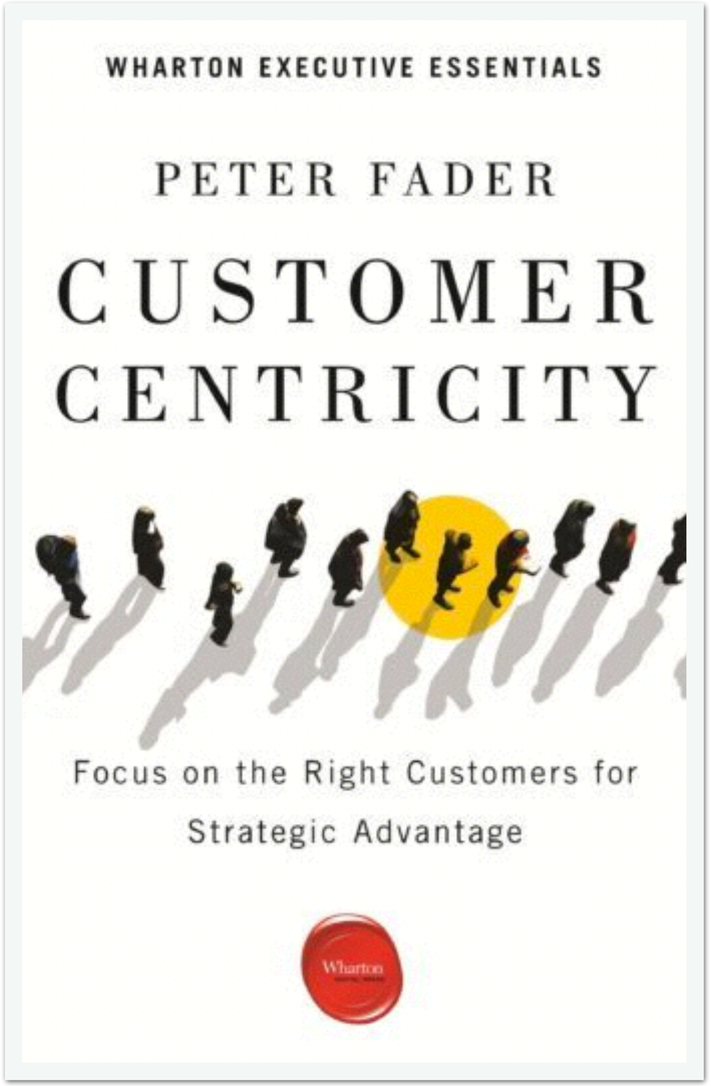
Customer centric data science

Ewan Nicolson
Principal data scientist
Skyscanner



Customers are individuals

Customers are different to each other

Not about customer service — that's separate dimension

Great approach for data science

- Wonderful dataset to work with
- Solve problems in novel ways
- Solve a real human problem

How to get a customer centric dataset

(user id_0, $timestamp_0$, event $details_0$)

(user id_1, $timestamp_1$, event $details_1$)

(user id_2, $timestamp_2$, event $details_2$)

....

(user id_N, $timestamp_N$, event $details_N$)

(user id_0, $timestamp_0$, event $details_0$)

(user id_1, $timestamp_1$, event $details_1$)

(user id_2, $timestamp_2$, event $details_2$)

....

(user id_N, $timestamp_N$, event $details_N$)

Strong user ID

(user id_0, $timestamp_0$, event $details_0$)

(user id_1, $timestamp_1$, event $details_1$)

(user id_2, $timestamp_2$, event $details_2$)

....

(user id_N, $timestamp_N$, event $details_N$)

Every interaction

(user id_0, $timestamp_0$, $event\ details_0$)

(user id_1, $timestamp_1$, $event\ details_1$)

(user id_2, $timestamp_2$, $event\ details_2$)

....

(user id_N, $timestamp_N$, $event\ details_N$)

Storage

(user id_0, $timestamp_0$, event $details_0$)

(user id_1, $timestamp_1$, event $details_1$)

(user id_2, $timestamp_2$, event $details_2$)

....

(user id_N, $timestamp_N$, event $details_N$)

No silos

(user id_0, $timestamp_0$, $event\ details_0$)

(user id_1, $timestamp_1$, $event\ details_1$)

(user id_2, $timestamp_2$, $event\ details_2$)

....

(user id_N, $timestamp_N$, $event\ details_N$)

Consent to use

(user id_0, $timestamp_0$, event $details_0$)

(user id_1, $timestamp_1$, event $details_1$)

(user id_2, $timestamp_2$, event $details_2$)

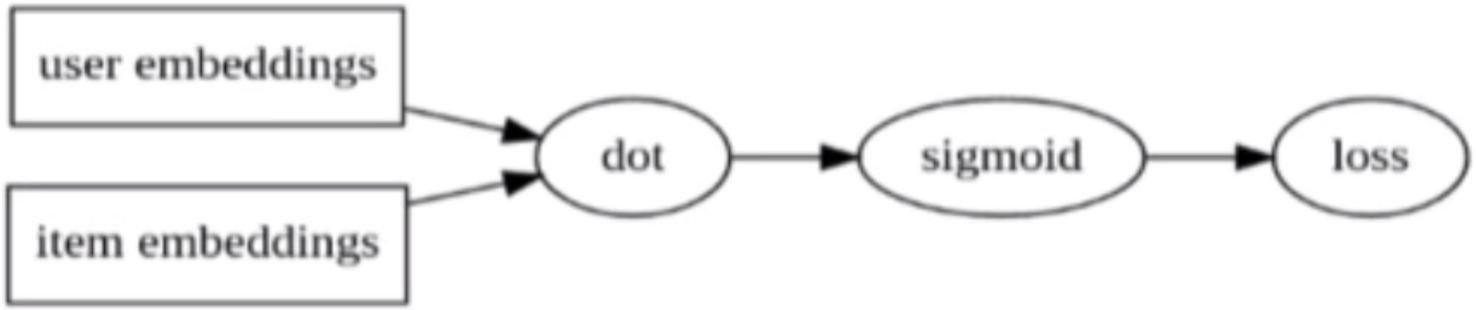
....

