
Data Analysis & Exploration on Retail Transaction

Instructions

In the dataset, you would have anonymised individual transactions data for our customers from Jan 19 to Sept 19. Please explore the data set, and answer the following questions:

- What are the key customer segments?
 - Customers ceasing to transact is often an early indicator that they are moving to another provider or closing their business. Suppose you are the analyst working with the customer retention team, how would you go about identifying such customers before they churn?
 - What would be your recommendations to retain these customers? How could the identification and retention of these customers be improved?
-

Customer Segments

→ Which variables could be useful for segmentation?

Looking at the table 1, it shows **transaction** data of customers through from Jan 2019 to Sep 2019. It also shows what **industry** each customer is in

→ How to segment customer using transaction and industry?

We can look at the transaction **value** and **frequency** by industry - table 2

→ Let's visualise table 2 after converting it to monthly data

customer_guid	transaction_guid	authorisation_datetime	transaction_type	transaction_amount_gbp	entry_mode_type	card_type	mcc_group
97bb79db4c5fc91	d05aa606df57c0ba66b66cce5aabcce7	2019-08-25 17:53:07	Sale	17.00	Ecom	CREDIT	F&B Off-Premise
f6fbddddd43651dd2	cc0ff2e88a3b175adee250a48acf1bf1	2019-08-25 18:30:21	Sale	15.00	Ecom	CREDIT	F&B Hospitality
10a98ca80da7041b	815507cde3b72ab3a497fb2cad247373	2019-08-25 18:16:03	Sale	4.50	Chip&Pin	DEBIT	F&B Off-Premise
10a98ca80da7041b	b1911c4b6e2440204fc131f8d16e6c7f1	2019-08-25 17:15:04	Sale	2.35	Chip&Pin	DEBIT	F&B Off-Premise
10a98ca80da7041b	c32478f613f44750be3fe848b1a978d7	2019-08-25 15:25:20	Sale	10.10	Chip&Pin	DEBIT	F&B Off-Premise

Table 1

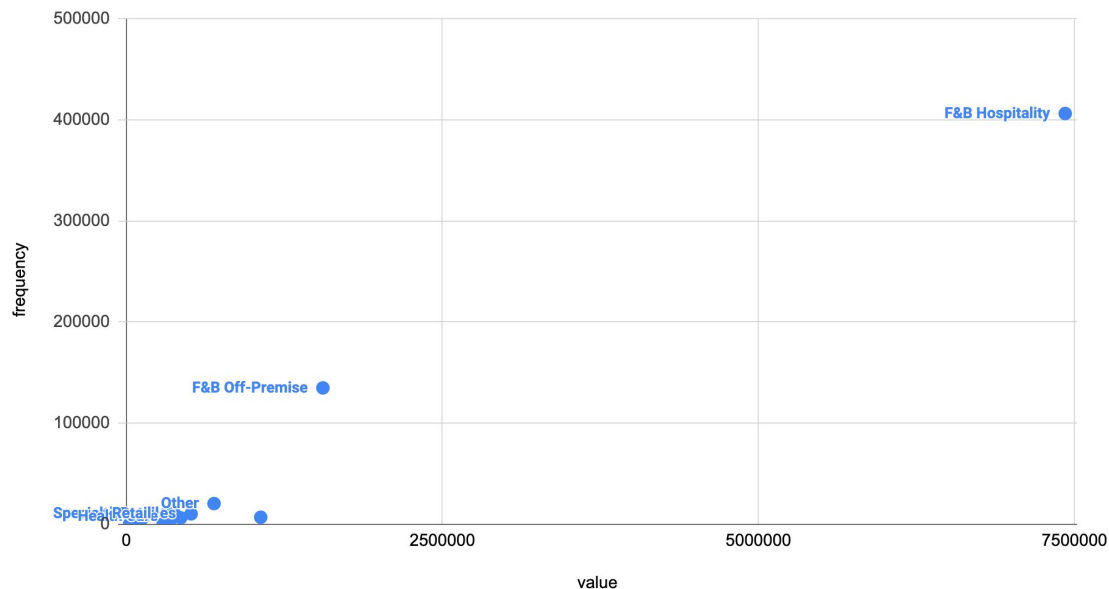


mcc_group	value	frequency
F&B Hospitality	7.427066e+06	406218.777778
F&B Off-Premise	1.555197e+06	134801.444444
Automotive	1.063433e+06	6828.777778
Other	6.940514e+05	20382.777778
Personal Services	5.128004e+05	10216.111111
Apparel/Clothing	4.311051e+05	6241.111111
Specialty Retail	3.790350e+05	10550.444444
Health Care	3.704707e+05	7952.000000
Accommodation & Travel	3.566764e+05	5346.555556
Retail	3.176362e+05	10411.666667
Trades & Trade Supplies	2.914130e+05	1074.222222
Entertainment / Recreational	1.310439e+05	6217.333333
Business Services	9.739372e+04	1888.666667
Home/Office Furnishing	9.335412e+04	273.333333
Garden & DIY	5.499779e+04	3302.777778
Professional Services	2.666312e+04	175.555556

