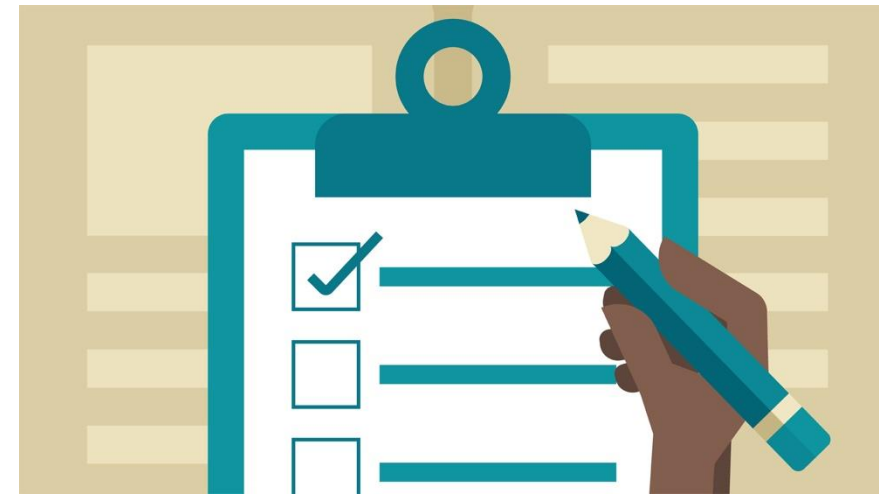# Marketing Analytics

Lecture 3 Transaction Analytics (I)

Dr. Bingqing Xiong

Senior Lecturer in Deakin Business School

1

Learning Objectives for Transaction Analytics

❖ Understand how to assess and characterize purchase dynamics

❖ Segmentation; RFM analysis

❖ Know how to perform relevant analysis and interpret the results – will be in next week

