Canada Oil Production Analysis

Group 3

- Bo Li (30212597)
- Israa Farouk (30076315)
- Jackie Yi (30220925)
- Junyi Jiang (30052672)

Introduction & Motivation

What the project is and why picked it

In the heart of Alberta and indeed across Canada, crude oil stands as not just a valuable resource but also as the cornerstone of the nation's economic stability. Its pivotal role in Canada's Gross Domestic Product (GDP) cannot be overstated. This research project aims to dissect the importance of crude oil to Canada within the global context. First and foremost, we seek to assess Canada's position in the international crude oil market by comparing its significance to that of its global competitors. By delving into extensive datasets, meticulously collected and curated, we intend to shine light on some of the factors that underscore Canada's reliance on this resource. This includes examining Canada's contribution to global oil production. What factors have impacts on crude oil prices. What does Canada's oil production mean for Canada interms of GDP, exports and investment and among other essential metrics.

There are four guiding questions will help us structure the report with dynamic flow and also how we should explore the data.

Guiding question :

1. What’s Canada’s contribution to Global oil production?

Canada has consistently been a prominent contributor to global oil production, primarily through its vast reserves of oil sands. Understanding the magnitude of its contribution compared to other global players is crucial in assessing its significance on the world stage. This question aims to ascertain the position of Canada in the global oil production landscape, helping us comprehend the country's role in meeting global energy demands and the economic implications of its production levels.

2. What factors impact oil prices?

Oil prices are subject to volatile fluctuations influenced by various external and internal factors. Identifying these factors is essential to decipher the dynamics of the oil market. By analyzing the fluctuations in West Texas Intermediate (WTI) crude oil prices, this question seeks to uncover the events, both international and domestic, as well as structural and technological changes that drive these price fluctuations. Understanding these influences can aid in forecasting and decision-making for stakeholders in the energy sector.

3. What does Canada's oil production mean for Canada?

Oil production plays a vital role in Canada's economy. It is crucial to evaluate the relative importance of this industry within the broader Canadian economic landscape. This question examines the economic impact of Canada's oil production. It seeks to uncover the extent to which oil production drives exports and investments, providing insights into the country's economic resilience.

In addressing these questions, we aim to provide a comprehensive perspective on Canada's oil production and its multifaceted implications, enabling stakeholders and decision-makers to navigate the complex landscape of energy production with greater clarity and foresight.

Summary of the Outcomes

- Project and data exploration :

Dataset 1:

1. Canada's overall oil production from 2000 to 2021 is ranked 6th globally.
2. Canada's oil production has shown a steady increase over the years.

Dataset 2:

After exploring each guiding question and dataset, we noticed that the Canada oil production and the industry as a whole plays a vital role in Canada's economy.

There are three main kinds affairs that can have impacts on oil price:

1. Oil production technology take into place examples like SAGD in 2003 and Shale Oil Boom in the USA in 2011.
2. political affairs happen in oil production regions examples like Ukrain war in 2022
3. OPEC production related annoucements example like OPEC+ Agrees To Cut 9.7 million b/d (2020)

Dataset 3:

1. The export of crude oil has became more and more important to the Candian yearly total export. From 4,68% of the yearly total export in 2000 to 12.21% of the yearly total export in 2020, curde oil has became the top export product.

Dataset 4:

1. Canada’s investment in Oil and Gas Industries have been steadily declining
2. 2020 was the lowest investment year and 2014 was the highest investment year
3. Historic events affected Canada’s investment

In summary, our analysis reveals the undeniable significance of Canada's oil industry in ensuring the stability of the country's economy. The data we have collected strongly underscores its vital role in supporting various economic sectors and contributing to Canada's overall fiscal health. However, it is important to note that despite its pivotal role, government investments and support for the oil industry lag behind those of other industries. In light of this, we firmly believe that it is imperative to increase investment in the Canadian oil sector. This move not only enhances the competitiveness of the industry but also aligns with Canada's commitment to the "Paris Climate Agreement." Increasing investments in the oil industry is, without a doubt, the most prudent and logical choice for Canada to navigate the complex landscape of economic stability and environmental responsibility.

Elevating investment in Canada's oil industry not only secures the nation's economic foundation but also represents a strategic step toward achieving the goals set forth in the "Paris Climate Agreement." By bolstering the competitiveness of this industry, Canada can better position itself to provide essential resources both domestically and on the international stage. Simultaneously, it can actively work towards reducing its environmental impact and transitioning to a more sustainable energy landscape. This dual-purpose investment strategy is undoubtedly a "no brainer" choice, as it balances economic stability with environmental responsibility, reinforcing Canada's commitment to both aspects for a prosperous and sustainable future.

- What we learned from this project:

- The findings have the potential to diverge from the initial expectations. This is an integral part of research as it involves identifying and addressing any discrepancies or challenges.
- For projects with sevral datasets, the most crucial part is the storyline. you have to make sure first you have a very clean and clear storyline for telling the story, then you have to make sure all your supporting data and reference will support your story at the right moment.

- Data Quality is important to asses before creating guiding questions and creating your data story, as the realiability of the data may shape the direction of the data story

Datasets

- Dataset 1 : Canada’s contribution to Global oil production

Answering Guiding Question 1 : What’s Canada’s contribution to Global oil production?

The source and licensing of the dataset : https://www.kaggle.com/datasets/alexandrepetit881234/oil-production-1900-2020 [1]

The data is sourced from Kaggle.com under the CC0: Public Domain License. The data is formatted in a .csv file.This dataset contains information about oil production by country from 1900 - 2020 in TWh. (Terawatt hour) The data consists of 4 columns (Entity, Code, Year, Oil production (TWh)) and 17328 rows.

Noteworthy challenges in working with the dataset :

1. In this dataset, the "Entity" column contains information about continents. However, since our research focuses on individual countries, I have removed the rows where the "Entity" column is empty (as only countries have country codes).
2. Additionally, our research specifically targets oil production from 2000 to 2021, so I have also removed any countries with missing values in this time range.

- Dataset 2 : Events impact on price

Answering Guiding Question 2 : What factors impact oil prices?

The source and licensing of the dataset : https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm [4] The other domain reference we used is https://www.eia.gov/finance/markets/crudeoil/spot_prices.php. [9] This report describes the seven key factors, from supply to demand, that could influence oil markets and explore possible links between each factor and crude oil prices. It includes regularly updated graphs that depict aspects of those relationships. The seven factors are Spot Prices , Supply Non-OPEC , Supply OPEC , Balance, Financial Markets, Demand Non-OECD, Demand OECD. All factors are chained together. An example will be oil price increase can result in oil production increase. But oil production increase can also drive oil prices down. Due to the complexity of the topic and also the limits of this project, we will only look at the Spot prices - Crude oil prices react to a variety of geopolitical and economic events. This report uses line graphs with pointers to show from 1970 till recent, how price fluctuated with the every specific geopolitical and economic events and result of it. This section of the source significantly bolsters our ability to address the second guiding question: "What factors impact oil prices?" by shedding light on the multifaceted interplay between international events and the oil market's price dynamics.

The structure and format of the dataset : This dataset is downloaed as a CSV file. it has 2 column with 9516 rows. the data inclues spot oil daily price from 1986 til 2023. all cells value are strings.

Noteworthy challenges in working with the dataset :

1. For other people's conveniece, I downloaded the CSV then uploaded to
2. github as public so people will also have access to the date. But later I noticed that the format is changed with all values merged into one column. and I took quite some time to figure how to do the data wrangling and cleaning.
3. For add the annotation on the line graph, I also noticed that the value of axies is not exactly as it shows as the number of year. I did lots of testing to find the percise number that representing the year I want.

- Dataset 3 : International merchandise trade by commodity, monthly (x 1,000,000)

Answering Guiding Question 3 : What does Canada's oil production mean for Canada?

The source and licensing of the dataset : https://doi.org/10.25318/1210012101-ENG [2]

The dataset include the monthly international merchandise trade by commodity from 2000 to 2020, which contains 9324 columns and 3 rows (Category, Date,Value), it was downloaded as CSV format.|

No noteworthy challenges in working with this dataset.

- Dataset 4 : Canada’s Investment by Industry

Answering Guiding Question 3 : What does Canada's oil production mean for Canada?

The source and licensing of the dataset : https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610009601&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=2.2&cubeTimeFrame.startYear=2000&cubeTimeFrame.endYear=2020&referencePeriods=20000101%2C20200101 [3]

This dataset is downloaded as a CSV file. It consists of all Canada's industry investment from the years of 2000 - 2020. It has 23 columns (Industry, Assets and the investment amount in CAD by year from the selected time period) and 297 rows. The columns of interest are the Year, Industry column and the investment amount column in CAD

There were no noteworthy challenges in working with this dataset.

