Métodos de machine learning con algoritmos de clúster no supervisados, una alternativa de segmentación de puntos kilométricos para planificar estrategias de acuerdo con sus características

Cecilia Inés Ramírez Mendoza

Universidad ICESI

Resumen

En el presente documento se hace aplicación de un PCA (Análisis de componentes principales) también conocido en castellano como ACP, ERC y ERA cuyo fundamentalmente de hacer reducción de la dimensionalidad y con algoritmos de clúster no supervisados como son los modelos K - Means y clúster jerárquico, para segmentar puntos kilométricos de acuerdo con sus características y así poder planificar estrategias de acuerdo con sus características. Se hace uso de una base de datos que contiene información de 2 años de 2016 y 2018 de 1.000 puntos kilométricos de una empresa concesionaria de vías en Colombia.

1 Importar librerías

```
import numpy as np
import pandas as pd
import matplotlib as mt
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import silhouette_score
from sklearn.metrics import pairwise_distances
from sklearn.metrics.cluster import adjusted_mutual_info_score
from sklearn.metrics.cluster import adjusted_rand_score
from sklearn.metrics.cluster import homogeneity_score
from sklearn.metrics.cluster import completeness_score
from sklearn.metrics.cluster import v_measure_score
from sklearn.metrics.cluster import fowlkes_mallows_score
from scipy import stats
from scipy import cluster
from scipy.cluster.hierarchy import *
from sklearn import preprocessing
from sklearn.preprocessing import LabelEncoder
%matplotlib inline
%matplotlib notebook
%matplotlib qtkagg
%matplotlib tk
%matplotlib inline
%matplotlib qt5
%matplotlib qt5agg
%matplotlib qt5inline
```

1.1 Carga de base de datos

```
pdata = pd.read_csv('/home/cecilia/Desktop/Datos_Kmeans.csv')
pdata.shape
```

```
(1000, 7)
```

