

CAR RESALE PRICE PREDICTION DATASET

GROUP 2: ABINO, ABOT, CO, DETABLAN

I. BUSINESS OBJECTIVES

Used Car Market

The often expensive prices of brand new cars - generally resulting in most consumers being unable to afford them - have resulted in a fairly lucrative second-hand car market in the Philippines. However, due to the many factors involved, attempting to determine the value of a given second-hand car is a nontrivial task, making things like sale forecasts, good purchasing and pricing decisions for such cars difficult for business owners.

Another concern in the second hand car industry is the prioritization of different characteristics of vehicles in purchases, repairs, and advertising. Due to the swathe of relevant factors involved, the relative preferences of consumers tend to be obscured under possibly irrelevant data.

The following dataset may be able to assist in accomplishing the above mentioned nontrivial tasks

II. TECHNICAL APPLICATIONS

Used Car Prices

The work of Collard (2022) has shown that using information on second hand cars that are readily available online and Random Forest Regression, it is possible to predict the resale price of a car based on its features with an R^2 of ~0.77, which is likely to be feasible to use in business decisions.

Meanwhile, another paper by Totakura & Kosuru (2021) not only shows similarly accurate regression (R^2 of ~85-91%) using Light Gradient Boosted Machine learning but also the extraction of feature importance, allowing one to see not only which factors influence resale pricing, but also by how much.

Given the relative accessibility of such data, this seems to be a feasible method of achieving the aforementioned business objectives.

III. PRIMARY WEBSITES DESCRIPTION

1. AUTODEAL WEBSITE: USED CARS

Using the website of AutoDeal (<https://www.autodeal.com.ph/>), the home screen has filters at the top bar as seen on the screenshot. Since our primary dataset contains used cars, we used the “Used Cars” filter, proceeded with “Used Cars for sale”, and searched for cars available.

The screenshot shows the homepage of the AutoDeal website. At the top, there is a navigation bar with the AD logo and links for New Cars, Car Promos, Used Cars, News & Reviews, Loans & Insurance, and Comparison. Below the navigation bar, the headline "Car Buying, Simplified." is displayed, followed by the subtext "Compare prices on new & used cars from official dealers across the Philippines." Two buttons are present: "Search New Cars" (orange) and "Search Used Cars" (white). The main feature is a large image of an orange MG ZS SUV. To the right of the car, a circular badge says "2024 BEST DEAL" and "AUTO DEAL.COM.PH". Above the car, the text "Low Monthly PHP" is shown, followed by a large "19K" in bold black letters. Below the car, the text "Available in all MG dealerships nationwide, the MG ZS MCE COM CVT is priced at P918,888." is displayed. At the bottom of the feature, the text "MG ZS MCE COM CVT" is shown. At the very bottom of the page, there are links for "The Mitsubishi XForce", "MG ZS MCE Promo" (which is highlighted in orange), "Nissan Terra", "AutoDeal App", and "Car Insurance".

1. AUTODEAL WEBSITE: USED CARS

After clicking those filters, it should display the used cars listed for sale (see the image).

The screenshot shows the AutoDeal website interface for searching used cars. On the left, there is a sidebar titled "Filters" with several dropdown menus:

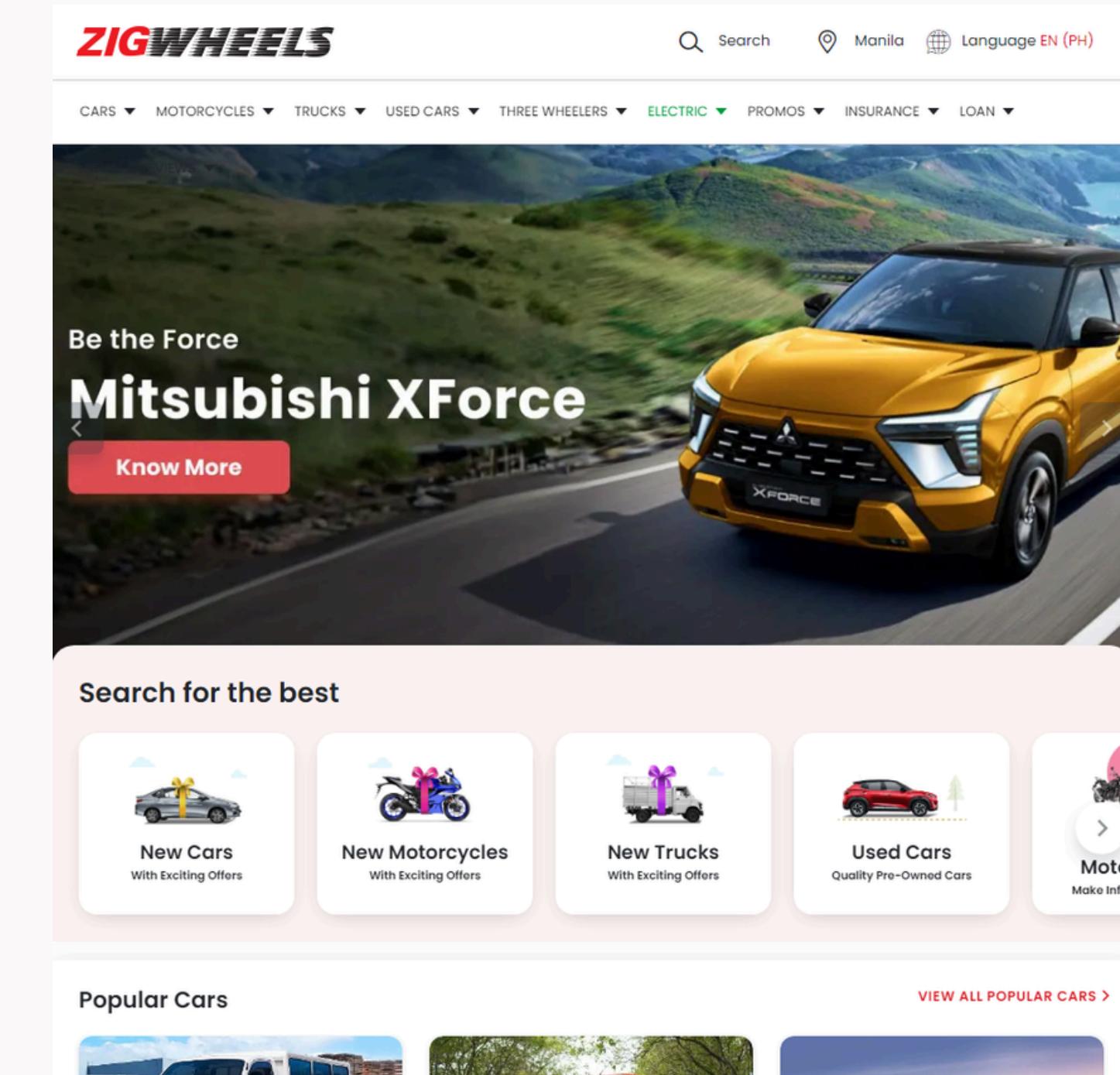
- Types of Car:** Used Car, Certified Pre-Owned, Repossessed.
- Nearly New / Premium Car:** Premium Car, Nearly New.
- Select Any Make:** A list of car brands including Abarth, Alfa Romeo, Aston Martin, Audi, BAIC, and Bentley.
- Select Any Model:** A dropdown menu.
- Select Any Year:** A dropdown menu.
- Location:** A text input field with placeholder text "Type a City, Province or Region".

In the center, the main content area displays the following information:

- 335 CERTIFIED PRE-OWNED, REPOSESSED AND USED CARS FOR SALE IN THE PHILIPPINES**
- A descriptive text: "Looking for a used car in the Philippines? The AutoDeal Used Car section contains vehicles from private sellers, used car dealers, certified pre-owned, and repossessed from banks. Narrow your search f... [Read More](#)"
- Your current search criteria: **Certified Pre-Owned ×**, **Repossessed ×**, **Used Car ×**
- Sort By: Relevance Show Hot Deals
- A "Try Car Advisor" button with a small icon.
- A large image of a white SUV parked in front of a building.

2. ZIGWHEELS WEBSITE: USED CARS

For the second primary website to web scrape from (<https://www.zigwheels.ph/>), the functionalities of the website is similar to AutoDeal. Using the top bar filter, clicking “Used Cars” drop down button, and clicking “Buy Used Cars” redirects the user to the used cars listing available to buy in the website.



2. ZIGWHEELS WEBSITE: USED CARS

After clicking those filters, it should display the image shown below.

The screenshot shows the Zigwheels website interface. At the top, there's a navigation bar with the Zigwheels logo, a search bar, and a wishlist icon. Below the navigation is a banner titled "Benefits of buying used cars from Zigwheels" featuring a woman in a yellow shirt holding a car key next to a silver car. The banner includes two red buttons: "Know More" and "Learn More". The main content area has a heading "Used & 2nd hand Cars for Sale Philippines" and a subtext about popular brands and models. Below this is a section titled "Newly Added Used Cars" displaying four car listings:

- 2015 Nissan X-Trail • 40,000 Km • Gasoline • Cvt
- 2010 Mitsubishi Montero S... • 140,000 Km • Diesel • Manual
- 2016 Honda Mobilio • 72,000 Km • Gasoline • Cvt
- 2020 Honda Civic • 29,000 Km • Gasoline • Cvt

2. ZIGWHEELS WEBSITE: USED CARS

Scroll down and click “View All Used Cars Under P100,000”.

The screenshot shows the ZigWheels website interface for used cars. At the top, there's a navigation bar with categories: CARS, MOTORCYCLES, TRUCKS, USED CARS, THREE WHEELERS, ELECTRIC, PROMOS, INSURANCE, LOAN, and NEWS & REVIEWS. Below the navigation, there are four featured car listings:

- 2015 Nissan X-Trail: 40,000 Km • Gasoline • Cvt, ₱568,000. [View Seller Details](#)
- 2010 Mitsubishi Montero S...: 140,000 Km • Diesel • Manual, ₱470,000. [View Seller Details](#)
- 2016 Honda Mobilio: 72,000 Km • Gasoline • Cvt, ₱565,000. [View Seller Details](#)
- 2020 Honda Civic: 29,000 Km • Gasoline • Cvt, ₱988,000. [View Seller Details](#)

Below these, there's a section titled "Choose Used Car by Budget" with a button labeled "Under ₱100,000" which is highlighted in red. Other budget options include Under ₱150,000, Under ₱200,000, Under ₱250,000, and Under ₱300,000. To the right of this section is a button labeled "VIEW ALL USED CARS UNDER ₱100,000 >" with navigation arrows.

Further down, there are four more car listings:

- 2000 Honda Accord: 60,000 Km • Gasoline • Cvt, ₱80,000. [View Seller Details](#)
- 1994 Nissan Vanette: 84,710 Km • Gasoline • Manual, ₱70,000. [View Seller Details](#)
- 1997 Mazda 3 Sedan: 90,000 Km • Gasoline • Automatic, ₱75,000. [View Seller Details](#)
- 2021 Suzuki Celerio: 12,000 Km • Gasoline • Cvt, ₱60,000. [View Seller Details](#)

At the bottom, there's a section titled "Used Cars by Body Type" with a "Crossover" filter selected. Other body type options include SUV, MPV, Sedan, and Pickup Trucks. There are four crossover car thumbnails shown:

- A white Nissan X-Trail
- A red Ford Ecosport
- A grey Honda HR-V
- A grey Ford EcoSport

On the far right of the page, there's a red circular arrow icon with a white upward-pointing arrow.

2. ZIGWHEELS WEBSITE: USED CARS

It will redirect the user to a webpage that contains useful filters on the left tab that can help control the price range, city where it is sold, brand, etc.

