AQUABLISS

STEFANO ROSSI 13 October 2017

CAPSTONE PROJECT

AQUABLISS

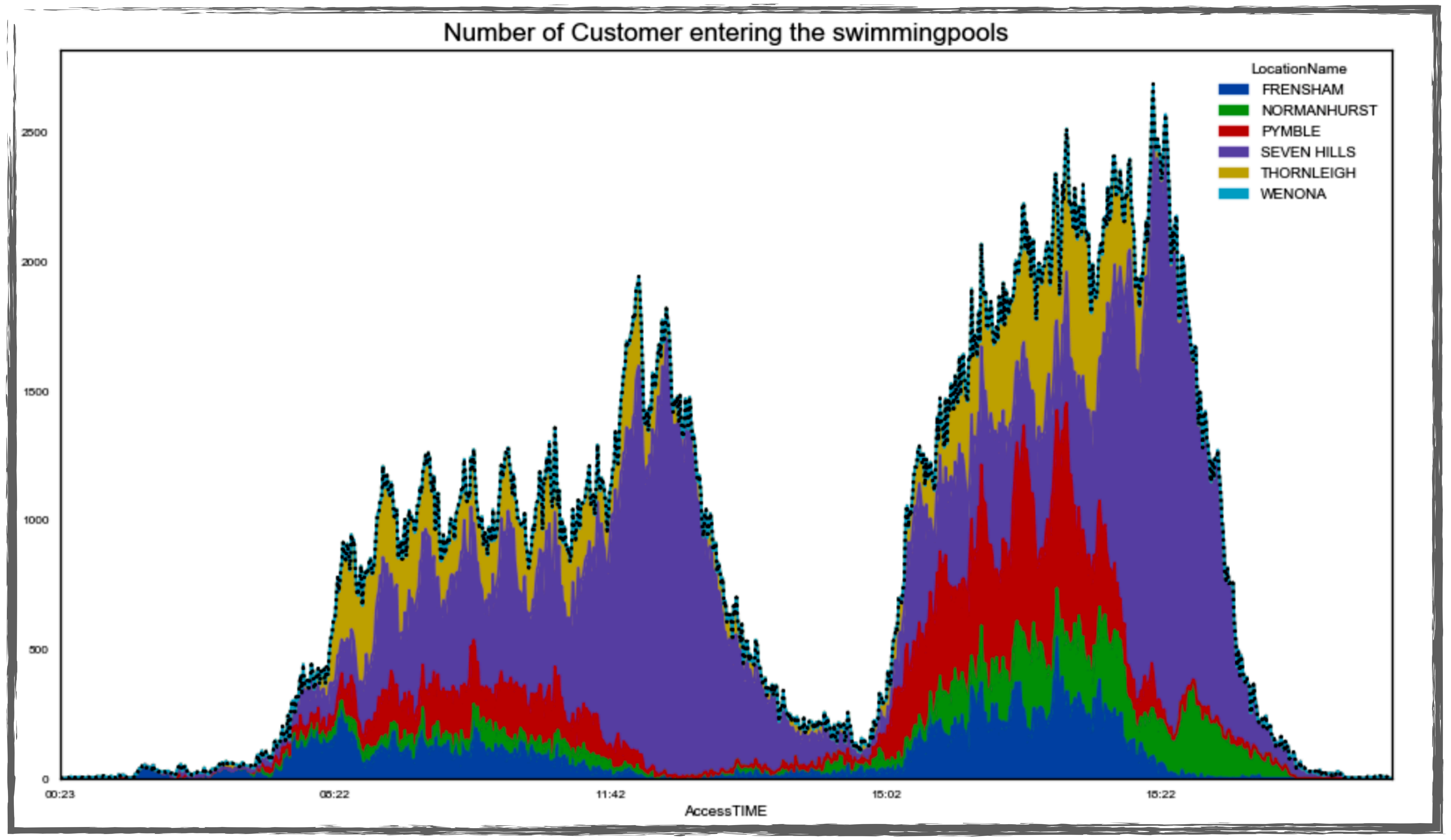
BUSINESS CASE

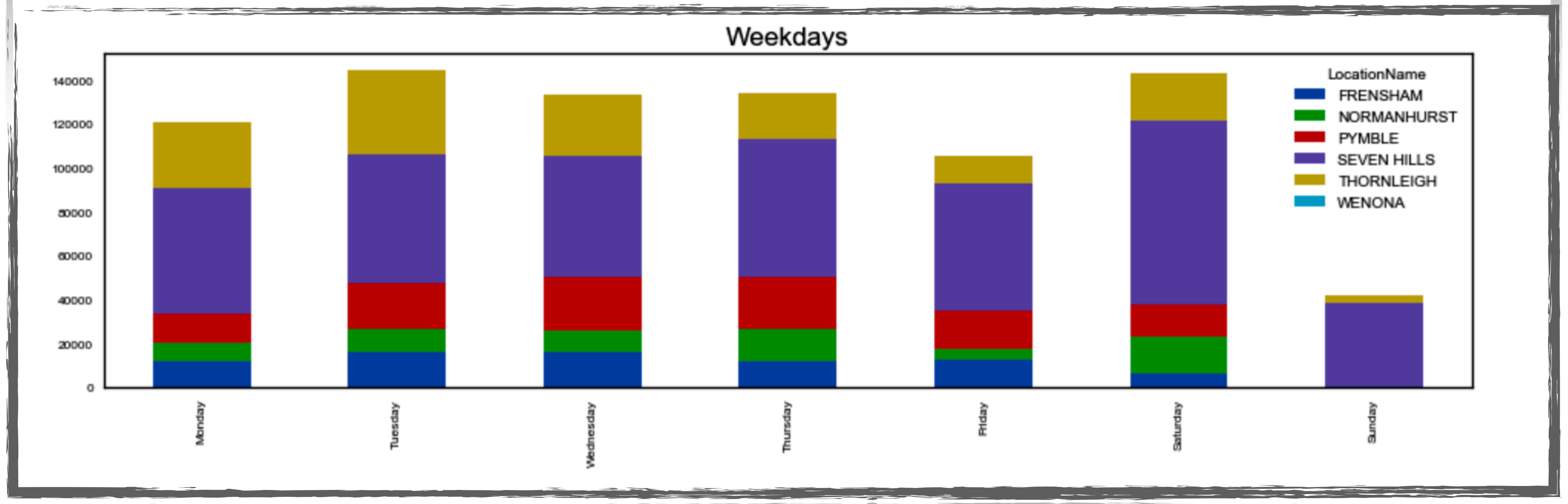
Is there seasonality in our swimming pool? Is it possible to predict customer a pool will have in advance?

