pds2136_HW1_submission

February 5, 2020

1 Importing libraries

[1]:
```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import math
from scipy.stats import kde

%matplotlib inline
```

2 Task 1

[2]:
```
data = pd.read_csv("fire_nrt_V1_96617.csv")
proportion = 0.10
```

[3]:
```
data.head()
```

[3]:

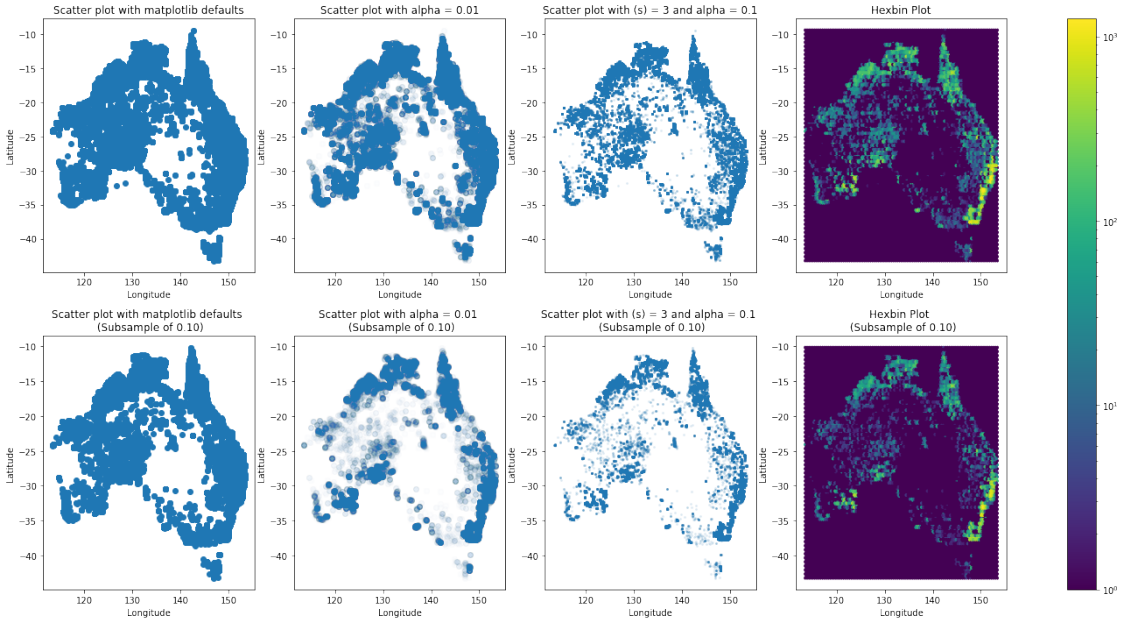
	latitude	longitude	bright_ti4	scan	track	acq_date	acq_time	satellite	instrument	confidence	version	bright_ti5	frp	daynight
0	-42.69706	147.70634	333.6	0.57	0.43	2019-10-01	336	N	VIIRS	n	1.0NRT	293.4	4.0	D
1	-42.26889	147.31104	336.2	0.39	0.44	2019-10-01	336	N	VIIRS	n	1.0NRT	296.9	3.9	D
2	-40.85202	145.38068	327.9	0.46	0.47	2019-10-01	336	N	VIIRS	n	1.0NRT	295.3	3.7	D
3	-42.39329	147.47144	346.9	0.38	0.43	2019-10-01	336	N	VIIRS	n	1.0NRT	294.2	6.2	D
4	-42.69701	147.70584	334.1	0.56	0.43	2019-10-01	336	N	VIIRS	n	1.0NRT	293.5	4.1	D

[4]:
```
x = data["longitude"]
y = data["latitude"]

num_samples = math.floor(proportion * len(x))
idx = np.random.choice(np.arange(len(x)), num_samples)
```

```
x_subsample = x[idx]
y_subsample = y[idx]
```

