showwcase

Yi-Ling (Carol) Huang

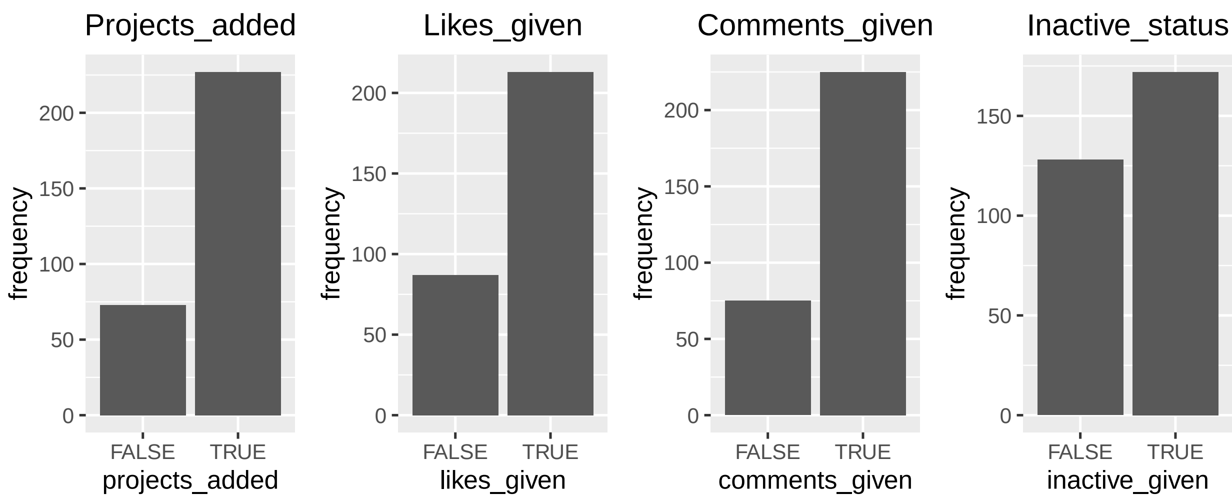
9/16/2020

The definition of "user engagement" can be different. It can be people log into the platform, upload files or share comments in the platform. Therefore, I will apply these variables to analyze "user engagement".

```
library(dplyr)
library(ggplot2)
library(grid)
library(gridExtra)
library(kableExtra)
library(knitr)
```

1. Overall Research: projects_added, likes_given, comment_given and inactive_status

```
a1 <- b %>%
        count(projects_added)
a2 <- b %>%
        count(likes_given)
a3 <- b %>%
        count(comment_given)
a4 <- b %>%
        count(inactive_status)

g1 <- ggplot(a1, aes(y=n, x=projects_added))+
geom_bar(position="dodge", stat="identity")+
ggtitle("Projects_added")+
theme(plot.title = element_text(hjust = 0.5))+
labs(x = "projects_added", y = "frequency")


g2 <- ggplot(a2, aes(y=n, x=likes_given))+
geom_bar(position="dodge", stat="identity")+
ggtitle("Likes_given")+
theme(plot.title = element_text(hjust = 0.5))+
labs(x = "likes_given", y = "frequency")

g3 <- ggplot(a3, aes(y=n, x=comment_given))+
geom_bar(position="dodge", stat="identity")+
ggtitle("Comments_given")+
theme(plot.title = element_text(hjust = 0.5))+
labs(x = "comments_given", y = "frequency")

g4 <- ggplot(a4, aes(y=n, x=inactive_status))+
geom_bar(position="dodge", stat="identity")+
ggtitle("Inactive_status")+
theme(plot.title = element_text(hjust = 0.5))+
labs(x = "inactive_given", y = "frequency")

grid.arrange(g1, g2, g3, g4, ncol = 4)
```

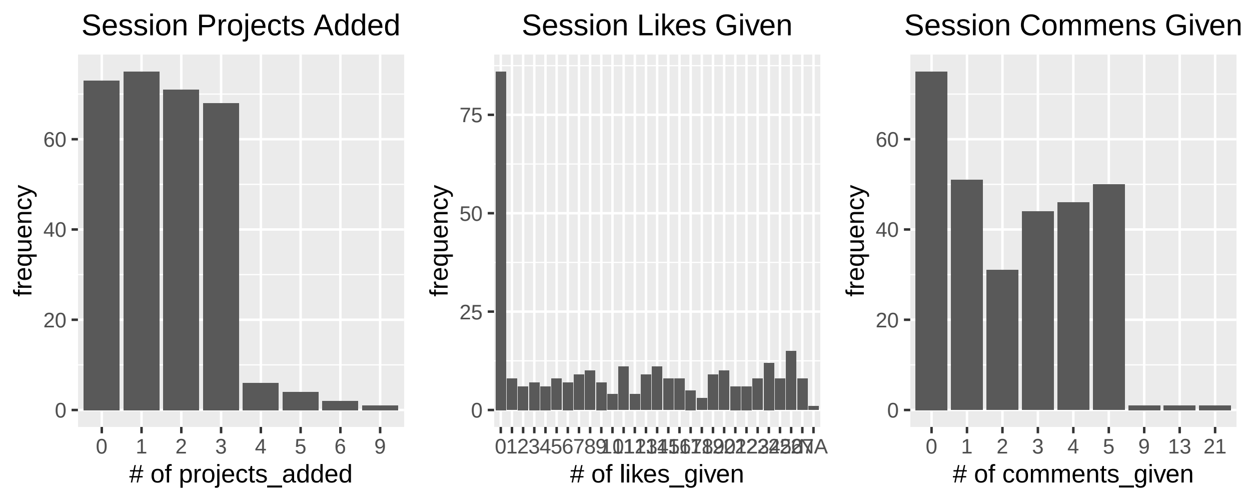


From bar charts above, we realize that customers tend to add project, give likes and give comment for each session in October 2019.

