**Programming languages:**

In 'A Study of External Community Contribution to open-source Projects on GitHub; This article USES the boxplot analysis of data, of which this article use the data features for developers that use the programming language, in our 2 data set as the **‘language’**, the data reflects the language used by the project development, and whether the developer as the core members, including developers can be divided into three kinds of article, However, we provide only one feature in the data set to determine whether the developer is a core member, namely **'core\_member'.**

In the article 'An Insight into the Pull Requests of GitHub', a bar chart was used to measure the number of developers using different languages and whether the merge was successful. In our data set, we have the **"“merged\_or\_not‘"** feature to determine whether the request was successfully merged.

In the article 'Acceptance Factors of Pull Requests in open-source Projects', association rules are used to quantify the influence of project language on Pull Requests, to calculate the correlation between project language and Pull Requests. And use a bar chart to display.

**Popularity of project:**

The ‘**fork\_num’** feature of a dataset can be used to determine the popularity of an item. The more branches an item has, the more popular it is. The number of branches can be used to determine the popularity of an item by looking for a relationship between the number of branches and pull request requests. At the same time, the relationship between these two features can be used to judge whether the success rate of pull request is related to the popularity of the project.

**Age of project:**

There is a feature called **'project\_age'** in the dataset that can be used to determine the time between the pull request and the project. The longer the project is created, the higher the maturity of the project, thus determining the relationship between the maturity of the project and the success rate of the pull request

**Workload of a project:**

A feature **‘open\_pr\_num’** exists in the dataset which represents the number of other pull requests that already exist when a new pull request is requested to be uploaded. This feature can be analysed in relation to whether the pull request has been merged or not to determine the impact of workload on the success rate of the pull request.

**Activeness of project:**

There is a feature **'push\_delta'** in the dataset, which refers to the time interval between two pull request open times, the shorter the time interval, the more active the item is, thus allowing the correlation between item activity and pull request acceptance to be explored.

**Openness of a project:**

The feature **'open\_issue\_num**' exists in the dataset and refers to the number of issues that exist when a new pull request is opened. The more issues that exist, the more open the project is and the relationship between the openness of the project and the likelihood of the pull request being accepted can be explored.

Dataset list:

**'project\_id':** The id of each item, using this feature all data for the same item can be aggregated for analysis

**'language':** The language in which the project is developed, and whether the use of a different language has an impact on the success of the pull request can be counted. This should be analysed in conjunction with 'merged\_or\_not' to show the number of pull requests merged between languages, using both bar and box plots.

**'project\_age':** The time interval between the creation of the project and the creation of the pull request, generally the longer the interval the more mature the project. This can also be combined with 'merged\_or\_not' to determine that the earlier the project is created, the less likely it is to be accepted. It can also be combined with 'open\_issue\_num', as the longer the project has been created, the less likely it is to be accepted.

**'pushed\_delta':** This can be used to determine the activity level of a project, generally the higher the activity level, the higher the likelihood of a pull request being accepted

**'pr\_succ\_rate':** Acceptance rate of pull requests in the project

**'open\_issue\_num':** This feature and 'pr\_succ\_rate' can be used to determine the openness of the project and thus explore the relationship between the openness of the project and whether the pull request will be accepted

**'open\_pr\_num'：**The number of pull requests opened when a pull request is submitted is used to demonstrate how much work the project is doing, and thus how the workload relates to whether the pull request will be accepted.

**'fork\_num':** The number of branches, generally the higher the number of branches, the more popular the project is and the relationship between the popularity of the project and whether the pull request will be accepted can be determined.

**'merged\_or\_not':** Used to record whether the pull request has been accepted and merged