Università degli studi di Milano-Bicocca

Decision Models

Final Project

PACKING SANTA'S SLEIGH

Authors:
Marco Ferrario - 795203 - m.ferrario39@campus.unimib.it
Giorgio Ottolina - 838017 - g.ottolina1@campus.unimib.it

July 7, 2019

Abstract

This project tries to address the problem of packing Santa's sleigh using R programming language. The best allocation for the gifts on the sleigh needs to be found, taking into account the insertion order of the presents. In order to solve the problem an approach based on heuristics was chosen and applied.

1 Introduction

The problem is taken from a challenge published on Kaggle in 2013, in which it is required, as underlined by the title itself, to pack Santa's sleigh with all the presents that will be delivered on Christmas night. In order to find the best possible allocation for the gifts, the most important value that has to be calculated is the final height reached by the sleigh itself. This value needs to be the smallest possible. The sleigh is 1000x1000 with infinite vertical extent and goes from 1 to 1000 for the length and width. Gravity is supposed to be absent, so a present doesn't necessarily have to be put right above the previous one. They have to be allocated in such a way that gifts with a low ID are put at the top, while the ones with a high ID are put at the bottom of the sleigh (the latter ones should ideally be the last to be delivered, in order to make their retrieval inside the sleigh easier). It is possible to rotate the gifts in any direction given their parallelism and perpendicularity to the x-y-z axes. It is already clear that the problem requires a solution with high computational complexity. The strategies for solving NP-hard Optimization Problems can be in the form of Heuristic Methods:

- Identify a good feasible initial solution (not necessarily the optimal one);
- It must be efficient to deal with large scale problems;
- Usually, it is an iterative algorithm that at each iteration searches for a better solution;
- When it stops, the last solution is the best one with respect to all the solutions found so far;
- Usually it is a solution found ad hoc.

In fact, an ad hoc solution based on heuristics and iterations was indeed chosen and applied to solve the problem.

2 Datasets

The dataset is still today publicly available on the Kaggle platform and contains the total amount of gifts needed to fill the sleigh (1 million). For each present, the following numeric variables are given:

- ID
- The three x-y-x dimensions

There are no NA values so no manipulations were applied. A preliminary analysis proves that gifts' dimensions for each axis go from 2 to 250 and that their distribution doesn't depend on the ID. External data wasn't added in order to solve this optimization problem.

3 The Methodological Approach

First things first, visualizations were applied to analyze the data in deeper ways. Scatter plots of the axes dimensions tested on different samples of various sizes gave some interesting results. A particular pattern was identified. Each present's coordinate dimensions appeared almost uniform to one another.

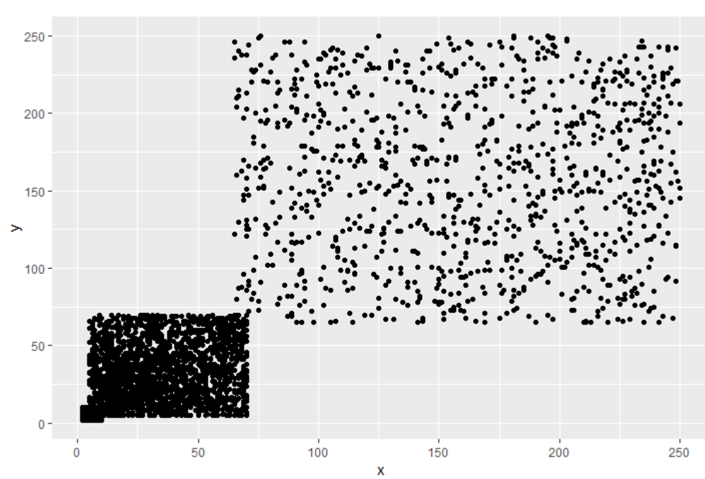

Fig 1 - Plot of present's x and y coordinates.

This observation led to the decision to divide the dataset into three main sub-groups: one for the small gifts, one for the medium ones and one for the big ones. In order to apply this division, the volume of each gift was calculated using the division based on quantiles. Another one of the key points of the analysis was how to orientate the gifts inside the sleigh. Several attempts were performed. The results obtained suggested that the gifts should be oriented keeping the longest side adjusted in the same direction. This operation was performed to preserve packs' uniformity. Talking about the sleigh's internal composition, its base was divided into three rectangular parts with the x-axis as width (333) and the y-axis as length (1000). This decision allowed the distribution of the gifts according to the aforementioned groups.

