

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

FILIPE MORANDI SOARES

DESENVOLVIMENTO DE UM PROGRAMA DE SUPORTE A ANÁLISE DE BALANÇO
COMPETITIVO DE CAMPEONATOS

RIO DE JANEIRO

2022

FILIPPE MORANDI SOARES

DESENVOLVIMENTO DE UM PROGRAMA DE SUPORTE A ANÁLISE DE BALANÇO
COMPETITIVO DE CAMPEONATOS

Trabalho de conclusão de curso de graduação apresentado ao Departamento de Ciência da Computação da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Geraldo Bonorino Xexeo

RIO DE JANEIRO

2022

CIP - Catalogação na Publicação

S676d Soares, Filipe Morandi
Desenvolvimento de um programa de suporte a
análise de balanço competitivo de campeonatos /
Filipe Morandi Soares. -- Rio de Janeiro, 2022.
48 f.

Orientador: Geraldo Bonorino Xexeo.
Trabalho de conclusão de curso (graduação) -
Universidade Federal do Rio de Janeiro, Instituto
de Computação, Bacharel em Ciência da Computação,
2022.

1. Balanço competitivo. 2. Gráfico. 3. Software.
4. Campeonato. 5. Análise de dados. I. Xexeo,
Geraldo Bonorino, orient. II. Título.

FILIPPE MORANDI SOARES

DESENVOLVIMENTO DE UM PROGRAMA DE SUPORTE A ANÁLISE DE BALANÇO
COMPETITIVO DE CAMPEONATOS

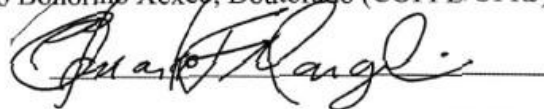
Trabalho de conclusão de curso de graduação
apresentado ao Departamento de Ciência da
Computação da Universidade Federal do Rio de
Janeiro como parte dos requisitos para obtenção
do grau de Bacharel em Ciência da Computação.

Aprovado em 6 de Maio de 2022.

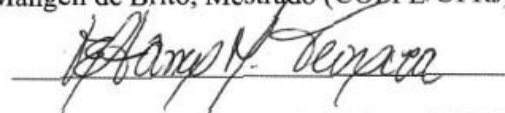
BANCA EXAMINADORA:



Geraldo Bonorino Xexeo, Doutorado (COPPE/UFRJ)



Eduardo Freitas Mangeli de Brito, Mestrado (COPPE/UFRJ)



Eldanae Nogueira Teixeira, Pós-Doutorado (COPPE/UFRJ)

Dedicatória: Para a Santíssima Virgem de
Guadalupe, porque sem ela nunca teria terminado
esse Trabalho de Conclusão de Curso.

AGRADECIMENTOS

Meus agradecimentos ao meu orientador Geraldo Bonorino Xéxeo, que me ajudou desde a escolha do tema do trabalho até a elaboração do projeto final e a sua formatação. Meus agradecimentos ao Eduardo Mangeli que me deu os recursos necessários, em termos de bibliotecas do Linguagem Python e código base para desenvolvimento desse projeto.

RESUMO

Esse trabalho pretende apresentar um programa que dê suporte a análise do balanço competitivo de campeonatos, desenvolvido com o paradigma de programação orientada a objeto. Foram selecionados algumas métricas de avaliação de balanço competitivo e uma de avaliação estética de jogos. Também foram selecionados um conjunto de dados contendo o histórico de partidas de campeonatos importantes de alguns países, os quais foram tratados e transformados em arquivos JSON. Cada arquivo armazena as mudanças do ranking de um campeonato ao longo desse. O programa desenvolvido abre esses arquivos, extrai esses dados, trata-os e armazena-os em uma classe que faz a interface entre os dados e a classes que calculam o valor das métricas. Então essas classes são usadas para calcular os valores das métricas que serão utilizadas para gerar histogramas, gráficos que representem a evolução do valor da métrica nos diferentes anos de um campeonato ou o histórico da edição de um campeonato em um ano por pontos ou posições, além de poder consultar o valor de uma métrica para um campeonato.

Utilizar a terceira pessoa do singular, os verbos na voz ativa e evitar o uso de símbolos e contrações que não sejam de uso corrente.

Palavras-chave: balanço competitivo; gráfico; programa; avaliação estética, campeonato; análise de dados

ABSTRACT

This work intends to present a program that supports competitive balance analysis of championship, developed with the object-oriented paradigm. Some metrics used for balance competitive evaluation and one for aesthetic evaluation of games were selected. Also several datasets were selected, containing the match history of important championships of some countries, and processed in JSON files. Each file stores the changes in ranking of a championship over the course of that one. The developed program open these files, extract these data, process and store them in a class that interfaces between data and classes that calculate the metric values. So, these classes are used to calculate the metrics' values that will be used to make histograms, graphs that represent the metric's value evolution on different years of a championship or the history, by points or by positions, of a edition of a championship at a year, in addition to being able to consult the value of a metric for a championship.

Keywords: competitive balance; graph; program; aesthetic evaluation; championship; data analysis

SUMÁRIO

INTRODUÇÃO.....	9
1.1 MOTIVAÇÃO.....	10
1.2 PARTES DESSE PROJETO FINAL.....	11
2 TECNOLOGIAS UTILIZADAS.....	12
2.1 PYTHON 3.8.....	12
2.1.1 wxPython.....	13
2.1.2 matplotlib.....	13
2.2 JSON : JAVASCRIPT OBJECT NOTATION.....	13
3 A ANÁLISE DE JOGOS.....	14
3.1 MÉTRICAS UTILIZADAS NO PROGRAMA.....	15
3.1.1 Accumulated Points Difference.....	15
3.1.2 CX Ratio.....	16
3.1.3 CX Index of Competitive Balance.....	16
3.1.4 Drama By Paths.....	18
3.1.5 Lead Change.....	18
3.1.6 Uncertainty by Probability Distributions Distance.....	19
3.1.7 Overall Aesthetic Metric.....	22
3.2 MOTIVAÇÃO DAS ESCOLHAS DAS MÉTRICAS.....	22
4 PROPOSTA DE ARQUITETURA.....	23
4.1 ARQUITETURA DO PROGRAMA.....	23
4.1.1 Classe Championship.....	24
4.1.2 Classe PontosCorridosCampeonato.....	25
4.1.3 Classe MeasureCall.....	26
4.1.4 Classe MeasureTemplate.....	27
4.2 DADOS.....	28
4.2.1 Conjunto de dados.....	28
4.3 TRATAMENTO DOS DADOS.....	29
4.4 EXEMPLO DE ARQUIVO JSON.....	29
4.4.1 Estrutura de diretório.....	30

5 EXEMPLO DO PROGRAMA ATUAL.....	31
6 ANÁLISE DE DADOS DOS CAMPEONATOS.....	35
6.1 CAMPEONATO ALEMÃO.....	36
6.2 CAMPEONATO ITALIANO.....	38
6.3 CAMPEONATO ESPANHOL.....	40
6.4 CAMPEONATO PORTUGUÊS.....	42
6.5 CAMPEONATO INGLÊS.....	44
7 CONCLUSÃO.....	46
REFERÊNCIAS.....	47

INTRODUÇÃO

Os jogos são um componente presente em toda cultura e em todo o planeta, existentes a muitos séculos. Foram essenciais para o divertimento e para a distração da população e em alguns casos, até mesmo para o preparo físico de soldados, como se relata um esporte muito semelhante ao futebol na China antiga (“Cuju”, 2019). Além de todo o divertimento proporcionado, eles possuem um impacto econômico e social, mesmo em tempos antigos. O mercado do esporte – venda de ingressos para partidas, comprar e vender jogadores, venda de itens com o símbolo do time etc. – é algo que esteve de maneira mais presente nos tempos modernos e continua popular atualmente. Sandy, Sloane and Rosentraub (SANDY; SLOANE; ROSENTRAUB, 2004) dizem:

(...) a indústria do esporte levanta questões econômicas fascinantes e (...) esportes têm sido um componente de alto perfil de todas as sociedades por mais de 4000 anos. Os Gregos e os Romanos fizeram do esporte uma parte central de sua sociedade. Eles construíram suas cidades com instalações esportivas como peças centrais. Civilização Maia na América Central também enfatizou o esporte. O entusiasmo público dos espectadores de esportes continua inabalável nos tempos modernos.

Uma parte do mercado do esporte são os campeonatos e esses, sejam de jogos eletrônicos ou não, sempre foram e continuam muito populares. De acordo com a FIFA, um pouco mais da metade do mundo assistiu a copa mundial de 2018 da FIFA (“More than half the world watched record-breaking 2018 World Cup”, [s.d.]); De acordo com a CNBC, cerca de 100 milhões de pessoas assistiram a final do Campeonato Mundial de League of Legends em 2018 (PEI, 2019); A final da NBA teve cerca de 15,14 milhões de espectadores americanos (“NBA vs WNBA”, [s.d.]); jogos de batalha jogador contra jogador como Call of Duty Mobile (“Call of Duty®”, 2019) e Brawl Stars (“Brawl Stars - Android app on AppBrain”, 2018) possuem mais de 100 milhões de downloads no Google Play.

Dependendo das circunstâncias, é muito importante para a manutenção do mercado que os participantes (ou times) não possuam nenhuma vantagem grande sobre os outros participantes (ou times). Isso é confirmado pelo artigos de Meehan et al (MEEHAN; NELSON; RICHARDSON, 2007) e outro de Rascher and Solmes (RASCHER; SOLMES, 2007), os quais mostram que partidas da Major League Baseball (MLB) e da NBA, respectivamente, possuíam menor presença de torcedores quando o resultado do jogo era mais certo. Porém, Buraimo mostra na conclusão do seu artigo que na English Premier League, os torcedores preferem uma partida com resultado previsível (BURAIMO; SIMMONS, 2008).

Em geral, o balanço competitivo, o oposto de excesso de vantagens por parte de um time, é necessário para manter a competitividade e a atratividade do esporte. Devido a essa necessidade,

procura-se analisar os campeonatos e medir a balanço competitivo destes. Para isso, são usados diversos métodos, os quais incluem analisar a parte financeira (fatia de mercado, dinheiro recebido para manter o clube et cetera), usar ferramentas estatísticas para avaliar as vitórias e derrotas de cada time e usar métricas para avaliação de balanço competitivo. Esse último método é o qual o presente trabalho final vai utilizar. Essas métricas se assemelham muito as métricas de produtividade, como a métrica de Herfindhal-Hirschman, que utiliza a fatia de mercado de cada empresa para tentar medir a monopolização de um mercado, e são muito simples de serem aplicadas. O desenvolvimento de métricas é uma das áreas trabalhadas pelo laboratório de pesquisa Ludes. Esse laboratório faz parte da PESC na COPPE da UFRJ, e tem como finalidade “investigar metodologias de desenvolvimento de jogos e como aplicar jogos para ajudar a sociedade, indo em direção à criação de jogos com propósito” (“LUDES | Laboratório de Ludologia, Engenharia e Simulação”, [s.d.]). O laboratório não trabalha apenas com métricas para avaliação de balanço competitivo, mas trabalha com pesquisas que busquem medir e avaliar as partidas de um jogo e a experiência de seus espectadores.

O presente trabalho relaciona-se a linha de pesquisa de Eduardo Mangeli, membro do Ludes, autor da dissertação “An Aesthetic Metric For Multiplayer Turn-Based Games”. Nessa dissertação, Mangeli procura desenvolver uma métrica para avaliar a estética de qualquer jogo multijogador baseado em turnos. Com a avaliação da estética de um jogo, procura-se verificar quais partidas de um jogo foram mais agradáveis ou não, procurando manter-se fiel a dinâmica das emoções dos espectadores diante da partida de um jogo. No caso do presente projeto final, pretende-se usar outras métricas para avaliação do balanço competitivo, além dessa métrica desenvolvida por Mangeli para o mesmo fim.

O presente projeto final procura apresentar um programa que permita a pessoas interessadas na avaliação do balanço competitivo ou detecção de ausência de balanço competitivo gerar visualizações dos dados sobre um ou mais campeonatos e, assim, dar suporte a seus trabalhos e ou pesquisas. Para atingir esse fim, pretende-se fazer um programa em Python que traga o necessário para o usuário poder gerar essas visualizações, mas sem limitá-lo as métricas e tipos de campeonatos já contemplados no programa, permitindo ao usuário adicionar outras métricas ao mesmo.

1.1 Motivação

A avaliação do balanço competitivo é muito importante para atratividade do esporte, porque um esporte sem balanço competitivo tende a gerar descontentamento aos torcedores, tornando o esporte cada vez mais previsível ou, ainda, monopolizado por um ou mais times. Ter meios de

identificar a ausência de balanço competitivo é importante para que se possa verificar quais campeonatos apontam um maior desbalanço competitivo e, assim, examinar mais a fundo as possíveis causas desse desbalanceamento. Para que os pesquisadores possam ter meios de avaliar campeonatos ou criar outras métricas e compará-las com outras já existentes, o presente projeto final se propõe a produzir um programa que dê uma ferramenta para esses pesquisadores trabalharem. A possibilidade de criar outras métricas poderá auxiliar especialmente o Ludes, permitindo que esse laboratório desenvolva novas métricas, comparar essas métricas com outras já implementadas e verificar se elas estão medindo bem o balanço competitivo de um campeonato e outro fator de um campeonato.

1.2 Partes desse Projeto Final

Esse projeto final é composto de seis capítulos além desse.

No segundo capítulo, as tecnologias utilizadas são brevemente apresentadas e é explicado a motivação pelo uso de cada uma delas.

No terceiro capítulo é apresentado um conceito de jogo, a linha de pesquisa sobre Competitividade e as métricas utilizadas no presente trabalho.

No quarto capítulo é apresentada a proposta de arquitetura do programa, apresentando os artefatos computacionais implementados e a funcionalidade de cada um dentro do programa. Também é apresentado como os dados estão armazenados dentro do programa e a estrutura de diretório dos arquivos que contém esses dados.

No quinto capítulo, algumas imagens são apresentadas para mostrar as funcionalidades do programa, além de mostrar como é a organização da tela e o comportamento da tela para cada campo.

