Identifying customers at risk of churn

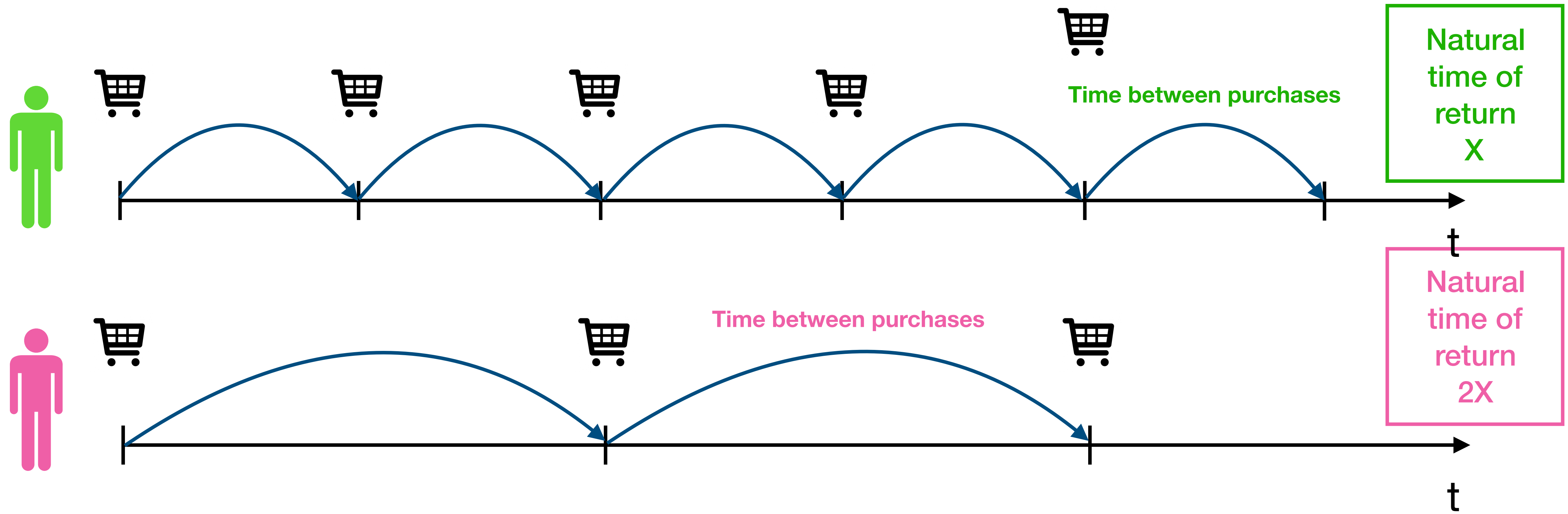
University of Essex - Profusion partnership

Agenda

- **Motivation**: Customer targeting
- **Methodology**: How it was done?
    - **Segmentation:** Customer behaviour
    - **Identification**: Customer Targeting
- **Results:** **Backtesting**

