

Indonesia_used_car

2025-06-12

by: Aldo Gadra P.S.

Introduction

Indonesia used car market have been growing fast in the last few years. In 2025, used car sold twice as much as new car. This new market shift is caused by the rise of new online used vehicle-oriented marketplace and economic downturn.

Goal: - Identify variables that significantly contribute to the used car's and its relevant contribution. - Furthermore, to make machine learning model to predict used car price.

Data & Package Preparation

```
## Data reading
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.3.3
```

```
## Data Handling
library(tidyverse)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v purrr      1.0.1
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(stringi)
```

```
## Warning: package 'stringi' was built under R version 4.3.3
```

```
library(dplyr)
```

About the dataset: - The dataset is uploaded by Indra in Kaggle. (<https://www.kaggle.com/datasets/indraputra21/used-car-listings-in-indonesia/data>). - The dataset made up of various listings scraped from <https://www.carsome.id/>.

Initial data checking

```
glimpse(df)
```

```
## Rows: 609
## Columns: 20
## $ car.name      <chr> "AYLA X 1.2", "AGYA TRD SPORTIVO 1.0", "X-TR~
## $ brand         <chr> "Daihatsu", "Toyota", "Nissan", "Toyota", "T~
## $ year          <int> 2018, 2015, 2015, 2020, 2019, 2021, 2022, 20~
## $ mileage..km.  <dbl> 10.508, 112.888, 118.429, 15.945, 30.404, 17~
## $ location      <chr> "Jakarta Utara", "Bogor", "Surabaya", "Tange~
## $ transmission  <chr> "Manual", "Manual", "Automatic", "Automatic"~
## $ plate.type    <chr> "even plate", "even plate", "odd plate", "od~
## $ rear.camera   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ sun.roof      <int> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,~
## $ auto.retract.mirror <int> 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0,~
## $ electric.parking.brake <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,~
## $ map.navigator <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ vehicle.stability.control <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ keyless.push.start <int> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ sports.mode    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ X360.camera.view <int> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ power.sliding.door <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ auto.cruise.control <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ price..Rp.     <int> 101000000, 82000000, 169000000, 218000000, 1~
## $ instalment..Rp.Monthly. <int> 2060000, 1670000, 3440000, 4440000, 2380000,~
```

The dataframe has 20 variables and 609 observations.

```
head(df, 5)
```

```
##           car.name    brand year mileage..km.      location
## 1          AYLA X 1.2 Daihatsu 2018      10.508 Jakarta Utara
## 2 AGYA TRD SPORTIVO 1.0  Toyota 2015      112.888          Bogor
## 3           X-TRAIL 2.5  Nissan 2015      118.429          Surabaya
## 4          YARIS S TRD 1.5 Toyota 2020       15.945 Tangerang Selatan
## 5           AGYA G 1.2  Toyota 2019       30.404 Jakarta Barat
## transmission plate.type rear.camera sun.roof auto.retract.mirror
## 1      Manual even plate          0          0          0
## 2      Manual even plate          0          0          0
## 3   Automatic odd plate          0          0          0
## 4   Automatic odd plate          0          0          0
## 5      Manual odd plate          0          0          0
## electric.parking.brake map.navigator vehicle.stability.control
```

```
## 1      0      0      0
## 2      0      0      0
## 3      0      0      0
## 4      0      0      0
## 5      0      0      0
##  keyless.push.start sports.mode X360.camera.view power.sliding.door
## 1      0      0      0      0
## 2      0      0      0      0
## 3      1      0      1      0
## 4      0      0      0      0
## 5      0      0      0      0
##  auto.cruise.control price..Rp. instalment..Rp.Monthly.
## 1      0 101000000      2060000
## 2      0  82000000      1670000
## 3      0 169000000      3440000
## 4      0 218000000      4440000
## 5      0 117000000      2380000
```

I need to rename some of the columns to ease further analysis process.

```
df <- df %>%
  rename(
    price = price..Rp.,
    instalment_month = instalment..Rp.Monthly.,
    mileage_km = mileage..km.
  )
```

Data Cleaning

```
all_cols <- colnames(df)
```

Data format checking & grouping

Format checking

Before data analysis, I must check for variable entry for correct formatting and additional info. But first, I need to check the columns' uniqueness level to identify for observation identifier (non continuous column who have high degree of uniqueness).

```
sapply(df, function(x){
  distinct_count <- n_distinct(x)
  row_count <- nrow(df)
  result <- (distinct_count / row_count) * 100
  result <- round(result, 5)
  return(result)
})
```

```
##      car.name      brand      year
##      28.24302      2.13465      2.29885
```

```
##          mileage_km          location          transmission
##          100.00000          2.13465          0.32841
##          plate.type          rear.camera          sun.roof
##          0.32841          0.32841          0.32841
##          auto.retract.mirror electric.parking.brake map.navigator
##          0.32841          0.32841          0.32841
## vehicle.stability.control keyless.push.start sports.mode
##          0.32841          0.32841          0.32841
##          X360.camera.view power.sliding.door auto.cruise.control
##          0.32841          0.32841          0.32841
##          price          instalment_month
##          33.49754          33.49754
```

There is no identifier column found (as “mileage_km” is a continuous variable despite it high level of uniqueness).

Checking columns entry

```
sample(df)
```

```
##          location          price auto.retract.mirror transmission rear.camera
## 1      Jakarta Utara 101000000          0      Manual          0
## 2          Bogor  82000000          0      Manual          0
## 3      Surabaya 169000000          0      Automatic          0
## 4  Tangerang Selatan 218000000          0      Automatic          0
## 5      Jakarta Barat 117000000          0      Manual          0
## 6  Tangerang Selatan 180000000          1      Manual          0
## 7      Surabaya 211000000          1      Manual          0
## 8          Bekasi 190000000          1      Automatic          0
## 9          Bandung 140000000          0      Manual          0
## 10         Malang 248000000          1      Automatic          0
## 11         Bandung 117000000          0      Manual          0
## 12          Depok 114000000          0      Manual          0
## 13          Bekasi 110000000          0      Manual          0
## 14  Jakarta Selatan 115000000          0      Automatic          0
## 15  Jakarta Utara 118000000          0      Manual          0
## 16         Malang 198000000          1      Automatic          0
## 17          Depok 141000000          1      Manual          0
## 18          Bekasi 117000000          0      Manual          0
## 19  Jakarta Timur 155000000          0      Manual          0
## 20          Bogor 108000000          0      Manual          0
## 21  Jakarta Timur 105500000          0      Manual          0
## 22          Bogor 111000000          0      Manual          0
## 23          Bogor 190000000          0      Manual          0
## 24  Tangerang Selatan 114000000          0      Manual          0
## 25  Jakarta Utara 340000000          0      Automatic          0
## 26         Malang 101500000          0      Manual          0
## 27  Jakarta Timur 202000000          1      Automatic          0
## 28  Tangerang Selatan 197000000          1      Automatic          0
## 29          Bandung 138000000          1      Automatic          0
## 30          Bandung 135000000          0      Manual          0
## 31          Bekasi 193000000          1      Manual          0
```

## 32	Surabaya	127000000	0	Manual	0
## 33	Bogor	208000000	1	Automatic	0
## 34	Jakarta Utara	216000000	1	Automatic	0
## 35	Bandung	204000000	1	Automatic	0
## 36	Tangerang Selatan	89000000	0	Manual	0
## 37	Depok	123000000	0	Manual	0
## 38	Bekasi	107000000	0	Manual	0
## 39	Bekasi	133000000	1	Manual	0
## 40	Jakarta Timur	248000000	0	Automatic	0
## 41	Jakarta Timur	228000000	1	Automatic	1
## 42	Jakarta Utara	157000000	0	Automatic	0
## 43	Bandung	142000000	0	Manual	0
## 44	Jakarta Barat	143000000	0	Manual	0
## 45	Bogor	195000000	1	Automatic	0
## 46	Bekasi	254000000	1	Automatic	0
## 47	Bandung	199000000	0	Automatic	1
## 48	Bandung	111500000	0	Manual	1
## 49	Bandung	203000000	0	Automatic	1
## 50	Jakarta Pusat	117000000	0	Automatic	0
## 51	Bogor	193000000	1	Manual	0
## 52	Bogor	213000000	1	Automatic	0
## 53	Jakarta Selatan	335000000	1	Automatic	0
## 54	Tangerang Selatan	255000000	1	Automatic	0
## 55	Jakarta Barat	242000000	0	Manual	0
## 56	Bogor	155000000	0	Automatic	1
## 57	Tangerang Selatan	193000000	0	Automatic	0
## 58	Bekasi	122000000	0	Manual	0
## 59	Tangerang Selatan	233000000	0	Automatic	0
## 60	Bandung	110000000	0	Manual	0
## 61	Bogor	143000000	1	Manual	0
## 62	Bogor	123000000	0	Manual	0
## 63	Tangerang Selatan	127000000	0	Manual	0
## 64	Bekasi	106000000	0	Manual	0
## 65	Depok	207000000	0	Automatic	1
## 66	Bogor	147000000	0	Manual	0
## 67	Tangerang Selatan	257000000	1	Automatic	0
## 68	Jakarta Barat	139000000	0	Automatic	0
## 69	Bogor	180000000	1	Automatic	0
## 70	Jakarta Selatan	365000000	1	Automatic	0
## 71	Tangerang Selatan	426000000	1	Automatic	0
## 72	Jakarta Utara	148000000	0	Automatic	1
## 73	Tangerang Selatan	163000000	0	Manual	0
## 74	Bogor	99000000	0	Manual	0
## 75	Jakarta Utara	128000000	0	Automatic	0
## 76	Bekasi	148000000	1	Manual	0
## 77	Jakarta Timur	231000000	1	Automatic	0
## 78	Surabaya	168000000	1	Automatic	0
## 79	Tangerang Selatan	273000000	0	Automatic	0
## 80	Jakarta Barat	104000000	0	Manual	0
## 81	Bekasi	100000000	0	Manual	0
## 82	Bekasi	204000000	1	Automatic	0
## 83	Bogor	117000000	0	Manual	0
## 84	Jakarta Utara	316000000	1	Automatic	0
## 85	Bandung	87000000	0	Manual	1

## 86	Jakarta Timur	176000000	1	Automatic	0
## 87	Bogor	160000000	1	Manual	0
## 88	Tangerang Selatan	114000000	0	Manual	0
## 89	Bogor	209000000	1	Automatic	0
## 90	Jakarta Barat	210000000	1	Automatic	0
## 91	Jakarta Utara	100000000	0	Manual	0
## 92	Tangerang Selatan	136000000	1	Automatic	1
## 93	Jakarta Barat	203000000	1	Automatic	0
## 94	Jakarta Utara	148000000	0	Manual	0
## 95	Bogor	214000000	1	Automatic	0
## 96	Malang	219000000	1	Manual	0
## 97	Bekasi	142000000	0	Manual	0
## 98	Bekasi	261000000	1	Automatic	0
## 99	Bogor	198000000	1	Automatic	0
## 100	Jakarta Utara	96000000	0	Manual	0
## 101	Jakarta Utara	360000000	1	Automatic	0
## 102	Bogor	116000000	0	Manual	0
## 103	Surabaya	135000000	0	Automatic	0
## 104	Tangerang Selatan	224000000	1	Automatic	0
## 105	Bogor	136000000	0	Manual	0
## 106	Surabaya	160000000	1	Manual	0
## 107	Tangerang Selatan	180000000	0	Automatic	1
## 108	Jakarta Utara	206000000	1	Automatic	0
## 109	Bogor	203000000	1	Automatic	0
## 110	Tangerang Selatan	232000000	1	Automatic	0
## 111	Bogor	365000000	1	Automatic	0
## 112	Jakarta Timur	390000000	1	Automatic	0
## 113	Bogor	246000000	1	Manual	0
## 114	Bekasi	201000000	1	Automatic	0
## 115	Jakarta Utara	210000000	1	Automatic	0
## 116	Jakarta Utara	100000000	0	Manual	0
## 117	Bekasi	156000000	0	Automatic	0
## 118	Bandung	172000000	0	Automatic	1
## 119	Jakarta Barat	191000000	0	Automatic	0
## 120	Bekasi	212000000	1	Automatic	0
## 121	Bogor	145000000	0	Automatic	1
## 122	Bekasi	201000000	1	Manual	0
## 123	Jakarta Utara	232000000	1	Automatic	0
## 124	Bandung	163000000	0	Manual	0
## 125	Jakarta Selatan	239000000	1	Automatic	1
## 126	Bogor	248000000	1	Automatic	0
## 127	Bogor	154000000	0	Automatic	0
## 128	Bandung	266000000	1	Automatic	0
## 129	Jakarta Barat	115000000	0	Manual	0
## 130	Tangerang Selatan	336000000	1	Automatic	0
## 131	Depok	218000000	1	Automatic	0
## 132	Bogor	216000000	1	Automatic	0
## 133	Bekasi	129000000	0	Manual	0
## 134	Bogor	336000000	1	Automatic	0
## 135	Tangerang Selatan	142000000	0	Automatic	0
## 136	Jakarta Timur	117000000	0	Manual	0
## 137	Bogor	232000000	1	Automatic	0
## 138	Bogor	240000000	1	Automatic	0
## 139	Tangerang Selatan	241000000	0	Automatic	0

## 140	Jakarta Timur	115000000	0	Manual	0
## 141	Jakarta Timur	95000000	0	Manual	0
## 142	Jakarta Pusat	93000000	0	Manual	0
## 143	Bekasi	98000000	0	Manual	0
## 144	Bandung	124000000	0	Manual	0
## 145	Depok	207000000	1	Automatic	0
## 146	Jakarta Timur	178000000	0	Automatic	0
## 147	Tangerang Selatan	198000000	1	Manual	1
## 148	Jakarta Timur	211000000	1	Automatic	0
## 149	Jakarta Timur	119000000	0	Manual	0
## 150	Tangerang Selatan	126000000	0	Automatic	0
## 151	Jakarta Selatan	152000000	1	Manual	0
## 152	Bekasi	105000000	0	Manual	0
## 153	Bogor	113000000	0	Manual	0
## 154	Bogor	355000000	1	Automatic	0
## 155	Bekasi	159000000	0	Automatic	0
## 156	Tangerang Selatan	154000000	1	Automatic	0
## 157	Bekasi	136000000	0	Automatic	0
## 158	Jakarta Barat	211000000	1	Automatic	0
## 159	Tangerang Selatan	194000000	1	Manual	0
## 160	Jakarta Timur	192000000	0	Manual	1
## 161	Jakarta Utara	210000000	1	Automatic	0
## 162	Tangerang Selatan	211000000	1	Automatic	0
## 163	Bogor	255000000	0	Automatic	0
## 164	Tangerang Selatan	184000000	1	Automatic	0
## 165	Tangerang Selatan	219000000	1	Automatic	0
## 166	Malang	248000000	0	Automatic	1
## 167	Bandung	120000000	0	Manual	0
## 168	Tangerang Selatan	121000000	0	Manual	0
## 169	Tangerang Selatan	190000000	0	Automatic	0
## 170	Depok	236000000	1	Automatic	0
## 171	Depok	213000000	1	Manual	1
## 172	Tangerang Selatan	183000000	1	Automatic	1
## 173	Tangerang Selatan	151000000	1	Automatic	0
## 174	Bogor	149000000	1	Manual	0
## 175	Jakarta Timur	218000000	1	Automatic	1
## 176	Depok	213000000	1	Automatic	0
## 177	Jakarta Barat	199000000	1	Automatic	0
## 178	Bandung	127000000	1	Manual	0
## 179	Depok	169000000	1	Automatic	0
## 180	Depok	211000000	1	Automatic	0
## 181	Tangerang Selatan	196000000	1	Automatic	0
## 182	Jakarta Pusat	129000000	1	Automatic	0
## 183	Bogor	111000000	0	Automatic	1
## 184	Surabaya	102000000	0	Manual	0
## 185	Tangerang Selatan	174000000	1	Automatic	0
## 186	Tangerang Selatan	185000000	0	Manual	1
## 187	Bekasi	256000000	1	Automatic	0
## 188	Jakarta Barat	258000000	1	Automatic	0
## 189	Jakarta Selatan	210000000	1	Automatic	0
## 190	Jakarta Utara	164000000	1	Automatic	0
## 191	Tangerang Selatan	205000000	1	Automatic	0
## 192	Bogor	261000000	0	Automatic	1
## 193	Bandung	120000000	0	Automatic	0

## 194	Bandung	151000000	1	Automatic	0
## 195	Bandung	162000000	0	Automatic	0
## 196	Jakarta Timur	258000000	1	Automatic	0
## 197	Bekasi	262000000	1	Automatic	0
## 198	Bandung	261000000	1	Automatic	0
## 199	Bekasi	162000000	1	Automatic	0
## 200	Jakarta Timur	166000000	1	Automatic	0
## 201	Jakarta Selatan	136000000	1	Manual	0
## 202	Jakarta Barat	212000000	1	Automatic	1
## 203	Tangerang Selatan	213000000	1	Automatic	1
## 204	Jakarta Pusat	199000000	1	Automatic	1
## 205	Bogor	221000000	1	Automatic	0
## 206	Jakarta Utara	219000000	1	Automatic	0
## 207	Malang	203000000	1	Automatic	0
## 208	Bogor	106000000	0	Automatic	1
## 209	Bekasi	182000000	0	Automatic	0
## 210	Bandung	129000000	0	Manual	0
## 211	Bandung	100000000	0	Automatic	0
## 212	Bandung	98000000	0	Manual	0
## 213	Bogor	224000000	1	Automatic	0
## 214	Bekasi	139000000	0	Automatic	0
## 215	Bekasi	101000000	0	Manual	0
## 216	Bandung	118000000	0	Automatic	0
## 217	Bandung	101000000	0	Manual	0
## 218	Bandung	161000000	0	Automatic	0
## 219	Bandung	111000000	0	Automatic	0
## 220	Bogor	77000000	0	Manual	0
## 221	Bandung	112000000	0	Automatic	0
## 222	Tangerang Selatan	189000000	1	Manual	0
## 223	Jakarta Selatan	102000000	0	Manual	0
## 224	Bandung	216000000	1	Automatic	0
## 225	Bogor	314000000	1	Automatic	0
## 226	Bogor	157000000	0	Manual	0
## 227	Tangerang Selatan	139000000	0	Manual	0
## 228	Bekasi	157000000	0	Automatic	1
## 229	Bekasi	201000000	1	Automatic	1
## 230	Jakarta Timur	209000000	1	Automatic	0
## 231	Jakarta Barat	323000000	1	Automatic	0
## 232	Jakarta Pusat	131000000	0	Automatic	0
## 233	Jakarta Timur	428000000	0	Automatic	1
## 234	Bogor	278000000	1	Automatic	0
## 235	Depok	186000000	1	Manual	0
## 236	Bandung	145000000	1	Automatic	0
## 237	Tangerang Selatan	221000000	1	Automatic	0
## 238	Tangerang Selatan	165000000	1	Automatic	0
## 239	Jakarta Barat	262000000	1	Automatic	0
## 240	Bekasi	255000000	1	Automatic	0
## 241	Jakarta Utara	213000000	1	Automatic	1
## 242	Tangerang Selatan	220000000	1	Automatic	0
## 243	Jakarta Pusat	125000000	0	Automatic	1
## 244	Bandung	319000000	0	Automatic	0
## 245	Bogor	194000000	0	Automatic	0
## 246	Jakarta Timur	224000000	1	Automatic	0
## 247	Tangerang Selatan	255000000	1	Automatic	1

## 248	Bogor	154000000	1	Manual	0
## 249	Tangerang Selatan	155000000	1	Automatic	0
## 250	Jakarta Barat	181000000	1	Automatic	0
## 251	Tangerang Selatan	143000000	1	Automatic	0
## 252	Jakarta Selatan	156000000	0	Automatic	0
## 253	Bogor	169000000	1	Automatic	1
## 254	Tangerang Selatan	158000000	1	Automatic	0
## 255	Tangerang Selatan	146000000	0	Automatic	1
## 256	Bekasi	417000000	0	Automatic	0
## 257	Bogor	212000000	1	Automatic	0
## 258	Jakarta Timur	199000000	0	Automatic	0
## 259	Jakarta Barat	137000000	0	Automatic	0
## 260	Jakarta Timur	156000000	0	Automatic	0
## 261	Jakarta Barat	106000000	0	Manual	0
## 262	Jakarta Barat	424000000	1	Automatic	0
## 263	Bandung	110000000	0	Automatic	0
## 264	Jakarta Utara	193000000	1	Automatic	0
## 265	Bekasi	162000000	0	Automatic	0
## 266	Jakarta Utara	209000000	1	Automatic	0
## 267	Bekasi	203000000	1	Automatic	0
## 268	Jakarta Selatan	187000000	1	Manual	0
## 269	Bogor	181000000	1	Automatic	0
## 270	Depok	212000000	1	Automatic	0
## 271	Tangerang Selatan	132000000	0	Automatic	0
## 272	Jakarta Barat	178000000	1	Automatic	0
## 273	Bandung	179000000	0	Automatic	1
## 274	Jakarta Utara	163000000	0	Automatic	0
## 275	Jakarta Timur	373000000	1	Automatic	0
## 276	Jakarta Barat	186000000	1	Automatic	0
## 277	Jakarta Utara	149000000	0	Automatic	0
## 278	Bekasi	324000000	1	Automatic	1
## 279	Tangerang Selatan	119000000	1	Manual	1
## 280	Tangerang Selatan	191000000	1	Manual	1
## 281	Depok	317000000	1	Automatic	0
## 282	Bekasi	228000000	1	Automatic	0
## 283	Jakarta Utara	127000000	0	Automatic	0
## 284	Jakarta Barat	198000000	1	Automatic	0
## 285	Bogor	183000000	1	Automatic	1
## 286	Jakarta Timur	251000000	1	Automatic	0
## 287	Tangerang Selatan	174000000	1	Automatic	0
## 288	Bogor	185000000	1	Automatic	0
## 289	Jakarta Pusat	162000000	0	Automatic	0
## 290	Jakarta Selatan	210000000	1	Automatic	0
## 291	Tangerang Selatan	270000000	1	Automatic	0
## 292	Jakarta Barat	282000000	1	Automatic	0
## 293	Bogor	120000000	1	Manual	0
## 294	Bogor	210000000	1	Automatic	0
## 295	Bogor	158000000	1	Automatic	0
## 296	Tangerang Selatan	223000000	1	Automatic	0
## 297	Bogor	116000000	0	Manual	0
## 298	Tangerang Selatan	278000000	1	Automatic	0
## 299	Bogor	214000000	1	Automatic	0
## 300	Bandung	121000000	0	Automatic	1
## 301	Jakarta Utara	138000000	0	Automatic	0

## 302	Bekasi	197000000	1	Automatic	0
## 303	Bekasi	197000000	1	Automatic	0
## 304	Tangerang Selatan	156000000	0	Automatic	1
## 305	Tangerang Selatan	179000000	1	Automatic	0
## 306	Tangerang Selatan	156000000	0	Automatic	0
## 307	Jakarta Timur	105000000	0	Manual	1
## 308	Tangerang Selatan	187000000	0	Automatic	0
## 309	Jakarta Timur	141000000	0	Automatic	0
## 310	Bogor	158000000	0	Automatic	0
## 311	Jakarta Timur	203000000	1	Automatic	0
## 312	Tangerang Selatan	138000000	0	Automatic	1
## 313	Depok	168000000	0	Automatic	1
## 314	Jakarta Timur	212000000	1	Automatic	0
## 315	Bogor	124000000	0	Automatic	0
## 316	Bogor	126000000	0	Automatic	0
## 317	Jakarta Timur	212000000	1	Automatic	0
## 318	Bogor	124000000	0	Automatic	0
## 319	Jakarta Timur	542000000	1	Automatic	0
## 320	Jakarta Pusat	315000000	1	Automatic	1
## 321	Bekasi	205000000	1	Automatic	0
## 322	Bekasi	177000000	1	Manual	0
## 323	Depok	319000000	0	Automatic	1
## 324	Jakarta Barat	132000000	0	Automatic	0
## 325	Bogor	145000000	0	Automatic	0
## 326	Jakarta Timur	180000000	0	Automatic	0
## 327	Tangerang Selatan	129000000	1	Automatic	0
## 328	Jakarta Utara	222000000	1	Automatic	0
## 329	Jakarta Timur	204000000	0	Automatic	1
## 330	Bogor	135000000	0	Automatic	0
## 331	Jakarta Utara	146000000	0	Automatic	0
## 332	Bekasi	203000000	1	Automatic	0
## 333	Bekasi	214000000	1	Automatic	0
## 334	Jakarta Selatan	215000000	1	Automatic	0
## 335	Bogor	224000000	1	Automatic	0
## 336	Tangerang Selatan	223000000	1	Automatic	0
## 337	Tangerang Selatan	285000000	1	Automatic	0
## 338	Bekasi	161000000	1	Automatic	0
## 339	Tangerang Selatan	277000000	1	Automatic	1
## 340	Bogor	180000000	0	Automatic	0
## 341	Bekasi	198000000	1	Automatic	0
## 342	Jakarta Pusat	206000000	1	Automatic	0
## 343	Bogor	211000000	1	Automatic	0
## 344	Jakarta Pusat	132000000	0	Automatic	0
## 345	Jakarta Timur	326000000	1	Automatic	0
## 346	Bogor	213000000	1	Automatic	1
## 347	Jakarta Timur	124000000	0	Manual	0
## 348	Tangerang Selatan	96000000	0	Automatic	0
## 349	Jakarta Barat	141000000	1	Automatic	1
## 350	Tangerang Selatan	380000000	0	Automatic	1
## 351	Tangerang Selatan	180000000	1	Automatic	0
## 352	Bogor	188000000	0	Automatic	1
## 353	Tangerang Selatan	211000000	1	Automatic	0
## 354	Bogor	154000000	1	Automatic	1
## 355	Tangerang Selatan	170000000	0	Automatic	0

## 356	Bogor	213000000	1	Automatic	0
## 357	Jakarta Utara	166000000	0	Automatic	0
## 358	Jakarta Utara	429000000	1	Automatic	0
## 359	Bekasi	317000000	0	Automatic	1
## 360	Jakarta Timur	217000000	1	Automatic	0
## 361	Bogor	152000000	1	Automatic	0
## 362	Bekasi	139000000	0	Automatic	0
## 363	Bekasi	201000000	1	Automatic	0
## 364	Depok	102000000	0	Manual	0
## 365	Bekasi	155000000	0	Automatic	0
## 366	Jakarta Selatan	144000000	0	Automatic	1
## 367	Depok	211000000	1	Automatic	0
## 368	Bekasi	213000000	1	Automatic	0
## 369	Depok	148000000	1	Automatic	0
## 370	Bogor	152000000	0	Manual	0
## 371	Jakarta Barat	172000000	0	Automatic	0
## 372	Bekasi	242000000	0	Automatic	0
## 373	Jakarta Barat	390000000	1	Automatic	0
## 374	Jakarta Selatan	261000000	0	Automatic	0
## 375	Bogor	131000000	0	Automatic	0
## 376	Tangerang Selatan	121000000	0	Automatic	0
## 377	Bandung	164000000	0	Automatic	0
## 378	Jakarta Utara	200000000	0	Automatic	1
## 379	Bekasi	201000000	1	Automatic	0
## 380	Bogor	172000000	0	Automatic	0
## 381	Depok	223000000	1	Automatic	0
## 382	Depok	130000000	0	Automatic	0
## 383	Bekasi	190000000	1	Automatic	0
## 384	Bogor	263000000	1	Automatic	1
## 385	Bandung	213000000	1	Automatic	0
## 386	Depok	160000000	0	Automatic	1
## 387	Bogor	129000000	1	Manual	0
## 388	Unknown	159000000	0	Automatic	0
## 389	Unknown	121000000	0	Automatic	0
## 390	Unknown	209000000	0	Automatic	0
## 391	Unknown	132000000	0	Automatic	0
## 392	Unknown	224000000	0	Automatic	0
## 393	Unknown	200000000	0	Automatic	0
## 394	Unknown	177000000	0	Automatic	0
## 395	Unknown	209000000	0	Automatic	0
## 396	Unknown	105000000	0	Automatic	0
## 397	Unknown	144000000	0	Automatic	0
## 398	Unknown	141000000	0	Automatic	0
## 399	Unknown	209000000	0	Automatic	0
## 400	Unknown	235000000	0	Automatic	0
## 401	Unknown	137000000	0	Automatic	0
## 402	Unknown	135000000	0	Automatic	0
## 403	Unknown	148000000	0	Automatic	0
## 404	Unknown	156000000	0	Automatic	0
## 405	Unknown	109000000	0	Manual	0
## 406	Unknown	258000000	0	Automatic	0
## 407	Unknown	443000000	0	Automatic	0
## 408	Unknown	141000000	0	Manual	0
## 409	Unknown	161000000	0	Automatic	0

## 410	Unknown	160000000	0	Automatic	0
## 411	Unknown	159000000	0	Automatic	0
## 412	Unknown	185000000	0	Automatic	0
## 413	Unknown	175000000	0	Automatic	0
## 414	Unknown	197000000	0	Manual	0
## 415	Unknown	212000000	0	Automatic	0
## 416	Unknown	212000000	0	Automatic	0
## 417	Unknown	212000000	0	Automatic	0
## 418	Unknown	212000000	0	Automatic	0
## 419	Unknown	122000000	0	Automatic	0
## 420	Unknown	161000000	0	Automatic	0
## 421	Unknown	181000000	0	Automatic	0
## 422	Unknown	182000000	0	Automatic	0
## 423	Unknown	256000000	0	Manual	0
## 424	Unknown	124000000	0	Automatic	0
## 425	Unknown	171000000	0	Automatic	0
## 426	Unknown	339000000	0	Automatic	0
## 427	Unknown	210000000	0	Automatic	0
## 428	Unknown	212000000	0	Automatic	0
## 429	Unknown	174000000	0	Automatic	0
## 430	Unknown	186000000	0	Automatic	0
## 431	Unknown	221000000	0	Automatic	0
## 432	Unknown	199000000	0	Automatic	0
## 433	Unknown	109000000	0	Manual	0
## 434	Unknown	120000000	0	Manual	0
## 435	Unknown	210000000	0	Automatic	0
## 436	Unknown	178000000	0	Manual	0
## 437	Unknown	281000000	0	Automatic	0
## 438	Unknown	78000000	0	Manual	0
## 439	Unknown	187000000	0	Automatic	0
## 440	Tangerang Selatan	210000000	0	Automatic	0
## 441	Jakarta Timur	179000000	0	Automatic	1
## 442	Tangerang Selatan	202000000	1	Automatic	0
## 443	Bogor	151000000	0	Automatic	0
## 444	Jakarta Selatan	195000000	0	Automatic	0
## 445	Jakarta Pusat	187000000	0	Automatic	0
## 446	Tangerang Selatan	174000000	1	Automatic	0
## 447	Jakarta Pusat	236000000	0	Manual	0
## 448	Jakarta Timur	206000000	1	Automatic	0
## 449	Tangerang Selatan	134000000	1	Manual	0
## 450	Jakarta Utara	173000000	1	Automatic	0
## 451	Jakarta Timur	198000000	1	Manual	0
## 452	Jakarta Timur	258000000	1	Automatic	0
## 453	Bekasi	195000000	1	Manual	1
## 454	Jakarta Utara	136000000	0	Automatic	0
## 455	Surabaya	129000000	0	Automatic	0
## 456	Bekasi	144000000	0	Automatic	1
## 457	Bekasi	128000000	0	Automatic	0
## 458	Bandung	92000000	0	Manual	0
## 459	Jakarta Utara	150000000	1	Manual	0
## 460	Tangerang Selatan	116000000	0	Manual	0
## 461	Jakarta Utara	122000000	0	Manual	0
## 462	Bandung	142000000	1	Automatic	0
## 463	Tangerang Selatan	282000000	1	Automatic	0

## 464	Bogor	370000000	1	Automatic	0
## 465	Jakarta Pusat	282000000	1	Automatic	1
## 466	Jakarta Timur	197000000	1	Automatic	0
## 467	Jakarta Utara	126000000	0	Manual	0
## 468	Jakarta Barat	192000000	1	Automatic	0
## 469	Jakarta Barat	92000000	0	Manual	0
## 470	Malang	160000000	1	Automatic	1
## 471	Tangerang Selatan	108000000	0	Manual	0
## 472	Jakarta Timur	141000000	1	Manual	1
## 473	Jakarta Selatan	131000000	0	Automatic	0
## 474	Jakarta Barat	203000000	1	Automatic	0
## 475	Bogor	116000000	0	Manual	0
## 476	Depok	111000000	0	Manual	0
## 477	Tangerang Selatan	262000000	1	Automatic	0
## 478	Bogor	194000000	1	Automatic	0
## 479	Bogor	113000000	0	Manual	0
## 480	Jakarta Utara	192000000	1	Automatic	0
## 481	Bogor	126000000	0	Automatic	0
## 482	Tangerang Selatan	106000000	0	Manual	1
## 483	Depok	209000000	1	Automatic	0
## 484	Bandung	197000000	1	Manual	0
## 485	Depok	222000000	1	Automatic	0
## 486	Jakarta Barat	164000000	0	Automatic	0
## 487	Jakarta Pusat	141000000	1	Automatic	0
## 488	Jakarta Utara	219000000	1	Automatic	0
## 489	Bogor	164000000	0	Manual	0
## 490	Jakarta Utara	122000000	0	Manual	0
## 491	Jakarta Selatan	183000000	1	Manual	0
## 492	Jakarta Barat	207000000	1	Automatic	0
## 493	Bogor	199000000	1	Automatic	0
## 494	Bogor	224000000	1	Automatic	0
## 495	Malang	113000000	0	Automatic	0
## 496	Bogor	128000000	0	Manual	0
## 497	Jakarta Utara	164000000	0	Automatic	1
## 498	Jakarta Utara	265000000	1	Automatic	0
## 499	Bekasi	200000000	1	Manual	0
## 500	Bogor	154000000	0	Automatic	0
## 501	Jakarta Barat	234000000	1	Manual	0
## 502	Jakarta Barat	151000000	1	Manual	0
## 503	Surabaya	130000000	0	Manual	1
## 504	Jakarta Selatan	242000000	1	Automatic	0
## 505	Tangerang Selatan	188000000	0	Automatic	1
## 506	Bogor	208000000	1	Automatic	0
## 507	Tangerang Selatan	128000000	0	Automatic	0
## 508	Jakarta Barat	126000000	0	Manual	0
## 509	Depok	160000000	0	Automatic	0
## 510	Depok	173000000	1	Automatic	0
## 511	Tangerang Selatan	151000000	1	Automatic	0
## 512	Bandung	99000000	0	Manual	0
## 513	Bekasi	137000000	1	Automatic	0
## 514	Depok	142000000	1	Automatic	0
## 515	Jakarta Timur	178000000	1	Automatic	0
## 516	Jakarta Selatan	219000000	1	Automatic	0
## 517	Bekasi	146000000	1	Automatic	1

## 518	Jakarta Timur	127000000	0	Automatic	0
## 519	Jakarta Timur	214000000	1	Automatic	0
## 520	Bekasi	141000000	0	Manual	0
## 521	Bekasi	108000000	0	Manual	0
## 522	Tangerang Selatan	84000000	0	Manual	0
## 523	Jakarta Selatan	213000000	1	Automatic	0
## 524	Jakarta Barat	191000000	1	Automatic	0
## 525	Bogor	145000000	0	Manual	0
## 526	Bandung	164000000	0	Automatic	1
## 527	Tangerang Selatan	134000000	1	Automatic	0
## 528	Depok	130000000	0	Automatic	0
## 529	Jakarta Barat	124000000	0	Automatic	0
## 530	Bogor	91000000	0	Manual	0
## 531	Jakarta Utara	88000000	0	Automatic	0
## 532	Bogor	154000000	0	Automatic	0
## 533	Jakarta Timur	264000000	1	Automatic	0
## 534	Bogor	340000000	0	Automatic	0
## 535	Tangerang Selatan	120000000	0	Automatic	0
## 536	Tangerang Selatan	156000000	0	Automatic	0
## 537	Jakarta Barat	168000000	1	Manual	1
## 538	Bogor	184000000	1	Automatic	1
## 539	Tangerang Selatan	222000000	1	Automatic	0
## 540	Bekasi	213000000	1	Automatic	1
## 541	Depok	274000000	1	Automatic	0
## 542	Bogor	112000000	0	Automatic	0
## 543	Tangerang Selatan	184000000	1	Automatic	1
## 544	Tangerang Selatan	244000000	1	Automatic	0
## 545	Jakarta Pusat	187000000	1	Automatic	0
## 546	Jakarta Utara	121000000	0	Manual	0
## 547	Tangerang Selatan	86000000	0	Automatic	0
## 548	Bandung	220000000	1	Automatic	1
## 549	Bogor	112000000	0	Automatic	0
## 550	Jakarta Timur	112000000	0	Manual	0
## 551	Bogor	95000000	0	Automatic	0
## 552	Tangerang Selatan	214000000	1	Automatic	0
## 553	Jakarta Barat	181000000	0	Automatic	1
## 554	Jakarta Pusat	210000000	1	Automatic	0
## 555	Bekasi	204000000	0	Automatic	0
## 556	Bogor	209000000	1	Automatic	0
## 557	Jakarta Utara	177000000	0	Manual	0
## 558	Jakarta Selatan	160000000	1	Automatic	0
## 559	Bekasi	132000000	0	Automatic	0
## 560	Tangerang Selatan	404000000	0	Automatic	0
## 561	Tangerang Selatan	129000000	1	Automatic	0
## 562	Tangerang Selatan	241000000	0	Automatic	1
## 563	Jakarta Barat	160000000	0	Automatic	0
## 564	Bandung	433000000	1	Automatic	0
## 565	Depok	163000000	0	Automatic	1
## 566	Jakarta Utara	109000000	0	Manual	0
## 567	Jakarta Pusat	136000000	0	Manual	1
## 568	Jakarta Pusat	118000000	0	Manual	0
## 569	Jakarta Utara	142000000	0	Automatic	0
## 570	Jakarta Utara	220000000	1	Automatic	0
## 571	Jakarta Barat	147000000	1	Automatic	0

