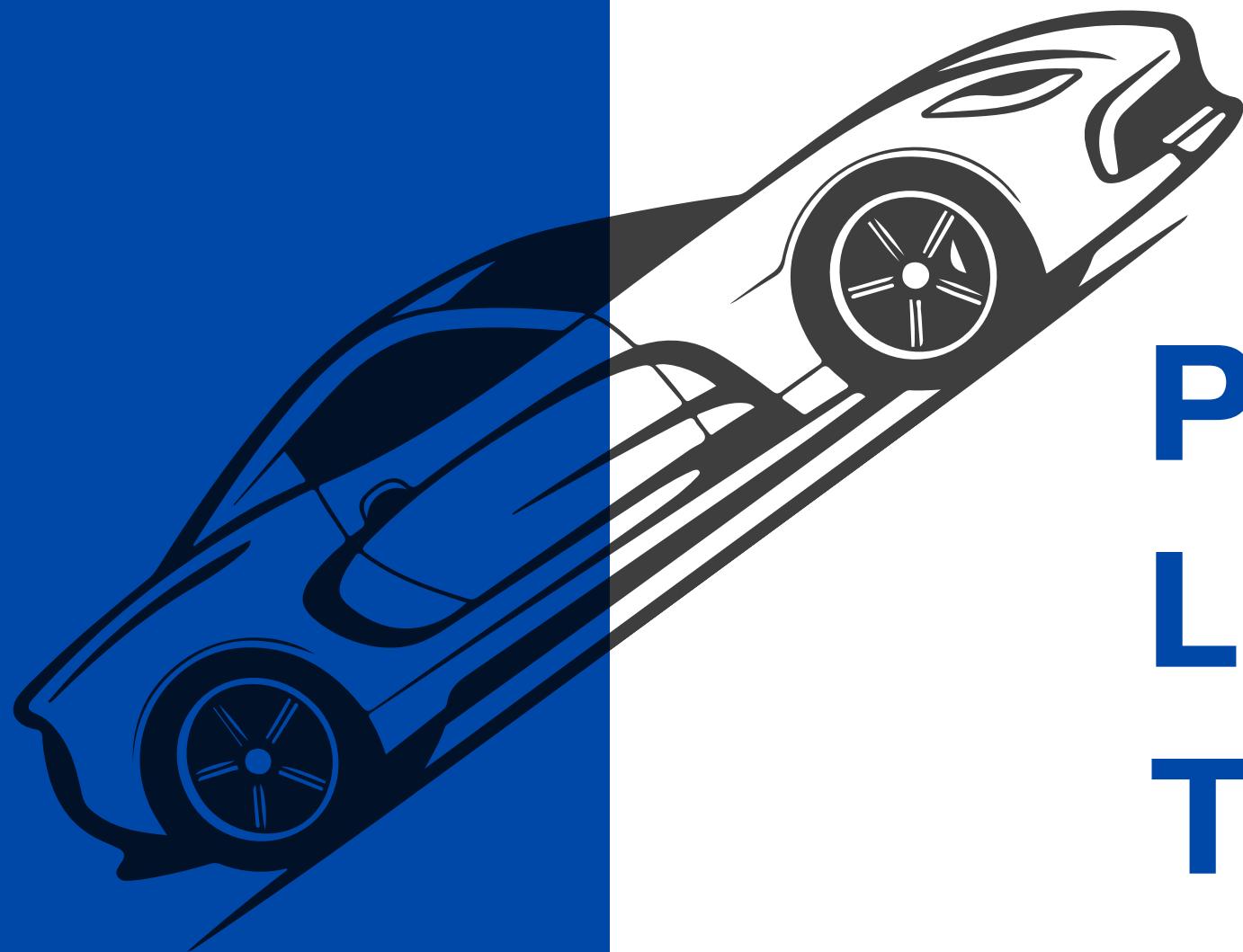




# BÁO CÁO ĐỒ ÁN THỰC HÀNH

**Môn học: Nhập môn Khoa học Dữ liệu  
GVHD: Lê Nhựt Nam**



# PHÂN TÍCH DỮ LIỆU XE HƠI BÁN TRÊN CARVAGO

Thực hiện quy trình khoa học dữ liệu và  
mô hình hóa dữ liệu với bộ dữ liệu liên quan  
đến xe hơi được lấy từ một website bán xe hơi

# NỘI DUNG

01

Giới thiệu

02

Thu thập dữ liệu

03

Tiền xử lý

04

EDA

05

Mô hình hóa dữ liệu



# NHÓM 19



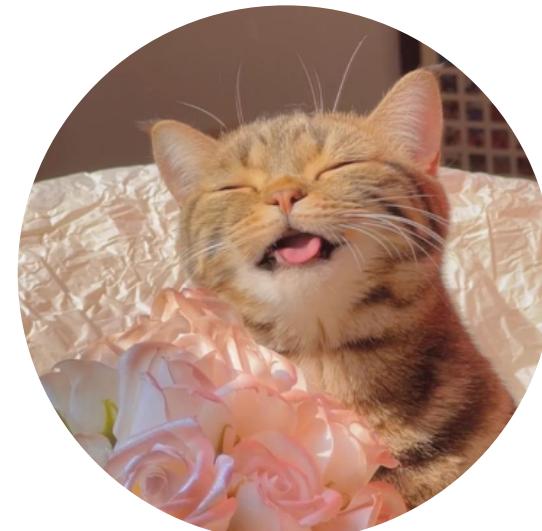
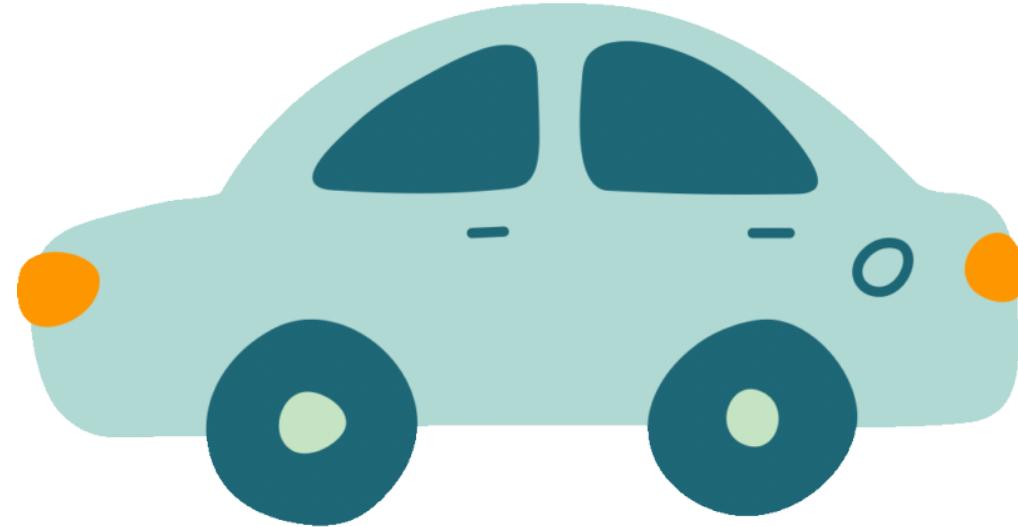
## Gồm các thành viên:



21120496 - Chu Hải Linh



21120574 - Nguyễn Minh Trí



21120580 - Trần Thị  
Kim Trinh



21120570 - Đặng Nguyễn  
Thanh Tín

01

# GIỚI THIỆU



Tổng quan đồ án



Lý do chọn đề tài



# Tổng quan đồ án



- Đồ án được thực hiện để **tổng hợp lại kiến thức và kĩ năng đã được học** ở môn “**Nhập môn Khoa học Dữ liệu**”.
- Mục tiêu của đồ án là giúp sinh viên xử lý những vấn đề thực tế bằng cách sử dụng mã nguồn dữ liệu đến từ Internet.

## Lý do chọn đề tài

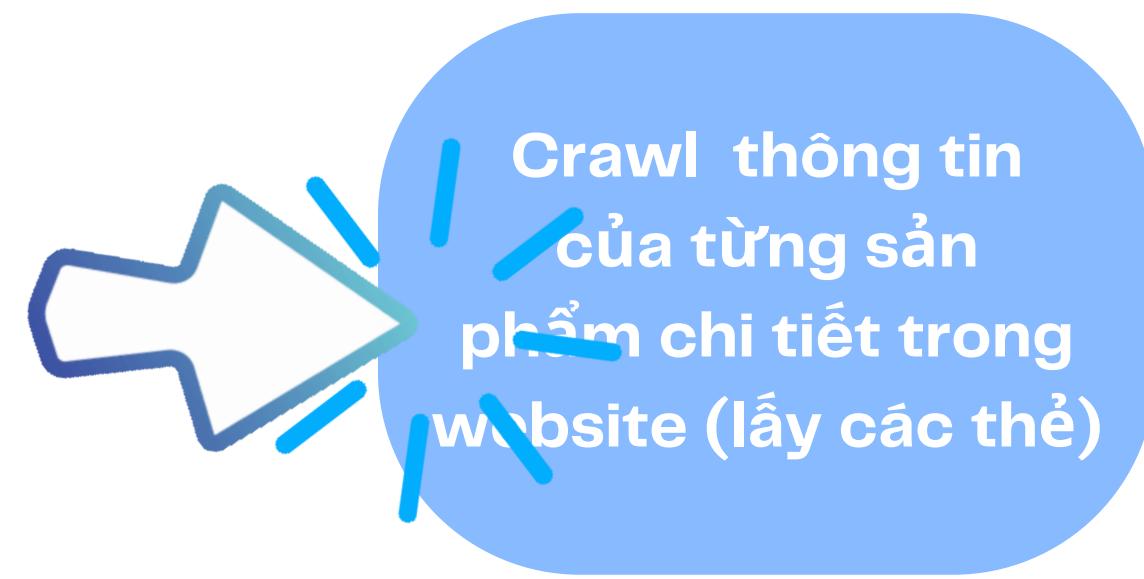
- Trong thời đại 4.0 ngày nay, việc mua bán trên mạng rất phổ biến, tuy nhiên nhóm cũng biết được, khách hàng vẫn **lo ngại về việc chọn xe phù hợp với nhu cầu và túi tiền.**
- Từ đó nhóm chúng em đã suy nghĩ nên việc ứng dụng phân tích dữ liệu vào việc mua – bán xe trên một website.
- Mục đích cuối cùng là **giúp cho khách hàng tìm được chiếc xe mà mình ưng ý**, cũng như **bán được xe cũ của mình với giá thích hợp**

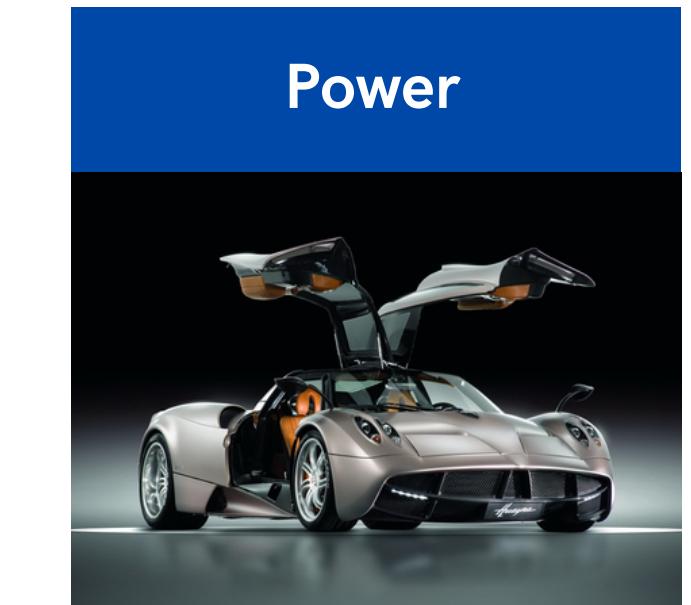
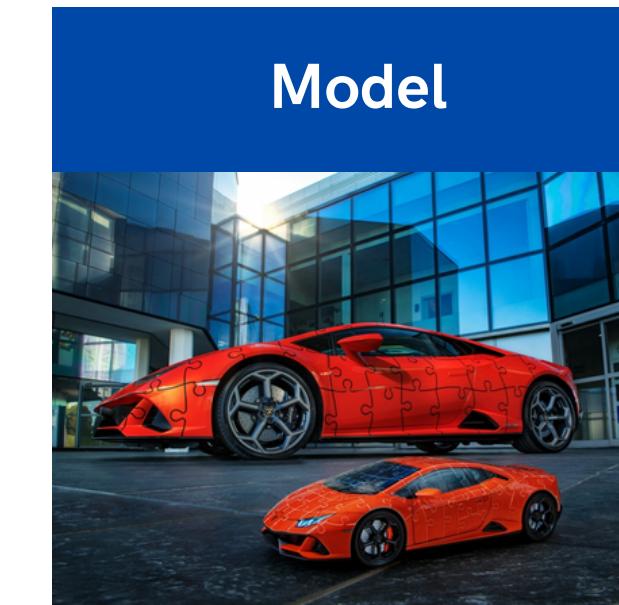
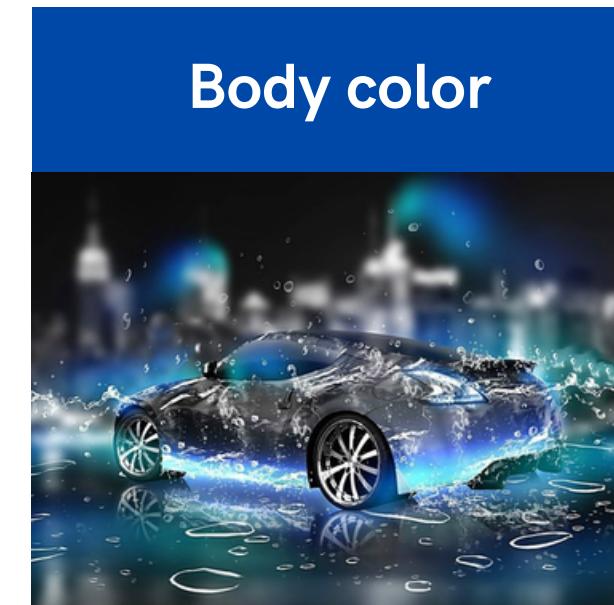
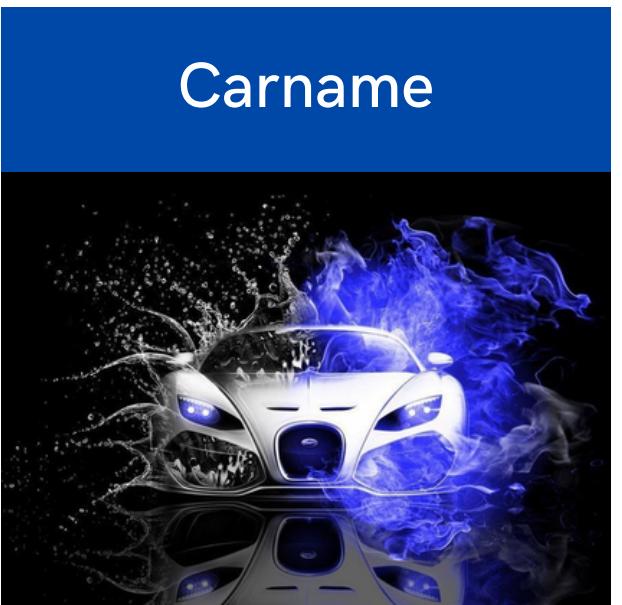


