

MLB_2

February 25, 2024

```
[ ]: import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline

import sys
sys.path.append('/home/raven/Documents/CS/CS543_ML/vscode')
from hroml import myfunctions as mf
```

```
[ ]: cardata = pd.read_csv('cardata.csv')
```

```
[ ]: cardata
```

```
[ ]:
```

	Make	Model	Year	Engine	Fuel Type	Engine HP \
0	BMW	1 Series M	2011	premium unleaded (required)		335.0
1	BMW	1 Series	2011	premium unleaded (required)		300.0
2	BMW	1 Series	2011	premium unleaded (required)		300.0
3	BMW	1 Series	2011	premium unleaded (required)		230.0
4	BMW	1 Series	2011	premium unleaded (required)		230.0
...
11909	Acura	ZDX	2012	premium unleaded (required)		300.0
11910	Acura	ZDX	2012	premium unleaded (required)		300.0
11911	Acura	ZDX	2012	premium unleaded (required)		300.0
11912	Acura	ZDX	2013	premium unleaded (recommended)		300.0
11913	Lincoln	Zephyr	2006	regular unleaded		221.0

	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors \
0	6.0	MANUAL	rear wheel drive	2.0
1	6.0	MANUAL	rear wheel drive	2.0
2	6.0	MANUAL	rear wheel drive	2.0
3	6.0	MANUAL	rear wheel drive	2.0
4	6.0	MANUAL	rear wheel drive	2.0
...
11909	6.0	AUTOMATIC	all wheel drive	4.0
11910	6.0	AUTOMATIC	all wheel drive	4.0
11911	6.0	AUTOMATIC	all wheel drive	4.0

11912	6.0	AUTOMATIC	all wheel drive	4.0
11913	6.0	AUTOMATIC	front wheel drive	4.0

		Market Category	Vehicle Size	Vehicle Style	\
0	Factory Tuner,	Luxury,High-Performance	Compact	Coupe	
1		Luxury,Performance	Compact	Convertible	
2		Luxury,High-Performance	Compact	Coupe	
3		Luxury,Performance	Compact	Coupe	
4		Luxury	Compact	Convertible	
...		
11909		Crossover,Hatchback,Luxury	Midsize	4dr Hatchback	
11910		Crossover,Hatchback,Luxury	Midsize	4dr Hatchback	
11911		Crossover,Hatchback,Luxury	Midsize	4dr Hatchback	
11912		Crossover,Hatchback,Luxury	Midsize	4dr Hatchback	
11913		Luxury	Midsize	Sedan	

	highway MPG	city mpg	Popularity	MSRP
0	26	19	3916	46135
1	28	19	3916	40650
2	28	20	3916	36350
3	28	18	3916	29450
4	28	18	3916	34500
...
11909	23	16	204	46120
11910	23	16	204	56670
11911	23	16	204	50620
11912	23	16	204	50920
11913	26	17	61	28995

[11914 rows x 16 columns]

```
[ ]: cardata.columns
```

```
[ ]: Index(['Make', 'Model', 'Year', 'Engine Fuel Type', 'Engine HP',
          'Engine Cylinders', 'Transmission Type', 'Driven_Wheels',
          'Number of Doors', 'Market Category', 'Vehicle Size', 'Vehicle Style',
          'highway MPG', 'city mpg', 'Popularity', 'MSRP'],
          dtype='object')
```

```
[ ]: cardata.columns = cardata.columns.str.lower().str.replace(' ', '_')
      string_columns = list(cardata.dtypes[cardata.dtypes == 'object'].index)
      for col in string_columns:
          cardata[col] = cardata[col].str.lower().str.replace(' ', '_')
```

```
[ ]: for i in cardata.columns:
      print(i)
      print(cardata[i].unique())
```

```
print("\n")
```

make

```
['bmw' 'audi' 'fiat' 'mercedes-benz' 'chrysler' 'nissan' 'volvo' 'mazda'  
'mitsubishi' 'ferrari' 'alfa_romeo' 'toyota' 'mclaren' 'maybach'  
'pontiac' 'porsche' 'saab' 'gmc' 'hyundai' 'plymouth' 'honda'  
'oldsmobile' 'suzuki' 'ford' 'cadillac' 'kia' 'bentley' 'chevrolet'  
'dodge' 'lamborghini' 'lincoln' 'subaru' 'volkswagen' 'spyker' 'buick'  
'acura' 'rolls-royce' 'maserati' 'lexus' 'aston_martin' 'land_rover'  
'lotus' 'infiniti' 'scion' 'genesis' 'hummer' 'tesla' 'bugatti']
```

