Modeling of Tailpipe CO2 Emissions on Vehicle Characteristics

Davood Aein

Shiley-Marcos School of Engineering, University of San Diego

Probability and Statistics for Data Science (ADS-500A-01)

Professor Matthew C Vanderbilt

Assignment due date: 2/26/2024

Modeling of Tailpipe CO2 Emissions on Vehicle Characteristics

The dataset contains detailed information about various automobiles, encompassing factors such as fuel efficiency, emissions, and vehicle specifications. Analyzing this data holds significant importance in the context of environmental sustainability, fuel consumption optimization, and regulatory compliance within the automotive industry. The study could shed light on trends in vehicle performance, guiding manufacturers towards developing more fuel-efficient and eco-friendly vehicles. Furthermore, understanding the correlation between different features and emissions can aid policymakers in formulating effective regulations to mitigate environmental impact. The practical implications of this study extend to consumers, as it may empower them to make informed choices based on both performance and environmental considerations when selecting a vehicle. Overall, the analysis of this automotive dataset has the potential to influence industry practices, regulatory frameworks, and consumer behaviors towards a more sustainable and environmentally conscious future.

The dataset comprises a diverse population of automobiles, representing various makes and engine sizes. It includes vehicles from renowned manufacturers such as Alfa Romeo, Subaru, and Ferrari, offering a broad spectrum of choices for analysis. The engine sizes range from compact 1.8L models like the Subaru Loyale to more powerful 5.2L engines found in vehicles like the Dodge B150/B250 Wagon. This diversity in makes and engine sizes provides a comprehensive overview of the automotive landscape, allowing for a nuanced exploration of the relationships between these characteristics and factors like fuel efficiency and emissions. The inclusion of different vehicle types, from two-seaters to vans, further enhances the dataset's richness, enabling researchers to examine population characteristics across a wide array of automotive segments.

While the average driver might expect a car to guzzle around 17.16 barrels per gallon and spew 463 grams of CO2 per mile, this snapshot reveals a wider reality. On one end, fuel efficiency ranges from a frugal 14.33 barrels per gallon all the way up to a thirsty 47.09, highlighting the significant variance in automotive resource consumption. Similarly, CO2 emissions paint a contrasting picture, with some vehicles emitting a concerning 1269 grams per mile compared to others' more modest 386 grams. These statistics, despite stemming from a limited sample, hint at a broader spectrum of vehicles on the road, each with its own unique impact on the environment. Notably, the typical engine displacement of 3.3 liters and a median combined fuel economy of 20.8 miles per gallon offer a glimpse into the average car's performance. However, the true story lies in the extremes, suggesting that conscious consumer choices and technological advancements in the automotive industry hold significant potential for reducing our collective environmental footprint (Table 1).

In figure 1, we create a boxplot which is based on Cylinder & Co2tailpipeGpm columns. Since it is only about vehicles which have cylinders, there is no electric vehicle. We can see less variation for 2-cylinder and 16-cylinder vehicles. On the other hand, 8-cylinderhas has notable variation among them. In terms of extreme outliers, 4-cylinder and 6-cylinder vehicles are the top ones. After that, we can 2-cylinder,3-cylinder, 8-cylinder, 10-cylinder, and 12-cylinder vehicles which have the most outliers. As the number of Cylinder increase, the amount of Co2tailpipeGpm increase as well.

A standard formula was developed to represent the likelihood of CO2 emissions as shown in Equation (1). The formula considers CO2 emissions as the variable. Upon examining the practicality of this formula, it became clear that any negative value for CO2 emissions would result in a probability of zero. This is because current technology limitations make it impossible for vehicles to have carbon dioxide emissions. Additionally, the probability distribution follows a bell-shaped curve allowing the formula to estimate the percentage of CO2 emissions expected at values as depicted in Figure 2. For research purposes we adjusted the data points to align with a distribution by manipulating the number of bins. With this normally distributed histogram and probability chart, we can anticipate that 99.7% of CO2 emission values will fall within the range of 200 and 800 grams per mile.

$$\boldsymbol{PDF_{CO2} = f(x; u = 465.38, \sigma = 119.88) = \begin{cases} \frac{1}{\sqrt{2\pi(119.88)}} e^{-[(x-465.538)^2/[(2)(119)^2]}, & x > 0 \\ 0, & x < 0 \end{cases}}$$

(1)

