

Analisi del Mercato Immobiliare del Texas

Descrizione del progetto

L'azienda **Texas Realty Insights** desidera analizzare le tendenze del mercato immobiliare nello stato del Texas, sfruttando i dati storici relativi alle vendite di immobili. L'obiettivo è fornire insight statistici e visivi che supportino le decisioni strategiche di vendita e ottimizzazione delle inserzioni immobiliari.

Obiettivi del progetto

- Identificare e interpretare i trend storici delle vendite immobiliari in Texas.
- Valutare l'efficacia delle strategie di marketing delle inserzioni immobiliari.
- Offrire una rappresentazione grafica dei dati che evidenzia la distribuzione dei prezzi e delle vendite tra città, mesi e anni.

Valore Aggiunto

L'analisi statistica proposta permetterà a **Texas Realty Insights** di ottimizzare le loro strategie di mercato, identificando città con opportunità di crescita e valutando l'efficacia delle inserzioni immobiliari nel tempo. Grazie a una visione chiara e strutturata dei dati, l'azienda potrà prendere decisioni basate su informazioni concrete, migliorando la gestione delle vendite immobiliari in Texas.

Dataset: "Real Estate Texas.csv"

Il dataset contiene le seguenti variabili:

- **city**: città di riferimento
- **year**: anno di riferimento
- **month**: mese di riferimento
- **sales**: numero totale di vendite
- **volume**: valore totale delle vendite (in milioni di dollari)
- **median_price**: prezzo mediano di vendita (in dollari)
- **listings**: numero totale di annunci attivi
- **months_inventory**: quantità di tempo necessaria per vendere tutte le inserzioni correnti, espresso in mesi

[Scarica il dataset](#)

Steps del progetto

1. Analisi delle variabili

Identifica e descrivi il tipo di variabili statistiche presenti nel dataset. Valuta come gestire le variabili che sottintendono una dimensione tempo e commenta sul tipo di analisi che può essere condotta su ciascuna variabile.

2. Indici di posizione, variabilità e forma

Calcola Indici di posizione, variabilità e forma per tutte le variabili per le quali ha senso farlo, per le altre crea una distribuzione di frequenza. Infine, commenta tutto brevemente.

3. Identificazione delle variabili con maggiore variabilità e asimmetria

Determina: - Qual è la variabile con la più alta variabilità - Qual è la variabile con la distribuzione più asimmetrica Spiega come sei giunto a queste conclusioni e fornisci considerazioni statistiche.

4. Creazione di classi per una variabile quantitativa

Seleziona una variabile quantitativa (es. sales o median_price) e suddividila in classi. Crea una distribuzione di frequenze e rappresenta i dati con un grafico a barre. Calcola l'**indice di eterogeneità Gini** e discuti i risultati.

5. Calcolo della probabilità

Qual è la probabilità che, presa una riga a caso di questo dataset, essa riporti la città "Beaumont"? E la probabilità che riporti il mese di Luglio? E la probabilità che riporti il mese di dicembre 2012?

6. Creazione di nuove variabili

- Crea una nuova colonna che calcoli il **prezzo medio** degli immobili utilizzando le variabili disponibili.
- Prova a creare una colonna che misuri l'**efficacia degli annunci di vendita**. Commenta e discuti i risultati.

7. Analisi condizionata

Usa il pacchetto dplyr o il linguaggio base di R per effettuare analisi statistiche condizionate per città, anno e mese. Genera dei **summary** (media, deviazione standard) e rappresenta graficamente i risultati.

8. Creazione di visualizzazioni con ggplot2

Utilizza ggplot2 per creare grafici personalizzati. Assicurati di esplorare: - Boxplot per confrontare la distribuzione del prezzo mediano tra le città. - Grafici a barre per confrontare il totale delle vendite per mese e città. - Line charts per confrontare l'andamento delle vendite in periodi storici differenti.

9. Conclusioni

Fornisci una sintesi dei risultati ottenuti, facendo riferimento alle principali tendenze emerse e fornendo raccomandazioni basate sull'analisi. Questo non è un progetto di programmazione, ma di statistica, e ci si aspetta di leggere commenti e considerazioni statistiche per i vari passaggi e risultati.

Osservazioni operative

Alcune considerazioni: 1. Utilizza i boxplot per confrontare la distribuzione del prezzo mediano delle case tra le varie città. Commenta il risultato. 2. Utilizza i boxplot o qualche variante per confrontare la distribuzione del valore totale delle vendite tra le varie città ma anche tra i vari anni. Qualche considerazione da fare? 3. Usa un grafico a barre sovrapposte per confrontare il totale delle vendite nei vari mesi, sempre considerando le città. Prova a commentare ciò che viene fuori. Già che ci sei prova anche il grafico a barre normalizzato. Consiglio: Stai attento alla differenza tra `geombar()` e `geomcol()`. PRO LEVEL: cerca un modo intelligente per inserire ANCHE la variabile Year allo stesso blocco di codice, senza però creare accrocchi nel grafico. 4. Prova a creare un line chart di una variabile a tua scelta per fare confronti commentati fra città e periodi storici. Ti avviso che probabilmente all'inizio ti verranno fuori linee storte e poco chiare, ma non demordere.

Consigli: Prova inserendo una variabile per volta. Prova a usare variabili esterne al dataset, tipo vettori creati da te appositamente.

Modalità di consegna:

Link a rpubs