## 572	Bogor	131000000	0	Automatic	1
## 573	Jakarta Barat	199000000	1	Automatic	0
## 574	Jakarta Selatan	382000000	0	Automatic	0
## 575	Bekasi	122000000	0	Manual	0
## 576	Jakarta Selatan	446000000	0	Automatic	0
## 577	Jakarta Utara	113000000	1	Manual	0
## 578	Bekasi	118000000	0	Manual	1
## 579	Bandung	100000000	0	Automatic	0
## 580	Jakarta Utara	145000000	0	Automatic	1
## 581	Jakarta Barat	122000000	0	Manual	0
## 582	Bogor	278000000	0	Automatic	1
## 583	Jakarta Barat	101000000	0	Manual	0
## 584	Jakarta Pusat	102000000	0	Manual	0
## 585	Bekasi	117000000	0	Manual	0
## 586	Jakarta Barat	222000000	1	Automatic	0
## 587	Tangerang Selatan	100000000	0	Manual	0
## 588	Tangerang Selatan	142000000	1	Automatic	0
## 589	Bekasi	155000000	1	Automatic	0
## 590	Jakarta Utara	159000000	0	Automatic	1
## 591	Bogor	224000000	1	Automatic	0
## 592	Jakarta Timur	191000000	1	Automatic	0
## 593	Depok	121000000	0	Automatic	0
## 594	Jakarta Utara	192000000	0	Automatic	1
## 595	Tangerang Selatan	165000000	0	Automatic	0
## 596	Jakarta Selatan	210000000	1	Automatic	0
## 597	Depok	110000000	0	Manual	0
## 598	Jakarta Utara	131000000	1	Automatic	0
## 599	Tangerang Selatan	184000000	1	Automatic	0
## 600	Jakarta Timur	218000000	1	Automatic	0
## 601	Jakarta Timur	202000000	1	Automatic	0
## 602	Tangerang Selatan	158000000	0	Automatic	0
## 603	Bogor	122000000	0	Automatic	0
## 604	Bogor	204000000	1	Automatic	1
## 605	Bogor	234000000	1	Automatic	0
## 606	Jakarta Utara	136000000	0	Automatic	0
## 607	Bogor	246000000	1	Automatic	0
## 608	Jakarta Utara	116000000	0	Manual	0
## 609	Bogor	149000000	0	Manual	1
##	year	plate.type	electric.parking.brake	X360.camera.view	power.sliding.door
## 1	2018	even plate	0	0	0
## 2	2015	even plate	0	0	0
## 3	2015	odd plate	0	1	0
## 4	2020	odd plate	0	0	0
## 5	2019	odd plate	0	0	0
## 6	2021	even plate	0	0	0
## 7	2022	even plate	0	0	0
## 8	2016	even plate	0	0	0
## 9	2019	odd plate	0	0	0
## 10	2021	odd plate	1	0	0
## 11	2018	even plate	0	0	0
## 12	2019	even plate	0	0	0
## 13	2019	odd plate	0	0	0
## 14	2017	odd plate	0	0	0
## 15	2019	odd plate	0	0	0

## 16	2019	even	plate	0	0	0
## 17	2015	odd	plate	0	0	0
## 18	2018	odd	plate	0	0	0
## 19	2018	odd	plate	0	0	0
## 20	2018	odd	plate	0	0	0
## 21	2017	odd	plate	0	0	0
## 22	2018	odd	plate	0	0	0
## 23	2022	even	plate	0	0	0
## 24	2019	odd	plate	0	0	0
## 25	2017	even	plate	1	0	0
## 26	2020	even	plate	0	0	0
## 27	2019	odd	plate	0	0	0
## 28	2019	even	plate	0	0	0
## 29	2021	even	plate	0	0	0
## 30	2018	even	plate	0	0	0
## 31	2019	even	plate	0	0	0
## 32	2015	odd	plate	0	0	0
## 33	2018	even	plate	1	0	0
## 34	2019	even	plate	0	0	0
## 35	2018	odd	plate	0	0	0
## 36	2016	odd	plate	0	0	0
## 37	2020	even	plate	0	0	0
## 38	2019	odd	plate	0	0	0
## 39	2020	even	plate	0	0	0
## 40	2019	odd	plate	1	0	0
## 41	2018	odd	plate	0	0	0
## 42	2014	odd	plate	0	0	0
## 43	2019	odd	plate	0	0	0
## 44	2016	even	plate	0	0	0
## 45	2016	odd	plate	0	0	0
## 46	2018	even	plate	0	0	0
## 47	2019	even	plate	1	0	0
## 48	2018	odd	plate	0	0	0
## 49	2019	odd	plate	0	0	0
## 50	2018	odd	plate	0	0	0
## 51	2019	even	plate	0	0	0
## 52	2019	even	plate	0	0	0
## 53	2017	odd	plate	0	0	0
## 54	2018	even	plate	0	0	0
## 55	2017	even	plate	0	0	0
## 56	2016	even	plate	0	0	0
## 57	2014	odd	plate	0	0	0
## 58	2019	even	plate	0	0	0
## 59	2016	odd	plate	0	0	0
## 60	2017	even	plate	0	0	0
## 61	2022	even	plate	0	0	0
## 62	2020	odd	plate	0	0	0
## 63	2014	even	plate	0	0	0
## 64	2018	odd	plate	0	0	0
## 65	2021	odd	plate	0	0	0
## 66	2020	odd	plate	0	0	0
## 67	2018	even	plate	0	0	0
## 68	2018	odd	plate	0	0	0
## 69	2020	odd	plate	0	0	0

## 70	2016	even	plate	0	0	0
## 71	2019	even	plate	0	0	0
## 72	2018	odd	plate	0	0	0
## 73	2019	even	plate	0	0	0
## 74	2017	even	plate	0	0	0
## 75	2019	even	plate	0	0	0
## 76	2021	odd	plate	0	0	0
## 77	2019	even	plate	0	0	0
## 78	2019	odd	plate	0	0	0
## 79	2021	odd	plate	0	1	0
## 80	2019	even	plate	0	0	0
## 81	2019	even	plate	0	0	0
## 82	2019	odd	plate	0	1	0
## 83	2019	odd	plate	0	0	0
## 84	2017	even	plate	0	0	0
## 85	2014	even	plate	0	0	0
## 86	2019	odd	plate	0	0	0
## 87	2019	even	plate	0	0	0
## 88	2018	odd	plate	0	0	0
## 89	2018	even	plate	0	0	0
## 90	2018	even	plate	0	0	0
## 91	2018	even	plate	0	0	0
## 92	2020	odd	plate	0	0	0
## 93	2019	even	plate	0	0	0
## 94	2015	even	plate	0	0	0
## 95	2019	even	plate	0	0	0
## 96	2022	even	plate	0	0	0
## 97	2019	odd	plate	0	0	0
## 98	2021	even	plate	0	0	0
## 99	2015	even	plate	0	0	0
## 100	2017	odd	plate	0	0	0
## 101	2022	even	plate	0	0	0
## 102	2018	even	plate	0	0	0
## 103	2017	odd	plate	0	0	0
## 104	2020	even	plate	0	0	0
## 105	2016	even	plate	0	0	0
## 106	2019	odd	plate	0	0	0
## 107	2021	even	plate	0	0	0
## 108	2021	even	plate	0	0	0
## 109	2019	odd	plate	0	1	0
## 110	2018	odd	plate	0	0	0
## 111	2016	even	plate	0	0	0
## 112	2016	even	plate	0	0	0
## 113	2021	odd	plate	0	0	0
## 114	2017	even	plate	0	0	0
## 115	2020	odd	plate	0	0	0
## 116	2020	odd	plate	0	0	0
## 117	2019	odd	plate	0	0	0
## 118	2019	even	plate	0	0	0
## 119	2019	even	plate	0	0	0
## 120	2019	even	plate	0	0	0
## 121	2016	even	plate	0	0	0
## 122	2018	odd	plate	0	0	0
## 123	2018	even	plate	1	0	0

## 124 2020 odd plate	0	0	0
## 125 2021 odd plate	0	0	0
## 126 2017 odd plate	0	0	0
## 127 2019 odd plate	0	0	0
## 128 2019 odd plate	1	0	0
## 129 2019 odd plate	0	0	0
## 130 2017 even plate	0	0	0
## 131 2019 odd plate	0	0	0
## 132 2021 even plate	0	0	0
## 133 2016 odd plate	0	0	0
## 134 2020 odd plate	0	0	0
## 135 2018 even plate	0	0	0
## 136 2019 even plate	0	0	0
## 137 2020 even plate	0	0	0
## 138 2021 odd plate	0	0	0
## 139 2021 odd plate	0	0	0
## 140 2019 odd plate	0	0	0
## 141 2016 even plate	0	0	0
## 142 2017 even plate	0	0	0
## 143 2018 even plate	0	0	0
## 144 2019 even plate	0	0	0
## 145 2018 odd plate	0	0	0
## 146 2019 odd plate	1	0	0
## 147 2019 even plate	0	0	0
## 148 2018 even plate	0	0	0
## 149 2019 even plate	0	0	0
## 150 2018 even plate	0	0	0
## 151 2017 even plate	0	0	0
## 152 2017 odd plate	0	0	0
## 153 2018 odd plate	0	0	0
## 154 2020 even plate	0	0	0
## 155 2020 odd plate	0	0	0
## 156 2022 even plate	0	0	0
## 157 2020 even plate	0	0	0
## 158 2018 odd plate	0	0	0
## 159 2018 even plate	0	0	0
## 160 2019 odd plate	0	0	0
## 161 2016 even plate	0	0	0
## 162 2016 even plate	0	0	0
## 163 2018 even plate	1	0	0
## 164 2015 even plate	0	0	0
## 165 2019 odd plate	0	0	0
## 166 2017 odd plate	0	0	0
## 167 2018 even plate	0	0	0
## 168 2018 odd plate	0	0	0
## 169 2016 even plate	0	1	0
## 170 2018 even plate	0	0	0
## 171 2019 odd plate	0	0	0
## 172 2020 even plate	0	0	0
## 173 2018 odd plate	0	0	0
## 174 2017 even plate	0	0	0
## 175 2021 even plate	0	0	0
## 176 2018 even plate	0	0	0
## 177 2018 even plate	0	0	1

## 178 2020 odd plate	0	0	0
## 179 2017 odd plate	0	0	0
## 180 2018 even plate	0	0	0
## 181 2019 odd plate	0	0	0
## 182 2017 even plate	0	0	0
## 183 2016 odd plate	0	0	0
## 184 2019 even plate	0	0	0
## 185 2016 odd plate	0	0	0
## 186 2018 odd plate	0	0	0
## 187 2018 even plate	0	0	0
## 188 2018 odd plate	0	0	0
## 189 2019 even plate	0	0	0
## 190 2016 even plate	0	0	0
## 191 2017 even plate	0	0	0
## 192 2018 even plate	0	0	0
## 193 2018 even plate	0	0	0
## 194 2018 even plate	0	0	0
## 195 2020 even plate	0	0	0
## 196 2018 even plate	1	0	0
## 197 2021 even plate	0	0	0
## 198 2019 even plate	0	0	0
## 199 2015 even plate	0	0	0
## 200 2015 even plate	0	0	0
## 201 2022 odd plate	0	0	0
## 202 2018 even plate	0	0	0
## 203 2018 even plate	0	0	0
## 204 2019 odd plate	0	0	0
## 205 2019 even plate	0	0	0
## 206 2019 odd plate	0	0	0
## 207 2019 even plate	0	1	0
## 208 2013 even plate	0	0	0
## 209 2018 even plate	0	0	0
## 210 2010 even plate	0	0	0
## 211 2014 even plate	0	0	0
## 212 2017 odd plate	0	0	0
## 213 2018 odd plate	1	0	0
## 214 2015 even plate	0	0	0
## 215 2016 even plate	0	0	0
## 216 2019 odd plate	0	0	0
## 217 2018 even plate	0	0	0
## 218 2017 even plate	0	0	0
## 219 2013 odd plate	0	0	0
## 220 2017 even plate	0	0	0
## 221 2018 even plate	0	0	0
## 222 2021 odd plate	0	0	0
## 223 2012 odd plate	0	0	0
## 224 2022 even plate	0	0	0
## 225 2018 even plate	0	0	0
## 226 2018 even plate	0	0	0
## 227 2016 odd plate	0	0	0
## 228 2017 odd plate	0	0	0
## 229 2018 odd plate	0	0	0
## 230 2018 odd plate	0	0	0
## 231 2019 odd plate	0	0	0

## 232 2016 odd plate	0	0	0
## 233 2018 even plate	0	0	0
## 234 2017 even plate	0	0	0
## 235 2019 odd plate	0	0	0
## 236 2021 even plate	0	0	0
## 237 2021 even plate	0	0	0
## 238 2018 odd plate	0	0	0
## 239 2021 even plate	0	0	0
## 240 2018 even plate	0	0	0
## 241 2018 even plate	0	0	0
## 242 2021 odd plate	1	0	0
## 243 2018 even plate	0	0	0
## 244 2022 even plate	0	0	0
## 245 2018 odd plate	0	0	0
## 246 2021 odd plate	0	0	0
## 247 2016 odd plate	0	0	0
## 248 2017 even plate	0	0	0
## 249 2017 odd plate	0	0	0
## 250 2019 odd plate	0	0	0
## 251 2017 odd plate	0	0	0
## 252 2019 odd plate	0	0	0
## 253 2017 odd plate	0	0	0
## 254 2016 odd plate	0	0	0
## 255 2014 even plate	0	0	0
## 256 2018 odd plate	1	0	0
## 257 2018 even plate	0	0	0
## 258 2015 odd plate	0	1	0
## 259 2017 even plate	0	0	0
## 260 2019 odd plate	0	0	0
## 261 2014 odd plate	0	0	0
## 262 2019 odd plate	0	0	0
## 263 2015 odd plate	0	0	0
## 264 2020 odd plate	0	0	0
## 265 2020 odd plate	0	0	0
## 266 2021 odd plate	0	0	0
## 267 2017 even plate	0	0	0
## 268 2019 odd plate	0	0	0
## 269 2020 odd plate	0	0	0
## 270 2018 odd plate	0	0	0
## 271 2019 odd plate	0	0	0
## 272 2013 odd plate	0	0	0
## 273 2021 odd plate	0	0	0
## 274 2020 odd plate	0	0	0
## 275 2017 odd plate	0	0	0
## 276 2017 odd plate	0	0	0
## 277 2018 odd plate	0	0	0
## 278 2018 even plate	0	0	0
## 279 2019 even plate	0	0	0
## 280 2018 odd plate	0	0	0
## 281 2017 even plate	0	0	0
## 282 2020 even plate	0	0	0
## 283 2016 odd plate	0	0	0
## 284 2018 odd plate	0	0	0
## 285 2019 even plate	0	0	0

## 286 2017 even plate	0	0	0
## 287 2017 odd plate	0	0	0
## 288 2020 even plate	0	0	0
## 289 2020 odd plate	0	0	0
## 290 2019 even plate	0	0	0
## 291 2021 even plate	0	0	0
## 292 2020 odd plate	1	0	0
## 293 2017 even plate	0	0	0
## 294 2016 even plate	0	0	0
## 295 2016 odd plate	0	0	0
## 296 2017 even plate	0	0	1
## 297 2015 odd plate	0	0	0
## 298 2018 odd plate	0	0	0
## 299 2018 even plate	0	0	0
## 300 2017 even plate	0	0	0
## 301 2017 odd plate	0	0	0
## 302 2017 odd plate	0	0	0
## 303 2020 even plate	0	0	0
## 304 2018 odd plate	0	0	0
## 305 2019 odd plate	0	0	0
## 306 2019 even plate	0	0	0
## 307 2018 even plate	0	0	0
## 308 2019 even plate	0	0	1
## 309 2018 even plate	0	0	0
## 310 2019 odd plate	0	0	0
## 311 2019 even plate	0	0	0
## 312 2014 even plate	0	0	0
## 313 2018 even plate	0	0	1
## 314 2018 even plate	0	0	0
## 315 2018 odd plate	0	0	0
## 316 2014 even plate	0	0	0
## 317 2018 odd plate	0	0	0
## 318 2017 even plate	0	0	0
## 319 2022 odd plate	0	0	0
## 320 2019 even plate	0	0	0
## 321 2019 even plate	0	0	0
## 322 2019 even plate	0	0	0
## 323 2021 odd plate	0	0	0
## 324 2014 even plate	0	0	0
## 325 2023 odd plate	0	0	0
## 326 2014 even plate	0	0	0
## 327 2017 even plate	0	0	0
## 328 2019 odd plate	0	0	0
## 329 2020 odd plate	0	0	0
## 330 2019 odd plate	0	0	0
## 331 2018 odd plate	0	0	0
## 332 2018 odd plate	0	0	0
## 333 2018 odd plate	0	0	0
## 334 2018 even plate	0	0	0
## 335 2019 even plate	0	0	0
## 336 2019 even plate	0	0	0
## 337 2016 even plate	0	0	0
## 338 2017 even plate	0	0	0
## 339 2017 odd plate	0	0	0

## 340 2019 even plate	0	0	0
## 341 2019 odd plate	0	0	1
## 342 2019 even plate	0	0	0
## 343 2018 even plate	0	0	0
## 344 2018 odd plate	0	0	0
## 345 2019 even plate	0	0	0
## 346 2018 even plate	0	0	0
## 347 2014 odd plate	0	0	0
## 348 2013 odd plate	0	0	0
## 349 2015 even plate	0	0	0
## 350 2017 even plate	0	0	0
## 351 2019 odd plate	0	0	0
## 352 2020 even plate	0	0	0
## 353 2018 odd plate	0	0	0
## 354 2015 even plate	0	0	0
## 355 2018 even plate	0	0	0
## 356 2018 odd plate	0	0	0
## 357 2020 odd plate	0	0	0
## 358 2018 even plate	0	0	0
## 359 2018 odd plate	0	0	0
## 360 2020 odd plate	0	1	0
## 361 2018 even plate	0	0	0
## 362 2019 even plate	0	0	0
## 363 2019 even plate	0	0	0
## 364 2021 odd plate	0	0	0
## 365 2018 odd plate	0	0	0
## 366 2018 odd plate	0	0	0
## 367 2017 odd plate	0	0	0
## 368 2018 even plate	0	0	0
## 369 2021 even plate	0	0	0
## 370 2017 even plate	0	0	0
## 371 2020 odd plate	0	0	0
## 372 2016 odd plate	0	0	0
## 373 2017 even plate	0	0	0
## 374 2017 even plate	0	0	0
## 375 2016 even plate	0	0	0
## 376 2014 odd plate	0	0	0
## 377 2020 odd plate	0	0	0
## 378 2014 even plate	0	0	0
## 379 2019 odd plate	0	0	0
## 380 2022 even plate	0	0	0
## 381 2021 even plate	0	0	0
## 382 2018 odd plate	0	0	0
## 383 2016 odd plate	0	0	0
## 384 2018 odd plate	0	0	0
## 385 2019 odd plate	0	0	0
## 386 2019 odd plate	0	0	0
## 387 2020 even plate	0	0	0
## 388 2019 even plate	0	0	0
## 389 2014 odd plate	0	0	0
## 390 2017 even plate	0	0	0
## 391 2015 odd plate	0	0	0
## 392 2016 even plate	0	0	0
## 393 2022 even plate	0	0	0

## 394 2021 even plate	0	0	0
## 395 2019 odd plate	0	0	0
## 396 2019 even plate	0	0	0
## 397 2015 even plate	0	0	0
## 398 2016 odd plate	0	0	0
## 399 2018 odd plate	0	0	0
## 400 2019 odd plate	0	0	0
## 401 2014 odd plate	0	0	0
## 402 2016 odd plate	0	0	0
## 403 2017 even plate	0	0	0
## 404 2017 even plate	0	0	0
## 405 2020 odd plate	0	0	0
## 406 2015 odd plate	0	0	0
## 407 2019 odd plate	0	0	0
## 408 2017 even plate	0	0	0
## 409 2019 even plate	0	0	0
## 410 2019 even plate	0	0	0
## 411 2019 odd plate	0	0	0
## 412 2020 even plate	0	0	0
## 413 2017 even plate	0	0	0
## 414 2018 even plate	0	0	0
## 415 2018 even plate	0	0	0
## 416 2018 even plate	0	0	0
## 417 2018 even plate	0	0	0
## 418 2018 odd plate	0	0	0
## 419 2014 odd plate	0	0	0
## 420 2018 odd plate	0	0	0
## 421 2019 even plate	0	0	0
## 422 2019 even plate	0	0	0
## 423 2016 odd plate	0	0	0
## 424 2018 odd plate	0	0	0
## 425 2013 odd plate	0	0	0
## 426 2019 even plate	0	0	0
## 427 2021 odd plate	0	0	0
## 428 2018 even plate	0	0	0
## 429 2016 odd plate	0	0	0
## 430 2018 even plate	0	0	0
## 431 2019 even plate	0	0	0
## 432 2017 even plate	0	0	0
## 433 2019 odd plate	0	0	0
## 434 2019 even plate	0	0	0
## 435 2018 odd plate	0	0	0
## 436 2019 odd plate	0	0	0
## 437 2019 odd plate	0	0	0
## 438 2016 even plate	0	0	0
## 439 2016 even plate	0	0	0
## 440 2017 odd plate	0	0	0
## 441 2019 even plate	0	0	0
## 442 2018 even plate	0	0	0
## 443 2020 even plate	0	0	0
## 444 2016 odd plate	1	0	0
## 445 2018 odd plate	0	0	0
## 446 2021 even plate	0	0	0
## 447 2017 even plate	0	0	0

## 448 2019 even plate	0	0	0
## 449 2018 even plate	0	0	0
## 450 2018 even plate	0	0	0
## 451 2018 even plate	0	0	0
## 452 2019 odd plate	1	0	0
## 453 2019 odd plate	0	0	0
## 454 2018 odd plate	0	0	0
## 455 2014 even plate	0	0	0
## 456 2016 odd plate	0	0	0
## 457 2019 odd plate	0	0	0
## 458 2019 odd plate	0	0	0
## 459 2014 odd plate	0	0	0
## 460 2019 odd plate	0	0	0
## 461 2019 even plate	0	0	0
## 462 2021 even plate	0	0	0
## 463 2021 odd plate	0	0	0
## 464 2021 even plate	0	0	0
## 465 2014 odd plate	0	0	0
## 466 2017 even plate	0	0	0
## 467 2019 even plate	0	0	0
## 468 2019 odd plate	0	0	0
## 469 2017 odd plate	0	0	0
## 470 2018 odd plate	0	0	0
## 471 2016 odd plate	0	0	0
## 472 2014 even plate	0	0	0
## 473 2020 even plate	0	0	0
## 474 2019 odd plate	0	1	0
## 475 2014 even plate	0	0	0
## 476 2018 odd plate	0	0	0
## 477 2021 even plate	0	0	0
## 478 2019 even plate	0	0	0
## 479 2017 odd plate	0	0	0
## 480 2019 odd plate	0	0	0
## 481 2018 odd plate	0	0	0
## 482 2016 even plate	0	0	0
## 483 2018 odd plate	0	0	0
## 484 2020 odd plate	0	0	0
## 485 2021 odd plate	0	0	0
## 486 2021 even plate	0	0	0
## 487 2019 even plate	0	0	0
## 488 2017 odd plate	0	0	0
## 489 2020 odd plate	0	0	0
## 490 2019 odd plate	0	0	0
## 491 2018 even plate	0	0	0
## 492 2017 even plate	0	0	0
## 493 2015 even plate	0	0	0
## 494 2021 even plate	0	0	0
## 495 2016 even plate	0	0	0
## 496 2018 odd plate	0	0	0
## 497 2018 even plate	0	0	0
## 498 2019 even plate	0	0	0
## 499 2019 even plate	0	0	0
## 500 2019 odd plate	0	0	0
## 501 2020 odd plate	0	0	0

## 502 2017 odd plate	0	0	0
## 503 2020 even plate	0	0	0
## 504 2021 even plate	0	0	0
## 505 2020 odd plate	0	0	0
## 506 2021 odd plate	0	0	0
## 507 2019 odd plate	0	0	0
## 508 2018 even plate	0	0	0
## 509 2020 odd plate	0	0	0
## 510 2014 odd plate	0	0	0
## 511 2022 even plate	0	0	0
## 512 2018 odd plate	0	0	0
## 513 2020 even plate	0	0	0
## 514 2020 even plate	0	0	0
## 515 2019 odd plate	0	0	0
## 516 2017 even plate	0	0	0
## 517 2014 even plate	0	0	0
## 518 2018 even plate	0	0	0
## 519 2018 odd plate	0	0	0
## 520 2019 odd plate	0	0	0
## 521 2016 even plate	0	0	0
## 522 2014 even plate	0	0	0
## 523 2018 even plate	0	0	0
## 524 2021 odd plate	0	0	0
## 525 2018 even plate	0	0	0
## 526 2017 odd plate	0	0	0
## 527 2019 odd plate	0	0	0
## 528 2018 odd plate	0	0	0
## 529 2017 odd plate	0	0	0
## 530 2018 even plate	0	0	0
## 531 2012 odd plate	0	0	0
## 532 2017 odd plate	0	0	0
## 533 2019 even plate	1	0	0
## 534 2019 even plate	0	0	0
## 535 2017 odd plate	0	0	0
## 536 2019 odd plate	0	0	0
## 537 2019 odd plate	0	0	0
## 538 2018 odd plate	0	0	0
## 539 2019 even plate	0	0	0
## 540 2018 odd plate	0	0	0
## 541 2014 even plate	0	0	0
## 542 2016 even plate	0	0	0
## 543 2018 even plate	0	0	0
## 544 2021 even plate	0	0	0
## 545 2017 even plate	0	0	1
## 546 2019 odd plate	0	0	0
## 547 2015 even plate	0	0	0
## 548 2021 odd plate	0	0	0
## 549 2018 odd plate	0	0	0
## 550 2016 odd plate	0	0	0
## 551 2013 even plate	0	0	0
## 552 2020 even plate	0	0	0
## 553 2015 even plate	0	0	0
## 554 2018 even plate	0	0	0
## 555 2019 odd plate	1	1	0

## 556 2015 odd plate	0	0	0
## 557 2022 odd plate	0	0	0
## 558 2016 even plate	0	0	0
## 559 2017 even plate	0	0	0
## 560 2016 even plate	1	0	0
## 561 2010 even plate	0	0	1
## 562 2022 odd plate	0	0	0
## 563 2018 even plate	0	0	0
## 564 2019 even plate	0	0	0
## 565 2017 even plate	0	0	1
## 566 2019 even plate	0	0	0
## 567 2022 even plate	0	0	0
## 568 2016 odd plate	0	0	0
## 569 2018 odd plate	0	0	0
## 570 2019 even plate	0	0	0
## 571 2021 even plate	0	0	0
## 572 2015 odd plate	0	0	0
## 573 2016 odd plate	0	0	0
## 574 2018 odd plate	0	0	0
## 575 2014 odd plate	0	0	0
## 576 2019 even plate	0	0	0
## 577 2015 even plate	0	0	0
## 578 2021 odd plate	0	0	0
## 579 2015 even plate	0	0	0
## 580 2019 odd plate	0	0	0
## 581 2019 even plate	0	0	0
## 582 2019 odd plate	0	0	0
## 583 2016 odd plate	0	0	0
## 584 2012 odd plate	0	0	0
## 585 2018 even plate	0	0	0
## 586 2019 even plate	0	0	0
## 587 2011 even plate	0	0	0
## 588 2016 even plate	0	0	0
## 589 2018 even plate	0	0	0
## 590 2017 even plate	0	0	0
## 591 2019 even plate	0	0	0
## 592 2017 odd plate	0	0	1
## 593 2017 even plate	0	0	0
## 594 2020 odd plate	0	0	0
## 595 2020 even plate	0	0	0
## 596 2018 even plate	0	0	0
## 597 2022 even plate	0	0	0
## 598 2017 odd plate	0	0	0
## 599 2016 even plate	0	0	0
## 600 2020 even plate	0	0	0
## 601 2019 odd plate	0	0	0
## 602 2013 even plate	0	0	0
## 603 2018 odd plate	0	0	0
## 604 2020 odd plate	0	0	0
## 605 2021 odd plate	0	0	0
## 606 2019 odd plate	0	0	0
## 607 2022 even plate	0	0	0
## 608 2017 even plate	0	0	0
## 609 2017 even plate	0	0	0

##	brand	sports.mode	sun.roof	instalment_month
## 1	Daihatsu	0	0	2060000
## 2	Toyota	0	0	1670000
## 3	Nissan	0	0	3440000
## 4	Toyota	0	0	4440000
## 5	Toyota	0	0	2380000
## 6	Toyota	0	0	3670000
## 7	Toyota	0	0	4300000
## 8	Mitsubishi	0	1	3870000
## 9	Honda	0	0	2850000
## 10	Honda	0	0	5050000
## 11	Toyota	0	0	2380000
## 12	Daihatsu	0	0	2320000
## 13	Daihatsu	0	0	2240000
## 14	Daihatsu	0	0	2340000
## 15	Toyota	0	0	2400000
## 16	Wuling	0	1	4030000
## 17	Daihatsu	0	0	2870000
## 18	Toyota	0	0	2380000
## 19	Honda	0	0	3160000
## 20	Daihatsu	0	0	2200000
## 21	Daihatsu	0	0	2150000
## 22	Daihatsu	0	0	2260000
## 23	Suzuki	0	0	3870000
## 24	Daihatsu	0	0	2320000
## 25	Honda	0	0	6930000
## 26	Wuling	0	0	2069999
## 27	Wuling	0	1	4120000
## 28	Nissan	0	0	4010000
## 29	Toyota	0	0	2810000
## 30	Honda	0	0	2750000
## 31	Daihatsu	0	0	3930000
## 32	Toyota	0	0	2590000
## 33	Honda	0	0	4240000
## 34	Mitsubishi	0	0	4400000
## 35	Mitsubishi	0	0	4160000
## 36	Daihatsu	0	0	1810000
## 37	Toyota	0	0	2510000
## 38	Daihatsu	0	0	2180000
## 39	Toyota	0	0	2710000
## 40	Honda	0	0	5050000
## 41	Mitsubishi	0	0	4650000
## 42	Mitsubishi	1	0	3200000
## 43	Honda	0	0	2890000
## 44	Toyota	0	0	2910000
## 45	Mitsubishi	0	1	3970000
## 46	Honda	0	0	5180000
## 47	Wuling	0	0	4050000
## 48	Suzuki	0	0	2270000
## 49	Wuling	1	0	4139999
## 50	Daihatsu	0	0	2380000
## 51	Daihatsu	0	0	3930000
## 52	Mitsubishi	0	0	4340000
## 53	Honda	0	0	6830000

## 54	Honda	0	0	5200000
## 55	Toyota	1	0	4930000
## 56	Honda	1	0	3160000
## 57	Toyota	0	0	3930000
## 58	Toyota	0	0	2490000
## 59	Toyota	1	0	4750000
## 60	Suzuki	0	0	2240000
## 61	Toyota	0	0	2910000
## 62	Daihatsu	0	0	2510000
## 63	Honda	0	0	2590000
## 64	Daihatsu	0	0	2160000
## 65	Daihatsu	0	0	4220000
## 66	Honda	0	0	3000000
## 67	Honda	0	0	5240000
## 68	Honda	1	0	2830000
## 69	Honda	1	0	3670000
## 70	Toyota	0	0	7440000
## 71	Mazda	0	0	8680000
## 72	Nissan	0	0	3020000
## 73	Suzuki	0	0	3320000
## 74	Toyota	0	0	2020000
## 75	Toyota	0	0	2610000
## 76	Suzuki	0	0	3020000
## 77	Honda	0	0	4710000
## 78	Wuling	0	0	3420000
## 79	Wuling	0	1	5560000
## 80	Wuling	0	0	2120000
## 81	Daihatsu	0	0	2040000
## 82	Wuling	0	0	4160000
## 83	Daihatsu	0	0	2380000
## 84	Toyota	1	0	6440000
## 85	Toyota	0	0	1770000
## 86	Suzuki	0	0	3590000
## 87	Suzuki	0	0	3260000
## 88	Daihatsu	0	0	2320000
## 89	Mitsubishi	0	0	4260000
## 90	Mitsubishi	0	0	4280000
## 91	Daihatsu	0	0	2040000
## 92	Daihatsu	0	0	2770000
## 93	Wuling	0	1	4139999
## 94	Daihatsu	0	0	3020000
## 95	Wuling	0	1	4360000
## 96	Toyota	0	0	4460000
## 97	Honda	0	0	2890000
## 98	Honda	0	0	5320000
## 99	Honda	0	0	4030000
## 100	Daihatsu	0	0	1960000
## 101	Toyota	0	0	7330000
## 102	Daihatsu	0	0	2360000
## 103	Toyota	0	0	2750000
## 104	Toyota	0	0	4560000
## 105	Honda	0	0	2770000
## 106	Daihatsu	0	0	3260000
## 107	Wuling	1	0	3670000

## 108	Wuling	0	1	4200000
## 109	Wuling	0	0	4139999
## 110	Honda	0	0	4730000
## 111	Toyota	0	0	7440000
## 112	Mitsubishi	0	1	7950000
## 113	Honda	0	0	5010000
## 114	Mitsubishi	0	0	4099999
## 115	Suzuki	0	0	4280000
## 116	Daihatsu	0	0	2040000
## 117	Honda	1	0	3180000
## 118	Honda	1	0	3500000
## 119	Toyota	1	0	3890000
## 120	Toyota	0	0	4320000
## 121	Honda	1	0	2950000
## 122	Toyota	0	0	4099999
## 123	Honda	0	0	4730000
## 124	Daihatsu	0	0	3320000
## 125	Toyota	0	0	4870000
## 126	Honda	0	0	5050000
## 127	Honda	1	0	3140000
## 128	Honda	0	0	5420000
## 129	Daihatsu	0	0	2340000
## 130	Honda	0	0	6850000
## 131	Toyota	0	0	4440000
## 132	Daihatsu	0	0	4400000
## 133	Suzuki	0	0	2630000
## 134	Toyota	1	0	6850000
## 135	Honda	1	0	2890000
## 136	Daihatsu	0	0	2380000
## 137	Toyota	0	0	4730000
## 138	Toyota	0	0	4890000
## 139	Wuling	0	1	4910000
## 140	Daihatsu	0	0	2340000
## 141	Toyota	0	0	1940000
## 142	Daihatsu	0	0	1890000
## 143	Daihatsu	0	0	2000000
## 144	Toyota	0	0	2530000
## 145	Toyota	0	0	4220000
## 146	Wuling	0	1	3630000
## 147	Mitsubishi	0	0	4030000
## 148	Mitsubishi	0	0	4300000
## 149	Suzuki	0	0	2420000
## 150	Toyota	0	0	2570000
## 151	Suzuki	0	0	3100000
## 152	Toyota	0	0	2140000
## 153	Daihatsu	0	0	2300000
## 154	Mitsubishi	0	1	7230000
## 155	Honda	1	0	3240000
## 156	Toyota	0	0	3140000
## 157	Daihatsu	0	0	2770000
## 158	Toyota	0	0	4300000
## 159	Daihatsu	0	0	3950000
## 160	Daihatsu	0	0	3910000
## 161	Honda	0	0	4280000

## 162	Honda	0	0	4300000
## 163	Honda	0	0	5200000
## 164	Mazda	0	0	3750000
## 165	Toyota	0	0	4460000
## 166	Toyota	1	0	5050000
## 167	Toyota	0	0	2440000
## 168	Toyota	0	0	2470000
## 169	Nissan	0	0	3870000
## 170	Honda	0	0	4810000
## 171	Toyota	0	0	4340000
## 172	Honda	0	0	3730000
## 173	Honda	0	0	3080000
## 174	Suzuki	0	0	3040000
## 175	Daihatsu	0	0	4440000
## 176	Toyota	0	0	4340000
## 177	Toyota	0	0	4050000
## 178	Wuling	0	0	2590000
## 179	Suzuki	0	0	3440000
## 180	Mitsubishi	0	0	4300000
## 181	Suzuki	0	0	3990000
## 182	Suzuki	0	0	2630000
## 183	Daihatsu	0	0	2260000
## 184	Daihatsu	0	0	2080000
## 185	Mazda	0	0	3550000
## 186	Daihatsu	0	0	3770000
## 187	Honda	0	0	5220000
## 188	Honda	0	0	5260000
## 189	Mitsubishi	0	0	4280000
## 190	Honda	1	0	3340000
## 191	Honda	0	0	4179999
## 192	Toyota	1	0	5320000
## 193	Daihatsu	0	0	2440000
## 194	Honda	1	0	3080000
## 195	Honda	1	0	3300000
## 196	Honda	0	0	5260000
## 197	Honda	0	0	5340000
## 198	Honda	0	0	5320000
## 199	Mazda	0	0	3300000
## 200	Mazda	0	0	3380000
## 201	Toyota	0	0	2770000
## 202	Mitsubishi	0	0	4320000
## 203	Mitsubishi	0	0	4340000
## 204	Nissan	0	0	4050000
## 205	Toyota	0	0	4500000
## 206	Toyota	0	0	4460000
## 207	Daihatsu	0	0	4139999
## 208	Daihatsu	0	0	2160000
## 209	Honda	0	0	3710000
## 210	Toyota	0	0	2630000
## 211	Nissan	0	0	2040000
## 212	Daihatsu	0	0	2000000
## 213	Honda	0	0	4560000
## 214	Honda	0	0	2830000
## 215	Toyota	0	0	2060000

## 216	Daihatsu	0	0	2400000
## 217	Daihatsu	0	0	2060000
## 218	Toyota	0	0	3280000
## 219	Suzuki	0	0	2260000
## 220	Datsun	0	0	1570000
## 221	Toyota	0	0	2280000
## 222	Daihatsu	0	0	3850000
## 223	Daihatsu	0	0	2080000
## 224	Toyota	0	0	4400000
## 225	Toyota	1	0	6400000
## 226	Suzuki	0	0	3200000
## 227	Toyota	0	0	2830000
## 228	Honda	1	0	3200000
## 229	Mitsubishi	0	0	4099999
## 230	Mitsubishi	0	0	4260000
## 231	Mazda	0	0	6580000
## 232	Honda	0	0	2670000
## 233	Mitsubishi	0	0	8720000
## 234	Toyota	1	0	5660000
## 235	Suzuki	0	0	3790000
## 236	Toyota	0	0	2950000
## 237	Suzuki	0	0	4500000
## 238	Suzuki	0	0	3360000
## 239	Honda	0	0	5340000
## 240	Toyota	1	0	5200000
## 241	Mitsubishi	0	0	4340000
## 242	Wuling	0	0	4480000
## 243	Suzuki	0	0	2550000
## 244	Hyundai	0	1	6500000
## 245	Mitsubishi	0	0	3950000
## 246	Toyota	0	0	4560000
## 247	Toyota	0	0	5200000
## 248	Suzuki	0	0	3140000
## 249	Suzuki	0	0	3160000
## 250	Suzuki	0	0	3690000
## 251	Honda	1	0	2910000
## 252	Honda	1	0	3180000
## 253	Honda	0	0	3440000
## 254	Suzuki	0	0	3220000
## 255	Honda	1	0	2970000
## 256	Hyundai	0	1	8500000
## 257	Mitsubishi	0	0	4320000
## 258	Nissan	0	0	4050000
## 259	Honda	1	0	2790000
## 260	Honda	1	0	3180000
## 261	Honda	0	0	2160000
## 262	Mazda	0	0	8640000
## 263	Toyota	0	0	2240000
## 264	Honda	1	0	3930000
## 265	Honda	1	0	3300000
## 266	Suzuki	0	0	4260000
## 267	Mitsubishi	0	0	4139999
## 268	Suzuki	0	0	3810000
## 269	Honda	1	0	3690000

