

Analysing an online retail dataset

A task for  bjss

by Gianmarco Mendiola Colan



Content



1. Data Preparation



2. Sales Performance



3. Customer Analysis



4. Recommendations





1. Data Preparation



2. Sales Performance



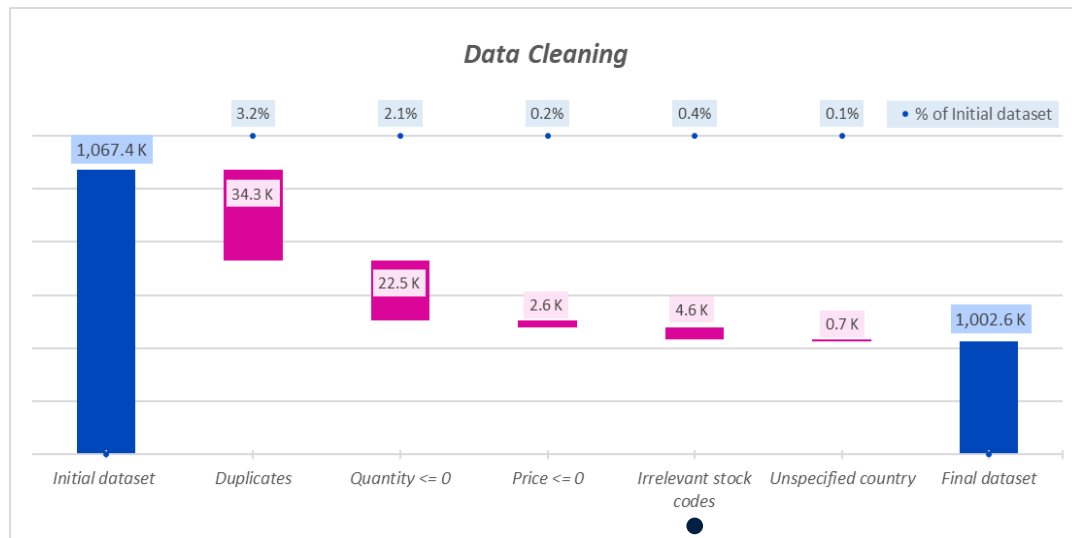
3. Customer Analysis



4. Recommendations



The initial dataset was reduced by 6% and had to be treated before reporting. Aggregated views were also constructed



Data treatments:

- ✓ Multiple descriptions for a single stock code
- ✓ Stock codes were case sensitive
- ✓ Abbreviated countries (i.e., 'EIRE', 'RSA')

Additional features:

- + Date columns (Year, Month, Day of Week)
- + Time (Hour of Day)
- + Recency

Aggregated views:

1. Transactions
2. Product
3. Customer

This included stock codes associated to: gift cards, bank transactions, commissions, debt adjustments, test entries, cancelations, etc.



