



Universidad
Provincial del Sudoeste
Promoviendo el Desarrollo Armónico de la Región

LABORATORIO

Clasificador zonas según el peligro

En el laboratorio, nos centraremos en un conjunto de datos que incluye registros de crímenes de 1973 en distintos estados de EE. UU., así como el porcentaje de la población que reside en zonas urbanas en esos estados. El propósito es clasificar los estados en diferentes categorías basadas en estos factores.

Integrantes: Grupo 20

- Lucia ALONSO
- Pastor PAREDES
- Viviana SUAREZ
- Heber ALONSO

<https://github.com/lu8dcf/UPSO-BigData-G20.git>

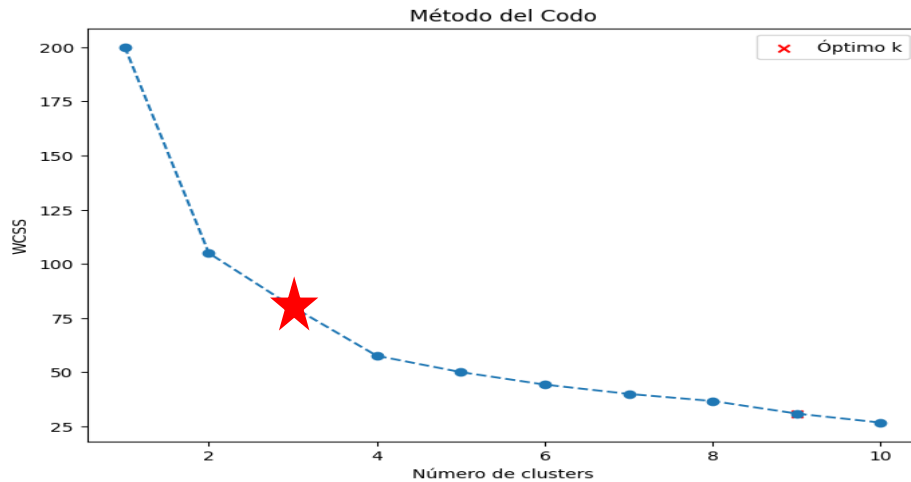
ARCHIVO clasificador_crmenes1.ipynb

Del dataframe original se realizaron dos análisis por separados.

CONCLUSION

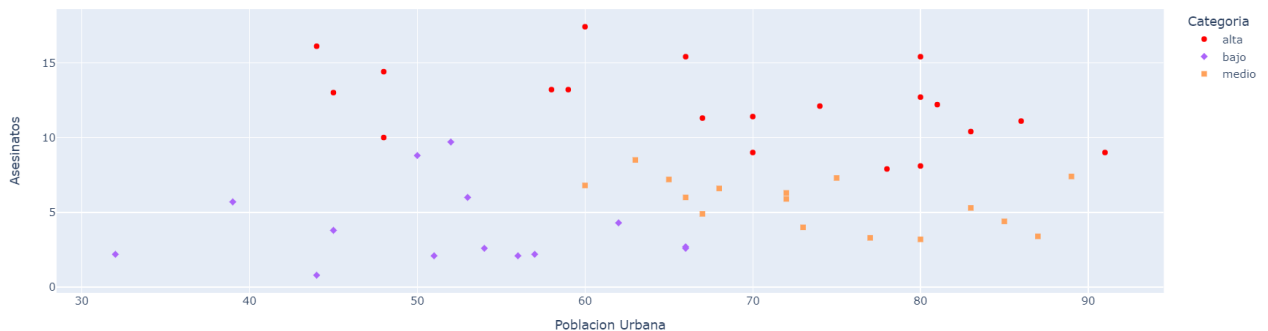
PROCEDIMIENTO 1:

Se separa del df los datos correspondientes a los 3 tipos de crímenes y la población urbana, y de ahí se grafica la curva a diferentes valores de k, y por el método del codo se selecciona k=3.



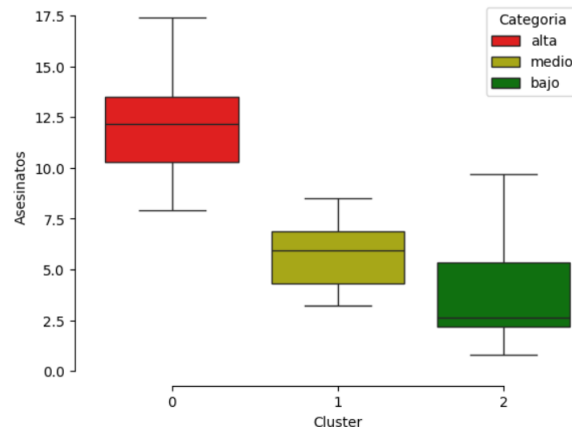
Se buscó una manera automática más eficiente, la cual toma la segunda derivada de la función para buscar el cambio de crecimiento, (este procedimiento está plasmado en el código, pero no se lo utiliza) el valor obtenido es k=9, el cual puede generar un alto costo computacional, y sobrepasa la expectativa de este análisis.

