Can studying be a game? (Gamification of eQuiz)

Petar Mishov

I. Introduction

Our attention is an extremely precious resource. Despite this, most people do not make the most of it. What attracts our attention? Usually it is something that has some elements of gamification. For this reason we want to add some game design elements. We can embed them into activities, a website or a platform to make the whole process more attractive for use. Game elements can greatly benefit e-learning platforms such as eQuiz by improving motivation as well as learning. The 3 main questions in this paper are:

- "What are the main types of students that exist on eQuiz?"
- "Can we predict if a student is a certain type of person?"
- "How can we appeal to each type?"

In the past, some studies have shown a moderate negative correlation between a lack of stimulus from the application and the user's desire to keep learning, see Gafni, Achituv, Eidelman and Chatsky [4]. Most studies compare their non-gamified and gamified version of the application by doing a personality test on users then analyzing how these users react to the different versions, see Voida, Jia and Karanam [6].

In this paper we take a different approach. In Section II we define 4 different types of users by using Bartle's taxonomy [3]. In Section III we briefly explain the methods and data used. In Section IV we analyse existing data and create models to predict students' Bartle type. In Section V we create and evaluate logistic regression models. Then, in Section VI, we use the k-means clustering algorithm to discover whether some extra types can be found.

Finally, in Sections VII and VIII, a decision is made on what to give to each type so that every user can have a more gamified experience in eQuiz. The reason for this is because previous studies, see Jia, Karanam and Voida [6], have shown that there is no one-size-fits-all approach to gamification, and contrasting personality types react differently to different types of gamification.

II. Bartle types

According to Bartle, players are classified into one of four types, based on their main motivation for playing games. They are often labeled according to the 4 suits of cards.

Bartle's types are graphically represented on Figure 1, by players' preference between other players and the world (x axis), and their preference between acting and interacting (y axis).

These types should serve as guidelines, because every game has a different profile of players.

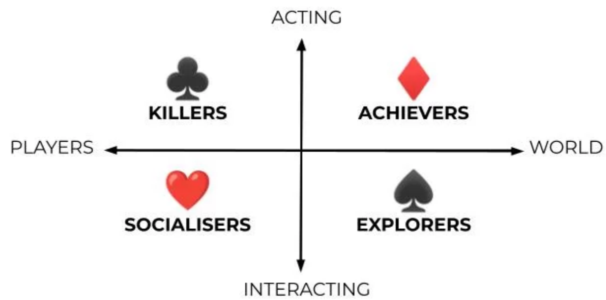

Fig. 1: Chart of Bartle types.

A. *Achievers (♢ , slov. perfekcionisti)*

They are the users that find joy in completing things. The main way to cater to those individuals is to implement an achievement system and a good visualization of their progress in order to motivate them.

On eQuiz, problems are split into topics. Achievers will be users that will have topics close to completion. They will have more than a percentage α completion. This will be elaborated upon in Section IV.

B. *Killers (♣, slov. prvaki)*

They are highly competitive individuals. They are the players that are very persistent and want to be the best or first in something. These are the players that enjoy competition and like to dominate other players.

On eQuiz, killers will be defined as users that have attempted more problems than most of the players with a very high accuracy.

C. *Explorers (♠, slov. raziskovalci)*

These individuals are the players that want to discover new places and learn about the story of the game's world.

In eQuiz, explorers would represent the students that enjoy studying probability and statistics.

D. *Socializers (♡, slov. družabne duše)*

These individuals are the players that enjoy talking to people or characters within the game. They differ from explorers because they enjoy interacting with characters rather than the world.

Explorers and socializers will be mainly discussed in Section VII, because they are difficult to define with the current data.

III. METHODOLOGY

To classify the students we will be looking at their:

- number of solved problems,
- accuracy (proportion of success) and
- number of attempts.

Proportion of success is the number of correctly solved problems divided by the total number of solved problems.

We will be analysing data that contains students' results on problems in the eQuiz group "Probability and Statistics" for years 2023 and 2024. This includes every attempt from every student on every problem in eQuiz, including the incorrect ones. Group "Probability and Statistics" contains different problems from the one-semester course of the same name, which are organized first into topics and then into subtopics.

~~Students do not need to study some subtopics on the exam.~~ Furthermore, the answers to some questions is a written text (essay questions). These 2 types of questions will be ignored when classifying achievers.

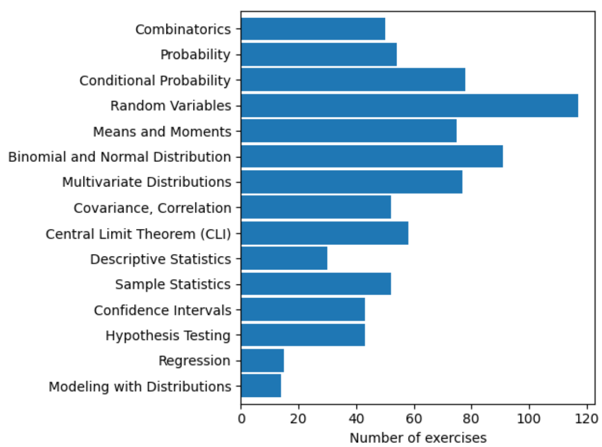

Fig. 2: Number of exercises per topic (see appendix A for exercises per subtopic).

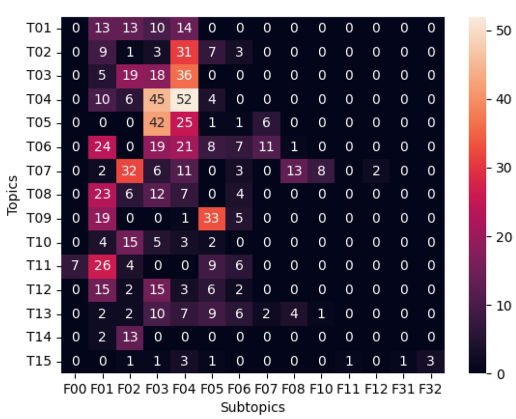

Fig. 3: Heatmap of the number of exercises per subtopic.

The included topics are:

- Probability,
- Conditional Probability,
- Random Variables,
- Means and Moments,
- Binomial and Normal Distribution:
  - Bernoulli Trials,
  - Stirling's Formula,
  - Error Function,
  - Standardization.
- Multivariate Distributions:
  - Definition of Random Vectors,
  - Discrete Random Vector,
  - Continuous Random Vector,
  - Polynomial Distribution,
  - Multivariate Normal Distribution,
  - Independence of Random Variables,
  - Functions of Random Variables.
- Covariance, Correlation,
- Central Limit Theorem (CLI):
  - Hereditery Property of Normal Distribution,
  - Central Limit Theorem for E(X),
  - Central Limit Theorem for Proportions.
- Descriptive Statistics,
- Sample Statistics,
- Confidence Intervals:
  - Definition, Confidence Level,
  - Small and Large Samples,
  - Confidence Interval for Expected Value,
  - Confidence Interval for Proportion.
- Hypothesis Testing,
- Regression:
  - Least Square Method,
  - First and Second Regression Lines.

In this paper two different prediction models will be used. The first model counts the number of a type after n weeks from the beginning of the semester and number of said type at the end of the semester. For example, if a student is not considered an achiever after 6 weeks, but is an achiever at the end of the semester, then our prediction is considered incorrect.

The second type of model will use logistic regression to predict if a student is of a certain type after n weeks from the beginning of the semester. The measurement of accuracy for both models will be the sum of the true positive and true negative, see Figure 4.

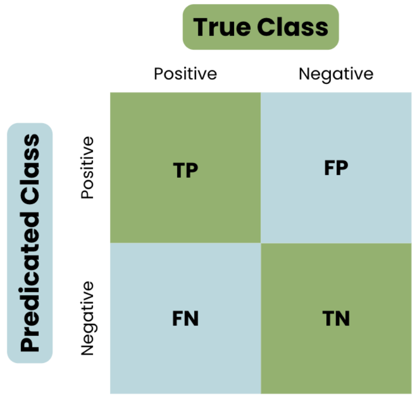

Fig. 4: Evaluation of the performance of classification methods [2]. Image from datacamp.com.

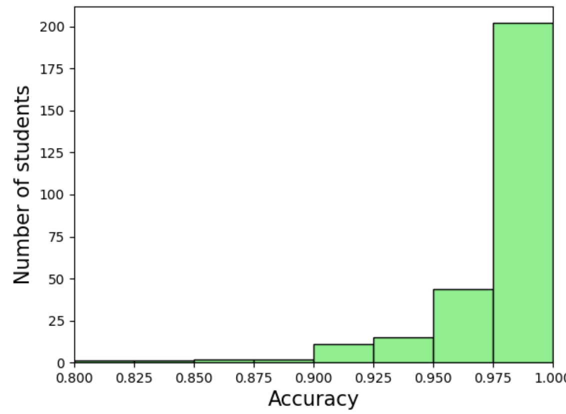

Fig. 6: Histogram of students' accuracy.

Data analysis was performed using the Python programming language (libraries: *numpy, matplotlib, pandas and sklearn*).

IV. Data Analysis

There were 368 students with 189.05 average number of problems solved, see Figure 5. Achievers and Killers will need to solve a lot of problems, in order to achieve high accuracy. For this reason, from here on, we will only be looking at students that have solved more than 50 problems. This will leave us with 278 students.

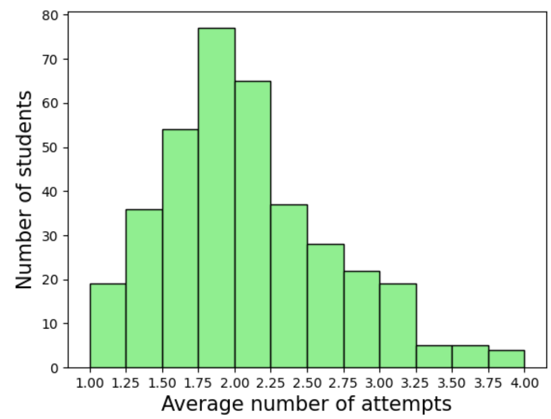

Fig. 7: Histogram of average number of attempts per problem.

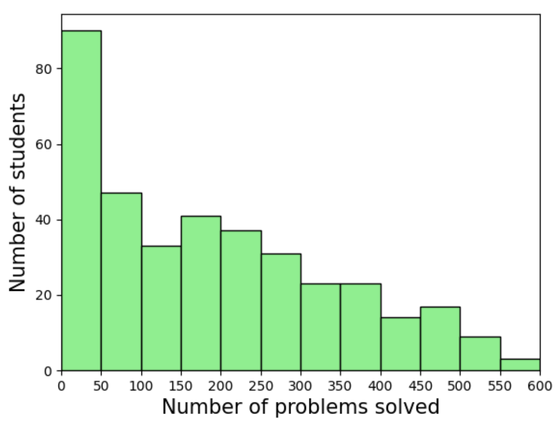

Fig. 5: Histogram of number of problems solved by students.

Next we want to follow the students' accuracy, see Figure 6. One can see that the distribution is very skewed to the left. This means that if we want to distinguish between higher and lower accuracy students, it should be done by creating very small intervals. For example, if we were to pick intervals of size 0.05, then the vast majority of the students would belong in the last interval between 0.95 and 1. The histograms were made using the pyplot.hist function from the matplotlib library.

A. Achievers

There was not a single student who had solved more than 50% of each group's problems, see Table I. Most topics had a low completion rate, see appendix B. There are no achievers for $\alpha = 60\%$ even if we only looked at topics T02, T03 and T04, see Table II.

α	20%	30%	40%	50%
Non-achievers	0.9160	0.9695	0.9962	1
Achievers	0.0840	0.0305	0.0038	0

TABLE I: Achiever and non-achiever percentage (T02-T14).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.4275	0.6870	0.7710	0.8702	0.9580	1
Achievers	0.5725	0.3130	0.2290	0.1298	0.0420	0

TABLE II: Achiever and non-achiever percentage (only topics T02, T03 and T04).

B. Killers

One of the requirements for a student to be considered a killer is to have solved a certain number of problems which is dependent on how much time has passed since the start of the semester. Because of this, a student is a killer if they have correctly solved more problems than 75% of the students at that moment in time and had a very high accuracy.

min accuracy	0.97	0.98	0.99	1
Killers	0.20	0.17	0.09	0.01
Non killers	0.80	0.83	0.91	0.99

TABLE III: Percentage of players classified as killers at the end of the semester.

weeks \ min accuracy	0.97	0.98	0.99	1
4 (77)	0.46	0.52	0.61	0.75
5 (84)	0.47	0.53	0.65	0.77
6 (121)	0.56	0.61	0.69	0.83

TABLE IV: Prediction accuracy of killers after n weeks. The numbers in parentheses represent the upper quartile of number of problems solved after n weeks.

V. Logistic Regression

The quality of the logistic regression model [7] will be measured using accuracy and R^2. The **coefficient of determination**, denoted by R^2, is the proportion of the variation in the dependent variable that is predictable from the independent variable(s) [7].

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}}$$

where

$$\text{SS}_{\text{res}} = \sum_i (y_i - f_i)^2$$

,

$$\text{SS}_{\text{tot}} = \sum_i (y_i - \overline{y})^2$$

and

$$\begin{aligned} y_i &= \text{ observed value,} \\ f_i &= \text{ predicted value,} \\ \overline{y} &= \text{ mean of observed } y. \end{aligned}$$

Due to the lack of students classified as achievers, there is no reason to make a model for them. There will be 2 logistic regression models for killers.

The independent variables for the first model are the number of problems solved and accuracy (the ratio between the number of correctly solved problems and total number of solved problems).

The independent variables for the second model are the same, but calculated for each group individually. This means that there are 6 parameters for testing periods of 4 or 5 weeks and 8 parameters for testing periods 6 weeks.

The dependent variable for both models will be a binary value which will tell us whether the student belongs to the type.

weeks \ min accuracy	0.97	0.98	0.99
4 (77)	0.79	0.82	0.86
5 (84)	0.79	0.82	0.86
6 (121)	0.79	0.82	0.86

TABLE V: Prediction accuracy of killers after n weeks (2 independent variables).

weeks \ min accuracy	0.97	0.98	0.99
4 (77)	-0.26	-0.23	-0.16
5 (84)	-0.26	-0.23	-0.16
6 (121)	-0.26	-0.23	-0.16

TABLE VI: R^2 score of killers after n weeks (2 independent variables).

weeks \ min accuracy	0.97	0.98	0.99
4 (77)	0.77	0.68	0.54
5 (84)	0.77	0.68	0.57
6 (121)	0.77	0.68	0.58

TABLE VII: Prediction accuracy of killers after n weeks (2 independent variables per topic).

weeks \ min accuracy	0.97	0.98	0.99
4 (77)	-0.31	-0.46	-0.85
5 (84)	-0.31	-0.46	-0.74
6 (121)	-0.27	-0.46	-0.69

TABLE VIII: R^2 score of killers after n weeks (2 independent variables per topic).

One can deduce from the negative values in Tables VI and VIII that the quality of the 2 logistic regression models is quite poor.

VI. K-Means Clustering

K-Means clustering is an unsupervised learning algorithm [9], that splits N observations into M clusters, each observation belonging to exactly 1 cluster. First select M points to be the centroids for the clusters. Usually, the initial centroids are M random observations. Observations are considered part of the same cluster if they share their closest centroid. In our case the measure for distance is squared Euclidean distance. After each observation has been assigned a cluster, the centroids are moved to the average position of the observations of their cluster. Finally, we repeat this process of calculating which cluster each observation belongs to and moving the centroids until the stop condition is met. It is not guaranteed that we will find the optimal positions, but we get very close after a few steps. The stop condition can be either reaching a maximum number of iterations or not reaching a minimum distance traveled by all centroids in a given iteration (tolerance).

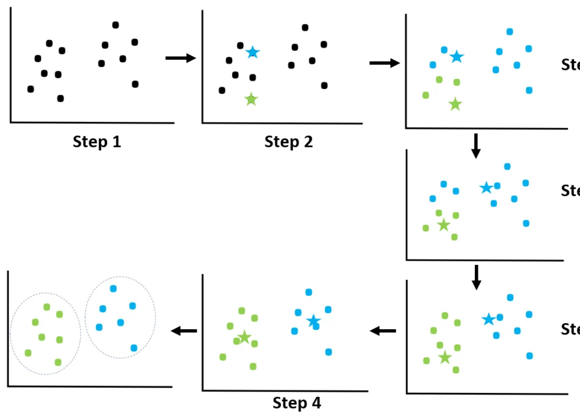

Fig. 8: Step by step representation of how the k-means algorithm works. Different colors represent different clusters. Image from r/learnmachinelearning [5].

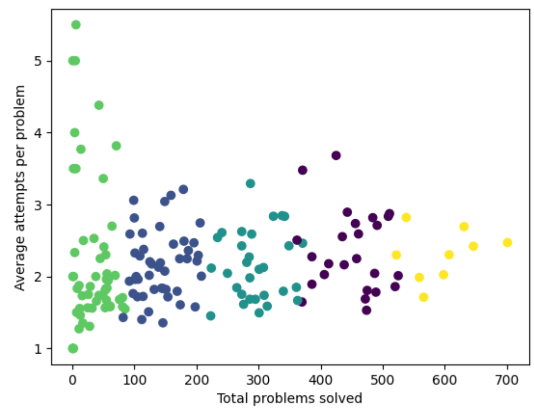

Fig. 10: Relationship between the number of problems solved and average attempts per problem (`tolerance` $= 0.0001$, `maximum iterations` $= 400$).

In the following figures, total problems solved, average difficulty, average attempts and accuracy were used. This means that the figures are two-dimensional visualizations of a four-dimensional space. The reason why 5 clusters were used in Figures 9 and 10 was because of the assumption that at least 4 types exist (Bartle's types). Selecting a different number of clusters yielded similar results as shown in Figures 11 and 12. The following visualizations were made using the pyplot.scatter function from the matplotlib library and the clustering was done using the cluster.KMeans function from the sklearn library.

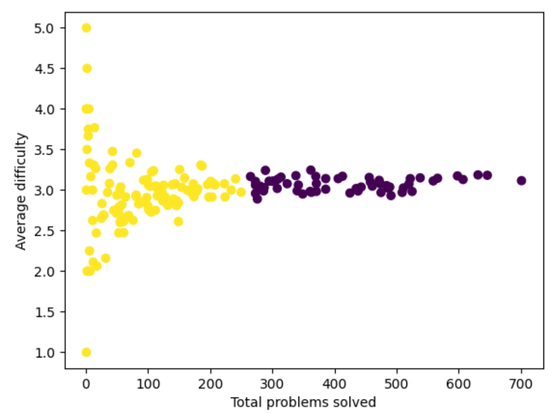

Fig. 11: Relationship between all the problems solved by each student and the average difficulty of those problems, 2 clusters (`tolerance` $= 0.0001$, `maximum iterations` $= 400$).

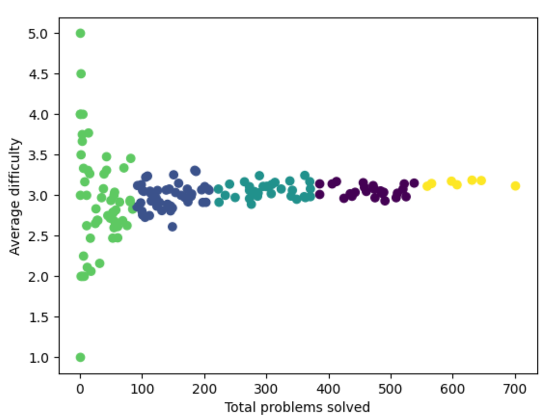

Fig. 9: Relationship between all the problems solved by each student and the average difficulty of those problems (`tolerance` $= 0.0001$, `maximum iterations` $= 400$). Each cluster is represented by a color.

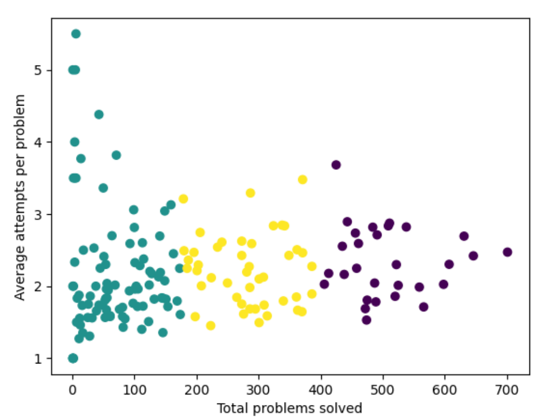

Fig. 12: Relationship between number of problems solved and average attempts per problem, 3 clusters (`tolerance` $= 0.0001$, `maximum iterations` $= 400$).

VII. Socializers and Explorers

The current major issue with adding personalized gamification by using achievers and killers is that from the previous data, we can expect a large portion of the students to not belong to any type. Because of this, it would be very beneficial to find other types of students. In order to classify users into these 2 (or more) other types, a broader range of data is needed.

To appeal to explorers, references and links to interesting articles, books, videos and journals can be added next to some problems. Furthermore, tracking how many people have clicked on these links will give us more information about students' interests and how successful each link is.

Furthermore, adding questions related to these links will tell us who actually read the material and will create a more interactive experience.

Adding avatars and badges to every user will only have a positive effect, see Jia et al. [8] and Sailer et al. [6]. This is especially important for socializers. Also have some avatars that are not initially available and can be bought with points earned from solving questions, reaching a certain milestone in a given month or interacting with other users.

To appeal to socializers, students can be notified when one of their friends gets a difficult achievement or levels up and be given the opportunity to interact with them by sending them a message. Human-to-human interaction is crucial for socializers. In addition, implementing a forum where students can have discussions and ask questions is an excellent way to improve eQuiz for socializers and give us new ways to measure activity on the application.

Adding these types of features is very important, not only because it will increase the quality of eQuiz, but will also give us different types of data to work with which can explain the profile of players that exist on the platform. Only 4 types are defined in this paper, but there might be many more.

VIII. Conclusion

A good approach to gamify eQuiz would be to wait 3-6 weeks or until a certain portion of the problems have been solved to assign proper types to students. To appeal to the killers type, leaderboards should be added, see Jia et al. [6]. For achievers, adding a better visualization of progress, see Jia et al. [6] and a performance graph, see Sailer, Hense, Mayr and Mandl [8], will be helpful. The completion bar that can be seen on eQuiz for each topic should be made more apparent by increasing its width and making it more animated. Furthermore, a line graph that plots performance should be added as well (to improve task meaningfulness). Because progress and performance graphs can make users with lower emotional stability feel pressured to do well, see Jia et al. [6], users should be given the option to set their desired activity level. Streaks should also be added for each consecutive day of using eQuiz because they promote habit formation.

A. Achievement ideas

- Just getting started. - Solve an entire topic of problems.
- On a roll. - Correctly solve 10 questions in a row.
- People person - Have at least 10 friends and send a total of 100 messages.
- Over-achiever - Have every other achievement.
- Yup, gravity still works! - Go down one or more rankings after getting to 1st place on the leaderboard.
- I Swear! I Didn't cheat! - Solve a quiz question in less than 5 seconds.

We can also use AI to create some interesting achievement and badge ideas.

B. Achievements and Badges using AI

- Daily Streak: Study every day for a week.
- Consistency King/Queen: Log in and complete work every day for a month.
- Night Owl: Study late at night (after 10 PM) 10 times.
- Early Bird: Study early in the morning (before 6 AM) 10 times.
- Quiz Master: Score 100% on 5 quizzes.
- Wizard : Achieve a 90% average in all quizzes.
- Probability Pro: Score 90% or higher on a probability exam.
- Champion: Finish 1st in this month's leaderboard.

Fig. 13: Badge for the "Wizard" achievement. Made using a stable diffusion model [1].

Fig. 14: Badge for the "Night Owl" achievement. Made using a stable diffusion model [1].

References

[1] URL: https://stablediffusionweb.com (visited on 07/04/2024).

[2] N. A. Ahmed. *What is A Confusion Matrix in Machine Learning? The Model Evaluation Tool Explained*. URL: https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning (visited on 09/13/2024).

[3] R. Bartle. “Hearts, clubs, diamonds, spades: Players who suit MUDs”. In: *Journal of MUD research* 1 (1996), p. 28.

[4] R. Gafni, D. B. Achituv, S. Eidelman and T. Chatsky. “The effects of gamification elements in e-learning platforms”. In: *Online Journal of Applied Knowledge Management* 6 (2018), p. 17.

[5] J. Grus. *K-means and hierarchical clustering with Python*. Sebastopol, CA: O’Reilly Media, Inc., 2016.

[6] Y. Jia, Y. Karanam and S. Voida. “Personality-targeted Gamification: A Survey Study on Personality Traits and Motivational Affordances”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 10–13.

[7] D. G. Kleinbaum and M. Klein. *Logistic Regression: A Self Learning Text*. 3rd ed. New York: Springer, 2010.

[8] M. Sailer, J. U. Hense, S. K. Mayr and H. Mandl. “How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction”. In: *Computers in Human Behaviour* 69 (2016), p. 10.

[9] A. A. Patel. *Hands-On Unsupervised Learning Using Python*. Sebastopol, CA: O’Reilly Media, Inc., 2019.

APPENDIX A
BARPLOT OF NUMBER OF TOPICS’ PROBLEMS

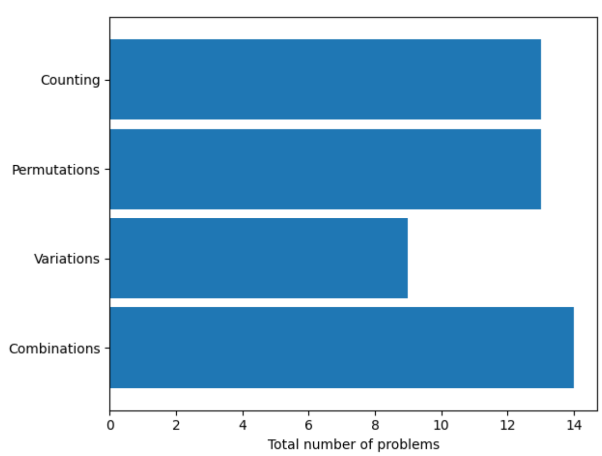

Fig. 15: Combinatorics (T01) with its subtopics.

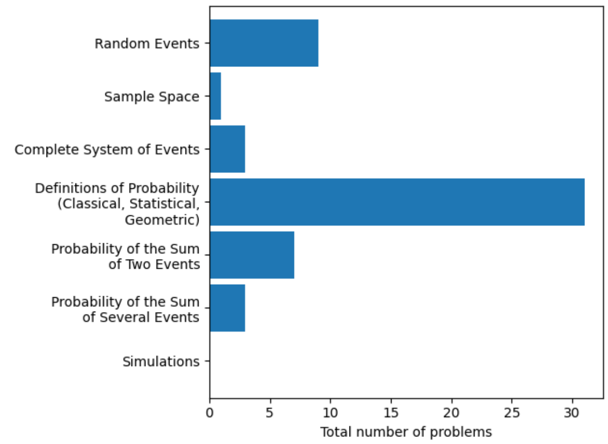

Fig. 16: Probability (T02).

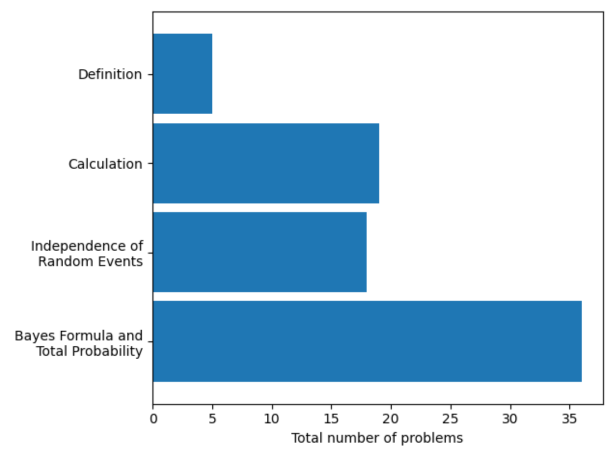

Fig. 17: Conditional Probability (T03).

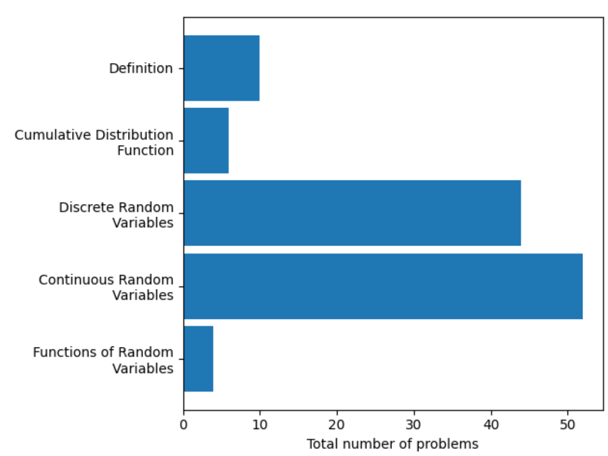

Fig. 18: Random Variables (T04).

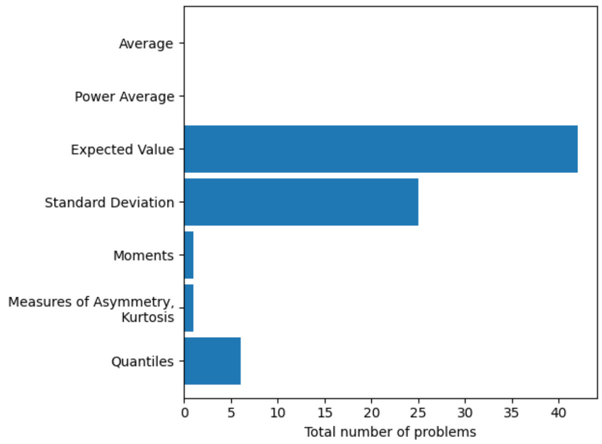

Fig. 19: Means and Moments (T05).

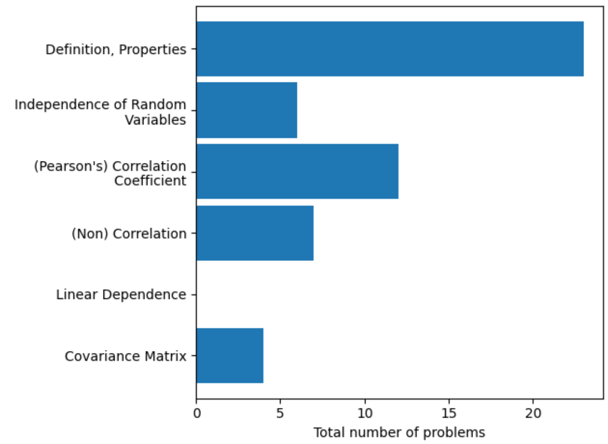

Fig. 22: Covariance, Correlation (T08).

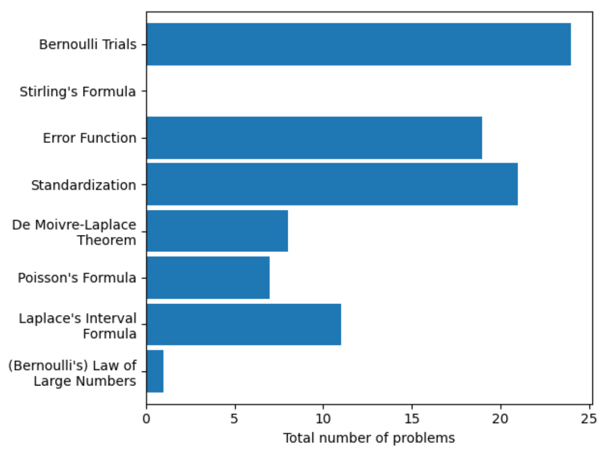

Fig. 20: Binomial and Normal Distribution (T06).

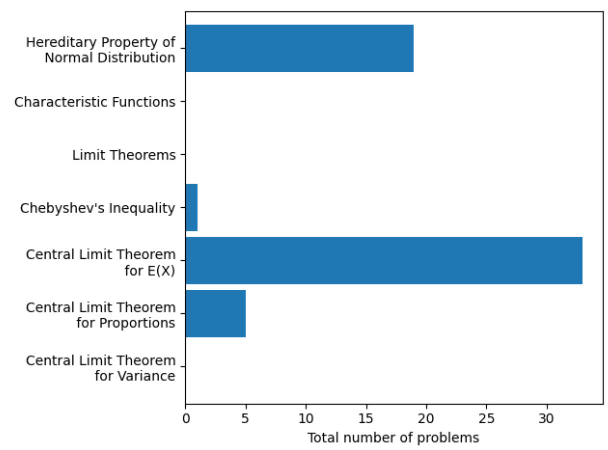

Fig. 23: Central Limit Theorem (CLI) (T09).

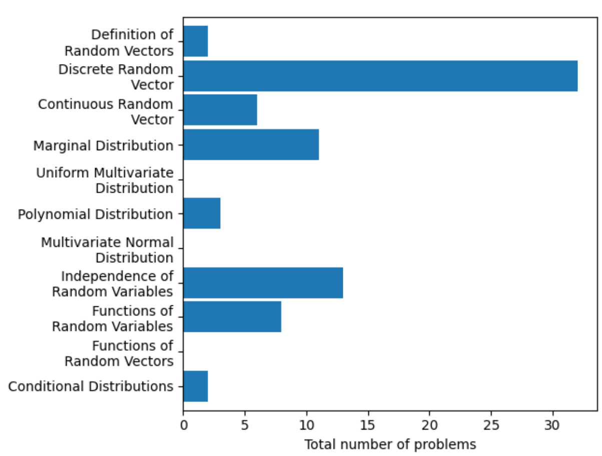

Fig. 21: Multivariate Distributions (T07).

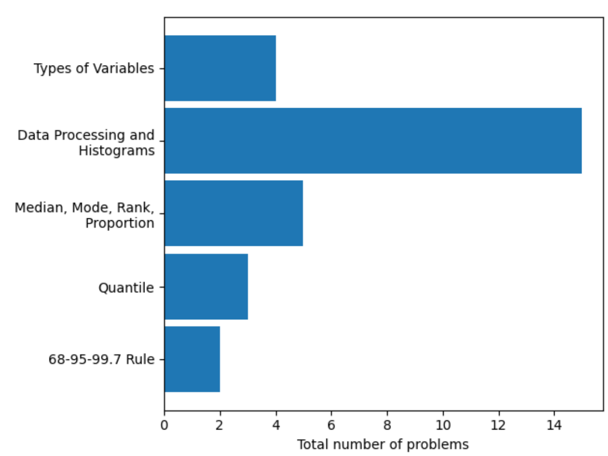

Fig. 24: Descriptive Statistics (T10).

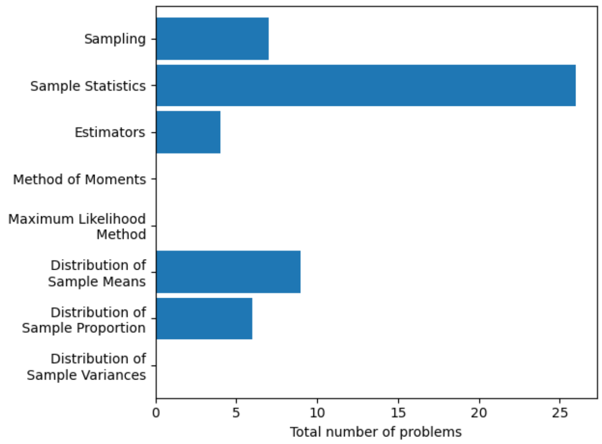

Fig. 25: Sample Statistics (T11).

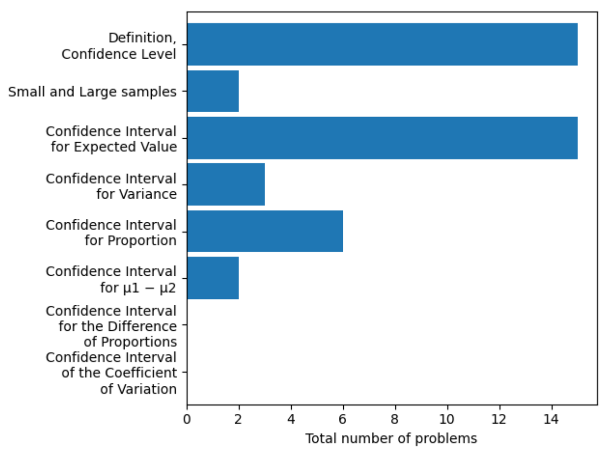

Fig. 26: Confidence Intervals (T12).

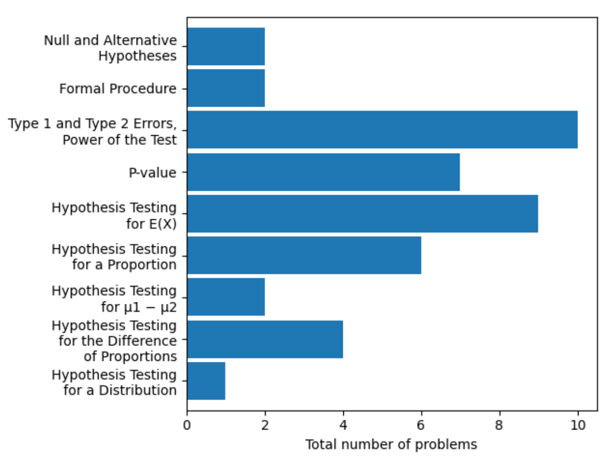

Fig. 27: Hypothesis Testing (T13).

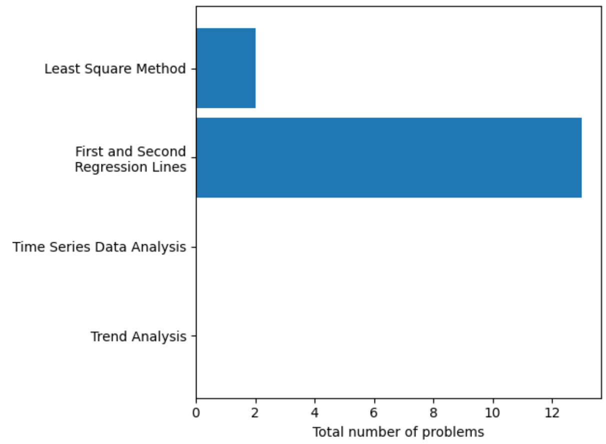

Fig. 28: Regression (T14).

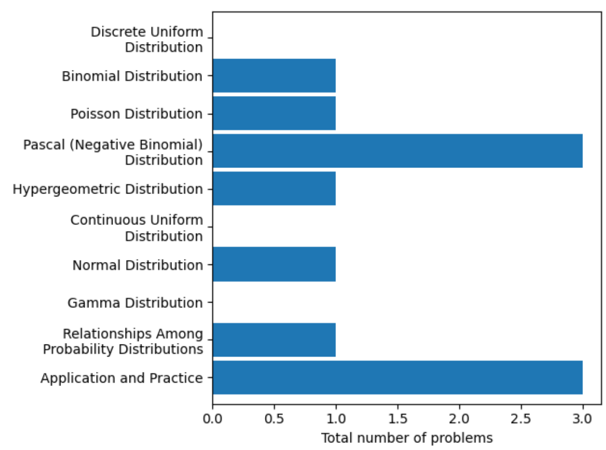

Fig. 29: Modeling with Distributions (Overview) (T15).

Appendix B
Proportion of Topic Completion

α	10%	20%	30%	40%	50%	60%
Achievers	0.8321	0.6947	0.4695	0.2519	0.1260	0.0191
Non-achievers	0.1679	0.3053	0.5305	0.7481	0.8740	0.9809

TABLE IX: Achiever and non-achiever percentage for Probability (T02).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3511	0.6221	0.7061	0.7901	0.8702	0.9389
Achievers	0.6489	0.3779	0.2939	0.2099	0.1298	0.0611

TABLE X: Conditional Probability (T03).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.0878	0.1908	0.3664	0.5573	0.7176	0.8969
Achievers	0.9122	0.8092	0.6336	0.4427	0.2824	0.1031

TABLE XI: Random Variables (T04).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3626	0.5382	0.6947	0.7366	0.7977	0.9084
Achievers	0.6374	0.4618	0.3053	0.2634	0.2023	0.0916

TABLE XII: Means and Moments (T05).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3969	0.6069	0.7939	0.8588	0.9046	0.9733
Achievers	0.6031	0.3931	0.2061	0.1412	0.0954	0.0267

TABLE XIII: Binomial and Normal Distribution (T06).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.4084	0.5305	0.6756	0.7901	0.8779	0.9351
Achievers	0.5916	0.4695	0.3244	0.2099	0.1221	0.0649

TABLE XIV: Multivariate Distributions (T07).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3626	0.5076	0.6870	0.7748	0.8511	0.9427
Achievers	0.6374	0.4924	0.3130	0.2252	0.1489	0.0573

TABLE XV: Covariance, Correlation (T08).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.2252	0.3550	0.6069	0.6985	0.7786	0.8817
Achievers	0.7748	0.6450	0.3931	0.3015	0.2214	0.1183

TABLE XVI: Central Limit Theorem (CLI) (T09).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3053	0.4962	0.6107	0.9160	0.9924	1
Achievers	0.6947	0.5038	0.3893	0.0840	0.0076	0

TABLE XVII: Descriptive Statistics (T10).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3740	0.5267	0.6908	0.8588	0.9885	1
Achievers	0.6260	0.4733	0.3092	0.1412	0.0115	0

TABLE XVIII: Sample Statistics (T11).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3053	0.4008	0.5725	0.6527	0.7672	0.9084
Achievers	0.6947	0.5992	0.4275	0.3473	0.2328	0.0916

TABLE XIX: Confidence Intervals (T12).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.3435	0.4237	0.5076	0.5992	0.6756	0.8397
Achievers	0.6565	0.5763	0.4924	0.4008	0.3244	0.1603

TABLE XX: Hypothesis Testing (T13).

α	10%	20%	30%	40%	50%	60%
Non-achievers	0.4122	0.5115	0.5954	0.6756	0.7443	0.8244
Achievers	0.5878	0.4885	0.4046	0.3244	0.2557	0.1756

TABLE XXI: Regression (T14).

α	10%	20%	30%	40%	50%	60%
Non-explorers	0.7328	0.8130	0.8931	0.9389	0.9771	1
Explorers	0.2672	0.1870	0.1069	0.0611	0.0229	0

TABLE XXII: Explorer and non-explorer percentage.