

Progetto Modulo 2 - Statistica descrittiva

Analisi del mercato immobiliare in Texas

Punto 2: Tipo di variabili

Variabili:

- city: qualitativa nominale
- year: quantitativa
- month: qualitativa (numerica codificata)
- sales: quantitativa
- volume: quantitativa
- median_price: quantitativa
- listings: quantitativa
- months_inventory: quantitativa

Punto 3: Nella seguente tabella sono riportati gli indici di posizione (media, mediana, minimo e massimo, 25° e 75° percentile), variabilità (range, range interquartile (IQR), deviazione standard e il coefficiente di variazione) e forma (curtosi e skewness) per le variabili quantitative: sales, volume, median_price, listings e months_inventory.

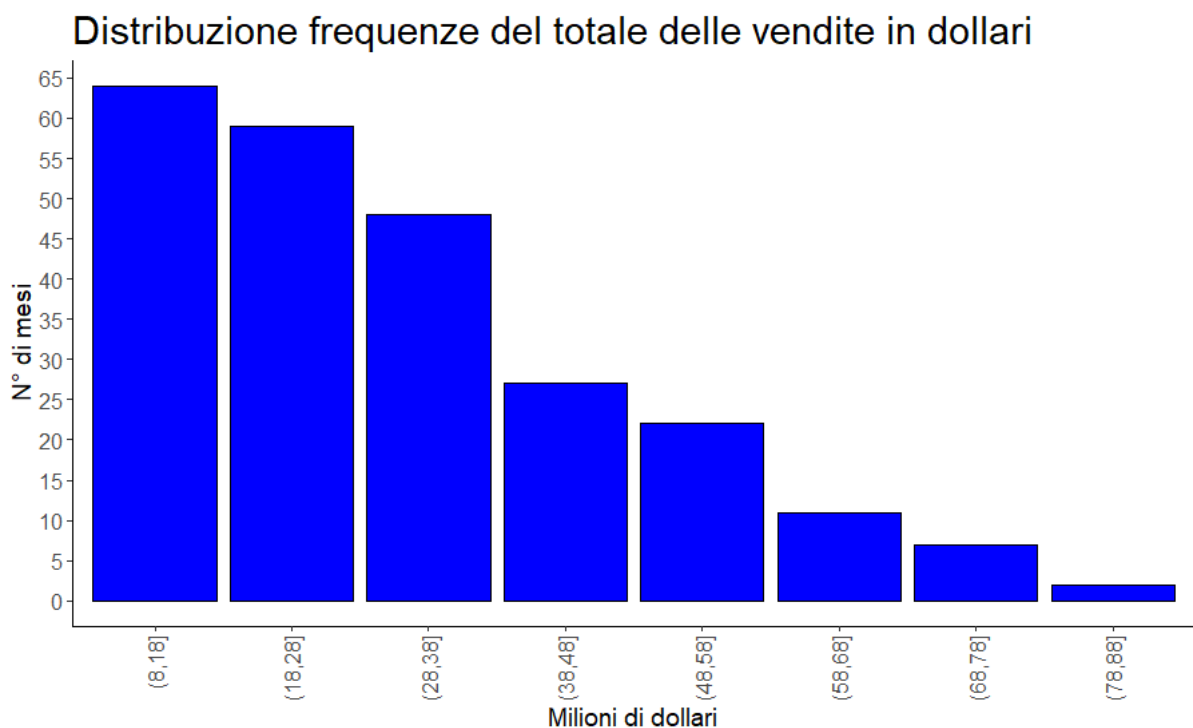
Variabile	media	mediana	min	max	perc_25	perc_75	range	IQR	dev_std	cv	curtosi	skewness
Sales	192.29	175.50	79.00	423.00	127.00	247.00	344.00	120.00	79.65	41.42	-0.31	0.72
Volume		27.06	8.17	83.55	17.66	40.89	75.38	23.23	16.65	53.71	0.18	0.88
Median_price	132,665.42	134,500.00	73,800.00	180,000.00	117,300.00	150,050.00	106,200.00	32,750.00	22,662.15	17.08	-0.62	-0.36
Listings	1,738.02	1,618.50	743.00	3,296.00	1,026.50	2,056.00	2,553.00	1,029.50	752.71	43.31	-0.79	0.65
Months_inventory	9.19	8.95	3.40	14.90	7.80	10.95	11.50	3.15	2.30	25.06	-0.17	0.04

