**KHOA CÔNG NGHỆ THÔNG TIN**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM**

Đồ án môn học  - TOÁN ỨNG DỤNG VÀ THỐNG KÊ

HỌC KỲ II – NĂM HỌC 2019-2020

**môn toán ứng dụng và thống kê**

**BẢNG THÔNG TIN CHI TIẾT NHÓM**

|  |  |  |
| --- | --- | --- |
| **Số lượng:** | **5** | |
| **MSSV** | **Họ tên** | **Email** |
| 1712003 | Lê Nguyễn Đức Anh | 1712003@student.hcmus.edu.vn |
| 1712061 | Lâm Quỳnh Hương | 1712061@student.hcmus.edu.vn |
| 1712096 | Đặng Hồng Minh | 1712096@student.hcmus.edu.vn |
| 1712103 | Ngô Minh Nghĩa | 1712103@student.hcmus.edu.vn |
| 1712112 | Cao Thành Nhân | 1712112@student.hcmus.edu.vn |

**PHÂN CÔNG CÔNG VIỆC**

|  |  |  |  |
| --- | --- | --- | --- |
| MSSV | Họ tên | Phân công đồ án 1 | Phân công đồ án 2 |
| 1712003 | Lê Nguyễn Đức Anh | Đọc dữ liệu | Train và test |
| 1712061 | Lâm Quỳnh Hương | Xử lí dữ liệu | Đọc dữ liệu |
| 1712096 | Đặng Hồng Minh | Xử lí dữ liệu, Viết báo cáo | Tiền xử lí ảnh |
| 1712103 | Ngô Minh Nghĩa | Xử lí, train và test | Phân tích dữ liệu |
| 1712112 | Cao Thành Nhân | Xử lý, Phân tích dữ liệu | Viết báo cáo |

# **Yêu cầu của Đồ án:**

## Đồ án 1:

- Sinh viên áp dụng các kiến thức về hồi quy tuyến tính để huấn luyện mô hình dựa vào tập dữ liệu ở trên và đưa ra dự đoán về giá xe.

- Việc chọn mô hình như thế nào cho phù hợp là do sinh viên tự quyết định. Tuy nhiên, sinh viên cần phải giải thích lý do tại sao bản thân lại chọn mô hình như vậy.

## Đồ án 2:

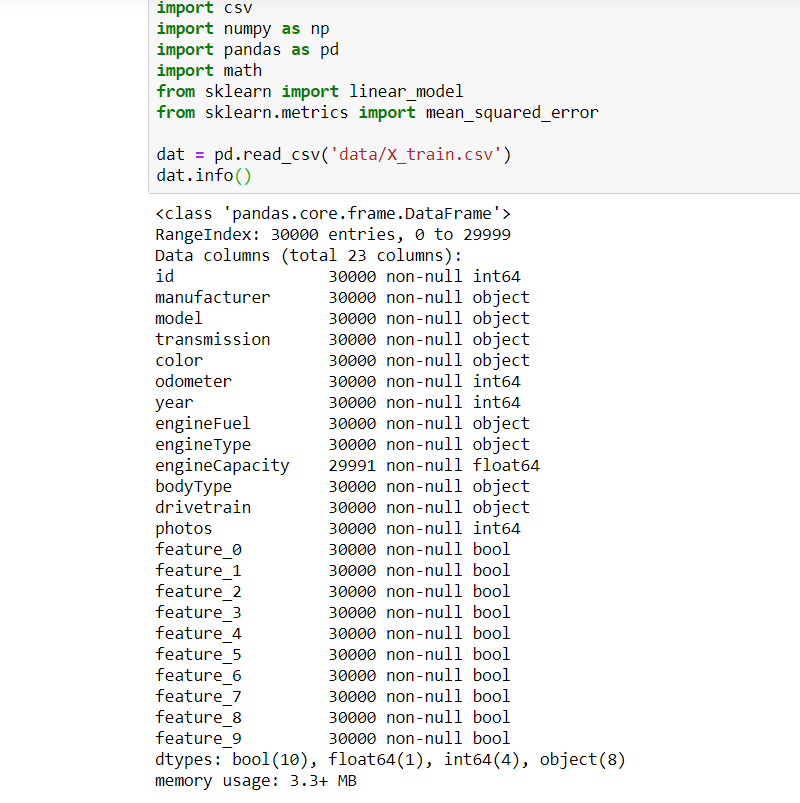
- Đọc được toàn bộ tập dữ liệu.

- Hiển thị hình ảnh và nhãn của bất kỳ mẫu nào trong tập dữ liệu. Hình ảnh cần được hiển thị như một bức ảnh grayscale, kích thược 28 × 28 pixel.

