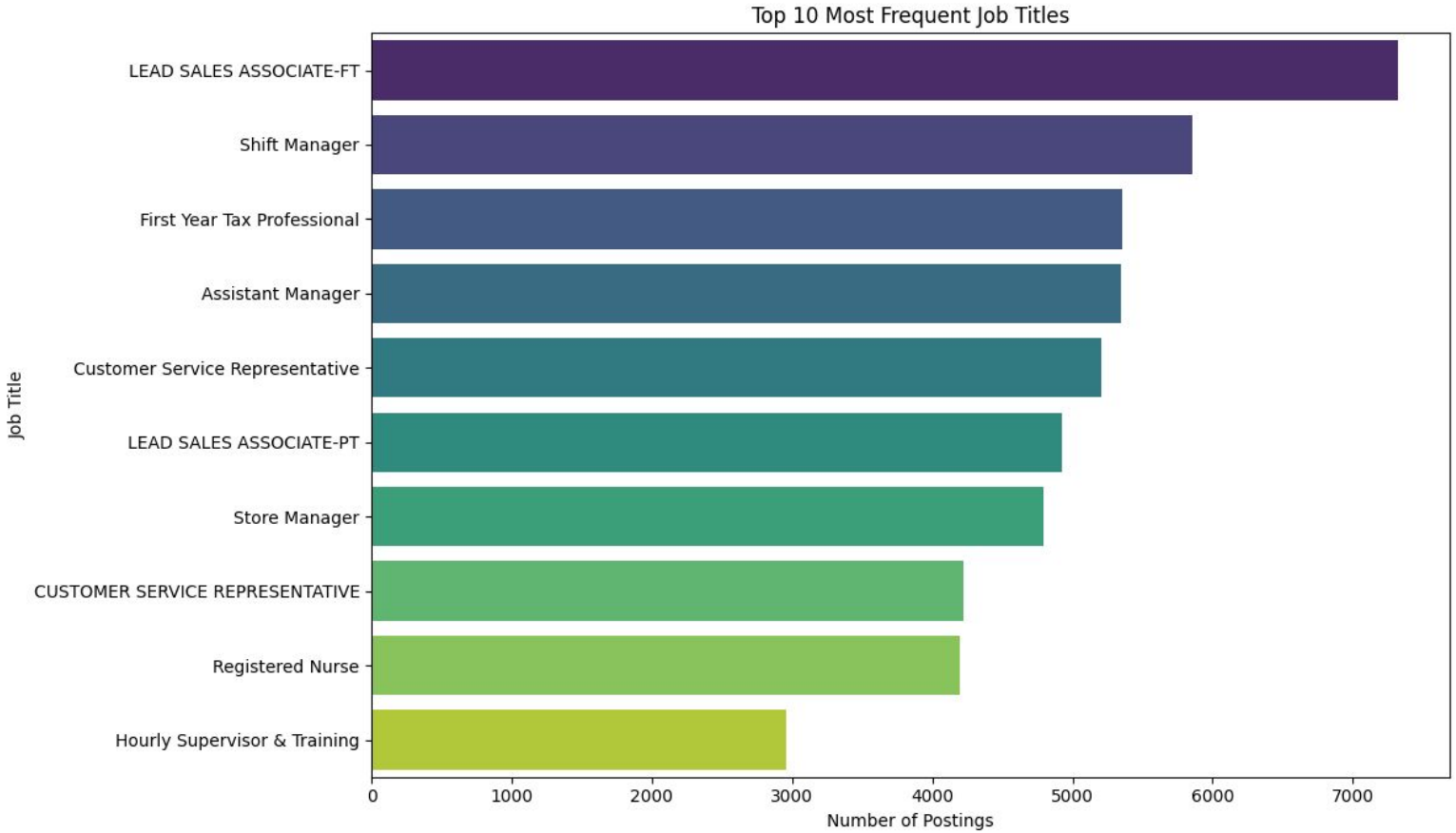
job_summary.csv

4.9 GB

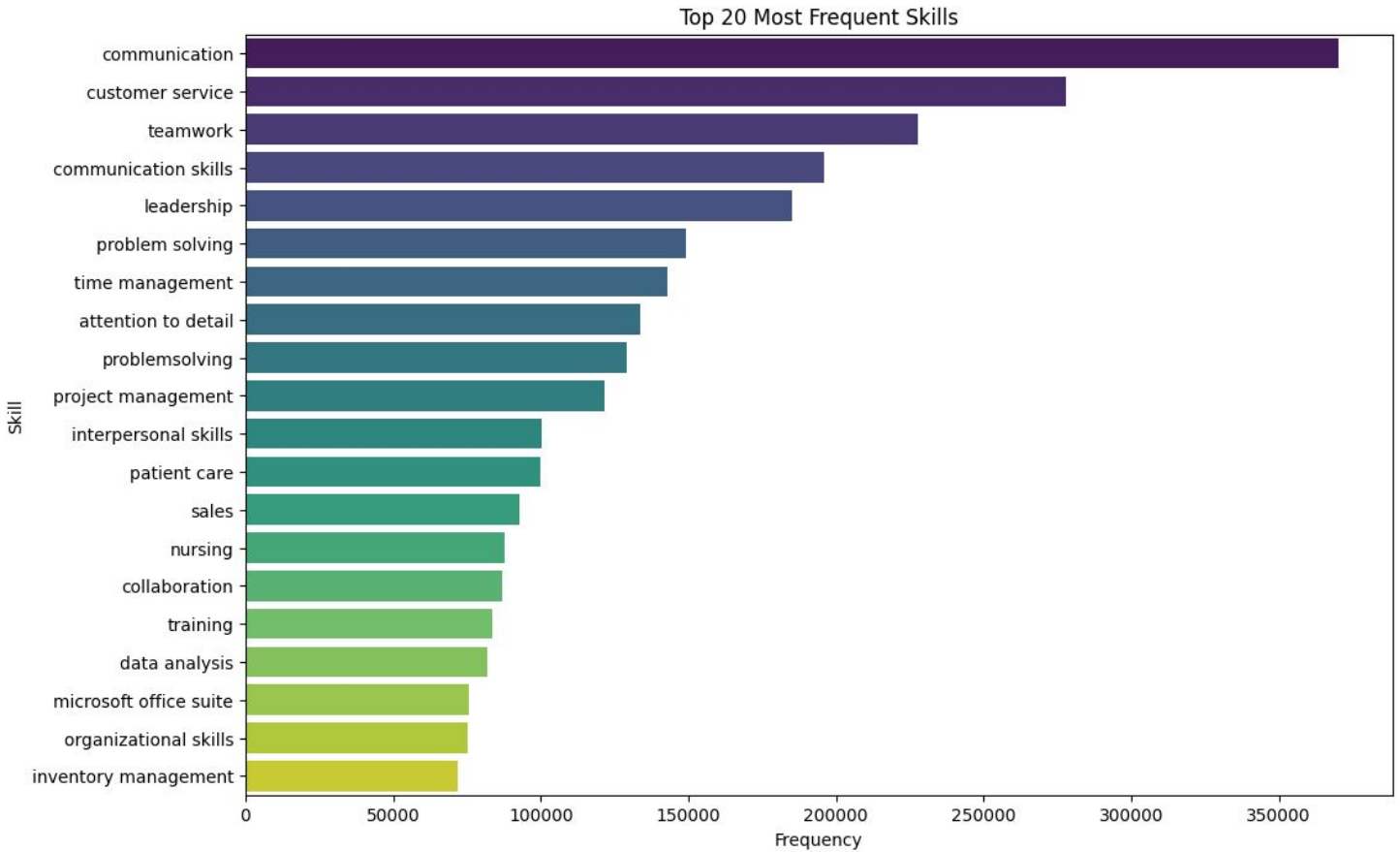
48M rows containing detailed job descriptions and comprehensive summary information