1. Data Preparation



2. Sales Performance



3. Customer Analysis

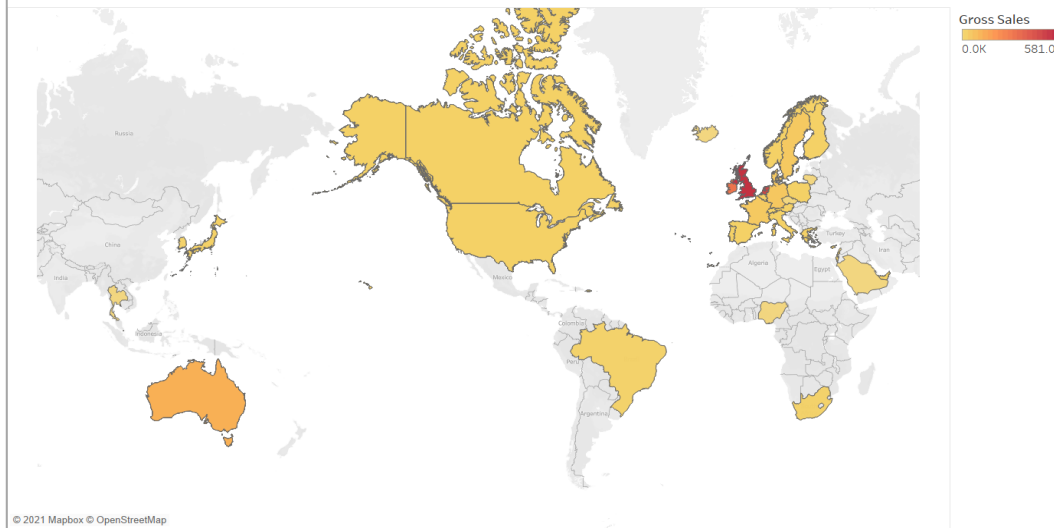


4. Recommendations



The UK comprises more than 85% of total gross sales and experiences a -0.5% decline year over year (YoY)

Gross sales heatmap



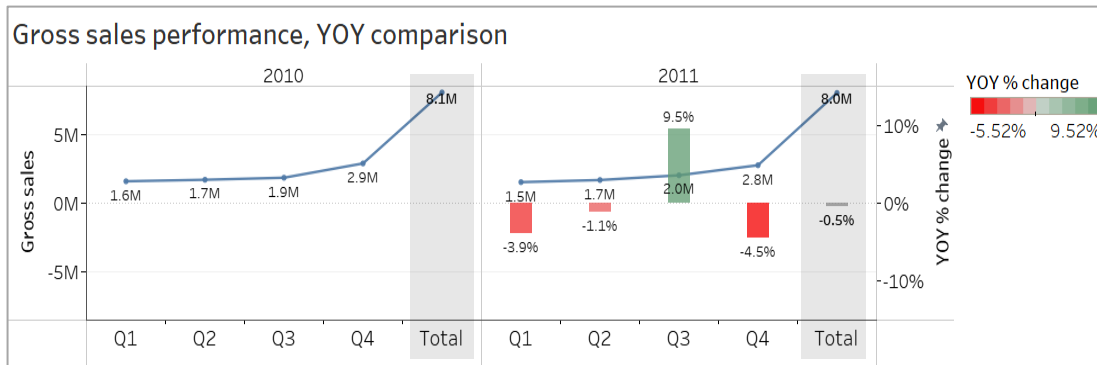
Countries: Gross sales year over year (YOY) changes