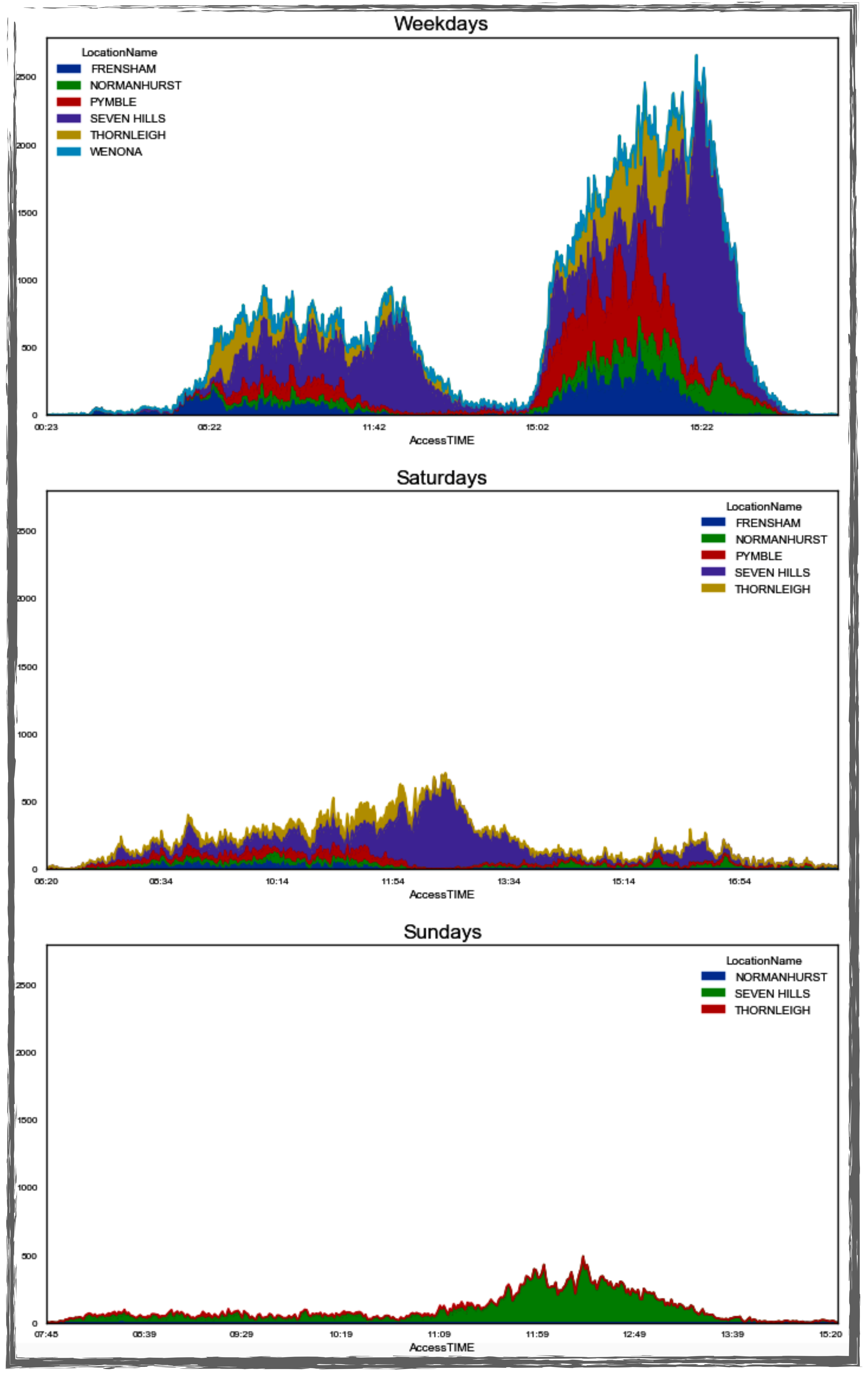
Time-series Analysis of foot-traffic

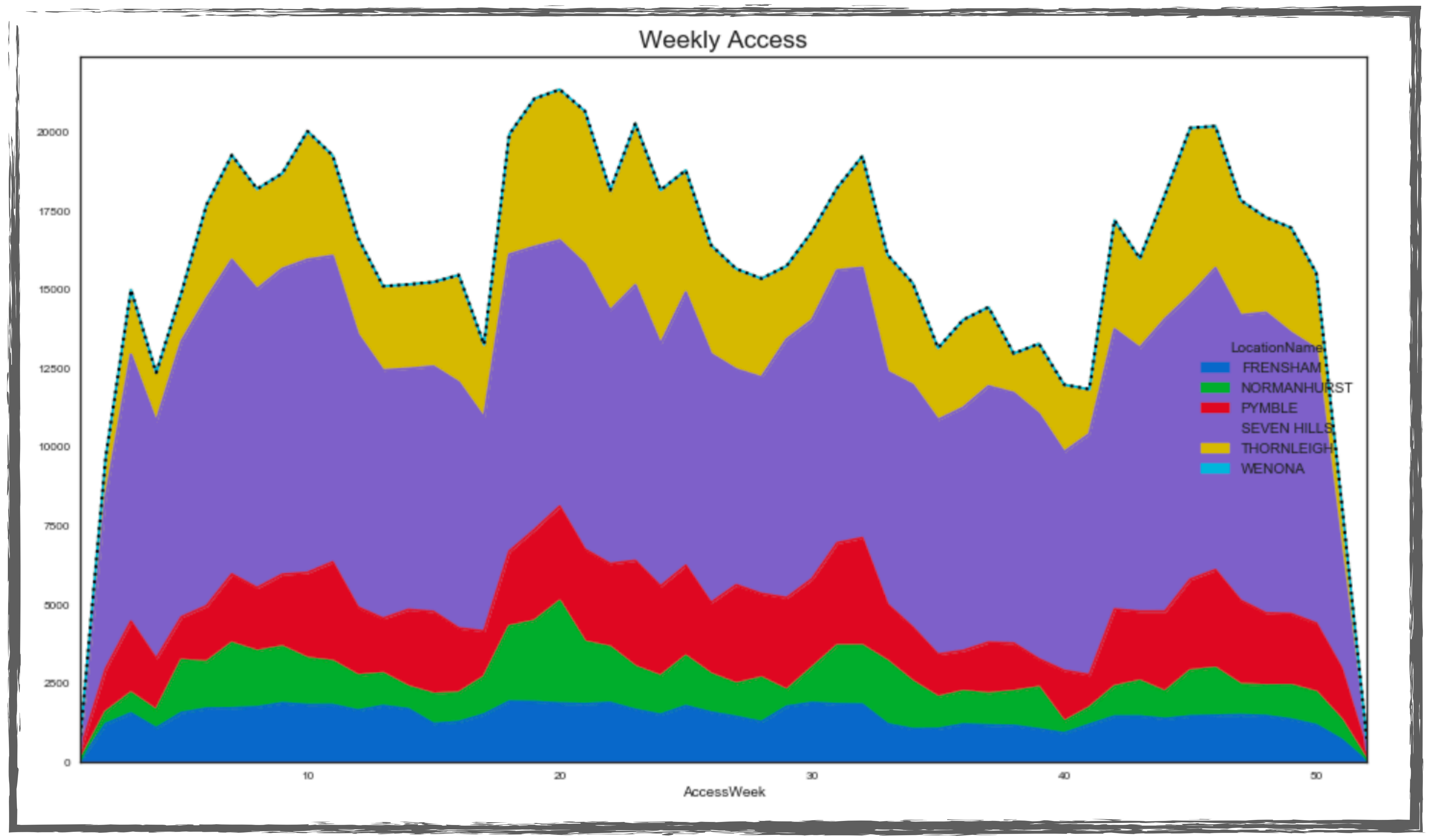
Can we cluster Aquabliss customers? What's a typical total customer lifetime value?

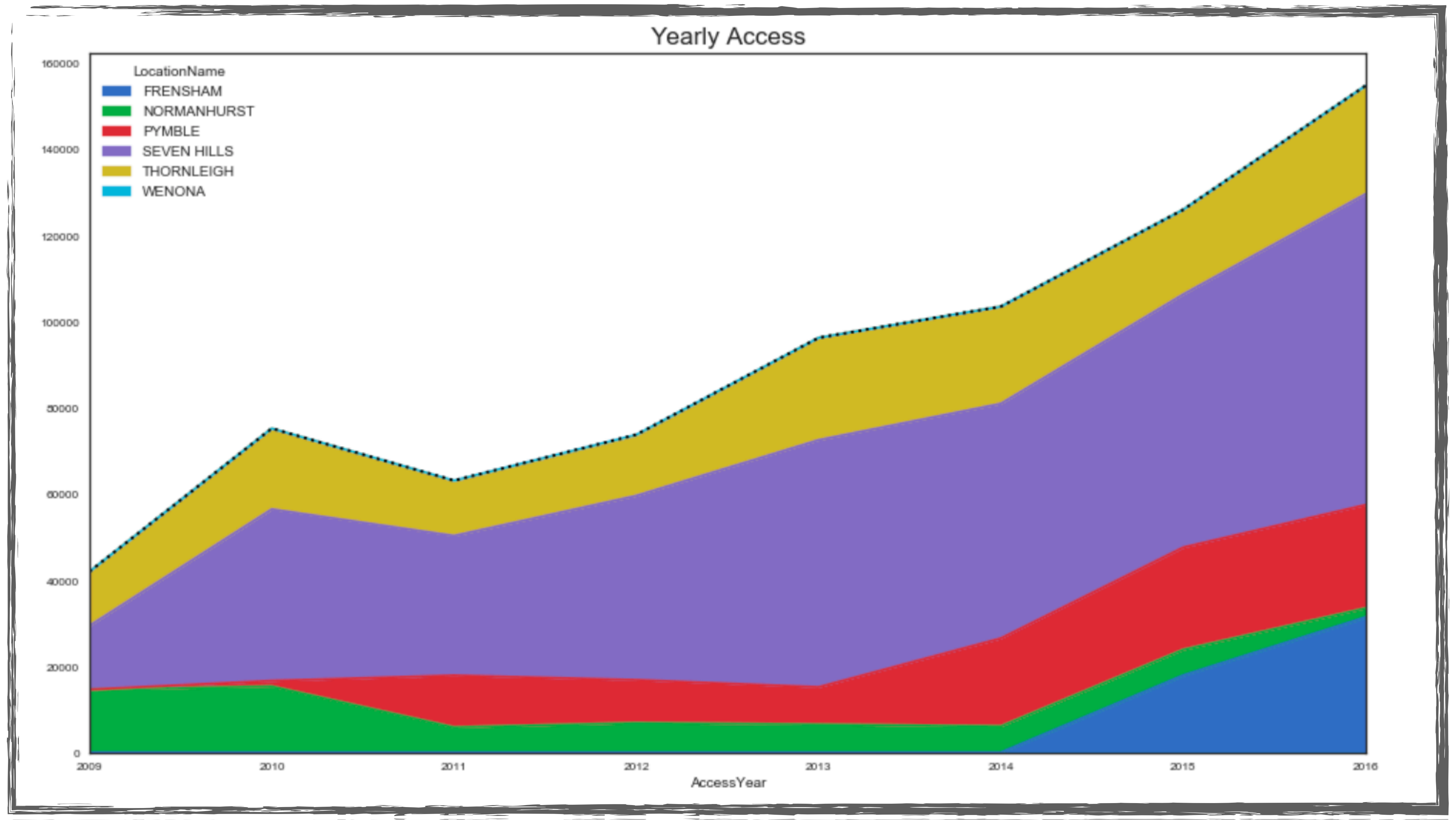
Customer lifetime value Analysis

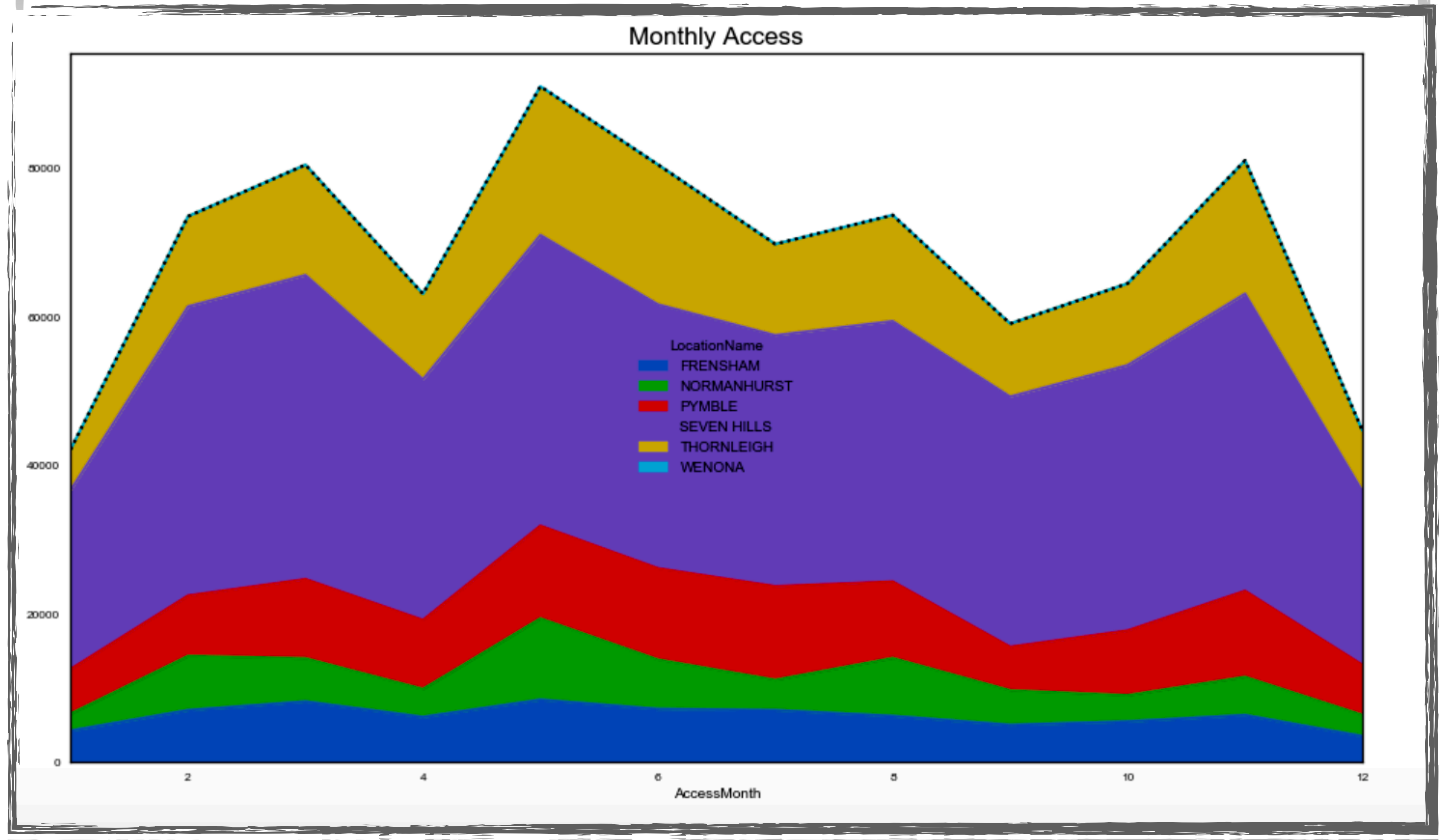

Number of Customer entering the swimmingpools
LocationName
FRENSHAM
NORMANHURST
PYMBLE
SEVEN HILLS
THORNLEIGH
WENONA
2500
2000
1500
1000
500
0
00:23
08:22
11:42
15:02
18:22
AccessTIME


