

PSA UnboXed

Data Analytics Business Challenge

Group Name: Hadilao

Group Members:

- **Yang Ming**
- **Natalie Lam**

Executive Summary

Introduction

The PSA UnboXed Data Analytics challenged us to help analyze a set of data to discover insights in redefining customers' digital experience on PORTNET, PSA's national-wide business-to-business port community digital solution.

A dataset of transactions of user data and web access logs has been provided. The dataset contains Web Access Logs totalling 112,605 records over a 7-day period, and Transactions Logs totalling 905,553 records over a two-month period.

Results, Conclusions and Recommendations

After our analysis, interesting patterns and relationships between users and their user behavior emerged.

LHS	=>	RHS
Delivery Order	=>	Container
Pregate Services	=>	Container
Container, Delivery Order, FlexiBook, Storing and Release Orders	=>	Pregate Services

We analyzed and observed a pattern of users consistently going to access Container Information Services on Haulier after Delivery Order and Pregating services. This insightful observation runs contrarian to our understanding of a Hauliers operation and its process flow. (Delivery Order -> Pregate Services) for imports (Pregate -> Arrival into Terminal) for exports

Our analysis built a user profile who's likely to access Container to check on Container Status after Delivery Order. This indicates potential areas of improvement in the process flow. We recommend putting the Container Status (Historical/Latest) and other commonly accessed data in a Dashboard overview can elevate the user experience on Portnet.

Analysis of user transactions revealed an association between consecutive transactions (i.e. transactions that are made together between 1 to 10 mins) and aborted transactions (i.e. transactions made within 10 seconds). We did not include transactions with a time difference of less than 1 minute in consecutive transactions as most of these transactions would fall under the aborted cases. It is understandable that the popular consecutive transactions also have a high abort rate as the web might not be able to cope with so many users. However, it is of concern when the number of aborted cases surpasses the number of consecutive transactions significantly. For example, we found that for "Delivery Order -> Delivery Order" transactions, there were **4,839** consecutive transactions and **20,415** aborted transactions which is 4 times that of consecutive transactions, a significant bottleneck. Hence, we recommend looking for targeted improvement in that aspect.

For Mobile/Web Access of the Portal, we found that the Web Portal usage is significantly higher than that of the Mobile Portal usage for most USERIPs. But interestingly, we found that countries like India used only the Mobile Portal. In India, the market is dominated by Mobile Phones, particularly budget smartphones over Desktop Computers which are considerably more expensive. This may be reflective of Indian users' preference for the Mobile Portal, but the general consensus for Portal Access is preferred through the web access.

We also realized there is no Mobile App for Portnet or PSA on the App Store and Android Play Store, and when we tried accessing Portnet through our mobile phone it failed to load the page. This is another area of improvement that can be looked into.

Limitation and Observations

Properties of the dataset:

1. Transaction data only contains one significant column of data for feature engineering. For example, it would have been useful to know the profits gained from each type of transaction to better measure the significance of each type of transaction.
2. Web Access data also only contains one significant column of data for feature engineering beyond identifying columns.

Over 99.6% of Access Logs come preminilarly from Singapore IPs, and this high number really stands out. A typical user's profile only accesses Portnet Haulier's services only in Singapore.

Full Report

(1) Rank the frequency of the transaction sequences (user-path)

-Outcome: to improve understanding of our user needs.

We analysed:

1. User sessions
2. Category frequencies

Firstly, we combined the two PN_TRANS files into one and extracted the category and subcategory from the USERPATH. We also split the LOGTIME column into date and time. We filtered out the transactions “webenv=Internet (Primary);entry;=” and “webenv=Internet (Secondary);entry;=” as these sessions seem to be idle.

This is how our cleaned Portnet Transaction History looks like:

	USERID	ORCODE_N	DATE	HOUR	MIN	SEC	CATEGORY	SUBCATEGORY	PRI_SEC	TIME
	<chr>	<fct>	<date>	<int>	<int>	<int>	<fct>	<fct>	<fct>	<dtm>
1	umar	ORG3	2020-01-07	0	8	37	Container	Cntr Status: Latest (DCON)	Pri	2020-01-07 00:08:37
2	etty	ORG11	2020-01-25	0	43	56	Container	Cntr Status: Latest (DCON)	Pri	2020-01-25 00:43:56
3	d8sen1	ORG1	2020-01-30	0	1	15	Container	Cntr Status: Latest (DCON)	Pri	2020-01-30 00:01:15
4	hencon	ORG4	2020-01-27	0	4	17	Container	Cntr Status: Latest (DCON)	Pri	2020-01-27 00:04:17
5	hencon	ORG4	2020-01-27	0	4	20	Container	Cntr Status: Latest (DCON)	Pri	2020-01-27 00:04:20
6	hencon	ORG4	2020-01-05	0	10	58	Container	Cntr Status: Latest (DCON)	Pri	2020-01-05 00:10:58
7	huye	ORG4	2020-01-01	0	43	34	Container	Cntr Status: Latest (DCON)	Pri	2020-01-01 00:43:34
8	eejin	ORG4	2020-01-12	0	48	3	Container	Cntr Status: Latest (DCON)	Pri	2020-01-12 00:48:03
9	eechan	ORG4	2020-01-05	0	42	39	Container	Cntr Status: Latest (DCON)	Pri	2020-01-05 00:42:39
10	asonhe	ORG8	2020-01-30	0	17	10	Container	Cntr Status: Latest (DCON)	Pri	2020-01-30 00:17:10

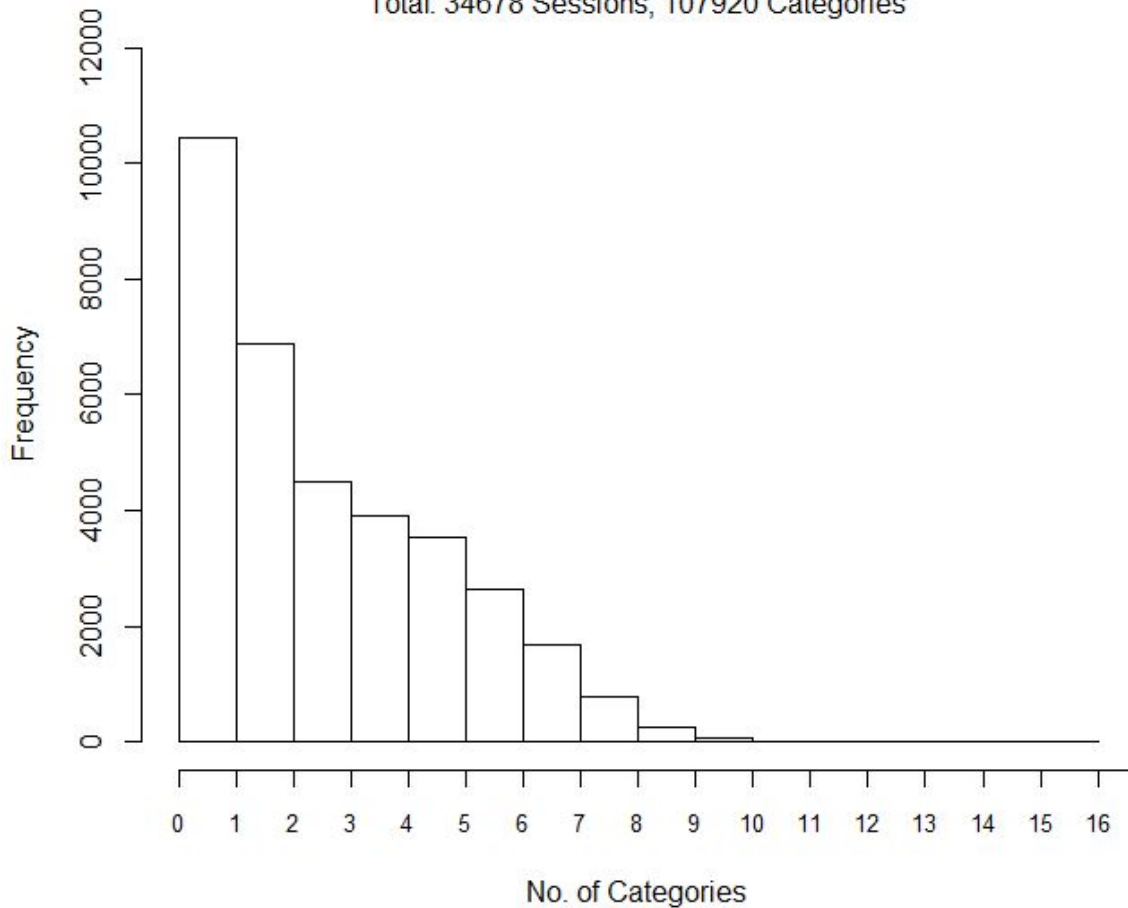
From our cleaned dataset, we logged **34,678 user sessions** from January to February 2020 who accessed the Haulier services. We took the definition of a single user session to be the same User ID making a series of transactions within the same day.

This is how the dataset looks like:

	USERID	DATE	basket
1	Qjas3	2020-01-02	c("Dangerous Goods", "Vessel", "Vessel", "Vessel")
2	Qjas3	2020-01-06	c("Container", "Delivery Order")
3	Qjas3	2020-01-07	c("Container", "Vessel", "Vessel", "Vessel")
4	Qjas3	2020-01-09	Vessel
5	Qjas3	2020-01-10	Vessel
6	Qjas3	2020-01-13	c("Dangerous Goods", "Container")
7	Qjas3	2020-01-14	Vessel
8	Qjas3	2020-01-15	c("Vessel", "Vessel", "Vessel")
9	Qjas3	2020-01-16	c("Vessel", "Container", "Container", "Vessel")
10	Qjas3	2020-01-17	c("Vessel", "Vessel", "Vessel", "Vessel")
11	Qjas3	2020-01-20	c("Container", "Vessel", "Vessel", "Container", "Vessel", "Vess...
12	Qjas3	2020-01-21	c("Vessel", "Vessel", "Container", "Container", "Container", "C...
13	Qjas3	2020-01-22	c("Vessel", "Container", "Vessel", "Vessel")

Number of Categories per session

Total: 34678 Sessions, 107920 Categories



This graph shows us the number of categories accessed in a transaction sequence each user makes per session. A user can access multiple categories within the same session in his transaction sequence. For example, a user can log on to Portnet, access categories such as Container, then further access other categories such as Pregate Services. We can see that most users access up to 6 categories per session.

This appears to be aligned with the high-level process flow for hauliers, which we understand to have a documentation flow of:

Documentation for Import

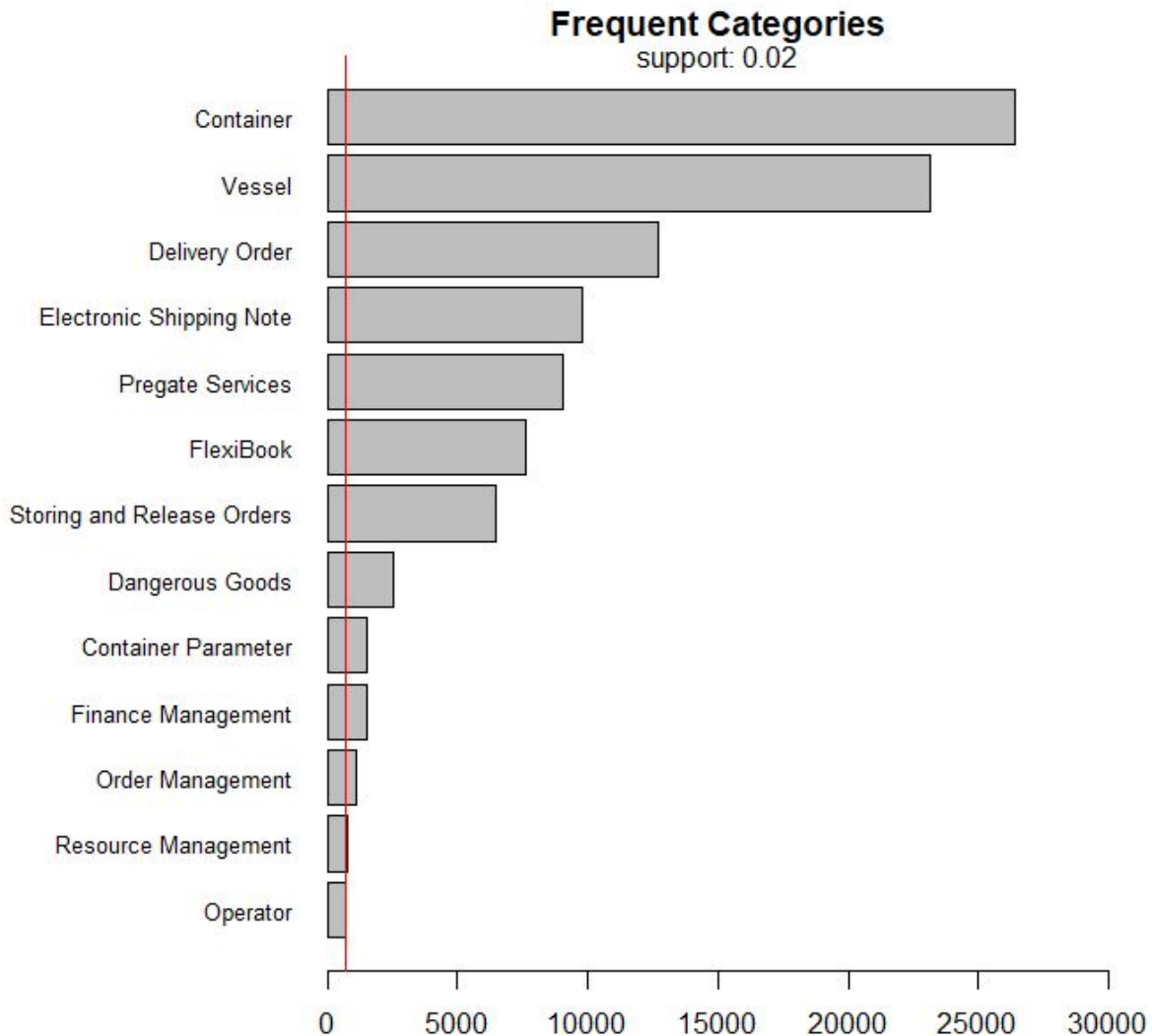
1. Delivery Order
2. Pregate
3. Arrive into terminal
4. Pick container
5. Exit out of terminal

Documentation for Export

1. Electronic Shipping Note
2. Flexibook
3. Pregate
4. Arrive into terminal
5. Drop container
6. Exit out of terminal

This insight helps us understand the customers journey better, if a high-level process flow ideally has 5 or 6 steps of documentation and a large proportion of user sessions have accessed 5 to 6 categories, it is a sign that the user experience is well designed. However we need to make sure that these 5 or 6 categories accessed are not repeated or aborted transactions, for example if a user only needs to create one Delivery Order but have to access 5 categories before he ends his user sessions. We will analyze this in greater detail in the next section

We will conduct further analysis on user behavior on Portnet, whether the choice of Categories they access is by chance, or not, by studying their transaction frequency.