The screenshot shows the ZigWheels website interface for finding used cars. The top navigation bar includes links for CARS, MOTORCYCLES, TRUCKS, USED CARS, THREE WHEELERS, ELECTRIC, PROMOS, INSURANCE, LOAN, and NEWS & REVIEWS. The main search bar has a 'Search' button and a 'Wishlist' icon. On the left, there are three filter panels: 'Price' (set to Under ₱100,000), 'Province' (search bar and dropdown for All City, Quezon City, Bacoor, Cebu City, Pasig), and 'Brand' (dropdown). The main content area displays a heading 'Used Cars for Sale Under ₱100,000 Philippines' with a note about 11 available cars. A car listing for a 2000 Honda Accord is shown with a price of ₱80,000. Below this, a section titled 'View Used Cars By Brand' lists logos for Honda, Nissan, Mazda, Mitsubishi, Suzuki, Toyota, and Tata. At the bottom, there are three small images of different vehicles.

ZIGWHEELS

CARS ▾ MOTORCYCLES ▾ TRUCKS ▾ USED CARS ▾ THREE WHEELERS ▾ ELECTRIC ▾ PROMOS ▾ INSURANCE ▾ LOAN ▾ NEWS & REVIEWS ▾

Used Cars for Sale Under ₱100,000 Philippines

As on Aug 06, 2024 there are 11 Used Under ₱100,000 cars available for sale in the Philippines. The most popular 2nd hand Under ₱100,000 car brands available for sale in Philippines are Honda, Nissan, Mazda, Mitsubishi and Suzuki and... [Read More](#)

Under ₱100,000 [Clear All](#) [Relevance](#)

Price

₱10,000 ₱100,000

Or What is your price range?

Under ₱100,000 9
 Under ₱150,000 6
 Under ₱200,000 18
 Under ₱250,000 29
 Under ₱300,000 63

Province

Q. City or Province

All City

Quezon City 2
 Bacoor 1
 Cebu City 1
 Pasig 1

Brand

Honda Nissan Mazda Mitsubishi Suzuki Toyota Tata

2000 Honda Accord 60,000 Km • Gasoline • Cvt ₱80,000

Dasmariñas 5 seat 1498 CC Accessory Pcs

[View Seller Details](#)

Compare Compare Compare

IV. EXTERNAL WEBSITES DESCRIPTION

1. AUTODEAL WEBSITE: NEW CARS

Using the same website of AutoDeal (<https://www.autodeal.com.ph/>). However in this case, instead of using “Used Cars” filter, “New Cars” button is used to extract the brand new cars.

The screenshot shows the AutoDeal website's homepage. At the top, there is a navigation bar with the AD logo and links for New Cars, Car Promos, Used Cars, News & Reviews, Loans & Insurance, and Comparison. Below the navigation, a large banner features the text "Car Buying, Simplified." and "Compare prices on new & used cars from official dealers across the Philippines." It includes two buttons: "Search New Cars" (highlighted in orange) and "Search Used Cars". The main content area displays a promotional offer for the MG ZS MCE COM CVT. It features a bright orange MG ZS SUV against an orange background. To the right of the car, a circular badge says "2024 BEST DEAL" and "AUTODEAL.COM.PH". Next to the badge, the text "Low Monthly PHP 19K" is displayed in large, bold letters. Below this, a smaller text states: "Available in all MG dealerships nationwide, the MG ZS MCE COM CVT is priced at P918,888." At the bottom of the offer section, the model name "MG ZS MCE COM CVT" is written. At the very bottom of the page, there are several small buttons: "The Mitsubishi XForce", "MG ZS MCE Promo" (which is highlighted in orange), "Nissan Terra", "AutoDeal App", and "Car Insurance".

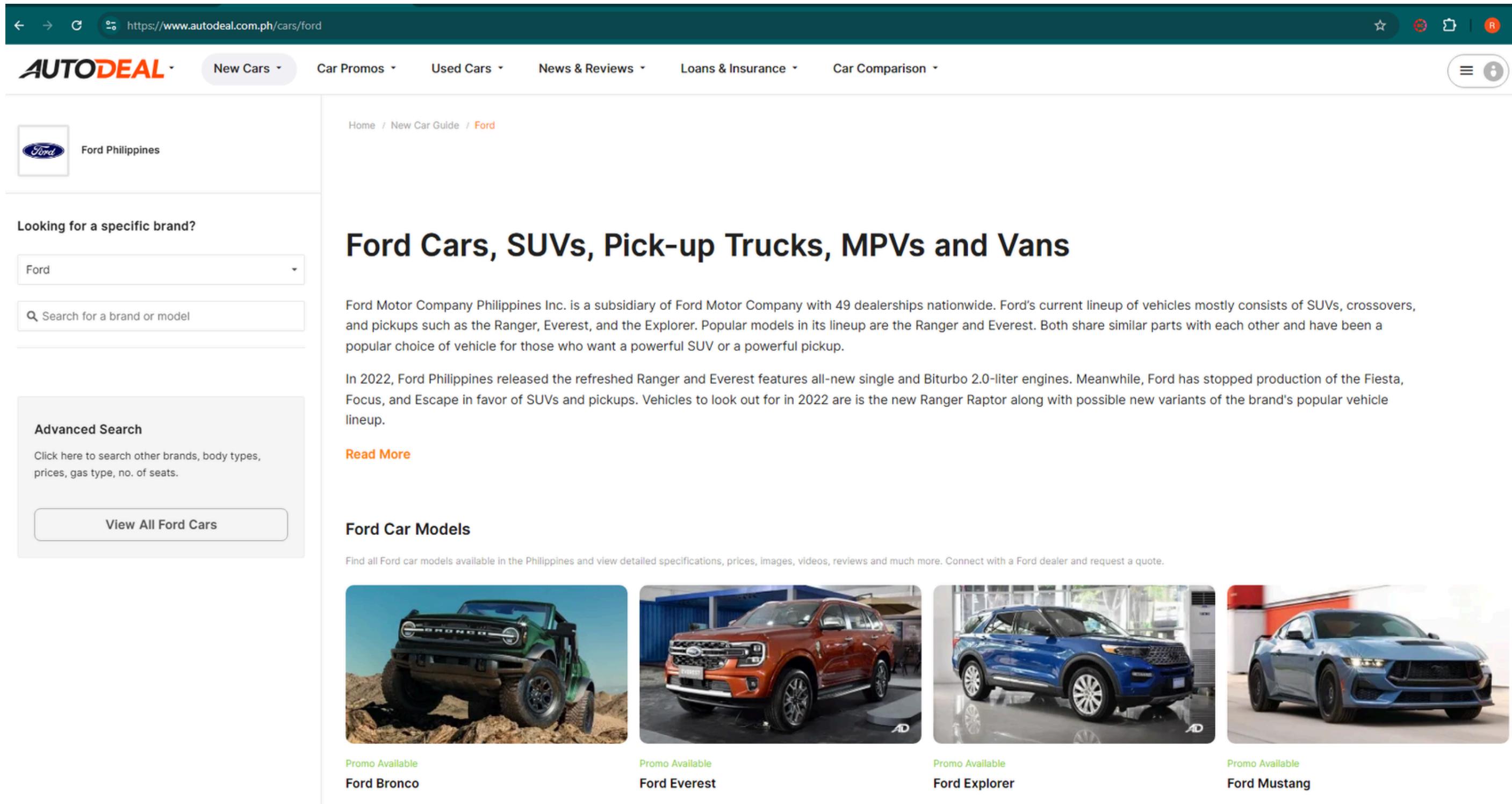
1. AUTODEAL WEBSITE: NEW CARS

From there, it displays car brands from which a user can choose from.

The screenshot shows the homepage of Autodeal.com.ph. At the top, the URL 'autodeal.com.ph/cars' is visible in the browser's address bar. Below the header, there are several navigation links: 'New Cars', 'Car Promos', 'Used Cars', 'News & Reviews', 'Loans & Insurance', and 'Car Comparison'. A search bar with the placeholder 'Search for a brand or model' is positioned below these links. The main content area features a section titled 'Search New Cars in the Philippines' with a sub-instruction: 'Not too sure what car to buy? Let AutoDeal.com.ph help you by searching car brand or the body type you're interested in. You can also narrow your search further by setting a price range.' Below this, there are three rows of car brand logos. The first row includes Ford, Nissan, Mitsubishi, MG, and Hyundai. The second row includes Mazda, Isuzu, GAC Motor, Toyota, and Suzuki. The third row includes Kia, Changan, Subaru, GWM, and Peugeot. Each logo is accompanied by its brand name. At the bottom of the grid, there is a horizontal ellipsis consisting of five small dots.

1. AUTODEAL WEBSITE: NEW CARS

As an example, this displays after clicking Ford.



The screenshot shows the Autodeal website's Ford section. At the top, there's a navigation bar with links for New Cars, Car Promos, Used Cars, News & Reviews, Loans & Insurance, and Car Comparison. Below the navigation, there's a sidebar on the left with a Ford logo and the text "Ford Philippines". It also includes a dropdown menu for "Looking for a specific brand?" set to "Ford", a search bar with placeholder text "Search for a brand or model", and a link to "Advanced Search" which says "Click here to search other brands, body types, prices, gas type, no. of seats." There's a button labeled "View All Ford Cars". The main content area has a heading "Ford Cars, SUVs, Pick-up Trucks, MPVs and Vans". It contains a paragraph about Ford Motor Company Philippines Inc., mentioning its 49 dealerships nationwide and popular models like the Ranger, Everest, and Explorer. It also notes that Ford has stopped production of the Fiesta, Focus, and Escape. A "Read More" link is present. Below this, there's a section titled "Ford Car Models" with a subtext about finding detailed specifications, prices, images, videos, reviews, and quotes. Four Ford vehicles are shown in thumbnail images: Ford Bronco (green), Ford Everest (orange), Ford Explorer (blue), and Ford Mustang (blue). Each thumbnail has a "Promo Available" label below it.

Ford Philippines

Looking for a specific brand?

Ford

Search for a brand or model

Advanced Search

Click here to search other brands, body types, prices, gas type, no. of seats.

View All Ford Cars

Home / New Car Guide / Ford

Ford Cars, SUVs, Pick-up Trucks, MPVs and Vans

Ford Motor Company Philippines Inc. is a subsidiary of Ford Motor Company with 49 dealerships nationwide. Ford's current lineup of vehicles mostly consists of SUVs, crossovers, and pickups such as the Ranger, Everest, and the Explorer. Popular models in its lineup are the Ranger and Everest. Both share similar parts with each other and have been a popular choice of vehicle for those who want a powerful SUV or a powerful pickup.

In 2022, Ford Philippines released the refreshed Ranger and Everest features all-new single and Biturbo 2.0-liter engines. Meanwhile, Ford has stopped production of the Fiesta, Focus, and Escape in favor of SUVs and pickups. Vehicles to look out for in 2022 are the new Ranger Raptor along with possible new variants of the brand's popular vehicle lineup.

[Read More](#)

Ford Car Models

Find all Ford car models available in the Philippines and view detailed specifications, prices, images, videos, reviews and much more. Connect with a Ford dealer and request a quote.

Promo Available

Ford Bronco

Promo Available

Ford Everest

Promo Available

Ford Explorer

Promo Available

Ford Mustang

2. ZIGWHEELS WEBSITE: CAR SERVICE CENTERS

Using the same website Zigwheels (<https://www.zigwheels.ph/>). After visiting the website, click “CARS” drop-down button, and select “Car Service Center”.

The screenshot shows the Zigwheels website homepage. At the top, there is a navigation bar with the Zigwheels logo, a search bar, and location and language settings. Below the navigation bar, there is a main menu with categories like CARS, MOTORCYCLES, TRUCKS, USED CARS, THREE WHEELERS, ELECTRIC, PROMOS, INSURANCE, and LOAN. A large banner features a blue BYD Atto 3 SUV driving on a road, with the text "IMPRESSIVE & DYNAMIC" and "BYD Atto 3". A red "Know More" button is visible. Below the banner, there is a section titled "Planning to buy a new Car?" with a brief description and a "Read More" link. At the bottom, there is a "Popular Cars" section with three thumbnail images of vehicles: a white van, a red hatchback, and a silver sedan, along with a "VIEW ALL POPULAR CARS >" link.

