E-commerce System User Behaviour Visualization

Emir Alaattin Yılmaz
Computer Science and Engineering
Sabancı University
emiralaattin@sabanciuniv.edu

Abstract

By incerase of using online retail web sites the competition between e-commerce web sites to attract customer has also increased. Therefore, it is important to have a strong data insights and analyze user behaviours to provide relevant products to the customers. In this project, an e-commerce web site data is analyzed and a tool is presented that visualizes the data as a force-directed graph with various types of filters such as country, timeline and calendar view to have insights.

Keywords: e-commerce, data visualization, force-directed graph

ACM Reference Format:
Emir Alaattin Yılmaz. .. E-commerce System User Behaviour Visualization. Istanbul, Turkey, 4 pages.

1 Introduction

In recent years, e-commerce web sites have been highly used by people due to convenience from homes the competition between these type of web sites to attract customer has increased. Therefore, it has become important to have a strong data insights and analyze user behaviours to provide relevant products to the customers in a time-efficient manner. To have such insights, data visualization techniques can be used as a practical tool.

In this project, an e-commerce web site data is analyzed and visualized as a force-directed graph with various types of filters.

2 Dataset Description

In this project, an online UK based retail web site data provided by UC Irvine Machine Learning Repository which has 541909 rows. This transnational dataset contains all user transactions on this web site since December 2010 to December 2011. All of these transactions are made across 4339 people and the number of products that is purchased is 3941.

3 Data Preprocessing

To preprocess the data was a challenging part of the project since the transactions are not grouped for each people. First, the cart information of each person is retrieved by grouping the data by invoice number. In addition, grouping also includes categorizing of transactions as monthly, daily and country-based. After dropping nan-values and cancelled transactions the cart transactions are obtained as 20726 distinct rows.

4 Data Sampling

After dropping nan-valuas we obtained 387875 user-product interaction, 4339 people with 3941 products. Since there are certain number of people and products, it can be worth to visualize these connections as a graph network. But for aesthetic visualization purposes, it is crucial to simplify large graphs and apply a sampling method to visualize and explore important connections [1].

There are used two thresholds to apply sampling in the network. In the relatively bigger network, the method was to apply sampling as obtaining people who have purchased more than 200 products and sampling of products by if this product has been added to cart more than 100 times. In the smaller network these values are changed into obtaining people who purchased more than 400 products and sampling products added to cart more than 1000 times. As a result, 454 people, 1541 products and 169740 edges in the bigger network and 139 people, 39 products and 10481 edges are obtained in the smaller network which is summarized as below.

Table 1. Non-Sampled and Sampled Network Properties

Network Name	# of People	# of Products	# of Edges
Non-Sampled Network	4339	3941	387875
Big Network	454	1541	169740
Small Network	139	39	10481

5 Data Visualization

5.1 Force-Directed Graph Visualization

Force-directed graphs can visualize the connections and cluster the relevant nodes in a natural way because it models the nodes as the charges repel each other.

Since there are many connections between users and products over time in the data set, force-directed graph representation is used for visualization of user-product interactions

, ,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/.

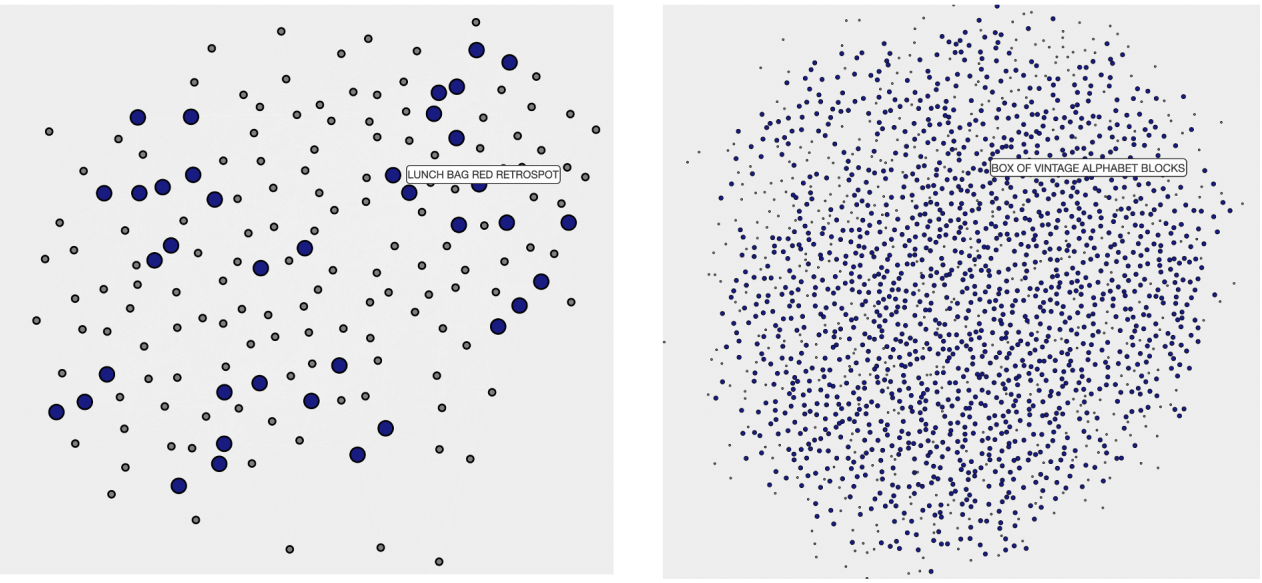

Figure 1. Force Directed Graph Visualization of Big Sampled (Left) and Small Sampled (Right) Networks

where people and products are different type of nodes and the interactions between them are shown as edges.