[5]:
```
fig, axes = plt.subplots(2, 4 , figsize=(20, 10))
# axes.titlesize : "x-large"
# axes.labelsize : "large"
axes[0,0].plot(x, y, 'o')
axes[1,0].plot(x_subsample, y_subsample, 'o')

axes[0,1].scatter(x, y, alpha = 0.01)
axes[1,1].scatter(x_subsample, y_subsample, alpha = 0.01)

axes[0,2].scatter(x, y, s = 3, alpha = 0.1)
axes[1,2].scatter(x_subsample, y_subsample, s = 3, alpha = 0.1)

im = axes[0,3].hexbin(x,y, bins = "log")
im = axes[1,3].hexbin(x_subsample,y_subsample, bins = "log")

axes[0,0].set_xlabel('Longitude')
axes[0,0].set_ylabel('Latitude')
axes[0,0].set_title("Scatter plot with matplotlib defaults")
axes[1,0].set_xlabel('Longitude')
axes[1,0].set_ylabel('Latitude')
axes[1,0].set_title("Scatter plot with matplotlib defaults \n (Subsample of 0.
↪10)")

axes[0,1].set_xlabel('Longitude')
axes[0,1].set_ylabel('Latitude')
axes[0,1].set_title("Scatter plot with alpha = 0.01")
axes[1,1].set_xlabel('Longitude')
axes[1,1].set_ylabel('Latitude')
axes[1,1].set_title("Scatter plot with alpha = 0.01 \n (Subsample of 0.10)")

axes[0,2].set_xlabel('Longitude')
axes[0,2].set_ylabel('Latitude')
axes[0,2].set_title("Scatter plot with (s) = 3 and alpha = 0.1")
axes[1,2].set_xlabel('Longitude')
axes[1,2].set_ylabel('Latitude')
axes[1,2].set_title("Scatter plot with (s) = 3 and alpha = 0.1 \n (Subsample of␣
↪0.10)")

axes[0,3].set_xlabel('Longitude')
axes[0,3].set_ylabel('Latitude')
axes[0,3].set_title("Hexbin Plot")
axes[1,3].set_xlabel('Longitude')
axes[1,3].set_ylabel('Latitude')
```

```
axes[1,3].set_title("Hexbin Plot \n (Subsample of 0.10)")

plt.tight_layout()
fig.colorbar(im, ax=axes.ravel().tolist())
plt.show()
```

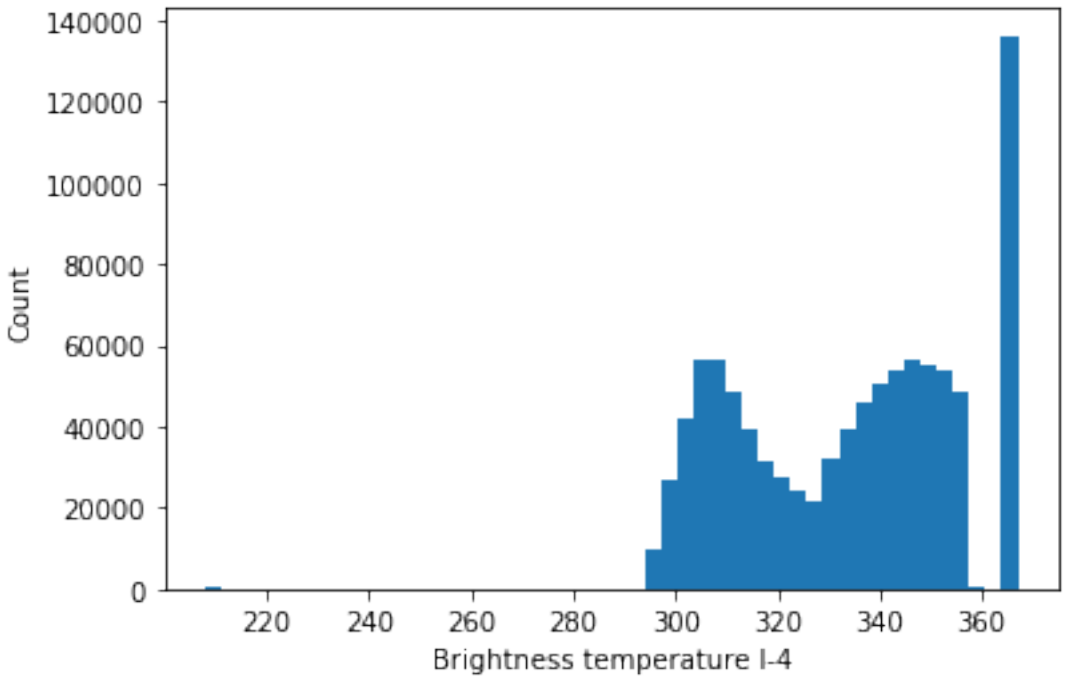


- From the plots it can be seen that the anomalies are spread out across the borders and aren't located in the central area of the continent. The scatter plot with specified alpha value also helps in determining that the central part of the continent has fewer anomalies.
- The Heatmap further shows that the anomalies in the bottom-right (i.e) the South-East is having higher concentration for the anomalies.
- The anomalies seem to be spreading out from the borders towards the central part of the continent.
- The plots in general show the prominanace of anomalies in the South-East region and somewhat on the top-left (i.e) the North-west part of the continent.

