Border Truck Traffic Forecasting

Suraj Mishra

Statistical Forecasting

STAT8041 - Spring 2024 - Section 1

Guided by : Professor Jonatham Plumbtree

Topics

- Introduction
- Problem Statement
- Dataset Overview
- Data Preprocessing
- Data Visualization(ACF,PACF) Visualization(ACF,PACF)
- Data Transformation
- Forecasting and Analysis
- Conclusion

Introduction

Eastern Border Transportation Coalition (EBTC), a non-profit group of representatives of the constituent state and provincial transportation agencies from Michigan

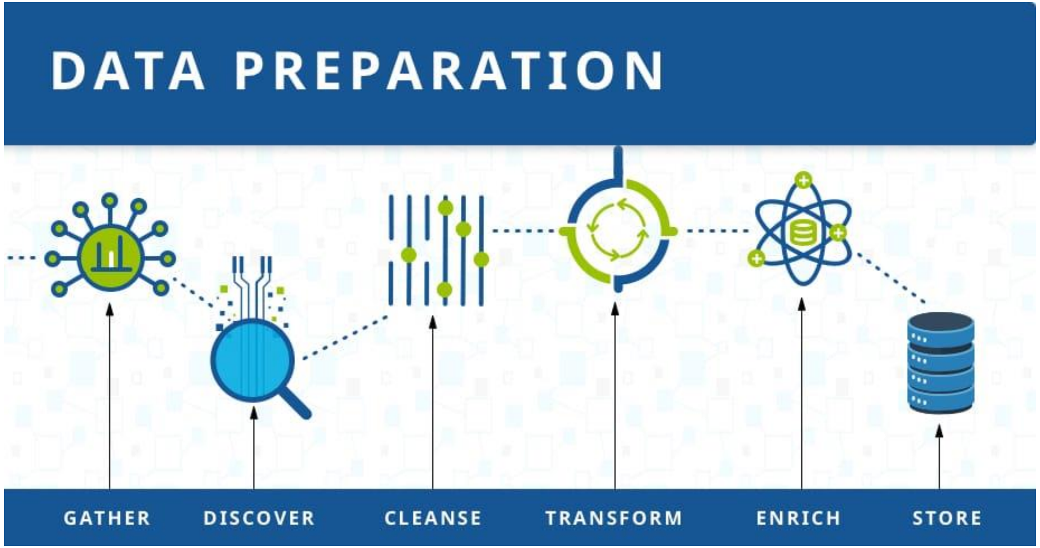
A key objective of the forecasting approach is to develop a straight-forward and efficient model that provides defensible forecasts based on historical trends.

To ensure that there are adequate facilities and personnel in place to meet the continued growth in trade between the two countries

About the Dataset

Port Name	State	Port Code	Border	Date	Measure	Value	Latitude	Longitude
Roma	Texas	2310	US-Mexico Border	Dec 2023	Buses	46	26.404	-99.019
Del Rio	Texas	2302	US-Mexico Border	Dec 2023	Trucks	6552	29.327	-100.928
Willow Creek	Montana	3325	US-Canada Border	Jan 2024	Pedestrians	2	49.000	-109.731
Whitlash	Montana	3321	US-Canada Border	Jan 2024	Personal Vehicles	29	48.997	-111.258
Ysleta	Texas	2401	US-Mexico Border	Jan 2024	Personal Vehicle Passengers	521714	31.673	-106.335

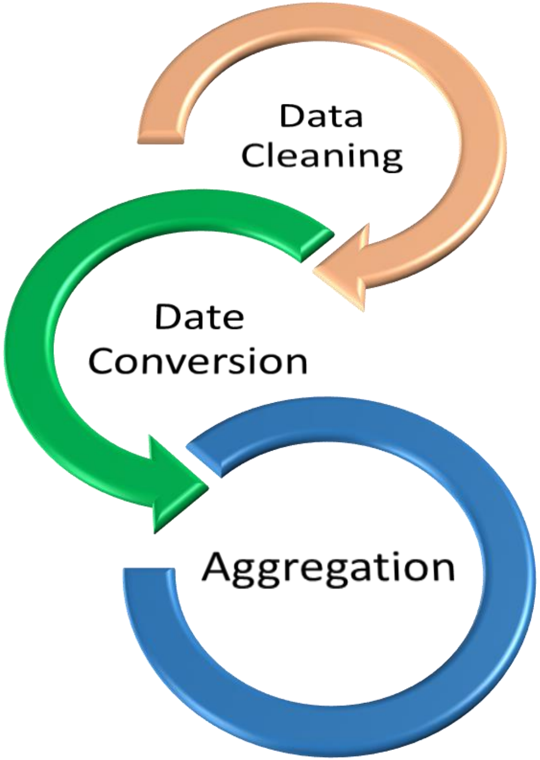
Number of Records: 393,000
Variables: 10



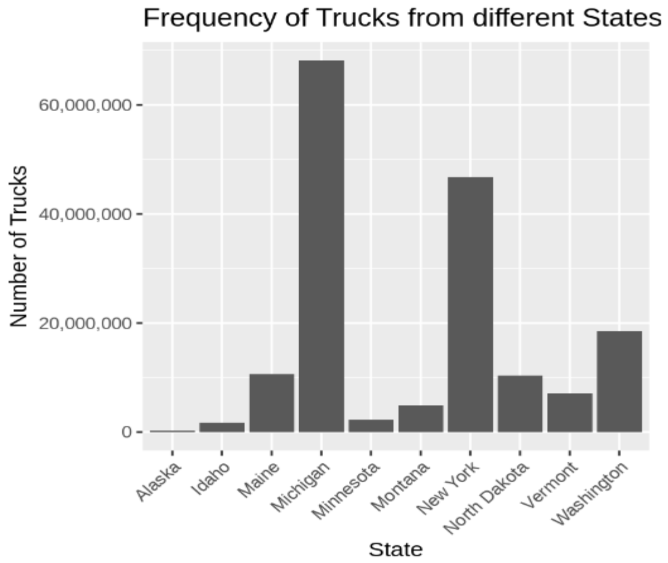
Date conversions using mutate()

