

A stylized, colorful illustration of a city skyline, featuring various skyscrapers in shades of orange, yellow, and blue. The illustration is framed by a circular border with splatter effects. The background is white with blue and orange splatters.

Superbonus 110%

Federico Reggiani

Homework Big Data & Analytics

Scraping & Data Quality

03/07/2020

Come Funziona



1.
Cos'è

2.
A chi è rivolto

3.
Chi può
usufruirne

4.
In che modo

❑ Il **Decreto Rilancio** (D.L. 19 maggio 2020 n.34) è il maxi provvedimento che contiene misure urgenti in materia di salute, sostegno al lavoro e all'economia, nonché di politiche sociali connesse all'emergenza epidemiologica da COVID-19

❑ Composto da oltre 260 articoli , tra cui il Bonus in edilizia al 110% (articolo 119 del DL Rilancio) per interventi di riqualificazione energetica (**Ecobonus**) e antisismica (**Sisma Bonus**) su prime e seconde case, installazione di **impianti fotovoltaici**, con spese sostenute dal 1 luglio 2020 al 31 dicembre 2021.

- Cos'è
- Quali provvedimenti
- Chi può usufruirne
- In che modo

❑ Ecobonus 110%

Interventi sulle unità abitative (prima casa, seconda casa, condomini) per il miglioramento di almeno due classi energetiche degli edifici:

- Interventi di **isolamento termico delle superfici**
Tetto massimo: € 60.000 moltiplicato per ogni unità immobiliare
- **Sostituzione degli impianti di riscaldamento esistenti**
Tetto massimo: € 30.000 moltiplicato per ogni unità immobiliare

❑ Sismabonus 110%

- Interventi per **l'adozione di misure antisismiche** su edifici ubicati in zone a rischio sismico
Tetto massimo: € 96.000 moltiplicato per ogni unità immobiliare

❑ Bonus Fotovoltaico 110%

- Interventi per **l'installazione di impianti solari fotovoltaici** connessi alla rete elettrica. La stessa detrazione spetta anche per l'installazione contestuale o successiva di sistemi di accumulo integrati negli impianti solari fotovoltaici. La detrazione vale solo nel caso di fruizione di uno dei superbonus previsti per l'efficientamento energetico e riduzione del rischio sismico.
Tetto massimo: € 48.000

- Cos'è
- Quali provvedimenti
- Chi può usufruirne
- In che modo

- ❑ Singoli proprietari (prima e seconda casa)
- ❑ Condomini (per ogni unità immobiliare)
- ❑ Imprese edili e/o immobiliari per immobili da destinare alla vendita o in affitto

- Cos'è
- Quali provvedimenti
- Chi può usufruirne
- In che modo

- ❑ Possibilità di usufruire della **detrazione del 110%** della spesa sostenuta in un periodo di **5 anni**
- ❑ L'interessato può **cedere il credito d'imposta** direttamente ad altri soggetti, quali ad esempio istituto di credito o altri soggetti finanziari
- ❑ **Sconto in fattura** da parte del fornitore che ha effettuato gli interventi e poi da questi recuperato sotto forma di credito d'imposta con possibilità anche successiva di cessione del credito a terzi senza la necessità di anticipare i soldi per i lavori

- Cos'è
- Quali provvedimenti
- Chi può usufruirne
- In che modo

Architettura BI

L'**obiettivo** è quello
di fornire dati e
analisi necessari a
chi può trarre
beneficio da questo
progetto

Sorgenti

Siti specializzati in annunci

- Immobiliare
- Idealista
- Cercacasa
- Wikicasa

Per individuare quali immobili sono disponibili in vendita e che possano essere interessati dagli interventi di superbonus

- **Listini immobiliari camera commercio**
- **Catasto**
- **Uffici tecnici comunali**

Per ottenere una fotografia completa della situazione e dello stato degli immobili esistenti su un dato territorio



Extract Transform Loading

Data Ingestion

- Scraping Selenium

Per acquisire i dati di vendita di immobili in tempo reale è necessario recuperarli direttamente dal sito web

Data Processing

- OpenRefine

Una volta recuperati i dati dai siti web bisogna fare data exploration e procedere alla pulizia

Data Integration

- Talend

Bisogna integrare tutti i dati dalle diverse fonti e caricarli nei database



Database



Database documentale poichè gli annunci delle case sono sostanzialmente tutti diversi tra loro e contengono una parte di testo libero (descrizione)



Database a grafo per analisi delle relazioni tra immobili.

Si possono individuare caratteristiche simili tra immobili venduti in una certa località.

Oppure, dopo statistiche sulle case più vendute, si può fare una ricerca e mostrare le case più «vicine» a quella tipologia e capire dove intervenire.



