3/3/2023

Exploring the Impact of Key Factors on the Sales Trend of New Vehicles in Australia from 1994 to 2017

BEO6000 - DATA ANALYSIS FOR BUSINESS (CT)
Lecturer - Salsabil Shaikh

Table of Contents

Executive summary

The aim of this research was to study the factors affecting sales of new vehicles in Australia. The analysis was per formed on data covering the period from 1994 to 2017, and included a comparison of sales between different vehicle types and states. The results showed that passenger vehicles had the highest median sales compared to other vehicle types, while Tasmania had the lowest sales and New South Wales had the highest among the states. The research also found that the total sales of new vehicles in Australia had increased gradually over the years, with the highest sales observed in the month of June. The results provide valuable insights for industry stakeholders and can inform future decisions about the automotive industry in Australia. The report concludes that passenger vehicles should be produced in higher quantities as sales are higher and more attention should be paid to the market in New South Wales, where sales are also higher. Additionally, production should be increased during the month of June when sales are higher.

Introduction

The automotive industry plays a significant role in the economies of many countries, and understanding the factors that affect sales of new vehicles is crucial for industry stakeholders (Chesterton, 2011). This research report aims to examine the factors affecting sales of new vehicles in Australia and identify the trend of sales growth from 1994 to 2017. The report will provide an in-depth analysis of the data to uncover the underlying drivers of the sales of new vehicles in Australia. By gaining a deeper understanding of these factors, the report will provide valuable insights for industry stakeholders and help inform decisions about the future of the automotive industry in Australia.

Research Objectives

The main aims of this research report are to understand the factors affecting the sales of new vehicles and to identify the trend of growth in sales from 1994 to 2017.

The main objective is portioned into subtopics as follows:

1. Does the type of vehicle affect the sales of new vehicles?
2. Does the state affect the sales of new vehicles?
3. Validating the impact of vehicle type and state on sales: A model to predict sales of new vehicles
4. Growth of sales of new vehicles from 1994 to 2017

Hypothesis statements for testing

In this report, we aim to address 3 main questions mentioned above. For each question, we formulate the null and alternative hypotheses.

- **Test 1: Does the vehicle type affect for the sales of new vehicles?**
  - H0: The median sales of Sports Utility Vehicles, Passenger Vehicles, and other vehicles are equal.
  - H1: At least one of the vehicle types, Sports Utility Vehicles, Passenger Vehicles, or other vehicles, has a different median sales value.

- **Test 2: Does the state affect the sales of new vehicles**
  - H0: The median sales of new vehicles in all states are equal.
  - H1: At least one state has a different median sale of new vehicles compared to the others.

- **Test 3: Validating impact of vehicle type and state for sales - A model to predict sales of new vehicles**
  - H0: The independent variables in the regression model do not have a significant effect on the sales of new vehicles.
  - H1: The independent variables in the regression model have a significant effect on the sales of new vehicles.

Methodology

Data Collection and Data Cleaning

Data was collected using the Australian Bureau of Statistics, https://www.abs.gov.au/, which is a website of Australia's national statistical agency. The dataset used for this analysis is included in this online platform named "Sales of New Motor Vehicles," which contains details on sales of new motor vehicles by state and vehicle type. The original dataset was arranged in a complicated way which may cause difficulties when analyzing. Therefore, the dataset was changed into an easily accessible format using Microsoft Excel. There were no missing values in the original dataset. After rearranging the dataset, there were five records in total and six variables. A preview of the arranged dataset can be seen in Figure 1.

Date	Sales	State	Type	Year	Month
Jan 1994	9925.0	NSW	Passenger vehicles	1994	Jan
Feb 1994	13552.0	NSW	Passenger vehicles	1994	Feb
Mar 1994	15385.0	NSW	Passenger vehicles	1994	Mar
Apr 1994	12265.0	NSW	Passenger vehicles	1994	Apr
May 1994	14440.0	NSW	Passenger vehicles	1994	May
Jun 1994	16570.0	NSW	Passenger vehicles	1994	Jun

Figure 1: Preview of the dataset

This dataset represents sales of new motor vehicles by vehicle type and state. Also, it has included the vehicle sold year and month. After importing the data, data were rearranged in the correct variable formats and extra column were created to get data for certain tests.

Statistical Analysis

For the statistical analysis, bar charts, histograms, and box plots were used in the descriptive statistics section. The data was checked on the normality assumption to see whether parametric tests could be applied. For that, skewness and kurtosis were used. To compare more than two group categories, the Kruskal-Wallis test (a non-parametric test) was used. The Kruskal-Wallis test is the non-parametric alternative to the ANOVA test. Finally, a regression model was used to build a model to predict sales of new motor vehicles. Statistical analysis was done using SPSS software.

Results and Interpretations

All the result outputs shown in this section were taken by using the SPSS software.

Descriptive statistics

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
Sales	5184	22612.0	31.0	22643.0	4198.304	60.1664	4331.9793	1.506	.034	1.505	.068
Valid N (listwise)	5184										

Table 1: Descriptive statistics of sales

The average number of sales was approximately 4198, and the minimum and maximum values were 31 and 22643. The range for sales variables is the difference between the maximum and minimum values for the sales data. In this case, the range is 22612, which means that the difference between the maximum statistic (22643) and the minimum statistic (31) is 22612. This information suggests that there is a significant spread in the sales data over the 1994–2017 period.

