1. 读取 *data* 中 *2023_6* 文件夹中的 *openrank* 数据集，分析美国排名前一百的项目的的 *value* 的最大值、最小值、均值以及中位数。

```
mysql> SELECT *
    -> FROM openrank
    -> ORDER BY rank
    -> LIMIT 100;
+-------------------------------------------------------------------------+------
---------+-----------+------------+
| item                                                                    | rank
 value   | rankDelta | valueDelta |
+-------------------------------------------------------------------------+------
---------+-----------+------------+
| home-assistant/core                                                     |    1
 1394.45 |         0 |     -23.01 |
| microsoft/vscode                                                        |    2
 1334.48 |         0 |     -36.42 |
| NixOS/nixpkgs                                                           |    3
 1240.31 |         0 |      -0.31 |
| flutter/flutter                                                         |    4
  886.61 |         0 |     -13.74 |
| pytorch/pytorch                                                         |    5
  803.78 |         1 |       8.29 |
| MicrosoftDocs/azure-docs                                                |    6
  793.99 |        -1 |      -5.21 |
| godotengine/godot                                                       |    7
  685.85 |         2 |       5.87 |
| dotnet/runtime                                                          |    8
  680.36 |         0 |      -7.54 |
| langchain-ai/langchain                                                  |    9
  592.57 |         6 |      74.53 |
| rust-lang/rust                                                          |   10
  565.62 |         1 |     -25.74 |
| vercel/next.js                                                          |   11
  536.36 |         5 |      26.43 |
| firstcontributions/first-contributions                                  |   12
  510.59 |         1 |     -24.73 |
| microsoft/PowerToys                                                     |   13
```

```
mysql> SELECT
    ->     MAX(value) AS max_value,
    ->     MIN(value) AS min_value,
    ->     AVG(value) AS avg_value
    -> FROM openrank
    -> ORDER BY rank
    -> LIMIT 100;
+-----------+-----------+-------------+
| max_value | min_value | avg_value   |
+-----------+-----------+-------------+
|   1394.45 |    200.68 | 346.967900 |
+-----------+-----------+-------------+
1 row in set (0.00 sec)
```

```
mysql> SELECT
    ->     AVG(value) AS median_value
    -> FROM (
    ->     SELECT
    ->         value,
    ->         @row_index := @row_index + 1 AS row_index
    ->     FROM openrank
    ->     ORDER BY value
    ->     LIMIT 100
    -> ) AS ranked_values
    -> WHERE row_index IN (FLOOR(@row_index / 2), CEIL(@row_index / 2));
+--------------+
| median_value |
+--------------+
|   273.670000 |
+--------------+
1 row in set (0.00 sec)
```

2. 读取 *data* 中 *2022* 文件夹下的 *activity_2020* 文件，分析美国排名前十的项目的平均增长率。

```
mysql> USE activity_database;
Database changed
mysql>
mysql> CREATE TABLE activity_2022 (
    ->     repo VARCHAR(255),
    ->     20221_value DECIMAL(10, 2),
    ->     20222_value DECIMAL(10, 2),
    ->     20223_value DECIMAL(10, 2),
    ->     20224_value DECIMAL(10, 2),
    ->     20225_value DECIMAL(10, 2),
    ->     20226_value DECIMAL(10, 2),
    ->     20227_value DECIMAL(10, 2),
    ->     20228_value DECIMAL(10, 2),
    ->     20229_value DECIMAL(10, 2),
    ->     202210_value DECIMAL(10, 2),
    ->     202211_value DECIMAL(10, 2),
    ->     202212_value DECIMAL(10, 2)
    -> );
Query OK, 0 rows affected (0.02 sec)

mysql> LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 5.7/Uploads/activ
ity_2022.csv"
    -> INTO TABLE activity_2022
    -> FIELDS TERMINATED BY ','
    -> ENCLOSED BY '"'
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 ROWS;
Query OK, 10 rows affected (0.00 sec)
Records: 10  Deleted: 0  Skipped: 0  Warnings: 0
```

```
    ->     SELECT
    ->         repo,
    ->         ((202212_value - 202211_value) / 202211_value)
    ->     FROM activity_2022
    -> ) AS growth_rate
    -> GROUP BY repo;
Query OK, 10 rows affected, 10 warnings (0.02 sec)
Records: 10  Duplicates: 0  Warnings: 10

mysql> SELECT *
    -> FROM project_growth_rates
    -> ORDER BY average_growth_rate DESC
    -> LIMIT 10;
+--------------------------------------+---------------------+
| repo                                 | average_growth_rate |
+--------------------------------------+---------------------+
| firstcontributions/first-contributions |        0.1408236364 |
| microsoft/vscode                     |        0.0229047273 |
| pytorch/pytorch                      |        0.0131602727 |
| home-assistant/core                  |        0.0084525455 |
| NixOS/nixpkgs                        |        0.0002944545 |
| elastic/kibana                       |       -0.0036507273 |
| rust-lang/rust                       |       -0.0037781818 |
| flutter/flutter                      |       -0.0124043636 |
| MicrosoftDocs/azure-docs             |       -0.0179278182 |
| dotnet/runtime                       |       -0.0228481818 |
+--------------------------------------+---------------------+
10 rows in set (0.00 sec)
```

```
mysql> SELECT AVG(average_growth_rate) AS average_growth_rate_top_10
    -> FROM (
    ->     SELECT average_growth_rate
    ->     FROM project_growth_rates
    ->     ORDER BY average_growth_rate DESC
    ->     LIMIT 10
    -> ) AS top_10_projects;
+---------------------------+
| average_growth_rate_top_10 |
+---------------------------+
|           0.01250263637000 |
+---------------------------+
1 row in set (0.00 sec)
```

