

**\*\*Is there a correlation between unemployment rates and crime incidents in the USA from 2004 to 2014?**

**\*\*How do the trends in unemployment rates and crime incidents change over the years from 2004 to 2014?**

Source1 : Kaggle (Dataset: "Unemployment in America Per US State"), URL: <https://www.kaggle.com/datasets/justin2028/unemployment-in-america-per-us-state> Content: State-wise unemployment rates for different years and months. License: Standard open-data license

Source2 : Kaggle (Dataset: "US Crime DataSet") URL: <https://www.kaggle.com/datasets/mrayushagrawal/us-crime-dataset/data> Content: Records of crime incidents across various states, detailing the nature and frequency of crimes.

Data Pipeline: The data pipeline is implemented using Python, leveraging libraries such as pandas for data manipulation, numpy for numerical operations, sqlite3 for database management, and kagglehub for data extraction from Kaggle.


First we will make necessary imports

```
import os
import pandas as pd
import sqlite3
import kagglehub
import seaborn as sns
import matplotlib.pyplot as plt
```

Now we will read the csv file using pandas which has numerous features to extract and transform data

```
df1=pd.read_csv('/content/Unemployment_in_America.csv')
```

```
df1.head()
```




	FIPS Code	State/Area	Year	Month	Total Civilian Non-Institutional Population in State/Area	Total Civilian Labor Force in State/Area	Percent (%) of State/Area's Population	Total Employment in State/Area	Percent (%) of Labor Force Employed in State/Area	Total Unemployment in State/Area	Un St
0	1	Alabama	1976	1	2,605,000	1,484,555	57.0	1,386,023	53.2	98,532	
1	2	Alaska	1976	1	232,000	160,183	69.0	148,820	64.1	11,363	
2	4	Arizona	1976	1	1,621,000	964,120	59.5	865,871	53.4	98,249	
3	5	Arkansas	1976	1	1,536,000	889,044	57.9	824,395	53.7	64,649	
4	6	California	1976	1	15,621,000	9,774,280	62.6	8,875,685	56.8	898,595	

Next steps:

[Generate code with df1](#)
[View recommended plots](#)
[New interactive sheet](#)

```
df2=pd.read_csv('/content/US_Crime_DataSet.csv')
```

 [Show hidden output](#)

```
df2.head()
```

	Record ID	Agency Code	Agency Name	Agency Type	City	State	Year	Month	Incident	Crime Type	...	Victim Ethnicity	Perpetrator
0	1	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	January	1	Murder or Manslaughter	...	Unknown	
1	2	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	March	1	Murder or Manslaughter	...	Unknown	
2	3	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	March	2	Murder or Manslaughter	...	Unknown	Unkr
3	4	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	April	1	Murder or Manslaughter	...	Unknown	
4	5	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	1980	April	2	Murder or Manslaughter	...	Unknown	Unkr

