Report on Sales Data Sample Analysis

Introduction:

In our sales data sample analysis, where we'll be examining various aspects of our sales transactions. Our dataset includes columns such as order number, quantity ordered, price each, order-line number, sales, order date, status, qtr_id, month_id, year_id, product line, msrp, product code, customer name, phone, addressline1, addressline2, city, state, postal code, country, territory, contact last-name, contact first-name, and deal size.

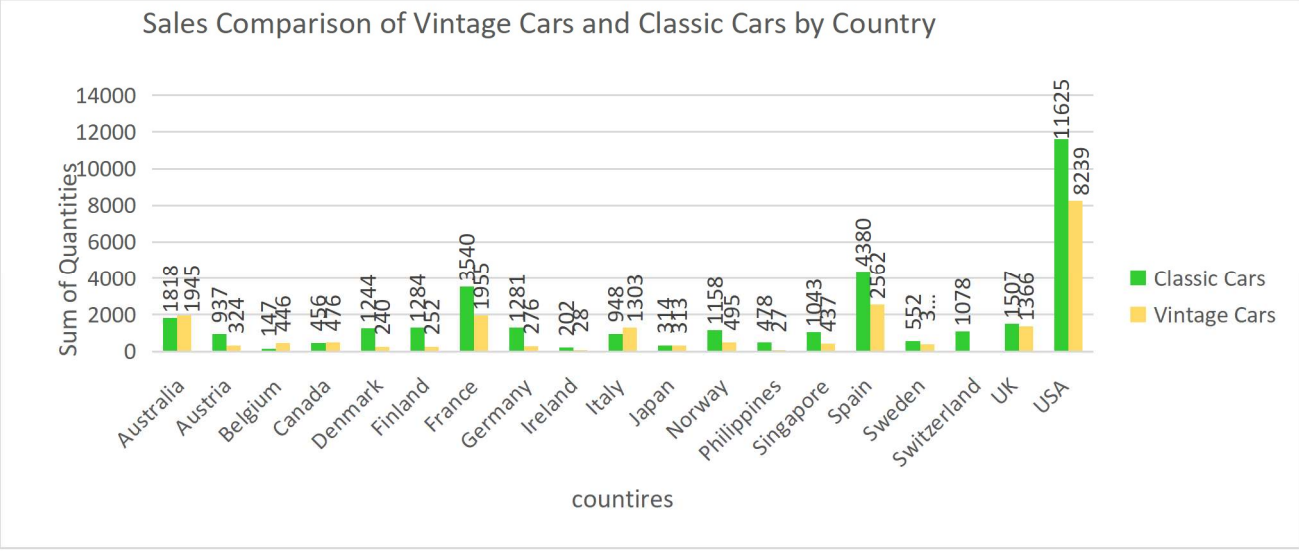
Each column holds valuable information about our sales operations, including details about the orders placed, quantities sold, pricing, product details, customer information, and sales performance. Through analysis of this data, we aim to gain insights into sales trends, customer preferences, and performance metrics to inform strategic decision-making and drive business growth.

Questionaries :

1. Compare the sale of Vintage cars and Classic cars for all the countries.
2. Find out average sales of all the products? which product yield most sale?
3. Which country yields most of the profit for Motorcycles, Trucks and buses?
4. Compare sales of all the items for the years of 2004, 2005.
5. Compare all the countries on the basis of deal size.

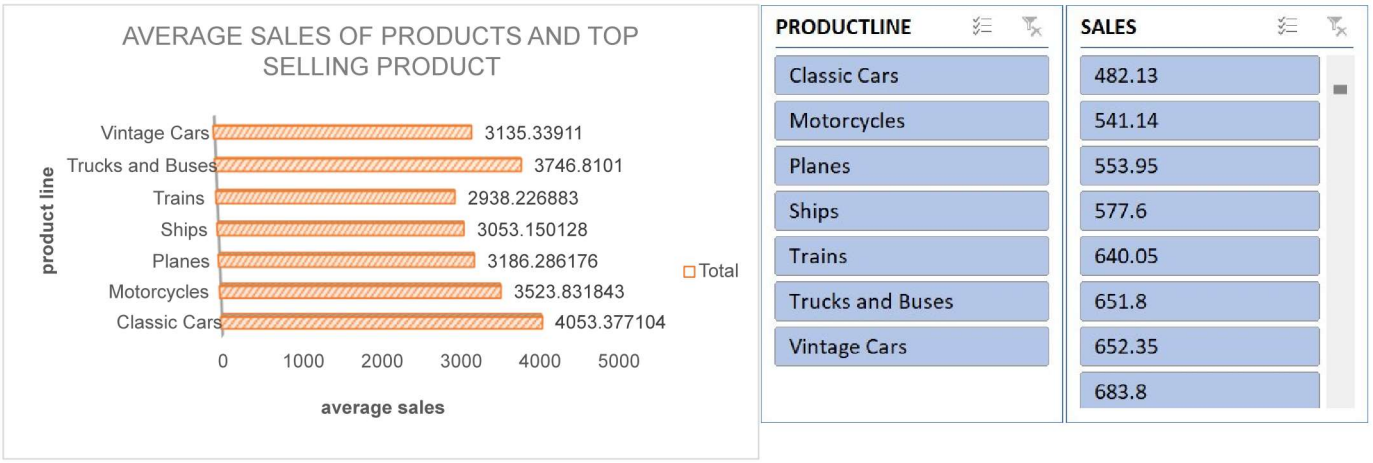
Analytics :

