

Credit Risk

Credit Risk	0
Introduction	0
Description des données	1
Variables figées	1
Variables évolutives	1
Analyse Descriptive	2
Les variables figées	2
Les variables binaires	2
Les variables continus	3
Variables Evolutives	5
Variable Target	7
Le profil type du mauvais payeur	8
Traitement des données	8
K-Plus-Proches Voisin	11
Régression Logistique	11
Random Forest	11
Classifieur Bayésien Naïf	12
Résultats	12
CONCLUSION :	14

Introduction

Dans le cadre de notre cours de risque de crédit nous avons appris à comprendre les problèmes inhérent à l'industrie du risque appliqué cette fois ci aux crédits. Pour l'examen de cette matière, nous allons travailler sur une base qui suit les prêts immobiliers. Dans un premier temps nous allons proposer une analyse des données accompagné d'une description et proposer une façon de les structurer. Nous allons ensuite nous attarder à sélectionner les variables qui nous paraissent pertinente pour, derrière construire un modèle. Ce modèle nous permettra d'évaluer la probabilité de défaut. Nous utiliserons ces probabilités de défaut pour construire des classes homogènes de risque.

Cette approche classique sera enrichie par des méthodes propre au machine learning pour apporter un angle d'étude différent à notre travail. Ces méthodes seront définis plus amplement dans notre papier.

Description des données

Nous travaillons ici sur une base de 50 000 résidents américains que l'on suit sur 60 périodes. A titre d'information il est important de rappeler que les prêts ne commencent pas nécessairement à la période 0. Il est également important de noter qu'une fois que le résident est rentré en défaut, le suivi est arrêté. De même

pour les gens qui ont payé leur obligation.

Pour suivre les résidents on a six variables qui suivent le client à travers le temps et 9 variables qui sont figées à la date d'origine du prêt. On va décomposer notre étude des variables entre les variables figées et les variables structurelles :

Variables figées

Ces variables figées donnent le portrait de la situation à l'origine de ce crédit.

Pour les 8 variables figées nous avons d'abord des variables binaires, valant 1 si le résident respecte ou non la condition :

- Pour savoir si le type de bien est un appartement via REtype_CO_orig_time
- Pour savoir si le bien fait partie d'un plan d'urbanisme REtype_PU_orig_time
- Pour savoir si il s'agit d'une maison pour une seule famille via REtype_SF_orig_time
- Pour savoir si il s'agit d'un investisseur emprunteur, c'est à dire un agent économique qui investit et emprunte dans notre situation via investor_orig_time

On a également un FICO_score_orig_time. Le FICO score est un modèle de scoring à l'échelle des Etats Unis qui renseigne le risque de l'emprunteur. A titre d'information il faut savoir que le score va entre 300 et 850 et qu'un score compris 670 et 739 est considéré comme une bonne note.

La note se construit approximativement avec les critères suivants , à savoir que la méthode évolue avec le temps : historique des paiements (35%), comptes détenus (30%), longueur de l'historique de crédit (15%), mix des crédits détenus (10%), créations récentes de crédits(10%).

On retrouve également certaines métriques des variables évolutives : Le prix de la maison indicé en base 100, l'encours restant dû, l'encours restant dû relativement au prix de la maison (non indicé) et le taux d'intérêt.

Variables évolutives

Les 6 variables évolutives comprennent d'un côté le bilan du prêt c'est à dire l'encours restant, la valeur indicé en base 100 du prix de la maison pour faire ressortir l'évolution du prix et éviter les effets de tailles, c'est à dire corriger une évolution de prix de 300 000\$ sur un bien qui valait 1 000 000, et le ratio de l'encours restant du prêt sur la valeur du bien (LTV) qui corrige donc l'encours du prêt restant par la valeur du prix du bien.

On a également sur ces 6 variables des informations sur le marché, c'est à dire sur la santé économique du pays. On trouve alors le taux de chômage à date et la croissance du PIB.

Sur des emprunts à taux variables on est également capable de suivre le taux d'intérêt. Ce taux d'intérêt qui découle de l'économie du pays mais qui est reporté à chaque individu.

Il nous paraît important ici de noter pour la suite de notre étude qu'il est fort à parier qu'on note une corrélation élevée entre la croissance du PIB et le chômage et aussi avec le taux d'intérêt. Il faudra sans doute effectuer un choix.

Nous ne considérons pas ici la variable défaut car elle fait partie de la variable cible qui est une variable binaire qui enregistre si le payer est entré en défaut ou non. De même la variable payoff, qui se rapproche d'un proxy de cette variable n'est pas ici pris en compte. De ce fait Status_time qui est une variable catégorielle construite sur payoff et status ne sera pas, non plus prise en compte dans notre analyse.

