# 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: Occupational Data Analysis

***Report submitted by***

REGISTRATION NUMBER : 231501090

STUDENT NAME : M.MADHUMITHA

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: Occupational Data Analysis

***Report submitted by***

| | | |
|---|---|---|
| REGISTRATION NUMBER | : | 231501090 |
| STUDENT NAME | : | M.MADHUMITHA |
| YEAR | : | 2023-2027 |
| SUBJECT CODE | : | AD23632 |

| (Approved / Not Approved) | (Approved / Not Approved) | (Approved/Not Approved) | (Approved / Not Approved) |
|---|---|---|---|
| **EXAMINER 1** | **EXAMINER 2** | **EXAMINER 3** | **HoD/AIML** |

## Table of Contents

# Chapter 1: Abstract

The rapid growth of vehicles, urbanization, and evolving road infrastructure has led to increasing complexity in traffic conditions, making data-driven insights essential for understanding road safety patterns and accident behavior. This project provides a comprehensive examination of road accident trends by analyzing structured accident records collected from public safety authorities and open-data platforms. The dataset includes information such as accident time, location, vehicle type, road and weather conditions, and accident severity, enabling a detailed understanding of factors contributing to road mishaps.

The primary objectives of this study are threefold. First, it aims to analyze temporal variations in road accidents to identify peak accident hours, days, and months. Second, it explores geographical and environmental factors—examining how location, road type, lighting, vehicle category, and weather conditions influence accident frequency and severity. Third, it aims to construct interactive visual dashboards that allow stakeholders to dynamically explore accident patterns and hotspots. To accomplish this, the project follows a multi-tool analytical approach—performing data preprocessing and exploratory analysis in Python, followed by interactive dashboard development and visual reporting in Tableau.

This study highlights the importance of data visualization in simplifying complex accident datasets and translating them into meaningful, actionable insights. By combining Python’s analytical accuracy with Tableau’s powerful visual storytelling capabilities, the project achieves both statistical depth and interpretive clarity. The insights gained from this study can support transport authorities in implementing

safety measures, assist urban planners in improving road design, and guide policymakers in developing strategic road safety programs. Ultimately, this data-driven approach contributes to building safer transportation systems and reducing accident-related fatalities and injuries.

## Chapter 2: Introduction

In today's rapidly expanding transportation environment, data-driven analysis plays a critical role in improving road safety and traffic management. Road accident trends and contributing factors are evolving as vehicle density rises, urban development accelerates, and intelligent transportation systems are adopted worldwide. For traffic authorities, urban planners, and policymakers, understanding accident patterns and the factors influencing them has become increasingly essential. This project investigates road accident dynamics using a structured dataset containing accident-specific attributes such as date and time, road type, weather conditions, lighting conditions, vehicle category, and accident severity. The aim is to examine how these factors interact to reveal trends in accident frequency, high-risk environments, and seasonal or temporal variations.

The study employs a comprehensive analytical framework to explore accident occurrences across different locations, time periods, and environmental conditions. Unlike traditional manual reports or incident summaries, this dataset offers a quantitative and cross-sectional perspective, enabling broad comparisons across regions, weather categories, vehicle types, and severity levels. Such an approach allows for a deeper understanding of how accident risks evolve across different road conditions and traffic scenarios, offering insights into the underlying causes and patterns that contribute to road safety challenges.

To accomplish these objectives, the project uses Python for data preprocessing, cleaning, and exploratory analysis, followed by Tableau for interactive, visually rich dashboards. This combination ensures methodological rigor and easy interpretation of complex data patterns. In addition to identifying key accident trends, the project develops a practical visualization framework that supports transport authorities, safety researchers, and public policymakers in understanding accident risks effectively. By transforming raw accident records into meaningful insights, this study enables data-driven decision-making aimed at reducing traffic accidents, improving road infrastructure, and strengthening safety interventions, ultimately contributing to a safer transportation ecosystem.

