

HackerRank Developer survey 15.06.2023

Aleksander Blok, Marcel Boxberger, Rafał Curyło, Wiktor Czetrybok, Marcel Dąbrowski

1. Analysis of the obtained results versus the goals

- to what degree the success criteria have been achieved

Identify trends and patterns

Geographic Distribution:

The majority of HackerRank users are from the United States and India.

Additionally, there is a notable presence of users from Canada, Russia, and various European countries.

These countries collectively form a significant user base that cannot be overlooked.

Age Distribution:

- The age of HackerRank users is predominantly concentrated in the 16-20 years old range.
- The number of users in this age group is more than twice the number in the 11-15 years old range.
- Users aged 21-25 years old follow, with nearly three times fewer users compared to the 16-20 age group.

Gender Distribution:

- In every age group, the number of male users significantly outweighs users of other genders.
- This trend is consistent across all age ranges, indicating a higher participation of males on the platform.

Educational Attainment:

- Most HackerRank users have completed college degrees.
- Users with postgraduate degrees follow closely behind, reflecting a commitment to higher education.
- Users who have completed some college and engaged in post-graduate work make up a smaller portion of the user base.
- High school graduates represent the smallest group among HackerRank users in terms of educational level.

Programming languages:

Most known programming languages are as follows

- Java
- C
- C++
- JavaScript
- Python
- PHP
- C#

- TypeScript
- Ruby
- R

Most liked programming languages to learn are as follow:

- Go
- Python
- Ruby
- Kotlin
- Swift
- R
- Javascript
- Typescript
- CSharp

Frameworks:

Most known frameworks are as follows

- NodeDotJS
- Angular
- Spring
- React
- ExpressJS
- Django
- ASP
- NetCore
- RubyOnRails
- JSF

Most liked frameworks to learn are as follow:

- React
- AngularJS
- NodeDotJS
- Djano
- React
- RubyOnRails
- VueJs
- Spring
- NetCore

- Validate findings -

- Can the results be used in practice?

This survey provided us data that certainly can be used in practice and further development of HackerRank Platform. There are ways of improving the platform.

1. **Targeted Marketing:** Understanding the geographic distribution of users allows for targeted marketing efforts in countries with a significant user base, such as the United States and India. It can help tailor marketing messages and campaigns to resonate with users from specific regions.
2. **User Engagement:** Recognizing the age distribution of users, particularly the dominance of younger individuals, can inform the development of engaging content,

challenges, and competitions that align with their interests and skill levels. This can enhance user engagement and encourage active participation on the platform.

3. **Content Creation:** The analysis of programming languages and frameworks that are popular can point the direction that the platform needs to expand. This ensures that the platform offers relevant and valuable resources for users of various user interests.
4. **Diversity and Inclusion:** Understanding the gender distribution highlights the predominance of male users. This insight can be used to develop initiatives aimed at increasing diversity and inclusion, such as targeted outreach programs, mentorship opportunities, or campaigns to encourage female participation in coding and programming challenges.
5. **Platform Expansion:** Recognizing the presence of users from countries like Canada, Russia, and various European countries highlights potential growth opportunities in these regions. The analysis can inform decisions related to platform expansion, localization efforts, and partnerships with educational institutions or organizations in those countries.

2. Review of the data mining process

- what have been done

Our project was done on the dataset containing 25000 responses from a 2016 survey conducted among developers worldwide. Goals of our research was to identify trends and patterns appearing in the development industry and understanding the developer community. EDA involves visually and statistically analyzing the dataset to understand its structure, distributions, and relationships between variables. Descriptive analytics focuses on summarizing and describing the dataset's key characteristics, such as central tendencies, variabilities, and distributions. Some commonly used algorithms for EDA and descriptive analytics include clustering algorithms , and visualization techniques (e.g., scatter plots, histograms). Evaluation methods can involve visual inspections of patterns and trends, statistical measures such as correlation coefficients, or comparing different subgroups or segments within the dataset.

- which steps gave useful results

Each distribution gave us insight into the developers industry. We can see that most HackerRank users are from the United States and India, the majority are male aged 16 to 20 with college degrees. Also we can check what programming languages are most known such as Java and c variation, which are most likable to learn: go and python , the same for frameworks most known are: NodeDotJS and Angular and which are most likable to learn React and AngularJS.

- what could have been possibly done better

Selection of analysis techniques: While the project mentions the use of clustering algorithms and visualization techniques, it would be beneficial to explore a wider range of analysis techniques to gain more comprehensive insights. Additional methods like regression analysis, decision trees, or association rule mining could have been employed to uncover hidden patterns and relationships within the dataset.

3. Discussion of importance of the results for further exploration

What additional data mining task could be performed for the dataset?

Additional and very useful data mining task that could be performed would be Time Series Analysis. We could find more datasets with information about programmers, their preferences and habits at different times. Of course it would require a lot of work to unify and merge different datasets but would give a lot of insight into knowledge about technology usage or job satisfaction changes over the years. There would be the possibility to uncover all types of trends, seasonality or patterns over time.

Results of which steps of the data mining process could be reused?

Data Cleaning and Preprocessing:

The initial data cleaning and preprocessing steps, such as handling missing values, removing duplicates, and formatting the data, are typically performed at the beginning of any data mining project. Once these steps are completed, the cleaned dataset can be reused for further analysis without repeating the cleaning process.

Data Visualization Techniques:

The visualization techniques applied in the previous data mining process, such as histograms, scatter plots, or heatmaps could be reused to present the results of further analyses to maintain consistency and similar graphical appearance.

Documentation and Reporting:

Similarly to the previous point, the documentation and reporting produced during further data mining process should be kept in similar form as the ones produced during previous process. This includes any notes, observations, or

insights captured during the process, which can help maintain consistency and track the progress of the overall analysis.