

ALGO 2

Lourdvic Alcindor

Armel Za-Bi

Xavier Le Cunff

Introduction

Les prêts constituent la base de notre économie actuelle. Une banque prête de l'argent en espérant toucher les intérêts sur cela en pourcentage de la somme prêtée à l'année. Si le créancier ne rembourse pas sa dette, il fait perdre de l'argent à la banque. Nous avons des bases de données des prêts qui ont été prêtés par les banques selon les revenus annuels des créanciers. Selon ces données nous pourrions savoir quelle somme d'argent un banquier pourra prêter à un créancier selon ses revenus annuels.

Plus les revenus du créancier sont élevés, plus il a des chances de rembourser sa dette à la banque.

Question

Quelle stratégie d'offre peut mettre en place la banque en fonction des revenus de ses clients ?

Clustering: algorithme k-means

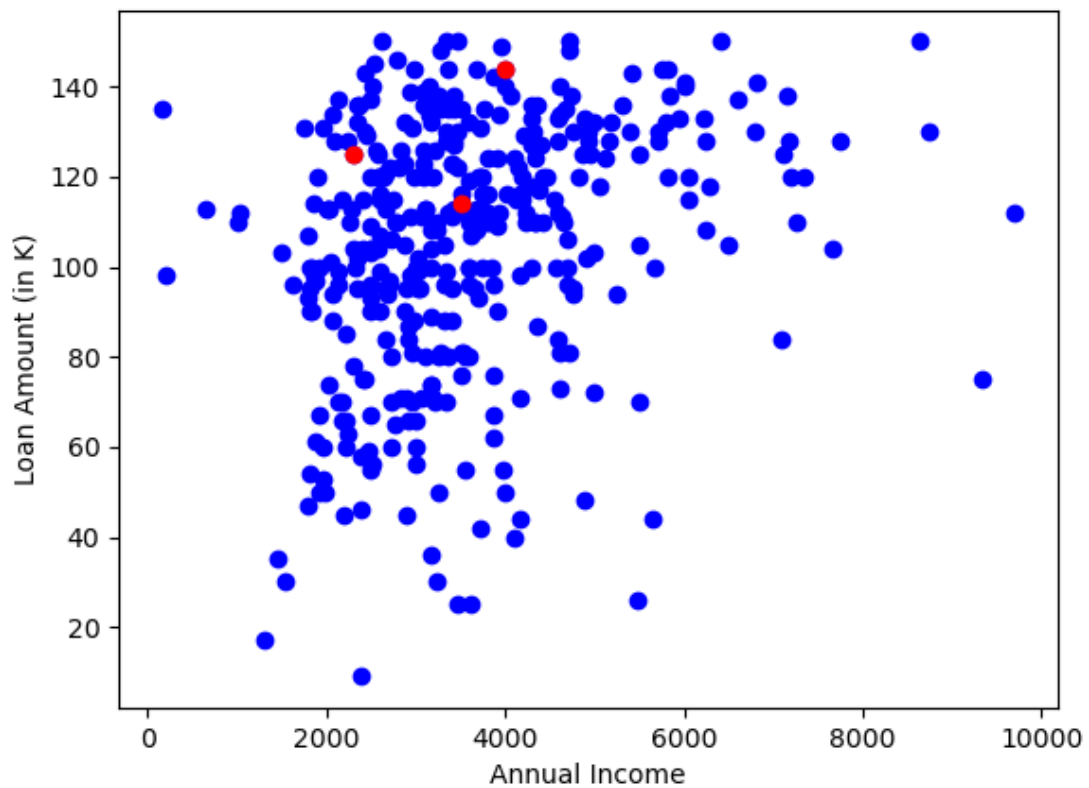
Nous décidons de faire une analyse de cluster afin de trouver si nous pouvons regrouper les clients en différents groupes afin de pouvoir proposer des offres ciblées selon le profil du client.