What is Transaction Data

❖ The time of purchase

❖ The location where it took place

❖ What digital channel or physical touchpoint was used

❖ The payment method

❖ Whether any discounts or promotions were used

❖ If the purchase came from a repeat customer or loyalty member

By Bingqing Xiong

Transaction Data: POS in offline stores
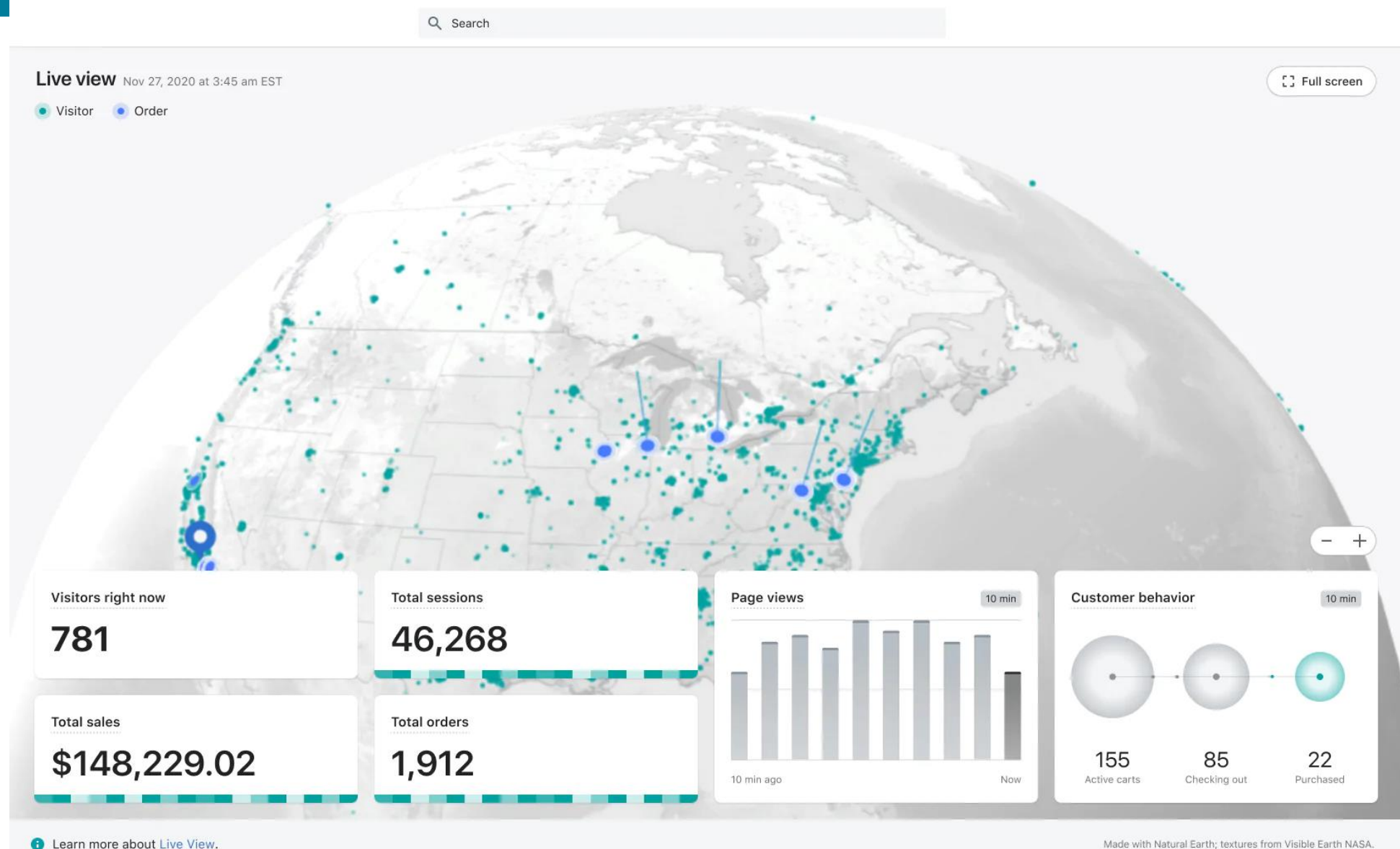

POS - Point-of-Sale Systems
The POS collects
- **date and time**
- **items purchased**
- **price of each item**
- **payment method used**
- **applicable taxes or fees**
- **Some customer information**

Offline Store: Manual record-keeping, barcode scanners and mobile payment systems are also used.

By Bingqing Xiong

Live View from Shopify (an e-commerce platform)

- Visitors
- Sales
- Checkout
- Metrics
- Traffic

Search

Live view Nov 27, 2020 at 3:45 am EST

● Visitor   ● Order

[ ] Full screen

Visitors right now
781

Total sessions
46,268

Page views   10 min

Customer behavior   10 min

155 Active carts   85 Checking out   22 Purchased

10 min ago   Now

Total sales
$148,229.02

Total orders
1,912

ⓘ Learn more about Live View.

Made with Natural Earth; textures from Visible Earth NASA.

5

The Role of Transaction Data

❖ Transaction and purchase data is useful for predicting sales

❖ Can also be used to reflect and/or understand the emotional or cognitive factors influencing customer behaviour

❖ Identifying meaningful and significant tendencies and characteristics in the data

Analysis of Transactions on Individual Level

❖ **Dynamics of Individual Transaction Data**

❖ **Loyalty**

- Purchase behaviour → Dynamics of repeat purchase (e.g., frequency)
- Attachment → Emotional bonds to a brand (e.g., sentiment from the review)

