A Deep Dive into Thermo and Gas systems

Filippo Santin, Gurjeet Singh, Francesca Zen

The Dataset

- **WHO** : a local company

- **WHAT** : data about thermoregulation, gas security and combustion analyzers

- **WHERE** : Italy and abroad

- **WHEN** : from 2009 to 2019

- **WHY** : analyze and forecast time series sales

- 153142 rows
- 10 columns
- 3 sectors

- 70 Commercial Groups
- 277 Commercial Families
- 2754 Articles
- 90 Nations

|	Nation	Region	Commercial Group	Family Group	Article	Year	Month	Date	Sales	Sector
0	ITALIA	Toscana	ELIOS	ELIOS MIDI	TDSA04M000AN	2009	1	2009-01-07	5.0	Termoregolazione
1	ITALIA	Campania	FUEGO	FUEGO2	CIE001MD00SE	2009	2	2009-02-27	6.0	Termoregolazione
67	BELGIO	BELGIO	ACCESSORI PORTATILI	SONDE TEMPERATURA	STS1C00000AN	2009	11	2009-11-12	1.0	Analizzatori e strumentazione portatile
68	ITALIA	Lombardia	ACCESSORI PORTATILI	SONDE TEMPERATURA	STSKZ00000SE	2009	7	2009-07-01	1.0	Analizzatori e strumentazione portatile
238	FRANCIA	FRANCIA	RIVELATORI GAS	SNIFFER	PORRDZBI00KI	2009	6	2009-06-05	50.0	Sicurezza Gas
239	ITALIA	Veneto	RIVELATORI GAS	SNIFFER	PORRDZBI00SE	2009	5	2009-05-28	1.0	Sicurezza Gas

Exploratory Analysis

Time Series of all Sales
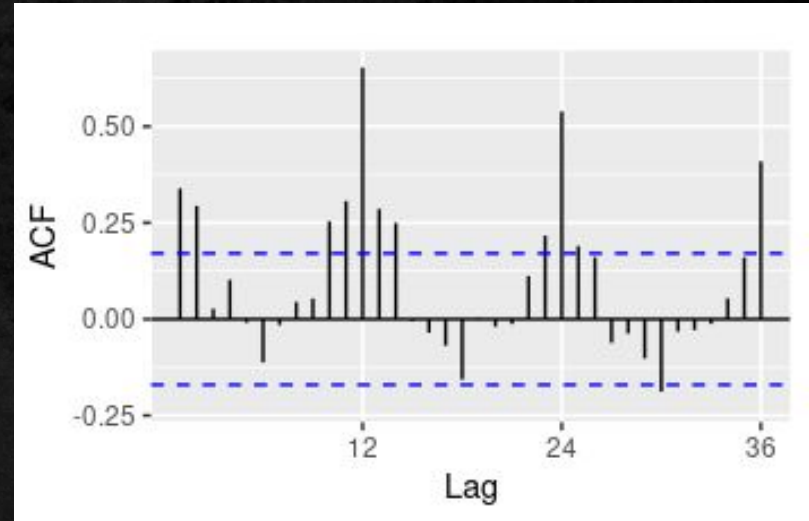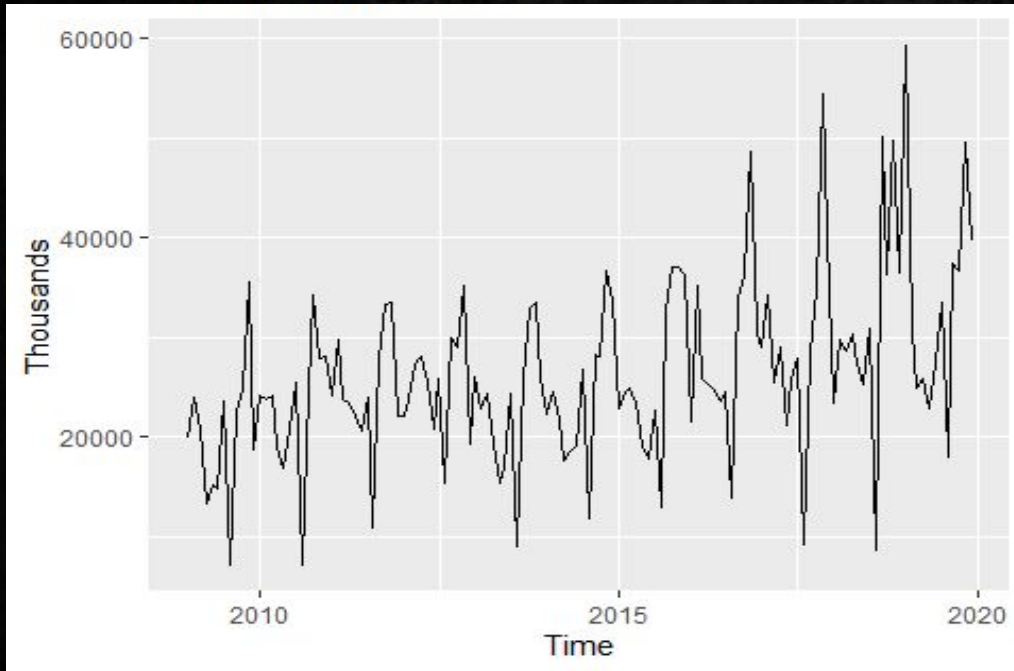
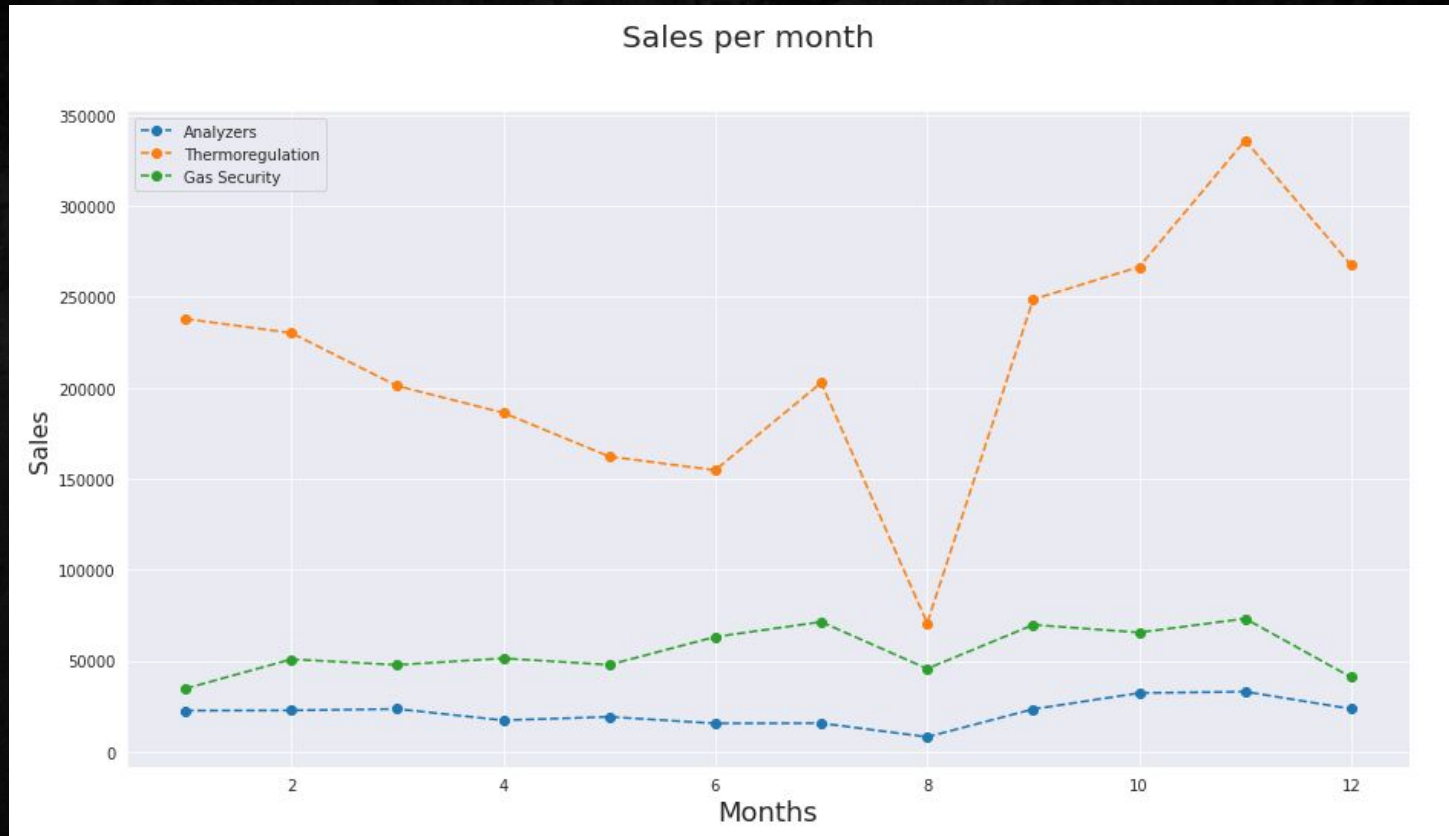

- Strong monthly seasonality
- Positive incrementing trending
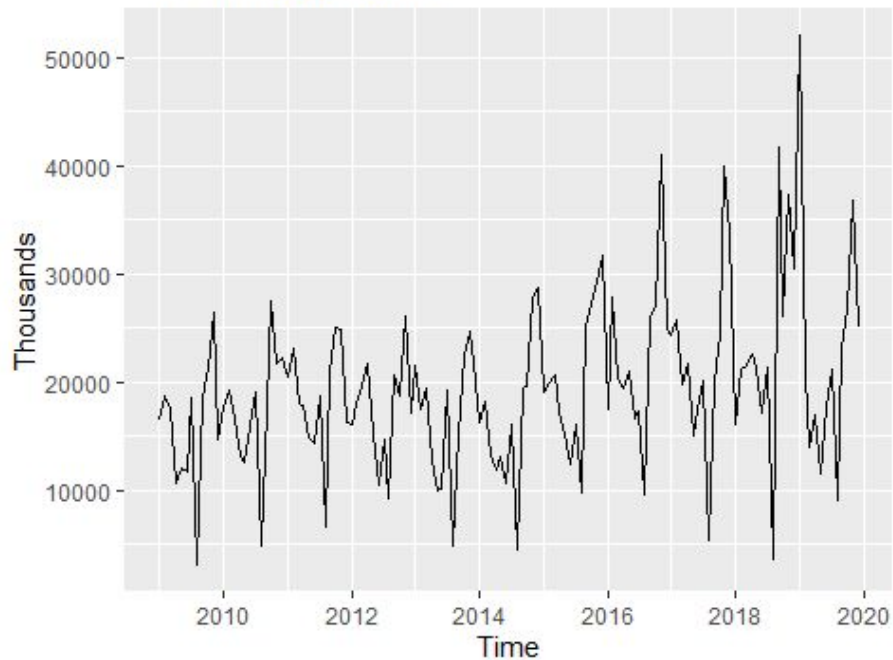
Sectors Sales Boxplot

