Rajalakshmi Engineering College (An Autonomous Institution) Rajalakshmi Nagar, Thandalam- 602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AD23632 - Framework for Data Visualization and Analytics

Mini Project: Job Market Data Analysis

Report submitted by

REGISTRATION NUMBER	:	231501164
STUDENT NAME	:	SUPRAJA R
YEAR	:	2023-2027
SUBJECT CODE	:	AD23632

Rajalakshmi Engineering College (An Autonomous Institution) Rajalakshmi Nagar, Thandalam- 602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AD23632 - Framework for Data Visualization and Analytics

Mini Project: Job Market Data Analysis

Report submitted by

REGISTRATION NUMBER	:	231501164
STUDENT NAME	:	SUPRAJA R
YEAR	:	2023-2027
SUBJECT CODE	:	AD23632

Table of Contents

Chapter 1: Abstract

The growing complexity of the global job market has made data-driven insights increasingly vital for understanding hiring patterns, workforce demands, and evolving employment trends. This project aims to provide a comprehensive analysis of the job market landscape by leveraging structured job listing data obtained from professional platforms such as LinkedIn. The dataset encompasses multiple attributes including company names, job titles, employment types, locations, posting years, and job counts, offering a detailed snapshot of employment dynamics across industries and geographies. Unlike qualitative surveys, this dataset is quantitative and cross-sectional in nature, allowing for the exploration of hiring patterns, company behavior, and temporal employment trends across diverse sectors.

The objective of this study is threefold. First, it seeks to identify temporal trends in job availability by examining variations in job postings across years and industries. Second, it explores spatial and organizational patterns, analyzing how company size, geographic location, and employment type influence job distribution. Third, it aims to provide a visual framework that enables users to interactively explore job insights through the use of data visualization tools. To achieve these objectives, the project adopts a multi-tool analytical approach: data preprocessing and exploratory analysis using Python, followed by interactive dashboard creation and storytelling in Tableau.

This study emphasizes the importance of data visualization in understanding large-scale employment datasets and translating complex patterns into actionable insights. It demonstrates the combined strength of Python's analytical capabilities and Tableau's visual storytelling features to deliver both quantitative accuracy and interpretive clarity. The findings derived from this analysis can support students in identifying emerging career opportunities, assist recruiters in understanding talent demand, and guide policymakers and educators in aligning skill development with market needs.

Chapter 2: Introduction

The modern workforce is increasingly shaped by data-driven technologies, with employment trends and hiring practices evolving rapidly across industries. As organizations adapt to global digital transformation, understanding the structure and behavior of the job market has become essential for both employers and job seekers. This project investigates the dynamics of the job market through a structured dataset capturing diverse employment attributes such as company names, job titles, locations, employment types, and posting years. The intent is to analyze how these variables interact to reveal patterns in hiring demand, industry preferences, and geographic concentration of opportunities.

The study adopts a comprehensive analytical approach by examining temporal trends in job postings, regional hiring patterns, and organizational recruitment behavior. Unlike traditional surveys or qualitative assessments, the dataset used here offers a quantitative, cross-sectional view that enables large-scale comparison across multiple dimensions—companies, cities, and job categories. This allows for a deeper understanding of how the employment landscape shifts over time and across sectors.

By leveraging Python for data preprocessing and exploratory analysis, and Tableau for interactive and visually rich dashboards, the project combines analytical precision with accessible visual storytelling. The ultimate goal is not only to uncover meaningful trends within the job market but also to create a practical visualization framework that assists students, professionals, and policymakers in interpreting these insights. In doing so, the project bridges the gap between raw employment data and actionable intelligence, empowering users to make informed career and strategic decisions.

.

Chapter 3 : Dataset Description

The dataset used in this project provides a structured overview of job market information, capturing multiple dimensions of employment dynamics such as job roles, company profiles, and industrial distribution. It serves as a valuable resource for analyzing workforce demand, hiring patterns, and regional employment concentration. The dataset was collected from professional job platforms such as LinkedIn and other recruitment sources, containing records of job postings across different companies and locations.

It is not a time-series dataset but rather a cross-sectional dataset, making it ideal for comparative analysis across sectors, locations, and experience levels. Each record represents a unique job posting with attributes that describe both organizational and positional characteristics.

Key variables include:

Job Title: Specifies the designation or role being offered (e.g., Data Analyst, Software Engineer, Marketing Executive).

Company Name: Indicates the organization or employer offering the job position.

Location: Provides the geographical location of the job, including city and country information.

