ArchiVehicle





Corso di Basi di Dati 2

Documentazione di Progetto

Docenti: Prof.ssa Genoveffa Tortora Prof. Luigi Di Biasi Studente: Daniele Fabiano 0522501738

Table of Contents

Introduzione	3
Descrizione del Sistema	3
Dataset	4
Informazioni iniziali sul dataset di partenza	4
Generazione del dataset dei produttori di veicoli	5
Data Cleaning	8
Database	
Scelta del Database	g
Schema dei dati	g
Struttura di un documento della collezione dei produttori	
Struttura di un documento della collezione dei veicoli	
Import dei dataset	
Applicazione Web	
Scelta delle tecnologie	
Pymongo	
Flask	
Bootstrap	
Descrizione delle pagine	
Pagina Home	
Pagina Vehicles	
Pagina Insert Vehicle	
Pagina Show Single Vehicle	
Pagina Manufacturers	
Pagina Stats	
Installazione di ArchiVehicle	

Introduzione

Questa documentazione intende descrivere il progetto realizzato per il corso di Basi Di Dati 2. All'interno di questo documento sono forniti i dettagli rispetto ai seguenti aspetti:

- Requisiti del progetto da realizzare;
- Scelta del dataset da utilizzare;
- · Analisi e pulizia del dataset;
- Scelta del DBMS da utilizzare;
- Creazione dello schema dei dati;
- Caricamento dei dati sul database;
- Scelta delle tecnologie web da utilizzare;
- Descrizione delle pagine realizzate e relative operazioni;
- Installazione dell'applicazione realizzata.

Descrizione del Sistema

ArchiVehicle nasce con l'obiettivo di semplificare e digitalizzare la gestione relative alle informazioni del parco veicoli di un concessionario. Questo software è destinato all'utenza di un'attività commerciale che si occuperà di recuperare le informazioni dei veicoli e di memorizzarle nel sistema, attraverso un'interfaccia grafica fornita da un applicazione web client/server. ArchiVehicle supporta le seguenti funzionalità:

- 1. Inserimento, modifica, cancellazione delle informazioni di un veicolo;
- 2. Inserimento, modifica, cancellazione delle informazioni di un produttore di veicoli;
- 3. Ricerca di un veicolo mediante l'applicazione di alcuni filtri;
- 4. Visualizzazione rapida del numero di veicoli da parte di ogni produttore, sulla base di varie caratteristiche per veicolo.

Dataset

Informazioni iniziali sul dataset di partenza

Non avendo direttamente a disposizione le informazioni dei veicoli, per simulare l'utilizzo di questa applicazione si è deciso di scegliere un dataset pubblico reperibile su Kaggle. Il dataset scelto è <u>Vehicle Dataset 2024</u>. Per ogni colonna è stata definita una descrizione per l'informazione mostrata e il tipo del dato da memorizzare. Laddove necessario, sono specificati i possibili valori inseribili per un elemento della colonna considerata:

Nome colonna	Descrizione	Tipo di dato
name	Il nome completo del veicolo	Stringa
description	Una breve descrizioine del veicolo	Stringa
make	Il nome del produttore del veicolo	Stringa
model	Il nome del modello del veicolo	Stringa
type	La tipologia del veicolo	Stringa: {"New", "Used"}
year	L'anno di produzione del veicolo	Intero
price	Il prezzo del veicolo in dollari	Decimale
engine	Informazioni sul motore del veicolo	Stringa
cylinders	Il numero di cilindri del motore del veicolo	Decimale
fuel	Il tipo di carburante usato dal veicolo	Stringa:
mileage	Il numero di miglia percorse dal veicolo	Decimale
transmission	Il tipo di trasmissione del veicolo	Stringa
trim	Il livello di accessori del veicolo	Stringa
body	Lo stile della carrozzeria del veicolo	Stringa: {"Convertible", "Hatchback", "Pickup Truck", "SUV", "Sedan", "Cargo Van", "Passegner Van", "Minivan"}
doors	Il numero di porte del veicolo	Decimale
exterior_color	Il colore esterno del veicolo	Stringa
interior_color	Il colore interno del veicolo	Stringa
drivetrain	Il tipo di trazione del veicolo	Stringa: {"All-wheel Drive", "Four-wheel Drive", "Front-wheel Drive",