## 02

# Thu thập dữ liệu

- Tập dữ liệu của đồ án này được lấy từ trang web: <https://carvago.com/cars>
- Các bước thu nhập dữ liệu:





# 03 TIỀN XỬ LÍ



# Dữ liệu có bao nhiêu dòng và cột?

```
raw_df.shape  
✓ 0.0s  
(2293, 46)
```

Dữ liệu ta thu thập về có 2293 dòng và 46 cột



# Sơ lược về các cột trong dữ liệu

## Hyundai Kona Long Range 160 kW



53 099 €

43 883 € without 21% VAT

Delivery	Enter ZIP code
Extended warranty	Free
Location	Germany
Mileage	10 km
First registration	9/2023
Transmission	Automatic
Power	160 kW (218 hp)
Fuel	Electric
Drive type	4x2

Buy



Unnamed:	CARNAME	ID
0	hyundai kona	61032325

Price      Currency

53099      EUR

Tags

Ventilated front seats; Heated rear seats; Par...

Mileage	First registration	Transmission	Power	Fuel	Drive type
10 km	9/2023	Automatic	160 kW	Electric	4x2

## Details

VEHICLE DETAIL	
Make	Hyundai
Model	Kona
Body color	<input checked="" type="radio"/> Grey
Type of finish	Metallic
Interior color	<input checked="" type="radio"/> Black interior
Interior material	Full leather interior
Body	SUV / offroad
Doors	4/5 doors
Seats	5
VIN	not published by the seller

ENGINE	
Fuel	Electric
Transmission	Automatic
Drive type	4x2
Power ⓘ	160 kW
El. motor power	150 kW
Consumption	0 kWh/100km (combined)
CO2 emissions	0 g/km
Emission class	No emission class
Battery type	Lithium polymer (Li-pol)
AC charging time ⓘ	6 h

Make	Model	Body color	Type of finish	Interior color	Interior material	Body	Doors	Seats
Hyundai	Kona	Grey	Metallic	Black interior	Full leather interior	SUV / offroad	4/5 doors	5

**Weight    Country of origin    Secondary drive**

NaN NaN NaN

**Energy efficiency**   **Full service history**   **Battery capacity** 1

NaN NaN NaN

DC charging time	Battery warranty (km)	Range extender	Condition
------------------	-----------------------	----------------	-----------

40 min                  160000 km                  No                  New

VIN	El. motor power	CO2 emissions	Emission class	Battery type	AC charging time
by ller	150 kW	0 g/km	No emission class	Lithium polymer (Li-pol)	6 h

# Làm sạch dữ liệu



# Kiểm tra trùng lắp dòng

- Dữ liệu được cào từ **mỗi link của mỗi xe khác nhau**, nên khó có việc sẽ có dòng bị duplicate
- Tuy nhiên, nhóm sẽ **dùng hàm duplicate()** để kiểm tra lại xem có dòng nào bị **trùng lắp** hay không

```
1 https://carvago.com/car/61023290/volkswagen-t-cross-1-5-tsi-dsg-110-kw
2 https://carvago.com/car/61023272/hyundai-i10-1-0-trend-49-kw
3 https://carvago.com/car/61023275/opel-adam-1-4-s-110-kw
4 https://carvago.com/car/61023278/mercedes-benz-glc-400-d-4matic-243-kw
5 https://carvago.com/car/61023291/bmw-x3-xdrive-135-kw
6 https://carvago.com/car/61023277/cupra-born-150-kw
7 https://carvago.com/car/61023276/volkswagen-golf-gtd-dsg-147-kw
8 https://carvago.com/car/61023284/mazda-cx-5-175-awd-129-kw
9 https://carvago.com/car/61023286/audi-a4-35-tfsi-s-tronic-110-kw
10 https://carvago.com/car/61023289/citroen-jumpy-m-180-eat8-driver-130-kw
11 https://carvago.com/car/61023285/seat-arona-1-0-tsi-opf-style-81-kw
12 https://carvago.com/car/61023287/kia-picanto-isg-opf-62-kw
13 https://carvago.com/car/61023282/volkswagen-passat-variant-2-0-tdi-elegance-147-kw-dsg
14 https://carvago.com/car/61023281/jeep-compass-longitude-96-kw
15 https://carvago.com/car/61023271/audi-a4-1-4-tfsi-s-line-avant-110-kw
16 https://carvago.com/car/61023273/volkswagen-t5-transporter-4motion-103-kw
17 https://carvago.com/car/61023263/audi-a4-40-tfsi-s-line-avant-140-kw
18 https://carvago.com/car/61023261/fiat-500-e-42-kwh-87-kw
19 https://carvago.com/car/61023267/mini-cooper-cabrio-100-kw
20 https://carvago.com/car/61023265/fiat-500x-96-kw
21 https://carvago.com/car/61023259/bmw-116d-85-kw
22 https://carvago.com/car/61023268/subaru-solterra-kwh-bev-awd-160-kw
23 https://carvago.com/car/61023264/volkswagen-polo-1-0-tsi-70-kw
24 https://carvago.com/car/61023266/ram-1500-5-7-hemi-v8-etorque-4wd-295-kw
25 https://carvago.com/car/61023260/ford-focus-ecoboost-st-line-114-kw
26 https://carvago.com/car/61023270/peugeot-2008-96-kw
27 https://carvago.com/car/61023269/jeep-renegade-1-3-111-kw
28 https://carvago.com/car/61023262/bmw-118i-advantage-103-kw
29 https://carvago.com/car/61023253/volkswagen-golf-e-100-kw
30 https://carvago.com/car/61023249/ford-transit-custom-320-11-va-trend-77-kw
31 https://carvago.com/car/61023250/peugeot-208-75-active-55-kw
32 https://carvago.com/car/61023258/volkswagen-golf-1-4-150-kw-dsg
```

## Kiểm tra kiểu dữ liệu

- Sau khi crawl xong, dữ liệu có các kiểu: int, object
  - Có rất nhiều giá trị Nan, một số cột có missing value khá cao
- ==> Ta tiến hành xóa các cột có tỉ lệ missing value cao, các cột không cần thiết

6	0	Unnamed: 0	2293 non-null	int64
7	1	CARNAME	2293 non-null	object
8	2	ID	2293 non-null	int64
9	3	Make	2293 non-null	object
10	4	Model	2293 non-null	object
11	5	Body color	2111 non-null	object
12	6	Type of finish	1204 non-null	object
13	7	Interior color	1894 non-null	object
14	8	Interior material	1930 non-null	object
15	9	Body	2293 non-null	object
16	10	Doors	2272 non-null	object
17	11	Seats	2080 non-null	object
18	12	VIN	2293 non-null	object
19	13	Fuel	2293 non-null	object
20	14	Transmission	2293 non-null	object
21	15	Drive type	2293 non-null	object
22	16	Power	2293 non-null	object
23	17	El. motor power	118 non-null	object
24	18	CO2 emissions	2293 non-null	object
25	19	Emission class	2145 non-null	object
26	20	Battery type	146 non-null	object
27	21	AC charging time	105 non-null	object
28	22	DC charging time	53 non-null	object
29	23	Battery warranty (km)	99 non-null	object
30	24	Range extender	316 non-null	object
31	25	Mileage	2293 non-null	object
32	26	First registration	2293 non-null	object
33	27	Condition	1376 non-null	object
34	28	Consumption	2025 non-null	object
35	29	Price	2293 non-null	int64
36	30	Currency	2293 non-null	object
37	31	Tags	2233 non-null	object
38	32	Engine capacity	2042 non-null	object
39	33	Valid MOT until	266 non-null	object
40	34	Previous owners	932 non-null	float64
41	35	Engine power	52 non-null	object
42	36	Battery capacity	177 non-null	object
43	37	Hybrid type	169 non-null	object
44	38	Electric range	28 non-null	object
45	39	Warranty until	312 non-null	object
46	40	Weight	564 non-null	float64
47	41	Country of origin	279 non-null	object
48	42	Secondary drive	10 non-null	object
49	43	Front wheel drive	5 non-null	object

## Loại bỏ các cột có missing cao và các cột không dùng

Unnamed: 0	Currency
0	EUR
1	EUR
2	EUR
3	EUR
4	EUR
5	EUR
6	EUR
7	EUR
8	EUR
9	EUR

'El. motor power': 94.85390318360227%  
'Battery type': 93.63279546445705%  
'AC charging time': 95.42084605320541%  
'DC charging time': 97.68861753161798%  
'Battery warranty (km)': 95.68251199302225%  
'Range extender': 86.21892716964675%  
'Valid MOT until': 88.39947666812037%  
'Engine power': 97.73222852158744%  
'Battery capacity': 92.2808547754034%  
'Hybrid type': 92.62974269515918%  
'Electric range': 98.77889228085478%  
'Warranty until': 86.39337112952464%  
'Weight': 75.40340165721761%  
'Country of origin': 87.83253379851723%  
'Secondary drive': 99.56389010030527%  
'Energy efficiency': 99.78194505015264%  
'Full service history': 99.65111208024422%  
'Battery capacity1': 99.91277802006105%

Xoá các cột không cần thiết

Loại bỏ các cột có tỉ lệ missing cao

## Tiến hành chuyển đổi kiểu dữ liệu cho các cột

Engine capacity: object

Power: object

Consumption: object

Mileage: object

CO2 emissions: object



```
'Engine capacity(ccm)': float64  
'Price(EUR)': int64  
'Power(kW)': int64  
'Mileage(km)': int64  
'Consumption(1/100km or kWh/100km)': float64  
'CO2 emissions(g/km)': int64
```

# Xử lý Outlier

- Do thị trường xe phức tạp và sự chênh lệch về giá cũng như sức mạnh giữa các xe, **hầu hết các outlier đều là dữ liệu thật.**
- Các **outlier do lỗi nhập liệu** sẽ thay thế bằng **giá trị trung bình của cột.**

## Toyota Land Cruiser D-4D 150 kW

Ventilated front seats  Air suspension  Keyless entry  Electric adjustable front seats  
 Heated front seats  Apple CarPlay  Android auto  Keyless ignition  Panoramic roof  
 Voice control



### ENGINE

Fuel Diesel

Transmission Automatic

Drive type 4x4

Power 150 kW

Engine capacity 280 ccm

Consumption 7.6 l/100km (combined)