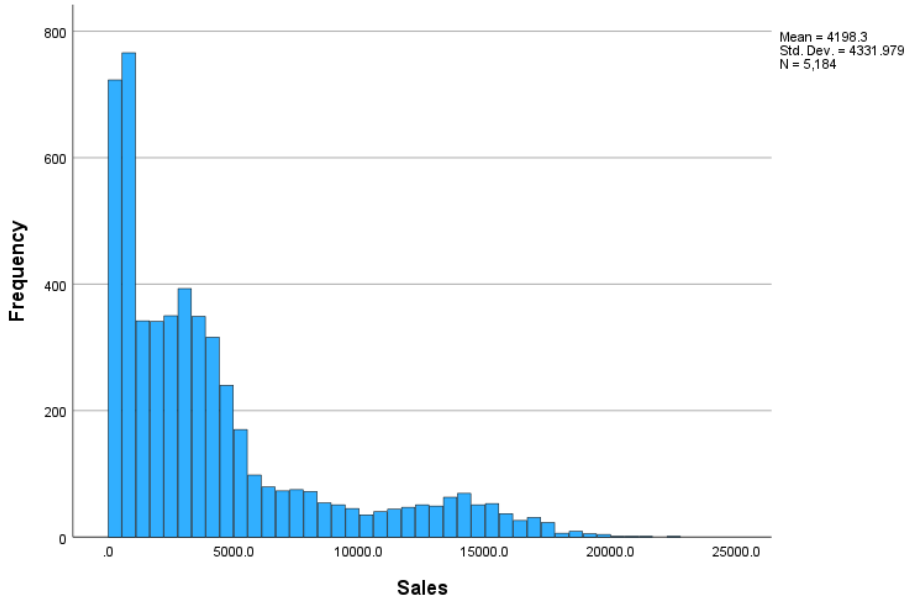
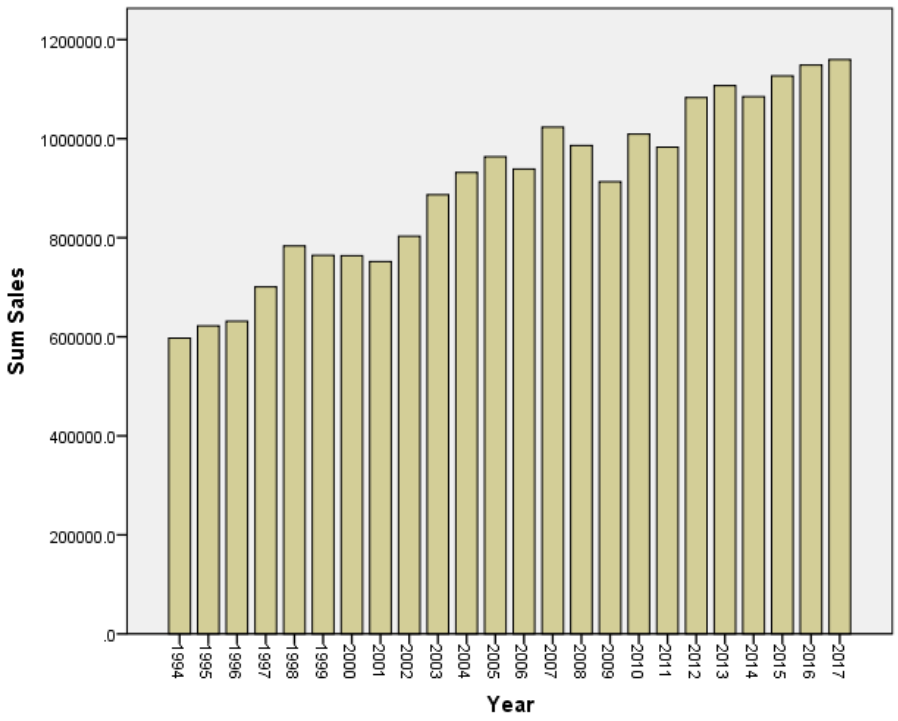


Figure 2: Histogram for sales

A positive skewness value of 1.506 means that the distribution is positively skewed, or skewed to the right. This means that the majority of the data is concentrated on the left side of the distribution, and there is a long tail to the right. A kurtosis value of 1.505 indicates that the distribution is slightly peaked compared to a normal distribution, which has a kurtosis of 0. A positive kurtosis value greater than 0 indicates a more peaked distribution, while a negative kurtosis value indicates a flatter distribution with most of the data concentrated in the lower range and a longer tail to the right. As verified by the histogram, the sales variable has a right-skewed distribution (positive skewness). The mean is typically better when the data follow a symmetric distribution. When the data are skewed, the median is more useful in statistics hence will be used in these tests because the mean will be distorted by outliers.

Growth of sales of new vehicles from 1994 to 2017.



According to Figure 3, the total sales of new vehicles have increased gradually from 1994 to 2017. But there are some points where it has decreased a little bit. When we take the era from 2010 to 2017, in 2011 and 2014, the sales of new vehicles decreased.

