SECOND HAND CARS IN THE PH: A Buyer's Guide

FTW Data Science Batch 2 - Team Money | Go | Lumagui | Misa | Santos |
10 August 2019

SECOND HAND CARS IN THE PH:
A Buyer's Guide

Problem

Context

Solution

Our Model

Why do we need this?
How does it work?
How does this fare with the other calculators out there?

Final Caveats and Recommendations

OUR TEAM

Elyse Go
elysekatrina.go
@ftwfoundation.org

Nicole Lumagui
nicole.lumagui
@ftwfoundation.org

Berna Misa
bernadette
@ftwfoundation.org

Jero Santos
jero
@ftwfoundation.org

WHY DO WE NEED THIS?


60
MAXIMUM
SPEED LIMIT

TRAIN LAW

Excise on Automobiles

2% for up to **Php600,000**

Php12,000 + 20% of excess over Php600,000

Php112,000 + 40% of excess over Php1.1 million

Php512,000 + 60% of excess over Php2.1 million

4% for up to **Php600,000**

10% for over **Php600,000 to Php1 million**

20% for over **Php1 million up to Php4 million**

50% of excess over **Php4 million**

AUTO TAX REFORM

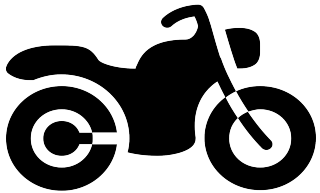
Public
Transportation

Public Transportation

Carpooling

Public Transportation

Carpooling



TNVS

CONVENIENCE
&
RELIABILITY

Market activity for second hand cars

Why buy used cars?

More savings

Cheaper insurance cost

Slower depreciation

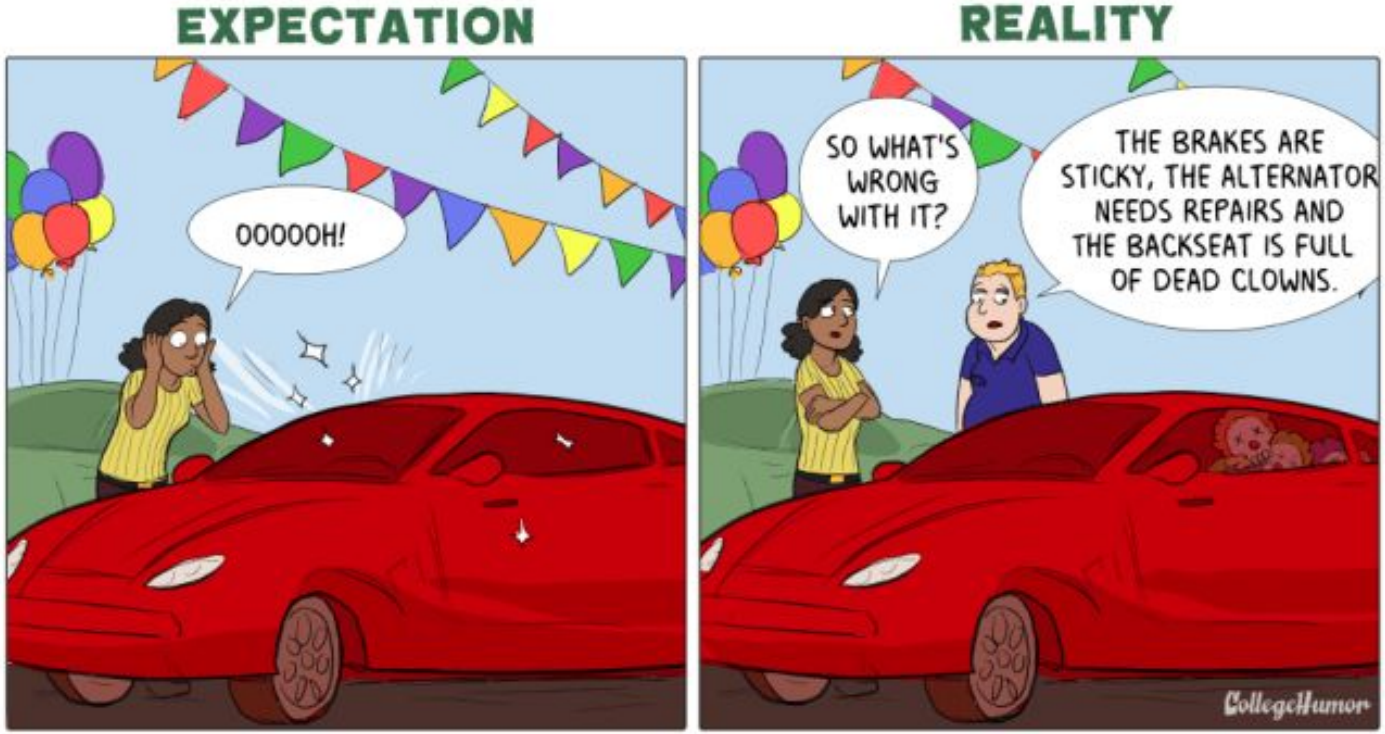
Extended warranty

Good for the environment

Buying used cars usually involves two steps:

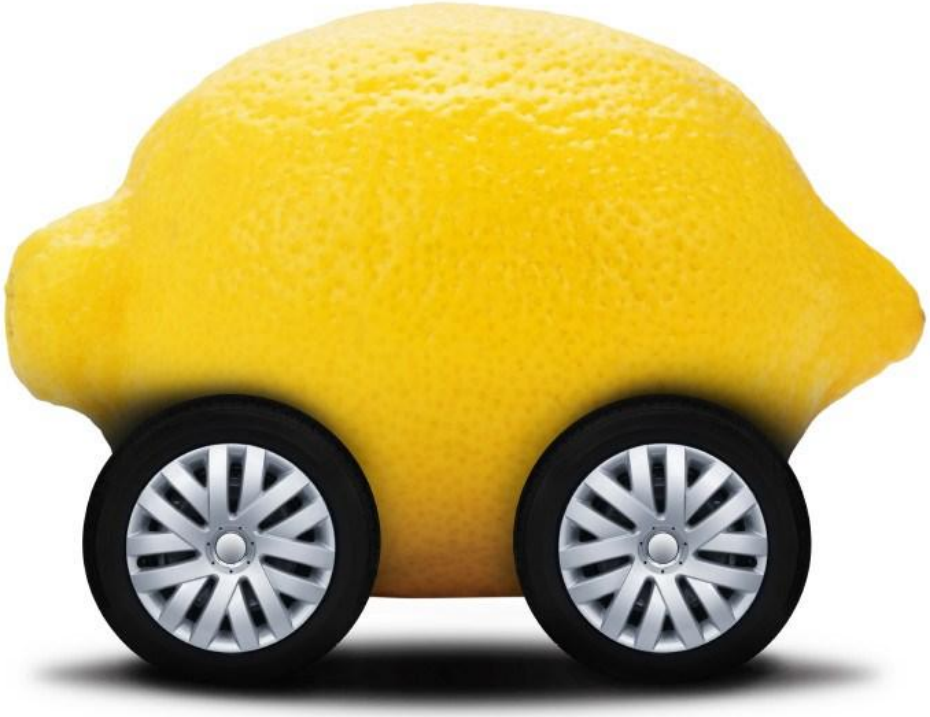
1 Consumer chooses preferred car-type
(Refers to used car buy/sell websites)

2 Consumer tries to find a special offer for sale (Looks for the best "price-quality-level")



What is a lemon car?