Check for null and duplicates

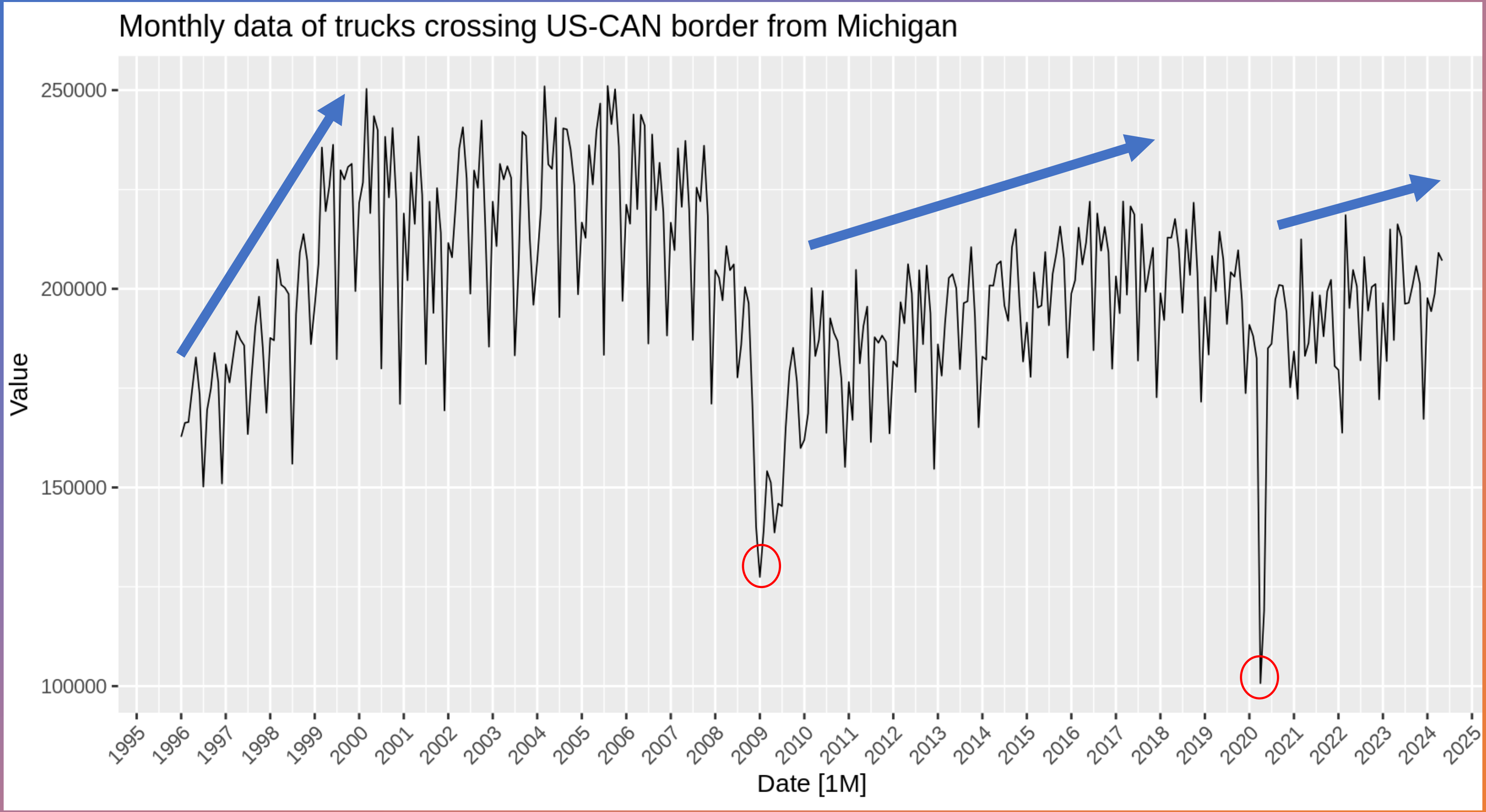
Combining ports data into monthly average



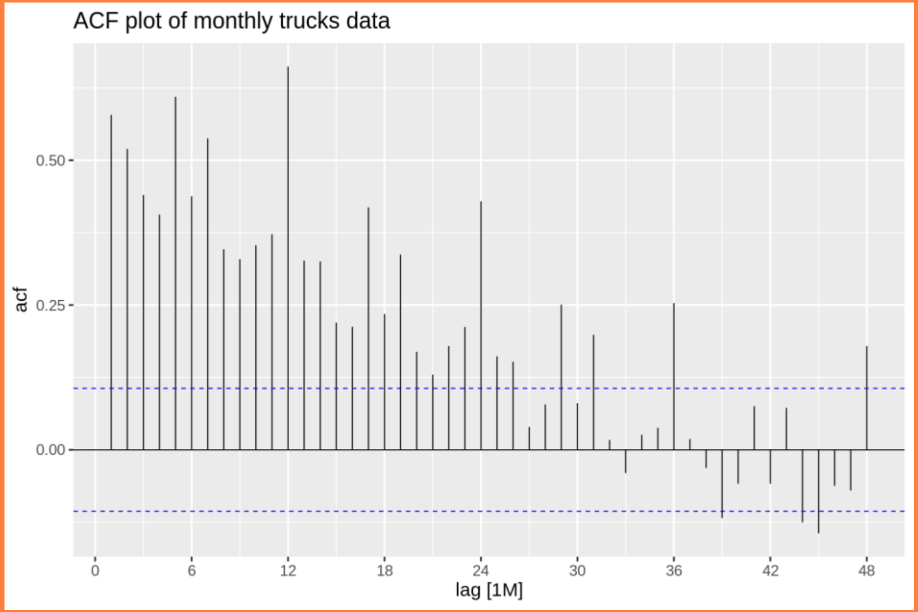
DATA DISTRIBUTION OF NUMBER OF TRUCKS



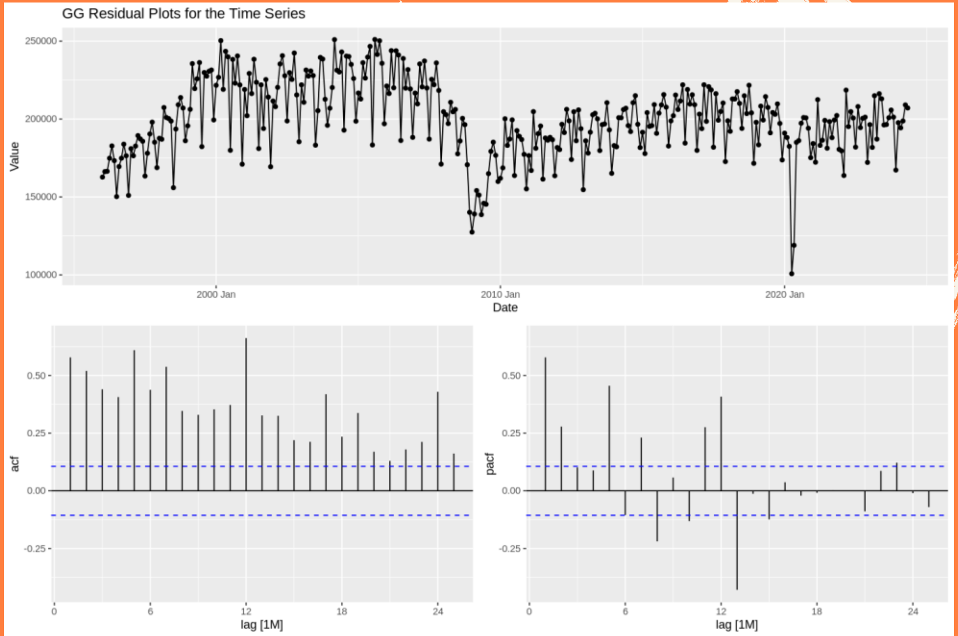
- ❑ Michigan seems to have the highest movement of trucks across the border
- ❑ Detroit handles the highest volume of truck traffic with considerable variability, Port Huron has moderate traffic with less variability, and Sault Sainte Marie has the lowest.


