MATLAB Analysis Project

Automobile Sample

NAME: OYERO OLUWATOBI ADEDOLAPO

STUDENT ID: 215924525

COURSE: AP/ITEC 2600 A

Sample: Automobile sample from UCI Machine Learning Repository

A picture containing screenshot

Description automatically generated

A = readtable('Automobiles1.xlsx')

A.Properties.VariableNames = {'symboling' 'normalized\_losses' 'make' 'fuel\_type' 'aspiration' 'num\_of\_doors' 'body\_style' 'drive\_wheels' 'engine\_location' 'wheel\_base' 'length' 'width' 'height' 'curb\_weight' 'engine\_type' 'num\_of\_cylinders' 'engine\_size' 'fuel\_system' 'bore' 'stroke' 'compression\_ratio' 'horsepower' 'peak\_rpm' 'city\_mpg' 'highway\_mpg' 'price'}

Explanation: This code was intended to import the sample above and add titles to the columns

Cleaning data

A.symboling = []

A.normalized\_losses = []

A.aspiration = []

A.drive\_wheels = []

A.wheel\_base = []

A.curb\_weight = []

A.engine\_type = []

A.fuel\_system = []

A.bore = []

A.stroke = []

A.compression\_ratio = []

A.city\_mpg = []

A.highway\_mpg = []

A.fuel\_type = []

A.num\_of\_doors = []

A.engine\_location = []

A.num\_of\_cylinders = []

A.peak\_rpm = []

A screenshot of a cell phone

Description automatically generated

Explanation: I removed the columns listed above because I didn’t find any specific use for them. The ones I chose to keep were easier to understand and think about in reality.

strcmp(A.horsepower, '?')

strcmp(A.peak\_rpm, '?')

strcmp(A.price, '?')

A(strcmp(A.horsepower, '?'), :) = []

A(strcmp(A.peak\_rpm, '?'),:) = []

A(strcmp(A.price, '?'),:) = []

A screenshot of a cell phone

Description automatically generated

Explanation: I used this code to remove the rows with unknown data fields

Analysis

Body Styles

sdncount = 0

hbcount = 0

cvtcount = 0

wagcount = 0

htcount = 0

B1 = string(A.body\_style)

for i = 1:length(B1)

if B1(i) == "sedan"

sdncount = sdncount + 1

else if B1(i) == "hatchback"

hbcount = hbcount + 1

else if B1(i) == "convertible"

cvtcount = cvtcount + 1

else if B1(i) == "wagon"

wagcount = wagcount + 1

else

htcount = htcount + 1

end

end

end

end

end

bs\_counts = [sdncount hbcount cvtcount wagcount htcount]

Output:

bs\_counts = 94 67 5 24 8

Explanation: I used the preceding code to count the number of cars under specific body styles to ascertain the popularity of the different styles.

pbs = pie(bs\_counts)

pText = findobj(pbs,'Type','text')

percentValues = get(pText,'String')

txt = {'Sedan: ';'Hatchback: ';'Convertible: ';'Wagon: ';'Hardtop: '}

combinedtxt = strcat(txt,percentValues)

labels = combinedtxt

pbs = pie(bs\_counts, labels)

title('Body Styles')

A picture containing screenshot

Description automatically generated

Explanation: The code above was used to make a pie chart to compare Body styles using the vectors created before.

Prices

A.price = str2double(A.price)

min(A.price)

max(A.price)

mean(A.price)

histogram(A.price)

title('Prices')

Output:

Min: 5118

Max: 45400

Mean: 1.3242e + 04

A picture containing screenshot

Description automatically generated

Explanation: Analysis of prices with minimum, maximum and mean values. Also, a histogram of the prices was created. This code was intended to check for patterns among prices.

Engine Sizes

min(A.engine\_size)

max(A.engine\_size)

mean(A.engine\_size)

histogram(A.engine\_size)

title('Engine Sizes')

Output:

Min: 61

Max: 326

Mean: 126.8081

A screenshot of a cell phone

Description automatically generated

Explanation: Analysis of engine sizes with minimum, maximum and mean values. A histogram of the engine sizes has also been created to check for patterns.