## 270	Mitsubishi	0	0	4320000
## 271	Toyota	0	0	2690000
## 272	Toyota	0	0	3630000
## 273	Wuling	0	0	3650000
## 274	Honda	1	0	3320000
## 275	Mazda	0	0	7600000
## 276	Toyota	0	0	3790000
## 277	Honda	1	0	3040000
## 278	Toyota	0	0	6600000
## 279	Wuling	0	0	2420000
## 280	Mitsubishi	0	0	3890000
## 281	Toyota	0	0	6460000
## 282	Toyota	0	0	4650000
## 283	Honda	1	0	2590000
## 284	Mitsubishi	0	0	4030000
## 285	Suzuki	0	0	3730000
## 286	Toyota	0	0	5110000
## 287	Toyota	0	0	3550000
## 288	Honda	1	0	3770000
## 289	Honda	1	0	3300000
## 290	Honda	0	0	4280000
## 291	Honda	0	0	5500000
## 292	Honda	0	0	5750000
## 293	Suzuki	0	0	2440000
## 294	Nissan	0	0	4280000
## 295	Suzuki	0	0	3220000
## 296	Mazda	0	0	4540000
## 297	Nissan	0	0	2360000
## 298	Mazda	0	0	5660000
## 299	Mitsubishi	0	0	4360000
## 300	Nissan	0	0	2470000
## 301	Daihatsu	0	0	2810000
## 302	Mitsubishi	0	0	4010000
## 303	Suzuki	0	0	4010000
## 304	Daihatsu	0	0	3180000
## 305	Toyota	0	0	3650000
## 306	Honda	1	0	3180000
## 307	Wuling	0	0	2140000
## 308	Toyota	1	0	3810000
## 309	Honda	1	0	2870000
## 310	Honda	1	0	3220000
## 311	Honda	1	0	4139999
## 312	Honda	1	0	2810000
## 313	Toyota	0	0	3420000
## 314	Mitsubishi	0	0	4320000
## 315	Daihatsu	0	0	2530000
## 316	Nissan	0	0	2570000
## 317	Toyota	0	0	4320000
## 318	Nissan	0	0	2530000
## 319	Toyota	0	0	11040000
## 320	Toyota	0	0	6420000
## 321	Mitsubishi	0	0	4179999
## 322	Toyota	0	0	3610000
## 323	Toyota	1	0	6500000

## 324	Honda	1	0	2690000
## 325	Suzuki	0	0	2950000
## 326	Toyota	0	0	3670000
## 327	Suzuki	0	0	2630000
## 328	Mitsubishi	0	0	4520000
## 329	Daihatsu	0	0	4160000
## 330	Toyota	0	0	2750000
## 331	Honda	1	0	2970000
## 332	Mitsubishi	0	0	4139999
## 333	Mitsubishi	0	0	4360000
## 334	Mitsubishi	0	0	4380000
## 335	Mitsubishi	0	0	4560000
## 336	Mitsubishi	0	0	4540000
## 337	Toyota	1	0	5810000
## 338	Suzuki	0	0	3280000
## 339	Toyota	0	0	5640000
## 340	Honda	0	0	3670000
## 341	Toyota	0	0	4030000
## 342	Mitsubishi	0	0	4200000
## 343	Mitsubishi	0	0	4300000
## 344	Toyota	0	0	2690000
## 345	Nissan	0	0	6640000
## 346	Toyota	0	0	4340000
## 347	Daihatsu	0	0	2530000
## 348	Nissan	0	0	1960000
## 349	Honda	0	0	2870000
## 350	Toyota	1	0	7740000
## 351	Honda	1	0	3670000
## 352	Honda	0	0	3830000
## 353	Toyota	0	0	4300000
## 354	Daihatsu	0	0	3140000
## 355	Toyota	0	0	3460000
## 356	Toyota	0	0	4340000
## 357	Honda	0	0	3380000
## 358	Mitsubishi	0	0	8740000
## 359	Toyota	1	0	6460000
## 360	Daihatsu	0	0	4420000
## 361	Honda	1	0	3100000
## 362	Chevrolet	0	0	2830000
## 363	Nissan	0	0	4099999
## 364	Suzuki	0	0	2080000
## 365	Honda	1	0	3160000
## 366	Honda	1	0	2930000
## 367	Mitsubishi	0	0	4300000
## 368	Mitsubishi	0	0	4340000
## 369	Toyota	0	0	3020000
## 370	Toyota	0	0	3100000
## 371	Toyota	0	0	3500000
## 372	Toyota	1	0	4930000
## 373	Toyota	0	0	7950000
## 374	Toyota	0	0	5320000
## 375	Honda	1	0	2670000
## 376	Honda	0	0	2470000
## 377	Honda	1	0	3340000

## 378	Mazda	0	1	4080000
## 379	Suzuki	0	0	4099999
## 380	Honda	1	0	3500000
## 381	Toyota	0	0	4540000
## 382	Suzuki	1	0	2650000
## 383	Mazda	0	0	3870000
## 384	Toyota	0	0	5360000
## 385	Honda	1	0	4340000
## 386	Honda	0	0	3260000
## 387	Daihatsu	0	0	2630000
## 388	Honda	0	0	3240000
## 389	Ford	0	0	2470000
## 390	Mitsubishi	0	0	4260000
## 391	Nissan	0	0	2690000
## 392	Nissan	0	0	4560000
## 393	Daihatsu	0	0	4080000
## 394	Daihatsu	0	0	3610000
## 395	Daihatsu	0	0	4260000
## 396	Datsun	0	0	2140000
## 397	Honda	0	0	2930000
## 398	Honda	0	0	2870000
## 399	Mitsubishi	0	0	4260000
## 400	Mitsubishi	0	0	4790000
## 401	Nissan	0	0	2790000
## 402	Nissan	0	0	2750000
## 403	Nissan	0	0	3020000
## 404	Nissan	0	0	3180000
## 405	Suzuki	0	0	2220000
## 406	Mazda	0	0	5260000
## 407	Mitsubishi	0	0	9030000
## 408	Suzuki	0	0	2870000
## 409	Honda	0	0	3280000
## 410	Honda	0	0	3260000
## 411	Honda	0	0	3240000
## 412	Honda	0	0	3770000
## 413	Honda	0	0	3570000
## 414	Mitsubishi	0	0	4010000
## 415	Mitsubishi	0	0	4320000
## 416	Mitsubishi	0	0	4320000
## 417	Mitsubishi	0	0	4320000
## 418	Mitsubishi	0	0	4320000
## 419	Toyota	0	0	2490000
## 420	Toyota	0	0	3280000
## 421	Toyota	0	0	3690000
## 422	Toyota	0	0	3710000
## 423	Toyota	0	0	5220000
## 424	Toyota	0	0	2530000
## 425	Toyota	0	0	3480000
## 426	Toyota	0	0	6910000
## 427	Toyota	0	0	4280000
## 428	Toyota	0	0	4320000
## 429	Toyota	0	0	3550000
## 430	Toyota	0	0	3790000
## 431	Toyota	0	0	4500000

## 432	Mitsubishi	0	0	4050000
## 433	Wuling	0	0	2220000
## 434	Toyota	0	0	2440000
## 435	Mitsubishi	0	0	4280000
## 436	Suzuki	0	0	3630000
## 437	Toyota	0	0	5730000
## 438	Datsun	0	0	1590000
## 439	Mazda	0	0	3810000
## 440	Honda	0	0	4280000
## 441	Nissan	0	0	3650000
## 442	Toyota	0	0	4120000
## 443	Honda	1	0	3080000
## 444	Honda	0	0	3970000
## 445	Daihatsu	0	0	3810000
## 446	Toyota	0	0	3550000
## 447	Toyota	1	0	4810000
## 448	Nissan	0	0	4200000
## 449	Honda	0	0	2730000
## 450	Honda	1	0	3520000
## 451	Toyota	0	0	4030000
## 452	Honda	0	0	5260000
## 453	Daihatsu	0	0	3970000
## 454	Honda	1	0	2770000
## 455	Suzuki	0	0	2630000
## 456	Honda	1	0	2930000
## 457	Toyota	0	0	2610000
## 458	Suzuki	0	0	1870000
## 459	Toyota	0	0	3060000
## 460	Daihatsu	0	0	2360000
## 461	Toyota	0	0	2490000
## 462	Toyota	0	0	2890000
## 463	Honda	0	0	5750000
## 464	Toyota	0	0	7540000
## 465	Toyota	0	0	5750000
## 466	Honda	0	0	4010000
## 467	Toyota	0	0	2570000
## 468	Suzuki	0	0	3910000
## 469	Suzuki	0	0	1870000
## 470	Suzuki	0	0	3260000
## 471	Toyota	0	0	2200000
## 472	Daihatsu	0	0	2870000
## 473	Daihatsu	0	0	2670000
## 474	Wuling	0	0	4139999
## 475	Daihatsu	0	0	2360000
## 476	Daihatsu	0	0	2260000
## 477	Honda	0	0	5340000
## 478	Honda	1	0	3950000
## 479	Toyota	0	0	2300000
## 480	Suzuki	0	0	3910000
## 481	Suzuki	1	0	2570000
## 482	Toyota	0	0	2160000
## 483	Mazda	0	0	4260000
## 484	Toyota	0	0	4010000
## 485	Toyota	0	0	4520000

## 486	Honda	1	0	3340000
## 487	Suzuki	1	0	2870000
## 488	Honda	0	0	4460000
## 489	Honda	0	0	3340000
## 490	Toyota	0	0	2490000
## 491	Toyota	0	0	3730000
## 492	Mitsubishi	0	0	4220000
## 493	Honda	0	0	4050000
## 494	Toyota	0	0	4560000
## 495	Toyota	0	0	2300000
## 496	Honda	0	0	2610000
## 497	Honda	1	0	3340000
## 498	Honda	0	0	5400000
## 499	Mitsubishi	0	0	4080000
## 500	Honda	1	0	3140000
## 501	Toyota	0	0	4770000
## 502	Suzuki	0	0	3080000
## 503	Toyota	0	0	2650000
## 504	Toyota	0	0	4930000
## 505	Suzuki	0	0	3830000
## 506	Suzuki	0	0	4240000
## 507	Toyota	0	0	2610000
## 508	Honda	0	0	2570000
## 509	Honda	1	0	3260000
## 510	Nissan	0	0	3520000
## 511	Toyota	0	0	3080000
## 512	Daihatsu	0	0	2020000
## 513	Toyota	0	0	2790000
## 514	Toyota	0	0	2890000
## 515	Honda	1	0	3630000
## 516	Honda	0	0	4460000
## 517	Nissan	0	0	2970000
## 518	Toyota	0	0	2590000
## 519	Toyota	0	0	4360000
## 520	Honda	0	0	2870000
## 521	Toyota	0	0	2200000
## 522	Daihatsu	0	0	1710000
## 523	Mitsubishi	0	0	4340000
## 524	Daihatsu	0	0	3890000
## 525	Honda	0	0	2950000
## 526	Toyota	0	0	3340000
## 527	Suzuki	0	0	2730000
## 528	Toyota	0	0	2650000
## 529	Daihatsu	0	0	2530000
## 530	Daihatsu	0	0	1850000
## 531	Daihatsu	0	0	1790000
## 532	Honda	1	0	3140000
## 533	Honda	0	0	5380000
## 534	Toyota	0	0	6930000
## 535	Toyota	0	0	2440000
## 536	Honda	1	0	3180000
## 537	Daihatsu	0	0	3420000
## 538	Honda	0	0	3750000
## 539	Mitsubishi	0	0	4520000

## 540	Mitsubishi	0	0	4340000
## 541	Toyota	0	0	5580000
## 542	Nissan	0	0	2280000
## 543	Suzuki	0	0	3750000
## 544	Toyota	0	0	4970000
## 545	Toyota	0	0	3810000
## 546	Toyota	0	0	2470000
## 547	Suzuki	0	0	1750000
## 548	Daihatsu	0	0	4480000
## 549	Datsun	0	0	2280000
## 550	Honda	0	0	2280000
## 551	Nissan	0	0	1940000
## 552	Honda	1	0	4360000
## 553	Mazda	0	0	3690000
## 554	Mitsubishi	0	0	4280000
## 555	Wuling	0	0	4160000
## 556	Nissan	0	0	4260000
## 557	Mitsubishi	0	0	3610000
## 558	Suzuki	0	0	3260000
## 559	Honda	1	0	2690000
## 560	Mitsubishi	0	1	8230000
## 561	Honda	0	0	2630000
## 562	Suzuki	0	0	4910000
## 563	Toyota	0	0	3260000
## 564	Mazda	0	1	8820000
## 565	Toyota	0	0	3320000
## 566	Daihatsu	0	0	2220000
## 567	Daihatsu	0	0	2770000
## 568	Honda	0	0	2400000
## 569	Honda	1	0	2890000
## 570	Mazda	0	0	4480000
## 571	Toyota	0	0	3000000
## 572	Nissan	0	0	2670000
## 573	Nissan	0	0	4050000
## 574	Honda	0	1	7780000
## 575	Toyota	0	0	2490000
## 576	BMW	1	0	9090000
## 577	Nissan	0	0	2300000
## 578	Suzuki	0	0	2400000
## 579	Toyota	0	0	2040000
## 580	Suzuki	0	0	2950000
## 581	Daihatsu	0	0	2490000
## 582	Toyota	1	0	5660000
## 583	Nissan	0	0	2060000
## 584	Honda	0	0	2080000
## 585	Daihatsu	0	0	2380000
## 586	Mitsubishi	0	0	4520000
## 587	Toyota	0	0	2040000
## 588	Honda	1	0	2890000
## 589	Honda	1	0	3160000
## 590	Honda	1	0	3240000
## 591	Toyota	0	0	4560000
## 592	Toyota	0	0	3890000
## 593	Daihatsu	0	0	2470000

## 594	Suzuki	0	0	3910000
## 595	Honda	1	0	3360000
## 596	Toyota	0	0	4280000
## 597	Suzuki	0	0	2240000
## 598	Suzuki	0	0	2670000
## 599	Mazda	0	0	3750000
## 600	Suzuki	0	0	4440000
## 601	Nissan	0	0	4120000
## 602	Honda	0	0	3220000
## 603	Toyota	0	0	2490000
## 604	Toyota	0	0	4160000
## 605	Toyota	0	0	4770000
## 606	Toyota	0	0	2770000
## 607	Hyundai	0	0	5010000
## 608	Toyota	0	0	2360000
## 609	Daihatsu	0	0	3040000
##				car.name vehicle.stability.control
## 1			AYLA X 1.2	0
## 2			AGYA TRD SPORTIVO 1.0	0
## 3			X-TRAIL 2.5	0
## 4			YARIS S TRD 1.5	0
## 5			AGYA G 1.2	0
## 6			AVANZA G 1.3	0
## 7			AVANZA G 1.5	0
## 8			OUTLANDER SPORT PX 2.0	0
## 9			BRIO SATYA E 1.2	0
## 10			HR-V S 1.5	0
## 11			AGYA G TRD 1.2	0
## 12			AYLA R 1.2	0
## 13			AYLA R 1.2	0
## 14			AYLA R DLX 1.2	0
## 15			AGYA G TRD 1.2	0
## 16			ALMAZ LT LUX 1.5	0
## 17			TERIOS TX 1.5	0
## 18			AGYA G TRD 1.2	0
## 19			MOBILIO E 1.5	0
## 20			AYLA R 1.2	0
## 21			AYLA R 1.2	0
## 22			AYLA R 1.2	0
## 23			ERTIGA GL 1.5	0
## 24			AYLA R 1.2	0
## 25			CR-V TURBO 1.5	0
## 26			CONFERO 2WD 1.5	0
## 27			ALMAZ LT LUX 1.5	0
## 28			LIVINA VL 1.5	1
## 29			CALYA G 1.2	0
## 30			BRIO SATYA E 1.2	0
## 31			TERIOS R 1.5	1
## 32			AVANZA G 1.3	0
## 33			HR-V S 1.5	0
## 34			XPANDER ULTIMATE 1.5	0
## 35			XPANDER ULTIMATE 1.5	1
## 36			AYLA X 1.0	0
## 37			AGYA G TRD 1.2	0

## 38	AYLA X 1.2	0
## 39	CALYA G 1.2	0
## 40	HR-V E 1.5	0
## 41	OUTLANDER SPORT PX 2.0	0
## 42	OUTLANDER SPORT GLS 2.0	0
## 43	BRIO SATYA E 1.2	0
## 44	AVANZA G 1.3	0
## 45	OUTLANDER SPORT PX 2.0	0
## 46	HR-V SE 1.5	0
## 47	ALMAZ S+T 1.5	0
## 48	IGNIS GL 1.2	0
## 49	ALMAZ S+T 1.5	0
## 50	AYLA R 1.2	0
## 51	TERIOS R 1.5	0
## 52	XPANDER ULTIMATE 1.5	0
## 53	CR-V TURBO 1.5	0
## 54	HR-V SE 1.5	0
## 55	KIJANG INNOVA REBORN G 2.0	0
## 56	MOBILIO RS 1.5	0
## 57	KIJANG INNOVA V 2.0	0
## 58	AGYA G TRD 1.2	0
## 59	KIJANG INNOVA G 2.0	0
## 60	IGNIS GL 1.2	0
## 61	AGYA GR SPORT 1.2	0
## 62	AYLA R 1.2	0
## 63	MOBILIO E 1.5	0
## 64	AYLA R 1.2	0
## 65	TERIOS X DLX 1.5	0
## 66	BRIO SATYA E 1.2	0
## 67	HR-V SE 1.5	0
## 68	BRIO SATYA E 1.2	0
## 69	BRIO RS 1.2	0
## 70	FORTUNER VRZ 2.4	0
## 71	CX-5 ELITE 2.5	0
## 72	GRAND LIVINA XV 1.5	0
## 73	ERTIGA GL 1.5	0
## 74	AGYA G 1.0	0
## 75	AGYA G TRD 1.2	0
## 76	IGNIS GX 1.2	0
## 77	HR-V S 1.5	1
## 78	CORTEZ C T LUX 1.5	1
## 79	ALMAZ RS LT LUX + SC CVT 1.5	0
## 80	CONFERO S 1.5	0
## 81	AYLA X 1.0	0
## 82	ALMAZ L TURBO 1.5	0
## 83	AYLA R 1.2	0
## 84	KIJANG INNOVA REBORN VENTURER GASOLINE 2.0	0
## 85	AGYA G 1.0	0
## 86	BALENO HATCHBACK 1.4	0
## 87	BALENO HATCHBACK 1.4	0
## 88	AYLA R DLX 1.2	0
## 89	XPANDER ULTIMATE 1.5	1
## 90	XPANDER ULTIMATE 1.5	1
## 91	AYLA X 1.0	0

## 92	AYLA R 1.2	0
## 93	ALMAZ LT LUX 1.5	0
## 94	TERIOS X 1.5	0
## 95	ALMAZ LT LUX + SC CVT 1.5	0
## 96	AVANZA G 1.5	1
## 97	BRIO SATYA E 1.2	0
## 98	CITY RS HATCHBACK 1.5	0
## 99	HR-V E 1.5	0
## 100	AYLA X ELEGANT 1.0	0
## 101	KIJANG INNOVA G 2.0	1
## 102	AYLA R DLX 1.2	0
## 103	AVANZA E 1.3	0
## 104	RUSH S TRD SPORTIVO 1.5	1
## 105	MOBILIO E 1.5	0
## 106	XENIA R STD 1.3	0
## 107	CORTEZ S T LUX 1.5	0
## 108	CORTEZ L T LUX 1.5	0
## 109	ALMAZ LT LUX 1.5	0
## 110	HR-V E 1.5	0
## 111	FORTUNER VRZ 2.4	0
## 112	PAJERO SPORT DAKAR 4X2 2.4	0
## 113	CITY RS HATCHBACK 1.5	0
## 114	XPANDER ULTIMATE 1.5	1
## 115	XL7 ALPHA 1.5	1
## 116	AYLA X 1.0	0
## 117	BRIO SATYA E 1.2	0
## 118	MOBILIO E 1.5	0
## 119	YARIS G 1.5	1
## 120	RUSH S TRD SPORTIVO 1.5	0
## 121	MOBILIO E 1.5	0
## 122	RUSH S 1.5	1
## 123	HR-V E 1.5	0
## 124	XENIA X 1.3	0
## 125	RUSH S GR SPORT 1.5	0
## 126	CR-V 2.0	0
## 127	BRIO SATYA E 1.2	0
## 128	HR-V SE 1.5	0
## 129	AYLA R 1.2	0
## 130	CR-V TURBO 1.5	0
## 131	YARIS S TRD 1.5	0
## 132	ROCKY R 1.0	1
## 133	ERTIGA GL 1.4	0
## 134	KIJANG INNOVA REBORN V 2.0	0
## 135	BRIO SATYA E 1.2	0
## 136	AYLA R 1.2	0
## 137	RUSH S TRD SPORTIVO 1.5	1
## 138	YARIS S TRD 1.5	0
## 139	ALMAZ LT LUX CVT 1.5	0
## 140	AYLA R DLX 1.2	0
## 141	AGYA G 1.0	0
## 142	AYLA X 1.0	0
## 143	AYLA X 1.0	0
## 144	CALYA G 1.2	0
## 145	RUSH S TRD SPORTIVO 1.5	1

## 146	CORTEZ T LUX 1.5	0
## 147	XPANDER EXCEED 1.5	0
## 148	XPANDER ULTIMATE 1.5	1
## 149	IGNIS GL 1.2	0
## 150	AGYA G TRD 1.2	0
## 151	BALENO HATCHBACK 1.4	0
## 152	AGYA G TRD 1.0	0
## 153	AYLA R 1.2	0
## 154	ECLIPSE CROSS ULTIMATE 1.5	0
## 155	BRIO SATYA E 1.2	0
## 156	AGYA GR SPORT 1.2	0
## 157	AYLA R 1.2	0
## 158	RUSH S TRD 1.5	1
## 159	TERIOS R 1.5	1
## 160	TERIOS X DLX 1.5	0
## 161	HR-V E 1.5	0
## 162	HR-V E 1.5	0
## 163	HR-V E PLUS 1.5	0
## 164	2 GT SKYACTIV 1.5	0
## 165	RUSH S TRD SPORTIVO 1.5	0
## 166	KIJANG INNOVA REBORN G 2.0	0
## 167	AGYA G TRD 1.2	0
## 168	AGYA G TRD 1.2	0
## 169	X-TRAIL 2.0	0
## 170	HR-V E 1.5	0
## 171	RUSH S TRD SPORTIVO 1.5	0
## 172	BRIO RS 1.2	0
## 173	BRIO RS 1.2	0
## 174	ERTIGA GX 1.4	0
## 175	ROCKY R 1.0	0
## 176	RUSH S TRD SPORTIVO 1.5	1
## 177	SIENTA Q 1.5	0
## 178	CONFERO S L LUX 1.5	0
## 179	SX4 S-CROSS 1.5	0
## 180	XPANDER ULTIMATE 1.5	1
## 181	ERTIGA SPORT GT 1.5	1
## 182	IGNIS GX 1.2	0
## 183	SIGRA R DLX 1.2	0
## 184	SIGRA D 1.0	0
## 185	2 R SKYACTIV 1.5	1
## 186	TERIOS X DLX 1.5	0
## 187	HR-V E PLUS 1.5	0
## 188	HR-V E PLUS 1.5	0
## 189	XPANDER SPORT 1.5	0
## 190	BR-V E 1.5	0
## 191	CITY E 1.5	0
## 192	KIJANG INNOVA REBORN G 2.0	0
## 193	AYLA R 1.2	0
## 194	BRIO RS 1.2	0
## 195	BRIO SATYA E 1.2	0
## 196	HR-V E PLUS 1.5	0
## 197	CITY RS HATCHBACK 1.5	0
## 198	HR-V E 1.5	0
## 199	2 R SKYACTIV 1.5	1

## 200	2 R SKYACTIV 1.5	1
## 201	AGYA G 1.2	0
## 202	XPANDER ULTIMATE 1.5	0
## 203	XPANDER ULTIMATE 1.5	0
## 204	LIVINA VL 1.5	0
## 205	RUSH S TRD SPORTIVO 1.5	1
## 206	RUSH S TRD SPORTIVO 1.5	1
## 207	TERIOS R DLX 1.5	0
## 208	XENIA R DLX 1.3	0
## 209	BR-V E PRESTIGE 1.5	0
## 210	KIJANG INNOVA G LUX 2.0	0
## 211	GRAND LIVINA XV 1.5	0
## 212	SIGRA R DLX 1.2	0
## 213	HR-V E 1.5	0
## 214	MOBILIO RS 1.5	0
## 215	CALYA G 1.2	0
## 216	AYLA R DLX 1.2	0
## 217	SIGRA R STD 1.2	0
## 218	RUSH G 1.5	0
## 219	ERTIGA GL 1.4	0
## 220	GO+ PANCA T 1.2	0
## 221	CALYA G 1.2	0
## 222	ROCKY R 1.0	1
## 223	XENIA X STD 1.3	0
## 224	RAIZE TURBO G 1.0	0
## 225	KIJANG INNOVA Q 2.0	0
## 226	ERTIGA GX 1.4	0
## 227	AVANZA VELOZ 1.3	0
## 228	MOBILIO E 1.5	0
## 229	XPANDER SPORT 1.5	0
## 230	XPANDER ULTIMATE 1.5	1
## 231	CX-3 GT 2.0	0
## 232	BRIO E 1.2	0
## 233	PAJERO SPORT DAKAR ULTIMATE 4X2 2.4	0
## 234	KIJANG INNOVA REBORN V 2.0	0
## 235	ERTIGA SPORT GT 1.5	0
## 236	AGYA G TRD SPORTIVO 1.2	0
## 237	XL7 ALPHA 1.5	1
## 238	BALENO HATCHBACK 1.4	0
## 239	CITY RS HATCHBACK 1.5	0
## 240	KIJANG INNOVA REBORN G 2.0	0
## 241	XPANDER ULTIMATE 1.5	0
## 242	ALMAZ S+T SMART ENJOY 1.5	0
## 243	IGNIS GX AGS 1.2	0
## 244	CRETA PRIME 1.5	0
## 245	XPANDER EXCEED 1.5	0
## 246	RAIZE GR SPORT 1.0	0
## 247	KIJANG INNOVA V 2.0	0
## 248	ERTIGA GX 1.4	0
## 249	ERTIGA DREZA 1.4	0
## 250	BALENO HATCHBACK 1.4	0
## 251	BRIO RS 1.2	0
## 252	BRIO SATYA E 1.2	0
## 253	BR-V E 1.5	0

## 254	SX4 S-CROSS 1.5	0
## 255	MOBILIO RS 1.5	0
## 256	SANTA FE CRDI 2.2	0
## 257	XPANDER ULTIMATE 1.5	1
## 258	X-TRAIL 2.5	0
## 259	BRIO SATYA E 1.2	0
## 260	BRIO SATYA E 1.2	0
## 261	BRIO SATYA E 1.2	0
## 262	CX-5 GT 2.5	0
## 263	AGYA G 1.0	0
## 264	MOBILIO RS 1.5	0
## 265	BRIO SATYA E 1.2	0
## 266	XL7 BETA 1.5	1
## 267	XPANDER ULTIMATE 1.5	1
## 268	ERTIGA SPORT GT 1.5	0
## 269	BRIO RS 1.2	0
## 270	XPANDER ULTIMATE 1.5	1
## 271	CALYA G 1.2	0
## 272	KIJANG INNOVA V 2.0	0
## 273	CORTEZ S T LUX 1.5	0
## 274	BRIO SATYA E 1.2	0
## 275	CX-5 ELITE 2.5	0
## 276	SIENTA Q 1.5	1
## 277	BRIO SATYA E 1.2	0
## 278	KIJANG INNOVA VENTURER 2.0	0
## 279	CONFERO S L 1.5	0
## 280	XPANDER EXCEED 1.5	0
## 281	KIJANG INNOVA REBORN VENTURER GASOLINE 2.0	0
## 282	YARIS S TRD 1.5	0
## 283	BRIO SATYA E 1.2	0
## 284	XPANDER SPORT 1.5	1
## 285	ERTIGA GX 1.5	0
## 286	KIJANG INNOVA REBORN G 2.0	0
## 287	SIENTA V 1.5	0
## 288	BRIO RS 1.2	0
## 289	BRIO SATYA E 1.2	0
## 290	BR-V PRESTIGE 1.5	1
## 291	CITY RS HATCHBACK 1.5	1
## 292	HR-V E PLUS 1.5	0
## 293	IGNIS GX 1.2	0
## 294	X-TRAIL 2.5	0
## 295	ERTIGA DREZA GS 1.4	0
## 296	BIANTE SKYACTIV 2.0	0
## 297	GRAND LIVINA SV 1.5	0
## 298	CX-3 TOURING 2.0	0
## 299	XPANDER ULTIMATE 1.5	0
## 300	MARCH 1.2	0
## 301	XENIA R 1.3	0
## 302	XPANDER SPORT 1.5	0
## 303	ERTIGA GX 1.5	0
## 304	XENIA R SPORTY 1.3	0
## 305	AVANZA G 1.3	0
## 306	BRIO SATYA E 1.2	0
## 307	CONFERO S 1.5	0

## 308	SIENTA G 1.5	0
## 309	BRIO SATYA E 1.2	0
## 310	BRIO SATYA E 1.2	0
## 311	BR-V E 1.5	0
## 312	MOBILIO E PRESTIGE 1.5	0
## 313	SIENTA G 1.5	0
## 314	XPANDER ULTIMATE 1.5	0
## 315	AYLA R 1.2	0
## 316	MARCH 1.5	1
## 317	RUSH S TRD SPORTIVO 1.5	1
## 318	MARCH 1.2	0
## 319	FORTUNER VRZ GR SPORT 2.8	0
## 320	KIJANG INNOVA REBORN V 2.0	0
## 321	XPANDER EXCEED 1.5	0
## 322	AVANZA G 1.3	0
## 323	KIJANG INNOVA G 2.0	0
## 324	MOBILIO E 1.5	0
## 325	S-PRESSO 1.0	1
## 326	KIJANG INNOVA G 2.0	0
## 327	IGNIS GX 1.2	0
## 328	XPANDER ULTIMATE 1.5	0
## 329	TERIOS X DLX 1.5	0
## 330	AGYA G TRD 1.2	0
## 331	BRIO SATYA E 1.2	0
## 332	XPANDER SPORT 1.5	0
## 333	XPANDER ULTIMATE 1.5	0
## 334	XPANDER ULTIMATE 1.5	1
## 335	XPANDER ULTIMATE 1.5	0
## 336	XPANDER ULTIMATE 1.5	0
## 337	KIJANG INNOVA Q 2.0	0
## 338	BALENO HATCHBACK 1.4	0
## 339	KIJANG INNOVA REBORN V 2.0	0
## 340	BRIO RS 1.2	0
## 341	SIENTA V 1.5	0
## 342	XPANDER EXCEED 1.5	0
## 343	XPANDER ULTIMATE 1.5	1
## 344	AGYA G TRD 1.2	0
## 345	SERENA HIGHWAY STAR 2.0	0
## 346	RUSH S TRD SPORTIVO 1.5	0
## 347	XENIA R DLX 1.3	0
## 348	MARCH 1.2	0
## 349	MOBILIO E 1.5	0
## 350	FORTUNER VRZ 4X2 2.4	0
## 351	BRIO RS 1.2	0
## 352	BRIO RS 1.2	0
## 353	YARIS S TRD 1.5	0
## 354	TERIOS ADVENTURE R 1.5	0
## 355	AVANZA VELOZ 1.3	0
## 356	RUSH S TRD SPORTIVO 1.5	1
## 357	BRIO SATYA E 1.2	0
## 358	PAJERO SPORT DAKAR 4X2 2.4	0
## 359	KIJANG INNOVA Q 2.0	0
## 360	TERIOS R DLX 1.5	0
## 361	BRIO RS 1.2	0

## 362	SPARK PREMIER 1.4	1
## 363	LIVINA VL 1.5	1
## 364	CARRY PICK UP 1.5	0
## 365	BRIO SATYA E 1.2	0
## 366	BRIO SATYA E 1.2	0
## 367	XPANDER ULTIMATE 1.5	0
## 368	XPANDER ULTIMATE 1.5	1
## 369	AGYA G TRD SPORTIVO 1.2	0
## 370	AVANZA G 1.3	0
## 371	AVANZA E 1.3	0
## 372	KIJANG INNOVA G 2.0	0
## 373	FORTUNER VRZ 4X2 2.4	0
## 374	KIJANG INNOVA REBORN G 2.0	0
## 375	BRIO SATYA E 1.2	0
## 376	BRIO E 1.2	0
## 377	BRIO SATYA E 1.2	0
## 378	CX-5 HIGH 2.5	0
## 379	ERTIGA SPORT GT 1.5	1
## 380	BRIO SATYA E 1.2	0
## 381	RAIZE GR 1.0	0
## 382	IGNIS GL 1.2	0
## 383	2 R SKYACTIV 1.5	0
## 384	KIJANG INNOVA G 2.0	0
## 385	BR-V PRESTIGE 1.5	0
## 386	BRIO SATYA E 1.2	0
## 387	SIGRA R STD 1.2	0
## 388	BRIO SATYA E 1.2	0
## 389	ECOSPORT TITANIUM 1.5	0
## 390	XPANDER ULTIMATE 1.5	0
## 391	GRAND LIVINA HIGHWAY STAR 1.5	0
## 392	SERENA HIGHWAY STAR 2.0	0
## 393	ROCKY X 1.2	0
## 394	SIRION 1.3	0
## 395	TERIOS R 1.5	0
## 396	GO PANCA T LIVE 1.2	0
## 397	MOBILIO E PRESTIGE 1.5	0
## 398	BRIO RS 1.2	0
## 399	XPANDER ULTIMATE 1.5	0
## 400	XPANDER CROSS 1.5	0
## 401	GRAND LIVINA HIGHWAY STAR AUTECH 1.5	0
## 402	GRAND LIVINA X-GEAR 1.5	0
## 403	GRAND LIVINA HIGHWAY STAR AUTECH 1.5	0
## 404	GRAND LIVINA HIGHWAY STAR AUTECH 1.5	0
## 405	CARRY PICK UP 1.5	0
## 406	CX-5 GT 2.5	0
## 407	PAJERO SPORT DAKAR ULTIMATE 4X2 2.4	0
## 408	ERTIGA GL 1.4	0
## 409	BRIO SATYA E 1.2	0
## 410	BRIO SATYA E 1.2	0
## 411	BRIO SATYA E 1.2	0
## 412	BRIO RS 1.2	0
## 413	BR-V E 1.5	0
## 414	XPANDER EXCEED 1.5	0
## 415	XPANDER ULTIMATE 1.5	0

## 416	XPANDER ULTIMATE 1.5	0
## 417	XPANDER ULTIMATE 1.5	0
## 418	XPANDER ULTIMATE 1.5	0
## 419	AVANZA G 1.3	0
## 420	AVANZA G 1.3	0
## 421	AVANZA G 1.3	0
## 422	AVANZA G 1.3	0
## 423	KIJANG INNOVA V 2.0	0
## 424	CALYA G 1.2	0
## 425	KIJANG INNOVA G 2.0	0
## 426	KIJANG INNOVA REBORN G 2.4	0
## 427	RAIZE TURBO G 1.0	0
## 428	RUSH S TRD SPORTIVO 1.5	0
## 429	SIENTA V 1.5	0
## 430	SIENTA V 1.5	0
## 431	YARIS S TRD 1.5	0
## 432	XPANDER SPORT 1.5	0
## 433	CONFERO S 1.5	0
## 434	CALYA E 1.2	0
## 435	XPANDER ULTIMATE 1.5	0
## 436	ERTIGA GX 1.5	0
## 437	KIJANG INNOVA REBORN G 2.0	0
## 438	GO+ PANCA T 1.2	0
## 439	2 GT 1.5	0
## 440	JAZZ RS 1.5	0
## 441	LIVINA VL 1.5	0
## 442	YARIS S TRD 1.5	0
## 443	BRIO SATYA E 1.2	0
## 444	HR-V E 1.5	0
## 445	TERIOS R 1.5	1
## 446	AVANZA G 1.3	0
## 447	KIJANG INNOVA REBORN G 2.0	0
## 448	LIVINA VL 1.5	0
## 449	BRIO RS 1.2	0
## 450	MOBILIO RS 1.5	0
## 451	RUSH S TRD SPORTIVO 1.5	0
## 452	HR-V SE 1.5	0
## 453	TERIOS R 1.5	0
## 454	BRIO SATYA E 1.2	0
## 455	ERTIGA GX 1.4	0
## 456	MOBILIO E 1.5	0
## 457	AGYA G TRD 1.2	0
## 458	CARRY PICK UP 1.5	0
## 459	RUSH S 1.5	0
## 460	AYLA R 1.2	0
## 461	AGYA G TRD 1.2	0
## 462	CALYA G 1.2	0
## 463	HR-V E 1.5	0
## 464	KIJANG INNOVA V 2.0	0
## 465	FORTUNER G TRD 2.5	0
## 466	HR-V S 1.5	1
## 467	CALYA G 1.2	0
## 468	SX4 S-CROSS 1.5	0
## 469	KARIMUN GS 1.0	0

## 470	ERTIGA GX 1.4	0
## 471	CALYA G 1.2	0
## 472	TERIOS TX 1.5	0
## 473	SIGRA X 1.2	0
## 474	ALMAZ LT LUX 1.5	0
## 475	XENIA R 1.3	0
## 476	AYLA R 1.2	0
## 477	CITY RS HATCHBACK 1.5	0
## 478	BR-V E 1.5	0
## 479	CALYA G 1.2	0
## 480	ERTIGA SPORT GT 1.5	1
## 481	IGNIS GL 1.2	0
## 482	CALYA G 1.2	0
## 483	2 GT SKYACTIV 1.5	0
## 484	AVANZA VELOZ 1.5	0
## 485	AVANZA VELOZ 1.5	0
## 486	BRIO SATYA E 1.2	0
## 487	IGNIS GX 1.2	0
## 488	HR-V E 1.5	0
## 489	MOBILIO S 1.5	0
## 490	CALYA G 1.2	0
## 491	YARIS S TRD 1.5	1
## 492	XPANDER ULTIMATE 1.5	1
## 493	CIVIC FB2 1.8	0
## 494	RAIZE GR TWO TONE 1.0	0
## 495	CALYA G 1.2	0
## 496	BRIO SATYA E 1.2	0
## 497	MOBILIO E 1.5	0
## 498	HR-V E 1.5	0
## 499	XPANDER SPORT 1.5	1
## 500	BRIO SATYA E 1.2	0
## 501	RUSH S TRD SPORTIVO 1.5	0
## 502	ERTIGA DREZA 1.4	0
## 503	AGYA G TRD 1.2	0
## 504	RUSH S GR SPORT 1.5	1
## 505	XL7 ZETA GL 1.5	0
## 506	XL7 BETA 1.5	1
## 507	AGYA G 1.2	0
## 508	BRIO SATYA E 1.2	0
## 509	BRIO SATYA E 1.2	0
## 510	X-TRAIL URBAN SELECTION 2.5	0
## 511	CALYA G 1.2	0
## 512	AYLA X 1.0	0
## 513	CALYA G 1.2	0
## 514	CALYA G 1.2	0
## 515	BRIO RS 1.2	0
## 516	HR-V E 1.5	0
## 517	JUKE 1.5	0
## 518	AGYA G TRD 1.2	0
## 519	RUSH S TRD SPORTIVO 1.5	1
## 520	BRIO SATYA E 1.2	0
## 521	CALYA G 1.2	0
## 522	AYLA X 1.0	0
## 523	XPANDER ULTIMATE 1.5	0

