

Auto **Scout24** **visualisation**

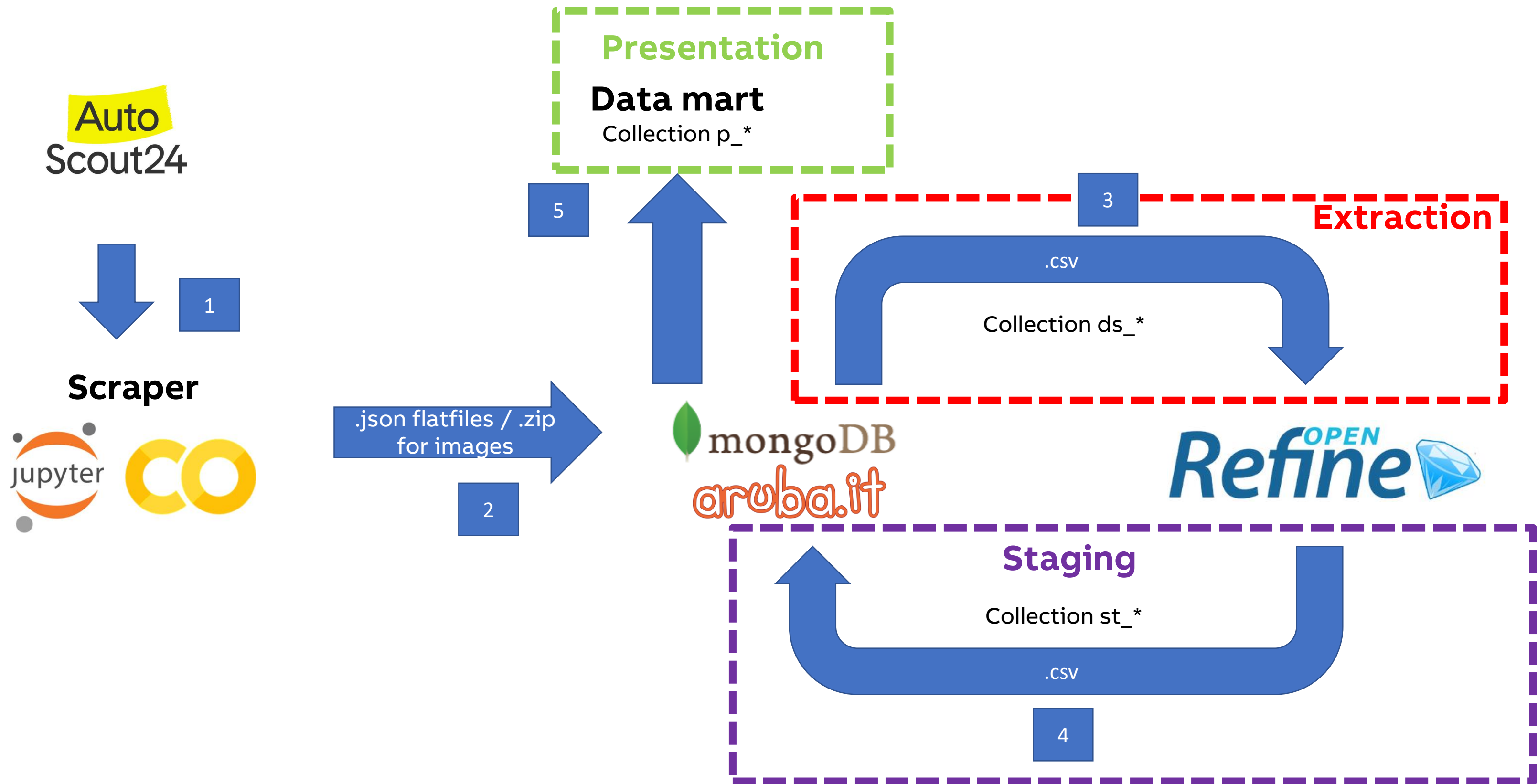
Simone Zambetti

Background economico

9/14/2020

Overall architecture

(1mo homework)