No sexto capítulo é feita uma breve análise sobre os campeonatos que foram usados para exemplificar o comportamento do programa.

No sétimo capítulo, o projeto final é concluído, contatando-se o que o presente projeto entregou como produto final.

2 TECNOLOGIAS UTILIZADAS

Neste capítulo, será apresentado as tecnologias utilizadas para desenvolver o programa proposto no presente artigo.

A seção 2.1 apresentará a linguagem de programação Python; a seção 2.1.1 apresentará a biblioteca wxPython, utilizada para implementar a interface; a seção 2.1.2 apresentará a biblioteca matplotlib, utilizada para gerar os gráficos. Na seção 2.2, será apresentado o formato de arquivo JSON e, por fim, será apresentado a razão do uso de cada tecnologia.

2.1 Python 3.8

Python é uma linguagem de programação com tipos fortes e dinâmicos e compilação em tempo de execução. (“Why is Python a dynamic language and also a strongly typed language - Python Wiki”, [s.d.]). Além dessas características, o interpretador dessa linguagem está disponível tanto para Windows como para Linux, o que geralmente permite a execução dos códigos escritos em ambas as plataformas.

A linguagem foi feita para que fosse uma linguagem para escrever scripts com o propósito de processarem de dados e/ou mostrar alguma visualização e terminar. Possui uma sintaxe simples, de forma que é rápido escrever um código na linguagem Python.

A versão da linguagem utilizada para o desenvolvimento do programa dá suporte ao paradigma de programação orientada a objeto, o qual foi utilizado nesse programa.

A linguagem Python foi escolhida por diversas razões: expansibilidade, liberdade, simplicidade e portabilidade.

Por ser uma linguagem compilada em tempo de execução permite que se ofereça uma arquitetura inicial para o usuário de forma que ele possa expandir o programa inserindo mais métricas e outros tipos de campeonato, porém, sem limitar que pessoas mais experientes com a linguagem e suas bibliotecas possam mudar a arquitetura do programa da maneira que desejarem.

Por ser uma linguagem de tipos fortes e dinâmicos, escrever o código de mais uma métrica ou de outro tipo de campeonato é algo mais simples, tornando o trabalho de escrevê-lo menos custoso.

Por ser uma linguagem multiplataforma, permitindo que o programa seja executado em múltiplas plataformas, apesar de o programa, no estado atual, permitir a execução somente na plataforma Windows.

2.1.1 wxPython

A biblioteca wxPython é um conjunto de ferramentas para criação de interface gráfica para a linguagem Python que funciona em múltiplas plataformas. Ela oferece a implementação de janela, botão, caixas de seleção, gerenciadores de layout e vários outros recursos que permitem criar janelas interativas, organizadas, altamente funcionais e robustas. (TEAM, 2017).

A biblioteca wxPython é uma biblioteca com muitos recursos que facilitam a organização dos elementos na tela e vários elementos de janela já implementados, além de permitir a implementação da detecção de eventos (como o clique de um botão) e a resposta a esses eventos.

2.1.2 matplotlib

A biblioteca matplotlib é uma biblioteca abrangente que permite a geração de visualizações estáticas, animadas e interativas em Python, como histogramas, gráficos de linha, gráficos de seção circular, entre outros, incluindo uma série de configurações (“Matplotlib — Visualization with Python”, [s.d.]).

Além de oferecer uma série de opções de gráficos diferentes com abertura a personalização dos mesmos, a biblioteca matplotlib possui uma integração com a biblioteca wxPython, permitindo implementar a inclusão do gráfico na janela do programa como um painel.

2.2 JSON : JavaScript Object Notation

JSON é um formato de troca de dados legível adequado para estruturas de dados simples derivado da linguagem de programação JavaScript. Apesar disso, é um formato independente de linguagem. Ele é razoavelmente flexível e facilmente tratável (“JSON”, [s.d.]).

Como os dados sobre os campeonatos guardados no programa são simplesmente uma série de conjuntos de tuplas, guardar essas informações no formato JSON é mais simples e direto.

3 A ANÁLISE DE JOGOS

Jogos, conforme o website do laboratório de pesquisa Ludes, podem ser definidos da seguinte maneira (“Pesquisa”, [s.d.]):

Jogos são atividades sociais e culturais voluntárias, significativas, fortemente absorventes, não produtivas, que se utilizam de um mundo abstrato, com efeitos negociados no mundo real, e cujo desenvolvimento e resultado final é incerto, onde um ou mais jogadores, ou equipes de jogadores, modificam interativamente e de forma quantificável o estado de um sistema artificial, possivelmente em busca de objetivos conflitantes, por meio de decisões e ações, algumas com a capacidade de atrapalhar o adversário, sendo todo o processo regulado, orientado e limitado, por regras aceitas, e obtendo, com isso, uma recompensa psicológica, normalmente na forma de diversão.

Apesar de sua realidade altamente subjetiva, os jogos possuem realidades objetivas e verificáveis como, por exemplo, a necessidade dos jogadores em possuírem oportunidades iguais de ganhar e isso é geralmente expresso como “balanço competitivo”. Esse termo não possui um conceito comumente definido, o que torna difícil trabalhar com ele (KRINGSTAD; GERRARD, 2004): O balanço competitivo é um termo genérico usado para se referir aos diferentes aspectos de uma competição (QUIRK; FORT, 1997); O balanço competitivo é um fenômeno complexo com muitas dimensões (ZIMBALIST, 2003); O balanço competitivo é um termo genérico para diferentes aspectos de incerteza de resultado e significância da competição (GERRARD, 2004). Porém, Kringstad and Gerrard identificam que “a ideia governante na literatura concernindo balanço competitivo é que esse tem muitas ligações com as diferenças de qualidade esportiva entre os times em um liga/torneio”(KRINGSTAD; GERRARD, 2004).

O balanço competitivo é algo muito importante para o mercado dos esportes, pois ele afeta a participação dos torcedores nos estádios e o envolvimento desses com o esporte. “Incerteza assim como qualidade do futebol cria interesse” (SANDY; SLOANE; ROSENTRAUB, 2004). Em um cenário monopolizado por um time, embora o time sempre ganhe, os torcedores do time vencedor começam a perder interesse pelo time vencedor, pois o resultado do jogo se torna demasiadamente previsível, então, em toda partida, será verificado uma vitória fácil de um time sobre o outro, tornando a partida pouco interessante. Como dizem Quirk e Fort, a “preservação do balanço competitivo é um problema legítimo para ligas de esportes” (QUIRK; FORT, 1997).

Esse tipo de pensamento é confirmado por diversos autores do século XX. Neale(1964) argumenta a favor do “paradoxo de Louis-Schmelling”, sugerindo que competição esportiva é mais lucrativo que monopólio esportivo. Szymanski afirma que “sem algum grau de balanço competitivo, os torcedores perderão o interesse na competição”. Dobson e Goddard afirmam que “... um nível

razoável de balanço competitivo ... é necessário para produzir uma genuína incerteza de resultado.” Zimbalist(2003b) afirma:

O sucesso de uma liga é, em parte, afetado pelo grau da incerteza de resultado das suas competições e campeonatos sazonais, ou, afirmando de uma maneira diferente, por um grau de balanceamento entre os times.

Por sua vez, Gerrard 2004 afirma:

Manutenção da incerteza suficiente de resultado é frequentemente visto como uma requisição necessária para viabilidade esportiva e financeira a longo prazo dos torneios e requer que os organizadores do torneio fomentem um grau relativamente alto de balanço competitivo.

No cenário dos esportes, vários pesquisadores procuram verificar a existência de um ou mais times dominantes ao longo de vários campeonatos ou a dominação de um time dentro de um campeonato. Para esse fim, eles procuram utilizar métricas que permitam identificar os campeonatos que sejam desbalanceados competitivamente, a fim de poder verificar quais campeonatos possuem melhor balanço competitivo.

Na próxima seção serão apresentadas algumas métricas que serão implementadas no programa e utilizadas para avaliar os campeonatos.

3.1 Métricas utilizadas no programa

3.1.1 Accumulated Points Difference

A métrica de diferenças acumuladas de pontos é uma métrica desenvolvida por Gasparetto et al (GASPARETTO; BARAJAS, 2015) e foi feita especificamente para campeonatos de Round-Robin em que o time vencedor recebe três pontos por vencer uma partida, um ponto para cada time em caso de empate e nenhum ponto para o time perdedor. O cálculo de desbalanceamento máximo (descrito abaixo com Unbalance) foi formulado para que a métrica possa ser usada para comparar campeonatos conforme descrição anterior com números diferentes de times.

$$Unbalance_{max}(N) = 6 * (N - 1)$$

$$APD(N) = \frac{\sum_{i=1}^N TP_i - TP_{i-1}}{Unbalance_{max}(N)} \times 100$$

Figura 1. Fórmula da métrica Accumulated Points Difference

Quando há o desbalanceamento máximo, a métrica é igual a um e quando há o balanceamento total, a métrica é igual a zero. Na primeira situação, o primeiro time teria que ganhar

de todos os outros times, o segundo de todos os outros times, exceto o primeiro, e assim por diante. O primeiro time faria então $6*(N-1)$ pontos, o segundo $6*(N-2)$ pontos, o terceiro $6*(N-3)$ e assim por diante até o último que teria zero pontos. Com isso, a diferença entre cada time daria exatamente seis. Como a diferença é calculada entre posições consecutivas, o valor total daria $6*(N-1)$, que é exatamente o valor calculado de desbalanceamento máximo. O caso de balanceamento total aconteceria havendo empate entre todos os times.

3.1.2 CX Ratio

Na literatura, esta métrica é apresentada como C4 Ratio ou como C5 Ratio, mas para fins de simplificação, ela será apresentada de maneira generalizada. A fórmula do CX Ratio é a seguinte, sendo que P_i é o número de pontos do time na i -ésima posição do campeonato e N o número de times:

$$C_X \text{ Ratio} = \frac{\sum_{i=1}^X P_i}{\sum_{i=1}^N P_i}$$

Figura 2. Fórmula da métrica CX Ratio

Essa métrica calcula a razão da soma dos pontos dos X primeiros times para a soma dos pontos de todos os times participantes. Quanto mais a razão se aproxima de 1, menor é o balanço competitivo. Quanto mais se aproxima de zero, melhor é o balanço competitivo.

Essa métrica muita usada na literatura. Ela aparece em diversos artigos como os de Naghsbandi et al (NAGHSHBANDI et al., 2011), Silva et al (SILVA et al., 2018), Dejonghe (DEJONGHE, 2006) e Feddersen (FEDDERSEN, 2006). Além disso, essa métrica é utilizada também para fazer análises no campo financeiro para quantificar a monopolização do mercado. Ela foi usada pela primeira vez por Michie & Oughton (MICHIE; OUGHTON, 2004) para campeonatos. Porém, enquanto a métrica antes poderia variar entre zero e um, entre um cenário altamente competitivo e um cenário altamente monopolizado, agora ele só pode variar dentro de um intervalo bem definido, pois depende de como são distribuídos pontos nas partidas do campeonato.

3.1.3 CX Index of Competitive Balance

Assim como no caso do CX Ratio, esta métrica é usada como C4 Index of Competitive Balance ou como C5 Index of Competitive Balance na literatura, mas como a ideia por baixo das duas é a mesma, ela será apresentado de maneira generalizada. A métrica CX Index of Competitive Balance é calculada da seguinte forma, sendo P_i o número de pontos do time na i -ésima posição do campeonato e N o número de times:

$$C_X Index = \frac{C_X Ratio}{\frac{X}{N}} \times 100$$

Figura 3. Fórmula da métrica CX Index of Competitive Balance

Mas, essa fórmula pode ser também escrita da seguinte forma:

$$C_X Index = \frac{\frac{\sum_{i=1}^X P_i}{N}}{\frac{X}{N}} \times 100 = \frac{\sum_{i=1}^X P_i}{\sum_{i=1}^X P_i} \div \frac{X}{N} \times 100 = \frac{\sum_{i=1}^X P_i}{\frac{X}{N}} \times 100$$

$$C_X Index = \frac{\text{média dos pontos dos X primeiros times}}{\text{média dos pontos de todos os times}} \times 100$$

Figura 4. Outra definição da fórmula da métrica CX Index of Competitive Balance

Essa fórmula calcula a razão entre a média dos X primeiros times em relação a média dos pontos de todos os times. Se a média é a mesma, significa que os X primeiros times estão afastados dos outros times na mesma proporção que todos os times estão afastados entre si, o que se trataria supostamente de uma posição de equilíbrio total. Nessa métrica, quanto mais próximo de cem, melhor é o balanço competitivo. Por outro lado, quanto cada vez mais é maior que cem, pior é o balanço competitivo. Essa métrica foi criada por Michie & Oughton a partir da métrica CX Ratio (MICHIE; OUGHTON, 2004) para avaliação do balanço competitivo de campeonatos e ela é usada com bastante frequência na literatura, como por Naghbandi et al (NAGHSHBANDI et al., 2011), Silva et al (SILVA et al., 2018), Dejonghe (DEJONGHE, 2006), Feddersen (FEDDERSEN, 2006) e Drummond et al (DRUMMOND; JÚNIOR; SHIKIDA, 2013).

3.1.4 Drama By Paths

A métrica de Drama By Paths é calculada comparando a evolução do ranking do vencedor em um campeonato observado com a evolução do ranking de um vencedor em um campeonato de drama máximo. Em um campeonato de drama máximo, o vencedor do campeonato começa do último lugar depois do primeiro turno, passa pelas posições do ranking uma por uma e vence em primeiro lugar após o último turno. Então, para definir essa função de Drama By Paths, é necessário primeiramente de uma função que forneça qual a posição que um time vencedor deve estar em um turno m em um campeonato com um conjunto P de times e um conjunto M de turnos, onde existe um vencedor P_w e uma função P_f que dado um time retorna a sua posição no campeonato observado. A definição dessas duas funções pode ser vistas a abaixo:

$$MDP(m) = \left\lceil |P| + \frac{(1-|P|)(m-1)}{|M|-1} \right\rceil$$

$$D(x) = \frac{| \{m | P_f(P_w, m) > 1\} |}{|M|-1} \left(1 - \sum_{m=1}^M \frac{|P_f(P_w, m) - MDP(m)|}{(|P|-1)(|M|-1)} \right)$$

Figura 5. Fórmula da métrica Drama By Paths junto com a fórmula da função auxiliar

Os dois campeonatos são comparados fazendo a diferença absoluta entre a posição do vencedor no campeonato observado e a posição do vencedor no campeonato de drama máximo. Também a função de drama dá um peso diferente a soma das diferenças absolutas quando o vencedor já estava no primeiro colocado em turnos anteriores ao último. De acordo com Mangeli (MANGELI, 2016), quando o time consegue atingir o primeiro lugar, que é o objetivo, antes do último turno, o drama do campeonato diminui.

