Steam Market Analysis Report

Team Participants:

Artur Kashnikov

Egor Lukjanenko

Arkadi Statsenko

Project Repository:

https://github.com/Artysuk/Steam-Project

1. Identifying Your Business Goals

Background

This project aims to analyze the Steam gaming market, focusing on understanding current market trends, genre popularity, and predicting the success of gaming projects. Steam, being a major platform for digital distribution of games, offers a wealth of data for analysis.

Business Goals

1. Analyze the state of the market based on current data, focusing on successful genres and games, and identifying factors that influence success.
2. Investigate genre popularity over an 11-year period to identify patterns correlating peak users with specific genres.
3. Develop predictive models to forecast the success of gaming projects based on existing market data and historical market dynamics.

Business Success Criteria

The success of the project will be measured by its ability to provide accurate and actionable insights into market trends, genre popularity, and the potential success of new games. This includes the development of reliable predictive models for project success.

2. Assessing Your Situation

Inventory of Resources

Resources include the datasets 'details.csv' (which contains current market data about large set of products: "name", "developer", "publisher", "positive" - positive reviews, "negative" -

negative reviews, "min_owners" - estimated minimal amount of purchases, "max_owners"- estimated maximum amount of purchases, "average_forever" - average amount of active users for whole project existing time, "average_2weeks" - average amount of active users for last two weeks, "median_forever" - median amount of active users for whole project existing time , "median_2weeks" - median amount of active users for last two weeks, "ccu" - peak CCU yesterday, "price" - current US price in cents, "initialprice" - initial US price in cents, "discount" - current discount percent, "genre", "tags", "languages") and 'timestamps.csv' (which contains amount of peak active users for some month, since July 2012 with step of one month), which were obtained by data crawling from SteamSpy and Steam Charts accordingly, containing comprehensive data about games on Steam. Additionally, our team has access to computational resources and data analysis tools. Most likely we are going to use Python and it's open source libraries for data analysis, and of course the Internet.

Requirements, Assumptions, and Constraints

The project assumes the availability and reliability of the Steam datasets. Constraints include the time and computational resources available for in-depth data analysis.

Risks and Contingencies

Potential risks include data inaccuracies or incompleteness. Contingency plans involve cross-validating with additional data sources or adjusting the analysis scope.

One of the reasons for potential inaccuracies is too many factors influencing the success of individual games. Most of the data that we are using is accurate only at time which it was captured, such as reviews and amount of owners, so it does not show full dynamic over time. Peak user activity over time does not show specific reasons, which affected this attribute, such as sales, updates, release of DLC and other factors.

Terminology

Key terminologies include market trends, genre popularity, user engagement metrics, and predictive modeling.

Costs and Benefits

The main cost is time invested in data analysis. Benefits include gaining insights into the Steam market, which can guide developers and marketers in decision-making. For

3. Defining Your Data-Mining Goals

Data-Mining Goals

The data-mining goals align with our business objectives: analyzing market state, genre popularity trends, and developing predictive models for game success. There may be different goals for different groups. For example for investors the program predicting the success of some title belonging to some company may be one of the factors influencing their behavior. Analyzing trends and market situations on Steams allows one to have a better understanding of the industry, which may also be of interest to different groups of people, including investors, social scientists, analysts.

Data-Mining Success Criteria

Success in data-mining will be assessed based on the accuracy and relevancy of the findings, as well as the predictive performance of the developed models.

Task 3. Data Understanding

Gathering Data:

- Data Requirements: For a comprehensive analysis of the gaming market, our project primarily targets data encompassing game genres, user ratings, game developers, number of purchases, and user engagement metrics. These elements are pivotal in painting a detailed picture of what drives success in this dynamic market.
- Verify Data Availability: We source our data from publicly available Steam datasets and APIs. Unfortunately not all data that we would like to have is covered by these datasets, for example information about download statistics and revenue from game is not publicly accessible, which is an obstacle to ensuring accuracy of our analysis and predictions as well as choosing some other, potentially more interesting aspects of the game market to analyze.
- Selection Criteria: Our selection process is guided by three key criteria: relevance, completeness, and recency. We focus on data from the past 11 years, ensuring that our analysis is both current and relevant to today's market dynamics. This time frame is critical in capturing the evolution of trends and preferences in the gaming industry.

Describing Data:

- The dataset at our disposal is rich and varied, comprising essential variables like game titles, genres, release dates, user ratings, and game developers. This diversity allows us to delve deep into the factors that govern game popularity and trends. By analyzing these variables, we can draw meaningful insights into how different genres have evolved over time and what factors most significantly impact a game's success.

Exploring Data:

- In our initial analysis, we delve into understanding the trends that dictate game popularity and genre preferences. A key focus is to investigate correlations between different attributes such as user reviews, number of purchases, playtime and pricing. This exploration is vital to unravel the nuances of what constitutes a successful game in the eyes of the consumers. By examining these correlations, we aim to uncover the underlying success factors that can guide future game development and marketing strategies.

Verifying Data Quality:

- Ensuring the quality of our data is of utmost importance. Considering that publicly available Steam datasets are most commonly generated using Steam's API the data is already pretty accurate, but it's still important to conduct checks. This process includes handling missing values and addressing outliers, which are critical steps in preparing the dataset for robust analysis. These initial cleaning steps not only refine the dataset but also bolster the reliability of our subsequent analysis.

Task 4. Planning Your Project

Project Plan:

Task	**Description**	**Time estimation**	**Assigned to**
Data Collection	Gathering and initial processing of Steam data	20h	all
Data Cleaning	Addressing missing values, outliers, and inconsistencies.	15h	all
Exploratory Data Analysis	Identifying patterns and trends.	25h	all
Predictive Modeling	Developing models to predict game success.	30h	all

Evaluation and Refinement	Testing and refining the models.	10h	all

Methods and Tools: The project will utilize Python for data analysis and machine learning, with libraries like pandas, scikit-learn, and matplotlib.