Osservando il dataset si può notare come ogni osservazione sia descritta dalle variabili qualitative city, month e year; per ognuna delle 4 città presenti sono state fatte un'osservazione al mese per 5 anni dal 2010 al 2014 per un totale di 240 osservazioni. Per questo motivo la frequenza assoluta e relativa delle modalità di ciascuna di queste variabili è la stessa.

Punto 4: La variabile volume è quella con variabilità più elevata, lo si può notare osservando il coefficiente di variabilità espresso in percentuale che è calcolato come rapporto tra la deviazione standard e la media; tale indice consente di confrontare tra loro variabili con range e ordini di grandezza molto diversi. La variabile volume è anche la variabile più asimmetrica con un valore di skewness pari a 0.88.

Punto 5: La variabile volume [milioni di \$] è stata divisa in 8 classi di ampiezza 10 milioni di \$, successivamente è stata calcolata la distribuzione di frequenze, riportata nella seguente tabella:

	ni	fi	Ni	Fi
(8,18]	64	0.27	64	0.27
(18,28]	59	0.25	123	0.51
(28,38]	48	0.20	171	0.71
(38,48]	27	0.11	198	0.83
(48,58]	22	0.09	220	0.92
(58,68]	11	0.05	231	0.96
(68,78]	7	0.03	238	0.99
(78,88]	2	0.01	240	1.00



L'istogramma mostra come la distribuzione sia asimmetrica con frequenze maggiori per le classi più con valori più bassi confermando il dato della skewness positiva discusso nel punto 4. L'indice di Gini è pari a 0.92

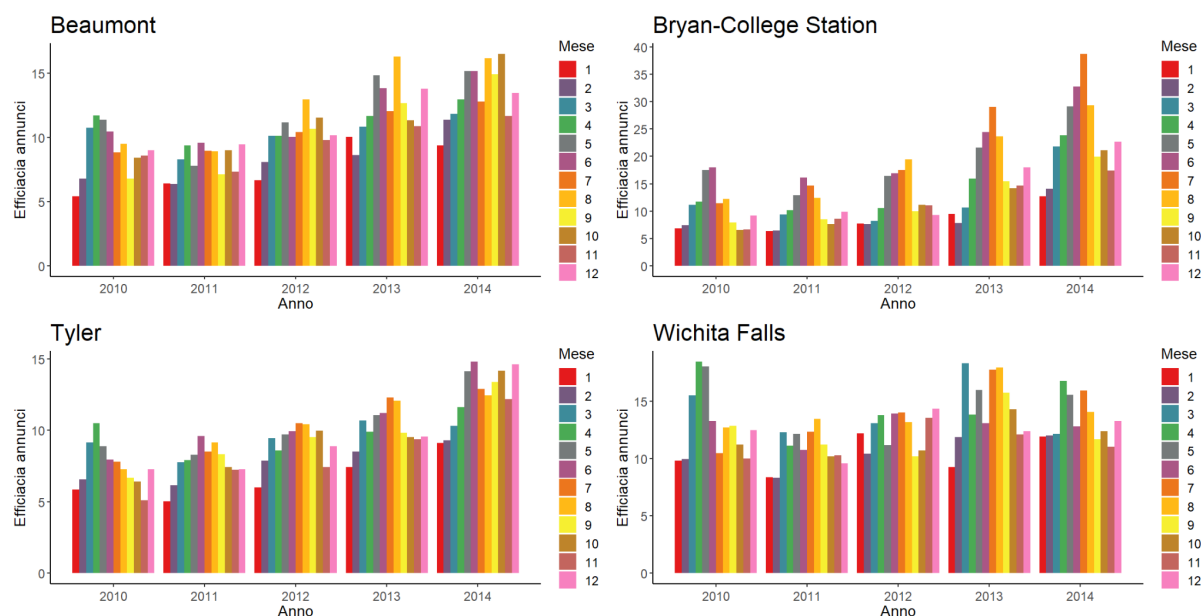
Punto 6: Per la variabile city l'indice di Gini sarà pari a 1 visto che le 4 modalità delle variabili si distribuiscono in modo perfettamente eterogeneo.