CO2 emissions 206 g/km

Emission class Euro 6

Ví dụ outlier và lỗi nhập liệu

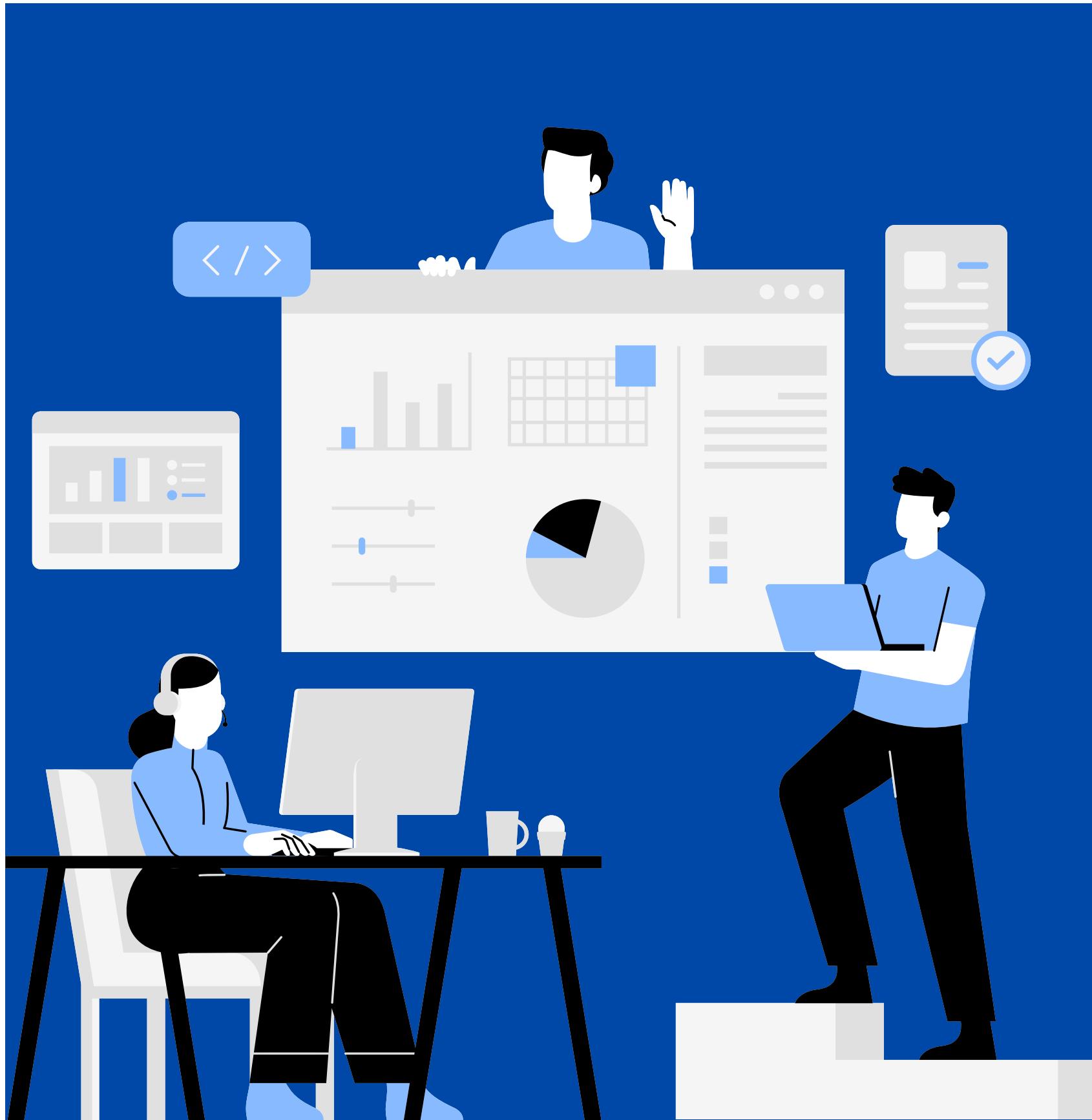
# EDA

04

Câu hỏi để khám phá mối quan hệ giữa các biến

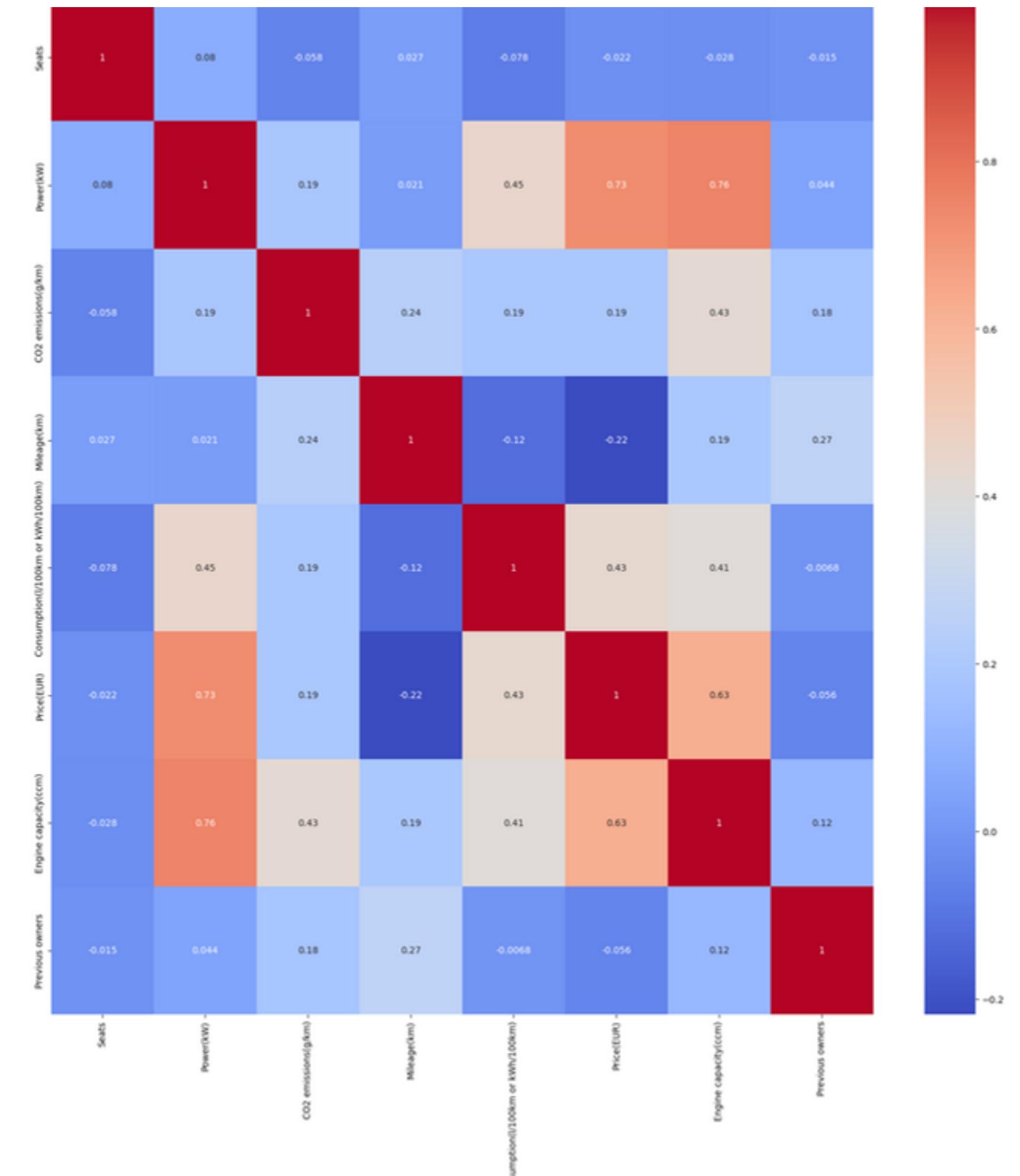
Câu hỏi để khai thác dữ liệu phục vụ người dùng

# **Khám phá mối quan hệ giữa các biến**



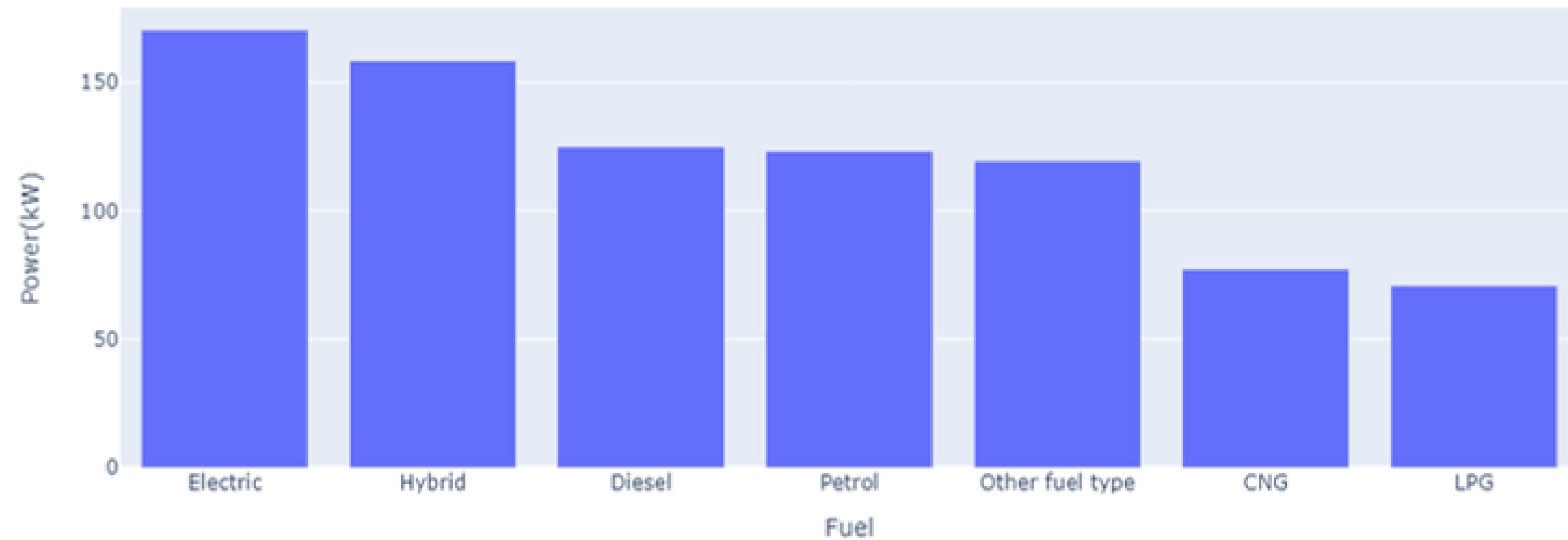
# Sự tương quan giữa các các cột

- Mỗi quan hệ tương quan dương mạnh mẽ nhất có thể kể đến là Engine capacity(ccm) với các biến còn lại
- Ta cũng nhận thấy Seats có mối tương hệ kém nhất, hầu như là mối quan hệ tương quan âm



# Câu hỏi 1: Với mỗi nhóm nhiên liệu, công suất trung bình của xe đối với từng nhóm là bao nhiêu?

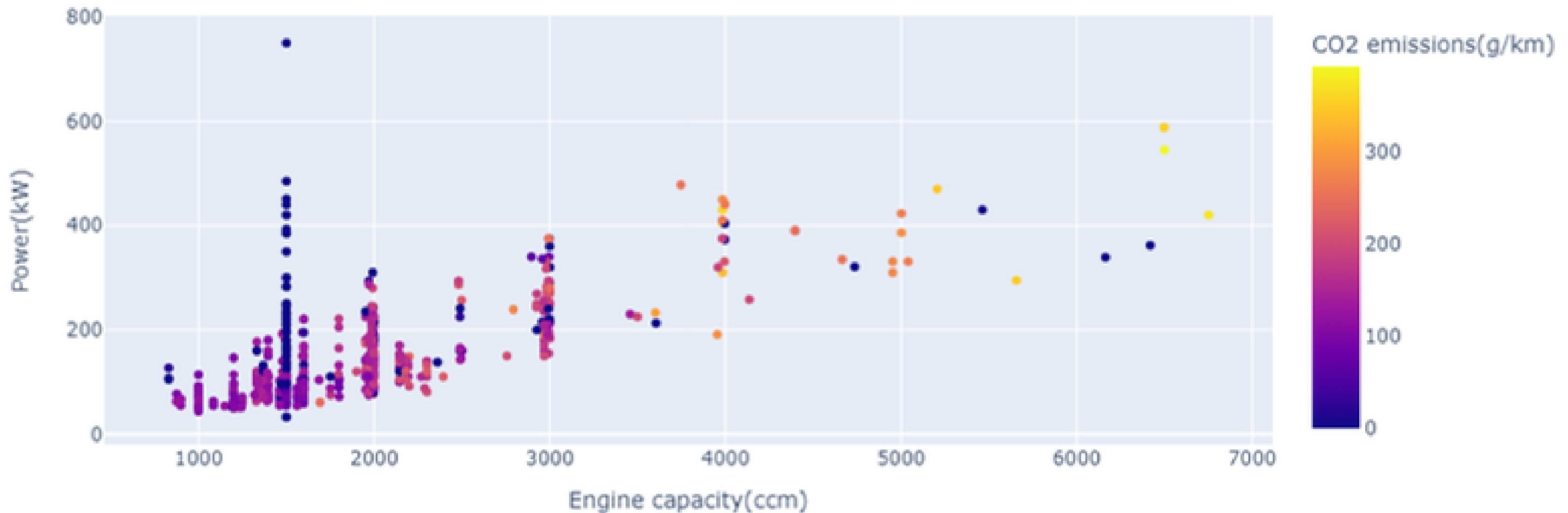
Average Power by Fuel Type



Biểu đồ cột thể hiện mối quan hệ giữa cột “Fuel” và “Power”

## Câu hỏi 2: Mối tương quan giữa dung tích máy và công suất động cơ với khí thải CO2 như thế nào?

Relationship between Engine Capacity, Power, and CO2 Emissions



Biểu đồ scatter thể hiện mối quan hệ giữa cột “Engine Capacity (ccm)”, “Power” và “CO2 Emissions”



# Khai thác dữ liệu để phục vụ người dùng

Các câu hỏi này được đặt ra sau khi chúng ta đã tìm hiểu được các mối liên hệ giữa các biến với nhau.

