Product Movement Categorized Analysis

Andhyka Cakrabuana Adhitama

Business Overview

One of the common problem on supply chain business is **storage efficiency**. In some experience in supply chain, most likely we would see there is a unnecessaries process in every warehouse that would increase the cycle time for manpower to fulfill the order. For example, best selling product placed on the farthest shelf and resulting the picker manpower should walk around for few second until they can deliver the goods to the packing area. With proper **SKUs movement categorized**, we can reduce common issues like this. This analysis tried to categorized the movement of SKUs, not only based on how much the SKUs ordered, but also how fast the retention rate and how big the monetary value we can save or gain.

What’s benefit of using Categorized SKU Movement?

BEST PRODUCT IDENTIFICATION

Using categorized movement, it will help us to identify characteristic of the SKUs. Not only best-selling on quantity, but also how fast the customer repeat their buying on SKUs, and also how much the potential value of the SKUs.

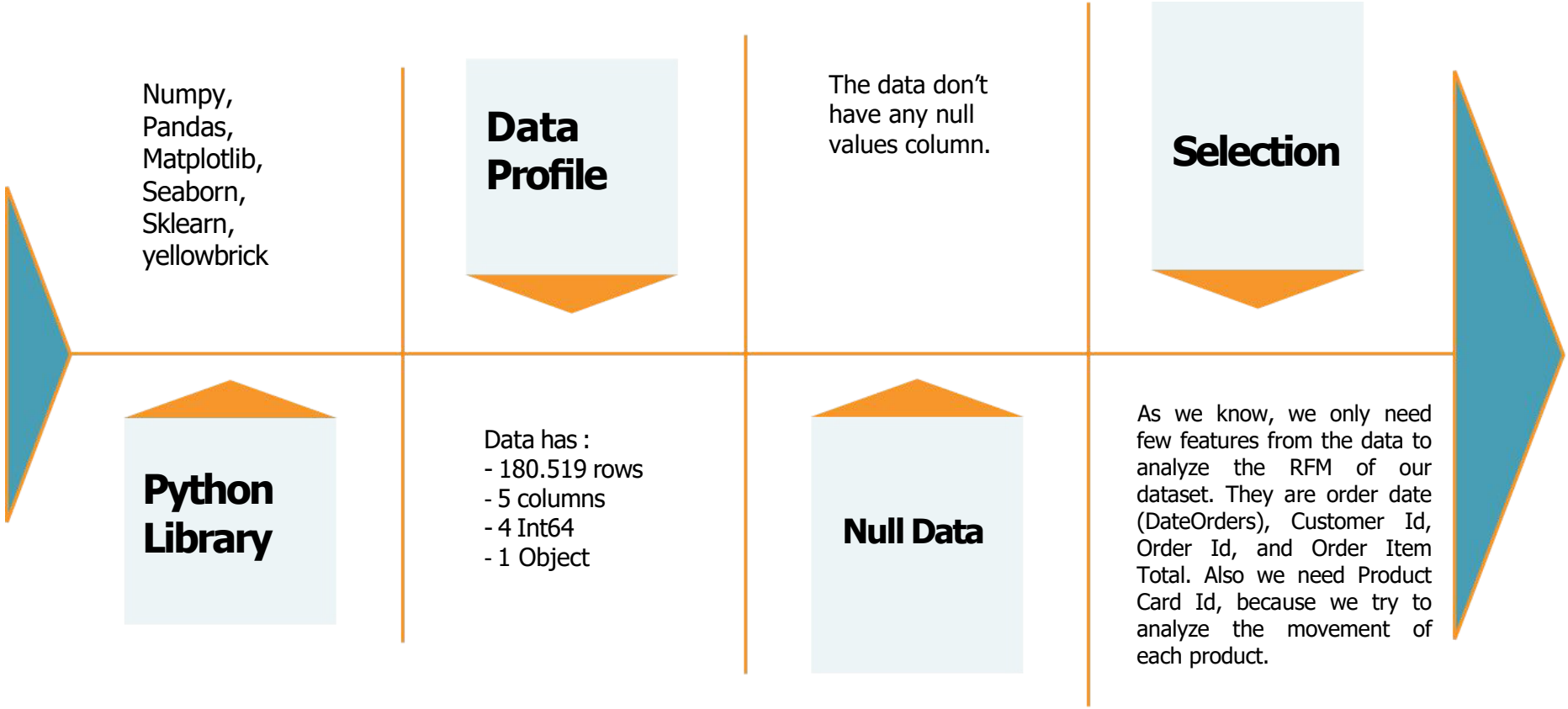
EFFICIENT WAREHOUSING

After categorizing the SKUs, we can formulate best option in term storage for the SKUs. For example we can put the FAST movement SKUs to the nearest shelf from the packaging area.

