

Selección de variables para la microcuenca de la quebrada La Valeria

Kevin Villegas Tamayo

Facultad de Minas, Universidad Nacional de Colombia

Cartografía Geotécnica

Edier Vicente Aristizábal Giraldo

Octubre de 2023

Selección de variables para la Microcuenca Quebrada La Valeria

Para la modelación, se tuvieron unas variables iniciales, a las cuales se les realizó un análisis exploratorio, el cual permitió conocer el comportamiento general de las variables, y posteriormente un análisis univariado y multivariado. Las variables consideradas preliminarmente fueron Pendiente, Aspectos, Curvatura, Flujo Acumulado, Elevación, Geología y Geomorfología.

Comenzando con estas dos últimas, la geomorfología no se consideró apropiada para el uso, pues se tienen otras variables morfométricas como pendiente, aspectos, elevación, curvatura, que representan mucho mejor la geomorfología, y evitaría un sobre entrenamiento del modelo. En el caso de la geología, pudimos observar que casi toda la totalidad de los MenM ocurrían en una misma unidad geológica, tal como se muestra en la siguiente imagen. Es por esto que se descartó como variable, pues esta no nos separaría de manera satisfactoria entre celdas con MenM y celdas sin MenM.

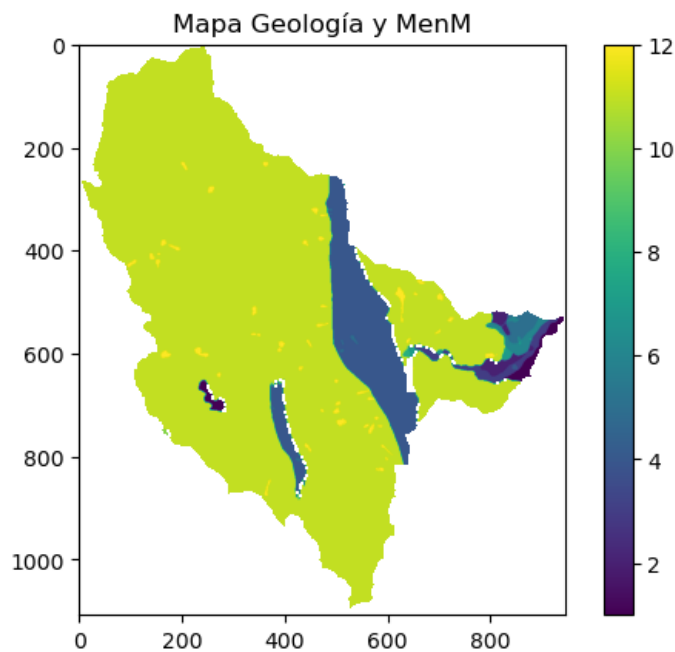


Imagen 1. Mapa Geología y MenM. Polígonos amarillos corresponden a MenM.

Por lo tanto, las variables a utilizar para los análisis y posterior selección de las mejores y más representativas, son Pendiente, Aspectos, Curvatura, Flujo Acumulado, Elevación.

El análisis exploratorio de las variables nos muestra información sobre las variables elegidas, tal como la media, desviación estándar, valores mínimos y máximos, entre otros.

	count	mean	std	min	25%	\
inventario	49380.0	0.103119	0.304117	0.000000	0.000000	
pendiente	49380.0	27.706602	12.164545	0.000000	19.218434	
aspectos	49380.0	165.081970	91.228210	-1.000000	93.376745	
curvatura	49380.0	-0.016501	9.144467	-66.649414	-4.430908	
Flujo Acumulado	49380.0	6.921527	9.295079	0.000000	1.000000	
elevacion	49380.0	2109.606932	198.661268	1730.989990	1960.134369	

	50%	75%	max
inventario	0.000000	0.000000	1.000000
pendiente	27.944465	36.187774	73.519409
aspectos	155.357010	239.020103	359.995636
curvatura	0.125000	4.604492	100.807617
Flujo Acumulado	3.000000	9.000000	50.000000
elevacion	2070.077515	2262.814453	2625.000000

Imagen 2. Análisis exploratorio de las variables.

El siguiente paso es obtener un Data Frame, el cuál nos organiza los datos y nos da mayor manejo de ellos. En la siguiente imagen se observa el Data Frame generado. En este punto cabe resaltar que las variables que se están utilizando son continuas.

