Postgraduate Diploma in Software Design with Artificial Intelligence

Assignment Part 1: Selecting a Data Set

Student ID: A00288399

Brief Description of the Data Set
Sourced from:
https://www.kaggle.com/justinas/housing-in-london?select=housing_in_london_yearly_variables.csv

This dataset is a combination of yearly and monthly data in regards to London City and the UK's housing prices, average salary and life satisfaction by area over a 24 year period (1996-2004).

Introduction

A look into housing in the U.K and London charting how areas, salaries, crime rates and more, have changed over the last 24 years. Using data collected from 44 areas monthly since 1996 and 50 areas yearly since 1999 there is a fantastic scope to chart changes in salaries, crime rates, housing prices and much more.

Data

Source is a combination of:
https://www.kaggle.com/justinas/housing-in-london?select=housing_in_london_yearly_variables.csv
https://www.kaggle.com/justinas/housing-in-london?select=housing_in_london_monthly_variables.csv

There are 6 total variables for monthly data (date, area, average price, area code, houses sold and no of crimes) and 14 total variables for the yearly data (all of monthly and median salary, life satisfaction, mean salary, recycling participation, population, number of jobs, size of area(hectares) and the number of houses).

- The date of the record is given
- Then the name of the area is given
- The average house price in the area
- The area code
- The amount of houses sold in the area that month
- The number of crimes that month
- The median salary in the area that year
- The average life satisfaction for a person living in the area that year
- The percentage of residents that recycle
- The population in the area
- The number of jobs in the area
- The size of the area in hectares
- And the number of houses in the area

1	date	area	average_price	code	houses_sold	no_of_crimes	median_salary	life_satisfaction	mean_salary	recycling_pct	population_size	number_of_jobs	area_size	no_of_houses
2755	2016-12-01	barnet		E09000003			32623	7.47	35315	37	386083	169000	8675	146730
2756	2016-12-01	bexley		E09000004				7.46	32666	53	244760	86000	6429	96864
2757	2016-12-01	brent		E09000005			30699	7.67	36579	36	328254	148000	4323	116649
2758	2016-12-01	bromley		E09000006			30861	7.6	35338	47	326889	134000	15013	137564
2759	2016-12-01	camden		E09000007			37817	7.56	45644	27	246181	394000	2179	102618
2760	2016-12-01	croydon		E09000008			31479	7.68	#	39	382304	149000	8650	154559
2761	2016-12-01	ealing		E09000009			29241	7.42	34468	51	343196	168000	5554	131249
2762	2016-12-01	enfield		E09000010			28918	7.29	33076	37	331395	129000	8220	124471
2763	2016-12-01	greenwich		E09000011			30180	7.28	33568	35	279766	96000	5044	108603
2764	2016-12-01	hackney		E09000012			32404	7.26	39387	27	273526	143000	1905	107574
2765	2016-12-01	hammersmith and fulham		E09000013			35000	7.47	40291	23	179654	157000	1715	85635
2766	2016-12-01	haringey		E09000014			29753	7.56	32799	36	278451	92000	2960	106879
2767	2016-12-01	harrow		E09000015			30134	7.64	34743	40	248752	93000	5046	89324
2768	2016-12-01	havering		E09000016			28906	7.53	33025	37	252783	100000	11446	101273
2769	2016-12-01	hillingdon		E09000017			33093	7.68	39729	43	302471	204000	11570	108171
2770	2016-12-01	hounslow		E09000018			36235	7.96	47577	30	271139	197000	5659	99267
2771	2016-12-01	islington		E09000019			39894	7.38	52885	32	232865	251000	1486	101783

fig1.1 yearly data example

1	date	area	average_price	code	houses_sold	no_of_crimes
12697	2016-11-01	outer london	417765	E13000002	5042	
12698	2016-11-01	north east	124910	E12000001	3366	
12699	2016-11-01	north west	149723	E12000002	10014	
12700	2016-11-01	yorks and the hu	151219	E12000003	7111	
12701	2016-11-01	east midlands	173872	E12000004	7172	
12702	2016-11-01	west midlands	179248	E12000005	7393	
12703	2016-11-01	east of england	273382	E12000006	9108	
12704	2016-11-01	london	470854	E12000007	8178	
12705	2016-11-01	south east	307867	E12000008	12669	
12706	2016-11-01	south west	237150	E12000009	9086	
12707	2016-11-01	england	231053	E92000001	74097	
12708	2016-12-01	city of london	811786	E09000001	13	
12709	2016-12-01	barking and dag	285622	E09000002	154	1349
12710	2016-12-01	barnet	522223	E09000003	383	2077
12711	2016-12-01	bexley	331025	E09000004	261	1240
12712	2016-12-01	brent	490418	E09000005	175	2362
12713	2016-12-01	bromley	429187	E09000006	458	1899
12714	2016-12-01	camden	832126	E09000007	209	2720
12715	2016-12-01	croydon	363571	E09000008	367	2518
12716	2016-12-01	ealing	485390	E09000009	272	2353
12717	2016-12-01	enfield	394465	E09000010	211	1835

fig1.2 monthly data example

With data from 44 areas over 24 years we have over 13000 data points for data taken monthly, and with data from 50 areas over 21 years with over 1000 data points for yearly data.

Reason for Selection

I think this is a very interesting data set that will show the growth of London over the past 24 years. I chose this dataset because I think it could provide very interesting insights into different aspects of not only London, but other cities with factors such as wealth distribution and crime rates affecting the data. Using this data we can provide suggestions as to where a local government might have to focus its resources in order to improve life for its area's inhabitants.

Proposed Analysis

Using this dataset I plan to:

- Examine how different areas grew based on population, salaries and crime rates.
- Analyse how salaries have grown and whether higher salaries equates to higher house prices in the area and whether crime rates are lower given higher salaries.
- Study the population distribution based on mean income and provide insights as to whether lower income areas would have higher population density or if it would be the opposite.
- Determine if their is any correlation between crime rates, population density and mean salaries
- Observe how crime rates affect house prices and house sales

and more.