

Measuring Transit Customers' Travel Regularity

A Metric of customer behaviour using smart card data

Graeme Brown, Senior Planner, TransLink, New Westminster, BC

Abdul Shehata, Consultant - Advanced Analytics R&D, InfoStages Advisors,
Vancouver, BC

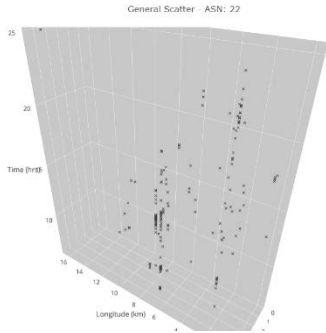
Introduction

Agenda

- Goal
- Methods
- Applications
- Lessons Learned

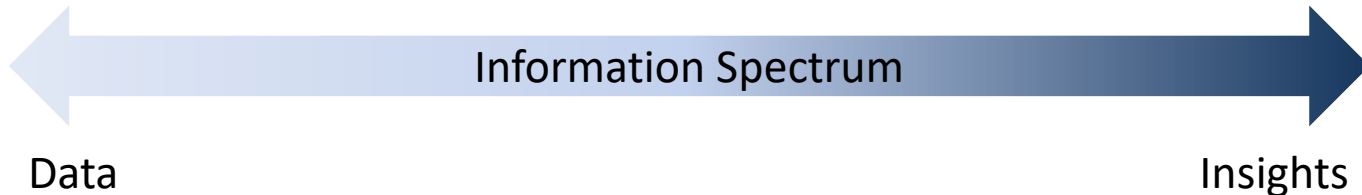
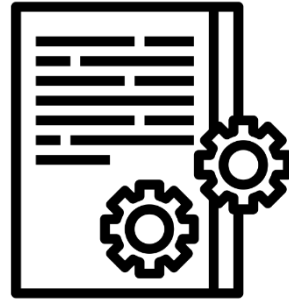
In a Nutshell

In a data driven world the core challenge we face is an abundance of data, and a lack of usable information



- POS Systems
- Supply Chain Management
- MES Systems

- Business Rules
- Strategic Decision Making
- Policy



Research Goal

To define and implement a metric that reflects a customer's tendency to make similar transit journeys over time - **Regularity**

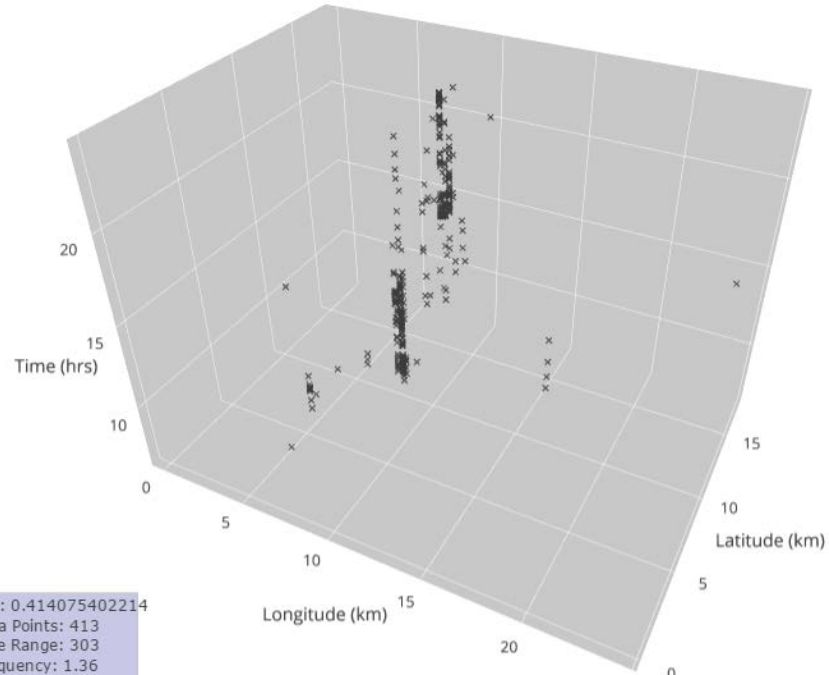
Potential applications:

- Market Segmentation
- Pattern Detection

What is the data?

- Travel data from the smart card (Compass) system
 - Location
 - Date
 - Time
 - Mode
 - Etc.
- User specific data

Sample travel data for one person



How do we examine the data?

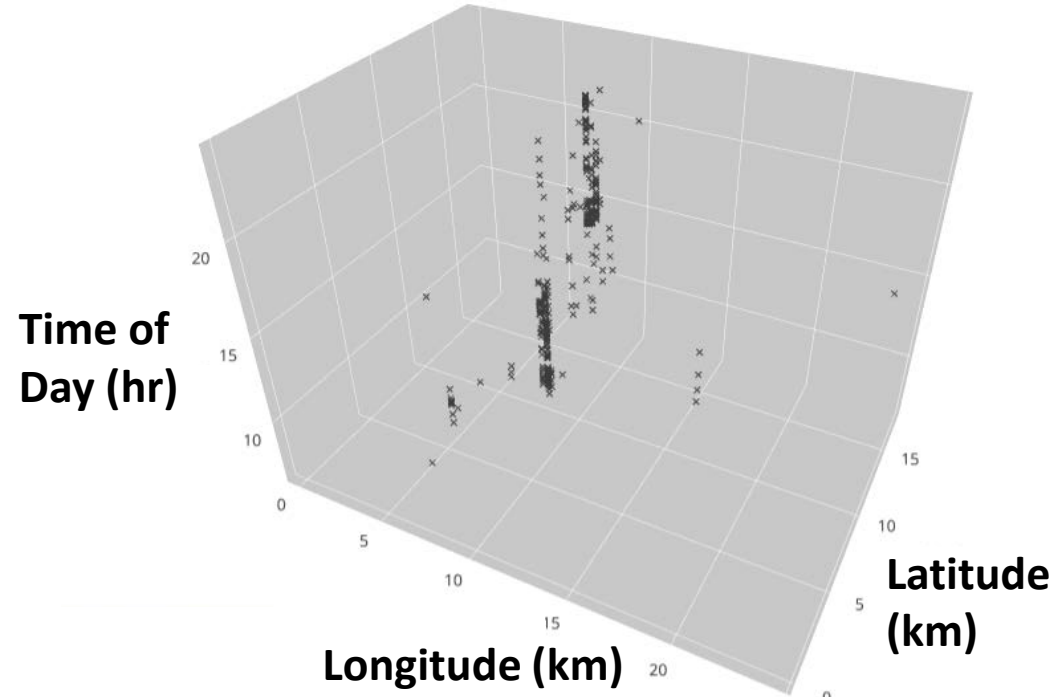
Research Goal

To define and implement a metric that reflects a customer's tendency to make similar transit journeys over time.

How do we examine the data?

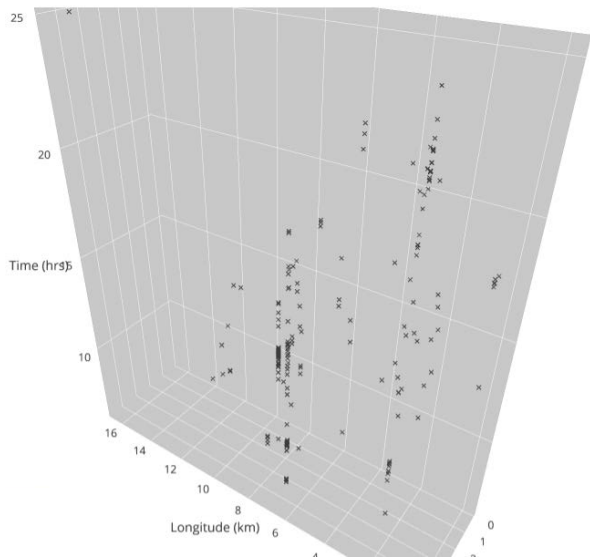
- Reduction of data to 3 relevant dimensions
 - Origin Longitude
 - Origin Latitude
 - Departure Time
- Framing it in the **Journey Space**
- Each point represents a Journey