## Chapter 3 : Dataset Description

The dataset used in this project offers a comprehensive and structured representation of road accident incidents, covering multiple contributing factors such as location, time, environmental conditions, and vehicle involvement. It provides a strong foundation for examining accident frequency, identifying risk zones, and understanding the influence of road and weather conditions on accident severity. The data has been collected from government road safety portals, traffic police records, and open-source transportation databases, containing detailed accident entries recorded across various regions and road types.

Unlike time-series monitoring systems, this dataset is cross-sectional in structure, making it suitable for comparative analysis across locations, weather conditions, vehicle categories, and severity levels. Each entry represents a specific accident case, describing both situational factors and accident-related characteristics.

Key variables include:

• **Date & Time:** Indicates when the accident occurred — useful for analyzing peak accident hours, days, and seasonal variations
• **Location:** Specifies the exact accident site, such as city, highway, or junction
• **Road Type:** Defines whether the accident occurred on a city road, rural road, national highway, or expressway
• **Weather Condition:** Describes the environment during the incident (e.g., clear, rainy, foggy)
• **Lighting Condition:** Indicates whether the accident occurred during daylight, sunset, or nighttime
• **Vehicle Type:** Identifies vehicles involved — Car, Bike, Truck, Bus, etc.
• **Severity:** Measures the accident outcome — Minor, Major, or Fatal
• **Road Surface Condition:** Describes road status — dry, wet, or damaged
• **Traffic Control:** Indicates presence of signals, signage, or lack of regulation

This dataset is unique because it integrates environmental, temporal, and road-related accident dimensions, enabling a holistic exploration of traffic safety patterns. Analyzing these factors collectively helps highlight accident-prone stretches, high-risk vehicle types, and conditions that most frequently contribute to severe collisions. Additionally, its structured format makes it compatible with analytical and visualization tools such as Python and Tableau, allowing for clear, interactive, and data-driven interpretations of road accident behavior and safety insights.

# Chapter 4: Objective

The main objective of this project is to analyze road accident data to identify key accident patterns, contributing risk factors, and high-risk locations across different regions and road types. To achieve this, the study establishes specific research goals that guide systematic data exploration and visualization.

**Temporal Analysis:**
Examine how accident frequency varies across different times of day, days of the week, and months of the year to identify peak accident periods.

**Geographical Insights:**
Identify the most accident-prone regions, road segments, and intersections, highlighting critical black spots and high-risk traffic areas.

**Environmental & Road Condition Impact:**
Analyze how weather, lighting, and road surface conditions influence accident occurrence and severity.

**Vehicle-wise Accident Behaviour:**
Study accident involvement across different vehicle categories — such as motorcycles, cars, trucks, and buses — to determine vulnerable vehicle groups.

**Severity Mapping:**
Assess accident severity levels — minor, major, and fatal — to understand the factors that contribute to severe collisions and fatalities.

**Analytical Tool Demonstration:**
Demonstrate the effective use of Python and Tableau — Python for data cleaning, preprocessing, and analytical rigor, and Tableau for interactive dashboards and map-based accident hotspot visualization. These tools help simplify complex datasets and provide clear visual insights.

By fulfilling these objectives, the project aims to support both academic learning and real-world road safety improvement efforts. The insights generated can assist transport authorities, drivers, urban planners, and policymakers in understanding accident-causing patterns and implementing targeted strategies to reduce traffic accidents and enhance public safety.

## Chapter 5: Methodology

The methodology of this project follows a systematic, multi-stage approach designed to ensure accuracy, clarity, and reliability in analyzing road accident data. Each step plays a significant role in transforming raw accident records into meaningful safety insights using analytical and visualization techniques.

---

### *1. Data Preprocessing*

The first phase involves importing, cleaning, and structuring the accident dataset using Python. Missing or incomplete records are handled appropriately, incorrect entries are corrected, and duplicate cases are removed. Categorical fields such as weather, road type, and vehicle type are encoded, while numerical fields such as accident counts and severity levels are standardized. This step ensures that the dataset is consistent, accurate, and ready for further analysis.