Hiring Status: Denotes whether the job posting is currently open, filled, or closed.

Date: Represents the posting date of the job, which helps in understanding temporal hiring trends.

Seniority Level: Identifies the experience level required for the position (e.g., Entry-level, Mid-level, Senior-level).

Job Function: Describes the primary functional domain of the role (e.g., Engineering, Management, Sales, Human Resources).

Employment Type: Specifies the nature of employment such as Full-time, Part-time, Internship, or Contract.

Industry: Refers to the sector or domain in which the company operates (e.g., Information Technology, Finance, Healthcare, Education).

This dataset is particularly valuable as it integrates organizational, functional, and geographical dimensions of the job market, enabling a holistic analysis of hiring practices. By examining these attributes together, the project identifies patterns such as top hiring industries, trending roles, and regional job concentrations. The dataset's structured format allows for seamless integration into analytical and visualization tools such as Python and Tableau, supporting clear and interactive interpretation of job market insights.

Chapter 4: Objective

The main objective of this project is to analyze job market data to identify key employment patterns, hiring trends, and organizational behaviors across different industries and locations. To achieve this, the study defines specific research aims that provide direction and clarity for systematic exploration and visualization.

Trend Analysis: Examine how job postings vary across different time periods to understand changes in employment demand and hiring activity.

Geographical Insights: Identify the regions and cities with the highest job concentrations, thereby highlighting potential employment hubs and industry clusters.

Organizational Behavior: Analyze company-wise hiring frequency to determine which organizations and industries are most active in recruitment.

Role and Experience Mapping: Investigate the distribution of job titles and seniority levels to understand the skill demand and experience patterns in the current job market.

Employment Type Evaluation: Compare the prevalence of various employment types—such as full-time, part-time, and internships—to reveal workforce structure and flexibility trends.

Tool Demonstration: Showcase how Python and Tableau can be effectively utilized—Python for data cleaning and analytical rigor, and Tableau for creating visually compelling and interactive dashboards that simplify complex data interpretation. By fulfilling these objectives, the project aims to deliver both

academic insights and **practical applications**. For students and professionals, it highlights in-demand industries and roles that can guide career planning.

Chapter 5: Methodology

The methodology of this project follows a structured, multi-step approach to ensure systematic and reliable analysis of the job market dataset. Each stage contributes to transforming raw data into meaningful insights through analytical and visualization tools.

Data Preprocessing:

Using Python, the raw dataset is cleaned and prepared for analysis. Missing values are handled appropriately, inconsistent entries are corrected, and duplicate rows are removed. Column data types are standardized (e.g., categorical for employment type and numerical for job counts), ensuring compatibility for further analysis. This step guarantees that the dataset is accurate, structured, and ready for visualization.

Exploratory Data Analysis (EDA):

Descriptive statistics and summary visualizations are generated to understand the overall distribution and characteristics of the dataset. Count plots, bar charts, and histograms are used to identify hiring patterns across industries, employment types, and seniority levels. Temporal analysis based on the posting date helps in tracking job market trends over time.

Feature Engineering:

New features are derived to enhance analytical depth. Examples include categorizing industries based on hiring frequency, classifying locations by job density, and grouping companies according to recruitment scale. These engineered features help in highlighting meaningful comparisons within the data.

Visualization Tools:

Python: Utilized for initial data cleaning, statistical summaries, and the creation of static visualizations that validate patterns within the dataset.

Tableau: Used for building interactive dashboards that combine multiple views—such as company-wise hiring, geographical job spread, and employment type distribution—into a cohesive and visually engaging story. Users can filter and interact with the dashboard to explore job data dynamically.

Interpretation:

The final step involves interpreting results in the context of job market behavior, industry demands, and workforce distribution. Insights are derived from both Python-based analysis and Tableau dashboards, providing a holistic understanding of the hiring landscape. The findings are contextualized to assist students, job seekers, and organizations in making data-informed decisions about employment trends and opportunities.

Chapter 6: Python Implementation

Python serves as the primary environment for data preprocessing and exploratory data analysis in this project. Libraries such as **pandas**, **numpy**, **matplotlib**, and **seaborn** are employed for cleaning, summarizing, and visualizing the job market dataset. The workflow begins by importing the dataset, standardizing column names, and converting relevant variables into suitable data types. Missing or inconsistent entries in fields such as *employment_type* and *industry* are systematically handled through imputation or removal to ensure data reliability.