3.1.5 Lead Change

A métrica de Lead Change é a média aritmética de duas fórmulas que variam entre zero e um. A primeira fórmula conta quantos líderes diferentes se teve ao longo do jogo, desconsiderando quem já era líder desde o início do jogo e a segunda conta quantas mudanças de liderança teve ao longo do jogo. A primeira fórmula divide o número contado pelo número de jogadores menos um e

a segunda divide o número contado pelo número de partidas menos um. Ambas calculam a raiz quadrada das divisões calculadas. Conforme Mangeli, um novo líder em um campeonato no qual houve muitos líderes não causa tanto apelo quanto um novo líder em um campeonato por enquanto havia apenas um só líder (MANGELI, 2016). Isso vale para as mudanças de liderança. Porém, como existe a possibilidade de todos os times terem sido líderes e em todo turno ter ocorrido a mudança de líder, a métrica continua a variar entre zero e um, como deve ser toda função de lógica fuzzy. A fórmula da métrica de Lead Change pode ser expressa da seguinte forma, sendo L o conjunto dos líderes ao longo do campeonato e Lchange o conjunto dos turnos nos quais ocorreram mudança de liderança.

$$L(x) = \frac{\sqrt{\frac{|L|-1}{|P|-1}} + \sqrt{\frac{|LChange|}{|M|-1}}}{2}$$

Figura 6. Fórmula da métrica Lead Change

3.1.6 Uncertainty by Probability Distributions Distance

Agora, a métrica de Uncertainty é calculada com base na distância das distribuições de probabilidade entre a distribuição de probabilidade de vitória de cada time em um turno e a distribuição de probabilidade ideal, na qual todos os times têm igual chance de ganhar. Para isso, existe distância de Hellinger, a qual é calculada da seguinte forma, dada duas distribuições de probabilidades discretas $W = (w_1, \dots, w_k)$ e $Q = (q_1, \dots, q_k)$:

$$H(W, Q) = \sqrt{\frac{1}{2} \sum_{i=1}^k (\sqrt{w_i} - \sqrt{q_i})^2}$$

Figura 7. Fórmula da distância de Hellinger

O cálculo do valor da métrica é feito calculando a distância de Hellinger para a situação na qual se há certeza do vencedor e a distância de Hellinger entre a distribuição discreta de probabilidade de um turno m e a distribuição discreta de probabilidade onde há total incerteza do vencedor. A distribuição de probabilidade na qual se há certeza do vencedor será representado por C , a outra na qual se há total incerteza será representado por U e a distribuição de probabilidade de um turno m será representado por Q_m , também será usada uma função ρ que retorna um valor entre zero e um dado o time p e um turno m , a qual será descrita com mais detalhes posteriormente nessa seção.

A distribuição C possui o valor um para apenas um time e para todos os outros times é retornado o valor zero. A distribuição U possui exatamente o valor da razão entre um e o número de times para todo o time. Ou seja, não importa o time, a chance de vitória é a mesma. Cada valor na distribuição Qm é simplesmente ditada pela função ρ , dado o time p e o turno m. O cálculo da distância de Hellinger entre U e C resulta no seguinte:

O cálculo da distância de Hellinger entre U e C resulta no seguinte:

$$H(U, C) = \sqrt{\frac{1}{2} \left[\left(\sqrt{1} - \frac{1}{\sqrt{|P|}} \right)^2 + (|P| - 1) \left(\frac{-1}{\sqrt{|P|}} \right)^2 \right]}$$

$$H(U, C) = \sqrt{\frac{1}{2} \left[1 - \frac{2}{\sqrt{|P|}} + \frac{1}{|P|} + (|P| - 1) \left(\frac{1}{|P|} \right) \right]}$$

$$H(U, C) = \sqrt{\frac{1}{2} \left(2 - \frac{2}{\sqrt{|P|}} \right)}$$

Figura 8. Desenvolvimento da aplicação da fórmula da distância de Hellinger para a distribuição ideal U e a pior distribuição C

O cálculo da distância de Hellinger entre U e Q resulta no seguinte:

$$H(U, Q) = \sqrt{\frac{1}{2} \sum_{p \in P} \left(\sqrt{\rho(p, m)} - \frac{1}{\sqrt{|P|}} \right)^2}$$

Figura 9. Aplicação da fórmula da distância de Hellinger

O cálculo da incerteza é nada mais do que a razão da primeira fórmula pela segunda. O resultado é a seguinte expressão:

$$\frac{H(U, Q)}{H(U, C)} = 1 - \sqrt{\frac{\sum_{p \in P} \left(\sqrt{\rho(p, m)} - \frac{1}{\sqrt{|P|}} \right)^2}{2 - \frac{2}{\sqrt{|P|}}}}$$

Figura 10. Fórmula da incerteza para um turno

A fórmula da métrica é a média da aplicação da fórmula acima para M turnos, com exceção do primeiro:

$$U(x) = \frac{\sum_{m=1}^{|M|-1} 1 - \sqrt{\frac{\sum_{p \in P} \left(\sqrt{\rho(p, m)} - \frac{1}{\sqrt{|P|}} \right)^2}{2 - \frac{2}{\sqrt{|P|}}}}}{|M|-1}$$

Figura 11. Fórmula da métrica Uncertainty

A função ρ definida por Mangeli está definida de duas formas: uma para quando não se tem um valor máximo de pontos que se pode conseguir em um turno e outra para quando tem (MANGELI, 2016). Como o presente artigo analisará apenas campeonatos de pontos corridos, será apresentado apenas a segunda forma. Para definir a fórmula dessa função, será preciso definir algumas funções: S_l é a função que retorna a pontuação do líder no turno m , S é a função que retorna a pontuação de um time p no turno m e MS é a função que retorna qual é a função que retorna qual é a pontuação máximo que se pode conseguir no turno m . A função A_f de teste de conquista pode ser definida na seguinte fórmula:

$$A_f(p, m_n) = \begin{cases} 0, & \text{if } S_l(m_n) - S(p, m_n) > \sum_{k=n+1}^{|M|} MS(m_k) \\ \frac{1}{1 + S_l(m_n) - S(p, m)}, & \text{otherwise} \end{cases}$$

Figura 12. Fórmula da função de teste de conquista A_f

O turno é representado por mn , pois m não necessariamente representa o número do turno, então o n indica que esse turno mn é o n -ésimo turno do conjunto M de turnos.

A fórmula de ρ é então definida da seguinte forma:

$$\rho(p, m_n) = \frac{S(p, m_n) \cdot A_f(p, m_n)}{\sum_{k \in P} S(k, m) \cdot A_f(k, m_n)}$$

Figura 13. Fórmula da função de probabilidade p

3.1.7 Overall Aesthetic Metric

Essa é uma métrica recente desenvolvida por Mangeli, baseado na ideia de métrica estética, que mede quanto um jogo de múltiplos turnos for “bonito”. Como ele mostra em seu artigo, os campeonatos de futebol podem ser facilmente adaptados a sua proposta, onde as partidas podem ser vistas como turnos de um jogo. Em seu artigo, ela pode ser identificada como Overall Aesthetic Metric, a qual é “uma operação de agregação sobre conjuntos fuzzy definidos pelas métricas por critérios estéticos Uncertainty, Drama and LeadChange” (MANGELI, 2016). Essa operação de agregação é realizada da seguinte forma, sendo OAM o símbolo para overall Aesthetic Metric, D para Drama, U para Uncertainty e L para Lead Change:

$$OAM(x) = \frac{D(x) + U(x) + L(x)}{3}$$

Figura 14. Fórmula da métrica Overall Aesthetic Metric

As funções de Drama, Uncertainty e Lead Change são chamadas “membership functions”, pois elas são funções de lógica fuzzy que compõe outra função de lógica fuzzy que é a Overall Aesthetic Metric. No artigo, Mangeli propõe mais de uma métrica de Drama e de Uncertainty, mas no presente artigo serão utilizadas as funções que ele observou como melhores em sua validação (MANGELI, 2016).

3.2 Motivação das escolhas das métricas

O objetivo de escolher várias métricas para o programa é que isso permite avaliar os campeonatos sob a perspectiva de visões diferentes, além de comparar métricas antigas com métricas novas. De acordo com Kahane (KAHANE, 2003):

Economistas do esporte e outros têm desenvolvido um número de diferentes métricas de balanço competitivo. Todas medem a distribuição de vitórias e derrotas dos times nas ligas esportivas; cada uma representa um diferente método de descrever a distribuição.

4 PROPOSTA DE ARQUITETURA

Nessa seção serão apresentados os conjuntos de dados, como eles são tratados e os artefatos computacionais utilizados pelo programa, seja aqueles artefatos responsáveis por ler os dados dos arquivos, seja aqueles responsáveis por calcular o valor das métricas.

O propósito do programa é oferecer uma interface simples com geração de gráficos mostrando os valores das métricas, seja por histograma, seja por um gráfico de linha, a fim de dar suporte a análise de balanço competitivo de campeonatos.

O desenvolvimento do programa foi feito em Python com o paradigma orientado a objeto, a fim de modularizar o código e fazer usufruto dos quatro objetivos da modularidade: estrutura explícita, desenvolvimento separado, ocultamento de informações e reuso de código (MÜLLER, 2002), que envolve a essência da programação orientada a objetos, definida dessa maneira (WEGNER, 2003):

The essence of object-oriented programming is to model systems of real entities with the goal of separating their internal structure from their external, visible interactions. It emphasizes the hiding or encapsulation of the "inner" state of entities and the specification of interactive properties of entities by an interface of operations (the events in which they may participate). This separates the inner functioning of entities like banks, aircraft, or people from their external behavior in interacting with other entities. The separation is realized by partitioning the state of a system of entities into chunks associated with objects so that each chunk is responsible for its own protection against access by unauthorized operations. In a concurrent environment, objects protect themselves against asynchronous access, removing the synchronization burden from processes that access the object's data.

4.1 Arquitetura do programa

Nessa subseção serão apresentados os artefatos computacionais mencionados. A arquitetura do programa está apresentada de maneira simplificada no diagrama UML a seguir para que os seus atributos e métodos possam ser mostrados de maneira mais detalhada posteriormente.

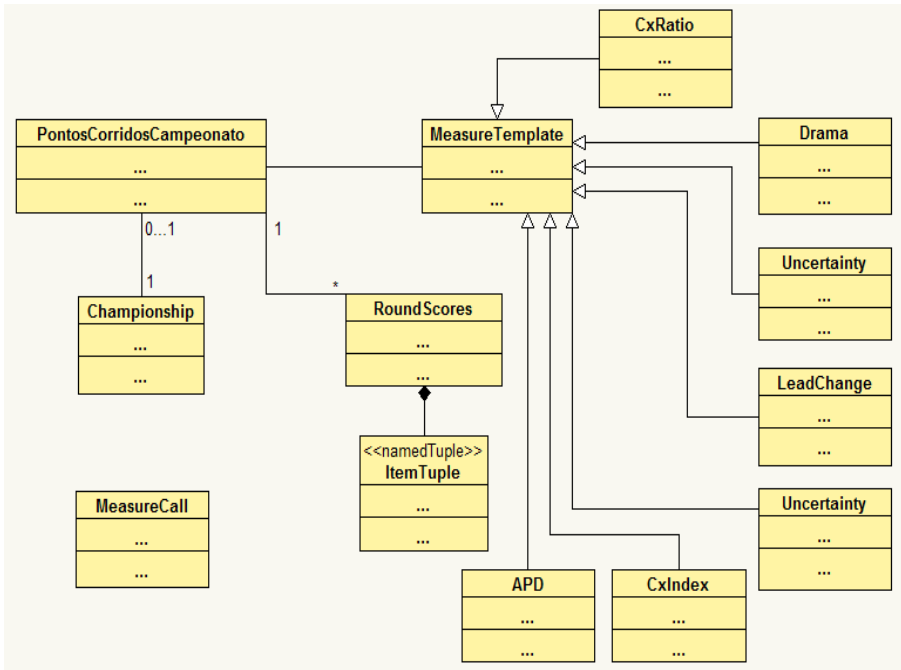


Figura 15. Diagrama de classes do programa com omissão dos atributos e métodos

As classes foram de tal forma implementadas que o trabalho de acesso aos dados e o trabalho de cálculo das métricas fosse dividida, com a intenção de facilmente expandir o programa para outras métricas e para outros campeonatos.

4.1.1 Classe Championship

A classe Championship é a classe responsável por extrair os dados de um campeonato qualquer armazenado em arquivo. Esses encontram-se armazenados de uma maneira sistemática e os dados desses estão organizados de maneira simples, tornando fácil recuperá-los e processar seu conteúdo.

O atributo country dessa classe guarda um código referente ao campeonato de um país, o atributo year armazena o ano que ocorreu uma instância desse campeonato, o atributo scoreLimit armazena o número de pontos máximo que pode ser feito em uma partida, mas quando o seu valor é nulo, indica que não existe um limite de pontuação. O atributo type indica o tipo do campeonato: é Double Round Robin ou Round Robin. Isso é importante porque, para implementação de algumas métricas isso pode fazer a diferença.

