LINKEDIN JOB DATA ANALYSIS

Data Analysis using
Python and its libraries

BASIC INTRODUCTION

This project focuses on extracting insights from a large dataset related to jobs, applications and companies recruitment. The project involves several steps, including data cleaning, exploratory data analysis (EDA), building a dashboard using Tableau and making a report. The dataset is about the job details data scraped from LinkedIn. Data consist of job descriptions for various roles like Java Developer, Data analyst, Software Testing in India-based companies.

STEP-1 DATA CLEANING

For cleaning the data , we will be using pandas and numpy libraries from python. Firstly, we loaded the dataset and we printed some of its data onto the screen and then we write the code to check for null values. During this, we got to know that columns named company_id and Column1 are totally empty and should be dropped. We also dropped the alumni and hiring person link columns to remove the discrepancies and make our data set more efficient to make accurate predictions.

FURTHER DATA CLEANING

Now, we duplicated the dataset ‘df’ into a new dataset which we called as job. In this data set, The work_type column is filled with the most frequently occurring value I.e the most demanding work type in that same column. When the hiring_person column has empty cells they are filled with NA. Also, the blanks in LinkedIn followers column is filled with NA.Finally, we used the dropna operation on the rest of the column (considered as subsets) to ensures that the DataFrame job includes only those rows that have complete information concerning the specified columns. This is a common preprocessing step to improve the quality and accuracy of further analyses or modeling.

STEP-2 EDA

Exploratory Data Analysis

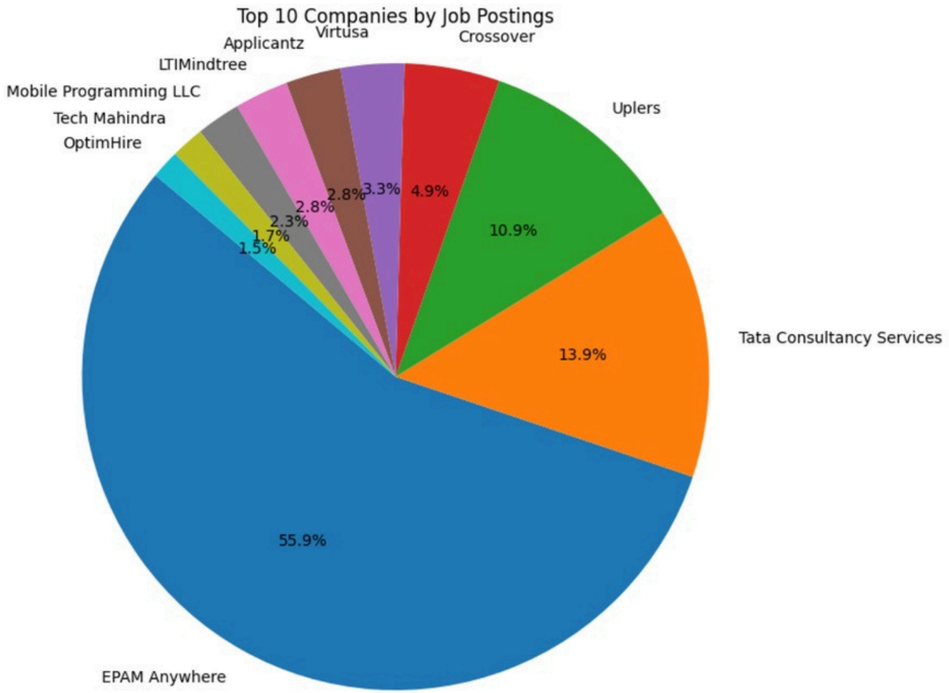
For Analysis and visualisation, we will be using Matplotlib, seaborn and pandas libraries . We used the describe function to generate some useful information including the total count, unique values, maximum, minimum, frequency, mean and percentiles.

