# Are you data freak?

## Then prove us your data analysis skills, testing abilities and programming proficiency!

## Here you have some data to dive into:

Link: https://egnyte.egnyte.com/dl/qvEcZIyFcb
Password: DataTask2019

## Data Description:

```
1  dataset_contains_multiple_log_file = True
2  each_row_is_a_JSON_object = True
3
4  if file_name.startswith('fse_data'):
5          print("This is action performed by Egnyte Cloud Server")
6  if file_name.startswith('domain_data'):
7          print("This is data about one Egnyte Connect domain")
```

## Task Description:

- ✔ Data analysis & engineering
  - Go through the provided data, focus on actions, workgroups, users, domains characteristics.
  - Create statistics and graphs regarding various features found in logs.
  - Automate above with a program that will create statistics, charts etc. without manual interference - imagine that source files will be updated everyday.

    - Focus on writing program in consistence with best coding practices - modularity, error-proof and correct behaviour in corner cases.
- ✔ Data QA
    - Suggest up to 10 test cases that will check data correctness, completeness, uniqueness (if applicable), etc.
    - Focus on unit testing, but If you have ideas provide info about what kind of other tests should be done.
- ✔ Report
    - Create a report in PDF or HTML.
    - Upload all source code and the report to newly created GitHub repo and send us a link.
    - Report should contain:
        - Explanation of basic characteristics, trends, and relations that you found
        - Appropriate graphs
        - All discrepancies that you found
        - All results of tasks you performed with step by step description what you did
        - Additional sources that you used when doing this task (bibliography of sorts)

Use any technology you want, but you will get additional points for Python or Scala and related technologies.

You have 7 days counting from today, but if you need more time, please let us know.

**Do not hesitate to contact us, Tymek (thossa@egnyte.com) can help you if you have any questions regarding data or task!**

# Good luck!