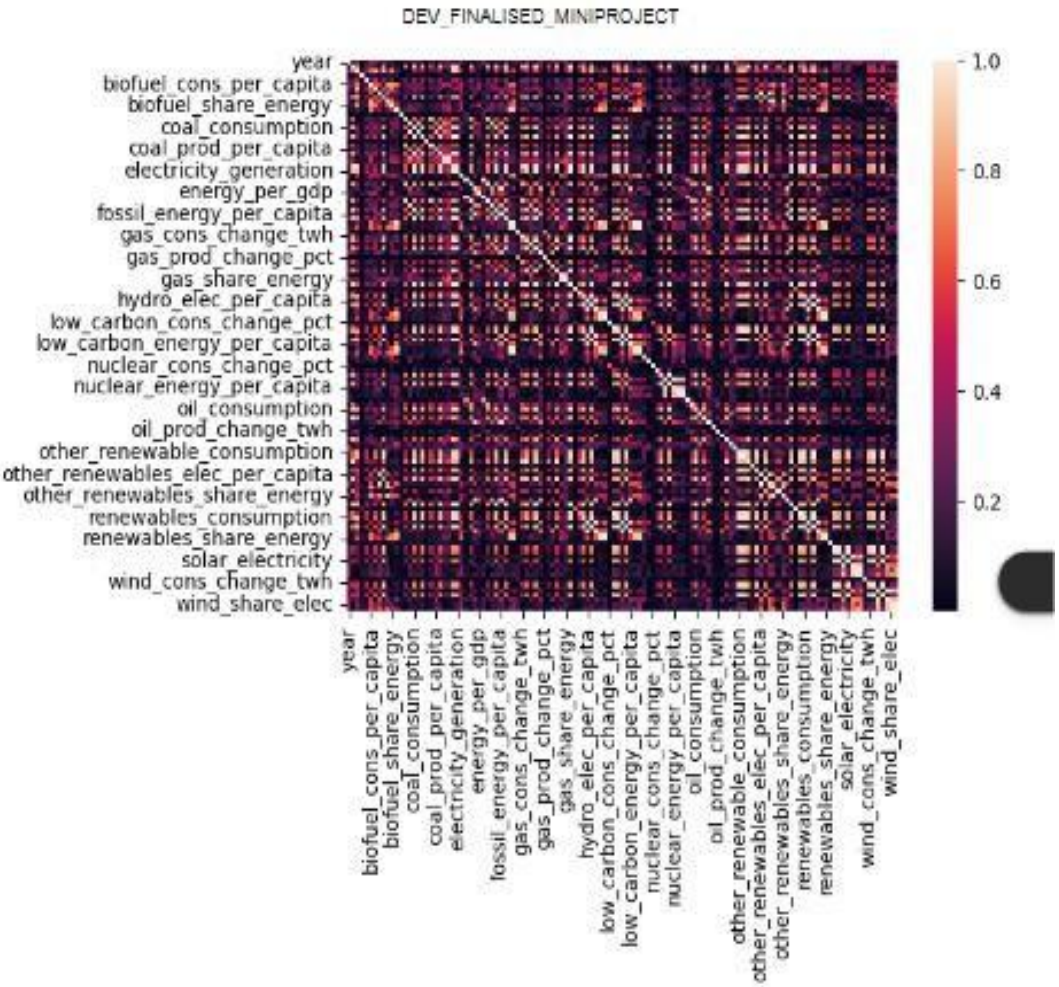
DEV MINI PROJECT – ENERGY CONSUMPTION DATASET

- **<u>Exploratory Data Analysis and Data Preprocessing (Overview):</u>**



From the given correlation matrix, one can infer that the dataset is extremely unclean and requires essential steps of preprocessing.

DROPPING ALL THE NULL VALUES OF SHEET 1 (Ember)

```
[ ]  drop_list = []

     for i in data.columns:
         if data[i].isnull().any():
             drop_list.append(i)

     print(drop_list)

     data.drop(columns=drop_list, inplace=True)
```

```
['population', 'gdp', 'biofuel_cons_change_pct', 'biofuel_cons_change_twh', 'biofuel
```

OVERVIEW OF THE DATA POINTS

```
[ ]  output = []
     for col in data.columns:
         unique = data[col].nunique()
         colType = str(data[col].dtype)
         categories=data[col].unique()

         output.append([col, unique, colType,categories])

     output = pd.DataFrame(output)
     output.columns = ['colName','unique','dtype','categories']
     output
```

