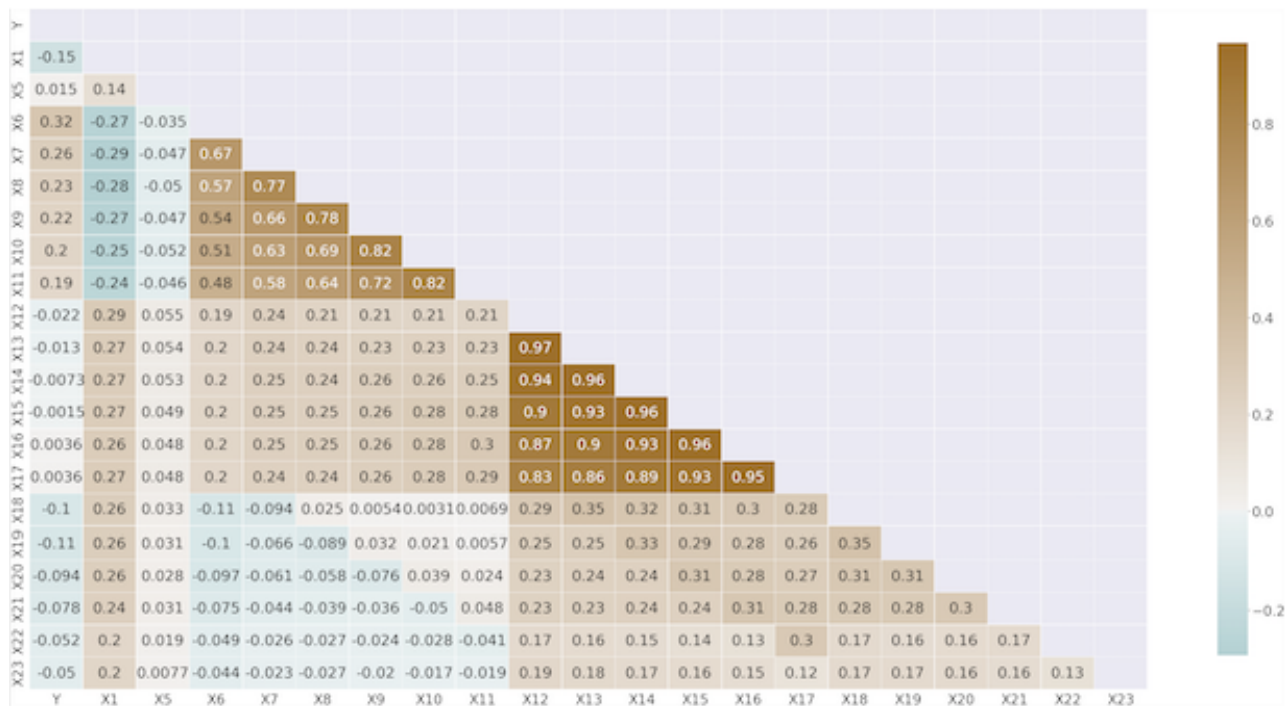
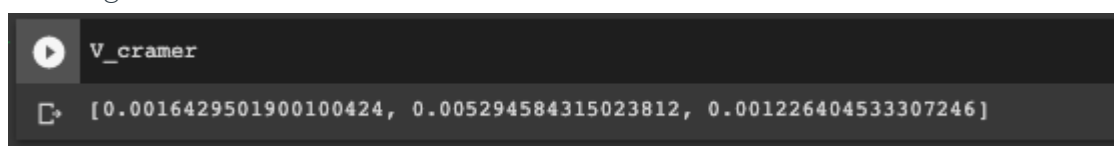


¿Qué datos considero más importantes? ¿Por qué?

Las variables continuas que considero más importantes son las variables X1,X6, X7, X8, X9, X10, X11 esto debido a que son las variables con una correlación de pearson más grande con la variable objetivo Y



Las variables categóricas, se utilizó una v de cramer es una medida de asociación entre dos variables nominales, dando un valor entre 0 y +1. Se basa en la estadística chi-cuadrado de Pearson para ver la relación de asociación entre la variable objetivo y las variables categóricas



No encontramos variables categóricas asociadas fuertemente a la variable Y

¿Se eliminan o reemplazan datos nulos? ¿Qué se hizo y por qué?

Se reemplazan valores nulos con la finalidad de las técnicas de asociación tomen el mayor número de observaciones, en este caso escogí la media para

¿Es necesario ordenar los datos para el análisis? Sí / No / ¿Por qué?

No es necesario por que este tipo de técnicas de asociación ve los valores de columna vs columna

¿Existen problemas de formato que deban solucionar antes del proceso de modelado? Sí / No / Por qué.

En mi caso todos los valores se tomaron como flotante o valores enteros, por lo que el manejo de técnicas estadísticas

¿Qué ajustes se realizaron en el proceso de limpieza de datos (agregar, integrar, eliminar, modificar registros (filas), cambiar atributos (columnas)?

En este caso solo observe un ID que fue excluido de todos lo análisis la primera parte consiste en identificar el tipo de variables que estamos trabajando para así saber que técnicas descriptivas podemos utilizar, o que que tipo de valor podemos imputar valores ausentes, mi forma de trabajar la información consistió en una metodología SEMMA, que dentro de los primeros pasos está la exploración, modificación (outliers y missing) y los pasos restantes no fueron necesarios por el momento