

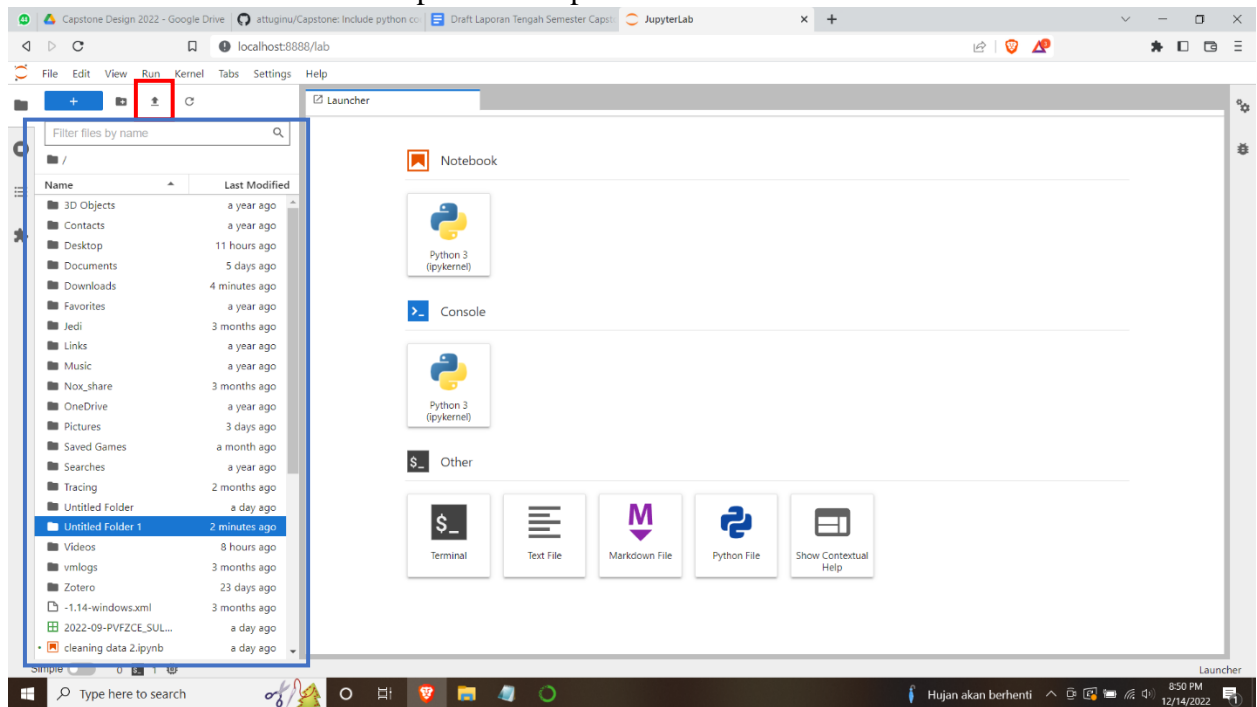
# ***Computational Geosciences :***

## **Aplikasi PCA dan K-Means Clustering pada Data Geokimia Batuan di Wilayah Sulawesi dan Sangihe**

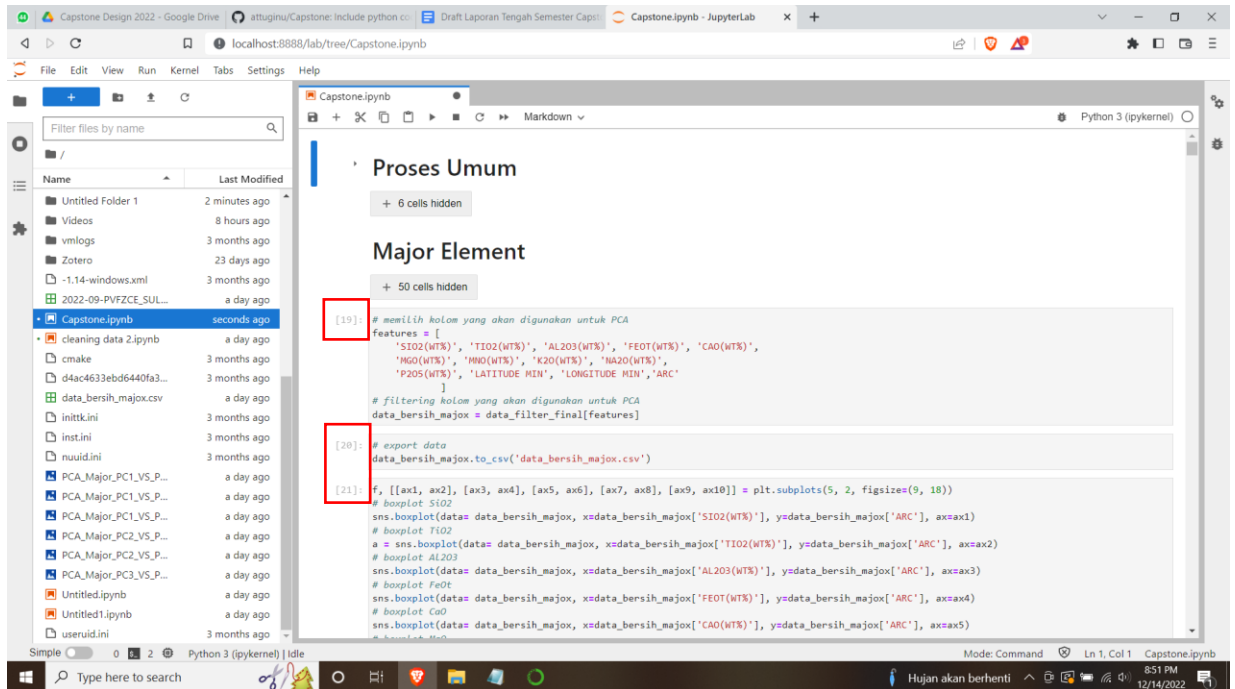
### **Petunjuk Penggunaan Kode**

Panduan ini menggunakan JupyterLab dalam menjalankan kode dan menggunakan Anaconda yang sudah membawa packages yang diperlukan untuk menjalankan program ini, jika tidak menggunakan Anaconda pastikan library numpy, pandas, matplotlib dan seaborn sudah terinstall pada environment