Sample travel data for one person

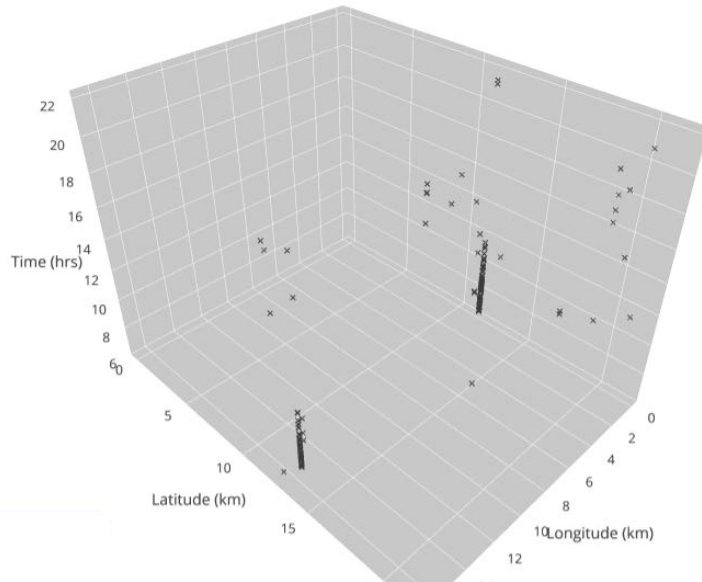


Comparing Users Journey Spaces

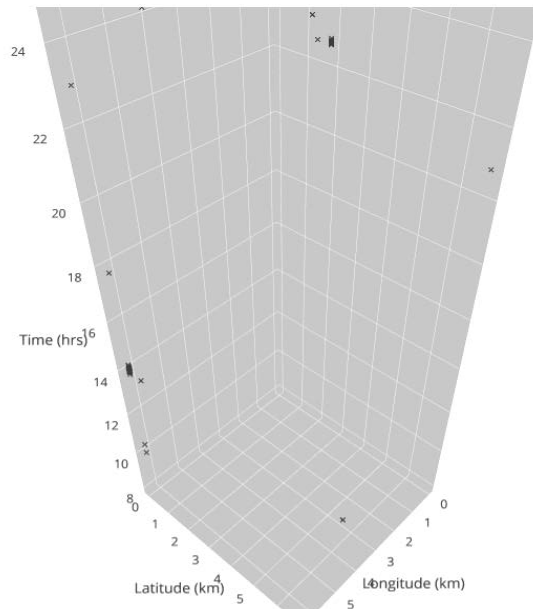
Transit User 1



Transit User 2



Transit User 3



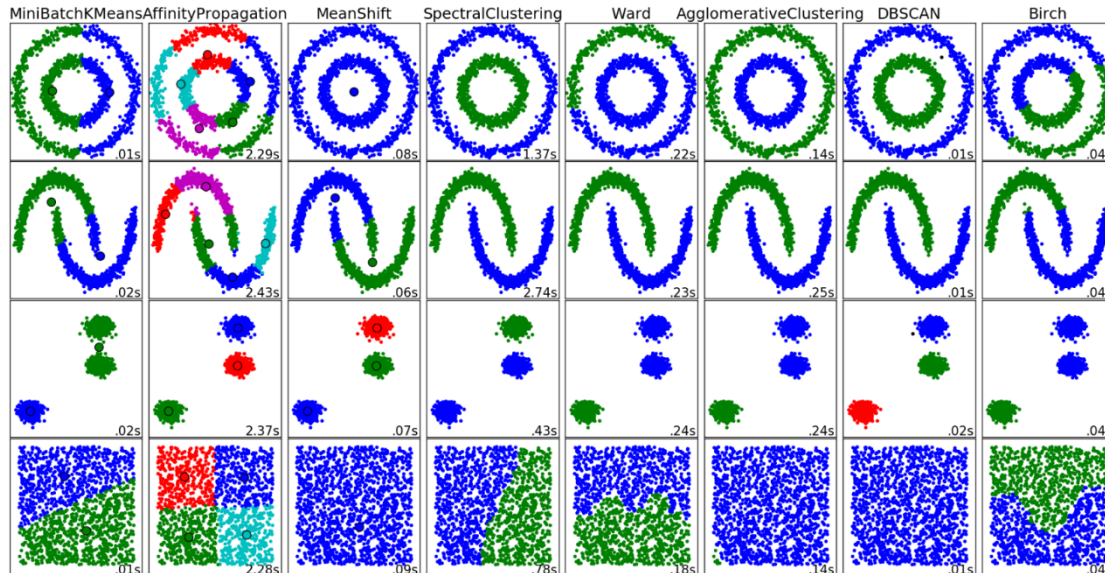
Increasing Regularity

Methods

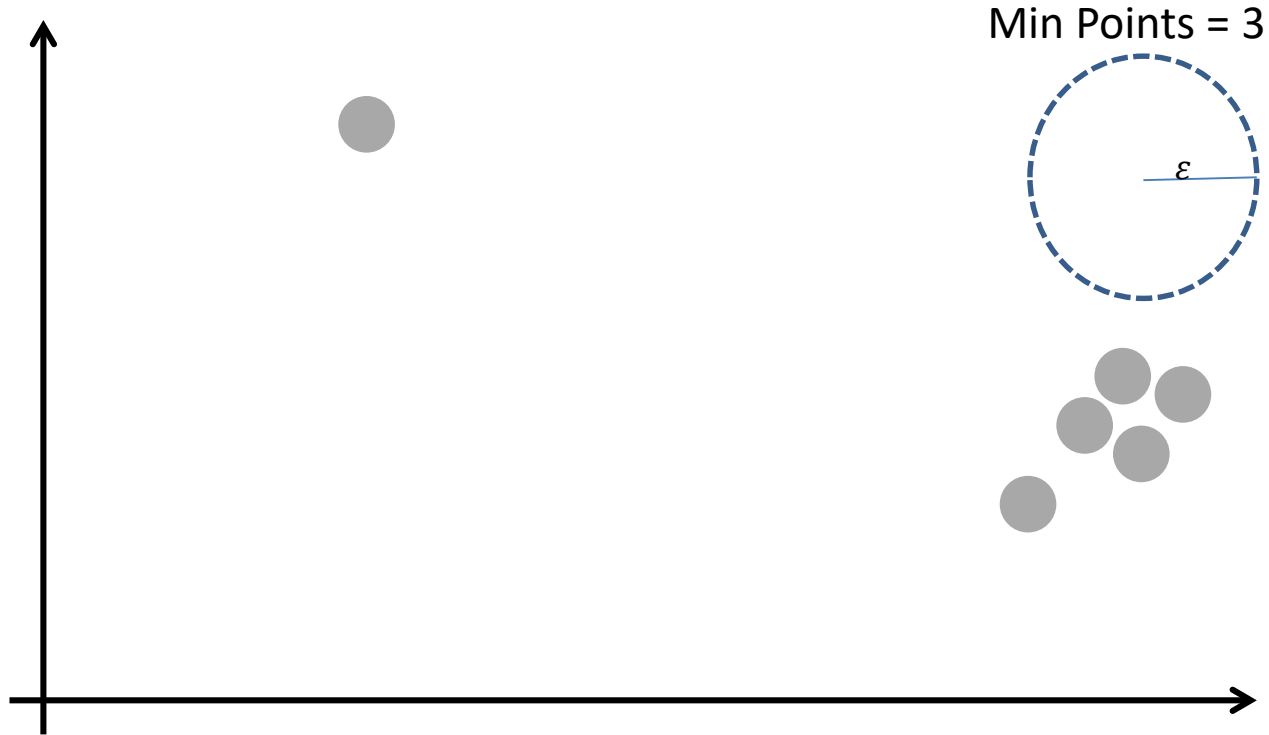
Algorithmic Clustering

Definition: **Clustering algorithms** are a subset of machine learning tools that group a set of data points into sets (clusters) of similar objects based on similarity of specific parameters