Visualizations form a key part of the Python implementation. Bar charts and count plots are used to illustrate hiring frequency across different **industries**, **companies**, and **employment types**. Line plots help to analyze **temporal trends** in job postings over various dates or years, while heatmaps provide an overview of correlations among variables such as **location**, **seniority level**, and **job function**. Comparative plots are also generated to identify which industries or job roles show the highest concentration of openings.

Feature engineering is incorporated to enhance analytical insight. New variables such as **job density by location**, **average hiring rate per company**, and **industry-wise distribution ratios** are derived to enable deeper comparisons. These refined metrics improve the interpretability of the data and support the subsequent visualization phase.

All visual outputs are saved and documented for use in **Tableau dashboards**, ensuring consistency between statistical findings and visual storytelling.

In essence, Python establishes a transparent, reproducible, and data-driven foundation for this analysis, enabling accurate preprocessing, reliable pattern detection, and smooth integration with advanced visualization platforms like Tableau.

Chapter 7: Power BI Dashboard

Power BI is used to create interactive dashboards, offering stakeholders a business friendly way to explore the dataset. Data is imported from the cleaned CSV generated through Python preprocessing. Within Power BI, fields are classified appropriately numeric values for social media time and productivity scores, categorical values for job type and platform preference.

Visualizations include:

- **Line and bar charts** comparing employee_type by company_name and hiring_status
- **Clustered column chart** displaying correlations between industry and employment_type
- **Stacked bar charts** showing differences in job title by employment typr
- **Pie chart** for users to explore the dataset by hiring status vs month

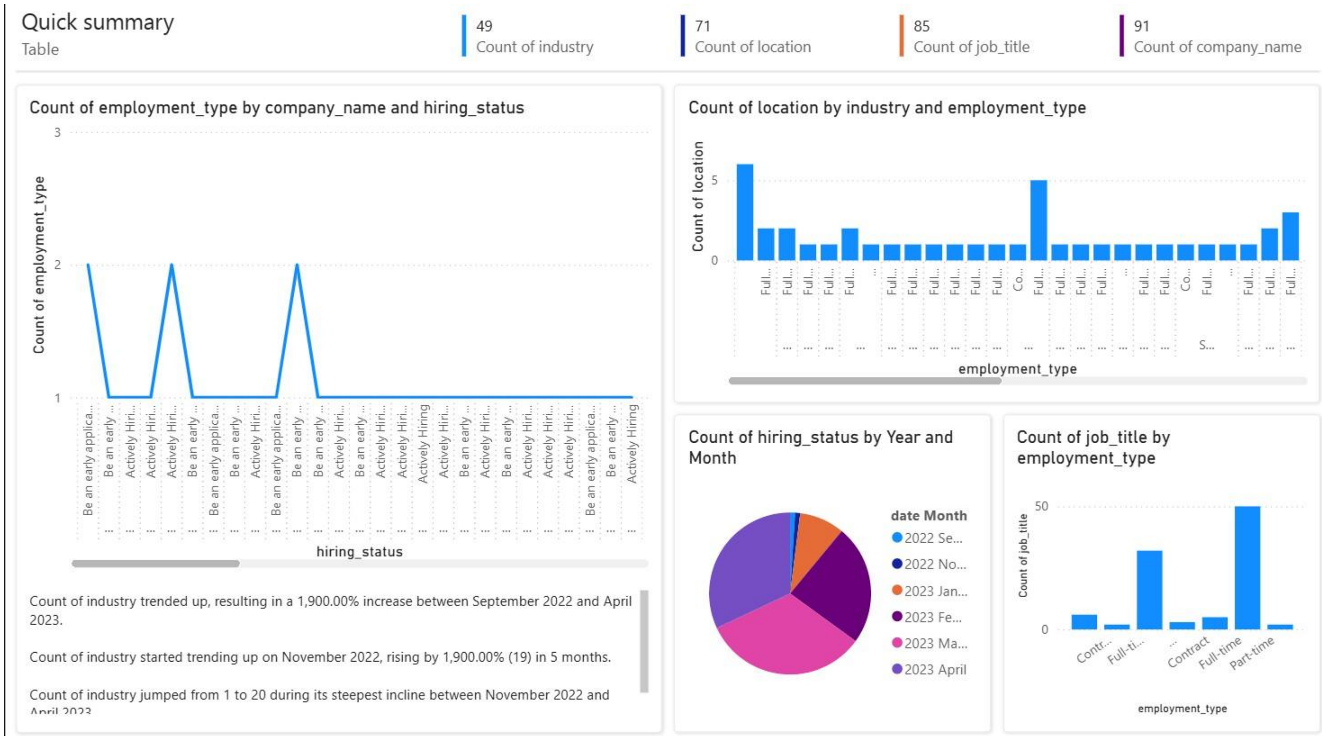


Fig 7.1: Power BI Dashboard

Chapter 8: Tableau Dashboard

Tableau complements Power BI by emphasizing visually compelling dashboards that are ideal for data-driven storytelling and professional presentations.

