

Python Take Home Challenge

Problem Statement:

You have been provided with a dataset of patients' medical records. Your task is to write a Python program that processes this data and generates a detailed report.

Steps:

1. **Load the Data:** The data is provided in a XML file format (`people_data.xml`). Use `pandas` to load the data into a DataFrame.
2. **Data Cleaning:** Handle any missing or incorrect data (e.g., missing values, malformed data, etc.). Assume missing or invalid values should be discarded.
3. **Generate a Report:**
 - The report should be a summary of people over the age of 18 by city.
 - For each city, provide the following information:
 - **Number of adults (Older than 18)** for each city.
 - **Number of children (Younger than 18)** for each city.
 - The report should be formatted as a table and output to a JSON file (`report.json`)
4. **Visualization (Optional):**
 - (Bonus) Generate a bar chart of average age by city and save it as an image (`age_by_city.png`).

Example Input (XML File `people_data.xml`):

```
Unset
<people>
  <person>
    <id>7966824611</id>
    <name>John Doe</name>
    <dob>1989-06-15</dob> <!-- Date of Birth -->
    <address>
      <street>123 Main St</street>
      <city>Springfield</city>
      <state>IL</state>
      <zipcode>62701</zipcode>
```

```
        <country>USA</country>
    </address>
</person>
<person>
    <id>8699099618</id>
    <name>Sally Doe</name>
    <dob>2010-04-03</dob>
    <address>
        <street>564 State St</street>
        <city>Springfield</city>
        <state>IL</state>
        <zipcode>77664</zipcode>
        <country>USA</country>
    </address>
</person>
<person>
    <id>8699099618</id>
    <name>Mario Rodríguez</name>
    <dob>2000-05-03</dob>
    <address>
        <street>334 N 712 E</street>
        <city>SLC</city>
        <state>UT</state>
        <zipcode>12324</zipcode>
        <country>USA</country>
    </address>
</person>
<person>
    <id>8699099618</id>
    <name>Margaret Doe</name>
    <!-- Some birth dates do not have dashes, this still needs to be
parsed and handled correctly -->
    <dob>19990101</dob>
    <address>
        <street>564 State St</street>
        <city>Springfield</city>
        <state>IL</state>
        <zipcode>46753</zipcode>
        <country>USA</country>
    </address>
</person>
<!-- This record is invalid as it does not contain the country field. It
should be discarded. -->
<person>
```

```
<id>6240350649</id>
<name>Zack Black</name>
<dob>1930-03-01</dob>
<address>
  <street>334 N 712 E</street>
  <city>SLC</city>
  <state>UT</state>
  <zipcode>12324</zipcode>
</address>
</person>
</people>
```

Deliverables:

- Python script to process the data and generate the report (`generate_report.py`).
- Generated report JSON file (`report.json`).
- (Optional) Image of the sales visualization (`age_by_city.png`).