Astana IT University

Abdireshov Zhandos
Mussanova Anel

Designing and executing a cross-sectional study of gaming and mobility behavior: demographic, sociocultural and psychological factors.

6B06103 — Big Data Analysis

Diploma project

Supervisor
Aidana Zhalgas
Master of Science

Kazakhstan Republic
Nur-Sultan, 2023

ABSTRACT

Video games are interactive electronic entertainment experiences that involve player participation to control and manipulate visual images or virtual worlds displayed on a screen. Machine learning is a field of artificial intelligence where computers learn from data and improve their performance without being explicitly programmed. K-means clustering is an unsupervised learning algorithm that groups data points into clusters based on their similarity, aiming to minimize the intra-cluster variance. Hierarchical clustering is a method of grouping data points into nested clusters based on their similarity, creating a hierarchical structure or tree-like representation. A portrait of the player is a comprehensive and detailed representation of an individual who engages in playing video games. The main problem is that no surveys and studies have been conducted in Kazakhstan related to the field of video games, and socio-cultural, psychological and demographic factors of video games. To solve this problem, data on the gaming behavior, demographic characteristics, and social factors of players was collected through surveys, the impact of video and mobile games on various sectors was analyzed and a clustering machine learning system that utilizes this information to group players into clusters with similar characteristics was developed. The purpose of this work is to contribute to the development of personalized and effective grouping of players into clusters with similar characteristics, based on their psychological, demographic, cultural, and social factors.

Keywords: video games, k-means clustering, hierarchical clustering, machine learning, portrait of the player.

CONTENTS

DEFINITIONS

Following terms are used in this work:

Istagram	Instagram is a well-known social networking site where users may post photographs, videos, and stories with their followers. It includes filters, editing tools, and hashtags to help users enhance and organize the work. Users may follow other accounts, interact with posts by like and commenting on them, and find new stuff by using the explore function.
Telegram	Telegram is a cloud-based messaging platform that allows for private and secure communication. Users may send text messages, voice messages, multimedia files, and make audio and video conversations using it. Telegram also enables group conversations, channels, and bots, making it a flexible communication and information sharing tool.
Google Sheet	Google Sheets is Google's internet-based spreadsheet software. Users may create, modify, and share spreadsheets online. Formulas, data validation, conditional formatting, collaborative tools, and the ability to import and export data are all available in Google Sheets. It provides a simple and easy-to-use tool for data analysis and modification.
Excel	Excel is a spreadsheet tool created by Microsoft. It is commonly used for data organization, analysis, and manipulation. Excel has a variety of tools and capabilities for calculating data, visualizing data, and reporting. It enables users to build formulae, macros, pivot tables, and charts, making it an effective data analysis and management tool.
Pandas	Pandas is a popular Python module for data manipulation and analysis. It includes data structures like dataframes that make dealing with structured data easier. Pandas provides several functions and techniques for data cleansing, transformation, merging, aggregation, and other tasks. It is a must-have for data scientists and analysts who work with tabular data in Python.

Python	Python is a famous language for programming that is noted for its ease of use and readability. It is commonly utilized in industries including data analysis, machine learning, web development, and automation. Python has a robust ecosystem of libraries and frameworks, making it adaptable to a wide range of applications. Its simplicity of use and widespread community support have made it the language of choice for data analytic jobs.
LabelEncoder	LabelEncoder is a machine learning preprocessing approach that converts category variables to numerical values. It gives each unique category in a categorical variable a distinct number designation. LabelEncoder is widely employed when dealing with machine learning algorithms that need numerical input, since it allows category data to be encoded into a format that the algorithms can understand.
One-hot Encoder	One-hot Encoder is Another preprocessing approach used to transform categorical variables into a numerical format appropriate for machine learning algorithms. It generates binary columns in a categorical variable for each distinct category, where each column denotes the existence or absence of a given category. One-shot encoding assures that the encoded variables are independent of one another and prevents any ordinal link between the categories from being introduced.
K-Means	K-Means is a common clustering technique for grouping or clustering data based on similarity. It attempts to reduce within-cluster variation by allocating data points to the cluster with the closest mean. The number of clusters, represented by 'K,' must be provided ahead of time. K-Means is an iterative technique that alternates between clustering data points and updating cluster centroids until convergence is reached.
Grid Search	Grid Search is a method for determining the best hyperparameters for a machine learning model. It entails searching through a predetermined grid of hyperparameter variables exhaustively and assessing the model's performance via cross-validation. Grid Search identifies the optimum hyperparameter combination that produces the best

	performance measure, such as accuracy or mean squared error.
Standart Scaler	Standard Scaler is a preprocessing method for standardizing numerical information by removing the mean and dividing by the standard variation. It changes the data to have a mean of 0 and a deviation from the mean of 1, making it acceptable for algorithms that demand standardized input or presume normally distributed characteristics. Standard Scaler prevents bigger-scale features from dominating the model's process of learning.
Silhouette Score	The Silhouette Score is a statistic for assessing the quality of clustering results. The distances between data points inside and between clusters are used to calculate the compactness and separation of clusters. The Silhouette Score is a number between -1 and 1, with higher values suggesting more distinct clusters. A score of around 1 indicates well-separated clusters, whereas a score near -1 shows overlapping or wrongly allocated data points.
Hierarchical clustering	Hierarchical Clustering is a clustering technique that creates a hierarchy of clusters by merging or dividing clusters repeatedly based on their similarity. It does not require a predetermined number of clusters because it generates a dendrogram that depicts the hierarchical structure. Hierarchical Clustering can be conducted using a variety of approaches, including agglomerative (bottom-up) or divisive (top-down) clustering, as well as various distance or linking criteria.

DESIGNATIONS AND ABBREVIATIONS

Following designations and abbreviations are used in this work:

AITU	Astana IT University
KBTU	Kazakh-British Technical University
NU	Nazarvayev University
AUES	Gumarbek Daukeev Almaty University of Energy and Communications
ENU	L.N. Gumilyov Eurasian National University.
SDU	Suleyman Demirel University
NIS CBD	Nazarbayev Intellectual School of Chemical and Biological direction
DSML	Data Science and Machine Learning
Power BI	Power Business Intelligence
ML	Machine Learning

INTRODUCTION

Rapid technological improvements and the growing popularity of gaming have had a huge impact on many parts of people's life, including their mobility behavior. Understanding the interaction between gaming and mobility behavior is critical to understanding the possible consequences for individuals and society as a whole. This theoretical section seeks to give a thorough framework for developing and carrying out a cross-sectional study that analyzes the link between gaming and mobility behavior while taking demographic, sociocultural, and psychological aspects into account.

The relevance of this research. The computer industry market is rapidly developing, because not only the turnover and profit indicators are increasing annually, but also the number of consumers and their interest in gaming products is growing. However, and this is especially true in Kazakhstan, gaming is still considered a frivolous activity, comparable to children's games, and is even perceived by some researchers as a threat to the full development of adolescents. That is why this study is relevant since it comprehensively examines the product of the gaming industry, its financial indicators, and development trends. Moreover, the technologies used in the gaming industry and developed for it can be used not only by gamers but also by people of other diverse professions – doctors, psychologists, architects, filmmakers, etc. Also, no research on this topic was found in Kazakhstan, which makes thesis unique.

Goal of the work. The purpose of this work is to contribute to the development of personalized and effective grouping of players into clusters with similar characteristics, based on their psychological, demographic, cultural, and social factors. This can help game developers create more targeted marketing campaigns and create games that appeal to specific groups of players.

Object and method of research The main component of this study is to study the gaming habits and preferences in the movement of people from different demographic groups. The study attempts to gain a complete understanding of how these factors influence gaming activity and future mobility decisions by examining numerous aspects such as age, gender, socioeconomic level, cultural background, and psychological characteristics. An intersectoral research strategy will be used for the study, which allows collecting data at an exact moment in time and analyzing correlations between variables.

Theoretical and methodological base This study's theoretical underpinning is based on a knowledge of human behavior, especially in the context of gaming and mobility. To provide a framework for understanding the motives, social influences, and phases of behavior change connected to gaming and mobility choices, the study draws on ideas such as the Uses and Gratifications Theory, the Social Cognitive Theory, and the Transtheoretical Model of Behavior Change.

Research methods This diploma study is extremely important in terms of comprehending the relationship between gaming habit and mobility choices. With the increasing popularity of gaming and its possible impact on sedentary behavior, it is critical to investigate the factors that influence individuals' gaming preferences and how these choices transfer into mobility patterns. The findings of this study can be used to develop treatments and strategies to promote healthy gaming habits and more active forms of transportation. Furthermore, the construction of a clustering machine learning system can help to tailor gaming experiences and create the portrait of the player.

Objectives:

1 Collect data on the gaming behavior, demographic characteristics, and social factors of players through surveys and interviews.

2 Analyze the impact of video and mobile games on various sectors (demography, culture, sociology, and psychology).

3 Identify the critical demographic, psychological, cultural, and social factors that are most relevant to gaming behavior and preferences.

4 Develop a clustering machine learning system that utilizes this information to group players into clusters with similar characteristics.

5 Provide insights and recommendations for improving the accuracy and usefulness of the clustering system

1 LITERATURE REVIEW OF THE GAMING FACTORS

1.1 Analysis of gaming industry

With the advent of video games in the world, it was found that they affect several factors in relation to a person and society, especially in the cultural and social, and psychological spheres. In particular, with the rapid development of video games and the emergence of new genres, scientists began to hypothesize that games have an impact on the psychological and mental factors of a person, social relations, and cultural aspects. For example, Quan-Hung Vuong, etc. came to the conclusion that "Nowadays, video gaming is the favorite entertainment of both young and older people around the world". Also, they found that "Playing video games have been reported to improve the problem-solving skill of children or enhance the visual short-term memory of action video game players" [1].

However, other authors of a similar topic, like Christopher J. Ferguson, of a similar topic converse that the issue of whether video games—violent or nonviolent—"harm" children and adolescents continue to be hotly contested in the scientific community, among politicians, and in the general public. Despite the fact that this is one of the topics discussed, no primary research with surveys of these four factors has been done, and if there were, then there are few of them. One good example of this is the work of Mark D. Griffiths, Mark N.O. Davies and Darren Chappel, called "Demographic Factors and Playing Variables in Online Computer Gaming". They conducted a survey on the game "Everquest where they learned the sociodemographic data of the players who passed this survey. According to their results, 81% of the players were men, and their average age was 28 years [2]. One of the main factors of the game for them was social factors. In addition, a very small part of the players sacrificed important daily activities (sleep, study time, time with family/friends) in order to play games, they spent an average of 80 hours a week

1.2 Review of demographic factors in the gaming industry.

Age.

According to research, age has a significant impact on gaming choices and mobility behavior. Younger people, for example, participate in more gaming activities and may have different movement patterns than older people [3]. Furthermore, age-related elements such as life stage and developmental considerations might have an impact on the balance between gaming and physical activity.

Gender.

Previous research has found gender disparities in gaming behavior and mobility patterns. Males, for example, are frequently observed to spend more time gaming and engaging in various sorts of games than females [4]. Gender

differences in mobility behavior, such as preferred means of transportation and activity levels, may also exist.

Education.

A possible demographic component impacting gaming and mobility behavior has been found as education level. Higher levels of education may be connected with different gaming and mobility preferences, as well as a higher knowledge of the possible influence of sedentary behavior on health consequences.

Socioeconomic status.

Factors such as income and employment can have an impact on both gaming behavior and mobility choices. Individuals with lower socioeconomic levels may have restricted access to gaming resources or modes of transportation, which might influence their gaming and mobility behaviors.

