

FIT5147 Visualisation Project with D3js

Student Name: Vinh Phan
Student ID: 27612937
Student Email: qvpha1@monash.edu.au

Contents

1. Introduction	2
2. Design	3
a. Data Overview	3
Reviews Data Distribution	3
Correlation: Top Sales vs Top Reviews.....	4
b. Customer Satisfaction.....	6
Are customers happy to buy product?	6
Where do the reviews come from?	7
c. Product Similarity	8
3. Implementation.....	9
4. Conclusion	10
5. References.....	10

1. Introduction

Amazon is one of the biggest e-commercial websites doing so well in this domain around the world. Nowadays, people come to Amazon to buy and sell products. The increasing demand has encouraged me to conduct a project to analyse the customer satisfaction on products and groups of similar products, so that the sellers can ease to select highly profitable products among million items on Amazon website. It's difficult to sellers to know which items they should stock, and those are more beneficial and fast movement.

To answer the question, **which items should sellers stock?**, we will look at 3 factors that make the success of items in Amazon

- ✓ Are buyers happy after purchasing the item?
- ✓ Should we buy top sales ranking items or top popular items?
- ✓ What are the “bought together” items of the original?

In the report, we will introduce 3 visualizations to partly answer those questions by analysing “Product Metadata” of products, and “Product Reviews”

- ✓ Cloud Word visualization can help us to understand what customers talk about the product they bought. After wrangling data to extract emotional words and highly frequent occurrences of each particular item, we will have the result in “Customer Satisfaction” report
- ✓ Scatter Plot visualization is the solution for measuring correlation between top sales rank and highly frequent reviewed items
- ✓ Network is used to visualize the related items

2. Design

a. Data Overview

Reviews Data Distribution

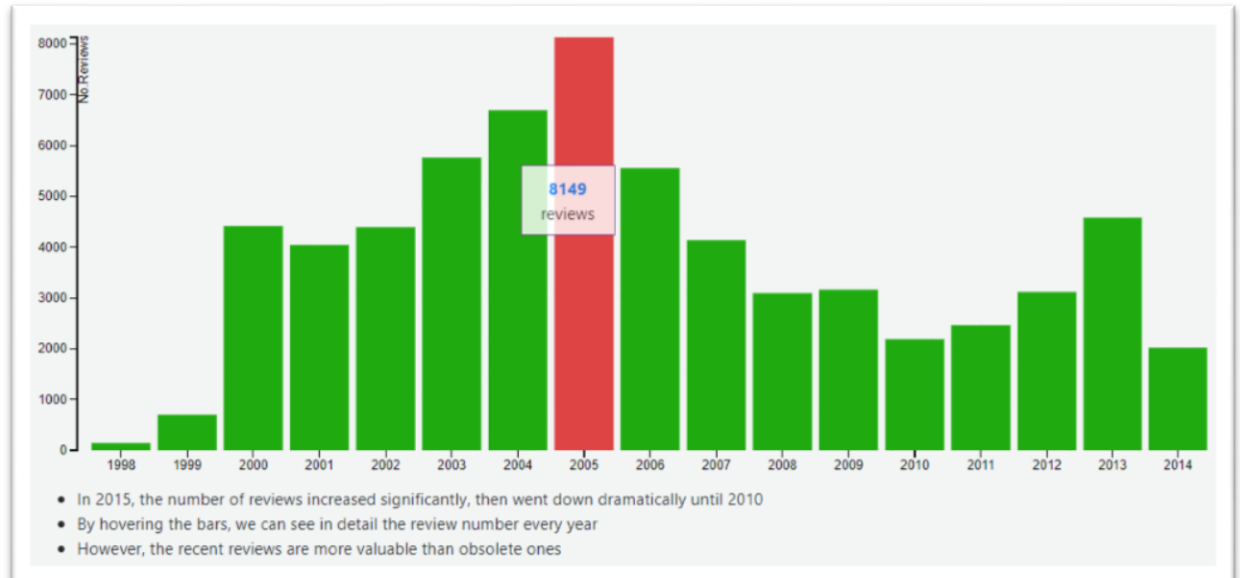


Figure 1: Dataset Distribution with Histogram

Five Sheets Design Principle	
Title	Number of Reviews Distribution
Data	Review dataset from Amazon (from 1998 to 2014)
Task Description	Give us a taste of how the whole dataset is distributed, and when the most reviews were made
Focus / Detail	By hovering the bars, we will see the number of reviews recorded each year
Advantages	✓ Quickly pick up the dataset for the first time
Disadvantages	✓ The data is obsolete, and we cannot see it in more detail
Alternative	Using Pie Chart