COUNTRY	2010			2011		
	Gross sales	% of Total	% YOY change	Gross sales	% of Total	% YOY change
United Kingdom	8,055,527	85.98%	-0.5%	8,019,066	84.71%	-0.5%
Ireland	343,300	3.66%	-23.9%	261,168	2.76%	-23.9%
Netherlands	259,593	2.77%	6.0%	275,135	2.91%	6.0%
Germany	182,383	1.95%	5.3%	192,137	2.03%	5.3%
France	129,254	1.38%	36.1%	175,963	1.86%	36.1%
Australia	30,390	0.32%	351.3%	137,138	1.45%	351.3%
Switzerland	41,715	0.45%	24.3%	51,841	0.55%	24.3%
Spain	36,010	0.38%	50.0%	54,003	0.57%	50.0%
Sweden	53,039	0.57%	-37.8%	32,995	0.35%	-37.8%
Denmark	48,983	0.52%	-65.1%	17,074	0.18%	-65.1%
Belgium	21,378	0.23%	65.3%	35,327	0.37%	65.3%
Portugal	19,670	0.21%	26.6%	24,903	0.26%	26.6%
Channel Islands	23,357	0.25%	-15.7%	19,684	0.21%	-15.7%
Japan	13,313	0.14%	123.2%	29,711	0.31%	123.2%
Norway	9,543	0.10%	200.4%	28,668	0.30%	200.4%
Italy	13,345	0.14%	13.7%	15,170	0.16%	13.7%
Finland	7,019	0.07%	152.1%	17,692	0.19%	152.1%
Cyprus	9,362	0.10%	23.7%	11,580	0.12%	23.7%
Austria	9,793	0.10%	-12.3%	8,585	0.09%	-12.3%
Greece	13,675	0.15%	-67.6%	4,426	0.05%	-67.6%
Hong Kong	3,922	0.04%	148.4%	9,744	0.10%	148.4%
Singapore	4,038	0.04%	125.9%	9,120	0.10%	125.9%
Israel	3,199	0.03%	154.1%	8,129	0.09%	154.1%
Poland	3,116	0.03%	115.9%	6,726	0.07%	115.9%
United Arab Emirates	7,891	0.08%	-76.4%	1,865	0.02%	-76.4%
USA	4,340	0.05%	-17.5%	3,580	0.04%	-17.5%
Iceland	1,323	0.01%	-171.4%	3,598	0.04%	-171.4%
Lithuania	4,893	0.05%	-100.0%			-100.0%
Malta	2,687	0.03%	-22.9%	2,071	0.02%	-22.9%
Canada	1,217	0.01%	156.1%	3,115	0.03%	156.1%
South Africa	2,367	0.03%	-57.6%	1,002	0.01%	-57.6%
Bahrain	2,561	0.03%	-78.6%	548	0.01%	-78.6%
Thailand	3,071	0.03%	-100.0%			-100.0%
Lebanon	212	0.00%	700.1%	1,694	0.02%	700.1%
Brazil	268	0.00%	326.3%	1,144	0.01%	326.3%
Bermuda	1,253	0.01%	-100.0%			-100.0%
European Community				1,159	0.01%	
Korea	1,119	0.01%	-100.0%			-100.0%
Czech Republic				787	0.01%	
West Indies	536	0.01%	-100.0%			-100.0%
Saudi Arabia				146	0.00%	
Nigeria	113	0.00%	-100.0%			-100.0%
Grand Total	9,368,773	100.00%		9,466,696	100.00%	1.0%

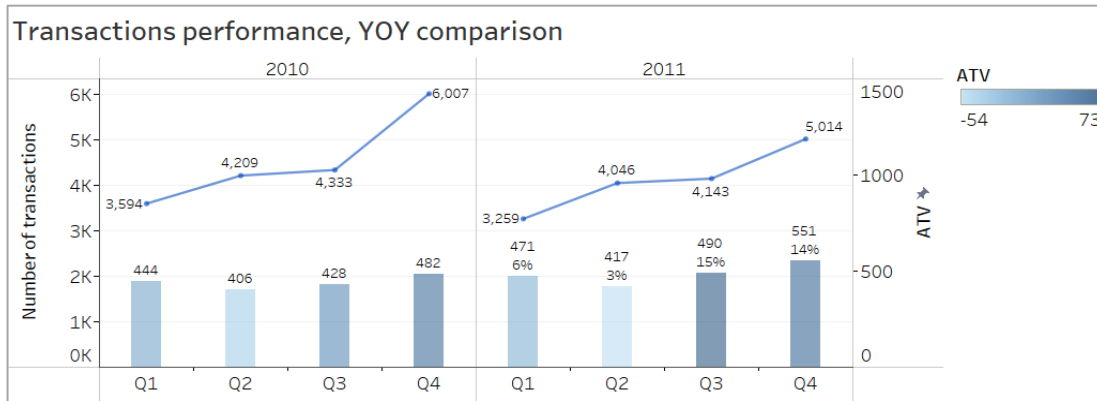
- ✓ The UK comprises **more than 85%** of gross sales and experiences a slight YOY slip
- ✓ Several countries appear to have incomplete data for either 2010 or 2011 and abnormal increments

How sales and transactions change over the two years?

