Data Science

Practice Visualization Project

Wildfire Activities in Australia

In this project, I have to perform as Data Analyst to explore patterns and trends, and create visualizations to gain insights into the behavior of wildfires in different Australia's regions and then create a dashboard wherein the user can select the Region and the Year.

Prepared by :
Smide ALEXIS

Supervised by :
Coursera Staff

Data Science with Python

21 Janvier 2024

Importing Required Packages

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sbn
import folium
```

Importing Data

```
URL = ("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM"
"DeveloperSkillsNetwork-DV0101EN-SkillsNetwork/Data%20Files/Historical_Wildfires.csv")
df = pd.read_csv(URL)
print(df.head())
```

```
  Region      Date  Estimated_fire_area  ...  Var_confidence  Count  Replaced
0    NSW  1/4/2005              8.68000  ...        8.333333      3         R
1    NSW  1/5/2005             16.61125  ...       65.428571      8         R
2    NSW  1/6/2005              5.52000  ...       10.333333      3         R
3    NSW  1/7/2005              6.26400  ...       56.700000      5         R
4    NSW  1/8/2005              5.40000  ...       63.000000      3         R

[5 rows x 10 columns]
```

This wildfire dataset contains data on fire activities in Australia starting from 2005.
Let's verify the column names and the type of each variables.

```
df.dtypes
```

```
Region                                 object
Date                                   object
Estimated_fire_area                   float64
Mean_estimated_fire_brightness        float64
Mean_estimated_fire_radiative_power   float64
Mean_confidence                       float64
Std_confidence                        float64
Var_confidence                        float64
Count                                   int64
Replaced                               object
dtype: object
```

Notice that the type Date variable is 'object', let's convert it to 'datatime' type and also let's extract 'Year' and 'Month' from date and include in the dataframe as separate columns.

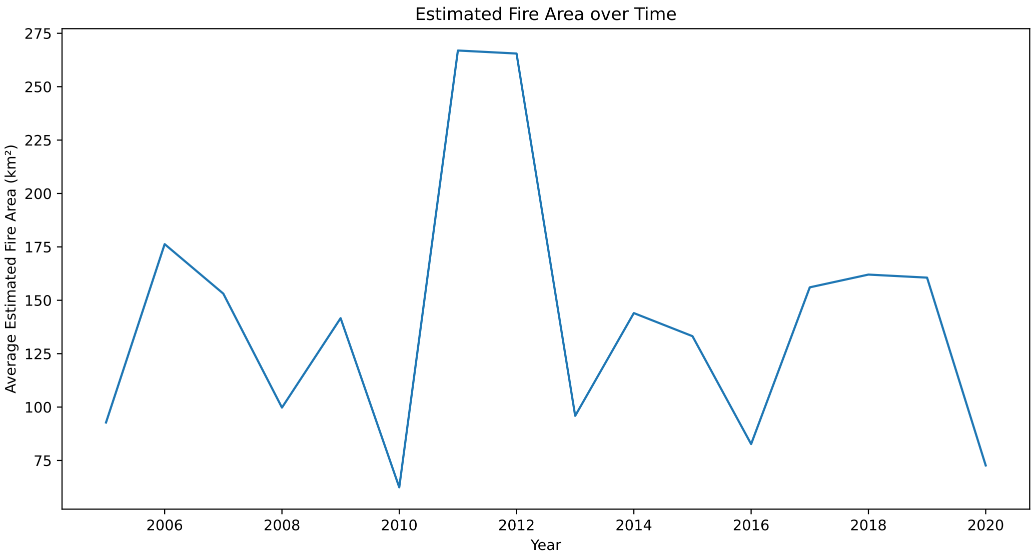
```
import datetime as dt

df['Year'] = pd.to_datetime(df['Date']).dt.year
df['Month'] = pd.to_datetime(df['Date']).dt.month
df.dtypes
```