[39]:

	pendiente	aspectos	curvatura	Flujo Acumulado	elevacion
0	0.757180	224.523972	0.260742	1.0	2609.993896
1	13.868361	169.804703	12.146484	0.0	2610.000000
2	14.844854	183.360001	7.830078	5.0	2609.767822
3	11.791483	195.952194	8.045898	1.0	2609.930908
4	6.822078	204.412933	4.381836	0.0	2610.000000

Imagen 3. DataFrame.

Se comienza comparando las variables continuas, que en este caso son todas las elegidas. Para esto se realiza una matriz de comparación, la cual en su diagonal contiene un histograma de densidad de Kernel, que nos muestra la distribución de los datos. Se desea observar si existe alguna correlación entre ellas, a partir de tendencias. En este caso, las celdas de la matriz diferentes a la diagonal, no muestra tendencias que den indicios de correlación entre variables.

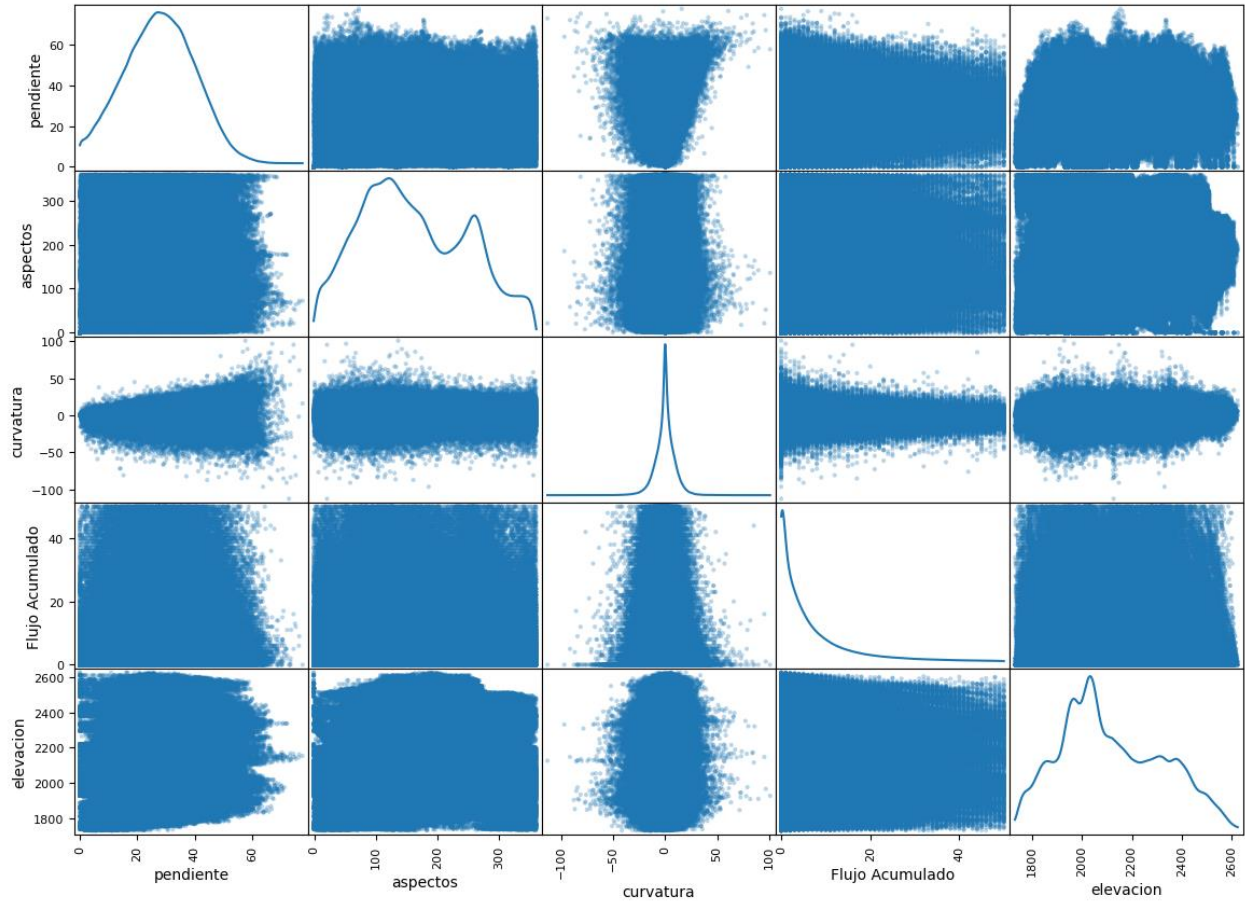


Imagen 4. Matriz de correlación 1.