## 524	ROCKY X 1.2	1
## 525	MOBILIO S 1.5	0
## 526	AVANZA VELOZ 1.5	0
## 527	IGNIS GX 1.2	0
## 528	AVANZA E 1.3	0
## 529	SIRION M602RS 1.3	0
## 530	AYLA X 1.0	0
## 531	SIRION M602RS 1.3	0
## 532	MOBILIO E 1.5	0
## 533	HR-V E 1.5	0
## 534	KIJANG INNOVA REBORN G 2.4	0
## 535	CALYA G 1.2	0
## 536	BRIO SATYA E 1.2	0
## 537	XENIA R STD 1.3	0
## 538	BR-V E 1.5	0
## 539	XPANDER ULTIMATE 1.5	0
## 540	XPANDER ULTIMATE 1.5	0
## 541	FORTUNER G 2.5	0
## 542	MARCH L 1.2	0
## 543	SX4 S-CROSS 1.5	0
## 544	YARIS S TRD 1.5	0
## 545	SIENTA Q 1.5	0
## 546	CALYA G 1.2	0
## 547	KARIMUN GL 1.0	0
## 548	ROCKY R 1.0	0
## 549	CROSS 1.2	1
## 550	BRIO SATYA S 1.2	0
## 551	GRAND LIVINA SV 1.5	0
## 552	BR-V PRESTIGE 1.5	0
## 553	2 GT SKYACTIV 1.5	0
## 554	XPANDER ULTIMATE 1.5	1
## 555	ALMAZ LT LUX 1.5	0
## 556	X-TRAIL 2.5	0
## 557	COLT L300 PICK UP 2.5	0
## 558	SX4 S-CROSS 1.5	0
## 559	BRIO SATYA E 1.2	0
## 560	PAJERO SPORT DAKAR 4X2 2.4	0
## 561	FREED E 1.5	0
## 562	SX4 S-CROSS 1.5	0
## 563	AVANZA G 1.3	0
## 564	CX-5 ELITE 2.5	0
## 565	SIENTA G 1.5	0
## 566	SIGRA M 1.0	0
## 567	SIGRA R 1.2	0
## 568	BRIO SATYA E 1.2	0
## 569	BRIO SATYA E 1.2	0
## 570	2 R 1.5	0
## 571	AGYA G TRD SPORTIVO 1.2	0
## 572	GRAND LIVINA HIGHWAY STAR 1.5	0
## 573	X-TRAIL 2.0	0
## 574	CR-V TURBO PRESTIGE 1.5	0
## 575	AVANZA G 1.3	0
## 576	3 20I (CKD) 2.0	0
## 577	MARCH 1.5	0

## 578	CARRY PICK UP 1.5	0
## 579	AGYA G 1.0	0
## 580	IGNIS GX 1.2	0
## 581	SIGRA R STD 1.2	0
## 582	KIJANG INNOVA REBORN G 2.0	0
## 583	MARCH 1.2	0
## 584	BRIO E 1.3	0
## 585	SIGRA R STD 1.2	0
## 586	XPANDER ULTIMATE 1.5	0
## 587	AVANZA G 1.3	0
## 588	BRIO RS 1.2	0
## 589	BRIO RS 1.2	0
## 590	MOBILIO E 1.5	0
## 591	RUSH S TRD SPORTIVO 1.5	0
## 592	SIENTA Q 1.5	0
## 593	SIGRA R DLX 1.2	0
## 594	XL7 ZETA GL 1.5	0
## 595	BRIO SATYA E 1.2	0
## 596	RUSH S TRD SPORTIVO 1.5	1
## 597	CARRY PICK UP 1.5	0
## 598	IGNIS GX 1.2	0
## 599	2 GT 1.5	0
## 600	XL7 ALPHA 1.5	0
## 601	LIVINA VL 1.5	1
## 602	FREED S 1.5	0
## 603	AGYA G 1.2	0
## 604	AVANZA VELOZ 1.5	0
## 605	RAIZE GR SPORT TSS 1.0	0
## 606	AGYA G TRD 1.2	0
## 607	STARGAZER PRIME 1.5	0
## 608	CALYA G 1.2	0
## 609	XENIA R SPORTY 1.3	0
##	mileage_km auto.cruise.control keyless.push.start map.navigator	
## 1	10.508 0 0 0	
## 2	112.888 0 0 0	
## 3	118.429 0 1 0	
## 4	15.945 0 0 0	
## 5	30.404 0 0 0	
## 6	17.306 0 0 0	
## 7	12.211 0 0 0	
## 8	126.885 0 0 0	
## 9	33.464 0 0 0	
## 10	33.364 0 0 0	
## 11	44.969 0 0 0	
## 12	43.292 0 0 0	
## 13	57.385 0 0 0	
## 14	53.151 0 0 0	
## 15	46.182 0 0 0	
## 16	58.152 0 0 0	
## 17	118.704 0 0 1	
## 18	78.385 0 0 0	
## 19	45.376 0 0 0	
## 20	37.343 0 0 0	
## 21	61.937 0 0 0	

## 22	47.085	0	0	0
## 23	178.000	0	0	0
## 24	10.301	0	0	0
## 25	75.302	0	1	0
## 26	18.298	0	0	0
## 27	50.387	0	0	0
## 28	64.235	0	0	0
## 29	44.473	0	0	0
## 30	24.577	0	0	0
## 31	42.320	0	0	0
## 32	81.158	0	0	0
## 33	64.753	0	0	0
## 34	72.887	1	0	0
## 35	77.425	0	0	0
## 36	63.175	0	0	0
## 37	43.689	0	0	0
## 38	85.583	0	0	0
## 39	54.512	0	0	0
## 40	35.177	0	1	0
## 41	57.230	0	0	0
## 42	133.127	0	0	0
## 43	34.815	0	0	0
## 44	98.478	0	0	0
## 45	91.926	0	0	0
## 46	62.209	0	0	0
## 47	76.950	0	0	0
## 48	43.945	0	0	0
## 49	27.804	0	0	0
## 50	85.184	0	0	0
## 51	40.717	0	1	0
## 52	63.130	1	0	0
## 53	64.558	0	0	0
## 54	43.740	0	0	0
## 55	51.254	0	0	0
## 56	84.060	0	0	0
## 57	87.616	0	0	0
## 58	53.576	0	0	0
## 59	74.730	0	0	0
## 60	79.289	0	0	0
## 61	5.751	0	0	0
## 62	15.552	0	0	0
## 63	84.575	0	0	0
## 64	8.691	0	0	0
## 65	29.897	0	0	0
## 66	15.855	0	0	0
## 67	83.286	0	0	0
## 68	36.455	0	0	0
## 69	46.441	0	0	0
## 70	94.919	0	0	0
## 71	45.271	0	0	0
## 72	98.416	0	0	0
## 73	49.988	0	0	0
## 74	78.175	0	0	0
## 75	56.303	0	0	0

## 76	27.044	0	1	0
## 77	60.978	0	0	0
## 78	55.534	0	0	0
## 79	38.223	0	0	0
## 80	32.630	0	0	0
## 81	73.528	0	0	0
## 82	43.439	0	0	0
## 83	25.400	0	0	0
## 84	89.898	0	0	0
## 85	95.007	0	0	0
## 86	72.458	0	1	0
## 87	41.091	0	1	0
## 88	21.880	0	0	0
## 89	66.471	0	0	0
## 90	57.815	0	0	0
## 91	28.113	0	0	0
## 92	54.801	0	0	0
## 93	41.998	0	0	0
## 94	115.866	0	0	0
## 95	64.335	0	0	0
## 96	7.835	0	0	0
## 97	42.586	0	0	0
## 98	18.936	0	0	0
## 99	122.090	0	0	0
## 100	34.014	0	0	0
## 101	25.226	0	0	0
## 102	6.011	0	0	0
## 103	70.052	0	0	0
## 104	53.096	0	0	0
## 105	66.129	0	0	0
## 106	34.359	0	0	0
## 107	42.302	0	0	0
## 108	32.320	0	0	0
## 109	55.346	0	0	0
## 110	85.936	0	0	0
## 111	112.148	0	0	0
## 112	95.178	0	0	0
## 113	25.160	1	0	0
## 114	53.266	0	0	0
## 115	51.694	0	0	0
## 116	56.960	0	0	0
## 117	48.434	0	0	0
## 118	52.561	0	0	0
## 119	65.003	0	0	0
## 120	74.934	0	1	0
## 121	82.350	0	0	0
## 122	86.251	0	0	0
## 123	41.320	0	0	0
## 124	36.205	0	0	0
## 125	58.598	0	0	0
## 126	104.079	0	0	0
## 127	81.726	0	0	0
## 128	95.389	0	0	0
## 129	57.138	0	0	0

## 130	117.511	0	0	1
## 131	30.898	0	0	0
## 132	21.304	0	0	0
## 133	132.270	0	0	0
## 134	67.327	0	0	0
## 135	55.240	0	0	0
## 136	17.916	0	0	0
## 137	69.785	0	0	0
## 138	39.191	0	0	0
## 139	9.775	0	0	0
## 140	33.635	0	0	0
## 141	76.931	0	0	0
## 142	93.445	0	0	0
## 143	86.462	0	0	0
## 144	40.358	0	0	0
## 145	47.810	0	0	0
## 146	72.650	0	0	0
## 147	34.947	0	0	0
## 148	98.742	0	0	0
## 149	26.233	0	0	0
## 150	55.010	0	0	0
## 151	24.032	0	1	0
## 152	10.656	0	0	0
## 153	78.944	0	0	0
## 154	42.516	0	0	0
## 155	64.684	0	0	0
## 156	22.803	0	1	0
## 157	42.587	0	0	0
## 158	67.595	0	0	0
## 159	28.869	0	0	0
## 160	70.234	0	0	0
## 161	65.219	0	0	0
## 162	74.516	0	0	0
## 163	71.135	1	0	0
## 164	94.312	0	0	0
## 165	66.029	0	1	0
## 166	103.601	0	0	0
## 167	30.738	0	0	0
## 168	45.824	0	0	0
## 169	117.889	0	0	0
## 170	54.437	0	0	0
## 171	51.860	0	0	0
## 172	28.884	0	0	0
## 173	21.701	0	0	0
## 174	110.889	0	0	0
## 175	22.709	0	0	0
## 176	51.410	0	0	0
## 177	64.808	0	0	0
## 178	12.790	0	0	0
## 179	96.199	0	0	0
## 180	67.803	0	0	0
## 181	78.472	0	0	0
## 182	103.106	0	1	0
## 183	99.317	0	0	0

## 184	29.119	0	0	0
## 185	83.081	0	0	0
## 186	41.027	0	0	0
## 187	73.712	0	0	0
## 188	27.308	0	0	0
## 189	70.969	0	1	0
## 190	119.436	0	0	0
## 191	61.556	0	0	0
## 192	74.520	0	0	0
## 193	81.739	0	0	0
## 194	83.538	0	0	0
## 195	52.723	0	0	0
## 196	94.946	0	0	0
## 197	22.670	0	0	0
## 198	68.990	0	0	0
## 199	126.721	0	0	0
## 200	117.682	0	0	0
## 201	9.881	0	0	0
## 202	61.532	0	0	0
## 203	47.180	0	0	0
## 204	83.691	0	0	0
## 205	76.928	0	0	0
## 206	57.056	0	0	0
## 207	72.957	0	0	0
## 208	140.903	0	0	0
## 209	85.650	0	0	0
## 210	169.237	0	0	0
## 211	142.175	0	0	0
## 212	122.574	0	0	0
## 213	92.961	0	0	0
## 214	73.033	0	0	0
## 215	59.581	0	0	0
## 216	80.201	0	0	0
## 217	140.428	0	0	0
## 218	89.405	0	0	0
## 219	77.093	0	0	0
## 220	46.497	0	0	0
## 221	169.351	0	0	0
## 222	24.619	0	0	0
## 223	44.746	0	0	0
## 224	19.489	0	0	0
## 225	53.528	0	0	0
## 226	79.198	0	0	0
## 227	172.865	0	0	0
## 228	59.147	0	0	0
## 229	58.752	0	0	0
## 230	93.142	0	0	0
## 231	32.130	0	1	0
## 232	48.751	0	0	0
## 233	73.296	0	1	0
## 234	34.960	0	0	0
## 235	18.426	0	0	0
## 236	36.037	0	1	0
## 237	26.534	0	0	0

## 238	79.665	0	1	0
## 239	46.694	0	0	0
## 240	80.809	0	0	0
## 241	82.596	0	0	0
## 242	39.239	0	0	0
## 243	125.114	0	1	0
## 244	28.896	0	1	0
## 245	51.387	0	0	0
## 246	18.599	0	0	0
## 247	104.055	0	0	0
## 248	28.466	0	0	0
## 249	77.648	0	0	0
## 250	44.979	0	1	0
## 251	61.782	0	0	0
## 252	46.912	0	0	0
## 253	109.995	0	0	0
## 254	91.268	0	0	0
## 255	120.026	0	0	0
## 256	52.497	0	0	0
## 257	51.925	0	0	0
## 258	92.536	0	0	1
## 259	24.658	0	0	0
## 260	55.430	0	0	0
## 261	40.516	0	0	0
## 262	16.763	0	0	0
## 263	48.952	0	0	0
## 264	70.213	0	0	0
## 265	22.059	0	0	0
## 266	42.044	0	0	0
## 267	68.413	0	0	0
## 268	25.127	0	1	0
## 269	55.434	0	0	0
## 270	50.764	0	0	0
## 271	90.659	0	0	0
## 272	158.669	0	0	0
## 273	26.527	0	1	0
## 274	22.928	0	0	0
## 275	44.889	0	0	0
## 276	112.207	0	0	0
## 277	35.961	0	0	0
## 278	97.032	0	0	0
## 279	67.690	0	0	0
## 280	81.128	0	0	0
## 281	79.013	0	1	0
## 282	61.487	0	0	0
## 283	97.128	0	0	0
## 284	80.941	0	0	0
## 285	78.531	0	0	0
## 286	89.457	0	0	0
## 287	111.609	0	1	0
## 288	50.261	0	0	0
## 289	43.337	0	0	0
## 290	41.412	0	0	0
## 291	30.492	0	0	0

## 292	42.850	0	0	0
## 293	88.669	0	1	0
## 294	110.956	0	0	0
## 295	95.539	0	0	0
## 296	48.097	0	0	0
## 297	104.054	0	0	0
## 298	73.975	0	0	0
## 299	60.960	1	0	0
## 300	57.201	0	0	0
## 301	109.214	0	0	0
## 302	58.277	0	1	0
## 303	35.385	0	1	0
## 304	70.926	0	0	1
## 305	78.434	0	0	0
## 306	79.325	0	0	0
## 307	91.501	0	0	0
## 308	35.480	0	0	0
## 309	81.964	0	0	0
## 310	34.001	0	0	0
## 311	21.707	0	0	0
## 312	96.773	0	0	0
## 313	94.182	0	0	0
## 314	98.215	1	0	0
## 315	74.372	0	0	0
## 316	74.562	0	1	0
## 317	30.977	0	0	0
## 318	19.223	0	0	0
## 319	15.713	0	0	0
## 320	65.076	0	0	0
## 321	56.425	0	0	0
## 322	58.385	0	0	0
## 323	32.582	0	0	0
## 324	91.739	0	0	0
## 325	245.000	0	0	0
## 326	139.609	0	0	0
## 327	113.568	0	1	0
## 328	71.110	1	0	0
## 329	38.952	0	0	0
## 330	34.243	0	0	0
## 331	40.661	0	0	0
## 332	89.474	0	1	0
## 333	85.438	0	1	0
## 334	65.729	0	0	0
## 335	66.213	1	0	0
## 336	76.021	1	0	0
## 337	91.500	0	0	0
## 338	91.044	0	1	0
## 339	97.991	0	0	0
## 340	77.208	0	0	0
## 341	43.520	0	0	0
## 342	64.707	0	0	0
## 343	98.408	0	0	0
## 344	44.496	0	0	0
## 345	31.620	0	1	0

## 346	77.776	0	0	0
## 347	48.788	0	0	0
## 348	135.982	0	0	0
## 349	112.025	0	0	0
## 350	95.495	0	0	0
## 351	64.136	0	0	0
## 352	11.320	0	0	0
## 353	32.892	0	1	0
## 354	110.660	0	0	0
## 355	62.536	0	0	0
## 356	66.337	0	0	0
## 357	28.648	0	0	0
## 358	60.880	0	1	0
## 359	64.583	0	0	0
## 360	68.981	0	0	0
## 361	41.984	0	0	0
## 362	16.685	0	0	0
## 363	30.482	0	0	0
## 364	1.699	0	0	0
## 365	69.603	0	0	0
## 366	43.996	0	0	0
## 367	41.790	0	1	0
## 368	84.753	0	0	0
## 369	40.040	0	1	0
## 370	88.066	0	0	0
## 371	52.800	0	0	0
## 372	119.413	0	0	0
## 373	51.475	0	0	0
## 374	46.010	0	0	0
## 375	15.393	0	0	0
## 376	30.466	0	0	0
## 377	11.054	0	0	0
## 378	81.916	0	0	0
## 379	56.760	0	0	0
## 380	3.574	0	0	0
## 381	19.809	0	0	0
## 382	60.258	0	0	0
## 383	72.706	0	1	0
## 384	71.225	0	0	0
## 385	67.177	0	0	0
## 386	18.602	0	0	0
## 387	11.448	0	0	0
## 388	56.645	0	0	0
## 389	123.661	0	0	0
## 390	87.660	0	0	0
## 391	63.625	0	0	0
## 392	108.956	0	0	0
## 393	19.031	0	0	0
## 394	13.371	0	0	0
## 395	66.848	0	0	0
## 396	40.119	0	0	0
## 397	88.166	0	0	0
## 398	65.321	0	0	0
## 399	109.127	0	0	0

## 400	40.631	0	0	0
## 401	107.143	0	0	0
## 402	116.892	0	0	0
## 403	117.692	0	0	0
## 404	72.181	0	0	0
## 405	12.639	0	0	0
## 406	133.948	0	0	0
## 407	43.295	0	0	0
## 408	56.500	0	0	0
## 409	38.479	0	0	0
## 410	44.459	0	0	0
## 411	68.667	0	0	0
## 412	39.029	0	0	0
## 413	102.450	0	0	0
## 414	39.319	0	0	0
## 415	44.905	0	0	0
## 416	46.354	0	0	0
## 417	57.312	0	0	0
## 418	47.160	0	0	0
## 419	99.455	0	0	0
## 420	48.183	0	0	0
## 421	58.620	0	0	0
## 422	38.337	0	0	0
## 423	104.396	0	0	0
## 424	87.820	0	0	0
## 425	135.311	0	0	0
## 426	28.368	0	0	0
## 427	38.998	0	0	0
## 428	78.657	0	0	0
## 429	84.583	0	0	0
## 430	104.099	0	0	0
## 431	58.435	0	0	0
## 432	41.420	0	0	0
## 433	50.614	0	0	0
## 434	77.701	0	0	0
## 435	60.933	0	0	0
## 436	58.082	0	0	0
## 437	86.381	0	0	0
## 438	24.909	0	0	0
## 439	82.938	0	0	0
## 440	83.487	1	0	0
## 441	71.142	0	1	0
## 442	78.114	0	0	0
## 443	51.347	0	0	0
## 444	93.048	0	0	0
## 445	93.935	0	1	0
## 446	40.655	0	0	0
## 447	63.309	0	0	0
## 448	51.618	0	1	0
## 449	80.163	0	0	0
## 450	36.060	0	0	0
## 451	60.929	0	1	0
## 452	59.647	0	0	0
## 453	53.125	0	0	0

## 454	71.172	0	0	0
## 455	121.801	0	0	0
## 456	82.137	0	0	0
## 457	98.857	0	0	0
## 458	17.670	0	0	0
## 459	74.334	0	0	0
## 460	32.662	0	0	0
## 461	16.097	0	0	0
## 462	18.150	0	0	0
## 463	38.935	1	0	0
## 464	10.962	0	0	0
## 465	89.433	0	0	0
## 466	70.953	0	0	0
## 467	71.766	0	0	0
## 468	35.009	0	0	0
## 469	93.534	0	0	0
## 470	85.331	0	0	0
## 471	84.104	0	0	0
## 472	101.930	0	0	0
## 473	38.327	0	0	0
## 474	53.501	0	0	0
## 475	94.869	0	0	0
## 476	60.702	0	0	0
## 477	41.113	0	0	0
## 478	74.754	0	0	0
## 479	67.611	0	0	0
## 480	96.327	0	0	0
## 481	30.115	0	0	0
## 482	74.535	0	0	0
## 483	45.975	1	0	0
## 484	28.514	0	1	0
## 485	61.458	0	1	0
## 486	49.424	0	0	0
## 487	64.346	0	0	0
## 488	61.785	0	0	0
## 489	50.241	0	0	0
## 490	54.702	0	0	0
## 491	68.648	0	0	0
## 492	59.729	0	0	0
## 493	57.589	0	0	0
## 494	15.274	0	0	0
## 495	103.238	0	0	0
## 496	69.956	0	0	0
## 497	51.833	0	0	0
## 498	39.444	0	0	0
## 499	50.068	0	0	0
## 500	40.564	0	0	0
## 501	9.823	0	0	0
## 502	28.515	0	0	0
## 503	21.982	0	0	0
## 504	34.441	0	0	0
## 505	37.694	0	0	0
## 506	55.142	0	0	0
## 507	68.632	0	0	0

## 508	50.958	0	0	0
## 509	55.045	0	0	0
## 510	79.564	0	1	0
## 511	4.994	0	0	0
## 512	64.067	0	0	0
## 513	46.550	0	0	0
## 514	36.219	0	0	0
## 515	48.518	0	0	0
## 516	73.730	0	0	0
## 517	107.104	0	0	0
## 518	45.947	0	0	0
## 519	34.555	0	0	0
## 520	16.469	0	0	0
## 521	65.073	0	0	0
## 522	100.418	0	0	0
## 523	80.976	1	0	0
## 524	40.066	0	0	0
## 525	34.932	0	0	0
## 526	93.580	0	0	0
## 527	54.231	0	0	0
## 528	128.262	0	0	0
## 529	86.506	0	0	0
## 530	72.413	0	0	0
## 531	87.631	0	0	0
## 532	63.267	0	0	0
## 533	58.950	0	0	0
## 534	57.910	0	0	0
## 535	92.204	0	0	0
## 536	46.521	0	0	0
## 537	33.937	0	0	0
## 538	66.961	0	0	0
## 539	28.437	1	0	0
## 540	54.260	0	0	0
## 541	95.458	0	0	0
## 542	89.734	0	0	0
## 543	93.404	0	0	0
## 544	14.736	0	1	0
## 545	58.365	0	0	0
## 546	53.807	0	0	0
## 547	67.265	0	0	0
## 548	37.472	0	0	0
## 549	48.865	0	0	0
## 550	28.625	0	0	0
## 551	109.026	0	0	0
## 552	43.749	0	0	0
## 553	66.182	0	0	1
## 554	63.203	0	0	0
## 555	49.088	0	0	0
## 556	67.756	0	0	0
## 557	1.065	0	0	0
## 558	76.606	0	1	0
## 559	64.556	0	0	0
## 560	28.906	0	0	0
## 561	170.021	0	0	0

## 562	10.206	0	0	0
## 563	24.873	0	0	0
## 564	29.440	0	0	0
## 565	70.529	0	0	0
## 566	18.826	0	0	0
## 567	8.817	0	0	0
## 568	97.829	0	0	0
## 569	61.043	0	0	0
## 570	37.929	1	0	0
## 571	39.599	0	1	0
## 572	60.480	0	0	0
## 573	62.427	0	1	0
## 574	84.641	0	0	0
## 575	104.117	0	0	0
## 576	47.554	0	0	1
## 577	37.414	0	1	0
## 578	12.191	0	0	0
## 579	57.783	0	0	0
## 580	20.613	0	0	0
## 581	40.938	0	0	0
## 582	44.322	0	0	0
## 583	72.140	0	0	0
## 584	88.614	0	0	0
## 585	12.934	0	0	0
## 586	51.777	1	0	0
## 587	66.609	0	0	0
## 588	40.559	0	0	0
## 589	54.763	0	0	0
## 590	93.915	0	0	0
## 591	39.522	0	1	0
## 592	66.910	0	0	0
## 593	58.351	0	0	0
## 594	35.449	0	0	0
## 595	19.331	0	0	0
## 596	81.269	0	0	0
## 597	14.081	0	0	0
## 598	81.783	0	1	0
## 599	53.270	0	1	0
## 600	52.960	0	1	0
## 601	39.145	0	0	0
## 602	91.567	0	0	0
## 603	60.588	0	0	0
## 604	50.137	0	0	0
## 605	32.288	0	0	0
## 606	26.474	0	0	0
## 607	2.651	0	0	0
## 608	34.349	0	0	0
## 609	90.740	0	0	0

The name variable has additional information about the car engine capacity (for example 1.5 means 1.500 cc engine capacity). I choose to extract the engine capacity into a new column considering that it might influenced the used car's listing price.

```
df$engine_cap <- str_sub(df$car.name, start= -3)

df$engine_cap <- as.numeric(df$engine_cap)
```

Based on the data dictionary, The “year” variable holds the information about the manufactured year of the car. This can be easily used to generate the “rough” age of the car (based on the latest year specify).

```
unique(df$year)
```

```
## [1] 2018 2015 2020 2019 2021 2022 2016 2017 2014 2013 2010 2012 2023 2011
```

```
df$age <- 2023 - df$year
```

Variables grouping

To ease the cleaning and further EDA process, I group up the variables based on their data type and characteristic.

```
num_cols <- names(df)[sapply(df,is.numeric)]
char_cols <- names(df)[sapply(df,is.character)]
print("Columns who are numeric")
```

```
## [1] "Columns who are numeric"
```

```
print(num_cols)
```

```
## [1] "year" "mileage_km"
## [3] "rear.camera" "sun.roof"
## [5] "auto.retract.mirror" "electric.parking.brake"
## [7] "map.navigator" "vehicle.stability.control"
## [9] "keyless.push.start" "sports.mode"
## [11] "X360.camera.view" "power.sliding.door"
## [13] "auto.cruise.control" "price"
## [15] "instalment_month" "engine_cap"
## [17] "age"
```

```
print("Columns who are character")
```

```
## [1] "Columns who are character"
```

```
print(char_cols)
```

```
## [1] "car.name" "brand" "location" "transmission" "plate.type"
```

Group up numeric data

```

get_disc <- function(data, col_names) {
  match_cols <- character(0)
  for (col_name in col_names) {
    if (col_name %in% names(data)) {
      num_unique <- length(unique(data[[col_name]]))
      if (num_unique > 2 && num_unique < 20) {
        match_cols <- c(match_cols, col_name)
      }
    }
  }
  return(match_cols)
}

```

```

disc_cols <- get_disc(df, num_cols)
print("Discrete variables:")

```

```
## [1] "Discrete variables:"
```

```
print(disc_cols)
```

```
## [1] "year"          "engine_cap" "age"
```

The year variable holds the same information as the age variable, hence the “year” variable needs to be excluded to avoid multicollinearity.

```

remove_var <- "year"

# Remove from the list
num_cols <- num_cols[!(num_cols %in% remove_var)]
disc_cols <- disc_cols[!(disc_cols %in% remove_var)]

```

```

get_dum <- function(data, col_names) {
  match_cols <- character(0)
  for (col_name in col_names) {
    if (col_name %in% names(data)) {
      if (length(unique(data[[col_name]])) == 2) {
        match_cols <- c(match_cols, col_name)
      }
    }
  }
  return(match_cols)
}

```

```

dum_cols <- get_dum(df, num_cols)
print("Dummy variables:")

```

```
## [1] "Dummy variables:"
```

```
print(dum_cols)
```

```
## [1] "rear.camera"          "sun.roof"
## [3] "auto.retract.mirror"   "electric.parking.brake"
## [5] "map.navigators"        "vehicle.stability.control"
## [7] "keyless.push.start"    "sports.mode"
## [9] "X360.camera.view"      "power.sliding.door"
## [11] "auto.cruise.control"
```

```
get_cont <- function(data, col_names) {
  match_cols <- character(0)
  for (col_name in col_names) {
    if (col_name %in% names(data)) {
      if (length(unique(data[[col_name]])) > 20) {
        match_cols <- c(match_cols, col_name)
      }
    }
  }
  return(match_cols)
}
```

```
cont_cols <- get_cont(df, num_cols)
print("Continuous variables:")
```

```
## [1] "Continuous variables:"
```

```
print(cont_cols)
```

```
## [1] "mileage_km"          "price"              "instalment_month"
```

Group up character variable

```
get_cat <- function(data, col_names) {
  match_cols <- character(0)
  for (col_name in col_names) {
    if (col_name %in% names(data)) {
      num_unique_values <- length(unique(data[[col_name]]))
      if (num_unique_values > 2 && num_unique_values <= 20) {
        match_cols <- c(match_cols, col_name)
      }
    }
  }
  return(match_cols)
}
```

```
cat_cols <- get_cat(df, char_cols)
print("Category columns:")
```

```
## [1] "Category columns:"
```

```
print(cat_cols)
```

```
## [1] "brand"      "location"
```

```
get_desc <- function(data, col_names) {  
  match_cols <- character(0)  
  for (col_name in col_names) {  
    if (col_name %in% names(data)) {  
      if (length(unique(data[[col_name]])) > 20) {  
        match_cols <- c(match_cols, col_name)  
      }  
    }  
  }  
  return(match_cols)  
}
```

```
desc_cols <- get_desc(df, char_cols)  
print("Description columns:")
```

```
## [1] "Description columns:"
```

```
print(desc_cols)
```

```
## [1] "car.name"
```

```
bin_cols <- get_dum(df, char_cols)  
print("Binary columns:")
```

```
## [1] "Binary columns:"
```

```
print(bin_cols)
```

```
## [1] "transmission" "plate.type"
```

Data Transformation: Prepare numerical matrices

The OLS and machine learning model can't handle processing character column as they only accept numerical matrices. Therefore, I need to convert categorical and discrete character column into numerical value first.

```
for (col in cat_cols){  
  print(paste0(col, " unique values"))  
  print(unique(df[[col]]))  
}
```

```
## [1] "brand unique values"  
## [1] "Daihatsu" "Toyota" "Nissan" "Mitsubishi" "Honda"  
## [6] "Wuling" "Suzuki" "Mazda" "Datsun" "Hyundai"  
## [11] "Chevrolet" "Ford" "BMW"
```



```
## [1] "location unique values"
## [1] "Jakarta Utara"      "Bogor"             "Surabaya"
## [4] "Tangerang Selatan" "Jakarta Barat"     "Bekasi"
## [7] "Bandung"           "Malang"            "Depok"
## [10] "Jakarta Selatan"   "Jakarta Timur"     "Jakarta Pusat"
## [13] "Unknown"
```

First, there are 2 columns that are categorical: 1. Car brand: There are too many brand values to be converted to dummies through one hot encoding. The OLS model would not be able to handle too many brand variable if I change it to dummies. In the other hand, the car brand may still influence the car price through the brand popularity and its subsequent quality recognition. To capture the effect of a brand popularity, I choose to make new “brand_popularity” column that contains the popularity percentage of the brand in the dataset: - brand_popularity = (brand count / total rows) * 100% 2. Location: The selling location can affected the price of used car by through the location preferability. But, it would be more appropriate to use model that incorporates them as geo points like GWR to analyze the “location” effect rather than converting it to dummies. Therefore, I choose to omit the “location” column for now.

```
df_ols = data.frame(df)
```

```
#Convert the brand into brand popularity
df_ols <- df_ols %>%
  group_by(brand) %>%
  mutate(brand_popularity_count = n()) %>% # Count of each brand
  ungroup() %>%
  mutate(brand_popularity = brand_popularity_count / nrow(.)*100)

# Remove the category columns from df_ols
df_ols <- df_ols %>% select(-any_of(c("brand_popularity_count")))
```

Beside the category columns, there is still binary columns that is characterized by only having 2 unique values. These columns can be easily converted into dummy variable (1 and 0).

```
df_ols$is.manual <- ifelse(df_ols$transmission == "Manual",1,0)
df_ols$is.odd_plate <- ifelse(df_ols$plate.type == "odd plate",1,0)

remove_char <- c(cat_cols,bin_cols)

## Remove from cols selection
all_cols <- all_cols[!(all_cols %in% remove_char)]
char_cols <- char_cols[!(char_cols %in% remove_char)]

## Input new dummy variables into the list
num_cols <- names(df_ols)[sapply(df_ols,is.numeric)]
dum_cols <- get_dum(df_ols, num_cols)
```

Checking duplicates

```
sum(duplicated(df))
```

```
## [1] 0
```

There is no duplicated entry (row) found.

Checking missing data (NA, empty string (“”), and”NULL” entry)

```
sum(is.na(df))
```

```
## [1] 0
```

```
## Checking for empty string  
print("Columns with empty string value")
```

```
## [1] "Columns with empty string value"
```

```
for (col in all_cols) {  
  n_empty <- sum(df[[col]]=="", na.rm = TRUE)  
  if (n_empty > 0) {  
    print(paste0("Number and percentages of empty string value in ", col))  
    print(n_empty)  
    print((n_empty/(nrow(df)))*100)  
  }  
}
```

```
print("Columns with NULL string value")
```

```
## [1] "Columns with NULL string value"
```

```
for (col in all_cols) {  
  n_null <- sum(is.null(df[[col]]))  
  if (n_null > 0) {  
    print(paste0("Number and percentages of NULL value in ", col))  
    print(n_null)  
    print((n_null/(nrow(df)))*100)  
  }  
}
```

There is no entry containing NA , empty string (“”), or”NULL” entry.

Exploratory Data Analysis (EDA)

```
## For plotting  
library(ggplot2)  
library(patchwork)
```

```
## Warning: package 'patchwork' was built under R version 4.3.3
```

Continuous Data’s summary

```
summary(df_ols[cont_cols])
```

```
##      mileage_km      price      instalment_month
## Min.   :  1.065   Min.   :7.70e+07   Min.    : 1570000
## 1st Qu.: 39.145   1st Qu.:1.32e+08   1st Qu.: 2690000
## Median : 58.365   Median :1.77e+08   Median : 3610000
## Mean   : 61.686   Mean   :1.83e+08   Mean    : 3728292
## 3rd Qu.: 81.726   3rd Qu.:2.12e+08   3rd Qu.: 4320000
## Max.   :245.000   Max.   :5.42e+08   Max.    :11040000
```

The max value of the continuous columns seems to be extremely far from the 3rd quartile and mean. This can indicate an outliers problem.

One of the goals in this analysis is to make a machine learning model to predict the used car price. Outliers can lower the model's accuracy, so I need to check for outliers first before EDA and model training.

```
for (col in cont_cols){

  var <- df_ols[[col]]
  #Calculate IQR, Upper, and Lower Bound
  iqr <- IQR(var)
  q1 <- quantile(var, 0.25)
  q3 <- quantile(var, 0.75)

  # Calculate Upper and lower bound
  lower_bound <- q1 - (1.5*iqr)
  upper_bound <- q3 + (1.5*iqr)

  # Count character that is outside the bounds
  outliers <- which(var < lower_bound | var > upper_bound)

  # Count percentages
  print(col)
  res = (length(outliers) / length(df[[col]]))*100
  print(res)
}
```

```
## [1] "mileage_km"
## [1] 1.149425
## [1] "price"
## [1] 4.269294
## [1] "instalment_month"
## [1] 4.269294
```

But all of the variables' outliers are less than 5% of the total data points, which is still in the acceptable range. But since I aim to make a prediction model, I choose to impute extreme outliers (more or less than 3 standard deviation from the mean).

```
## Duplicate Dataframe
df2 <- data.frame(df_ols)

## Remove outliers
```

```

for (col in cont_cols){
  avg <- mean(df2[[col]], na.rm = TRUE)
  std <- sd(df2[[col]], na.rm = TRUE)

  upper_limit <- avg + (3*std)
  lower_limit <- avg - (3*std)

  df2 <- df2[which(df2[[col]] <= upper_limit & df2[[col]] >= lower_limit),]
}

```

```
print("With outliers")
```

```
## [1] "With outliers"
```

```
print(dim(df_ols))
```

```
## [1] 609 25
```

```
print("Without outliers")
```

```
## [1] "Without outliers"
```

```
print(dim(df2))
```

```
## [1] 582 25
```

Imputed 12 rows of outliers.

```
print("With outliers")
```

```
## [1] "With outliers"
```

```
print(summary(df_ols[cont_cols]))
```

```
##      mileage_km      price      instalment_month
## Min.   : 1.065   Min.   :7.70e+07   Min.    : 1570000
## 1st Qu.: 39.145  1st Qu.:1.32e+08   1st Qu.: 2690000
## Median : 58.365  Median :1.77e+08   Median : 3610000
## Mean   : 61.686  Mean   :1.83e+08   Mean    : 3728292
## 3rd Qu.: 81.726  3rd Qu.:2.12e+08   3rd Qu.: 4320000
## Max.   :245.000  Max.   :5.42e+08   Max.    :11040000
```

```
print("Without outliers")
```

```
## [1] "Without outliers"
```

```
print(summary(df2[cont_cols]))
```

```
##      mileage_km      price      instalment_month
## Min.   : 1.065   Min.   : 77000000   Min.    :1570000
## 1st Qu.: 39.058   1st Qu.:132000000   1st Qu.:2690000
## Median : 58.358   Median :172500000   Median :3510000
## Mean   : 60.537   Mean   :175755155   Mean    :3581134
## 3rd Qu.: 80.908   3rd Qu.:211000000   3rd Qu.:4300000
## Max.   :142.175   Max.   :340000000   Max.    :6930000
```

The min and 1st quartile value doesn't change after I imputed the outliers. The outliers removed are the value exceeding the over the 3 standard deviation of the mean on the positive side. Moreover, the continuous data seems to be still positively skewed since the media value is less than the mean.