Types of Clustering

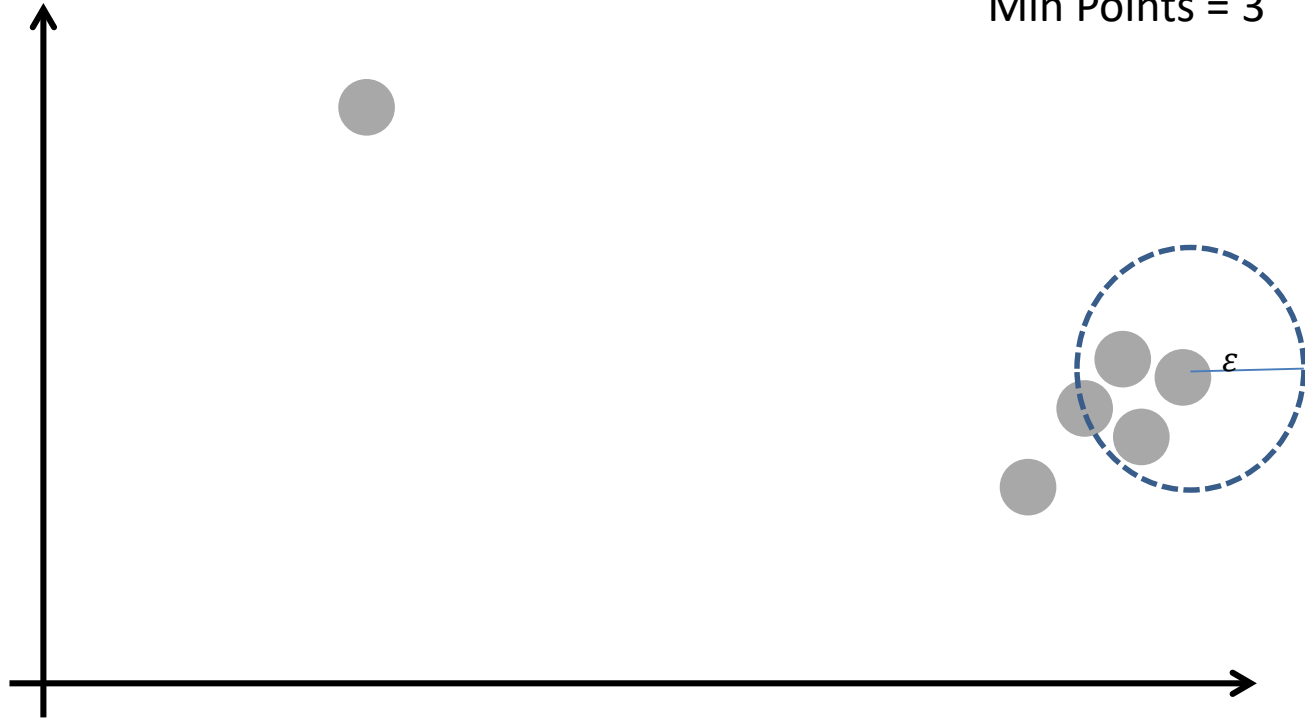


Density Based Clustering



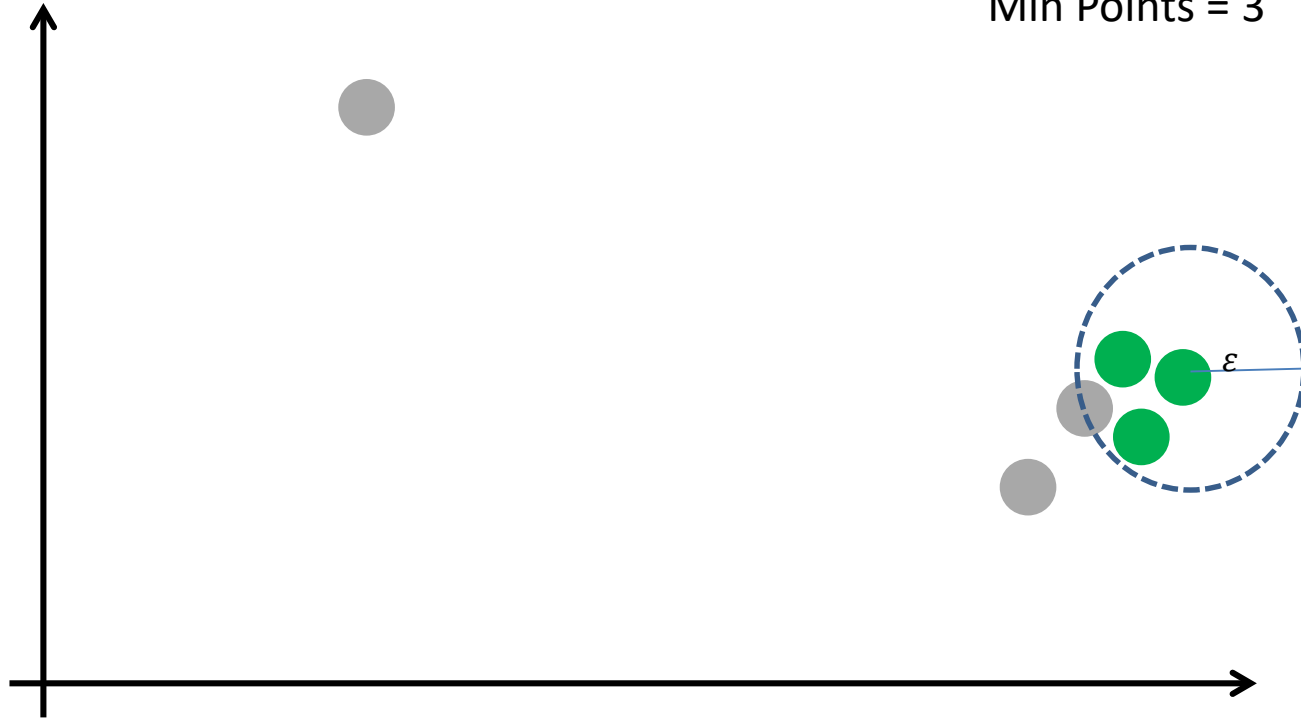
Density Based Clustering

Min Points = 3



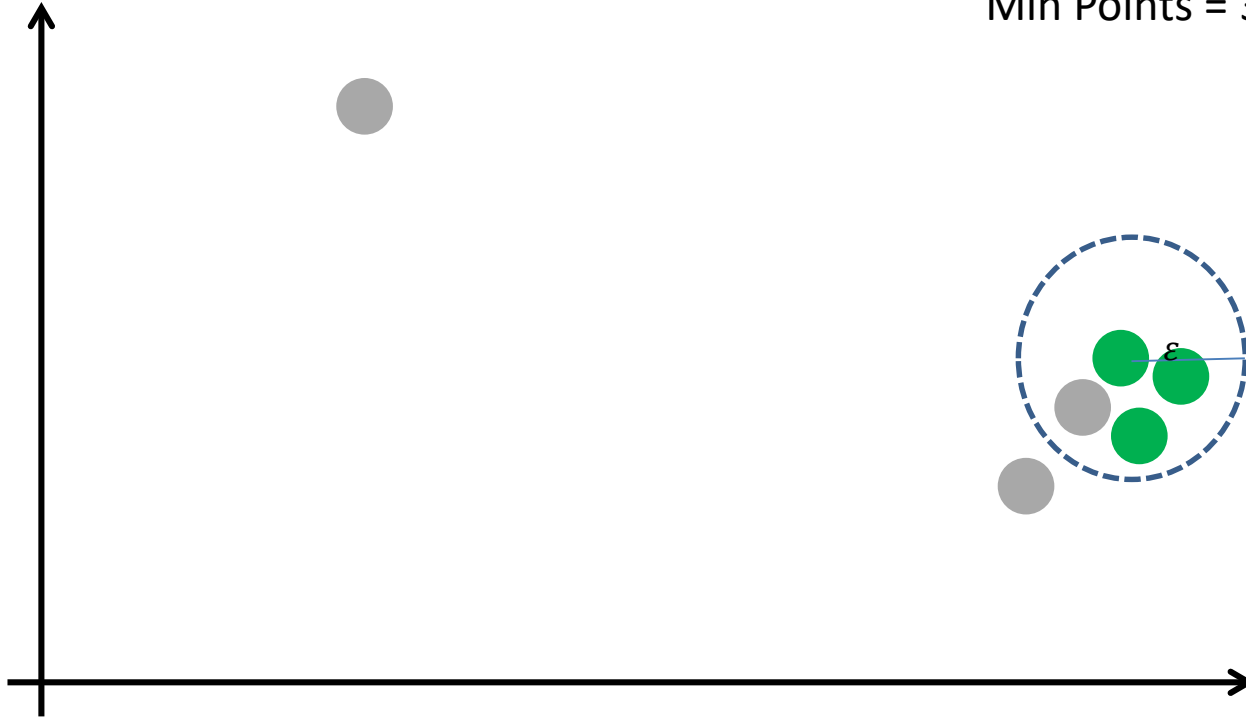
Density Based Clustering

Min Points = 3



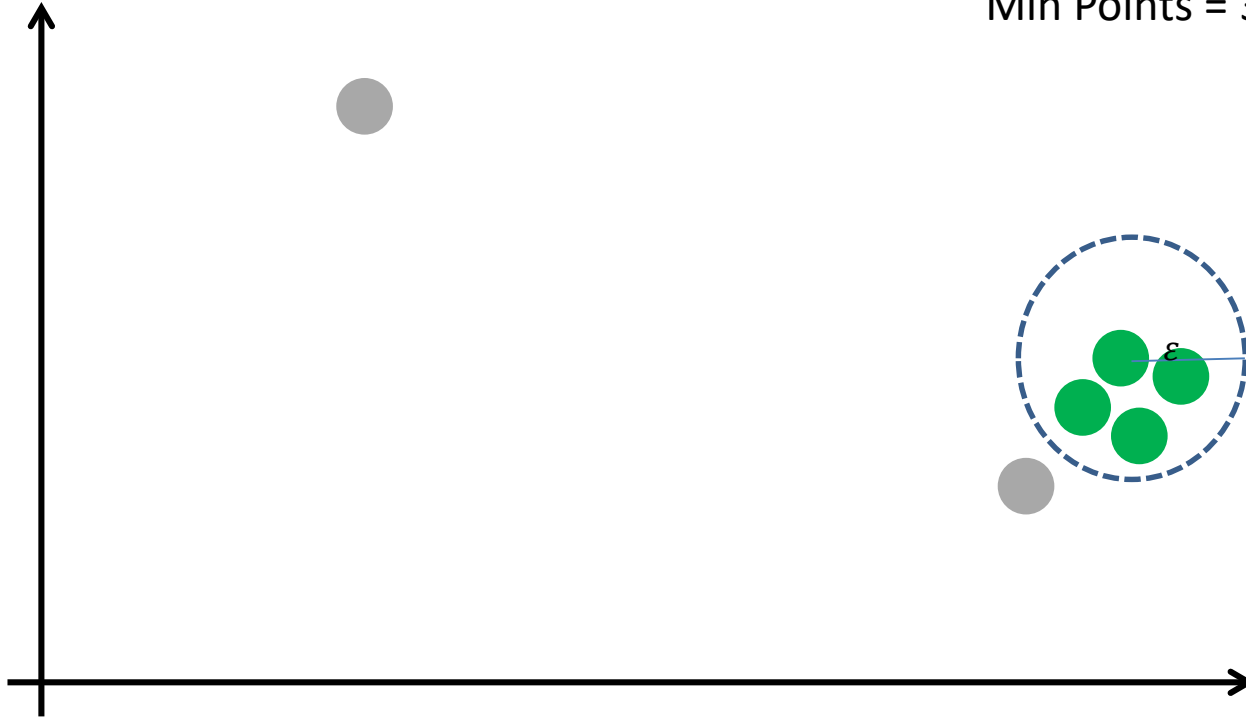
Density Based Clustering

Min Points = 3



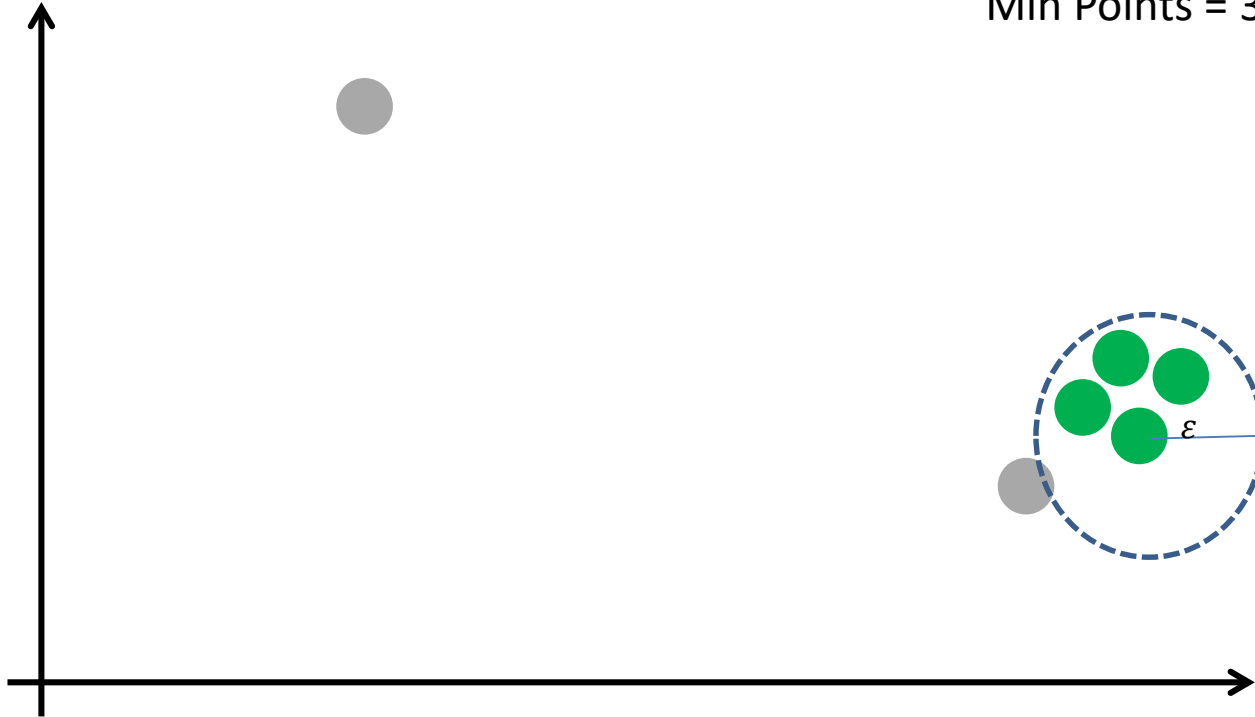
Density Based Clustering

