Analysis of Open Source Simulation Applications

Balaji Janamanchi
Associate Professor of Management
Division of International Business and Technology Studies
A.R. Sanchez Jr. School of Business, Texas A & M International University

ABSTRACT

This study seeks to understanding the growth and popularity of Open Source Software (OSS) development for simulation modeling under Science and Engineering domain on SourceForge online service platform. The objectives include understanding the importance, if any, of the features such as, license type, programming language, Operating System version, review rating in influencing the popularity of a given software application as measured by downloads. For obvious reasons, end-users select and adopt specific applications based on their usage needs and the suitability of a given application in meeting their needs. Nevertheless, some applications may be easily adoptable than the other given the different options available for customization as well as scalability opportunities present in an application. So, other factors remaining common, popularity gained by an application with the increased downloads tends to draw prospective users to a specific application more than the others. This study analyses the availability and accessibility of simulation modeling software on the Source Forge website, undeniably, a very large repository of open source software resources. Limiting the search to "Simulations" within "Science and Engineering" domain /category. The top 250 applications' data is downloaded and analyzed. Useful insights and metrics are developed and compiled to assist end-users to understand the profile of open source simulation applications available for users for free downloads.

Keywords: Open Source Software, Science and Engineering, Simulation Modeling, Open Source Licensing.

INTRODUCTION

What is Open Source?

The term "open source" refers to something people can modify and share because its design is publicly accessible [1]. Essentially, the term open source denotes the accessibility of 'source code" underlying any application software that is made available to the end-users so that they modify the application software to suit their usage needs under rather unrestrictive set of terms of usage license [1].

As may be logically inferred, users are able to modify, extend, alter, amend and improve the open source software, as they need or desire, to suit their specific usage requirements. Contrast this with the most common proprietary software that comes already installed or users purchase and install on their computers for their use; proprietary software providers do not allow access to the source course of the software to the end-users. A very limited to no option is given to end-users to customize functionality and/or fine-tune the features of proprietary software applications. This, obviously, leads to a dependence on proprietary software sellers for almost all issues with the use and implementation of applications, whereas the users of open source software applications are free from such dependence on developers.

Another interesting fact about open source software is that, most open source software is available "free of cost" to interested end-users as long as they are willing to abide by the usage license. This means, all end-users big and small can acquire it at practically free of cost and

possibly, benefit from using the software application for their business needs. Typically, small and medium enterprise that are more often are cash strapped than not will be able to afford the open source software as long as they are willing to invest enough time and effort in getting to know how best to use it to meet their decision support requirements.

What is Source Forge?

"SourceForge is a web-based service that offers software developers a centralized online location to control and manage free and open-source software projects. It provides a source code repository, bug tracking, mirroring of downloads for load balancing, a wiki for documentation, developer and user mailing lists, user-support forums, user-written reviews and ratings, a news bulletin, micro-blog for publishing project updates, and other features." [2].

"SourceForge offers free access to hosting and tools for developers of free / open-source software. As of March 2014, the SourceForge repository claimed to host more than 430,000 projects and had more than 3.7 million registered users. The domain sourceforge.net attracted at least 33 million visitors by August 2009 according to a Compete.com survey." [2]

From the above it may be noted that SourceForge is a common shared forum where independent software developers can upload and host software applications they created for others to see and appreciate and possibly download and use, improve share further with others. This may be likened to artifact emerging from a loosely put together collaborative project effort by team members that may not ever meet or know each other, of the team. Currently, SourceForge is not the top ranked forum for open source software [3]. Nevertheless, SourceForge is the first such forums to provide a centralized location to host opensource software for developers and users free of cost some twenty years in November of 1999. As per table 1 below, based on Alexa rankings SourceForge ranks #2 among the source code hosting facilities, second to GitHub [3].

Table 1: Source Code Hosting facility by Popularity [3]

Name	Users	Projects	Alexa rank (lower = more popular)
GitHub	31,000,000[50]	100,000,000[50]	65 as of 9 September 2019[51]
SourceForge	3,700,000[67]	500,000[67]	407 as of 9 September 2019[68]
Bitbucket	5,000,000[45]	Unknown	989 as of 9 September 2019[46]
GitLab	100,000[52]	546,000[53][k]	2,146 as of 9 September 2019[54]
OSDN	54,826[60]	6,294[60]	8,529 as of 9 September 2019[61]
Launchpad	3,965,288[57]	40,881[58]	12,344 as of 9 September 2019[59]
Assembla	Unknown	526,581+[43]	23,052 as of 9 September 2019[44]
Rosetta code	Unknown	Unknown	62,045 as of 9 September 2019[65]
Buddy	Unknown	Unknown	73,518 as of 9 September 2019[47]
GNU Savannah	93,346[55]	3,848[55]	100,244 as of 9 September 2019[56]
Gitea	Unknown	Unknown	209,697 as of 9 September 2019[49]
CloudForge	Unknown	Unknown	339,271 as of 9 September 2019[48]
OW2 Consortium	Unknown	Unknown	610,052 as of 9 September 2019[64]
Ourproject.org	6,353[62]	1,846[62]	1,191,954 as of 9 September 2019[63]
SEUL	Unknown	Unknown	1,268,571 as of 9 September 2019[66]
Name	Users	Projects	Alexa rank (lower = more popular)

What is Simulation?

"Simulation refers to a broad collection of methods and applications to mimic the behavior of real systems, usually on a computer with appropriate software" [4]. In fact, simulation is used and applied in many fields and contexts to get a better understanding of complex behavior of systems in order to manage them better. Typically, simulation is useful in situations or context where an analytical solution does not exist due to complexity of the behavior system variables or other attendant features. The author has used and been using a type of simulation modeling in researching Supply Chain Management [5] and Project Management [6] related issues.

Simulation modeling is creation of computer models of some activity, process or phenomena for the purpose of study by manipulating it logically to understand how the real physical world works. Simulation helps to answer questions that are otherwise to be learned by experimenting with real systems and in many situations, conducting experiments on real system is impossible, impractical, or prohibitively expensive. For example: consider using flight simulators instead of flying real aircraft at the beginning stages of a pilot's training.

Who, when, where, and why of Simulation.

Any standard simulation book provides in its introduction section or the first chapter answers to the basic questions of who uses simulation. When is simulation useful? Alternatively, what context is simulation most useful? Why to use simulation? What are the limitations of simulation? Suffice it to recount here that simulation is used in a wide variety of fields and industries. It is most useful where analytical solutions do not exist. A curious reader wanting know more details about simulations and its use, is well advised to refer to a fairly effective contribution to the Winder Simulation Conference 2004 by Carson [7]. As is well known, all managers, decision makers need as much decision support as they can get when dealing with complex business situations. Whoever needs to make effective decisions need decision support tools in making better decisions will most likely at some point in time to simulation for help and guidance.

CURRENT STUDY- PURPOSE AND OBJECTIVE

As stated earlier most business enterprises as well as individual users can benefit from using simulation models, tools for decision support in their daily routines. However, lack of suitable and affordable tools/options and the financial costs in many cases become a big hurdle. Not much attention is given in academic literature to the availability of and characteristics of open source Simulation Software (OSSS). Further, this study also seeks to understand the user activity on the SourceForge in respect of the simulation modeling to learn about success of OSS simulation modeling applications. This study explores these issues with an analysis of Simulation Software under Open source projects on Source-Forge, a large repository of open source projects. The idea is to assess the available cost-free resources for simulation that may be accessible and useful to small and medium business organizations that may not be able to afford expensive proprietary software alternatives.

