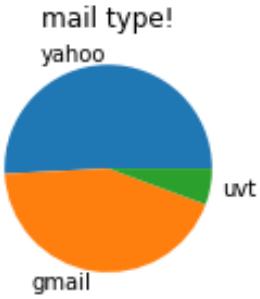
Statistics report

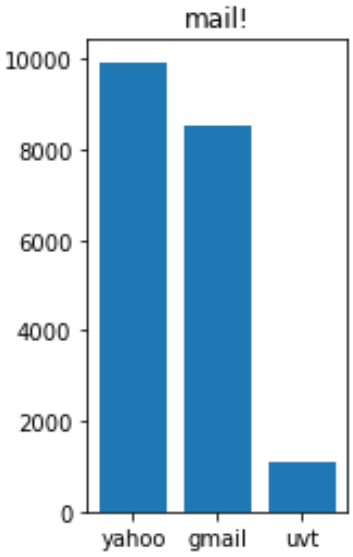
Statistics report

Date: 2023-07-14

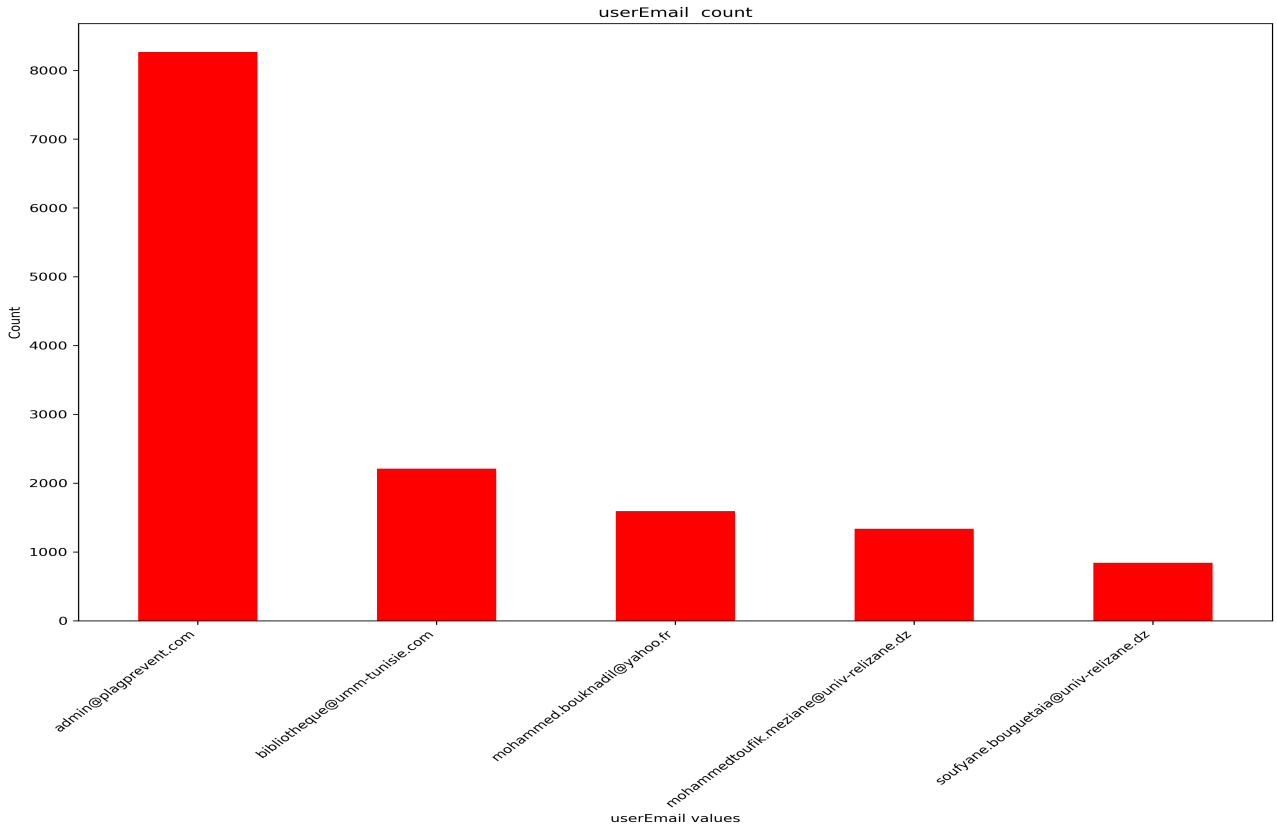
Author: admin

In this report, we present the analysis of the dataset collected from our research team. The dataset comprises agenda jobs. Our objective is to explore the dataset and uncover key insights and patterns through statistical analysis and visualization techniques.

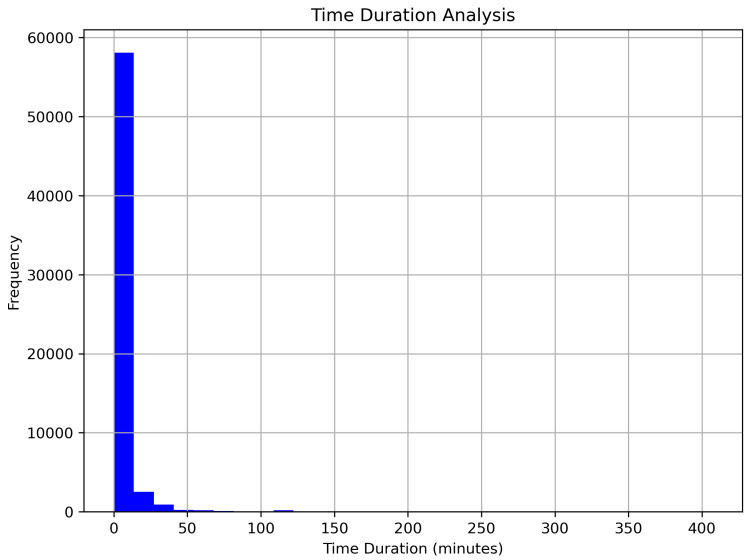




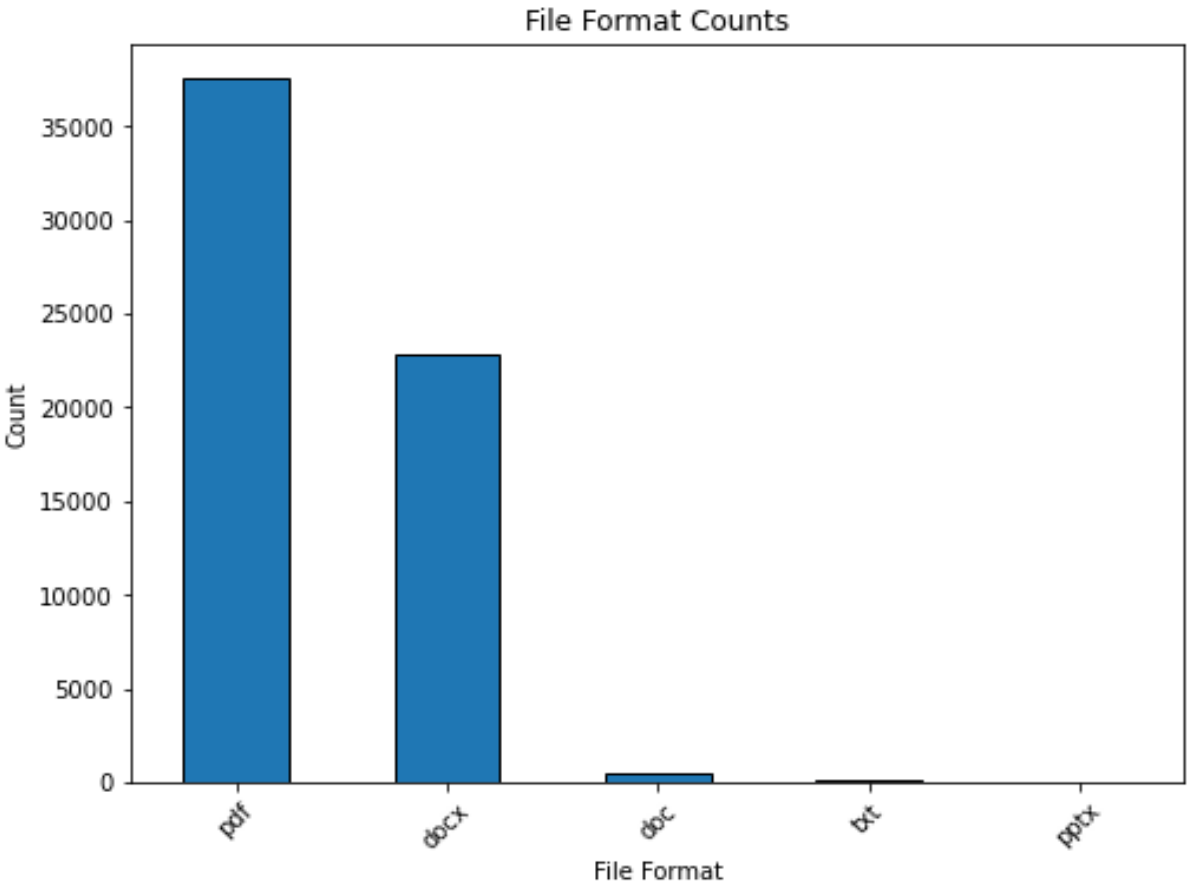
using those 2 figures, we can see the most used types of emails between those 3:gmail,yahoo,uvt . we choosed them because they are the most known 3 types of emails nowadays.



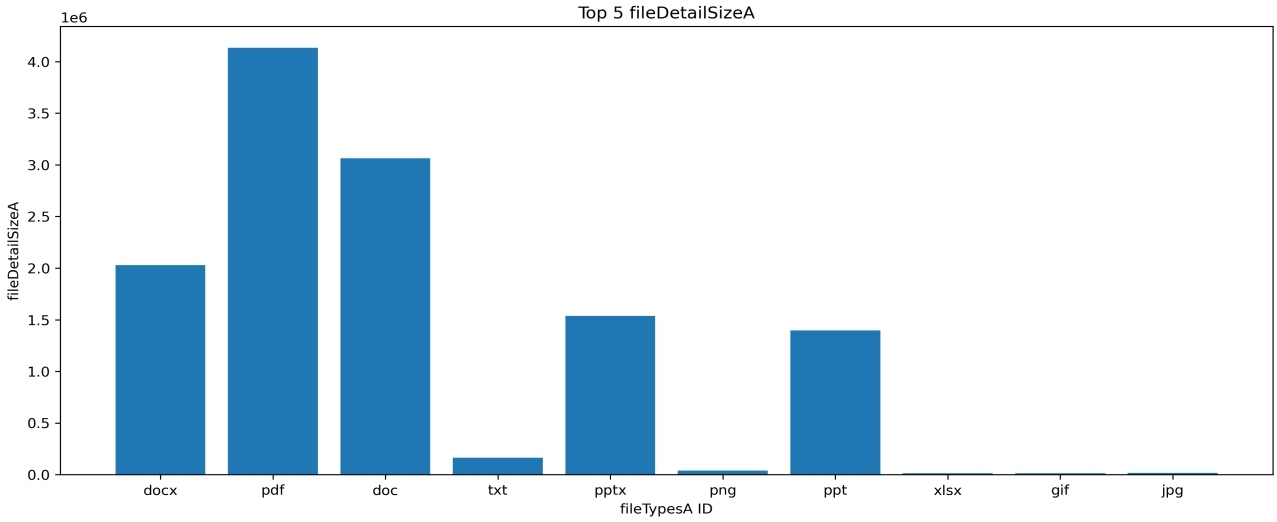
here, we can see the most used emails in the dataset.This information can be valuable for understanding the email landscape and identifying potential communication channels.



