PSA UnboXed Data Analytics Business Challenge Report

Team Corona:

Richard Lee
Teo Jao Ming
Zhuang Bao Kun

Contents:

1. Context and Objectives

Dataset:
9 different files were released to us at the start of this project. 2 of which were text files and 7 were .csv files.

The text files represented the bulk of the information that we would be working on throughout the project, consisting of the UserID, OrgCode, Actions taken in Portnet and their respective trigger times. Whereas, the .csv files consist of extra information during a week of January, to relay the platforms individuals/organisations used to access Portnet.

It was also made known to us that the files given to us mainly represented the Haulier demographic of those that used Portnet.

date	userID	orgCode	urlType	category1	category2	login
03/02/2020 09:26:27	aixin1	ORG1	webenv	Internet (Primary)	nan	0
03/02/2020 09:26:33	aixin1	ORG1	link	Vessel	Berthing (QVOY / VEDA)	0
03/02/2020 09:37:37	aixin1	ORG1	link	Storing and Release Orders	Storing Order : Enquire	0
03/02/2020 09:37:41	aixin1	ORG1	link	Storing and Release Orders	Storing Order : Enquire	0
03/02/2020 11:28:15	aixin1	ORG1	webenv	Internet (Primary)	nan	1
03/02/2020 11:28:41	aixin1	ORG1	link	Electronic Shipping Note	Cntr Terminal : Create	1
03/02/2020 16:19:09	aixin1	ORG1	webenv	Internet (Primary)	nan	2
03/02/2020 16:19:16	aixin1	ORG1	link	Electronic Shipping Note	Cntr Terminal : Update / Delete	2
03/02/2020 17:22:08	aixin1	ORG1	webenv	Internet (Primary)	nan	3
03/02/2020 17:22:10	aixin1	ORG1	link	Vessel	Berthing (QVOY / VEDA)	3
03/02/2020 17:22:12	aixin1	ORG1	link	Vessel	Berthing (QVOY / VEDA)	3
06/02/2020 09:18:08	aixin1	ORG1	webenv	Internet (Primary)	nan	4
06/02/2020 09:18:46	aixin1	ORG1	link	Vessel	Berthing (QVOY / VEDA)	4
06/02/2020 09:19:28	aixin1	ORG1	link	Container	Cntr Status: Historical (DCON)	4
06/02/2020 09:19:33	aixin1	ORG1	link	Container	Cntr Status: Latest (DCON)	4
06/02/2020 10:24:26	aixin1	ORG1	link	Vessel	Berthing (QVOY / VEDA)	4
06/02/2020 10:25:01	aixin1	ORG1	link	Electronic Shipping Note	Cntr Terminal : Create	4
06/02/2020 10:25:05	aixin1	ORG1	link	Electronic Shipping Note	Cntr Terminal : Create	4
06/02/2020 10:25:20	aixin1	ORG1	link	Electronic Shipping Note	Cntr Terminal : Create	4
06/02/2020 10:31:29	aixin1	ORG1	link	Electronic Shipping Note	Cntr Terminal : Create	4
06/02/2020 11:51:42	aixin1	ORG1	webenv	Internet (Primary)	nan	5
06/02/2020 11:53:05	aixin1	ORG1	link	Vessel	Berthing (QVOY / VEDA)	5
06/02/2020 13:29:28	aixin1	ORG1	webenv	Internet (Primary)	nan	6

The figure above shows the way we have cleaned and organized the text data that was released to us.

Objective:
The main objective for us was to explore the data and uncover patterns that might not be so obvious from superficial viewing.

2. Exploratory Data Analysis

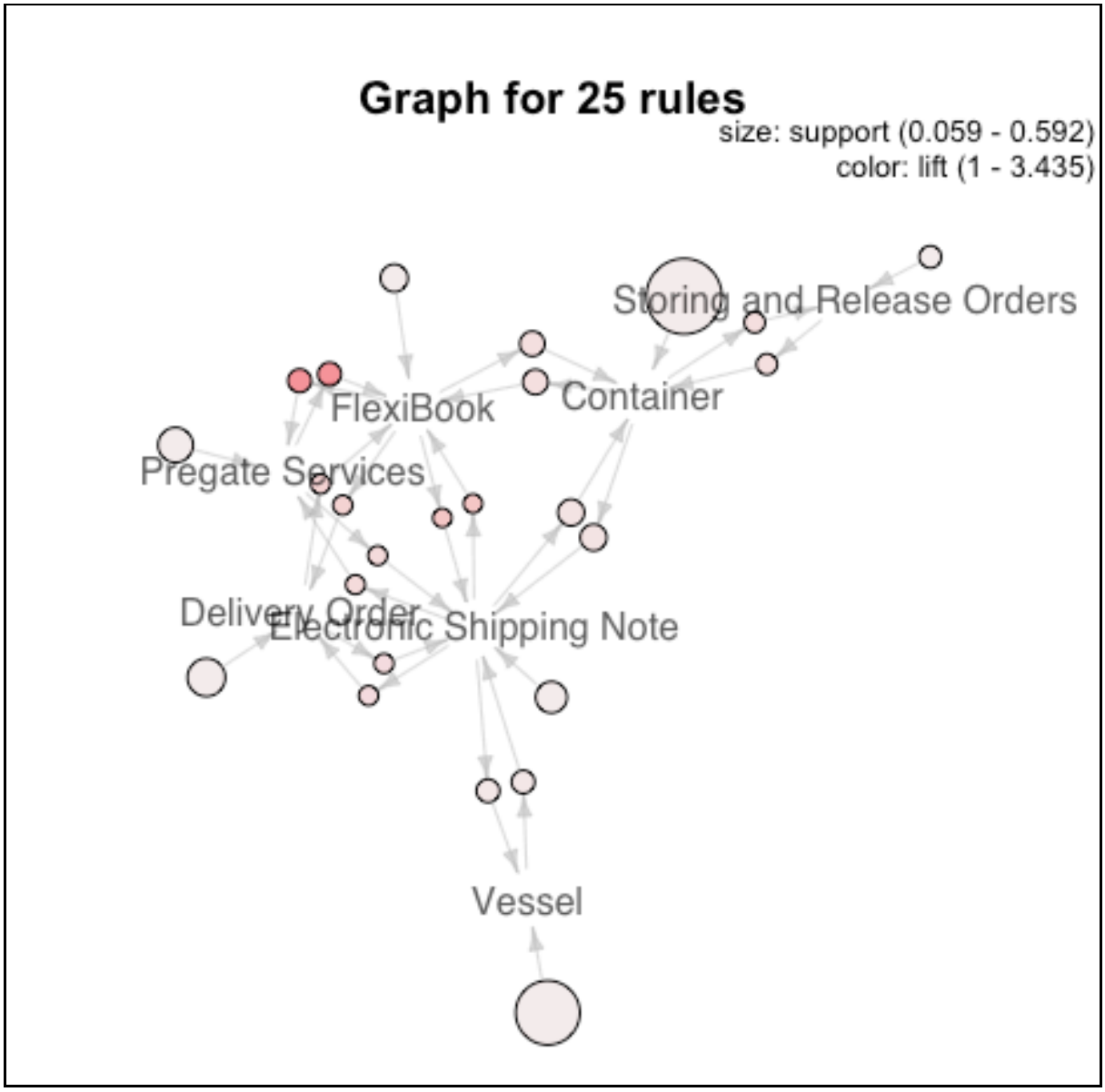
2.1 User Patterns:

Analysis:
For finding out user patterns, we took an association rule approach. Meaning to say, we found all the unique categories and subcategories throughout the whole dataset and have them be variables to indicate whether a user has accessed it during a log-in or not. A log-in is defined by when a unique user logs in to the portal till their last action from that log-in.

Based on these associations, we would use the apriori algorithm to determine which categories were more frequently accessed and how it compares and relates to the others. The figure below shows a list of relationships between categories alongside their support, confidence and lift. For example, those that accessed the `Container` and `Pregate Services` category are likely to also access the `Delivery Order` and `FlexiBook` category. Based on this example, the 2 pairs occur together 1040692*0.055(support) = 57238 number of times, and either of the pair is ~5(lift) times more likely to happen if the other of the pair occurs.

antecedents	consequents	support	confidence	lift
frozenset({'Container', 'Pregate Services'})	frozenset({'Delivery Order', 'FlexiBook'})	0.05469771630972279	0.3368712998026562	5.163794751030467
frozenset({'Delivery Order', 'FlexiBook'})	frozenset({'Container', 'Pregate Services'})	0.05469771630972279	0.838444178945971	5.1637947510304665
frozenset({'Container', 'FlexiBook'})	frozenset({'Delivery Order', 'Pregate Services'})	0.05469771630972279	0.4669870609981516	4.67274375394433
frozenset({'Delivery Order', 'Pregate Services'})	frozenset({'Container', 'FlexiBook'})	0.05469771630972279	0.547313691507799	4.67274375394433
frozenset({'FlexiBook'})	frozenset({'Delivery Order', 'Container', 'Pregate Services'})	0.05469771630972279	0.39021376498208327	4.325465456105034
frozenset({'Delivery Order', 'Container', 'Pregate Services'})	frozenset({'FlexiBook'})	0.05469771630972279	0.6063165978688683	4.325465456105034
frozenset({'Pregate Services'})	frozenset({'Delivery Order', 'Container', 'FlexiBook'})	0.05469771630972279	0.2589373565103313	4.15101422929397
frozenset({'Delivery Order', 'Container', 'FlexiBook'})	frozenset({'Pregate Services'})	0.05469771630972279	0.8768568651950577	4.15101422929397
frozenset({'Delivery Order', 'FlexiBook'})	frozenset({'Pregate Services'})	0.05673285933264629	0.8696402495685649	4.116850986304188
frozenset({'Pregate Services'})	frozenset({'Delivery Order', 'FlexiBook'})	0.05673285933264629	0.2685716628402755	4.116850986304188
frozenset({'FlexiBook'})	frozenset({'Delivery Order', 'Pregate Services'})	0.05673285933264629	0.4047324848634622	4.0498149705085655
frozenset({'Delivery Order', 'Pregate Services'})	frozenset({'FlexiBook'})	0.05673285933264629	0.5676776429809358	4.0498149705085655
frozenset({'FlexiBook'})	frozenset({'Container', 'Pregate Services'})	0.09069809735777815	0.6470406524156679	3.984982195055181
frozenset({'Container', 'Pregate Services'})	frozenset({'FlexiBook'})	0.09069809735777815	0.5585897914555443	3.984982195055181
frozenset({'Container', 'FlexiBook'})	frozenset({'Pregate Services'})	0.09069809735777815	0.7743438077634012	3.665720475001956
frozenset({'Pregate Services'})	frozenset({'Container', 'FlexiBook'})	0.09069809735777815	0.42936208592981306	3.6657204750019554
frozenset({'FlexiBook'})	frozenset({'Pregate Services'})	0.10170519004771761	0.7255653033485728	3.434804490938138
frozenset({'Pregate Services'})	frozenset({'FlexiBook'})	0.10170519004771761	0.4814693342079371	3.434804490938138
frozenset({'FlexiBook'})	frozenset({'Electronic Shipping Note', 'Container'})	0.05332074720059582	0.38039046089212897	2.894120505348555
frozenset({'Electronic Shipping Note', 'Container'})	frozenset({'FlexiBook'})	0.05332074720059582	0.40567964683402513	2.894120505348555
frozenset({'FlexiBook', 'Pregate Services'})	frozenset({'Delivery Order', 'Container'})	0.05469771630972279	0.5378065395095368	2.85050302596648
frozenset({'Delivery Order', 'Container'})	frozenset({'FlexiBook', 'Pregate Services'})	0.05469771630972279	0.28991095198751493	2.8505030259664794
frozenset({'Electronic Shipping Note'})	frozenset({'Container', 'FlexiBook'})	0.05332074720059582	0.3006934948232077	2.5671998921057755
frozenset({'Container', 'FlexiBook'})	frozenset({'Electronic Shipping Note'})	0.05332074720059582	0.4552310536044362	2.5671998921057755
frozenset({'Delivery Order'})	frozenset({'Container', 'FlexiBook', 'Pregate Services'})	0.05469771630972279	0.22929751316028318	2.528140279015665
frozenset({'Container', 'FlexiBook', 'Pregate Services'})	frozenset({'Delivery Order'})	0.05469771630972279	0.6030745727107801	2.528140279015665
frozenset({'FlexiBook'})	frozenset({'Delivery Order', 'Container'})	0.06237929869837448	0.4450142098109477	2.358681530390156
frozenset({'Delivery Order', 'Container'})	frozenset({'FlexiBook'})	0.06237929869837448	0.3306251721288901	2.358681530390156
frozenset({'FlexiBook'})	frozenset({'Electronic Shipping Note'})	0.05855149777866304	0.4177066600766094	2.355587309323411
frozenset({'Electronic Shipping Note'})	frozenset({'FlexiBook'})	0.05855149777866304	0.3301914436413362	2.3555873093234108
frozenset({'FlexiBook', 'Pregate Services'})	frozenset({'Delivery Order'})	0.05673285933264629	0.5578167574931879	2.338415640025264 3

