



ONLINE RETAILS

Data Analysis Report

BUSI4389 Data at Scale: Management Processing and Visualization
2019 - 2020

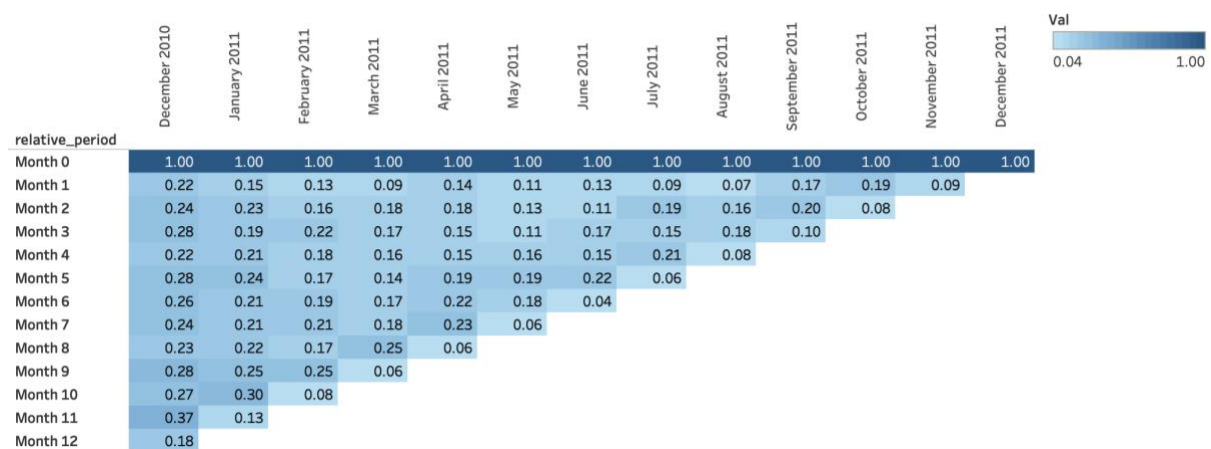
Shiqi BAI (20219140)

Section 1: Cohort Analysis

Data cleaning summary

- Delete the rows with missing values (Description, CustomerId)
- Delete the rows where Quantity with negative values
- Add the column named Revenue (= UnitPrice*Quantity)
- Delete the rows where UnitPrice equals to 0
- Delete the rows where the description is ineffective for analysis (e.g. POSTAGE)

Cohort Analysis



In this section we are going to build time-based cohorts that will assign the acquisition month to each customer i.e. the month when each customer made their first paid transaction with this company. We divide customers who purchase popular items each month into cohorts for a cohort analysis. From the analysis, we can see that the purchase of popular products is on the decline, but the overall sales in the fall are better. This shows that the customer groups of this e-commerce have high-frequency purchases, also the loyalty of the old customers who are retained is very high.

Section 2: The KPIs

KPI Description: Top 5 customers with most number of orders and highest money spent
KPI formula:
Steps to realize KPI: <pre>SELECT customer_id, country, COUNT(invoice_no) AS invoice_num FROM shopping_retails_clean GROUP BY customer_id, country ORDER BY invoice_num DESC LIMIT 5;</pre> <pre>SELECT customer_id, country, SUM(revenue) AS amount_spent FROM shopping_retails_clean GROUP BY customer_id, country ORDER BY amount_spent DESC LIMIT 5;</pre>
Additional Notes:

KPI Description: Top 5 countries with most number of orders and highest money spent
KPI formula:
Steps to realize KPI: <pre>SELECT country, COUNT(invoice_no) AS invoice_num FROM shopping_retails_clean GROUP BY country ORDER BY invoice_num DESC LIMIT 5;</pre> <pre>SELECT country, SUM(revenue) AS amount_spent FROM shopping_retails_clean GROUP BY country ORDER BY amount_spent DESC LIMIT 5;</pre>
Additional Notes:

KPI Description: Top 10 items with most number of orders
KPI formula:
Steps to realize KPI: <pre>SELECT stock_code, description, COUNT(invoice)no) AS invoice_num FROM shopping_retails_clean GROUP BY stock_code, description ORDER BY invoice_num DESC LIMIT 10;</pre>
Additional Notes:

KPI Description: Customer retention for popular items
KPI formula:
Steps to realize KPI: <pre>CREATE TABLE cohort_analysis AS WITH cohort_assignment AS (SELECT customer_id, DATE_TRUNC('month', MIN(invoice_date))::DATE as cohort_date FROM pop_items GROUP BY customer_id),</pre>

```

cohort_mth_cts AS (
    SELECT cohort_date,
           EXTRACT(month FROM AGE(invoice_date, cohort_date))
           + 12*EXTRACT(year FROM AGE(invoice_date, cohort_date)) AS
relative_period,
           COUNT(DISTINCT customer_id) AS active_ct
FROM pop_items
JOIN cohort_assignment
USING (customer_id)
GROUP BY cohort_date, relative_period
),
cohort_totals AS (
    SELECT cohort_date, COUNT(DISTINCT customer_id) AS
cohort_total
FROM cohort_assignment
GROUP BY cohort_date
),
cohort_mth_percent AS (
    SELECT cohort_date, relative_period, active_ct::NUMERIC
/cohort_total AS active_percent
FROM cohort_mth_cts
JOIN cohort_totals
USING (cohort_date)
)
SELECT cohort_date AS row_id, relative_period::TEXT AS col_id, active_percent AS val
FROM cohort_mth_percent
UNION ALL
SELECT cohort_date AS row_id, 'total'::TEXT AS col_id, cohort_total AS val
FROM cohort_totals;

```

Additional Notes: popular items -> top 10 items with most number of orders

Section 3: Executive Summary

This report was commissioned to examine the online retailers situation and their customers behaviour from December 2010 to December 2011. The analysis was explored in 4 dimensions: orders, items, countries and customers, along with the recommendation of the target locations to increasing the marketing spend.

The research draws attention to the sales status of the online retailers at the first. For the orders: in the observation after effective data cleaning, a total of 15,091 valid orders were generated during the transactional period, with a mean revenue of 21.17 pounds and a mean of product quantity was about 12. It indicated that most of the orders were wholesales. In addition, the average values of revenue and product quantity were both higher than the middle values, of which the mean revenue was even higher than 75% of the whole. It could be demonstrated that there were some customers with powerful purchasing ability among all the orders. As for the items sold, we discovered that the e-commerce position is mainly in the commodity market with low prices. The products less than 4 pounds were more popular than other items with high price, the low-price products brought the most revenue and sales. Finally in country level, it inferred that the company was based on UK since the sales was significant than other locations. The main sources of overseas income are mostly countries around the UK, which basically conforms to the situation where the UK is the centre of radiation. This phenomenon may be related to transportation costs and language, or it may be that the influence decreases with distance.

This retail company has excellent sales performance in the UK market, and the current challenge is how to increase sales in other surrounding areas. Fortunately, such as Netherlands, Australia and Ireland in the surrounding areas all have good customer groups, so we are targeting to increase the purchase and repurchase of customers in such areas, and the cultivation and maintenance of loyal customers.

It is recommended:

- Since May 2011, the transaction amount has generally increased and reached a peak in November. It is speculated that it may be caused by the increase in the demand for gifts or promotional activities after the mid-year festival. The power of promotion should be enhanced in the normal days.
- The United Kingdom is the main source area for customers, and Germany, France, the Netherlands, Ireland and Australia have good transaction values. You can try to expand the markets of these countries.
- The proportion of active customers and returning customers increases with time, and the loyalty of old customers in the later period is good, so the repurchase rate and repurchase rate perform well; customer value can be classified through machine

learning classification models, and different value customer take different services decision making.

Section 4: Comparative Analysis

By analysing sales data to understand the consumption of online retail business, analyse customer consumption data and analyse customer consumption behaviour.

Store consumption performance

- Monthly sales amount:

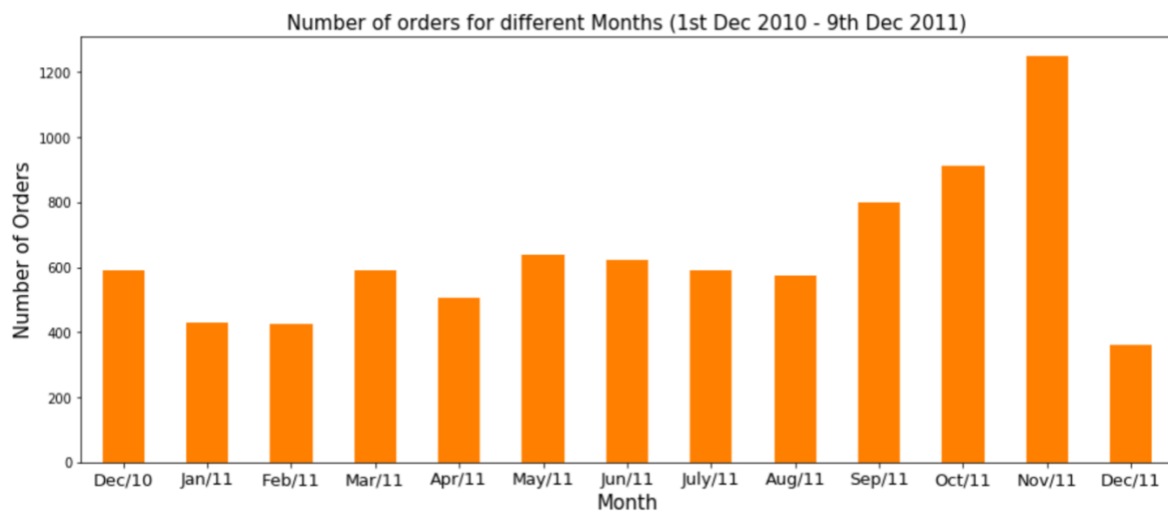


Figure 1: Number of orders for different months

- From December 2010 to November 2011, the monthly turnover generally showed an upward trend, and it began to rise from August 2011. It may be because the company's main products include gifts, and there are various festivals abroad in the second half of the year leading to a surge in demand for gifts. It is also possible that there are various promotional activities near the end of the year.
- Consumption country distribution:

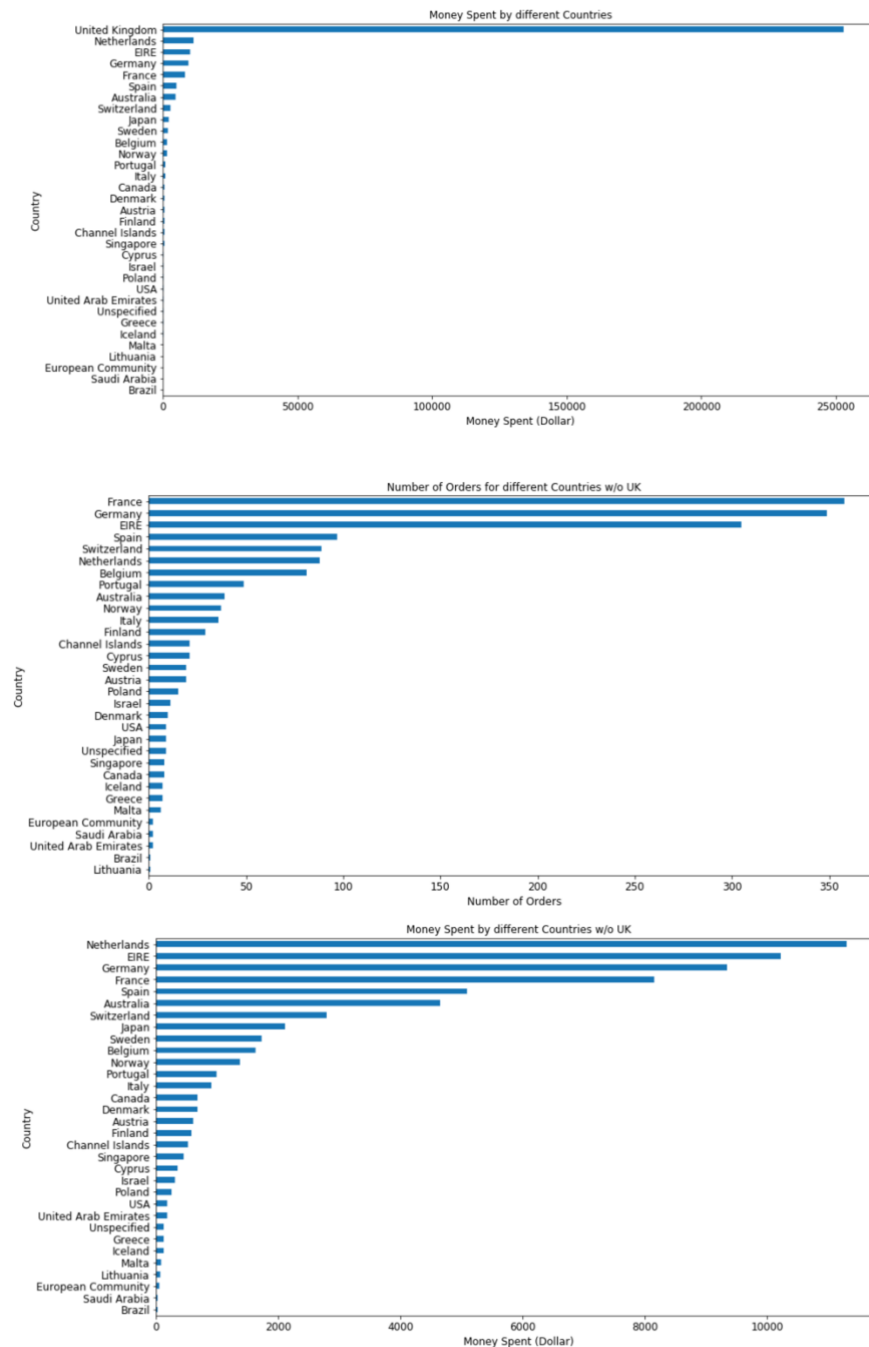


Figure 2: Money spent and number of orders among the countries

- Because the company is based in the UK, it can see that the number of customers, orders and transaction values in the UK are far more than in other countries. The UK is the main country of customer origin and is an important value country. Except for the United Kingdom, Germany and France are among the top 5 in terms of number of customers, number of orders, and transaction value, and are among the key maintaining countries.
- Although the number of customers and orders is small in the Netherlands, the transaction value is in the top 6. This country has higher customer unit prices and belong to key developing countries. Countries that focus on maintaining