Các câu hỏi được dùng để phục vụ các vấn đề mà người dùng gặp phải.

## Câu 1

với mỗi phân khúc giá , sự khác biệt về công suất giữa xe sử dụng điện, hybrid và xe sử dụng các nhiên liệu truyền thống là như thế nào?

### Lợi ích của câu hỏi:

Hỗ trợ người dùng

- so sánh công suất động cơ của ba loại phương tiện **cùng tầm giá**
- tìm kiếm phương tiện phù hợp với công suất và nhu cầu mong muốn.

Xử lý câu hỏi

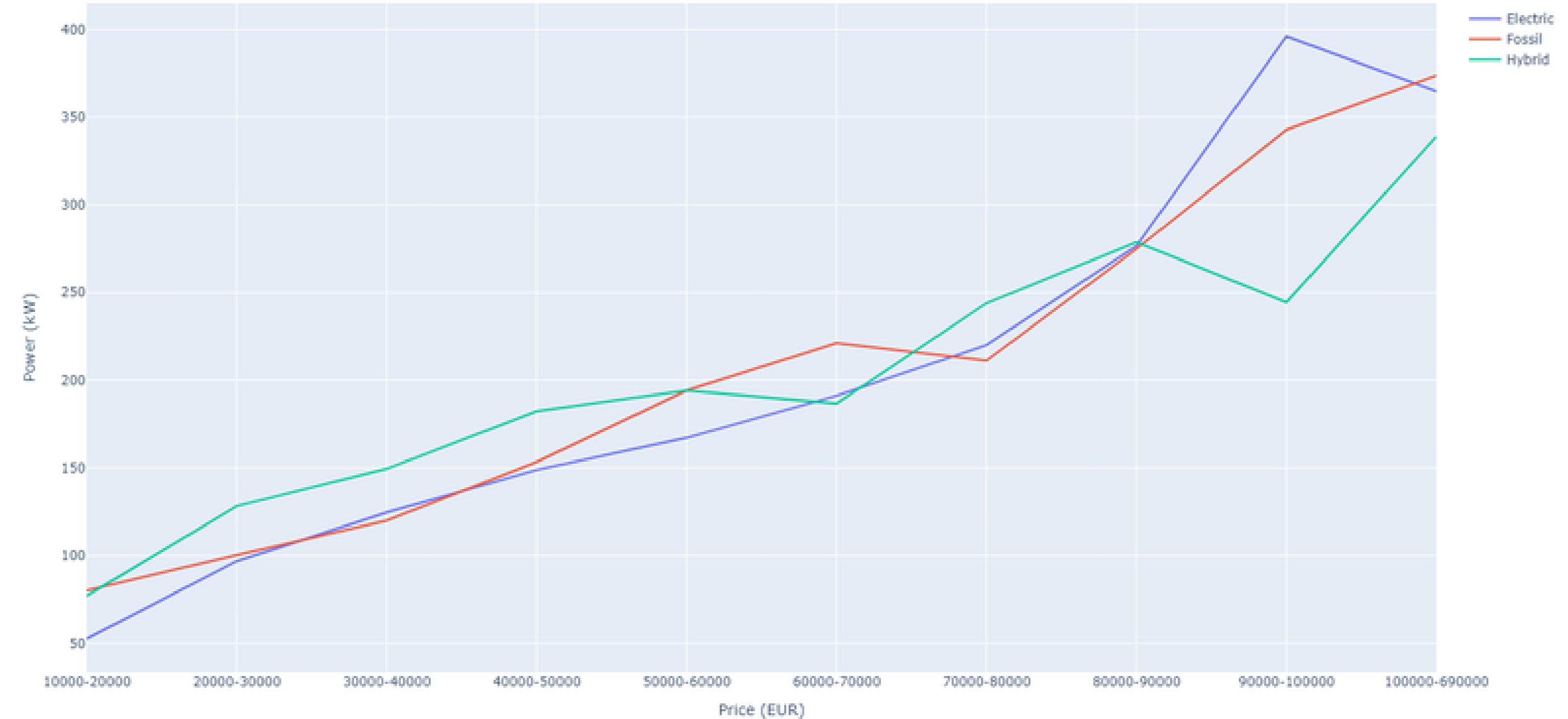
### Cách trả lời câu hỏi:

#### • Bước 1: Tiền xử lý

- Chỉ tập trung vào nhiên liệu **Electric, Hybrid và Fossil Fuel**, loại bỏ ô tô ‘Other fuel type’.
- Chuyển ‘Petrol’, ‘Diesel’, ‘LPG’, ‘CNG’ thành “Fossil”.
- **Loại bỏ giá trị ngoài khoảng tin cậy 95%** để giữ thông tin quan trọng.

## Bước 2: Minh họa

AVG Power for each type of Vehicle in each Price range



Biểu đồ đường thể hiện “Công suất trung bình của mỗi loại phương tiện trong từng phân khúc giá”

## Câu 2

### Phân tích 6 tính năng đặc biệt phổ biến nhất và giá cả trung bình của xe có trang bị tính năng đó.

#### Lợi ích của câu hỏi:

- Hỗ trợ xác định **tính năng hiện đại cần thiết** cho xe ô tô
- Tìm hiểu giá trung bình của các xe có chức năng này, giúp người dùng đưa ra quyết định mua xe phù hợp với kinh tế

Xử lý câu hỏi

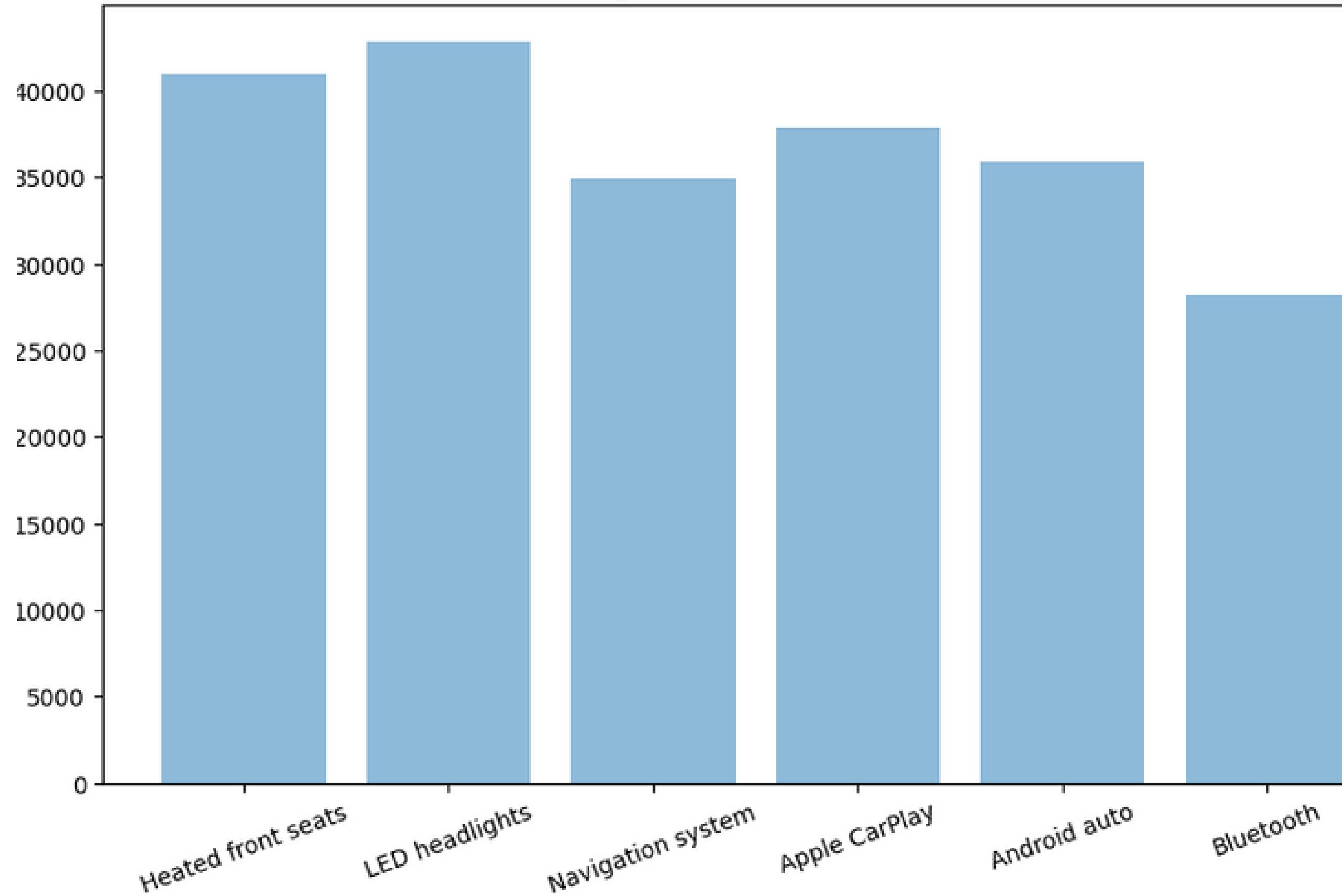
#### ■ Cách trả lời câu hỏi bằng cách sử dụng dữ liệu:

##### • **Bước 1: Tiền xử lý**

- Từ cột Tags, tách và mã hóa one-hot cho 6 tags phổ biến.
- Tính giá trung bình cho xe có các tính năng đó.

## Bước 2: Minh họa

Average Price for each Function



Bar chart thể hiện “6 chức năng đặc biệt phổ biến nhất  
và tác động của chúng lên giá trị trung bình”

## Câu 3

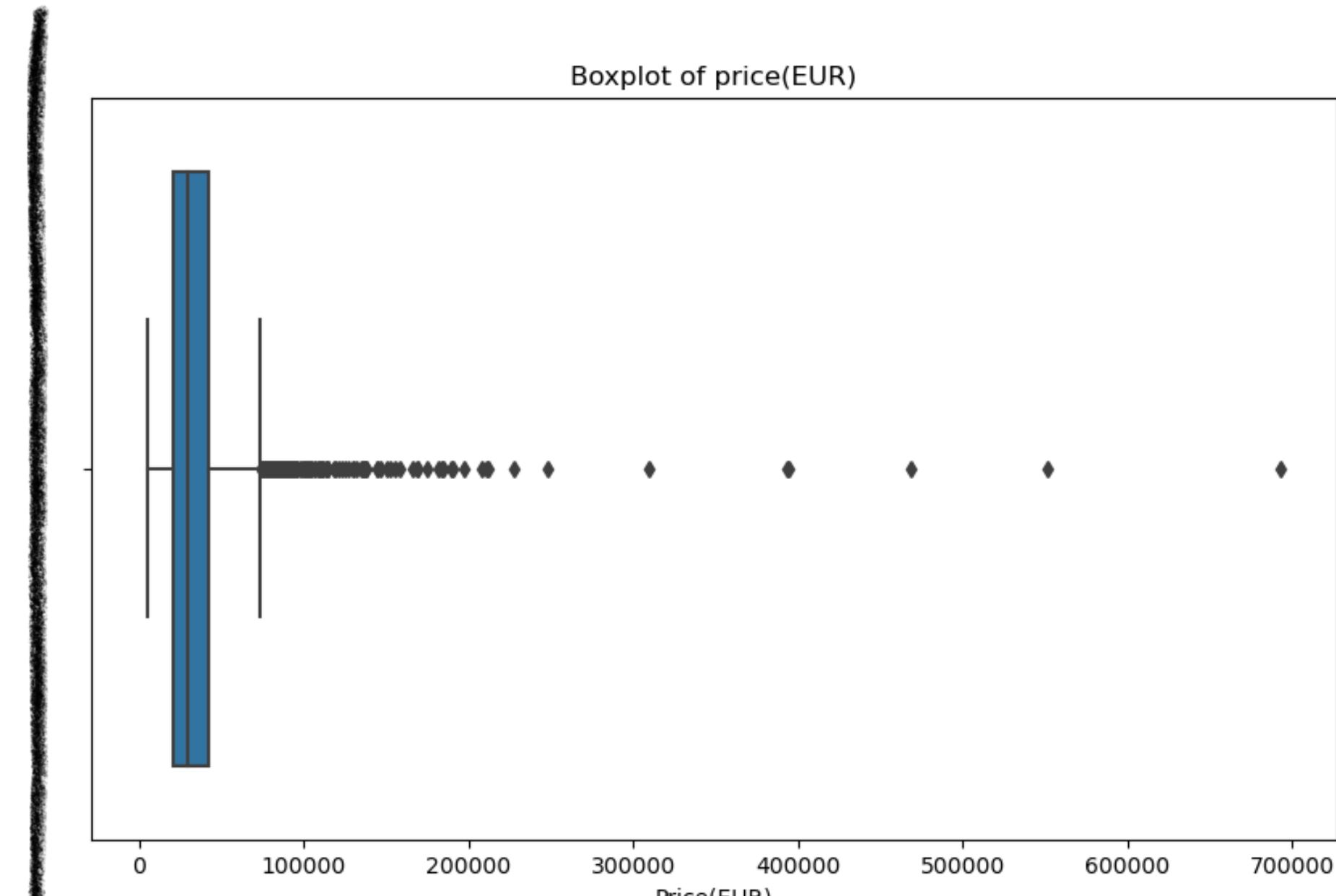
# Có sự khác biệt đáng kể về hiệu suất giữa các xe ở các phân khúc giá khác nhau không?

