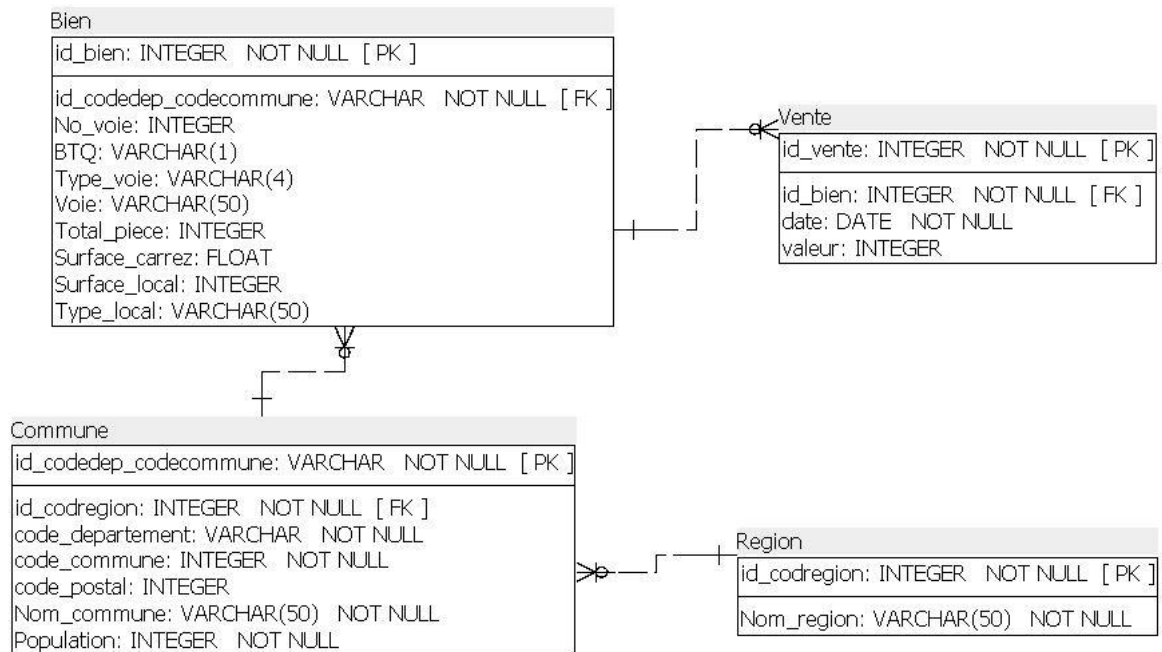


Formation DATA ANALYST

Projet SQL : Créez et utiliser une base de données immobilière avec SQL (Postgresql)

1. SCHEMA RELATIONNEL

Modélisation d'une base de données immobilière avec SQL POWER ARCHITECT



2. Import des données dans Postgresql

J'ai premièrement Créé des tables dans Posgresql en respectant toutes les contraintes des colonnes et deuxièmement traiter les fichier excel (Power query et les fonctions excel) contenant les données. En fin j'ai importé les fichiers csv obtenu dans la base de donnée postgresql.

Création des tables :

```
1 create table Vente (  
2 Id_vente Integer not null primary key,  
3 foreign key(Id_bien) references Bien(Id_bien),  
4 Date date not null,  
5 Valeur integer  
6 );  
7  
8 create table Region (  
9 Id_codregion Integer not null primary key,  
10 Nom_region Varchar(50) not null  
11 );
```

```

13 create table Commune (
14   Id_codedep_codecommune varchar not null primary key,
15   Id_coderegion integer not null references region(Id_coderegion),
16   Code_departement varchar not null,
17   Code_commune integer not null,
18   code_postal integer,
19   Nom_commune varchar(50) not null,
20   Population integer not null
21 );
22
23
24 create table Bien (
25   Id_bien integer not null primary key,
26   Id_codedep_codecommune varchar not null references Commune(Id_codedep_codecommune),
27   No_voie integer,
28   BTQ varchar(1),
29   Type_voie varchar(4),
30   Voie varchar(50),
31   Total_piece Integer,
32   Surface_carrez float,
33   Surface_local integer,
34   Type_local varchar(50)
35 );
36
37
38 create table Vente (
39   Id_vente integer not null primary key,
40   Id_bien integer not null references Bien(Id_bien),
41   Date date not null,
42   Valeur integer
43 );
44

