

## Problem statement:

### Identifying the Country with the Highest Renewable Energy Production:

Determine the country that has consistently exhibited the highest electricity or energy production from renewable sources over the years.

### Analyzing the Proportion of Renewable Energy in Each Country Over Time:

Examine the proportion or percentage of renewable energy within the total energy production for each country across multiple years.

### Determining the Most Produced/Consumed Renewable Energy Source by Year and country:

Identify the renewable energy source (e.g., solar, wind, hydro, etc.) that has consistently shown the highest production or consumption levels on a yearly basis.

# Data cleaning/ pre processing

## Cleaning the Entity category

making a function for cleaning the values of 'entity' column.

```
def get_unique_col(df):
    unique_col = []
    waste_catecter = ['(BP)', '(27)', '(Ember)']
    unique_val = ''
    for i, val in enumerate(df['Entity'].unique()):
        for j in val.split():
            if j in unique_col:
                unique_val = ''
                break
            if j not in waste_catecter:
                unique_val += j
        if unique_val:
            unique_col.append(unique_val)
        unique_val = ''

    return unique_col

def clean_entity(df):
    unique_col = get_unique_col(df)
    drop_index = []
    df_dummy = df
    waste_carecters = ['(BP)', '(27)', '(Ember)']

    for i, val in enumerate(df_dummy['Entity']):
        val_new = val
        for waste_carecter in waste_carecters:
            val_new = val_new.replace(waste_carecter, '')
        if val_new not in unique_col:
            drop_index.append(i)
        if val != val_new:
            df_dummy.loc[df_dummy.index[i], 'Entity'] = val_new

    df_dummy = df_dummy.drop(drop_index)
    return df_dummy

for df in df_list:
    df = clean_entity(df)
    df.reset_index(drop=True)
```

## Dealing with outliers

checking for outliers.

as the data dosnt contain proper seasonality, this step redundant.

