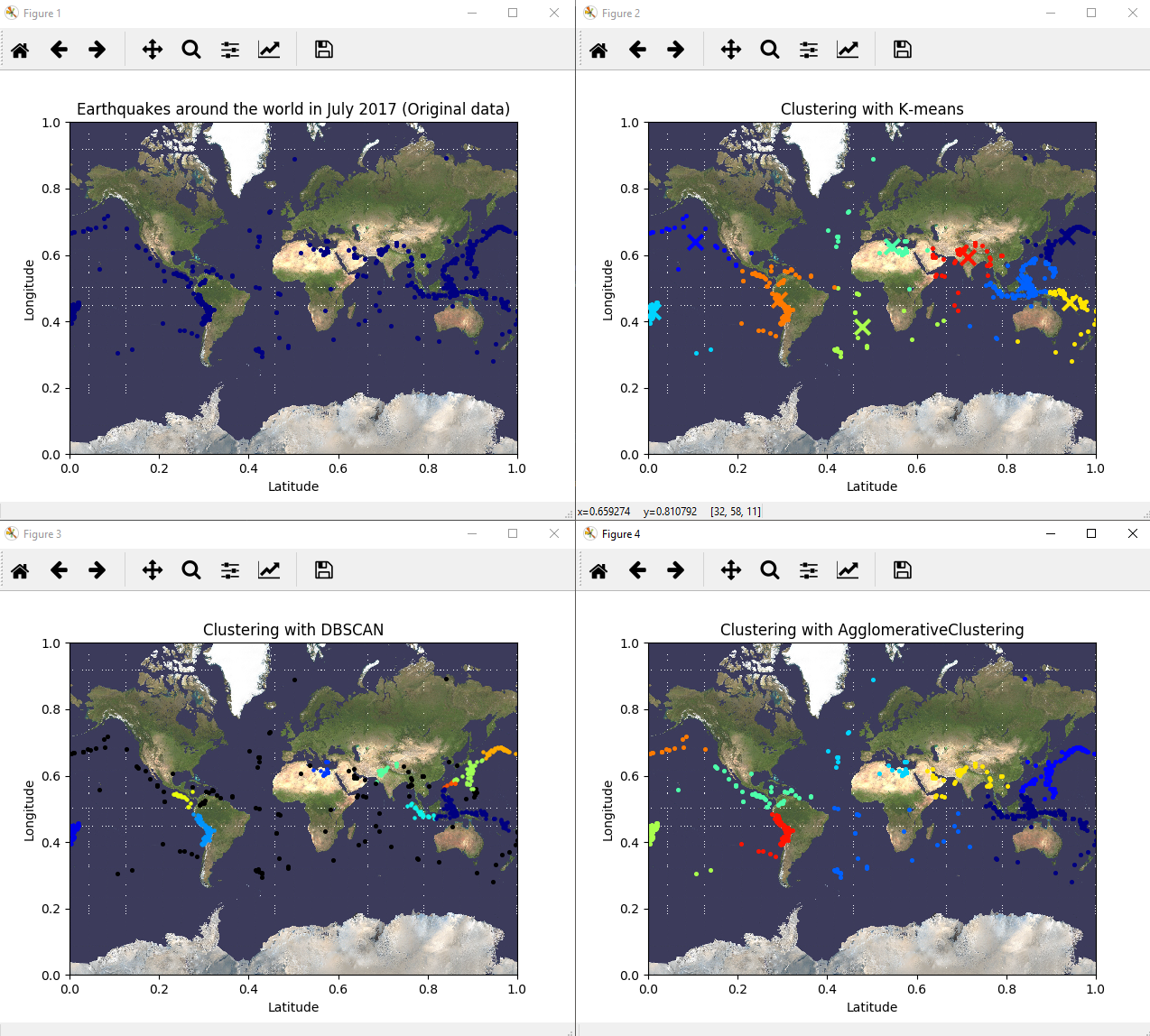
Johdatus tekoälyyn Harjoitus 4 ryhmä 7

Tehtävä 1:

  
**Kuva 1: Klusterit, silhouette score = 9 ja dbscan klusterien määrä 10.**

**Kysymys 1: Mitä eroavaisuuksia käytetyillä klusterointimenetelmillä on kuvaajien perusteella?**DBscan tuottaa poikkeavimman tuloksen, mikä johtuu siitä, että se jättää outlierit pois klustereista (mustat pisteet). Klusterit ovat myös huomattavasti pienempiä. Kokoava hierarkkinen klusterointi ja k-means saavat klustereiden määrän silhouette scorena, joka oli suorituskerrasta riippuen 8 tai 9. Näiden kahden menetelmän ero on ainakin siinä, että k-means:lla klusteroidessa hyvinkin lähellä toisiaan olevat datapisteet voivat joutua eri klustereihin (kuvassa esim. oranssi ja vihreä piste Etelä-Amerikan itäpuolella).

**Kysymys 2: Mikä menetelmistä on paras, kun halutaan selvittää tarkasti, missä maanjäristyskeskukset sijaitsevat:**

K-means sopii tähän tarkoitukseen parhaiten, koska se tekee klusterointia keskiarvojen perusteella. Tässä tapauksessa lopulliset klusterikeskiarvot ovat maanjäristyskeskuksia ja ne on merkattu karttaan rasteilla.