Online Retail Sales Insights

Group members

- Zachary Breen
- Samra Vatan Parast
- Miley Kotamee
- Nicholas Chua

Table of contents

Project overview

We are looking at sales from a UK based non-store online retailer that sells unique all occasion gifts mainly to wholesalers and determining whether time affects purchasing behaviour across 42 countries.

Reasoning behind this project

Analysing sales data will allow us to better understand the customers' behaviour, the products they are buying and the reasoning behind this behaviour.

In doing so, we are able to determine the weakest sales period and highlight the most profitable target markets.

The result of this analysis will help us to implement targeted marketing activities to improve sales and take hold of sales opportunities during those weaker periods.

Our Data Source

URL: https://archive.ics.uci.edu/ml/datasets/online+retail

Date retrieved: Tuesday 14th September 2021

Data format: XLSX (converted to CSV)

Author: Dr Daqing Chen, Director: Public Analytics group.

Organisation: School of Engineering, London South Bank University, London SE1 0AA, UK.

How was the data retrieved?: Downloaded the CSV file

Data period: From 2009 to 2011

Data points: Over 1 million

Our hypothesis

- Customers are more likely to shop online during a certain time of the day or week of the day.
- Customers are more likely to shop in certain months of the year (i.e. Christmas season or end of financial year) and purchases will increase during these seasons.
- Customers are more likely to buy in higher quantities at lower prices, thus, the lower price items are more popular.
- Customers are more likely to buy from local companies due to having better access (In this case, we expect the UK to have the most purchases)

Data wrangling process

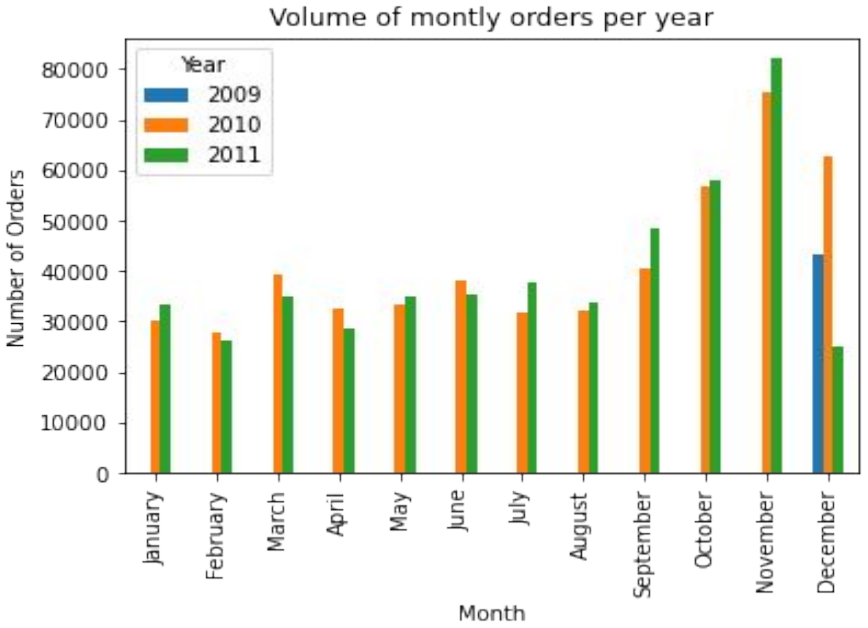
- Separated two tabs from Excel and saving it as two separate CSV files
- Merged two CSV files
- Removing any data points that have:
  - Quantity less than zero
  - Price less than zero
  - Removing unspecified countries
- Rename column data
- Renamed column names
- Changed data type to appropriate type
- Split ‘Date and time’ into separate columns