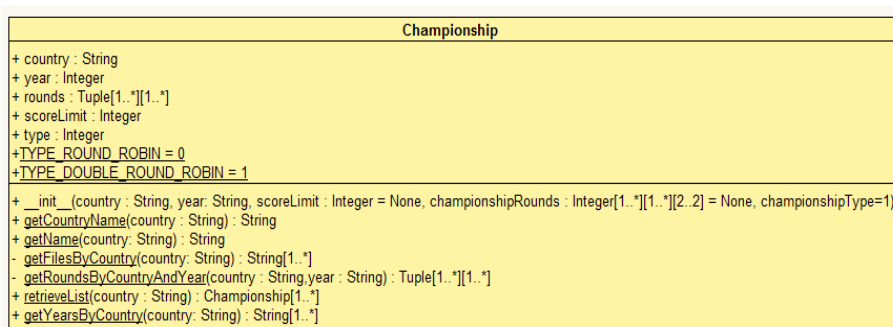


Figura 16. Representação UML da classe Championship

Os atributos `country` e `year` são atributos chave na recuperação do arquivo correto. O atributo `rounds` é simplesmente a lista das rodadas ordenadas da primeira para a última e cada rodada guarda uma ou mais tuplas com o nome do time e o número de pontos feitos no final daquele turno. Esse atributo é quem recebe o que está contido dentro do arquivo recuperado. Tanto para o campeonato de Double Round Robin quanto para um campeonato de Round Robin, o número de partidas que acontecem em um turno é equivalente a metade do número de times.

O método da classe `getCountryName` retorna o nome do país no qual é realizado o campeonato. É uma classe que foi implementado com o propósito de se usar nas telas do programa.

O método da classe `getName` retorna o nome do campeonato, também foi implementado com propósito de usar nas telas do programa.

O método da classe privado `getFilesByCountry` é um método usado mais de uma vez dentro da própria classe que pega os arquivos que tem como código o valor do parâmetro `country`. É um método utilitário.

O método da classe privado `getRoundsByCountryAndYear` é um método utilitário utilizado pelo método da classe `retrieveList`. Ele retorna a lista de tuplas, conforme descrito sobre o atributo `rounds` acima, extraídos do arquivo correspondente ao código e ao ano contidos nos parâmetros `country` e `year`, respectivamente.

O método da classe `retrieveList` retorna uma ou mais instâncias de `Championship` e é muito importante na hora de gerar os gráficos que aparecem na tela do programa.

O método da classe `getYearsByCountry` é um método que foi feito também com o propósito para ser usado em telas. Ela retorna uma lista de anos que tem um arquivo armazenado correspondente com o código contido no parâmetro `country`.

4.1.2 Classe PontosCorridosCampeonato

A classe `PontosCorridosCampeonato` é a classe que as métricas utilizam para fazer o cálculo do valor das mesmas. Cada métrica é representada por uma classe que é filha da classe `MeasureTemplate`.

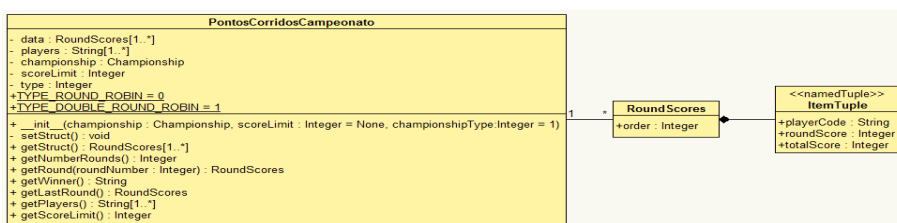


Figura 17. Representação UML das classes `PontosCorridosCampeonato`, `RoundScores` e `ItemTuple`

O atributo `data` armazena uma ou mais instâncias da classe `RoundScores` que é simplesmente uma classe que tem como atributos a ordem daquele turno (primeiro, segundo, entre outros) e um

conjunto de objetos da classe ItemTuple que armazena o nome do time (playerCode), o número de pontos feitos naquele turno (roundScore) e o número de pontos do time até o momento (totalScore). As instâncias já se encontram na ordem do primeiro para o último turno.

O atributo players armazena o nome de cada time participante de uma instância do campeonato, o qual corresponde a instância de Championship armazenada no atributo championship.

Os atributos scoreLimit e type, em significado, são equivalentes ao atributo scoreLimit e type da classe Championship, respectivamente

O método privado setStruct pega o atributo rounds do atributo championship e utiliza-o para preparar o atributo data. O método getStruct simplesmente retorna o atributo data e o método getNumberRounds retorna o número de instâncias de RoundScores contidas nesse atributo.

O método getRound retorna a instância de RoundScores, a qual tem o atributo order igual ao parâmetro roundNumber.

O método getWinner retorna o nome do time que venceu o campeonato.

O método getLastRound retorna a última rodada do campeonato, que é uma instância de RoundScores.

O método getPlayers retorna os nomes dos jogadores.

O método getScoreLimit retorna o valor armazenado no atributo scoreLimit.

4.1.3 Classe MeasureCall

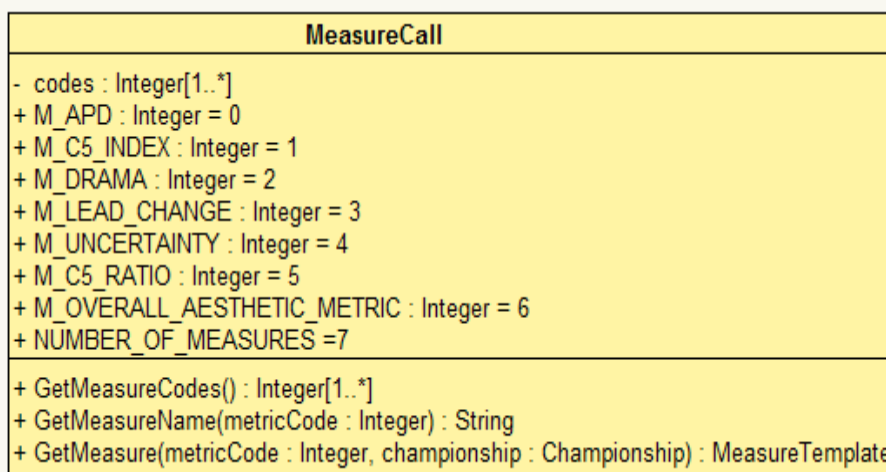


Figura 18. Representação UML da classe MeasureCall

A classe MeasureCall é uma classe utilitária que mapeia um código número para uma chamada de uma métrica. Por exemplo, um dos códigos chama-se M_C5_INDEX. Observando as classes filhas de MeasureTemplate é possível notar que não existe uma classe chamada M_C5_INDEX, isso porque esse código refere-se a declaração de CXIndex com o parâmetro x valendo cinco.

O atributo codes armazena todos os códigos possíveis de métricas. Os demais atributos são os códigos que são usados para identificar qual métrica deve ser chamada ou de qual deseja-se saber o nome. O método GetMeasureCodes retorna o atributo codes.

O método GetMeasureName retorna o nome da métrica em função do código da métrica passado.

O método GetMeasure retorna uma instância de uma das métricas filhas de MeasureTemplate em função do código passado por parâmetro e a métrica é avaliada a partir da instância de PontosCorridosCampeonato passada por parâmetro.

4.1.4 Classe MeasureTemplate

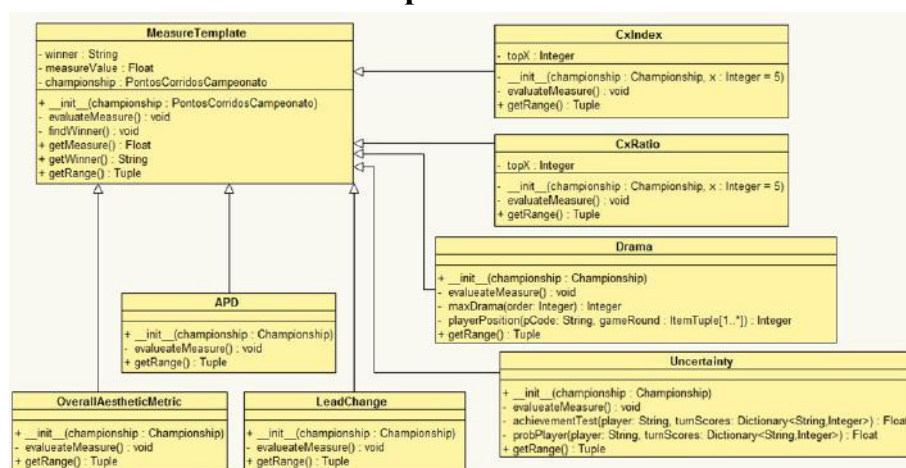


Figura 19. Representação UML da classe MeasureTemplate

A classe MeasureTemplate é a classe base para a implementação de todas as outras métricas. Nela tem métodos importantes para o cálculo das métricas, especialmente as de Drama, Uncertainty, LeadChange e OverallAestheticMetric.

O atributo championship armazena a instância de um campeonato, o qual é dado pela instância de PontosCorridosCampeonato.

O atributo winner armazena o nome do vencedor da instância do campeonato, conforme parágrafo anterior. Ele é determinado a partir do método privado findWinner e seu valor pode ser obtido a partir do método getWinner.

O atributo measureValue armazena o valor da métrica para a instância do campeonato em questão. O valor é determinado a partir do método privado evaluateMeasure e pode ser obtido a partir do método getMeasure. O primeiro método é implementado pela classe filha.

O método getRange é um método implementado também pela classe filha e determina o intervalo em que o valor da métrica pode variar. Esse método é usado para a tela na qual apresentam-se os gráficos das métricas. É especialmente importante para os histogramas, pois com

o intervalo definido é possível avaliar, apenas olhando para o gráfico, se os valores estão mais próximos de um cenário balanceado ou desbalanceado em geral.

A explicação de “evaluateMeasure” de cada classe filha da classe MeasureTemplate é correspondente a explicação do cálculo do valor das métricas apresentadas na seção <da revisão de métricas>. É possível observar que as classes Drama e Uncertainty implementam métodos extras. Esses métodos são métodos auxiliares para facilitar o cálculo do valor da métrica.

4.2 Dados

Nessa seção serão mostrados os conjuntos de dados que serão utilizados para demonstrar o programa e também a maneira que foram processados e armazenados.

4.2.1 Conjunto de dados

Para testar o programa, foram utilizados dados da Premier League, Fußball Bundesliga, Primeira Liga, Primera División da liga de Fútbol Profesional, Lega Nazionale Professionisti – Serie A, correspondentes ao campeonato de primeira divisão da Inglaterra, Alemanha, Portugal, Espanha e Itália, respectivamente. Foram recuperados dados de campeonatos de 1994 a 2018.

Na tabela abaixo está apresentado um sumário dos conjuntos de dados mencionados no início dessa seção:

Quadro 1. Apresentação sumarizada do conjunto de dados

Pais	Ano(s)	Número de Times	Número de Rodadas	Vencedor Mais Frequente
Alemanha	1994-2018	18	34	Bayern Munich
Espanha	1994	20	38	Real Madrid
Espanha	1995-1996	22	42	Ath Madrid
Espanha	1997-2018	20	38	Barcelona
Inglaterra	1994	22	42	Blackburn
Inglaterra	1995-2018	20	38	Man United
Itália	1994-2003	18	34	Juventus
Itália	2004-2018	20	38	Juventus
Portugal	1994-2005	18	34	Porto
Portugal	2006-2010	16	30	Porto
Portugal	2011	17	30	Porto
Portugal	2012-2013	16	30	Porto
Portugal	2014-2018	18	34	Benfica

Os históricos de partidas foram encontrados no website <https://www.football-data.co.uk/data.php> disponíveis em planilhas, havendo planilhas para vários campeonatos de vários países contendo vários dados estatísticos, mas apenas as colunas com o nome do time da casa, o nome do time visitante e o resultado da partida foram considerados no tratamento dos dados.

4.3 Tratamento dos dados

Os dados foram processados de tal maneira que produziram o ranking dos times em cada turno da instância do campeonato. Como todos os campeonatos processados são de modalidade Double Round-Robin, um turno é constituído de um número de partidas correspondente a metade do número de times, então a cada vez que eram processadas esse número de partidas do histórico, o programa anexava o ranking dos times tal como ele se apresentava a uma lista de rankings. Essa lista, por fim era salva em um arquivo JSON.

4.4 Exemplo de arquivo JSON

Um exemplo da estrutura do arquivo JSON encontra-se a frente, utilizando o arquivo do campeonato Fußball Bundesliga de 2012. O número da rodada não encontra-se no arquivo JSON, mas ele é apresentado aqui para explicitar que o placar do fim de cada rodada estão armazenados em ordem, assim como os times em relação ao placar, semelhante ao ranqueamento dos times de um campeonato após uma rodada.

Tabela 1. Exemplo da estrutura do arquivo JSON

Número da rodada	Nome do time	Placar
1	Dortmund	3
	Fortuna Dusseldorf	3
	Ein Frankfurt	3
	Bayern Munich	3
	Nurnberg	3
	M'gladbach	3
	Wolfsburg	3
	Freiburg	1
	Mainz	1
	Hannover	1
	Schalke 04	1
	Werder Bremen	0
	Augsburg	0
	Leverkusen	0
	Greuther Furth	0
	Hamburg	0
	Hoffenheim	0
Stuttgart	0	

Número da rodada	Nome do time	Placar
2	Ein Frankfurt	6
	Bayern Munich	6
	Dortmund	4
	Fortuna Dusseldorf	4
	Nurnberg	4
	M'gladbach	4
	Hannover	4
	Schalke 04	4
	Werder Bremen	3
	Leverkusen	3
	Greuther Furth	3
	Wolfsburg	3
	Freiburg	1
	Mainz	1
	Augsburg	0

	Hamburg	0
	Hoffenheim	0
	Stuttgart	0

4.4.1 Estrutura de diretório

Os arquivos JSON são armazenados da seguinte forma: a partir da pasta raiz do programa, existe uma pasta chamada “data”, como é possível ver na imagem abaixo. Nessa pasta, existem várias pastas, uma correspondente a um campeonato diferente e o nome da pasta o código do campeonato, citado anteriormente. Dentro de cada pasta, encontram-se arquivos JSON com o prefixo “simples” seguido do ano da instância do campeonato. Então, por exemplo, para pegar a evolução do ranqueamento dos times ao longo do campeonato de 2013 da Premier League, a partir da pasta raiz do programa, abre-se a pasta “data”, depois a pasta “ingles”(esse é o código usado no programa para a Premier League) e, por fim, o arquivo JSON “simples2013”.