1. **Compare the sale of Vintage cars and Classic cars for all the countries**.



Ans: The United States is a big player in the vintage and classic car scene worldwide. They've got a huge variety of cars from different times, and they host fancy events like the Pebble Beach Concours elegance where people show off these beauties

2. Find out average sales of all the products? which product yield most sale?

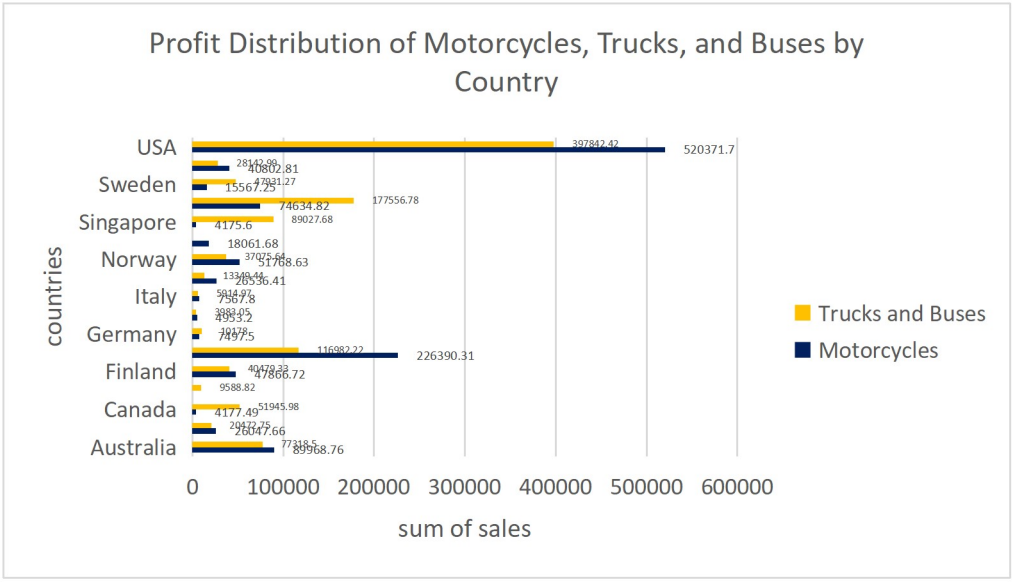


Ans: The Average Sales of all products are given below:-

i. Motorcycles :- 3523.831843
ii. Vintage Cars :- 3135.33911
iii. Classic Cars :- 4053.377104
iv. Truck and Buses :- 3746.8101
v. Trains :- 2938.226883
vi. Ships :- 3053.150128
vii. Plane :- 3186.286176

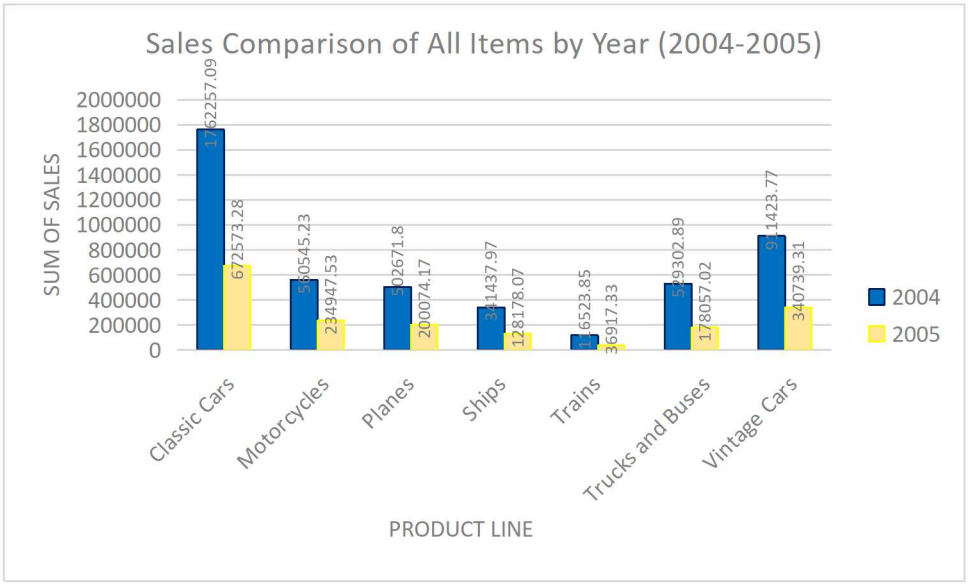
Trucks and buses are the heavy hitters in terms of sales worldwide. Their utility and versatility make them essential in various industries, from transportation and logistics to construction and public services.