1.3 Inquiry of cultural factors in the gaming industry.

Individuals' gaming habits and mobility choices are heavily influenced by culture. Cultural elements include common conventions, values, beliefs, and behaviors within a culture or group. Understanding the impact of culture on gaming and mobility behavior is critical for developing culturally relevant treatments and policies [5].

Cultural norms and attitudes.

Cultural norms and attitudes toward gaming might differ between nations and communities. Some cultures accept gaming as a popular form of entertainment and leisure activities, while others do not [6]. Cultural norms and attitudes might influence individuals' gaming habits and subsequent mobility behavior.

Gaming preferences.

Cultural elements, such as game genres, themes, and platforms, can influence people's gaming choices. For example, certain cultures may choose traditional or culturally distinct games, whilst others may favor popular worldwide gaming trends [7]. These preferences can influence how much time people spend gaming and potentially influence their mobility choices.

Social gaming and collectivist cultures. Social gaming may be more prominent in collectivist societies, where group cohesion and social relationships are highly prized. Interactions with people are involved in this type of gaming, such as multiplayer games or gaming within social networks. Social gaming can influence Individuals' mobility behavior by changing their social contacts and possibilities for physical exercise [8].

Gaming regulation and cultural context

Cultural norms and values may also have an impact on gaming regulation and laws within society. Some cultures may have stronger rules regarding gaming activities, while others may have more lax rules [9]. These legal frameworks can have an influence on the availability and accessibility of gaming platforms, which

in turn can have an impact on people's gaming habits and subsequent mobility behavior.

1.4 Investigation of psychological factors in the gaming industry.

Knowledge of the link between gaming and mobility behavior requires knowledge of psychological aspects. Individuals' psychological well-being, motives, and preferences can all have an impact on their participation in gaming activities and subsequent mobility decisions [10]. Examining these psychological aspects gives important insights into the underlying mechanisms underpinning the link between gaming and mobility behavior.

Psychological well-being.

Psychological well-being includes a variety of factors such as self-esteem, life happiness, and mental health. According to research, gaming may have both beneficial and bad impacts on psychological well-being. For example, gaming may provide enjoyment, stress reduction, and social connection, all of which can improve an individual's overall well-being [11]. Excessive gaming, on the other hand, might result in negative psychological effects such as increased tension, anxiety, and depression [12].

Motivation for gaming.

Understanding people's gaming motives is critical in understanding the link between gaming and mobility behavior. Different motivators, including enjoyment, social contact, competitiveness, skill improvement, and escapism, might influence how much time people spend gaming and, as a result, their physical activity participation or mobility choices.

Gaming preferences and intrinsic needs.

Individuals' gaming preferences and intrinsic requirements can also influence their mobility behavior. Individuals who are driven by discovery or accomplishment in gaming, for example, may choose outdoor activities or physically demanding types of mobility. Individuals who enjoy sedentary or relaxation-focused gaming experiences, on the other hand, may be more predisposed to sedentary activities or less physically demanding mobility options.

Gaming disorder and problematic gaming behaviors.

The idea of gaming disorder has gained prominence in recent years, underlining the potentially harmful impact of excessive or problematic gaming behaviors on individuals' psychological well-being and functioning [13]. Understanding the psychological aspects involved with gaming disorder can give insights into the mechanics behind the gaming-mobility association.

1.5 Scrutiny of social factors in the gaming industry.

Social considerations have a huge impact on people's gaming and mobility choices. The association between gaming and mobility behavior can be influenced

by social contacts, societal norms, peer influence, and social support networks [14]. Examining these social aspects reveals social dynamics that impact people's gaming habits and subsequent mobility decisions.

Social interactions and Gaming.

Gaming frequently involves social contacts, both offline and online, which might impact people's gaming habits and later mobility decisions. Multiplayer gaming, gaming groups, and social networks centered on gaming can foster social engagement and connection [15]. The social side of gaming can influence individuals' preferences, motives, and the amount of time they devote to gaming, impacting their participation in physical activities or mobility choices.

Social Norms and Peer Influence.

Social norms, which are described as shared expectations and values within a social group, can affect people's gaming behavior and mobility choices. Peer influence, especially among teenagers and young adults, has been shown to have a considerable impact on game choices and mobility patterns [16]. Individuals' decisions and actions might be influenced by societal approval or condemnation of their gaming habits and physical activity levels.

Social support and encouragement.

Individuals' gaming activity and subsequent mobility decisions might be influenced by social support networks such as family, friends, or gaming groups. Positive social support can motivate people to engage in physical activities, promote active gaming alternatives, or promote a healthy balance of gaming and physical activity [17]. Negative or unsupportive social settings, on the other hand, may lead to sedentary behavior or impede participation in physical activity.

Online gaming and social isolation.

While gaming can give opportunities for social connection, excessive or harmful gaming activity can lead to social isolation or disengagement from offline social activities [18]. The link between internet gaming and social isolation is complicated and varies by individual. It is critical to investigate the effects of excessive gaming on social interactions and subsequent mobility behavior.

2 Creating the portrait

2.1 Data preparation

In modern society, the gaming industry occupies an important place and is becoming an increasingly relevant topic for research. With the advent of new technologies, the availability of gaming platforms, and the widespread use of mobile devices, games have become an integral part of culture. They attract millions of people of different ages, genders, and social statuses, creating a huge potential for research [19].

However, despite the general interest in the gaming industry, research, especially in the context of Kazakhstan, remains relatively limited. The lack of research on this topic in Kazakhstan presents an opportunity for a thesis that will focus on the analysis of gaming behavior and its relationship with mobility in the context of demographic, socio-cultural, and psychological factors.

The accuracy of the data utilized can have a big impact on how well the text summarization model performs, therefore data preparation is essential. To extract the most crucial information from a huge body of text during text summarization, an algorithm or model is required. However, the model may have trouble locating the important details and producing a helpful summary if the text contains noise, mistakes, or unrelated information. As a result, data preparation aids in making sure the language is clear, organized, and pertinent to the work at hand. This includes trimming extraneous characters, fixing grammatical and spelling issues, and making sure the content is correctly arranged into sentences and paragraphs. Additionally, text must be transformed into a format that the summarization model can understand, such as numerical vectors or embeddings, as part of the data preparation process. As a result, the model is better able to comprehend the text's meaning and recognize its most crucial details. Data preparation is crucial for text summarizing since it helps to increase the relevance and quality of the text data, which eventually yields better summary outcomes. Several crucial phases are involved in data preparation for text summarizing, such as:

1 Data collection.
2 Data cleaning.
3 Text splitting.

In summary, the process of data preparation plays a vital role in text summarization as it guarantees the quality, relevance, and efficiency of the summarization model. By eliminating unnecessary information, standardizing the text, and establishing a cohesive and representative dataset, data preparation sets the foundation for accurate and meaningful summaries. It involves removing irrelevant noise, such as redundant or repetitive text, and transforming the data into a consistent format, ensuring that the summarization model can effectively extract the key information and generate concise summaries. Additionally, data

preparation enables the model to capture the essential elements of the original text, enhancing the overall summarization process and delivering more valuable insights.

2.1.1 Data collection

Data collection is the procedure of collecting pertinent data for research or study from multiple sources, including interviews, questionnaires, or databases [20].

In order to conduct this research, a survey was conducted in Google Forms among 350 people, regardless of whether these people played games or not. In order to collect answers from different audiences, several distribution methods were used, such as sending telegram groups to AITU, KBTU, NU, AUES, NARXOZ, ENU universities, school conversations of the Daryn school in Karaganda, NIS CBD Almaty, as well as secondary schools from different cities of Kazakhstan (Karaganda, Astana, Almaty, Pavlodar), audience coverage from Instagram Stories, telegram bot - mailing list of Astana IT University.

Collected data was collected through online surveys called Google Forms from different groups of people at universities (Astana IT University, Kazakh-British Technological University, Nazarbayev University, Eurasian National University, and students from abroad universities), students at schools (Nazarbayev Intellectual schools, Daryn school and different providing general education schools from other cities in Kazakhstan), employed people (TOO "Eurasian bank", "Code.me" company, TOO "Jysan Bank" workers etc.). The survey consists of 53 different open-ended, close-ended, multiple-choice questions, demographic and rating questions related to demographic, cultural, social, and psychological factors of the respondents regardless of whether they play video games or not. The methods of distributing the survey are actively forwarding messages to various schools, university conversations (3rd-course AITU), and groups of different IT communities (Zhana DSML, Almaty IT community, SDU community, NU community) in which the authors of this research are members. The survey was also actively distributed on the social network Instagram via Stories (@prcrstnation, @qapqara.bala). The survey consists of 5 parts: demographic, cultural, social, demographic, and game-based questions. The survey covered various aspects related to gaming, including demographics, culture, social preferences, and psychological factors. The demographic section included questions about the respondents' gender, age, place of birth, and education level. The cultural section focused on language preferences and communication with teammates, as well as players' preferences regarding teammates' nationalities and languages. The social section explored players' preferences for playing partners and their methods of finding games. The psychological section included questions about players' motivations for playing games, as well as their emotional

responses to winning or losing. Finally, the game-based questions covered players' preferences for game types, such as multiplayer or single-player, their preferences for playing as good, evil, or neutral characters, and their preferred game genres, such as action, arcade, or sports.

During the conversion of the data, there were unforeseen problems related to the language of respondents' responses, which would affect the data analysis and modeling. To cover more answers, questions, and answers were written in both Russian and English (separated by the "/"symbol).

Below are the questions that were asked of the respondents to answer:

1 What is your age? / Какой ваш возраст?
2 What is your gender? / Какой ваш пол?
3 What is your nationality? / Какая ваша национальность?
4 Do you study or work? / Вы учитесь или работаете?
5 What is your educational background? / Какое у вас образование?
6 What region of Kazakhstan were you born in? / В каком регионе Казахстана вы родились?
7 If you were not born in Kazakhstan, where exactly? (Write with a capital letter) / Если вы не родились в Казахстане, то где именно? (Напишите с заглавной буквы)
8 Have you ever played video games? / Вы когда-либо играли в видеоигры?
9 Do you think video games can be used to teach skills and knowledge? / Считаете ли вы, что видеоигры могут использоваться для обучения навыкам и знаниям?
10 If yes, what skills? / Если да, то какие навыки?
11 Do you think video games can be used to bring people from different cultures and backgrounds together? / Считаете ли вы, что видеоигры могут объединять людей разных культур и происхождения?
12 Have you ever attended a gaming convention or event? / Вы когда-нибудь посещали игровую конвенцию или мероприятие?
13 Do you belong to any clubs, sections, or university/school clubs? / Вы состоите в каких-либо клубах, секциях или университетских/школьных клубах?
14 Do you have trouble sleeping? / У вас есть проблемы со сном?
15 How often do you get sick? / Как часто вы болеете?
16 Is the nationality/religion/ethnicity of your teammates important to you? / Важна ли для вас национальность/религия/этническая принадлежность ваших товарищей по команде?
17 Do you play exclusively with people from your nation? / Вы играете исключительно с людьми из вашей страны?
18 What language do you play the video game in? / На каком языке вы играете в видеоигры?

