METODOLOGÍA

*GET\_DATA*

*R TO*

*PRE-PROCESSING*

*R TO*

*PROCESSING*

* OBTENCIÓN DE DATOS

La Ciudad de México cuenta con un banco de datos de carácter público, accesible en la dirección <https://datos.cdmx.gob.mx/pages/home/>.

Los datos están clasificados por tema principal y publicados en diferentes formatos (JSON, csv, XML, etc), además algunos pueden ser manipulados de forma automática usando API’s integradas.

Obtuvimos los datos que se encuentran en la categoría de Medio Ambiente y Cambio Climático, en este grupo se obtuvo la información del periodo de 2000 a 2020 para datos de Medición Meteorológica (<https://datos.cdmx.gob.mx/explore/dataset/medicion-de-meteorologia/information/>) y Concentración de Contaminantes en el Aire ([http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBhnmI=%27&opcion=Zg](http://www.aire.cdmx.gob.mx/default.php?opc='aKBhnmI='&opcion=Zg)==)

* CREACIÓN DE FUNCIONES

Los datos de forma bruta contienen información de diferentes variables meteorológicas (temperatura, velocidad del aire, presión, etc) y contaminantes (CO2, PM10, PM2.5, CO, etc) registrados cada hora por cada estación instalada, sin embargo, no existen datos de forma continua pues hay estaciones que son nuevas o sufrieron algún inconveniente que imposibilito el registro de información.

Dado lo anterior, los datos contienen valores del tipo Not A Number (NAN) que imposibilita la manipulación práctica de los datos, por tal motivo se desarrollo una función que elimina este tipo de datos de la fuente de datos y a la vez, selecciona una únca variable para facilitar el análisis

filtroNA: *base de datos, variable*

*por cada dato de tipo variable en base de datos:*

*si dato es NAN:*

*retirar dato*

*en otro caso:*

*mantener dato*

A la par se crearon varias funciones que obtienen promedios en diferentes escalas de tiempo, a saber diario, mensual, anual.

* APLICACIÓN DE LAS FUNCIONES

Las funciones fueron realizadas para manipular de forma compacta los datos y obtener un único archivo csv que almacena los promedios anuales de cada estación siguiendo la siguiente lógica

por cada año:

por cada variable de interes:

df ← leer (basededatos.csv)

df2csv(filtroNA(df,variable),año)

* GRAFICA DE RESULTADOS

Usando las funciones *plot, char.Correlation* se grafico el promedio anual de las estaciones con más información contenida.