```

Import des données contenu dans les fichiers Excel dans les tables :

The screenshot shows the pgAdmin 4 interface. On the left is the 'Browser' pane with a tree view of the database structure. The 'Tables (4)' folder is expanded, showing 'bien', 'commune', 'region', and 'vente'. The 'vente' table is selected. The main pane shows a SQL query in the 'Query' tab:

```

39 (66,'Mayotte'),(28,'Normandie'),(75,'Nouvelle-Aquitaine'),
40 (76,'Occitanie'),(52,'Pays-de-la-Loire'),(93,'Provence-Alpes-Côte-d'Azur');
41
42 select * from region;
43
44 select * from commune;
45
46 select * from vente;
47
48 select * from bien
49

```

Below the query editor is the 'Data output' pane, which displays the results of the query in a table format. The table has 9 columns: id_codedep, codecommu, id_coderegion, code_departement, code_commune, code_postal, nom_commune, and population. The results show 7 rows of data for the 'vente' table.

	id_codedep	codecommu	id_coderegion	code_departement	code_commune	code_postal	nom_commune	population
1	1001		84	1		1	L'Abergement-Clémén...	779
2	1002		84	1		2	L'Abergement-de-Varey	256
3	1004		84	1		4	Ambérieu-en-Bugey	14134
4	1005		84	1		5	Ambérieu-en-Dombes	1751
5	1006		84	1		6	Amblién	112
6	1007		84	1		7	Ambronay	2800
7	1008		84	1		8	Ambutrix	762

At the bottom of the 'Data output' pane, it says 'Total rows: 1000 of 34991' and 'Query complete 00:00:00.449'.

3. Requête pour répondre aux questions Métiers (business)

Question 1 : Nombre total d'appartements vendus au 1er semestre 2020.

Réponse : A partir de la requête (voir figure ci-dessous) nous pouvons dire qu'au premier semestre 2020 il y a **15442 appartements** qui ont été vendus.

```

--Question 1
select
    count("id_vente") as Nombre_total_appartements_vendus
from "vente"
inner join bien using("id_bien")
where date between '2019/12/31' and '2020/04/01' and type_local='Appartement';

```

nombre_total_appartements_vendus
bigint
15442

Question 2 : Le nombre de ventes d'appartement par région pour le 1er semestre 2020

Réponse : ile-de-France(7100), Provence-Alpes-cote(2129),Auvergne-Rhone-Alpes(1410),Nouvelle-Aquitaine(956), Occitanie(762), Pays-de-la-Loire(607),Hauts-de-France(590),Bretagne(422), Grand-Est (401), Normandie(369), Centre-Val-de-Loire(305), Bourgogne-Franche-Comté(189), Corse (123), Martinique(36), Guyane(23), La-Réunion(18), Guadeloupe(2)

```
--Question 2
with premier_semestre2020 as (
  select
    *
  from vente
  where date between '2019/12/31' and '2020/04/01'
)
select
  b_c_r."nom_region",
  count("id_vente") as Nombre_total_appartements_vendus_parRegion
from premier_semestre2020
left join ( select "id_bien","nom_region","type_local" from bien
  left join commune using("id_codedep_codecommune")
  left join region using("id_coderegion")) as b_c_r using(id_bien)
where "type_local"='Appartement'
group by "nom_region"
order by Nombre_total_appartements_vendus_parRegion desc;
```

	nom_region character varying (50)	nombre_total_appartements_vendus_parregion bigint
1	Ile-de-France	7100
2	Provence-Alpes-Côte-...	2129
3	Auvergne-Rhône-Alpes	1410
4	Nouvelle-Aquitaine	956
5	Occitanie	762
6	Pays-de-la-Loire	607
7	Hauts-de-France	590
Total rows: 17 of 17		Query complete 00:00:00.196

Question 3 : Proportion des ventes d'appartements par le nombre de pièces.