The graph plot below shows a graphical view of what's happening in the table above. We can see clusters forming in the graph of the top 25 associations; `Electronic Shipping Note`, `Pregate Services`, `FlexiBook`, `Delivery Order`, and `Container`. These seem to be the main categories that users circle around and we would be safe to say or at least assume that they may be the main reasons for the Portnet platform, at least for the Haulier demographic.

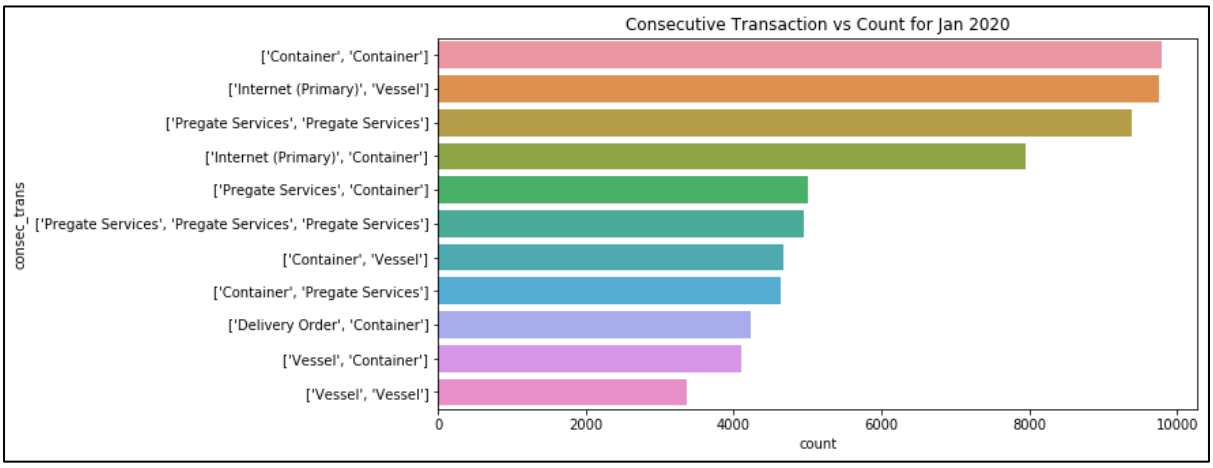


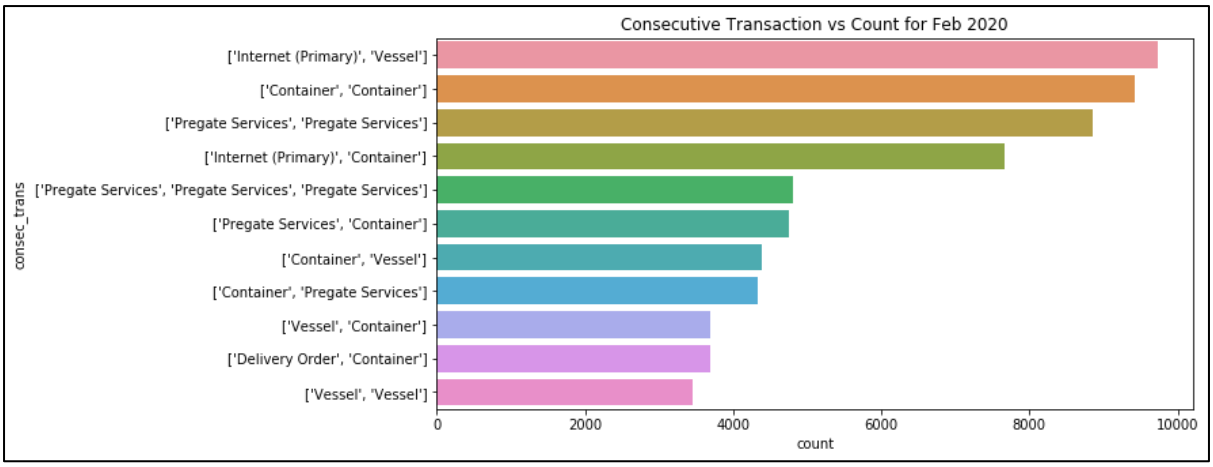
Recommendation:
With this evaluation, we will be able to link the categories together to come up with a way to have the user experience be more efficient. For example, we could bundle 5 categories that were listed above and place them in a **specialized tab** with revamped functions that help each user get whatever they want done quicker. Moreover, if a specialized tab is not possible, it may also be beneficial to push for **recommendation categories** for easier access, especially for those that use any of the 5 categories. Take for instance someone that is using the `Container` tab. Whilst using the `Container` tab, there could be express buttons that would lead the user to access the other 4 main categories that were associated with each other based on the apriori algorithm.

