

Vehicle Analysis

<https://www.kaiandkaro.com/>

Introduction

Data Scrapped form <https://www.kaiandkaro.com/>

```
1 library(tidyverse)
2 library(pander)
3 library(broom)
```

```
1 all_vehicles_data <- read_csv('vehicles_data_202412141810.csv')
```

Data

```
1 #
2 analysis_data <- all_vehicles_data %>%
3   select(where(~100*sum(is.na(.x))/length(.x) < 0.01)) |>
4   select(price, source, current_location, mileage, mileage_unit,
5          annual_insurance_currency, annual_insurance, year_of_manufacture,
6          availability, purchase_status, model_make_name, model_make_vehicle_type) |>
7   mutate(
8     current_location = str_replace_all(
9       current_location, c("[[:punct:]]" = '', 'Port of' = '')) |>
10    str_to_lower() |> trimws()
11 ) |>
12 mutate(
13   price = as.numeric(price),
14   mileage = as.numeric(mileage),
15   annual_insurance = as.numeric(annual_insurance),
16   current_location = str_replace_all(
17     current_location,
18     c(
19       "nairobi kenya"= "kenya", "nairobi" = "kenya" ,
20       "mombasa" = "kenya", "kenyakenya" = "kenya",
21       "enroute" = "overseas", "high seas" = "overseas"
22     ),
23   model_make_name = str_replace_all(
24     model_make_name, c("[[:punct:]]" = '', 'Port of' = '')) |>
25     str_to_lower() |> trimws()
26 ) |>
27 mutate(
28   mileage = case_when(
```

```

29     mileage_unit == "Miles" ~ mileage*1.6094,
30     .default = mileage
31   )
32 ) |>
33 select(-mileage_unit, -annual_insurance_currency)
34 #

```

```

1 glimpse(analysis_data)

```

```

1 Rows: 2,819
2 Columns: 10
3 $ price                <dbl> 7990000, 2299000, 1850000, 570000, 7990000, 14~
4 $ source               <chr> "Kenyan Used", "Kenyan Used", "Kenyan Used", "~
5 $ current_location     <chr> "kenya", "kenya", "kenya", "kenya", "kenya", "~
6 $ mileage              <dbl> 27000, 69362, 56909, 145000, 154307, 153000, 1~
7 $ annual_insurance     <dbl> 330000, 84000, 74000, 30000, 316000, 60000, 58~
8 $ year_of_manufacture  <dbl> 2023, 2015, 2012, 2004, 2014, 2006, 2011, 2005~
9 $ availability         <chr> "available", "available", "available", "availa~
10 $ purchase_status     <chr> "Available", "Available", "Available", "Availa~
11 $ model_make_name      <chr> "toyota", "toyota", "volkswagen", "bmw", "land~
12 $ model_make_vehicle_type <chr> "Automobile", "Automobile", "Automobile", "Aut~