General Analysis

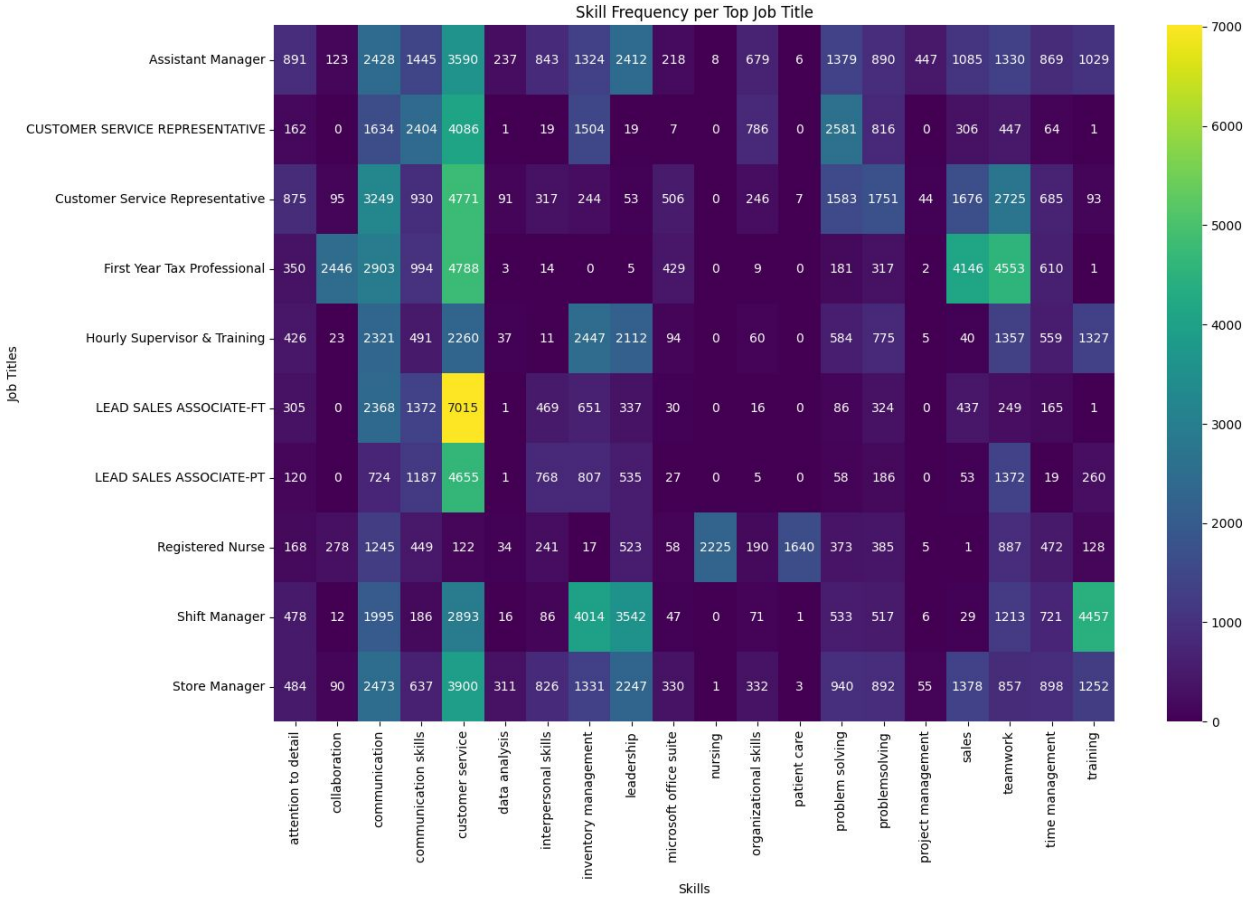
Job title Analysis



Skills Analysis

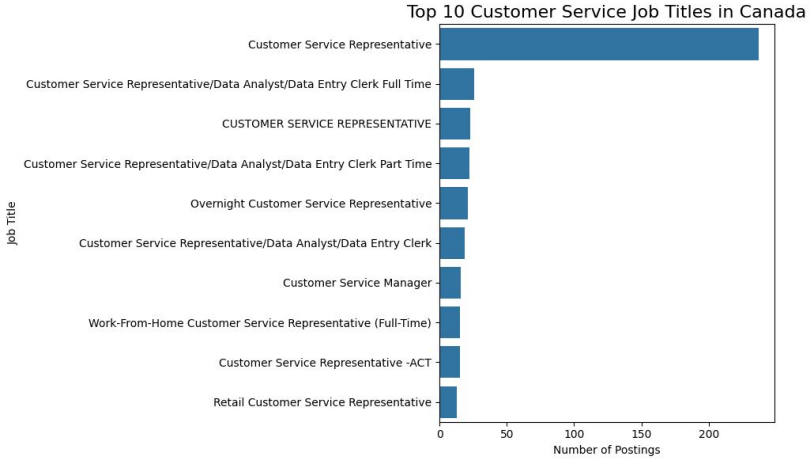


Skills & Job titles



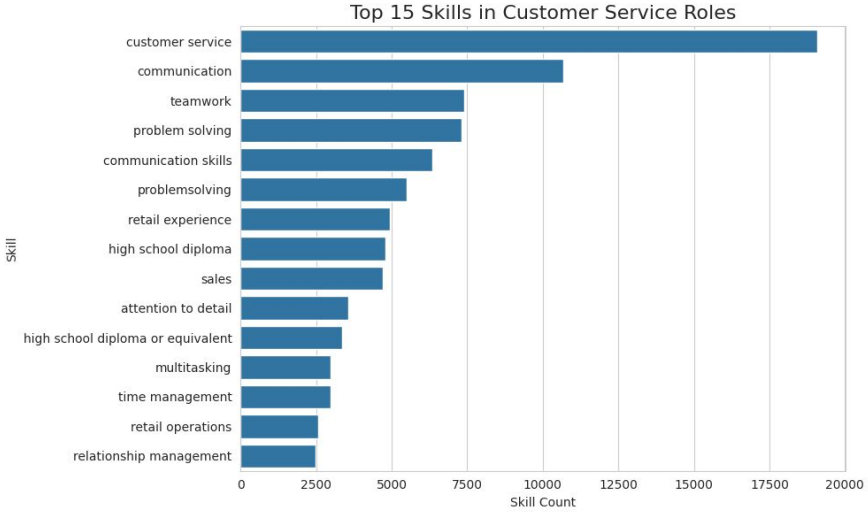