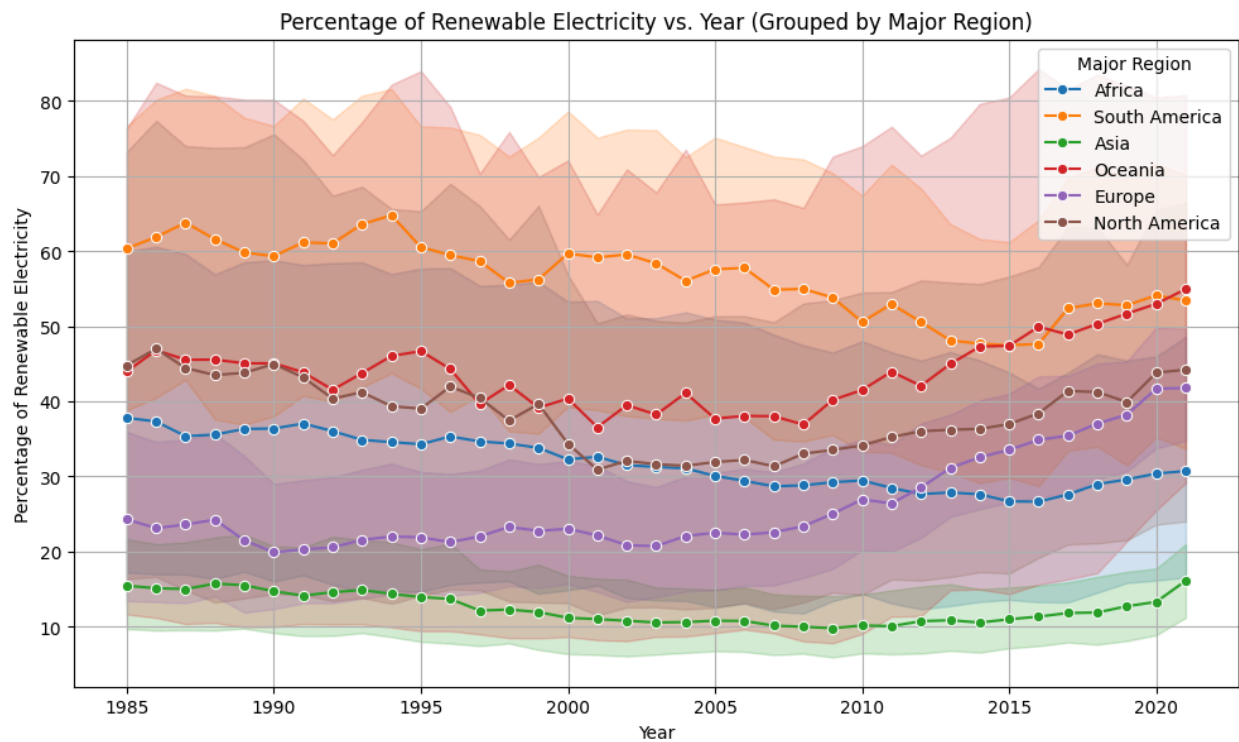
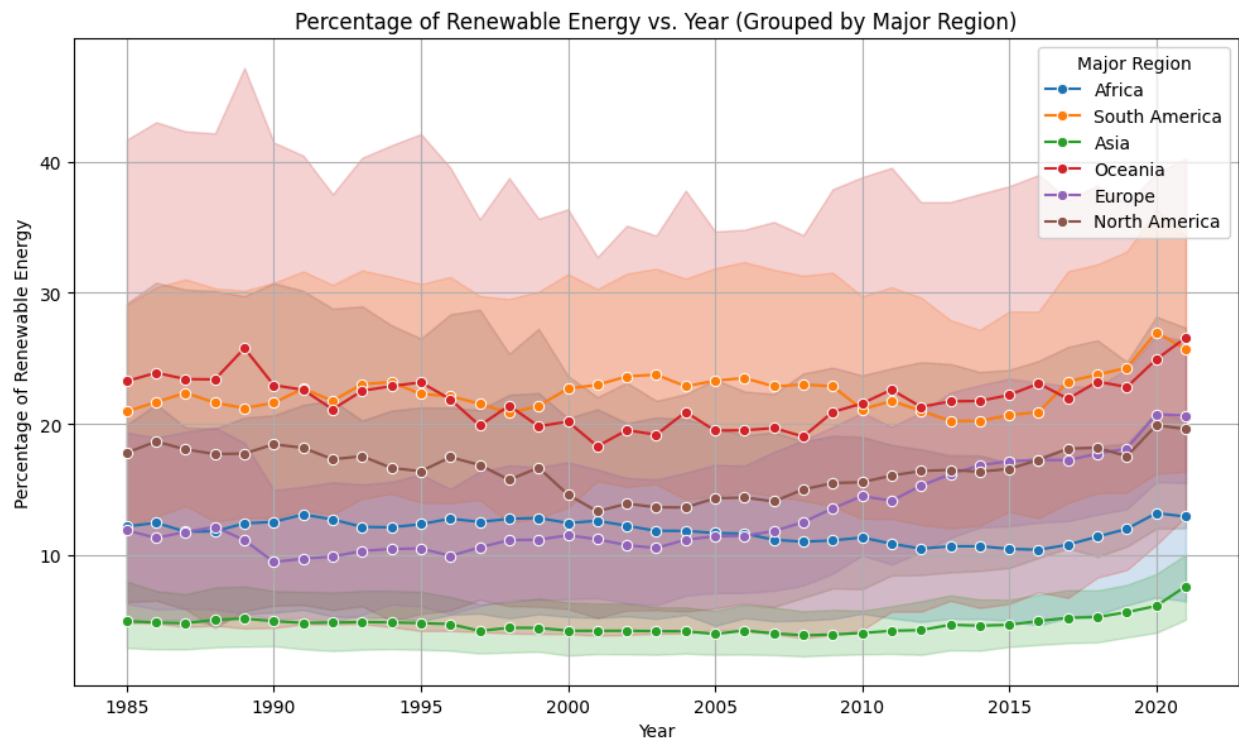
## Combining data for eda

combining datagames requiered the problem statement.

```
df_share_renewable = combine_df([df_01,df_04])  
df_mordern_renewable = combine_df([df_02,df_03])  
df_indivijual_energy_electricity = combine_df([df_06,df_07,df_10,df_11,df_14,df_15])
```

# EDA

1)



thw first graph shows the transition of % renewable energy over years for these 6 regions.

we can see that europe and asia has remained consistent with the energy production over time,  
while the rest of the rejons have improved frpm 2010 onwards.

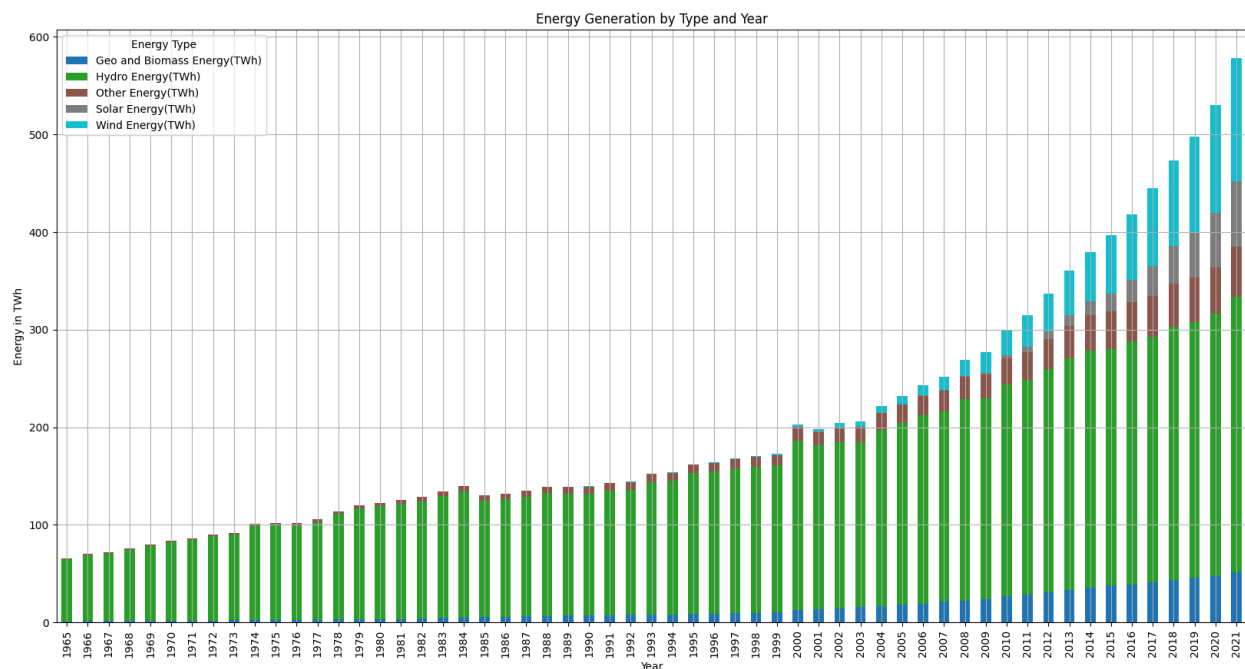
Europe is the contry with highest rewnable energy production.

the second graph shows the transition of % renewable electricity over years for these 6 regions.

we can see that asia, north america and africa and norway has remained consistent with the energy production over time,  
while europe and oceania have increased there reliance on renewable electricity.

Europe is the contry with highest rewnable energy production.