FILL THE NaN VALUES WITH THE MEDIAN OF ALL THE DATA POINTS

Main Dataset

In [14]:
```
df.replace(0, np.nan, inplace=True)

for country, group in df.groupby('country'):
    for col in group.columns[2:]:  # Exclude the 'country' column and any other non-numeric columns
        median_val = group[col].median()
        df.loc[group.index, col] = group[col].fillna(median_val)

# Now df contains NaN values replaced with the median of each column grouped by country
```

CREATE EXCEL SHEETS FOR EACH SPECIFIC ORGANISATION DATA

In [15]:
```
with pd.ExcelWriter('International_Org.xlsx') as writer:

    ember_countries = ['Africa (Ember)', 'ASEAN (Ember)', 'Asia (Ember)', 'Europe (Ember)', 'G20 (Ember)', 'G7 (Ember)', 'North America (Ember)', 'Latin Amer
    df[df['country'].isin(ember_countries)].to_excel(writer, sheet_name='Ember', index=False)

    shift_countries = ['Africa (Shift)', 'Asia and Oceania (Shift)', 'Central and South America (Shift)', 'EU28 (Shift)', 'Eurasia (Shift)', 'Europe (Shift)'
    df[df['country'].isin(shift_countries)].to_excel(writer, sheet_name='Shift', index=False)

    eia_countries = ['Africa (EIA)', 'Asia & Oceania (EIA)', 'Australia and New Zealand (EIA)', 'Central & South America (EIA)', 'Eurasia (EIA)', 'Europe (EIA
    df[df['country'].isin(eia_countries)].to_excel(writer, sheet_name='EIA', index=False)

    ei_countries = ['Africa (EI)', 'Asia Pacific (EI)', 'CIS (EI)', 'Central America (EI)', 'Eastern America (EI)', 'Europe (EI)', 'Middle Africa (EI)', 'Midd
    df[df['country'].isin(ei_countries)].to_excel(writer, sheet_name='EI', index=False)
```

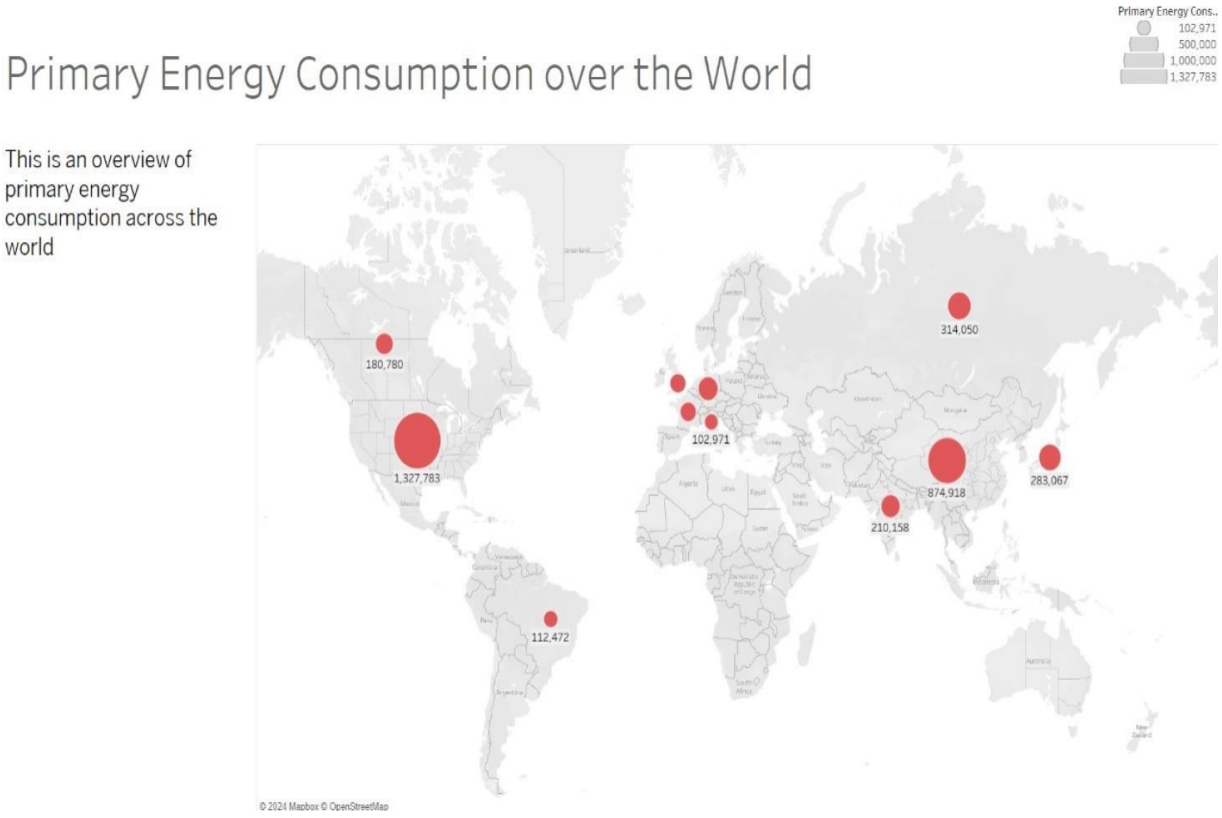
In [16]:
```
countries_to_delete = ember_countries + shift_countries + eia_countries + ei_countries

df = df[~df['country'].isin(countries_to_delete)]

df.to_excel('updated_dataset.xlsx', index=False)
```