Sales Growth Analysis

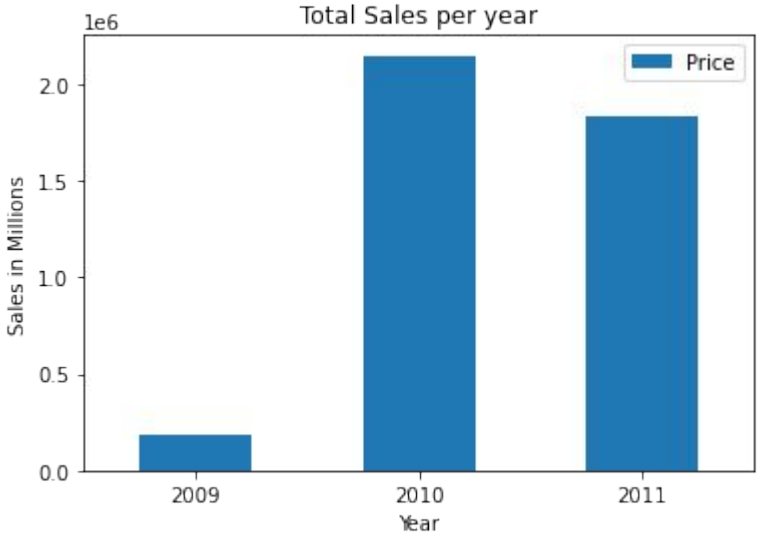
Volume of monthly orders per year

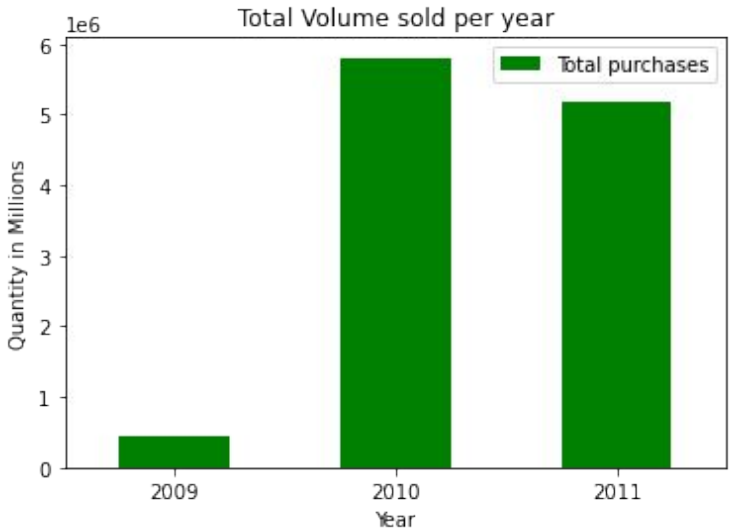
- As shown in the graph, we have noticeable movements between monthly orders for each year in our data.

- Fluctuations in the the number of orders could be due to variety of factors such as successful marketing campaigns (i.e. increase in marketing activities), expanding the market, improving customer service, etc.

