OIL PRODUCTION ANALYSIS

Tools & Technologies

- Python (via IDLE)
- Libraries: numpy, pandas, sqlite3, pathlib
- Database: SQLite
- Data Visualization: Tableau
- Version Control: Git + GitHub

File Overview

OIL_PRODUCTION.csv	Original raw dataset
Oil _production_analysis.py	Data cleaning, analysis, SQLite storage
oil_production_database_file.py	Converts SQLite tables to CSV
Global_production.csv	Derived table for global trend
top_producers.csv	Top 10 countries by latest year
Oil_production.db	SQLite DB with all structured data
oil_production.csv	Derived table for all countries oil production

Step-by-Step Execution

Step 1: Data Cleaning & Analysis

File: oil_production_analysis.py

```
import pandas as pd
import numpy as np
import sqlite3
from pathlib import Path
file_path = r"P:\DOCUMENTS\PROJECTS-1\OIL_PRODUCTION.csv"
df = pd.read_csv(file_path)
print(df.head())
print(df.info())
print(df.describe())
df.fillna(0, inplace=True)
global_production = df.groupby('Year')['Oil production (TWh)'].sum().reset_index()
global_production['Growth Rate'] = global_production['Oil production (TWh)'].pct_change() * 100
latest_year = df['Year'].max()
top_producers = df[df['Year'] == latest_year].nlargest(10, 'Oil production (TWh)')
conn = sqlite3.connect('oil_production.db')
df.to_sql('oil_production', conn, if_exists='replace', index=False)
global_production.to_sql('global_production', conn, if_exists='replace', index=False)
```

```
top_producers.to_sql('top_producers', conn, if_exists='replace', index=False)
cursor = conn.cursor()
cursor.execute("CREATE VIEW IF NOT EXISTS country_trends AS SELECT Entity, Code, Year, [Oil production (TWh)] FROM
oil_production WHERE Code != ''")
cursor.execute("CREATE VIEW IF NOT EXISTS regional_production AS SELECT Entity, Year, [Oil production (TWh)] FROM
oil_production WHERE Code = ''")
conn.commit()
conn.close()
```

