

Usando los datos del censo nacional de población y vivienda 2010 y 2020 y centrándonos únicamente en el municipio de Ensenada, localidad Ensenada, se hará un clustering con el método k-means de las variables por AGEB (32) de la zona norte de la ciudad

	AGEB	P_TOTAL	POB65_MAS	GRAPROES	PROM_HNV	VIVTOT	PRO_OCUP_C	PROM_OCUP	VPH_PC	OTRAENT	P_0A5
0	6006	5443.0	225.0	8.844595	2.155405	1581.0	0.993784	3.735405	624.0	1987.0	548.0
1	0435	3632.0	325.0	9.617302	2.193333	1321.0	0.853016	3.352540	584.0	1241.0	419.0
2	0524	3040.0	276.0	9.127049	2.500164	1083.0	0.816885	3.296066	497.0	1074.0	364.0
3	1005	2987.0	4.0	11.066000	1.524500	1228.0	0.829500	3.247000	423.0	1286.0	483.0
4	7714	2879.0	172.0	8.219375	2.270313	905.0	0.869375	3.345625	329.0	1057.0	367.0

En cada ageb se toman en cuenta un total de 10 variables:

P_TOTAL: Población total

POB65_MAS: Población de 65 años o más

GRAPROES: Grado promedio de escolaridad

PROM_HNV: Promedio de hijos nacidos vivos

VIVTOT: Viviendas totales

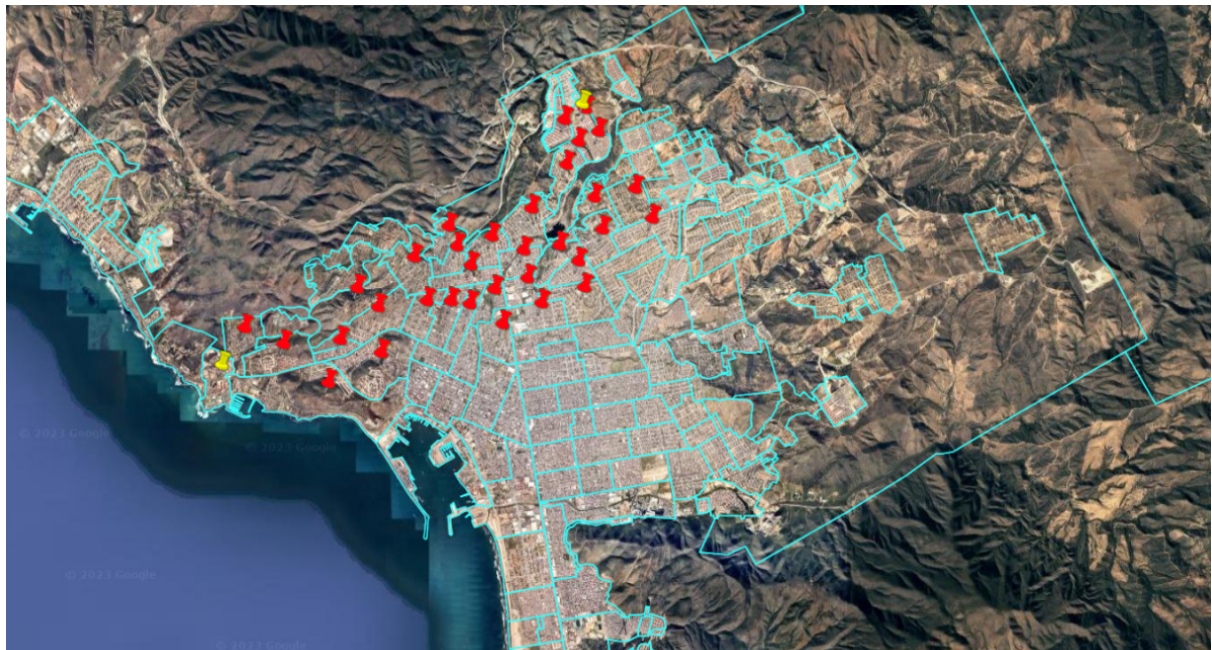
PRO_OCUP_C: Promedio de ocupantes por cuarto