2.2 Consecutive Transaction:

Analysis:

Consecutive transactions are defined to be transactions that are within the time difference of 10 minutes with respect to the first click. To obtain consecutive transactions, we first have to filter by each userID. Next, we compile each instance of consecutive transactions. For example, if the transaction of the first click is related to “Container” and the next click that is performed within 10 minutes is related to “Container” , the consecutive transaction will be reflected as [“Container”, “Container”].





We tabulated the total count of consecutive transactions for January 2020 and February 2020 to determine if there is any significance difference in the total count and types of consecutive transactions in different months. From the above bar plot, we can see that [“Container”, “Container”] and [“Internet (Primary)”, “Vessel”] are consistently the top 2 types of consecutive transactions. Next, we further zoom into the sub-category of these transactions to understand the breakdown of transactions.

category2	count
Cntr Status: Latest (DCON)	275392
Pending Export Containers	144614
Cntr Status: Historical (DCON)	18851
Haulier Trucking Events	5363
Cntr In Port	5276

The table above shows the top 3 most common category2 transactions for the case of ["Container", "Container"] consecutive transactions. "Cntr Status" and "Pending Export Containers" have significantly high counts in the consecutive transactions.

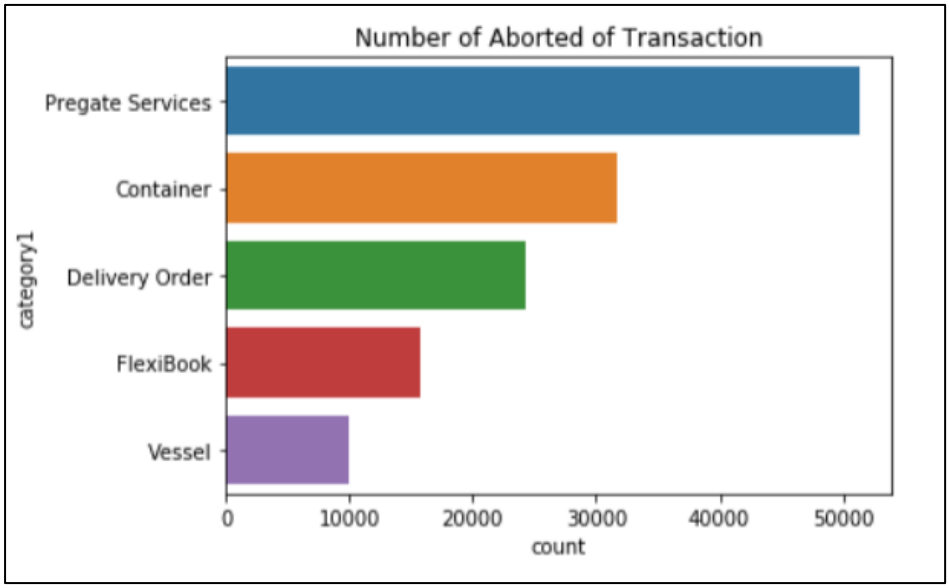
category2	count
Berthing (QVOY / VEDA)	25820
Berthing Schedule (VSCH)	2978
Sailing Schedule	40

The table above shows that for the consecutive transactions of ["Internet (Primary)", "Vessel"], the "category2" transactions that occur the most frequently are related to "Berthing".

Recommendations:
Based on the high frequency of consecutive transactions that contain status of container, PSA can create a regular auto-update system that can inform the clients the status of the container so that clients do not need to manually check for the container status. The auto-update system could be in the form of message or email directly to the client.

2.3 Abortion of Transactions:

Analysis:
As there is no explicit column that clearly indicates the case of an aborted transaction, we set the following conditions for rows to be identified as aborted transactions. The conditions are when there is a back to back transaction from belong to the same category (i.e. “category1”) within 1 minute for an unique user.



The table above shows the top 5 most common aborted transactions which are Pregate Services, Container, Delivery Order, Flexibook and Vessel.

Recommendation:

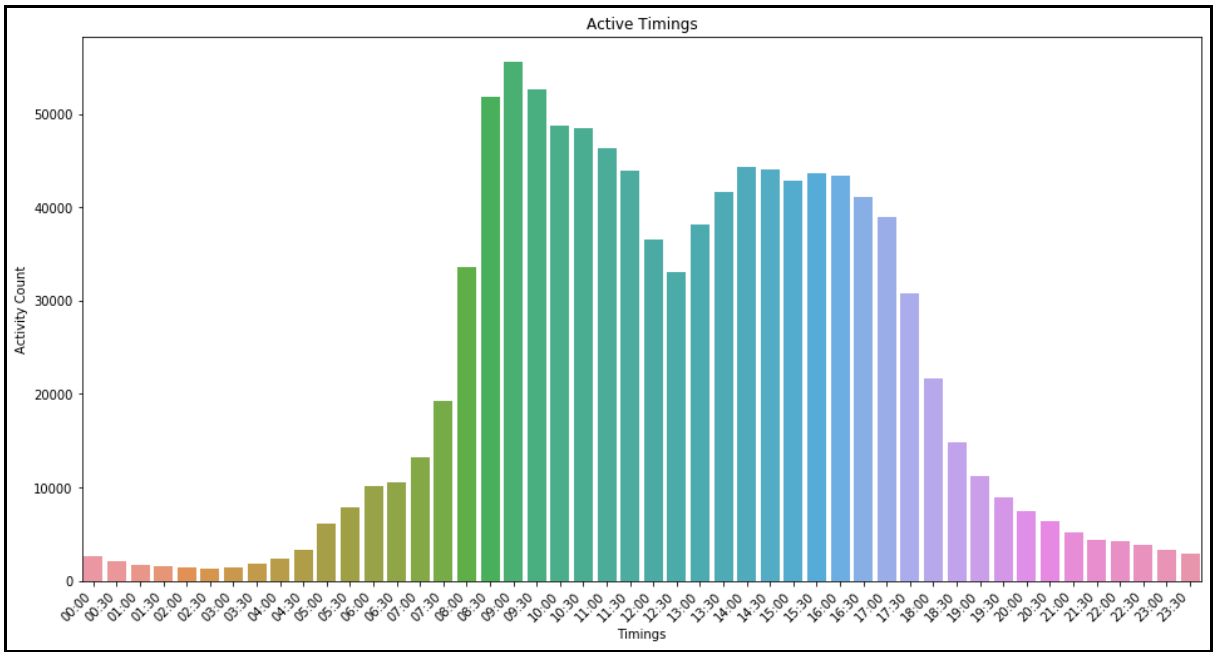
category1	category2	count
Delivery Order	Process Delivery / Release	23110
Pregate Services	Report : Vehicle In Port	22733
Pregate Services	Report : Pregate Summary	19246
Container	Cntr Status: Latest (DCON)	15805
Container	Pending Export Containers	11159

