**Ejecución algoritmo Kmeans en paralelo**

**Requisitos previos**

* Python 3.6
* Programa anaconda navigator para ejecutar Júpyter Notebook

Librerias necesarias para ejecución:

numpy

matplotlib

matplotlib

pandas

* Archivos .py necesarios

from KMeansBase import KMeansBase

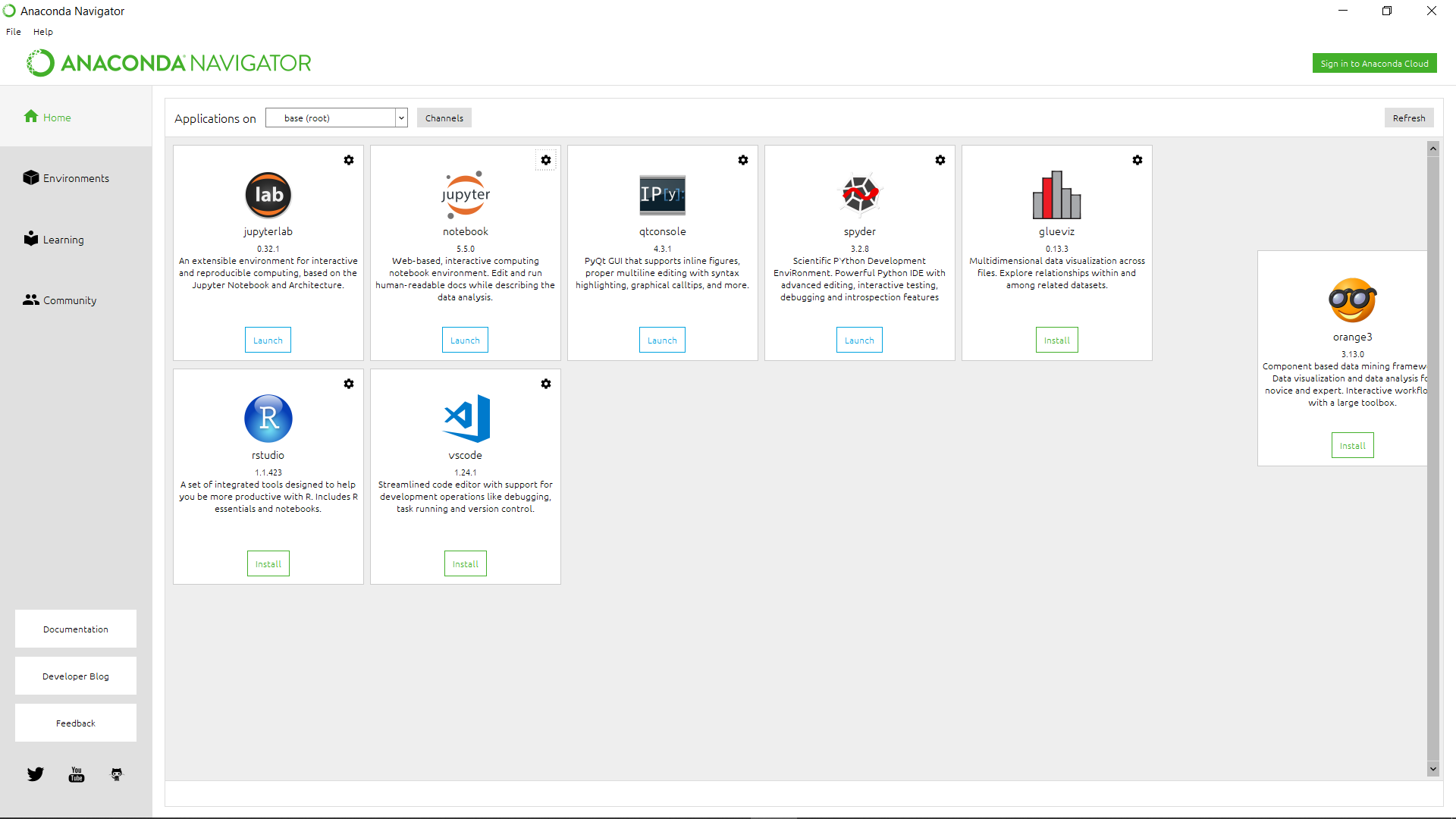
from KMeansPP import KMeansPP

from ScalableKMeansPP import ScalableKMeansPP

para instalar estas librerias ejecutar en la consola **pip install <NOMBRE DE LA LIBRERIA>**

