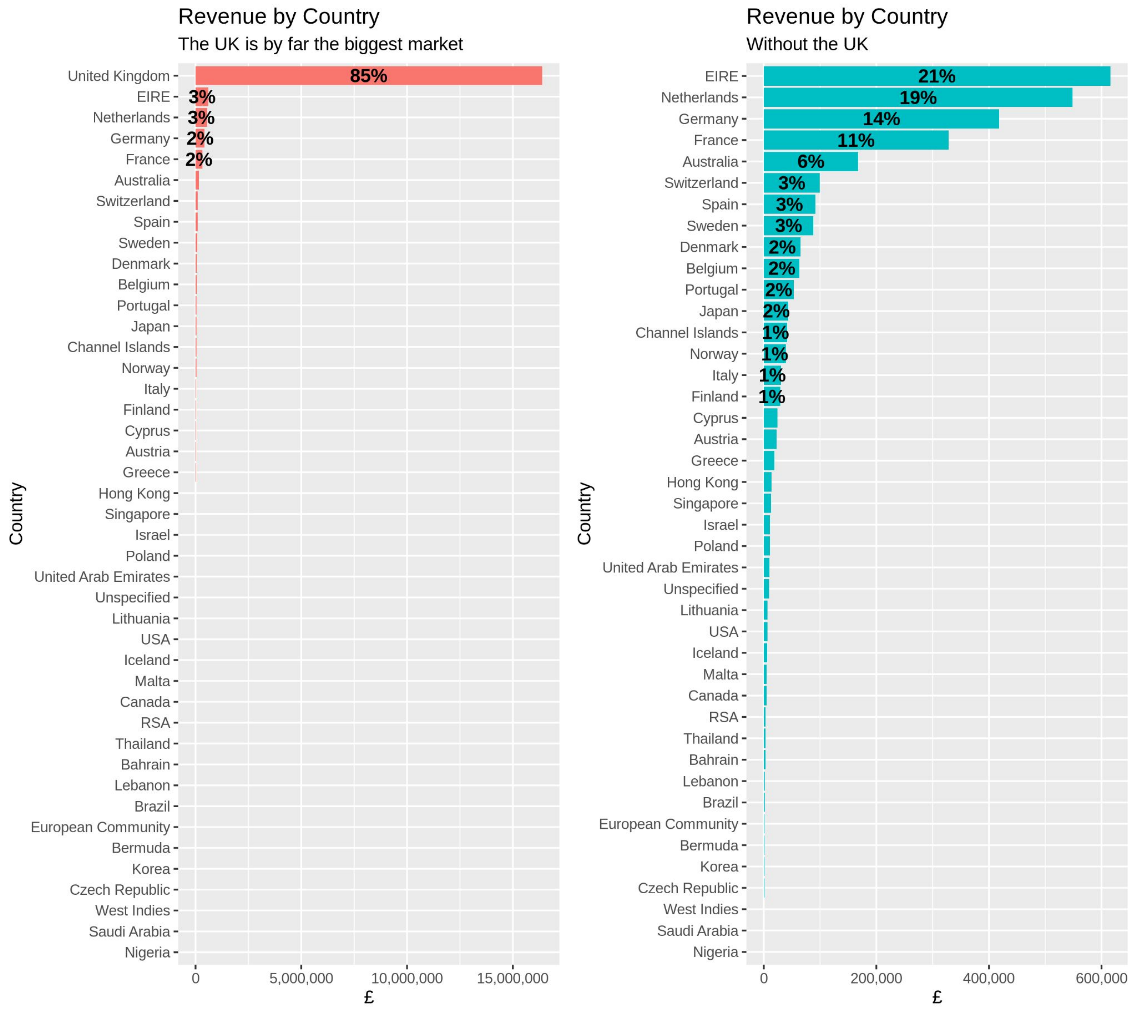
Online Retail II UCI

Waseem Khalifa

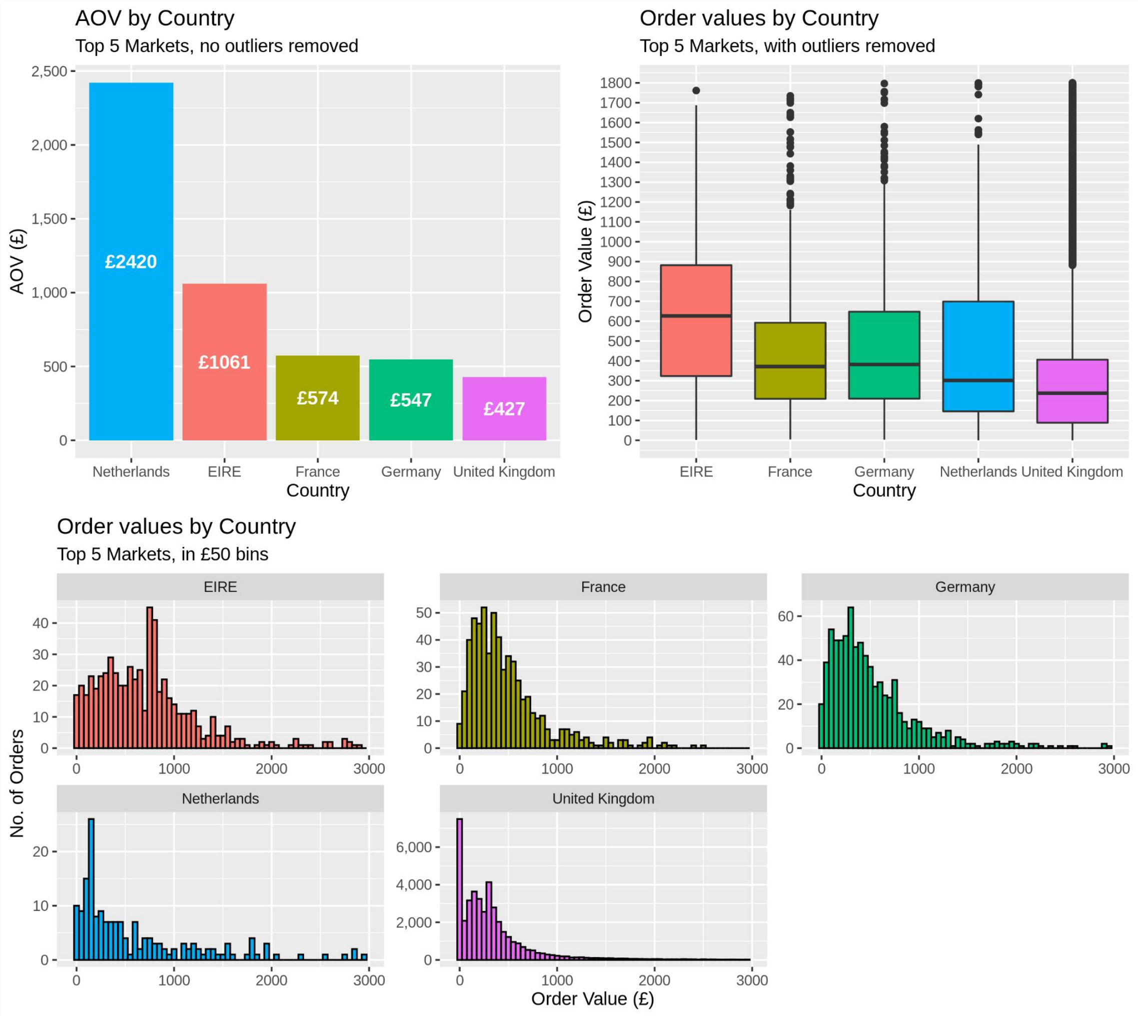
Markets

- UK is the by far the biggest market, holding 85% market share within the data
- The next 4 biggest markets are Eire, Netherlands, Germany and France (all western Europe)