	Nacionales	Extranjeros	Nacionales	Extranjeros	SUMCICLOS2016	SUMPPR2016	AVP_PORCENT2016	Agencia_operacional
0	0	Tramo de concesión	0	0	0	0	0	0
1	0	Tramo de concesión	0	0	0	0	0	0
2	0	Tramo de concesión	0	0	0	0	0	0
3	0	Tramo de concesión	0	0	0	0	0	0
4	0	Tramo de concesión	0	0	0	0	0	0
5	0	Tramo de concesión	0	0	0	0	0	0
6	0	Tramo de concesión	0	0	0	0	0	0
7	0	Tramo de concesión	0	0	0	0	0	0
8	0	Tramo de concesión	0	0	0	0	0	0
9	0	Tramo de concesión	0	0	0	0	0	0
10	0	Tramo de concesión	0	0	0	0	0	0
11	0	Tramo de concesión	0	0	0	0	0	0
12	0	Tramo de concesión	0	0	0	0	0	0
13	0	Tramo de concesión	0	0	0	0	0	0
14	0	Tramo de concesión	0	0	0	0	0	0
15	0	Tramo de concesión	0	0	0	0	0	0
16	0	Tramo de concesión	0	0	0	0	0	0
17	0	Tramo de concesión	0	0	0	0	0	0
18	0	Tramo de concesión	0	0	0	0	0	0
19	0	Tramo de concesión	0	0	0	0	0	0
20	0	Tramo de concesión	0	0	0	0	0	0
21	0	Tramo de concesión	0	0	0	0	0	0
22	0	Tramo de concesión	0	0	0	0	0	0
23	0	Tramo de concesión	0	0	0	0	0	0
24	0	Tramo de concesión	0	0	0	0	0	0
25	0	Tramo de concesión	0	0	0	0	0	0
26	0	Tramo de concesión	0	0	0	0	0	0
27	0	Tramo de concesión	0	0	0	0	0	0
28	0	Tramo de concesión	0	0	0	0	0	0
29	0	Tramo de concesión	0	0	0	0	0	0
30	0	Tramo de concesión	0	0	0	0	0	0
31	0	Tramo de concesión	0	0	0	0	0	0
32	0	Tramo de concesión	0	0	0	0	0	0
33	0	Tramo de concesión	0	0	0	0	0	0
34	0	Tramo de concesión	0	0	0	0	0	0
35	0	Tramo de concesión	0	0	0	0	0	0
36	0	Tramo de concesión	0	0	0	0	0	0
37	0	Tramo de concesión	0	0	0	0	0	0
38	0	Tramo de concesión	0	0	0	0	0	0
39	0	Tramo de concesión	0	0	0	0	0	0
40	0	Tramo de concesión	0	0	0	0	0	0
41	0	Tramo de concesión	0	0	0	0	0	0
42	0	Tramo de concesión	0	0	0	0	0	0
43	0	Tramo de concesión	0	0	0	0	0	0
44	0	Tramo de concesión	0	0	0	0	0	0
45	0	Tramo de concesión	0	0	0	0	0	0
46	0	Tramo de concesión	0	0	0	0	0	0
47	0	Tramo de concesión	0	0	0	0	0	0
48	0	Tramo de concesión	0	0	0	0	0	0
49	0	Tramo de concesión	0	0	0	0	0	0
50	0	Tramo de concesión	0	0	0	0	0	0
51	0	Tramo de concesión	0	0	0	0	0	0
52	0	Tramo de concesión	0	0	0	0	0	0
53	0	Tramo de concesión	0	0	0	0	0	0
54	0	Tramo de concesión	0	0	0	0	0	0
55	0	Tramo de concesión	0	0	0	0	0	0
56	0	Tramo de concesión	0	0	0	0	0	0
57	0	Tramo de concesión	0	0	0	0	0	0
58	0	Tramo de concesión	0	0	0	0	0	0
59	0	Tramo de concesión	0	0	0	0	0	0
60	0	Tramo de concesión	0	0	0	0	0	0
61	0	Tramo de concesión	0	0	0	0	0	0
62	0	Tramo de concesión	0	0	0	0	0	0
63	0	Tramo de concesión	0	0	0	0	0	0
64	0	Tramo de concesión	0	0	0	0	0	0
65	0	Tramo de concesión	0	0	0	0	0	0
66	0	Tramo de concesión	0	0	0	0	0	0
67	0	Tramo de concesión	0	0	0	0	0	0
68	0	Tramo de concesión	0	0	0	0	0	0
69	0	Tramo de concesión	0	0	0	0	0	0
70	0	Tramo de concesión	0	0	0	0	0	0
71	0	Tramo de concesión	0	0	0	0	0	0
72	0	Tramo de concesión	0	0	0	0	0	0
73	0	Tramo de concesión	0	0	0	0	0	0