Monthly data of trucks crossing US-CAN border from Michigan
Value
250000
200000
150000
100000
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Date [1M]

AUTOCORRELATION PLOTS

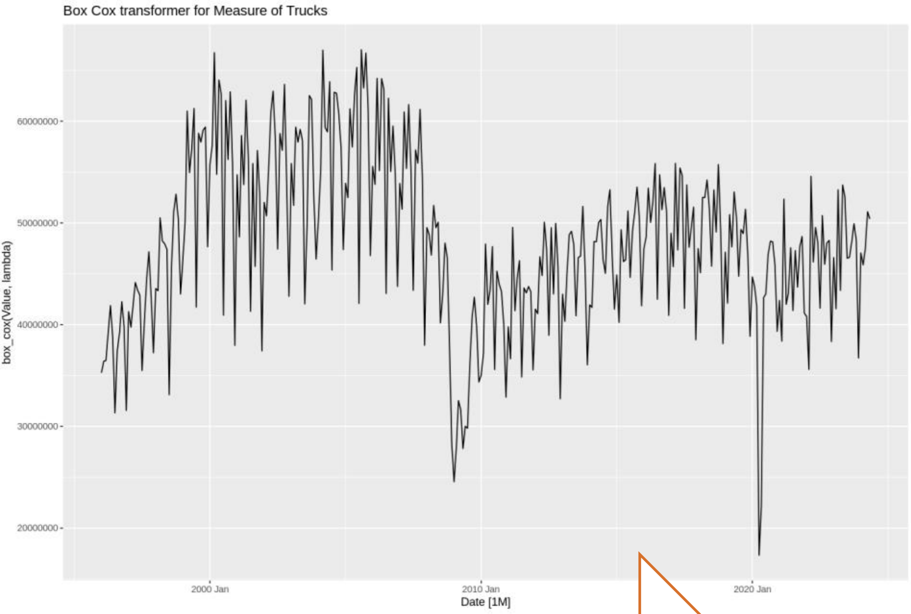


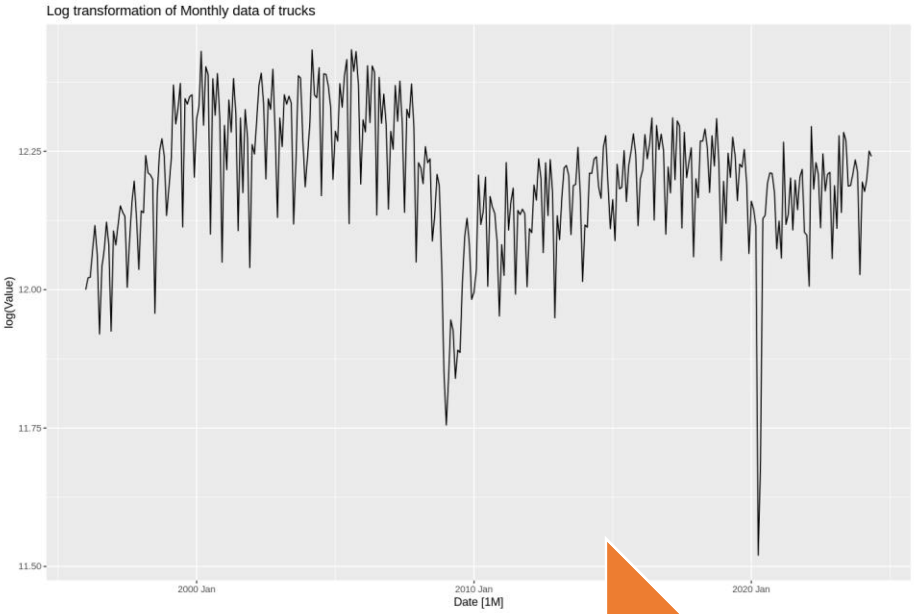
ACF PLOT



GS Display Plots

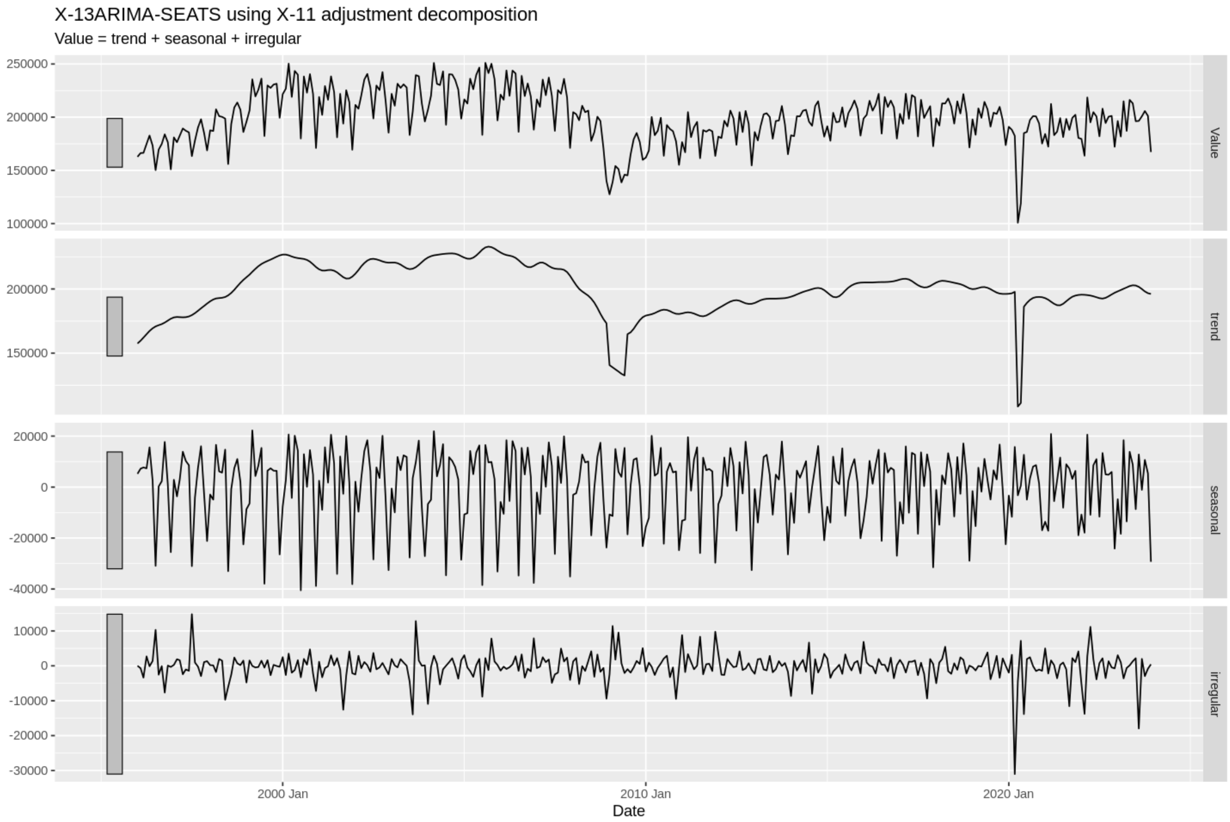
DATA TRANSFORMATIONS





Box Cox transformations hardly had any effect variance in the values

Log transformations had shown some significant improvement in capturing the data variance



