TRUCK SALES TIME SERIES ANALYSIS AND PREDICTION

Author: Nguyen Ngoc Tue Minh

2024

Table of Contents

Table of Figures

Introduction and project purposes

This project use time series analysis techniques to analyse monthly sales of a truck company over a period of 12 years and build time series models to predict sales for the last 12 months.

Dataset overview

[This dataset](#) from Kaggle has 144 rows containing information on monthly sales from 2003 to 2014. Sales range went from 152 to 958 units. 50% of monthly sales stayed between 273 and 560. There is no missing values and outliers in this dataset.

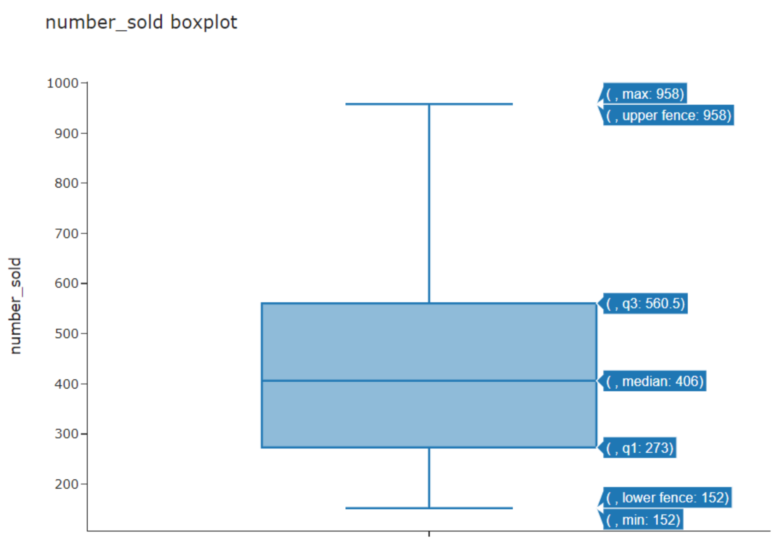

Figure 1

Time series analysis

The line chart illustrates a yearly pattern. This pattern shows the sales gradually climbing and peaking toward the end of the year but ending with a sudden reduction. This peak occurs in the months of July and August.

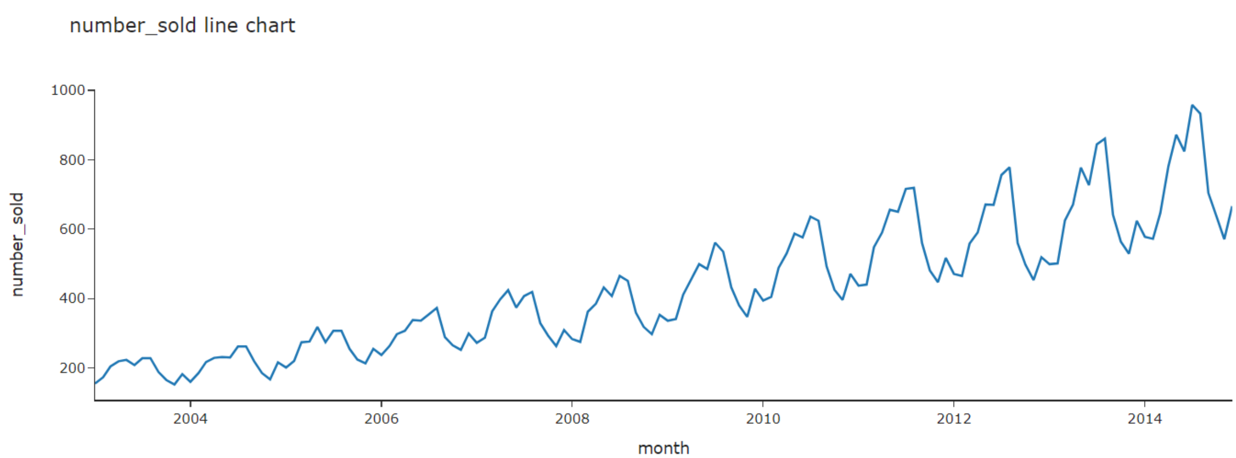

Figure 2

Identifying patterns and decomposition

The time series is multiplicative because the magnitude of the seasonality component is dependent of the trend (Sigmundo Preissler Jr, 2018). The line plot above shows there is a 12-month-seasonality. Decomposition as below:

- The original data reflects the combined effects of trend, seasonality, and noise.
- The trend indicates a consistent long-term increase.
- Seasonality shows regular peaks and drops every year.
- Residuals mostly shows small random changes, but sometimes there are bigger jumps or drops that aren't explained by the model

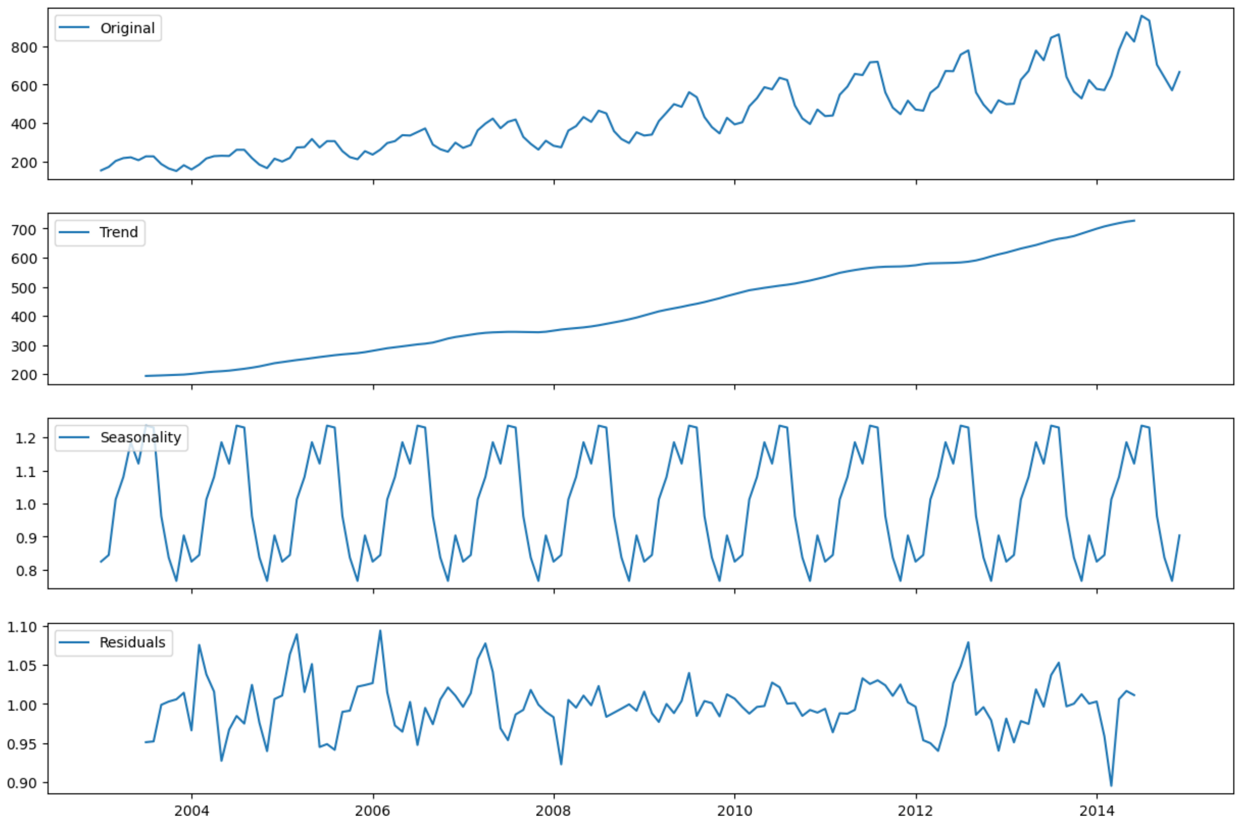

Figure 3

Stationarity check and preliminary visualizations

The ADF test on the original series shows a high p-value (0.995), indicating non-stationarity. Therefore, we need to differencing to make it stationary.

```
ADF Test Statistic: 1.1158932574252591
p-value: 0.9953500083802601
Number of Lags Used: 14
Number of Observations Used: 129
crit values:  {'1%': -3.482087964046026, '5%': -2.8842185101614626, '10%': -2.578864381347275}
Weak evidence against null hypothesis, time series has a unit root, indicating it is non-stationary
```

Figure 4: ADF test on original data

Order of non-seasonal differencing

ACF and PACF plot show a trend and uncaptured seasonality after the first difference. Second differencing shows signs of over-differenced by many negative lags.