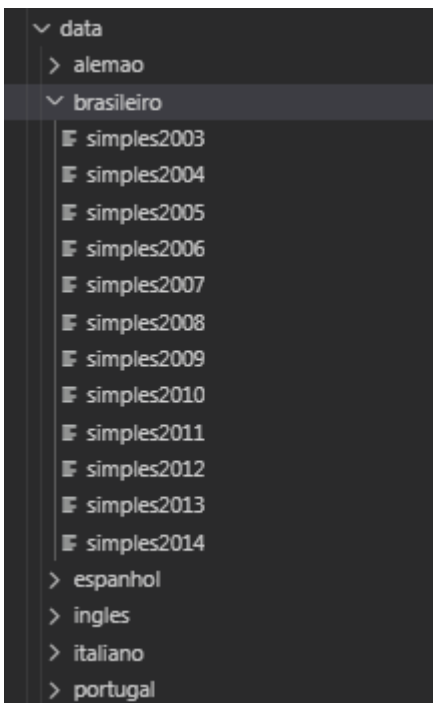


Figura 20. Captura de tela da estrutura de armazenamento dos dados

5 EXEMPLO DO PROGRAMA ATUAL

O programa possui cinco caixas de seleção: Championship, Year, Metric, Information e Graph Type, como pode ser observado na primeira imagem abaixo.

Na caixa Championship é selecionado o conjunto de dados correspondente a um dos campeonatos dos países, no caso do programa atual, Itália, Alemanha, Portugal, Espanha ou Inglaterra.

As caixas correspondentes a label Year filtram ainda mais o conjunto de dados, podendo selecionar uma ou mais instâncias de um campeonato previamente selecionado. As caixas da label Year influenciam também o comportamento das caixas Metric, Information e Graph Type: ao selecionar um ano igual nas duas caixas, a caixa Metric é trocada pela caixa Information e as opções dessa são “Positions” e “Points”; a caixa Graph Type tem suas opções trocadas e passa a oferecer apenas a opção histórico, como é possível ver na segunda imagem abaixo. Quando se seleciona valores diferentes para cada caixa, a caixa Information volta a ser caixa Metric. Enquanto isso, a caixa Graph Type apresenta as opções de histograma, evolução ao longo dos anos e evolução da média dos três últimos anos ao longo dos anos. Além disso, selecionar um mesmo ano para as duas caixas faz aparecer o botão “Metric Value” que permite olhar o valor de uma das métricas para a instância do campeonato do ano selecionado.

Ao selecionar uma das opções da caixa Metric é determinado qual métrica será aplicado para os dados do campeonato selecionado. Será calculado o valor da métrica para o conjunto de dados de cada instância desse campeonato ao pressionar o botão Make. As métricas disponíveis, por enquanto são Accumulated Points Difference, C5 Index of Competitive Balance, Drama, Lead Change, Uncertainty, C5 Ratio e Overall Aesthetic Metric.

No caso da caixa Information, a opção “Positions” e a opção “Points” implicam em mostrar a posição e o número de pontos de cada time ao longo da instância de um campeonato, respectivamente.

A caixa Graph Type permite selecionar em que tipo de gráfico os valores da métrica calculados serão apresentadas.

Na imagem abaixo é possível verificar um exemplo: Para todos os anos do campeonato Primeira Liga (nesse caso, de Portugal), é calculado o valor da métrica Drama para cada ano e esses valores são apresentados em um gráfico de linhas que mostra o valor da métrica para o campeonato em questão ao longo dos anos.

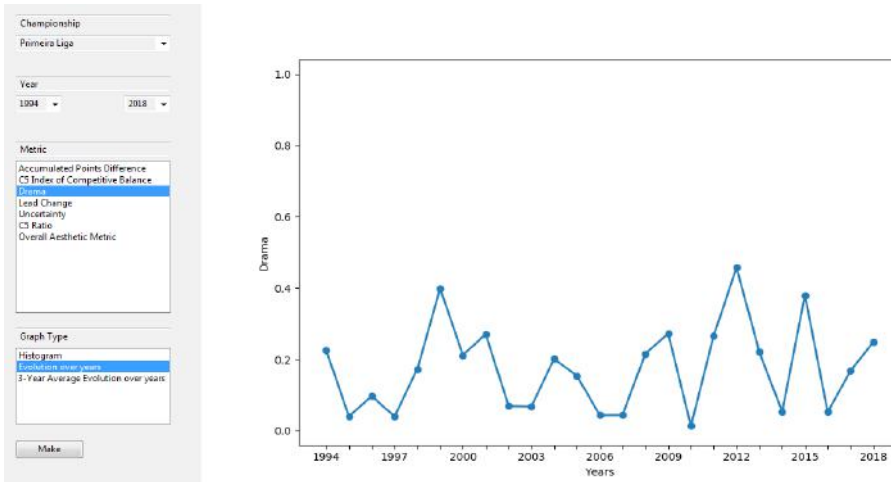


Figura 21. Gráfico de Drama para Primeira Liga

Abaixo, há um exemplo da tela quando é mostrado o histórico das posições dos times de uma instância do campeonato, a qual corresponde, nesse exemplo, a instância de 1997 do campeonato Primeira Liga, de Portugal.

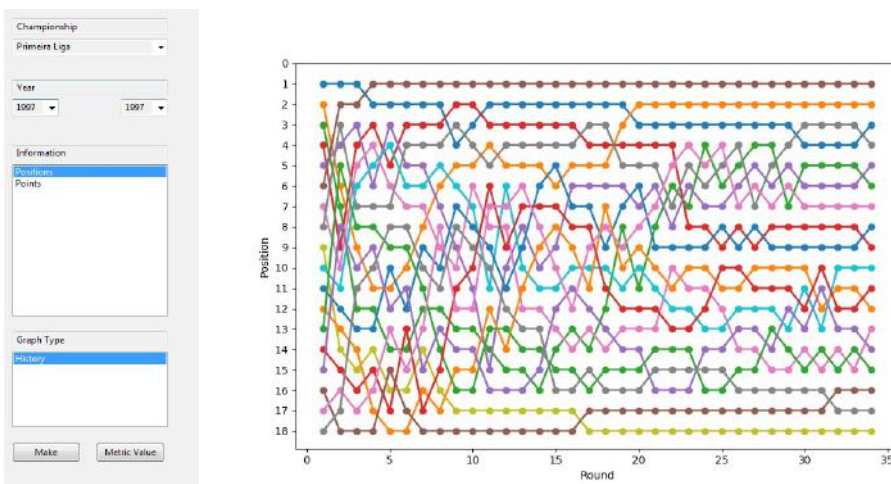


Figura 22. Gráfico de posicionamento dos times ao longo do campeonato

Na imagem abaixo encontra-se um exemplo da janela que abre ao clicar no botão “Metric Value”. A janela é simples com uma caixa de seleção para selecionar a métrica da qual se deseja ver o valor da instância do campeonato desejada e um campo para ver o valor dito.

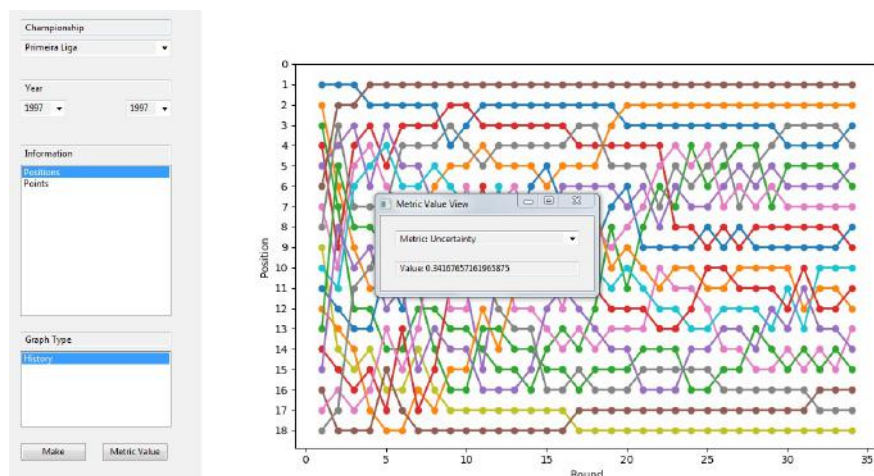


Figura 23. Caixa de diálogo mostrando o valor de Uncertainty para o campeonato do ano 1997 da Primeira Liga

As opções de gráficos disponíveis no programa são histograma, evolução ao longo dos anos, evolução da média dos três últimos anos ao longo dos anos e histórico.

O histograma é um gráfico de barras onde a barra representa a frequência com que os valores da métrica caem em um intervalo, esse podendo ser um intervalo entre 0% e 4%, 4% e 8%, 8% e 12% e assim por diante, da diferença entre o valor mínimo e o valor máximo da métrica. O propósito dele é verificar onde os valores se concentram mais ou se não há concentração nenhuma, permitindo ter uma noção da média, mas sem deixar de mostrar a variação existente.

A evolução ao longo dos anos é um gráfico de linha no qual o eixo das abcissas é o dos anos e o eixo das ordenadas é o eixo do valor da métrica. O propósito desse gráfico é verificar como um campeonato se comportou, de acordo com aquela métrica, ao longo dos anos, sendo possível verificar o crescimento, a diminuição ou a oscilação do balanço competitivo.

A evolução da média dos três últimos anos ao longo dos anos é composto de dois gráficos de linhas: um corresponde ao gráfico descrito no parágrafo anterior, no outro o eixo das ordenadas é o eixo das médias dos valores da métrica para o ano naquele ponto e para os dois anos anteriores. O propósito desse gráfico é ajudar a verificar o impacto do valor da métrica naquele ano em comparação ao dois anos anteriores. Essa opção de gráfico é especialmente importante para uma métrica estética como a de Drama, na qual procura-se medir quanta tensão houve na determinação do vencedor de um campeonato.

O histórico é um gráfico de linhas no qual mostra ou as posições ou a pontuação dos times ao longo da instância do campeonato. O propósito desse gráfico é permitir ao usuário visualizar a realidade da evolução da ranking dessa instância.

Nas figuras abaixo, estão apresentados exemplos de histograma, evolução da média dos três últimos anos ao longo dos anos e histórico de pontos, respectivamente.

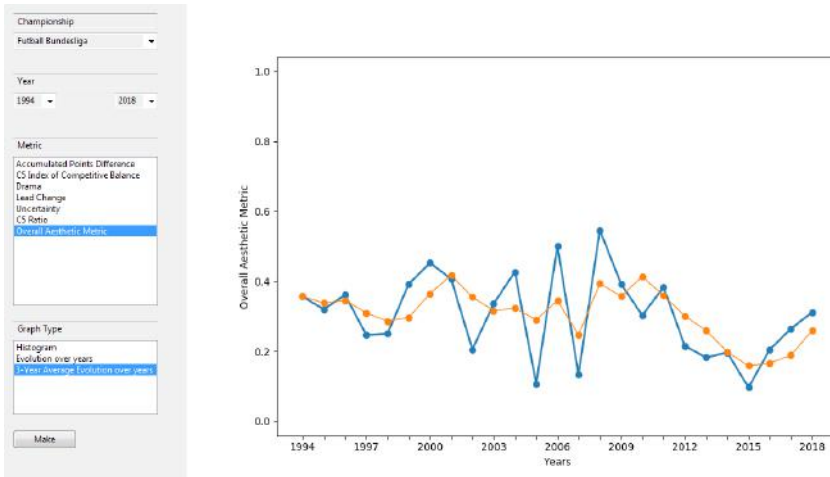


Figura 24. Descrição da Figura

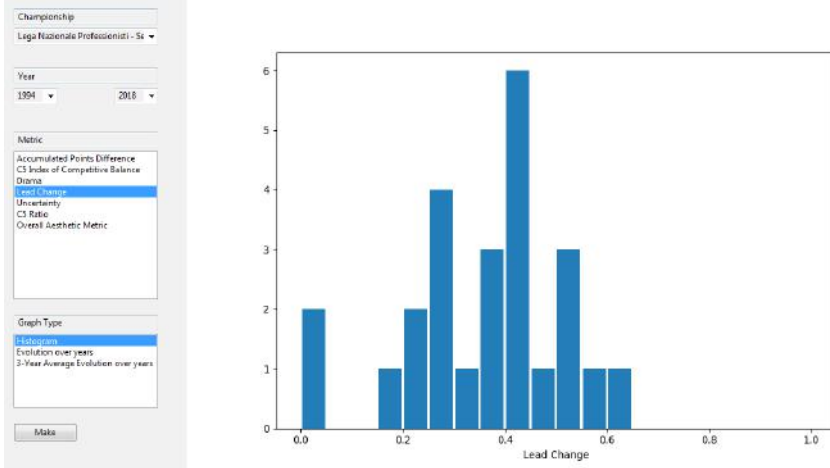


Figura 25. Descrição da Figura

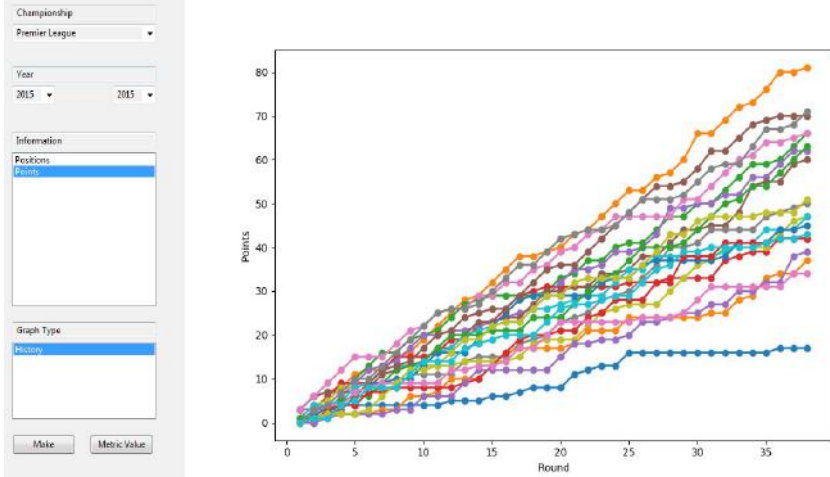


Figura 26. Descrição da Figura

6 ANÁLISE DE DADOS DOS CAMPEONATOS

Nessa seção será apresentado os valores médios das métricas para cada um dos seus campeonatos, assim como a evolução do valor das métricas ao longo dos campeonatos. Em seguida, será feita uma análise do balanço competitivo por campeonato e por ano.