Q3 and Q4 experienced notorious changes in gross sales, while the overall transaction value increased



- While sales soar on Q3 (**9.5%**) 2011, it drops **4.5%** during on Q4.
- Q1 and Q2 also experience a decrease in gross sales.

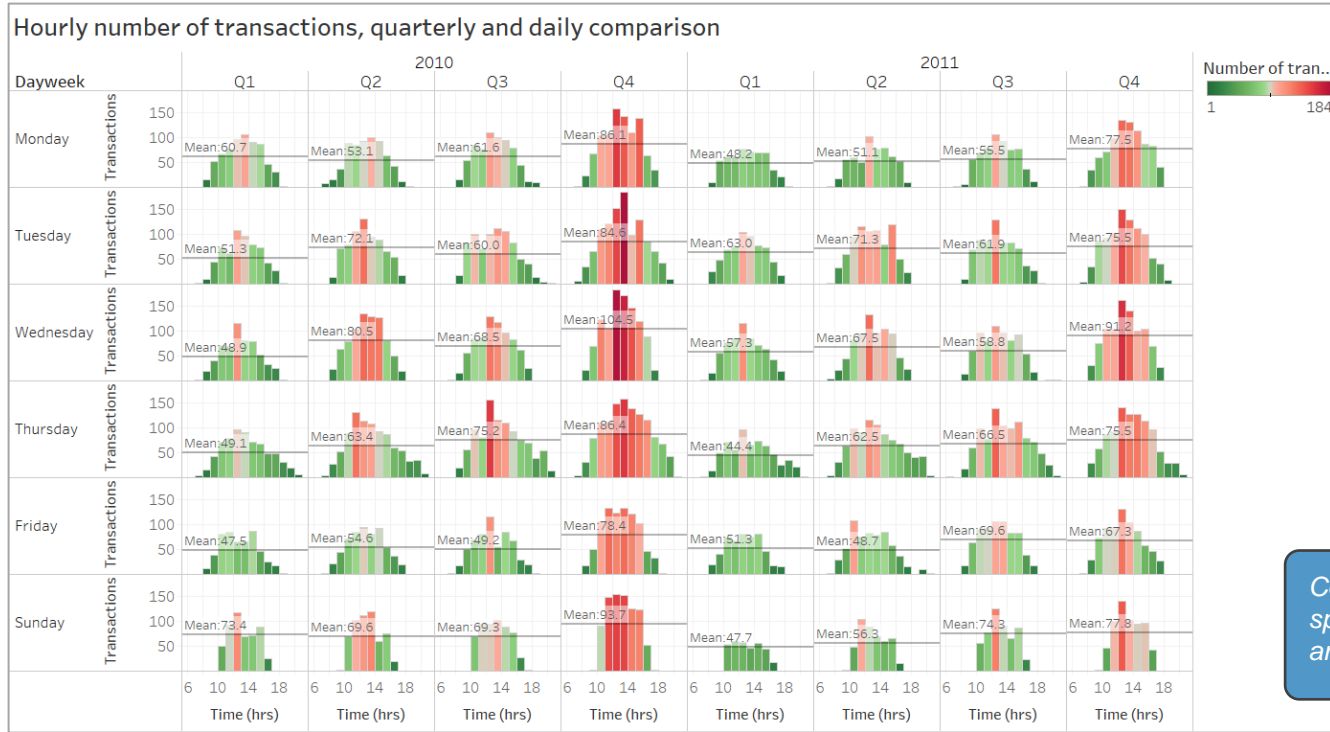


- Despite the gross sales' declines, the **average transaction value (ATV) increased** on every quarter in 2011

*What caused sales to increase on Q3 2011?
How can we maintain the ATV?*

How the number of transactions fluctuates during a certain day?

Peak hours occur during the afternoon and most sales happen during the last quarter



- Q4 experience the highest number of transactions as a result of Christmas.
- Peak hours occur during the afternoon.
- On average, Monday and Friday are the days with less hourly transactions.