The gifts are inserted inside the sleigh along its x-axis, then along its y-axis, and finally along its z-axis. The gifts have initially been distributed along the x-axis until the end of the segment, keeping track of the y and z values. This procedure allows to calculate their maximum reached value, in order to start the new filling cycle starting from that point. After the end of the x-axis, a new cycle was started from the highest obtained y value. This procedure was performed until the maximum value of y (1000) was reached. The process has then been repeated for each one of the other groups until every rectangle's area was successfully filled.

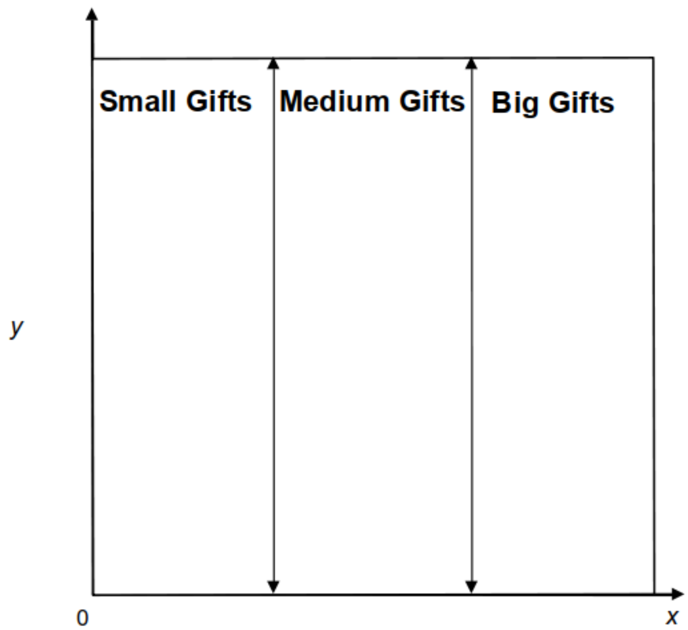

Fig 2 - X and Y axis division of the sleigh.

Each layer of gifts has its opposite one located above itself starting from the maximum value of z obtained by filling the area in analysis. The big presents are located above the small ones, the small gifts are above the big ones and finally the medium packs are above the other medium ones. The end of the first macro layer is reached when the final value of z is obtained, and at this point the second layer starts (in the same way, but assigning the height to the maximum z value). Initially, all the gifts were inserted dividing their total amount into small, medium and big groups only once. It was clear that in this way the small presents would end very quickly because of their smaller dimensions. Then it was decided to insert the packs in groups of 6000 elements, where this kind of division must have been made after any filling cycle was completed along the x and y dimensions. While trying to obtain the best solution possible, the decision to run the code by iterating the loop for the 6 possible combinations in terms of packs' orientation (combining X, Y and Z together) was taken. This choice's goal was to find and finally select the combination with the minimum height.

Adopting this procedure, it can be seen that the maximum score depended on the number of groups of gifts used to divide the presents into small, medium and big. If the observed samples did have more than 1000 elements, the height value would be smaller but the second metric would have very high values. This is due to the fact that the sides have big dimensions and are not immediately filled, while the remaining gifts are reinserted into the dataset (and so are positioned late in time). It must also be noted that by keeping the sample too small, the result would end up being the opposite one. If the group was made of a few presents, the space on the x and y-axis could be used not at its fullest when the allocation ended too soon (small gifts in particular). A medium value, which had not been found yet until this point, now needed to be calculated in order to balance the two aforementioned metrics.

4 Results and Evaluation

The obtained results are shown in this section. It must be taken into consideration that the evaluation metrics of the final score concern both the height reached by the pack which sits in the highest sleigh's position (multiplied by 2) and the insertion order of the gifts. Grouping the presents by samples of 6000 elements, the obtained scores for each section were:

X-Y-Z	**X-Z-Y**	**Y-X-Z**	**Y-Z-X**	**Z-X-Y**	**Z-Y-X**
1056105	1236146	1050789	1196627	1525841	1442467

Analyzing samples of 500 gifts, the results were:

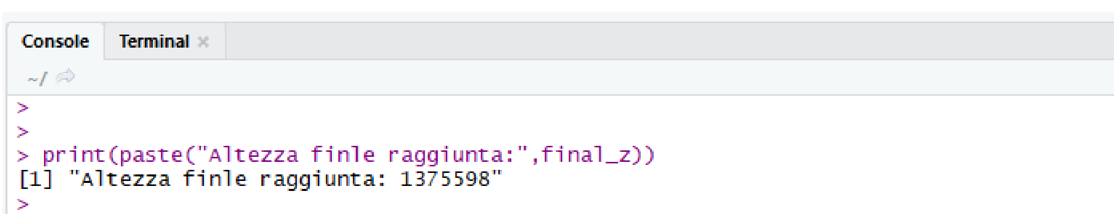

Fig 3 - Results with groups of 500 gifts.

In order to process this huge amount of data, a key point to consider was the execution time of the code. The processing time taking into account the

IDs of the packs and their maximum height reached and using vectorial data structures was about 30 minutes. Anyway, the project asked to keep track of all the coordinates of each inserted gift, so using a matrix as structure in which to save the data the execution time was significantly slower than the one before (16 hours).

	PresentId	x1	y1	z1	x2	y2	z2	x3	y3	z3	x4	y4
999990	8	334	647	1377371	334	789	1377371	540	647	1377371	540	789
999989	22	334	462	1377371	334	647	1377371	529	462	1377371	529	647
999988	91	334	300	1377371	334	462	1377371	519	300	1377371	519	462
999987	270	334	165	1377371	334	300	1377371	568	165	1377371	568	300
999986	292	334	1	1377371	334	165	1377371	550	1	1377371	550	165
999985	68	1	788	1377371	1	905	1377371	198	788	1377371	198	905
999984	1	874	715	1377115	874	717	1377115	877	715	1377115	877	717
999983	2	667	715	1377115	667	788	1377115	874	715	1377115	874	788
999982	3	667	637	1377115	667	715	1377115	827	637	1377115	827	715
999981	4	962	517	1377115	962	520	1377115	970	517	1377115	970	520
999980	5	953	517	1377115	953	526	1377115	962	517	1377115	962	526
999979	6	783	517	1377115	783	637	1377115	953	517	1377115	953	637
999978	7	667	517	1377115	667	608	1377115	783	517	1377115	783	608
999977	51	800	392	1377115	800	495	1377115	973	392	1377115	973	495
999976	66	667	392	1377115	667	517	1377115	800	392	1377115	800	517
999975	94	810	293	1377115	810	392	1377115	986	293	1377115	986	392
999974	100	667	293	1377115	667	391	1377115	810	293	1377115	810	391

Fig 4 - Part of the dataset with final coordinates of the packages.

5 Discussion

Trying to minimize the height value, the maximum value reached was 1050789. Trying all the aforementioned combinations on the entire dataset, it was clear that the "Y-X-Z" one was the best. Talking about the initial configuration, the Z axis remained the biggest dimension, the Y axis was the smallest one and the X axis was the medium dimension. After several approaches, 500 ended up being the optimal number to choose regarding the gifts' division. Doing so, the final height (still using the "Y-X-Z" combination) resulted equal to 1375598; computational time and presents' order were in line with expectations. Using heuristics and a high number of iterations in order to find the best solution possible, one way to improve the obtained scores could be keeping gravity into account too. This could lead to limit wasting of space. Each present would be positioned exactly in contact with the one in the layer below leaving no unused space on the sleigh. Finally, by incrementing the computational complexity of the code, gifts could be rotated on the x-y-z axes even before each insertion or at the end of each layer's completion. This would allow to order the gifts in the best way possible.

6 Conclusion

The project's scores were obtained by trying to minimize the sleigh's height the most, and then by balancing the second evaluation criterion too. By creating samples made of 500 presents, a moderate height value could indeed be found and calculated. Finally, the distribution of the packs is consistent with expectations, since the gifts with higher delivery priority are located at the top of the sleigh while the ones with lower priority are located at its bottom.