```

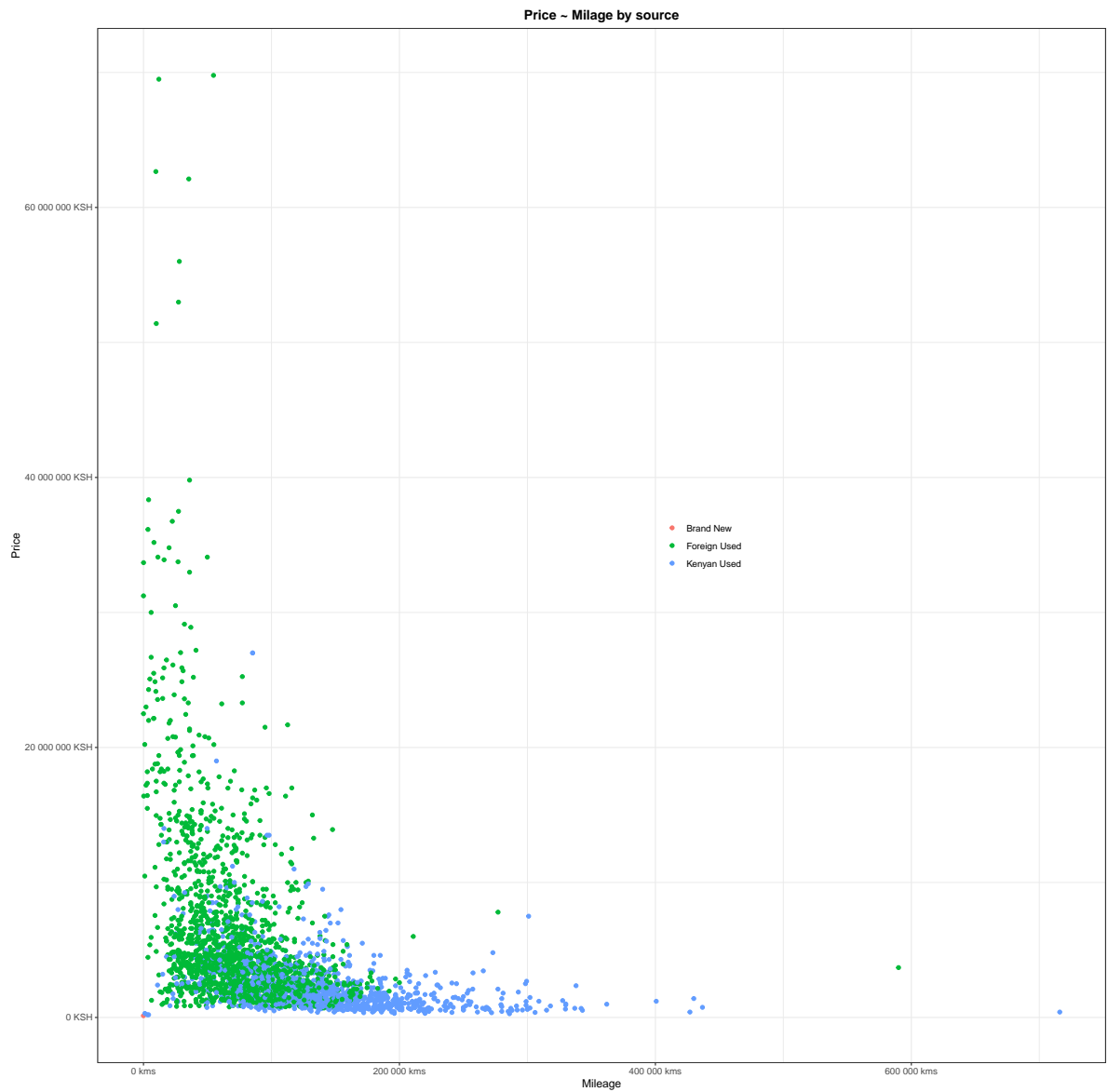
EDA

```

1 mile_price_source_plot = analysis_data |>
2   ggplot(aes(x = mileage, y = price, colour = source)) +
3   geom_point() +
4   labs(
5     x = "Mileage",
6     y = "Price",
7     title = "Price ~ Milage by source",
8     caption = "https://github.com/akhapwoyaco"
9   ) +
10  scale_y_continuous(
11    labels = scales::unit_format(unit = 'KSH')
12  ) +
13  scale_x_continuous(
14    labels = scales::unit_format(unit = 'kms')
15  ) +
16  theme_bw() +

```

```
17 theme(  
18   plot.title = element_text(face = 'bold', hjust = 0.5),  
19   legend.title = element_blank(),  
20   legend.position = 'inside',  
21   legend.position.inside = c(0.6, 0.5),  
22   legend.background = element_blank()  
23 )  
24 mile_price_source_plot
```



