

PROJET ALGORITHMIQUE POUR LE BIG DATA

Présenté par Amélie PSYCHE et Ulrick BLAVIN

SOMMAIRE

01

Introduction

02

Segmentation –
Premier jeu de
données

03

Recommandation
de Films

04

Conclusion

INTRODUCTION

- Plusieurs bases de données
- Trouver une structure sur des données non étiquetées
- Recommander des films

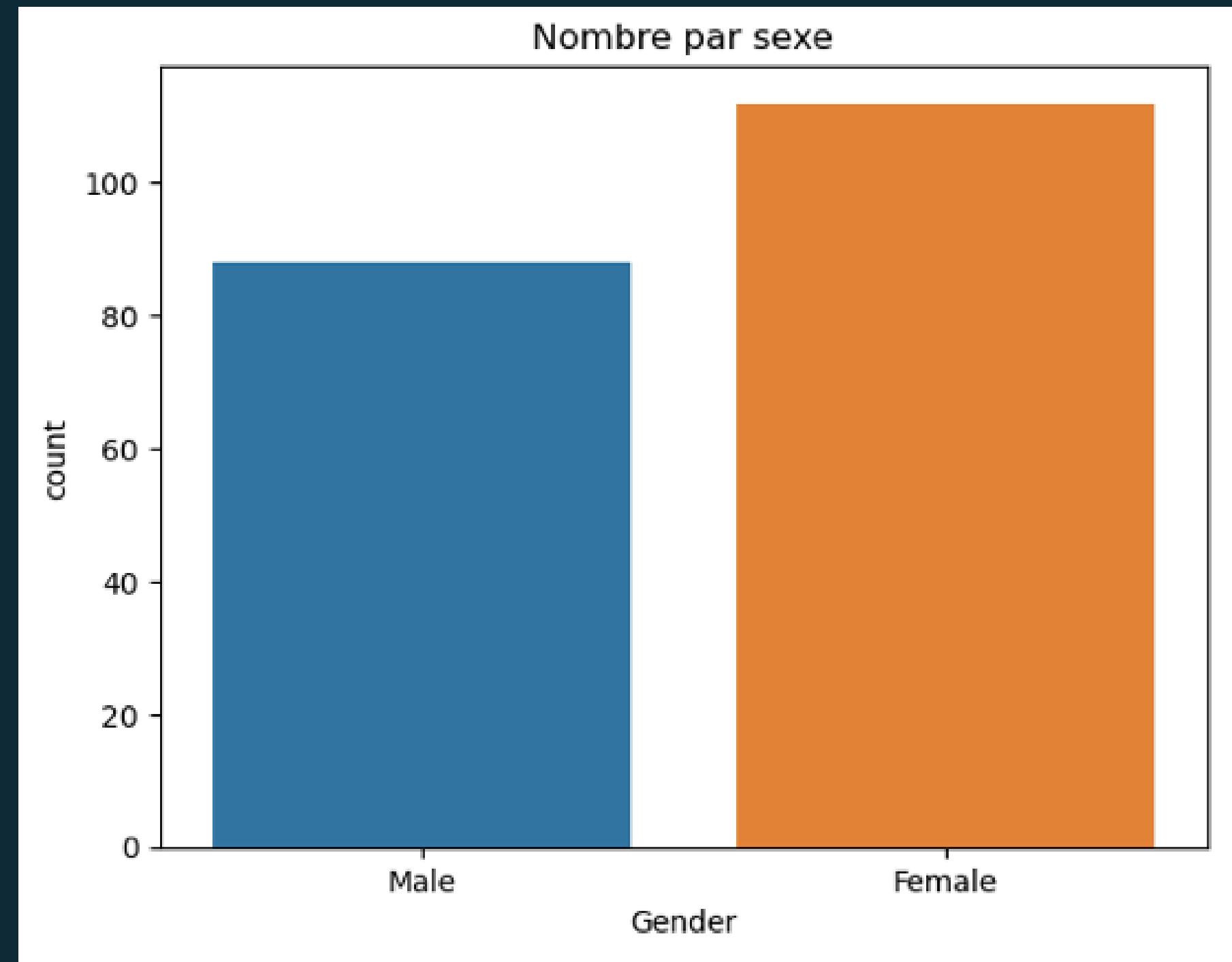
SEGMENTATION

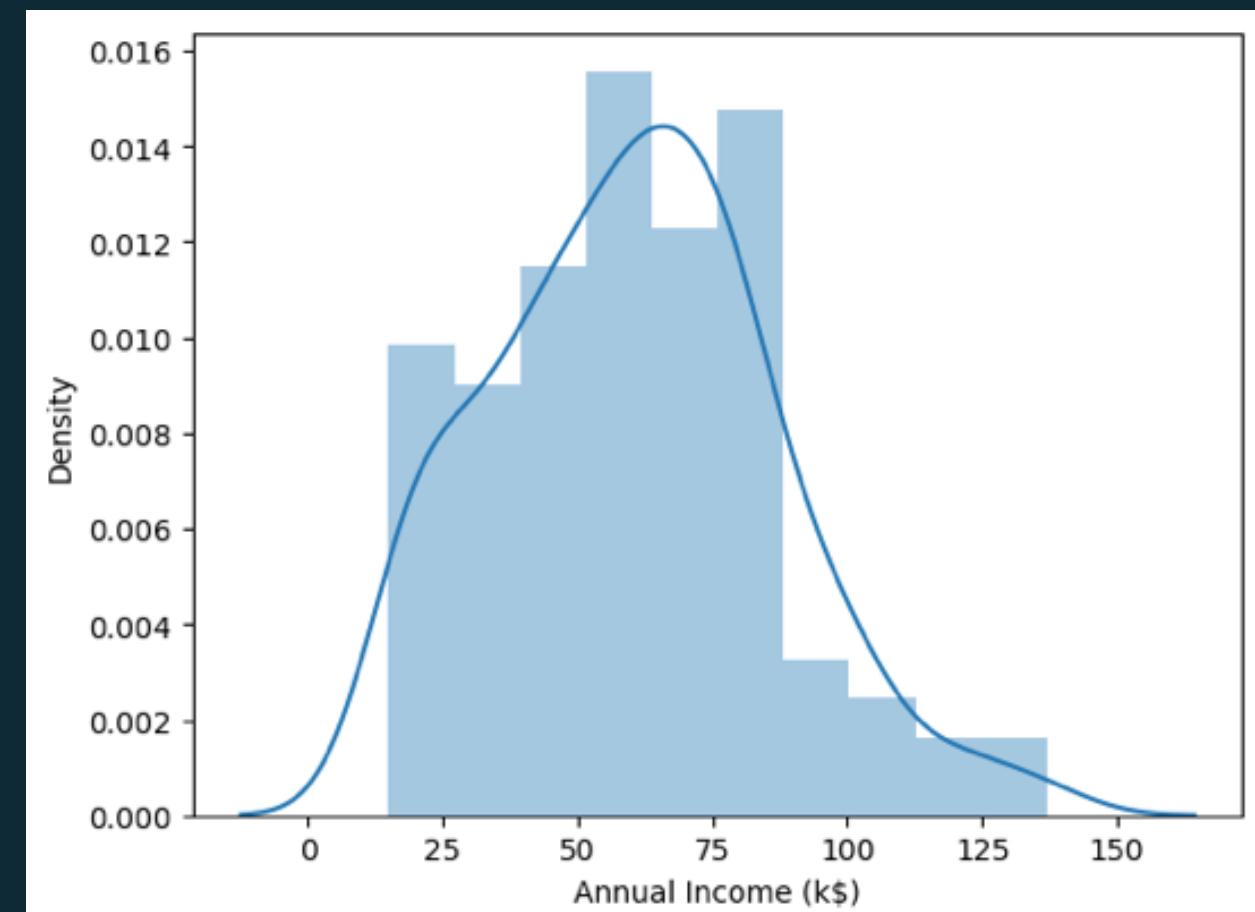
Premier jeu de données

SCHEMA

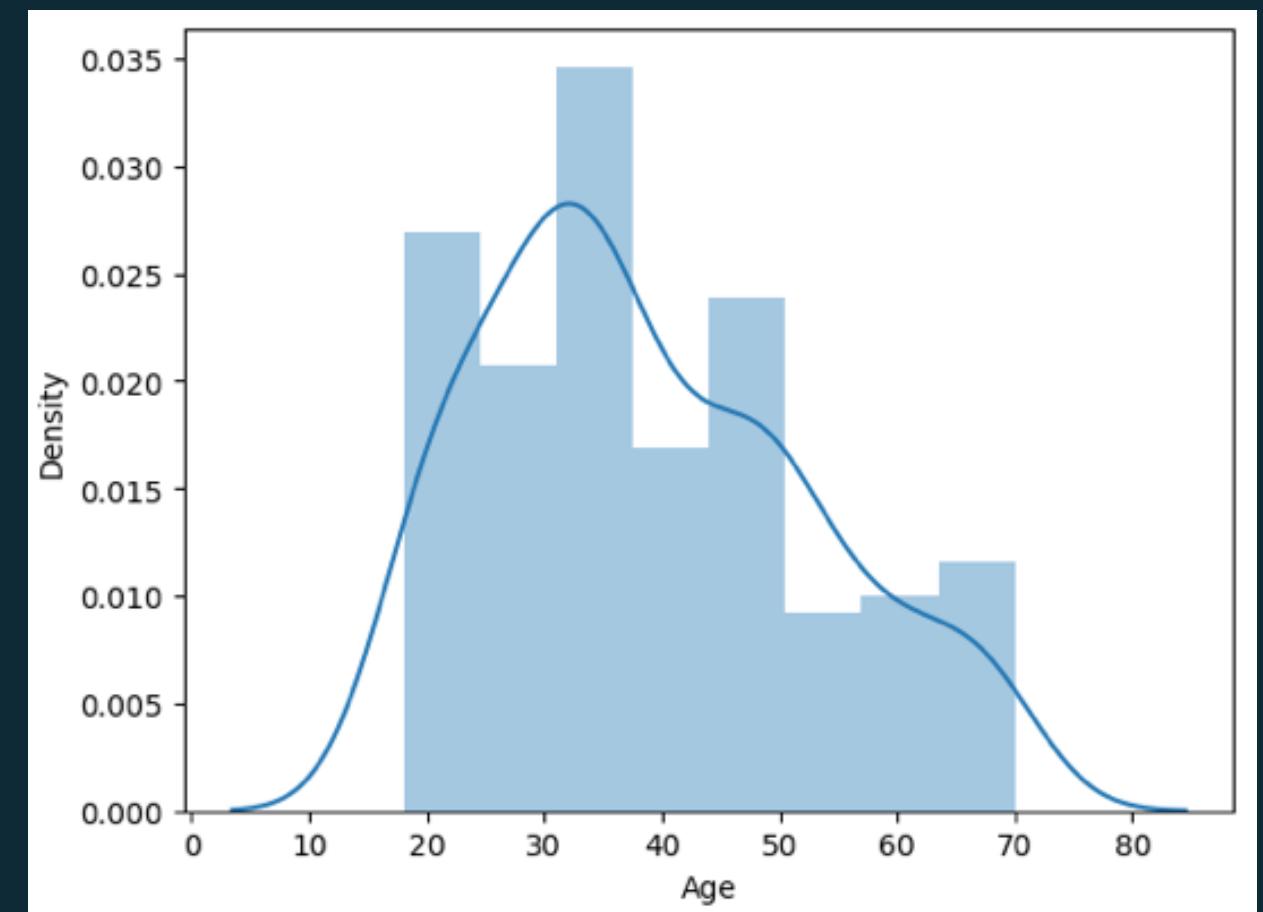
```
root
|--- CustomerID: integer (nullable = true)
|--- Gender: string (nullable = true)
|--- Age: integer (nullable = true)
|--- Annual Income (k$): integer (nullable = true)
|--- Spending Score (1-100): integer (nullable = true)
```