Step 2: Export SQLite Tables to CSV

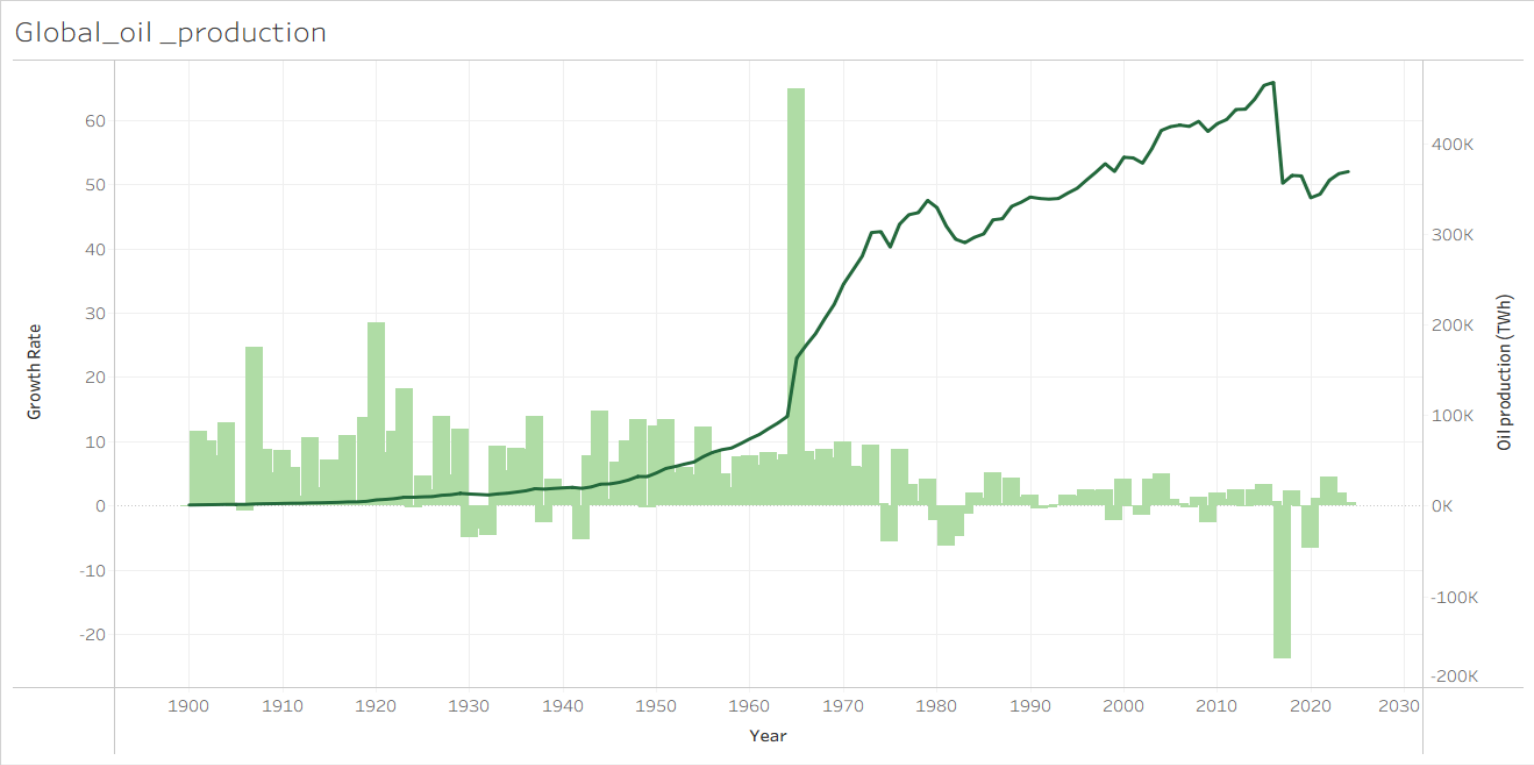
File: oil_production_database_file.py

```
import sqlite3
import pandas as pd
from pathlib import Path
db_path = 'oil_production.db'
conn = sqlite3.connect(db_path)
query = "SELECT name FROM sqlite_master WHERE type='table';"
tables = pd.read_sql(query, conn)['name'].tolist()
output_folder = Path('csv_exports')
output_folder.mkdir(exist_ok=True)
for table in tables:
    df = pd.read_sql(f'SELECT * FROM {table}', conn)
    csv_path = output_folder / f'{table}.csv'
    df.to_csv(csv_path, index=False)
    print(f"Exported {table} to {csv_path}")
conn.close()
```