Correlation: Top Sales vs Top Reviews

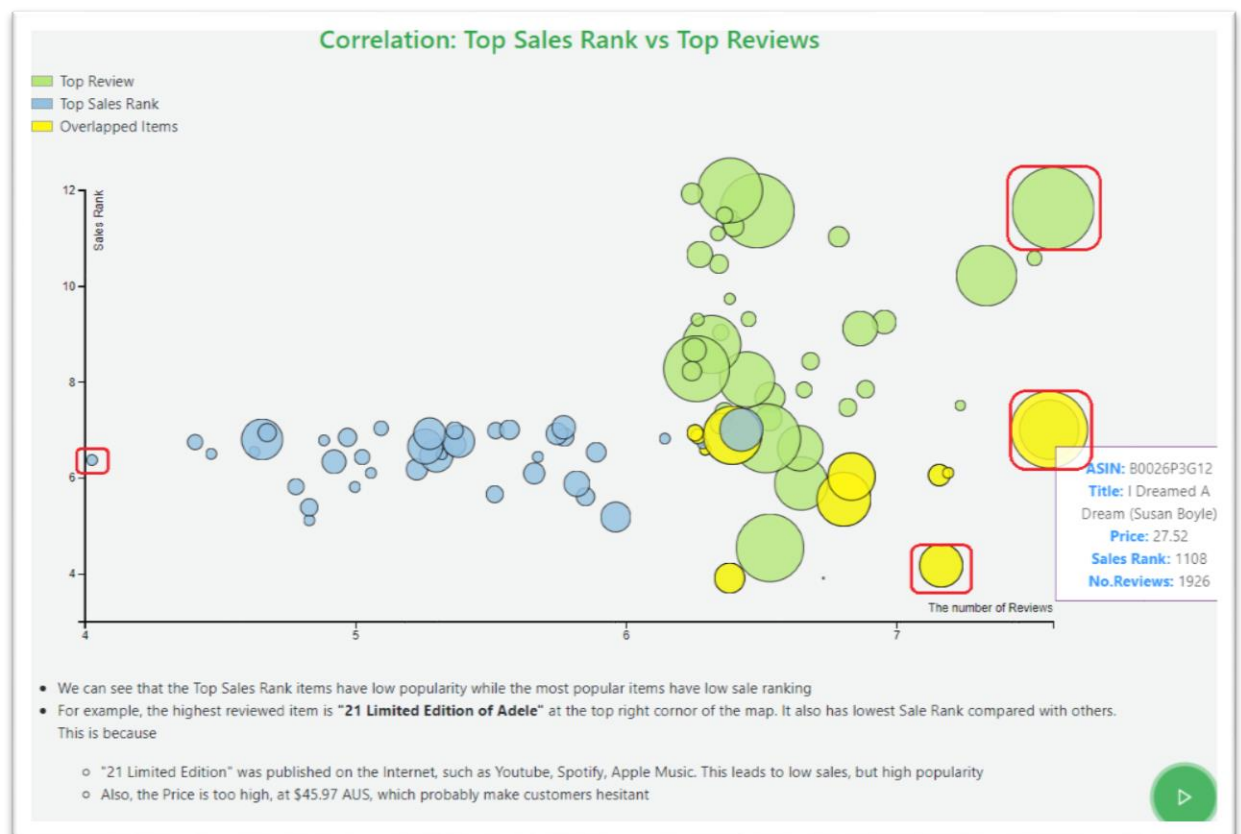


Figure 2: Top Sales Rank vs Top Reviews with Scatter Plot

Five Sheets Design Principle	
Title	The correlation between Top Sales Rank and Top Reviews
Question	Whether or not the items with top ranking are also well-known in popularity
Task Description	We filter top 50 highest ranking items, in over 200,000 items in the dataset, and 50 highest reviewed items
Focus / Detail	Categorising items to 3 groups. The "overlapped items" is those overlapped between top review and top sales rank
Operation	Hover the circle, the box will pop up with detail of that item

Advantages	Have a free to explore the data with scatter plot Multi-dimensional support <ul style="list-style-type: none"> ✓ Sales Rank (y-axis) ✓ Number of Reviews (x-axis) ✓ Price of item (circle size) ✓ Category (colour)
Disadvantages	<ul style="list-style-type: none"> ✓ The data is obsolete, and we cannot see it in more detail
Alternative Solution	Correlation Matrix with colour and circle size. However, we use 100 items, it is not a good idea to use Correlation matrix with big dataset due to readability, and beauty

