

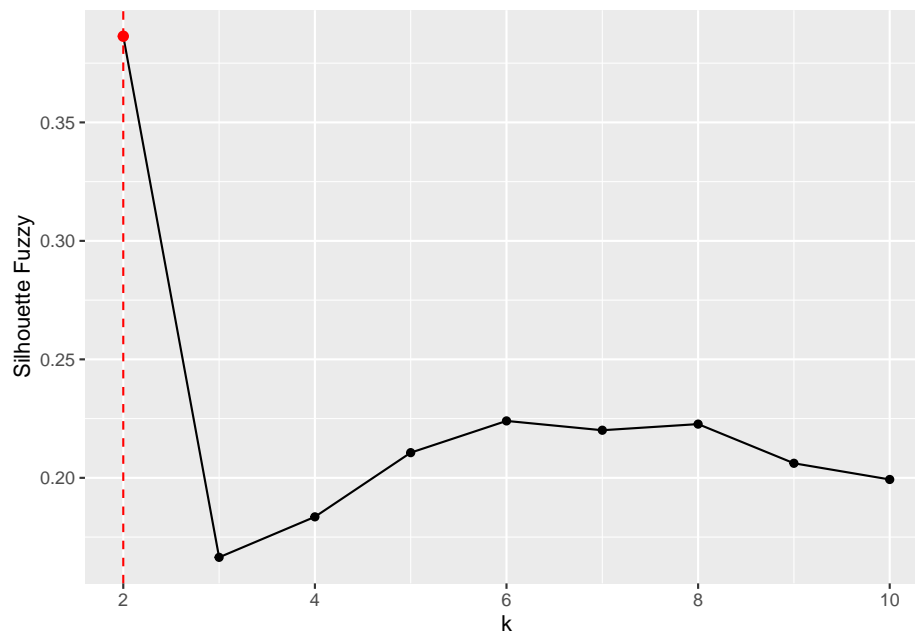
Fuzzy Clustering

La característica distintiva del Fuzzy Clustering en comparación con otros algoritmos radica en su capacidad para permitir que una observación pertenezca a más de una agrupación. En otras palabras, posibilita que los elementos o individuos tengan grados de pertenencia a varios grupos simultáneamente.

En el contexto de la ejecución del algoritmo, la función FKM() se encarga de realizar el Fuzzy Clustering para las K especificadas y, de manera automática, selecciona el valor óptimo de k .

La propia función FKM guarda los valores Silhouette fuzzy para cada k . Este es un índice específico del Fuzzy y cuanto mayor sea el índice del Silhouette mejor. Se grafican los valores SIL.F para cada k :

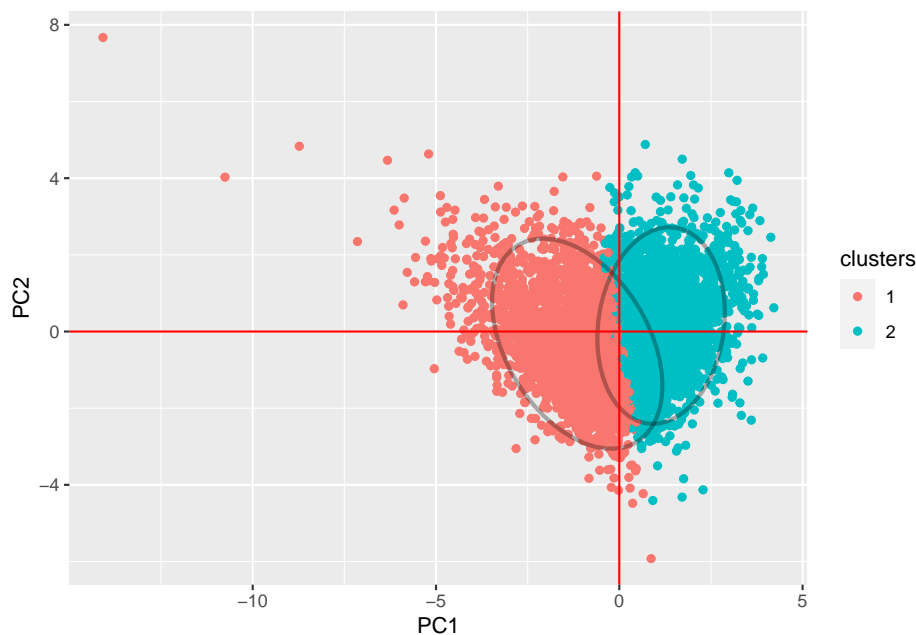
Figura 112: Índice de Silhouette Fuzzy



Como se puede ver en el gráfico, $k=2$ es la k óptima para el algoritmo fuzzy.