As métricas que serão utilizadas para fazer a análise são as métricas de Accumulated Points Difference (APD), C5 Index of Competitive Balance (C5ICB) e Overall Aesthetic Metric (OAM). A razão para não usar todas as métricas está no fato de que a segunda métrica mencionada utiliza a C5 Ratio, como explicado na seção 3.2.3, e a terceira métrica mencionada é a média das métricas de Drama, Uncertainty e Lead Change, conforme seção 3.2.7. Além disso, a análise torna-se mais simples de ser feita quando tomando por referência apenas três métricas, em vez de sete.

Para maior compreensão das análises, uma breve explicação sobre os valores das métricas: quanto maior é o valor das métricas APD e C5ICB, menor é o balanço competitivo; quanto menor é o valor da métrica OAM, menor é o balanço competitivo.

Apesar de a métrica OAM ser uma métrica para avaliar a estética de um jogo, ela será utilizada para avaliar quão balanceado é o campeonato, porque essa métrica envolve três outras métricas que envolvem a ideia de chances iguais para todos os jogadores.

A métrica Drama compara o caminho do vencedor com um caminho de Drama máximo, no qual vencedor começaria do último lugar e lentamente escalaria até o primeiro lugar. Essa ideia de métrica dá um destaque maior a campeonatos nos quais um time que começou mal e que ainda conseguiu alcançar o primeiro lugar, mostrando uma possível igualdade de chances.

A métrica Uncertainty envolve calcular a distância entre a probabilidade de um time qualquer ser vencedor em um turno do jogo e a probabilidade ideal de um time qualquer vencer, novamente envolvendo uma ideia de chance igual para todos.

A métrica Lead Change envolve calcular o número de mudanças de líder e o número de líderes diferentes. Se existem muitos líderes diferentes e muitas mudanças de líder, é um sinal de que o campeonato foi balanceado competitivamente.

Sendo a métrica OAM uma média desses três métricas, ela expressará conseqüentemente as ideias de igualdade de chance contidas nessas métricas.

6.1 Campeonato alemão

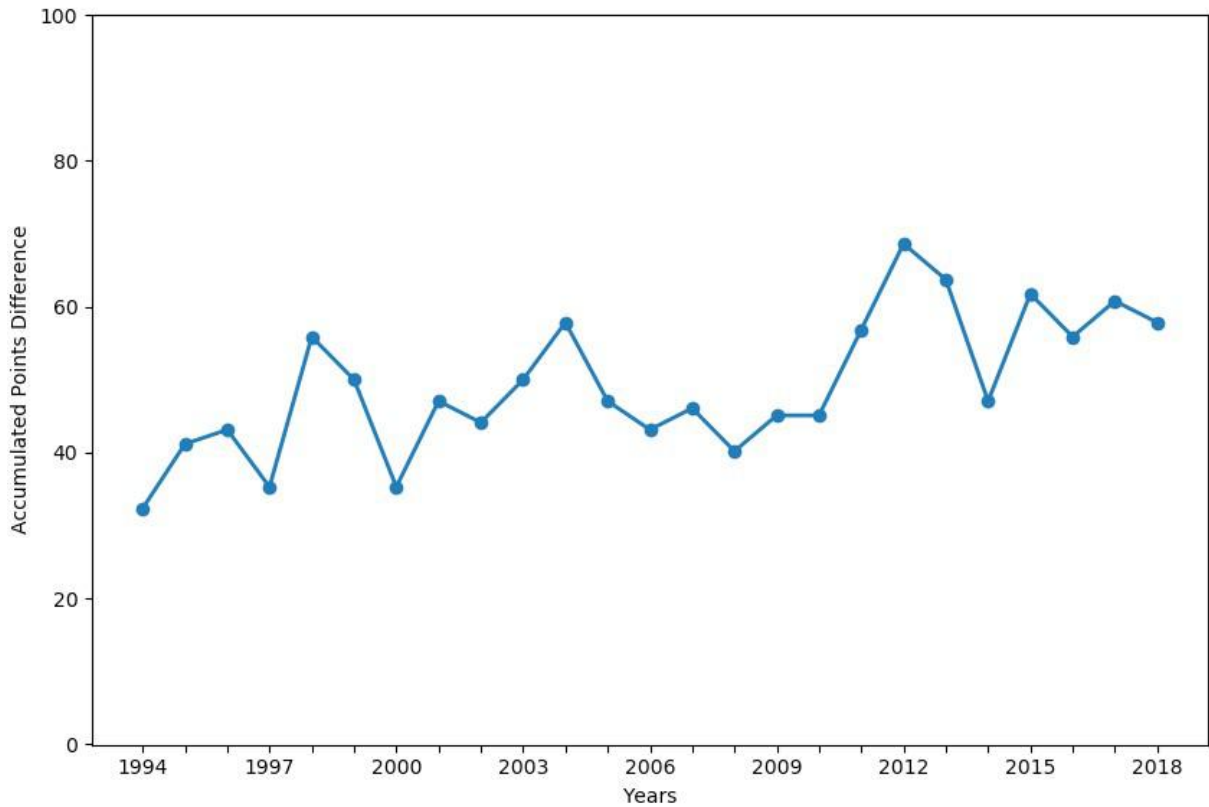
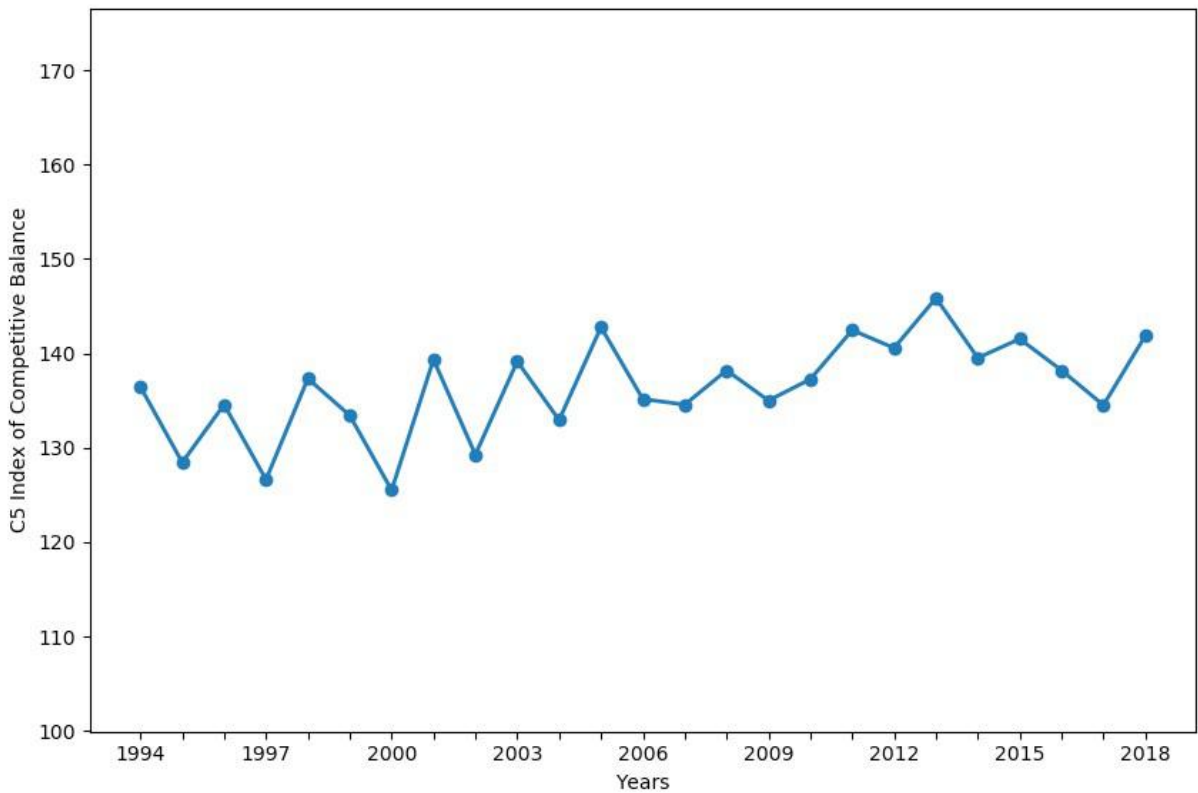
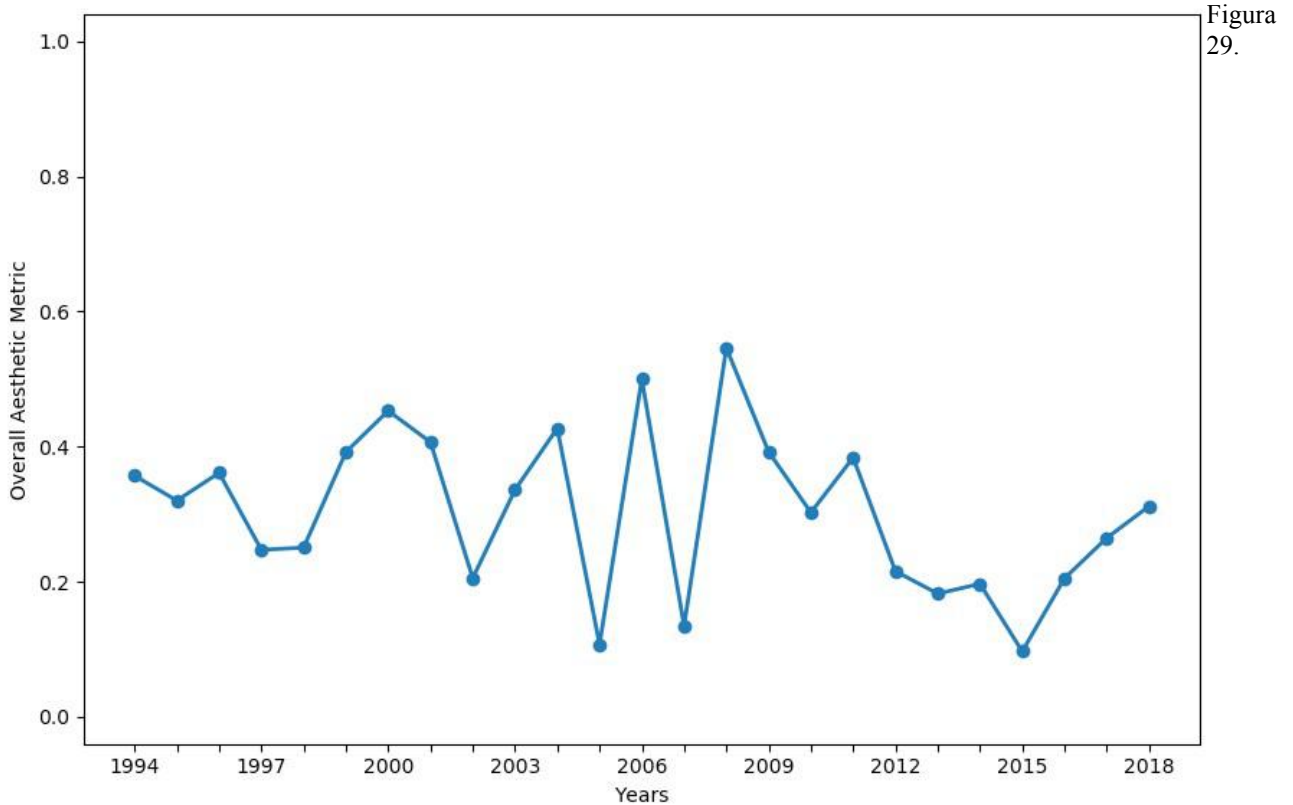


Figura 27. Evolução do valor da métrica Accumulated Points Difference no período de 1994 a 2018



28. Evolução do valor da métrica C5 Index of Competitive Balance no período de 1994 a 2018

Figura



Evolução do valor da métrica Overall Aesthetic Metric no período de 1994 a 2018

As métricas indicam três comportamentos distintos, mas similares, para o campeonato: A métrica APD indica uma queda gradual do balanço competitivo, a métrica C5ICB indica uma queda do balanço competitivo, mas que se estabilizou e a métrica OAM indica uma queda no balanço competitivo, mas com uma indicação de crescimento desse ao final. É possível observar que as métricas mostram realidades diferentes em cada ano, mas mostram praticamente a mesma realidade ao longo do tempo. Para o ano de 2006, as três métricas indicam valores que indicam um bom balanço competitivo, especialmente a métrica OAM.

