Name: Haolin Feng

UID: 964150056

Option 1

1.Files count: 5944, total size: 1107544667 bytes

2.

Biggest file: brochure-outside.ai Size: 59381544 bytes

Smallest file: format Size: 2 bytes

Empty file count: 15

3.

mean: 186206.8795626577 bytes

standard deviation: 2213503.710254146 bytes

median: 1369 bytes

mode: 155 bytes

4.

图形用户界面, 文本, 应用程序

描述已自动生成

图表, 折线图

描述已自动生成

From pdf and cdf we can observe that most of the file size is around hundred KB to a few MB. There are only several file size is more than 10MB.

5.

File type distribution

|  |  |  |
| --- | --- | --- |
| File type | Type size(in bytes) | Size Fraction |
| pdf | 545146440 | 49.25% |
| jpg | 337287389 | 30.47% |
| ai | 116690557 | 10.54% |
| txt | 32110376 | 2.90% |
| rar | 21585946 | 1.95% |
| emz | 15244265 | 1.38% |
| png | 11812026 | 1.07% |
| html | 7279458 | 0.66% |
| gif | 3003468 | 0.27% |
| mov | 2314790 | 0.21% |
| others | 14525184 | 1.31% |

File count distribution

|  |  |  |
| --- | --- | --- |
| File type | Type count | Count fraction |
| png | 2695 | 45.33% |
| jpg | 905 | 15.22% |
| gif | 656 | 11.03% |
| php | 532 | 8.95% |
| pdf | 362 | 6.09% |
| txt | 251 | 4.22% |
| wmz | 133 | 2.23% |
| html | 121 | 2.04% |
| css | 91 | 1.53% |
| pl | 42 | 0.71% |
| others | 157 | 2.64% |

From the table above, we can observe that the number of that file type and the size of that file type is not in a proportional relationship. We can conclude that there are some large files which is much larger than most of the files in the website.

6.

图表, 直方图

描述已自动生成

图表, 直方图

描述已自动生成

图表, 折线图

描述已自动生成

From the graph above, we can observe that the largest size of poster are much smaller than the file size of paper. Most of the file size of paper and poster are small, around hundred KB. From the cdf graph we can conclude that the overall file size of paper is larger than poster.

7.

There are 18 oldest files which are acmlogo.gif, BalconyView.jpg, banff.gif, BanffSpringsHotel.jpg, BanquetRoom.jpg, Gardens,jpg, iw3c2.gif, Lobby.gif, mapairport.gif, mapairportsm.gif, mapbanffsm.gif, MeetingRoom.jpg, MoraineLake.jpg, SignatureRoom.jpg, SpanishWalk.jpg, Thumbs.db, Whirlpool.jpg and WWW2007logo.gif. They are 6000 days ago

There are 2 newest file which are index.php and indexold.html. They are 5291 days ago

mean: 5655.236709286675days

median: 5639.0days

mode: 5639days

8.

图表, 折线图

描述已自动生成

From the cdf graph above, we can observe that the most active period of the website is around 5600-5650 days before. From 5300 to 5550 days before, the website still has some active user. In the beginning of the website, the active rate is likely to be linear.