Motivation

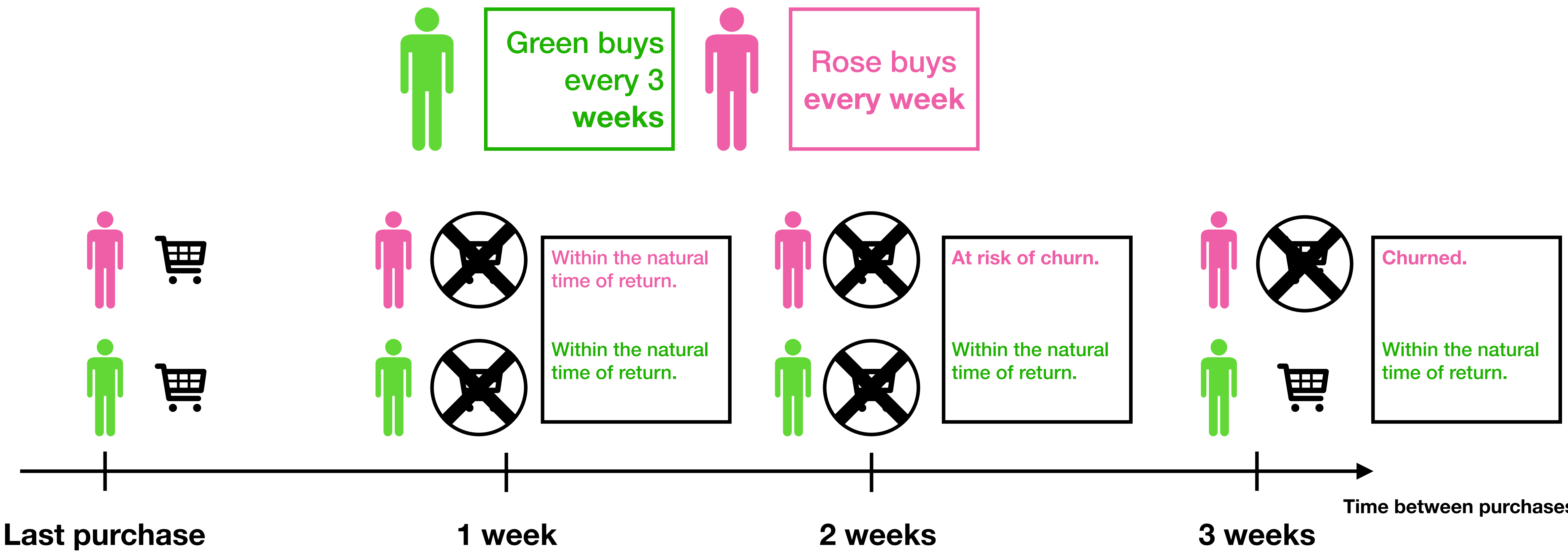
Each **customer** tends to follow a **distinct behaviour;** then, it can be represented as a **natural time (rate) of return.**



Based on the **time between purchases** we can find the **natural time of return**. It will tell us when one of our customers is at **risk of churning** to take a timely action (target those customers).

Motivation

Each **customer** tends to follow a **distinct behaviour;** then, it can be represented as a **natural time (rate) of return.**



Based on the **time between purchases** we can find the **natural time of return**. It will tell us when one of our customers is at **risk of churning** to take a timely action **(target those customers)**.

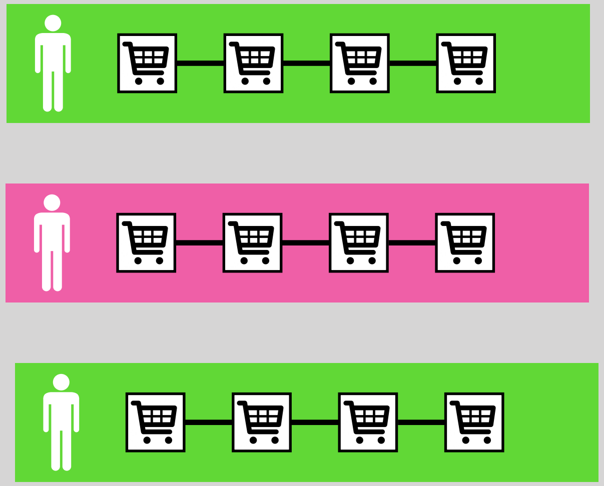
Motivation

We can **see** the **data** in many **different ways**; There are several **approaches** to use the same data

One **vase** or two **faces**?