19 In what language do you communicate with teammates? / На каком языке вы общаетесь с товарищами по команде?
20 Is there a language (or another) barrier that prevents you from playing certain games? / Есть ли языковое (или другое) препятствие, которое мешает вам играть в определенные игры?
21 If there is, what is the barrier? / Если да, то какое это препятствие?
22 In which games does this barrier occur? / В каких играх возникает это препятствие?
23 Do you prefer to play video games alone or with others? / Вы предпочитаете играть в видеоигры в одиночку или с другими людьми?
24 Do you play games more with friends, the community, or a family member? / Вы играете в игры больше с друзьями, сообществом или с членами семьи?
25 How do you communicate with your teammates during the game? / Как вы общаетесь с товарищами по команде во время игры?
26 What platforms or devices do you use to play games? / На каких платформах или устройствах вы играете в игры?
27 What types of video games do you play? / Какие виды видеоигр вы играете?
28 How do you discover new games (e.g., recommendations, social media, ads, etc.)? / Как вы находите новые игры (например, рекомендации, социальные медиа, реклама и т.д.)?
29 What factors influence your decision to buy or play a game (e.g., graphics, storyline, price, reviews, etc.)? / Какие факторы влияют на ваше решение купить или сыграть в игру (например, графика, сюжет, цена, отзывы и т.д.)?
30 Do games affect your relationships with friends/family? (If you play together) / Влияют ли игры на ваши отношения с друзьями/семьей? (Если вы играете вместе)
31 How do you think gaming affects your social and psychological well-being? / Как вы думаете, как влияют игры на ваше социальное и психологическое благополучие?
32 Do you feel closer to your teammate friends after the game? / Вы чувствуете близость к своим друзьям-товарищам после игры?
33 Do you discuss games with your loved ones? / Вы обсуждаете игры со своими близкими?
34 Are you a member of the gaming community? / Вы являетесь членом игрового сообщества?
35 What is the name of your community? / Как называется ваше сообщество?
36 What game is your community based in? / В какой игре основано ваше сообщество?
37 Why do you play games? / Почему вы играете в игры?

38 What motivates you to continue playing these games? / Что мотивирует вас продолжать играть в эти игры?
39 Do you play for the sake of competition? / Играете ли вы ради соревнования?
40 How do you feel when you play games? / Как вы себя чувствуете, когда играете в игры?
41 How do these emotions affect you while playing? / Как эти эмоции влияют на вас во время игры?
42 Were there any consequences after playing the game? / Были ли какие-либо последствия после игры?
43 Were these effects positive or negative? / Были ли эти эффекты положительными или отрицательными?
44 What were they like? Describe in detail. / Какие они были? Опишите подробно.
45 How do video games affect your mental health? / Как видеоигры влияют на ваше психическое здоровье?
46 How exactly? / Как именно?
47 I feel drained after playing the game. / Я чувствую усталость после игры.
48 How often do you play video games? / Как часто вы играете в видеоигры?
49 Do you play games on specific days or whenever you want? / Вы играете в игры в определенные дни или когда хотите?
50 What time of day do you usually play? / В какое время дня вы обычно играете?
51 What types of games do you prefer (e.g., action, puzzle, strategy, sports, etc.)? / Какие виды игр вы предпочитаете (например, экшн, головоломки, стратегии, спортивные и т.д.)?
52 Who do you like to play for? / За кого вы любите играть?
53 Why do you like to play for this or that character? / Почему вам нравится играть за этого или того персонажа?

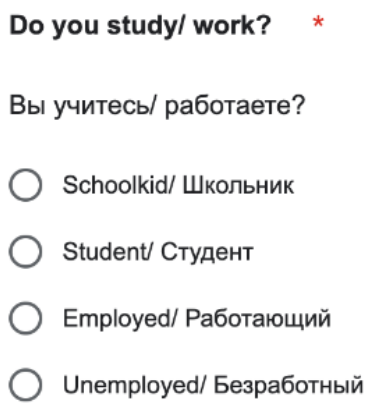


Figure 2.1 – Sample of question

2.1.2 Data cleaning in Excel

Data cleaning is the procedure of locating and fixing or eradicating mistakes, discrepancies, and errors from the obtained data to assure its accuracy and dependability [10].

Microsoft's Excel spreadsheet program is a part of the Office family of business software programs [21]. Users of Microsoft Excel may format, arrange, and compute data in a spreadsheet (fig.2.2). All the answers and data cleaning have already been initiated in Excel. The respondents answered the question in both Russian and English. To translate into English, Google Sheets functions as LEFT, FIND were used in fig.2.3.

	What is your gender? Каков Ваш пол?	What is your gender	What is your nationality? Какова Ваша национальность	what is your nationality	Do you study/ work? Вы учитесь/ работаете?	Do you study work?	What is your educational background? Каково Ваше образование?
2	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Incomplete higher professional education/ Не
3	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Employed/ Работающий	Employed	Average (full) general/ Среднее (полное) об
4	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Higher professional education/ Высшее проф
5	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Incomplete higher professional education/ Не
6	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Initial professional/ Начальное профессиона
7	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Incomplete higher professional education/ Не
8	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Employed/ Работающий	Employed	Higher professional education/ Высшее проф
9	Male / Мужской	Male	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Secondary vocational/ Среднее профессион
10	Male / Мужской	Male	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Initial professional/ Начальное профессиона
11	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Higher professional education/ Высшее проф
12	Male / Мужской	Male	Kazakh/ Казах(-шка)	Kazakh	Employed/ Работающий	Employed	Higher professional education/ Высшее проф
13	Male / Мужской	Male	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Average (full) general/ Среднее (полное) об
14	Male / Мужской	Male	Kazakh/ Казах(-шка)	Kazakh	Schoolkid/ Школьник	Schoolkid	Initial general/ Начальное общее
15	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Average (full) general/ Среднее (полное) об
16	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Employed/ Работающий	Employed	Higher professional education/ Высшее проф
17	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Incomplete higher professional education/ Не
18	Male / Мужской	Male	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Higher professional education/ Высшее проф
19	Male / Мужской	Male	Kazakh/ Казах(-шка)	Kazakh	Student/ Студент	Student	Average (full) general/ Среднее (полное) об
20	Female / Женский	Female	Kazakh/ Казах(-шка)	Kazakh	Schoolkid/ Школьник	Schoolkid	Basic General (incomplete secondary)/ Основ

Figure 2.2 – Raw dataset in Google Sheets

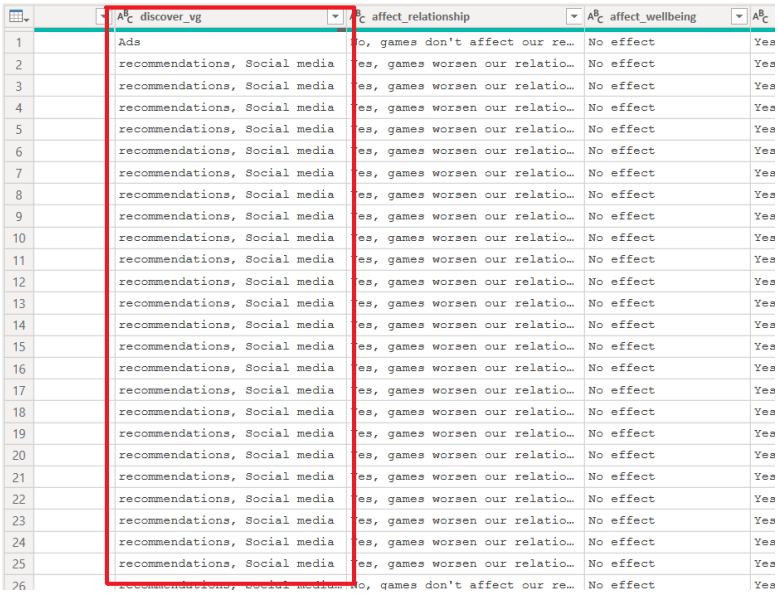
```
=ЛЕВСИМВ(C3;НАЙТИ("/";C3)-1)
```

Figure 2.3 – Formula in Google Sheets

And also manually shortened the answers to one or two words, for a more understandable presentation.

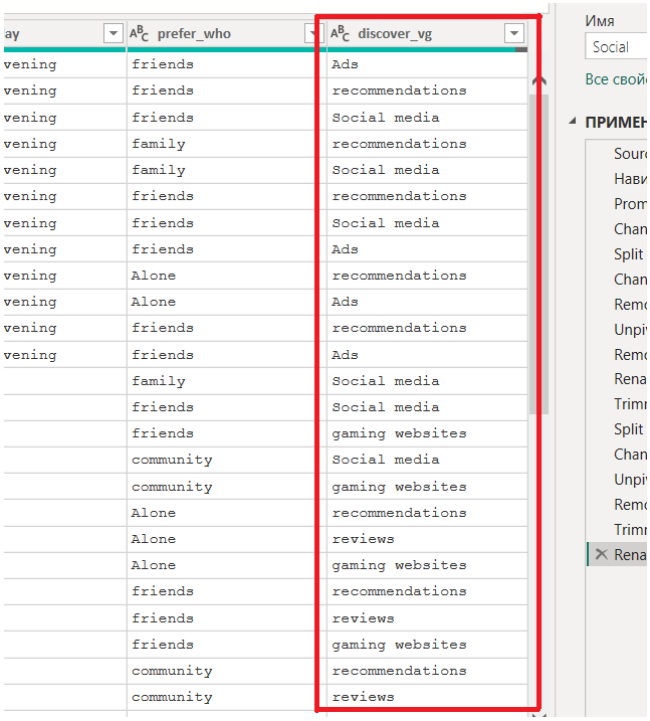
2.1.3 Text splitting

To display the visualization, the separation of all the answers to the multi-choice questions is performed using Text splitting. For example, the question "How do you discover new games?"In the figure 2.4 the answer options were Ads, Recommendations, Social Media, etc. Therefore, by utilizing the tools in Power BI, the separation of the answers was achieved after each "resulting in individual answers. Where can see in figure 2.5. A total of 7 columns underwent Text Splitting individually to separate the answers using the mentioned technique.



	discover_vg	affect_relationship	affect_wellbeing	
1	Ads	No, games don't affect our re…	No effect	Yes
2	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
3	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
4	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
5	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
6	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
7	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
8	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
9	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
10	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
11	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
12	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
13	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
14	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
15	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
16	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
17	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
18	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
19	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
20	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
21	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
22	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
23	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
24	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
25	recommendations, Social media	Yes, games worsen our relatio…	No effect	Yes
26	[illegible]	No, games don't affect our re…	No effect	Yes

Figure 2.4 – Before splitting



ay	prefer_who	discover_vg
vening	friends	Ads
vening	friends	recommendations
vening	friends	Social media
vening	family	recommendations
vening	family	Social media
vening	friends	recommendations
vening	friends	Social media
vening	friends	Ads
vening	Alone	recommendations
vening	Alone	Ads
vening	friends	recommendations
vening	friends	Ads
	family	Social media
	friends	Social media
	friends	gaming websites
	community	Social media
	community	gaming websites
	Alone	recommendations
	Alone	reviews
	Alone	gaming websites
	friends	recommendations
	friends	reviews
	friends	gaming websites
	community	recommendations
	community	reviews

Figure 2.5 – After splitting

2.1.4 Visualization

Power BI is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. Visualization in Power BI was used to analyze and understand the behavior of the data, and charts were divided into 5 sections: demography, culture, psychology, game, and society.

