



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IX **Month of publication:** September 2026

DOI: <https://doi.org/10.22214/ijraset.2026.84923>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

What Holds Its Value? Price Determinants in India's Used Motorcycle Market and the Absence of the Yamaha RX100

Perseus Bhavnagri

Department of Information Technology, Wilson College, University of Mumbai

Abstract: *The Yamaha RX100 retains a strong reputation among Indian motorcycle enthusiasts, yet it does not appear once in 7,324 listings of used motorcycles offered for sale online in India. This paper starts from that absence and asks a broader question: what determines the asking price of a used two-wheeler in India? The data are a public, CC0-licensed file of listings scraped from the marketplace droom.in. The file holds 32,648 rows, but 25,324 of them are exact duplicates, leaving 7,324 distinct listings; the duplication is not disclosed by the publisher and is documented here. A hedonic regression explains 84.7 per cent of the variation in log asking price. Engine displacement, age and brand each carry a large partial effect, and asking price falls by about 7.1 per cent for each year of age. Ownership history matters much less than its place in every listing suggests: holding other attributes constant, a second-owner machine is listed 4.1 per cent below a comparable first-owner one (95 per cent confidence interval 1.4 to 6.7 per cent), and ownership has a partial η^2 of 0.002. A deliberate check for Simpson's paradox found none: the negative relationship between age and price holds within every brand and every displacement band examined. Depreciation slows with age, and the fitted curve turns upward at about 25 years (bootstrap 95 per cent interval 19.7 to 30.0 years), but that upturn rests on 17 to 20 listings, most of them Royal Enfields, and is reported as suggestive. Royal Enfield is listed 33.5 per cent above the reference brand, Bajaj, on otherwise comparable machines. A search of 173 further public datasets found only twelve RX100 sale listings among them. All results describe asking prices, not transaction prices.*

Keywords: *Used motorcycle market, hedonic pricing, resale value, India, two-wheeler, Yamaha RX100, effect size, Simpson's paradox, secondary data*

I. INTRODUCTION

Two-wheelers dominate personal motorised transport in India. In a public sample of Indian vehicle registration records from 2020 to 2025 they account for 73.6 per cent of the vehicles registered, and a single online marketplace, the one studied here, listed 7,324 distinct used two-wheelers for sale. Yet how second-hand prices are set is poorly documented. No public index of used two-wheeler values was found in the course of this study, and the published work on used-vehicle prices, reviewed below, deals with cars far more often than with motorcycles.

This paper asks a plain question: when someone lists a used motorcycle for sale in India, what explains the price they ask? The candidate answers are the obvious ones—how old the machine is, how far it has been ridden, how big its engine is, how many people have owned it, who made it and where it is being sold. That these matter is not in doubt. What this review did not find is an estimate of how much each matters, relative to the others, for Indian two-wheeler listings. Providing one is the paper's first contribution. Its second is to show that ownership history, which every listing in the data reports, carries far less weight than its prominence suggests.

A. The Starting Point: A Motorcycle That Is Not There

The study began with a different question. The Yamaha RX100 is a small motorcycle with a 98cc engine, made from 1985 to 1996 according to a public specification dataset. It was chosen as the starting point because of the standing it retains among Indian motorcycle enthusiasts. The original intention was to test whether it holds its value better than an ordinary motorcycle does.

That study could not be done, and the reason is informative. The dataset used here, which holds 7,324 distinct Indian listings, contains no RX100 at all, and a search of 173 other public datasets, reported with the findings, found only twelve sale listings. The machine is well remembered but is rarely offered for sale through channels that leave a public record. That gap frames the analysis that follows.

B. What This Paper Does

The paper builds a hedonic price model on the listings that do exist, reports the contribution of each attribute as an effect size rather than by significance alone, and tests explicitly for the aggregation error known as Simpson's paradox. It closes by returning to the RX100. The data are public, every number reported comes from a script run on those data, and the scripts are retained by the author.

II. PROBLEM STATEMENT AND NEED FOR THE STUDY

A. Problem Statement

Prices in India's used two-wheeler market are hard to read from the outside. A listing reports a machine's age, distance covered, engine size, ownership history, make and location, but it does not say which of these the asking price reflects, or by how much. Sellers cannot easily tell which attributes of their machine carry value, and buyers cannot easily tell whether a brand's higher asking prices reflect its machines or its badge.

A further difficulty is statistical. Comparisons between groups of vehicles—one brand against another, older machines against newer ones—are easily made on raw averages, which is unsafe when the groups differ in composition. A brand that sells mostly large machines will show higher average prices whatever its machines are worth, like for like.

B. Need for the Study

The need arises on four fronts.

- *Practical.* One marketplace alone listed 7,324 distinct used two-wheelers. Knowing which attributes move asking price, and by how much, is directly useful to both sides of such sales.
- *Methodological.* Recent work on used-vehicle prices concentrates on prediction accuracy. Some studies rank attributes by their importance to a fitted model, but few express each attribute's contribution as an effect size that can be compared across studies.
- *Statistical.* Whether aggregation distorts group comparisons in this market is a question that can be answered rather than assumed, and answering it honestly means reporting the result whichever way it falls.
- *Documentary.* The dataset used here is distributed as a file of 32,648 rows, of which only 7,324 are distinct. Placing that on record is a small but real contribution for anyone else who uses it.

III. LITERATURE REVIEW

A. Used Vehicles as an Economic Category

The theoretical starting point for work on second-hand vehicles in developing economies is Grubel's price-theoretic account of why cars in such economies lose value more slowly and stay in service longer than in industrial ones [1]. The main reason he gives is that repair labour is cheaper, so an old car can be kept running at lower cost; a second is that buyers with lower incomes place less value on the features of new models. The present study observes a single country and cannot test that comparison, but the argument frames what a slow rate of depreciation would mean.

Nakamoto approaches used cars from an environmental direction [2]. Modelling the Japanese car market, he finds that longer vehicle lifetimes and a larger market for used cars can lower life-cycle carbon emissions. The relevance here is that the used-vehicle market is treated as one of the mechanisms that decide how long vehicles remain in use.

B. Hedonic Price Modelling

The hedonic approach treats a price as the sum of what buyers pay for each attribute of a good. Koç and Kostak apply it to online advertisements for two popular car models in Turkey during a period of sharp price rises [3]. Their linear models, fitted separately to each car model and each period, use equipment level, model year, mileage, transmission and fuel type, and explain between 69 and 80 per cent of the variation in price. Because they too work with asking prices from online listings, theirs is the closest published parallel to this study, although they model price in levels rather than logarithms and each of their models covers a single make and model.

Francke and Van de Minne address a standing weakness of the method: characteristics that affect price are often missing from the data, which leaves unobserved differences between the goods being compared [4]. They develop their treatment for property markets, but the problem applies directly here, where mechanical condition, service history, accident record and modification status are all unobserved.

C. Prediction-Oriented Studies of Used-Vehicle Price

A large recent literature applies machine learning to used-vehicle prices. Gegic and colleagues combine an artificial neural network, a support vector machine and a random forest into an ensemble trained on listings scraped from a Bosnian web portal, and report a test accuracy of 87.38 per cent [5]. Three studies published in this journal apply similar methods: Bharambe and colleagues compare algorithms on a public dataset of 302 used cars [6], Kalpana and colleagues work with about 50,000 listings from an online marketplace [7], and Ranjith and colleagues study the resale value of used luxury cars in India [8]. Budiono and colleagues train models on 504 listings scraped from an Indonesian marketplace [9], and Barlybayev and colleagues frame the task explicitly as a developing-market problem, using data from Kazakhstan [10].

Together these studies show, across several countries, that used-vehicle price can be predicted well from a small set of observable attributes. Two also rank the attributes. Ranjith and colleagues find engine power, seller type, vehicle age and distance driven to be the most influential factors for Indian luxury cars [8], and Barlybayev and colleagues use feature importance and a correlation matrix to choose their predictors [10]. What none of the six reports is the size of each attribute's contribution in a form that can be compared across studies, or a check of whether a relationship seen in the pooled data holds within its subgroups.