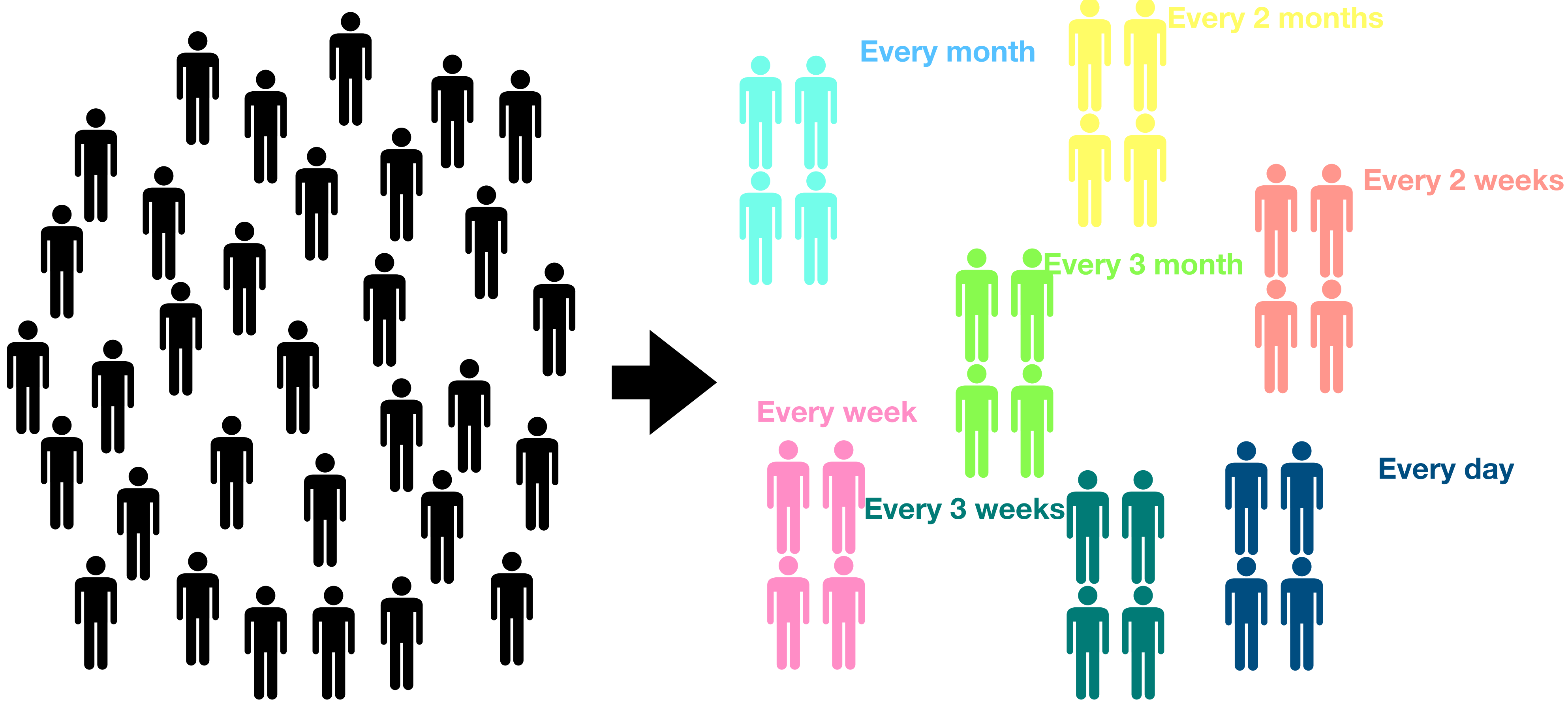
Each customer can be seen as a **collection of individual transactions.**

Each customer can be seen as a **function made of transactions.**



Methodology: Segmentation

Cluster the **customers** based on **their behaviour** (time between purchases)



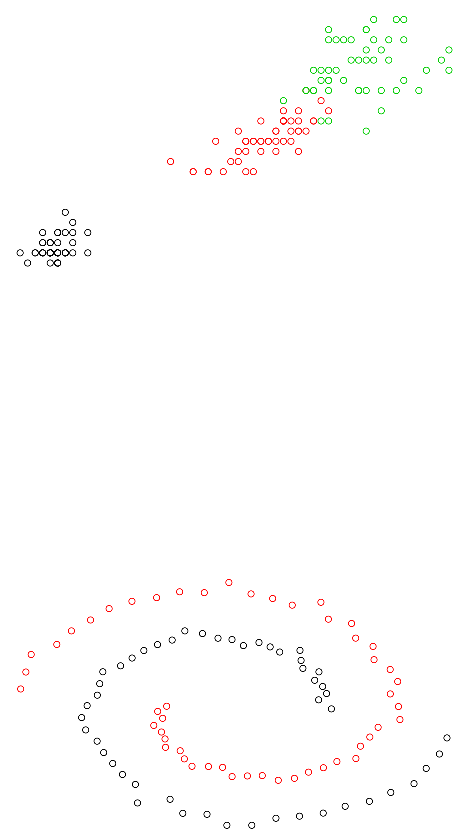
Each resulting **cluster** will contain customers with **similar behaviour**, and the **natural time of return** can identify it.