Ý tưởng về câu hỏi:

- Các mức giá xe ô tô khác nhau **đáng kể**. 75% số xe có giá dưới 42 nghìn EUR, còn 25% còn lại có xe lên đến 600 nghìn EUR.
- Điều này không phải lỗi dữ liệu mà là **đặc điểm của thị trường ô tô**, nơi có sự biến động rất lớn giữa các thương hiệu và mẫu xe.

Lợi ích của câu hỏi:

Giúp người dùng trả lời “**Liệu chất lượng có đi đôi với giá tiền?**”



Boxplot để xem mức giá khác nhau của các chiếc xe

## Xử lý câu hỏi

Sử dụng biểu đồ boxplot để trực quan hóa phân phối của các xe theo giá của chúng.

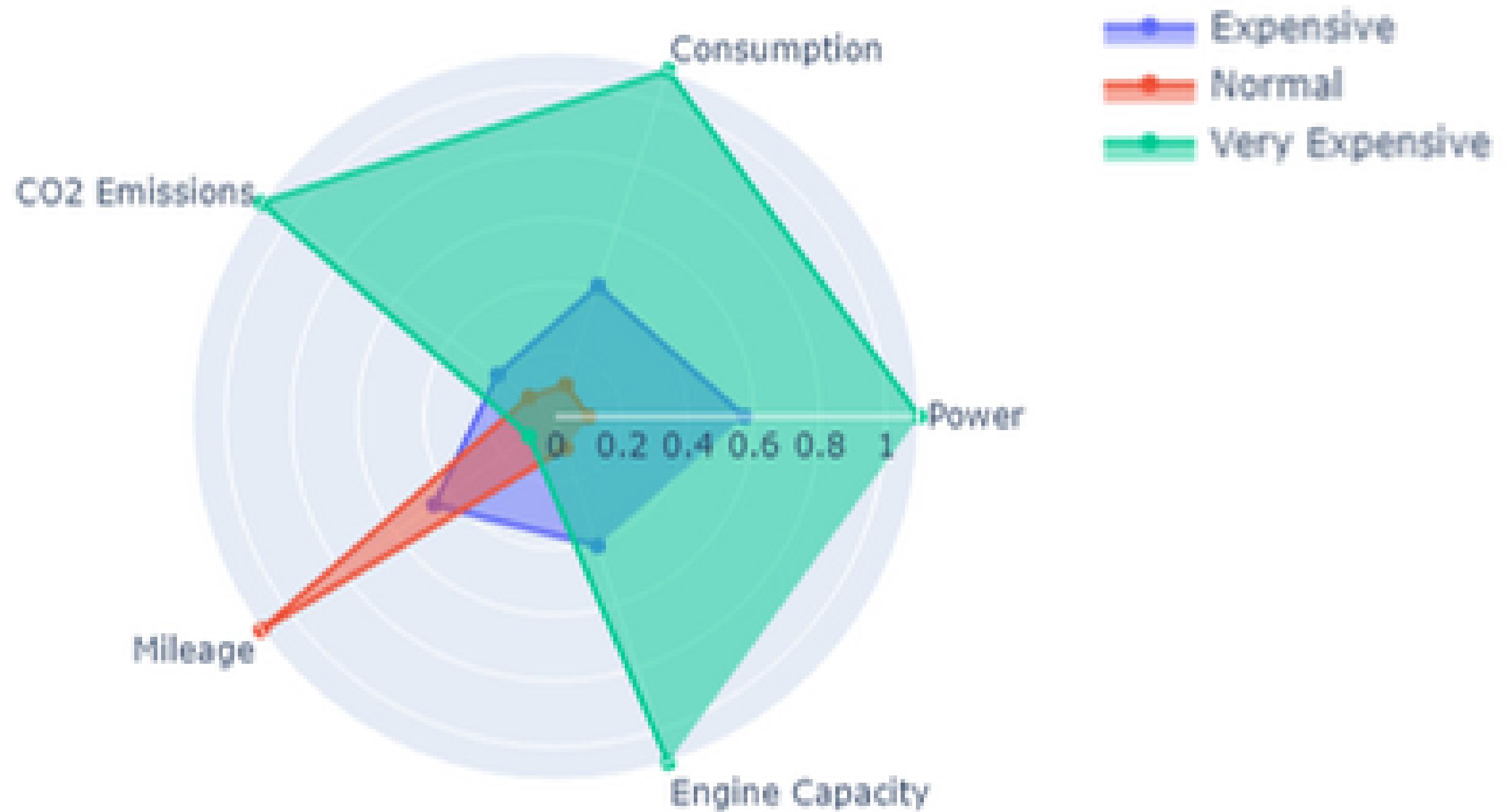


Phân chia ra từng phân khúc giá tiền

Phân tích các thông số bổ sung để xem xét xu hướng phân phối của các xe trong các nhóm khác nhau

Tính toán hiệu suất trung bình của các xe trong mỗi nhóm và trực quan hóa kết quả

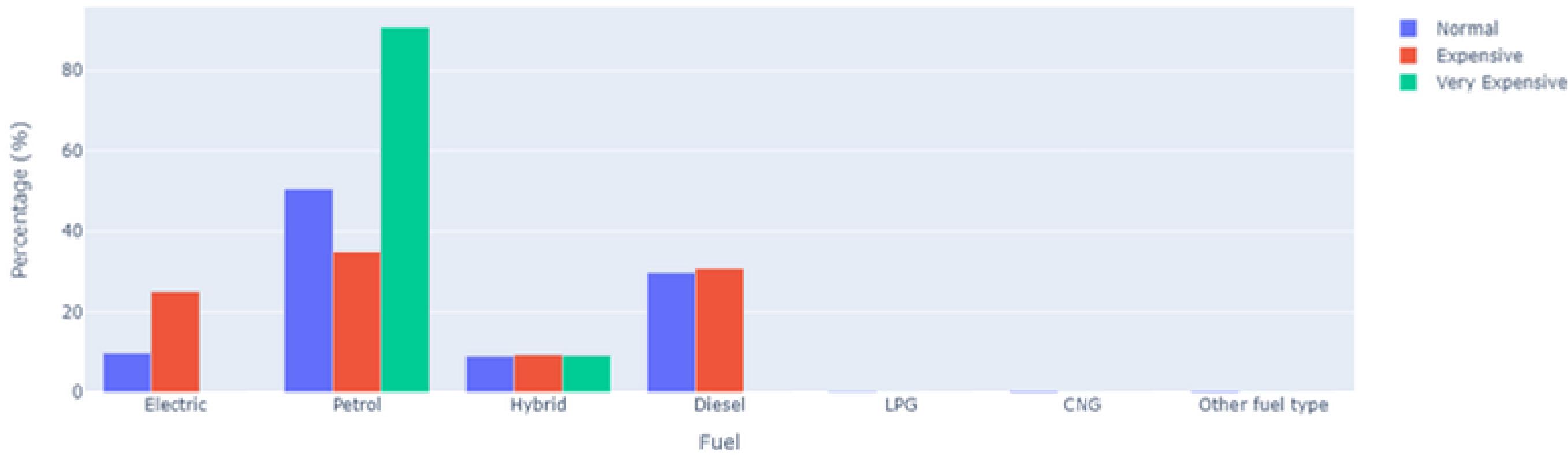
## Average Specifications of Each Type of Car



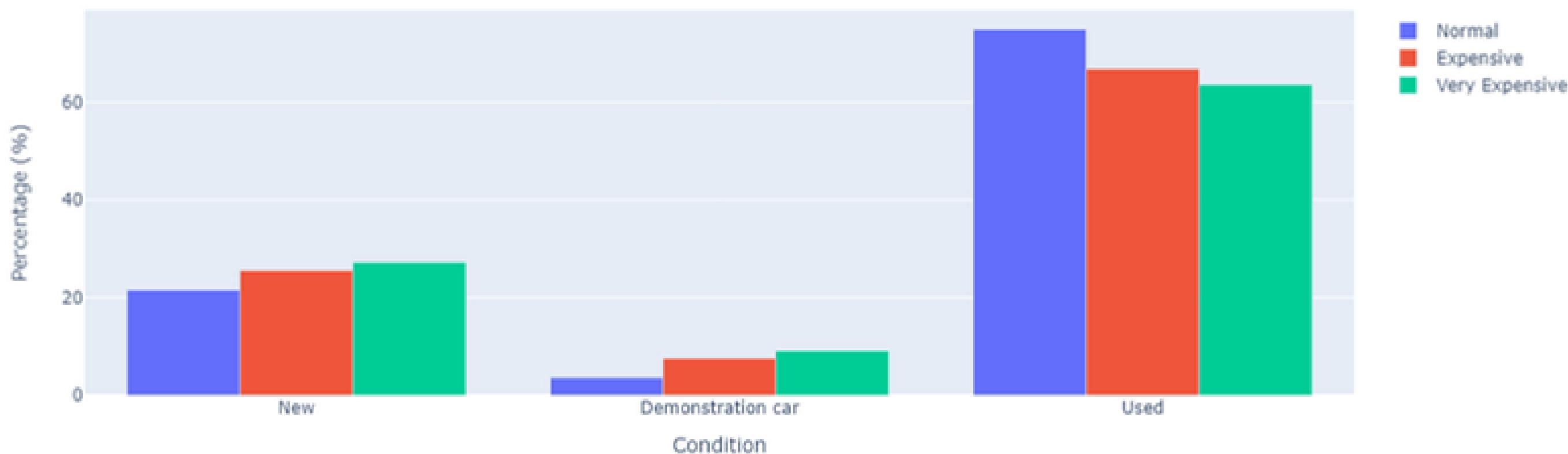
**Biểu đồ radar thể hiện hiệu năng trung bình  
của các loại xe trong từng phân khúc giá**

**Các biểu đồ  
cột thể hiện  
mối quan hệ  
hiệu năng  
trong từng  
phân khúc  
giá với các  
ứng với  
từng cột  
"Fuel",  
"Condition"**

