

# DC3 FINAL HANDIN

## INTRODUCTION:

In this assignment, we were asked to examine the “Tree of Stuff” and give interesting insights based on the categories dataset that we were provided.

We have used all-nodes.csv dataset to implement our visualization.

We assumed marketing leaders, businessmen, product managers to be our target audience. There are many categories with the same name. This makes understanding the product catalog challenging and sometimes confusing to identify the correct hierarchy. .

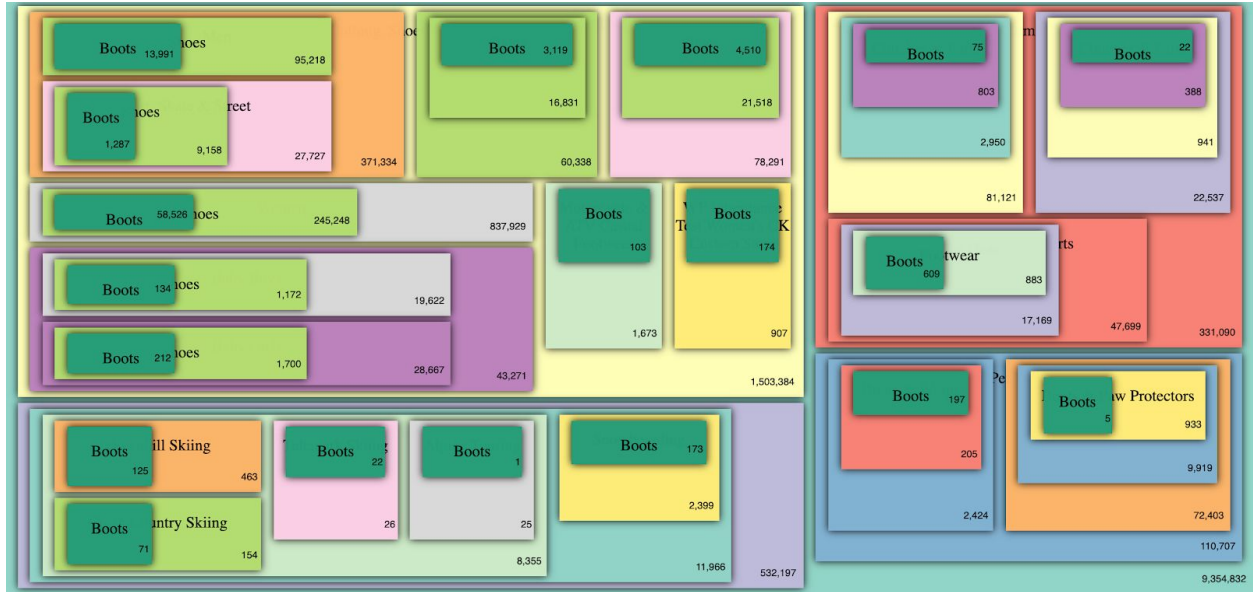
With this visualization tool, we are trying to address the following problems or questions:

1. How many categories are present with the same name ?
2. What are the different hierarchies of those categories ?
3. How many products are present within each category in those hierarchies?
4. How many children does a particular category have?

## TOOL OVERVIEW:

We have developed an interactive tool to address all the questions, ease of accessing information and navigate across the product category graph. This tool can also be used for explorative analysis. We selected zoomable treemap to illustrate the tree structure of categories.

Web Access : <https://cs765dc3.herokuapp.com/>



The user can search for a category, and the treemap displays only the hierarchy of categories with the searched category as leaf nodes (on top of the stack). The hierarchy of categories stacked on each other with visible layers. This helps scale the categories to a large extent and also avoids over-clutter.

Unlike node link diagrams, this visualization is highly scalable, since the treemap displays the hierarchy of categories with the searched category as leaf nodes that pop out, and thereby avoids over-clutter.

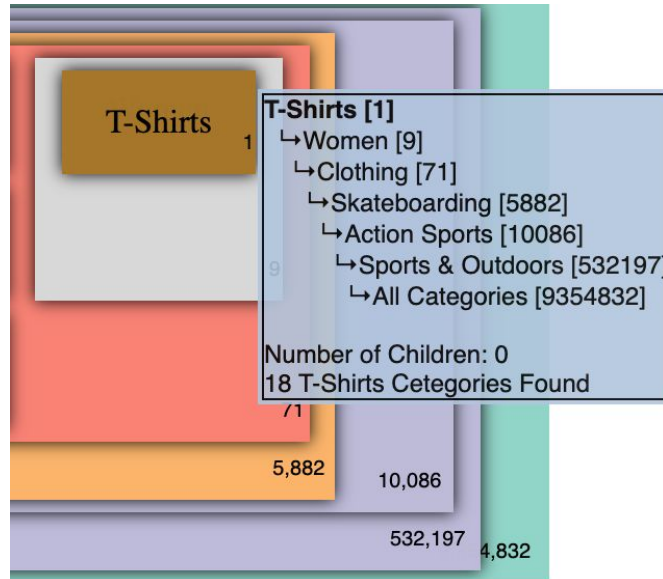


Figure 2: Tooltip showing the number of items in each category. Also shows the number of children in that category node and how many categories in the database with the same name.

Also, when you click the category, it zooms in completely using the zoom interaction, and shows you the number of items belonging to that category. This feature may be helpful for future extension of the tool and additional information about the product categories could be provided at different zoom level. You can zoom out of each category step by step either by using the 'UP' arrow on the top left corner, or the right click menu.



Figure 3: Functionality of 'UP'

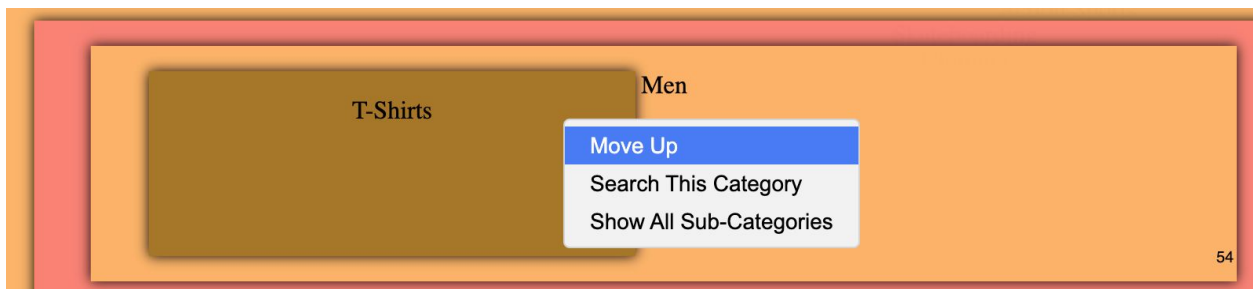


Figure 4: Same functionality as 'UP'

User can also explore any of the shown categories and visualize its sub-categories by using 'Show All Sub-Categories' option from the right click menu. This will show a supplementary visualization as shown in Figure 5 below:

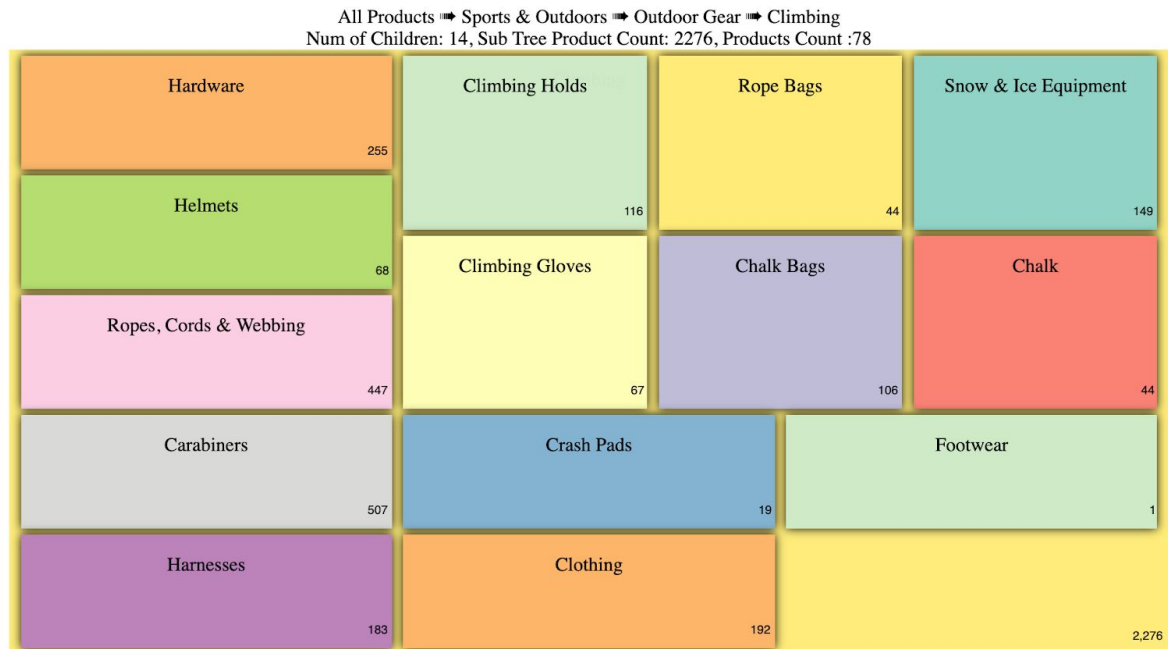


Figure 5: Shows all Subcategories of Climbing Outdoor Gear, with the hierarchy of the path leading to the searched category, Climbing, and details such as number of children of Climbing, sub-tree product count and products count are displayed.

## DESIGN RESTRICTIONS/LIMITATIONS:

**Size Encoding:** We used size encoding for subtreeProductCount. We didn't use size encoding strictly for the comparison purpose, but for visually identifying parent-child relationship. That means the rectangles are not strictly proportional to subtreeProductCount. This is intentional, as our purpose was to highlight all the hierarchies of the searched category belongs to. By forcing size encoding it would not be possible as the area of the intended rectangle significantly small as compared to the categories higher in the hierarchy. For example Figure 6 demonstrates the result with size encoding (strictly proportional to subtreeProductCount) after searching 'Boots'. This is not effective. To address the issue, we have provided subtreeProductCount information on mouse hover and also on the rectangles so that user can easily access this information.

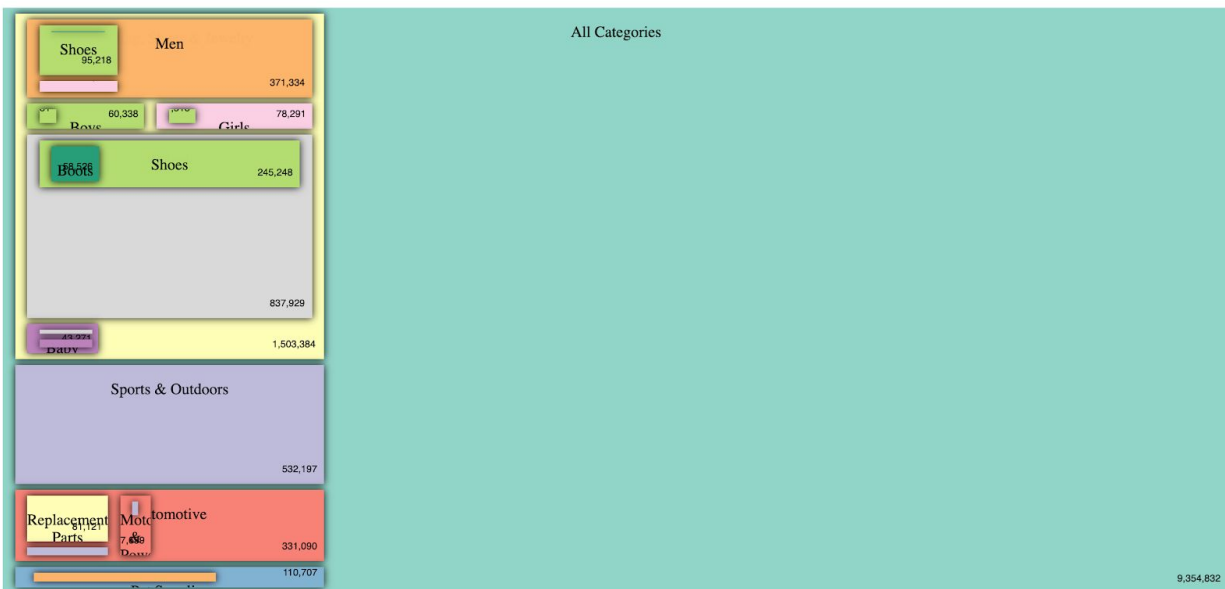


Figure 6: Example demonstrating the ineffectiveness of the size encoding

**Color Encoding:** Our intention was to use different colors for different categories. However, due to lack of the amount of colors available, it is difficult to establish this.

Since boxes have been used to denote parent and child categories, they overlap each other and it affects user readability. To address this we provided the hierarchy information on tooltip. We had thought of providing additional visualization of concentric circles to denote the parent and child categories of a single hierarchy, with the innermost circle representing the child category and the outermost circle representing the parent category.

## DISCUSSION OF IMPLEMENTATION:

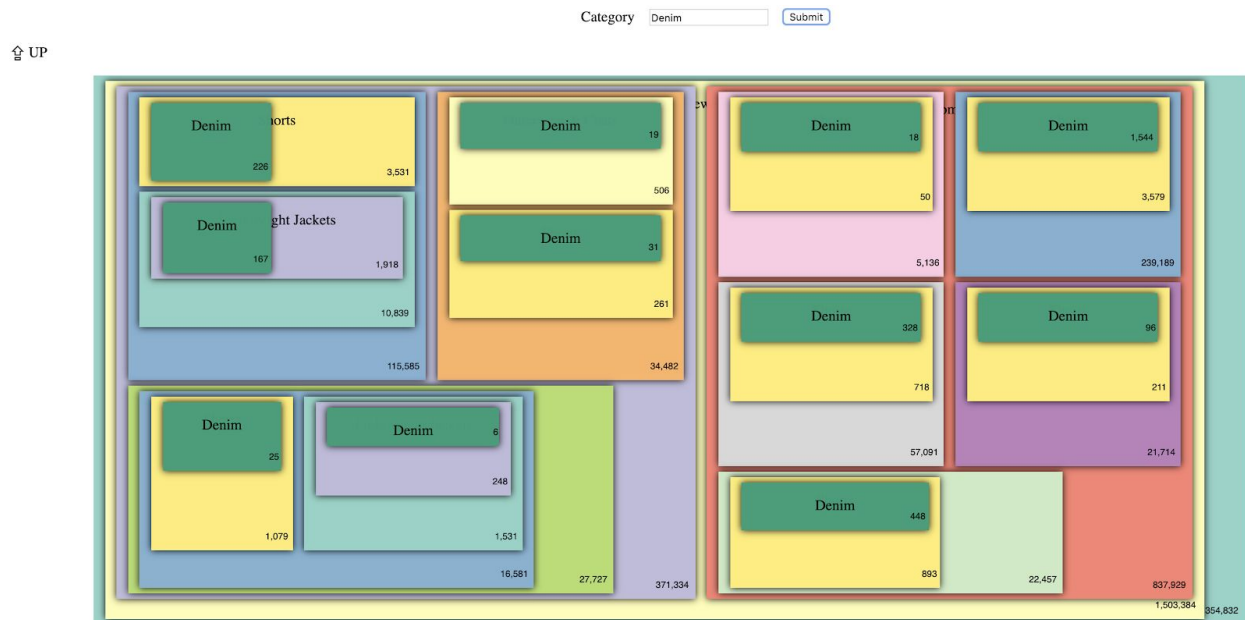


Figure 7: Result after searching for the category 'Denim'