Methodology: Segmentation

Two different approaches to **cluster** the customers:

K-Means clustering

Treats each customer can be seen as a **collection** of **individual transactions.**

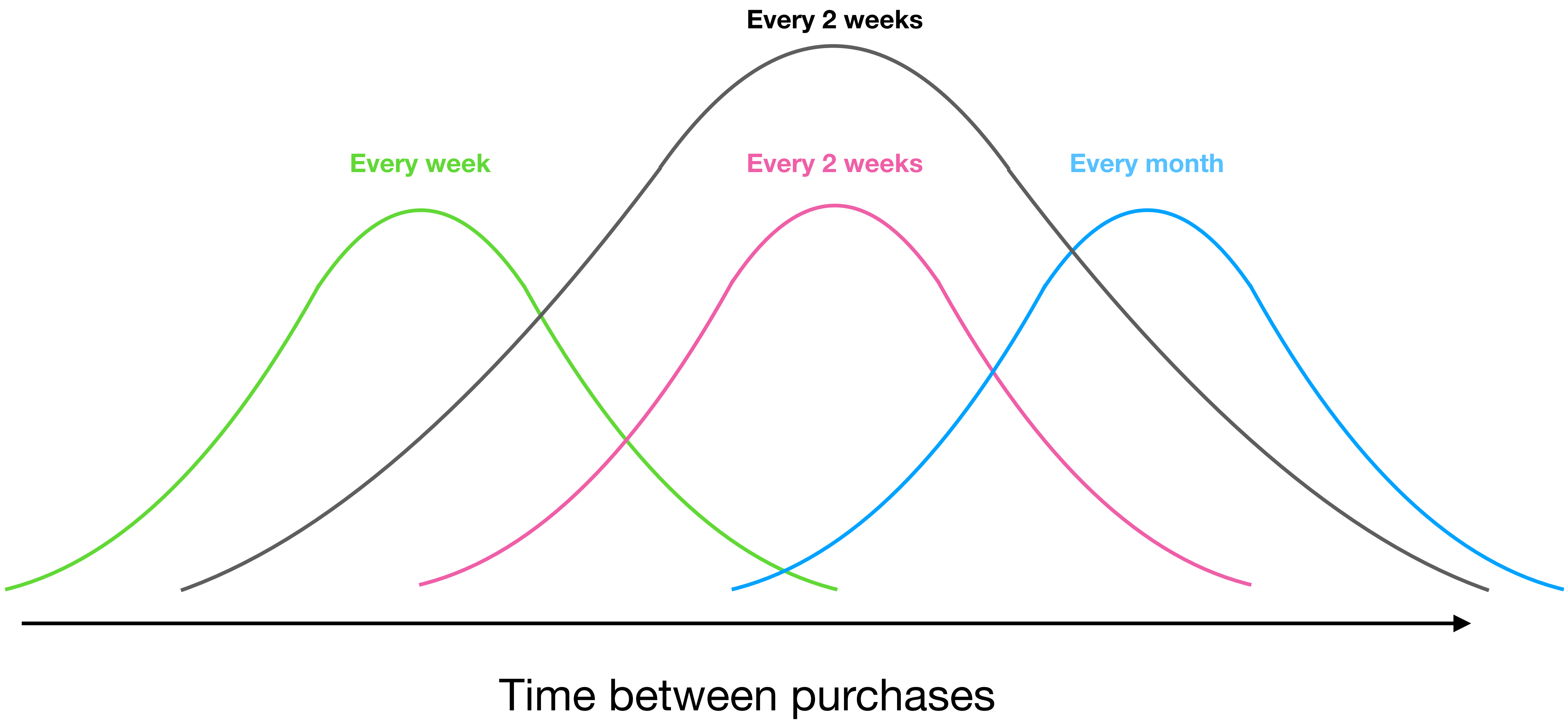


Functional Clustering

Treats each customer can be seen as a **function made of transactions.**

Methodology: Segmentation

The **resulting clusters** will **represent** each **behaviour**.