74	0	Tramo de concesión	0	0	0	0	0	0
75	0	Tramo de concesión	0	0	0	0	0	0
76	0	Tramo de concesión	0	0	0	0	0	0
77	0	Tramo de concesión	0	0	0	0	0	0
78	0	Tramo de concesión	0	0	0	0	0	0
79	0	Tramo de concesión	0	0	0	0	0	0
80	0	Tramo de concesión	0	0	0	0	0	0
81	0	Tramo de concesión	0	0	0	0	0	0
82	0	Tramo de concesión	0	0	0	0	0	0
83	0	Tramo de concesión	0	0	0	0	0	0
84	0	Tramo de concesión	0	0	0	0	0	0
85	0	Tramo de concesión	0	0	0	0	0	0
86	0	Tramo de concesión	0	0	0	0	0	0
87	0	Tramo de concesión	0	0	0	0	0	0
88	0	Tramo de concesión	0	0	0	0	0	0
89	0	Tramo de concesión	0	0	0	0	0	0
90	0	Tramo de concesión	0	0	0	0	0	0
91	0	Tramo de concesión	0	0	0	0	0	0
92	0	Tramo de concesión	0	0	0	0	0	0
93	0	Tramo de concesión	0	0	0	0	0	0
94	0	Tramo de concesión	0	0	0	0	0	0
95	0	Tramo de concesión	0	0	0	0	0	0
96	0	Tramo de concesión	0	0	0	0	0	0
97	0	Tramo de concesión	0	0	0	0	0	0
98	0	Tramo de concesión	0	0	0	0	0	0
99	0	Tramo de concesión	0	0	0	0	0	0
100	0	Tramo de concesión	0	0	0	0	0	0
101	0	Tramo de concesión	0	0	0	0	0	0
102	0	Tramo de concesión	0	0	0	0	0	0
103	0	Tramo de concesión	0	0	0	0	0	0
104	0	Tramo de concesión	0	0	0	0	0	0
105	0	Tramo de concesión	0	0	0	0	0	0
106	0	Tramo de concesión	0	0	0	0	0	0
107	0	Tramo de concesión	0	0	0	0	0	0
108	0	Tramo de concesión	0	0	0	0	0	0
109	0	Tramo de concesión	0	0	0	0	0	0
110	0	Tramo de concesión	0	0	0	0	0	0
111	0	Tramo de concesión	0	0	0	0	0	0
112	0	Tramo de concesión	0	0	0	0	0	0
113	0	Tramo de concesión	0	0	0	0	0	0
114	0	Tramo de concesión	0	0	0	0	0	0
115	0	Tramo de concesión	0	0	0	0	0	0
116	0	Tramo de concesión	0	0	0	0	0	0
117	0	Tramo de concesión	0	0	0	0	0	0
118	0	Tramo de concesión	0	0	0	0	0	0
119	0	Tramo de concesión	0	0	0	0	0	0
120	0	Tramo de concesión	0	0	0	0	0	0
121	0	Tramo de concesión	0	0	0	0	0	0
122	0	Tramo de concesión	0	0	0	0	0	0
123	0	Tramo de concesión	0	0	0	0	0	0
124	0	Tramo de concesión	0	0	0	0	0	0
125	0	Tramo de concesión	0	0	0	0	0	0
126	0	Tramo de concesión	0	0	0	0	0	0
127	0	Tramo de concesión	0	0	0	0	0	0
128	0	Tramo de concesión	0	0	0	0	0	0
129	0	Tramo de concesión	0	0	0	0	0	0
130	0	Tramo de concesión	0	0	0	0	0	0
131	0	Tramo de concesión	0	0	0	0	0	0
132	0	Tramo de concesión	0	0	0	0	0	0
133	0	Tramo de concesión	0	0	0	0	0	0
134	0	Tramo de concesión	0	0	0	0	0	0
135	0	Tramo de concesión	0	0	0	0	0	0
136	0	Tramo de concesión	0	0	0	0	0	0
137	0	Tramo de concesión	0	0	0	0	0	0
138	0	Tramo de concesión	0	0	0	0	0	0
139	0	Tramo de concesión	0	0	0	0	0	0
140	0	Tramo de concesión	0	0	0	0	0	0
141	0	Tramo de concesión	0	0	0	0	0	0
142	0	Tramo de concesión	0	0	0	0	0	0
143	0	Tramo de concesión	0	0	0	0	0	0
144	0	Tramo de concesión	0	0	0	0	0	0
145	0	Tramo de concesión	0	0	0	0	0	0
146	0	Tramo de concesión	0	0	0	0	0	0
147	0	Tramo de concesión	0	0	0	0	0	0
148	0	Tramo de concesión	0	0	0	0	0	0
149	0	Tramo de concesión	0	0	0	0	0	0