Acto seguido, visualizamos los clústeres en el gráfico factorial de las dos primeras dimensiones:

Figura 113: Representación de las clases en PC1-PC2



Predicción de un individuo aleatorio en el fuzzy

En esta sección, creamos un individuo hipotético y aplicamos el algoritmo Fuzzy C-means con el propósito de determinar a qué clúster se asignaría y con qué probabilidad. A continuación, se presenta la información detallada del individuo recién creado:

Individuo:

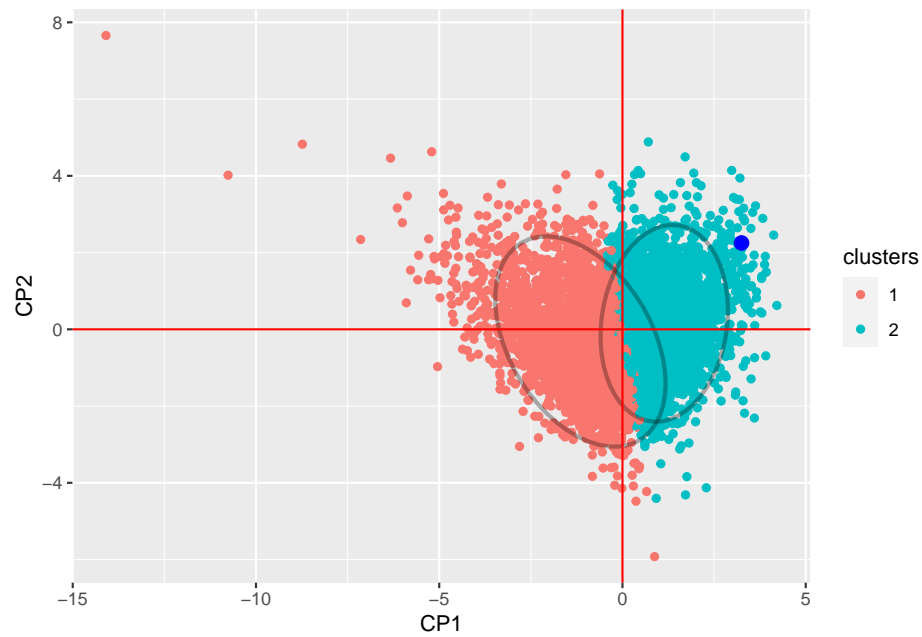
Sexo: Masculino Edad: 56 años Ingresos: \$100,000 Número de personas en la familia: 4 Años de posesión del automóvil: 7 Monto del crédito: \$12,000 Proporción de crédito: 0.12 Proporción de anualidad: 0.05 Proporción de deuda a ingresos: 0.15

Cuadro 41: Individuo generado

| | |
|----------------------|-----------|
| | 5001 |
| OWN_CAR_AGE | 7.000000 |
| CNT_FAM_MEMBERS | 4.000000 |
| log_AMT_INCOME_TOTAL | 11.512925 |
| log_AMT_CREDIT | 9.392662 |
| AGE_YEARS | 56.000000 |
| RATIO_CREDIT_INCOME | 0.120000 |
| RATIO_ANNUITY_CREDIT | 0.050000 |
| DTI_RATIO | 0.150000 |

En este apartado, aplicamos el algoritmo Fuzzy C-means a la base de datos, incorporando el individuo recién añadido. Después de la ejecución del algoritmo, visualizamos los resultados mediante la representación gráfica en el plano factorial del PCA, resaltando el individuo en cuestión con un tono azul.

Figura 114: Representación de las clases en PC1-PC2 con el individuo añadido



```
##      Clus 1      Clus 2
## 0.4477844 0.5522156
```

Vemos como, a pesar de su ubicación distante de los puntos identificados en el Clúster 1, a este individuo se le asigna a este grupo debido a una alta probabilidad de pertenencia. Esto resalta la capacidad del algoritmo para reconocer patrones y adaptarse a la variabilidad de los datos.

Profiling Fuzzy

Con el objetivo de perfilar los grupos conseguidos mediante el algoritmo Fuzzy primero veremos la significación de las variables para los grupos y después se graficarán para identificar las características definitorias de cada grupo.

A continuación se muestran los p-valores para evaluar la significación de cada variable. Primeramente de las variables categóricas y seguidamente las numéricas.