Discoveries:

- ✓ In the picture, the red rectangle are important items we should consider.
- ✓ For example, album “I Dreamed A Dream of Susan Boyle”, yellow circle,
 - ✓ High sales rank => high sales in Amazon
 - ✓ Very famous with 1926 reviews, just a bit less than “21 Limited Edition”
 - ✓ High Price, \$27.52 per album
 - ✓ To sum up, if we can stock it with lower price, we can quickly get return due to its popularity, sales rank, and high profit

b. Customer Satisfaction

- Are customers happy to buy product?

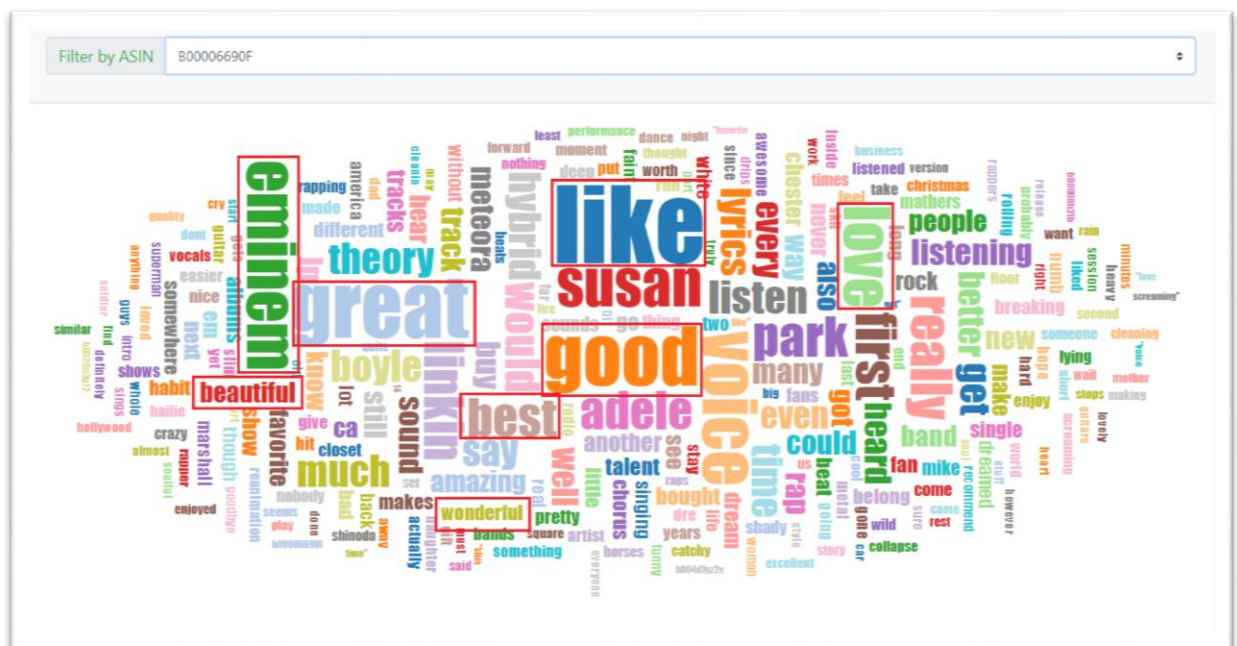


Figure 3: Word Analysis with Cloud Word

Five Sheets Design Principle	
Title	The cloud word of each album
Question	Are customers happy to buy the item?
Task Description	<p>We collect reviews for each items in the highly reviewed group, then wrangle it to have the final <i>bag word</i></p> <p>After that we used D3js to visualize the bag word with the words' density</p>
Focus / Detail	Hover on the word, we will see the occurrences of that word in the corpus
Operation	On the top, there is a combo box to change the album
Advantages	Quickly and easily pick up the keywords from the corpus. It gives a general view about the album we're analysing

Alternative Solution	Not available at the moment
----------------------	-----------------------------

Discoveries:

- ✓ In the picture, we can see the name of the singer, Eminem – a rapper. We can also see that the most buyers like the album, and they are happy after bought. The positive words are highlighted with red rectangles

Further:

- ✓ The word cloud visualisation is actually a part of “*item analysis*”. What we’re trying to support an “*auto complete search with Ajax*”. The word cloud gives sellers a deeper understanding about the product via its reviews. However, due to client site development, it’s not practical to support search api in the big dataset purely with javascript.

