



DREAM HOUSING FINANCE CO LTD.

YOUR ROAD TO A DREAM HOME

Décision rapide d'admissibilité au prêt

**Démarches préalables au choix et
rédaction d'une étude**

Data Analyst







Projet 8

Marc Sellam

Décembre 2020

SOMMAIRE :



	Réflexion préliminaire	P1
	Choix du thème	P2
	Choix du dataset	P3
	Choix du plan	P9
	Analyses graphiques et harmonie	P10
	Conclusion	P12

Réflexion préliminaire:



Mes objectifs :

- Mesurer mes connaissances
- Pratiquer au travers d'un exercice d'analyse et de classification binaire (typique qui est demandé au sein des organisations).
- Dans un domaine faisant partie de mes centre d'intérêts professionnels.

Choix du thème



Domaines de prédilection :

- Jeux vidéo consoles
- Secteur banque d'affaires
- Médical (prévention)

Choix du dataset



Critères auxquels le dataset doit répondre:

- Concerne l'un des domaines de prédilection
- Problématique intéressante
- Données récentes
- Se prête à la classification
- Suffisamment complet
- Contenir suffisamment d'informations
- Variables variées



Ressources utilisées lors de la recherche du dataset:

Moteurs de recherche principaux:

Kaggle.com

data.world

researchportal.port.ac.uk

<https://www.stat4decision.com/fr/10-sites-de-reference-open-data/>

Exemple de datasets examinés (liste non exhaustive) :

https://www.data.gouv.fr/fr/datasets/open-food-facts-produits-alimentaires-ingredients-nutrition-labels/#_

<https://data.sncf.com/explore/dataset/ponctualite-mensuelle-transilien/table/?sort=date>

<https://github.com/GregorUT/vgchartzScrape> (scraping base de donnée JV)

<https://github.com/hpanwar08/data-analysis-videogames>

<https://www.data.gouv.fr/fr/datasets/marche-du-jeu-video/>

<https://www.kaggle.com/mohalim/video-games-sales>

<https://medium.com/@vishnumbaprof/case-study-loan-prediction-ac035f3ec9e4>

<https://data.world/scuttlemonkey/starcraft-2-skillcraft-1>



Dataset sélectionné pour l'étude

Source: analyticsvidhya.com

- <https://datahack.analyticsvidhya.com/contest/practice-problem-loan-prediction-iii/#ProblemStatement>

- un site internet d'exercices en data science

- les données datent du 26-05-2016 à 12 :01.

- correspond au critères fixés



Description sommaire du dataset

Le jeu de données fourni est un tableau de 614 lignes, chacune représentant une candidature à l'emprunt.

Chaque ligne est composée de 13 colonnes :

- 1 colonne contenant l'identification unique de la demande de prêt.
- 11 colonnes numériques, contenant les caractéristiques des candidatures.
- 1 colonne contenant la décision de la banque Y/N (accordé/refusé).



Thématique et problématique du Dataset sélectionné

- Thématique : Comment donner un avantage concurrentiel aux banques traditionnelles en les dotant d'un outil décisionnel correspondant à leur besoin business.
- Problématique : Comment donner rapidement un accord de principe à une candidature au prêt sans que le client ait le temps d'aller à la concurrence.
- Solution proposée : Utilisation d'un algorithme de classification afin de créer un modèle de prédiction.

Choix du plan



Plan défini :

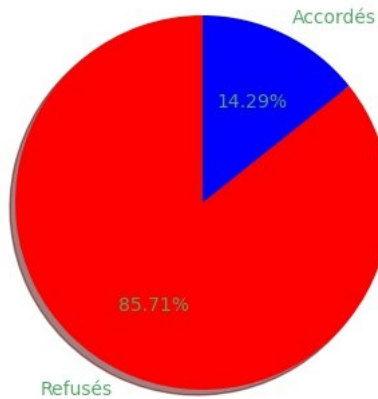
- Sommaire.
- Introduction.
- Description des données.
- Préparation des données.
- Analyses graphiques.
- Création d'un modèle.
- Conclusion.