3 Task 2

[6]:
```
temp_i4 = data["bright_ti4"]
```

[7]:
```
plt.hist(temp_i4,bins=50)
plt.xlabel('Brightness temperature I-4')
plt.ylabel('Count')
plt.show()
```



[8]:
```
temp = temp_i4[data["bright_ti4"]>=360]
```

[9]:
```
unique_temp_i4 = temp.unique()[0]
print(unique_temp_i4)
```

```
367.0
```

- As visible in the histogram, the point at 367 is having a peek and hence we can identify that 367 is the value of Brightness temperature I-4 which corresponds to the value which is saturated as visible from the histogram.

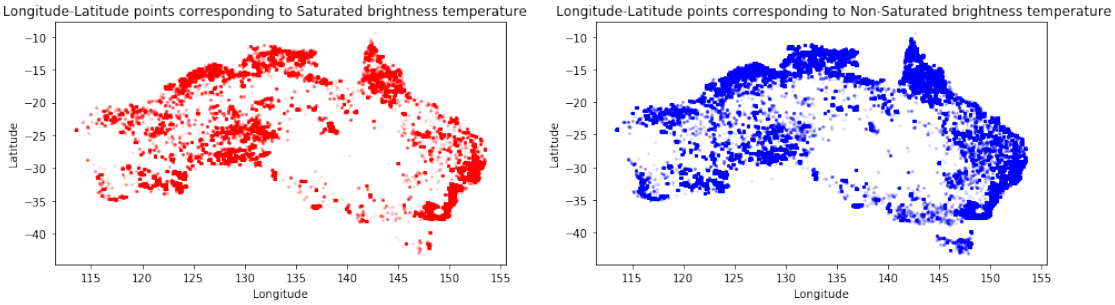
[10]:
```
temp_i4_saturated = temp_i4[data["bright_ti4"] == unique_temp_i4]
temp_i4_not_saturated = temp_i4[data["bright_ti4"] != unique_temp_i4]
```

[11]:
```
data_saturated = data.loc[data["bright_ti4"].isin(temp_i4_saturated)]
data_not_saturated = data.loc[data["bright_ti4"].isin(temp_i4_not_saturated)]
```

[12]:
```
long_saturated = data_saturated["longitude"]
lat_saturated = data_saturated["latitude"]

long_not_saturated = data_not_saturated["longitude"]
lat_not_saturated = data_not_saturated["latitude"]
```

[13]:
```
fig, ax = plt.subplots(1, 2, figsize=(16, 4))
ax[0].scatter(long_saturated, lat_saturated, s=3, alpha = 0.1, color = "red")
```

```
ax[1].scatter(long_not_saturated, lat_not_saturated, s=3, alpha = 0.1, color =␣
↪"blue")

ax[0].set_xlabel('Longitude')
ax[0].set_ylabel('Latitude')

ax[1].set_xlabel('Longitude')
ax[1].set_ylabel('Latitude')

ax[0].set_title("Longitude-Latitude points corresponding to Saturated␣
↪brightness temperature       ")
ax[1].set_title("      Longitude-Latitude points corresponding to Non-Saturated␣
↪brightness temperature")

plt.show()
```



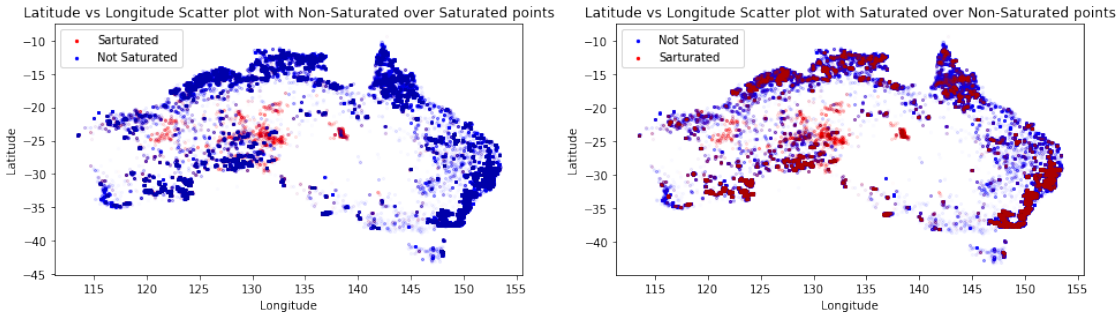
The above two plots differentiate the points having Saturated Brightness temperature (i.e) the one we are certain of a fire from the non-saturated ones.

The difference that are visible from the above plots

- The non-saturated brightness points seems to be more concentrated in the South-East Region.
- The concentration of non-saturated points in general is more than that of the saturated ones. However, the concentration of the saturated ones seems to be more in the western side of the central part of the continent.