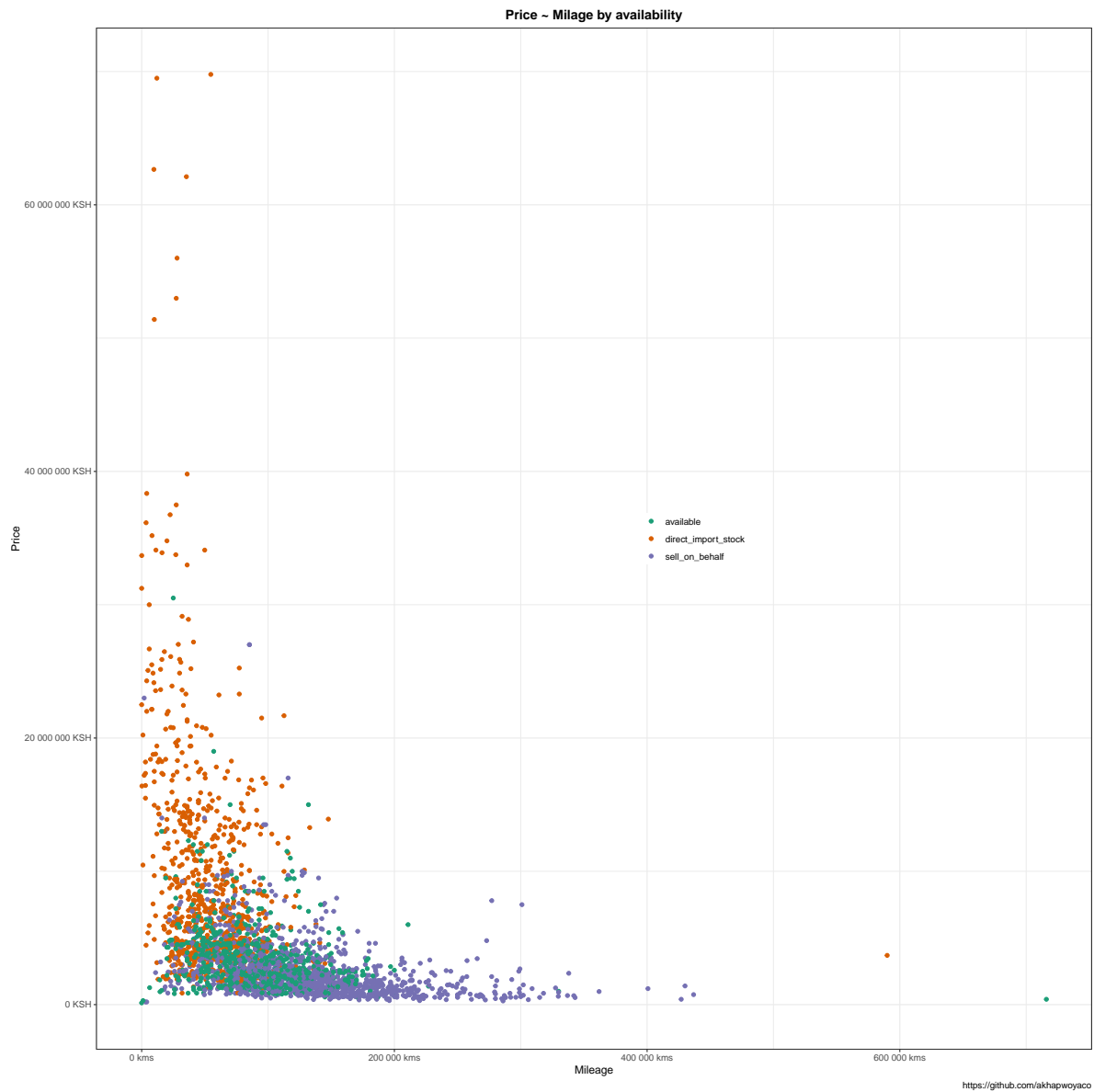
```
1 ggsave("mile_price_source_plot.png", plot = mile_price_source_plot,
2       width = 22, height = 15, unit = "cm", dpi = 450)
```

```
1 mile_price_availability_plot <- analysis_data |>
2   ggplot(aes(x = mileage, y = price, colour = availability)) +
3   geom_point() +
4   scale_color_brewer(palette = 'Dark2') +
5   labs(
```

```

6     x = "Mileage",
7     y = "Price",
8     title = "Price ~ Milage by availability",
9     caption = "https://github.com/akhapwoyaco"
10  ) +
11  scale_y_continuous(
12    labels = scales::unit_format(unit = 'KSH')
13  ) +
14  scale_x_continuous(
15    labels = scales::unit_format(unit = 'kms')
16  ) +
17  theme_bw() +
18  theme(
19    plot.title = element_text(face = 'bold', hjust = 0.5),
20    legend.title = element_blank(),
21    legend.position = 'inside',
22    legend.position.inside = c(0.6, 0.5),
23    legend.background = element_blank()
24  )
25  mile_price_availability_plot