VISUALISATION DES DONNÉES

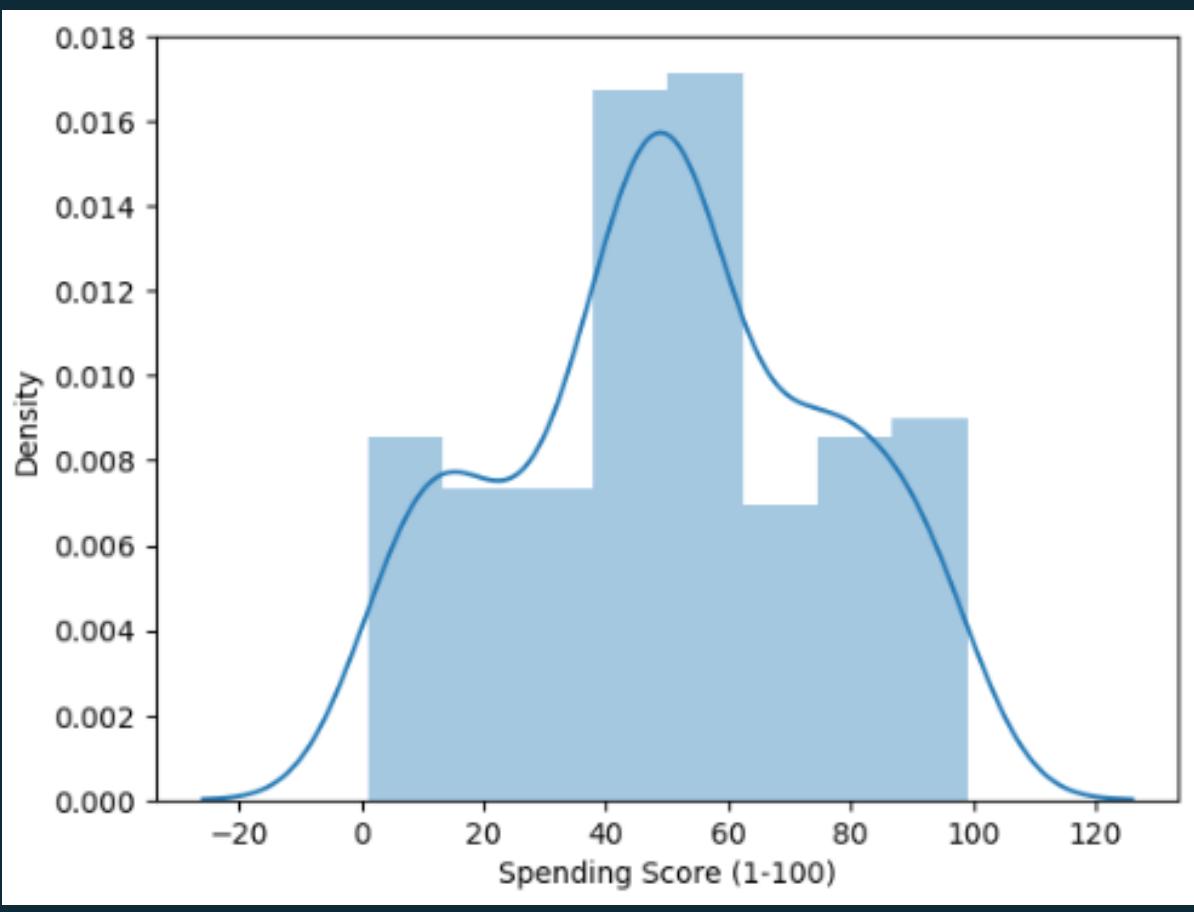




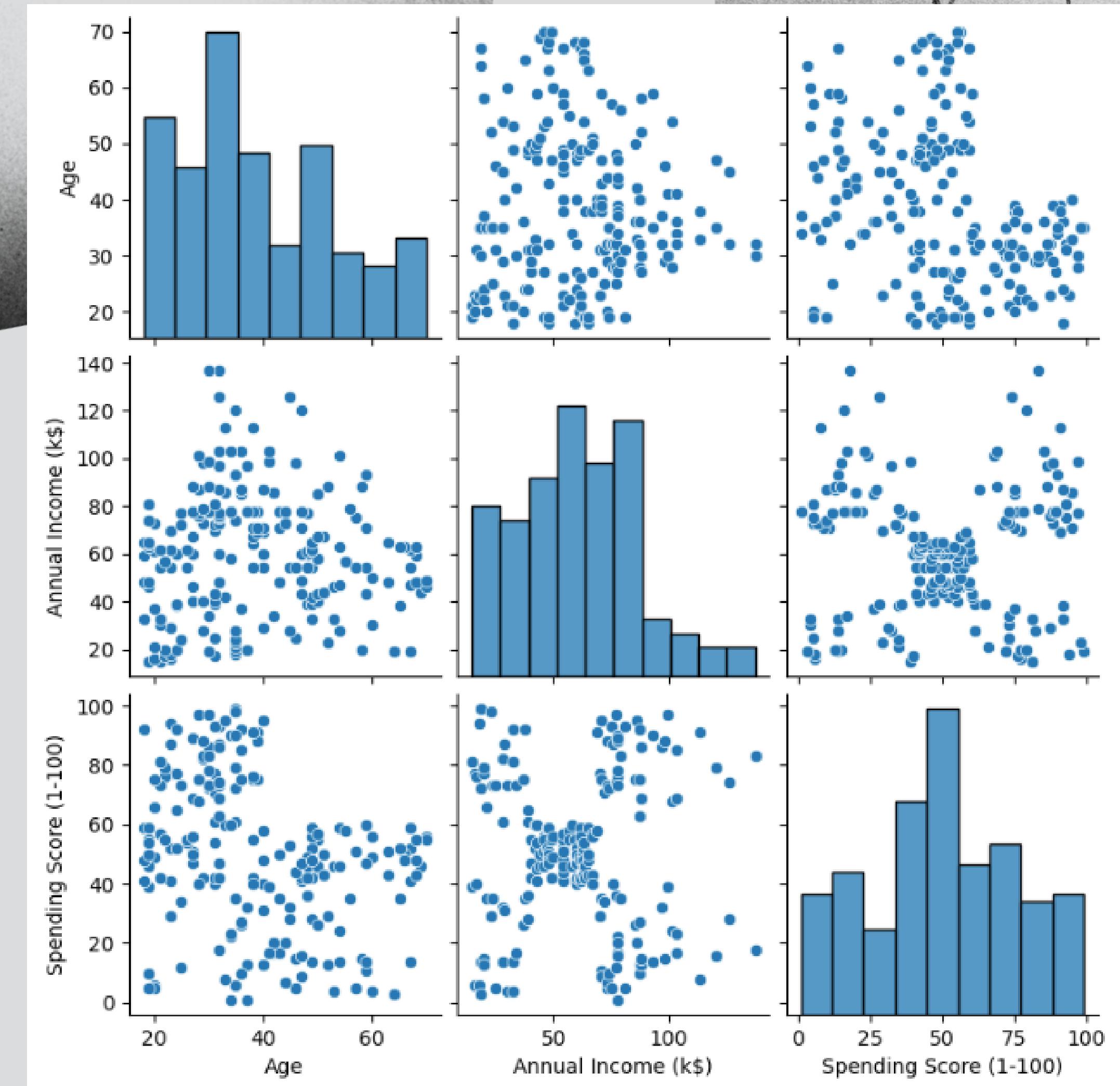
Revenu annuel des clients



Âge des clients