Analyse Descriptive

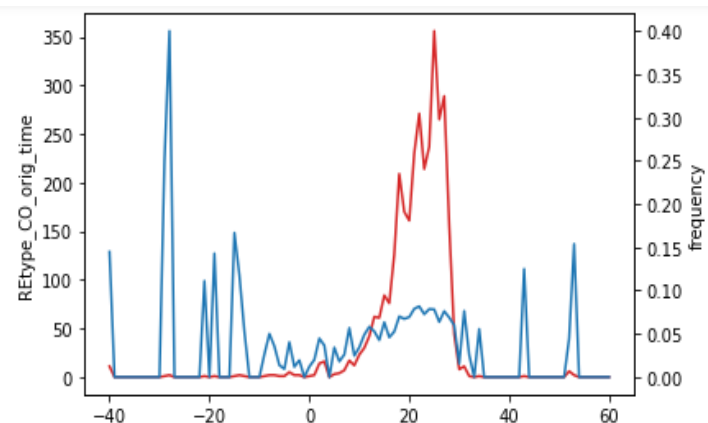
Pour mieux restituer les ordres de grandeurs ici présent nous accompagnerons cette section par des éléments graphiques ainsi que des tableaux.

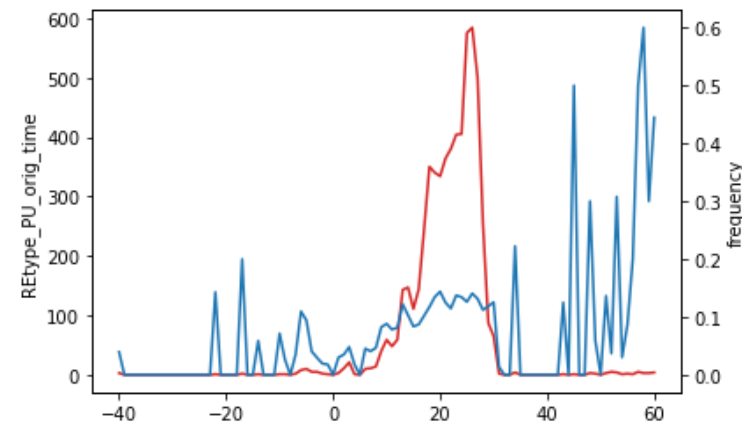
Nous nous attacherons à d'abord restituer les variables figées qui font l'état des emprunteurs à l'origine en figeant à la première date et ensuite nous nous concentrerons sur les variables fluctuantes pour conclure pour les variables cibles. Nous proposerons nos approches 'restrictives' dans un deuxième temps.

Les variables figées

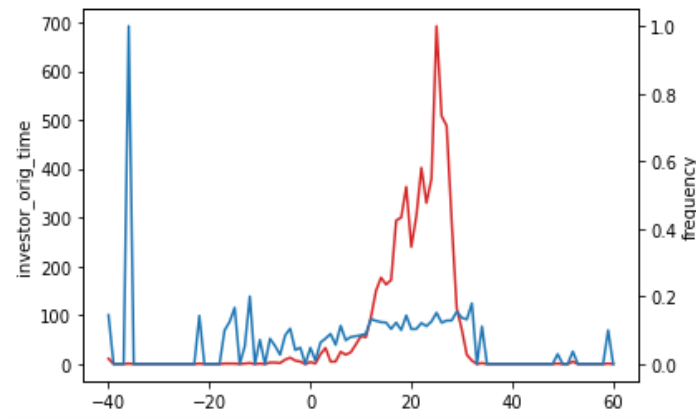
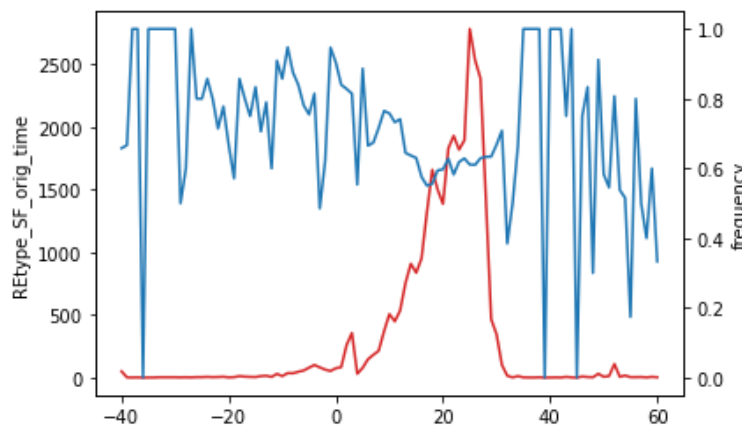
Les variables binaires

Dans une chronologie similaire, commençons par étudier la fréquence des variables binaire. On va mettre en parallèle les évolutions des différentes variables binaire en fonction de la croissance du nombre de prêt qui commencent.





D'un côté l'axe orange représente le nombre de prêt octroyé qui répondent à la condition, c'est à dire la dummy sélectionnée vaut 1. Et de l'autre côté on a la fréquence qui représente donc le pourcentage de prêt octroyé qui répondent à la condition. Attention il est important ici de remarquer qu'il y a un effet taille qui laisse à croire que le



nombre de prêt qui répond à chacune des conditions est croissant, mais si on regarde la fréquence cette règle ne se confirme pas. En effet les prêts octroyés augmentent drastiquement pour atteindre leur maximum à la période 20. Ainsi dans notre analyse on s'attachera à étudier les 4 variables binaires ci dessus que dans l'intervalle entre 0 et 20+ pour avoir un échantillon suffisamment significatif pour parler réellement de fréquence.