Group Bar Chart for Fuel

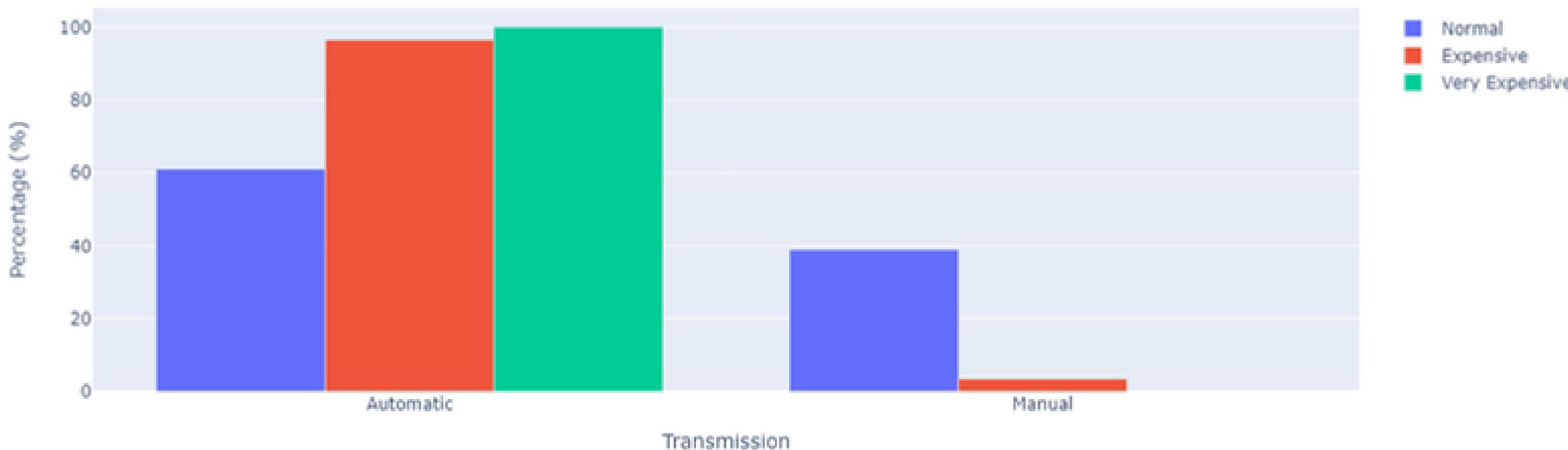


Group Bar Chart for Condition

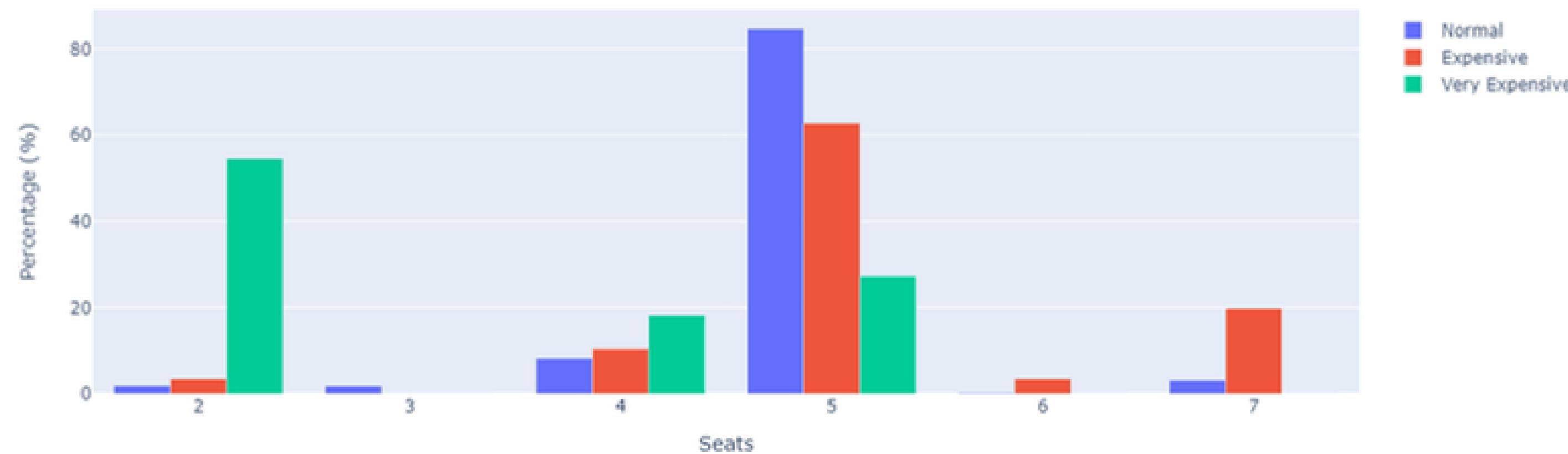


**Các biểu đồ cột  
thể hiện mối  
quan hệ hiệu  
năng trong từng  
phân khúc giá  
với các ứng với  
từng cột  
"Transmission",  
"Seats"**

Group Bar Chart for Transmission



Group Bar Chart for Seats



## Câu 4

**Có mối tương quan giữa giá bán và các tính năng đặc biệt (tag) của xe không?**

**Những tính năng nào có tác động đáng kể đến giá bán của một chiếc xe?**

**Lợi ích của câu hỏi:**

- Hiểu mối **liên hệ giữa giá và đặc tính xe**
- Xác định **đặc tính ảnh hưởng đến giá** giúp người bán định giá tối ưu cho xe của họ.

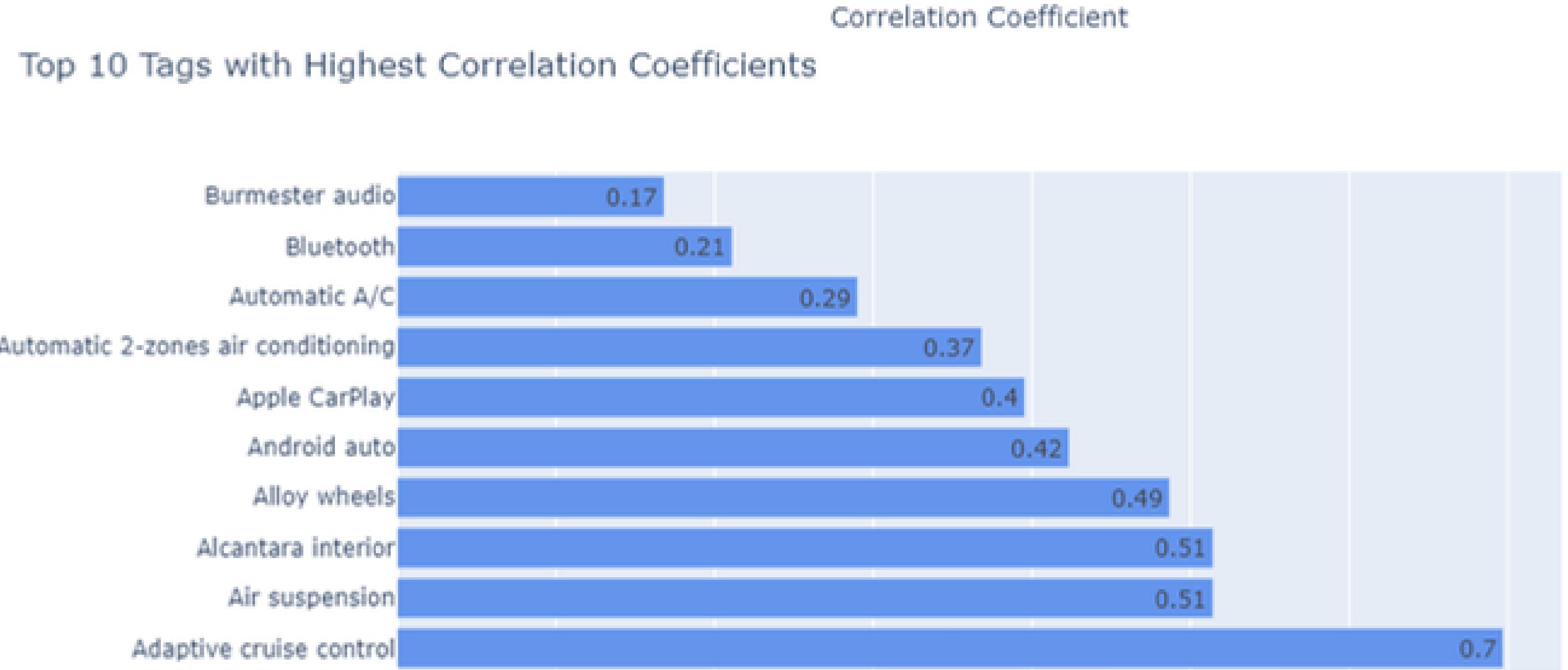
Xử lý câu hỏi



**■ Cách trả lời câu hỏi:**

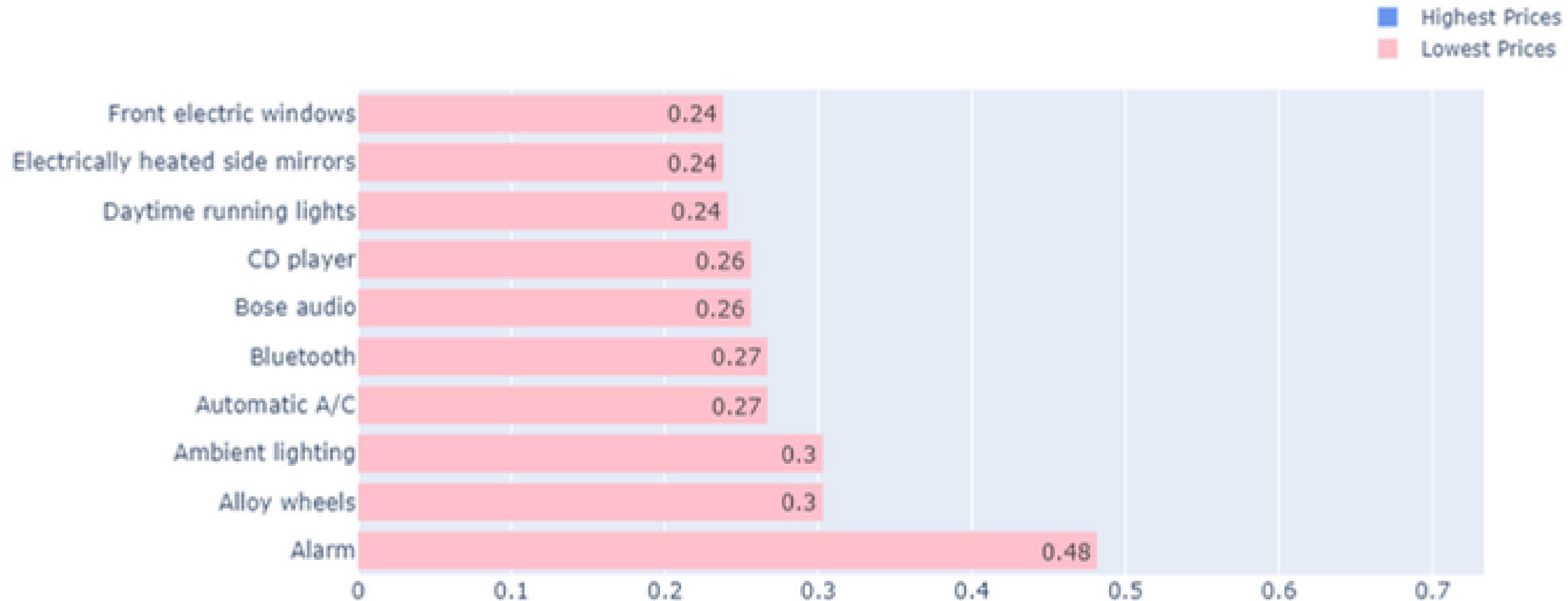
- **Bước 1: Tiền xử lý**
  - Chọn 10 xe **giá cao nhất** và 10 xe **giá thấp nhất**.
  - Tạo hàm **mã hóa one-hot** cho cột “**Tags**” và áp dụng cho bộ dữ liệu xe được chọn.
  - **Tính ma trận tương quan** cho **đặc tính và giá bán** của xe giá cao và thấp.

## Bước 2: Minh họa



Bar chart thể hiện Top 10 thẻ có hệ số  
tương quan cao Nhất với các xe mắc nhất

## Bước 2: Minh họa



Bar chart thể hiện Top 10 thẻ có hệ số  
tương quan cao nhất và xe rẻ nhất

## Câu 5

**Mức độ khí thải CO<sub>2</sub> (g/km) và tiêu thụ nhiên liệu thay đổi như thế nào giữa các kiểu dáng thân xe và loại nhiên liệu khác nhau?**

### Lợi ích của câu hỏi:

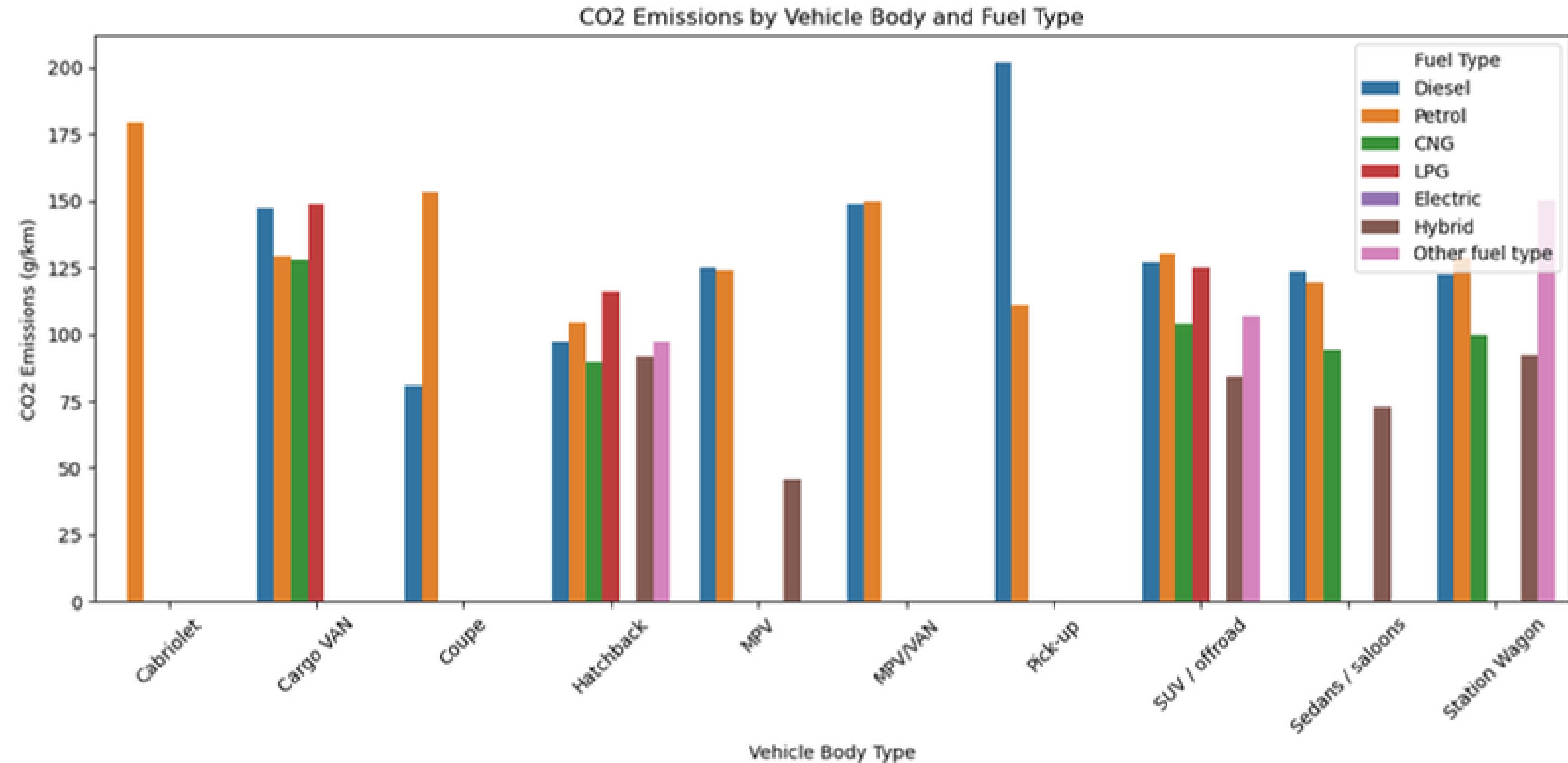
- Nhận biết các **phương tiện thân thiện môi trường**, các **phương tiện gây ô nhiễm cao**
- Hỗ trợ người dùng lựa chọn phương tiện thân thiện môi trường, góp phần bảo vệ môi trường.