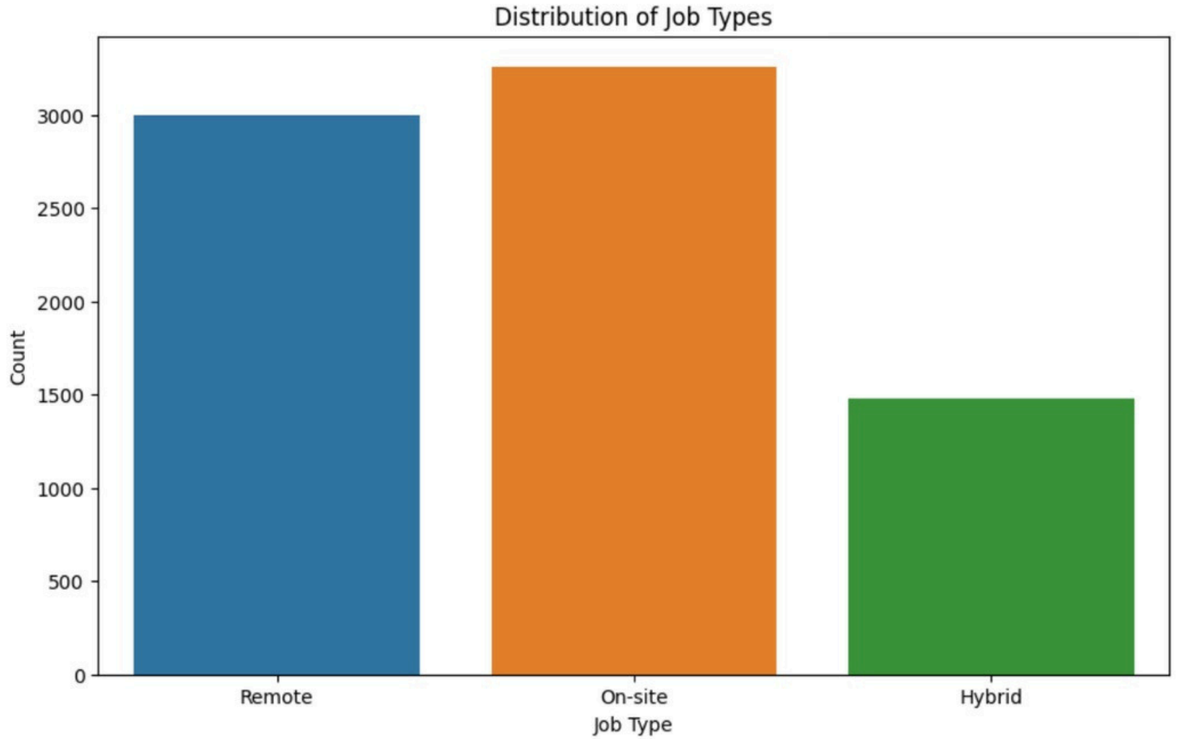
USEFUL INSIGHTS

- **Total Number of Rows:** The DataFrame contains **7,583** job listings.
- **Feature Summary:**
  - **job:** Presumably this column contains job titles, with **2,903** unique job titles.
  - **location:** There are **139** unique locations noted in the dataset, indicating a diverse geographical coverage.
  - **company_name:** Specifies **2,342** different companies listed in the DataFrame, showcasing a variety of potential employers.
  - **work_type:** This column has **3** unique values, indicating a limited number of categories regarding the type of work (e.g., "On-site," "Remote," "Hybrid").
  - **full_time_remote:** The work arrangements show **23** unique values, indicating flexibility in work arrangements.
  - **no_of_employ:** This reflects the size range of companies, with **269** unique values indicating a range in company sizes.
  - **no_of_application:** It shows **202** unique values, indicating variability in the number of applications received for jobs.
  - **posted_day_ago:** Limits to **91** unique values suggest a distribution of how recently jobs have been posted.
  - **Hiring_person:** Represents the number of individuals involved but shows all missing or not applicable.
  - **linkedin_followers:** A significant number, **2,671**, indicates companies range widely in their online presence and may suggest their employer branding influence.
  - **job_details:** Captures **4,388** unique details about job expectations, responsibilities, etc., implying a varied job description structure across companies.