2. ZIGWHEELS WEBSITE: CAR SERVICE CENTERS

It should redirect the user to this web page where a user can select any car brands and city to view the available car service centers.

The screenshot shows the ZigWheels website's homepage with a focus on car service centers. At the top, the ZigWheels logo is on the left, and a search bar with 'Search' and location options for 'Manila' and 'Language EN (PH)' are on the right. Below the header, a banner features three people in professional attire (two men and one woman) shaking hands in front of a blurred background of a car showroom or service center. The banner text reads 'Find Car Service Centers in Philippines' and 'We will help you to find a service center near you.' Below the banner is a search form titled 'Car Service Centers in Philippines' with fields for 'Select Brand' (dropdown), 'Enter your city' (text input), and a red 'Search' button. At the bottom, there is a section titled 'Car Service Centers by Popular Brands' featuring logos for Toyota, Honda, Mitsubishi, Suzuki, Nissan, Isuzu, Hyundai, Ford, Mazda, Chevrolet, and Kia.

V. DATA PREPARATION PROCESS DOCUMENTATION

1. DATA COLLECTION

A. PRIMARY DATASETS

- AutoDeal Used Cars
- Zigwheels Used Cars

PRIMARY DATASET
CONCATENATE

B. EXTERNAL DATASETS

- AutoDeal Brand New Cars
- Zigwheels Car Service Centers

A. PRIMARY DATASETS

- AutoDeal Used Cars
- Zigwheels Used Cars

AutoDeal Used Cars

AutoDeal Website

Used Cars ▾ News & Reviews ▾ Loans & Insurance ▾ Car Comparison ▾

Contact Dealer Add To Comparison

USED CAR / FEATURED LISTING
2021 Nissan NV350 Urvan 2.5L MT Diesel
P930,000
30,000 Km Manual Diesel
Car Empire Las Pinas City, Metro Manila, NCR Dealer
Contact Dealer Add To Comparison

USED CAR / FEATURED LISTING
2013 Toyota Rav4 (4X4) AT
P595,000
70,000 Km Automatic Gasoline
Car Empire Las Pinas City, Metro Manila, NCR Dealer
Contact Dealer Add To Comparison

Prev | 1 2 3 4 5 6 7 8 9 ... 17 | Next

Web scraped dataset

	Make	Model	Price	Vehicle	Fuel	Transmission	Mileage	Engine	Body
0	Suzuki	S-Presso	415000	2021	Gasoline	Manual	30,000	1.0	Hatchback
1	Ford	Everest	930000	2017	Diesel	Automatic	60,000	2.2	SUV
2	Honda	Civic	760000	2017	Gasoline	Automatic	50,000	1.8	Sedan
3	Toyota	Innova	1120000	2023	Diesel	Automatic	17,000	2.8	MPV
4	Toyota	Camry	520000	2013	Gasoline	Automatic	60,000	2.5	Sedan
...
314	Suzuki	Ertiga	600000	2021	Gasoline	Manual	2,500	NaN	MPV
315	Kia	Sportage	950000	2017	Gasoline	Automatic	38,971	2.0	Crossover
316	Hyundai	Accent	385000	2016	Diesel	Automatic	105,000	1.6	Sedan
317	Mazda	CX-5	520000	2012	Gasoline	Automatic	48,000	2.0	SUV
318	Foton	Gratour	350000	2019	Gasoline	Manual	17,000	1.5	Minivan

319 rows × 9 columns

See autoDealWebScraping.ipynb
for the codes.

AutoDeal Used Cars

Upon observing and seeing patterns from AutoDeal, we noticed that after clicking the filter “Used Cars”, the website has a page number in the url that iterates upon clicking onto the succeeding pages (see the **bold** text below).

<https://www.autodeal.com.ph/used-cars/search/certified-pre-owned+repossessed+used-car-status/page-1?sort-by=relevance>

With that, the for loop was utilized to ensure that after collecting all the cars and its data from each page, it moves on to the next. After getting all the cards from the cars in each page, we took the necessary car features and proceeded to contain them in a Pandas data frame, and saved as csv.

Zigwheels Used Cars

Zigwheels Website

ZIGWHEELS

CARS ▾ MOTORCYCLES ▾ TRUCKS ▾ USED CARS ▾ THREE WHEELERS ▾ ELECTRIC ▾ PROMOS ▾ INSURANCE ▾ LOAN ▾ NEWS & REVIEWS ▾

[Collapse All Filters](#)

Price

₱200,001 ₱250,000

Or What is your price range?

- Under ₱100,000 9
- Under ₱150,000 6
- Under ₱200,000 18
- Under ₱250,000 29
- Under ₱300,000 63

Province

Q. City or Province

All City

- Quezon City 10
- Antipolo, Rizal 3
- Cebu City 3
- Mandaluyong City, Metro Manila 3

Second Hand Cars for Sale in Philippines

As on Aug 05, 2024 there are 42 Used cars available for sale in the Philippines. The most popular 2nd hand car brands available for sale in Philippines are Honda, Hyundai, Toyota, Ford and Mitsubishi and the most popular second hand car brands are Toyota, Honda, Suzuki, Ford, and Mitsubishi.

[Read More](#)

Clear All Relevance

Compare 
2012 Toyota Vios
218,600 Km • Gasoline • Manual
₱220,000
Davao City 1293 CC

Compare 
2007 Suzuki APV
109,724 Km • Gasoline • Manual
₱220,000
Antipolo

Compare 
2008 Toyota Vios
74,000 Km • Gasoline • Manual
₱250,000
San Jose del Monte 1329 CC

View Seller Details

View Seller Details

View Seller Details

Compare 
2008 Toyota Vios
115,000 Km • Gasoline • Cvt
₱220,000

Compare 
2004 Toyota Corolla Altis
115,000 Km • Gasoline • Cvt
₱240,000

Compare 
2008 Honda City
90,000 Km • Gasoline • Manual
₱220,000

View Seller Details

View Seller Details

View Seller Details

Zigwheels Used Cars

Problems

The data from the website are categorized by a certain price range, i.e. Under PHP100,000, above PHP40,000,000 etc. We had trouble in scraping the data from this Website due to several reasons:

- In each webpage, there are initially 21 class of car cards in which we can scrape the data from and we have to click the “View More” button in order for the other car cards to pop-up. However, with a certain price range, say PHP 1 Million - PHP 1.1 Million, there are 116 cars under that price range. Because of the “View More” button, it restricts us to from getting the 95 car data.
- Frequency of cars that have the same number of resale price also affects our data extraction. If there are more than 21 cars that has the same resale price, we are forced to extract the first 21 cars only in the page.

Zigwheels Used Cars

Solutions

1. "View More" Button

We overcame this problem by implementing Selenium in the code. We learned it through this YouTube tutorial:

<https://www.youtube.com/watch?v=Dt66BsSIIgM>

For the Action class and other debugging in the code was solved by simply searching on the internet - specifically the ones posted in stackoverflow.com/. However, when we implemented this method, it seemed that the View More button isn't a button class and always produced an error in which it says that another button or class is receiving the clicks instead of the View More. We didn't get to solve this issue but upon further research, we assume that it has something to do with hidden elements or wrapper classes.

Zigwheels Used Cars

View More Button

View Used Cars By Model

Mitsubishi Montero Sport Toyota Fortuner Toyota Hiace Toyota Innova Ford Everest Ford Ranger



2017 Toyota Fortuner 
50,000 Km • Diesel • Automatic
₱1.1 Million

Manila 7 seat 2393 CC Voice Control

[View Seller Details](#)



2016 Mitsubishi Montero Sport 
38,900 Km • Diesel • Automatic
₱1.1 Million

Las Piñas 2398 CC Hill-Start Assist Control

[View Seller Details](#)



2018 Toyota Fortuner 
35,000 Km • Diesel • Manual
₱1.065 Million

Tanay 7 seat 2393 CC Accessory Power C

[View Seller Details](#)

VIEW MORE ▾

Automotive News and Reviews

News Feature Stories Road Test Used Car News

Zigwheels Used Cars

Solutions

1. Alternative Solution for “View More” Button

We realized that, we can manually adjust the resale price range in zigwheels by applying this format in the URL:

<https://www.zigwheels.ph/used-cars?p=1000001-1100000>

Where, 1000001 is the lower bound and 1100000 is the upper bound of the price range. After this, we will sort the car prices from low to high using Selenium and Action class for clicking the Relevance → Price Low to High buttons. From there, we extract 21 car data at a time and get the price of the last car extracted to set as the new lower bound for the webpage URL. In this way, we were able to maximize the number of data we could extract while minimizing the duplicates during the webscraping process.

Zigwheels Used Cars

Sorting Method

Second Hand Cars for Sale in Philippines

As on Aug 05, 2024 there are 116 Used cars available for sale in the Philippines. The most popular 2nd hand car brands available for sale in Philippines are Toyota, Mitsubishi, Ford, Nissan and Hyundai and the most popular second hand car brands are Toyota, Mitsubishi, Ford, Nissan and Hyundai.

₱1 - 1.1 Million Clear All



Compare

2019 Toyota Fortuner

55,000 Km • Diesel • Manual

₱1.1 Million

Butuan 7 seat 2393 CC Voice Control

[View Seller Details](#)



Compare

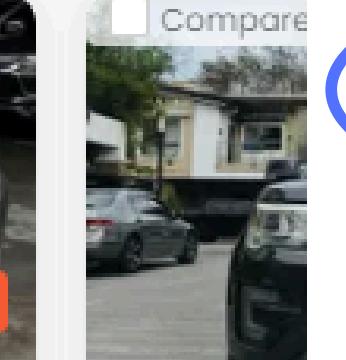
2017 Mitsubishi Montero Sport

40,000 Km • Diesel • Automatic

₱1.02 Million

Las Piñas 7 seat 2477 CC Accessory Power

[View Seller Details](#)



Compare

2016 Ford

18,000 Km • Diesel • Manual

₱1.095 M

Pasig 7 seat 2296 CC Leather Seats

[View Seller Details](#)

