

Case Study

Introduzione al caso di studio

- Da che cosa dipendono i prezzi delle case?
- **Hedonic price modelling**: i prezzi delle case dipendono dalle loro caratteristiche (dimensione, numero di stanze, tipo di abitazione . . .)
- ma è proprio così? possiamo pensare che ci siano altri fattori che danno valore ad una casa, e quindi al suo prezzo.

Introduzione al caso di studio

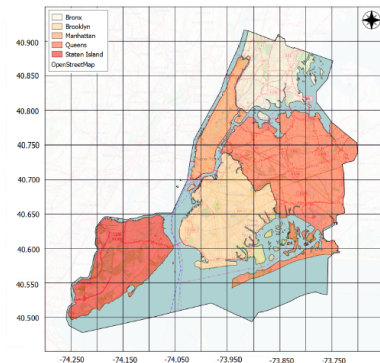
L'approccio 'edonico' è stato utilizzato anche per spiegare le differenze di prezzi degli hotel ... e più di recente anche per studiare il crescente fenomeno di **Airbnb** in varie città del mondo

- Airbnb: uno dei primi esempi di **sharing economy**
- Opportunità di business o minaccia?
- Competizione 'sleale' nei confronti degli hotel?
- Aumento dei prezzi degli affitti?
- Che cosa influenza i prezzi delle case su Airbnb?

Introduzione al caso di studio

Supponiamo di essere un hotel in una delle città più importanti del mondo
New York

Abbiamo il sospetto che il business di Airbnb ci abbia rubato mercato e vogliamo capire quali sono i fattori che influenzano i prezzi delle case su Airbnb. . .



Introduzione al caso di studio

Per capire la situazione vogliamo tenere conto di tre aspetti:

- diffusione di Airbnb
- presenza di distretti molto eterogenei, con caratteristiche differenti
- possibilità di reperire **Open Data**

Open Data: perchè?

Vogliamo infatti tenere conto anche di informazioni 'esterne'

Gli Open Data sono

- **accessibili** agli utenti in formato aperto e corredati da adeguati metadati
- **disponibili** gratuitamente o a costi marginali per la loro divulgazione e riproduzione, senza limiti di copyright o altre licenze che ne limitino la diffusione
- **integrabili** con altre fonti al fine di creare nuove risorse, applicazioni o servizi di pubblica utilità → **interoperabilità**
- **aggiornati** periodicamente al fine di garantire una continuità nelle analisi
- **machine readable**, ovvero processabili in automatico

Open Data a New York

Tramite il sito <https://opendata.cityofnewyork.us/> è possibile reperire informazioni, ad esempio, su:

<i>variabili</i>	<i>fonte</i>
attrazioni principali	Dept Finance
hotel	NYC open data
ristoranti	NYC open data
metropolitana	Metro trans authority
tempo libero	NYC open data
servizi sanitari	NYC open data
criminalità	NY Police Dept

... e molte altre ancora.

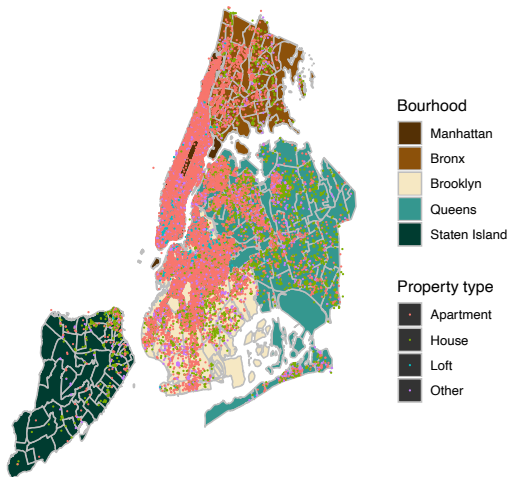
Airbnb a New York

- sul sito `insideairbnb.com` sono contenuti 'listings' di case su Airbnb per molte città nel mondo.
Per ogni listing sono disponibili molte variabili relative
 - alle caratteristiche delle **proprietà**,
 - alle caratteristiche del **proprietario** (host),
 - alle **reviews** degli ospiti,
 - ai termini di **servizio**.
- naturalmente è presente anche la variabile '**prezzo per notte**'
- siamo interessati a capire quali di queste variabili abbiano maggior impatto nella definizione del prezzo ...