Weekdays
LocationName
FRENSHAM
NORMANHURST
PYMBLE
SEVEN HILLS
THORNLEIGH
WENONA
140000
120000
100000
80000
60000
40000
20000
0
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday


Weekdays
LocationName
FRENSHAM
NORMANHURST
PYMBLE
SEVEN HILLS
THORNLEIGH
WENONA
AccessTIME
Saturdays
LocationName
FRENSHAM
NORMANHURST
PYMBLE
SEVEN HILLS
THORNLEIGH
AccessTIME
Sundays
LocationName
NORMANHURST
SEVEN HILLS
THORNLEIGH
AccessTIME







LocationName AccessYear	FRENSHAM	NORMANHURST	PYMBLE	SEVEN HILLS	THORNLEIGH	WENONA
2009	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2010	0.000000	0.050000	inf	1.694042	0.489249	0.000000
2011	0.000000	-0.618072	8.301391	-0.186173	-0.320635	inf
2012	0.000000	0.170669	-0.169990	0.319365	0.097404	1.115789
2013	0.000000	-0.050196	-0.132132	0.341556	0.697635	-1.000000
2014	0.000000	-0.063113	1.354095	-0.050351	-0.050516	0.000000
2015	inf	-0.022564	0.163253	0.078428	-0.127466	0.000000
2016	0.753506	-0.647708	0.010530	0.228480	0.278150	0.000000

Holt Winters

Exponential smoothing is a rule of thumb technique for smoothing time series data. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality.

ARIMA

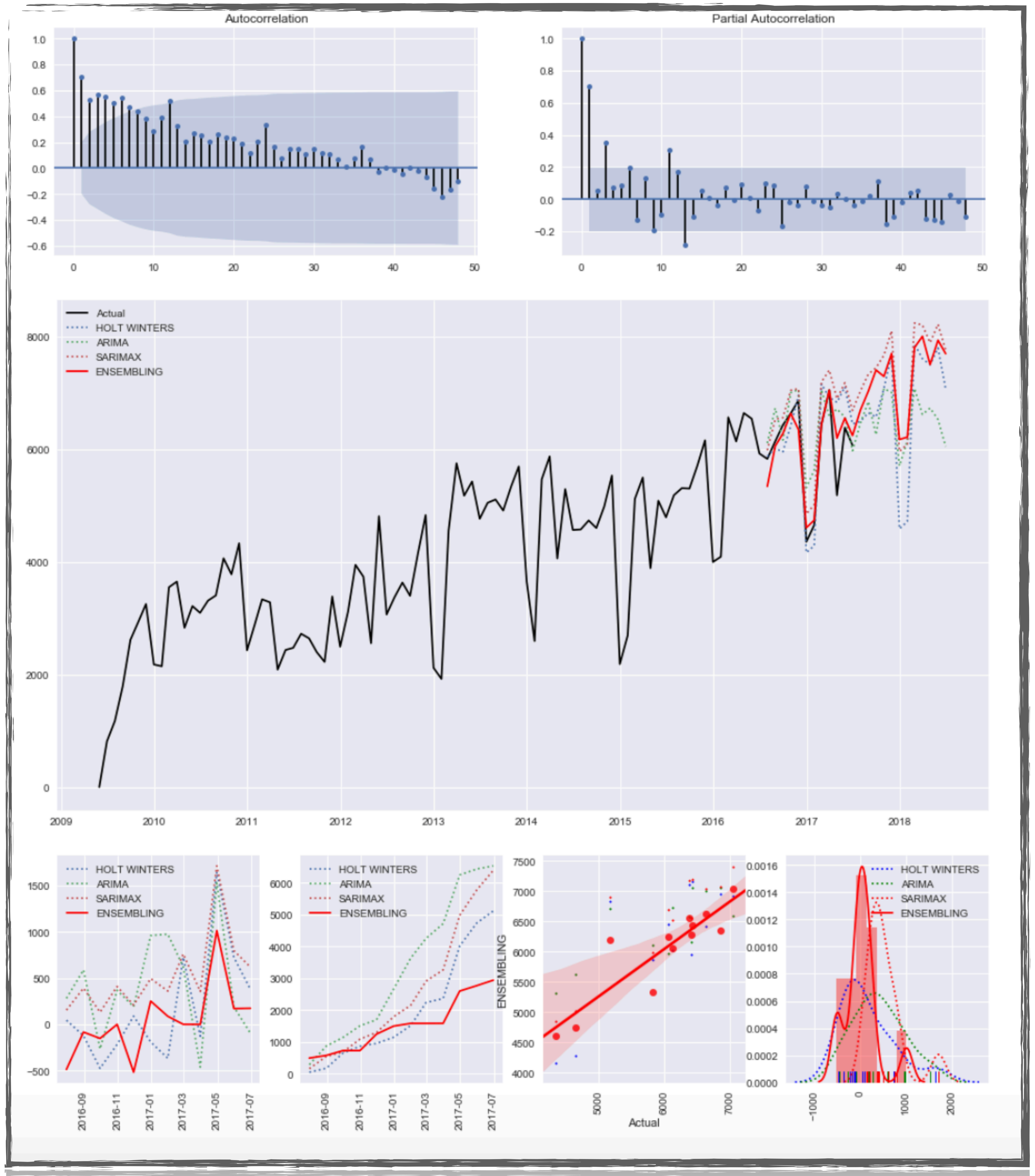
In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model. Where AR indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model.

ENSEMBLING

SARIMA

Seasonal ARIMA models are usually denoted ARIMA (p,d,q) (P,D,Q) m, where m refers to the number of periods in each season, and the uppercase P,D,Q refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model.



The average entire during the overlapping period is: **5998**

The mean absolute error for the Holt Winter model in the overlapping period is: **425**

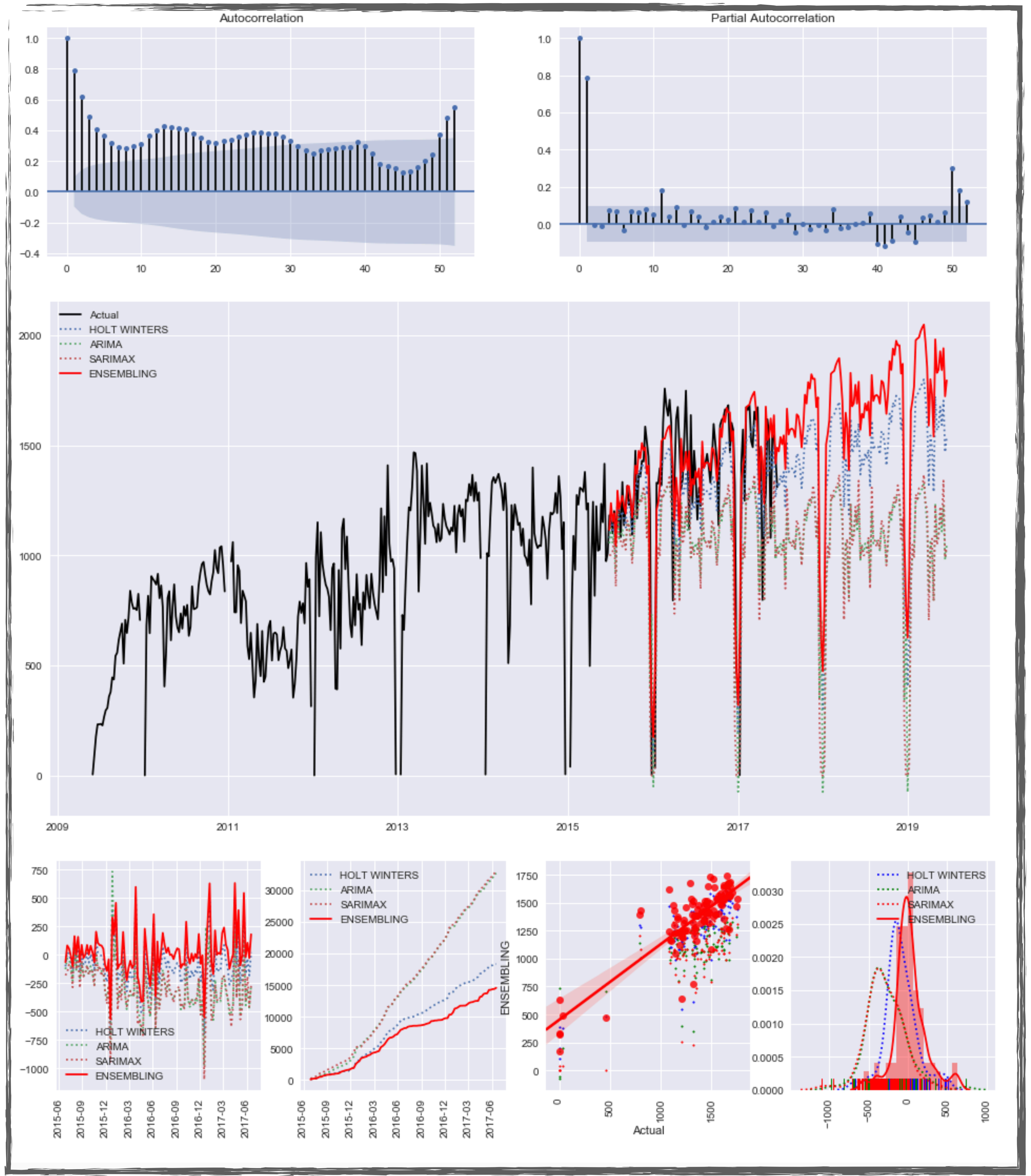
The mean absolute error for the ARIMA in the overlapping period is: **719**

The mean absolute error for the SARIMAX model in the overlapping period is: **226**

The mean absolute error for the **ENSEMBLING** model in the overlapping period is: **201**

The predicted average **GROWTH** from 2016 to 2017 is: **10%**

The predicted average **GROWTH** from 2017 to 2018 is: **26%**



The average entire during the overlapping period is: **1332**

The mean absolute error for the Holt Winter model in the overlapping period is: **176**

The mean absolute error for the ARIMA in the overlapping period is: **314**

The mean absolute error for the SARIMAX model in the overlapping period is: **316**

The mean absolute error for the **ENSEMBLING** model in the overlapping period is: **140**

The predicted average **GROWTH** from 2016 to 2017 is: **05%**

The predicted average **GROWTH** from 2017 to 2018 is: **19%**

The average entire during the overlapping period is: **1938**

The mean absolute error for the Holt Winter model in the overlapping period is: **405**