D. Two-Wheelers as a Distinct Market

Two-wheelers are not small cars, and the literature treats them separately. Das and colleagues examine the Indian two-wheeler fleet from the standpoint of emissions inspection and maintenance [11]. Assuming 5,000 km of riding a year, they estimate that most machines in their Delhi sample were six to nine years old, younger than they had expected for a developing country, and they record Delhi legislation to ban two-wheelers more than 15 years old. Shaharudin and colleagues study purchase intention for the products of Malaysia's national motorcycle manufacturer and find that customers' perceptions of product quality had no significant effect on their purchase decisions; buyers, they conclude, weigh considerations beyond perceived quality [12].

More recent two-wheeler research has concentrated on electrification. Balijepalli and colleagues model preferences for electric motorcycles in Bandung, Indonesia [13], and Bakker sets out the part electric two-wheelers can play in sustainable urban mobility [14]. Cirolia and colleagues take a different angle, examining how motorcycle taxis in Kigali and Nairobi are being drawn onto digital platforms [15]. This work establishes the two-wheeler as a serious research object, but none of it addresses how second-hand prices are formed.

E. Aggregation, Effect Size and Honest Reporting

Two bodies of methodological work inform the analysis. The first concerns Simpson's paradox, in which a relationship visible in pooled data reverses within the subgroups that make up the pool. Kievit and colleagues give a practical guide for researchers [16], Armstrong and Wattenberg show how such mix effects can be exposed visually [17], and Walton and Walton show how the paradox can distort the 'leaky pipeline' plots used to inform gender-equality policy [18]. The shared advice is to check for a reversal rather than assume there is none.

The second concerns effect size. Lakens treats the effect size, rather than the p -value, as the main result a study produces, because it conveys practical importance and lets findings accumulate across studies [19]. Hanel and Mehler compare how informative lay readers find different effect-size statistics and argue that reporting them alongside significance improves communication with the public [20]. With several thousand observations, small differences reach significance routinely, so this guidance matters here: one of the results below is statistically significant but small, and is reported as such.

F. Gap Analysis

Table I summarises the twenty sources read for this study, what each contributes and what it leaves open. The reference numbers match the list at the end of the paper.

TABLE I
GAP ANALYSIS OF THE REVIEWED LITERATURE

Ref.	Author and year	Area	What it leaves open
[1]	Grubel (1980)	Used cars and economic development	Theoretical; no vehicle-level price data
[2]	Nakamoto (2017)	Used-car lifetimes and CO2 emissions, Japan	Emissions focus; prices not modelled

[3]	Koç and Kostak (2021)	Hedonic model, Turkish used cars	Cars only; two car models; linear in price
[4]	Francke and Van de Minne (2021)	Unobserved heterogeneity in hedonic models	Method paper; property market
[5]	Gegic et al. (2019)	Machine-learning price prediction, Bosnia	Accuracy reported; no attribute effect sizes
[6]	Bharambe et al. (2022)	Used-car price prediction	Small public dataset; prediction only
[7]	Kalpana et al. (2022)	Pre-owned car price prediction	Cars; prediction only
[8]	Ranjith et al. (2024)	Resale value, Indian luxury cars	Luxury cars; importance ranking, not effect sizes
[9]	Budiono et al. (2024)	Used-car price model, Indonesia	Cars; prediction only
[10]	Barlybayev et al. (2023)	Resale value, developing markets	Cars in Kazakhstan; no two-wheelers
[11]	Das et al. (2001)	Two-wheeler inspection and fleet age, India	Emissions focus; prices not addressed
[12]	Shaharudin et al. (2011)	Motorcycle quality and purchase intention	New purchases; not resale value
[13]	Balijepalli et al. (2023)	Electric motorcycle preferences, Indonesia	Stated preference for new electric machines
[14]	Bakker (2019)	Electric two-wheelers and cities	Policy discussion; no price data
[15]	Cirolia et al. (2023)	Motorcycle taxis and digital platforms	Platforms and labour; no pricing
[16]	Kievit et al. (2013)	Simpson's paradox, practical guide	Method; no vehicle application
[17]	Armstrong and Wattenberg (2014)	Visualising mix effects	Visualisation method only
[18]	Walton and Walton (2016)	Simpson's paradox in policy data	Different domain
[19]	Lakens (2013)	Calculating and reporting effect sizes	Method; experimental designs
[20]	Hanel and Mehler (2019)	Communicating effect sizes	Method; communication focus

The pattern across the table is clear. Studies of used-vehicle prices concentrate on cars and on predictive accuracy. Two-wheeler research is active but concerned with fleet condition, adoption and platforms rather than resale price. The methodological literature on aggregation and effect size is mature but, among the sources reviewed, has not been applied to this market. No study located models second-hand two-wheeler prices in India, reports comparable effect sizes for the attributes that drive them and tests for aggregation bias. That is the gap this paper addresses.

IV. OBJECTIVES, VARIABLES AND HYPOTHESES

A. Objectives

- To measure the association between a used two-wheeler's asking price and its age, distance covered, engine displacement and ownership history in the Indian market.
- To estimate the price premium attributable to the manufacturer once machine characteristics are held constant.
- To rank these factors by the size of their contribution rather than by statistical significance alone.

- To test whether the pooled age–price relationship survives disaggregation by brand and by engine displacement, and so establish whether Simpson’s paradox is present.
- To examine whether price falls steadily with age or whether the oldest surviving machines recover value.
- To document the presence or absence of the Yamaha RX100 in the Indian online resale market.

B. Identification of Variables

Table II sets out every variable used in the analysis and how each was defined.

TABLE II
VARIABLES USED IN THE STUDY

Type	Variable	Operational definition
Dependent	Asking price	Listed price in Indian rupees; modelled as its natural logarithm
Independent	Age	Years since manufacture at the time the listing was collected
Independent	Distance covered	Kilometres driven; modelled as its natural logarithm
Independent	Engine displacement	Cubic capacity in cc
Independent	Ownership history	First, second, third, or fourth and later owner
Independent	Brand	Manufacturer, as a fixed effect with Bajaj as reference
Contextual	City	City of listing; 443 distinct values, used descriptively
Derived	Age band	Seven bands from 1–3 years to 21 years and above
Derived	Displacement band	Five bands from 110cc and below to above 400cc

C. Hypotheses

The hypotheses were written down, each as a null and an alternative, before any test was run, and the file is retained with the analysis scripts. They are summarised here by their alternative hypotheses.

- H1. Older motorcycles are listed at lower prices.
- H2. Motorcycles with more kilometres covered are listed at lower prices.
- H3. Larger-displacement motorcycles are listed at higher prices.
- H4. Motorcycles with more previous owners are listed at lower prices.
- H5. Manufacturers differ in price after machine characteristics are controlled, with Royal Enfield commanding a premium.
- H6. The direction of the age–price relationship is the same in the pooled data as within brand and within displacement class. This is a check for confounding, not a result hoped for.
- H7. Price declines with age up to a point and then rises again for the oldest machines. It was recorded in advance that this test would be underpowered.

An eighth item in the same file, the presence of the RX100, was designated in advance as an observation rather than a hypothesis, because no inference can be drawn from a count.

D. Departures from the Stated Plan

Three departures from the plan as written are reported so that a reader can weigh them. First, H6 was specified as a comparison of pooled and within-stratum regression slopes; it was carried out with Spearman rank correlations, which test the same question, whether the direction of the relationship survives disaggregation, without assuming it is linear. Second, the four robustness checks on H7 were designed after the quadratic result had been seen, and should be read as exploratory.

Third, the adjusted ownership coefficients reported under H4 were added during a review of the paper before submission, because the bivariate effect size alone could not show how much of the raw price difference between owner groups came from their composition.

V. RESEARCH METHODOLOGY

A. Research Design

This is a quantitative, descriptive and inferential study based entirely on secondary data. No primary data was collected and no respondents were surveyed. The unit of analysis is a single published listing for a used motorcycle, not a transaction and not a machine. This distinction matters and is returned to in the limitations.

B. Data Sources

The main data are a public file of used motorcycle listings scraped from droom.in, an Indian online vehicle marketplace, and published on Kaggle by Sai Saathvik Domala under a CC0 Public Domain dedication (kaggle.com/datasets/saisaathvik/used-bikes-prices-in-india). The file, *Used_Bikes.csv*, is a single comma-separated table of 2.49 MB with eight fields: *bike_name*, *price*, *city*, *kms_driven*, *owner*, *age*, *power* and *brand*. It was downloaded on 28 August 2026, and a second download on 22 September 2026 was byte-identical to it. The listings themselves carry no dates. The file inside the published archive is time-stamped 1 June 2021, so the listings were collected on or before that date, and each machine's age is measured at the time of collection.