MORE FASTER ORDER FULFILLMENT

The more efficient storaging system = the more less lead time of processing. The categorized SKUs movement will cut-off the lead time most of unnecessary process and will impact more order can be fulfilled at the same times.

DataCo Smart Supply Chain for Big Data Analytics (only selected features)



Lets Engineering our dataset!

01

Rename few columns to make analysis easier and change date order type into datetime64 format.

02

Make Recency function. The column makes using newest invoice date – invoice date of each order.

03

Make Recency, Frequency, and Monetary column. We use groupby function with recency function for Recency, count invoice for Frequency, and sum for Monetary.

04

For K-Means clustering, we use Standarization for Recency and Normalization for both Frequency and Monetary.

Let’s check RFM Scoring!

	SKU	Recency	Frequency	Monetary	R	F	M	rfm_score
0	19	143	64	7256.879885	2	1	1	211
1	24	145	74	16653.069852	2	2	2	222
2	35	142	65	9386.549972	2	1	1	211
3	37	284	262	24626.510213	2	2	2	222
4	44	285	305	50354.210668	1	3	3	133
...	...	...	...	...	...	...	...	...
109	1355	29	484	231765.460292	3	3	3	333
110	1356	27	362	95358.110288	3	3	3	333
111	1357	27	208	39365.709993	3	2	3	323
112	1358	25	434	101687.289866	3	3	3	333
113	1359	21	492	37318.299847	3	3	3	333

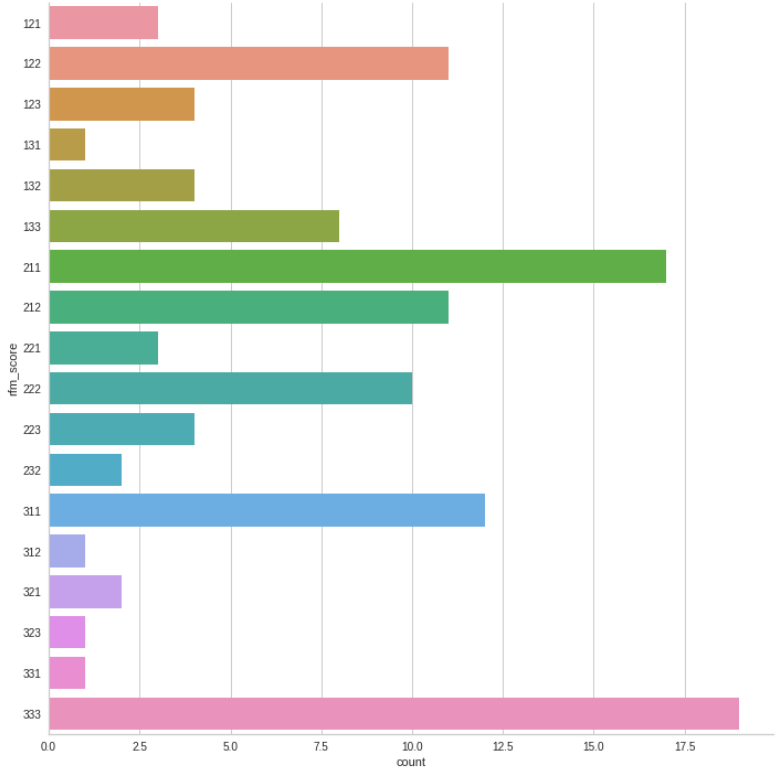
114 rows × 8 columns

The RFM analysis is method to create segment of feature based on quantile criteria. For this chance, we tried to segment the SKUs into 3 segment based on scoring from R, F, and M.

For the Frequency and Monetary, more higher the values will makes higher score (3). Different case for the Recency, the lower value will makes higher score (3). The difference is because in terms of Recency, the lower value is because the range between SKUs first buy and the next buy is more shorter. That means the customer repeat to buy the SKUs more often than the other SKUs.

