

# Projet 3: Requêtez une base de données avec SQL

Analyse du marché des assurances  
habitation:

Méthodologie, résultats, et  
tendances.

# Marché des assurances habitation: objectifs de l'analyse

- L'objectif de la présente étude est de fournir une analyse des données du marché des assurances habitation en France en décrivant les étapes clefs suivantes :
  - Analyse des données brutes provenant des CSV Contrats et Régions.
  - Etablissement du dictionnaire des données.
  - Elaboration du schéma relationnel.
  - Transcription du schéma en Base de Données.
  - Exploitation de la base de données pour l'analyse des tendances par requêtes SQL.
  - Conclusion.

# Marché des assurances habitation: des données brutes au dictionnaire de données.

- L'analyse des données d'entrée des **contrats d'assurance** croisées avec les **informations géographiques** des régions françaises permettent d'établir le **dictionnaire des données**, 1 ère étapes nécessaire à l'élaboration de la base pour stocker ces données clients (BDD)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Contrat_ID	No_voie	B_T_Q	Type_de_voie	Voie	Code_dep_coi	Code_postal	Surface	Type_local	Occupation	Type_contrat	Formule	Valeur_declar	Prix_cotisation_mensuel	
2	100601	190	A	RUE	CENTRALE	1350	1370	50	Appartement	Locataire	Residence pri Integral	0-25000		25	
3	100602	347		RUE	DU CHATEAU	1103	1170	48	Appartement	Locataire	Residence pri Classique	0-25000		30	
4	100603	58		AV	DU MONT BLA	1143	1220	131	Appartement	Proprietaire	Residence pri Integral	25000-50000		57	
5	100604	140		RUE	DE L'ABBE JOL	1288	1630	109	Maison	Locataire	Residence pri Integral	25000-50000		43	
6	100605	39		RUE	BUFFON	1033	1200	109	Appartement	Locataire	Residence pri Classique	0-25000		33	
7	100606	8		RUE	DE GENEVE	1354	1630	53	Appartement	Proprietaire	Residence pri Classique	0-25000		19	
8	100607	2		RUE	DU RECULET	1354	1630	59	Appartement	Proprietaire	Residence pri Integral	0-25000		15	
9	100608	1403		RUE	JEAN DE GING	1143	1220	93	Maison	Proprietaire	Mise en locati Integral	25000-50000		34	
10	100609	226		ALL	DES CAPUCIN	1354	1630	117	Maison	Proprietaire	Residence pri Classique	25000-50000		32	
11	100610	276		RTE	DE POUIGNY	1288	1630	36	Appartement	Proprietaire	Residence pri Integral	25000-50000		22	
12	100611	79		CRS	DE VERDUN	1283	1100	138	Appartement	Proprietaire	Residence pri Classique	0-25000		11	
13	100612	240		RUE	DE PRE BAILL	1173	1170	45	Appartement	Locataire	Residence pri Classique	0-25000		16	
14	100613	3		RUE	TURENNE	1033	1200	83	Appartement	Locataire	Residence pri Classique	0-25000		14	
15	100614	44												34	
16	100615	59												24	
17	100616	282												17	
18	100617	54												10	
19	100618	1												20	
20	100619	17												12	
21	100620	329												11	
22	100621	369												12	
23	100622	6												12	
24	100623	2												20	