model

```
['1_series_m' '1_series' '100' '124_spider' '190-class' '2_series' '200'  
'200sx' '240sx' '240' '2' '3_series_gran_turismo' '3_series' '300-class'  
'3000gt' '300' '300m' '300zx' '323' '350-class' '350z' '360' '370z' '3'  
'4_series_gran_coupe' '4_series' '400-class' '420-class' '456m'  
'458_italia' '4c' '4runner' '5_series_gran_turismo' '5_series'  
'500-class' '500e' '500' '500l' '500x' '550' '560-class' '570s' '575m'  
'57' '599' '5' '6_series_gran_coupe' '6_series' '600-class' '6000'  
'612_scaglietti' '626' '62' '650s_coupe' '650s_spider' '6' '7_series'  
'718_cayman' '740' '760' '780' '8_series' '80' '850' '86' '9-2x'  
'9-3_griffin' '9-3' '9-4x' '9-5' '9-7x' '9000' '900' '90' '911' '928'  
'929' '940' '944' '960' '968' 'a3' 'a4_allroad' 'a4' 'a5' 'a6' 'a7' 'a8'  
'acadia_limited' 'acadia' 'accent' 'acclaim' 'accord_crosstour'  
'accord_hybrid' 'accord_plug-in_hybrid' 'accord' 'achieva'  
'activehybrid_5' 'activehybrid_7' 'activehybrid_x6' 'aerio' 'aerostar'  
'alero' 'allante' 'allroad_quattro' 'allroad' 'alpina_b6_gran_coupe'  
'alpina_b7' 'alpina' 'altima_hybrid' 'altima' 'amanti' 'amg_gt' 'armada'  
'arnage' 'aspen' 'aspire' 'astro_cargo' 'astro' 'ats_coupe' 'ats-v' 'ats'  
'aurora' 'avalanche' 'avalon_hybrid' 'avalon' 'avenger' 'aventador'  
'aveo' 'aviator' 'axxess' 'azera' 'aztek' 'azure_t' 'azure'  
'b-class_electric_drive' 'b-series_pickup' 'b-series_truck' 'b-series'  
'b9_tribeca' 'baja' 'beetle_convertible' 'beetle' 'beretta'  
'black_diamond_avalanche' 'blackwood' 'blazer' 'bolt_ev' 'bonneville'  
'borrego' 'boxster' 'bravada' 'breeze' 'bronco_ii' 'bronco' 'brooklands'  
'brougham' 'brz' 'c-class' 'c-max_hybrid' 'c30' 'c36_amg' 'c43_amg' 'c70'  
'c8' 'cabriolet' 'cabrio' 'cadenza' 'caliber' 'california_t' 'california'  
'camaro' 'camry_hybrid' 'camry_solara' 'camry' 'canyon' 'caprice'  
'captiva_sport' 'caravan' 'carrera_gt' 'cascada' 'catera' 'cavalier'  
'cayenne' 'cayman_s' 'cayman' 'cc' 'celebrity' 'celica' 'century'  
'challenger' 'charger' 'chevy_van' 'ciera' 'cirrus' 'city_express'  
'civic_crx' 'civic_del_sol' 'civic' 'c/k_1500_series' 'c/k_2500_series'  
'cl-class' 'cla-class' 'cl' 'classic' 'clk-class' 'cls-class' 'cobalt'  
'colorado' 'colt' 'concorde' 'continental_flying_spur_speed'  
'continental_flying_spur' 'continental_gt_speed_convertible'  
'continental_gt_speed' 'continental_gt3-r' 'continental_gt']
```

'continental_gtc_speed' 'continental_gtc'
 'continental_supersports_convertible' 'continental_supersports'
 'continental' 'contour_svt' 'contour' 'corniche' 'corolla_im' 'corolla'
 'corrado' 'corsica' 'corvette_stingray' 'corvette' 'coupe' 'cr-v' 'cr-z'
 'cressida' 'crossfire' 'crosstour' 'crosstrek' 'crown_victoria'
 'cruze_limited' 'cruze' 'ct_200h' 'ct6' 'cts_coupe' 'cts-v_coupe'
 'cts-v_wagon' 'cts-v' 'cts_wagon' 'cts' 'cube' 'custom_cruiser'
 'cutlass_calais' 'cutlass_ciera' 'cutlass_supreme' 'cutlass' 'cx-3'
 'cx-5' 'cx-7' 'cx-9' 'dakota' 'dart' 'dawn' 'daytona' 'db7' 'db9_gt'
 'db9' 'dbs' 'defender' 'deville' 'diablo' 'diamante'
 'discovery_series_ii' 'discovery_sport' 'discovery' 'dts' 'durango'
 'dynasty' 'e-150' 'e-250' 'e-class' 'e-golf' 'e-series_van'
 'e-series_wagon' 'e55_amg' 'echo' 'eclipse_spyder' 'eclipse' 'edge'
 'eighty-eight_royale' 'eighty-eight' 'elantra_coupe' 'elantra_gt'
 'elantra_touring' 'elantra' 'eldorado' 'electra' 'element' 'elise'
 'enclave' 'encore' 'endeavor' 'entourage' 'envision' 'envoy_xl'
 'envoy_xuv' 'envoy' 'enzo' 'eos' 'equator' 'equinox' 'equus' 'es_250'
 'es_300h' 'es_300' 'es_330' 'es_350' 'escalade_esv' 'escalade_ext'
 'escalade_hybrid' 'escalade' 'escape_hybrid' 'escape' 'escort' 'esprit'
 'estate_wagon' 'esteem' 'eurovan' 'evora_400' 'evora' 'ex35' 'excel'
 'exige' 'ex' 'expedition' 'explorer_sport_trac' 'explorer_sport'
 'explorer' 'expo' 'express_cargo' 'express' 'f-150_heritage'
 'f-150_svt_lightning' 'f-150' 'f-250' 'f12_berlinetta' 'f430' 'festiva'
 'ff' 'fiesta' 'firebird' 'fit_ev' 'fit' 'five_hundred' 'fj_cruiser'
 'fleetwood' 'flex' 'flying_spur' 'focus_rs' 'focus_st' 'focus' 'forenza'
 'forester' 'forte' 'fox' 'fr-s' 'freelander' 'freestar' 'freestyle'
 'frontier' 'fusion_hybrid' 'fusion' 'fx35' 'fx45' 'fx50' 'fx' 'g-class'
 'g_convertible' 'g_coupe' 'g_sedan' 'g20' 'g35' 'g37_convertible'
 'g37_coupe' 'g37_sedan' 'g37' 'g3' 'g5' 'g6' 'g80' 'g8' 'galant'
 'gallardo' 'genesis_coupe' 'genesis' 'ghibli' 'ghost_series_ii' 'ghost'
 'gl-class' 'gla-class' 'glc-class' 'gle-class_coupe' 'gle-class' 'gli'
 'glk-class' 'gls-class' 'golf_alltrack' 'golf_gti' 'golf_r'
 'golf_sportswagen' 'golf' 'grand_am' 'grand_caravan' 'grand_prix'
 'grand_vitara' 'grand_voyager' 'gransport' 'granturismo_convertible'
 'granturismo' 'gs_200t' 'gs_300' 'gs_350' 'gs_400' 'gs_430' 'gs_450h'
 'gs_460' 'gs_f' 'gt-r' 'gt' 'gti' 'gto' 'gx_460' 'gx_470' 'h3' 'h3t'
 'hhr' 'highlander_hybrid' 'highlander' 'horizon' 'hr-v' 'hs_250h'
 'huracan' 'i-miev' 'i30' 'i35' 'i3' 'ia' 'ilx_hybrid' 'ilx'
 'impala_limited' 'impala' 'imperial' 'impreza_wrx' 'impreza' 'im'
 'insight' 'integra' 'intrepid' 'intrigue' 'iq' 'is_200t' 'is_250_c'
 'is_250' 'is_300' 'is_350_c' 'is_350' 'is_f' 'j30' 'jetta_gli'
 'jetta_hybrid' 'jetta_sportswagen' 'jetta' 'jimmy' 'journey' 'juke'
 'justy' 'jx' 'k900' 'kizashi' 'lacrosse' 'lancer_evolution'
 'lancer_sportback' 'lancer' 'land_cruiser' 'landaulet' 'laser' 'le_baron'
 'le_mans' 'leaf' 'legacy' 'legend' 'lesabre' 'levante' 'lfa' 'lhs'
 'loyale' 'lr2' 'lr3' 'lr4' 'ls_400' 'ls_430' 'ls_460' 'ls_600h_1' 'ls'
 'lss' 'ltd_crown_victoria' 'lucerne' 'lumina_minivan' 'lumina' 'lx_450'
 'lx_470' 'lx_570' 'm-class' 'm2' 'm30' 'm35' 'm37' 'm3' 'm4_gts' 'm45'