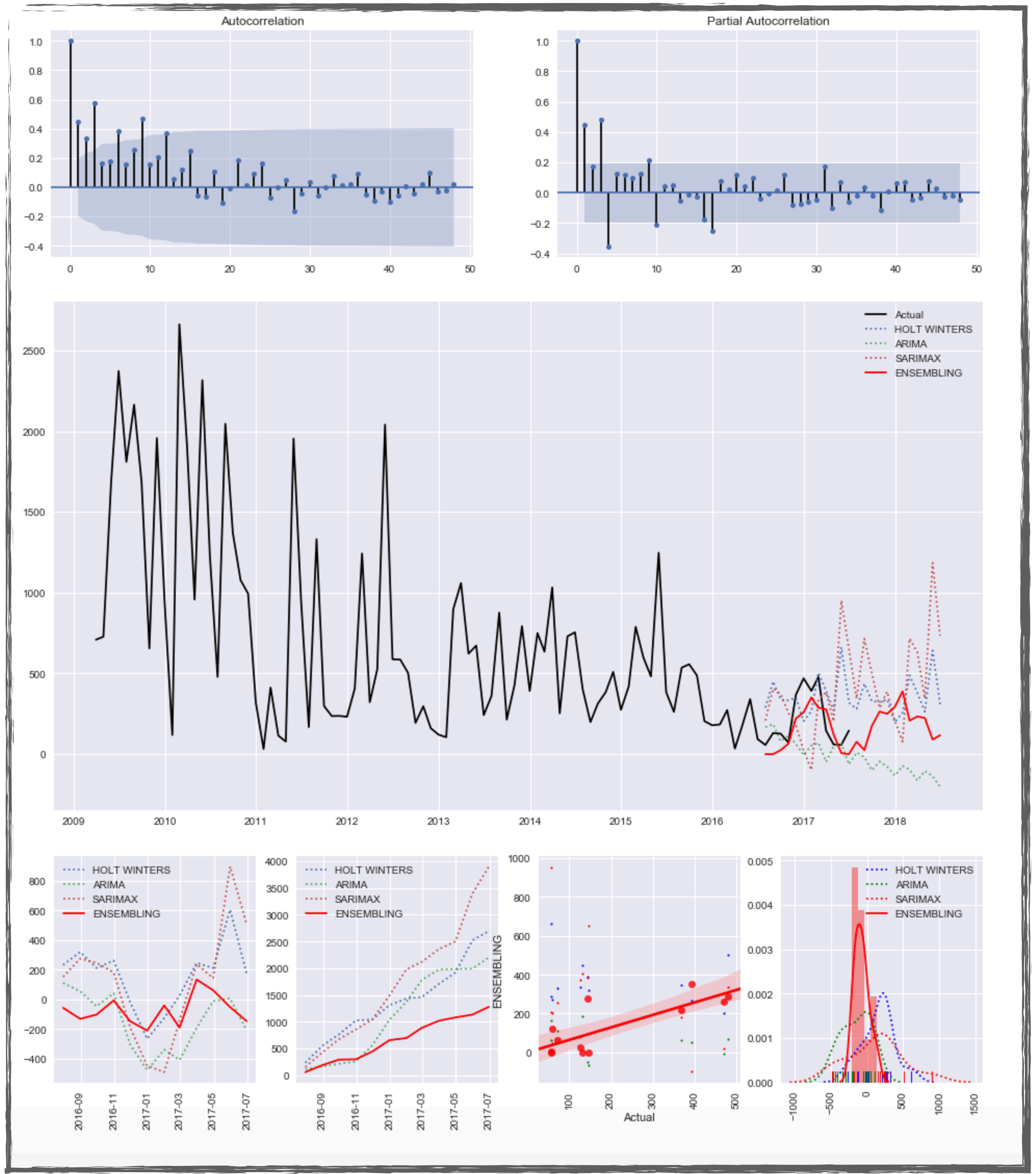
The mean absolute error for the ARIMA in the overlapping period is: **499**

The mean absolute error for the SARIMAX model in the overlapping period is: **266**

The mean absolute error for the **ENSEMBLING** model in the overlapping period is: **93**

The predicted average **GROWTH** from 2016 to 2017 is: **02%**

The predicted average **GROWTH** from 2017 to 2018 is: **10%**



The average entire during the overlapping period is: **208**

The mean absolute error for the Holt Winter model in the overlapping period is: **224**

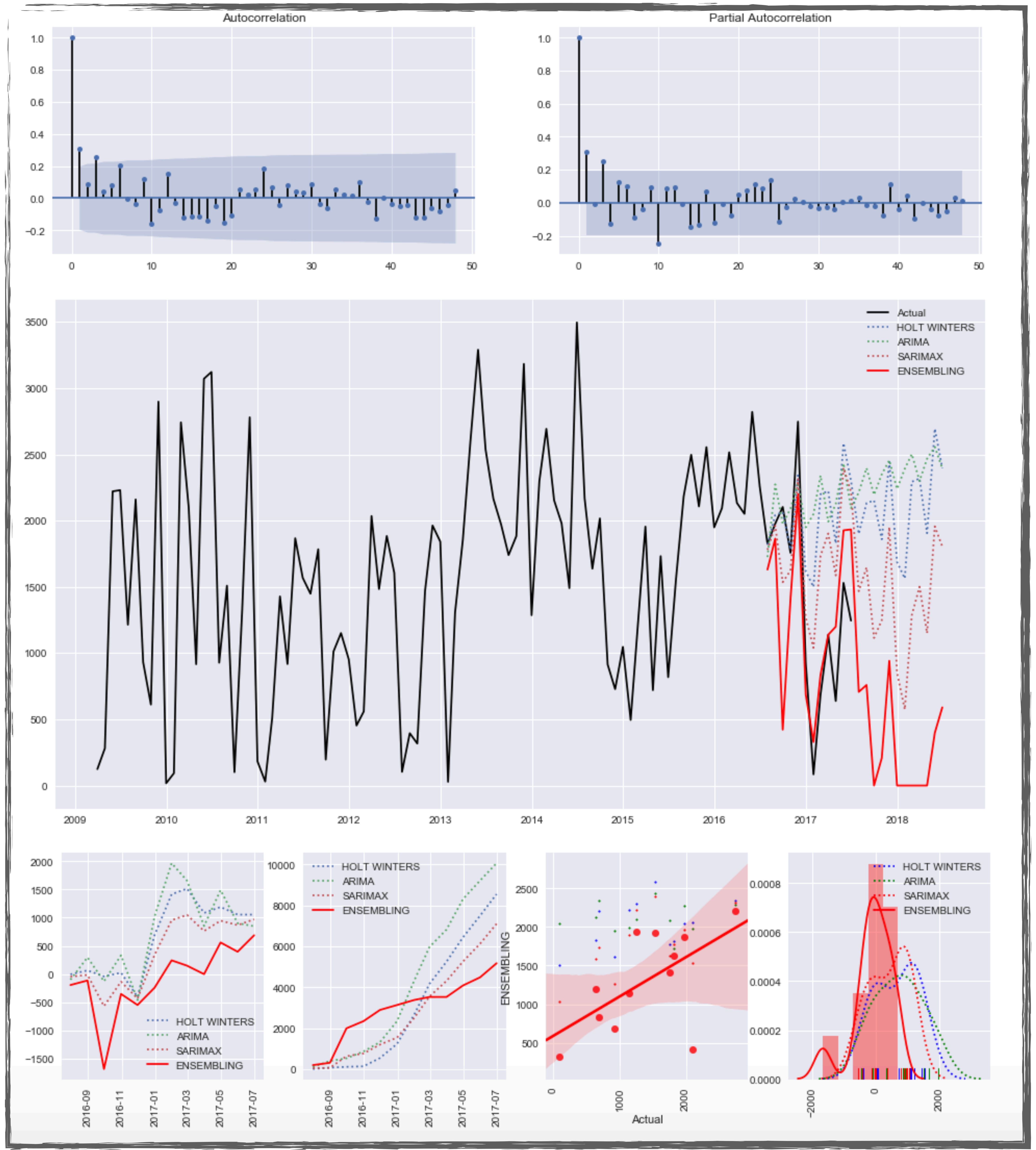
The mean absolute error for the ARIMA in the overlapping period is: **183**

The mean absolute error for the SARIMAX model in the overlapping period is: **324**

The mean absolute error for the **ENSEMBLING** model in the overlapping period is: **106**

The predicted average **GROWTH** from 2016 to 2017 is: **-08%**

The predicted average **GROWTH** from 2017 to 2018 is: **-01%**



The average entire during the overlapping period is: **1387**

The mean absolute error for the Holt Winter model in the overlapping period is: **711**

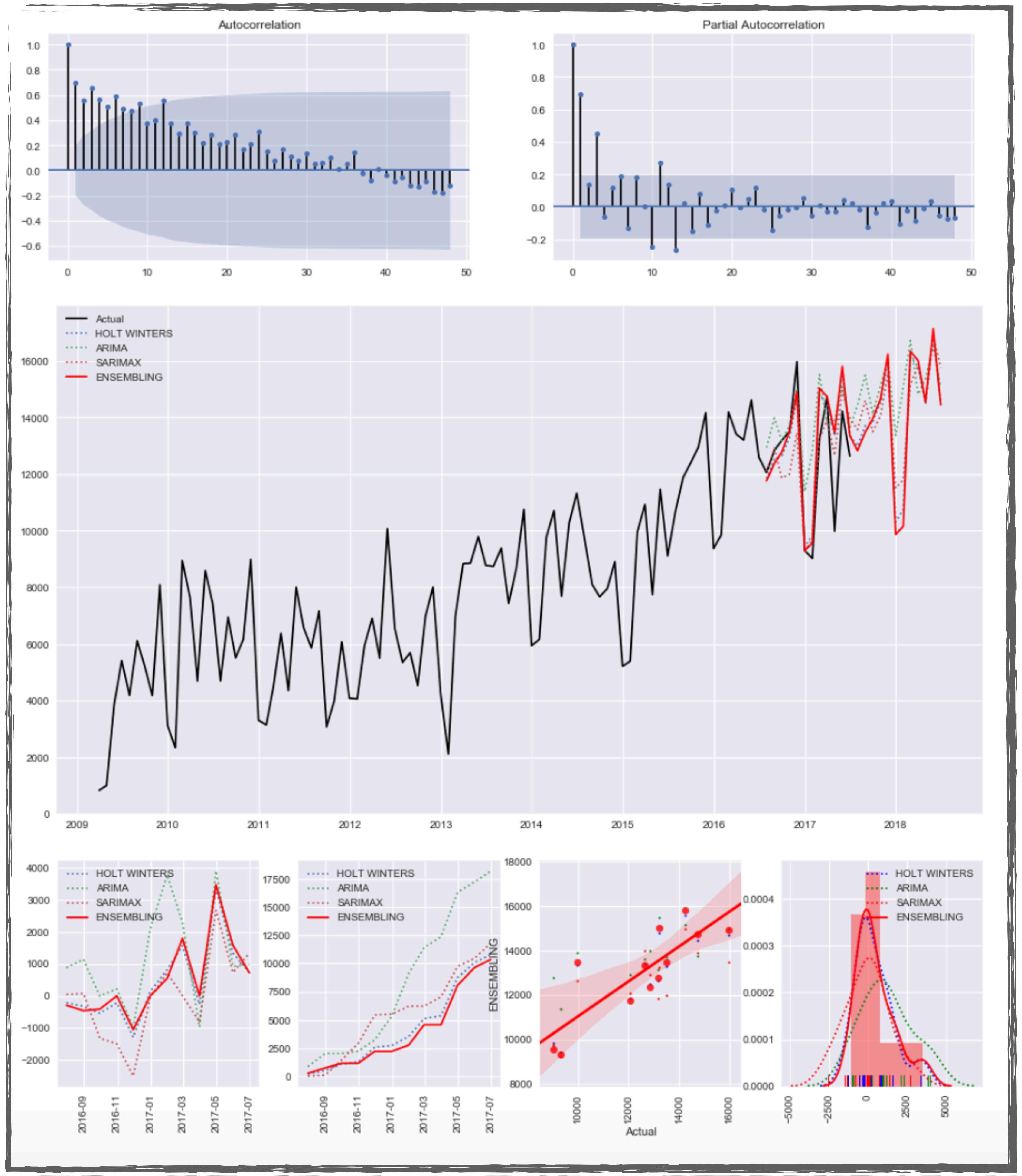
The mean absolute error for the ARIMA in the overlapping period is: **837**

