PROJECT REPORT

Predicting Profits of Online Sport Bettors

1. Business overview & data description

Gambling is one of the most lucrative and flourishing industries worldwide. At the height of the COVID pandemic, the global casino industry recorded a revenue of $227 billion, 30% of which came from online casinos [1]. With a fast increase in Internet penetration and mobile applications, the global online gambling market has been and will be an extremely promising playground for companies.

bwin Interactive Entertainment (bwin), an online betting agency, has collaborated with Division of Addictions (DOA) to conduct a gambling research project in which the demographic and sports betting records of 46,339 bwin's customers have been summarized and published on The Transparency Project [2]. Acting as a bwin's consultant, I am tasked with a mission: <u>how to predict the profits of sports gamblers on bwin's website in order to develop specialized customer strategies and targeted measures for specific individuals.</u>

The objective of this project is to formulate and compare the efficiency of 2 machine learning models that closely predict the profits of internet sports gamblers based on their gambling tendencies. These models will be compared based on mean absolute error to identify the superior option.

Source:

[1] 10 Online Gambling statistics for 2021 and Beyond Available at: https://playtoday.co/blog/online-gambling-statistics

[2] Dataset: Population Trends in Internet Sports Gambling Available at: http://thetransparencyproject.org/Availabledataset.htm

2. Data understanding - variables

| Feature	Description	Type
user_ID	ID number bwin assigned to each user upon registration	Continuous
date_registered	The subscriber's registration date	Date
date_of_1st_deposit	The date the subscriber first deposited funds into their bwin account.	Date
date_of_1st_bet	The date the subscriber first placed a sports bet	Date
stake_fixed	The total amount in Euros bet on fixed odds sports bets	Numeric
stake_live	The total amount in Euros bet on live action sports bets.	Numeric
win_fixed	The total amount in Euros won in fixed odds sports bets	Numeric
win_live	The total amount in Euros won in live action sports bets	Numeric
bets_fixed	The number of fixed odds sports bets placed by the subscriber	Numeric
bets_live	The number of live action sports bets placed by the subscriber	Numeric
days_fixed	The number of different days that the subscriber placed a fixed odds sports bet	Numeric
days_live	The number of different days that the subscriber placed a live action sports bet	Numeric
profit_live	The profit made from live odds sports bet	Numeric
Profit_fixed	The profit made from fixed odds sports bet	Numeric
Profit_total	The total profit made from sports bet	Numeric