ADF test for first differencing doesn't fully stabilize the series (p-value > 0.05) but second differencing achieved stationary (p-value < 0.05), but possible over-differencing because of many negative spikes.

Given these observations and the clear seasonality in the data, we tried seasonal differencing to avoid too much non-seasonal differencing.

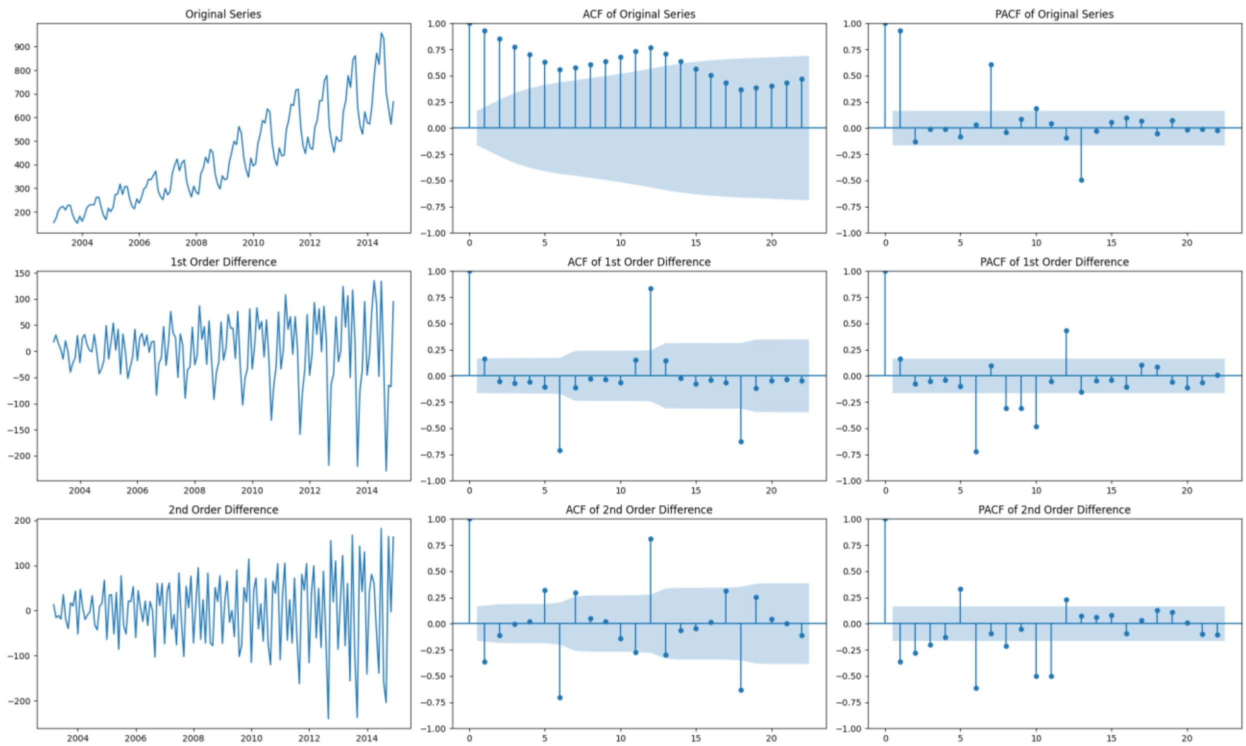

Figure 5

```
ADF Test Statistic: -2.541355326769916
p-value: 0.10573354923819672
Number of Lags Used: 14
Number of Observations Used: 128
crit values:  {'1%': -3.4825006939887997, '5%': -2.884397984161377, '10%': -2.578960197753906}
Weak evidence against null hypothesis, time series has a unit root, indicating it is non-stationary
```

Figure 6: ADF test on 1st non-seasonal difference

```
ADF Test Statistic: -8.64960599310258
p-value: 5.158038063798361e-14
Number of Lags Used: 13
Number of Observations Used: 128
crit values:  {'1%': -3.4825006939887997, '5%': -2.884397984161377, '10%': -2.578960197753906}
Strong evidence against the null hypothesis, reject the null hypothesis. Data has no unit root and is stationary
```

Figure 7: ADF test on 2nd non-seasonal difference

Order of seasonal differencing

The 1st order seasonal difference removes some of the trend and seasonality but still shows short-term correlation. Result from ADF test shows stationarity.

