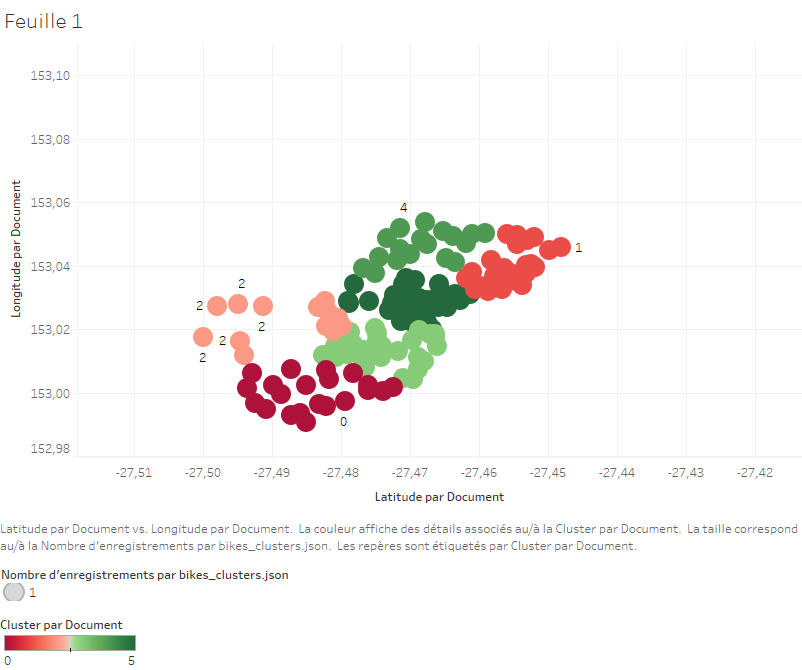
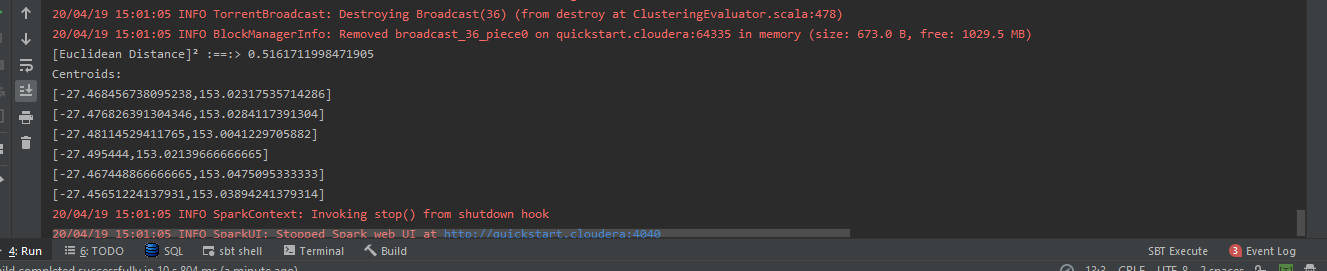
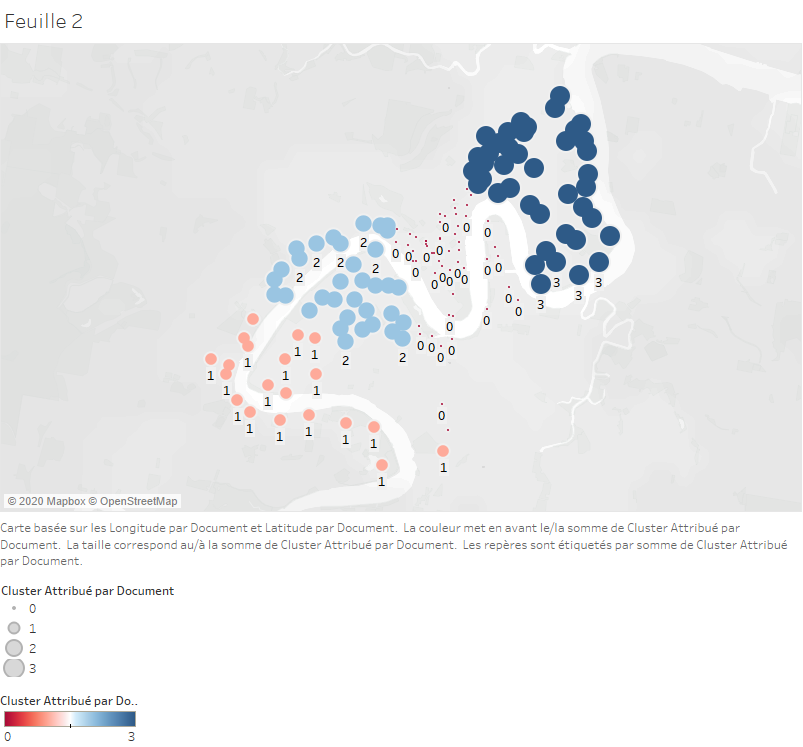
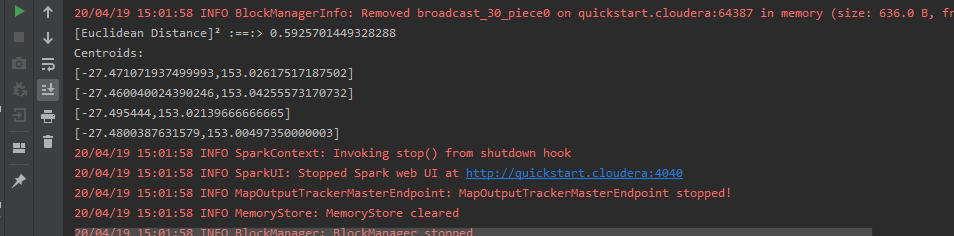
**DATA TEST SUPPORT**