The image at the top shows the visualization in Power BI Desktop of the main columns in the field of Demography. Below are similar diagrams in different

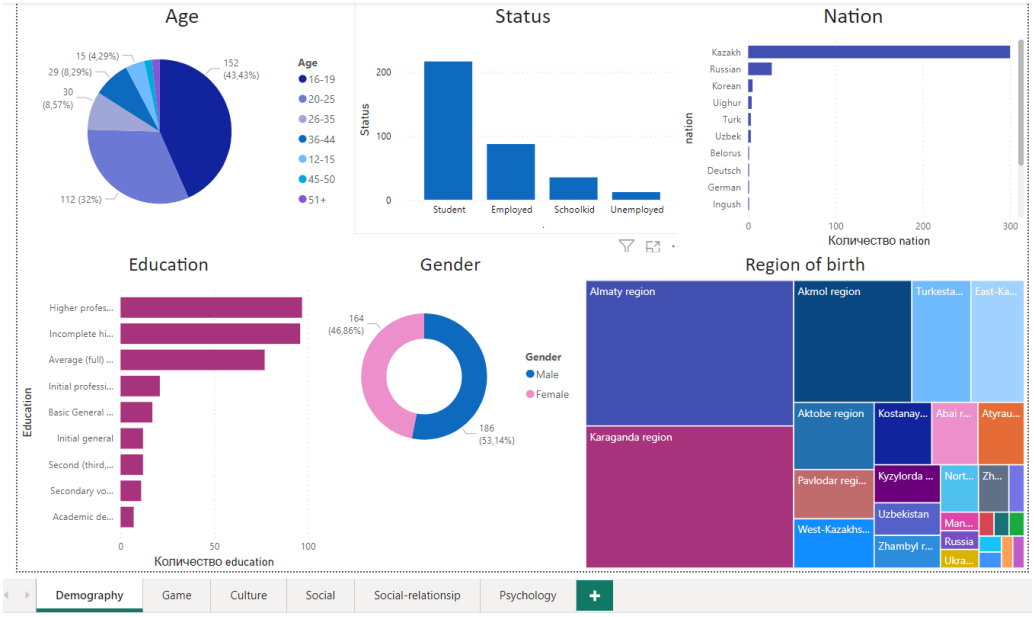


Figure 2.6 – Visualization in Power BI. Demography.

spheres (Culture, Social, Social-relationship, Psychology, Game).

Looking at visualization 2.6 in the field of demography, you can see that most of them are men (more than half), aged from 16 to 19 (43,4%), students, Kazakhs, were born in Almaty region and with higher professional education.

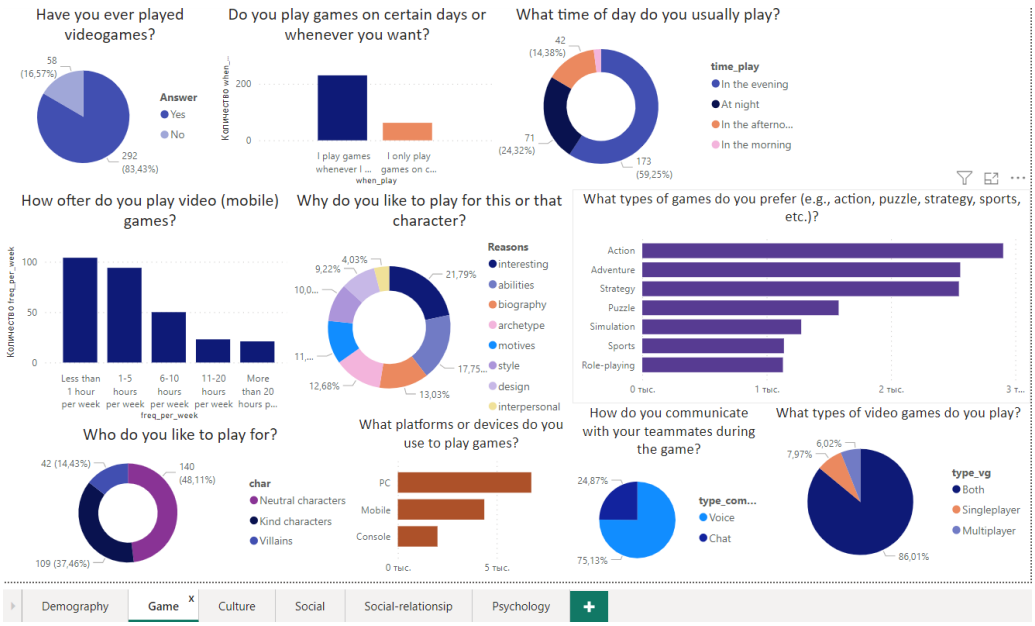


Figure 2.7 – Visualization in Power BI. Game.

In the gaming part of the data, 83.43% of respondents play games, which gives us a big advantage, 230 people from the respondents play when they want, without making a game schedule in the figure 2.7. More than half (59.25%) prefer to play

in the evening and the least in the morning (2.05%). 104 respondents prefer to play less than 1 hour a week, which makes the maximum amount. On multiple choice questions "Why do you like to play for this or that character?"21.79% of respondents chose interesting, 17.75% - abilities, 13.03% - biography. When choosing which genre of game people prefer, most have chosen Action, Adventure, and Strategy. Despite the fact that people choose many different genres, 85% of players like to play both multiplayer and single-player games. Also, when choosing a platform on which players play, a personal computer was in the first place, then a mobile phone and a console. 75% of the players choose to communicate with teammates via voice, while the remaining 25% via chat.

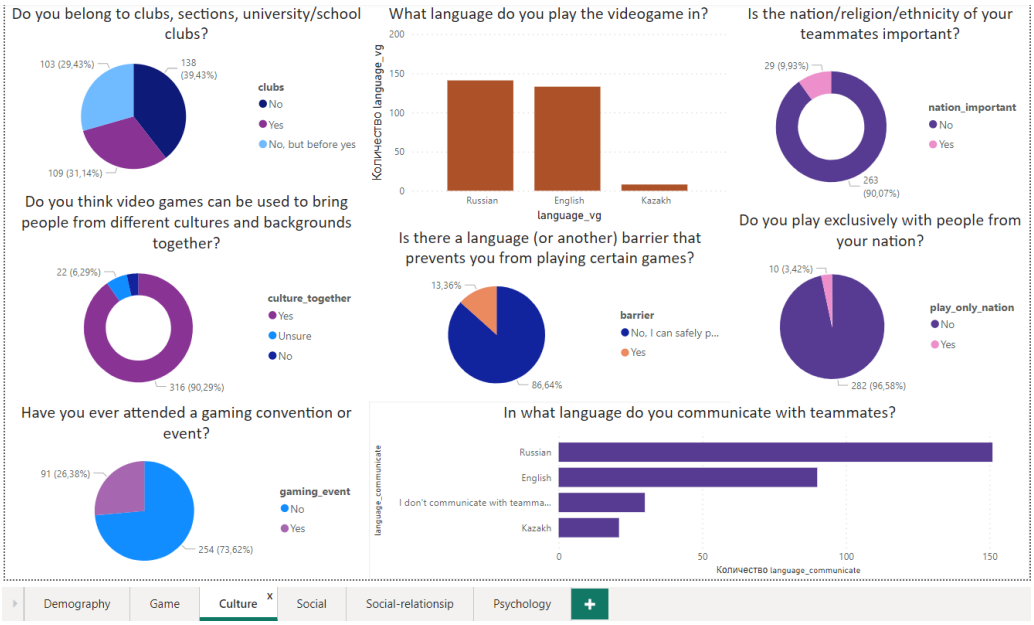


Figure 2.8 – Visualization in Power BI. Culture.

For the cultural part of the survey, 1/3 of people participated in school or university clubs, others 1/3 have not and others belong to clubs. 141 people prefer to play video games in the Russian language, 133 in English, and other 8 in Kazakh in the figure above 2.7. 90% of people think that teammates' nation is not important while playing multiplayer games, while 10% of respondents prefer to play only with people of their nation. Nevertheless, 90% of people think that video games can be used to bring people from different cultures and backgrounds together. Despite the fact that 290 people play games, only 26% of people have been to gaming conferences and events. For communication with teammates while gameplay, respondents speak with them in Russian, English, and Kazakh, with counts of 151, 90, and 21, respectively. The other 30 players do not communicate with their gaming partners.

In the social part of the survey (fig. 2.9 and fig. 2.10), for 68% of people games do not affect their social relationships, for 24% games make their relationships

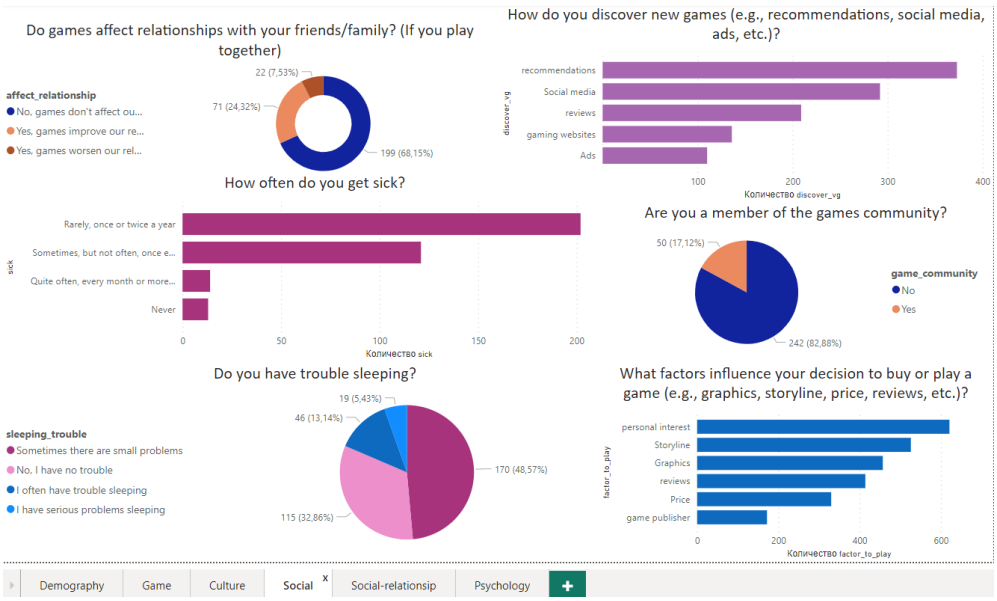


Figure 2.9 – Visualization in Power BI. Sociality.

Figure 2.10 – Visualization in Power BI. Social-relationship.

exceptional and for others, 7% video games worsen the relationships with their people. The health part of the players looks like this: 202 people get sick very rarely, 1-2 times a year, 121 people get sick not often, a couple of times a season, 14 people get sick almost every month and the remaining 13 people have never been sick. One-third of people have no problems with sleep at all, while half of people have problems with sleeping occasionally. 13% of people answered that they often do not sleep at night, and the remaining 5% of people have serious sleeping problems. Mostly, respondents prefer to play games with someone, especially with their friends or alone. Nevertheless, 71% of people feel closer to their teammates during the game.

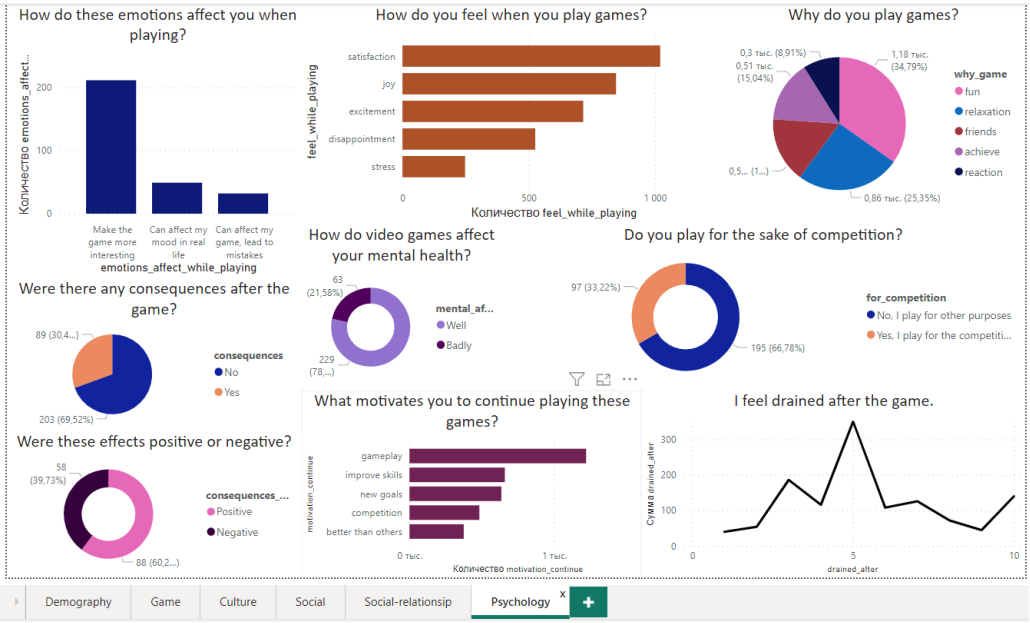


Figure 2.11 – Visualization in Power BI. Psychology.