(user id_N, $timestamp_N$, event $details_N$)

Avoid personal data

Why is this so great?

Perfect dataset for personal recommendations



Embeddings

https://www.youtube.com/watch?v=ZkBQ6YA9E40

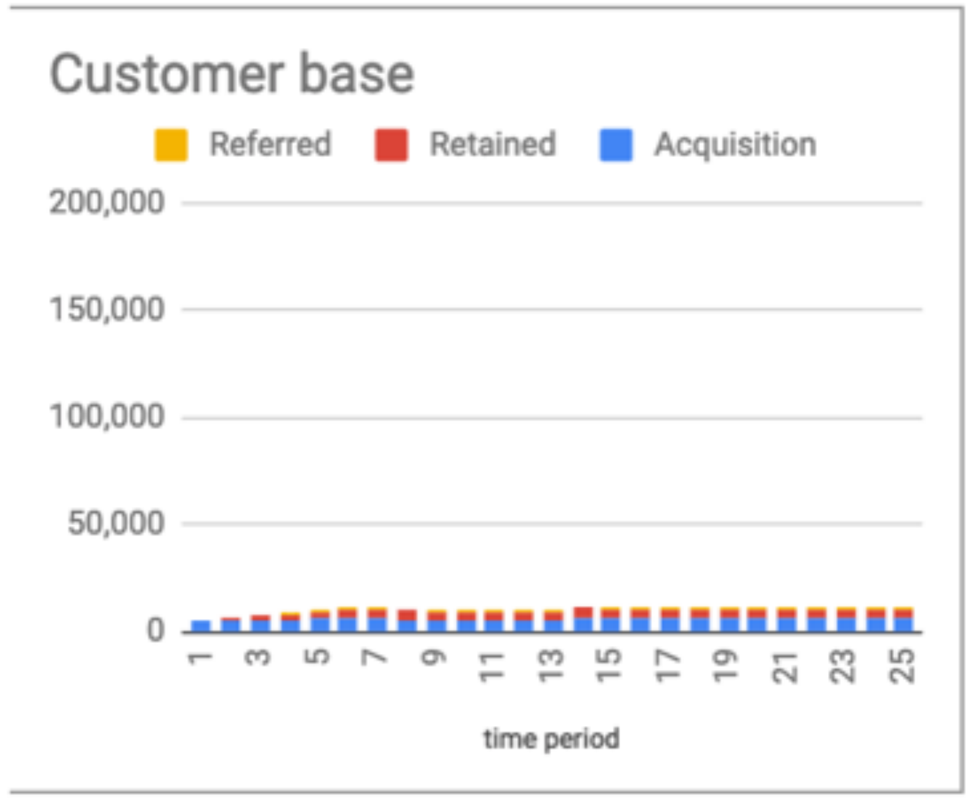
In [19]:

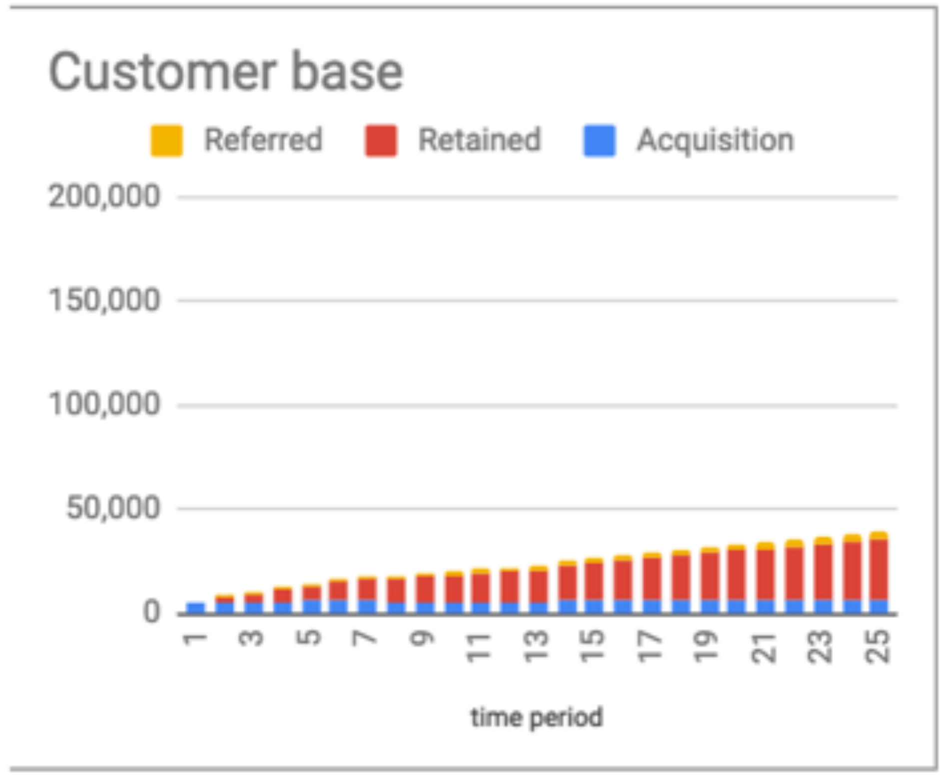
```
movie = model.get_layer('movie_embedding')
movie_weights = movie.get_weights()[0]
movie_lengths = np.linalg.norm(movie_weights, axis=1)
normalized_movies = (movie_weights.T / movie_lengths).T

def similar_movies(movie):
    dists = np.dot(normalized_movies, normalized_movies[movie_to_idx[movie]])
    closest = np.argsort(dists)[-10:]
    for c in reversed(closest):
        print(c, movies[c][0], dists[c])

similar_movies('Rogue One')
```