Customer Service Analysis

Customer Service (CS) Roles.



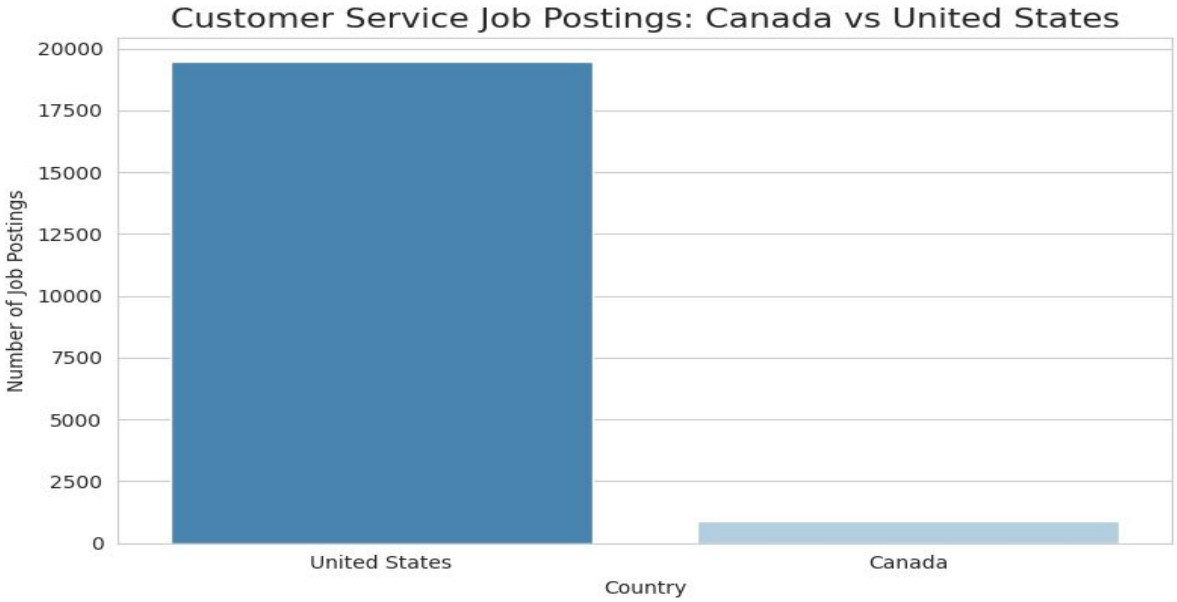
Customer Service roles represent a high-volume and stable job category across many industries.