2)



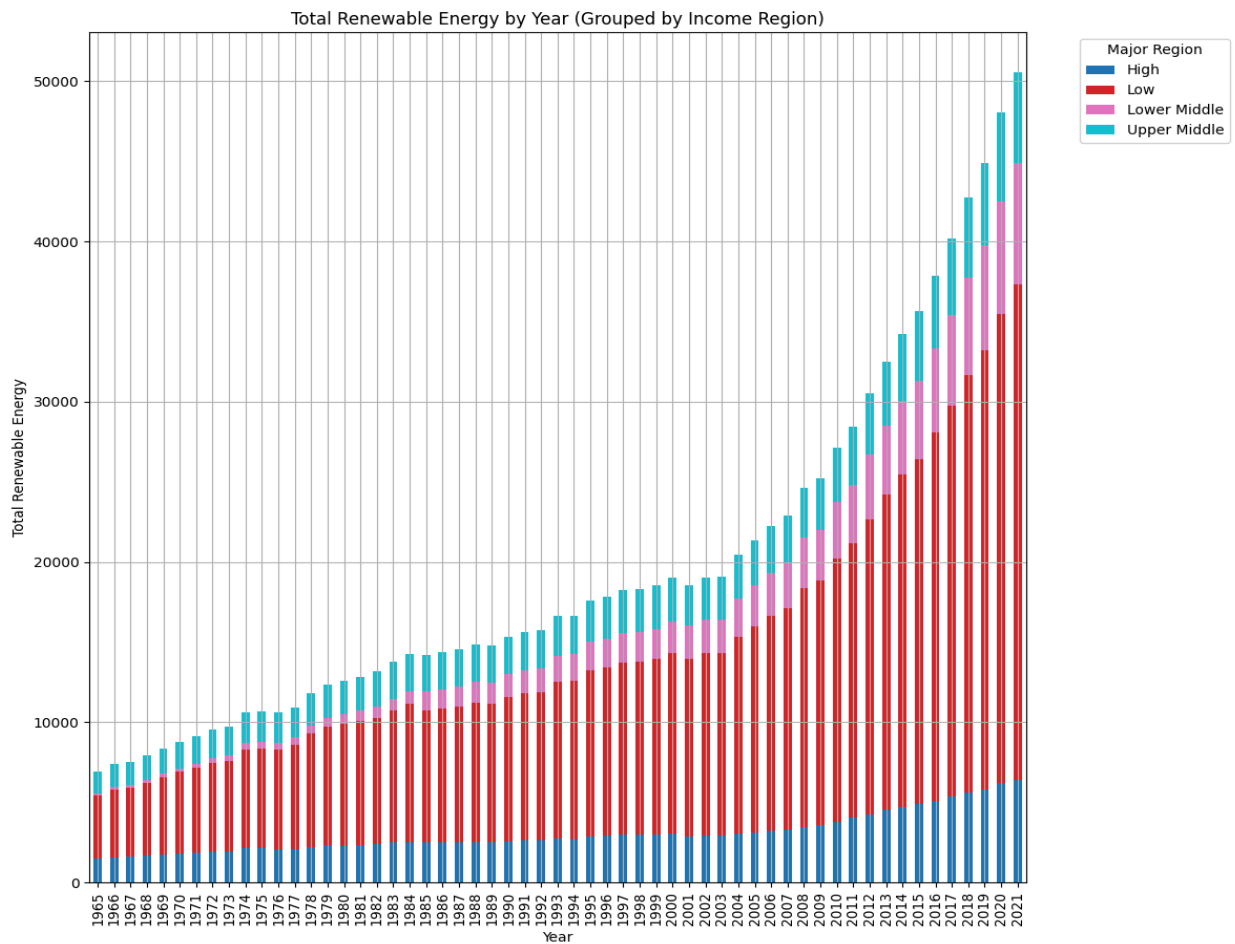
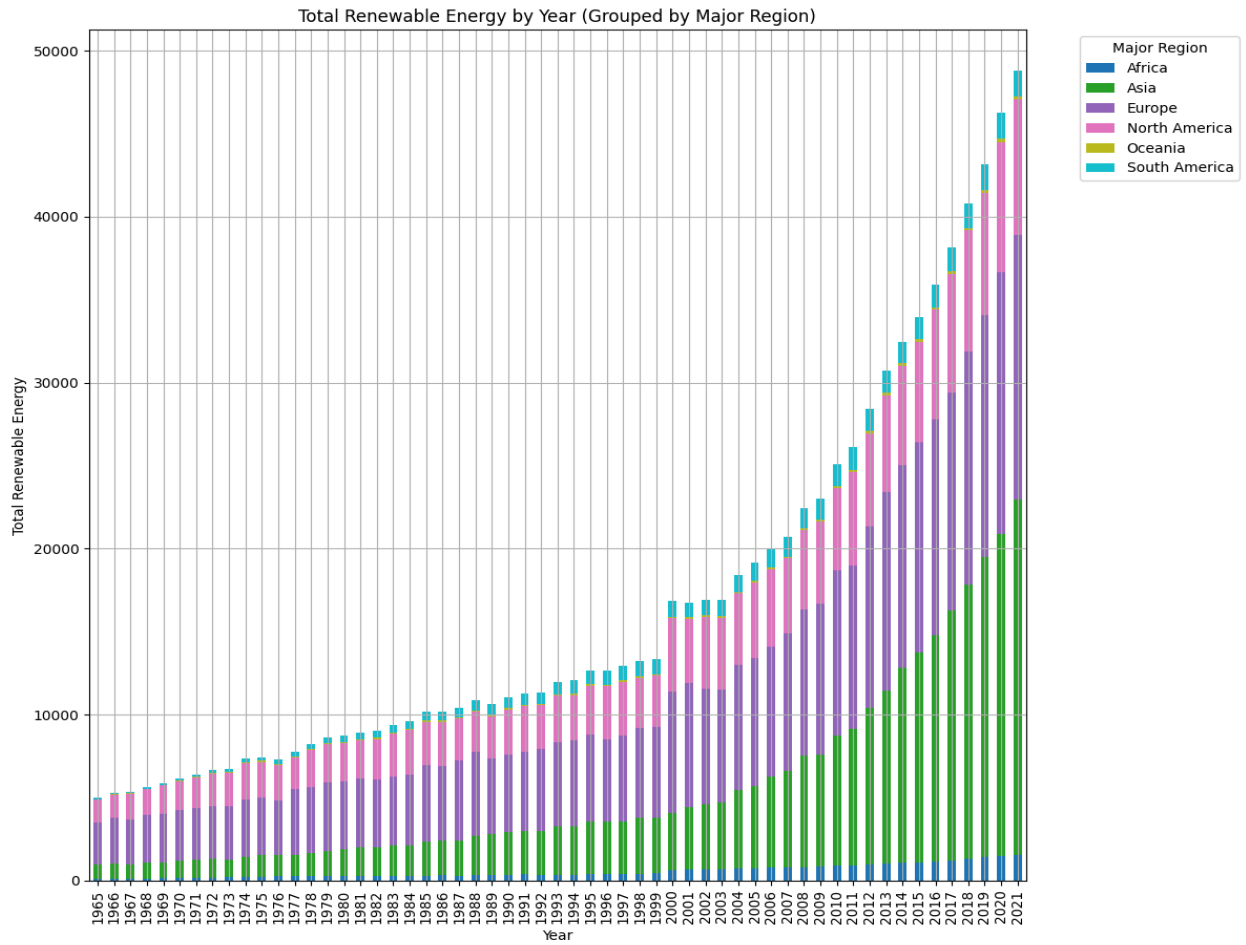
this graph shows that during 1965 the primary source of energy was hydro,  
geo is evolving slowly from 1980

