

MFTBC CS Data Scientist Hiring – Case Study Problem

Problem 1:

How can Vidya be helped to understand the data better and derive strategic insights from it to answer his problem?

Vidya is expecting:

1. A clear and easy understanding of the data she has acquired.
2. Hidden insights that can be derived statistically from the data which he can use to solve his problem.
3. Recommendations and way forward for increasing fill rate.

Please indicate all assumptions with logic.

Findings and insights:

Firstly, I removed few unwanted columns and cleared the data. Then I perform left join on store_view and lead_time dataset on the basis of supply.site and store.

```
store_view_new = merge(store_view,lead_time,by.x = c("Supply.Site","Store"), by.y =  
c("Source","Destination"),all.x = TRUE)
```

then I calculated the difference between the actual DC to store Lead_time and proposed Lead_time.

```
Diff_lead_time = store_view_new$DC.to.Store.Lead.Time..in.Days. -
```

```
store_view_new$Lead.Time..in.Days.
```

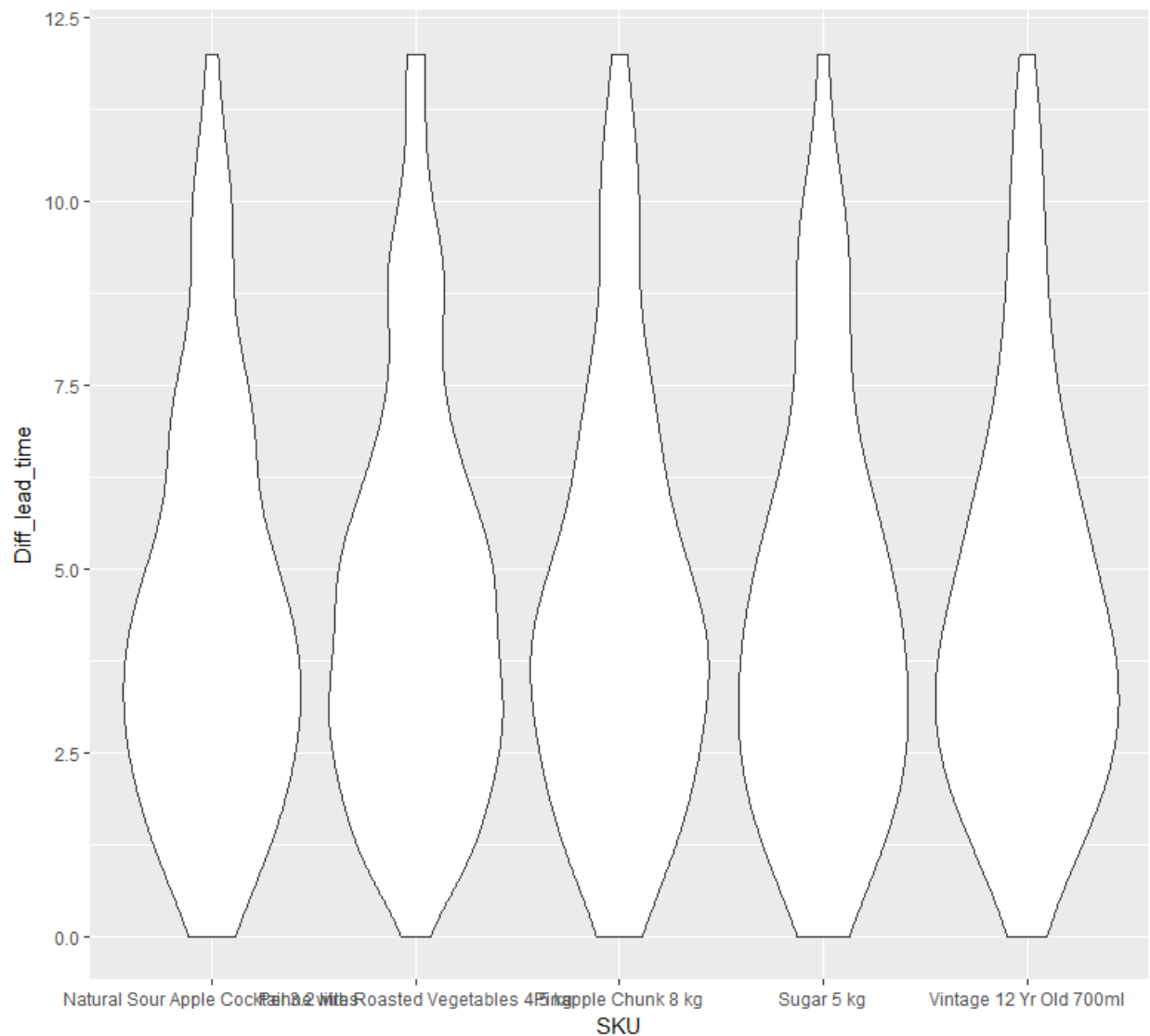
```
store_view_new = data.frame(store_view_new[,],Diff_lead_time)
```

```
unique(store_view_new$Diff_lead_time)
```

```
length(Diff_lead_time[Diff_lead_time==0])
```