Next, we zoomed into subcategory(i.e. “category2) to identify which specific area that has the highest frequency of aborted. These are the specific areas PSA can look into to channel its resources to improve the user interface of the transaction system.

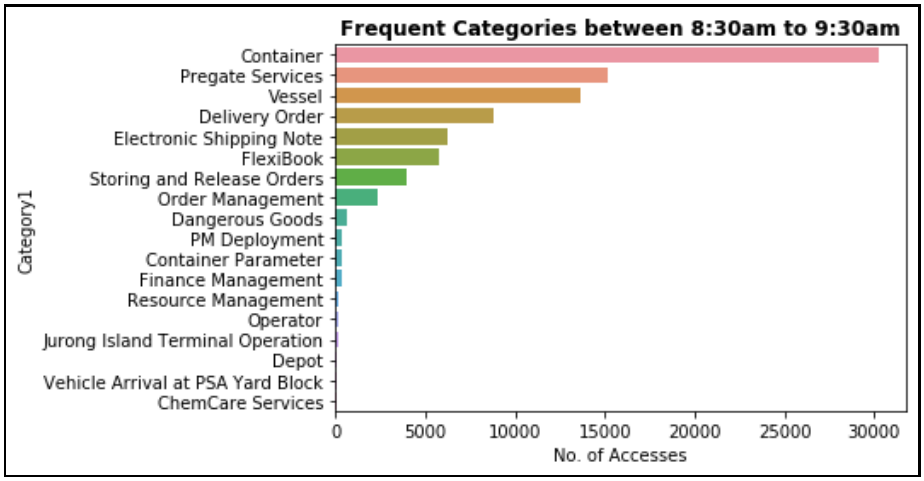
2.4 Peak and Non-Peak Transaction Periods:

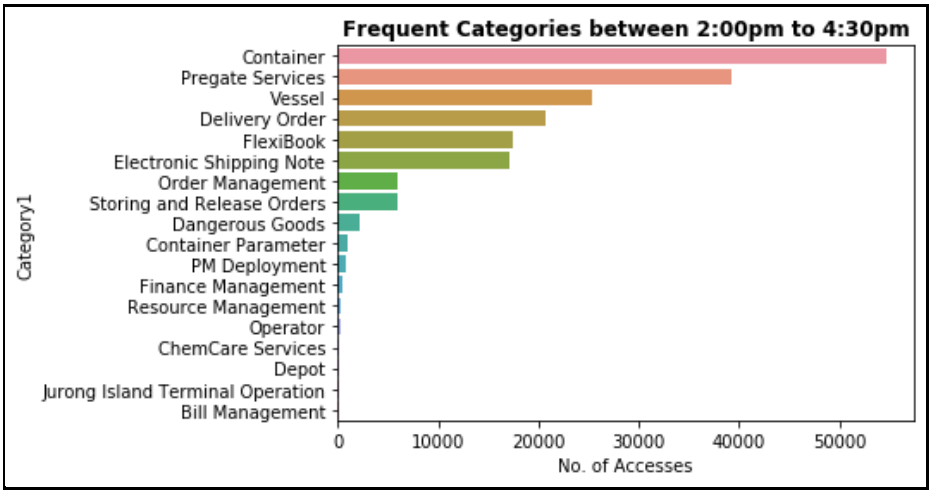
Analysis:

We analyse the timings of which each user would access Portnet. The figure below shows all the active timings of all users and organisations.