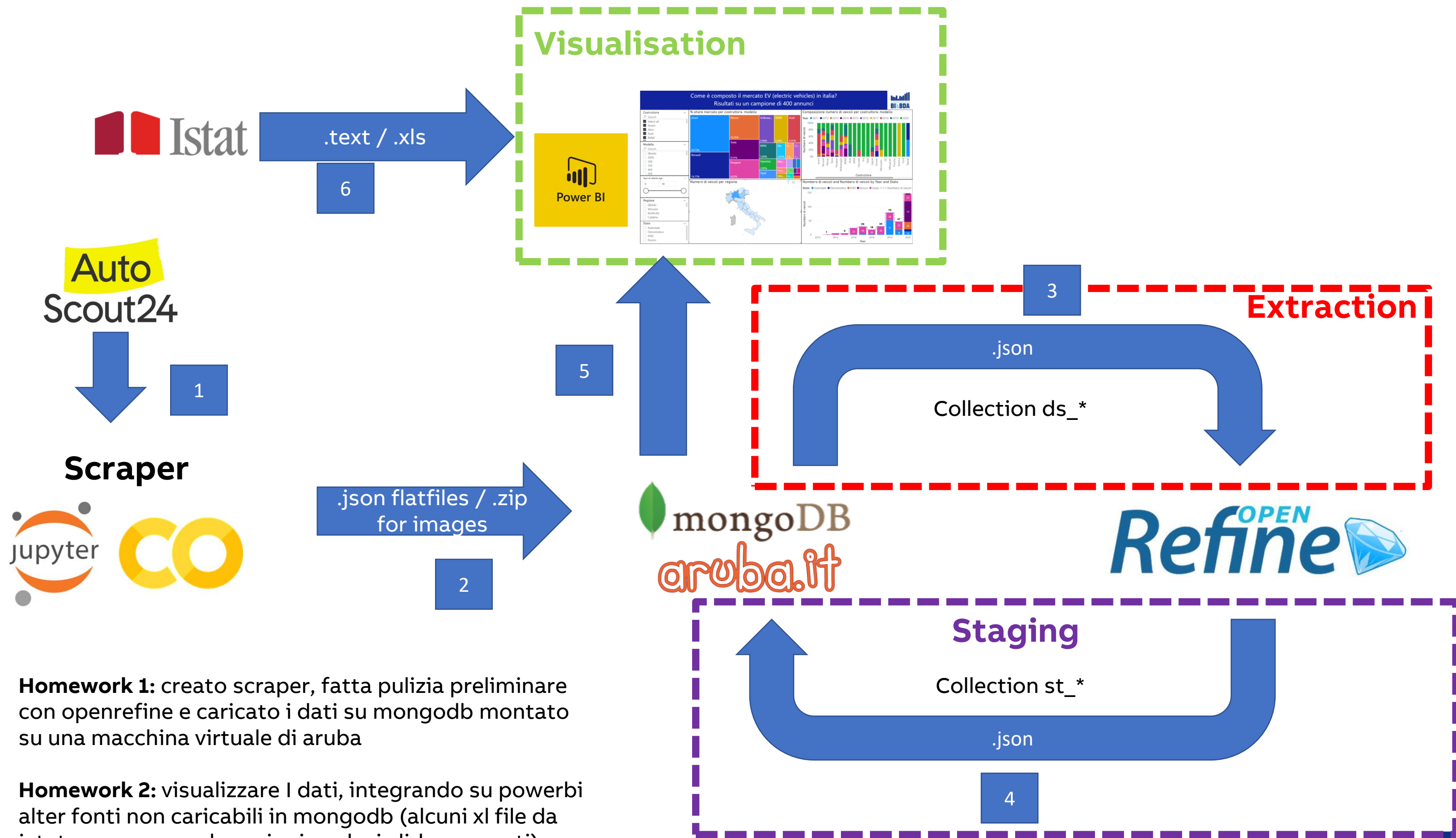
Business target

(aggiornamento 1mo homework)

- **Target:** partendo dalla parte I del compito (l'architettura sta nella slide precedente) costruire una dashboard che contenga tutte le risposte alle domande riportate sempre nella parte 1ma dell'homework.
- **Descrizione generale dell'obiettivo di business:** focalizzandosi sul mercato emergente dei **veicoli elettrici (EV)**, recuperare tutte le informazioni disponibili sulle pagine degli annunci di Autoscout24 elaborando anche i singoli annunci, rispondendo alle seguenti domande:
 - E' vero che il deprezzamento delle auto elettriche è inferiore rispetto alle endotermiche? Una domanda particolarmente interessante visto che, intuitivamente, il tasso di evoluzione tecnologico delle EV è molto più veloce oggi rispetto alle ICE (internal combustion engine), dunque si ci aspetterebbe il contrario.
 - Quale è lo share di mercato dei diversi marchi?
 - Quali variabili influenzano il valore di una vettura elettrica?

