1.

Table

Description automatically generatedGraphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated

2.

a. Godzilla height is 108 meters/ 10800 cm.

b. No the column Height represents this incorrectly.

Graphical user interface, text, application, email

Description automatically generated

Formula for height modification from “meters” to “cm”

Graphical user interface, application, table

Description automatically generated

screen shot of Godzilla’s height and the 9 rows before it.

3. Graphical user interface, table

Description automatically generated

screen shot of the Columns list on the left for those columns. (used hide and exclude options from column menu)

4. There are 10 outliers, they have exceeded the numeric threshold of 792 to be called outliers.

Graphical user interface, application

Description automatically generated

Outliers

31 Anti-Monitor

123 Bloodwraith

254 Fin Fang Foom

270 Galactus

281 Giganta

284 Godzilla

390 King Kong

654 Surtur

692 Utgard-Loki

739 Ymir

5.  screen shot of the first 10 rows after Normalising/standarsizing:

Graphical user interface

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| 6. id Comic height Standardized Height in cm | | | |
| 31 | Anti-Monitor | 6100 | 3.521757 |
| 123 | Bloodwraith | 3050 | 1.68391 |
| 254 | Fin Fang Foom | 975 | 0.433571 |
| 270 | Galactus | 876 | 0.373916 |
| 281 | Giganta | 6250 | 3.612142 |
| 284 | Godzilla | 10800 | 6.353849 |
| 390 | King Kong | 3050 | 1.68391 |
| 654 | Surtur | 30480 | 18.21248 |
| 692 | Utgard-Loki | 1520 | 0.761973 |
| 739 | Ymir | 30480 | 18.21248 |

Check the standardized height for outliers.  Did these results differ from the previous outliers question?  If so, how did it differ?  If not, why not?

Usually the standardization/ normalizing is to bring all the variables on a same scale, after standardization the outliers remain the same, there are the same 10 outliers as we have in previous question, but the scale seems to be compressed, when represented on a graph its easier to represent after standardizing, since the larger values would dominate the output.

7. Text, table

Description automatically generated

 Do any of these variables have missing values?  If so, how many?  Which attribute are the characters highest in on average?

Yes, these variables do have missing values, there are 78 missing values, Intelligence has the highest average among all the attributes.

8.

Chart, box and whisker chart

Description automatically generated

screen shot of the bins.

b. There are 67 characters that have a strength of 100.

9.

a. There are 16 categories in the skin color column.

b. there are 669 missing values in the skin color column.

c. The best way to deal with missing values is Omission, as there are many missing values on a small data set its better to drop those variables.

Screenshot is attached below for Q9.

Graphical user interface, application, Word

Description automatically generated

10. screen shot showing the row counts of each partition.

Graphical user interface, application

Description automatically generated