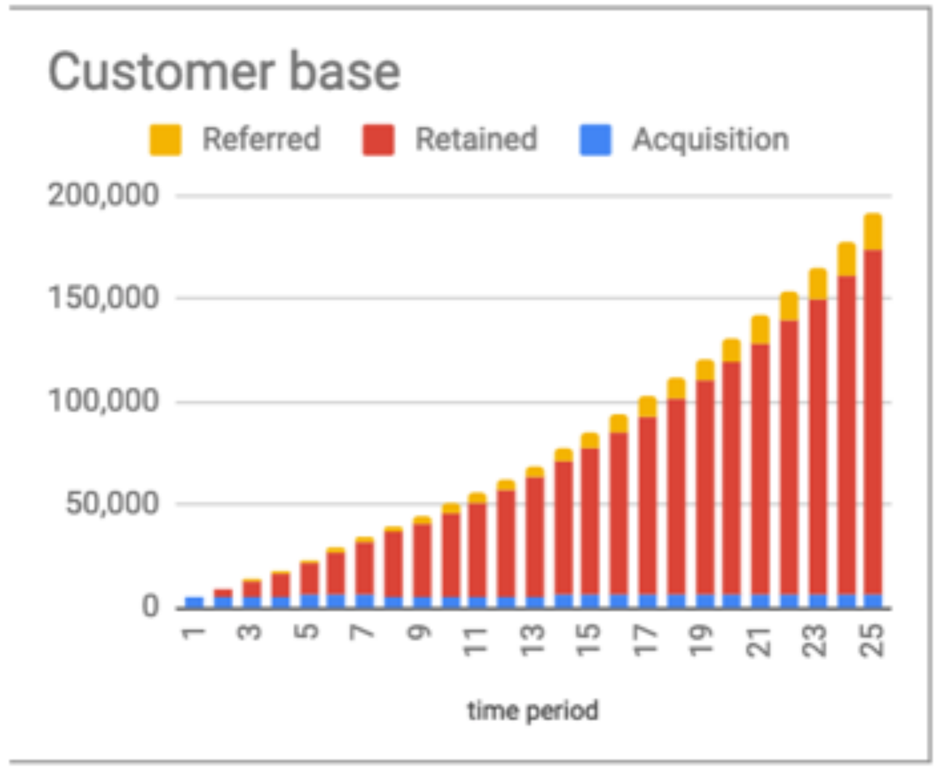
```
29 Rogue One 0.9999999
3349 Star Wars: The Force Awakens 0.9722805
101 Prometheus (2012 film) 0.9653338
140 Star Trek Into Darkness 0.9635347
22 Jurassic World 0.962336
25 Star Wars sequel trilogy 0.95218825
659 Rise of the Planet of the Apes 0.9516557
62 Fantastic Beasts and Where to Find Them (film) 0.94662267
42 The Avengers (2012 film) 0.94634
37 Avatar (2009 film) 0.9460137
```

https://github.com/dataewan/filmrecommendations-trai

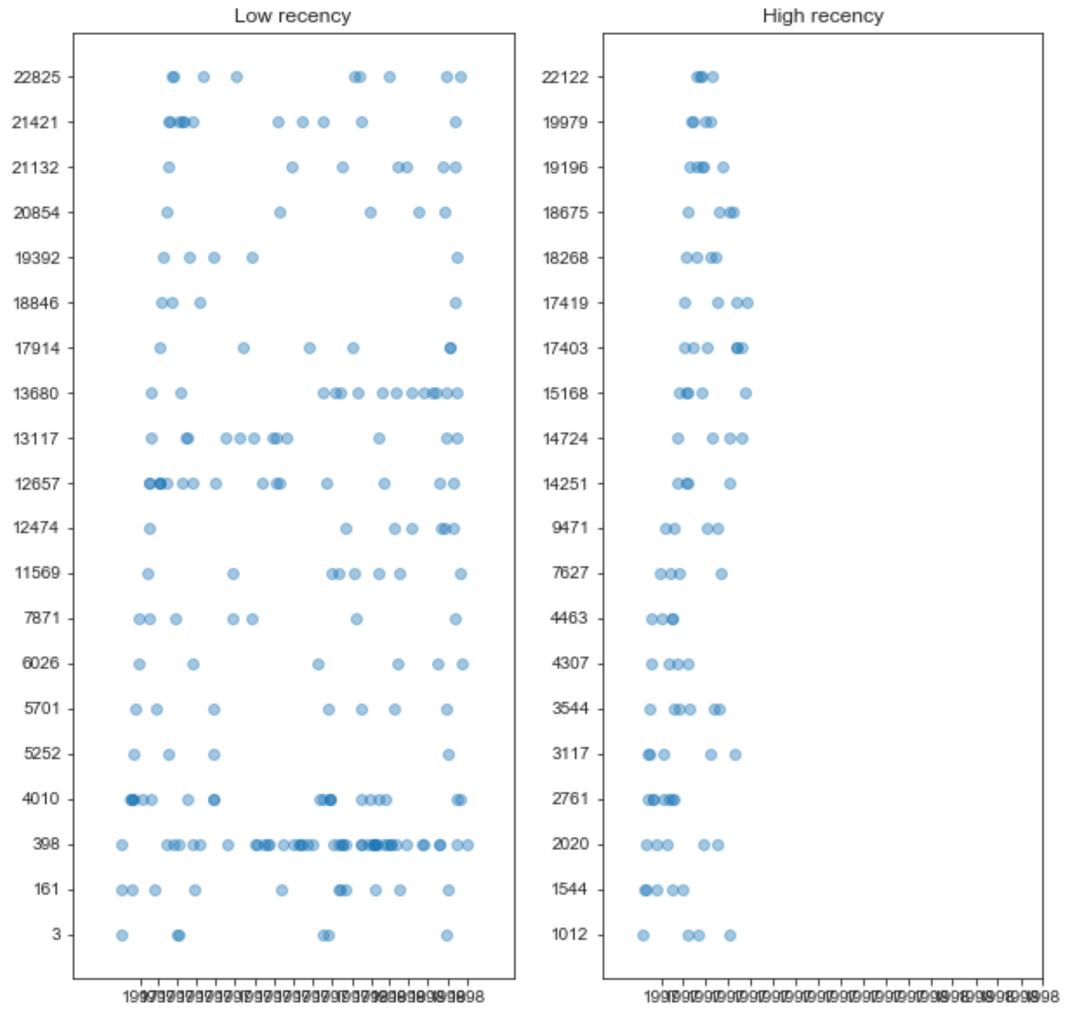
Understand different longitudinal behaviours





