**Kubernetes - Pull Request Analysis**

1. This analysis is on the data from Kubernetes’s master repository for all its closed pull requests which are either merged or rejected, containing a total of 54,645records
2. GitHub repository comparison:

Closed pull request’s count: 55,235

Closed pull request’s count which are merged: 42,701

1. Data Description:

|  |  |
| --- | --- |
| pull\_request\_id | Is a unique id within a repository representing a pull request made |
| loc\_additions | Lines of code added in the pull request |
| loc\_deletions | Lines of code deleted in the pull request |
| num\_files\_changed | Number of files changed in the pull request |
| pr\_src\_state | Pull request’s state |
| pr\_src\_title | Pull request’s label/description |
| merged\_status | Closed PR’s detailed state – Merged/Rejected |

1. Outliers Analysis:

Number of files changed:

A screenshot of a social media post

Description automatically generated

Lines of code deleted:

A screenshot of a social media post

Description automatically generated

Lines of code added:

A screenshot of a social media post

Description automatically generated

1. Logistic regression models:

Descriptive statistics for entire data with outliers

Dataset count: 54,645

A picture containing battery

Description automatically generated

Logistic Regression Model for the entire data with outliers

A screenshot of a cell phone

Description automatically generated

----------------------------------------------------------------------------------------------------------------

Descriptive statistics for normal range – data without any outliers

Dataset count: 42, 602

A close up of a device

Description automatically generated

Logistic Regression Model for normal range - data without any outliers

A screenshot of a cell phone

Description automatically generated

----------------------------------------------------------------------------------------------------------------

Descriptive Statistics of All data excluding Num of files added Outliers

Dataset count: 47,487

A close up of a device

Description automatically generated

Regression Logistic for All data excluding Num of files added Outliers

A screenshot of a cell phone

Description automatically generated

Descriptive statistics for All data excluding lines of code deleted Outliers

Dataset count: 46,251

A screenshot of a cell phone

Description automatically generated

Logistic Regression Model for All data excluding lines of code deleted Outliers

A screenshot of a cell phone

Description automatically generated

Descriptive statistics for All data excluding lines of code added Outliers

Dataset count: 47,128

A picture containing battery

Description automatically generated

Logistic Regression Model for All data excluding lines of code added Outliers

A screenshot of a cell phone

Description automatically generated

----------------------------------------------------------------------------------------------------------------

Descriptive statistics for data with number of files added Outliers only

Dataset count: 7,158

A close up of a battery

Description automatically generated

Logistic Regression Model for data with number of files added Outliers only

A screenshot of a cell phone

Description automatically generated

Descriptive statistics for data with lines of code deleted Outliers only

Dataset count: 8,394

A picture containing battery

Description automatically generated

Logistic Regression Model for data with lines of code deleted Outliers only

A screenshot of a cell phone

Description automatically generated

Descriptive statistics for data with lines of code added Outliers only

Dataset count: 7,517

A picture containing battery

Description automatically generated

Logistic Regression Model for data with lines of code added Outliers only

A screenshot of a cell phone

Description automatically generated

All Outliers combined

Descriptive Statistics

A picture containing battery

Description automatically generated

Logistic Regression model

A screenshot of a cell phone

Description automatically generated

**Complexity Analysis**:

Merged Status’s Top 10 **num\_files\_changed** by their pull request ids

Merged Status’s Top 10 **lines of code deleted** by their pull request ids

Merged Status’s Top 10 **lines of code added** by their pull request ids