PROM_OCUP: Promedio de ocupantes por vivienda

VPH_PC: Viviendas particulares habitadas que disponen de computadora

OTRAENT: Habitantes provenientes de otra entidad federativa

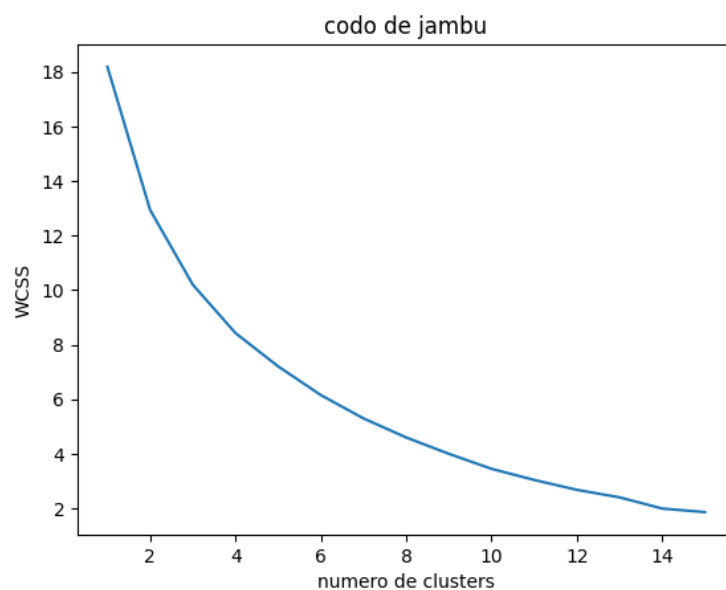
P_0A5: Población de 0 a 5 años.



Al observar los datos, se nota una diferencia significativa entre los valores de las variables, que van desde números muy grandes hasta muy pequeños, es por eso que se optó por normalizar todas las variables de forma que tomaran valores entre 0 y 1 utilizando el método *MinMaxScaler* de la librería scikit learn.

	P_TOTAL	POB65_MAS	GRAPROES	PROM_HNV	VIVTOT	PRO_OCUP_C	PROM_OCUP	VPH_PC	OTRAENT	P_0A5
0	1.000000	0.692308	0.337502	0.782807	1.000000	0.726155	1.000000	0.915225	1.000000	1.000000
1	0.629349	1.000000	0.436298	0.806701	0.800000	0.561597	0.816560	0.846021	0.565268	0.755682
2	0.508187	0.849231	0.373616	1.000000	0.616923	0.519360	0.789502	0.695502	0.467949	0.651515
3	0.497339	0.012308	0.621524	0.385346	0.728462	0.534107	0.765993	0.567474	0.591492	0.876894
4	0.475235	0.529231	0.257564	0.855197	0.480000	0.580721	0.813247	0.404844	0.458042	0.657197

Para definir el número de clusters, se utilizó el método del Codo de Jambú y uno de los artículos de referencia(*Bournazou Marcou*), que clasifica los AGEBS en 6 grupos.



El codo de jambú no resultó muy útil en la toma de la decisión.

Aplicando el clustering a los datos tanto de 2010 y 2020 obtenemos los siguientes resultados::

```

CLUSTERS 2010
CLUSTER 0: ['0492']
CLUSTER 1: ['0647', '0558', '0454', '7729', '7771', '7733', '0628', '0562', '044A', '0577']
CLUSTER 2: ['1005', '1024', '1255', '7678', '101A']
CLUSTER 3: ['6006', '0435']
CLUSTER 4: ['0524', '7714', '7752', '7786', '7697', '0543']
CLUSTER 5: ['7748', '6025', '7451', '7663', '770A', '7767', '0581', '8089', '0505']

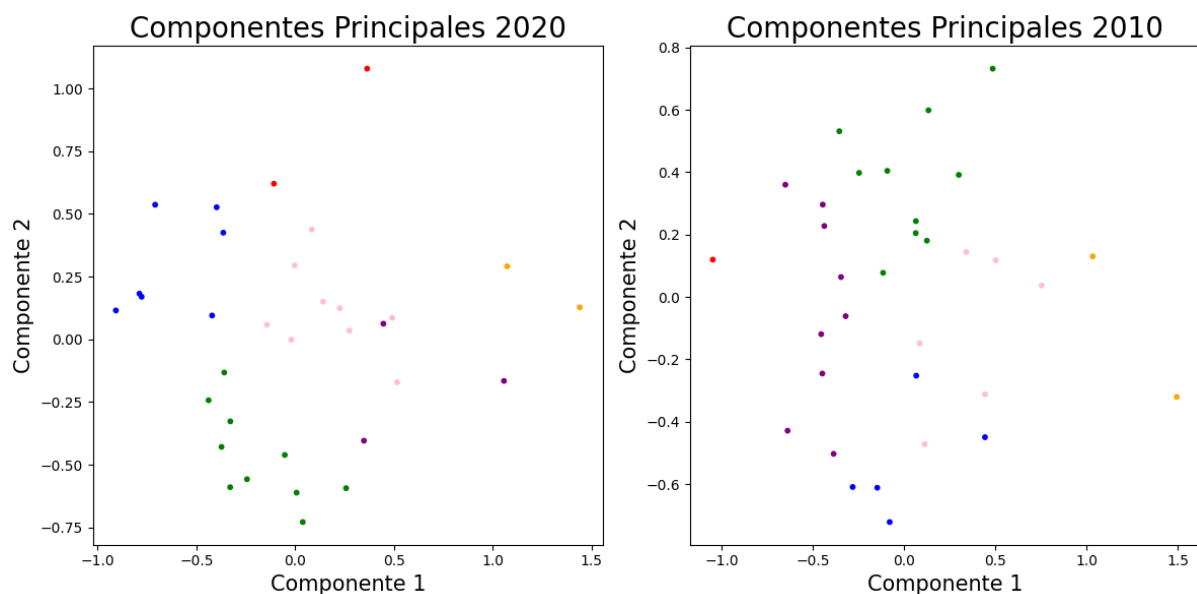
```