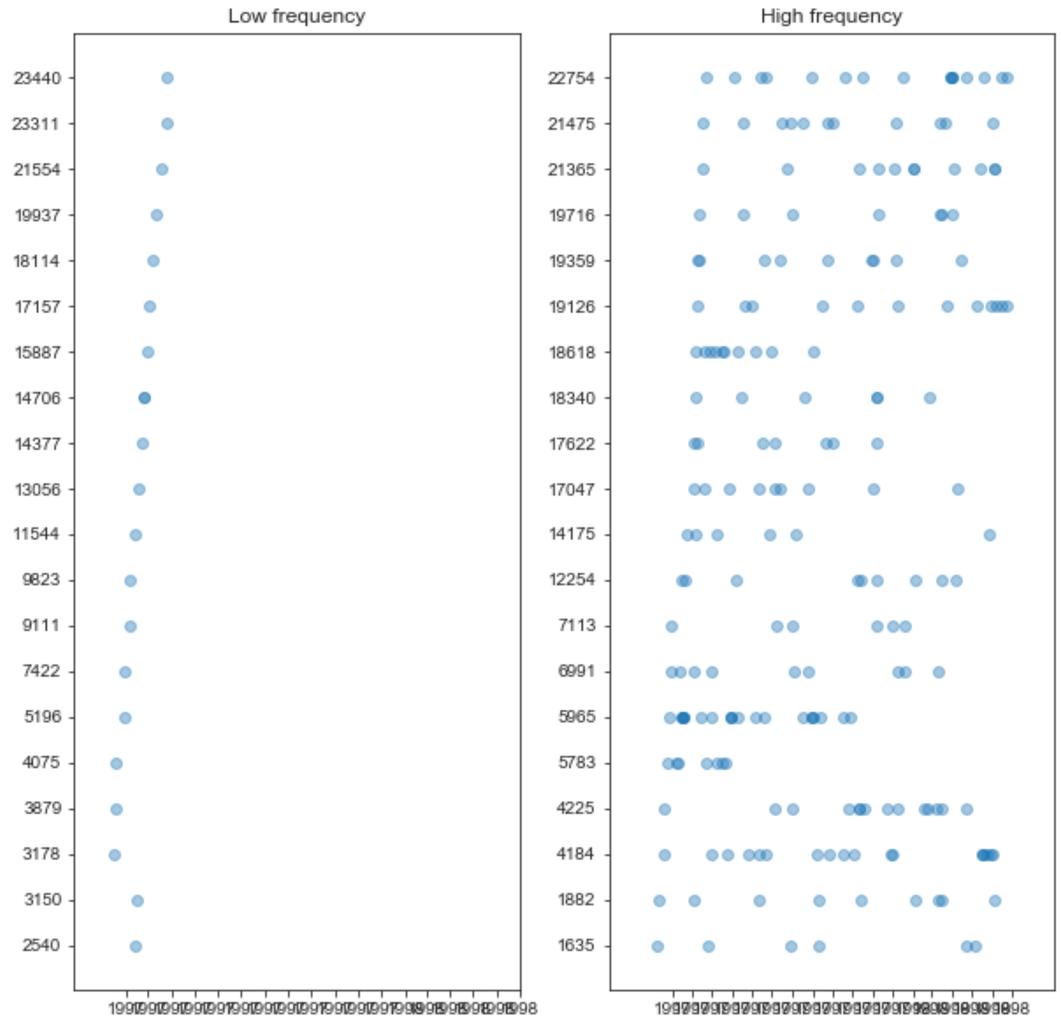


Long term retention means growth



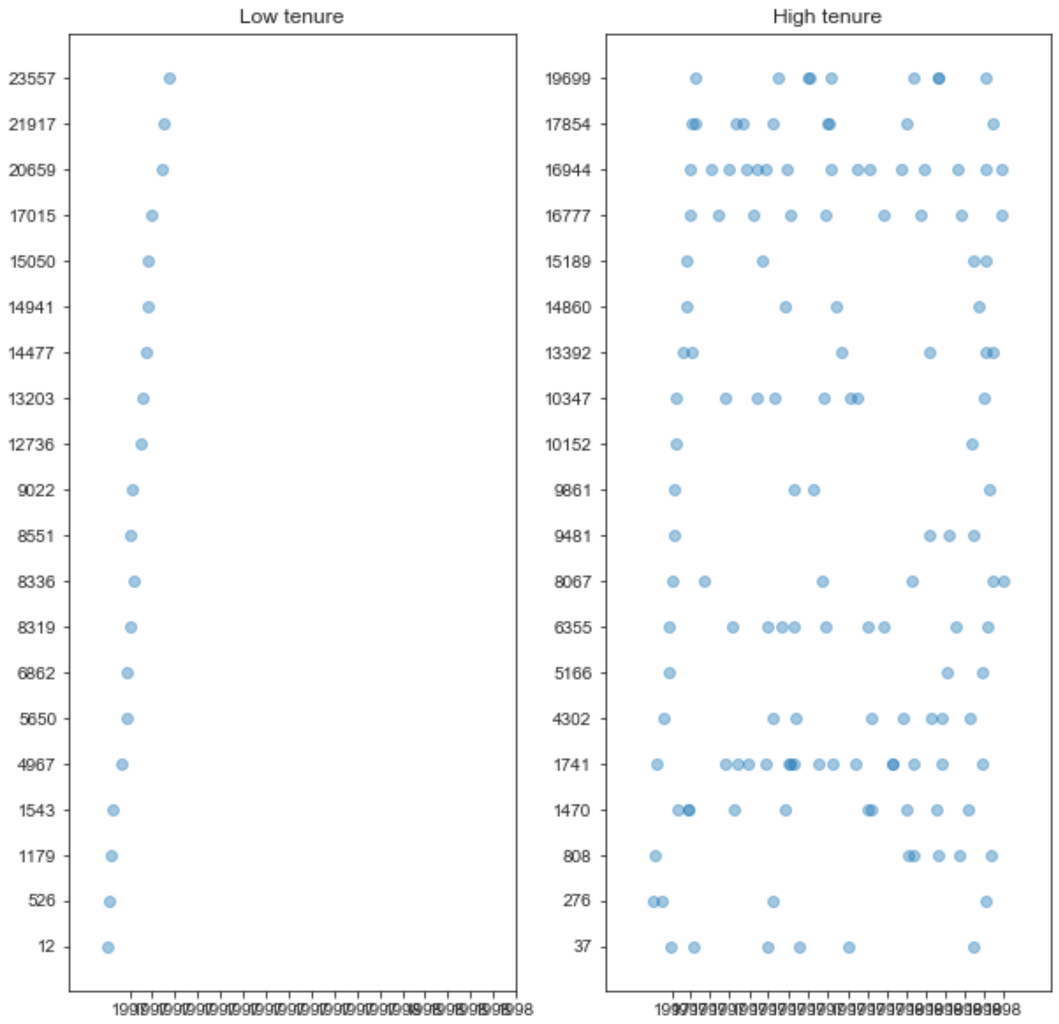
Recency

https://github.com/dataewan/customer-centric-ds/blob/master/notebooks/exploratory_cdwow.ipynb