Réponse : Nous pouvons dire à partir du résultat que les appartements de 2 pièce représentent 31% des ventes, ceux de 3 pièces représentent 28% des ventes, ceux d'une pièce 22% des ventes, ceux de 4 pièces 13% et ceux de 5 pièces 3%.

```
--Question 3
select
  "total_piece",
  count("id_vente")*100/(select count("id_vente") from vente
  inner join bien using("id_bien") where "type_local"='Appartement' and
  (date between '2019/12/31' and '2020/04/01')) as propotion_enourcentage
from vente
inner join bien on vente.id_bien=bien.id_bien
where "type_local"='Appartement' and date between '2019/12/31' and '2020/04/01'
group by "total_piece"
```


	total_piece integer	propotion_enourcentage bigint
1	8	0
2	10	0
3	9	0
4	7	0
5	1	22
6	5	3
7	4	13
Total rows: 11 of 11		Query complete 00:00:00.158

Question 4 : Liste des 10 départements où le prix du mètre carré est le plus élevé.

Réponse : Les Départements ayant pour code : 75, 94, 91, 6, 92, 69, 59, 11, 93, 78.

```
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    "code_departement",
    ROUND(("valeur")/("surface_carrez")) as prix_m2
from premier_semestre2020
inner join bien using("id_bien")
inner join commune using("id_codedep_codecommune")
order by prix_m2 desc
Limit 50;
```

	code_departement character varying	prix_m2 double precision
8	75	417407
9	94	262319
10	75	178162
11	75	166688
12	91	134375
13	75	117800
14	6	88073
Total rows: 50 of 50		Query complete 00:00:00.159

Question 5 : Prix moyen du mètre carré d'une maison en Île-de-France.

Réponse : Le prix moyen du mètre carré d'une maison en Île-de-France est : **3738 euros**

```
--Question 5
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    avg(("valeur")/("surface_carrez")) as prix_moyen_maison_ilefrance
from premier_semestre2020
left join bien using("id_bien")
left join commune using("id_codedep_codecommune")
where ("id_coderegion"=11) and (type_local='Maison');
```

prix_moyen_maison_ilefrance
double precision
3738.757362916054

Question 6 : Liste des 10 appartements les plus chers avec la région et le nombre de mètres carrés.

Réponse : Les appartements les plus chère se trouvent en Ile de France et en Provence-Alpes-Côte-d'Azur.

```
--Question 6
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    b_c_r."nom_region",
    "valeur",
    b_c_r."surface_carrez"
from premier_semestre2020
left join ( select "id_bien","nom_region","type_local","surface_carrez" from bien
             left join commune using("id_codedep_codecommune")
             left join region using("id_coderegion")) as b_c_r using(id_bien)
where "type_local"='Appartement'
order by "valeur" desc
Limit 20;
```

nom_region	valeur	surface_carrez
character varying (50)	integer	double precision
Ile-de-France	8600000	64
Ile-de-France	8577713	20.55
Ile-de-France	7620000	42.77
Ile-de-France	7600000	253.3
Ile-de-France	7420000	360.95
Ile-de-France	7200000	595
Ile-de-France	7050000	122.56

rows: 20 of 20 Query complete 00:00:00.210

Question 7 : Taux d'évolution du nombre de ventes entre le premier et le second trimestre de 2020.

Réponse : Le taux d'évolution est de 4%.

```
--Question 7
select
    cast((-count("id_vente")+(select count("id_vente")
    from vente where date between '2020/03/31' and '2020/07/01'))*100/count("id_vente")
    as float) as taux_croissance
from vente
where date between '2019/12/31' and '2020/04/01';
```

	taux_croissance double precision
1	4

Question 8 : Le classement des régions par rapport au prix au mètre carré des appartement de plus de 4 pièces.