Answered by Oussama DEBBOU :   
data visualization according to the number of clusters “K”   
  
K = 6 :

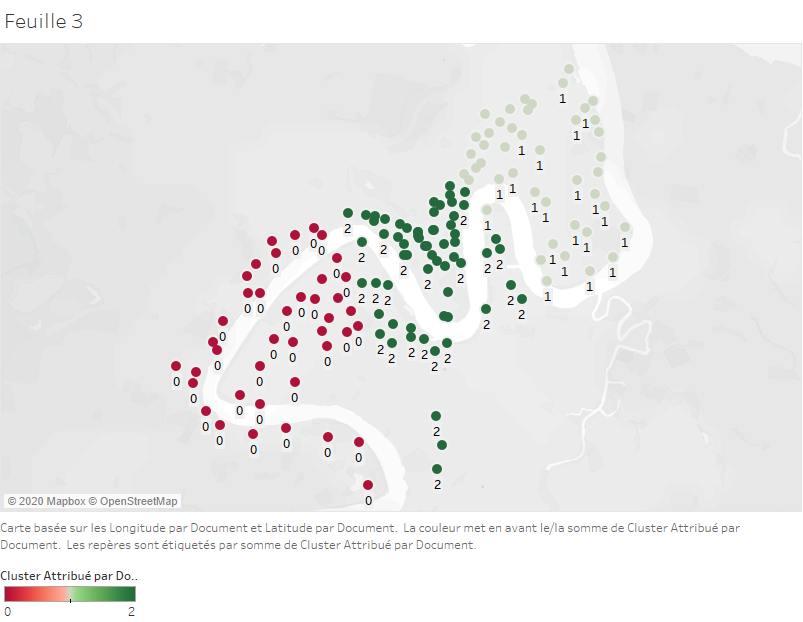


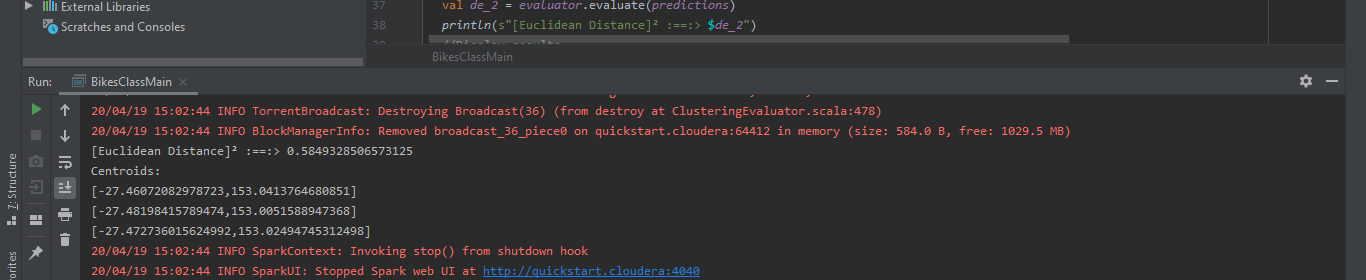


K = 4 :   
  