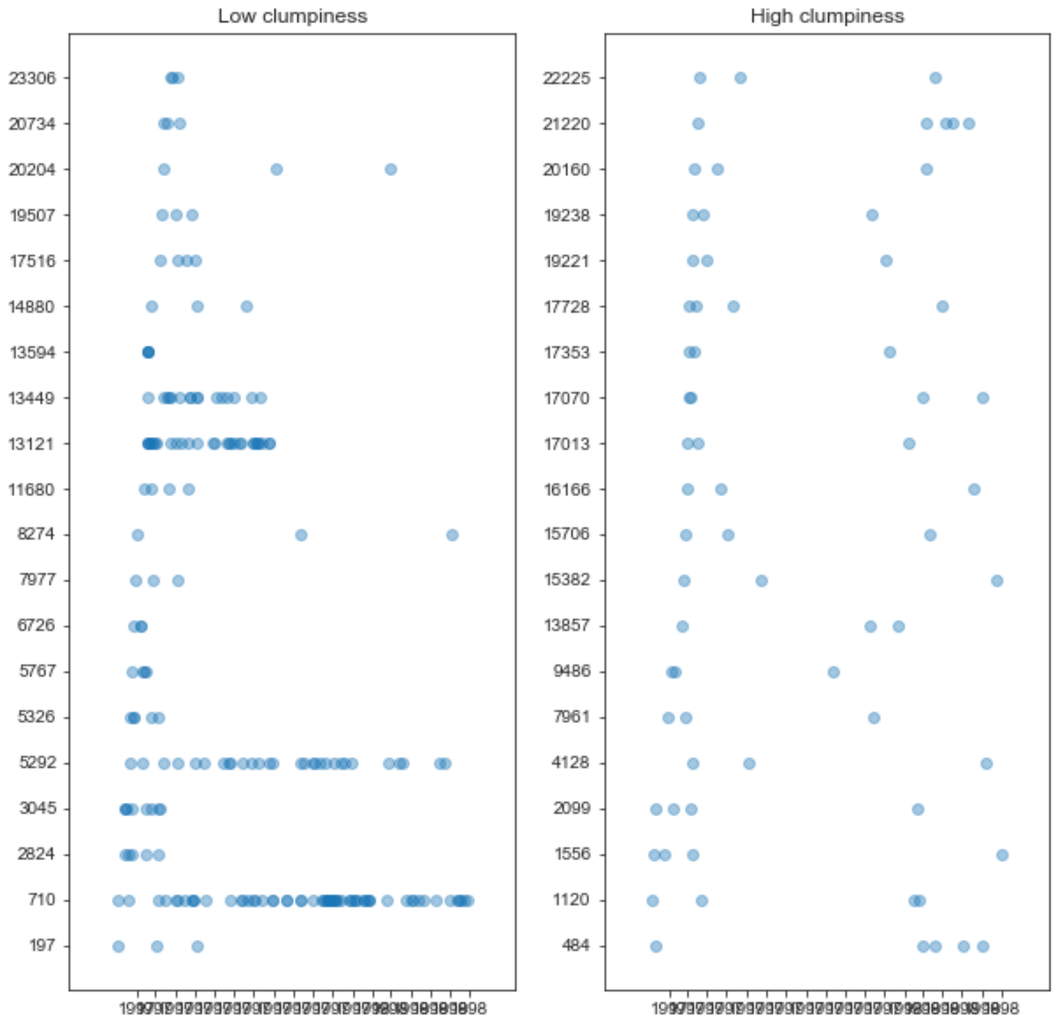
Frequency

https://github.com/dataewan/customer-centric-ds/blob/master/notebooks/exploratory_cdwow.ipynb