```
Region                                 object
Date                                   object
Estimated_fire_area                   float64
Mean_estimated_fire_brightness        float64
Mean_estimated_fire_radiative_power   float64
Mean_confidence                       float64
Std_confidence                        float64
Var_confidence                        float64
Count                                   int64
Replaced                               object
Year                                    int32
Month                                   int32
dtype: object
```

Line Plot

Let’s try to understand the change in average estimated fire area over time.

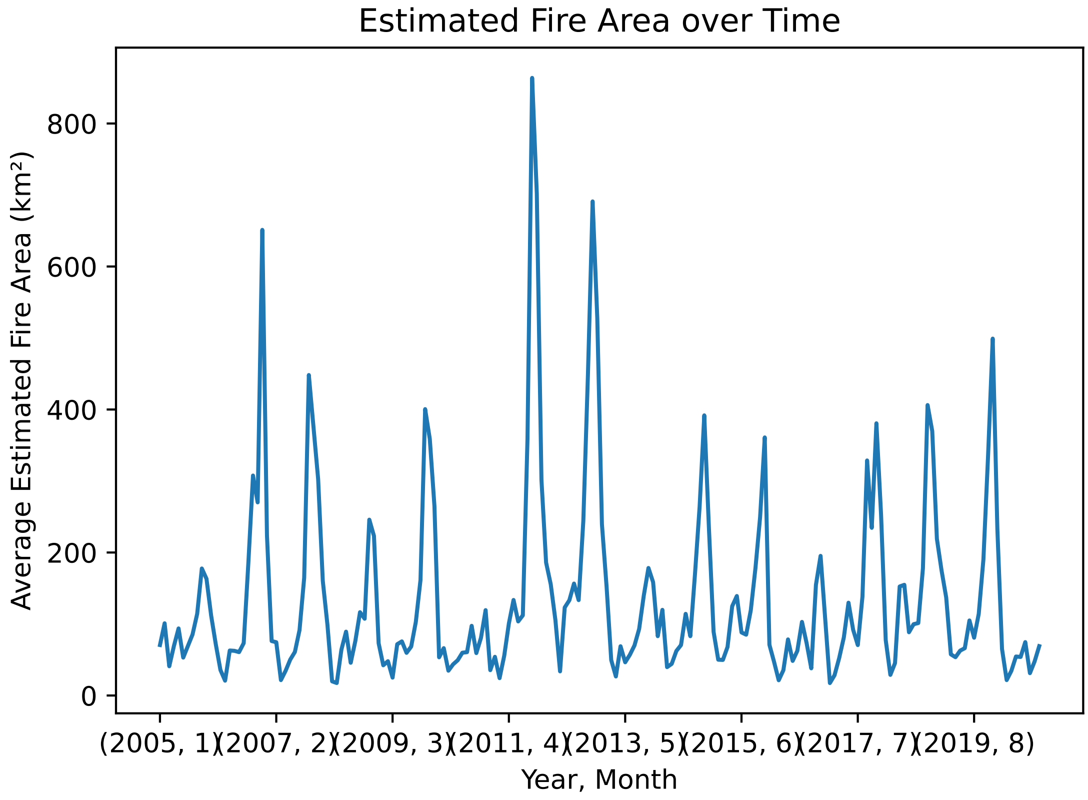
```
plt.figure(figsize=(12, 6))

df_new=df.groupby('Year')['Estimated_fire_area'].mean()

df_new.plot(x=df_new.index, y=df_new.values)
plt.xlabel('Year')
plt.ylabel('Average Estimated Fire Area (km²)')
plt.title('Estimated Fire Area over Time')
plt.show()
```



We can notice that the peak of the Estimated Fire Area in the plot is between 2010 and 2013 in Australia. Let’s narrow down our finding, by plotting the estimated fire area for year grouped together with month.

```
df_new=df.groupby(['Year','Month'])['Estimated_fire_area'].mean()
```

```
df_new.plot(x=df_new.index, y=df_new.values)
plt.xlabel('Year, Month')
plt.ylabel('Average Estimated Fire Area (km²)')
plt.title('Estimated Fire Area over Time')
plt.show()
```



This plot shows that the estimated fire area was on its peak after 2011, April and before 2012. This was the time of maximum wildfire hit in Austrailia.

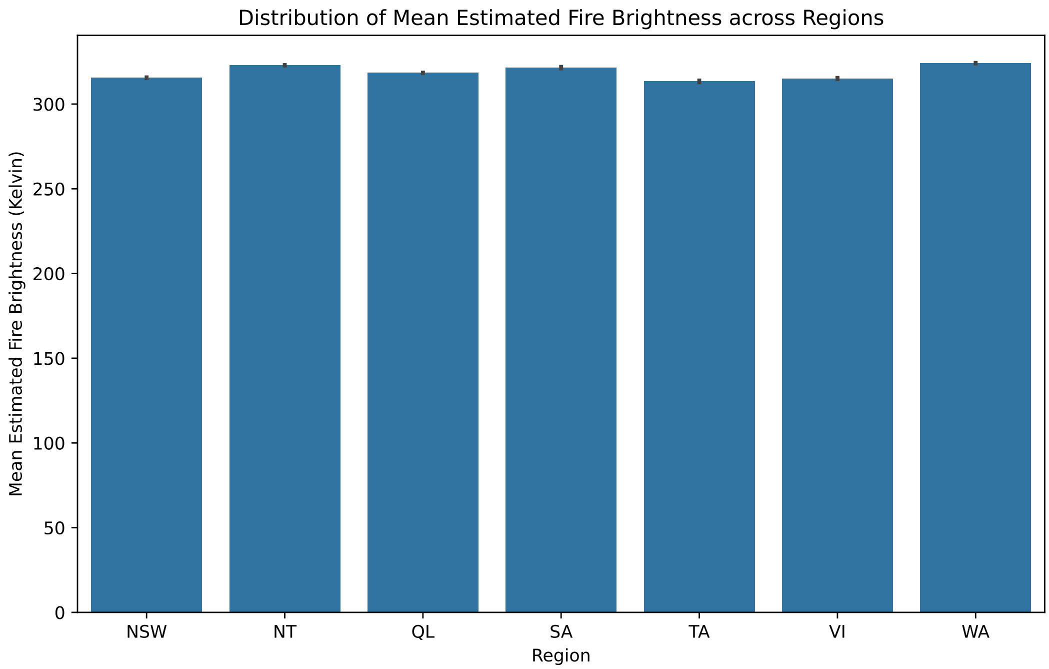
Bar chart

Let's have an insight on the distribution of mean estimated fire brightness across the regions. Before that, let's know the regions mentioned in the dataset.