	A	B	C	D	E	F	G	H
1	Code_dep_coi	reg_code	reg_nom	aca_nom	dep_nom	com_nom_ma	dep_code	dep_nom_num
2	1001	84	Auvergne-Rhône	Lyon	Ain	LABERGEMEN	1	Ain (01)
3	1002	84	Auvergne-Rhône	Lyon	Ain	LABERGEMEN	1	Ain (01)
4	1003	84	Auvergne-Rhône	Lyon	Ain	AMAREINS	1	Ain (01)
5	1004	84	Auvergne-Rhône	Lyon	Ain	AMBERIEU EN	1	Ain (01)
6	1005	84	Auvergne-Rhône	Lyon	Ain	AMBERIEUX EI	1	Ain (01)
7	1006	84	Auvergne-Rhône	Lyon	Ain	AMBLEON	1	Ain (01)
8	1007	84	Auvergne-Rhône	Lyon	Ain	AMBRONAY	1	Ain (01)
9	1008	84	Auvergne-Rhône	Lyon	Ain	AMBUTRIX	1	Ain (01)
10	1009	84	Auvergne-Rhône	Lyon	Ain	ANDERT ET CC	1	Ain (01)
11	1010	84	Auvergne-Rhône	Lyon	Ain	ANGLEFORT	1	Ain (01)
12	1011	84	Auvergne-Rhône	Lyon	Ain	APREMONT	1	Ain (01)
13	1012	84	Auvergne-Rhône	Lyon	Ain	ARANC	1	Ain (01)
14	1013	84	Auvergne-Rhône	Lyon	Ain	ARANDAS	1	Ain (01)
15	1014	84	Auvergne-Rhône	Lyon	Ain	ARBENT	1	Ain (01)
16	1015	84	Auvergne-Rhône	Lyon	Ain	ARBIGNIEU	1	Ain (01)
17	1016	84	Auvergne-Rhône	Lyon	Ain	ARBIGNY	1	Ain (01)
18	1017	84	Auvergne-Rhône	Lyon	Ain	ARGIS	1	Ain (01)
19	1018	84	Auvergne-Rhône	Lyon	Ain	ARLUD	1	Ain (01)
20	1019	84	Auvergne-Rhône	Lyon	Ain	ARMIX	1	Ain (01)
21	1020	84	Auvergne-Rhône	Lyon	Ain	ARNANS	1	Ain (01)
22	1021	84	Auvergne-Rhône	Lyon	Ain	ARS SUR FOR	1	Ain (01)
23	1022	84	Auvergne-Rhône	Lyon	Ain	ARTEMARE	1	Ain (01)
24	1023	84	Auvergne-Rhône	Lyon	Ain	ASNIERES SUF	1	Ain (01)

	Nom des colonnes	Type de données	Taille	Clé	Description
CONTRAT.CSV	Contrat_ID	INT	4	Clé primaire	Id unique pour les contrats
	No_voie	INT	4		Numéro dans la voie pour l'adresse du logement assuré
	B_T_Q	CHAR	1		Indicateur éventuel de répétition pour l'adresse du logement assuré sur un caractère
	Type_de_voie	VARCHAR	20		Type de voie pour l'adresse du logement assuré: rue, av (Avenue), rte (Route), ...
	Voie	VARCHAR	30		Libellé de la voie pour l'adresse du logement assuré
	Code_dep_code_commune	VARCHAR	10	Clé secondaire	Concaténation du code département et code commune pour avoir une clé unique
	Code_postal	INT	4		Code postal pour l'adresse du logement assuré
	Surface	INT	4		Surface en m² du logement assuré
	Type_local	CVARCHAR	20		Type de logement (soit appartement ou maison)
	Occupation	CVARCHAR	20		Qui occupe le logement (soit un locataire ou un propriétaire)
	Type_contrat	CVARCHAR	20		Type d'occupation du logement à assuré(soit residence principal ou secondaire ou en location)
	Formule	CVARCHAR	20		Formule d'assurance souscrite (commercial soit classique ou integrale) pour le logement assuré
REGION.CSV	Valeur_declaree_biens	CVARCHAR	20		La tranche de valeur en euro des biens déclarés dans le logement assuré, 4 tranches fixes.
	Prix_cotisation_mensuel	INT	4		La cotisation à payer par mois pour l'assurance du logement.
	Code_dep_code_commune	VARCHAR	10	Clé primaire	Concaténation du code département et code commune pour avoir une clé unique
	reg_code	INT	4		Code de la Région ou se trouve le département
	reg_nom	VARCHAR	20		Nom de la Région ou si situe le Département
	aca_nom	VARCHAR	20		Nom de l'Académie d'où dépend le Département et donc la commune
	dep_nom	VARCHAR	20		Nom du Département
	com_nom_maj_court	VARCHAR	20		Nom de la Commune en Majuscule, raccourci
	dep_code	VARCHAR	10		Code du département (attention à la corse 2A /2B)
	dep_nom_num	VARCHAR	20		Concaténation du Nom du Département et du code de Département entre (), séparé par un espace.