1 Relevance

2 Price-Low to High

Price-High to Low

Kilometer-Low to High

Year-High to Low

Year-Low to High

Newest

Zigwheels Used Cars

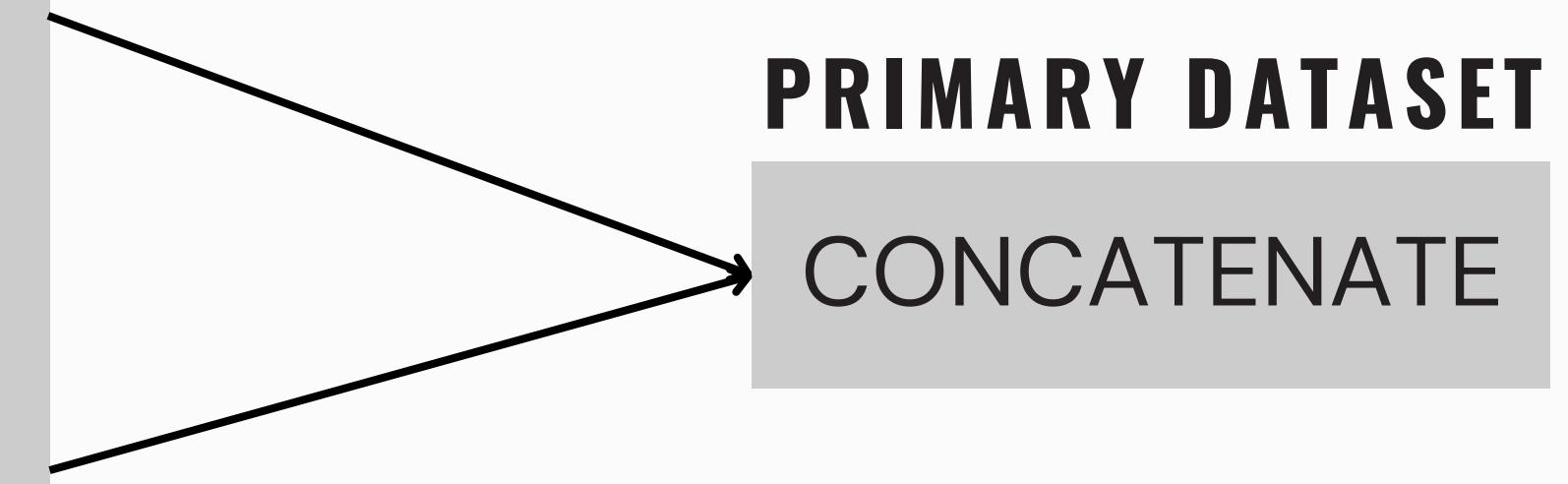
Web Scrapped Dataset

	model	price	registration_year	make_year	fuel_type	body_type	kms_driven	ownership	transmission	city	seats	engine_displacement
0	1990 Nissan Sentra 1.6L SGX MT Detail	₱45,000	1990.0	1990	Gasoline	Sedan	61,233 Km	First Owner	Manual	Quezon City	Nan	Nan
1	2021 Suzuki Celerio CVT 1.0L Detail	₱60,000	2024.0	2021	Gasoline	Hatchback	12,000 Km	First Owner	Cvt	Pasig	5 seat	998 CC
2	1994 Nissan Vanette 2.0L Vanette MT Detail	₱70,000	2023.0	1994	Gasoline	Van	84,710 Km	Second Owner	Manual	Quezon City	Nan	Nan
3	2000 Nissan Sentra 1.6L Super Saloon MT Detail	₱70,000	2024.0	2000	Gasoline	Sedan	150,000 Km	Second Owner	Manual	Mandaluyong City	Nan	Nan
4	1997 Mazda 3 Sedan 2.0L Premium Detail	₱75,000	2019.0	1997	Gasoline	Sedan	90,000 Km	Third Owner	Automatic	sanfernando	5 seat	1998 CC
...
2232	2018 Rolls-Royce Ghost 6.6L Detail	₱18.5 Million	2018.0	2018	Gasoline	Sedan	20,000 Km	First Owner	Automatic	Manila	Nan	Nan
2233	2022 Cadillac Escalade ESV Platinum Detail	₱19 Million	2022.0	2022	Gasoline	Suv	10 Km	First Owner	Automatic	Manila	7 seat	5998 CC
2234	2024 Ferrari F8 Tributo 3.9L Detail	₱20.5 Million	2024.0	2024	Gasoline	Coupe	16,000 Km	First Owner	Automatic	Imus	2 seat	3902 CC
2235	2023 Bentley Bentayga V8 Detail	₱21 Million	2023.0	2023	Gasoline	Suv	30,000 Km	First Owner	Automatic	Olongapo City	5 seat	3996 CC
2236	2023 Lamborghini Urus 4.0 L Detail	₱29.5 Million	2023.0	2023	Gasoline	Suv	1,300 Km	First Owner	Automatic	Muntinlupa	5 seat	3996 CC

See zigwheelsWebScraping.ipynb
for the codes.

A. PRIMARY DATASETS

- AutoDeal Used Cars
- Zigwheels Used Cars



PRIMARY DATASET CONCATENATION

AutoDeal Used Cars

'primaryData1: 319'

	Make	Model	Price	Vehicle	Fuel	Transmission	Mileage	Engine	Body
0	Suzuki	S-Presso	415000	2021	Gasoline	Manual	30,000	1.0	Hatchback
1	Ford	Everest	930000	2017	Diesel	Automatic	60,000	2.2	SUV
2	Honda	Civic	760000	2017	Gasoline	Automatic	50,000	1.8	Sedan
3	Toyota	Innova	1120000	2023	Diesel	Automatic	17,000	2.8	MPV
4	Toyota	Camry	520000	2013	Gasoline	Automatic	60,000	2.5	Sedan

Zigwheels Used Cars

'primaryData2: 2237'

	Unnamed: 0	model	price	registration_year	make_year	fuel_type	body_type	kms_driven	ownership	transmission	city	seats	engine_displacement
0	0	1990 Nissan Sentra 1.6L SGX MT Detail	₱45,000	1990.0	1990	Gasoline	Sedan	61,233 Km	First Owner	Manual	Quezon City	Nan	Nan
1	1	2021 Suzuki Celerio CVT 1.0L Detail	₱60,000	2024.0	2021	Gasoline	Hatchback	12,000 Km	First Owner	Cvt	Pasig	5 seat	998 CC
2	2	1994 Nissan Vanette 2.0L Vanette MT Detail	₱70,000	2023.0	1994	Gasoline	Van	84,710 Km	Second Owner	Manual	Quezon City	Nan	Nan
3	3	2000 Nissan Sentra 1.6L Super Saloon MT Detail	₱70,000	2024.0	2000	Gasoline	Sedan	150,000 Km	Second Owner	Manual	Mandaluyong City	Nan	Nan
4	4	1997 Mazda 3 Sedan 2.0L Premium Detail	₱75,000	2019.0	1997	Gasoline	Sedan	90,000 Km	Third Owner	Automatic	sanfernando	5 seat	1998 CC

PRIMARY DATASET CONCATENATION

A little bit cleaning, rearrangement, and renaming was done for Zigwheels dataset to match the column names of AutoDeal dataset to concatenate the two (See Concatenation and Rearrangement.ipynb for full detail of the procedure).

Afterwards, we now have the concatenated dataset (see the screenshot).

	make	model	make_year	fuel_type	transmission	kms_driven	engine_size	body_type	price
0	Suzuki	S-Presso	2021	Gasoline	Manual	30,000	1.000	Hatchback	415000
1	Ford	Everest	2017	Diesel	Automatic	60,000	2.200	SUV	930000
2	Honda	Civic	2017	Gasoline	Automatic	50,000	1.800	Sedan	760000
3	Toyota	Innova	2023	Diesel	Automatic	17,000	2.800	MPV	1120000
4	Toyota	Camry	2013	Gasoline	Automatic	60,000	2.500	Sedan	520000
...
2232	Rolls-Royce	Ghost	2018	Gasoline	Automatic	20,000 Km	6.600	Sedan	₱18.5 Million
2233	Cadillac	Escalade	2022	Gasoline	Automatic	10 Km	5.998	Suv	₱19 Million
2234	Ferrari	F8	2024	Gasoline	Automatic	16,000 Km	3.900	Coupe	₱20.5 Million
2235	Bentley	Bentayga	2023	Gasoline	Automatic	30,000 Km	3.996	Suv	₱21 Million
2236	Lamborghini	Urus	2023	Gasoline	Automatic	1,300 Km	4.000	Suv	₱29.5 Million

2444 rows × 9 columns

PRIMARY DATASET

The following columns that make up the primary dataset were chosen to be mostly consistent with the datasets used in the aforementioned sources, with a few minor additions, and insignificant removals:

Independent Variables

- make_year – Year the car was manufactured
- make – manufacturer of the vehicle
- model – model of the vehicle
- transmission – type of transmission (eg. manual, automatic, etc.)
- body_type – eg. sedan, hatchback, pick-up, SUV, etc.
- fuel_type – eg. gasoline, diesel, electric, etc.
- kms_driven – distance travelled since first purchase in kilometers
- engine_size – volume of engine in liters

Dependent Variable

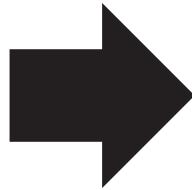
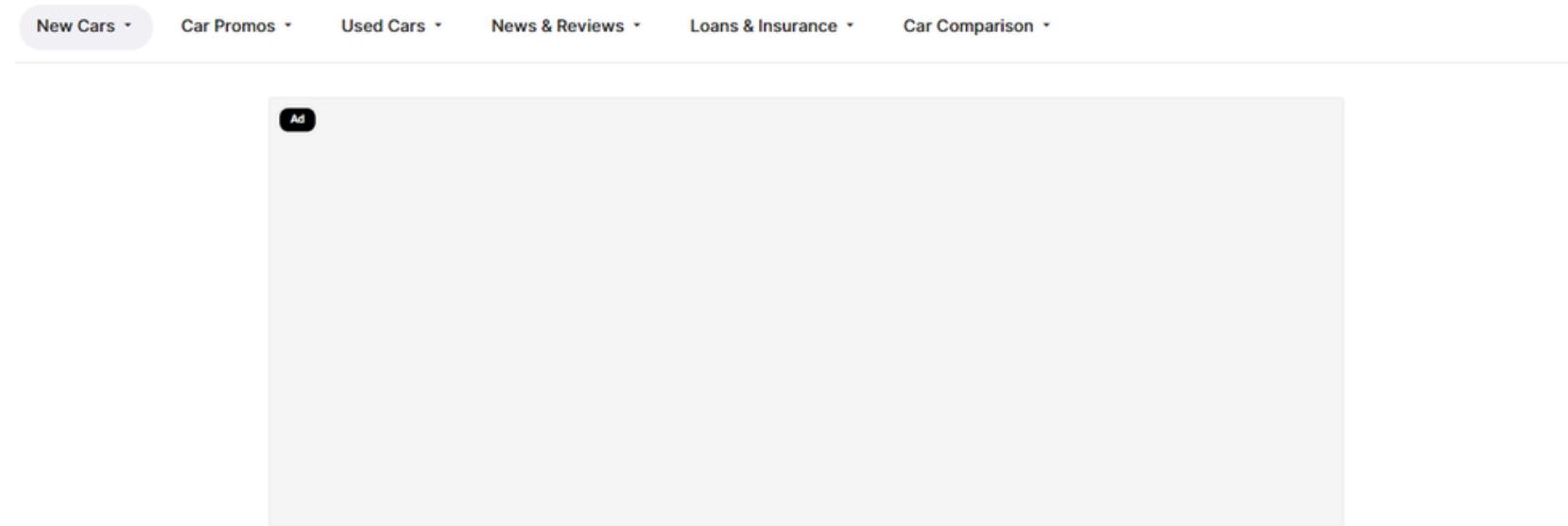
- resale_price – selling price on website as second-hand car in PHP (TARGET)

B. EXTERNAL DATASETS

- AutoDeal Brand New Cars
- Zigwheels Car Service Centers

AutoDeal Brand New Cars

AutoDeal Website



Webscraped Dataset

	Make	Model	Year	SRP
0	Suzuki	APV	2024	743000
1	Suzuki	Carry	2024	614000
2	Suzuki	Celerio	2024	708000
3	Suzuki	Dzire	2024	609000
4	Suzuki	Ertiga	2024	954000
...
103	Foton	Tornado	2024	915000
104	Foton	Transvan	2024	1290000
105	Foton	Traveller	2024	1330000
106	Foton	Tunland	2024	1758000
107	Foton	Tunland	2024	1788000

223 rows × 4 columns

See autoDealExternal.ipynb for the codes.

AutoDeal Brand New Cars

We decided to make use of the price of brand new cars from AutoDeal to merge it with the primary dataset so that we have an additional column called SRP (price of a brand new car) that could provide a valuable use in comparison to the resale price from the used cars dataset. To make the data collection faster, we looked for the unique Makes and stored it in a list, then iterated through it to only web scrape the Makes that are in the primary dataset.

```
uniqueMakes = primary['Make'].unique()  
print(uniqueMakes)  
i = 1  
for make in uniqueMakes:  
    print(f'{i}. {make}')  
    i += 1
```

Code 1. uniqueMakes

```
card_listings = []  
for i in range(11, len(uniqueMakes)):  
    print(uniqueMakes[i])  
  
    parent_url =  
f"https://www.autodeal.com.ph/cars/{uniqueMakes[i]}"  
"  
    print(parent_url)
```

Code 2. uniqueMakes Filter

AutoDeal Brand New Cars

There was a problem during web scraping that stopped the code because one car had no listed price that caused an error. A simple solution we made is that we web scraped first from Suzuki to Kia, then Isuzu to Foton to avoid the error.