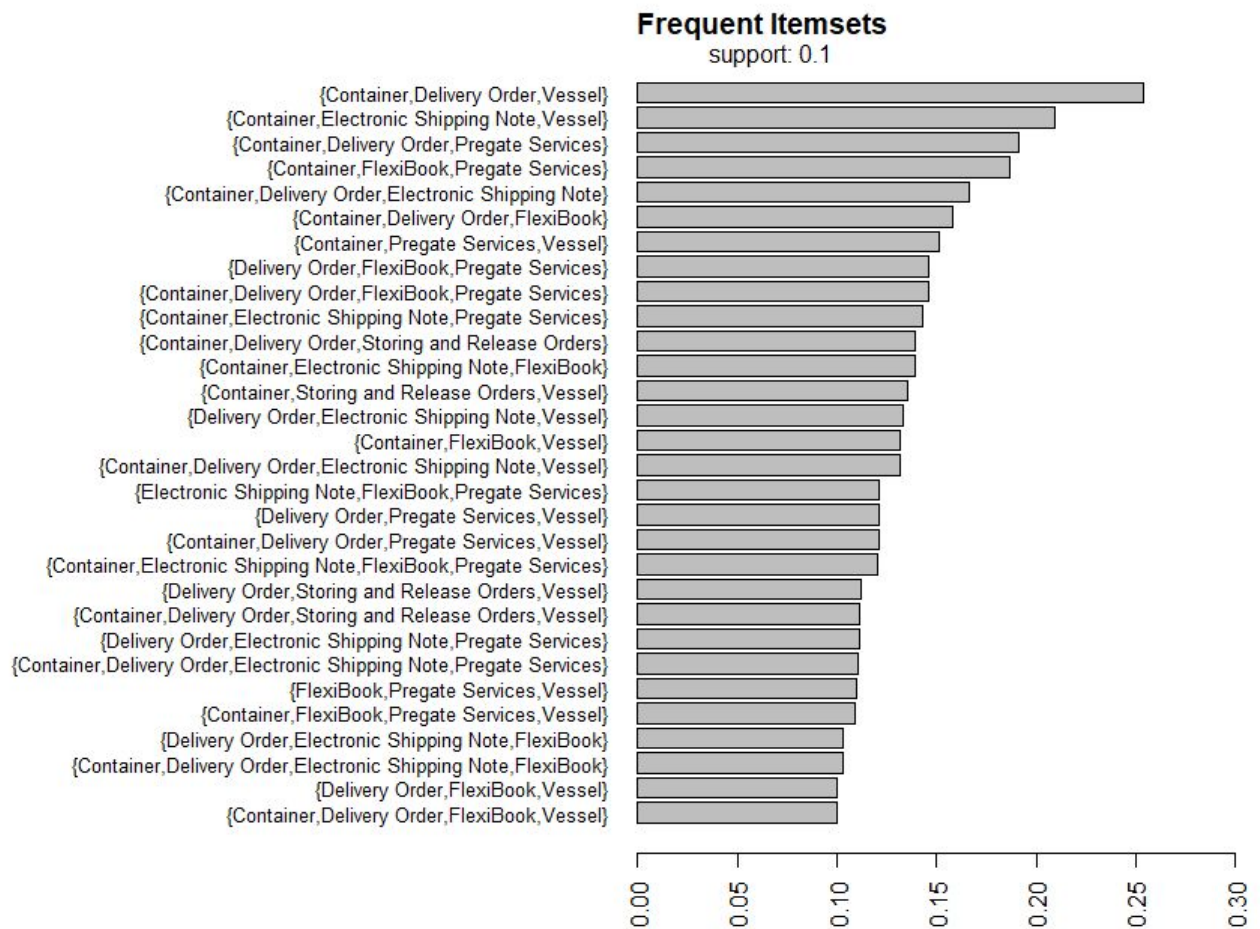


The above plot shows the frequent transactions made in a single user session. The most frequent transactions came from Information Services (Container and Vessel) followed by Container Services (Pregate Services, Delivery Order, Electronic Shipping Note, FlexiBook).

We used a support threshold of 0.02, meaning that a Category will be considered frequently accessed if at least 2 percent of all user sessions contain it in its transaction sequence. In this case, a category will be considered frequent if it appears in more than 694 user sessions.

We observed that a lot of Container Services were being accessed. Perhaps this has to do with the High Level Process Flow for Hauliers, which a lot of steps involves documenting Containers. We investigate this later on in the report when we mined the data for association rules.

We increase the support threshold to take into account the relatively large probability of observing a frequent itemset of at least size 3.



This shows us the relative frequency of transaction baskets made by each user, with the *{Container, Delivery Order, Vessel}* basket combination being the most frequent, followed closely by the *{Container, Electronic Shipping Note, Vessel}* basket combination.

Definitions:

Association Rules: Predict the occurrence of a category based on the occurrences of other categories accessed in user session.

$\{X\} \rightarrow \{Y\}$: If Category X accessed, THEN category Y co-accessed.

Confidence: Fraction of a rule $\{X\} \rightarrow \{Y\}$ in the total number of sessions with X

Lift: Probability of having $\{X\} \rightarrow \{Y\}$ / Probability of having $\{X\}$ & $\{Y\}$ in a random chance

Close to 1 indicates a rule $\{X\} \rightarrow \{Y\}$ occurring more than a random chance

Support: Proportion of items occurs in the dataset

We mined the Data for Association Rules with the Apriori Algorithm

We started with a low support threshold and high confidence to get strong rules even for less frequent categories. Then we inspect those rules with high lift indicating strong associations between categories, as well as rules with high confidence, indicating strong predictability in our association rule being more than the probability of happening in a random chance. After tuning, these are our support and confidence values.

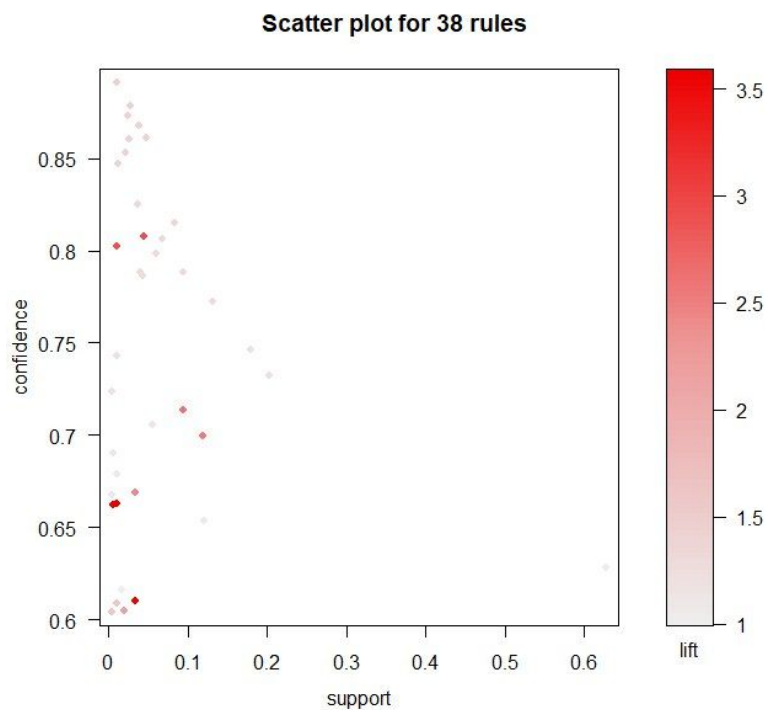
mining info:

data	ntransactions	support	confidence
transactions	184233	0.005	0.6

Summary of quality measures:

support	confidence	lift	count
Min. :0.005032	Min. :0.6042	Min. :0.9811	Min. : 927
1st Qu.:0.011442	1st Qu.:0.6678	1st Qu.:1.1849	1st Qu.: 2108
Median :0.033496	Median :0.7449	Median :1.3311	Median : 6171
Mean :0.063114	Mean :0.7459	Mean :1.6469	Mean : 11628
3rd Qu.:0.065896	3rd Qu.:0.8133	3rd Qu.:1.5564	3rd Qu.: 12140
Max. :0.627960	Max. :0.8911	Max. :3.5777	Max. :115691

Overall, we see that Flexibook and Pregate Services are categories that are often accessed together, and the predictability of a user accessing Container Category is very high.



From the above scatterplot, there are some rules with a heavy lift indicating a strong association between the categories. We shall investigate this later on, but first to interpret this scatterplot: here is an association rule that was mined:

Association Rule:

{Delivery Order, Vessel} -> {Container} Support: 0.2535 Confidence: 0.962, Lift: 1.265

Support: 0.2535 means that 0.2535 proportion of all user sessions accessed Delivery Order and Vessel.

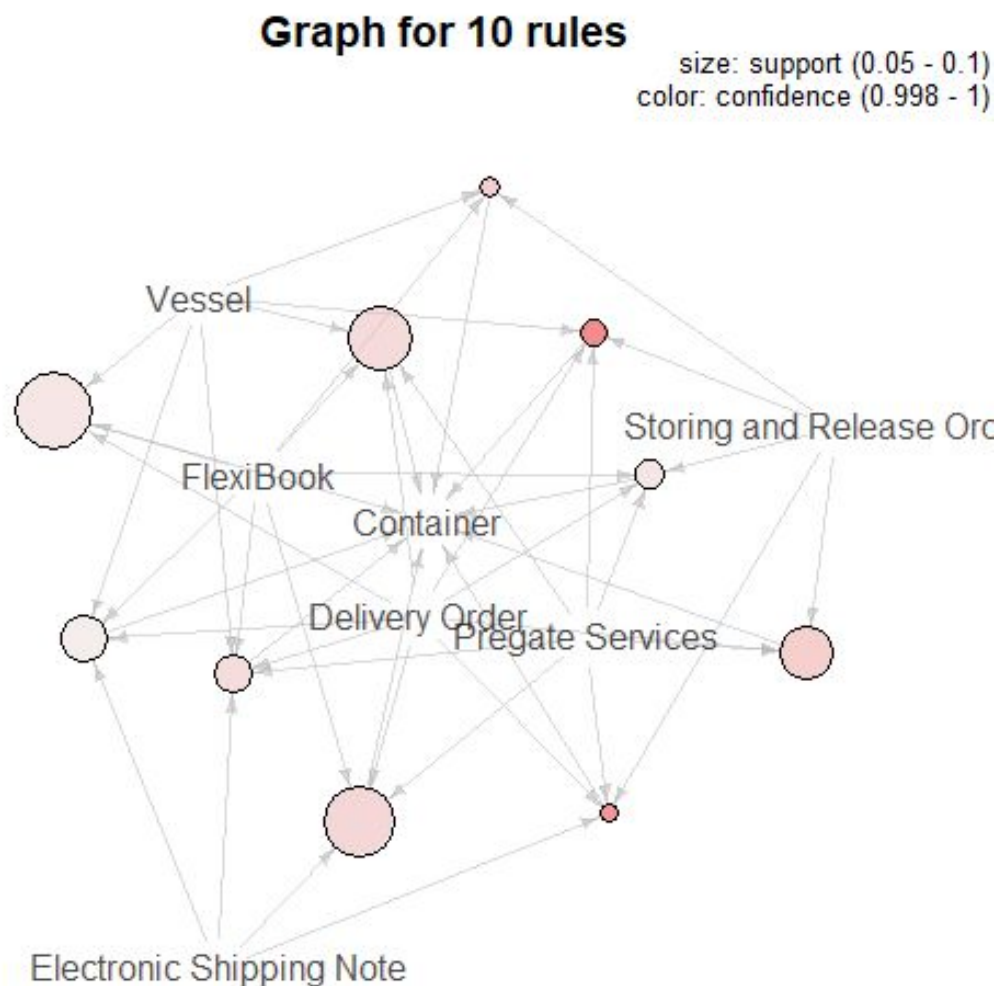
Lift: 1.265 means that a person accessing categories Delivery Order and Vessel is 1.399 more likely to access Container than a regular user.

Confidence: 0.962 refers to the predictability of a person accessing categories FlexiBook and Vessel accessing Container.