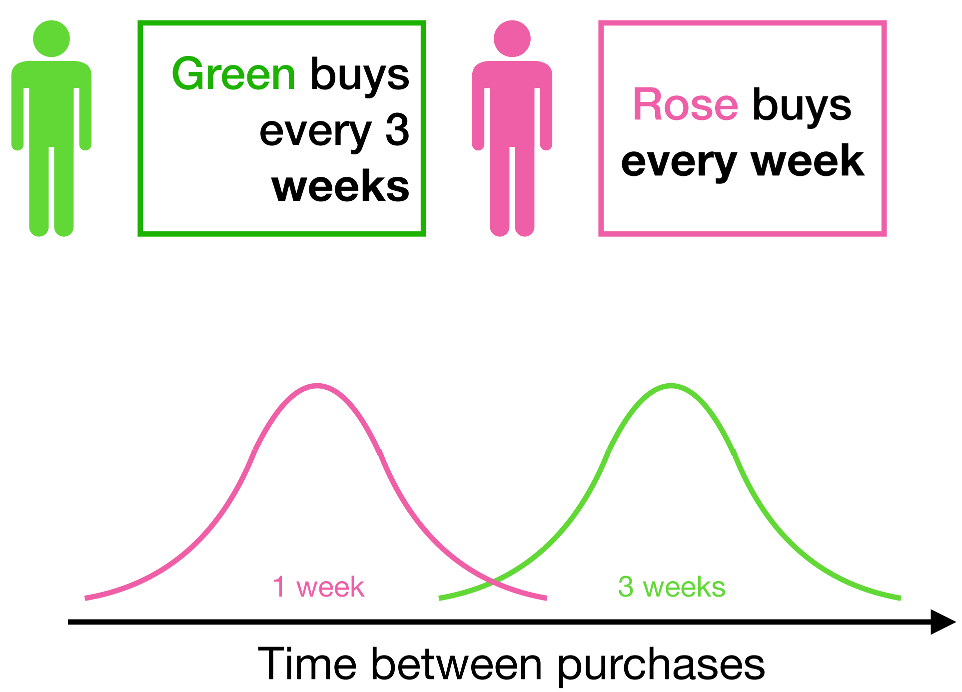


Methodology: Identification

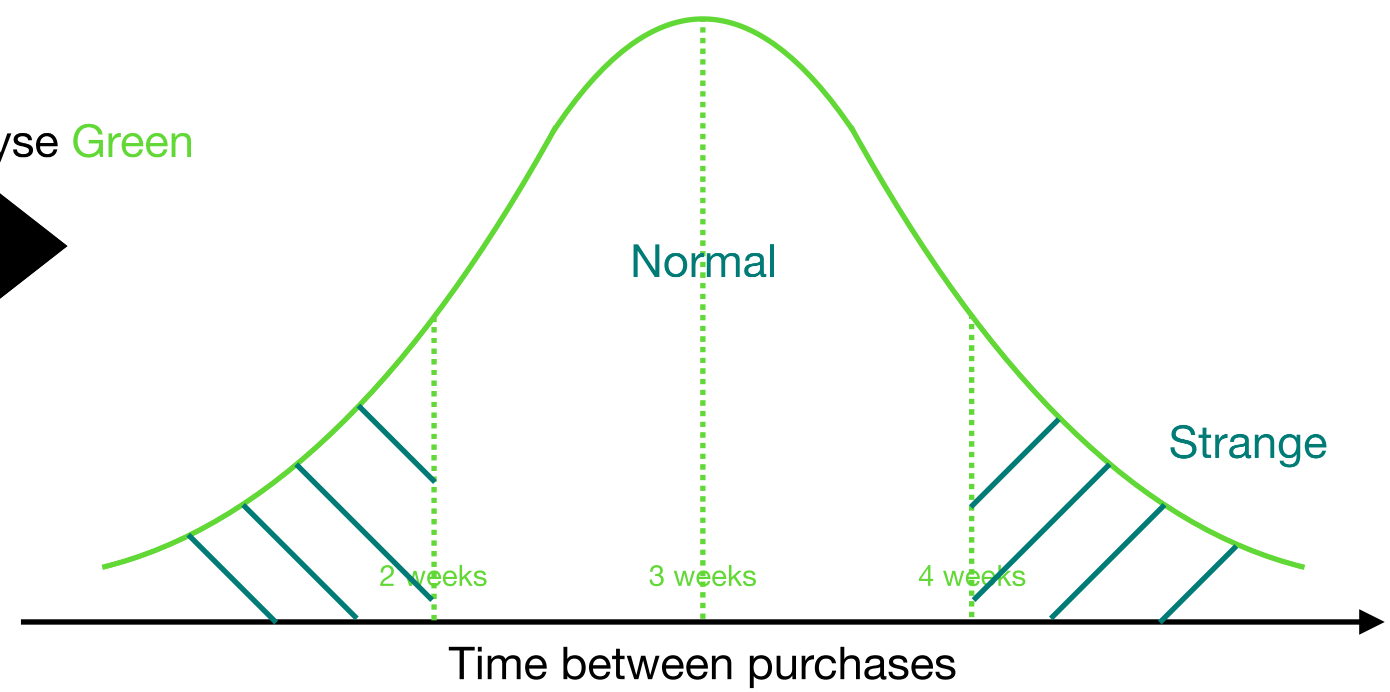
Each resulting **cluster** will contain customers with **similar behaviour.** Then, a customer will be **at risk** of churned when it **behaviour** is **“strange”**.