In terms of recency, we get rid the SKUs with 0 Recency value, because this means SKUs just bought once in the time of analysis so it will be not valid to analyze. And we get rid the Monetary value 0 or minus since it didn’t give any value for the company and we just could put these SKUs on the farthest shelf.

RFM Characteristic per SKU

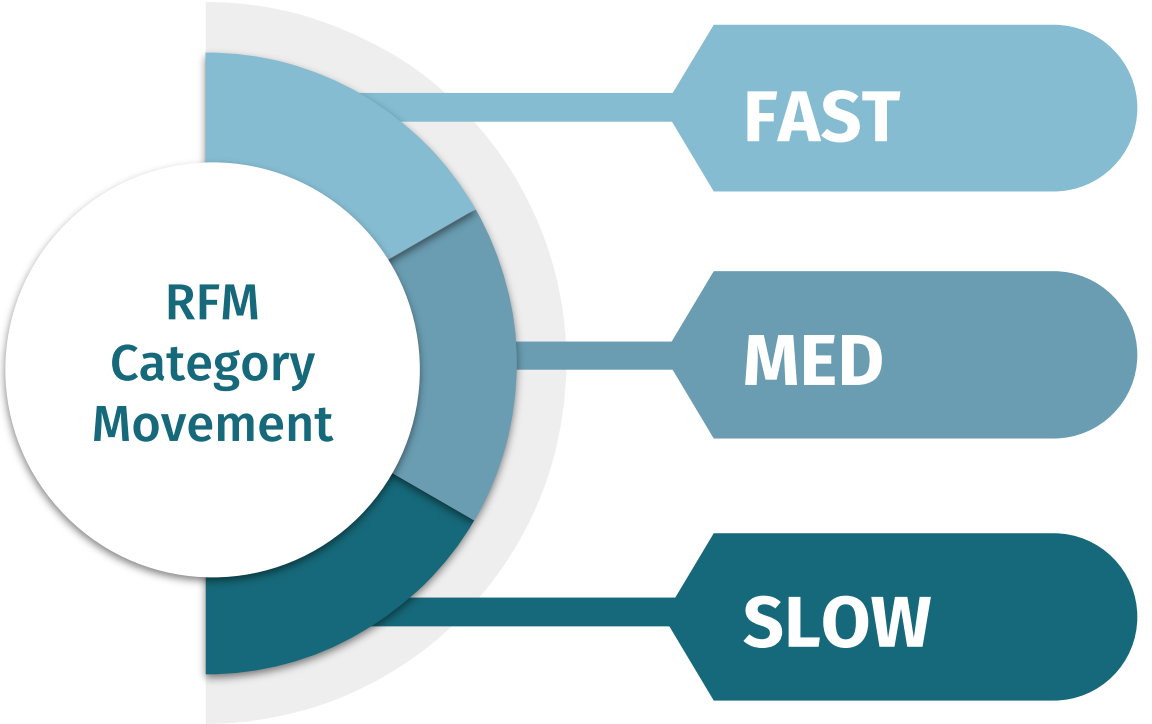


As we can see from the plot, there different characteristic in R,F,M of all SKUs. Sometimes the SKUs have one or two category with high score, while on the same time have low score on another category. After we create the conditions below :

- **FAST** for SKU with RFM score 333, 233, 323, 332
- **MED** for SKU with RFM score 322, 232, 223, 331, 313, 133
- **SLOW** for SKU with RFM score asidae from FAST and MED

We get 20 FAST movement category SKUs, 15 MED movement category SKUs, and 79 SLOW movement category SKUs.