The 3 Sectors

Sales per month
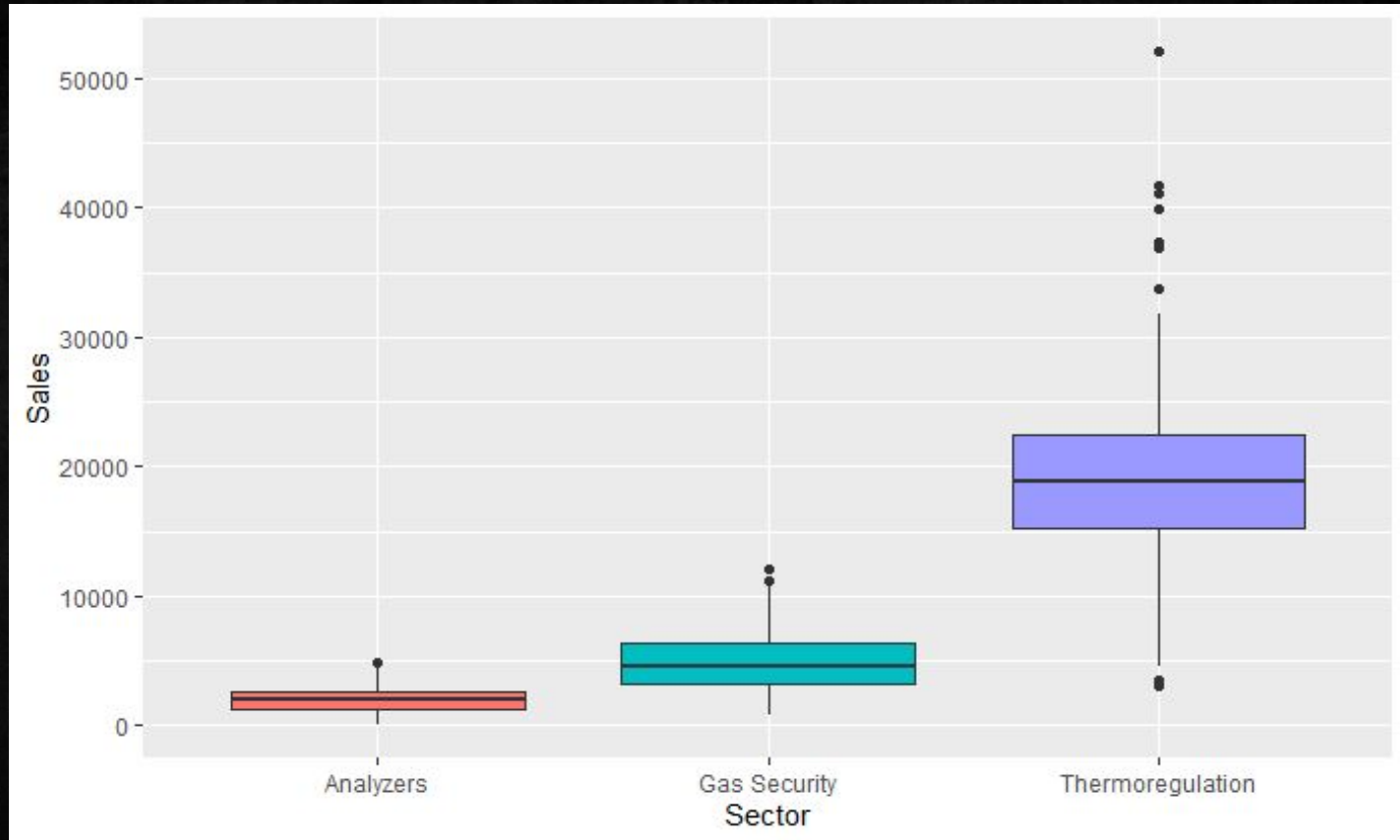
Thermoregulation

Thermoregulation

- ❏ chrono-thermostats
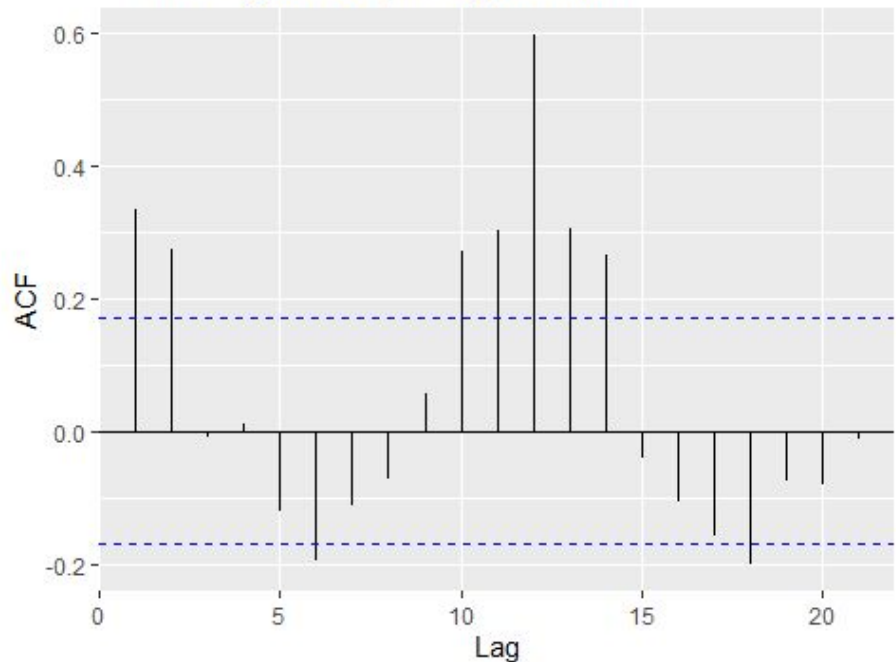
- ❏ thermostats & central heating

- ❏ accessories

Thermoregulation

1. Time Series Linear Model

Time Series Linear Model

termo_sales ~ trend + season

| PREDICTORS	COEFFICIENT	P-VALUE
Intercept	17704.33	<2 e-16
Trend	64.33	1.28 e-08
May	-7136.26	3.95 e-04
June	-7883.86	9.93 e-05
August	-15651.74	9.46 e-13
November	8275.96	4.72 e-05

Time Series Linear Model
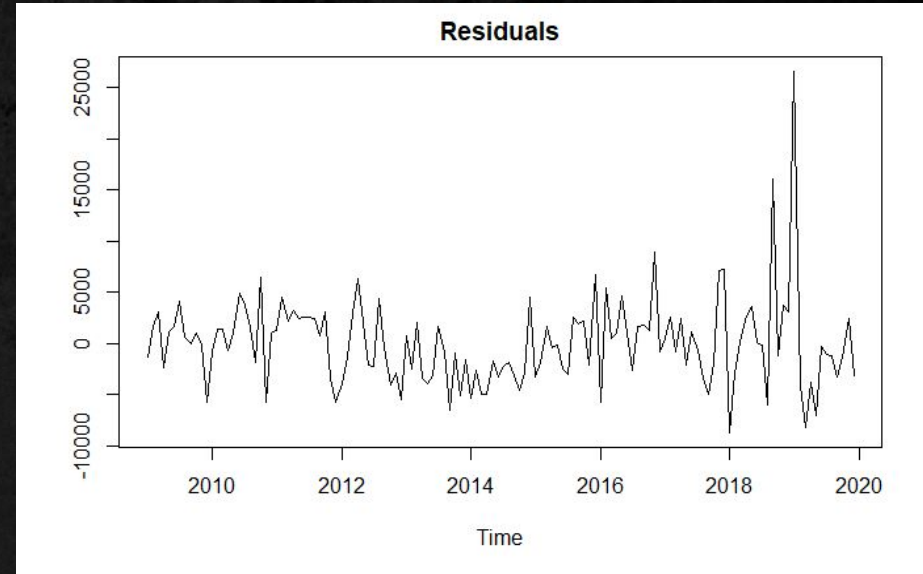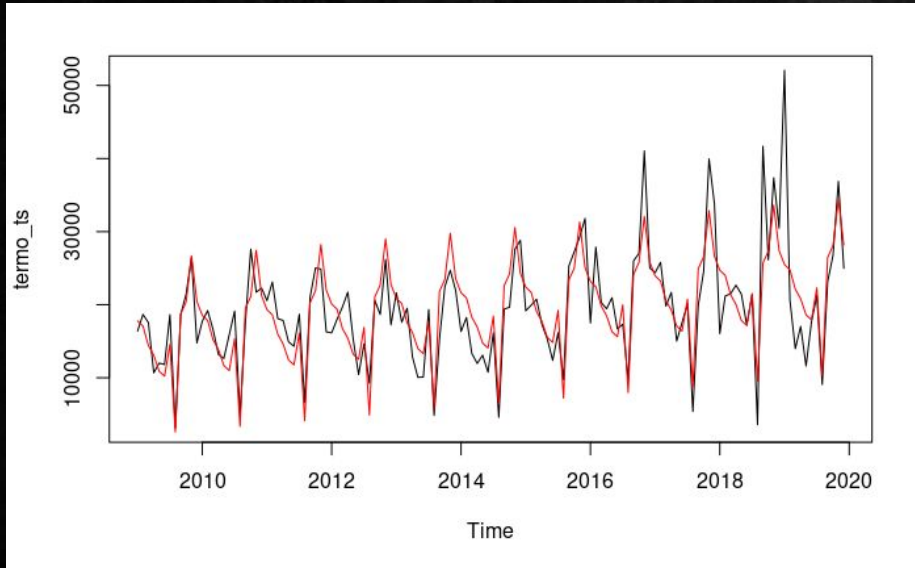
DW = 1.8439, **p-value = 0.3426.**
Almost stationary

2. Generalized Bass Model

Generalized Bass Model

Predictors	coefficient	p-value
m	2.86 e+06	2.11 e-55
p	5.61 e-03	1.04 e-58
q	9.64 e-03	8.12 e-30
a_1	7.13 e+01	6.60 e-76
b_1	4.43 e-02	1.87 e-18
c_1	2.32 e-01	5.20 e-15

Fitted values
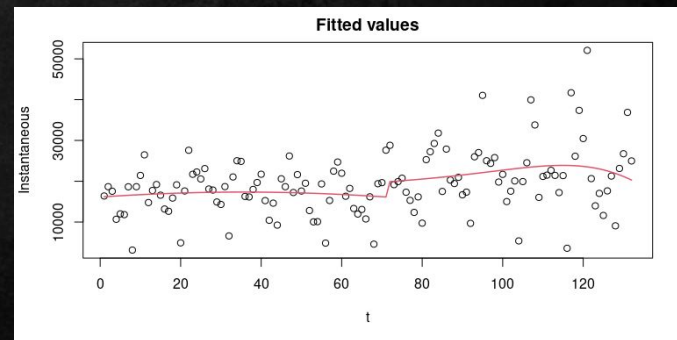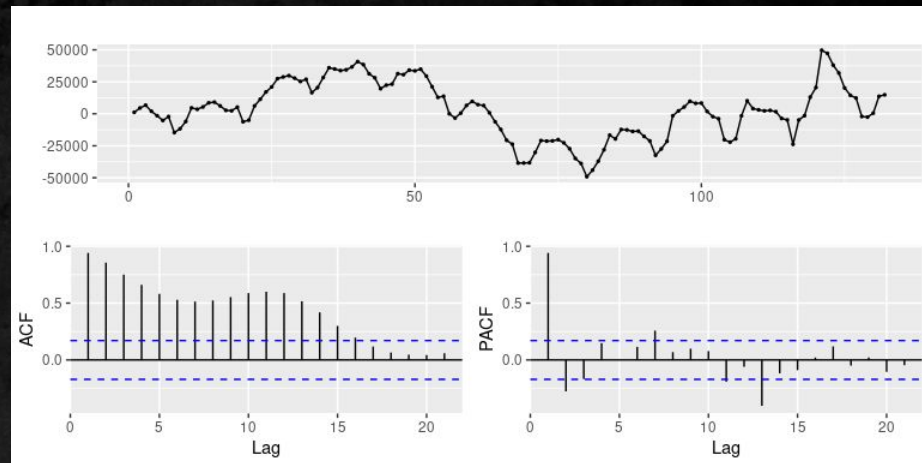
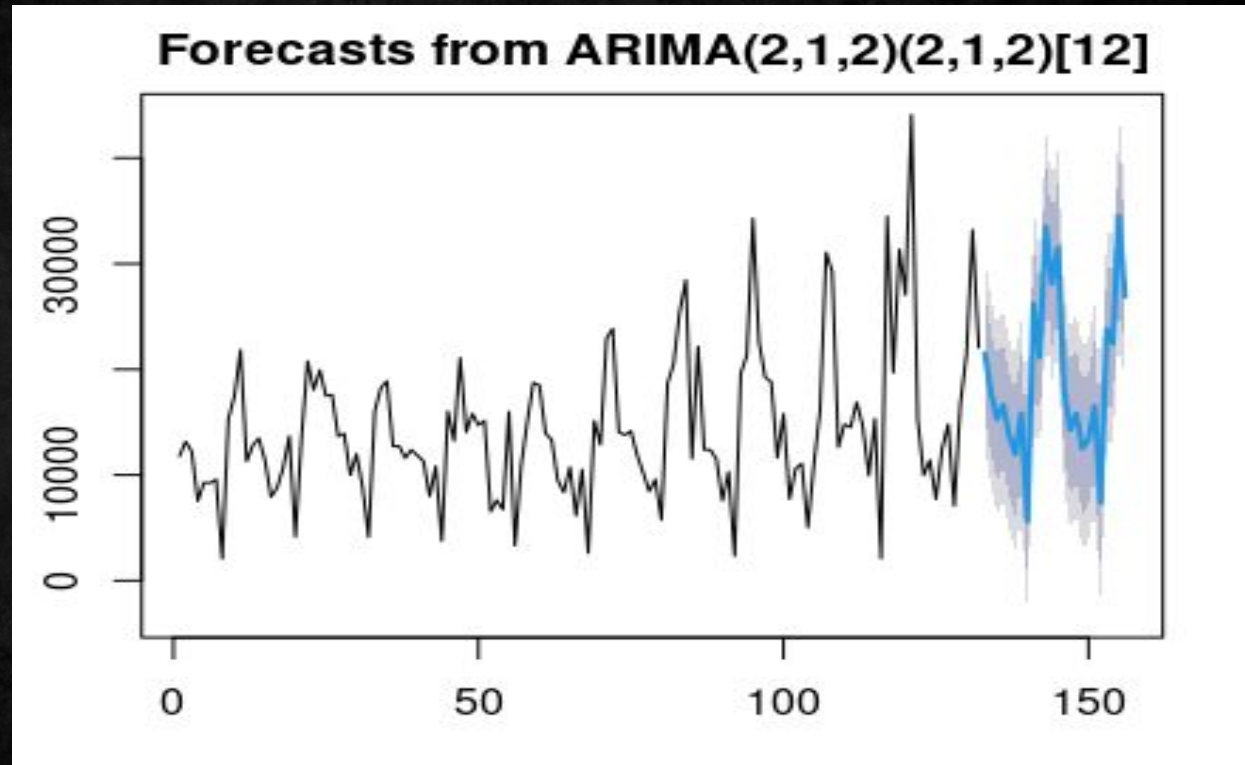# 3. Seasonal ARIMA Model