With the help of this process, it is easier to group the significant and essential data together which will then be applied to visualize in Tableau for drawing impactful conclusions.

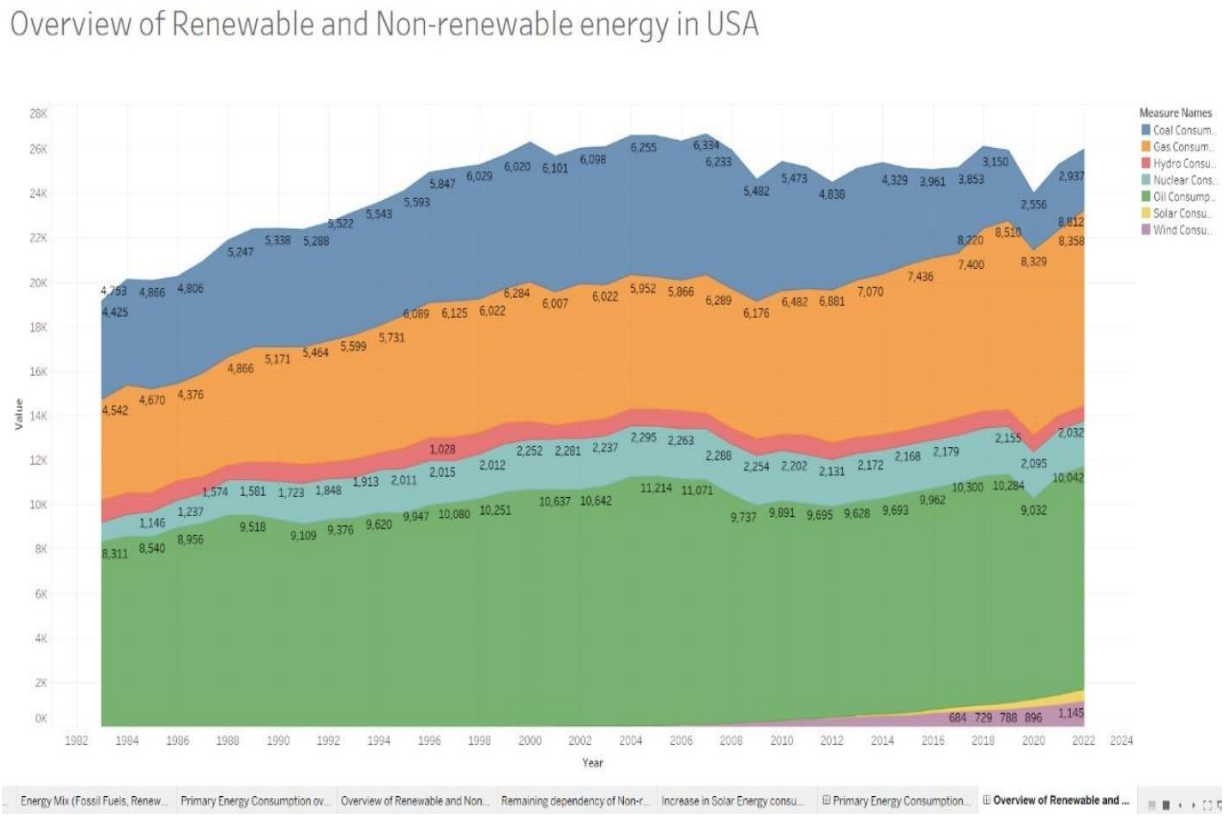
Conclusion: Thus, the data preprocessing steps that were taken are:
1. The dataset contained a lot of null values that needed to be treated so they were filled using the median from the previous values
2. The Dataset contained values from various organizations that were split into different excel files, various other pre processing was taken place
3. The new dataset created was processed and plots were made to check the potency of the new data

- **<u>Data Visualizations with Tableau:</u>**
  Aim: To analyze and compare the consumption of different types of energies such as non – renewable (coal, oil, gas, etc.) and non – renewable (solar, wind, etc.) by the developed and developed countries of the World. Thus, to draw meaningful insights, we will consider United States of America (the superpower) and Africa (developing country) as a baseline for our analysis.

