

# Mentees\_Updates\_Report

Marwa El Awik

2023-04-16

---

```
library(tidyverse)
library(RColorBrewer)
library(plotly)
library(plyr)
library(gt)
library(stringr)
library(stringi)
library(mapquestr)
library(leaflet)
```

**About Dataset** The dataset refers to 18 data about our SUPERGirls. It includes monthly data collected by those Girls regarding their feedback over the mentorship program and their relationship with their mentors. Looking ahead, of interest in this project will be to apply data preparation to be ready for further analysis, then to apply some EDA, to get all the information about our variable of interest, in addition to visualizing the data.

Here is a glimpse of what we will be working with.

```
data<- readr::read_csv("40_40_MENTEES_MONTHLY_Updates2023.csv")
colnames(data)
```

```
## [1] "Submission Date"
## [2] "Indicate What Month This Is"
## [3] "First Name"
## [4] "Last Name"
## [5] "Your School"
## [6] "Zip Code of School"
## [7] "Your Age"
## [8] "Your Current Grade"
## [9] "Your Email"
## [10] "Your Cell Number"
## [11] "Which area of STEM are you currently interested in?"
## [12] "What are this month's goals?"
## [13] "On a scale of 1 to 5, with 5 being the highest, how would you rate your confidence in speaking"
## [14] "Is there an area, subject or idea you would like your mentor to focus more on? If yes, please"
## [15] "On a scale of 1 to 5, with 5 being the highest, how would you rate your mentor mentee relation"
## [16] "Upload this month's mentor mentee photos. Place where you met, selfies, Zoom or Teams picture"
## [17] "Anything you'd like to share? Highlights. Celebrations. Upcoming Events. Or Suggestions? I"
## [18] "Today's Date"
```

```
data<- data[c(2:8,11:13,14:15)]
```

```
##Fixing the colnames
names(data)[1] <- 'Month'
names(data)[4] <- 'School name'
names(data)[5] <- 'School zipCode'
names(data)[6] <- 'Age'
names(data)[7] <- 'Current grade'
names(data)[8] <- 'Area intrested in'
names(data)[9] <- 'Month goals'
names(data)[10] <- 'confidence in speaking with others'
names(data)[11] <- 'Need mentor to focus more on'
names(data)[12] <- 'Mentor-Mentee relationship rating'
```

```
datanona<-subset(data, !is.na(data$`Area intrested in`))
datanona <- filter(datanona, datanona$Month == "Dec 2022")
```

```
Professional_Interest0 <- data.frame(Interest = c('Sciences/Research','Engineering','Technology','Science/Healthcare',
', 'Aerospace','Mathematics/Finance', 'Biomedical Research','Marketing/Advertisement in STEM'),
Dec_counting = c(sum(str_count(datanona$`Area intrested in` , regex("\\bSciences/Research")),
sum(str_count(datanona$`Area intrested in` , regex("\\bEngineering")),
sum(str_count(datanona$`Area intrested in` , regex("\\bTechnology")),
sum(str_count(datanona$`Area intrested in` , regex("\\bScience/Healthcare")),
sum(str_count(datanona$`Area intrested in` , regex("\\bBusiness & Economics")),
sum(str_count(datanona$`Area intrested in` , regex("\\bComputer Science")),
sum(str_count(datanona$`Area intrested in` , regex("\\bAerospace")),
sum(str_count(datanona$`Area intrested in` , regex("\\bMathematics/Finance")),
sum(str_count(datanona$`Area intrested in` , regex("\\bBiomedical Research")),
sum(str_count(datanona$`Area intrested in` , regex("\\bMarketing/Advertisement in STEM"))
)
)
```

```
datanona<-subset(data, !is.na(data$`Area intrested in`))
datanona <- filter(datanona, datanona$Month == "Jan 2023")
```

```
Professional_Interest1 <- data.frame(Interest = c('Sciences/Research','Engineering','Technology','Science/Healthcare',
', 'Aerospace','Mathematics/Finance', 'Biomedical Research','Marketing/Advertisement in STEM'),
Jan_counting = c(sum(str_count(datanona$`Area intrested in` , regex("\\bSciences/Research")),
sum(str_count(datanona$`Area intrested in` , regex("\\bEngineering")),
sum(str_count(datanona$`Area intrested in` , regex("\\bTechnology")),
sum(str_count(datanona$`Area intrested in` , regex("\\bScience/Healthcare")),
sum(str_count(datanona$`Area intrested in` , regex("\\bBusiness & Economics")),
sum(str_count(datanona$`Area intrested in` , regex("\\bComputer Science")),
sum(str_count(datanona$`Area intrested in` , regex("\\bAerospace")),
sum(str_count(datanona$`Area intrested in` , regex("\\bMathematics/Finance")),
sum(str_count(datanona$`Area intrested in` , regex("\\bBiomedical Research")),
sum(str_count(datanona$`Area intrested in` , regex("\\bMarketing/Advertisement in STEM"))
)
)
```

```

datanona<-subset(data, !is.na(data$`Area intrested in`))
datanona <- filter(datanona, datanona$Month == "Feb 2023")

Professional_Interest2 <- data.frame(Interest = c('Sciences/Research','Engineering','Technology','Scien
','Aerospace','Mathematics/Finance', 'Biomedical Research','Marketing/Advertisement in STEM'),
Feb_counting = c(sum(str_count(datanona$`Area intrested in` , regex("\\bScien
sum(str_count(datanona$`Area intrested in` , regex("\\bEngineerin
sum(str_count(datanona$`Area intrested in` , regex("\\bTechnolog
sum(str_count(datanona$`Area intrested in` , regex("\\bScience/Ma
sum(str_count(datanona$`Area intrested in` , regex("\\bHealthcar
sum(str_count(datanona$`Area intrested in` , regex("\\bBusiness &
sum(str_count(datanona$`Area intrested in` , regex("\\bComputer S
sum(str_count(datanona$`Area intrested in` , regex("\\bAerospace'
sum(str_count(datanona$`Area intrested in` , regex("\\bMathemati
sum(str_count(datanona$`Area intrested in` , regex("\\bBiomedical
sum(str_count(datanona$`Area intrested in` , regex("\\bMarketing/

)

)