- ❖ Where do the reviews come from?

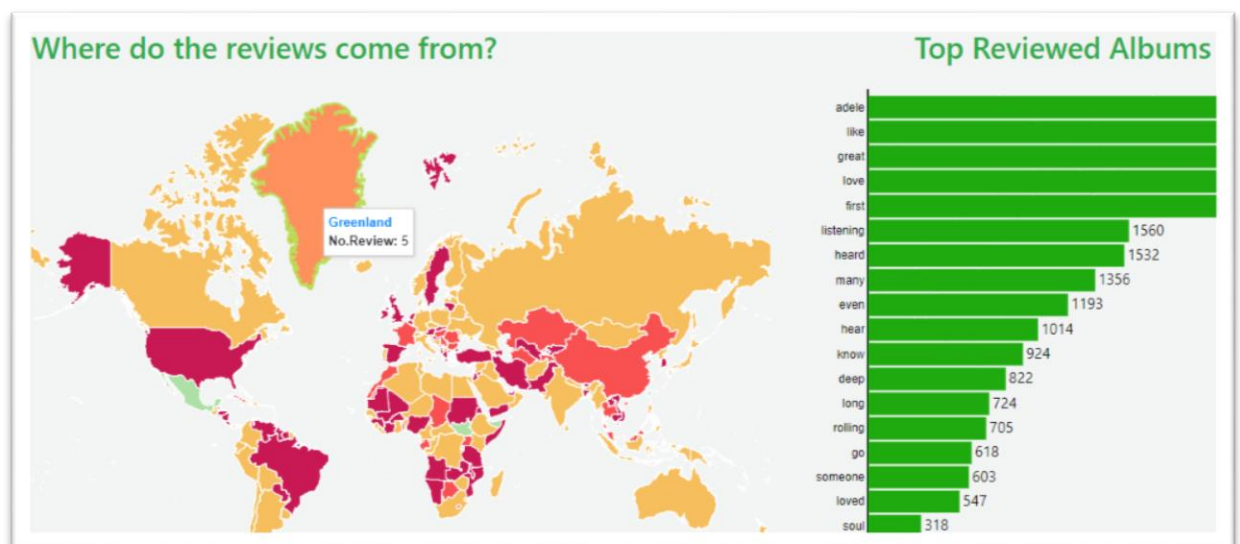


Figure 4: Map and Horizontal bar chart with D3.js

It is useful to know which countries contribute mostly to the reviews. Looking at the picture, we can see China and America are the big contributors.