One field name is misleading and is corrected throughout. The column named *power* contains engine displacement in cubic centimetres, from 100 to 1,800, not power output; it is referred to here as *displacement*.

Two contextual statements, the share of two-wheelers among registered vehicles and the count of RX100 registrations, draw on a second public file: a random sample of 500,000 aggregated Indian vehicle registration records from 2020 onwards, published on Kaggle by Aatif Ahmad under a CC BY 4.0 licence (kaggle.com/datasets/aatifahmad123/indian-vehicle-registration-data-202025). Its publisher describes it as derived from 'RTO-style administrative records' without naming the original source, so it is used only for context and never for testing.

C. Data Quality and Cleaning

The supplied file contains 32,648 rows. Checking for exact row-level duplication found that 25,324 of these rows, or 77.6 per cent, are exact copies of another row. The duplication is not random. Thirty-eight individual listings each appear between 608 and 620 times, and these alone account for the majority of the excess rows. Fig. 1 shows the distribution.

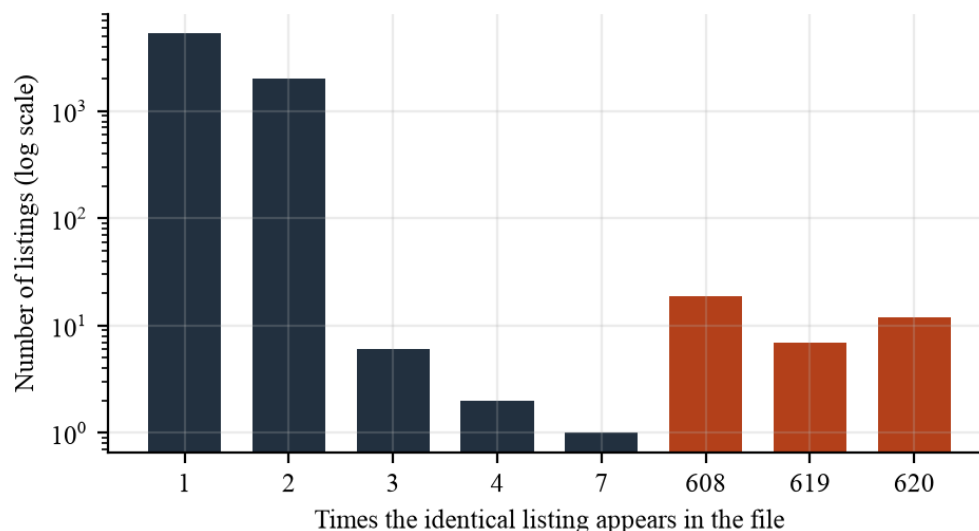


Fig. 1 Repetition of identical listings in the source file. Thirty-eight listings appear more than 600 times each.

This is a fault in the original scrape rather than a property of the market, and the dataset's description does not mention it. Keeping the duplicates would let thirty-eight machines dominate every estimate in the study. All duplicates were therefore removed, leaving 7,324 distinct listings, which form the analytical sample throughout. Every number in this paper refers to that deduplicated sample unless stated otherwise.

No values were missing in any field. Nine listings record distances above 300,000 kilometres, including a 100cc commuter at 750,000 kilometres, which is not credible. These were flagged rather than deleted; because the model uses the logarithm of distance, their influence is limited. No observation was removed on the basis of its price.

D. Analytical Approach

Price is strongly right-skewed, with a skewness of 6.27. Taking natural logarithms reduces this to 0.54, so log price is the dependent variable throughout and non-parametric tests are used for bivariate association. Fig. 2 shows both distributions.

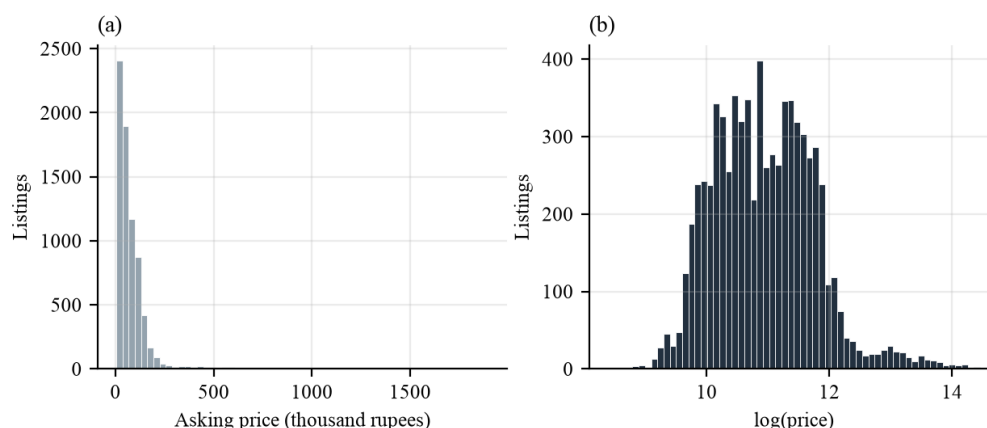


Fig. 2 Distribution of asking price (a) in rupees and (b) after logarithmic transformation

Bivariate associations use Spearman rank correlation, whose coefficient ρ serves as its own effect size. Group comparisons use the Kruskal–Wallis and Mann–Whitney U tests, reported with ε^2 , the rank-biserial correlation and Cohen’s d . The main model is an ordinary least squares hedonic regression of log price on age, log distance, displacement, ownership category and brand fixed effects. Partial η^2 , computed from Type II sums of squares, compares the contribution of predictors measured on different scales. The significance level is 0.05 throughout. The analysis was carried out in Python with pandas, SciPy and statsmodels; the scripts are retained by the author, and every number in this paper can be traced to their saved output.

E. Sample Profile

Table III describes the analytical sample.

TABLE III
DESCRIPTIVE STATISTICS FOR THE ANALYTICAL SAMPLE (N = 7,324)

Statistic	Price (Rs)	Distance (km)	Age (years)	Displacement (cc)
Mean	84,884	23,911	6.7	228.1
Standard deviation	120,966	27,318	3.6	158.3
Minimum	4,400	1	1	100
First quartile	30,000	10,156	4	125
Median	55,000	19,000	6	160
Third quartile	100,000	30,112	8	350
Maximum	1,900,000	750,000	63	1,800

The sample covers 471 distinct model names across 443 cities. Bajaj is the most common manufacturer with 2,081 listings, followed by Royal Enfield with 1,346, Hero with 1,142, Honda with 676 and Yamaha with 651. Ownership is heavily concentrated: 6,642 listings, or 90.7 per cent, are advertised as first-owner machines, with 588 second-owner, 84 third-owner and 10 fourth-owner or later.

VI. DATA ANALYSIS AND FINDINGS

A. What Moves Price: Bivariate Associations

Table IV reports the rank correlation between asking price and each of the three continuous attributes. All three hypotheses are supported, all three coefficients are large, and each has the predicted sign.

TABLE IV
BIVARIATE ASSOCIATIONS WITH ASKING PRICE (SPEARMAN, N = 7,324)

Hypothesis	Variable	ρ	p	Effect size
H1	Age (years)	-0.6028	< 0.001	Strong
H2	Distance covered (km)	-0.5318	< 0.001	Strong
H3	Engine displacement (cc)	+0.8017	< 0.001	Strong

The strongest single association is not age but engine displacement, at $\rho = 0.802$. What the Indian second-hand market prices first is the class of machine, and only then how worn it is. In the data, the median ten-year-old 350cc machine is listed at Rs 71,750 and the median three-year-old 100cc machine at Rs 41,650. Age and distance covered matter, and matter strongly, but they operate largely within a segment rather than across segments.

B. Ownership History: Significant but Small

H4 predicted that machines with more previous owners would be listed at lower prices. The prediction holds statistically, but the effect is small, and most of the raw difference between owner groups comes from their composition.

