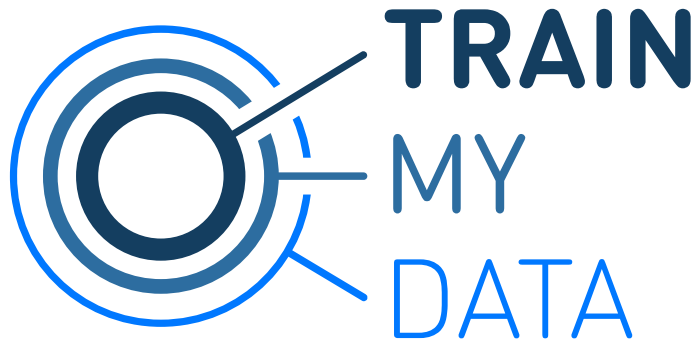


# Ascott Group Sales Forecasting

On TrainMyData.com platform



# Ascott group

## WE CREATE INTERIORS

We provide our customers with quality products and high level of service in Russia and the CIS countries.



### Wallpaper glue KLEO

One of the leaders of wallpaper glue market in Russia

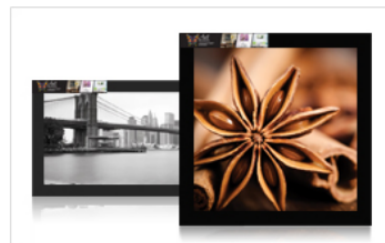
[kleo.com](http://kleo.com)



### Decorative stickers Decoretto

Bright, colorful vinyl stickers for every taste, perfect for decoration of your home or office.

[decoretto.ru](http://decoretto.ru)



### Decoretto Art and photo wallpapers

Decorative canvas and glass art

[decoretto.ru/art](http://decoretto.ru/art)

Vivid collection of wallpapers

[decoretto.ru/fotooboi](http://decoretto.ru/fotooboi)

Turn-key project for wall decor

[b2b.decoretto.ru](http://b2b.decoretto.ru)



### Goods for repair

Accessory materials for repair and final finishing

[kleo.com/kleo-pro](http://kleo.com/kleo-pro)

[www.ascottgroup.eu](http://www.ascottgroup.eu)

# Competition Platform

- New in town
- Fast web site and checker
- Excellent support
- Telegram Channel for discussions
- Opened private score for all submissions
- Promised to turn this competition into a playground

# Data

	wk	N wk	idFilial	KanalDB	idSubGrp	value
id						
0	201401	1	9	2	3	3560.0
183	201401	1	9	2	4	4608.0
366	201401	1	9	2	5	2280.0
14144	201402	2	17	2	3	24920.0
4163	201402	2	6	2	3	-1780.0
8447	201402	2	11	2	3	10680.0
7905	201402	2	10	1	6	3672.0
13432	201402	2	15	1	3	10680.0
13295	201402	2	15	2	1	3460.0

Train 183    9        2        6

train\_set\_weeks.csv

	wk	N wk	idFilial	KanalDB	idSubGrp
id					
0	201729	184	9	2	3
1	201730	185	9	2	3
2	201731	186	9	2	3
3	201732	187	9	2	3
4	201733	188	9	2	3
5	201734	189	9	2	3
6	201735	190	9	2	3
7	201736	191	9	2	3
8	201737	192	9	2	3

Test 11

test\_set\_weeks.csv

108 combinations

# Data

	Date	wk	N wk	idFilial	idBP	idSlp	KanalDB	City	idTypeItem	idGrp	idSubGrp	idItem	value
id													
0	2014-01-03	201401	1	9	3936	370	2	Астрахань	3	5	3	4825	3560.0
5433	2014-01-03	201401	1	9	3936	370	2	Астрахань	3	5	5	4826	2280.0
2250	2014-01-03	201401	1	9	3936	370	2	Астрахань	3	5	4	14	4608.0
3	2014-01-09	201402	2	9	3492	95	2	Краснодар	3	5	3	4827	3560.0
191727	2014-01-09	201402	2	11	67334	81	1	Нижний Новгород	3	5	3	4824	10680.0
191728	2014-01-09	201402	2	11	17394	81	1	Нижний Новгород	3	5	3	4825	1780.0
191729	2014-01-09	201402	2	11	67334	81	1	Нижний Новгород	3	5	3	4825	5340.0