Nous utiliserons donc les colonnes :

« LoanAmount » qui comprends les montant de prêt en milliers

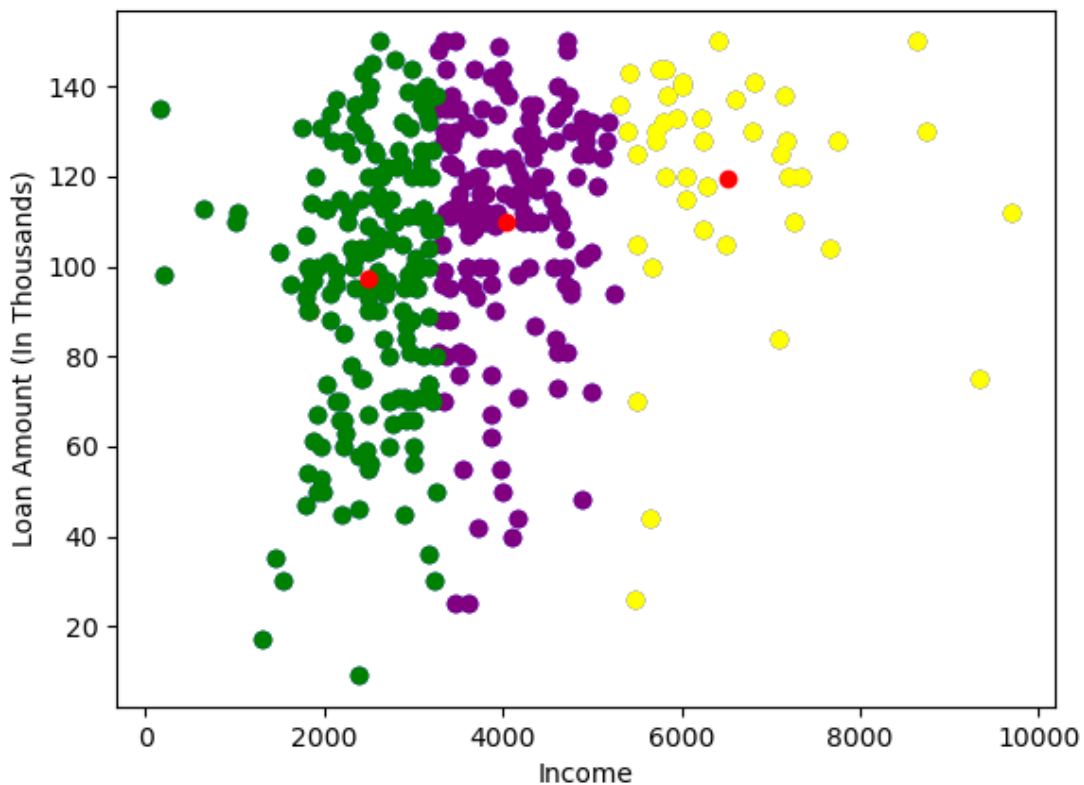
« AnnualIncome » qui comprends le revenu annuel de chaque client

Tout d'abord nous sélectionnons des points aléatoirement. Ce nombre de points est égal au nombre de clusters. Le but étant de trouver la distance moyenne la plus petite entre tous les point de ce cluster. Ces points sont appelés des centroïdes.



On calcule ensuite la distance moyenne des points de chaque cluster avec leur centroïdes correspondant.

Le centroïde change de coordonnées durant chaque itération, jusqu'à que ces derniers ne bougent plus. Si les centroïdes ne bougent plus l'algorithme se termine



Conclusion

Sur l'image générée, on a trois clusters, représentant:

- les personnes à bas revenus
- les personnes à moyens revenus
- les personnes à hauts revenus

On observe que globalement que les bas et moyens revenus empruntent environ les mêmes sommes d'argent. Les hauts revenus n'empruntent que très rarement en dessous de 100000\$, et ils sont bien moins à emprunter. On peut en conclure que les personnes à haut revenu n'empruntent qu'à partir d'une certaine somme d'argent.

On peut proposer une réduction d'intérêts à partir de 100000\$ pour attirer les personnes à hauts revenus et encourager les prêts élevés.

Data Columns

Variable Description

Loan_ID

Unique Loan ID

Gender

Male/ Female

Married

Applicant married (Y/N)

Dependents

Number of dependents

Education

Applicant Education (Graduate/ Under Graduate)

Self_Employed

Self employed (Y/N)

ApplicantIncome

Applicant income

CoapplicantIncome

Coapplicant income

LoanAmount

Loan amount in thousands

Loan_Amount_Term

Term of loan in months

Credit_History

credit history meets guidelines

Property_Area

Urban/ Semi Urban/ Rural

Loan_Status

Loan approved (Y/N)