BUSINESS CASE ANALYSIS

Analyzing Retail Information

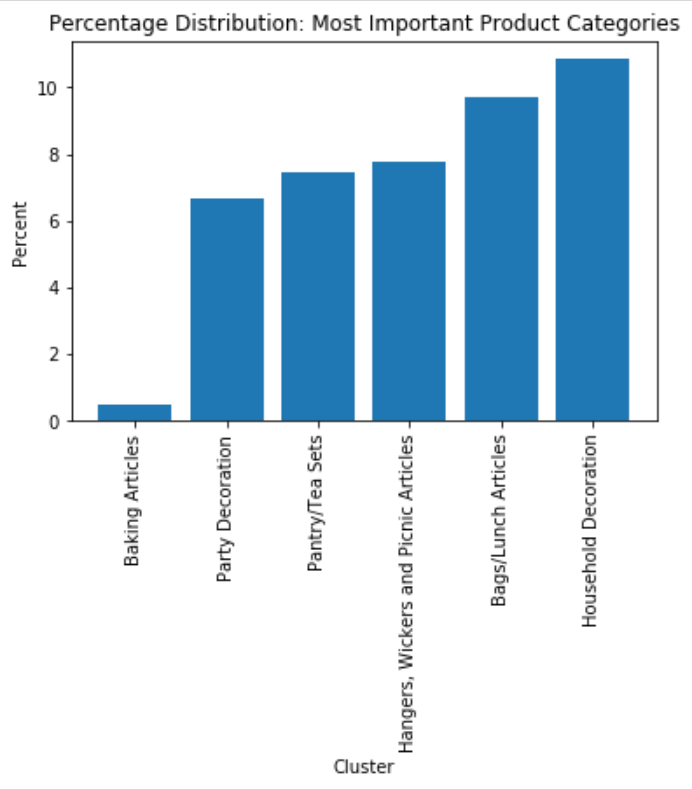
Dataset direct description

- The dataset represents purchases of products on a retail store. The attributes available are:
  - InvoiceNo: the number of the ticket
  - StockCode: SKU from the item
  - Description: description of the article
  - Quantity: amount of units bought by the customer
  - InvoiceDate: date of the purchase with hour
  - UnitPrice: price of the product by unit
  - CustomerID: ID of the customer
  - Country: country where the purchase took place
- Amount of unique products: 2663
- Customer with more purchases: 7983
- Customer with the least purchases: 1
- The timeline available is from December 2010 to December 2011. Thus, we are not able to detect properly seasonality condition, we need at least 2-3 years to be able to do so.

Dataset statistical aspects

- The average amount of times that a customer goes into the website and buys something is 93.
- The average amount of money spent per customer is 322.
- The average amount of articles that each customer buys is 199.
- We are analyzing 38 different countries in total.
- The country with more total sales is **United Kingdom**. This country has a lot of **good wholesales**, because the mean and median of the historical purchases are far apart from each other.
- The country with less sales is Saudi Arabia.
- The country with the highest average sales is The **Netherlands**. This country has a lot of **direct sales** because the mean and the median are close to each other.
- The country with the least average sales is USA.
- The price of the product that people tend to buy the most is 2.51.

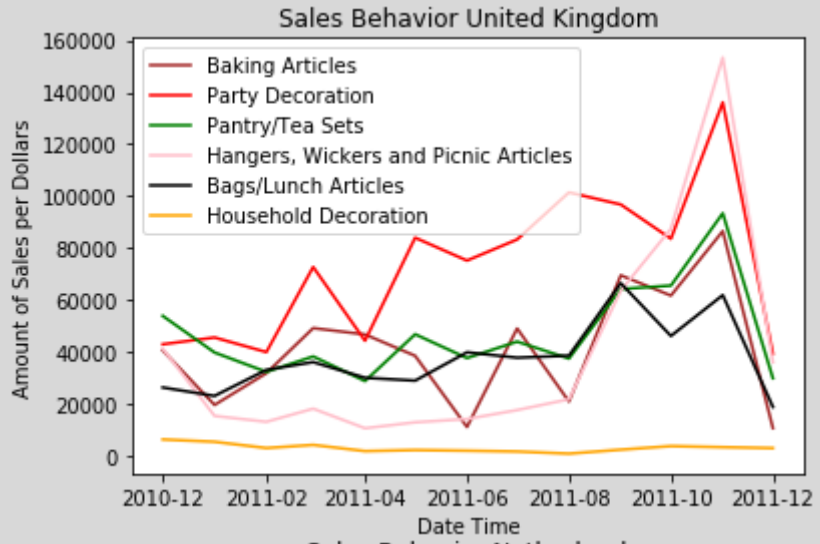
Text Mining Analysis



- From the information available on the product description. It was possible to find groups of products.
- In the graph you can appreciate the top 5 most important groups of products which represent the 45% of the whole sales.
- The household decoration are the products that we tend to sell the most.
- The baking articles are the products that we sell the least.