Data Exploration

Dataset 1

Here is a step-by-step guide on how to clean and wrangle the dataset and visualize the data:

- Start by dropping the entities that do not have a country code, as we are specifically interested in analyzing countries.
- Filter the dataset to only include data from the years 2000 to 2021, as this is the time range of interest.
- Remove any countries that have missing values for oil production during the years 2000-2021.
- Calculate the cumulative oil production for each country during the time period and sort them in descending order.
- Reflect the cumulative oil production values on a geographic map to visualize the distribution of oil production by country.
- Create a plot showcasing the oil production distribution of the top 10 countries by year, which will provide insight into each country's contribution over time.

By following these steps, you will have a clean and manageable dataset for analysis and visualizations that explore the oil production trends of different countries.

```
In [ ]: # To export our project as html
        import plotly.io as pio
        pio.renderers.default = 'notebook'
```

```
In [ ]: import pandas as pd
        country=pd.read_csv('oil-production-by-country.csv')
```

In [ ]:

```
countrya=country.dropna()#drop the entity without country code,filter all country
countryb=countrya[~countrya['Entity'].str.contains('World')]
display(countryb)
```

	Entity	Code	Year	Oil production (TWh)
0	Afghanistan	AFG	1980	0.0
1	Afghanistan	AFG	1981	0.0
2	Afghanistan	AFG	1982	0.0
3	Afghanistan	AFG	1983	0.0
4	Afghanistan	AFG	1984	0.0
...	...	...	...	...
17322	Zimbabwe	ZWE	2012	0.0
17323	Zimbabwe	ZWE	2013	0.0
17324	Zimbabwe	ZWE	2014	0.0
17325	Zimbabwe	ZWE	2015	0.0
17326	Zimbabwe	ZWE	2016	0.0

13838 rows × 4 columns

In [ ]:

```
aimyear=list(range(2000,2022))
country4=countryb[countryb["Year"].isin(aimyear)]#filter the data during 2000-2021
display(country4)
```

	Entity	Code	Year	Oil production (TWh)
20	Afghanistan	AFG	2000	0.0
21	Afghanistan	AFG	2001	0.0
22	Afghanistan	AFG	2002	0.0
23	Afghanistan	AFG	2003	0.0
24	Afghanistan	AFG	2004	0.0
...	...	...	...	...
17322	Zimbabwe	ZWE	2012	0.0
17323	Zimbabwe	ZWE	2013	0.0
17324	Zimbabwe	ZWE	2014	0.0
17325	Zimbabwe	ZWE	2015	0.0
17326	Zimbabwe	ZWE	2016	0.0

3905 rows × 4 columns

In [ ]:

```
country5 = country4.groupby('Entity').filter(lambda x: len(x) == 22)#focus on the countries
#that have values of oil production for 2000-2021
display(country5)
```

	Entity	Code	Year	Oil production (TWh)
550	Algeria	DZA	2000	776.732800
551	Algeria	DZA	2001	764.684140
552	Algeria	DZA	2002	824.497200
553	Algeria	DZA	2003	918.828100
554	Algeria	DZA	2004	971.698100
...	...	...	...	...
17156	Yemen	YEM	2017	34.364900
17157	Yemen	YEM	2018	47.119152
17158	Yemen	YEM	2019	47.393547
17159	Yemen	YEM	2020	43.814594
17160	Yemen	YEM	2021	32.096210

1056 rows × 4 columns

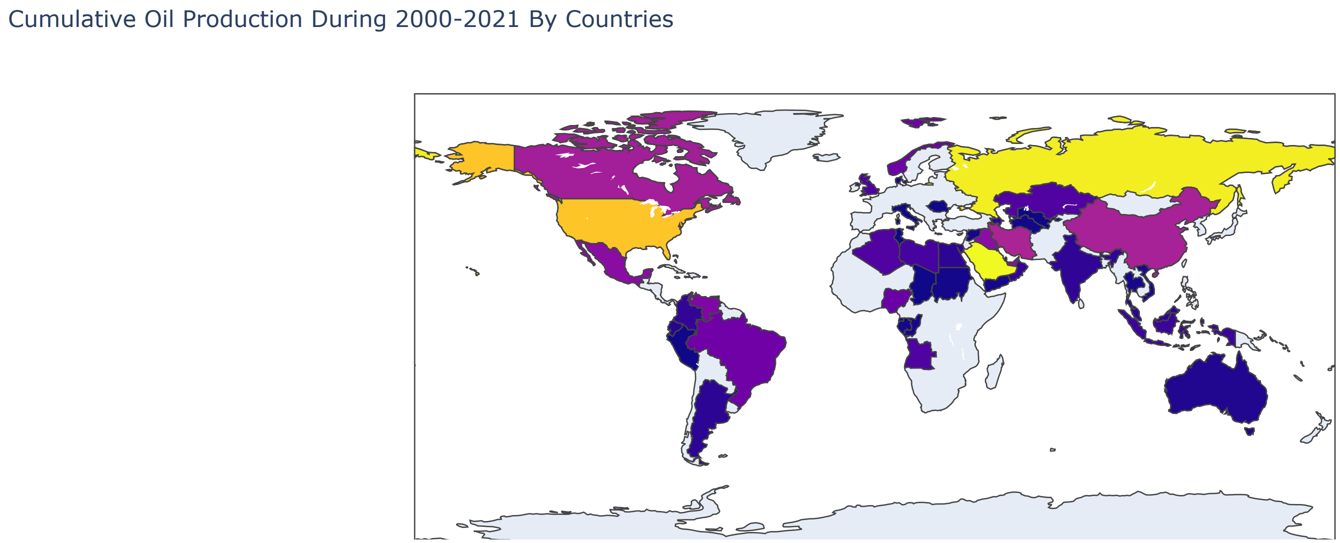
```
In [ ]: #calculate the cumulative oil production during 2000-2021 by countries
        xygroup = country5['Oil production (TWh)'].groupby(country5['Entity'])
        mapcountry=pd.DataFrame(xygroup.sum().sort_values(ascending=False))
        display(mapcountry)
```

	Oil production (TWh)
Entity	
Saudi Arabia	130169.472400
Russia	126731.098100
United States	114076.510400
Iran	49909.931000
China	48676.518300
Canada	47177.302500
United Arab Emirates	38353.586800
Iraq	38064.844300
Mexico	37312.835900
Kuwait	33995.016900
Venezuela	33969.915790
Norway	28539.850470
Brazil	28516.216560
Nigeria	27069.312500
Algeria	18386.607260
Kazakhstan	18331.840330
Angola	18029.974330
United Kingdom	17562.997800
Qatar	16381.602030
Libya	15212.615420
Indonesia	12338.607440
Oman	11142.103100
Colombia	10079.376310
India	9738.166310
Azerbaijan	8916.247570
Egypt	8796.330300
Argentina	8790.302880
Malaysia	8185.857020
Ecuador	6814.381870
Australia	5849.947620
Vietnam	3995.324630
Thailand	3603.589106
Syria	3563.475039
Congo	3459.937090
Equatorial Guinea	3165.185009

	Oil production (TWh)
Entity	
Yemen	3006.924007
Gabon	2986.118534
Denmark	2975.041936
Sudan	2793.623122
Turkmenistan	2781.337810
Brunei	2026.009072
Peru	1514.665381
Trinidad and Tobago	1484.113940
Chad	1364.862030
Italy	1307.776237
Romania	1170.245490
Uzbekistan	1137.535012
Tunisia	790.946527

In [ ]:

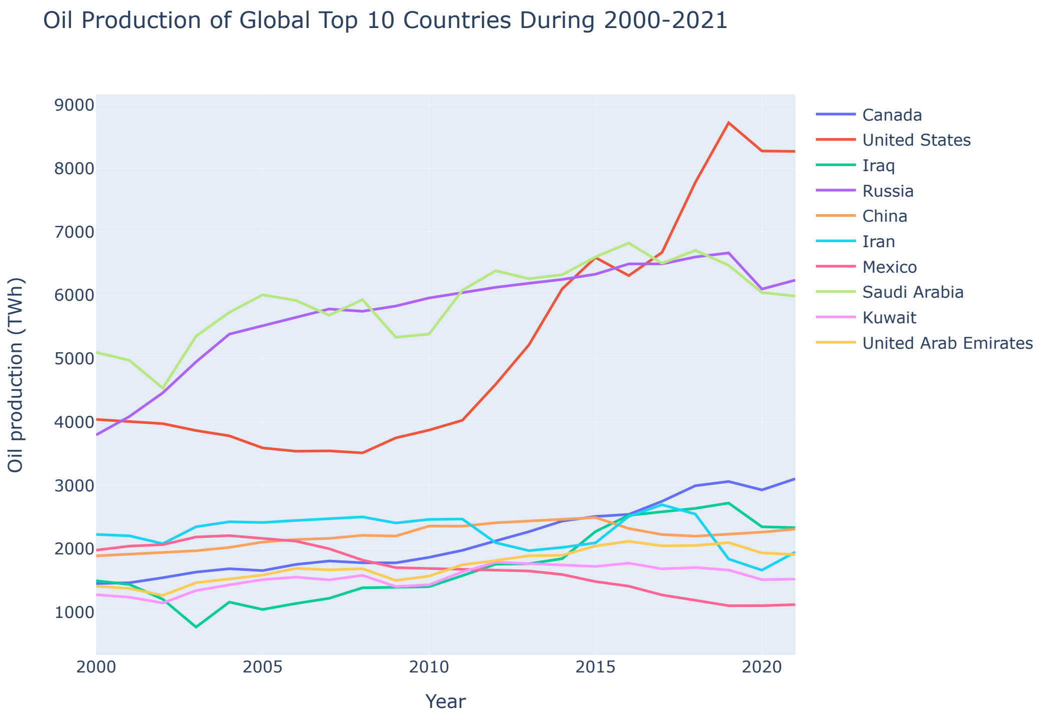
```
#using plotly package for this visualiztion [7]
import plotly.express as px

fig = px.choropleth(mapcountry, locations=mapcountry.index,
                    color="Oil production (TWh)",
                   hover_name=mapcountry.index,locationmode="country names",
                    title="Cumulative Oil Production During 2000-2021 By Countries"
                   )
fig.show()
```



Cumulative Oil Production During 2000-2021 By Countries

In [ ]:

```
import plotly.graph_objects as go
fig=go.Figure()
fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="CAN"]['Year'],
        y=country5[country5['Code']=="CAN"]['Oil production (TWh)'],
        mode="lines",
        name="Canada"

    )
)
fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="USA"]['Year'],
        y=country5[country5['Code']=="USA"]['Oil production (TWh)'],
        mode="lines",
        name="United States"
    )
)
fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="IRQ"]['Year'],
        y=country5[country5['Code']=="IRQ"]['Oil production (TWh)'],
        mode="lines",
        name="Iraq"

    )
)
fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="RUS"]['Year'],
        y=country5[country5['Code']=="RUS"]['Oil production (TWh)'],
        mode="lines",
        name="Russia"

    )
)
fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="CHN"]['Year'],
        y=country5[country5['Code']=="CHN"]['Oil production (TWh)'],
        mode="lines",
        name="China"


    )
)

fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="IRN"]['Year'],
        y=country5[country5['Code']=="IRN"]['Oil production (TWh)'],
        mode="lines",
        name="Iran"


    )
)

fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="MEX"]['Year'],
        y=country5[country5['Code']=="MEX"]['Oil production (TWh)'],
        mode="lines",
        name="Mexico"


    )
)

fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="SAU"]['Year'],
        y=country5[country5['Code']=="SAU"]['Oil production (TWh)'],
        mode="lines",
        name="Saudi Arabia"


    )
)

fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="KWT"]['Year'],
        y=country5[country5['Code']=="KWT"]['Oil production (TWh)'],
        mode="lines",
        name="Kuwait"


    )
)

fig.add_trace(
    go.Scatter(
        x=country5[country5['Code']=="ARE"]['Year'],
        y=country5[country5['Code']=="ARE"]['Oil production (TWh)'],
        mode="lines",
        name="United Arab Emirates"


    )
)


fig.update_layout(
    title="Oil Production of Global Top 10 Countries During 2000-2021",
    xaxis=dict(title="Year"),
    yaxis=dict(title="Oil production (TWh)"),
    width=800,
    height=600

)
fig.show()
```



Dataset 2

- Data import: Download the data from reference link: https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm, the uploaded it github as public file (convenice for readers to check)
- cleaning and wrangling of the dataset : Noticing the first 4 rows are irrlavant, I used the skiprows function to slice them out. Because of the values of all column are "str", I used the pd.to_datetime to chage the value of column[0] to date format="%m/%d/%Y", so it can used for later for visualization. I also used the below_zero_prices = df[df["price"] < 0] to make sure all price value are

greater then "0".For the date as 2020-04-20,the WTI spot price shows as negative. Then I refer back to the reference page and noticed that they also excluded that negative number in their visualization. So I used the df = df[df["Day"] != "2020-04-20"] to created a new data without that row.

- Visualizatin: Line graph with annotation for specific events that has impact on the spot oil prices.

In [ ]:

```
#using numpy package for this visualiztion [5]
import numpy as np
#using Pandas package for this visualiztion [6]
import pandas as pd
#using matplotlib package for this visualiztion [8]
import matplotlib.pyplot as plt
import datetime
# Load the data from the CSV file but I also upload the file to github for the convenice of the reader
url = "https://raw.githubusercontent.com/boli3ucalgary/602/main/Cushing_OK_WTI_Spot_Price_FOB.csv"
df = pd.read_csv(url, skiprows=4) #with unexpected input, for python abel to read, we have to skip the first 4 rows
display(df.head())
display(df.tail())
display(type(df["Day"][1]))
df["Day"] = pd.to_datetime(df["Day"], format="%m/%d/%Y")
df.rename(columns={df.columns[1]: "price"}, inplace=True)

df['Year'] = df['Day'].dt.strftime('%Y')

display(df.head())
display(df.tail())
#now we need to check the value make sure all values are correct
below_zero_prices = df[df["price"] < 0]
print(below_zero_prices)
#after printing we noticed there is one value for 2020-04-20
df = df[df["Day"] != "2020-04-20"]

below_zero_prices = df[df["price"] < 0]
print(below_zero_prices)
```

	Day	Cushing OK WTI Spot Price FOB Dollars per Barrel
0	10/2/2023	88.81
1	09/29/2023	90.77
2	09/28/2023	91.65
3	09/27/2023	93.67
4	09/26/2023	91.43

	Day	Cushing OK WTI Spot Price FOB Dollars per Barrel
9506	01/8/1986	25.87
9507	01/7/1986	25.85
9508	01/6/1986	26.53
9509	01/3/1986	26.00
9510	01/2/1986	25.56

```
str
```

	Day	price	Year
0	2023-10-02	88.81	2023
1	2023-09-29	90.77	2023
2	2023-09-28	91.65	2023
3	2023-09-27	93.67	2023
4	2023-09-26	91.43	2023

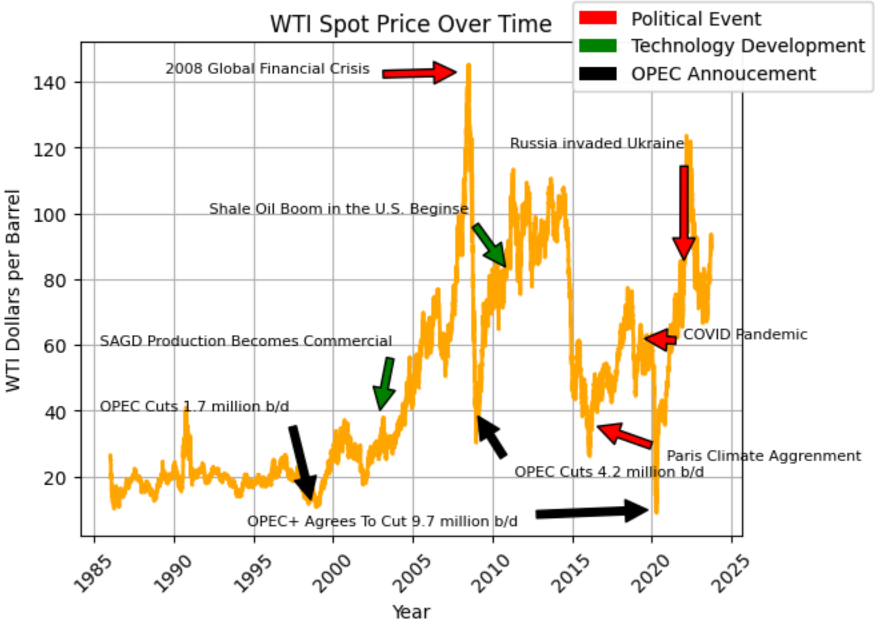
	Day	price	Year
9506	1986-01-08	25.87	1986
9507	1986-01-07	25.85	1986
9508	1986-01-06	26.53	1986
9509	1986-01-03	26.00	1986
9510	1986-01-02	25.56	1986

```
           Day  price  Year
867 2020-04-20 -36.98  2020
Empty DataFrame
Columns: [Day, price, Year]
Index: []
```