'm4' 'm56' 'm5' 'm6_gran_coupe' 'm6' 'macan' 'magnum' 'malibu_classic'
'malibu_hybrid' 'malibu_limited' 'malibu_maxx' 'malibu' 'mark_lt'
'mark_viii' 'mark_vii' 'matrix' 'maxima' 'maybach' 'mazdaspeed_3'
'mazdaspeed_6' 'mazdaspeed_mx-5_miata' 'mazdaspeed_protege' 'm' 'mdx'
'metris' 'metro' 'mighty_max_pickup' 'millenia' 'mirage_g4' 'mirage'
'mkc' 'mks' 'mkt' 'mkx' 'mkz_hybrid' 'mkz' 'ml55_amg' 'model_s' 'monaco'
'montana_sv6' 'montana' 'monte_carlo' 'montero_sport' 'montero' 'mp4-12c'
'mpv' 'mr2_spyder' 'mr2' 'mulsanne' 'murano_crosscabriolet' 'murano'
'murcielago' 'mustang_svt_cobra' 'mustang' 'mx-3' 'mx-5_miata' 'mx-6'
'navajo' 'navigator' 'neon' 'new_beetle' 'new_yorker' 'ninety-eight'
'nitro' 'nsx' 'nv200' 'nx_200t' 'nx_300h' 'nx' 'odyssey' 'omni'
'optima_hybrid' 'optima' 'outback' 'outlander_sport' 'outlander'
'pacific' 'panamera' 'park_avenue' 'park_ward' 'paseo' 'passat'
'passport' 'pathfinder' 'phaeton' 'phantom_coupe'
'phantom_drophead_coupe' 'phantom' 'pickup' 'pilot' 'precis' 'prelude'
'previa' 'prius_c' 'prius_prime' 'prius_v' 'prius' 'prizm' 'probe'
'protege5' 'protege' 'prowler' 'pt_cruiser' 'pulsar' 'q3' 'q40' 'q45'
'q50' 'q5' 'q60_convertible' 'q60_coupe' 'q70' 'q7' 'quattroporte'
'quest' 'qx4' 'qx50' 'qx56' 'qx60' 'qx70' 'qx80' 'qx' 'r-class' 'r32'
'r8' 'rabbit' 'raider' 'rainier' 'rally_wagon' 'ram_150' 'ram_250'
'ram_50_pickup' 'ram_cargo' 'ram_pickup_1500' 'ram_van' 'ram_wagon'
'ramcharger' 'range_rover_evoque' 'range_rover_sport' 'range_rover'
'ranger' 'rapide_s' 'rapide' 'rav4_ev' 'rav4_hybrid' 'rav4' 'rc_200t'
'rc_300' 'rc_350' 'rc_f' 'rdx' 'reatta' 'regal' 'regency' 'rendezvous'
'reno' 'reventon' 'ridgeline' 'rio' 'riviera' 'rl' 'rlx' 'roadmaster'
'rogue_select' 'rogue' 'rondo' 'routan' 'rs_4' 'rs_5' 'rs_6' 'rs_7' 'rsx'
'rx_300' 'rx_330' 'rx_350' 'rx_400h' 'rx_450h' 'rx-7' 'rx-8'
's-10_blazer' 's-10' 's-15_jimmy' 's-15' 's-class' 's2000' 's3' 's40'
's4' 's5' 's60_cross_country' 's60' 's6' 's70' 's7' 's80' 's8' 's90'
'safari_cargo' 'safari' 'samurai' 'santa_fe_sport' 'santa_fe'
'savana_cargo' 'savana' 'sc_300' 'sc_400' 'sc_430' 'scoupe' 'sebring'
'sedona' 'sentra' 'sephia' 'sequoia' 'seville' 'shadow' 'shelby_gt350'
'shelby_gt500' 'sidekick' 'sienna' 'sierra_1500_classic'
'sierra_1500_hybrid' 'sierra_1500' 'sierra_1500hd' 'sierra_c3'
'sierra_classic_1500' 'sigma' 'silhouette' 'silver_seraph'
'silverado_1500_classic' 'silverado_1500_hybrid' 'silverado_1500'
'sixty_special' 'skylark' 'sl-class' 'slc-class' 'slk-class'
'slr_mclaren' 'sls_amg_gt_final_edition' 'sls_amg_gt' 'sls_amg' 'slx'
'solstice' 'sonata_hybrid' 'sonata' 'sonic' 'sonoma' 'sorento' 'soul_ev'
'soul' 'spark_ev' 'spark' 'spectra' 'spirit' 'sportage' 'sportvan'
'spyder' 'sq5' 'srt_viper' 'srx' 'ss' 'ssr' 'stanza' 'stealth' 'stratus'
'sts-v' 'sts' 'suburban' 'sunbird' 'sundance' 'sunfire' 'superamerica'
'supersports_convertible_isr' 'supra' 'svx' 'swift' 'sx4' 'syclone'
't100' 'tacoma' 'tahoe_hybrid' 'tahoe_limited/z71' 'tahoe' 'taurus_x'
'taurus' 'tc' 'tempo' 'tercel' 'terrain' 'terrazza' 'thunderbird'
'tiburon' 'tiguan' 'titan' 'tl' 'tlx' 'toronado' 'torrent' 'touareg_2'
'touareg' 'town_and_country' 'town_car' 'tracker' 'trailblazer_ext'
'trailblazer' 'trans_sport' 'transit_connect' 'transit_wagon' 'traverse'

```
'trax' 'tribeca' 'tribute_hybrid' 'tribute' 'truck' 'tsx_sport_wagon'
'tsx' 'tt_rs' 'tt' 'tts' 'tucson' 'tundra' 'typhoon' 'uplander'
'v12_vanquish' 'v12_vantage_s' 'v12_vantage' 'v40' 'v50'
'v60_cross_country' 'v60' 'v70' 'v8_vantage' 'v8' 'v90' 'vanagon'
'vandura' 'van' 'vanquish' 'vanwagon' 'veloster' 'venture' 'venza'
'veracruz' 'verano' 'verona' 'versa_note' 'versa' 'veyron_16.4' 'vibe'
'vigor' 'viper' 'virage' 'vitara' 'voyager' 'windstar_cargo' 'windstar'
'wraith' 'wrx' 'x-90' 'x1' 'x3' 'x4' 'x5_m' 'x5' 'x6_m' 'x6' 'xa' 'xb'
'xc60' 'xc70' 'xc90' 'xc' 'xd' 'xg300' 'xg350' 'xl-7' 'xl7' 'xlr-v' 'xlr'
'xt5' 'xterra' 'xts' 'xt' 'xv_crosstrek' 'yaris_ia' 'yaris'
'yukon_denali' 'yukon_hybrid' 'yukon_xl' 'yukon' 'z3' 'z4_m' 'z4' 'z8'
'zdx' 'zephyr']
```