		"Rear-wheel Drive"}
--	--	---------------------

Generazione del dataset dei produttori di veicoli

Avendo a disposizione tutte le informazioni sui veicoli, ma la mancanza di quelli dei produttori, si è deciso di sfruttare l'utilizzo di ChatGPT per richiedere la generazione di un dataset sui veicoli. Dopo aver estratto l'elenco dei nomi di tutti i produttori, sfruttando le tecniche che saranno mostrate nella sezione **Data Cleaning**, è stato generato e scaricato il dataset in formato csv. Nelle figure seguenti viene mostrata la conversazione realizzata:

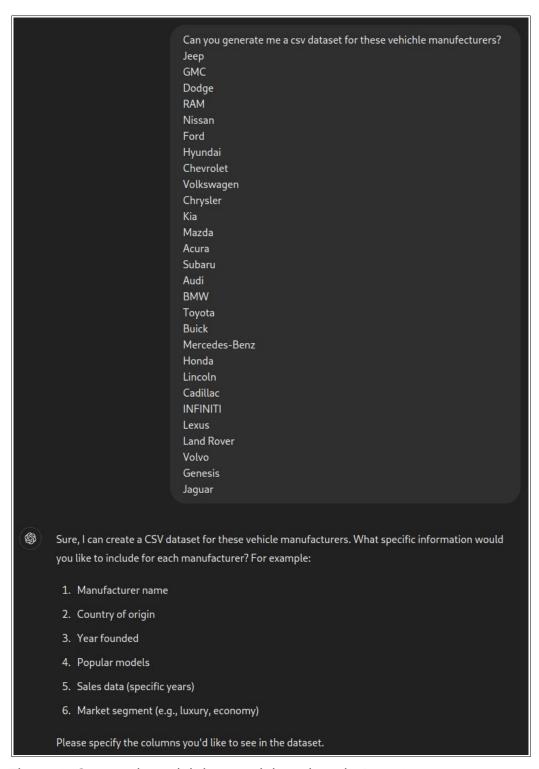


Figure 1: Generazione del dataset dei produttori - Parte1



Figure 2: Generazione del dataset dei produttori - Parte2



Figure 3: Generazione del dataset dei produttori - Parte3

Il dataset così ottenuto presenta le seguenti colonne:

Nome colonna	Descrizione	Tipo di dato
name	Il nome dell'azienda di produzione	Stringa
country	Il paese dove è stato fondata l'azienda di produzione	Stringa
year	L'anno di fondazione dell'azienda di produzione	Intero
founder	Il nome del fondatore dell'azienda di produzione	Stringa
description	Una descrizione dell'azienda di produzione	Stringa

Data Cleaning

Una volta ottenuti i dataset necessari, prima di caricarli nel db è stato necessario effettuare alcune operazioni di data cleaning. Per questo motivo sono stati realizzati due notebook che grazie all'utilizzo della libreria di data processing *pandas*, hanno contributo a realizzare le operazioni di pulizia e trasformazione del dataset, correttamente importato nel notebook come un oggetto dataframe. Sono di seguito riportate le principali operazioni effettuate:

vehicles_dataset_manufacturer_extract.ipynb:

- 1. Estrazione della colonna make che contiene i nomi dei produttori;
- 2. Rimozione dei duplicati dalla colonna estratta;
- 3. Salvataggio della colonna estratta su un file csv.

vehicles_dataset_cleaning.ipynb:

- 1. Rimozione dell'anno di produzione dagli elementi della colonna name;
- 2. Rimozione delle righe che presentano righe con valori nulli e/o vuoti;
- 3. Rimozione di caratteri di newline/spacing in eccesso dagli elementi della; colonna description;
- 4. Cambio del tipo da float a int per le colonne cylinders e doors;
- 5. Recupero degli indici delle righe che presentano valori numerici non significativi pari a "zero" (0 oppure 0.0) e rimozione delle righe individuate;
- 6. Recupero degli indici delle righe dove gli elementi della colonna description hanno un numero di caratteri inferiori a 100 e rimozione delle righe individuate;
- 7. Rimozione delle righe che presentano un valore duplicato per gli elementi della colonna name;
- 8. Cambio nome della colonna body in chassis;
- 9. Semplificazione di alcun tipi di chassis, raggruppando in una unica categoria "Van" le tre possibile sotto-categorie indicate nella descrizione del dataset;
- 10. Creazione della colonna image e inserimento dei path di un immagine esemplificativa per ogni categoria di veicolo come elementi della colonna;
- 11. Cambio del nome della colonna make in manufactuer_id e sostituzione dei nomi dei produttori con l'id del rispettivo documento del produttore caricato nel database;
- 12. Cambio nome delle colonne e riposizionamento di alcune di esse, per garantire una corretta importazione del dataset sul database;
- 13. Salvataggio del nuovo dataset ottenuto su un file csv.