Generalized Object Formula

The generalized object formula represented by Equation (2) is a multiple linear regression model used to predict carbon dioxide (CO2) emissions from vehicles. In this equation, CO2 is the dependent variable, and the right-hand side includes various independent variables denoted by β coefficients. β0 represents the y-intercept of the equation. Each β coefficient signifies the contribution of the corresponding independent variable to the overall prediction of CO2 emissions and ε represents the error term. This generalized object formula allows researchers to assess the impact of various factors on CO2 emissions, supporting hypothesis testing and analysis in the context of vehicle emissions.

$$CO2 = \beta 0 + \beta 1 barrels08 + \beta 2\, co2TailpipeGpm + \beta 3 comb08 + \beta 4 displ + \beta 5 cylinders + \beta 6 volume + \beta 7 emissionscat + \beta 8 makeid + \beta 8 fueltype + \beta 9 transmissiontype + \beta 10 vehicletype + \varepsilon \quad (2)$$

Bivariate Frequency

Table 2 encapsulates a comprehensive bivariate frequency analysis, examining the relationships between four distinct vehicle types and two transmission types across six fuel categories within a dataset comprising 43,177 samples. The independent variables in the table encompass "Unknown," "Hatchback," "Passenger Two-Door," and "Passenger Four-Door" vehicle types, as well as "Automatic" and "Manual" transmission types. This detailed breakdown allows for a nuanced exploration of how different vehicle and transmission types align with specific fuel preferences. The inclusion of the six fuel categories adds an additional layer of complexity to the analysis, providing insights into the diverse patterns of fuel usage within each combination of vehicle and transmission types. The table serves as a valuable resource for researchers seeking to understand the interplay between these categorical variables and contributes essential information for further statistical investigation and interpretation.

Table 2 is indicative of a structured investigation into the relationships between vehicle characteristics, specifically "Vehicle Type" and "Transmission Type." Each cell in the table provides the count (n) and percentage (%) of vehicles falling into distinct categories, shedding light on the prevalence and distribution of transmission types within various vehicle types. For instance, by scrutinizing the entries corresponding to "Hatchback" vehicles, researchers can discern the proportion of manual and automatic transmissions, unraveling potential trends in transmission preferences based on the vehicle's structural design. This bivariate frequency analysis acts as a preliminary exploration, offering valuable insights into the intricate patterns of vehicle characteristics that can subsequently inform more sophisticated statistical examinations.

Moreover, when examining the fuel type distribution, among the population we can observe some trends. Premium gasoline accounts for 12,801 samples, which makes up 30% of the population. Among these passenger four-door vehicles have the representation with a proportion of 38%. Unknown vehicles and passenger two-door vehicles have shares at around 27% and 24%, respectively. Moving on to gasoline, it comprises 130 samples representing a mere 0.30% of the total population. Interestingly, unknown vehicle types dominate this category with a proportion of 69%. Passenger four-door vehicles follow closely behind with a share of 21%. Regular gasoline emerges as the fuel type with an impressive count of 28,733 samples (equivalent to approximately 66% of the total population). Within this category unknown vehicle types take up the majority at around 53% while passenger four-door vehicles hold a share of 21%. In terms of diesel fuel consumption, there are a total of 1,196 samples (around 2.77% of the population). Again, unknown vehicle types dominate with a proportion reaching about 57% while passenger four-door vehicles rank second at 21%. Lastly, natural gas has a sample count with about 60 instances (which is equivalent, to roughly0.14% of the total population). In this case, the majority is represented by unknown vehicle types around 56%, followed by passenger four vehicles share of about 38%. Finally, out of the population, there are a total of 257 vehicles, which accounts for approximately 0.60%. Among these vehicles, hatchbacks make up the majority with a proportion of 41% followed by a category of vehicle types at 33%.

Associations

Table 3 explores the association of emissions categories with key vehicle characteristics, such as primary fuel type, vehicle type, and transmission type. CO2 emissions are converted into categorical data, highlighting a significant relationship between the dependent variable (CO2 emission levels) and independent variables (primary fuel type, vehicle type, and transmission type). Notably, p-values below .0001 underscore the statistical significance of these associations. A detailed breakdown of primary fuel type, vehicle type, and transmission type based on emission categories reveals intriguing patterns. For instance, unknown vehicles show high associations with polluter and gross polluter categories, while regular gasoline displays the strongest association with gross pollution. This breakdown provides valuable insights into the distribution of vehicle characteristics within different emission categories. Further examination of the distribution of variables within emission categories uncovers noteworthy insights. Premium gasoline vehicles tend to have fewer emissions below the standard but more standard and gross polluter vehicles. This pattern varies for other fuel types, showcasing different proportions for emission categories below and above the standard. These observations contribute to a comprehensive understanding of the relationships between fuel types and emission levels.