2. Data understanding - correlations

Correlation matrix

Correlation of numerical features to total profit
Correlation Heatmap with Total Profit
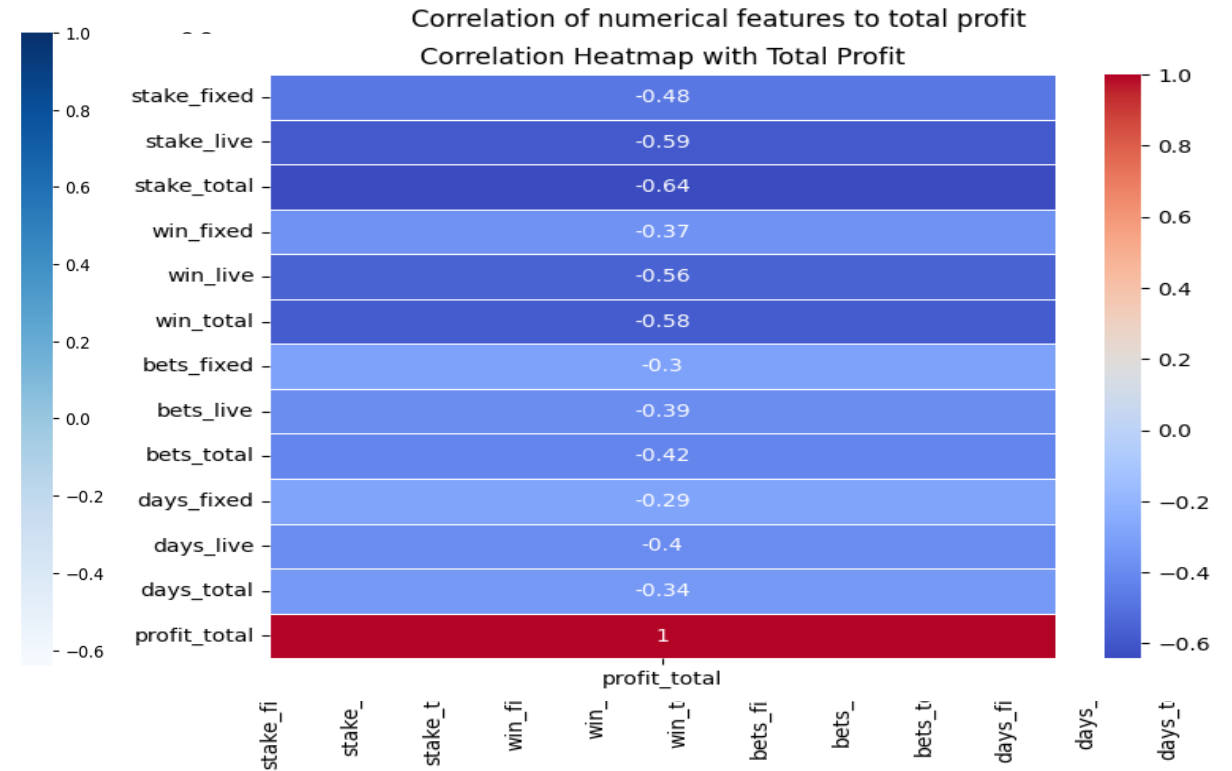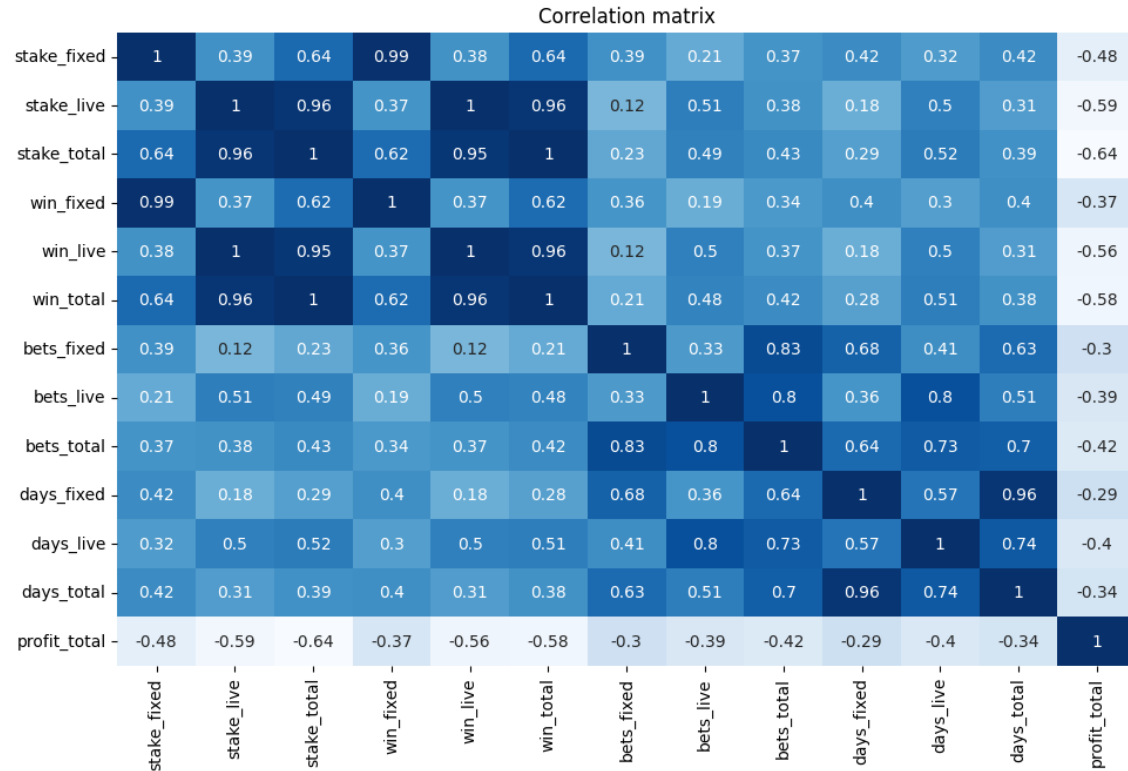
- **Correlations**: In general, there are weak correlations between major features
- **Correlated to total profit**: stake_live, stake_total, win_live, win_total show high correlation to profit_total