Overall architecture

(2ndo homework)



Homework 1: creato scraper, fatta pulizia preliminare con openrefine e caricato i dati su mongodb montato su una macchina virtuale di aruba

Homework 2: visualizzare i dati, integrando su powerbi alter fonti non caricabili in mongodb (alcuni xl file da istat per mappare le regioni, vedasi slide seguenti)

Scraping



- Veloce recap di cosa è stato scaricato da Autoscout24:

Pagina di ricerca annunci

Singolo annuncio

Dalla pagina di ricerca degli annunci sono stati scaricati i dati sommari di un annuncio, ad esempio prezzo, stato (nuova, km 0, etc), anno di immatricolazione...

Dalle pagine dei singoli annunci sono stati scaricati invece gli equipaggiamenti e descrizione dell'annuncio, immagini degli annunci etc



Visualisation



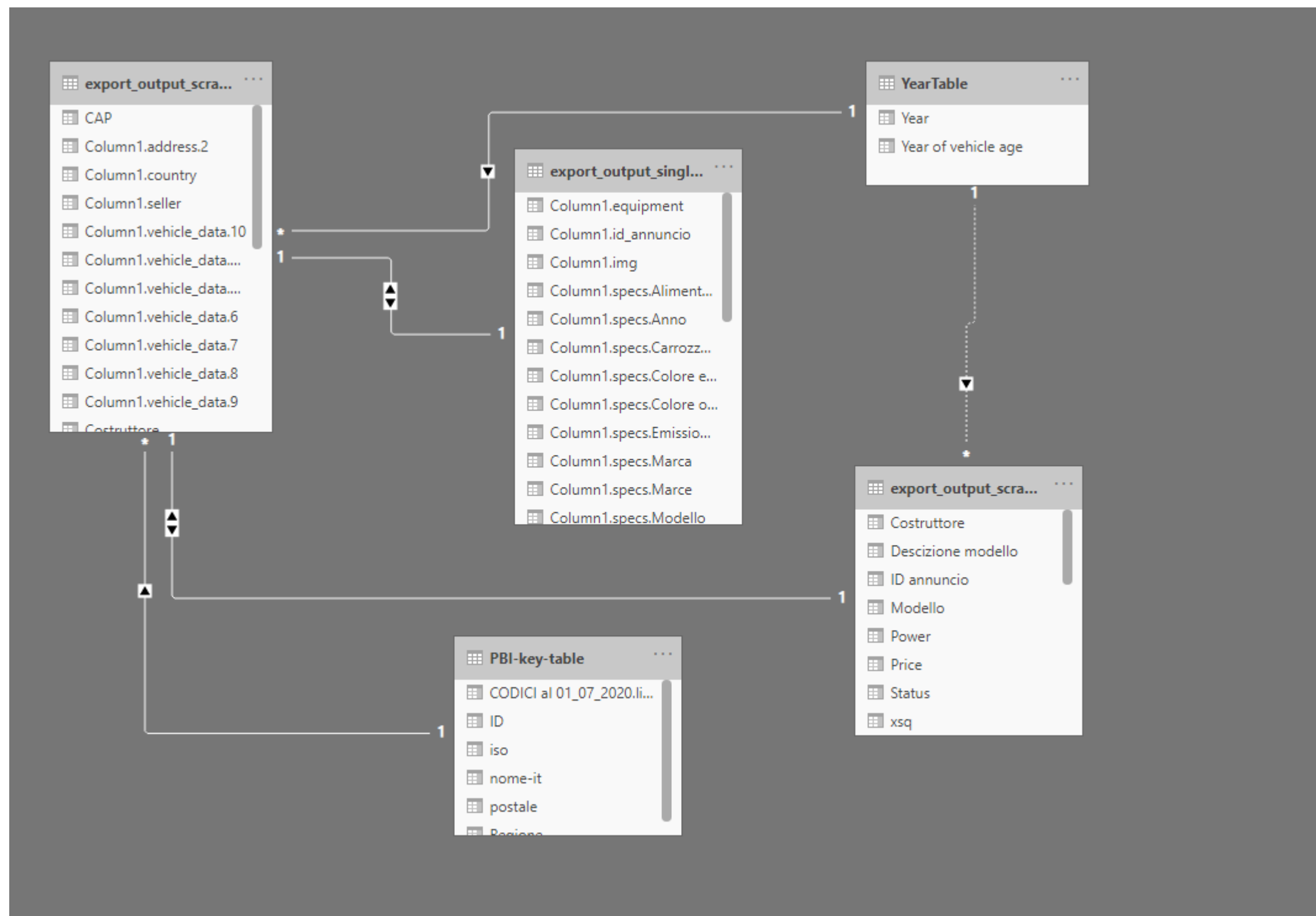
- Utilizzando powerBI, le collection st_* sono state caricate da mongoDB (che si trova su una macchina di aruba) tramite driver ODBC