Exploring the percent frequency of vehicle and transmission types by primary fuel type sheds light on additional nuances. For instance, unknown vehicles exhibit a higher proportion of polluter and gross polluter categories, while hatchback vehicles excel in emission categories below the standard. The dominance of automatic transmission vehicles in the ultra-low emission category contrasts with manual transmission vehicles, which have a higher proportion in the low-emission category. The associations between primary fuel types, vehicle types, and transmission types with emission categories are further detailed. Regular gasoline is notably associated with gross pollution, while electric vehicles overwhelmingly contribute to the ultra-low emission category. Unknown vehicles tend to have average to high emission levels, while hatchback vehicles excel in low to ultra-low emission levels. Automatic vehicles consistently show higher CO_2 emissions proportionately compared to the population, with manual vehicles emitting less.

Correlation Coefficients

The correlation coefficient is a central focus in current research, drawing extensive attention from scholars. Early investigations utilized scalar products to establish variance and covariance, while later studies extended into probability spaces. Researchers have explored correlation coefficients by considering collections of fuzzy sets and developing novel methods tailored to specific types of sets. The evolving nature of these studies underscores the importance of understanding and quantifying relationships between variables, reflecting the ongoing interest and advancements in correlation coefficient research (Ganie, 2020).

In Table 4, the Pearson correlation coefficient is a measure of how strong and in which direction two variables are related linearly. The correlation coefficient is a value that ranges from -1 to +1 and does not have any units. A correlation coefficient nearing 1 implies a strong positive linear relationship whereas a value close to 1 suggests a strong negative linear relationship. With p-value <0.0001, we can realize that variable was statistically significant. The strongest positive relationship is between co2TailpipeGpm and barrels08 which is (0.9885). This means that as the amount of CO2 increases, so does fuel consumption measure in barrels. Also, cylinders and displ have a strong positive correlation which is (0.9046). It means as the number of engine cylinders increases, engine displacement in liters increases as well. On the contrary, the correlation between comb08 and co2TailpipeGpm is (-0.9184) revealing a strong negative correlation. This inverse shows that as combined MPG increases, the amount of CO2 decreases and vice versa. Moreover, there is a strong negative correlation between comb08 and barrels08 which is (-0.9050). This shows that as combined MPG increases, the amount of annual petroleum consumption in barrels decreases.

We can divide the weakest correlations into two different sections. The weakest correlations which are meaningful and the ones which are meaningless. For the first part, we can talk about volume (-0.4323) and vehtype (-0.3626) which show weak correlation with co2TailpipeGpm. Also, for the second part, we can see the weakest correlation between prifueltype and vehtype (-0.0340) which is meaningless. It makes sense since we are looking at two nominal variables.

Chi-Square Test

A chi-squared test of independence assesses whether there is a significant association or dependence between the variables. The analysis of the chi-squared table in the course data file involved investigating the association between emissions categories and vehicle types, focusing on tests of independence and homogeneity. The chi-squared test was deemed necessary due to lack of correlation, indicating a weak correlation between emissions category and vehicle types. Indeed, the chi-squared test is one of the most widely used statistical hypothesis tests. Among other objectives, it can be applied to test goodness of fit, homogeneity, independence, etc (Du, 2020). The null hypothesis of homogeneity posited that the proportion of emissions categories is the same for each population. In contrast, the alternative hypothesis suggested that emissions categories are not homogeneously related to vehicle types. The test involved creating contingency tables, expected values, and calculating chi-squared statistics. Results indicated a rejection of the null hypothesis because of low P-value (0.001), affirming that the distribution across variables is different. With a p-value below 0.05, the null hypothesis is rejected at any reasonable significance level. Thus, we reject the hypothesis of homogeneity in favor of the alternative hypothesis, indicating that the distribution of reasons for nonconformity differs among the emission categories. It was emphasized that large sample sizes can inflate chi-squared statistics, prompting the need for additional research to determine independence between emission categories and vehicle types. The significance of chi-squared values, degrees of freedom, and p-values were crucial in making conclusions about homogeneity and independence, with a strong indication that the null hypotheses were rejected.