Figure 3: Sales by year

Impact of selling month for sales of new motor vehicles

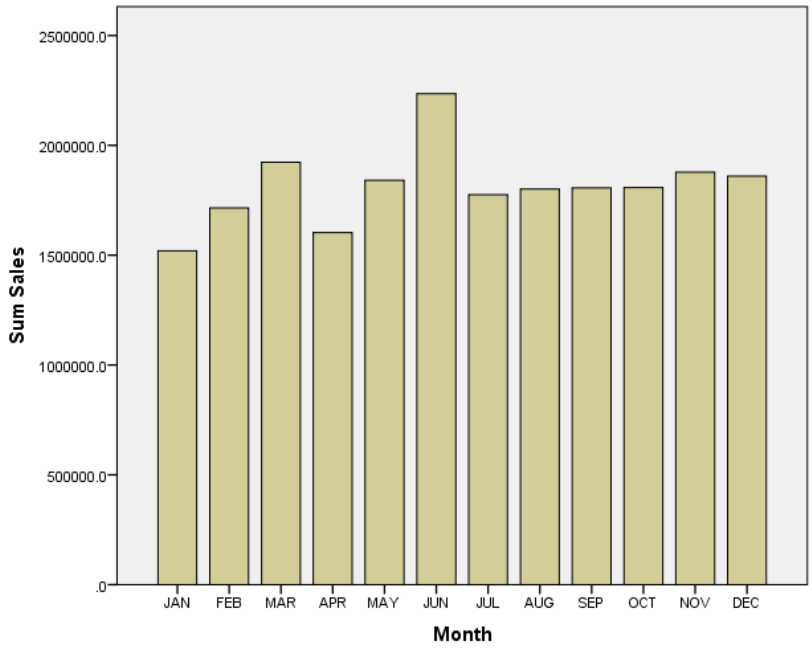


Figure 4 shows a bar plot for total sales by month. According that plot, the sales of June month is higher than the other months. January and April show a smaller number of sales compared to other months. July through December, sales do not show any considerable difference.

Figure 4: Total sales by month

Research problems

Test 1: How does the vehicle type affect for the sales of new vehicles?

Descriptive Analysis:

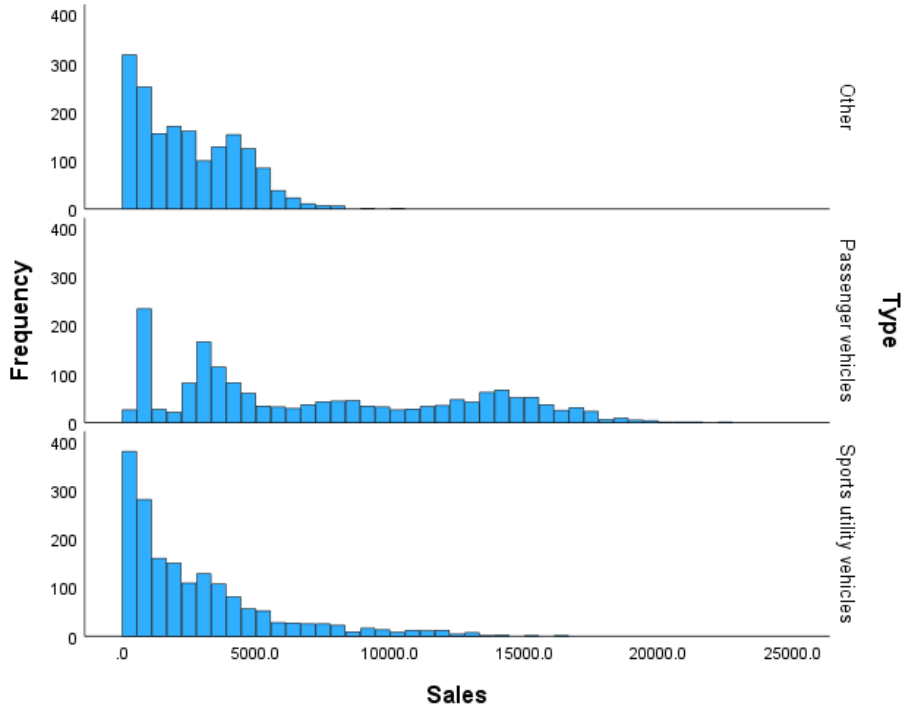
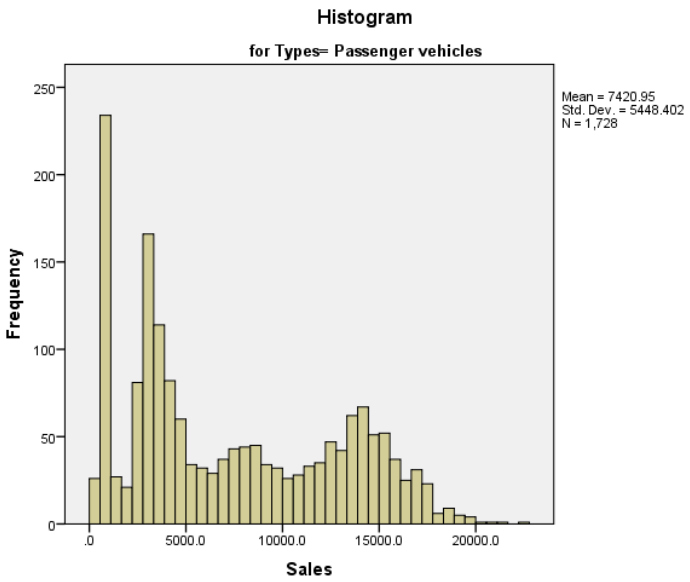
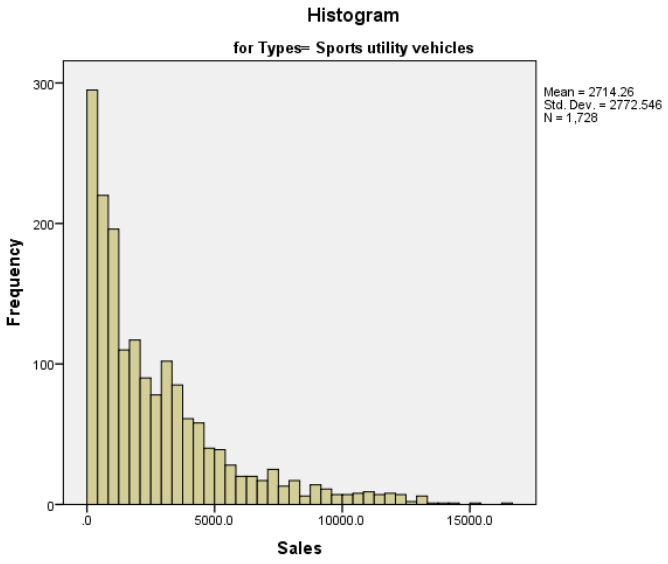


