SampleExercise

Shonte

5/1/2019

Context

Many times, you will want to go beyond creating rule sets based on validating single rows in a column. Expanding your rule set to validate aggregates with functions like sum, mean and median will help you consolidate vast amounts of detailed data in order to extract and expect higher quality from your data.

Instructions

Consider this example that creates a rule to validate the mean of column `x` against the data df with respect to a numerical criteria: Said differently, we are testing that the mean of column x is greater than or equal to 2.

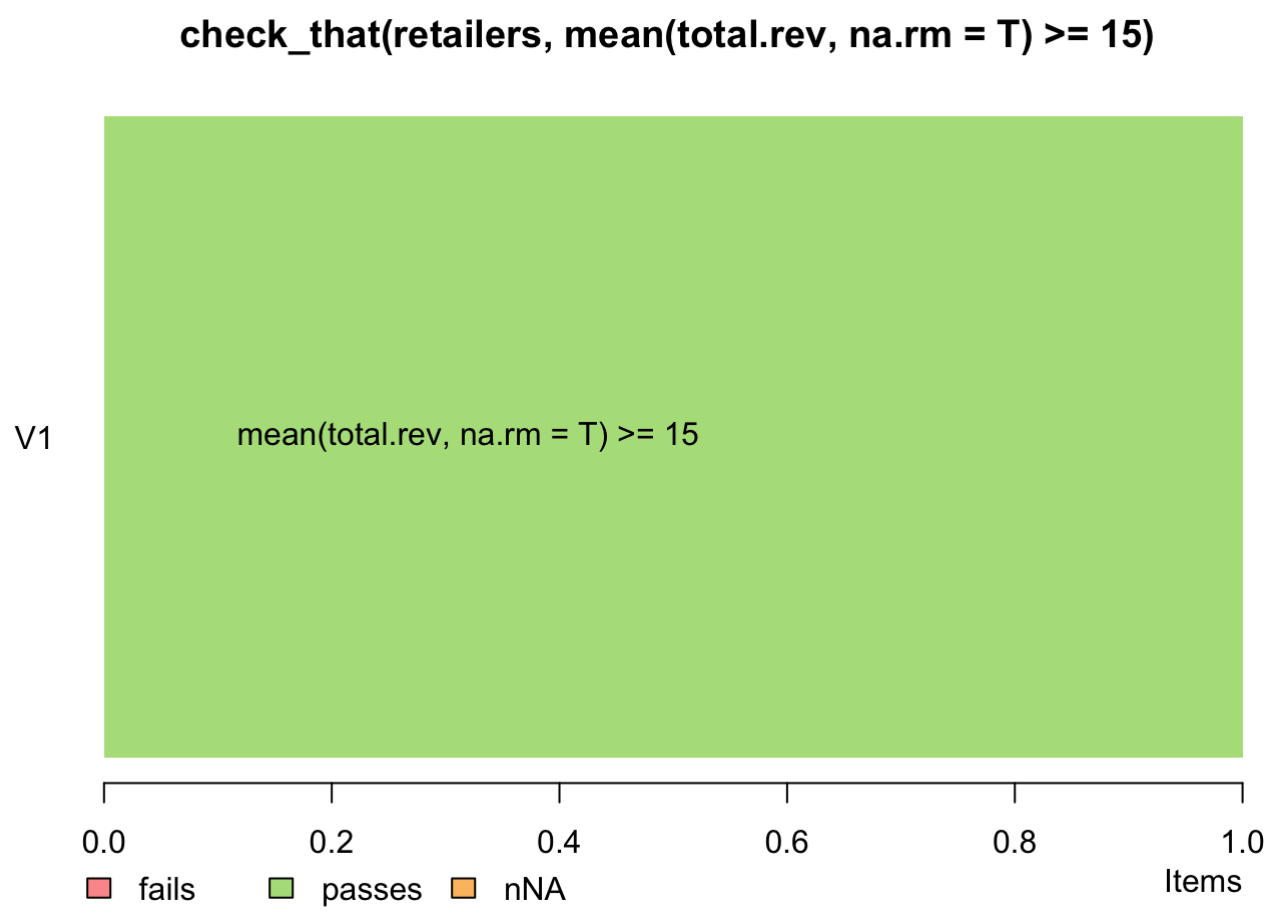
```
check_that(df, mean(x, na.rm = T) >= 2)
```

For this Exercise:

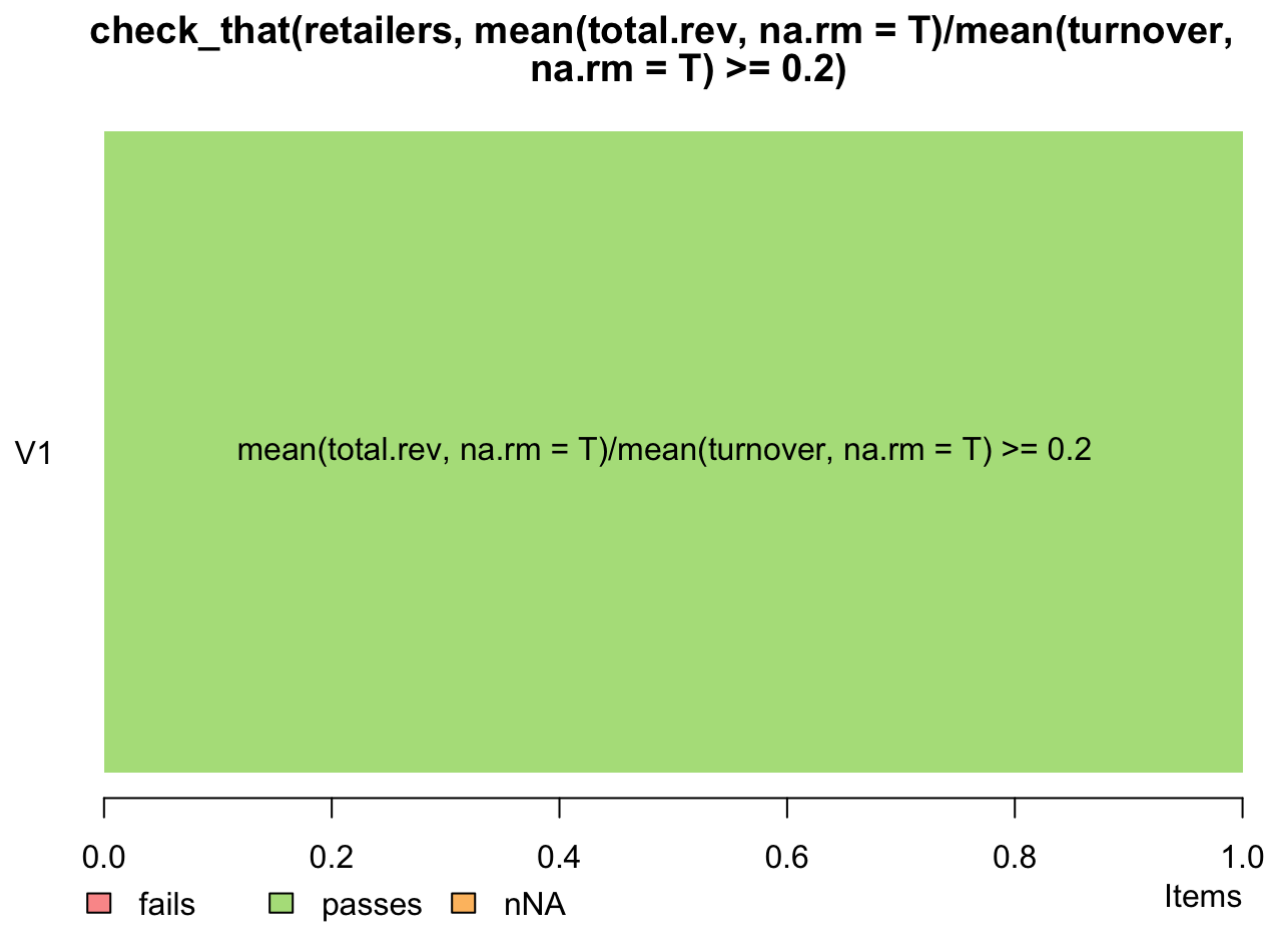
- Use `mean` to check that that the average of `total.rev` is > 15. Save the resulting dataset to `v1`
- Specify a new rule that validates the ratio of two columns: Check that the ratio of the average of `total.rev` and `turnover` is at least `.2` . Save the resulting dataset to `v2`
- Add a new rule that expects `turnover` to be b/t 0.1 and 10 times its median. Save the resulting dataset to `v3`
- Visualize each rule.