We used XTL decomposition since the data has seasonal variations , hence we decomposed to get trend and seasonality

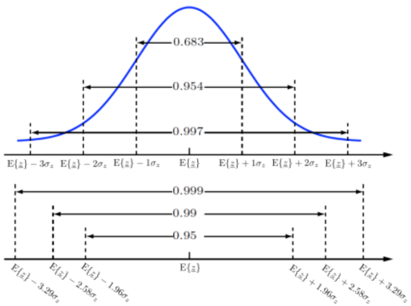
Time Series Model

ERROR ANALYSIS

Identify and rectify forecast errors errors

VALIDATION METRICS

Assess forecast performance using using metrics

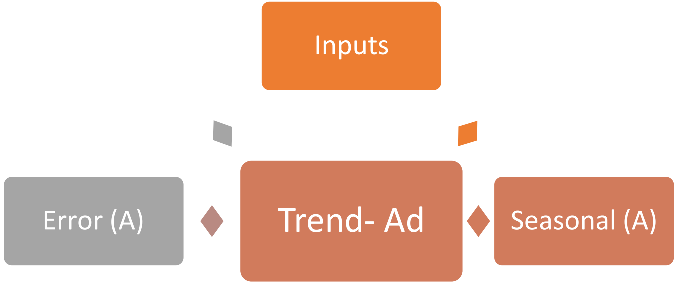


CONFIDENCE INTERVALS

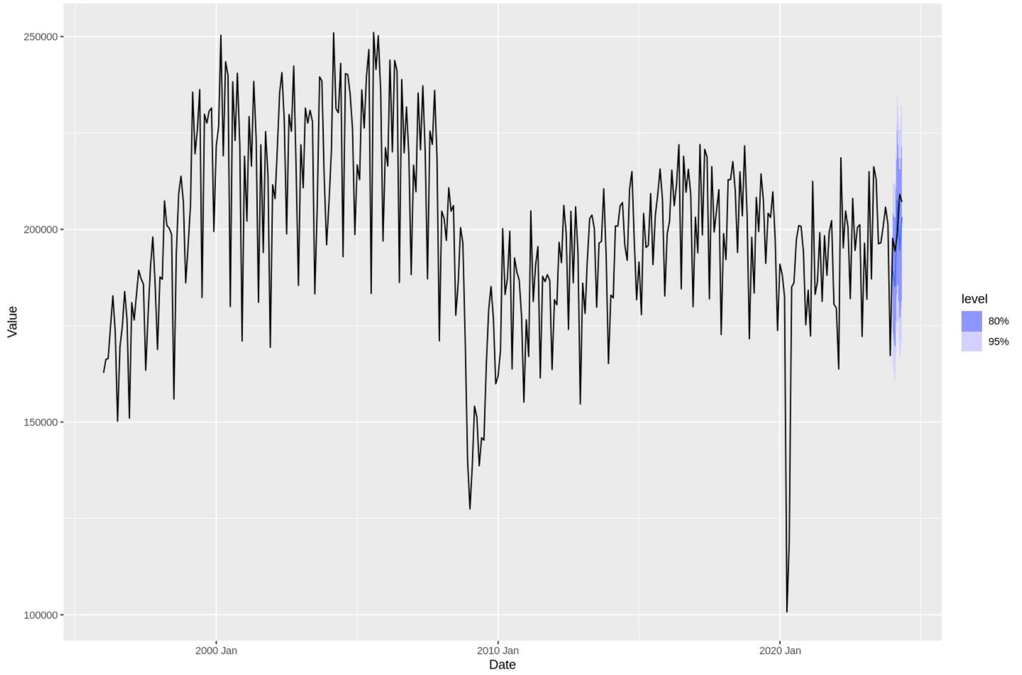
Estimate range of forecasted values values