Punto 7: Probabilità

- La probabilità che riporti la città di Beaumont è pari al 25% (casi favorevoli/casi totali = $60/240$)
- La probabilità che riporti il mese di luglio è pari al 8,33% (casi favorevoli/casi totali = $20/240$)
- La probabilità che riporti dicembre 2012 è pari al 1,67% (casi favorevoli/casi totali = $4/240$)

Punto 8: La colonna con il prezzo medio è stata creata dividendo il volume totale di vendite (variabile volume) per il numero di vendite (variabile sales), è stata salvata nel dataframe con il nome mean_price.

Punto 9: La colonna per esprimere l'efficacia degli annunci è stata creata dividendo il numero delle vendite (variabile sales) per il numero degli annunci attivi (listings), moltiplicato per 100. Sarà un numero compreso tra 0 e 100, tanto più sarà vicino a 100 tanto più saranno efficaci gli annunci per quel determinato mese.



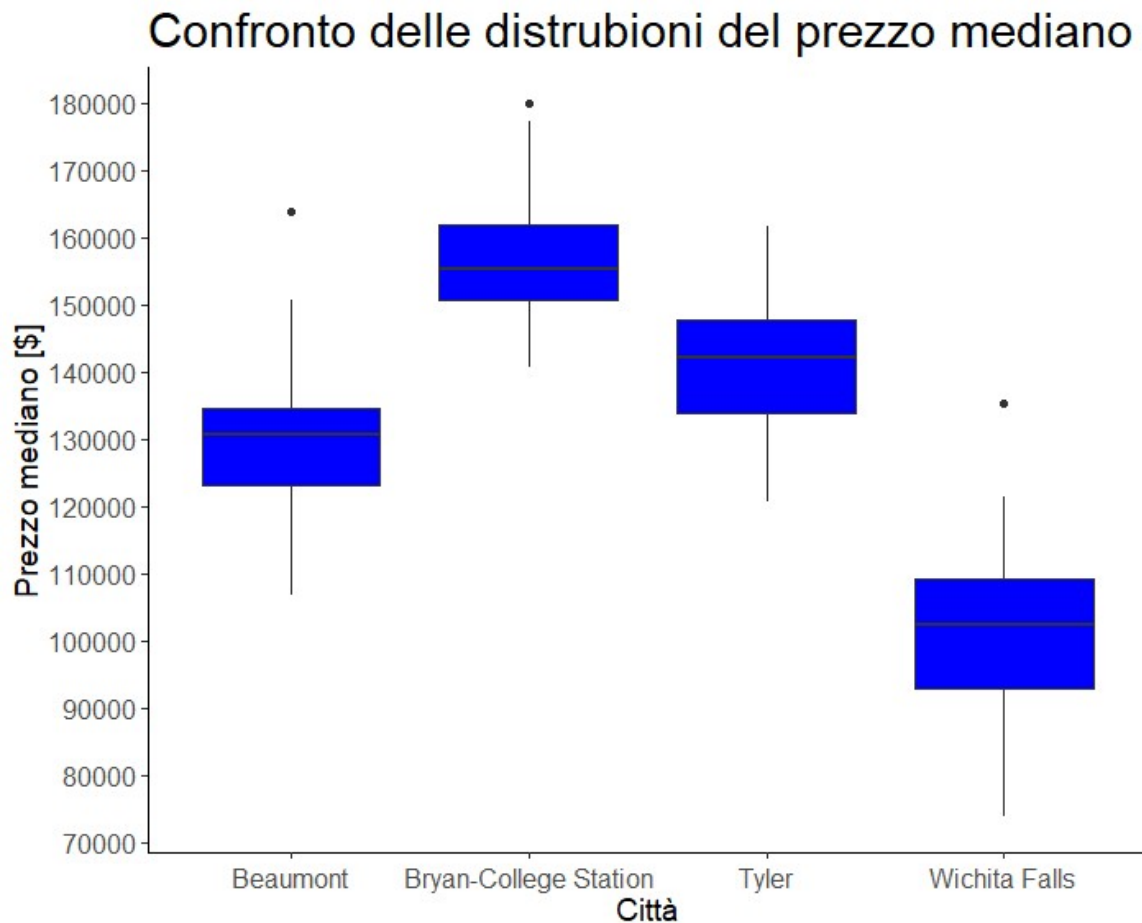
Il grafico mostra tale indici per ciascun mese dal 2010 al 2014 nelle diverse città, si può vedere come, ad esempio, la città di Bryan-College Station sia quella con l'indice più alto nel 2014, con il valore massimo raggiunto nel mese di luglio.