Seasonal ARIMA Model (Final)

ARIMA parameters:

- order = (2, 1, 2)

- seasonal = (2, 1, 2)

- period = 12

Forecasts from ARIMA(2,1,2)(2,1,2)[12]

4. Thermoregulation summary

Thermoregulation summary

| MODEL	RMSE	ADJR^2	MAPE	AIC
tslm	4356.24	0.649	18.13	2614.75
Bass Model	21700.01	0.999	41.10	NA
Generalized BM	11976.98	0.999	41.06	NA
ARIMA	4161.013	NA	20.62	2361.61

Combustion Analyzers

Combustion Analyzers & Co.

- Portable instruments (e.g. gas leak detector and manometer)

- Instruments for the analysis of boiler fumes and of emissions and industrial combustion
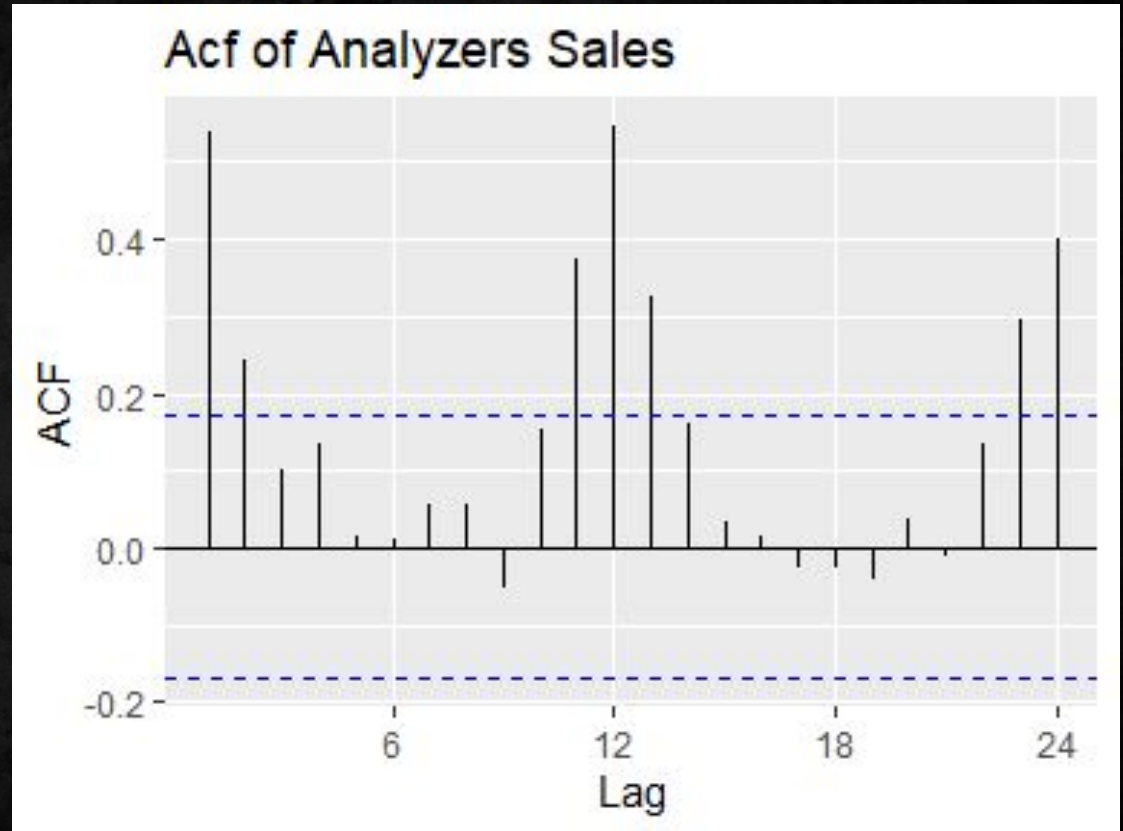
Combustion Analyzers
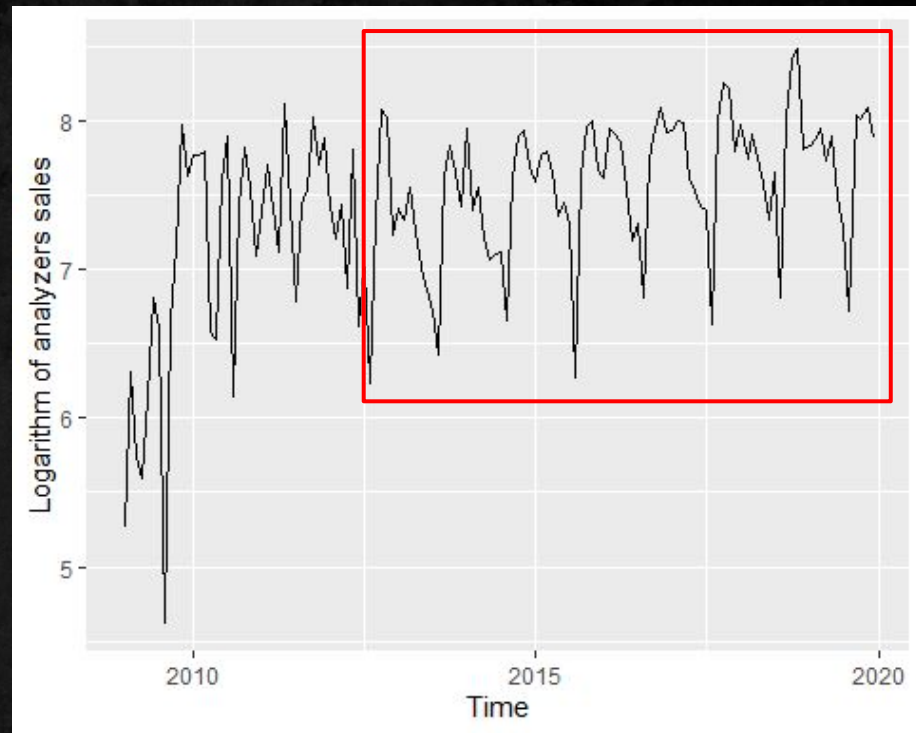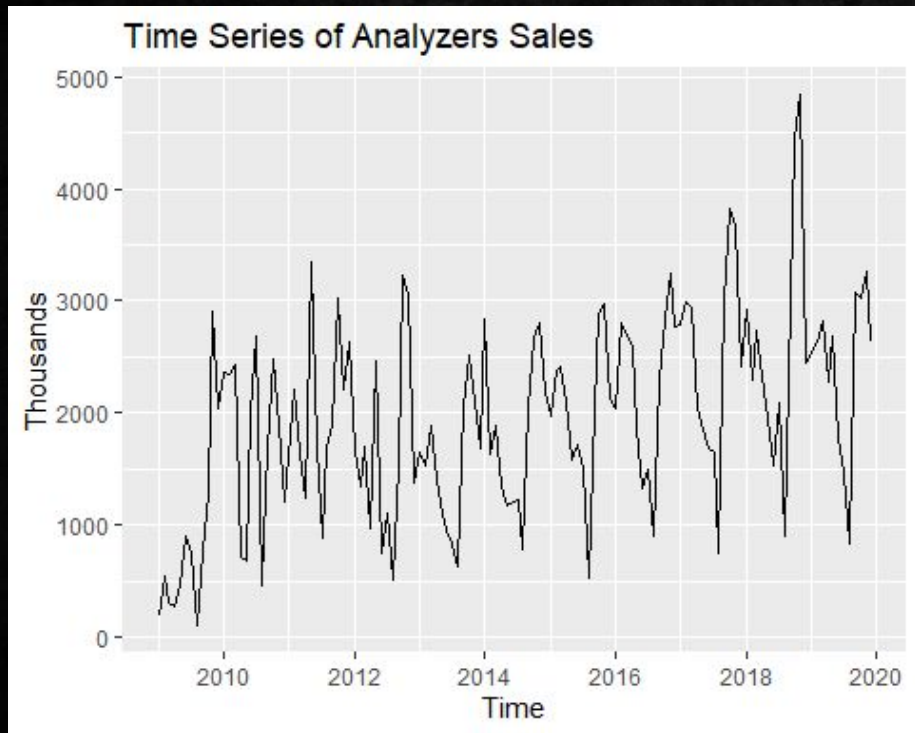
Combustion Analyzers