whilethere was a rapid improvement in wind from 2000, solar from 2010.

conclusion:

wind energy is the most promising and evolving renewable energy as per 2021.

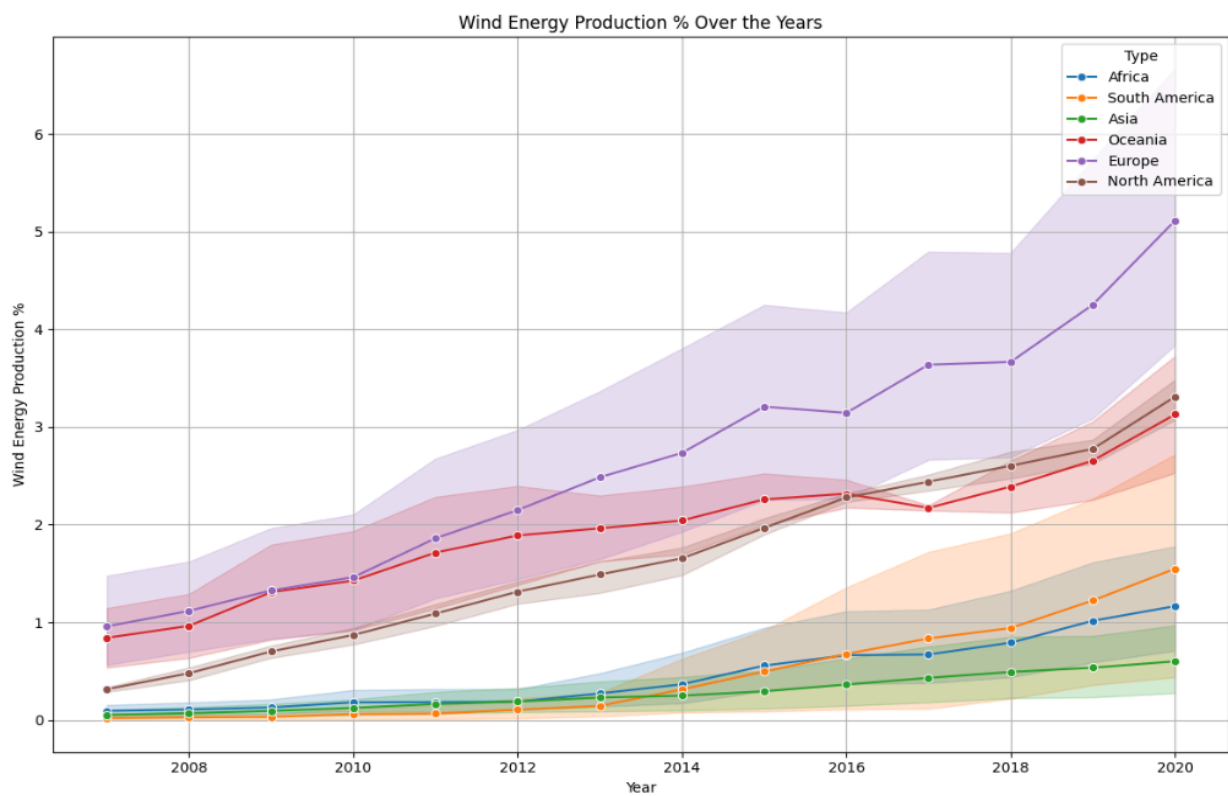
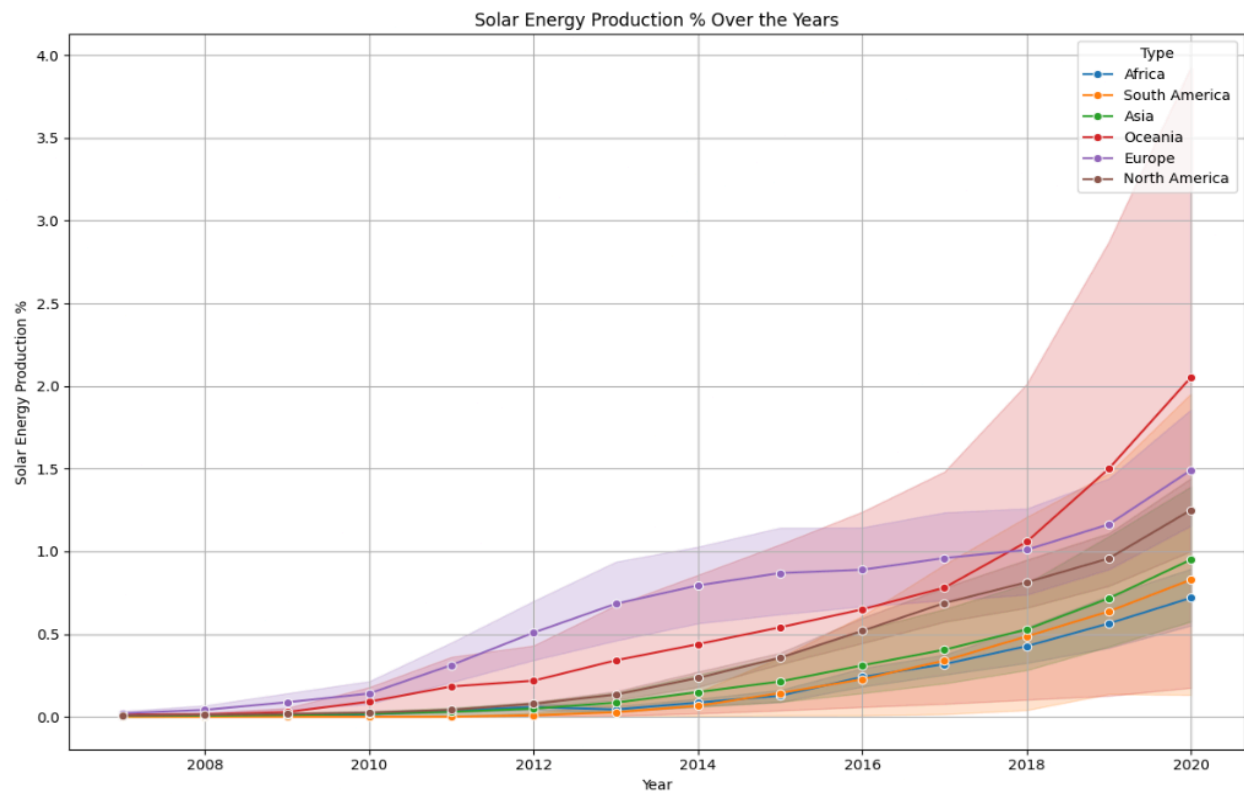
hydro energy has remained the highest source of renewable energy.