and developing countries can carry out appropriate promotion, improve logistics services, and so on.

Customer consumption behaviour

- Amount of customer consumption, number of products purchased by customers:
 - o More than half of the customers have spent at least two times, and even some customers have spent more than 100 times, which are very loyal customers. More than 75% of customers purchase between 10 and 50 times.
 - o The highest customer consumption cost is 3,621 pounds, the average customer consumption amount is greater than 12, and there are some high-volume consumer users.

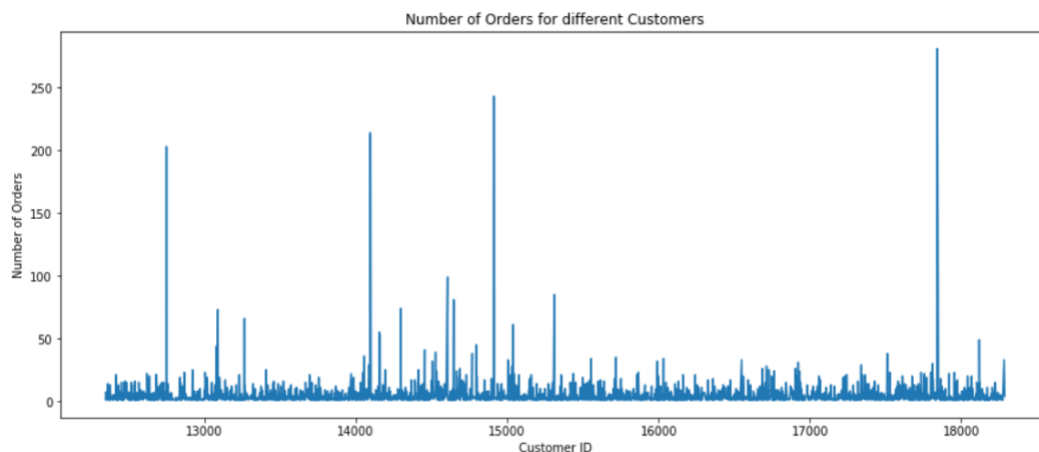


Figure 3: Number of orders for different customers

- Customer purchase cycle:
 - o There is a weak correlation between the number of consumptions and the amount of consumption. The larger the number of consumptions by the customer, the larger the amount of consumption, which can lead the customer to consume multiple times.
 - o Taking a look at the customer's consumption gap. For customers who spent at least twice, half of the consumption interval is within one month, and 75% of the consumption interval is within two months.
 - o Most customers have spent at least two times. The more customers consume, the larger the consumption amount can be in three periods of half a month, one month, and two months. Recall customers and guide customers to make multiple purchases.

Popular items performance

- Top 10 order items and Top 10 reorder items:

Description	Reorder	Description	Reorder
Small Red Retrospot Windmill	187	Spotty Bunting	49
White Hanging Heart T-light Holder	85	Jumbo Bag Red Retrospot	47
Regency Cakestand 3 Tier	70	Popcorn Holder	44
Lunch Bag Red Retrospot	59	Wooden Frame Antique White	44
Party Bunting	57	Assorted Colour Bird Ornament	43

Table 1: Top 10 order items

Description	Reorder	Description	Reorder
Small Red Retrospot Windmill	109	Number Tile Cottage Garden 1	25
Edwardian Parasol Natural	85	Party Bunting	20
Five Cats Hanging Decoration	54	Lunch Bag Spaceboy Design	19
Jumbo Bag Red Retrospot	33	Red Toadstool Led Night Light	17
Ninja Rabbit Pink	27	Suki Shoulder Bag	17

Table 2: Top 10 reorder items

- From the table, we can analyse the top 10 sales and reorder products as the most popular product groups. Among them, Small Red Retrospot Windmill, Jumbo Bag Red Retrospot, Party Bunting and Lunch Bag have excellent performance in both purchase and repurchase, and marketing can consider preparing more corresponding inventory.
- For some products with high repurchase value, such as Edwardian Parasol Natural and Five Cats Hanging Decoration, we should increase publicity efforts to increase customers' first purchase of these products.