#below plot shows how much is the Delay in lead time for each SKU.

```
ggplot(store_view_new, aes(y=Diff_lead_time, x = SKU))+geom_violin()
```



From the above ggplot, it looks like for every SKU on an average there is delayed by 3 days, very few are having 0 days as difference stating that they were on time and few even have delay of 12 days. It is also clear that none of the sku is within the lead_time, either they are on last day or delayed. Due to this, it seems like Lead_time has to be increased which can be done either at DC end, by scheduling the stock transfer earlier or at the store end, by placing the order little more earlier which is not a good thing to do as from data it is clear that they are already giving order much earlier. So, DC needs to schedule the stock transit earlier so the fill rate can be increased which will be a good option.

Further analysis is done by creating below tables:

```
table(store_view_new$Store,store_view_new$Diff_lead_time)
```

```
table(store_view_new$SKU,store_view_new$Diff_lead_time)
```

```
> table(store_view_new$Store,store_view_new$Diff_lead_time)
```

```

      0  1  2  3  4  5  6  7  8  9 10 11 12
Chandannagore Depot  0 36 41 42 25 27 24  0  0  0  0  0  0
Howrah Depot        0 39 45 32 36 43  0  0  0  0  0  0  0
Hugli Depot         0  0  0 17 25 19 20 18 22 24 13 20 17
Kolkata Depot       31 36 43 40 45  0  0  0  0  0  0  0  0
Nadia Depot         0  0 33 28 28 37 32 37  0  0  0  0  0
North 24 Parganas Depot 0  0  0 28 22 21 21 23 26 23 31  0  0
> table(store_view_new$SKU,store_view_new$Diff_lead_time)

      0  1  2  3  4  5  6  7  8  9 10 11 12
Natural Sour Apple Cocktail 3.2 litres  7 23 32 38 37 28 14 22  9  7 11  3  3
Penne with Roasted Vegetables 4.5 kg    2 24 31 38 32 34 23 13  9 14  6  3  5
Pinapple Chunk 8 kg                     7 22 31 35 41 28 18 17 10  6 10  6  3
Sugar 5 kg                              9 23 34 36 34 29 20 12 10 12 10  2  3
Vintage 12 Yr Old 700ml                  6 19 34 40 37 28 22 14 10  8  7  6  3

```

From first table, only Kolkata Depot is receiving stock on time or just few days late which may be because the store manager might have some good order placing techniques or good business model which can be shared among other depots to increase the fill rate at all store.

then daily sales for last month on rolling average is converted to factor levels to see sales for each SKU

```

store_view_new$daily_sales_last_month =
cut(store_view_new$Daily.SALES.QTY...Rolling.Average.for.Last.Month,
     breaks = c(75,105,135,165,195,225,255), labels = c("75-105","106-135","136-
165","166-195","196-225","226-255"))

```

```
> table(store_view_new$SKU,store_view_new$daily_sales_last_month)
```

	75-105	106-135	136-165	166-195	196-225	226-255
Natural Sour Apple Cocktail 3.2 litres	57	64	35	34	25	19
Penne with Roasted Vegetables 4.5 kg	56	68	32	32	29	17
Pinapple Chunk 8 kg	59	66	31	30	26	22
Sugar 5 kg	61	59	36	31	27	20
Vintage 12 Yr Old 700ml	56	71	29	30	28	20

From above table, it looks like mostly daily sales lies in 106-135 interval.

```
> table(store_view_new$Store,store_view_new$daily_sales_last_month)
```

	75-105	106-135	136-165	166-195	196-225	226-255
Chandannagore Depot	36	159	0	0	0	0
Howrah Depot	0	0	0	157	38	0
Hugli Depot	87	108	0	0	0	0
Kolkata Depot	0	0	0	0	97	98
Nadia Depot	166	29	0	0	0	0
North 24 Parganas Depot	0	32	163	0	0	0

The above table shows how many times sales falls in each group but it is not clear that which quantity of each SKU so created below tables taking each depot and looking for each SKU quantity group.

