



Università
Ca' Foscari
Venezia

Data and Web mining

Prof. Claudio Lucchese

Zillow's Home Value Prediction

Matteo Spanio
Matricola: 877485

Settembre 2022

1 The dataset

Di fronte al dataset fornito da Zillow la prima problematica da affrontare è scegliere come aggregare i dati: i file contenenti le transazioni con *logerror* e data infatti trattano più volte la vendita della stessa proprietà. Unire i dati sulla chiave *parcelid* creerebbe diverse righe duplicate tranne che per pochi valori, mentre includere questi duplicati permetterebbe di monitorare, a parità di condizioni della proprietà, cambiamenti di *logerror* nel tempo, un dato che si intuisce essere importante dal format del file .csv per la consegna delle predizioni.

2 Features

ciao

3 Model selection

ciao

4 About outliers

5 Conclusions

ciao