In Figure 7, sampled big network and sampled small networks are visualized as force-directed graphs where small grey nodes represent users and bigger blue nodes represent products.

5.1.1 User-Product Interactions. The tool which is developed for this project allows to analyze user-products interactions on the force directed graph to analyze a user to buy which products and user interests. This information can be used to offer special campaigns specific for these users. Moreover it can reveal the common interests between users.

5.1.2 Product-User Interactions. The tool also allows to analyze product-user interactions to show who purchased that product that can show which user groups have interests in these products.

5.2 Scatter Visualization

Scatter plot visualization is used to represent user activity over time to analyze how many products that user purchased on timeline. It can be used to send special campaign emails to attract passive users.

5.3 Line Chart Visualization

Line chart visualization is used to understand change in number of interactions in a timeline and can reveal times of the year when people were more active to make purchases.

5.4 Calendar Visualization

The data is same with line-chart but in a different form that shows the which days of the week when people have made more purchases as a heat map.

5.5 Bar Chart Visualization

To reveal which countries have more users who made purchases is represented as a bar chart except United Kingdom since it has 18784 instances among total number of transactions which is 20726.

6 Visualization Tool

Aforementioned visualizations are combined and presented in a tool that is developed for this project. The tool includes force-directed graph visualization that shows user-product and product-user interactions by highlighting the connections on clicking actions. Moreover, scatter visualization of a user activity clicking on a user node twice.

6.1 Country Filter

Users can be filtered according to their countries by use of country filter that can help to analyze which countries are more active in purchasing. This can also help to analyze products which the people are interested in specific countries. It can be used to increase stocks of these products in these countries.

6.2 Time Filter

Another important feature of the developed tool is the ability of the filtering of people according to a timeline. Selecting an area of the timeline can allow in understanding which users made transactions and which products were purchased.