Punto 10: La seguente tabella riporta media e deviazione standard delle variabili mean_price, sales e listings condizionatamente alla, città, all'anno e al mese.

	media_median_price	std_mean_price	media_sales	std_sales	media_listings	std_listings
<u>City</u>						
Beaumont	129988.33	10104.99	177.38	41.48	1679.32	91.13
Bryan-College Station	157488.33	8852.24	205.97	84.98	1458.13	252.53
Tyler	141441.67	9336.54	269.75	61.96	2905.05	226.75
Wichita Falls	101743.33	11320.03	116.07	22.15	909.58	73.76
<u>Year</u>						
2010	130,191.67	21,821.76	168.67	60.54	1,826.00	785.02
2011	127,854.17	21,317.80	164.13	63.87	1,849.65	780.38
2012	130,077.08	21,431.52	186.15	70.91	1,776.81	738.45
2013	135,722.92	21,708.08	211.92	84.00	1,677.60	743.52
2014	139,481.25	25,625.41	230.60	95.51	1,560.04	706.71
<u>Month</u>						
1	124,250.00	25,151.28	127.40	43.38	1,647.05	704.61
2	130,075.00	22,822.59	140.85	51.07	1,692.50	711.20
3	127,415.00	23,442.03	189.45	59.18	1,756.70	727.35
4	131,490.00	21,458.40	211.70	65.40	1,825.70	770.43
5	134,485.00	18,796.26	238.85	83.12	1,823.85	790.22
6	137,620.00	19,231.02	243.55	95.00	1,833.25	811.63
7	134,750.00	21,944.78	235.75	96.27	1,821.20	826.72
8	136,675.00	22,488.38	231.45	79.23	1,786.30	815.87
9	134,040.00	24,344.10	182.35	72.52	1,748.90	802.66
10	133,480.00	26,358.07	179.90	74.95	1,710.35	779.16
11	134,305.00	24,691.47	156.85	55.47	1,652.70	741.25
12	133,400.00	22,809.76	169.40	60.75	1,557.75	692.57