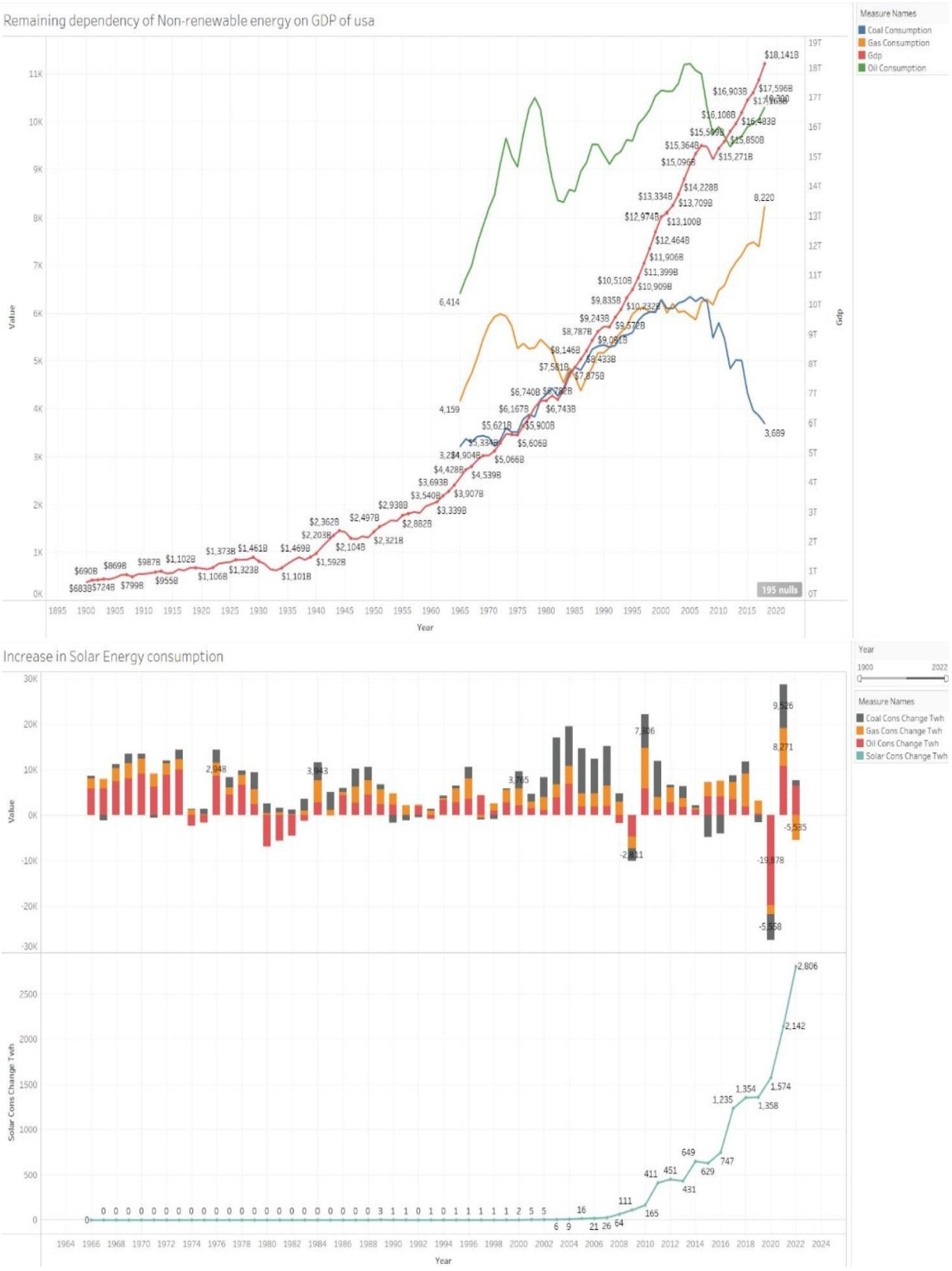
1) America –



- **It is an overview of primary energy consumption across the world.**
- **In which, US consumes the highest number of primary energy resources as compared to other continents**