```



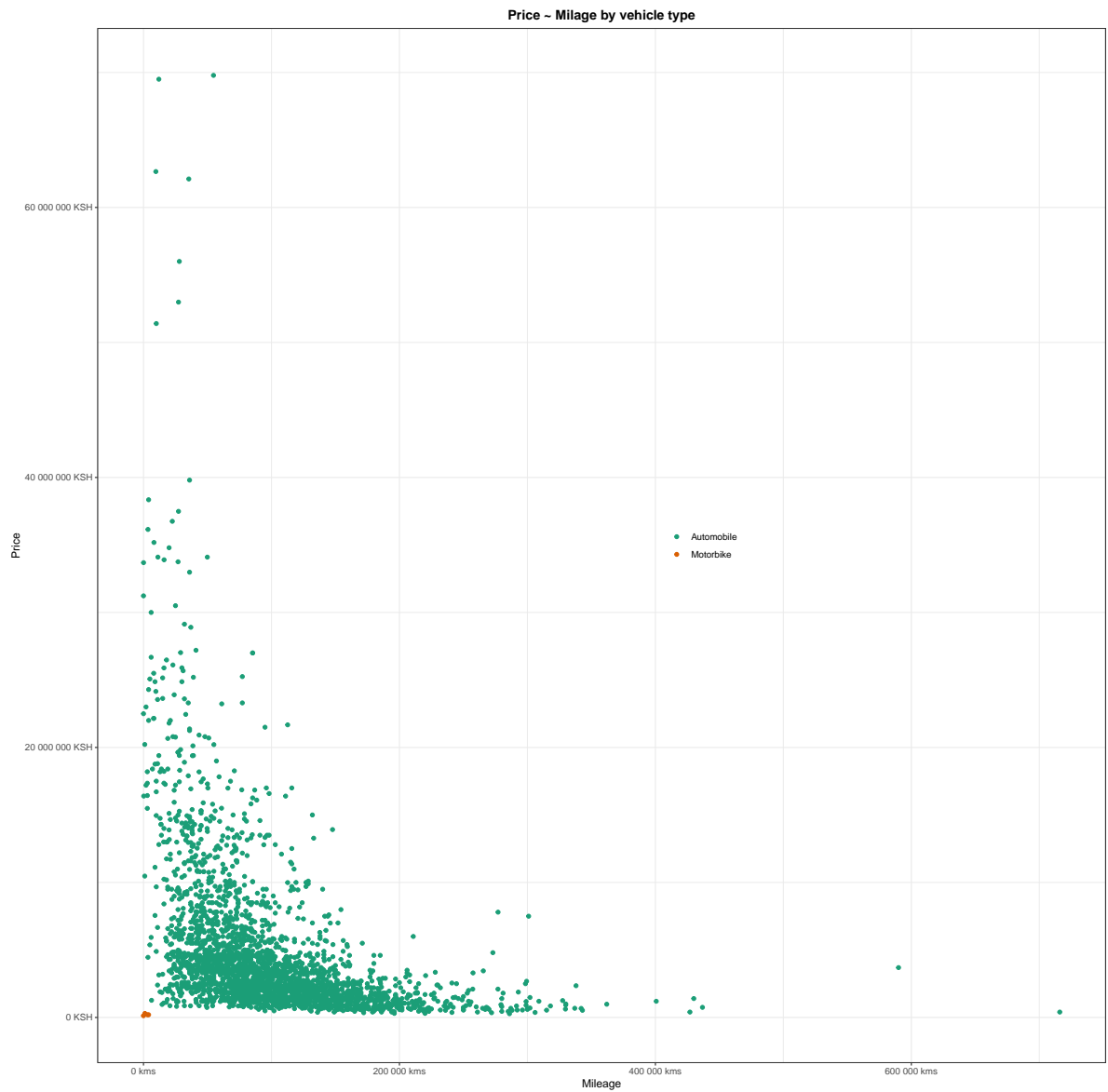
```
1 ggsave("mile_price_availability_plot.png",plot = mile_price_availability_plot,  
2       width = 22, height = 15, unit = "cm", dpi = 450)
```

```
1 mile_price_make_plot = analysis_data |>  
2   ggplot(aes(x = mileage, y = price, colour = model_make_vehicle_type)) +  
3   geom_point() +  
4   scale_color_brewer(palette = 'Dark2') +  
5   labs(
```

```

6     x = "Mileage",
7     y = "Price",
8     title = "Price ~ Milage by vehicle type",
9     caption = "https://github.com/akhapwoyaco"
10  ) +
11  scale_y_continuous(
12    labels = scales::unit_format(unit = 'KSH')
13  ) +
14  scale_x_continuous(
15    labels = scales::unit_format(unit = 'kms')
16  ) +
17  theme_bw() +
18  theme(
19    plot.title = element_text(face = 'bold', hjust = 0.5),
20    legend.title = element_blank(),
21    legend.position = 'inside',
22    legend.position.inside = c(0.6, 0.5),
23    legend.background = element_blank()
24  )
25  mile_price_make_plot