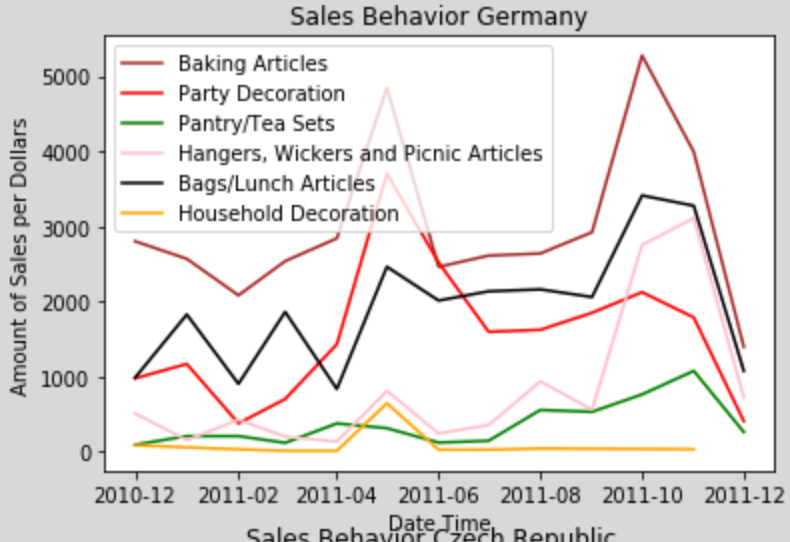
Analyzing Sales Behavior per most Important Countries

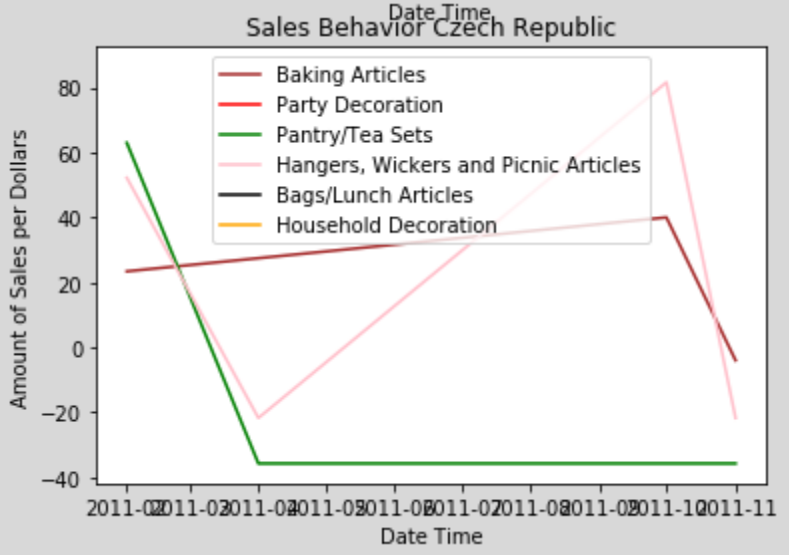
- In the UK (#1 sales) we can appreciate that the most important category is the "Party Decoration", followed by the "Hangers, Wickers and Picnic articles".
- In the Netherlands (#2 sales) we can appreciate that the "Party Decoration" is the most popular product, followed by the "Bags/Lunch articles".



Analyzing Sales Behavior per most Important Countries

- In Germany (#3 sales) people love to buy “Baking Articles” online, followed by “Party Decoration”.
- One of the worst countries in sales is the Check Republic with some profitable sales in “Baking Articles” and “Hangers, Wickers and Picnic articles”. It can be appreciated the money lost in the Pantry/Tea Sets, most probably due to problems (Delivery issue, damage article, return of article in bad conditions, etc).