year

```
[2011 2012 2013 1992 1993 1994 2017 1991 2016 1990 2015 1996 1997 1998
2014 1999 2002 2003 2004 1995 2007 2008 2009 2001 2010 2000 2005 2006]
```

engine_fuel_type

```
['premium_unleaded_(required)' 'regular_unleaded'
'premium_unleaded_(recommended)' 'flex-fuel_(unleaded/e85)' 'diesel'
'electric' 'flex-fuel_(premium_unleaded_recommended/e85)' 'natural_gas'
'flex-fuel_(premium_unleaded_required/e85)'
'flex-fuel_(unleaded/natural_gas)' nan]
```

engine_hp

```
[ 335.  300.  230.  320.  172.  160.  130.  158.  240.  248.  162.  217.
 184.  295.  115.  140.  155.  114.  100.  241.  180.  177.  228.  121.
 148.  194.  218.  161.  292.  250.  255.  222.   82.  134.  306.  400.
 425.  350.  332.  268.  282.  275.  201.  442.  562.  597.  237.  270.
 445.  443.  302.  322.  315.   nan  101.  135.  485.  238.  515.  543.
 631.  604.  620.  611.  661.  157.  402.  389.  110.  532.  170.  165.
 125.  641.  535.  153.  144.  188.  372.  108.  168.  190.  205.  200.
 227.  173.  220.  210.  280.  207.  265.  260.  290.  285.  390.  225.
 185.  150.  430.  520.  560.  475.  500.  540.  370.  580.  420.  345.
 195.  193.  208.  181.  236.  186.  252.  310.  333.  340.  450.  281.
 288.  138.  137.  106.  271.  196.  212.  278.  189.  480.  152.  600.
 375.  198.  182.  179.  264.  503.  456.  317.  235.  385.  303.   63.
 321.  272.  464.  202.  215.  283.  700.  720.  750.  107.  293.  119.
 143.  245.  120.  337.  276.  330.  132.  199.  530.  451.  329.  469.
 362.   94.  553.  453.  483.  323.  426.  505.  455.  650.  178.  242.
 305.  605.  440.  570.  325.  175.  707.  131.   62.   92.  102.  127.
 174.  621.  510.  429.  536.  355.  382.  577.  113.  136.  234.  552.
 626.  616.  572.  521.  567.  582.  460.  164.  192.  224.  239.  404.
 318.  556.  640.  122.  146.  244.  273.  563.  141.  435.  550.  360.
 145.  349.  166.  147.  128.  197.  291.  660.  261.  156.  403.   95.]
```

```

297.  81.  257.  365.  203.  231.  731.  651.  287.  123.  126.  416.
343.  348.  328.  298.  171.  219.  221.  311.  361.  256.  415.  274.
449.  395.  401.  454.  444.  338.  342.  467.  545.  565.  301.  263.
 93.  187.  610.  66.  111.  98.  204.  211.  73.  304.  381.  142.
 74.  424.  253.  90.  386.  359.  438.  232.  383.  518.  493.  259.
523.  55.  79.  116.  151.  78.  191.  592.  632.  670.  88.  167.
118.  380.  214.  573.  284.  99.  103.  525.  254.  470.  154.  176.
279.  377.  251.  223.  308.  105.  316.  124.  526.  662.  266.  296.
557.  617.  583.  622.  84.  163.  354.  159.  96.  206.  169.  133.
568.  109. 1001.  645.  490.  624.  410.  97.  394.]

```

engine_cylinders

```
[ 6.  4.  5.  8. 12.  0. nan 10.  3. 16.]
```

transmission_type

```
['manual' 'automatic' 'automated_manual' 'direct_drive' 'unknown']
```

driven_wheels

```
['rear_wheel_drive' 'front_wheel_drive' 'all_wheel_drive'
'four_wheel_drive']
```

number_of_doors

```
[ 2.  4.  3. nan]
```

market_category

```

['factory_tuner,luxury,high-performance' 'luxury,performance'
'luxury,high-performance' 'luxury' 'performance' 'flex_fuel'
'flex_fuel,performance' nan 'hatchback' 'hatchback,luxury,performance'
'hatchback,luxury' 'luxury,high-performance,hybrid' 'diesel,luxury'
'hatchback,performance' 'hatchback,factory_tuner,performance'
'high-performance' 'factory_tuner,high-performance'
'exotic,high-performance' 'exotic,factory_tuner,high-performance'
'factory_tuner,performance' 'crossover' 'exotic,luxury'
'exotic,luxury,high-performance' 'exotic,luxury,performance'
'factory_tuner,luxury,performance' 'flex_fuel,luxury' 'crossover,luxury'
'hatchback,factory_tuner,luxury,performance' 'crossover,hatchback'
'hybrid' 'luxury,performance,hybrid'
'crossover,luxury,performance,hybrid' 'crossover,luxury,performance'
'exotic,factory_tuner,luxury,high-performance'
'flex_fuel,luxury,high-performance' 'crossover,flex_fuel' 'diesel'
'hatchback,diesel' 'crossover,luxury,diesel'
'crossover,luxury,high-performance'
'exotic,flex_fuel,factory_tuner,luxury,high-performance'