= Table.ExpandRecordColumn("#Converted to Table", "Column1", {"id_annuncio", "link_annuncio", "vehicle", "vehicle_user_desc", "asking_price", "vehicle_data", "seller", "country", "address"},							
Column1.id_annuncio	Column1.link_annuncio	Column1.vehicle	Column1.vehicle_user_desc	Column1.asking_price	Column1.vehicle_data	Column1.seller	Column1.country
084037c1-3fbc-48ce-b5b2-f1a278522e22	https://www.autoscout24.it/annunci/smart-fortwo-electric-drive-cou...	smart forTwo	electric drive coupé	€ 9.000,- 12	List	Unique	
9dc886a9-480b-d914-e053-010007f3eeb	https://www.autoscout24.it/annunci/smart-fortwo-eq-edition-one-22...	smart forTwo	EQ Edition One (22kW)	€ 31.900,- 2	List	New Ce	
36179b32-ce8a-490a-9ac0-4135dc143a5c	https://www.autoscout24.it/annunci/volkswagen-up-elettrica-grigio-3...	Volkswagen up!		€ 12.600,- 12	List	Grimal	
2587ee6d-6d0e-48a2-86f7-ca85503b369c	https://www.autoscout24.it/annunci/smart-fortwo-electric-drive-pass...	smart forTwo	electric drive Passion	€ 14.890,- 12	List	Autom	
7b2985a9-bc13-3b49-e053-010007fc0b	https://www.autoscout24.it/annunci/renault-twizy-80-technic-elettric...	Renault Twizy	80 Technic	€ 5.500,-	List	Privato	
be8549b3-18da-4098-9f61-e99094246fd2	https://www.autoscout24.it/annunci/smart-fortwo-eq-parisblue-22k...	smart forTwo	EQ Parisblue (22kW)	€ 24.900,- 12	List	City Car	
ce0193ba-0632-4c4b-b369-d40ebdd8f48a	https://www.autoscout24.it/annunci/peugeot-ion-active-elettrica-bia...	Peugeot iOn	Active	€ 13.500,- 2	List	Autolec	
9074bb95-8f77-48a6-875f-e362562b71a7	https://www.autoscout24.it/annunci/smart-forfour-eq-electric-drive...	smart forFour	EQ Electric Drive Prime	€ 16.900,- 12	List	Desaca	
04ad05b0-97eb-4732-a7d5-3ee56a2da3...	https://www.autoscout24.it/annunci/volkswagen-up-e-5p-my20-elett...	Volkswagen up!	e- 5p my20	€ 17.800,- 12	List	Eurocal	
be59766e-2dac-40d0-9072-59782d6952...	https://www.autoscout24.it/annunci/audi-e-tron-spb-55-quattro-s-lin...	Audi e-tron	SPB 55 quattro S line edition	€ 118.860,- 12	List	Frav Sri	
c7861535-b64b-45a2-8f74-6d157046e451	https://www.autoscout24.it/annunci/nissan-leaf-i-leaf-tekna-30kw-10...	Nissan Leaf	I leaf Tekna 30kW 109cv my17	€ 17.900,- 12	List	Comer	
28baa281-a23b-4e42-9f22-c457f161fda1	https://www.autoscout24.it/annunci/smart-fortwo-eq-passion-elettri...	smart forTwo	EQ PASSION ELETTRICA ITALIANA UNICO PROPORIETA...	€ 17.900,- 12	List	Emaut	
ae198071-67f0-4246-9e03-ce82ded9e7a0	https://www.autoscout24.it/annunci/nissan-leaf-40kwh-tekna-elettric...	Nissan Leaf	40KWH TEKNA	€ 29.900,- 12	List	Gruppo	
57e5e60e-be9e-43b7-a8ae-fa6cbf80c11a	https://www.autoscout24.it/annunci/peugeot-2008-2008-motore-elet...	Peugeot 2008	2008 motore elettrico 136 CV GT Line	€ 32.900,- 12	List	Mattar	
02b2ea18-4101-4e45-85d2-ba5cdd99c6d5	https://www.autoscout24.it/annunci/peugeot-2008-2008-motore-elet...	Peugeot 2008	2008 motore elettrico 136 CV GT	€ 33.900,- 12	List	Mattar	

- PowerBI è molto flessibile e consente di operare anche su dati non strutturati (strutturandoli) e liste
- I file dello scraper sono stati ripuliti ulteriormente, per consentire di leggerli con il “data type” corretto
- Sono stati anche importati dei txt e xl da locale di Istat, per mappare gli indirizzi CAP alle rispettive regioni, per visualizzare su mappa italiana i risultati dello scraping