Lemon Car
=
Overpriced car

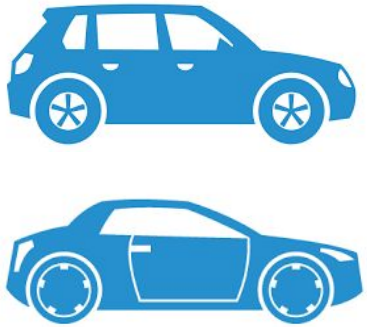


Sellers know more than the buyer

Risks of buying used cars

Unknown reliability or treatment

More frequent maintenance



Hard to find an exact match of what you want

Untouched warranty

How can we help ***consumers*** *in their journey of* ***buying a second hand car****?*

*By **empowering** them with **information***
*derived from **Machine Learning***

SOLUTION

1. **Predict prices of used cars** in the Philippines based on data from top local buy-and-sell websites for cars

2. Propose a comprehensive and scalable **“fair value” pricing model** for used cars sold online

HOW DOES THIS FARE WITH THE OTHER CALCULATORS OUT THERE?

How do consumers currently compute the price of used cars?

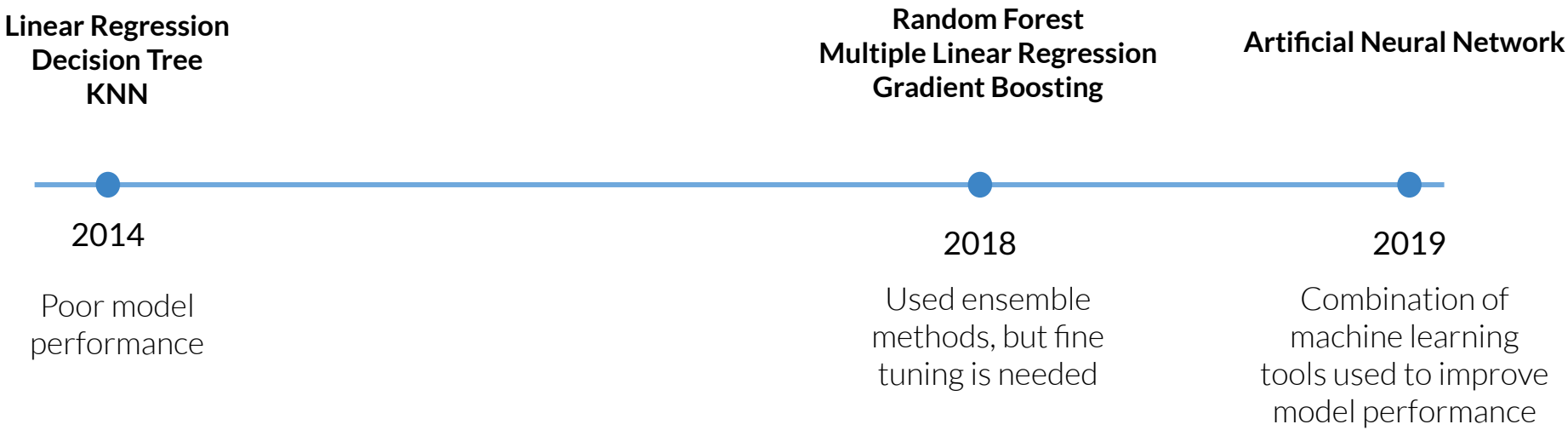
Custom Car depreciation calculator

Straight-line depreciation

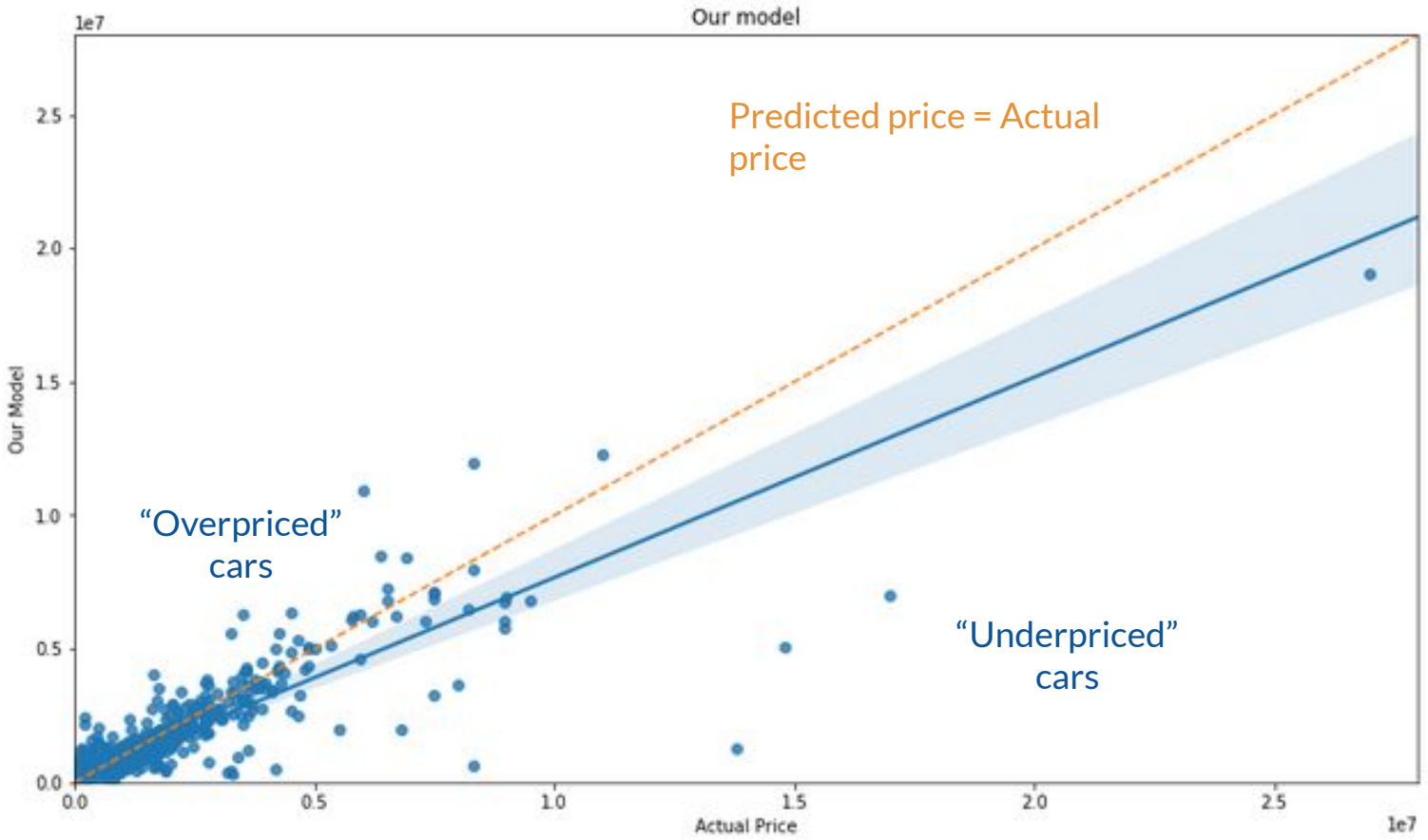
Third party appraisal

Checking price of second hand or repossessed vehicles (banks, online marketplace, local car dealers)

Employing machine learning (ML) tools is becoming a trend for predicting used car prices



