



UNIVERSIDAD AUTÓNOMA DE
NUEVO LEÓN



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Minería de Datos

Avance I Proyecto Integrador

Grupo 003, Viernes 18:00 a 21:00 hrs

Equipo 13

1588292 Laura Estefany Rodríguez de los Reyes

Maestra: Mayra Cristina Berrones Reyes

Título de la base de datos

a) Nombre con el que se encuentra en la página encontrada.

House price prediction

b) Url de la página.

https://www.kaggle.com/shree1992/housedata

Descripción de los datos

a) Qué tipo de datos son (tablas, imágenes, encuestas, etc.)

El tipo de datos son tablas, donde nos brindan las especificaciones de la vivienda.

data																	
date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	sqft_above	sqft_basement	yr_built	yr_renovated	street	city	statezip	country
2014-05-02 00:00:00	313000.0	3.0	1.5	1340	7912	1.5	0	0	3	1340	0	1955	2005	18810 Densmore Ave N	Shoreline	WA 98133	USA
2014-05-02 00:00:00	2384000.0	5.0	2.5	3650	9050	2.0	0	4	5	3370	280	1921	0	709 W Blaine St	Seattle	WA 98119	USA
2014-05-02 00:00:00	342000.0	3.0	2.0	1930	11947	1.0	0	0	4	1930	0	1966	0	26206-26214 143rd Ave SE	Kent	WA 98042	USA
2014-05-02 00:00:00	420000.0	3.0	2.25	2000	8030	1.0	0	0	4	1000	1000	1963	0	857 170th Pl NE	Bellevue	WA 98008	USA
2014-05-02 00:00:00	550000.0	4.0	2.5	1940	10500	1.0	0	0	4	1140	800	1976	1992	9105 170th Ave NE	Redmond	WA 98052	USA
2014-05-02 00:00:00	490000.0	2.0	1.0	880	6380	1.0	0	0	3	880	0	1938	1994	522 NE 88th St	Seattle	WA 98115	USA
2014-05-02 00:00:00	335000.0	2.0	2.0	1350	2560	1.0	0	0	3	1350	0	1976	0	2616 174th Ave NE	Redmond	WA 98052	USA
2014-05-02 00:00:00	482000.0	4.0	2.5	2710	35868	2.0	0	0	3	2710	0	1989	0	23762 SE 253rd Pl	Maple Valley	WA 98038	USA
2014-05-02 00:00:00	452500.0	3.0	2.5	2430	88426	1.0	0	0	4	1570	860	1985	0	46611-46625 SE 129th St	North Bend	WA 98045	USA
2014-05-02 00:00:00	640000.0	4.0	2.0	1520	6200	1.5	0	0	3	1520	0	1945	2010	6811 55th Ave NE	Seattle	WA 98115	USA
2014-05-02 00:00:00	463000.0	3.0	1.75	1710	7320	1.0	0	0	3	1710	0	1948	1994	Burke-Gilman Trail	Lake Forest Park	WA 98155	USA
2014-05-02 00:00:00	1400000.0	4.0	2.5	2920	4000	1.5	0	0	5	1910	1010	1909	1988	3838-4098 44th Ave NE	Seattle	WA 98105	USA
2014-05-02 00:00:00	588500.0	3.0	1.75	2330	14892	1.0	0	0	3	1970	360	1980	0	1833 220th Pl NE	Sammamish	WA 98074	USA
2014-05-02 00:00:00	365000.0	3.0	1.0	1090	6435	1.0	0	0	4	1090	0	1955	2009	2504 SW Portland Ct	Seattle	WA 98106	USA
2014-05-02 00:00:00	1200000.0	5.0	2.75	2910	9480	1.5	0	0	3	2910	0	1939	1969	3534 46th Ave NE	Seattle	WA 98105	USA
2014-05-02 00:00:00	242500.0	3.0	1.5	1200	9720	1.0	0	0	4	1200	0	1965	0	14034 SE 201st St	Kent	WA 98042	USA
2014-05-02 00:00:00	419000.0	3.0	1.5	1570	6700	1.0	0	0	4	1570	0	1956	0	15424 SE 9th St	Bellevue	WA 98007	USA
2014-05-02 00:00:00	367500.0	4.0	3.0	3110	7231	2.0	0	0	3	3110	0	1997	0	11224 SE 306th Pl	Auburn	WA 98092	USA
2014-05-02 00:00:00	257950.0	3.0	1.75	1370	5858	1.0	0	0	3	1370	0	1987	2000	1605 S 245th Pl	Des Moines	WA 98198	USA
2014-05-02 00:00:00	275000.0	3.0	1.5	1180	10277	1.0	0	0	3	1180	0	1983	2009	12425 415th Ave SE	North Bend	WA 98045	USA
2014-05-02 00:00:00	750000.0	3.0	1.75	2240	10578	2.0	0	0	5	1550	690	1923	0	3225 NE 92nd St	Seattle	WA 98115	USA
2014-05-02 00:00:00	435000.0	4.0	1.0	1450	8800	1.0	0	0	4	1450	0	1954	1979	3922 154th Ave SE	Bellevue	WA 98006	USA
2014-05-02 00:00:00	626000.0	3.0	2.25	1750	1572	2.5	0	0	3	1470	280	2005	0	3140 Franklin Ave E	Seattle	WA 98102	USA
2014-05-02 00:00:00	612500.0	4.0	2.5	2730	12261	2.0	0	0	3	2730	0	1991	0	10212 NE 156th Pl	Bothell	WA 98011	USA
2014-05-02 00:00:00	495000.0	4.0	1.75	1600	6380	1.0	0	0	3	1130	470	1959	1989	2021 NE 100th St	Seattle	WA 98125	USA
2014-05-02 00:00:00	285000.0	3.0	2.5	2090	10834	1.0	0	0	4	1360	730	1987	0	27736 23rd Avenue South	Federal Way	WA 98003	USA
2014-05-02 00:00:00	615000.0	3.0	1.75	2360	7291	1.0	0	0	4	1360	1000	1948	0	8436-8438 41st Ave SW	Seattle	WA 98136	USA
2014-05-02 00:00:00	698000.0	4.0	2.25	2200	11250	1.5	0	0	5	1300	900	1920	0	1036 4th St	Kirkland	WA 98033	USA
2014-05-02 00:00:00	675000.0	5.0	2.5	2820	67518	2.0	0	0	3	2820	0	1979	2014	23525 SE 32nd Way	Issaquah	WA 98029	USA
2014-05-02 00:00:00	790000.0	3.0	2.5	2600	4750	1.0	0	0	4	1700	900	1951	1999	3314 NW 75th St	Seattle	WA 98117	USA