3. Which country yields most of the profit for Motorcycles, Trucks and buses?



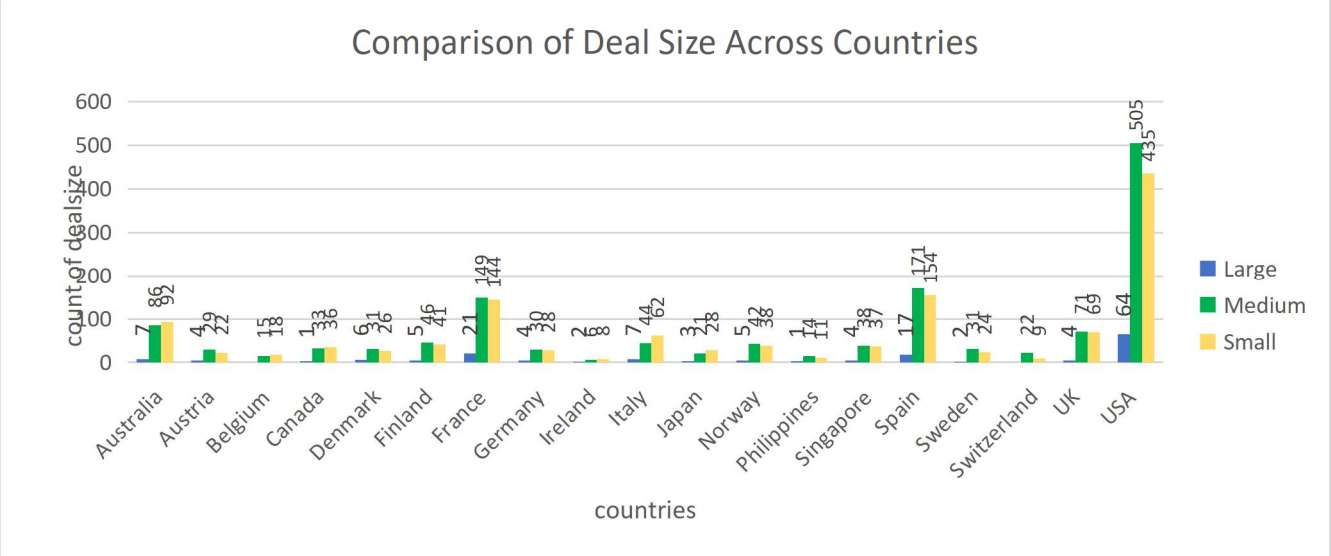
Ans: the USA is a big player in both motorcycles and trucks/buses. They rake in a hefty profit of $40,802.81 from motorcycles, showing how much Americans love hitting the open road on two wheels. And when it comes to trucks and buses, they're not far behind, making $28,142.99 in profit. Put together, that's a total profit of $68,945.8.

4. Compare sales of all the items for the years of 2004, 2005.?



Ans: During 2004 and 2005, classic cars were the top sellers in the automotive world. In 2004, there was a big craze for vintage cars, with lots of people excited about getting their hands on those iconic models from way back when. This excitement carried on into 2005, as classic cars remained super popular among folks who loved their timeless style and the nostalgic vibes they brought. So, for those two years, classic cars were definitely the stars of the show in the automotive market.

5. Compare all the countries on the basis of deal size?



Ans: When it comes to deal sizes, the USA leads the pack across the board. Whether it's small, medium, or large deals, the USA comes out on top. In small deals, they're making the most significant transactions, Moving up to medium deals, once again, the USA is leading the way, And when it comes to large deals, they're still the champions, sealing the biggest and most impactful agreements.

Conclusion and Review :

our analysis of the sales data sample has provided valuable insights into our sales operations, customer preferences, and performance metrics. While the report effectively outlined the dataset and objectives, it could benefit from more detailed analysis and visualizations to enhance understanding and presentation. Nevertheless, the insights gained will inform strategic decision-making and drive improvements in our sales processes. Continued analysis and refinement of our sales data will be crucial for achieving our business goals.