Try Pitch

EXPONENTIAL SMOOTHENING MODEL

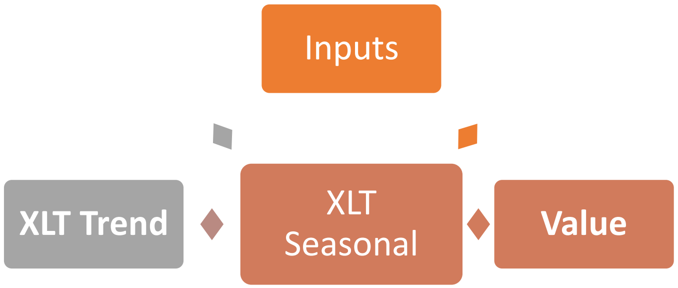


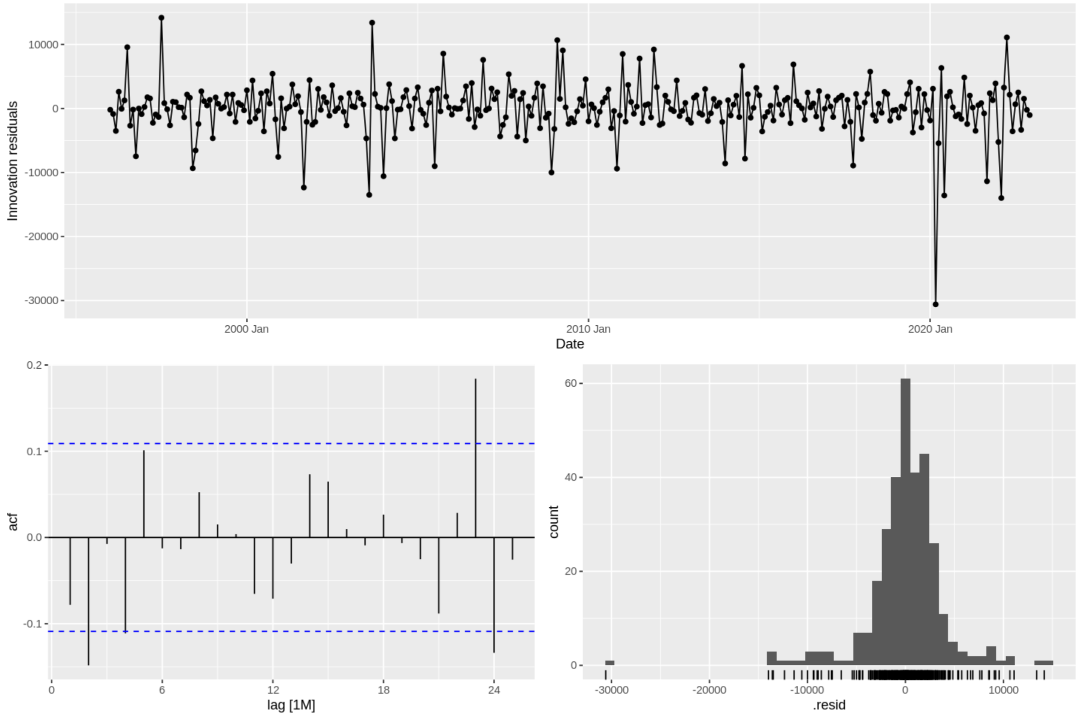
Model fit Parameters	**Model**: ETS(A, Ad, A) **Smoothing parameters**: alpha = 0.4308 beta = 0.0001000294 gamma = 0.0001001801 phi = 0.9709474	AIC 8281.540 AICc 8283.698 BIC 8350.248



MSE: 91303809.3420198
RMSE: 9555.30268186308
MAPE: 0.0444392199410917
MAE: 8955.87852854038

MULTIPLE LINEAR REGRESSION

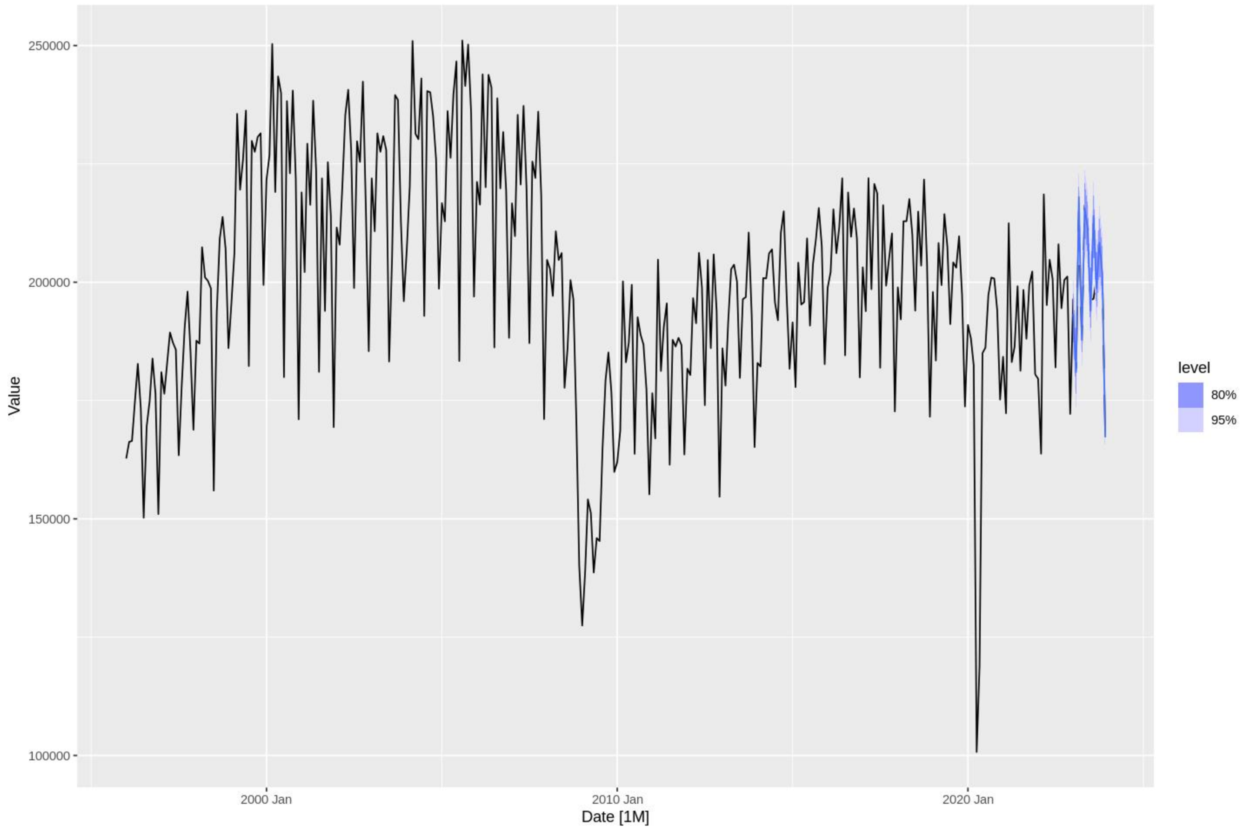




Residual standard error: 4045 on 321 degrees of freedom
Multiple R-squared: 0.973,
Adjusted R-squared: 0.9728
F-statistic: 5782 on 2 and 321 DF,
p-value: < 0.0000000000000000222

Residuals show stationary and are normally distributed

Multiple_Linear_Predict
<dbl>
193296.7
181009.9
218033.2
187772.7
215648.3
211317.8
193988.3
214006.2
198687.6
208360.1
201801.7
167222.8