```

```

datanona<-subset(data, !is.na(data$`Area intrested in`))
datanona <- filter(datanona, datanona$Month == "March 2023")

Professional_Interest3 <- data.frame(Interest = c('Sciences/Research','Engineering','Technology','Scien
','Aerospace','Mathematics/Finance', 'Biomedical Research','Marketing/Advertisement in STEM'),
Mar_counting = c(sum(str_count(datanona$`Area intrested in` , regex("\\bScien
sum(str_count(datanona$`Area intrested in` , regex("\\bEngineerin
sum(str_count(datanona$`Area intrested in` , regex("\\bTechnolog
sum(str_count(datanona$`Area intrested in` , regex("\\bScience/Ma
sum(str_count(datanona$`Area intrested in` , regex("\\bHealthcar
sum(str_count(datanona$`Area intrested in` , regex("\\bBusiness &
sum(str_count(datanona$`Area intrested in` , regex("\\bComputer S
sum(str_count(datanona$`Area intrested in` , regex("\\bAerospace'
sum(str_count(datanona$`Area intrested in` , regex("\\bMathemati
sum(str_count(datanona$`Area intrested in` , regex("\\bBiomedical
sum(str_count(datanona$`Area intrested in` , regex("\\bMarketing/

)

)

```

```

Total <-

mutate(

Professional_Interest0,

Dec_counting= round(Professional_Interest0$Dec_counting /sum(Professional_Interest0$Dec_counting )*100)

Jan_counting= round(Professional_Interest1$Jan_counting/sum(Professional_Interest1$Jan_counting)*100),
Feb_counting=round(Professional_Interest2$Feb_counting/sum(Professional_Interest2$Feb_counting)*100),
Mar_counting=round(Professional_Interest3$Mar_counting/sum(Professional_Interest3$Mar_counting)*100)

