Un chercheur Australien, Murray Cox, met en partage des données d'usage de AIRBNB sur le site insideairbnb.com. On y apprend qu'il y a plus de 60.000 logements proposés dans Paris (à comparer avec les 80.000 chambres d'hôtels) et que parmi les nombreux casse-têtes de la ville, il y a celui de l'usage abusif des logements sociaux dont certains sont proposés sur airbnb.
Je me suis livré à une expérience sur ces données, combinées avec celles de la ville de Paris (en open data) qui recense les immeubles de son parc de logements sociaux.
L'objectif est de rapprocher deux sources de données afin d'identifier l'usage anormal d'un logement social.

Méthode

1. Télécharger le set de données le plus récent pour la ville de Paris sur le site insideairbnb.com/paris (60.000 logements environ en juillet 2019)

2. Télécharger le set de données contenant la liste des logements sociaux de la ville de paris depuis le site open data (3100 immeubles, environ)

3. Charger les deux fichiers CSV dans une base relationnelle (Oracle dans mon cas avec le produit Oracle APEX)

4. Géolocaliser plus pécisément les adresse des logements sociaux à partir des APIs Google Maps. Les coordonnées fournies dans le fichier open Data présentent trop d'écart avec la localisation réelle sur Google Maps, déduite de l'adresse.

5. Effectuer un appariement (simple produit cartésien. cf requête en annexe)

afin de détecter les localisations très proches (distance géodésique inférieure à 5m)

Extraire les couples répondant au critère ci-dessus et ne retenir que ceux présentant la caractéristique d'être proposés en location complète (ce qui est illégal pour une sous-location d'un appartement HLM, par exemple)

On arrive à un total de 334 logements, dont 39 sont très pratiquement confondues (moins de 2 m).
Il ne faut surtout pas conclure que ces propositions de logement sont illégales ! Il s'agit simplement d'une base de travail à partir de laquelle il reste à effectuer d'autres vérifications pour écarter toute suspicion.
Pour info: un id de logement airbnb, qui est unique, peut être examiné en utilisant l'url de la forme: https://www.airbnb.fr/rooms/XXXXX

Interface d'analyse des données

Les *snaps* d'écran suivants illustrent l'interface graphique que l'on peut produire rapidement avec Oracle APEX, et que l'on peut mettre à disposition d'analystes fonctionnels.

-
-
-
-
-
  Heat Map montrant les prix moyens par quartier de Paris
-
  Recherche par facettes sur les logements sociaux de Paris

-
-
-
-
-

Annexes

Requetes SQL utilisées

```
insert into air_map (id_ag,id_ta,la_ag,lg_ag,la_ta,lg_ta,geo_dist)
select
    A.id,
    T.id,
    to_char(A.LATITUDE),
    to_char(A.LONGITUDE),
    to_char(T.LATITUDE),
    to_char(T.LONGITUDE),
    trunc(
         sdo_geom.sdo_distance(
              sdo_geometry(2001, 4326, sdo_point_type(
a.longitude,a.latitude, null), null, null),
              sdo_geometry(2001, 4326,
sdo_point_type(t.longitude,t.latitude, null), null, null)
         )
    )
from AIR_PARIS A, AIR_PARIS_HLM T
where trunc(A.latitude,3) = trunc(T.latitude,3) and
trunc(A.longitude,3) = trunc(T.longitude,3)
```