RFM Output

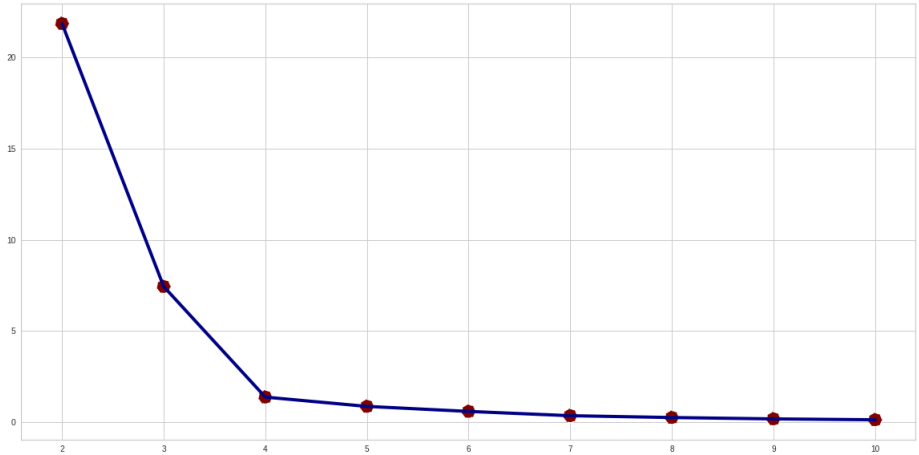


FAST for SKU with RFM score 333, 233, 323, 332

MED for SKU with RFM score 322, 232, 223, 331, 313, 133

SLOW for SKU with RFM score aside from FAST and MED

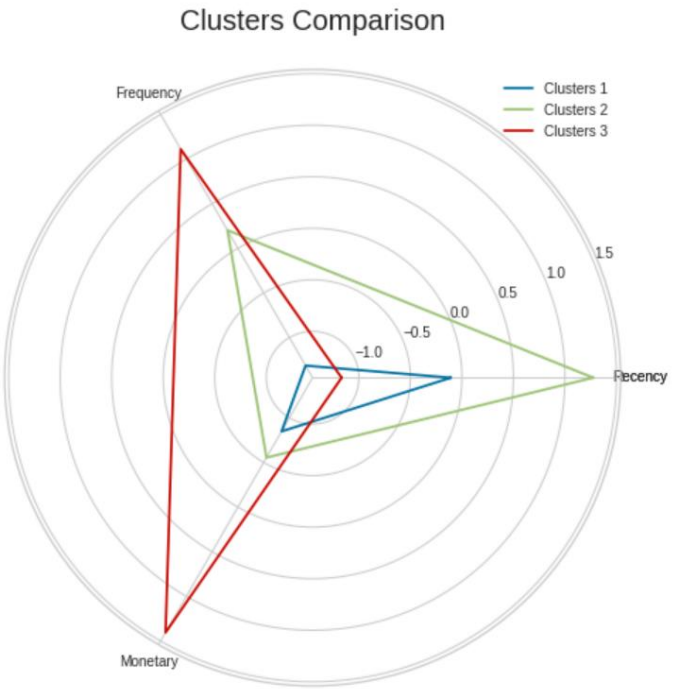
Let’s try Unsupervised Machine Learning!



After using scaling for our data. We can tried Unsupervised Machine Learning method to create cluster between SKUs. At this chance, we use K-Means clustering. Before start to train the model, we need to find the best K for our datasets.

To find the best K, we need to search the lowest Inertia. Inertia it’s a value that measure how good the data clustered by K-Means, It calculated by measuring the distance between each data point and centroid, squaring it and sum it to one cluster. The more lower inertia and lower number K will results on good model. At the plot above, we found that the best K is 4. However, in this chance we tried to clustering using K = 3 to see the difference RFM and K-Means clustering.

