Orb Screening Problems

Problem #1

Companies often have two names: legal name and a DBA name, which means "Doing Business As". We need to split legal and DBA name.

Write a Python function `clean_names(raw_names)` that gets a list of raw names, processes it and returns the list of pairs equal to `CLEANED_NAME_PAIRS` list.

```
RAW_NAMES = [
    'SPV  Inc., DBA:   Super  Company',
    'Michael Forsky LLC d.b.a F/B Burgers .',
    '*** Youthful You Aesthetics ***',
    'Aruna Indika (dba. NGXess)',
    'Diot SA,  -  D. B. A.   *Diot-Technologies*',
    'PERFECT PRIVACY, LLC, d-b-a Perfection,',
    'PostgreSQL DB Analytics',
    '/JAYE INC/',
    ' ETABLISSEMENTS  SCHEPENS /D.B.A./ ETS_SCHEPENS',
    'DUIKERSTRAINING OOSTENDE | D.B.A.:  D.T.O. '
]

CLEANED_NAME_PAIRS = [
    ('SPV Inc',                  'Super Company'),
    ('Michael Forsky LLC',       'F/B Burgers'),
    ('Youthful You Aesthetics',  None),
    ('Aruna Indika',             'NGXess'),
    ('Diot SA',                  'Diot-Technologies'),
    ('PERFECT PRIVACY, LLC',     'Perfection'),
```

```python
    ('PostgreSQL DB Analytics',  None),
    ('JAYE INC',                 None),
    ('ETABLISSEMENTS SCHEPENS',  'ETS SCHEPENS'),
    ('DUIKERSTRAINING OOSTENDE', 'D.T.O'),
]

def clean_names(raw_names):
    # write your code here

assert clean_names(RAW_NAMES) == CLEANED_NAME_PAIRS
```

Problem #2

```sql
CREATE TABLE companies (
  company_id INT,
  name       TEXT,
  city       TEXT,
  country    TEXT,
  revenue    INT
);

CREATE TABLE offices (
  location_id INT,
  company_id  INT REFERENCES companies (company_id),
  name        TEXT,
  city        TEXT,
  country     TEXT
);
```

We have separate tables for storing information about the companies and their offices. Some companies don't have any office, some have many offices.

Write a SQL query that returns name, revenue and number of offices for all companies that have less than 5 offices. Order the result by companies' number of offices.

Problem #3

Write a Python script that:

- fetches Caltrans offices page: https://dot.ca.gov/contact-us
- extracts offices information
- and returns office info in JSON form as shown below:

```
[
  {
    "office_name": "Headquarters",
    "office_link": "http://www.dot.ca.gov/",
    "office_address": "1120 N Street",
    "office_city": "Sacramento",
    "office_state": null,
    "office_zip": null,
    "office_phone": "916-654-5266",
    "mail_address": null,
    "mail_pobox": "P.O. Box 942873",
    "mail_city": "Sacramento",
    "mail_state": "CA",
    "mail_zip": "94273-0001",
    "mail_phone": null
  },
  ...
  {
    "office_name": "District 12",
    "office_link": "http://www.dot.ca.gov/dist12/",
```

```json
    "office_address": "1750 East 4th Street",
    "office_city": "Santa Ana",
    "office_state": "CA",
    "office_zip": "92705",
    "office_phone": "657-328-6000",
    "mail_address": "1750 East 4th Street",
    "mail_pobox": null,
    "mail_city": "Santa Ana",
    "mail_state": "CA",
    "mail_zip": "92705",
    "mail_phone": null
  }
]
```