Figure 1: Histograms for vehicle types

Before analyzing the data, to get an idea about the three categories, histograms were plotted for all three categories separately.





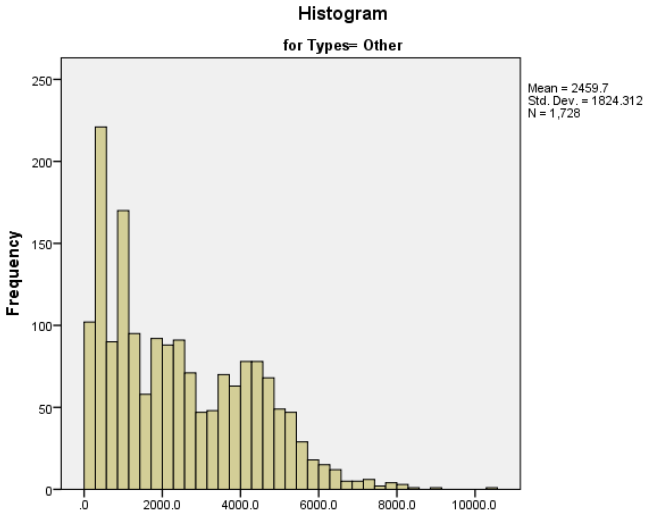


Figure 2: Histograms for vehicle types

Comparison of three vehicle types Before applying a test, it is easy to get a general idea of the sales of three vehicle types. For that, a box plot was created for the sales of all three categories.

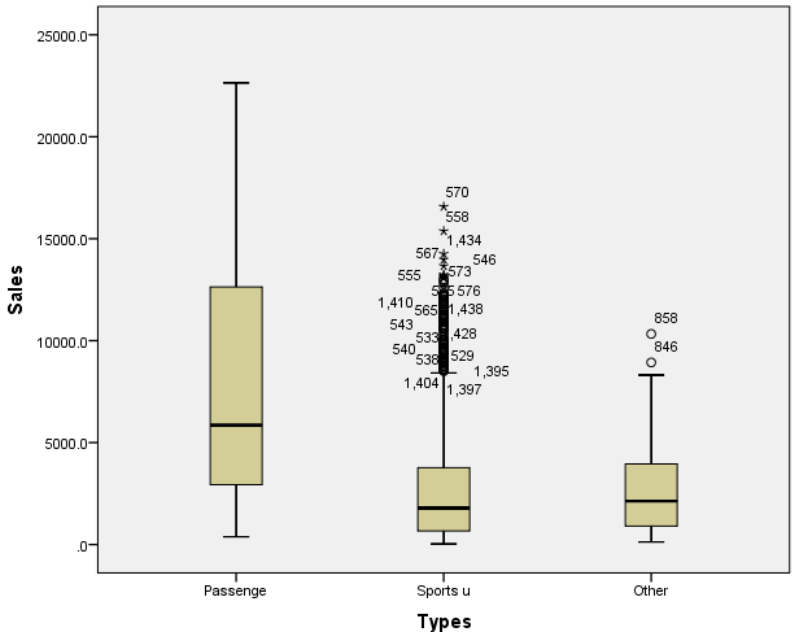


Figure 3: Boxplot for sales of vehicle types

According to Figure 3, the median sales of sports utility vehicles are the lowest and passenger vehicles are the highest. The median sales refer to the middle value in a set of data. If the median sales of Sports Utility Vehicles (SUVs) are the lowest and the median sales of Passenger Vehicles (PVs) are the highest, it means that the middle value of sales for SUVs is lower than the middle value of sales for PVs. This suggests that, on average, fewer SUVs are being sold compared to PVs. The median is a commonly used measure of central tendency and can be useful for getting an understanding of the "typical" or average sales in a data set when the data may have outliers or a skewed distribution. To clarify whether there is a significant difference among these three categories, we have to use a statistical test.

Checking the normality of data:

To determine the sort of test that must be used, it is necessary to examine a number of conditions. Parametric tests require normally distributed data. Skewness and kurtosis must be examined to confirm this.