Si on s'intéresse à la dummy qui renseigne le plan urbain d'aménagement dont bénéficie ou non l'emprunteur on se rend compte que ce pourcentage est nettement croissant à travers le temps. Cette analyse se vérifie sensiblement aussi sur le type de bien acheté avec le prêt, on voit ici que plus le temps passe plus la fréquence d'achat des appartements augmente. Cette analyse n'est pas vraie en revanche pour le nombre de famille seule. Les gens achètent de plus en plus avec des familles nombreuses même si la tendance n'est pas très nette sur l'intervalle, elle n'est pas non plus négligeable. Finalement pour ce qui est de l'emprunteur on voit aussi une fréquence croissante sur l'intervalle 0:20+ mais ce n'est pas nécessairement flagrant, on préférera ne pas en tirer de conclusion hâtives.

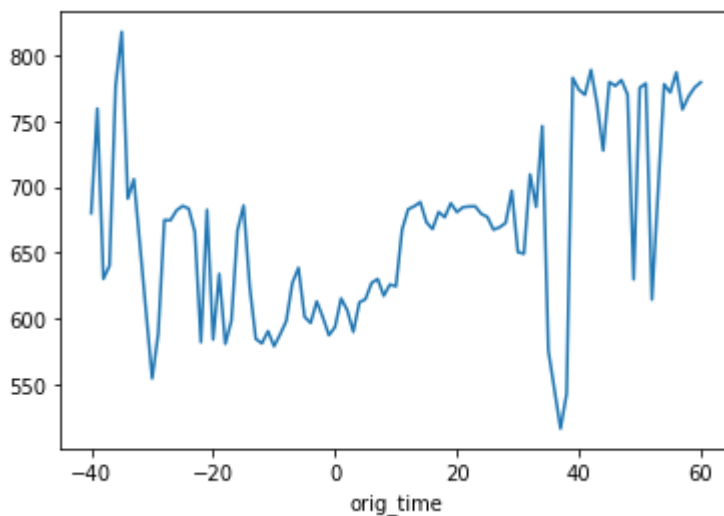
Les variables continus

Nous allons maintenant étudier les variables continus de notre table:

FICO :

Le FICO score vaut en moyenne 670 ce qui en moyenne est donc une bonne note pour notre fond de commerce. Ce dernier est compris entre 516 et 818 avec 50% du fond de

commerce compris entre 614 et 703.

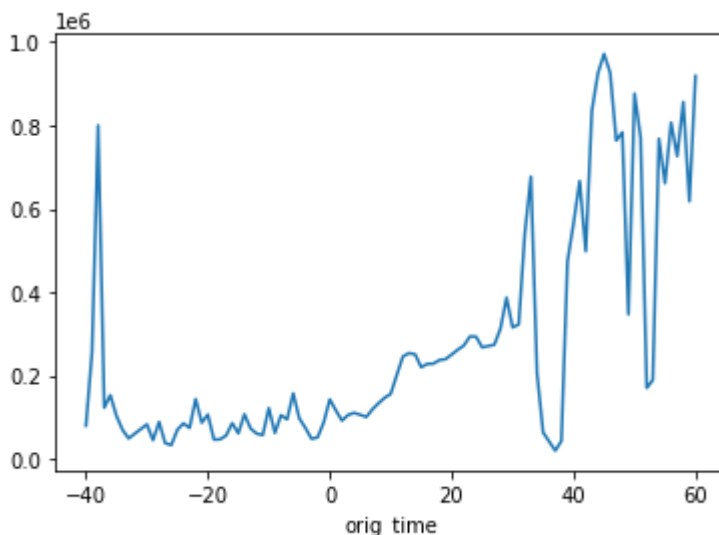


Si on s'attache à garder l'intervalle d'étude entre 0 et 20+ on voit que le FICO des clients augmentent avec le temps.

Atteignant au pic 700 en moyenne ce qui est pour rappel la tranche haute des bonnes notes alors qu'au début on observe un niveau raisonnablement faible.

Balance :

Si on étudie l'encours du des prêts on observe qu'en moyenne il est à \$281.000. Après avoir observé des piques à \$8.000.000 nous avons pris la décision d'éliminer les valeurs dépassant le quantile d'ordre 0.999 qui dépassait les \$2.000.000. On voit que l'encours est croissant soit car le prix de l'immobilier augmente soit parce que les gens achètent des biens de plus en plus cher. Ce qui serait cohérent avec le niveau des FICO qui augmente. Nous arrivons sur des clients avec une gamme nettement plus élevée.