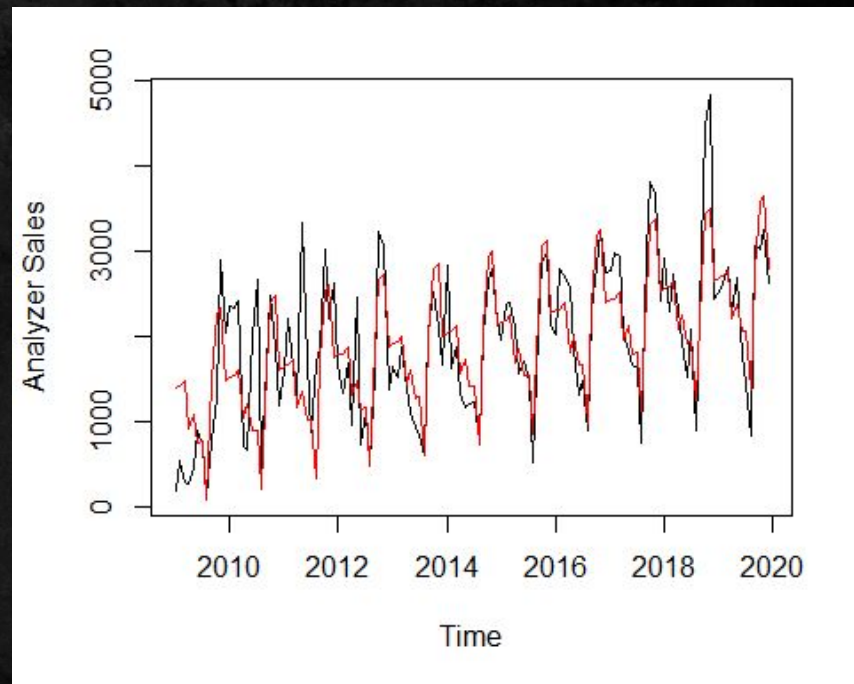- seasonality at lags 1, 12 and 24

- local trend towards January
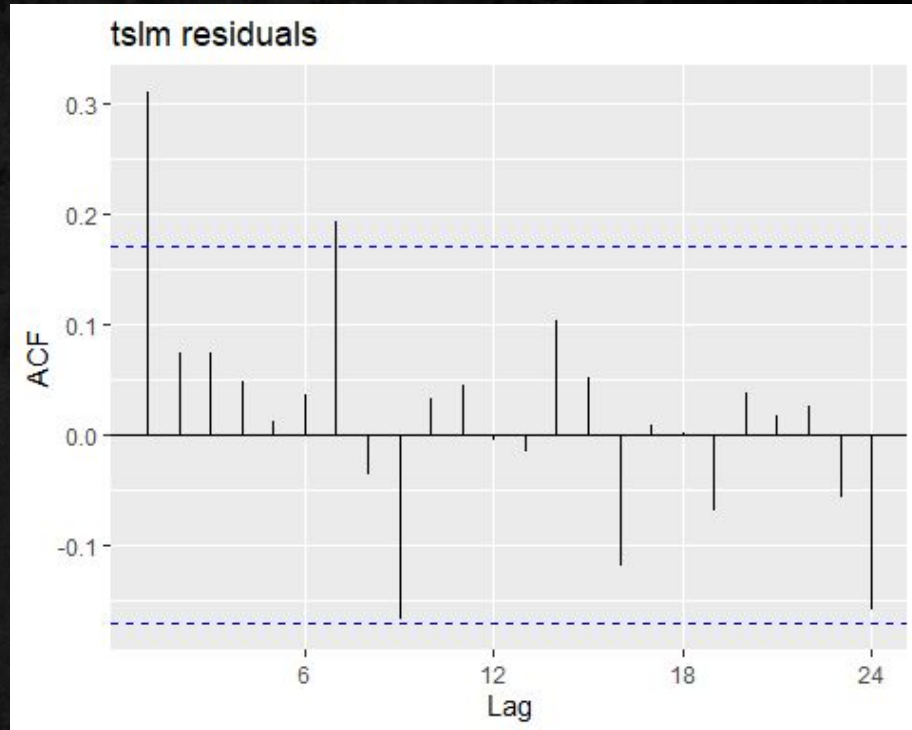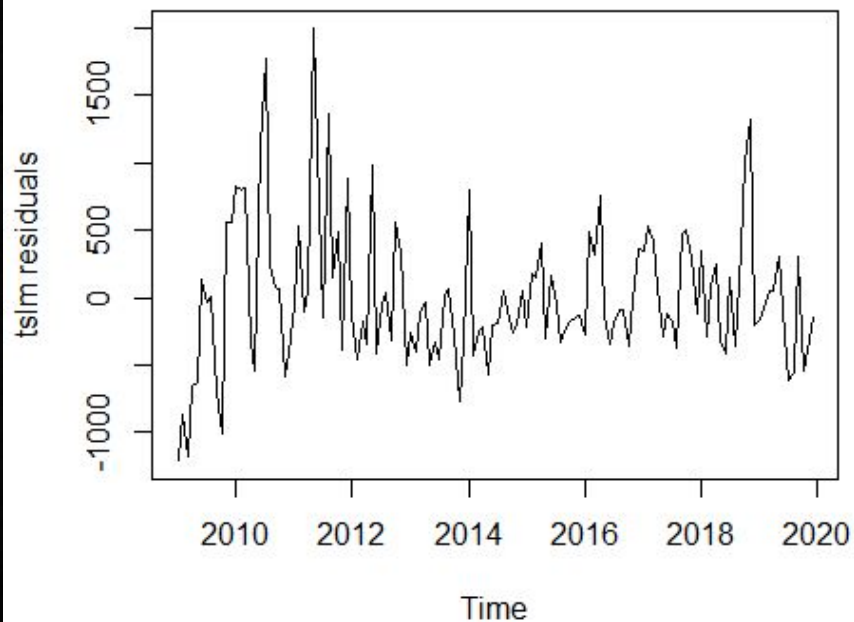
1. Time Series Linear Model

Time Series Linear Model

`sales ~ trend + season`

Predictors	coefficient	p-value
Intercept	1393.614	5.34 e-12
Trend	10.802	3.56 e-14
August	-1391.09	2.64 e-08
October	780.442	1.11 e-03
November	838.216	4.84 e-04

Time Series Linear Model

DW = 0.9201, p-value = 2.431 e-09
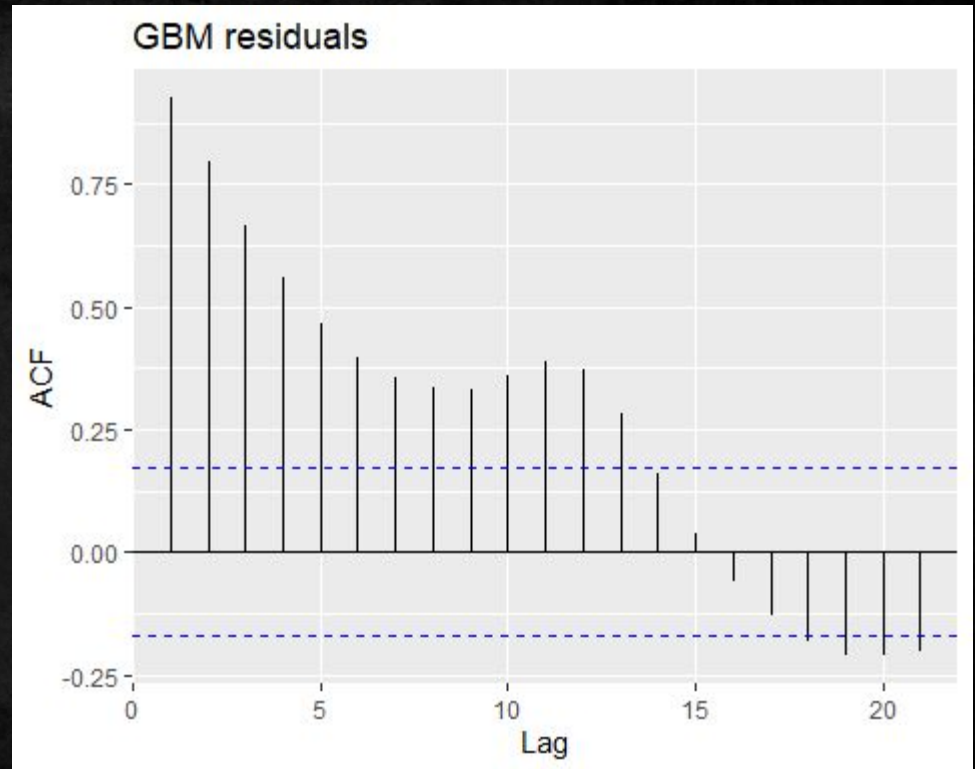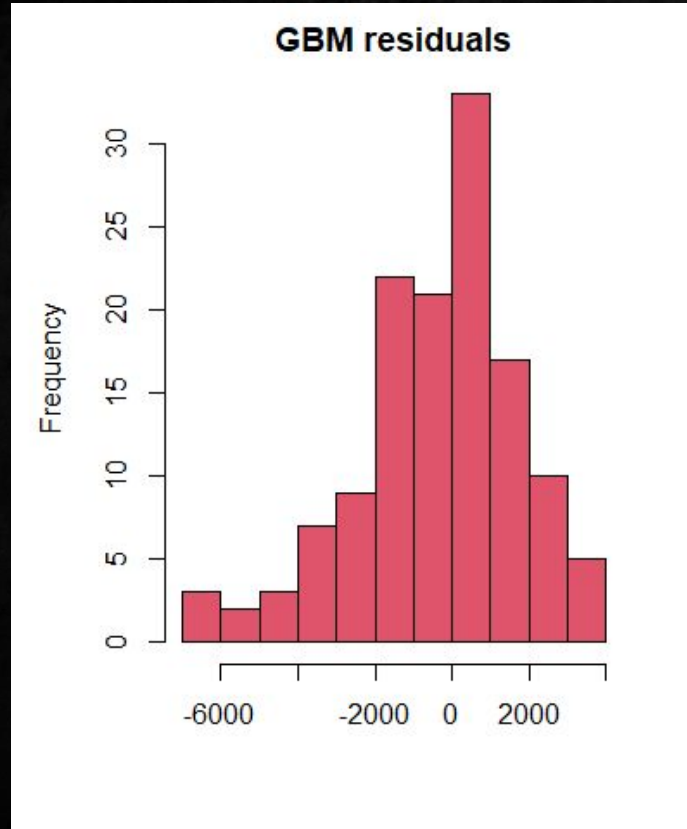
2. Generalized Bass Model

Generalized Bass Model

Predictors	coefficient	p-value
m	2.72 e+05	8.56 e-59
p	4.341 e-03	1.43 e-61
q	1.717 e-02	2.96 e-42
a_1	7.852 e+01	1.11 e-03
b_1	83.822 e-01	9.79 e-13
c_1	1.588 e-01	1.61 e-03