In addition, from the bar chart of inactive_status, we also understand that most of the customers are inactive. That being said, specific customers engaged this platform quite often in October 2019.

```r
# How's the projects_added distribution?
b1 <- b %>%
  count(session_projects_added)

b1$session_projects_added <- as.factor(b1$session_projects_added)

# Plot histogram to see distribution of likes_given
b2 <- b %>%
  count(session_likes_given)

b2$session_likes_given <- as.factor(b2$session_likes_given)

# Plot histogram to see distribution of comments_given
b3 <- b %>%
  count(session_comments_given)

b3$session_comments_given <- as.factor(b3$session_comments_given)

gb1 <- ggplot(b1, aes(y=n, x=session_projects_added))+
geom_bar(position="dodge", stat="identity")+
ggtitle("Session Projects Added")+
theme(plot.title = element_text(hjust = 0.5))+
labs(x = "# of projects_added", y = "frequency")

gb2 <- ggplot(b2, aes(y=n, x=session_likes_given))+
geom_bar(position="dodge", stat="identity")+
ggtitle("Session Likes Given")+
theme(plot.title = element_text(hjust = 0.5))+
labs(x = "# of likes_given", y = "frequency")

gb3 <- ggplot(b3, aes(y=n, x=session_comments_given))+
geom_bar(position="dodge", stat="identity")+
```

```r
ggtitle("Session Commens Given")+
theme(plot.title = element_text(hjust = 0.5))+
labs(x = "# of comments_given", y = "frequency")

grid.arrange(gb1, gb2, gb3, ncol = 3)
```



From graphs above, we understand that most of customers add less than 4 projects in each session. In addition, most of customers give 0 likes in each session. Last but not least, most of customers give less or equal to 5 comments in each session.

2. Which customer has the most session_project_added, session_likes_given and session_comments_given?

```r
x <- b %>%
  filter(session_projects_added >= 4) %>%
  arrange(desc(session_projects_added)) %>%
  top_n(10) %>%
  select(customer_id)

y <- b %>%
  filter(session_likes_given >= 14) %>%
  arrange(desc(session_likes_given)) %>%
  top_n(10) %>%
  select(customer_id)

z <- b %>%
  filter(session_comments_given >= 4) %>%
  arrange(desc(session_comments_given)) %>%
  top_n(10) %>%
  select(customer_id)
```

```r
# customers have most comments, most likes and most projects.
intersect(x, y, z)
```

```
##   customer_id
## 1       51243
## 2       87265
## 3       23985
```

From the outcome, we understand that customers who leave the most comments and likes and add the most projects are customer_id 51243, customer_id 87265 and customer_id 23985.

3. How many customers using this platform in October 2019?

```
# How many unique customers?
length(unique(b$customer_id))
```

```
## [1] 48
```

There are 48 customers using this platform in October 2019.

4. Which customer enters most frequently? (Top 10)

```
e1 <- b %>%
  count(customer_id)%>%
  arrange(desc(n)) %>%
  top_n(10) %>%
  select(customer_id)

e1 <- as.data.frame(e1)
e1
```

```
##    customer_id
## 1        29375
## 2        23404
## 3        38459
## 4        40235
## 5        87323
## 6        40572
## 7        14354
## 8        73245
## 9        51243
## 10       87265
```

The outcome shows the most frequent customers using this platform.

```
# count total number of session_projects_added, session_likes_given and session_comments_given for most
b %>%
  select(ï..session_id, session_projects_added, session_comments_given) %>%
  summarize(sum_project = sum(session_projects_added), sum_comments = sum(session_comments_given))

b %>%
  select(session_likes_given)%>%
  summarize_all(sum, na.rm = TRUE)

# top 10 most frequent customers session_projects_added
c1 <- c %>%
        select(customer_id, session_projects_added)
c11 <- aggregate(c1$session_projects_added, by=list(Customer=c1$customer_id), FUN=sum)

c11$projects_percentage <- c11$x/486
c11$projects_percentage <- round(c11$projects_percentage, digits = 4)

# top 10 most frequent customers total session_projects_added_percentage
sum(c11$projects_percentage)
```

```r
# top 10 most frequent customers session_likes_given
c2 <- c %>%
        select(customer_id, session_likes_given)
c22 <- c2 %>%
  group_by(customer_id) %>%
  summarize_all(sum, na.rm = TRUE)

c22$likes_percentage <- c22$session_likes_given/3127
c22$likes_percentage <- round(c22$likes_percentage, digits = 4)
# top 10 most frequent customers total session_likes_given_percentage
sum(c22$likes_percentage)


# top 10 most frequent customers session_comments_given
c3 <- c %>%
        select(customer_id, session_comments_given)
c33 <- aggregate(c3$session_comments_given, by=list(Customer=c3$customer_id), FUN=sum)

c33$comments_percentage <- c33$x/722
c33$comments_percentage <- round(c33$comments_percentage, digits = 4)

# top 10 most frequent customers total session_comments_given_percentage
sum(c33$comments_percentage)
```

```r
f <- cbind(Projects_added = 0.5371, Likes_given = 0.5586, Comments_given = 0.5609)
rownames(f) <- "Total percentage"
f
```

```
##                  Projects_added Likes_given Comments_given
## Total percentage         0.5371      0.5586         0.5609
```

The outcome shows the top 10 most frequent customers in projects_added, likes_given and comments_given with percentage.