However, the features will be further considered to be included in the models

2. Data understanding – ratio of winning players, bets & stakes
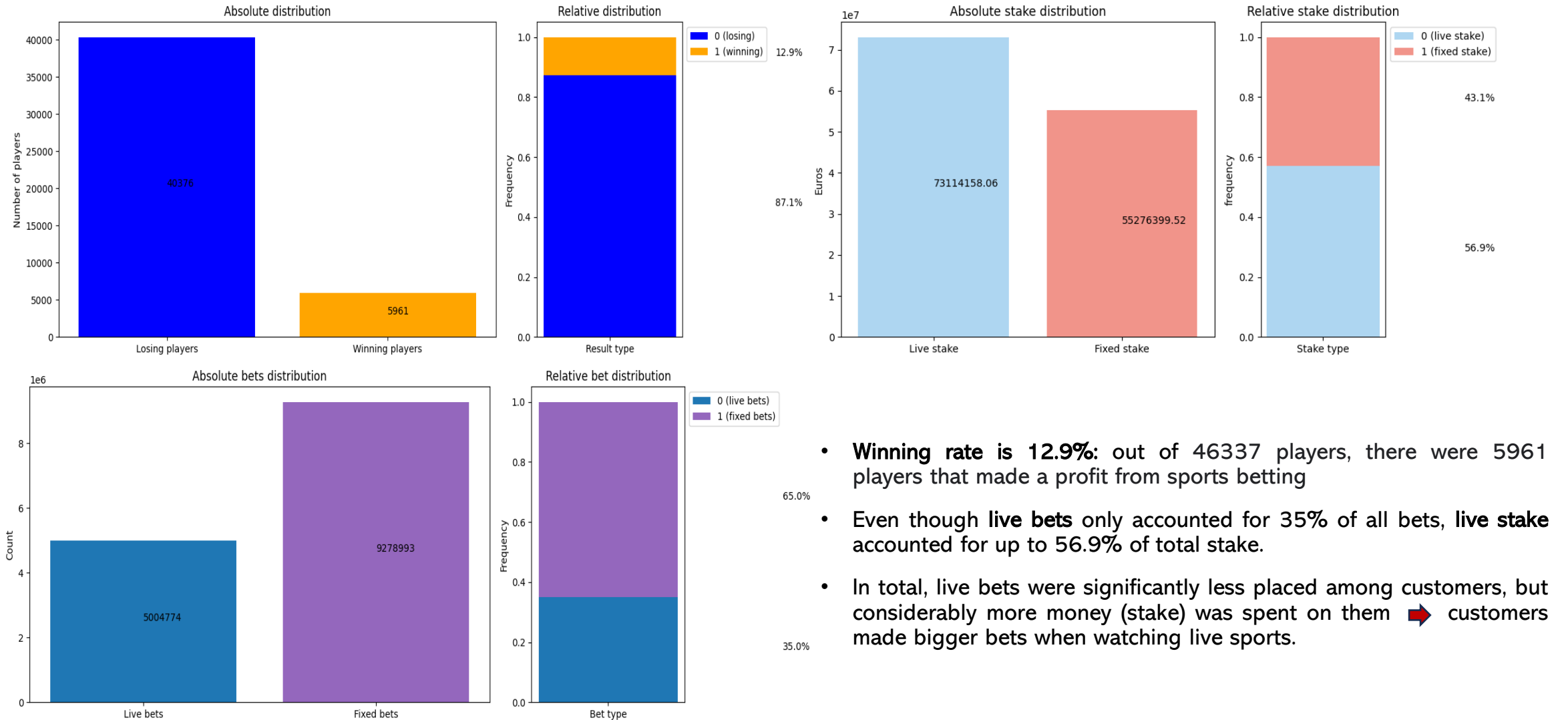
- **Winning rate is 12.9%:** out of 46337 players, there were 5961 players that made a profit from sports betting

- Even though **live bets** only accounted for 35% of all bets, **live stake** accounted for up to 56.9% of total stake.

- In total, live bets were significantly less placed among customers, but considerably more money (stake) was spent on them ➡ customers made bigger bets when watching live sports.