6.2 Campeonato italiano

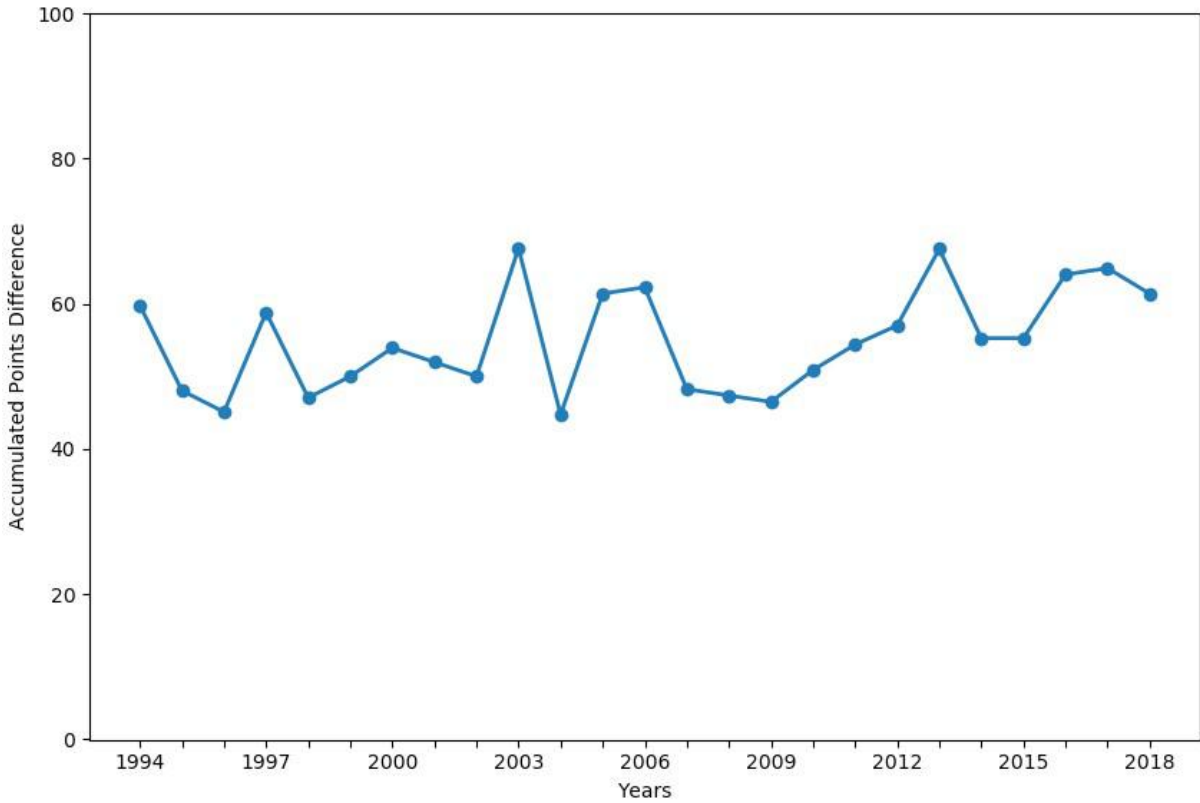


Figura 30. Evolução do valor da métrica Accumulated Points Difference no período de 1994 a 2018

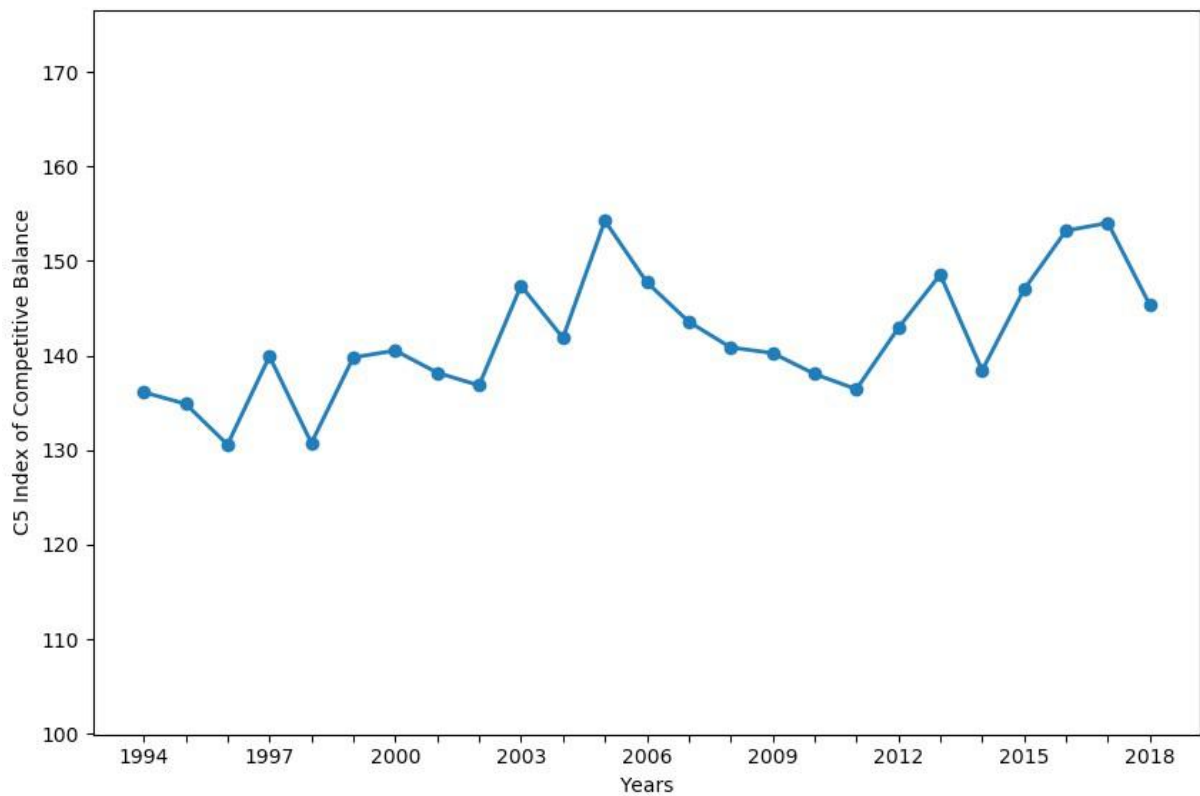


Figura 31. Evolução do valor da métrica C5 Index of Competitive Balance no período de 1994 a 2018

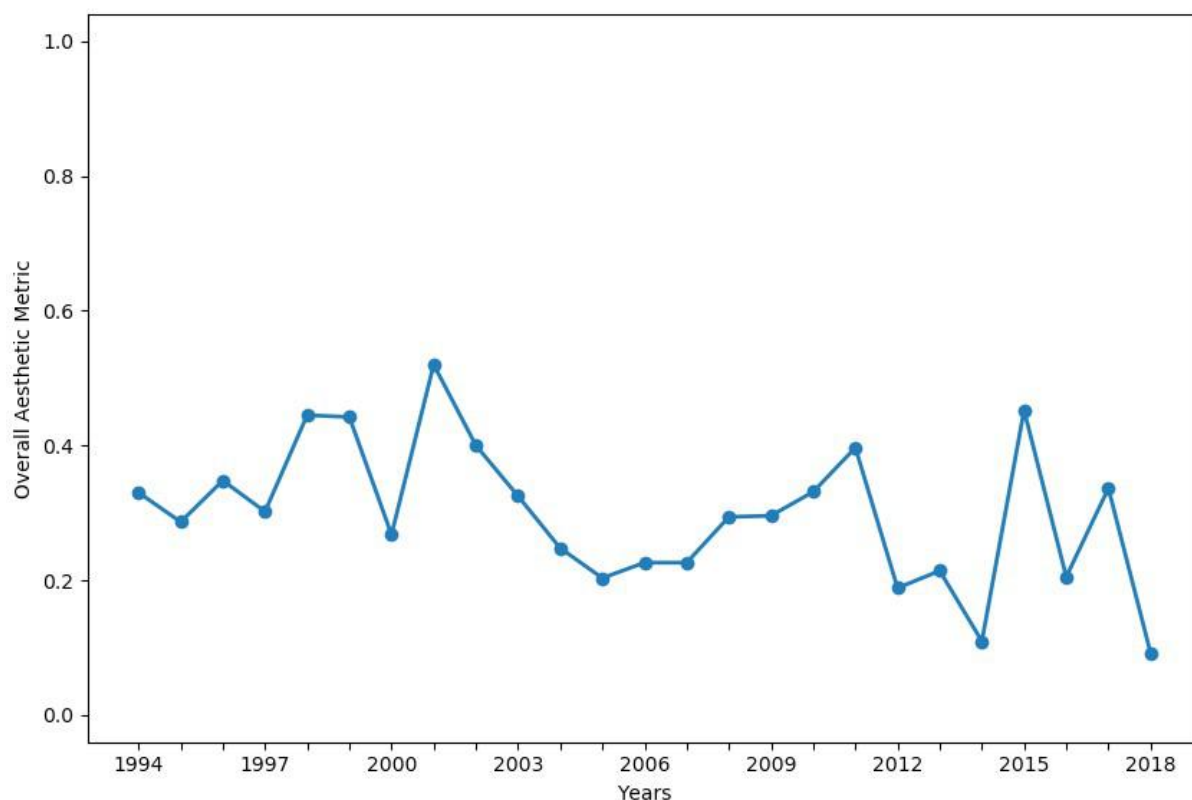


Figura 32. Evolução do valor da métrica Overall Aesthetic Metric no período de 1994 a 2018

Apesar de as métricas mostrarem uma evolução diferente para cada um dos campeonatos, a realidade apontada é praticamente a mesma: em 1998 há uma melhoria visível do balanço competitivo, mas logo o balanço competitivo começa a cair, cada uma indica uma realidade a partir de 2008, mas em 2012 todas apontam uma queda visível no balanço competitivo e a situação não melhora muito depois disso.

6.3 Campeonato espanhol

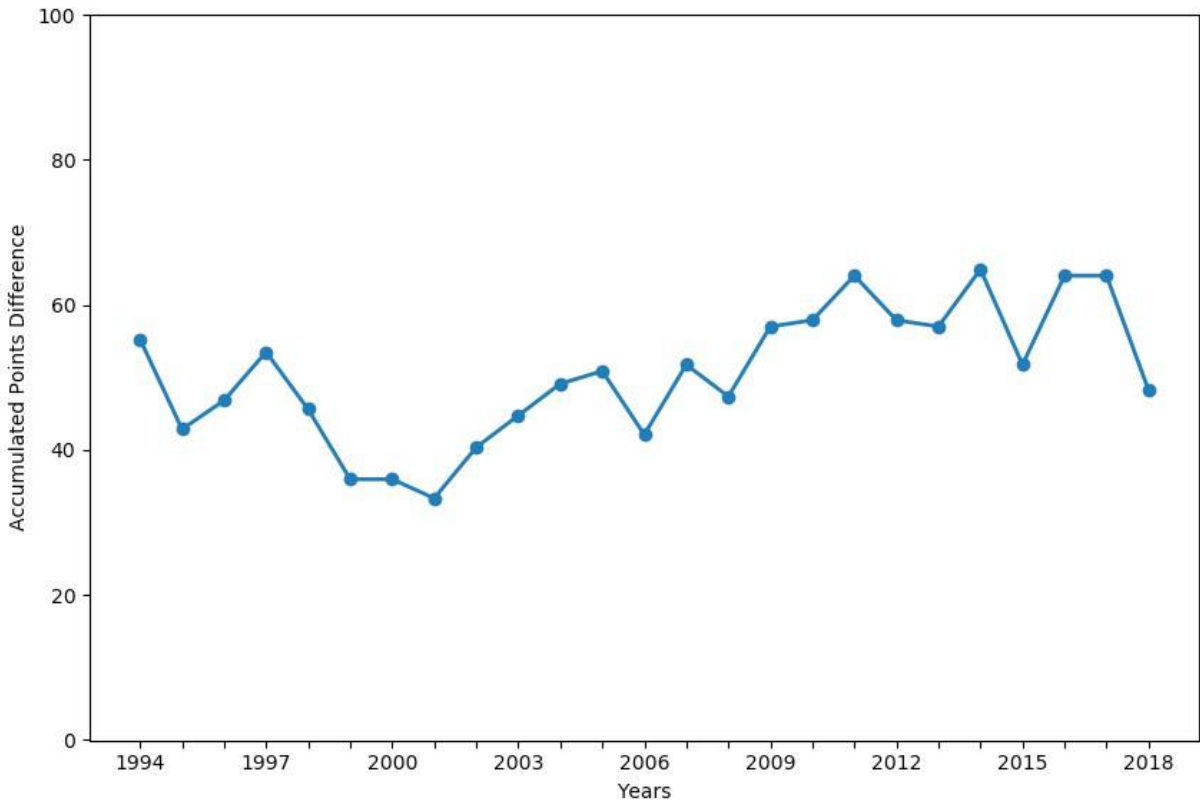


Figura 33. Evolução dos valores da métrica Accumulated Points Difference no período de 1994 a 2018

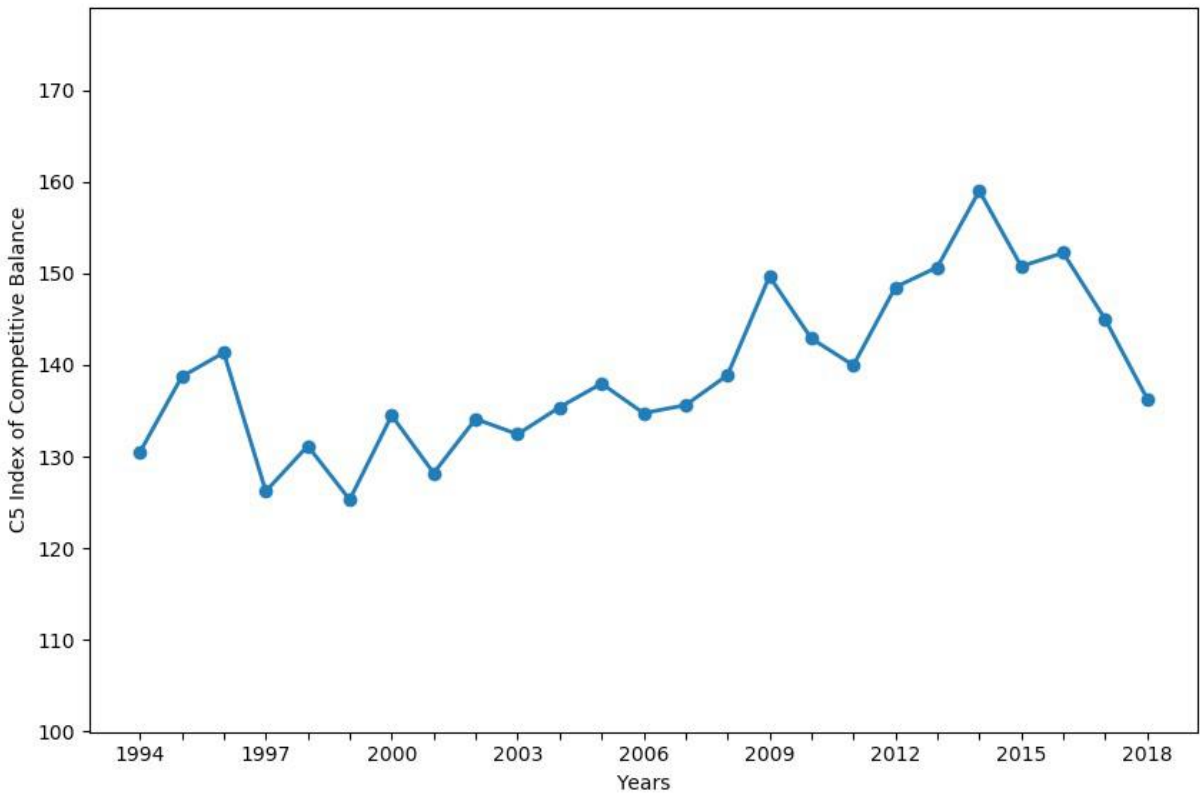


Figura 34. Evolução dos valores da métrica C5 Index of Competitive Balance no período de 1994 a 2018

