Descripción del caso de estudio:

Para realizar este proyecto, se escogió la temática del covid-19, ya que vimos que, en este año, la enfermedad ha cambiado la vida de todos drásticamente. Hemos querido utilizar los datos que se han estado reuniendo a lo largo del año para poder administrar la información y hacer predicciones, gráficos y deducciones con distintos algoritmos existentes a partir de estos.

Procedimiento:

Nuestro proyecto consiste en administrar la información de las bases de datos para poder tener una visión más amplia de lo que se vive en la pandemia hoy en día. Todo esto se logra a partir de la preparación de datos (muestro, normalización, imputación y eliminación de valores anómalos y outliers), modelación de los datos estructurados (modelado, implementación, poblado y serialización), transformación y consultas exploratorias (selección, unión, división, filtrado y uso de algunos valores de estadística descriptiva), exploración visual de datos (diagramas de dispersión, barras, cajas, serie de tiempo, etc) y modelado (regresión lineal, regresión multivariable, SVM, PCA, KNN y Kmeans).

Para lograr este caso, hemos recogido de distintas bases de datos de distintas plataformas de base nacional e internacional de datos abiertos y le hemos aplicado todas estas funcionalidades.

Conclusiones:

Mediante al manejo, manipulación y administración de los datos, es posible realizar todo tipo de predicciones futuras, tener registros en tiempo real de lo que está pasando a nuestro alrededor y poder evaluar las distintas situaciones para poder actuar en algún ámbito en nuestra vida cotidiana o en caso de gobiernos, realizar acciones que ayuden a su pueblo, en caso de empresas, realizar movimientos estratégicos para ser exitosos sin tener que repetir errores de pasadas empresas. En otras palabras, el uso de la manipulación de datos tiene muchísimos campos que pueden servir a todos de distintas maneras.