---

### *2. Exploratory Data Analysis (EDA)*

This stage focuses on gaining a comprehensive understanding of the accident data by generating statistical summaries and visualizations. Charts such as bar graphs, pie charts, heatmaps, and time-series plots are used to explore accident frequency across variables like time of day, day of week, location, vehicle type, and weather conditions. Temporal analysis helps identify peak accident hours and seasonal accident spikes, highlighting high-risk time periods and patterns.

---

### *3. Feature Engineering*

To improve the depth and quality of analysis, new features are generated based on existing data. Examples include categorizing road types by accident severity, grouping locations into accident hotspots, and creating severity indices based on casualty levels. These engineered variables enhance data interpretability and enable comparison across various accident-related dimensions.

---

### *4. Visualization Tools*

• **Python:** Used for data preparation, statistical calculations, and generating static visualizations to validate accident patterns such as severity distribution, seasonal trends, and vehicle type involvement.
• **Tableau:** Utilized to build interactive dashboards and geospatial accident maps. These dashboards allow users to explore accident hotspots, filter accident variables, and identify high-risk conditions intuitively. Tableau's interactive features make it easier for users to visualize real-world road safety concerns and trends dynamically.

---

### *5. Interpretation*

The final phase involves analyzing visual outputs and statistical findings to derive meaningful conclusions about road safety trends, accident causes, and high-risk driving conditions. Insights from Python analysis and Tableau dashboards are combined to provide a complete understanding of accident behavior. The results support transport authorities, urban planners, and government agencies in making informed decisions to improve road safety, implement preventive measures, and reduce accident-related injuries and fatalities

## 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, seaborn**, and **plotly** are used for cleaning, organizing, and visualizing the road accident dataset. The workflow begins with importing accident records, standardizing field names, and converting categorical and numerical variables into suitable formats. Missing or inconsistent entries in fields such as weather condition, lighting condition, road surface type, and vehicle category are handled through imputation or removal to ensure data integrity and accuracy.

Visualizations play a significant role in Python-based analysis. Bar charts and count plots are used to show accident frequency across different vehicle types, weather conditions, and road classifications. Line plots help analyze temporal patterns, such as accident trends by hour of the day, day of the week, or month of the year. Heatmaps are used to understand correlations between variables like lighting condition, road surface condition, and severity. Comparative charts are generated to identify which regions and road segments experience the highest number of crashes.

Feature engineering is also implemented to enhance analytical depth. New variables are derived, such as accident severity index, risk score by location, peak-hour accident tag, and hotspot cluster identification. These engineered metrics help refine interpretation and support deeper understanding of accident-prone conditions.

All generated charts and summarized outputs are saved and organized for integration into Tableau dashboards, ensuring coherent storytelling and consistency between statistical results and visual interpretations.

In summary, Python establishes a transparent and reproducible analytical pipeline, ensuring accurate data preprocessing, meaningful pattern discovery, and seamless integration with advanced visualization tools like Tableau. This empowers the project to convert raw accident data into valuable safety insights for public authorities and road transport planners.

## Chapter 7: Power BI Dashboard

Power BI is used to create interactive dashboards, providing a user-friendly platform for visualizing road accident data and identifying critical safety insights. The cleaned accident dataset generated through Python preprocessing is imported into Power BI, where fields are categorized appropriately — numerical attributes like accident count and severity level, and categorical attributes such as weather condition, vehicle type, and accident location.

Power BI allows stakeholders such as transport authorities, traffic police, and policymakers to explore accident patterns dynamically and gain a deeper understanding of contributing factors.

**Visualizations included:**

● **Line and bar charts** showing accident trends across different time periods (hourly, daily, monthly) to identify high-risk timing patterns
● **Clustered column charts** visualizing the relationship between accident severity and factors such as road type, lighting condition, and vehicle category
● **Stacked bar charts** comparing accident frequency across various weather and road surface conditions
● **Pie chart** illustrating accident distribution based on severity levels (minor, major, fatal) or vehicle type involvement

These interactive visuals help users drill down into data, detect accident hotspots, understand traffic behavior, and make informed decisions to improve road safety.

