Portefolio

Mette

2024-01-18

Table of contents

1 Introduction

In the dynamic world of sports, the ability to anticipate game attendance is crucial for effective stadium management, marketing strategies, and fan engagement. This project presents a comprehensive prediction model specifically designed for estimating the number of viewers attending home games of Viborg FF, a prominent Danish soccer club. By leveraging data-driven approaches and advanced analytical techniques, the model aims to provide accurate forecasts that can assist in optimizing the overall matchday experience.

Viborg FF, with its rich history and dedicated fanbase, presents an interesting case study for attendance prediction. The club's performance, league standings, opponent strength, ticket pricing, weather conditions, and special events are among the numerous factors that can influence game attendance. This project delves into these aspects, utilizing historical data and statistical methods to build a robust predictive framework.

The objective of this endeavor is not only to aid Viborg FF in planning and resource allocation but also to contribute to the broader field of sports analytics. By offering insights into fan behavior and attendance patterns, this model seeks to enhance the understanding of factors driving spectator turnout in soccer, particularly within the context of Danish football.

In the following sections, the methodology, data sources, analytical techniques, and results of the prediction model will be thoroughly discussed, demonstrating the model's effectiveness and potential applications.

2 Analysis

The project embarked on developing a predictive model for estimating the attendance at Viborg FF home games with a structured and methodical approach. The initial stage involved importing data through API calls and web scraping, a technique that proved essential in gathering a diverse range of public data. This included team performance statistics, league standing, TV channels, weather conditions, and details of special events.

2.1 Reading libraries

```
pacman::p_load("rvest", "purrr", "data.table", "stringr", "tidyr", "zoo",
               "DT", "janitor", "chron", "rlist", "rjstat",
               "rjson", "Rcrawler", "magrittr", "gapminder",
               "Lahman", "maps", "lubridate", "pryr", "hms", "hexbin",
               "feather", "htmlwidgets", "broom", "pander", "modelr",
               "XML", "httr", "jsonlite", "microbenchmark",
               "splines", "ISLR2", "MASS", "testthat", "leaps", "caret",
               "RSQLite", "class", "nasaweather",
               "fueleconomy", "viridis", "boot", "glmnet", "pls", "tidyverse", "dplyr","co
```

2.2 Importing data

2.3 CRISP-DM metodology

2.3.1 Business understanding

2.3.2 Data Understanding

2.3.3 Data Preparation

2.3.4 Modeling

2.3.5 Evaluation

2.3.6 Implementation