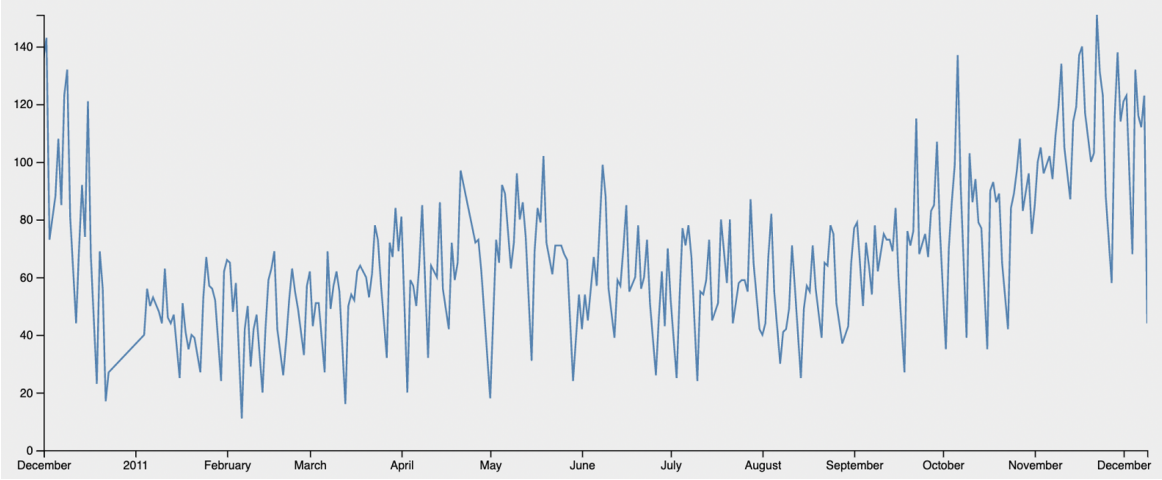

Figure 2. Line Chart Visualization of number of transactions over time

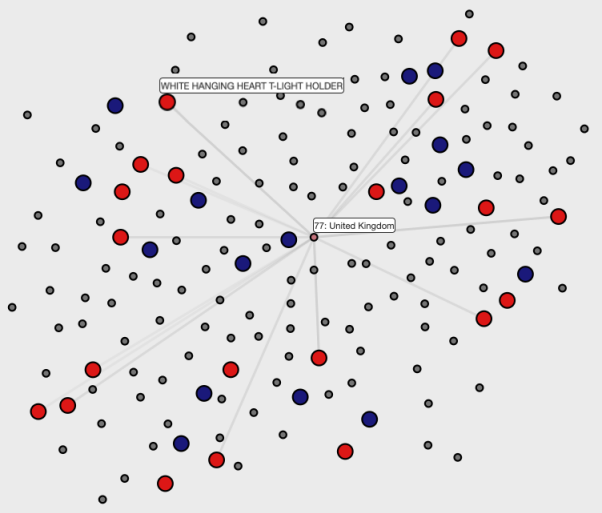

Figure 3. User-Product Interactions

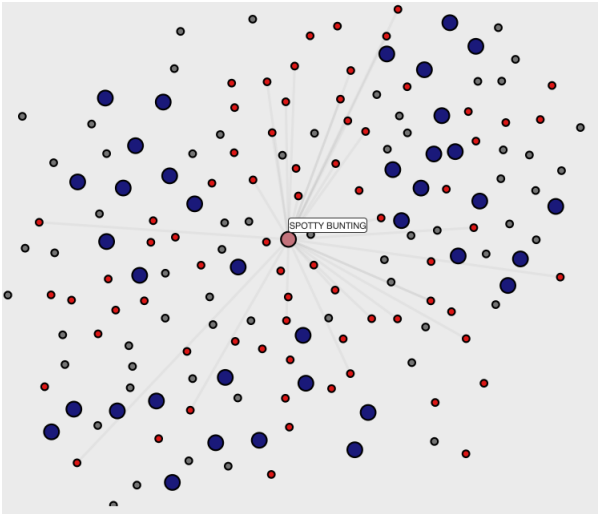

Figure 4. Product-User Interactions

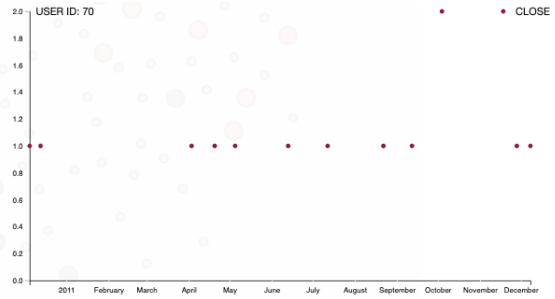

Figure 5. User Activity Scatter Visualization

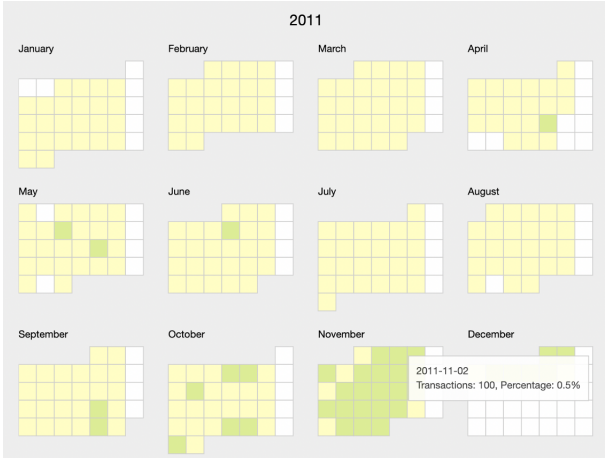

Figure 6. Bar Chart Visualization of number of transactions over countries except United Kingdom

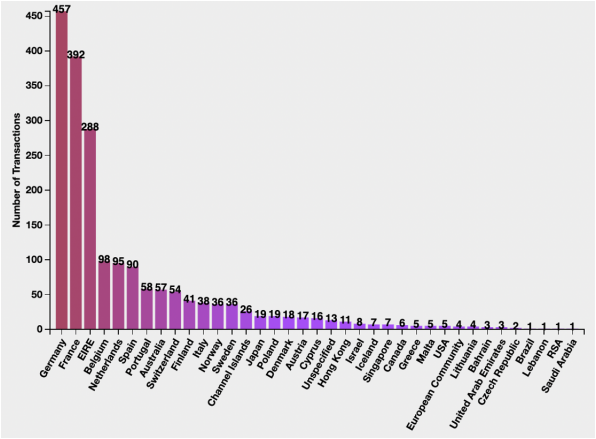

Figure 7. Bar Chart Visualization of number of transactions over countries except United Kingdom

6.3 Implementation

For the implemention of the tool, d3.js which is a JavaScript library to manipulate documents by use of data.
Implementation:
https://github.com/alaattinyilmaz/ecommerce-data-visualization/
Demo: https://alaattinyilmaz.github.io/ecommerce-tool/

7 Conclusion

In this project an online retail website data is visualized on a force directed graph and a tool is developed for exploration of user-product, product-user interactions that can filter the user and product data according to their country and transaction time.

References

[1] Z. Zhou, C. Shi, X. Shen, L. Cai, H. Wang, Y. Liu, Y. Zhao, and W. Chen. 2020. Context-aware Sampling of Large Networks via Graph Representation Learning. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. https://doi.org/10.1109/TVCG.2020.3030440