Xử lý câu hỏi:

### Cách trả lời câu hỏi

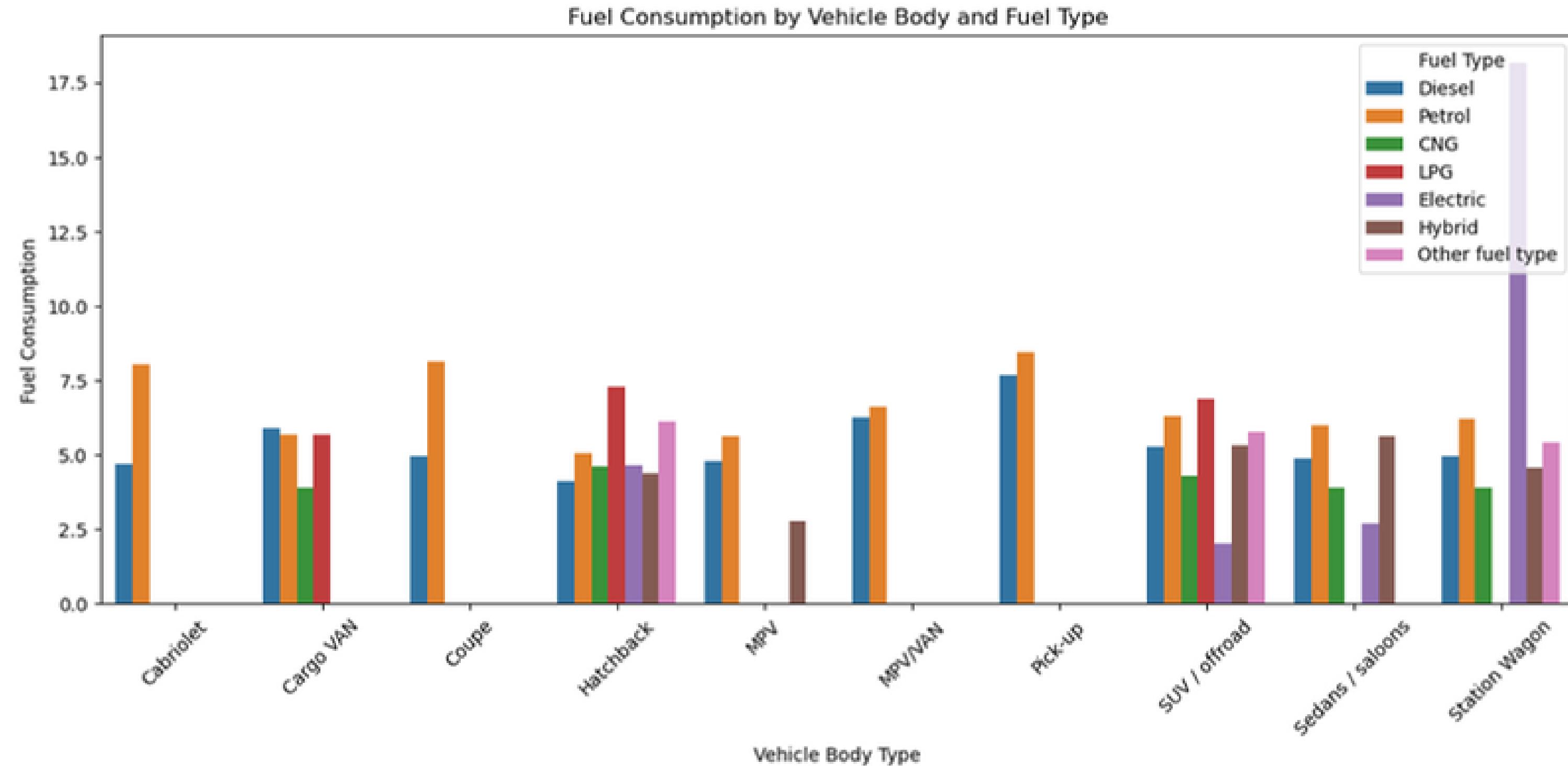
- **Bước 1: Tiền xử lý**
  - Nhóm các phương tiện **theo loại thân xe và loại nhiên liệu**
  - Tính toán lượng khí thải CO<sub>2</sub> trung bình và tiêu thụ nhiên liệu trung bình cho mỗi nhóm

## Bước 2: Minh họa



Bar Chart thể hiện sự ảnh hưởng của loại thân xe và loại nhiên liệu lên sự thải CO2

## Bước 2: Minh họa



Bar Chart thể hiện sự ảnh hưởng của loại thân xe và loại nhiên liệu lên sự tiêu thụ nhiên liệu

05

# Mô hình hóa dữ liệu

Mô hình sử dụng



Xây dựng mô hình



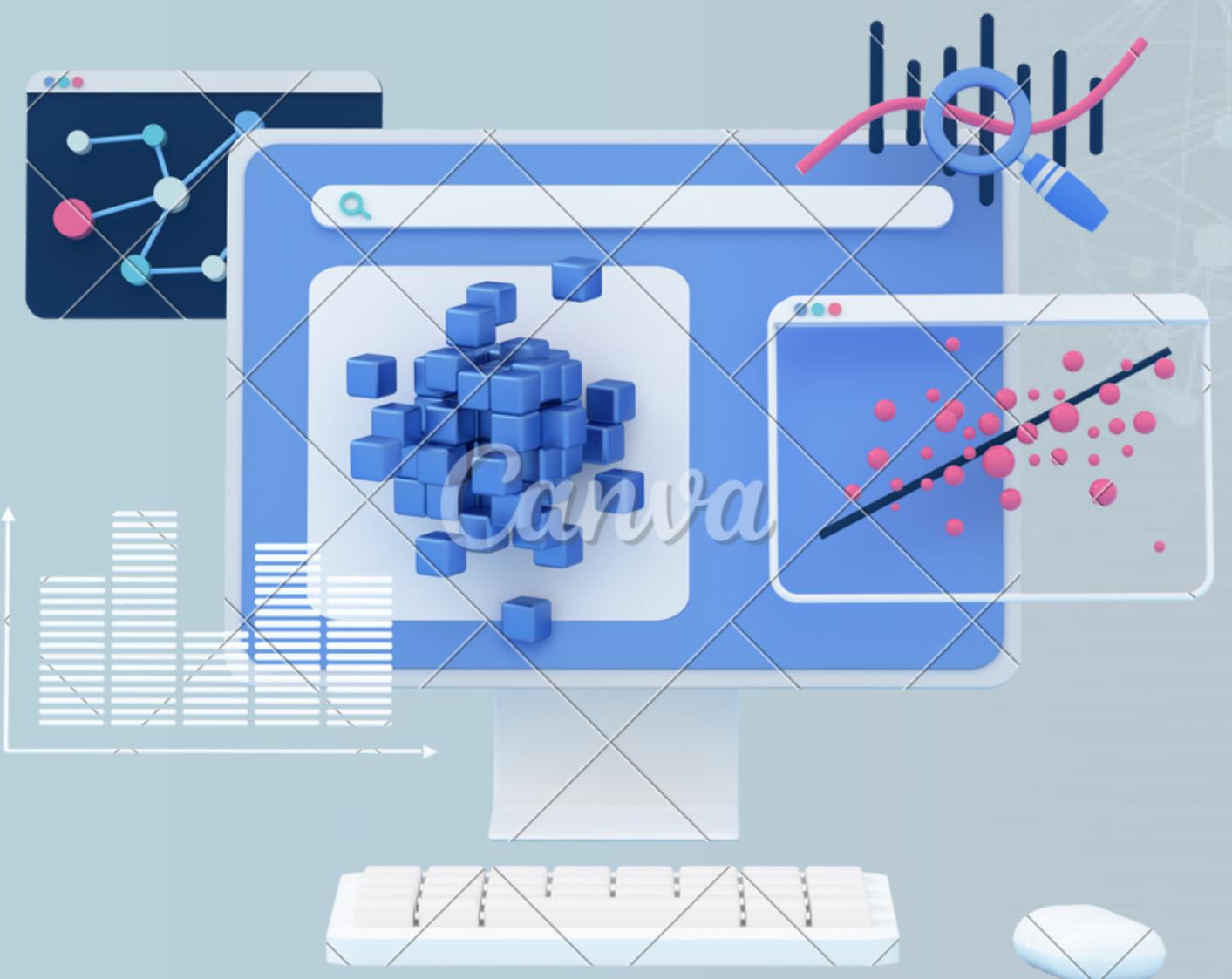
Xử lý Outlier



Tinh chỉnh tham số



Thực thi mô hình



# Mô hình sử dụng

Business Model  
Canvas

- Nhóm chọn mô hình “Random Forest” để sử dụng

- Lý do chọn mô hình “Random Forest”:

- Dễ tiếp cận sử dụng cho người mới
- Dễ cài đặt

- Giới thiệu sơ về mô hình “Random Forest”:

- Nếu như decision tree là dựa vào kết quả của một cây quyết định, thì Thuật toán này **dựa vào kết quả của nhiều cây quyết định mà mỗi cây** quyết định được huấn luyện **dựa trên nhiều mẫu con khác nhau.**
- Random forest regression có thể được áp dụng vào cả phân loại và hồi quy.

# Xây dựng mô hình

Nhóm sử dụng mô hình để dự đoán giá của một chiếc xe thông qua các thông tin mà người dùng cần cho một chiếc xe

# Xây dựng ở các cột category

- Tìm những cột có số giá trị khác nhau (**unique value**) < 10
- Nhận thấy hãng xe cũng đôi lúc ảnh hưởng đến giá trị của chiếc xe, nhóm chọn ra **các hãng có độ tương quan cao nhất (>0.25) với giá cả**

## Mercedes-Benz GLS 400 d 4Matic 243 kW

Ventilated front seats  Heated rear seats  Air suspension  Parking assist system self-steering  
 Digital cockpit  Keyless entry  Electric adjustable front seats  Heated front seats  
 Head-up display  Apple CarPlay

118 699 €

98 098 € without 21% VAT



Very good price	
Delivery	Enter ZIP code
Extended warranty	Free
Location	Germany
Mileage	8 300 km
First registration	8/2023
Transmission	Automatic
Power	243 kW (330 hp)
Fuel	Diesel
Consumption	9.6 l/100km
Drive type	4x4

Buy

## Audi S3 2.0 TFSI quattro Sportback 228 kW

Parking assist system self-steering  Digital cockpit  Heated front seats  Head-up display  
 Apple CarPlay  Android auto  Panoramic roof  Navigation system  Electric tailgate  
 Paddle shifters

