

PROJECT LUTHER PROPOSAL

PREDICTING PRICE OF HOUSE IN KING COUNTY

-Hiranya Krishna Kumar

DOMAIN

In Project Luther, I will predict house prices in King County using linear regression by web scraping data from Zillow and US Census Data. Zillow is a real estate listing website that has school-rating, which is typically a critical decision making criteria for home buyers in urban areas like King County. Prediction of house prices will be based on a certain set of features like zipcode, school-rating, year built, condition, home-type, sqft-living, sqft lot, year built ,no of days on zillow,views since listing,No of floors, structure type,market temperature ,walk score , transit score, parking type, monthly mortgage etc . Furthermore, using US-census data, median income data in the zip-codes of King county will be added to the feature set. Zillow has approximately 3100 listings in King County.

I will use selenium to scrape the data from zillow, as the website does not give all the html tags using beautifulsoup since the site is written in Javascript.

An MVP would predict the value of the house based on the features listed above using linear regression . Test all the features for homoscedasticity .If they are heteroskedastic then apply feature transformation like log, sqrt , linear combination to achieve residuals with constant variance across all values of the target variable.

DATA

Variable	Type	Description	Used for model
Price	int	Price of house	Target
Price per sqft	int		Alternate Target
Bedrooms	float	No of beds	Y
Bathrooms	float	No of baths	Y
Sqft living	int		Y
Sqft Lot	int		Y
High School Rating	float	Rating of schools	Y
Middle School Rating	float		Y
Elementary School Rating	float		Y

PROJECT LUTHER PROPOSAL

PREDICTING PRICE OF HOUSE IN KING COUNTY

-Hiranya Krishna Kumar

Walk Score	float		Y
Median Income	float	Median income in zip	Y
Zipcode	str		Y
Year Built	str/int		Y
Transit Score	int	Public transport connectivity	Y
Parking Type	str	On street, paid	N
Market Temp	str	Very hot to cold	Y
Views on zillow	int		N
No of Days on zillow	int		N
Condition	str	fixer/refurbished/new	Y
Home Type	str	Single family/ Condo/ townhome/ apartment	Y
City	str		Y
Lat and Long	float		N
Exterior features	str/binary	pool/no pool, patio/no patio Garden/No garden	Y
View type	str/binary	waterfront/non waterfront	Y