train\_set\_days.csv

# Data

	index	idBP	idFilial	visit_date	calc_share	visits_count	calls_count
id							
3165	119	69610	6	2015-08-03	NaN	1	0
3164	106	68386	6	2015-08-03	NaN	1	0
3166	110	68760	6	2015-08-03	NaN	1	0
3167	106	68386	6	2015-08-03	NaN	1	0
3168	119	69610	6	2015-08-03	NaN	1	0

train\_set\_sfa.csv

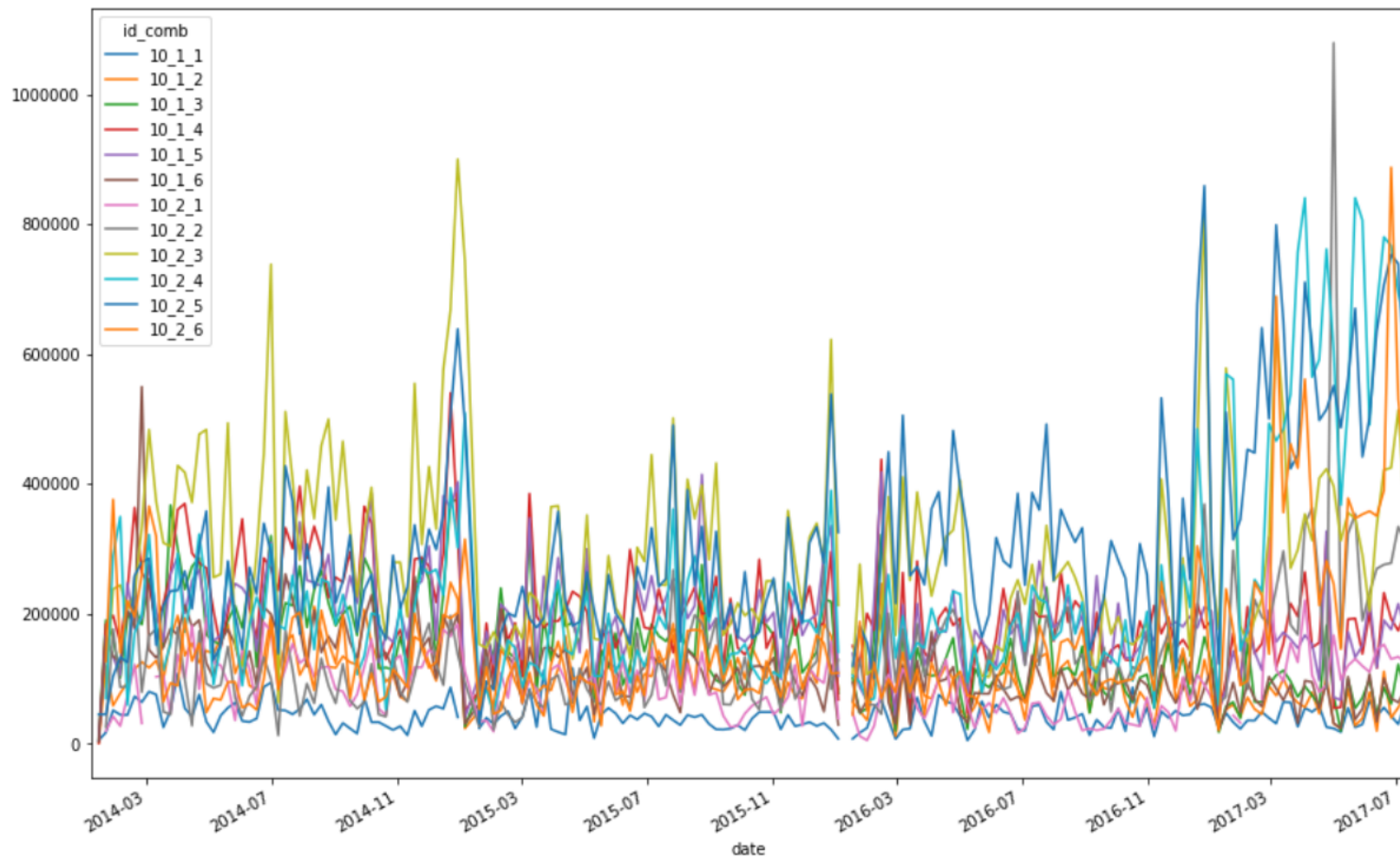
	idBP	size	size_value
0	75	Средний	100-199 м2
1	132	Большой	500-1999 м2
2	146	Малый	10-99 м2
3	150	Малый	10-99 м2
4	163	Малый	10-99 м2

