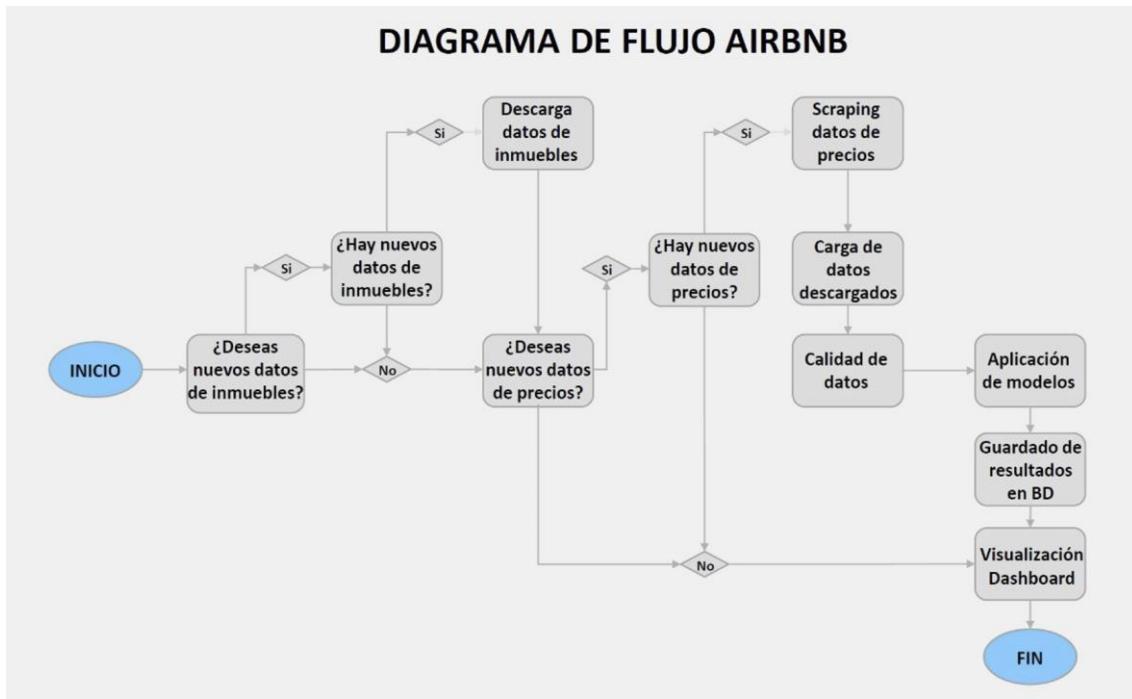


## DIAGRAMA DE FLUJO PROYECTO AIRBNB



La explicación que voy a redactar es una explicación funcional para la comprensión del flujo.

El punto de partida desde el que voy a empezar la explicación es que ya hay un dashboard creado con información y se va a evaluar si se renueva la información contenida o se sigue consumiendo la información que tiene actualmente.

El dashboard está alimentado por dos orígenes de datos:

- Datos sobre inmuebles de Airbnb
- Datos sobre precios medios de Idealista

### 1. Flujo funcional del MVP

Usuario decide si quiere actualizar los datos de los inmuebles y precios y debe comprobar si existen datos nuevos

En caso negativo en ambas decisiones, o si no hubiera datos nuevos para actualizar la información, pasaríamos directamente a la visualización del dashboard y se finalizaría el flujo

En caso de querer y poder actualizar los datos contenidos en el dashboard, el flujo del proceso sería descargar los datos de los inmuebles de Airbnb y hacer el Scraping de los precios de idealista

A continuación, se cargarían los datos en una base de datos, se haría un proceso de calidad de datos a través de un pipeline, aplicaríamos los modelos definidos previamente de Machine Learning, guardaríamos los resultados en la base de datos y se refrescaría el dashboard para su visualización

## **2. Componentes técnicos por etapa del flujo**

1. Datos de inmuebles:
  - Interfaz de usuario: Web de Airbnb
  - Almacenamiento o bases de datos propia creada para ad-hoc
  - Modelo de datos: Existe un modelo de datos ya que todos los inmuebles están identificados con su id y también existe un id de propietario e id de cliente
2. Datos de precios:
  - Interfaz de usuario: Scraping a csv realizado con la extensión de Google Instant Data Scraper
  - Almacenamiento o bases de datos propia creada ad-hoc para el propósito
  - Modelo de datos: Los precios de Idealista se identifican en el datamarkt analítico según el distrito identificado en la tabla de Airbnb
3. Carga de datos en base de datos propia creada ad-hoc para el propósito.
  - Modelo de datos: Se mantienen todas las características de las tablas de origen
4. Calidad de datos y Aplicación de modelos:
  - Lógica de negocio aplicada para la parametrización de la información contenida sobre los inmuebles
5. Guardado de resultados en BD:
  - Almacenamiento o bases de datos propia creada ad-hoc para el propósito
  - Modelo de datos: El datamark analítico es una sola tabla que contiene toda la información necesaria para su consumo
6. Visualización dashboard:
  - Interfaz de usuario: Power BI
  - Almacenamiento o bases de datos propia creada ad-hoc para el propósito

## **3. Datos necesarios en cada etapa**

1. Datos de inmuebles: Datos externos públicos descargados de la web de Airbnb. Fáciles de conseguir
2. Datos de precios: Datos externos públicos scrapeados de la web de Idealista. Fáciles de conseguir
3. Carga de datos en base de datos:
4. Calidad de datos y Aplicación de modelos: Aquí hay datos introducidos por el usuario porque la información de Airbnb no contiene los metros cuadrados de los inmuebles y nos hace falta ese dato para calcular el precio de la vivienda según el precio del distrito marcado en Idealista
5. Guardado de resultados en BD:
6. Visualización dashboard: