***CASO DI STUDIO DEL CORSO IN***

***INGEGNERIA DELLA CONOSCENZA***

***-----------------------------------------------------------------------------------------------------------------------***

*SmartInvest NYC:*

*Strategie Smart per Investimenti Immobiliari a New York: Apprendimento, Analisi e Opportunità nei Prezzi Convenienti.*

***-----------------------------------------------------------------------------------------------------------------------***

Immagine che contiene edificio, Palazzo, Metropoli, Area metropolitana

Descrizione generata automaticamente

***----------------- -----------------***

*Autori: Docente:*

*Mauro Carlucci Prof. Nicola Fanizzi*

*Matricola:*

*Paola Campaniello*

*Matricola: 758277*

**INTRODUZIONE**

Ciao a tutti!

Siamo qui per presentarvi un progetto che avrà il potenziale di trasformare il nostro approccio agli investimenti immobiliari a New York.

Il nostro progetto si concentra sull'analisi approfondita dei prezzi degli affitti delle case e delle stanze di Airbnb nella regione, con l'obiettivo di identificare le opportunità più promettenti per un investimento redditizio. Siamo consapevoli che la scelta di una casa non riguarda solo l'acquisto di una proprietà, ma anche la creazione di una strategia per massimizzare il rendimento dell'investimento nel minor tempo possibile.

L'approccio innovativo che stiamo adottando è basato su dati accurati e algoritmi avanzati che analizzano non solo il costo iniziale di acquisto, ma anche la previsione degli affitti futuri e il rendimento potenziale delle stanze su piattaforme di condivisione come Airbnb.

Questo ci consentirà di fornire ai nostri investitori un'analisi dettagliata e personalizzata, guidandoli verso la scelta della proprietà che massimizzerà i loro rendimenti nel breve e lungo termine.

Siamo super emozionati di darvi l'accesso a questa nuova era degli investimenti immobiliari, dove la tecnologia e l'esperienza si incontrano per risultati concreti.

Vi invitiamo ad esplorare più a fondo il nostro progetto e a porre domande in modo che possiamo condividere ulteriori dettagli sulla nostra metodologia e visione.

Grazie per essere qui, e per essere incuriositi dal mondo degli investimenti immobiliari a New York.

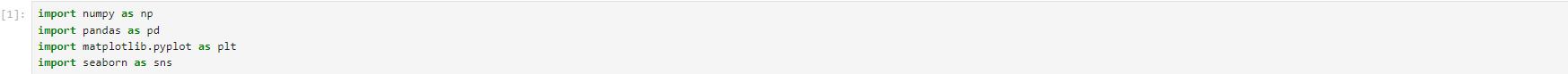
Immagine che contiene clipart, simbolo, Elementi grafici, silhouette

Descrizione generata automaticamente

**GitHub Repository:** <https://github.com/maurocarlu/InvestimentoImmobiliare_ICON>

1. **OBIETTIVO**

La fase introduttiva consiste nel leggere il dataset e comprenderne la struttura, dove la nostra analisi si focalizza principalmente sulla modellazione di alcune variabili indipendenti disponibili. Tali variabili rivestono un ruolo essenziale per gli imprenditori operanti nel settore immobiliare, poiché forniscono orientamento nella loro decisione imprenditoriale, basandosi sui prezzi proposti da diversi fornitori di servizi. Questo approccio mira a indirizzare l'imprenditore verso una strategia funzionale capace di garantire un significativo ritorno economico.

****

1. **LETTURA E COMPRENSIONE DEL DATASET**

Successivamente alla fase introduttiva, per comprendere al meglio la struttura del dataset, implementiamo la libreria pandas.

****

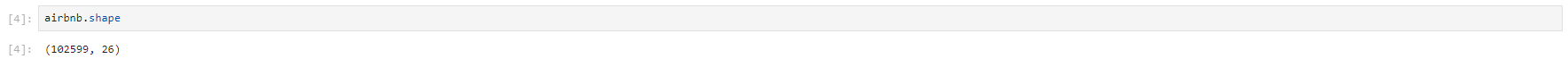
Immagine che contiene testo, ricevuta, algebra

Descrizione generata automaticamente

Immagine che contiene testo, ricevuta, algebra

Descrizione generata automaticamenteImmagine che contiene testo, ricevuta, algebra

Descrizione generata automaticamente



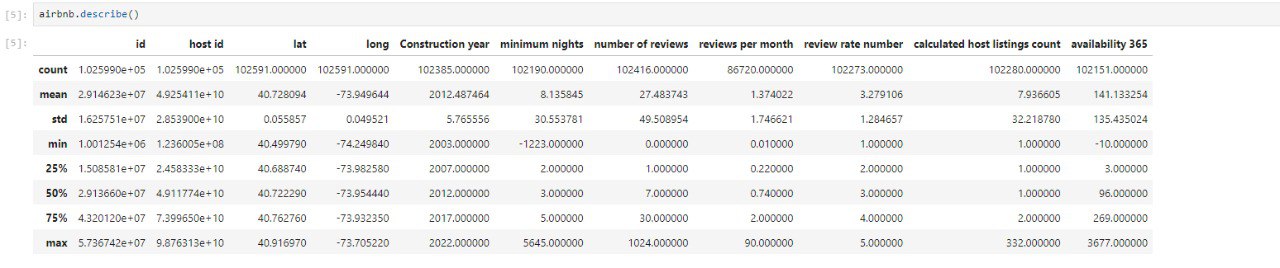


Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamenteImmagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Immagine che contiene testo, schermata

Descrizione generata automaticamente

1. **PRE-PROCESSING DEL DATASET**

Dopo aver concluso la fase inziale e la fase successiva, relativa all’analisi strutturale del dataset, passiamo alla fase del pre-processing, nella quale andremo ad effettuare un cleaning dei dati.  
  
Inizio cancellando le colonne che so a prescindere non essere rilevanti per la mia analisi:

Immagine che contiene testo, Carattere, numero, schermata

Descrizione generata automaticamente

Immagine che contiene testo, Carattere, numero, schermata

Descrizione generata automaticamenteImmagine che contiene testo, schermata, numero, Carattere

Descrizione generata automaticamente

Le colonne eliminate sono le seguenti:

*‘country’ e ‘country\_code’* in quanto il dataset fa riferimento a tutti annunci sul territorio americano e dunque questi dati non servono per il nostro caso.  
  
 *‘NAME’* In quanto essa rappresenta la descrizione dell’annuncio della casa/stanza in affitto e dunque cioè non serve per la nostra analisi.  
  
*‘license’* in quanto tale colonna è sempre vuota

*‘host name’* in quanto essa rappresenta il nome dell’host es: Mauro, Paola, Michele e dunque tale dato non è necessario per la nostra analisi

*‘house\_rules’* in quanto tale campo rappresenta le regole della casa es: Non fumare, non portare animali, no feste. Dunque questo campo non ci è utile.

Sono state rimosse tutte le righe duplicate



Nelle celle delle colonne *‘price’* e *‘service fee’* è stato rimosso il simbolo *‘ $ ‘,* sono stati rimossi gli spazi e sono state convertite in colonne numeriche.

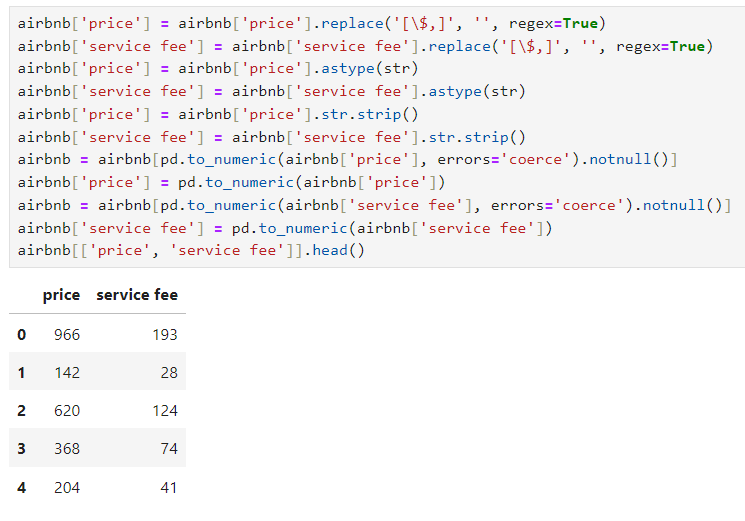


Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Converto la colonna *room type* da tipo object a tipo stringa:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Converto la colonna *last review* da object a tipo data:



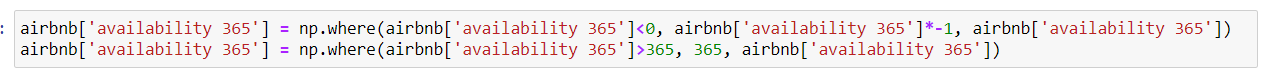
* La metrica *‘availability 365’*indica il numero massimo di giorni in cui una stanza di un B&B può essere affittata. Durante l'analisi dei dati nel nostro dataset, abbiamo osservato la presenza di alcuni valori negativi o superiori a 365 associati a questa metrica. I valori superiori sono stati regolati a 365, rappresentando il limite massimo, mentre quelli negativi sono stati convertiti in valori positivi.