info\_business\_points.csv

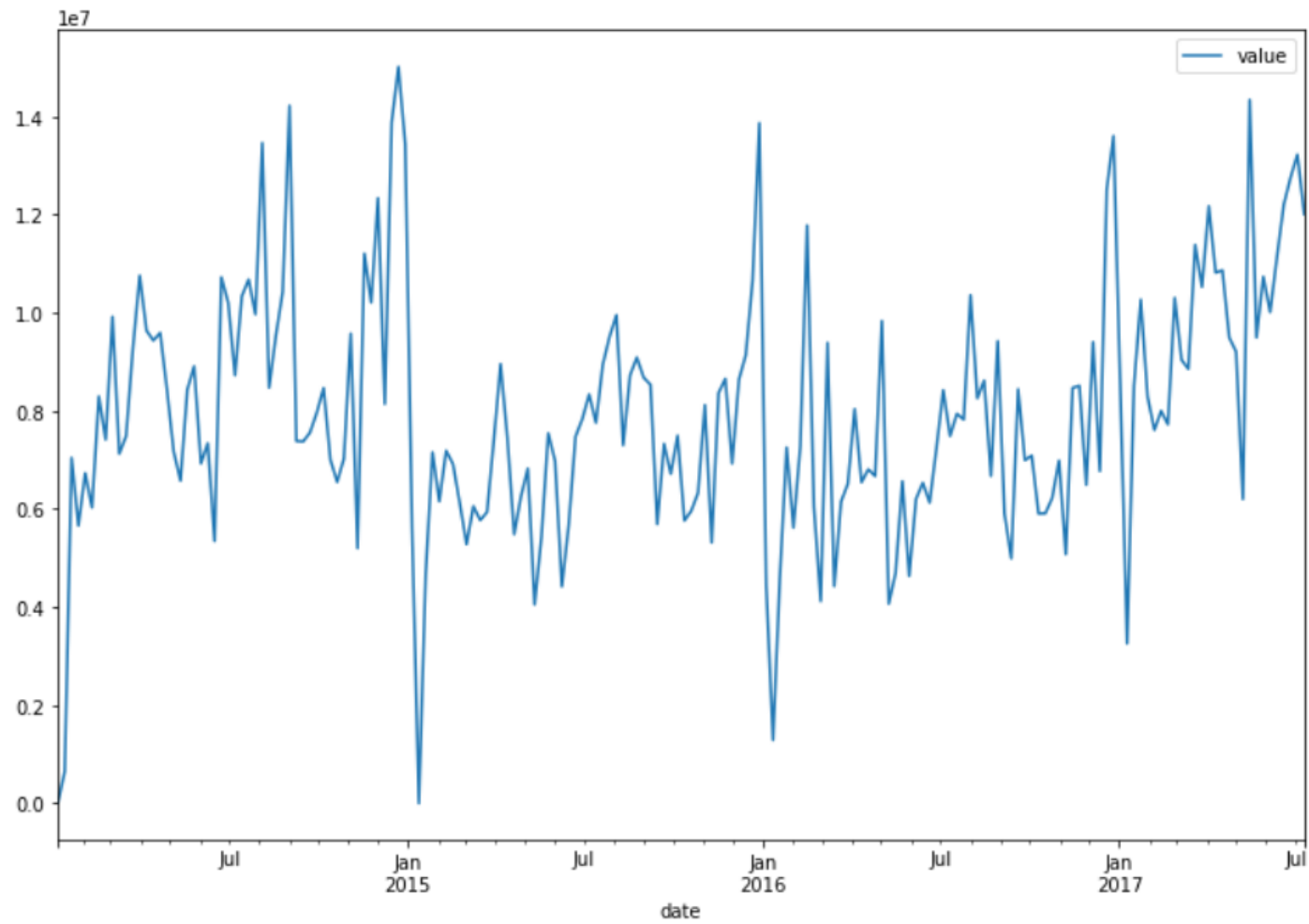
	idGrp	idSubGrp	idTypeItem
0	5	2	1
1	5	2	2
2	5	2	3
3	5	3	1
4	5	3	2

