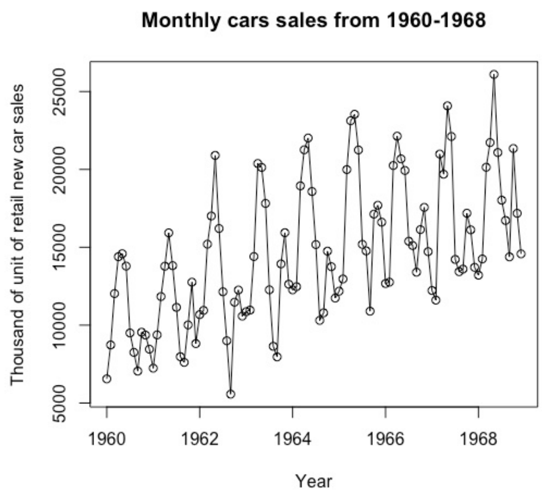
CAR SALES PREDICTION FOR NEXT 10 YEARS PREDICTING SALES IN QUEBEC CANADA

All those cars were once just a dream in Somebody's head– (Peter Gabriel)

This report outlines the methodology, various statistical analysis, concerns, recommendations, forecast of car sales.

The data of 108 observations between January 1960 to December 1968 was examined with variables like Year, Month and Sales.

Exhibit1

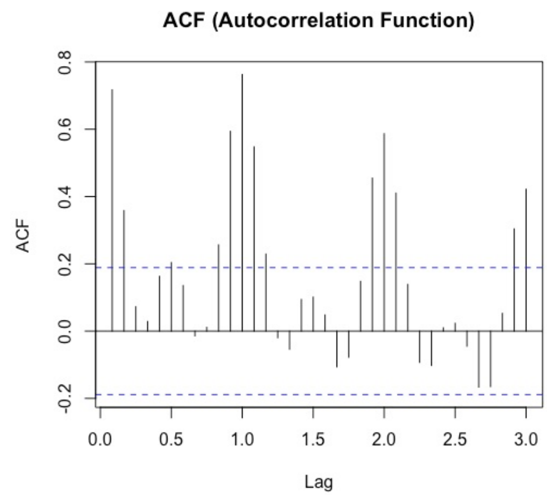


Volatility in the car industry is what makes TIME the most important factor.

About the Exhibit 1:

The Canadian province of Quebec's car sales are detailed in the data collection Monthly Car Sales. There are 108 observations overall, and the units represent counts of the number of sales. Given that Canada's prime winter months are November through February, and its summer season begins in April or May, it stands to reason that most automobile purchasers prefer to acquire their vehicles during the warmer months because the cold weather and accompanying snow make it difficult for people to get outside and enjoy themselves.

ACF plot:



ACF plots provide insights into the correlation between different elements within a time series. Lags in the ACF plot represent the correlation between current observations and their corresponding previous time stamps. This analysis helps to identify and understand patterns and relationships within the data.

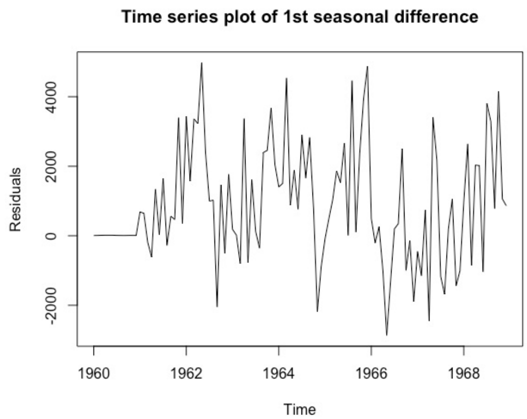
The presence of strong correlation at lags 12, 24, 36, and so on suggests the existence of a seasonal auto-correlation relationship within the time series. This indicates that there are recurring patterns or cycles that occur at regular intervals, likely corresponding to seasonal variations.

Automotive companies are transforming old beat-up cars into lucrative assets through modern-day alchemy.