```

```

)

data_ggp <- data.frame(Interest = Total$Interest,                                     # Reshape data frame
                      Number_of_SUPERGirls = c(Total$Dec_counting,Total$Jan_counting,Total$Feb_counting),
                      Month = c(rep("Dec 2022", nrow(Total)),
                                rep("Jan 2023", nrow(Total)),
                                rep("Feb 2023", nrow(Total)),
                                rep("March 2023", nrow(Total))))

getPalette = colorRampPalette(brewer.pal(3, "Accent"))

data_ggp$Month = factor(data_ggp$Month, levels = c('March 2023', 'Feb 2023', 'Jan 2023', 'Dec 2022'))

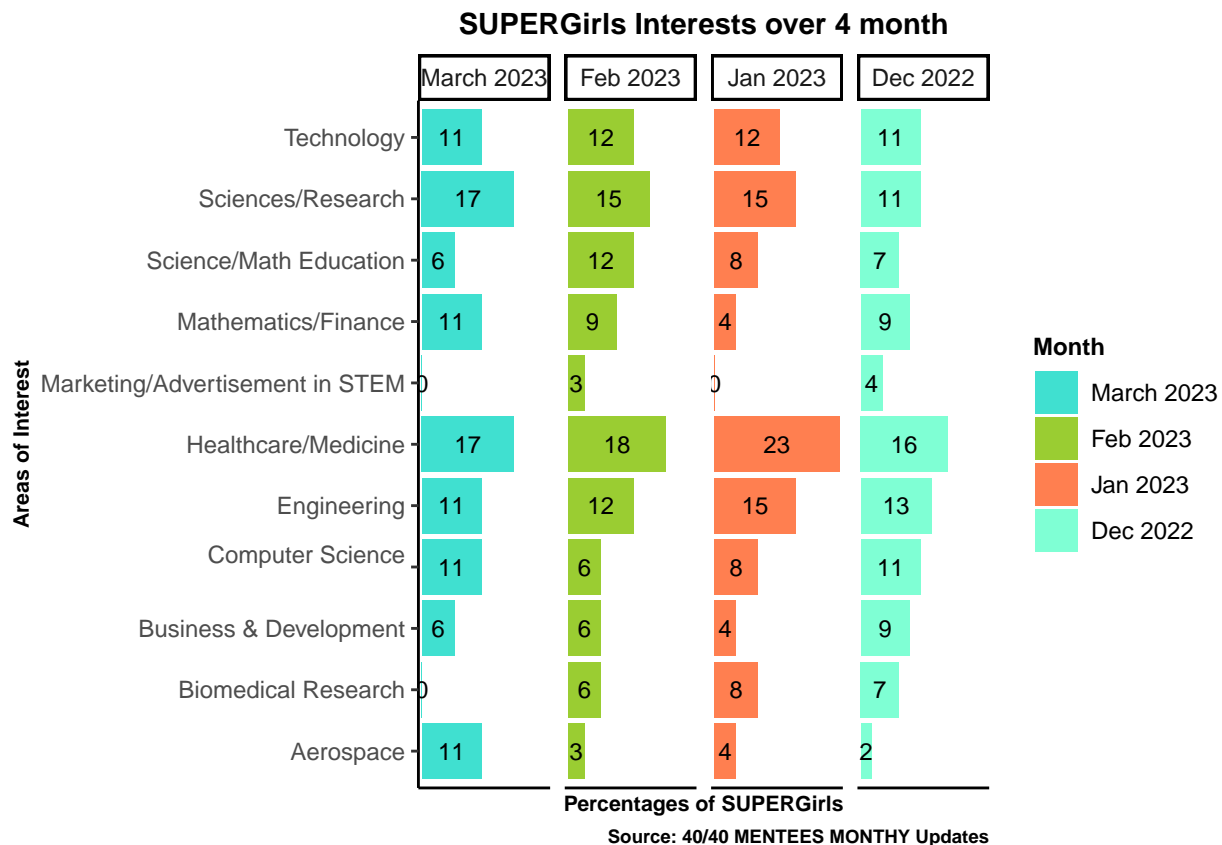
p1<-ggplot(data_ggp, aes(x= Interest, y=Number_of_SUPERGirls, fill= Month))+

  geom_bar(stat = 'identity',width=0.9)+
  ggtitle("SUPERGirls Interests over 4 month") +
  theme_classic()+
  coord_flip()+
  geom_text(aes(label=Number_of_SUPERGirls),size = 3, position = position_stack(vjust = 0.5))+
  labs(y="Percentages of SUPERGirls",x="Areas of Interest", caption = "Source: 40/40 MENTEES MONTHLY UPDATES")
  facet_grid(~Month)+
  scale_y_discrete(labels = NULL, breaks = NULL)+

  theme(
    title = element_text(size = 9, face = "bold"),
    plot.title = element_text(hjust = 0.5),
    axis.title.x = element_text(size = 8, face = "bold"),
    axis.title.y = element_text(size = 8, face = "bold"),
    axis.ticks.x = element_blank(),

    panel.grid.minor = element_blank())
p1 +scale_fill_manual(values = c("turquoise","yellowgreen","coral","aquamarine"))