Esta no correlación es también observable a partir de una matriz como la de la siguiente imagen. No se observan anomalías que indiquen lo contrario.

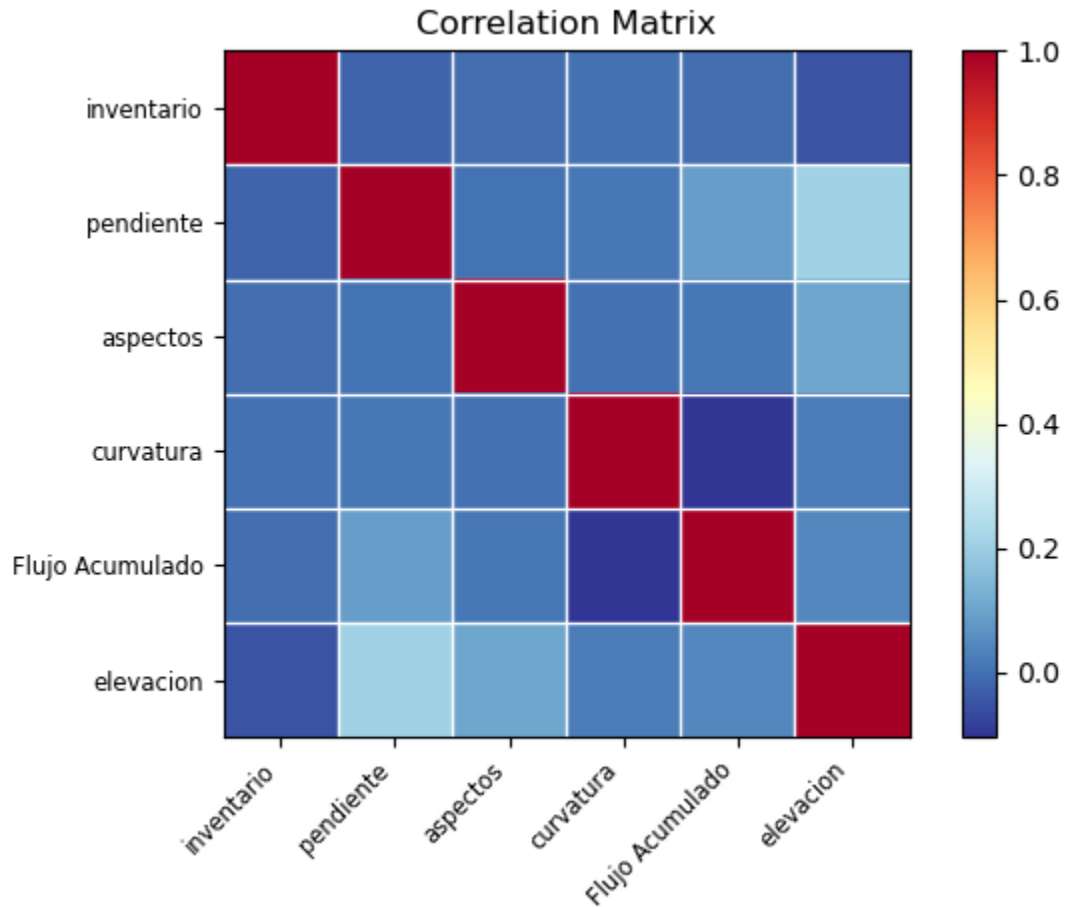


Imagen 5. Matriz de correlación.

Hay otra matriz que nos permite observar si las variables nos distinguen entre ocurrencia de MenM o no. Se hace una relación de datos por densidad y MenM. La diagonal principal nos expone una comparación entre histogramas de la variable, en donde hay y no hay MenM. Como vemos que estos, en todas las variables, son distintos, es decir, que no se tapan uno con el otro, podemos decir que las variables nos permiten hacer la distinción entre ocurrencia y no ocurrencia.