The 2nd order seasonal difference may be too much, as indicated by the ACF and PACF patterns, which no longer show a clear seasonal pattern and suggest possible over-differencing.

Based on these observations, we modelled the dataset with a 1st order seasonal difference.

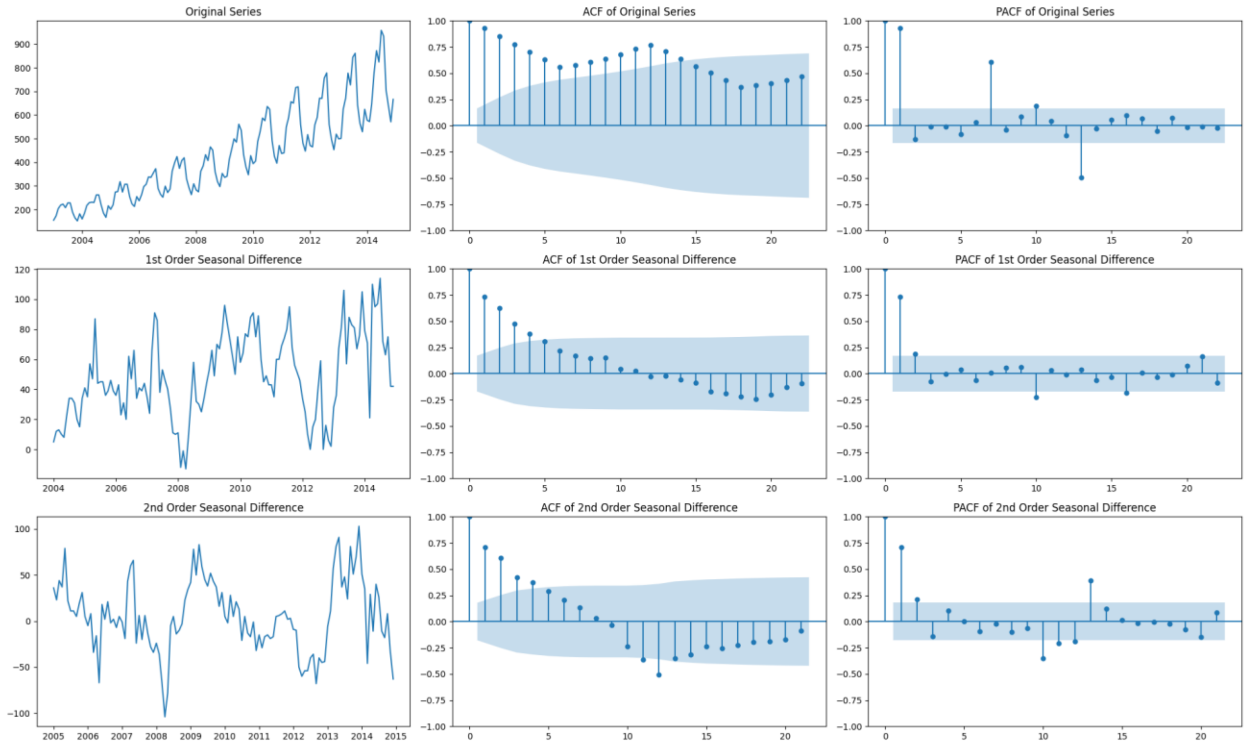

Figure 8

```
ADF Test Statistic: -3.5079300409395344
p-value: 0.007785754185827427
Number of Lags Used: 1
Number of Observations Used: 130
crit values:  {'1%': -3.4816817173418295, '5%': -2.8840418343195267, '10%': -2.578770059171598}
Strong evidence against the null hypothesis, reject the null hypothesis. Data has no unit root and is stationary
```

Figure 9: ADF test on 1st seasonal difference

Data modelling

We created five models (manual_ARIMA, manual_SARIMA, Facebook Prophet, auto_ARIMA, auto_SARIMA) to forecast sales for the last 12 months, following these steps for each model:

- Analyzed the train dataset with ACF and PACF plots.
- Conducted ADF and KPSS tests for stationarity.
- Applied logarithmic transformation if initial differences didn't stabilize the time series.
- Identified the appropriate differencing and model terms (AR, MA) using ACF and PACF.
- Fitted the model to the training set and forecasted the test dataset of 12 months.