Visualisation II

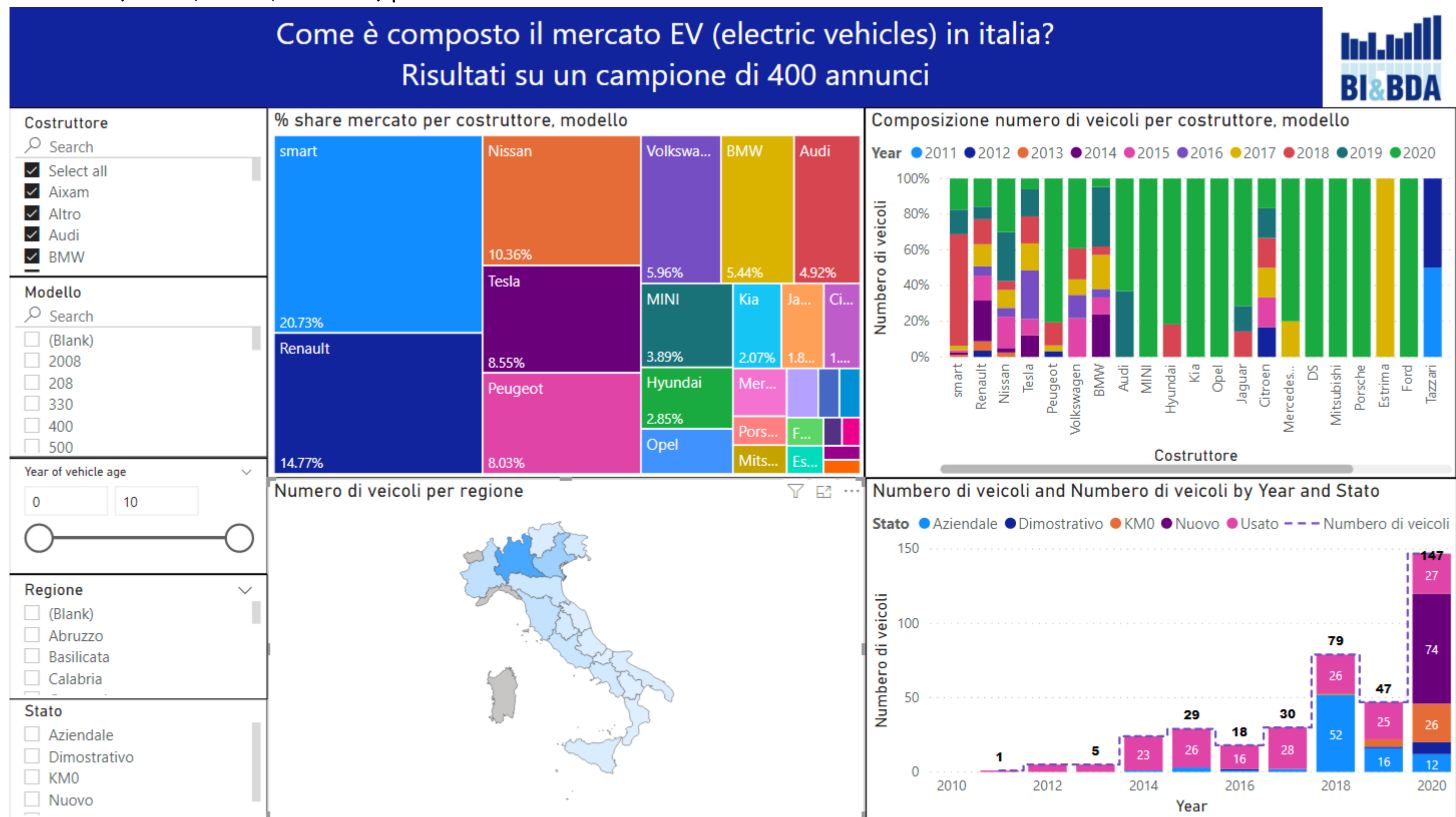
- Data model:



Visualisation III



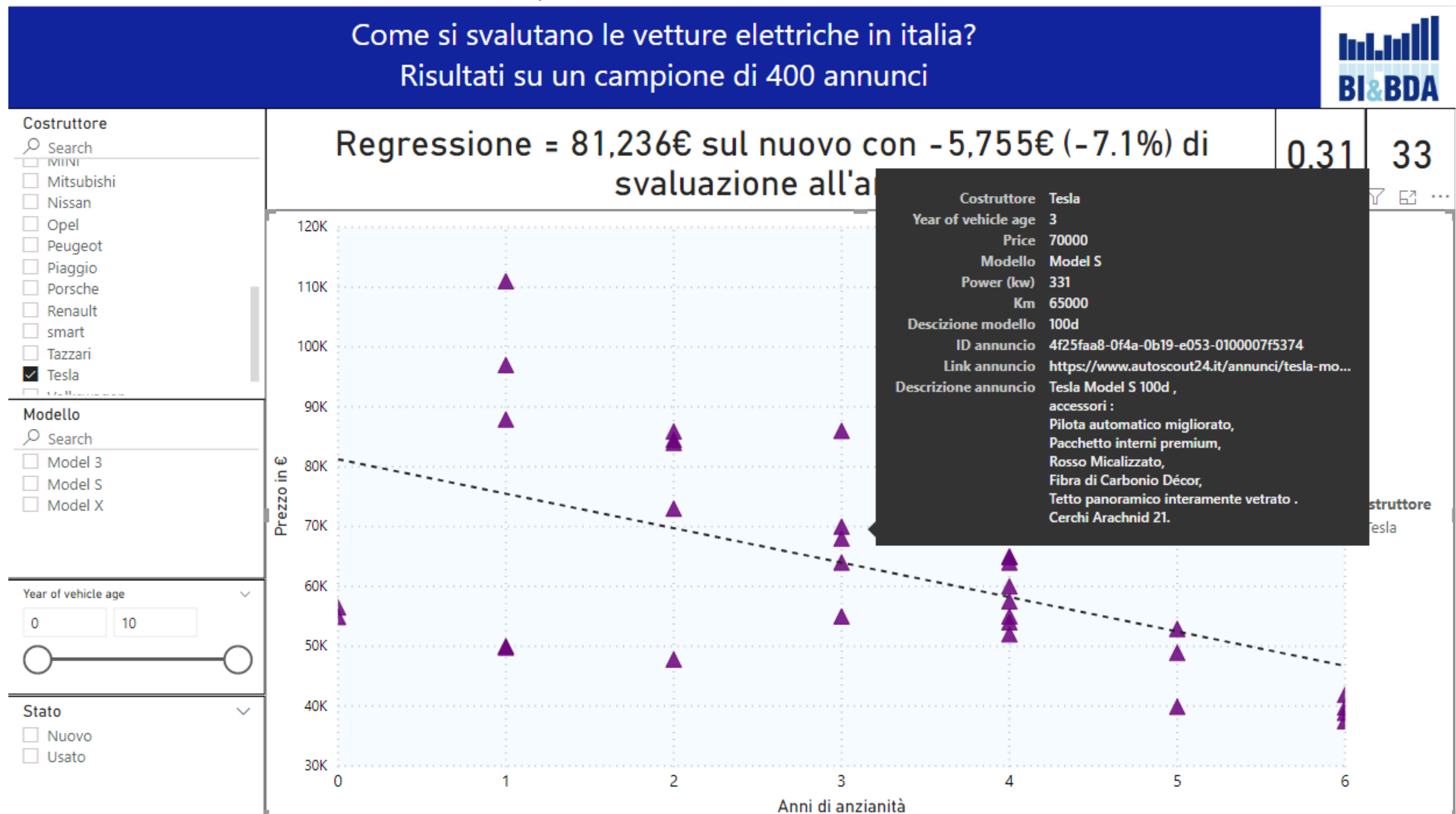
- Sono state create 2 viste in PowerBI.
- La prima (screenshot qui sotto) è una sommario che descrive gli annunci di autoscout24. In senso orario, da sinistra abbiamo una sezione per applicare eventuali filtri, un treemap con la composizione di share di mercato per costruttore (con possibilità di drilldown al singolo modello), a sinistra la % di immatricolazione per costruttore (sempre drill down a modello)
- Più sotto abbiamo una heatmap per il numero di immatricolazioni per regione, ed a sinistra una barchart riguardante lo stato (nuovo, usato, KM0 etc) per anno dei veicoli in vendita



Visualisation IV



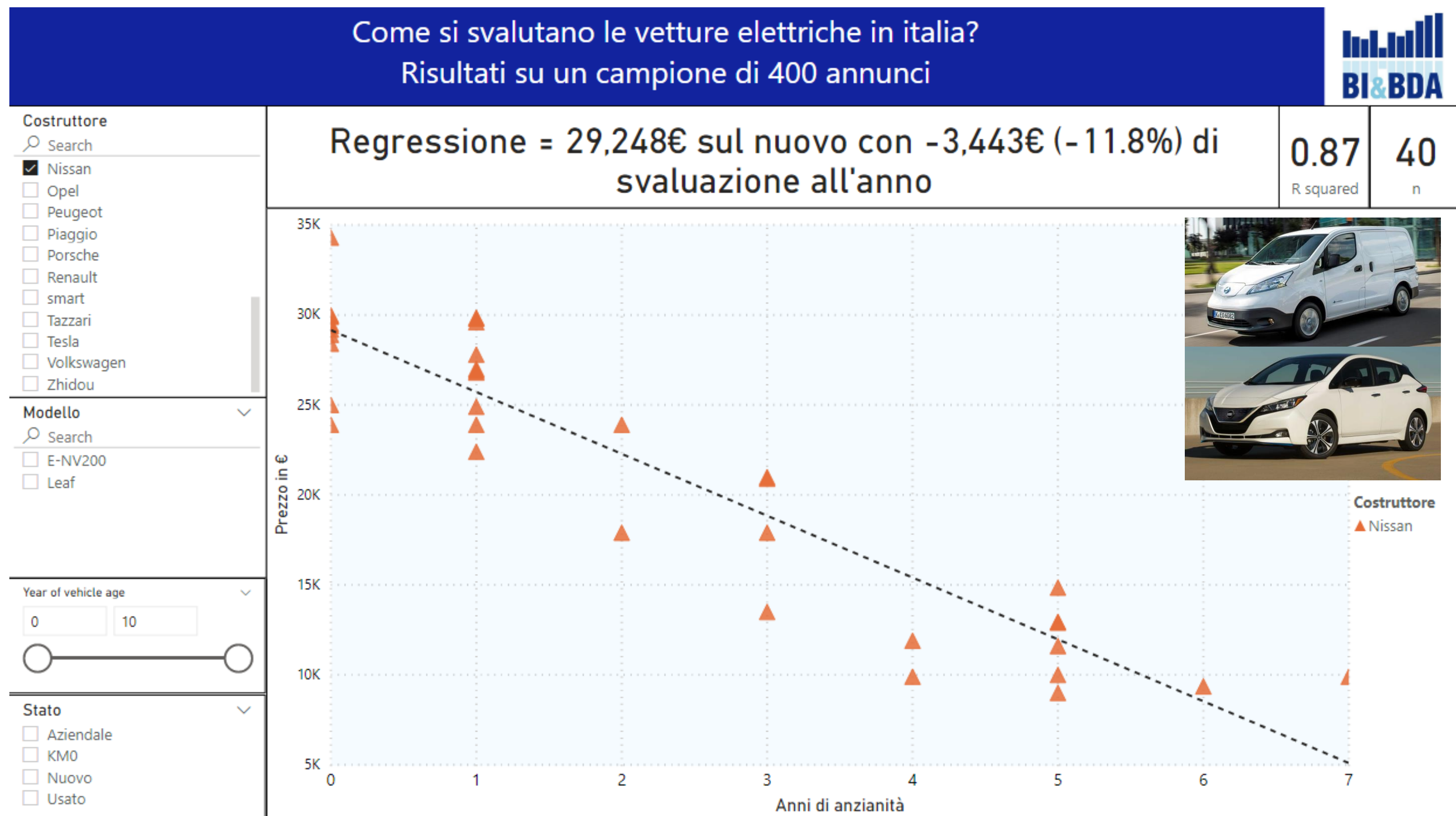
- Nella seconda vista, ho fatto una regressione lineare per calcolare il costo di partenza e la svalutazione per anno di anzianità di un veicolo.
- In più sono disponibili una serie di filtri per costruttore/modello e tutti i dettagli dei singoli annunci possono essere consultati in hover come nello screenshot qui sotto



Visualisation III



- Per esempio, Nissan conta 4 osservazioni nel campione scrapato, con un prezzo di partenza di 29K€ e una svalutazione annua di circa 3.4K€ (parti a 12% del costo iniziale). Ottima goodness of fit (R quadrato) di 0.87
- Non c'è una differenza sostanziale tra nissan leaf (la berlina) e E-NV200, sebbene uno sia un veicolo da privati e l'altro commerciale. Questo si può ipotizzare sia dovuto alla tecnologia dei veicoli (batteria, motore elettrico e pianale sono gli stessi).