Descriptives

	Type			Statistic	Std. Error
Sales	Other	Mean		2459.698	43.8861
		95% Confidence Interval for Mean	Lower Bound	2373.622	
			Upper Bound	2545.773	
		5% Trimmed Mean		2360.654	
		Median		2130.000	
		Variance		3328113.357	
		Std. Deviation		1824.3117	
		Minimum		124.0	
		Maximum		10330.0	
		Range		10206.0	
		Interquartile Range		3050.0	
		Skewness		.627	.059
		Kurtosis		-.391	.118
	Passenger vehicles	Mean		7420.954	131.0682
		95% Confidence Interval for Mean	Lower Bound	7163.885	
			Upper Bound	7678.023	
		5% Trimmed Mean		7227.222	
		Median		5854.500	
		Variance		29685086.147	
		Std. Deviation		5448.4022	
		Minimum		382.0	
		Maximum		22643.0	
		Range		22261.0	
		Interquartile Range		9699.5	
		Skewness		.421	.059
		Kurtosis		-1.166	.118
	Sports utility vehicles	Mean		2714.259	66.6971
		95% Confidence Interval for Mean	Lower Bound	2583.443	
			Upper Bound	2845.074	
		5% Trimmed Mean		2398.846	
		Median		1788.500	
		Variance		7687010.235	
		Std. Deviation		2772.5458	
		Minimum		31.0	
		Maximum		16570.0	
		Range		16539.0	
		Interquartile Range		3101.0	
		Skewness		1.656	.059
		Kurtosis		2.762	.118

Table 2: Descriptive statistics of sales by Vehicle type

Skewness statistics of all three categories are considerably higher than 0. Also, the kurtosis values are not closer to 0. Therefore, data is not normally distributed. Thus, a non-parametric test is required.

Applying the non-parametric test:

Because there are three categories, Kruskal-Walli's test can be used. The hypothesis conditions for this analysis are as follows:

- H0: The median sales of sports utility vehicles, passenger vehicles, and other vehicles are equal.
- H1: At least one of the vehicle types, such as sports utility vehicles, passenger vehicles, or other vehicles, has a different median sales value.

The Kruskal-Wallis results were as follows,

Test Statistics[a,b]

	Sales
Chi-Square	1042.372
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test
b. Grouping Variable: Type_code

Table 3: Results of Kruskal-wallis test

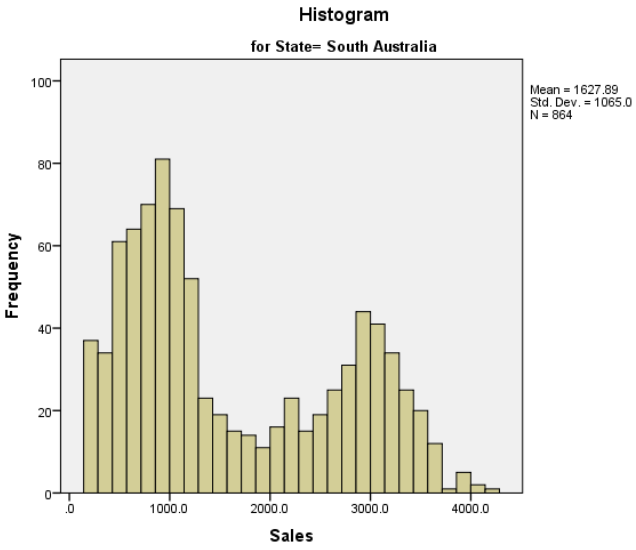
Statistical decision: According to the results, the p-value (0.000) is less than the significance level of 0.05 rejecting the null hypothesis that the median sales of sport utility vehicles, passenger vehicles, and other vehicles are equal.

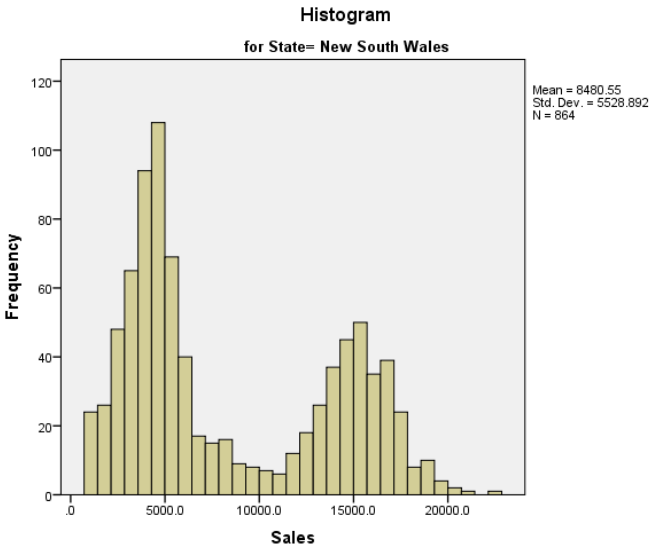
Conclusion: It is concluded that there are significant differences between the sales of vehicle types at 5% significance level.

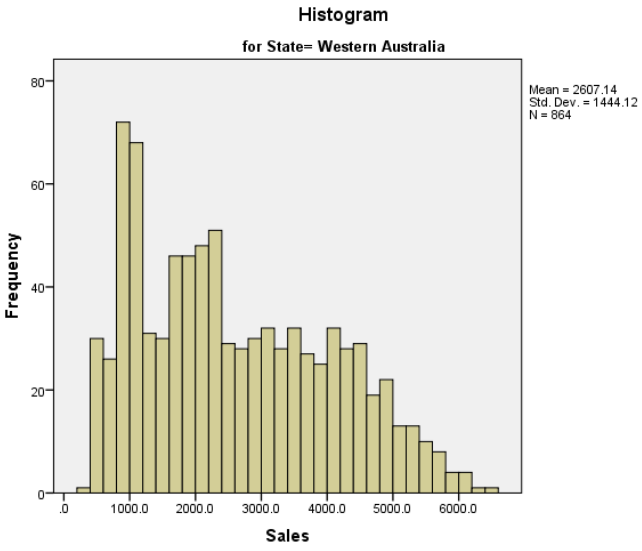
Test 2: Does the state affect the sales of new vehicles?