```

```

'exotic,flex_fuel,luxury,high-performance'
'exotic,factory_tuner,luxury,performance' 'hatchback,hybrid'
'crossover,hybrid' 'hatchback,luxury,hybrid'
'flex_fuel,luxury,performance' 'crossover,performance' 'luxury,hybrid'
'crossover,flex_fuel,luxury,performance' 'crossover,flex_fuel,luxury'
'crossover,flex_fuel,performance'
'hatchback,factory_tuner,high-performance' 'hatchback,flex_fuel'
'factory_tuner,luxury' 'crossover,factory_tuner,luxury,high-performance'
'crossover,factory_tuner,luxury,performance'
'crossover,hatchback,factory_tuner,performance'
'crossover,hatchback,performance' 'flex_fuel,hybrid'
'flex_fuel,performance,hybrid' 'crossover,exotic,luxury,high-performance'
'crossover,exotic,luxury,performance' 'exotic,performance'
'exotic,luxury,high-performance,hybrid' 'crossover,luxury,hybrid'
'flex_fuel,factory_tuner,luxury,high-performance' 'performance,hybrid'
'crossover,factory_tuner,performance' 'crossover,diesel'
'flex_fuel,diesel' 'crossover,hatchback,luxury']

```

vehicle_size

```
['compact' 'midsize' 'large']
```

vehicle_style

```

['coupe' 'convertible' 'sedan' 'wagon' '4dr_hatchback' '2dr_hatchback'
'4dr_suv' 'passenger_minivan' 'cargo_minivan' 'crew_cab_pickup'
'regular_cab_pickup' 'extended_cab_pickup' '2dr_suv' 'cargo_van'
'convertible_suv' 'passenger_van']

```

highway_mpg

```

[ 26  28  27  25  24  20  21  22  35  34  31  30  32  33  23  36  29  45
 43  40  42  19  18  17  15  37  39  41  16  14  38 108 103  12 354  47
 46  82 110  44  13 105  99 102 111 106  48 101  90  97  94  98 100 107
 53  50  74  92 109]

```

city_mpg

```

[ 19  20  18  17  16  26  23  22  21  24  15  25  29  28  32  31  30  14
 10  27  12  13   9  11 122 121   8  50  49  47  35  33  40  85 128  42
 43  36  44 126   7  34 132 110 137 129  39  41 124  37  88  94  86 101
 95  89 102  98  92  97  91  53  55  51  54  58  78  38 120]

```

popularity

```

[3916 3105  819  617 1013 2009  870  586  436 2774  113 2031  416   67
 210 1715  376  549 1439  535 2202   26  481 5657 1624 1720  520 1385
1851 1158   61  640  873    2  155  204   86  238  454  259  258  613]

```



```
190 105 21 130 1391 820]
```

```
msrp  
[46135 40650 36350 ... 46120 50620 50920]
```

```
[ ]: # sns.pairplot(cardata)
```

```
[ ]: cardata.corr()
```

```
/tmp/ipykernel_17812/2059604511.py:1: FutureWarning: The default value of  
numeric_only in DataFrame.corr is deprecated. In a future version, it will  
default to False. Select only valid columns or specify the value of numeric_only  
to silence this warning.
```

```
cardata.corr()
```

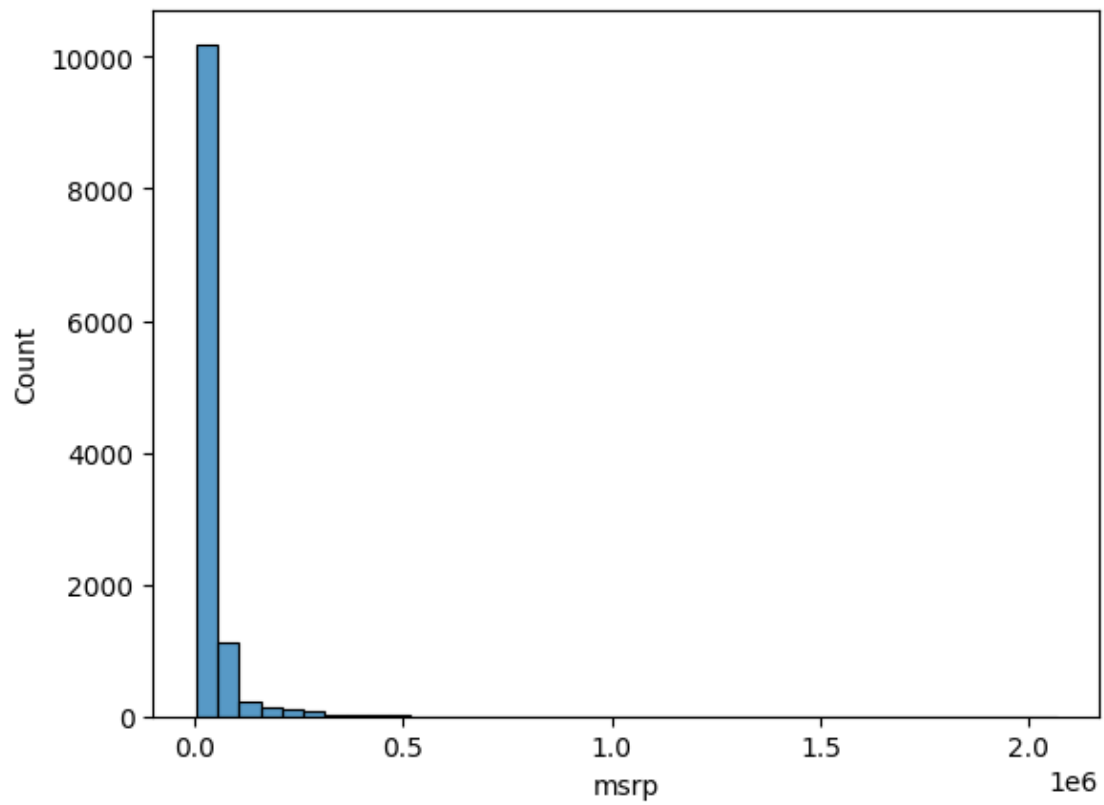
```
[ ]:
```

	year	engine_hp	engine_cylinders	number_of_doors	\
year	1.000000	0.351794	-0.041479	0.263787	
engine_hp	0.351794	1.000000	0.779988	-0.102713	
engine_cylinders	-0.041479	0.779988	1.000000	-0.140088	
number_of_doors	0.263787	-0.102713	-0.140088	1.000000	
highway_mpg	0.258240	-0.406563	-0.621606	0.118570	
city_mpg	0.198171	-0.439371	-0.600776	0.120881	
popularity	0.073049	0.037501	0.041145	-0.048272	
msrp	0.227590	0.662008	0.531312	-0.126635	