The mean absolute error for the SARIMAX model in the overlapping period is: **591**

The mean absolute error for the **ENSEMBLING** model in the overlapping period is: **430**

The predicted average **GROWTH** from 2016 to 2017 is: **-60%**

The predicted average **GROWTH** from 2017 to 2018 is: **-85%**



The average entire during the overlapping period is: **12559**

The mean absolute error for the Holt Winter model in the overlapping period is: **901**

The mean absolute error for the ARIMA in the overlapping period is: **1513**

The mean absolute error for the SARIMAX model in the overlapping period is: **983**

The mean absolute error for the **ENSEMBLING** model in the overlapping period is: **863**

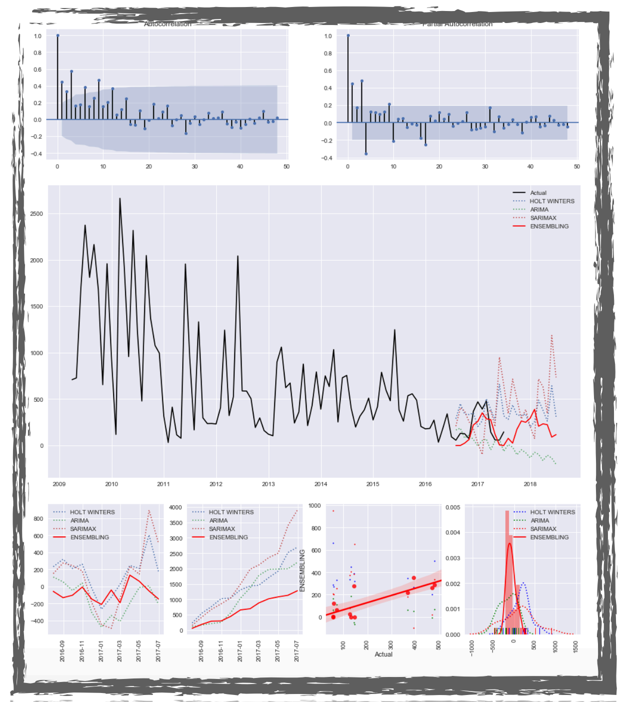
The predicted average **GROWTH** from 2016 to 2017 is: **05%**

The predicted average **GROWTH** from 2017 to 2018 is: **20%**

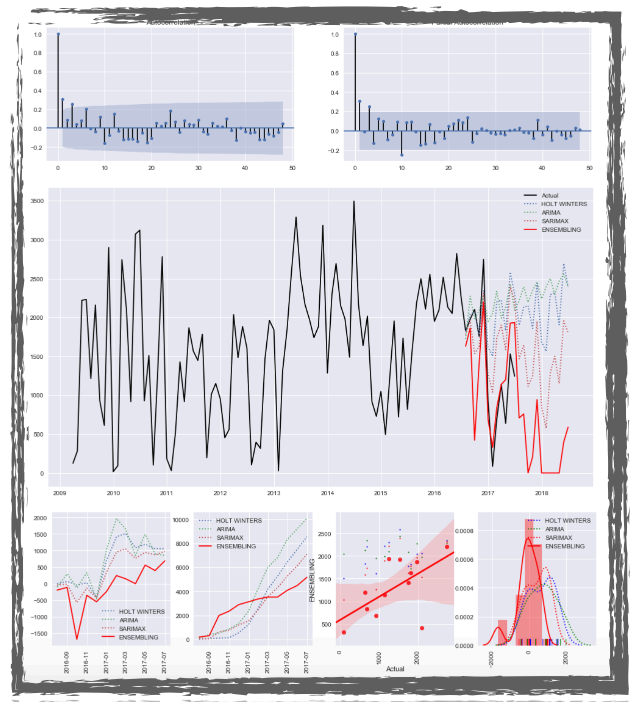
	2017	2018
SEVEN HILLS	10%	26%
PYMBLE	2%	10%
NORMANHURST	-8%	-2%
THORNLEIGH	-60%	-75%
FRENSHAM & WENONA	Not enough data	Not enough data
TOTAL	5%	20%