VISUALISATIONS

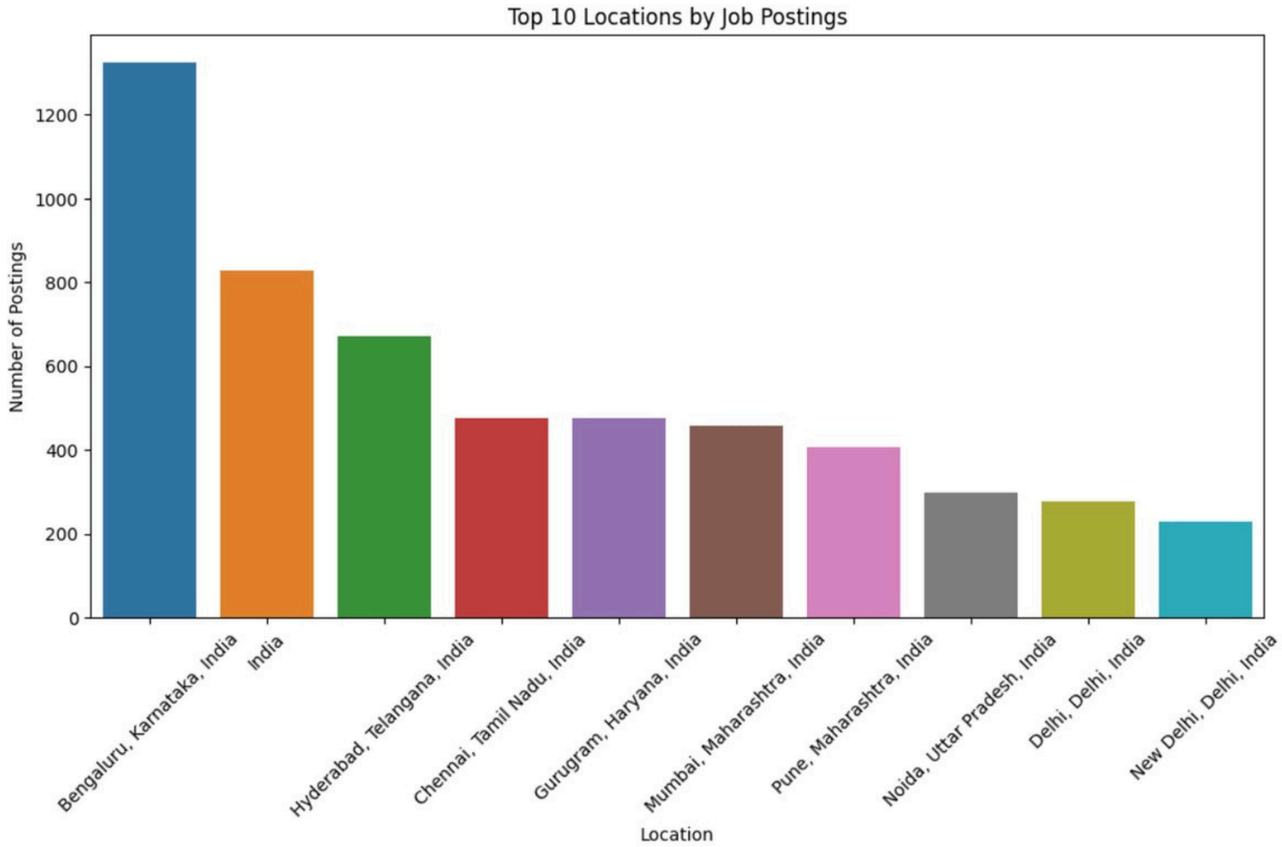
The given pie chart shows the top 10 companies according to the number of job postings.





