

Steam Games Data Analysis

This presentation will explore the key findings from an exploratory data analysis project on Steam games. The analysis will focus on uncovering patterns and insights related to game pricing, popularity, playtime, and user sentiment.

by Sri Ram Prabu Elenchezian





Introduction

■ This project delves into the fascinating world of Steam games, leveraging data analytics to understand the complex relationships between game characteristics and player engagement.

■ By analyzing a vast dataset of Steam games, we aim to gain valuable insights into the factors that contribute to game popularity, pricing strategies, user sentiment, and playtime.

Dataset Description

The dataset comprises information on over 97,000 Steam games, providing a rich and diverse collection of data for analysis.

Key columns include game name, release date, price, playtime, user reviews, and genres, offering a comprehensive overview of each game's characteristics.

Data sources include the Steam API and third-party data providers, ensuring a reliable and comprehensive dataset for analysis.

Data cleaning is crucial to ensure data quality and consistency, addressing missing values, inconsistent formatting, and other data anomalies.

Data Cleaning Process

The release date column was converted to a date format, ensuring consistent data representation and enabling temporal analysis.

Column names were standardized by removing spaces and special characters, improving data readability and consistency across the dataset.

Columns with over 60% missing data were removed to minimize the impact of missing values on the analysis.

Numeric conversion was performed for the price and playtime columns, enabling quantitative analysis and statistical calculations.

Feature Engineering



A 'Price Category' feature was created by grouping games into categories: Free, Low, Mid, and High.



A 'Playtime Category' feature was engineered to segment games based on engagement levels: No Engagement, Casual, Moderate, and Highly Engaged.



A 'Sentiment Score' feature was calculated by determining the percentage of positive reviews for each game.

Rew Game Drates

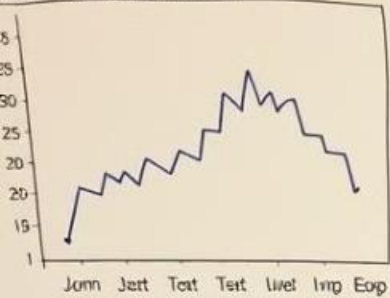
		Devieiw Text		
Player ID:	Dar	Turt	\$7.0	\$65sec
Game Title	-	\$	50.45	3173ec
Date Price:	-	\$	\$0.45	\$6.7sec
Drta Price:	-	\$	50.43	\$35sec
Rayttime:	-	\$	20.27	\$705ec
Relytime:	-	\$	50.05	\$155ec

N:w Featureess

		Features:	
Price Cartegory		Plibytime Cartagory	
→ SentimentScore*		\$155	2067
		62	3035
		\$1.60	40.15
		340	96.75

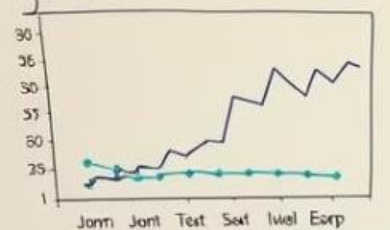
Pro EDA - MI!

Statistics of Alacrities



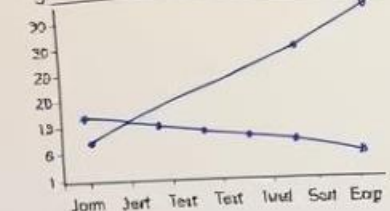
- Foulking Prices
- Your hogue bocke
- Bellior byationes
- Over buy face
- Medline's gnating Comony of et's lach.

laytime will Price



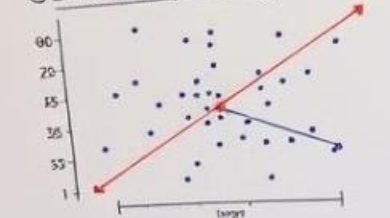
- + Jrycified maske
- Foul disparing
- ① 3 3 5
- My farbs law
- ② 3 3 5
- In far fast

laytime will depormen)



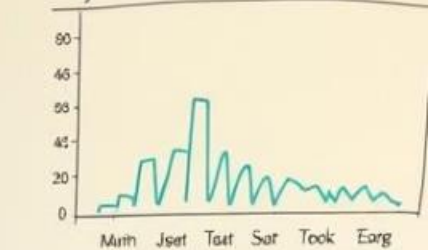
- + Scatter piece)
- End bnice lopolaring, incouts were fast)
- (Love many led) with the fast

Genre. Popularity

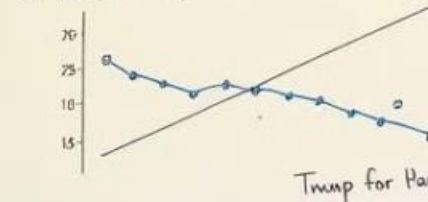


- Exebro (tonel) Plank Hen boack to befire)
- genre chacke.
- Eval be withien Positionette foriation in Pac!
- You hjarises
- Conckring Pertheris for neeased is
- Seorcal with Far the chagmols
- your haaget

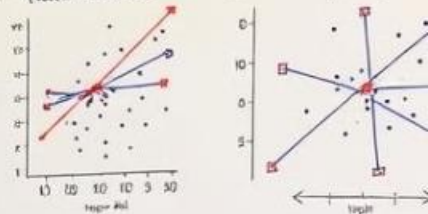
Playnice Levere Blatime



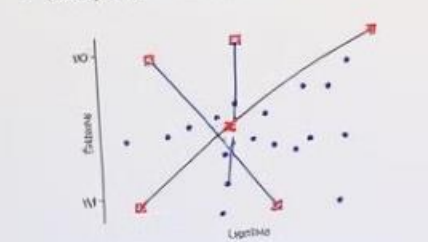
3. Farttre Populmie)



Raid ofices



5. Door, Genires fartbrick



Exploratory Data Analysis (EDA)

1

The analysis examined the distribution of game prices and playtime, revealing patterns in pricing strategies and player engagement.

2

The most common game genres and categories were identified, providing insights into the popularity of different game types.

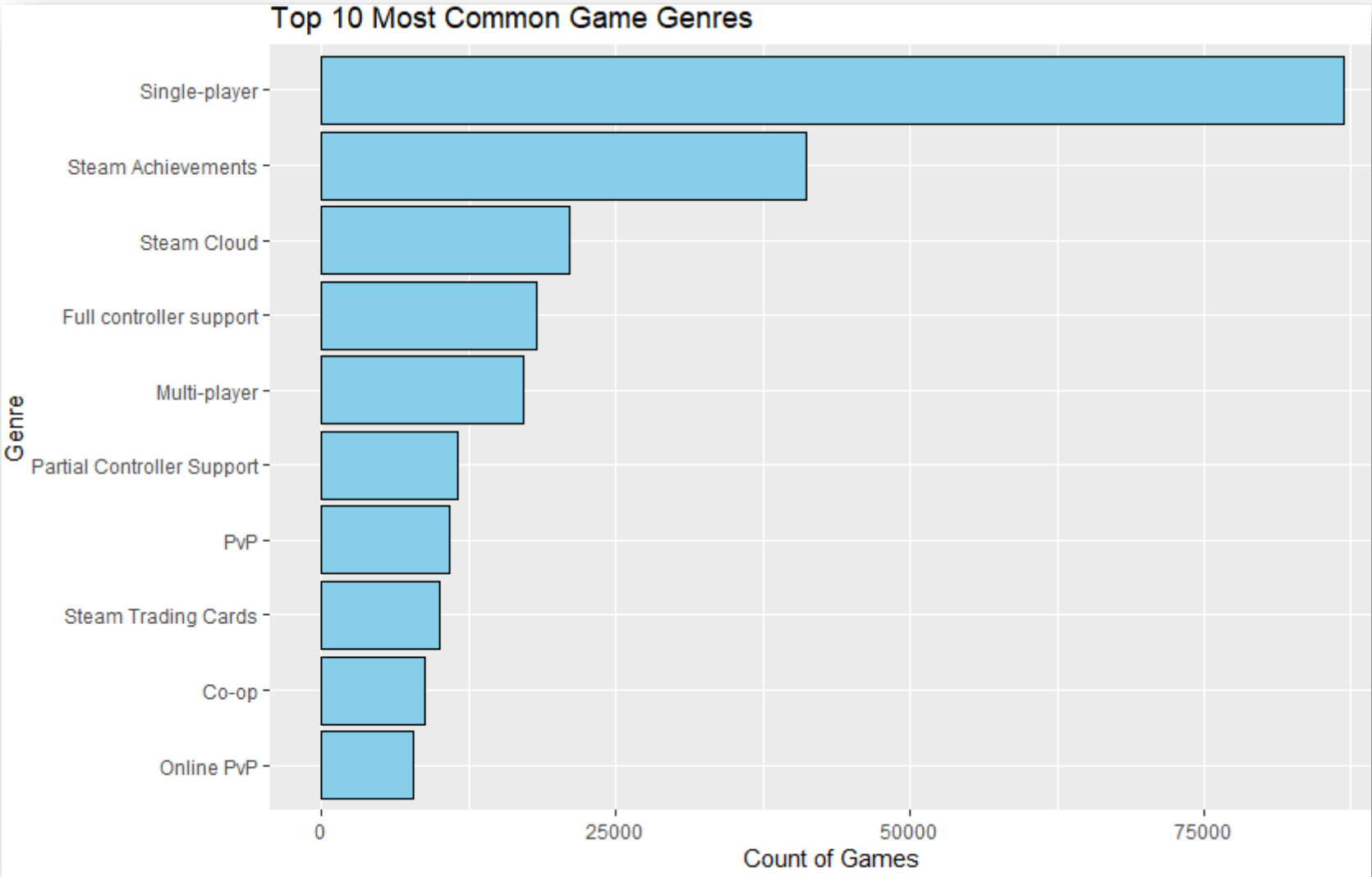
3

The correlation between price and playtime was explored to understand the relationship between pricing and player engagement.

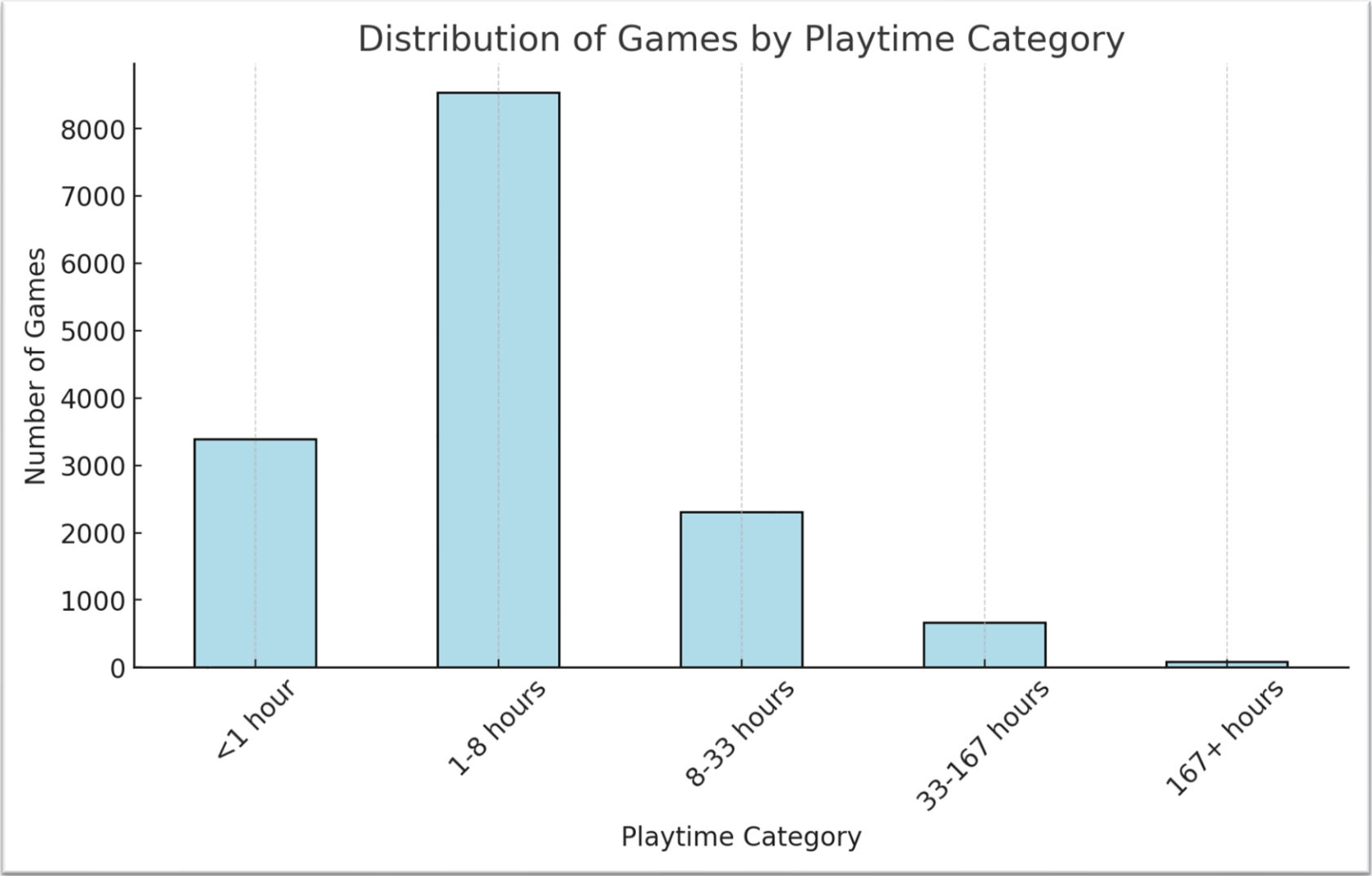
4

User review sentiment scores were visualized to identify trends in player satisfaction and the factors that influence it.

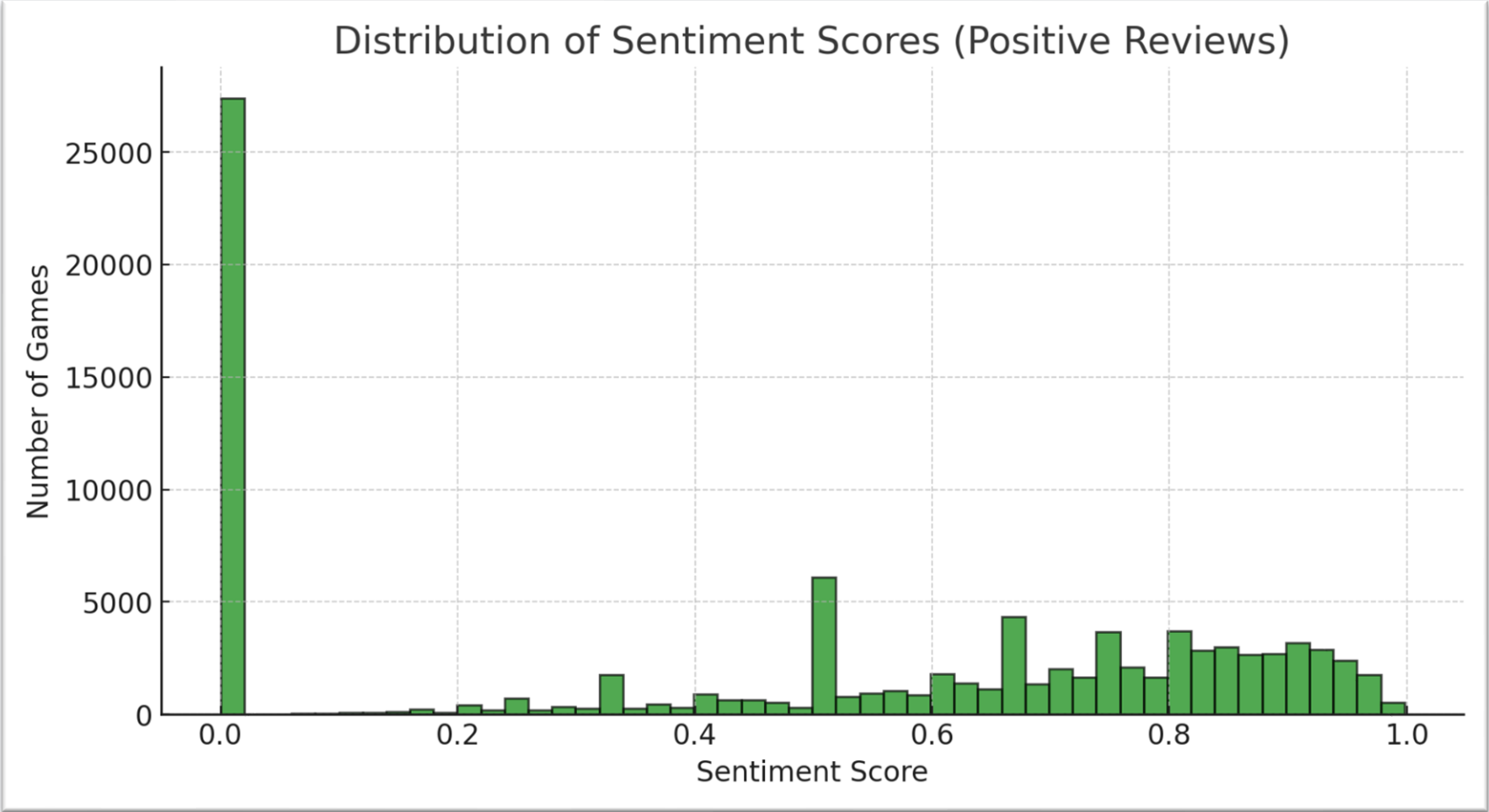
Key Visualizations



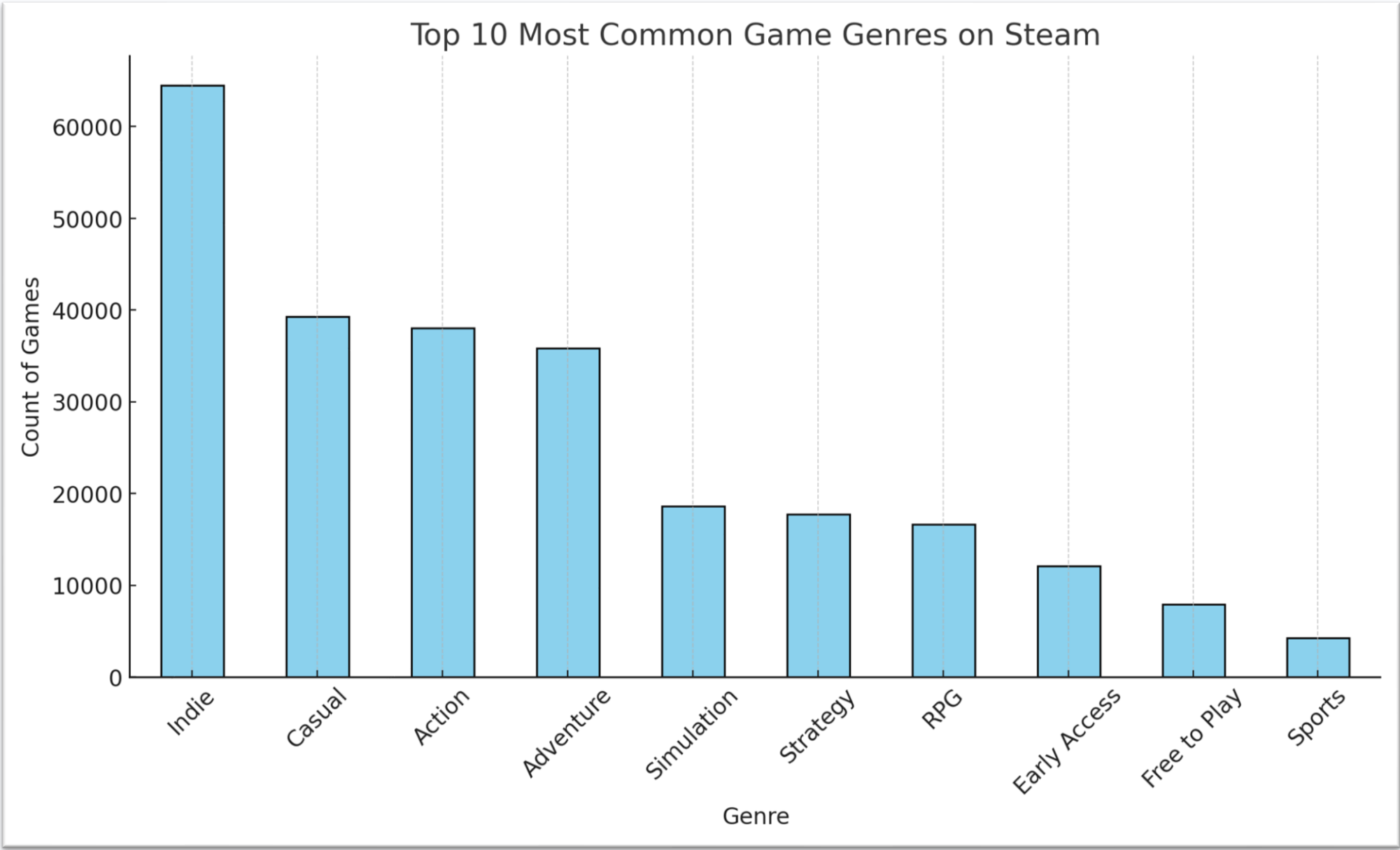
Key Visualizations



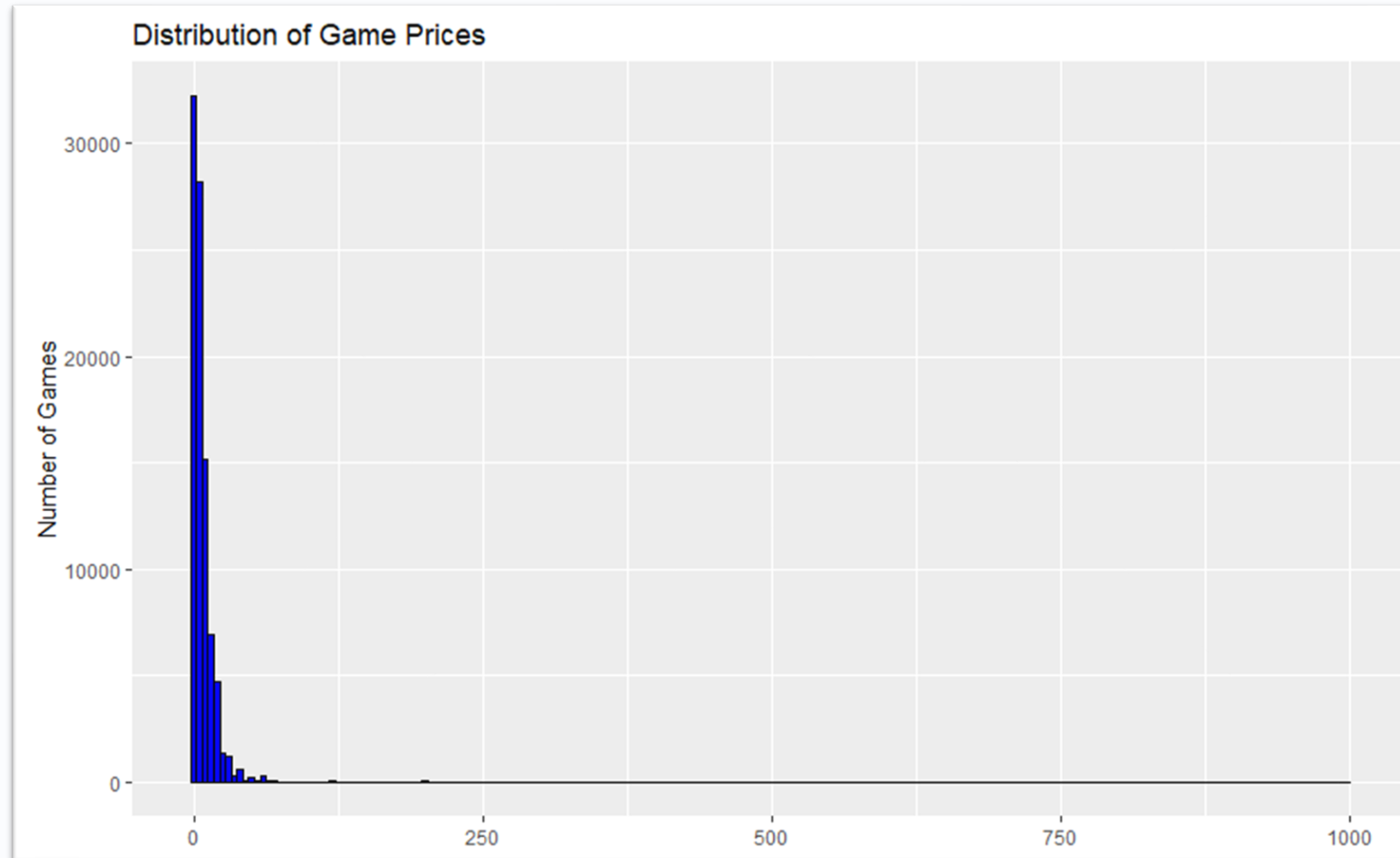
Key Visualizations



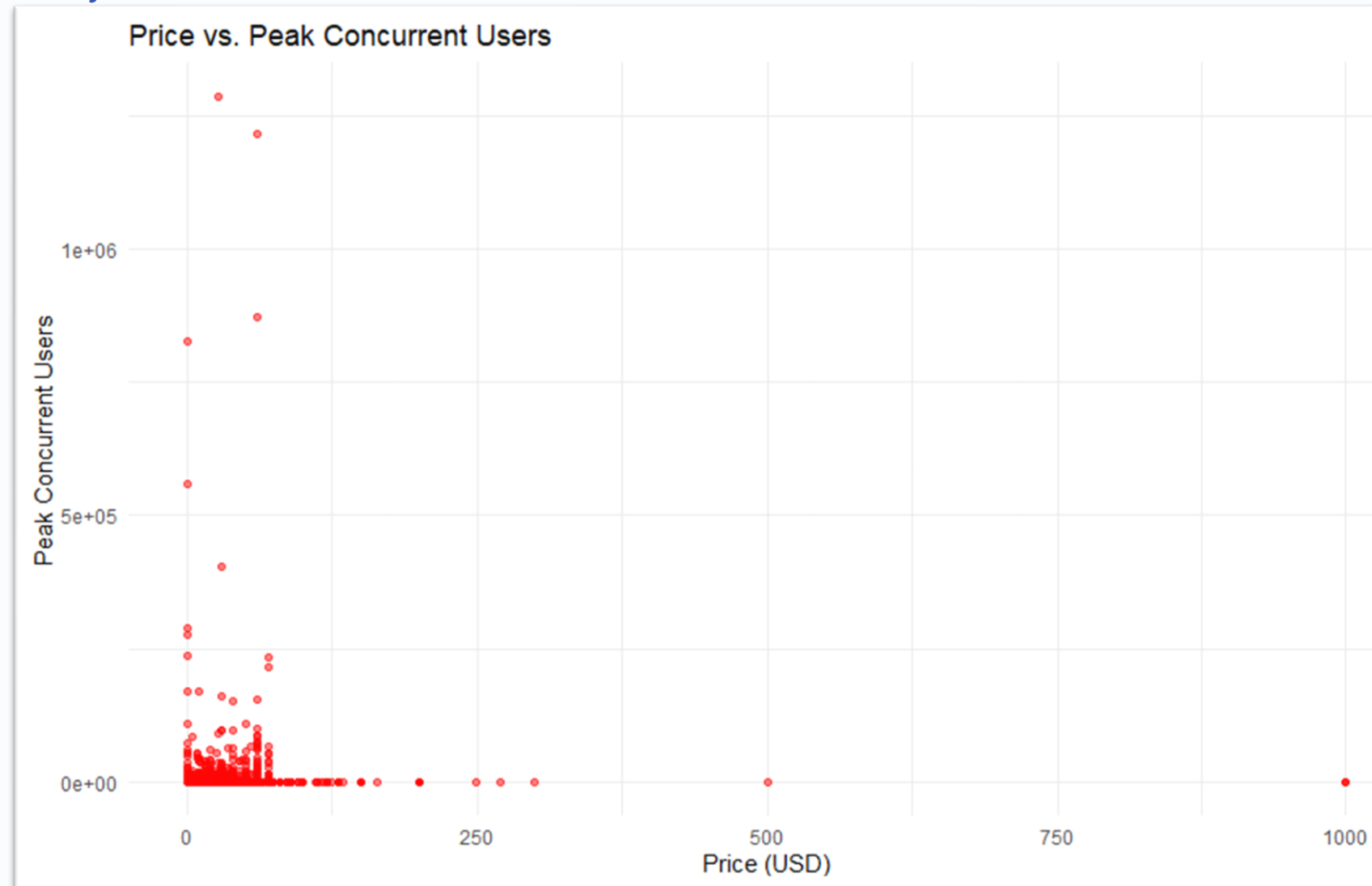
Key Visualizations



Key Visualizations



Key Visualizations



Conclusion & Insights

The analysis reveals that free and low-priced games dominate the Steam market, reflecting the prevalence of accessible gaming options.

Indie and casual games are the most common genres on Steam, showcasing the popularity of these game categories.

High user engagement is often linked to multiplayer and competitive games, emphasizing the social and interactive aspects of gaming.

Positive sentiment is higher for well-rated and frequently updated games, indicating the importance of continuous improvement and player satisfaction.

References:

- Kaggle. (n.d.). *Steam Games Dataset*. Retrieved from <https://www.kaggle.com/>
- R Project for Statistical Computing. (n.d.). *CRAN - Comprehensive R Archive Network*. Retrieved from <https://www.r-project.org/>
- Data.gov. (n.d.). *Open Government Data*. Retrieved from <https://www.data.gov/>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.