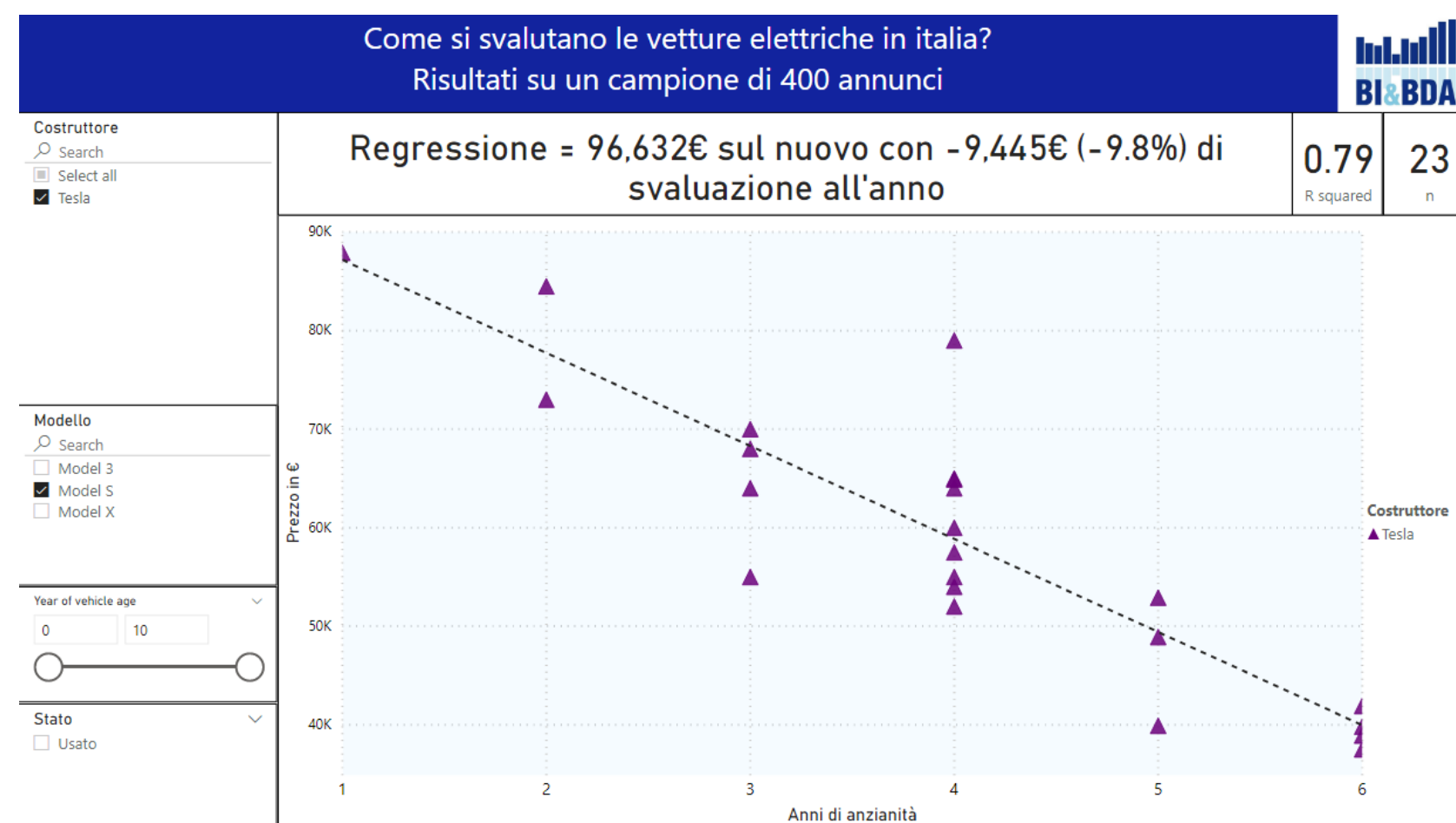


Visualisation V

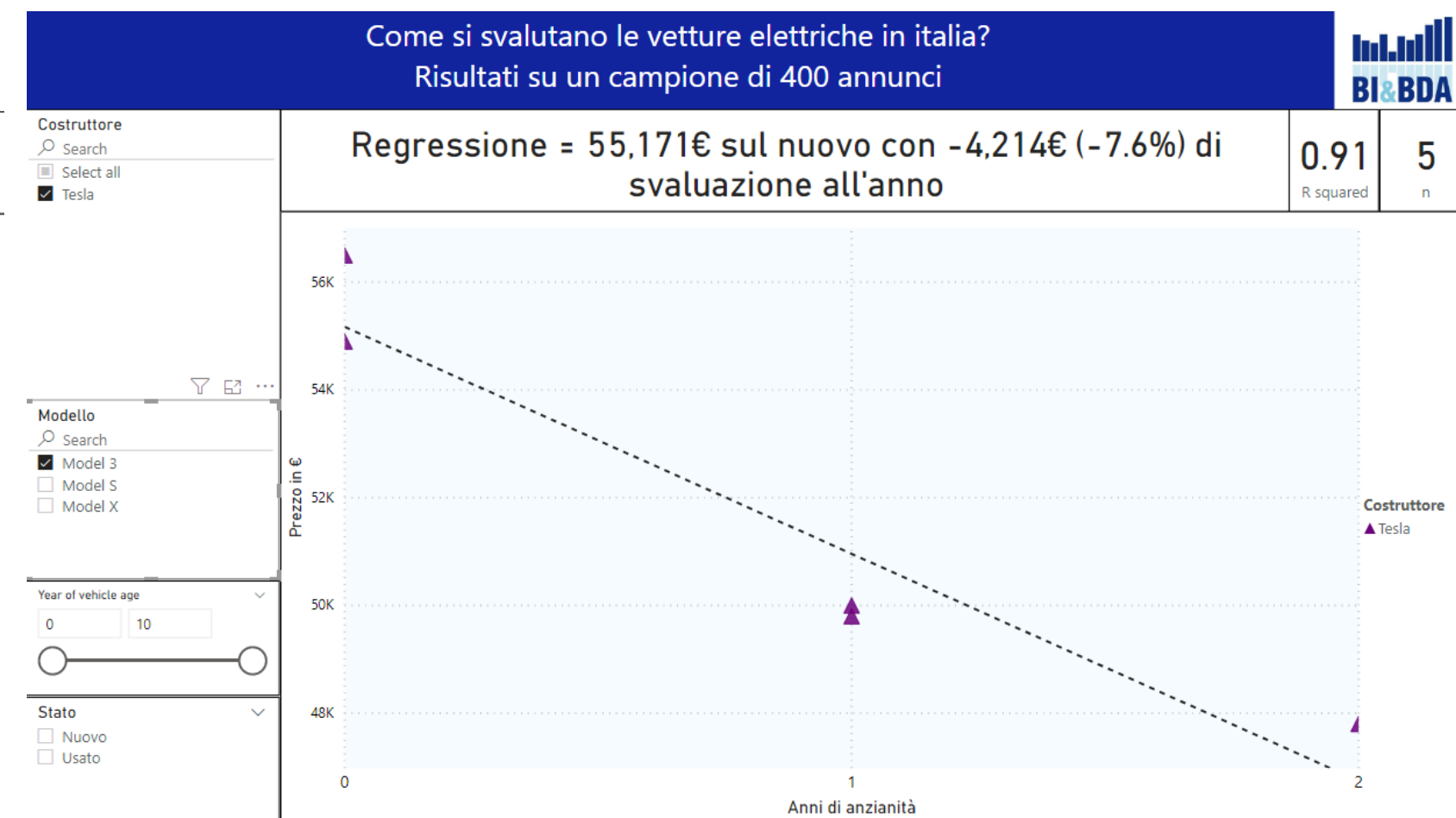


- Tesla invece ha un deprezzamento % inferiore agli altri competitor, confermando la forte domanda per il marchio e il mantenimento del prezzo nel tempo, in particolare model 3 si svaluta del 7.6% annuo, confermando articoli giornalistici come il seguente ([link](#))