DATA COLLECTION PROCESS

Data Search on Sourceforege

The data was gathered using the open source software website www.sourceforge.net. Collection of the data started on September 28 and finished on October 2, 2016. By using the browse option on the website, it was possible to limit the search to "Science and Engineering" domain and by looking up the available options by typing in the key words "Simulation Software" in the search field. A whopping 101 pages of results showed up with 25 items per page totaling a little over 2,500 different item of software. We then limited our analysis to the top 250 results from these. There are different options to sort the results and we used the most

intuitive option of sorting by popularity. The analysis and discussions presented herein under the sections to follow is based on this shortlist of 250 software applications.

Metrics recorded and date of collected

For each of these 250 Simulation software in the Science and Engineering field, the following data was collected: Software Name, Registered Date, Last Updated Date, Intended Audience, Programming Language, Languages Available, OS Versions Available, Review & Rating, Downloads, Top Downloaded Country, Top Downloaded OS, and License Type (subsequently) totaling 12 variables each. Any missing data was highlighted in the spreadsheet, but was not imputed. Instead, the analyses are performed based on the available data under the respective fields.

ANALYSIS AND FINDINGS

As stated under the data collection section, for each of the 250 Simulation Software Applications found under "Science and Engineering" classification, details of twelve attributes were collected. The idea in gathering these items of data about each of the application is to provide the potential end-users a reasonably good understanding of the profile of successful OS Simulation applications so that they may choose the right application for their purposes.

Figure 1: Histogram of Intended Audience
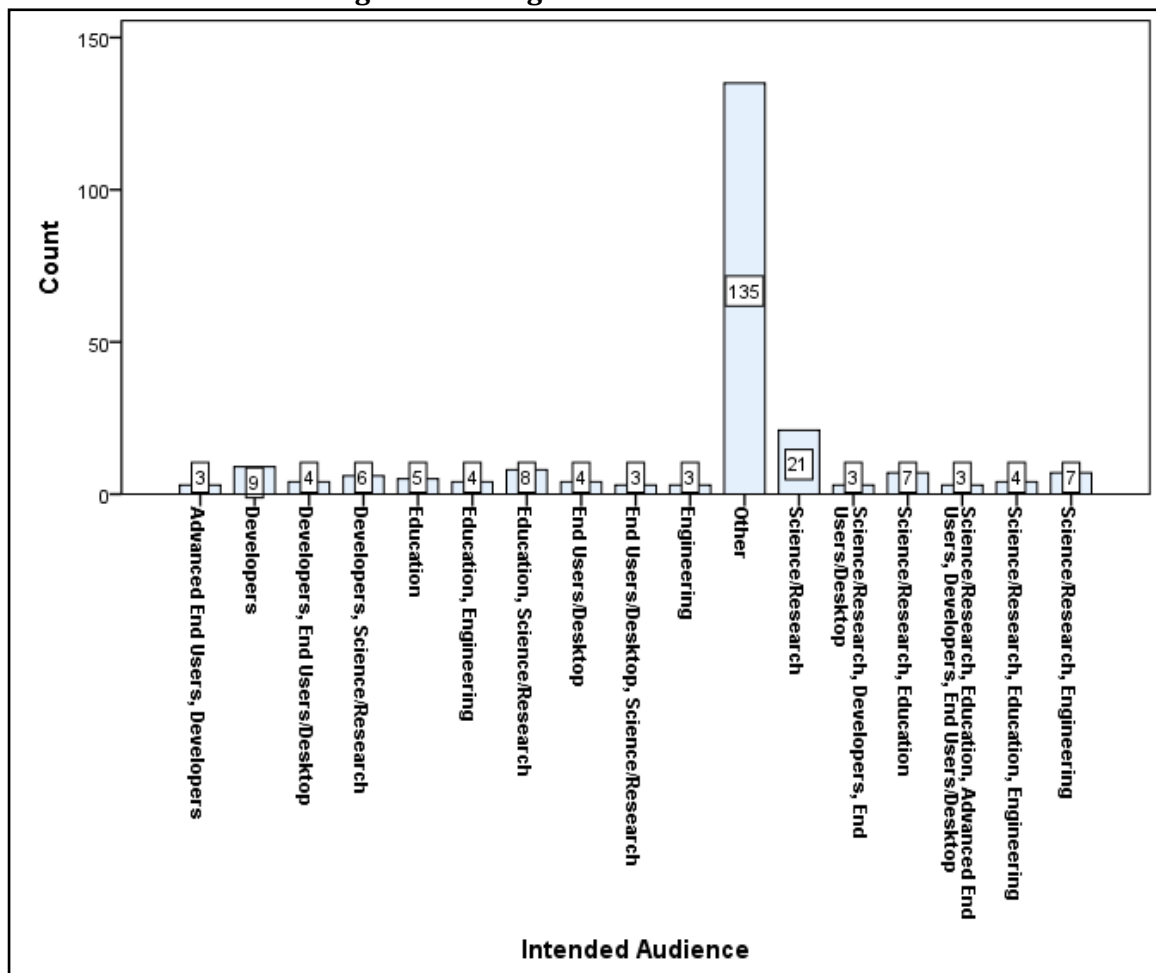
In developing the above chart, only those categories that had > 1% out of 250 observations were captured distinctly. All other designations of intended audience with 2 or 1 (being <1% in 250) were combined under the broad category of "other" which turned out to be largest group

with 135 out of 250 observations. This is very interesting in that, these applications appear to target very specific and distinct intended audience from each other.

Amongst the distinctly captured groups of intended audience, "scientific research" is the single largest. Further, scientific research in combination with other categories of audience accounts for 62 observations, which is to be expected since the observations were selected from the "Science and Engineering" domain.

The other noteworthy point in the above histogram is the group of "Developers" and combinations containing developers. On careful review, one might conclude that this is not surprising given that most developers would expect other developers to extend their applications by taking from where they left off and build the application to the next level of user-friendliness.