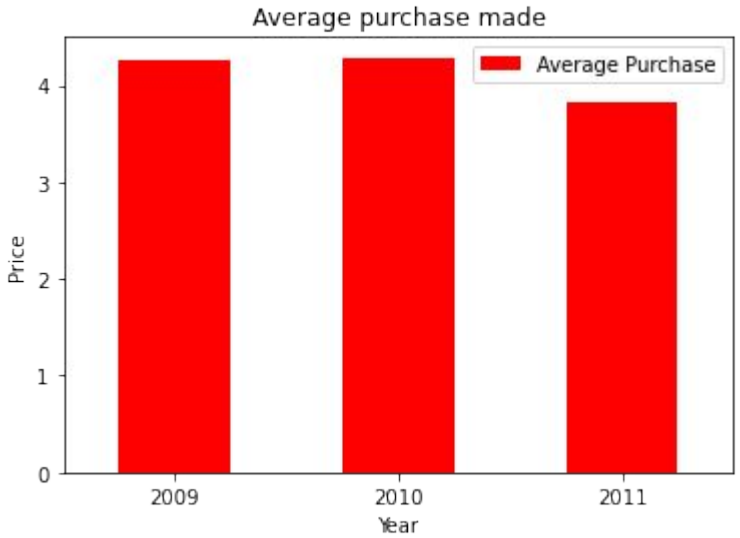


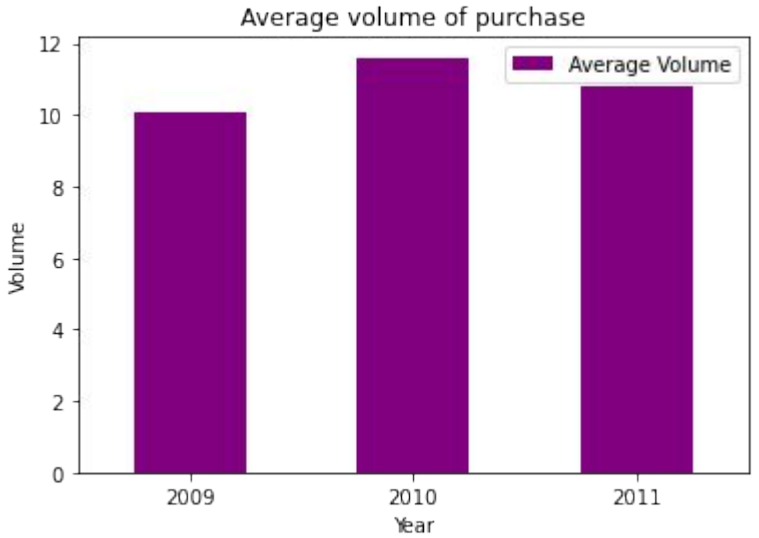
Annual sale increase/decrease between 3 year of the data





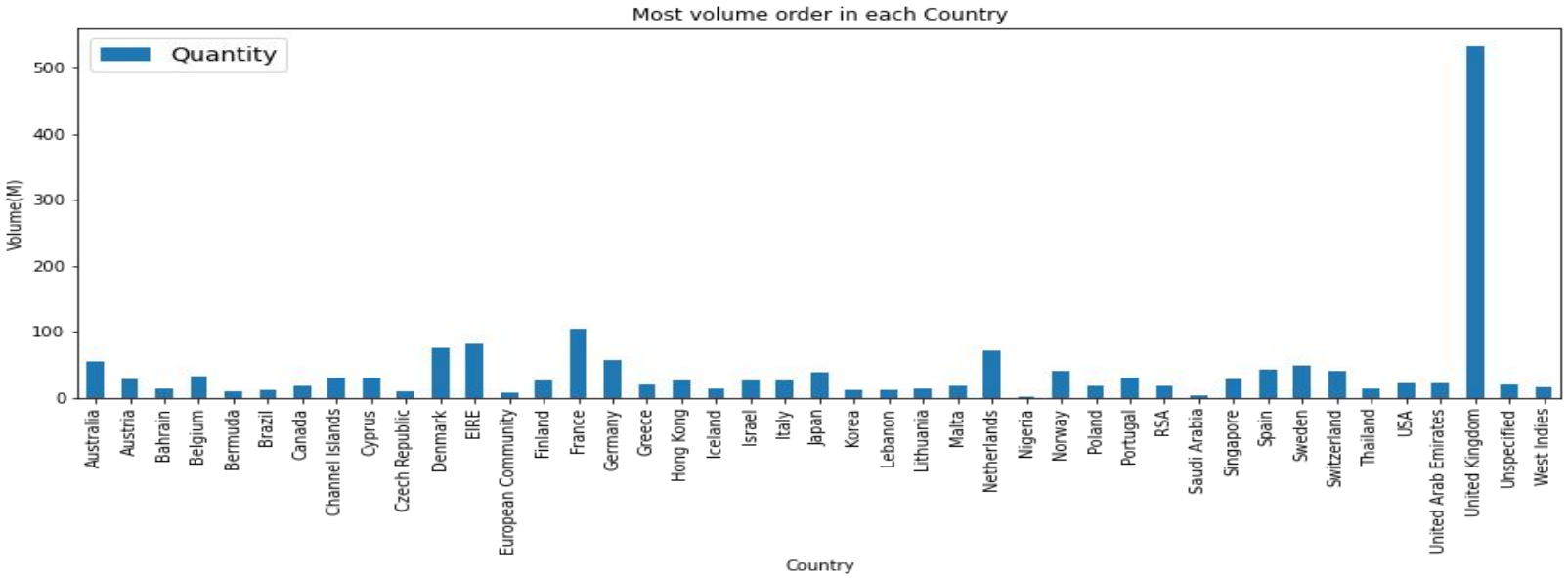
Total average volume increase/decrease between 3 year of the data