Database

Scelta del Database

Per la gestione dei dati persistenti di questa applicazione, si è deciso di utilizzare MongoDB. Sono vari i motivi che hanno portato alla scelta dell'utilizzo di questo database:

- 1. Strutturazione dei dati: grazie alla flessibilità offerta da un database NoSQL, è possibile memorizzare in maniera meno rigida le informazioni parziali di un'entità. Nel nostro caso, il concessionario potrebbe ricevere dal produttore solo le informazioni essenziali di alcuni veicoli ma voler comunque iniziare a memorizzarle e poi integrarle in un secondo momento con quelle mancanti.
- 2. Scalabilità: un database NoSQL permette un'espandibilità maggiore, potendo così facilitare la distribuzione dei dati su più nodi di una rete. Nel nostro, il concessionario potrebbe avere più sedi sparse sul territorio, avendo quindi la necessità di usare l'applicazione su più dispositivi e dovendo definire una soluzione distribuita per il database.

Schema dei dati

}

Anche se il database utilizzato è di tipologia NoSQL, per le due collezioni individuate sono state comunque definite le strutture dei documenti che costuiscono lo schema dei dati.

Struttura di un documento della collezione dei produttori

Per un documento della collezione dei produttori, essendo presenti pochi attributi si è optato per una struttura su un singolo livello. Di seguito viene riportato un esempio di un documento:

Struttura di un documento della collezione dei veicoli

Per un documento della collezione dei veicoli, sono presenti molti più attributi e c'è la necessità di raggrupparli per definire una struttura su due livelli e con sottodocumenti. Sulla base delle considerazioni effettuate nella sezione <u>Scelta del Database</u>, si è scelto di suddividere la struttura del documento come mostrato nel seguente esempio:

```
{
   "id": {
       "$oid": "66957f0b62d99ecd714e267b"
   "name": "Jeep Grand Cherokee 4xe Trailhawk",
   "description": "Diamond Black Crystal Pearlcoat 2023 Jeep Grand Cherokee Trailhawk 4xe 4WD 8-
                   Speed Automatic 2.0L I4 DOHC 2.0L I4 DOHC, 4WD. Clement Chrysler Dodge Jeep Ram
                   of Columbia is pleased to offer this 2023 Jeep Grand Cherokee Trailhawk 4xe in
                   Diamond Black Crystal Pearlcoat with Global Black Leather. This Grand Chero..."
   "model": "Grand Cherokee 4xe",
   "chassis": "SUV",
   "year": 2023,
   "price": 59547 (in dollari),
   "image": "../static/img/vehicle_sample_suv.png",
   "manufacturer_id": {
       "$oid": "6687ea0b2d6d295df5c1f188"
   },
   "mechanical_details": {
       "engine": "16V GDI DOHC Turbo Hybrid",
       "cylinders": 4,
       "fuel": "Hybrid",
       "mileage": 185,
       "transmission": "8-Speed Automatic",
       "drivetrain": "Four-wheel Drive",
   },
   "body_details": {
       "trim": "Trailhawk",
       "doors": 4,
       "exterior_color": "Diamond Black",
       "interior color": "Global Black"
   }
```

}

Import dei dataset

Completate le operazioni di generazione, trasformazione, pulizia dei dataset e una volta definita la struttura dei documenti, si è effettuato l'import dei dataset all'interno del database, grazie a Compass, la GUI del DBMS MongoDB. Per consentire la corretta costruzione delle collezioni di documenti, è stato importato per primo il dataset dei produttori, così da permettere ai documenti della collezione dei veicoli la definizione del riferimento con il rispettivo documento della collezione dei produttori. Attraverso le seguenti figure, viene mostrato come sono effettuate le operazioni di import, avendo l'accortezza di selezionare il tipo corretto per ogni colonna del dataset:

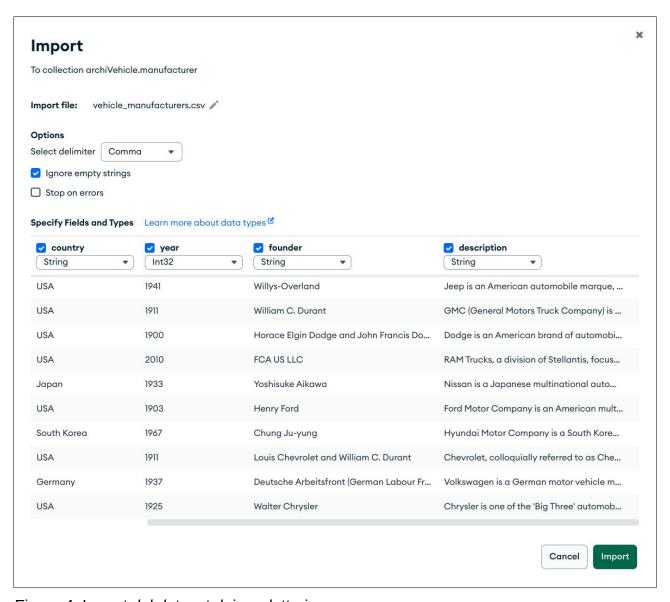


Figure 4: Import del dataset dei produttori

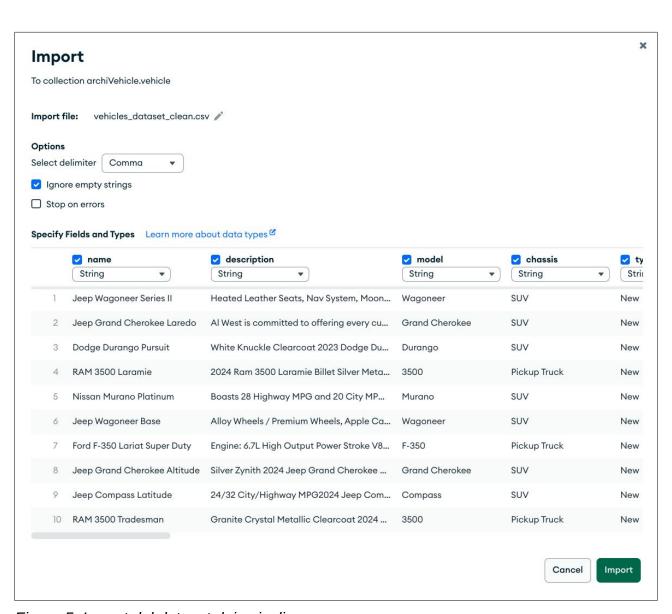


Figure 5: Import del dataset dei veicoli

Applicazione Web

Scelta delle tecnologie

Considerati i tempi ristretti per la realizzazione di questa applicazione, si è deciso di usare tecnologie che non prevedono una curva di apprendimento troppo ampia e permettono di scrivere poco codice, facendo riuso di componenti già definite. Per questi motivi si è deciso di scegliere le seguenti tecnologie:

Database Driver: PyMongo

Logica applicativa / back-end: Flask

Logica di presentazione / front-end: Bootsrap

Pymongo

PyMongo è il driver raccomandato dagli sviluppatori di MongoDB per sfruttare le potenzialità del loro database all'interno di un'applicazione web. PyMongo mette a disposizione tutte le operazioni fondamentali per interagire con il database. Una volta installato il pacchetto del driver, il primo passaggio è quello di importare il pacchetto e inizializzare una connessione con il database:

```
from pymongo import MongoClient
client = MongoClient('localhost', 27017)
```

Flask

Flask è un micro-framework ideato per lo sviluppo di applicazioni web in maniera semplice ed efficace. Essendo un micro-framework, sono di base messe a disposizione solo le funzionalità essenziali come il server di sviluppo e debug, il routing, elaborazione di richieste e risposte. Una volta installato il relativo pacchetto, il primo passaggio è quello di definire creare un modulo app. py e definire la prima route per la pagina index della nostra applicazione web, che inizialmente risponderà solamente con il classico messaggio di "Hello, World!":

```
from flask import Flask
@app.route("/")
def index():
    return "Hello, World!"
```

Bootstrap

Bootstrap è un framework ideato per creare pagine web responsive, fornendo dei template per ogni componente necessaria. Insieme alle funzionalità offerte da Flask per includere componenti html, è possibile riusare le varie componenti sulle pagine dell'applicazione. Per iniziare ad usare Bootstrap, bisogna includere i relativi tag nell'head di una pagina.