b) Descripción de las columnas. (Que datos se describen en cada columna. Mediciones, tiempo, texto largo, etc.)

Las columnas contienen los siguientes datos de las especificaciones de la casa:

- Fecha en que se vendió la casa
- Precio de venta
- Número de recamaras
- Número de baños
- Pies cuadrados del espacio interior de la vivienda
- Pies cuadrados del espacio del lote
- Número de pisos de la casa

- Variable ficticia para determinar si la vivienda tenía vista al mar o no (0 si no tiene y 1 si sí tiene vista al mar)
- Un índice de 0 a 4 de qué tan buena era la vista de la propiedad
- Pies cuadrados que tiene la vivienda de construcción sobre el nivel del suelo
- Pies cuadrados que tiene la vivienda bajo el nivel del suelo, esto puede ser lo que corresponde al sótano.
- El año en que se construyó la casa inicialmente
- El año de la última renovación de la casa
- Dirección de la casa
- Ciudad donde se encuentra la casa
- Código postal
- País

Justificación del uso de datos

Lo que mas me llamó la atención de los datos eran los precios en los cuales se vendían las casas por lo cual se intentará apreciar cual o cuales son las características principales para fijar el precio en dicha casa. Una vez hecho los calculos del modelo podremos saber cuales son estos factores que tienen mayor impacto en el precio de una casa.

Planteamiento del problema

La problemática que se pretende resolver es en como una empresa constructora puede aprovechar las características que brindan mayor valor a una casa para así incluir en sus próximos desarrollos estas características en las casas y por lo tanto poder obtener mayor beneficio economico.

Objetivo final

El objetivo principal es saber cuales son los factores clave y que tienen mayor impacto en el precio de una casa.

Como objetivo secundario también me gustaría realizar un mapa donde podemos ver cuales son las zonas con precios elevados.

Planeación de la herramienta a utilizar

Para poder cumplir con el objetivo propuesto, se utilizarán las técnicas de regresión lineal para saber cuales son las variables que influyen más en el precio de la vivienda; para esto se usará Excel o R.

Adicionalmente se usará la técnica de visualización para desarrollar un mapa con las zonas con precios más elevados.