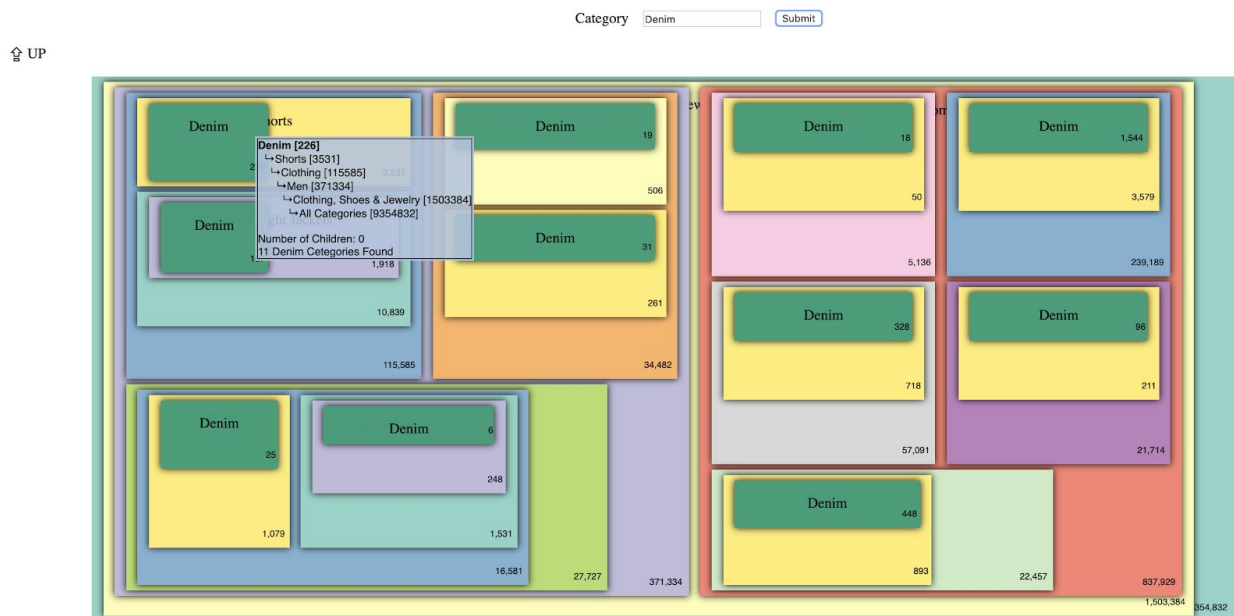


Figure 8: When the user hovers over the Denim Category that he wants, the tooltip shows the hierarchy and the number of items present in each category

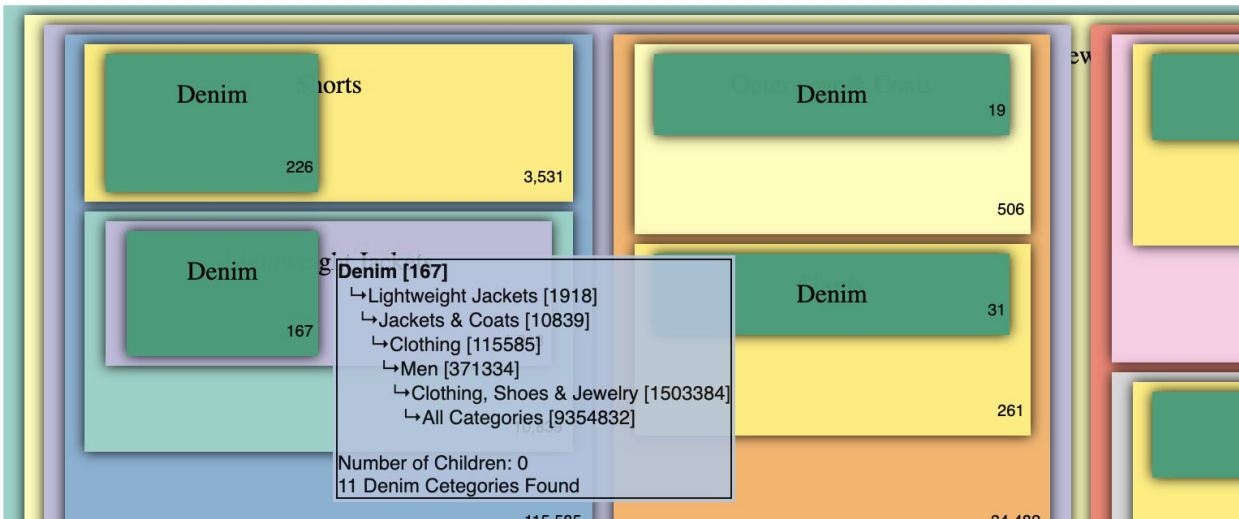


Figure 9: This picture shows the particular node named Denim is a subcategory of Men's Lightweight jackets

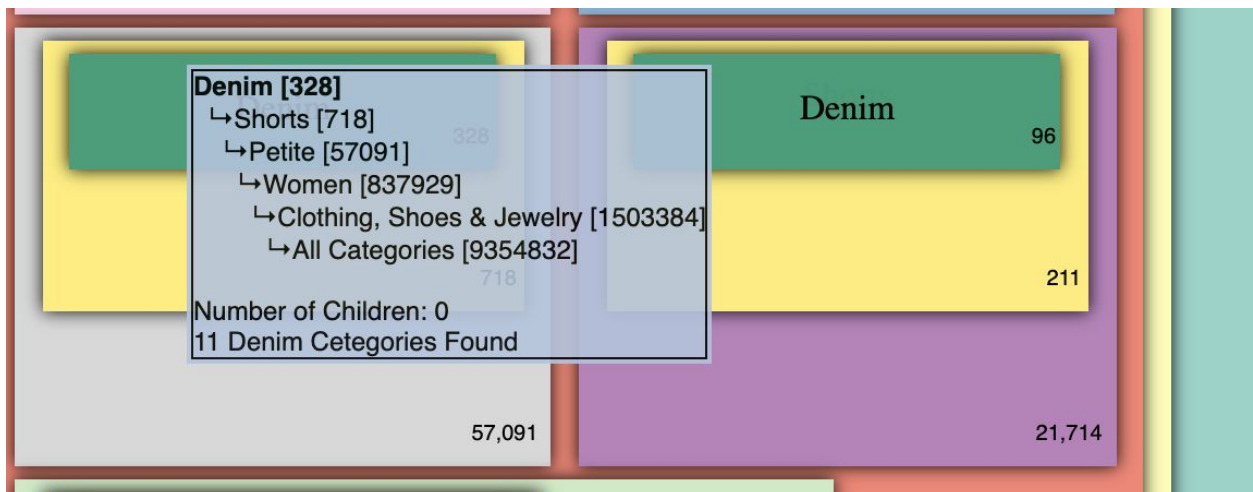


Figure 10: Similarly one can hover other nodes and get the hierarchical information. Example shows Denim belonging to Women's Petite Shorts



Figure 11: When the user clicks on the desired category from the multiple categories, one can see the total items in that category