	highway_mpg	city_mpg	popularity	msrp
year	0.258240	0.198171	0.073049	0.227590
engine_hp	-0.406563	-0.439371	0.037501	0.662008
engine_cylinders	-0.621606	-0.600776	0.041145	0.531312
number_of_doors	0.118570	0.120881	-0.048272	-0.126635
highway_mpg	1.000000	0.886829	-0.020991	-0.160043
city_mpg	0.886829	1.000000	-0.003217	-0.157676
popularity	-0.020991	-0.003217	1.000000	-0.048476
msrp	-0.160043	-0.157676	-0.048476	1.000000

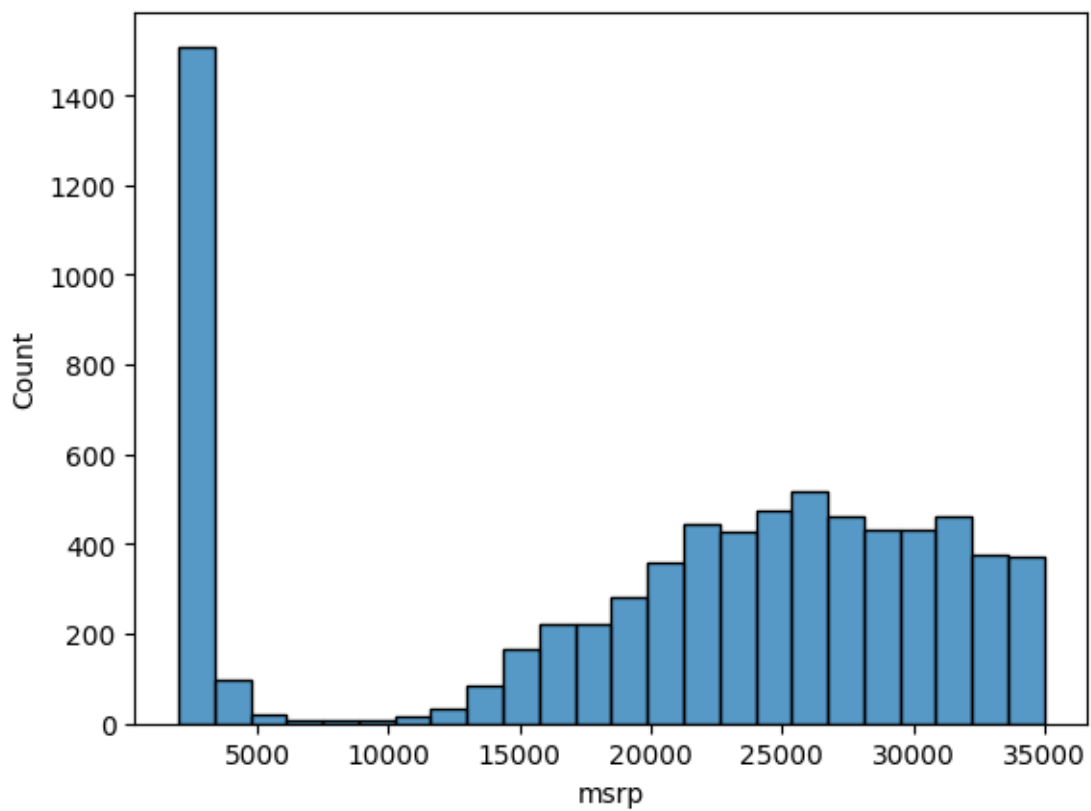
```
[ ]: sns.histplot(cardata.msrp, bins=40)
```

```
[ ]: <Axes: xlabel='msrp', ylabel='Count'>
```



```
[ ]: sns.histplot(cardata.msrp[cardata.msrp < 35000])
```