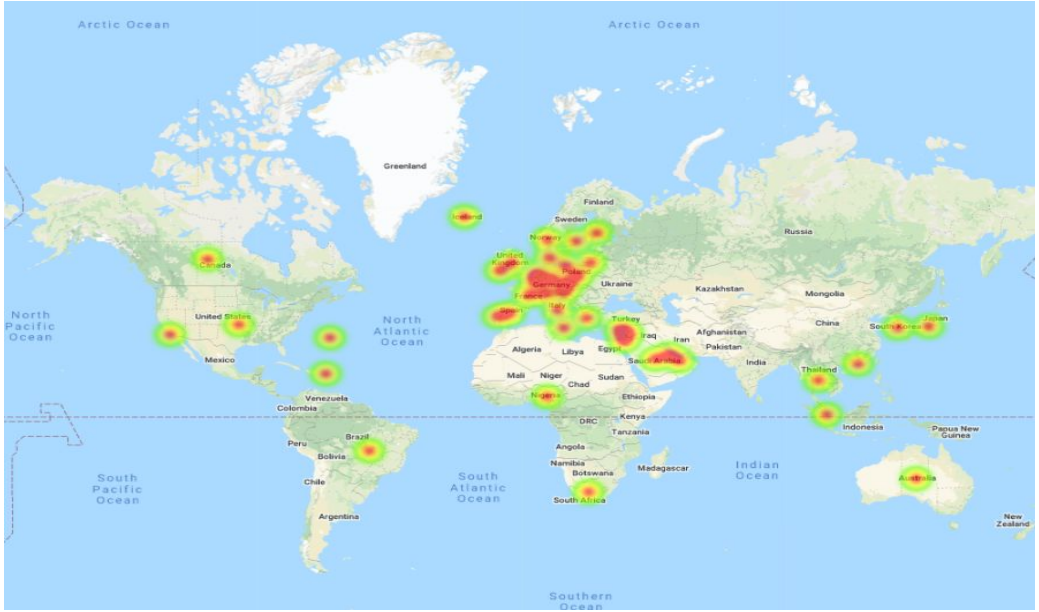
Market Research

Most volume orders between countries



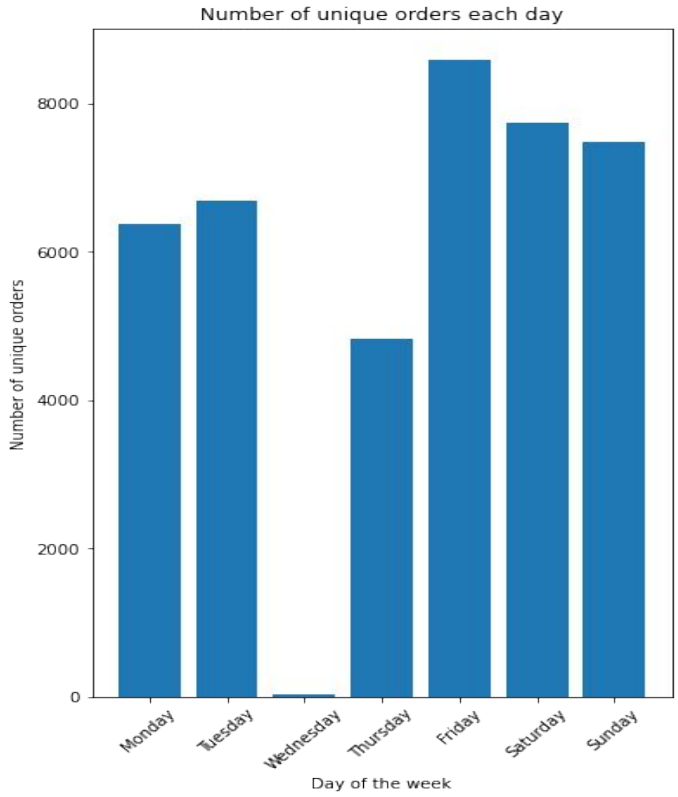
Heatmap volume of sale in each country

- The heatmap present the sale correlation in each country.
- The heavily sale is more likely in Europe
- And some present in Asia including Middle eastern , Australia and North America.