Numerical Variables

Continuous variable

```
summary(df2[cont_cols])
```

Univariate analysis

```
##      mileage_km      price      instalment_month
## Min.   : 1.065   Min.   : 77000000   Min.    :1570000
## 1st Qu.: 39.058   1st Qu.:132000000   1st Qu.:2690000
## Median : 58.358   Median :172500000   Median :3510000
## Mean   : 60.537   Mean   :175755155   Mean    :3581134
## 3rd Qu.: 80.908   3rd Qu.:211000000   3rd Qu.:4300000
## Max.   :142.175   Max.   :340000000   Max.    :6930000
```

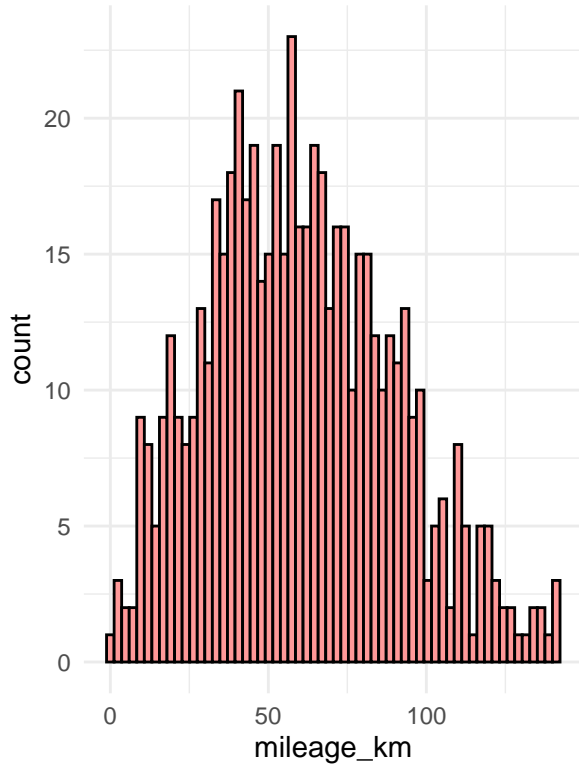
The mileage's min value of only 1 km is odd, but it is still possible for a barely used car to be sold.

```
for (col in cont_cols) {
  hist_plot <- ggplot(df2, aes(x = .data[[col]])) +
    geom_histogram(bins = 60, fill = '#FF9999', color = 'black') + # Added color for bin outlines
    labs(title = paste("Histogram of", col), x = col, y = "count") +
    theme_minimal()

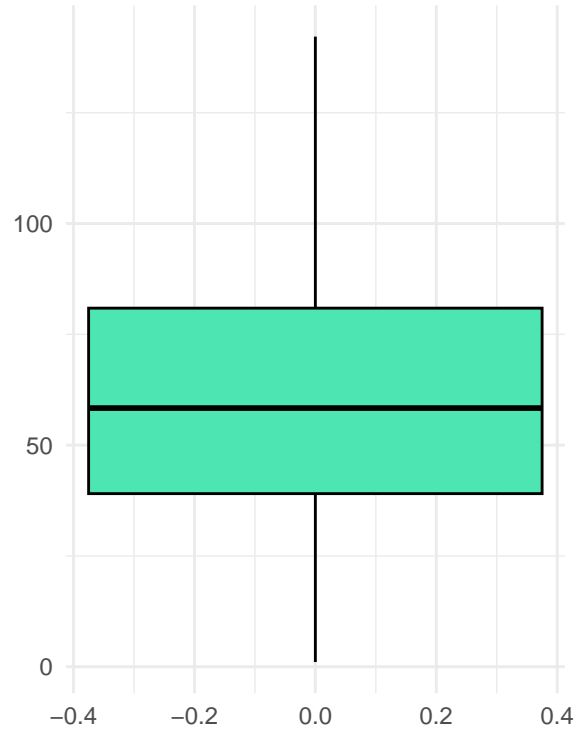
  # Boxplot
  box_plot <- ggplot(df2, aes(y = .data[[col]])) +
    geom_boxplot(fill = '#4DE5B1', color = 'black') + # Added color for boxplot outlines
    labs(title = paste("Boxplot of", col), y = "") +
    theme_minimal()

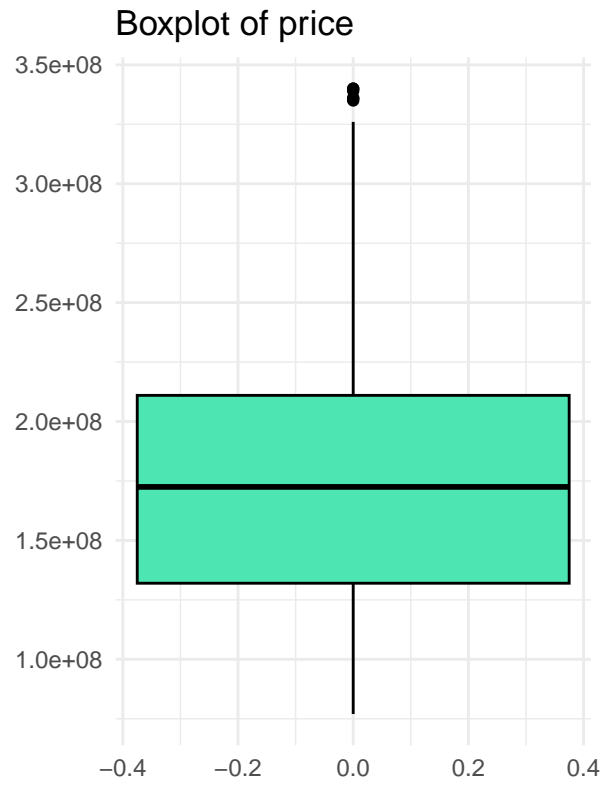
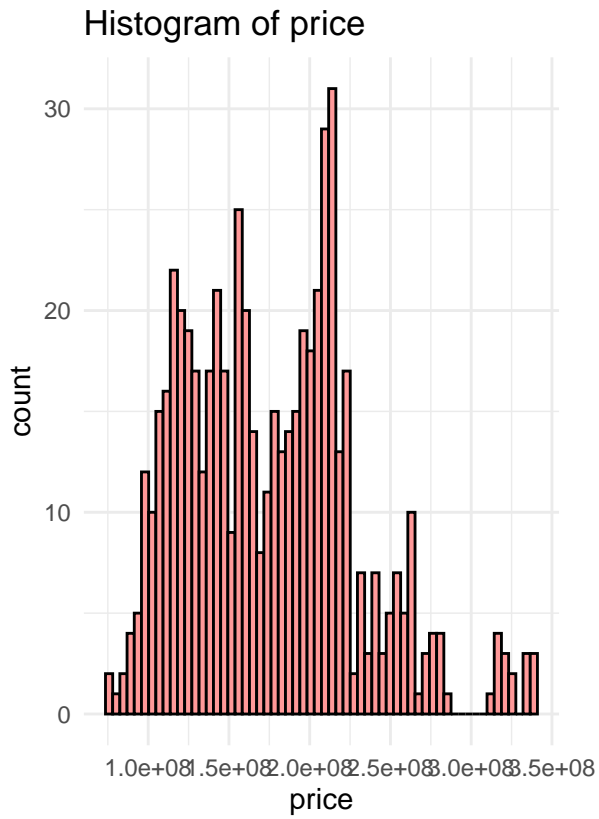
  # Combine and print the plots
  print(hist_plot + box_plot)
}
```

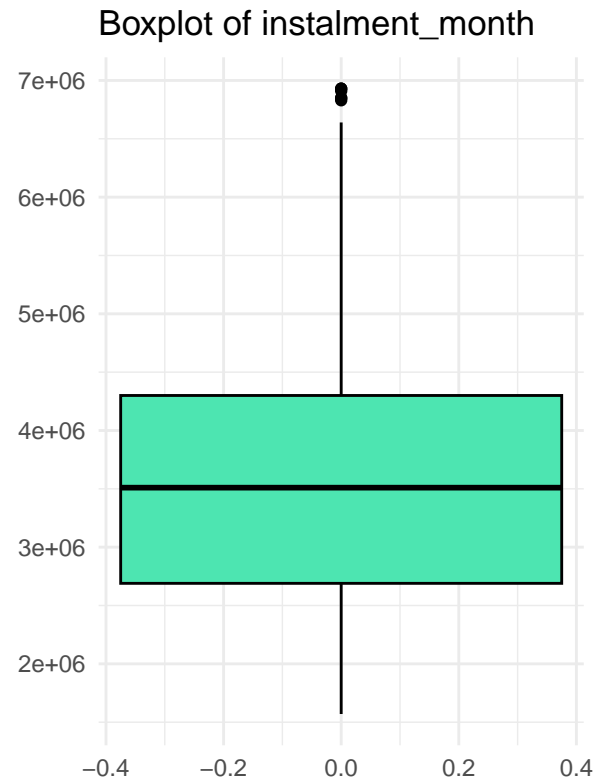
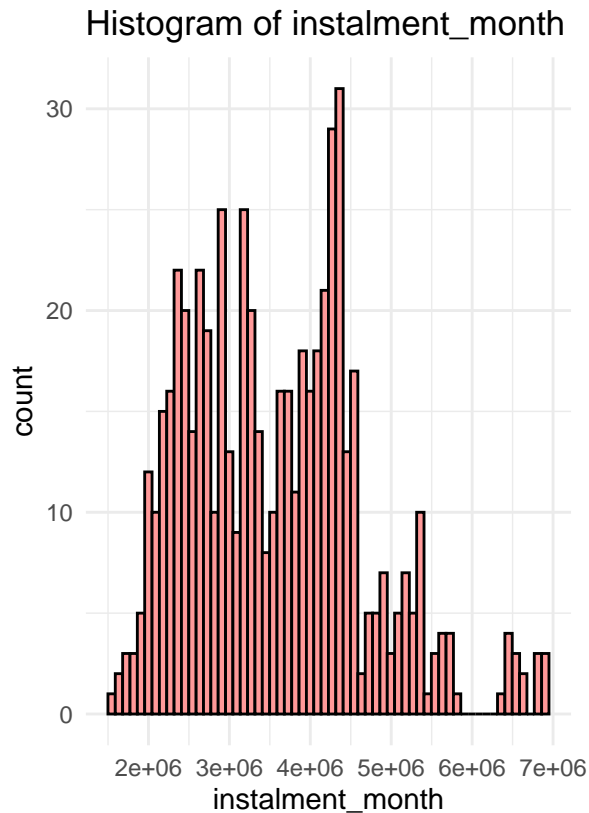
Histogram of mileage_km



Boxplot of mileage_km







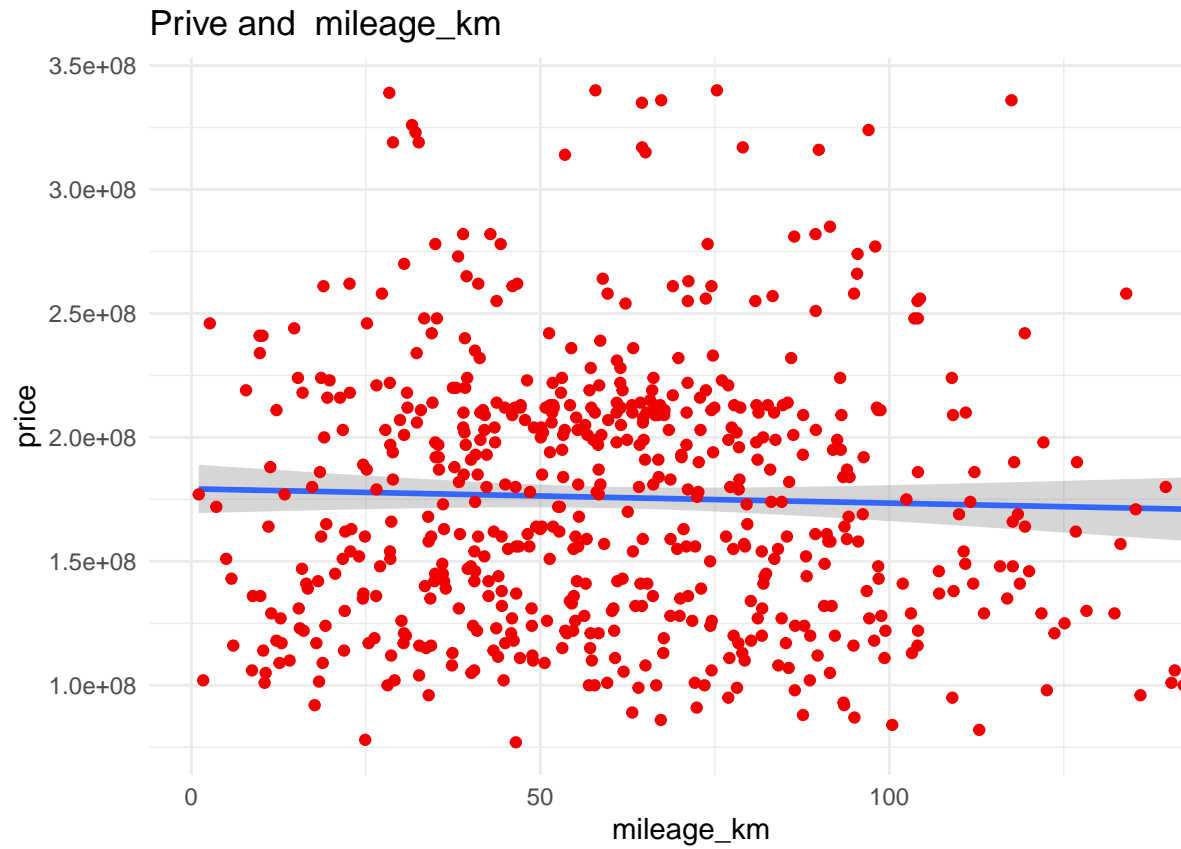
Insights: - All of the continuous variables are positively skewed. - Boxplot revealed that the skewness is caused by outliers in the right side.

```
dep_var <- "price"

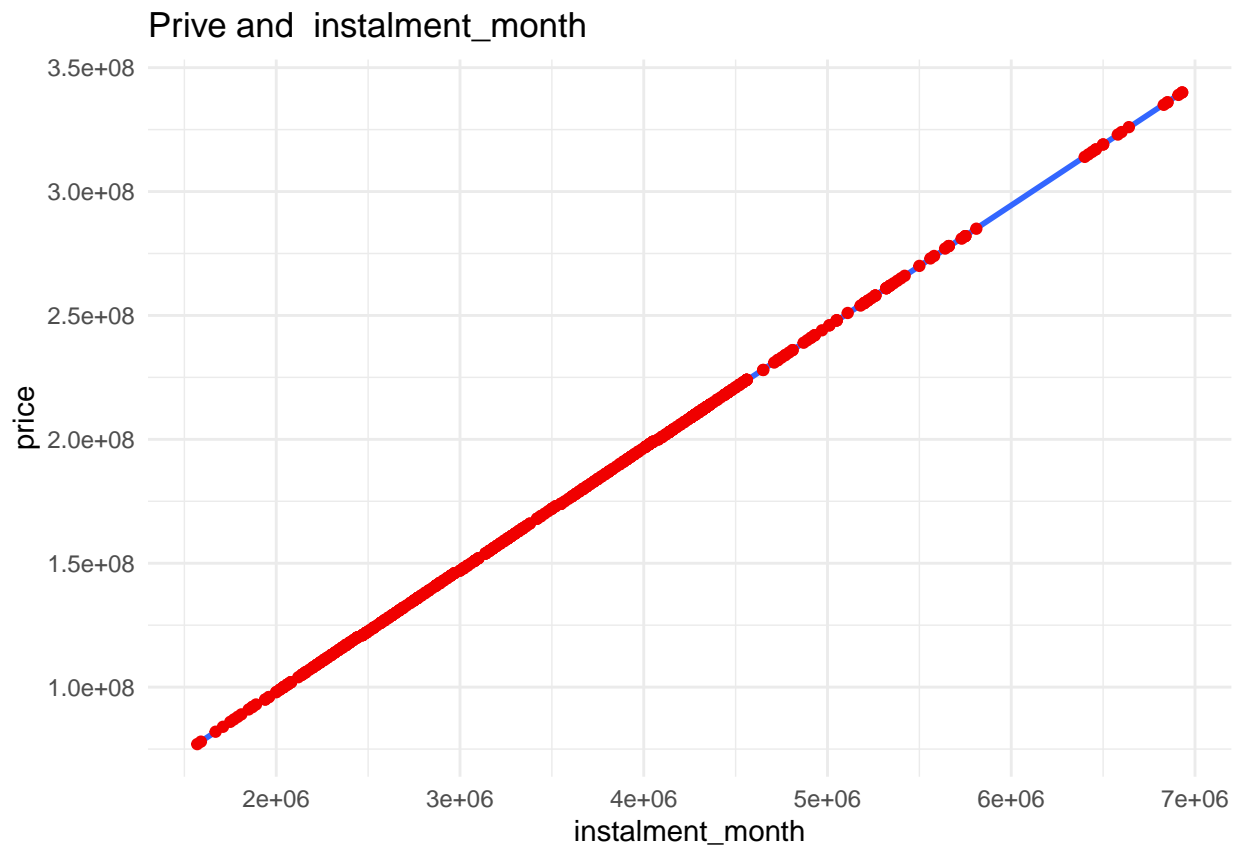
# Prepare independent variables
num_cols <- num_cols[num_cols != dep_var]
cont_cols <- cont_cols[cont_cols != dep_var]
```

```
for (col in cont_cols){
  plot <- ggplot(df2, aes(x =.data[[col]], y =price, na.rm =TRUE)) +
    geom_smooth(formula = y ~ x, method = "lm")+
    geom_point(color='#f00000')+
    labs(title= paste("Price and ", col))+
    theme_minimal()

  print(plot)
}
```

Bivariate analysis



Insights: - The milage_km seems to have low relationship with price, as the correlation line is almost horizontal. - Meanwhile, monthly installment has a perfect linear relationship with price. This is most likely caused by the monthly installment value being determined from price value. Therefore, I will omit "instalment_month" from further analysis to avoid multicollinearity.

```
remove_var <- c("instalment_month")

# Remove from list
num_cols <- num_cols[num_cols != remove_var]
cont_cols <- cont_cols[cont_cols != remove_var]
```

```
cont_cols <- c(cont_cols, "price")
```

Discrete Variable

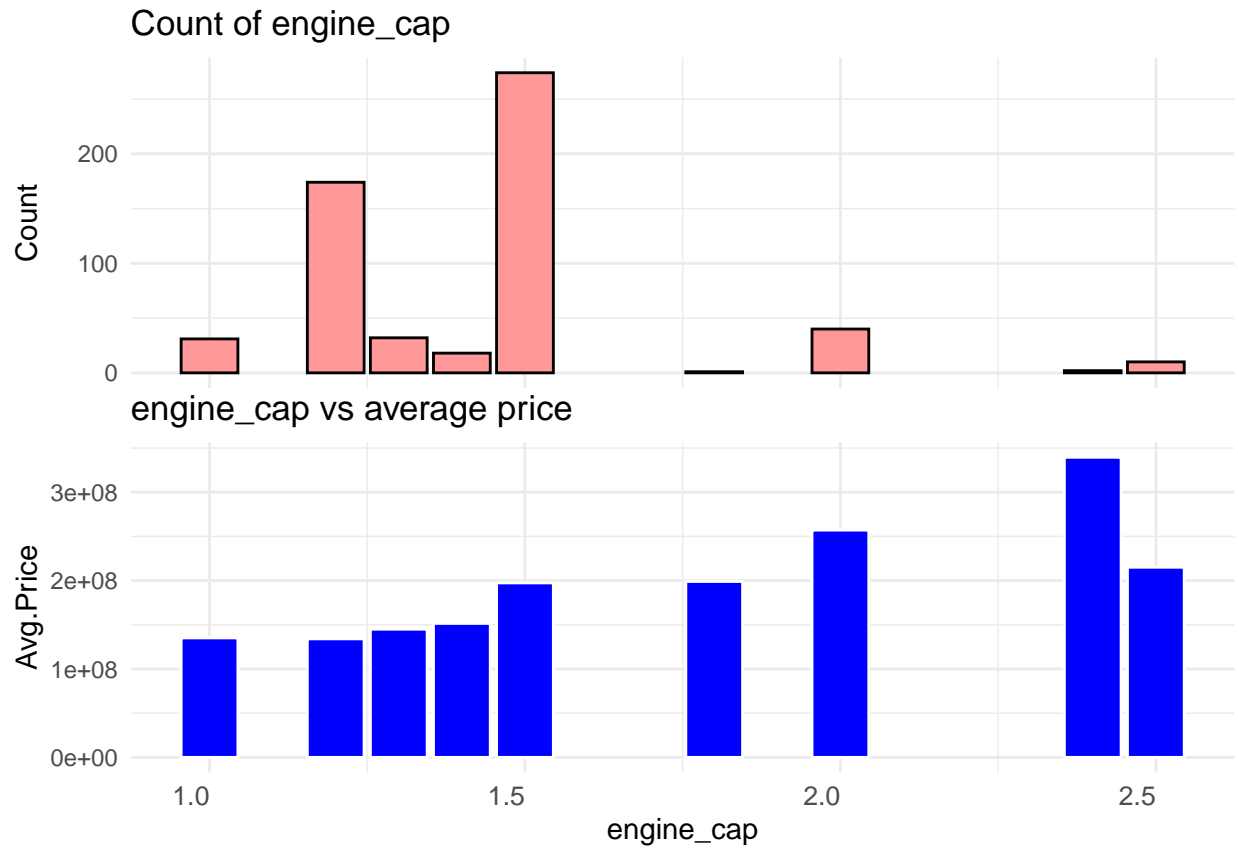
```
for (col in disc_cols) {
  options(repr.plot.width = 20, repr.plot.height = 8)

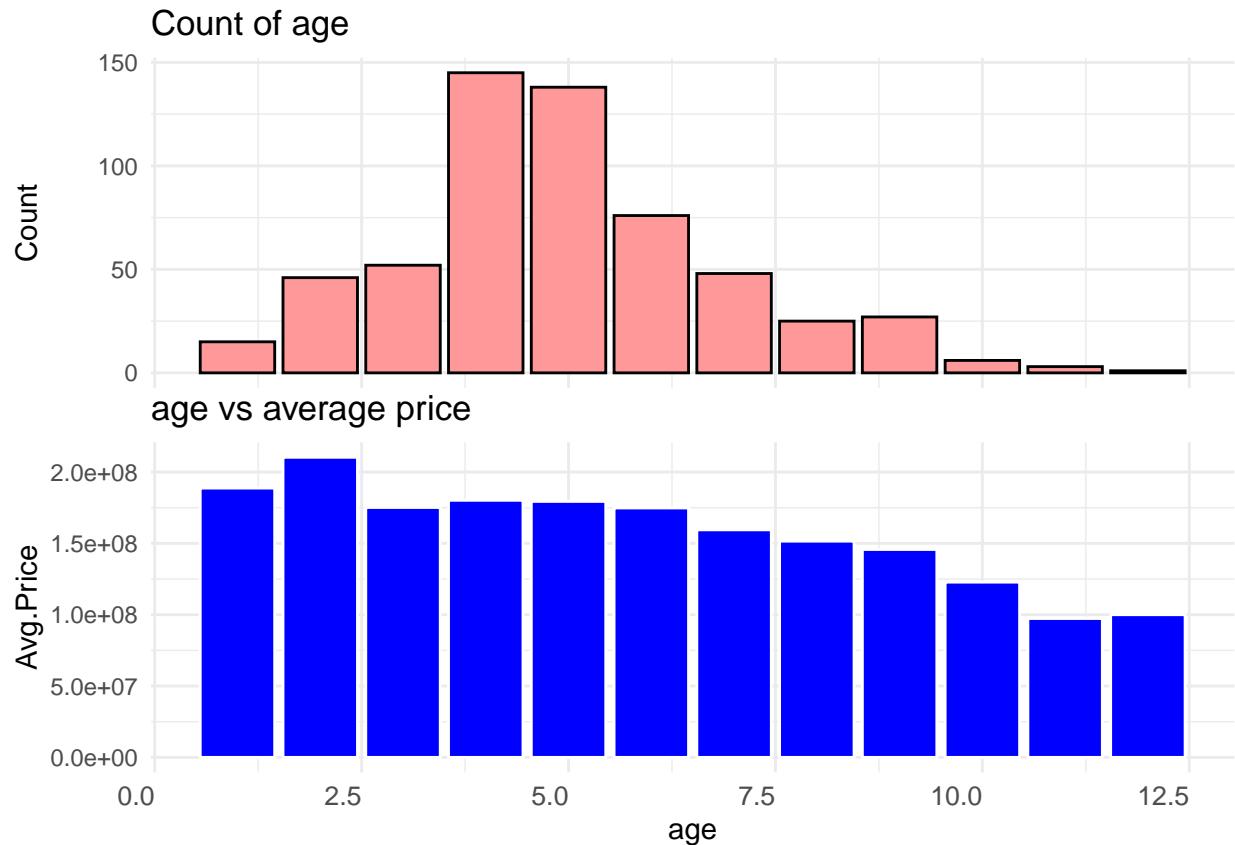
  # Histogram
  hist_plot <- ggplot(df2, aes(x = .data[[col]])) +
    geom_bar(fill = '#FF9999', color = 'black') +
    labs(title = paste("Count of", col), y = "Count") +
    xlab(col) + # Keep xlab here, but we'll blank it out in theme
    theme_minimal() +
    theme(
      # Hide x-axis title, text (labels), and ticks for the top plot
      axis.title.x = element_blank(), # Hides the "xlab(col)" title
      axis.text.x = element_blank(), # Hides the tick labels (A, B, C, etc.)
      axis.ticks.x = element_blank(), # Hides the small tick marks
      plot.margin = margin(b = 0, unit = "pt") # Reduce bottom margin to minimize space between plots
    )

  # Bar plot (dependent variable)
  data <- df2 %>%
    group_by(.data[[col]]) %>%
    summarise(Avg.Price = mean(price), .groups = 'drop')

  price_plot <- ggplot(data, aes(x = .data[[col]], y = Avg.Price)) +
    geom_bar(stat = 'identity', fill = 'blue', color = 'white') +
    labs(title = paste(col, "vs average price")) +
    xlab(col) + # This plot will keep its x-axis label
    theme_minimal() +
    theme(
      axis.text.x = element_text(hjust = 1, vjust = 1, size = 10),
      plot.margin = margin(t = 0, unit = "pt") # Reduce top margin to minimize space between plots
    )

  print(hist_plot / price_plot + plot_layout(guides = "collect", axis_titles = "collect"))
}
```





Insights: - Most of the car in the dataframe are 4 and 5 year old. - Furthermore, the average price tends to increase as it gets older. - To test this, I need to take a closer look at the car models. - Most of the car listed have engine capacity 1.500cc and lower. - As the engine capacity gets higher, the price tends to increase.

Let's see a detailed look on yearly effect based on specific car model.

```
model_count <- aggregate(data.frame(count = df2$car.name), list(value = df2$car.name), length) %>%
  arrange(desc(count))
```

model_count

##	value	count
## 1	BRIO SATYA E 1.2	48
## 2	XPANDER ULTIMATE 1.5	38
## 3	CALYA G 1.2	20
## 4	AYLA R 1.2	19
## 5	RUSH S TRD SPORTIVO 1.5	18
## 6	BRIO RS 1.2	17
## 7	AGYA G TRD 1.2	16
## 8	HR-V E 1.5	15
## 9	AVANZA G 1.3	14
## 10	MOBILIO E 1.5	12
## 11	AYLA X 1.0	9
## 12	KIJANG INNOVA REBORN G 2.0	9
## 13	YARIS S TRD 1.5	9
## 14	IGNIS GX 1.2	8
## 15	XPANDER SPORT 1.5	7

## 16	ALMAZ LT LUX 1.5	6
## 17	BALENO HATCHBACK 1.4	6
## 18	BR-V E 1.5	6
## 19	CITY RS HATCHBACK 1.5	6
## 20	KIJANG INNOVA G 2.0	6
## 21	LIVINA VL 1.5	6
## 22	SX4 S-CROSS 1.5	6
## 23	TERIOS R 1.5	6
## 24	XPANDER EXCEED 1.5	6
## 25	AGYA G 1.0	5
## 26	AYLA R DLX 1.2	5
## 27	CARRY PICK UP 1.5	5
## 28	ERTIGA GX 1.4	5
## 29	ERTIGA SPORT GT 1.5	5
## 30	HR-V E PLUS 1.5	5
## 31	HR-V SE 1.5	5
## 32	IGNIS GL 1.2	5
## 33	MOBILIO RS 1.5	5
## 34	2 R SKYACTIV 1.5	4
## 35	AGYA G 1.2	4
## 36	AVANZA VELOZ 1.5	4
## 37	HR-V S 1.5	4
## 38	KIJANG INNOVA REBORN V 2.0	4
## 39	MARCH 1.2	4
## 40	ROCKY R 1.0	4
## 41	SIENTA Q 1.5	4
## 42	SIENTA V 1.5	4
## 43	SIGRA R STD 1.2	4
## 44	TERIOS X DLX 1.5	4
## 45	X-TRAIL 2.5	4
## 46	2 GT SKYACTIV 1.5	3
## 47	AGYA G TRD SPORTIVO 1.2	3
## 48	AVANZA E 1.3	3
## 49	BR-V PRESTIGE 1.5	3
## 50	CONFERO S 1.5	3
## 51	CR-V TURBO 1.5	3
## 52	ERTIGA GL 1.4	3
## 53	ERTIGA GX 1.5	3
## 54	GRAND LIVINA HIGHWAY STAR AUTECH 1.5	3
## 55	KIJANG INNOVA Q 2.0	3
## 56	KIJANG INNOVA V 2.0	3
## 57	OUTLANDER SPORT PX 2.0	3
## 58	SIENTA G 1.5	3
## 59	SIGRA R DLX 1.2	3
## 60	XL7 ALPHA 1.5	3
## 61	2 GT 1.5	2
## 62	AGYA GR SPORT 1.2	2
## 63	ALMAZ S+T 1.5	2
## 64	AVANZA G 1.5	2
## 65	AYLA X 1.2	2
## 66	BRIO E 1.2	2
## 67	CORTEZ S T LUX 1.5	2
## 68	ERTIGA DREZA 1.4	2
## 69	GO+ PANCA T 1.2	2

## 70	GRAND LIVINA HIGHWAY STAR 1.5	2
## 71	GRAND LIVINA SV 1.5	2
## 72	GRAND LIVINA XV 1.5	2
## 73	KIJANG INNOVA REBORN G 2.4	2
## 74	KIJANG INNOVA REBORN VENTURER GASOLINE 2.0	2
## 75	MARCH 1.5	2
## 76	MOBILIO E PRESTIGE 1.5	2
## 77	MOBILIO S 1.5	2
## 78	RAIZE TURBO G 1.0	2
## 79	ROCKY X 1.2	2
## 80	RUSH S 1.5	2
## 81	RUSH S GR SPORT 1.5	2
## 82	SERENA HIGHWAY STAR 2.0	2
## 83	SIRION M602RS 1.3	2
## 84	TERIOS R DLX 1.5	2
## 85	TERIOS TX 1.5	2
## 86	X-TRAIL 2.0	2
## 87	XENIA R 1.3	2
## 88	XENIA R DLX 1.3	2
## 89	XENIA R SPORTY 1.3	2
## 90	XENIA R STD 1.3	2
## 91	XL7 BETA 1.5	2
## 92	XL7 ZETA GL 1.5	2
## 93	2 R 1.5	1
## 94	AGYA G TRD 1.0	1
## 95	AGYA TRD SPORTIVO 1.0	1
## 96	ALMAZ L TURBO 1.5	1
## 97	ALMAZ LT LUX + SC CVT 1.5	1
## 98	ALMAZ LT LUX CVT 1.5	1
## 99	ALMAZ RS LT LUX + SC CVT 1.5	1
## 100	ALMAZ S+T SMART ENJOY 1.5	1
## 101	AVANZA VELOZ 1.3	1
## 102	AYLA X ELEGANT 1.0	1
## 103	BIANTE SKYACTIV 2.0	1
## 104	BR-V E PRESTIGE 1.5	1
## 105	BRIO E 1.3	1
## 106	BRIO SATYA S 1.2	1
## 107	CALYA E 1.2	1
## 108	CITY E 1.5	1
## 109	CIVIC FB2 1.8	1
## 110	COLT L300 PICK UP 2.5	1
## 111	CONFERO 2WD 1.5	1
## 112	CONFERO S L 1.5	1
## 113	CONFERO S L LUX 1.5	1
## 114	CORTEZ C T LUX 1.5	1
## 115	CORTEZ L T LUX 1.5	1
## 116	CORTEZ T LUX 1.5	1
## 117	CR-V 2.0	1
## 118	CRETA PRIME 1.5	1
## 119	CROSS 1.2	1
## 120	CX-3 GT 2.0	1
## 121	CX-3 TOURING 2.0	1
## 122	CX-5 GT 2.5	1
## 123	CX-5 HIGH 2.5	1

```

## 124          ECOSPORT TITANIUM 1.5      1
## 125          ERTIGA DREZA GS 1.4        1
## 126          ERTIGA GL 1.5              1
## 127          FORTUNER G 2.5             1
## 128          FORTUNER G TRD 2.5         1
## 129          FREED S 1.5                1
## 130          GO PANCA T LIVE 1.2        1
## 131          GRAND LIVINA X-GEAR 1.5    1
## 132          IGNIS GX AGS 1.2           1
## 133          JAZZ RS 1.5                1
## 134          JUKE 1.5                   1
## 135          KARIMUN GL 1.0              1
## 136          KARIMUN GS 1.0             1
## 137          KIJANG INNOVA VENTURER 2.0 1
## 138          MARCH L 1.2                1
## 139          OUTLANDER SPORT GLS 2.0    1
## 140          RAIZE GR 1.0                1
## 141          RAIZE GR SPORT 1.0          1
## 142          RAIZE GR SPORT TSS 1.0      1
## 143          RAIZE GR TWO TONE 1.0       1
## 144          RUSH G 1.5                  1
## 145          RUSH S TRD 1.5              1
## 146          SIGRA D 1.0                 1
## 147          SIGRA M 1.0                 1
## 148          SIGRA R 1.2                 1
## 149          SIGRA X 1.2                 1
## 150          SIRION 1.3                  1
## 151          SPARK PREMIER 1.4           1
## 152          STARGAZER PRIME 1.5         1
## 153          TERIOS ADVENTURE R 1.5      1
## 154          TERIOS X 1.5                1
## 155          X-TRAIL URBAN SELECTION 2.5 1
## 156          XENIA X 1.3                 1
## 157          XENIA X STD 1.3             1
## 158          XPANDER CROSS 1.5           1
## 159          YARIS G 1.5                 1

```

```

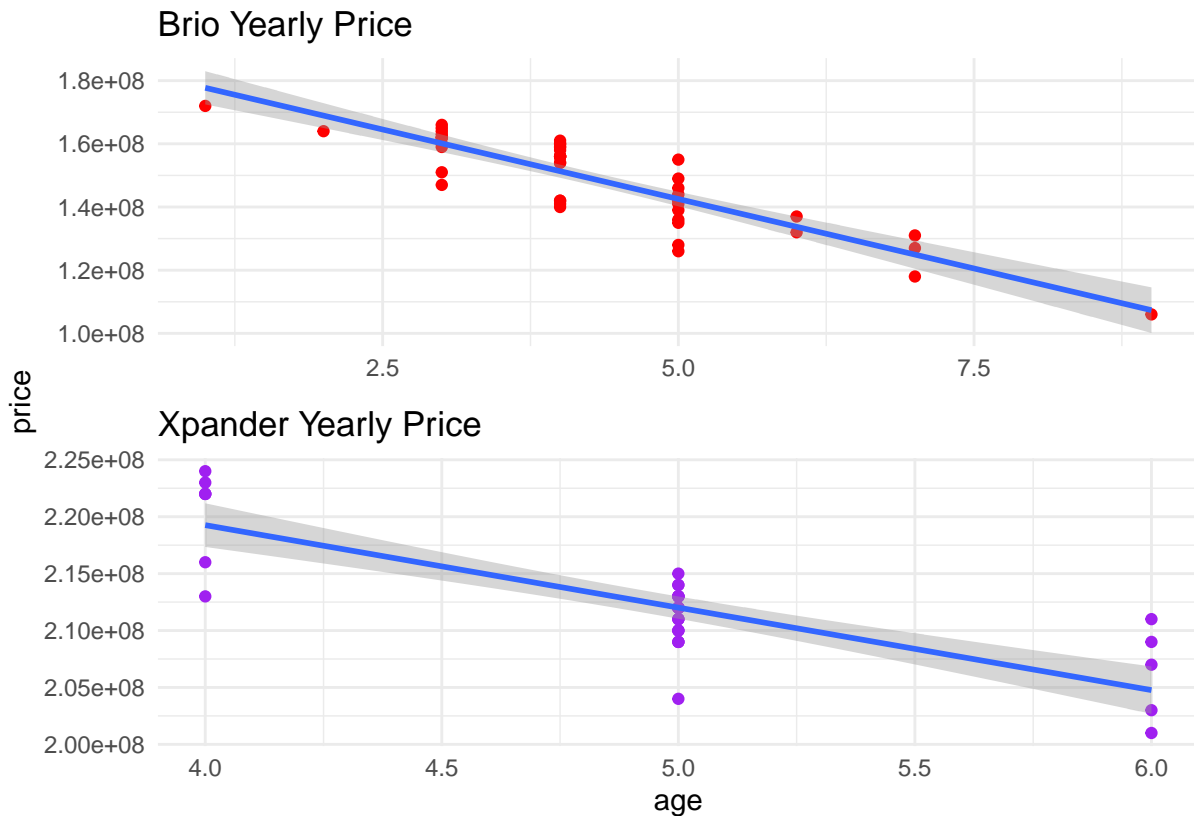
# Filter based on models
df_brio <- filter(df2, car.name=='BRIO SATYA E 1.2')
df_xpander <- filter(df2, car.name=='XPANDER ULTIMATE 1.5')

# Plot
a <- ggplot(df_brio, aes(x= age, y =price)) +
  geom_point(color='red')+
  geom_smooth(formula = y ~ x, method = "lm")+
  labs(title= 'Brio Yearly Price') +
  theme_minimal()

b <- ggplot(df_xpander, aes(x= age, y =price)) +
  geom_point(color='purple')+
  geom_smooth(formula = y ~ x, method = "lm")+
  labs(title= 'Xpander Yearly Price') +
  theme_minimal()

```

```
print(a / b + plot_layout(guides = "collect", axis_titles = "collect"))
```



Insight: - Yes, the specified investigation shows that the price of a car decrease as it get older.