Table 2

Customer Segments

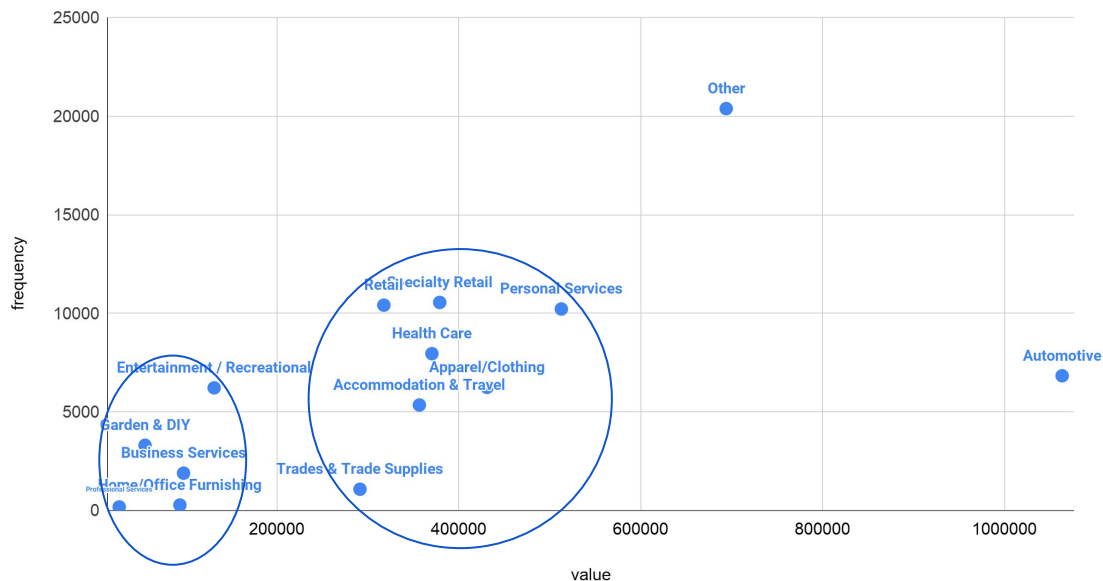
Monthly average transaction frequency vs value by industry



- **Highlight 1**
F&B Hospitality and **Off-Premise** with high frequency and high value
- **Highlight 2**
Everything else is centered in the left bottom
- **Let's take a closer look**

Customer Segments

Monthly average transaction frequency vs value by industry



→ **Highlight 3**
Services and entertainment form a cluster with low frequency and value (under 200k)

→ **Highlight 4**
Retail & Trade carries a higher frequency and value (400k - 600k)

→ **Highlight 5**
Other let's leave it out for now

Customer Segments

- **Segment 1: High value, High frequency**
F&B including: F&B Hospitality, F&B Off-Premise
- **Segment 2: Comparatively high value and high frequency**
Auto including: Automotive
- **Segment 3: Medium value, Medium frequency**
Retail & Trade including: Personal Services, Specialty Retail, Retail, Health Care, Apparel/Clothing, Accommodation & Travel, Trade Supplies
- **Segment 4: Low value and low frequency**
Service & Entertainment including: Entertainment/Recreational, Garden & DIY, Business Services, Home/Office Furnishing, Professional Service
- **Segment 5: Comparatively high value and high frequency**
Other including: Other

Churn Identification

→ Early indicator of churning

Most often, Customers cease to transact

→ How can data help?

1. We have record of each transaction. Time and value are super helpful. We could manipulate the data to see *the number of daily transactions* each customer makes - **frequency** and the *total volume each customer transacts daily* - **value**

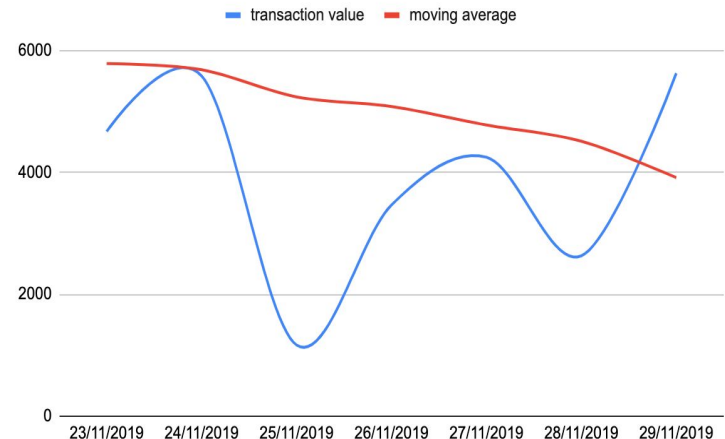
2. At this stage, we noticed that the frequency and value fluctuate so it's hard to us to say if a customer shows signs of churning. However, we could calculate [Moving Averages](#) for **value** and **frequency**. This should reduce the fluctuations and give a smoother trend whether it's upward, stable or downward. Let's look at an example of a customer and it's recent 7 days total transaction value vs. moving average total value (see graph)

3. If we look at the moving average for frequency and value of each customer, we can tell if there is an **downward trend**. And we add two additional columns for to denote it

4. At last, we turn "declines" and "increases" into numeric value <1, 0> and add the scores in one column. If we look at **the total for most 7 recent days* for each customer**, we can tell if there is an downward trend - the higher the score, the more downward trend is. And we come up with a list of customers whose scores are above 12 (the maximum is 14 for each customer)

* most 7 days is my educated guess but we can look at the churn data to decide the cut-off days

transaction value and moving average



Recommendation

→ Now we have a list of risky customers.