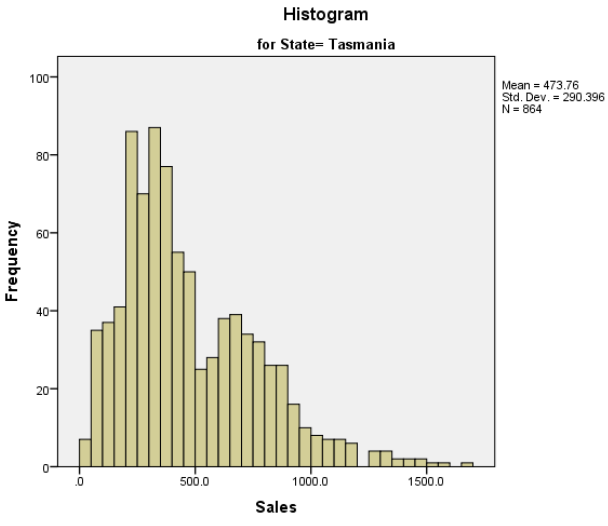
Descriptive Analysis:

To get an idea about the 6 states, histograms were plotted for all states separately before applying any tests.









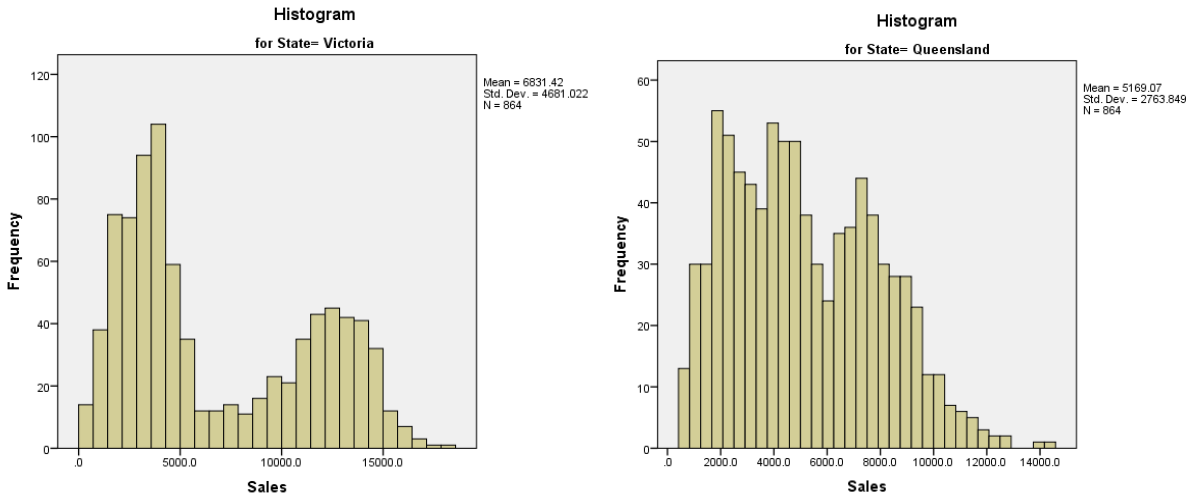


Figure 1: Histograms of sales by state

Comparison of states before applying a test is easy to get a general idea on the sales of each state. For that, a boxplot was created for the sales of new vehicles.

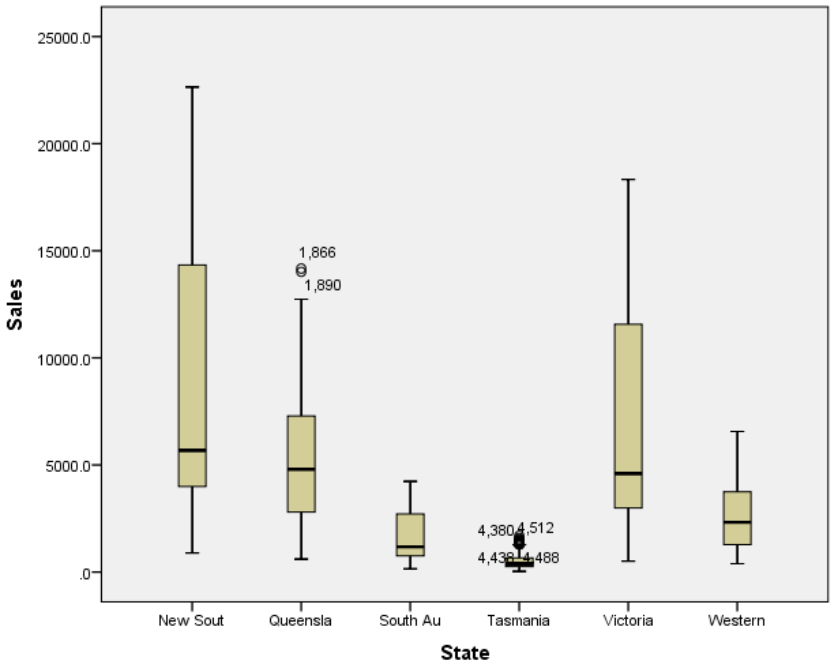


Figure 2: Boxplot for sales by states

According to Figure 6, the median sales in Tasmania are the lowest and those in New South Wales are the highest. South Australia likewise sells few new cars. A statistical test is needed to determine whether there is a significant difference between the six states.