In this Job Market Analysis, the cleaned dataset is imported into Tableau, and calculated fields are created — such as the **employment-to-population ratio**, **average salary by sector**, and **unemployment trend by education level**.

This Tableau dashboard focuses on **analyzing job market patterns and employment trends** across regions, industries, and demographics. Multiple sheets are integrated into an interactive dashboard that highlights key insights like **in-demand skills**, **salary distribution**, and **sectoral growth**.

The storytelling capability of Tableau makes it highly effective for presenting job market insights to stakeholders, policymakers, and students. **Bar charts, heat maps, and trend lines** are used to visualize shifts in employment rates, helping users understand the current workforce landscape in a clear, engaging way.

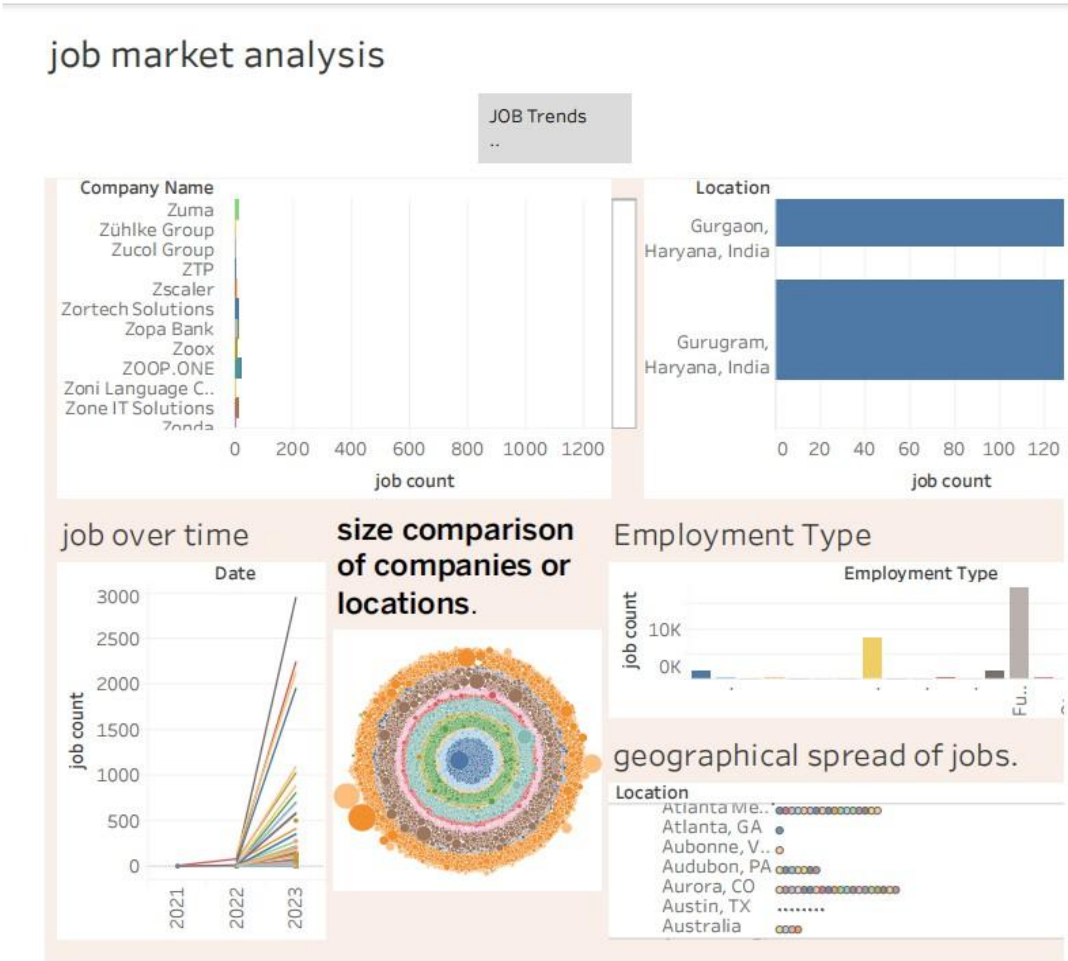


Fig 8.1: Tableau Dashboard

Chapter 9: Analysis

The analysis of the job market dataset reveals several key patterns and insights. First, **education level and technical proficiency** show a strong positive correlation with both **employability** and **average salary**. Candidates with specialized skills in emerging technologies—such as data analytics, cloud computing, and AI—tend to secure higher-paying roles with shorter job search durations. Conversely, individuals lacking digital skills or industry certifications experience longer periods of unemployment and lower wage growth.