Dummy variable

```
for (col in dum_cols) {
  options(repr.plot.width = 20, repr.plot.height = 8)

  # Histogram
  hist_plot <- ggplot(df2, aes(x = .data[[col]])) +
    geom_bar(fill = '#FF9999', color = 'black') +
    labs(title = paste("Histogram of", col), y = "Count") +
    xlab(col) + # Keep xlab here, but we'll blank it out in theme
    theme_minimal() +
    theme(
      # Hide x-axis title, text (labels), and ticks for the top plot
      axis.title.x = element_blank(), # Hides the "xlab(col)" title
      axis.text.x = element_blank(),  # Hides the tick labels (A, B, C, etc.)
      axis.ticks.x = element_blank(), # Hides the small tick marks
      plot.margin = margin(b = 0, unit = "pt") # Reduce bottom margin to minimize space between plots
    )
}
```



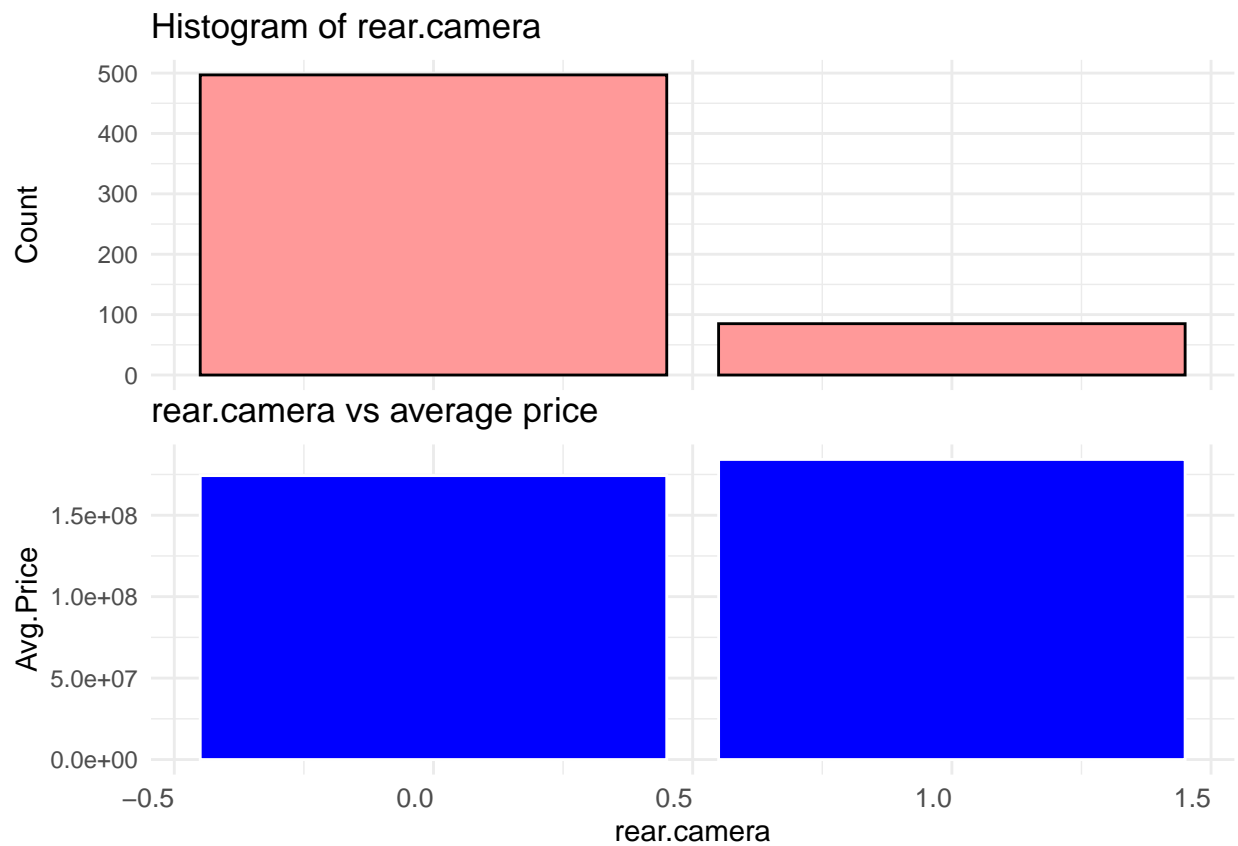
```

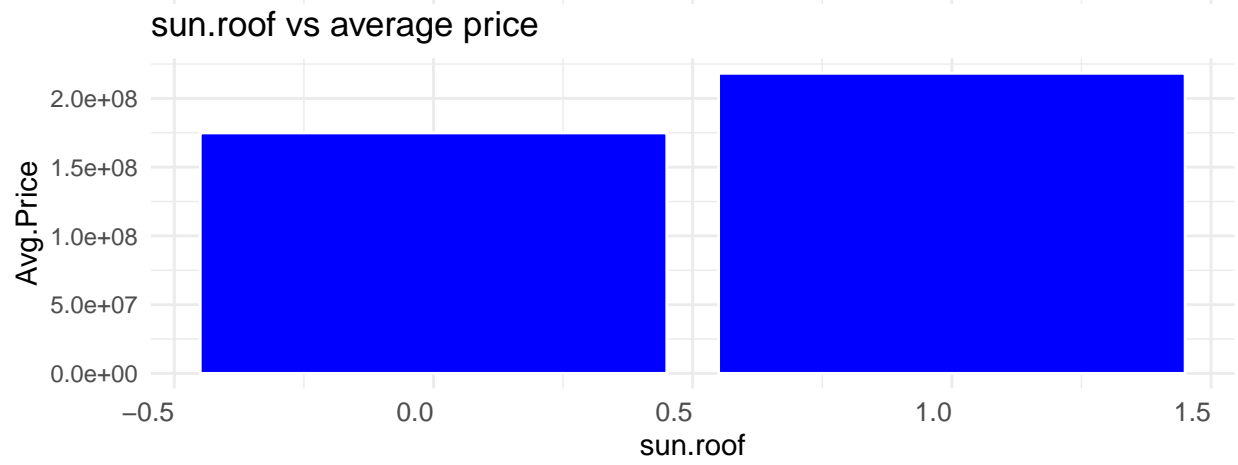
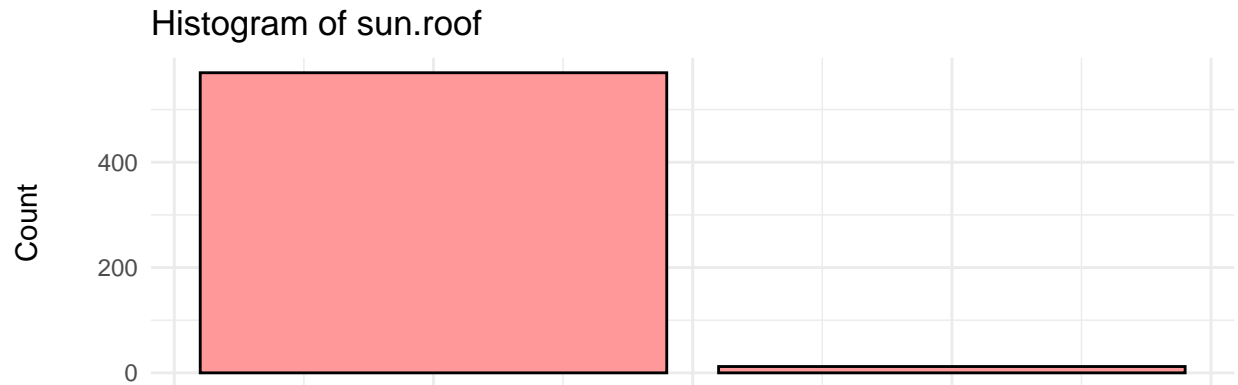
# Bar plot (dependent variable)
data <- df2 %>%
  group_by(.data[[col]]) %>%
  summarise(Avg.Price = mean(price), .groups = 'drop')

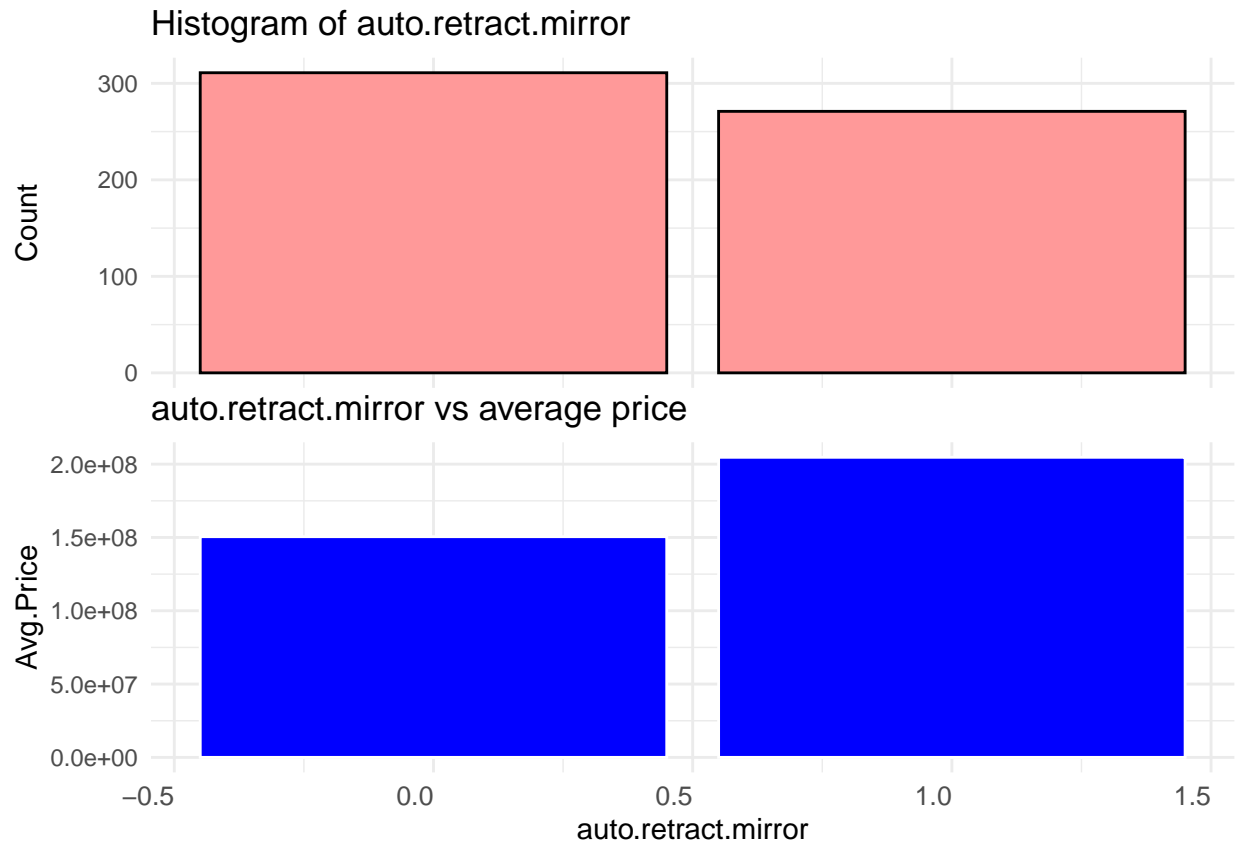
price_plot <- ggplot(data, aes(x = .data[[col]], y = Avg.Price)) +
  geom_bar(stat = 'identity', fill = 'blue', color = 'white') +
  labs(title = paste(col, "vs average price")) +
  xlab(col) + # This plot will keep its x-axis label
  theme_minimal() +
  theme(
    axis.text.x = element_text(hjust = 1, vjust = 1, size = 10),
    plot.margin = margin(t = 0, unit = "pt") # Reduce top margin to minimize space between plots
  )

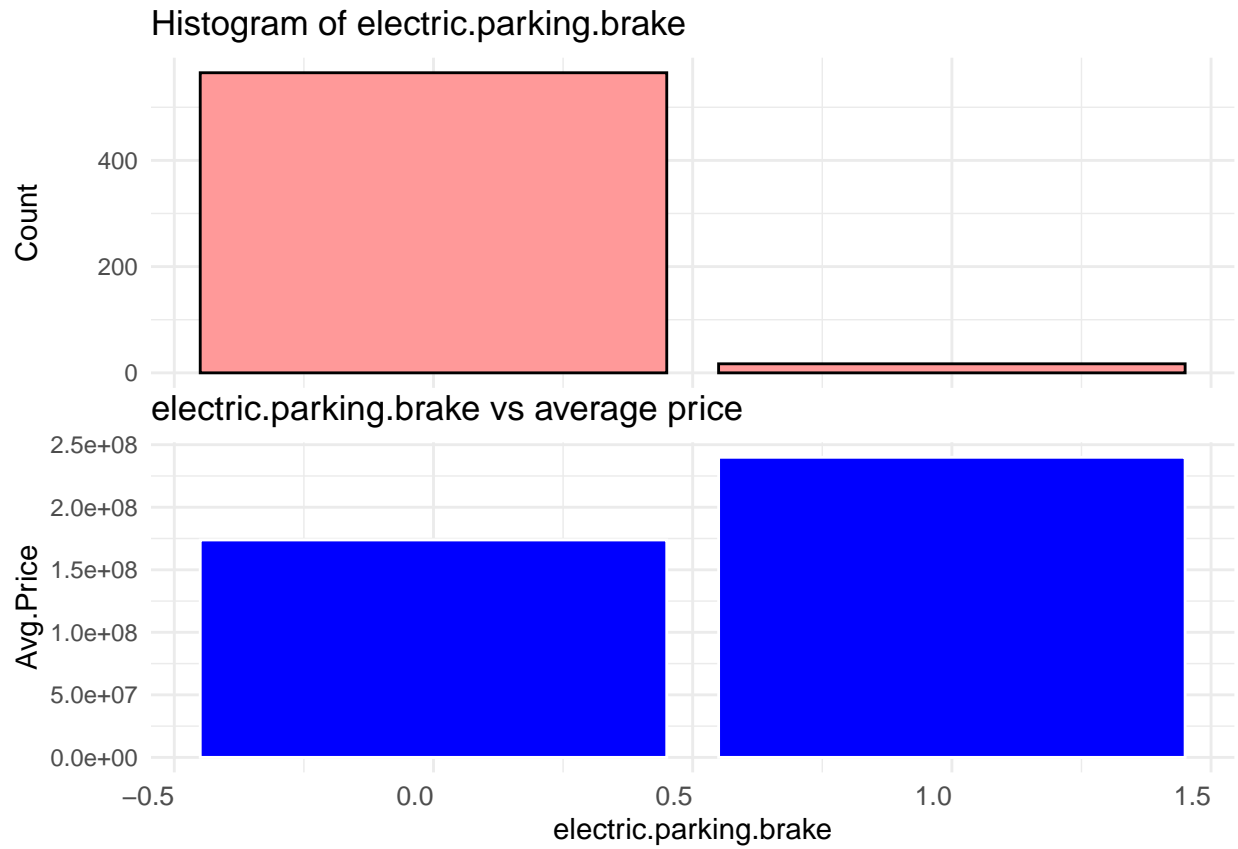
print(hist_plot / price_plot + plot_layout(guides = "collect", axis_titles = "collect"))
}

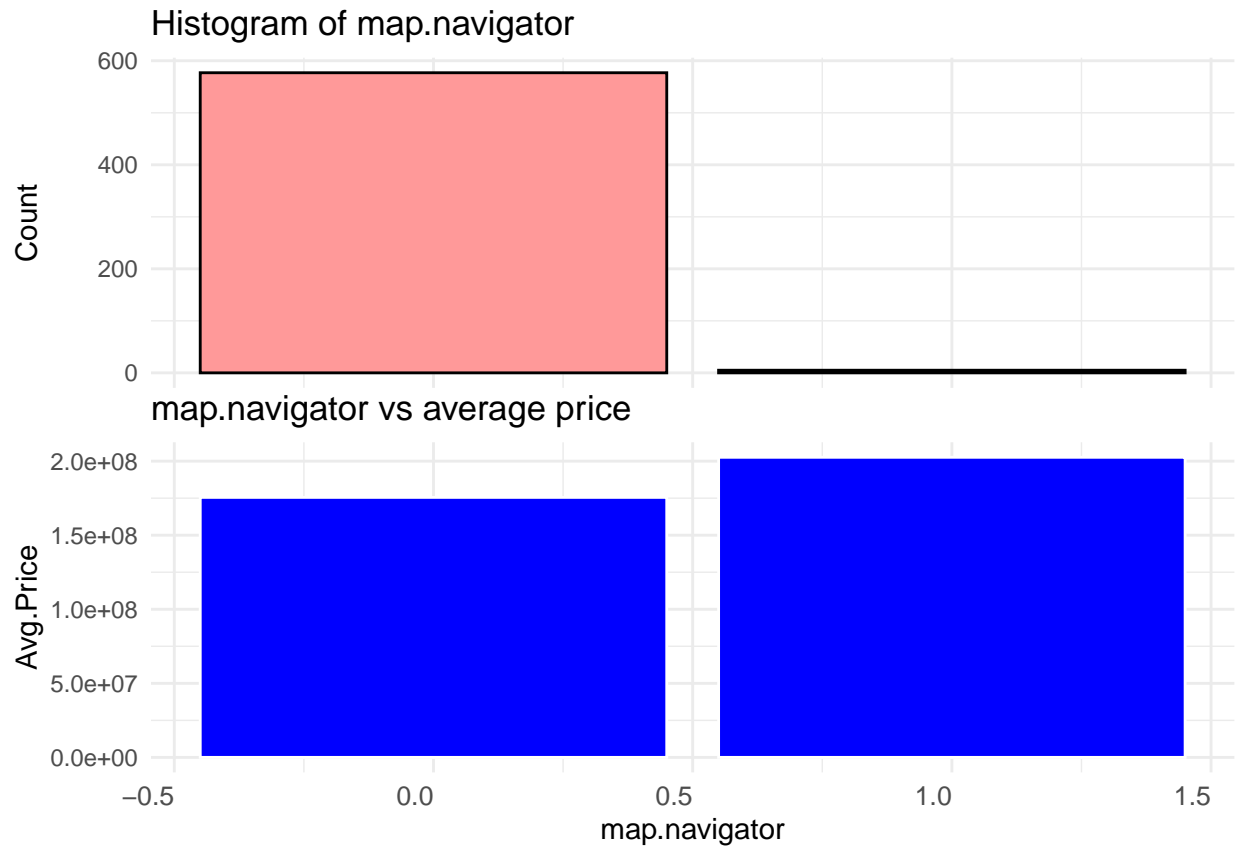
```



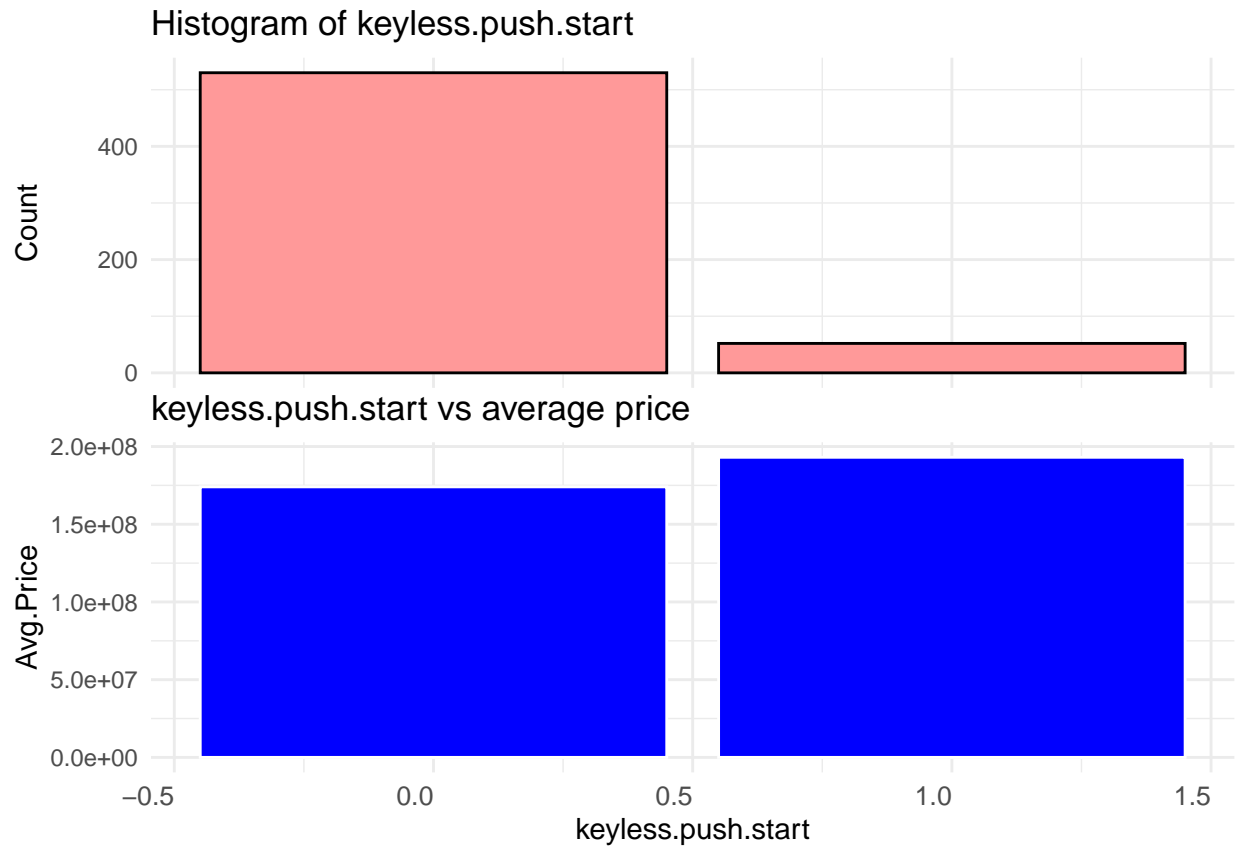


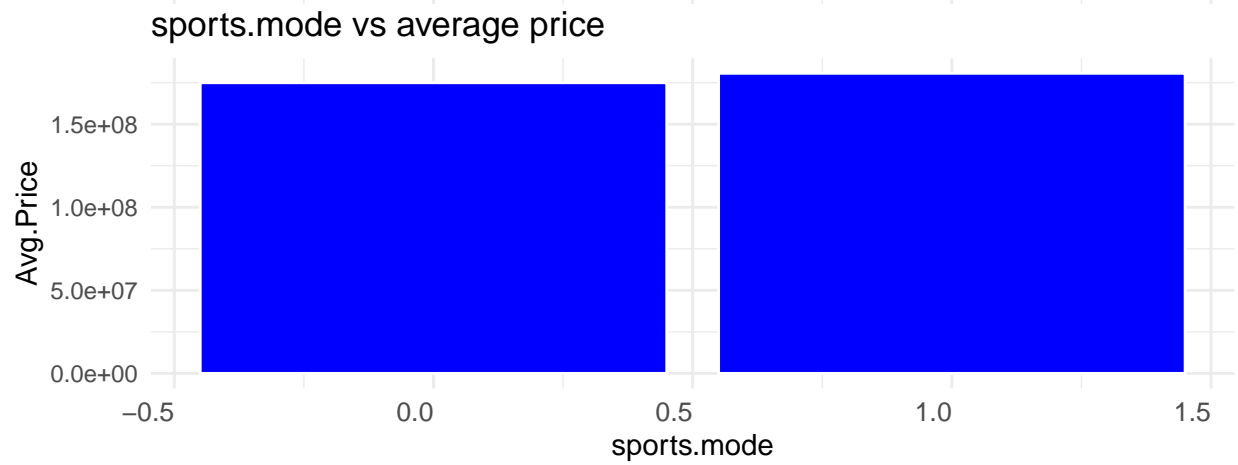
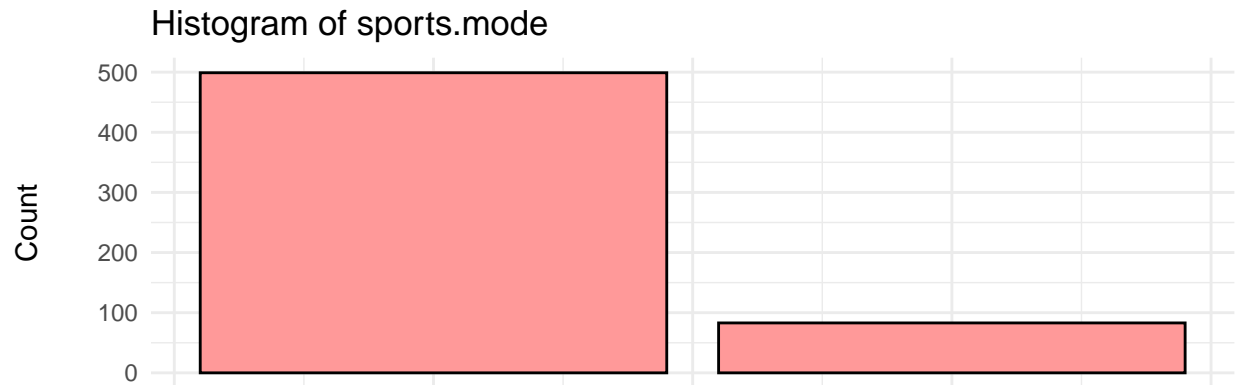


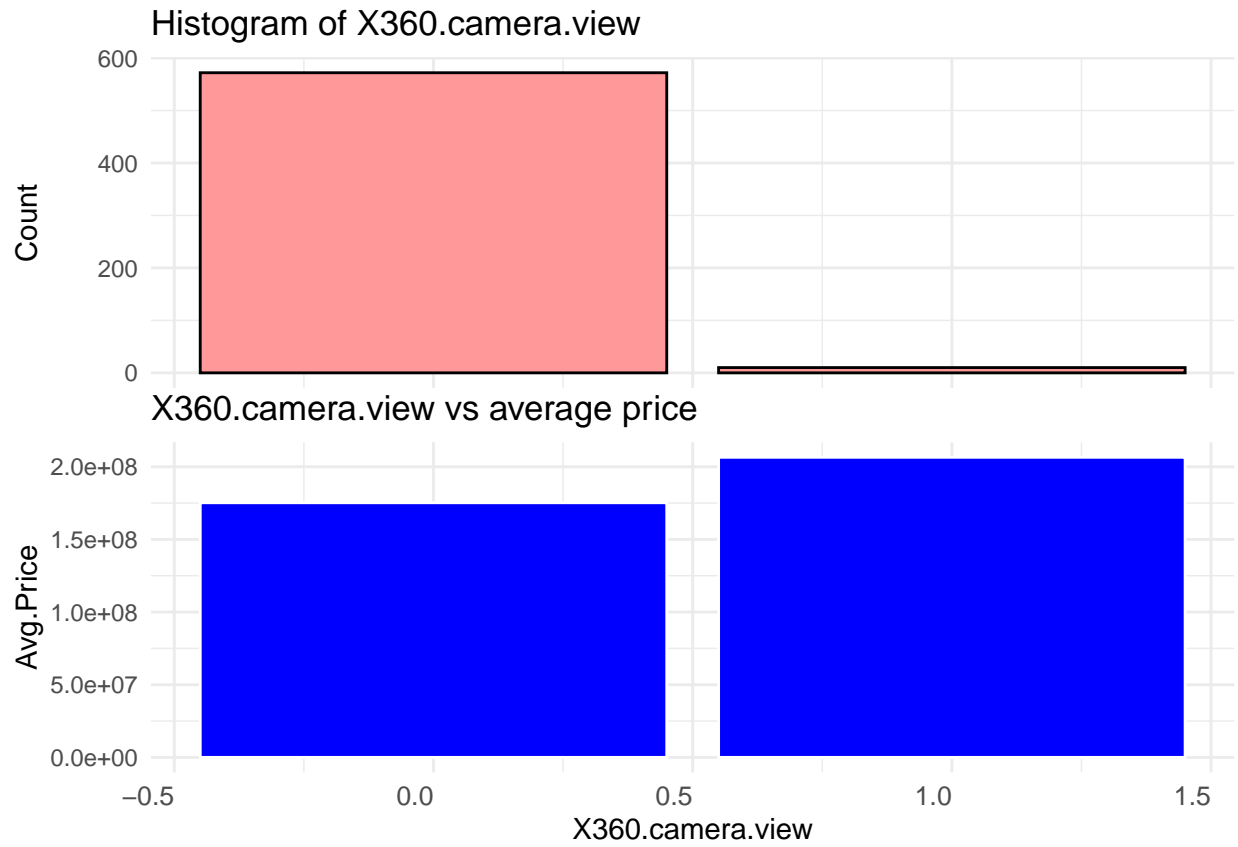


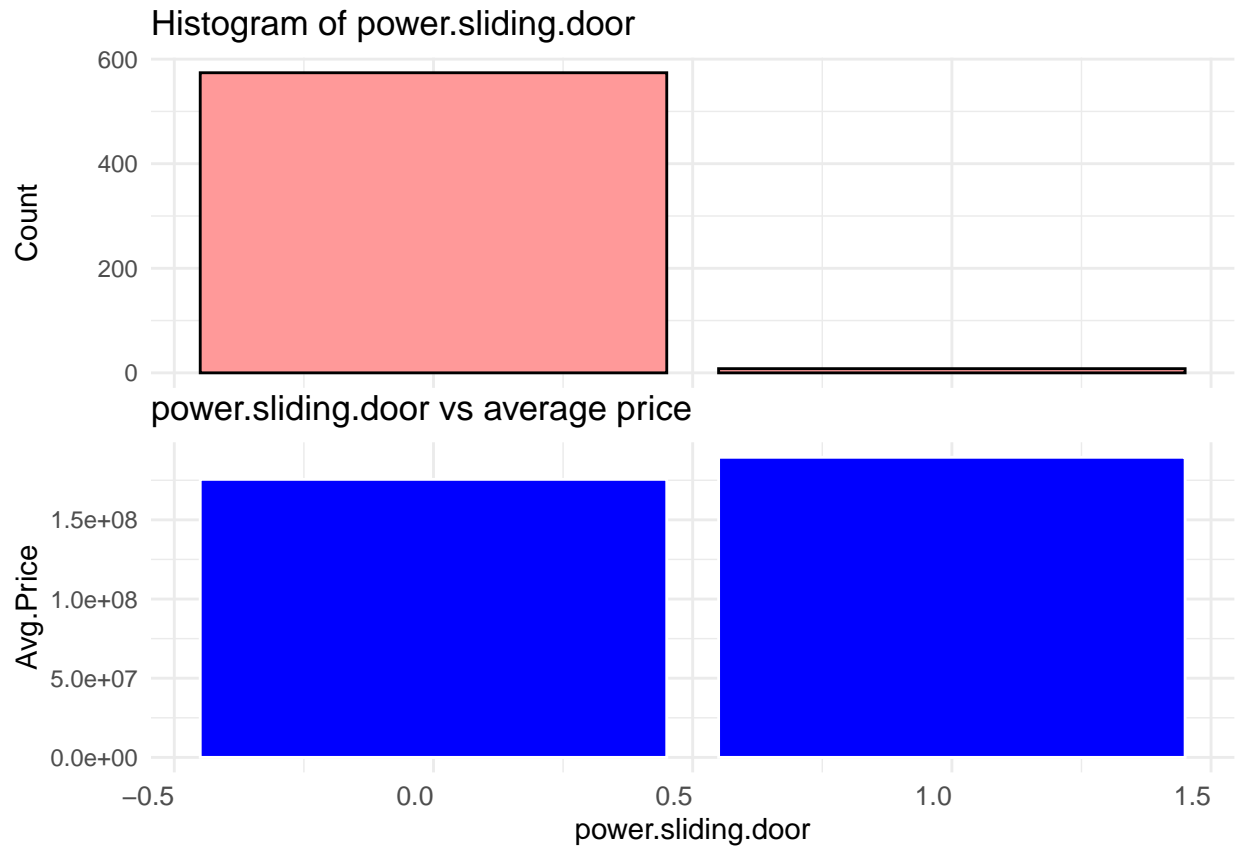


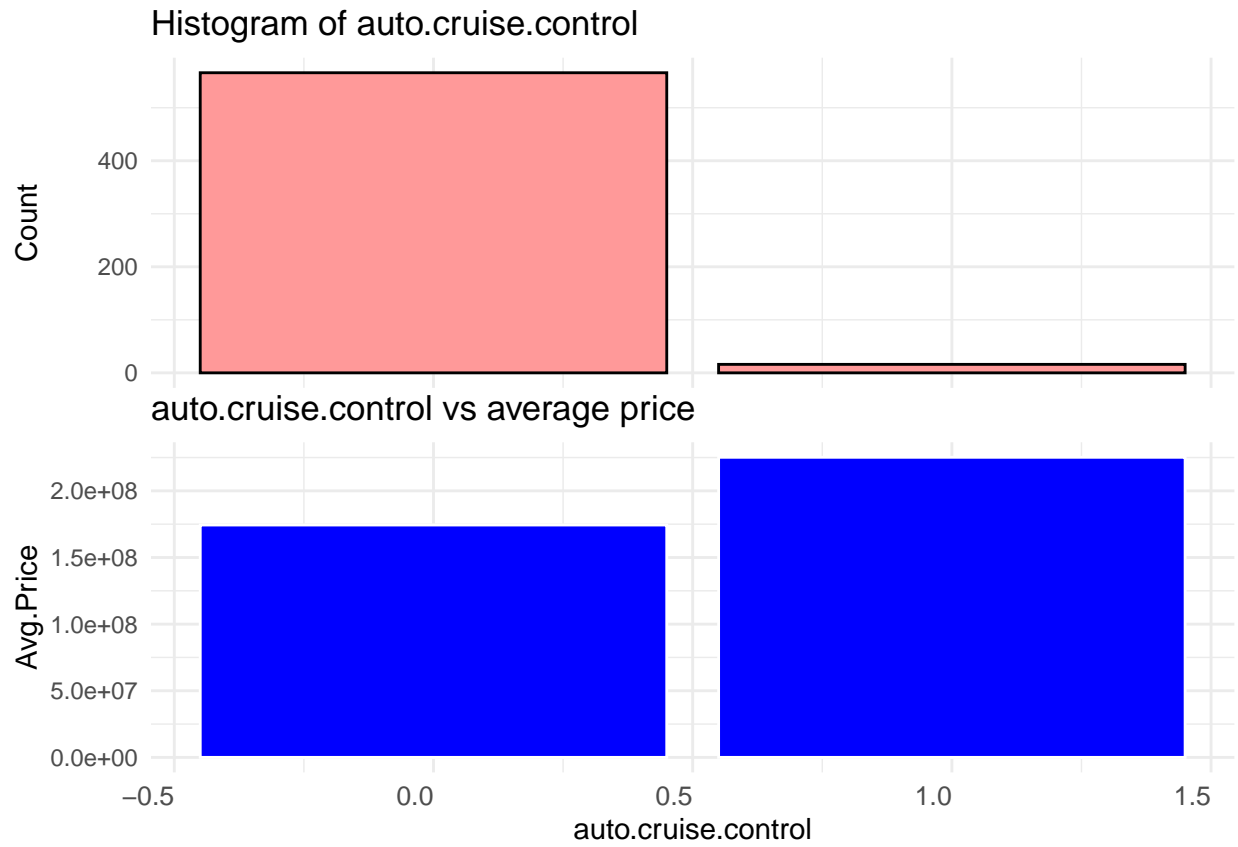


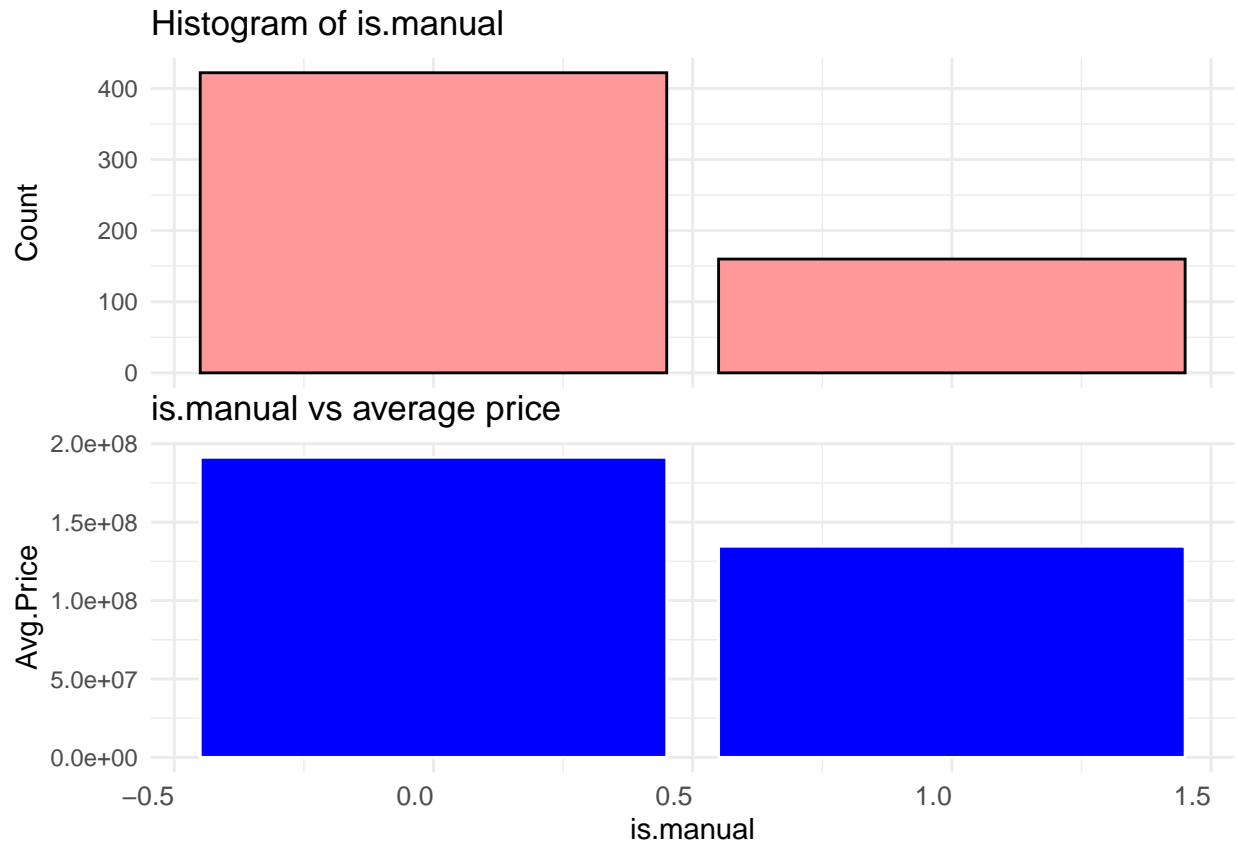


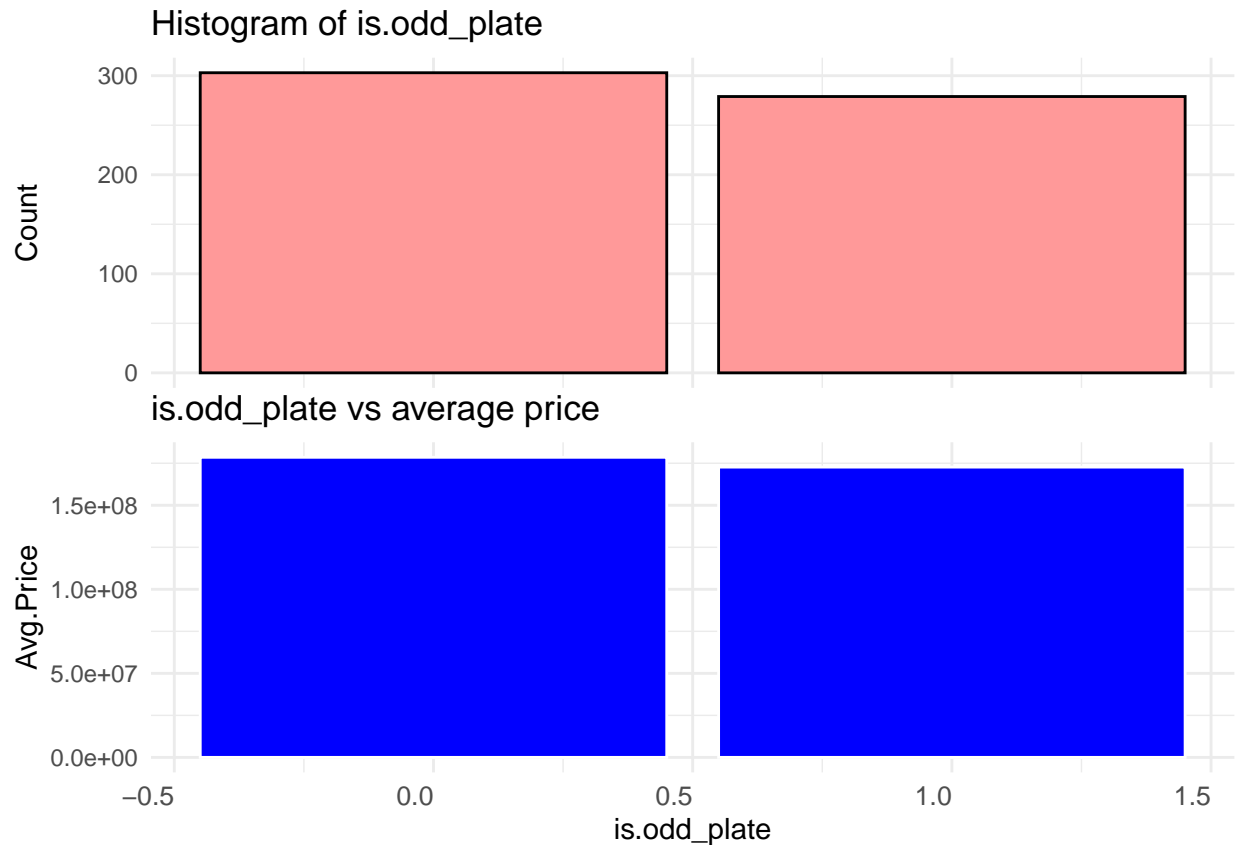












Categorical Variables EDA

```
cat_cols
```

```
## [1] "brand" "location"
```

```
for (col in cat_cols) {
  options(repr.plot.width = 20, repr.plot.height = 8)

  # Histogram
  hist_plot <- ggplot(df2, aes(x = fct_infreq(.data[[col]]))) +
    geom_bar(fill = '#FF9999', color = 'black') +
    labs(title = paste("Count of", col), y = "Count") +
    xlab(col) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, size = 10),
          plot.margin = margin(t = 0, unit = "pt"))

  # Bar plot (dependent variable)
  data <- df2 %>%
    group_by(.data[[col]]) %>%
    summarise(Avg.Price = mean(price), .groups = 'drop') %>%
```