How do we go about it?

Churning is inevitable but we need to retain as many customers as can. Resources such as man power is limited so We need to be smart when allocating resources. We can **rank customers by previous transaction value** and **prioritise customers who have higher value**

→ What could have helped us understand?

I would check for:

1. The number of times listed customers have contacted Paymentsense
2. What's our response time
3. Are they happy with the service
4. Was there transaction failure. If so, how many times
5. How much time they have spent on our FAQ page etc.

customerid	tran_amt	industry
25f8d7777a	632,024	Accommodation & Travel
a2a716c6f35	426,663	F&B Hospitality
cf1f3b24471	177,391	Accommodation & Travel
6c36d6fef9d	164,366	Automotive
988cdca7fa5	130,406	Personal Services
d4006d7a76	102,180	F&B Hospitality
63caaf030bf	100,033	Automotive
231972c7ba	91,576	Specialty Retail
8f3b18669a3	83,186	F&B Hospitality
44d530ddc7	72,422	Other
da5be23e70	71,524	Other
d3dabae7c	63,998	F&B Hospitality
4e6acf0d34e	50,493	Trades & Trade Supplies
8d9c3b5ae8	48,960	Apparel/Clothing
6647b247cc	48,505	F&B Hospitality
7b8e618abf7	45,510	Personal Services
b26ba14f4c5	39,737	Entertainment / Recreational
9378e96979	38,001	F&B Off-Premise
3907e15114	37,334	F&B Hospitality
4dff40e8c6c	21,092	Other
a223e8fd7b1	19,829	Accommodation & Travel
e122d8d2eb	19,378	Automotive
5e37252bd2	16,731	Garden & DIY
c5e961e2ad	13,637	F&B Hospitality
373a27c673	12,305	Automotive
1cd0588012	11,461	F&B Hospitality
e71e7955da	9,464	Retail
3a04532a5b	9,393	Entertainment / Recreational
286d06d766	7,124	Automotive
fbe8e3b19f6	3,695	Automotive
94e786cdfd2	2,782	F&B Off-Premise
a7b60012f7a	2,489	Personal Services
5d2e6dea88	1,788	F&B Off-Premise
526dd5448	1,160	Entertainment / Recreational
5e6a6fe4e73	269	Automotive
b5a9458aec5	129	Entertainment / Recreational

Recommendation

Deeper look into the list:

→ Have we served each industry equally good?

Have we served each industry equally good? Let's see **the number of customers we serve** in each industry vs. **the number of risky customer** in each industry.

I would like to find out why the **risky rate for industries like Entertainment / Recreational and Garden & DIY** is so much higher than average of 1.86%

→ Is risky rate high in our high-value segments?

Yes, **the risky rate is high Automotive and Accommodation & Travel**. It's also worth investigating as these two industry carry high business weight to us

industry	risky customer	total customer	risky rate
F&B Hospitality	8	803	1.00%
Automotive	7	155	4.52%
Entertainment / Recreational	4	30	13.33%
Personal Services	3	116	2.59%
Other	3	93	3.23%
F&B Off-Premise	3	341	0.88%
Accommodation & Travel	3	48	6.25%
Trades & Trade Supplies	1	24	4.17%
Specialty Retail	1	77	1.30%
Retail	1	60	1.67%
Garden & DIY	1	11	9.09%
Apparel/Clothing	1	74	1.35%
Grand Total	36	1935	1.86%

Recommendation

→ Understand why the customer is transacting less

We have come up with a list of risky customers and then we need understand why. Are they closing down their business? Are they switching to our competitor? Or it's just natural cycle of the business

→ How we can incentivise them?

For customers with high frequency (make many transactions), we can consider **lowering the commission rate depending on their transaction value** so customers want to stay with us

For customers with high value, we can offer them **tiered pricing** so customers will want to continue using our service for more value

industry	risky customer	total customer	risky rate
F&B Hospitality	8	803	1.00%
Automotive	7	155	4.52%
Entertainment / Recreational	4	30	13.33%
Personal Services	3	116	2.59%
Other	3	93	3.23%
F&B Off-Premise	3	341	0.88%
Accommodation & Travel	3	48	6.25%
Trades & Trade Supplies	1	24	4.17%
Specialty Retail	1	77	1.30%
Retail	1	60	1.67%
Garden & DIY	1	11	9.09%
Apparel/Clothing	1	74	1.35%
Grand Total	36	1935	1.86%