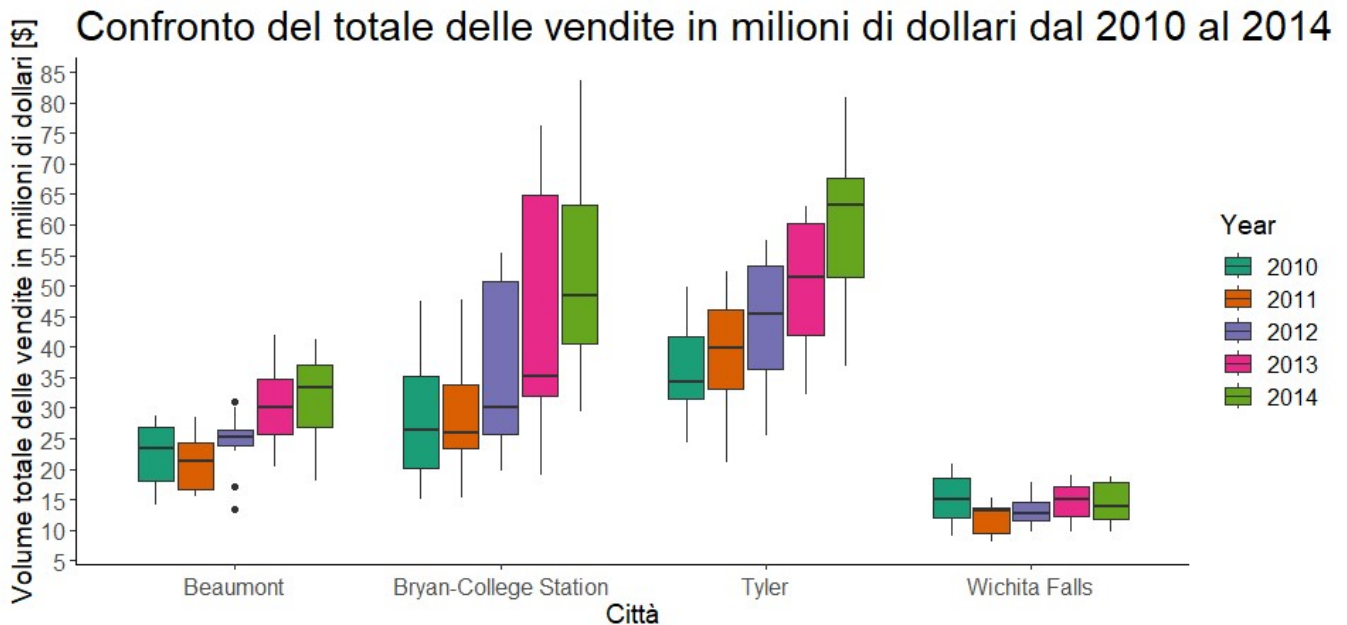
GRAFICI

Punto 1: Il seguente grafico mostra la distribuzione del prezzo mediano nelle diverse città del dataset, si può notare come la città di Bryan-College Station sia quella con il prezzo mediano più elevato, superiore di oltre 50000\$ rispetto al prezzo mediano delle case di Wichita Falls, la località con il prezzo minore.



Punto 2: Il grafico seguente mostra la distribuzione del valore totale delle vendite in milioni di dollari nelle diverse città, suddiviso per anni, dal 2010 al 2014.

Si può notare come in generale, ad eccezione del 2011, per le città di Beaumont, Bryan-College Station e Tyler il volume delle vendite sia sempre aumentato rispetto all'anno precedente, con la città di Tyler che è quella con il volume totale delle vendite maggiore. Per quanto riguarda, invece, la città di Wichita Falls il volume di vendite è pressoché costante nei 5 anni analizzati.