Descrizione delle pagine

Pagina Home

Alla pagina Home è possibile effettuare una ricerca dei veicoli e applicare alcuni filtri per delle proprietà dei veicoli.

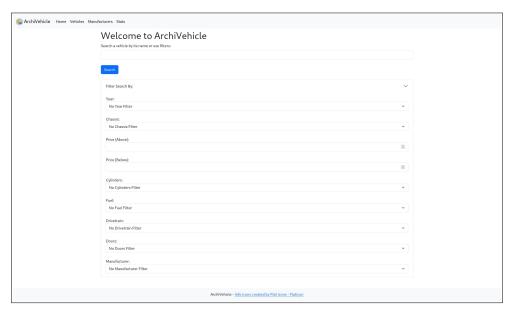


Figure 6: Pagina Home

Pagina Vehicles

Alla pagina Vehicles è possibile visualizzare i vari veicoli caricati e selezionarne uno per accedere alla pagina dedicata al singolo veicolo con maggiori dettagli. Inoltre è possibile inserire un nuovo veicolo.

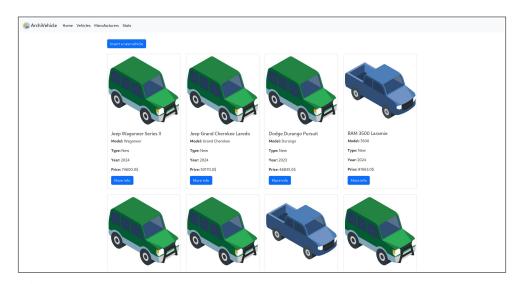


Figure 7: Pagina Vehicles

Pagina Insert Vehicle

Alla pagina Insert Vehicle è possibile inserire le informazioni di un nuovo veicolo. I campi contrassegnati con un asterisco rosso sono considerati obbligatori da inserire, in quanto definiscono le informazioni essenziali del veicolo. Attraverso le due dropdown list è possibile inserire anche le informazioni relative ai dettagli meccanici e ai dettagli della carrozzeria, seguendo la modellazione dello schema dati del documento del veicolo.

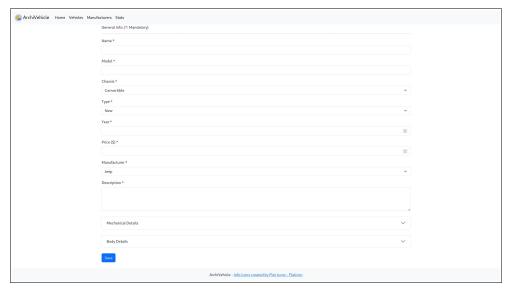


Figure 8: Pagina Insert Vehicle

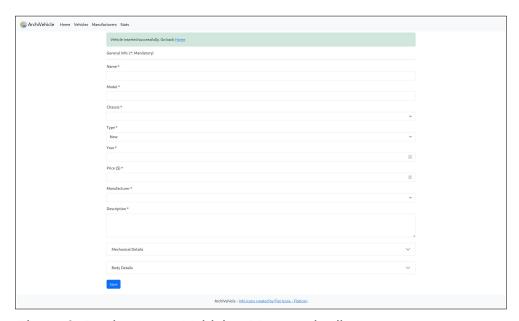


Figure 9: Pagina Insert Vehicle - Messaggio di successo

Pagina Show Single Vehicle

Alla pagina Show Single Vehicle è possibile visualizzare tutte le informazioni del veicolo. Inoltre è possibile accedere alla pagina Update Vehicle per aggiornare le informazioni del veicolo ed effettuare la cancellazione di un veicolo in seguito alla conferma di un messaggio alert.