Cuadro 42: Significación de las categóricas

| Variable | P_Value |
|----------------------|-----------|
| CODE_GENDER | 0.0000000 |
| NAME_INCOME_TYPE | 0.0000000 |
| NAME_EDUCATION_TYPE | 0.0043287 |
| NAME_FAMILY_STATUS | 0.0000000 |
| OCCUPATION_TYPE | 0.0000134 |
| ORGANIZATION_TYPE | 0.0000000 |
| REGION_RATING_CLIENT | 0.0258849 |

Cuadro 43: Significación de las numéricas

| | |
|----------------------|------------------------------------|
| OWN_CAR_AGE | c(Cluster = 0.00165538345012159) |
| CNT_FAM_MEMBERS | c(Cluster = 0.866438547924501) |
| log_AMT_INCOME_TOTAL | c(Cluster = 0.826792734435659) |
| log_AMT_CREDIT | c(Cluster = 0) |
| AGE_YEARS | c(Cluster = 3.82997233028708e-48) |
| RATIO_CREDIT_INCOME | c(Cluster = 0) |
| RATIO_ANNUITY_CREDIT | c(Cluster = 0) |
| DTI_RATIO | c(Cluster = 1.88660547874547e-317) |

Elaboramos una tabla donde se indica con 1 si se considera variable significativa para el clúster y 0 en caso contrario.

Cuadro 44: Significancia de p-valores para variables numéricas:

| | x |
|----------------------|---|
| OWN_CAR_AGE | 1 |
| CNT_FAM_MEMBERS | 0 |
| log_AMT_INCOME_TOTAL | 0 |
| log_AMT_CREDIT | 1 |
| AGE_YEARS | 1 |
| RATIO_CREDIT_INCOME | 1 |
| RATIO_ANNUITY_CREDIT | 1 |
| DTI_RATIO | 1 |

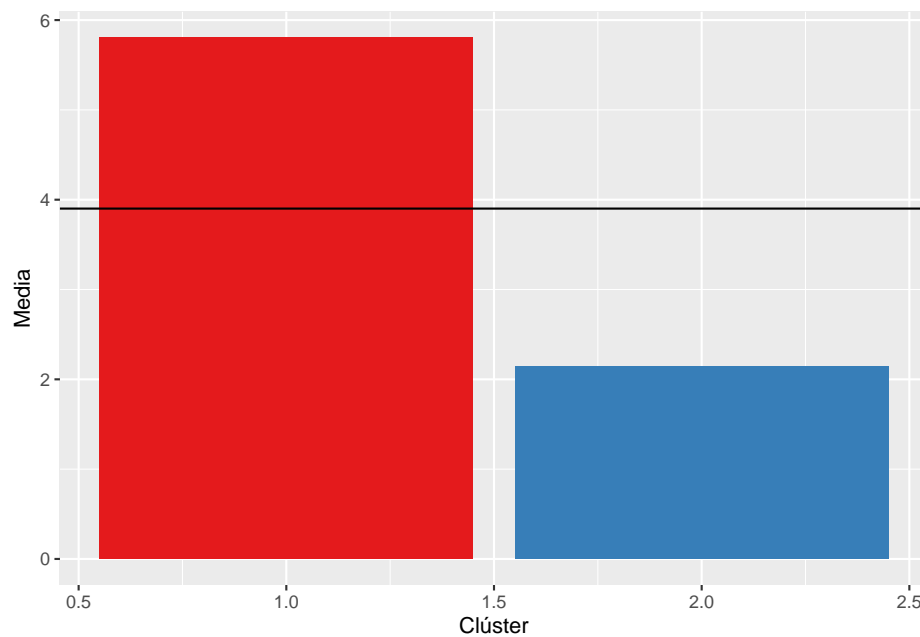
Cuadro 45: Significancia de p-valores para variables categóricas:

| | x |
|----------------------|---|
| CODE_GENDER | 1 |
| NAME_INCOME_TYPE | 1 |
| NAME_EDUCATION_TYPE | 1 |
| NAME_FAMILY_STATUS | 1 |
| OCCUPATION_TYPE | 1 |
| ORGANIZATION_TYPE | 1 |
| REGION_RATING_CLIENT | 1 |

Vemos como solo nos descarta 2 variables, pero gráficamente muy pocas aportan información que muestren diferencias grandes entre clúster.