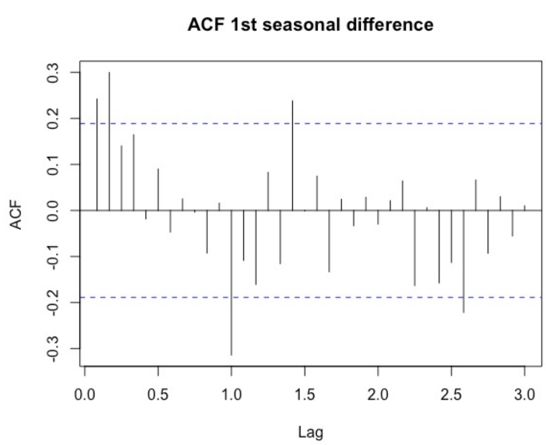
A Seasonal ARIMA model uses data values and errors to predict X_t at times with lags that are multiples of S (i.e., the span of seasonality).

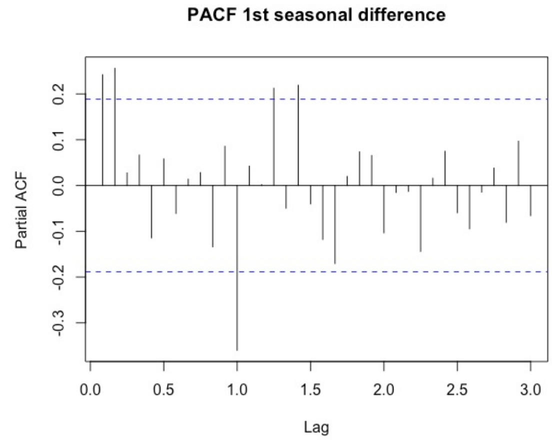
Residual Approach When there is seasonality, it is vital to investigate the **differences** in the data. Because the average value at some points during the seasonal period may differ from the average values at other points, seasonality typically results in series that are non-stationary.

Seasonal ARIMA differencing:
Both non-seasonal and seasonal components are included in the multiplicative model of the seasonal ARIMA model.



- Firstly, we do seasonal differencing to get rid of seasonal trends and fitting a plain model until time series and ACF/PACF plots of residuals show no sign of seasonality.
- Then we determine the orders of P & Q of the seasonal part based on final ACF/PACF plots of residuals.
- We do so by fitting the ARIMA(0,0,0)x(0,1,0) model.



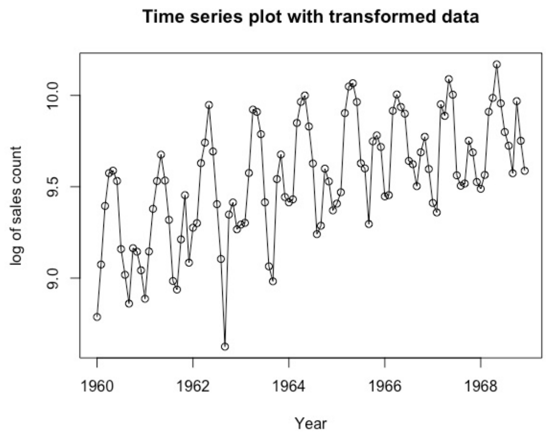


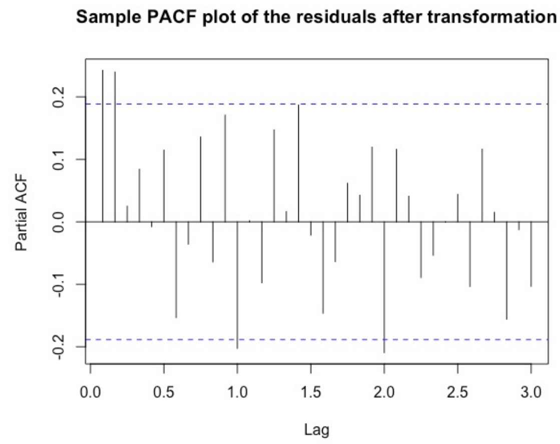
The ACF/PACF plots show seasonal trends are filtered out.