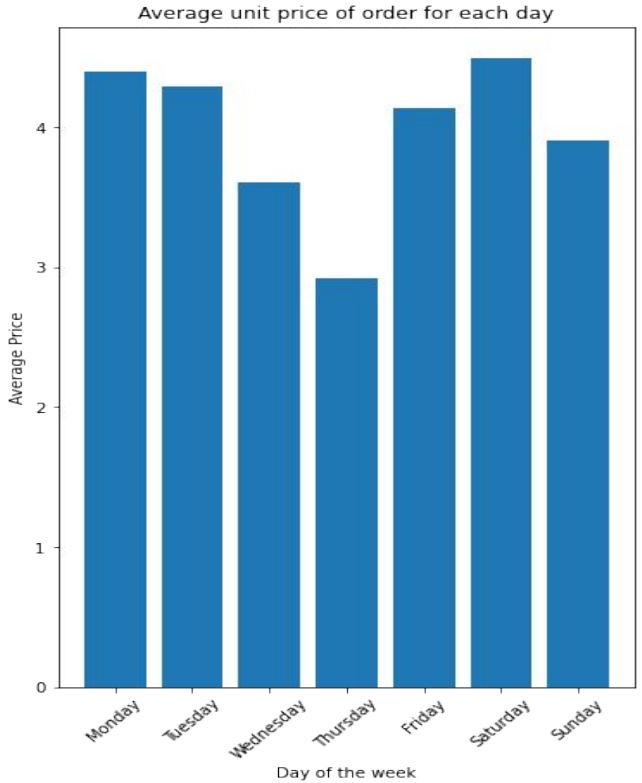
Number of unique orders each day

- Wednesdays seems to have the lowest unique sales volume.
- Friday has the highest unique sales volume.



Average unit price of order for each day

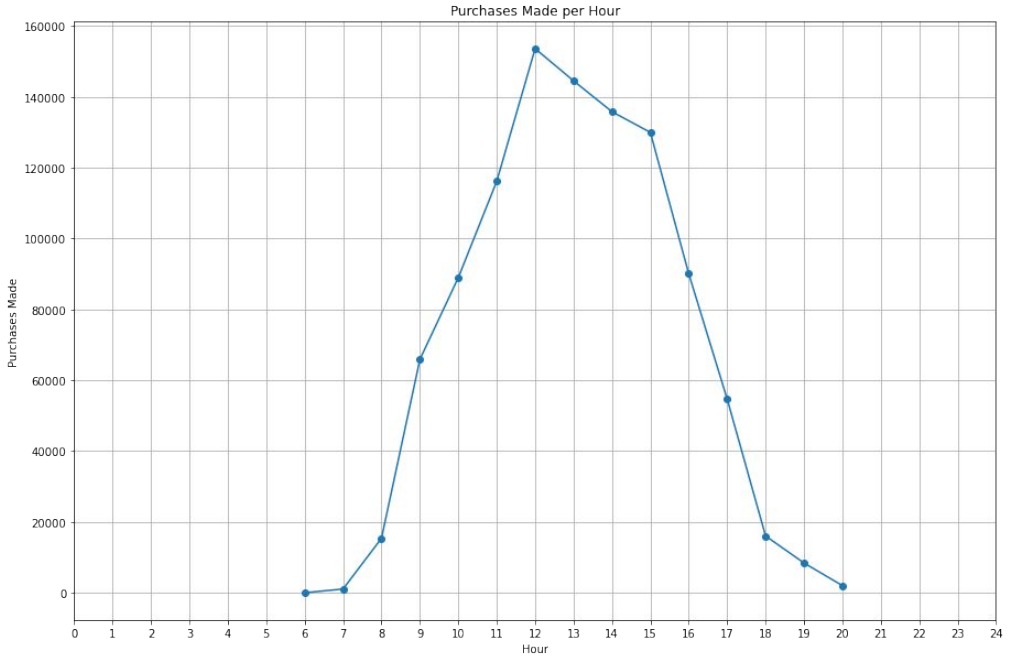
- Buyers tend to purchase lower value items on Thursdays.
- On Saturday buyers tend to purchase the highest value items.
- On other days, the value of items is fairly similar.