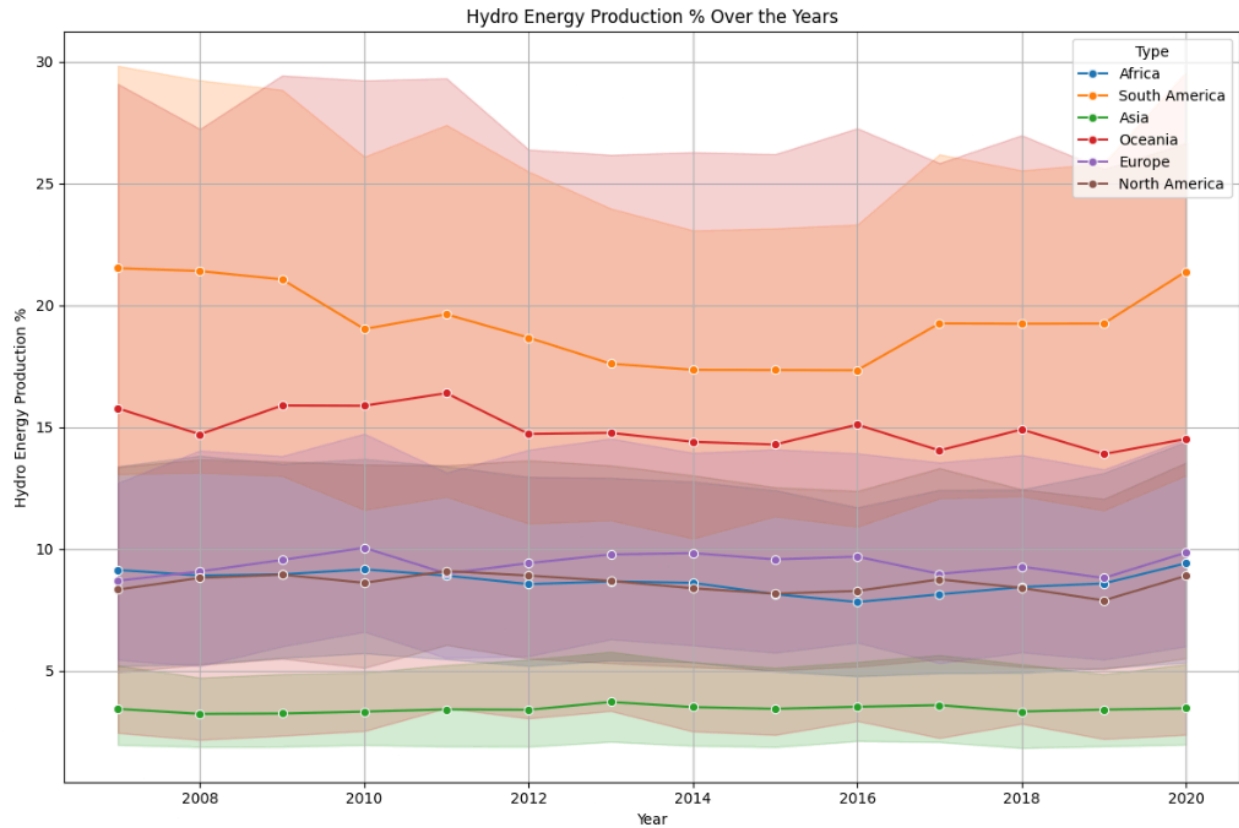
in first graph from 2010 there is significant growth in energy production in asia and europe.  
asia is the highest energy producer.

