General Practice Prescribing Analysis

Louis Fisher

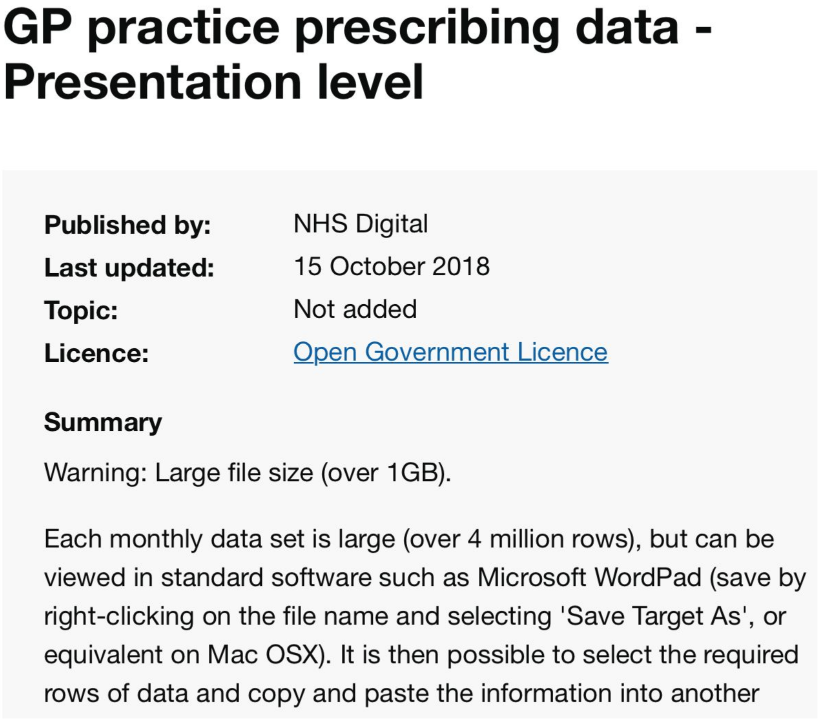
Dataset Exploration

Dataset

Dataset of General Practice (GP) prescribing data. For each GP in England, the following information is available:

- The total number of items prescribed and dispensed.
- The total net ingredient cost.
- The total actual cost.
- The total quantity.

Also provided are two additional data files containing practice information and further details on chemical names and codes.



GP practice prescribing data - Presentation level

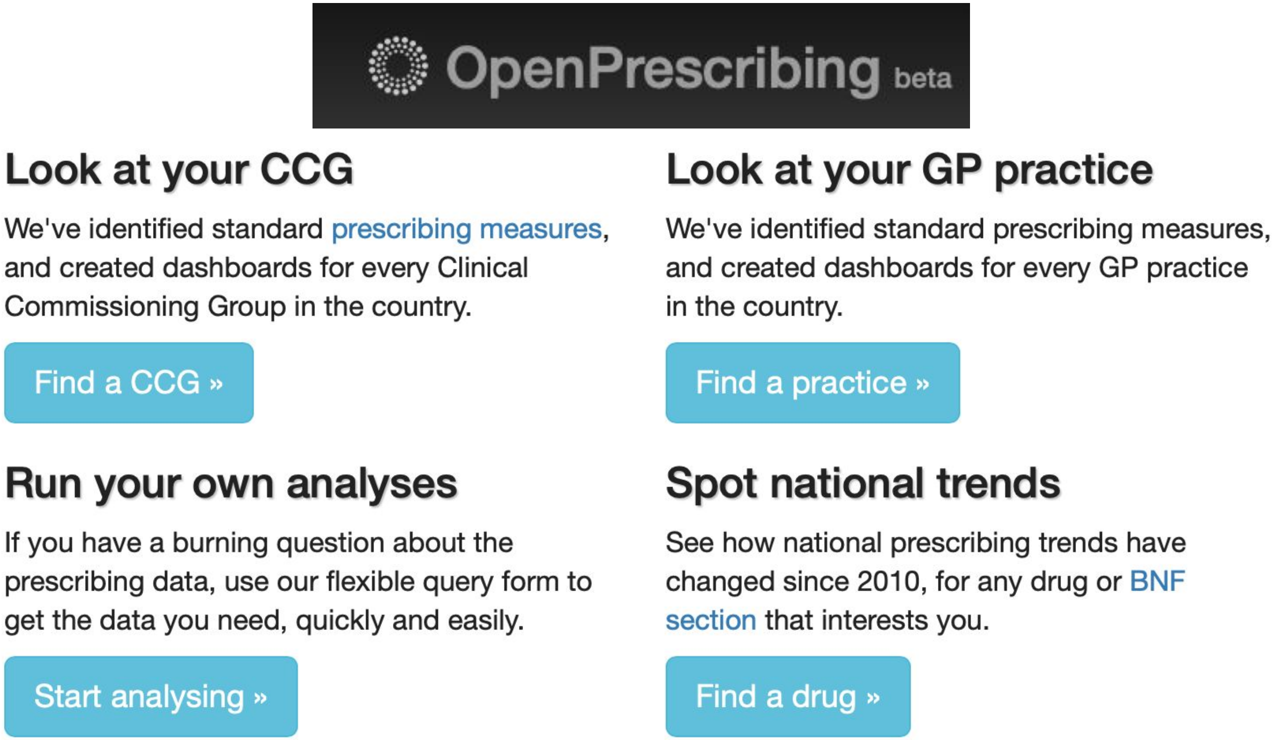
Published by:	NHS Digital
Last updated:	15 October 2018
Topic:	Not added
Licence:	Open Government Licence

Summary

Warning: Large file size (over 1GB).

Each monthly data set is large (over 4 million rows), but can be viewed in standard software such as Microsoft WordPad (save by right-clicking on the file name and selecting 'Save Target As', or equivalent on Mac OSX). It is then possible to select the required rows of data and copy and paste the information into another

Open Prescribing



Look at your CCG

We've identified standard prescribing measures, and created dashboards for every Clinical Commissioning Group in the country.

Look at your GP practice

We've identified standard prescribing measures, and created dashboards for every GP practice in the country.

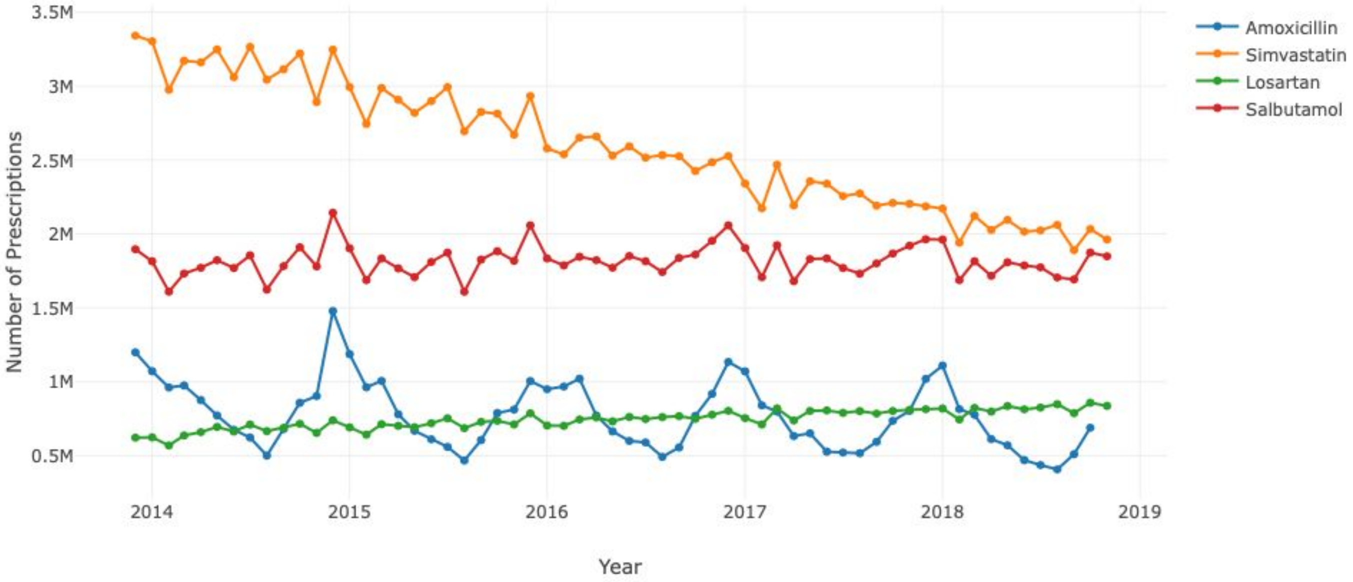
Run your own analyses

If you have a burning question about the prescribing data, use our flexible query form to get the data you need, quickly and easily.

Spot national trends

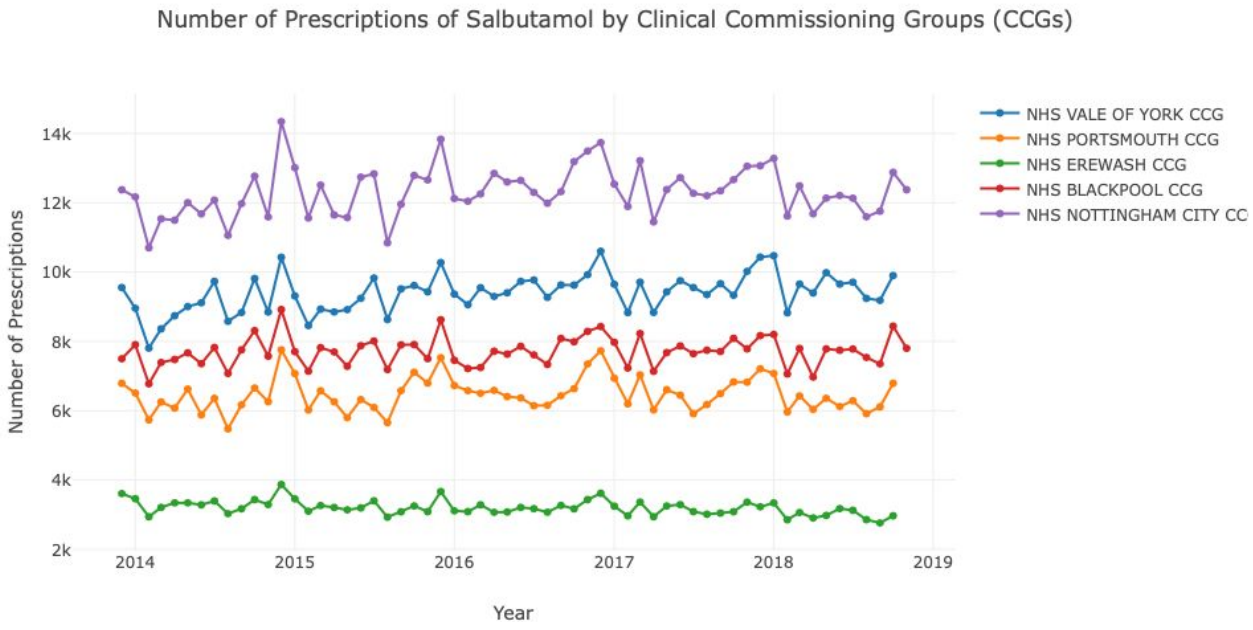
See how national prescribing trends have changed since 2010, for any drug or BNF section that interests you.

Data Exploration - Prescribing Trends by Drug



Data Exploration - Prescribing Trends by CCG

- Every GP practices belongs to one of 211 Clinical Commissioning Groups.
- Data at CCG aggregate level is available through OpenPrescribing - data for 197 CCGs.

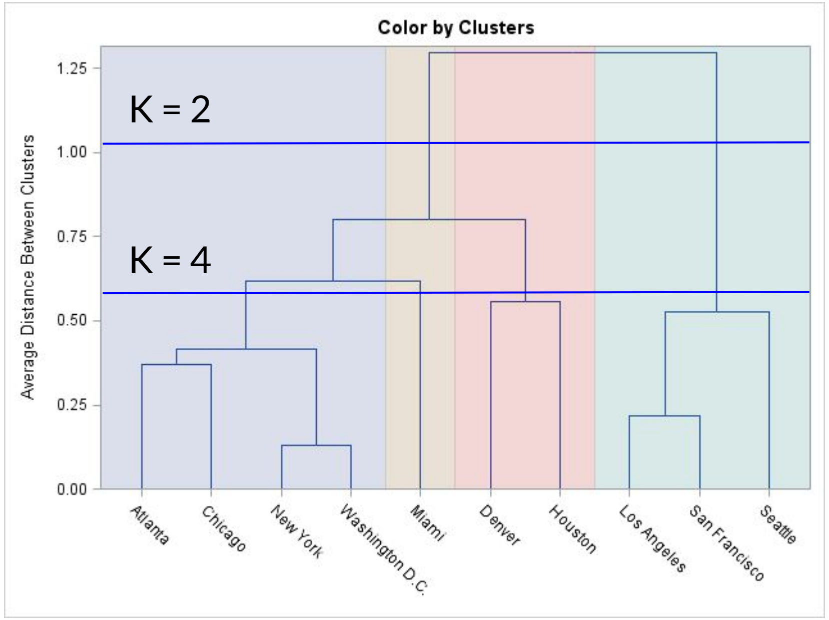


Unsupervised Approach

Hierarchical Clustering

Unsupervised Approach

- Can the prescribing trends of individual GP practices for given drugs be clustered into sub-groups?
- Can be done using hierarchical clustering - an approach that seeks to build a hierarchy of clusters.
- Can take an agglomerative or divisive approach.
- Agglomerative is “bottom-up” - each sample starts in its own cluster and pairs merge as you move up the hierarchy.
- Divisive is “top-down” - all samples start in a single cluster and are split as you move down the hierarchy.



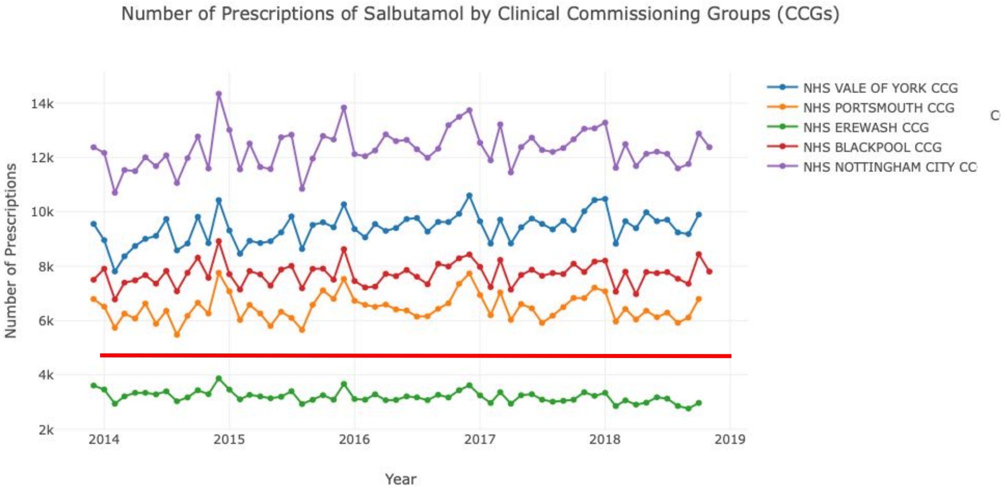
- Can be represented as a dendogram - this can be used to split into k clusters.

Data Preprocessing - Normalisation

- Without normalisation, clustering would group according to the number of prescriptions. I.e. Those with large number of prescriptions would be clustered together.
- Want to cluster by trend, so need to normalise.
- Use z-normalisation.

$$z = \frac{x - \mu}{\sigma}$$

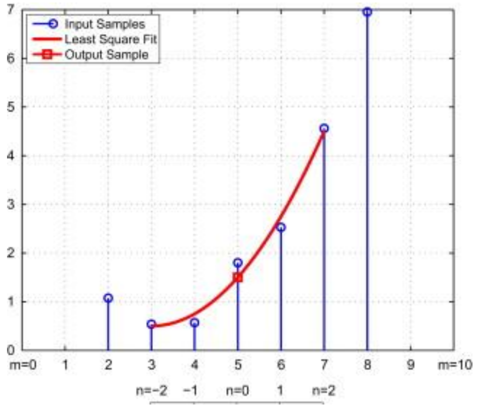
Normalised Prescriptions by CCG

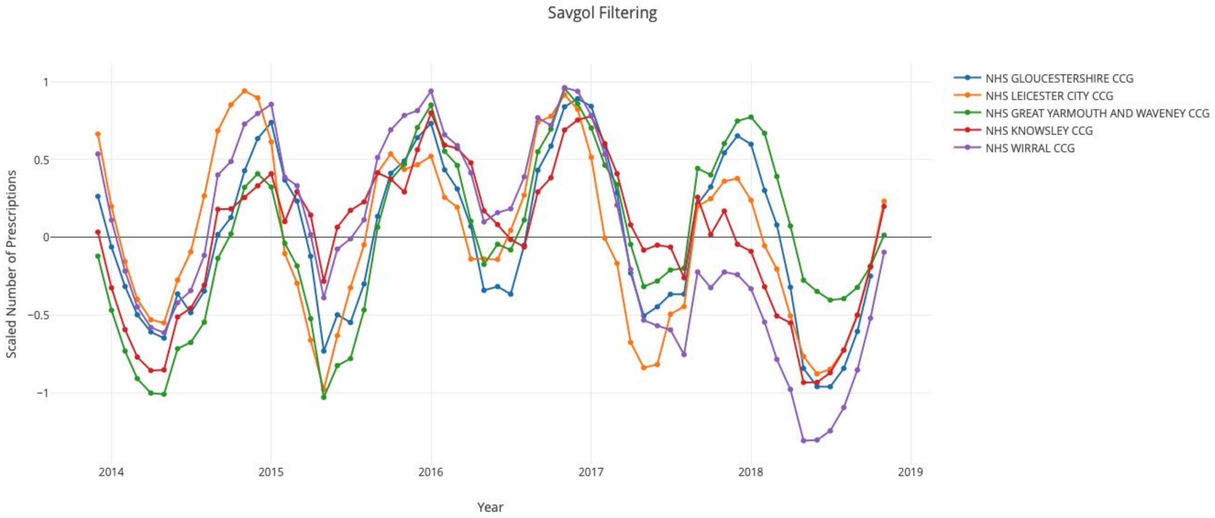


Number of Prescriptions of Salbutamol by Clinical Commissioning Groups (CCGs)

Data Preprocessing - Smoothing

- Less interested in short term fluctuations in number of prescriptions.
- More interested in the long-term trend - smoothing should help elucidate this.
- Saviztky-Golay filter - perform least squares fit of *n* consecutive data points to a polynomial and take calculated central point of fitted polynomial as new data point.



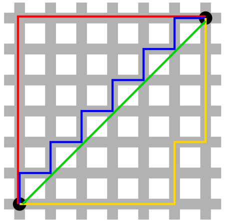


Hierarchical Clustering - Distance Metric

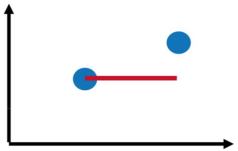
In order to decide which clusters should be combined, you have to use an appropriate distance metric - different metrics may cluster a given data point differently.

Euclidean - ordinary straight line distance between two points.

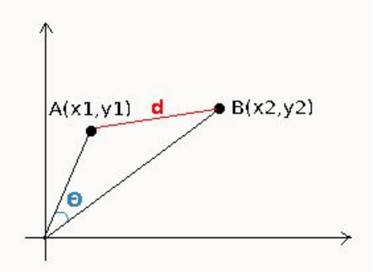
Manhattan - sum of the absolute differences of Cartesian coordinates



Chebyshev - greatest distance along individual axis.



Cosine - Measure of orientation not magnitude



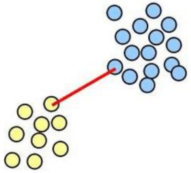
Hierarchical Clustering - Linkage Criteria

The linkage criterion determines the distance between sets of observations as a function of the pairwise distances between observations.

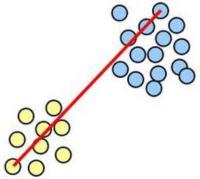
Single-linkage - combine two clusters that contain the closest pair of elements not yet belonging to the same cluster as each other.

Complete linkage - combine two clusters according to the longest distance between two points in each cluster.

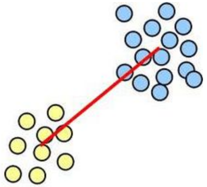
Average - look at the average distance between each point.



single-link



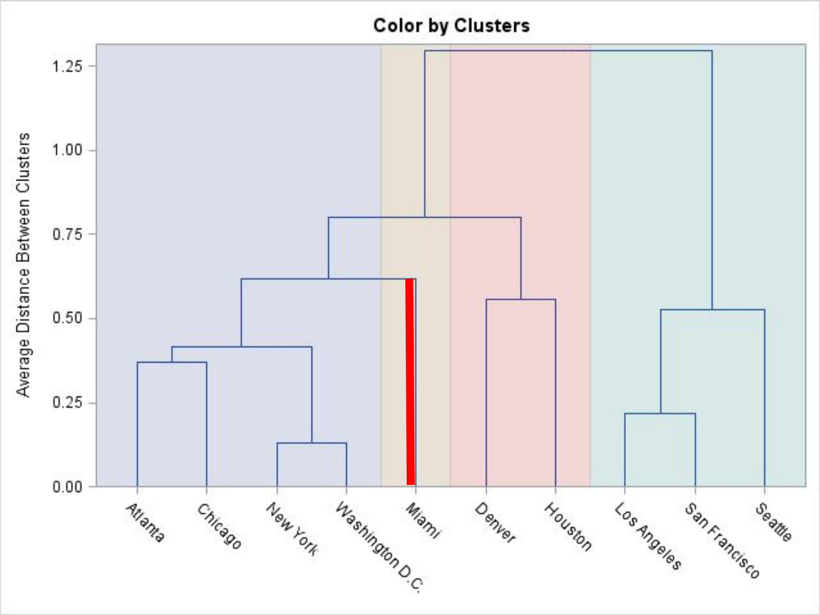
complete-link



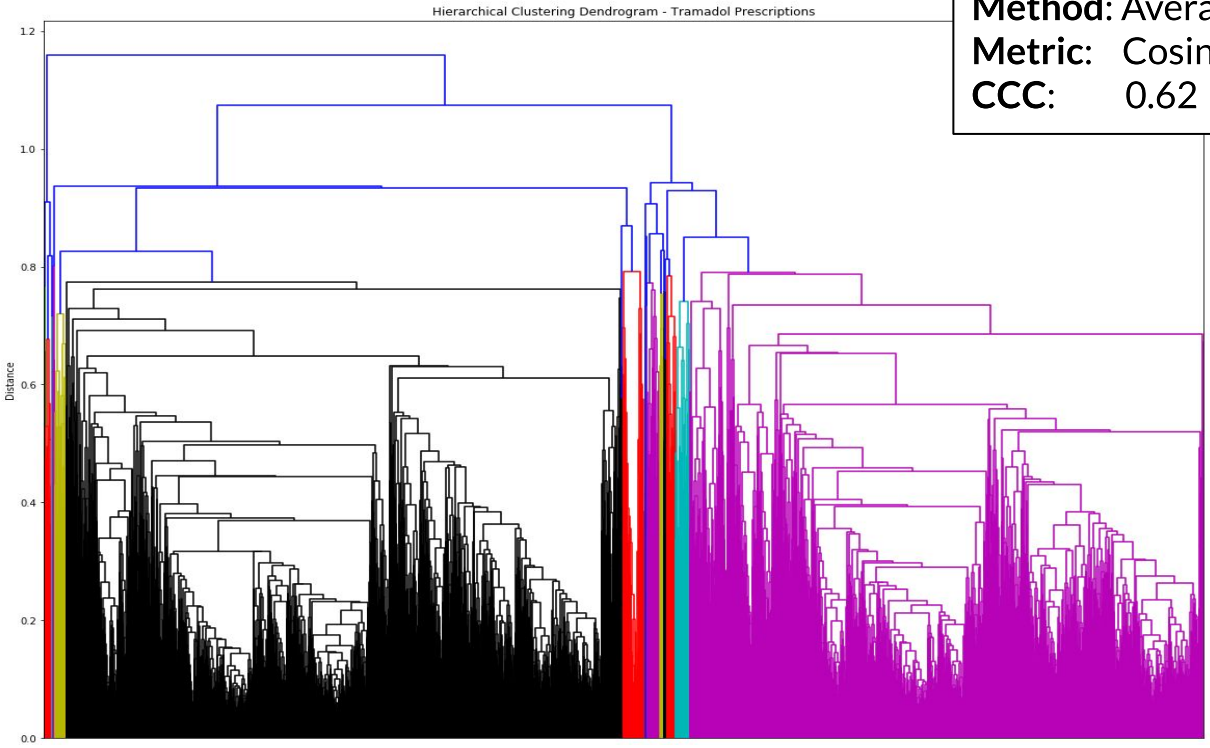
average-link

Cophenetic Correlation Coefficient (CCC)

- Cophenetic distance is given by the height at which observations are first joined in a dendogram.
- Cophenetic correlation is the linear correlation between the cophenetic distances from the dendrogram and the original distances used to build the dendrogram.
- Higher quality solution given by greater cophenetic correlation coefficient.



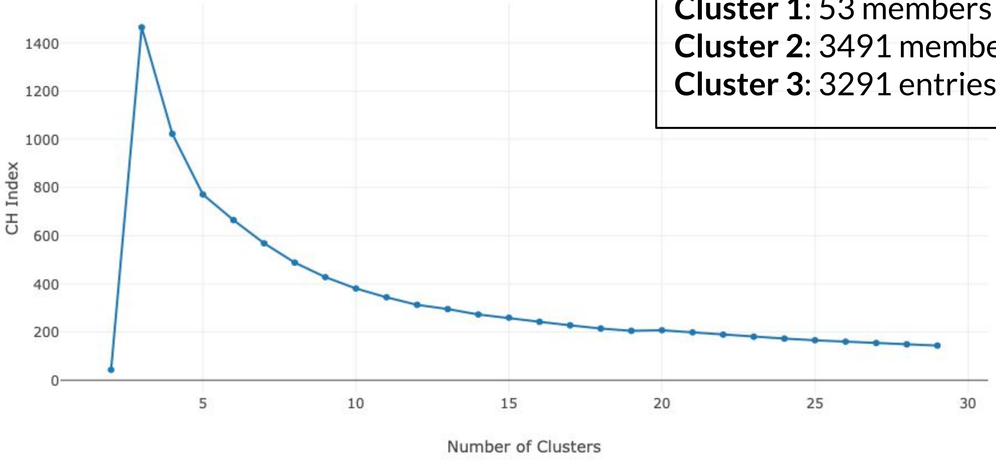
Clustering at Practice Level



Method: Average
Metric: Cosine
CCC: 0.62

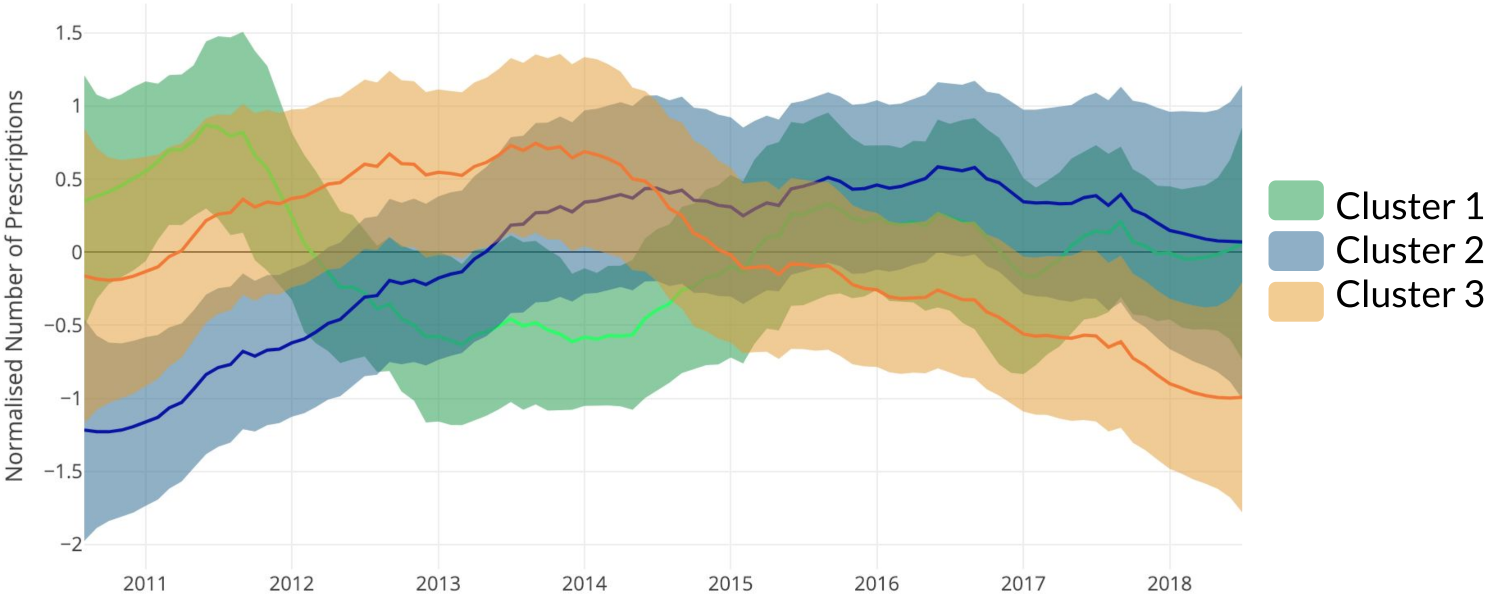
Calinski & Harabasz (CH) Index

- Looks at within cluster variation, *W*, and between cluster variation *B*.
- W measured based on sum of distances between objects and cluster center.
- B measured based on maximum distance between cluster centers.
- Looking to maximise the CH index.

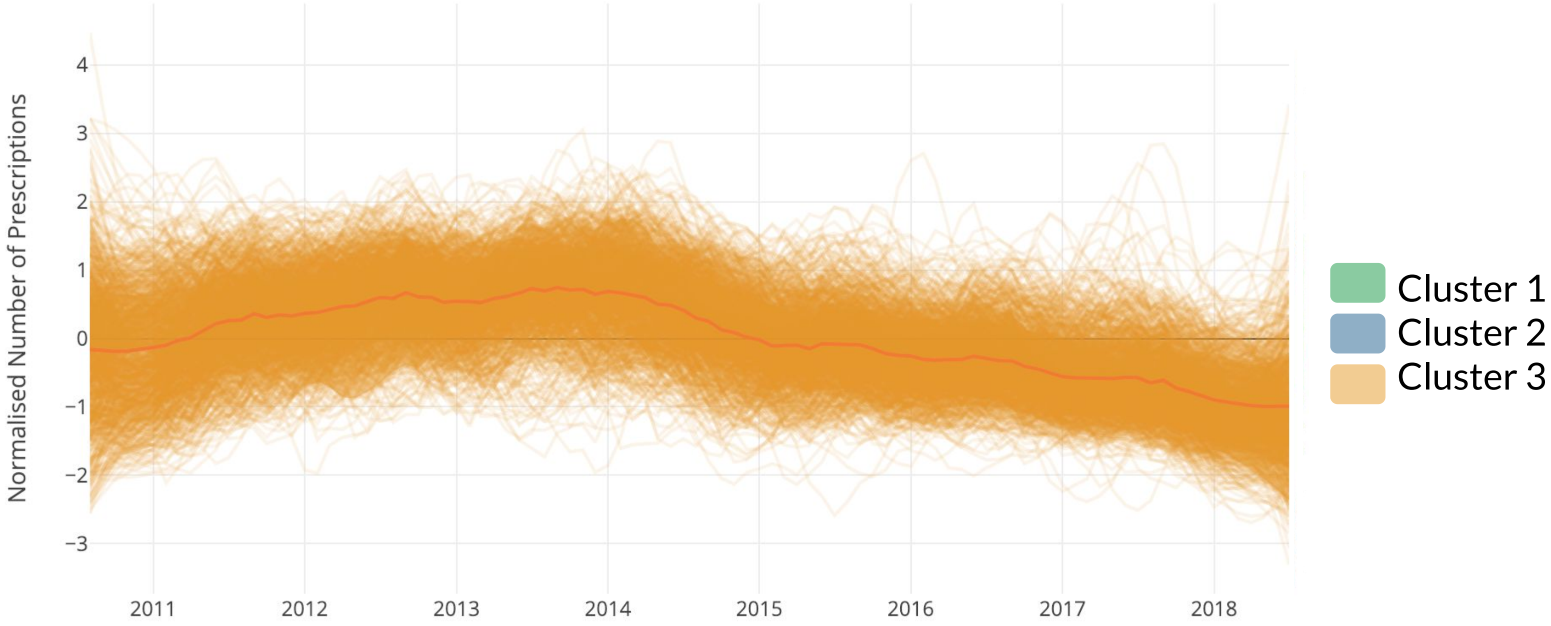


Cluster 1: 53 members
Cluster 2: 3491 members
Cluster 3: 3291 entries

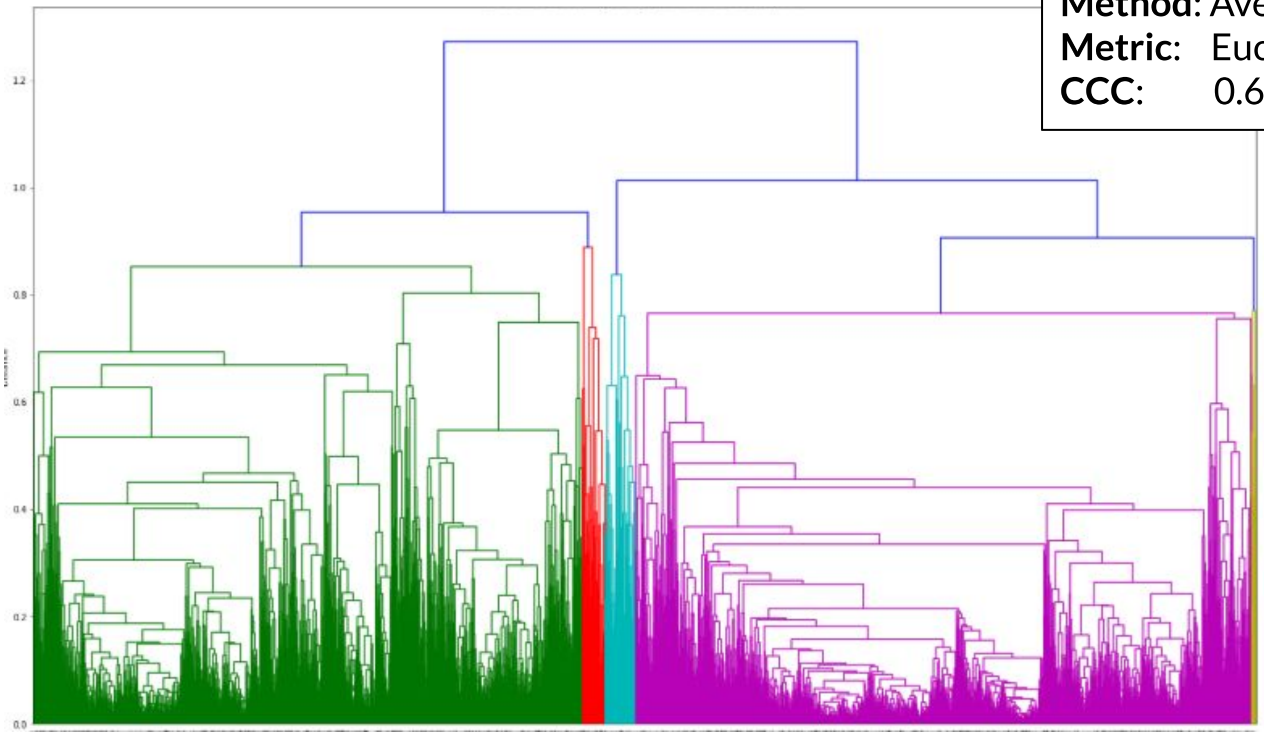
Visualising the Clusters (k=3)



Visualising the Clusters - k=3

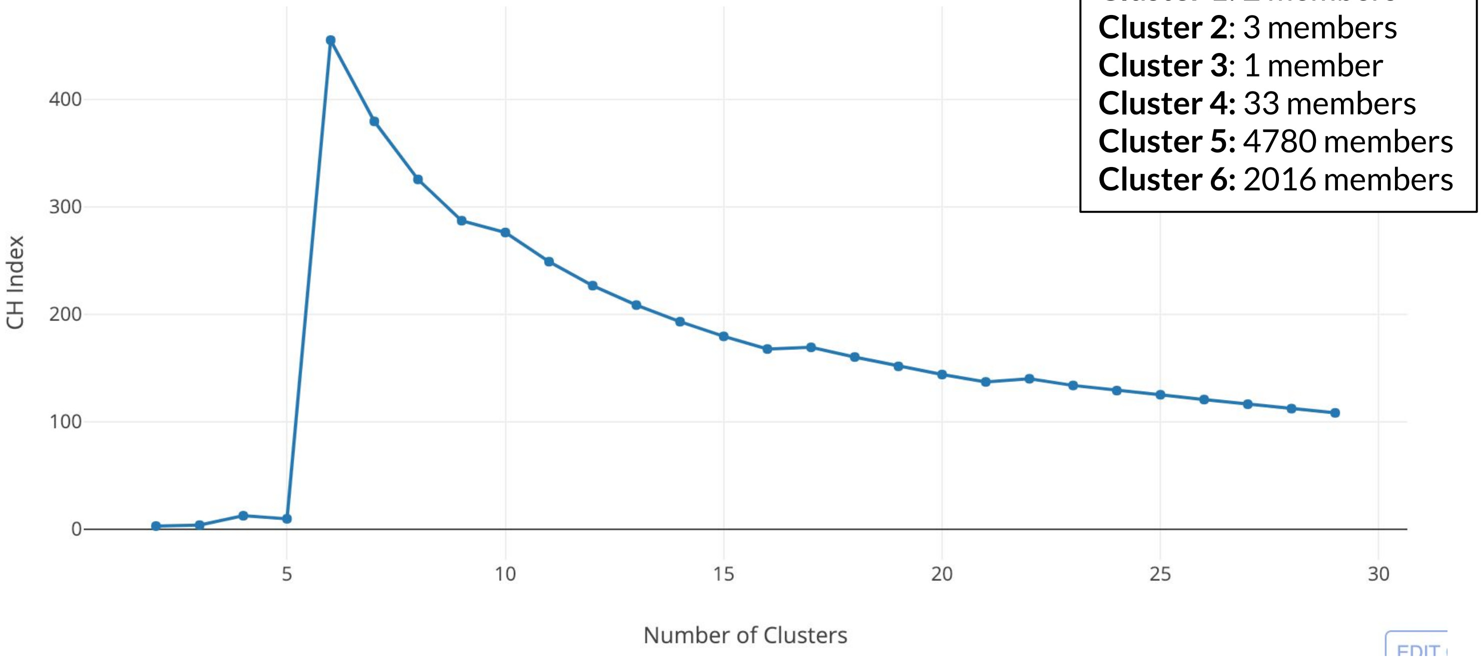


Clustering Shorter Period



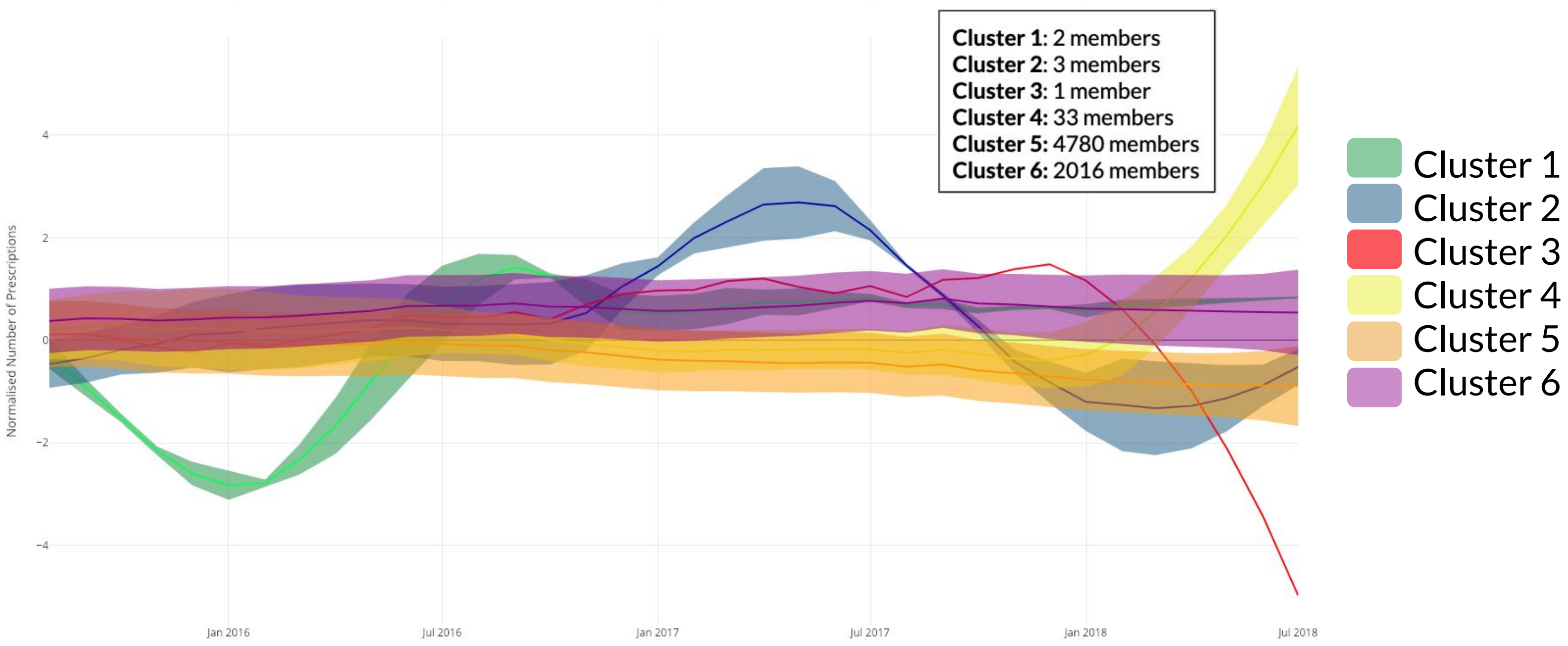
Method: Average
Metric: Euclidean
CCC: 0.66

CH Index



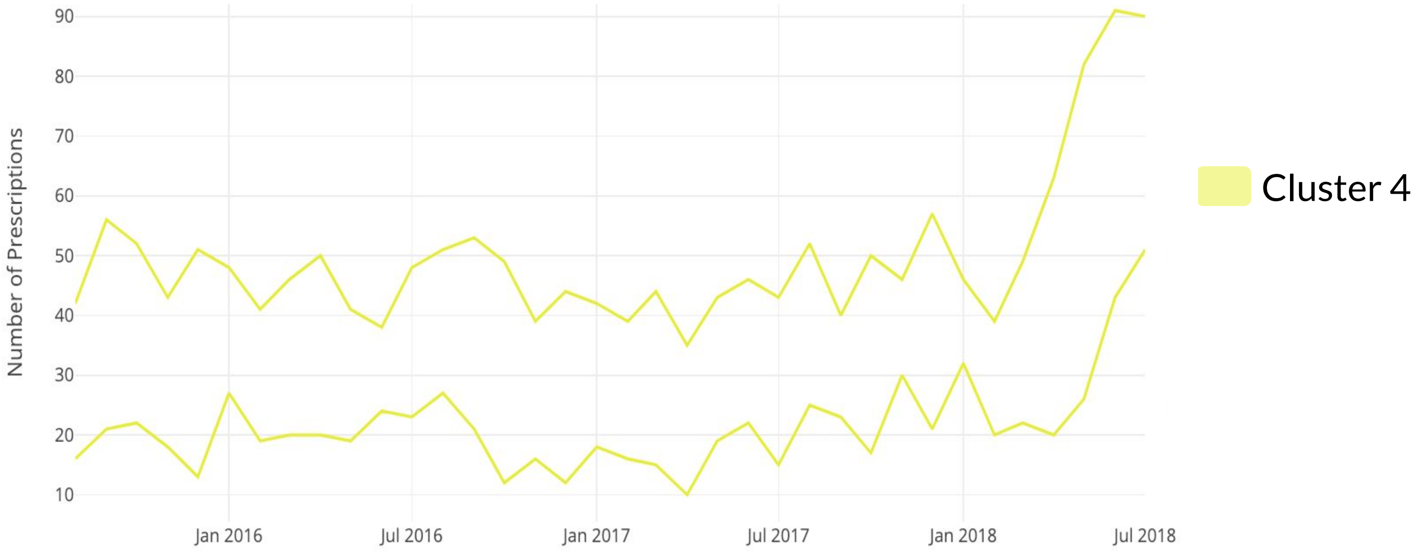
Cluster 1: 2 members
Cluster 2: 3 members
Cluster 3: 1 member
Cluster 4: 33 members
Cluster 5: 4780 members
Cluster 6: 2016 members

Visualising Clusters (k=6)



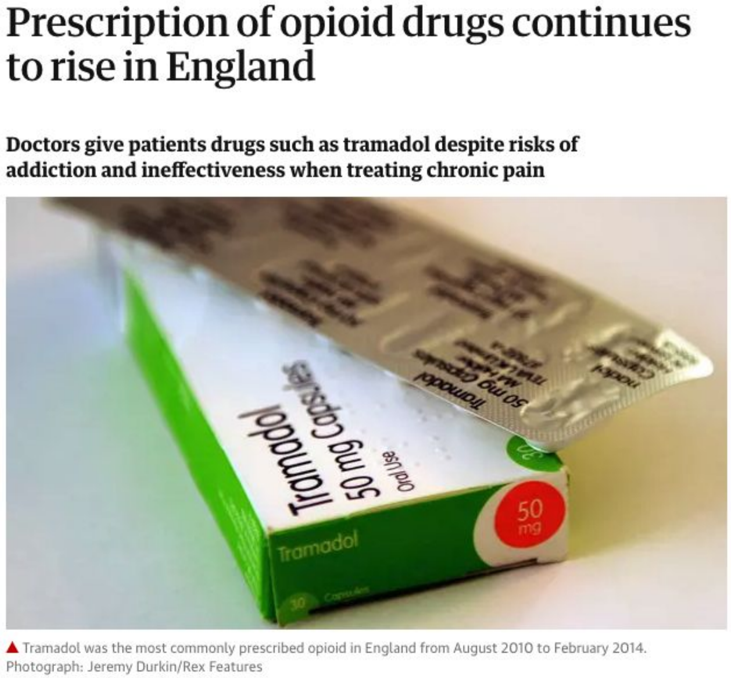
Cluster 1: 2 members
Cluster 2: 3 members
Cluster 3: 1 member
Cluster 4: 33 members
Cluster 5: 4780 members
Cluster 6: 2016 members

Cluster 4


Number of Prescriptions
90
80
70
60
50
40
30
20
10
Jan 2016
Jul 2016
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Cluster 4

Clinical Importance

“The prescription of opioid drugs by GPs in England is steadily rising, especially in more deprived communities” - *Guardian, Feb 2018*



Prescription of opioid drugs continues to rise in England

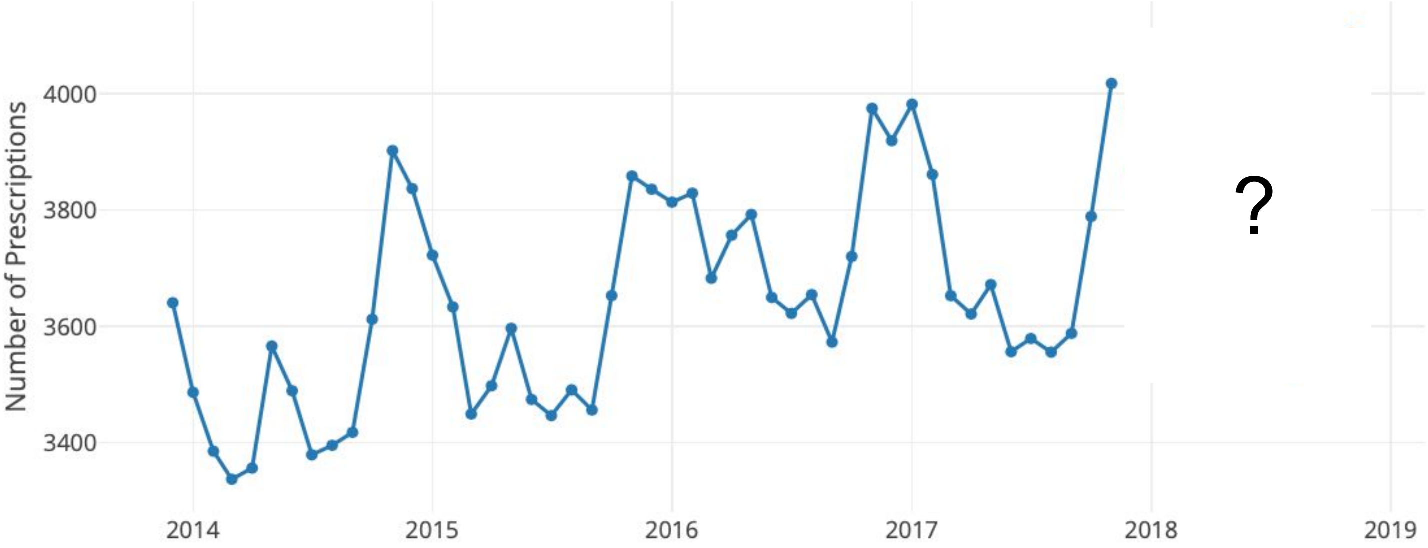
Doctors give patients drugs such as tramadol despite risks of addiction and ineffectiveness when treating chronic pain

▲ Tramadol was the most commonly prescribed opioid in England from August 2010 to February 2014. Photograph: Jeremy Durkin/Rex Features

Supervised Approach

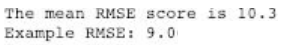
Time Series Forecasting

Framing as a Supervised Problem

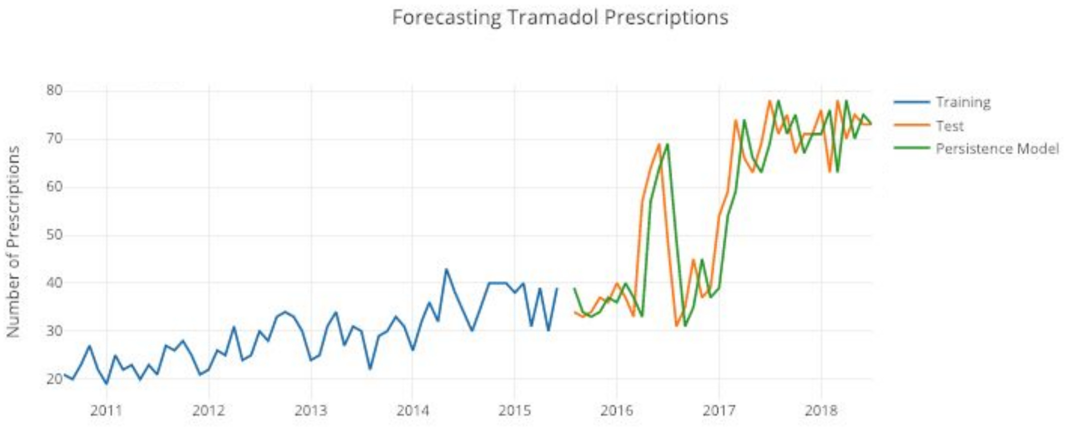


Setting a Baseline - Persistence Model

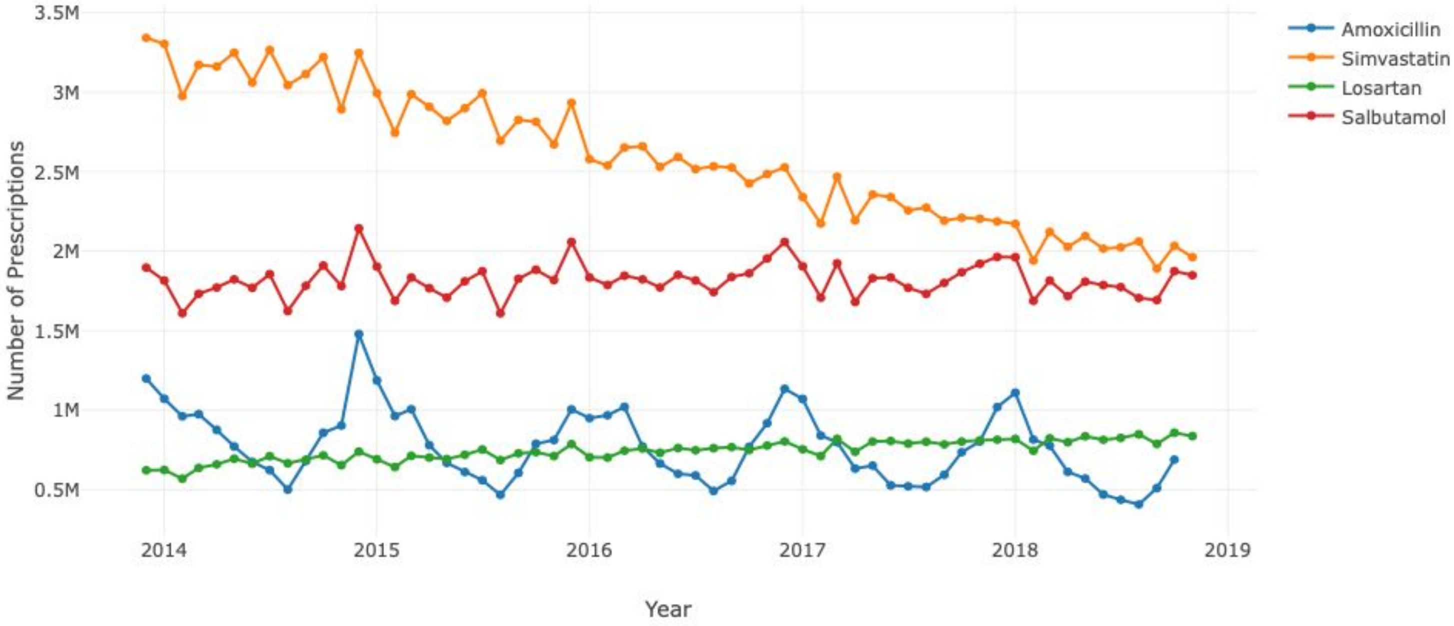
- The persistence forecast is where the observation from the prior time step (t-1) is used to predict the observation at the current time step (t).



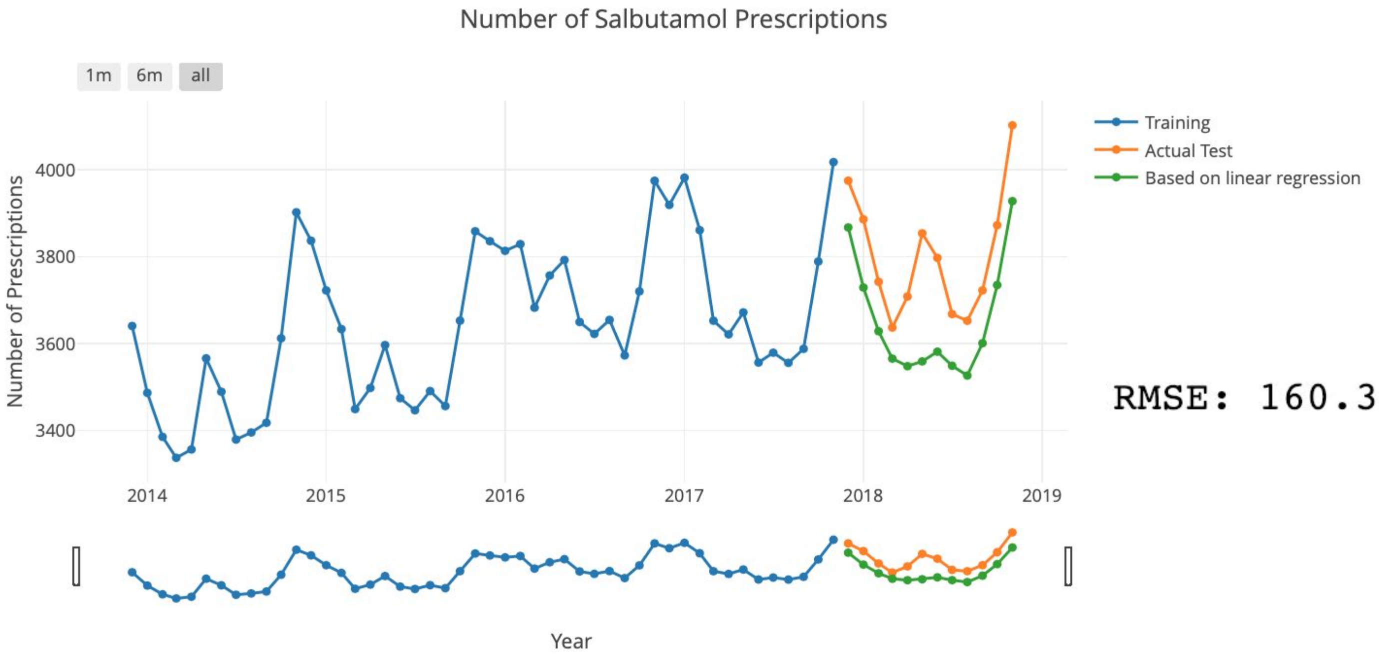
```
The mean RMSE score is 10.3
Example RMSE: 9.0
```



Linear Regression on Month

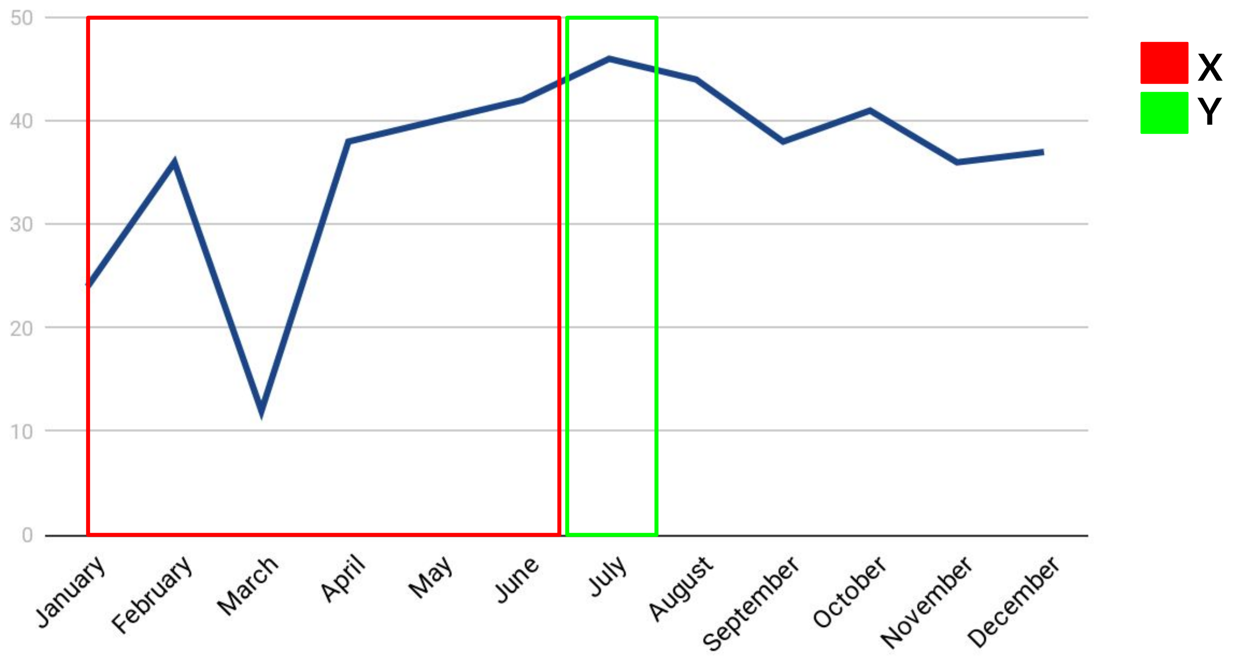


Linear Regression on Month

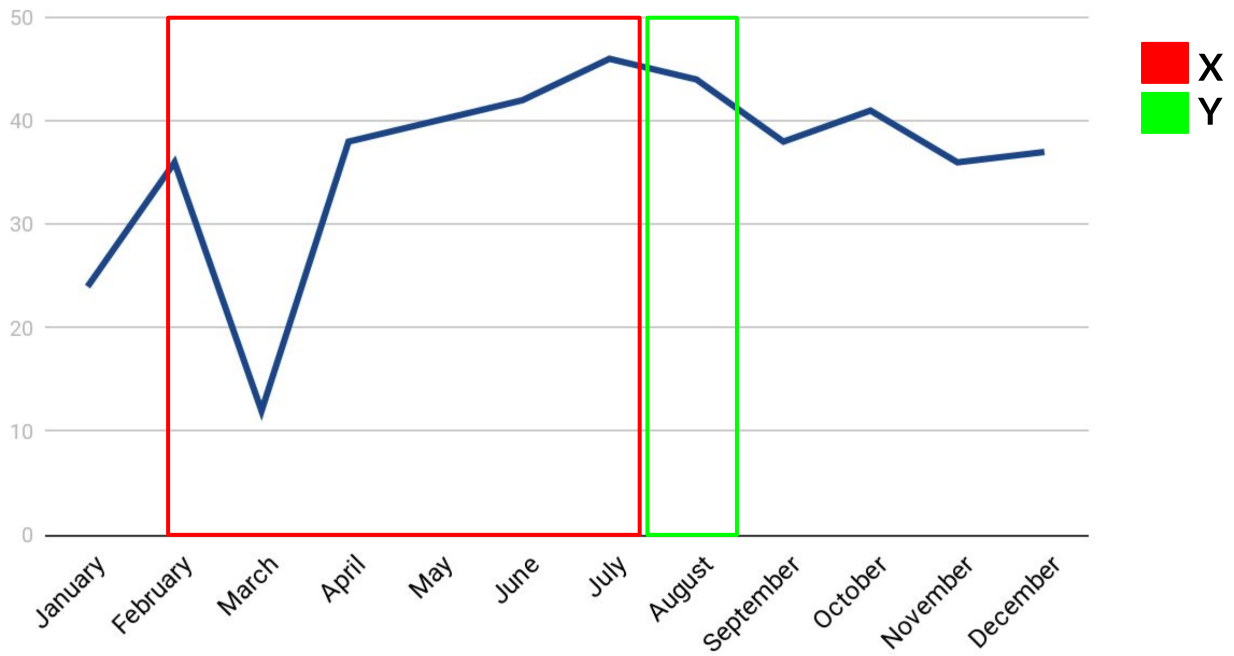


RMSE: 160.349716

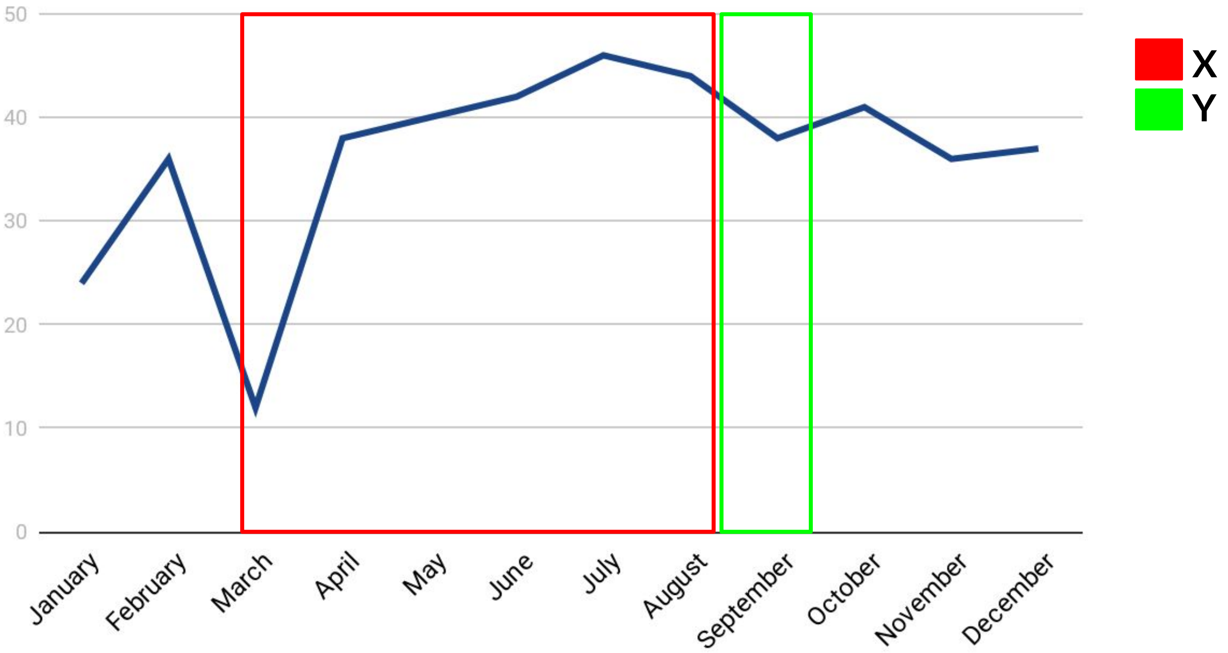
Sliding Window



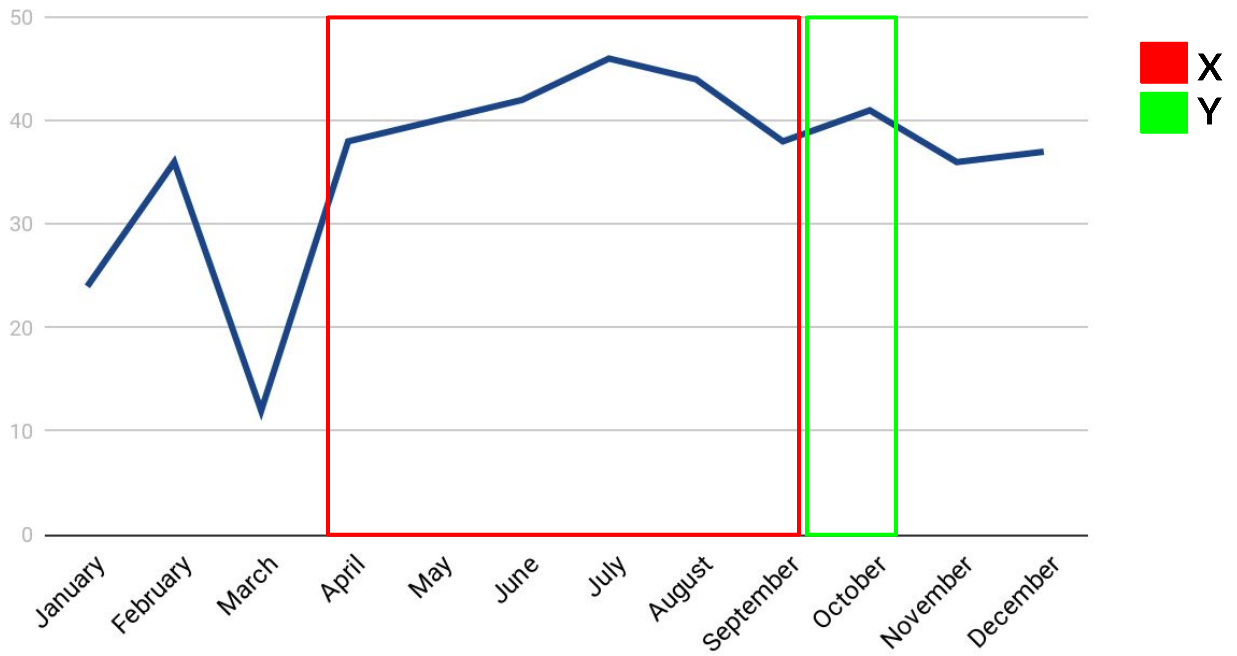
Sliding Window



Sliding Window



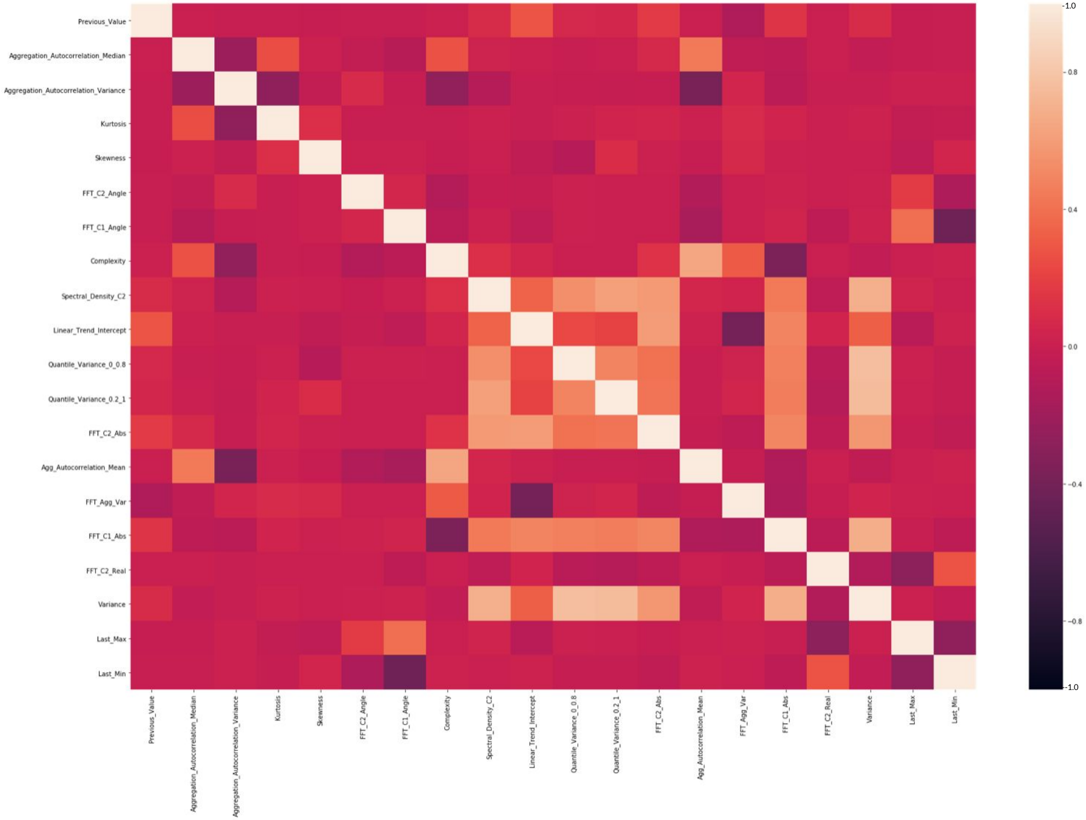
Sliding Window



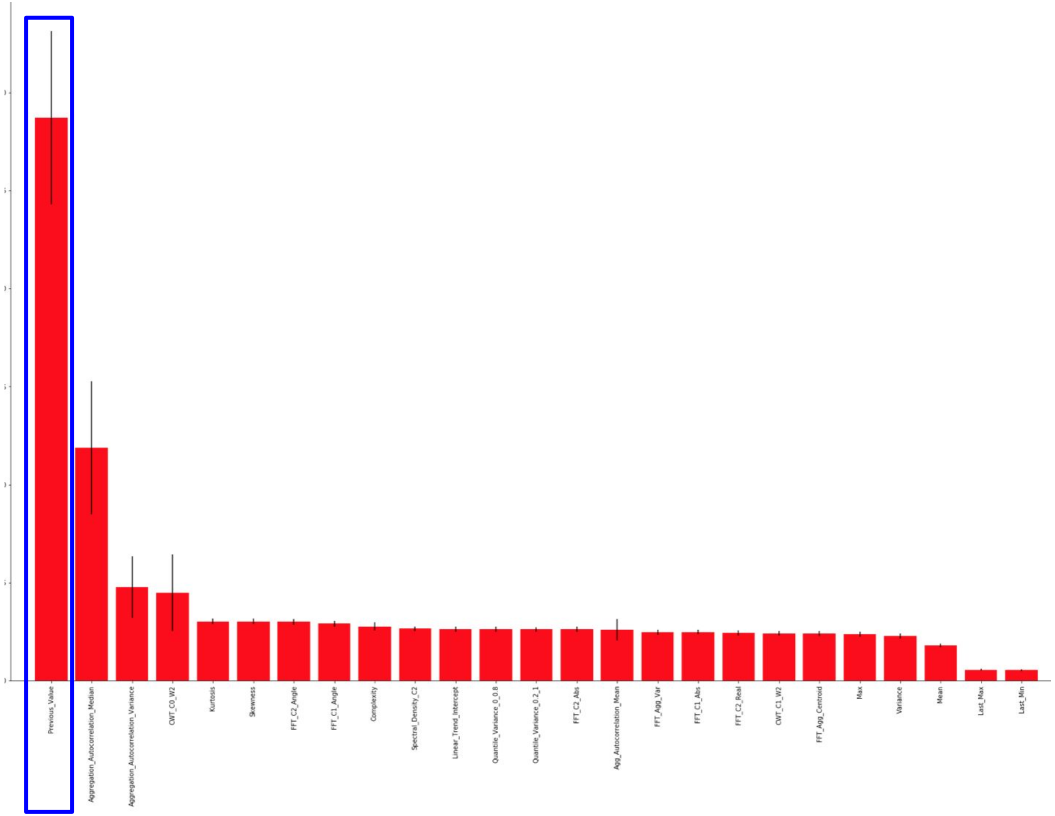
Feature Extraction

- 6835 time series for Tramadol prescriptions.
- For each, extracted 90 6-month windows - resulted in X with dimensions (615150, 6).
- Applied feature extraction to each 6 month period.
- Extracted features such as the mean, skewness, value of the 6th month, kurtosis, variance, number of peaks, continuous wavelet transform (decomposition into highly localised oscillations)
- Could then use these as feature inputs to supervised approaches.

Feature Correlation



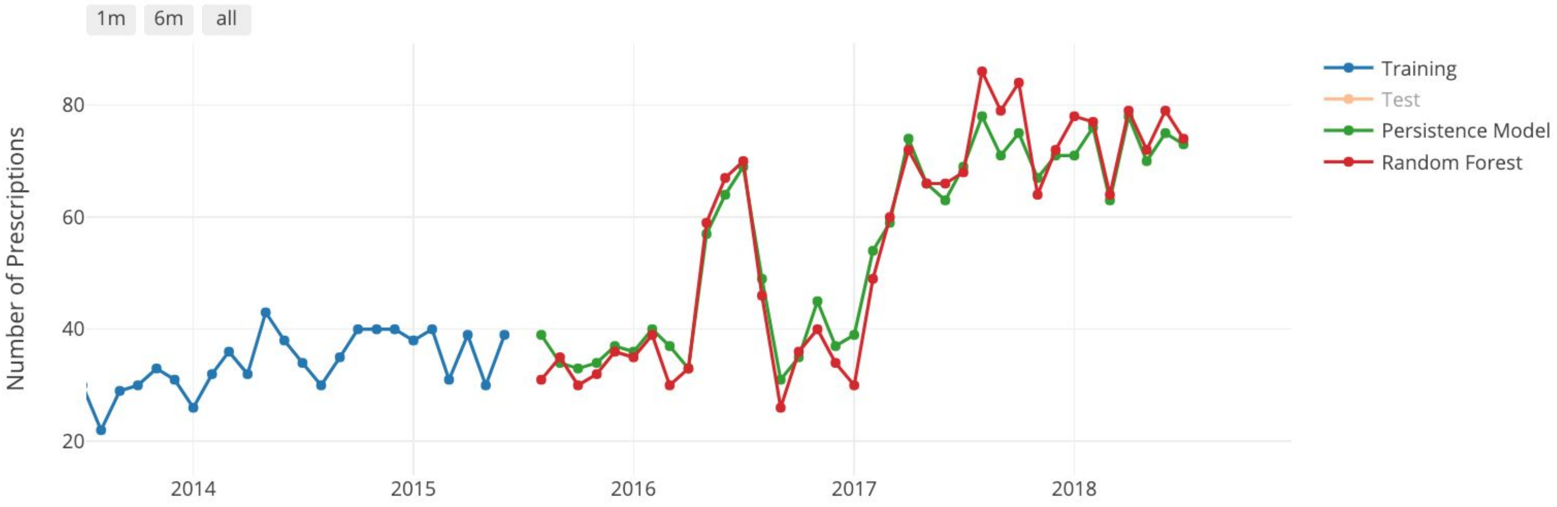
Feature Importance



Random Forest - 1 Month Forecast

RMSE:

Persistence:	10.3
Random Forest:	17.3



Alternative Approaches

Feature extraction

- Extract features for previous 3 years as well as previous six months.
- Look at multi-step forecasting (more than one month ahead).

Prediction Method

- Long-short-term-memory recurrent neural networks.
- 1D convolution for feature extraction.

Summary

- Using appropriate metrics and methods, you can successfully cluster drug prescribing trends.
- Normalising and smoothing of time series allows you to look at clustering by long-term trend.
- Clustering can elucidate potentially interesting findings in drug prescribing trends.