PYMBLE

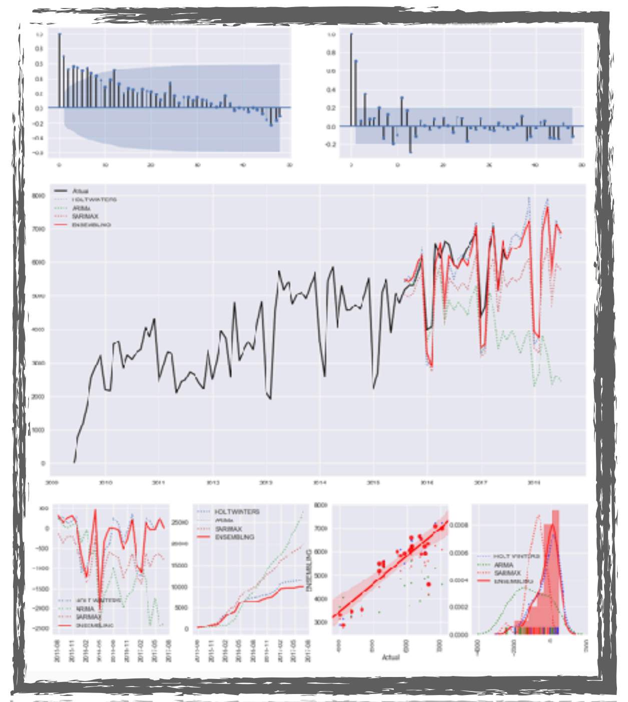
NORMANHURST



THORNLEIGH

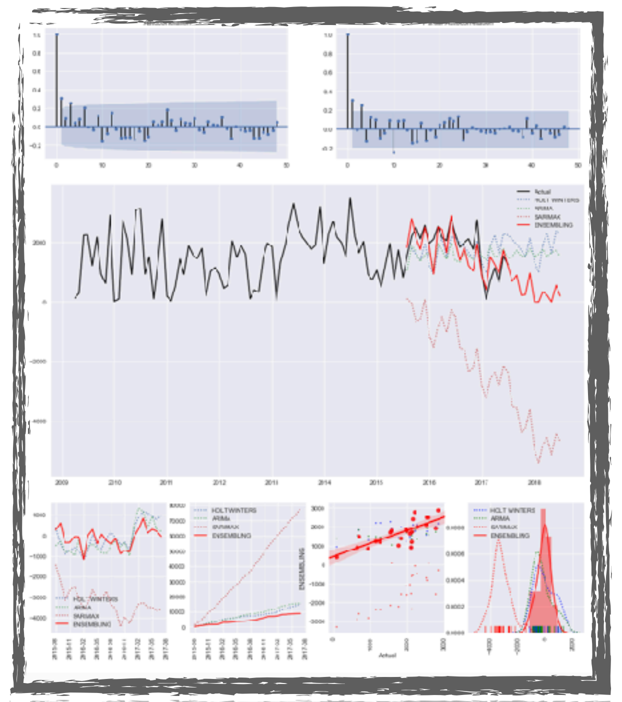


SEVEN HILLS
3 YEARS

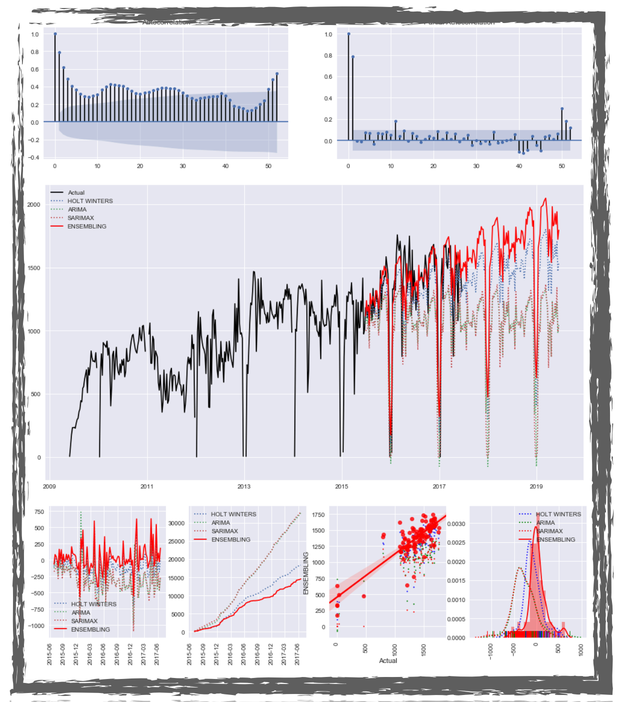


PYMBLE
3 YEARS

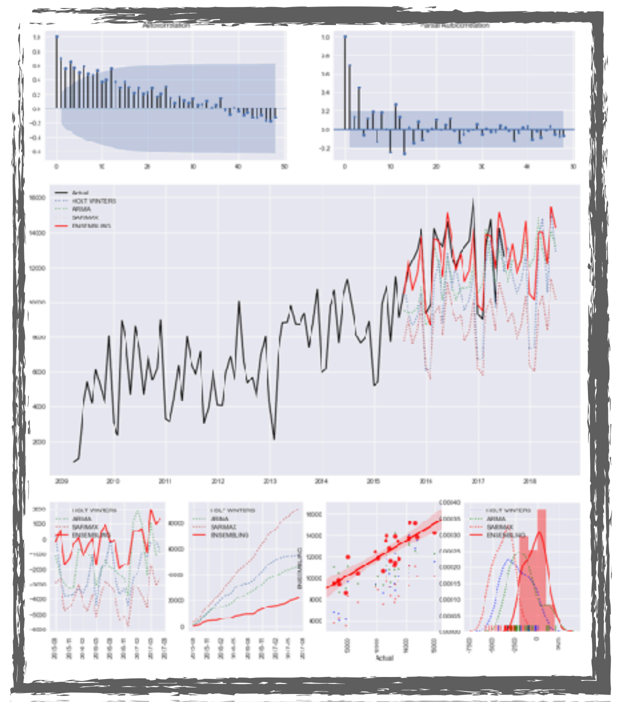
THORNLEIGH
3 YEARS



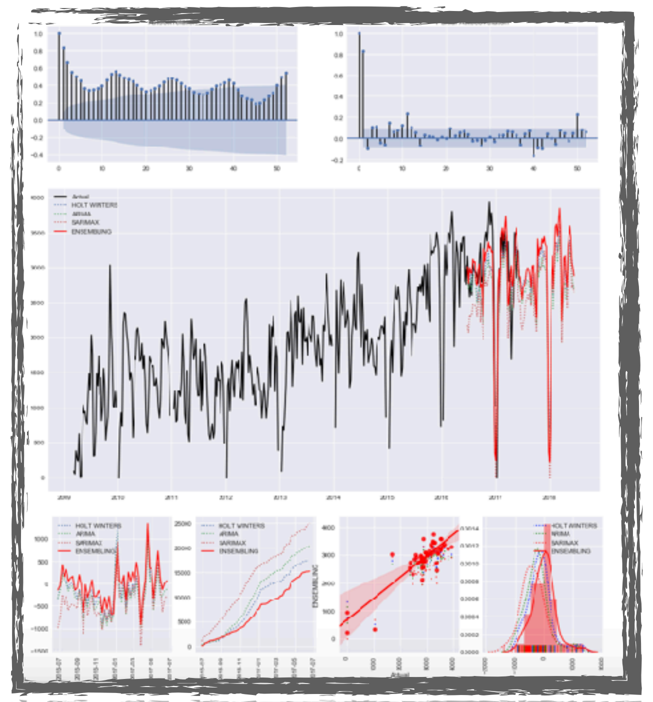
SEVEN HILLS
4 YEARS
WEEKLY



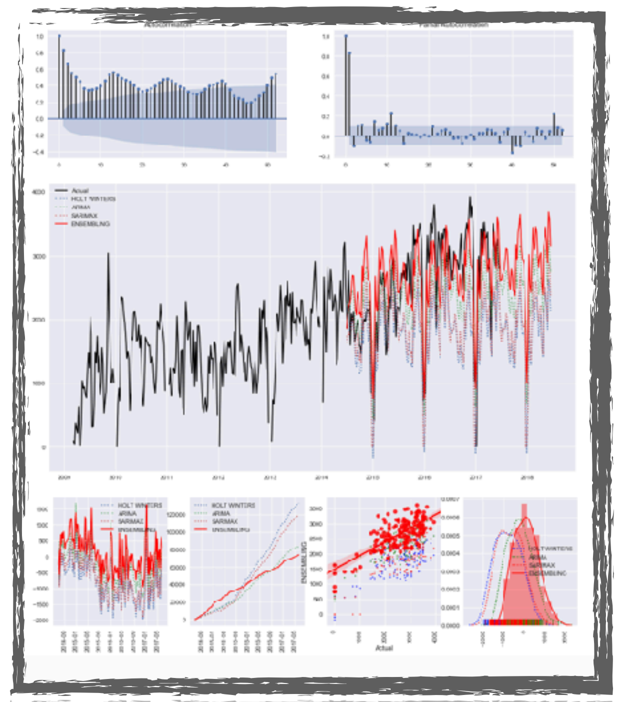
TOTAL
3 YEARS



TOTAL
WEEKLY

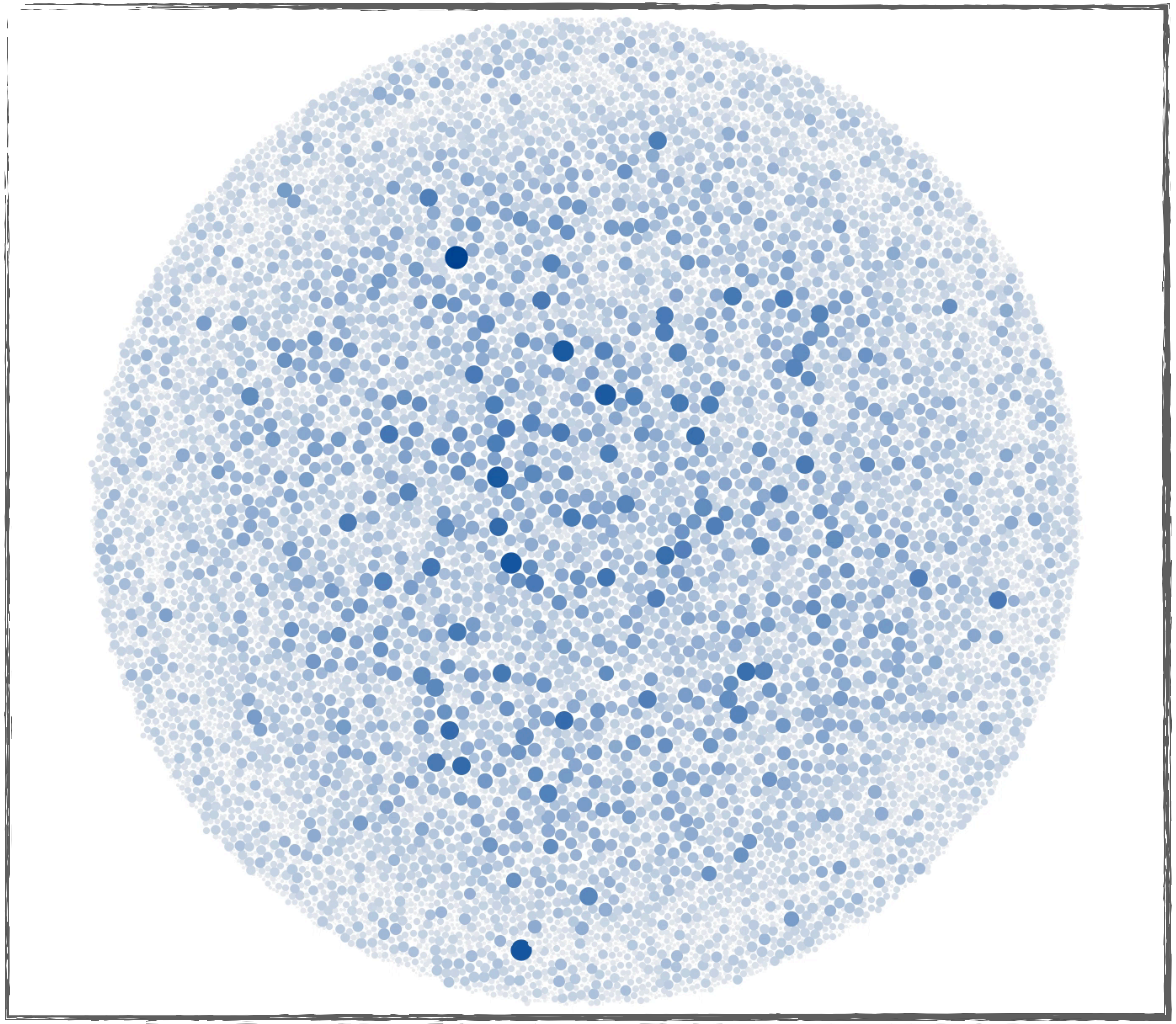


TOTAL 3 YEARS
WEEKLY



ARE OUR CUSTOMER ALL THE SAME ?