- ma non ci basta! vogliamo tenere conto anche di **variabili Open**
- criminalità, vicinanza di attrazioni turistiche, presenza della metropolitana ... hanno un ruolo nella definizione dei prezzi? e quale?
- interessante problema di **Data Integration**

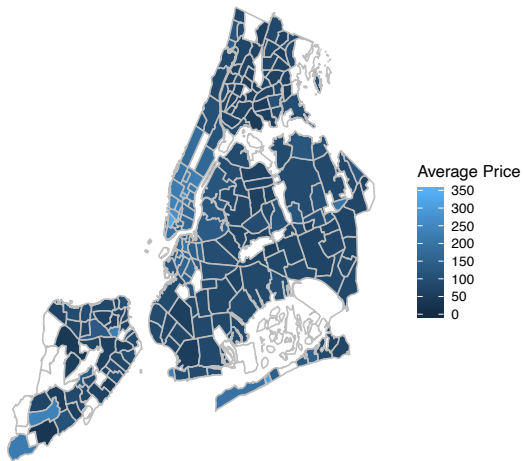
Airbnb a New York

Posizione delle proprietà



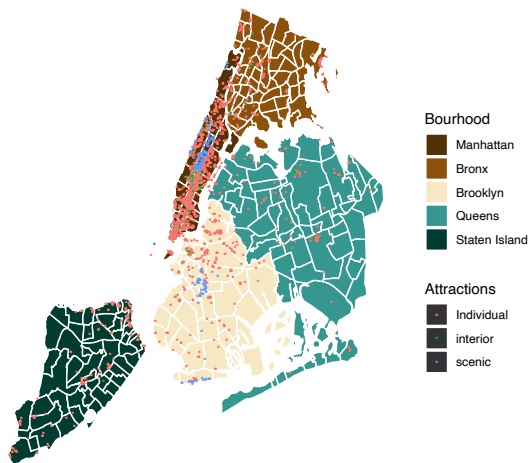
Airbnb a New York

Prezzo medio per quartiere



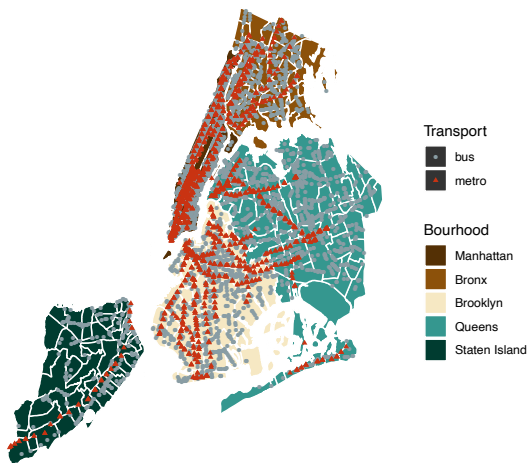
Open Data a New York

Attrazioni principali per quartiere



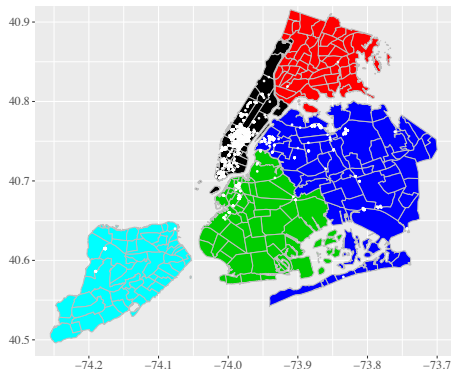
Open Data a New York

Trasporti pubblici



Open Data a New York

Posizione degli hotel



Prevedere i prezzi su Airbnb

- visualizzare i dati ci ha permesso di formulare qualche ipotesi ...
- ora è il momento di fare previsioni ...
- possiamo usare i modelli che conosciamo ... modello lineare, foresta causale, rete neurale, mars, bagging, boosting ...
- ma anche usare un nuovo strumento ... Gradient Boosting