In [ ]:

```python
# Create a line graph with "Day" on the x-axis and "price" on the y-axis
plt.figure(figsize=(10, 6))  # Optional: Set the figure size

# Create a figure and axis
fig, ax = plt.subplots()

# Assuming the column names are as mentioned
x = df["Day"]
y = df["price"]

# Create color bars for the legend
red_bar = plt.Rectangle((0, 0), 0.1, 0.1, fc="red")
green_bar = plt.Rectangle((0, 0), 0.1, 0.1, fc="green")
black_bar = plt.Rectangle((0, 0), 0.1, 0.1, fc="black")

# Create the legend with labels
legend = ax.legend([red_bar, green_bar, black_bar], ['Political Event', 'Technology Development', 'OPEC Annoucement'],loc='upper left', bbox_to_anchor=(0.73, 1.1))

# Add the legend to the plot
ax.add_artist(legend)

plt.plot(x, y, marker='o', linestyle='-', markersize=0.1, linewidth=2, color='orange', solid_capstyle='round')
plt.xlabel("Year")  # Set the x-axis label
plt.ylabel("WTI Dollars per Barrel")  # Set the y-axis label
plt.title("WTI Spot Price Over Time")
plt.grid(True)

#add annotate for events that has impact on oil price
# black is oil related, green is technology related, red is political related
plt.annotate('OPEC+ Agrees To Cut 9.7 million b/d', xy=(18500, 10), xytext=(9000, 5),arrowprops=dict(facecolor='black', shrink=0.1),fontsize=8)
plt.annotate('COVID Pandemic', xy=(18000, 62), xytext=(19000, 62),arrowprops=dict(facecolor='red', shrink=0.1),fontsize=8)
plt.annotate('Russia invaded Ukraine', xy=(19000, 82), xytext=(15000, 120),arrowprops=dict(facecolor='red', shrink=0.1),fontsize=8)
plt.annotate(' Shale Oil Boom in the U.S. Beginse', xy=(15000, 82), xytext=(8000, 100),arrowprops=dict(facecolor='green', shrink=0.1),fontsize=8)
plt.annotate(' 2008 Global Financial Crisis', xy=(14000, 143), xytext=(7000, 143),arrowprops=dict(facecolor='red', shrink=0.1),fontsize=8)
plt.annotate(' SAGD Production Becomes Commercial', xy=(12000, 38), xytext=(5500, 60),arrowprops=dict(facecolor='green', shrink=0.1),fontsize=8)
plt.annotate(' Paris Climate Aggrenment', xy=(16850, 36), xytext=(18500, 25),arrowprops=dict(facecolor='red', shrink=0.1),fontsize=8)
plt.annotate(' OPEC Cuts 1.7 million b/d', xy=(10500, 10), xytext=(5500, 40),arrowprops=dict(facecolor='black', shrink=0.1),fontsize=8)
plt.annotate(' OPEC Cuts 4.2 million b/d', xy=(14200, 40), xytext=(15000, 20),arrowprops=dict(facecolor='black', shrink=0.1),fontsize=8)



# Optional: Rotate x-axis labels for better readability
plt.xticks(rotation=45)
# Show the plot
plt.show()
```

```
<Figure size 1000x600 with 0 Axes>
```



Dataset 3

Task:

- Dataset import
- Data manipulation
  - Extract the year in an appropriate format using regular expression
  - Convert string formatted value as numeric using replace(), strip(), to_numeric() functions
- Create grouped Pandas data frame using groupby() function
- Create data visualizations using interactive data visulization package: Plotly

Preface: In order to display the pie charts appropreatiately, we labelled the categories which export values out of TOP 5 as all other products, the specific categories as follow:

'Tires; motor vehicle engines and motor vehicle parts', 'Industrial machinery, equipment and parts', 'Pulp and paper', 'Special transactions trade', 'Natural gas, natural gas liquids and related products', 'Intermediate metal products', 'Other electronic and electrical machinery, equipment and parts', 'Communication, and audio and video equipment', 'Food, beverage and tobacco products', 'Farm and fishing products', 'Aircraft, aircraft engines and aircraft parts', 'Basic chemicals and industrial chemical products', 'Plastic and rubber products', 'Furniture and fixtures', 'Cleaning products, appliances, and miscellaneous goods and supplies', 'Clothing, footwear and textile products', 'Medium and heavy trucks, buses, and other motor vehicles', 'Refined petroleum energy products', 'Paper and published products', 'Fabricated metal products', 'Electricity', 'Non-metallic minerals', 'Computers and computer peripherals', 'Non-metallic mineral products', 'Metal ores and concentrates', 'Other transportation equipment and parts', 'Pharmaceutical and medicinal products', 'Coal', 'Intermediate food products', 'Waste and scrap of metal and glass', 'Nuclear fuel and other energy products', 'Logs, pulpwood and other forestry products', 'Waste and scrap of wood, wood by-products, paper and paperboard', 'Waste and scrap of plastic and rubber'

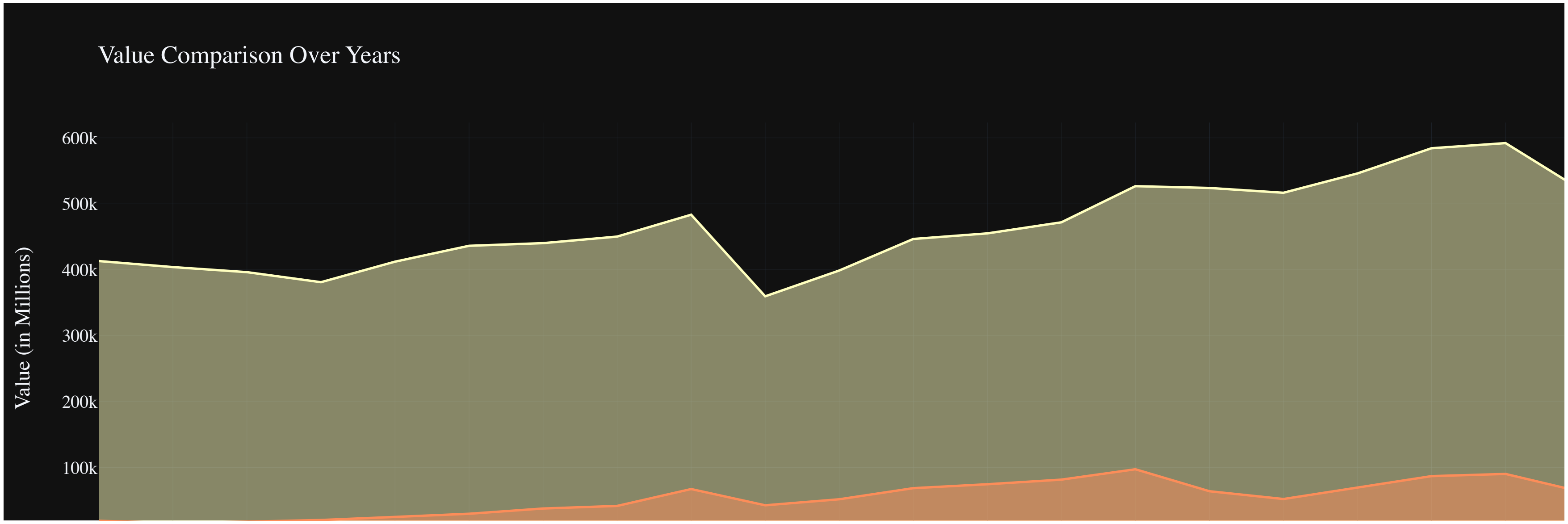
In [ ]:

```
#Dataset Import
df00_20 = pd.read_csv("exportnew.csv")
dfoil_00_20 =  pd.read_csv("oilexport.csv")
df2 = pd.read_csv("1210012101-eng (5).csv")
df3 = pd.read_csv("1210012101-eng (6).csv")


#Data Manipulation
pattern = r'\d{2}-(\d{2}|\w{3})'
df00_20['Year'] = df00_20['Date'].str.extract(r'(\d{2}|\d{2}-\w{3})')
df00_20['Year'] = "20" + df00_20['Year']
df00_20["Value"] = pd.to_numeric(df00_20["Value"].str.replace(",",""))
df00_20_group = df00_20.groupby("Year")["Value"].sum().reset_index()

dfoil_00_20['Year'] = dfoil_00_20['Date'].str.extract(r'(\d{2}|\d{2}-\w{3})')
dfoil_00_20['Year'] = "20" + dfoil_00_20['Year']
dfoil_00_20["Value"] = pd.to_numeric(dfoil_00_20["Value"].str.replace(",",""))
dfoil_00_20_group = dfoil_00_20.groupby("Year")["Value"].sum().reset_index()


df2['Year'] = df2['Date'].str.extract(r'(\d{2}|\d{2}-\w{3})')
df2['Year'] = "20" + df2['Year']

for i in df2.columns[1:-1]:
    df2[i] = pd.to_numeric(df2[i].str.replace(",",""))

df2["Trade_Balance"] = df2["Export_Customs_Unadjusted"] - df2["Import_Custom_Unadjusted"]

df2_grouped = df2.groupby("Year")[df2.columns.difference(["Date","Year"])].sum().reset_index()
df2_grouped[df2_grouped.columns.difference(['Year'])] = df2_grouped[df2_grouped.columns.difference(['Year'])]*100000

df3.dropna(inplace=True)
df3['Year'] = df3['Date'].str.extract(r'(\d{2}|\d{2}-\w{3})')
df3['Year'] = "20" + df3['Year']
df3["Value"] = pd.to_numeric(df3["Value"].str.replace(",",""))
df3['Category'] = df3['Category'].str.replace(r'\[.*\]', '', regex=True).str.strip()

df3_oil = df3[df3["Category"] == "Crude oil and crude bitumen"].groupby('Year')['Value'].sum()
oil_yearly_percentage = df3_oil / df3.groupby('Year')['Value'].sum()
oil_yearly_percentage

other_percentage = 1- df3_oil / df3.groupby('Year')['Value'].sum()

percentage_df = pd.DataFrame({'Oil Yearly Percentage': round(oil_yearly_percentage*100,2), 'Other Percentage': round(other_percentage*100,2)}).reset_index()

#Filter the data in 2000
df3_2000 = df3[df3["Year"]=="2000"]
df3_2000_group = df3_2000.groupby('Category')['Value'].sum().reset_index().sort_values(by='Value', ascending=False)

top_categories = df3_2000_group.head(2)
# Create a list of the top 3 categories and 'Crude oil and crude bitumen'
top_category_list = top_categories['Category'].tolist() + ['Crude oil and crude bitumen']

# Replace categories not in the top list with 'Other'
df3_2000_group['Category'] = df3_2000_group['Category'].apply(lambda x: x if x in top_category_list else 'All Other Products')


df3_2020 = df3[df3["Year"]=="2020"]
```

```
df3_2020_group = df3_2020.groupby('Category')['Value'].sum().reset_index().sort_values(by='Value', ascending=False)

top_categories = df3_2020_group.head(3)
# Create a list of the top 3 categories and 'Crude oil and crude bitumen'
top_category_list = top_categories['Category'].tolist() + ['Crude oil and crude bitumen']

# Replace categories not in the top list with 'Other'
df3_2020_group['Category'] = df3_2020_group['Category'].apply(lambda x: x if x in top_category_list else 'All Other Products')
```

