

Business Intelligence per i Big Data

Quaderno #2 – MongoDB

Introduzione

L'obiettivo del quaderno è quello di analizzare e sintetizzare informazioni utili da un database Mongo. Lo studio è condotto su un dataset reale che raccoglie le informazioni sui parcheggi e sui noleggi di veicoli per il *Car Sharing*.

Dataset

L'analisi verte su dei dati reali relativi al servizio di Car Sharing nell'area metropolitana di Torino. Il dataset originale contiene ~160.000 documenti, divisi in due collezioni. Lo schema delle due collezioni è descritto di seguito.

Per accedere al database utilizzare le seguenti credenziali su MongoDB Compass:

- Hostname: bigdatadb.polito.it
- Port: 27017
- Authentication: Username/Password
- Username: Compass
- Password: Compass19!
- Authentication database: dbdmg
- SSL: Unvalidated (insecure)

Parkings

Il database Parkings contiene informazioni relative ai parcheggi di auto associate ad un servizio di *Car Sharing*. Gli attributi più importanti per ogni documento sono elencati nella tabella 1.

Nome	Tipo	Descrizione
_id	objectid	identificativo del documento
address	string	Indirizzo di parcheggio del veicolo
city	string	Città nella quale è avvenuto il parcheggio
engineType	string	Identificativo del tipo di motore del veicolo
exterior	string	Stringa identificativa delle condizioni esterne del veicolo
final_date	date	Data e ora di fine stazionamento del veicolo
fuel	int32	Livello di carburante del veicolo durante il periodo di stazionamento

init_date	date	Data e ora di inizio stazionamento del veicolo
interior	string	Stringa identificativa delle condizioni interne del veicolo
loc	coordinates	Coordinate della posizione di parcheggio del veicolo
plate	int32	Identificativo della targa del veicolo
smartphoneRequired	boolean	Valore binario (Vero/Falso) che indica qualora sia necessario lo smartphone per il noleggio del veicolo
vendor	string	Azienda proprietaria del veicolo da noleggiare
vin	string	Identificativo del telaio del veicolo

Tabella 1: descrizione del database Parkings.

Bookings

Il database Bookings contiene informazioni relative al noleggio di veicoli per un servizio di *Car Sharing*. Gli attributi più importanti per ogni documento sono elencati nella tabella 2.

Nome	Tipo	Descrizione		
_id	objectid	identificativo del documento		
car_name	string	Modello del veicolo		
city	string	Città nella quale è avvenuto il noleggio		
distance	int32	Distanza percorsa durante il noleggio (in metri)		
driving	object	distance	int32	Distanza percorsa durante il noleggio (in metri)
		duration	int32	Durata del noleggio (in secondi)
engineType	string	Identificativo del tipo di motore del veicolo		
exterior	string	Stringa identificativa delle condizioni esterne del veicolo		
final_address	string	Indirizzo della posizione di fine noleggio		
final_date	date	Data e ora di fine noleggio		
final_fuel	int32	Livello del carburante a fine noleggio		
init_address	int32	Indirizzo della posizione di inizio noleggio		
init_date	date	Data e ora di inizio noleggio		
init_fuel	int32	Livello del carburante a inizio noleggio		
interior	string	Stringa identificativa delle condizioni interne del veicolo		
plate	int32	Identificativo della targa del veicolo		

smartphoneRequired	boolean	Valore binario (Vero/Falso) che indica qualora sia necessario lo smartphone per il noleggio del veicolo		
vendor	string	Azienda proprietaria del veicolo da noleggiare		
walking	object	distance	int32	Distanza percorsa a piedi per raggiungere il veicolo.
		duration	int32	Durata del percorso a piedi per raggiungere il veicolo.

Tabella 2: descrizione del database Bookings.

Specifiche per la generazione del report

Preparare una breve relazione (max 2 pagine) per descrivere i risultati dell'analisi includendo:

- la risposta a ciascuna domanda e, ove necessario, la query - o le query - usate per rispondere alla domanda.
- gli screenshot dei risultati ottenuti (gli screenshot non rientrano nel conteggio del numero massimo di pagine).

Domande

Rispondere alle seguenti domande:

1. Quale periodo di tempo è coperto dalla base dati? Riportare uno screenshot che supporti la risposta.
2. Qual è il giorno della settimana in cui sono stati effettuati più noleggi? Riportare uno screenshot che supporti la risposta.
3. Qual è l'intervallo orario della settimana in cui sono stati effettuati più noleggi? Riportare uno screenshot che supporti la risposta.
4. Quanti sono i parcheggi effettuati nel cerchio con centro nel punto A = [7.694652260765925, 45.07449636396572] e raggio R = 0.00005117446809534804?
5. Rispondere alle domande 2 e 3 considerando solo i parcheggi selezionati al punto 4.
6. Quanti sono i veicoli che, considerando tutti i periodi di stazionamento, sono stati parcheggiati con un livello di carburante medio maggiore o uguale a 97. Visualizzare le targhe e il livello medio di carburante di questi veicoli in ordine decrescente di livello carburante.
7. Quali sono i primi 3 veicoli che hanno percorso una minore distanza considerando tutti i noleggi? Visualizzarne la targa e il conteggio in ordine crescente per conteggio.
8. Quali sono i fornitori distinti presenti nella base dati? Quanti noleggi sono stati registrati per ciascuno di essi?
9. Al punto 4 abbiamo analizzato l'attività nell'area in termini di parcheggi. Ora guardiamo l'affluenza in termine di noleggio che termina in quell'area. Scegliere un secondo punto nella mappa a piacimento (ad es. B = [7.695856, 45.064539]), lasciando R invariato. Quanti sono i viaggi effettuati che iniziano in B e terminano in A? (*Hint: le coordinate del luogo di partenza/arrivo del record si trovano in "origin_destination.coordinates.0/1"*)
10. Rispondere alle domande 2 e 3 considerando solo i noleggi ottenuti al punto 9.

N.B.: le prime tre domande fanno riferimento al sample di 1000 documenti considerato da MongoDB Compass.