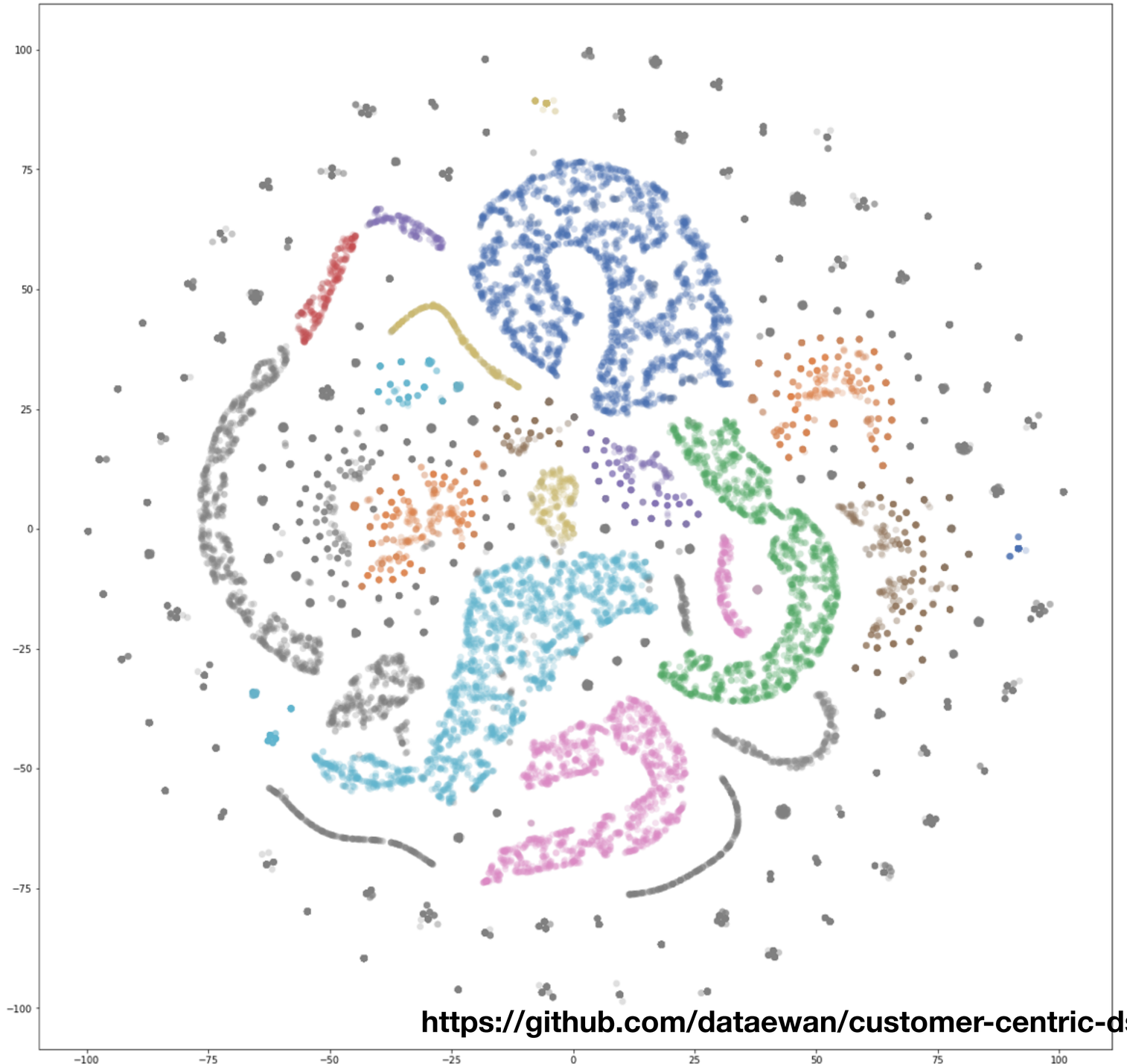
Tenure

https://github.com/dataewan/customer-centric-ds/blob/master/notebooks/exploratory_cdwow.ipynb



