

25402226_OnlineRetailTrends

1 Assignment 2:

2 Introduction: Online Sales Analysis (Project 2)

2.1 1) Introduction:

This project seeks to first analyse the general spending patterns and proportions of spending between key customer segments, then drilling down, focusing on the year with the greatest total spend and the customer segment with the greatest net spend on the online retail website, and uncovering trends relating to spending at certain times of day and spending contextualized by discounts redeemed by time of day and top products by hour in the selected market segment and time period.

2.2 2) Data Wrangling and Preparation:

As was done in the first assignment, this project starts the data cleaning process by cleaning the data, removing any cells with special characters in them (such as ? for example). The need for this was found upon doing an initial inspection of the dataset in excel, before I transformed it into a csv file better efficiency on my laptop. Invoice Date is also reformatted using a lubridate package to be easier to parse for future data-frame iterations. All NA, “ ” and 0 values in the data are also removed from the initial data-frame (the 0 values from unit price). Manual Items are also removed from the data set, while a total spend column is added to the dataset so values of transactions can more clearly be seen. The data is prepared for the first visualisation by grouping the data by country, year and total spend in order to visualize the trends in net spending across the 5 largest customer segments over years.

The next data preparation step sum's the net spend of all the years across the customer segments and then selects the year with the highest value and uses this as the reference year thereafter for the creation of a tree map visualation showing the percentage of total spend per country in the year with the highest total spending, whereafter the UK is identified as the majority market on this platform.

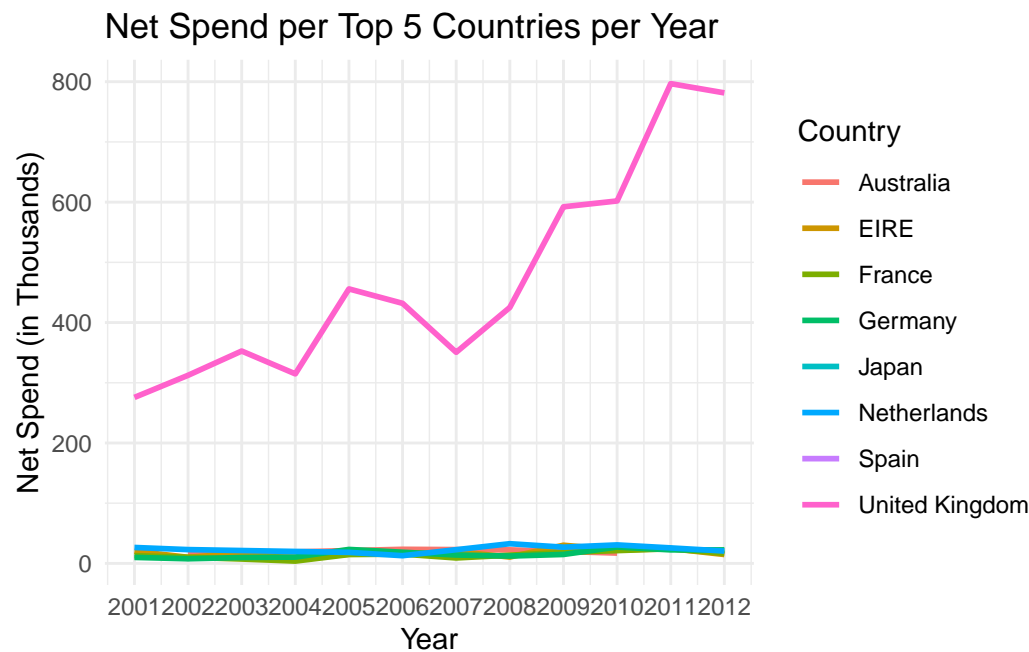
The next data wrangling and data preparation steps set time as a factor along conventional 12am - 11pm time conventions, rounds the time values up or down and then orders the time bins so a heatmap can be made to indicate the times in the day where the largest amount is spent in the United Kingdom market segment in 2011. The data is further filtered looking at the heatmap and selecting the 3 most prolific times for product purchases as 5am, 7am, and 8pm. A tibble is then created to show the top 10 products with the highest net spend on them and their time of purchase.

The data is then prepared for the final visualisations by grouping the data by unique product description and selects the top 3 best selling products across hours of the day, with the specific market segment

and year being filtered for explicitly too. Lastly, a discount df is created to to compare discount times and values they can be compared to spending by hour in the heatmap visualisation.