Top Skills in Customer Service (CS) Roles.



These Customer Service skills, are the most required in this industry, so you can be competitive for job hunting

Customer Service (CS) job postings: Canada vs EE.UU

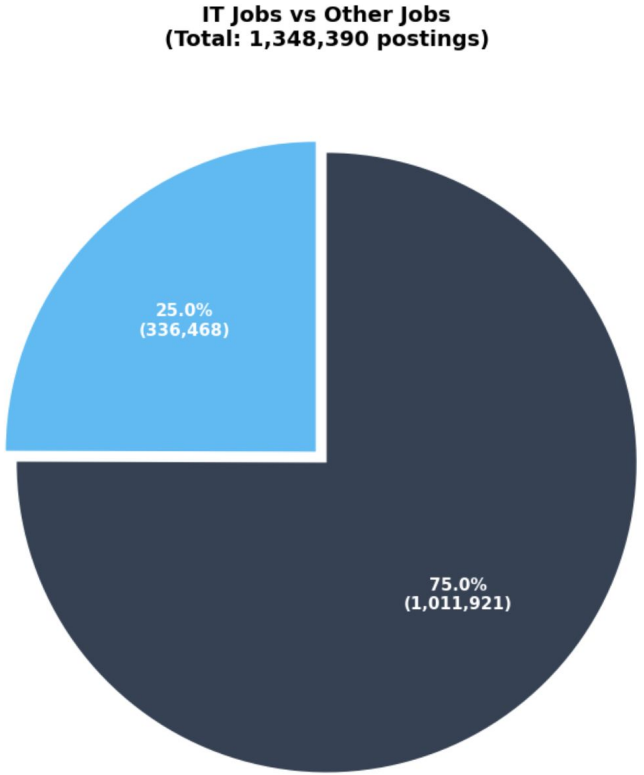


IT Job Analysis

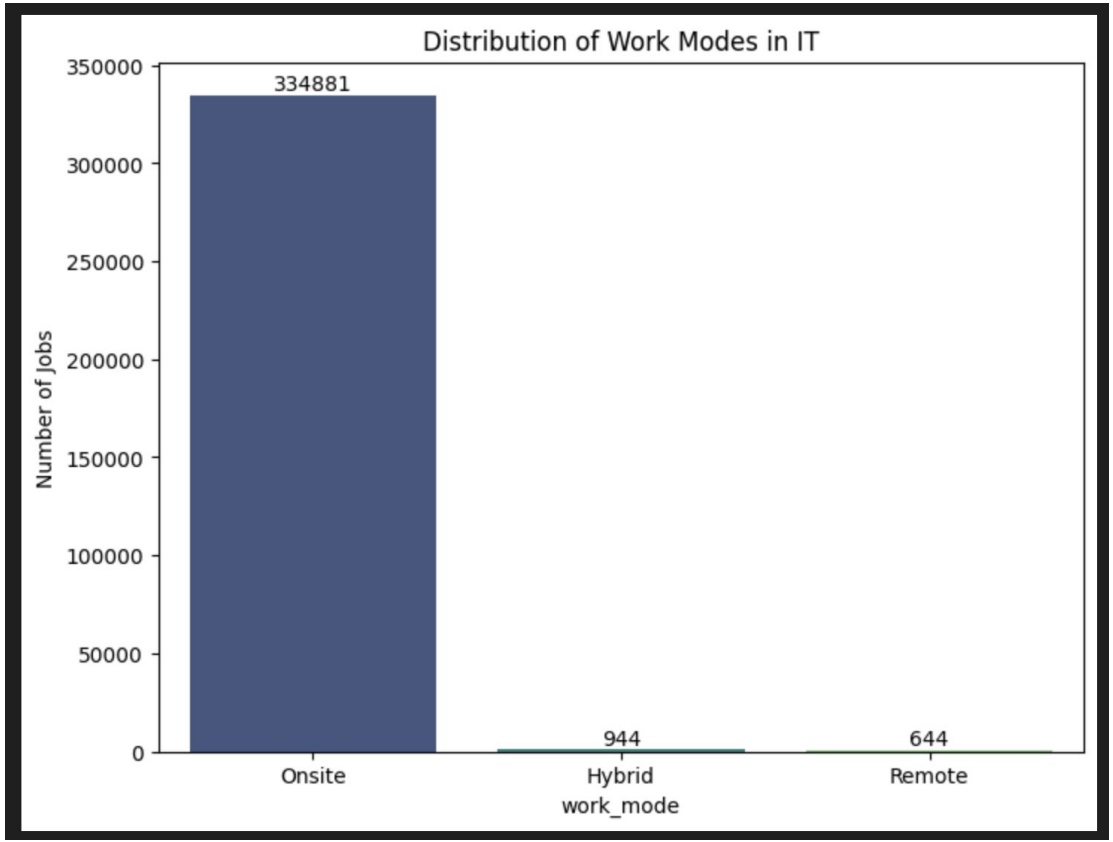
Filter IT Job

IT Keyword List

```
# IT Keyword Filter Definition
strict_keywords = [
    "AI", "IT", "Web", "QA", "UI", "UX", "ML", "SRE", "NLP", "IOT", "ERP", "SAAS",
    "API", "SDK", "ETL", "BI", "DBA", "AR", "VR", "XR", "GIS"
]
broad_keywords = [
    "Software", "Developer", "Engineer", "Data", "Analyst", "Cyber",
    "Cloud", "Network", "System", "Database", "Machine Learning",
    "DevOps", "Programmer", "Technical", "Information Technology",
    "Security", "Architect", "Scrum", "Product Owner", "Java", "Python", "Javascript",
    "Full Stack", "Backend", "Frontend", "Mobile", "Firmware", "Blockchain",
    "Data Scientist", "Robotics", "Automation", "Reliability", "Platform", "Infrastructure",
    "Support", "Technician", "Help Desk", "Service Desk", "Admin", "SysAdmin", "Technology", "IT",
    "Automation", "Developer", "Architect"
]
tech_keywords = (
    "React", "Angular", "Vue", "Node", "AWS", "Azure", "GCP", "Linux", "Docker",
    "Kubernetes", "Terraform", "Ansible", "Jenkins", "Salesforce", "Oracle", "SAP",
    "C#", "C++", "Golang", "Rust", "Swift", "Kotlin", "Flutter", "TensorFlow", "PyTorch",
    "Python", "SQL", "HTML", "CSS", "JavaScript", "TypeScript", "PHP", "Ruby", "Perl", "Go", "R",
    "API", "Agile", "Cloud Computing", "Machine Learning",  "Jira", "Software Engineering",
    "Microservices", "Power BI", "Active Directory",
    "NoSQL", "CI/CD", "Statistics", "Scripting", "SQL Server", "Git", "Spark",
    "Hadoop", "Tableau", "Bash"
)
```