info\_groups.csv

# Plot of 10 series



# Total Weekly Sales





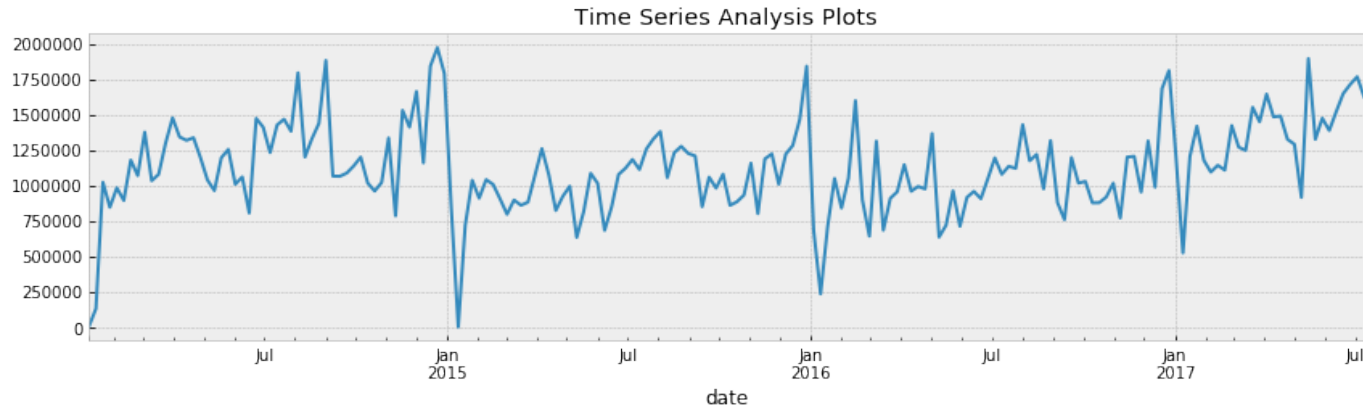
# Your choice of a model?

- Neural Networks
- Gradient boosting
  - Xgboost, lightgbm, catboost
- Enterprise forecasting tools
  - SAS, Facebook Prophet, ...
- Linear models
- Classic Time Series models
  - EMA, Holt-Winters, ARIMA, ...
- Ad hoc (your own model)
- Can't decide before doing some more EDA.