62 599 €

Special VAT arrangements

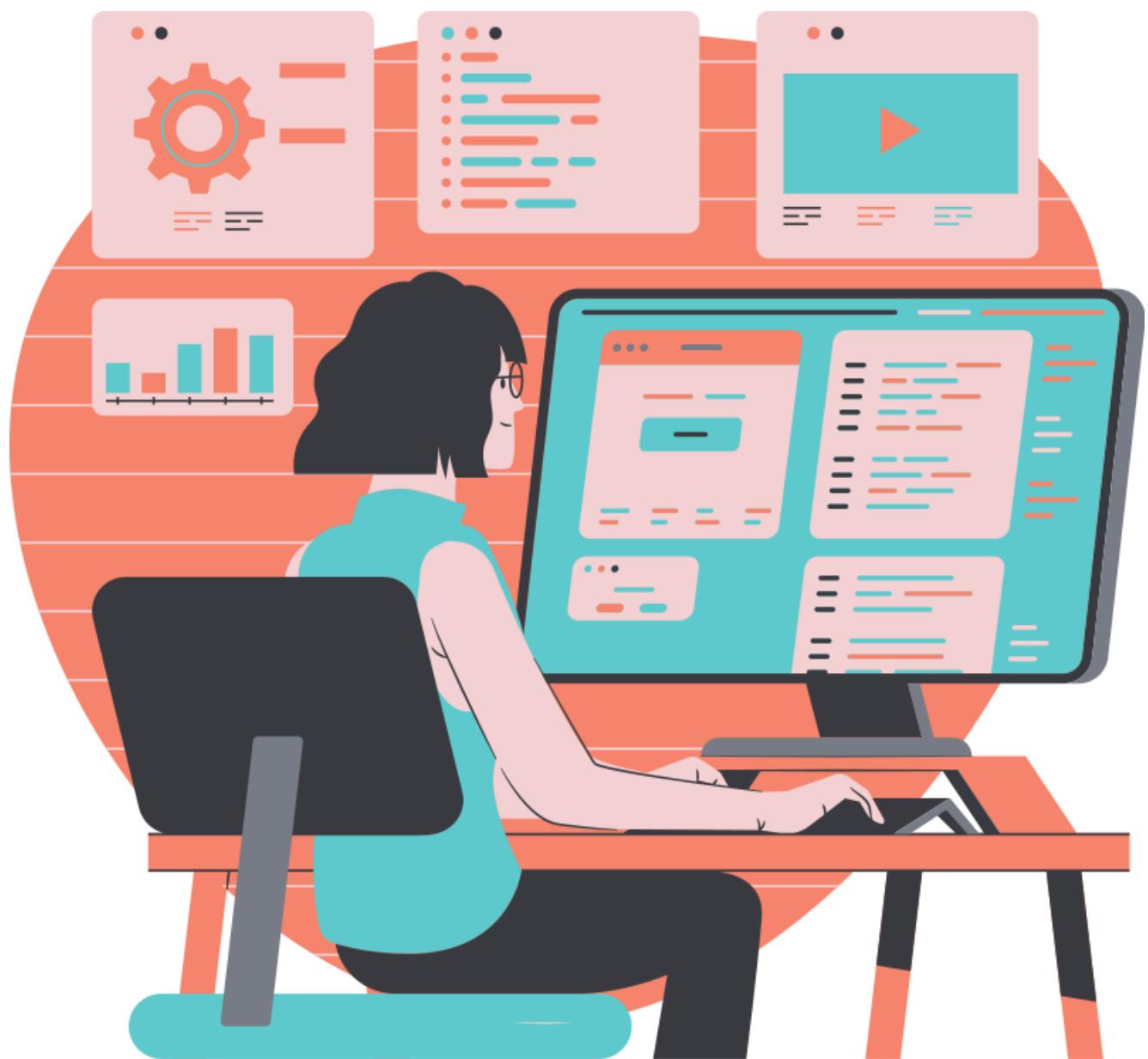


POUW EXCLUSIVE	
Delivery	Enter ZIP code
Extended warranty	Free
Location	Netherlands
Mileage	28 165 km
First registration	2/2021
Transmission	Automatic
Power	228 kW (310 hp)
Fuel	Petrol
Consumption	8 l/100km
Drive type	4x4

Buy

## Xây dựng ở các cột multiple value

- Tìm những đặc điểm nổi bật (tag) có **hệ số tương quan với Price cao.**
- Cuối cùng, thực hiện onehot encoding cho các Tag nổi bật và Price



# Xây dựng ở các cột number

Thay thế các giá trị outlier bằng phương pháp robustscaler

$$\text{Scaled Value} = \frac{\text{Original Value} - \text{Input's Median}}{\text{Input's IQR}}$$

## Tại sao chúng ta lại phải scale data?

Do dữ liệu có nhiều outliers, các giá trị trong một cột dữ liệu chênh lệch quá lớn

### Opel Astra GTC 1.4 92 kW

Apple CarPlay, Android auto, Bluetooth, Automatic lights, Rain sensor, Light sensor, Hill-start assist, Multifunctional steering wheel, Side mirrors with electric adjustment, Daytime running lights

13 299 €

Special VAT arrangements



### Mercedes-Benz GLS 400 d 4Matic 243 kW

Ventilated front seats, Heated rear seats, Air suspension, Parking assist system self-steering, Digital cockpit, Keyless entry, Electric adjustable front seats, Heated front seats, Head-up display, Apple CarPlay

118 699 €

98 098 € without 21% VAT



Very good price

Delivery: Enter ZIP code: Free

Extended warranty: Free

Location: Poland

Mileage: 70 000 km

1/2017

First registration: Manual

Transmission: Petrol

Power: 92 kW (125 hp)

Fuel: 5.4 l/100km

Consumption: 4x2

Drive type: Buy

Buy

Very good price

Delivery: Enter ZIP code: Free

Extended warranty: Free

Location: Germany

Mileage: 8 300 km

8/2023

First registration: Automatic

Transmission: Diesel

Power: 243 kW (330 hp)

9.6 l/100km

Fuel: 4x4

Consumption: Buy

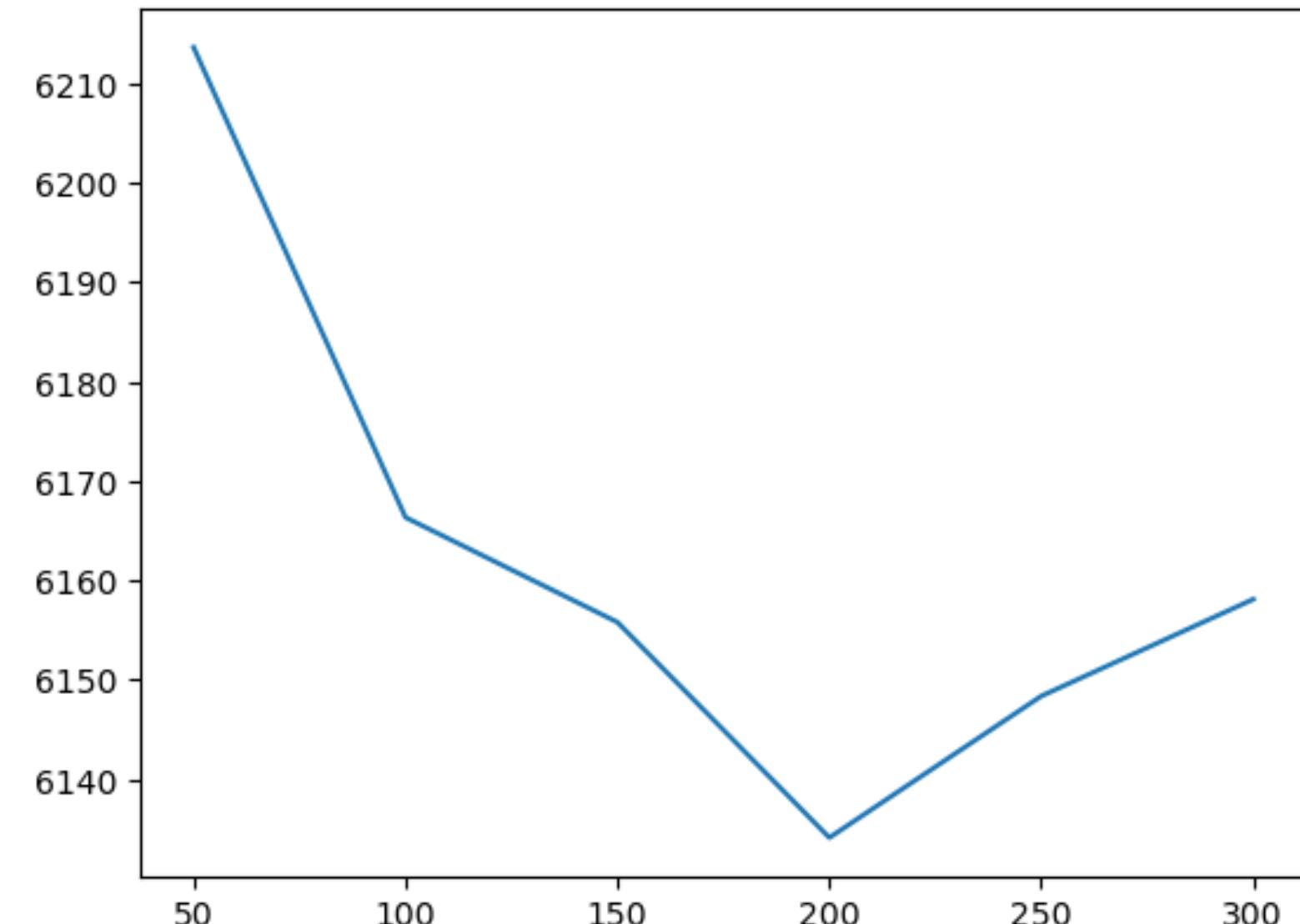
Drive type: Buy

Buy

# Tinh chỉnh tham số

Business Icons

- **Lựa chọn n\_estimator – số lượng cây trong rừng phù hợp**



Biểu đồ thể hiện sự số lượng n\_estimator

# Thực thi mô hình

- Sử dụng thư viện **flask** để xây dựng một website đơn giản chạy trong localhost.
- Người dùng sau khi truy cập trang web, sẽ **điền các thông tin một chiếc xe mà mình muốn** dự đoán, sau đó nhờ mô hình mà ta **đã xây dựng trước** đó, sẽ dự đoán ra giá của chiếc xe đó.

The screenshot shows a web-based car configuration tool. At the top, there are several tabs: 'Giải quyết tham nh...' (Solve conflicts), 'Booking.com', 'Express VPN', and a 'New' tab. Below the tabs is a large input form for car details. The fields include:

- Số ghế: [Input field]
- Công suất (kW): [Input field]
- Lượng khí thải CO2 (g/km): [Input field]
- Số km đã đi: [Input field]
- Mức tiêu thụ nhiên liệu (l/100km hoặc kWh/100km): [Input field]
- Dung tích động cơ (ccm): [Input field]
- Số chủ xe trước: [Input field]
- Số năm kể từ ngày đăng ký lần đầu: [Input field]
- Đặc điểm nổi bật:
  - Air suspension
  - Ventilated front seats
  - Electric adjustable front seats
  - Digital cockpit
- Màu nội thất: [Dropdown menu: Black interior]
- Chất liệu nội thất: [Dropdown menu: Full leather interior]
- Cửa: [Dropdown menu: 4/5 doors]
- Nhiên liệu: [Dropdown menu: Electric]
- Hộp số: [Dropdown menu: Automatic]
- Loại dẫn động: [Dropdown menu: 4x2]
- Lớp khí thải: [Dropdown menu: No emission class]
- Tình trạng: [Dropdown menu: New]
- Nhà sản xuất: [Dropdown menu: Ferrari]

At the bottom right of the form is a large blue button labeled 'Gửi' (Send).

**Nội dung của trang web**

Số ghế: 4

Công suất (kW): 250

Lượng khí thải CO2 (g/km):  
200

Số km đã đi: 10000

Mức tiêu thụ nhiên liệu (l/100km hoặc kWh/100km): 9

Dung tích động cơ (ccm):  
2500

Số chủ xe trước: 1

Số năm kể từ ngày đăng ký lần đầu:  
2022

Đặc điểm nổi bật:

- Air suspension
- Ventilated front seats
- Electric adjustable front seats
- Digital cockpit

Màu nội thất: Black interior

Chất liệu nội thất: Full leather interior

Cửa: 4/5 doors

Nhiên liệu: Electric

Hộp số: Automatic

Loại dẫn động: 4x2

Lớp khí thải: No emission class

Tình trạng: Used

Nhà sản xuất: Others

**Gửi**



**Kết quả dự đoán**

**Các thông số của một chiếc xe  
cần dự đoán**



127.0.0.1:5000



# Giá xe dự đoán bằng RandomForestRegression là: 98 186.792 EUR

## Kết quả dự đoán

**BMW X7 xDrive40i 280 kW**

- Driver's seat with massage function
- Ventilated front seats
- Heated rear seats
- Air suspension
- Parking assist system self-steering
- Digital cockpit
- Keyless entry
- Electric adjustable front seats
- Heated front seats
- Active steering

**110 049 €**

90 950 € without 21% VAT

Top offer

Delivery	Enter ZIP code
Extended warranty	Free
Location	Germany
Mileage	10 001 km
First registration	9/2023
Transmission	Automatic
Power	280 kW (381 hp)
Fuel	Petrol
Consumption	9.7 l/100km
Drive type	4x4

Buy

**Giá thực tế của một chiếc xe có thông số tương tự**

# Những khó khăn gặp phải

- Preprocess: phân vân cách xử lí outlier, liệu có nên remove outlier hay không, và nên xử lí nó thế nào.
- Tham số model: chọn tham số nào để tinh chỉnh vì mô hình này có rất nhiều tham số?
- Suy nghĩ về việc đặt câu hỏi và ứng dụng giải quyết vấn đề thực tế



# Tự đánh giá

Business Ideas

## Học được

- Các crawl data thực tế, tìm hiểu qua selenium
- Cách deploy một mô hình học máy
- Cách scale data, xử lý outlier



## Ưu điểm

- Chủ đề gần gũi với thời đại
- Đồ án được thực hiện theo quy trình rõ ràng
- Biết cách sử dụng các công cụ và môi trường làm việc



## Khuyết điểm

- Bắt đầu làm đồ án tương đối muộn
- Bộ dữ liệu còn ít



Cảm ơn Thầy  
đã lắng nghe