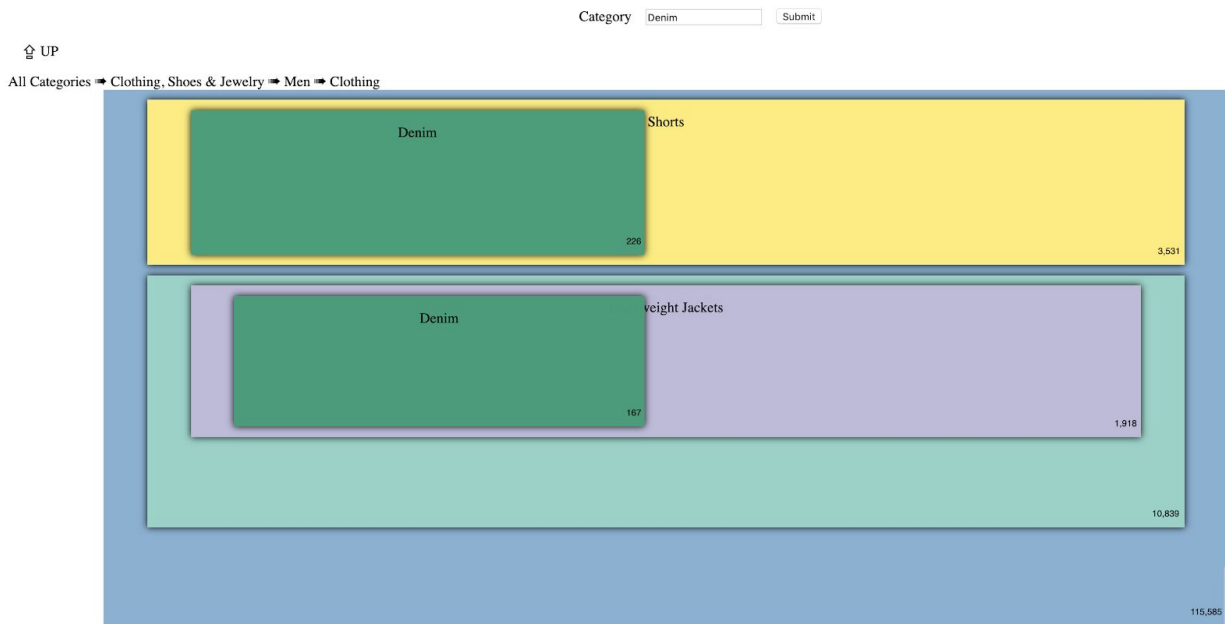


Figure 12: When the user zooms out from the category, one can see the hierarchy on the top left

Since the treemap clearly shows the nested structure present within the categories, one can view multiple categories based on a broad category.





The user can choose between specific boots used in snow sports such as cross-country skiing, downhill skiing, etc.

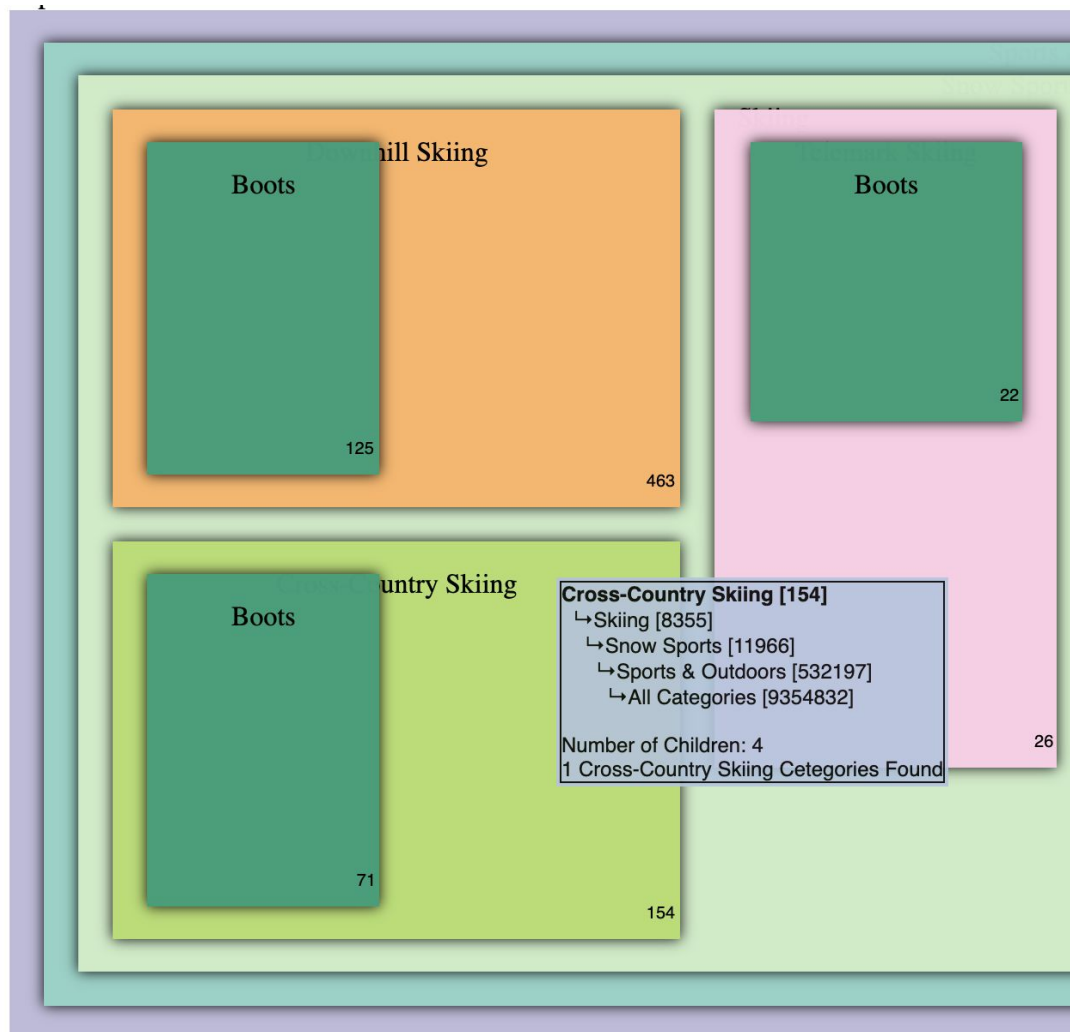


Figure 15: Boots belonging to Cross-Country Skiing

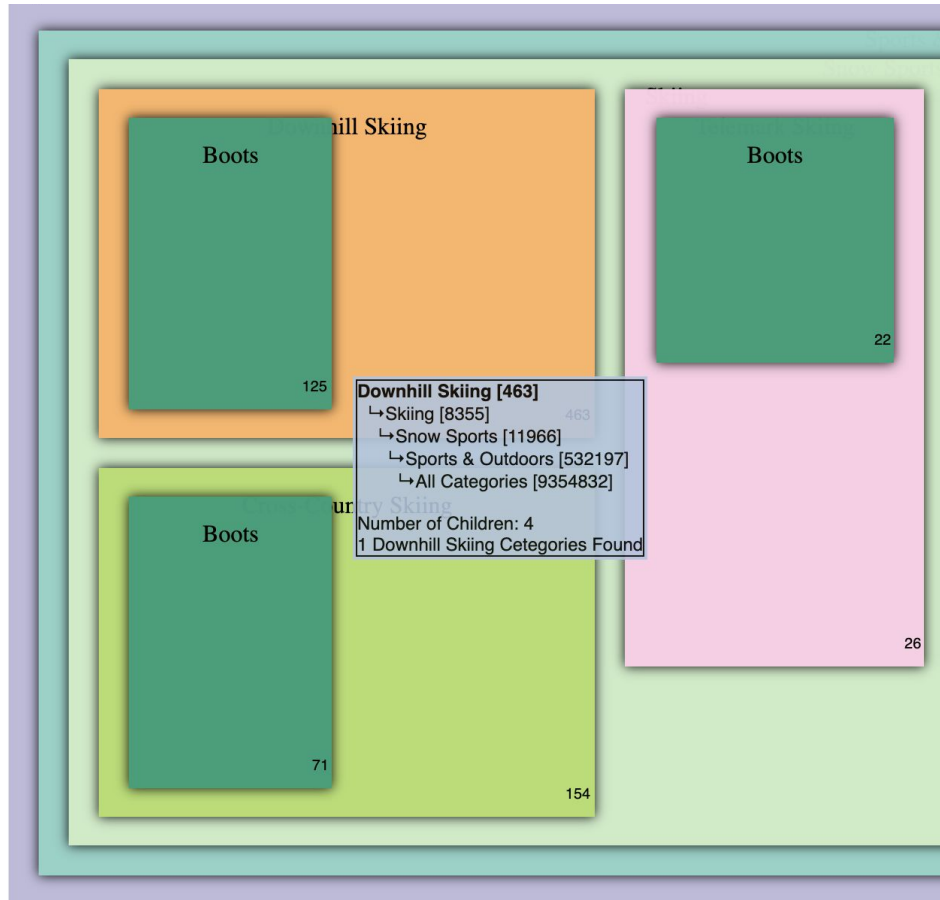


Figure 16: Boots belonging to Downhill Skiing

The containment between each category and the categories nested within it can be shown effectively using color as an encoding, for instance, the turquoise color indicates Snow Sports, which contains the pale green and purple boxes that indicate Snowboarding and Skiing respectively. And within the purple box(Skiing), it contains subcategories of Skiing denoted by the pink, bright green, lavender blue and orange boxes that indicate Alpine Touring, Telemark Skiing, Downhill Skiing and Cross-Country Skiing respectively, as shown in the below figure:

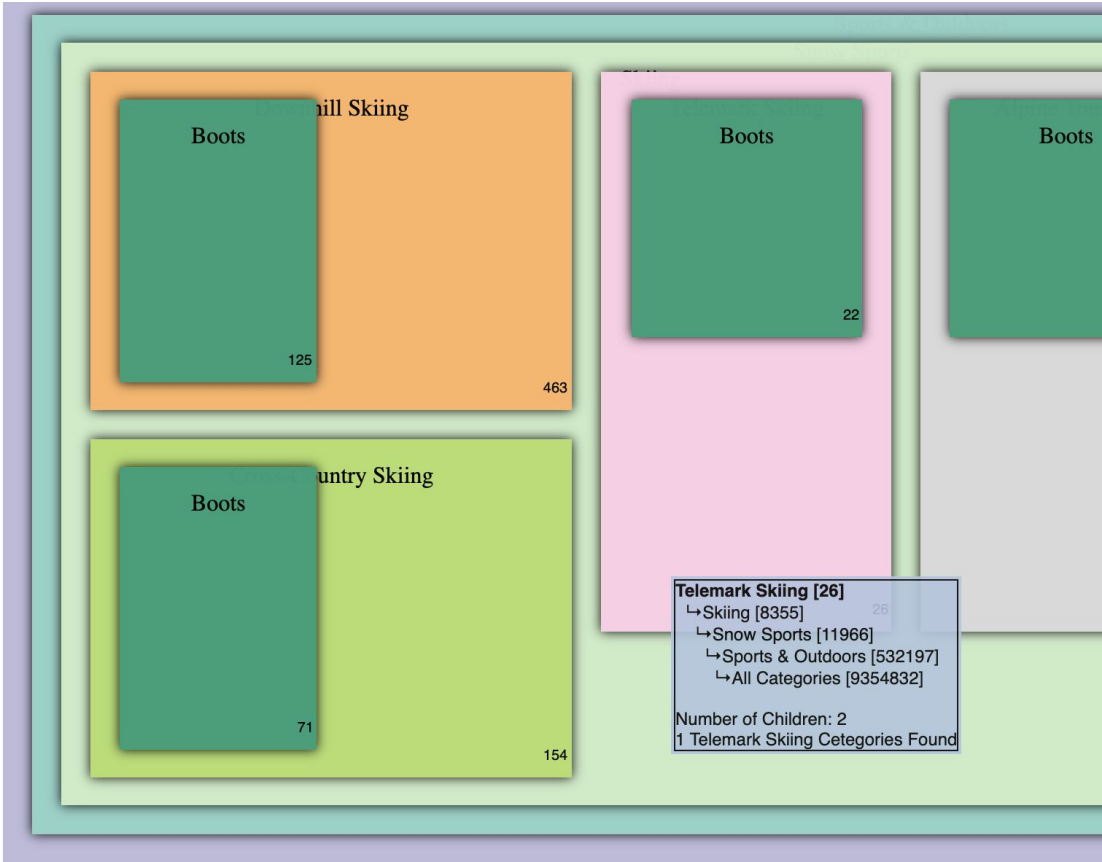


Figure 17: Categories showing containment

TASKS:

- ❑ If the goal was to determine which Denim category to choose, the tooltip shows details on the parent hierarchy of the Denim category, treated as a leaf node.

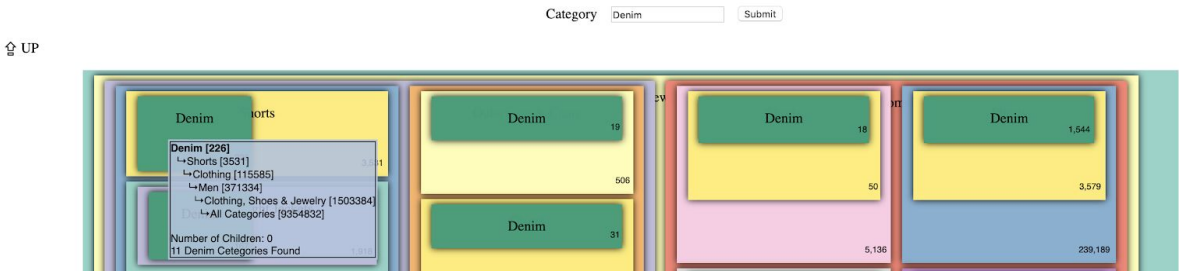


Figure 18: This figure shows Denim belonging to Men’s Clothing, as indicated by the details of the parent on the tooltip.

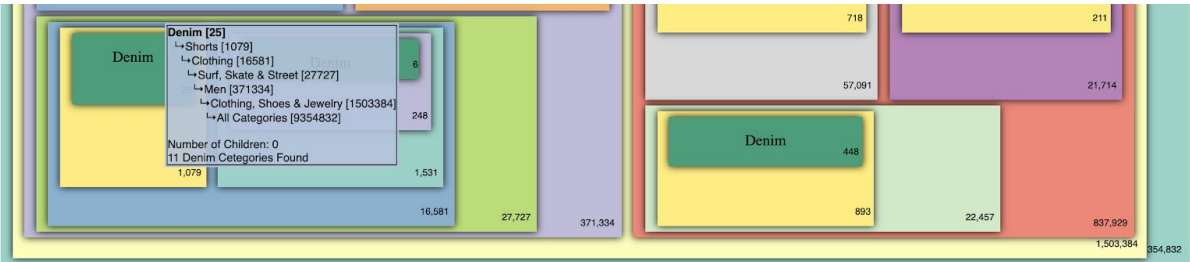


Figure 19: This figure shows Denim belonging to Surfing and Skating, as indicated by the details of the parent on the tooltip.



Figure 20: If the user clicks on the favorable category, this is how it appears using the zoom interaction, with the complete hierarchy information on the top.

If the user wants to trace back one category at a time, it can be done using the UP functionality on the top left that uses zoom-out interaction and helps the user examine same or different Denim products based on the broader categories like Jackets, Shorts, etc, as shown below:

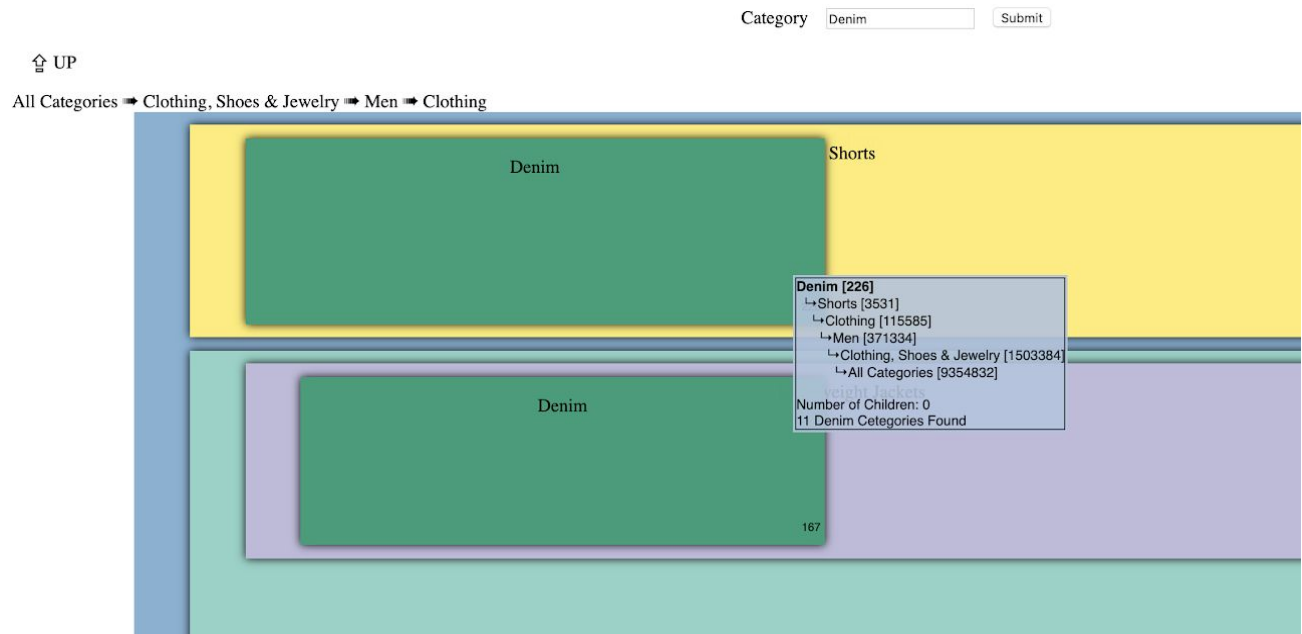


Figure 21: Denim category belonging to Mens' Shorts

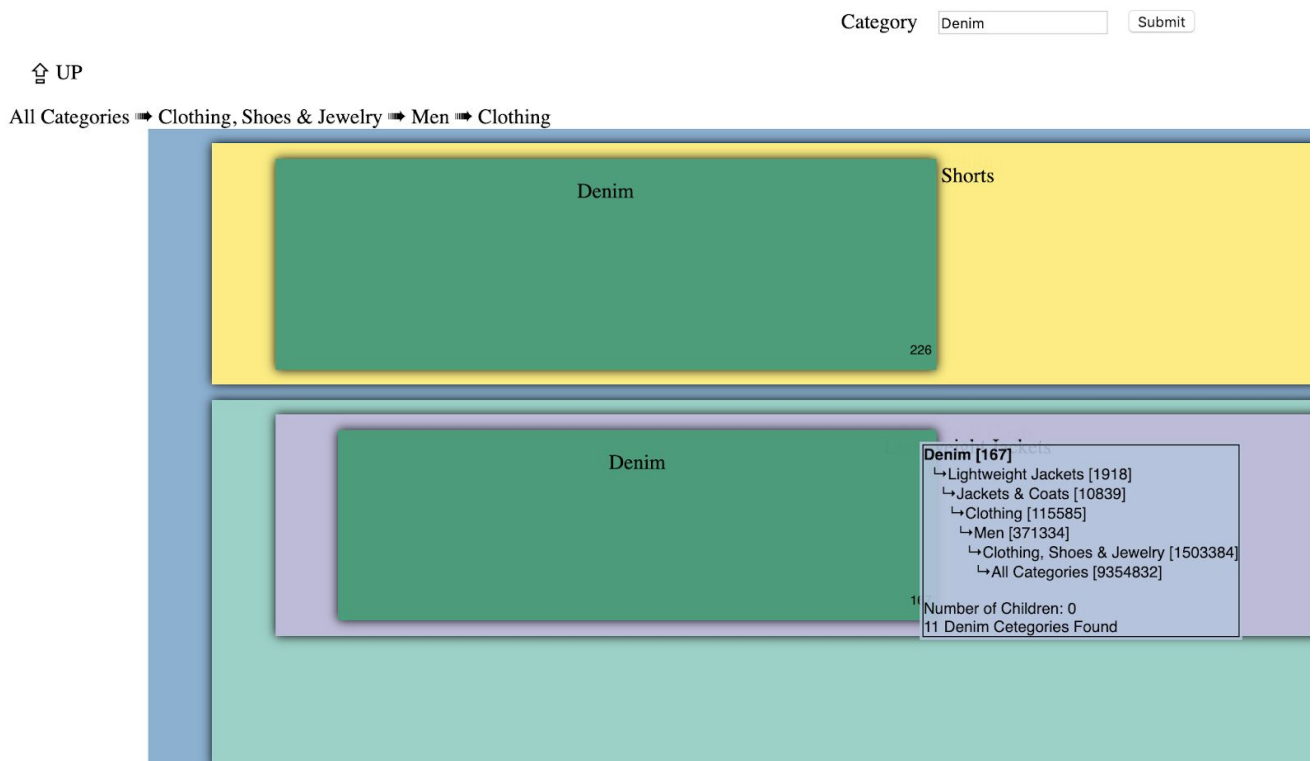


Figure 22: Denim category belonging to Mens' Lightweight Jackets

- ❑ If the goal was to determine the number of items present within each Denim Category, since Denim was the searched category, the details within each green block, indicates the total number of items. Also, the tooltip shows the total number of items of each Denim category.

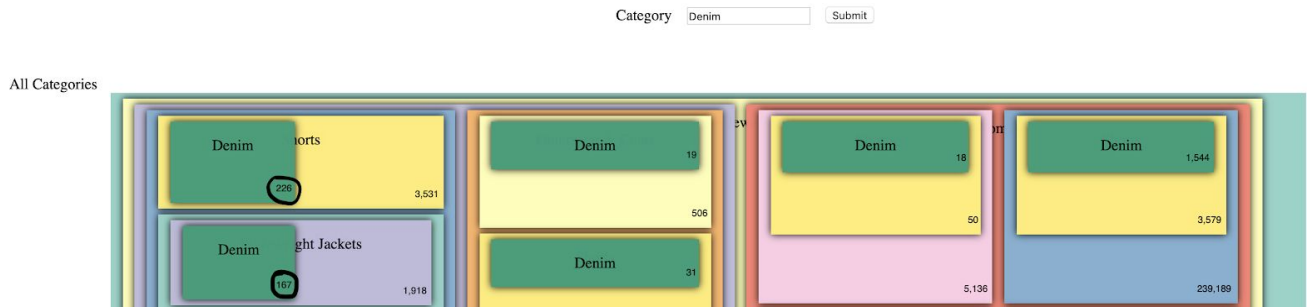


Figure 23: This figure shows the product count present in the Denim categories, as shown by the black circle above.

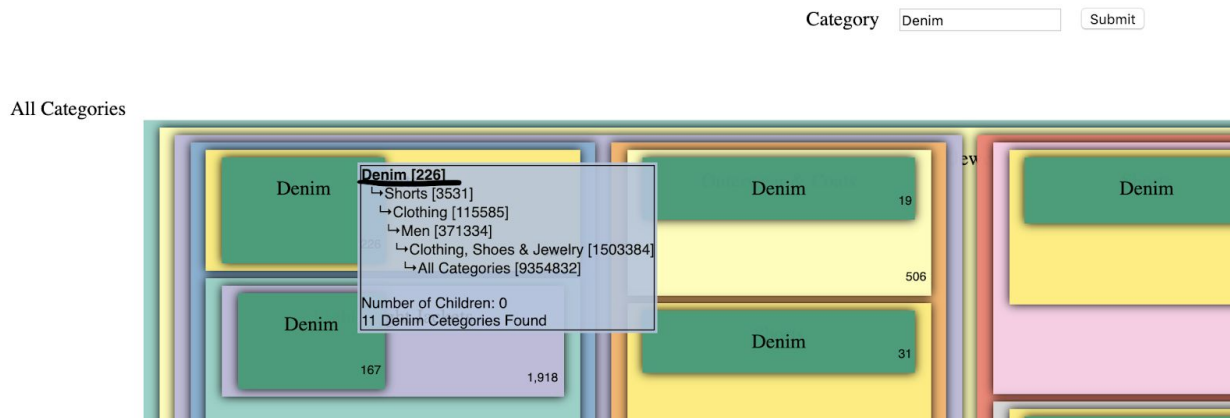


Figure 24: This figure shows the product count in the Denim category, as indicated by the tooltip below:

- ❑ If the goal was to determine the product count of the parents belonging to the category being searched for, the tooltip shows the count of products present within each category hierarchy.