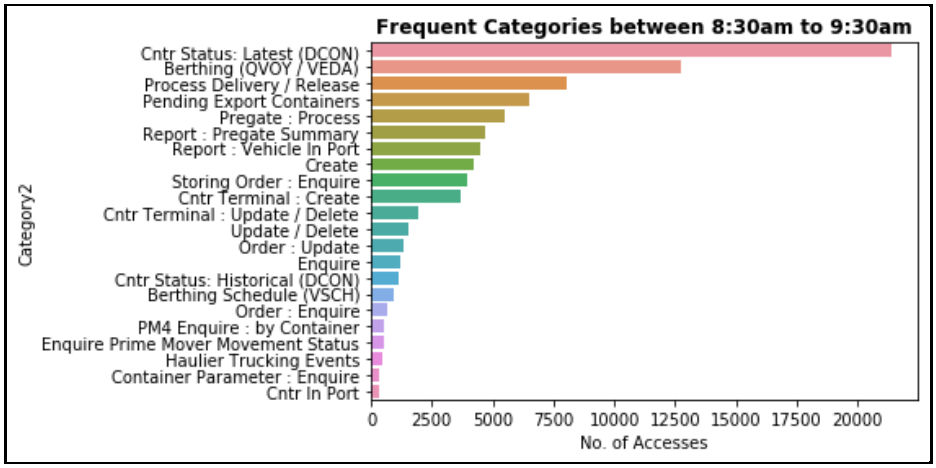


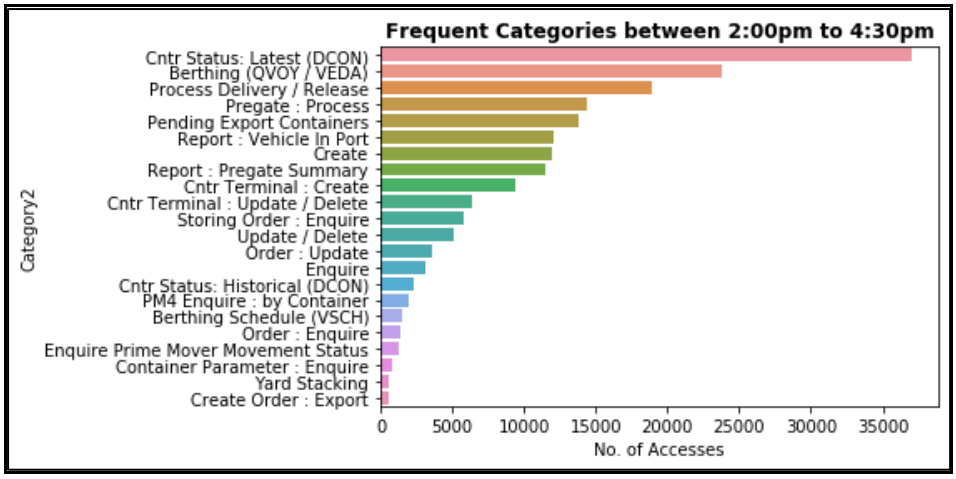
It can be seen that the active timings spike at 2 points throughout the day; 8:30am - 9:30am and 2:00pm to 4:30pm. Now that we know the peak timings, we decide to look at what categories users usually engage during these periods.





The 2 figures above display the most frequent main categories that were accessed during the peak periods of the day. It is not surprisingly that the categories accessed during both periods are similar. Hence, we decide to take a look at the subcategories instead.



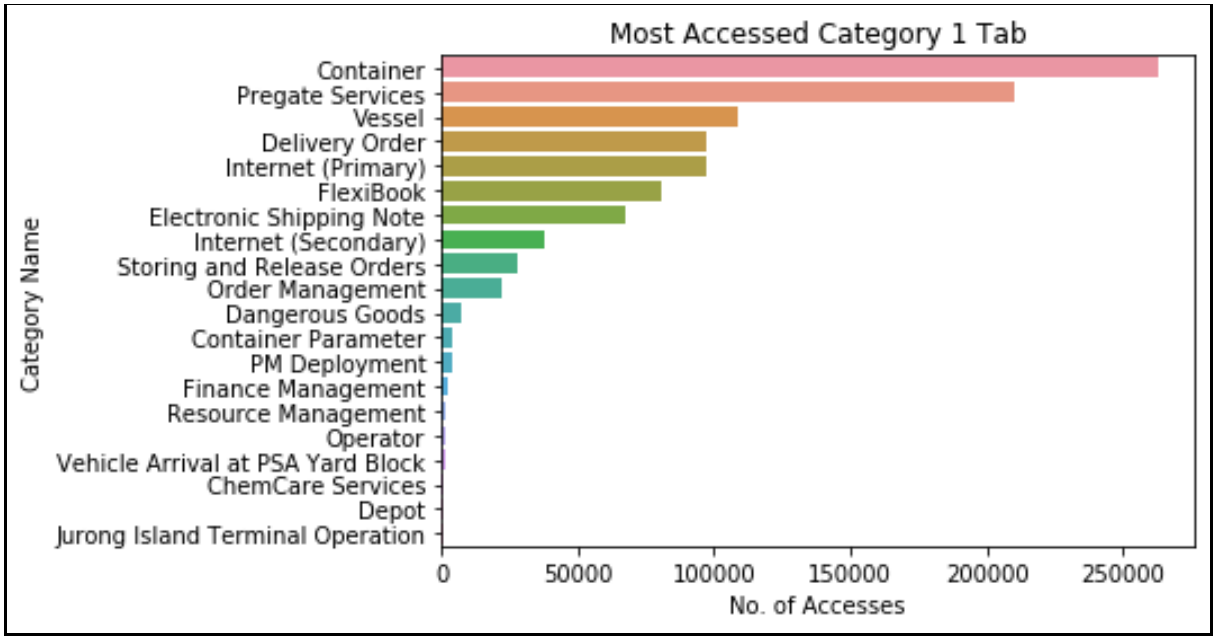


Something to note from this evaluation is that `Cntr Status` seems to be a very very high in use category. This seems to be consistent throughout all the other evaluations that we did as well. Hence, it might be worth looking into how we can make it more easily accessible and simpler to engage for the users.

Recommendations:
Although we may not have found striking insights from the timings, there is potentially more that we could look into with regards to the peak and non-peak hours if we were given more time.

2.5 Transaction Frequencies:

Analysis:
Looking into the frequency of categories accessed by each user during a session, we can find out what the majority of users do during a particular session. Based on the text data that we've been given, we can break down the categories that each user has accessed into category 1 and 2; a main category and a sub-category.