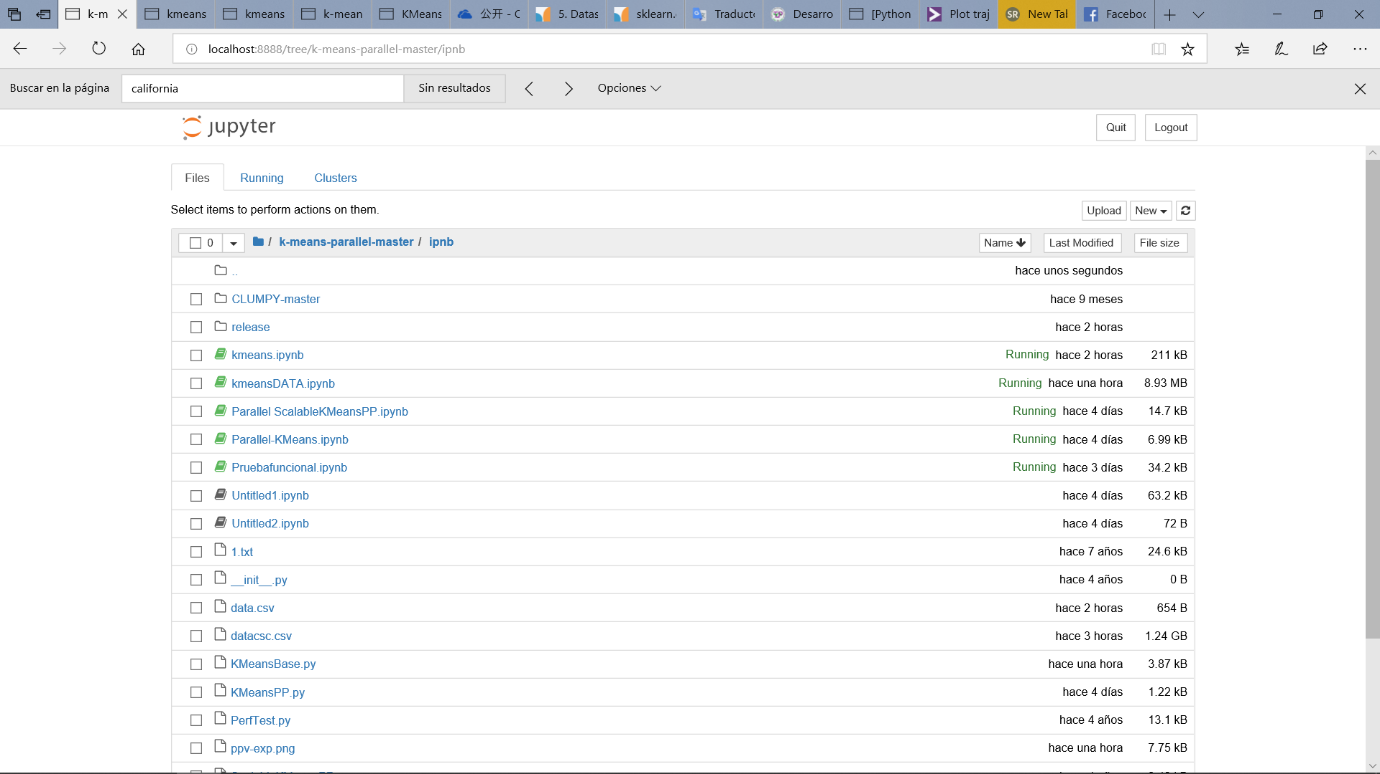
**Ejecución:**

**Abrir el programa Anaconda navigator y hacer click en el icono de jupyter.**



Se abrirá una aplicación web en la siguiente dirección <http://localhost:8888/>

Abrir el archivo [kmeansDATA.ipynb](http://localhost:8888/notebooks/k-means-parallel-master/ipnb/kmeansDATA.ipynb)



En el archivo la función principal es **accuracy\_spam()**

En esta función el método **pandas.read\_csv('spambase.csv')**  recibe un archivo con extensión .csv en el formato del archivo de ejemplo: **spambase.csv** ubicado en la carpeta del proyecto.

La función hace uso del archivo KMeansPP.py el cual se encuentra en la carpeta del proyecto.

**OUTPUT DEL ARCHIVO KMeansPP.py**

min\_location = para averiguar a qué grupo pertenece cada punto de datos y agruparlos en el cluster **. j\_values**

centroids = centroide

j\_values = CLUSTER RESULTANTE

EL Algoritmo devuelve el CLUSTER RESULTANTE por cada iteración.