Second, **regional and industry-specific variations** emerge. Metropolitan regions and technology-driven industries (e.g., IT, finance, and consulting) exhibit higher employment rates and stronger demand for hybrid and remote roles. In contrast, traditional sectors like manufacturing and agriculture show slower job growth but greater stability in long-term employment.

Third, **demographic factors** influence participation and outcomes. Younger professionals report higher job mobility and openness to contract or freelance work, while mid-career professionals tend to value job security and career progression opportunities. Gender differences are evident in certain sectors—particularly STEM—where female participation remains comparatively lower despite similar qualification levels.

Furthermore, **soft skills** such as communication, adaptability, and teamwork are increasingly cited by recruiters as critical for employability. The gap between **academic qualifications and job-ready skills** persists, particularly among recent graduates, underscoring the need for continuous upskilling and real-world project exposure.

.

Chapter 9: Conclusion

The study concludes that the **job market is undergoing a major transition** toward skill-based hiring and digital readiness. While education and experience remain important, employers increasingly prioritize **technical competence, adaptability, and domain-specific certifications**. Job seekers who engage in upskilling—through online courses, workshops, and internships—achieve higher employability and better career outcomes.

Regional and industry variations emphasize the importance of **geographic and sectoral alignment** between skills and job availability. Policymakers and institutions should therefore focus on bridging the gap between academia and industry through **curriculum modernization and internship-driven programs**.

From a workforce perspective, individuals should actively pursue **data literacy, digital communication, and analytical thinking**, as these are becoming baseline requirements in nearly all professions. Employers, on the other hand, should invest in **reskilling and employee development programs** to maintain competitiveness in an AI-driven economy.

For future research, expanding the dataset to include **time-series employment data** would enable analysis of trends in job creation and automation impact. Integrating **machine learning prediction models** could help forecast **in-demand skills and emerging job sectors**, while cross-country comparisons could provide a global perspective on workforce adaptation.

Ultimately, this analysis underscores the shift from degree-based employment to **skill-driven employability**, marking a transformative phase in how individuals and organizations prepare for the future of work.