We averaged the three metrics and produced the top 10 Association Rules:

	lhs	Var.2	rhs	support	confidence	lift	average
	{Delivery Order}	=>	{Container}	0.35195225	0.9589095	1.260636	1.5173372
	{Pregate Services}	=>	{Container}	0.25468597	0.9778565	1.285545	0.9170181
{Container,Delivery Order,FlexiBook,Storing and Release Orders}		=>	{Pregate Services}	0.05934598	0.9320652	3.578627	0.7887913
{Delivery Order,FlexiBook,Storing and Release Orders}		=>	{Pregate Services}	0.05943249	0.9317360	3.577363	0.7854327
	{Delivery Order,Vessel}	=>	{Container}	0.25350366	0.9621320	1.264873	0.7458808
{Container,Delivery Order,Electronic Shipping Note,Pregate Services}		=>	{FlexiBook}	0.09513236	0.8588909	3.900553	0.6742247
{Delivery Order,Electronic Shipping Note,Pregate Services}		=>	{FlexiBook}	0.09524771	0.8581450	3.897165	0.6655247
{Container,Electronic Shipping Note,Pregate Services}		=>	{FlexiBook}	0.12001846	0.8414881	3.821520	0.6526369
{Electronic Shipping Note,Pregate Services}		=>	{FlexiBook}	0.12140262	0.8391469	3.810887	0.6336847
{Electronic Shipping Note}		=>	{Container}	0.26558625	0.9414290	1.237655	0.6320238

The below graph is a visualization of the most common transaction sequences.



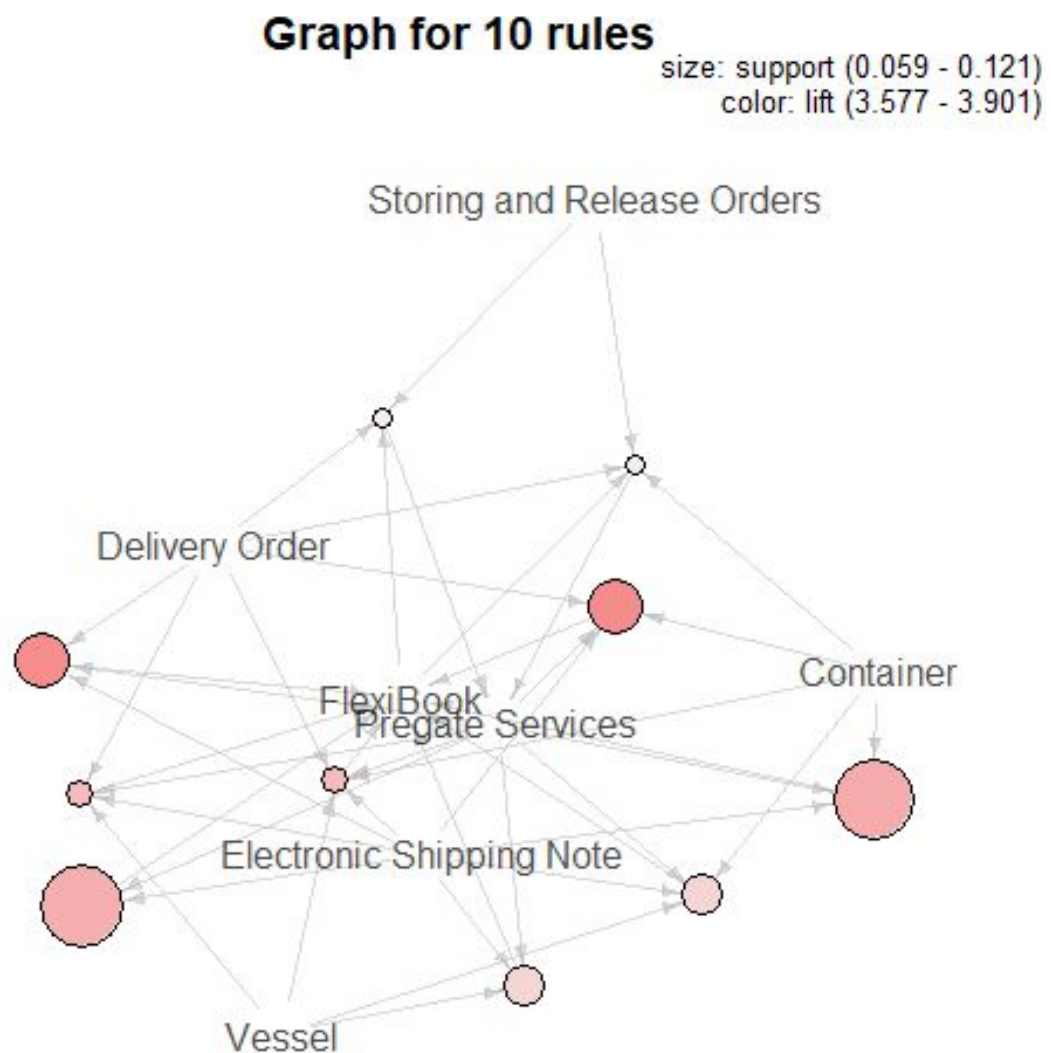
The unlabelled dots represent Confidence Level, whose values are determined by the shading (0.988-1).

The sizes of the unlabelled dots represent the Support, the proportion of categories that are accessed in all sessions in the dataset (0.05-0.1).

Directed arrows from Categories to unlabelled dots refer to the LHS in the association rule.

Directed arrows from unlabelled dots refer to the RHS in the association rule.

The below graph is the same for the most common transaction sequences, sorted by lift instead.



Proposal:

From the plots, Vessel and Container are the most commonly accessed categories, and represent potential cross-selling and upselling opportunities as well as areas of focus for improvement

(2) Using timestamp data, rank the most common consecutive transactions where the time gap between the first click of each transaction is within 10 mins.

-Outcome: to identify which transactions can be completed in one setting and reduce users' turnaround time spent on the portal

Here is the code and brief description of how we went about finding consecutive transactions where the time gap is within 1 to 10 mins. We chose the minimum to be a 1 min difference as we noticed that those with less than 1 minute difference looked like the aborted cases (repeat of the same transaction).

```
# I will mainly focus on the columns USERID, DATE, TIME, CATEGORY and SUBCATEGORY
picked <- trans_combined %>% select(USERID, ORGCODE_N, DATE, TIME, CATEGORY, SUBCATEGORY)

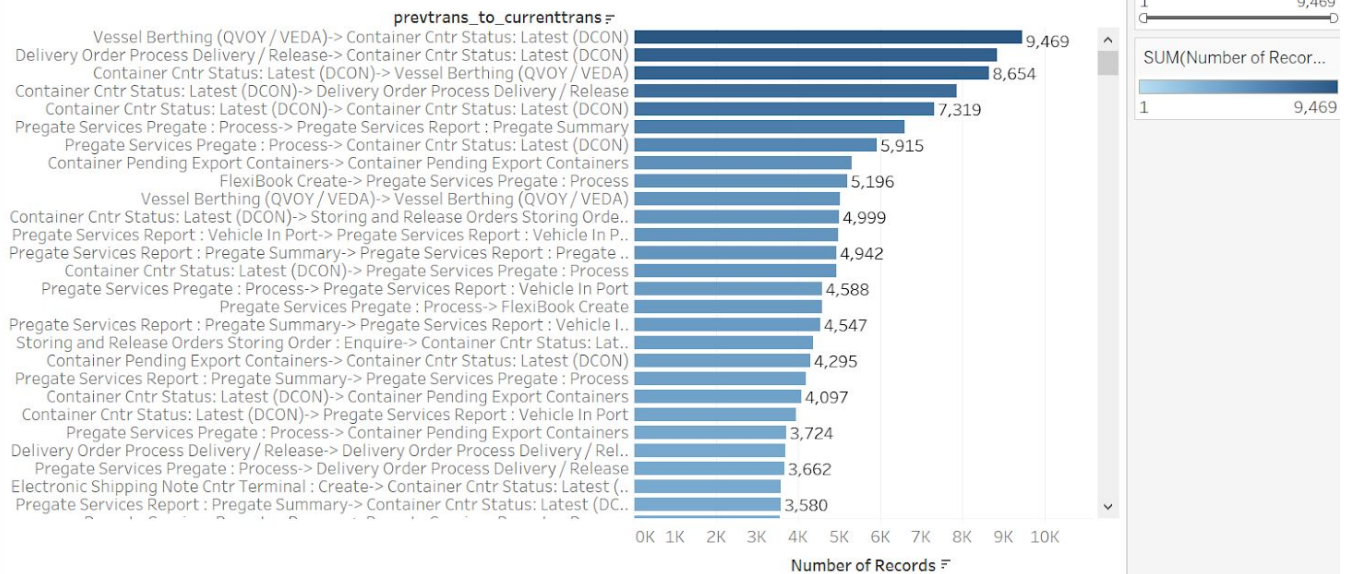
flow <- picked %>%
  group_by(USERID, DATE) %>%
  arrange(TIME) %>%
  unite("transaction", c("CATEGORY", "SUBCATEGORY"), sep=" ") %>%
  mutate(timediff = as.numeric(difftime(TIME, lag(TIME), units = "mins")), #adding new column that shows the
time difference between two consecutive transactions
  prev_cat = lag(transaction)) %>% #adding a new column that shows the previous transaction
  filter(between(timediff,1,10)) %>% # filtering out the rows that have time difference in range [1,10]
  unite("prevtrans_to_currenttrans",c("prev_cat", "transaction"), sep = "-> ", remove = T) %>%
  select(DATE,USERID, prevtrans_to_currenttrans, timediff)
```

Here is how the data set looks like now:

```
# A tibble: 335,892 x 4
# Groups:   USERID, DATE [21,801]
  DATE      USERID prevtrans_to_currenttrans      timediff
  <date>    <chr>    <chr>                                     <dbl>
1 2020-01-01 ip161  Container Cntr Status: Latest (DCON)-> Pregate Services Report : Pregate Summary 1.25
2 2020-01-01 ip161  Pregate Services Report : Pregate Summary-> Container Cntr Status: Latest (DCON) 1.37
3 2020-01-01 okwai  Delivery Order Process Delivery / Release-> Container Cntr Status: Latest (DCON) 1.57
4 2020-01-01 okwai  Container Cntr Status: Latest (DCON)-> Electronic Shipping Note Cntr Terminal : Update ~ 1.13
5 2020-01-01 annan1 FlexiBook Enquire-> Container Pending Export Containers 2.92
6 2020-01-01 ip161  FlexiBook Create-> Pregate Services Report : Vehicle In Port 2.73
7 2020-01-01 hawk   Delivery Order Process Delivery / Release-> Delivery Order Process Delivery / Release 5.6
8 2020-01-01 annan1 Pregate Services Report : Pregate Summary-> FlexiBook Enquire 9.13
9 2020-01-01 hawk   Delivery Order Process Delivery / Release-> Pregate Services Report : Pregate Summary 2.07
10 2020-01-01 annan1 FlexiBook Enquire-> Container Pending Export Containers 2.95
# ... with 335,882 more rows
```

For example, in row 1, a particular **“Container: Cntr Status: Latest (DCON) -> Pregate Services: Report: Pregate Summary”** consecutive transaction takes 1.25 mins, with Container and Pregate Services being the main categories.

Most Common Consecutive Transactions

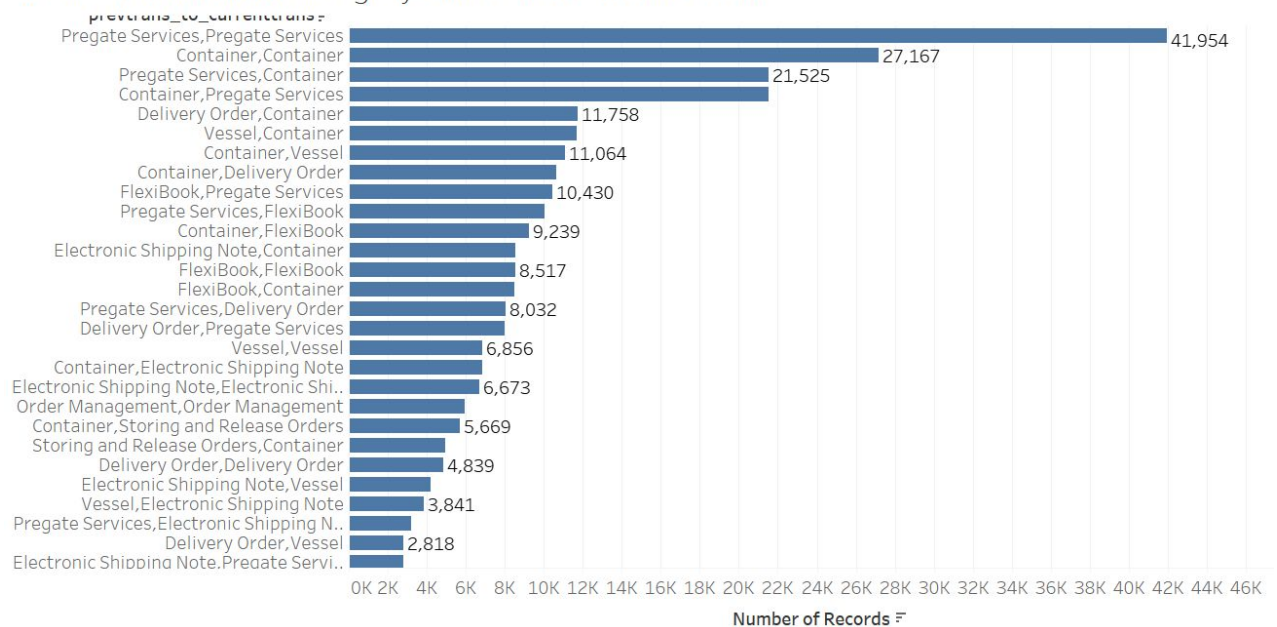


This shows us the ranking of most common consecutive transactions, ranked from most to least common.

From the graph, we can see the most common transaction is from “**Vessel: Berthing (QVOY / VEDA) -> Container: Cntr Status: Latest (DCON)**” as well as “**Container: Cntr Status: Latest (DCON) -> Vessel: Berthing (QVOY / VEDA) -> ”**. Vessel and Container are the main categories. This supports the proposal in question (1) where we stated that Vessel and Container are the most commonly accessed categories and that they are linked!

This looks very cluttered and so we will group by the **main** category transaction (e.g. Container, Pregate Services, Vessel etc.) instead.

