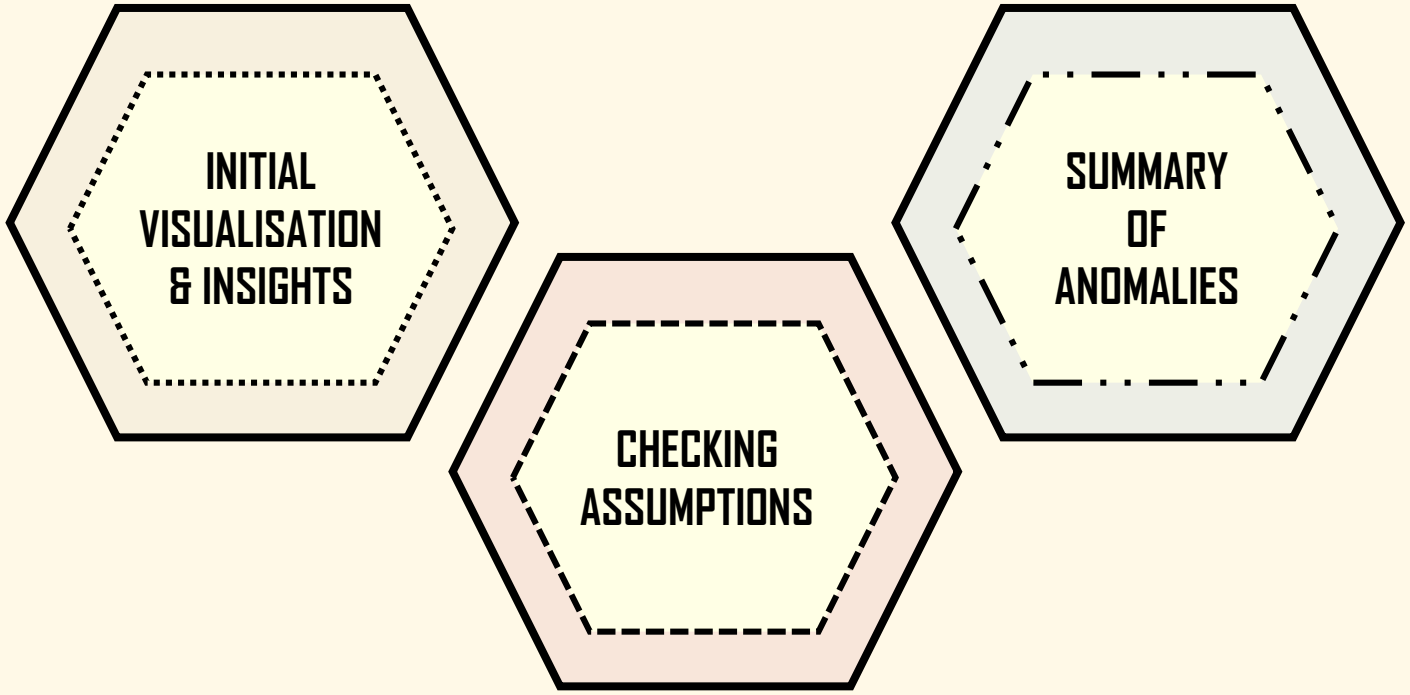
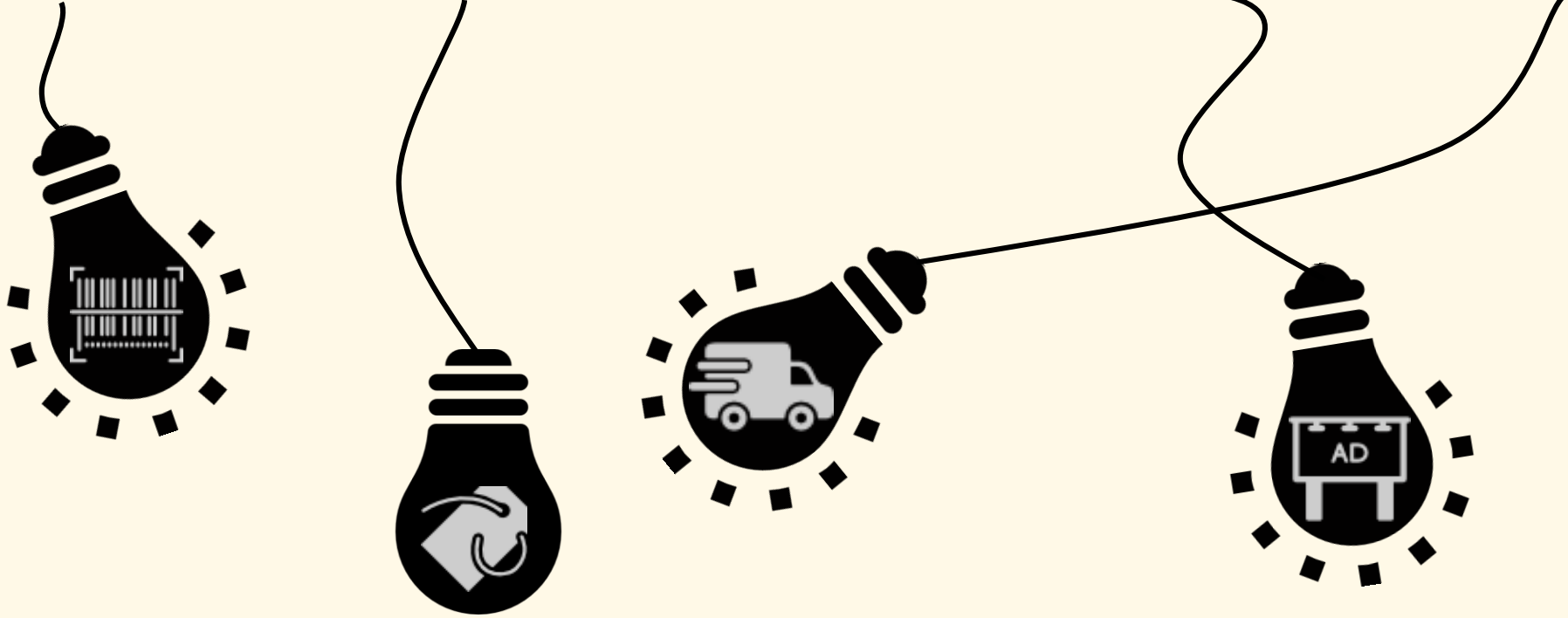
PROJECT UPDATES

6 OCT 2016

Liu Bowei | Teo Hui Min | Pong Chong Xin

AGENDA


INITIAL VISUALISATION & INSIGHTS
CHECKING ASSUMPTIONS
SUMMARY OF ANOMALIES



INITIAL VISUALISATION

WITH TABLEAU WORKBOOK

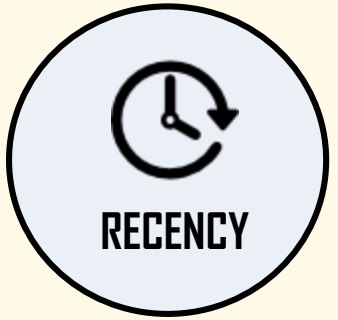
RFM ANALYSIS IN MARKETING

Company

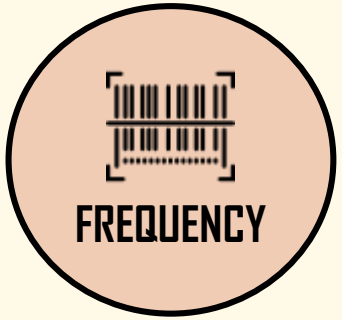
Supported by

Consumer Base

Sorted based on



and



and

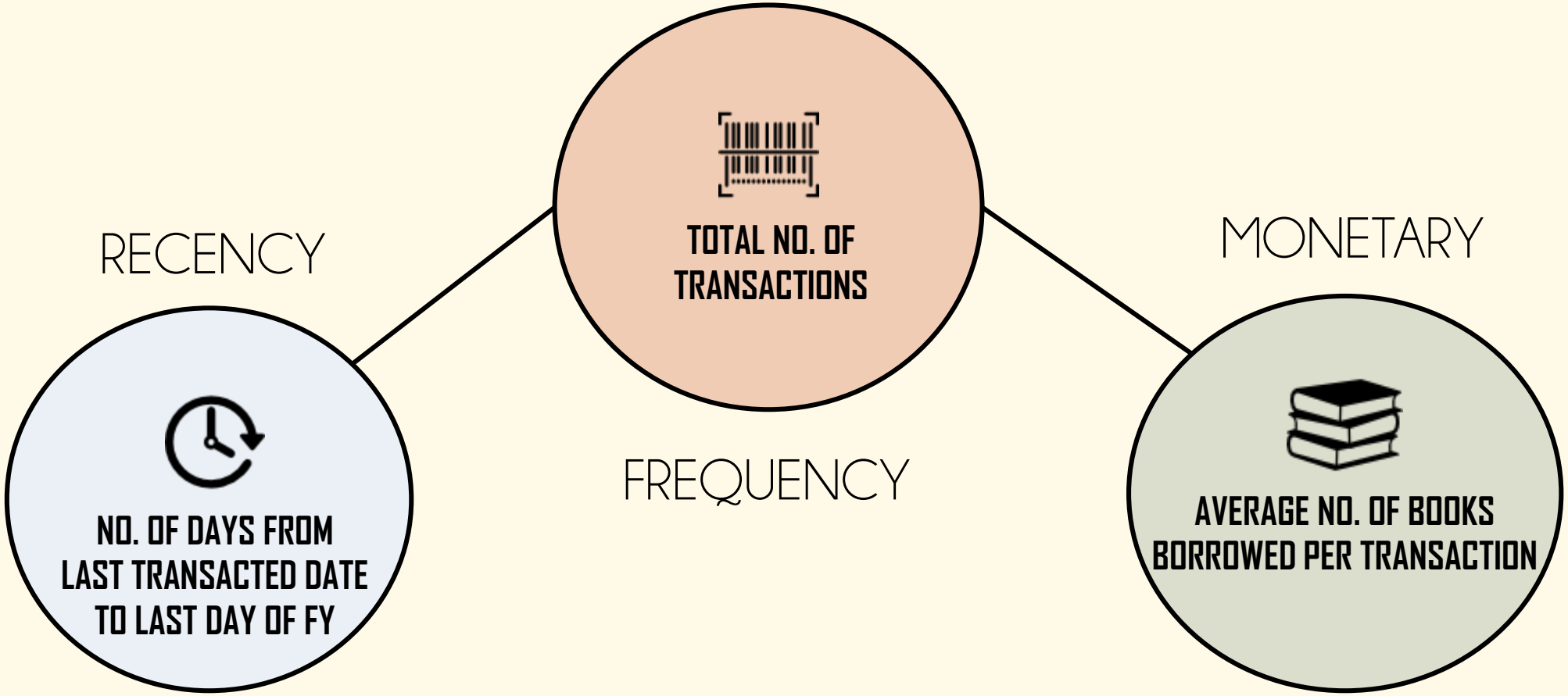
Scoring 1-5 on each metric

5-5-5

Best customers

ADJUST MARKETING PLANS ACCORDINGLY

PATRON PROFILING



RFM WORKFLOW

Raw data → Removed anomalies → Calculate **Recency** by: *365 - (End Date – Last TXN Date)* → Calculate **Frequency** by: *Total no. of TXN* → Calculate **Monetary** by: *(Total no. of books borrowed) / (Total no. of TXN)* → Performed a log-transformation on the data based on R, F, and M → Standardized the log-transformed data → Conducted cluster analysis *(k-means algorithm)* → 6 distinct clusters were formed

PATRON PROFILES

Cluster 1: 166,768 patrons

- Active, Frequent, Low Volume
- High R, High F, Low M

Cluster 2: 117,572 patrons

- Inactive, Non-Frequent, High Volume
- Low R, Low F, High M

Cluster 3: 179,666 patrons

- Active, Non-Frequent, High Volume
- High R, Low F, High M

Cluster 4: 117,799 patrons

- Inactive, Non-Frequent, Low Volume
- Low R, Low F, Low M

Cluster 5: 178,875 patrons

- Active, Frequent, High Volume
- High R, High F, High M

Cluster 6: 134,859 patrons

- Active, Non-Frequent, Low Volume
- High R, Low F, Low M

PATRON PROFILES

Active
- Clusters 1, 3, 5 & 6

High Frequency
- Clusters 1 & 5

High Volume
- Clusters 2, 3 & 5

Inactive
- Clusters 2 & 4

Low Frequency
- Clusters 2, 3, 4 & 6

Low Volume
- Clusters 1, 4 & 6

CHECKING OF ASSUMPTIONS

- Definition of 1 transaction
  - All records in the TXN_FY13 and TXN_FY14 csv files with the same:
    - Txn Date
    - Patron UID
    - Branch Code
- There is a maximum of 32 books in 1 transaction
- Patron location is set as the centroid of the subzone the patron is in

ANOMALY 1

Locale Planning ADZID contains *"Bad Value"* and *"Missing Value"*

- Patron Dataset

  2013 *Locale Planning ADZID* Column (2,355,651)
  - Count of Missing Value (23,822)
  - Count of Bad Value (35,866)
  - Count Missing & Bad Value (59,688)
  - % Missing & Bad Value (2.534%)

  2014 *Locale Planning ADZID* Column (2,462,270)
  - Count of Missing Value (33,808)
  - Count of Bad Value (37,010)
  - Count Missing & Bad Value (70,818)
  - % Missing & Bad Value (2.876%)

Patron Gender	Patron Activ...	Locale Planning ADZID
Female	0	Bad Value
Female	1	Bad Value
Female	1	Bad Value
Female	1	Bad Value
Female	0	Bad Value
Male	0	Bad Value

ANOMALY 2

Patron UID contains *"0"*

- Transaction Dataset

  TXN_FY13

  - 19,828 records have attribute *Patron UID = "0"*

  TXN_FY14

  - 641 records have attribute *Patron UID = "0"*

	Txn Date Time	Branch Code	Circulation Type Code	Item Barcode	Patron Borrower Category Code	Patron UID
1	2013/04/01 12:00 AM	TRL	CH	A00700688G	NLTSRT	0
2	2013/04/01 12:00 AM	TPPL	CH	A00585013K	SCJB	0

ANOMALY 3

Branch Code contains *"Bad Value"* and *"Missing Value"*

- Transaction Dataset

  TXN_FY13

  - 3 records with *Branch Code* set to "*Bad Value*".
  - 3,065 records with *Branch Code* set to "*Missing Value*".

  TXN_FY14

  - 9,467 records with *Branch Code* set to "*Bad Value*".
  - 1,600 records with *Branch Code* set to "*Missing Value*".

	Txn Date Time	Txn Date	Branch Code	Circulation Type Code	Item Barcode	Patron Borrower Category Code	Patron UID
1	2013/05/27 12:00:00 AM	2013/05/27	Bad Value	RN	B22588560C	SCAB	503867
2	2013/05/31 12:00:00 AM	2013/05/31	Bad Value	RN	B10603300B	SCAB	30718
3	2013/06/13 12:00:00 AM	2013/06/13	Bad Value	RN	B25842032D	SPPARTNERA	72019
4	2013/09/01 12:00:00 AM	2013/09/01	Missing Value	RN	B22524371D	SPPARTNERA	31284
5	2013/09/01 12:00:00 AM	2013/09/01	Missing Value	RN	B25529504J	SPPARTNERA	153726

ANOMALY 4

Records have *Avg. No. of Books Borrowed Per Transaction > 32*

- Transaction Dataset

  TXN_FY13

  - 893 patrons with *Avg. No. of Books Borrowed* > 32

  TXN_FY14

  - 779 patrons with *Avg. No. of Books Borrowed* > 32

	Patron UID	Patron Borrower Category Code	Locale Planning ADZID	Total No. of TXN (F)	Total No. of Books Borrowed	Avg No. of Books ...
1	1311394	DEAR	TMSZ04084	1	7998	7998
2	1322024	DEAR	TMSZ04084	1	6000	6000
3	1456682	DEAR	TMSZ02305	1	4000	4000

ANOMALY 4

Records have *Avg. No. of Books Borrowed Per Transaction > 32*

- From the records with *Avg. No. of Books Borrowed* > 32, most of the records have attribute *Patron Borrower Category Code = "DEAR"*
- 775 patrons with attribute *Patron Borrower Category Code = "DEAR"* (2013)
- 705 patrons with attribute *Patron Borrower Category Code = "DEAR"* (2014)
- The records have unrealistic values in attribute *Patron Birthyear* i.e. *"1900"*, *"2015"*, *"2016"* etc.
- The records have value *"Others"* as *Patron Citizenship*, *Patron Race*, *Patron Gender*

	Patron UID	Patron Borrower Category Code	Locale Planning ADZID	Total No. of TXN (F)	Total No. of Books Borrowed	Avg No. of Books ...
1	1311394	DEAR	TMSZ04084	1	7998	7998
2	1322024	DEAR	TMSZ04084	1	6000	6000
3	1456682	DEAR	TMSZ02305	1	4000	4000

ANOMALY 5

Presence of *Branch Codes* that do not exist in Collection Dataset

- Transaction Dataset
  - Branch codes do not exist in *Collection_Dataset_FY13 and FY14*

	Branch Code	N Rows
1	07LKCRL	422
2	08LKCRL	2137
3	09LKCRL	989
4	11LKCRL	1013
5	AMKPL	1583222
6	Bad Value	3
7	BBPL	1241287
8	BEPL	1443306
9	BIPL	1914124
10	BMPL	1171022
11	BPPL	1238467
12	CCKPL	1491890
13	CMPL	1840523
14	CNPL	178149
15	CSPL	1258112
16	CTPL	1578072
17	EPPL	307836
18	GEPL	1190552

SUMMARY OF ANOMALIES

- Patron Dataset
  - Under column heading *Locale Planning ADZID*, records have values set to *"Bad Value"* and *"Missing Value"*

- Transaction Dataset
  - Under column heading *Patron UID*, records have values set to *"0"*
  - Under column heading Branch Code, records have values set to *"Bad Value"* and *"Missing Value"*
  - When aggregating transaction records, some records have Avg. No. of Books Borrowed Per Transaction exceeding 32
  - Presence of *Branch Codes* that are not listed in the Collection Dataset