



Universidad Autónoma de Nuevo León

Facultad de ciencias fisico matematicas

Maestría en ciencia de datos

Aprendizaje Automatizado

Práctica 4 Agrupamiento

Catedrático: Jose Anastacio Hernandez Saldaña

Equipo : Cynthia Selene Martinez Espinoza

Matrícula : 1011238

San nicolás de los Garza, Nuevo León, a 24 de Julio del 2024

Datos

Exploración de datos

Datos tomados de: [Repositorio Datos ML UCI](#)

En la actualidad, en muchas ciudades se utilizan bicicletas de alquiler para mejorar la comodidad de la movilidad. Es importante que la bicicleta de alquiler esté disponible y sea accesible para el público en el momento adecuado, ya que esto reduce el tiempo de espera.

Finalmente, proporcionar a la ciudad un suministro estable de bicicletas de alquiler se convierte en una preocupación importante. La parte crucial es la predicción del número de bicicletas necesarias a cada hora para el suministro estable de bicicletas de alquiler.

El conjunto de datos contiene información meteorológica (temperatura, humedad, velocidad del viento, visibilidad, punto de rocío, radiación solar, nevadas, precipitaciones), el número de bicicletas alquiladas por hora e información sobre la fecha.

Exploración de datos

	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)
0	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0
1	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0
2	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0
3	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0
4	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0

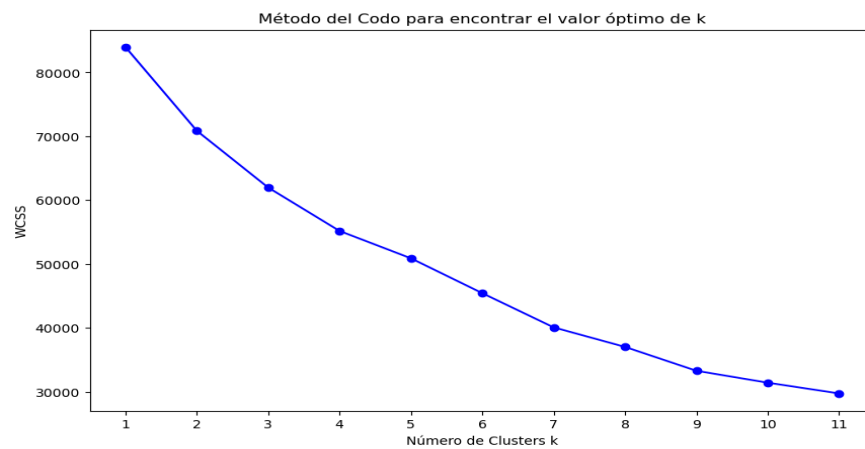
Mejor Modelo con Validación cruzada

Se aplica la validación cruzada encontrando el mejor modelo de árbol de decisión.

```
DecisionTreeClassifier
DecisionTreeClassifier()
```

Determinar el número óptimo de clusters k utilizando el método del codo

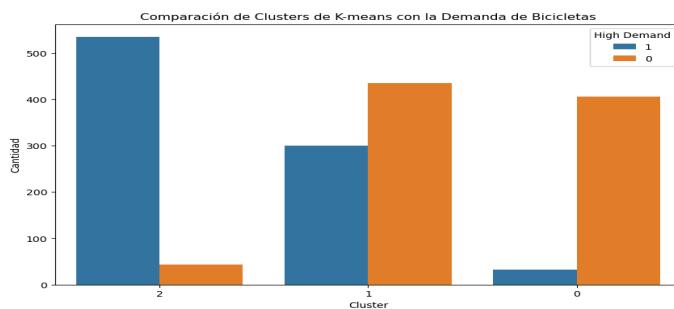
Se determinó la variable de Alta demanda de renta de bicicletas. Entrenando el modelo con K Means, para encontrar los factores con la gráfica de codo.



Entrenar el modelo K-means con el número óptimo de clusters.

Encontramos el k óptimo $=3$, con este valor entrenamos el modelo de k means, generamos los resultados de los valores de prueba proyectados.