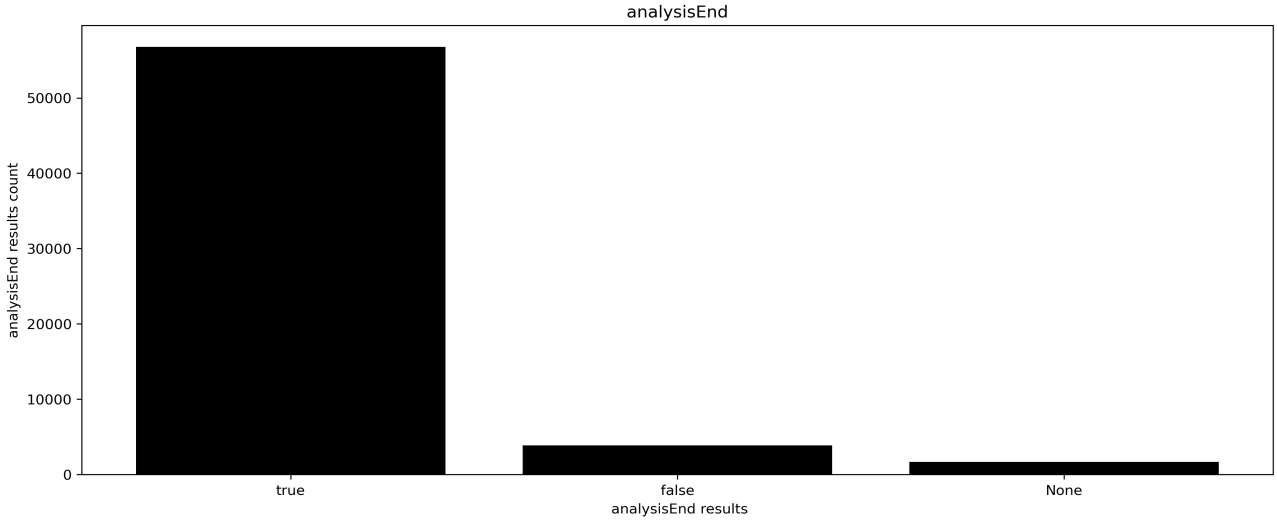
In order to gain insights into the duration of various processes and runs, we have conducted a comprehensive time duration analysis.



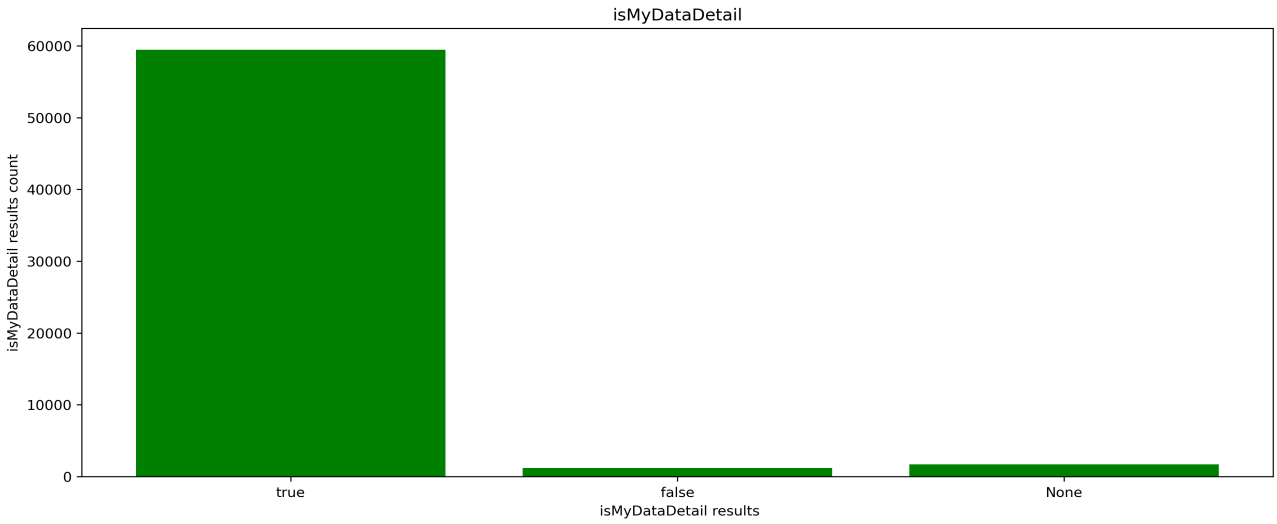
This distribution provides insights about the most used type of files in the database.



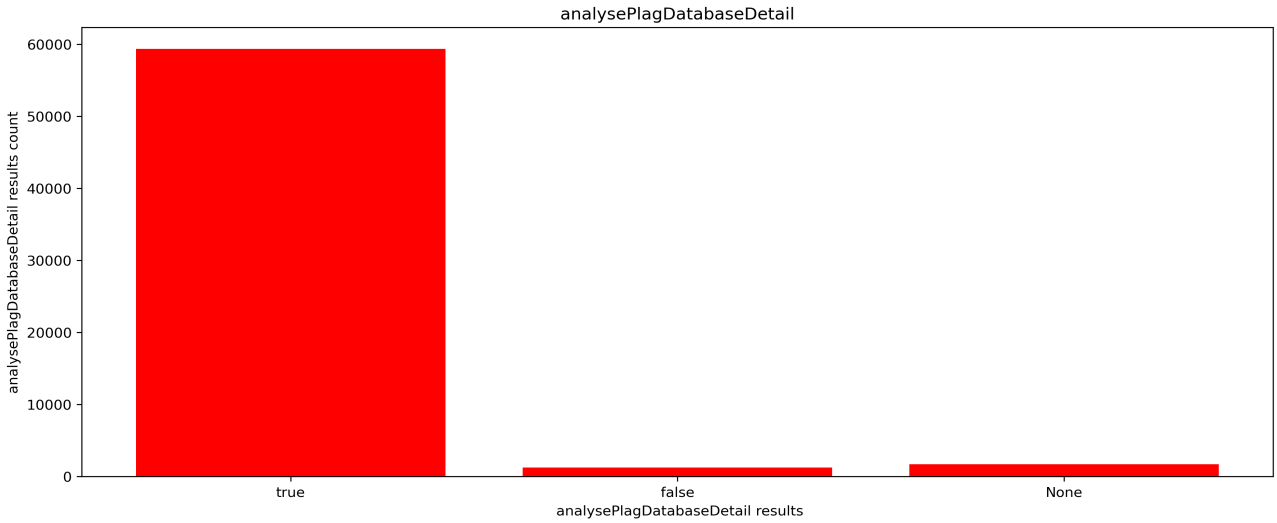
in this plot, we used the most known file types and we calculated the total size of each of them in the dataset.



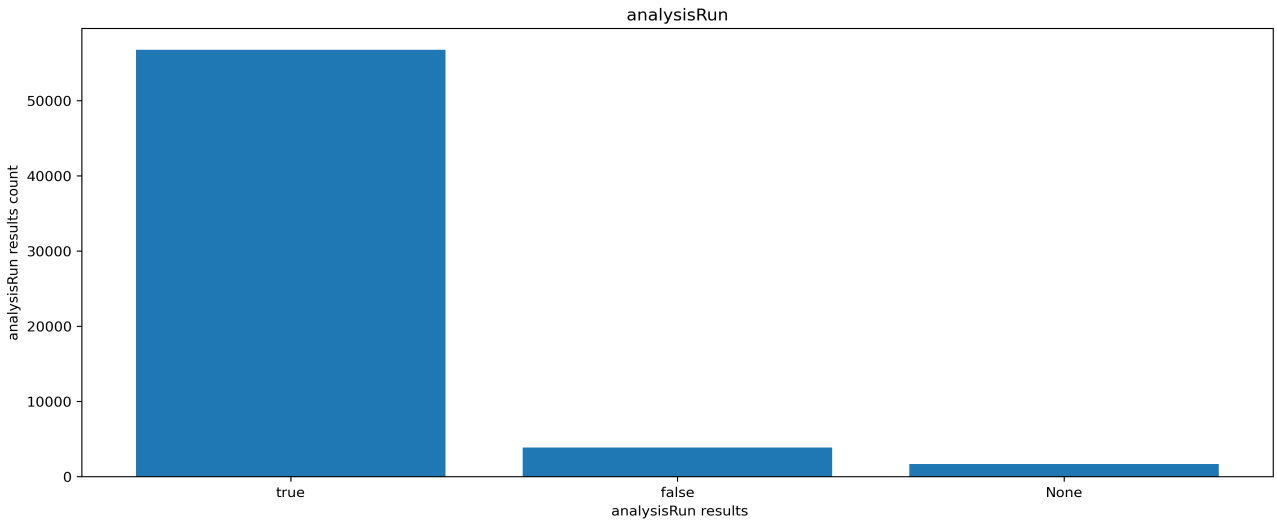
this plots shows the values of -isanalyseEnd- .



and this shows the values of -isMyData- .



those are the values of -analysePlagDatabase- .