A Kruskal-Wallis test across the four ownership categories rejects the null hypothesis with $H = 18.99$ on 3 degrees of freedom and $p = 0.00027$, with $\epsilon^2 = 0.0022$. Comparing first-owner listings with all later-owner listings on log price gives $U = 2,487,989.5$ with $p = 0.000022$, a rank-biserial correlation of 0.099 and a Cohen's d of 0.143. The median asking price is Rs 55,000 for first-owner machines and Rs 47,000 for later-owner machines, 14.5 per cent lower.

By the usual benchmarks a Cohen's d of 0.143 is negligible, below the 0.2 conventionally taken as the threshold of a small effect. The raw gap is also misleading on its own, because later-owner machines are older (a median of eight years against six) and more heavily used (a median of 24,600 km against 18,338 km). The hedonic model separates these. Holding age, distance, displacement and brand constant, a second-owner machine is listed 4.1 per cent below a comparable first-owner one (95 per cent confidence interval 1.4 to 6.7 per cent; $p = 0.003$; $n = 588$). Third-owner machines show no significant difference (+3.4 per cent; $p = 0.35$; $n = 84$), and the ten listings with four or more previous owners show a positive coefficient that, with ten observations, is best treated as noise. Following the reporting guidance discussed earlier [19], [20], the fair summary is that the first-owner label is associated with a premium of about four per cent, far smaller than the raw comparison suggests.

C. The Hedonic Model

The regression of log price on age, log distance, displacement, ownership and brand explains 84.7 per cent of the variation in log asking price, with an adjusted R^2 of 0.846 and $F(28, 7295) = 1441.5$, $p < 0.001$. Table V reports the continuous coefficients.

TABLE V
HEDONIC REGRESSION OF LOG ASKING PRICE (N = 7,324)

Term	Coefficient	Std. error	t	p	95% confidence interval
Age (years)	-0.07356	0.00124	-59.23	< 0.001	[-0.0760, -0.0711]
Log distance covered	-0.10422	0.00434	-24.02	< 0.001	[-0.1127, -0.0957]
Displacement (cc)	+0.00289	0.00005	63.35	< 0.001	[0.00280, 0.00298]

The coefficients translate into plain quantities. Each additional year of age is associated with a 7.1 per cent lower asking price, holding everything else constant. A one per cent increase in distance covered is associated with a 0.10 per cent lower price. Each additional cubic centimetre of displacement is associated with a 0.29 per cent higher price, which compounds across segments: an otherwise identical 350cc machine is listed at about twice the price of a 100cc one (105.8 per cent higher).

For comparison, the linear models of Koç and Kostak explain between 69 and 80 per cent of price variation for single car models in Turkey [3]. The specifications differ too much for the figures to be compared closely, but both studies find that a handful of observable attributes account for most of the variation in online asking prices.

D. Ranking the Factors by Size, Not by Significance

Every predictor in the model is significant at any conventional threshold, so significance cannot rank them. Partial η^2 , computed from Type II sums of squares, gives a comparison that does not depend on the units of each predictor. Fig. 3 shows the result and Table VI reports the underlying values.

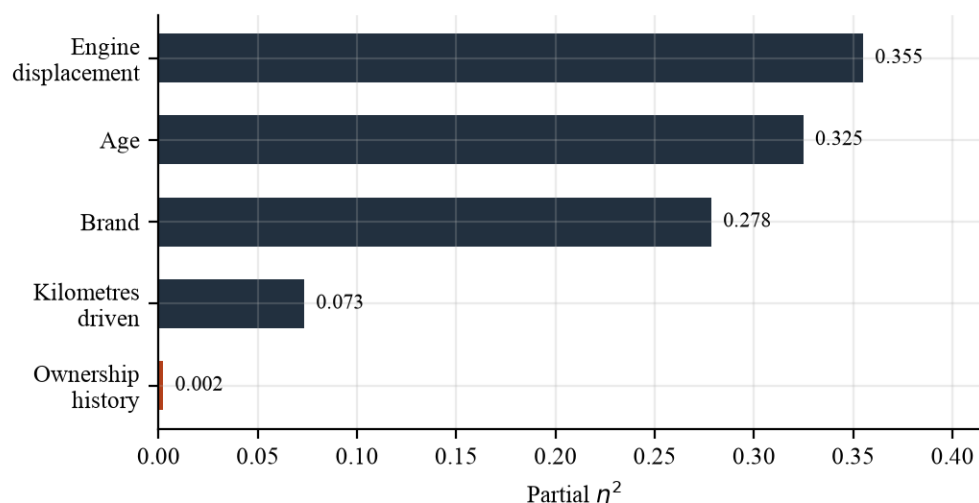


Fig. 3 Partial η^2 for each predictor in the hedonic model. Ownership history is negligible despite being statistically significant.

TABLE VI

PARTIAL EFFECT SIZES FROM THE HEDONIC MODEL

Predictor	<i>F</i>	<i>p</i>	Partial η^2	Magnitude
Engine displacement	4013.57	< 0.001	0.3549	Large
Age	3508.30	< 0.001	0.3247	Large
Brand	127.96	< 0.001	0.2784	Large
Distance covered	577.06	< 0.001	0.0733	Medium
Ownership history	5.49	0.0009	0.0023	Negligible

The ordering is informative. Displacement, age and brand each carry a large effect. Distance covered carries a medium one, about a quarter the size of the effect of age. Ownership history carries a negligible one. Most of the variation in asking price is associated with the class of machine, its age and its maker; the two attributes that describe its history, distance covered and previous owners, account for much less.

E. The Brand Premium

Brand effects are estimated relative to Bajaj, the most common manufacturer in the sample. Fig. 4 shows the estimated premium for every brand with at least fifty listings, after age, distance covered, displacement and ownership are held constant.

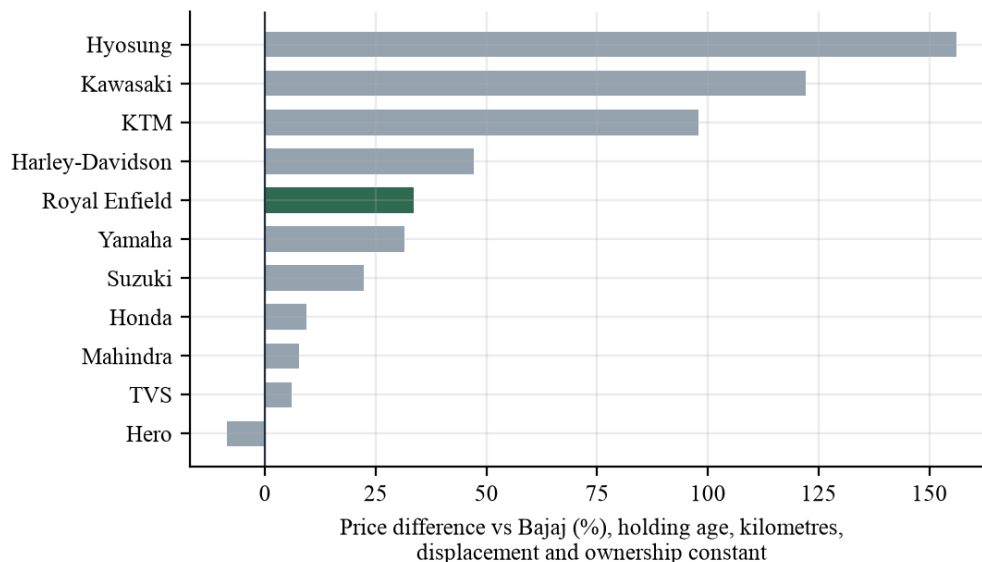


Fig. 4 Brand premium relative to Bajaj after controlling for age, distance covered, displacement and ownership.

Royal Enfield is listed 33.5 per cent above Bajaj on otherwise comparable machines, with $p < 0.001$ across 1,346 listings. Hyosung, Kawasaki and KTM show larger percentage premiums but rest on far fewer listings and occupy a different market segment. At the other end, Hero is listed 8.6 per cent below Bajaj on comparable machines. H5 is supported: brand carries a large partial effect of 0.278, and the specific prediction about Royal Enfield holds.