Clumpiness

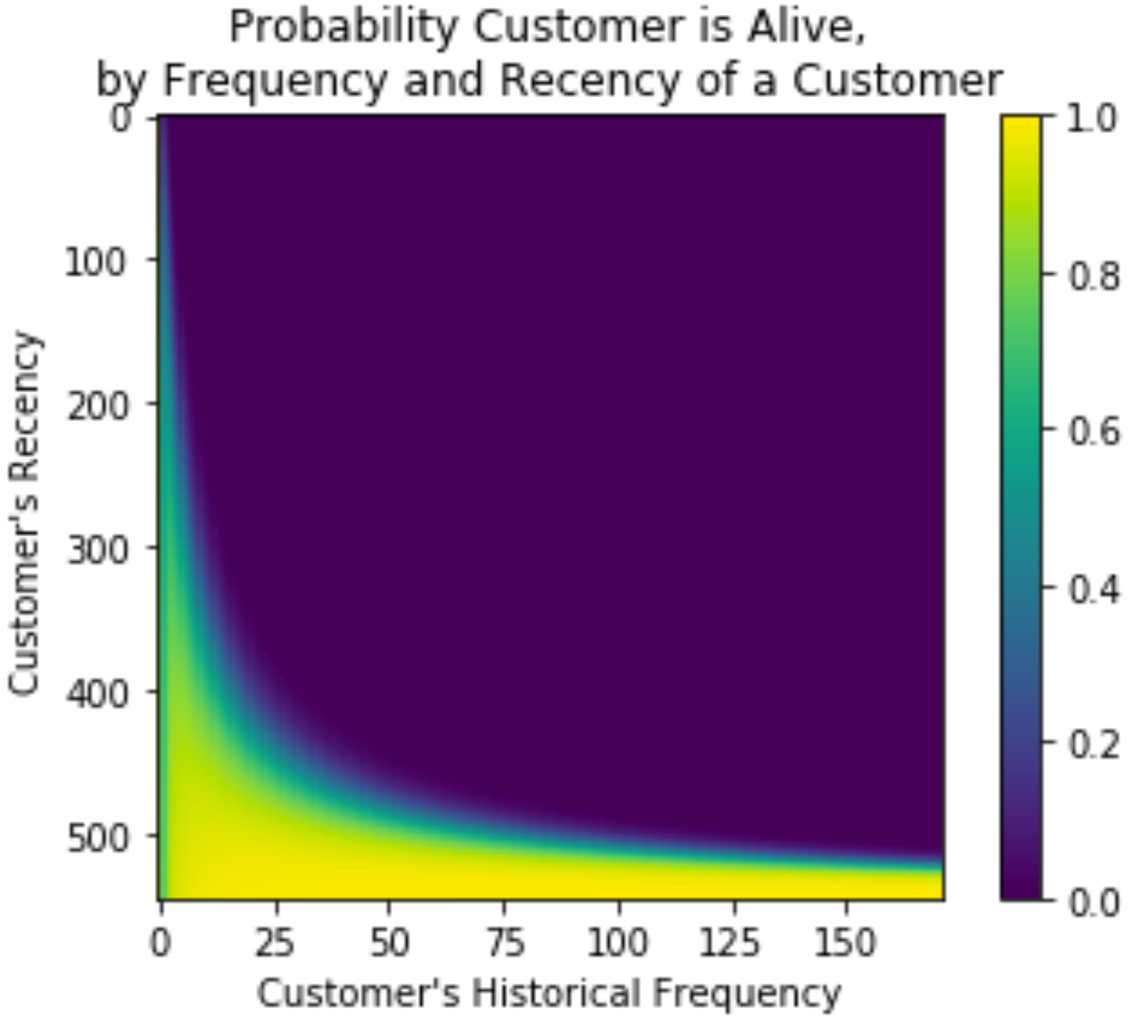
https://github.com/dataewan/customer-centric-ds/blob/master/notebooks/exploratory_cdwow.ipynb



Customer segments

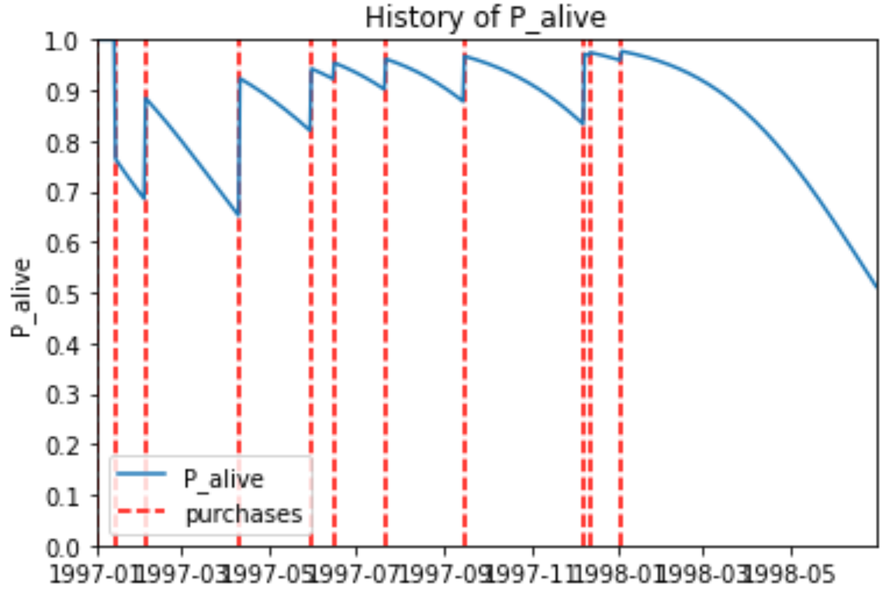
https://github.com/dataewan/customer-centric-ds/blob/master/notebooks/segmentation_cdwow.ipynb

