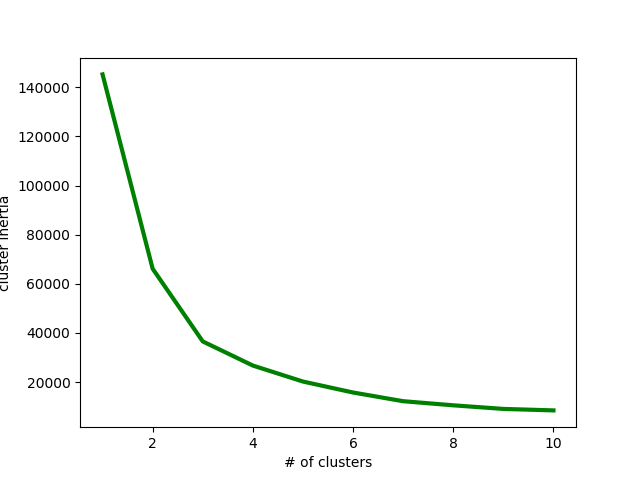
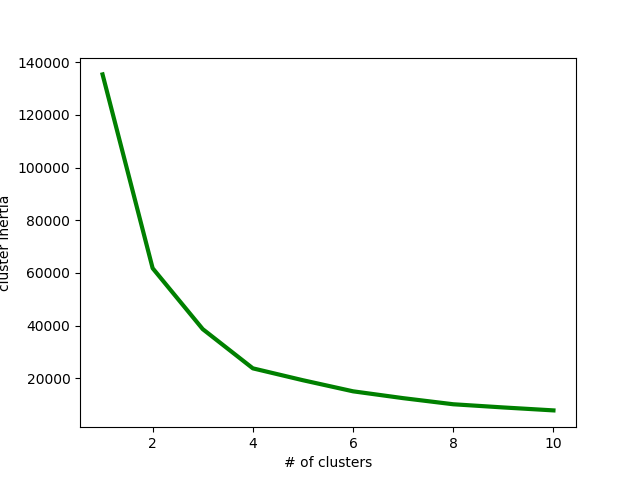
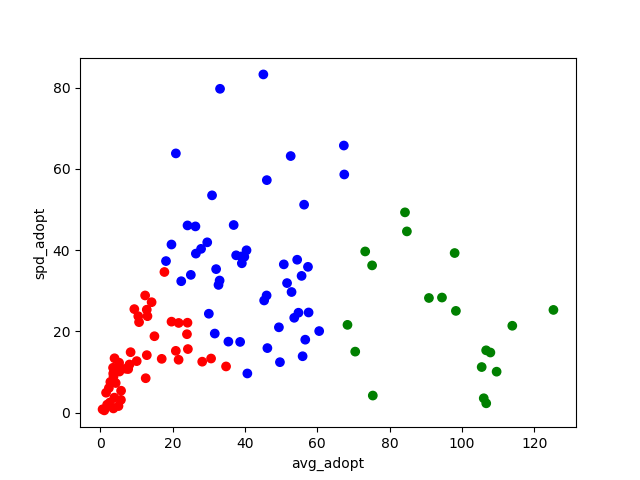
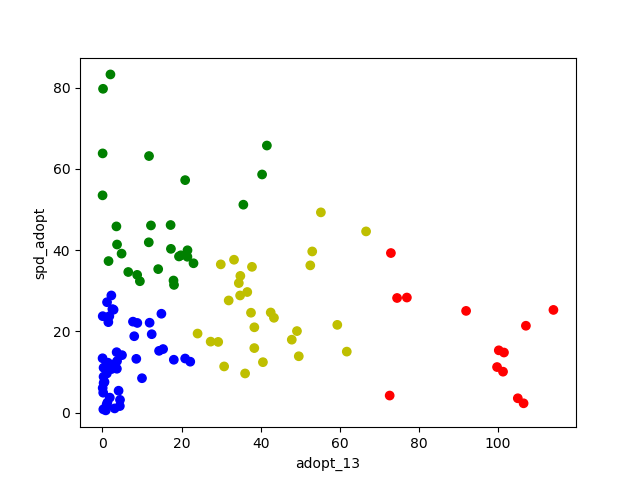
**Clustering with DBSCAN & KMeans**

1. **Summary**
   1. Dataset = A4data(mobile broadband adoption rate)
   2. features: avg\_adopt, spd\_adopt, adopt\_13
   3. 13년 adoption rate를 사용한 dataframe을 df\_13,  
      average adoption rate을 사용한 dataframe을 df\_avg라 칭함.
2. **KM TEST**