JupyterLab. Variable yang dituliskan di panduan ini dapat diubah sesuai keperluan.

1. Setelah menjalankan JupiterLab upload script dengan tombol di kotak merah, jika upload berhasil maka file akan muncul pada folder pada sisi kiri dalam kotak biru



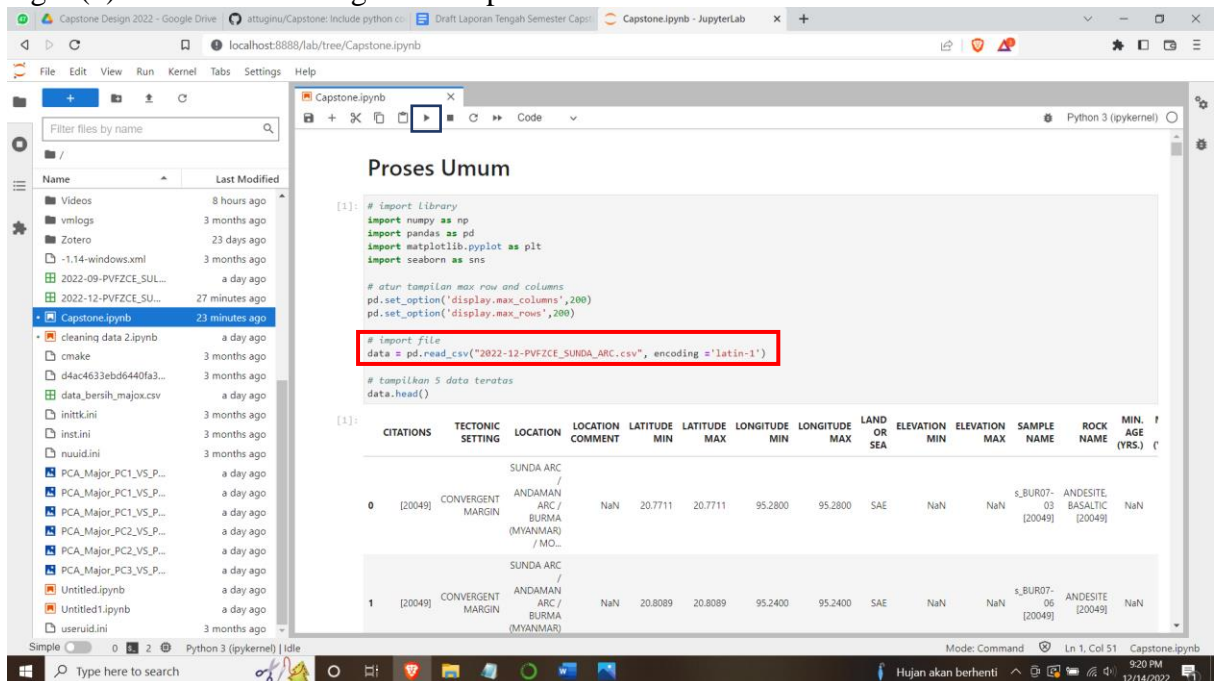
2. Klik 2 kali pada file untuk membuka kode, akan terlihat tampilan sebagai berikut. Tekan cell hidden untuk menampilkan seluruh kode, upload data geokimia dengan langkah yang sama pada langkah 1. Angka pada kotak merah merupakan nomor cell.