Del análisis con k=3, se muestra la distribución del Crimen "Asesinatos" en función de la Población Urbana

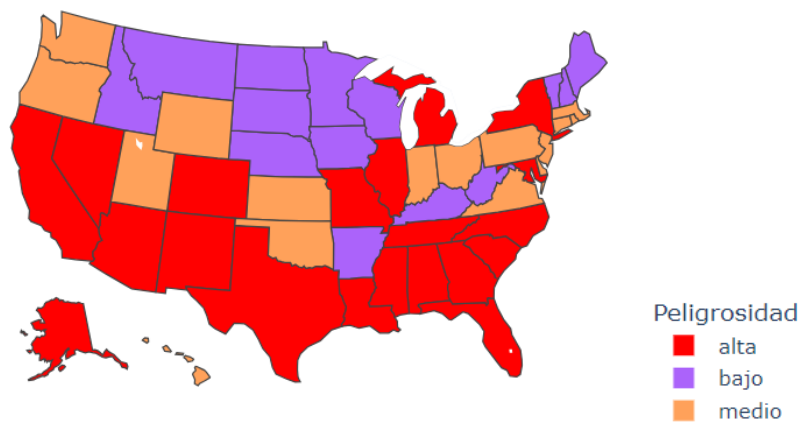


Se genera un diagrama de caja (boxplot) que representa la distribución estadística de los asesinatos en diferentes niveles de clúster. Y manteniendo los colores representativos

CONCLUSION



Como resultado final se analizan todos los crimenes en conjunto agregando al df, los tres niveles de peligrosidad por estado tomando en cuenta todos los crímenes en función de la Población Urbana.



PROCEDIMIENTO 2:

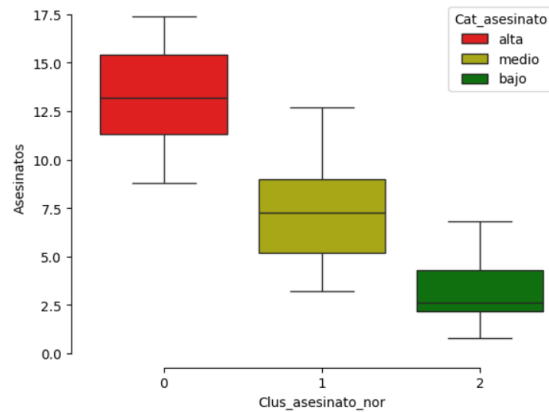
Se separa del df los datos correspondientes a cada tipo de crímenes y la población urbana, y de ahí se grafica la curva a diferentes valores de k, y por el método del codo se selecciona k=3.

Para mantener una misma clasificación se denomina los nombres de los clúster según los valores para poder compararlos entre los análisis definiendo como NORMALIZADO de esta manera clúster: “0=alta 1=media 2=baja”, refiriéndose a la peligrosidad y del mismo modo los colores Rojo, Amarillo y Verde respectivamente.

Se analizan los 3 tipos de crímenes por separado, cada uno en función de la población urbana de esta manera se obtienen 3 columnas independientes de cada tipo de crimen por estado.

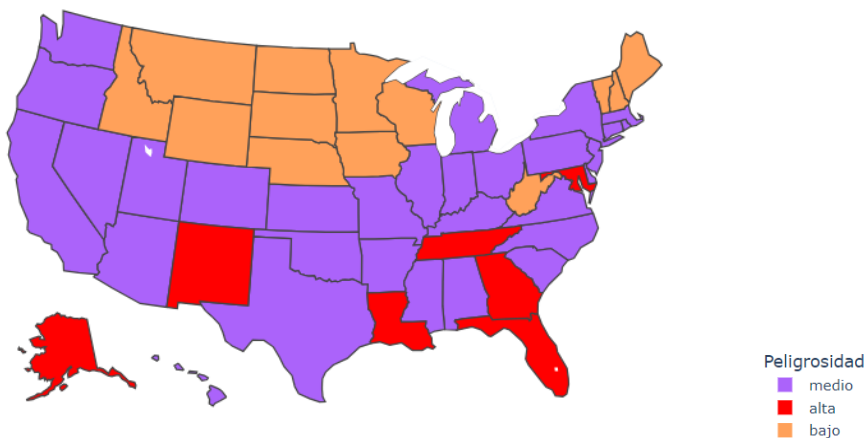
En el siguiente grafico se muestra los valores estadísticos de “Asesinatos” de los 3 niveles de cluster normalizados.

CONCLUSION



Del mismo modo se obtienen de “Violaciones” y “Asaltos”.

En el siguiente paso se suman los valores de cada análisis por estado obteniendo una nueva escala de 0 a 6, en la cual se busca el valor promedio y redondeo obteniendo un campo con el valor de todos los análisis normalizados y promediados por estado



Conclusión:

- Ambos gráficos de mapas muestran una distribución general de crímenes uniforme, en referencia a los de “baja”, sin embargo en el análisis por partes se notan más marcados, los estados más peligrosos.
- Se puede desprender que el primer procedimiento con todos los datos, devuelve una distribución más uniforme.
- Viendo en general, los mapas, resuelve en una mayor tasa de delitos en el sur de los EEUU, donde tiene la frontera con México.