In [ ]:

```
fig = go.Figure()

fig.add_trace(go.Scatter(x=df00_20_group['Year'], y=df00_20_group['Value'], fill='tozeroy', mode='lines', name='Total export 2000 - 2020',marker_color = "#ffffbf"))
fig.add_trace(go.Scatter(x=dfoil_00_20_group['Year'], y=dfoil_00_20_group['Value'], fill='tozeroy', mode='lines', name='Crude Oil export 2000 - 2020',marker_color = "#f

fig.layout.template = "plotly_dark"
fig.update_layout(
    title='Value Comparison Over Years',
    xaxis=dict(title='Year'),
    yaxis=dict(title='Value (in Millions)'),
    showlegend=True,
    font = dict(size= 15, family="Franklin Gothic")
)

fig.show()
```

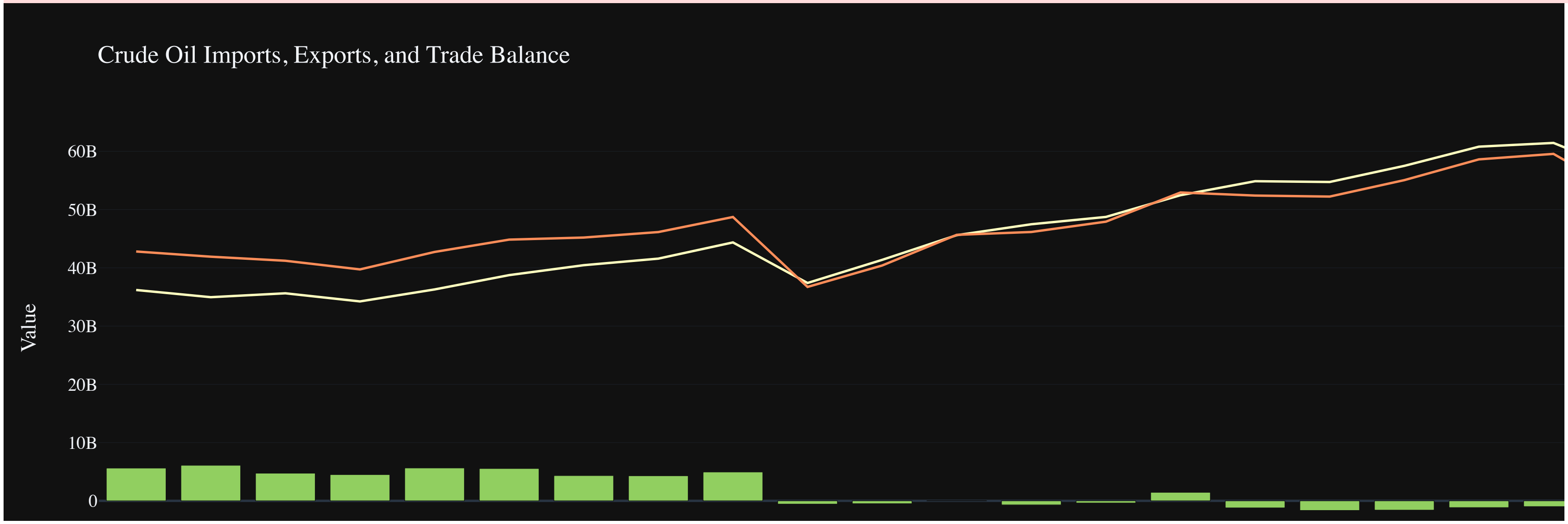


In [ ]:

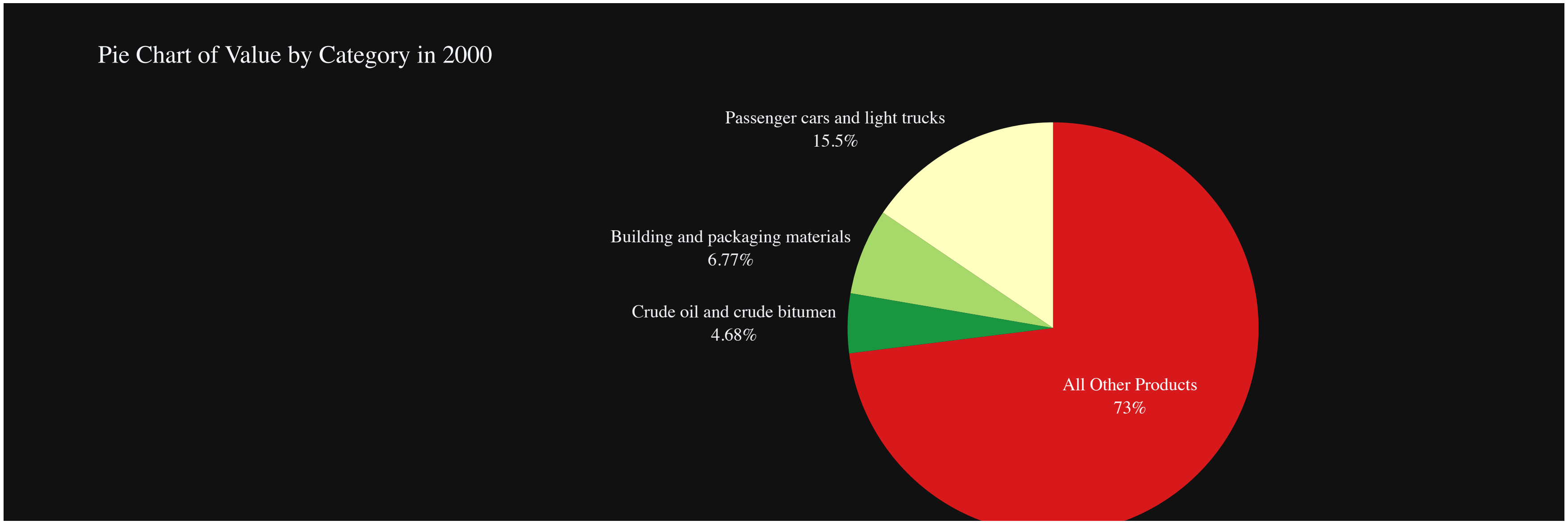
```
fig2 = go.Figure()

fig2.add_trace(go.Line(x=df2_grouped['Year'], y=df2_grouped['Import_Balance_Unadjusted'], mode='lines', name='Crude Oil Import',marker_color = "#ffffbf"))
fig2.add_trace(go.Line(x=df2_grouped['Year'], y=df2_grouped['Export_Balance_Unadjusted'], mode='lines', name='Crude Oil Export',marker_color = "#fc8d59"))
fig2.add_trace(go.Bar(x=df2_grouped['Year'],y=df2_grouped["Trade_Balance"],name="Trade Balance",marker_color="#91cf60"))
```

```
fig2.layout.template = "plotly_dark"
fig2.update_layout(
    title='Crude Oil Imports, Exports, and Trade Balance',
    yaxis=dict(title='Value'),
    showlegend=True,
    font = dict(size= 15, family="Franklin Gothic")
)

fig2.show()
```