150	0	Tramo de concesión	0	0	0	0	0	0
151	0	Tramo de concesión	0	0	0	0	0	0
152	0	Tramo de concesión	0	0	0	0	0	0
153	0	Tramo de concesión	0	0	0	0	0	0
154	0	Tramo de concesión	0	0	0	0	0	0
155	0	Tramo de concesión	0	0	0	0	0	0
156	0	Tramo de concesión	0	0	0	0	0	0
157	0	Tramo de concesión	0	0	0	0	0	0
158	0	Tramo de concesión	0	0	0	0	0	0
159	0	Tramo de concesión	0	0	0	0	0	0
160	0	Tramo de concesión	0	0	0	0	0	0
161	0	Tramo de concesión	0	0	0	0	0	0
162	0	Tramo de concesión	0	0	0	0	0	0
163	0	Tramo de concesión	0	0	0	0	0	0
164	0	Tramo de concesión	0	0	0	0	0	0
165	0	Tramo de concesión	0	0	0	0	0	0
166	0	Tramo de concesión	0	0	0	0	0	0
167	0	Tramo de concesión	0	0	0	0	0	0
168	0	Tramo de concesión	0	0	0	0	0	0
169	0	Tramo de concesión	0	0	0	0	0	0
170	0	Tramo de concesión	0	0	0	0	0	0
171	0	Tramo de concesión	0	0	0	0	0	0
172	0	Tramo de concesión	0	0	0	0	0	0
173	0	Tramo de concesión	0	0	0	0	0	0
174	0	Tramo de concesión	0	0	0	0	0	0
175	0	Tramo de concesión	0	0	0	0	0	0
176	0	Tramo de concesión	0	0	0	0	0	0
177	0	Tramo de concesión	0	0	0	0	0	0
178	0	Tramo de concesión	0	0	0	0	0	0
179	0	Tramo de concesión	0	0	0	0	0	0
180	0	Tramo de concesión	0	0	0	0	0	0
181	0	Tramo de concesión	0	0	0	0	0	0
182	0	Tramo de concesión	0	0	0	0	0	0
183	0	Tramo de concesión	0	0	0	0	0	0
184	0	Tramo de concesión	0	0	0	0	0	0
185	0	Tramo de concesión	0	0	0	0	0	0
186	0	Tramo de concesión	0	0	0	0	0	0
187	0	Tramo de concesión	0	0	0	0	0	0
188	0	Tramo de concesión	0	0	0	0	0	0
189	0	Tramo de concesión	0	0	0	0	0	0
190	0	Tramo de concesión	0	0	0	0	0	0
191	0	Tramo de concesión	0	0	0	0	0	0
192	0	Tramo de concesión	0	0	0	0	0	0
193	0	Tramo de concesión	0	0	0	0	0	0
194	0	Tramo de concesión	0	0	0	0	0	0
195	0	Tramo de concesión	0	0	0	0	0	0
196	0	Tramo de concesión	0	0	0	0	0	0
197	0	Tramo de concesión	0	0	0	0	0	0
198	0	Tramo de concesión	0	0	0	0	0	0
199	0	Tramo de concesión	0	0	0	0	0	0
200	0	Tramo de concesión	0	0	0	0	0	0
201	0	Tramo de concesión	0	0	0	0	0	0
202	0	Tramo de concesión	0	0	0	0	0	0
203	0	Tramo de concesión	0	0	0	0	0	0
204	0	Tramo de concesión	0	0	0	0	0	0
205	0	Tramo de concesión	0	0	0	0	0	0
206	0	Tramo de concesión	0	0	0	0	0	0
207	0	Tramo de concesión	0	0	0	0	0	0
208	0	Tramo de concesión	0	0	0	0	0	0
209	0	Tramo de concesión	0	0	0	0	0	0
210	0	Tramo de concesión	0	0	0	0	0	0
211	0	Tramo de concesión	0	0	0	0	0	0
212	0	Tramo de concesión	0	0	0	0	0	0
213	0	Tramo de concesión	0	0	0	0	0	0
214	0	Tramo de concesión	0	0	0	0	0	0
215	0	Tramo de concesión	0	0	0	0	0	0
216	0	Tramo de concesión	0	0	0	0	0	0
217	0	Tramo de concesión	0	0	0	0	0	0
218	0	Tramo de concesión	0	0	0	0	0	0
219	0	Tramo de concesión	0	0	0	0	0	0
220	0	Tramo de concesión	0	0	0	0	0	0
221	0	Tramo de concesión	0	0	0	0	0	0
222	0	Tramo de concesión	0	0	0	0	0	0
223	0	Tramo de concesión	0	0	0	0	0	0
224	0	Tramo de concesión	0	0	0	0	0	0