The premium is not explained by Royal Enfield machines being larger, newer or less used, because those attributes are in the model. It attaches to the brand rather than to the attributes observed here. It may reflect things the data cannot see, such as build quality or the availability of parts, as well as reputation. It is at least consistent with the conclusion of Shaharudin and colleagues that motorcycle buyers weigh considerations beyond perceived product quality [12], although this data cannot say what those considerations are.

F. Testing for Simpson's Paradox

H6 asked whether the pooled negative relationship between age and price survives disaggregation. Following the practical guidance in the aggregation literature [16]–[18], this was tested rather than assumed. The pooled Spearman coefficient is -0.6028 . The same coefficient was recomputed within every brand with at least fifty listings and within every displacement band, and Fig. 5 shows the result for brands.

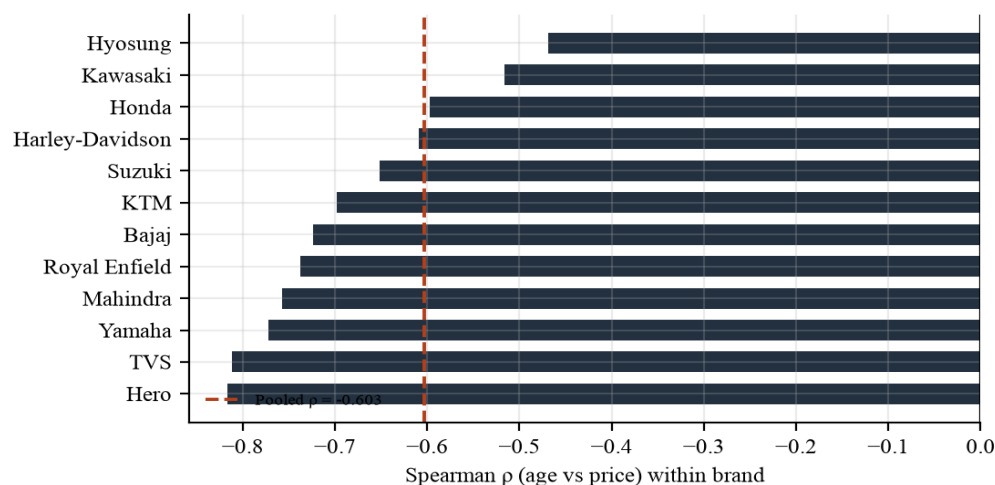


Fig. 5 Age–price rank correlation computed separately within each brand. No brand reverses the pooled direction.

No reversal was found. Every one of the twelve brands examined shows a negative age–price correlation, from -0.469 for Hyosung to -0.817 for Hero, and every one of the five displacement bands does the same, from -0.449 above 400cc to -0.802 at 110cc and below. H6 is not rejected: there is no Simpson’s paradox in the age–price relationship in this data.

This is a null result and is reported as one. It is also a useful one, because it means the pooled rate of 7.1 per cent per year can be quoted without the caveat that its direction might reverse inside some segment.

A related caution does survive. The brands in this market differ greatly in composition. Harley-Davidson listings average 799cc and have a median price of Rs 439,000, while Hero listings average 118cc with a median of Rs 28,000. A comparison of raw average prices between manufacturers would therefore measure segment mix rather than value retention. The absence of a paradox in the age relationship does not license naive comparison of group averages, and Table VII gives the composition figures that make this concrete.

TABLE VII
BRAND COMPOSITION (BRANDS WITH AT LEAST FIFTY LISTINGS)

Brand	Listings	Mean age (years)	Mean displacement (cc)	Median price (Rs)
Bajaj	2,081	6.79	174.7	42,000
Royal Enfield	1,346	6.46	389.8	110,000
Hero	1,142	8.19	118.2	28,000
Honda	676	6.91	150.2	39,950
Yamaha	651	6.76	167.4	52,096
TVS	481	5.74	162.8	48,000
KTM	375	4.81	292.9	142,600
Suzuki	203	4.80	199.7	60,000
Harley-Davidson	91	5.41	799.3	439,000
Kawasaki	61	5.00	569.7	415,000
Hyosung	53	5.68	378.3	220,000
Mahindra	50	5.52	189.8	35,000

G. Does Value Ever Come Back? The Vintage Turn

H7 asked whether price falls continuously with age or whether the oldest surviving machines recover value. Fig. 6 gives the descriptive answer.

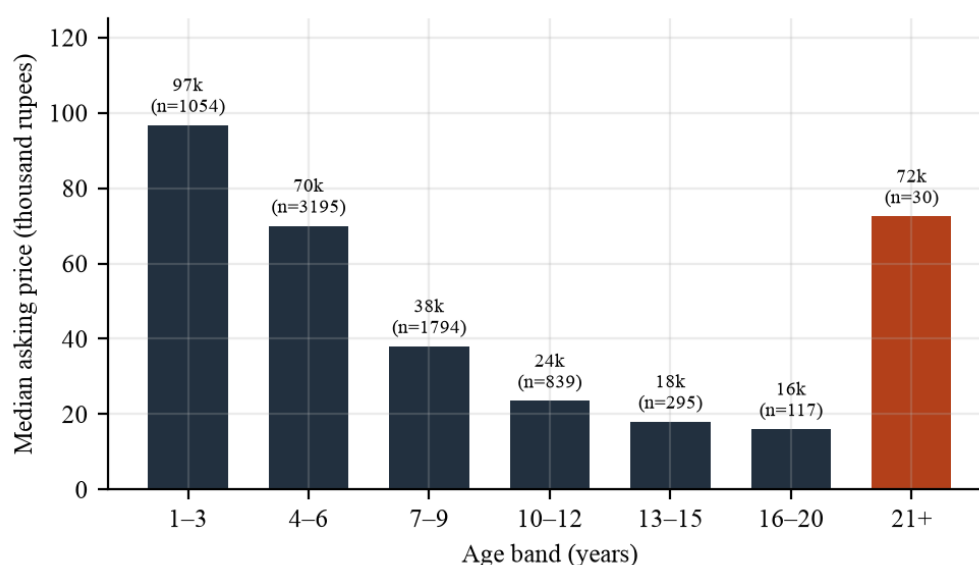


Fig. 6 Median asking price by age band. The decline reverses in the oldest band, which contains thirty listings.

The median asking price falls from Rs 96,725 in the one-to-three-year band to Rs 16,000 in the sixteen-to-twenty-year band, a decline of 83.5 per cent. In the band of twenty-one years and above it rises to Rs 72,500. The reversal is visually striking, and it would be easy to present it as proof of a collector premium. Thirty listings cannot support that, so the claim was tested rather than asserted.

Adding a quadratic age term to the hedonic model raises R^2 from 0.847 to 0.870. The squared term is positive at 0.00279 with $p < 0.001$, and a likelihood-ratio test against the linear model gives $LR = 1196.0$ on one degree of freedom. The fitted curve is convex, with an implied turning point at 25.1 years. Fig. 7 shows the fitted relationship against the observed listings.

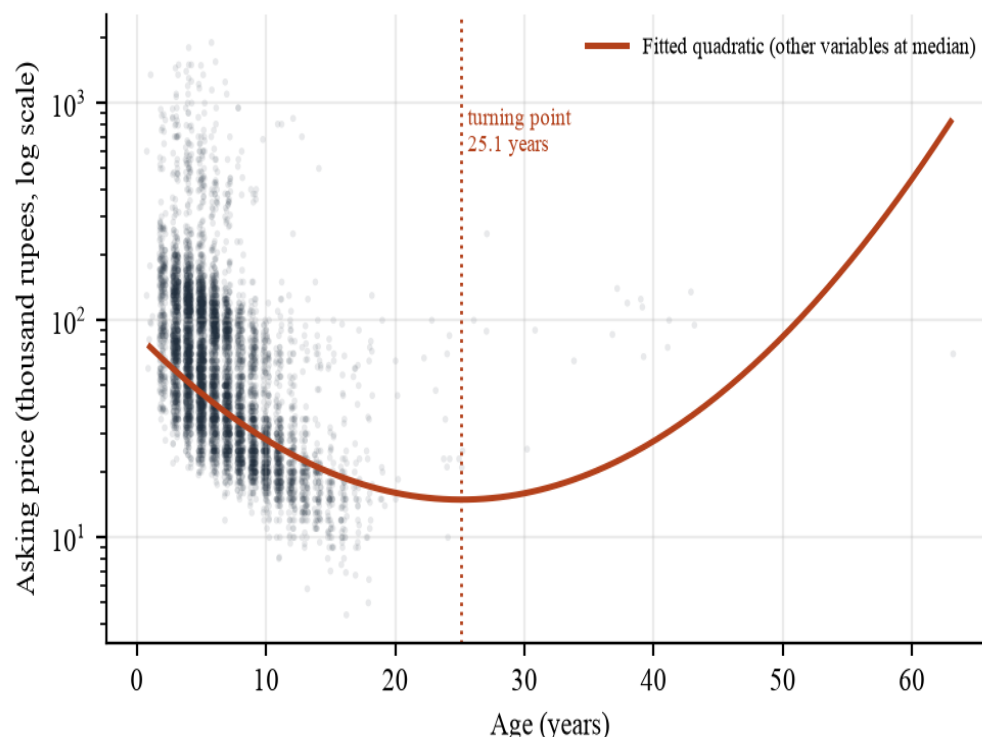


Fig. 7 Asking price against age with the fitted quadratic curve, drawn for a first-owner Bajaj with median distance and displacement.

