Big Data Engineer Assessment

For this assignment, the datasets you will be working with consist of UK reported street crimes from January 2019 to March 2020.

1. You can download the relevant datasets here: https://data.police.uk/data/
2. On the page select the following for the respective fields:
    - Date Range: Choose any date range with a least 3 months.
    - Forces: All forces
    - Data sets: Include crime data and Include outcomes data
3. Once you have downloaded the zip file and extracted the data, you will notice for each district has two files.
    - As an example for “Avon and Somerset” we have:
        - 2019-01-avon-and-somerset-street.csv and;
        - 2019-01-avon-and-somerset-outcomes.csv

Your task is to:

1. Extract the following fields from each csv:
    a. crimeID
    b. districtName
        i. Can be extracted from the filename
    c. latitude
    d. longitude
    e. crimeType
    f. lastOutcome
        i. The last outcome should be taken from the <district>-outcomes.csv file where the crime IDs match. If there is no matching data use the data listed in the original <district>.csv file.
    g. As an example the final data structure should look like **Table_1** below 2.

Store the final structured data (could be either local Parquet files or MongoDB,

Elasticsearch)

3. The data should be accessible either through

    a. an API call (documented in the README.md file);

b. scripts, in case of using Apache Spark (documented in the README.md file) c. visual tool like Kibana (optional)

4. Provide KPIs (JSON structure) to relevant end-users that are examining crime statistics (e.g: different crimes, location hotspots, crimes by location, crimes by crimeTypes, etc). KPIs should be accessible in the same way as in task 3.

5. Use docker-compose to orchestrate the set up of your final solution project. (bonus)

6. Make sure to include a detailed README.md file outlining the setup instructions and description of what steps you took to reach your final solution.

7. Upload your project to GitHub/GitLab and include a link when responding.

crimeId	districtName	latitude	longitude	crimeType	lastOutcome
98096d1a6920 5691a56b89c1 182eadd6aaf1 5400ea18da13 4e0023f20aba 5cdb	Avon and Somerset Constabulary	-2.56417	51.508139	Criminal damage and arson	Under investigation

Table 1

If there are any details that are unclear please do not hesitate to ask any questions.

Good luck! We look forward to reviewing your solution.