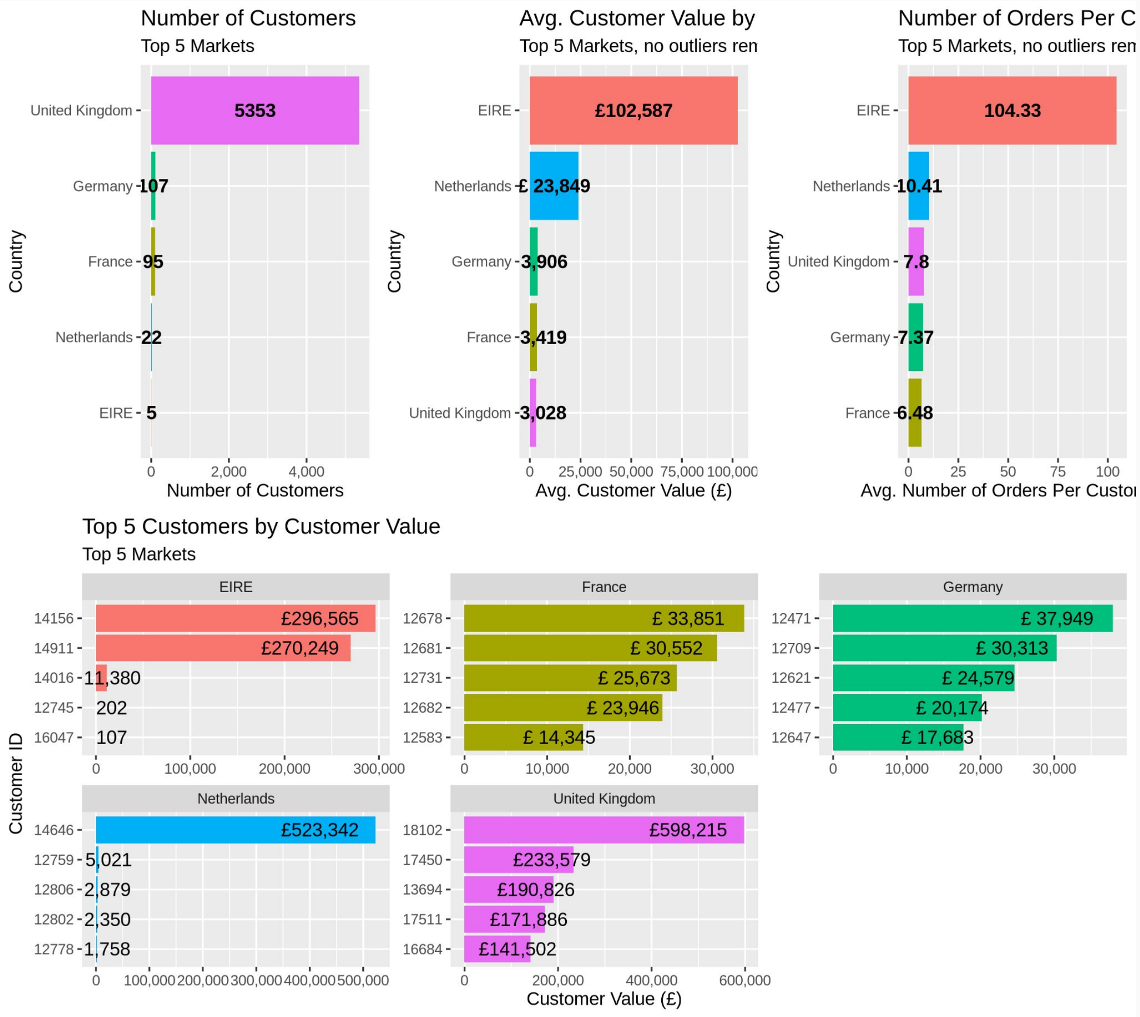
Orders

- Even though UK is the biggest market, it's mainly made up by a vast number of small orders (most orders are around £50)
- Ireland usually has the biggest orders, with most orders close to £1000