225	0	Tramo de concesión	0	0	0	0	0	0
226	0	Tramo de concesión	0	0	0	0	0	0
227	0	Tramo de concesión	0	0	0	0	0	0
228	0	Tramo de concesión	0	0	0	0	0	0
229	0	Tramo de concesión	0	0	0	0	0	0
230	0	Tramo de concesión	0	0	0	0	0	0
231	0	Tramo de concesión	0	0	0	0	0	0
232	0	Tramo de concesión	0	0	0	0	0	0
233	0	Tramo de concesión	0	0	0	0	0	0
234	0	Tramo de concesión	0	0	0	0	0	0
235	0	Tramo de concesión	0	0	0	0	0	0
236	0	Tramo de concesión	0	0	0	0	0	0
237	0	Tramo de concesión	0	0	0	0	0	0
238	0	Tramo de concesión	0	0	0	0	0	0
239	0	Tramo de concesión	0	0	0	0	0	0
240	0	Tramo de concesión	0	0	0	0	0	0
241	0	Tramo de concesión	0	0	0	0	0	0
242	0	Tramo de concesión	0	0	0	0	0	0
243	0	Tramo de concesión	0	0	0	0	0	0
244	0	Tramo de concesión	0	0	0	0	0	0
245	0	Tramo de concesión	0	0	0	0	0	0
246	0	Tramo de concesión	0	0	0	0	0	0
247	0	Tramo de concesión	0	0	0	0	0	0
248	0	Tramo de concesión	0	0	0	0	0	0
249	0	Tramo de concesión	0	0	0	0	0	0
250	0	Tramo de concesión	0	0	0	0	0	0
251	0	Tramo de concesión	0	0	0	0	0	0
252	0	Tramo de concesión	0	0	0	0	0	0
253	0	Tramo de concesión	0	0	0	0	0	0
254	0	Tramo de concesión	0	0	0	0	0	0
255	0	Tramo de concesión	0	0	0	0	0	0
256	0	Tramo de concesión	0	0	0	0	0	0
257	0	Tramo de concesión	0	0	0	0	0	0
258	0	Tramo de concesión	0	0	0	0	0	0
259	0	Tramo de concesión	0	0	0	0	0	0
260	0	Tramo de concesión	0	0	0	0	0	0
261	0	Tramo de concesión	0	0	0	0	0	0
262	0	Tramo de concesión	0	0	0	0	0	0
263	0	Tramo de concesión	0	0	0	0	0	0
264	0	Tramo de concesión	0	0	0	0	0	0
265	0	Tramo de concesión	0	0	0	0	0	0
266	0	Tramo de concesión	0	0	0	0	0	0
267	0	Tramo de concesión	0	0	0	0	0	0
268	0	Tramo de concesión	0	0	0	0	0	0
269	0	Tramo de concesión	0	0	0	0	0	0
270	0	Tramo de concesión	0	0	0	0	0	0
271	0	Tramo de concesión	0	0	0	0	0	0
272	0	Tramo de concesión	0	0	0	0	0	0
273	0	Tramo de concesión	0	0	0	0	0	0
274	0	Tramo de concesión	0	0	0	0	0	0
275	0	Tramo de concesión	0	0	0	0	0	0
276	0	Tramo de concesión	0	0	0	0	0	0
277	0	Tramo de concesión	0	0	0	0	0	0
278	0	Tramo de concesión	0	0	0	0	0	0
279	0	Tramo de concesión	0	0	0	0	0	0
280	0	Tramo de concesión	0	0	0	0	0	0
281	0	Tramo de concesión	0	0	0	0	0	0
282	0	Tramo de concesión	0	0	0	0	0	0
283	0	Tramo de concesión	0	0	0	0	0	0
284	0	Tramo de concesión	0	0	0	0	0	0
285	0	Tramo de concesión	0	0	0	0	0	0
286	0	Tramo de concesión	0	0	0	0	0	0
287	0	Tramo de concesión	0	0	0	0	0	0
288	0	Tramo de concesión	0	0	0	0	0	0
289	0	Tramo de concesión	0	0	0	0	0	0
290	0	Tramo de concesión	0	0	0	0	0	0
291	0	Tramo de concesión	0	0	0	0	0	0
292	0	Tramo de concesión	0	0	0	0	0	0
293	0	Tramo de concesión	0	0	0	0	0	0
294	0	Tramo de concesión	0	0	0	0	0	0
295	0	Tramo de concesión	0	0	0	0	0	0
296	0	Tramo de concesión	0	0	0	0	0	0
297	0	Tramo de concesión	0	0	0	0	0	0
298	0	Tramo de concesión	0	0	0	0	0	0
299	0	Tramo de concesión	0	0	0	0	0	0