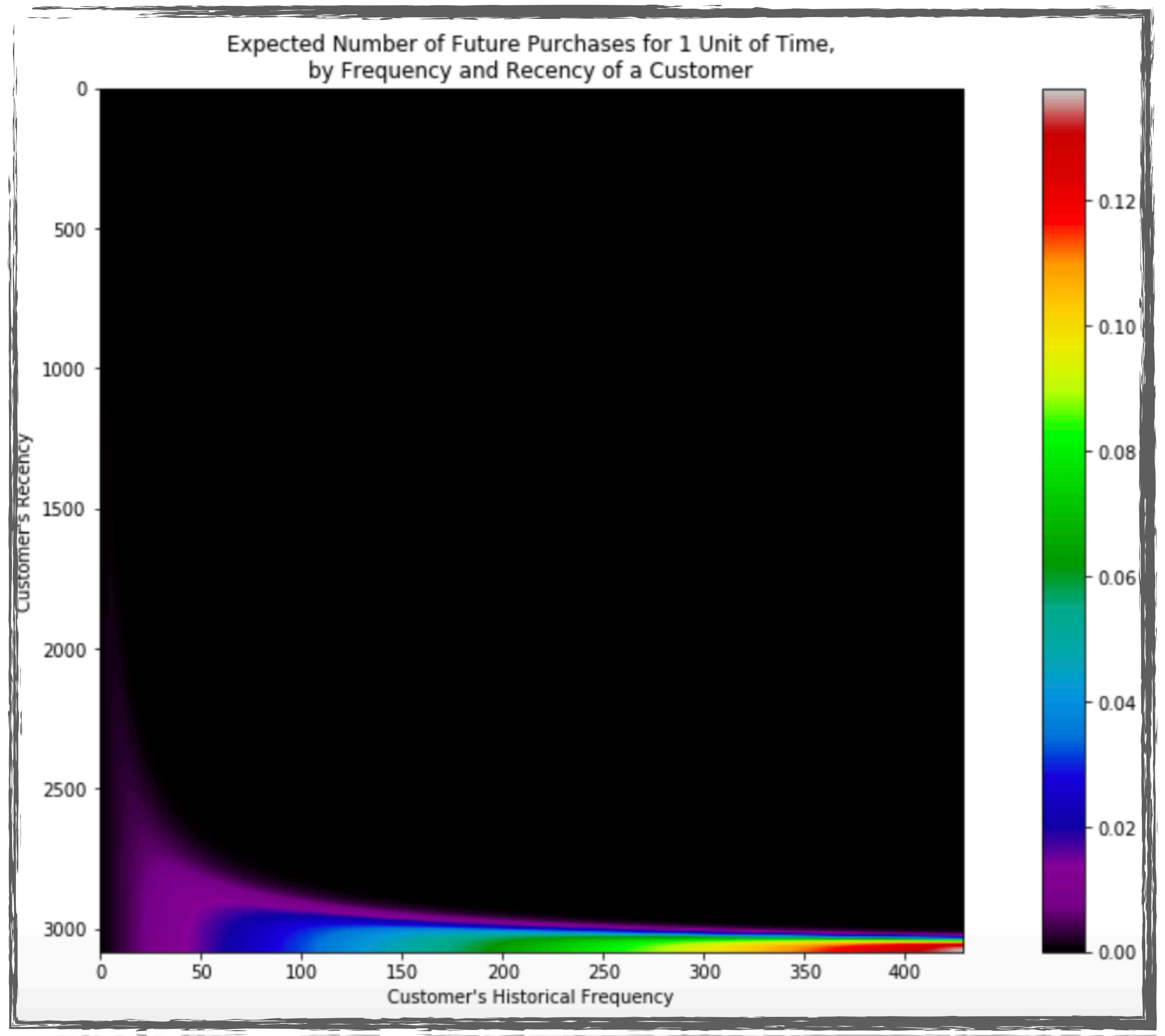
Customer Analysis & lifetime evaluation



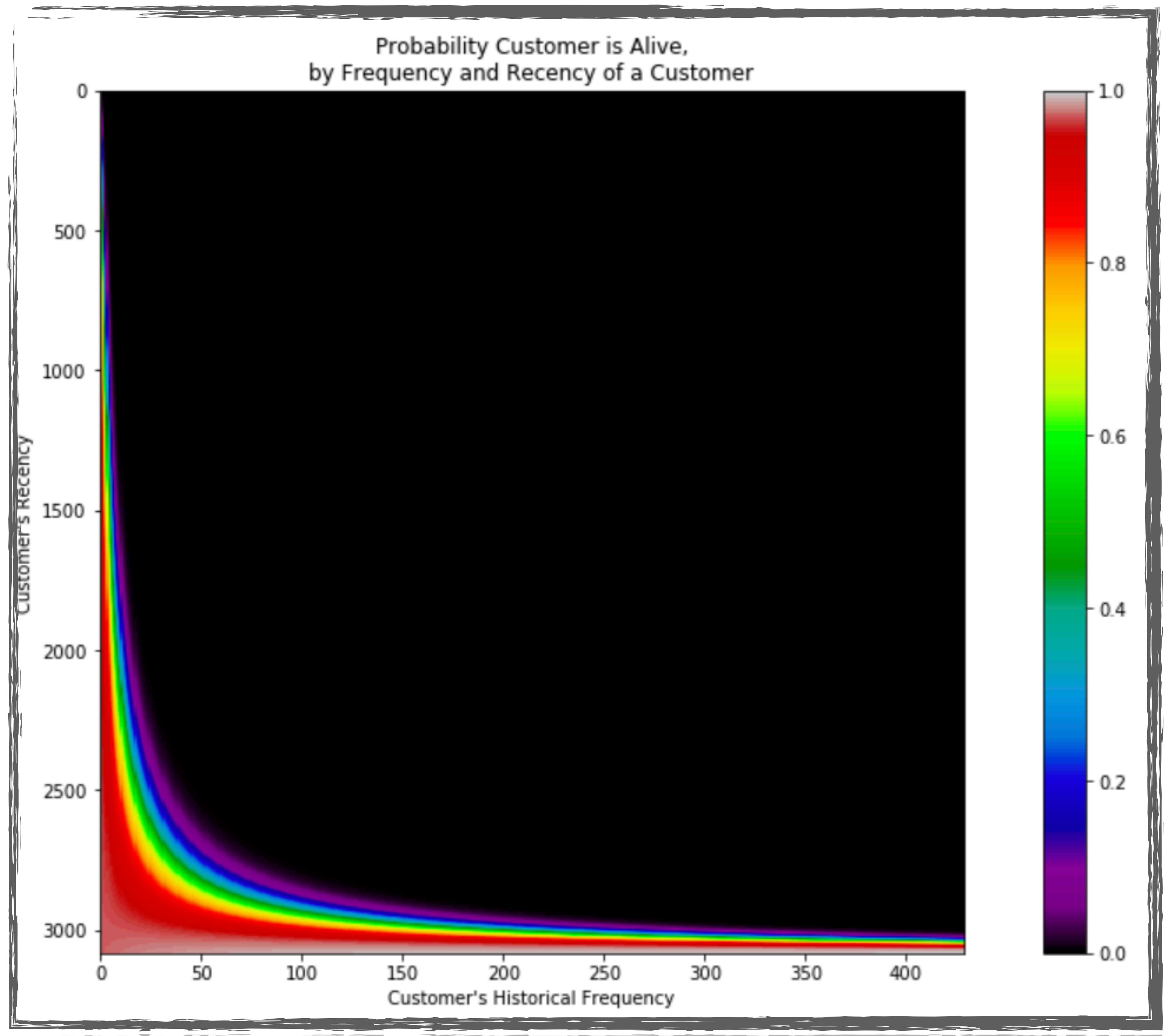
Here we have an overview of all your **CUSTOMER BASE**, which express how different customer behave.

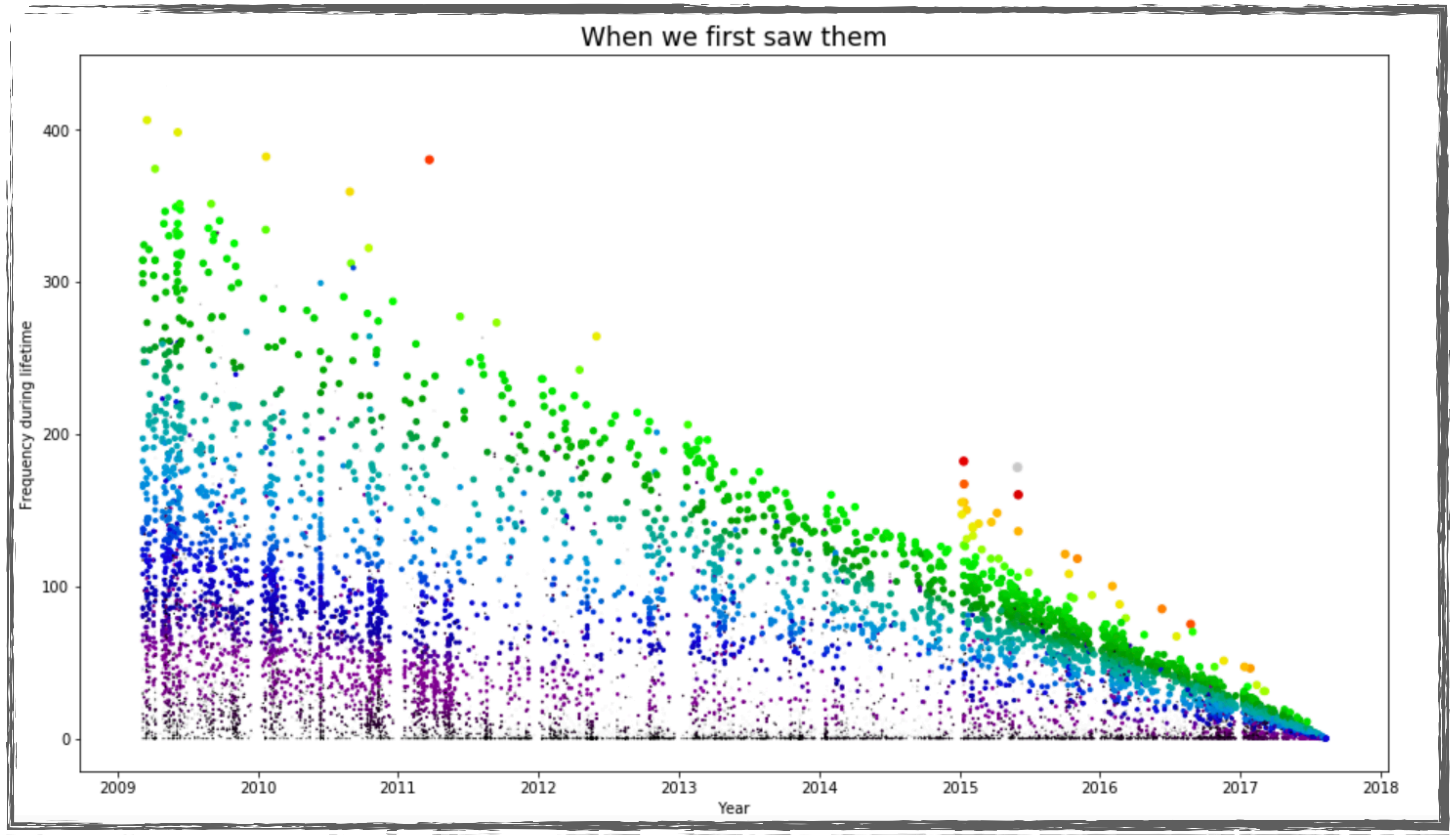
The darkens and size of the point indicate the the **NUMBER OF ENTRY** that each customer has reach during his lifetime.

Finally we can clearly see that we have some really **FREQUENT USERS** that have a high fidelity to Aquabliss .



Basic Frequency /Recency analysis using the BG/NBD model.