Strengths and Weaknesses

The study exhibits several strengths that contribute to its credibility and reliability. Foremost among these strengths is the substantial sample size of 43,177, which enhances the study's representativeness of the population. This large dataset allows for a more accurate estimation of population parameters and a closer approximation to a normal distribution, making the results more robust and less sensitive to outliers. Additionally, the continuous updates to the dataset, such as the inclusion of MPG estimates for various model years and weekly updates to fuel prices, enhance the accuracy of the findings. However, the sheer size of the dataset introduces challenges, such as missing information, particularly in interior volume dimensions and cylinder values, and unrounded MPG values for certain vehicles.

Despite the strengths, the study is not without its weaknesses, which may impact the precision and accuracy of the results. The existence of 19,731 vehicles in an unknown category, the absence of interior volume dimensions for certain vehicle types, and the reliance on EPA emission factors for older model years pose limitations. The study acknowledges potential data limitation issues, emphasizing the importance of addressing missing values for a more precise final regression model. The implications for future research or application lie in refining the dataset by filling missing values, defining unknown vehicle categories, and incorporating categorical variables as dummy variables to enhance the overall accuracy of the model.

When assessing the strengths and weaknesses of the study, it becomes evident that thoughtful consideration is essential for guiding subsequent research and practical applications. The study's merits, exemplified by its expansive sample size and comprehensive dataset, facilitate a meticulous exploration of the variables impacting CO2 emissions. Nevertheless, it is imperative to acknowledge and rectify weaknesses such as presumptions related to normal distribution and potential confounding variables spanning an extended data collection period. Future research endeavors stand to gain from embracing larger and more representative sample sizes, especially considering the dynamic nature of the automotive industry. While a cautious approach is recommended in the interpretation of findings, the study nonetheless furnishes valuable insights into the determinants of greenhouse gas emissions from vehicles, thereby indicating pathways for further enhancement and optimization in subsequent research endeavors.

Multicollinearity

Multicollinearity occurs when predictor variables have correlations with each other and the response variable. Multicollinearity signifies the presence of linear dependence among regression variables, leading to challenges in obtaining individual parameter estimates and, consequently, a distorted relationship between explanatory variables and the response variable (Singh, 2023). In table 4 (correlation table) predictor variables that are strongly correlated specifically mentioning the correlation between emissions category with annual petroleum consumption and tailpipe CO2, tailpipe CO2 with annual petroleum consumption and combined MPG, engine displacement with engine cylinders, and combined MPG with annual petroleum consumption. It is important to note that strong correlations don't automatically imply an estimation issue but hint at the possibility of multicollinearity. Strong positive or negative correlations do not necessarily mean that there are multicollinearities, but also more investigation is needed to prove it. To get better results for multicollinearity, we need to remove the variables that are useless.

Reference

Du, L., & Hu, I. (2020). An empirical Bayes method for chi-Squared Data. *Journal of the American Statistical Association*, *117*(537), 334–347. https://doi.org/10.1080/01621459.2020.1777137

Ganie, A. H., Singh, S., & Bhatia, P. K. (2020). Some new correlation coefficients of picture fuzzy sets with applications. *Neural Computing and Applications*, *32*(16), 12609–12625. https://doi.org/10.1007/s00521-020-04715-y

Singh, P., Singh, S., & Paprzycki, M. (2023). Detection and elimination of multicollinearity in regression analysis. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, *27*(1), 105–111. https://doi.org/10.3233/kes-221622

Table 1

descriptive statistics associated with key characteristics.

	barrels08	co2TailpipeGpm	comb08	displ	volume
count	43177.000000	43177.000000	43177.000000	42919.000000	43177.000000
mean	17.153166	462.766843	20.845913	3.286738	66.928967
std	4.662888	124.772017	8.224150	1.357021	69.041682
min	0.060000	0.000000	7.000000	0.000000	0.000000
25%	14.330870	386.391304	17.000000	2.200000	0.000000
50%	16.480500	444.350000	20.000000	3.000000	86.000000
75%	19.388824	522.764706	23.000000	4.300000	113.000000
max	47.087143	1269.571429	141.000000	8.400000	538.000000