Analyses graphiques et harmonie

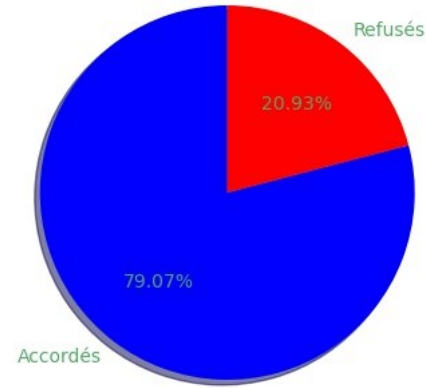


Avant amélioration :

Proportion des réponses aux candidats incapables de rembourser un emprunt

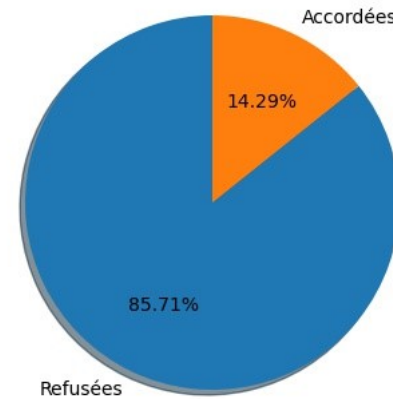


Proportion des réponses aux candidats ayant la capacité à rembourser un emprunt

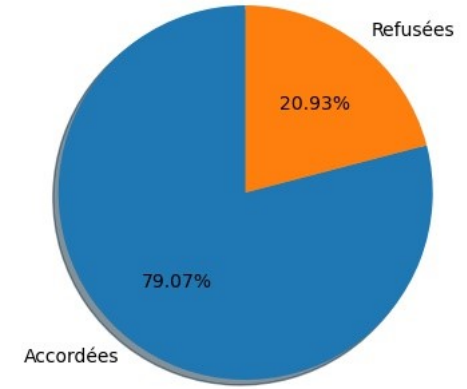


Après amélioration :

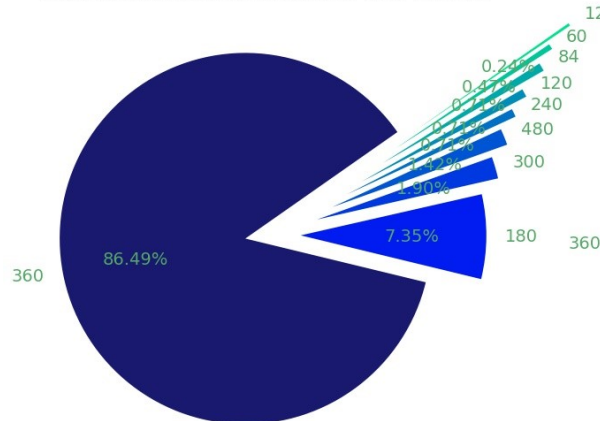
Répartition des réponses aux candidats dont le Crédit_History est défavorable



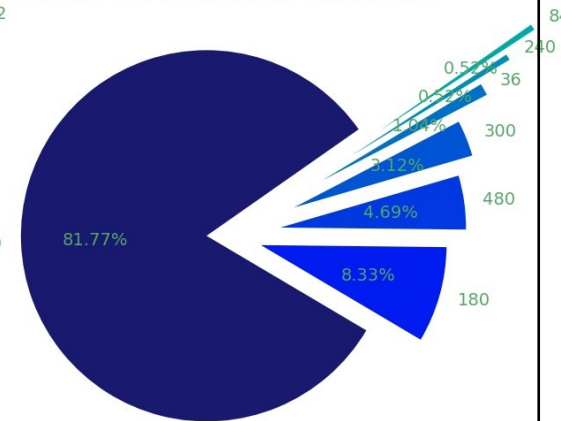
Répartition des réponses aux candidats dont le Crédit_History est favorable