Engine Sizes and Prices

corrcoef(A.engine\_size, A.price)

Output:

1.0000 0.8739

0.8739 1.0000

Explanation: This code was intended to test for any correlation between engine size and price. From the results we can see that there is a positive correlation.

plot(A.engine\_size, A.price, 'linestyle', 'none', 'marker', '.')

xlabel('Engine Size')

ylabel('Price')

title('Engine Sizes and Prices')

A screenshot of a social media post

Description automatically generated

Explanation: The previous code created this scatter plot. From the scatter plot we can see that most of the data is converged near the mean engine sizes and prices. We also see that the sizes of engines is directly connected to the price.

Height

min(A.height)

max(A.height)

mean(A.height)

histogram(A.height)

title('Car Heights')

Output:

Min: 47.8000

Max: 59.8000

Mean: 53.8010

A picture containing screenshot

Description automatically generated

Explanation: Analysis of heights with minimum, maximum and mean values. A histogram has been created to show heights and trends among them.

Width

min(A.width)

max(A.width)

mean(A.width)

histogram(A.width)

title('Car Widths')

Output:

Min: 60.3000

Max: 72

Mean: 65.8914

A picture containing screenshot

Description automatically generated

Explanation: Analysis of widths with minimum, maximum and mean values. A histogram has been created to show widths and trends among them.

Length

min(A.length)

max(A.length)

mean(A.length)

histogram(A.length)

title('Car Lengths')

Output:

Min: 141.1000

Max: 208.1000

Mean: 174.1783

A picture containing screenshot

Description automatically generated

Explanation: Analysis of lengths with minimum, maximum and mean values. A histogram has been created to show lengths and trends among them.

Length, Height and Width

corrcoef(A.height, A.width)

Output:

1.0000 0.3043

0.3043 1.0000

corrcoef(A.height, A.length)

Output:

1.0000 0.4959

0.4959 1.0000

corrcoef(A.width, A.length)

Output:

1.0000 0.8573

0.8573 1.0000

Explanation: The lines of code above are intended to test for correlation between length, height and width. From the output we can see that the correlation between height and width/length is relatively weak (and positive). We can also see that the correlation between width and length is relatively strong (and positive).

plot(A.width, A.length, 'linestyle', 'none', 'marker', '.')

xlabel('Width')

ylabel('Length')

title('Width and Length')

A screenshot of a social media post

Description automatically generated

Explanation: The above code created this scatter plot. It shows the relationship between width and length as positive. This means that the trend suggests that car width is positively connected to car length.

Horsepower

A.horsepower = str2double(A.horsepower) %changing horsepower from cell to double

min(A.horsepower)

max(A.horsepower)

mean(A.horsepower)

histogram(A.horsepower)

title('Horsepower')

Output:

Min: 48

Max: 262

Mean: 103.3586

A screenshot of a video game

Description automatically generated

Explanation: Analysis of horsepower with minimum, maximum and mean values. A histogram has been created to show horsepower and trends among them.

Horsepower and Engine Size

corrcoef(A.horsepower, A.engine\_size)

Output:

1.0000 0.8227

0.8227 1.0000

Explanation: This code is intended to test the correlation between horsepower and engine size. We can see from the output that there’s a strong positive correlation between horsepower and engine size.

plot(A.horsepower, A.engine\_size, 'linestyle', 'none', 'marker', '+')

xlabel('Horsepower')

ylabel('Engine Size')

title('Horsepower and Engine Size')

A screenshot of a social media post

Description automatically generated

Explanation: The above code created this scatter plot. We can see that most of the data is converging near the mean of horsepower/engine size. The positive relationship can also be seen.

Horsepower and Price

corrcoef(A.horsepower, A.price)

Output:

1.0000 0.8106

0.8106 1.0000

Explanation: The code above was intended to test the correlation between horsepower and price. From the output we can see that there is a strong positive relation between horsepower and price. The higher the horsepower, the higher the price

plot(A.horsepower, A.price, 'linestyle', 'none', 'marker', '\*')

xlabel('Horsepower')

ylabel('Price')

title('Horsepower and Price')

A screenshot of a social media post

Description automatically generated

Explanation: The above code creates this scatter plot. The plot shows the positive relationship between horsepower and price. The data is also centered near the mean of horsepower/price.