Table 2

Characteristics of sample vehicle model

TABLE 1. Characteristics of 43,166 Sample Vehicle Models by Primary Fuel Type

Variable	Population *N* (%) (N=43,177)	Premium Gasoline *n* (%) (n=12,801)	Midgrade Gasoline *n* (%) (n=130)	Regular Gasoline *n* (%) (n=28,733)	Diesel *n* (%) (n=1,196)	Natural Gas *n* (%) (n=60)	Electricity *n* (%) (n=257)	*p* value*
Vehicle Type								<.0001
Unknown (0)	19,730 #####	3,491 (27.3%)	90 (69.2%)	15,346 (53.4%)	685 (57.3%)	34 (56.7%)	84 (32.7%)	
Hatchback (1)	5,070 #####	1,313 (10.3%)	0 (0.0%)	3,535 (12.3%)	115 (9.6%)	2 (3.3%)	105 (40.9%)	
Passenger 2-Door (2)	6,394 #####	3,157 (24.7%)	12 (9.2%)	3,120 (10.9%)	103 (8.6%)	1 (1.7%)	1 (0.4%)	
Passenger 4-Door (3)	11,983 #####	4,840 (37.8%)	28 (21.5%)	6,732 (23.4%)	293 (24.5%)	23 (38.3%)	67 (26.1%)	
Transmission Type								<.0001
Automatic (1)	30,210 #####	9,411 (73.5%)	130 (100.0%)	19,588 (68.2%)	773 (64.6%)	60 (100.0%)	248 (100.0%)	
Manual (2)	12,956 #####	3,390 (26.5%)	0 (0.0%)	9,143 (31.8%)	423 (35.4%)	0 (0.0%)	0 (0.0%)	

* *p* values based on Pearson chi-square test of association.

Note: P-values based on Pearson chi-squared test of association

Table 3

Association of emission category by fuel type and other characteristics

TABLE 2: Association of Emissions Category by Fuel Type and Other Characteristics

Variable	Population *N* (%) (N=43,177)	Ultra-Low Emission *n* (%) (n=321)	Very-Low Emission *n* (%) (n=384)	Low Emission *n* (%) (n=5,556)	Standard *n* (%) (n=29,543)	Polluter *n* (%) (n=5,899)	Gross Polluter *n* (%) (n=1,474)	*p* value*
Primary Fuel Type								<.0001
Premium Gasoline (1)	12,801 ######	24 (7.5%)	70 (18.2%)	1,169 (21.0%)	9,798 (33.2%)	1,262 (21.4%)	478 (32.4%)	
Midgrade Gasoline (2	130 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	124 (0.4%)	6 (0.1%)	0 (0.0%)	
Regular Gasoline (3)	28,733 ######	40 (12.5%)	311 (81.0%)	4,066 (73.2%)	18,971 (64.2%)	4,358 (73.9%)	987 (67.0%)	
Diesel (4)	1,196 (2.8%)	0 (0.0%)	0 (0.0%)	303 (5.5%)	629 (2.1%)	259 (4.4%)	5 (0.3%)	
Natural Gas (5)	60 (0.1%)	0 (0.0%)	3 (0.8%)	18 (0.3%)	21 (0.1%)	14 (0.2%)	4 (0.3%)	
Electricity (6)	257 (0.6%)	257 (80.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Vehicle Type								<.0001
Unknown (0)	19,730 ######	91 (28.3%)	50 (13.0%)	739 (13.3%)	12,579 (42.6%)	5,119 (86.8%)	1,152 (78.2%)	
Hatchback (1)	5,070 ######	122 (38.0%)	128 (33.3%)	1,820 (32.8%)	2,952 (10.0%)	47 (0.8%)	1 (0.1%)	
Passenger 2-Door (2)	6,394 ######	7 (2.2%)	11 (2.9%)	703 (12.7%)	5,193 (17.6%)	339 (5.7%)	141 (9.6%)	
Passenger 4-Door (3)	11,983 ######	101 (31.5%)	195 (50.8%)	2,294 (41.3%)	8,819 (29.9%)	394 (6.7%)	180 (12.2%)	
Transmission Type								<.0001
Automatic (1)	30,210 ######	312 (100.0%)	301 (78.4%)	3,202 (57.6%)	20,730 (70.2%)	4,557 (77.3%)	1,108 (75.2%)	
Manual (2)	12,956 ######	0 (0.0%)	83 (21.6%)	2,354 (42.4%)	8,813 (29.8%)	1,341 (22.7%)	365 (24.8%)	

* *p* values based on Pearson chi-square test of association.

