Case study

Done by: Erokhin Evgenii, Davydov Danil
Group: DS-01
Email: d.davydov@innopolis.university, e.erokhin@innopolis.university

Innopolis

Introduction

Mobile Legends Bang Bang (MLBB) is a popular mobile game that has gained immense popularity worldwide. In 2022, MLBB emerged as the second most popular game in terms of viewer count[1][2]. With millions of players from different countries, the game offers a unique platform for players to showcase their skills and earn money through various competitions and tournaments.

The primary objective of this case study is to provide a comprehensive analysis of the data related to MLBB players, their positions, earnings, nationality, and retirement. By examining the relationship between these factors, we aim to gain insights into the game's ecosystem and the factors that influence the success and longevity of players.

To achieve this objective, we will use a dataset[3] containing information about MLBB Pro players, their positions, earnings, nationality, and retirement. The dataset will be analyzed using various statistical methods and visualizations to identify patterns and trends. The findings from this analysis will help us answer the following questions:

1) Is there a relationship between a player's position and the amount of money they earn?
2) Is there a connection between a team's earnings and the nationality of its members?
3) Is there a link between a player’s status and their age and nationality?

The results of this case study will provide valuable insights into the factors that contribute to the success and retirement of MLBB players.

Data

In this case study, we will analyze the relationship between the position and the money earned, the relationship between the team's earned money and nationality, and the dependency of retirement on age and nationality. To address these questions, we will use a dataset containing information about MLBB players, their positions, earnings, nationality, and retirement.

The dataset contains information about professional players in South East Asia as of November 1, 2022. The dataset includes two files: Player_details.csv and ML_players.csv.

These files can be merged on the index column provided. The Player_details.csv file contains information about the players:

- **Column0(index)**: Basically unique numbers or row numbering.
- **Players**: Name of the players there are known by (IGN, In game name).
- **Name**: Players real name.
- **Team**: Professional team the player plays for.

The ML_players.csv file contains:

- **Column0(index)**: Basically unique numbers or row numbering.
- **Nationality**: Country the player belongs to. (Birth country)
- **Birth_date**: Day the player was born.
- **Status**: Players current status as a professional player.
- **Laner**: Role of the player in professional tournaments.
- **Earnings**: Money earned as a professional player.

By analyzing this dataset, we aim to gain insights into the factors that influence the success and retirement of MLBB players.

Study protocol:

1) Merge two files into one dataset
2) Preprocess the data
3) Exploring the dataset
4) Answering the stated questions
5) Making conclusions

Statistical tools, other software

In this case study, we have used various statistical tools and software to analyze the dataset of MLBB players. The following are the software tools and methods used in the study:

1. Pandas: This is a Python library that is used for data analysis and manipulation. We have used Pandas to import, clean, and manipulate the dataset.
2. NumPy: This is another Python library that is used for scientific computing. We have used NumPy to perform numerical calculations and generate random numbers.
3. Matplotlib: This is a Python library for creating visualizations and plots. We have used Matplotlib to create graphs and charts to visualize the data.
4. SciPy: This is a Python library that provides various scientific algorithms and functions. We have used SciPy to perform statistical tests and calculations.
5. Chi-squared test: This is a statistical test used to determine if there is a significant association between two categorical variables. We have used the Chi-squared test to analyze if retirement is dependent on age and nationality, and the relationship between the team's earned money and nationality.
6. Kruskal-Wallis test: This is a non-parametric test used to compare the means of multiple groups. We have used the Kruskal-Wallis test to analyze the relationship between the position and the money earned.

Results

Upon consolidating multiple files into a single dataset, we obtained a dataframe comprising 429 entries with eight distinct features (Fig. 1). Subsequently, a decision was made to retain players whose roles correspond to one of the five lanes or their combinations, namely Roamer, Jungle, EXP Laner, Mid Laner, and Golder (Fig. 2), which resulted in a total of 428 players. Players with combined roles were categorized as "2_plus_lanes."

index	Players	Name	Team	Nationality	Birth_date	Status	Laner	Earnings
0	Wizzking	Mohd Zulkarnain Hj Mohd Zulkifli	NaN	Brunei	April 11, 1997 (age 25)	Inactive	Gold Laner	$70,849
1	ATEV	Ly Kimhong	Logic Esports	Cambodia	NaN	Active	Roamer	$6,399
2	ATM	Kosal Piseth	Burn x Team Flash	Cambodia	NaN	Active	EXP Laner	$8,473
3	Arishem	Kunn Chankakada	Burn x Team Flash	Cambodia	NaN	Active	NaN	$8,933
4	BOXI	Sok Viera	See You Soon	Cambodia	NaN	Active	Roamer	$12,475

Fig 1. Merged players dataframe

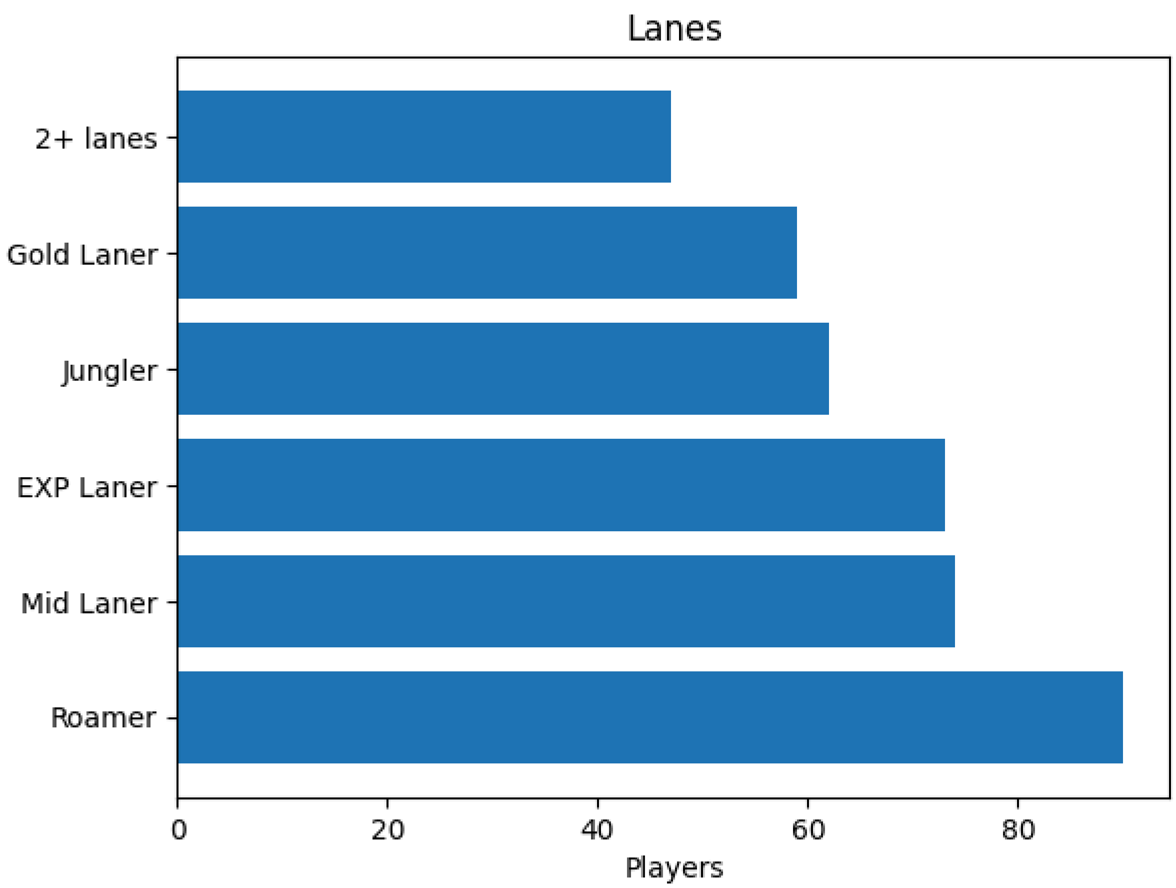

Fig 2. Players’ lanes

An analysis of the dataset revealed that the players represented 12 different countries (Fig. 3) and were classified into three categories (Fig. 4): active, nonactive, and retired.

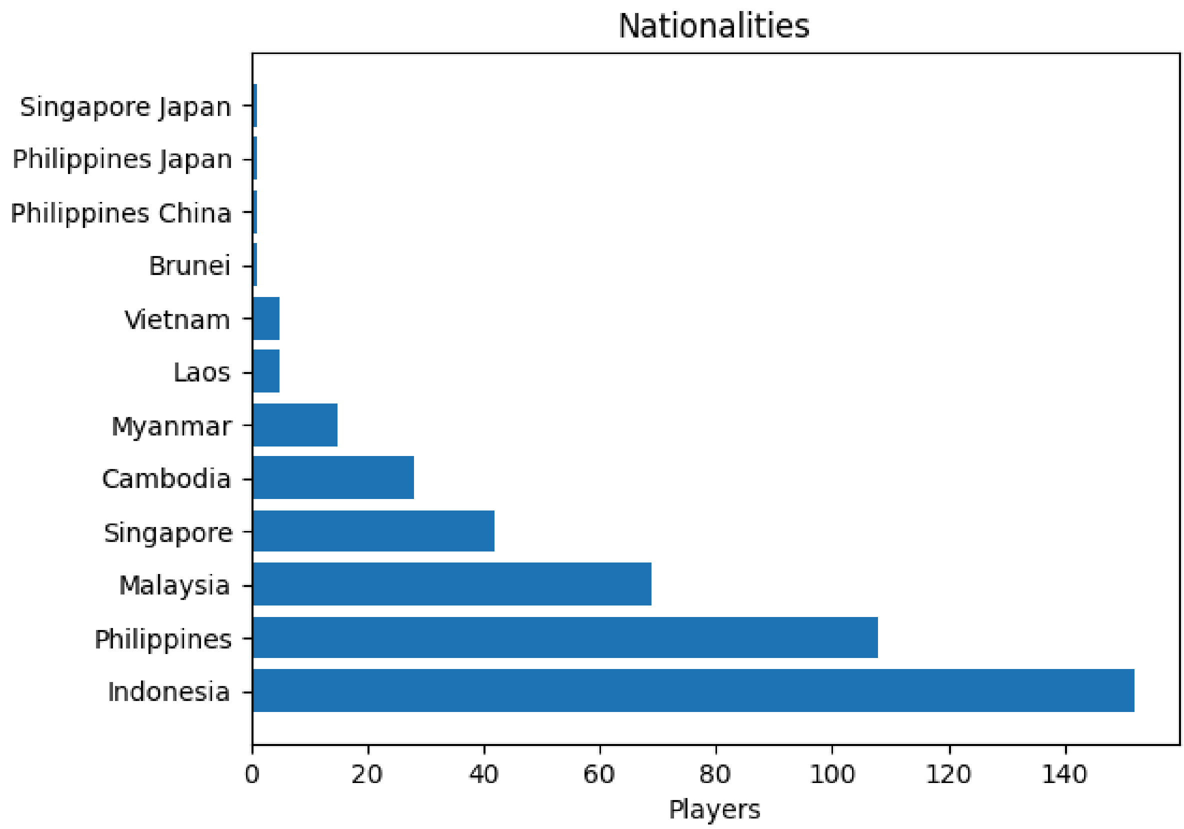

Fig 3. Players’ nationalities

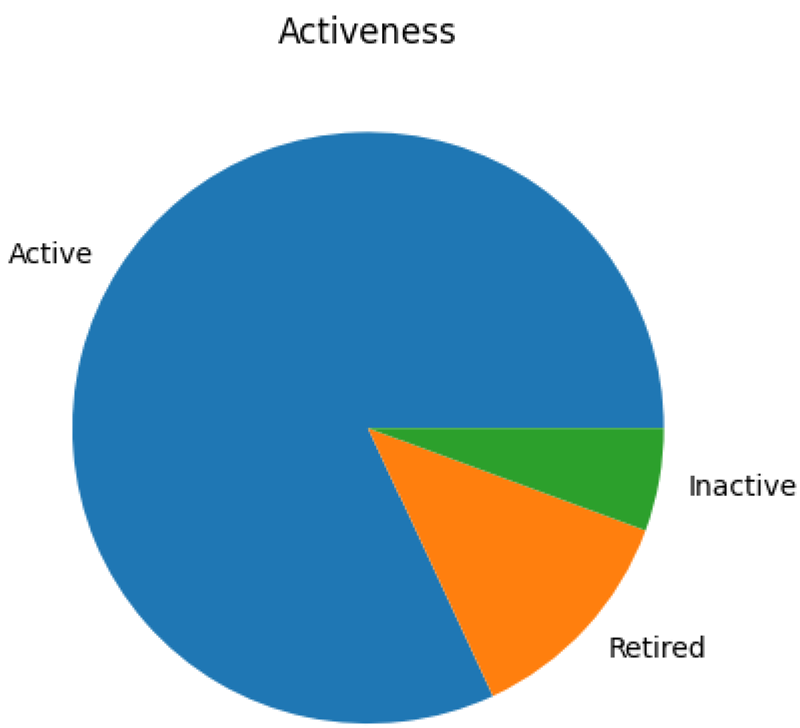

Fig 4. Players status

To address the first research question, players were grouped based on their roles and subjected to a Kruskal-Wallis test with a significance level of 0.05 to ascertain the relationship between player positions and earnings. The utilization of a nonparametric test was deemed appropriate due to the non-normal distribution of the data(Appendix Fig.-s 8-13).

Kruskal-Wallis test
Null Hypothesis (**H0**): The medians are equal, indicating no dependency.

Alternative Hypothesis (**H1**): The medians are not equal, suggesting a dependency.

Gained statistics:
KruskalResult(statistic=12.631058400813776, pvalue=0.027092692805994557)

Looking at p-value < significance level = 0.05 we reject Null hypothesis, thus the player's earnings are dependent from their lane.

In response to the second question, nationalities were consolidated into five regions: 'Cambodia and Vietnam', 'Indonesian archipelago', 'Myanmar and Laos', 'Malay Peninsula', and 'Philippines_reg'. Teams were categorized as top teams or non-top teams based on the total earnings of their players, with a prerequisite that teams must consist of at least four players to ensure a diverse national team structure. A chi-squared contingency test was conducted on a contingency table (Fig. 5) to determine the impact of national structure on a team's earnings status.

Chi-squared contingency test

Null Hypothesis (**H0**): There is no dependency between team equality and the nationality of players.

Alternative Hypothesis (**H1**): There is a significant dependency between team equality and the nationality of players.

Gained statistics:
Chi2ContingencyResult(statistic=56.113494699315126, pvalue=1.8982632636193304e-11, dof=4, expected_freq=array([[12.65822785, 55.69620253, 38.48101266, 8.60759494, 44.55696203],
[12.34177215, 54.30379747, 37.51898734, 8.39240506, 43.44303797]]))

p-value < significance level = 0.05, thus relationship between national structure of a team and the fact that team is in top-list by earnings is significant. Notably, the analysis revealed that professional teams tend to have a higher representation of players from the Philippines and Indonesian archipelago compared to other regions (Fig. 5).

Nationality	Cambodia and Vietnam	Indonesian archipelago	Malay Peninsula	Myanmar and Laos	Philippines_reg
IsTop					
IsInTop	0	64	28	4	64
NotIsInTop	25	46	48	13	24

Fig. 5. Contingency table for teams types and nationality

For the third question, a similar chi-squared test was employed, this time focusing on the dependency of nationality and age on player's status. The countries were further divided into subregions, and players were categorized as Juniors (younger than 21) or Seniors. Contingency table as on Fig. 6. Due to some frequency values falling below the recommended threshold of five, adjustments were made to reduce the number of categories within each group. The revised analysis involved two subregions (Philippines Region and Indonesia Region), two status categories (active and inactive), and two age categories (Juniors and Seniors). Now, the contingency table looks like on Fig. 7.

Chi-squared contingency test:
Null Hypothesis (**H0**): There is no dependency between player status and their nationality and age.

Alternative Hypothesis (**H1**): There is dependency between player status and their nationality and age.

Gained statistics:
Chi2ContingencyResult(statistic=30.12009920925356, pvalue=1.302065209996061 5e-06, dof=3, expected_freq=array([[48.93693694, 50.68468468, 59.42342342, 34.95495495],
 [7.06306306, 7.31531532, 8.57657658, 5.04504505]]))

The chi-squared test results indicated a dependency between player status, nationality, and age, since p-value < significance level = 0.05.

age_category	Nationality	Active	Inactive	Retired
Junior	Cambodia and Vietnam	1	0	0
	Indonesian archipelago	46	0	0
	Malay Peninsula	10	0	0
	Philippines_reg	55	2	0
Senior	Indonesian archipelago	35	5	10
	Malay Peninsula	13	1	4
	Myanmar and Laos	4	0	0
	Philippines_reg	30	6	0

Fig. 6. Contingency table for age, nationality, and status

age_category	Junior		Senior	
Nationality	Indonesia region	Philippines region	Indonesia region	Philippines region
Status				
Active	56	56	48	34
Inactive	0	2	20	6

Fig. 7. Contingency table for age, nationality, and status after reduction

Conclusion

In this case study, we analyzed the relationship between various factors in MLBB player ecosystem. Our findings indicate that there is a relationship between a player's position and the amount of money they earn, a connection between a team's type(top or not top) and the nationality of its members, and a link between a player's status and their age and nationality.

Our analysis shows that players in certain positions tend to earn more money, suggesting that positional skills and performance play a significant role in determining earnings. Additionally, teams with higher earnings tend to have more players from a specific region compared to other teams. This suggests that having a higher concentration of players from a particular region can contribute to a team's success and financial performance. However, it is essential to note that this does not necessarily mean that the team is not diverse in terms of nationalities, as players from different countries can still be part of the same region. Lastly, our findings suggest that players' status is dependent on their age and nationality.

These insights provide valuable information for players, teams, and game developers to improve their strategies and create a more balanced and enjoyable gaming experience. By understanding the factors that contribute to success and retirement in MLBB, stakeholders can make informed decisions and adapt their approaches accordingly.

Point on Bias

In this case study, we analyzed the relationship between various factors in the MLBB player ecosystem. Our analysis was based on a dataset containing information about 428 professional players from 12 different countries. The dataset included information about the players' positions, earnings, nationality, and retirement status.

It is essential to acknowledge some potential biases in the dataset and the study. Firstly, the dataset is not recent, as it was collected almost two years ago. This may limit the generalizability of the findings, as the game's ecosystem and player dynamics may have changed since then.

Secondly, the dataset includes a limited number of players, which may not represent the entire population of MLBB players. This could lead to a sampling bias, as the sample may not be representative of the entire population.

Thirdly, the dataset contains different proportions of laners, nationalities, and other factors. This could introduce a selection bias, as the sample may not be representative of the entire population of MLBB players.

Despite these potential biases, the findings of this case study provide valuable insights into the factors that contribute to the success and retirement of MLBB players. These insights can be used by players, teams, and game developers to improve their strategies and create a more balanced and enjoyable gaming experience. However, it is essential to consider the limitations of the dataset and the study when interpreting the findings.

Appendix

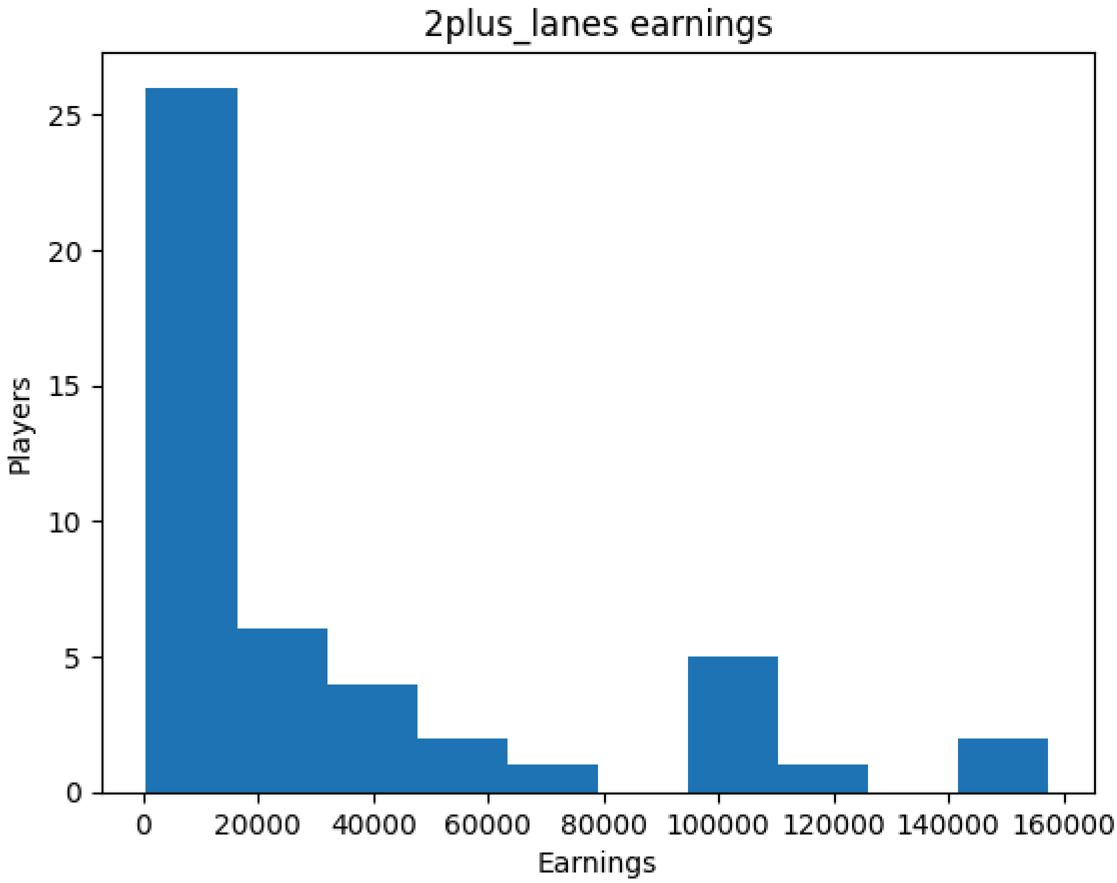

Fig. 8. Distribution of multi-lanes players by earnings

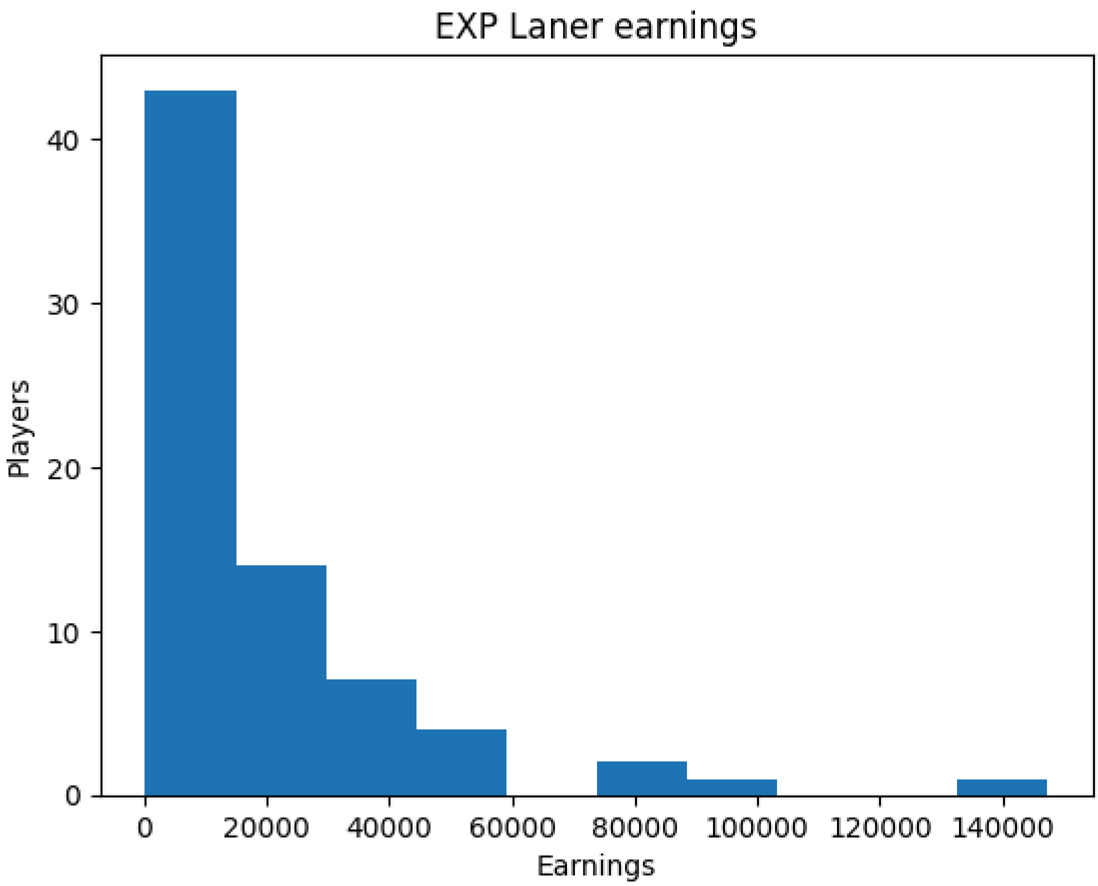

Fig. 9. Distribution of EXP Lane players by earnings

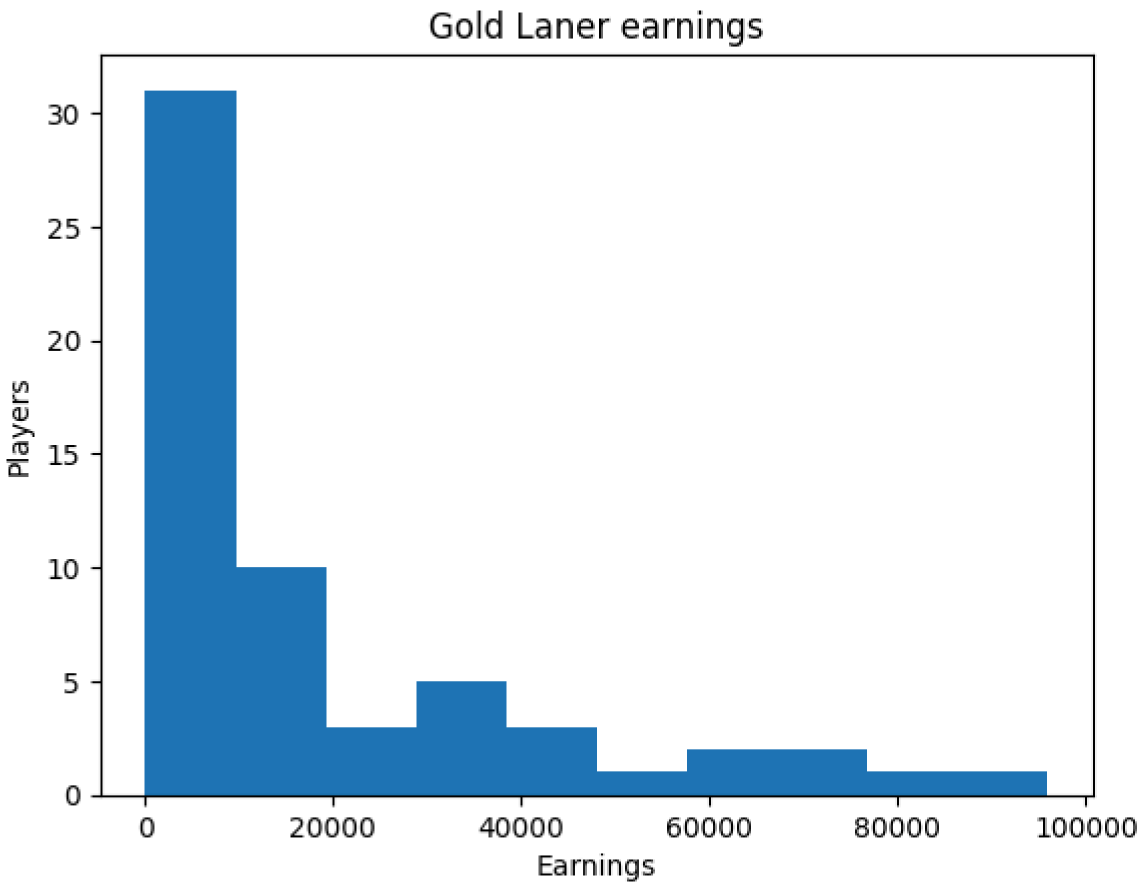

Fig. 10. Distribution of Gold Lane players by earnings

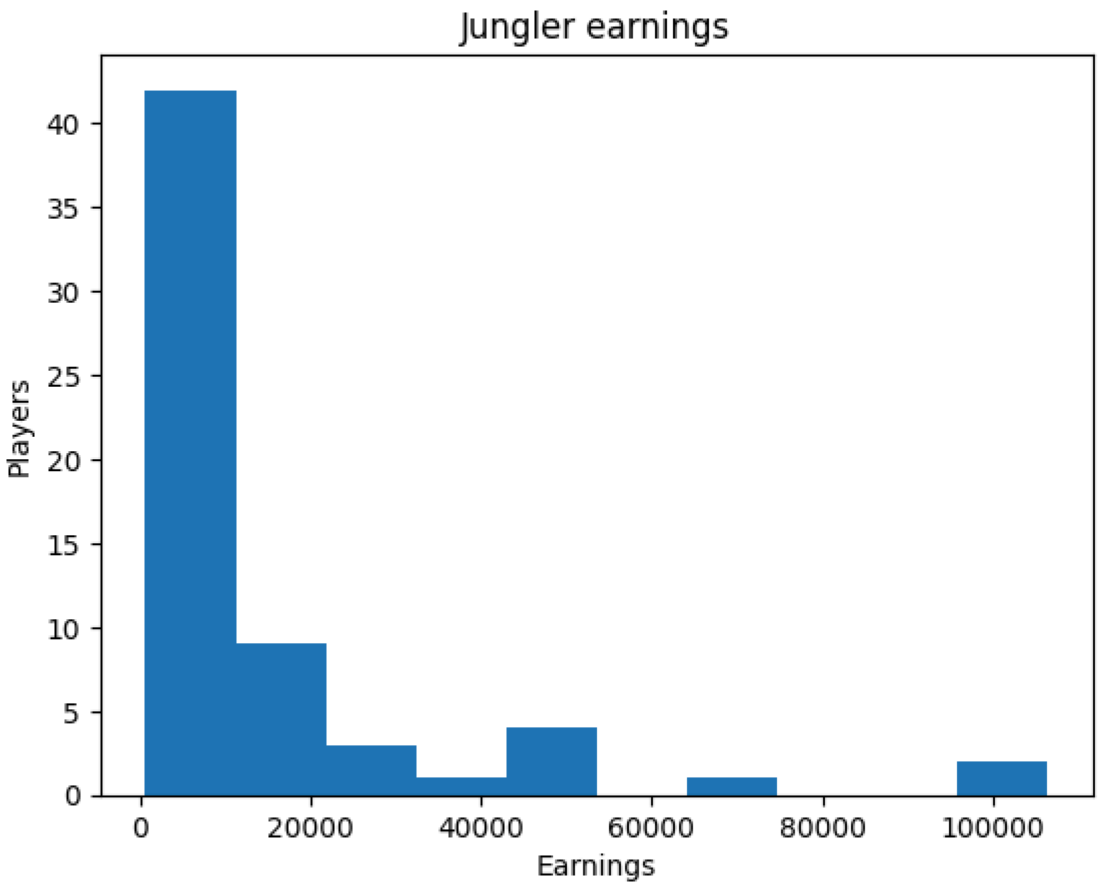

Fig. 11. Distribution of Jungle players by earnings

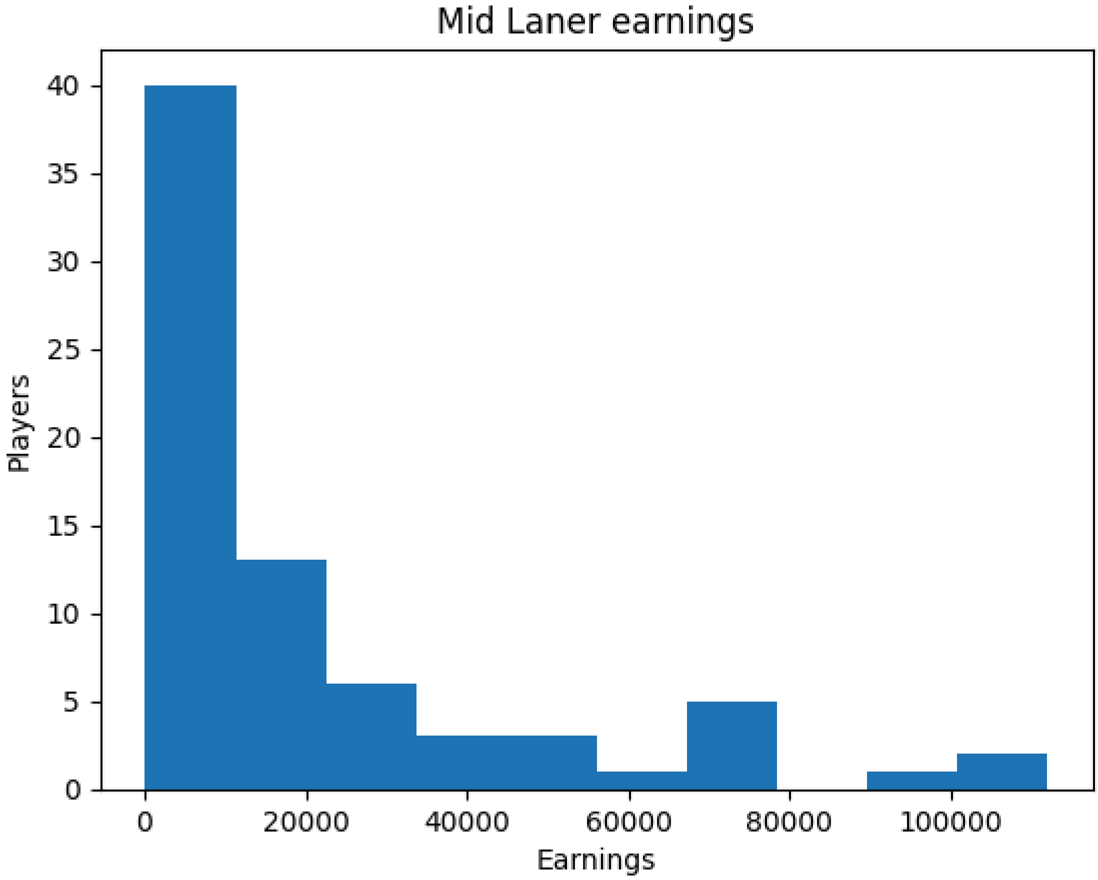

Fig. 12. Distribution of Mid Lane players by earnings

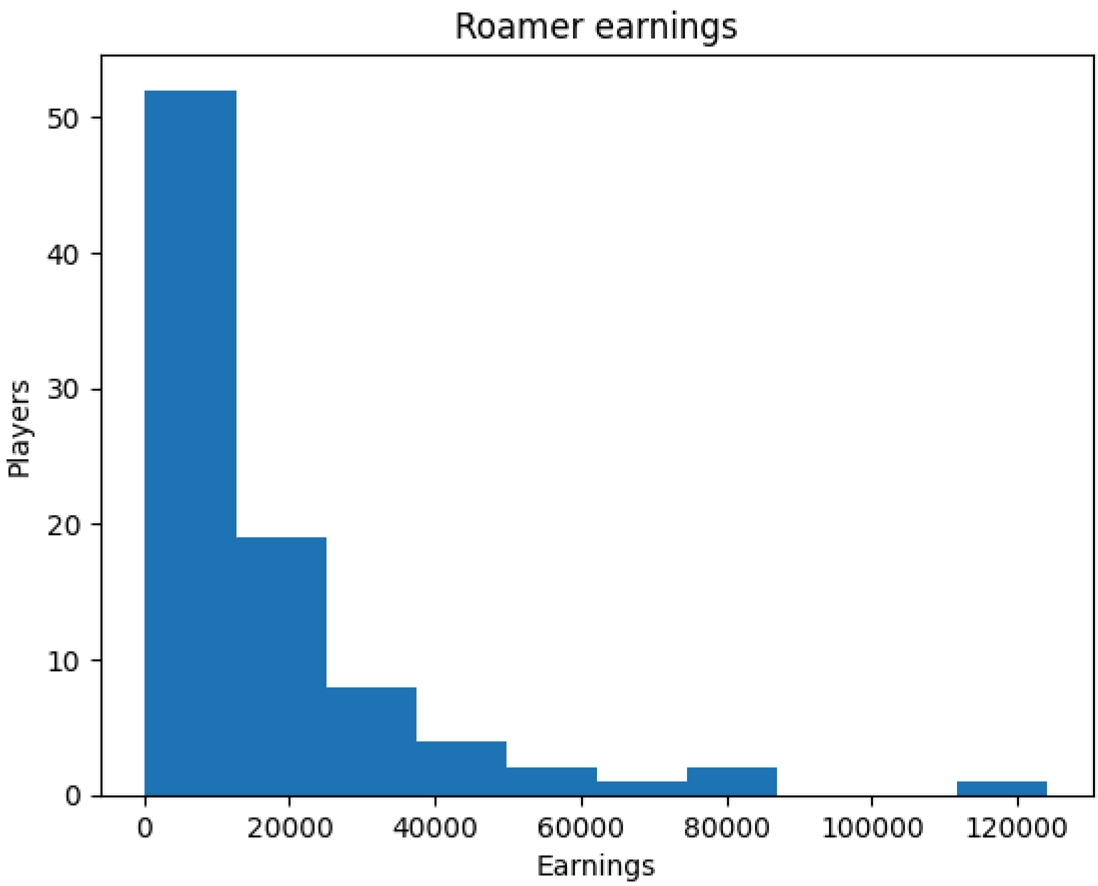

Fig. 13. Distribution of Roam players by earnings

Contributions of co-authors

We worked together both doing code and report, then checking ourselves.

References

[1]
https://www.igromania.ru/news/121946/League_of_Legends_stala_liderom_po_prosmotram_sredi_kibersportivnyh_igr_v_2022-m.html

[2]
https/www.sportskeeda.com/esports/5-esports-games-highest-viewership

[3]
https://www.kaggle.com/datasets/kishan9044/mobile-legends-professional-players