ArchiVehicle





Database II Class

Project Documentation

Superivsors: Prof.ssa Genoveffa Tortora Prof. Luigi Di Biasi Author: 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	

Introduction

This documentation describes the project realized for the Database II Class. In this document you can find more details about:

- Requirements of the project;
- Dataset choice:
- Dataset analysis and cleaning;
- DBMS choice;
- Data Scheme creation;
- · Uploading data to the database;
- Technologies choice for the webapp;
- Description of the page realized and the possible actions;
- Webapp installation.

Description of the system

ArchiVehicle has the goal to simplify and digitalize the the information management about a vehicle park of a car dealer. This software is intended for the userd of a business that have to retrieve the information about the vehicles and memorize them into the system, thanks to a graphical user interface served by a client/server web application. ArchiVehicle offers the following functionalities:

- 1. Insert, update, delete the information about a vehicle;
- 2. Insert, update, delete the information about a manufacturer of vehicles;
- 3. Searching for a vehicle by applying some filters;
- 4. Quick view of the number of vehicles by each manufacturer, based on various characteristics by vehicle.

Dataset

Information about the initial dataset

Not having available information about vehicles, in order to simulate the use of this application, a public dataset from Kaggle has been picked. The choosed dataset is **Vehicle Dataset 2024.** For each column it has been defined a description about the showed information e the data type to be memorized. If necessary, it has been specified the possible insertable value of the picked column:

Column name	Description	Data type
name	Full name of the vehicle	String
description	A brief description of the vehicle	String
make	The vehicle's manufacturer name	String
model	The model name of the vehicle	String
type	The vehicle type	String: {"New", "Used"}
year	The year of production of the vehicle	Intero
price	The vehicle price	Decimal
engine	Info about the vehicle engine	String
cylinders	The number of cylinders in the vehicle engine	Decimal
fuel	The type of fuel used by the vehicle	String
mileage	The number of miles driven by the vehicle	Decimal
transmission	The type of vehicle transmission	String
trim	The level of vehicle accessories	String
body	The body style of the vehicle	String: {"Convertible", "Hatchback", "Pickup Truck", "SUV", "Sedan", "Cargo Van", "Passegner Van", "Minivan"}
doors	The number of vehicle doors	Decimal
exterior_color	The exterior color of the vehicle	String
interior_color	The interior color of the vehicle	String
drivetrain	The type of vehicle traction	String: {"All-wheel Drive", "Four-wheel Drive", "Front-wheel Drive", "Rear-wheel Drive"}

Generation of vehicle manufacturer dataset

Now having all the information about the vehicles, but still missing the ones about the manufacturers, it has been decided to use ChatGPT to request the generation of a dataset about the manufacturers. After extracting the name of all the manufacturers, with the techniques showed in the **Data Cleaning** section, the dataset has been generated and downloaded in a csy format. The conversation is showed next:

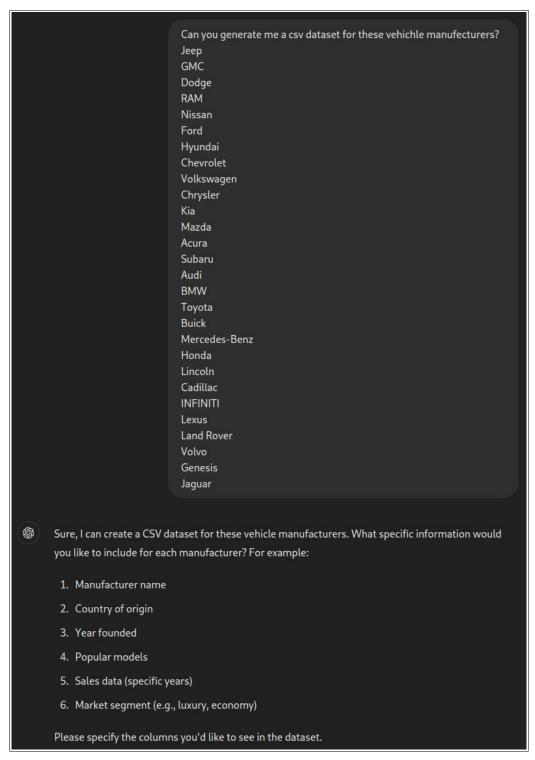


Figure 1: Generation of vehicle manufacturer dataset - Part 1

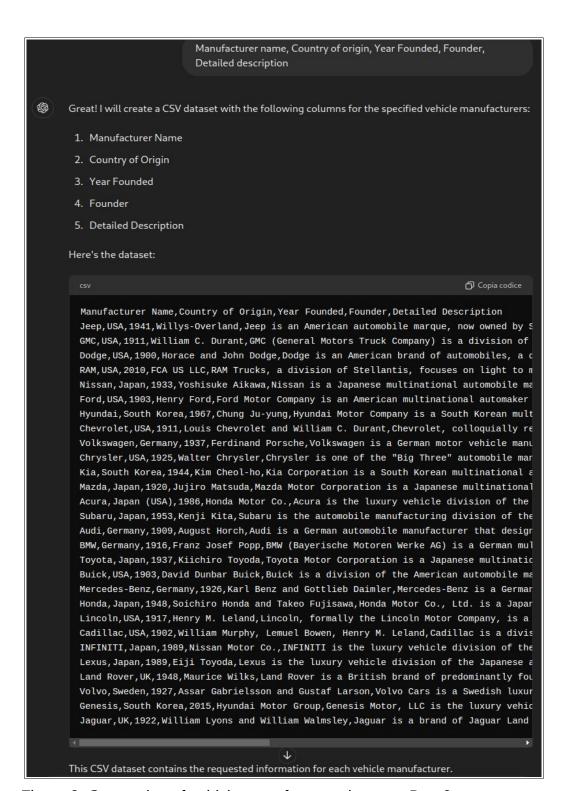


Figure 2: Generation of vehicle manufacturer dataset - Part 2



Figure 3: Generation of vehicle manufacturer dataset - Part 3

The result dataset has the following columns:

Nome colonna	Descrizione	Tipo di dato
name	The manufacturer name	String
country	The country where the production company was founded	String
year	The year the production company was founded	Integer
founder	The name of the founder of the production company	String
description	A brief description of the production company	String

Data Cleaning

After obtaining all the dataset, before is possibile to load them on the db, some data cleaning operations have been necessary. For this reason, two notebooks has been realized, using also the pandas library. The main operation are showed next:

vehicles_dataset_manufacturer_extract.ipynb:

- 1. Extract the make column that contains the manufacturers name;
- 2. Remove the duplicates from the extracted column;
- Save the column on a csy file.

vehicles_dataset_cleaning.ipynb:

- 1. Remove the production year from the elements of the name column;
- 2. Remove the rows whit null/empty values;
- 3. Remove extra newline/spacing characters from the element of description column;
- 4. Change the type of cylinders and doors column from float to int;
- 5. Retrieve the indexes of the rows with meaningless numerical values (0 or 0.0) and the remove these rows;
- 6. Retrieve the indexes of the rows where the description is under 100 characters length and remove these rows;
- 7. Remove the duplicated rows by the name column;
- 8. Change the name of body column to chassis;
- 9. Simplify all the chassis Van types, grouping them up;
- Create a new column image wiht default value "img";
- 11. Change the make column name to manufacturer_id and change the manufacturers name with the respective id in the database;
- 12. Change the name of some columns e move them, in order to correctly import the dataset into the database;
- 13. Save the cleaned dataset on a csv file.

Database

Database Choose

In order to manage the persistent data of this application, it has been decided to use MongoDB and there are several reason:

- 1. Data structure: thanks to the "flexibility" of a NoSQL databse, is possibile to save the partial info of an entity in a less rigid way. In this case, the car dealer, could have received only the essential vehicles information from the manufacturers and still starts to save them, so they can be later integrated with the missing ones.
- **2. Scalability:** a NoSQL database offers an increased expandability, so it's easier to distribute the data on different nodes on the network. In this case, the car dealer could have multiple shops in the country, having the necessity to use the application on more machines and to define a distribute solution for the database.

Data Scheme

Even if the database been used is a NoSQL, for each collection has been defined a document structure in order to define the data scheme.