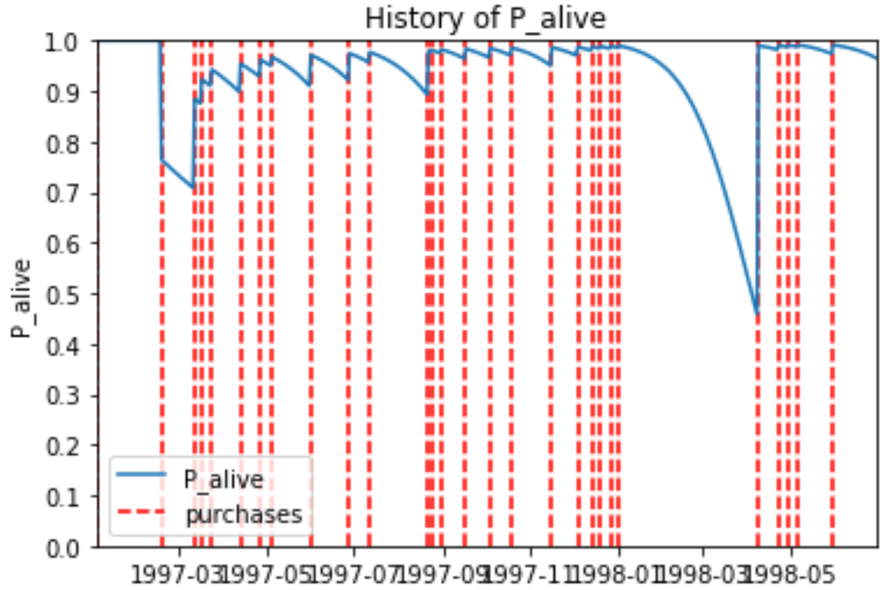
Customer lifetimes



Predicting probability of churn

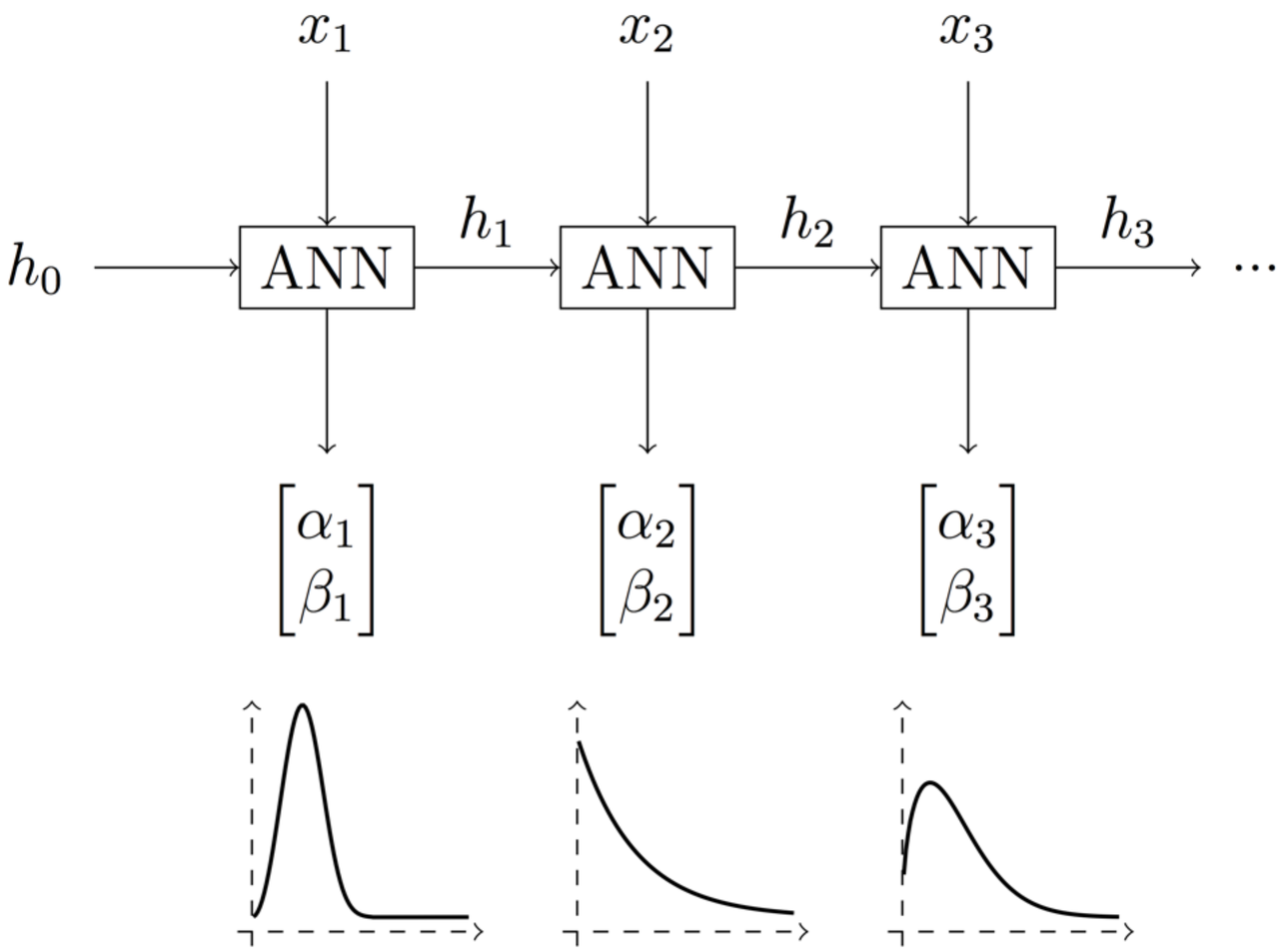
https://github.com/dataewan/customer-centric-ds/blob/master/notebooks/lifetimes.ipynb





Predicting probability of churn

https://github.com/dataewan/customer-centric-ds/blob/master/notebooks/lifetimes.ipynb



Recurrent neural networks for churn

https://ragulpr.github.io/2016/12/22/WTTE-RNN-Hackless-churn-modeling/

Customer Centric Data Science

Good data science perspective

Solves a real human problem