Fitted values
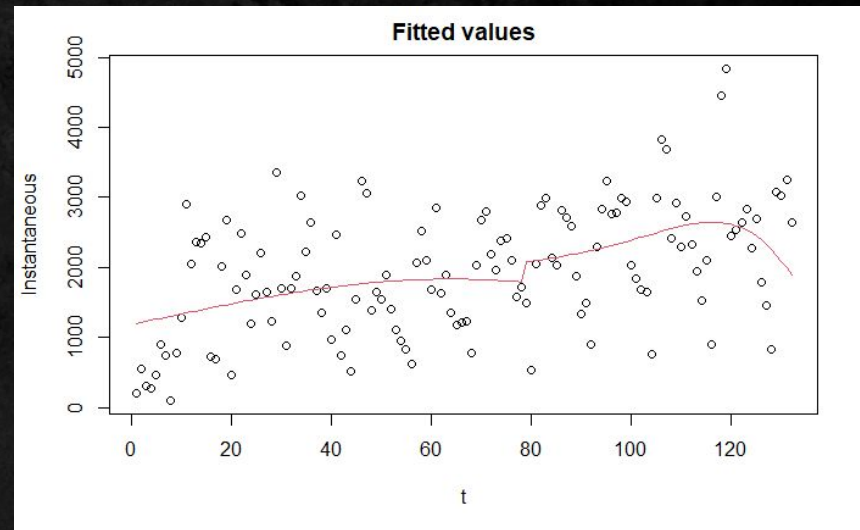
Generalized Bass Model

3. Seasonal ARIMA Model

Seasonal ARIMA Model

Forecasts from ARIMA(2,1,2)(1,1,1)[12]
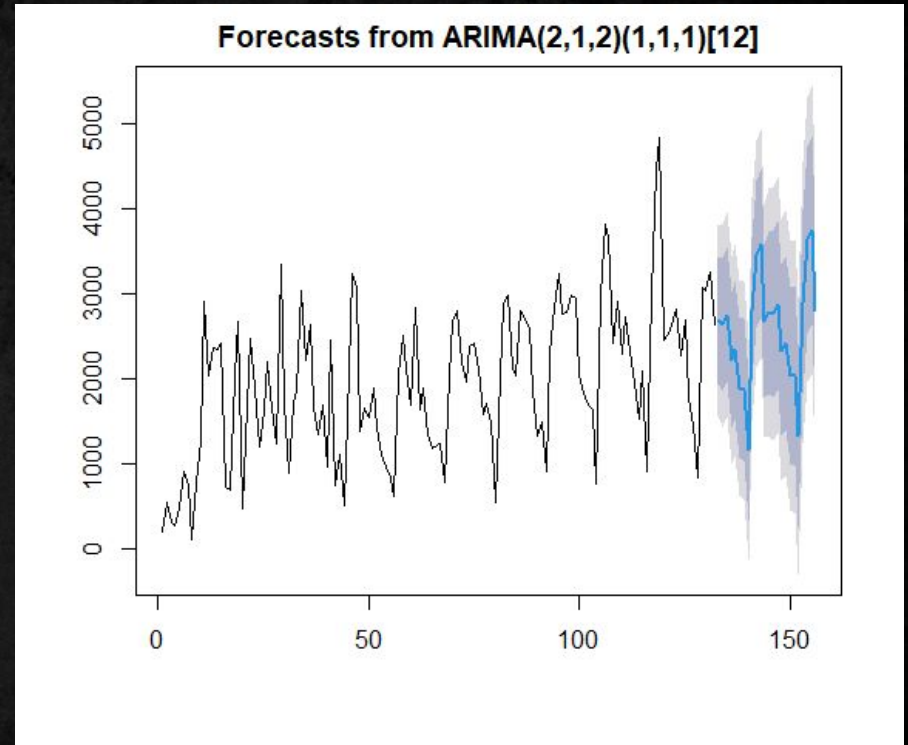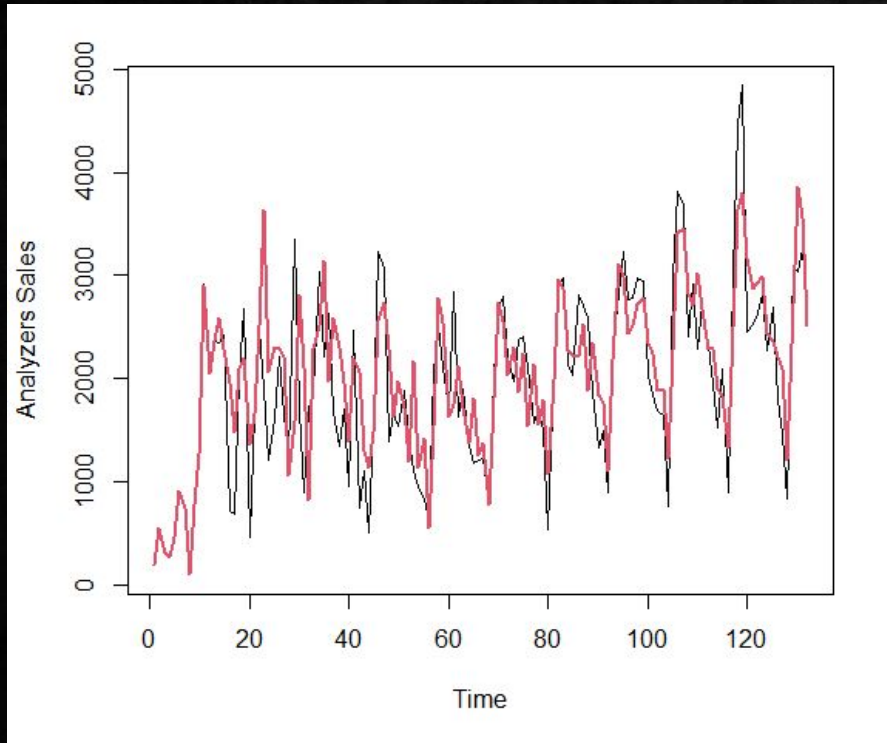
4. Combustion Analyzers summary

Combustion Analyzers summary

MODEL	RMSE	ADJR^2	MAPE	AIC
tslm	519.343	0.631	31.80	2053.277
Bass Model	2465.109	0.998	61.61	NA
Generalized BM	2138.139	0.999	60.53	NA
Seasonal ARIMA	527.611	NA	25.823	1870.635

Gas Security
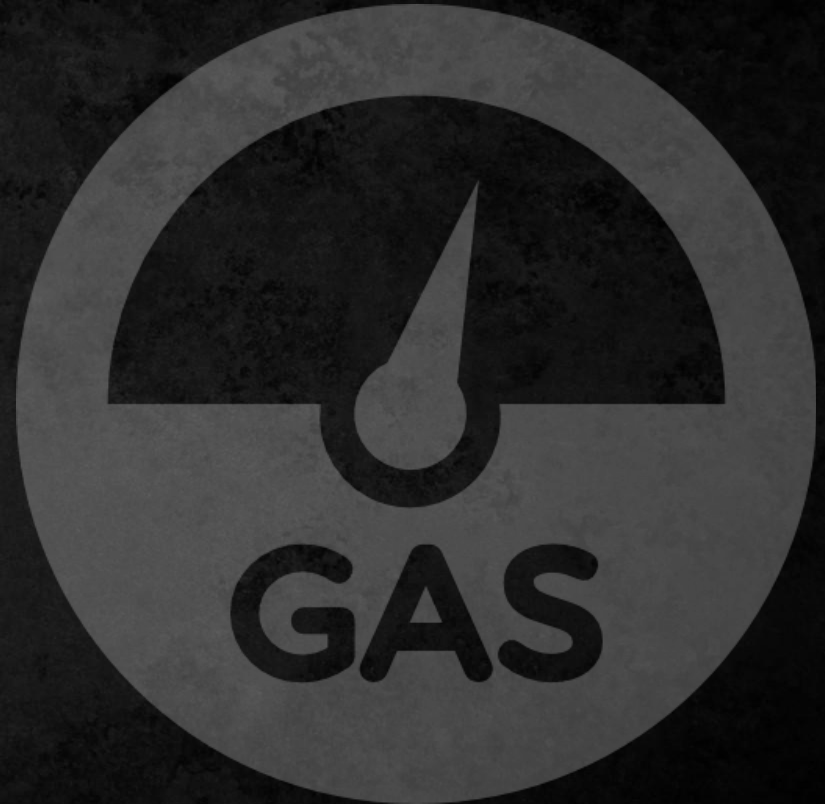
What about this sector?

Safety at home, in the office and in the company is very important.
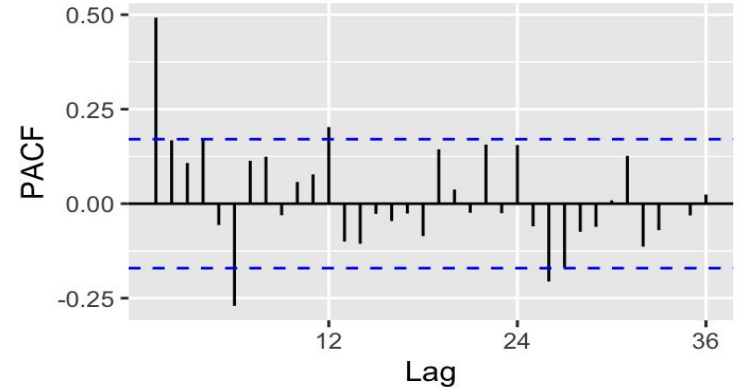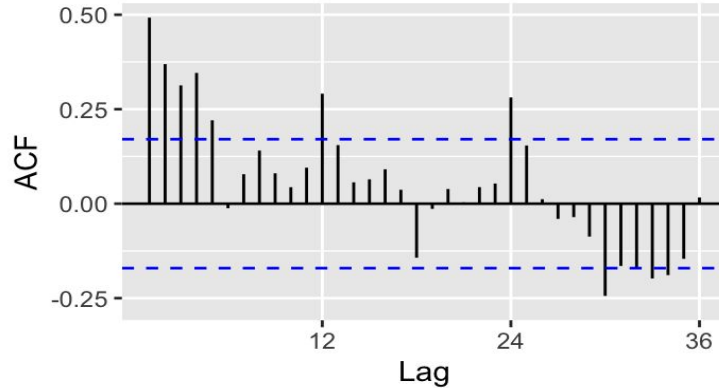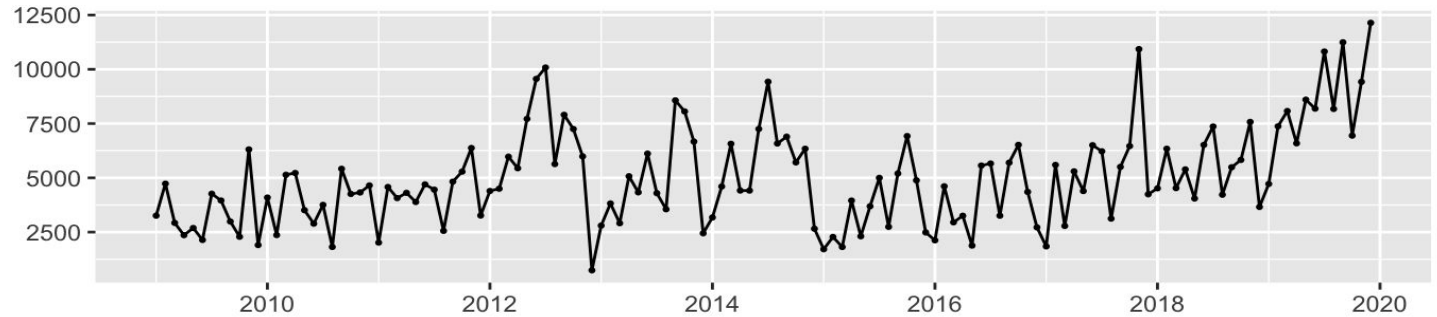Innovative tools are used on the safety of gas user systems.

- household gas detector
- industrial gas detector
- gas safety accessories
- gas solenoid valves

Gas Security Time Series

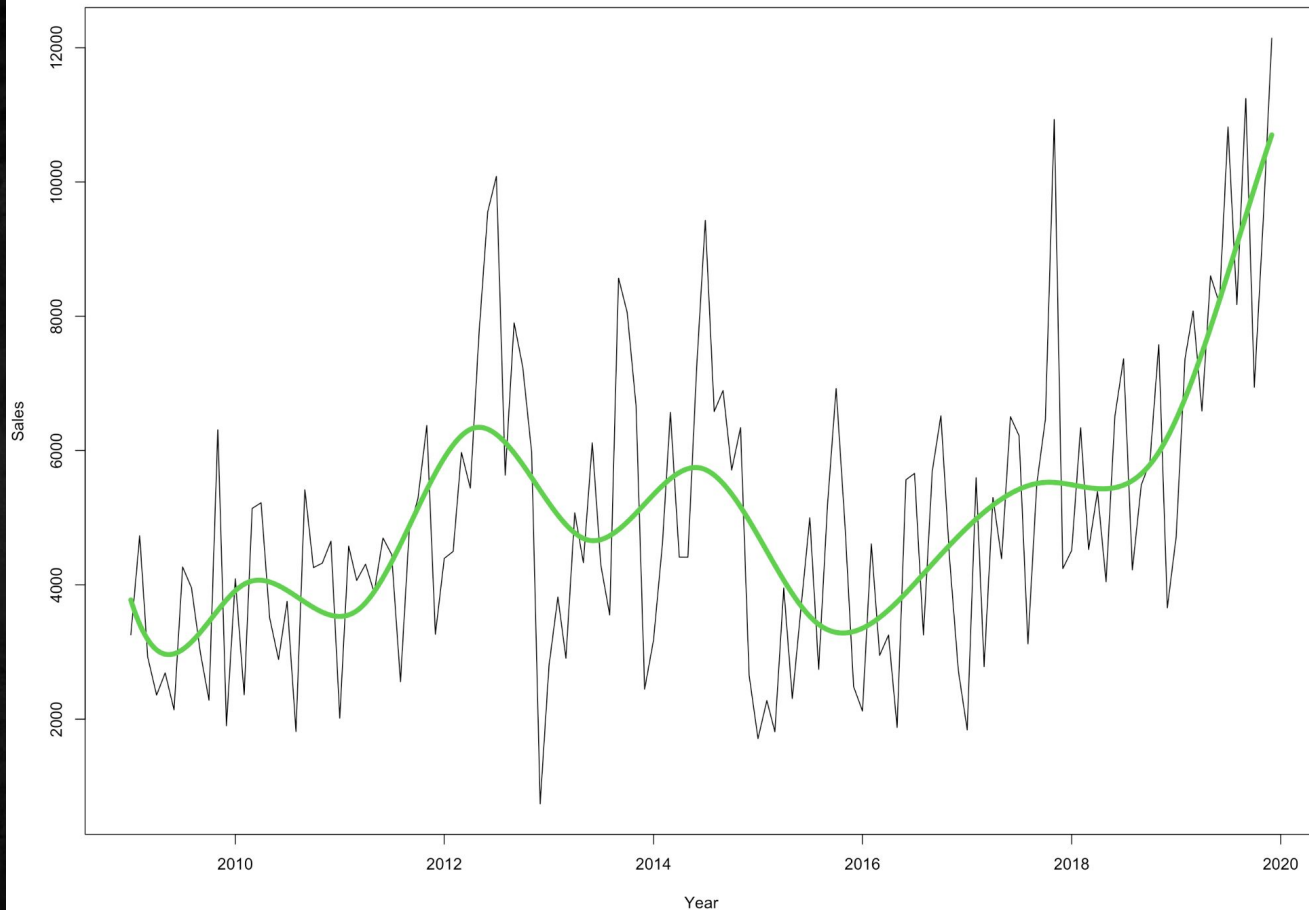- Growing trend over the long term

- No clear seasonal pattern

Regression Spline

degree = 3

length = 11

internal knots = 2

The sales are described well by a cubic interpolation curve. Highlighting the non-linear relationship.