LTV_orig_time :

Les clients empruntent en général pour 77% de la valeur du bien avec un minimum à 52% et un maximum à 94%. Il est fort à parier que le 94% s'explique par une maison qui aurait baissé en prix augmentant alors le LTV. En effet comme expliqué auparavant le LTV défini

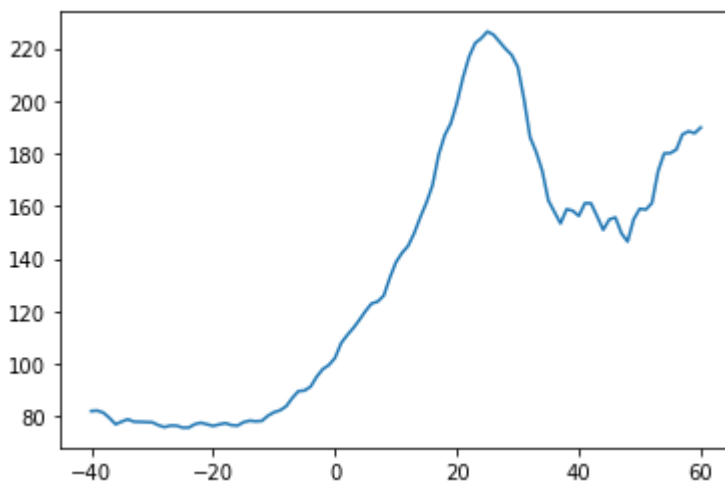
l'encours restant sur le prix de la maison. Ainsi si ce dernier baisse le LTV augmente ce qui correspond à une dégradation de l'emprunt.

La LTV de façon cohérente avec nos analyses ci dessus diminue, on fait face à des emprunteur de meilleur qualité.

Interest_rate :

De façon cohérente avec le marché américain le taux d'intérêt moyen à l'origine est aux environs de 6,5% ce qui paraît conséquent dans le contexte de taux actuel mais est représentatif de la situation à ce moment là. Toutefois il paraît intéressant de noter qu'au vu du maximum soit 12,75%, il est probable que l'on soit en capacité d'identifier la période à laquelle fait partie cette base de donnée. Des taux aussi conséquents font directement allusion aux subprimes. Nous étudierons les défauts pour voir si cette hypothèse est cohérente.

Hpi_orig_time :



En moyenne à l'origination du prêt les biens avaient un prix dépassant les 130% de valeur initiale. Cette indice pouvant atteindre jusqu'à 225% et qui pour plus de 50% des foyers dépasse les 100%.

Si on suit l'évolution on voit que l'immobilier a dans l'ensemble vraiment perdu de la valeur mais la chute du prix de l'immobilier (pour presque 50 points) après le pic d'octroi des prêts évoque encore une fois les subprimes.

Variables Evolutives

Ici on s'attache à décrire les variables qu'on appelle 'evolutives' c'est à dire qui sont enregistrées à chaque date et qui évoluent en fonction du temps. Pour procéder à leur analyse on ne part plus cette fois-ci de orig_time mais de time. De façon similaire on va figer notre base de donnée à chaque date en agrégeant par moyenne pour chaque variable.

'Balance_time':

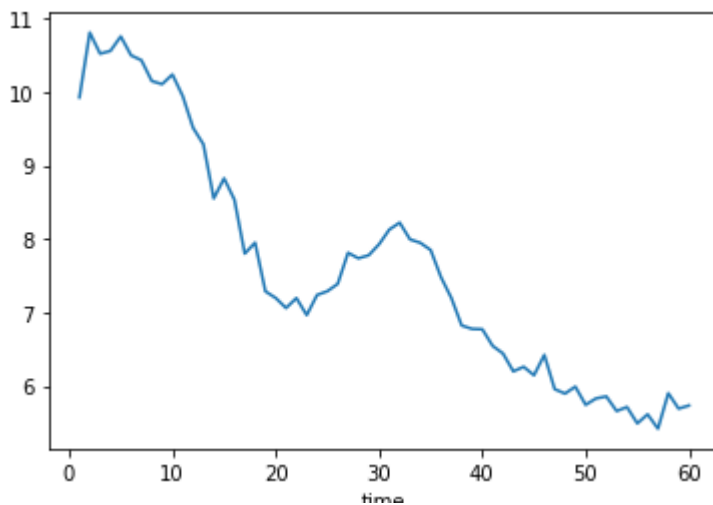
L'encours restant est en moyenne inférieur à celui à l'origine, cette conclusion est évidemment rassurante car elle signifie que les choses se sont bien passées, les gens ont en moyenne plutôt payé leur prêt. Cependant cet encours est croissant avec le temps, en effet si les gens ont payés leur nombre de personnes qui sont rentrés en défaut est par définition croissant, plus le temps passe plus les gens ont des chances de rentrer en défaut ce qui explique que l'encours en moyenne augmente.

LTV:

L'analyse ci-dessus est sensiblement identique à celle de la loan to value, pour autant en moyenne la variable nous indique que les gens ont des emprunts qui ont soit pas suffisamment baissé en encours soit qui se sont vu piégé par une baisse de la valeur de leur bien avec un LTV en moyenne à 78% contre 77% à l'origine. On assiste à une très forte croissance de la métrique à travers le temps dans un premier lieu jusqu'à la période 40 pour ensuite suivre à un épisode d'affaissement de cette LTV.