```

```
1 ggsave("mile_price_make_plot.png",plot = mile_price_make_plot,  
2       width = 22, height = 15, unit = "cm", dpi = 450)
```

MODELS

Individual Variables

```
1 # model of numerics
2 model_1 <- lm(
3   price~mileage+annual_insurance+year_of_manufacture,
4   data = analysis_data)
5 glance(model_1) |> t() |> pander()
```

r.squared	0.9246
adj.r.squared	0.9245
sigma	1571028
statistic	11508
p.value	0
df	3
logLik	-44217
AIC	88445
BIC	88474
deviance	6.948e+15
df.residual	2815
nobs	2819

```

1 rhs <- c('price') #Right hand side of model
2 lhs <- c("mileage","annual_insurance","year_of_manufacture") #Left hand side Y-Variables
3
4 #Model List Initiation
5 model = list()
6 for (i in lhs){
7   for (j in rhs){
8     model[[paste(i, "vs", j)]] <- lm(as.formula(paste(i, "~", j)),
9                                     data = analysis_data)
10  }
11 }
12 model %>% pander::pander()

```

- **mileage vs price:**

Table 2: Fitting linear model: `as.formula(paste(i, "~", j))`

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	118435	1332	88.92	0
price	-0.00435	0.0001793	-24.26	3.014e-118

- **annual_insurance vs price:**

Table 3: Fitting linear model: `as.formula(paste(i, "~", j))`

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4548	1606	2.833	0.004649
price	0.03967	0.0002161	183.5	0

- **year_of_manufacture vs price:**

Table 4: Fitting linear model: `as.formula(paste(i, "~", j))`

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2014	0.08806	22868	0
price	3.244e-07	1.186e-08	27.36	2.292e-146

All Variables

```
1 multi_lm = lm(price~., data = analysis_data)
2 # summary(multi_lm)
```

```
1 multi_lm |>
2   pander::pander()
```

Table 5: Fitting linear model: price ~ . (continued below)

	Estimate	Std. Error	t value
(Intercept)	-156183145	22762448	-6.861
sourceForeign Used	1199081	2569832	0.4666
sourceKenyan Used	1181957	2572212	0.4595
current_locationjapan	-1465129	669939	-2.187
current_locationkenya	-1038502	907699	-1.144
current_locationoverseas	-1122312	918705	-1.222
current_locationsouth africa	109981	938246	0.1172
current_locationthailand	-1499363	939413	-1.596
current_locationunited kingdom	-178932	683347	-0.2618
mileage	-1.904	0.7034	-2.706
annual_insurance	19.8	0.1996	99.22
year_of_manufacture	77732	11173	6.957
availabilitydirect_import_stock	821233	614448	1.337
availabilitysell_on_behalf	233645	104526	2.235
purchase_statusReserved	-409821	862568	-0.4751
purchase_statusSold	-21912	142930	-0.1533
model_make_nameashok leyland	-366413	2096438	-0.1748
model_make_nameaudi	112773	1484186	0.07598
model_make_namebentley	3344541	1543944	2.166
model_make_namebmw	352250	1486420	0.237
model_make_namechevrolet	411631	2100050	0.196
model_make_namecitreon	-413447	1709651	-0.2418
model_make_namedaihatsu	-460683	1552473	-0.2967
model_make_namedodge	523763	2096053	0.2499
model_make_namefaw	-156344	2094586	-0.07464
model_make_nameferrari	4313724	1823576	2.366
model_make_nameford	29054	1505960	0.01929
model_make_namehino	308023	2096683	0.1469
model_make_namehonda	-201179	1490946	-0.1349
model_make_namehyundai	-317642	1623277	-0.1957