Reponse : Hauts-de-France, Ile-de-France, Bourgogne-Franche-Comté, Provence-Alpes-Côte-d'Azur, Provence-Alpes-Côte-d'Azur, Bretagne.

```
--Question 8
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    b_c_r."nom_region",
    ("valeur")/("surface_carrez") as prix_m2,
    "total_piece"
from premier_semestre2020
left join ( select "id_bien","nom_region","type_local","surface_carrez","total_piece" from bien
    left join commune using("id_codedep_codecommune")
    left join region using("id_coderegion")) as b_c_r using("id_bien")
where ("total_piece">4) and (type_local='Appartement');
```

	nom_region character varying (50)	prix_m2 double precision	total_piece integer
1	Hauts-de-France	1281.13769380839	5
2	Ile-de-France	4502.88162727187	5
3	Ile-de-France	7401.25726485588	5
4	Ile-de-France	2149.75845410628	5
5	Hauts-de-France	860.032705469081	6
6	Bourgogne-Franche-C...	1241.01281650515	5
7	Ile-de-France	9177.17086834733	6
Total rows: 682 of 682		Query complete 00:00:00.285	

Question 9 : Liste des communes ayant eu au moins 50 ventes au 1er trimestre

Réponse : Paris 17^e arrondissement, Paris 15^e arrondissement, Paris 18^e arrondissement, Nice, Paris 11^e arrondissement, Paris 16^e arrondissement, Bordeaux.

```
--Question 9
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    "nom_commune",
    count("id_vente") as Nombre_vente_commune
from premier_semestre2020
left join bien using("id_bien")
left join commune using("id_codedep_codecommune")
group by "nom_commune"
having count("id_vente")>=50
order by Nombre_vente_commune desc;
```

	nom_commune character varying (50) 🔒	nombre_vente_commune bigint 🔒
1	Paris 17e Arrondisse...	228
2	Paris 15e Arrondisse...	215
3	Paris 18e Arrondisse...	210
4	Nice	173
5	Paris 11e Arrondisse...	170
6	Paris 16e Arrondisse...	166
7	Bordeaux	157
Total rows: 48 of 48		Query complete 00:00:00.162

Question 10 : Différence en pourcentage du prix au mètre carré entre un appartement de 2 pièces et un appartement de 3 pièces.

Reponse : Cette différence est de 22,18%


```
--Question 10
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    (-percentile_disc(.5)within group (order by "valeur")+
    (select percentile_disc(.5)within group(order by "valeur")
    from premier_semestre2020 inner join bien using("id_bien")
where total_piece=3))*100/avg("valeur")as pourcen_difference23
from premier_semestre2020
inner join bien using("id_bien")
where total_piece=2;
```

pourcen_difference23
numeric
22.1854652520205380

Question 11 : Les moyennes de valeurs foncières pour le top 3 des communes des départements 6, 13, 33, 59 et 69.

Reponse :

```
--Question 11
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    "code_departement",
    "nom_commune",
    avg("valeur") over (partition by "nom_commune") as moyenne
from premier_semestre2020
left join "bien" using("id_bien")
left join "commune" using("id_codedep_codecommune")
where "code_departement" in ('6','13','33','59','69')
order by moyenne desc;
```

	code_departement character varying	nom_commune character varying (50)	moyenne numeric
1	6	Saint-Jean-Cap-Ferrat	1315000.000
2	6	Saint-Jean-Cap-Ferrat	1315000.000
3	6	Eze	980000.0000
4	59	Halluin	970000.0000
5	69	Genas	621300.0000
6	69	Lyon 6e Arrondissement	600084.2500
7	69	Lyon 6e Arrondissement	600084.2500
Total rows: 1000 of 2971		Query complete 00:00:00.281	

Question 12 : Les 20 communes avec le plus de transactions pour 1000 habitants pour les communes qui dépassent les 10 000 habitants.

Reponse : Paris 17^e arrondissement, Paris 15^e arrondissement, Paris 18^e arrondissement, Nice, Paris 11^e arrondissement, Paris 16^e arrondissement, Bordeaux.

```
--Question 12
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    "nom_commune",
    sum("population") as population,
    count("id_vente")*100/(sum("population")/1000) as propor_transaction
from premier_semestre2020
left join "bien" using("id_bien")
left join "commune" using("id_codedep_codecommune")
group by "nom_commune"
having sum("population")>10000
order by propor_transaction desc
Limit 20
```

	nom_commune character varying (50)	population bigint	propor_transaction bigint
1	Huez	14366	78
2	Saint-Etienne-de-Tinée	10458	70
3	Blonville-sur-Mer	21742	66
4	Vieux-Boucau-les-Bains	14526	64
5	Houlgate	30798	60
6	Les Deux Alpes	25077	52
7	Notre-Dame-de-Monts	15043	46
Total rows: 20 of 20		Query complete 00:00:00.200	

