

ex - 1a

October 31, 2025

```
[ ]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
import numpy as np

sns.set(style="whitegrid")

def load_job_data(csv_path=None):
    if csv_path:
        df = pd.read_csv(csv_path, parse_dates=['date_posted'])
        return df
    rng = pd.date_range(start='2015-01-01', end='2024-12-31', freq='D')
    years = rng.year
    base_by_year = {y: 50 + (y - 2015) * 60 for y in range(2015, 2025)}
    counts = [np.random.poisson(lam=max(1, base_by_year[y]/30)) for y in years]
    df = pd.DataFrame({'date_posted': rng, 'postings': counts})
    return df

def aggregate_by_year(df):
    df['date_posted'] = pd.to_datetime(df['date_posted'])
    df['year'] = df['date_posted'].dt.year
    if 'postings' in df.columns:
        yearly = df.groupby('year')['postings'].sum().
↪reset_index(name='num_postings')
    else:
        yearly = df.groupby('year').size().reset_index(name='num_postings')
    return yearly

def plot_trend(yearly_df, title="Data Science Job Postings by Year"):
    plt.figure(figsize=(10,5))
    ax = sns.lineplot(data=yearly_df, x='year', y='num_postings', marker='o')
    ax.set_title(title)
    ax.set_xlabel("Year")
    ax.set_ylabel("Number of Job Postings")
    plt.xticks(yearly_df['year'])
    plt.tight_layout()
```

```
plt.show()

if __name__ == "__main__":
    df = load_job_data(csv_path=None)
    yearly = aggregate_by_year(df)
    print(yearly)
    plot_trend(yearly)
```