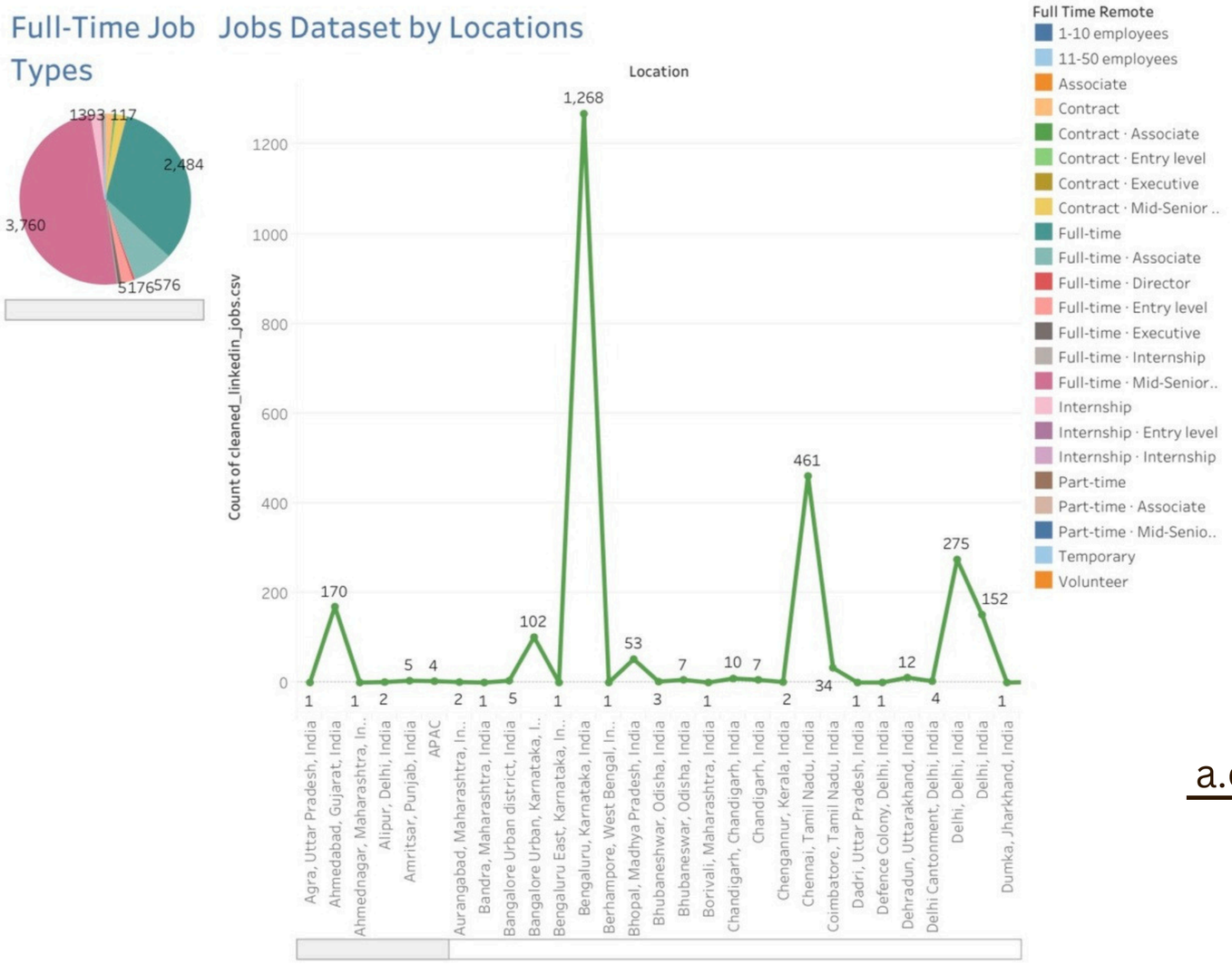
The given bar graph shows the distribution of the three job types namely: Remote, on-site and hybrid respectively.

VISUALISATIONS



The given bar graph shows the top ten cities or locations where there are highest job postings.

TABLEAU DASHBOARD VIEW



Link to the dashboard:

https://prod-apnortheast-a.online.tableau.com/#/site/vanshisethi1815-9b883a84be/views/Linkedin-Job-Analysis/Dashboard1?%3Aiid=3

THANK YOU