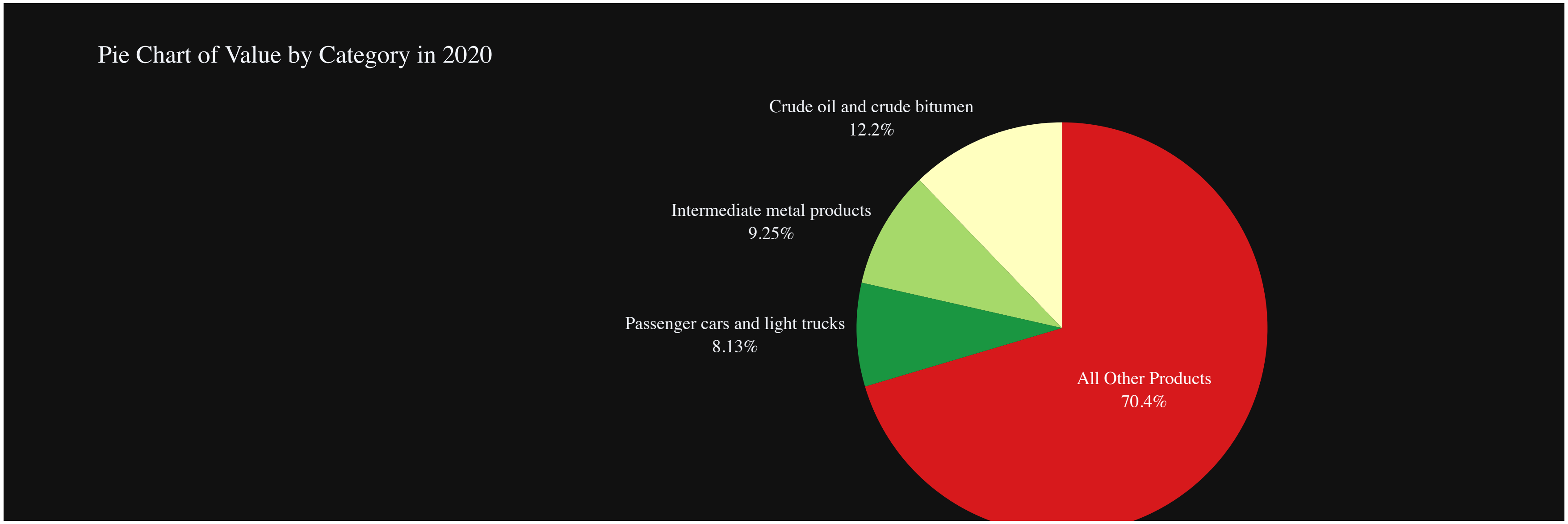
```
C:\Users\israa\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\LocalCache\local-packages\Python310\site-packages\plotly\graph_objs\_deprecatio
ns.py:378: DeprecationWarning:

plotly.graph_objs.Line is deprecated.
Please replace it with one of the following more specific types
  - plotly.graph_objs.scatter.Line
  - plotly.graph_objs.layout.shape.Line
  - etc.
```



```
In [ ]: colors =  [
            "#d7191c",
            "#ffffbf",
            "#a6d96a",
            "#1a9641",
            "#fdae61"
             ]
fig3_5 = px.pie(df3_2000_group, values='Value', names='Category', title='Pie Chart of Value by Category in 2000',hover_data=['Category'],template="plotly_dark",color_di
fig3_5.update_traces(textinfo='label+percent')
fig3_5.update_layout(font = dict(size= 15, family="Franklin Gothic"))
fig3_5.show()
```

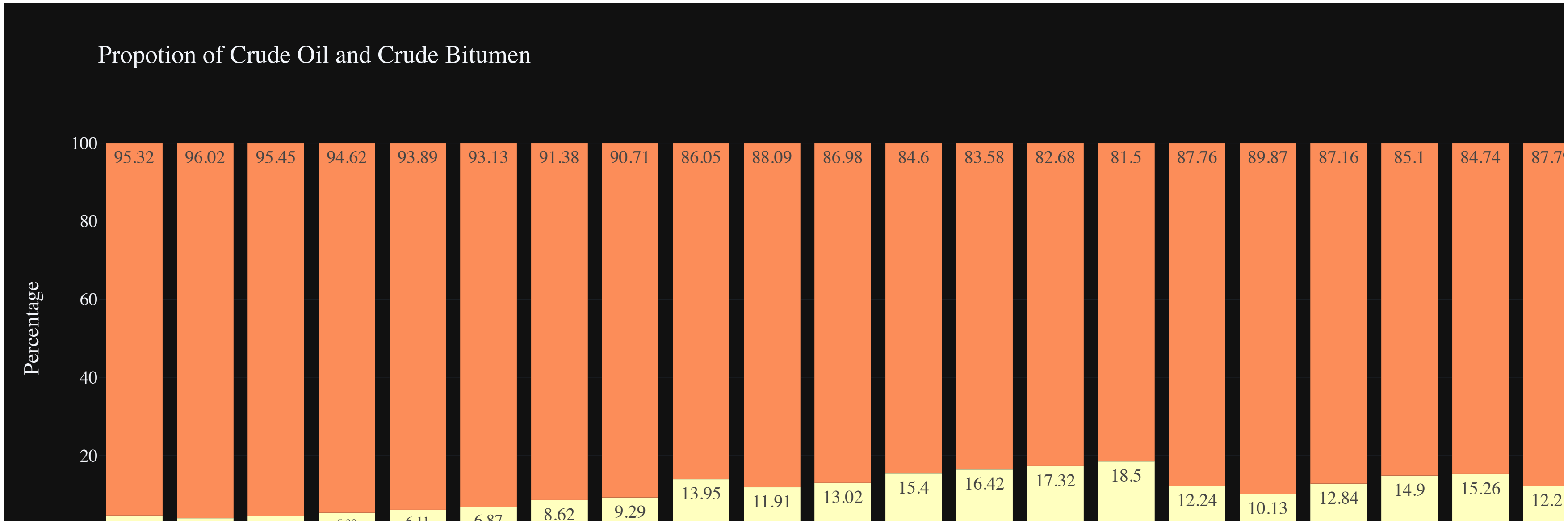


```
In [ ]: fig4 = px.pie(df3_2020_group, values='Value', names='Category', title='Pie Chart of Value by Category in 2020',hover_data=['Category'],template="plotly_dark",color_disc
        fig4.update_traces(textinfo='label+percent')
        fig4.update_layout(font = dict(size= 15, family="Franklin Gothic"))
        fig4.show()
```



In [ ]:

```
fig5 = go.Figure(data=
[go.Bar(name="Crude oil and crude bitumen",x=percentage_df["Year"],y=percentage_df["Oil Yearly Percentage"],text=percentage_df["Oil Yearly Percentage"],marker_color = "
go.Bar(name="All other products",x=percentage_df["Year"],y=percentage_df["Other Percentage"],text=percentage_df["Other Percentage"],marker_color = "#fc8d59")])
fig5.update_layout(title="Propotion of Crude Oil and Crude Bitumen",xaxis_tickangle=45,xaxis_title="Year",yaxis_title="Percentage")
fig5.layout.template = "plotly_dark"
fig5.update_layout(font=dict(size = 15,family="Franklin Gothic"))
fig5.update_layout(barmode="stack",xaxis_categoryorder="total ascending")
fig5.show()
```



