Contenido

[PLANIFICACIÓN 2](#_Toc137399040)

[NOTAS 3](#_Toc137399041)

# PLANIFICACIÓN

Los pasos a seguir han sido y van a ser:

1. Búsqueda por las diferentes páginas gubernamentales sin saber que datos vamos a analizar
2. Extracción de información candidata
3. Comparación entre DataSets y se decide hacerlo de la ciudad de Alicante
4. Fase Extracción = Repetimos fase 1, 2 y 3. Finalmente nos decantamos por los datos que se encuentran en el fichero DataSet\_formatoInicial.zip extraídos de la web \_\_\_\_\_\_\_\_
5. Se toman el formato del año 2022 como el definitivo
6. Fase Transformación = Se crean los bots necesarios para transformar los datos
7. ¿Realizar borrado de datos NULOS o sin campos o dejarlos?
8. Fase Carga = Tras tener los datos normalizados, se crea un bot para cada tipo de contrato, cuya función es cargarlo en nuestra base de datos o DataMart

# NOTAS

**04/11/2022**

Comienzo la búsqueda en el portal del gobierno de España:

<https://datos.gob.es/es/catalogo>

Consigo encontrar diferentes datasets pero sólo uno es más o menos grande, los demás son pequeños y ninguno de ellos tiene buenas categorías, solo contienen números sin tener claro a que se refieren.

**05/11/2022**

Por ello decido buscar en la página web del ayuntamiento de Alicante y consigo llegar al portal de transparencia y al apartado donde se recogen todos los contratos públicos que el ayuntamiento permite descargar.

<https://www.alicante.es/es/gobierno-abierto/transparencia/contratación-pública/contratos>

Tras navegar por el portal decido descargar todos los datos e intentar juntarlos para comprobar si toda la información que encuentro aquí es mi dataset final.

Justo antes de empezar a juntar los datos me planteo que a lo mejor tengo que probar la API del portal del gobierno de España antes de descartar toda la información allí.

**07/12/2022**

Junto todos los contratos menores por año que están en pdf. Falta por investigar los datos en Excel.

**03/04/2023**

Vuelvo a buscar datos porque los que tengo no parecen ni correctos ni suficientes. Para ello, me he informado a través de ChatGPT sobre conocimientos de la administración pública para poder entender los contratos y los datos que vamos a recolectar.

Tras ello, voy a intentar agrupar los datos que encuentro.

He tomado la decisión de que nuestro DataSet va a estar formado por los contratos formalizados y los menores en cada año. Tendré que hacer una tabla también con los desistimientos, renuncias, desiertos y modificados para actualizar los nuevos o directamente, actualizarlos antes de entrar en la base de datos. En mi opinión sería más sencillo meterlo todo en la base de datos y luego actualizarlo o hacer búsquedas que comprueben que no está modificado\*.

**05/04/2023**

He descargado todos los datos desde 2017 a 2022, en los últimos años encuentro todos los datos en formato Excel pero a partir de 2020 ya encontramos algún fichero en PDF. Tras eso se empieza a notar un cambio hasta llegar a 2017 con todos los datos en PDF, por lo que debo realizar algún programa en Python utilizando la librería Pandas para normalizar los formatos de los contratos e introducirlos a la base de datos.

He decidido hacerlo con Pandas porque tras consultar diferentes opciones, veo que es la más actual y sencilla. Creo que podría realizar todas extracciones con formatos diferentes con la misma tecnología.

**11/06/2023**

Aquí voy a mantener un histórico de como estaban los datos para ver de donde vengo y a donde voy. He guardado un ZIP con los datos iniciales, voy a describirlos:

**2022 (.ODS)**

Formalizados

Renuncias

Modificados

Prorrogados

Menores

**2021 (.ODS)**

Formalizados

Renuncias

Modificados

Prorrogados

Menores

**2020**

Formalizados **(.ODS)**

Renuncias **(.ODS)**

Modificados **(.ODS) y (.PDF)**

Prorrogados **(.ODS) y (.PDF)**

Menores **(.ODS) y (4trim)**

**2019**

Formalizados **(.ODS)**

Renuncias **(.ODS)**

Modificados **(.ODS) y (.PDF)**

Prorrogados **(.ODS) y (.PDF)**

Menores **(2.ODS) y (2.PDF)**

**2018 (.PDF)**

Formalizados

Renuncias

Modificados

Prorrogados

Menores

**2017 (.PDF)**

Formalizados

Renuncias

Modificados

Prorrogados

Menores

Voy a coger el formato de ODS del año 2022 como el definitivo y crear un bot que coja esos datos de los archivos en PDF.

Empezaré con los Formalizados de todos los años.