Question 13 : *Corrélation entre la variable valeur et la variable surface_carrez.*

Réponse : *La corrélation entre les deux variables est faible.*

```
--Question 13
select
    corr("valeur","surface_carrez")
from vente as v
inner join "bien" as b on v.id_bien=b.id_bien
```

corr	double precision
	0.30291257633327

Question 14 : *Classement des appartements les plus luxueux, les moins couteux et ceux ayant un cout moyen.*

Reponse : *Nous pouvons remarquer que l'appartements le plus luxueux cout 8 millions 600000 euros et tous les appartements luxueux sont situés en majorité ile de France*

```
--Question 14
with premier_semestre2020 as (
    select * from vente
    where date between '2019/12/31' and '2020/04/01'
)
select
    b_c_r."nom_region",
    "type_local",
    case when "valeur"<10000 then 'Très moins couteux'
         when "valeur">1000000 then 'Très luxueux'
         else 'Moyen' end as standing
    , "valeur" as cout
from premier_semestre2020
inner join ( select "id_bien","nom_region","type_local" from bien
            inner join commune using("id_codedep_codecommune")
            inner join region using("id_coderegion")) as b_c_r using(id_bien)
where "type_local"='Appartement'
order by "valeur" desc
```

nom_region character varying (50)	type_local character varying (50)	standing text	cout integer
Ile-de-France	Appartement	Très luxie...	8600000
Ile-de-France	Appartement	Très luxie...	8577713
Ile-de-France	Appartement	Très luxie...	7620000
Ile-de-France	Appartement	Très luxie...	7600000
Ile-de-France	Appartement	Très luxie...	7420000
Ile-de-France	Appartement	Très luxie...	7200000
Ile-de-France	Appartement	Très luxie...	7050000
rows: 1000 of 15442 Query complete 00:00:00.311 Rows selected: 1			

4. Dictionnaire des données

1- Fichier Données communes

Noms de la colonne	Description	Type
CODREG	Code région	entier
CODDEP	Code département	Chaine de caractère
CODARR	Code arrondissement	entier
CODCAN	Code canton	entier
CODCOM	Code Commune	entier
COM	Nom de la commune	Chaine de caractère
PMUN	La population municipale comprend les personnes ayant residant sur le territoire de la commune, dans un logement ou une communauté;	entier
PCAP	La population comptée à part comprend certaines personnes dont la résidence habituelle est dans une autre commune mais qui ont conservé une résidence sur le territoire de la commune .	entier
PTOT	La population totale d'une commune est égale à la	entier

	somme de la population municipale et de la population comptée à part de la commune.	
--	---	--

2- Valeurs foncières

Noms de la colonne	Description	Type
N° de disposition	Dans le cas des actes comprenant plusieurs mutations – appelées « dispositions » –, chacune d’entre elles est identifiée dans les fichiers par un numéro de disposition. Seules les dispositions concernant les mutations à titre onéreux sont restituées dans le fichier.	entier
Date de mutation	Date de signature de l’acte (au format JJ/MM/AAAA) – Une restitution au format AAAA/MM/JJ (norme ISO 8601) est prévue à compter de la mise à jour d’octobre 2019.	Date
Nature de la mutation	Vente, vente en l’état futur d’achèvement, vente de terrain à bâtir, adjudication, expropriation ou échange	Chaine de caractères
Valeur foncière	Il s’agit du montant ou de l’évaluation déclaré(e)	réel