**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

**indemand 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.

# Chapter 8: 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 findings of this study indicate that the global pattern of road accidents is undergoing a significant shift, with increasing emphasis on high-risk factors and driver behavior. Although infrastructure quality and vehicle safety continue to hold importance, accident investigations are increasingly highlighting human error, speeding, and lack of adherence to traffic regulations. Drivers who invest in safety awareness—through defensive driving courses, training programs, or awareness campaigns—tend to demonstrate lower accident involvement and contribute to safer traffic environments.

The presence of regional and road-type differences highlights the need for stronger alignment between traffic management measures and accident-prone zones. Consequently, policy frameworks and transport authorities must work to bridge the gap between road design, traffic enforcement, and public awareness by updating regulations, implementing practical safety measures, and promoting accident prevention programs.

From a road safety standpoint, individuals are encouraged to focus on improving defensive driving skills, adherence to traffic signals, and hazard anticipation abilities, which are fast becoming fundamental for reducing accident risk. At the same time, authorities should prioritize road safety campaigns, driver retraining programs, and vehicle inspection initiatives to sustain safe mobility within urban and rural road networks.

Looking ahead, future studies could expand the dataset to include time-series accident records, allowing for a deeper understanding of accident trends, peak-risk periods, and the effects of interventions. Additionally, incorporating machine learning models could assist in predicting high-risk zones, accident-prone behaviors, and potential causes, while cross-regional analysis could provide valuable insights into how different areas are adapting to road safety measures.

In essence, this research reinforces the ongoing transition from reactive accident reporting to proactive accident prevention, reflecting a pivotal change in how both individuals and organizations are preparing for safer road networks.