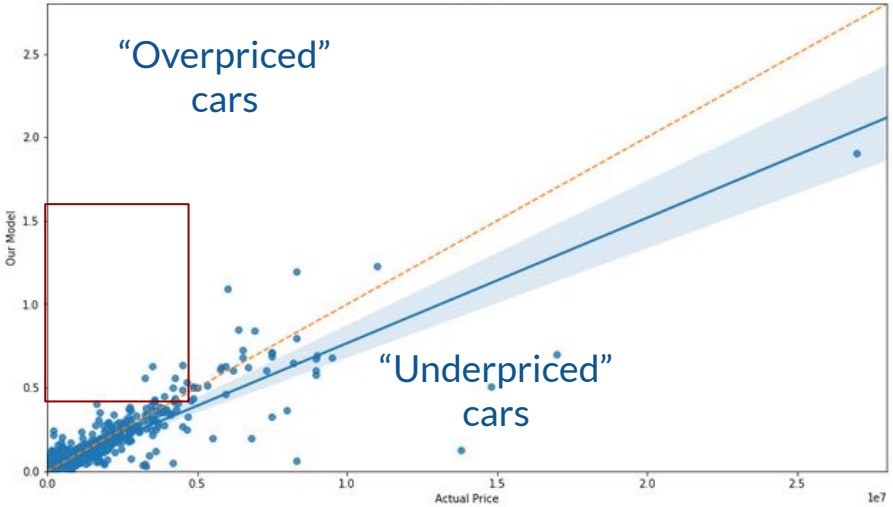
How about our model?



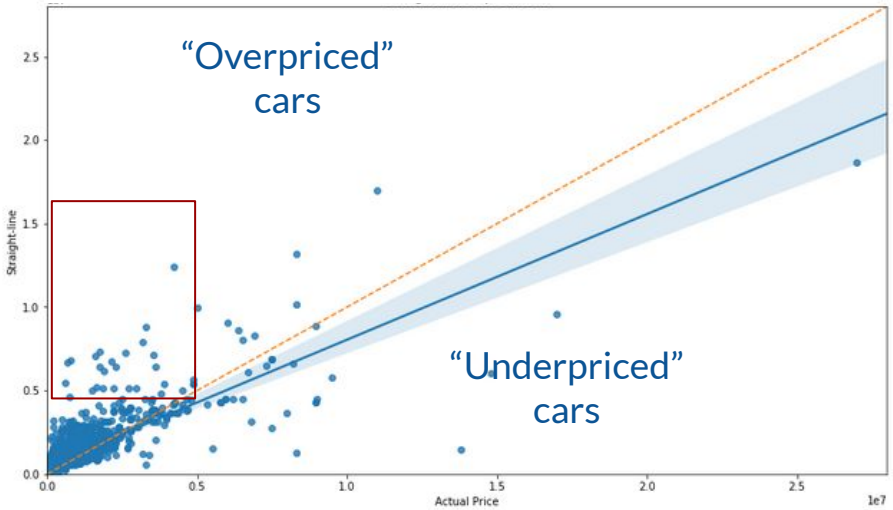
OUR MODEL

How about our model?

Our model



Straight-line depreciation



Overpricing of cars has been **reduced** with our model.

HOW DOES IT WORK?

What it can do	What it can't do
Predict price of used cars on a **local context**	Does not account for the markup of the seller when posting in a used car marketplace (**haggling room for sellers**)
Accounts for **12 features or factors** in predicting the price	Does not account for **issues of that is not captured** in the data points we considered
Conservative estimations in the used car price	Data only represents information scraped on a **limited time frame**

Data was scraped from the used car listings online and their brand new counterparts

Used car listings

PHILKOTSE.COM

carmudi PHILIPPINES

Retail prices of brand new cars

Priceprice Auto

AutoSearchManila

Final dataset

Brand	Age of Car
Body type	Location
Model	Color
Price	Transmission type
Mileage (in km)	Fuel type
Retail price (new)	Age of the post (in days)
Seller type	

Data was scraped from the used car listings online and their brand new counterparts

Philkotse

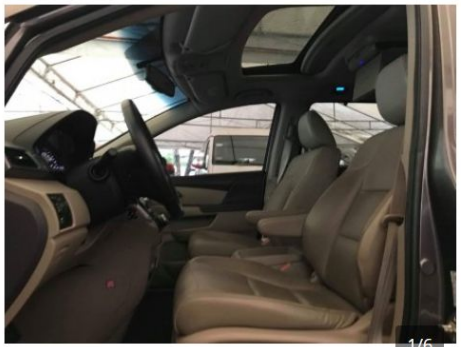
Sell 2nd Hand 2013 Honda Odyssey at 59000 km in Makati

Makati | Aug 02, 2019

Product: Honda Odyssey 2013

Tweet
Share

Price drop alert



Compare Save listing Report ad

₱ 1,098,000

Make an offer

Brand	*Honda*
Model	*Odyssey*
Version	*Full option*
Year of manufacture	*2013*
Condition	*Used*
Transmission	*Automatic*
Mileage	*59,000 km*
Color	*Grey*
Body type	*Wagon (Estate)*