1. Time Series Linear Model

What did we do?

- Arima

p = 1
d = 1
q = 1

	ar1	**ma1**
coef.	0.2502	-0.8338
s.e.	0.1214	0.0771

- Arima with seasonality

p = 1   P = 1
d = 1   D = 1
q = 1   Q = 1

	ar1	**ma1**	**sar1**	**sma1**
coef.	0.2285	-0.7491	0.0198	-0.995
s.e.	0.1916	0.1493	0.1151	0.2270

Time Series Linear Model

gas_sales ~ trend + season

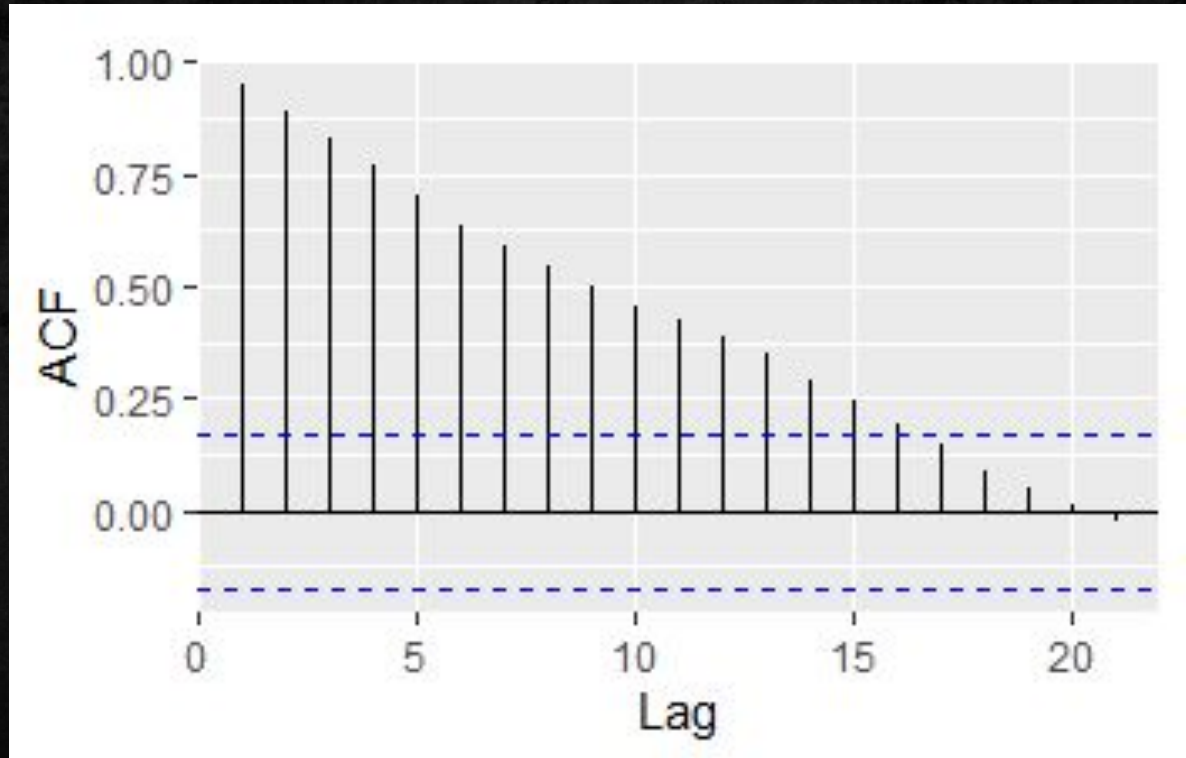| Predictors	coefficient	p-value
Intercept	1787.381	0.003765
Trend	22.30	4.36 e-07
July	3203.079	6.86 e-05
September	3011.712	0.000173
October	2606.301	0.001064
November	3279.799	4.77 e-05

DW = 0.77212
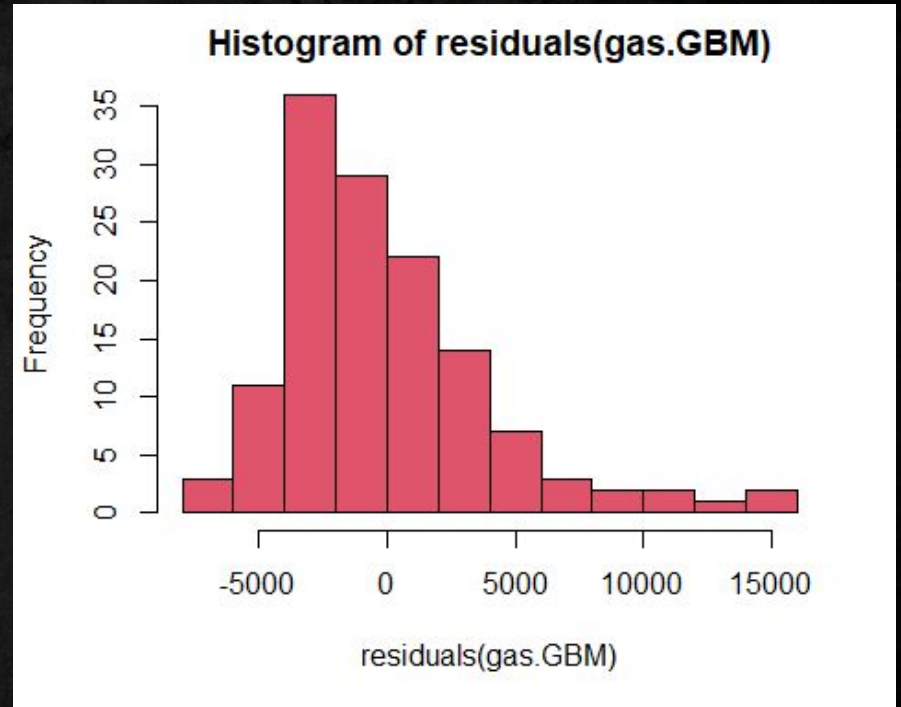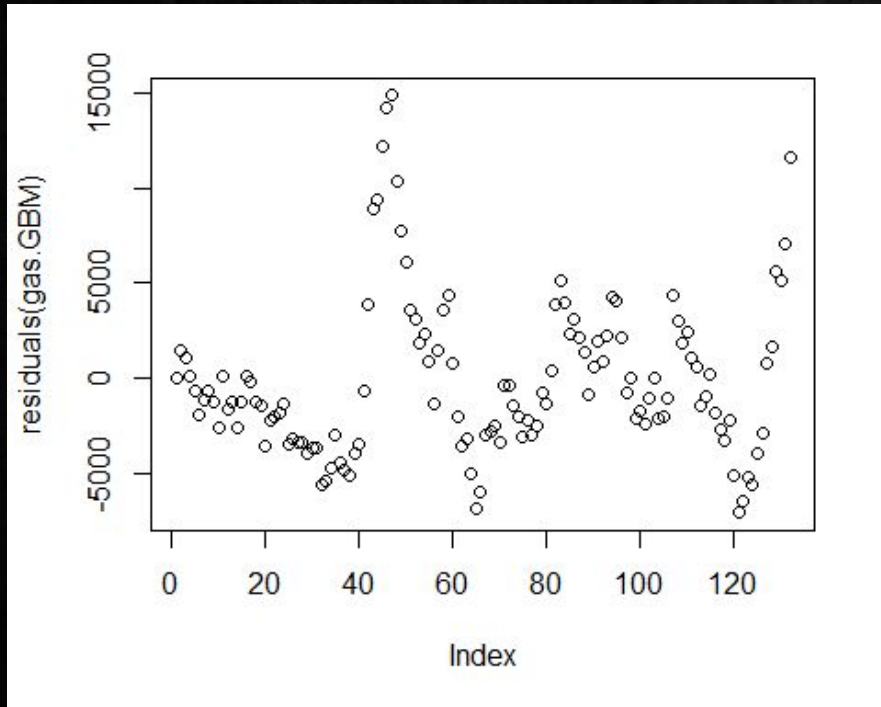
p-value = 1.17 e-11

2. Bass Model

3. Generalized Bass Model

Residuals Bass Model

Residuals show a decreasing pattern.
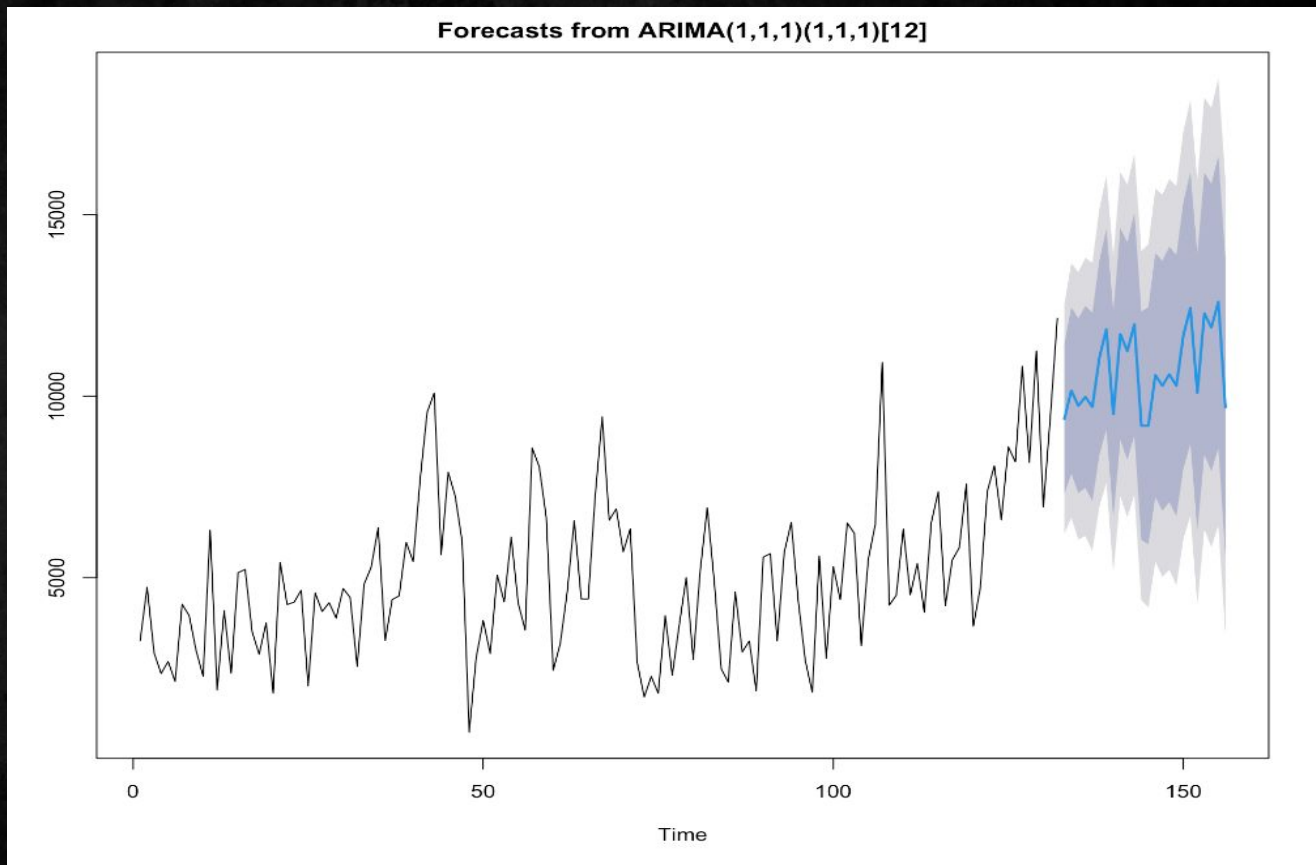
Residuals Generalized Bass Model