Most Common Main Category Consecutive Transactions

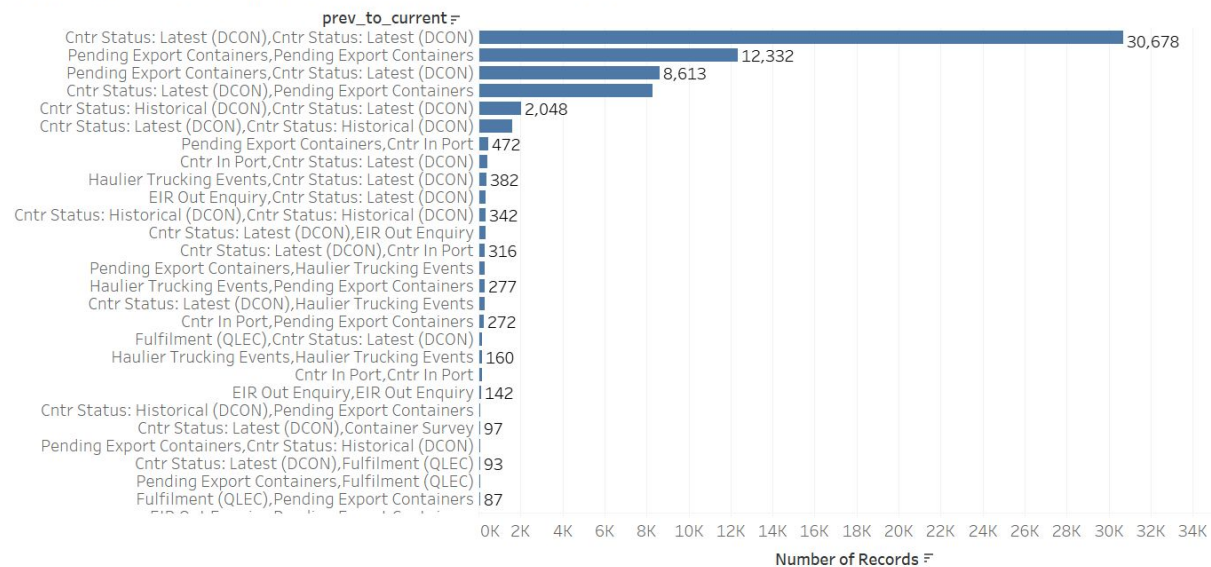


This is a bar plot to show the most common **main** category consecutive transactions made.

There seems to be a lot of **“Pregate Services -> Pregate Services”** and **“Container -> Container”** consecutive transactions. Following that, there are also significant **“Container -> Pregate Services”** and **“Pregate Services -> Container”** transactions.

We then delve deeper into the **Container** subcategories consecutive transactions:

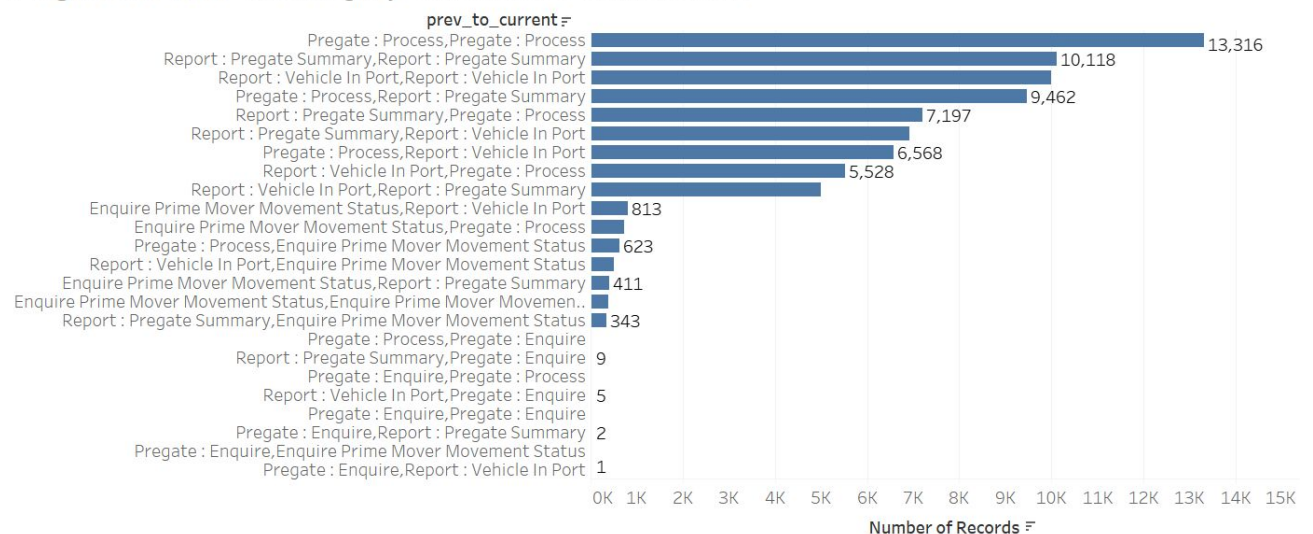
Container subcategory consecutive transactions



Here, we can see that the main subcategory transaction that contributed to “Container -> Container” consecutive transactions was the **“Cntr Status: Latest (DCON) -> Cntr Status: Latest (DCON)”** subcategory transaction followed by **“Pending Export Containers -> Pending Export Containers”**.

We next study the **Pregate Services** subcategories consecutive transactions:

Pregate Services subcategory consecutive transactions



The **“Pregate: Process -> Pregate:Process”** subcategory contributed most to the “Pregate Services -> Pregate Services” consecutive transactions, followed by the **“Report: Pregate Summary -> Report: Pregate Summary”** subcategory transaction.

(3) Identify and rank the transactions that are aborted with user having to repeat the transaction again

-Outcome: to understand if the user interface of the particular transaction system needs to be improved

Here, we filtered for consecutive transactions that had less than or equal to 10 seconds difference (i.e. aborted transaction).

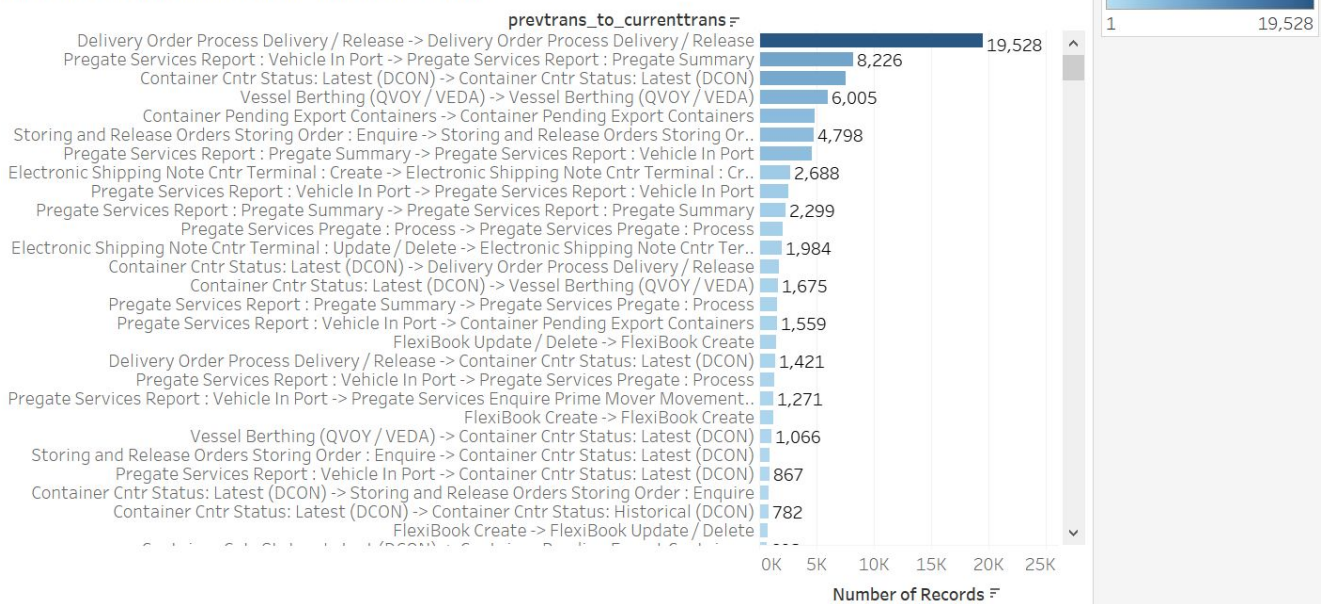
```
abort <- picked %>%
  group_by(USERID, DATE) %>%
  arrange(TIME) %>%
  unite("transaction", c("CATEGORY", "SUBCATEGORY"), sep=" ") %>%
  mutate(timediff = as.numeric(difftime(TIME, lag(TIME), units = "secs")), #adding new column that shows the
time difference between two consecutive transactions in seconds
  prev_cat = lag(transaction)) %>% #adding a new column that shows the previous transaction
  filter(timediff <= 10) %>% # filtering out the rows that have time difference of less than 10 seconds
  unite("prevtrans_to_currenttrans", c("prev_cat", "transaction"), sep = " -> ", remove = T) %>%
  select(DATE, USERID, prevtrans_to_currenttrans, timediff)
```

This is how the dataset looks like:

```
# A tibble: 120,212 x 4
# Groups:   USERID, DATE [18,215]
  DATE      USERID prevtrans_to_currenttrans      timediff
  <date>    <chr>    <chr>                                     <dbl>
1 2020-01-01 okwai  Delivery Order Process Delivery / Release -> Delivery Order Process Delivery / Release      2
2 2020-01-01 ip161  FlexiBook Create -> FlexiBook Create                                     6
3 2020-01-01 annan1 Container Pending Export Containers -> Pregate Services Report : Vehicle In Port             5
4 2020-01-01 angsh  Delivery Order Process Delivery / Release -> Delivery Order Process Delivery / Release      3
5 2020-01-01 ip161  Container Pending Export Containers -> Pregate Services Report : Pregate Summary             8
6 2020-01-01 ip161  Delivery Order Process Delivery / Release -> Container Cntr In Port                2
7 2020-01-01 annan1 Delivery Order Enquire -> Electronic Shipping Note Cntr Terminal : Update / Delete           9
8 2020-01-01 d766  Delivery Order Yard Stacking -> Delivery Order Yard Stacking                  1
9 2020-01-01 hawk  Delivery Order Process Delivery / Release -> Delivery Order Process Delivery / Release      5
10 2020-01-01 ip161 Pregate Services Report : Vehicle In Port -> Container Cntr In Port                6
# ... with 120,202 more rows
```

Row 1 tells us that a particular **“Delivery Order Process Delivery / Release -> Delivery Order Process Delivery / Release”** transaction takes 2 seconds. Delivery order is the main category.

Most Common Aborted Transactions

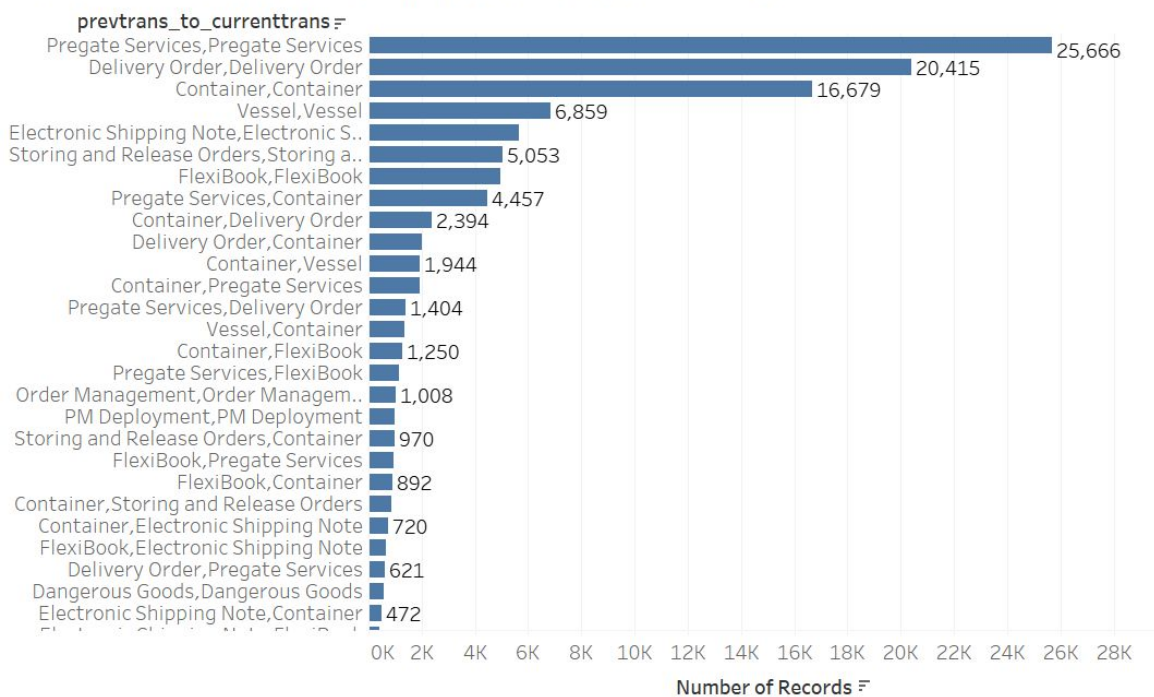


This graph shows us the ranking of the most common aborted transactions.

The transaction “**Delivery Order Process Delivery / Release -> Delivery Order Process Delivery / Release**” has a significantly high abort rate. This could be an area for improvement.

This looks very cluttered and so we will group by the **main** category transaction (e.g. Container, Pregate Services, Vessel etc.) instead.

Most Common Main Category Aborted Transactions



This is a bar plot to show the most common **main** category aborted transactions.

We can see that the top 3 most common aborted transactions are the “**Pregate Services -> Pregate Services**”, “**Delivery Order -> Delivery Order**” and “**Container -> Container**” transactions.