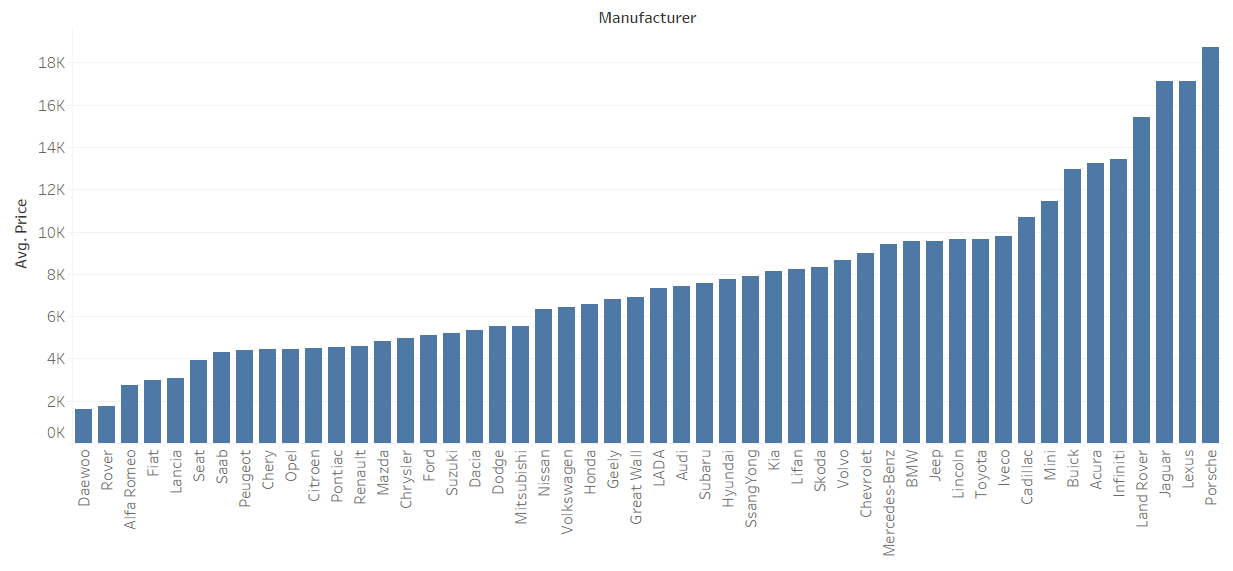
# **Kết quả:**

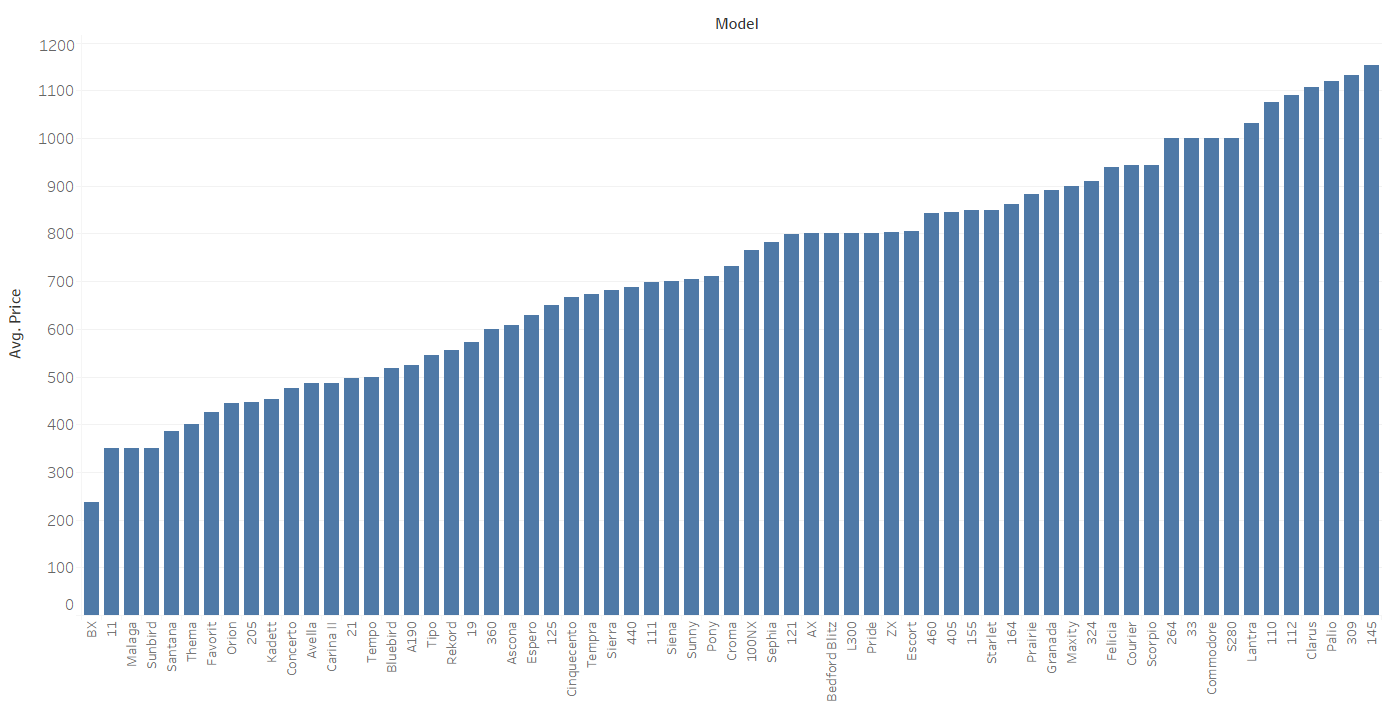
## I . Đồ án 1:

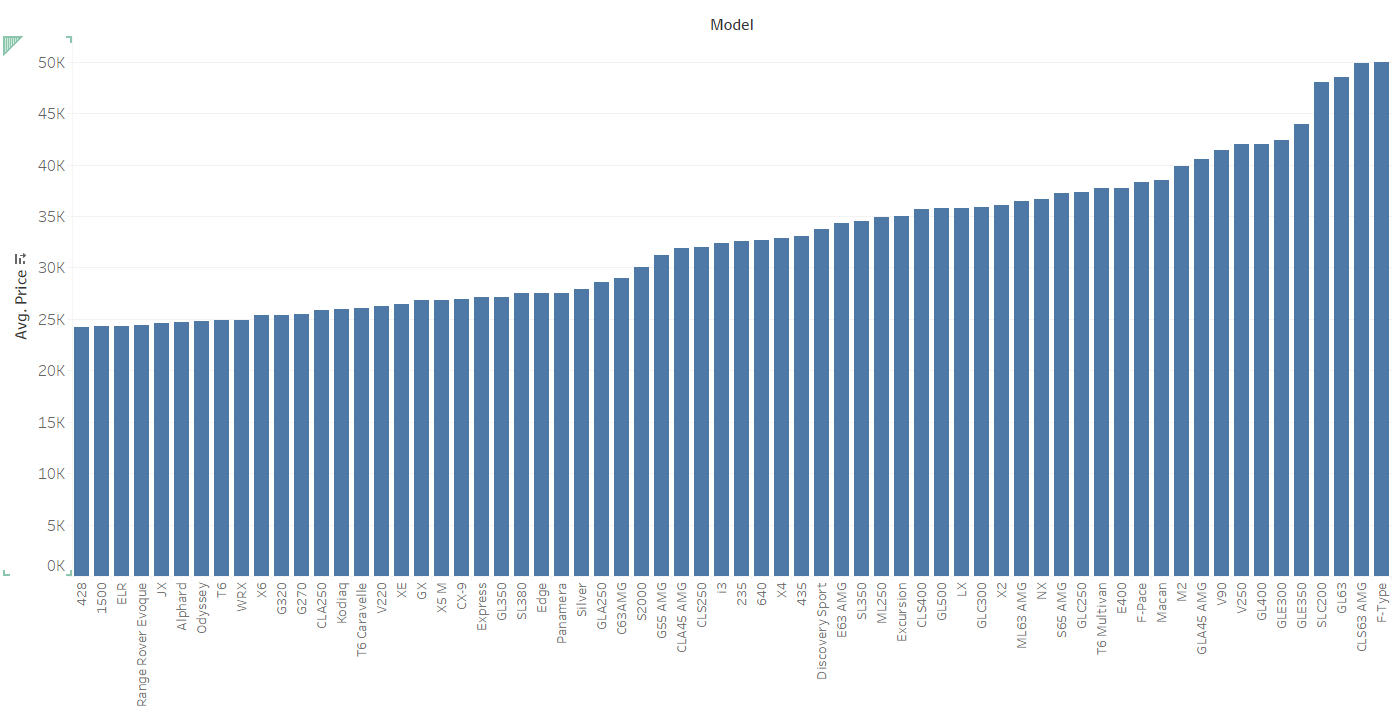
1. Thể hiện dữ liệu:

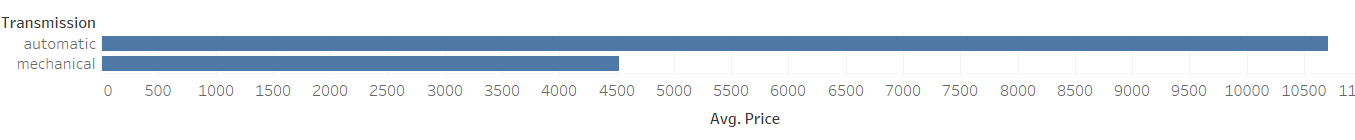


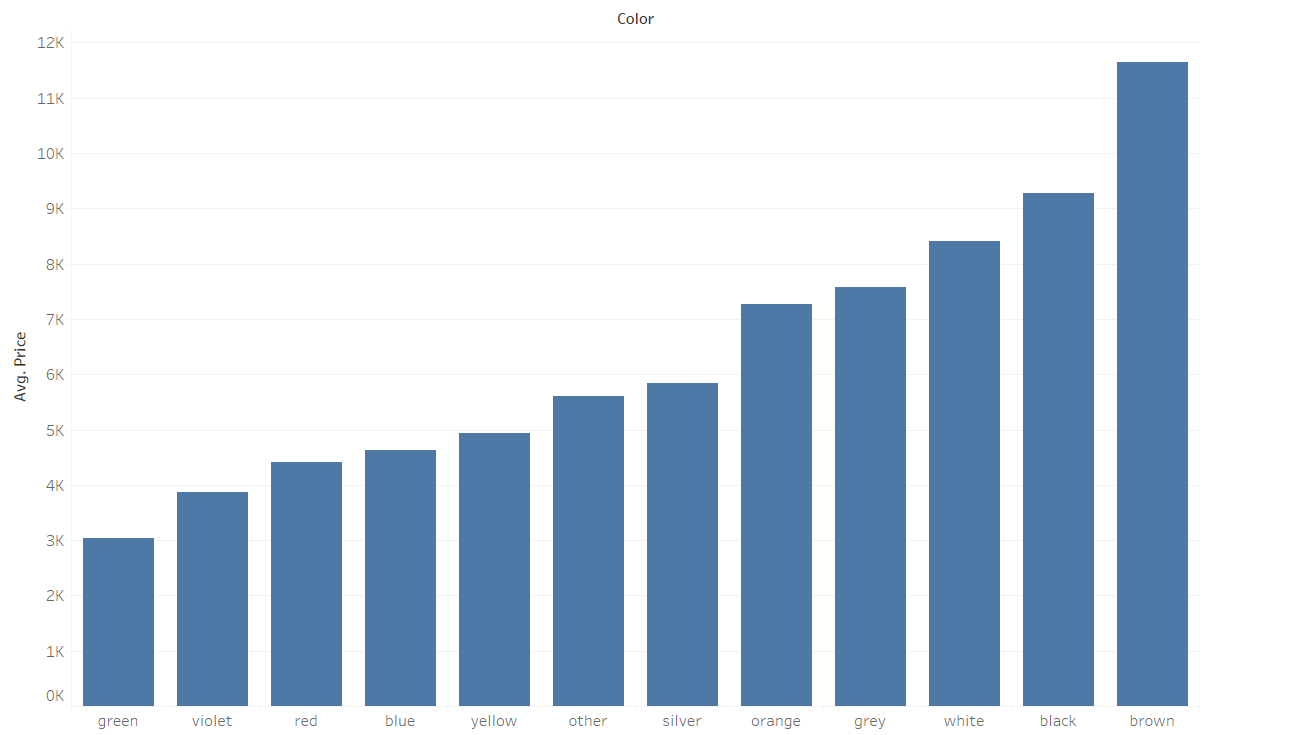
2. Ảnh hưởng dữ liệu đến giá:

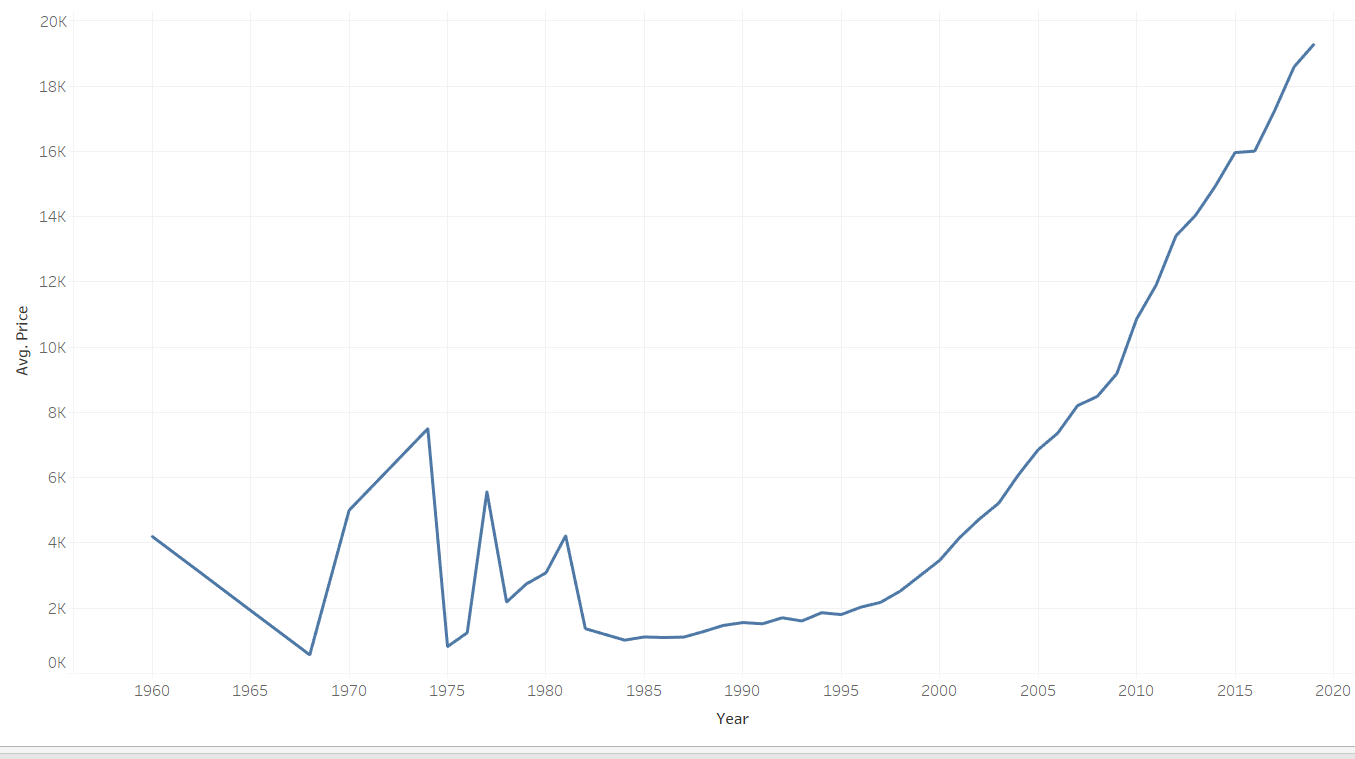


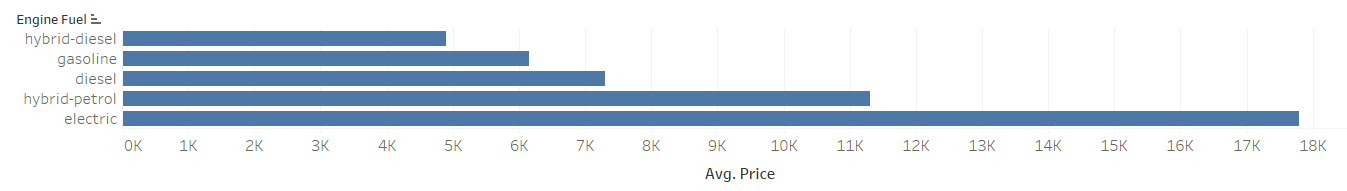


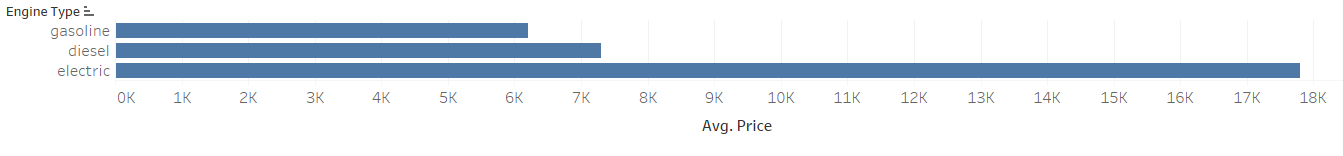


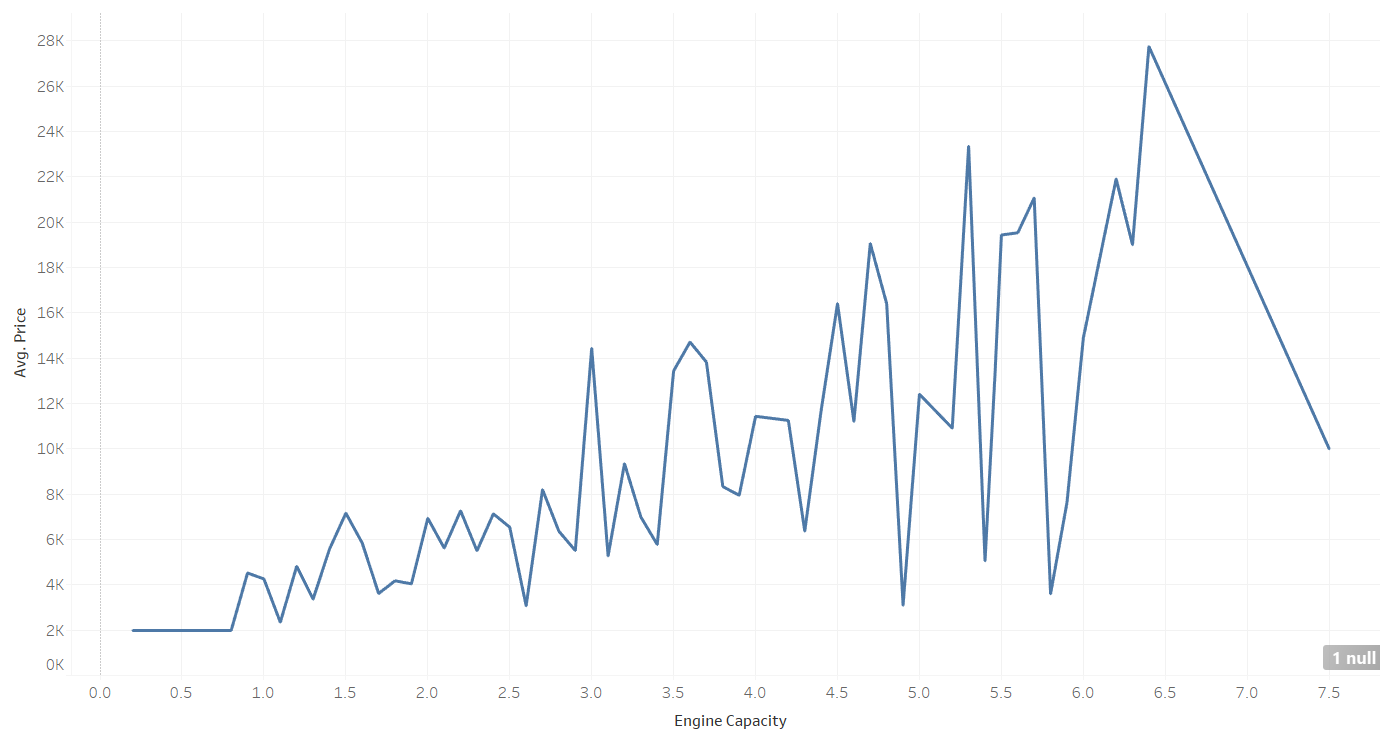


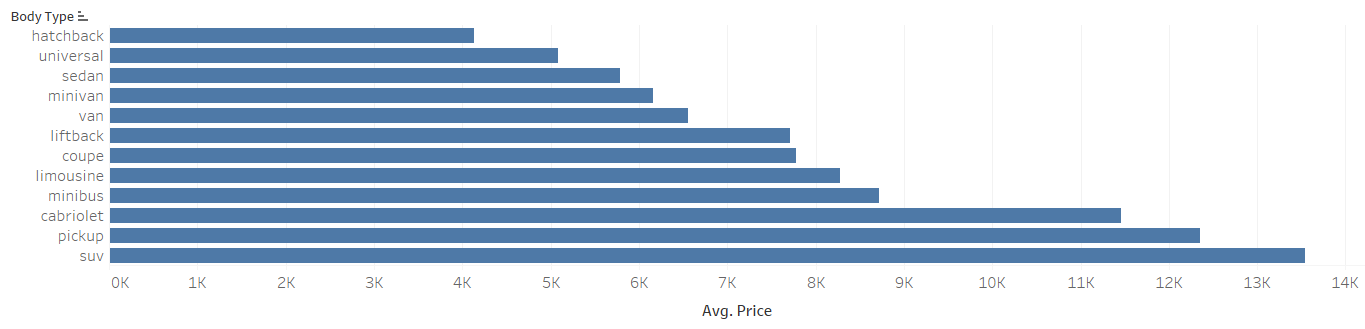


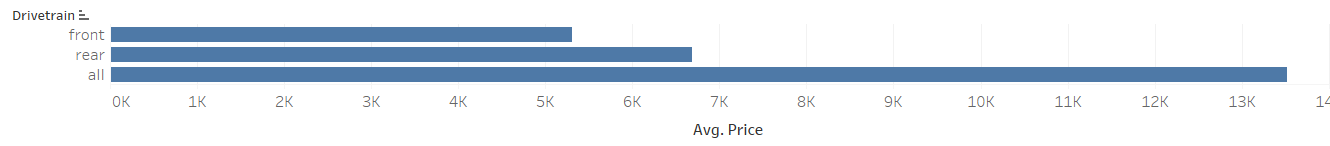


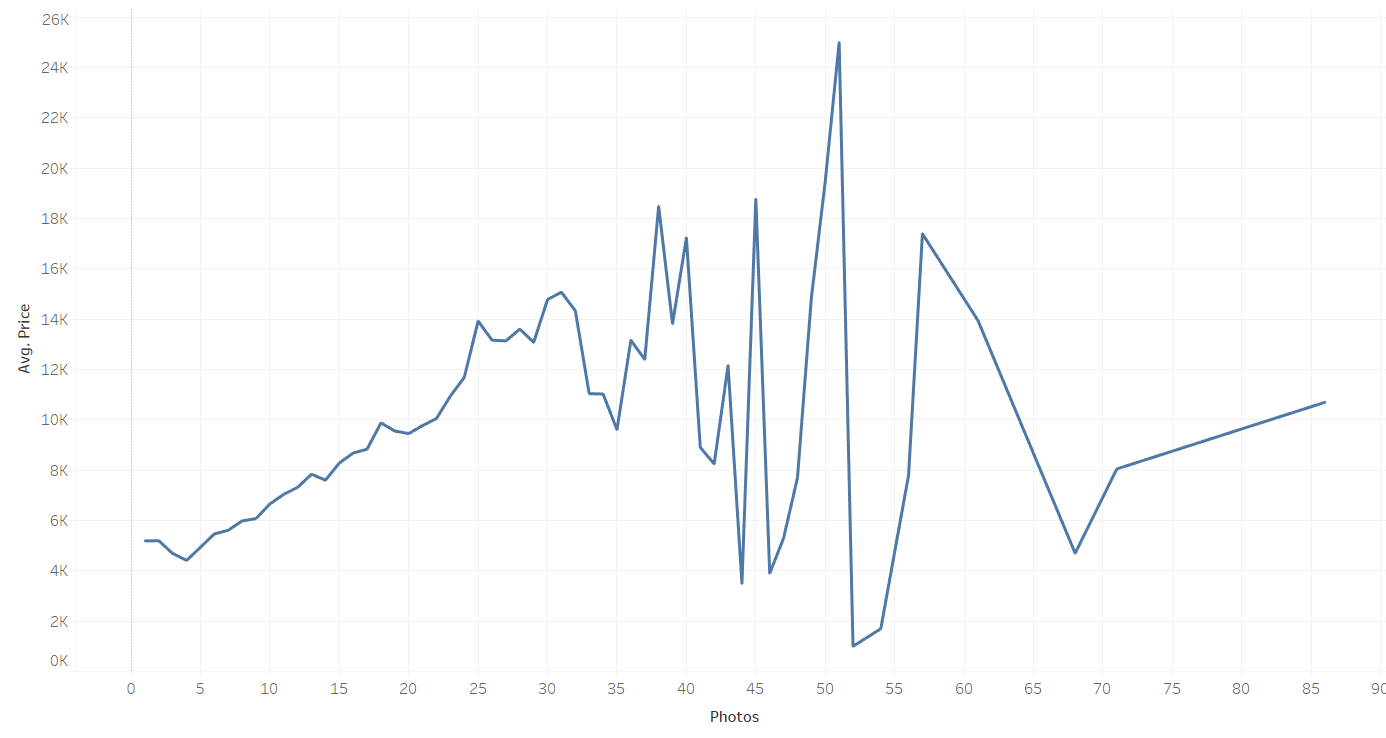


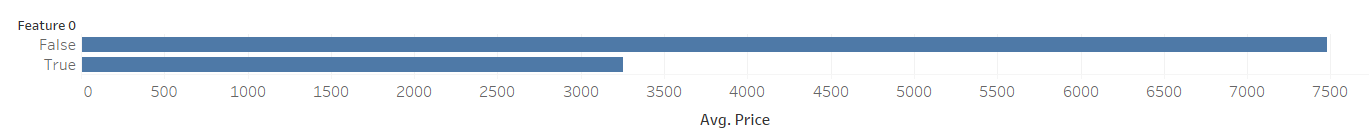


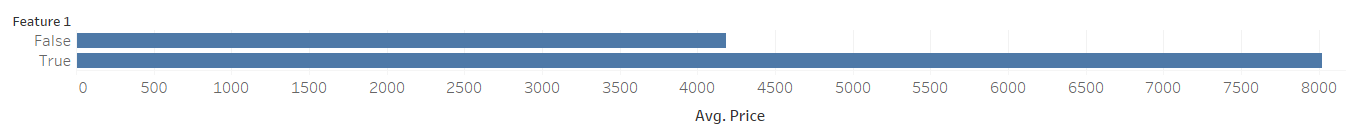


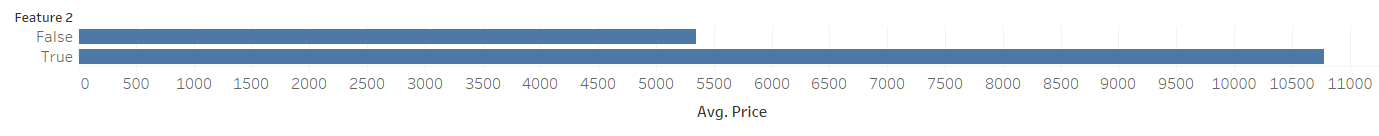


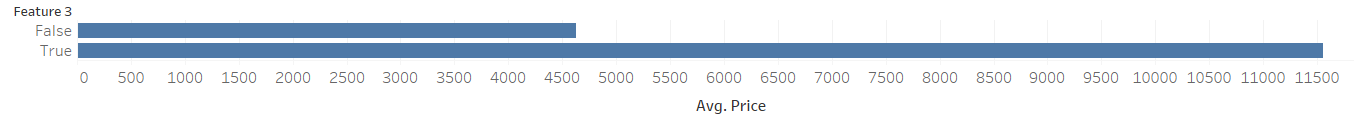


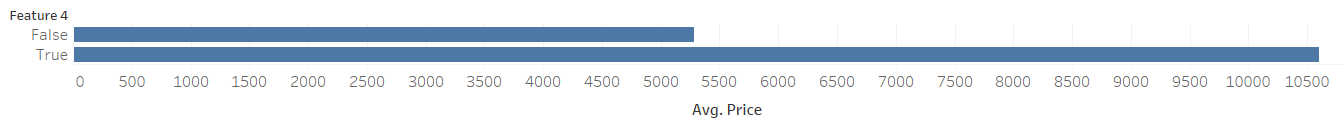