Psychological factors of gamers are very positive (fig. 2.11). 211 respondents think that gaining emotions during the gameplay makes it more interesting and even gambling, and for 49 people games have influenced their mood in real life. The emotions that they feel are satisfaction, joy, and excitement. Moreover, the main motivation for playing the games are the same fun, relaxation, and friends with 34%, 25%, and 16%, respectively. For 69% of people after the games, there are no consequences, and for the other 31% consequences are mainly beneficial. Gameplay, improving cognitive skills, and achieving new goals are the main reasons why they continue to play games.

2.2 Data cleaning in Python

2.2.1 Deleting the values

Initially, the dataset consisted of 350 rows and 53 columns [2.12]. Since the audience of the study is only those people who have played or continue to play video games, the answers of people who do not play have been deleted using the ‘pandas’ library. Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. After deleting these values from the dataset, it decreased from 350 to 292 [2.13].

2.2.2 Merging the region column

There were 20 unique answers in the ’regions’ column, such as "Karaganda region "Almaty region"and other regions of Kazakhstan. All of them were divided and replaced into 5 main regions of Kazakhstan: Western, Eastern, Central, Southern and Northern [2.14].

```
dfm.shape
```

```
(350, 53)
```

Figure 2.12 – Initial shape of the dataset

```
played_vg_yes = dfm[dfm['played_vg']=='Yes']
dfm = dfm.loc[dfm['played_vg'].isin(played_vg_yes['played_vg'])]
```

```
dfm.shape
```

```
(292, 53)
```

Figure 2.13 – Shape of the dataset

The implementation of the replacement method is presented in Appendix A.

```
dfm['region'].value_counts()
```

```
South KZ          104
North KZ           63
Central KZ         59
West KZ            33
East KZ            20
Other country      13
Name: region, dtype: int64
```

Figure 2.14 – Regions of Kazakhstan

2.2.3 Converting the "age"column to numbers

In the "age"column, the data is written with a range, as 16-20, 21-25. Using the split and mean function of pandas library, the average data of the ages were taken [2.15].

The implementation of the split and mean method is presented in Appendix B.

2.3 Data preprocessing

2.3.1 LabelEncoder and One-hot Encoder

Further, using the LabelEncoder function from the sklearn.preprocessing library, most of the categorical data was transformed into numerical data, since for cluster machine learning, the data must be numeric. Label encoding is a

```
data['age'].value_counts()
```

```
18    141
22    105
30     19
14     13
40     12
63      2
Name: age, dtype: int64
```

Figure 2.15 – Age values

process of assigning numerical labels to categorical data values. For example, in the "gender"column, there are values such as Male/Female, and they were converted to 1/0 [2.16]. However, not all data was converted, since there are 8 questions (columns) where the respondent chose several options of the answers, because of this, the number of unique answers in these columns varies from 10 to 70. However, the number of unique selections varies from 4 to 7. Using this property of columns, the one-hot encoding function of the pandas library was used [2.17]. One-hot encoding is a popular technique for converting category variables into a binary form that machine learning algorithms may use. It is the transformation of categorical data into a binary vector representation. After this method, columns of the dataset are increased from 53 to 84 [2.18].

The implementation for data preprocessing techniques is provided in Appendix C.

```
Value counts for column 'gender':
1    182
0    110
Name: gender, dtype: int64
```

Figure 2.16 – Label encoder of "gender"column

char_desc_archetype	char_desc_biography	char_desc_design	char_desc_interesting	char_desc_interpersonal	char_desc_motives	char_desc_style
0	0	0	0	0	0	0
1	1	0	1	0	1	0
0	0	0	1	0	0	0
0	0	0	0	0	1	1
1	0	0	1	0	0	0

Figure 2.17 – Results of One-hot encoding

```
data.shape
```

```
(292, 84)
```

Figure 2.18 – Shape of the dataset after data preprocessing

3 RESULTS

3.1 Application of machine learning

3.1.1 Dividing dataset

The next step to build a model is to divide the dataset into 5 parts, since there are too many columns: demography, society, game, psychology and culture. Based on the questions in the survey, the dataset was divided and 2 cluster machine learning (ML) was conducted for each of them: K-means and Hierarchical clustering.

K-means is a machine learning and unsupervised learning clustering technique. It is used to divide a dataset into discrete groups or clusters based on data point similarity.

Hierarchical clustering is an unsupervised clustering technique that groups related data points into nested clusters. It generates a hierarchical structure of clusters, which is frequently depicted as a dendrogram, with the top level clusters being more generic and the bottom level clusters being more particular.

3.2 Grid Search, Standard Scaler and Silhouette score

To standardize the features of the dataset, the StandardScaler method of scikit-learn library was used. Standardization is a standard preprocessing technique that removes the mean and scales the data to unit variance to ensure that all features have the same scale. StandardScaler() is a scikit-learn (a popular machine learning toolkit) class that is used to standardize or scale numerical values in a dataset.

In order to improve the results of machine learning (ML), Grid Search techniques were applied for the most appropriate number of clusters. To do this, from 2 to 11 clusters were sent to the function, and with the help of the highest Silhouette Score from -1 to 1, in Grid Search, the most appropriate number of clusters was identified for K-Means algorithm.

Grid Search is a machine learning hyperparameter tuning approach. Hyperparameters are model configuration parameters that are not learned from data but are determined prior to training the model. Silhouette score is a metric used in unsupervised learning to assess the quality of clustering. It assesses how well samples within a cluster are similar to samples from other clusters.

For another clustering algorithm, hierarchical clustering, to the grid search function were added another hyperparameter - linkage. Linkage is the method used to measure the distance between clusters, by determining how the clusters are merged during the process of hierarchical clustering. Grid Search presented that the 'ward' method of linkage hyperparameter is the best option. Ward method minimizes the variance within clusters. After the grid search and implementing machine learning (ML) algorithms, the silhouette score shows the best algorithm

	Demography	Social	Game	Psychology	Cultural
K-Means	0.3	0.07	0.08	0.1	0.22
Hierarchical clustering	0.28	0.05	0.02	0.06	0.21

Table 3.1 – Silhouette score of the algorithms

for each dataset. The obtained results are provided in Table 3.1

The implementation of models is provided in Appendix D.

As interpreted in Table 3.1, for each dataset K-Means shows the best results than Hierarchical clustering. Nevertheless, values are similar to each other. However, the results of both algorithms will be shown for comparison in Table 3.2 for demography portrait, Table 3.3 for social portrait, Table 3.4 for game portrait, Table 3.5 for psychological portrait and Table 3.6 for cultural portrait. For each dataset, to understand the behavior of the features among themselves, a correlation was carried out. The results of the correlation analysis show [17-21] that not all columns are connected to each other, but nevertheless, in order to understand the full picture of the player's portrait by all factors, none of the columns were removed.

Correlation analysis is a statistical approach for determining the strength and direction of a relationship between two or more features. It helps in determining whether and show variables are related to each other.

	K-Means	Hierarchical clustering
age	22	22
gender	Male	Male
nation	Kazakh	Kazakh
status	Student	Student
education	Higher professional education	Higher professional education
region	South KZ	South KZ

Table 3.2 – Demography portrait of the player

	age	gender	nation	who	education	region
age	1.000000	-0.027297	0.239280	-0.388937	0.116157	-0.209865
gender	-0.027297	1.000000	0.062344	0.133593	-0.002508	0.137532
nation	0.239280	0.062344	1.000000	-0.097060	-0.024735	-0.157315
who	-0.388937	0.133593	-0.097060	1.000000	0.025817	0.158870
education	0.116157	-0.002508	-0.024735	0.025817	1.000000	-0.021503
region	-0.209865	0.137532	-0.157315	0.158870	-0.021503	1.000000

Figure 3.1 – Correlation matrix of demography dataset

	Hierarchical clustering	K-Means
belong to clubs	no, but before yes	no, but before yes
communication of language in games	Russian	Russian
gaming affect to relationship	No	No
feel closer to teammates	yes	yes
discuss the videogame	yes	yes
belong to gaming communities	no	no
sleeping trouble	no	no
how often they get sick	Rarely, once or twice a year	Rarely, once or twice a year
language of videogames	Russian	Russian
with who they prefer to play	Alone/Friends	Friends
how often they play games	6-10 hours per week	6-10 hours per week
when they play games	on certain days	whenever they want
at what part of the day they play game	in the afternoon	in the evening
how they discover videogames	recommendations/social media	recommendations/social media

Table 3.3 – Social portrait of the player

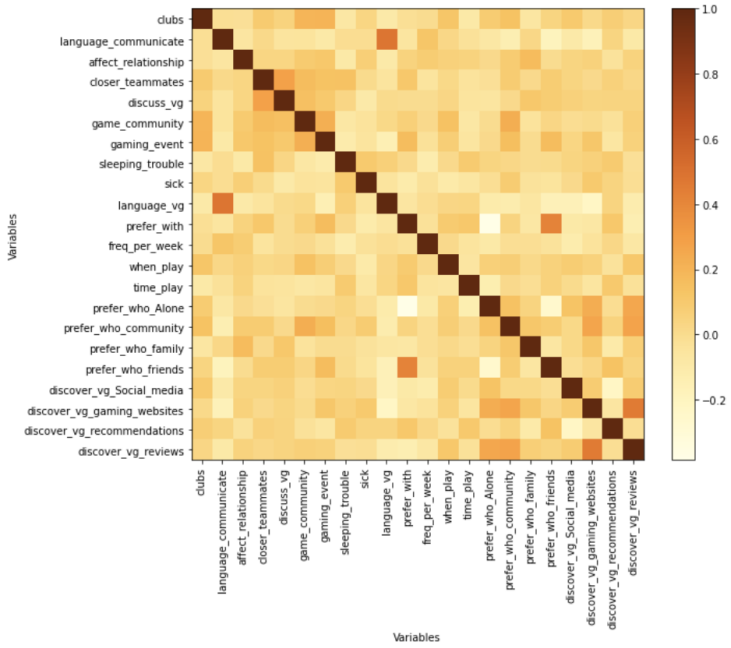


Figure 3.2 – Correlation matrix of social dataset

	Hierarchical clustering	K-Means
type of videogame	Multiplayer	multiplayer and singleplayer
genre	Action, Adventure	Action
platfrom	PC	PC. mobile
factor to play	personal interest, graphic storyline	graphics
character in videogame	neutral	neutral
why they choose the character	interesting style of game	interesting style of game