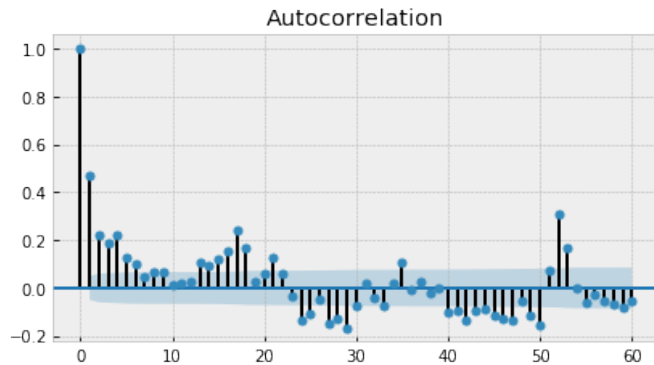
# (S)ARIMA

- Autoregressive – fitted using lags of target variable.
- We need lags to bring a TS to stationary
  - Mean, dispersions, covariance don't depend on time
- Can account for seasonality
- Should not be overfitting
- Parameters
  - $(p,d,q)$  order of the model for the number of AR parameters, differences, and MA parameters.
    - $d$  must be an integer indicating the integration order of the process
    - $p$  and  $q$  are integers indicating the AR and MA orders
  - $(P,D,Q)$  – same for seasonality component
  - $S$  – season length

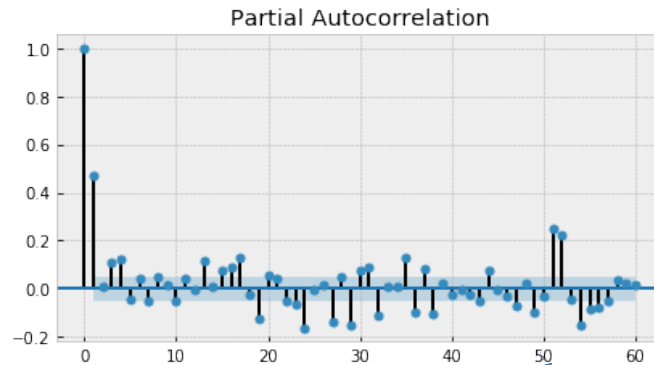
# ARIMA - analysis



Критерий Дики-Фуллера:  $p=0.000924$

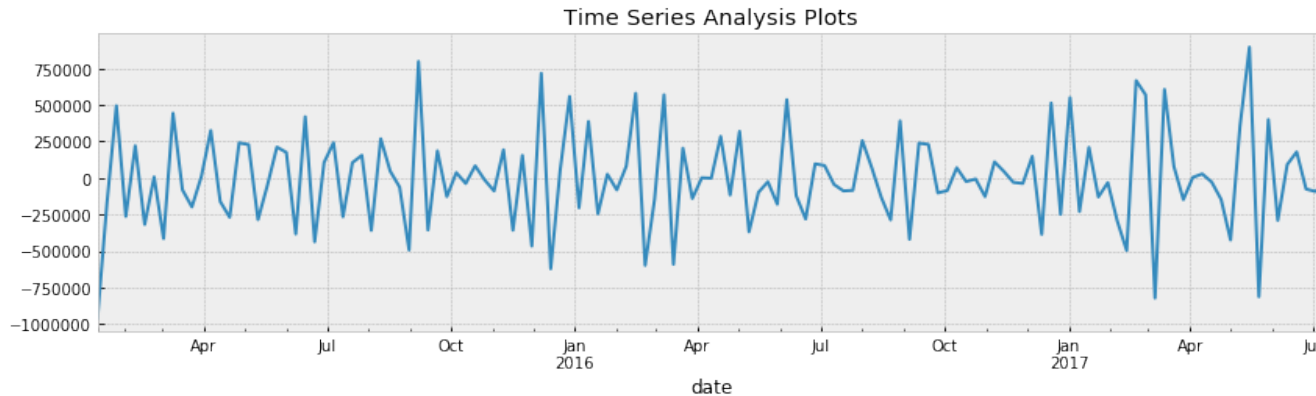


Depends on time!