Four further checks were run. As noted earlier, they were designed after this result was seen and are exploratory. First, refitting the quadratic model without any listing aged twenty-one or more leaves the curvature intact: the squared term remains positive at 0.00345 with $p < 0.001$ across 7,294 listings, and the implied turning point of 22.7 years now lies outside the range of the data. The curvature is therefore a property of the main body of the data, not an artefact of a handful of vintage machines. What the data show clearly is that depreciation slows down: a motorcycle loses value fastest in its early years and more slowly thereafter.

Second, a bootstrap with 2,000 resamples places the turning point at a median of 24.9 years with a 95 per cent confidence interval of 19.7 to 30.0 years, and the fitted curve was convex in all 2,000 resamples. Third, the age-price correlation is negative both within Royal Enfield, at -0.738 across 1,346 listings, and across all other brands, at -0.695 across 5,978, so the pattern is not produced by one manufacturer alone.

Fourth, and most important, the composition of the oldest group was examined directly. Of the twenty listings aged twenty-five years or more, thirteen are Royal Enfields, twelve of them Bullets. The median displacement in the band of twenty-one years and above is 300cc, against 150cc in the middle bands. Fig. 8 separates Royal Enfield from everything else.

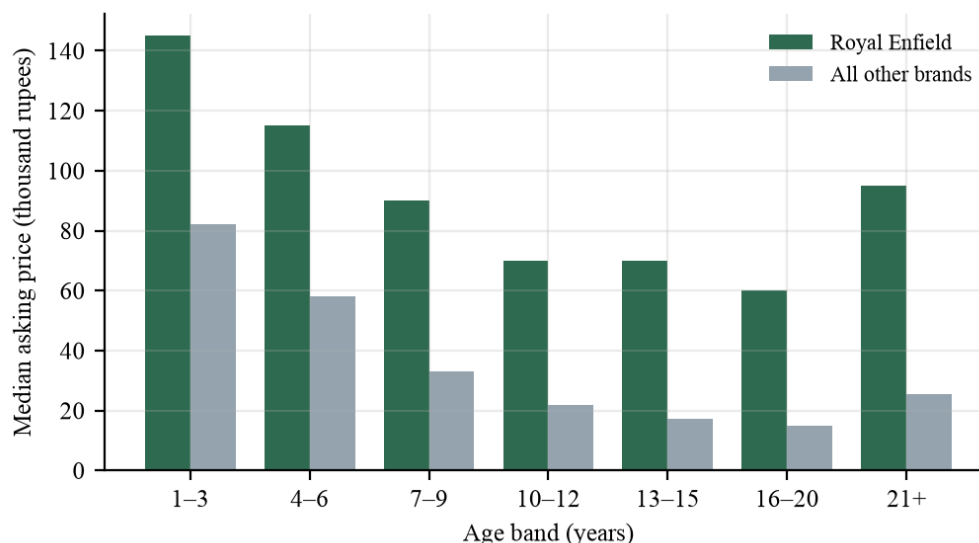


Fig. 8 Median asking price by age band, Royal Enfield compared with all other manufacturers.

The separation is clear. Royal Enfield machines are listed above other brands in every age band, and in the oldest band the medians are Rs 95,000 and Rs 25,500. The upturn in the pooled curve therefore appears to be largely a composition effect: the machines still listed after twenty-five years are mostly large-engined Royal Enfields, from a brand listed higher than others at every age. It is not evidence that an ordinary commuter motorcycle gains value if kept long enough. Regulation may also play a part. Das and colleagues record Delhi legislation to ban two-wheelers more than 15 years old [11], and rules of that kind would thin out the oldest part of the market in the places they apply.

The result for H7 is therefore split. The slowing of depreciation with age is robust. The upturn at about twenty-five years is suggestive: it rests on seventeen to twenty listings, is confounded with brand and displacement, and is reported as a lead for further work rather than as a finding.

H. The Yamaha RX100

The dataset was searched for the Yamaha RX100. Across 7,324 distinct listings, 651 of them Yamaha machines spanning 49 distinct Yamaha models, there is not a single RX100.

The absence is not explained by the dataset excluding old machines. It contains a sixty-three-year-old Royal Enfield Bullet, a forty-one-year-old Rajdoot GTX and two Yezdis aged thirty-nine and forty-one. Nor is it explained by an absence of old Yamahas: the sample includes two RX135 machines aged twenty-four years, one listed at Rs 85,000 with 300 kilometres recorded, and an RX-Z aged twenty-three. Close relatives of the RX100 are present. The RX100 itself is not.

Nor is the absence peculiar to this dataset. On 22 September 2026, eighteen motorcycle and used-vehicle search terms on Kaggle returned 243 datasets. The 175 of 80 MB or less were downloaded, 173 of them successfully, and searched for the RX100. Eight contained the name. Once specification tables, registration records and one false match on the model code KRX1000 are set aside, they hold twelve RX100 sale listings in all: ten Indian, three of which appear to be a single advertisement repeated, and two from Sri Lanka. The registration sample described in the methodology contains two RX100 records among its 202,856 two-wheeler records, both resales in Karnataka, one still carrying a Bharat Stage I emission classification. The specification entry, from the Indian Motorcycle Specifications Dataset published on Kaggle by VanshRajput-dev under a CC BY-NC-SA 4.0 licence, is the source of the production years and displacement given in the introduction.

No statistical inference is drawn from a count of zero. What can be said is descriptive. A machine with a strong following among enthusiasts does not appear in the online resale market that this dataset records, and appears only rarely in any public dataset searched. One explanation is that the surviving machines now change hands privately, between people who already know one another, and so leave no public record. That is a hypothesis, not a finding, and the data cannot test it.

VII. RESULTS

Table VIII collects the outcome of every hypothesis stated in advance.

TABLE VIII
SUMMARY OF HYPOTHESIS TESTS

H	Test	Statistic	p	Effect size	Decision
H1	Spearman, age	$\rho = -0.6028$	< 0.001	Strong	Supported
H2	Spearman, distance	$\rho = -0.5318$	< 0.001	Strong	Supported
H3	Spearman, displacement	$\rho = +0.8017$	< 0.001	Strong	Supported
H4	Mann–Whitney, ownership	$U = 2,487,989.5$	0.000022	$d = 0.143$; adjusted -4.1%	Supported; effect small
H5	Partial F , brand	$F = 127.96$	< 0.001	Partial $\eta^2 = 0.278$, large	Supported
H6	Stratified ρ comparison	No reversal found	–	–	Not rejected
H7	Quadratic age term	$LR = 1196.0$	< 0.001	$\Delta R^2 = 0.023$	Supported, weakly

For H1 to H5 and H7, ‘supported’ means the null hypothesis was rejected at the 0.05 level. For H6 the null hypothesis is the absence of a reversal, and it was not rejected. H4 is supported statistically, but the adjusted effect is about four per cent.