Min Points = 3



Density Based Clustering

Min Points = 3



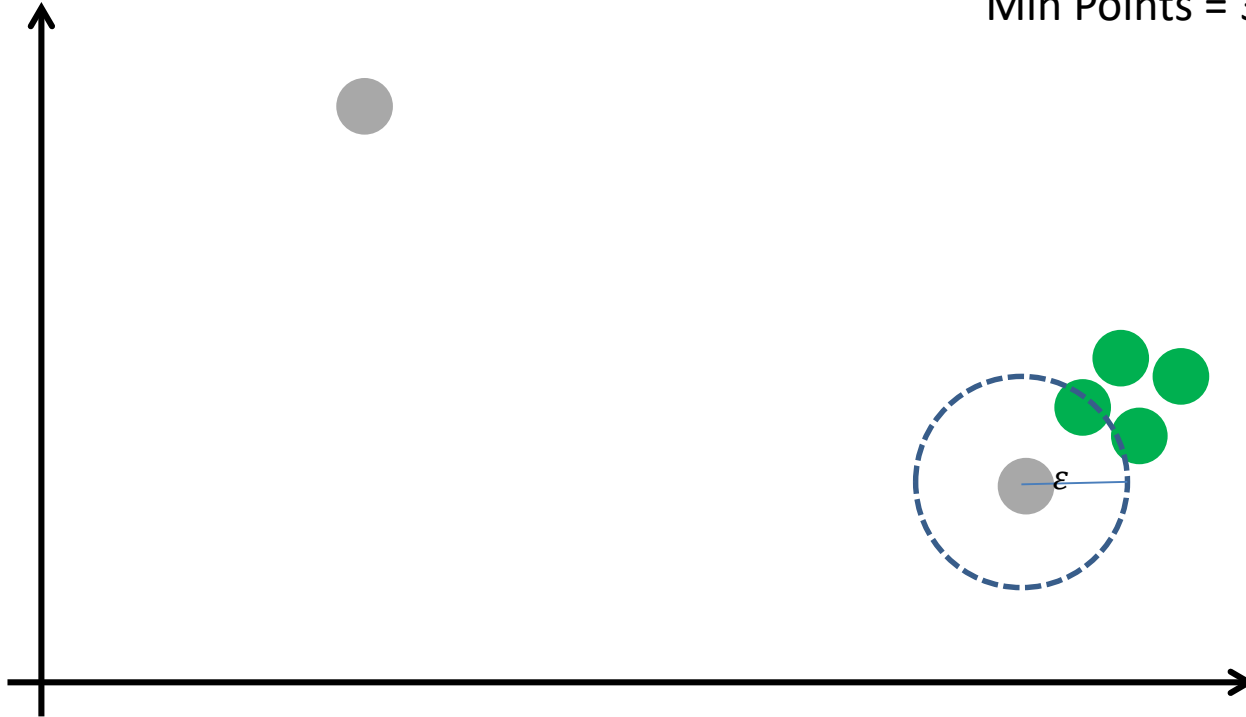
Density Based Clustering

Min Points = 3



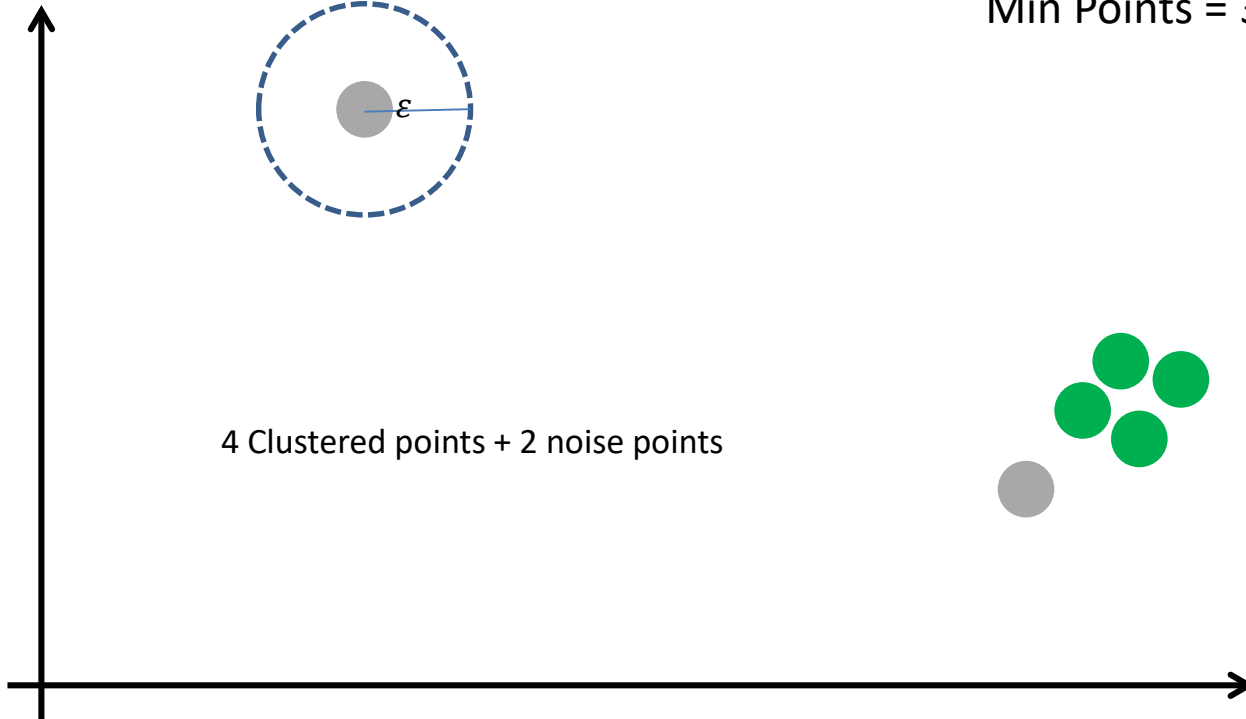
Density Based Clustering

Min Points = 3

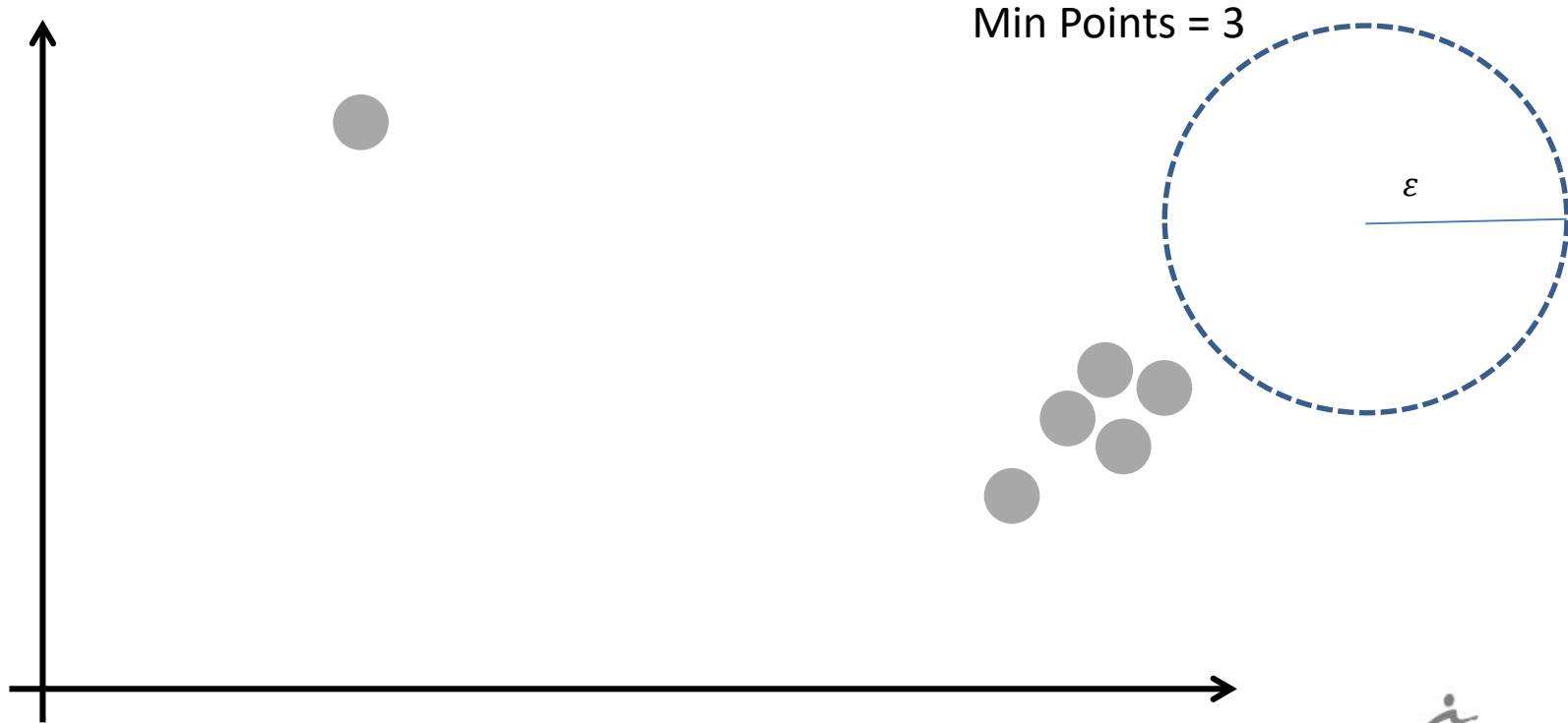


Density Based Clustering

Min Points = 3

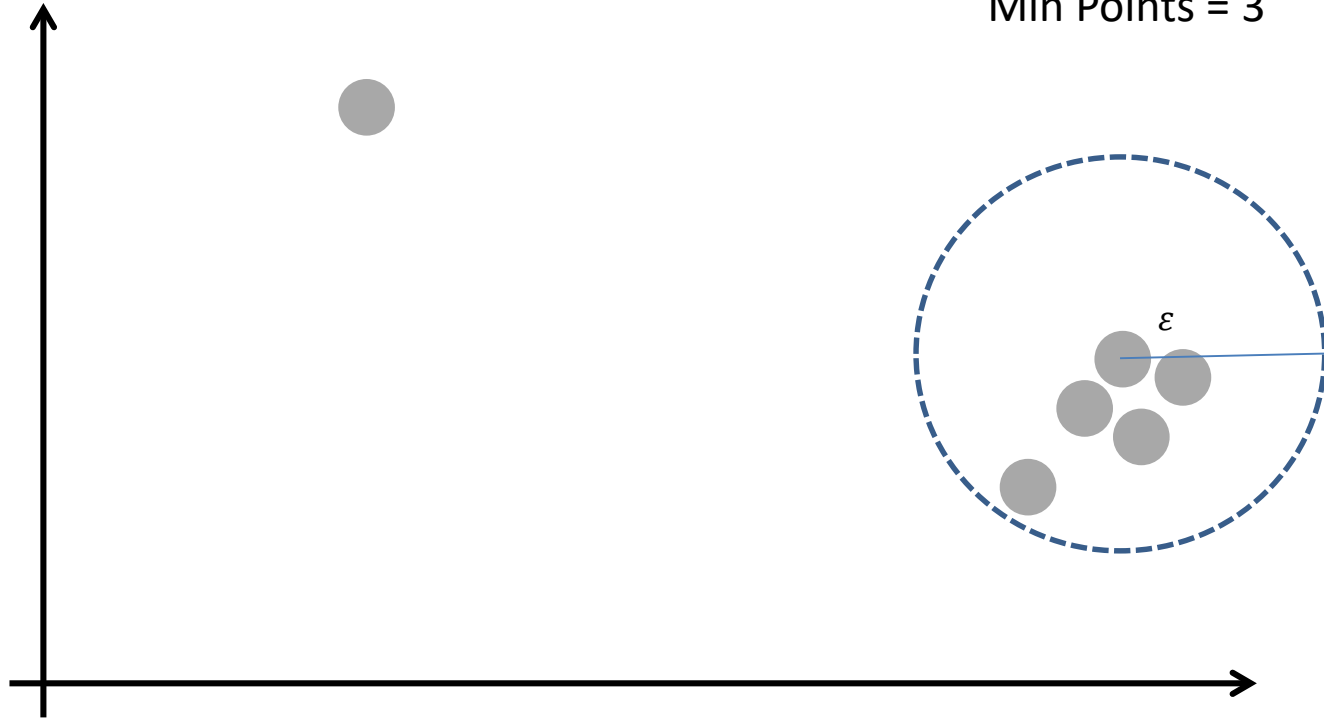


Density Based Clustering



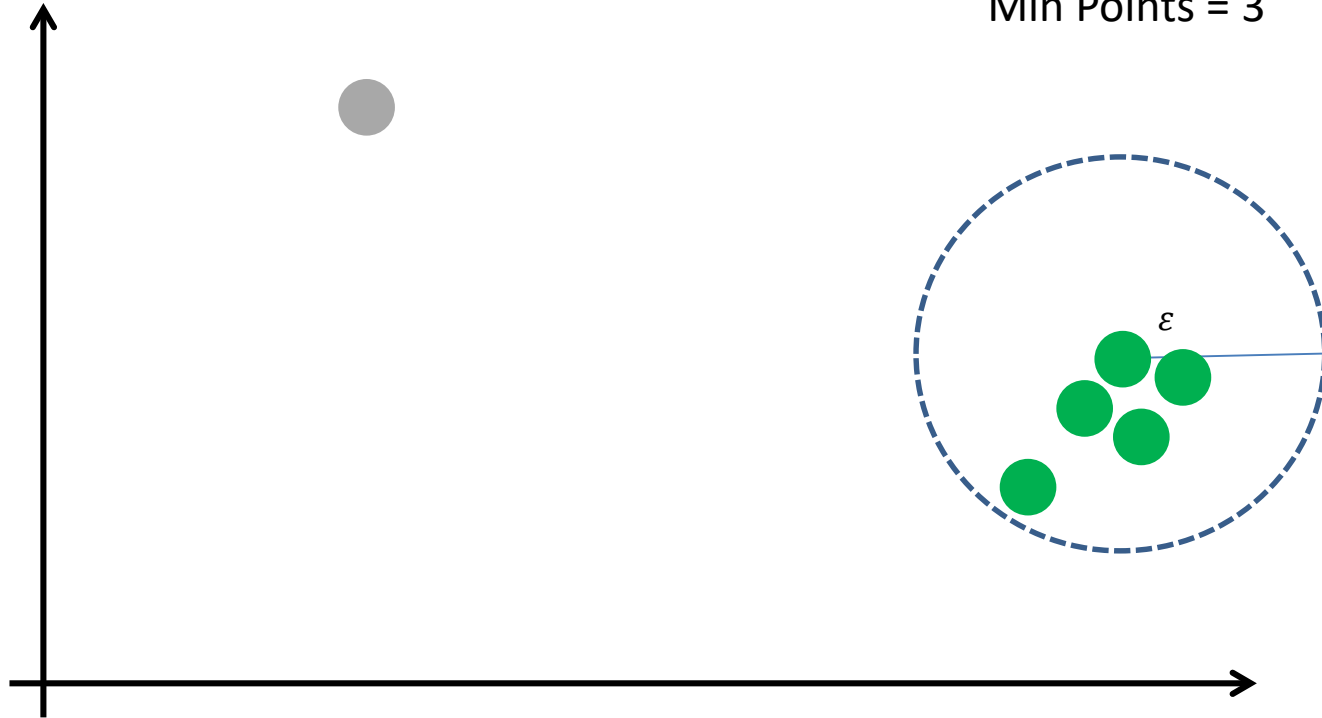