Some statistics:

“Pregate Services -> Pregate Services”:

- **41,954** consecutive transactions
- **25,666** aborted transactions

“Container -> Container”:

- **27,167** consecutive transactions
- **16, 679** aborted transactions

“Delivery Order -> Delivery Order”:

- **4,839** consecutive transactions
- **20,415** aborted transactions

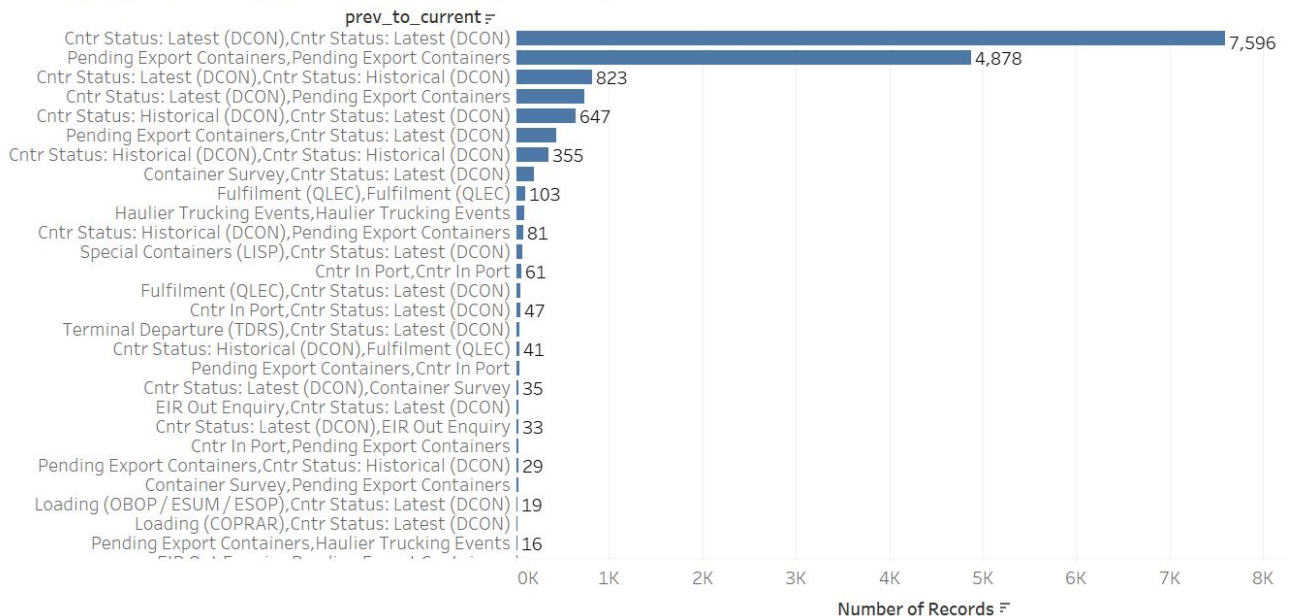
For “**Pregate Services -> Pregate Services**” and “**Container -> Container**”, their consecutive transactions are very high. Similarly, their aborted transactions are very high as well. This could be due to the hanging of the website since many users are doing the same transaction. There might be some association between consecutive and aborted transactions.

There is a concern for “**Delivery Order -> Delivery Order**” transactions as their aborted transactions are 4 times their consecutive transactions.

Since there are many **“Container -> Container”** and **“Pregate Services -> Pregate Services”** abort transactions, we decided to study their subcategories more closely.

We first study the subcategories for the **“Container -> Container”** aborted transactions.

Container subcategories aborted transactions

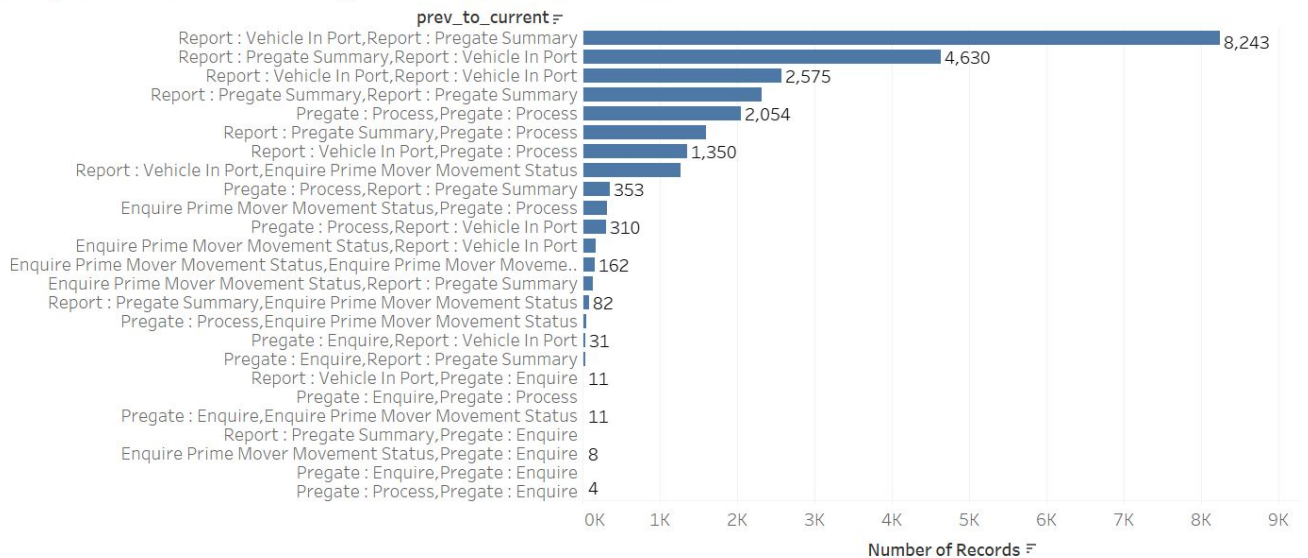


Here, we can see that the main culprit that contributed to “Container -> Container” aborted transactions was the **“Cntr Status: Latest (DCON) -> Cntr Status: Latest (DCON)”** subcategory transaction followed by **“Pending Export Containers -> Pending Export Containers”**.

This shows a similar pattern to the consecutive subcategory transactions for Container. Hence there might be association with consecutive and aborted transactions.

Secondly, we study the subcategories of the “**Pregate Services -> Pregate Services**” aborted transactions.

Pregate Services subcategories of aborted transactions

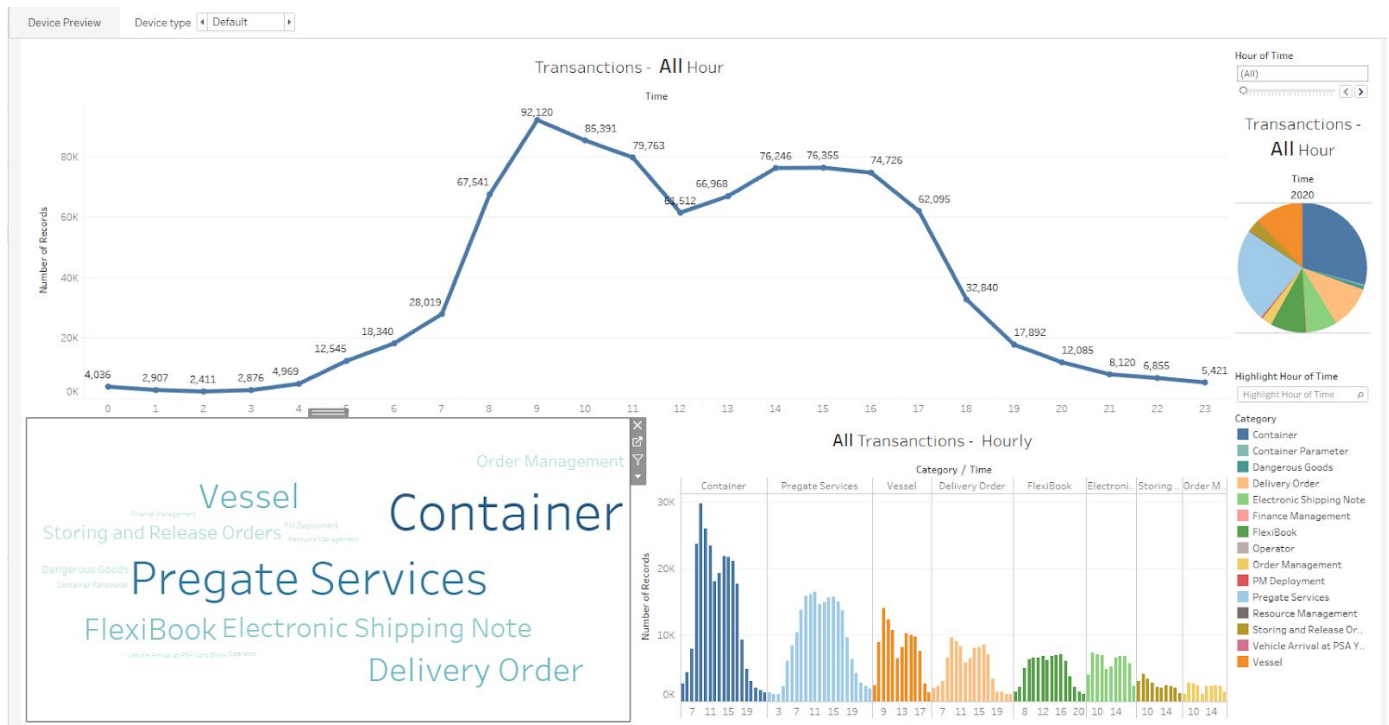


The “**Report: Vehicle in Port -> Report: Pregate Summary**” subcategory contributed most to the “Pregate Services -> Pregate Services” aborted transactions, followed by the “**Report: Pregate Summary -> Report: Vehicle in Port**” subcategory transaction.

(4) Based on no. of transactions, rank the hourly timings where the portal is most utilised. For each hourly timing, rank the frequency of the transactions that are submitted.

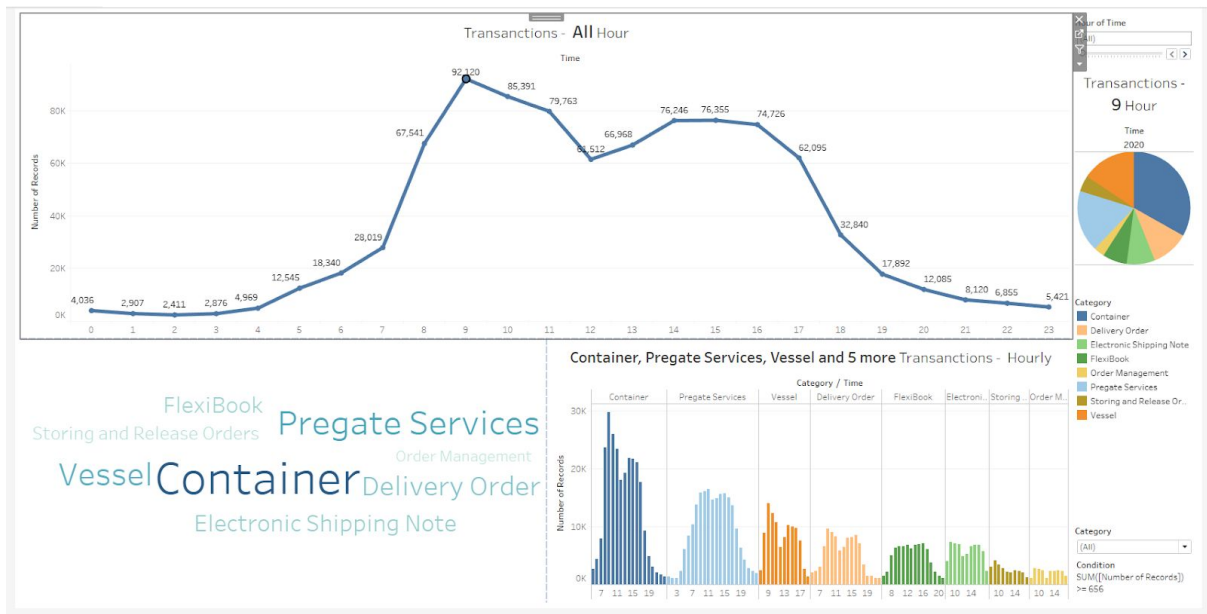
- Outcome: To understand the peak hours of the portal as well as the needs of the user during their peak hours

(*) For granularity, records showing less than 656 (20% percentile of all records) will not be shown in the Dashboard



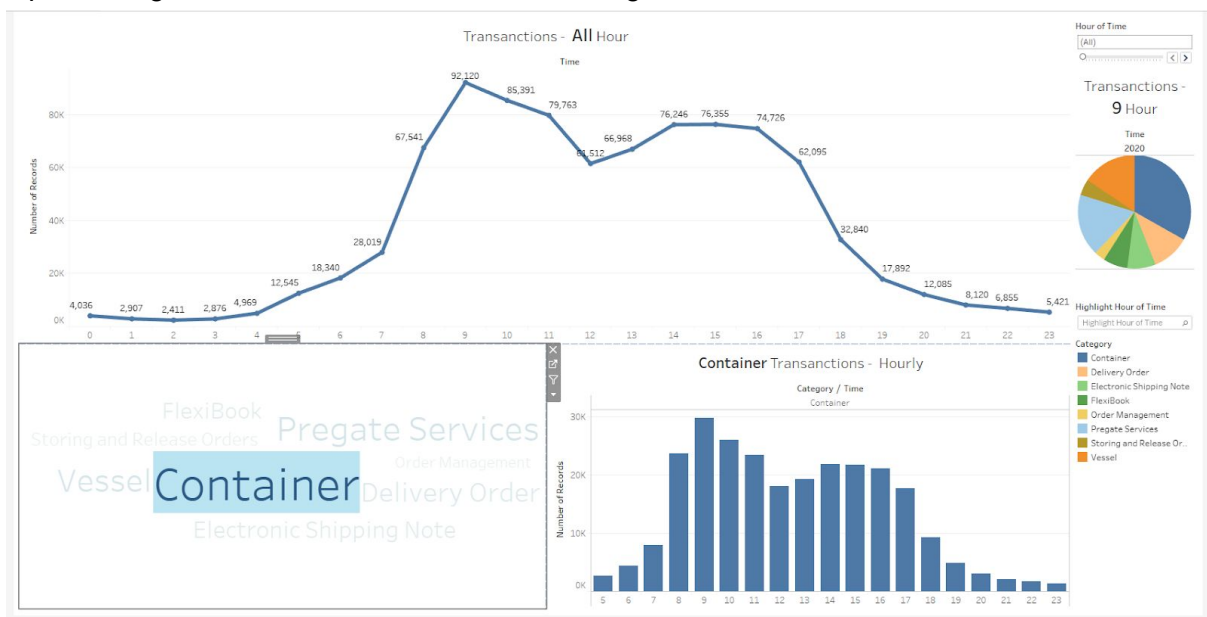
We can see the peak usage of the Portal at 9am, and activity remains high through the rest of the day. The most common Categories accessed through the Portal are shown in the word cloud to be Container and Pregate Services.

