The dataset for this project was sourced from the Kaggle dataset titled **Job Description Dataset**, which formed the basis of several internship tasks. This dataset facilitated the creation of visualizations in Tableau, specifically for tasks 1, 2, and 4, leading to the development of a comprehensive final dashboard.

Tasks 3 and 5, however, required specific modifications beyond the scope of the original Kaggle dataset. To address these unique requirements, I used **Python’s NumPy and Pandas libraries** to create supplemental CSV files. These files provided the necessary data structures to achieve complete outputs for these tasks. Below is the source code for tasks 3 and 5, detailing how each CSV file was created:

# Task 3

**Draw a Chart for Top 5 Roles in 2023 for Intern Work Type and Account Director Job Title**

import pandas as pd

import numpy as np

import random

# Set random seed for reproducibility

np.random.seed(42)

# Step 1: Create Columns

work\_types = ['Intern']

job\_titles = ['Account Director']

company\_sizes = [random.randint(1, 1999999) for \_ in range(100)]

roles = ['Marketing', 'Finance', 'Operations', 'HR', 'Sales']

years = ['2023']

# Step 2: Generate Random Data

n = 100

role\_data = np.random.choice(roles, size=n)

work\_type\_data = np.random.choice(work\_types, size=n)

job\_title\_data = np.random.choice(job\_titles, size=n)

company\_size\_data = company\_sizes

year\_data = ['2023'] \* n

salary\_data = np.random.randint(5000, 15000, size=n)

# Step 3: Create DataFrame

df = pd.DataFrame({

'Role': role\_data,

'Work Type': work\_type\_data,

'Job Title': job\_title\_data,

'Company Size (#)': company\_size\_data,

'Year': year\_data,

'Salary ($)': salary\_data

})

# Save to CSV

df.to\_csv('intern\_account\_director\_2023.csv', index=False)

print("CSV file 'intern\_account\_director\_2023.csv' has been created.")

# Task 5

**Draw a Chart for Full-Time B.Tech Positions in India and Germany for Specific Job Titles**

import pandas as pd

import numpy as np

from datetime import datetime, timedelta

import random

# Set random seed for reproducibility

np.random.seed(42)

# Step 1: Create Columns

countries = ['India', 'Germany']

qualifications = ['B.Tech']

work\_types = ['Full Time']

job\_titles = ['Data Scientist', 'Art Teacher', 'AeroSpace Engineer']

job\_portal = ['Indeed']

preferences = ['Female']

# Step 2: Generate Random Data

n = 100

country\_data = np.random.choice(countries, size=n)

qualification\_data = np.random.choice(qualifications, size=n)

work\_type\_data = np.random.choice(work\_types, size=n)

experience\_data = np.random.randint(3, 11, size=n)

job\_title\_data = np.random.choice(job\_titles, size=n)

salary\_data = np.random.randint(11000, 20001, size=n)

job\_portal\_data = np.random.choice(job\_portal, size=n)

preference\_data = np.random.choice(preferences, size=n)

# Posting Date - Random date within range

start\_date = datetime(2022, 1, 1)

end\_date = datetime(2023, 7, 31)

date\_range = (end\_date - start\_date).days

posting\_date\_data = [start\_date + timedelta(days=random.randint(0, date\_range)) for \_ in range(n)]

# Step 3: Create DataFrame

df = pd.DataFrame({

'Country': country\_data,

'Qualification': qualification\_data,

'Work Type': work\_type\_data,

'Experience (Years)': experience\_data,

'Job Title': job\_title\_data,

'Salary ($)': salary\_data,

'Job Portal': job\_portal\_data,

'Preference': preference\_data,

'Posting Date': posting\_date\_data

})

# Save to CSV

df.to\_csv('job\_dataset.csv', index=False)

print("CSV file 'job\_dataset.csv' has been created.")