Géocodage

```
create or replace FUNCTION  "AIR_GPS"
return VARCHAR2
is

-- P MONACO - GPM FACTORY
-- Aug 2019

l_clob    CLOB;
t_temp varchar2(2000);
tnb INTEGER;
tlat VARCHAR2(30);
tlong VARCHAR2(30);
taddr VARCHAR2(2000);
tcode VARCHAR2(2000);
l_values apex_json.t_values;
l_count       pls_integer;
tkey  varchar2(100);
CURSOR poi_cur
IS
SELECT * from AIR_PARIS_HLM
where
glong is null;
l_poi   poi_cur%ROWTYPE;
BEGIN
-- Retrieve API Key
select key into tkey
from API_KEY
where name = 'Google Maps API';
OPEN poi_cur;
LOOP
FETCH poi_cur INTO l_poi;
```

```
EXIT WHEN poi_cur%NOTFOUND;  -- We go away

apex_web_service.g_request_headers(1).name  := 'Content-Type';
apex_web_service.g_request_headers(1).value := 'application/json';

l_clob := APEX_WEB_SERVICE.make_rest_request(
p_url         => 'https://maps.googleapis.com/maps/api/geocode/json',
p_parm_name => apex_util.string_to_table(
'address:key'),
p_parm_value => apex_util.string_to_table(
l_poi.adresse||', Paris France:'||tkey
),
p_http_method => 'GET'
);

apex_json.parse(p_values => l_values,
p_source => l_clob );
l_count := apex_json.get_count(
p_path   => '.',
p_values => l_values
);

l_count :=1;
for idx in 1 .. l_count
loop
tlat :=
apex_json.get_varchar2(p_path=>'results[%d].geometry.location.lat',p0=
> 1,p_values=>l_values);
tlong :=
apex_json.get_varchar2(p_path=>'results[%d].geometry.location.lng',p0=
> 1,p_values=>l_values);

end loop;
if tlat is null then tlat := 0;end if;
```

```
update air_paris_hlm set glong=tlong, glat = tlat
where id = l_poi.id;
COMMIT;
END LOOP;  -- End cursor
COMMIT;
CLOSE poi_cur;
htp.p('Terminé');
return('0');
EXCEPTION
when others then
htp.p('Undefined problem');
END;
```

Tables

```
CREATE TABLE  "AIR_PARIS"
    (    "ID" NUMBER,
     "NAME" VARCHAR2(255),
     "HOST_ID" NUMBER,
     "HOST_NAME" VARCHAR2(255),
     "NEIGHBOURHOOD_GROUP" VARCHAR2(1),
     "NEIGHBOURHOOD" VARCHAR2(30),
     "LATITUDE" NUMBER,
     "LONGITUDE" NUMBER,
     "ROOM_TYPE" VARCHAR2(30),
     "PRICE" NUMBER,
     "MINIMUM_NIGHTS" NUMBER,
     "NUMBER_OF_REVIEWS" NUMBER,
     "LAST_REVIEW" DATE,
     "REVIEWS_PER_MONTH" NUMBER,
     "CALCULATED_HOST_LISTINGS_COUNT" NUMBER,
```

```
      "AVAILABILITY_365" NUMBER,
       CONSTRAINT "AIR_PARIS_PK" PRIMARY KEY ("ID")
   USING INDEX  ENABLE
    )

CREATE TABLE  "AIR_PARIS_HLM"
    (    "ID" NUMBER,
     "ADRESSE" VARCHAR2(255),
     "CODE_POSTAL" NUMBER,
     "VILLE" VARCHAR2(30),
     "ANNÉE_DU_FINANCEMENT___AGRÉMENT" NUMBER,
     "BAILLEUR_SOCIAL" VARCHAR2(30),
     "NOMBRE_LOGEMENTS" NUMBER,
     "PLA_I" NUMBER,
     "PLUS" NUMBER,
     "PLUS_CD" NUMBER,
     "PLS" NUMBER,
     "MODE_DE_RÉALISATION" VARCHAR2(100),
     "COMMENTAIRES" VARCHAR2(1000),
     "ARRONDISSEMENT" NUMBER,
     "NATURE_DE_PROGRAMME" VARCHAR2(255),
     "COORD_X_L93" NUMBER,
     "COORD_Y_L93" NUMBER,
     "GEO_SHAPE" VARCHAR2(300),
     "GEO_POINT_2D" VARCHAR2(300),
     "LATITUDE" NUMBER,
     "LONGITUDE" NUMBER,
     "GLAT" NUMBER,
     "GLONG" NUMBER,
      CONSTRAINT "AIR_PARIS_HLM_PK" PRIMARY KEY ("ID")
   USING INDEX  ENABLE
    )

CREATE TABLE  "AIR_MAP"
```

```
(    "ID" NUMBER(10,0),
 "ID_AG" NUMBER,
 "ID_TA" NUMBER,
 "LA_AG" VARCHAR2(30),
 "LG_AG" VARCHAR2(30),
 "LA_TA" VARCHAR2(30),
 "LG_TA" VARCHAR2(30),
 "GEO_DIST" NUMBER(10,0),
 "ROAD_DIST" NUMBER(10,0),
 "ROAD_TIME" NUMBER(10,0),
 "ROAD_TIME_TEXT" VARCHAR2(100),
 "PREFERED" NUMBER(1,0)
)
```

Author

- 

  Patrick

  GPM Factory