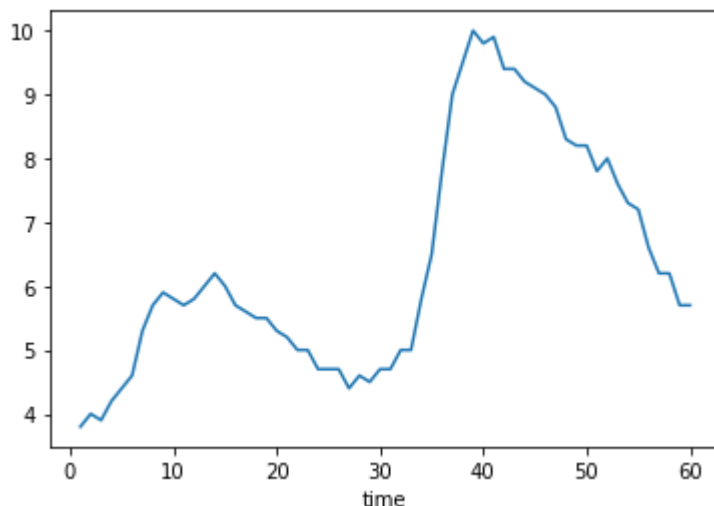
Interest_rate_time

Lorsqu'on suit l'évolution du taux d'intérêt on voit qu'en moyenne ce dernier s'affaisse drastiquement. On passe d'un taux d'intérêt moyen à +10% pour retomber à 6%. A savoir qu'en moyenne on approche les 7,5%



Uer_time :

Lorsqu'on suit le graphique du chômage on voit qu'il explose à la 40ème période. Pour froter les 10%, un scénario qui est impressionnant d'autant plus qu'on observe une explosion du chômage aux états unis en 5 à 10 périodes passant du simple au double.

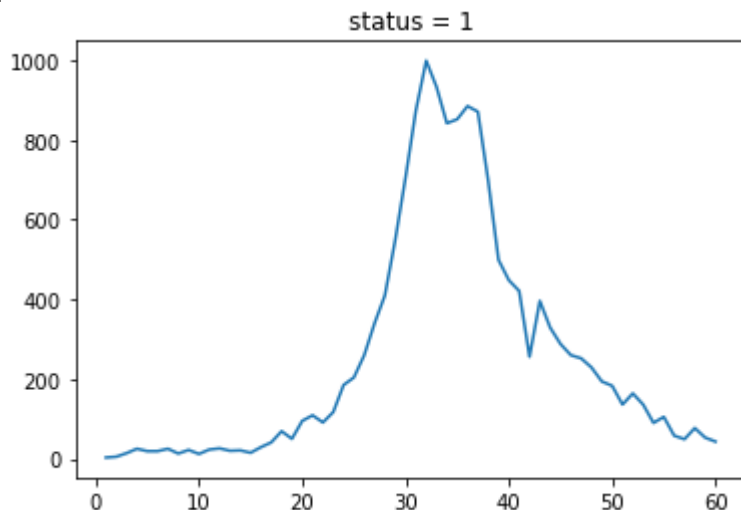


Pour le house price index et le taux de croissance du pib les comportements vont être identiques à ceux observés en orig_time, ainsi nous n'iront pas dans le détail.

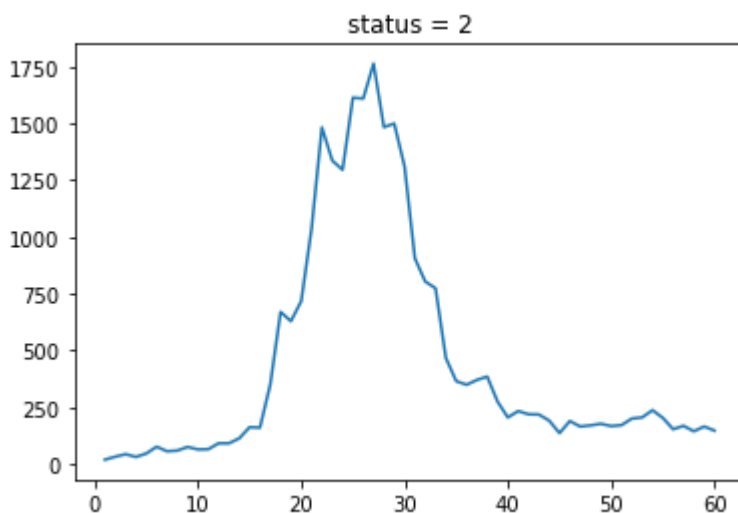
Variable Target

La base est compos e de 3 variables qu'on caract risera de cibles, l'exercice ici est de se concentrer sur les d faits des payeurs car ils sont ceux qui ont le plus d'impact sur le bilan de l'entreprise.

On observe qu'il y a pour 8250 personnes qui n'ont pas encore fini leur pr t   la derni re date enregistr e, il y a 15137 personnes en d faut et 26545 personnes qui ont r gl  leur pr t.



Le pic de d faut se fait vers la 30 me p riode avec 1000 d faits presque par p riode soit 2% des clients qui entrent en d faut en une p riode et ceux de fa on r currente pendant presque 10 p riodes.



On observe que les gens  taient de relativement p riode avant la p riode ou on enregistre le pic catastrophique de d faut.

Ces deux graphiques nous incitent   penser que notre analyse issue des variables fig es para t tr s plausible, d'abord confirm e par un d faut de 36% relativement aux gens qui ont pay  ou qui sont entr s en d faut et puis parce que ce d faut c'est enregistr  de fa on brusque dans une p riode qui para t troubl e avec un contexte de taux qui est reparti   la hausse et surtout un taux de ch mage qui explose.