Density Based Clustering

Min Points = 3



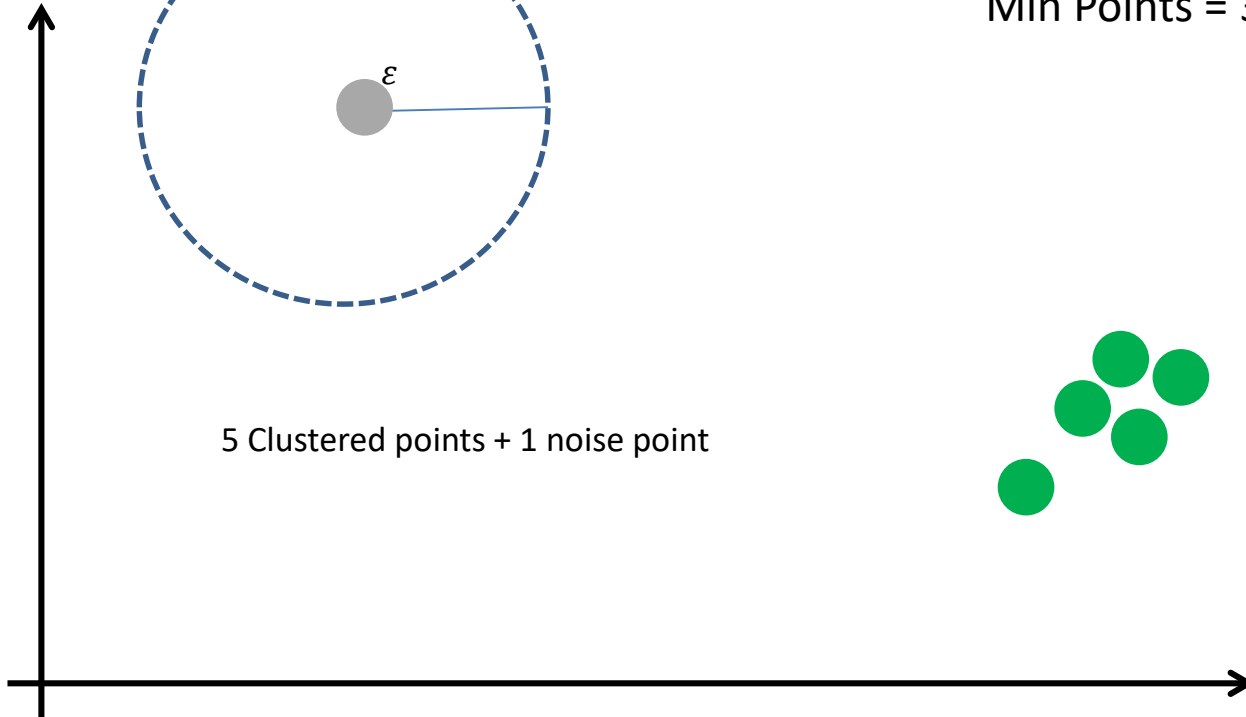
Density Based Clustering

Min Points = 3

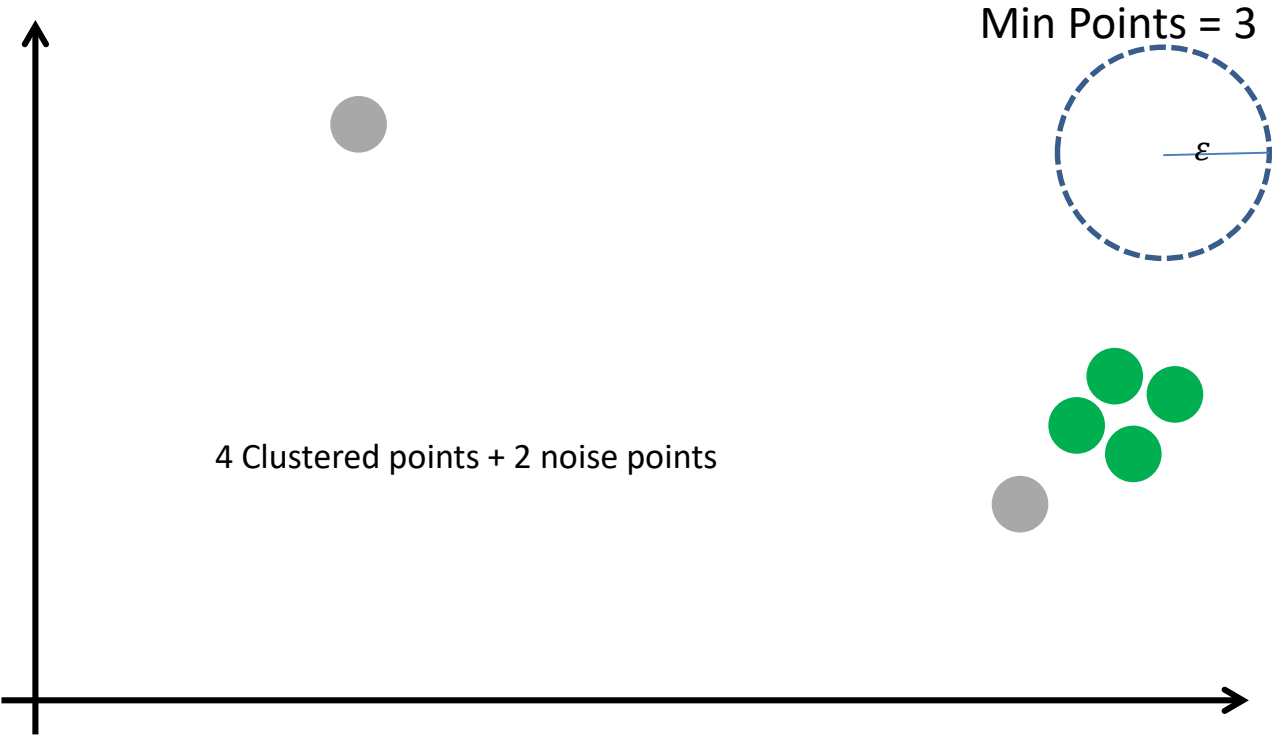


Density Based Clustering

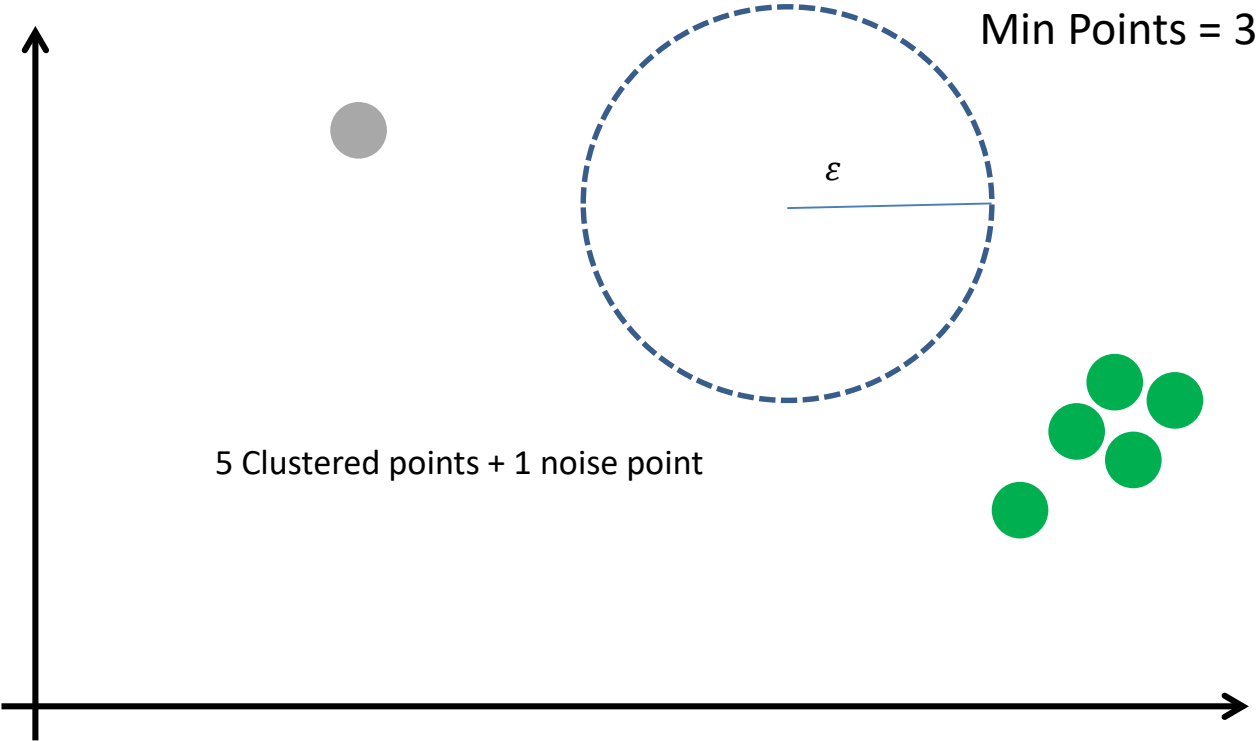
Min Points = 3



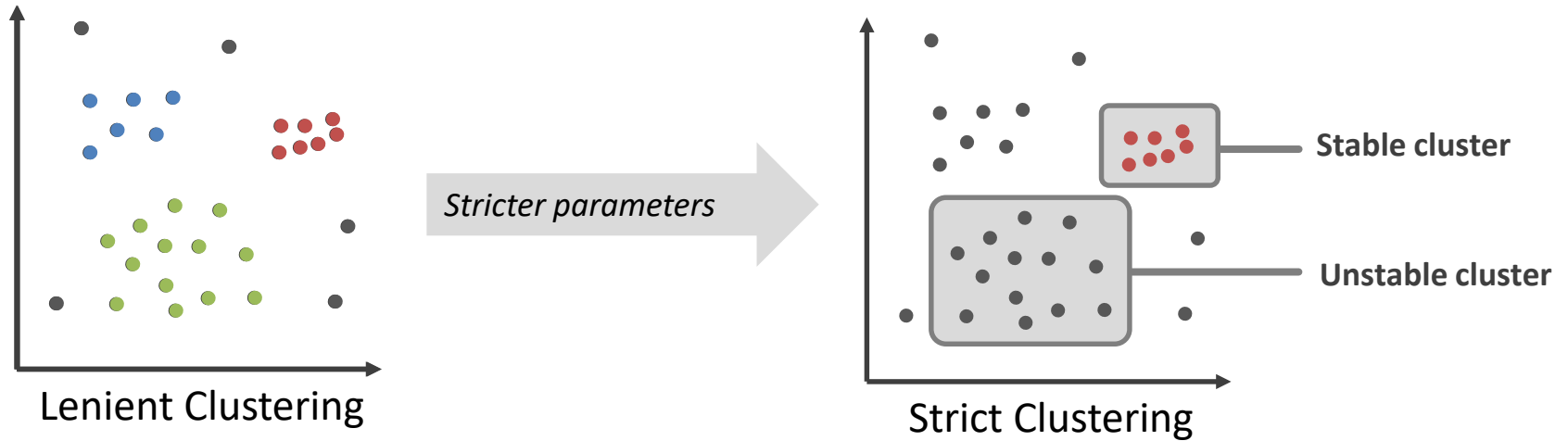
Density Based Clustering



Density Based Clustering



Cluster Stability



Regularity: customer's tendency to make *similar journeys* over time