Carmudi

2013 Audi A6 3.0 AT

₱ 1,670,000 Negotiable

₱ 48,879 monthly (Sample computation only) Apply for a loan

Mileage	Transmission	Fuel type	Engine size
49,000 Km	Automatic	Diesel	3.0 L

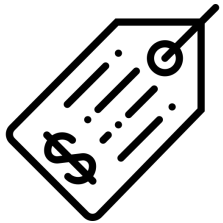
Primary Details

Color Family	Black	**Drive Type**	Rear wheel drive
Doors	4		

2nd Hand Car Price

CURRENT ONLINE CALCULATORS

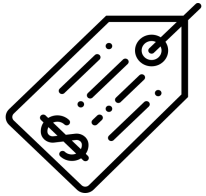
Age



Retail Price

WHAT WE ACCOUNTED FOR

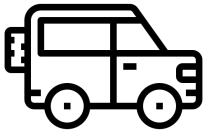
Age



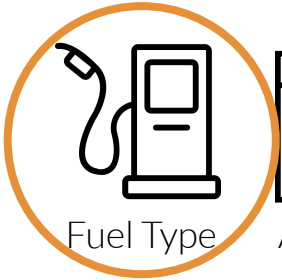
Retail Price

Mileage

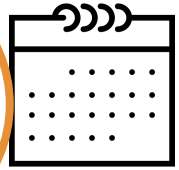
Brand/Make



Body Type



Fuel Type



Age of post

Individual / Dealer

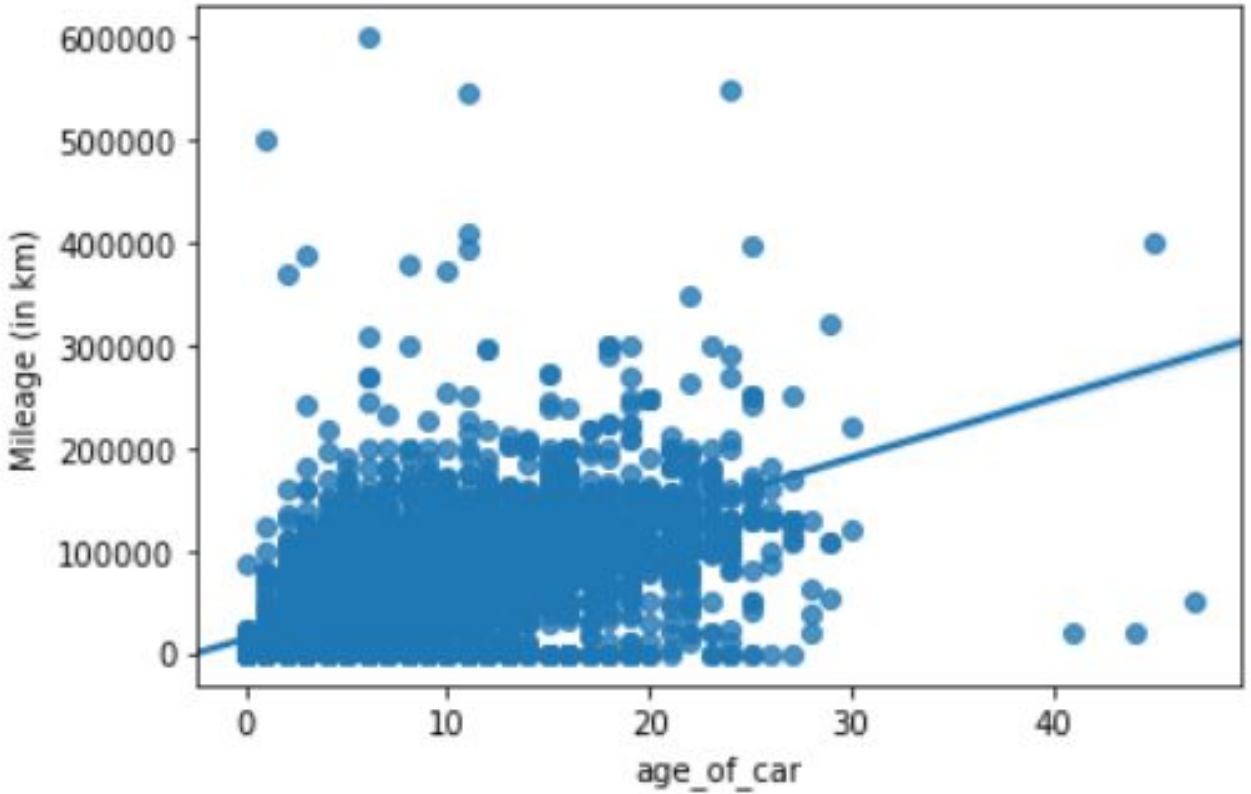
Color Family

Location

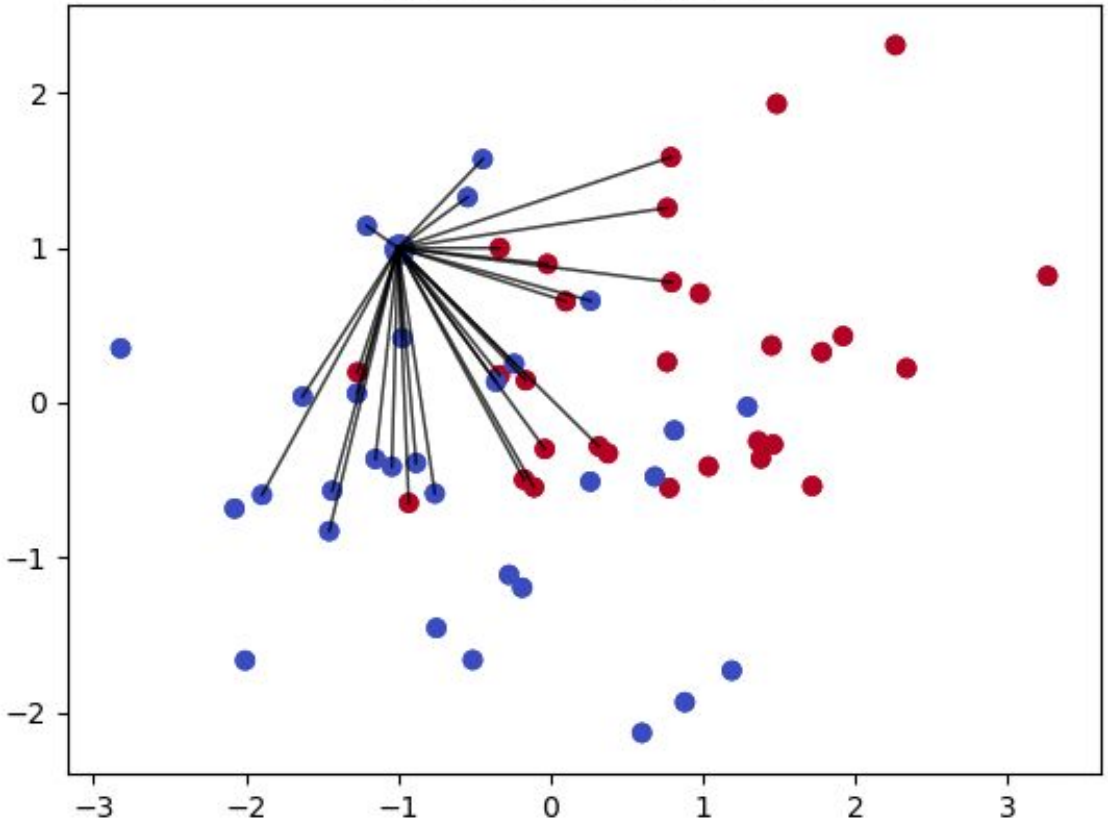
Transmission Type

Model

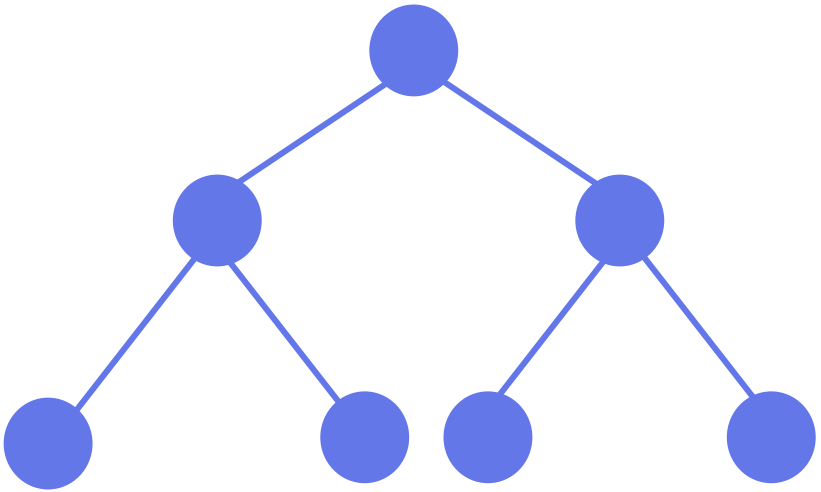
Exploratory Data Analysis show that the age of car and mileage have…



K-Nearest Neighbors was used to impute mileage values based on Age of Car



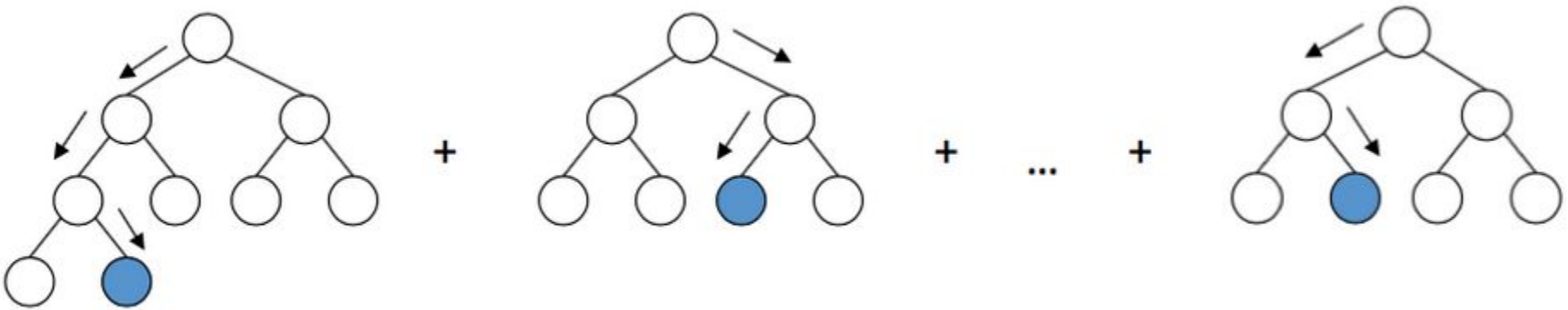
Model #1: Decision Tree was used to determine the key component affecting car price



Feature Importance:

- **Retail**
- **Age of the Car**
- **Mileage**
- **Model**
- **Brand**

Model #2: XGBoost

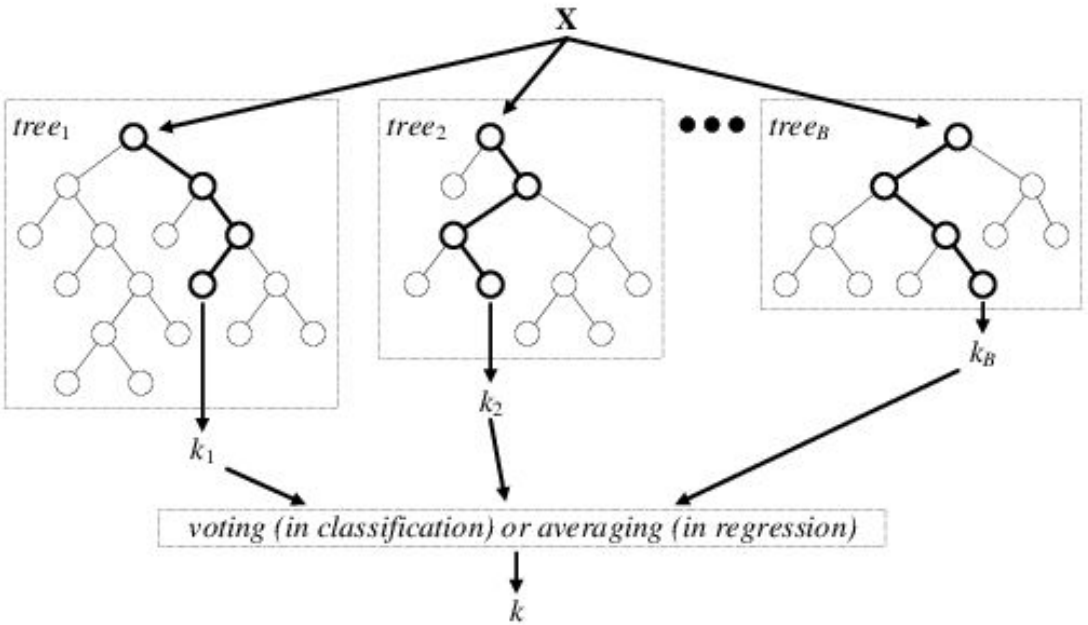


Cross Validation Score

r-squared

79.63%

Model #3: Random Forest Regressor



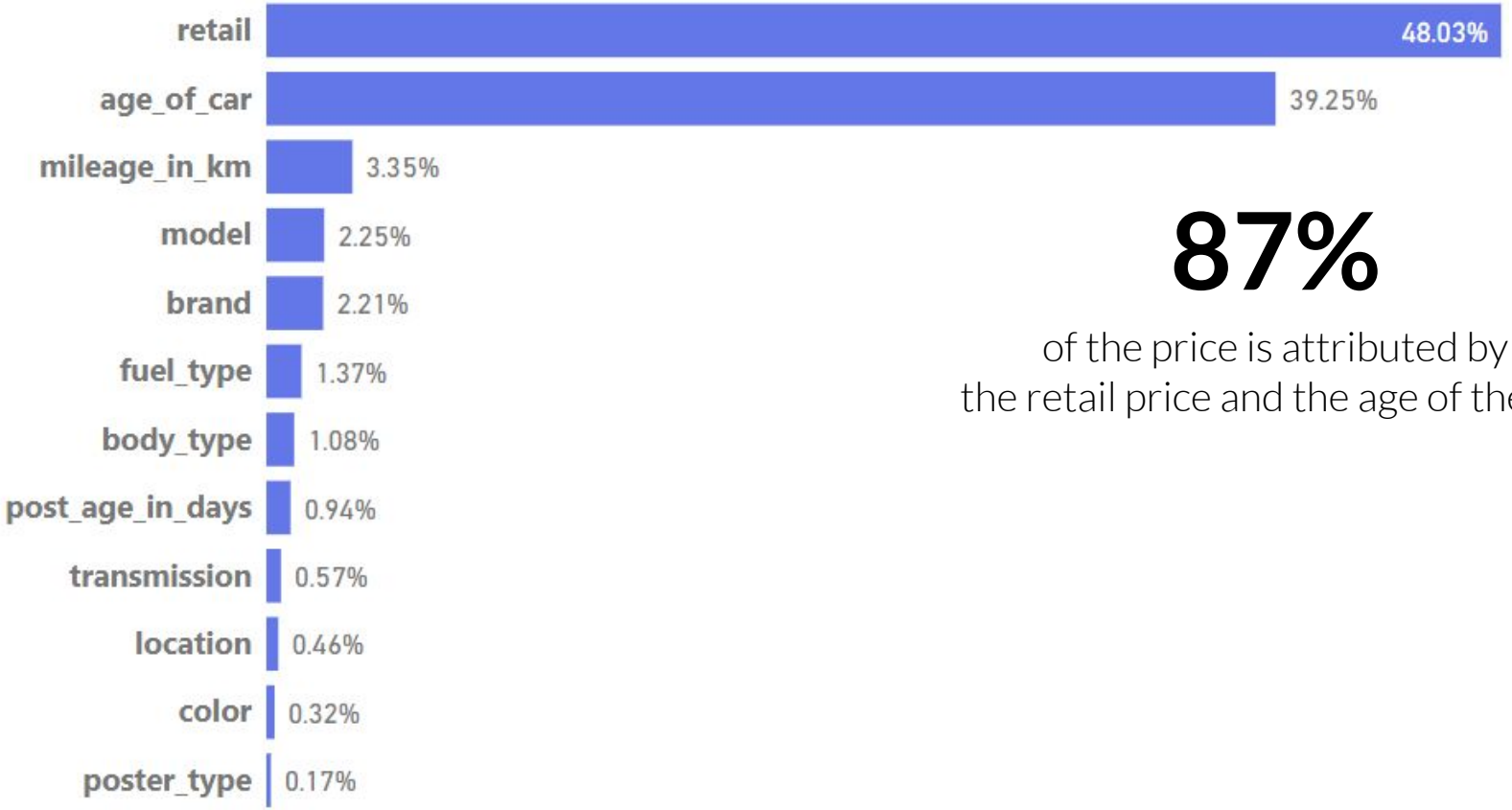
Out-of-Bag R^2 Score

84%

Cross Validation Score
r-squared

80%

Retail price and age of car greatly affect the price of used cars

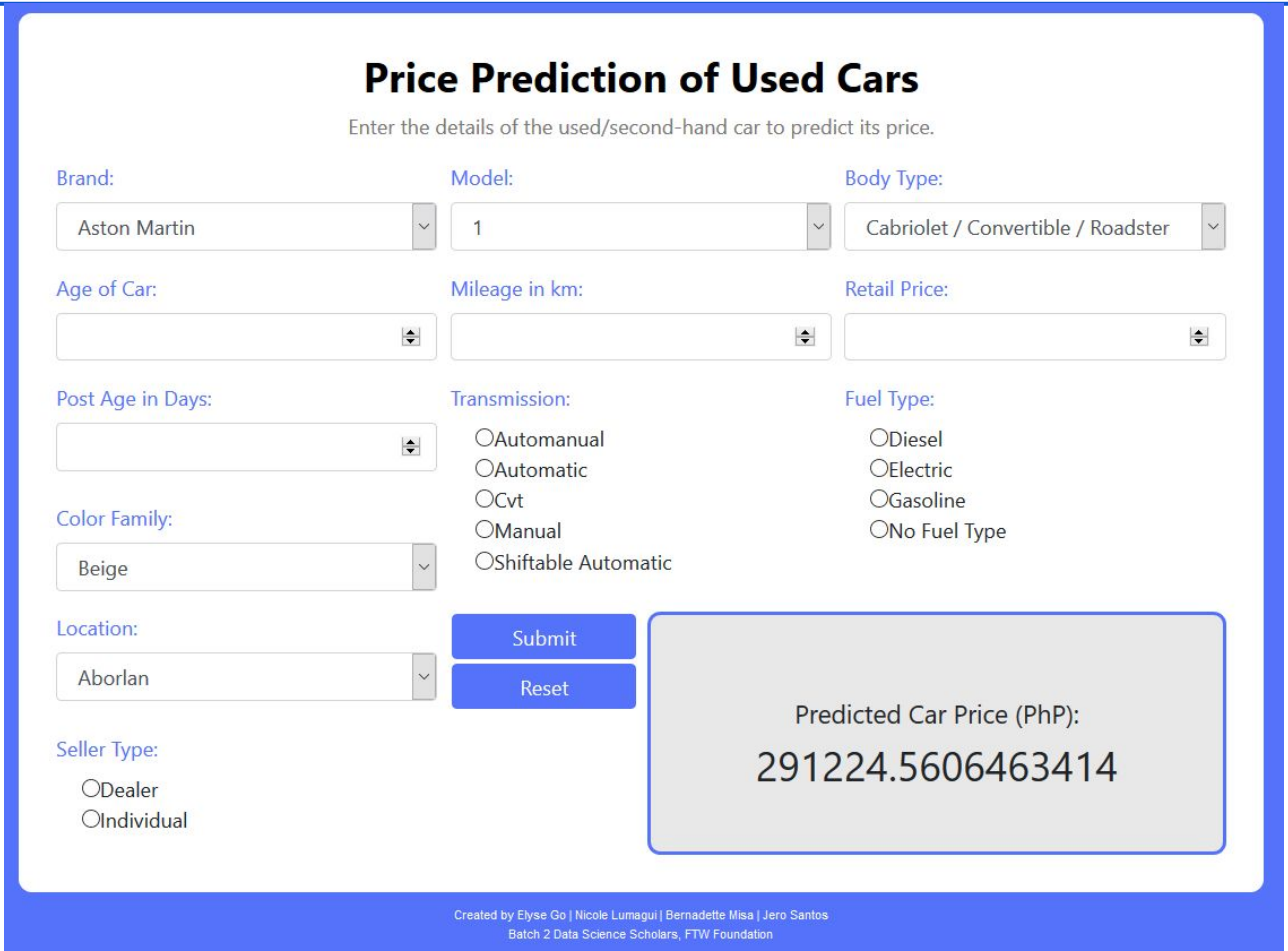


87%

of the price is attributed by
the retail price and the age of the car

FINAL OUTPUT

Our Model in Action (Prototype/Alpha version)

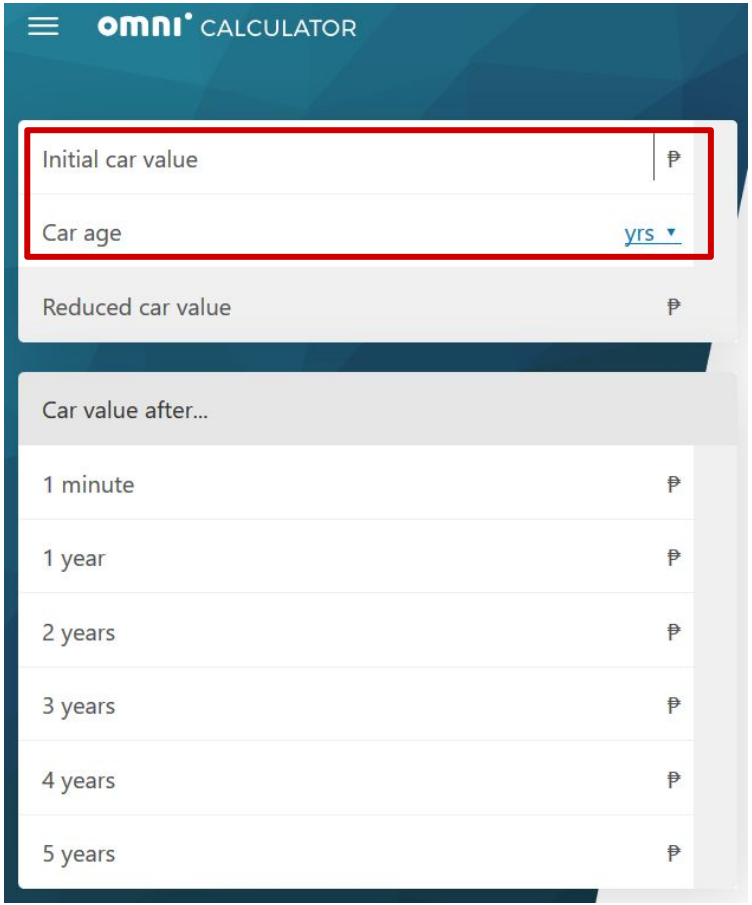


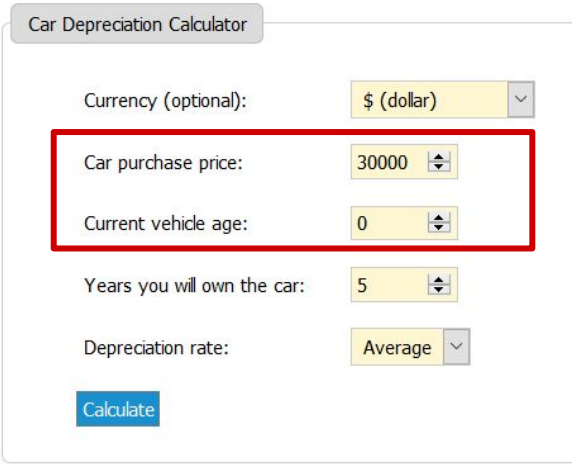
SECOND HAND CARS IN THE PH: A Buyer's Guide

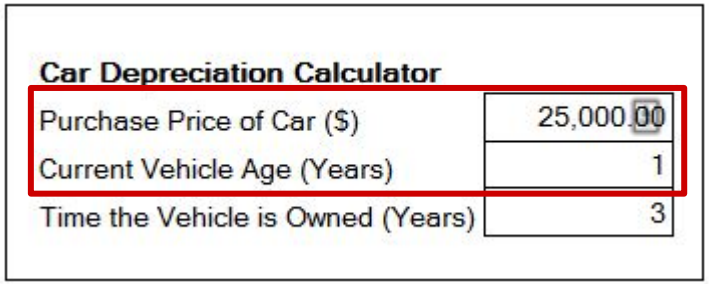
FTW Data Science Batch 2 - Team Money | Go | Lumagui | Misa | Santos |
10 August 2019

THANK YOU!

Available online calculators usually cover the basic features and are not localized

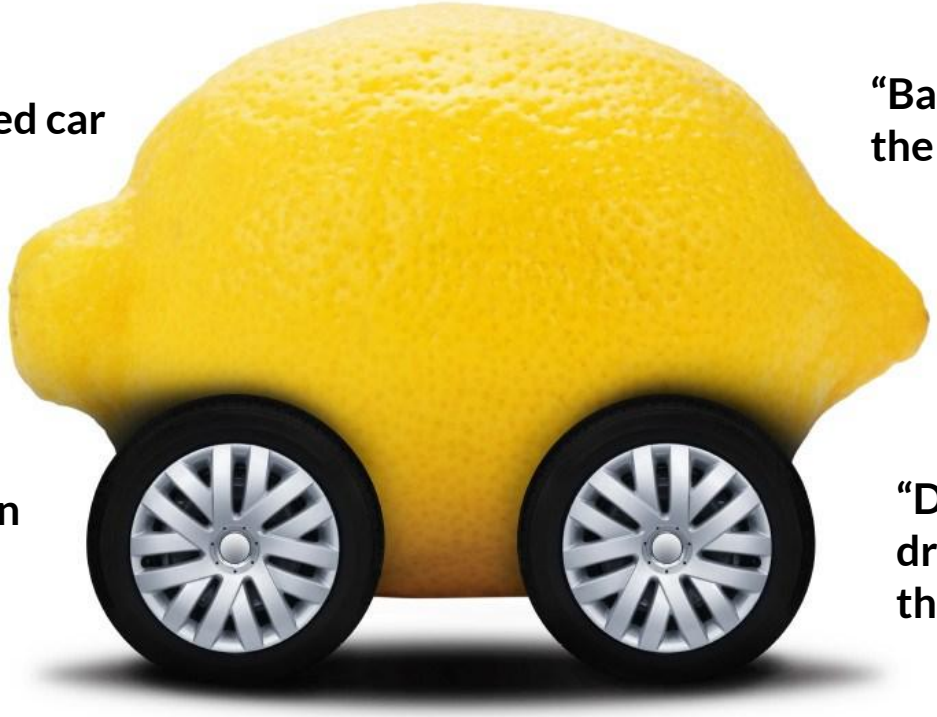






What is a lemon car?

Lemon Car = Overpriced car



"Bad cars tend to drive out the good [cars]"

Sellers know more than the buyer

"Dishonest dealings tend to drive honest dealings out of the market"

Recommendations

Next Steps:

1. Adding more data from other marketplaces
2. Talk about what the model can and can't do (make a slide)
3. Use data from bank's prices on cars (repossessed)
4. Highlight what's surprising (poor relevance of dealer/individual)

References

1. Akerlof, G. “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism”. The Quarterly Journal of Economics, vol. 84, no. 3, 1970, pp. 488-500.
2. Pudaruth,S. “Predicting the Price of Used Cars using Machine Learning Techniques”. International Journal of Information & Computation Technology, vol. 4, no. 7, 2014. pp. 753-764.
3. Monborinon et. al. “Prediction of Prices for Used Car by Using Regression Models “. 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, pp. 115-119.
4. Gegic et. al. “Car Price Prediction using Machine Learning Techniques”. TEM Journal, vol. 8, issue 1, 2019. pp. 113-118.
5. “Hedonic Methods of Price Measurement for Used Cars”. German Federal Statistics Office. 20 Oct 2003. Web. 29 July 2019.
6. Venturi, D. “The Data Science Process”. 27 Jan 2017. Web. 30 Jul 2019.
7. “5 Reasons Why You Should Buy a Used Car”. Business Standard. 20 May 2016. Web. 31 July 2019.
8. Smith, L. “Car Shopping: New or Used?” Investopedia. 22 Jul 2019. Web. 02 Aug 2019.

Image Sources:

1. https://www.carlove.ph/
2. https://st.motortrend.com/uploads/sites/11/2018/07/Mulsanne-WO-Edition-hero.jpg
3. http://www.autoinsurancecompanion.com/wp-content/uploads/2015/06/car-insurance.png
4. https://pngimage.net/wp-content/uploads/2018/06/money-blue-icon-png-5.png
5. https://www.pngkey.com/detail/u2w7q8y3t4e6y3t4_how-best-to-manage-vehicle-depreciation-icon/
6. https://ii.alatest.com/css/b2b/bootstrap/icon-shieldblue.svg
7. https://img.playbuzz.com/image/upload/q_auto:good,f_auto,fl_lossy,w_640,c_limit/v1539199075/malv1khwwkzctorcw0k5.png
8. https://img.playbuzz.com/image/upload/q_auto:good,f_auto,fl_lossy,w_640,c_limit/v1539199075/x5anm65ggltf1m5perig.png
9. https://business.inquirer.net/files/2018/01/Carmudi-logo.jpg
10. https://pbs.twimg.com/profile_images/827469003886956544/6VecITV0_400x400.jpg
11. https://www.autodeal.com.ph/images/essentials/logo_black.svg
12. https://business.inquirer.net/files/2018/01/Carmudi-logo.jpg
13. https://pbs.twimg.com/profile_images/827469003886956544/6VecITV0_400x400.jpg
14. https://www.autodeal.com.ph/images/essentials/logo_black.svg
15. https://www.holmanparts.com/wp-content/uploads/maintenance-icon.png
16. http://pngimg.com/uploads/question_mark/question_mark_PNG49.png
17. https://dumielauxepices.net/wallpaper-2207748
18. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcT10rMqofw_MgHLeHXUOqeNQwZRt0y4rx8hvLTTU3ZfSBH8cHTB
19. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcS4KaKKH8ypAelm02kz_IM5MKUlFe5pecAgH2iottJaY-docVU7
20. https://www.kisspng.com/png-computer-icons-tree-garden-arborist-deciduous-tree-5105096/
21. https://i0.wp.com/www.usedcarguys.com.au/wp-content/uploads/2017/08/avoid-buying-bad-car.jpg
22. https://miro.medium.com/max/1280/1*b9BXv0uAkbSAn8MJIa4-_Q.gif

Glossary of Terms

Make - brand of the car

Mileage - total distance traveled for a given time

Age - the time between the car's year of manufacture and the present

Body type - the design or style of a car

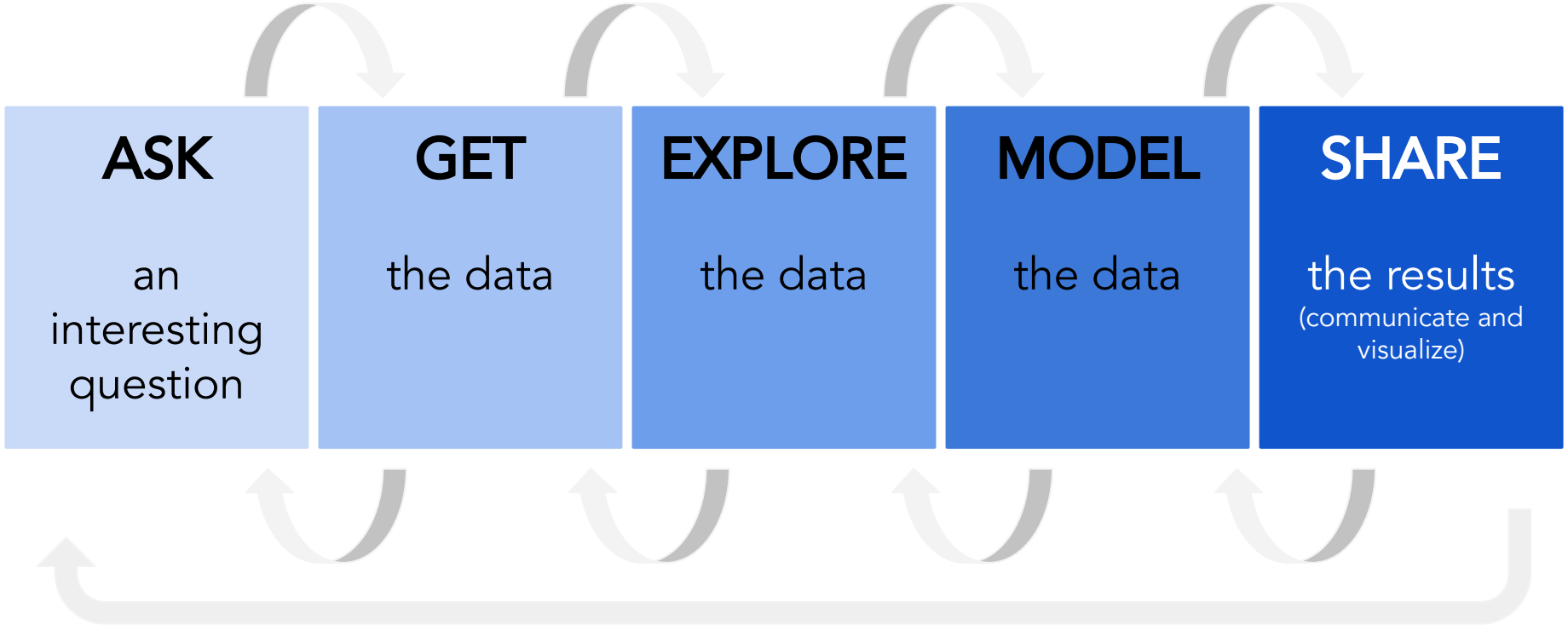
Location - the user's (creator of the classified ad) location

Transmission - how a car shifts gears (e.g. automatic or manual)

Fuel Type - what type of fuel it uses

Engine Size - how large a space the engine's pistons operate in

The Data Science process has five major and continuous steps



Venturi, D. “The Data Science Process”. 27 Jan 2017. Web. 30 Jul 2019.

Assumptions were made to simplify analysis and modeling

Dropped less relevant columns - name of user (who posted the ad), description (free text info about the listing),

Data was scraped from the used car listings in top buy and sell websites in the Philippines

	WEBSITE TRAFFIC [1]
PHILKOTSE.COM	2.33 M
AUTODEAL	2.20 M
carmudi PHILIPPINES	796.89 K

[1] Obtained from SimilarWeb.com and total visits figures are for Jun 2019 only.

Hedonic regression is used to predict prices based on the “adjustments in quality” of a product

Measures the impact of product features on sale price

Hedonic regression is used to calculate for the price index, expressed as a logarithmic function:

$$\ln(SP) = \alpha + \beta_1 \cdot age + \beta_2 \cdot kil + \beta_3 \cdot \ln(NP) + \gamma_1 \cdot D_{1\ brand} + \ldots + \gamma_{15} \cdot D_{15\ brand} + \delta \cdot D_{time} + \varepsilon$$

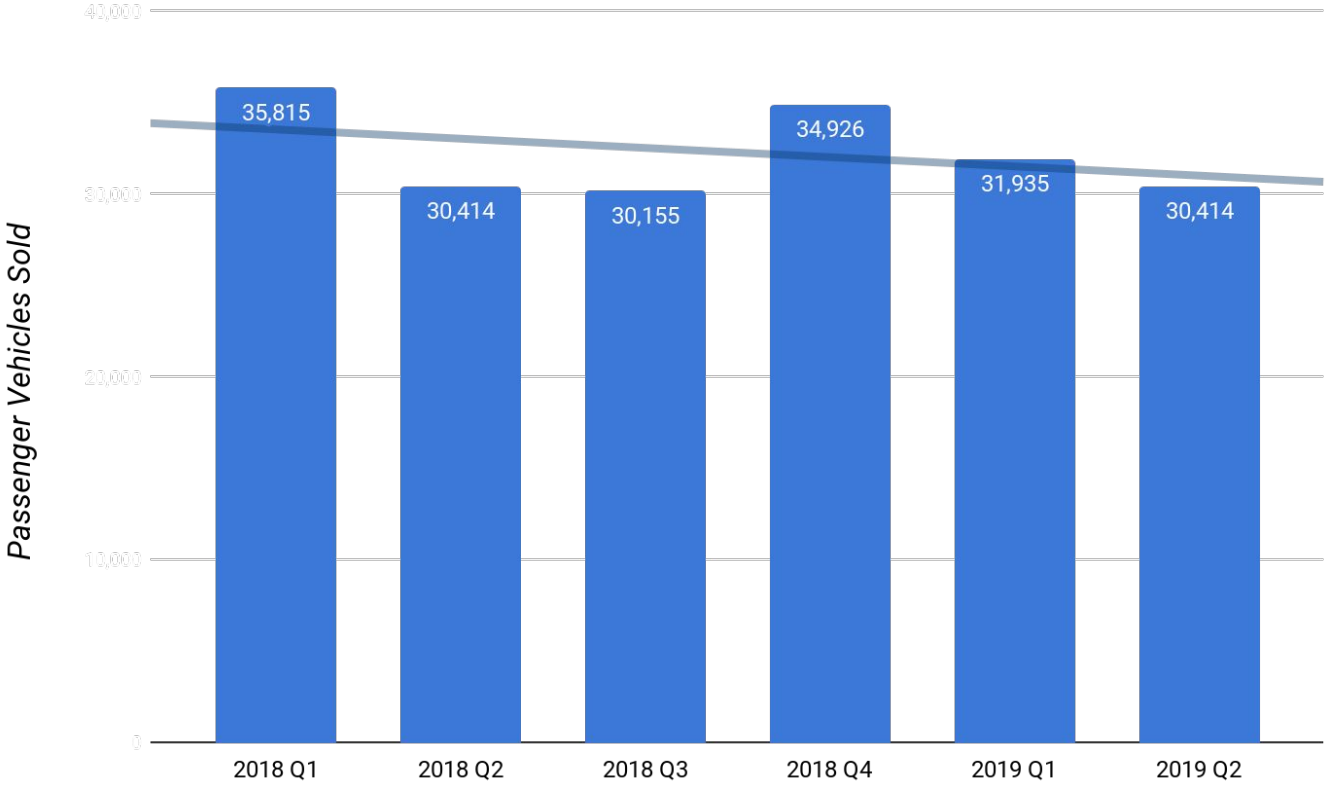
where

Figure 4: Variables in the regression function

Symbol	*Designation*
SP	sale price
NP	deflated original price of the new car
age	age of car in months
kil	relative mileage (kilometres travelled per month of age)
D_{brand}	15 dummy brand variables (Audi, BMW, Mercedes Benz, VW, etc.)
D_{time}	time dummy variable
α	absolute term
β, γ, δ	coefficient estimator
ε	random variable

“Hedonic Methods of Price Measurement for Used Cars”. German Federal Statistics Office. 20 Oct 2003. Web. 29 July 2019.

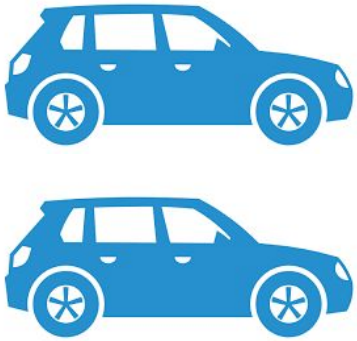
Passenger vehicle sales in 2018 to 2019



Have the right information to make an informed decision

Is this priced correctly?

Am I getting my value for money?



How does it compare with those available in the market?