[14]:
```
fig, ax = plt.subplots(1, 2, figsize=(16, 4))

ax[0].scatter(long_saturated, lat_saturated, s=5, alpha = 0.01, color = "red" ,␣
↪label = "Sarturated")
ax[0].scatter(long_not_saturated, lat_not_saturated, s=5, alpha = 0.01, color =␣
↪"blue", label = "Not Saturated")
leg = ax[0].legend()
ax[0].set_xlabel('Longitude')
ax[0].set_ylabel('Latitude')
```

```
for lh in leg.legendHandles:
    lh.set_alpha(1)

ax[1].scatter(long_not_saturated, lat_not_saturated, s=5, alpha = 0.01, color =␣
 ↪"blue", label = "Not Saturated")
ax[1].scatter(long_saturated, lat_saturated, s=5, alpha = 0.01, color = "red" ,␣
 ↪label = "Sarturated")
leg = ax[1].legend()
ax[1].set_xlabel('Longitude')
ax[1].set_ylabel('Latitude')

ax[0].set_title("Latitude vs Longitude Scatter plot with Non-Saturated over␣
 ↪Saturated points")
ax[1].set_title("Latitude vs Longitude Scatter plot with Saturated over␣
 ↪Non-Saturated points")
for lh in leg.legendHandles:
    lh.set_alpha(1)
plt.show()
```

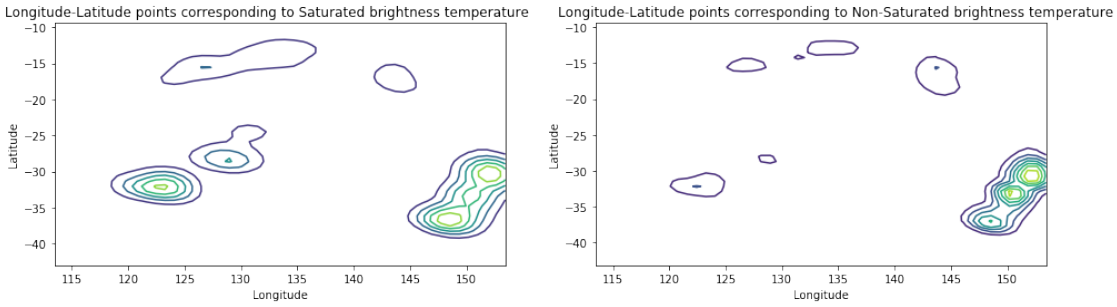


The two subplots formed by interchanging the order of plotting impacts the results a lot and it really is dependent on the order of plotting to get the information out from the plots.

- When the saturated were plotted before non-saturated ones they were getting overlapped and it was not quite possible to differentiate the plots. When the order was changed the points came out and it was possible to notice that the non-saturated points are surrounding the saturated ones. Thus, it can be said that the saturated ones might be the epicenter for the fire points which spread out and become non saturated as it further propogates.

[16]:
```
nbins = 50
x,y = long_saturated, lat_saturated
k = kde.gaussian_kde([x,y])
xi, yi = np.mgrid[x.min():x.max():nbins*1j, y.min():y.max():nbins*1j]
zi = k(np.vstack([xi.flatten(), yi.flatten()]))

xj,yj = long_not_saturated, lat_not_saturated
```

```
kj = kde.gaussian_kde([xj,yj])
xii, yii = np.mgrid[xj.min():xj.max():nbins*1j, yj.min():yj.max():nbins*1j]
zii = kj(np.vstack([xii.flatten(), yii.flatten()]))

fig, ax = plt.subplots(1, 2, figsize=(16, 4))

ax[0].contour(xi, yi, zi.reshape(xi.shape))
ax[0].set_xlabel('Longitude')
ax[0].set_ylabel('Latitude')

ax[1].contour(xii, yii, zii.reshape(xii.shape))
ax[1].set_xlabel('Longitude')
ax[1].set_ylabel('Latitude')

ax[0].set_title("Longitude-Latitude points corresponding to Saturated␣
 ↪brightness temperature      ")
ax[1].set_title("      Longitude-Latitude points corresponding to␣
 ↪Non-Saturated brightness temperature")

plt.show()
```



The above 2 contour plots for the saturated and unsaturated points can easily show the concentration of Saturated and Non-Saturated points.

These are better than the 2 seperated scatter plots as they easily differentaiate the concentration while also removing the ambiguity of determining by the darkness of the colour points.

- The contours thus, show that the saturated points have 2 different major concentration areas in South-East, central-South on the westerrn side.
- While the non-saturated points have a major concentration near the South-East region which was not that clear using the 2 scatter plots plotted earlier
- This concentrations mentioned above are with respect to the total number of points belonging to each category.