Hafta 3 – Spark Modül Projesi

Teslim Tarihi: 12.05.2024 23.59

Bu projeyi tamamlayarak PySpark ile veri işleme konusunda tecrübe kazanmayı amaçlıyoruz.

1. Kişisel bilgisayarınıza ***Jupyter Notebook*** kurulumu yapın.
2. *Spark-Installation.pptx* dokümanından faydalanarak kişisel bilgisayarınıza PySpark kurulumu yapın (Kaggle, Google Colab gibi platformları da kullanabilirsiniz)
3. *PySpark\_practice\_notebook.*ipynb dosyasını Jupyter Notebook ile açtığınızda, içerisinde örnek bir veri seti, diyagram ve örnek veri işleme adımları göreceksiniz. Dokümanı inceleyerek ve sizin için önceden hazırlanmış kodları çalıştırarak anlamaya çalışın.
4. Notebook dosyasının en altındaki “Case Studies” kısmını tamamlayın (**100p**)
   1. Assignment 1: Jacket sales per region (50p)
      * Spark SQL çözümü (25p)
      * PySpark çözümü (25p)
   2. Assignment 2: Maximum turnover of the retailer regions (50p)
      * Spark SQL çözümü (25p)
      * PySpark çözümü (25p)

**Beklenen dosyalar:**

1. Kodları içeren Jupyter Notebook dosyası (ipynb formatında)