Could we adjust some prices during specific time slots to meet the demand and increase revenue?

What are the best selling products?

Some items exhibit significant changes in sales vs the previous year

Best selling (top 20) products and their YOY % change				
Description	2010		2011	
	Total Quantity	% Change Q	Total Quantity	% Change Q
WORLD WAR 2 GLIDERS ASSTD DESIGNS	49,888		44,859	8.81%
JUMBO BAG RED RETROSPOT	42,262		42,100	6.69%
WHITE HANGING HEART T-LIGHT HOLDER	49,267		32,032	-33.90%
SMALL POPCORN HOLDER	30,233		50,847	67.25%
PAPER CRAFT , LITTLE BIRDIE			80,995	
MEDIUM CERAMIC TOP STORAGE JAR			77,036	
ASSORTED COLOUR BIRD ORNAMENT	38,339		31,441	-17.75%
BROCADE RING PURSE	47,511		21,653	-34.32%
PACK OF 72 RETROSPOT CAKE CASES	44,732		21,697	-44.96%
PACK OF 60 PINK PAISLEY CAKE CASES	25,624		19,231	-27.90%
60 TEATIME FAIRY CAKE CASES	27,277		12,458	-53.34%
ASSORTED COLOURS SILK FAN	18,285		20,917	26.49%
VICTORIAN GLASS HANGING T-LIGHT	16,035		22,308	55.70%
PACK OF 12 LONDON TISSUES	15,000		22,513	37.73%
COLOUR GLASS T-LIGHT HOLDER HANGING	21,841		14,965	-28.82%
JUMBO BAG PINK POLKADOT	16,261		18,912	21.21%
PACK OF 72 SKULL CAKE CASES	21,972		11,529	-45.15%
LUNCH BAG RED RETROSPOT	16,929		14,409	-5.35%
STRAWBERRY CERAMIC TRINKET BOX	22,531		8,679	-57.75%
MINI PAINT SET VINTAGE	13,574		11,538	-10.26%

- Some products are vastly sold even in different packaging or styles.
- Small items usually have more sales.

Review more impactful drops in sales, maybe articles are overpriced or need discount