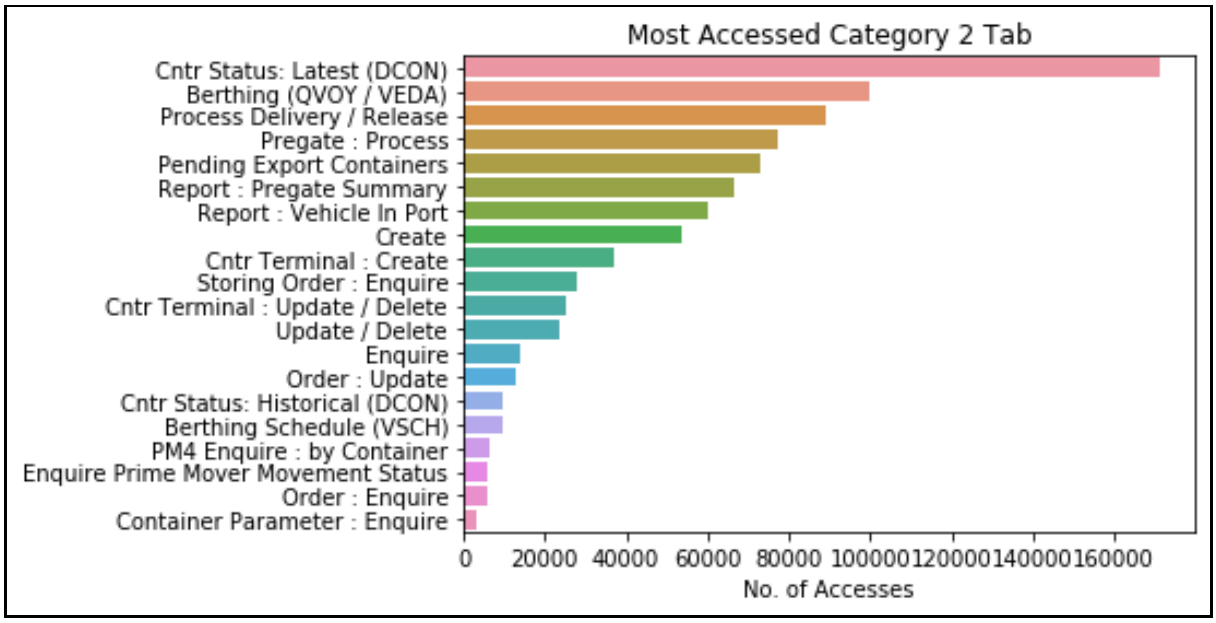


As seen from the figure above, the most frequently visited main category at Portnet by Hauliers is the `Container` tab; followed by `Pregate Services` and `Vessel`. This is unsurprising as containers are the key focus of the Haulier demographic. Taking a deeper look into the top 2 most accessed categories, `Container` and `Pregate Services`, we can further understand the reasons as to why these 2 categories are frequently accessed.

		userID
category1	category2	
Container	Cntr In Port	2686
	Cntr Status: Historical (DCON)	9589
	Cntr Status: Latest (DCON)	171138
	Container Survey	572
	EIR Out Enquiry	1369
	Fulfilment (QLEC)	1038
	Haulier Trucking Events	3088
	KD Movements	108
	Loading (COPRAR)	38
	Loading (OBOP / ESUM / ESOP)	53
	Pending Export Containers	73087
	Planning Statement (FIST)	49
	Planning Summary (FINS)	6
	Short/Overlanded (QCOS)	10
	Special Containers (LISP)	111
	Terminal Departure (TDRS)	73
	Transhipment (SAST)	23
	Yard Stacking	7
Pregate Services	Enquire Prime Mover Movement Status	6074
	Pregate : Enquire	200
	Pregate : Process	77508
	Report : Pregate Summary	66502
	Report : Vehicle In Port	60132

Looking at the figure above, we can see that users would usually access `Cntr Status: Latest (DCON)` at 171138 accesses and `Pending Export Containers` at 73087 accesses for the main category of `Container`. `Pregate : Process` at 77508 accesses, `Report : Pregate Summary` at 66502 accesses, and `Report : Vehicle In Port` at 60132 accesses.

Next, we do further exploration with the frequency of accesses of the sub-category, or category 2. The figure below shows the number of accesses of each sub-category tab, regardless of the main category that's been accessed.



Looking at the aforementioned figure, we notice that there are 2 sub-categories that were not within the subcategories of the top 2 main categories shown previously. They are `Berthing (QVOY / VEDA)` and `Process Delivery / Release`. This could possibly mean that although the users did not click on their main tabs often, they accessed these 2 sub-category tabs very often whenever they're within their main tabs already.

Recommendation:
All in all, from looking at the frequency pattern of the main categories, it could be useful to create a **dashboard** on the main page of Portnet. Given that the request to look at the status of containers and pending containers are of high demand, it would be useful to have those data at the forefront of the platform. Assuming `Pregate : Process` is an action base subcategory, there's not much we can do other than to create a quick-link to that action trigger. Lastly for the main category aspect, it seems that the generating of the reports, namely the Pregate Summary and the Vehicles in Port, seem to be used frequently. If it were common for these reports to be accessed within the same session, it would prove useful to be able to somehow merge the process of generating both reports at the same time. Moreover, in terms of efficiency and convenience, it is also possible to **recommend** the viewing of the Pregate Summary report if one is requesting for a report on the Vehicles in Port, and vice versa. Either of the reports could be pre-loaded as well for efficiency sake.

On the other hand, when it comes to the sub-categories, now that we know the sub-categories that are frequently clicked on given that its respective main category that are not, we can consider **automating** those requests. For example, given that the processing of delivery and release is clicked on so frequently, it might be more convenient and faster if all the user had to do was to scan something or upload an excel sheet for the platform to automate the processing of delivery and release for multiple containers at a time.

2.6 User Platform Access Preferences:

Analysis:
A total of 1001 unique users have used (both) platforms across the 1 week of data that was provided. Of the 1001 users, 119 (11.9%) users have accessed the portal through a mobile device.. The usage of the platform by device is summarised below:

```
Total unique users: 1001
--- Platform usage (total) ---
Windows: 99.00%
iOS: 5.19%
Mac OS X: 0.70%
Android: 7.09%
Linux: 0.10%
```

The percentages do not add up to 100% since a user may use more than 1 platform. It can be seen that a large proportion of the users use Windows to access the portal, with the mobile options of Android and iOS trailing behind with single digit percentages. This large usage of Windows can be attributed to employees of the various companies accessing the portal through their office computers, which most likely are Windows workstations.