TS requires differencing

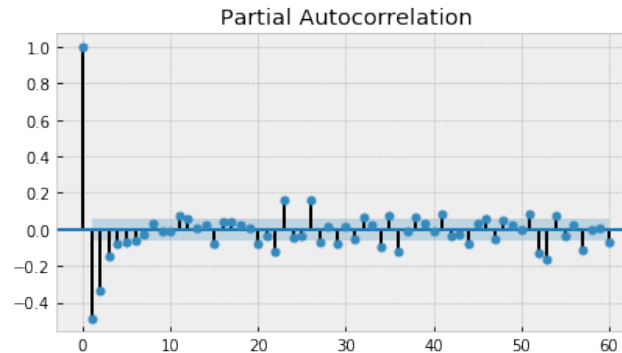
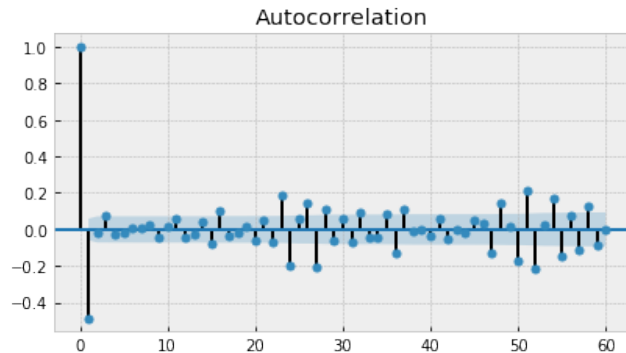
# ARIMA – analysis



Критерий Дики-Фуллера:  $p=0.000000$

Transformations:

- box-cox
- season difference ( $D=1$ )
- single difference ( $d=1$ )



Still need to chose  $p, q, P, Q$   
and season length  
In order to fit ARIMA model

Looks much better!

# How many weeks in a year?

- ?
- ?
- ?

# Model parameter search

- Brute force over parameter grid
  - $D=1$ ,  $D=1$
  - $P=[0,3]$ ,  $p=[0,3]$
  - $Q=[0,3]$ ,  $p=[0,3]$
  - $s=51$  or  $52$
- Model selection - AIC (Akaike Information Criteria)
- Splitting weekly sums back to 108 series
  - Used average historic proportions for the last 17 weeks (found experimentally)
- Blending top models with 51,52 weeks
- Blending top-3 or top-4 models

# Use of exogenous data.

- Macrostatistics
  - Dollar, euro rate
  - Inflation
  - Payroll
  - Producer price index
  - ...
- Allows to find some correlation between sales and external data
- Can be used for stress-testing
  - If Rouble plunges, what will be our sales? (Hedge accordingly)

## 2<sup>nd</sup> and 3<sup>rd</sup> place solution

- Делал прогноз на 2017 суммы value по неделям по среднему значению между 2015-2016, добавив рост 2017 к 15-16 на комбинации делил по их средней доле в 2017 (2-28 неделя)
- я взяла сколько-то последних недель (20 кажется) - откинула top3 и tail 3 значений - взяла от этого среднее и потом все значения увеличила на 10%. Затем я сделала точно такой же прогноз для периода год назад - и посмотрела, где сильно ошибаюсь и скорректировала решение. (это было 5 место). Потом я выделила лучший месяц и увеличила там продажи и худший месяц - а там уменьшила. На сколько увеличивать-уменьшать смотрела по паблику.



# Further reading

- Mlcourse\_open, Time Series Analysis with Python
  - <https://habrahabr.ru/company/ods/blog/327242/>
  - [https://www.youtube.com/watch?v=nQjul-5\\_0\\_M](https://www.youtube.com/watch?v=nQjul-5_0_M)
  - [https://github.com/Yorko/mlcourse\\_open/tree/master/jupyter\\_russian/topic09\\_time\\_series](https://github.com/Yorko/mlcourse_open/tree/master/jupyter_russian/topic09_time_series)
- Hyndman and Athanasopoulos, Forecasting Principles and Practice, 2<sup>nd</sup> edition
  - <http://otexts.org/fpp2/>



Starting  
05.02.2018