Data structure of the manufacturers collection

For each document of the manufacturers collection, since there are a few attributes the structure is on a single level, as showed in the next example:

Data structure of the vehicles collection

}

For each document of the vehicles collection, since there are a lot of attributes, the structure is on multiple levels with subdocuments. Considering the choices took in the **Database Choose** section, an example of a document is showed next:

```
{
   "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"
   }
```

Datasets import

After completing all the operation about generating, transforming, cleaning the datasets and defined the structure of the documents, the datasets have been imported into the database, thankns to the DBMS GUI, Compass. In order to correctly create the documents, it has been first imported the manufacturers dataset, so in this way is possible to define the reference between the vehicle document and the respective manufacturer document. In the next figures, this process is showed:

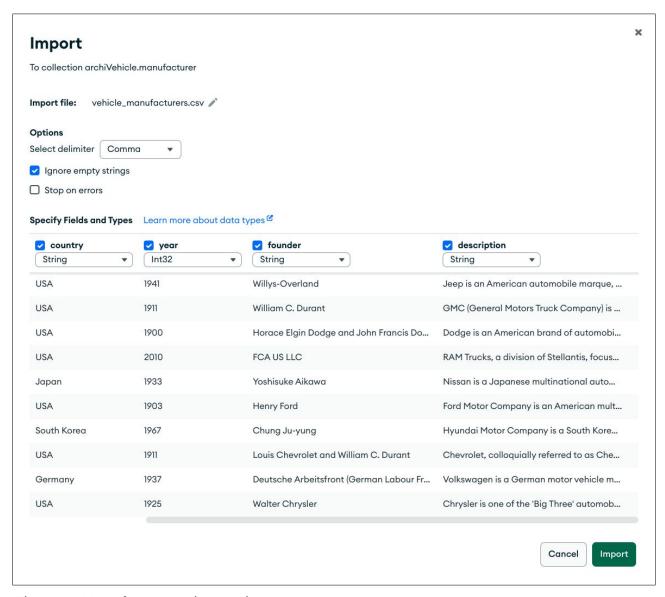


Figure 4: Manufacturers dataset import

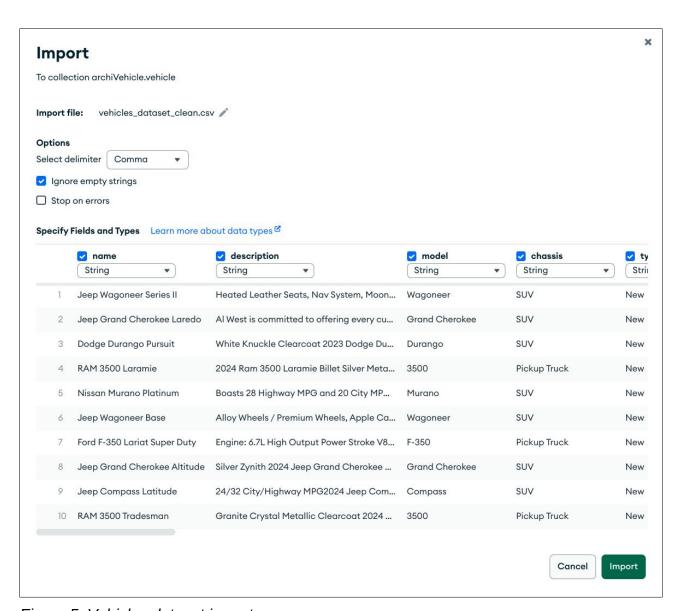


Figure 5: Vehicles dataset import

Web Application

Technologies choose

Having short times to realize this application, it has been decided to use technologies with a low learning curve and writing not too much code, using also already defined components where is possible to do it. For these reasons, the picked technologies are:

Database Driver: PyMongo

Application logic / back-end: Flask

Presentation logic / front-end: Bootsrap

PyMongo

PyMongo is the reccomended driver by the MongoDB developers, in order to exploit the potential of their database within a web application. PyMongo offers the main operations to interact with the database. Once the package has been installed, the first step is to import the package and start the connection with the database:

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

Flask

Flask is a micro-framework created for developing web-application in a fast and easy way. Being a micro-framework, there are few functionalities out of the box, such as the server, the debugger, the router, the component for processing requests and responses. Once the package has been installed, the first step is to create a new module named app.py and define the first route for the index page, that will only initially gave a response with a classic "Hello World!" message:

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

Bootstrap

Bootstrap is a framework created for realizing responsive web pages, offering templates for each necessary component. Thanks to the import html component functionality offered by Flask, is possible to reuse the different components on the web application pages. In order to start use Bootstrap, it's necessary to include the repsective tags inside the page header.