Solution Exercise with Respect to Instructions

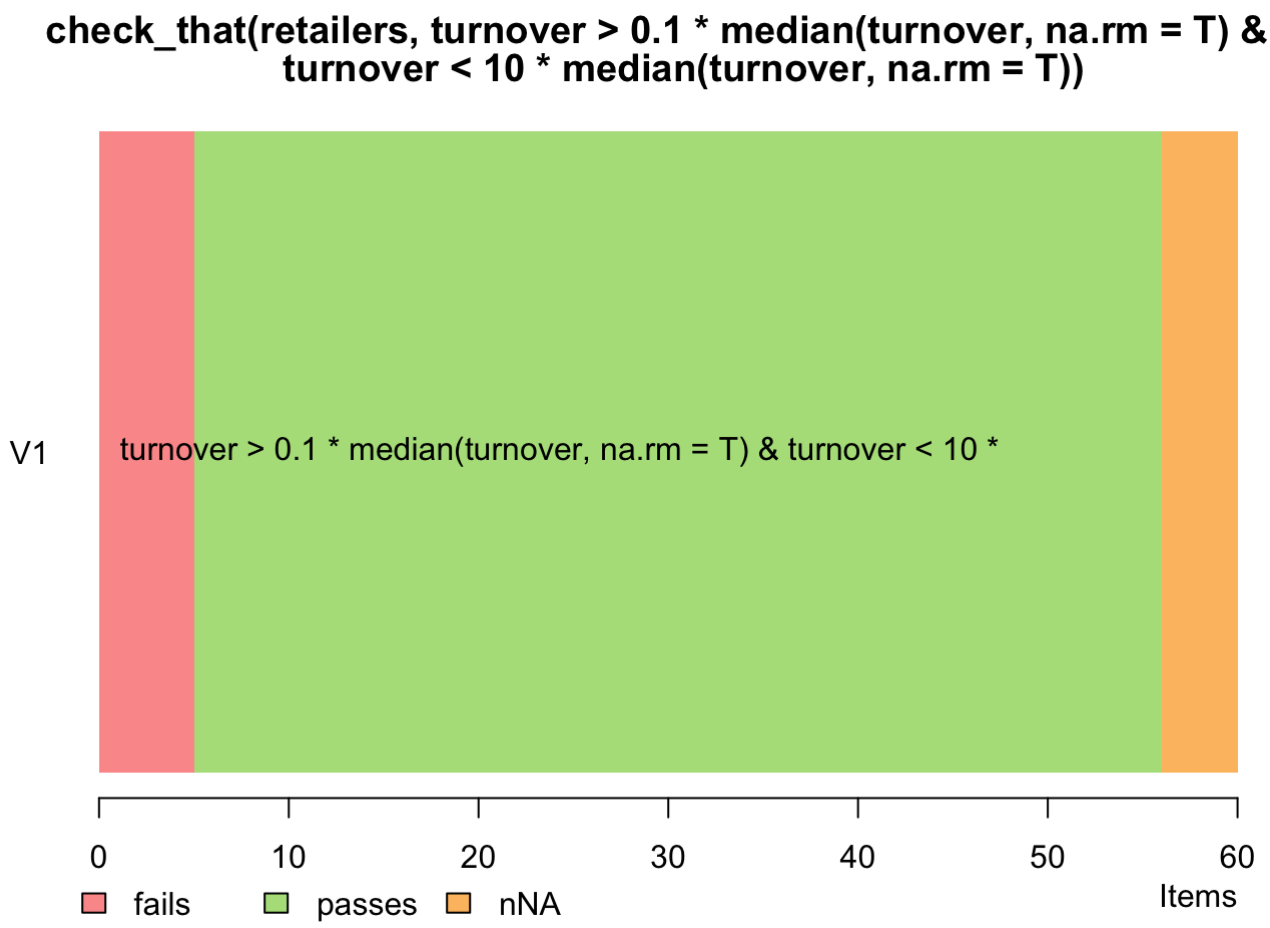
```
# The Validate and Retailers Data are already loaded

library(validate)
data("retailers")

# Check that the average of `total.rev` is > 15: v1
 v1 <- check_that(retailers, mean(total.rev, na.rm = T) >=15)

# Check that the ratio of the average of `total.rev` and `turnover` is at least `.2`:  v
2
 v2 <- check_that(retailers, mean(total.rev, na.rm = T) / mean(turnover, na.rm = T) >=.2
)

# Check that `turnover` is b/t 0.1 and 10 times its median: v3
 v3 <- check_that(retailers, turnover > 0.1*median(turnover, na.rm = T) & turnover < 10*
median(turnover, na.rm = T))

# visualize each rule set
barplot(v1)
```


check_that(retailers, mean(total.rev, na.rm = T) >= 15)
V1
mean(total.rev, na.rm = T) >= 15
0.0
0.2
0.4
0.6
0.8
1.0
Items
fails
passes
nNA

```
barplot(v2)
```


check_that(retailers, mean(total.rev, na.rm = T)/mean(turnover, na.rm = T) >= 0.2)
V1
mean(total.rev, na.rm = T)/mean(turnover, na.rm = T) >= 0.2
0.0
0.2
0.4
0.6
0.8
1.0
Items
fails
passes
nNA

```
barplot(v3)
```


check_that(retailers, turnover > 0.1 * median(turnover, na.rm = T) & turnover < 10 * median(turnover, na.rm = T))
V1
turnover > 0.1 * median(turnover, na.rm = T) & turnover < 10 *
0
10
20
30
40
50
60
Items
fails
passes
nNA