- ACF & PACF shows 1 seasonal lag, indicates SARMA (1,1)

Transformation:

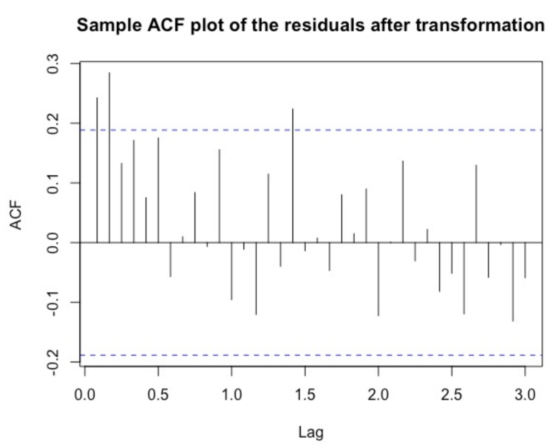
Data transformation is an important tool for proper statistical analysis.
After applying a log transformation to the time series data and plotting it on a graph, we observe a consistent upward trend. However, there is an intervention point in the plot that stands out from the rest of the time series. Aside from this point, the overall pattern of the time series remains relatively unchanged, indicating a sustained upward trend.

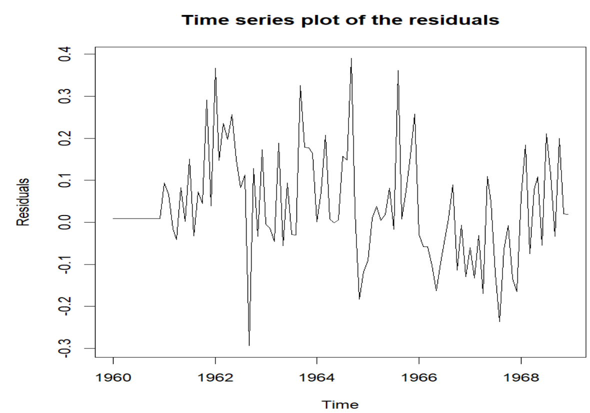




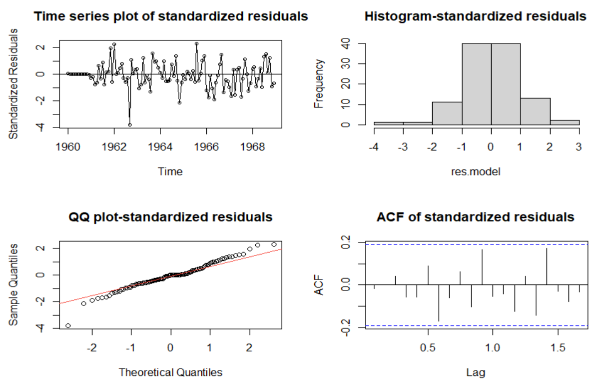
After fitting the model with the log-transformed data, we analyzed the sample ACF and PACF plots to draw conclusions. The ACF plot indicates that there are two significant lags before the first seasonal lag. This suggests a potential non-seasonal correlation between the current observation and its immediately preceding two observations.

On the other hand, the PACF plot shows two significant lags. This suggests that there is a direct relationship between the current observation and its two previous observations after removing the effect of intervening observations.