	<p>dans le cadre d'une mutation à titre onéreux. La valeur foncière :</p> <ul style="list-style-type: none"> • inclut : <ul style="list-style-type: none"> • les frais d'agence, s'ils sont à la charge du vendeur. • l'éventuelle TVA. • exclut : <ul style="list-style-type: none"> • les frais d'agence, s'ils sont à la charge de l'acquéreur. • les frais de notaires. • la valeur des biens meubles stipulée dans l'acte de mutation. 	
N° de voie	Numéro dans la voie	entier
B/T/Q	Indice de répétition	caractère
Code type voie		entier
Type de voie	Exemple : Rue, avenue, etc.	Chaine de caractères
Code voie	Code Rivoli (répertoire informatisé codifiant, par commune, les voies, les lieux dits et les ensembles immobiliers)	Chaine de caractères
Voie	Libellé de la voie	Chaine de caractères
Code ID commune		entier
Code postal		entier
Commune	Libellé de la commune	Chaine de caractères
Code département	Référence cadastrale de la parcelle	entier
Code commune		entier
Préfixe de section		entier

Section		Chaine de caractères
No plan		entier
1er lot	Un lot de copropriété est constitué d'une partie privative (appartement, cave, etc.) et d'une quote-part de partie commune (tantièmes). Seuls les 5 premiers lots sont mentionnés. Si le nombre de lots est supérieur à 5, ils ne sont pas restitués.	Chaine de caractères
Surface Carrez du 1er lot		réel
2e lot		
Surface Carrez du 2e lot		réel
3e lot		
Surface Carrez du 3e lot		réel
4e lot		
Surface Carrez du 4e lot		réel
5e lot		
Surface Carrez du 5e lot		réel
Nombre de lots	Nombre total de lots par disposition.	entier
Code type local	1 : maison ; 2 : appartement ; 3 : dépendance (isolée) ; 4 : local industriel et commercial ou assimilés	
Type local		Chaine de caractères
Identifiant local	Donnée non restituée en application du décret n° 2018-1350 du 28 décembre 2018	
Surface réelle bâti	La surface réelle est la surface mesurée au sol entre les murs ou séparations et arrondie au mètre carré inférieur. Les surfaces des dépendances ne sont pas prises en compte.	entier

Nombre de pièces principales	Les cuisines, salles d'eau et dépendances ne sont pas prises en compte.	entier
Code nature culture	Voir dans le présente document la table de références « nature de culture »	Chaine de caractères
Nature culture spéciale	Voir dans le présent document la table de références « nature de culture spéciale »	Chaine de caractères
Surface terrain	Contenance du terrain	entier

3- Référentiel géographique France

Noms de la colonne	Description	Type
regrgp_nom	Groupe de region : Province, ile France et outre mer	Chaine de caractère
reg_nom	Nom de la région	Chaine de caractère
reg_nom_old	Ancien nom de la région	Chaine de caractère
aca_nom	Nom académie	Chaine de caractère
dep_nom	Nom département	Chaine de caractère
com_code	Code commune	entier
com_id	Identifiant commune	Chaine de caractère
com_nom	Nom commune	Chaine de caractère
uu_code	Code unité urbaine	Chaine de caractère
uu_id	Identifiant unite urbaine	Chaine de caractère
uucr_id	Identifiant unités urbaines et communes rurales	Chaine de caractère
uucr_nom	Nom unités urbaines et communes rurales	Chaine de caractère
ze_id	Identifiant zone d'emploi	Chaine de caractère
dep_code	Code département	Chaine de caractère
dep_id	Identifiant département	Chaine de caractère
dep_num_nom	Nom et numéro département	Chaine de caractère

aca_code	Code academie	entier
aca_id	Identifiant académie	Chaine de caractère
reg_code	Code region	entier
reg_id	Identifiant de la region	Chaine de caractère
reg_code_old	Ancien code région	Chaine de caractère
reg_id_old	Ancien identifiant colonne	Chaine de caractère
fd_id	Identifiant France détaillée	Chaine de caractère
fr_id	Identifiant France regroupée	Chaine de caractère
fe_id	Identifiant France entière	Chaine de caractère
uu_id_99	Identifaint unité urbaine	Chaine de caractère
au_code	Code aire urbaine	Chaine de caractère
au_id	Identifiant aire urbaine	Chaine de caractère
auc_id	identifiant aire urbaine et communes	Chaine de caractère
auc_nom	Nom aire urbaine et communes	Chaine de caractère
uu_id_10	Identifiant unite urbaine	Chaine de caractère
geolocalisation	Coordonnées géographique	réel