❖ **Habit**

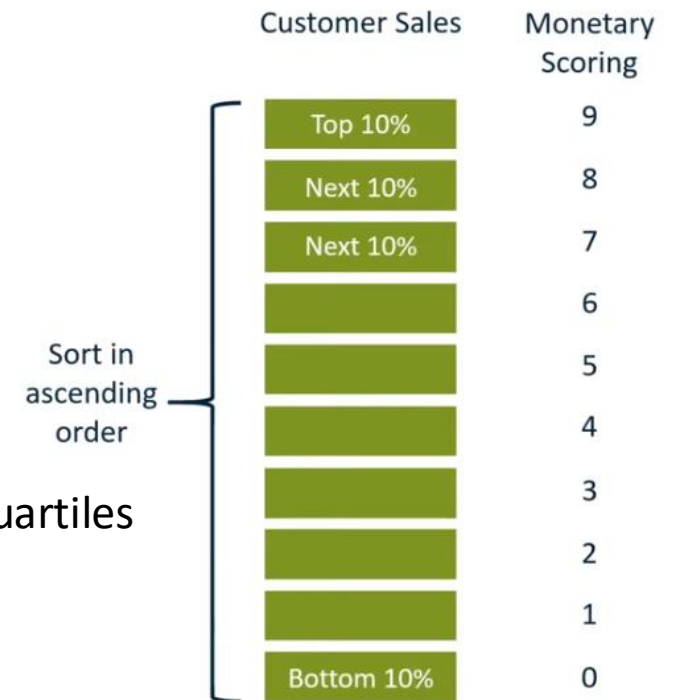- Purchase behaviour → Frequency + Routine (e.g., Time/Space)

❖ **Variety Seeking**

- Product switching
- Variance of products

By Bingqing Xiong

❖ Recency, Frequency, Monetary (RFM) Analysis

❑ Recency: How recent was the customer's last purchase?
❑ Frequency: How often did this customer make a purchase in a given period?
❑ Monetary: How much money did this customer spend in a given period

- Consumers can be given a score according to the actual value or be assigned to quartiles of groups based on their scores

By Bingqing Xiong

Analysis of Transactions on Individual Level

❖ Recency, Frequency, Monetary (RFM) Analysis

- Consumers who rank highly in RFM are more likely to be repeat buyers

- RFM scores can then be cross-referenced with other variables (e.g., demographics)

- RFM can also be about other behaviours, and may set alert of customer churning

- However, each business has different interpretations of what a good recency score is

- Provide your RFM criteria is important as it varies in different contexts

By Bingqing Xiong

❖ Customer Lifetime Value (CLV)

Analysis of Transactions on Individual Level

❖ Customer Lifetime Value (CLV)

- Extends the logic of RFM

- Measure show valuable a customer is to the company, not just on a purchase-by-purchase basis but across the whole relationship

- The actual formula vary from company to company and from market to market (**Customer lifetime value formula: Easy ways to calculate it**)

Identify the touchpoints where the customer creates the value

+ Integrate records to create the customer journey

+ Measure revenue at each touchpoint

Add together over the lifetime of that customer

11

By Bingqing Xiong

❖ Basket Analysis

- Which products are brought together?

- Associate products bought at the same time (by the same purchaser)

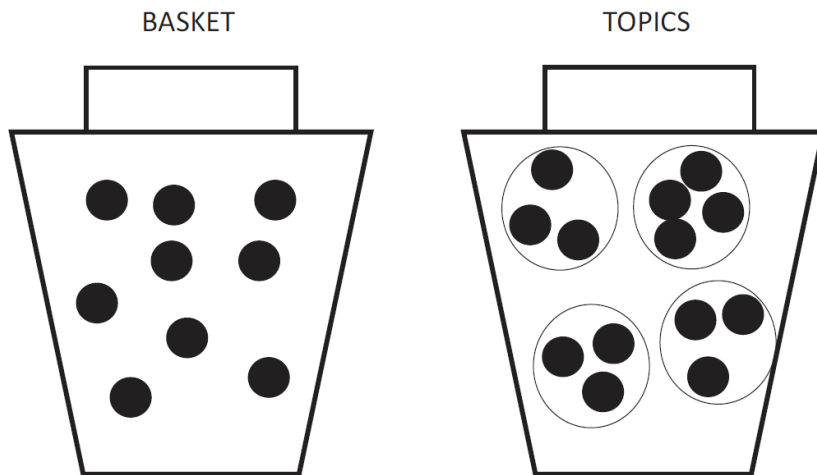| ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
...	...