Le **champs commun**, unique et distinct permettant de faire **le lien** (jointure) entre les contrats et les informations de régions est la concaténation du Code département avec le code commune (colonne **Code\_dep\_code\_commune**).

# Marché des assurances habitation: du schéma Relationnel à la création de la base de données

- 1 même **Région** identifiée par son Code\_dep\_code\_commune (clef primaire) pouvant concerné **plusieurs Contrat** d'assurance (identifié par sa Contrat\_ID, Clef primaire ), la **définition des tables** avec le type de chaque attribut ainsi que la **relation entre les contrats et les Région** (via la contrainte de clef étrangère Code\_dep\_code\_commune de la table contrat) peuvent être **modélisés afin d'établir le schéma relationnel** d'entrée nécessaire à la création de la **base de données finale**.

Contrat	
Contrat_ID: INTEGER NOT NULL [ PK ]	
No_voie: INTEGER NOT NULL	
B_T_Q: CHAR(1)	
Type_de_voie: VARCHAR(20)	
Voie: VARCHAR(30)	
Code_dep_code_commune: VARCHAR(10) NOT NULL [ FK ]	
Code_postal: INTEGER NOT NULL	
Surface: INTEGER NOT NULL	
Type_local: VARCHAR(20) NOT NULL	
Occupation: VARCHAR(20) NOT NULL	
Type_contrat: VARCHAR(20) NOT NULL	
Formule: VARCHAR(20) NOT NULL	
Valeur_declaree_biens: VARCHAR(20) NOT NULL	
Prix_cotisation_mensuel: INTEGER NOT NULL	

Region	
Code_dep_code_commune: VARCHAR(10) NOT NULL [ PK ]	
reg_code: INTEGER NOT NULL	
reg_nom: VARCHAR(20) NOT NULL	
aca_nom: VARCHAR(20) NOT NULL	
dep_nom: VARCHAR(20) NOT NULL	
com_nom_maj_court: VARCHAR(20) NOT NULL	
dep_code: VARCHAR(10) NOT NULL	
dep_nom_num: VARCHAR(20) NOT NULL	

La **transcription** du schéma relationnel en Base de données est établie via les **commandes SQL** « create table » afin d'être chargée dans le **SGBD** avec ensuite toutes les **données** d'entrée correspondantes de contrats et d'info géographique.

```
CREATE TABLE Region (  
    Code_dep_code_commune VARCHAR(10) NOT NULL,  
    reg_code INTEGER NOT NULL,  
    reg_nom VARCHAR(20) NOT NULL,  
    aca_nom VARCHAR(20) NOT NULL,  
    dep_nom VARCHAR(20) NOT NULL,  
    com_nom_maj_court VARCHAR(20) NOT NULL,  
    dep_code VARCHAR(10) NOT NULL,  
    dep_nom_num VARCHAR(20) NOT NULL,  
    CONSTRAINT Region_pk PRIMARY KEY (Code_dep_code_commune)  
);  
  
CREATE TABLE Contrat (  
    Contrat_ID INTEGER NOT NULL,  
    No_voie INTEGER NOT NULL,  
    B_T_Q CHAR(1),  
    Type_de_voie VARCHAR(20),  
    Voie VARCHAR(30),  
    Code_dep_code_commune VARCHAR(10) NOT NULL,  
    Code_postal INTEGER NOT NULL,  
    Surface INTEGER NOT NULL,  
    Type_local VARCHAR(20) NOT NULL,  
    Occupation VARCHAR(20) NOT NULL,  
    Type_contrat VARCHAR(20) NOT NULL,  
    Formule VARCHAR(20) NOT NULL,  
    Valeur_declaree_biens VARCHAR(20) NOT NULL,  
    Prix_cotisation_mensuel INTEGER NOT NULL,  
    CONSTRAINT Contrat_pk PRIMARY KEY (Contrat_ID)  
    FOREIGN KEY (Code_dep_code_commune) REFERENCES Region (Code_dep_code_commune)  
);
```