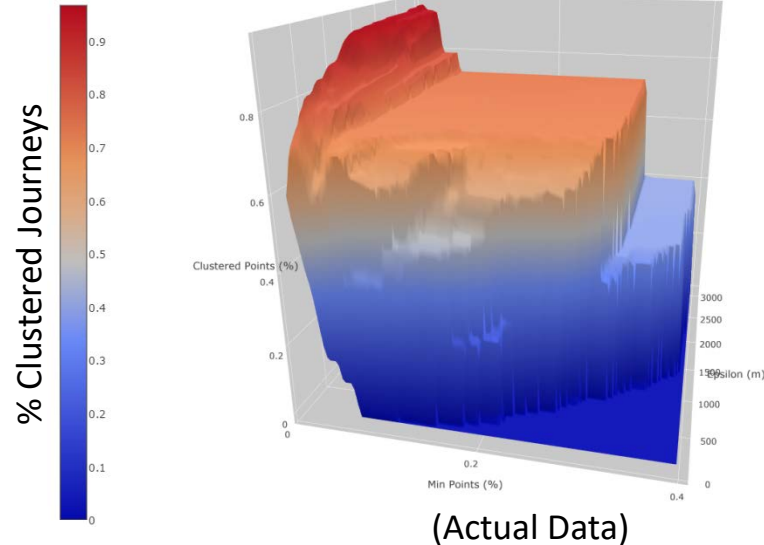
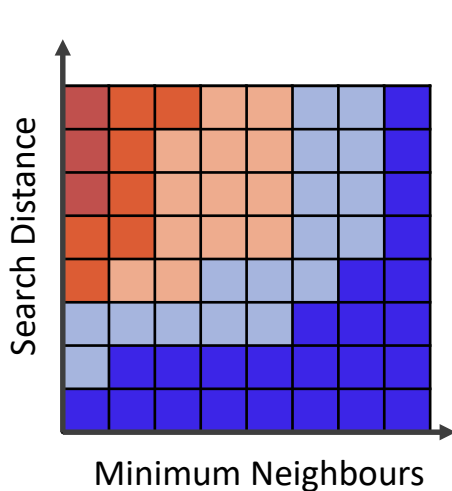
If a large proportion of a customer's journeys fall into clusters, AND these clusters survive strict parameters, the customer can be considered "regular".

Methods – Calculating Regularity

Selected clustering algorithm (DBSCAN) has 2 parameters:

- Search distance
- Minimum neighbours

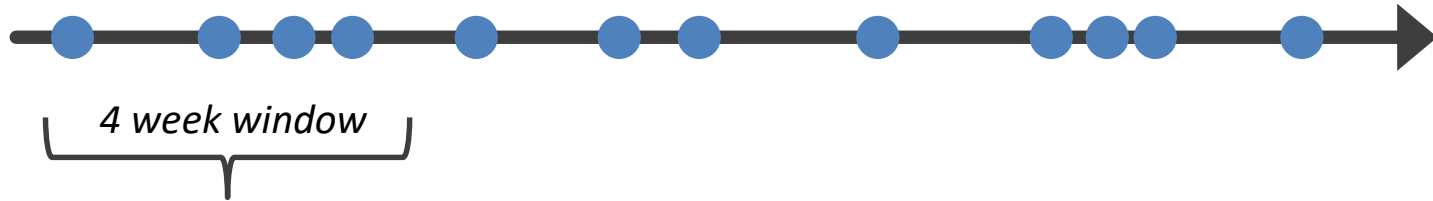