Comparar los Resultados del Modelo K-means con el Modelo de Clasificación



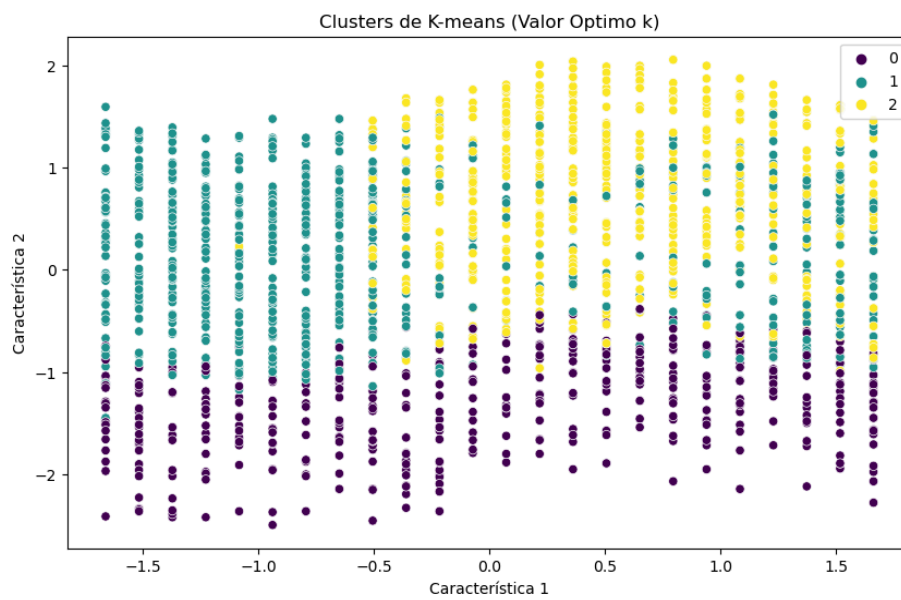
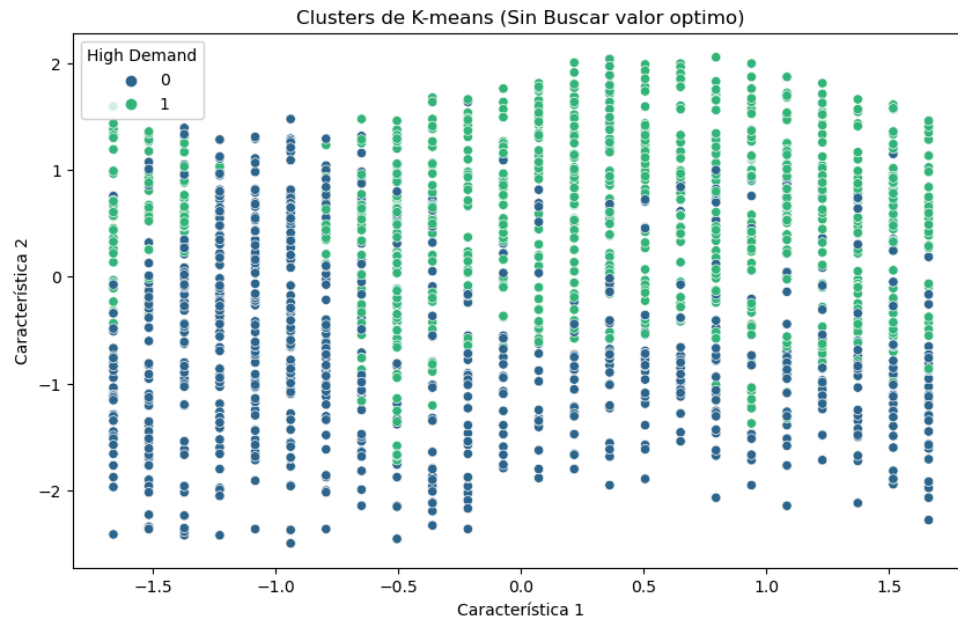
Evaluando la calidad de agrupamiento

Adjusted Rand Index: 0.2595

Homogeneity: 0.3682

Completeness: 0.2370

V-measure: 0.2884



Análisis de resultados

El modelo K-means con $k=3$ muestra una correspondencia moderada con las etiquetas de alta demanda, pero sus métricas de evaluación indican que hay mezcla de clases dentro de los clusters y que los datos de una misma clase están dispersos en varios clusters.

El modelo K-means con $k = 3$ proporciona una visión agrupada de los datos, sus métricas de evaluación sugieren que no es tan eficaz para capturar la estructura de la demanda de bicicletas en comparación con el modelo de clasificación.

El modelo K-means puede ser útil para análisis exploratorio y para identificar patrones generales en los datos, pero no es la mejor opción para predicciones precisas de demanda.