in second graph low income contries are the major contributors of global renewable energy.

3)







in first graph oceane has the highest growth in solar energy production.

in second graph europe has the highest growth in wind energy production.

in third graph southaffrica has the highest hydro energy production.

# Feature selection

## **the csv files we are working on:**

for problem statement 1, (1 and 4)

for problem statement 2, (2 and 3)

for problem statement 3, (6,7,10,11,14,15)

## **combining the eda results:**

1) south america is the king of hydro energy production (based on third problem statement graphs),

from second problem statement we know that hydro market share is the highest

these 2 together show why south africa is the top overall.

though south america is not growing.

2) europe has the highest growth in renewable energy (based on first problem statement graph),

europe is the king of wind energy and is a good competitor of solar,

this is further justified as we see that wind energy is the highest growing renewable energy (shown in second problem statement)

as we proved that wind energy was the highest growing renewable energy since 2010.

## **Answering the problem statements:**

Identifying the Country with the Highest Renewable Energy Production:

south america is the country with the highest renewable energy production (specifically in hydro)

while europe is the top growing regions for renewable energy (specifically wind)

ocean has the highest solar while second in wind.

Analyzing the Proportion of Renewable Energy in Each Country Over Time:

south america major energy source is hydro while there is no growth.

europe has a mixture of solar and wind.

ocean is the top in solar energy, and is one of the top in hydro and wind.

africa, asia and south america has a mix of hydro, wind and solar.

Determining the Most Produced/Consumed Renewable Energy Source by Year and country:

the most produced and consumed renewable energy has and remained to be hydro.