- Ubah script pada kotak merah menjadi nama file dari data yang akan digunakan misal nama file data geokimia adalah 2020\_10\_Pacific\_Ocean.csv maka ganti script menjadi  

```
data = pd.read_csv("2020_10_Pacific_Ocean.csv", encoding='latin-1')
```

Setelah itu jalankan cell dengan tombol pada kotak biru untuk memastikan data terbaca. setelah cell ini terdapat cell dengan kode untuk membaca hal tertentu, baca script dengan tagar (#) untuk melihat fungsi dari tiap cell



4. Setelah memastikan data terbaca proses data terlebih dahulu. Untuk memproses data FeO menjadi FeO total jalankan cell nomor 7 seperti pada gambar. Jalankan cell 8 jika ingin memastikan data sudah diproses dengan benar, akan muncul tabel dibawah cell 8.

```
[7]: # mengisi nilai FeOt dengan nilai
# : Fe2O3*0.899
data_vol_land_Fe2O3 = data_vol_land.copy() # menghindari masalah copy
data_vol_land_Fe2O3['FEOT(WT%)'].fillna(value= data_vol_land_Fe2O3['FE2O3(WT%)'] * 0.899, inplace=True)
```

```
[8]: # tampilkan deskripsi data
data_vol_land_Fe2O3.describe()
```

	LATITUDE MIN	LATITUDE MAX	LONGITUDE MIN	LONGITUDE MAX	ELEVATION MIN	ELEVATION MAX	ERUPTION DAY	ERUPTION MONTH	ERUPTION YEAR	DRILL DEPTH MIN	DRILL DEPTH MAX	SiO2(WT%)	TiO2(WT%)
count	531.000000	531.000000	531.000000	531.000000	40.000000	44.000000	0.0	0.0	8.000000	0.0	0.0	344.000000	340.000000
mean	-0.479562	-0.510824	121.757052	128.470423	796.275000	811.931818	NaN	NaN	1943.625000	NaN	NaN	55.811453	0.826979
std	2.966091	3.105084	4.131276	85.249530	518.933571	528.364497	NaN	NaN	47.551288	NaN	NaN	7.648432	0.291063
min	-6.000000	-6.000000	100.240300	100.240300	0.000000	0.000000	NaN	NaN	1900.000000	NaN	NaN	42.280000	0.100000
25%	-3.100000	-3.500000	119.880000	120.000000	330.000000	400.000000	NaN	NaN	1900.000000	NaN	NaN	50.215000	0.670000
50%	0.500000	0.748430	120.700000	121.300000	832.000000	832.000000	NaN	NaN	1933.000000	NaN	NaN	53.535000	0.830000
75%	1.360890	1.470000	124.847920	125.000000	1278.000000	1286.500000	NaN	NaN	1994.000000	NaN	NaN	59.357500	0.970000
max	4.800000	4.800000	125.900000	125.700000	1708.000000	1762.000000	NaN	NaN	1995.000000	NaN	NaN	76.500000	2.450000

Lalu jalankan kode pada cell 9 untuk menghilangkan data yang kosong

5. Jika ingin memfilter data berdasarkan lokasi buka cell 11, 13 dan 15 dan masukkan lintang dan bujur sesuai dengan lokasi yang ditentukan contoh pada cell 13 data yang akan digunakan memiliki lintang lebih dari 2 dan bujur kurang dari 120 jika ingin menggunakan area dengan lintang 3 – 5 LU dan bujur 130 – 140 BT maka ganti script menjadi

```
outlier = (data_vol_land_Fe2O3_majox_loc['LATITUDE MIN'] > 3) &
          (data_vol_land_Fe2O3_majox_loc['LATITUDE MAX'] < 5) &
          (data_vol_land_Fe2O3_majox_loc['LONGITUDE MIN'] > 130) &
          (data_vol_land_Fe2O3_majox_loc['LONGITUDE MAX'] > 140)
```

```
[13]: # menghilangkan satu data yang nampaknya outlier
# yakni di koordinat lintang +2 dan bujur 119 BT (kiri atas)
outlier1 = (data_vol_land_Fe2O3_majox_loc['LATITUDE MIN'] > 2) & (data_vol_land_Fe2O3_majox_loc['LONGITUDE MIN'] < 120)

# filtering
data_majox = data_vol_land_Fe2O3_majox_loc.loc[~outlier1]

# deskripsi data
data_majox.describe()
```

Jika tidak perlu melakukan filtering lokasi maka pada cell 9 dapat langsung melakukan export data dengan menambahkan script berikut di bagian paling bawah atau menambahkan cell baru setelah cell 9

```
features = [  
    'SIO2(WT%)', 'TIO2(WT%)', 'AL2O3(WT%)', 'FEOT(WT%)', 'CAO(WT%)',  
    'MGO(WT%)', 'MNO(WT%)', 'K2O(WT%)', 'NA2O(WT%)',  
    'P2O5(WT%)', 'LATITUDE MIN', 'LONGITUDE MIN', 'ARC'  
]  
data_vol_land_majox_loc = data_vol_land_majox[features]  
data_vol_land_majox_loc.to_csv('nama file.csv')
```

6. Pemodelan data dapat dilakukan dengan cell 22 dan seterusnya, script dapat diubah sesuai dengan kebutuhan.