2. Data understanding – bets/stake distribution among winning/ losing players
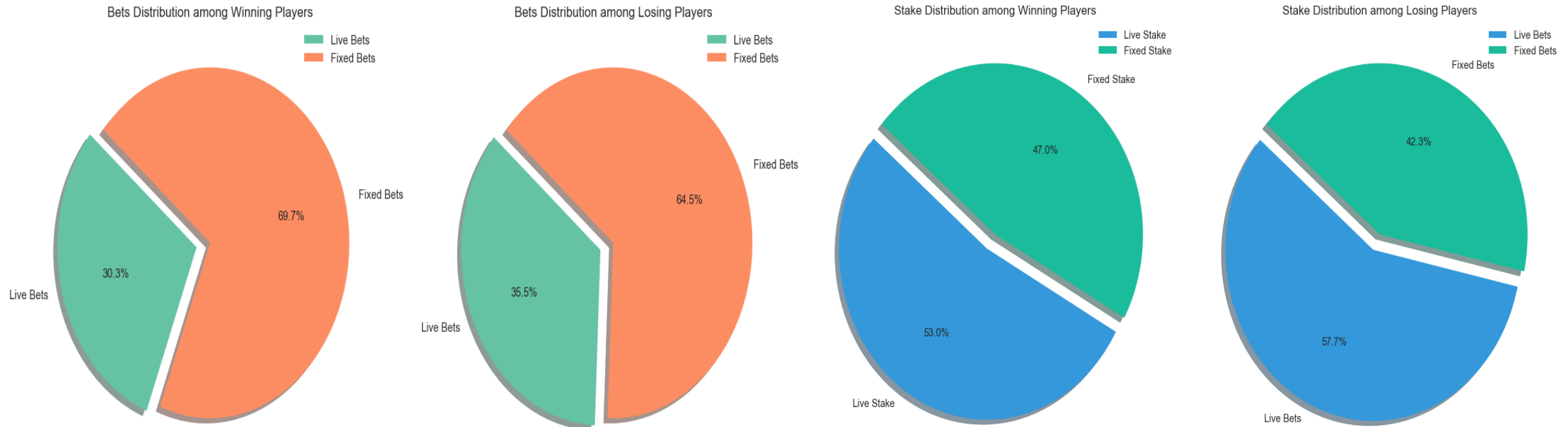