Prezzi su Airbnb: Gradient Boosting



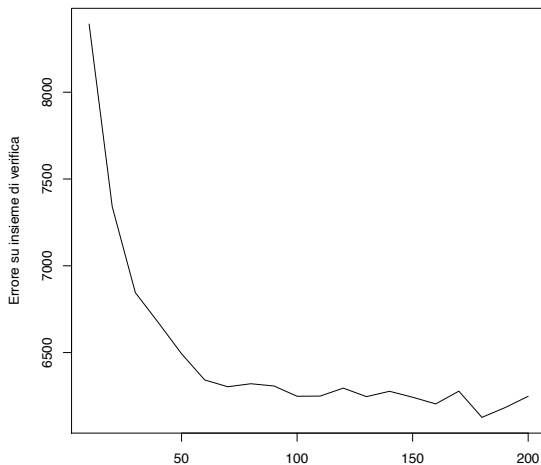
Prezzi su Airbnb: Gradient Boosting

Studiamo i prezzi delle case su Airbnb a New York con un GB

- grande dataset: 77000 osservazioni, 107 variabili
- variabile risposta: prezzo per notte di una casa su Airbnb
- insieme di stima: 50000
- consideriamo come ipotesi iniziale un modello con iterazioni= 100, profondità alberi= 1 (stump), shrinkage= 0.1
- successivamente proviamo altre configurazioni, modificando i parametri di regolazione

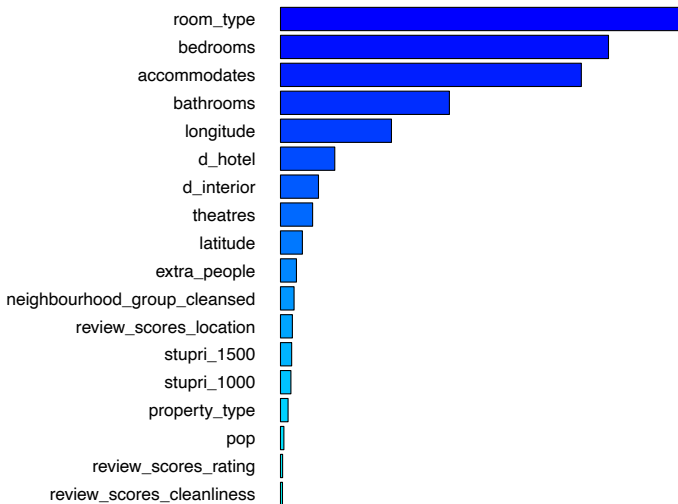
Prezzi su Airbnb: Gradient Boosting

Performance del modello al crescere delle iterazioni



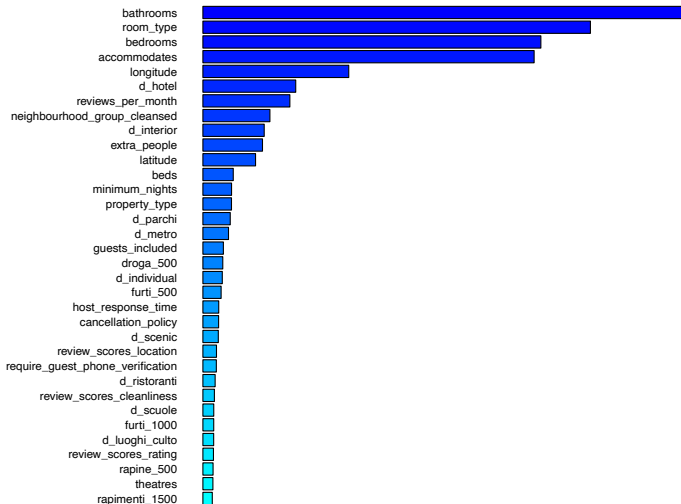
Prezzi su Airbnb: Gradient Boosting

iterazioni= 100, profondità alberi= 1, shrinkage= 0.1



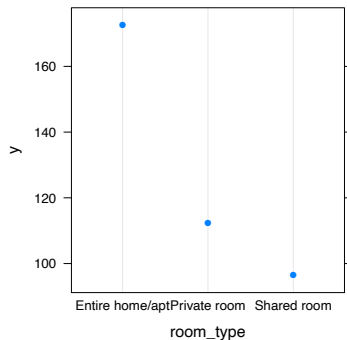
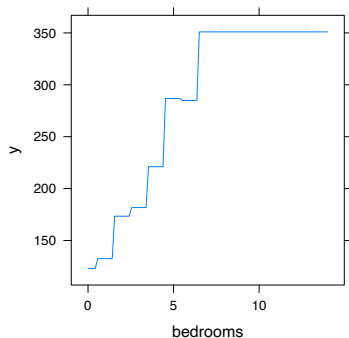
Prezzi su Airbnb: Gradient Boosting