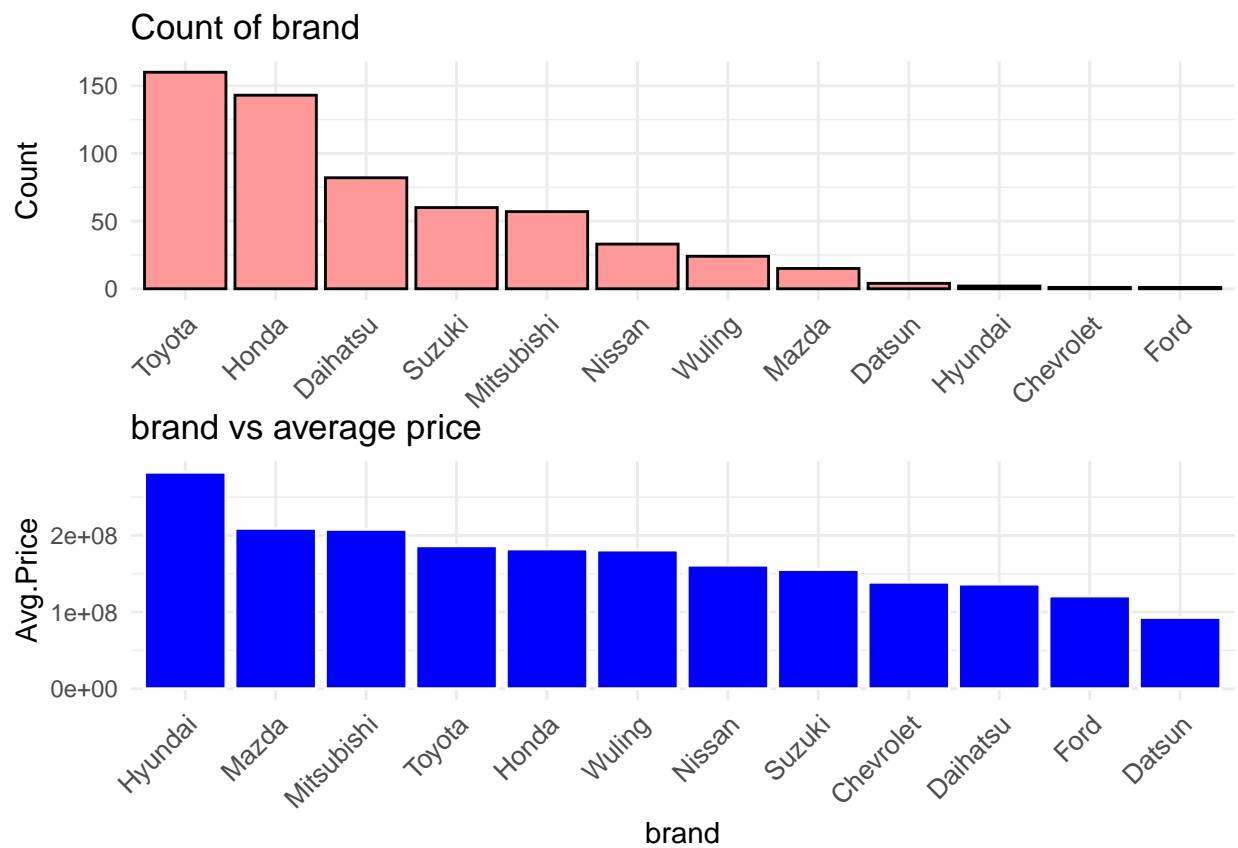
```

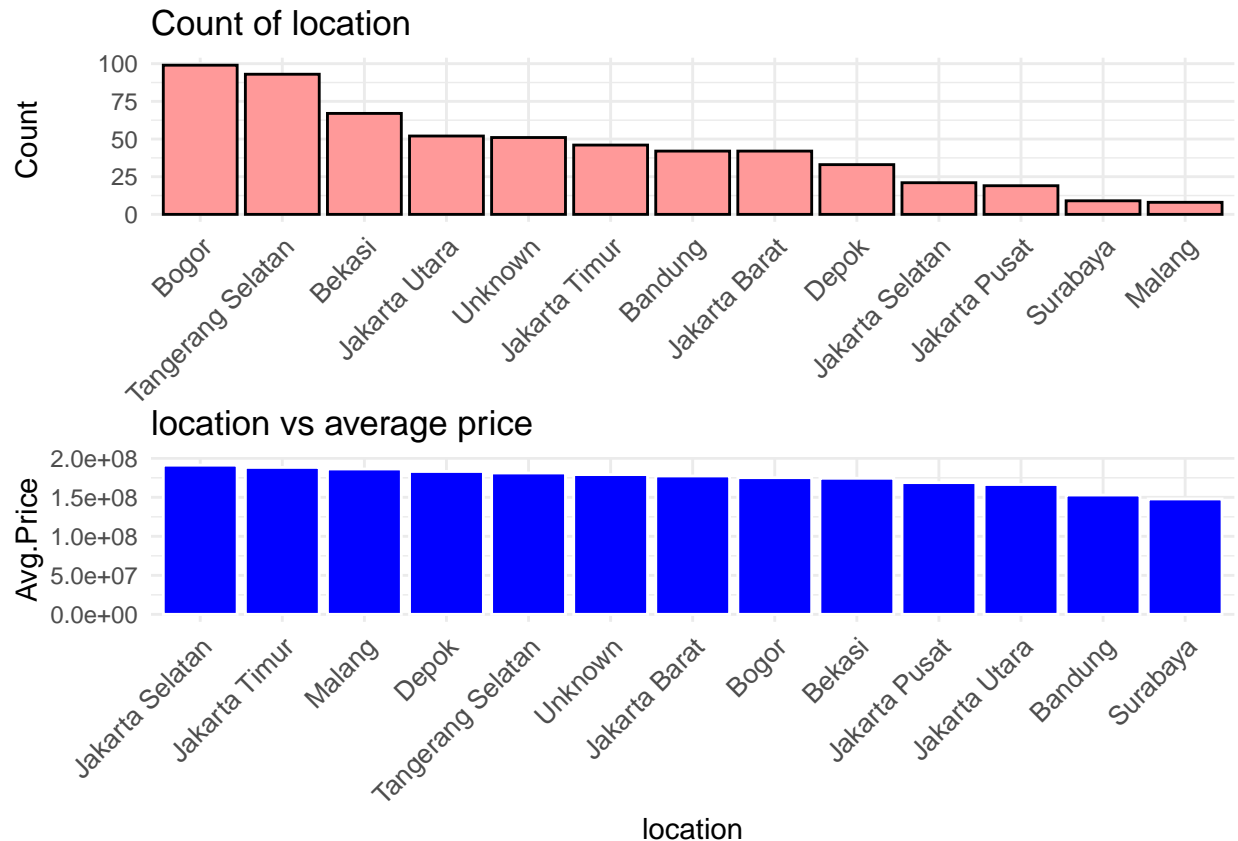
arrange(desc(Avg.Price)) %>%
mutate(!sym(col) := forcats::fct_reorder(!sym(col), Avg.Price, .desc = TRUE))

price_plot <- ggplot(data, aes(x = .data[[col]], y = Avg.Price)) +
  geom_bar(stat = 'identity', fill = 'blue', color = 'white') +
  labs(title = paste(col, "vs average price")) +
  xlab(col) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, size = 10),
        plot.margin = margin(t = 0, unit = "pt"))

print(hist_plot / price_plot + plot_layout(guides = "collect", axis_titles = "collect"))
}

```





Insights: - Toyota, Daihatsu, and Hyundai are the most common brand listed in the website. But interestingly, they don't hold the top highest average price. Hyundai, Mazda, and Mitsubishi are the top 3 brand with the highest average price despite their relatively low listing popularity. This can indicate that the aforementioned brands are the luxury brands. - The location variable doesn't seem to be affected the local price, as there isn't highly noticeable disparities of average car price based on locations.

Hedonic Model (Regression)

```
## For multicollinearity check
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.94 loaded
```

```
library(reshape2)
```

```
##
```

```
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
## smiths
```

```
## For assumption check
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
# For error distribution check
library("ggpubr")
```

```
## Warning: package 'ggpubr' was built under R version 4.3.3
```

For the hedonic model and machine learning, the dependent or predicted variable will be the “price” column.

```
# Prepare dependent variable
dep_var <- "price"
```

```
# Remove excluded numeric columns
remove_var <- c("instalment_month", "year")
```

```
# Prepare independent variables
ind_vars <- num_cols[num_cols != dep_var]
ind_vars <- ind_vars[!(ind_vars %in% remove_var)]
```

```
# For variable collection
reg_vars <- c(dep_var, ind_vars)
```

```
ind_vars
```

```
## [1] "mileage_km"          "rear.camera"
## [3] "sun.roof"            "auto.retract.mirror"
## [5] "electric.parking.brake" "map.navigator"
## [7] "vehicle.stability.control" "keyless.push.start"
## [9] "sports.mode"          "X360.camera.view"
## [11] "power.sliding.door"    "auto.cruise.control"
## [13] "engine_cap"           "age"
## [15] "brand_popularity"      "is.manual"
## [17] "is.odd_plate"
```

Checking Correlation

```
ind_df <- df2 %>%
  as.data.frame() %>% dplyr :: select(all_of(ind_vars))

res1 <- cor.mtest(ind_df, conf.level=0.95)
corrmatrix <-cor(ind_df)
```



```

melted_corr <- melt(corrmatrix)
colnames(melted_corr) <- c("Variable1", "Variable2", "Correlation")

filtered_corr <- subset(melted_corr, abs(Correlation) > 0.5 & abs(Correlation) < 1)
sorted_corr <- filtered_corr[order(-abs(filtered_corr$Correlation)), ]

print(sorted_corr)

```

```

##      Variable1 Variable2 Correlation
## 14          age mileage_km  0.6346377
## 222 mileage_km          age  0.6346377

```

The age and mileage variable are highly correlated, which is understandable as both variable can be a proxy for car use. I choose to keep the mileage variable as it is a more accurate variable to reflect car use.

```

remove_var <- c("age")
ind_vars <- ind_vars[!(ind_vars %in% remove_var)]
reg_vars <- c(dep_var, ind_vars)

```

Model selection using regression subset method

```

## Variable Selection
library(olsrr)

```

```
## Warning: package 'olsrr' was built under R version 4.3.3
```

```

##
## Attaching package: 'olsrr'

```

```

## The following object is masked from 'package:datasets':
##
##      rivers

```

```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.3.3
```

```

get_ols <- function(dataframe, vars, dep) {
  df1 <- dataframe[vars]
  fml <- as.formula(paste0(dep, " ~ ."))
  ols_res <- lm(data = df1, fml)
  return(ols_res)
}

```

```

ols_1 <- get_ols(df2, reg_vars, dep_var)
ols_regress(ols_1)

```

```

##                                     Model Summary
## -----
## R                                0.843      RMSE          28607393.173
## R-Squared                       0.710      MSE           8.183829e+14
## Adj. R-Squared                  0.702      Coef. Var       16.520
## Pred R-Squared                  0.677      AIC            21672.565
## MAE                             20736144.261    SBC            21751.161
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
## AIC: Akaike Information Criteria
## SBC: Schwarz Bayesian Criteria
##
##                                     ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression    1.166826e+18        16      7.292662e+16    86.508    0.0000
## Residual      4.762989e+17       565      8.430069e+14
## Total        1.643125e+18       581
## -----
##
##                                     Parameter Estimates
## -----
##                               model      Beta      Std. Error      Std. Beta      t      Sig.
## -----
## (Intercept)    16347002.691      7940841.481          2.059      0.040      74
## mileage_km     -401796.808       43532.633         -0.223     -9.230      0.000     -48
## rear.camera    5851665.858       3587997.888          0.039      1.631      0.103    -119
## sun.roof       20178651.759       8917743.683          0.054      2.263      0.024     266
## auto.retract.mirror 28408413.571       2844986.008          0.267      9.985      0.000    2282
## electric.parking.brake 44966344.536       7295819.246          0.143      6.163      0.000    3063
## map.navigator  23833040.279      13299964.484          0.041      1.792      0.074    -229
## vehicle.stability.control 10406505.712       4436621.428          0.060      2.346      0.019     169
## keyless.push.start 738803.882       4562749.691          0.004      0.162      0.871    -822
## sports.mode    -3600669.613       3834892.505         -0.024     -0.939      0.348   -1113
## X360.camera.view 1139100.255       9745015.650          0.003      0.117      0.907   -1800
## power.sliding.door -20585814.250      10499068.617         -0.045     -1.961      0.050   -4120
## auto.cruise.control 30728799.318       7717187.341          0.095      3.982      0.000    1557
## engine_cap     107410702.617       4975927.102          0.551     21.586      0.000    9763
## brand_popularity 1276021.735       151846.491          0.212      8.403      0.000      97
## is.manual      -31480460.010      3109348.295         -0.265    -10.124      0.000   -3758
## is.odd_plate   -2015523.343       2445100.843         -0.019     -0.824      0.410   -681
## -----

```

Insights: - The initial model using all of the selected variable can explain for 72% of the total price variation.
- Not all of the independent variable is significant to the price variable, which raises the concern of the model effectiveness.

Regression model selection - One of the main goal of this analysis is to make a prediction model using the hedonic model (OLS) and machine learning model, so I need to choose model with the best variable combination to predict the price variable. - I will use the best subsets regression using the exhaustive method; since the data is still relatively small (< 5000 rows) and the independent variables is still acceptable for an

exhaustive search (<20 variables). - Because the goal is to make a prediction model, I choose to select the model best the highest adjusted R squared value.

```
df_reg <- df2[reg_vars]

best_subset <-
  regsubsets(price~., data = df_reg, nbest = 1, nvmax = length(ind_vars),
             method="exhaustive")

summary_best_subset <- summary(best_subset)
as.data.frame(summary_best_subset$outmat)
```

```
##          mileage_km rear.camera sun.roof auto.retract.mirror
## 1 ( 1 )
## 2 ( 1 )
## 3 ( 1 )
## 4 ( 1 )
## 5 ( 1 )
## 6 ( 1 )
## 7 ( 1 )
## 8 ( 1 )
## 9 ( 1 )
## 10 ( 1 )
## 11 ( 1 )
## 12 ( 1 )
## 13 ( 1 )
## 14 ( 1 )
## 15 ( 1 )
## 16 ( 1 )
##          electric.parking.brake map.navigator vehicle.stability.control
## 1 ( 1 )
## 2 ( 1 )
## 3 ( 1 )
## 4 ( 1 )
## 5 ( 1 )
## 6 ( 1 )
## 7 ( 1 )
## 8 ( 1 )
## 9 ( 1 )
## 10 ( 1 )
## 11 ( 1 )
## 12 ( 1 )
## 13 ( 1 )
## 14 ( 1 )
## 15 ( 1 )
## 16 ( 1 )
##          keyless.push.start sports.mode X360.camera.view power.sliding.door
## 1 ( 1 )
## 2 ( 1 )
## 3 ( 1 )
## 4 ( 1 )
## 5 ( 1 )
## 6 ( 1 )
## 7 ( 1 )
```

```

## 8 ( 1 )
## 9 ( 1 )
## 10 ( 1 )
## 11 ( 1 )
## 12 ( 1 )
## 13 ( 1 )
## 14 ( 1 )
## 15 ( 1 )
## 16 ( 1 )
##      auto.cruise.control engine_cap brand_popularity is.manual
## 1 ( 1 )
## 2 ( 1 )
## 3 ( 1 )
## 4 ( 1 )
## 5 ( 1 )
## 6 ( 1 )
## 7 ( 1 )
## 8 ( 1 )
## 9 ( 1 )
## 10 ( 1 )
## 11 ( 1 )
## 12 ( 1 )
## 13 ( 1 )
## 14 ( 1 )
## 15 ( 1 )
## 16 ( 1 )
##      is.odd_plate
## 1 ( 1 )
## 2 ( 1 )
## 3 ( 1 )
## 4 ( 1 )
## 5 ( 1 )
## 6 ( 1 )
## 7 ( 1 )
## 8 ( 1 )
## 9 ( 1 )
## 10 ( 1 )
## 11 ( 1 )
## 12 ( 1 )
## 13 ( 1 )
## 14 ( 1 )
## 15 ( 1 )
## 16 ( 1 )

```

```
which.max(summary_best_subset$adjr2)
```

```
## [1] 12
```

```
summary_best_subset$which[12,]
```

```

##      (Intercept)      mileage_km      rear.camera
##      TRUE      TRUE      TRUE
##      sun.roof      auto.retract.mirror      electric.parking.brake

```

```
##                TRUE                TRUE                TRUE
##      map.navigator vehicle.stability.control      keyless.push.start
##                TRUE                TRUE                FALSE
##      sports.mode      X360.camera.view      power.sliding.door
##                FALSE                FALSE                TRUE
##      auto.cruise.control      engine_cap      brand_popularity
##                TRUE                TRUE                TRUE
##      is.manual      is.odd_plate
##                TRUE                FALSE

best_vars <- c("price", "mileage_km", "rear.camera", "sun.roof",
              "auto.retract.mirror", "electric.parking.brake", "map.navigator",
              "vehicle.stability.control", "power.sliding.door",
              "auto.cruise.control", "engine_cap", "brand_popularity", "is.manual")

best_model1 <- (price ~ mileage_km + rear.camera
              + sun.roof + auto.retract.mirror + electric.parking.brake
              + map.navigator + vehicle.stability.control
              + power.sliding.door + auto.cruise.control + engine_cap
              + brand_popularity + is.manual)

best_reg1 <- lm(df2, formula = price ~ mileage_km + rear.camera
              + sun.roof + auto.retract.mirror + electric.parking.brake
              + map.navigator + vehicle.stability.control
              + power.sliding.door + auto.cruise.control + engine_cap
              + brand_popularity + is.manual)
```

Checking for BLUE assumption

```
library("tseries")

## Warning: package 'tseries' was built under R version 4.3.3

## Registered S3 method overwritten by 'quantmod':
##   method      from
##   as.zoo.data.frame zoo

## Pooled model regression
library(plm)

## Warning: package 'plm' was built under R version 4.3.3

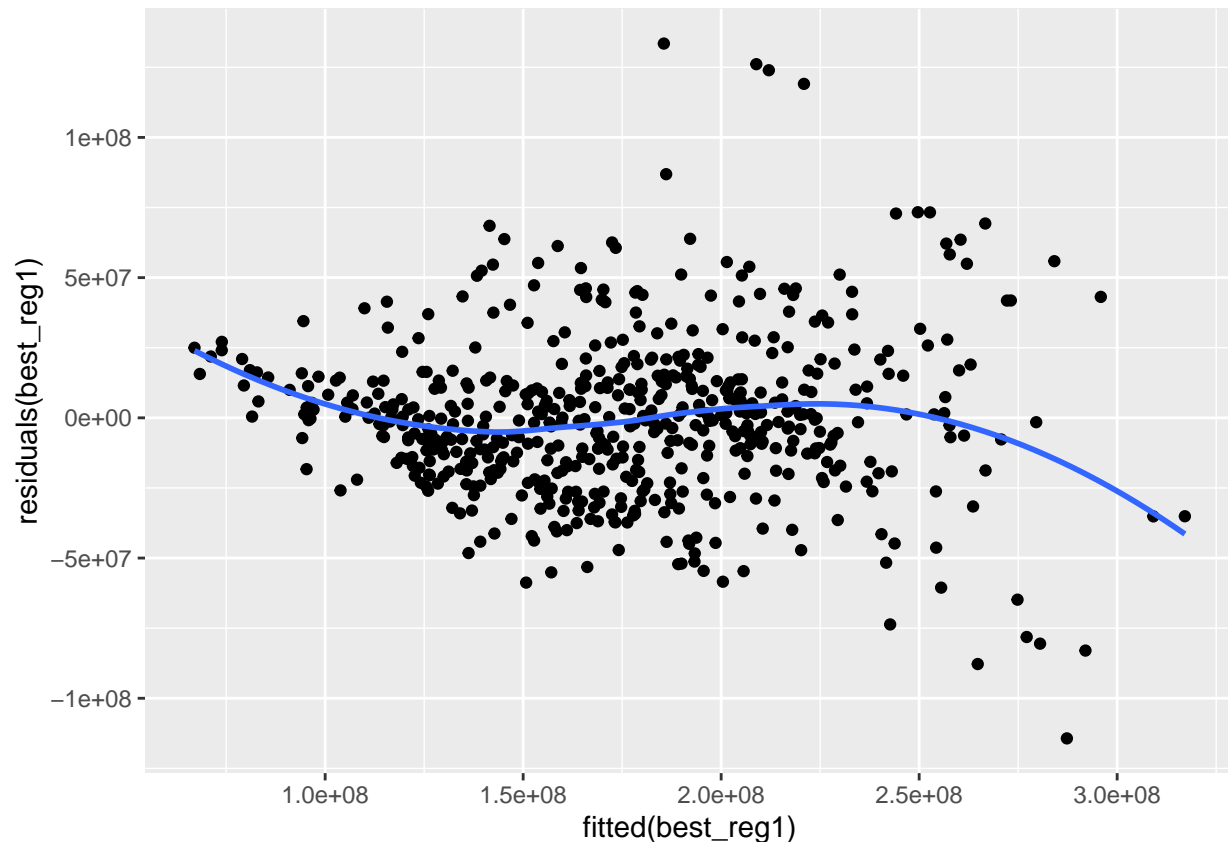
##
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':
##
##   between, lag, lead
```

Linearity

```
ggplot(df2, aes(x=fitted(best_reg1), y= residuals(best_reg1)))+  
  geom_point() +  
  geom_smooth(method = "loess", se = FALSE)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



The residuals dispersion have an U concave pattern; This suggest non-linearity between the dependent and independent variables's relationship.

```
resettest(best_reg1, power=2:3, type="fitted")
```

```
##  
## RESET test  
##  
## data: best_reg1  
## RESET = 12.691, df1 = 2, df2 = 567, p-value = 4.058e-06
```

The test shows that th relationship between the dependent and independent variables is not linear. But the test result could be caused by other BLUE violation and not because of linearity.

Multicollinearity check

```
car::vif(best_reg1)
```

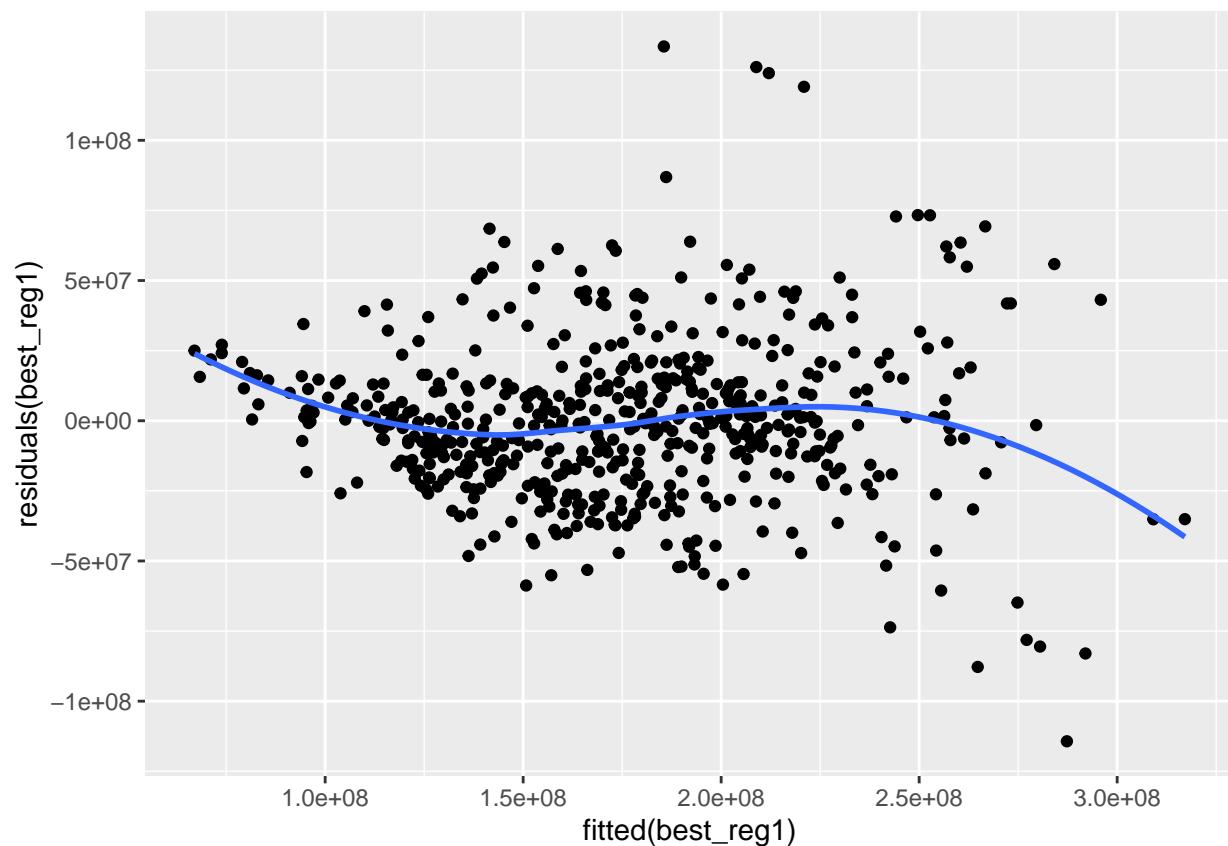
```
##           mileage_km           rear.camera           sun.roof
##           1.125649           1.074461           1.096639
##   auto.retract.mirror   electric.parking.brake   map.navigator
##           1.261161           1.025654           1.023198
## vehicle.stability.control   power.sliding.door   auto.cruise.control
##           1.191824           1.022036           1.063158
##           engine_cap           brand_popularity           is.manual
##           1.241499           1.117748           1.195983
```

There is no variable with VIF value > 5 , so it can be concluded that there is no multicollinearity.

Homoskedasticity

```
ggplot(df2, aes(x=fitted(best_reg1), y= residuals(best_reg1)))+
  geom_point() +
  geom_smooth(method = "loess", se = FALSE)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



Insights: - There residuals distribution is not constant along the fitted value. - There is a down concave

pattern of the residuals and fitted value correlation. - The models residuals seems to increase in the relatively low and high fitted price.

```
ols_test_breusch_pagan(best_reg1, rhs = TRUE, multiple = TRUE)
```

```
##
## Breusch Pagan Test for Heteroskedasticity
## -----
## Ho: the variance is constant
## Ha: the variance is not constant
##
##
## -----
## Response : price
## Variables: mileage_km rear.camera sun.roof auto.retract.mirror electric.parking.brake map.navigator
##
## Test Summary (Unadjusted p values)
## -----
## Variable chi2 df p
## -----
## mileage_km 0.10720062 1 7.433540e-01
## rear.camera 0.02541785 1 8.733303e-01
## sun.roof 65.98746303 1 4.538003e-16
## auto.retract.mirror 1.31532534 1 2.514327e-01
## electric.parking.brake 2.73675725 1 9.806358e-02
## map.navigator 50.60662575 1 1.128634e-12
## vehicle.stability.control 14.29135435 1 1.565824e-04
## power.sliding.door 3.61932648 1 5.711199e-02
## auto.cruise.control 5.64246442 1 1.753055e-02
## engine_cap 128.13458395 1 1.048841e-29
## brand_popularity 4.54774923 1 3.296205e-02
## is.manual 15.17130596 1 9.818412e-05
## -----
## simultaneous 268.55382508 12 1.825811e-50
## -----
```

- The p value is less than 0.05, which means the null hypothesis (homoscedasticity) is rejected and therefore the model have a heteroscedasticity problem.

Conclusion and Method Selection

```
ols_regress(best_reg1)
```

```
## Model Summary
## -----
## R 0.842 RMSE 28649295.236
## R-Squared 0.709 MSE 8.207821e+14
## Adj. R-Squared 0.703 Coef. Var 16.486
## Pred R-Squared 0.684 AIC 21666.269
## MAE 20783773.203 SBC 21727.399
## -----
```



```

## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
## AIC: Akaike Information Criteria
## SBC: Schwarz Bayesian Criteria
##
##
## ANOVA
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression    1.16543e+18      12      9.711914e+16    115.682    0.0000
## Residual      4.776952e+17     569      8.395346e+14
## Total         1.643125e+18     581
## -----
##
## Parameter Estimates
## -----
##              model              Beta      Std. Error      Std. Beta      t      Sig.
## -----
##              (Intercept)    14889178.724      7779713.422              1.914    0.056    -39
##              mileage_km     -397208.337       43237.269             -0.220    -9.187    0.000    -48
##              rear.camera     5561032.522      3525237.413              0.037     1.577    0.115   -136
##              sun.roof        20653135.804      8850840.541              0.055     2.333    0.020    326
##              auto.retract.mirror 29197176.119      2703965.670              0.274    10.798    0.000   2388
##              electric.parking.brake 45501429.112      7223247.750              0.144     6.299    0.000   3131
##              map.navigator    24198955.560      13163992.343              0.042     1.838    0.067   -165
##              vehicle.stability.control 10674936.424      4280584.852              0.062     2.494    0.013    226
##              power.sliding.door -19664019.571      10428286.297             -0.043    -1.886    0.060  -4014
##              auto.cruise.control 31193800.914       7573757.787              0.096     4.119    0.000   1631
##              engine_cap      107236668.392      4907160.860              0.550    21.853    0.000   9759
##              brand_popularity  1242257.577       143591.798              0.207     8.651    0.000    96
##              is.manual       -30616286.240      2941894.255             -0.257   -10.407    0.000  -3639
## -----

```

Insights: - The model can explain 73% of the total data variations. - The non - significant variable is rear.camera - rear.camera

From the BLUE assumption checking, I conclude that the model have: - Non-linear Relationship - Heteroscedasticity - Heteroscedasticity don't affect the coefficient value.

Regression Option: - Variable transformation - WLS

Alternative Model

Double Log Model

```
best_vars
```

```

## [1] "price"              "mileage_km"
## [3] "rear.camera"        "sun.roof"
## [5] "auto.retract.mirror" "electric.parking.brake"

```

```
## [7] "map.navigators" "vehicle.stability.control"
## [9] "power.sliding.door" "auto.cruise.control"
## [11] "engine_cap" "brand_popularity"
## [13] "is.manual"
```

```
for (col in cont_cols){
  if (col %in% best_vars == TRUE){
    print(paste0(col, "'s summary"))
    print(summary(df[[col]]))
    cat("\n")
  }else{
    print(paste0(col, "is not in the model"))
    cat("\n")
  }
}
```

```
## [1] "mileage_km's summary"
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.065  39.145  58.365  61.686  81.726 245.000
##
## [1] "price's summary"
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 7.70e+07 1.32e+08 1.77e+08 1.83e+08 2.12e+08 5.42e+08
```

There is only 2 variable in the variable that is continuous: “mileage_km’s summary” + “price’s summary” and both: - doesn’t have 0 - not a percentage

Which means that they are eligible for natural log transformation.

```
## Set dataframe
df_log <- data.frame(df_reg)

## Prepare log variable model
dbl_log_vars <- best_vars

dbl_log_vars
```

```
## [1] "price" "mileage_km"
## [3] "rear.camera" "sun.roof"
## [5] "auto.retract.mirror" "electric.parking.brake"
## [7] "map.navigators" "vehicle.stability.control"
## [9] "power.sliding.door" "auto.cruise.control"
## [11] "engine_cap" "brand_popularity"
## [13] "is.manual"
```

```
for (col in cont_cols){
  if (col %in% best_vars == TRUE){
    ## Natural Log transformation
    new_var <- paste0("ln_", col)
    ## Load to new dataframe
    df_log[[new_var]] <- log(df_log[[col]])
    df_log %>% select(-any_of(c(col)))
    print(paste0(col, "'transformed to ", new_var))
  }
```

```

## Input to new model variable
dbl_log_vars <- c(dbl_log_vars, new_var)
dbl_log_vars <- dbl_log_vars[dbl_log_vars != col]
cat("\n")
}else{
  print(paste0(col,"is not in the model"))
  cat("\n")
}
}

```

```

## [1] "mileage_km'transformed to ln_mileage_km"
##
## [1] "price'transformed to ln_price"

```

```

# Prepare dependent variable
ln_dep_var <- "ln_price"

# Prepare independent variables
ln_ind_vars <- dbl_log_vars[!(dbl_log_vars %in% ln_dep_var)]

## Combine vars
ln_vars <- c(ln_dep_var, ln_ind_vars)

```

```

ln_ols <- get_ols(df_log, dbl_log_vars, ln_dep_var)
ols_regress(ln_ols)

```

```

##                               Model Summary
## -----
## R                               0.848      RMSE                0.161
## R-Squared                       0.720      MSE                0.026
## Adj. R-Squared                   0.714      Coef. Var           0.860
## Pred R-Squared                   0.701      AIC                 -446.161
## MAE                             0.123      SBC                 -385.030
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
## AIC: Akaike Information Criteria
## SBC: Schwarz Bayesian Criteria
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      38.721      12      3.227      121.689      0.0000
## Residual        15.088      569      0.027
## Total           53.809      581
## -----
##
##                               Parameter Estimates
## -----

```

	model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
	(Intercept)	18.286	0.058		317.281	0.000	18.172	18.399
	rear.camera	0.043	0.020	0.050	2.162	0.031	0.004	0.081
	sun.roof	0.136	0.050	0.064	2.744	0.006	0.039	0.233
	auto.retract.mirror	0.190	0.015	0.312	12.557	0.000	0.160	0.220
	electric.parking.brake	0.239	0.041	0.132	5.881	0.000	0.159	0.319
	map.navigators	0.109	0.074	0.033	1.478	0.140	-0.036	0.254
	vehicle.stability.control	0.085	0.024	0.085	3.520	0.000	0.037	0.133
	power.sliding.door	-0.065	0.059	-0.025	-1.110	0.267	-0.180	0.050
	auto.cruise.control	0.189	0.043	0.102	4.435	0.000	0.105	0.273
	engine_cap	0.546	0.027	0.490	20.142	0.000	0.493	0.600
	brand_popularity	0.007	0.001	0.207	8.832	0.000	0.006	0.008
	is.manual	-0.203	0.017	-0.298	-12.198	0.000	-0.236	-0.170
	ln_mileage_km	-0.080	0.011	-0.171	-7.355	0.000	-0.102	-0.058

The double log model can explain about 72% of the data. The double logarithmic model accuracy decreases about 10% from the ols. The non significant variables are: - map.navigators - power sliding door

```
ols_test_breusch_pagan(ln_ols, rhs = TRUE, multiple = TRUE)
```

```
##
## Breusch Pagan Test for Heteroskedasticity
## -----
## Ho: the variance is constant
## Ha: the variance is not constant
##
##
## -----
## Response : ln_price
## Variables: rear.camera sun.roof auto.retract.mirror electric.parking.brake map.navigators vehicle.stability.control
##
## Test Summary (Unadjusted p values)
## -----
## Variable chi2 df p
## -----
## rear.camera 0.06546925 1 7.980515e-01
## sun.roof 30.94859828 1 2.649534e-08
## auto.retract.mirror 5.30181905 1 2.130317e-02
## electric.parking.brake 0.18183007 1 6.698052e-01
## map.navigators 6.04722873 1 1.392814e-02
## vehicle.stability.control 13.85108062 1 1.978829e-04
## power.sliding.door 3.49694779 1 6.148204e-02
## auto.cruise.control 6.61813314 1 1.009456e-02
## engine_cap 41.08808954 1 1.455204e-10
## brand_popularity 14.60382143 1 1.326453e-04
## is.manual 0.15122787 1 6.973647e-01
## ln_mileage_km 13.47687925 1 2.415212e-04
## -----
## simultaneous 137.23971447 12 2.158642e-23
## -----
```

Although Reduced, there is still a heteroskedasticity problem.

```
resettest(ln_ols, power=2:3, type="fitted")
```

```
##
## RESET test
##
## data: ln_ols
## RESET = 18.429, df1 = 2, df2 = 567, p-value = 1.762e-08
```

Moreover, the RESET test showed that the relationship between the independent and dependent variable is still non-linear. But this could be a result of heteroskedasticity.

Conclusion

The double log model improves the model accuracy, but it still has heteroskedasticity and non linearity.

Alternative Model: - Weighted Least Squares (WLS) - Robust Standard Errors (RBS)

Weighted Least Squares (WLS) Regression

```
## Define Weight
wt <- 1 / lm(abs(ln_ols$residuals) ~ ln_ols$fitted.values)$fitted.values^2

df_1 <- df_log[dbl_log_vars]

#perform weighted least squares regression
wls_model <- lm(ln_price ~ ., data = df_1, weights=wt)

summary(wls_model)
```

```
##
## Call:
## lm(formula = ln_price ~ ., data = df_1, weights = wt)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0869 -0.8277  0.0204  0.7432  4.7458
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    18.2504333   0.0584187  312.407 < 2e-16 ***
## rear.camera      0.0423550   0.0198593   2.133 0.033373 *
## sun.roof         0.1541410   0.0526193   2.929 0.003533 **
## auto.retract.mirror 0.1910083   0.0152678  12.511 < 2e-16 ***
## electric.parking.brake 0.2403460   0.0442277   5.434 8.17e-08 ***
## map.navigator    0.1089611   0.0760995   1.432 0.152743
## vehicle.stability.control 0.0857689   0.0249902   3.432 0.000642 ***
## power.sliding.door -0.0663132   0.0602876  -1.100 0.271820
## auto.cruise.control 0.1876771   0.0457142   4.105 4.63e-05 ***
## engine_cap       0.5716018   0.0286411  19.957 < 2e-16 ***
## brand_popularity  0.0071417   0.0008016   8.909 < 2e-16 ***
```

```
## is.manual          -0.1993128  0.0158314 -12.590 < 2e-16 ***
## ln_mileage_km      -0.0810379  0.0107947  -7.507 2.35e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.323 on 569 degrees of freedom
## Multiple R-squared:  0.7303, Adjusted R-squared:  0.7246
## F-statistic: 128.4 on 12 and 569 DF,  p-value: < 2.2e-16
```

The model can explain 72% of the data variance.

