Objetivo General: El problema de trabajo contempla predecir los precios de propiedades en la ciudad de Melborne, Australia. Para ello se cuenta con un *DataSet Base*, que es complementado con datos sobre precios de alquileres obtenidos de la plataforma *AirBnB*.

1. En una primer etapa se configura una base de datos SQL para contener la información de orígen.
2. Para enriquecer con datos la tabla original de Melbourne, incopora información de la DB AirBnB respecto de los precios de alquileres diarios, semanales y mensuales. Se obtiene un archivo .csv final llamado *melb_data_extended.csv*.
3. Se eliminan columnas que no se consideran relevantes al problema de predicción de precios. Ellas son:
   - *Propertycount*: creemos que es solo un registro de numeración de propiedades por suburbio
   - *index.1*: una columna de indices que proviene de un Join anterior
   - *zipcode*: ya tenemos esa información en la variable Postcode
   - *Date*: existen solo 58 registros de esta fecha, con lo cual suponemos es la fecha de carga del dato. Es irrelevante al precio.
   - *SellerG*: figuran apellidos, deben ser los responsables de administrar la venta de la propiedad. Irrelevante al precio.
   - *Address*: la dirección específica de cada propiedad. No describe entorno.
   - Bedroom2: es una variable que contiene información redundante y de menor densidad que la variable *'Room'*.
4. Se analiza la presencia de outliers en el dataframe combinado: Se elimina 1 valor extremo en la variable 'Landsize' y 'BuildingArea'.
5. Se analiza de manera crítica la variable 'YearBuilt'. Se eliminan datos de propiedades con fecha de construcción anterior al año 1900.
6. Se identifican necesidades de imputación de datos, ya que hay faltantes en diversas columnas. La acciones de imputación realizadas sobre las distintas columnas son:
   - *Car*: solo existen 48 datos a ser imputados. Se asumen que esos casos **"no tienen cochera"**, se hace una **imputación simple asignando valor = 0**.
   - Se realizaron imputaciones sobre las variables *'BuildingArea'*, *'YearBuilt'*, *'avg_weekly_price'* y *'avg_monthly_price'* empleando el método IterativeImpter, basado en dos estimadores:
     a) *KNeighborsRegressor()*
     b) *BayesianRidge()*
7. Se realizaron encodings de variables categóricas para fines posteriores de predicció de precios. Previo a esos encoding se:
   - Redujeron categorías en las columnas *'Suburb'* y '*CouncilArea'*. Se dejaron como el original las columnas que tienen mas de 120 registros idénticos para *'Suburb'*y 250 registros para '*CouncilArea'*. Las que tienen menos se unificaron bajo 'Other'.

- Se reliza codificación del tipo OneHotEncoder sobre las variables categóricas.

8. Luego se procedió a reducir la dimensionalidad del DataFrame, empleando Principal Component Analysis (PCA): Se incorporaron los 10 primeros componentes principales al DataFrame, los cuales acumulan mas del 60 % de la varianza.