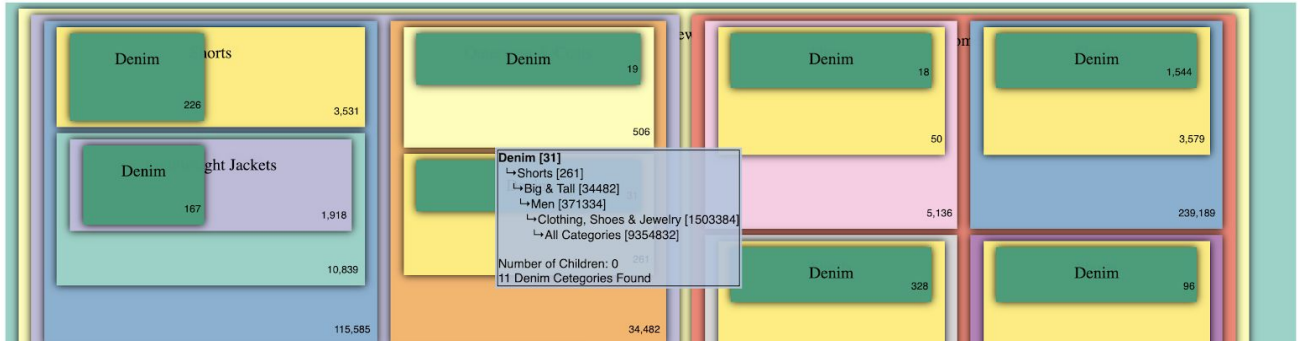


Figure 25: Tooltip showing the product counts

- ❑ If the goal was to find the total number of products present in the category searched, the user can hover over each category to find the number of products and sum it up accordingly.

For instance, if the user searches for Denim, the user can hover over every Denim category that pops out, as shown in Figure 25, and sum up the products from every Denim category and find out the total number of products of every Denim category. An even more effective approach would be to calculate the total number of products of the searched category at the back end and display it next to the search bar, to provide more insights.

- ❑ If the goal was to determine the proportion of the product counts of a category with respect to the broader categories, it can be easily done with the help of the tooltip that appears while hovering over the category being searched for.

For instance, based on the figure below, we can understand that there are 3119 boys' boots out of 16831 shoes.

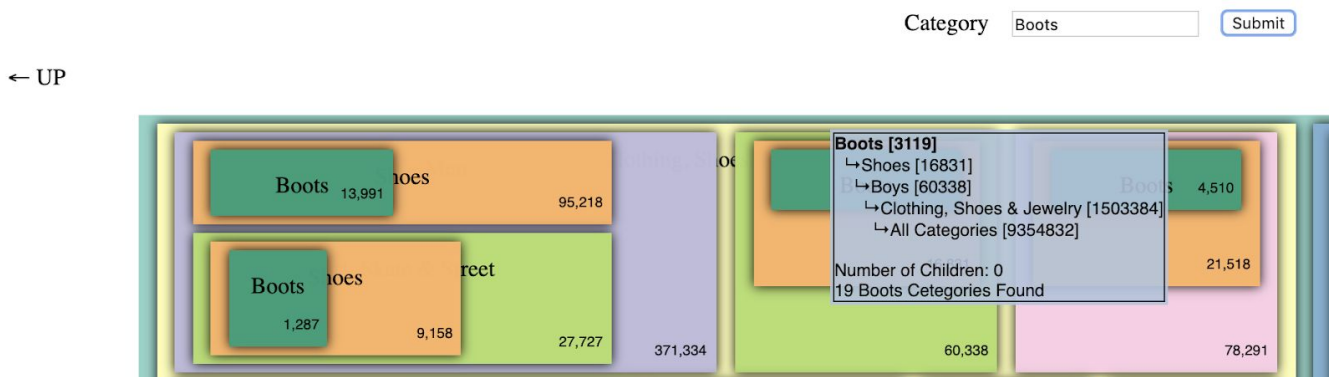


Figure 26: Tooltip showing boys' boots:



- ❑ If the goal was to find which categories of Boots are more/less than the other, then the tooltips information makes it easy because we can determine how many products are present in girls' boots vs. boys' boots. A better approach would be to assign different colors to girls' boots and boys' boots, but since we don't have so many colors to distinguish between categories, this serves as a limitation.

Here, we can see that there are 3119 boys' boots and 4510 girls' boots. So, we can clearly see that there are more girls' boots when compared to the boys' boots and the store has more stock of girls' boots. As shown above, Figure 26 shows us the tooltip of the boys' boots.

Category



Figure 27: Tooltip showing girls' boots:

- ❑ If the goal was to draw a comparison between the percentages of boots among girls' shoes and boys' shoes, it can be done effectively using the details provided in the tooltip, as shown in Figures 26 and 27.

We can see that there are 3119 boys' boots out of 16831 boys' shoes, which constitute of about 18.5% of the boys' shoes. Similarly, we can see that there are 4510 girls' boots out of 21518 girls' shoes, which constitute of about 20.9% of the girls' shoes.

- ❑ If the goal was to find out the count of how many categories are present based on what the user search, it can be easily determined with the details on the tooltip. For instance, if the user wants to know how many Boots categories are present, it might be overwhelming to count each one of them, though they pop out. Instead, if the user just hovers over any one Boots category, one can see information on the tooltip that 19 Boots categories were found, as shown in Figure 27.
- ❑ If the goal was to determine the number of children of the parental category, it can be done so, using the category count information present in the tooltip. In Figure 28 below, we see that baby girls' shoes represented by the orange box, has 6 children, out of which Boots is one child, that is contained within the green box. Here, only the required child gets displayed based on the user search, which makes the design highly scalable.

Since the boxes overlap each other, it makes it difficult to view that Shoes is the parental category of Boots and so on. It would be a better approach to use concentric circles with the leaf node present as the innermost circle and the parent nodes as outermost circles, thereby the names won't be hidden and the layout would be clear too.

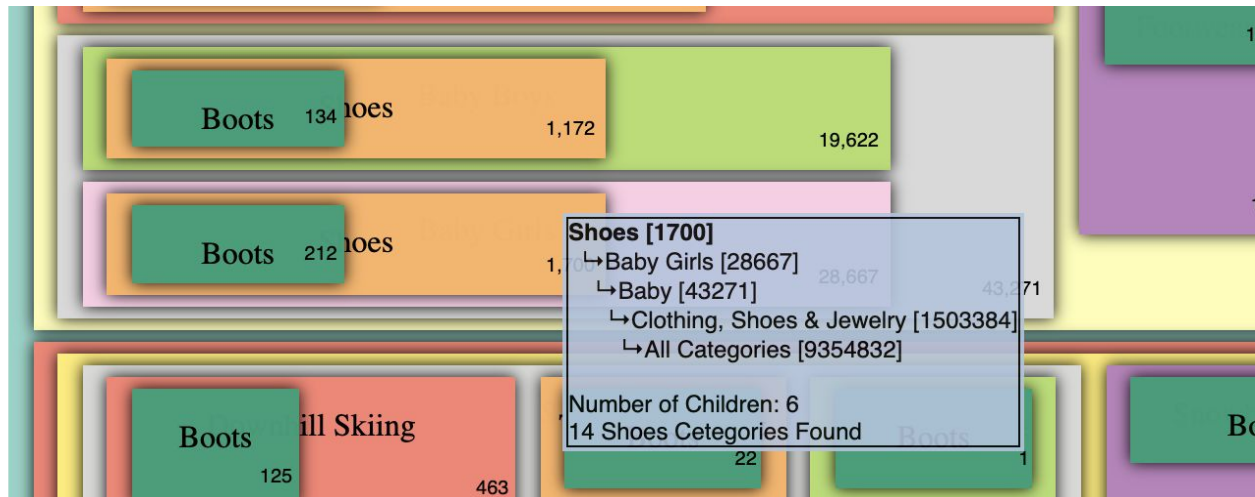


Figure 28: Tooltip showing information on baby girls' shoes:

- ❑ If the goal was to compare between categories, based on the user search, that belong to the same broad category, the treemap is highly effective as it shows the containment of children nodes within the parent node effectively, using size as an encoding. That is, parent nodes are indicated by larger boxes and child nodes are indicated by smaller boxes. Also, the color encoding is highly effective in distinguishing between the parent and child categories and shows the containment efficiently.

