

Transformación

Precio_Venta:

Para crear esta tabla se han utilizado las tablas Precio_Venta e Historico_Venta. Se han modificado los nombres de los distritos, nombres de columnas, cambiado el formato de los registros, entre otras transformaciones. Finalmente, se han concatenado ambas tablas y se han creado nuevas columnas mes y año para facilitar el trabajo posterior.

Precio_Alquiler:

Se ha modificado el nombre de la columna 'zona' para pasar a llamarla 'distrito'. Una vez hecho eso, se han adaptado los nombres de los distritos para ajustarlo a nuestro estándar. Se han cambiado los valores para eliminar el % de los números, convirtiéndolos en 'float'.

Renta_Capita:

Limpieza de datos, escribir los distritos de la misma forma que en las otras tablas y pivotar la tabla para poner de columnas los distritos. Cree una regresión lineal para calcular los años que nos faltaban creando una nueva tabla con todos los años entre 2002-2022. Cambio de la columna year por tipo fecha DD/MM/YY

Padrón:

Partimos de 108 tablas conteniendo el padrón en Madrid, por distrito, de los 9 últimos años, mes a mes, acabando en diciembre de 2022.

Hemos tenido que realizar operaciones aritméticas, ya que en cada una de las tablas no se contaban las personas empadronadas por cada uno de los distritos (que daría como resultado una tabla con 21 filas). En su lugar, había una serie grande de instancias que, al sumarlas, sí nos daba el total de nacionales y extranjeros por distrito.

Se incluye un valor 'total', suma de nacionales y extranjeros, para conseguir finalmente un registro de empadronados por distrito, mes a mes, desde enero de 2014.

Datos_Demográficos:

Se cargan los archivos, se aplican modificaciones para convertir los datos en formatos adecuados igualando a su vez nombres de columnas para facilitar la conexión con las demás tablas y se utilizan funciones para completar los valores faltantes mediante la media de años cercanos. Luego, se realiza una combinación de los DataFrames para crear una tabla completa de valores sociodemográficos y se redondean los valores para una presentación más limpia. El resultado es una tabla con información relevante sobre paro, calidad de vida, seguridad y zonas verdes en Madrid para los años 2016 a 2022

Interes_Busqueda:

Esta tabla ha requerido pocas transformaciones, tan sólo se ha creado columna mes y año para que sea más sencillo trabajar en SQL

Euribor:

Pocas transformaciones, renombramos columnas para que sean más fácilmente interpretables.

Tabla Master Distrito:

Unimos las rentas medias de los distritos en 2023, y dividimos en distritos con rentas altas, medias altas, medias bajas y bajas.