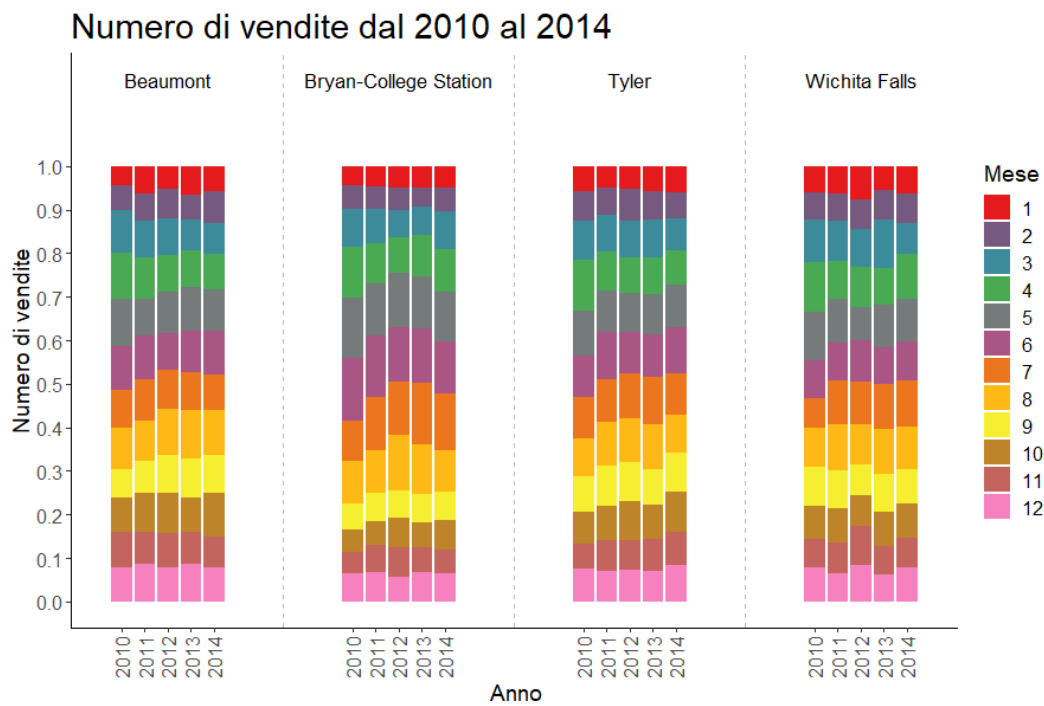
Punto 3:

I seguenti grafici riportano il numero di vendite suddiviso per mesi dal 2010 al 2014 per le diverse città, il primo con valori assoluti e il secondo con valori normalizzati.

Dal primo grafico è possibile notare come il trend del numero delle vendite sia molto simile a quello del volume di dollari generato, con un generale aumento rispetto all'anno precedente, ad eccezione del 2011, per le città di Beaumont, Bryan-College Station. La città di Tayler registra un aumento del numero di vendite per tutti gli anni presi in esame mentre quella di Wichita Falls risulta avere un andamento piuttosto costante.

Da entrambi i grafici si può vedere come i mesi iniziali e finali dell'anno siano i mesi con il minore numero di vendite, tuttavia la disparità rispetto ai mesi centrali dell'anno non sembra essere così marcata.

The figure displays four stacked bar charts, one for each location: Beaumont, Bryan-College Station, Tyler, and Wichita Falls. The y-axis represents the 'Numero di vendite' (Number of sales) ranging from 0 to 4350. The x-axis represents the 'Anno' (Year) from 2010 to 2014. Each bar is composed of 12 segments representing the months of the year, color-coded according to the legend on the right, labeled 'Mese' (Month). The colors range from red (1) to pink (12). The charts show a general upward trend in sales for all locations over the five-year period, with Tyler consistently having the highest sales volume and Wichita Falls the lowest.



Punto 4:

Il seguente grafico riporta l'andamento del numero di annunci attivi dal 2010 al 2014, per crearlo è stata aggiunta una colonna al data set iniziale riportante la data per esteso in modo da poter ordinare le osservazioni per ciascuna città in modo univoco, utilizzando solo una colonna. Le linee verticali mostrano la divisione del periodo temporale in anni.

Si nota come il numero di annunci sia diminuito dal 2010 al 2014 nelle città di Tyler e Bryan-College Station mentre risulta essere pressoché costante nelle città di Beaumont e Wichita Falls.

Si può notare anche come l'andamento nei 5 anni di osservazione sia tendenzialmente periodico per ciascun anno, con un picco del numero di annunci attivi nei mesi centrali dell'anno e due minimi in corrispondenza dei mesi iniziali e finali. Per la città di Bryan-College Station il picco è anticipato tendenzialmente nei mesi di Marzo e Aprile.

Numero di annunci dal 2010 al 2014