For instance, as shown in Figure 29 below, if one wants to choose between different Boots belonging to the Skiing category, one can do it easily with the help of this visualization, as it shows the type of each Boots on the tooltip. The size of the boxes, along with the category hierarchy information on the tooltip, make it easy to trace back the broad category of these boots, that is Skiing.

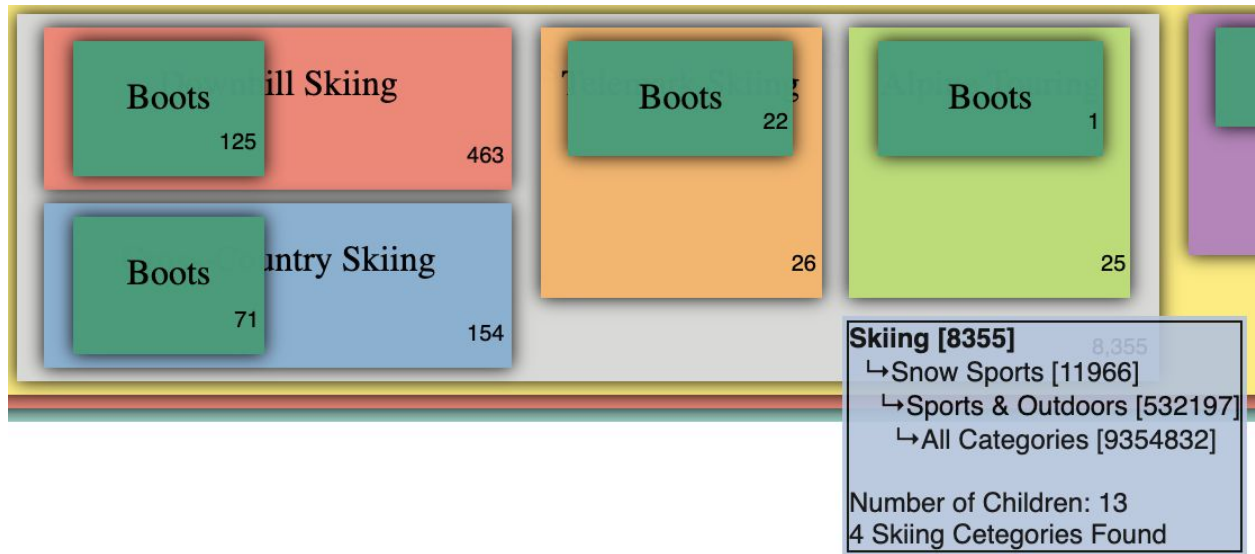


Figure 29: Boots belonging to Skiing category

The user can view and choose between Telemark Skiing boots and Downhill skiing boots easily, as shown in Figures 30 and 31 below:

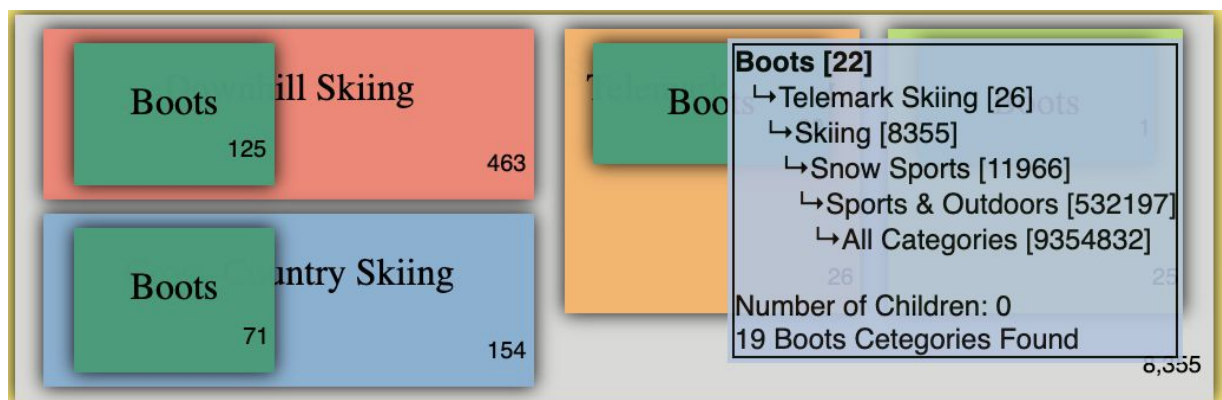


Figure 30: Boots belonging to Telemark Skiing

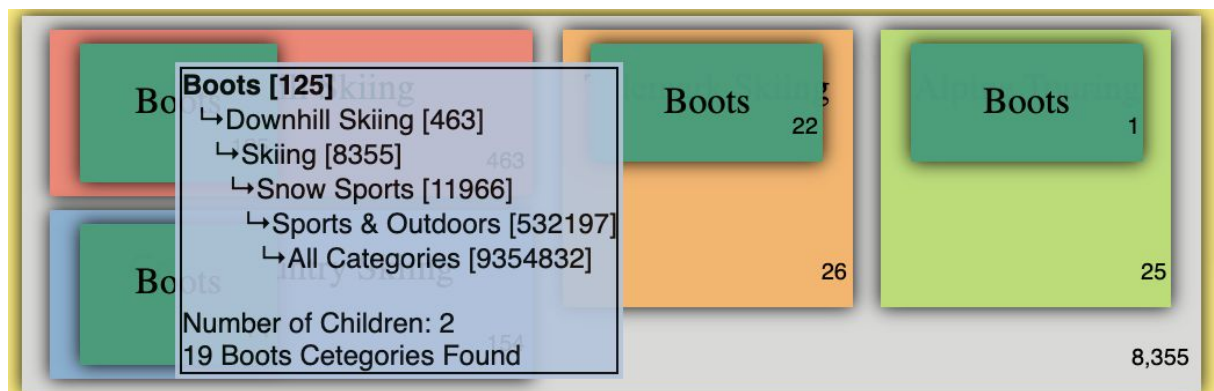


Figure 31: Boots belonging to Downhill Skiing

- ❑ If the goal was to view the product count, subtree product count and the number of children of the category based on the user's search, the 'Show all Sub-Categories' option that appears when the user right clicks on the category, can be used to display all the sub-categories, along with the details mentioned above, as shown in Figures 4 and 5.

## INSTRUCTIONS TO RUN THE PROGRAM:

Information about the programs

Languages and libraries used :

Python : numpy, pandas, Flask

Javascript L D3, d3-context-menu

HTML

CSS

1. Code Access : [https://github.com/chitcode/DataViz\\_Proj](https://github.com/chitcode/DataViz_Proj)
2. To run locally Python and mentioned python libraries need to be installed. A web browser is required to access it (Tool is tested only on Google Chrome browser). No additional software is required to install. Once the required python packages are installed, the following command has to be used from the root folder of the source code to run the tool.  
    > python implementation/viz\_serve.py
3. Web browser address : <http://127.0.0.1:5000/>
4. Additionally, the tool can also be accessed from the internet  
<https://cs765dc3.herokuapp.com/>