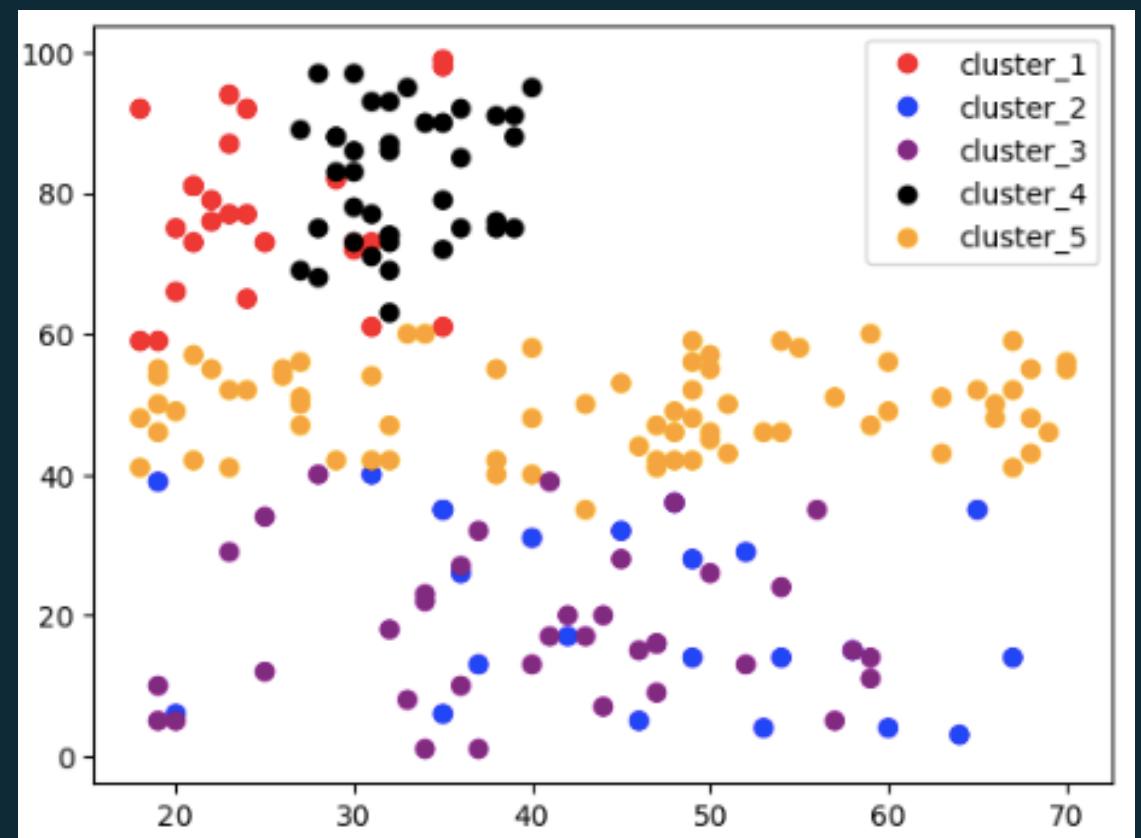


Note attribuée par le centre commercial pour les dépenses

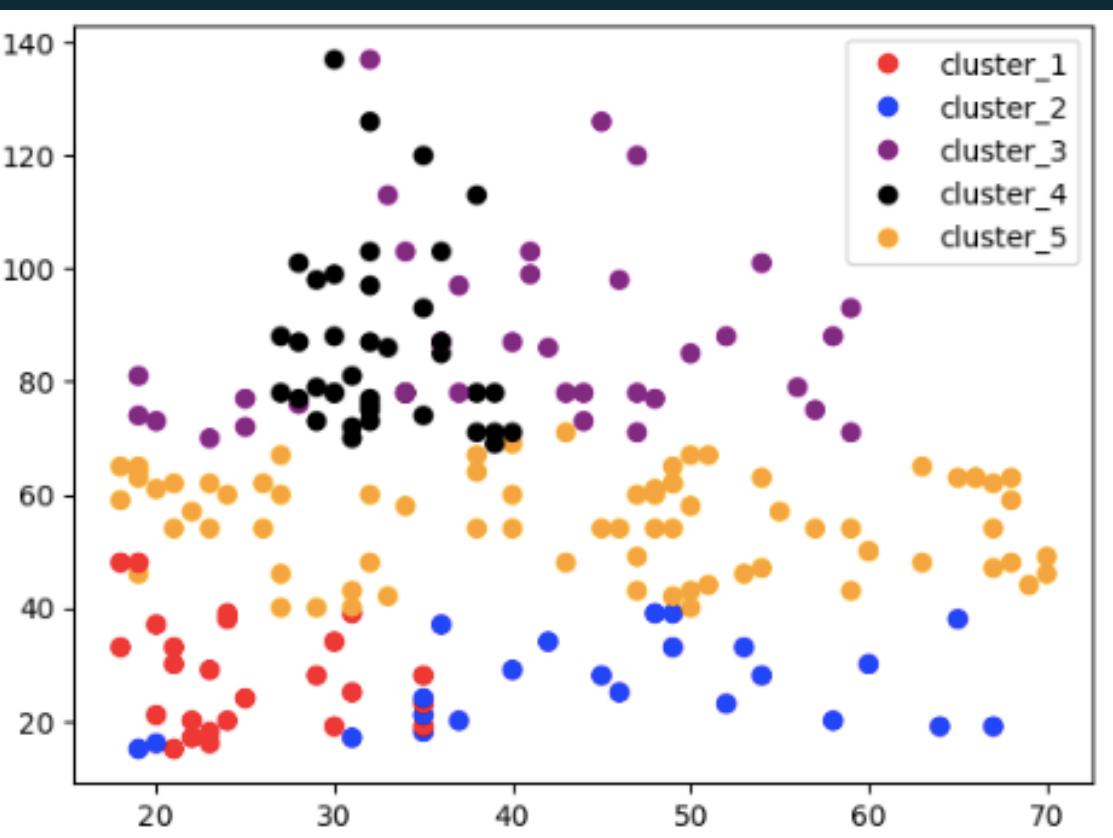


prediction	count
0	25
1	23
2	37
3	39
4	76