```



- There are switches in SUPERGirls interest between Dec 22 and Mar 23
- The biggest interest is in Healthcare/Medicine field

```
data1 <- filter(data, data$Month == "Dec 2022")
relationship_rating_Dec<-count(data1$`Mentor-Mentee relationship rating`)%>%
  mutate(Dec_percent = round(freq/sum(freq)*100))
names(relationship_rating_Dec)[1]<-"Rating"
relationship_rating_Dec = subset(relationship_rating_Dec, select = -c(freq))

data2 <- filter(data, data$Month == "Jan 2023")
relationship_rating_Jan<-count(data2$`Mentor-Mentee relationship rating`)%>%
  mutate(Jan_percent = round(freq/sum(freq)*100))
names(relationship_rating_Jan)[1]<-"Rating"
relationship_rating_Jan = subset(relationship_rating_Jan, select = -c(freq))

data3 <- filter(data, data$Month == "Feb 2023")
relationship_rating_Feb<-count(data3$`Mentor-Mentee relationship rating`)%>%
  mutate(Feb_percent = round(freq/sum(freq)*100))
names(relationship_rating_Feb)[1]<-"Rating"
relationship_rating_Feb = subset(relationship_rating_Feb, select = -c(freq))

data4 <- filter(data, data$Month == "March 2023")
relationship_rating_Mar<-count(data4$`Mentor-Mentee relationship rating`)%>%
```

```

mutate(Mar_percent = round(freq/sum(freq)*100))
names(relationship_rating_Mar)[1]<-"Rating"
relationship_rating_Mar = subset(relationship_rating_Mar, select = -c(freq))

All_ratings<- full_join(relationship_rating_Dec,relationship_rating_Jan, by="Rating")%>%