- Calculated accuracy metrics (MAE, MAPE, MSE, R2) and plotted residual diagnostics.

We compared their performance and selected SARIMA(0,1,1)(0,1,1,12) as the top performer. Then we evaluated this model via time series cross-validation to assess its reliability over time.

Due to the assignment's word-limitation, we only presented the result of the best model.

```
                                      SARIMAX Results                                       
==========================================================================================
Dep. Variable:                    number_sold_log   No. Observations:                  132
Model:             SARIMAX(0, 1, 1)x(0, 1, 1, 12)   Log Likelihood                 223.622
Date:                            Fri, 29 Mar 2024   AIC                           -441.245
Time:                                    19:09:32   BIC                           -432.907
Sample:                                01-01-2003   HQIC                          -437.859
                                     - 12-01-2013                                         
Covariance Type:                              opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.3571      0.080     -4.463      0.000      -0.514      -0.200
ma.S.L12      -0.5563      0.094     -5.915      0.000      -0.741      -0.372
sigma2         0.0013      0.000      8.680      0.000       0.001       0.002
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 1.68
Prob(Q):                              0.96   Prob(JB):                         0.43
Heteroskedasticity (H):               0.38   Skew:                             0.01
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.58
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 10: SARIMA(0, 1, 1)(0, 1, 1, 12) result

Evaluation measures:

- MAE: 20.14 units off on average from actual sales.
- MAPE: approximately 2.90% average deviation from actual values, indicating high accuracy.
- MSE: roughly 794.19, highlighting typical squared errors in predictions.
- R2: 95.63% variance in truck sales explained by this model, demonstrating excellent explanatory power.