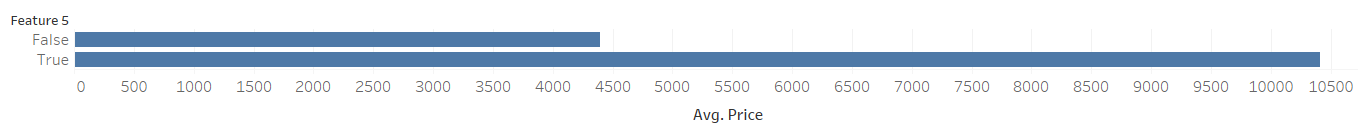


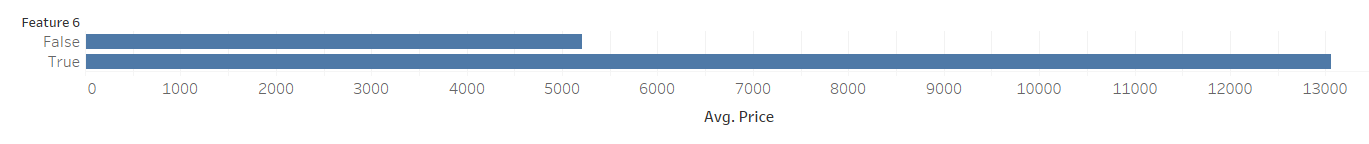


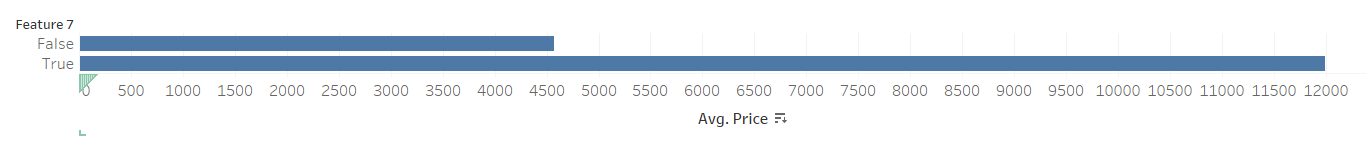


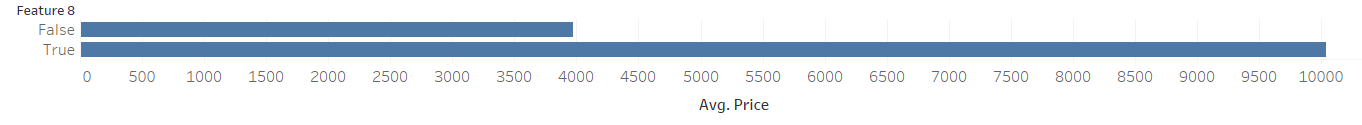


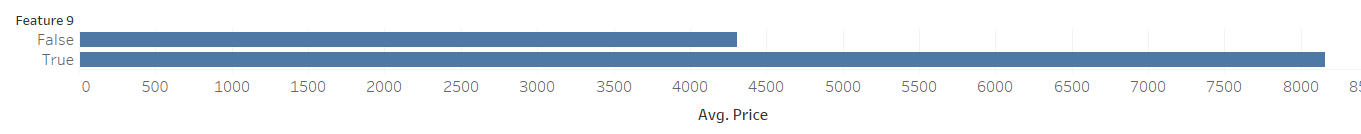








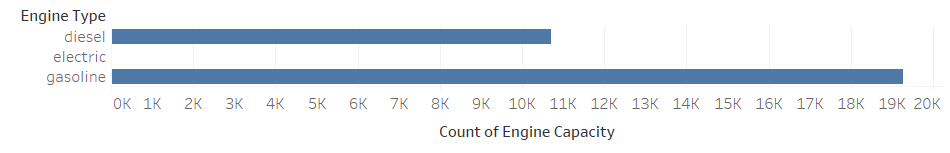




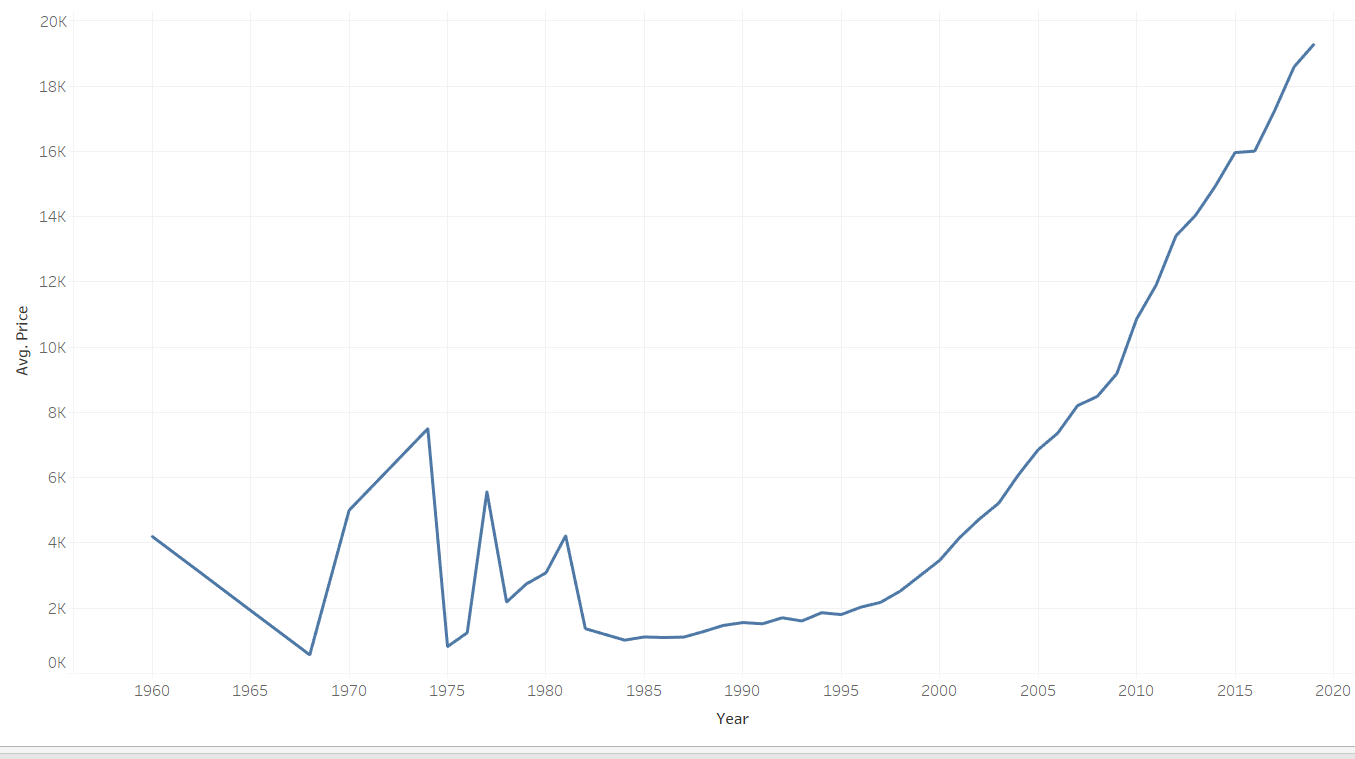
3. Các điểm đặc biệt trong dữ liệu:

-Feature 0 = True Thì Feature khác = False.

-Engine Capacity = Null khi và chỉ khi Engine Type= Electric.



-Year càng lớn thì giá càng cao từ năm 1985 trước 1985 giá trị có dấu hiệu tăng nhưng không ổn định.