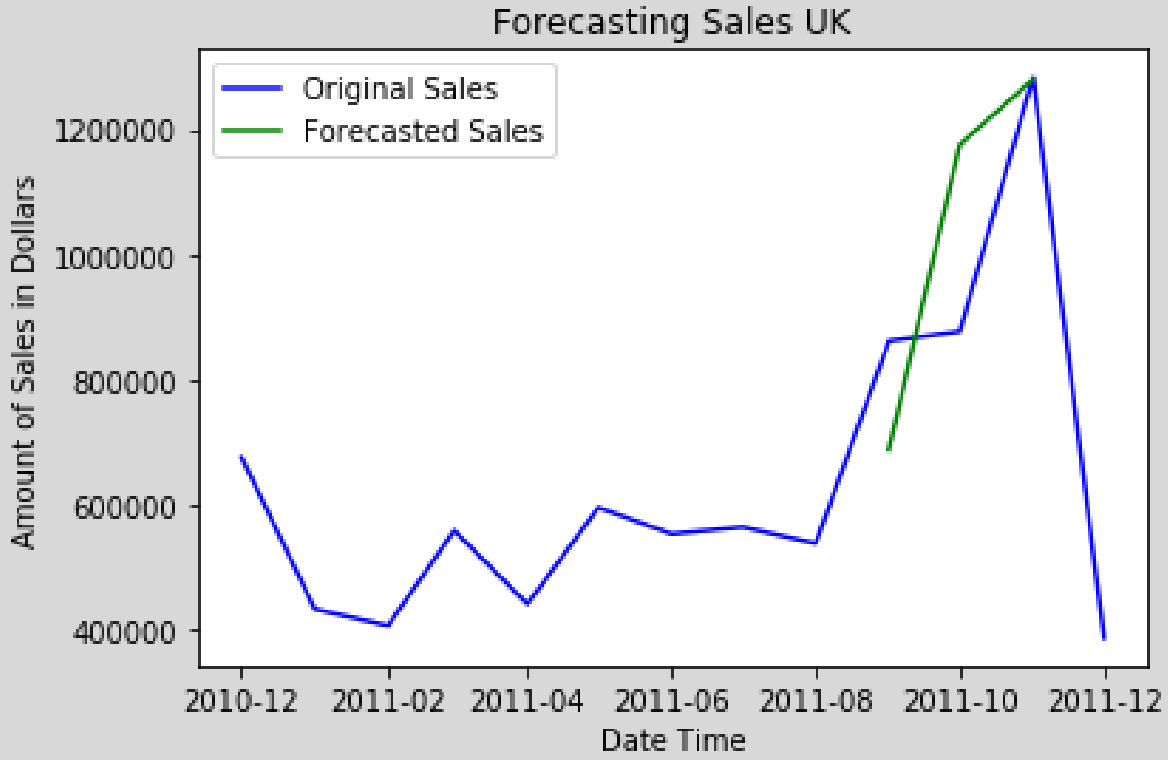




Analyzing Sales Behavior per most Important Countries

- The most baking article are bought by Germany.
- Overall, the household items are the ones with more sales worldwide and are the most stable in terms of purchase behavior. The trend of this kind or product tend to be almost null. But always taking place constantly.
- The party decoration has a pick on December and during carnival seasons. UK, April to June. The Netherlands April to august, due to king's day, plus carnivals, summer techno events. Germany April to August, open airs during the summer, der Karneval der kulturen, koeln Karneval etc.
- Overall, the bags category is the one which people tend to buy in the third place.
- The pantry and tea sets, tend to be a good option for the British people.

Forecasting Sales in UK



- Another important application that we can do is to forecast sales by country. Here is an example of the forecast for the UK sales.
- Thanks to the data available it is possible to create a forecast from January till November. However, we will need more years to be able to predict properly December.

Next Steps

This example illustrates what can be done with data available from a retail company.

We can check marketing promotions.

We can create a reward program base on the number of purchases per customer. Create customer segmentation, track purchase history, combinations of purchases.

We can check which are the most and least buy products and base on it, create combo offers. To move faster merchandise.

We can check how the sales products moves worldwide and base on it, store products which are more likely to be bought in the most likely location to be buy. (Logistics savings)

We can check forecasting to keep in store only what we need. Optimal use of our warehouse.

Thank You!

Feel free to reach me out for any other advance analytics case. Benazir de la Rosa, Data Scientist Consultant.