```
MAE: 20.139720828726855
MAPE: 0.029009642078918197
MSE: 794.1861794298269
R2: 0.9563295374626942
```

Figure 11: SARIMA(0, 1, 1)(0, 1, 1, 12) result

The diagnostic plots show the model effectively captures the time series trends:

- Standardized_Residual: Lack of patterns in residuals implies a good model fit.
- Histogram_plus_Estimated Density: The close match with the normal distribution suggests accurate residual modelling.
- Normal_Q-Q_Plot: Some deviations at the ends, but the dots follow the line quite closely, supports residual normality.
- Correlogram: Autocorrelation is within acceptable limits, indicating no significant residual dependencies.

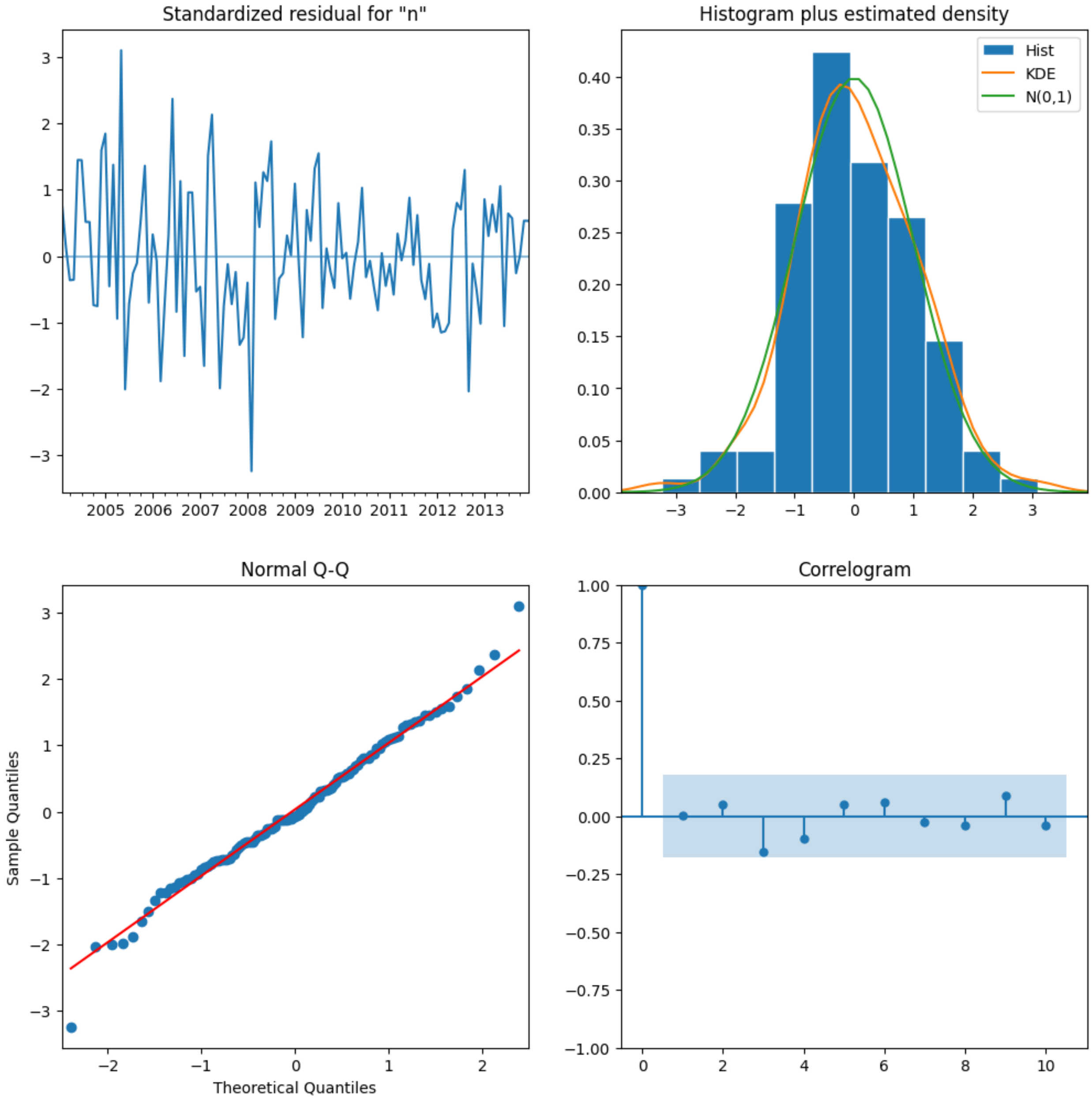

Figure 12: SARIMA(0, 1, 1)(0, 1, 1, 12) result

Compare models' performance

Models' performance comparison:

- Manual_ARIMA(0,1,0): has a negative R2 of -0.60, indicating worse performance than a naive average model, and an AIC of -161.65, suggesting poor fit.
- Manual_SARIMA(0,1,0)(0,1,0,12): shows improvement with an R2 of 0.77, explaining about 77% of variability, and a better AIC of -398.61.
- Manual_Prophet: R2 of 0.64, indicating reasonable fit. Performing better than manual ARIMA but not as well as manual SARIMA, possibly due to its preference for additive models.
- Auto_ARIMA(4,0,4): performs poorly with the lowest R2 of -0.76 and an AIC of -182.95, indicating a less accurate model than simple averages.
- Auto_SARIMA(0,1,1)(0,1,1,12): outperforms all with the highest R2 of 0.96, showing excellent model accuracy, and the lowest AIC of -441.24, indicating **the best fit**.

	Method	Model	MAE	MAPE	MSE	R2	AIC
0	Manual	ARIMA(0, 1, 0)	129.833333	0.158793	29141.000000	-0.602396	-161.652677
1	Manual	SARIMA(0, 1, 0)(0, 1, 0, 12)	59.148360	0.082034	4210.505061	0.768474	-398.612106
2	Manual	Prophet	NaN	NaN	NaN	0.637861	NaN
3	Auto	ARIMA(4, 0, 4)	140.096362	0.173203	32075.059539	-0.763733	-182.946629
4	Auto	SARIMA(0, 1, 1)(0, 1, 1, 12)	20.139721	0.029010	794.186179	0.956330	-441.244785

Figure 13

The graph shows SARIMA(0,1,1)(0,1,1,12) matches the closest, indicating the highest accuracy.

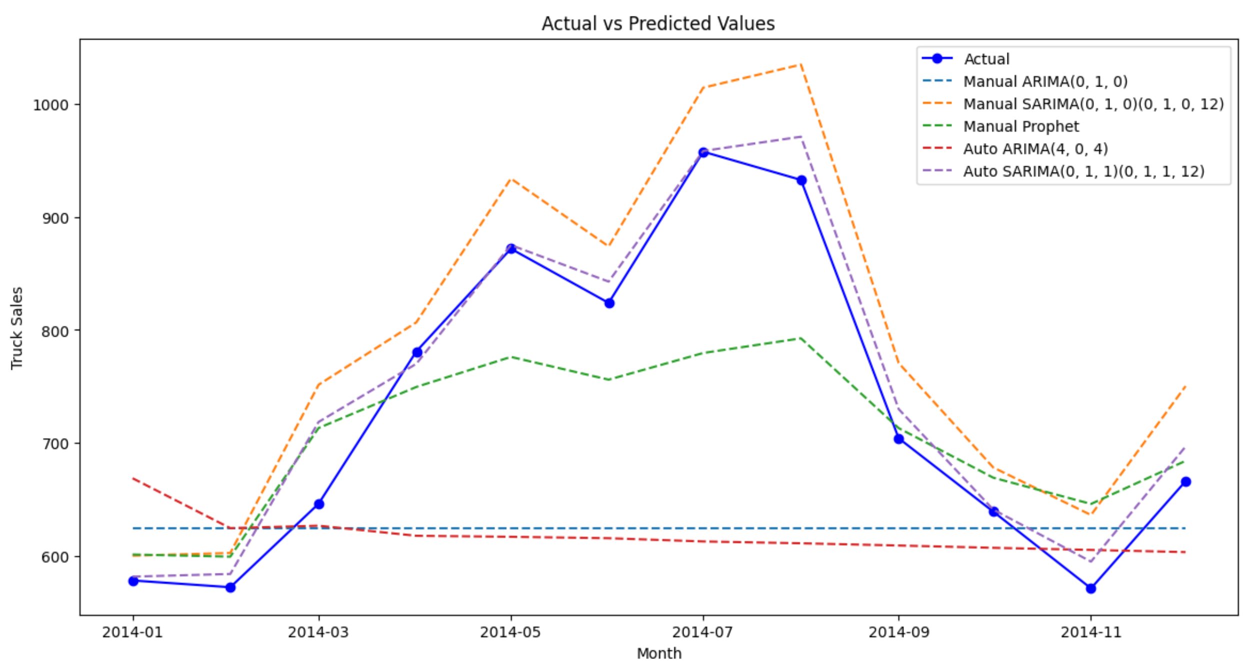

Figure 14