Note: P-values based on Pearson chi-squared test of association

Table 4

Correlation coefficients

TABLE 1. Pearson Correlation Coefficients (*N* =42,917)

	co2TailpipeGpm	barrels08	comb08	make_id	displ	cylinders	volume	vehtype	emissionscat	prifueltype
co2TailpipeGpm	1.0000	.9885	-.9184	-.2157	.7954	.7438	-.4323	-.3626	.8894	-.1128
barrels08	.9885	1.0000	-.9050	-.2117	.7843	.7337	-.4266	-.3580	.8791	-.1084
comb08	-.9184	-.9050	1.0000	.2072	-.7327	-.6863	.4161	.3313	-.8415	.1234
make_id	-.2157	-.2117	.2072	1.0000	-.2823	-.2670	.1165	.0940	-.1755	.0710
displ	.7954	.7843	-.7327	-.2823	1.0000	.9046	-.3628	-.2631	.6703	-.2149
cylinders	.7438	.7337	-.6863	-.2670	.9046	1.0000	-.2648	-.1524	.6185	-.2181
volume	-.4323	-.4266	.4161	.1165	-.3628	-.2648	1.0000	.7418	-.3627	.0498
vehtype	-.3626	-.3580	.3313	.0940	-.2631	-.1524	.7418	1.0000	-.3054	-.0340
emissionscat	.8894	.8791	-.8415	-.1755	.6703	.6185	-.3627	-.3054	1.0000	-.0874
prifueltype	-.1128	-.1084	.1234	.0710	-.2149	-.2181	.0498	-.0340	-.0874	1.0000

Note: All correlation values resulted in a *p* -value < .0001.

Note: All correlation values resulted in a P-values < 0.001

Table 5

Output of Chi-Squared, Two-Way Contingency, and Expected Value

Output of Chi-Squared Test - HINT: Convert to Contingency Table

		Hatchback	Passenger 2-Door	Passenger 4-Door	Unknown	Total
GROSS POLLUTER	Frequency	1.00	141.00	180.00	1,152.00	1,474.00
	Expected	173.08	218.28	409.08	673.55	-
	Chi-squared	171.08	27.36	128.28	339.86	666.58
	Percent	-	0.33	0.42	2.67	3.41
	Row Pct	0.07	9.57	12.21	78.15	-
	Col Pct	0.02	2.21	1.50	5.84	-
LOW EMISSION	Frequency	1,820.00	703.00	2,294.00	739.00	5,556.00
	Expected	652.41	822.78	1,542.00	2,538.80	-
	Chi-squared	2,089.61	17.44	366.77	1,275.95	3,749.77
	Percent	4.22	1.63	5.31	1.71	12.87
	Row Pct	32.76	12.65	41.29	13.30	-
	Col Pct	35.90	10.99	19.14	3.75	-
POLLUTER	Frequency	47.00	339.00	394.00	5,119.00	5,899.00
	Expected	692.68	873.57	1,637.20	2,695.60	-
	Chi-squared	601.87	327.12	943.98	2,178.73	4,051.70
	Percent	0.11	0.79	0.91	11.86	13.66
	Row Pct	0.80	5.75	6.68	86.78	-
	Col Pct	0.93	5.30	3.29	25.95	-
STANDARD	Frequency	2,952.00	5,193.00	8,819.00	12,579.00	29,543.00
	Expected	3,469.00	4,375.00	8,199.10	13,500.00	-
	Chi-squared	77.06	152.96	46.86	62.81	339.69
	Percent	6.84	12.03	20.43	29.13	68.42
	Row Pct	9.99	17.58	29.85	42.58	-
	Col Pct	58.22	81.22	73.60	63.76	-
ULTRA-LOW EMISSION	Frequency	122.00	7.00	101.00	91.00	321.00
	Expected	37.69	47.54	89.09	146.68	-
	Chi-squared	188.57	34.57	1.59	21.14	245.87
	Percent	0.28	0.02	0.23	0.21	0.74
	Row Pct	38.01	2.18	31.46	28.35	-
	Col Pct	2.41	0.11	0.84	0.46	-
VERY-LOW EMISSION	Frequency	128.00	11.00	195.00	50.00	384.00
	Expected	45.09	56.87	106.57	175.47	-
	Chi-squared	152.45	36.99	73.37	89.72	352.53
	Percent	0.30	0.03	0.45	0.12	0.89
	Row Pct	33.33	2.86	50.78	13.02	-
	Col Pct	2.52	0.17	1.63	0.25	-
TOTAL	Frequency	5,070.00	6,394.00	11,983.00	19,730.00	43,177.00
	Chi-squared	3,280.64	596.44	1,560.85	3,968.21	9,406.14
	Percent	11.74	14.81	27.75	45.70	100.00