En conclusion nous travaillons sur une base très volatile qui change d'un contexte familial qui a tendance à se rapprocher à un ordre compris entre 2 et 5%.

Le profil type du mauvais payeur

Pour définir ce qu'est un mauvais payeur nous allons figer une image de leur état à la date où ils sont entrés en défaut. Nous raisonnerons ensuite en moyenne pour définir le profil type.

Ainsi un mauvais payeur détient un encours assez faible de l'ordre de \$250.000 avec un écart assez faible entre le min et le 75%. Il a en moyenne un score FICO de 650 à savoir que son FICO peut monter à 810 mais qu'il est à 50+% des cas inférieur à une bonne note.

En moyenne il a acheté sa maison à un moment où les prix étaient très haut 210 de l'indice de prix qui signifie qu'il a acheté à un moment sur un marché potentiellement une bulle. Le contexte de taux dans lequel il a fait l'acquisition est plutôt élevé de l'ordre de 6%.

63% des mauvais payeurs sont des familles seuls mais seul 7% ont acheté un appartement ce qui est un peu en deçà de la moyenne habituelle.

En moyenne il est entré en défaut à une période où l'économie n'avait quasiment plus de croissance (0.5%) et 75% des cas étaient dans un cas où l'économie était à bas régime <2%. Le taux de chômage était relativement élevé aux alentours des 7% et toujours au-dessus des 5% de chômage.

On note également que le hpi_time est bien plus dégradé que le hpi_orig_time ce qui traduit l'acquisition d'un bien qui s'est largement déprécié.

De même en moyenne les encours n'ont été quasiment pas remboursés, c'est à dire que sur les \$251.500 d'encours moyen on est à \$250.000 restant en moyenne.

On note également que ce sont des gens avec une maturité très longue sur leur emploi pour 120 périodes convaincu que le marché immobilier est en bonne santé ils se permettent d'acheter de façon très exposée souvent à taux variable ce qui expose d'autant plus le foyer. On comprend ainsi mieux l'encours restant qui est significativement élevé.

Traitement des données

Pour sélectionner les données on va d'abord décider de traiter les données par le biais de l'information value. On définira un seuil minimum à partir duquel on conserve la variable. Suite à quoi on convertira les variables en classes. Ces classes représenteront des déciles de la variable. Ces variables seront réassemblées pour former des classes plus homogènes. Pour quantifier l'homogénéité des classes nous allons procéder par le WoE. Ce weight of evidence représente l'impact de la classe sur la variable cible, ici le défaut du payeur. Nous avons d'abord calculé l'IV de chaque variable sauf les variables binaires qu'on traitera à côté.

$$IV = \sum (\text{Event\%} - \text{Non Event\%}) * \ln \left(\frac{\text{Event\%}}{\text{Non Event\%}} \right)$$

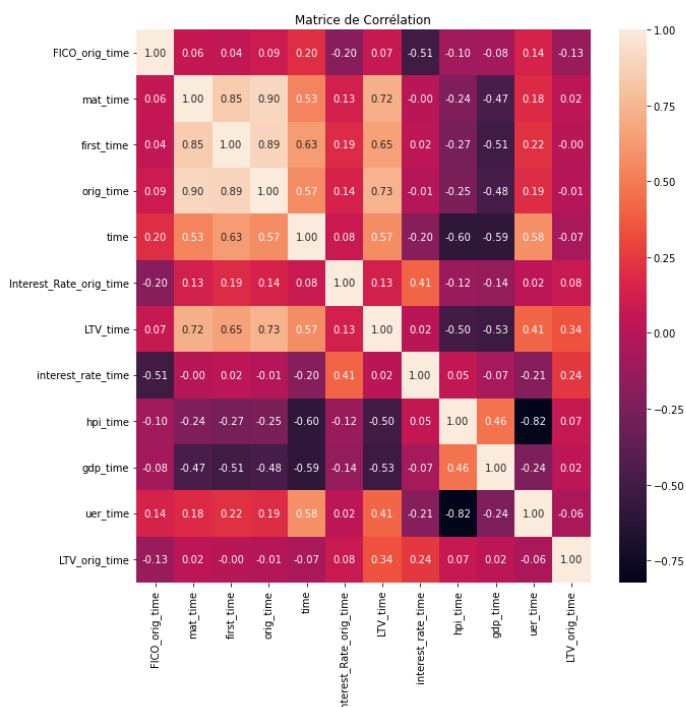
$$WOE = \ln \left(\frac{\text{Distribution of Goods}}{\text{Distribution of Bads}} \right)$$

On a d cid  de ne garder que les IV qui d passaient 0.05, en effet il est de coutume de traiter les IV avec les intervalles suivant :

Information Value (IV)	Variable's Predictive Power
< 0.02	Not useful for prediction
0.02 – 0.1	Weak predictive power
0.1 – 0.3	Medium predictive power
0.3 – 0.5	Strong predictive power
> 0.5	Too good to be true! Check the variable

Un rapide controle de l'information value des variables binaires nous a amen   les  liminer de notre cadre d' tude.