```
df['Region'].unique()
```

```
array(['NSW', 'NT', 'QL', 'SA', 'TA', 'VI', 'WA'], dtype=object)
```

```
plt.figure(figsize=(10, 6))

sbn.barplot(data=df, x='Region', y='Mean_estimated_fire_brightness')
plt.xlabel('Region')
plt.ylabel('Mean Estimated Fire Brightness (Kelvin)')
plt.title('Distribution of Mean Estimated Fire Brightness across Regions')
plt.show()
```



All the regions hit the 300° Kelvin mean estimated fire brightness.

Pie chart

Let's find the portion of count of pixels for presumed vegetation fires vary across regions.

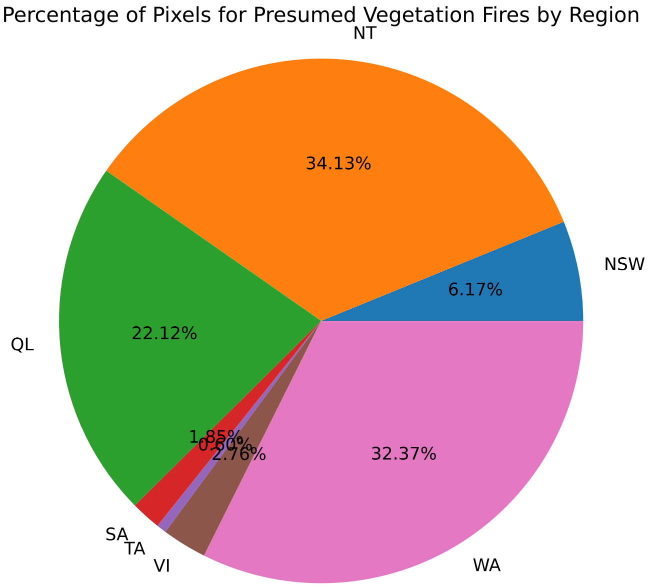
```
plt.figure(figsize=(10,6))

region_count=df.groupby('Region')['Count'].sum()

plt.pie(region_count, labels=region_count.index, autopct='%1.2f%%')
```

```
([<matplotlib.patches.Wedge object at 0x0000028782C1B380>, <matplotlib.patches.Wedge object at 0x0000028
```

```
plt.title('Percentage of Pixels for Presumed Vegetation Fires by Region')
plt.axis('equal')
```