Figure 1

Boxplot for Cylinder & Co2tailpipeGpm

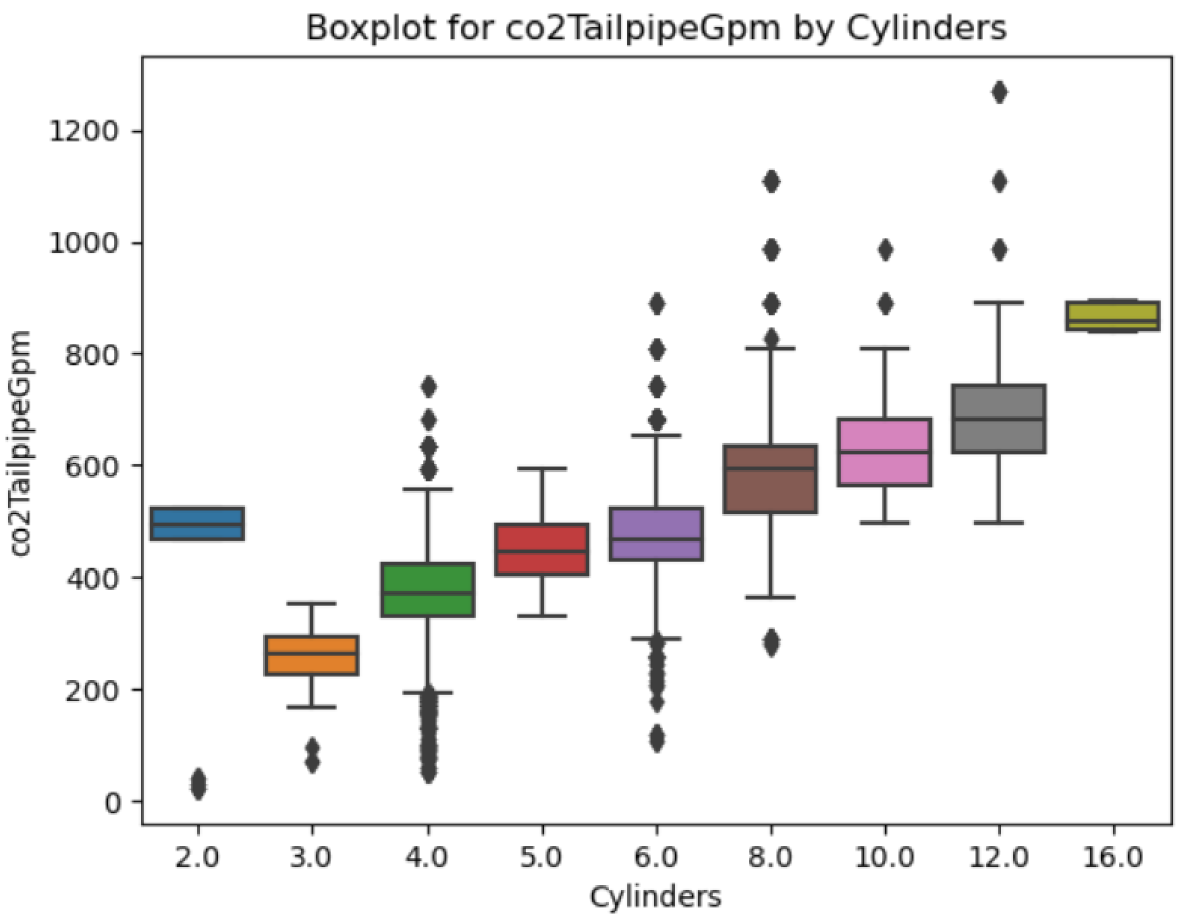

Figure 2

Probability of Density function carbon dioxide