To measure regularity for a card, we run DBSCAN with various parameter combinations...



Regularity = average % of journeys that fall within clusters.

Regularity Over Time

How does regularity change over time?



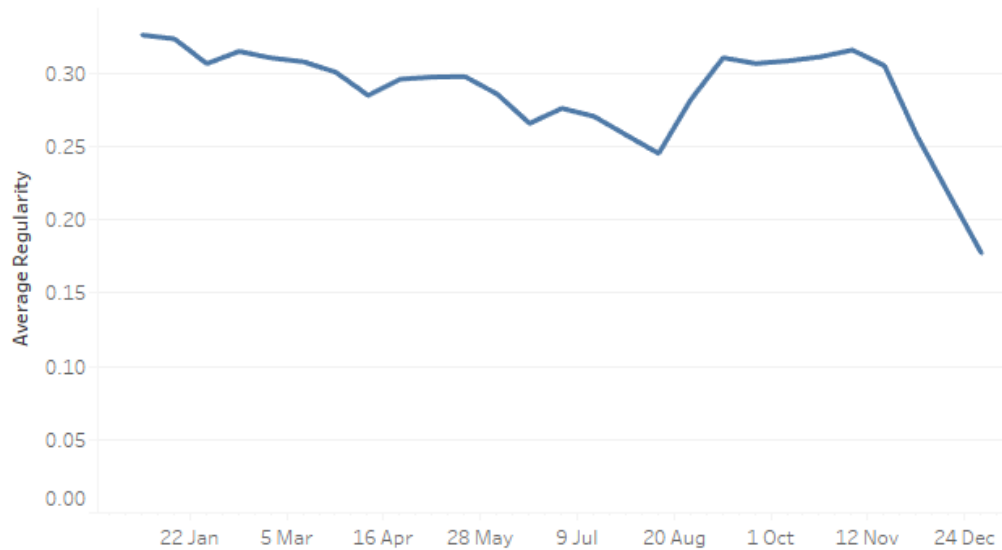
Regularity = 0.65
Regularity = 0.58

Card	Period					
	1	2	3	4	5	6
A	0.65	0.58	0.55	0.48	0.58	0.6
B	0.70	0.75	0.72	0.65	0.60	0.53
C	0.20	0.25	0.27	0.29	0.30	0.30