market basket transaction

{Diapers, Beer}   **Example of a frequent itemset**

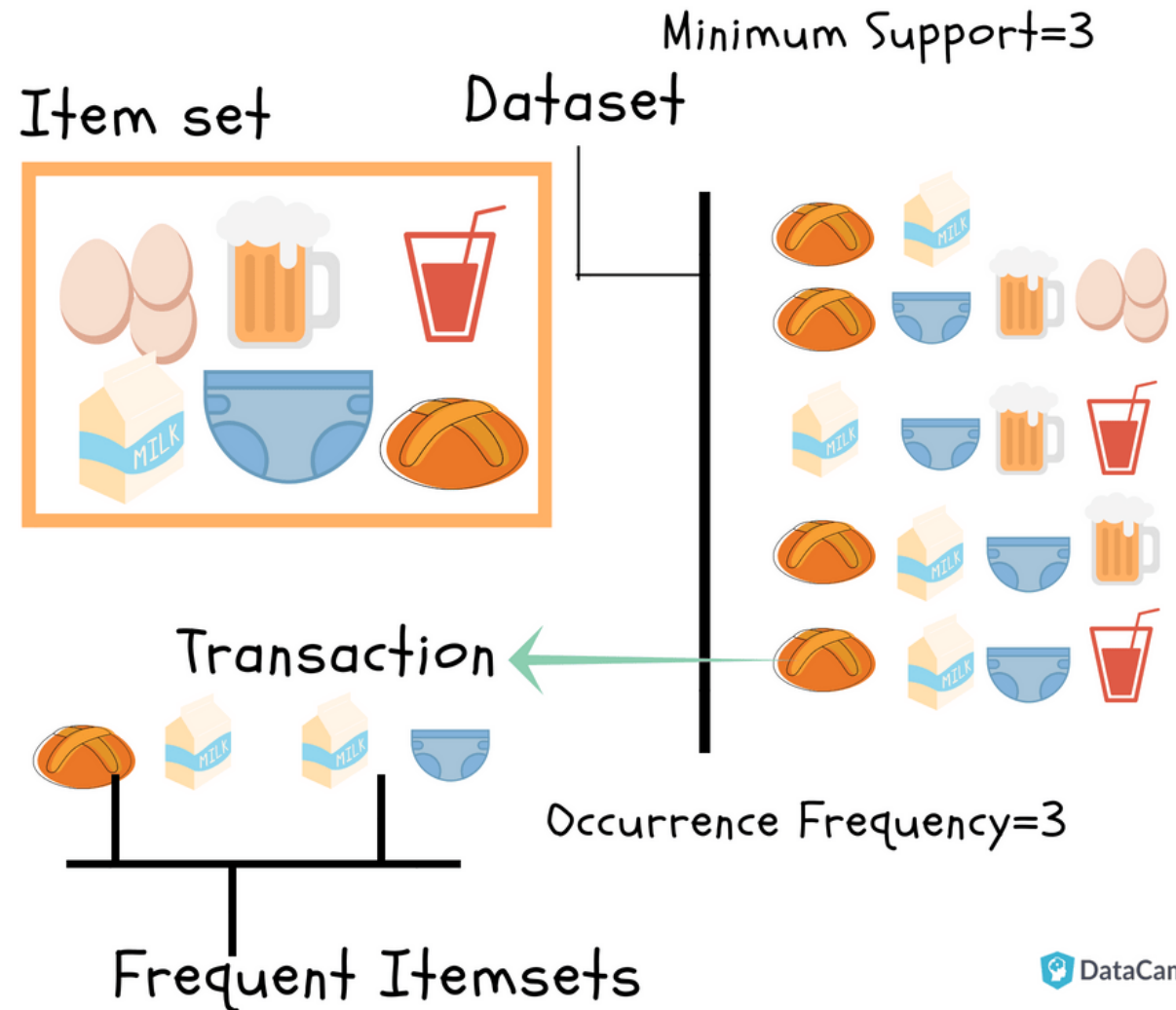{Diapers} → {Beer}   **Example of an association rule**

BASKET   TOPICS

Figure 2.9 Basket vs. topics

By Bingqing Xiong

12

❖ Basket Analysis

- Itemset: Collection of one or more items.

- Support Count: Frequency of occurrence of an item-set

- Support(s): Fraction of transactions that contain the itemset X

- Bread & Milk vs. Diapers & Milk

Analysis of Transactions on Aggregate Level

- ❖ Frequency & Proportion

- ❖ Central Tendency

- ❖ Dispersion

- ❖ Segmentation

By Bingqing Xiong

Describing Frequency & Proportion

❖ Count

❖ Percentage

Measure of Central Tendency

Mode

❖ The most frequently occurring value.

❖ If more than one score occurs most frequently then all values should be reported.

Median

❖ Central value of a set of scores (i.e. has as many scores above it as below it).

❖ NB : if there are an even number of scores then the median is half way between the two middle points.

Mean

❖ Sum of all the scores divided by the number of scores

$$\frac{\sum x}{n}$$

By Bingqing Xiong

Measure of Dispersion

Range

❖ Range = Maximum value - Minimum value

❖ This value is sensitive only to the two extreme values in a data set, and provides no information about the variability of the scores in between - so, if one data point is unusually large or small (i.e. an outlier) then the picture of the data set as a whole could be distorted.

❖ Outlier = a value which is unusually extreme

Measure of Dispersion

Variance

❖ The variance (S2) is a value which takes account of each of the scores in a sample, and looks at the difference between each of those scores and the mean.

❖ The formula used to calculate the variance is:

$$S^2 = \frac{\Sigma(x - \bar{x})^2}{(N-1)}$$, $\bar{x}$ = mean, N = number of observations

❖ But this can be cumbersome to calculate, so an easier way is to use the formula:

$$S^2 = \frac{\Sigma x^2 - (\Sigma x)^2 / N}{N - 1}$$

❖ The variance describes the amount of spread of a data sample. The larger the spread of the data the larger the variance

By Bingqing Xiong

Measure of Dispersion

Variance

❖ The value (N-1) is used so that data obtained from a sample can be approximated to the value of the population variance.

❖ If only the sample variance was required, then (N) would be substituted for (N-1).

❖ However, sample variance value is only useful for describing the data and is not as useful or rigorous as the population approximation value, and therefore when we mention the variance or any formula requiring the variance without other specification, then the two equations in the previous slide should be used.

Measure of Dispersion

Standard Deviation

❖ The variance uses a squared value, so the numbers obtained are high and not so useful.

❖ Therefore it is more appropriate to refer to the Standard Deviation (S or SD), which is the square root of the variance:

$$S = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{\left(N - 1\right)}} \quad \text{Or} \quad S = \sqrt{\frac{\sum x^2 - \left(\sum x\right)^2 / N}{N - 1}}$$

❖ (N-1) is used to make an estimate of the population SD on the basis of information from a sample.

❖ Standard deviation serves as the indication of how the values of a variable are spread around the mean and the indication of the usefulness of the mean as typical value.

An Introduction to Market Segmentation

Segmentation

❖ Businesses usually segment their customers using

- Demographic information (age, gender, income, education etc..)

- Geographical information (region, city, country)

- Psychographics (lifestyle, values, personality)

- Behavioural data (RFM, product usage)

Segmentation Using RFM

Steps:

1. <u>Filter the dataset</u>

2. <u>Exam the dataset</u> (data cleaning is necessary !!!)

3. <u>Summarize the dataset</u>

4. <u>Put together the RFM Report</u>

5. <u>Interpreting the results</u>

Note: Detailed instructions will be shared at Lab 4

by Bingqing Xiong

Segmentation Using RFM

❖ Interpreting the results

- Look at recency, frequency and monetary separately

  - X, High F, X: frequent buyers – loyalty programmes

  - X, X, High M: highest paying customers – premium offers, luxury products

  - … …

- Taking these three aspects together

  - High R, High F, High M: highly engaged customers – loyalty program, focusing on value-added offers

  - … …

Manual Bivariate Segmentation/Two-Dimensional Segmentation

Pro: a useful technique for businesses that have limited resources or data analysis capabilities

Con: less precise than more advanced segmentation methods
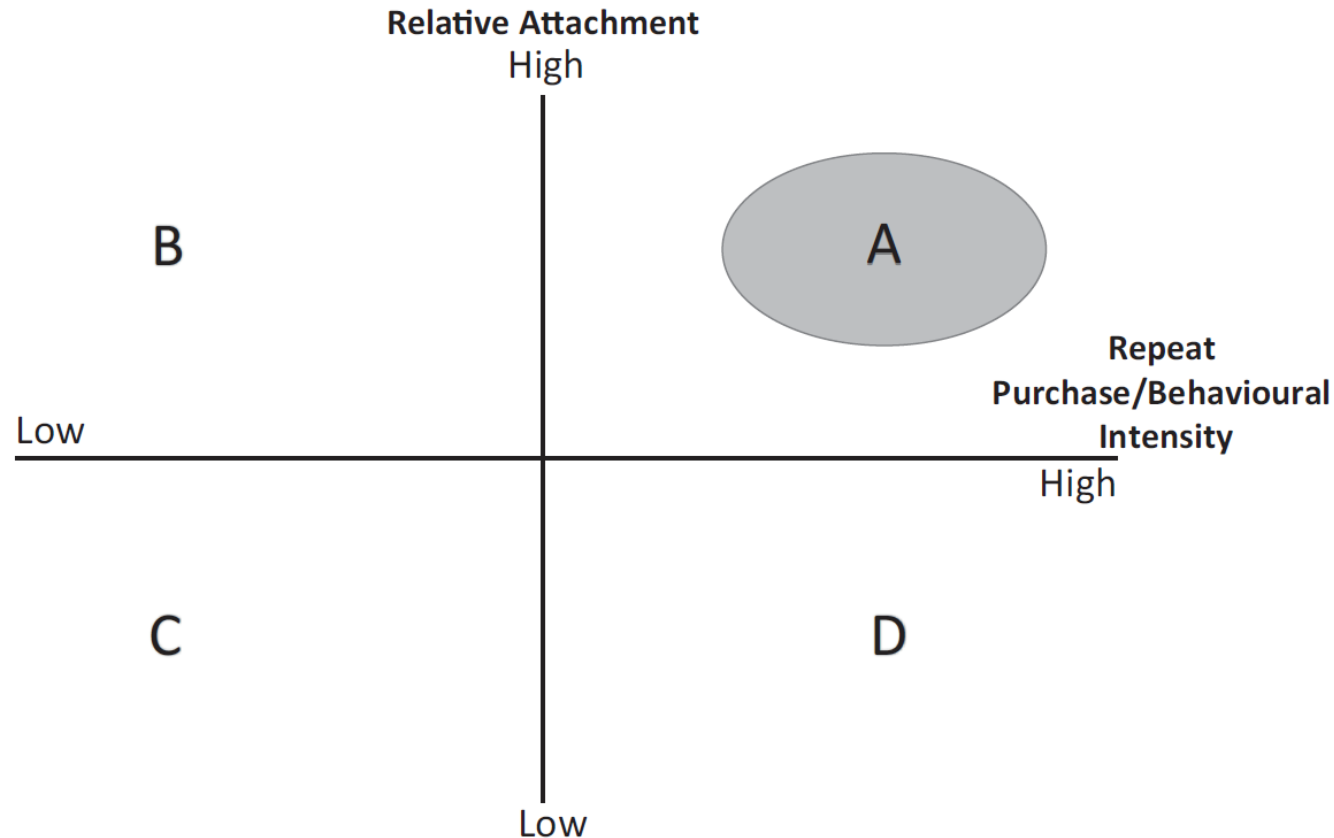
Relative Attachment
High

B

A

Low

Repeat Purchase/Behavioural Intensity
High

C

D

Low

Source: Andrew Smith (2009) Consumer Behaviour and Analytics, Routledge.

By Bingqing Xiong

Manual Bivariate Segmentation
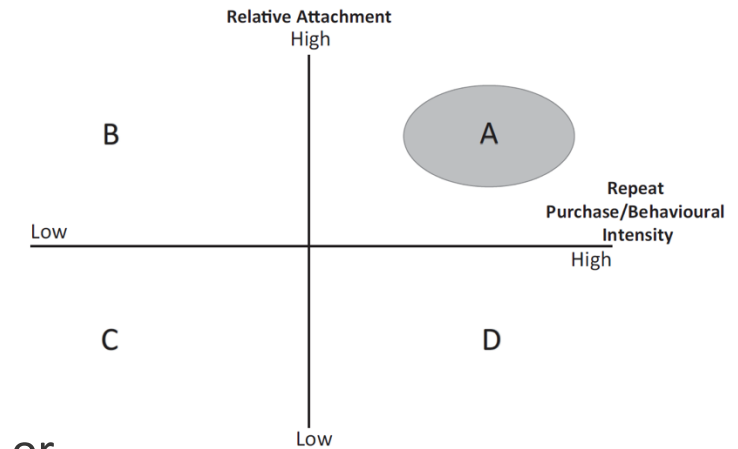
Source: Andrew Smith (2009) Consumer Behaviour and Analytics, Routledge.

❖ **A - Composite loyal**

➢ Consumers buy the product at relatively high rates of repeat purchase or intensity; positively attached to the product

❖ **B - Constrained loyal**

➢ Consumers have positive attachment to the product, but this does not manifest in behaviour and purchase

❖ **C - Disloyal**

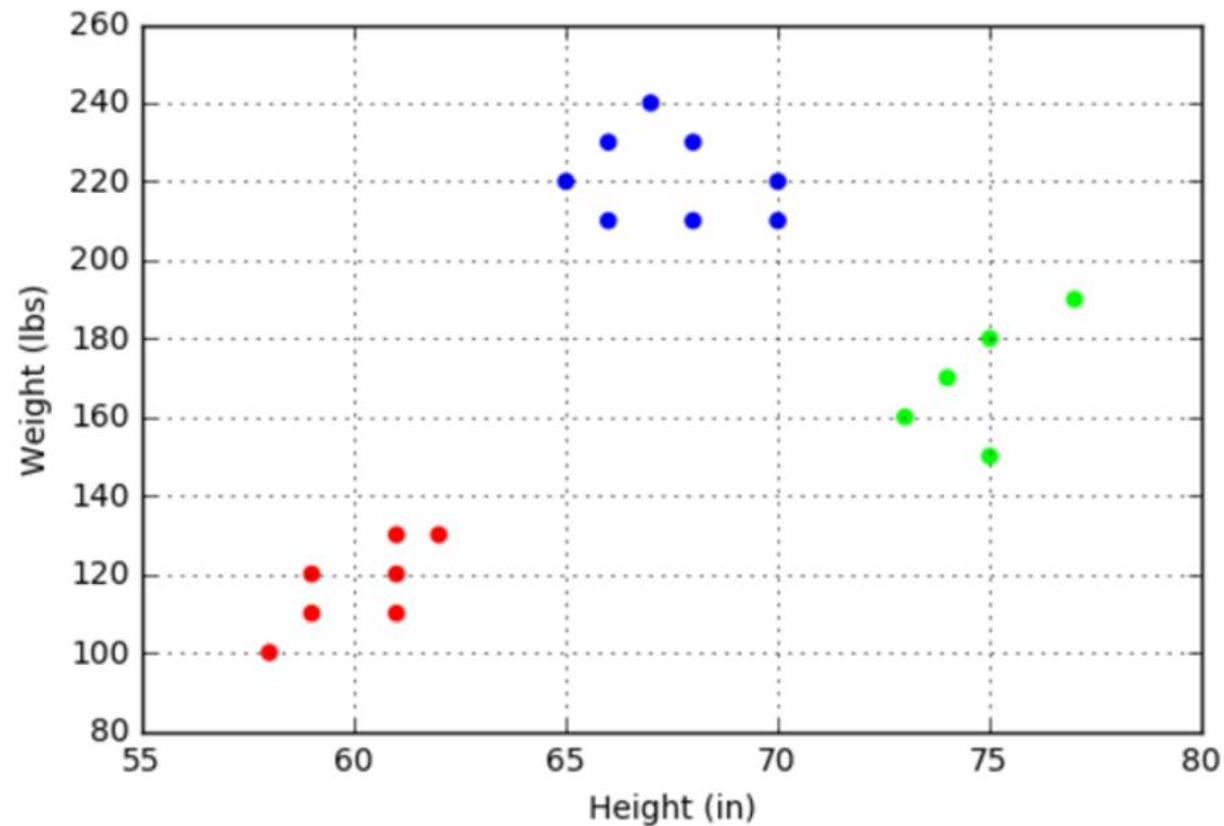➢ Consumers do not like or buy the product a great deal

❖ **D - Specious loyal**

➢ Consumers do not like the product or brand buy buys it nonetheless

26

Discussion

Reasons behind why specious loyal

Loyalty Programs/ Social Pressure/ Limited options

Segmentation by Scatter Plot

❖ T-shirt company

By Bingqing Xiong

28

Clustering Analysis for Segmentation

❖ Segment customers to achieve more effective customer marketing via personalization

❖ K-means cluster ana...

- An unsupervised learning m... overlapping clusters where c...

- Identifies clusters by minimi...

- To decide the number of clu... within-cluster sum of square... level off

- Interpret each cluster by con...

fig 1: before applying
k-means clustering

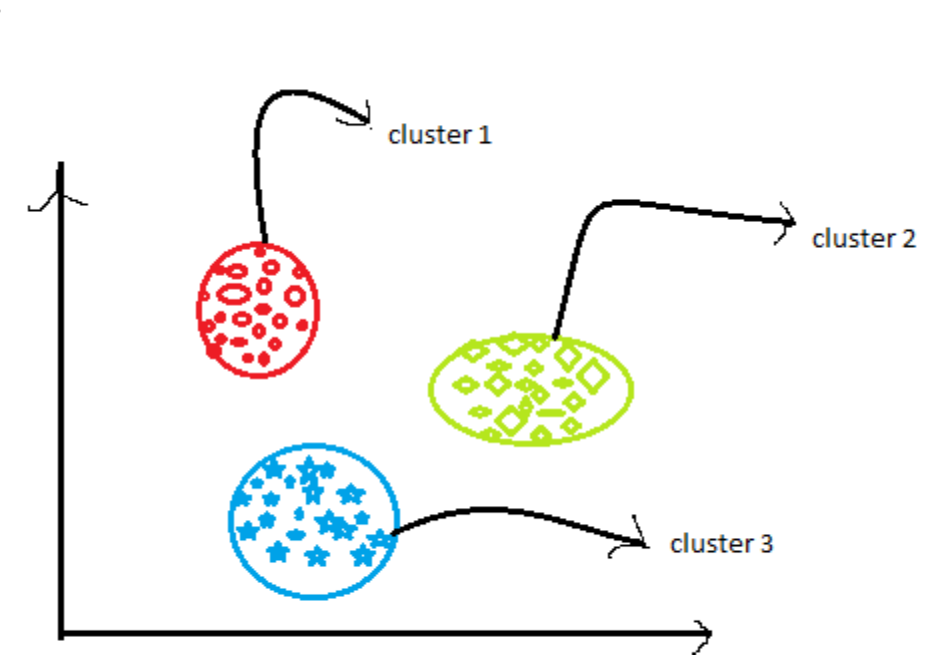
cluster 1

cluster 2

cluster 3

fig 2: After applying K-
means clustering

By B

To be continued in the next week …

❖ Transaction Data

❖ Analysis Levels: Individuals & Aggregation

❖ Segmentation