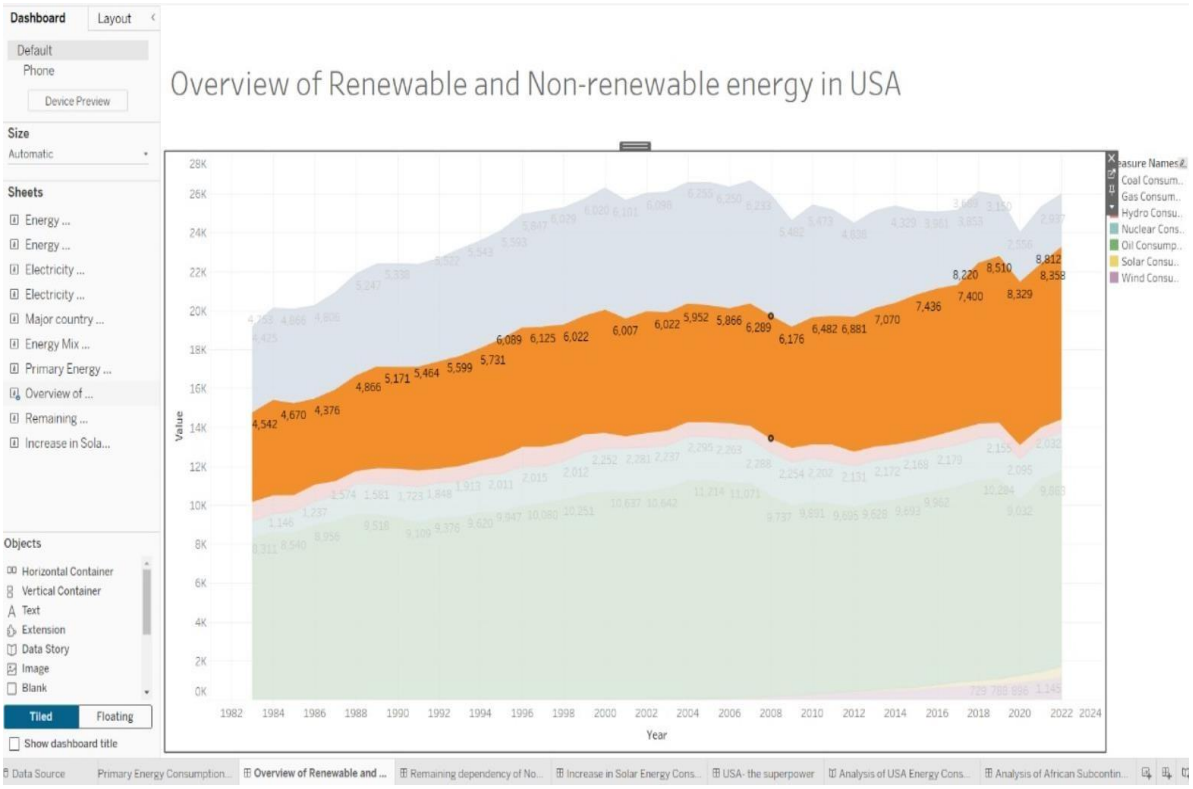
- **This gives us an idea how much USA consumes Non-Renewable resources like Coal, Gas, Hydro , Nuclear, Oil .**
- **However, the adoption of Renewable Resources like Solar and Wind can be seen rising in the recent years.**
- **The depletion in the consumption of all resources during 2020 is due to the pandemic occurrence of COVID – 19 across the world which impacted the USA resources as well.**