Toutefois avec une IV nous conversions 13 variables ce qui  tait suffisant   notre avis. Apr  s avoir gard   ses 13 variable nous avons observ   les corr  lations suivantes :



On d cide de poser des filtres, seuil   partir duquel la corr  lation est consid  r   comme trop significative pour conserver les deux variables : ici ce seuil   t   fix      0.75, ce qui est tr  s significatif. Ce filtre    mis en avant 4 duo qui pouvait avoir une corr  lation mettant en danger le mod  le:

uer_time/hpi_time
mat_time/orig_time
mat_time/first_time

Avec une approche qualitative et en gardant comme r f rentiel de comparaison l'information value de chaque variable, nous avons donc d cid  de garder la maturit  et l'house price index et d' liminer les autres variables.

Nous avons donc affaire   maintenant plus que 10 variables continues, certaines  tant des valeurs enti res et d'autres des nombres   virgule.

Afin de traiter les variable avec 10 classes il est recommand  de recomposer les variables par WoE. En effet le WoE transcrit la fa on dont laquelle la cat gorie diff rencie les bons des mauvais payeurs c'est   dire   quel degr  et surtout dans quel sens. En faisant des rapprochements on peut alors reconstruire les cat gories en entre 2 et 4 classes.

Le WoE pour une variable continu est lin aire c'est   dire qu'en moyenne il cro t ou d cro t de fa on continu. Ainsi les classes sont par n cessit  des s ries de d ciles.

```
1.FICO : 0-5,6-7,8-9 ->3
2.mat_time : 0-3,4-6,7-9 ->3
4.Interest_Rate_orig_time : 0-2,3,4-7 -> 3
5.LTV_time : 0-4,5-6,7-10 -> 3
6.hpi_time : 0-2,3-6,7-9 -> 3
7.gdp_time : 0-1,2-4,5-9 -> 3
8.time : 0-3,4,5-9 ->3
9.interest_rate_time : 0-2,3-9 ->2
10.LTV_orig_time : 0-1,2-6 -> 2|

SUM : 25 Dummies
```

Ces 25 variables binaires ne sont pas gard s tel quel. En effet  tant

des variables binaires il est important d' viter le challenge de la multicollin arit . Pour ce faire on retire une des trois segmentations. Ici le but  tant de cr er des classes qui apporte un maximum d'information et ayant un profil tr s discriminant on aura comme coutume de retirer la classe du 'milieu' sous reserve qu'il y en ai 3 et sinon celle avec le WoE qui se rapproche le plus de 0.

Au final nous recr ons la base suivante :

```
0 FICO : 399-681 .
1 FICO : 726+ - 840 .
2 mat : <=139 .
3 mat : >=145 .
4 rate int ot : <=6.4 .
5 rate int ot : >6.9 .
6 LTV: <=74 .
7 LTV : >=95 .
8 hpi : <=162 .
9 hpi : >=218 .
10 gdp : <=0.8 .
11 gdp : >=2.3 .
12 time : <=28 .
13 time : >=31 .
14 rate int t : <=6.4 .
15 LTV ot : <=75 .
dtypes: int32(16)
```

Puisqu'on avait 25 classes pour 10 variables et que par soucis de multi colinearit  on doit retirer une classe on obtient alors un total de 15 classes.

Cette base sera r utilis e pour le mod le.

Modélisation

K-Plus-Proches Voisin

L'algorithme des K-Plus-Proches Voisin est un algorithme de classification et de régression. La méthode des K-plus proches voisins consiste à affecter, à un point à classer, la classe majoritaire de ses K plus proches voisins. On précalcule la distance entre le point à classer et l'ensemble des autres points de toutes les classes en se rappelant d'où vient le point (index de la classe et l'index dans la classe) et on prend les k premiers.

Grâce à cet algorithme, on obtient le résultat suivant : (TN étant la part de non-défauts bien prédits et TP la part de défauts bien prédits).

```
KNeighborsClassifier
```

```
score AUC = 0.7499485405340451
```

```
TN = 0.7047892430880798 & TP = 0.7951078379800105
```

Régression Logistique

La régression logistique est un modèle statistique permettant d'étudier les relations entre un ensemble de variables qualitatives X_i et une variable qualitative Y . Il s'agit d'un modèle linéaire généralisé utilisant une fonction logistique comme fonction de lien.

Grâce à cet algorithme, on obtient le résultat suivant : (TN étant la part de non-défauts bien prédits et TP la part de défauts bien prédits).

Un modèle de régression logistique permet aussi de prédire la probabilité qu'un événement arrive (valeur de 1) ou non (valeur de 0) à partir de l'optimisation des coefficients de régression. Ce résultat varie toujours entre 0 et 1. Lorsque la valeur prédite est supérieure à un seuil, l'événement est susceptible de se produire, alors que lorsque cette valeur est inférieure au même seuil, il ne l'est pas.

```
LogisticRegression
```

```
score AUC = 0.7680628487372244
```

```
TN = 0.7134008158332075 & TP = 0.8227248816412415
```

Random Forest

Le random forest est composé de plusieurs arbres de décision, travaillant de manière indépendante sur une vision d'un problème. Chacun produit une estimation, et c'est

l'assemblage des arbres de décision et de leurs analyses, qui va donner une estimation globale. En somme, il s'agit de s'inspirer de différents avis, traitant un même problème, pour mieux l'appréhender. Chaque modèle est distribué de façon aléatoire aux sous-ensembles d'arbres décisionnels.