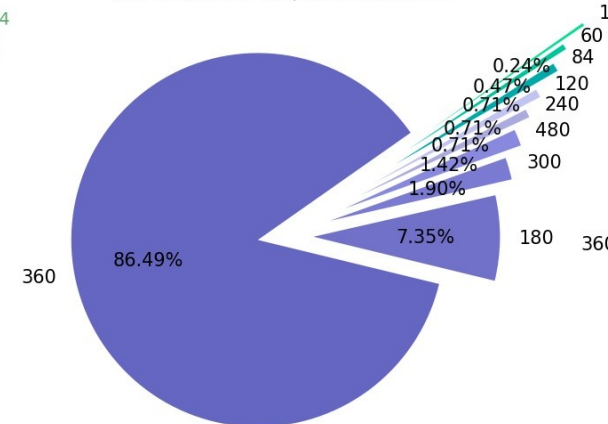
Proportions des mensualités pour les prêts accordés



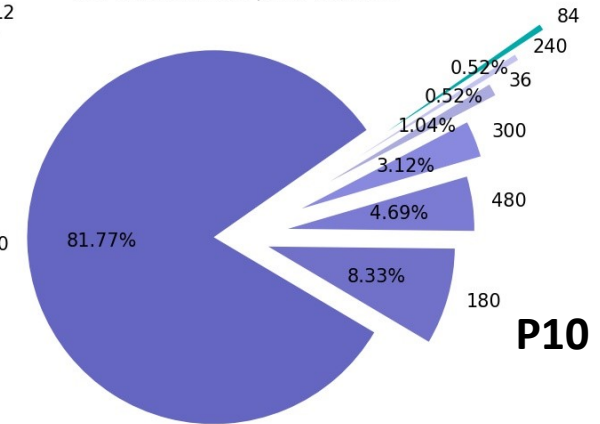
Proportions des mensualités pour les prêts refusés



Répartition du nombre de mensualités des demandes de prêts accordées



Répartition du nombre de mensualités des demandes de prêts refusés





Harmonie graphique avec le texte

■ Exemple d'harmonisation

4. ANALYSES GRAPHIQUES

Maintenant que nous avons décrit nos données, passons aux analyses graphiques.

4.1 Analyses univariées

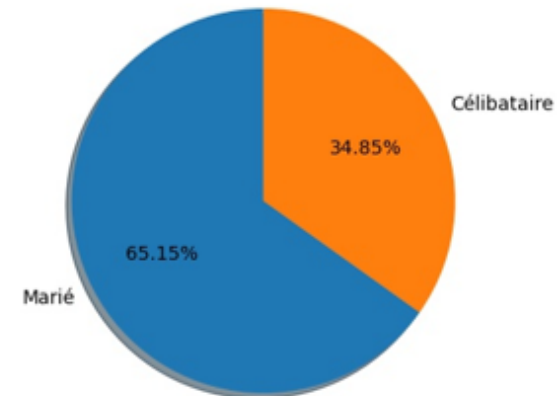
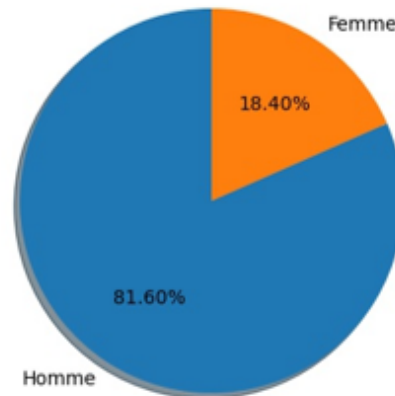
4.1.1 Les variables qualitatives

Variable 'Gender' : plus de 80% des demandes sont faites par des hommes.

Variable 'Married' : plus de 65% des demandes sont faites par des demandeurs mariés.

Genre du demandeur de crédit

Situation conjugale du demandeur de crédit





Conclusion

- Une étude bien sympathique ma foi !!!

Merci

DES QUESTIONS ?

WHAT ELSE ?