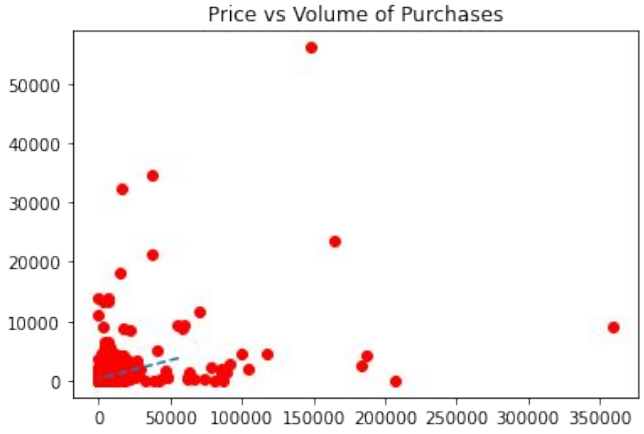
Time of the day with highest sale volume

As shown, most of the purchases were made during the afternoon.

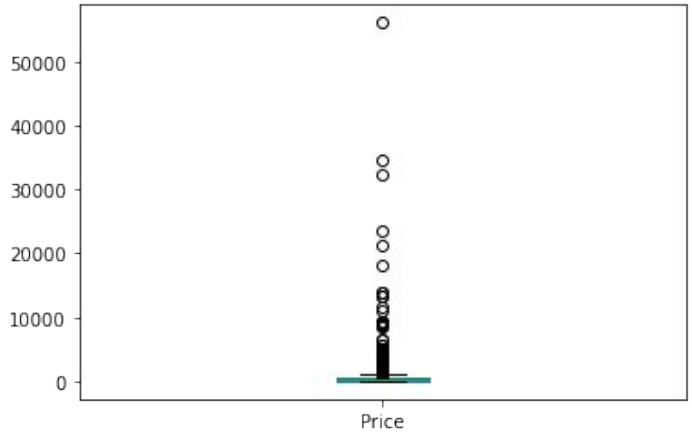
This might be due to many of the customers being wholesalers, scheduled overtime website maintenance cycles to allow time for stocktake and reconciliation, etc.

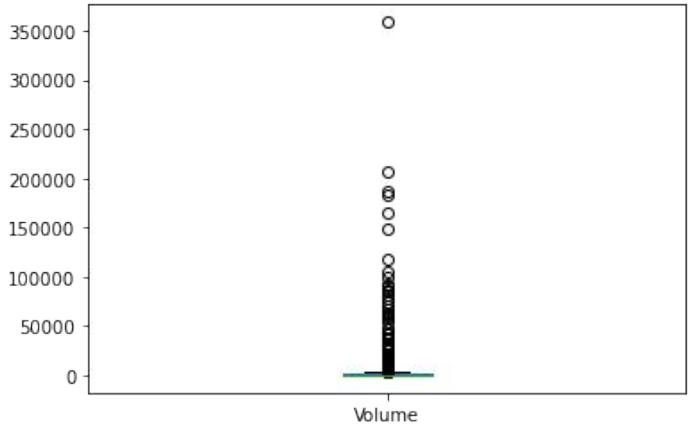


Money spent vs quantity purchased for all customers and trend



Highest Price and Volume for Customers





Massive outliers! It's a positive.

Data Limitations

1. Time limitation of the dataset, it is only showing information from December 2009 to the end of 2011
2. Data is only from the UK from an unknown wholesaler, restricting deeper data analysis.
3. Customer information is limited to just a customer ID, restricting gender, age or demographic analysis
4. Geographic information was limited to country, which meant we could only look at country vs country
5. Finding the impacts of returns and refunds against the total sales and volume information.

Conclusion

- We found that there was no particular correlation between time of day or day of the week
- We found that time of year had a slight correlation with an increase sales, with November being the most popular month.
- The UK as a country has significantly the most sales, which means that local customers are the most common.
- There is a slight correlation with an increase in volume at lower prices
- We believe most customers for the wholesaler are commercial, which has lead to an interesting dataset. As a result, we expect that most purchases were during working hours and as a result of stocktake.

Further considerations

- Obtain data including detailed information such as discounts, refunds, and specials that may be applied, this would allow for understanding if a loyalty program or sales is a better way to increase profitability.
- Obtain data including cost and profitability so we can find other inefficiencies in sales to match the market.
- Obtain more data that gives a more detailed breakdown on customers, from Commercial vs Private, as well as age, gender or location, rather than just Country and Customer ID.

Thank you for listening!

Does anyone have a question?