Chapter 10: Appendix

```
# ============================
# STEP 1: MANUAL UPLOAD
# ============================

from google.colab import files
import pandas as pd
import numpy as np
import os
```

```
print("📂 Please upload your Excel file (e.g., linkedin_job_posts_insights.xlsx)")
uploaded = files.upload()
```

```
# Automatically detect uploaded file name
excel_path = list(uploaded.keys())[0]
print(f"✅ Uploaded file: {excel_path}")
```

```
# ============================
# STEP 2: LOAD DATA
# ============================
```

```
df = pd.read_excel(excel_path)
print("✅ Dataset Loaded Successfully!\n")
```

```
print("◇ First 5 rows:")
display(df.head())
```

```
# ============================
# STEP 3: DATA OVERVIEW
# ============================
```

```
print("\nShape:", df.shape)
print("\nData Types:\n", df.dtypes)
print("\nMissing Values:\n", df.isnull().sum())
print("\nBasic Statistics:\n", df.describe(include='all'))
```

```
# ============================
# STEP 4: DATA CLEANING
# ============================
```

```
# 1️⃣ Remove duplicates
df = df.drop_duplicates()
```

```
# 2️⃣ Handle missing values
fill_values = {
    'company_name': 'Unknown',
    'location': 'Unknown',
    'job_title': 'Unknown',
    'industry': 'Unknown',
    'employment_type': 'Unknown'
}
df = df.fillna(fill_values)
```

```
# Drop columns with more than 40% missing data
threshold = 0.4 * len(df)
df = df.dropna(axis=1, thresh=threshold)
```

```python
# 3 Convert data types (ensure all strings)
for col in df.select_dtypes(include='object'):
    df[col] = df[col].astype(str)

print("\n✅ After Cleaning:")
df.info()

# ===========================
# STEP 5: FILTERING / SUBSETTING
# ===========================

india_jobs = df[df['location'].str.contains('India', case=False, na=False)]
print(f"\n🌏 Jobs in India: {india_jobs.shape[0]}")

data_roles = df[df['job_title'].str.contains('Data|ML|AI', case=False, na=False)]
print(f"🤖 Data-related roles: {data_roles.shape[0]}")

# ===========================
# STEP 6: NORMALIZATION & ENCODING
# ===========================

from sklearn.preprocessing import LabelEncoder, MinMaxScaler

# Encode categorical columns
cat_cols = ['company_name', 'location', 'seniority_level', 'job_function',
            'employment_type', 'industry', 'hiring_status']
le = LabelEncoder()

for col in cat_cols:
    if col in df.columns:
        df[col] = le.fit_transform(df[col].astype(str))

# Normalize numeric columns (if any)
numeric_cols = df.select_dtypes(include=['int64', 'float64']).columns
scaler = MinMaxScaler()
df[numeric_cols] = scaler.fit_transform(df[numeric_cols])

print("\n✅ Normalization & Encoding Completed!")

# ===========================
# STEP 7: EXPLORATORY DATA ANALYSIS (EDA)
# ===========================

import seaborn as sns
import matplotlib.pyplot as plt
from collections import Counter
import plotly.express as px

print("\nD Performing Exploratory Data Analysis...")

# Re-load original for readable plots
original_df = pd.read_excel(excel_path)
original_df['job_title'] = original_df['job_title'].astype(str)

# --- 1 Top Hiring Companies ---
top_companies = original_df['company_name'].value_counts().head(10)
plt.figure(figsize=(10,5))
sns.barplot(x=top_companies.values, y=top_companies.index, palette="Blues_r")
plt.title("Top 10 Hiring Companies")
plt.xlabel("Number of Jobs")
plt.show()

# --- 2 Top Job Locations ---
top_locations = original_df['location'].value_counts().head(10)
plt.figure(figsize=(10,5))
```

```python
sns.barplot(x=top_locations.values, y=top_locations.index, palette="coolwarm")
plt.title("Top 10 Job Locations")
plt.show()
```

```python
# --- 3️⃣ Word Frequency in Job Titles ---
words = ' '.join(original_df['job_title']).split()
common_words = Counter(words).most_common(15)
plt.figure(figsize=(10,5))
sns.barplot(x=[w[1] for w in common_words], y=[w[0] for w in common_words], palette="Purples_r")
plt.title("Most Frequent Words in Job Titles")
plt.xlabel("Frequency")
plt.ylabel("Word")
plt.show()
```

```python
# --- 4️⃣ Seniority Level Distribution ---
if 'seniority_level' in original_df.columns:
    plt.figure(figsize=(8,4))
    sns.countplot(y='seniority_level', data=original_df, order=original_df['seniority_level'].value_counts().index,
palette="Greens_r")
    plt.title("Seniority Level Distribution")
    plt.show()
```

```python
# --- 5️⃣ Job Function Distribution ---
if 'job_function' in original_df.columns:
    plt.figure(figsize=(8,4))
    sns.countplot(y='job_function', data=original_df, order=original_df['job_function'].value_counts().index, palette="Oranges_r")
    plt.title("Job Function Distribution")
    plt.show()
```

```python
# --- 6️⃣ Employment Type Distribution ---
if 'employment_type' in original_df.columns:
    plt.figure(figsize=(8,4))
    sns.countplot(y='employment_type', data=original_df, order=original_df['employment_type'].value_counts().index,
palette="Reds_r")
    plt.title("Employment Type Distribution")
    plt.show()
```

```python
# --- 7️⃣ Industry Distribution ---
if 'industry' in original_df.columns:
    plt.figure(figsize=(8,4))
    sns.countplot(y='industry', data=original_df, order=original_df['industry'].value_counts().head(10).index, palette="mako")
    plt.title("Top 10 Industries Hiring")
    plt.show()
```

```python
# --- 8️⃣ Correlation Heatmap ---
numeric_df = df.select_dtypes(include=['float64','int64'])
if not numeric_df.empty:
    corr = numeric_df.corr()
    plt.figure(figsize=(10,8))
    sns.heatmap(corr, annot=True, cmap='viridis')
    plt.title("Correlation Heatmap")
    plt.show()
```

```python
# --- 9️⃣ Interactive Plots ---
fig = px.bar(top_companies, title="Top Hiring Companies (Interactive)",
             labels={'index':'Company', 'value':'Job Count'})
fig.show()
```

```python
fig = px.bar(top_locations, title="Top Job Locations (Interactive)",
             labels={'index':'Location', 'value':'Job Count'})
fig.show()
```

```python
# ===========================
# STEP 8: EXPORT CLEANED DATASET
# ===========================
```

```python
clean_path = "/content/cleaned_linkedin_jobs.xlsx"
```