Descrizione delle pagine

Home Page

In the Home Page, is possible to do a search of the vehicles and apply filters related to the property of the vehicles.

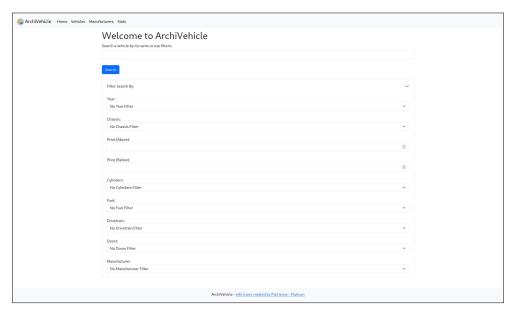


Figure 6: Home Page

Vehicles Page

In the Vehicles Page, is possible to display all the various vehicles and pick one in order to go to the dedicated vehicle page where more details are showed. It is also possible to insert a new vehicle.

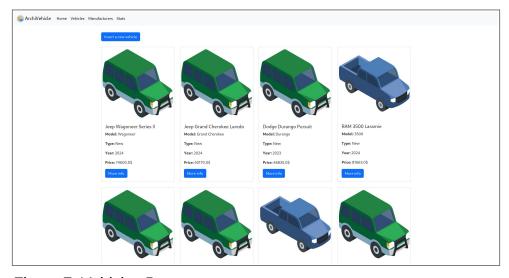


Figure 7: Vehicles Page

Insert Vehicle Page

In the Insert Vehicle Page, is possile to insert the info about a new vehicle. The fields marked with a red asterisk are mandatory because they define the main info about the vehicle. Through the two dropdown lists is possible to insert the info related to the details of the vehicle, following the document structure and the data scheme for the vehicle.

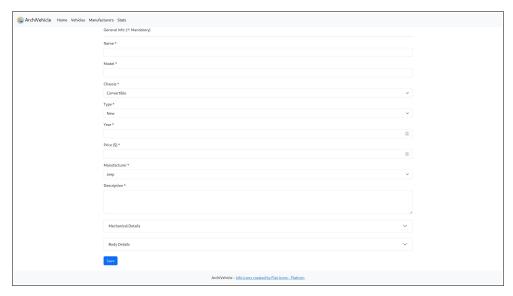


Figure 8: Insert Vehicle Page

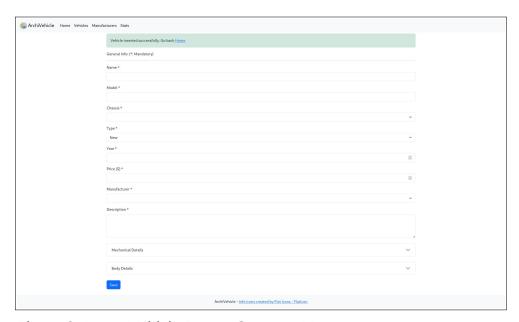


Figure 9: Insert Vehicle Page - Success Message

Show Single Vehicle Page

In the Show Single Vehicle Page, is possible to display all the information about the vehicle. Also, is possible to go to the Update Vehicle Page for updating the info about the vehicle and to delete the info about the vehicle after conferming the alert message.