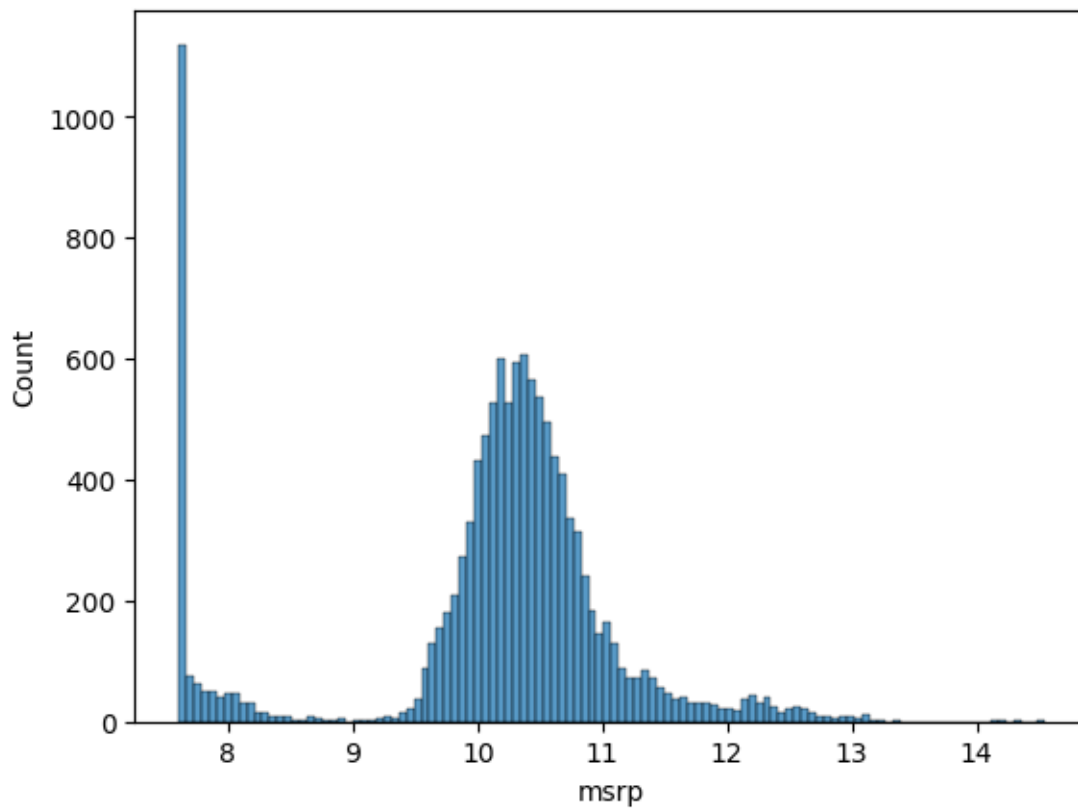
```
[ ]: <Axes: xlabel='msrp', ylabel='Count'>
```



```
[ ]: log_price = np.log1p(cardata.msrp)
```

```
[ ]: sns.histplot(log_price)
```

```
[ ]: <Axes: xlabel='msrp', ylabel='Count'>
```



```
[ ]: cardata.isnull().sum()
```

```
[ ]: make          0
      model         0
      year          0
      engine_fuel_type  3
      engine_hp      69
      engine_cylinders 30
      transmission_type 0
      driven_wheels   0
      number_of_doors  6
      market_category 3742
      vehicle_size     0
      vehicle_style    0
      highway_mpg      0
      city_mpg         0
      popularity       0
      msrp             0
      dtype: int64
```

```
[ ]: cardata_train, cardata_val, cardata_test = mf.train_val_test_split(cardata,0.
    ↪6,0.20,0.20,'msrp',5)

y_train = cardata_train.msrp
y_val = cardata_val.msrp
y_test = cardata_test.msrp

del cardata_train['msrp']
del cardata_val['msrp']
del cardata_test['msrp']

# length = len(cardata)
# constant = 5
# val_p = 0.2
# test_p = 0.2
# train_p = 0.6

# n_val = int(val_p * length)
# n_test = int(test_p * length)
# n_train = int(train_p * length)

# np.random.seed(constant)
# idx = np.arange(length)
# np.random.shuffle(idx)

# cardata_shuffled = cardata.iloc[idx]

# cardata_train = cardata_shuffled.iloc[:n_train].copy()
# cardata_val = cardata_shuffled.iloc[n_train:n_train+n_val].copy()
# cardata_test = cardata_shuffled.iloc[n_train+n_val:].copy()

print(y_val.shape)
print(y_test.shape)
print(y_train.shape)
print(cardata_val.shape)
```

```
(2382,)
(2384,)
(7148,)
(2382, 15)
```

```
[ ]: cardata_train
```

```
[ ]:      make      model  year      engine_fuel_type \
3863  hyundai    elantra  2016      regular_unleaded
7113   mazda    mx-5_miata  2015  premium_unleaded_(recommended)
917    audi      90    1995      regular_unleaded
```

3833	hyundai	elantra_coupe	2013	regular_unleaded
1312	suzuki	aerio	2005	regular_unleaded
...
2860	bentley	continental_gt	2014	premium_unleaded_(required)
10347	gmc	terrain	2015	regular_unleaded
6815	infiniti	m	2012	premium_unleaded_(required)
2055	subaru	brz	2017	premium_unleaded_(required)
1531	cadillac	ats	2015	flex-fuel_(unleaded/e85)

	engine_hp	engine_cylinders	transmission_type	driven_wheels	\
3863	173.0	4.0	automatic	front_wheel_drive	
7113	167.0	4.0	manual	rear_wheel_drive	
917	172.0	6.0	manual	all_wheel_drive	
3833	145.0	4.0	automatic	front_wheel_drive	
1312	155.0	4.0	automatic	front_wheel_drive	
...	
2860	521.0	8.0	automatic	all_wheel_drive	
10347	182.0	4.0	automatic	all_wheel_drive	
6815	330.0	6.0	automatic	all_wheel_drive	
2055	205.0	4.0	manual	rear_wheel_drive	
1531	321.0	6.0	automatic	rear_wheel_drive	

	number_of_doors	market_category	\
3863	4.0	performance	
7113	2.0	performance	
917	4.0	luxury	
3833	2.0	NaN	
1312	4.0	NaN	
...	
2860	2.0	exotic,factory_tuner,luxury,high-performance	
10347	4.0	crossover	
6815	4.0	luxury,performance	
2055	2.0	performance	
1531	4.0	flex_fuel,luxury,high-performance	

	vehicle_size	vehicle_style	highway_mpg	city_mpg	popularity
3863	compact	sedan	35	24	1439
7113	compact	convertible	28	22	586
917	compact	sedan	23	17	3105
3833	compact	coupe	37	27	1439
1312	compact	sedan	28	22	481
...
2860	midsize	coupe	24	15	520
10347	compact	4dr_suv	29	20	549
6815	large	sedan	24	17	190
2055	compact	coupe	29	21	640
1531	compact	sedan	28	18	1624

[7148 rows x 15 columns]

```
[ ]: base = ['engine_hp', 'engine_cylinders', 'highway_mpg', 'city_mpg',  
           ↪ 'popularity']  
df_num = cardata_train[base]
```

```
[ ]: df_num = df_num.fillna(0)
```

```
[ ]: X_train = df_num.values
```