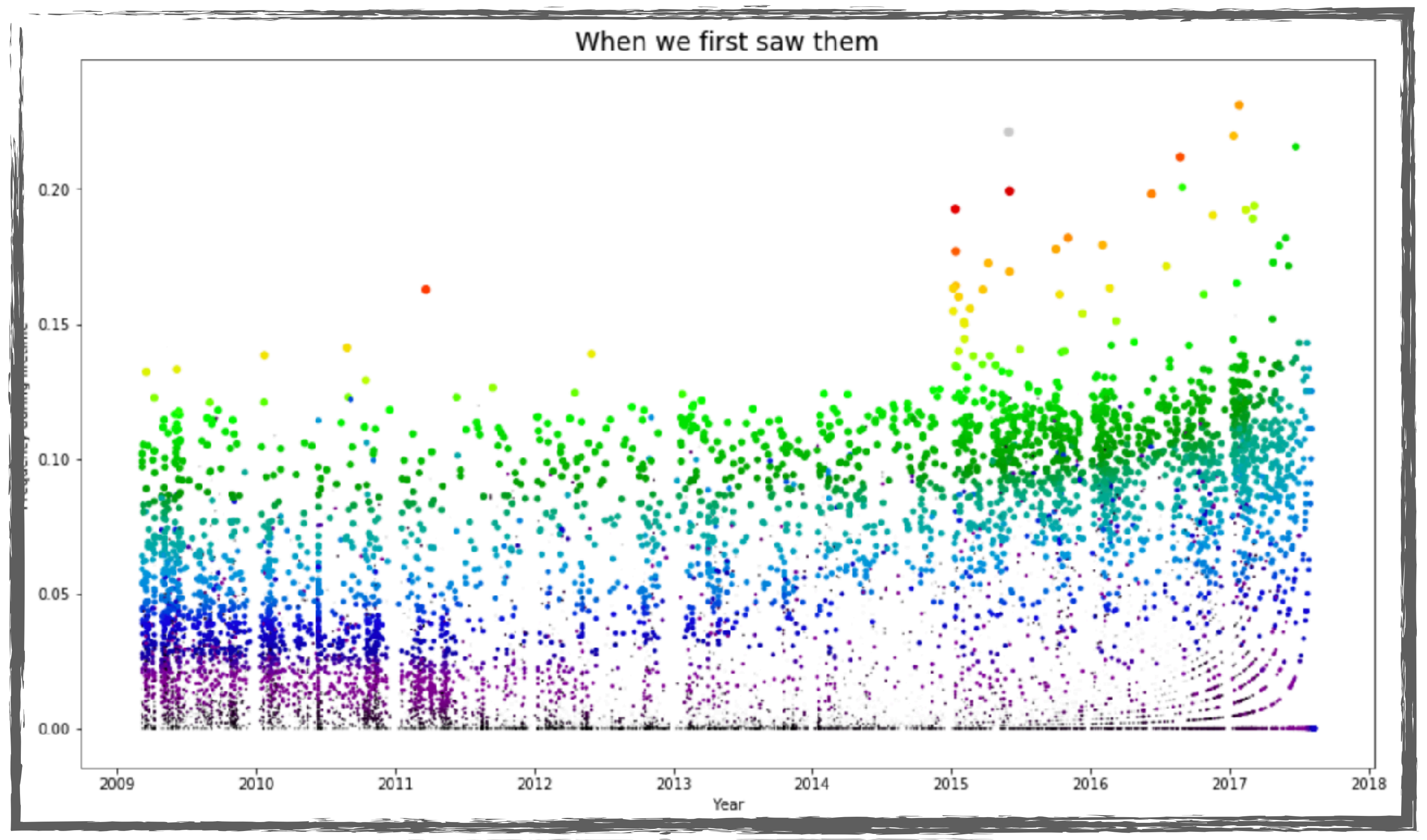


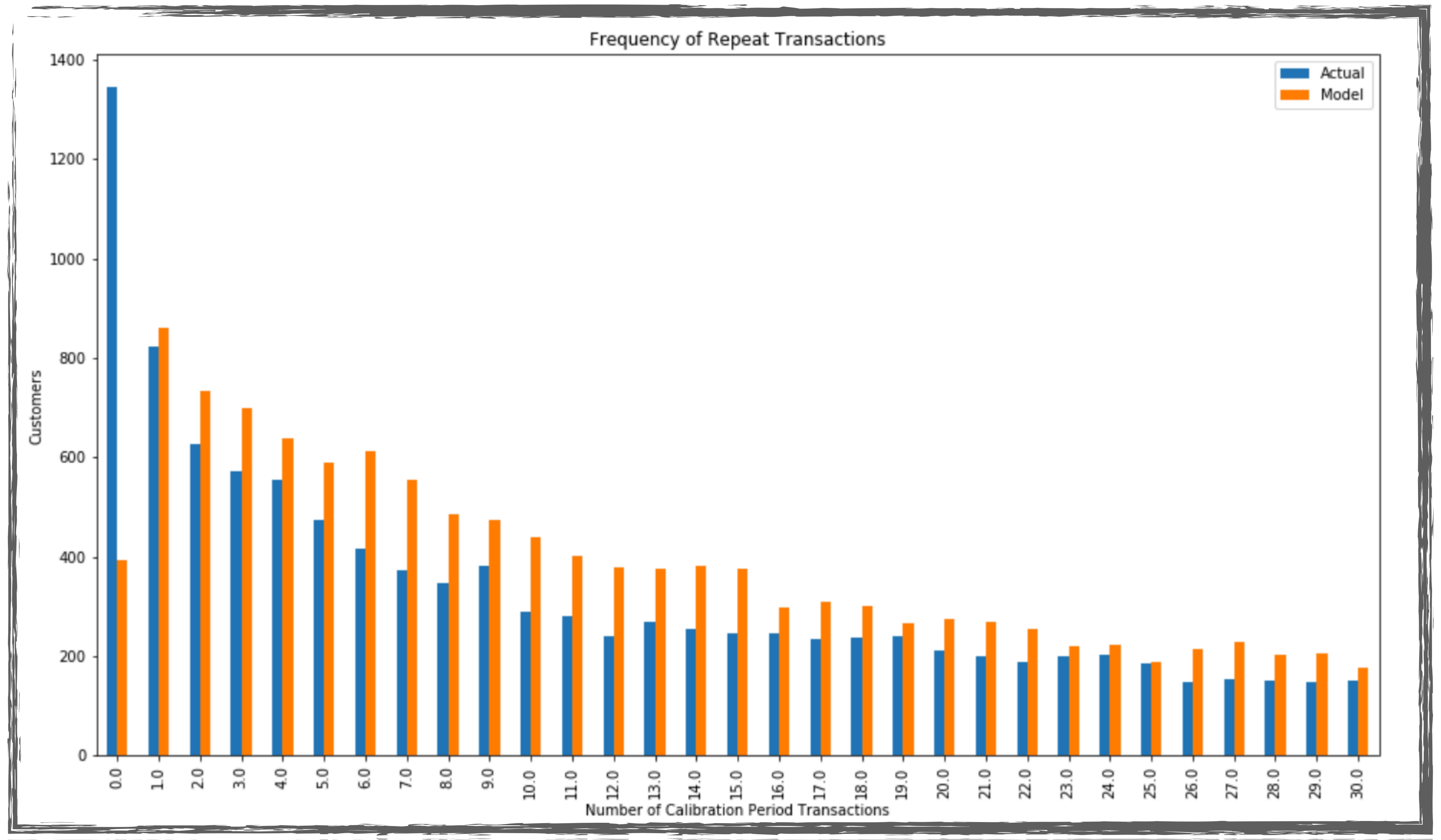


PeopleId	frequency	recency	T	predicted_entries
805	178.0	805.0	805.0	66.879704
808	160.0	803.0	803.0	60.396388
12	182.0	941.0	945.0	59.373957
13781	75.0	350.0	354.0	55.199979
14812	380.0	2336.0	2336.0	55.023547
42	167.0	940.0	944.0	54.641204
39742	85.0	427.0	429.0	54.149065
31835	118.0	649.0	649.0	53.599611
2850	46.0	196.0	199.0	52.978429
11834	121.0	677.0	681.0	52.445508

PeopleId	frequency	recency	T	predicted_entries
4491	324.0	3073.0	3079.0	36.162935
14812	380.0	2336.0	2336.0	55.023547
975	335.0	2907.0	2911.0	39.450469
355	349.0	2991.0	2996.0	39.972602
3089	359.0	2538.0	2543.0	47.921547
4285	374.0	3047.0	3050.0	42.148637
1320	340.0	2857.0	2882.0	38.587658
12740	216.0	2443.0	2490.0	23.831388
1238	332.0	2813.0	2887.0	2.490044
1131	331.0	2890.0	2896.0	39.127847

Clustering the client base
upon historical frequency
and predicted frequency.



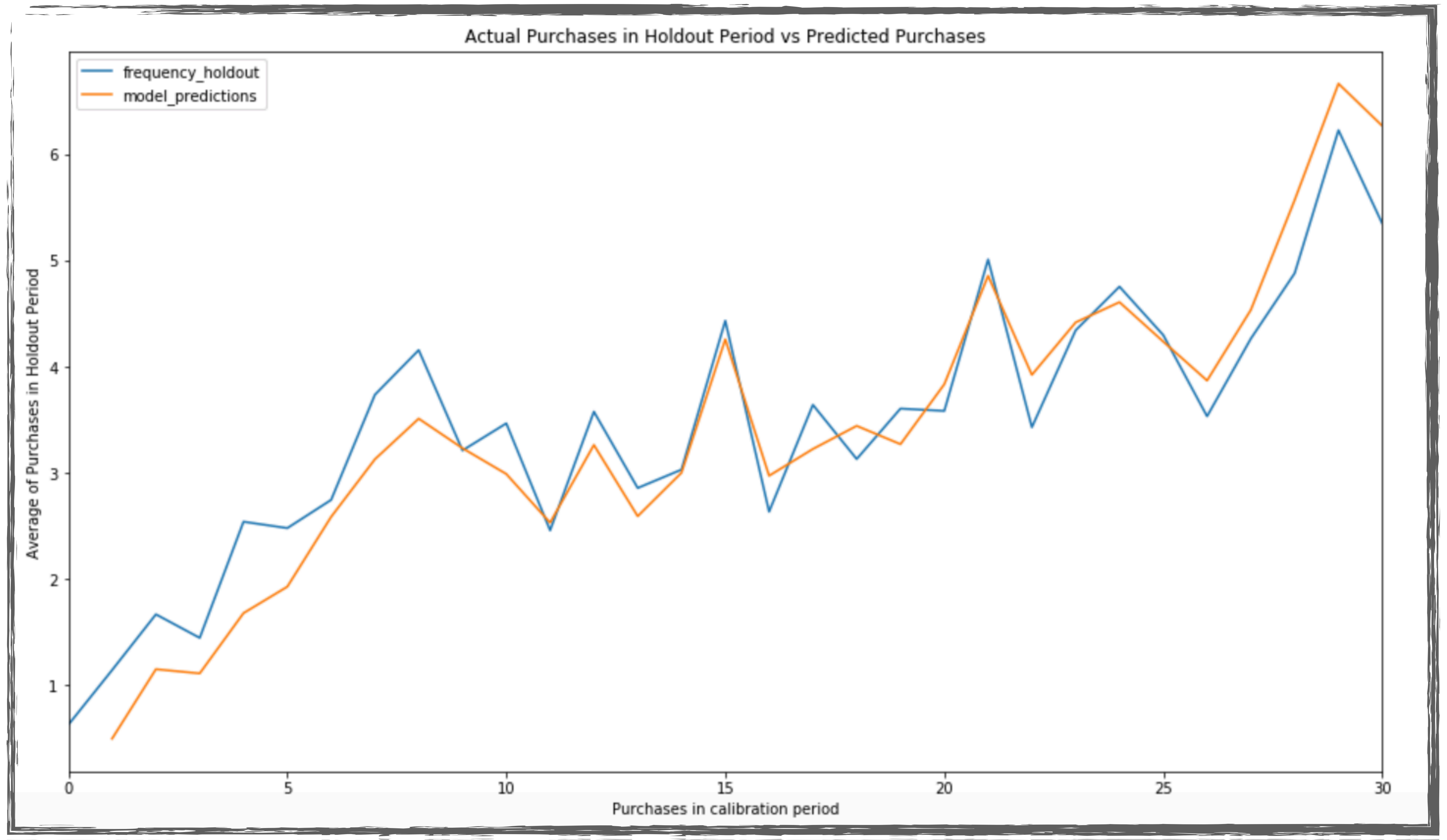


Assessing model fit

We can predict and visualise our customers' behaviour, but is our model correct? There are a few ways to assess the model's correctness. The first is to compare your data versus artificial data simulated with your fitted model's parameters.

More model fitting

With transactional data, we can partition the dataset into a calibration period dataset and a holdout dataset. This is important as we want to test how our model performs on data not yet seen (think cross-validation in standard machine learning literature).



"PORTA ITINERIS DICITUR LONGISSIMA ESSE"

Stefano Rossi