1.2 Carga de base de datos

```
pdata = pd.read_csv('/home/cecilia/Desktop/Datos_Kmeans.csv')
pdata.shape
```

```
(1000, 7)
```

2 Análisis Exploratorio de Datos

2.1 Nombre de datos y variables

La base de datos corresponde a 999 entradas y filas y 30 variables y columnas

```
pdata.shape
```

```
(1000, 7)
```

2.2 Nombre de empresas

Las 4004 entradas corresponden 1000 empresas en los años 2016, 2017, 2018, 2019.

```
'Concesión ACF' unique()
```

```
[nan]
```

2.3 Tipos de variables

```
pdata.dtypes
```

```
Unnamed: 0                 int64
Nacionales                 int64
Extranjeros                int64
Nacionales.1               int64
Extranjeros.1              int64
SUMCICLOS2016              int64
SUMPPR2016                 int64
AVP_PORCENT2016            int64
Agencia_operacional        int64
Tramo_del_corredor         int64
Tramo_Activo               int64
Mongo_Activo               int64
Tramo_Activo.1             int64
Mongo_Activo.1             int64
Tramo_Activo.2             int64
Mongo_Activo.2             int64
Tramo_Activo.3             int64
Mongo_Activo.3             int64
Tramo_Activo.4             int64
Mongo_Activo.4             int64
Tramo_Activo.5             int64
Mongo_Activo.5             int64
Tramo_Activo.6             int64
Mongo_Activo.6             int64
Tramo_Activo.7             int64
Mongo_Activo.7             int64
Tramo_Activo.8             int64
Mongo_Activo.8             int64
Tramo_Activo.9             int64
Mongo_Activo.9             int64
dtype: object
```

```
# de cualquier dato en tipo de texto pasa a valor binario cuando a quienes
pdata.dtypes.sample(30)
```

2.4 Valores Nulos

Se presentan valores nulos en las variables como son de ventas, costos de distribución, gastos administrativos, otros gastos, otros ganancias y ganancias ingresos financieros, costos financieros y gastos de impuestos.

```
pdata.isnull().sum()/len(pdata)*100
```

```
Unnamed: 0                 0
Nacionales                 0
Extranjeros                0
Nacionales.1               0
Extranjeros.1              0
SUMCICLOS2016              0
SUMPPR2016                 0
AVP_PORCENT2016            0
Agencia_operacional        0
Tramo_del_corredor         0
Tramo_Activo               0
Mongo_Activo               0
Tramo_Activo.1             0
Mongo_Activo.1             0
Tramo_Activo.2             0
Mongo_Activo.2             0
Tramo_Activo.3             0
Mongo_Activo.3             0
Tramo_Activo.4             0
Mongo_Activo.4             0
Tramo_Activo.5             0
Mongo_Activo.5             0
Tramo_Activo.6             0
Mongo_Activo.6             0
Tramo_Activo.7             0
Mongo_Activo.7             0
Tramo_Activo.8             0
Mongo_Activo.8             0
Tramo_Activo.9             0
Mongo_Activo.9             0
dtype: object
```

2.5 Transformación de las valores nulos

Los valores nulos se convierten en cero

```
pdata.isnull()sum()/len(pdata)*100
```

```
Unnamed: 0                 False
Nacionales                 False
Extranjeros                False
Nacionales.1               False
Extranjeros.1              False
SUMCICLOS2016              False
SUMPPR2016                 False
AVP_PORCENT2016            False
Agencia_operacional        False
Tramo_del_corredor         False
Tramo_Activo               False
Mongo_Activo               False
Tramo_Activo.1             False
Mongo_Activo.1             False
Tramo_Activo.2             False
Mongo_Activo.2             False
Tramo_Activo.3             False
Mongo_Activo.3             False
Tramo_Activo.4             False
Mongo_Activo.4             False
Tramo_Activo.5             False
Mongo_Activo.5             False
Tramo_Activo.6             False
Mongo_Activo.6             False
Tramo_Activo.7             False
Mongo_Activo.7             False
Tramo_Activo.8             False
Mongo_Activo.8             False
Tramo_Activo.9             False
Mongo_Activo.9             False
dtype: object
```

```
pdata.dtypes.sample(30)
```

```
Unnamed: 0                 int64
Nacionales                 int64
Extranjeros                int64
Nacionales.1               int64
Extranjeros.1              int64
SUMCICLOS2016              int64
SUMPPR2016                 int64
AVP_PORCENT2016            int64
Agencia_operacional        int64
Tramo_del_corredor         int64
Tramo_Activo               int64
Mongo_Activo               int64
Tramo_Activo.1             int64
Mongo_Activo.1             int64
Tramo_Activo.2             int64
Mongo_Activo.2             int64
Tramo_Activo.3             int64
Mongo_Activo.3             int64
Tramo_Activo.4             int64
Mongo_Activo.4             int64
Tramo_Activo.5             int64
Mongo_Activo.5             int64
Tramo_Activo.6             int64
Mongo_Activo.6             int64
Tramo_Activo.7             int64
Mongo_Activo.7             int64
Tramo_Activo.8             int64
Mongo_Activo.8             int64
Tramo_Activo.9             int64
Mongo_Activo.9             int64
dtype: object
```

```
p
```

```
# Gráfico descriptivo
plot_data_frame(df[,c("Sexo")],
                fun.numeric=function(x) {x[is.na(x)] <- -1},
                ncol=2,nrow=2,ylab="Frecuencia")
```

```
# Gráfico Cualitativas
plot_data_frame(df[,c("Ocupacion")],
                fun.numeric=function(x) {x[is.na(x)] <- -1},
                ncol=2,nrow=2,ylab="Frecuencia")
```

```
# Gráfico de Cuantitativas
plot_data_frame(df[,c("Edad")],
                fun.numeric=function(x) {x[is.na(x)] <- -1},
                ncol=2,nrow=2,ylab="Frecuencia")
```

```
# Gráfico
plot_data_frame(df[,c("Edad")],
                fun.numeric=function(x) {x[is.na(x)] <- -1},
                ncol=2,nrow=2,ylab="Frecuencia")
```

```
# Gráfico
plot_data_frame(df[,c("Edad")],
                fun.numeric=function(x) {x[is.na(x)] <- -1},
                ncol=2,nrow=2,ylab="Frecuencia")
```