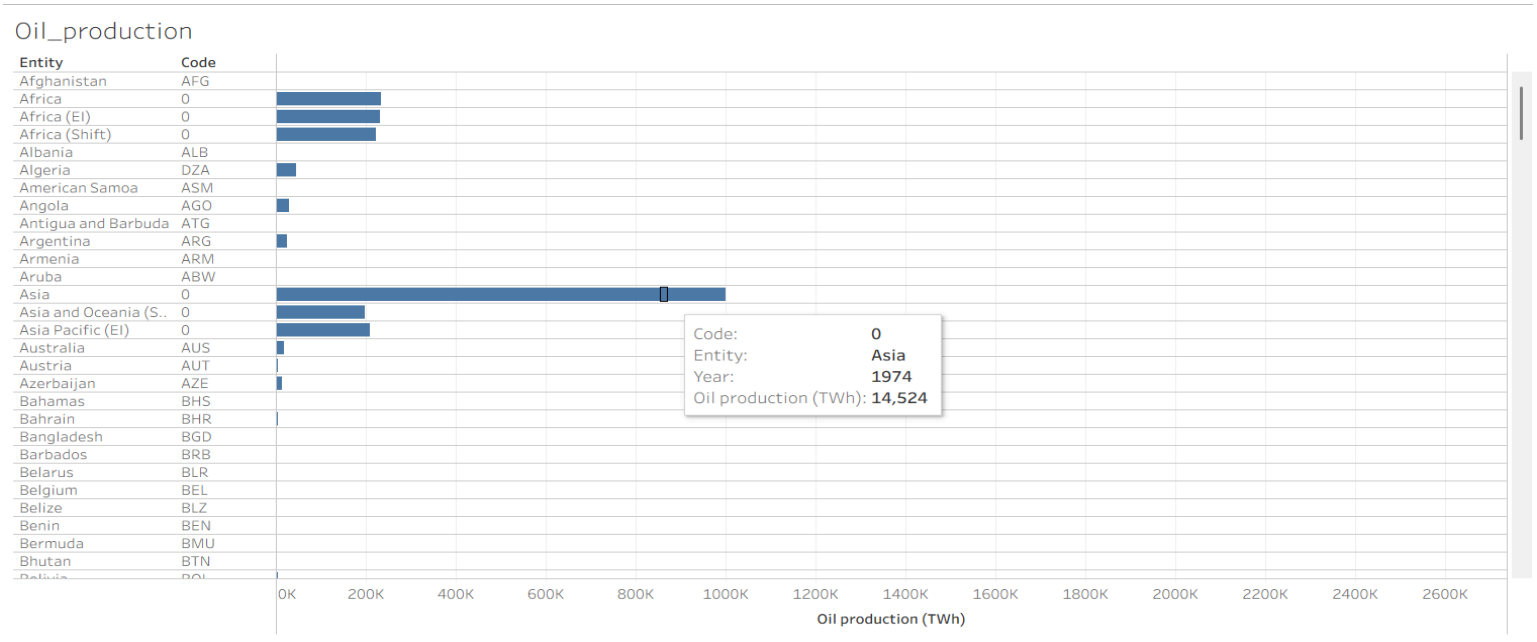
Step 3: Visualize in Tableau

1. Open Tableau Desktop / Public
2. Go to Microsoft excel
3. Select 'oil_production.csv'
4. Create charts

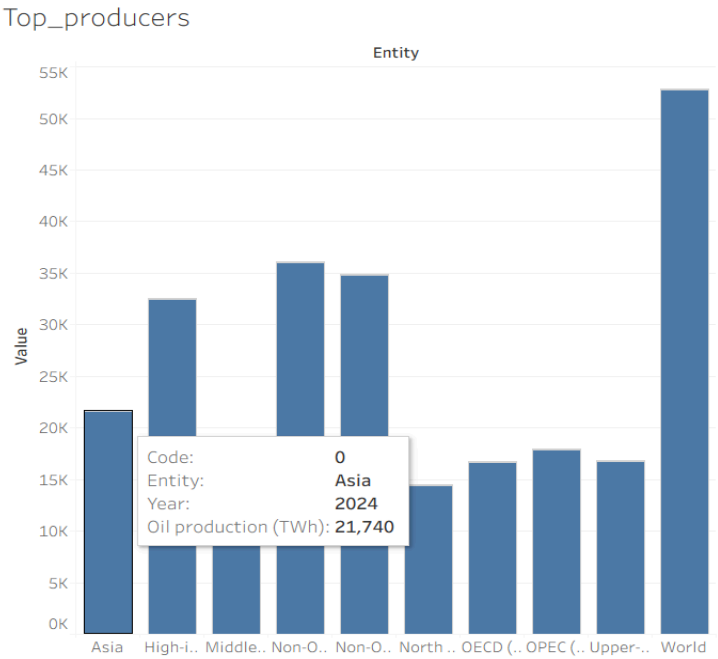
Global Oil_production chart:



Oil_production chart:



Top_Producers chart:



Step 4: Upload Project to GitHub

cd path_to_project_folder
git init
git add .
git commit -m "Initial commit - Oil Production Analysis"
git remote add origin https://github.com/your-username/oil-production-analysis.git
git push -u origin master

Key Insights and Recommendations

- Analyze long-term oil production trends
- Identify top producing countries
- Compare regional vs national output
- Keep 'csv_exports' for Tableau
- Add README.md in GitHub with overview and visuals