We delved further into user usage at the 9am peak hour by hovering over the 9am time in the top half of the dashboard:



9am has 92,120 records of user logs, and it appears that Container and Pregate Services are still the most commonly accessed categories, and this pattern persists through the rest of the hours.

We further investigate the access rates of the most common Category: Container. We can do so by either clicking “Container” in the Word Cloud, or the Blue part of the Pie Chart representing Container in the Pie Chart on the right



We also explored the access rates of Pregate services as it appears during off-peak hours, that transactions are largely dominated by Pregate services, however the data suggests that Pregate Services also are not frequently accessed during off-peak hours, it is just that Container Services are totally not accessed at all during off-peak hours, which is why the data appears skewed towards Pregate Services access during off-peak hours.

Further analysis and finding can be explored on your own in our attached Tableau File, Dashboard Q4. Hover your mouse over the line graph at the top of the Dashboard to get more insights into the specific hour. The Word Cloud shows the frequency of the most commonly accessed Categories, filtered by the Hour. Clicking on the Word Cloud or Pie Chart gives you more insights into that particular Category Access

(5) Based on Company ID, identify and rank the most common transactions
-Outcome: to find cross-selling and up-selling opportunities

CATEGORY	ORG1	ORG2	ORG3	ORG4	ORG5	ORG6	ORG7	ORG8	ORG9	ORG10	ORG11	ORG12
Container	16,370	10,342	14,036	23,837	13,046	10,365	3,510	9,901	4,153	9,140	17,085	6,140
Pregate Servic..	8,019	9,733	3,938	14,877	9,044	7,224	4,591	11,121	1,531	23,022	16,058	5,140
Delivery Order	5,608	4,691	5,477	3,806	5,421	2,431	1,782	4,340	1,655	2,504	2,191	2,191
FlexiBook	3,422	5,679	3,506	11,315	4,385	4,823	2,649	1,946	1,137	3,584	23	2,191
Electronic Ship..	3,264	3,812	1,904	5,382	4,430	5,012	1,701	2,607	1,078	2,379	346	1,140
Vessel	2,818	4,085	5,906	10,144	5,399	3,291	1,422	2,301	1,134	1,145	1,982	1,140
Storing and Re..	2,799	1,153	1,727	796	1,952	1,408	334	679	1,509	571	1,549	1,140
Container Para..	405	37	68	139	196	16	7	153	26	56		
Jurong Island ..	336	248			109							
Dangerous Go..	317	49	372	864	330	623	851	121		407		
Finance Manag..	92	33	81	7	6	121	4	184	73	34	85	
Resource Man..	49	96	32	23	7	50	2	4	10	117	3	
Operator	39	10	97	79	24	23	13	20		6	113	
Depot	19	67	55	1	64			14	2	11		
ChemCare Serv..	19	3	64	128	66	9	52	53		1		
Vehicle Arrival ..	18	147	43	102	34	81	8	43		90		
PORTNET Char..	12	5	4	2	2	10	3	6	3	5		
Arrival Bayplan	7	1		2	2	3			1			
Bill Manageme..	6		1	2			1					
Container Boo..	5	3	1	4	5			4				
Location	4	3	2	87	2	17	4		1	8	6	
Currency / Exc..	4			1		1						
Company Infor..	4	6	16	17	4	15	1	2	6	7	1	

We have tried to rank the most common transactions made by each organisation. For example, the most common transaction for ORG1 is Container, followed by Pregate Services and Delivery Order. For ORG10, the most common transaction was Pregate Services, followed by Container then FlexiBook.

Category and Subcategory transactions for All Organisations

CATEGORY	SUBCATEGORY	ORG1	ORG2	ORG3	ORG4	ORG5	ORG6	ORG7	ORG8	ORG9	ORG10	ORG11	ORG12
Container	Cntr Status: Lat..	10,262	6,412	10,760	14,932	9,254	7,488	2,353	7,955	2,917	5,644	9,790	2,191
	Pending Export ..	4,348	3,351	2,337	7,186	2,947	2,422	896	982	874	2,991	1,140	2,191
	Cntr Status: His..	1,252	267	680	470	671	68	56	545	253	82	451	2,191
	Haulier Truckin..	318	216	178	154	39	180		210	248	327		2,191
	Cntr In Port	126	36	1	36	38	141	187	147	39	48		2,191
	EIR Out Enquiry	27	18	15	69	38	21	11	25		24	660	2,191
	Container Survey	23	19	26	41	43	34	6	14	2	21	31	2,191
	Planning State..	5	4	3	4	4	4						2,191
	Special Contain..	4	4	15	5	4	1	1	7	1		8	2,191
	Fulfilment (QLE..	3	4	3	907	4				3			2,191
	Loading (OBOP..	1	3	5	4	3			1	10			2,191
	Loading (COPR..	1		5	3				5	2		1	2,191
	Yard Stacking			1			4						2,191
	Transhipment (..		2	2	3				2	2		2	2,191
	Terminal Depar..		4	5	21		2		7	1		2	2,191
Pregate Services	Short/Overland..					1				1	2		2,191
	Planning Summ..			1	1								2,191
	KD Movements		1		1				1		1		2,191
	Pregate : Proce..	3,894	3,317	1,868	6,808	4,632	4,618	1,460	2,438	1,003	5,327	2,500	2,191
	Report : Pregat..	2,511	3,843	1,615	3,561	2,176	589	1,472	2,293	340	8,479	5,359	2,191
Pregate Services	Report : Vehicle..	1,465	2,357	378	3,809	2,028	1,554	1,619	6,351	186	8,610	7,010	2,191
	Enquire Prime ..	144	202	69	691	198	461	40	36	2	601	1,186	2,191
	Pregate : Enquire	5	14	8	8	10	2		3		5	3	2,191

We can also see the ranking of transactions for each subcategory of a category for every organisation.

Question 6: Based on geographical location (using IP) of the User, identify their preference to mobile and web portal.

-Outcome: To understand which communication channel is more effective to different users and different regions to ensure better outreach

We combined all the PN_WEB_ACCESS files into one and we categorised them into Mobile/Web using the Browser Information column. We created a new column called “*channel*” which states whether the mobile or web portal was used. Those that had iPhone/Android in the Browser Information were counted as Mobile. We also created a new column “*Country*” where we mapped the USERIP to the actual geographical locations.

This is how our cleaned combined WEB files look like:

B	C	D	E	F	G	H	I	J	K	L	M	N	O
User.IP	User.ID	OrgCode	Browser.I	PN.URL	channel	Date	Time	hour	count	V1	V1.1	Country A	Country
14.100.24.3	enish9	ORG26	Mozilla/5.	www.port Web		7/1/2020	10H 40M 15S	10	1801	2.41E+08	2.42E+08	SG	Singapore
14.100.29.241	enish9	ORG26	Mozilla/5.	www.port Web		10/1/2020	8H 20M 55S	8	1801	2.41E+08	2.42E+08	SG	Singapore
27.54.62.49	ich	ORG37	Mozilla/5.	www.port Web		7/1/2020	11H 15M 40S	11	1774	4.57E+08	4.57E+08	SG	Singapore
27.54.62.49	ich	ORG37	Mozilla/5.	www.port Web		8/1/2020	9H 35M 52S	9	1774	4.57E+08	4.57E+08	SG	Singapore
27.54.62.49	ich	ORG37	Mozilla/5.	www.port Web		8/1/2020	9H 38M 44S	9	1774	4.57E+08	4.57E+08	SG	Singapore
27.54.62.49	ich	ORG37	Mozilla/5.	www.port Web		8/1/2020	9H 43M 14S	9	1774	4.57E+08	4.57E+08	SG	Singapore
27.54.62.49	ich	ORG37	Mozilla/5.	www.port Web		9/1/2020	10H 22M 47S	10	1774	4.57E+08	4.57E+08	SG	Singapore
27.104.164.25	ohkl0	ORG16	Mozilla/5.	www.port Mobile		8/1/2020	20H 14M 41S	20	1998	4.6E+08	4.6E+08	SG	Singapore
27.104.164.25	ohkl0	ORG16	Mozilla/5.	www.port Mobile		8/1/2020	20H 14M 58S	20	1998	4.6E+08	4.6E+08	SG	Singapore
27.104.164.25	ohkl0	ORG16	Mozilla/5.	www.port Mobile		8/1/2020	20H 15M 0S	20	1998	4.6E+08	4.6E+08	SG	Singapore
42.60.77.87	okwai	ORG13	Mozilla/5.	www.port Web		6/1/2020	4H 31M 12S	4	2152	7.09E+08	7.09E+08	SG	Singapore
42.60.77.87	okwai	ORG13	Mozilla/5.	www.port Web		6/1/2020	4H 31M 15S	4	2152	7.09E+08	7.09E+08	SG	Singapore
42.60.77.87	okwai	ORG13	Mozilla/5.	www.port Web		6/1/2020	4H 32M 13S	4	2152	7.09E+08	7.09E+08	SG	Singapore
42.60.77.87	okwai	ORG13	Mozilla/5.	www.port Web		6/1/2020	4H 32M 42S	4	2152	7.09E+08	7.09E+08	SG	Singapore
42.60.77.87	okwai	ORG13	Mozilla/5.	www.port Web		6/1/2020	4H 32M 53S	4	2152	7.09E+08	7.09E+08	SG	Singapore

Here are some of the statistics that we found:

Portal Access **Mobile/Web** (All Countries)

Country	Channel	
	Mobile	Web
Hong Kong		50
India	18	
Indonesia		18
Malaysia	90	164
Philippines		86
Singapore	3,351	113,624
United States of America		10
Viet Nam		28
Grand Total	3,459	113,980

Sum of Number of Records broken down by Channel vs. Country. Color shows details about Channel. The view is filtered on Country, which keeps 8 of 8 members.

The use of Web Access is significantly higher overall.

Singapore: A large proportion of both Mobile and Web Access comes from Singapore, with Web Access being significantly higher than Mobile Access

India: Uses only Mobile Access

Hong Kong, Indonesia, Philippines, USA, Viet Nam: Uses only Web Access

Malaysia: Uses both Mobile and Web access with preference for Web access being almost double of Mobile Access

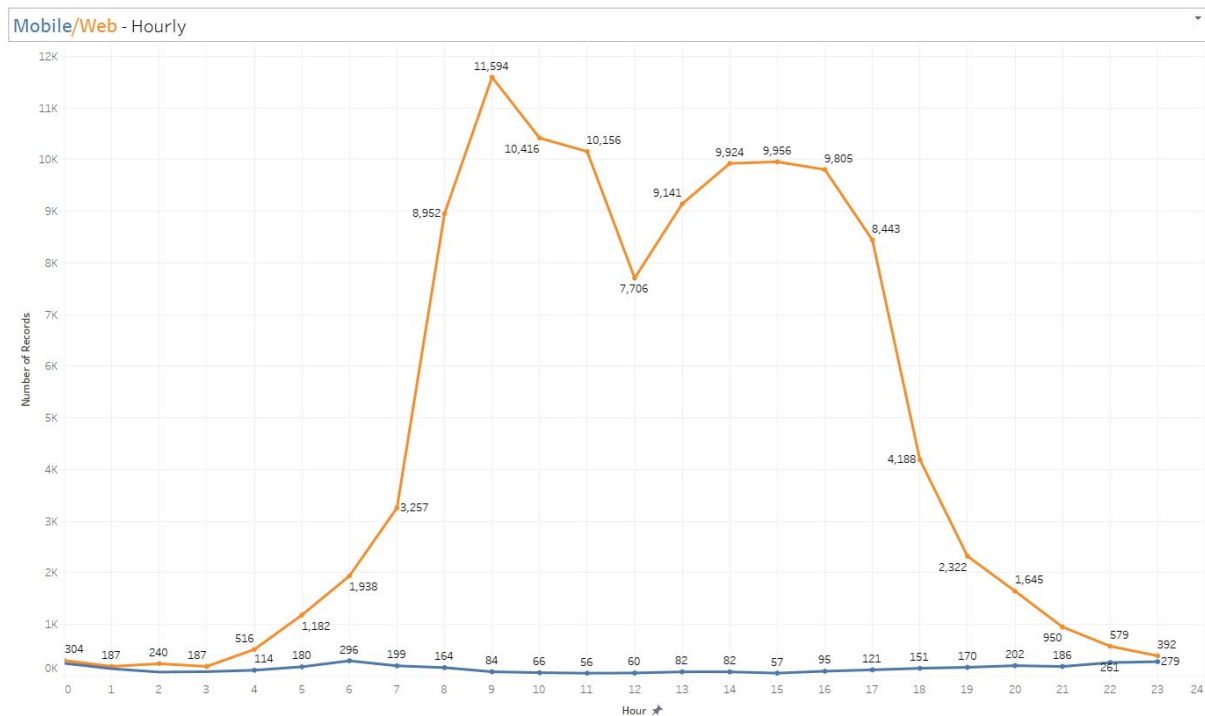
We studied the preference of Mobile/Web for each USERIP:

User IP - # of Access All Countries

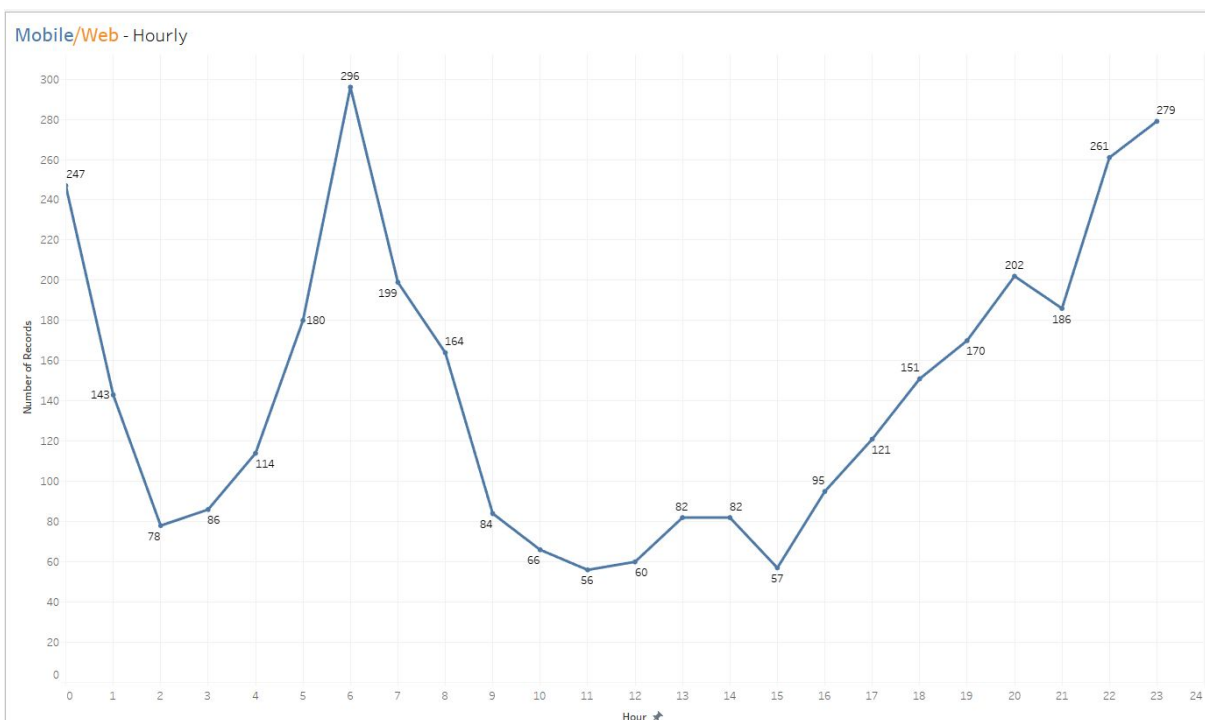
User.IP	OrgCode N	Country Abb	Channel	
			Mobile	Web
202.55.76.220	ORG4	SG		8,199
203.126.7.94	ORG5	SG		6,616
101.127.239.130	ORG2	SG		5,660
202.69.158.81	ORG10	SG		5,105
203.126.243.62	ORG24	SG		4,844
219.74.176.112	ORG8	SG		4,601
219.74.117.173	ORG6	SG		4,538
119.75.14.5	ORG3	SG		4,506
203.126.122.20	ORG11	SG		4,160
115.42.203.142	ORG1	SG		4,106
58.185.183.60	ORG7	SG		3,100
118.200.131.28	ORG17	SG		2,180
	ORG37	SG		809
58.185.82.22	ORG27	SG		2,739
118.200.217.174	ORG21	SG		2,349
116.12.179.114	ORG30	SG		2,152
128.106.99.187	ORG25	SG	3	2,029
121.6.240.230	ORG31	SG		1,996
58.185.171.50	ORG33	SG		1,988
202.73.55.206	ORG20	SG		1,949
119.73.205.34	ORG16	SG		1,932
128.106.218.85	ORG28	SG		1,784
203.117.84.4	ORG13	SG		1,743
202.14.8.26	ORG26	SG		1,706
122.11.161.167	ORG22	SG	2	1,698
118.200.177.34	ORG39	SG		1,636
118.200.119.51	ORG35	SG	2	1,587
42.61.85.90	ORG38	SG		1,519
129.126.173.131	ORG32	SG		1,417
118.200.217.79	ORG34	SG		1,331
58.185.90.50	ORG14	SG	1	1,325
103.250.59.78	ORG9	SG		1,294
203.127.61.198	ORG12	SG		1,280
203.125.56.19	ORG40	SG		1,231
122.11.147.164	ORG15	SG	9	1,214
218.212.59.16	ORG18	SG		1,177
202.79.183.46	ORG36	SG		961

We can see again, near unanimous Web Access from Singapore IP addresses.

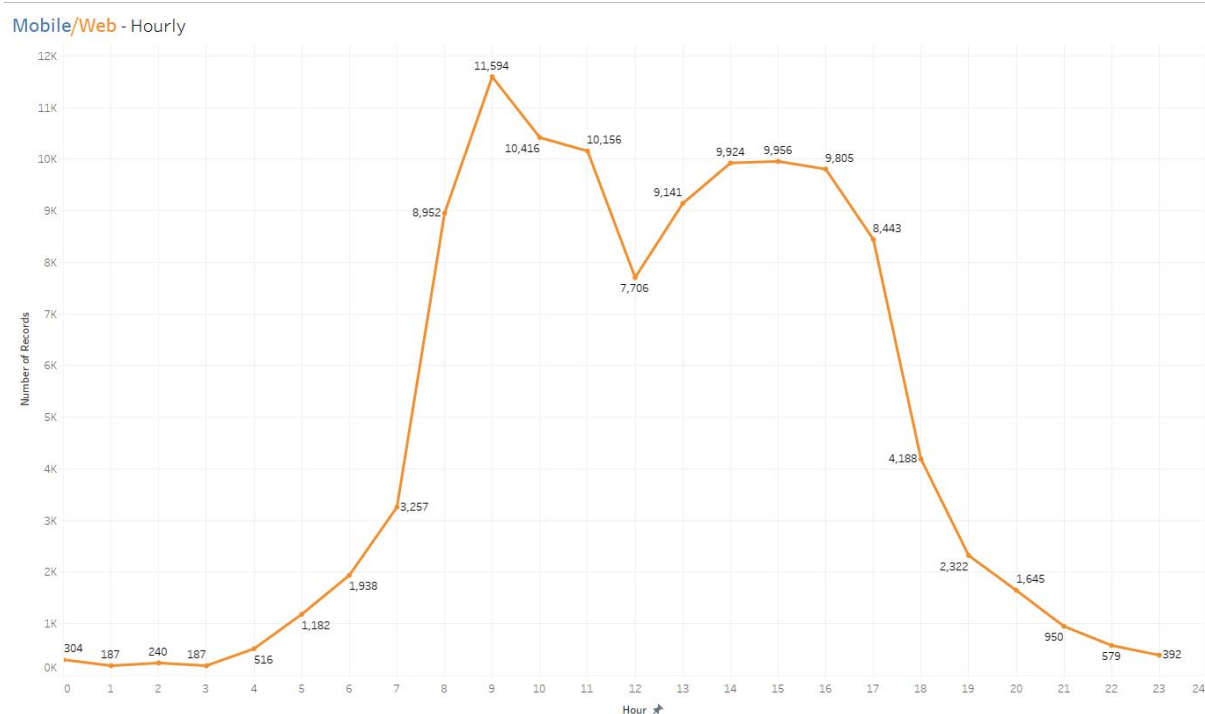
We then plotted the hourly Portal Access:



This graph shows the use of the Portal for both Mobile and Web. To study the pattern of Mobile Portal usage more clearly, we plot Mobile by itself.

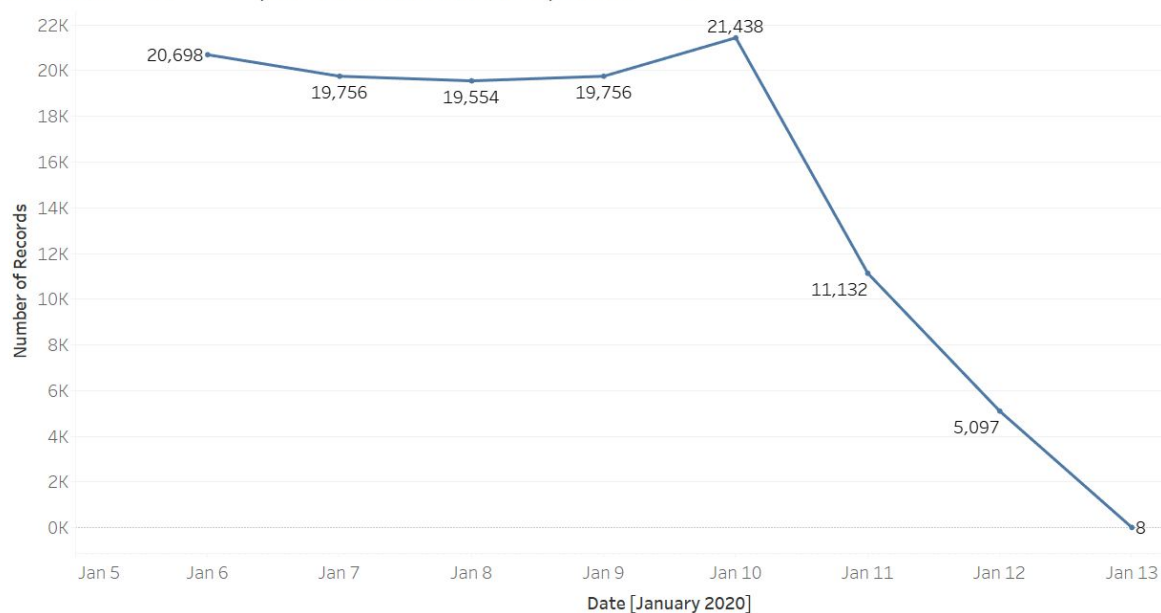


We can see from the graph that the use of the Mobile Portal is most popular in the early mornings (6am) and at night (non-office hours).



For the use of the Web Portal, they are more popular during office hours, especially at 9am and 4pm.

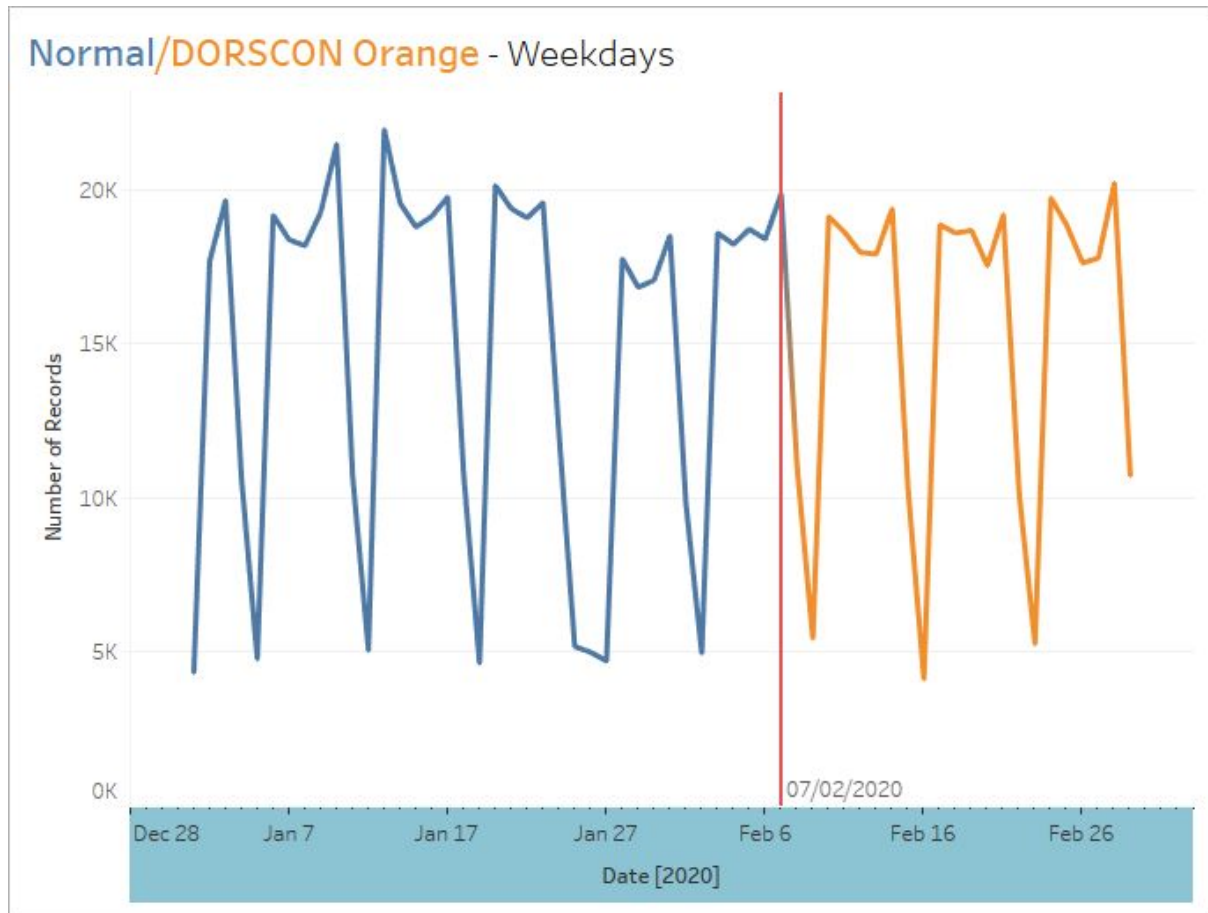
Total use of Portal (both Mobile and Web) from 6-12 Jan



This shows us the total use of the Portal from 6 Jan to 12 Jan 2020. This is a reasonable pattern as there should be more Mobile and Web Web Accesses combined in the weekdays (6-10 Jan) versus the weekend (11-12 Jan).