Should we focus on the group of users that require information on the go and hence have accessed the platform on their mobile devices to acquire information, we have the following summary data:

```
Total unique users that use a mobile platform: 119
--- Platform usage (mobile) ---
Windows: 92.44%
iOS: 43.70%
Mac OS X: 1.68%
Android: 59.66%
Linux: 0.00%
```

iOS sees usage of about 43%, with Android having a usage of 60%. This indicates a slightly larger proportion of the users that rely on their mobile phones for information.

Recommendation:
It can be seen that quite a small percentage (10%) of the total users access the platform via their mobile phones. We attribute this to two main reasons:

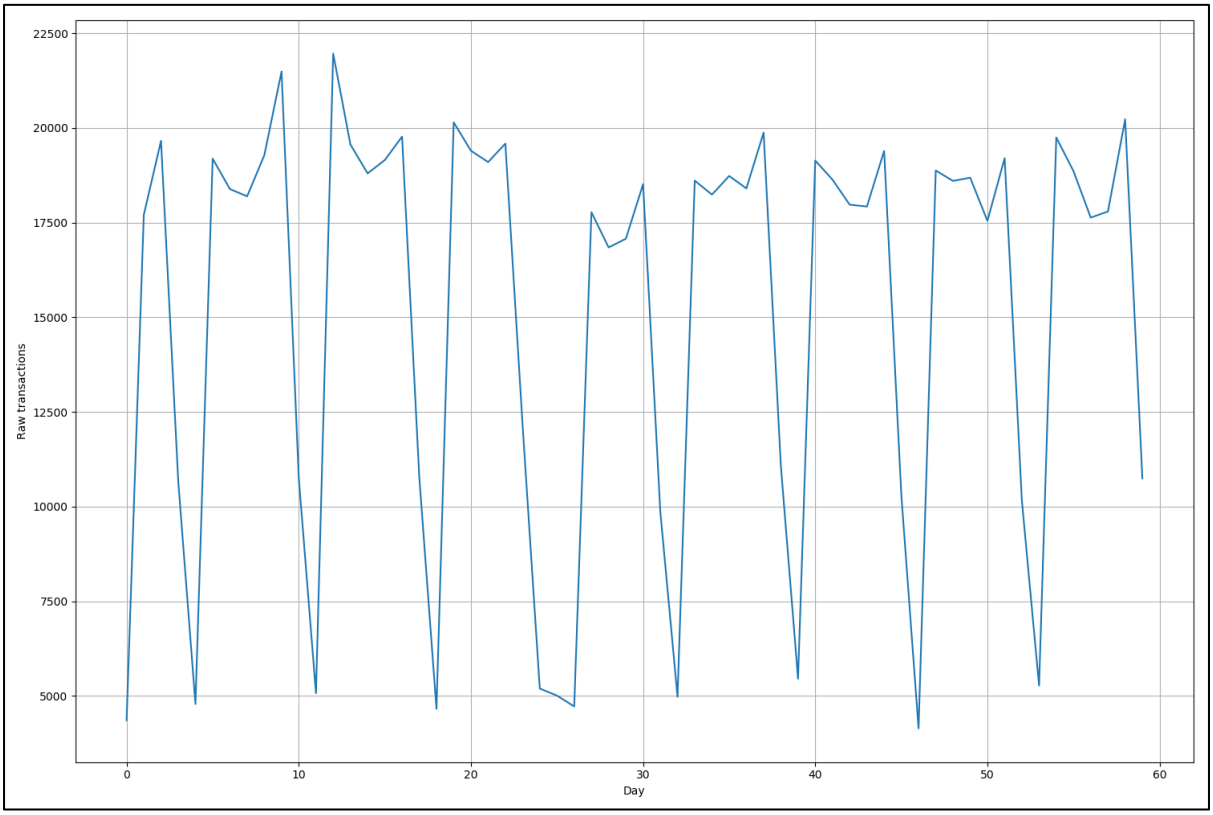
1. Users do not actually need to access the platform via their mobile devices
2. Users may find accessing the platform via their mobile phones a hassle

More work needs to be done to figure out which one of those reasons may be, and a simple **survey** may be conducted.

If development for the mobile platform is deemed to provide convenience to the users and is considered necessary, then development should be targeted towards the **Android platform** with its (slightly) higher user base.

2.7 Impacts of COVID-19:

Analysis:



We have plotted activity against day in terms of the number of (raw) transactions. Apart from the seasonal (weekly) trend of having very little activity on the weekends, the peaks reach about the same height at each interval. While there can be somewhat of a visible spike on day ~25 (roughly the time when Singapore announced its first Covid-19 infection case), the number of raw transactions can be seen to be slowly climbing post-initial announcement, and can be regarded as relatively healthy.

Recommendation:

A dip in platform usage can be expected with a sudden decrease in demand for shipment ever since the first case of Covid-19 being announced, and with the gradual increase of raw transactions over time **no further action** needs to be taken. However, it can be somewhat hard to tell from the raw transactions alone since it does not tell us any information about the actual shipments being transferred. Should a deeper analysis with regard to the impact of Covid-19 on the platform usage be required then more data relating to the movement of the actual containers may be more useful in this analysis.

Summary:

Overall, we would say that the insights generated from these datasets seem to prove useful as we managed to come up with relatively concrete paths on helping out the users of the platform. However, having said that, the lack of domain knowledge and abbreviations has hindered us from creating more concrete steps into actually helping the users. All in all, the most common recommendations across the board would be to create a dashboard for easy viewing of information that is commonly accessed, coming up with surveys to extract more information and definitely a recommendations machine for each category.