5 rows × 24 columns

class Pipeline:

```
def __init__(self):
    pass
```

```
def transform_data(self):
```

```
    unemployment_data_path = '/content/Unemployment_in_America.csv'
    crime_data_path = '/content/US_Crime_DataSet.csv'
```

```
    df_unemployment = pd.read_csv(unemployment_data_path, delimiter=',')
    df_crime = pd.read_csv(crime_data_path, delimiter=',', low_memory=False)
    df_unemployment = df_unemployment[(df_unemployment['Year'] >= 2004) & (df_unemployment['Year'] <= 2014)]
    df_crime = df_crime[(df_crime['Year'] >= 2004) & (df_crime['Year'] <= 2014) & (df_crime['Perpetrator Age'] >
```

```
    # Selecting only the necessary columns for the analysis
```

```
    df_unemployment = df_unemployment[['State', 'Year', 'Month', 'Percent (%) of Labor Force Unemployed in State/Area']]
    df_crime = df_crime[['State', 'Year', 'Month', 'Incident']]
    merged_df = pd.merge(df_unemployment, df_crime, how='outer', on=['State', 'Year', 'Month'])
    # print("Data transformation complete.")
```

```
    # Sum up incidents for each combination of State, Year, and Month to get the total incidents
```

```
    merged_df = merged_df.groupby(['State', 'Year', 'Month', 'Percent (%) of Labor Force Unemployed in State/Area'])
```

```
    # Aggregate incidents by State and Year
```

```
    yearly_incidents_df = merged_df.groupby(['State', 'Year'], as_index=False).agg({
        'Percent (%) of Labor Force Unemployed in State/Area': 'mean',
        'Incident': 'sum'
    })
```

```
    # Formatting the 'Percent (%) of Labor Force Unemployed in State/Area' column to two decimal places as they are
    yearly_incidents_df['Percent (%) of Labor Force Unemployed in State/Area'] = yearly_incidents_df['Percent (%)
```

```
    # print(yearly_incidents_df.head())
```

```
    yearly_incidents_df.to_csv('/content/Yearly_Aggregated_Unemployment_Crime_Data.csv', index=False)
```

```
    # Save to SQLite
```

```
    conn = sqlite3.connect('/content/Unemployment_Crime_Data.sqlite')
    yearly_incidents_df.to_sql('YearlyAggregatedUnemploymentCrimeData', conn, if_exists='replace', index=False)
    cursor = conn.cursor()
    cursor.execute("SELECT * FROM YearlyAggregatedUnemploymentCrimeData")
    rows = cursor.fetchall()
```

```
    cursor.close()
```

```
    conn.close()
```

```
def query_merged_data(self):
```

```
    conn = sqlite3.connect('/content/Unemployment_Crime_Data.sqlite')
    query = "SELECT * FROM YearlyAggregatedUnemploymentCrimeData"
    merged_df = pd.read_sql_query(query, conn)
    print(merged_df.head())
    conn.close()
```

```
if __name__ == '__main__':
```

```
    pipeline = Pipeline()
    pipeline.transform_data()
    pipeline.query_merged_data()
```

	State	Year	Percent (%) of Labor Force Unemployed in State/Area \
0	Alabama	2004	5.58
1	Alabama	2005	4.46
2	Alabama	2006	4.00
3	Alabama	2007	4.04
4	Alabama	2008	5.88

	Incident
0	245.0
1	583.0
2	621.0
3	502.0
4	542.0

```
# Load the merged and aggregated dataset
```

```
yearly_incidents_df = pd.read_csv('/content/Yearly_Aggregated_Unemployment_Crime_Data.csv')
```

```
# Create a figure with subplots
```

```
fig, axes = plt.subplots(1, 2, figsize=(15, 5)) # 1 row, 2 columns
```

```
# Bar Graph: Total number of crime incidents per state in a specific year (e.g., 2010)
```

```
specific_year = 2010
```

```
year_data = yearly_incidents_df[yearly_incidents_df['Year'] == specific_year]
```

```
sns.barplot(ax=axes[0], x='State', y='Incident', data=year_data)
```

```
axes[0].set_title(f'Total Number of Crime Incidents per State in {specific_year}')
```

```
axes[0].set_xlabel('State')
```

```
axes[0].set_ylabel('Number of Crime Incidents')
```

```
axes[0].tick_params(axis='x', rotation=90)
```

```
axes[0].grid(True)
```

```
# Historical Line Graph: Trend of unemployment rates and crime incidents over the years
```

```
sns.lineplot(ax=axes[1], x='Year', y='Percent (%) of Labor Force Unemployed in State/Area', data=yearly_incidents_df,
```

```
sns.lineplot(ax=axes[1], x='Year', y='Incident', data=yearly_incidents_df, label='Crime Incidents')
```

```
axes[1].set_title('Trend of Unemployment Rates and Crime Incidents (2004-2014)')
```

```
axes[1].set_xlabel('Year')
```

```
axes[1].set_ylabel('Values')
```

```
axes[1].legend()
```

```
axes[1].grid(True)
```

```
plt.tight_layout() # Adjust layout to make it more compact
```

```
plt.savefig('/content/Visualizations_Side_by_Side.png')
```

```
plt.show()
```