(7) Effects of COVID'19 on Usage Patterns

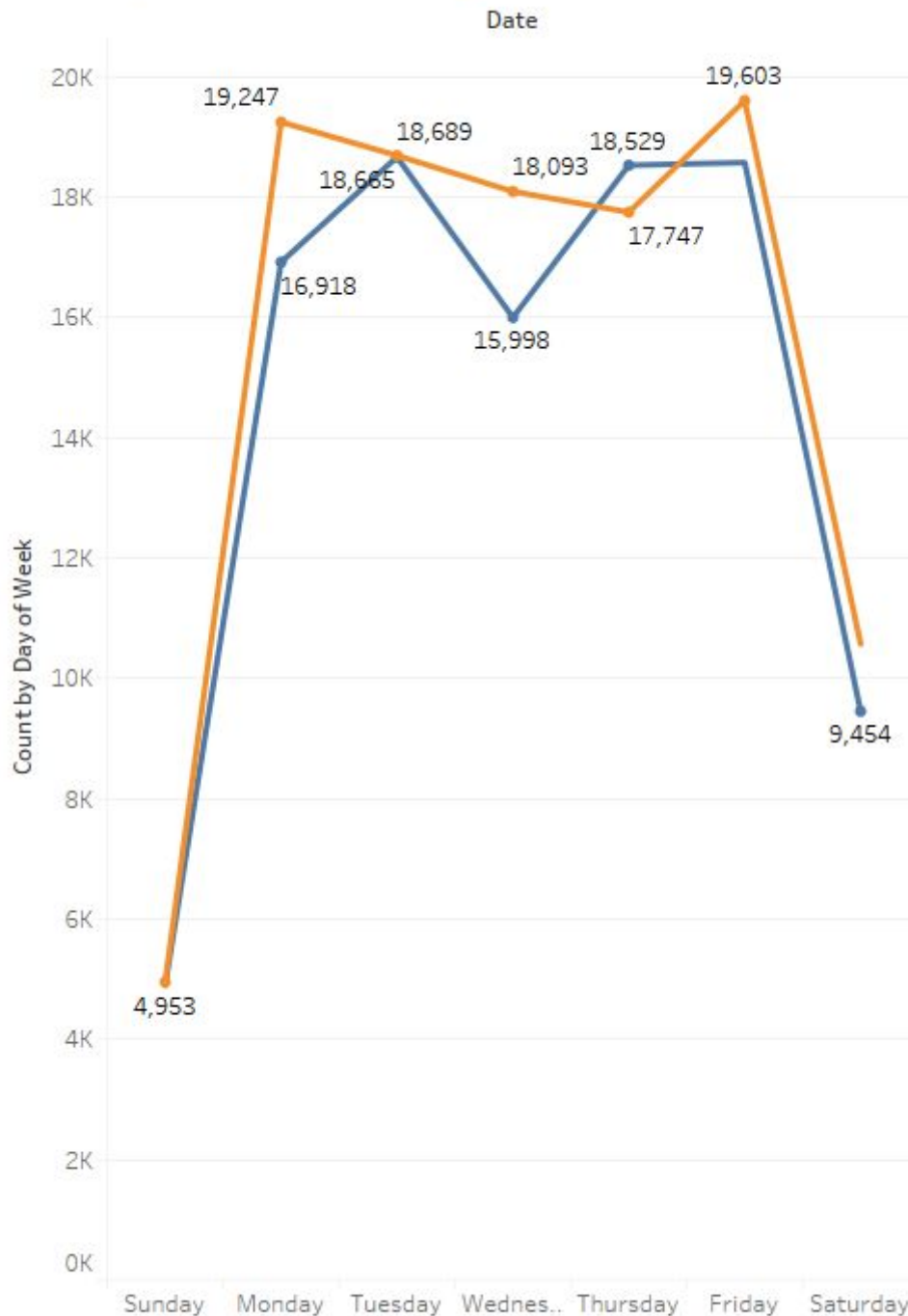
-Outcome: To understand if there are any changes in user behavior patterns with the recent outbreak



The trend of sum of Number of Records for Date Weekday. Orange color shows details after DORSCON change. The marks are labeled by sum of Number of Records.

There appears to be no significant changes in user behavior after the DORSCON change was implemented on 7th February. However we wish to investigate further:

Normal/DORSCON Orange - Weekdays Average

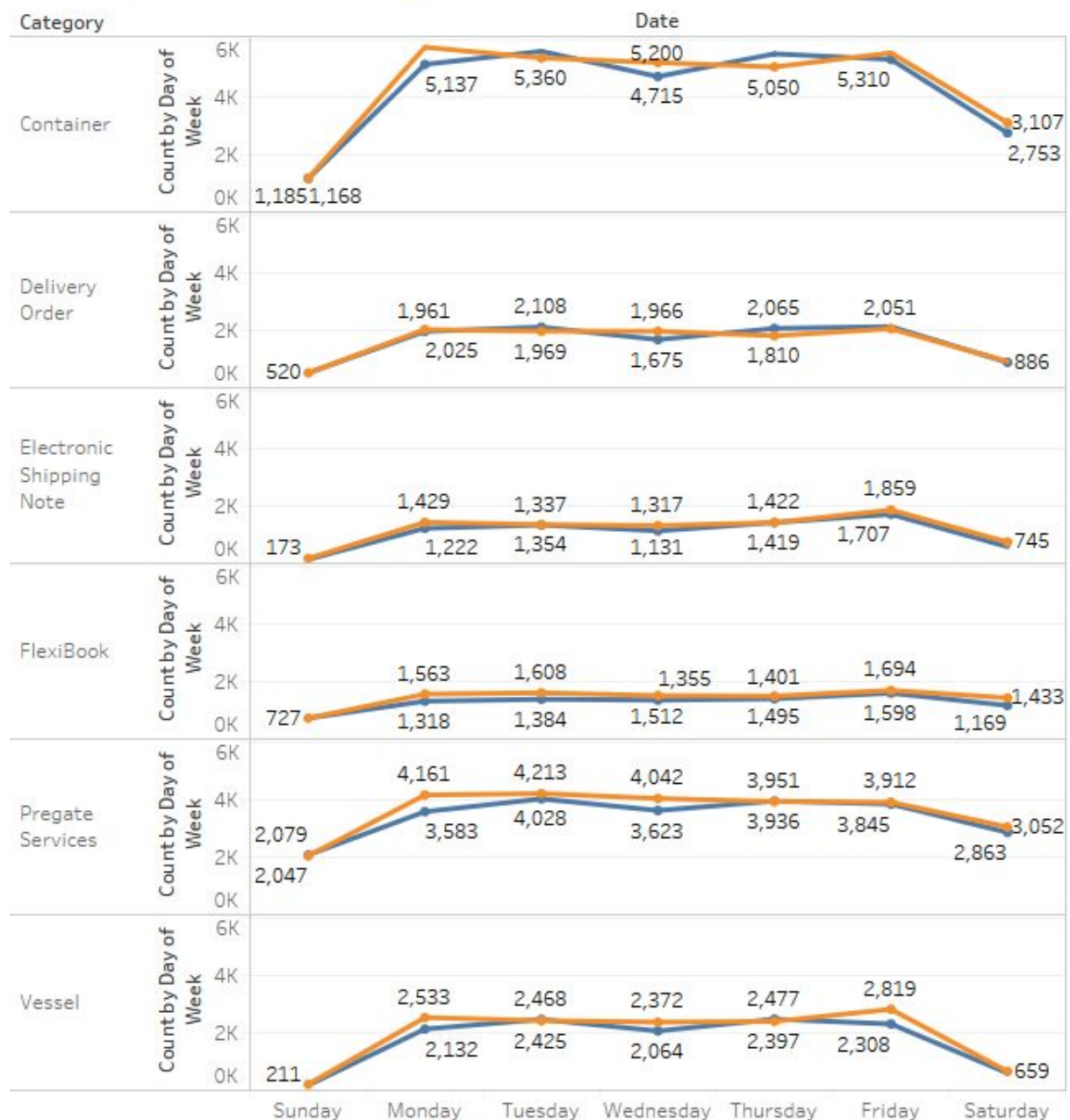


The trend of Average of Sum of Number of Records for Date per Weekday. Orange color shows details after DORSCON change. The marks are labeled by sum of Number of Records.

We plotted the Avg. Total Number of Records / day. The Orange Line shows there the Avg. Total Number of Records after the DORSCON Orange Change was announced. The Blue Line is the Avg. Total Number of Records / day before the DORSCON Change. There appears to be no change in user behavior pattern before and after the DORSCON change

We investigated this for the different most frequently accessed Categories and there also appears to be no changes in user behavior.

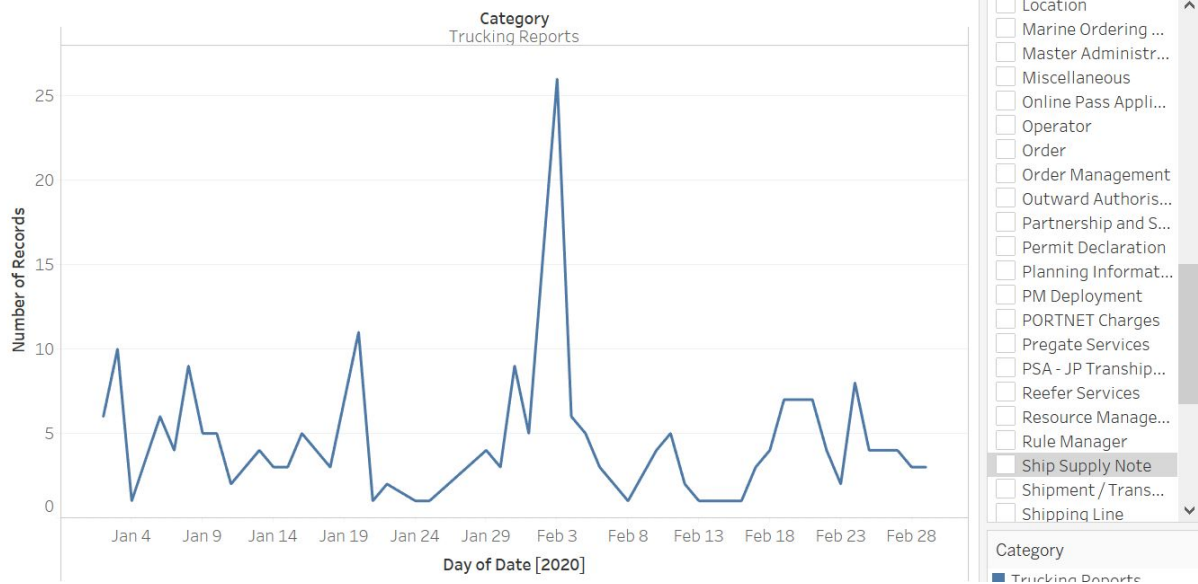
Normal/DORSCON Orange - Weekdays Average



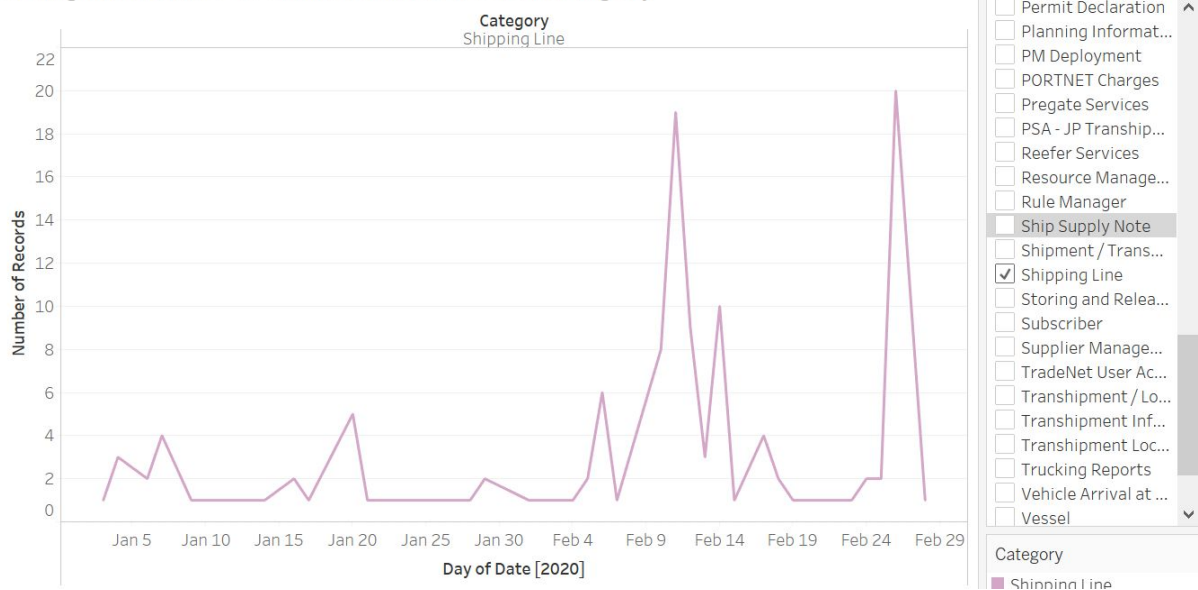
The trend of Average of Sum of Number of Records for Date per Weekday. Orange color shows details after DORSCON change. The marks are labeled by sum of Number of Records.

Interestingly, this means that the Port Community remains largely unaffected so far with regards to CoVID. However, it could also be that the effects on the Port Community is delayed by a few weeks and more recent data will reflect an impact on User Activity on Portnet.

Change in number of transactions for each category



Change in number of transactions for each category



We noticed some unusual spikes in some of the categories too such as in Trucking Reports and Shipping Line. These might have been linked to the DORSCON Orange Change, more analysis is needed.