However, what is **“strange”**?
standard deviation approach

Given Green´s **behaviour**, we know that he comes often **every 3 weeks**. Sometimes **1 week before**, sometimes **1 week after.**



Lets analyse Green



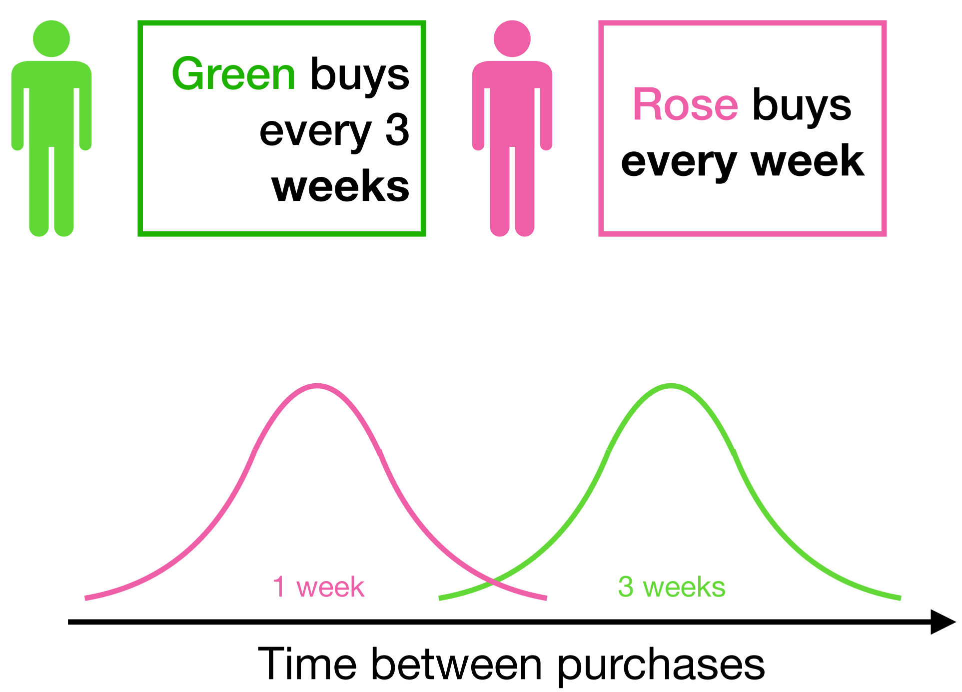
Then, it will be **“strange”** if Green is **out** of the **range.**