Table 3.4 – Portrait of the player by gaming preferences

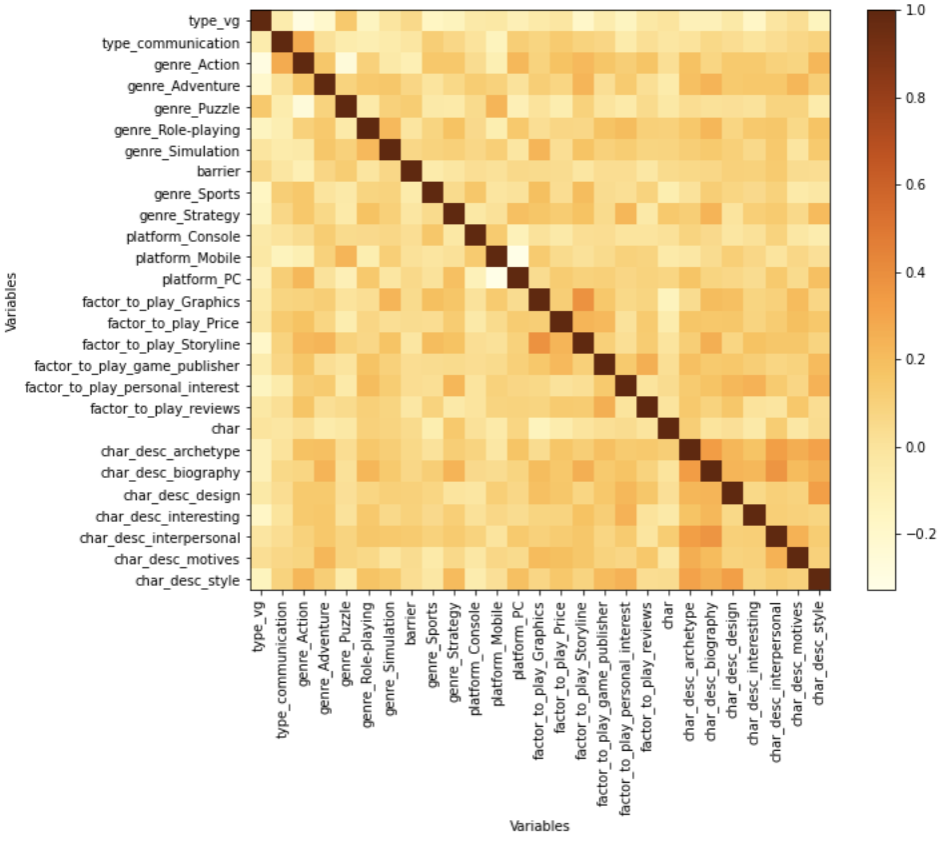


Figure 3.3 – Correlation matrix of game dataset

	Hierarchical clustering	K-Means
how games affect to wellbeing	positively	positively
do they play for competition	no	no
how emotions affect while playing	make the game more interesting	make the game more interesting
are there any consequences	no	no
do video games affect to their mental health	Well	Well
how much they feel themselves drained after the game	5	4
why they play games	for fun and relaxation	for fun and relaxation
what is the motivation to continue to play the games	gameplay is interesting	gameplay is interesting
what they feel while playing	satisfaction and joy	satisfaction

Table 3.5 – Psychological portrait of the player

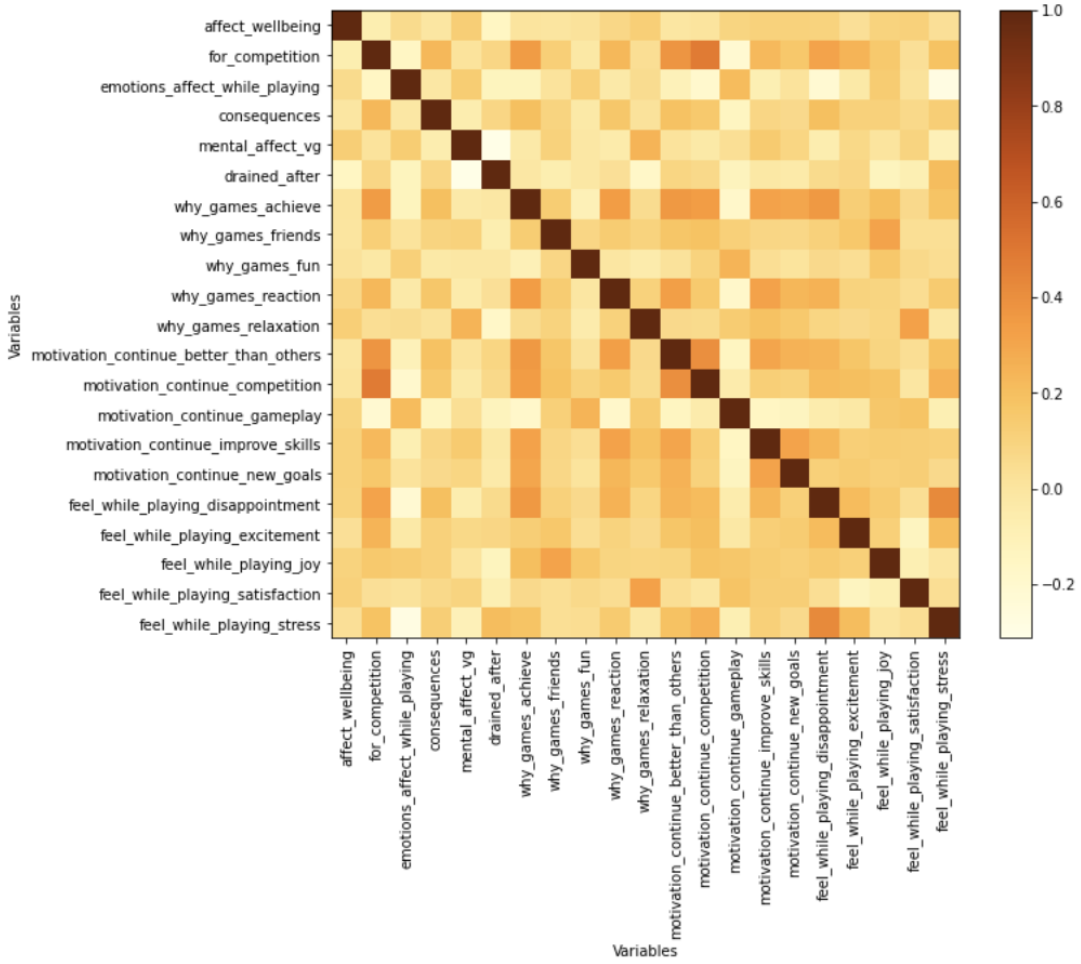


Figure 3.4 – Correlation matrix of psychology dataset

	Hierarchical clustering	K-Means
nation	Kazakh	Kazakh
region	North KZ	North KZ
game can bring cultures together	yes	yes
have they attended gaming events	no	no
do they belong to clubs	No, but before yes	No, but before yes
nation is important while gaming	no	no
play only with teammates from their nation	no	no
language of videogame	Russian	Russian
language of communication	Kazakh, Russian	Kazakh

Table 3.6 – Cultural portrait of the player

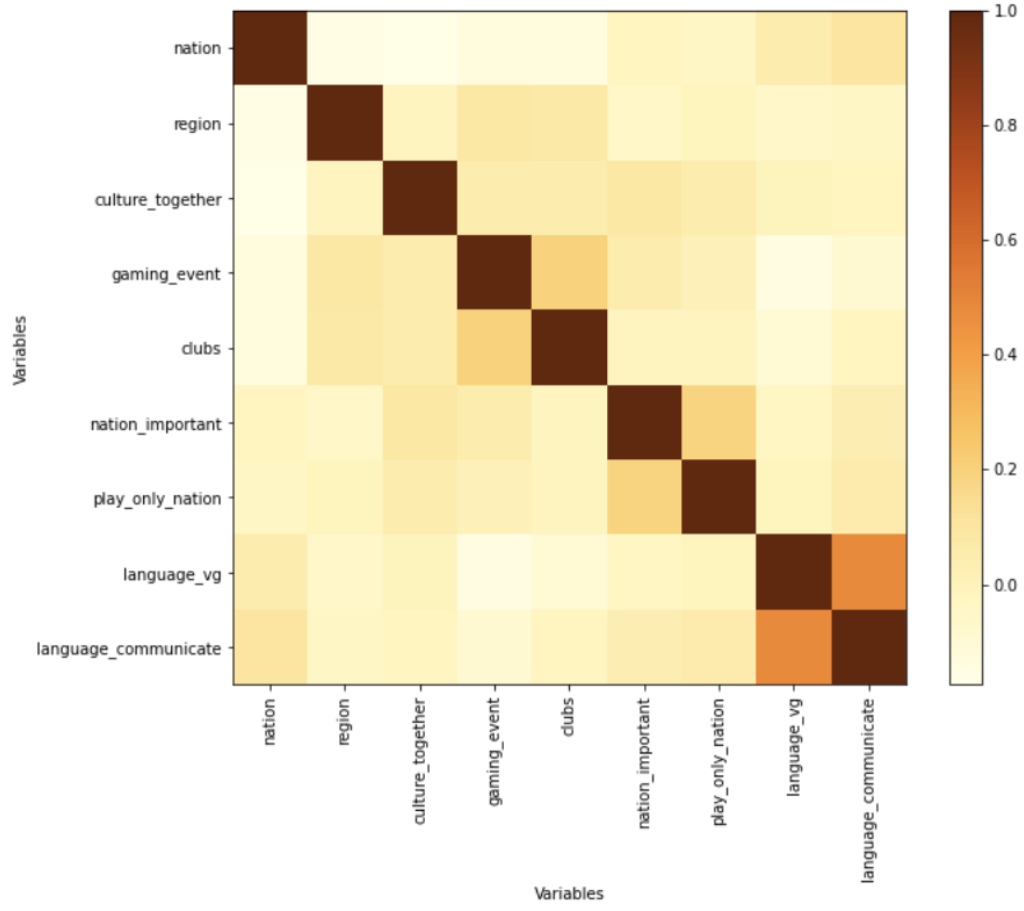


Figure 3.5 – Correlation matrix of culture dataset

Results of the machine learning (ML) algorithms are as it were expected, but it is not appropriate to understand the portrait of the player according to these results. During the building of machine learning (ML) algorithms, it was revealed that the features of datasets that are not related to each other could have a strong connection, for example, how the frequency of spending time playing the game affects mental health. Several features that have connections between each other were found in the same way through correlation analysis, and 3 separate datasets were created.

3.3 Related features

3.3.1 Mental health

After clustering the players separately by these 4 factors, it was decided to do additional clustering in order to understand the connections between the data. Since video games are a very discussed and researched area in psychology, it was decided to look at the behavior of columns in psycho-social factors. According to Dr. Daniel Johnson's research, positive mental wellbeing has been associated with video game play as a means of relaxation and stress reduction, and frequency of play does not significantly relate to body mass index or mental health. After the discussions, researches and results of the last clustering algorithms, it was decided that it would be appropriate to see whether the number of hours spent on video games affect a person's mental health and other features. To do this, the same two models in the previous datasets were used, as well as the correlation between the columns [22]. To understand the factors that could affect to the mental health of the player, next features were chosen:'age', 'sick', 'sleeping_trouble', 'affect_wellbeing', 'emotions_affect_while_playing', 'mental_affect_vg', 'freq_per_week', 'when_play', 'time_play', 'feel_while_playing_disappointment', 'feel_while_playing_excitement', 'feel_while_playing_joy', 'feel_while_playing_satisfaction', 'feel_while_playing_stress'.

Silhouette scores for each dataset are very different. If K-means before showed the best results than Hierarchical clustering, then for this dataset second algorithm's score is 0.06, while K-Means shows -0.06, that's why the results of the clustering is based on the output of the Hierarchical clustering.