Steps to apply forecasting for Auto Regressive Models

PLOT DATA

DATA TRANSFORMATION

DATA STATIONARITY

EXAMINE ACF AND PACF AND VERIFY RESIDUALS

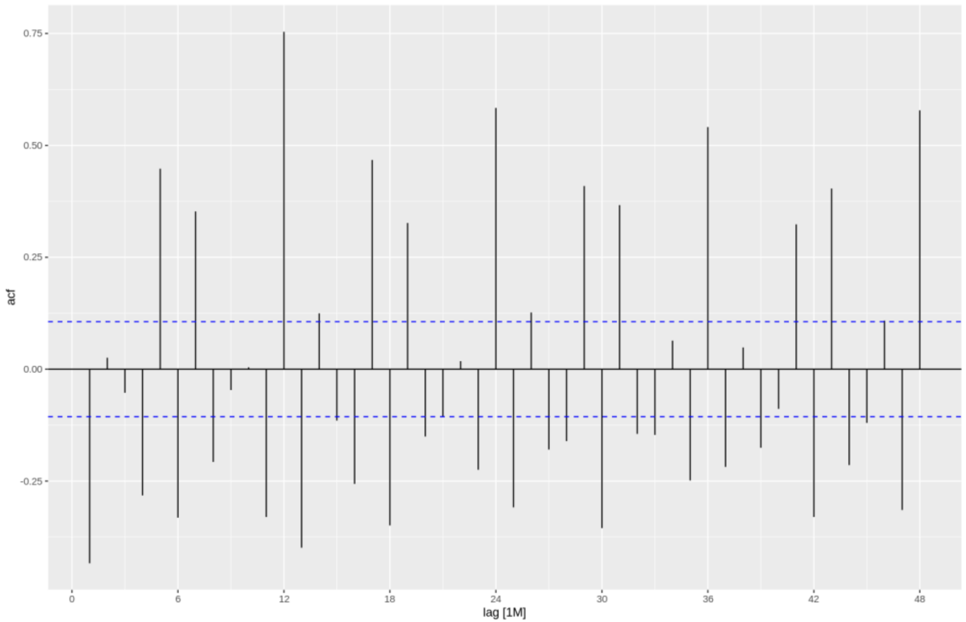
KPSS TEST

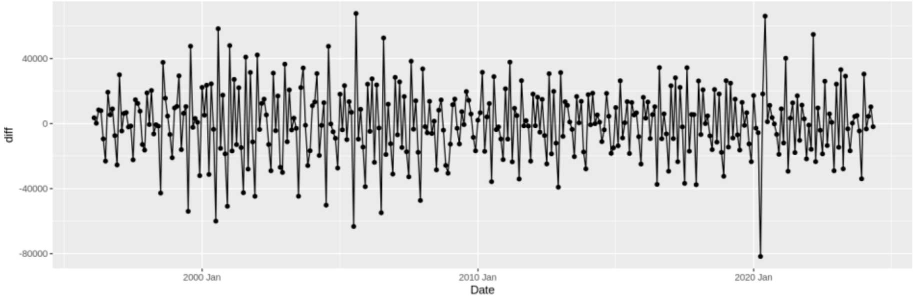
CHOOSE NUMBER OF AR AND MA TERMS FOR SEASONAL AND NON-SEASONAL ARIMA MODELS

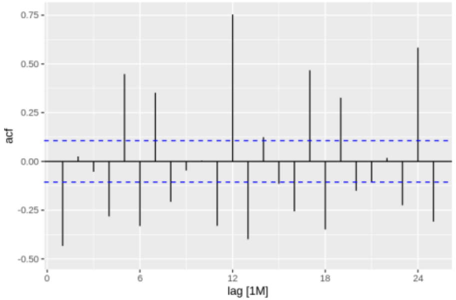
MODEL FITTING AND EVALUATION

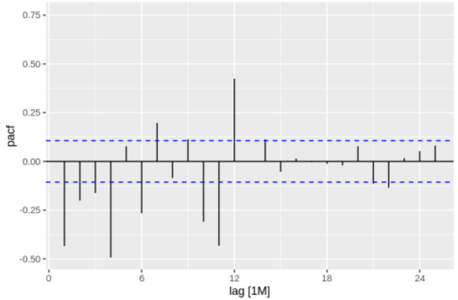
STATIONARITY OF THE DATA

ACF plot of first difference of trucks data





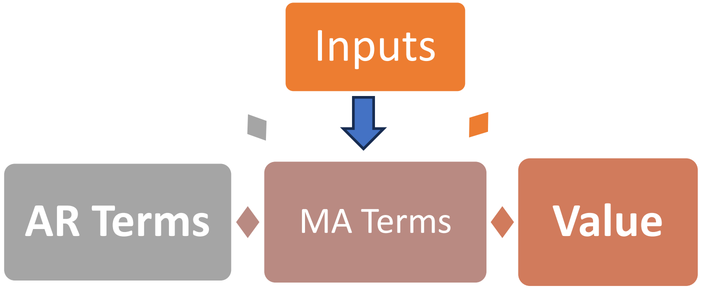




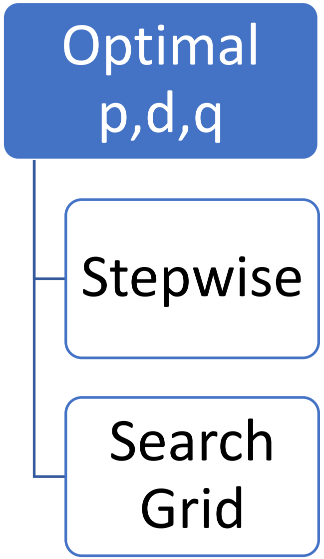
A tibble: 1 × 2

kpss_stat	kpss_pvalue
<dbl>	<dbl>
0.02943175	0.1

ARIMA MODELS



p	d	q	Model
1	1	1	ARIMA(1,1,1)
2	1	2	ARIMA()
0	1	2	ARIMA()

