TO DO:

- Análisis exploratorio: analizar en profundidad tanto variables numéricas como categóricas. Probad Databot de Positron que va muy bien para ver estas cosas.
    - Revisar calidad del dato, atípicos, etc.
    - Detectar variables que sean redundantes y eliminar aquellas que no aporten información.
    - Variables numéricas: analizar correlaciones
    - Variables categoricas:
        - Contrastes chi cuadrado de independencia entre pares de variables categóricas -> detectar variables categóricas redundantes
        - ANOVA/Kruskal Wallis + HSD Tukey + Boxplot de precio x variable categórica -> detectar variables categóricas relacionadas con la respuesta (precio).
    - Al finalizar el exploratorio podremos descartar aquellas variables redundantes y aquellas que no tienen relacion con la respuesta.
    - PCA: rehacer habiendo quitado las variables redundantes
- Modelos (semana que viene):
    - Hacer una partición de validación dentro de train (es la que utilizaremos para emular a la particion test y ver que tal funcionan los modelos entrenados en train)
        - Train_old = Train_new + Validation
    - Valorar si hacer transformaciones de la variable respuesta
    - *Lasso: hacer un lasso inicial para ver si hay variables cuyo coeficiente -> 0 (no significativas) y quizás descartarlas de algunos modelos (OLS, etc)
    - Benchmark comparativo de modelos: Entrenar los siguientes modelos en train, los que requieran optimizar hiperparámetros hacerlo mediante CV
        - Modelos a comparar:
            - LM forward, backward, stepwise
            - PCA + LM
            - Lasso + LM
            - Ridge + LM
            - Elasticnet + LM
            - Regresión local
            - GAM's (cubic splines, B-splines, P-splines, etc)
        - Metodologia:
            - Entrenar en train_new, optimizar hiperparametros mediante CV en train_new y evaluar en validation set
            - Escoger mejor modelo de cada categoría (mediante R^2_adj, AIC, etc) y comparar luego entre modelos mediante RMSE en validation set
            - Mejor modelo: aquel que minimiza RMSE en validation set (proxy de test set)
            - TEST (Único uso). Una vez seleccionado el mejor modelo hacer la predicción de test set y evaluar las métricas que salen.