	Make	Model	Year	SRP
0	Suzuki	APV	2024	743000
1	Suzuki	Carry	2024	614000
2	Suzuki	Celerio	2024	708000
3	Suzuki	Dzire	2024	609000
4	Suzuki	Ertiga	2024	954000
...
110	Kia	K2500	2024	1198000
111	Kia	Seltos	2024	1198000
112	Kia	Soluto	2024	780000
113	Kia	Sonet	2024	758000
114	Kia	Sorento	2024	2448000
115 rows × 4 columns				

	Make	Model	Year	SRP
0	Isuzu	C-Series	2024	NaN
1	Isuzu	D-MAX	2024	938000
2	Isuzu	E-Series	2024	NaN
3	Isuzu	F-Series	2024	1570000
4	Isuzu	mu-X	2024	1675000
...
103	Foton	Tornado	2024	915000
104	Foton	Transvan	2024	1290000
105	Foton	Traveller	2024	1330000
106	Foton	Tunland	2024	1758000
107	Foton	Tunland	2024	1788000
108 rows × 4 columns				

AutoDeal Brand New Cars

Then, we concatenated the two to finally get the first external dataset and saved as a csv file, autoDealExternal.csv in folder 2.

	Make	Model	Year	SRP
0	Suzuki	APV	2024	743000
1	Suzuki	Carry	2024	614000
2	Suzuki	Celerio	2024	708000
3	Suzuki	Dzire	2024	609000
4	Suzuki	Ertiga	2024	954000
...
103	Foton	Tornado	2024	915000
104	Foton	Transvan	2024	1290000
105	Foton	Traveller	2024	1330000
106	Foton	Tunland	2024	1758000
107	Foton	Tunland	2024	1788000
223 rows × 4 columns				

AutoDeal Brand New Cars

External Dataset 1

Under the assumption that the current SRP of a car of the same make and model of an older, to-be-resold car has some bearing on the latter's value/pricing, the resulting first external dataset consists of the following columns:

- Make – manufacturer
- Model – (self explanatory)
- Year – year of manufacture (all 2024, the current year when the resale prices were set), later renamed to ext_make_year
- SRP – corresponding initial price as indexed by the above three columns

The scope of said first two columns was derived from the corresponding scope in the primary dataset, and these two columns were concatenated into a third “key” column on which this external dataset was merged with the primary, on a similar “key” column.

Zigwheels Car Service Centers

Zigwheels Website

The screenshot shows a search interface for Toyota service centers. At the top, it says "Toyota Car Service Centers in Philippines" and "will help you find the best Toyota Cars service centers". Below this is a large image of two people in business attire shaking hands. The search form has "Select Brand" dropdown set to "Toyota", an "Enter your city" input field, and a red "Search" button. Below the search form is a section titled "Toyota Car Service Centers in Popular Cities" with four buttons: "Quezon City 3 Toyota Car Service Centers", "Makati City 2 Toyota Car Service Centers", "Cebu City 1 Toyota Car Service Centers", and "Manila 2 Toyota Car Service Centers".

There is no notebook for this because the data collection for this dataset is done manually. This was saved as a csv file, make_mshops.csv in the datasets folder.

Dataset

	make	no_sc_mnn1
0	Suzuki	9
1	Ford	2
2	Honda	4
3	Toyota	9
4	Hyundai	6
5	Mitsubishi	4
6	Mazda	1
7	Geely	4
8	Nissan	2
9	Chevrolet	4
10	Isuzu	6
11	Subaru	6
12	BMW	4
13	Mercedes-Benz	1
14	Kia	6
15	MG	4
16	MINI	1
17	Chery	2
18	Lexus	1
19	Foton	0
20	Volkswagen	0
21	Audi	1
22	Jeep	0

Zigwheels Car Service Centers

External Dataset 2

Though no source had yet used such data for prediction, it seems plausible that the number of service centers in Metro Manila for each manufacturer is a relevant factor to pricing, since the access to maintenance facilities is a likely concern for consumers, especially for second-hand cars. Thus, the second external dataset consists of two columns:

- make - manufacturer
- no_sc_mmnl - the corresponding number of service centers in Metro Manila of the manufacturer in the column above

This external dataset was merged with the primary on the former (Make) column.

2. DATA PROCESSING

PART 1. DATA CLEANING

PART 2. DATA TRANSFORMATION

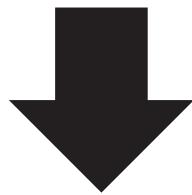
PART 3. MERGE

**PART 4. CATEGORICAL PROFILING AND
ENCODING VARIABLES**

PART 1. DATA CLEANING

```
make      0  
model     0  
make_year 0  
fuel_type 0  
transmission 5  
kms_driven 185  
engine_size 33  
body_type 0  
price     0  
key       0  
dtype: int64
```

We imported **primaryData.csv**, and proceeded to check the number of nulls per column in the primary dataset first.

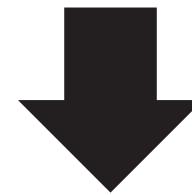


```
make      0  
model     0  
make_year 0  
fuel_type 0  
transmission 0  
kms_driven 0  
engine_size 0  
body_type 0  
price     0  
key       0  
dtype: int64
```

We then proceed to drop nulls in the columns `kms_driven`, `transmission`, and `engine_size` because no other data cleaning techniques taught in class besides dropping rows are reasonable to do. Now, that leaves us having zero nulls for the primary dataset.

```
make      0  
model     0  
make_year 0  
srp      19  
key       0  
dtype: int64
```

We imported **externalData.csv**, and proceeded to check the number of nulls per column in the external dataset.



```
make      0  
model     0  
make_year 0  
srp      0  
key       0  
dtype: int64
```

Dropping all the rows of `srp` was necessary since those nulls are the cars in AutoDeal that do not have a current price listing from the website. Now, we have no nulls anymore for the `externalDataset`.

PART 1. DATA CLEANING

Then we saved these two cleaned datasets and exported them as:
cleanedPrimaryData.csv, and
cleanedExternalData.csv.

PART 2. DATA TRANSFORMATION

```
make          object
model         object
make_year     int64
fuel_type    object
transmission  object
kms_driven   object
engine_size  float64
body_type    object
price        object
key          object
dtype: object
```

We imported **cleanedPrimaryData.csv**, and proceeded with the data transformation of the primary dataset (see the summary of the data types of the primary data).

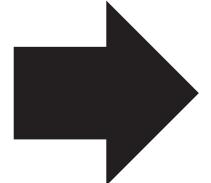
	make	model	make_year	fuel_type	transmission	kms_driven	engine_size	body_type	resale_price	key
0	Suzuki	S-Presso	2021	Gasoline	Manual	300000.0	1.0	Hatchback	415000.0	Suzuki S-Presso
1	Ford	Everest	2017	Diesel	Automatic	600000.0	2.2	SUV	930000.0	Ford Everest
2	Honda	Civic	2017	Gasoline	Automatic	500000.0	1.8	Sedan	760000.0	Honda Civic
3	Toyota	Innova	2023	Diesel	Automatic	170000.0	2.8	MPV	1120000.0	Toyota Innova
4	Toyota	Camry	2013	Gasoline	Automatic	600000.0	2.5	Sedan	520000.0	Toyota Camry
...
2216	Rolls-Royce	Ghost	2018	Gasoline	Automatic	200000.0	6.6	Sedan	18500000.0	Rolls-Royce Ghost
2217	Cadillac	Escalade	2022	Gasoline	Automatic	10.0	6.0	Suv	19000000.0	Cadillac Escalade
2218	Ferrari	F8	2024	Gasoline	Automatic	16000.0	3.9	Coupe	20500000.0	Ferrari F8
2219	Bentley	Bentayga	2023	Gasoline	Automatic	30000.0	4.0	Suv	21000000.0	Bentley Bentayga
2220	Lamborghini	Urus	2023	Gasoline	Automatic	1300.0	4.0	Suv	29500000.0	Lamborghini Urus

We then converted the column name price into resale_price then transformed it from an object to float, then kms_driven (truncated “km”), and engine_size (rounded it by two decimals) (kindly see Part 2. Data Transformation.ipynb for the full details).

PART 2. DATA TRANSFORMATION

After the procedure, we now have these newly converted data types.

```
make          object
model         object
make_year     int64
fuel_type     object
transmission  object
kms_driven   object
engine_size   float64
body_type    object
price         object
key           object
dtype: object
```



```
make          object
model         object
make_year     int64
fuel_type     object
transmission  object
kms_driven   float64
engine_size   float64
body_type    object
resale_price float64
key           object
dtype: object
```

PART 3. MERGE

In this portion of the data processing, we merged transformedCleanedPrimaryData.csv, and transformedCleanedExternalData.csv by inner join, retaining all the columns from the primary dataset, and only getting the columns: "ext_make_year", "srp", and "key".

```
# Merge

mergedDataset = pd.merge(primaryData, externalData[["ext_make_year", "srp", "key"]], on="key")
display(len(primaryData))
display(mergedDataset)
display(len(mergedDataset))
```

	key	make	model	fuel_type	transmission	body_type	make_year	ext_make_year	kms_driven	engine_size	resale_price	srp
0	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2021	2024	30000.0	1.0	415000.0	620000.0
1	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2021	2024	35000.0	1.0	410000.0	620000.0
2	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2021	2024	9000.0	1.0	448000.0	620000.0
3	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2022	2024	6000.0	1.0	528000.0	620000.0
4	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2023	2024	800.0	1.0	560000.0	620000.0
...
1636	Chevrolet Camaro	Chevrolet	Camaro	Gasoline	Manual	Coupe	2019	2024	900.0	6.2	5280000.0	4266888.0
1637	Hyundai Palisade	Hyundai	Palisade	Diesel	Automatic	Suv	2023	2024	15000.0	2.2	3250000.0	3780000.0
1638	BMW 4	BMW	4	Diesel	Automatic	Coupe	2020	2024	16000.0	2.0	3350000.0	4190000.0
1639	Nissan GT-R	Nissan	GT-R	Gasoline	Automatic	Coupe	2012	2024	16000.0	3.8	4980000.0	12445000.0
1640	Ford Bronco	Ford	Bronco	Gasoline	Automatic	Suv	2022	2024	100.0	2.0	6800000.0	4998000.0

PART 3. MERGE

Now, with the second external dataset to merge it with, make_mshops.csv was imported into this part's notebook and used inner join for two datasets: mergedDataset (the primary dataset, and the first external dataset), and make_mshops. We kept all the columns for the first merged dataset, and took the columns no_sc_mmnl (number of car service centers in Metro Manila), and make as the key for merging.

```
mergedDataset2 = pd.merge(mergedDataset,  
externalDataset2[["make", "no_sc_mmnl"]], on="make")  
  
display(mergedDataset2)  
display(len(mergedDataset))
```

	key	make	model	fuel_type	transmission	body_type	make_year	ext_make_year	kms_driven	engine_size	resale_price	srp	no_sc_mmnl
0	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2021	2024	30000.0	1.0	415000.0	620000.0	9
1	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2021	2024	35000.0	1.0	410000.0	620000.0	9
2	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2021	2024	9000.0	1.0	448000.0	620000.0	9
3	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2022	2024	6000.0	1.0	528000.0	620000.0	9
4	Suzuki S-Presso	Suzuki	S-Presso	Gasoline	Manual	Hatchback	2023	2024	800.0	1.0	560000.0	620000.0	9

1636	Jeep Grand	Jeep	Grand	Diesel	Automatic	Suv	2014	2024	55000.0	3.0	1800000.0	5490000.0	0
1637	Jeep Grand	Jeep	Grand	Gasoline	Automatic	Suv	2013	2024	51000.0	6.4	1950000.0	5490000.0	0
1638	Jeep Grand	Jeep	Grand	Gasoline	Automatic	Suv	2017	2024	30000.0	3.6	2180000.0	5490000.0	0
1639	Jeep Grand	Jeep	Grand	Gasoline	Automatic	Suv	2017	2024	38000.0	6.4	2500000.0	5490000.0	0
1640	Jeep Wrangler	Jeep	Wrangler	Gasoline	Automatic	Suv	2008	2024	88000.0	3.8	1998000.0	5090000.0	0

1641 rows × 13 columns

PART 4. CATEGORICAL PROFILING AND ENCODING CATEGORICAL VARIABLES

```
[ ] # Categorical Profiling

data = pd.read_csv("mergedDataset.csv")

make_data = data['make'].value_counts()
model_data = data['model'].value_counts()
transmission_data = data['transmission'].value_counts()
body_type_data = data['body_type'].value_counts()

make_data.to_csv('make_cp.csv')
model_data.to_csv('model_cp.csv')
transmission_data.to_csv('transmission_cp.csv')
body_type_data.to_csv('body_type_cp.csv')
```

For Categorical Profiling, we got the frequency count of each values in the following variables:

- make
- model
- transmission
- body_type

We, then, saved the frequency data with respect to each of the variables

PART 4. CATEGORICAL PROFILING AND ENCODING CATEGORICAL VARIABLES

```
# Data Cleaning

df = pd.read_csv("mergedDataset_updated.csv")

def cvt(x):
    if x == "Cvt":
        return "CVT"
    else:
        return x

def body(x):
    if x == "Minivans" or x == "Mpv" or x == "Minivan":
        return "MPV"
    elif x == "Pickup-trucks" or x == "Pickup":
        return "Pickup Truck"
    elif x == "Suv":
        return "SUV"
    elif x == "Luxury-vehicles":
        return "Luxury Vehicle"
    elif x == "Station-wagon":
        return "Station Wagon"
    elif x == "Convertibles":
        return "Convertible"
    else:
        return x

un = df['body_type'].unique()
df['transmission'] = df['transmission'].apply(cvt)
df['body_type'] = df['body_type'].apply(body)
```

Before encoding the Categorical Variables, we had to do some data cleaning first in order to get more accurate data encoding.