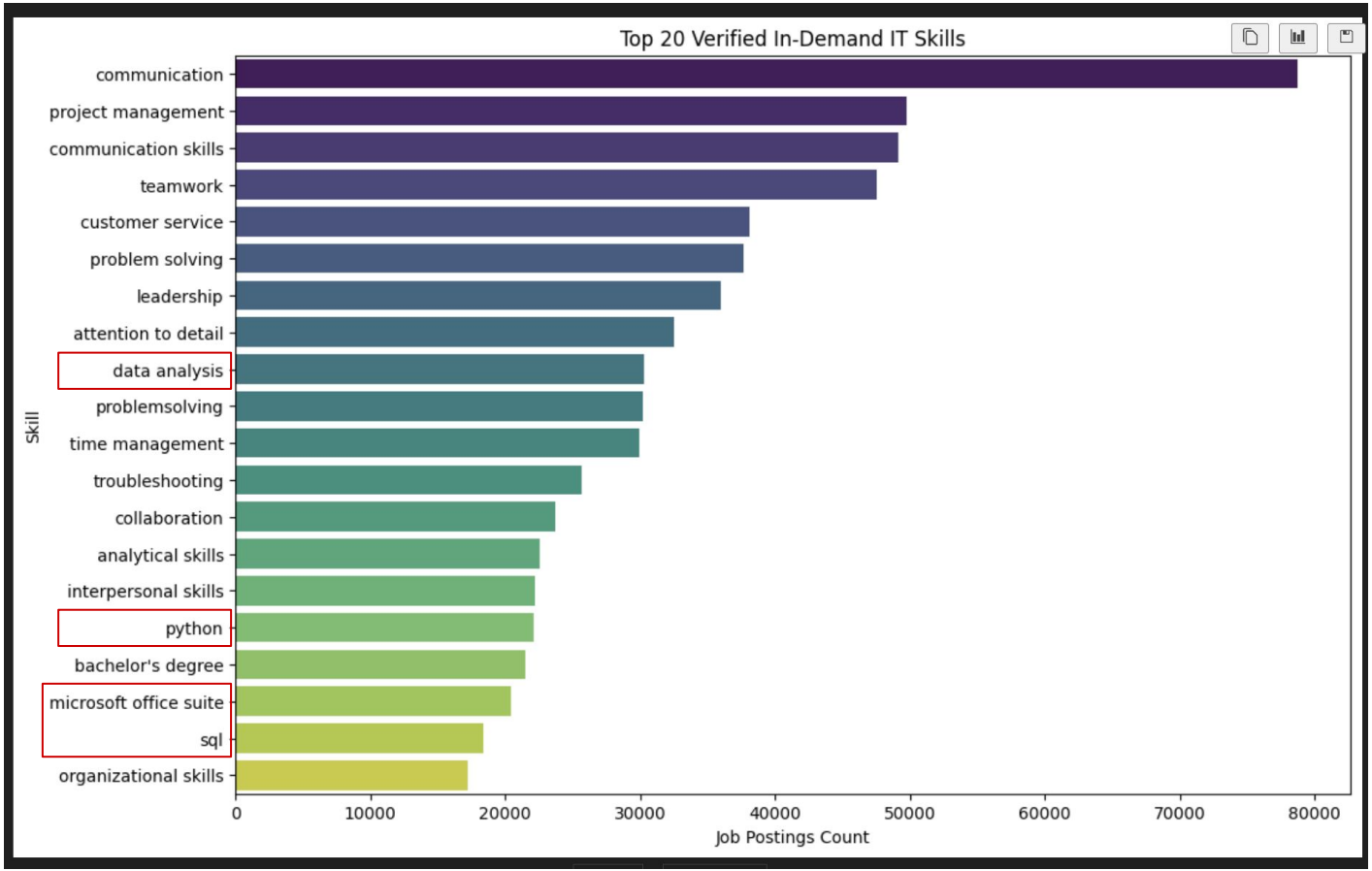


Work Mode Analysis

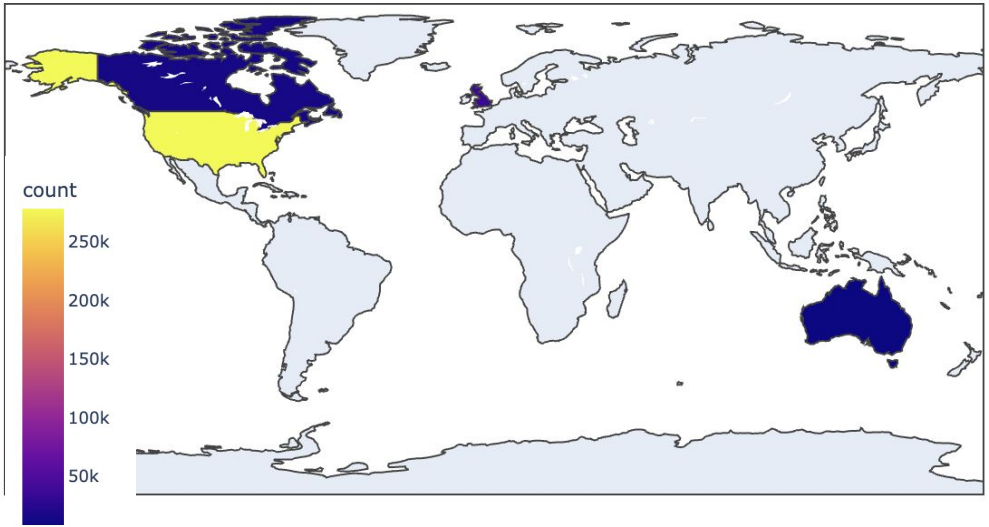


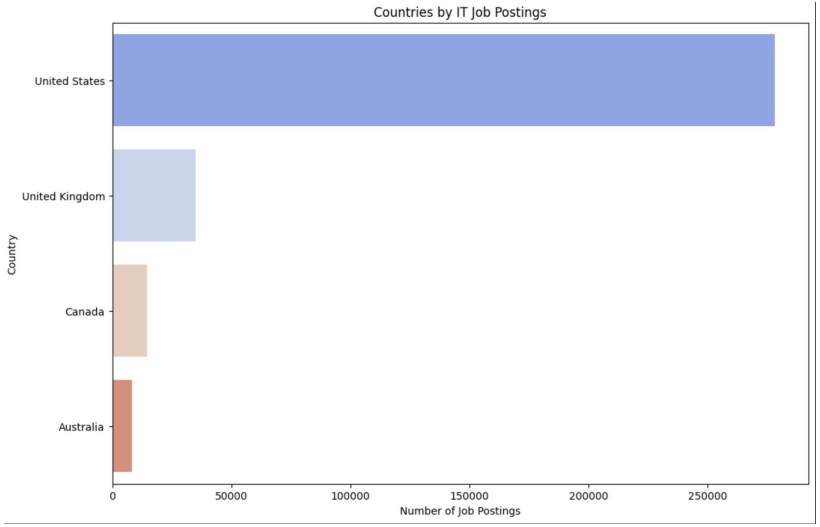
99.5%: Onsite

Skills Analysis

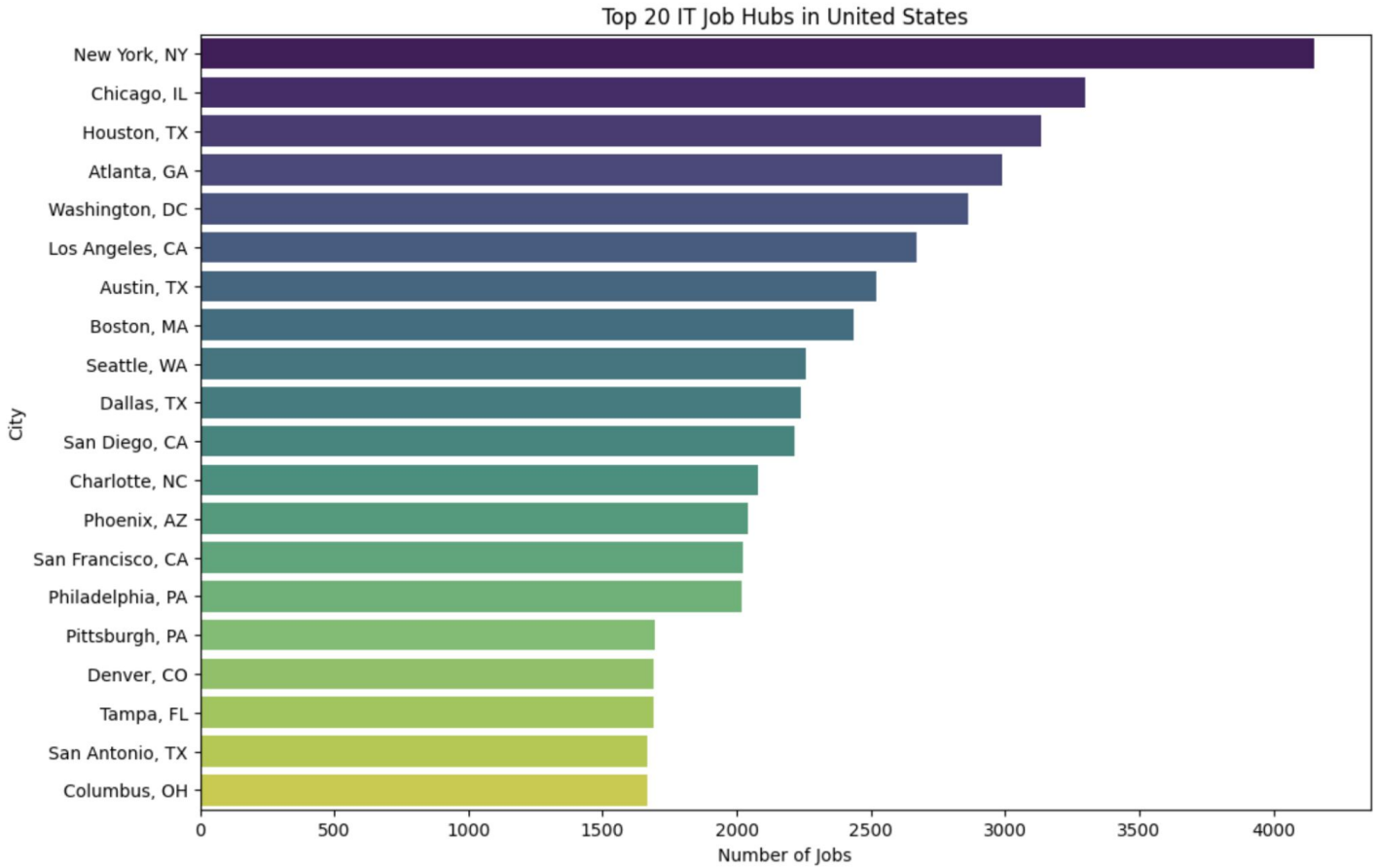