2.3 3 & 4) Exploratory Data Analysis (EDA) and Reporting :

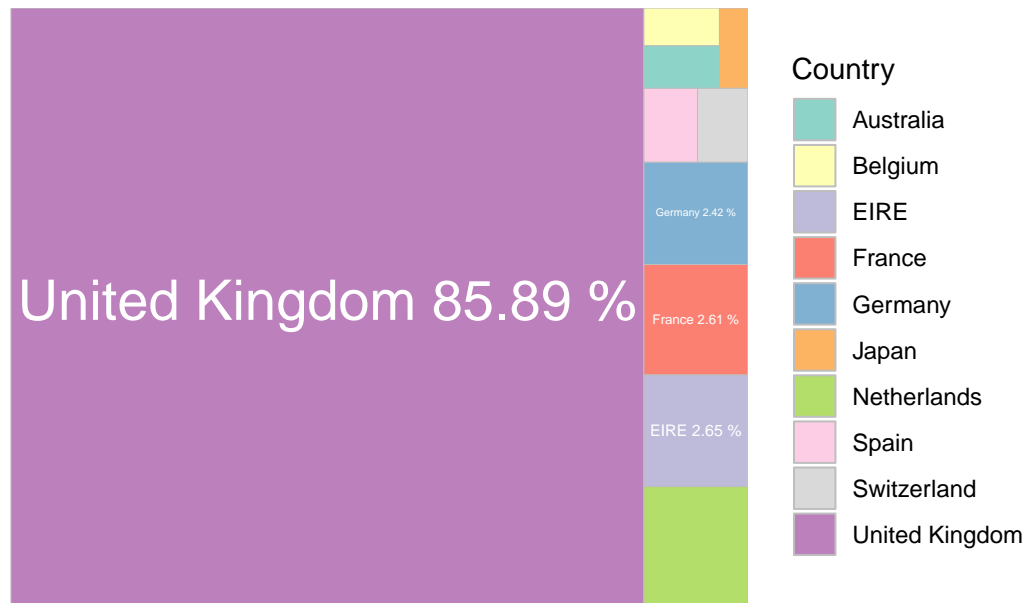
2.3.0.1 Graph 1: Net Spend per Top 5 Countries



2.3.0.2 Graph 2: Percentage of Total Spend by Country in 2011

Research Question 1: What is the general distribution of net spending amongst the largest market segments ?

Percentage of Total Spend by Country in 2011

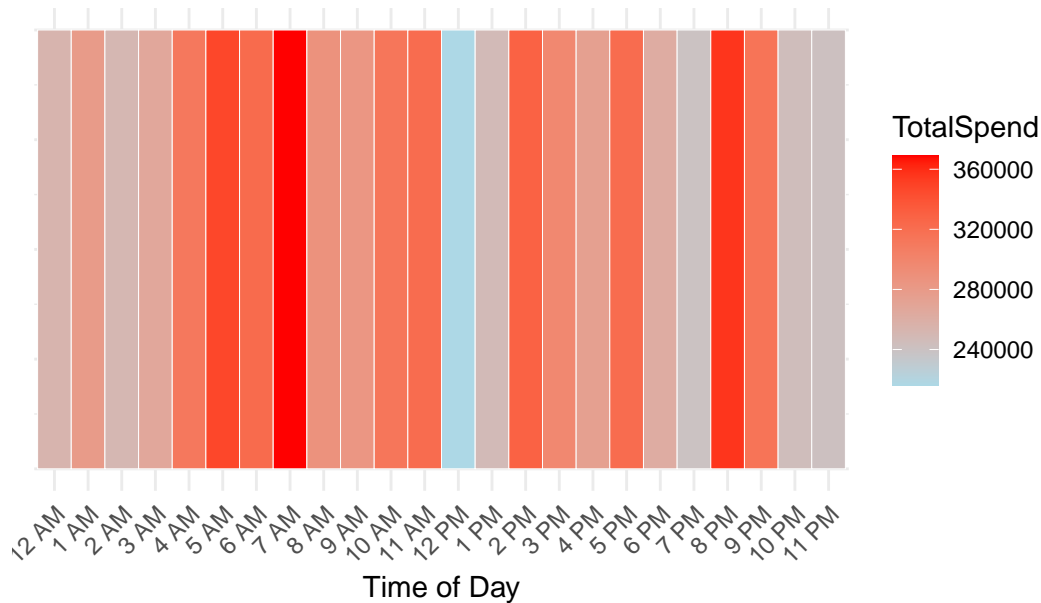


The visualisation indicates that the United Kingdom has a significant majority of total spend on this particular online retail website. In fact, the UK customer segment has a more than 4 times greater proportion of total spending in this time period than all other countries combined, while having a much smaller population comparatively. This suggests that the online retailer is likely based in or targeted towards the UK customer segment, or indicates a much higher population involvement in online shopping relative to population size when compared to other nations. Due to this overwhelming majority, the UK is selected as the key market segment for analysis as it constitutes the most significant data points which are being investigated.

2.3.0.3 Graph 3: Heatmap of Spending by hour of the day

Research Question 2: How does spending change throughout the day and can any “hotspots” be observed ? What does this suggest about those making the purchasing?

