La métrica que utilizaría sería MSE o RMSE dado que penalizan más los errores grandes (en comparación con MAE), y en este caso, los errores grandes son los que más nos interesan pues `time2event` está en días. Sin embargo, me parece que dicha métrica se relaciona más con modelos tradicionales, y respecto a la pregunta principal, se usaría un modelo de supervivencia dado que se desconoce el tiempo del evento para las propiedades que no se han vendido. Si utilizáramos un modelo de XGBoost, se tendría que descartar todos los datos donde `evento=0`.

El modelo de supervivencia modelaría no sólo si el evento ocurrirá, sino cuándo es más probable que ocurra. En este caso, el evento es la venta de la propiedad, y el tiempo que tarda en venderse es `time2event` lo cuál es un caso de censura derecha en análisis de supervivencia. (Censoring). En nuestros datos, tenemos 1534 observaciones de censura derecha.

Siendo honesto, no tengo experiencia con modelos de supervivenvia. Sin embargo, como se muestra en la Figura 1 no se ve un sesgo importante para los individuos *right-censored* representados con líneas azules. (La relación se mantiene con más datos en la muestra de la gráfica).

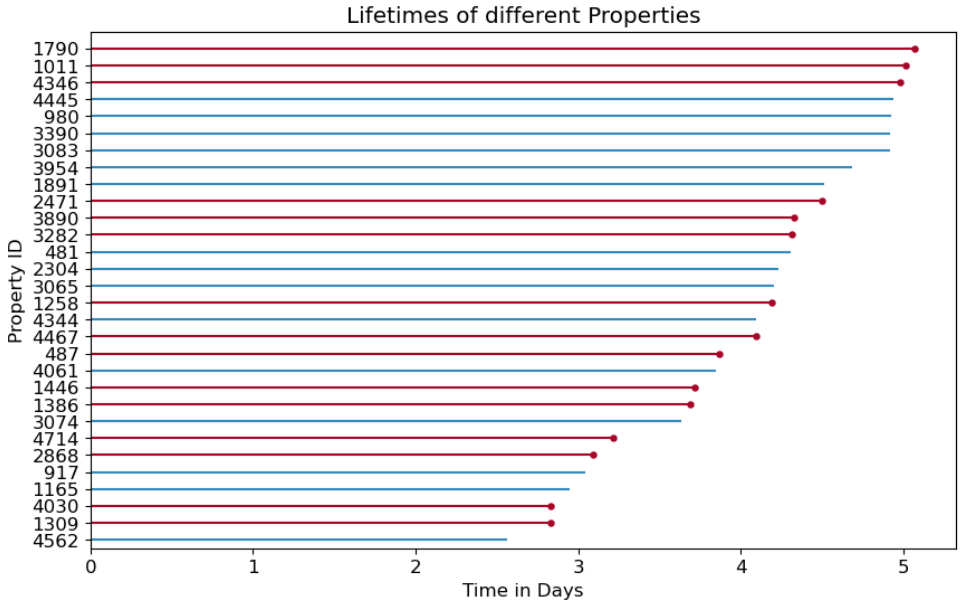

Figura 1: Lifelines

Si bien esto se tendría que analizar con más detalle, el análisis de supervivencia sería inevitable si los datos reflejaran gran diferencia en la distribución para las propiedades que no se han vendido. En este caso, dado la muestra en la gráfica, se puede entender que el evento que estudiamos no está realmente asociado con la variable que intentamos pronosticar.

Dado lo anterior, se podría argumentar que se puede proceder, a manera de proxy, con un modelo basado en árboles de decisón como ejemplo. Para dicho modelo, se obtuvo un `RMSE` de 0.093694.