Methodology: Identification

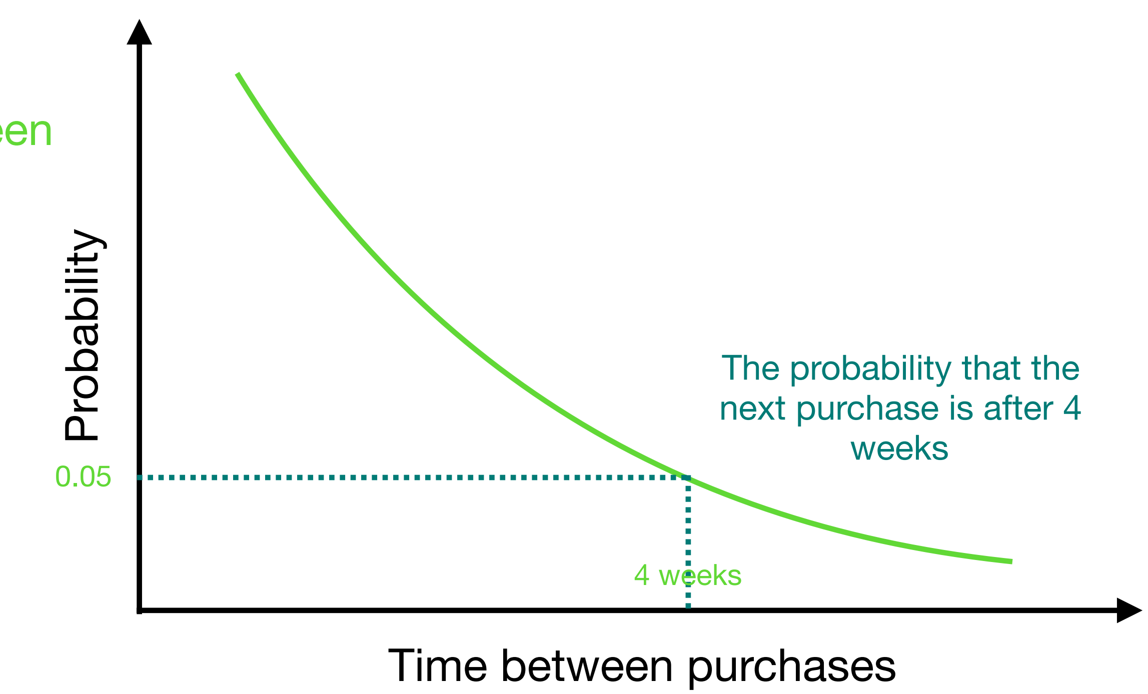
Each resulting **cluster** will contain customers with **similar behaviour.** Then, a customer will be **at risk** of churned when it **behaviour** is **“strange”**.

However, what is **“strange”**?
survival analysis approach

We can find the **probability** that the **next purchase** will be after a **particular time**.



Lets analyse Green



Small probabilities are associated with a **"strange"** behaviours

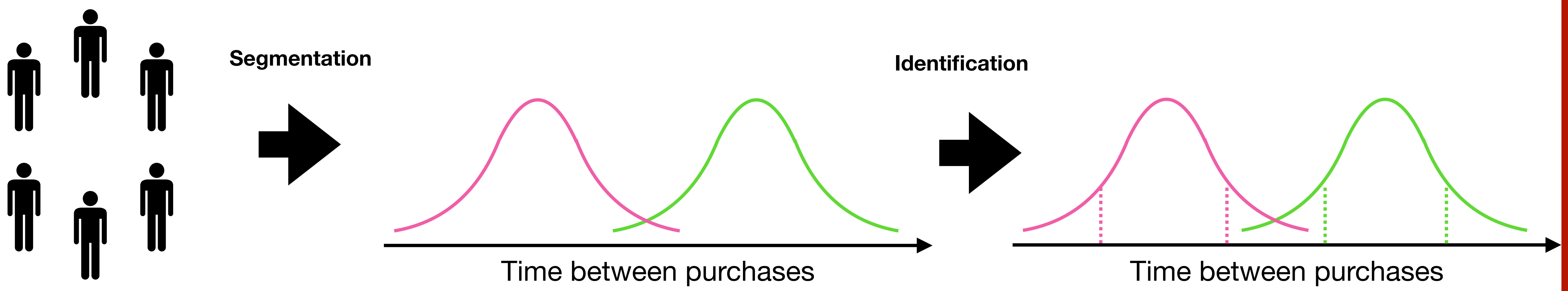
Methodology: Flow

Two different **approaches**

Segmentation	K-Means clustering	Functional Clustering
Identification	Standard deviation	Survival analysis
Testing	Backtesting	Backtesting

Results: Backtesting

We used the **transactions** for **10 thousand** customers between **2016** and **2017**.



As a very **conservative** test, the customers that were **marked** as **churned** were **tracked** over 1 year to check if they **came back**.

Results: Backtesting

K-Means clustering + Standard deviation =

22% of the **customers** that were marked as **churned didn't return**.

42% of **churned customers** were **correctly identified**.

Functional Clustering + Survival analysis =

21% of the **customers** that were marked as **churned didn't return**.

47% of **churned customers** were **correctly identified**.

Identifying customers at risk of churn

University of Essex - Profusion partnership