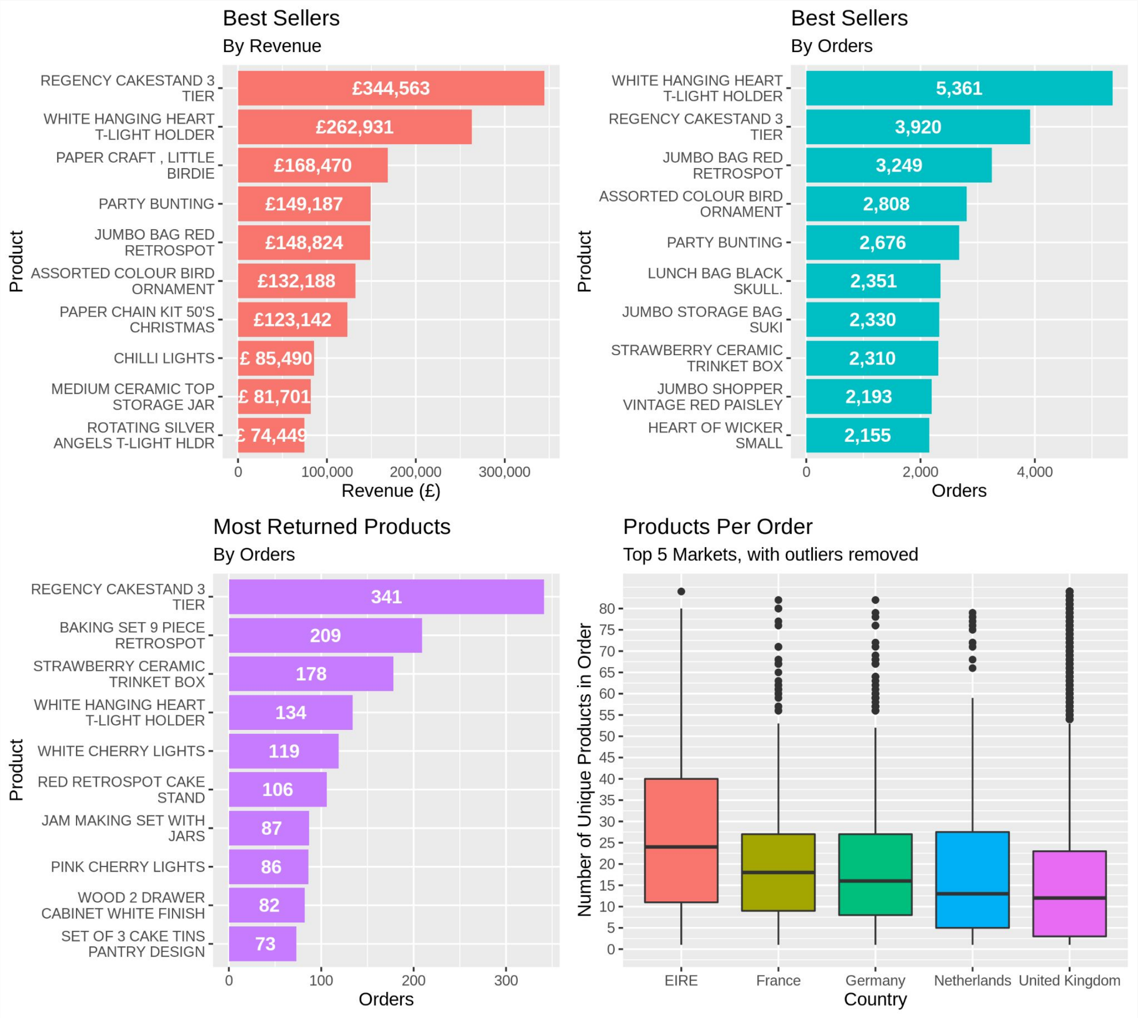
Customers

- Unsurprisingly the UK being the biggest market, has the most customers
- To note, EIRE the second biggest market has only 5 customers in the entire dataset, with most of their sales coming from two customers (who have a higher value than any UK customer bar the top 1)
- Eire also has the highest Avg. Customer Value & the most number of Orders per customer
- The Netherlands sales is dominated by one customer (who nearly rivals the top customer in the UK)



Products

- The Regency Cake Stand 3 Tier is the best selling product (by revenue), but also the most returned product
- The UK typically has the lowest number of products per order, whilst EIRE has the most



Market Basket Analysis Using Apriori Algorithm

I’ve ran a ML algorithm (Apriori) on the dataset, to uncover products frequently bought together. The data uncovered could be used for cross-selling on the website or improve site navigation, amongst other recommendation system tactics. An example of a recommendation system is Netflix’s “people who enjoyed X also enjoyed Y” or Amazon’s “customers who viewed X, also viewed Y”.

Sample output of the Apriori Algorithm:

	lhs		rhs	support	confidence	coverage	lift	count
[1]	{PINK REGENCY TEACUP AND SAUCER}	=>	{GREEN REGENCY TEACUP AND SAUCER}	0.017560341	0.8302961	0.021149492	29.79163	729
[2]	{PINK REGENCY TEACUP AND SAUCER, ROSES REGENCY TEACUP AND SAUCER}	=>	{GREEN REGENCY TEACUP AND SAUCER}	0.014958809	0.9000000	0.016620899	32.29265	621
[3]	{GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER}	=>	{ROSES REGENCY TEACUP AND SAUCER}	0.014958809	0.8518519	0.017560341	29.29890	621
[4]	{POPPY'S PLAYHOUSE BEDROOM}	=>	{POPPY'S PLAYHOUSE KITCHEN}	0.011369658	0.8027211	0.014163897	52.23223	472
[5]	{PINK REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER}	=>	{GREEN REGENCY TEACUP AND SAUCER}	0.009828010	0.8812095	0.011152864	31.61844	408
[6]	{POPPY'S PLAYHOUSE LIVINGROOM}	=>	{POPPY'S PLAYHOUSE KITCHEN}	0.009755745	0.8472803	0.011514188	55.13165	405
[7]	{PINK REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER}	=>	{ROSES REGENCY TEACUP AND SAUCER}	0.009370333	0.8401728	0.011152864	28.89721	389
[8]	{POPPY'S PLAYHOUSE LIVINGROOM}	=>	{POPPY'S PLAYHOUSE BEDROOM}	0.009322156	0.8096234	0.011514188	57.16107	387
[9]	{JUMBO BAG PINK POLKADOT, JUMBO SHOPPER VINTAGE RED PAISLEY}	=>	{JUMBO BAG RED RETROSPOT}	0.008984921	0.8021505	0.011201041	11.35760	373
[10]	{PINK REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER, ROSES REGENCY TEACUP AND SAUCER}	=>	{GREEN REGENCY TEACUP AND SAUCER}	0.008527244	0.9100257	0.009370333	32.65238	354
[11]	{GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER}	=>	{ROSES REGENCY TEACUP AND SAUCER}	0.008527244	0.8676471	0.009828010	29.84217	354
[12]	{CHARLOTTE BAG PINK POLKADOT, CHARLOTTE BAG SUKI DESIGN}	=>	{RED RETROSPOT CHARLOTTE BAG}	0.008334538	0.8046512	0.010357952	31.45413	346
[13]	{POPPY'S PLAYHOUSE BEDROOM, POPPY'S PLAYHOUSE LIVINGROOM}	=>	{POPPY'S PLAYHOUSE KITCHEN}	0.008310449	0.8914729	0.009322156	58.00722	345
[14]	{POPPY'S PLAYHOUSE KITCHEN, POPPY'S PLAYHOUSE LIVINGROOM}	=>	{POPPY'S PLAYHOUSE BEDROOM}	0.008310449	0.8518519	0.009755745	60.14248	345
[15]	{KITCHEN METAL SIGN, TOILET METAL SIGN}	=>	{BATHROOM METAL SIGN}	0.008117743	0.8753247	0.009273980	26.33205	337
[16]	{CHARLOTTE BAG PINK POLKADOT, STRAWBERRY CHARLOTTE BAG}	=>	{RED RETROSPOT CHARLOTTE BAG}	0.007925037	0.8680739	0.009129450	33.93335	329
[17]	{JUMBO BAG PINK POLKADOT, JUMBO BAG STRAWBERRY}	=>	{JUMBO BAG RED RETROSPOT}	0.007828684	0.8044554	0.009731657	11.39023	325
[18]	{CHARLOTTE BAG PINK POLKADOT, WOODLAND CHARLOTTE BAG}	=>	{RED RETROSPOT CHARLOTTE BAG}	0.007443272	0.8328841	0.008936744	32.55777	309
[19]	{POPPY'S PLAYHOUSE BATHROOM}	=>	{POPPY'S PLAYHOUSE KITCHEN}	0.007395096	0.8950437	0.008262273	58.23957	307
[20]	{POPPY'S PLAYHOUSE BATHROOM}	=>	{POPPY'S PLAYHOUSE LIVINGROOM}	0.007274654	0.8804665	0.008262273	76.46796	302

From our sample output above, looking at number 15, we find that when customer have ‘Kitchen Metal Sign’ & ‘Toilet Metal Sign’ in their basket, they also buy ‘Bathroom Metal Sign’.

Customer Segmentation using RFM & K-Means Clustering

I've segmented the customers in the dataset, to find 'clusters' using the K-Means Algorithm based on R-F-M:

[R] Recency: Time since customer's last transaction (based on max date in the dataset)

[F] Frequency: Total number of transactions

[M] Monetary: Total money spent by the customer (Lifetime value)

We can use the output to market or 'talk' to each customer segment differently to increase customer satisfaction which in turn increases sales! We can also run specific marketing campaigns for each customer segment.

The 4 Clusters the K-Means Algorithm found:

Cluster	Cluster Name	No. of Customers	% of Total Customers	Avg. Recency	Avg. Frequency	Avg. Monetary
1	Occasional Buyers	3689	62.9%	72	6	£2,187
2	Superstars	11	0.2%	3	190	£242,169
3	Churned	1935	33.0%	468	2	£725
4	Great Customers	230	3.9%	27	40	£23,128

Our Clusters:

1. **Occasional Buyers**: These make up most of our customers, they don't buy as often as some of the other customers, but due their volume, bring in a sizable amount of revenue. We need to ensure these customers don't churn, we could potentially due this with a loyalty programme (discounts after a certain number of orders within a timeframe - this would increase frequency, recency and monetary).
2. **Superstars**: These are our superstar customers, they spend the most and buy very frequently. Our aim is to always keep them happy and ensure they don't leave for a competitor.
3. **Churned**: These are customers who haven't shopped with us for a long time, and when they did, they didn't buy as often as the other clusters. Look to run a reactivation/welcome back campaign, to see if some of these customers return and potentially move into the other clusters, as they make up the second largest cluster (33% of customers)
4. **Great Customers**: Our second best customers, behind the Superstars. Look to see if there is potential for some of them to become 'Superstars'.

The End