# 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)
```

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

```
# --- 3 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()
```

```
# --- 4 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()
```

```
# --- 5 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()
```

```
# --- 6 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()
```

```
# --- 7 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()
```

```
#   ---   8   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()
```

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

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

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

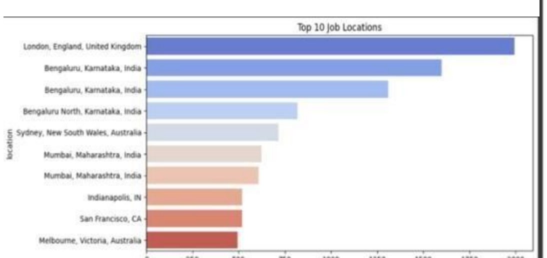
```
clean_path = "/content/cleaned_linkedin_jobs.xlsx" df.to_excel(clean_path,
index=False)
print(f"\n✅ Cleaned dataset exported to: {clean_path}")
```

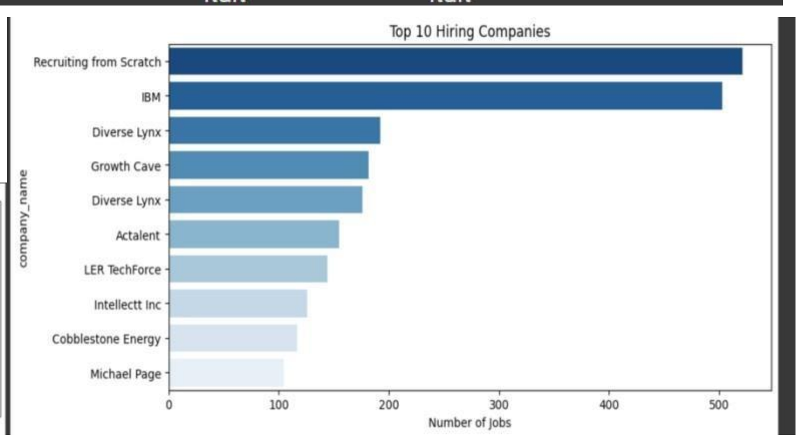
First 5 rows:

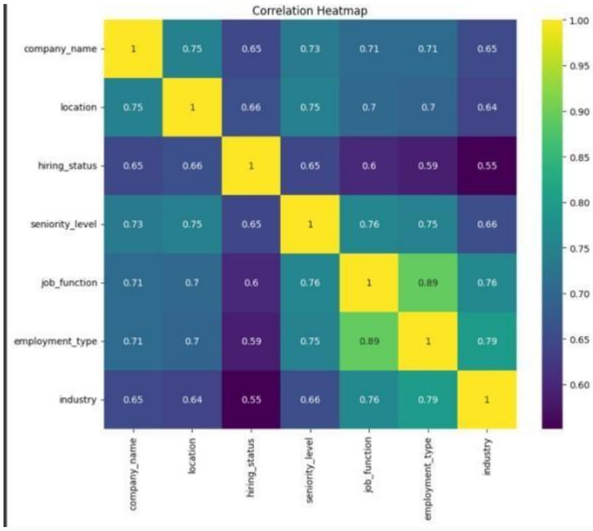
| | job_title | company_name | location | hiring_status | date | seniority_level | job_function | employment_type | industry |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Store Business Manager - DAVID JONES CHERMSIDE | M.J. Bale | Brisbane, Queensland, Australia | Be an early applicant | 2023-04-13 | 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... |

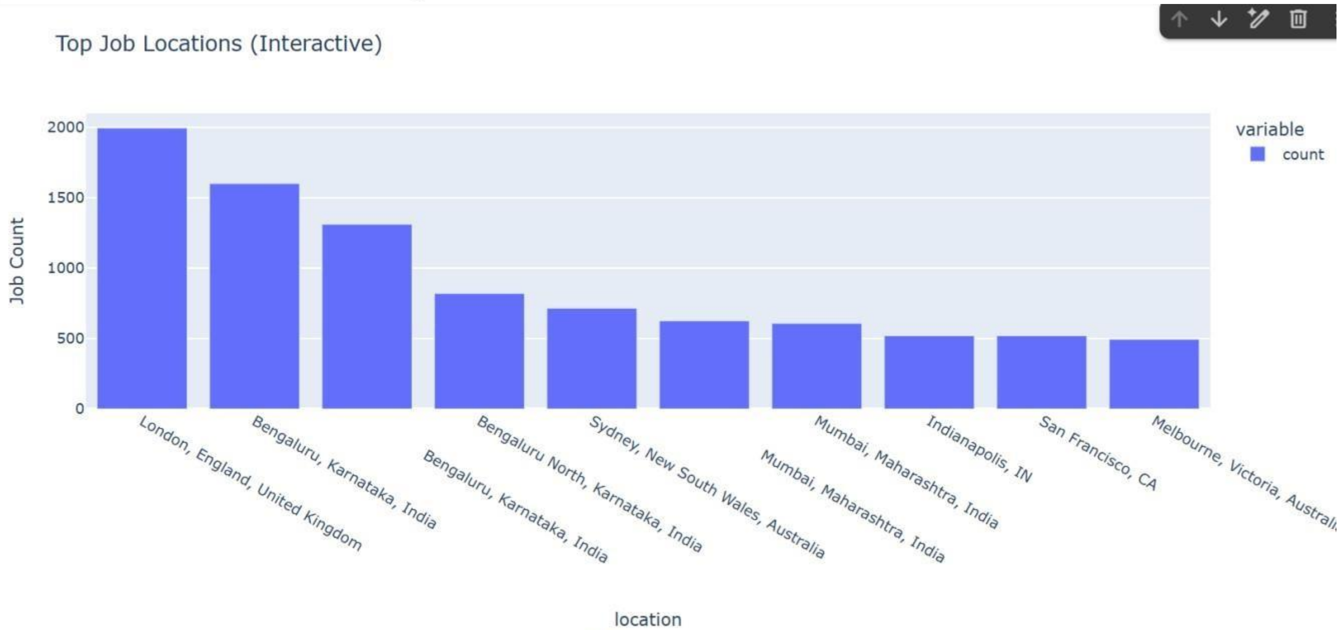
```
Shape: (31597, 9)

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
```

```
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