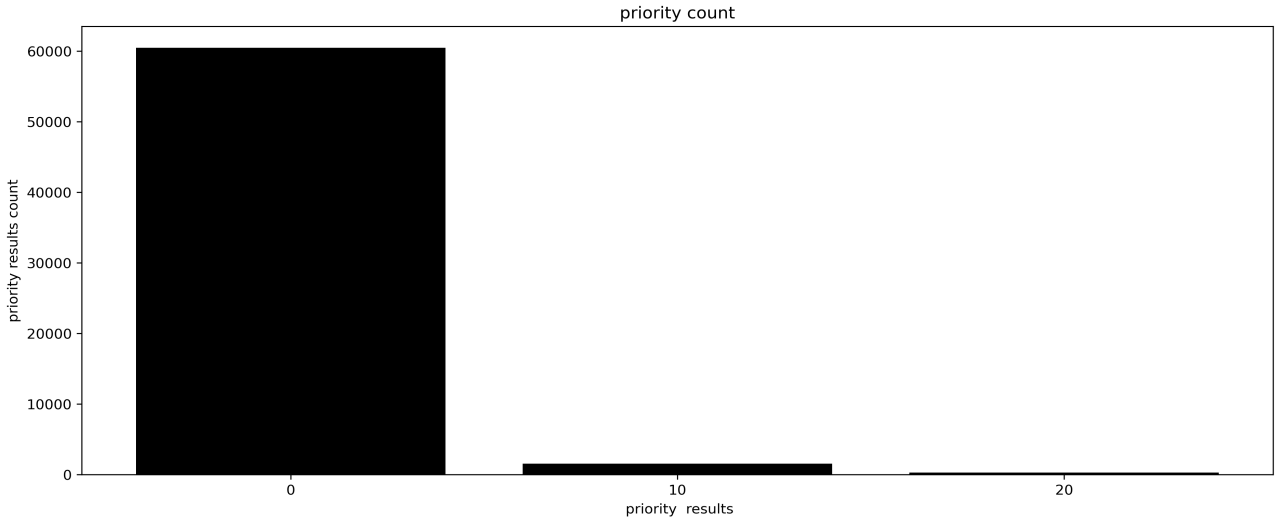
and those are the values of -analyseRun- .

name	True	False	None
isAnalysedEnd	56724	57	0
isPlagDatabase	55877	904	0
isMyData	56016	765	0

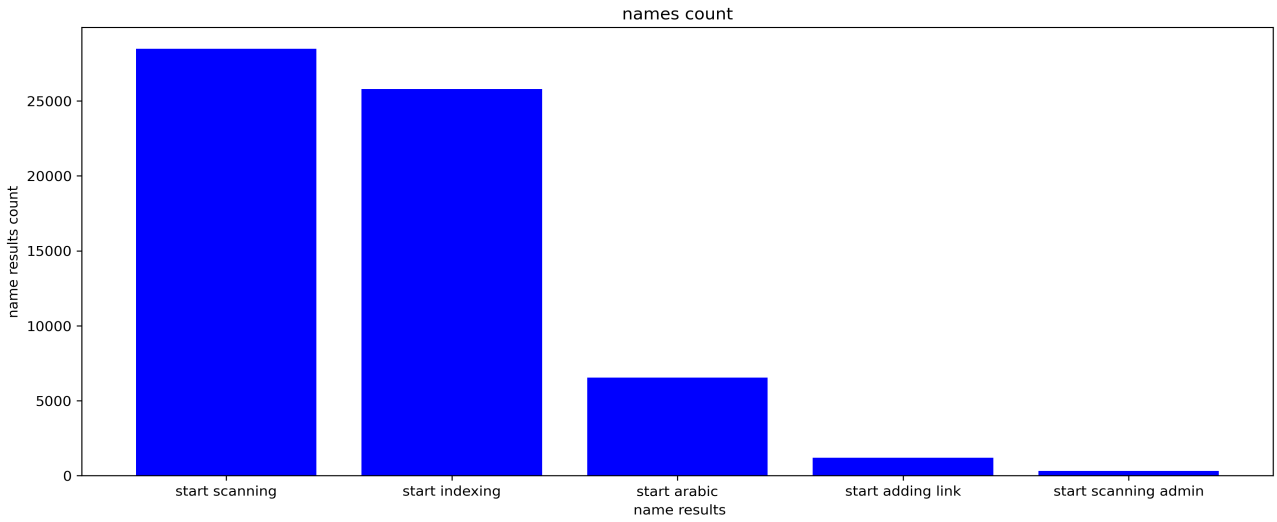
this table shows us the values of other data if -analyseRun- is true , maybe it can helps in some patterns.

name	True	False	None
isAnalysedEnd	62	3795	0
isPlagDatabase	3500	357	0
isMyData	3438	419	0

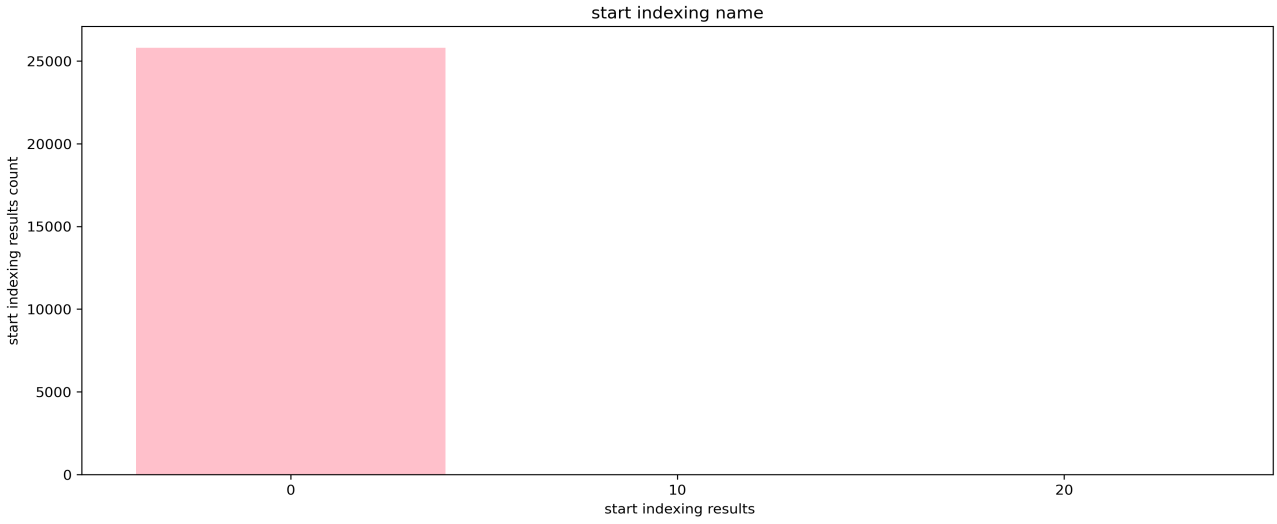
this table shows the values of data when -analyseRun- is False.



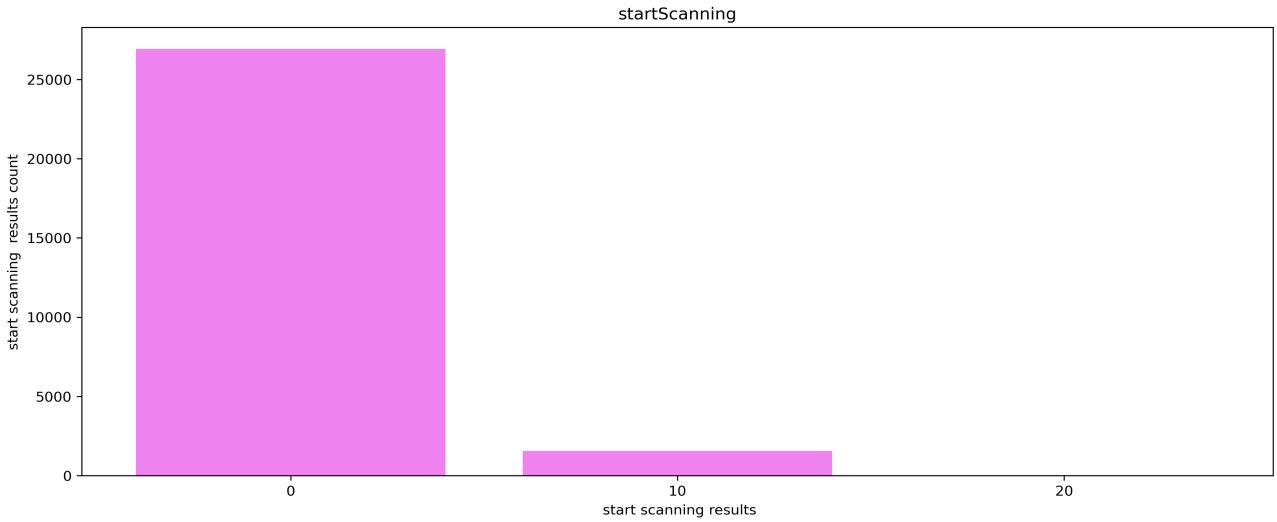
By considering the priority level of each processus, we can effectively allocate resources and attention to the most critical data sources.



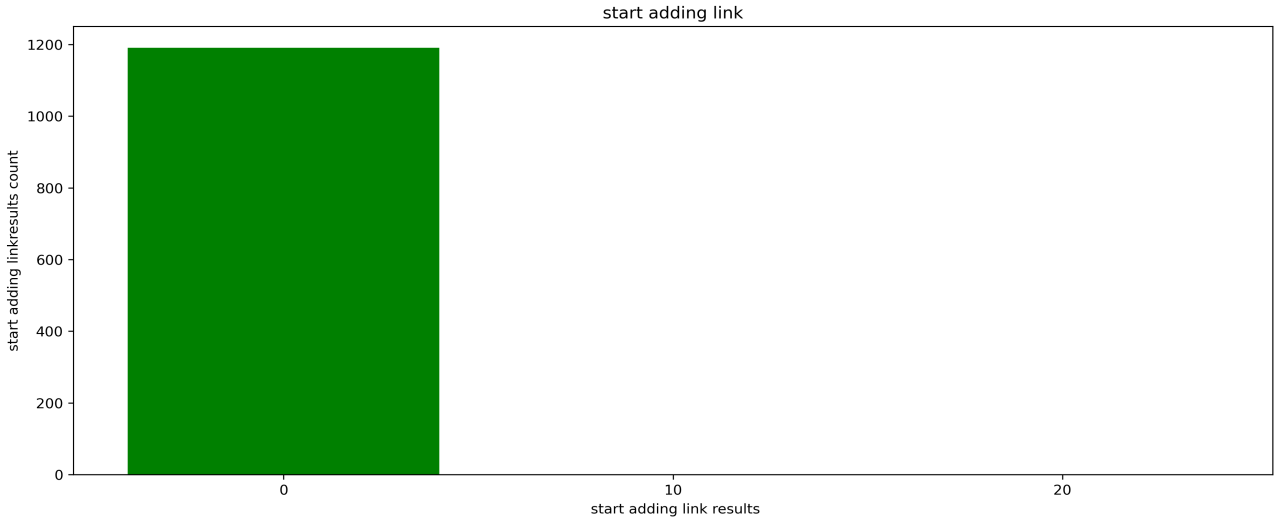
we can see through this plot the most common documents names we have.