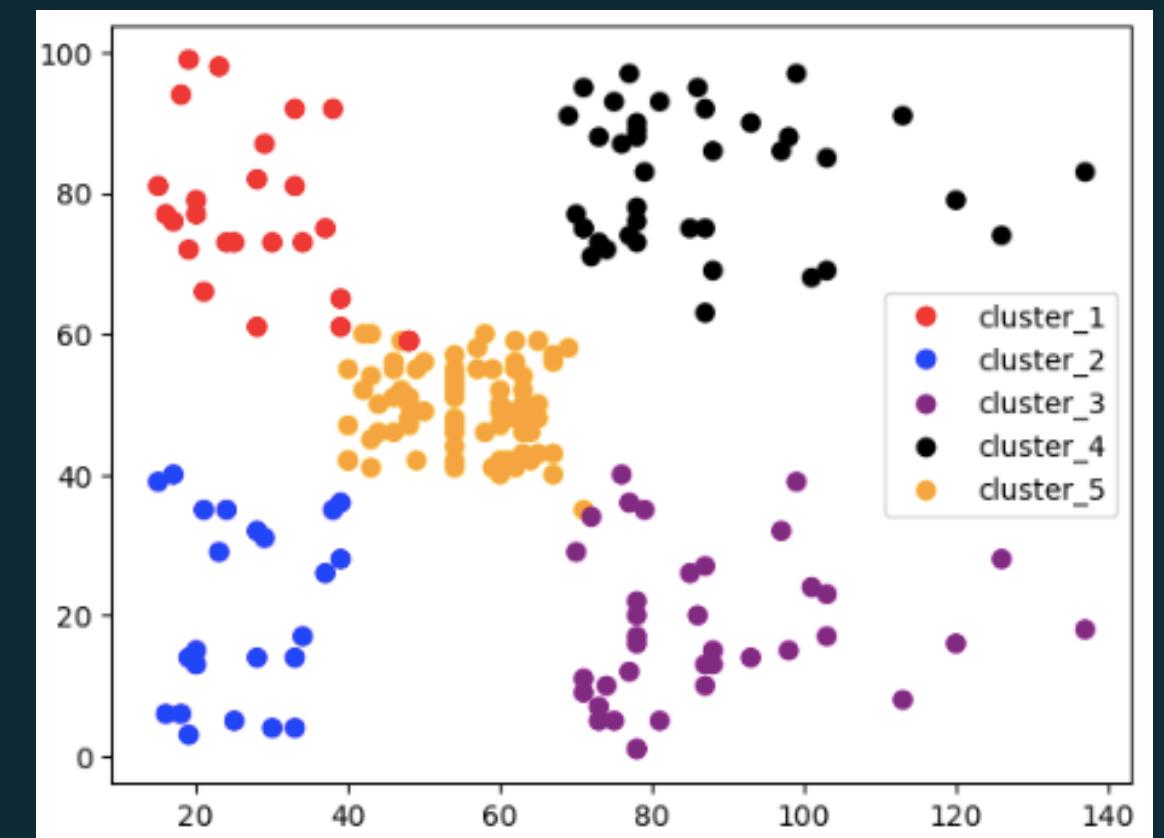
Résultats après Kmeans



L'âge en fonction du revenu annuel



L'âge en fonction de la note attribuée



Revenu annuel en fonction de la note attribué

RECOMMANDATION DE FILMS

```
root
```

```
|--- movieId: string (nullable = true)  
|--- title: string (nullable = true)  
|--- genres: string (nullable = true)
```