Checking the normality of data:

Descriptives

	State			Statistic	Std. Error
Sales	New South Wales	Mean		8480.546	188.0967
		95% Confidence Interval for Mean	Lower Bound	8111.366	
			Upper Bound	8849.727	
		5% Trimmed Mean		8315.184	
		Median		5676.000	
		Variance		30568646.668	
		Std. Deviation		5528.8920	
		Minimum		888.0	
		Maximum		22643.0	
		Range		21755.0	
		Interquartile Range		10348.3	
		Skewness		.488	.083
		Kurtosis		-1.315	.166
	Queensland	Mean		5169.071	94.0281
		95% Confidence Interval for Mean	Lower Bound	4984.520	
			Upper Bound	5353.621	
		5% Trimmed Mean		5081.693	
		Median		4794.000	
		Variance		7638862.161	
		Std. Deviation		2763.8492	
		Minimum		607.0	
		Maximum		14170.0	
		Range		13563.0	
		Interquartile Range		4488.5	
		Skewness		.397	.083
		Kurtosis		-.629	.166
	South Australia	Mean		1627.888	36.2333
		95% Confidence Interval for Mean	Lower Bound	1556.772	
			Upper Bound	1699.003	
		5% Trimmed Mean		1594.519	
		Median		1178.500	
		Variance		1134305.245	
		Std. Deviation		1065.0377	
		Minimum		149.0	
		Maximum		4235.0	
		Range		4086.0	
		Interquartile Range		1960.3	
		Skewness		.498	.083
		Kurtosis		-1.189	.166
	Tasmania	Mean		473.758	9.8795
		95% Confidence Interval for Mean	Lower Bound	454.368	
			Upper Bound	493.149	
		5% Trimmed Mean		455.108	

	Median		392.000	
	Variance		84329.574	
	Std. Deviation		290.3955	
	Minimum		31.0	
	Maximum		1697.0	
	Range		1666.0	
	Interquartile Range		409.8	
	Skewness		.984	.083
	Kurtosis		.857	.166
Victoria	Mean		6831.418	159.2516
	95% Confidence Interval for Mean	Lower Bound	6518.852	
		Upper Bound	7143.984	
	5% Trimmed Mean		6691.307	
	Median		4599.500	
	Variance		21911970.721	
	Std. Deviation		4681.0224	
	Minimum		507.0	
	Maximum		18332.0	
	Range		17825.0	
	Interquartile Range		8581.0	
	Skewness		.480	.083
	Kurtosis		-1.287	.166
Western Australia	Mean		2607.141	49.1300
	95% Confidence Interval for Mean	Lower Bound	2510.713	
		Upper Bound	2703.569	
	5% Trimmed Mean		2555.658	
	Median		2325.500	
	Variance		2085481.982	
	Std. Deviation		1444.1198	
	Minimum		387.0	
	Maximum		6565.0	
	Range		6178.0	
	Interquartile Range		2472.8	
	Skewness		.435	.083
	Kurtosis		-.843	.166

Table 4: Descriptive statistics of sales by state

The results of the skewness statistics indicate that all categories have considerably higher values than 0. Additionally, the kurtosis values are not close to 0, indicating that the data is not normally distributed. It is important to note that a normal distribution cannot be assumed in this case. As a result, a non-parametric test must be used in the analysis. Given that there are six states in the data set, the appropriate test to use would be the Kruskal-Wallis test.

Applying the non-parametric test:

Because there are more than two categories, the Kruskal-Wallis test can be used. The hypothesis conditions for this analysis are as follows.

- H0: The median sales of new vehicles in all states are equal.
- H1: At least one state has a different median sale of new vehicles compared to the others.

After applying the Kruskal-Wallis test, the results obtained were as follows.

Test Statistics[a,b]

	Sales
Chi-Square	3239.024
df	5
Asymp. Sig.	.000

a. Kruskal Wallis Test
b. Grouping Variable: State

Table 5: Results of Kruskal-Wallis test

Statistical decision: According to the results, p-value (0.000) is less than the significance level 0.05 rejecting the null hypothesis that the median sales of new vehicles in all states are equal.

Conclusion: It is concluded that there are significant differences between the sales of states at 5% significance level. The results of the Kruskal-Wallis test indicate that there is a low probability (less than 0.05) that the difference in sales between states could have occurred by chance. Therefore, the conclusion is that there is evidence of a real difference in sales between the states. This information could be useful for business purposes, for example, to identify which states might be the most promising markets for future sales growth.

Test 3: A regression model to predict sales of new vehicles

To predict the sales of new vehicles, a regression model was built. The dependent variable was sales, and the independent variables were states and vehicle type. The results obtained are as follows:

- H0: The independent variables in the regression model do not have a significant effect on the sales of new vehicles.
- H1: The independent variables in the regression model have a significant effect on the sales of new vehicles.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.787[a]	.620	.620	2670.4443

a. Predictors: (Constant), State, Type

Table 6: Model summary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14784.355	124.125		119.108	.000
	Type	-2480.628	45.425	-.468	-54.609	.000
	State	-1607.084	21.717	-.634	-74.000	.000

a. Dependent Variable: Sales

Table 7: Coefficients of variables

Statistical Decision: The R-squared value for the model is 0.62 and the two independent variables were significant at 5% significance level. The coefficient for Vehicle type was -0.468 and the coefficient for state variable was -0.634. A p-value of 0.000 rejects the null hypothesis of the independent variables in the regression model do not have a significant effect on the sales of new vehicles.