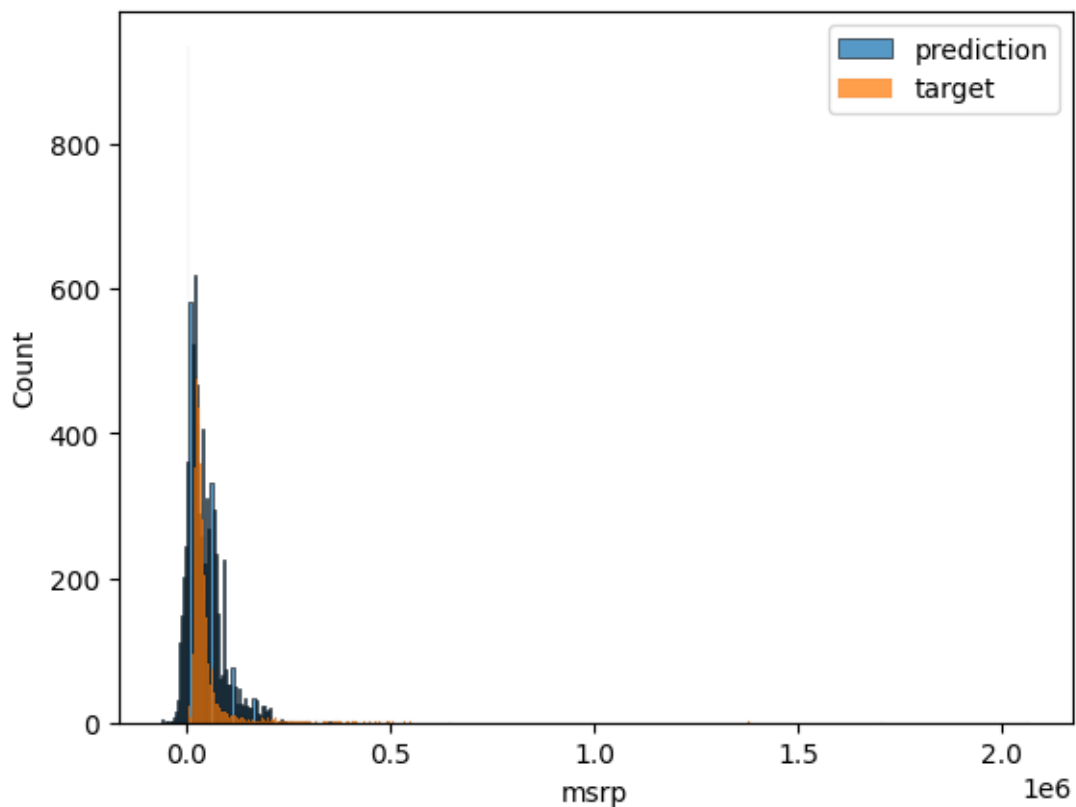
```
[ ]: w_0, w = mf.train_linear_regression(X_train, y_train, r=0.001)  
print(w.shape)
```

(5,)

```
[ ]: y_pred = w_0 + X_train.dot(w)
```

```
[ ]: sns.histplot(y_pred, label='prediction')  
sns.histplot(y_train, label='target')  
plt.legend()
```

```
[ ]: <matplotlib.legend.Legend at 0x7fa8b1ed3a90>
```



RMSE

```
[ ]: mf.rsme(y_train, y_pred)
      print(y_train.shape)
      print(y_pred.shape)
```

(7148,)

(7148,)

Validation

```
[ ]: df_num_val = cardata_val[base]
      df_num_val = df_num_val.fillna(0)
      X_val = df_num_val.values
```

```
[ ]: y_pred_val = mf.lin_regress(w_0, w, X_val)
```

```
[ ]: mf.rsme(y_val, y_pred_val)
```

```
[ ]: 30680.895412192694
```

```
[ ]: def prepare_X(input_df, base_columns):
      df_num_prep = input_df[base_columns]
      df_num_prep = df_num_prep.fillna(0)
      X_prep = df_num_prep.values
      return X_prep
```

Simplified and put together

```
[ ]: X_train = prepare_X(cardata_train, base)
      w_0, w = mf.train_linear_regression(X_train, y_train, r=0.001)

      X_val = prepare_X(cardata_val, base)
      y_pred = mf.lin_regress(w_0, w, X_val)
      print('vallidation',mf.rsme(y_val, y_pred))
```

vallidation 30680.895412192694

Feature Engineering

```
[ ]: X_train = mf.MLB_ch2_prepare_X(cardata_train, base)
      w_0, w = mf.train_linear_regression(X_train, y_train, r=0.001)

      X_val = mf.MLB_ch2_prepare_X(cardata_val, base)
      y_pred = mf.lin_regress(w_0, w, X_val)
      print('validation:',mf.rsme(y_val, y_pred))
```

validation: 27728.416985928914


```
[ ]: X_train = mf.MLB_ch2_prepare_X(cardata_train, base)
w_0, w = mf.train_linear_regression(X_train, y_train, r=0.01)

X_val = mf.MLB_ch2_prepare_X(cardata_val, base)
y_pred = mf.lin_regress(w_0, w, X_val)
print('validation:',mf.rsme(y_val, y_pred))
```

validation: 27728.373488809757

```
[ ]: ad = {
    'city_mpg':18,
    'driven_wheels': 'all_wheel_drive',
    'engine_cylinders': 6.0,
    'engine_fuel_type': 'regular_unleaded',
    'engine_hp': 268.0,
    'highway_mpg':25,
    'make': 'toyota',
    'market_category': 'crossover,performance',
    'model': 'venza',
    'number_of_doors': 4.0,
    'popularity': 2031,
    'transmission_type': 'automatic',
    'vehicle_size': 'midsize',
    'vehicle_style': 'wagon',
    'year': 2013
}

df_test = pd.DataFrame([ad])
X_test = mf.MLB_ch2_prepare_X(df_test,base)
y_pred = mf.lin_regress(w_0, w, X_test)
suggestion = y_pred
suggestion
```

```
[ ]: mytruck = {
    'city_mpg':22,
    'driven_wheels': 'rear_wheel_drive',
    'engine_cylinders': 4.0,
    'engine_fuel_type': 'regular_unleaded',
    'engine_hp': 143.0,
    'highway_mpg':28,
    'make': 'ford',
    'market_category': 'crossover',
    'model': 'venza',
    'number_of_doors': 2.0,
    'popularity': 7385,
    'transmission_type': 'automatic',
    'vehicle_size': 'midsize',
```

```
'vehicle_style': 'wagon',  
'year': 2009  
}
```

```
df_test = pd.DataFrame([mytruck])  
X_test = mf.MLB_ch2_prepare_X(df_test,base)  
y_pred = mf.lin_regress(w_0, w, X_test)  
suggestion = y_pred  
suggestion
```

```
[ ]: array([-28783.25850217])
```

```
[ ]:
```