Players of the survey rarely get sick, once or twice a year, and also do not have any sleeping troubles. Video games positively affect their wellbeing, their emotions make the game more interesting. The key result is that gaming positively affects their mental health, while they play games 6-10 hours per week on certain days. There is one cluster where people play games 11-20 hours per week, although they do not have any mental problems. Moreover, all clusters show that people play games in the afternoon, and only 1 cluster shows that people play games in the evening, which is normal, because as it is mentioned before, people do not have

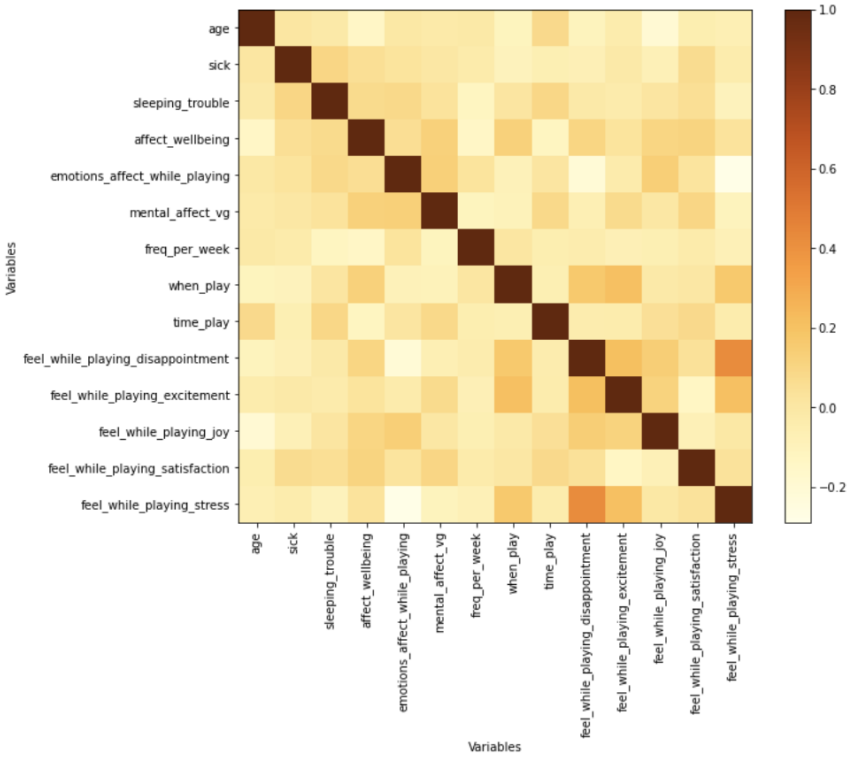


Figure 3.6 – Correlation matrix of mental dataset

any sleeping troubles. Also, all clusters display that people feel joy and satisfaction while playing the games, and there are no disappointments or stress during the game. Results of this dataset, mental health, are the same as the research of Dr. Daniel Johnson: “There are many creative, social and emotional benefits from playing video games, including violent games.”

3.3.2 Community

Video games represent a distinct world, often referred to as a "virtual life,"which parallels the real world. Specifically, the social factors that influence players in virtual life can be as in reality. In order to prove this hypothesis, a dedicated dataset, “community”, was collected to seizure the socio-cultural factors of players through correlation analysis [23] and two clustering machine learning algorithms.

The features for this dataset are: ’type_communication’, ’language_communicate’, ’prefer_with’, ’closer_teammates’, ’discuss_vg’, ’game_community’, ’clubs’, ’why_games_friends’, ’gaming_event’. Same two machine learning (ML) algorithms, K-Means and Hierarchical, were used to capture the output of the clustering for this dataset.

K-Means algorithm’s silhouette score is higher than the Hierarchical clustering, 0.8 and 0.6, respectively. Therefore, the portrait of the socio-cultural factors of the player is based on K-Means algorithm. Results for this portrait are next: despite the fact that gamers prefer to play alone and with someone, they play

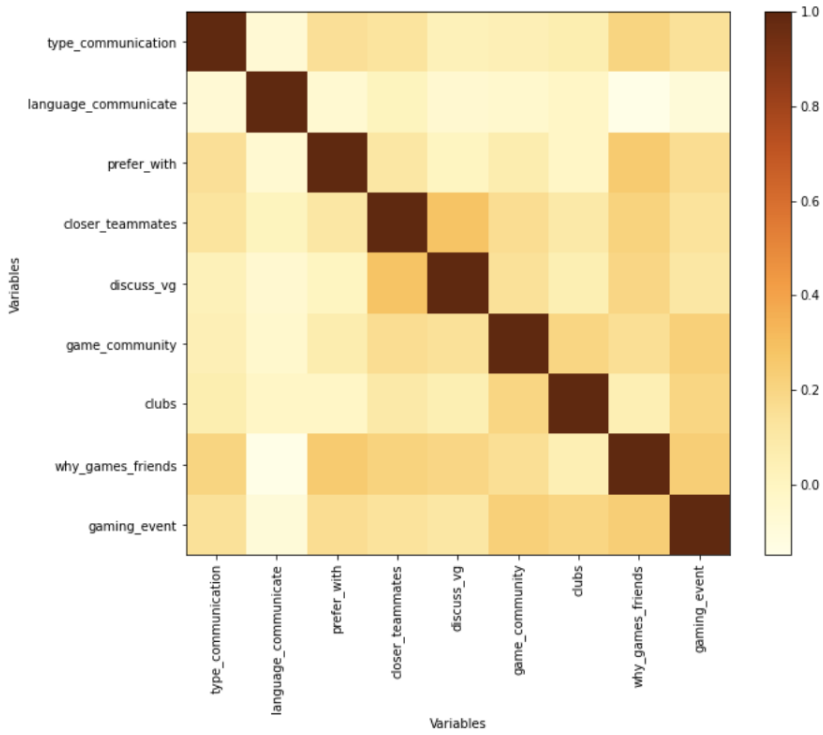


Figure 3.7 – Correlation matrix of community dataset

more multiplayer games and communicate in games via voice and in Russian language. After the game they discuss the events and process of gaming with teammates, and thanks to this, they feel closer to them, neither teammates are strange, family members or friends. However, friends are not the main reason why they play games, they rather play for fun and relaxation. Gamers were once in student or school clubs, but now they are not, also do not belong to gaming communities or have not attended any gaming events.

3.3.3 Gaming

In order to better understand the social factors and the overall game market for answered players, also through correlation analysis [24] and clustering machine learning(ML) techniques, several features were selected separately. The features are: 'type_vg', 'genre', 'prefer', 'platform', 'discover', 'factor_to_play'.

For this dataset, silhouette score for K-Means is 0.75, and for Hierarchical clustering is 0.25, so the output is based on K-Means algorithm.

Player's prefer both types of games: multiplayer and singleplayer, and to be more precise, the genre of video games is action. Even if the type of game is singleplayer or multiplayer, gamers prefer to play with friends on mobile or personal computer (PC). Video game discovery is frequently done by social media platforms and the effect of recommendations from friends or family. The decision to play video games is influenced by a variety of factors, including personal interest and the graphics.

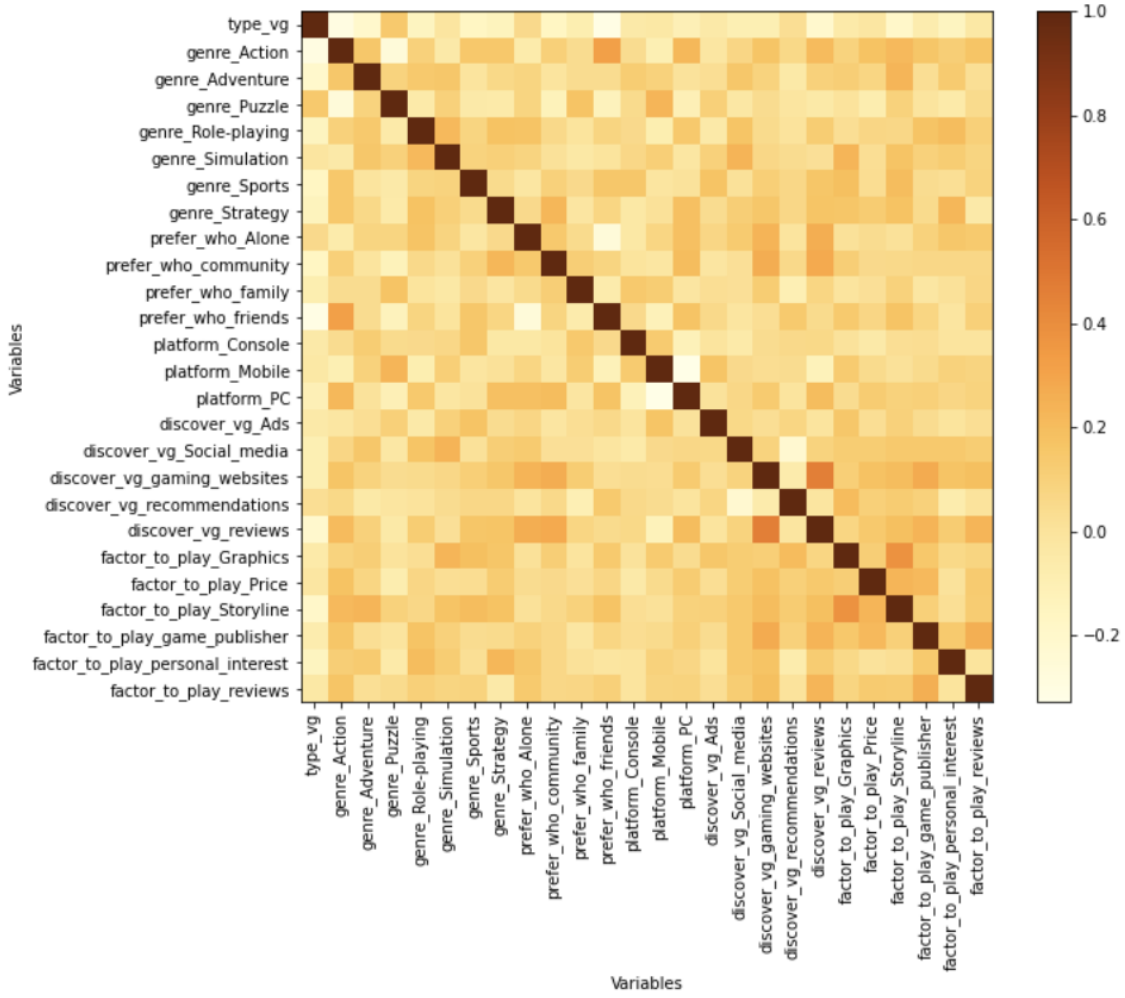


Figure 3.8 – Correlation matrix of gaming dataset

CONCLUSION

All the goals of researching gaming and mobility behavior in several factors were achieved. The image of players has been designed by developing personalized and effective clusters with similar characteristics, based on their psychological, demographic, cultural, and social factors.

1 Firstly, a literary review was conducted on the factors of video games that affect demographics, socio-cultural and psychological factors. In depth, separate studies were conducted for each 4 factors and their relationship to each other. Based on the work of other authors, questions were created for the survey that include the same main 4 factors: demography, social, psychology and culture.

2 Secondly, after the final list of 53 questions, a survey was created, which was forwarded using social networks, IT communities in messengers, PR and universities. As a result, 350 people with different backgrounds took the survey and a dataset was created on which analysis and clustering was carried out to create a portrait of the player.

3 Thirdly, using tools for data analysis and visualization and for cluster machine learning, the main and detailed demographic, social, cultural and psychological factors of video games were identified. To create a portrait of the player, two types of clustering were applied, K-Means and Hierarchical, for each factor. In most cases, by the Silhouette score, K-Means shows the best result of the clustering, but nevertheless, the results of each algorithm were compared.