1. Data Preparation



2. Sales Performance



3. Customer Analysis

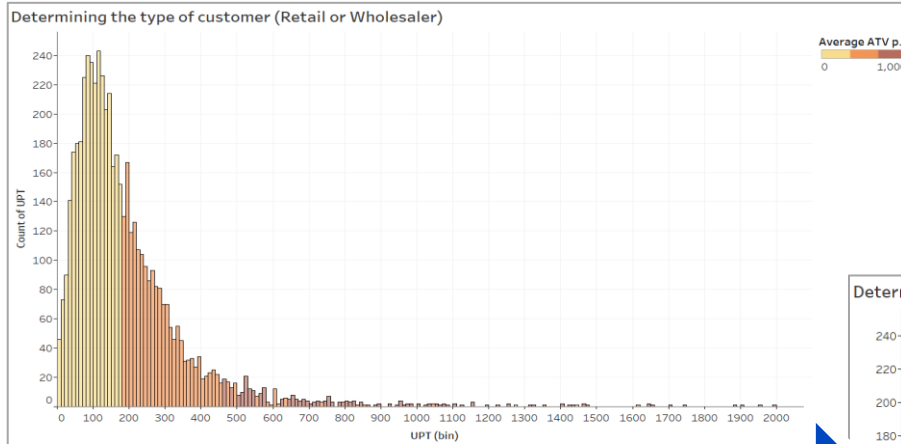


4. Recommendations



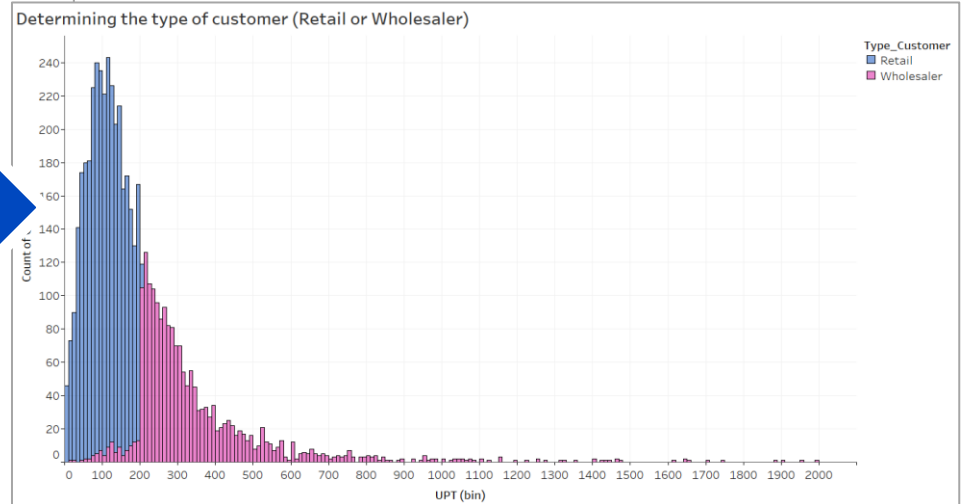
Skewed data denotes the presence of two types of customers

Key factors to estimate both types were: unit per transaction (UPT), ATV and number of transactions



At first glance, differences based on UPT and ATV were observed.

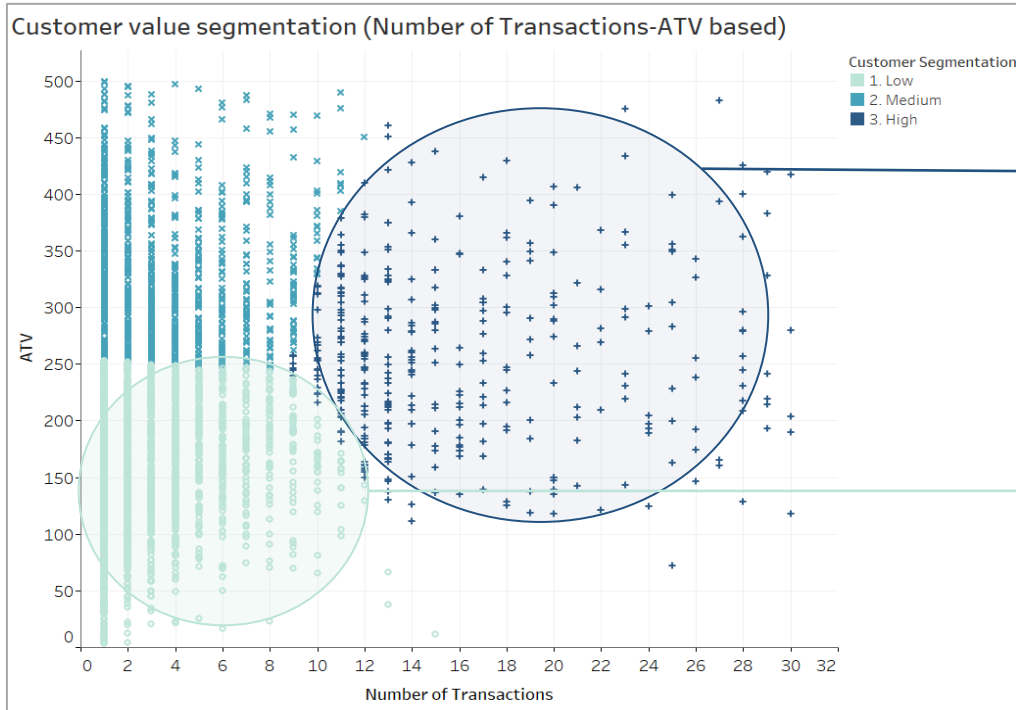
Final customer distribution



- UTP data was skewed, this could be **explained by the mix of retail and wholesale customers** in the same dataset
- Conditions to approximate a retail customer:
 $[UPT] \leq 200$ and $[ATV] \leq 500$ and $[Number\ Invoices] \leq 30$