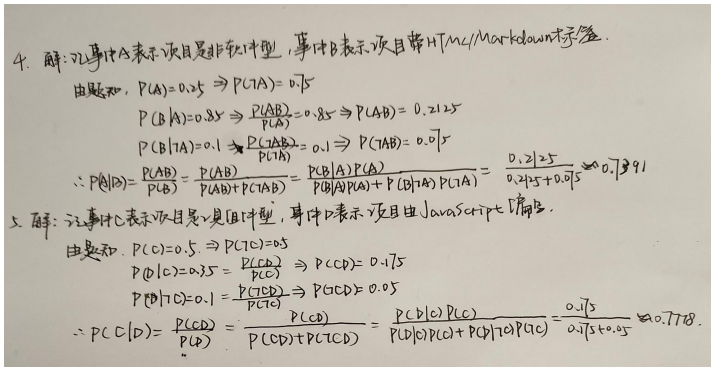
3. *data/2022/china_2022.csv* 表示中国开源领域排名前十的企业。*data/2022/global_2022.csv* 表示开源领域全球前十的的企业，请通过各种统计指标比较两者的各种数据差异。

```
mysql> SELECT * FROM comparison_results;
+---------+----------------+-------------+--------------+------------+
| name    | metric         | china_value | global_value | difference |
+---------+----------------+-------------+--------------+------------+
| Alibaba | issue_comment  |      111972 |       111972 |          0 |
| Baidu   | issue_comment  |       79370 |        79370 |          0 |
| Alibaba | open_issue     |       22397 |        22397 |          0 |
| Baidu   | open_issue     |       15580 |        15580 |          0 |
| Alibaba | open_pull      |       35266 |        35266 |          0 |
| Baidu   | open_pull      |       27414 |        27414 |          0 |
| Alibaba | review_comment |       48412 |        48412 |          0 |
| Baidu   | review_comment |       35072 |        35072 |          0 |
| Alibaba | merged_pull    |       26732 |        26732 |          0 |
| Baidu   | merged_pull    |       21680 |        21680 |          0 |
| Alibaba | rank           |           1 |            5 |         -4 |
| Baidu   | rank           |           2 |           10 |         -8 |
| Alibaba | value          |      103368 |       103368 |          0 |
| Baidu   | value          |     71636.8 |      71636.8 |          0 |
| Alibaba | rankDelta      |           0 |            1 |         -1 |
| Baidu   | rankDelta      |           0 |            1 |         -1 |
| Alibaba | valueDelta     |     21093.1 |      21093.1 |          0 |
| Baidu   | valueDelta     |     10032.1 |      10032.1 |          0 |
+---------+----------------+-------------+--------------+------------+
18 rows in set (0.00 sec)
```