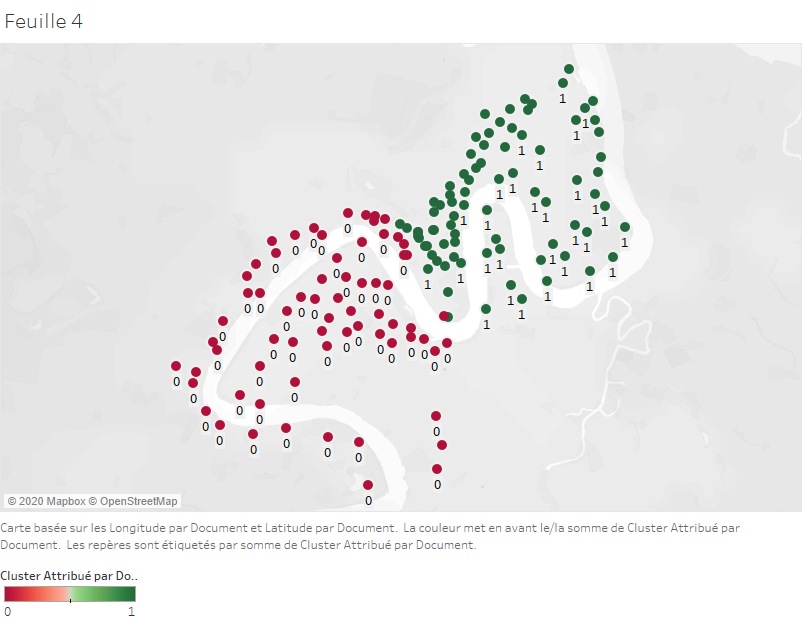


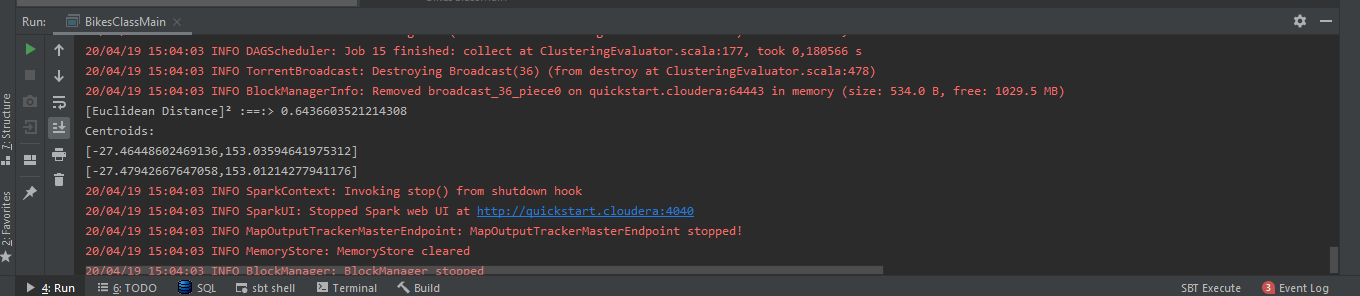
K = 3 :





K = 2 :





Conclusion :

After long reflection and visualization of the graphs generated by Tableau, the closest model in my opinion is either 6 or 4, however, in 6 the clusters are not very well separated so I will look for model 4.

For the execution in my case it is done locally on a Spark cluster.  
  
However if the data was stored elsewhere for example on hdfs:  
we would have made a spark-submit which will point to the data in hdfs after generating the jar “Bike.jar”

spark-submit \

--class BikesClassMain \

--master local/YARN/MESOS \

--deploy-mode cluster \

Bike.jar