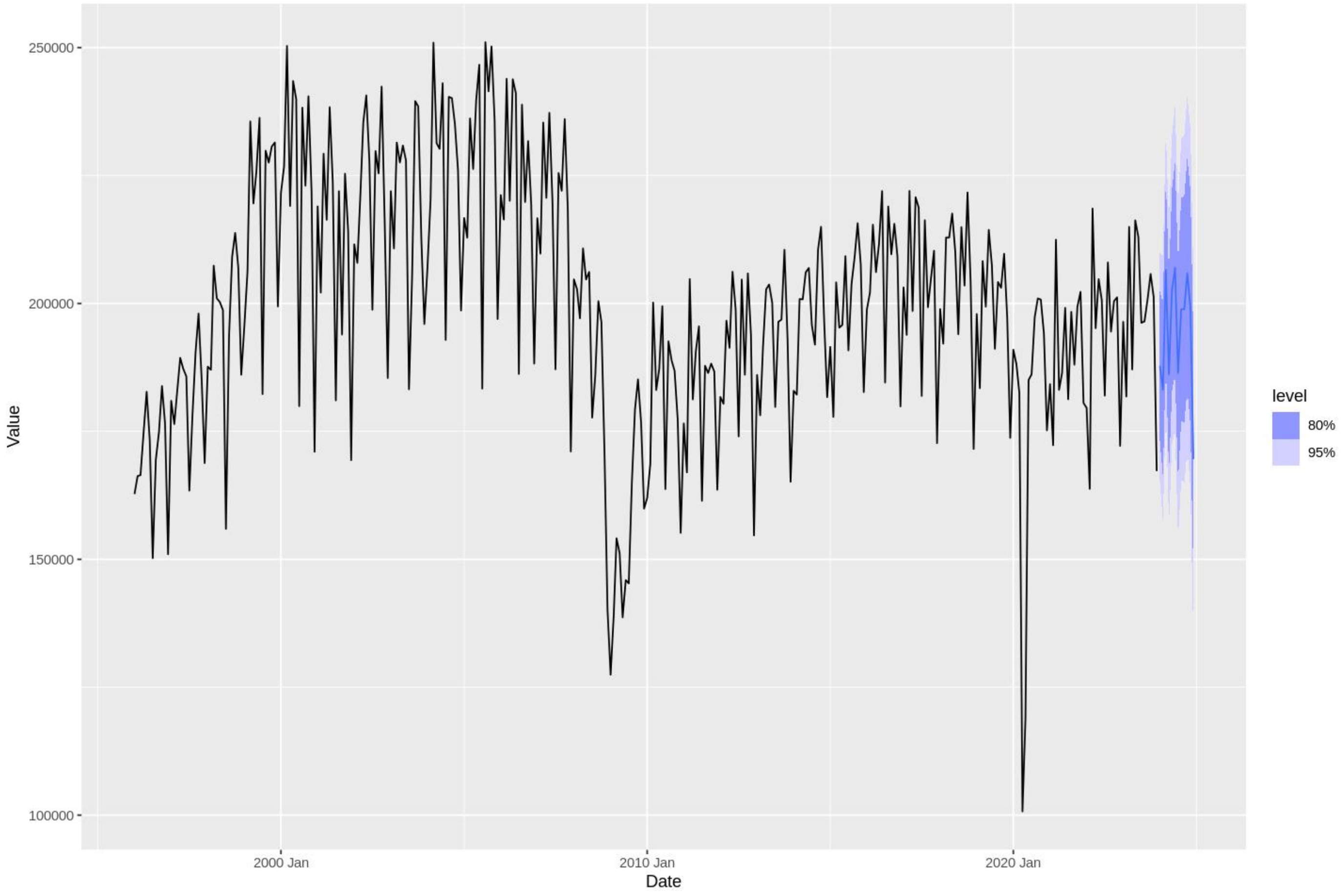


Optimal Model: ARIMA(2,0,2)

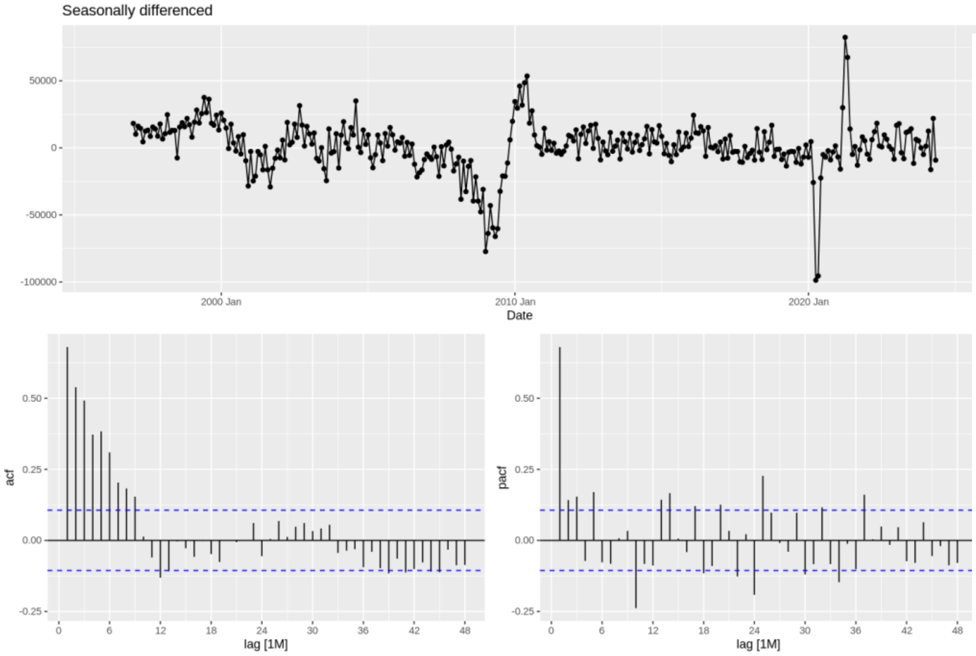
sigma^2 estimated as 128278747:

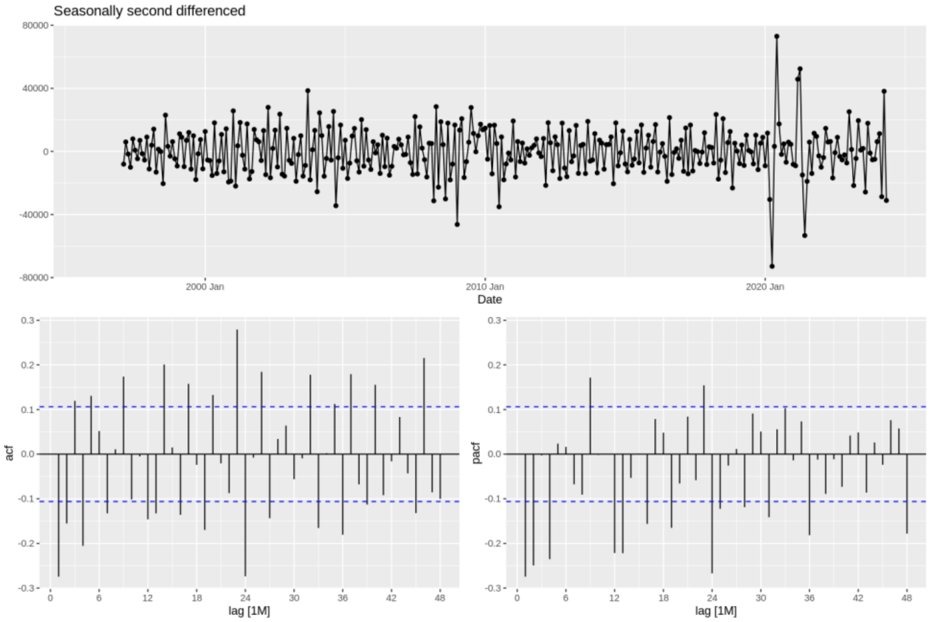
log likelihood=-3487.68

AIC=6989.37
AICc=6989.72
BIC=7015.83



SEASONALLY ADJUSTED DATA





Seasonal differenced data does not look stationarity

Seasonal second differenced data does look stationarity

SEASONAL ARIMA MODELS

p	d	q	Model
0	1	2	ARIMA(1,1,1)
2	1	0	ARIMA(2,1,0)

P	D	Q	Seasonal Model
1	1	1	(1,1,1)
2	1	2	(0,1,2)
1	1	2	(1,1,2)

Optimal (p,d,q)