5.Análisis de cluster K-Means

El algoritmo K-Means busca agrupar en clústeres a grupos de puntos de datos que tienen una alta similitud entre ellos. En los términos de aprendizaje automático, los clústeres son grupos de puntos de datos que tienen una alta similitud entre ellos (como los puntos de datos, más cercanos) y son más proporcionales de pertenecer a nivelo cruzar vecin (Kumar, 2019).

El algoritmo de agrupamiento K-Means es un proceso iterativo que trata de centrar a los de los agrupamientos (centroides) a partirdo de puntos de datos que tienen una alta similitud entre ellos. Los centroides se mueven hasta que ya no es posible mejorar la similitud significativa en el número de clústeres posibles y centros de tenencias alcanzada (Google, 2017).

La función de K-Means está determinada por la distancia euclidiana (como puntos de observaciones pertenecientes a los centroides) y se puede expresar como:

Fuente del Algoritmo Google (Kumar, 2019)

- x_i Grupo, observaciones (Grupo 1, el número de clústeres que queremos que nos encuentre)
- c_j Grupo, el centroide asociado (Grupo1(clústeres) Con centroides de cada grupo)
- $\|x_i - c_j\|^2$ Es la distancia entre el punto de datos y el centroide del clúster al que pertenece (EUCLIDEAN) Es el cuadrado la norma del vector diferencial.
- $\sum_{i=1}^{n}\sum_{j=1}^{k}$ Es la suma de todas las distancias euclidianas, tanto de los puntos que pertenecen al centroide del mismo clúster. Es la suma cuadrada, calculada al otro cuadrado entre los puntos de datos y los centroides de los clústeres a los que pertenecen.
- $\sum_{i=1}^{n}$ Es la suma de todos los puntos de datos.

5.1 Número de clusters y calculo de inercia

Se realiza el análisis exploratorio de cuanto K-Means puede predecir la inercia conocimiento el número óptimo de clusters

```
# Cargando del paquete
library(ggplot2) library(ggpubr) library(
```

```
# Para plot
set.seed(123)
ggplot(data=df,aes(x=Edad,
       y=Ingresos,color=Sexo))+
  geom_point()+theme_bw()+xlab("Edad")+ylab("Ingresos")+
  ggtitle("Diagrama de Dispersión")+
  theme(plot.title=
        element_text(hjust=0.5,
                     face="bold"))+
  scale_color_manual(values=c("orange","orange"))
```

```
# Para plot_kmeans() del paquete factoextra
set.seed(123)
km.res <- kmeans(df[,c("Edad","Ingresos")],
                 centers=2,nstart=25)
# Gráfico
library(factoextra)
fviz_cluster(km.res, data = df[,c("Edad",
                                  "Ingresos")],
             palette = c("#2E9FDF", "#00AFBB",
             "#E7B800"),
             geom = c("point"),
             ellipse.type = "convex",
             ggtheme = theme_bw()
)
```