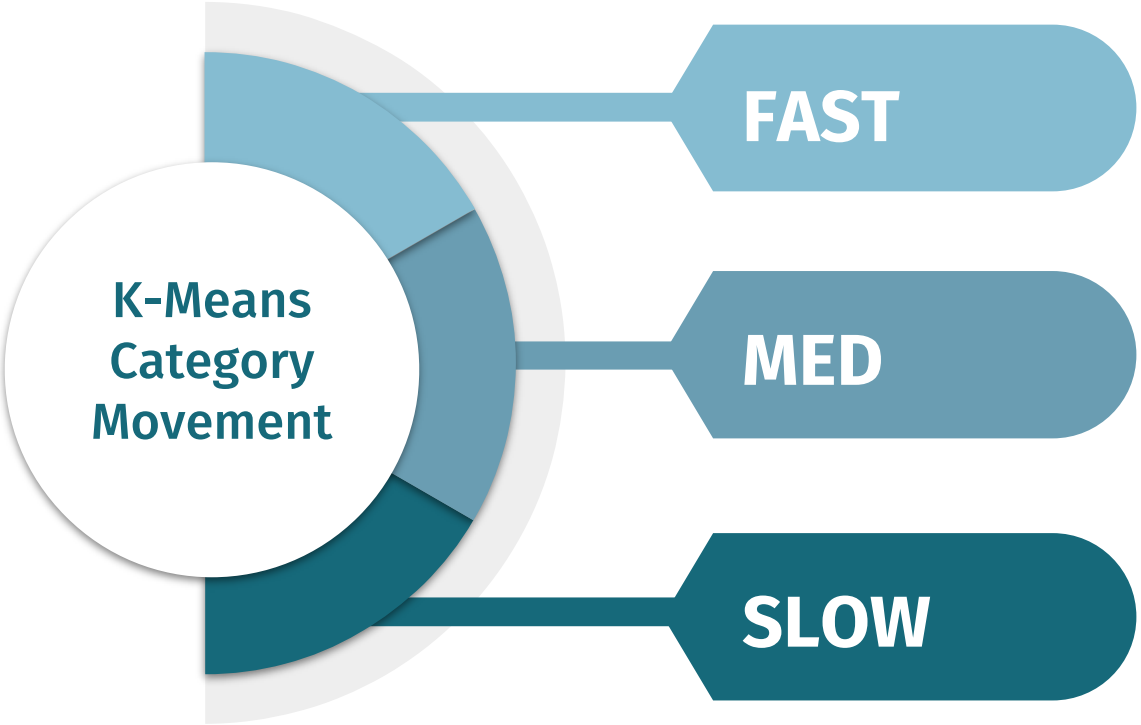
RFM vs K-Means



There are difference on output of RFM analysis and K-Means machine learning for categorized movement. In the K-Means clustering we got 14 FAST movement category SKUs (-6 from RFM), 45 MED movement category SKUs (+30 from RFM), and 55 SLOW movement category SKUs (-24 from RFM).

However, there are the recommendation from K-Means that we actually could separate from 3 cluster to 4 cluster movement. It will create more detailed movement but also we should consider the warehouse space in order to accomodate 4 different categories movement.

K-Means Output



For the FAST movement SKUs have all best criteria in 3 point of our analysis : high frequency order, huge monetary value, and fastest retention rate. These SKUs should be put as near as possible from the packing area so it will reduce the cycle time very much in order to cut lead time and produce more benefit to the company.

These category filled with SKUs with MED movement category, in which have average in value in monetary value and order frequencies but have more slower retention rate than SLOW movement. These SKUs can put on the middle-farest storage.

This SLOW movement filled by SKUs with slow retentional activities and low frequency of order. We could place all SKUs on this category on the farest storage.

Now see the impact and the recommendation.

SKU	Recency	Frequency	Monetary	rfm_score
1346	45.0	405.0	11303.420033	331.0
1347	43.0	207.0	10957.400143	321.0
1348	41.0	271.0	2750.029998	321.0
1349	41.0	592.0	240496.679017	333.0
1350	39.0	652.0	209268.380555	333.0
1351	38.0	442.0	595395.000000	333.0
1352	36.0	431.0	97937.559815	333.0
1353	32.0	484.0	200704.870333	333.0
1354	30.0	483.0	71319.379686	333.0
1355	29.0	484.0	231765.460292	333.0
1356	27.0	362.0	95358.110288	333.0
1357	27.0	208.0	39365.709993	323.0
1358	25.0	434.0	101687.289866	333.0
1359	21.0	492.0	37318.299847	333.0

SKU	Recency	Frequency	Monetary	rfm_score
1350	39.0	652.0	209268.380555	333.0

After we get the clustering sample, we found that one of the FAST category is SKU "1350". For one piece of SKU 1350 we will get USD $321, with assume that working hour for one shift is 8 hours. We can process 160 pcs of the SKU. And then we get monetary value at USD $ 33,482,941 from order in one shift. However, if we put it on the FAST category and reduce the processing leadtime we can process 240 pcs of the SKU. Impact of that, we can get monetary value at USD $50,224,412. **It's $16,741,471 more than previous method, which is raise 50% of previous SKU monetary value.**

At the end of analysis, we recommended to use the categorized movement (either RFM or K-Means method) because this will increase the efficiency of order fulfill and indirectly raise company income around **50%**.

Let's connect to discuss!



Full Python Code