Findings & Applications

Trends in Regularity

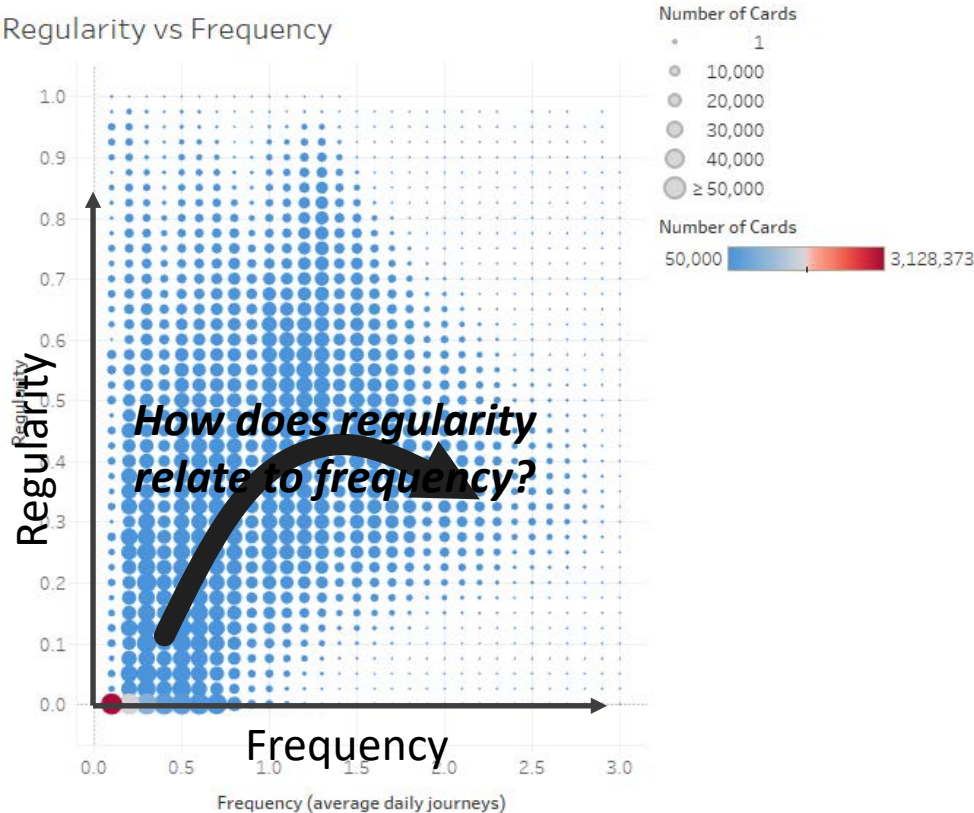
Average Regularity by Period



Regularity declined through first 8 months of 2017, then recovered in the Fall.

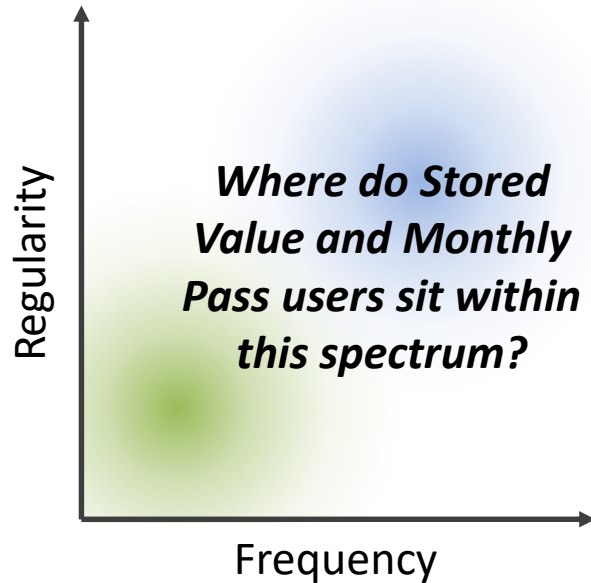
Regularity vs. Frequency

Regularity vs Frequency



- Wide variation in regularity for all frequencies.
- Regularity tends to increase with frequency, up to ~1.3 journeys per day, then declines.

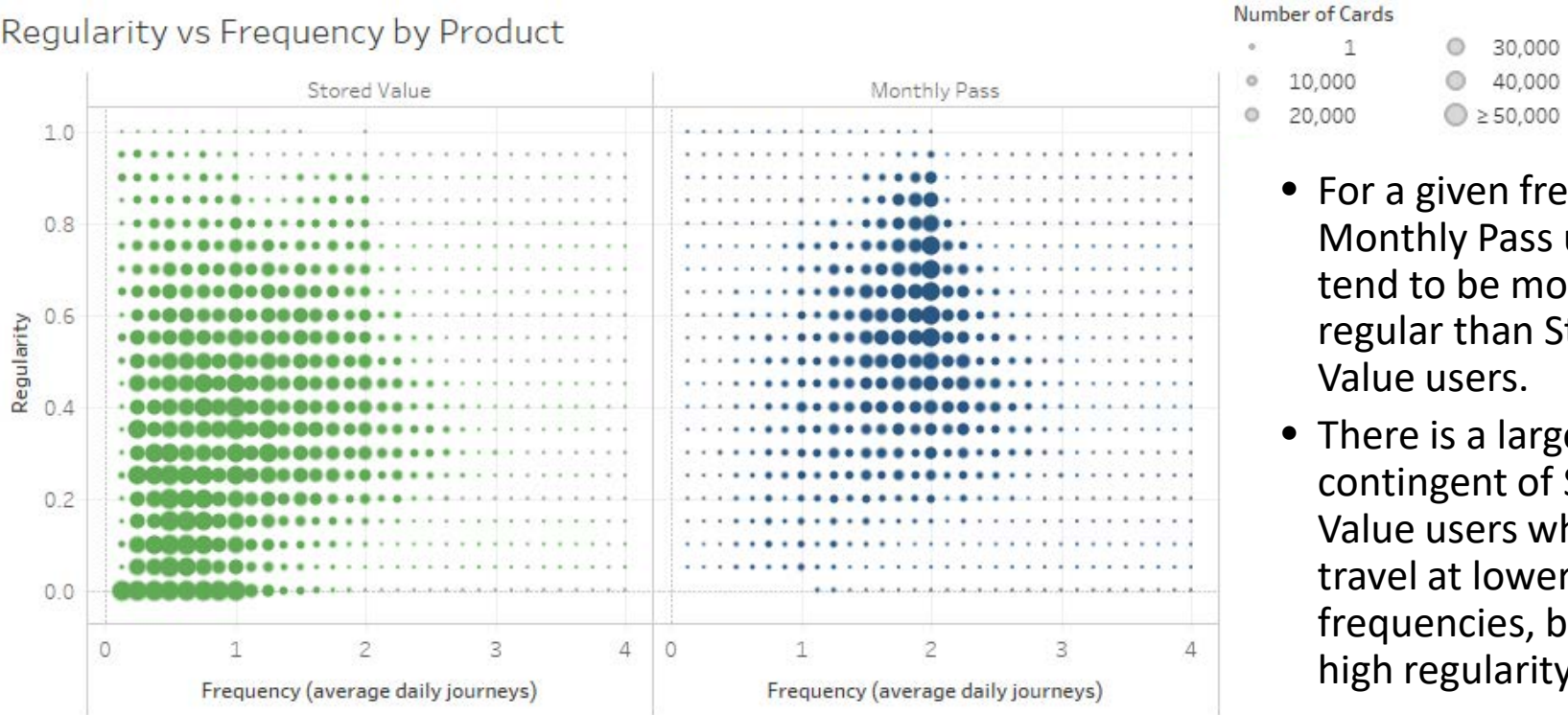
Findings – Unexpected Transit Market



Conventional wisdom is that Stored Value users are “irregular” users.

Regularity vs Frequency

Regularity vs Frequency by Product



- For a given frequency, Monthly Pass users tend to be more regular than Stored Value users.
- There is a large contingent of Stored Value users who travel at lower frequencies, but with high regularity.

Future Applications

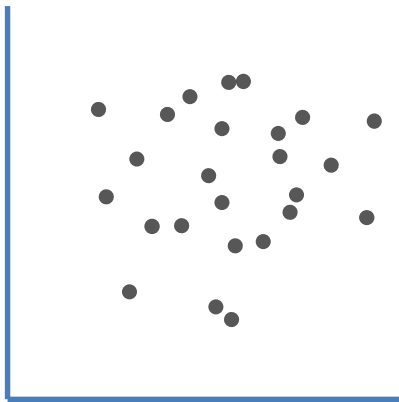
- Market Segmentation
- Pattern detection / behaviour trends

Thank you!

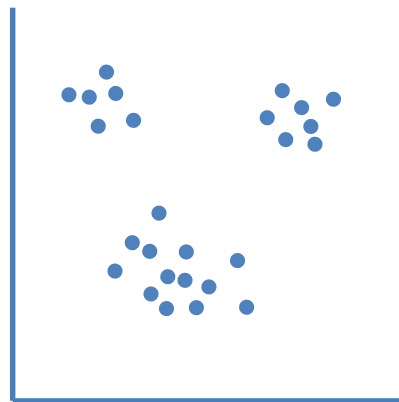
Contact Information

Graeme: graeme.brown@translink.ca

Abdul: abdul.shehata@infostages.co



No clusters (all noise)



Strong Clusters