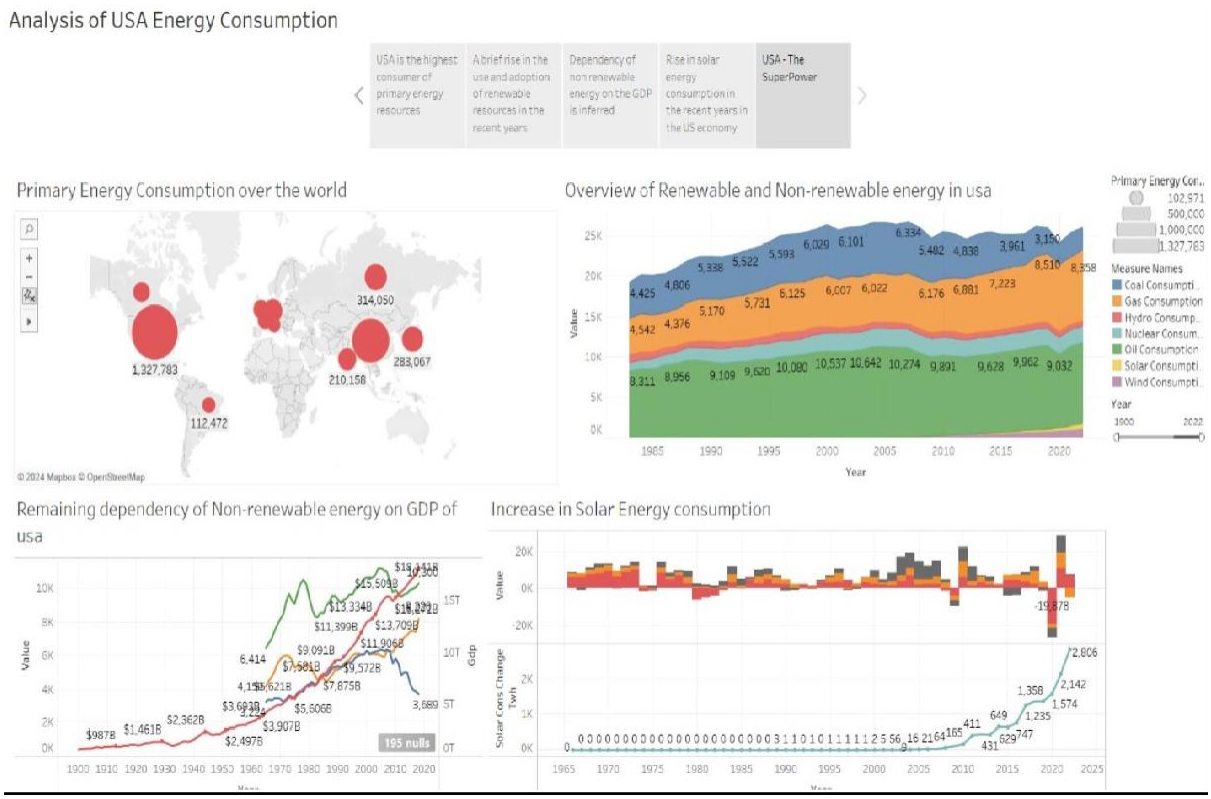
- **In the above graphs, when we compare the consumptions with the American GDP, there is a decline in the use of coal resources however the GDP still continues to rise**

- **The reason behind this, is the maximum consumption of Solar resources since the past few years due to the introduction of TESLA Solar Panels in the American Market**

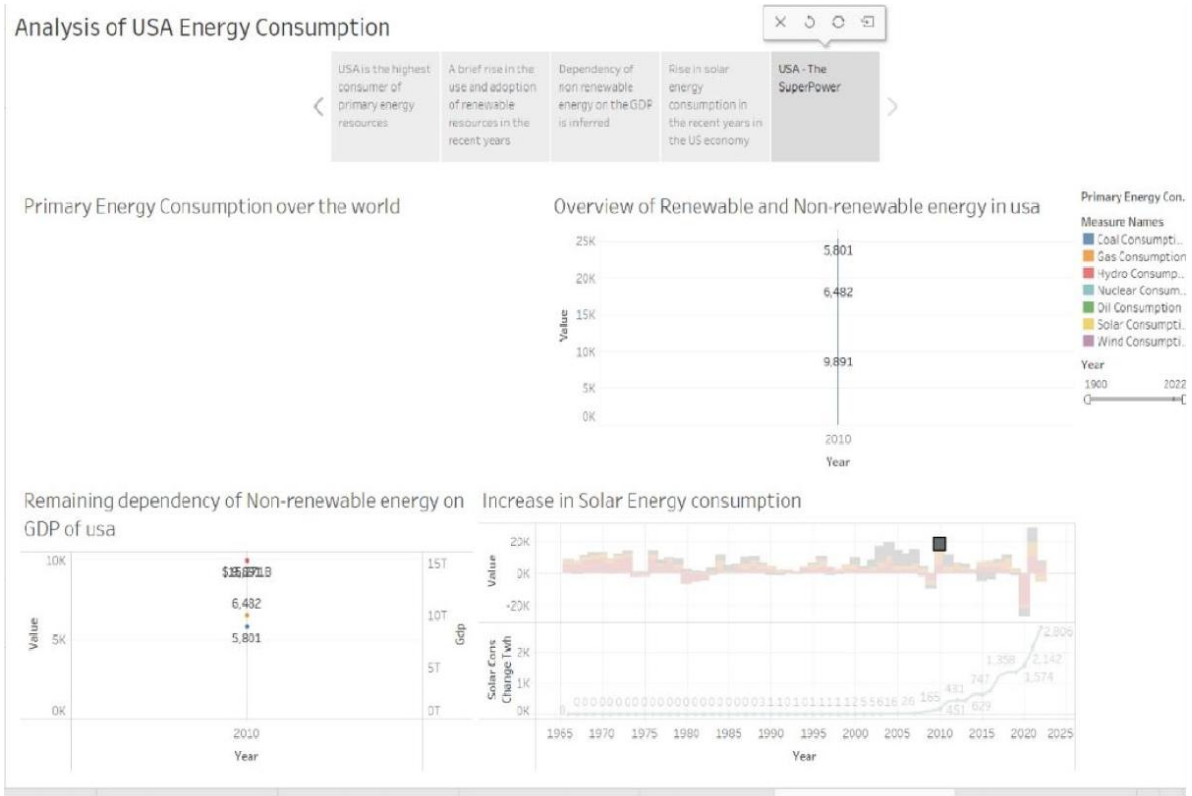
- **<u>Interactive Dashboard:</u>** Analysis of solar consumption