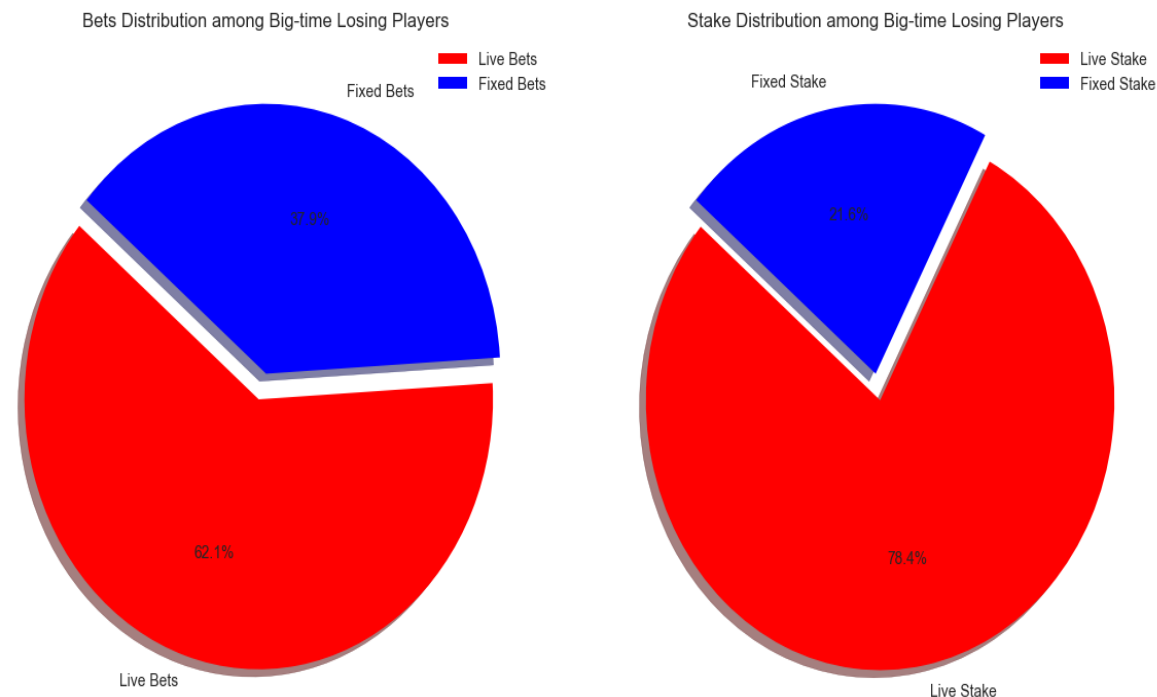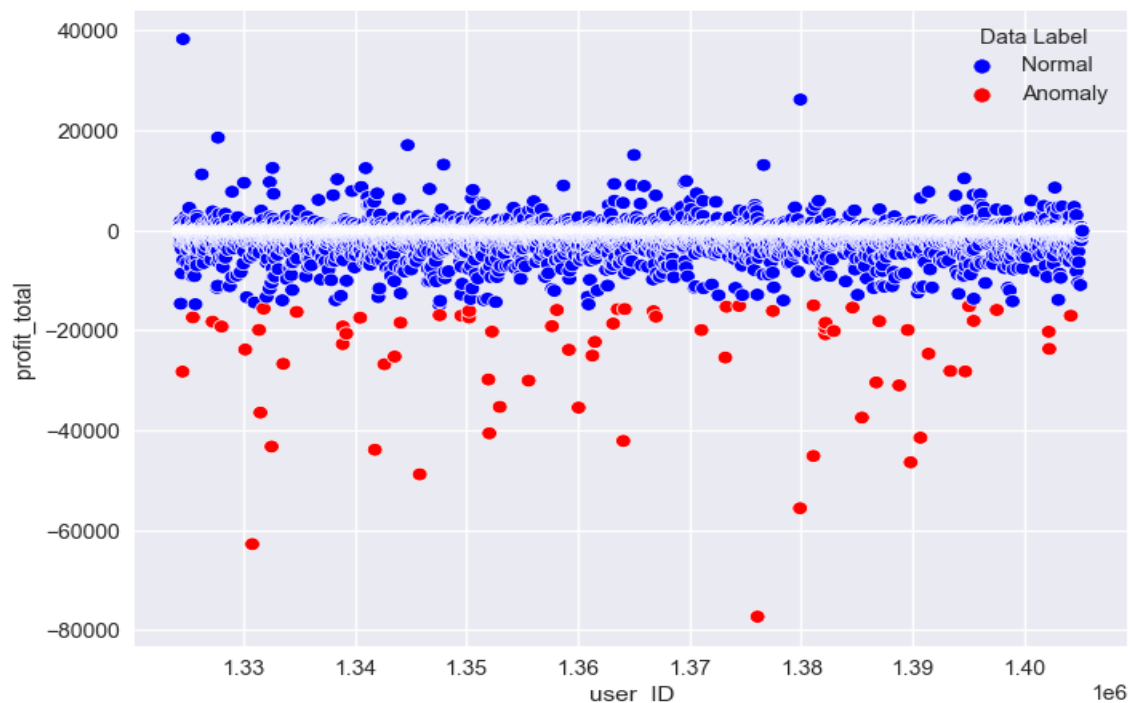