# Marché des assurances habitation: analyse et interrogation de la base de données

- La base de données étant établie, nous pouvons maintenant interroger, récupérer, et croiser les données nécessaires à l'analyse des marchés d'assurances habitation !
- Nous pouvons filtrer les données souhaitées et ainsi nous concentrer sur une **Region particulière** par exemple les **Pays de la Loire**, afin de croiser les données des contrats qui ont une formule de type « **Intégrale** » grâce à la **relation établie** entre les 2 tables suivant le bon **Code\_dep\_code\_commune** commun correspondant:

```
select count(Contrat_ID), Formule,  
Region.reg_nom from Contrat,Region  
  
where  
Contrat.Code_dep_code_commune=Region.C  
ode_dep_code_commune  
  
And Region.reg_nom="Pays de la Loire"  
And Contrat.Formule="Integral";
```

	count(Contrat_ID)	Formule	reg_nom
1	589	Integral	Pays de la Loire

- Cette analyse nous permet ainsi de **déduire le nombre de contrat Intégral** pour cette région particulière des **Pays de la Loire** s'établissant à **598 contrats**!

# Marché des assurances habitation: analyse et interrogation de la base de données

- Nous pouvons maintenant **interroger, récupérer, et croiser les données** nécessaires à l'analyse des marchés d'assurances habitation dans chaque **Régions française**.
- La table **Régions** nous permet d'avoir une vue d'ensemble **exhaustive de toutes les Régions de France** en sélectionnant **toutes** les Régions distinctes de la table :

```
Select DISTINCT reg_nom from  
Region;
```

reg_nom
1 Auvergne-Rhône-Alpes
2 Hauts-de-France
3 Provence-Alpes-Côte d'Azur
4 Grand Est
5 Occitanie
6 Normandie
7 Nouvelle-Aquitaine
8 Centre-Val de Loire
9 Bourgogne-Franche-Comté
10 Bretagne
11 Pays de la Loire
12 Ile-de-France
13 Guadeloupe
14 Martinique
15 Guyane
16 La Réunion
17 Collectivités d'outre-mer
18 Mayotte
19 Corse

- Ainsi nous pouvons comptabiliser le **nombre de contrat par Région de France** en **croisant** les données avec la table définissant les *Contrats* grâce à la relation établie **entre les 2 tables** suivant le bon Code\_dep\_code\_commune commun correspondant , en **regroupant au préalable** par chaque **Nom de Region** et en **comptant** ensuite le **nombre de contrat** pour chacune d'entre elles, en présentant l'information finale classée par ordre **décroissant**:

```
Select Region.reg_nom,count(Contrat_ID) from  
Contrat,Region  
WHERE  
Contrat.Code_dep_code_commune=Region.Code  
_dep_code_commune  
GROUP BY Region.reg_nom  
ORDER BY count(Contrat_ID) DESC;
```

reg_nom	count(Contrat_ID)
1 Ile-de-France	14177
2 Provence-Alpes-Côte d'Azur	3279
3 Auvergne-Rhône-Alpes	3042
4 Nouvelle-Aquitaine	2038
5 Occitanie	1609
6 Pays de la Loire	1196
7 Hauts-de-France	1189
8 Bretagne	947
9 Normandie	824
10 Grand Est	769
11 Centre-Val de Loire	598
12 Bourgogne-Franche-Comté	293
13 Corse	247
14 Martinique	73
15 Guyane	37
16 La Réunion	3

- On distingue clairement et largement la Région « **Ile de France** » en tête du classement présentant **le plus de contrat d'assurances parmi toutes les régions de France** !

# Marché des assurances habitation: analyse et interrogation de la base de données

- En analysant plus en détail, cette fois ci à l'échelle des Communes et ayant plus de 150 contrats, toujours en croisant les 2 tables en **regroupant au préalable** par chaque **Nom de Commune** et en restreignant **le nombre de contrat supérieur ou égale à 150** pour chacune d'entre elles, en présentant l'information finale classée par ordre **décroissant**:

```
Select DISTINCT Region.com_nom_maj_court,  
count(Contrat_ID) from Contrat,Region  
WHERE  
Contrat.Code_dep_code_commune=Region.Code_dep_  
_code_commune  
GROUP BY Region.com_nom_maj_court  
HAVING count(Contrat_ID)>=150  
ORDER BY count(Contrat_ID) DESC;
```