Geographical Analysis

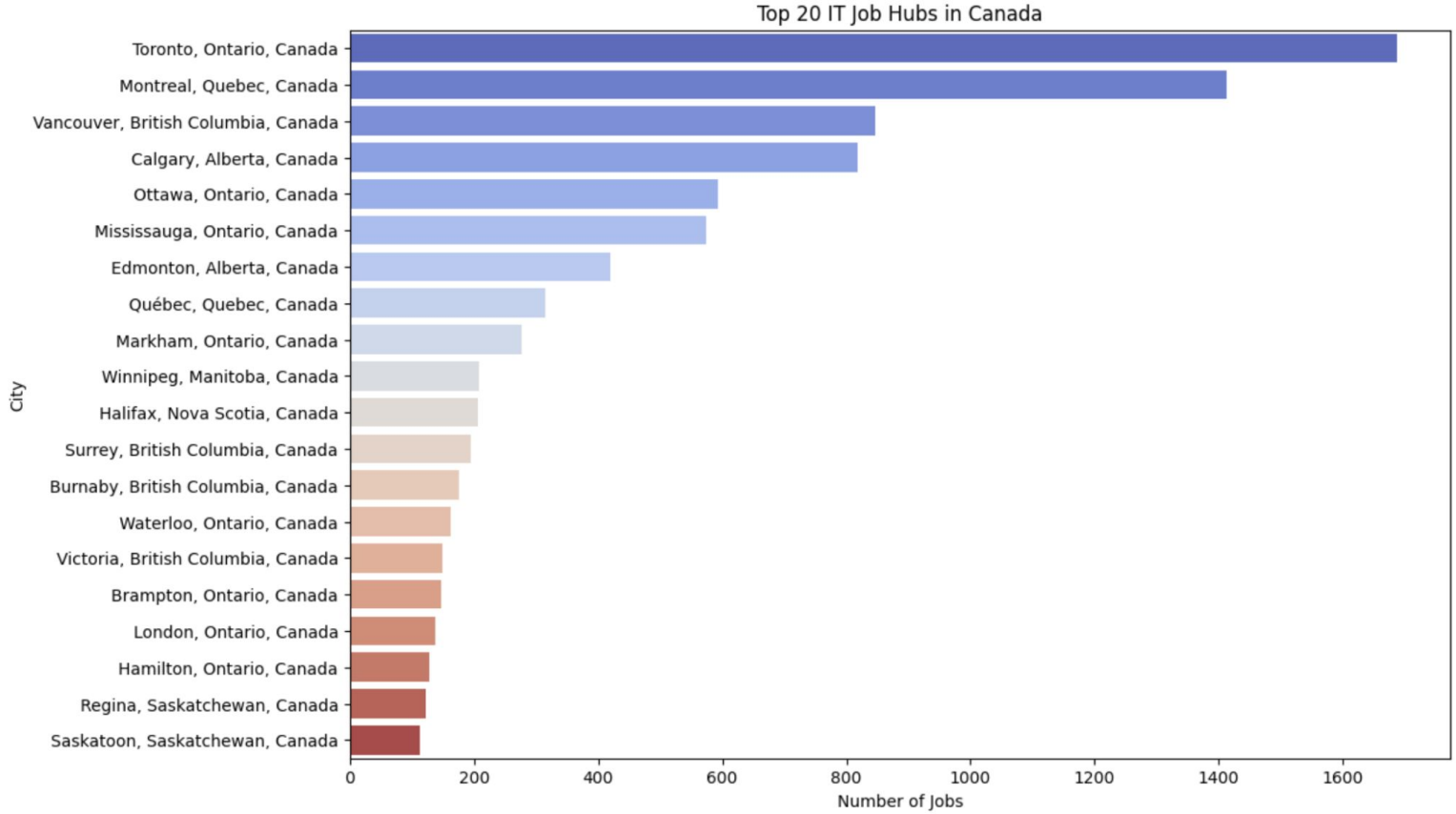




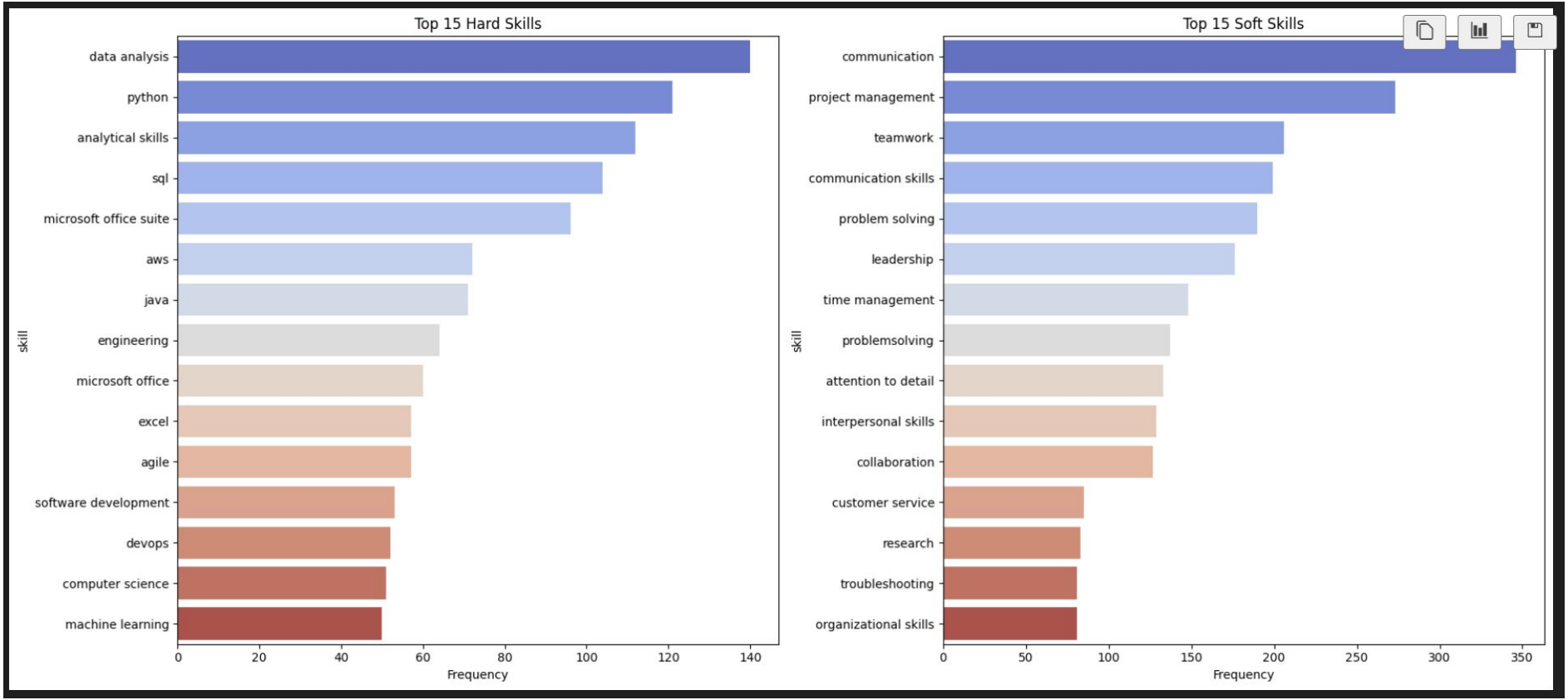
Geographical Analysis - United States



Geographical Analysis - Canada



Skills Analysis - Vancouver



Interactive Web Report

Public Access

Deployed on GitHub Pages for seamless sharing and portfolio demonstration

Conclusion

Conclusion

- Analyzed 1.3 million Linkedin job posting data

- Customer Service Job
  - Recognized ‘Customer service’, ‘Communication’, ‘Leadership’, ‘Problem Serving’ skills are required

- IT Job
  - Recognized data analytics skills are required

Thank you

Tetsuya

Maco

Johny