Total Spend by Time of Day in 2011 (UK)



The heatmap provides a useful visualization of spending trends through times of the day. What can be cleaned from this visualisation, is a much larger amount of money is spent online during specific hours of the day (5am, 7am and 8pm). This spending trend suggests that online shopping occurs mostly in the morning and again at around 8pm, with scant or relatively small spending at other times of the day. This insight could prove usefl for advertisers, as advertising at certain times of the day could allow them to access a customer base who is more likely to purchase their product during certain times of the day. The timing of the spending trends also suggests that online shopping occurs largely before standard working hours and could suggest online retailers as being a first stop for purchasing products (given how early these times are, it is unlikely that the online shopping was done after checking physical retail shops first).

A tibble: 10 x 4

	Time	Description	TotalQuantity	TotalSpend
	<ord>	<chr>	<int>	<dbl>
1	7 AM	VINTAGE UNION JACK MEMOBOARD	2028	12981.
2	8 PM	SET OF TEA COFFEE SUGAR TINS PANTRY	1522	7632.
3	7 AM	CREAM HEART CARD HOLDER	3137	7412.
4	7 AM	WOOD BLACK BOARD ANT WHITE FINISH	1416	7080.
5	7 AM	REGENCY CAKESTAND 3 TIER	527	6023.
6	5 AM	REGENCY CAKESTAND 3 TIER	504	5771.
7	7 AM	JUMBO BAG RED RETROSPOT	3012	5383.
8	8 PM	PARTY BUNTING	1137	4959.
9	8 PM	REGENCY CAKESTAND 3 TIER	427	4955.
10	8 PM	PANTRY CHOPPING BOARD	943	4788.

This tibble provides a glimpse of the top 10 products, their rounded time of purchase, quantity and the total amount spent on it. This provides tabular proof of the vastly larger spending trends in the early morning and later evening than compared ot other times of the day (corroborating the fidnings of the heatmap analysis).

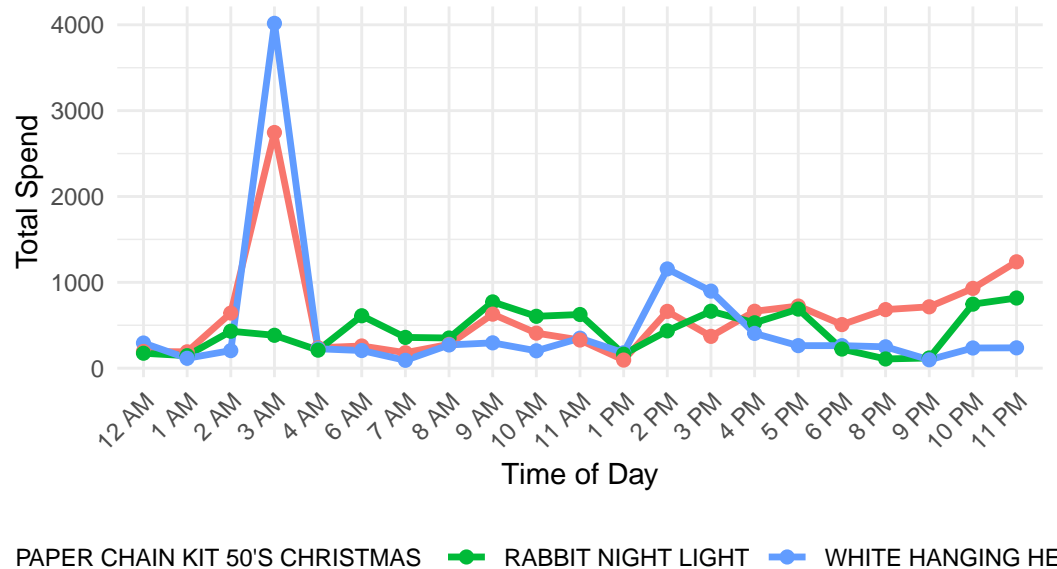
A tibble: 10 x 4

	Country	Description	TotalQuantity	TotalSpend
	<chr>	<chr>	<int>	<dbl>
1	United Kingdom	"REGENCY CAKESTAND 3 TIER"	7520	86600.
2	United Kingdom	"WHITE HANGING HEART T-LIGHT HOLDER"	26921	73659.
3	United Kingdom	"JUMBO BAG RED RETROSPOT"	35534	65413.
4	United Kingdom	"PARTY BUNTING"	11747	52838.
5	United Kingdom	"ASSORTED COLOUR BIRD ORNAMENT"	28281	45151.
6	United Kingdom	"PICNIC BASKET WICKER 60 PIECES"	61	39620.
7	United Kingdom	"CHILLI LIGHTS"	7709	36724.
8	United Kingdom	"BLACK RECORD COVER FRAME"	9763	33381.
9	United Kingdom	"PAPER CHAIN KIT 50'S CHRISTMAS "	12062	33004.
10	United Kingdom	"WOOD BLACK BOARD ANT WHITE FINISH"	4822	27888.

