

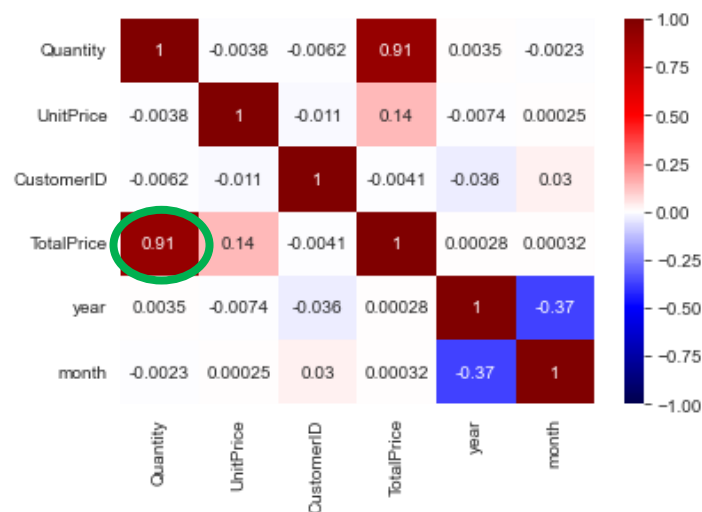
## Data Description:

This dataset can be found at Kaggle, which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

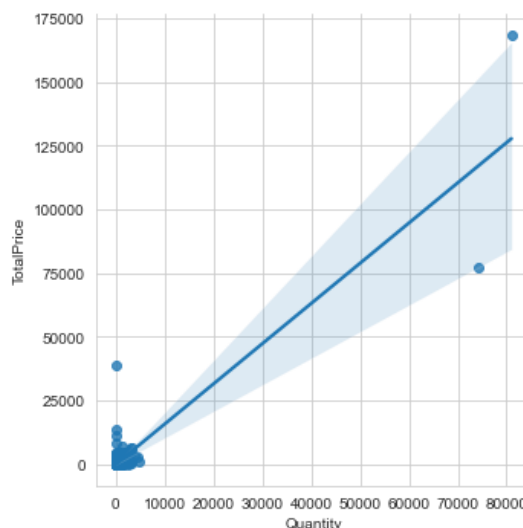
Data shape (542k,8)

- **InvoiceNo:** Invoice number that consists of 6 digits.
- **StockCode:** Product code that consists of 5 digits.
- **Description:** Product name.
- **Quantity:** The quantities of each product per transaction.
- **InvoiceDate:** Represents the day and time when each transaction was generated.
- **UnitPrice:** Product price per unit.
- **CustomerID:** Customer number that consists of 5 digits. Each customer has a unique customer ID.
- **Country:** Name of the country where customer receive the item.

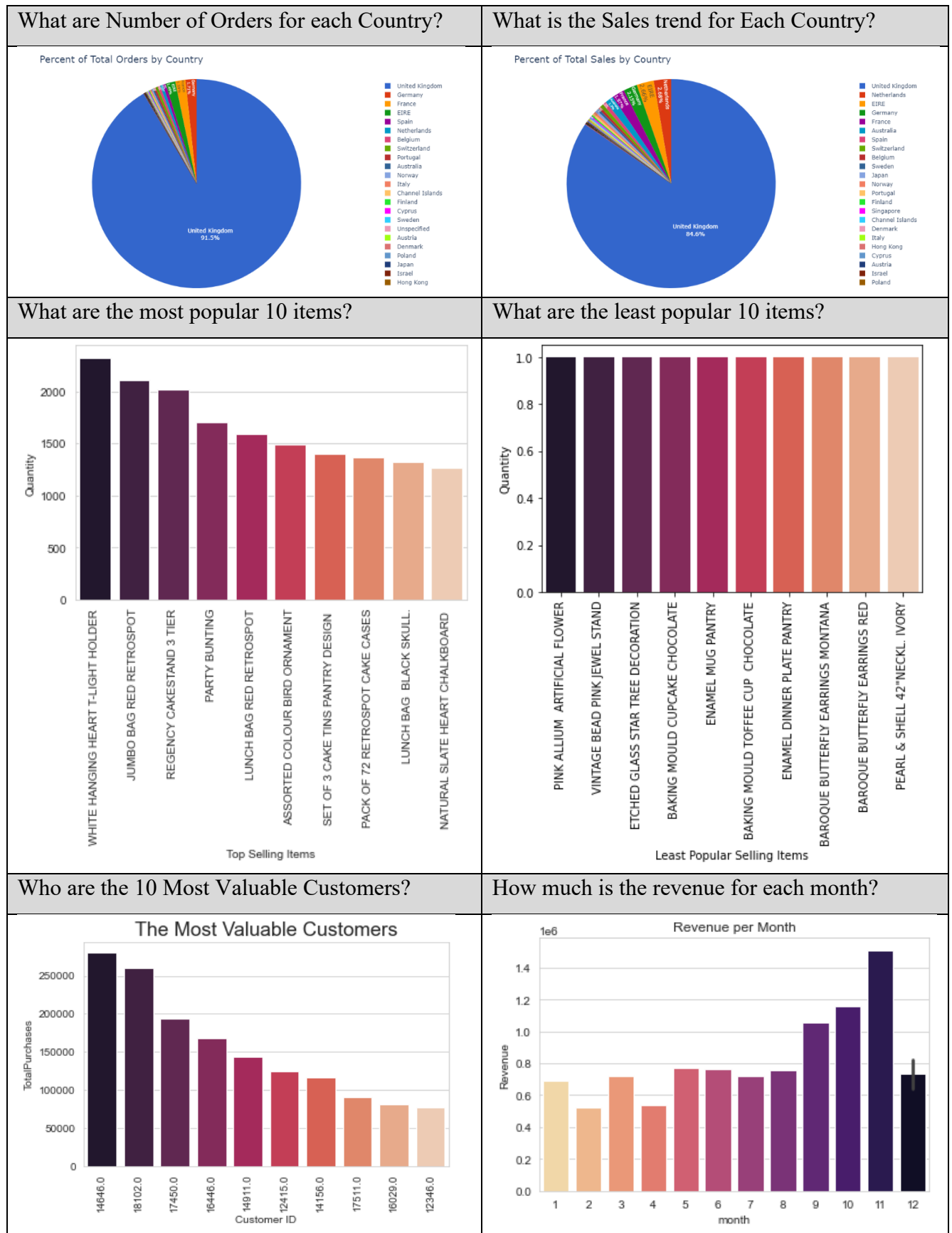
## Correlation matrix:



It seems that there is a Strong Relationship between the Quantity and Total Price



## Questions:



**Awatif Alshehri**

|                                   |                              |
|-----------------------------------|------------------------------|
| Best Customers: 456               | Loyal Customers: 872         |
| Customers at risk of churning: 70 | Almost Churned Customers: 10 |
| Churned Customers: 444            | Big Spenders: 1085           |

| Modeling Scores:   |     |
|--|-----|
| A. KMeans  | 33% |
| B. MiniBatch KMeans  | 29% |
| C. DBSCAN  | 14% |
| D. Spectral Clustering   | 32% |
| E. Mean Shift  | 43% |
| Conclusion   |     |
| The Mean Shift Model with n_clusters = 3 was the best performing models, based on Silloutte score !! |     |