Immagine che contiene testo, Carattere, numero, software

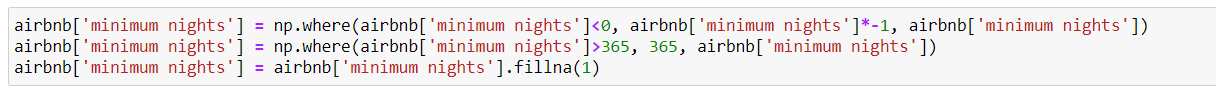
Descrizione generata automaticamente

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente



* La metrica *‘minimum nights’* indica il numero massimo di giorni in cui una stanza di un B&B può essere affittata. Durante l'analisi dei dati nel nostro dataset, abbiamo osservato la presenza di alcuni valori negativi, valori nulli, o superiori a 365 associati a questa metrica. I valori superiori sono stati regolati a 365, rappresentando il limite massimo di giorni durante l’anno, quelli negativi sono stati convertiti in valori positivi, mentre i valori nulli, sono stati posti tutti a 1, dato che si parte da un valore minimo di 1

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Controllo dei valori unici di neighbourhood group

* In questo caso è stato modificato il nome di *Brooklyn*, dato che abbiamo rilevato un errore di battitura cambiando il nome in *“brookln”*

Immagine che contiene testo, Carattere, linea, schermata

Descrizione generata automaticamente

Controllo quanti valori nulli ci sono per ogni colonna

Immagine che contiene testo, schermata, numero, Carattere

Descrizione generata automaticamente

Mi occupo della gestione delle celle null per ogni colonna :

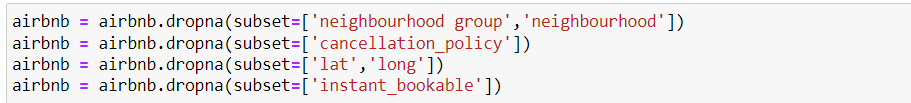
avendo le seguenti voci:

* *host\_identity\_verified*, ove all’interno delle colonne troveremo i valori: unconfirmed e confirmed, se il valore non è presente, avremo per certo *“unconfirmed”;*
* *calculated host listings count*, tale voce rappresenta quanti immobili appartengono ad un determinato host, di conseguenza, sostituiamo i valori null col “*count”* di quante volte si presenta l’host id.

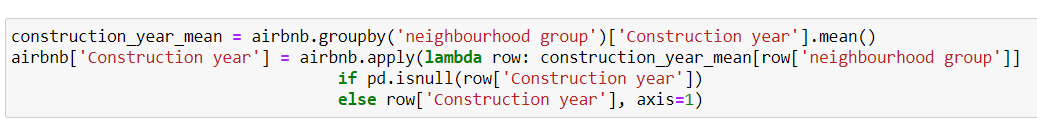




* *neighbourhood group* e *neighbourhood*, entrambi i dati sono stati eliminati, poiché ci potrebbero essere molte occorrenze null, e non abbiamo modi alternativi per effettuare una modifica, in quanto si causerebbero delle problematiche per la media per la posizione della casa;
* *cancellation\_policy,* è stato eliminato in quanto un raggiungibile attraverso qualsiasi modifica, dunque la modifica, avrebbe portato all’inserimento di valori non veritieri che avrebbero causato problematiche;
* *lat e long,* i valori rappresentano latitudine e longitudine, dati che anche in questo caso, attraverso la media potremmo avere dei risultati non corretti, la miglior soluzione è eliminarli per evitare sbalzi nei risulatati
* *instant\_bookable,* anche in questo caso il dato è stato eliminato poiché non abbiamo una certa accuratezza attraverso la media che andremo a ricavare per gli eventuali valori nulli.



Sostituisco i valori null di *“Construction year”* con la media del quartiere corrispondente



Divido le colonne in categoriche e numeriche





Immagine che contiene testo, Carattere, schermata

Descrizione generata automaticamente

Immagine che contiene testo, diagramma, schermata, Parallelo

Descrizione generata automaticamente

Immagine che contiene schizzo, diagramma

Descrizione generata automaticamente

Immagine che contiene schizzo

Descrizione generata automaticamente

Immagine che contiene testo, schizzo, diagramma

Descrizione generata automaticamente

* il seguente grafico rappresenta la correlazione ottima tra *“reviews per month”* e *“number of reviews”*



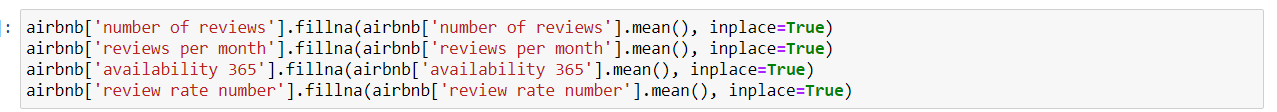
Immagine che contiene testo, schermata

Descrizione generata automaticamente

* a seguito della visione del grafici precedenti, abbiamo potuto osservare, che per i dati :

*number of reviews, reviews per month, availability 365, review rate number,*

è più conveniente effettuare la media, anziché la mediana;



L'unico elemento che rimane da verificare è 'last review'



Immagine che contiene testo, Carattere, schermata, numero

Descrizione generata automaticamente

Poiché ci sono delle date errate, prima pongo le date posteriori al giorno attuale, poi sostituisco le celle vuote con la media

Immagine che contiene testo, Carattere, schermata, linea

Descrizione generata automaticamente

Immagine che contiene testo, schermata, numero, software

Descrizione generata automaticamente