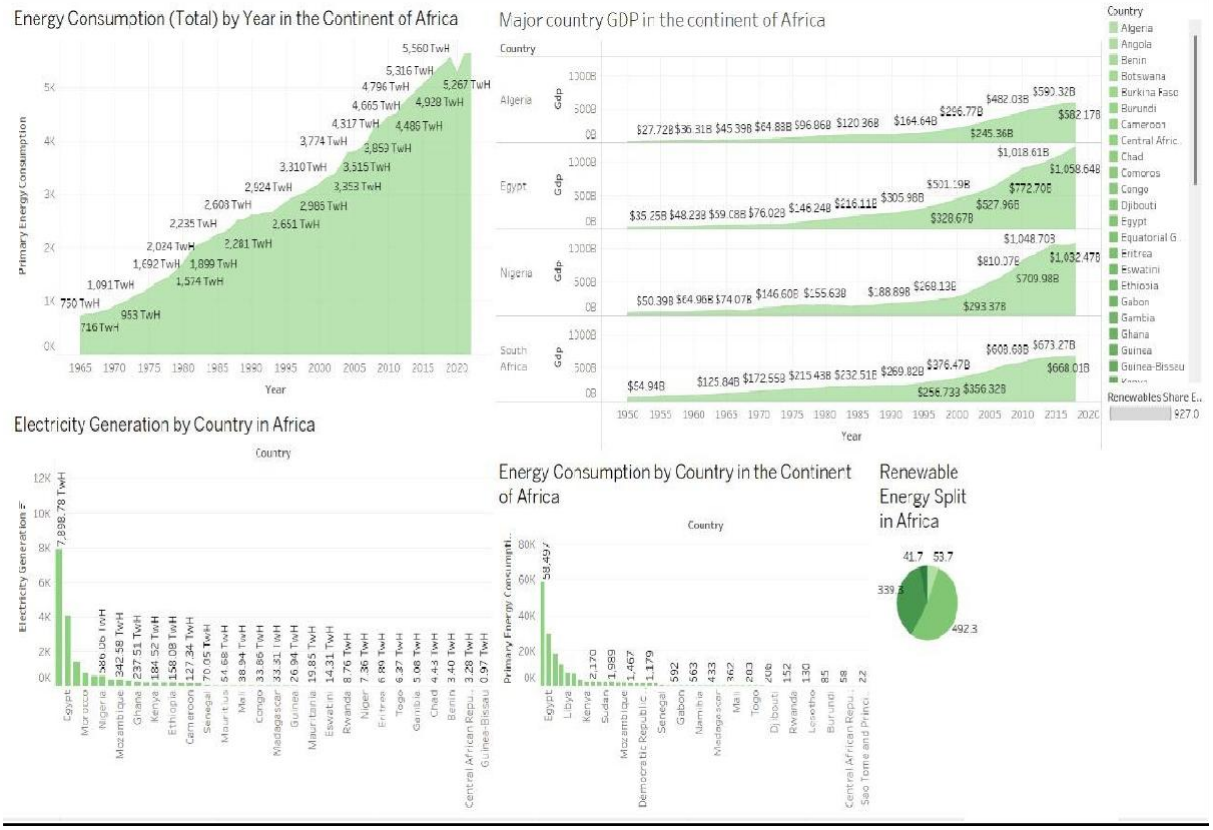
- **Story:**



Analysis of year 2010

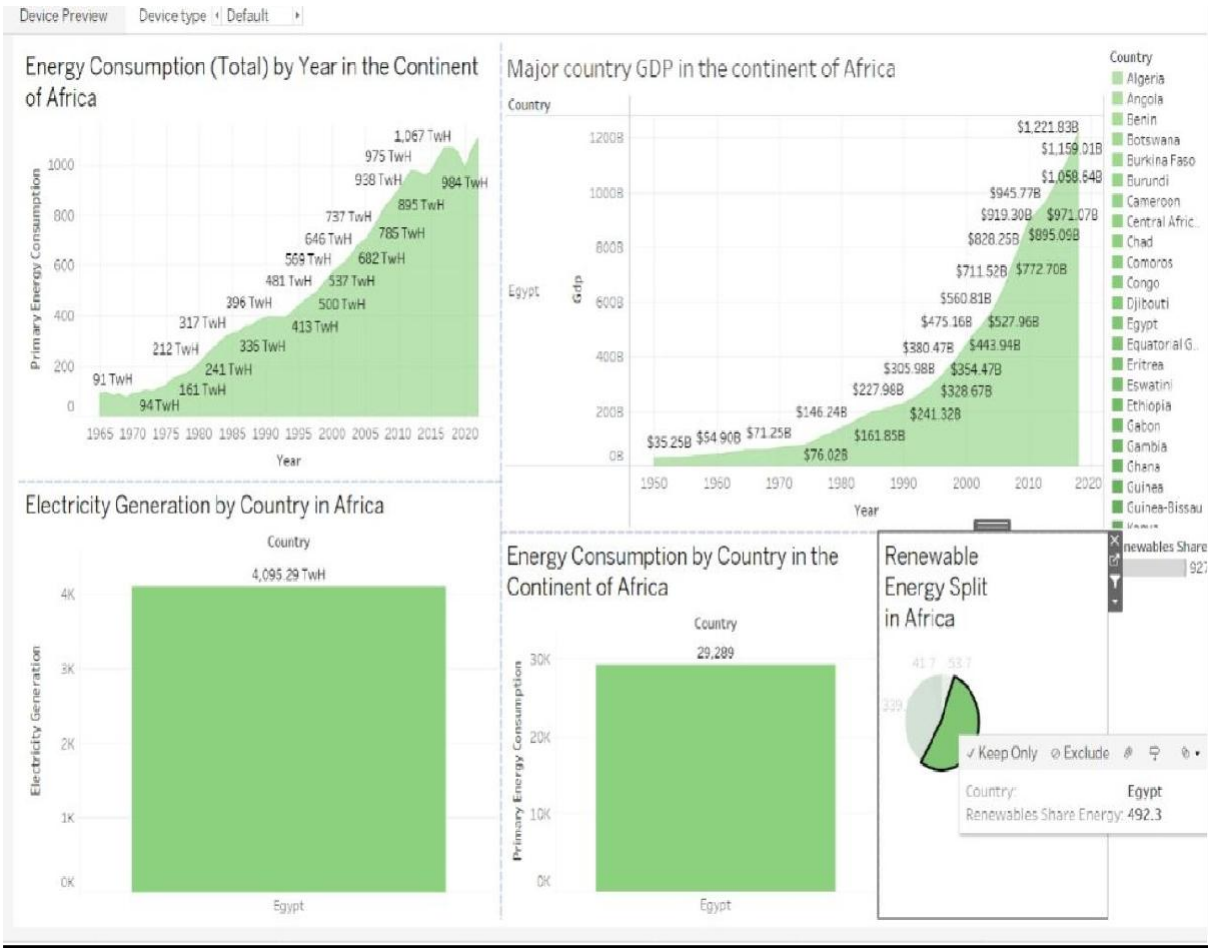


2) <u>Africa:</u>



- **Focusing on Africa, we understand that continents major GDP income is dependent on the countries like South Africa, Egypt, Nigeria, Algeria. It majorly depends consuming energy through renewable resources**
- **Throughout the years, the renewable resources based consumption keeps increasing , however if we observe carefully in the year 2010 there is a slight decline in the GDP of Egypt**
- **This was lead due the government policy that reduced the deficit gradually to 3 percent of GDP by 2010/11, which would put public debt on a firmly declining path.**

- **<u>Interactive Dashboard:</u>**



<u>Conclusion:</u> Henceforth, considering the GDP and all other types of energy consumptions led by USA and Africa we can conclude that even though America is the super power and the developed country in the entire world, it still has not completely made a progressive approach towards using renewable energy while Africa being a developing country has led the use of renewable resources energy consumption since many years in order to reduce the depletion of non – renewable resources in nature.