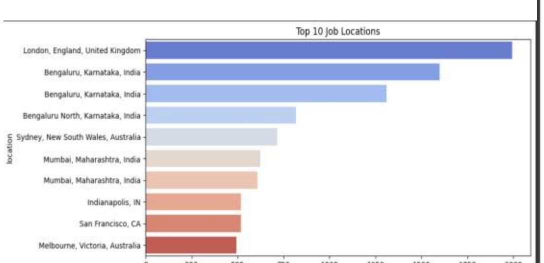
```python
df.to_excel(clean_path, index=False)
print(f"\n✅ Cleaned dataset exported to: {clean_path}")
```

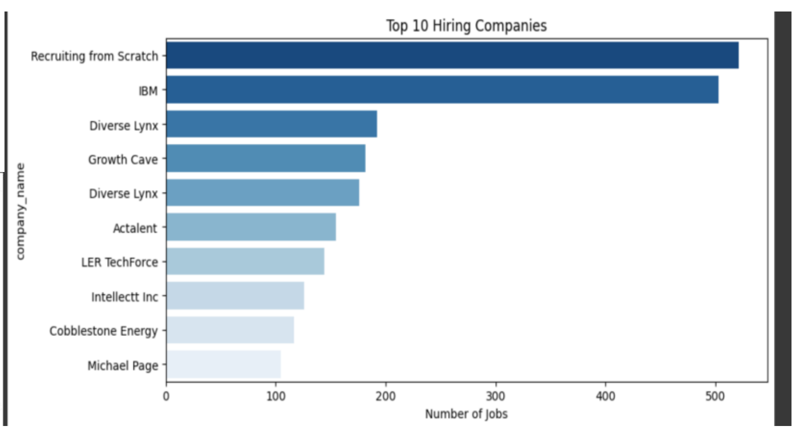
🔹 First 5 rows:

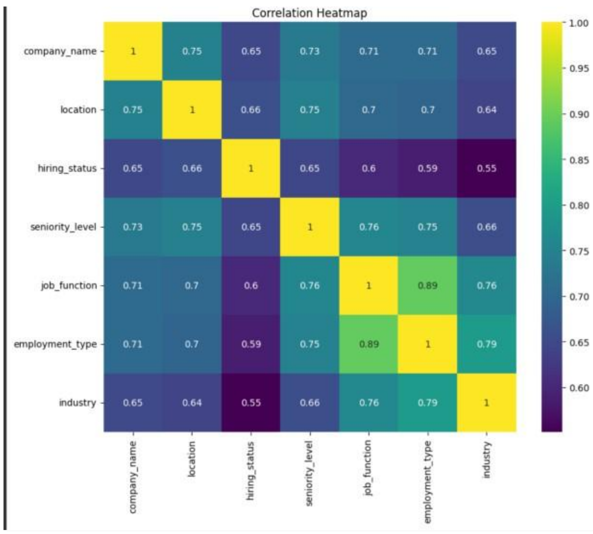
	job_title	company_name	location	hiring_status	date	seniority_level	job_function	employment_type	industry
0	Store Business Manager - DAVID JONES CHERMSIDE	M.J. Bale	Brisbane, Queensland, Australia	Be an early applicant	2023-04-13	Not Applicable	Sales and Business Development	Full-time	Government Administration
1	Full-time	Gatesman	Chicago, IL	Be an early applicant	2023-03-31	NaN	NaN	NaN	NaN
2	Senior Machine Learning Engineer	Redwolf + Rosch	Adelaide, South Australia, Australia	Be an early applicant	2023-04-25	Mid-Senior level	Engineering and Information Technology	Part-time	Internet Publishing
3	Senior Data Scientist	Bupa	Melbourne, Victoria, Australia	Be an early applicant	2023-04-29	Entry level	Engineering and Information Technology	Full-time	Technology, Information and Internet
4	\n \n Solution Architect\n ...	\n Xybion Digital\n	\n Chennai, Tamil Nadu, India\n ...	\n Be an early applicant\n	2023-01-26	\n Mid-Senior level\n	\n Engineering and Information Tech...	\n Full-time\n	\n IT Services and IT Consulting, Sof...