c. Product Similarity

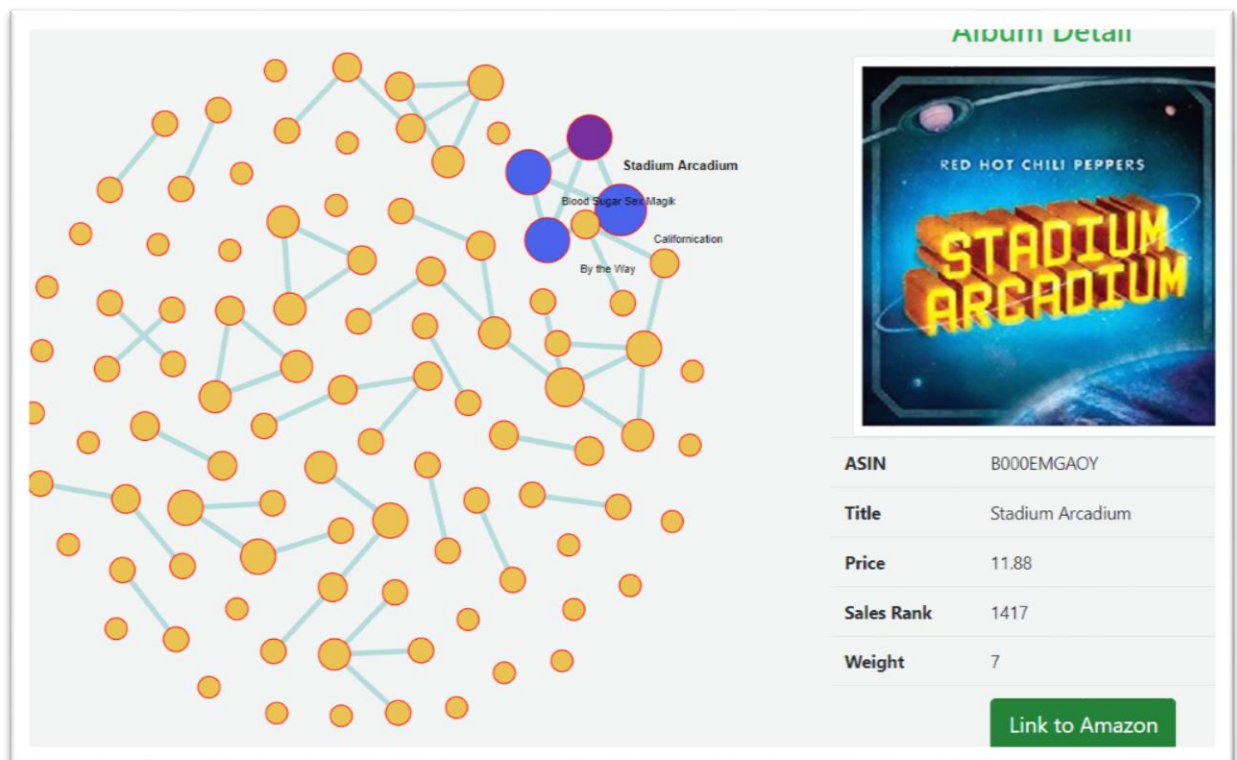


Figure 5: Similar Items visualised with Forced Network Graph D3js

Five Sheets Design Principle	
Title	Force network graph for Similar Albums
Question	What products should we buy together?
Task Description	<p><u>There are 3 steps for this Visualisation</u></p> <ol style="list-style-type: none"> Running “<u>Apriori</u>” algorithm to derive the list of similar items Wrangling data, and <u>convert it to Network graph</u> format (nodes, edges) Use D3js to <u>visualize forced network graph</u> with animation, and interaction <p><u>Notes</u>: detail of data wrangling is mentioned in Exploration Document</p>
Focus / Detail	We can see detail of the album crawling from Amazon website, such as album images, title, price, sales rank, and hyperlink to Amazon

Operation	<p><u>Hovering</u>: to see the title of node quickly</p> <p><u>Clickable</u>: to see album information</p>
Advantages	<p>Derive insight from data, it gives sellers a hint to extent their view</p> <p>We can quickly pickup similar items. A reference link to Amazon website if we want to have more detail</p>
Disadvantage	It takes time to see the final result. We have to run the algorithm, wrangling data, converting to nodes and edges, finally visualize with D3js
Alternative	<p>Sequences sunburst graph with D3js</p> <p>However, it's more interactive with Network graph, and easier to understand.</p>

3. Implementation

Javascript	D3js version 3, jQuery
HTML / CSS	Bootstrap V4, CSS3 and HTML
Viz Technology	Bar chart with D3js Scatter Plot with D3js Word Cloud with D3js Choropleth Map with D3js Forced Network Graph with D3js
Development Environment	Visual Studio Code with Live Review plug-in Chrome browser Windows 10 Home
Online Version	https://phanvinh0526.github.io/amazon-reviews-d3js/d3js_data_overview.html

4. Conclusion

The project gives me a chance to work in different stage of analysing data from the scratch. We try to use different technologies, mainly D3js, to communicate the results to audience which is Sellers on Amazon

I try to build a complete solution to support the idea of “*which albums should Amazon sellers buy?*” based on the factors of profit and stock movement. However, the client site is not powerful enough to execute sophisticated query, such as auto complete search on huge dataset.

5. References

- Rodden, K. (2018). *Sequences sunburst*. [online] Bl.ocks.org. Available at: <https://bl.ocks.org/kerryrodden/7090426> [Accessed 4 Jun. 2018].
- Palmer', J. (2018). *Using d3-tip to add tooltips to a d3 bar chart*. [online] Bl.ocks.org. Available at: <http://bl.ocks.org/Caged/6476579> [Accessed 4 Jun. 2018].
- Villegas, P. (2018). *Untitled*. [online] Bl.ocks.org. Available at: <http://bl.ocks.org/paulovn/9686202> [Accessed 4 Jun. 2018].
- Mark Otto, a. (2018). *Introduction*. [online] Getbootstrap.com. Available at: <https://getbootstrap.com/docs/4.1/getting-started/introduction/> [Accessed 4 Jun. 2018].
- Agarwal, P., Lal Yadav, M. and Anand, N. (2013). Study on Apriori Algorithm and its Application in Grocery Store. *International Journal of Computer Applications*, 74(14), pp.1-8.