	com_nom_maj_court	count(Contrat_ID)
1	PARIS 18	515
2	PARIS 17	468
3	PARIS 15	407
4	PARIS 16	394
5	NICE	387
6	PARIS 11	381
7	PARIS 20	302
8	BORDEAUX	302
9	NANTES	291
10	PARIS 19	266
11	PARIS 10	263
12	PARIS 12	252
13	PARIS 14	222
14	GRENOBLE	220
15	PARIS 9	204
16	TOULOUSE	187
17	TOULON	170
18	COURBEVOIE	163
19	LILLE	161
20	PARIS 3	159

- On distingue plus précisément les 4 communes (ou plutôt arrondissements) de la Région « **Ile de France** » vu précédemment que sont **Paris 18-ème / Paris 17-ème / Paris 15-ème / Paris 16-ème** en tête du classement **du plus grand nombre de contrat d'assurances!**
- Suivant le même concept d'analyse, le **prix moyen ou la cotisation est le plus élevé** se trouve également dans ce département d'Ile de France: **Paris / Hauts de Seine / Val-de-Marnes / Yvelines**

```
Select Region.dep_nom,AVG(Prix_cotisation_mensuel)  
from Contrat,Region  
WHERE  
Contrat.Code_dep_code_commune=Region.Code_dep_  
_code_commune  
GROUP BY Region.dep_nom  
ORDER BY AVG(Prix_cotisation_mensuel) DESC  
LIMIT 10;
```

	dep_nom	AVG(Prix_cotisation_mensuel)
1	Paris	36.39908732628086
2	Hauts-de-Seine	26.26935840707965
3	Val-de-Marne	19.82416396979504
4	Yvelines	18.88806970509383
5	Rhône	18.48551829268293
6	Ain	18.23888888888889
7	Alpes-Maritimes	18.13762287756926
8	La Réunion	17.33333333333333
9	Charente-Maritime	17.31730769230769
10	Haute-Savoie	17.14873417721519



# Marché des assurances habitation: analyse et interrogation de la base de données

- Si l'on compare nos résultats précédents de cotisation les plus élevés avec **la moyenne National des cotisations mensuelles** tous contrats confondus, on constate que les **3 départements d'Ile De France** (Paris / Hauts de Seine / Val-de-Marnes) sont bien au-dessus de cette moyenne Nationale s'établissant autour des **19 Euros**

```
select AVG(Contrat.Prix_cotisation_mensuel)
from Contrat;
```

	AVG(Contrat.Prix_cotisation_mensuel)
1	19.33066191748293

- En continuant notre analyse sur le département de Paris, nous pouvons estimer également que la surface moyenne des contrats à Paris est d'environ **52 m<sup>2</sup>**

```
select AVG(Surface) from Contrat,Region
where
Contrat.Code_dep_code_commune=Region.Code_dep_
code_commune and Region.dep_nom="Paris";
```

	AVG(Surface)
1	51.76954988591579



# Marché des assurances habitation:

## conclusion

- Après avoir intégré dans notre base de données les données brutes d'entrée csv des contrats d'assurance et des régions de France, nous pouvons exploiter ces données et croiser ces informations de « Contrat – Region » pour définir les principales tendances en France :
  - la Région « Ile de France » arrive en tête du classement présentant le plus de contrat d'assurances parmi toutes les régions de France
  - Plus précisément 4 arrondissements de la Région « Ile de France » Paris 18-ème / Paris 17-ème / Paris 15-ème / Paris 16-ème arrive en tête du classement du plus grand nombre de contrat d'assurances
  - Les 3 départements d'Ile De France (Paris / Hauts de Seine / Val-de-Marnes) sont bien au-dessus de la moyenne Nationale des cotisations mensuelles s'établissant à 19 Euros
  - La surface moyenne des contrats à Paris est d'environ 52 m<sup>2</sup>
- => **Dans la région Ile de France et ses départements se concentrent donc les marchés les plus significatifs et importants dans le domaine de l'assurance habitation!**