Regression:

Multiple R	0.551426
R Square	0.304071
Adjusted R Square	0.303824
Standard Error	8.127982
Observations	2823

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	81428.86	81428.86	1232.574	2.4E-224
Residual	2821	186366.8	66.0641		
Total	2822	267795.7			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	24.72811	0.332504	74.36941	0	24.07613	25.38008	24.07613	25.38008
SALES	0.002916	8.31E-05	35.10803	2.4E-224	0.002754	0.003079	0.002754	0.003079

The analysis shows that there's a strong connection between sales and the outcome we're looking at, with a p-value so low it's basically zero. This means the relationship is very likely real, not just due to chance. The model explains about 30.41% of what's going on, which is pretty good. And the standard error, which tells us how much our predictions might be off by, is around 8.128 units.

.Correlation:

	ORDERLINENUMBER	*SALES*
ORDERLINENUMBER	1	-0.0584
SALES	-0.0584	1

The correlation coefficient between ORDERLINENUMBER and SALES is -0.0584, which indicates a weak negative correlation between these two variables

Anova (single Factor) :

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
30	2822	99037	35.09461	94.92015
95.7	2822	236072.4	83.65428	407.0943
2	2822	18252	6.467753	17.85699
2871	2822	10029758	3554.131	3393504

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.61E+10	3	8.71E+09	10261.03	0	2.605696
Within Groups	9.57E+09	11284	848506			
Total	3.57E+10	11287				

The single-factor ANOVA analysis unveils significant variations among the groups, with a high F-value of 10261.03 and an ultra-low p-value close to zero, indicating a strong impact of the factor being analyzed. The degrees of freedom (df) for the between-groups factor are 3, representing the variability in means across the groups. Within the groups, the df is 11284, reflecting the variation within each group,and an error (standard error of the residuals) of approximately 848506.0368.

Anova without Replication :

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	3.24E+09	2822	1146956	1.021361	0.257009	1.054809
Columns	2.32E+10	2	1.16E+10	10320.57	0	2.997323
Error	6.34E+09	5644	1122968			
Total	3.28E+10	8468				

The ANOVA analysis reveals no significant differences in the means across rows, as indicated by the non-significant F-value of 1.021 ($p = 0.257$) and the degrees of freedom (df) of 2822. Similarly, for columns, a highly significant difference is observed among the means, with an F-value of 10320.57 ($p < 0.001$) and df of 2. The error term, representing variability within groups, has an MS of approximately 1122968.007, reflecting the average amount of variation not explained by the model

Descriptive Statistics :

Mean	35.09281	Mean	83.65854	Mean	6.466171	Mean	
Standard Error	0.183344	Standard Error	0.379702	Standard Error	0.079535	Standard Error	
Median	35	Median	95.7	Median	6	Median	
Mode	34	Mode	100	Mode	1	Mode	
Standard Deviation	9.741443	Standard Deviation	20.17428	Standard Deviation	4.225841	Standard Deviation	
Sample Variance	94.89571	Sample Variance	407.0014	Sample Variance	17.85773	Sample Variance	
Kurtosis	0.415744	Kurtosis	-0.37482	Kurtosis	-0.56115	Kurtosis	
Skewness	0.362585	Skewness	-0.94665	Skewness	0.590741	Skewness	
Range	91	Range	73.12	Range	17	Range	
Minimum	6	Minimum	26.88	Minimum	1	Minimum	
Maximum	97	Maximum	100	Maximum	18	Maximum	
Sum	99067	Sum	236168.1	Sum	18254	Sum	
Count	2823	Count	2823	Count	2823	Count	
Largest(1)	97	Largest(1)	100	Largest(1)	18	Largest(1)	
Smallest(1)	6	Smallest(1)	26.88	Smallest(1)	1	Smallest(1)	
Confidence Level(95.0%)	0.359503	Confidence Level(95.0%)	0.744521	Confidence Level(95.0%)	0.155952	Confidence Level(95.0%)	

The descriptive statistics provide valuable insights into the dataset's variables. Notably, QUANTITYORDERED has a mean of 35.09 and a standard deviation of 9.74, indicating moderate variability. PRICEEACH shows wider variability, with a mean of 83.66 and a standard deviation of 20.17. ORDERLINENUMBER and SALES follow similar patterns, with means of 6.47 and 3553.89, and standard deviations of 4.23 and 1841.87, respectively. The largest QUANTITYORDERED is 97, while the smallest is 6. Similarly, the largest PRICEEACH is 100, and the smallest is 26.88. For ORDERLINENUMBER, the largest value is 18, and the smallest is 1. Finally, the largest SALES amount to 14082.8, while the smallest is 482.13. These statistics shed light on the central tendencies and variability within the dataset, providing crucial insights for analysis.