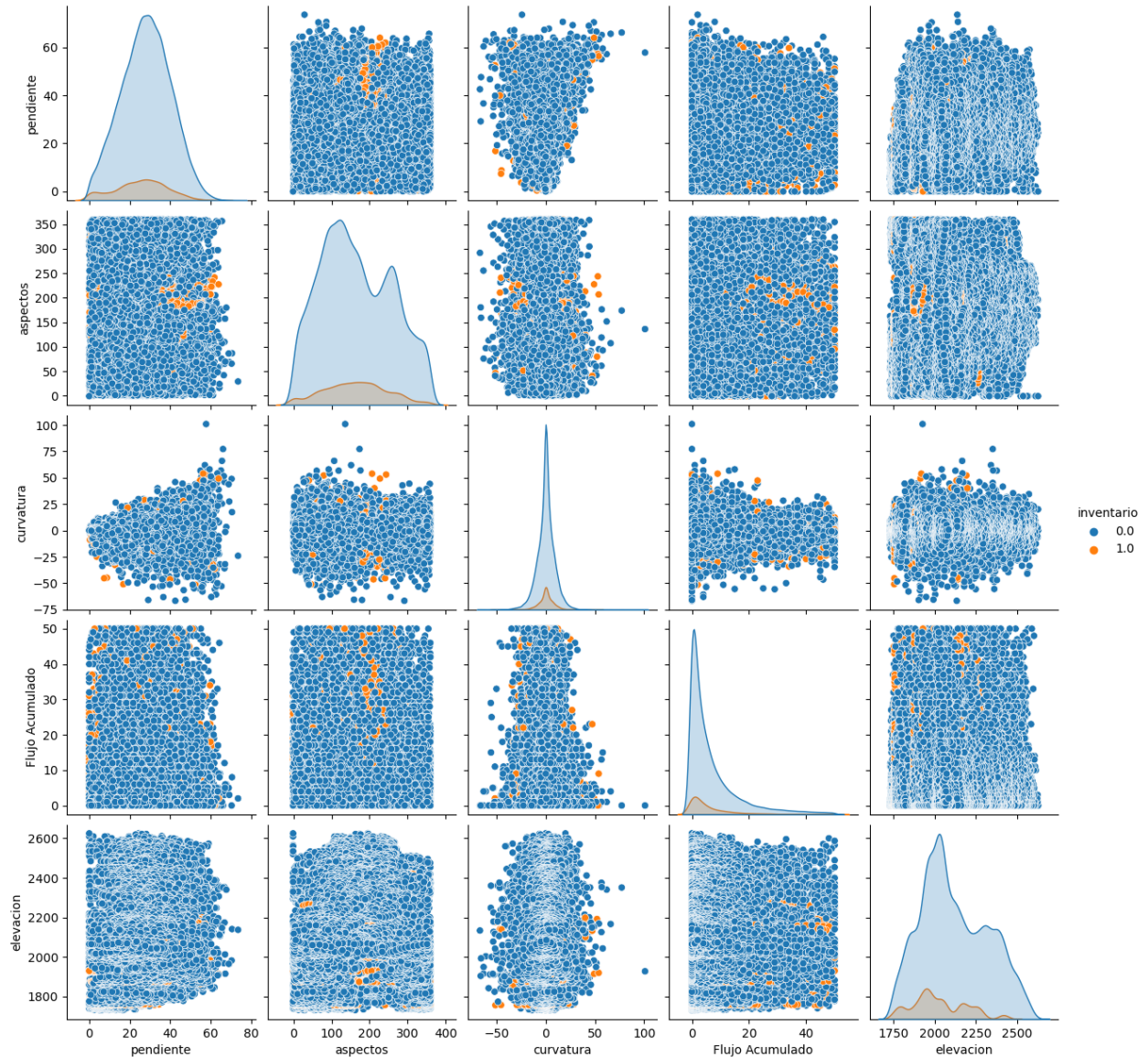


Imagen 6. Matriz de relación de datos y MenM

Análisis univariado.

Para este caso, se trata de analizar cada una de las variables elegidas, en pro de lograr conocer si estas pueden o no mostrar la ocurrencia de MenM.

- Pendiente:

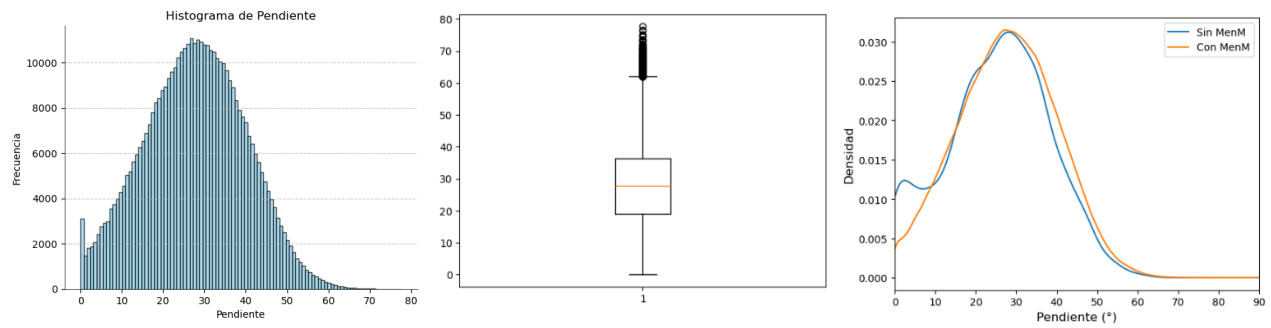


Imagen 7. Análisis univariado – Pendiente

- Elevación:

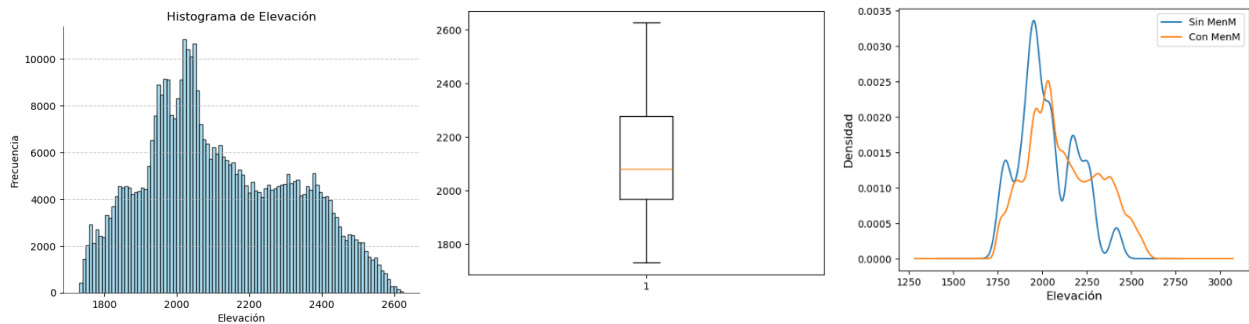


Imagen 8. Análisis univariado – Elevación

- Curvatura:

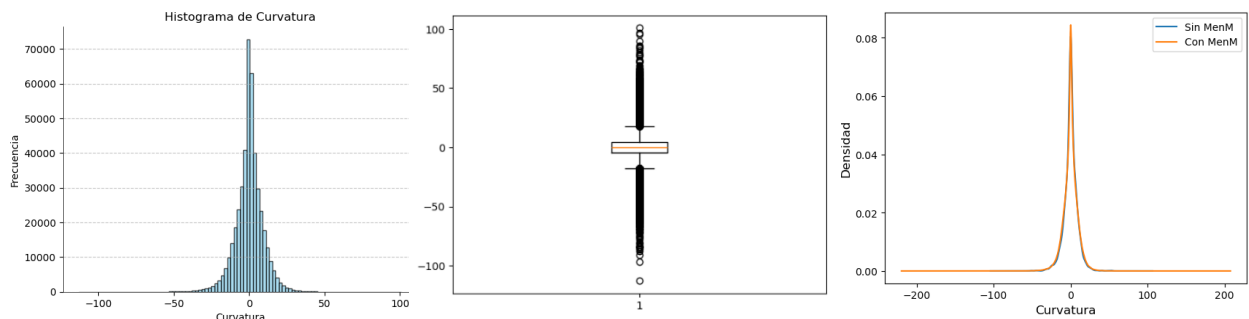


Imagen 9. Análisis univariado – Curvatura

- Aspectos:

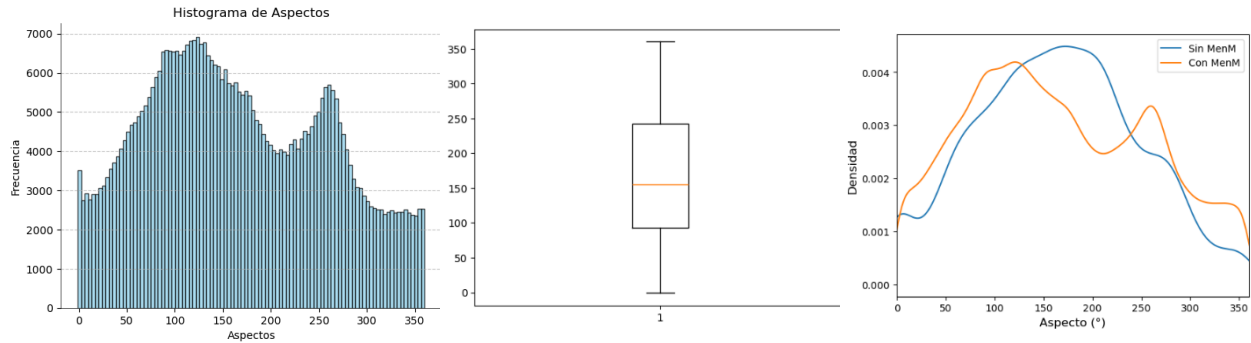


Imagen 10. Análisis univariado – Aspectos

- Flujo Acumulado:

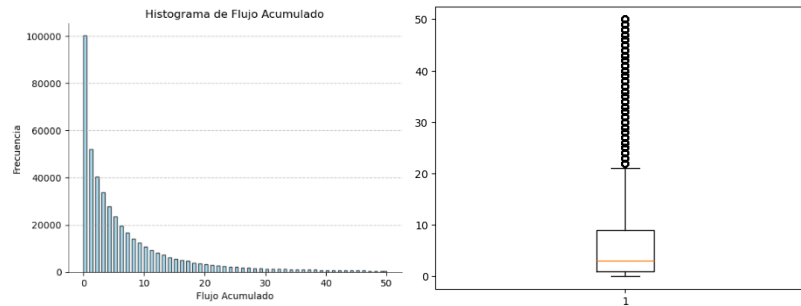


Imagen 11. Análisis univariado – Flujo Acumulado

Análisis multivariado

En el análisis multivariado, a diferencia del univariado, se trata de comparar unas variables con otras, para así ver si existe alguna correlación, a partir de tendencias o comportamientos.

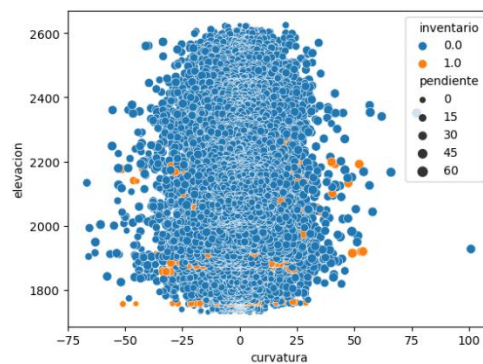
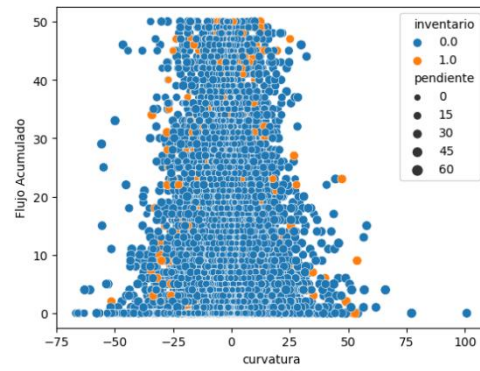
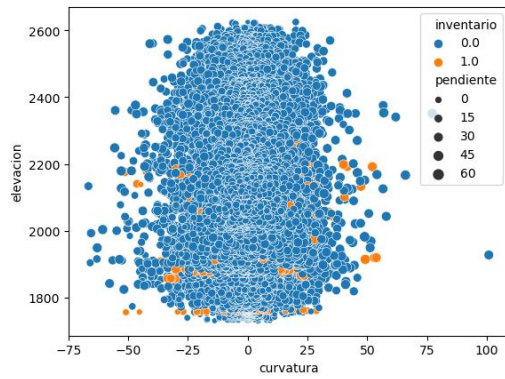
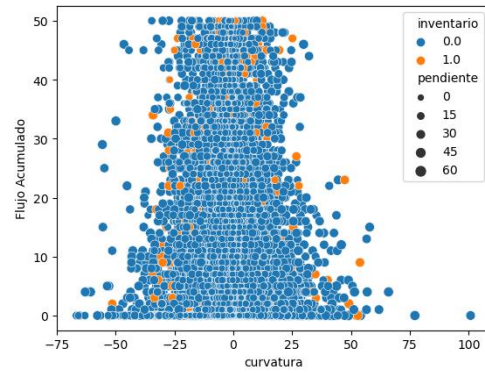
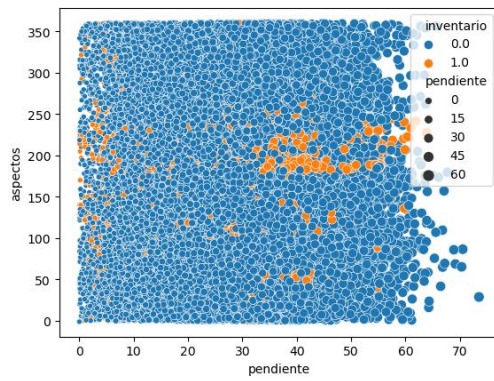
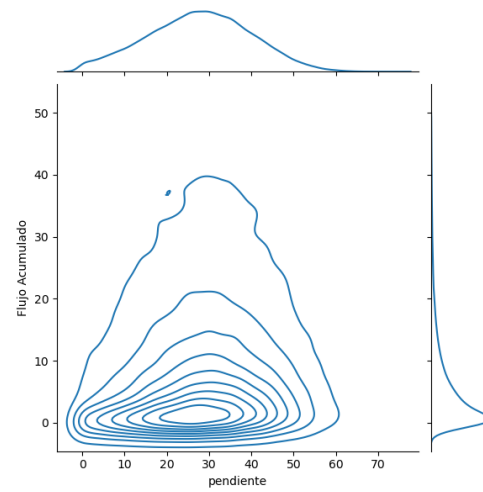
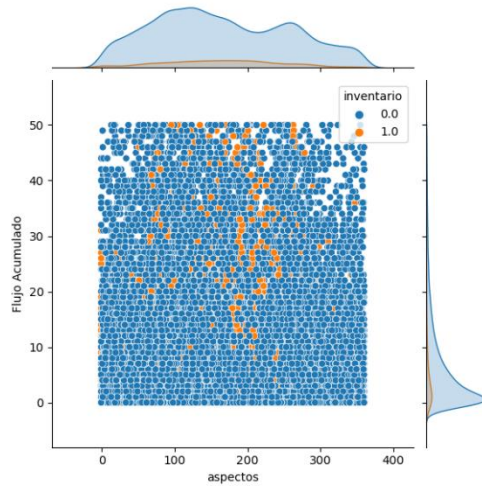