```
(-1.099998843542985, 1.0999999449306184, -1.0999998833294313, 1.0999991346181728)
```

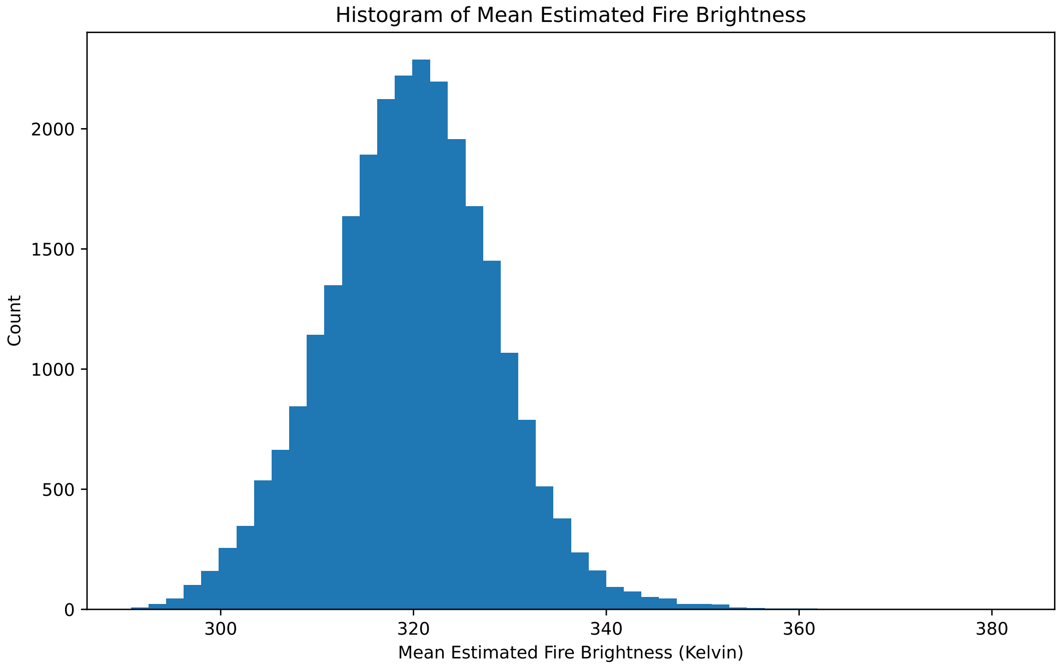
```
plt.show()
```



Histogram

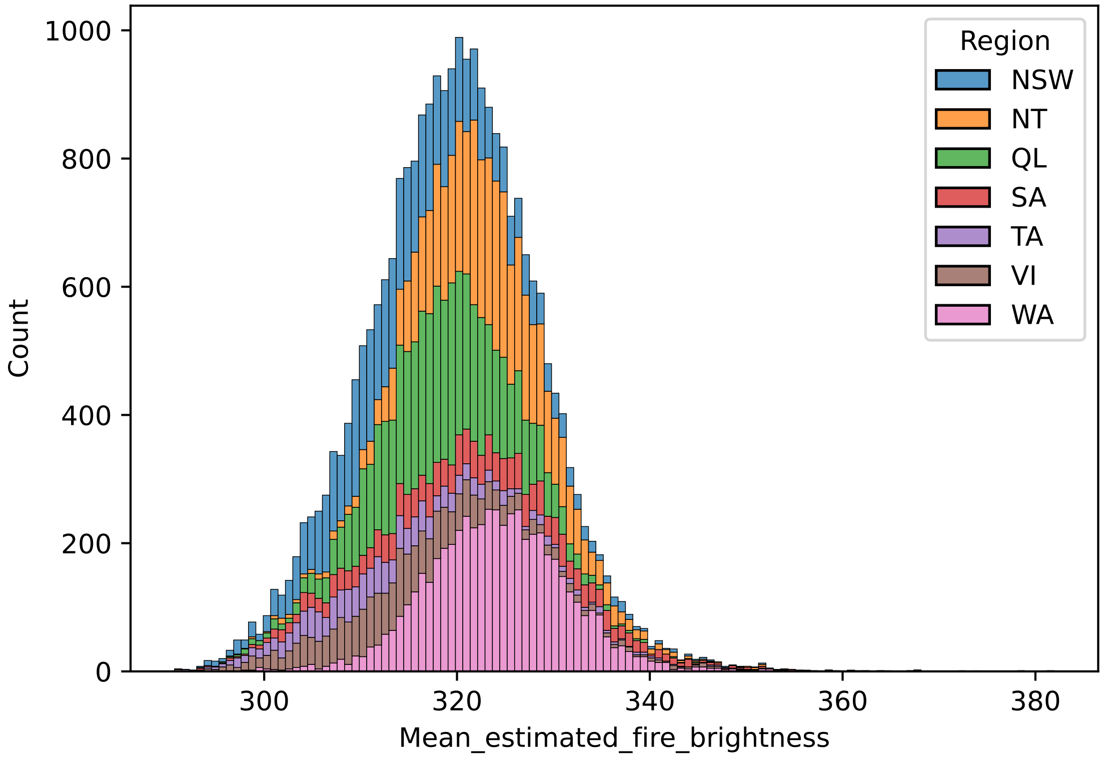
Let’s try to develop a histogram of the mean estimated fire brightness.