Conclusion: In this case, the null hypothesis is rejected and the alternative hypothesis indicating that both variables are significant predictors of sales of new vehicles. Overall, these results suggest that the regression model is a good fit for the data and provides information on the relationship between the independent variables (state and vehicle type) and the dependent variable (sales of new vehicles).

Findings and discussion

Test 1

Kruskal-test Wally's can be used to find out if there are differences in median sales between different types of vehicles. Automotive companies may utilize this data to guide product development, pricing, marketing, and other strategic choices. PVs have greater median sales than other vehicle types, thus the businesses may produce and promote more PVs. The data may also be used to analyze market trends and anticipate sales. The study might help manufacturers, dealers, and marketers understand customer preferences and make educated production and marketing choices.

Test 2

The findings of the Kruskal-Wallis test offer organizations with vital information for identifying the states with the most potential for future sales growth. If one state has considerably greater median sales than the others, it could be beneficial to focus marketing efforts accordingly. New South Wales has the greatest median sales of new automobiles, while Tasmania has the lowest. South Australia sells less than other states. This data may help companies target their marketing and sales. For instance, they may aim to expand into New South Wales, where sales are strong, or boost sales in Tasmania and South Australia, where they are low. Additionally, it may be beneficial to conduct further analysis to understand the factors that contribute to the differences in sales between states, such as demographic differences, local economic conditions, or differences in consumer preferences.

Test 3

State and car type anticipated sales. With an R-squared value of 0.62, the model's independent variables explain 62% of the dependent variable's variation (sales). The model's independent variables' predictive power is measured by R-squared. A value close to 1 indicates a good fit of the model, while a value close to 0 indicates a poor fit. The results indicate that the two independent variables, state and vehicle type, were significant at a 5% significance level. This indicates a statistically significant relationship between these independent factors and the dependent variable. The coefficients for the two independent variables show their link to the dependent variable. A negative coefficient for vehicle type of -0.468 means that for each unit increase in vehicle type, the sales of new vehicles is expected to decrease by 0.468 units on average. A negative coefficient for the state variable of -0.634 means that for each unit rise in state, new car sales are predicted to fall by 0.634 units on average.

Recommendations

The following table was created to illustrate the recommendations for each research question.

Test	**Recommendation**
Test 1 (Kruskal-Wallis test for Vehicle Type)	• Focus on producing and promoting PVs based on higher median sales compared to other vehicle types. • Use the data to analyze market trends and make predictions about future sales. • Conduct further research to understand the factors that contribute to differences in sales between vehicle types.
Test 2 (Kruskal-Wallis test for States)	• Focus marketing and sales efforts in New South Wales where sales are high, and find ways to improve sales in Tasmania and South Australia where they are low. • Conduct further analysis to understand the factors that contribute to differences in sales between states.
Test 3 (Regression model for Sales prediction)	The independent variables, state and vehicle type, have a statistically significant relationship with the dependent variable (sales) and explain 62% of the variation in sales. Negative coefficients for the two independent variables indicate a decrease in sales for each unit increase in the independent variable.

Limitations of this study

1. Data limitations: The findings may not be typical of the total population due to the study's limited sample. Data utilized in the study may have been imprecisely obtained or prone to measurement error.

2. Model limitations: The study's regression model may not incorporate all variables affecting new car sales. The model may overlook consumer habits, market developments, and competition.

3. Significance vs practical importance: The findings may be statistically significant but not practical. The magnitude of the impacts may be little and negligible in terms of real sales.

4. Contextual factors: The research may neglect contextual elements including macroeconomic situations, technological improvements, and cultural variations that may affect new car sales.

5. Limitations of generalization: Due to varying market circumstances, consumer habits, and legislation, the study's findings may not applicable to other countries or regions.

Conclusion

This research discovered several significant conclusions. The purpose of this study was to identify the variables that influence new car sales. The SPSS results revealed a considerable difference in sales between the three different vehicle types, with sports utility vehicles selling for the least amount of money and passenger cars selling for the most. At a 5% significance level, Tasmania had the lowest sales and New South Wales the most. From 1994 through 2017, new car sales increased gradually, with occasional modest declines. June sales were greatest, with January and April sales lowest. By concentrating on generating more passenger cars, paying more attention to New South Wales, and producing more vehicles in June, these results may be utilized to boost sales in the future.

References

- ABS (2018). *Sales of New Motor Vehicles, December 2017 | Australian Bureau of Statistics*. [online] www.abs.gov.au. Available at: https://www.abs.gov.au/statistics/industry/tourism-and-transport/sales-new-motor-vehicles/latest-release.

- Australian car sales statistics 2019 (2019). *Australian car sales statistics 2019*. [online] Budget Direct. Available at: https://www.budgetdirect.com.au/car-insurance/research/australian-car-sales-statistics.html.

- Chesterton, A. (2011). *Australian car market: Car sales, statistics and figures*. [online] CarsGuide. Available at: https://www.carsguide.com.au/car-advice/australian-car-market-car-sales-statistics-and-figures-70982.

- Statistics, A.B. of (n.d.). Sales of New Motor Vehicles. *data.gov.au*. [online] Available at: https://data.gov.au/data/dataset/sales-of-new-motor-vehicles [Accessed 3 Feb. 2023].