Imagen 12. Análisis multivariado

Correlación con la variable categórica

En este caso se realiza una correlación de las variables independientes elegidas, teniendo en cuenta el inventario de MenM.

	pendiente	aspectos	curvatura	Flujo Acumulado	elevacion
inventario					
0.0	27.756807	165.919617	-0.037623	6.875036	2118.929147
1.0	25.356510	163.149689	0.169285	6.765711	2025.687026

Imagen 13. Correlación con la variable categórica

Si la diferencia entre la media de si hay MenM o no es menor al 5%, nos indica que la variable tiene una significancia estadística. Como podemos observar, todas nuestras variables están en este rango.

Análisis de componentes principales (PCA)

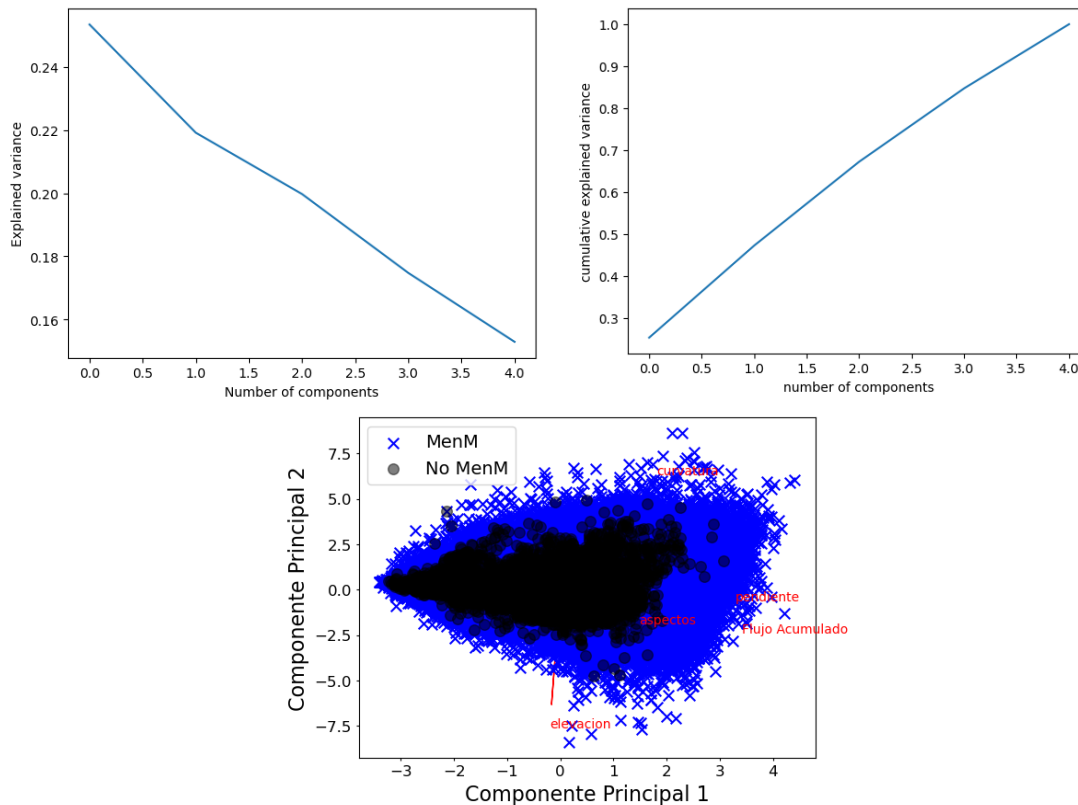


Imagen 14. PCA

A partir del PCA podemos observar que las variables que más nos permiten diferenciar entre ocurrencia o no de MenM son aspectos, pendiente, flujo acumulado. Sin embargo, tanto elevación como curvatura tienen una tendencia que muestra que con el aumento, la apariencia de MenM se hace mayor, por lo cual se hacen igualmente relevantes.

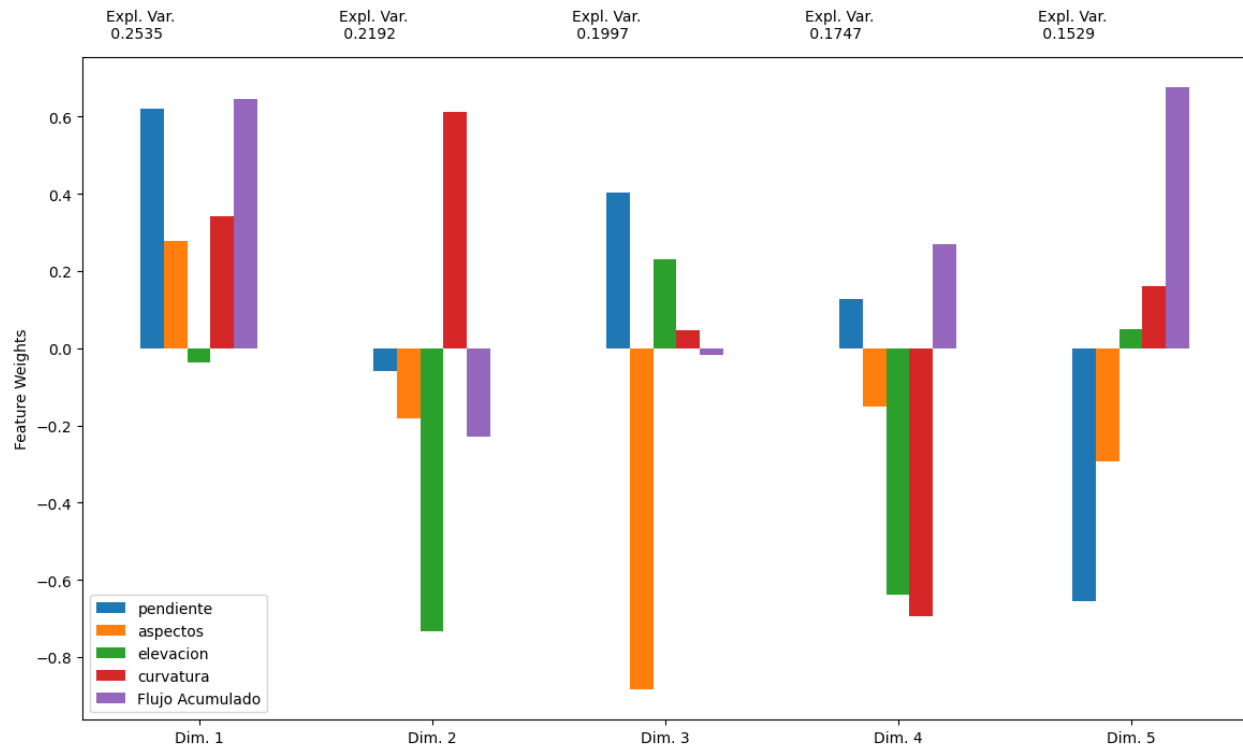


Imagen 15. PCA

A partir del análisis realizado, se puede concluir que las variables a utilizar son las propuestas en un inicio: Pendiente, Aspectos, Curvatura, Flujo Acumulado, Elevación.