  full_join(relationship_rating_Feb)%>%full_join(relationship_rating_Mar, by="Rating")%>% arrange(Rating)

names(All_ratings)[2]<-"Dec"
names(All_ratings)[3]<-"Jan"
names(All_ratings)[4]<-"Feb"
names(All_ratings)[5]<-"Mar"

data_ggp1 <- data.frame(Rating = All_ratings$Rating, # Reshape data frame
  counts = c(All_ratings$Dec,All_ratings$Jan,All_ratings$Feb,All_ratings$Mar),
  Month = c(rep("Dec 2022", nrow(All_ratings)),
    rep("Jan 2023", nrow(All_ratings)),
    rep("Feb 2023", nrow(All_ratings)),
    rep("Mar 2023", nrow(All_ratings))))

data_ggp1$Month = factor(data_ggp1$Month, levels = c('Dec 2022','Jan 2023','Feb 2023','Mar 2023'))

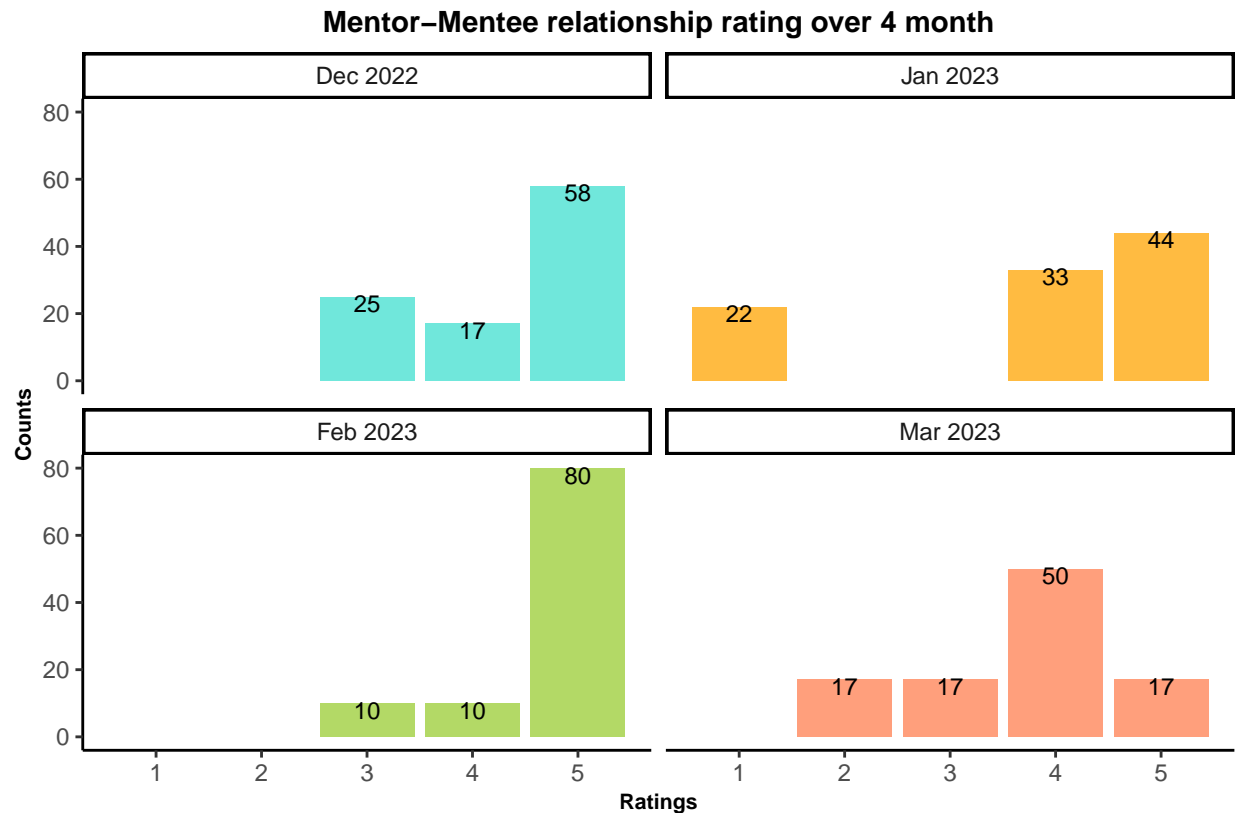
getPalette = colorRampPalette(brewer.pal(3, "Spectral"))

p2<-ggplot(data_ggp1, aes(x= Rating, y=counts, fill= Month))+

  geom_bar(stat = 'identity',position=position_dodge(), alpha= 0.75)+
  ggtitle("Mentor-Mentee relationship rating over 4 month") +
  theme_classic()+
  facet_wrap(~Month)+
  geom_text(aes(label=counts), vjust=0.9,
    position=position_dodge(.9), size=3)+
  labs(y="Counts",x="Ratings", caption = "Source: 40/40 MENTEES MONTHY Updates")+

  theme(legend.position = "none",
    title = element_text(size = 9, face = "bold"),
    plot.title = element_text(hjust = 0.5),
    axis.title.x = element_text(size = 8, face = "bold"),
    axis.title.y = element_text(size = 8, face = "bold"),
    panel.grid.minor = element_blank())
p2 +scale_fill_manual(values = c("turquoise","orange1","yellowgreen","coral"))