```
plt.figure(figsize=(10,6))

plt.hist(x=df['Mean_estimated_fire_brightness'], bins=50)
plt.xlabel('Mean Estimated Fire Brightness (Kelvin)')
plt.ylabel('Count')
plt.title('Histogram of Mean Estimated Fire Brightness')
plt.show()
```



If we need to understand the distribution of estimated fire brightness across regions, We can use the functionality of seaborn and pass region as hue.

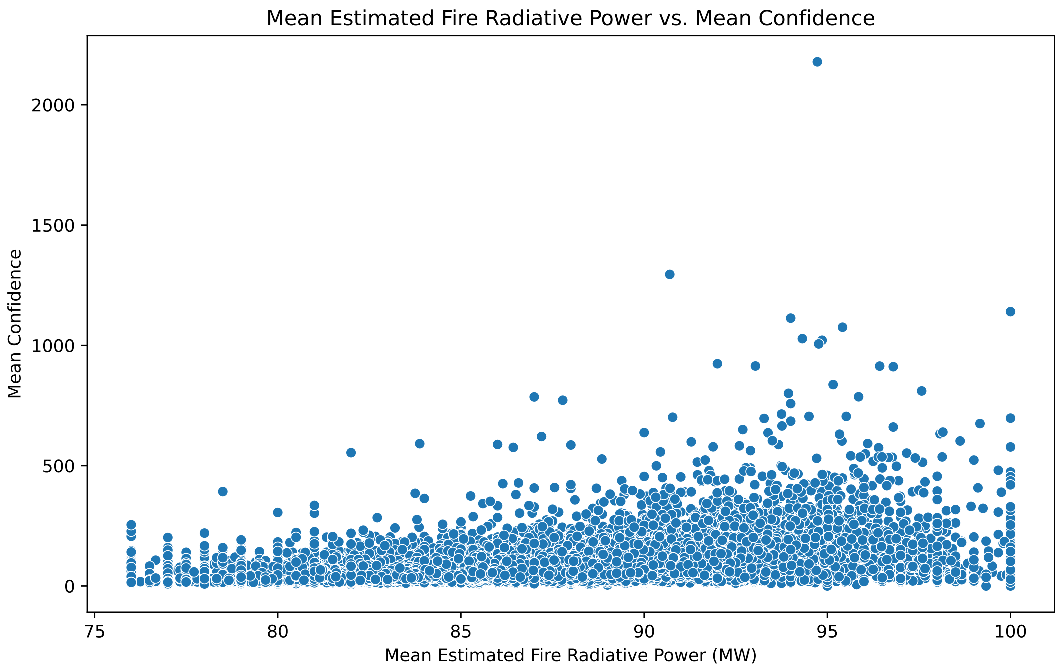
```
a=sbn.histplot(data=df, x='Mean_estimated_fire_brightness', hue='Region', multiple='stack')
plt.show()
```



Scatter Plot

Let's try to find if there is any correlation between mean estimated fire radiative power and mean confidence level.

```
plt.figure(figsize=(10,6))
sbn.scatterplot(data=df, x='Mean_confidence',
y='Mean_estimated_fire_radiative_power')
plt.xlabel('Mean Estimated Fire Radiative Power (MW)')
plt.ylabel('Mean Confidence')
plt.title('Mean Estimated Fire Radiative Power vs. Mean Confidence')
plt.show()
```



Map chart

Let's mark these seven regions on the Map of Australia, creating a dataframe containing the regions, their latitudes and longitudes.
For australia I use [-25, 135] as location to create the map

```
region_data = {'region':['NSW','QL','SA','TA','VI','WA','NT'], 'Lat':[-31.8759835,-22.1646782,-30.53436
               'Lon':[147.2869493,144.5844903,135.6301212,146.6366887,144.6780052,121.0187246,132.55096
reg=pd.DataFrame(region_data)

aust_reg = folium.map.FeatureGroup()

aust_map = folium.Map(location=[-25, 135],zoom_start=4)

# loop through the region and add to feature group
for lat, lng, lab in zip(reg.Lat, reg.Lon, reg.region):
    aust_reg.add_child(
        folium.features.CircleMarker(
            [lat, lng],
            popup=lab,
            radius=5, # define how big you want the circle markers to be
            color='red',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )
```

```
<folium.map.FeatureGroup object at 0x0000028785F53350>
```

```
<folium.map.FeatureGroup object at 0x0000028785F53350>
<folium.map.FeatureGroup object at 0x0000028785F53350>
<folium.map.FeatureGroup object at 0x0000028785F53350>
<folium.map.FeatureGroup object at 0x0000028785F53350>
<folium.map.FeatureGroup object at 0x0000028785F53350>
<folium.map.FeatureGroup object at 0x0000028785F53350>
```

```
# add incidents to map
aust_map.add_child(aust_reg)
```

```
<folium.folium.Map object at 0x00000287857E32C0>
```