According to the initial dataset, “Cvt” and “CVT” are two different transmission type (which are not). So we applied the function *cvt()* on the *transmission* column to ensure that all the values are unique.

Moreover, in the *body_type* column, we observed that “Pickup-trucks” and “Pickup” are treated as different values so we got to fix that. And upon further research, we found out that “Minivan” and “Mpv” are the same. Thus, along with the “Minivans” value, we re-wrote it under “MPV”.

Lastly, we refactored the other unique values to make it more “readable”.

PART 4. CATEGORICAL PROFILING AND ENCODING CATEGORICAL VARIABLES

Now, for the `fuel_type` variable, we used one-hot encoding since there are only 3 unique values under it. Meanwhile, for the variables `make`, `model`, `transmission`, and `body_type`, we used both frequency encoding and target encoding. For the target encoding, we imported the `category_encoders` from `sklearn` and made use of `TargetEncoder` class. This makes the code fairly straight-forward and easy rather than encoding it from scratch.

	key	make	model	transmission	body_type	make_year	ext_make_year	make_target	model_target	transmission_target	body_type_target	fuel_type_Diesel	fuel_type_Gasoline	kms_driven	engine_size	resale_price	srp	no_sc_mmnl
0	Suzuki S-Presso	Suzuki	S-Presso	Manual	Hatchback	2021	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0	1	30000.0	1.0	415000.0	620000.0	9
1	Suzuki S-Presso	Suzuki	S-Presso	Manual	Hatchback	2021	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0	1	35000.0	1.0	410000.0	620000.0	9
2	Suzuki S-Presso	Suzuki	S-Presso	Manual	Hatchback	2021	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0	1	9000.0	1.0	448000.0	620000.0	9
3	Suzuki S-Presso	Suzuki	S-Presso	Manual	Hatchback	2022	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0	1	6000.0	1.0	528000.0	620000.0	9
4	Suzuki S-Presso	Suzuki	S-Presso	Manual	Hatchback	2023	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0	1	800.0	1.0	560000.0	620000.0	9
...	
1636	Jeep Grand	Jeep	Grand	Automatic	SUV	2014	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	1	0	55000.0	3.0	1800000.0	5490000.0	0
1637	Jeep Grand	Jeep	Grand	Automatic	SUV	2013	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	0	1	51000.0	6.4	1950000.0	5490000.0	0
1638	Jeep Grand	Jeep	Grand	Automatic	SUV	2017	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	0	1	30000.0	3.6	2180000.0	5490000.0	0
1639	Jeep Grand	Jeep	Grand	Automatic	SUV	2017	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	0	1	38000.0	6.4	2500000.0	5490000.0	0
1640	Jeep Wrangler	Jeep	Wrangler	Automatic	SUV	2008	2024	1.159390e+06	1.092448e+06	1.102811e+06	1.086440e+06	0	1	88000.0	3.8	1998000.0	5090000.0	0

1641 rows × 18 columns

PART 4. CATEGORICAL PROFILING AND ENCODING CATEGORICAL VARIABLES

Now, for the frequency encoding, we just took the frequency count of each unique values and normalize it within its respective column/variable. We, then, mapped the values in their respective categorical values.

	key	make_year	ext_make_year	make_target	model_target	transmission_target	body_type_target	make_freq	model_freq	transmission_freq	body_type_freq	fuel_type_Diesel	fuel_type_Gasoline	kms_driven	engine_size	resale_price	srp	no_sc_mmn
0	Suzuki S-Presso	2021	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0.039610	0.003047	0.232785	0.060329	0	1	30000.0	1.0	415000.0	620000.0	9
1	Suzuki S-Presso	2021	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0.039610	0.003047	0.232785	0.060329	0	1	35000.0	1.0	410000.0	620000.0	9
2	Suzuki S-Presso	2021	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0.039610	0.003047	0.232785	0.060329	0	1	9000.0	1.0	448000.0	620000.0	9
3	Suzuki S-Presso	2022	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0.039610	0.003047	0.232785	0.060329	0	1	6000.0	1.0	528000.0	620000.0	9
4	Suzuki S-Presso	2023	2024	5.488572e+05	8.685647e+05	7.532528e+05	5.400233e+05	0.039610	0.003047	0.232785	0.060329	0	1	800.0	1.0	560000.0	620000.0	9
...	
1636	Jeep Grand	2014	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	0.004266	0.003656	0.606338	0.371115	1	0	55000.0	3.0	1800000.0	5490000.0	0
1637	Jeep Grand	2013	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	0.004266	0.003656	0.606338	0.371115	0	1	51000.0	6.4	1950000.0	5490000.0	0
1638	Jeep Grand	2017	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	0.004266	0.003656	0.606338	0.371115	0	1	30000.0	3.6	2180000.0	5490000.0	0
1639	Jeep Grand	2017	2024	1.159390e+06	1.140776e+06	1.102811e+06	1.086440e+06	0.004266	0.003656	0.606338	0.371115	0	1	38000.0	6.4	2500000.0	5490000.0	0
1640	Jeep Wrangler	2008	2024	1.159390e+06	1.092448e+06	1.102811e+06	1.086440e+06	0.004266	0.000609	0.606338	0.371115	0	1	88000.0	3.8	1998000.0	5090000.0	0

1641 rows × 18 columns

PART 4. CATEGORICAL PROFILING AND ENCODING CATEGORICAL VARIABLES

```
make_year          int64
ext_make_year      int64
make_target        float64
make_freq          float64
model_target       float64
model_freq         float64
transmission_target float64
transmission_freq  float64
body_type_target   float64
body_type_target   float64
fuel_type_Diesel   int64
fuel_type_Gasoline int64
kms_driven         float64
engine_size        float64
resale_price       float64
srp                float64
no_sc_mmn1         int64
dtype: object
```

After dropping the key column (which are the brand and model of the cars), we now have a purely numerical dataset which can be used for modelling and data analysis.

OUR FINAL DATASET

	make_year	ext_make_year	make_target	make_freq	model_target	model_freq	transmission_target	transmission_freq	body_type_target	body_type_freq	fuel_type_Diesel	fuel_type_Gasoline	kms_driven	engine_size	resale_price	srp	no_sc_mmn1
0	2021	2024	5.488572e+05	0.039610	8.685647e+05	0.003047	7.532528e+05	0.232785	5.400233e+05	0.060329	0	1	30000.0	1.0	415000.0	620000.0	9
1	2021	2024	5.488572e+05	0.039610	8.685647e+05	0.003047	7.532528e+05	0.232785	5.400233e+05	0.060329	0	1	35000.0	1.0	410000.0	620000.0	9
2	2021	2024	5.488572e+05	0.039610	8.685647e+05	0.003047	7.532528e+05	0.232785	5.400233e+05	0.060329	0	1	9000.0	1.0	448000.0	620000.0	9
3	2022	2024	5.488572e+05	0.039610	8.685647e+05	0.003047	7.532528e+05	0.232785	5.400233e+05	0.060329	0	1	6000.0	1.0	528000.0	620000.0	9
4	2023	2024	5.488572e+05	0.039610	8.685647e+05	0.003047	7.532528e+05	0.232785	5.400233e+05	0.060329	0	1	800.0	1.0	560000.0	620000.0	9
...
1636	2014	2024	1.159390e+06	0.004266	1.140776e+06	0.003656	1.102811e+06	0.606338	1.086440e+06	0.371115	1	0	55000.0	3.0	1800000.0	5490000.0	0
1637	2013	2024	1.159390e+06	0.004266	1.140776e+06	0.003656	1.102811e+06	0.606338	1.086440e+06	0.371115	0	1	51000.0	6.4	1950000.0	5490000.0	0
1638	2017	2024	1.159390e+06	0.004266	1.140776e+06	0.003656	1.102811e+06	0.606338	1.086440e+06	0.371115	0	1	30000.0	3.6	2180000.0	5490000.0	0
1639	2017	2024	1.159390e+06	0.004266	1.140776e+06	0.003656	1.102811e+06	0.606338	1.086440e+06	0.371115	0	1	38000.0	6.4	2500000.0	5490000.0	0
1640	2008	2024	1.159390e+06	0.004266	1.092448e+06	0.000609	1.102811e+06	0.606338	1.086440e+06	0.371115	0	1	88000.0	3.8	1998000.0	5090000.0	0

1641 rows × 17 columns

OUR FINAL DATASET

make_year
ext_make_year
make_target
make_freq
model_target
model_freq

- the year in which a particular car is made (Independent Variable)
- the year in which (brand new) cars from the external dataset are made (Independent Variable)
- target encoded values of brands that made a given car (Independent Variable)
- frequency encoded values of the brands that made a given car (Independent Variable)
- target encoded values of the model of each car (Independent Variable)
- frequency encoded values of the model of each car (Independent Variable)

OUR FINAL DATASET

transmission_target
transmission_freq
body_type_target
body_type_freq
fuel_type_Diesel
fuel_type_Gasoline

- target encoded values of a certain car's transmission (Independent Variable)
- frequency encoded values of a certain car's transmission (Independent Variable)
- target encoded values of a car's body type (Independent Variable)
- frequency encoded values of a car's body type (Independent Variable)
- one-hot encoded value that indicates whether a car's fuel type is Diesel or not (Independent Variable)
- one-hot encoded value that indicates whether a car's fuel type is Gasoline or not (Independent Variable)

OUR FINAL DATASET

kms_driven
engine_size
resale_price
srp
no_sc_mmnl

- the total distance (in km) that a car has traveled prior its sale (**Independent Variable**)
- the engine size of a particular car (**Independent Variable**)
- a car's resale price in the market (**Dependent Variable**)
- a car's suggested retail price in the market (**Independent Variable**)
- number of service centers in metro manila, that is, per manufacturer/make (**Independent Variable**)