Se elegió 10 clusters como número óptimo. Existe tal número que cuando ya no hay ganancia significativa en la suma de los cuadrados dentro de los clústeres (inercia) y que se ve (Codo de Kaminsky) de la inercia se estabiliza.

```
# Para Kmeans() paquete ClusterSim
library(ClusterSim)
# Para plot_kmeans() paquete Data.table
library(Rtsne)
# Para plot_kmeans() paquete Rtsne
library(data.table)
# Para plot_kmeans() paquete ggplot2
library(ggplot2)
# Para plot_kmeans() paquete ggrepel
library(ggrepel)
```

```
# Para Kmeans
set.seed(123)
km.res <- Kmeans(df[,c("Edad",
                       "Ingresos")],
                 10,method="cityblock")
km.res
```

```
Data Set: df[, c("Edad", "Ingresos")]
Number of Clusters: 10
Method: cityblock
Clustering Vector:
    1     2     3     4     5     6     7     8     9    10
 1000  1000  1000  1000  1000  1000  1000  1000  1000  1000
Within Cluster Sum of Distances:
    1     2     3     4     5     6     7     8     9    10
 1000  1000  1000  1000  1000  1000  1000  1000  1000  1000
Between Cluster Sum of Distances: 0
Cluster Sizes:
   1    2    3    4    5    6    7    8    9   10
1000 1000 1000 1000 1000 1000 1000 1000 1000 1000
Cluster Medoids:
     Edad Ingresos
[1,]   20    1000
[2,]   21    1000
[3,]   22    1000
[4,]   23    1000
[5,]   24    1000
[6,]   25    1000
[7,]   26    1000
[8,]   27    1000
[9,]   28    1000
[10,]  29    1000
Clustering Algorithm: Kmeans
Clustering Time: 0.0 seconds
Clustering Quality:
Dunn: 0.0
Davies-Bouldin: 0.0
Silhouette: 0.0
Calinski-Harabasz: Inf
Rand: 0.8181818
Corrected Rand: 0.0
Jaccard: 0.3333333
Hubert: 0.0
Hubert Normalized: 0.0
Minkowski: 0.0
Minkowski Normalized: 0.0
Mutual Information: 0.0
Mutual Information Normalized: 0.0
Variation of Information: 0.0
Normalized Variation of Information: 0.0
Gamma: 0.0
Gamma Plus: 0.0
Tau: 0.0
Tau A: 0.0
Tau B: 0.0
Tau C: 0.0
Somers' D: 0.0
Somers' D A: 0.0
Somers' D B: 0.0
Somers' D C: 0.0
Pearson Correlation: 0.0
Spearman Correlation: 0.0
Kendall Correlation: 0.0
Mantel Test: 0.0
Mantel Test Normalized: 0.0
Mantel Test Correlation: 0.0
Mantel Test Significance: 0.0
Mantel Test Significance Normalized: 0.0
Mantel Test Standardized Statistic: 0.0
Mantel Test Standardized Statistic Normalized: 0.0
Mantel Test Standardized Statistic Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance: 0.0
Mantel Test Standardized Statistic Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized Correlation Significance Normalized: 0.0
```

```
# Gráfico
plot_kmeans(km.res,1:2,data=df,
            geom=c("point"),
            ellipse=TRUE,
            ellipse.type="convex",
            ellipse.alpha=0.2,
            ellipse.size=0.1,
            ellipse.show.legend=FALSE,
            ellipse.linetype=1,
            ellipse.color=NULL,
            ellipse.fill=NULL,
            ellipse.show.text=FALSE,
            ellipse.text.size=1.5,
            ellipse.text.color=NULL,
            ellipse.text.font=NULL,
            ellipse.text.fill=NULL,
            ellipse.text.alpha=0.2,
            ellipse.text.linetype=1
```

```
cluster    NaT
0          2.0
1         26.0
2         18.0
3         16.0
4         15.0
5         14.0
6         13.0
7         12.0
8         11.0
9         10.0
10         9.0
11         8.0
12         7.0
13         6.0
14         5.0
15         4.0
16         3.0
17         2.0
18         1.0
19         1.0
20         1.0
21         1.0
22         1.0
```