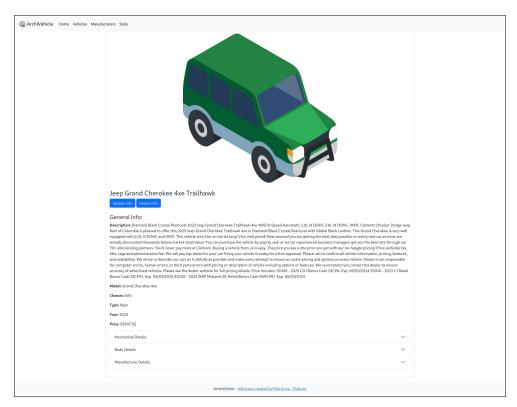


Figure 10: Show Single Vehicle Page

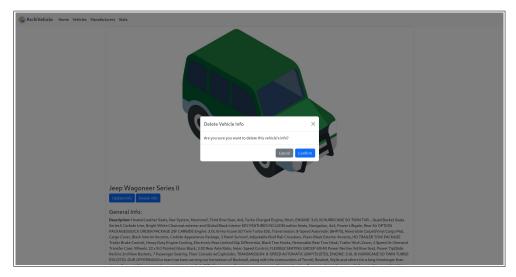


Figure 11: Show Single Vehicle Page - Confirm Message

Manufacturers Page

In the Manufacturers Page, is possile to display the info about all the loaded manufacturers. Also, is possible to insert a new manufacturer, update the info of an already existent manufacturer or delete an existing one.

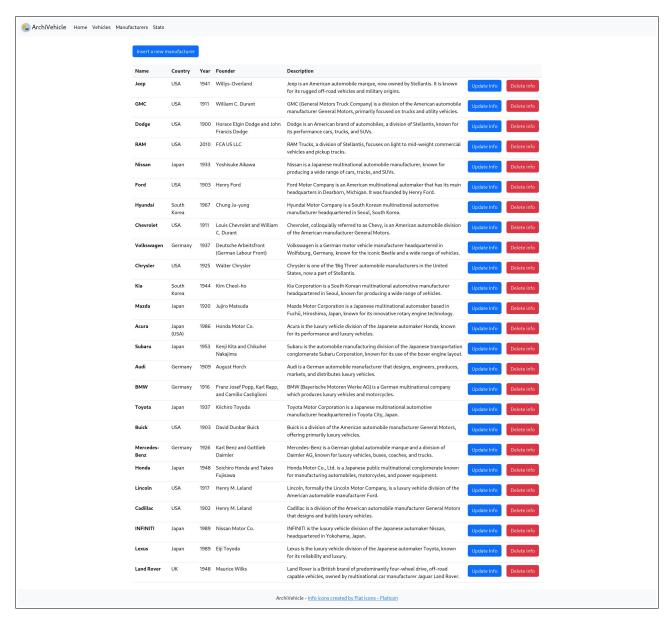


Figure 12: Manufacturers Page

Stats Page

In the Stats Page, is possible to display the number of all the vehicles for the respective manufacturer, counting the vehicles given the different features.

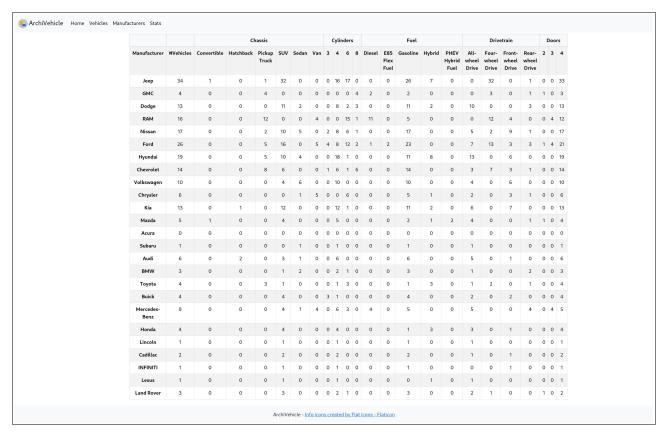


Figure 13: Stats Page

ArchiVehicle Installation

In order to install and use ArchiVehicle you need to follow the steps:

- 1. Import through MongoDB Compass these collections of documents:
 - archiVehicle.manufacturer.json
 - archiVehicle.yehicle.json
- 2. Change the current directory to the root of the project;
- 3. Create a virtual environment with the command:
 - python -m venv .venv
- 4. Activate the virtual environment;
- 5. Install the needed dependencies with the command:
 - pip install flask pymongo
- 6. Execute the flask server with the command:
 - flask run