```



Source: 40/40 MENTEES MONTHY Updates

There are switches in SUPERGirls relationship between Dec 22 and Mar 23.

- 80% of the SUPERGirls rated their relationship with their mentor as 5 in Feb. This percentage has decreased to 17% in Mar 2023.
- No SUPERGirls rated their relationship with their mentor as 2 in Feb. However, 17% gave 2 or “Bad” ratings to their relationship with their mentors in Mar 2023

```
data1 <- filter(data, data$Month == "March 2023")
condifence_rating_Mar<-count(data1$`confidence in speaking with others`)%>%
  mutate(Mar_percent = round(freq/sum(freq)*100))
names(condifence_rating_Mar)[1]<-"Rating"
condifence_rating_Mar = subset(condifence_rating_Mar, select = -c(freq))

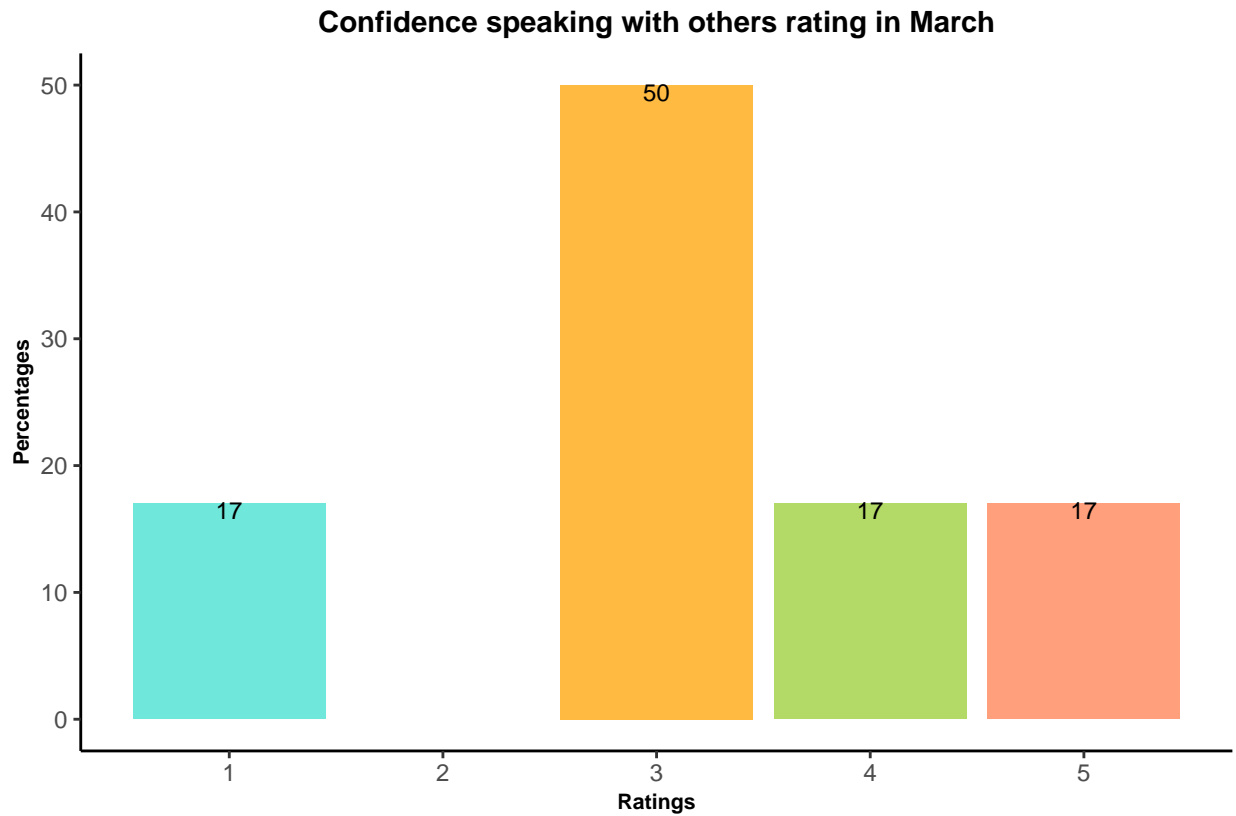
p2<-ggplot(condifence_rating_Mar, aes(x= Rating, y=Mar_percent, fill= factor(Rating)))+
  geom_bar(stat = 'identity',position=position_dodge(), alpha= 0.75)+
  ggtitle("Confidence speaking with others rating in March") +
  theme_classic()+
  geom_text(aes(label=Mar_percent), vjust=0.9,
            position=position_dodge(.9), size=3)+
  labs(y="Percentages",x="Ratings", caption = "Source: 40/40 MENTEES MONTHY Updates")+

  theme(legend.position = "none",
        title = element_text(size = 9, face = "bold"),
        plot.title = element_text(hjust = 0.5),
        axis.title.x = element_text(size = 8, face = "bold"),
        axis.title.y = element_text(size = 8, face = "bold"),
```

```

panel.grid.minor = element_blank()
p2 +scale_fill_manual(values = c("turquoise","orange1","yellowgreen","coral","aquamarine"))

```



Half of the SUPERGirls rated their confidence when speaking with others as 3 or Neutral. The other percentages as equally distributed between very bad, good and fantastic.

#### Recommendations/Room for improvement

- Need to have all the SUPERGirls submit their monthly report (Dec: 12 submissions/Jan: 9 submissions/Feb: 10 submissions/Mar: 6 submissions/Apr: 0 submission)
- Follow up with the girls that rated their mentor-mentee relationship less than 4 to check for any issues to solve.
- Follow up with the mentors of the SUPERGirls who rated their confidence while speaking with others less than 4 to focus on improving this skill.