****

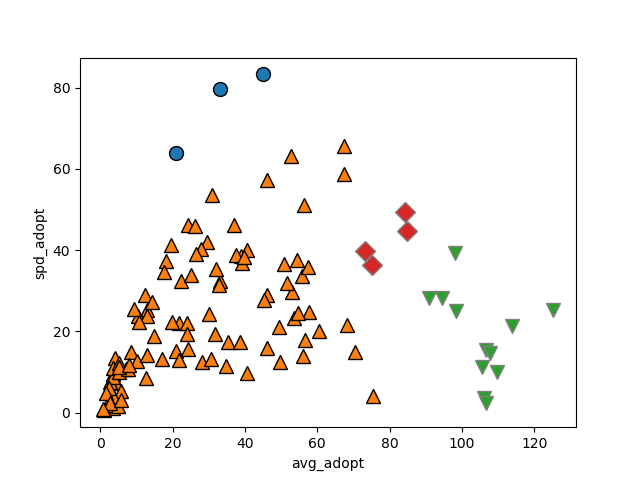
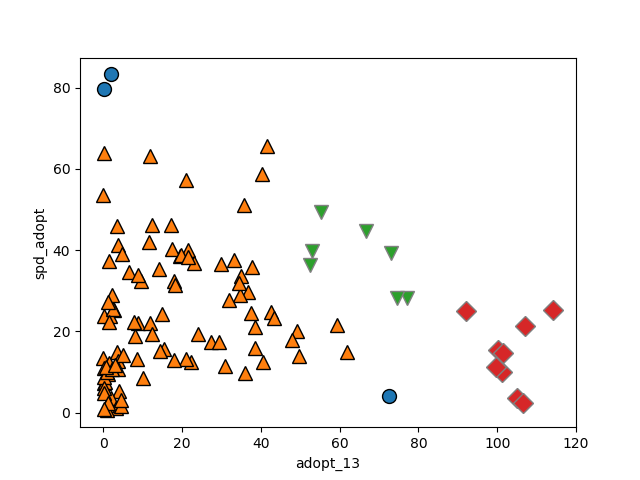
* 좌측은 df\_13, 우측은 df\_avg의 데이터의 cluster개수 별 inertia를 측정한 그래프
* df\_13의 경우 cluster가 4개부터 효과가 현저히 감소
* df\_avg의 경우 cluster가 3개부터 효과가 현저히 감소

1. **K-means clustering**



* 2번의 결과에 따라 df\_13은 cluster를 4개로, df\_avg는 cluster를 3개로 구분함.
* df\_13
  + 파랑: 초기단계의 adoption rate, 느린 성장속도
  + 초록: adoption rate는 초기단계이지만 빠른 성장이 예상되는 그룹
  + 노랑: 중간단계의 adoption rate그룹
  + 빨강: 이미 성숙한 adoption rate라 성장속도가 높을 수 없음.
* df\_avg
  + 빨강: 발전속도가 느린 탓에 꾸준히 낮은 adoption rate를 보이는 그룹
  + 파랑: 비교적 높은 발전속도로 평균 adoption rate가 중간단계인 그룹
  + 초록: 이미 발전하여 높은 성장속도를 볼 수 없는 그룹

1. **DBSCAN clustering**



* 몇 차례 직접 df를 구분해보며 적정한 eps와 min\_sample값을 추정함
* 그룹을 위한 최소 조건은 3개의 샘플이며(min\_sampe=3), sample당 거리는 14로 설정(eps=14)
* df\_13
  + 주황: 초기단계의 adoption rate, 속도에 따른 구분은 되지 않음
  + 초록: 높은 adoption rate와 높은 성장속도를 보이는 그룹
  + 빨강:
  + 파랑: 그룹에 포함되지 못한 국가
* df\_avg
  + 주황: 초기-중기단계의 adoption rate, 속도에 따른 구분은 되지 않음
  + 빨강: 높은 성숙도를 가졌음에도 빠른 성장속도를 보이는 그룹
  + 초록: 이미 성숙한 adoption rate를 가진 그룹으로 성장속도가 높기 힘듦
  + 파랑: 성장속도가 아주 빠른 그룹

1. **Conclusion**

* Country의 feature별 분포가 부채꼴 형태로 특별한 도형의 양상을 띄지 않음.
* 즉 circle형태로 나눠도 특별히 문제가 생기지 않음.
* 때문에 dbscan보다 k-means가 보다 정확한 구분을 하고 있는 것으로 보임.