4. ARIMA Models

Forecast of Seasonal ARIMA Model

ARIMA parameters:

- Order = (1, 1, 1)
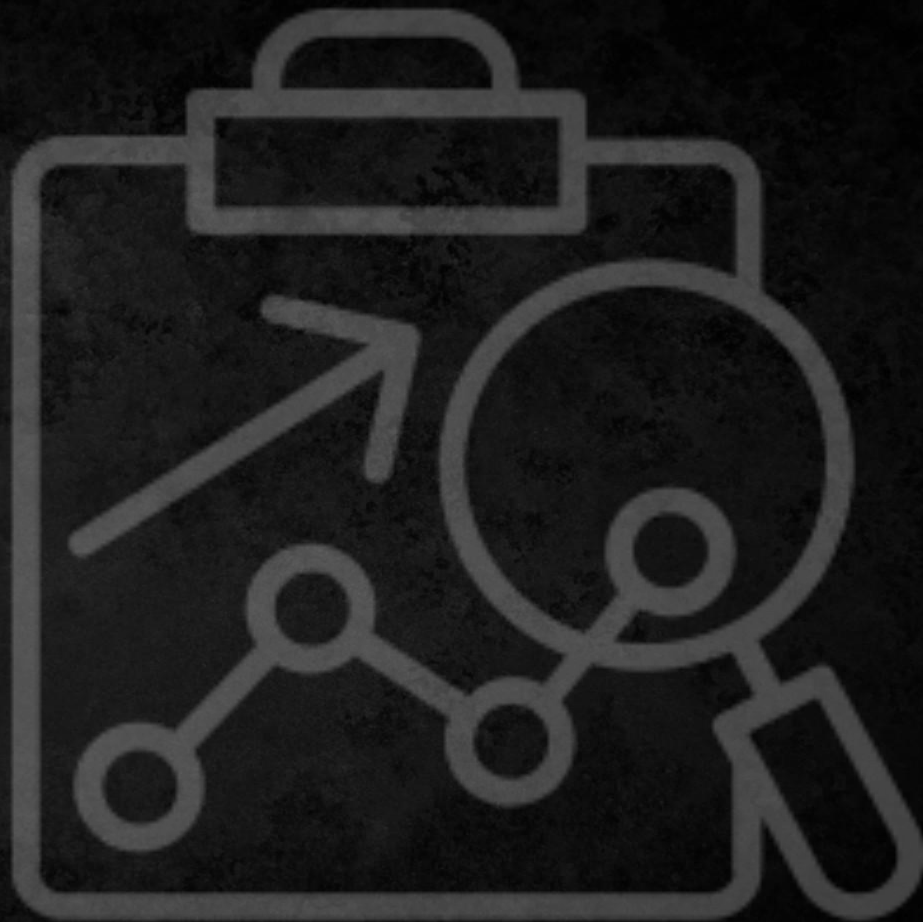- Seasonal = (1, 1, 1)
- Period = 12

Trend will gradually grow

Forecasts from ARIMA(1,1,1)(1,1,1)[12]

5. Gas Security summary

Gas Security summary

MODEL	RMSE	ADJR^2	MAPE	AIC
tslm	1727.53	0.338	33.59	2370.57
Bass Model	10752.50	0.997	41.86	NA
Generalized BM	4116.43	0.999	36.30	NA
ARIMA	1775.66	NA	37.88	2348.23
Seasonal ARIMA	1438.281	NA	26.87	2119.34

Further analysis

Sales & Gas price Italy

Gas price dataset

- Introduction of a new dataset with italian natural gas prices

- Prices given in dollars

- Correlated with Thermoregulation and Gas security sectors sales.

GAS PRICE

Natural gas is a gas mixture, composed almost by methane, to which are added very small quantities of other gases.
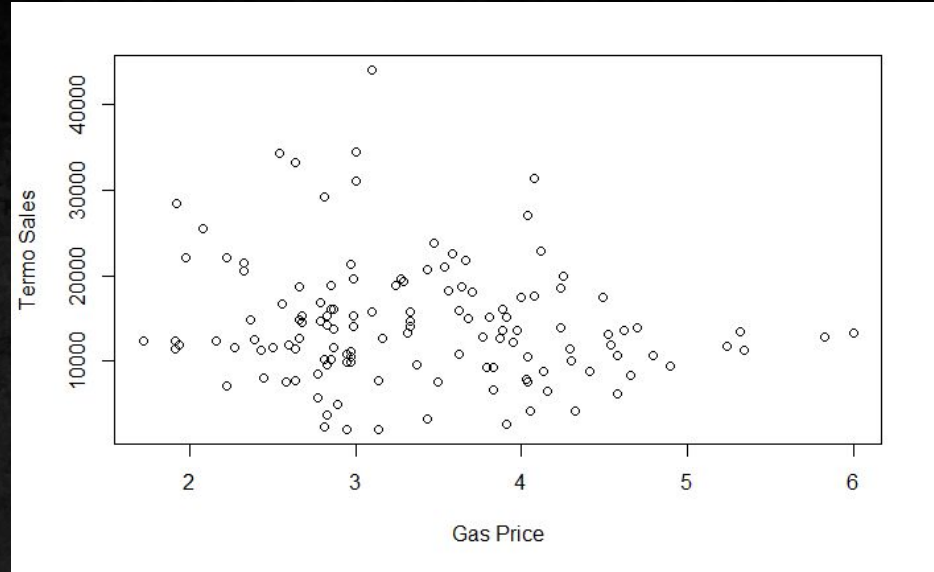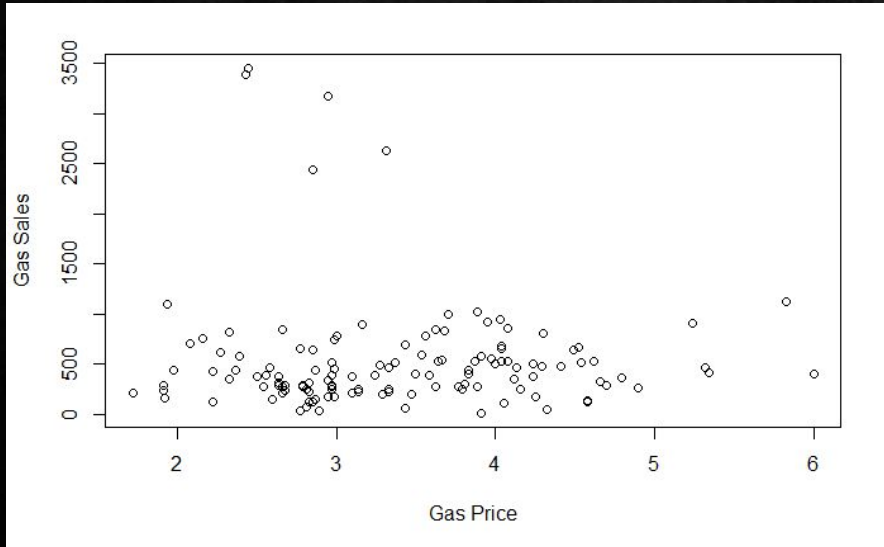It is generated by geological formations and is connected to coke or oil fields.
Moreover, it is less polluting than other fossil fuels.

Gas is a source of energy that has spread to Italy and Europe since the 1950s thanks to its cheapness and the abundance of its reserves.
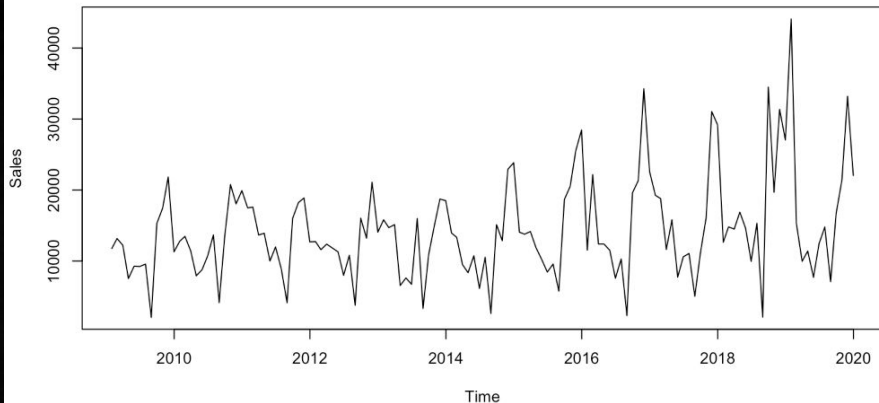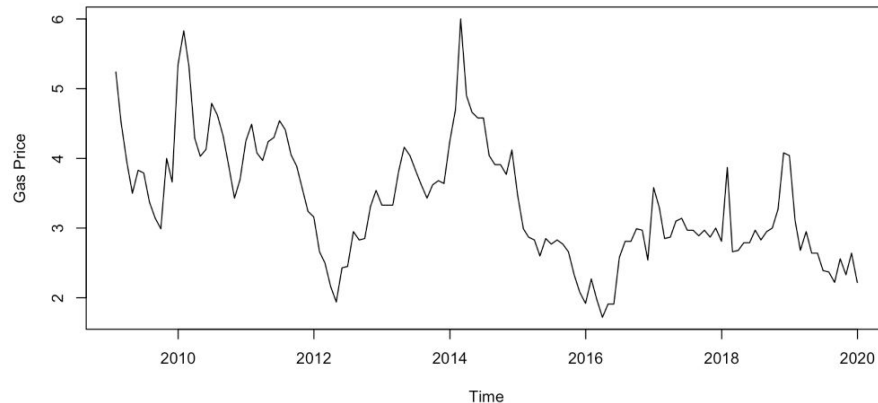
GAS PRICE AND THERMO SALES