## II . Đồ án 2:

1. Đọc dữ liệu:

Đọc dữ liệu vào và chuyển sang ma trận ảnh thứ i (28x28) để xử lí nhiễu

A screenshot of a cell phone

Description automatically generatedA screenshot of text

Description automatically generatedTiếp theo chuyển ma trận sau khi xử lí nhiễu thàng mảng 1 chiều X

A screenshot of a cell phone

Description automatically generatedCuối cùng là đọc nhãn vào mảng 1 chiều Y

1. Ý nghĩa dữ liệu:

Hàm xuất ảnh thứ i

A screenshot of a cell phone

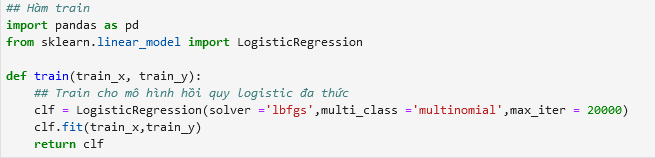
Description automatically generated Chuyển dữ liệu từ mảng 1 chiều dòng thứ i thành ma trận pixel 28x28

Chuyển ma trận thành dạng hình ảnh

 Kết quả chạy thử ảnh thứ 26999

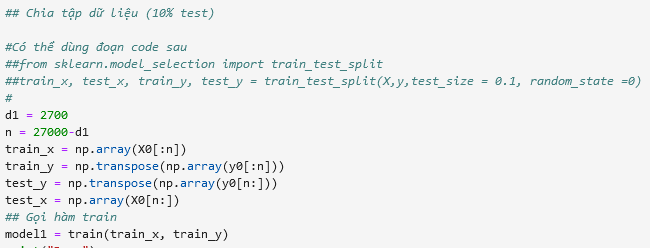
Ý nghĩa dữ liệu: Các thủ ngữ dành cho người khiếm thính

1. Hàm train cho mô hình:

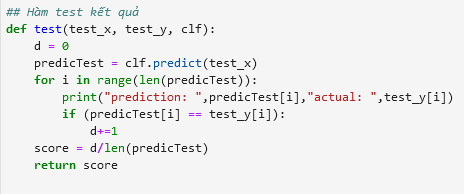


Dữ liệu vào train\_x, train\_y, test\_x, test\_y (10% test)

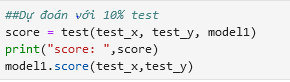
Gọi hàm train



Hàm test kết quả:



Thử dự đoán kết quả test\_x, test\_y (10% tập dữ liệu)



Kết quả trả về



# Phương pháp: Plot multinomial

Có J giá trị đầu ra (24 label)

Dùng J-1 mô hình: so các xác suất các giá trị đầu ra còn lại với xác suất giá trị đầu ra thứ J

Ta có: = 1

⬄

* :=P[Y=J|=,…,=]=
* :=P[Y=j|=,…,=]=

Tìm max(, ) (j = 1…J-1)

1. Nhận xét

Kết quả dự đoán đạt được rất tốt do train được tập dữ liệu rất lớn.

Không có trường hợp sai trong tập 10% test.

**Dự đoán trường hợp khiến kết quả sai:**

* Ảnh kém chất lượng, độ sáng ảnh không phù hợp (quá chói hoặc quá tối khiến cho các giá trị điểm ảnh bị sai lêch)