VI. EXPLORATORY DATA ANALYSIS

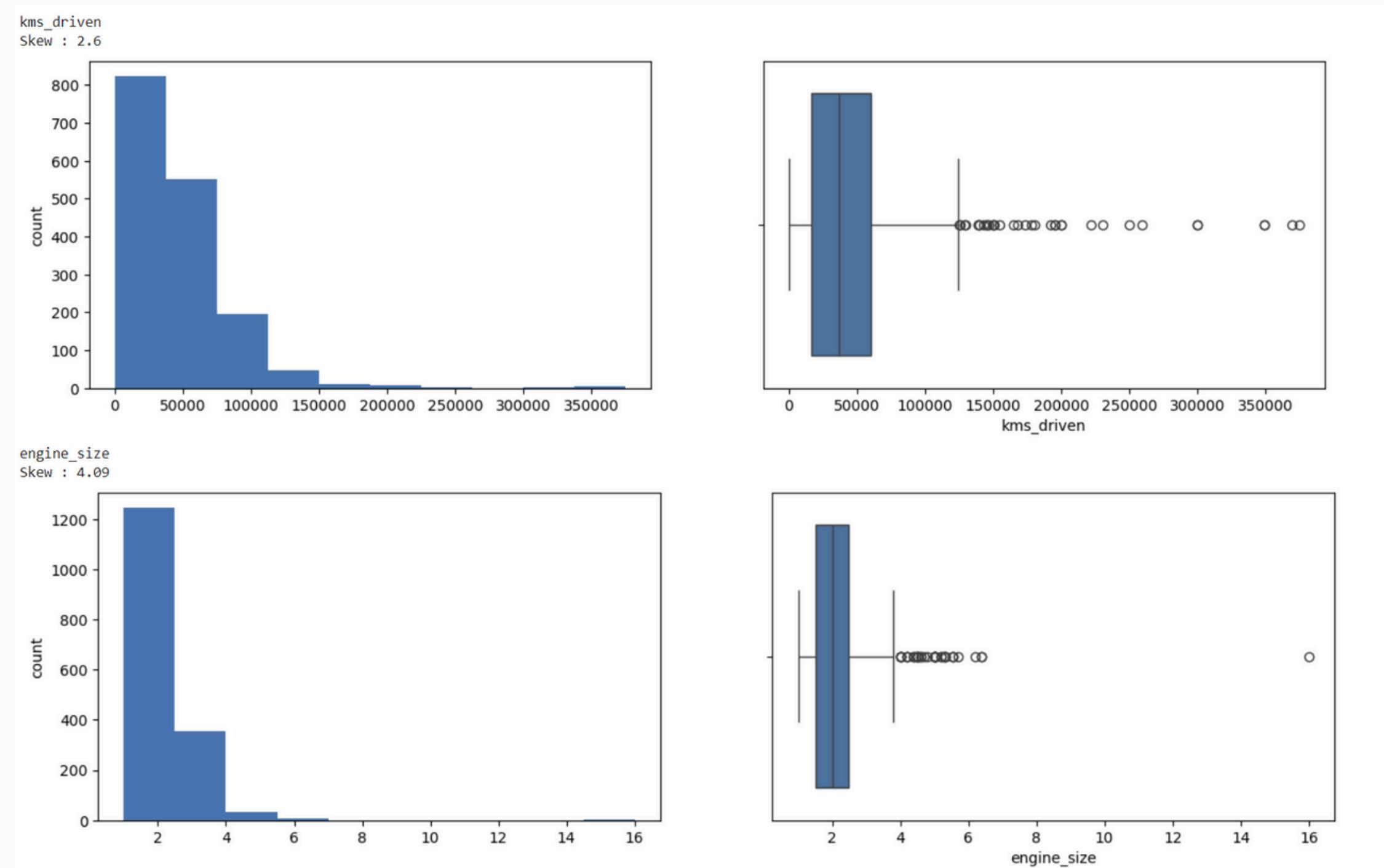
PART 1. STATISTICS SUMMARY

	count	mean	std	min	25%	50%	75%	max
make_year	1,641.00	2,017.03	4.07	1,993.00	2,016.00	2,018.00	2,020.00	2,024.00
ext_make_year	1,641.00	2,024.00	0.00	2,024.00	2,024.00	2,024.00	2,024.00	2,024.00
make_target	1,641.00	934,818.89	216,081.22	548,857.20	762,397.43	959,523.65	959,523.65	1,731,480.99
make_freq	1,641.00	0.17	0.14	0.00	0.05	0.10	0.35	0.35
model_target	1,641.00	893,155.50	245,507.34	463,102.42	758,945.67	874,949.60	1,082,464.66	2,014,394.24
model_freq	1,641.00	0.03	0.02	0.00	0.01	0.02	0.04	0.07
transmission_target	1,641.00	956,088.24	185,803.15	678,251.55	753,252.82	1,102,810.83	1,102,810.83	1,102,810.83
transmission_freq	1,641.00	0.45	0.20	0.00	0.23	0.61	0.61	0.61
body_type_target	1,641.00	940,732.63	221,466.48	540,023.28	704,313.63	1,086,439.88	1,086,439.88	1,982,512.02
body_type_freq	1,641.00	0.22	0.13	0.00	0.07	0.21	0.37	0.37
fuel_type_Diesel	1,641.00	0.44	0.50	0.00	0.00	0.00	1.00	1.00
fuel_type_Gasoline	1,641.00	0.55	0.50	0.00	0.00	1.00	1.00	1.00
kms_driven	1,641.00	43,297.57	38,790.32	1.00	16,563.00	37,000.00	60,000.00	374,683.00
engine_size	1,641.00	2.08	0.82	1.00	1.50	2.00	2.49	16.00
resale_price	1,641.00	957,005.63	797,191.79	60,000.00	588,000.00	788,000.00	1,059,000.00	11,500,000.00
srp	1,641.00	1,608,335.87	1,204,063.05	570,000.00	954,000.00	1,261,000.00	1,769,000.00	14,790,000.00
no_sc_mmnl	1,641.00	5.64	2.94	0.00	4.00	4.00	9.00	9.00

There are 1641 cars. Based on the statistics summary, years range from 1993–2024 for both used and brand new cars. 25% of the cars were created from 2020 onwards. 55% of cars are running on gasoline while 44% are running on diesel, and the remaining 1% are hybrid. Cars can be driven on an average of around 43000 km. Car prices range from 60K to 11.5 M. 25% of the cars can be brought to 9 service centers in Metro Manila, which is the maximum value of the said variable.

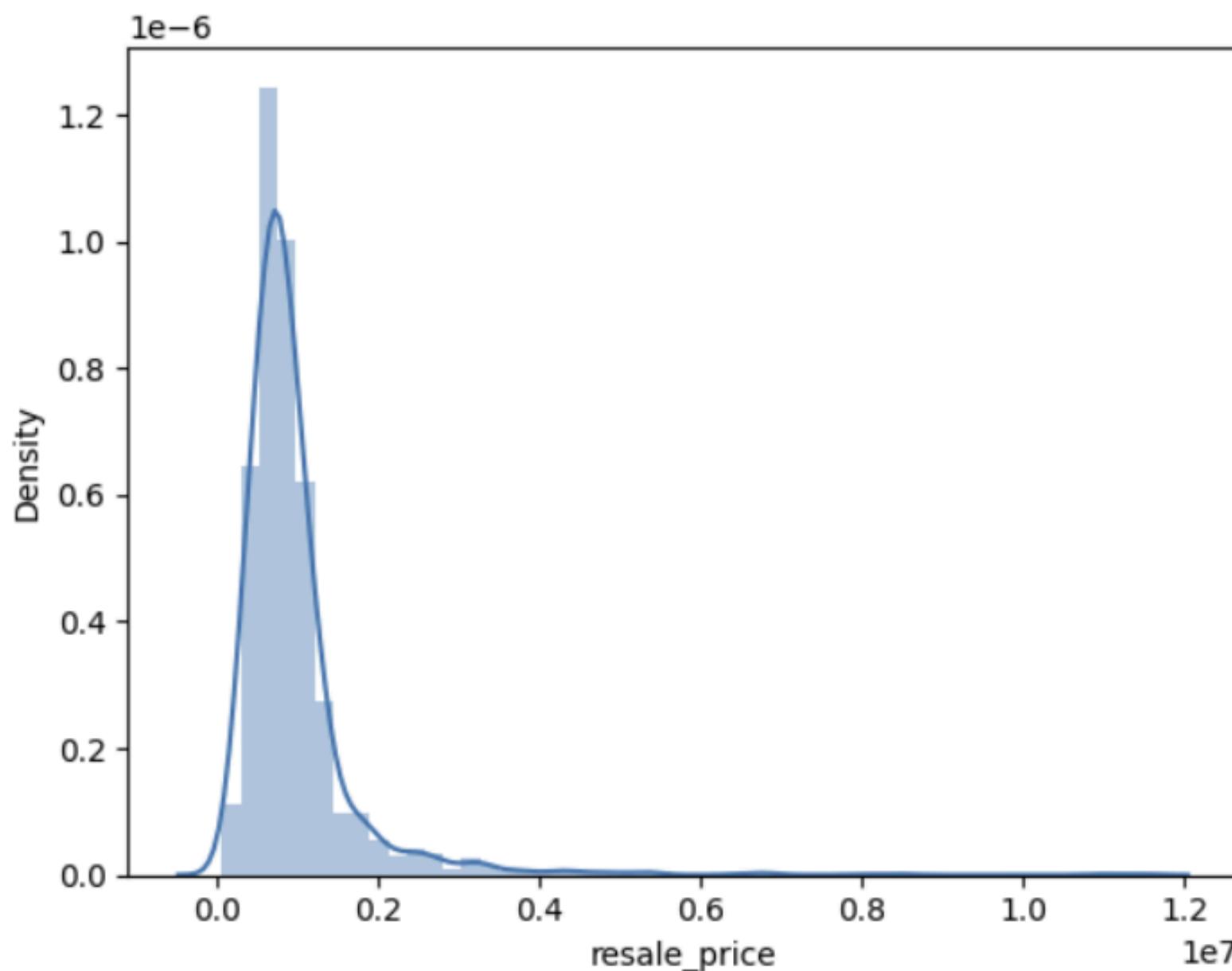
PART 2. UNIVARIATE ANALYSIS

Numerical variables are visualized into histograms and boxplots, and some of the variables are shown to have high skewness. There is no further visualization since there are no categorical variables in the dataset.



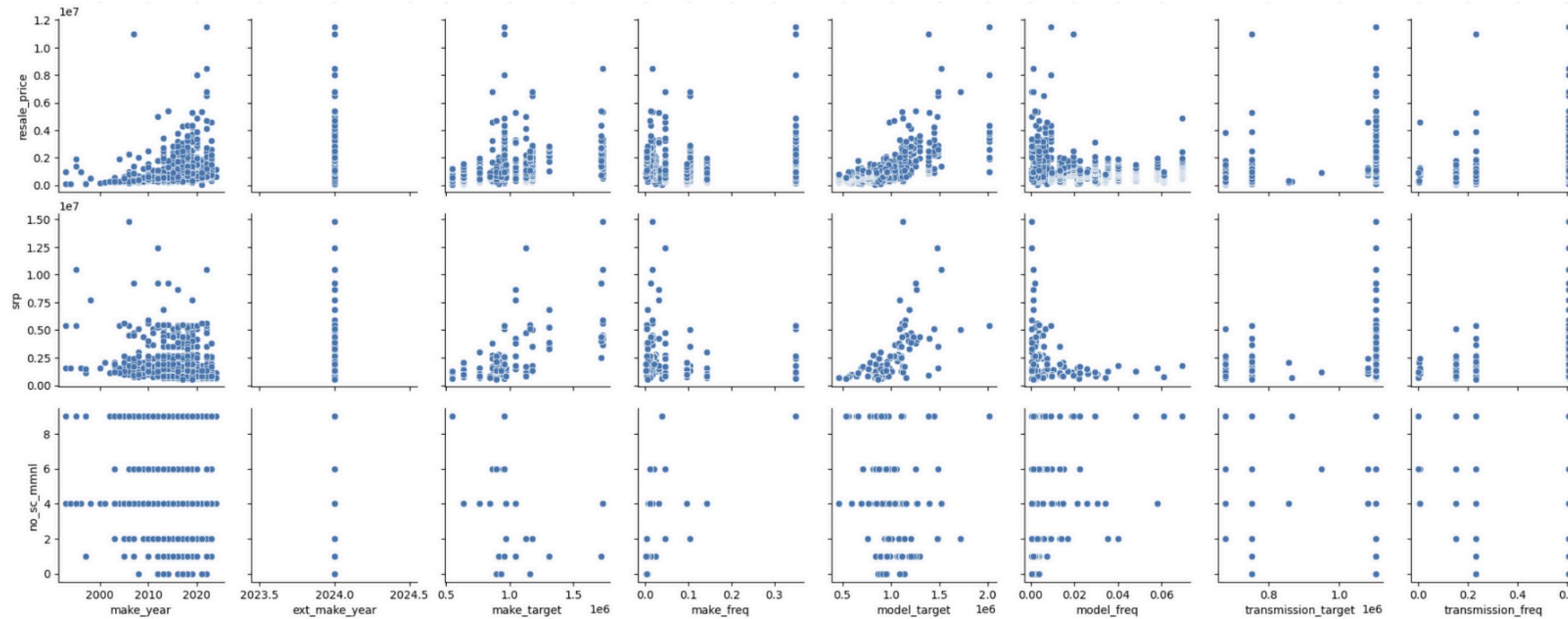
PART 2. UNIVARIATE ANALYSIS

Since the resale price is highly skewed, **log transformation** could be used in normalizing the variables.