The non significant variables are: - map navigator - power sliding doors.

```
ols_test_breusch_pagan(wls_model, rhs = TRUE, multiple = TRUE)
```

```
##
## Breusch Pagan Test for Heteroskedasticity
## -----
## Ho: the variance is constant
## Ha: the variance is not constant
##
## -----
## Response : ln_price
## Variables: rear.camera sun.roof auto.retract.mirror electric.parking.brake map.navigator vehicle.st
##
## Test Summary (Unadjusted p values)
## -----
## Variable                chi2      df      p
## -----
## rear.camera              0.09506431    1 7.578350e-01
## sun.roof                 34.59175702    1 4.066245e-09
## auto.retract.mirror      4.45433742    1 3.481272e-02
## electric.parking.brake   0.14629461    1 7.021018e-01
## map.navigator            7.08560872    1 7.770542e-03
## vehicle.stability.control 13.75758001    1 2.079799e-04
## power.sliding.door       3.49645000    1 6.150053e-02
## auto.cruise.control      6.66247056    1 9.846431e-03
## engine_cap              51.69194733    1 6.492851e-13
## brand_popularity        16.76108054    1 4.239397e-05
## is.manual                0.20848716    1 6.479557e-01
## ln_mileage_km           14.25276529    1 1.598258e-04
## -----
## simultaneous           154.57404609   12 6.678311e-27
## -----
```

Hedonic Model Recap

```
get_rmse <- function(model) {
  ssr <- sum(residuals(model)^2)
  df_residual <- model$df.residual
  rmse <- sqrt(ssr/df_residual)
```

```

    return(rmse)
}

print(paste0("Initial model's adj.R2: ", summary(ols_1)$adj.r.squared))

## [1] "Initial model's adj.R2: 0.701917367781985"

print(paste0("Best model's adj.R2: ", summary(best_reg1)$adj.r.squared))

## [1] "Best model's adj.R2: 0.70314513539943"

print(paste0("Double log + best model's adj.R2: ", summary(ln_ols)$adj.r.squared))

## [1] "Double log + best model's adj.R2: 0.713689473676195"

print(paste0("WLS + Double log + Best OLS Model's adj.R2: ", summary(wls_model)$adj.r.squared))

## [1] "WLS + Double log + Best OLS Model's adj.R2: 0.724565499658563"

print(paste0("Initial model's MSE: ", get_rmse(ols_1)))

## [1] "Initial model's MSE: 29034580.34404"

print(paste0("Best model's MSE: ", get_rmse(best_reg1)))

## [1] "Best model's MSE: 28974723.6186467"

print(paste0("Double log + best model's MSE: ", get_rmse(ln_ols)))

## [1] "Double log + best model's MSE: 0.162838182708786"

print(paste0("WLS + Double log + Best OLS Model's MSE: ", get_rmse(wls_model)))

## [1] "WLS + Double log + Best OLS Model's MSE: 0.163010223967787"

```

The best performing model is WLS model using log transformed's best variable selection. But all of the model have these problems: - Heteroskedasticity - Non Linear Relationship

Machine Learning

Train test split

```

set.seed(1111)
### Shuffle the dataframe
df3 <- df_log[ln_vars]

df3 <- df3[sample(1:nrow(df3)), ]
## Split to test and train
sample <- sample(c(TRUE, FALSE), nrow(df3), replace=TRUE, prob=c(0.8,0.2))
train  <- df3[sample, ]
test   <- df3[!sample, ]

```

Regression Tree

```
library(MASS)
```

```

##
## Attaching package: 'MASS'

## The following object is masked from 'package:olsrr':
##
##      cement

## The following object is masked from 'package:patchwork':
##
##      area

## The following object is masked from 'package:dplyr':
##
##      select

```

```

## Regression Tree
library(rpart)
library(rpart.plot)

```

```
## Warning: package 'rpart.plot' was built under R version 4.3.3
```

```

##For parameter tuning
library(caret)

```

```

## Warning: package 'caret' was built under R version 4.3.3

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##      lift

```



```

## Set seed
set.seed(1111)

## Pre pruning
tree_1 <- rpart(ln_price ~ ., data = train, method = "anova",
                control = rpart.control(cp = 0.01, minsplit = 20, maxdepth=30))
summary(tree_1)

## Call:
## rpart(formula = ln_price ~ ., data = train, method = "anova",
##       control = rpart.control(cp = 0.01, minsplit = 20, maxdepth = 30))
##      n= 462
##
##              CP nsplit rel error      xerror      xstd
## 1  0.43479397    0 1.0000000 1.0035593 0.05372238
## 2  0.07405191    1 0.5652060 0.5675610 0.04314683
## 3  0.05698300    2 0.4911541 0.5192641 0.04188884
## 4  0.04448144    3 0.4341711 0.4845430 0.03644664
## 5  0.03103467    4 0.3896897 0.4104847 0.03118646
## 6  0.02050210    5 0.3586550 0.3880502 0.02971467
## 7  0.02033584    6 0.3381529 0.3727853 0.02803416
## 8  0.01814884    8 0.2974812 0.3606252 0.02735292
## 9  0.01534137    9 0.2793324 0.3417444 0.02567744
## 10 0.01322525   10 0.2639910 0.3270809 0.02603548
## 11 0.01225727   11 0.2507658 0.3150886 0.02572817
## 12 0.01086730   12 0.2385085 0.3025617 0.02551938
## 13 0.01000000   13 0.2276412 0.3028684 0.02578135
##
## Variable importance
##      engine_cap      is.manual auto.retract.mirror  brand_popularity
##              42              19              15              13
##      ln_mileage_km
##              11
##
## Node number 1: 462 observations,      complexity param=0.434794
## mean=18.93294, MSE=0.09143041
## left son=2 (206 obs) right son=3 (256 obs)
## Primary splits:
##      engine_cap      < 1.45      to the left, improve=0.43479400, (0 missing)
##      auto.retract.mirror      < 0.5      to the left, improve=0.30401790, (0 missing)
##      is.manual      < 0.5      to the right, improve=0.29498020, (0 missing)
##      vehicle.stability.control      < 0.5      to the left, improve=0.04685959, (0 missing)
##      electric.parking.brake      < 0.5      to the left, improve=0.03702442, (0 missing)
## Surrogate splits:
##      auto.retract.mirror      < 0.5      to the left, agree=0.693, adj=0.311, (0 split)
##      is.manual      < 0.5      to the right, agree=0.658, adj=0.233, (0 split)
##      brand_popularity      < 11.90476 to the right, agree=0.643, adj=0.199, (0 split)
##      ln_mileage_km      < 3.891338 to the left, agree=0.617, adj=0.141, (0 split)
##
## Node number 2: 206 observations,      complexity param=0.07405191
## mean=18.71067, MSE=0.04811162
## left son=4 (90 obs) right son=5 (116 obs)
## Primary splits:

```

```

##      is.manual          < 0.5      to the right, improve=0.31561070, (0 missing)
##      auto.retract.mirror < 0.5      to the left,  improve=0.26490930, (0 missing)
##      ln_mileage_km       < 4.039576 to the right, improve=0.10600460, (0 missing)
##      brand_popularity    < 18.63711 to the left,  improve=0.09261388, (0 missing)
##      engine_cap          < 1.25     to the left,  improve=0.03529204, (0 missing)
## Surrogate splits:
##      brand_popularity    < 18.63711 to the left,  agree=0.612, adj=0.111, (0 split)
##      engine_cap          < 1.1      to the left,  agree=0.602, adj=0.089, (0 split)
##      ln_mileage_km       < 3.530127 to the left,  agree=0.587, adj=0.056, (0 split)
##      auto.retract.mirror < 0.5      to the left,  agree=0.573, adj=0.022, (0 split)
##
## Node number 3: 256 observations,      complexity param=0.056983
## mean=19.11179, MSE=0.05454605
## left son=6 (42 obs) right son=7 (214 obs)
## Primary splits:
##      is.manual          < 0.5      to the right, improve=0.17237520, (0 missing)
##      brand_popularity    < 10.26273 to the left,  improve=0.15997790, (0 missing)
##      auto.retract.mirror < 0.5      to the left,  improve=0.12688950, (0 missing)
##      engine_cap          < 1.9      to the left,  improve=0.10616260, (0 missing)
##      ln_mileage_km       < 2.892815 to the left,  improve=0.05319086, (0 missing)
## Surrogate splits:
##      ln_mileage_km < 2.927414 to the left,  agree=0.863, adj=0.167, (0 split)
##
## Node number 4: 90 observations,      complexity param=0.0205021
## mean=18.57078, MSE=0.03023072
## left son=8 (79 obs) right son=9 (11 obs)
## Primary splits:
##      auto.retract.mirror < 0.5      to the left,  improve=0.31830240, (0 missing)
##      engine_cap          < 1.25     to the left,  improve=0.20068530, (0 missing)
##      ln_mileage_km       < 4.077198 to the right, improve=0.09042888, (0 missing)
##      brand_popularity    < 18.63711 to the left,  improve=0.03873248, (0 missing)
## Surrogate splits:
##      engine_cap < 1.35      to the left,  agree=0.9, adj=0.182, (0 split)
##
## Node number 5: 116 observations,      complexity param=0.03103467
## mean=18.81921, MSE=0.03501908
## left son=10 (108 obs) right son=11 (8 obs)
## Primary splits:
##      engine_cap          < 1.1      to the right, improve=0.322713400, (0 missing)
##      ln_mileage_km       < 4.381183 to the right, improve=0.223371400, (0 missing)
##      auto.retract.mirror < 0.5      to the left,  improve=0.182815100, (0 missing)
##      brand_popularity    < 11.82266 to the left,  improve=0.070881780, (0 missing)
##      rear.camera         < 0.5      to the left,  improve=0.006980555, (0 missing)
##
## Node number 6: 42 observations,      complexity param=0.04448144
## mean=18.89292, MSE=0.08176673
## left son=12 (15 obs) right son=13 (27 obs)
## Primary splits:
##      brand_popularity    < 10.26273 to the left,  improve=5.471237e-01, (0 missing)
##      auto.retract.mirror < 0.5      to the left,  improve=1.456307e-01, (0 missing)
##      ln_mileage_km       < 2.892815 to the left,  improve=9.547085e-02, (0 missing)
##      rear.camera         < 0.5      to the left,  improve=3.707127e-05, (0 missing)
## Surrogate splits:
##      ln_mileage_km       < 3.069509 to the left,  agree=0.738, adj=0.267, (0 split)

```

```

##      auto.retract.mirror < 0.5      to the left,  agree=0.667, adj=0.067, (0 split)
##
## Node number 7: 214 observations,      complexity param=0.02033584
##   mean=19.15475, MSE=0.03795596
##   left son=14 (71 obs) right son=15 (143 obs)
##   Primary splits:
##     auto.retract.mirror    < 0.5      to the left,  improve=0.10565850, (0 missing)
##     engine_cap             < 1.9      to the left,  improve=0.10435160, (0 missing)
##     brand_popularity       < 10.26273 to the left,  improve=0.10370360, (0 missing)
##     ln_mileage_km          < 4.645246 to the right, improve=0.08136821, (0 missing)
##     electric.parking.brake < 0.5      to the left,  improve=0.02123946, (0 missing)
##   Surrogate splits:
##     ln_mileage_km          < 4.589163 to the right, agree=0.696, adj=0.085, (0 split)
##     rear.camera           < 0.5      to the right, agree=0.678, adj=0.028, (0 split)
##     brand_popularity       < 1.80624  to the left,  agree=0.678, adj=0.028, (0 split)
##     engine_cap            < 2.2      to the right, agree=0.673, adj=0.014, (0 split)
##
## Node number 8: 79 observations,      complexity param=0.01225727
##   mean=18.53417, MSE=0.02071521
##   left son=16 (15 obs) right son=17 (64 obs)
##   Primary splits:
##     engine_cap             < 1.1      to the left,  improve=0.31638090, (0 missing)
##     brand_popularity       < 18.63711 to the left,  improve=0.10358930, (0 missing)
##     ln_mileage_km          < 4.038295 to the right, improve=0.05122813, (0 missing)
##
## Node number 9: 11 observations
##   mean=18.83366, MSE=0.01983977
##
## Node number 10: 108 observations,      complexity param=0.01534137
##   mean=18.79028, MSE=0.02540657
##   left son=20 (23 obs) right son=21 (85 obs)
##   Primary splits:
##     ln_mileage_km          < 4.381183 to the right, improve=2.361712e-01, (0 missing)
##     brand_popularity       < 18.63711 to the left,  improve=1.242778e-01, (0 missing)
##     auto.retract.mirror    < 0.5      to the left,  improve=1.035239e-01, (0 missing)
##     engine_cap             < 1.25     to the left,  improve=4.473439e-02, (0 missing)
##     rear.camera           < 0.5      to the left,  improve=1.955941e-05, (0 missing)
##
## Node number 11: 8 observations
##   mean=19.20981, MSE=0.0009215885
##
## Node number 12: 15 observations
##   mean=18.60915, MSE=0.04069782
##
## Node number 13: 27 observations
##   mean=19.05057, MSE=0.03499266
##
## Node number 14: 71 observations,      complexity param=0.02033584
##   mean=19.06488, MSE=0.05924149
##   left son=28 (55 obs) right son=29 (16 obs)
##   Primary splits:
##     ln_mileage_km          < 3.864961 to the right, improve=0.20441200, (0 missing)
##     engine_cap             < 1.75     to the left,  improve=0.13931970, (0 missing)
##     brand_popularity       < 25.94417 to the left,  improve=0.07224559, (0 missing)

```

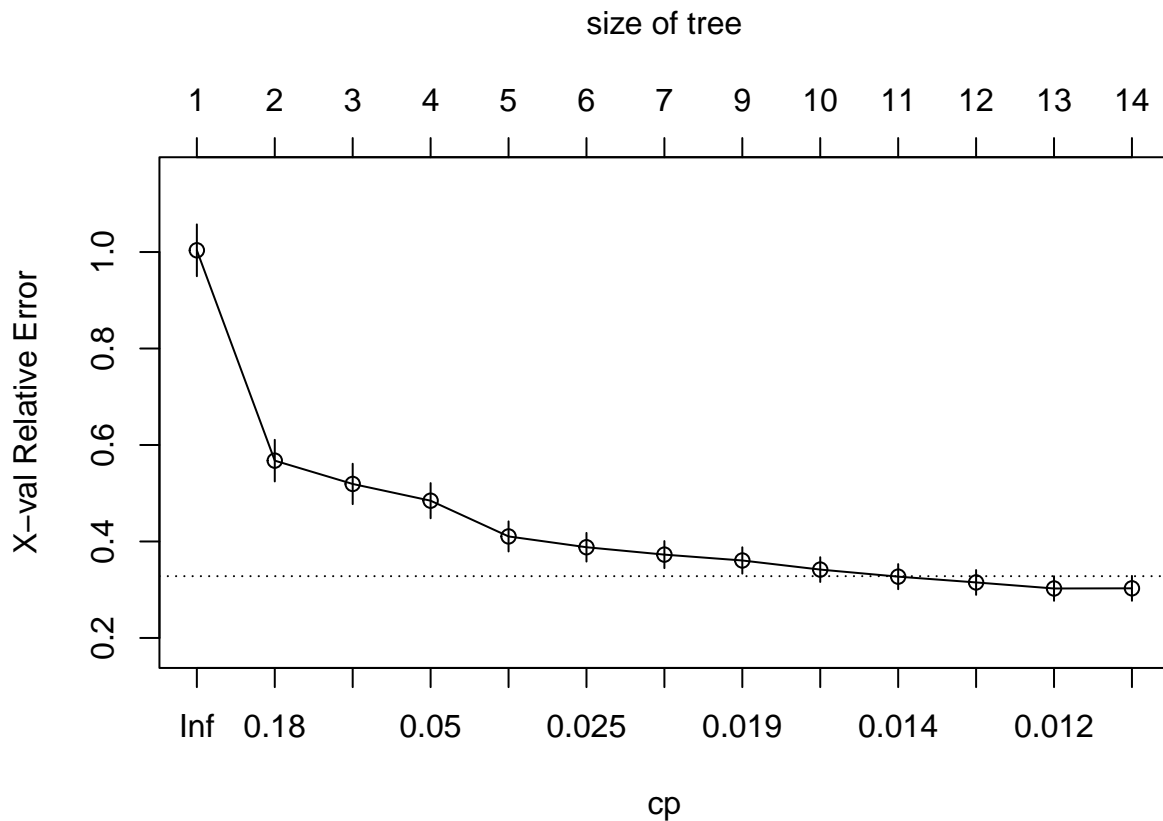
```

##      electric.parking.brake < 0.5      to the left,  improve=0.05464403, (0 missing)
##      rear.camera           < 0.5      to the right, improve=0.01716920, (0 missing)
##  Surrogate splits:
##      sun.roof < 0.5      to the left,  agree=0.789, adj=0.062, (0 split)
##
## Node number 15: 143 observations,      complexity param=0.01814884
##  mean=19.19937, MSE=0.02138609
##  left son=30 (72 obs) right son=31 (71 obs)
##  Primary splits:
##      brand_popularity      < 18.63711 to the left,  improve=0.25067670, (0 missing)
##      engine_cap            < 1.9      to the left,  improve=0.14471810, (0 missing)
##      ln_mileage_km         < 4.659475 to the right, improve=0.03627974, (0 missing)
##      vehicle.stability.control < 0.5    to the right, improve=0.02831441, (0 missing)
##      electric.parking.brake < 0.5    to the left,  improve=0.01857159, (0 missing)
##  Surrogate splits:
##      ln_mileage_km         < 3.728114 to the right, agree=0.587, adj=0.169, (0 split)
##      auto.cruise.control   < 0.5      to the right, agree=0.566, adj=0.127, (0 split)
##      sun.roof              < 0.5      to the right, agree=0.545, adj=0.085, (0 split)
##      vehicle.stability.control < 0.5    to the right, agree=0.545, adj=0.085, (0 split)
##      engine_cap            < 1.65     to the left,  agree=0.545, adj=0.085, (0 split)
##
## Node number 16: 15 observations
##  mean=18.36695, MSE=0.004950495
##
## Node number 17: 64 observations
##  mean=18.57337, MSE=0.0163201
##
## Node number 20: 23 observations
##  mean=18.64137, MSE=0.01841116
##
## Node number 21: 85 observations
##  mean=18.83058, MSE=0.01967554
##
## Node number 28: 55 observations,      complexity param=0.0108673
##  mean=19.00553, MSE=0.05068436
##  left son=56 (43 obs) right son=57 (12 obs)
##  Primary splits:
##      engine_cap            < 1.75     to the left,  improve=0.16467120, (0 missing)
##      brand_popularity      < 25.94417 to the left,  improve=0.07506844, (0 missing)
##      ln_mileage_km         < 4.407041 to the right, improve=0.06419887, (0 missing)
##      rear.camera           < 0.5      to the right, improve=0.01498974, (0 missing)
##  Surrogate splits:
##      ln_mileage_km < 4.768907 to the left,  agree=0.836, adj=0.25, (0 split)
##
## Node number 29: 16 observations
##  mean=19.26891, MSE=0.03491998
##
## Node number 30: 72 observations
##  mean=19.12667, MSE=0.007021014
##
## Node number 31: 71 observations,      complexity param=0.01322525
##  mean=19.27311, MSE=0.025156
##  left son=62 (58 obs) right son=63 (13 obs)
##  Primary splits:

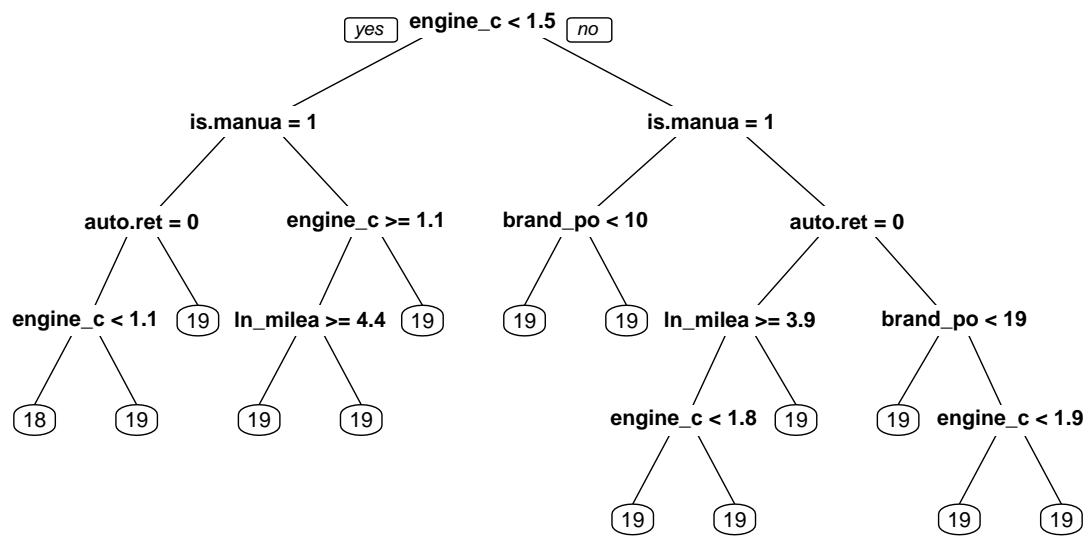
```

```
##      engine_cap          < 1.9      to the left,  improve=0.3127782000, (0 missing)
##      ln_mileage_km       < 4.363891 to the left,  improve=0.0834965200, (0 missing)
##      vehicle.stability.control < 0.5      to the right, improve=0.0474577100, (0 missing)
##      rear.camera         < 0.5      to the left,  improve=0.0460286900, (0 missing)
##      brand_popularity     < 25.94417 to the left,  improve=0.0009718159, (0 missing)
## Surrogate splits:
##      rear.camera  < 0.5      to the left,  agree=0.859, adj=0.231, (0 split)
##      ln_mileage_km < 4.363891 to the left,  agree=0.831, adj=0.077, (0 split)
##
## Node number 56: 43 observations
##   mean=18.95726, MSE=0.04807941
##
## Node number 57: 12 observations
##   mean=19.17846, MSE=0.0217651
##
## Node number 62: 58 observations
##   mean=19.23111, MSE=0.01900195
##
## Node number 63: 13 observations
##   mean=19.46047, MSE=0.009639761
```

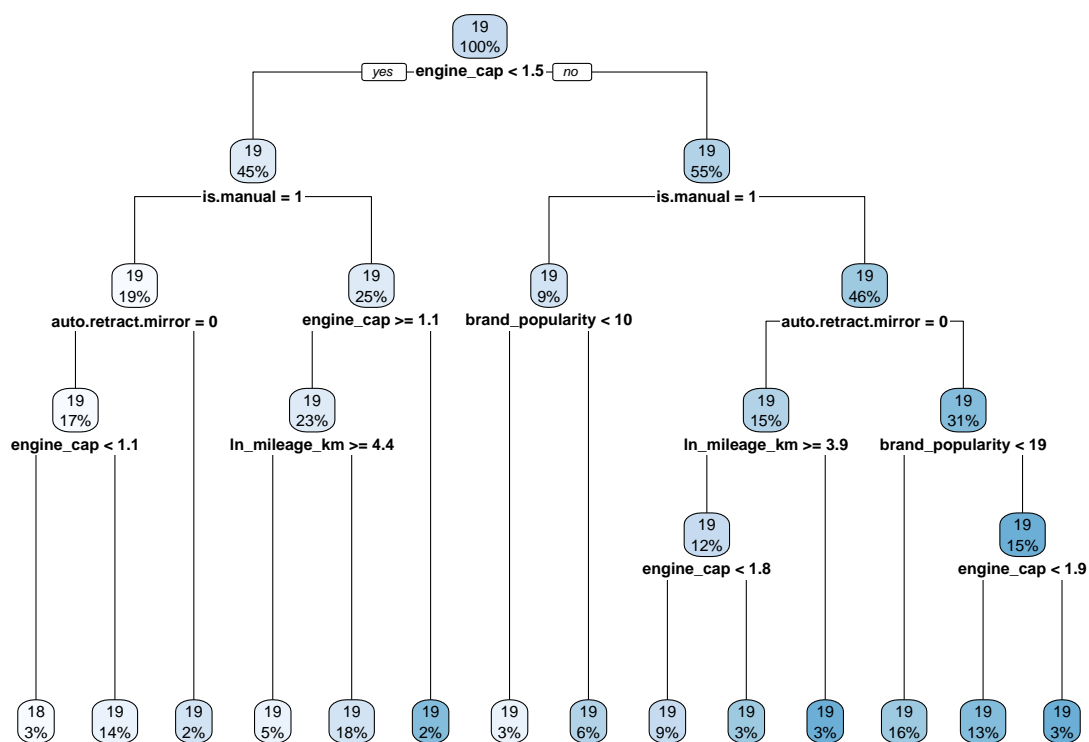
```
plotcp(tree_1)
```



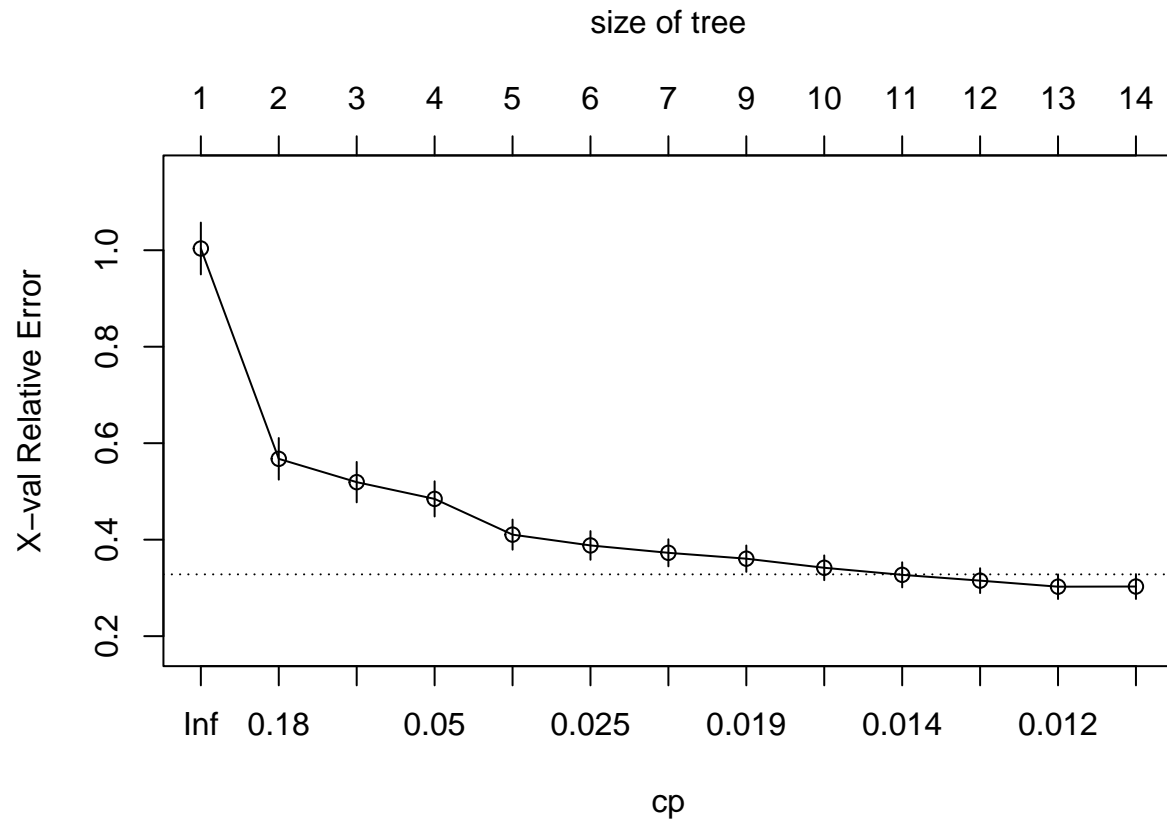
```
prp(tree_1)
```



```
rpart.plot(tree_1)
```



```
plotcp(tree_1)
```



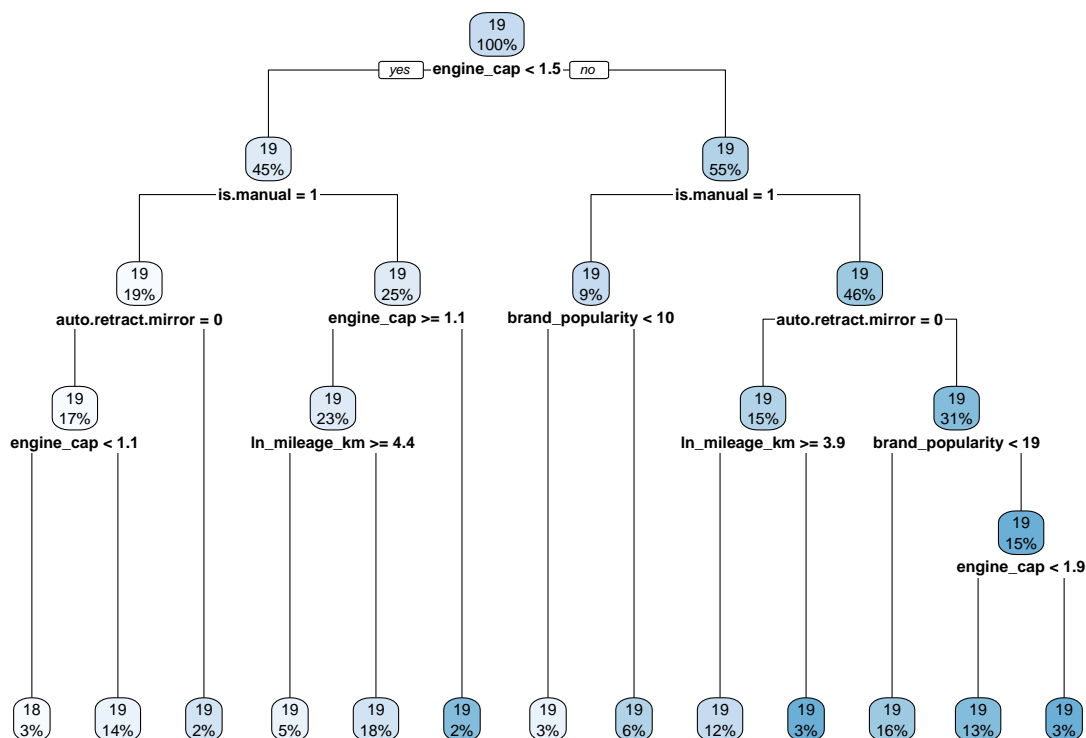
```
pred1 <- predict(tree_1, test)
rmse1 <- RMSE(pred1, test$ln_price)
```

Post-pruning

```
min_cp = tree_1$cptable[which.min(tree_1$cptable[, "xerror"]), "CP"]
min_cp
```

```
## [1] 0.0108673
```

```
tree_pruned <- prune(tree_1, cp = min_cp)
rpart.plot(tree_pruned)
```

The regression tree uses: - engine_cap - is.manual - brand_popularity - ln_mileage_km - auto.retract mirror
From this regression tree, it seems that engine capacity has the highest information in predicting car price.

```
pred2 <- predict(tree_pruned, test)
rmse2 <- RMSE(pred2, test$ln_price)
```

```
print(paste0("Pre-pruned regression tree: ", rmse1))
```

```
## [1] "Pre-pruned regression tree: 0.219521928561908"
```

```
print(paste0("Post-pruned regression tree: ", rmse2))
```

```
## [1] "Post-pruned regression tree: 0.229030298725043"
```

Prepruning regression tree have lower RMSE score and thus is better at predicting the price value. This is an interesting result as it is rare to see prepruning regression tree version do better than postpruning.

Random Forest

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.3.3
```

```
## randomForest 4.7-1.2

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':
##
##      combine

## The following object is masked from 'package:ggplot2':
##
##      margin
```

Parameter tuning

```
p <- ncol(df3) - 1
p
```

```
## [1] 12
```

The standard size of random subspace subset for a random forest regression is $p/3$ so 4 variables is the middle range. To be safe, I'm gonna set the range to 2 : 6.

```
grid_rf <- expand.grid( mtry=c(2:6) )

grid_rf
```

```
##      mtry
## 1      2
## 2      3
## 3      4
## 4      5
## 5      6
```

```
train_ctrl <- trainControl(method="cv", # type of resampling in this case Cross-Validated
                           number=10, # number of folds
                           search = "grid" # we are performing a "grid-search"
                           )

model_cv_grid <- train(ln_price ~ .,
                      data = train,
                      method = "rf", # this will use the randomForest::randomForest function
                      metric = "RMSE", # which metric should be optimized for
                      trControl = train_ctrl,
                      tuneGrid = grid_rf,
                      # options to be passed to randomForest
```

```

        ntree = 501,
        keep.forest=TRUE,
        importance=TRUE
    )

model_cv_grid

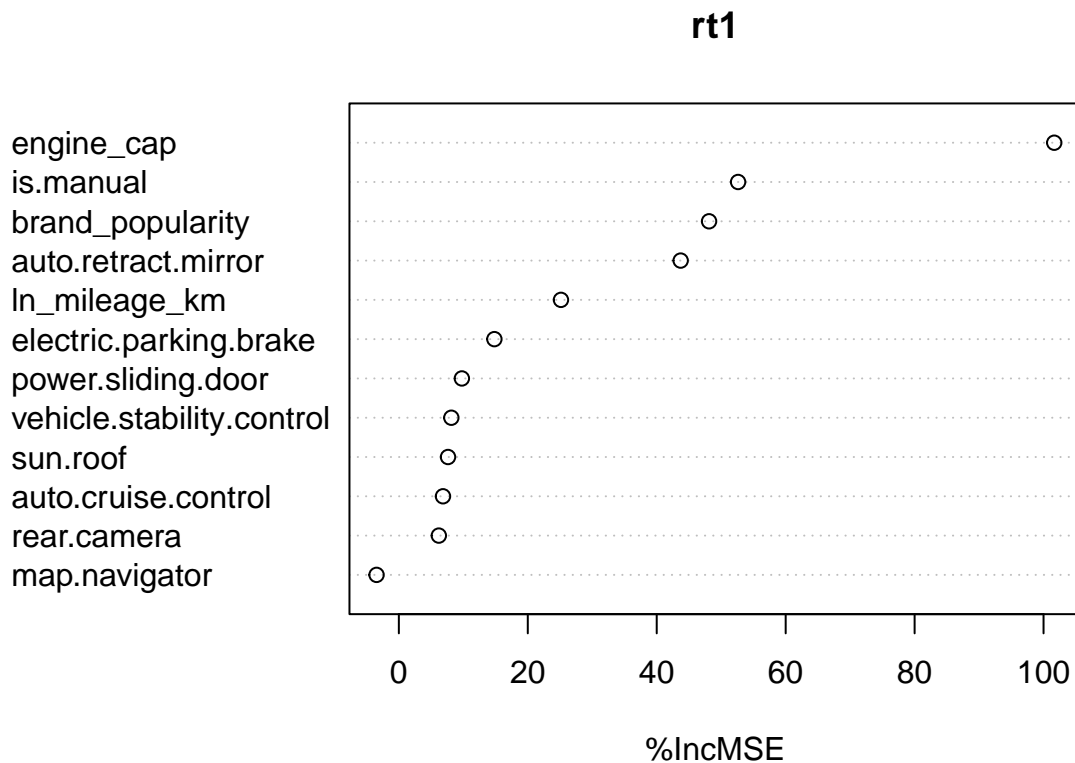
## Random Forest
##
## 462 samples
## 12 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 417, 414, 417, 415, 415, 416, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared  MAE
##   2     0.1635922  0.7484645  0.12619148
##   3     0.1449922  0.7851546  0.11036160
##   4     0.1366772  0.8036694  0.10232767
##   5     0.1324781  0.8130378  0.09772935
##   6     0.1312407  0.8154400  0.09598864
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 6.

rt1 <- randomForest(ln_price~., data = train, mtry = 6, importance = TRUE, ntrees = 501)
summary(rt1)

##               Length Class  Mode
## call              6    -none- call
## type              1    -none- character
## predicted         462    -none- numeric
## mse               500    -none- numeric
## rsq               500    -none- numeric
## oob.times         462    -none- numeric
## importance         24    -none- numeric
## importanceSD       12    -none- numeric
## localImportance    0    -none- NULL
## proximity          0    -none- NULL
## ntree              1    -none- numeric
## mtry               1    -none- numeric
## forest            11    -none- list
## coefs              0    -none- NULL
## y                 462    -none- numeric
## test              0    -none- NULL
## inbag             0    -none- NULL
## terms             3     terms  call

varImpPlot(rt1, type=1)

```



The top 3 variables are - engine_cap - is.manual - brand_popularity

```
print(rt1)
```

```
##
## Call:
##  randomForest(formula = ln_price ~ ., data = train, mtry = 6,      importance = TRUE, ntrees = 501)
##              Type of random forest: regression
##              Number of trees: 500
## No. of variables tried at each split: 6
##
##              Mean of squared residuals: 0.01715234
##              % Var explained: 81.24
```

```
pred2 <- predict(rt1, test)
rmse3 <- RMSE(pred2, test$ln_price)
rmse3
```

```
## [1] 0.1534842
```

The random forest method is more accurate (lower RMSE).

Conclusion

```
print("Regression")

## [1] "Regression"

print(paste0("Initial OLS: ", get_rmse(ols_1)))

## [1] "Initial OLS: 29034580.34404"

print(paste0("Best performanced OLS: ", get_rmse(best_reg1)))

## [1] "Best performanced OLS: 28974723.6186467"

print(paste0("Double Log Model: ", get_rmse(ln_ols)))

## [1] "Double Log Model: 0.162838182708786"

print(paste0("WLS model: ", get_rmse(wls_model)))

## [1] "WLS model: 0.163010223967787"

print("Machine Learning")

## [1] "Machine Learning"

print(paste0("Pre-pruned regression tree: ", rmse1))

## [1] "Pre-pruned regression tree: 0.219521928561908"

print(paste0("Post-pruned regression tree: ", rmse2))

## [1] "Post-pruned regression tree: 0.229030298725043"

print(paste0("Random Forest: ", rmse3))

## [1] "Random Forest: 0.153484179682873"
```

Random forest have the lowest RMSE score, which means that it is the best performing machine learning model.

Surprisingly, the double log and wls model outperformed regression tree.