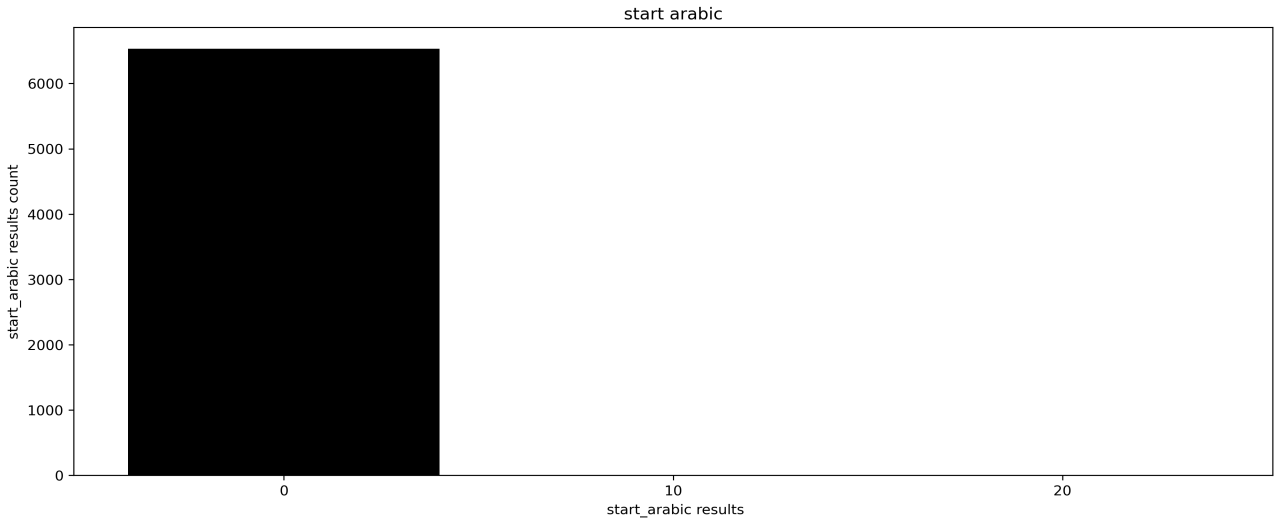
this figure shows us the -priority- count when the name of the collection is -start indexing- .



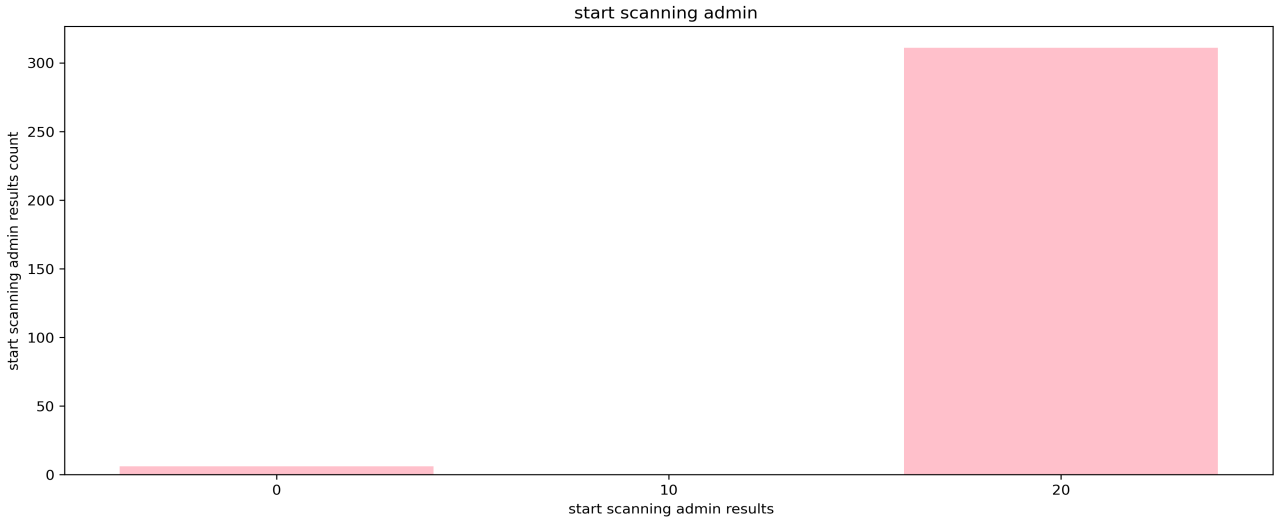
this figure shows us the -priority- count when the name of the collection is -start scanning- .



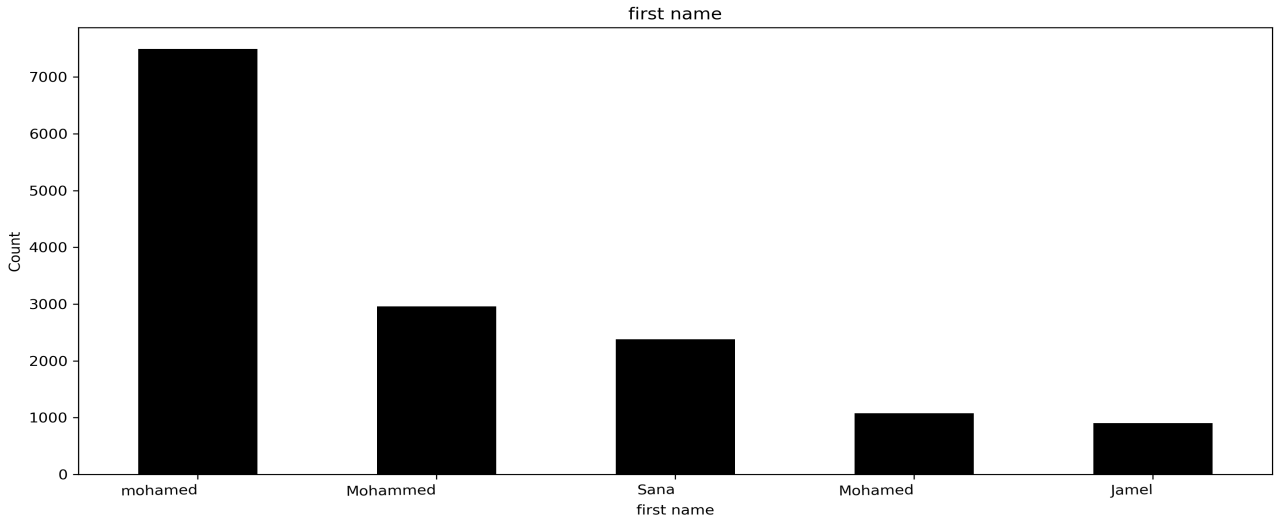
this figure shows us the -priority- count when the name of the collection is -start adding link- .



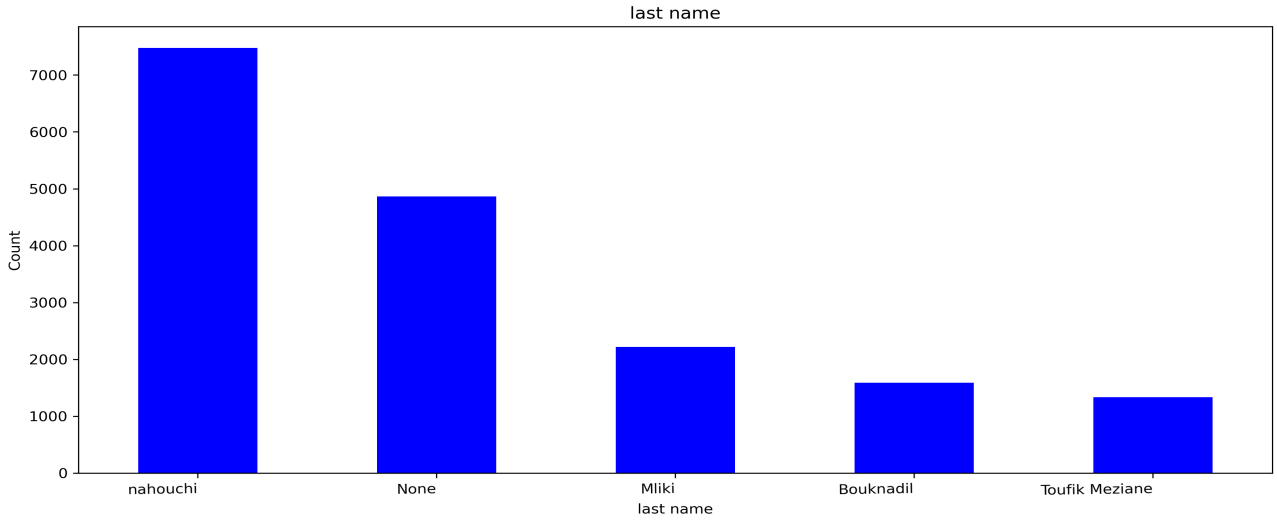
this figure shows us the -priority- count when the name of the collection is -start arabic- .



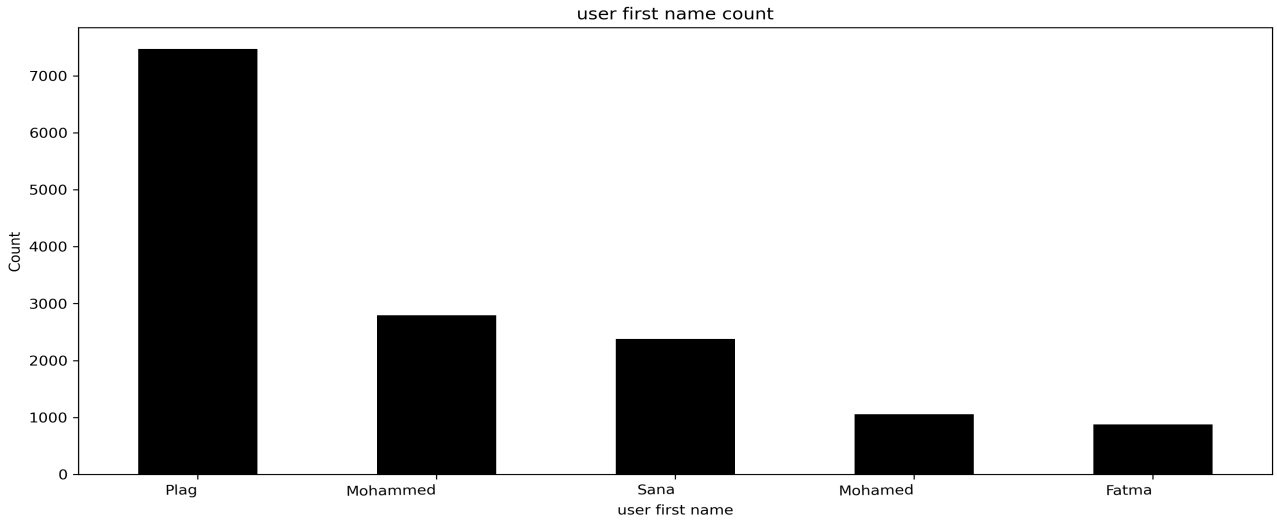
this figure shows us the -priority- count when the name of the collection is -start scanning admin- .