Dataset 4

- Data import: Download the data from reference link: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610009601&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=2.2&cubeTimeFrame.startYear=2000&cubeTimeFrame.endYear=2020&referencePeriods=20000101%2C20200101
- Cleaning and wrangling of the dataset : The data was filtered and grouped, the oil and gas Industry data is comprised of Support activities for mining and oil and gas extraction, Mining, quarrying and oil and gas extraction and Conventional oil and gas extraction The Investment in other industries was calculated by subtracting the total investment from the oil and gas investment. The other industries in this data includes Government sector, Transportation and warehousing, Finance, insurance, real estate, rental and leasing, Utilities, Manufacturing, Other municipal government services, Educational services, Other provincial and territorial government services, Information and cultural industries, Mining and quarrying (except oil and gas), Finance and insurance, Real estate and rental and leasing, Retail trade, Professional, scientific and technical services, Other federal government services, Agriculture, forestry, fishing and hunting, Wholesale trade, Hospitals, Construction, Transportation equipment manufacturing, Crop production, Primary metal manufacturing, Accomodation and food services, Chemical manufacturing, Defence services, Administrative and support, waste management and remediation services, Food manufacturing, Computer and electronic product manufacturing, Petroleum and coal products manufacturing, Arts, entertainment and recreation, Paper manufacturing, Animal production, Health care and social assistance, Non-profit institutions serving households, Wood product manufacturing, Holding companies, Machinery manufacturing, Other services (except public administration), Beverage and tobacco products manufacturing, Nursing and residential care facilities Other aboriginal government services, Fabricated metal product manufacturing, Non-metallic mineral product manufacturing, Plastics and rubber products manufacturing, Electrical equipment, appliance and component manufacturing, Printing and related support activities, Miscellaneous manufacturing, Forestry and logging, Furniture and related product manufacturing, Textile and textile product mills, Support activities for agriculture and forestry, Fishing, hunting and trapping, Clothing and leather and allied product manufacturing

For the purposes of visualization the data was then compiled into on data frame that consisted of the oil and gas investment data, the total investment data and the investment into all other industries.

- Visualizatipn: Multi Line graphs showcasing data over the time period and Investment amount shown in millions.

In [ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
```

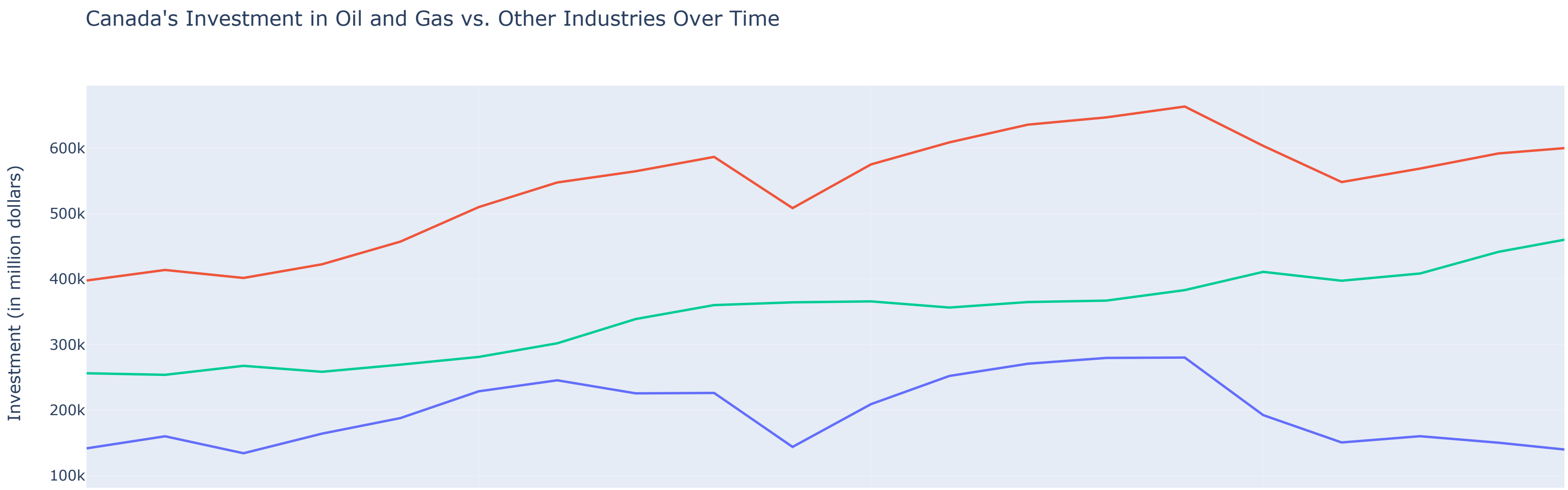
```
# Load the data from the CSV file
# Removing irrelavant fields
df = pd.read_csv("3610009601_databaseLoadingData.csv")
df = df.drop(['GEO', 'Assets', 'DGUID', 'Prices', 'UOM_ID', 'SCALAR_ID', 'SCALAR_FACTOR','Flows and stocks', 'UOM', 'VECTOR', 'COORDINATE', 'STATUS', 'SYMBOL', 'TERMINA
df.rename(columns={"REF_DATE": "Year", "VALUE": "Investment"}, inplace = True)

# Filter data for the two industries
oil_gas0 = df[df['Industry'] == 'Conventional oil and gas extraction']
oil_gas0 = oil_gas0.groupby('Year')['Investment'].sum().reset_index()
oil_gas0.insert(1, "Industry", "Conventional oil and gas extraction", True)
oil_gas1 = df[df['Industry'] == 'Mining, quarrying and oil and gas extraction']
oil_gas1 = oil_gas1.groupby('Year')['Investment'].sum().reset_index()
oil_gas1.insert(1, "Industry", "Mining, quarrying and oil and gas extraction", True)
oil_gas2 = df[df['Industry'] == 'Support activities for mining and oil and gas extraction']
oil_gas2 = oil_gas2.groupby('Year')['Investment'].sum().reset_index()
oil_gas2.insert(1, "Industry", "Support activities for mining and oil and gas extraction", True)
frames0 = [oil_gas0, oil_gas1, oil_gas2]
oil_gas = pd.concat(frames0)
oil_gas = oil_gas.groupby('Year')['Investment'].sum().reset_index()
oil_gas['Industry'] = 'Oil & Gas'

total_industries = df[df['Industry'] == 'Total all industries']
total_industries = total_industries.groupby('Year')['Investment'].sum().reset_index()
total_industries.insert(1, "Industry", "Total all industries", True)

# Calculate investment for all other industries
other_industries0 = total_industries['Investment'] - oil_gas['Investment']
other_industries1 = oil_gas[['Year']].copy()
frames1 = [other_industries0, other_industries1]
other_industries = pd.concat(frames1, axis=1, join='inner')
other_industries.insert(1, "Industry", "All other industries", True)



frames = [oil_gas, total_industries, other_industries]
result = pd.concat(frames)
display(result)
```

	Year	Investment	Industry
0	2000	141600.0	Oil & Gas
1	2001	160040.0	Oil & Gas
2	2002	134135.0	Oil & Gas
3	2003	164059.0	Oil & Gas
4	2004	187816.0	Oil & Gas
...	...	...	...
16	2016	397479.0	All other industries
17	2017	408519.0	All other industries
18	2018	441586.0	All other industries
19	2019	463571.0	All other industries
20	2020	468710.0	All other industries

63 rows × 3 columns

In [ ]:

```
# Multiine graph with "Year" on the x-axis and "Investment amount" on the y-axis
plt.figure(figsize=(10, 6))
fig = px.line(result, x="Year", y="Investment", title="Canada's Investment in Oil and Gas vs. Other Industries Over Time", color="Industry", labels={
                     "Investment": "Investment (in million dollars)",},)
#fig.layout.template = "plotly_dark"
fig.show()
```