Optimal (P,D,Q)

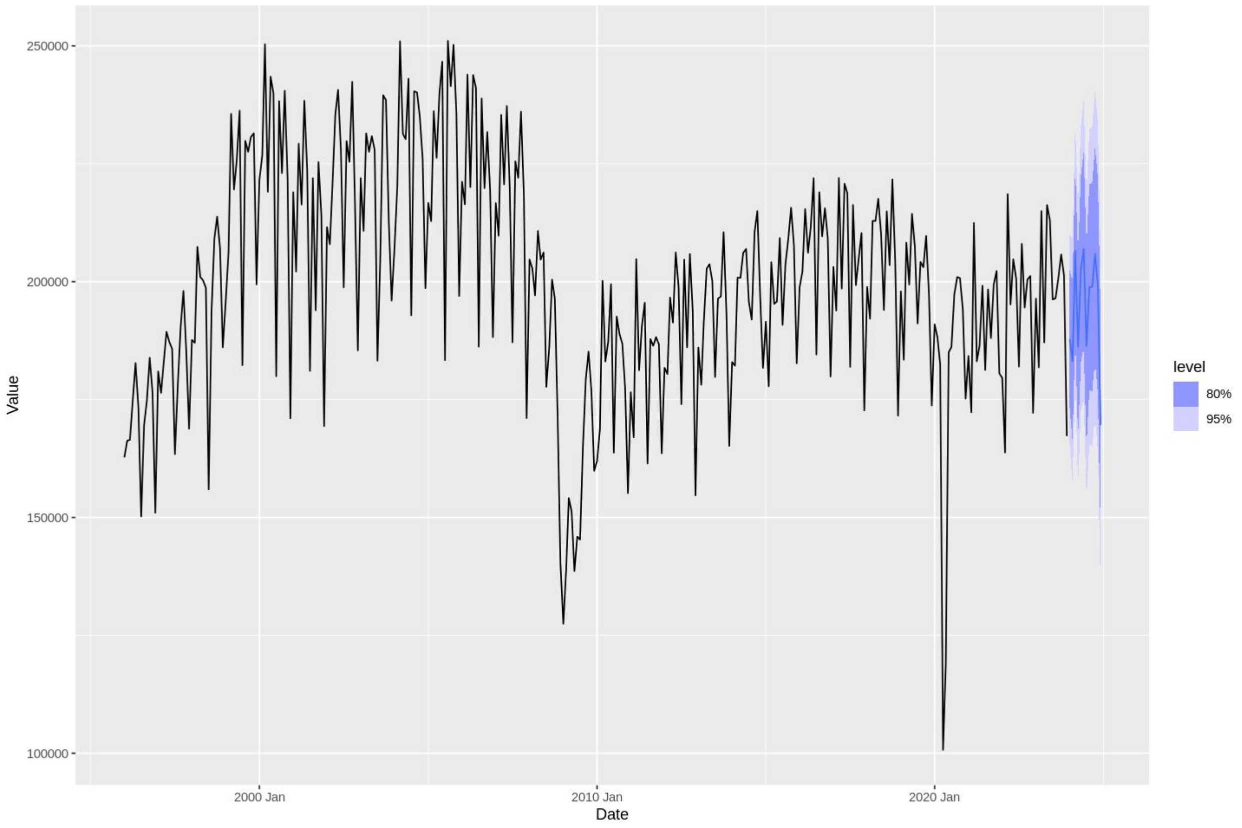
- Stepwise
- Search Grid

Optimal Model:
ARIMA(2,0,2) (0,1,2)

sigma^2 estimated as 128278747:

log likelihood=-3487.68

AIC=6989.37
AICc=6989.72
BIC=7015.83



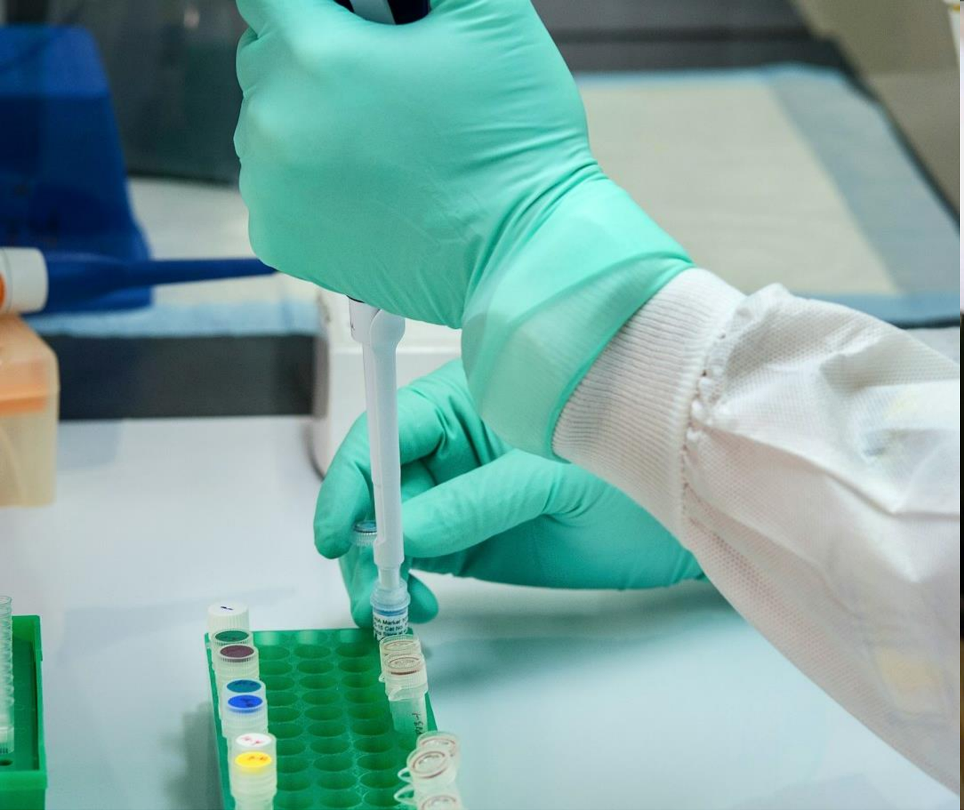
PROJECT SUMMARY

MODEL	RMSE	MAE
ETS	9555.302	8955.878
Multiple Linear	5378.155	2919.457
ARIMA	12508.005	11063.411
SARIMA	12341.961	10920.49

- ✓ ARIMA AND SARIMA show similar accuracy but overall Multiple Linear Regression shows the best accuracy in terms of RMSE and MAE.
- ✓ We recommend use of Linear regression since its easy to implement with less computational costs.

DATA INSIGHTS

Leverage trends for forecasts forecasts



Empowering insights for better decisions and customer relations.