PART 3. BIVARIATE ANALYSIS

A **pair plot** was used to visualize the relationship between two variables from the dataset.



Note: The visualization above is NOT the full pair plot due to the limited space of a slide. Kindly see the full pair plot in the EDA python notebook in the Google drive.

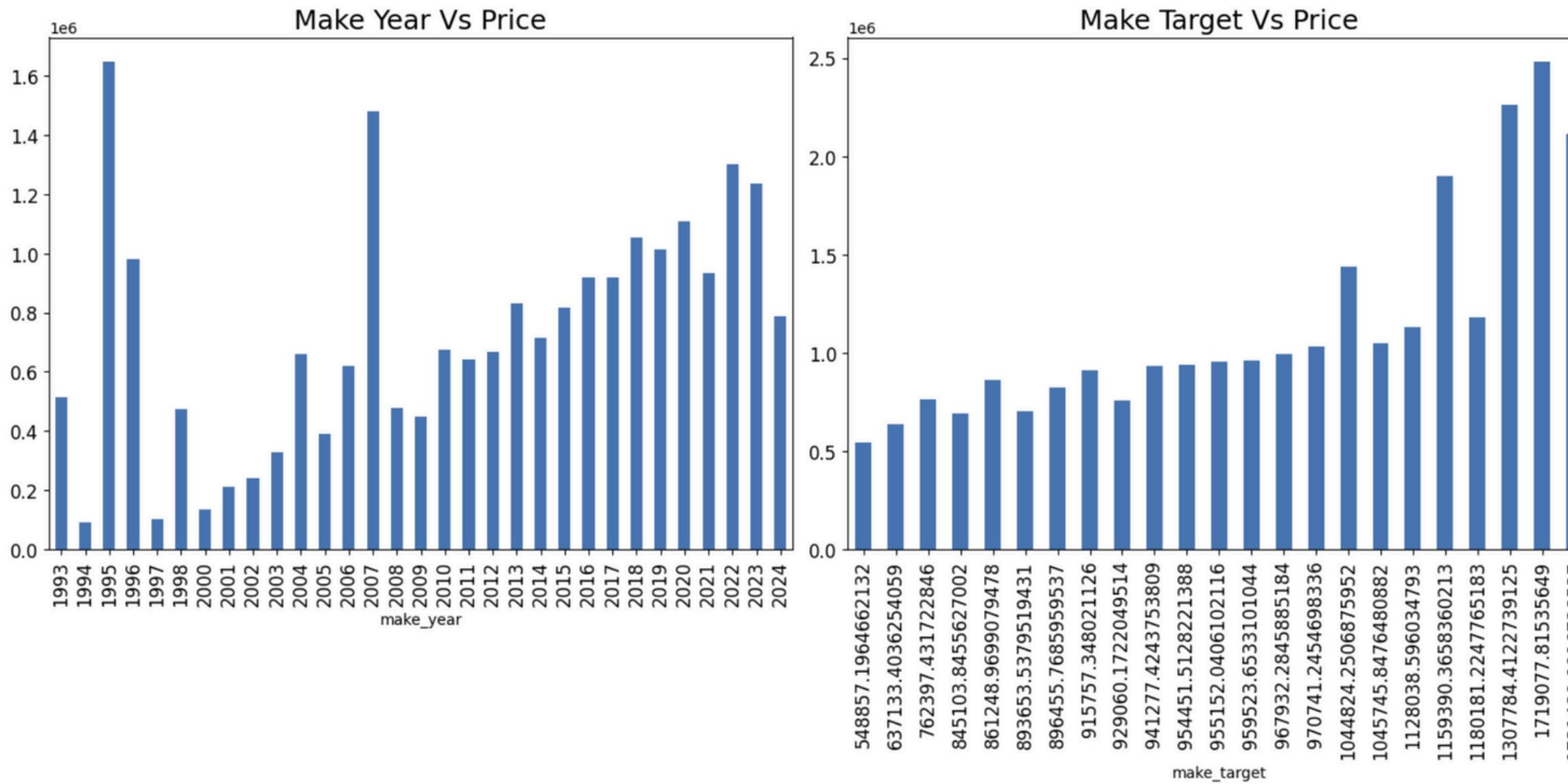
PART 3. BIVARIATE ANALYSIS

Some observations from the pair plot:

- The resale price has a positive correlation of make year, model target, engine size, and SRP. That shows that a car's price is higher if a specific car is made recently.
- Kilometers-driven has a negative correlation with price.
- SRP is positively correlated with model target and engine size.
- As make frequency increases, the model frequency increases.

PART 3. BIVARIATE ANALYSIS

Bar plots were also used to visualize the relationship between two numerical variables and the resale price since it is a continuous variable. Numerical variables include make year, make (both target and frequency), model frequency, transmission (both target and frequency), body type, and no. of service centers in Metro Manila.



Note: The visualization above is two of the eight bar plots made for EDA.

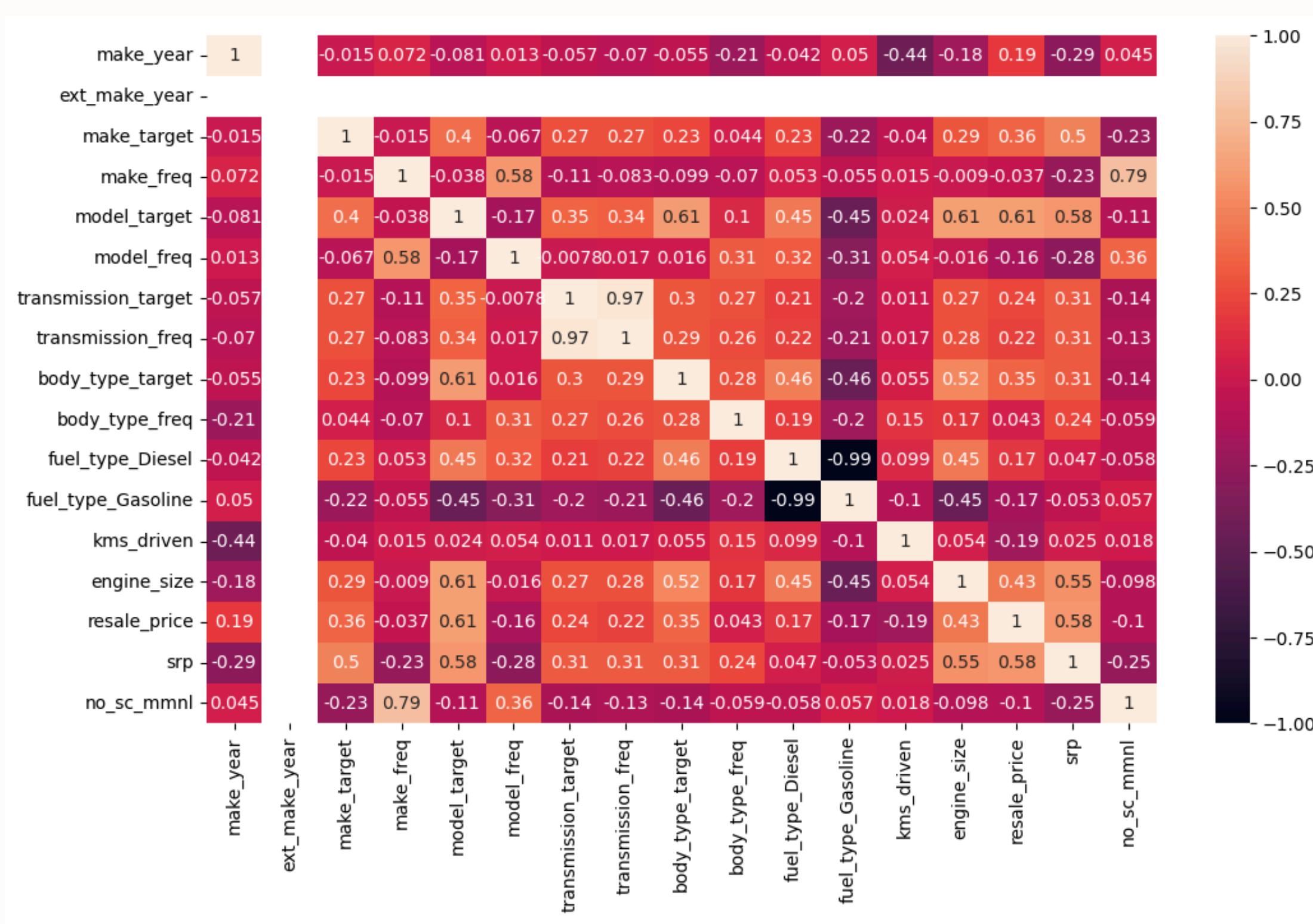
PART 3. BIVARIATE ANALYSIS

Some observations from the bar plots:

- Manufactured cars in 1995 have the highest resale price even though cars made in 1994 and 1997 are sold in the lowest price.
- As the make target increases, the price increases.
- Cars with a transmission from around 800k-900k are low in price.
- A car with a high body type is high in price.
- Cars that can be brought to 0-2 service centers in Metro Manila are usually high in price.

PART 4. MULTIVARIATE ANALYSIS

A **heat map** was used to show the correlation between two variables from the dataset.



PART 4. MULTIVARIATE ANALYSIS



Make year and kilometers-Driven are very weakly correlated to most of the variables from the dataset. Resale price has a positive correlation to diesel cars and a negative correlation to gasoline cars. Model frequency does not create much impact on body type, engine size, and transmission (both target and frequency encoding). However, model target has a fairly strong positive correlation to body type, engine size, and the price. One fuel type has a very strong negative correlation to one other fuel type (-0.99). Make frequency has a strong positive correlation to the number of service centers in Metro Manila. SRP is fairly strongly positive in correlation to make target, model target, engine size, and price.

REFERENCES

- AutoDeal. (n.d.). AutoDeal. <https://www.autodeal.com.ph/>
- ZigWheels. (n.d.). ZigWheels. <https://www.zigwheels.ph/>
- Collard, M. (2022). Price Prediction for Used Cars: A Comparison of Machine Learning Regression Models. *Mittuniversitetet*. Retrieved from: <https://www.diva-portal.org/smash/get/diva2:1674070/FULLTEXT01.pdf>
- Totakura, S. S. G., Kosuru, H. (2021). Comparison of Supervised Learning Models for predicting prices of Used Cars. *Blekinge Institute of Technology*. Retrieved from: https://www.diva-portal.org/smash/get/diva2:1609361/FULLTEXT02?fbclid=IwZXhObgNhZWOCMTEAAR3yUWYBnT5AnLIPvY_zewsNCA17EvXv_zdrV41S9eilkBA82FJNDuwphkY_aem_FHnXVVSOhM7AcY6tBvV_wQ