
DÖNEM PROJESİ

PYSPARK KULLANARAK SOSYAL MEDYA

UYGULAMA KULLANICILARININ ANALİZİ

Bu projede, sizden randomuser.me/api adlı açık kaynak API'sini kullanarak bir sosyal medya uygulamasının kullanıcı verilerini 10 dakika boyunca 2 saniye aralıklarla alan bir Python betiği geliştirmenizi istiyoruz (10 dakika boyunca GET HTTP istekleri göndermeniz gerekmektedir). Bu veriler, Pandas kütüphanesi kullanılarak Parquet dosyası olarak yazılmalıdır. Daha sonra, aşağıdaki soruları cevaplamak için PYSPARK kullanarak başka bir Python betiği geliştirin. Her soru için lütfen Spark SQL ve DataFrame API kodu geliştirin.

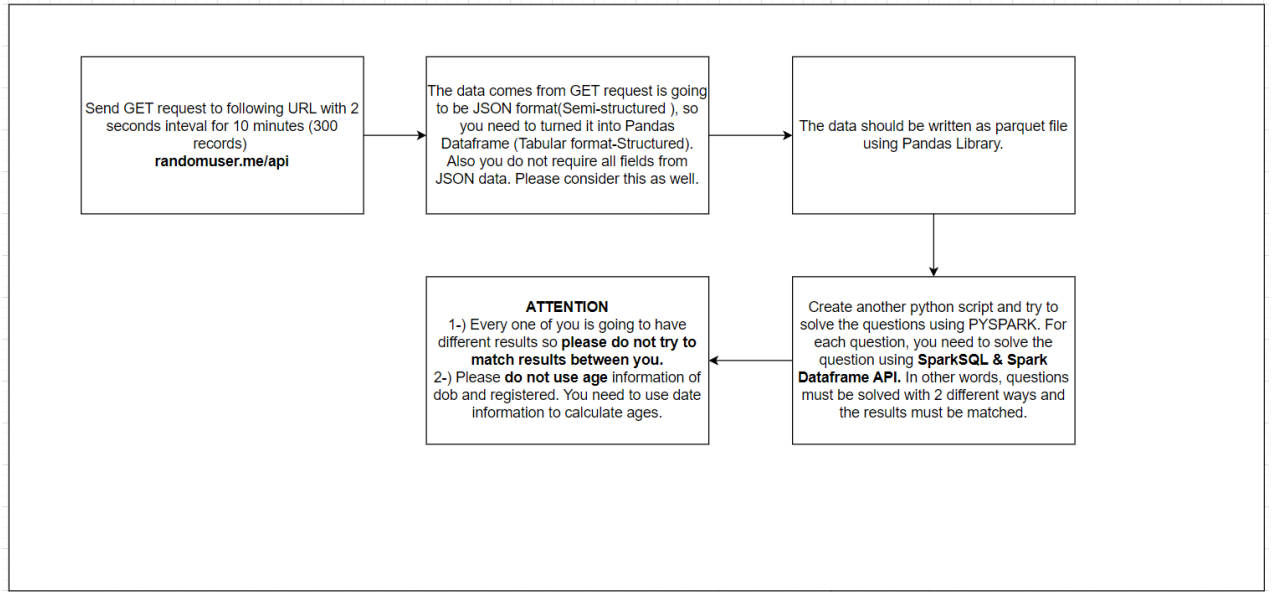
In this project, we would like you to develop python script that get user's data of social media application using an open-source API called randomuser.me/api with 2 seconds interval for 10 minutes (You need to send GET HTTP requests for 10 minutes). This data should be written as parquet file using **Pandas library**. After that develop another python script using **PYSPARK** in order to answer the following questions. For each question please develop **Spark SQL & DataFrame API** code.

Not: Her öğrencinin sonuçları farklı olacaktır, bu yüzden sonuçlarınızı birbirinizle karşılaştırmaya çalışmayın.

Dikkat: Lütfen doğum tarihi (dob) ve kayıt tarihi (registered) bilgilerindeki yaş bilgilerini kullanmayın. Yaşları hesaplamak için "date" bilgisini kullanmanız gerekmektedir.

Note: Every student will have different results, so please **do not try** to match results between you.

Attention: Please **do not use** age information of dob (date of birth) and registered. You need to use "date" to calculate ages.



Talimatlar:

1. Erkek ve kadın kullanıcıların yaş ortalamasını bulun. Ayrıca, sosyal medya uygulamasını ortalama ne kadar süredir kullandıklarını bulun.
2. Ükelere göre erkek ve kadın kullanıcıların yaş ortalamasını bulun. Ayrıca, sosyal medya uygulamasını ortalama ne kadar süredir kullandıklarını bulun.
3. Ükelere göre en yaşlı 3 erkek ve kadın kullanıcıyı bulun.

Instructions:

1. Find average age of male and female users. In addition, how long have they been using the social media application on average?
2. Find average age of male and female users **by countries**. In addition, how long have they been using the social media application on average?
3. Find top 3 oldest male and female users **by countries**.

“get_and_write_data.py”

```
import pandas as pd
import requests
from time import time
from datetime import datetime, timedelta

def get_api_data():
    #Write your code here

def create_parquet_file():
    #Write your code here

if __name__ == "__main__":
    #Write your code here
```

“spark_analysis.py”

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import *
from pyspark.sql.types import *
from pyspark.sql.window import Window
```

```
def create_spark_session():
    # Write your code here
```

```
def read_parquet_file():
    # Write your code here
```

```
def question_1_spark_sql():
    # Write your code here
```

```
def question_1_dataframe_api():
    # Write your code here
```

```
def question_2_spark_sql():
    # Write your code here
```

```
def question_2_dataframe_api():
    # Write your code here
```

```
def question_3_spark_sql():
    # Write your code here
```

```
def question_3_dataframe_api():
    # Write your code here
```

```
if __name__ == "__main__":
    # Write your code here
```