Figure 2: Histogram of projects by the OS version available
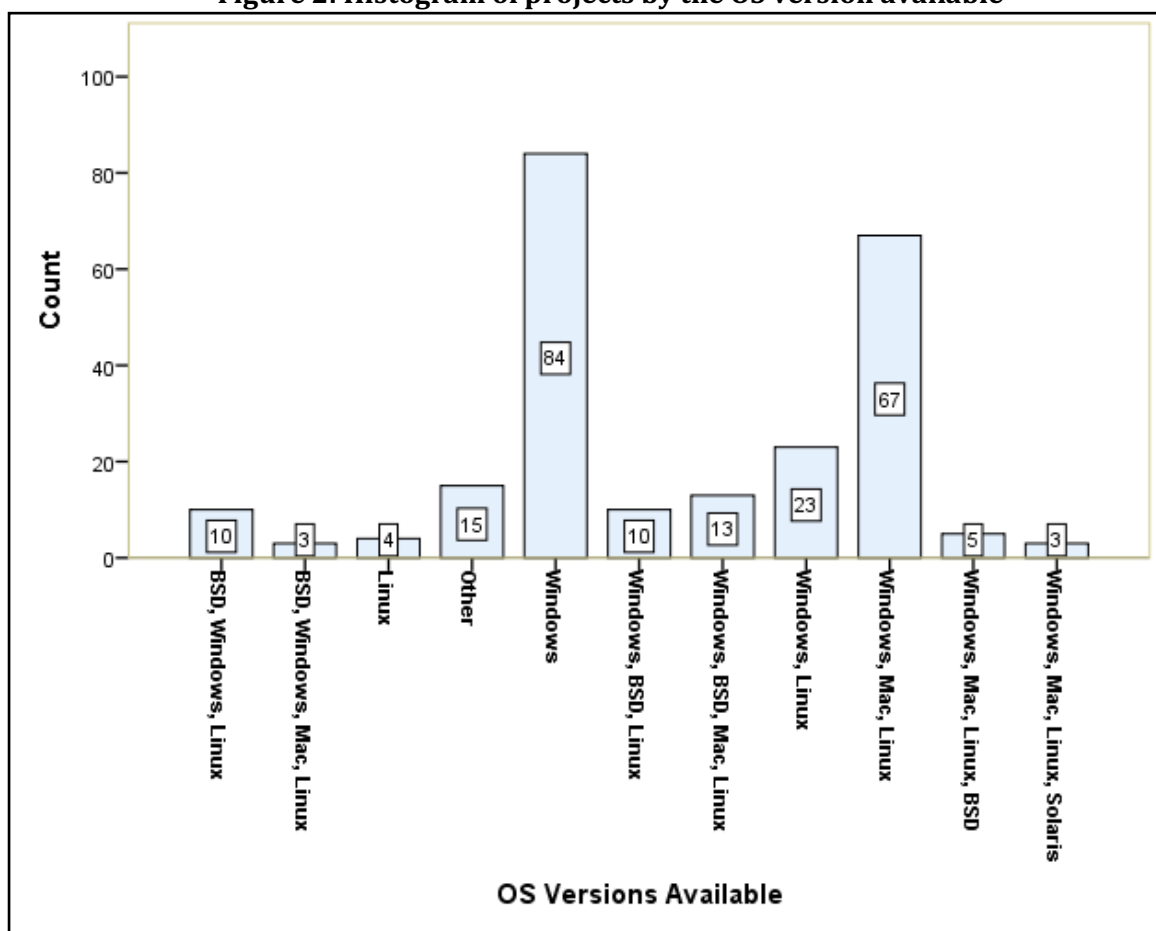
As shown in Figure 2, Windows and Windows combined with other OS appear to dominate the field, followed by Linux and Mac. It's interesting to note that while applications are open source, the preferred OS for developing the applications is proprietary Windows (by Microsoft Corpn) rather than Linux or other open source Operating systems.

Generally speaking Open Source software license categories form a spectrum ranging from highly restrictive e.g. GNU AGPLv3 through moderately restrictive e.g. GNU LGPLv3, moderately permissive e.g. Apache License 2.0, highly permissive e.g. MIT License, and ending with Unlicense e.g. public domain [8]. Not surprisingly, restrictive licenses tend to attract

developers more than end-users because restrictive licenses place restrictions and obligations on the user of the respective application that only seasoned developers can navigate. On the other hand, highly permissive and unlicense types will be of great interest to end-users more than developers for obvious reasons.

Detailed explanation for various license is available at the Open Source Initiative website [9]. Figure 3 below shows the distribution of license types among the Open Source Simulation applications. As may be seen from it, GNU GPLv2 is the most popular license type. There is a widespread of license types other than GNU GPL v2 and its variations.

**Figure 3: Count of project by their license type**
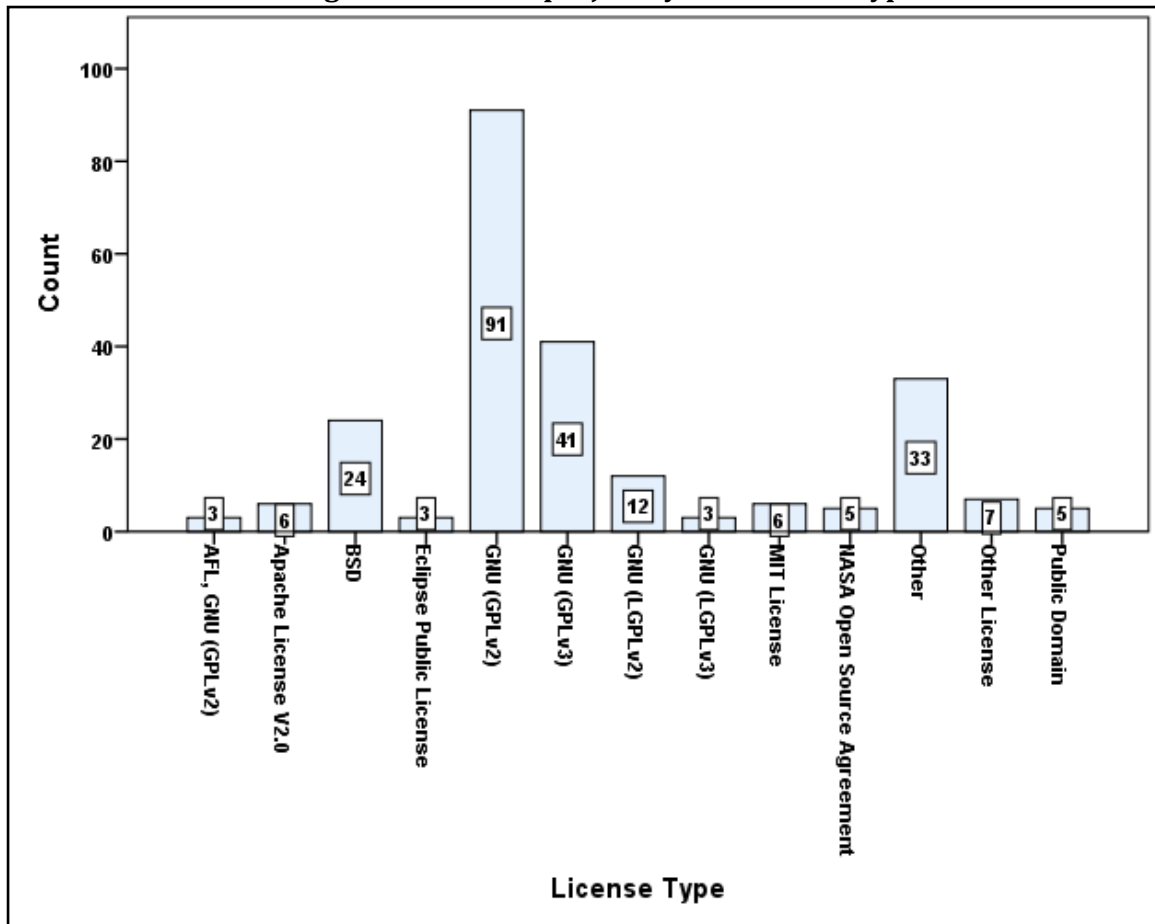
As far as the preferred programming language is concerned, Figure 4 below shows Java to be the single most popular with 65 out of 250 accounting for slightly more than 25% of the top 250 applications. However, C++ and its combinations combined add up to 71 out of 250. Python appears to emerge as significant third with 26 application of the 250 using it as their preferred language for application development. Given that the domain is Science and Engineering, MATLAB also found fair following in OS Simulation application development.