GAS PRICE AND GAS SECURITY SALES
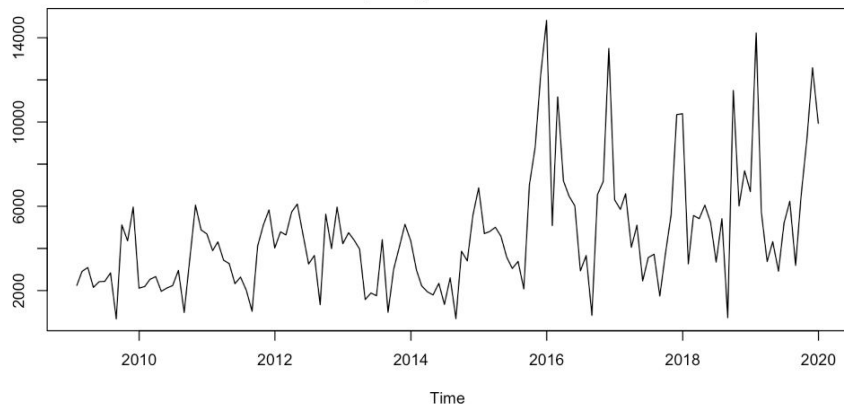
Relationship between gas price & thermo sales
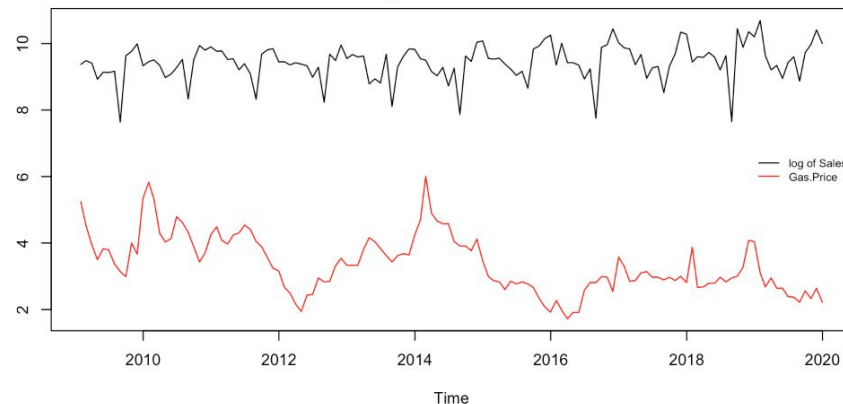
Time Series Linear Regression Model
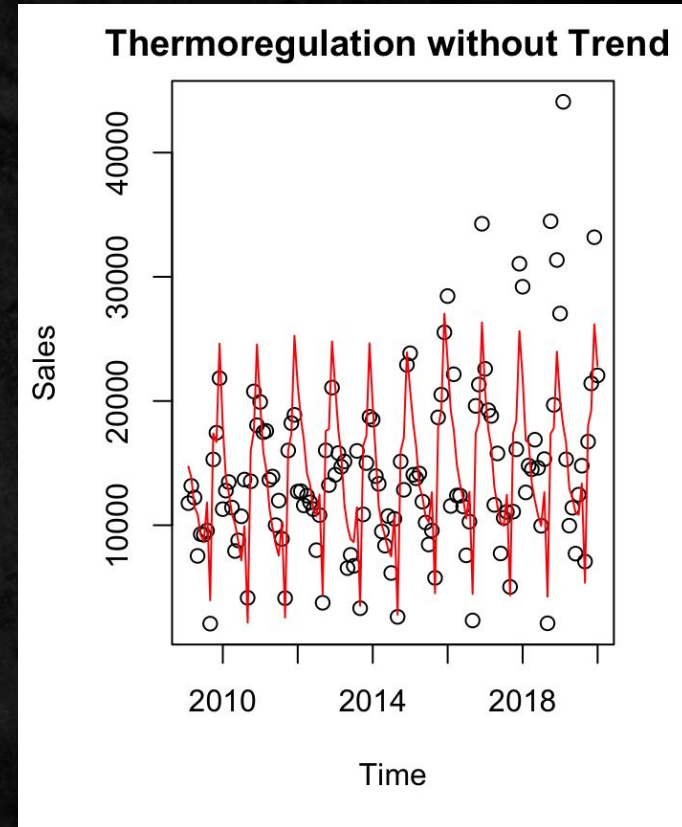
➤ with trend

➤ without trend

Gas price
is significant
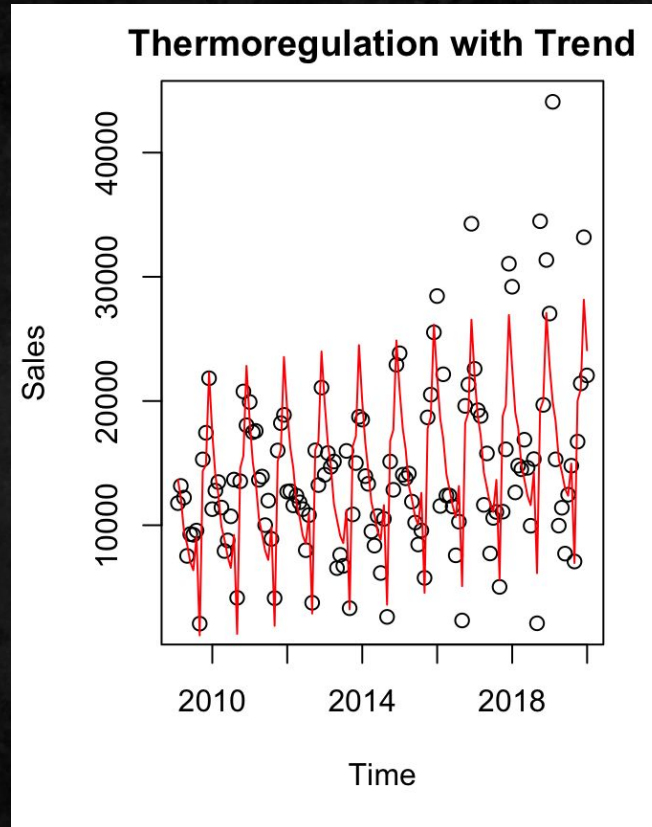
Sales ~ GasPrice + season

Predictors	coefficient	p-value
Intercept	22723.1	2e-16
Gas Price	-1527.4	1.259 e-03
April	-6468	9.96 e-04
May	-7681.7	1.02 e-04
June	-8237.3	3.24 e-05
August	-13968.1	3.84 e-11
November	7480.8	1.53 e-04

DW = 1.8439

p-value = 0.34

Train & Test

Train & Test

Training Set: from 2009 to 2017
→ ~80% sales
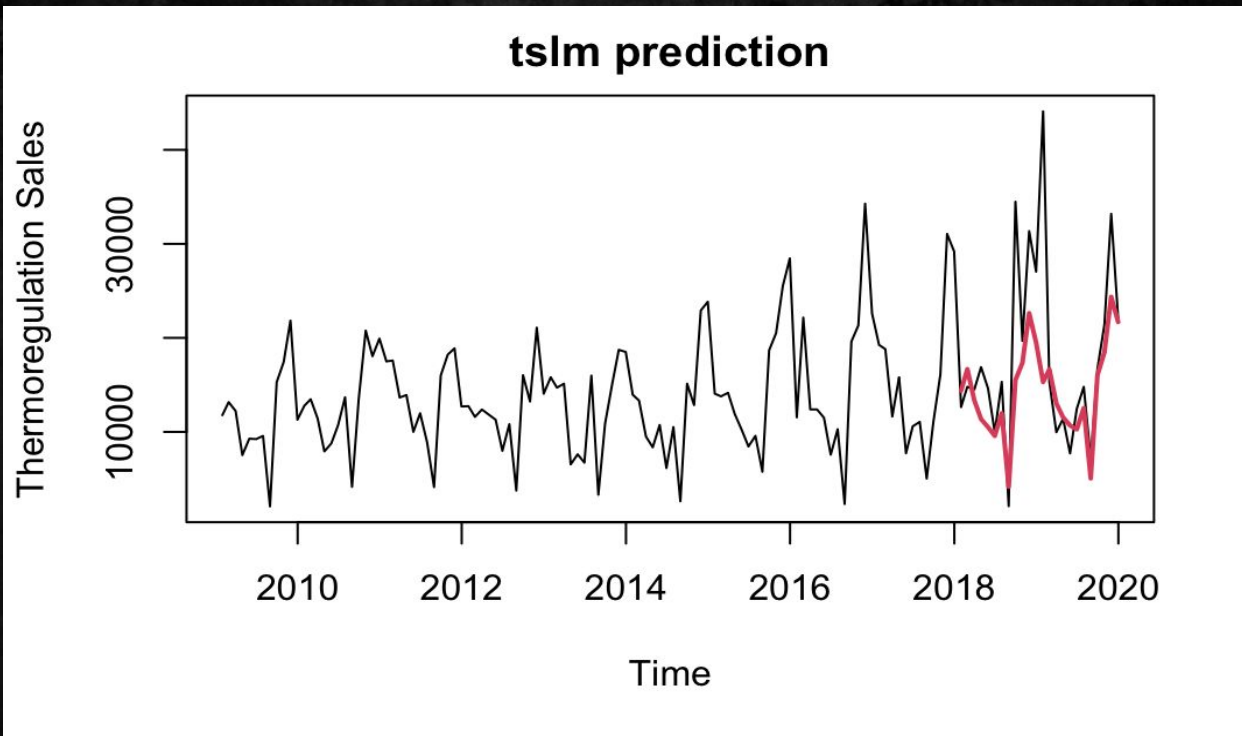
Test Set: 2018 - 2019
→ ~20% sales

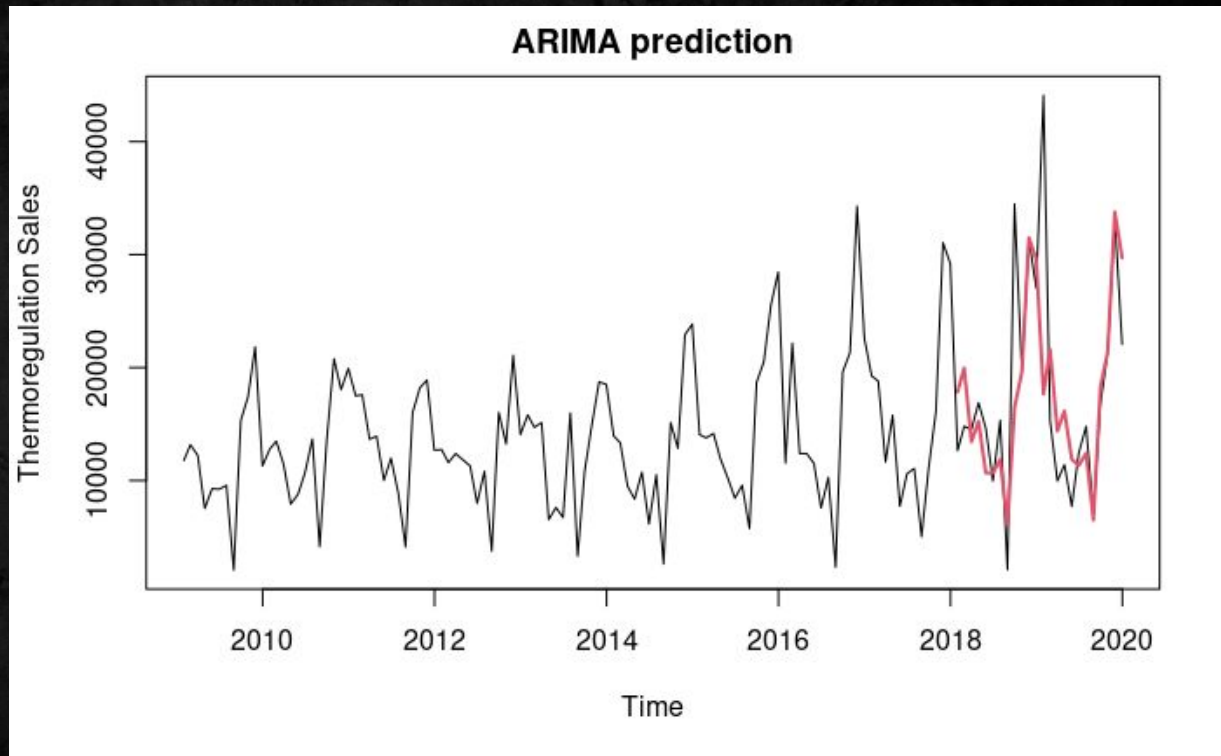**Models**: `tslm` and `ARIMA`

Train & Test - tslm

Sales ~ GasPrice + season

tslm prediction

Train & Test - ARIMA

ARIMA parameters:

- order = (2, 1, 2)

- seasonal = (2, 1, 2)

- period = 12

Train & Test

MODEL	TRAIN RMSE	TEST RMSE	TRAIN MAPE	TEST MAPE
tslm	3035.16	7955.65	18.69	24.751
ARIMA	3045.01	7354.36	18.82	30.05

- `tslm` errors are lower than `ARIMA` errors
- overfitting problem

Conclusions

- Seasonality and Trends both contribute in the predictions, but the former is more relevant.

- Sales are concentrated in cold months but with slightly different effect of seasonality for each sector.

- ARIMA and tslm are the best models and they are able to capture different aspects of the data, but the last one has higher statistical explanatory power.

- Gas Price shows relevant contribution in the forecast.

- The given data was limited by few useful features. Therefore, further analysis should exploit the usage of other predictors like Economic indicators (e.g. Gas Price, Inflation, etc) or climate temperature. Unfortunately, open data are not easily available for the considered countries