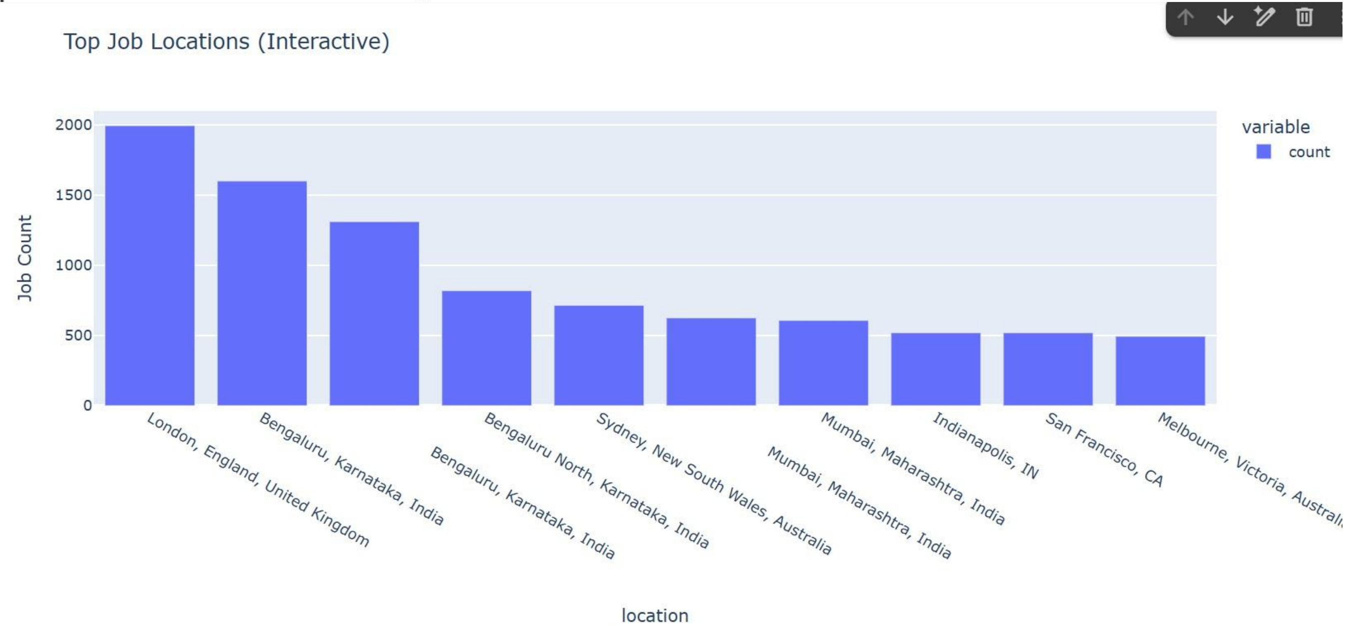
```
Shape: (31597, 9)

Data Types:
 job_title                  object
company_name               object
location                   object
hiring_status              object
date               datetime64[ns]
seniority_level            object
job_function               object
employment_type            object
industry                   object
dtype: object

Missing Values:
 job_title           26
company_name       940
location             9
hiring_status        0
date                 0
seniority_level   1308
job_function      1590
employment_type   1591
industry          2011
dtype: int64
```

```
Basic Statistics:
                   job_title            company_name  \
count                  31571                   30657
unique                  6112                    7201
top     Full Stack Developer  Recruiting from Scratch
freq                     718                     521
mean                     NaN                     NaN
min                      NaN                     NaN
25%                      NaN                     NaN
50%                      NaN                     NaN
75%                      NaN                     NaN
max                      NaN                     NaN

                              location          hiring_status  \
count                            31588                  31597
unique                            2043                     47
top     London, England, United Kingdom  Be an early applicant
freq                              1994                  14423
mean                               NaN                    NaN
min                                NaN                    NaN
25%                                NaN                    NaN
50%                                NaN                    NaN
75%                                NaN                    NaN
max                                NaN                    NaN

                                 date   seniority_level  \
count                           31597             30289
unique                            NaN                25
top                               NaN  Mid-Senior level
freq                              NaN              8651
mean    2023-03-12 20:52:44.199132928               NaN
min               2021-05-27 00:00:00               NaN
25%               2023-02-23 00:00:00               NaN
50%               2023-03-20 00:00:00               NaN
75%               2023-04-08 00:00:00               NaN
max               2023-04-29 00:00:00               NaN

                                  job_function employment_type  \
count                                    30007           30006
unique                                     542              18
top     Engineering and Information Technology       Full-time
freq                                      7141           18305
mean                                       NaN             NaN
min                                        NaN             NaN
```









✅ Cleaned dataset exported to: /content/cleaned_linkedin_jobs.xlsx