	Estimate	Std. Error	t value
model_make_nameisuzu	-21412	1570062	-0.01364
model_make_namejaguar	100571	1518006	0.06625
model_make_namejeep	216888	1504735	0.1441
model_make_namejincheng	-1047549	2347353	-0.4463
model_make_namekawasaki	-402824	2569082	-0.1568
model_make_namekia	263551	2094416	0.1258
model_make_namelamborghini	9426409	1646766	5.724
model_make_nameland rover	1206211	1488446	0.8104
model_make_namelexus	1232991	1494426	0.8251
model_make_namemazda	-213200	1483581	-0.1437
model_make_namemercedes	559383	1484414	0.3768
model_make_namemini	64201	1621645	0.03959
model_make_namemitsubishi	35806	1492778	0.02399
model_make_namenissan	-70984	1484327	-0.04782
model_make_namepeugeot	-380289	1495026	-0.2544
model_make_nameporsche	893136	1493410	0.5981
model_make_namerenault	-288928	1582694	-0.1826
model_make_namerolls royce	11865689	2148642	5.522
model_make_nameskygo	-755729	2569850	-0.2941
model_make_namesmart	-342408	1600042	-0.214
model_make_namesubaru	-73936	1485278	-0.04978
model_make_namesuzuki	-386423	1501395	-0.2574
model_make_nametata	-214402	2095159	-0.1023
model_make_nametoyota	45346	1482557	0.03059
model_make_namevolkswagen	-128076	1485554	-0.08621
model_make_namevolvo	3381	1490084	0.002269
model_make_vehicle_typeMotorbike	-97024	1491151	-0.06507

	Pr(> t)
(Intercept)	8.383e-12
sourceForeign Used	0.6408
sourceKenyan Used	0.6459
current_locationjapan	0.02883
current_locationkenya	0.2527
current_locationoverseas	0.222
current_locationsouth africa	0.9067
current_locationthailand	0.1106
current_locationunited kingdom	0.7935
mileage	0.006848

	Pr(> t)
annual_insurance	0
year_of_manufacture	4.313e-12
availabilitydirect_import_stock	0.1815
availabilitysell_on_behalf	0.02548
purchase_statusReserved	0.6347
purchase_statusSold	0.8782
model_make_nameashok leyland	0.8613
model_make_nameaudi	0.9394
model_make_namebentley	0.03038
model_make_namebmw	0.8127
model_make_namechevrolet	0.8446
model_make_namecitreon	0.8089
model_make_namedaihatsu	0.7667
model_make_namedodge	0.8027
model_make_namefaw	0.9405
model_make_nameferrari	0.01807
model_make_nameford	0.9846
model_make_namehino	0.8832
model_make_namehonda	0.8927
model_make_namehyundai	0.8449
model_make_nameisuzu	0.9891
model_make_namejaguar	0.9472
model_make_namejeep	0.8854
model_make_namejincheng	0.6554
model_make_namekawasaki	0.8754
model_make_namekia	0.8999
model_make_namelamborghini	1.151e-08
model_make_nameland rover	0.4178
model_make_namelexus	0.4094
model_make_namemazda	0.8857
model_make_namemercedes	0.7063
model_make_namemini	0.9684
model_make_namemitsubishi	0.9809
model_make_namenissan	0.9619
model_make_namepeugeot	0.7992
model_make_nameporsche	0.5499
model_make_namerenault	0.8552
model_make_namerolls royce	3.655e-08
model_make_nameskygo	0.7687
model_make_namesmart	0.8306
model_make_namesubaru	0.9603

	$\Pr(> t)$
model_make_namesuzuki	0.7969
model_make_nametata	0.9185
model_make_nametoyota	0.9756
model_make_namevolkswagen	0.9313
model_make_namevolvo	0.9982
model_make_vehicle_typeMotorbike	0.9481