4 In the end, according to the results of the literature review, the relationship and impact of the factors between each other were conducted. Despite the different demographic elements such as age, gender, region, education, games have positive impact on the physical and mental health, social and cultural aspects of life and all theories that video games are the main reason for all gloomy situations and problems are untrue.

Throughout the process of research, some difficulties and challenges have been encountered.

1 Data collection. Due to the big number of questions which are related to the big 4 factors, it was hard to collect the appropriate and necessary data. There has never been a survey done on the topic of video games in Kazakhstan, so for the first time, it was quite difficult to find the ground to start. However, with the help of the Astana IT University's mailing bot in Telegram and school communities, the number of respondents increased to 350.

2 Data cleaning. Due to the fact that this survey was conducted in both English and Russian, it was difficult to clean the data and leave only the answers in English. In addition, there were 8 open-ended questions and 7 multiple-choice questions in the survey. This led to the decision that the open answers were

cleaned by hand, and the answers with multiple options were divided through one-hot encoding.

Several recommendations might be made to improve this study work further:

1 Interview. Besides the survey to collect the data and research this topic, there is one more option to understand the background of video games: interview. With the help of the interview of 15-20 different demographic, psychological, and socio-cultural backgrounds, it is possible to understand the factors of video games even more deeply, because the answers of respondents will be honest and truthful and it will be clear to see the relationship of the factors between each other.

2 Different range of respondents. The target audience of this survey is people in Kazakhstan, but in the future, with the help of the different worldwide communities, entertainment organizations of different games, and esports players, it is possible to gain more data from several countries and create a portrait of the player for each type of games or popular video games. This could bring valuable results to be used by game publishers.

To sum up, this research paper provides new insights and valuable results to understand the demographic, socio-cultural, and psychological factors of video games. The results of the work are beneficial not only for humanitarian scientists for further exploration but also for potential Kazakhstani game publishers.

BIBLIOGRAPHY

1 *Vuong, Q.H. et al.* A multinational data set of game players' behaviors in a virtual world and environmental perceptions / Q.H. et al. Vuong // *Data Intelligence*. — 2021. — Vol. 3, no. 4. — Pp. 606–630. https://doi.org/10.1162/dint$_{a0}$0111.

2 *Griffiths, Mark D.* Demographic Factors and Playing Variables in Online Computer Gaming / Mark D. Griffiths, Mark N.O. Davies, Darren Chappell // *CyberPsychology & Behavior*. — 2004. — Aug. — Pp. 479–487. http://doi.org/10.1089/cpb.2004.7.479.

3 *Anderson, C. A.* Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life / C. A. Anderson, K. E. Dill // *Journal of Personality and Social Psychology*. — 2000. — Vol. 78, no. 4. — Pp. 772–790.

4 *Charlton, J. P. et al.* Internet addiction: A brief summary of research and practice / J. P. et al. Charlton // *Current Psychiatry Reviews*. — 2002. — Vol. 1, no. 1. — Pp. 59–69.

5 *Ferguson, C. J.* Video games and youth violence: A prospective analysis in adolescents / C. J. Ferguson // *Journal of Youth and Adolescence*. — 2017. — Vol. 46, no. 1. — Pp. 204–217.

6 *Huynh, Q. L. et al.* A systematic review of cultural perspectives on video gaming: Implications for interventions and research / Q. L. et al. Huynh // *Frontiers in Psychology*. — 2020. — Vol. 11. — P. 560023.

7 *Li, J.* Understanding cross-cultural play: A review of gaming research / J. Li, M. Liu // *Frontiers in Psychology*. — 2019. — Vol. 10. — P. 88.

8 *Poels, K. et al.* The influence of culture on digital gaming preferences and behavior / K. et al. Poels // *International Journal of Communication*. — 2018. — Vol. 12. — Pp. 3133–3152.

9 *Charlton, J. P.* Distinguishing addiction and high engagement in the context of online game playing / J. P. Charlton, I. D. W. Danforth // *Computers in Human Behavior*. — 2007. — Vol. 23, no. 3. — Pp. 1531–1548.

10 *Ferguson, C. J.* Understanding digital playability and its effects on gaming behavior / C. J. Ferguson, J. Colwell // *Journal of Media Psychology*. — 2017. — Vol. 29, no. 1. — Pp. 1–8.

11 *Griffiths, M. D. et al.* A systematic review of empirical research on Internet addiction in online gaming / M. D. et al. Griffiths // *Journal of Behavioral Addictions*. — 2016. — Vol. 5, no. 3. — Pp. 561–566.

12 *Kuss, D. J. et al.* Gaming addiction and its association with personality traits, psychopathology, and comorbidity / D. J. et al. Kuss // *International Journal of Mental Health and Addiction*. — 2018. — Vol. 16, no. 4. — Pp. 905–918.

13 *Ryan, R. M. et al.* Motivational patterns in online gaming: A self-determination theory perspective / R. M. et al. Ryan // *Motivation and Emotion*. — 2006. — Vol. 30, no. 4. — Pp. 347–364.

14 *Chan, A. S. et al.* Online gaming addiction in children and adolescents: A review of empirical research / A. S. et al. Chan // *Journal of Behavioral Addictions*. — 2016. — Vol. 5, no. 4. — Pp. 518–528.

15 *Griffiths, M. D. et al.* The role of online gaming and social capital in adolescents and adults: A review / M. D. et al. Griffiths // *International Journal of Mental Health and Addiction*. — 2019. — Vol. 17, no. 4. — Pp. 954–978.

16 *Hussain, Z. et al.* The impact of online multiplayer video games on social interactions: An overview / Z. et al. Hussain // *Computers in Human Behavior*. — 2019. — Vol. 92. — Pp. 246–257.

17 *Przybylski, A. K. et al.* Motivation for play in online games / A. K. et al. Przybylski // *Cyberpsychology, Behavior, and Social Networking*. — 2010. — Vol. 13, no. 6. — Pp. 771–775.

18 *Wang, C. K. et al.* The relationship between gaming disorder and physical health outcomes / C. K. et al. Wang // *Mental Health and Physical Activity*. — 2017. — Vol. 13. — Pp. 141–149.

19 *López-Fernández, O.* Internet gaming disorder in adolescence: A systematic review / O. López-Fernández, A. J. Williams, M. D. Griffiths // *Adolescent Research Review*. — 2020. — Vol. 5, no. 2. — Pp. 153–162.

20 Prevalence of Internet gaming disorder in German adolescents: diagnostic contribution of the nine DSM-5 criteria in a state-wide representative sample / F. Rehbein, S. Klim, D. Baier et al. // *Addiction*. — 2015. — Vol. 110, no. 5. — Pp. 842–851.

21 *Yee, N.* Motivations for play in online games / N. Yee // *CyberPsychology & Behavior*. — 2006. — Vol. 9, no. 6. — Pp. 772–775.

Appendix A Replacement

```
north = ['North-Kazakhstan region','Kostanay region',
'Pavlodar region','Akmol region']
central = ['Karaganda region']
south = ['Almaty region','Turkestan region','Zhambyl region',
'Zhetysu region','Kyzylorda region']
west = ['Aktobe region','Atyrau region','West-Kazakhstan region',
'Mangystau region']
east = ['East-Kazakhstan region', 'Abai region']
other = ['Uzbekistan','Canada','Ukraine','Russia','South Korea',
'USA','China','UAE', "Kyrgyzstan","Italy","Kyrgystan"]

dfm['region'] = dfm['region'].replace(north, "North KZ")
dfm['region'] = dfm['region'].replace(central, "Central KZ")
dfm['region'] = dfm['region'].replace(south, "South KZ")
dfm['region'] = dfm['region'].replace(west, "West KZ")
dfm['region'] = dfm['region'].replace(east, "East KZ")
dfm['region'] = dfm['region'].replace(other, "Other country")
```

Appendix B Split method

```
data['age'] = data['age'].replace({'51+': '51-75'})

lower_bounds = data['age'].str.split('-',
expand=True)[0].astype(int)
upper_bounds = data['age'].str.split('-',
expand=True)[1].astype(int)

mean_age = (lower_bounds + upper_bounds) / 2

mean_age_rounded = mean_age.round().astype(int)

data['age'] = mean_age_rounded
```

Appendix C Label Encoding and One-hot Encoding

```
encoder = LabelEncoder()

columns_to_encode = ['nation','who','education','region','skills',
'culture_together','clubs','sleeping_trouble',
        'sick','language_vg','language_communicate','prefer_with',
        'type_vg','affect_relationship',
        'affect_wellbeing','emotions_affect_while_playing',
        'freq_per_week','when_play','time_play',
         'char']
for column in columns_to_encode:
    encoder.fit(data[column])
    data[column] = encoder.transform(data[column])


encoded_data = data['genre_vg'].str.get_dummies(', ')

encoded_data = encoded_data.add_prefix('genre_')

data = pd.concat([pd.DataFrame(data), encoded_data], axis=1)

data = data.drop(columns = ['genre_vg'])

encoded_data = None
```

Appendix D K-Means and Hierarchical Clustering

```
def perform_grid_search(data, labels, min_clusters, max_clusters):
    scaler = StandardScaler()
    scaled_data = scaler.fit_transform(data)

    param_grid = {'n_clusters': range(min_clusters, max_clusters+1)}
    kmeans = KMeans(random_state=42)
    grid_search = GridSearchCV(kmeans, param_grid,
    scoring=make_scorer(silhouette_score, greater_is_better=True),
    cv=5)
    grid_search.fit(scaled_data, labels)

    best_num_clusters = grid_search.best_params_['n_clusters']
    best_score = grid_search.best_score_

    return best_num_clusters, best_score


def perform_kmeans_clustering(data, labels, num_clusters):
    scaler = StandardScaler()
    scaled_data = scaler.fit_transform(data)

    kmeans = KMeans(n_clusters=num_clusters, random_state=42)
    cluster_labels = kmeans.fit_predict(scaled_data)

    cluster_results = pd.DataFrame({'Cluster': cluster_labels})
    clustered_data = pd.concat([data, cluster_results], axis=1)

    cluster_distribution = clustered_data['Cluster'].
    value_counts().sort_index().astype(int)
    print("Cluster distribution:")
    print(cluster_distribution)

    cluster_centroids = pd.DataFrame(
        scaler.inverse_transform(kmeans.cluster_centers_),
        columns=data.columns
    ).astype(int)
    print("\nCluster centroids:")
    print(cluster_centroids)
```

```
    cluster_demographics = clustered_data.groupby('Cluster').
    mean().round().astype(int)
    print("\nCluster:")
    print(cluster_demographics)



scaler = StandardScaler()
scaled_subset = scaler.fit_transform(subset_columns)

param_grid = {
    'n_clusters': range(2, 10),
    'linkage': ['ward', 'complete', 'average', 'single']
}

def silhouette_scorer(estimator, X, y_true=None):
    labels = estimator.labels_
    score = silhouette_score(X, labels)
    return score

hierarchical = AgglomerativeClustering(n_clusters=5, linkage='ward')

hierarchical.fit(scaled_subset)

silhouette_avg = silhouette_score(scaled_subset,
hierarchical.labels_)
print("Silhouette Score:", silhouette_avg)


cluster_labels = hierarchical.labels_

cluster_results = pd.DataFrame({'Cluster': cluster_labels})

clustered_data = pd.concat([dg, cluster_results], axis=1)

cluster_distribution = clustered_data['Cluster'].
value_counts().sort_index()
print("Cluster distribution:")
print(cluster_distribution)
```

```
cluster_demographics = clustered_data.groupby('Cluster').
mean().round().astype(int)
print("\nCluster demographics:")
print(cluster_demographics)
```