Brand and Price

B2 = string(A.make)

volprice = []

vkwprice = []

toyprice = []

subprice = []

saabprice = []

porprice = []

plyprice = []

peugprice = []

nisprice = []

mitprice = []

merprice = []

mazprice = []

jagprice = []

isuprice = []

honprice = []

dodprice = []

chevprice = []

bmwprice = []

audiprice = []

arprice = []

for i = 1:length(A.make)

if B2(i) == "volvo"

volprice = [volprice, A.price(i)]

else if B2(i) == "volkswagen"

vkwprice = [vkwprice, A.price(i)]

else if B2(i) == "toyota"

toyprice = [toyprice, A.price(i)]

else if B2(i) == "subaru"

subprice = [subprice, A.price(i)]

else if B2(i) == "saab"

saabprice = [saabprice, A.price(i)]

else if B2(i) == "porsche"

porprice = [porprice, A.price(i)]

else if B2(i) == "plymouth"

plyprice = [plyprice, A.price(i)]

else if B2(i) == "peugot"

peugprice = [peugprice, A.price(i)]

else if B2(i) == "nissan"

nisprice = [nisprice, A.price(i)]

else if B2(i) == "mitsubishi"

mitprice = [mitprice, A.price(i)]

else if B2(i) == "mercedes-benz"

merprice = [merprice, A.price(i)]

else if B2(i) == "mazda"

mazprice = [mazprice, A.price(i)]

else if B2(i) == "jaguar"

jagprice = [jagprice, A.price(i)]

else if B2(i) == "isuzu"

isuprice = [isuprice, A.price(i)]

else if B2(i) == "honda"

honprice = [honprice, A.price(i)]

else if B2(i) == "dodge"

dodprice = [dodprice, A.price(i)]

else if B2(i) == "chevrolet"

chevprice = [chevprice, A.price(i)]

else if B2(i) == "bmw"

bmwprice = [bmwprice, A.price(i)]

else if B2(i) == "audi"

audiprice = [audiprice, A.price(i)]

else

arprice = [arprice, A.price(i)]

end

end

end

end

end

end

end

end

end

end

end

end

end

end

end

end

end

end

end

end

volmean = mean(volprice)

vkwmean = mean(vkwprice)

toymean = mean(toyprice)

submean = mean(subprice)

saabmean = mean(saabprice)

pormean = mean(porprice)

plymean = mean(plyprice)

peugmean = mean(peugprice)

nismean = mean(nisprice)

mitmean = mean(mitprice)

mermean = mean(merprice)

mazmean = mean(mazprice)

jagmean = mean(jagprice)

isumean = mean(isuprice)

honmean = mean(honprice)

dodmean = mean(dodprice)

chevmean = mean(chevprice)

bmwmean = mean(bmwprice)

audimean = mean(audiprice)

armean = mean(arprice)

car\_make\_means = [volmean, vkwmean, toymean, submean, saabmean, pormean, plymean, peugmean, nismean, mitmean, mermean, mazmean, jagmean, isumean, honmean, dodmean, chevmean, bmwmean, audimean, armean]

min(car\_make\_means)

max(car\_make\_means)

Output:

car\_make\_means:

1.0e+04 \*

[ 1.8063 1.0077 0.9886 0.8541 1.5223 3.1401 0.7963 1.5489 1.0416 0.9240 3.3647 1.0653 3.4600 0.8917 0.8185 0.7875 0.6007 2.6119 1.7859 1.6501]

Min: 6007

Max: 34600

Explanation: The code up to this point is intended to categorize the makes (brands) of cars by price. After which I found the average price of cars by brand and sorted them into vectors.

makes = ["volvo", "volkswagen", "toyota", "subaru", "saab", "porsche", "plymouth", "peugot", "nissan", "mitsubishi", "mercedes", "mazda", "jaguar", "isuzu", "honda", "dodge", "chevrolet", "bmw", "audi", "alfa-romero" ]'

make\_means = car\_make\_means'

MT = table(makes, make\_means)

A screenshot of a cell phone

Description automatically generated

Explanation: A table was created so that one could see the makes and their average prices side by side for comparison. From the previous output we see that (on average) Chevrolet makes the cheapest cars and Jaguar makes the most expensive ones.

bar(car\_make\_means)

ylabel('Mean Price')

Xticklabel = { 'volvo', 'volkswagen', 'toyota', 'subaru', 'saab', 'porsche', 'plymouth', 'peugot', 'nissan', 'mitsubishi', 'mercedes', 'mazda', 'jaguar', 'isuzu', 'honda', 'dodge', 'chevrolet', 'bmw', 'audi', 'alfa-romero' }

set(gca,'XtickLabel',Xticklabel)

xlabel('Brands')

title('Car Brands by Mean Price')

A picture containing screenshot

Description automatically generated

Explanation: The above code plots a bar graph to visually represent the different makes of cars by average cost.