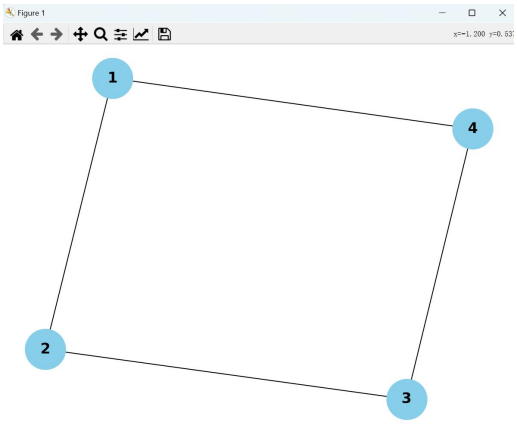
```
mysql> SELECT * FROM comparison_2;
+---------------+----------------+-----------+-------------+--------------+------------+
| name          | metric         | statistic | china_value | global_value | difference |
+---------------+----------------+-----------+-------------+--------------+------------+
| All Companies | issue_comment  | max       |      167814 |      1437320 |   -1269500 |
| All Companies | issue_comment  | min       |       11741 |        78530 |     -66789 |
| All Companies | issue_comment  | avg       |     61205.5 |       341409 |    -280204 |
| All Companies | open_issue     | max       |       22397 |       189185 |    -166788 |
| All Companies | open_issue     | min       |         752 |        13162 |     -12410 |
| All Companies | open_issue     | avg       |      9169.2 |      43300.2 |     -34131 |
| All Companies | open_pull      | max       |       35266 |       309685 |    -274419 |
| All Companies | open_pull      | min       |        1823 |        27414 |     -25591 |
| All Companies | open_pull      | avg       |     16912.7 |        83423 |    -66510.3 |
| All Companies | review_comment | max       |       60402 |       456166 |    -395764 |
| All Companies | review_comment | min       |        2113 |        35072 |     -32959 |
| All Companies | review_comment | avg       |     19857.5 |       120392 |    -100535 |
| All Companies | merged_pull    | max       |       26732 |       257123 |    -230391 |
| All Companies | merged_pull    | min       |        1165 |        15418 |     -14253 |
| All Companies | merged_pull    | avg       |     13764.1 |        62472 |    -48707.9 |
| All Companies | rank           | max       |          10 |           10 |          0 |
| All Companies | rank           | min       |           1 |            1 |          0 |
| All Companies | rank           | avg       |         5.5 |          5.5 |          0 |
| All Companies | value          | max       |      103368 |       824849 |    -721480 |
| All Companies | value          | min       |     12033.7 |      71636.8 |    -59603.1 |
| All Companies | value          | avg       |     40269.5 |       215855 |    -175586 |
| All Companies | rankDelta      | max       |          25 |            1 |         24 |
| All Companies | rankDelta      | min       |           0 |           -2 |          2 |
| All Companies | rankDelta      | avg       |         5.3 |          0.1 |        5.2 |
| All Companies | valueDelta     | max       |     21093.1 |      57536.1 |    -36443 |
| All Companies | valueDelta     | min       |     2329.36 |     -47388.6 |    49717.9 |
| All Companies | valueDelta     | avg       |     9265.01 |      11906.9 |    -2641.91 |
+---------------+----------------+-----------+-------------+--------------+------------+
27 rows in set (0.00 sec)
```

4. 已知一个项目带有 *HTML/Markdown* 标签，那么该项目是非软件型项目的概率是多少？

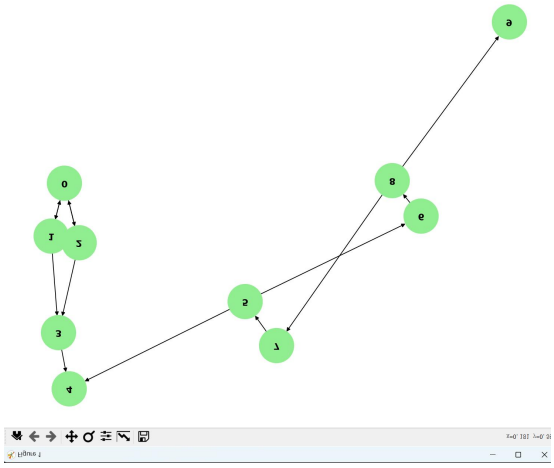
5. 接上文，已知一个项目是由 *JavaScript* 语言编写的，那么它是工具组件型项目的概率是多少？



6. 根据以下数据建立可视化无向图。*user = [1, 2, 3, 4]*，*edge = [(1, 2), (2, 3), (3, 4), (4, 1)]*



7. 根据以下数据建立可视化**有向图。**users = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]，edges = [(0, 1), (1, 0), (0, 2), (2, 0), (1, 2), (2, 1), (1, 3), (2, 3), (3, 4), (5, 4), (5, 6), (7, 5), (6, 8), (8, 7), (8, 9)]



8. 针对第七题构建的有向图，计算并输出每个节点的 pagerank 值。同时根据 pagerank 调整可视化图的大小，使得 PageRank 越大的节点在可视化结果中也越大。