```
<Figure size 1000x600 with 0 Axes>
```

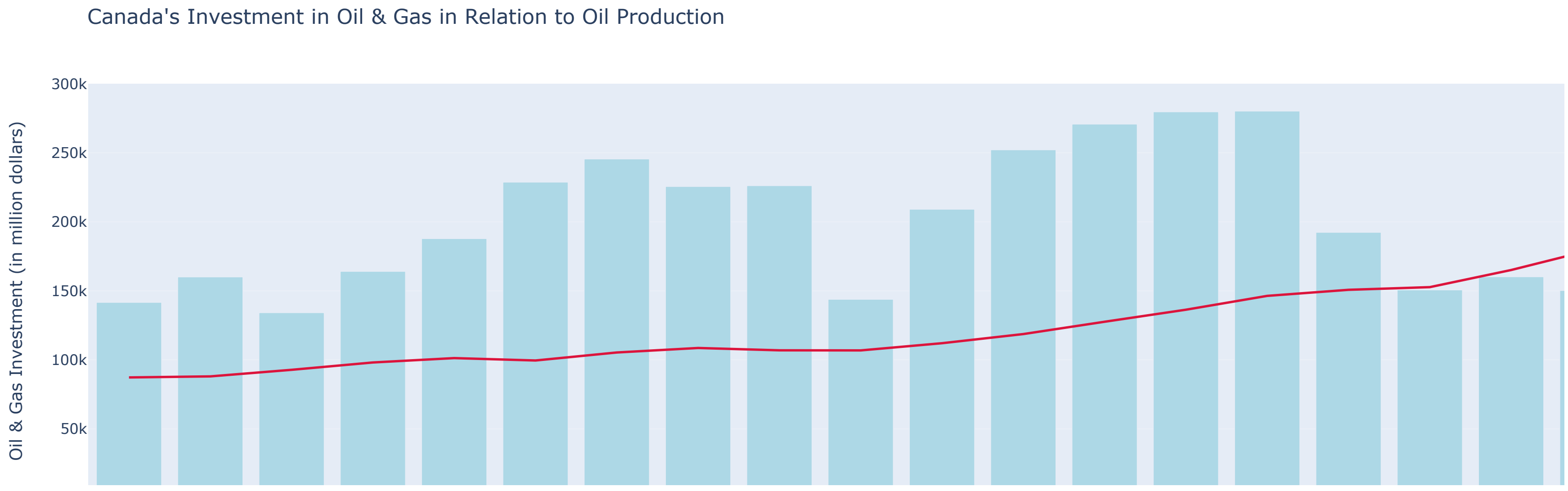
In [ ]:

```
canada_oil = country5[country5['Entity'] == 'Canada']
canada_oil = canada_oil.drop(canada_oil.index[len(canada_oil)-1])
result2 = pd.merge(oil_gas,canada_oil, on="Year")
# result2 = oil_gas.join(canada_oil["Oil production (TWh)"])
display(result2)
```

	Year	Investment	Industry	Entity	Code	Oil production (TWh)
0	2000	141600.0	Oil & Gas	Canada	CAN	1454.7484
1	2001	160040.0	Oil & Gas	Canada	CAN	1467.3573
2	2002	134135.0	Oil & Gas	Canada	CAN	1546.8190
3	2003	164059.0	Oil & Gas	Canada	CAN	1635.5127
4	2004	187816.0	Oil & Gas	Canada	CAN	1688.1901
5	2005	228733.0	Oil & Gas	Canada	CAN	1659.3553
6	2006	245505.0	Oil & Gas	Canada	CAN	1756.1113
7	2007	225581.0	Oil & Gas	Canada	CAN	1810.4688
8	2008	226186.0	Oil & Gas	Canada	CAN	1782.1812
9	2009	143826.0	Oil & Gas	Canada	CAN	1781.1637
10	2010	209094.0	Oil & Gas	Canada	CAN	1868.1313
11	2011	252155.0	Oil & Gas	Canada	CAN	1978.6215
12	2012	270790.0	Oil & Gas	Canada	CAN	2127.2776
13	2013	279675.0	Oil & Gas	Canada	CAN	2272.6140
14	2014	280216.0	Oil & Gas	Canada	CAN	2439.7065
15	2015	192356.0	Oil & Gas	Canada	CAN	2512.9456
16	2016	150625.0	Oil & Gas	Canada	CAN	2545.1667
17	2017	160163.0	Oil & Gas	Canada	CAN	2751.9353
18	2018	150227.0	Oil & Gas	Canada	CAN	2997.5752
19	2019	137873.0	Oil & Gas	Canada	CAN	3064.0650
20	2020	93642.0	Oil & Gas	Canada	CAN	2931.0240

In [ ]:

```
import plotly.graph_objects as go

df = result2
summed_values = df.groupby(by="Year", as_index=False).sum(numeric_only=True)
years = summed_values["Year"].values
oil_prod = summed_values["Oil production (TWh)"].values
invest_amt = summed_values["Investment"].values


fig = go.Figure(
    data=go.Bar(
        x=years,
        y=invest_amt,
        name="Oil & Gas Investment (in million dollars)",
        marker=dict(color="lightblue"),
    )
)

fig.add_trace(
    go.Scatter(
        x=years,
        y=oil_prod,
        yaxis="y2",
        name="Oil production (TWh)",
        marker=dict(color="crimson"),
    )
)

fig.update_layout(
    legend=dict(orientation="h"),
    yaxis=dict(
        title=dict(text="Oil & Gas Investment (in million dollars)"),
        side="left",
        range=[0, 300000],
    ),
    yaxis2=dict(
        title=dict(text="Oil production (TWh)"),
        side="right",
        range=[0, 5000],
        overlaying="y",
        tickmode="sync",
    ),
    title="Canada's Investment in Oil & Gas in Relation to Oil Production"
)

fig.show()
```



Dataset & Guiding Question Conclusion

Dataset 1

Based on our findings, Canada's overall oil production from 2000 to 2021 is ranked 6th globally. It holds a significant position among the top countries, and its oil production has shown a steady increase over the years. Canada's consistent contribution to global oil production highlights its important role in meeting global energy demands.

While it was already anticipated that Canada would be among the top countries in terms of oil production, we discovered some differences from our expectations. Originally, we anticipated that the top five countries in terms of oil production would be Russia, United States, Canada, United Arab Emirates, and Iraq. However, the actual ranking demonstrated a slight deviation from our predictions.

Dataset 2

Diverse Impact Factors: The data suggests that oil prices are significantly influenced by three main categories of events.

a. Green Arrow: The introduction of new oil production technologies, such as the SAGD method in 2002 and the shale oil boom in the USA in 2011, has consistently shown a positive correlation with oil prices. These technological advancements have had a favorable impact on oil prices.

b. Red Arrow: Political affairs have also proven to be influential. Certain political events, such as the "Paris Climate Agreement," have resulted in a positive impact on oil prices. Conversely, events like the Ukraine war or the Covid pandemic have had a negative impact or a delayed negative effect on oil prices.

c. Black Arrow: Observations from the data also indicate that OPEC-related announcements play a significant role. When oil prices remain at the bottom for an extended period, OPEC tends to announce production cuts, which, in turn, have an immediate and substantial positive impact on oil prices.

This analysis highlights the multifaceted nature of factors affecting oil prices and underscores the importance of considering these diverse influences when making predictions or decisions related to the oil market.

Dataset 3

According to the data analysis result, we can conclude that the export of crude oil has became more and more important to the Candian yearly total export. From 4,68% of the yearly total export in 2000 to 12.21% of the yearly total export in 2020, curde oil has became the top export product.

Dataset 4

The data suggests that the investment into oil and gas industry has been steadily declining since 2014. This aligns with our predictions that Canada is investing into other renewable sources of energy.

This data also showed us that despite Canada being ranked as the 6th highest global oil producer, Canada's interests lie elsewhere as it is moving away from investing into it. The production of rate has not decreased due to this decrease investment and we can conclude that there isn't a strong correlation between the two at this time. This could change due to a variety of factors such as historic events, depletion of oil reserves, infrastructure issues etc.

References

[1] Petit, A. (2021). Oil Production (1900-2020) [Dataset]. Kaggle. https://www.kaggle.com/datasets/alexandrepetit881234/oil-production-1900-2020

[2] Statistics Canada. (2018). International merchandise trade by commodity, monthly [Dataset]. Government of Canada. https://doi.org/10.25318/1210012101-ENG

[3] Statistics Canada. (2018). Flows and stocks of fixed non-residential capital, by industry and type of asset, Canada, provinces and territories [Dataset]. Government of Canada. https://doi.org/10.25318/3610009601-ENG

[4] Spot Prices for Crude Oil and Petroleum Products. (2020). Eia.gov. https://www.eia.gov/dnav/pet/PET_PRI_SPT_S1_D.htm

[5] Overview — NumPy v1.21 Manual. (n.d.). Numpy.org. https://numpy.org/doc/stable/index.html

[6] Pandas documentation — pandas 1.0.3 documentation. (n.d.). Pandas.pydata.org. https://pandas.pydata.org/docs/index.html

[7] Modern Analytic Apps for the Enterprise - Plotly. (n.d.). Plotly.com. https://plotly.com/

[8] Matplotlib documentation — Matplotlib 3.5.0 documentation. (n.d.). Matplotlib.org. https://matplotlib.org/stable/

[9] ‘Energy & Financial Markets: What Drives Crude Oil Prices? - Energy Information Administration’. Accessed 2 October 2023. https://www.eia.gov/finance/markets/crudeoil/spot_prices.php