Each Store and its quantity group is highlighted and helps to project the range for purchasing order.

```
depots = unique(store_view_new$Store)
for (i in depots){
  depot = i
  print(table(store_view_new$SKU[store_view_new$Store ==
depot],store_view_new$daily_sales_last_month[store_view_new$Store == depot]))
  print(paste("Depot = ",i))
}
```

	75-105	106-135	136-165	166-195	196-225	226-255
Natural Sour Apple Cocktail 3.2 litres	8	31	0	0	0	0
Penne with Roasted Vegetables 4.5 kg	5	34	0	0	0	0
Pinapple Chunk 8 kg	7	32	0	0	0	0
Sugar 5 kg	9	30	0	0	0	0
Vintage 12 Yr Old 700ml	7	32	0	0	0	0

[1] "Depot = Chandannagore Depot"

	75-105	106-135	136-165	166-195	196-225	226-255
Natural Sour Apple Cocktail 3.2 litres	0	0	0	34	5	0
Penne with Roasted Vegetables 4.5 kg	0	0	0	32	7	0
Pinapple Chunk 8 kg	0	0	0	30	9	0
Sugar 5 kg	0	0	0	31	8	0
Vintage 12 Yr Old 700ml	0	0	0	30	9	0

[1] "Depot = Howrah Depot"

	75-105	106-135	136-165	166-195	196-225	226-255
Natural Sour Apple Cocktail 3.2 litres	15	24	0	0	0	0
Penne with Roasted Vegetables 4.5 kg	19	20	0	0	0	0
Pinapple Chunk 8 kg	17	22	0	0	0	0
Sugar 5 kg	19	20	0	0	0	0
Vintage 12 Yr Old 700ml	17	22	0	0	0	0

[1] "Depot = Hugli Depot"

	75-105	106-135	136-165	166-195	196-225	226-255
Natural Sour Apple Cocktail 3.2 litres	0	0	0	0	20	19
Penne with Roasted Vegetables 4.5 kg	0	0	0	0	22	17
Pinapple Chunk 8 kg	0	0	0	0	17	22
Sugar 5 kg	0	0	0	0	19	20
Vintage 12 Yr Old 700ml	0	0	0	0	19	20

[1] "Depot = Kolkata Depot"

	75-105	106-135	136-165	166-195	196-225	226-255
Natural Sour Apple Cocktail 3.2 litres	34	5	0	0	0	0
Penne with Roasted Vegetables 4.5 kg	32	7	0	0	0	0
Pinapple Chunk 8 kg	35	4	0	0	0	0
Sugar 5 kg	33	6	0	0	0	0
Vintage 12 Yr Old 700ml	32	7	0	0	0	0

[1] "Depot = Nadia Depot"

	75-105	106-135	136-165	166-195	196-225	226-255
Natural Sour Apple Cocktail 3.2 litres	0	4	35	0	0	0
Penne with Roasted Vegetables 4.5 kg	0	7	32	0	0	0
Pinapple Chunk 8 kg	0	8	31	0	0	0
Sugar 5 kg	0	3	36	0	0	0
Vintage 12 Yr Old 700ml	0	10	29	0	0	0

[1] "Depot = North 24 Parganas Depot"

From above tables, it looks like each store places nearly same quantity for each SKU.

Summarizing above findings to find below answers:

1. A clear and easy understanding of the data she has acquired.

The data which she acquired is having details of the time taken from DC to store and quantity which can be used to find out the problems at various store, DC. Also, other variables like Available stock, stock in transit is not having any mathematical connection to carry out operation. May be there is some discrepancy in data or require some other counter variables to go with how often store manager places an order and how much is the quantity of each purchasing order and when this order is needed. Also, from the data it is seen that after receiving the GRN, store also takes time to process the stock and finally get out for customer.

2. Hidden insights that can be derived statistically from the data which he can use to solve his problem.

- None of the stock is reaching within the lead_time either it is arriving at store on last day (very few cases) or late (by up to 12 days).
- Purchasing orders are not predicted corrected means each purchasing order is having less quantity when compared with average daily sales.
- After receiving stock, store also takes time to process the stock which is quite high.

3. Recommendations and way forward for increasing fill rate.

- The time taken by store after GRN should be reduced to increase the fill rate.
- DC should schedule the stock transit little earlier if the difference in Lead_time is increasing so much which can increase the fill rate at store.
- As Kolkata Depot is almost receiving stock on time, another depot should look for its business model or demand forecast model while placing purchasing order.