Reporte de “Extracción de Datos”

En una visión general del análisis que se realizó pudimos ver diversos puntos a comentar en el siguiente documento. Teniendo en cuenta las mecánicas de limpieza, sustitución y visualización es que se realizó un análisis de una base de datos elaborada por alumnos del Tecnológico de Monterrey respecto a las tiendas de retail.

En primera instancia vemos que la base de datos cuenta con 106 columnas de datos cuantitativos así como cualitativos, con 171 entradas. Haciendo un análisis en cuestión de los datos se puede ver que hay un rango amplio de datos nulos o falta de datos en diversas de las columnas por las cuales se hace una limpieza de toda la base de datos utilizando la media de cada una de las columnas así como también para los datos cualitativos se tome el valor posterior al suyo para poder llenarlo.

Realizamos una limpieza de los datos atípicos con los cuartiles de cada una de las columnas para eliminar esos outliers y que la muestra en la que hagamos el análisis visual sea lo más óptimo posible. Por lo tanto, el rango intercuartílico es útil para resumir la variabilidad de un conjunto de datos y proporcionar una medida robusta de la dispersión que no se ve afectada por valores extremos o atípicos. Finalmente, el rango intercuartílico es una herramienta útil para analizar y describir bases de datos.

En la parte de visualización utilizamos 20 de las columnas cualitativas, en las cuales realizamos tablas de frecuencias para poder tener datos numéricos y poder hacer gráficas para tener una visualización de cada una de las respuestas de la manera más adecuada. Como podemos ver en esta parte muchas de las columnas cualitativas que se tienen realmente no era tienen todas la misma relevancia o importancia como para realizar un análisis visual de esas variantes.