Model S



Model 3

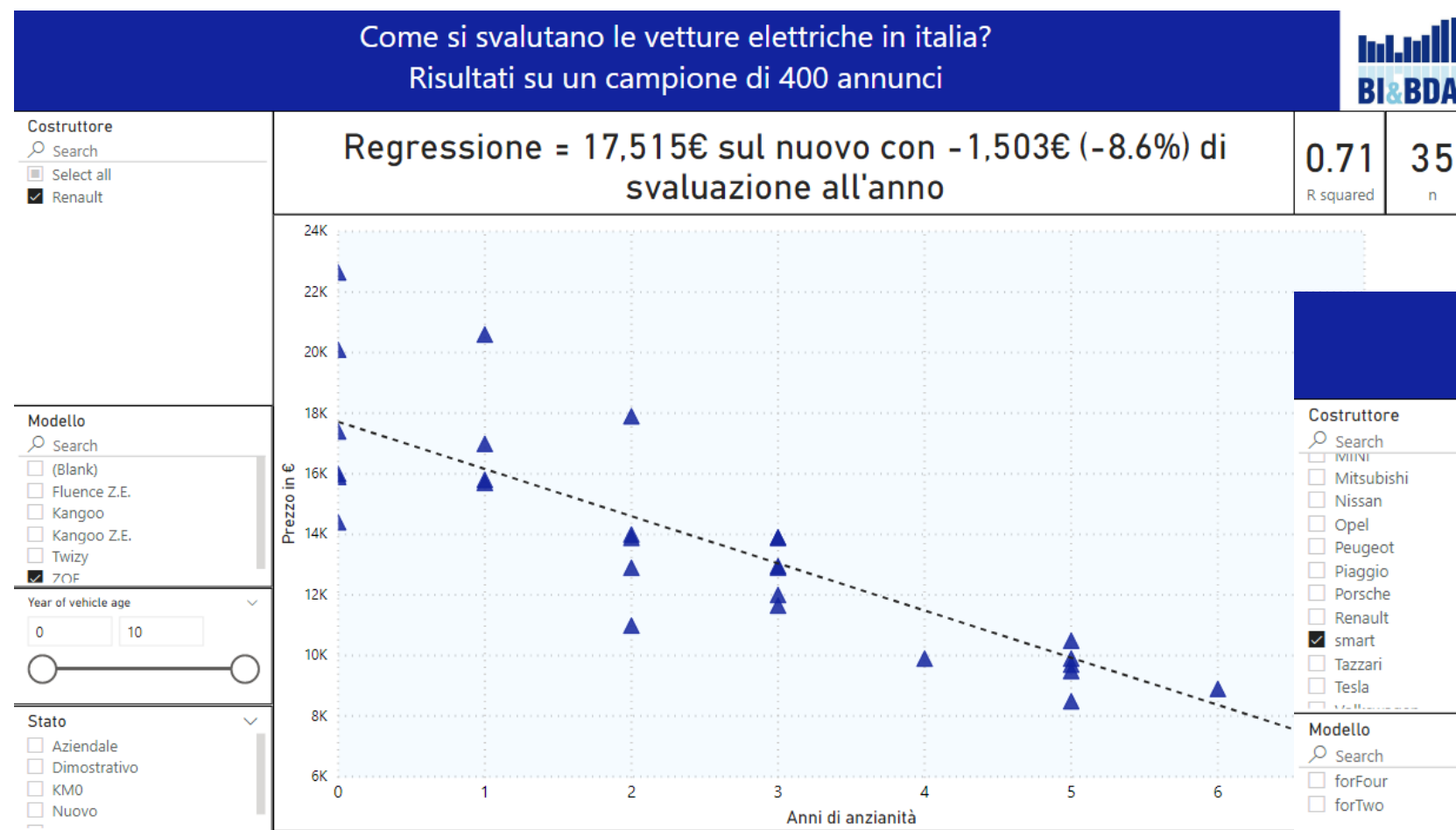


Visualisation VI

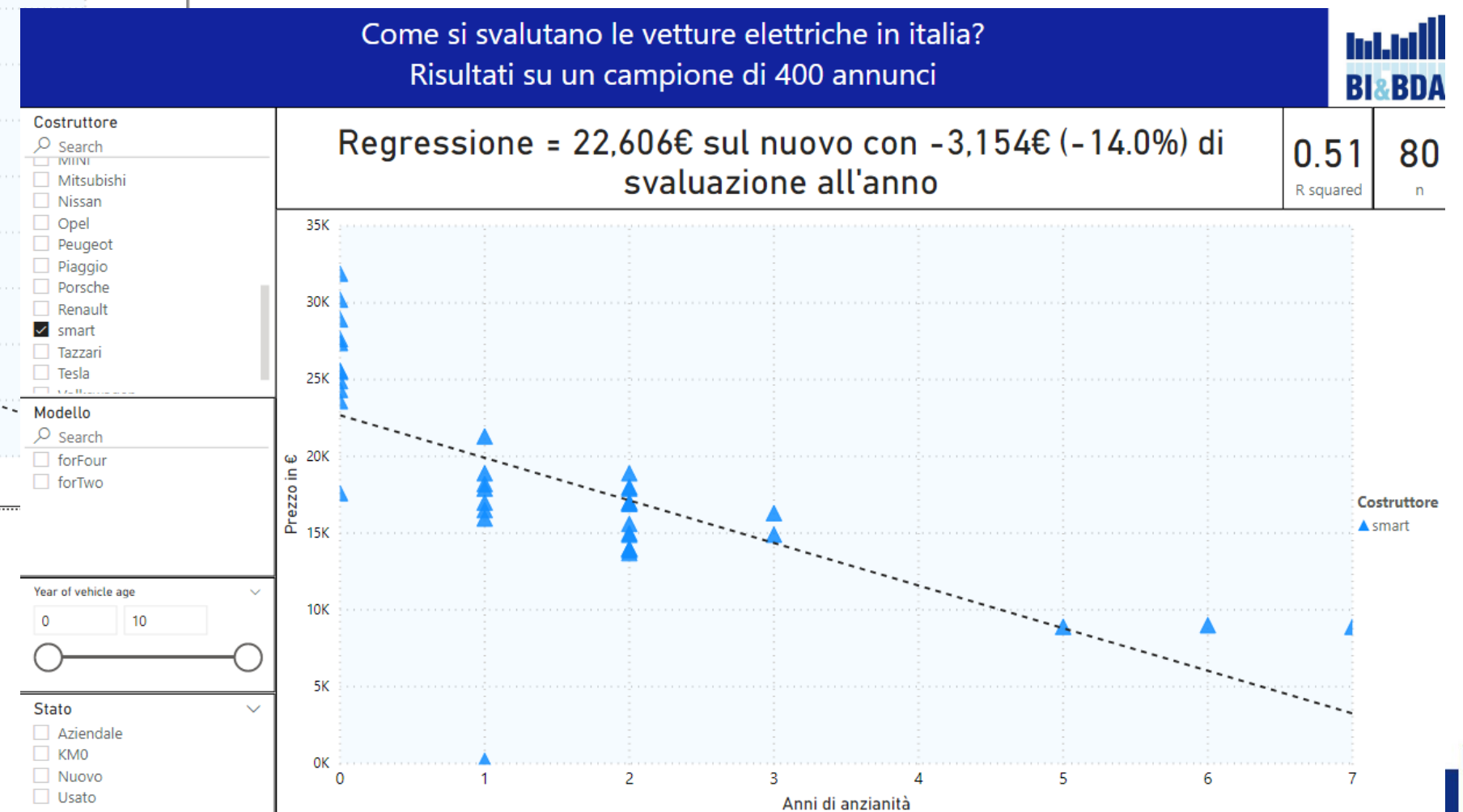


- Altre marche si deprezzano maggiormente:

Renault ZOE



Smart

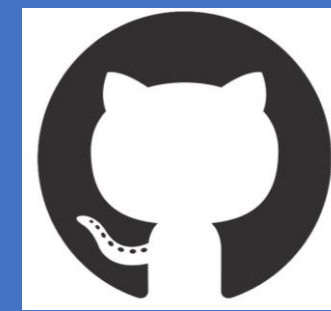


Conclusioni



- La visualizzazione è stata efficace ed ha confermato sia la composizione di mercato che la svalutazione dei veicoli in base all'anzianità del veicolo, sebbene si basi solo su un campione di 400 annunci da autoscout24.
- Reperire i dati effettivi di svalutazione può essere molto utile, soprattutto per veicoli puramente elettrici.
- Sebbene la visualizzazione sia efficace, il modello di regressione non prende in considerazione allestimenti differenti, kilometraggio etc, quindi può essere migliorato ulteriormente.
- Dal campione (di autoscout) si dovrebbe passare alla superpopolazione (sempre di autoscout) per infierire la popolazione vera !!
- Siccome c'è stato ulteriore trasformazione e pulizia del dato in powerBI, questi cambiamenti andrebbero applicati direttamente in OpenRefine.
- Gli immagini degli annunci non state utilizzate !

Github



- Tutto il materiale, includendo la dashboard di PowerBI, è disponibile sulla repo, cartella «homework – parte II - visualisation»:

<https://github.com/szambetti/Data-science-Master-UNIMIB/tree/master/Module%204%20-%20Big%20Data%20%26%20Analytics/>