iterazioni= 180, profondità alberi= 4, shrinkage= 0.2



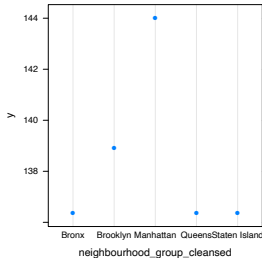
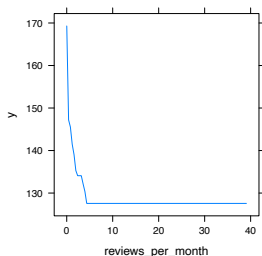
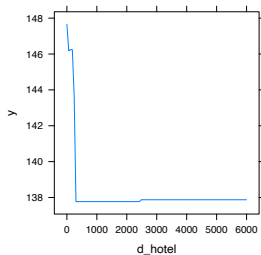
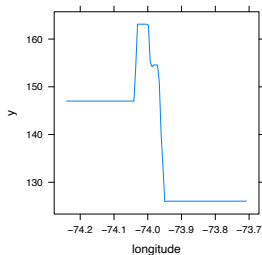
Prezzi su Airbnb: Gradient Boosting

Partial Dependence Plots



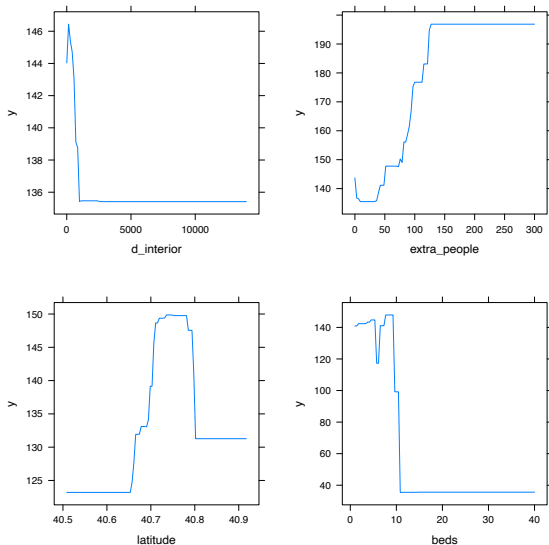
Prezzi su Airbnb: Gradient Boosting

Partial Dependence Plots



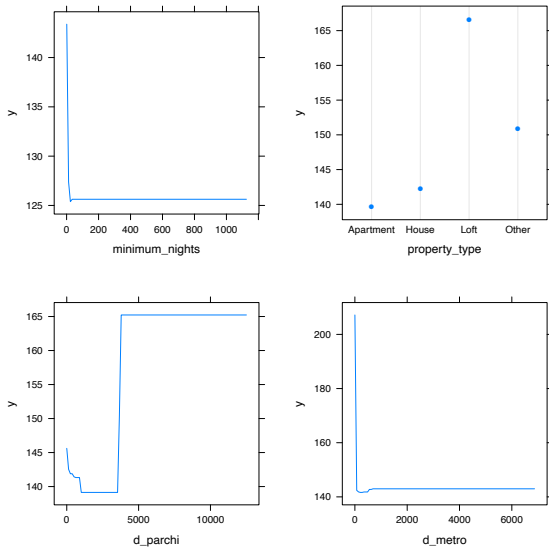
Prezzi su Airbnb: Gradient Boosting

Partial Dependence Plots



Prezzi su Airbnb: Gradient Boosting

Partial Dependence Plots



Prezzi su Airbnb: Gradient Boosting

Quali conclusioni dovrebbe trarre il manager dell'hotel?