

**Name: Naveen Aradhyamath**

**Data Analyst**

**Contact no: 9731047593**

**Title:** Data Analyst test (**kwalee**)

**Task:** To find on which level are players most likely to fail?

**Solution:** level\_number 3

**Method 1:**

**R code:**

```
install.packages("readr")
install.packages("magrittr")
install.packages("dplyr")
install.packages("ggplot2")
library(readr)
players <- read.csv("level_progress.csv")
players
library(magrittr)
library(dplyr)

players_failed <- players %>% filter(status == "fail") # filtering the data that shows only the failed status

most_failed_level <- players_failed %>% group_by(level_number) %>%

summarise(freq = n()) %>% arrange(desc(freq)) # summarising the number of frequency that the total level number occurred in a dataset and arranging it in a descending order.

head(most_failed_level) # Records showing only top 6 counts which are in descending order.
```

**Output:**

	level_number	freq
	<int>	<int>
1	3	865
2	5	537
3	4	475
4	6	400
5	7	255
6	8	255

**Conclusion:** we can see that the total frequency of level number 3 is high that is 865, hence we can conclude that maximum people failed at level number 3.

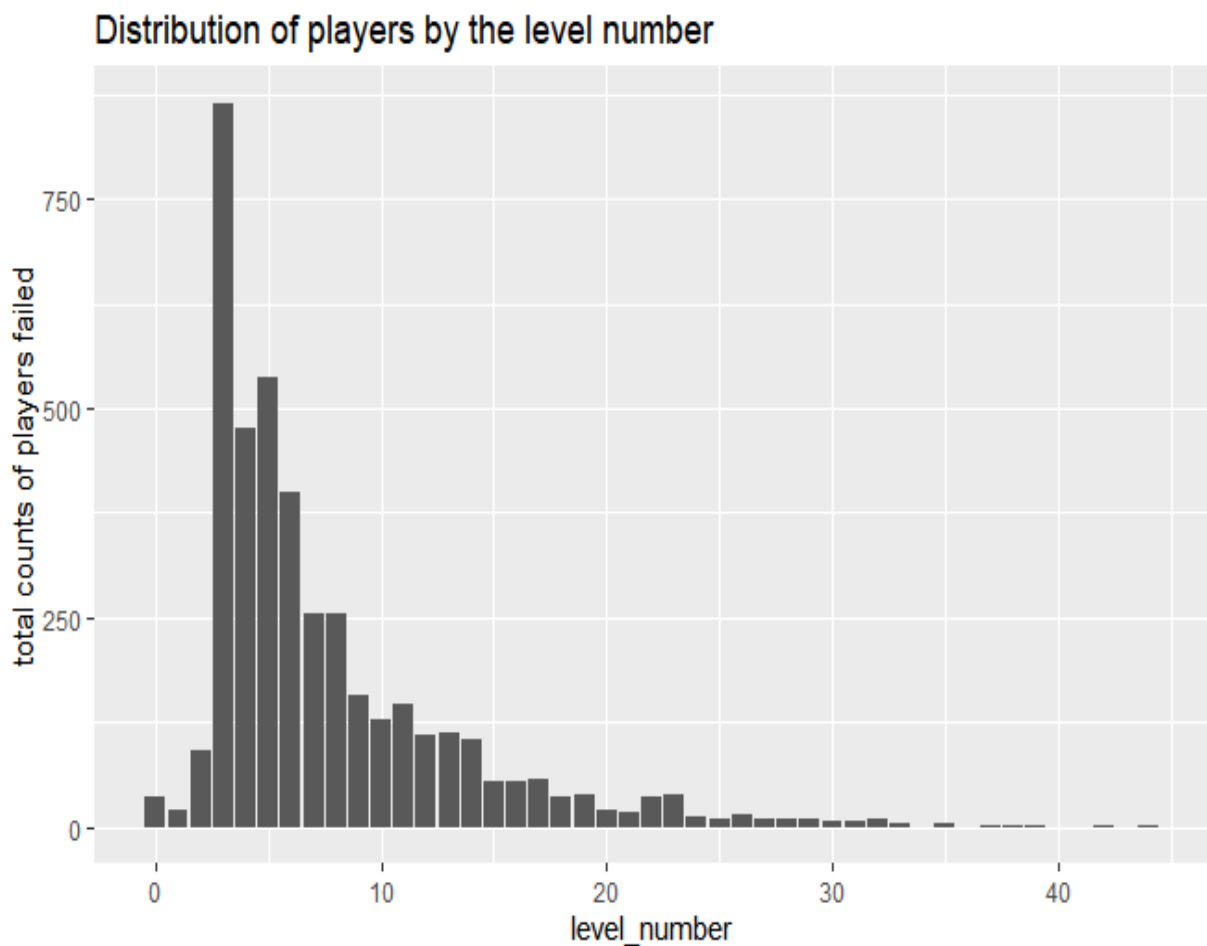
## Method 2:

### Data Visualization:

```
library(ggplot2)

ggplot(data = players_failed, aes(x = level_number)) +
  geom_bar() +
  labs(x = "level_number", y = "total counts of players failed",
       title = "Distribution of players by the level number")
```

### Output:



**Conclusion:** Above bar graph shows that at level 3 there is a increase in number of players failed. Hence level 3 is the where most players failed.

### Method 3:

#### R code:

```
players_failed <- players %>% filter (status == "fail")
```

```
count(players_failed)
```

```
mean(players_failed$level_number)
```

o/p 7.616015

```
players_failed_more <- players_failed %>% filter(level_number <= 8)
```

```
players_failed_less <- players_failed %>% filter(level_number >= 8)
```

```
count(players_failed_more)
```

```
count(players_failed_less)
```

```
players_failed_more %>% mutate(player_stratum= cut(level_number, breaks = 5)) %>%  
group_by(player_stratum) %>% summarise(n = n())
```

#### O/p:

player_stratum <fct>	n <int>
1 (-0.007,1.4]	57
2 (1.4,2.8]	91
3 (2.8,4.2]	1340
4 (4.2,5.6]	537
5 (5.6,7.01]	655

**Conclusion:** We can see that the total number of players failed is higher in the level that falls between the range 2.8 - 4.2, therefore we can conclude that players are most likely to fail in the range 2.8 – 4.2.