Per dati catastali e dati degli uffici tecnici comunali che probabilmente hanno archiviato su database relazionali



Database chiave – valore per archiviare immagini e planimetrie di immobili oltre che a ulteriore documentazione ufficiale (certificazioni,...)



Analisi dati & Data Visualization



Sviluppare algoritmi e analisi statistiche per capire quali tipologie di immobili sono più richieste con quali tipologie di lavori da effettuare e/o in quale area concentrarsi



Poter visualizzare graficamente su mappa i dati raccolti, puliti e analizzati nelle aree precedenti





Caso di studio

Scraping

Da Immobiliare.it filtrando per gli annunci in provincia di Bergamo

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1550 entries, 0 to 1549
```

```
Data columns (total 40 columns):
```

#	Column	Non-Null Count	Dtype
0	id_annuncio	1550 non-null	object
1	url	1550 non-null	object
2	titolo	1550 non-null	object
3	IMMOBILE GARANTITO	838 non-null	object
4	RIFERIMENTO E DATA ANNUNCIO	1550 non-null	object
5	CONTRATTO	1550 non-null	object
6	TIPOLOGIA	1521 non-null	object
7	SUPERFICIE	1519 non-null	object
8	LOCALI	1498 non-null	object
9	PIANO	1439 non-null	object
10	TOTALE PIANI EDIFICIO	1438 non-null	object
11	DISPONIBILITÀ	1213 non-null	object
12	TIPO PROPRIETÀ	1526 non-null	object
13	ALTRE CARATTERISTICHE	1484 non-null	object
14	PREZZO	1544 non-null	object
15	SPESE CONDOMINIO	721 non-null	object
16	INFORMAZIONI CATASTALI	907 non-null	object
17	ANNO DI COSTRUZIONE	1179 non-null	object
18	STATO	1510 non-null	object

19	RISCALDAMENTO	1454 non-null	object
20	CLASSE ENERGETICA	1404 non-null	object
21	posizione	1518 non-null	object
22	descrizione	1518 non-null	object
23	immobiliare	1473 non-null	object
24	POSTI AUTO	1184 non-null	object
25	CLIMATIZZAZIONE	740 non-null	object
26	UNITÀ	25 non-null	object
27	DATA DI INIZIO LAVORI E DI CONSEGNA PREVISTA	26 non-null	object
28	INDICE PREST. ENERGETICA RINNOVABILE	56 non-null	object
29	PRESTAZIONE ENERGETICA DEL FABBRICATO	221 non-null	object
30	CERTIFICAZIONE ENERGETICA	145 non-null	object
31	TIPO VENDITA	2 non-null	object
32	DATA VENDITA	2 non-null	object
33	OFFERTA MINIMA	4 non-null	object
34	SPESA PRENOTA DEBITO	4 non-null	object
35	CONTRIBUTO NON DOVUTO	4 non-null	object
36	NUMERO IMMOBILI	5 non-null	object
37	AGGIORNATO IL	5 non-null	object
38	PROCEDURA	1 non-null	object
39	RIALZO MINIMO	1 non-null	object

dtypes: object(40)
memory usage: 484.5+ KB



Scraping

Criticità emerse

❑ Numero di pagina non selezionabile dall'url:

- Click
- Aggiornamento pagine

❑ Struttura pagina diverse

❑ Gestione errori



Caratteristiche Dataset

- ❑ Potenzialmente una grande mole di dati: solo per la provincia di Bergamo e su un solo sito web di riferimento sono presenti più di 22k annunci e ogni annuncio contiene foto, planimetrie, certificazioni...
- ❑ Dato attendibile e generalmente corretto
- ❑ Grande varietà del dato: testo della descrizione, foto dell'immobile, planimetrie, tipologia immobile con caratteristiche e informazioni molto diverse
- ❑ Lenta variazione dei dati, limitati nuovi annunci al giorno e persistenza per lunghi periodi.



Data profiling & processing

Le modalità di trattamento adottate

- ❑ Nessun annuncio duplicato, titoli doppi perché stessi immobili ma unità diverse
- ❑ Valori «null» per la colonna Tipologia, perché riferiti ad immobili con più tipologie al suo interno ed eliminati recuperando informazioni dal Titolo
- ❑ Puliti dati colonne Contratto, Superficie, Data Annuncio
- ❑ Valori «null» per prezzi recuperati da Titolo e valorizzati con «Asta» e «Su richiesta»
- ❑ Cancellato varie colonne poiché poco utili o pochissime righe valorizzate



Conclusioni

Effetti
miglioramento
patrimonio
edilizio

Monitorare
evoluzione e
interpretazione
normativa per
implementare il
modello

Creazione
opportunità di
business per gli
operatori del
settore edilizio e
finanziario



Grazie per l'attenzione

Federico Reggiani

<https://github.com/FedericoReggiani/Superbonus-110-.git>