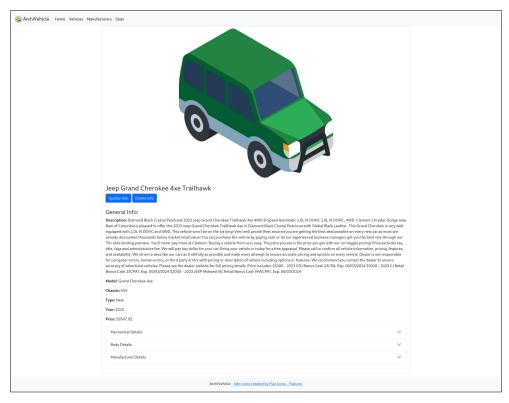


Figure 10: Pagina Show Single Vehicle

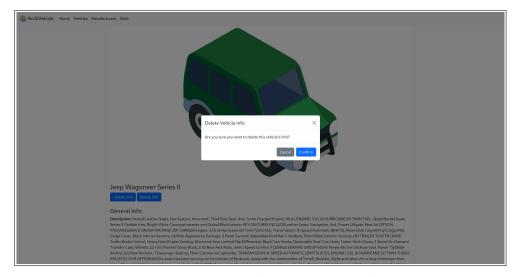


Figure 11: Pagina Show Single Vehicle - Messaggio di conferma

Pagina Manufacturers

Alla pagina Manufacturers è possibile visualizzare i vari produttori caricati. Inoltre è possibile inserire un nuovo produttore, aggiornare le informazioni di un produttore già esistente oppure cancellarne uno già esistente.

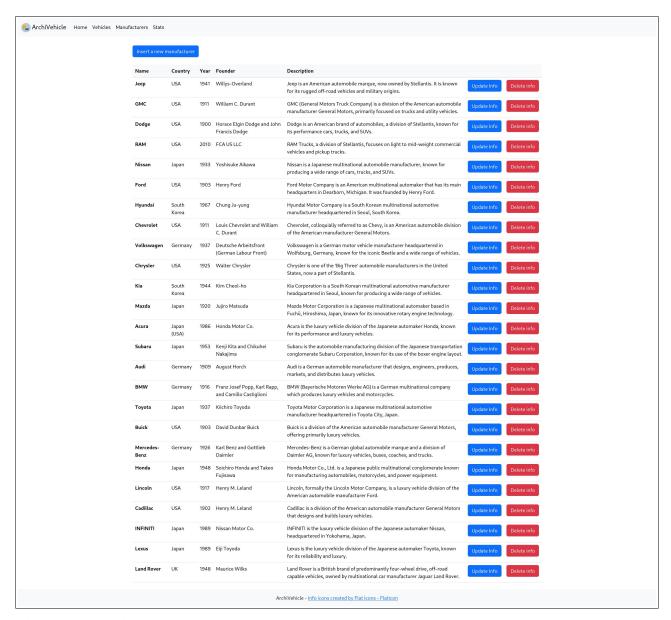


Figure 12: Pagina Manufacturers

Pagina Stats

Alla pagina Stats è possibile visualizzare il numero complessivo di veicoli per produttore e il conteggio dei veicoli che presentano sulla base di varie caratteristiche.

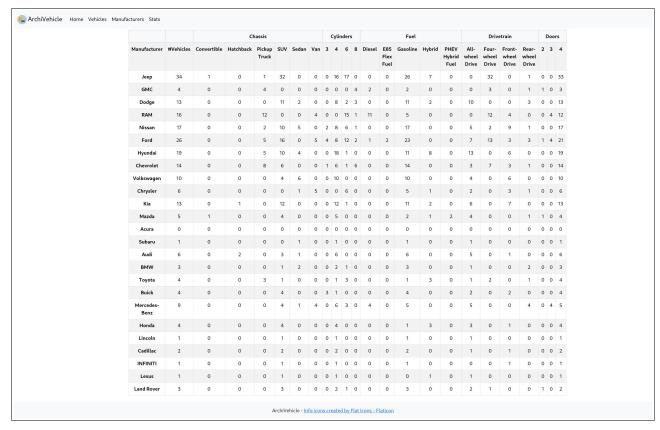


Figure 13: Pagina Stats

Installazione di ArchiVehicle

Per installare e utilizzare ArchiVehicle è sufficiente seguire questi passaggi:

- 1. Importare tramite Compass le due collezioni di documenti:
 - archiVehicle.manufacturer.json
 - archiVehicle.vehicle.json
- 2. Posizionarsi nella cartella root del progetto
- 3. Creare un virtual environment con il seguente comando:
 - python -m venv .venv
- 4. Attivare il virtual environment
- 5. Installare le dipendenze necessarie con il seguente comando:
 - pip install flask pymongo
- 6. Lanciare l'esecuzione del server flask con il seguente comando:
 - flask run