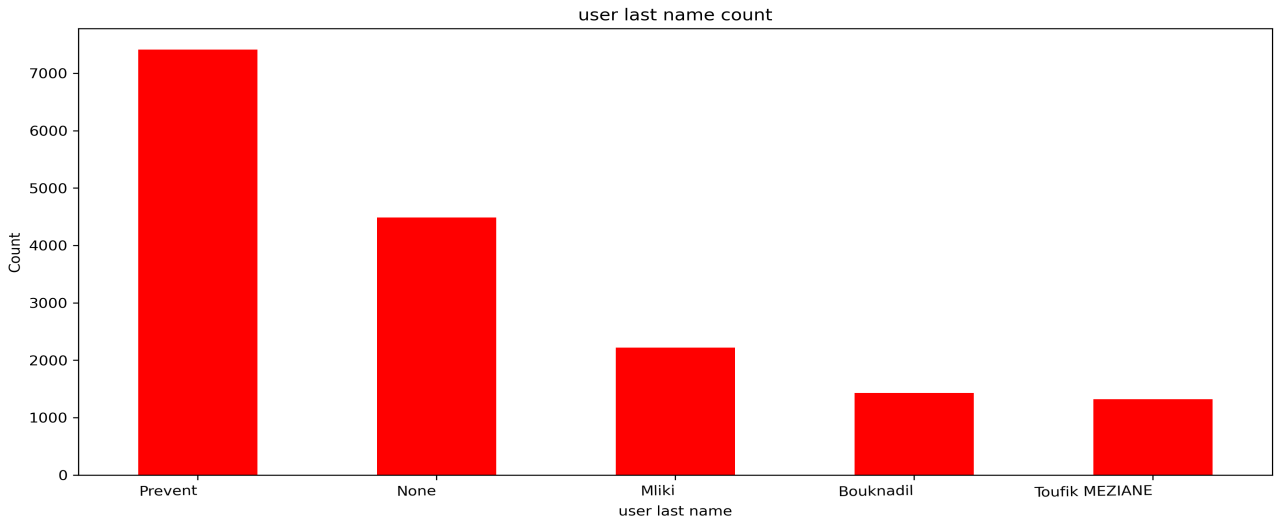
here we have the most common -Firstnames- in the collection.



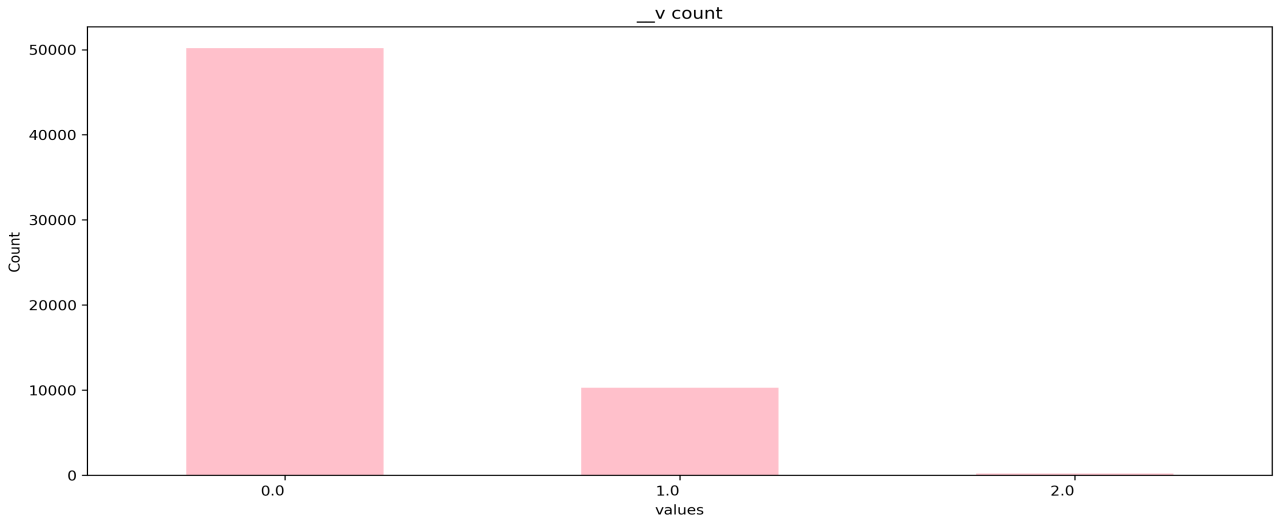
and here we have the most common -lastnames- .



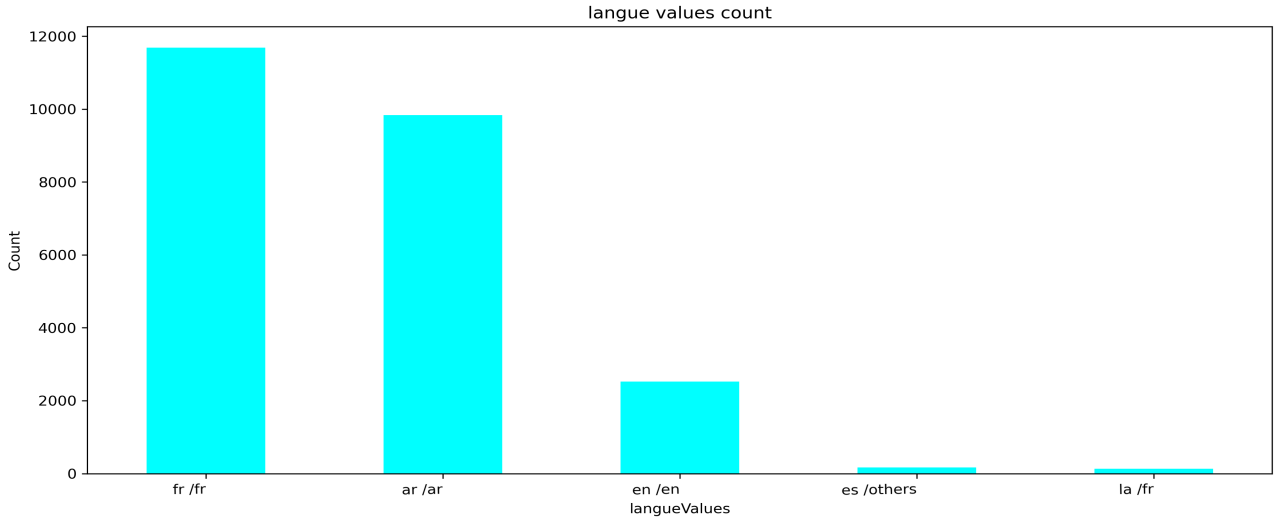
this plot is for -userFirstname-,it shows the most common userfirstnames.



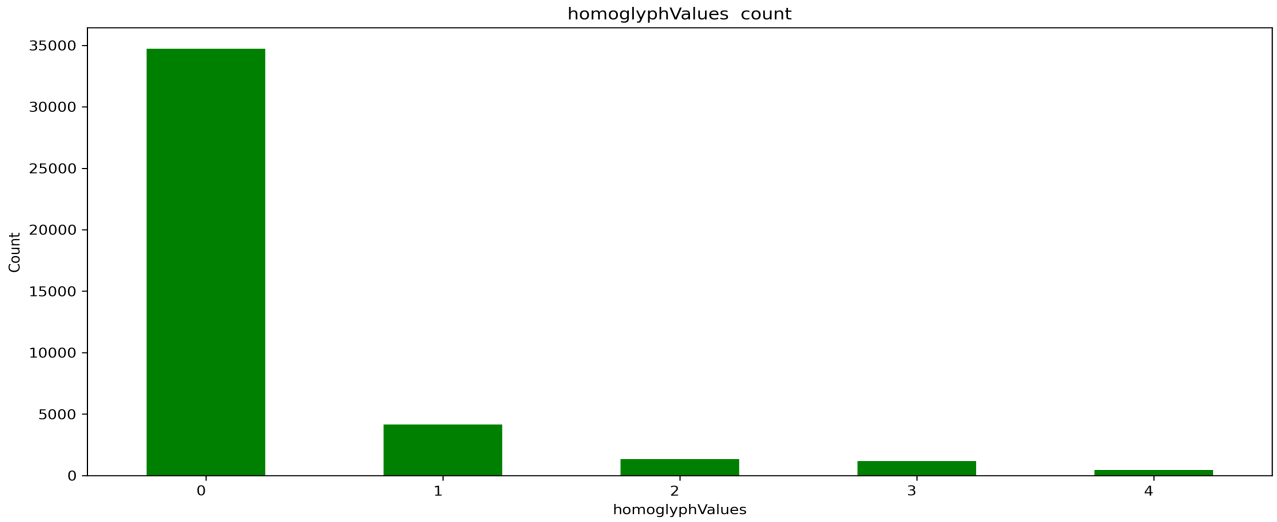
and this is for the most common userlastnames.



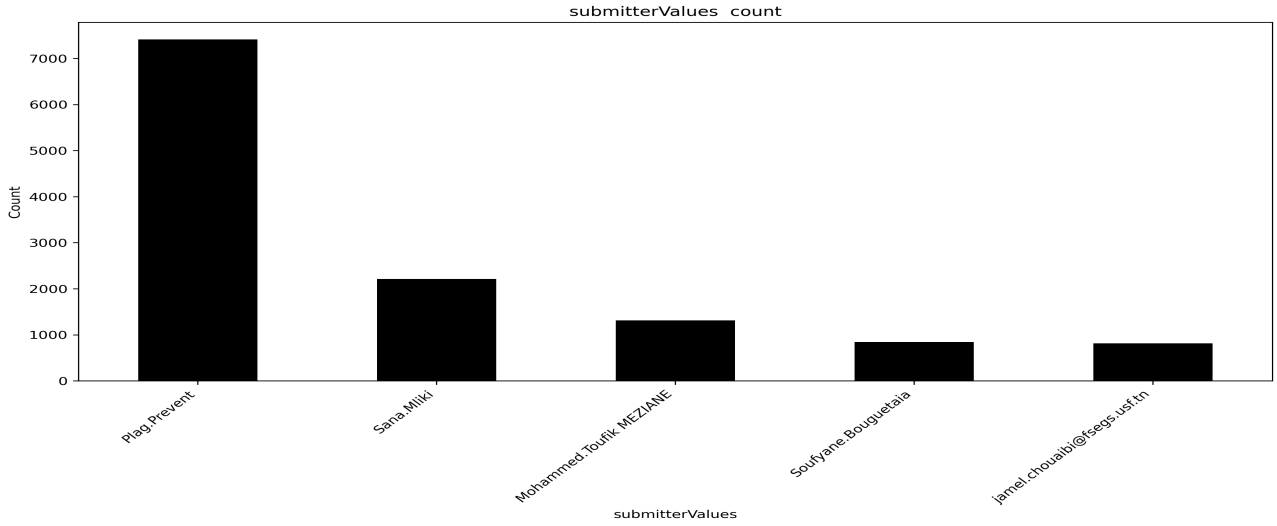
this plot presents the value counts for - __v - variable.



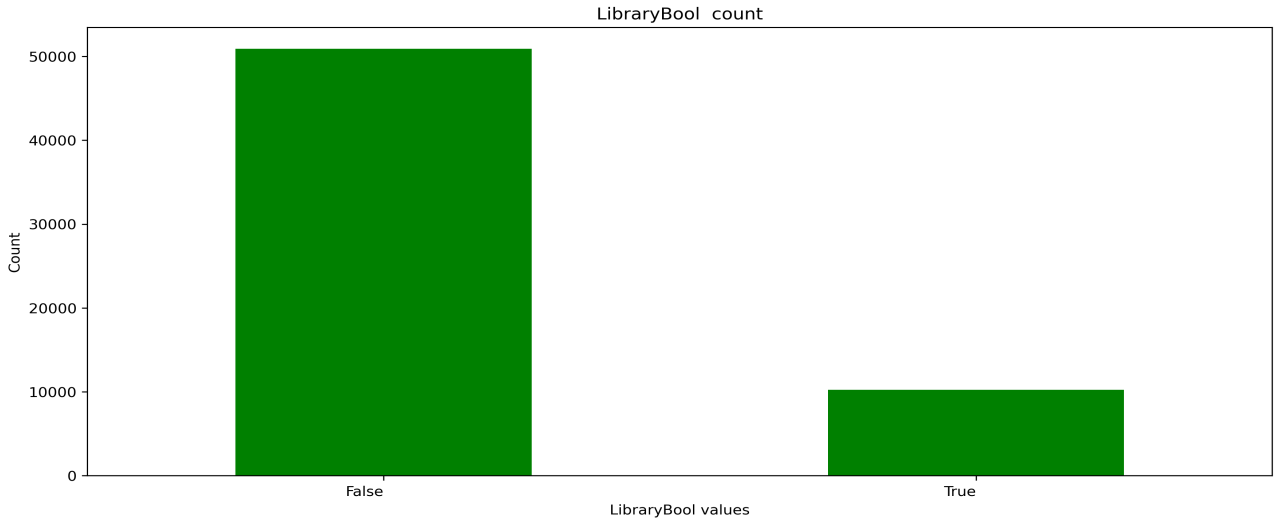
By identifying the most commonly used languages, we can optimize our content localization efforts and ensure effective communication with our target audience.



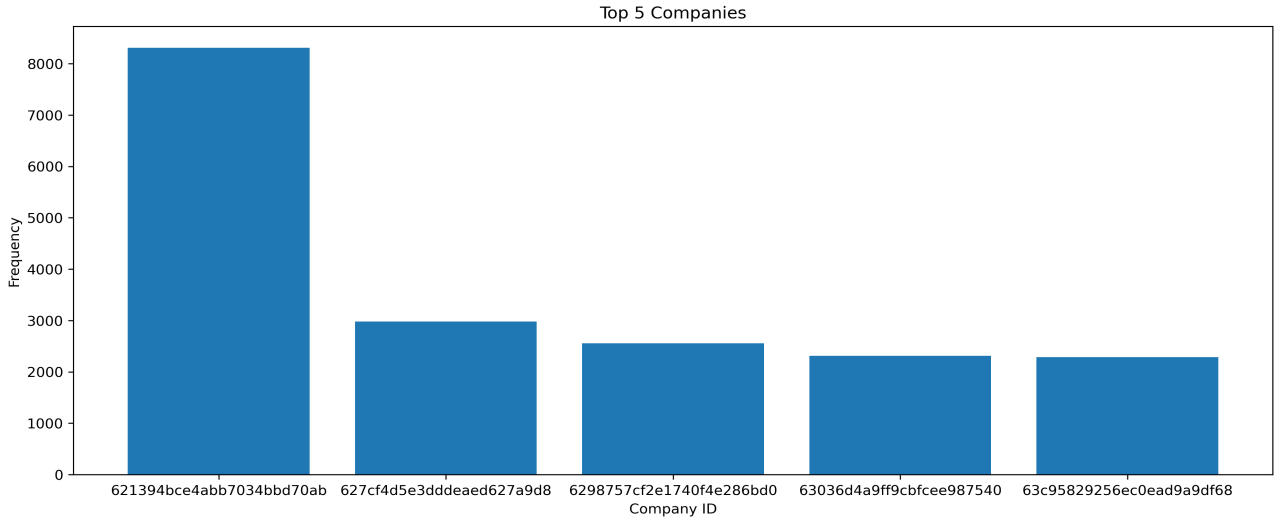
this figure shows us the values of the variable -homoglyph- .



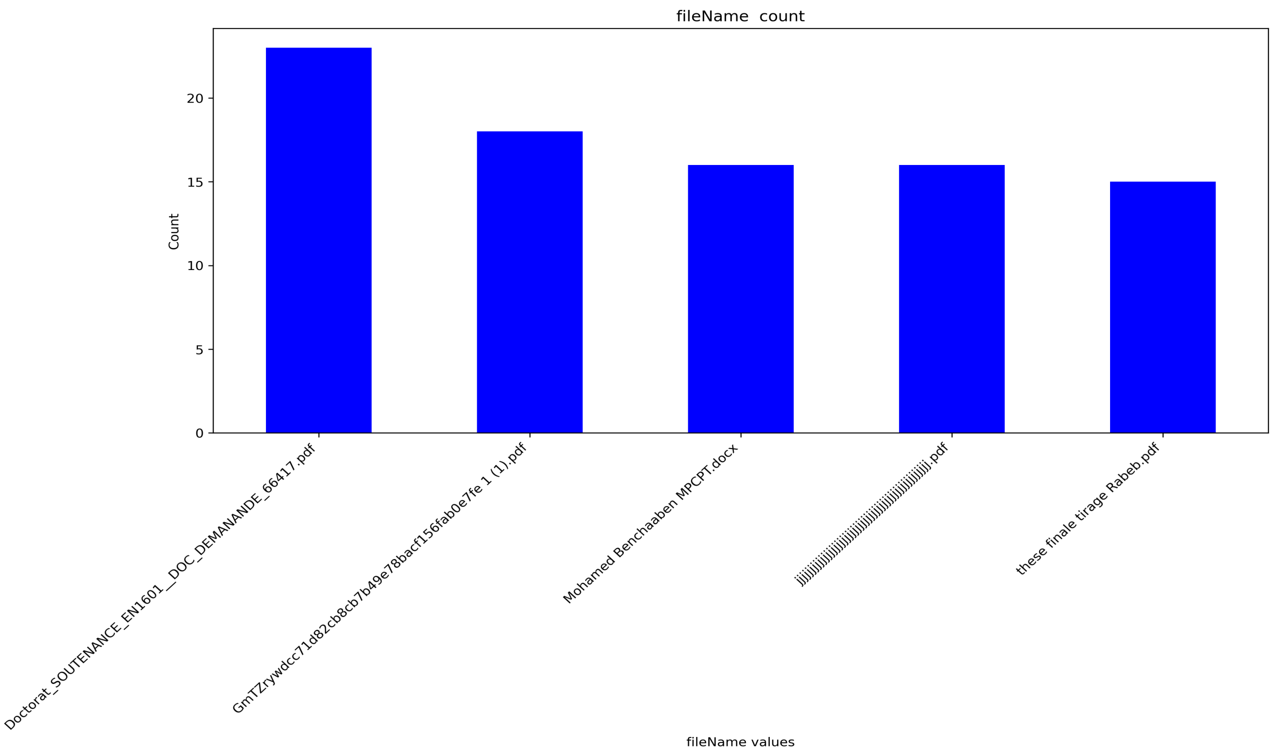
those are the most common submitters in the dataset.



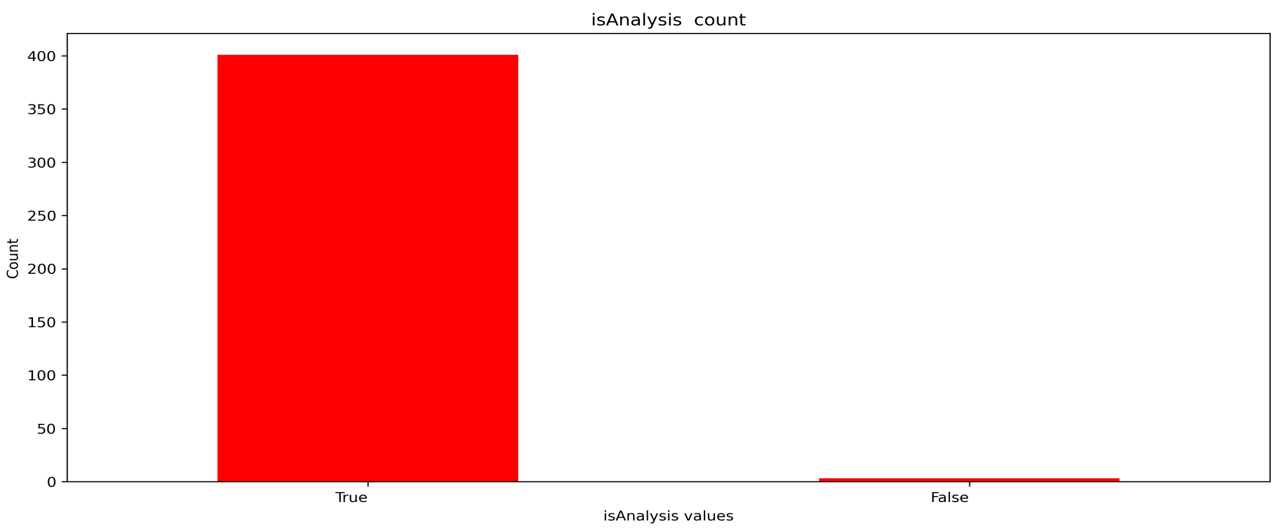
this plot shows us the values of -libraryBool- variable.



The plot provides insights into the occurrence of companies in the dataset, allowing for easy identification of the most common or prevalent companies.

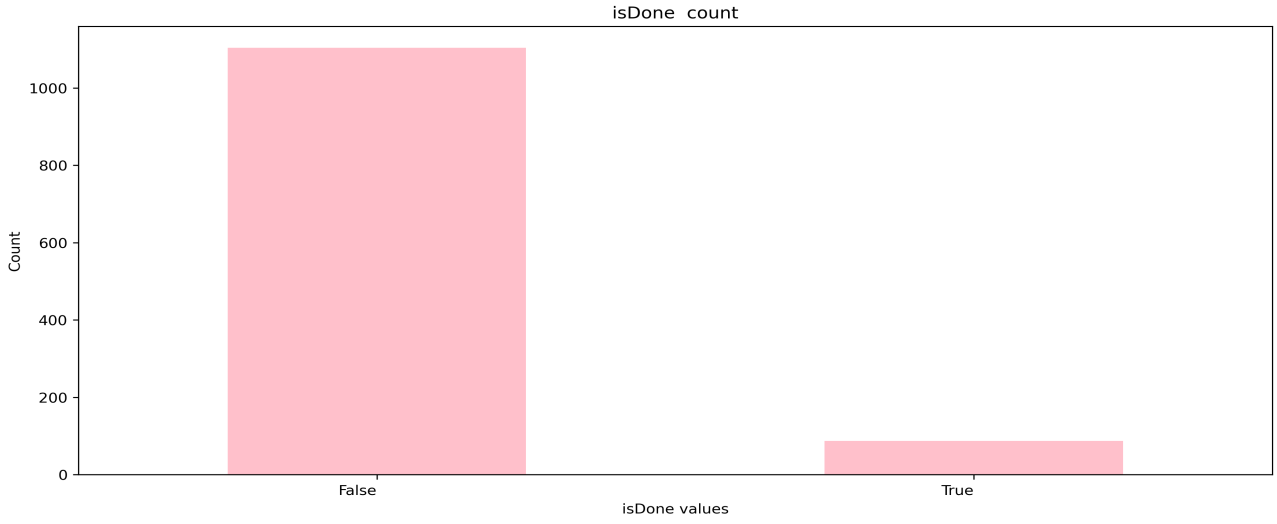


By examining the -filename- plot, you can quickly identify the most common or popular file names in the dataset. This information can be useful for understanding naming conventions, identifying naming trends, or making decisions related to file management or organization.

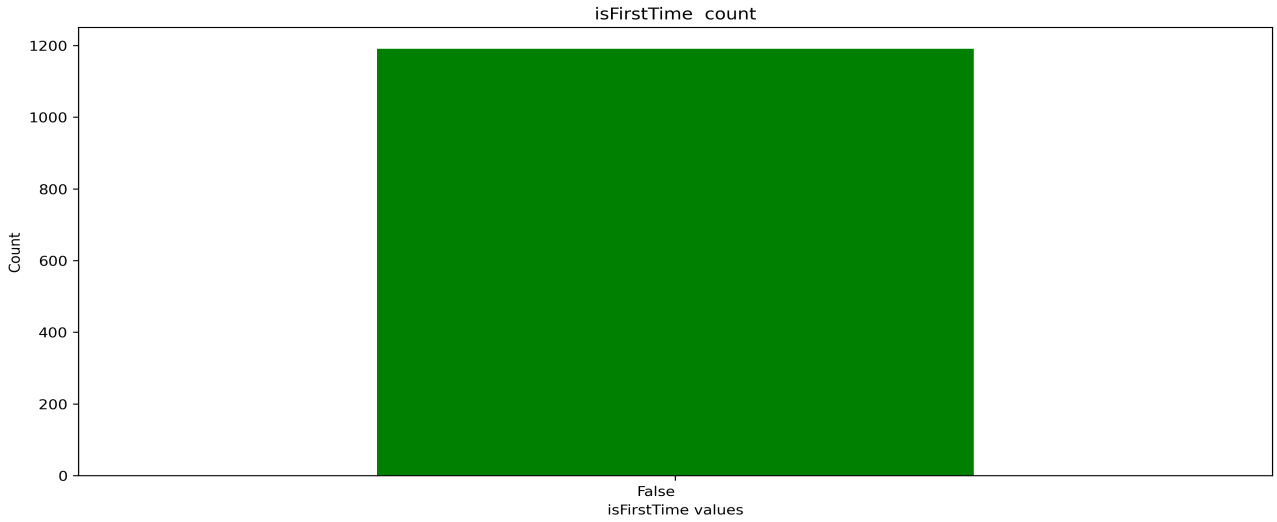


from here, we will examine the data in the archive array ,this plot shows the values of -isAnalysis- Variable.

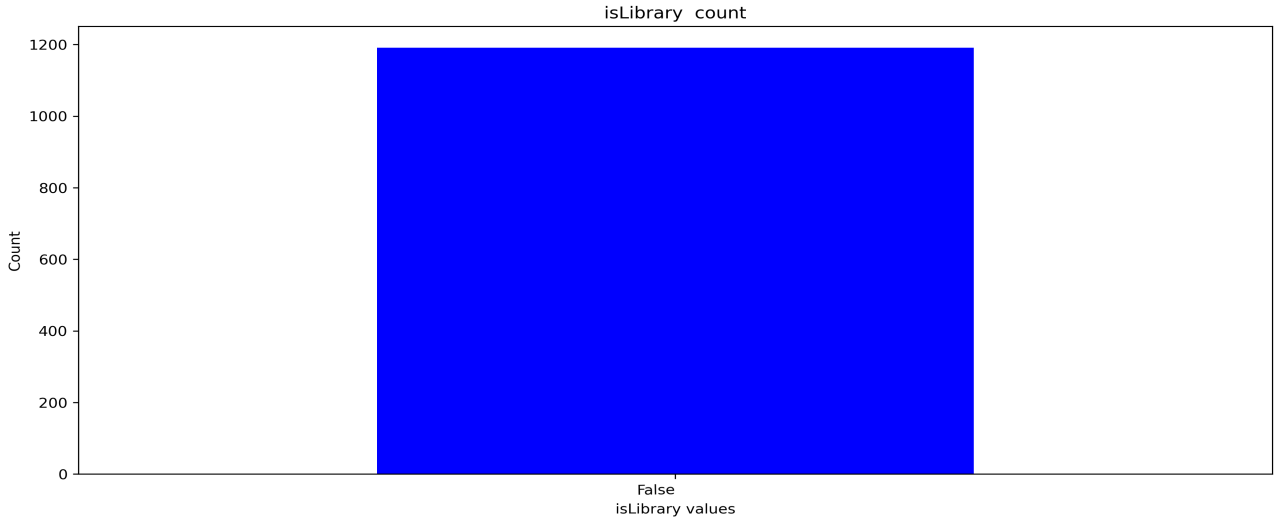
this plot is for -isDeleted- variable.



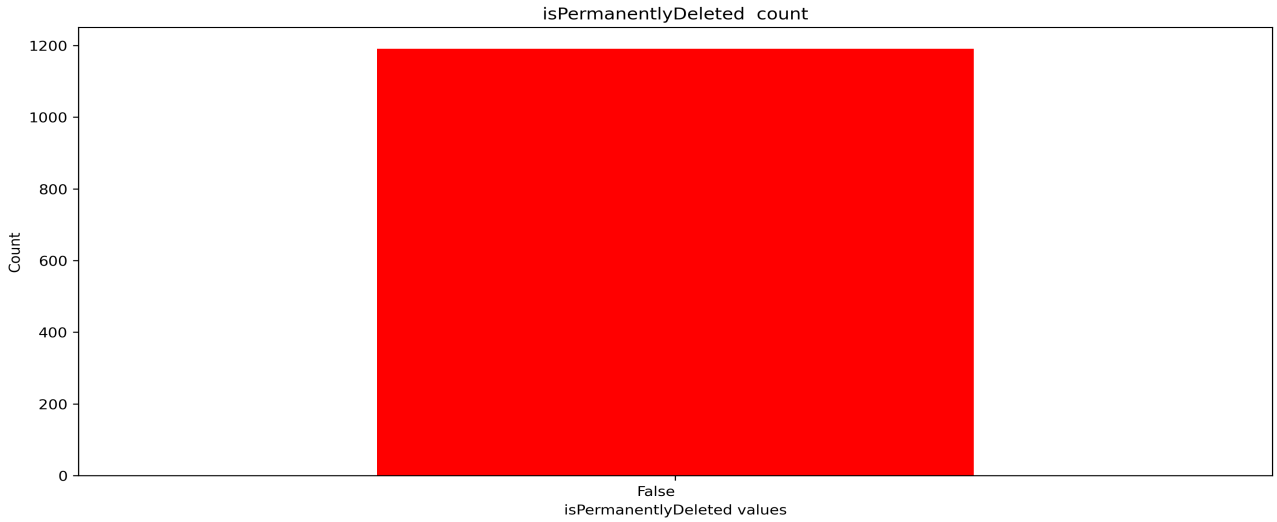
here we have the -isDone- value counts.



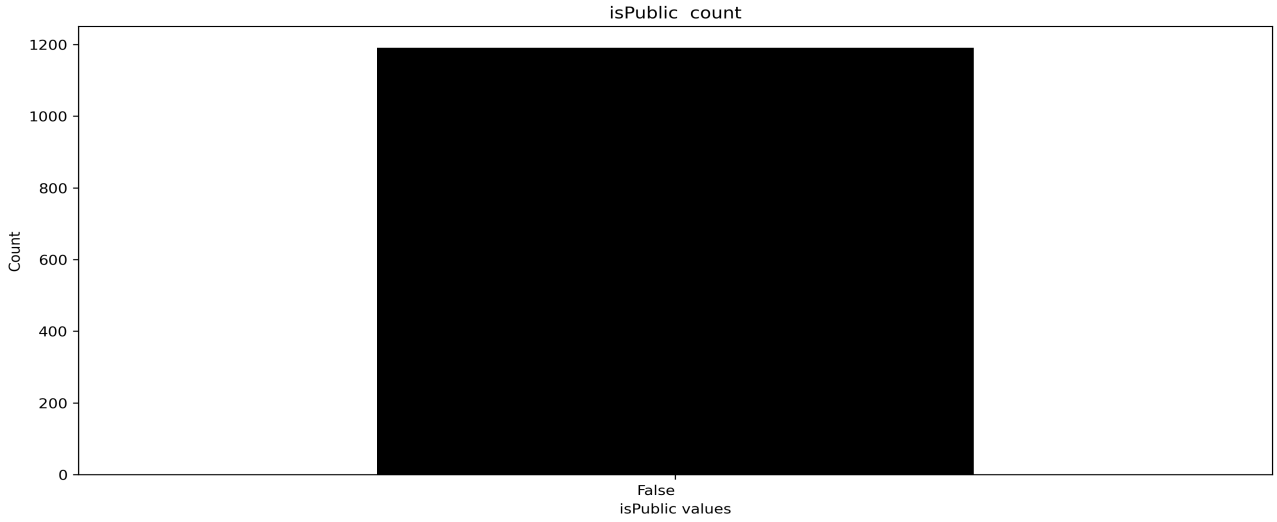
this plot shows us the variable -isFirstTime- values in archive.



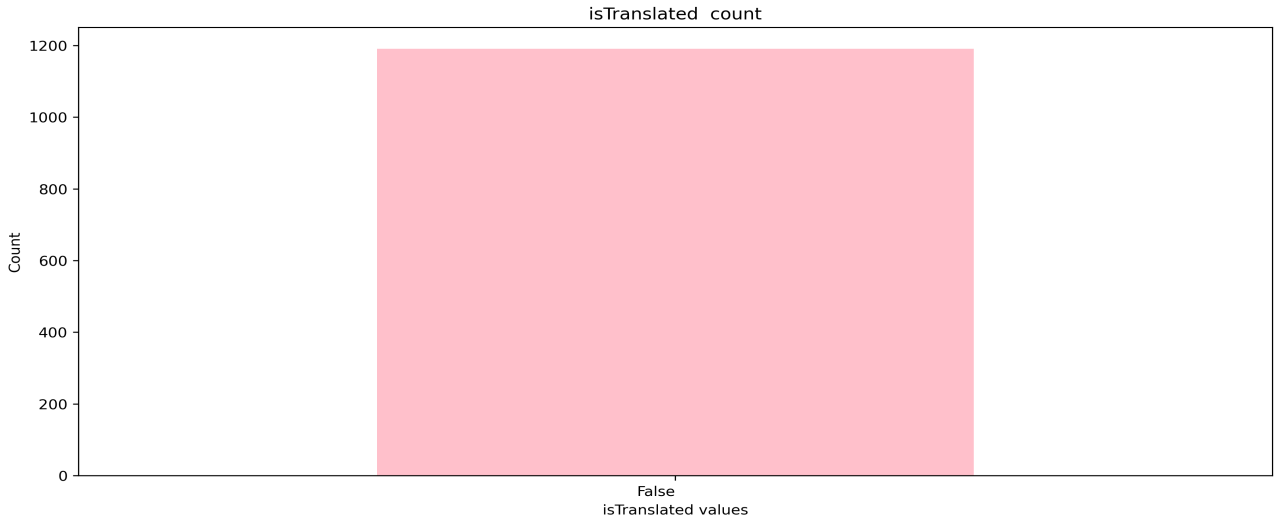
here we have the values of -islibrary- .



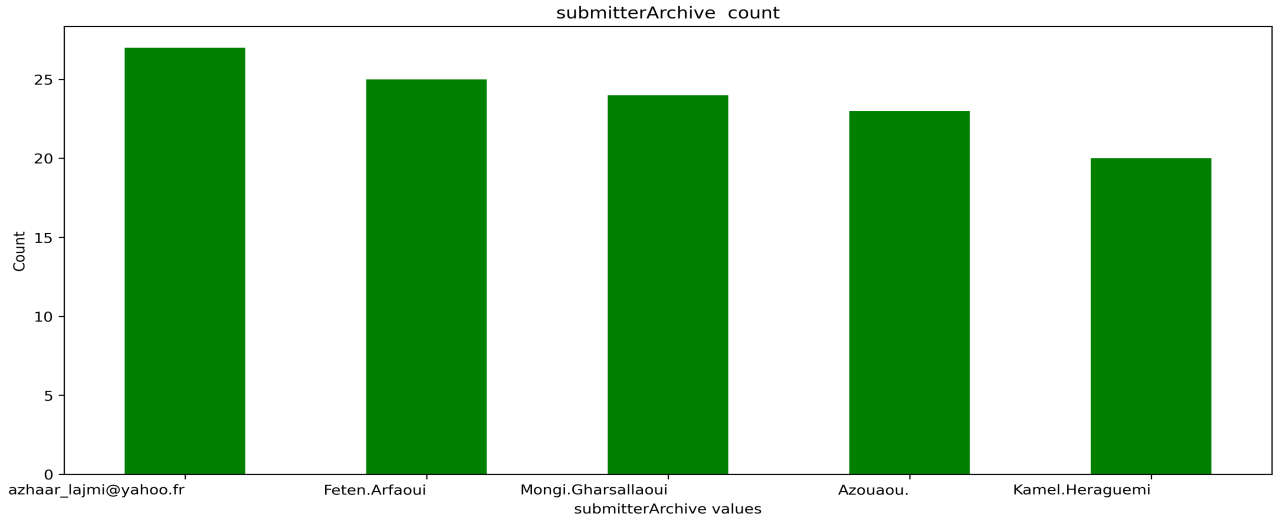
through this figure we can the values of -isPermanentlyDeleted- .



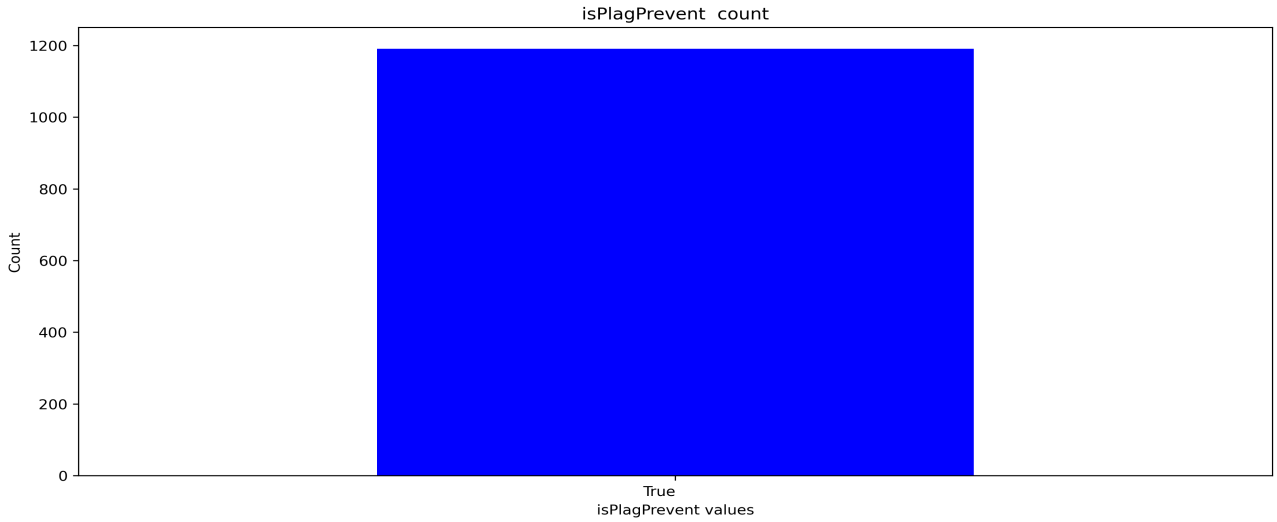
and here we have the values of -is public- .



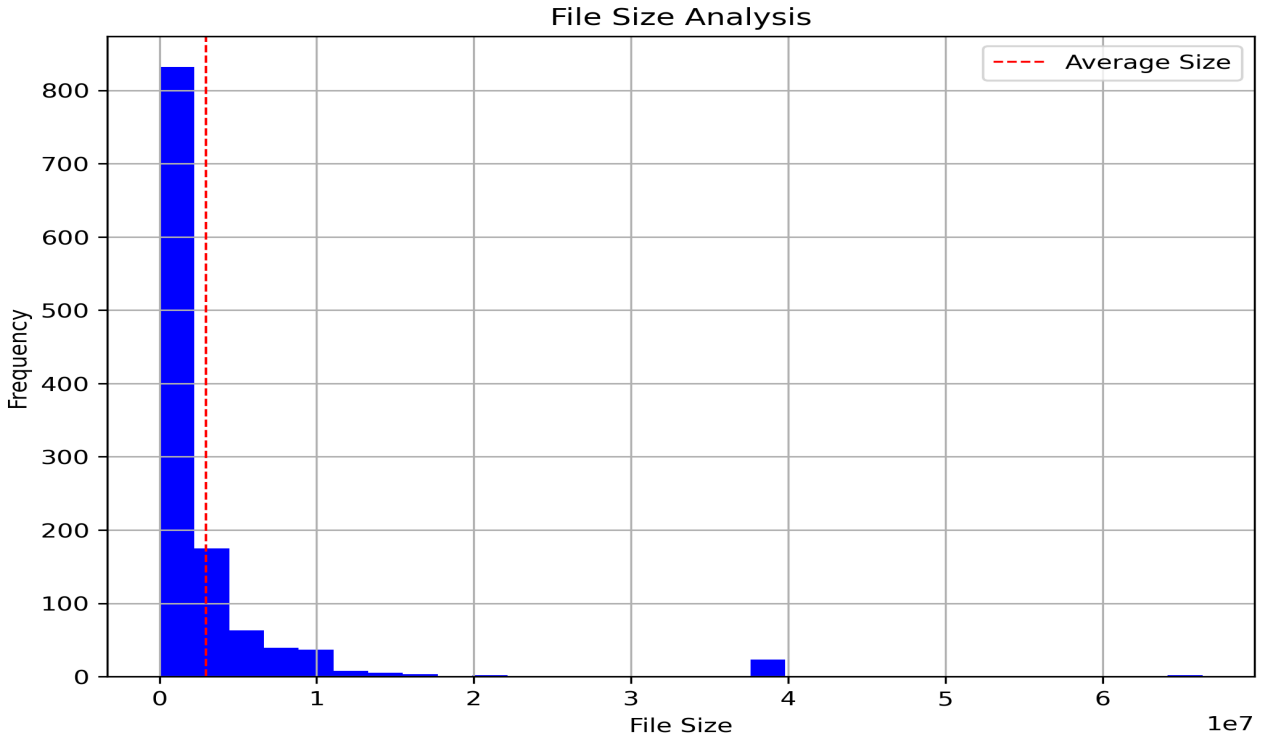
those are the values of -isTranslated- .



and this plot shows us the top submitters in the archive.



this figure presents the values of -isPlagPrevent- variable.



finally we have this histogramme that shows the average value of the variable -size- in archive.