CLUSTERS 2020

```
CLUSTER 0: ['0647', '7771']  
CLUSTER 1: ['7697', '7678', '0543', '101A', '044A', '8089', '770A', '7663', '7451', '7767']  
CLUSTER 2: ['0628', '6025', '0562', '0577', '0581', '0492', '0505']  
CLUSTER 3: ['6006', '0435']  
CLUSTER 4: ['7752', '7714', '0524', '7786', '0558', '0454', '7729', '7733', '7748']  
CLUSTER 5: ['1005', '1024', '1255']
```

Esto es un resultado suficiente y útil para nuestro objetivo. Nótese que de momento no es posible identificar el nivel de cambio al que corresponde cada cluster. La idea es un análisis de las AGBS que brincaron de un cluster a otro entre cada censo para identificar cuales sufrieron cambios, aunque no sabemos por ahora que tipo de cambio es sea y hacia qué dirección.

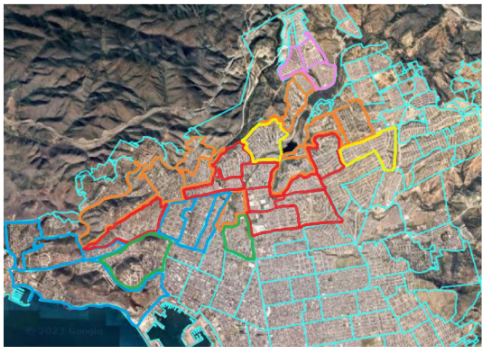
Aunque nos es suficiente conocer que agebs corresponden a cada cluster formado, vamos a visualizar los resultados aplicando un análisis de componentes principales (PCA por sus siglas en inglés) para reducir dimensionalidad utilizando la librería scikit learn.



Una forma más adecuada de visualizar los clusters es en su representación geográfica, como se muestra a continuación:



2010



2020

Las estadísticas asociadas a cada cluster por año son las siguientes:

	P_TOTAL	POB65_MAS	GRAPROES	PROM_HNV	VIVTOT	PRO_OCUP_C	PROM_OCUP	VPH_PC	OTRAENT	P_0A5
KMeans_Clusters										
0	8618.0	1004.0	11.609367	1.697222	3551.0	0.576350	2.764259	2155.0	3032.0	647.0
1	8233.0	31.0	10.426899	1.443611	3698.0	0.948207	2.990379	1233.0	3523.0	1509.0
2	17845.0	1240.0	9.123893	2.222757	5928.0	0.863812	3.380499	2526.0	6267.0	1985.0
3	20759.0	2415.0	10.631591	1.968677	7715.0	0.684362	3.058848	4262.0	7070.0	1641.0
4	8620.0	391.0	8.639790	1.930985	2786.0	0.916431	3.305852	1096.0	3498.0	1045.0
5	557.0	15.0	7.670000	0.912826	281.0	0.372609	1.648261	189.0	271.0	39.0

2010

2020

	P_TOTAL	POB65_MAS	GRAPROES	PROM_HNV	VIVTOT	PRO_OCUP_C	PROM_OCUP	VPH_PC	OTRAENT	P_0A5
KMeans_clusters										
0	7091.0	1097.0	10.807374	1.972372	2643.0	0.708516	2.948372	1507.0	2288.0	427.0
1	26547.0	3739.0	11.414890	1.747738	10255.0	0.648873	2.779205	6061.0	8495.0	1926.0
2	8483.0	854.0	10.162255	2.003602	3101.0	0.793305	3.113156	1467.0	2600.0	694.0
3	8887.0	270.0	10.806502	1.494946	3076.0	0.860434	2.982545	1668.0	3361.0	953.0
4	8554.0	1546.0	12.777090	1.546560	3905.0	0.535869	2.425435	2720.0	2949.0	549.0
5	8243.0	605.0	9.513133	1.810212	2804.0	0.858545	2.976123	1353.0	3065.0	892.0