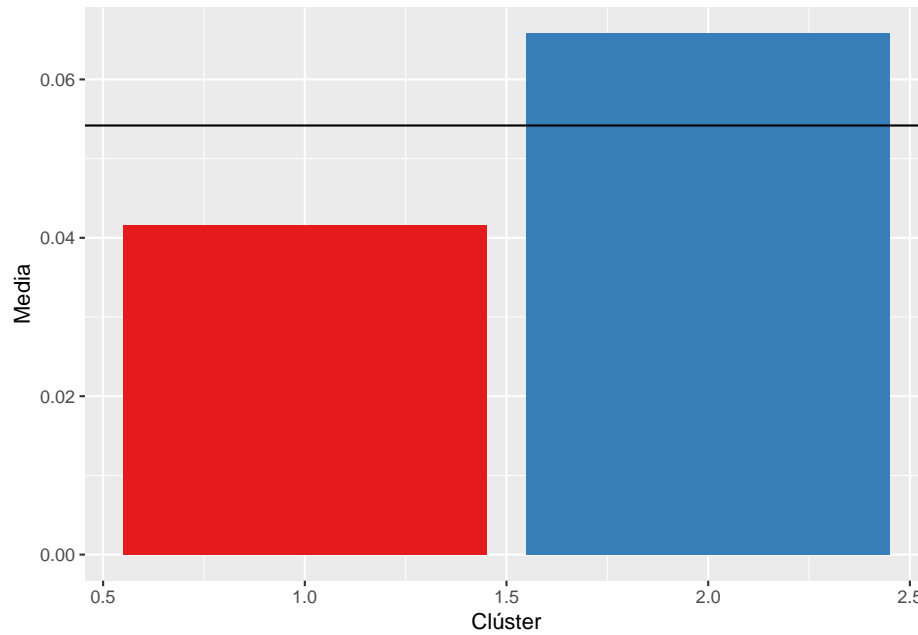
Se grafican las variables según clúster. Para las variables numéricas se mostrará la media grupal y la media global; para las variables categóricas se mostrarán las cantidades de cada nivel de la variable categórica por clúster.

Figura 115: Medias del Ratio del Importe del préstamo por clúster respecto la media global



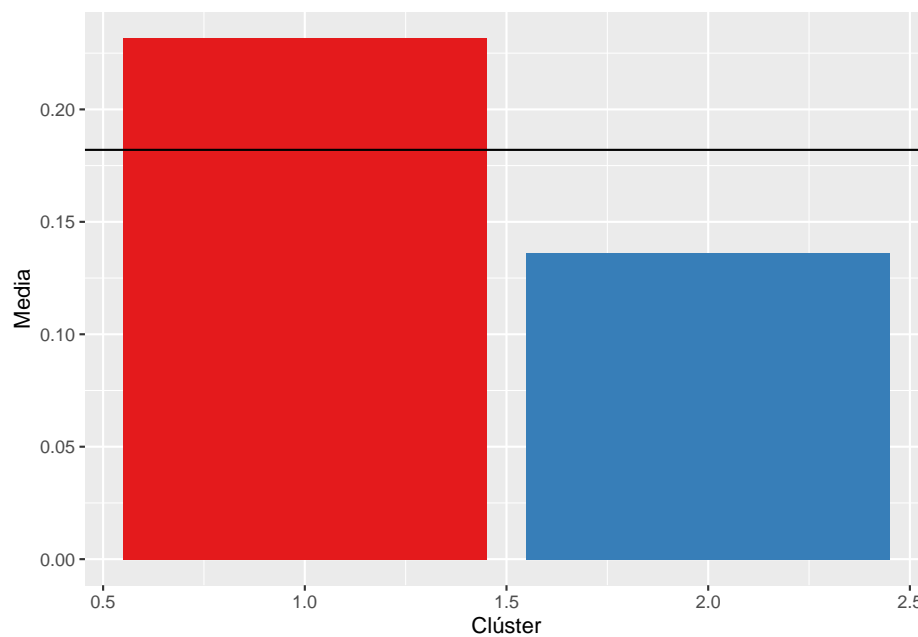
Observamos que el Clúster 1 presenta un período de reembolso más breve en comparación con el Clúster 2, en términos de la cantidad de años que los clientes tardan en devolver el préstamo.

Figura 116: Medias del Ratio de la Anualidad del préstamo por clúster respecto la media global



En el Clúster 1, se destaca que la proporción de anualidad con respecto al crédito es más alta en comparación con los otros grupos.

Figura 117: Medias de la Capacidad de cliente para pagar la annuity con sus ingresos por clúster respecto la media global



En el Clúster 1, se evidencia una mayor capacidad para reembolsar el préstamo en función de los ingresos

de los individuos.

Conclusiones:

La distinción fundamental entre nuestros clústeres reside en sus capacidades económicas. En particular, el Cluster 1 se caracteriza como el conjunto de clientes con la mejor capacidad para reembolsar el préstamo. Este grupo exhibe una proporción de anualidad más elevada y se destaca por devolver el préstamo de manera más rápida, con una media de 2 años.

Por otro lado, el Cluster 2 se identifica como el grupo de clientes con menor capacidad para reembolsar el préstamo, reflejándose en una proporción de anualidad más baja. Estos usuarios tienden a demorarse más en la devolución del préstamo, con una media de alrededor de 5-6 años.

Se adjuntan en el anexo otros gráficos relacionados con las variables analizadas, aunque no se han incluido en la presentación principal debido a su limitada relevancia informativa.