- Winning players and losing players had relatively similar bets and stake distribution
- Winning players slightly preferred fixed bets compared to losing players

2. Data understanding – number of days playing on bwin's website
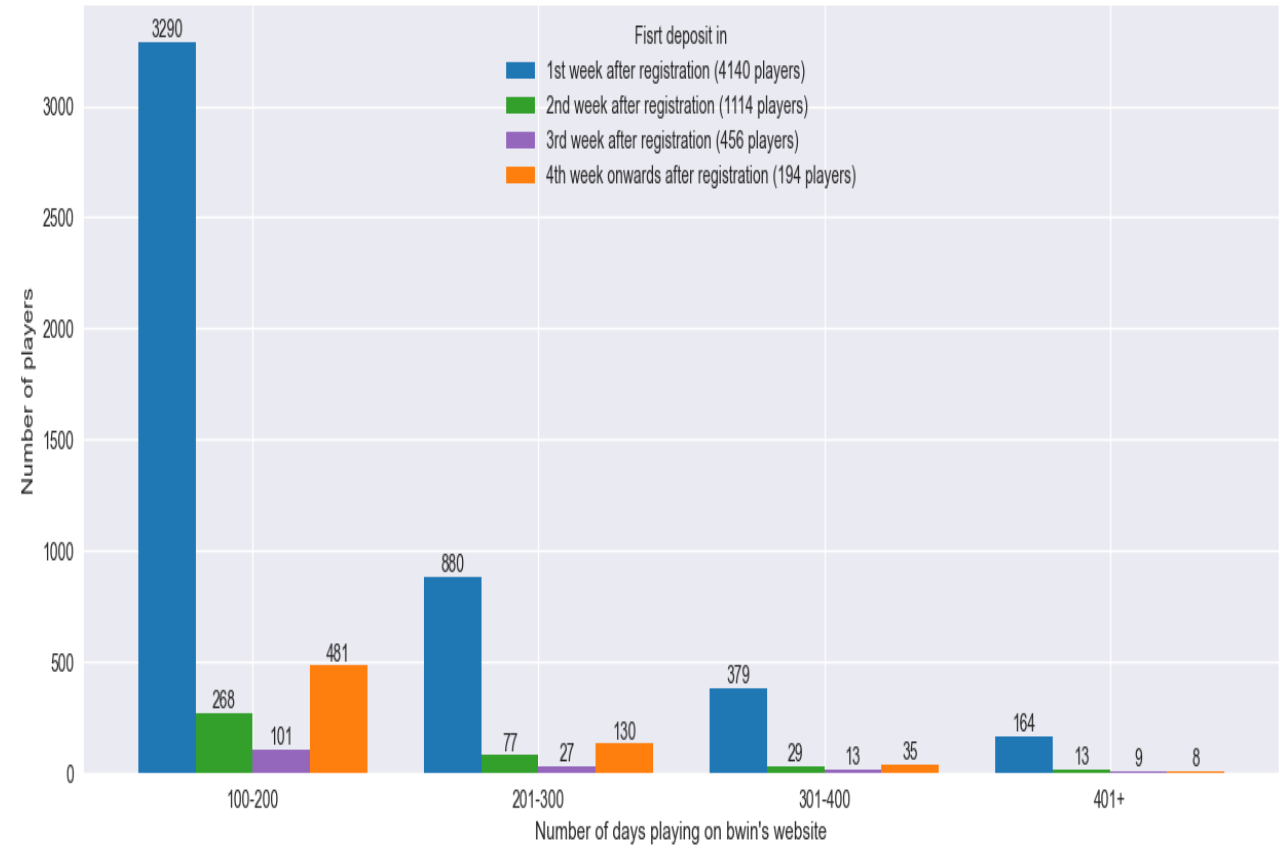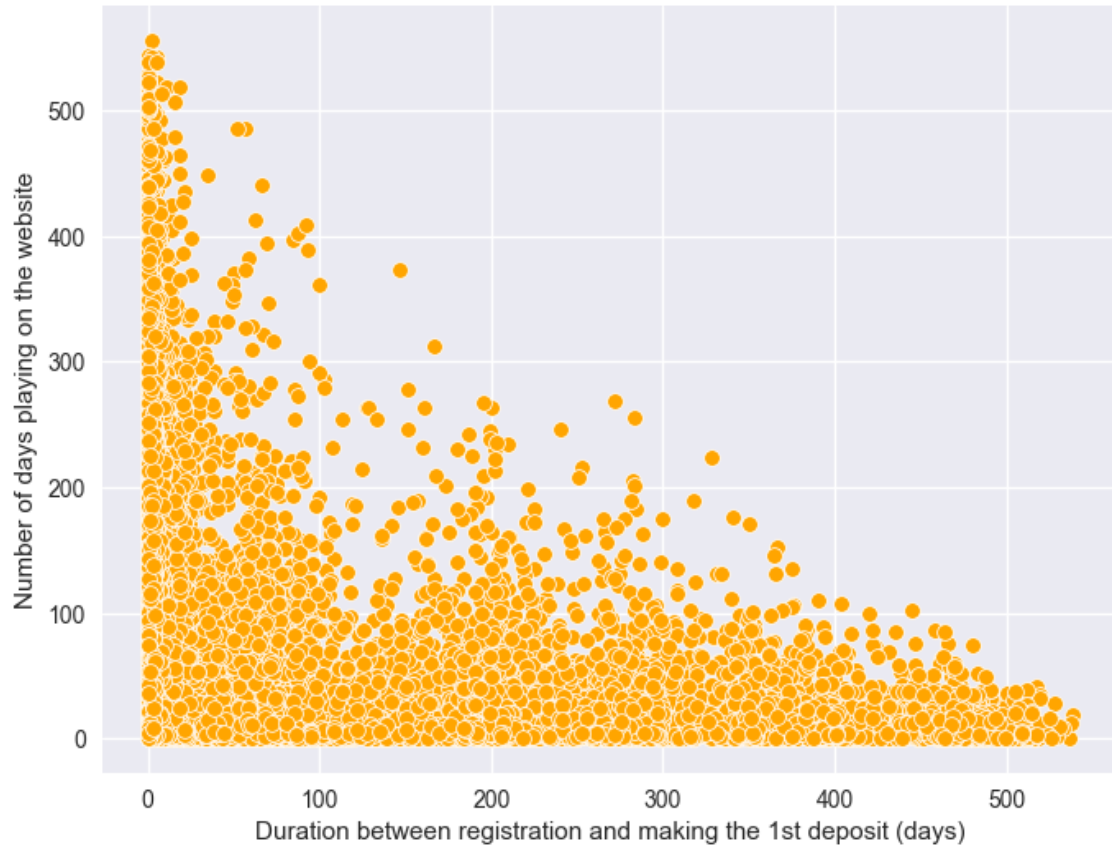
- 12.7% (5904) of customers played on the website for **more than 100 days**
- 80% of customers who played for over 100 days made their **1st deposits** within **1 week** after registering their accounts.