Retail customers could be clustered based on their ATV and purchase frequency

Each segment can be treated differently through marketing campaigns



High value customers:



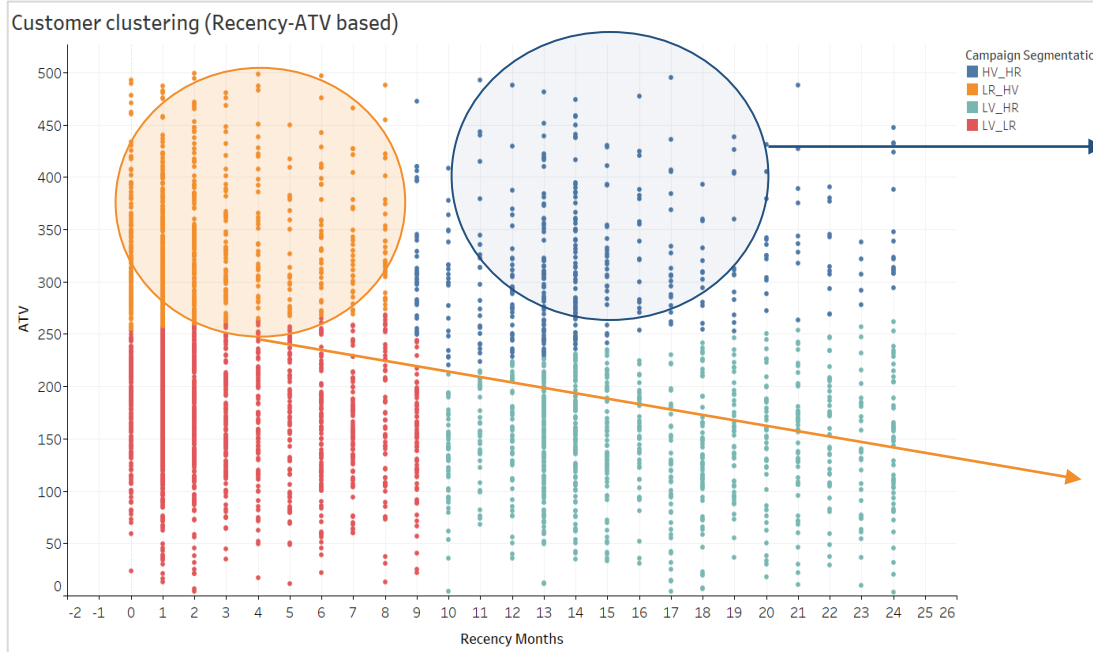
- Reward their loyalty by offering occasional discounts.
- Offer other benefits such as one-day delivery, preferential customer service, etc.

Low value customers:

- Increase number of transactions by offering next purchase discounts.
- Promote upselling at checkout to boost their ATV.

Campaigns to retain current and past high-value customers

Clusters are based on recency and ATR



High value – High recency:



- Win back past customers with special benefits (Next purchase discounts, product-specific offers).
- Investigate what makes them stop buying, to prevent the same behaviour in current customers.

High value – Low recency:

- Retain current customers by improving their satisfaction/NPS, through benefits which can involve shipping or preferential customer service.



1. Data Preparation



2. Sales Performance



3. Customer Analysis



4. Recommendations



What can be improved?

Product profitability

Including cost information would allow performing a more comprehensive analysis

Product classes

Incorporating product categories would significantly improve the interpretation of this data through visualisations

Customer type

Since B2C and B2B sales need to be analysed separately, adding this information to the dataset would facilitate the analysis

Incentivise customer sign-ups

Registrations mean that more customers could be profiled and included in mailing databases. One way to promote this could be through “first purchase” discounts.