Le random forest est un modèle d'apprentissage, dont l'efficacité dépend fortement de la qualité de l'échantillon de données de départ.

RandomForestClassifier

score AUC = 0.7681411307047388

TN = 0.7146094576219972 & TP = 0.8216728037874803

Classifieur Bayésien Naïf

En termes simples, un classifieur bayésien naïf suppose que l'existence d'une caractéristique pour une classe, est indépendante de l'existence d'autres caractéristiques. Un fruit peut être considéré comme une pomme s'il est rouge, arrondi, et fait une dizaine de centimètres. Même si ces caractéristiques sont liées dans la réalité, un classifieur bayésien naïf déterminera que le fruit est une pomme en considérant indépendamment ces caractéristiques de couleur, de forme et de taille.

GaussianNB

score AUC = 0.769636713654656

TN = 0.7407463363045778 & TP = 0.7985270910047344

Suite à ces différents algorithmes, nous décidons d'utiliser la régression logistique, qui malgré le fait qu'elle n'ait pas le meilleur score de classification, prédit le mieux l'évènement de défaut tout en gardant une prédiction correcte pour ceux qui vont aller au bout de leur remboursement. Le défaut est le plus important à bien classer car il fait perdre beaucoup d'argent à la banque. Il reste important de ne pas discriminer les bons payeurs sinon, la banque perdrait de l'argent accordé moins de crédit.

Avec cet modèle, nous trouvons les coefficients suivants :

Résultats

Intercept = -0.95771213

FICO : 399-681 = 0.4449040617728028

FICO : 726+ - 840 = -0.6542081772155476

mat : <=139 = -0.23688121930635844

mat : >=145 = 0.3233517883131619

rate int ot : <=6.4 = 0.0452493628367293

rate int ot : >6.9 = 0.2086029599822234

LTV: <=74 = -0.7219853715441019

LTV : >=95 = 0.7426361731615214

hpi : <=162 = 0.534682331357623

hpi : >=218 = 0.1396978877422722

gdp : <=0.8 = 0.21939204455638675

gdp : >=2.3 = -0.17707637033873386

time : <=28 = -0.6537032644336747

time : >=31 = 0.5188041275070271

rate int t : <=6.4 = -0.40812601130274906

LTV ot : <=75 = -0.13174061828102313

On remarque qu'un indice FICO haut diminue fortement la probabilité de défaut, à l'inverse un FICO l'augmente un peu moins fortement. On remarque aussi très logiquement, que plus la maturité est longue, plus le taux de défaut est important. A l'origine, quand les taux d'intérêt sont soit très hauts, soit très bas, la probabilité de défaut augmente. On peut remarquer qu'un défaut aura plus tendance à arriver quand le Loan to Value Ratio est plus grand que 95% du prix de la maison.

Grâce à l'estimation des probabilités de défauts, nous pouvons construire différentes notations de risque.

CHR	Probabilité de défaut
0	0.041245

1	0.081061
2	0.109163
3	0.154407
4	0.232496
5	0.337607
6	0.476761
7	0.616263
8	0.742933
9	0.832351

CONCLUSION :

Dans un contexte macro économique très instable, cette table recense les prêts de 50 000 résidents américains sur une période de 60 échéance. Si le niveau de défaut atteint un seuil critique de 35%, il se comprend à travers l'analyse qu'on peut donner des subprimes.

Il devient alors essentiel en tant que banque de réussir à se prémunir face à de tel risque surtout quand ils peuvent exposer les fonds propres des groupes à de tel niveau. C'est ce qui explique aujourd'hui le comportement de provisionnement des banques qui couvrent pour jusqu'à 200% de leur exposition aux défauts aux Etats Unis.

Par le biais de modèle d'étude des fonds de commerces de prêt sur les métriques à la disposition des banques, nous sommes capables d'appréhender de façon tangible et concrète le niveau d'exposition de la banque avec un intervalle de confiance suffisamment large. Dans le cadre de notre projet nous nous sommes attachés à étudier comment les

métriques interagissaient elles les unes entre les autres et à quel seuil certaines métriques pouvaient exposer de façon plus ou moins flagrantes les portefeuilles de prêts des banques. Dans notre modèle nous avons abouti à une estimation de presque 80% des mauvais payeur ce qui signifie que notre modèle estime à 80% correctement un mauvais payeur, soit la personne qui présente le plus de risque pour la banque. Ce niveau correspond aux ordres de grandeur des modèles qu'on peut trouver dans l'activité de credit risk modeling, surtout dans un contexte où l'accès aux données est limité pour les entreprises.

Il permettra alors à la banque de se provisionner correctement pour se prémunir des risques auquel elle peut s'exposer en cas de défaillance de marché pour ces banques qu'on qualifie aujourd'hui de 'too big to fail' et qui se voient imposer des régulations de plus en plus impactante pour leur business.

Toutefois si ce travail et ces régulations sont essentiels à la survie de notre modèle économique il peut être intéressant de s'interroger à la place que ces réglementations pourront avoir à l'avenir sur l'acquisition de biens immobiliers pour les particuliers.