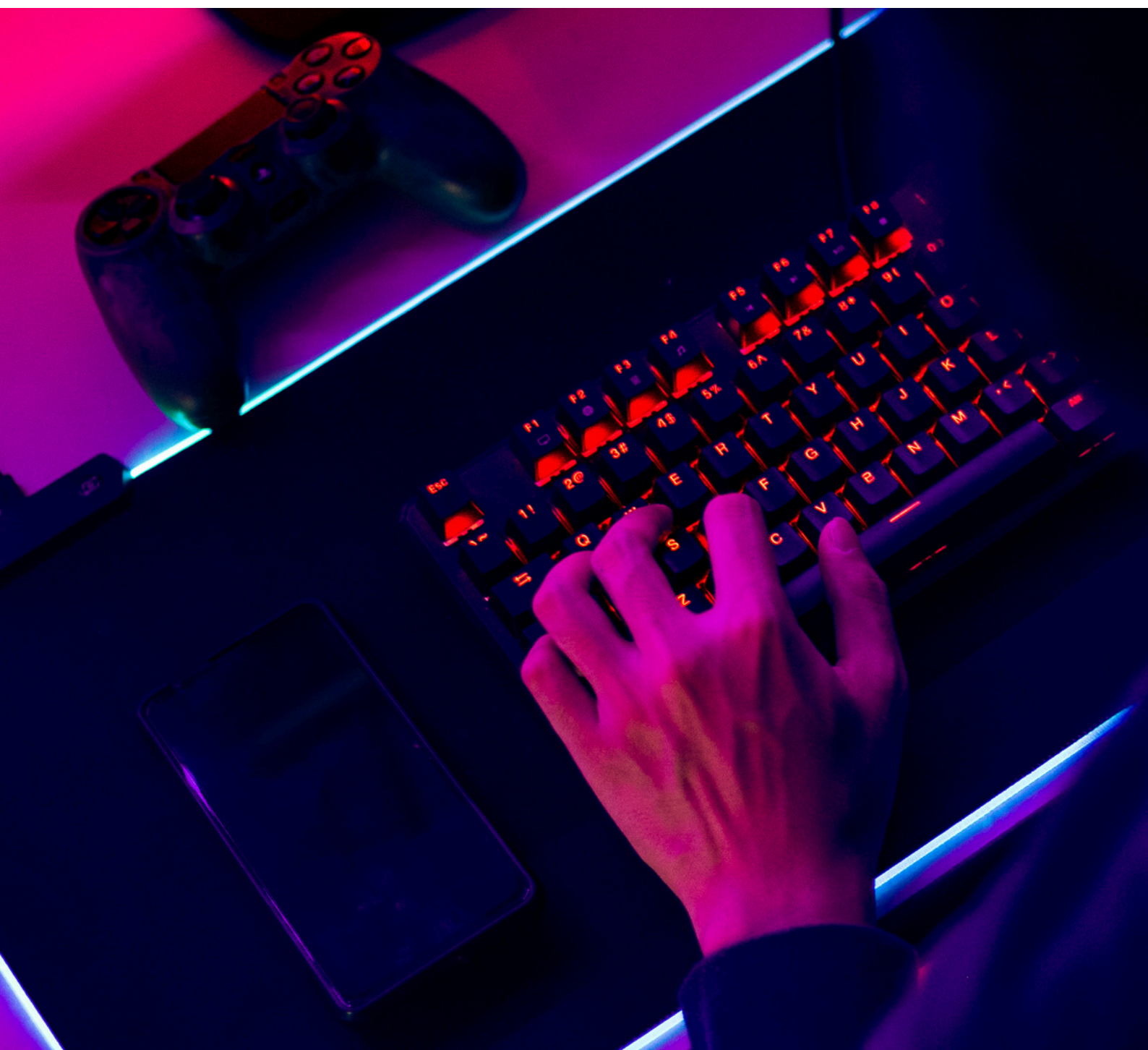


Exploratory Data Analysis (EDA): Player Behaviour in Online Games



INDEX

01 Introduction

01.A Context

01.B Hypothesis

02 Data Exploration

03 Global Data Analysis

03.A Univariate, Bivariate, and
Multivariate Analysis

04 Specific Analysis

05 Conclusions

01 Introduction

This project analyses player behaviour in online games using a dataset sourced from Kaggle.

Source:

Rabie El Kharoua. (2024).

 Predict Online Gaming Behavior Dataset [Data set].

Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/8742674>.

<https://www.kaggle.com/datasets/rabieelkharoua/predict-online-gaming-behavior-dataset>

01.A Context

The starting point is this dataset (online_gaming_behavior_dataset.csv) containing a sample of 40034 entries and 13 columns. Apparently the dataset has already been processed and conditioned by the author, ready to be analysed. At a more advanced point in the analysis, the dataframe is filtered and the 'PlayerID' column is removed.

The study begins with a global analysis of the data in order to find information about how the collected variables influence the purchase behaviour and the level of engagement of the player. We found very homogeneous and condensed data, which made it difficult to find significant differences when relating variables. For this reason, we decided to carry out a more specific analysis by applying filters to the dataset and segmenting the sample, with the aim of finding significant relationships and answers to our questions and hypotheses.

To answer the questions initially raised and generate a storytelling for our presentation, we consider the hypothetical case in which we seek to guide decision-making to create an online video game, which integrates purchases within it.

From here, the following questions and hypotheses arise.



01.B Key questions and hypotheses

“Players with higher engagement levels tend to make more in-game purchases.”

- What variables influence purchasing decisions?
- Do certain game genres attract more in-game purchases?

“Engagement level affects the time spent playing.”

- Which game genres generate higher engagement?
- Which genres see the highest number of logins per week?
- Are games with higher engagement levels more popular?

New questions and areas of study emerged during the analysis.

02 Data Exploration

Key observations:

- Missing values: None found.
- Frequency analysis: Some variables exhibit very similar frequency distributions.
- Variable classification: Variables are categorised based on cardinality, percentage of cardinality, and data type. The numeric variables have low cardinality.
- Target variables: InGamePurchases and EngagementLevel are selected for in-depth analysis.
- Central tendency measures: The dataset's mean, mode, and median provide insights into the average player profile and their time investment.
- Outliers: None detected.

03 Global Data Analysis

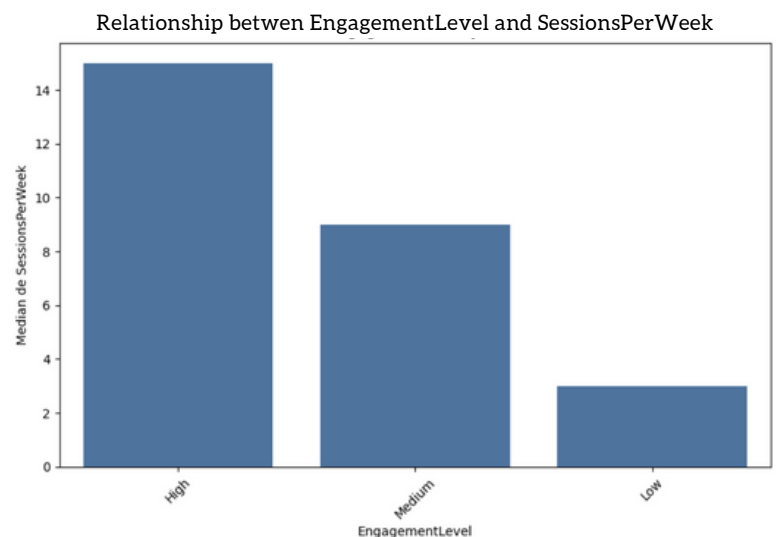
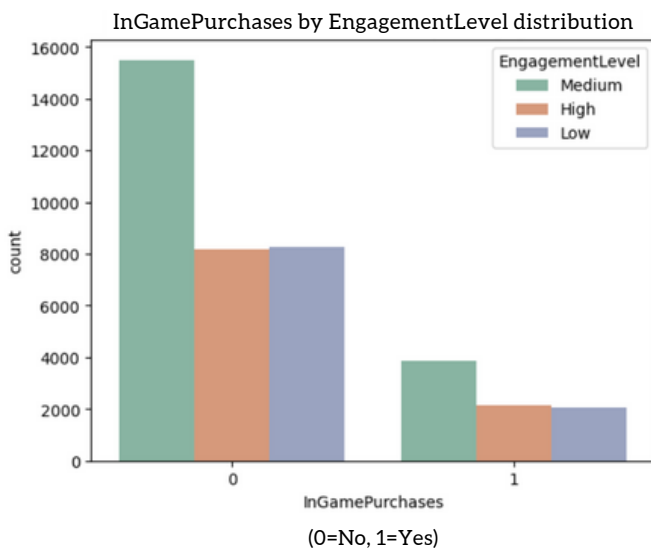
03.A Univariate, Bivariate, and Multivariate Analysis

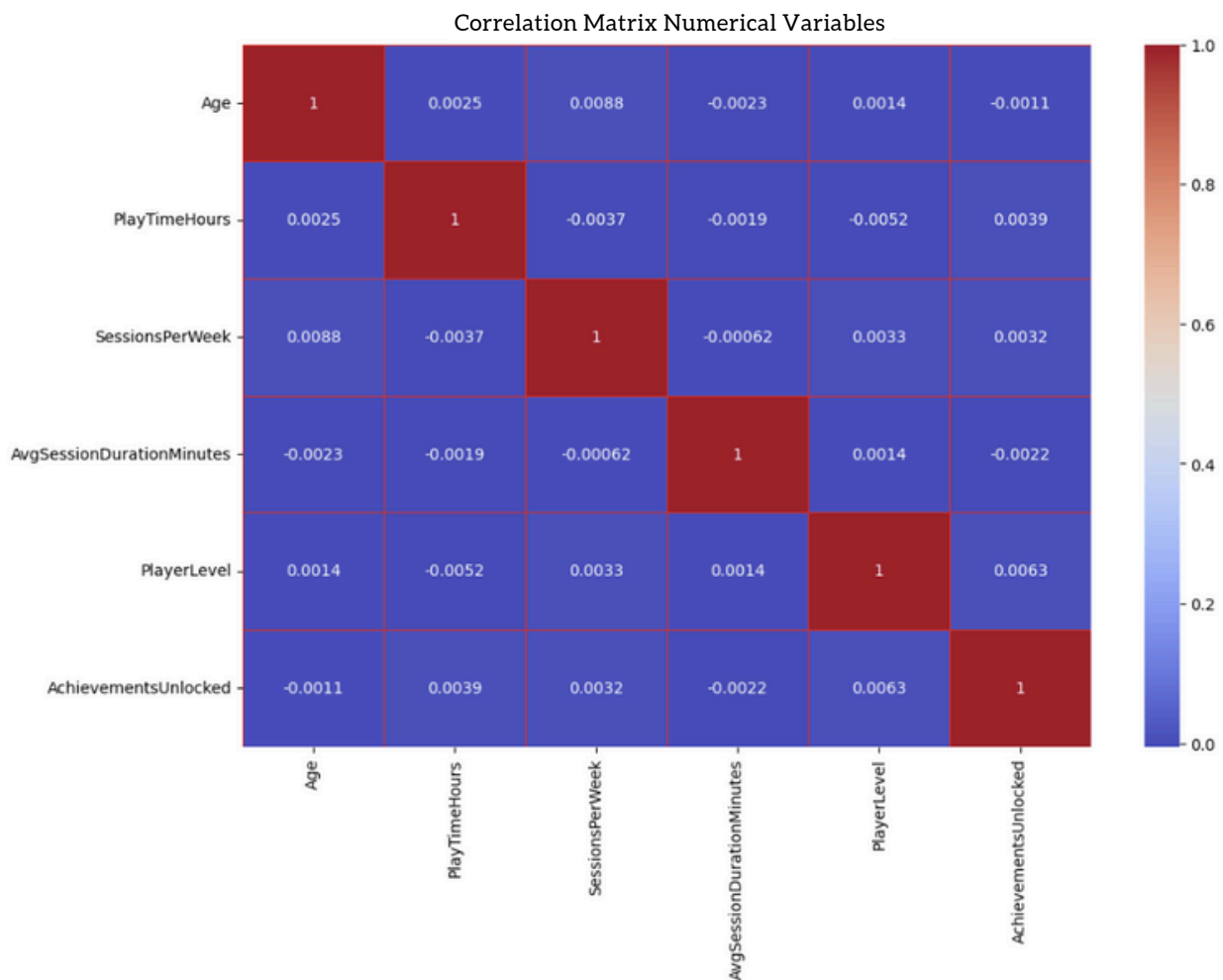
We first performed a bivariate analysis of InGamePurchases to identify behavioural relationships:

- Statistical tests (Chi-square and ANOVA) indicate that in-game purchases are independent of other variables.
- Key Finding: Higher engagement does not necessarily lead to more in-game purchases.
- Implication: Marketing strategies should focus on alternative methods to increase in-game purchases.

Next, we analysed EngagementLevel in relation to GameGenre and SessionsPerWeek:

- No strong correlation was found between EngagementLevel and GameGenre.
- However, players with higher engagement log in more frequently each week.
- The lack of a strong relationship between engagement and game genre could be due to a homogeneous player base.
- A heatmap analysis showed weak correlations between numeric variables.

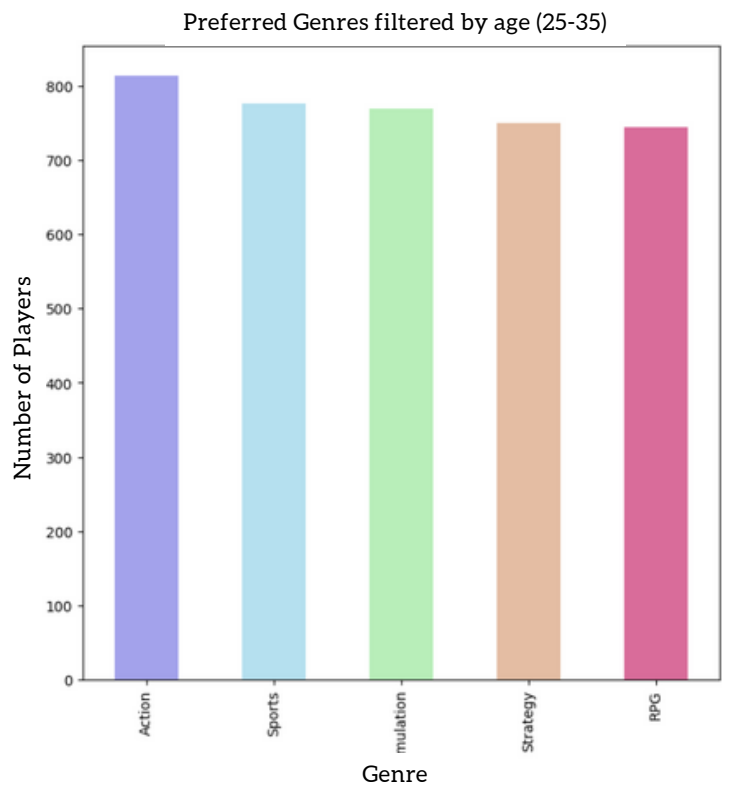
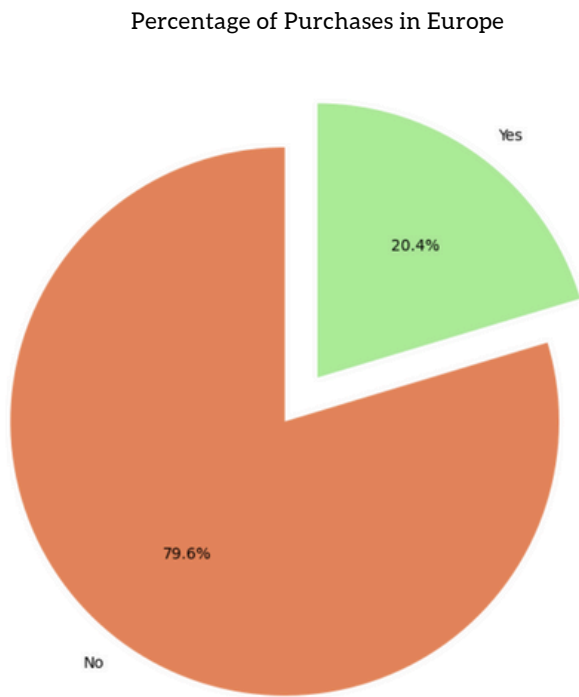
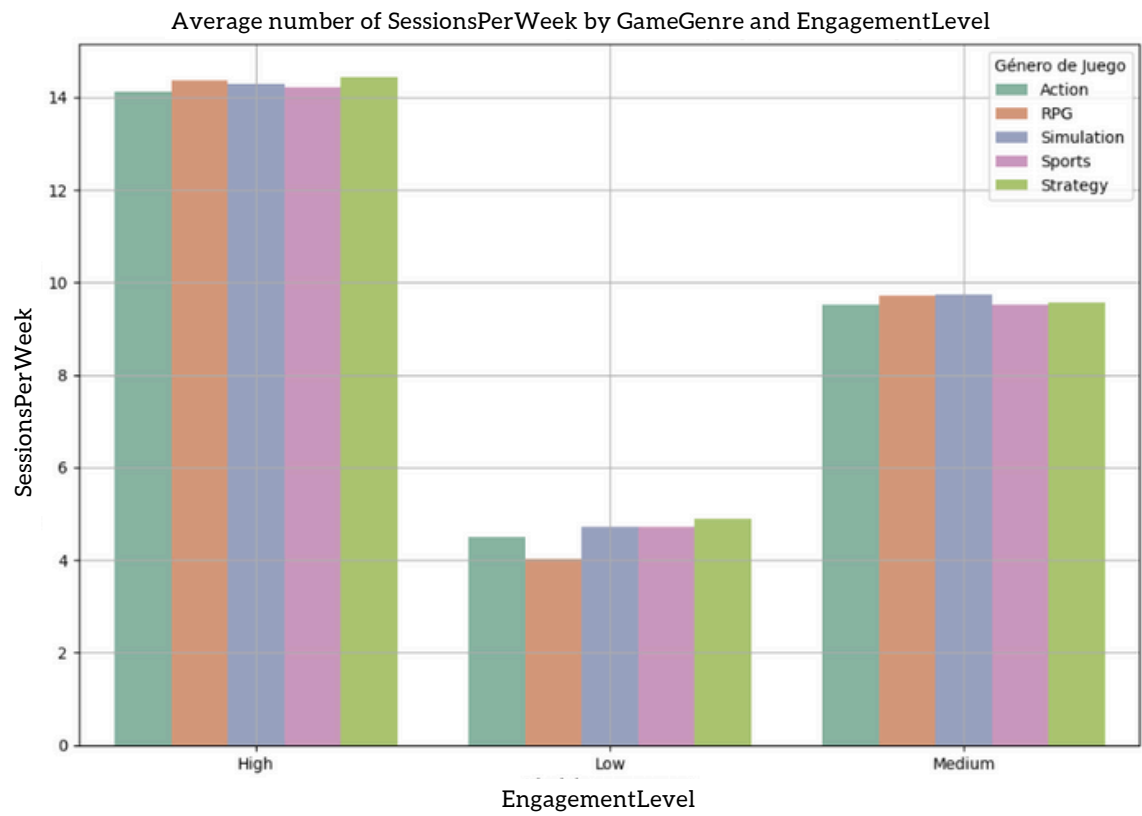




04 Specific Analysis

With the aim of answering our initial questions a little more precisely, we focused on the European market and an age range around the average player age (32 years):

- The dataset was filtered by age (25-35 years), accounting for 32% of the sample.
- Only 20% of players make in-game purchases.
- Players with high engagement log in more frequently per week, particularly in Strategy and RPG games.
- The most popular genres were Action and Sports, though engagement levels remained moderate.
- The average player profile:
 - Age: 32 years
 - Preferred genres: Sports and Action
 - Weekly logins: 9.5
 - Engagement level: Medium



05 Conclusions

For a hypothetical scenario where decisions are made to develop an online game with in-game purchases, the following points are relevant:

- In-game purchases are not a primary player behaviour. Most players do not make purchases.
- Since engagement does not influence purchases, game development should focus on genre popularity rather than engagement levels to attract a larger audience.
- Marketing strategies should:
 - a. Leverage logins as engagement opportunities by introducing special offers, battle passes, or time-limited missions.
 - b. Segment campaigns more effectively to target the right player demographics.

These findings provide valuable insights for designing a game that maximises both engagement and revenue opportunities.



Bootcamp Data Science
Blanca Novella Serrano
2024

Análisis Exploratorio de Datos (EDA):
El comportamiento del jugador en juegos online