Figure 4: Programming Language
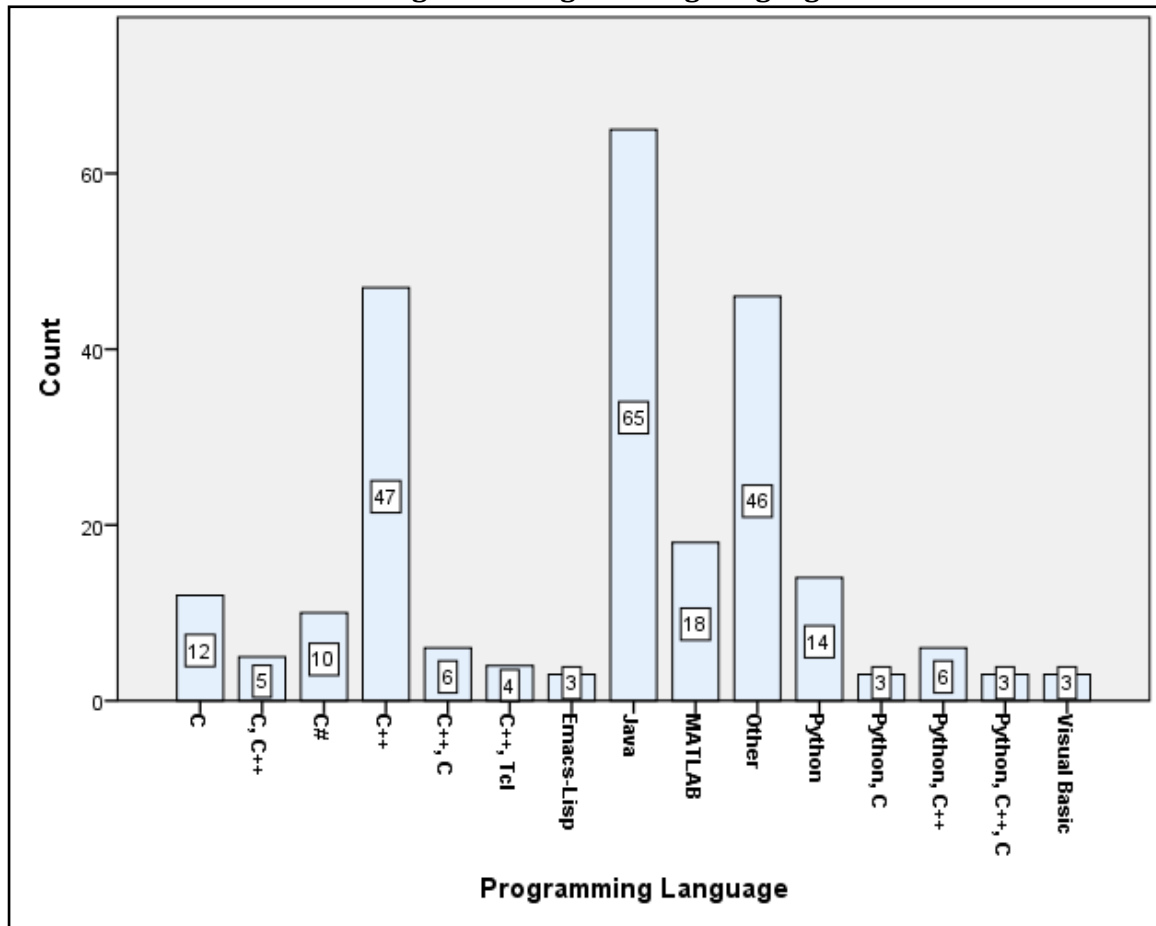
Review and rating of an application by the users is a reasonably good measure of the usefulness and ease-of-use features of the application. Figure 5 depicts the review ratings of the top 250 applications under the study. As may be noted, 60% of the applications received a rating of three or higher out of a possible five score. Nearly 42% of the 250 applications received 5 out of 5 perfect score from their users. This is a very encouraging sign in that not only the software applications are available under open source initiative (presumably, free-of-cost) but are of top quality applications as rated by the discerning end-users.

However, one point of concern could be that as many as 94 applications have not been reviewed or rated among the top 250. It is not clear if they were not rated because the users didn't like them but didn't want to leave negative comments/ or poor ratings. Alternatively, that the software application is in trial version and not yet open for reviews and rating. In either case, this could potentially be a concern given that these 94 applications already found their place in the top 250-simulation software on the forum without yet receiving ratings from the end-users.

Figure 5: Count of projects by their review ratings
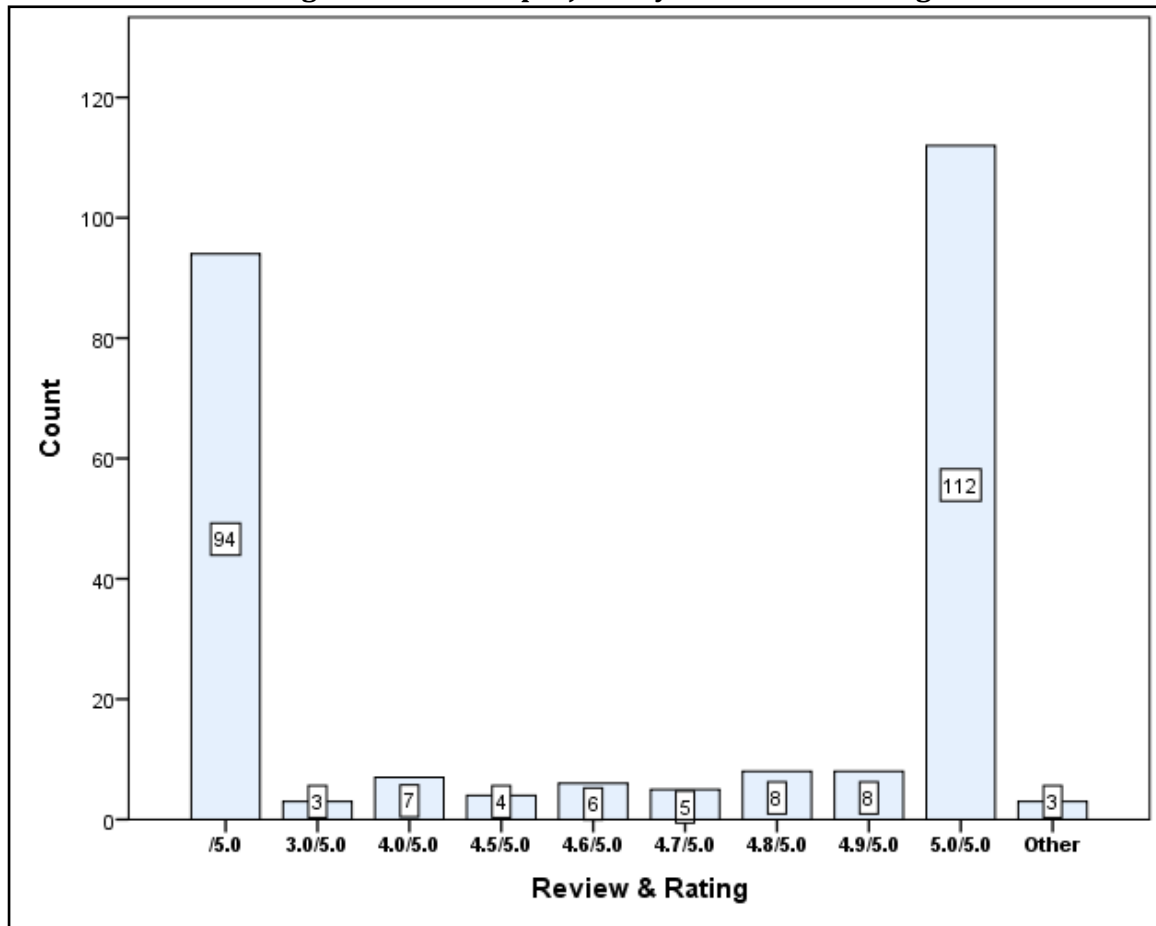
Although downloads metric doesn't always guarantee the usage of downloaded software, in the context of an OSS community downloads could be an effective metric for the popularity of an application in that downloads evidences interest in the applications by the target audience. So downloads metric is arguably a good proxy to popularity of an application or that of a programming language used or license type etc. We analyzed the data in respect of downloads in multiple perspectives and results are depicted as cross tabulations and discussed below.

Figure 6: Count of projects by the top downloaded country
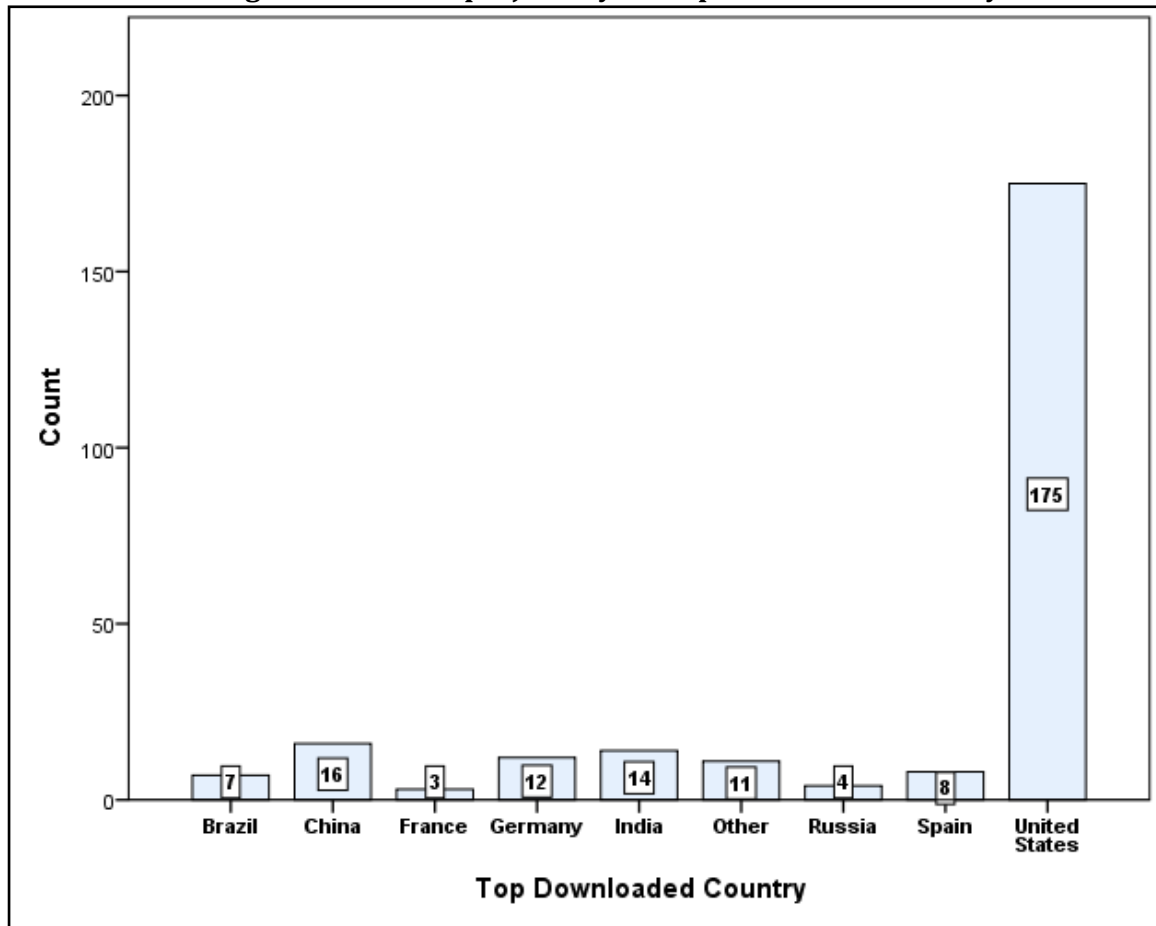
We begin with the presentation of top downloaded application count by country in Figure 6 above. As may be seen from Figure 6, the Open source software usage and awareness appears to be widespread in the US more than any other country in the world. As a matter of fact the total number of applications downloaded the most (top downloaded) in US at 175 out of 250 is more than twice as many as all other countries combined, which is only 75 out of 250.

Noteworthy from the figure 6 is the fact that China and India are #2 and #3 in downloading 16 and 14 distinct simulation applications the most number of times compared to any other country in the world having access to these Open source simulation software applications. Given that this chart is for the count of "most downloaded country" one cannot read too much into it without overlooking other important aspects of open source software usage.

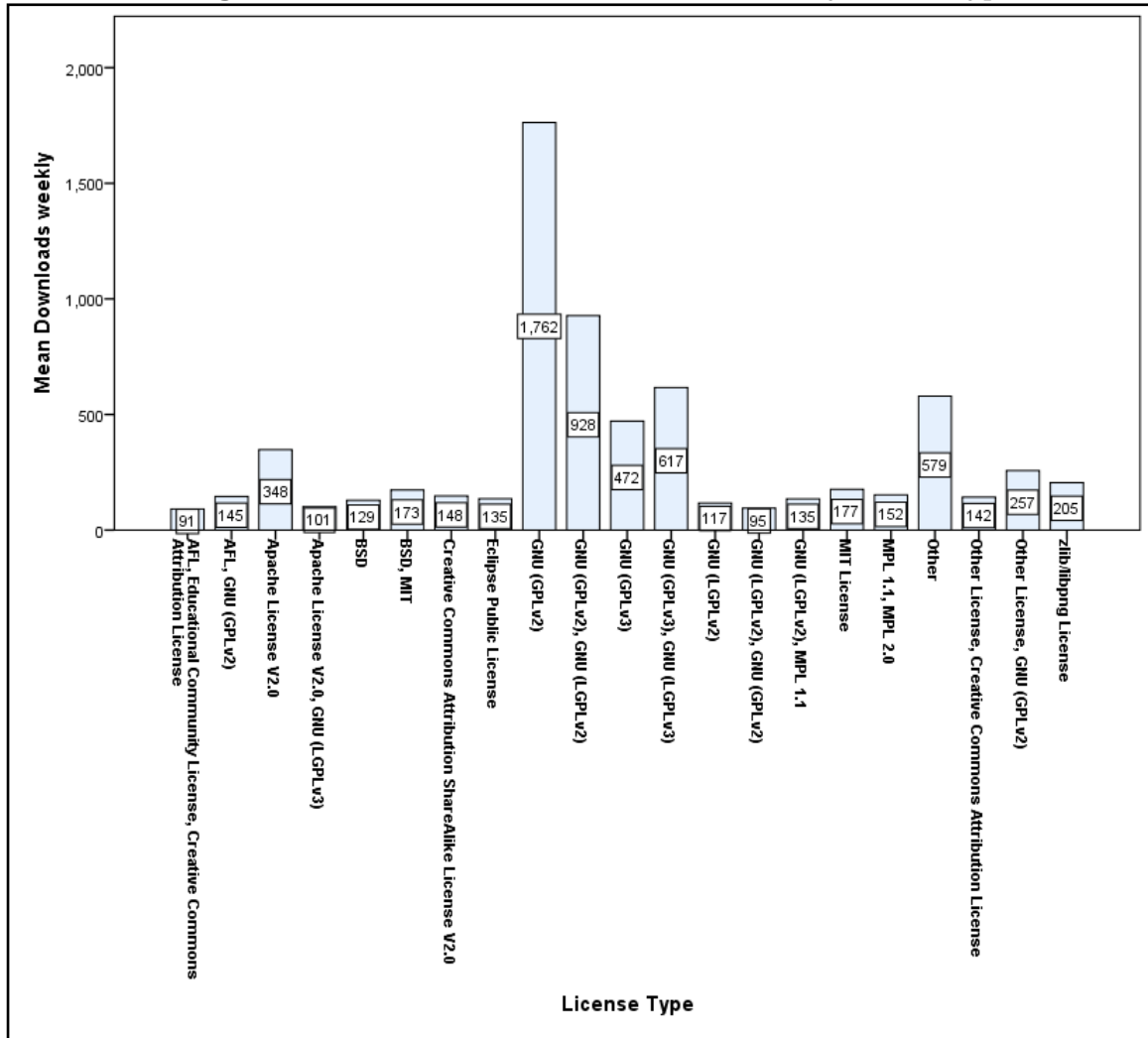**Figure 7: Cross tabulation of Mean Downloads by License type**

Figure 7 captures the mean weekly downloads by the license type. By far the downloads of software provided under restrictive license types such as GNU GPL v2 and its variations appear to dominate the scenario. However, aside of that there is a reasonable distribution among moderately and highly permissive license types. This is a little counterintuitive in that it is commonly believed that only programmers and advanced end-users prefer the restrictive license than a novice end-user or a person not very adept at modifying source code files. So one would expect more downloads of unrestrictive and highly permissive license categories of application than the restrictive ones. Therefore, this phenomenon needing further exploration to understand the underlying dynamics.

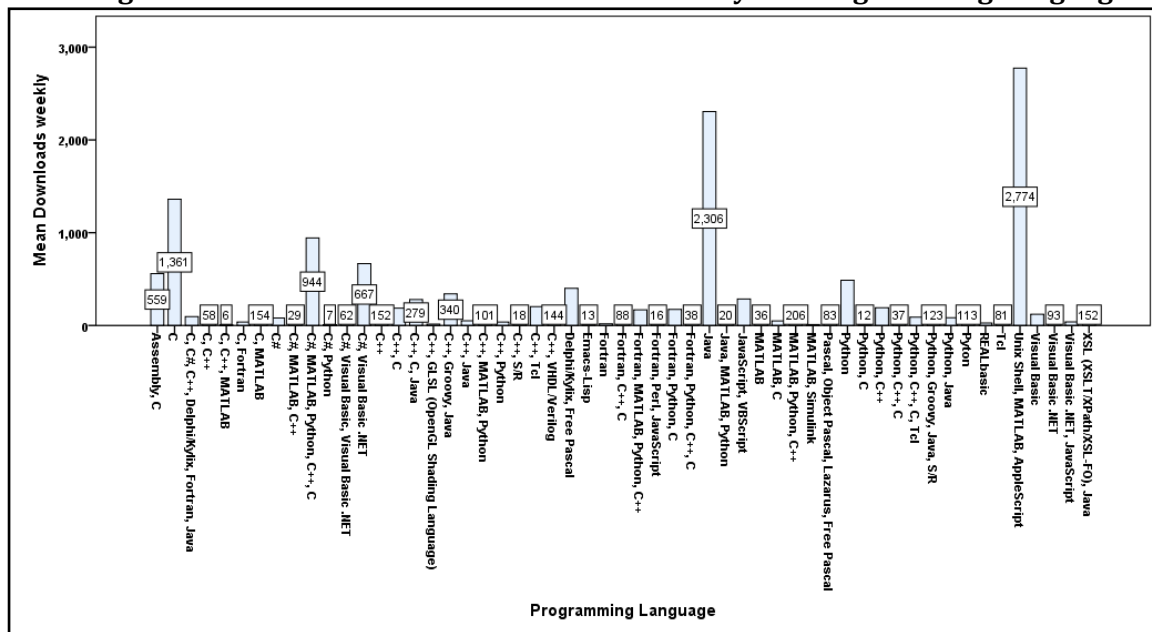**Figure 8: Cross tabulation of Mean Downloads by the Programming Language**

Figure 8 above clearly establishes that Java is the single most popular software for developing these Open Source Simulation software by leading the average weekly. The other bar that is bigger than java is, actually, a group of four different software. These four seem to combine well in that the developers decide to release their application in all four at the same time. Further programming language C and C++, C# and combinations seem to account for sizeable downloads as well.

While the use of programming language by the count of projects may indicate the developers' preference for a language, the weekly downloads tabulated by the programming languages may indicate the end-user preference for a specific software. It may also indicate the feature richness of the specific software as well as the support for on online forums and other aspects.

Over a period, the developers may be influenced by the end user's choice of software and decide to switch to those rather than to stick to their personal preferences. Or a new section of developers may emerge that will convert existing Open source applications from less popular software languages into applications developed in more popular software. Given that some big name corporations from IT industry are actively sponsoring new projects in Open Source platforms, we may see more and more of the latter of above said two outcomes.

In figure 9 depicts the top 10 project by the weekly downloads in October 2017. As may the seen, the top most software's downloads are way more than the weekly downloads of the rest of the nine software downloads combined.

In order to provide the readers a more recent state of the Open source software in simulation, figure 10 depicts the top 10 projects by the weekly downloads in December 2019. While the popularity of the top of the list appears to wane a little bit (down from 131,283 to 87,670) it nevertheless continues to dominate this segment of open source domain. Seven of the top 10 from October 2017 continue to be in top 10 of December 2019, some with slight change of ranks while there are three new names at # 3, #7, and #10 in the top 10 of December 2019.

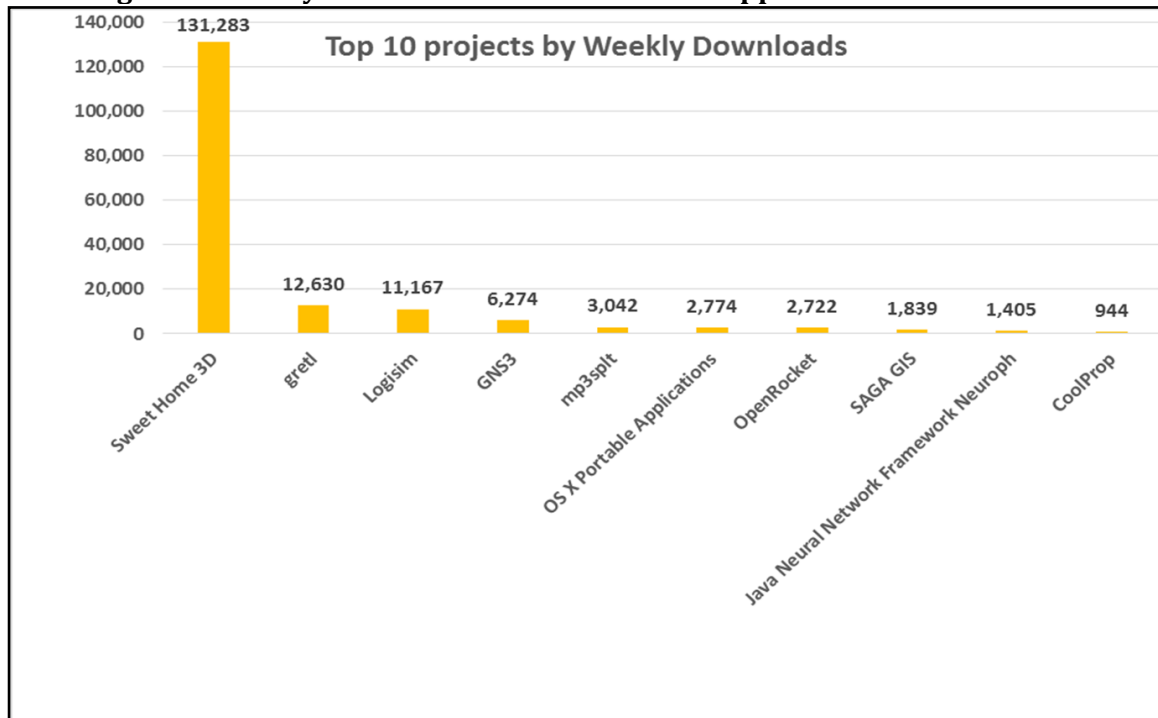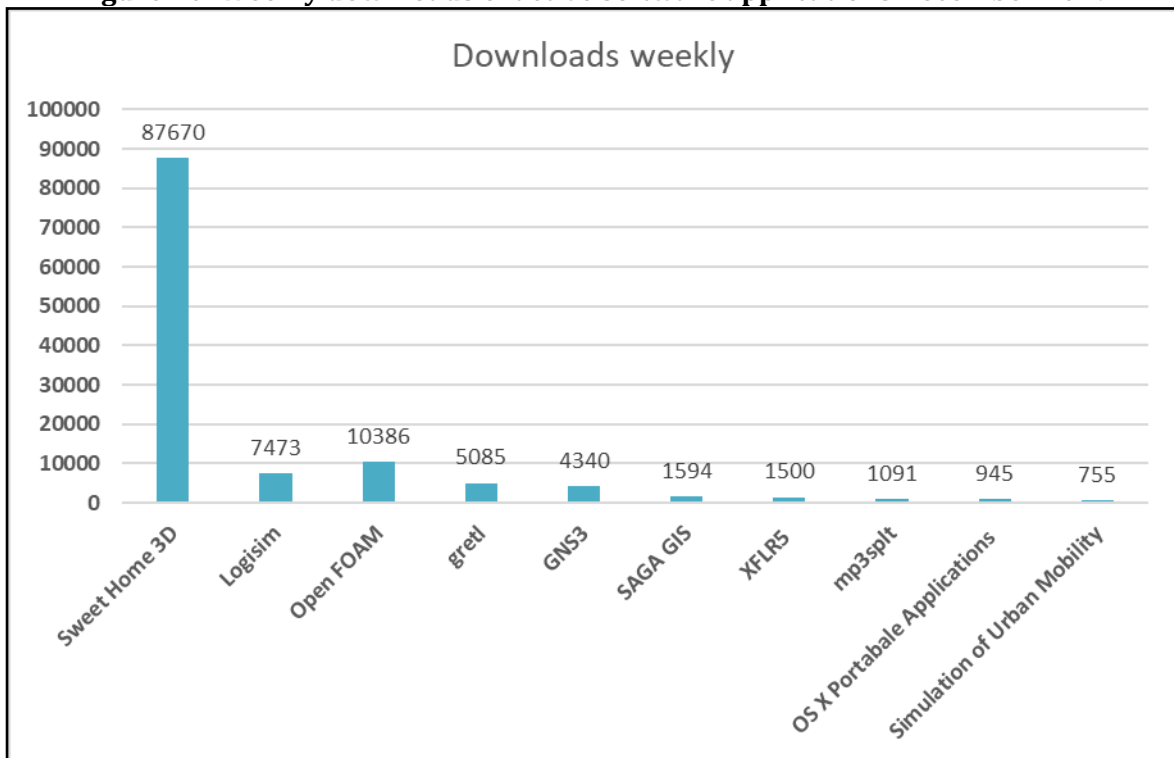**Figure 9: Weekly downloads of active software applications October 2017**

Figure 10: Weekly downloads of active software applications December 2019

The Top Ten Applications

Sweet Home 3D (GPLv2): Sweet Home 3D is an interior design application that helps you to quickly draw the floor plan of your house, arrange furniture on it, and visit the results in 3D. With the help of Sweet Home 3D users can draw walls and rooms upon the image of an existing plan, on one or more levels, drag and drop doors, windows and furniture from a catalog onto the plan. This application is available in over 25 languages viz. Korean, French, Ukrainian, Dutch, Polish, Slovene, Czech, Finnish, Italian, Greek, Vietnamese, English, Portuguese, Serbian, Chinese (Traditional), Bulgarian, Swedish, Turkish, Brazilian Portuguese, Chinese (Simplified),

German, Japanese, Spanish, Russian, and Hungarian. This might explain the immense popularity of this application in contrast to other items in this top 10 list because most others are available in fewer languages with some software limited to just a language or two at most. The above details and more about this application are given on its webpage in SourceForge pages [10].

gretl (GPLv3): gretl is a cross-platform software package for econometric analysis, written in the C programming language. It is claimed to be easy and intuitive GUI based. Offers sophisticated features of OLS, MLE, GMM and system estimation. Further it is internationalized with offering in 10 languages besides English. The above details and more about gretl are given on its webpage in SourceForge pages [11].

Logisim (GPLv2): Logisim, is an educational tool for designing and simulating digital logic circuits. Featuring simple-to-learn interface, hierarchical circuits, wire bundles, and a large component library, as a Java application, it can run on many platforms. However, with the last update recorded as 4/29/2013, the project appears to be abandoned by the developers and could be categorized as 'abandonware!' Surprisingly, Logisim find a place in the December 2019 list of top 10 as well. The above details and more about Logisim are given on its webpage in SourceForge pages [12].

GNS3 (GNU GPLv2): GNS3 is a graphical network simulator that allows you to design complex network topologies. You may run simulations or configure devices ranging from simple workstations to powerful Cisco routers. It is based on Dynamips, Pemu/Qemu and Dynagen. It's available in 18 different world laguages that contributes its popularity. The above details and more about Logisim are given on its webpage in SourceForge pages [13].

mp3splt (GNU GPLv2): mp3splt is a free utility to split mp3, ogg vorbis and FLAC files without decoding, selecting begin/end time. It also splits Mp3Wrap and AlbumWrap archives. Supports splitting and trimming by silence detection. It's available in five different world languages. The above details and more about mp3splt are given on its webpage in SourceForge pages [14].

OS X Portable Applications (GPLv2): OS X FOSS portable applications are packaged so one can carry around on any portable device, USB thumb drive, iPod, portable hard drive, memory card, other portable device (or also on one's internal hard disk), taking their preferences with them. All these details and more about OSX Portable Applications are given on its webpage in SourceForge pages [15].

OpenRocket (GPLv3): An Open Source model rocket simulator. Claims to provide, everything you need to design, simulate and fly better rockets. More of hobby project but available in five different world languages contributing to its success all over the world. All these details and more about OSX Portable Applications are given on its webpage in SourceForge pages [16].

SAGA GIS (GPLv2):
SAGA - System for Automated Geoscientific Analyses - is a Geographic Information System (GIS) software with immense capabilities for geodata processing and analysis. SAGA is programmed in the object oriented C++ language and supports the implementation of new functions with a very effective Application Programming Interface (API). Functions are organised as modules in framework independent Module Libraries and can be accessed via SAGA's Graphical User Interface (GUI) or various scripting environments (shell scripts, Python, R ...). [17].

Java Neural Network Framework Neuroph (Apache License v2.0): Neuroph is lightweight Java Neural Network Framework which can be used to develop common neural network architectures. Small number of basic classes which correspond to basic NN concepts, and GUI editor makes it easy to learn and use. All these details and more about OSX Portable Applications are given on its webpage in SourceForge pages [18].

CoolProp (MIT License): A cross-platform, open-source, alternative to NIST REFPROP. Based on reference-accuracy equations of state and transport property correlations for refrigerants like Water, CO_2, R134a, Nitrogen, Argon, Ammonia, Air, R404a, R410a, Propane and many others. All these details and more about OSX Portable Applications are given on its webpage in SourceForge pages [19].

LIMITATIONS AND CONCLUSIONS

Limitations
As stated in the introduction, this study is limited to Open Source Simulation Software from SorceForge repository alone. While Sourceforge is the oldest such repositories, it is currently second in Alexa ratings with GitHub occupying the first place. Further, this study is limited to simulation software in the Science and Engineering domain. It is quite likely the simulation software in other domains may rather different characteristics.

Conclusions:
Based on the charts created, descriptive statistics computed and the analyses performed from the data collected, the following important observations may be made.

- Eight of the top ten most downloaded Open Source Simulation software on Sourceforge downloaded in USA (3 to 10). Top project is most downloaded in Russia and second ranked project is most downloaded in Germany.
- Nine of the top ten are most downloaded on Windows OS versions except for one ranked 6th which is most downloaded in its MAC OS version.
- First Eight of top ten offer GNU (GPLv2 or GPl3) while the ninth offers Apache License V 2.0 and tenth MIT license
- Not surprisingly, most applications appear to target "End-users, Advanced end users, scientific research"
- Java, Java script, C and C++ appear to be most preferred programming language
- A very active OSS community in developing Simulation Modeling software with updates as recent as the week of data collection.
- Despite losing its place of pride as the top forum for OSS projects to GitHub, SourceForge continues to attract large number of developers and users
- Perhaps, users are accustomed to the SourceForge interface or they come for the old projects that are continued to be hosted here

Depending upon their respective objectives, readers may find these observations and conclusions useful for their future use of open source simulation software. The readers may also find these observations useful for planning their own participation as developers, sponsors, end-users or other stakeholders on SourceForge and other similar open source repositories.

References
[1]. Opensource 2019 at URL: https://opensource.org/faq#osd accessed on 12/1/2019

[2]. Wikipedia page on Sourceforge at URL: https://en.wikipedia.org/wiki/SourceForge accessed on 12/1/2019

[3]. Wikipedia page on source code hosting facilities at URL: https://en.wikipedia.org/wiki/Comparison_of_source-code-hosting_facilities accessed on 12/1/2019

[4]. Kelton W.D., Sadowski, R.P., and Sturrock, D.T., *Simulation with Arena,* 4th edition, 2007 McGraw Hill, New York, USA. pp1.

[5]. Janamanchi, B and Burns, J., "Control Theory Concepts Applied to Retail Supply Chain: A System Dynamics Modeling Environment Study" - Journal of Modeling and Simulation in Engineering- Volume 2013, Article ID 421350, 14 pages.

[6]. Burns, J., and Janamanchi, B. "The Dynamics of Change Management in a Technology Project Context," The International Journal of Information Systems and Change Management 2006, Volume 1, No.2, pp 115-137

[7]. Carson, J.S. "Introduction to Modeling and Simulation," Proceedings of the 2004 Winter Simulation Conference, at URL: https://dl.acm.org/citation.cfm?id=1161740 accessed on 12/1/2019.

[8]. Choosealicense.com at URL https://choosealicense.com/licenses/ accessed on 12/1/2019.

[9]. Opensource 2019 at URL: https://opensource.org/licenses/category accessed on 12/1/2019

[10]. Sweet Home 3D details on Sourceforge at URL: https://sourceforge.net/projects/sweethome3d/ accessed on 10/2/2017

[11]. gretl on Sourceforge at URL: https://sourceforge.net/projects/gretl/?source=directory accessed accessed on 10/2/2017

[12]. Logisim on Sourceforge at URL: https://sourceforge.net/projects/circuit/?source=directory accessed on 10/2/2017

[13]. GNS3 on Sourceforge at URL: https://sourceforge.net/projects/gns-3/?source=directory accessed on 10/2/2017

[14]. mp3splt on Sourceforge at URL: https://sourceforge.net/projects/mp3splt/?source=directory accessed on 10/2/2017

[15]. OSX Portable applications on Sourceforge at URL: https://sourceforge.net/projects/osxportableapps/?source=directory accessed on 10/2/2017

[16]. OpenRocket on Sourceforge at URL: https://sourceforge.net/projects/openrocket/?source=directory accessed on 10/2/2017

[17]. SAGA GIS on Sourceforge at URL: https://sourceforge.net/projects/saga-gis/?source=directory accessed on 10/2/2017

[18]. Java Neural Network Framework Neuroph on Sourceforge at URL: https://sourceforge.net/projects/neuroph/?source=directory accessed on 10/2/2017.

[19]. CoolProp on Sourceforge at URL: https://sourceforge.net/projects/phpcollab/?source=directory accessed on 10/2/2017.