Données movies.csv

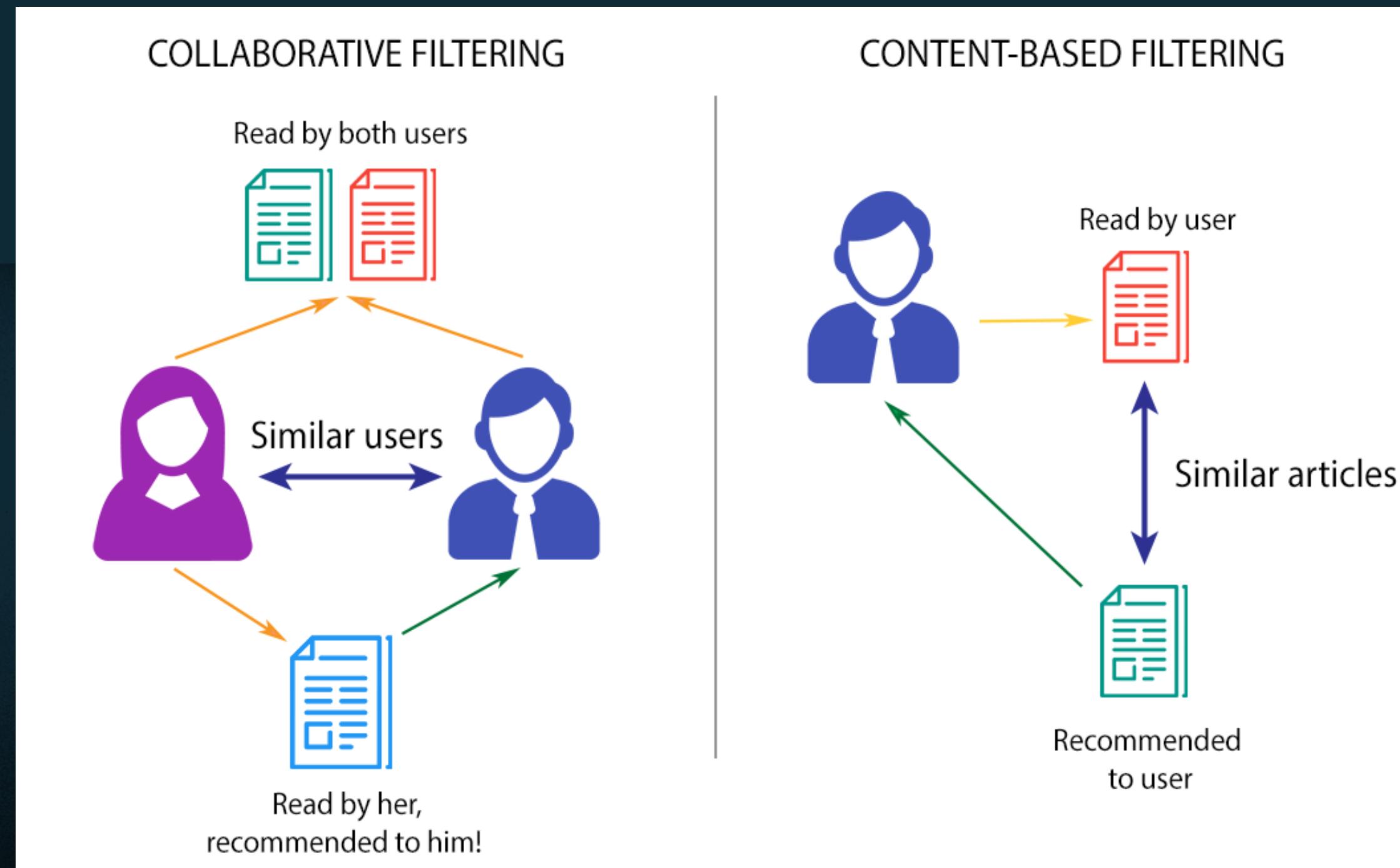
SCHEMA

```
root
```

```
|--- userId: string (nullable = true)  
|--- movieId: string (nullable = true)  
|--- rating: string (nullable = true)  
|--- timestamp: string (nullable = true)
```

Données ratings.csv

ALGORITHME ALS



RECOMMANDATION

FILMS NOTÉES

movieId	userId	rating	timestamp	title	genres
1193	100	5.0	862169731	One Flew Over the...	Drama
714	100	5.0	862169065	Dead Man (1995)	Drama Mystery Wes...
1358	100	5.0	862169132	Sling Blade (1996)	Drama
1466	100	5.0	862169186	Donnie Brasco (1997)	Crime Drama
1233	100	5.0	862168955	Boot, Das (Boat, ...	Action Drama War
781	100	5.0	862168912	Stealing Beauty (...	Drama
1354	100	5.0	862169101	Breaking the Wave...	Drama Mystery
85	100	4.0	862168835	Angels and Insect...	Drama Romance
213	100	4.0	862169731	Burnt by the Sun ...	Drama
260	100	4.0	862168435	Star Wars: Episod...	Action Adventure ...

RECOMMANDATION

FILMS RECOMMANDÉS

movieId	title	genres	userId	rating
193257	Familie Brasch (2000)	Documentary	100	5.6120677
203086	Truth and Justice... (1999)	Drama	100	5.5448117
194434	Adrenaline (1990)	(no genres listed)	100	5.469316
203882	Dead in the Water... (1999)	Horror	100	5.4482026
194334	Les Luthiers: El Show (1999)	(no genres listed)	100	5.445797
173871	I'll Take You There (1972)	Comedy Drama Romance	100	5.3621836
151989	The Thorn (1971)	Comedy	100	5.213784
183947	NOFX Backstage Pass (1999)	(no genres listed)	100	5.18946
192089	National Theatre Live: King Lear (2012)	Comedy	100	5.175419
155923	Sing (1989)	(no genres listed)	100	5.001781

CONCLUSION

- Visualisation est une étape importante
- Formuler des hypothèses
- Adaptabilité aux besoins

The background features a dark teal color with abstract white shapes. On the left, there is a large circle containing a complex network of black lines forming a web-like pattern. To its right and above, several overlapping circles of varying sizes create a sense of depth. The overall aesthetic is minimalist and modern.

MERCI DE VOTRE
ATTENTION