Evaluate model performance with Time Series Cross Validation

Time series cross-validation helps to verify model performance on unseen data, reducing overfitting risk, and confirm the model's reliability over different periods (Shrivastava, 2020).

The results suggest that SARIMA(0,1,1)(0,1,1,12) generally performs well on the given data.

- R2 averages 0.745, lower than the single split's 0.956, but still shows a strong fit.
- AIC averages -204.33, less negative than the single split's -441.24, indicating a less optimal fit compared to the single split model.
- MAE, MAPE and MSE average 34.79, 7.24% and 1911.75 respectively, all are higher than the single split of 20.14, 2.90% and 794.18, indicating larger errors across data subsets.

	Fold	MAE	MAPE	MSE	R2	AIC
0	1	21.701611	0.073348	691.169783	0.646535	-36.082426
1	2	29.053262	0.085037	1186.800469	0.675153	-113.310342
2	3	28.636509	0.058130	1128.516156	0.846401	-191.691568
3	4	33.374239	0.059450	1977.472218	0.817251	-291.432551
4	5	61.173189	0.085818	4574.803550	0.740631	-389.111422

Figure 16

	Measure	Mean
0	MAE	34.787762
1	MAPE	0.072357
2	MSE	1911.752435
3	R2	0.745194
4	AIC	-204.325662

Figure 15

Conclusion

We applied time series decomposition, stationarity tests, and forecasting models to predict sales data. Cross-validation shows the SARIMA(0,1,1)(0,1,1,12) model fits well but with higher errors (MAE, MAPE, MSE) than in single-split testing. This suggests the model's general performance can vary with different data segments, highlighting cross-validation's importance in assessing the model's predictive strength across time.

References

Hyndman, R. J. & Athanasopoulos, G., n.d. *Forecasting: Principles and Practice.* [Online] Available at: https://otexts.com/fpp2/stationarity.html
[Accessed 29 March 2024].

Shrivastava, S., 2020. *Cross Validation in Time Series.* [Online] Available at: https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4
[Accessed 29 March 2024].

Sigmundo Preissler Jr, P., 2018. *Seasonality in Python: additive or multiplicative model?.* [Online] Available at: https://sigmundojr.medium.com/seasonality-in-python-additive-or-multiplicative-model-d4b9cf1f48a7
[Accessed 29 March 2024].