# 2. Data understanding – the anomalies (big-time losing players)

- 72 big-time losing players (highly beneficial for bwin) were identified using the Z-score method

- Overall, big-time losing players had profits below -17,000 euros

- The ratio of live bets is double that of winning players (62.1% compared to 30.3%), and the ratio of live stake is 78.4%, compared to 53% of winning players ➡ very different betting tendencies

3. Modelling – Feature Selection

3 features are selected for the final modeling, all showing a high correlation to the label:

- days_total

- bets_total,

- stake_total
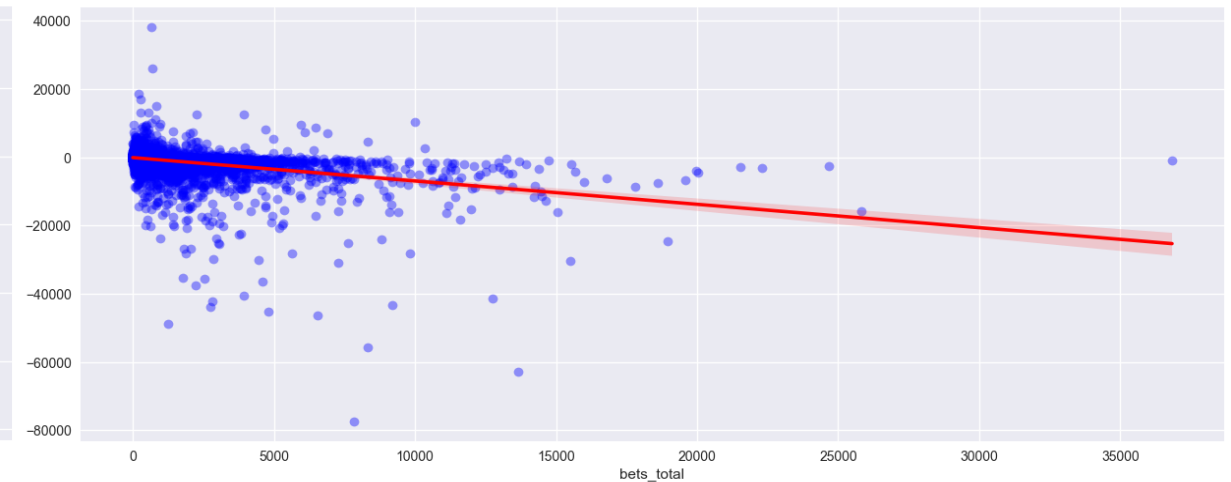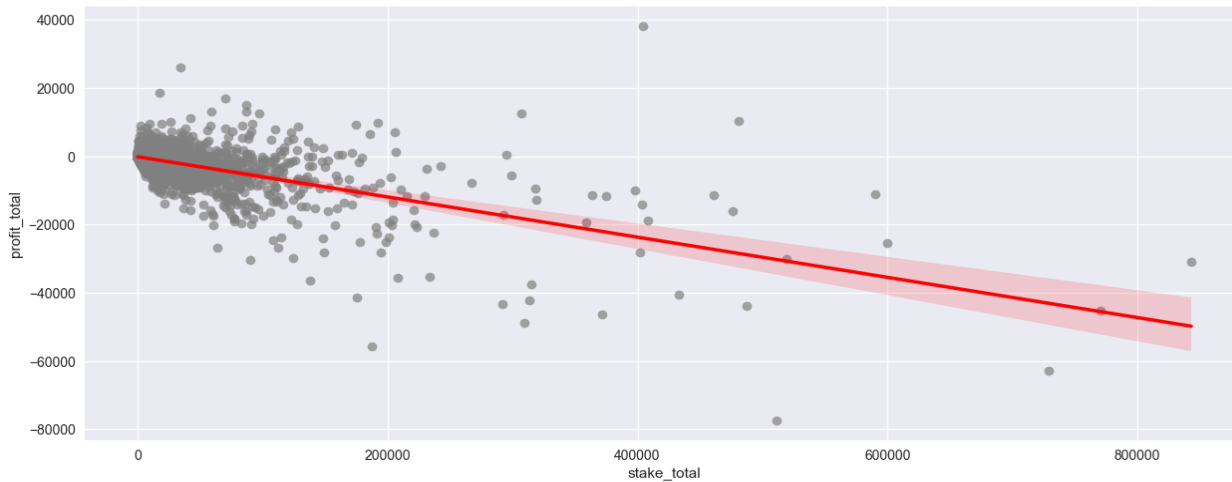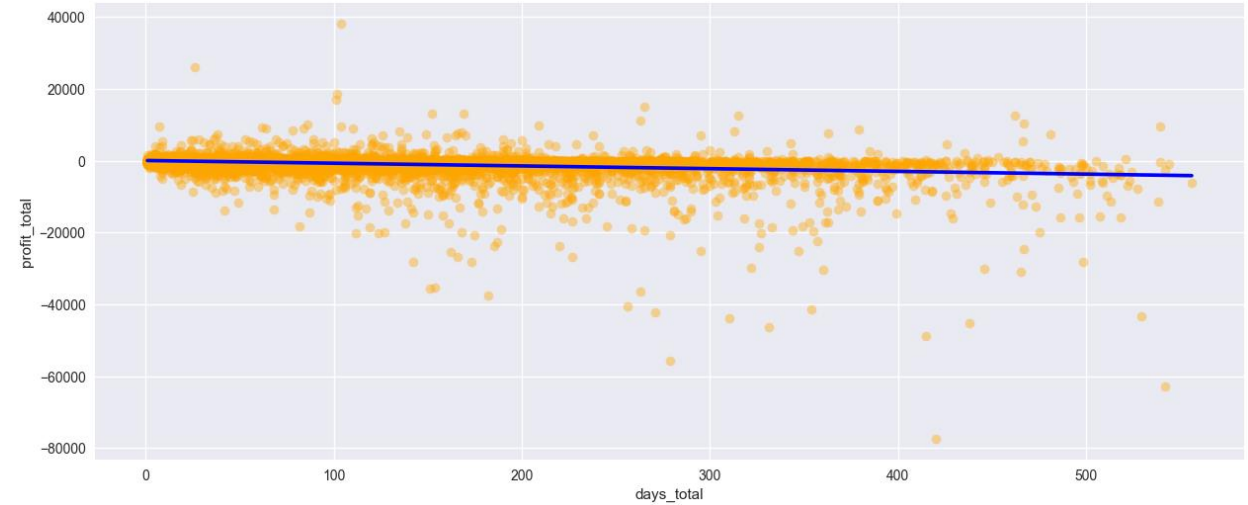
3. Modelling - Results

The dataset is divided into training data (80%) and testing data (20%) using train_test_split function from sklearn.

Polynomial Regression and Random Forest Regression are used.

Results

Polynomial Regression (degree 3):

- Training error: 253.66

- Test error: 240.1252

Random Forest Regression (n = 150):

- Training error: 98

- Test error: 256.3

Learning curve of 2 models: