Kennesaw State University
Center for Statistics and Analytical Research

Corrugated Box Optimization Project

Developed for the Industrial Packing Corporation

June 13, 2017

Shashank Hebbar
PhD student in Analytics and Data Science
Kennesaw State University

Problem Description

The Industrial Packing Corporation (IPC) is interested in "optimizing" the configuration of boxes to accommodate a series of unique items which could be shipped in different combinations. For the purposes of the current project, the established number of boxes is four and costs associated with the production of the boxes are not relevant. The primary metric of "success" is the minimization of the excess cumulative cubic space across the boxes shipped.

IPC has forwarded several datasets for potential consideration for this project. After several phone and email communications, it was determined that the following variables will be of particular relevance:

- ✓ Carton_ID
- ✓ Carton_Type
- ✓ Item_Number
- ✓ Quantity
- ✓ Length, Width, Height of Items
- ✓ Conversion Factor of Items
- ✓ Length, Width, Height of existing boxes

Methodology

A master dataset (Table 1) was provided with fields for carton type and item number combination. The item table was reduced to account only for unique dimensions. The container table included dimensions of the existing 24 carton box sizes – box sizes which were recorded as having "zero" volume were removed. The final data set included 13 unique carton box sizes. Frequency counts of these 13 box configurations are provided below:

CARTON_ID	CARTON_TYPE	ITEM_NUMBER	QUANTITY	SHIP_TO_ZIP
C200000004885217	PP2704	7380000	2	04005-9496
C200000004885217	PP2704	7380000	2	04005-9496
C200000005022265	PP2704	10114	24	19390
C200000005022265	PP2704	0165SI16	12	19390
C200000005022265	PP2704	0166L18	36	19390
C200000005022265	PP2704	72211	10	19390
C200000005022265	PP2704	72225	10	19390
C200000005022265	PP2704	33616	10	19390
C200000005022265	PP2704	996101	3	19390
CSP3726821658013	PP2166	119316	12	32901-3278
CSP3726821658013	PP2166	119316	12	32901-3278
CSP3726921658018	PP2166	119316	12	94304-2200
CSP3726921658018	PP2166	119316	12	94304-2200
CSP3727021658022	PP2166	119316	12	92408
CSP3727021658022	PP2166	119316	12	92408
C200000005279958	PP2704	FEL12060	1	92691-6426
C200000005279958	PP2704	FEL12120	1	92691-6426
C200000005279958	PP2704	FEL14060	1	92691-6426
C200000005279958	PP2704	FEL14120	1	92691-6426
C200000005279959	PP2704	U81503H30	1	60160-1692

Table 1: Master data set

The list of container dimensions is given below:

CONTAINER_TYPE	DESCRIPTION	WEIGHT	VOLUME	FULL_LENGTH	FULL_WIDTH	FULL_HEIGHT
PP2704	PP2704	1	100	23	13	6
V3	MDVV3	1	100	26	23	4
V4	MDVV4	1	100	26	23	8
V5	MDVV5	1	100	26	23	16
PP2115	PP2115	2	100	23	13	16
PP2166	PP2166	1	100	23	10	8
PP2699	PP2699	2	100	23	18	16
PRBOX	PRBOX	1	100	25	20	20
PK0300189	PK0300189	2	100	36	12	9
CB300005	Longtriangle	2	90	62	8	7
STENTXS	STENTXS	3	0	48	13	6
PP2485	Shorttriangle	2	100	48	8	7
FXLG	FXLG	1	0	12	18	3

Table 2: Container table

The process to optimize the configuration of four boxes followed three steps:

(1) Estimate the current wasted volume using these 13 box configurations
(2) Estimate the projected wasted volume using the "best" subset of four boxes from the current 13 available boxes.
(3) Estimate the projected wasted volume by optimizing four box configurations which may or may not be currently available.

The first step was to estimate the wasted cubic space based upon the current use of these 13 boxes.

This was done by calculating the volume of the item and subtracting this volume from the available cubic space available in the shipping box:

The wasted space was calculated for each row (i.e., the volume of the item was subtracted from the volume of the container multiplied by the quantity packed).

Following this logic, the total volume of wasted space across 114,697 orders and the current 13 boxes was estimated to be 420,679,698 cubic inches.

This wasted volume represents the "base" case that will inform the final solution.

The next step was to estimate the projected wasted volume using the "best" subset of four boxes from the current 13 available boxes.

The optimization function generated a permutation matrix of 4 items from 13 existing container items. This matrix was sequenced through two "*for*" loops calculating for wasted space for each set of 4 containers sizes whose values were stored in a list. Conditions were given to avoid negative values. The minimum value of that list is returned as the "best set".

For example, for the of set of (2,10,4,6) the total wasted space was calculated by subtracting the volume of each item from each of those boxes provided they could accommodate the item. The largest length of 48 was replaced by 25 for erector spec constraints.

The container ID and dimensions are provided below.

	Length	Breadth	Height
Set 1	23	13	6
Set 2	26	23	4
Set 3	23	10	8
Set 4	12	18	3

Table 3: Reduced set

The total wasted space using this four box solution is 152,203,592 cubic inches which is almost 40% reduction from the base case.

The final step in the project was to estimate the total wasted cubic space by optimizing four box configurations which may or may not be currently available.

The two key parameters which guided this step were the smallest and largest items by dimension from table order above:

(1) The smallest current item is with dimensions of 11x8x3.
(2) The largest item to be shipped had dimensions of 48x13x6. This had to be accommodated.

To determine this optimized configuration, a clustering analysis was executed to develop a new set of carton boxes that included all the item sets. The k means method was used, with k=4 clusters.

The k means method computes the three-dimensional Euclidean distance between the centroid of that cluster and the individual dimensions, and assigns them to the closet cluster. Similar dimension types are grouped together to form a cluster.

Adjustments:
Since the clustering process does not take properly take into account the outliers as well as the upper and lower bounds of dimensions. Using the dimensions given by kmeans as a guide , I made some changes to the dimensions by looking at the maximum and minimum sizes. The methodology for determined by comparing the centroid table given by the clustering procedure, and the frequency table(called freqtable) of top 10 items. For length , each length from the centroid table is subtracted from the length in the top 10 frequency table. The index with the least difference is taken as the new length from the frequency table. In case of width, due to the wide variation in its value, I multiplied the width from the output set by a factor of 2 to encompass the variation. For height , the dimension set with the highest frequency was assigned a height of 6 ,and the set with the lowest frequency a height of 18. This conforms to the erector spec limits as well ensure all the variation is captured.

A visualization of this solution is provided in Figure 1 below:

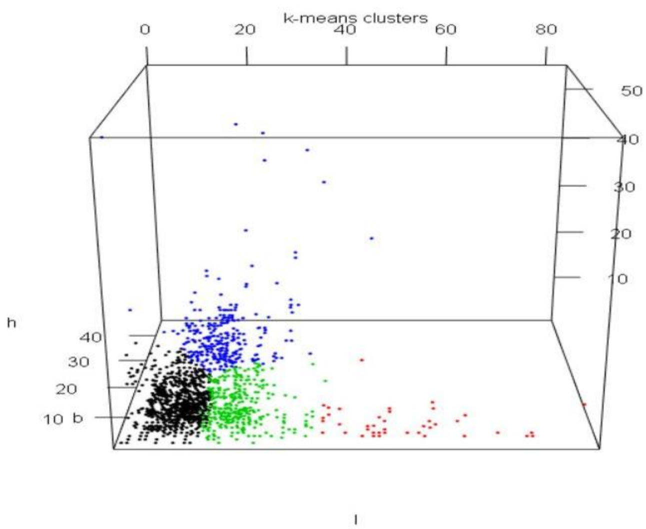

Figure 1: Clustering Distribution

Informed by this result, we determined that the optimal solution set of four boxes is:

	Length	Breadth	Height
Set 1	24	10	6
Set 2	19	14	18
Set 3	13	20	6
Set 4	13	20	6

Table 4: Best set

This optimal solution of boxes generates 424,187,186 cubic inches of wasted space, which is a reasonable result given the reduction in number of sizes. For this data, the script generates three dimension sets as opposed to four. This is due to the erector spec constraints.

Summary of Results

The current project sought to minimize the wasted cubic space associated with shipping 114,697 orders. The process to execute this project required the estimation of the wasted space under the current 13 box option, the optimal four box subset and the theoretical four box subset. The reduced set had the least wasted space but was not included here because it didn't conform to the erector specs. These results are summarized below in Figure 2 and Table 5 gives the percentage distribution of these dimensions on the order table.

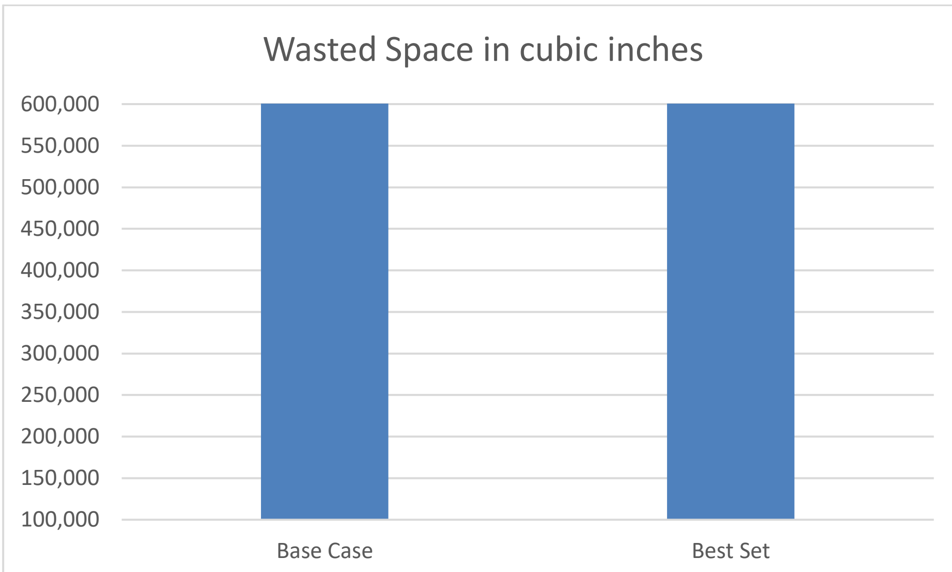

Figure 2: Total wasted space (volume) by solution

Length	Width	Height	Percentage%
24	10	6	35.9
19	14	18	27.77
13	20	6	36.31

Table 5: Percentage distribution