2.3.0.4 Graph 4: Total Spend of Top 3 Products by Time of Day (2011, UK), by Time of Day

Research Question 3: What is the distribution of spending based on the top 3 best selling/ highest sale value products according to time of the day?

Total Spend of Top 3 Products by Time of Day (2011,



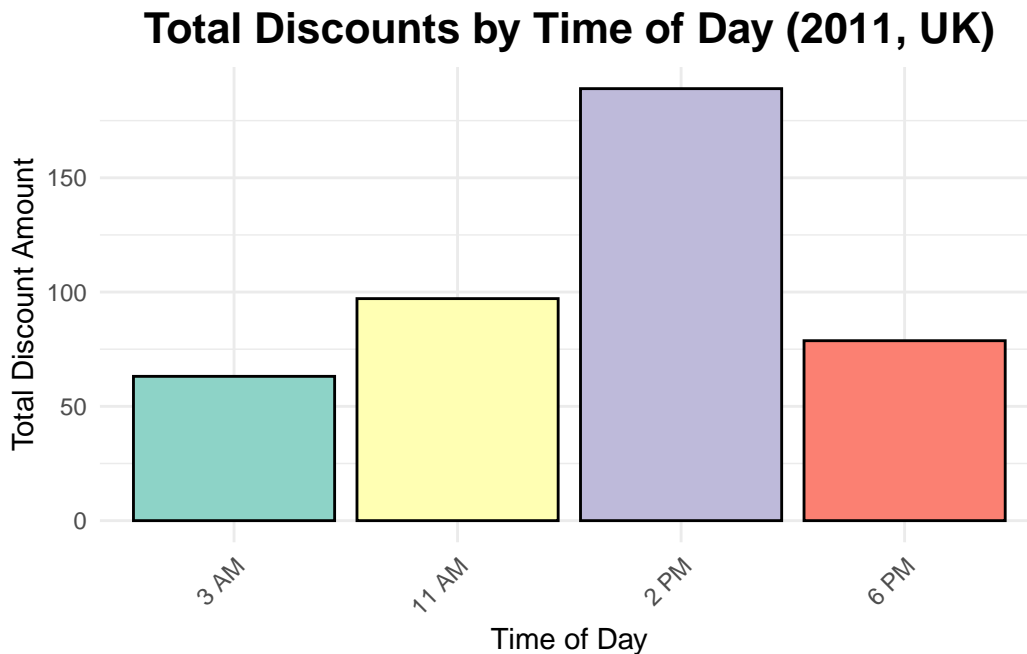
This visualisation further shows the massive spike in online purchasing at certain times of the day, particularly in the early morning. In fact, this graph shows that the period from 2am - 4am as being the time period in which the largest amount is spent on 2 of the 3 top 3 products in the UK by total sale value. This purchasing trend suggests a degree of impulsivity to these purchases, as these times are outside the normal operating hours for most individuals and indicates how the ease of purchasing access at all times continues to be a useful factor for online retailers in maintaining high levels of sales throughout the day.

2.3.0.5 Graph 5: Total Discounts by Time of Day (2011, UK)

Research Question 4 and 3 (extended): After reviewing previous visualizations, what times of the day are the greatest value of discounts redeemed and what does this suggest about online retailers marketing strategy?

A tibble: 4 x 3

	Time	TotalSpend	TotalDiscount
	<ord>	<dbl>	<dbl>
1	3 AM	267177.	63.1
2	11 AM	321230.	97.1
3	2 PM	328994.	189
4	6 PM	262281.	78.7



Upon analysis of the graph above, its clear that online retailers offer discounts at the times of day when they have the lowest average total spend. This suggests that retailers are aiming to both get more stable sales throughout the day by incentivising users to buy products at more varying hours of the day, while also minimizing the cost to company discounts play as they are largely redeemed only during the least active hours of the day, when contextualized by the previous 2 visualisations.

2.4 5) Conclusion of Reporting:

From detailed analysis it is clear that buyers in the UK market segment (filtered into only the niggst net spend year, 2011) have clear spending trends based on the time of day, with the largest unique product sales occuring at very early hours of the morning, suggesting high levels of impulsivity in this market segment, and with the generalised spending trends grouping around early morning and middle evening. All of this is contextualized by online retailers discount timing and the values of discounts redeemed which are the polar opposite of the spending trends, occurring in the hours of the day with the least generalised toal spending and product specific spending. As such, it is evident that online

retailers are attempting to leverage recorded user spending trends to expand and stabilize their most profitable business hours, while taking advantage of users impulsivity and offering the fewest discounts when the users are most likely to be purchasing products of the store.