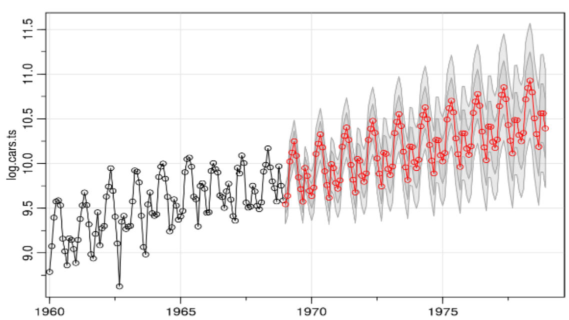




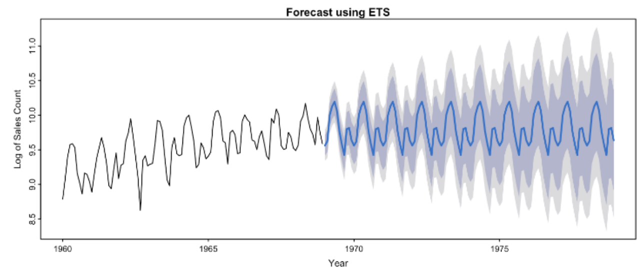
This is the graph of residuals derived from fitting an ARIMA model to the log-transformed car sales data. Residuals represent the differences between observed and predicted values. The plot shows the spread and patterns of residuals over time, allowing assessment of model adequacy and the need for further refinement.



The graph shows if standardization is required and proceeds to create a 3x2 grid for plots. The plots include a time series plot showing the fluctuation of standardized residuals over time, a histogram to assess their distribution, and a QQ plot for comparison with a standard normal distribution. Additionally, an autocorrelation function (ACF) plot detects potential autocorrelation, while the Shapiro-Wilk test evaluates normality. The function concludes by assigning a value of zero to the variable "k." Overall, this comprehensive analysis aids in assessing assumptions, distribution, autocorrelation, and normality of the residuals obtained from the model.



The increasing demand for cars aligns with predictions due to factors such as technological advancements, population growth, and other driving forces in the automobile industry. However, it's important to note that the growth in car sales may not follow a straight or steady path. Economic conditions, consumer choices, government regulations, and external factors can impact the growth rate and direction of car sales over time. Therefore, it is essential to regularly review and update forecasts as new information emerges.



We conducted an analysis using both the seasonal ARIMA and ETS models to study time series data. Our methodology involved specifying the models, evaluating residuals, fitting the models, and performing diagnostic checks. When we utilized our model to forecast the next 10 years, we determined that the SARIMA (0,1,3)x(1,1,2) model provided the most accurate results.

Conclusion:

In conclusion, this executive summary provides an overview of a report that focuses on the methodology, statistical analysis, concerns, recommendations, and forecast of car sales. The report examines a dataset comprising 108 observations from January 1960 to December 1968, featuring variables such as Year, Month, and Sales.

Exhibit 1 highlights the importance of time in the volatile car industry. The data reveals a significant and consistent upward trend, indicating potential growth or improvement in the underlying phenomenon. A clear seasonal pattern is observed, with regular intervals of rise and fall in values.

The ACF plot analyzes the correlation between different elements in the time series and identifies seasonal autocorrelation at lags 12, 24, 36, and so on. This suggests recurring patterns or cycles corresponding to seasonal variations.

The report discusses the transformation of data using a log transformation, revealing a sustained upward trend with an intervention point. The ACF and PACF plots indicate non-seasonal correlations and relationships between observations.

Furthermore, the report includes a graph of residuals derived from fitting an ARIMA model to the log-transformed car sales data. The graph, generated by a comprehensive analysis function, assesses assumptions, distribution, autocorrelation, and normality of the residuals.

Overall, this report provides valuable insights into the car sales industry, utilizing statistical analysis techniques to understand trends, patterns, and relationships within the data. The findings offer recommendations and forecasts to aid decision-making in this dynamic sector.

References:

Automotive News - http://www.autonews.com/

International Journal of Forecasting - http://www.journals.elsevier.com/international-journal-of-forecasting

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

"Forecasting: Principles and Practice" by Rob J. Hyndman and George Athanasopoulos https://otexts.com/fpp2/arima.html

"Arima Modeling in R" by DataCamp Website: https://www.datacamp.com/community/tutorials/arima-models-in-r

MADE BY GROUP 6:

Mohil Jain
Taha Talwala
Vedant Chhibber
Xinyu Li