Stated as findings rather than decisions, the results are these. Engine displacement is the strongest correlate of asking price in this market, ahead of age. Age comes second, at about 7.1 per cent of asking price per year. Brand carries a large independent effect, with Royal Enfield listed 33.5 per cent above the reference manufacturer on otherwise comparable machines. Distance covered matters, but considerably less than age or displacement. Ownership history matters only a little: a second-owner machine is listed about four per cent below a comparable first-owner one. The age–price relationship does not reverse in any subgroup. Depreciation slows with age, and there is a suggestion, no more, that it reverses for the oldest surviving machines, most of which are Royal Enfields. The Yamaha RX100 does not appear in the data at all.

VIII. DISCUSSION

A. The Market Prices the Machine More Than Its History

The clearest pattern in the results is that attributes describing what a motorcycle is carry more weight than attributes describing what has happened to it. Displacement, brand and age account for most of the explained variation; distance covered carries about a quarter of the weight of age, and ownership history very little.

One possible explanation is verification. Ownership count and odometer reading are declared by the seller, and a buyer browsing listings has no easy way to check either, whereas make and displacement are hard to misstate. A market in which some claims cannot be checked may discount them. The data cannot test this, and the small ownership effect is equally consistent with ownership history simply mattering little to buyers.

B. Depreciation over a Long Service Life

An average decline of 7.1 per cent a year that slows as machines age means that a used motorcycle in this data keeps part of its asking price well into its second decade: the median listing aged sixteen to twenty years is priced at about a sixth of the median listing aged one to three. That pattern is compatible with Grubel’s argument that vehicles in lower-income economies lose value slowly because they are cheap to keep running [1], although a single-country study cannot test it. The age profile of the listings is also close to what Das and colleagues estimated for Delhi two decades earlier, where most two-wheelers were six to nine years old [11]; the median listing here is six years old.

C. What the Brand Premium Is Buying

Royal Enfield's 33.5 per cent premium survives controls for age, distance, displacement and ownership, so it is not a mechanical consequence of selling larger or newer machines. Some of it may reflect attributes the data cannot see, such as build quality or the availability of parts; some may be reputation. The evidence on this is thin. The one study of motorcycle purchase behaviour reviewed here found that perceived quality did not predict purchase intention, and concluded that buyers weigh other considerations it did not identify [12]. For a buyer, the practical point is that part of the asking price attaches to the brand rather than to the attributes listed.

D. Why the Absence of a Paradox Is Worth Reporting

Finding no Simpson's paradox is a negative result, but it is what allows the paper's central estimate to be read as general. Without the stratified check, the pooled depreciation rate might or might not describe any actual segment of the market. With it, the direction of the relationship is known to hold in every brand and displacement band examined. Table VII shows how easily a misleading conclusion could be drawn by comparing raw brand averages instead.

E. A Motorcycle Outside the Market

The RX100 result inverts the paper's method. Every other finding comes from counting listings; this one comes from their absence, and no statistical technique can extract more from a zero than the fact of it. What the absence does show is that a machine's standing among enthusiasts is no evidence that it is traded in the open market. Enthusiasm and liquidity are different things. Cirolia and colleagues describe motorcycle taxis in Kigali and Nairobi being drawn onto digital platforms [15]; if the surviving RX100s are indeed changing hands privately, they have moved the other way. Any quantitative study of the RX100 will have to reach those private exchanges, because the public data do not contain them.

IX. LIMITATIONS

The limitations below are substantial and are stated plainly, including those that weaken the paper's own conclusions.

- • These are asking prices, not transaction prices. Every figure in this paper describes what sellers hoped to receive, not what buyers paid. The gap between the two is unknown and could be large, and it may differ by brand or by age. This is the most serious limitation of the study and it affects every result.
- • The unit of observation is a listing, not a machine or a sale. A machine listed twice appears twice, and a machine that never sells counts the same as one that sells at once. Unsold, optimistically priced listings may therefore be over-represented.
- • The data come from one marketplace and were collected on or before 1 June 2021. Droom is one of several Indian platforms and its stock is not a random sample of the national market. Machines sold through other platforms, local classifieds or private networks are absent, which is also the explanation offered for the RX100's absence. That explanation is therefore consistent with the data but cannot be tested with it.
- • The source file was 77.6 per cent duplicated, and the duplication was concentrated rather than random. Removing it was necessary, but a fault this severe, left undocumented, raises questions about the scrape's coverage that the file alone cannot answer.
- • Important attributes are missing. Mechanical condition, service history, accident damage, modification status, colour and photographs all influence real prices, and none is observed. The model's high explained variance should not be read as evidence that these do not matter.
- • The vintage upturn rests on seventeen to twenty listings and is confounded with brand and displacement, and its robustness checks were designed after the result was seen. It should not be cited as an established collector premium.
- • Ownership history and odometer readings are declared by sellers. The small ownership effect may reflect unreliable reporting rather than indifference among buyers, and the study cannot distinguish the two.
- • The RX100 conclusion is an inference from absence. Zero listings are consistent with a machine that has left the online market, but also with a scrape that missed it. The wider search covered Kaggle datasets only; other platforms were not examined.
- • The registration sample used for context is described by its publisher only as 'RTO-style' records, and its original source is not documented.

X. RECOMMENDATIONS

- 1) • *For buyers.* On this evidence, asking prices are driven mainly by engine displacement, age and brand. A first-owner label is associated with a premium of about four per cent on otherwise comparable machines, far less than the raw difference between owner groups suggests, and distance covered deserves roughly a quarter of the weight given to age.
- 2) • *For sellers.* Expect asking prices for comparable machines to fall by about seven per cent a year, and more slowly as the machine ages. A brand premium exists and is large, but it is fixed when the machine is bought, not when it is sold.
- 3) • *For marketplace operators.* Ownership history carries little price signal in this data, possibly because buyers cannot verify it. Verified service and ownership records would add information that a self-declared field does not.
- 4) • *For researchers.* Do not compare raw group averages in this market: the brands differ so much in composition that such comparisons measure segment mix rather than value. Stratified checks for aggregation effects should be routine, and effect sizes should accompany every significance test; with samples of this size, significance alone says little.
- 5) • *For data publishers.* The dataset used here is distributed as 32,648 rows and contains 7,324 distinct listings. Scraped data should be checked for duplication, and the result disclosed, before publication, and checked again by anyone who uses it.
- 6) • *For future work on the RX100.* The public datasets searched here hold only a handful of RX100 records. Progress requires primary data from owners' clubs, restorers and specialist dealers. That is a different study, and a worthwhile one.

XI. CONCLUSION

This study set out to explain what determines the asking price of a used motorcycle in India, using 7,324 distinct listings recovered from a public file that was more than three-quarters duplicated. A hedonic model explains 84.7 per cent of the variation in log asking price. Engine displacement is the strongest factor, followed by age at about 7.1 per cent of asking price per year, then brand, with Royal Enfield listed 33.5 per cent above the reference manufacturer on comparable machines. Distance covered matters, but far less than age or displacement.

Two of the results are useful precisely because they are small or negative. Ownership history is statistically significant but modest: once age, distance and brand are accounted for, a second-owner machine is listed about four per cent below a first-owner one, and ownership has a partial η^2 of 0.002. The paper reports that plainly rather than letting a p -value of 0.000022 stand in for importance. A deliberate search for Simpson's paradox found none: the negative age-price relationship holds within every brand and every displacement band, so the headline depreciation figure describes the market rather than an accident of aggregation.

