About this file

1. Executive Summary

Online property companies offer valuations of houses using machine learning techniques. The aim of this report is to predict the house sales in King County, Washington State, USA using Multiple Linear Regression (MLR). The dataset consisted of historic data of houses sold between May 2014 to May 2015.

1. Problem Statement

> "Predict the sales of houses in King County with an accuracy of at least 75-80% and understand which factors are responsible for higher property value - $650K and above."

---

1. Database Description

The dataset consists of house prices from King County an area in the US State of Washington, this data also covers Seattle. The dataset was obtained from Kaggle. *This data was published/released under CC0*: Public Domain. Unfortunately, the user has not indicated the source of the data. Please find the citation and database description in the Glossary and Bibliography.

♣ The dataset consisted of 21 variables and 21613 observations.

Due to the "Grade" variable which is based on a grading system specific to King County, it seems likely that the data originated from an official source. On the other hand, the existence of the "View" variable relates to the number of viewings the property received. This suggests that the data originated from a real estate agent/company.

In either case, there is no reason to call into question the general accuracy of the data because values such as prices, area etc. do not appear random. Nevertheless, some caution is appropriate before applying models trained on this data to more general cases.

Variables Description Data Type

id a notation for a house Numeric

date Date house was sold String

price Price is prediction target Numeric

bedrooms Number of Bedrooms/House Numeric

bathrooms Number of bathrooms/bedrooms Numeric

sqft*living square footage of the home Numeric sqft*lot square footage of the lot Numeric

floors Total floors (levels) in house Numeric

waterfront House which has a view to a waterfront Numeric

view Has been viewed Numeric

condition How good the condition is (Overall). 1 indicates worn out property and 5
excellent.(http://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r#g) Numeric

grade overall grade given to the housing unit, based on King County grading system. 1 poor ,13 excellent.
Numeric

sqft*above square footage of house apart from basement Numeric sqft*basement square footage of the basement
Numeric

yr*built Built Year Numeric yr*renovated Year when house was renovated Numeric

zipcode zip Numeric

lat Latitude coordinate Numeric

long Longitude coordinate Numeric

sqft*living15 Living room area in 2015(implies-- some renovations) This might or might not have affected the lotsize area
Numeric sqft*lot15 lotSize area in 2015(implies-- some renovations) Numeric

About this file

1. Executive Summary

Online property companies offer valuations of houses using machine learning techniques. The aim of this report is to predict the house sales in King County, Washington State, USA using Multiple Linear Regression (MLR). The dataset consisted of historic data of houses sold between May 2014 to May 2015.

∞ id	🗓 date	# price	# bedrooms	# bathrooms	# sqft_living	# sqft_lot	# floors	# waterfront	# view	# condition
a notation for a house	Date house was sold	Price is prediction target	Number of Bedrooms/House	Number of bathrooms/bedrooms	square footage of the home	square footage of the lot	Total floors (levels) in house	House which has a view to a waterfront	Has been viewed	How good the con (Overall). 1 indica worn out property excellent. (http://info.kingcc
7129300520	10/13/2014	221900	3	1	1180	5650	1	0	0	3
6414100192	12/9/2014	538000	3	2.25	2570	7242	2	0	0	3
5631500400	2/25/2015	180000	2	1	770	10000	1	0	0	3
2487200875	12/9/2014	604000	4	3	1960	5000	1	0	0	5
1954400510	2/18/2015	510000	3	2	1680	8080	1	0	0	3
7237550310	5/12/2014	1.23E+06	4	4.5	5420	101930	1	0	0	3
1321400060	6/27/2014	257500	3	2.25	1715	6819	2	0	0	3
2008000270	1/15/2015	291850	3	1.5	1060	9711	1	0	0	3
2414600126	4/15/2015	229500	3	1	1780	7470	1	0	0	3
3793500160	3/12/2015	323000	3	2.5	1890	6560	2	0	0	3