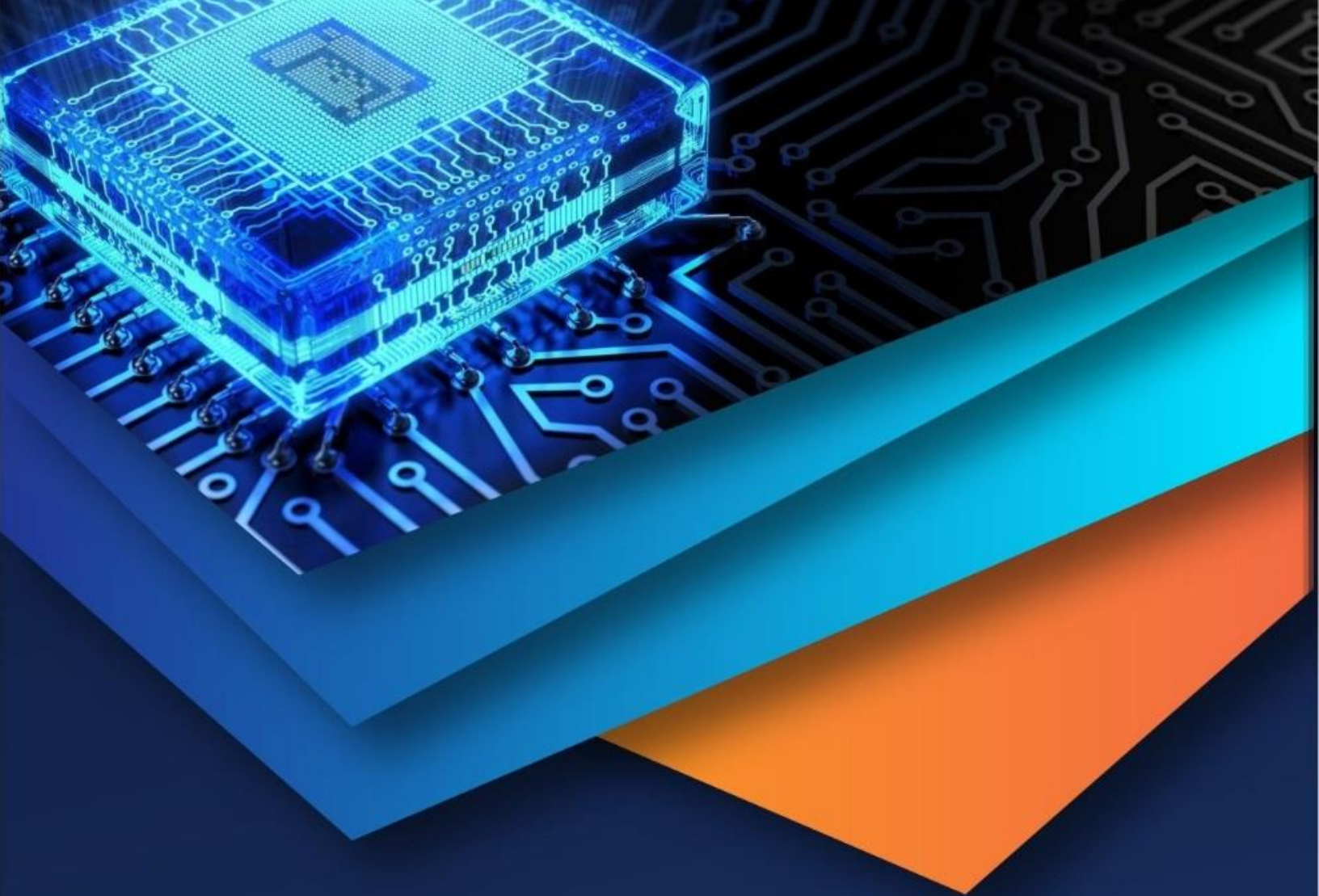
The evidence on whether old motorcycles recover value is mixed, and is reported that way. Depreciation clearly slows with age, and that result survives every check applied. The apparent upturn beyond twenty-five years is present in the fitted curve but rests on seventeen to twenty listings and is confounded with brand: the machines that remain listed at that age are mostly Royal Enfields.

The study began with the Yamaha RX100 and ends with it, having found none. In 7,324 listings, among 651 Yamahas of 49 distinct models, in a dataset that contains a sixty-three-year-old Bullet and a forty-one-year-old Rajdoot, the RX100 does not appear once. A motorcycle that enthusiasts still remember is missing from a marketplace that lists almost everything else. Whatever value it holds is being exchanged somewhere the data cannot see. That is a finding of a kind, even if it is not one that can be given a p -value.

REFERENCES

- [1] H. G. Grubel, "International trade in used cars and problems of economic development," *World Development*, vol. 8, no. 10, pp. 781–788, Oct. 1980.
- [2] Y. Nakamoto, "CO2 reduction potentials through the market expansion and lifetime extension of used cars," *Journal of Economic Structures*, vol. 6, no. 1, Art. no. 17, 2017.
- [3] M. Koç and N. Kostak, "What is happening to used car prices in Turkey? Evidence from hedonic price model (HPM)," *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, no. 46, pp. 225–240, 2021.
- [4] M. Francke and A. Van de Minne, "Modeling unobserved heterogeneity in hedonic price models," *Real Estate Economics*, vol. 49, no. 4, pp. 1315–1339, 2021.
- [5] E. Gegic, B. Isakovic, D. Keco, Z. Masetic, and J. Kevric, "Car price prediction using machine learning techniques," *TEM Journal*, vol. 8, no. 1, pp. 113–118, Feb. 2019.
- [6] P. Bharambe, B. Bagul, S. Dandekar, and P. Ingle, "Used car price prediction using different machine learning algorithms," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 4, pp. 773–778, Apr. 2022.
- [7] G. Kalpana, A. Kanaka Durga, T. A. Reddy, and G. Karuna, "Predicting the price of pre-owned cars using machine learning and data science," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 7, pp. 1468–1476, Jul. 2022.
- [8] K. Ranjith, B. K. Rishi Sagar, and U. Sharma, "Prediction of resale value of pre-owned luxury cars in the Indian market employing machine learning techniques," *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 7, pp. 1026–1037, Jul. 2024.
- [9] D. A. Budiono, K. S. Utomo, K. J. Wibowo, and M. J. Wiradinata, "Used car price prediction model: A machine learning approach," *International Journal of Computer and Information System*, vol. 5, no. 1, pp. 59–66, Mar. 2024.

- [10] A. Barlybayev, A. Sankibayev, Y. Kadyr, N. Amangeldy, and T. Sabyrov, "Predicting used-vehicle resale value in developing markets: Application of machine learning models to the Kazakhstan car market," *Ingénierie des Systèmes d'Information*, vol. 28, no. 5, pp. 1237–1246, Oct. 2023.
- [11] S. Das, R. Schmoyer, G. Harrison, and K. Hausker, "Prospects of inspection and maintenance of two-wheelers in India," *Journal of the Air & Waste Management Association*, vol. 51, no. 10, pp. 1391–1400, Oct. 2001.
- [12] M. R. Shaharudin, S. Wan Mansor, A. Abu Hassan, M. Wan Omar, and E. H. Harun, "The relationship between product quality and purchase intention: The case of Malaysia's national motorcycle/scooter manufacturer," *African Journal of Business Management*, vol. 5, no. 20, pp. 8163–8176, Sep. 2011.
- [13] C. Balijepalli, S. Shepherd, R. Crastes dit Sourd, M. Farda, T. Praesha, and H. A.-R. Lubis, "Preferences for electric motorcycle adoption in Bandung, Indonesia," *Urban, Planning and Transport Research*, vol. 11, no. 1, Art. no. 2238033, 2023.
- [14] S. Bakker, "Electric two-wheelers, sustainable mobility and the city," in *Sustainable Cities – Authenticity, Ambition and Dream*. IntechOpen, 2019, ch. 6, doi: 10.5772/intechopen.81460.
- [15] L. R. Cirolia, R. Sitas, A. Pollio, A. G. Sebarenzi, and P. K. Guma, "Silicon Savannahs and motorcycle taxis: A Southern perspective on the frontiers of platform urbanism," *Environment and Planning A: Economy and Space*, vol. 55, no. 8, pp. 1989–2008, 2023.
- [16] R. A. Kievit, W. E. Frankenhuis, L. J. Waldorp, and D. Borsboom, "Simpson's paradox in psychological science: A practical guide," *Frontiers in Psychology*, vol. 4, Art. no. 513, Aug. 2013.
- [17] Z. Armstrong and M. Wattenberg, "Visualizing statistical mix effects and Simpson's paradox," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2132–2141, Dec. 2014.
- [18] P. H. Walton and D. J. Walton, "Simpson's paradox in the interpretation of 'leaky pipeline' data," *International Journal for Transformative Research*, vol. 3, no. 2, pp. 1–7, 2016.
- [19] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," *Frontiers in Psychology*, vol. 4, Art. no. 863, Nov. 2013.
- [20] P. H. P. Hanel and D. M. A. Mehler, "Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication," *Public Understanding of Science*, vol. 28, no. 4, pp. 468–485, 2019.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)