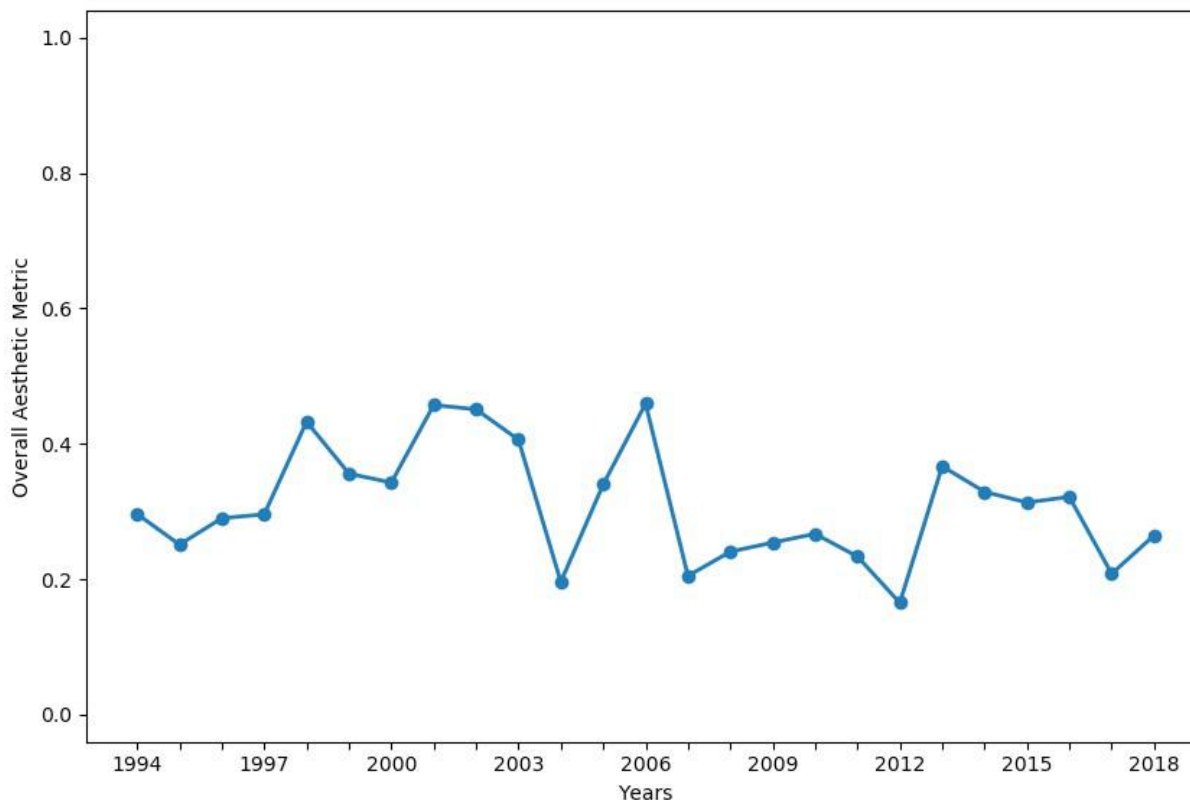


Figura 35. Evolução dos valores da métrica Overall Aesthetic Metric no período de 1994 a 2018

A métrica OAM mostra um cenário diferente das outras duas métricas: enquanto as métricas APD e C5ICB mostram inicialmente um campeonato balanceado, mas que perde seu balanço competitivo ao longo dos anos e começa a recuperá-lo, a métrica OAM mostra um campeonato que manteve o balanço competitivo, mas situando-se um pouco abaixo em comparação a outros campeonatos, mas que torna-se desbalanceado ao longo do tempo e, mesmo passando por uma fase de maior balanço competitivo, mostra-se mais fraco nesse aspecto. Enquanto as duas primeiras métricas mostram um cenário muito positivo que tornou-se ruim, a terceira métrica mostra um cenário que era positivo, mas não tanto quanto outros campeonatos e não conseguiu manter esse cenário no fim.

O que pode ser observado que o ano de 2001 é considerado, por todas as métricas, como um cenário positivo para balanço competitivo, mas que definiu-se, mesmo que levando alguns anos para a métrica mostrar isso.

6.4 Campeonato português

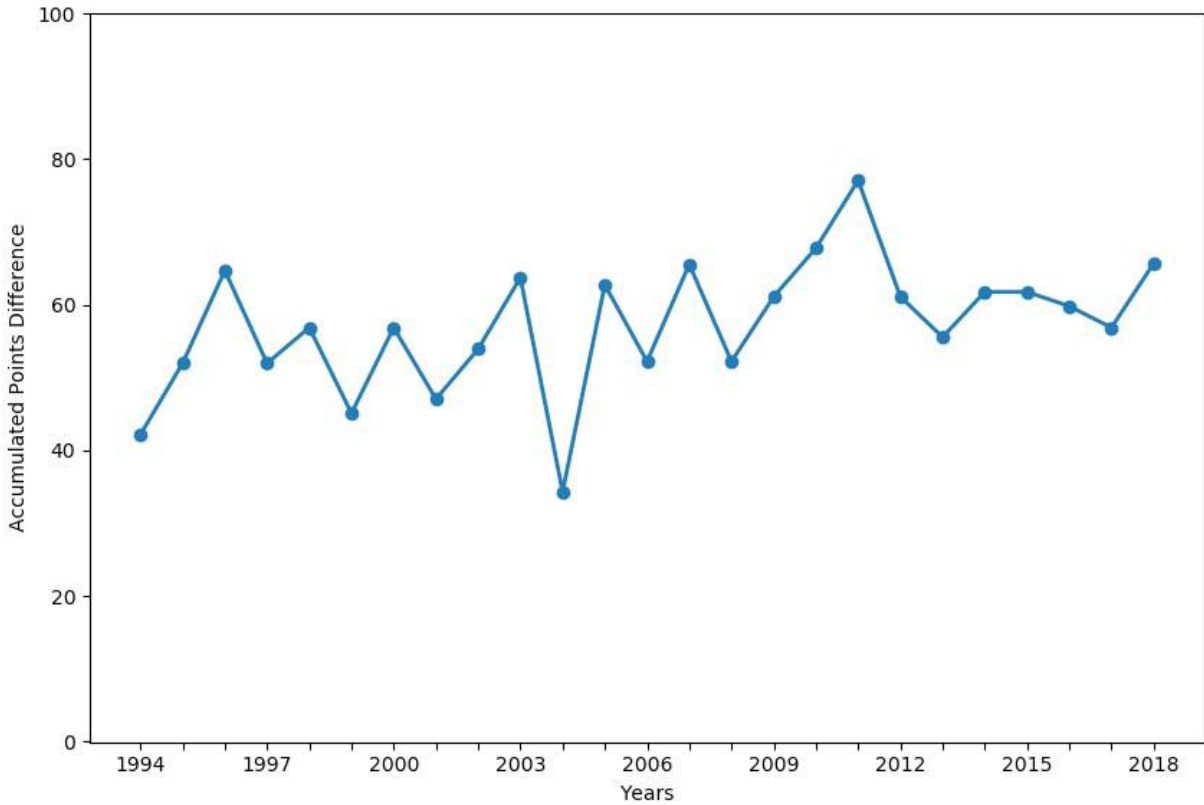
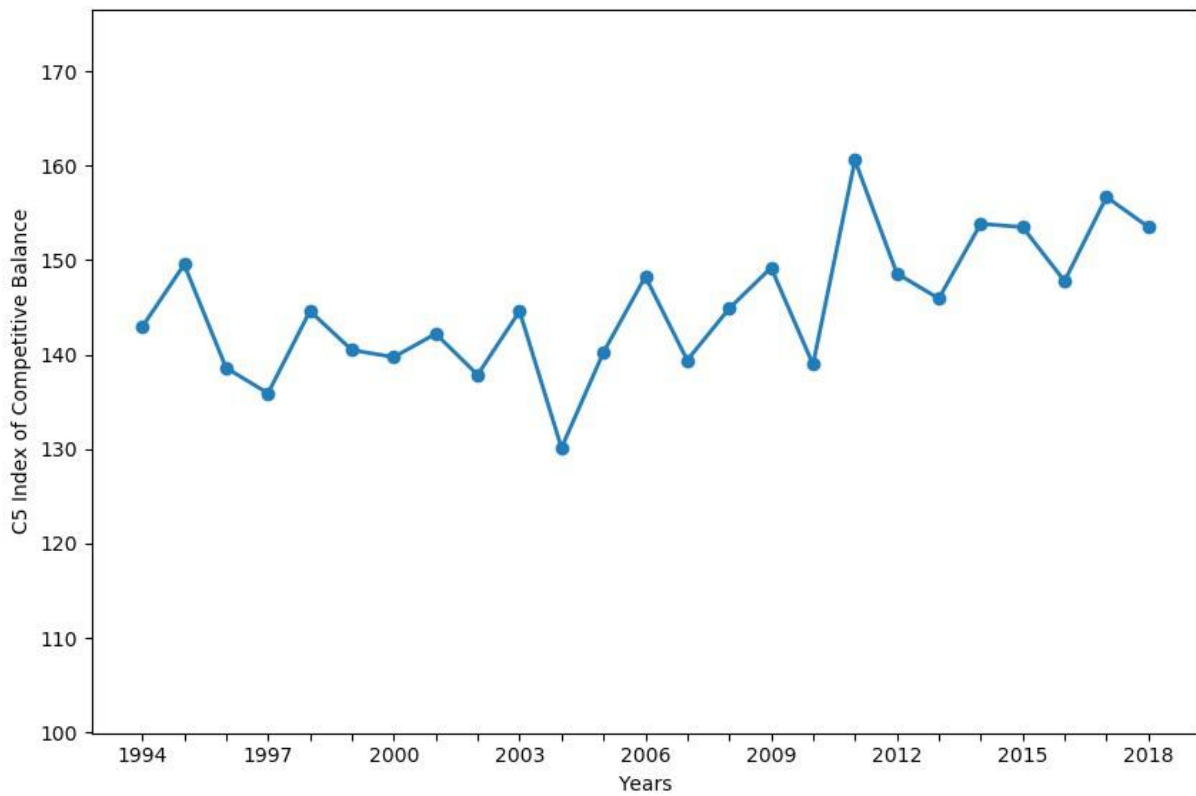
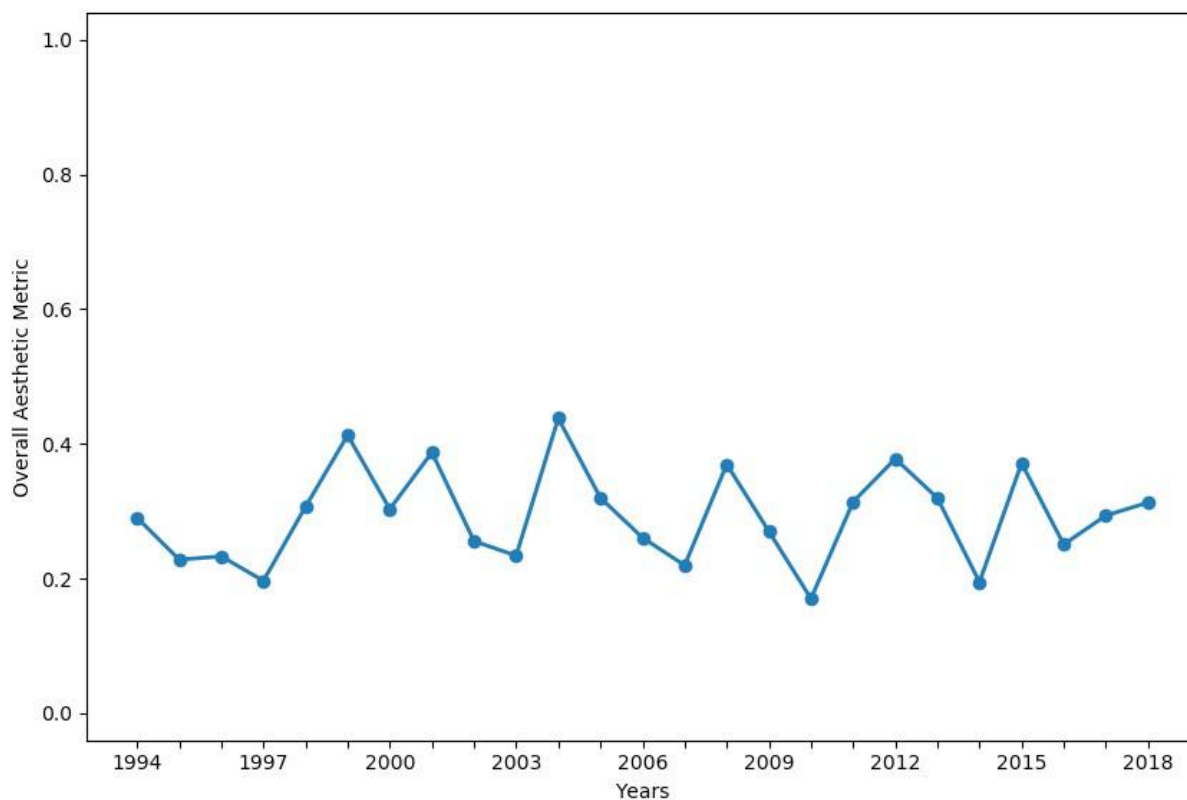


Figura 36. Evolução do valor da métrica Accumulated Points Difference para o período de 1994 a 2018



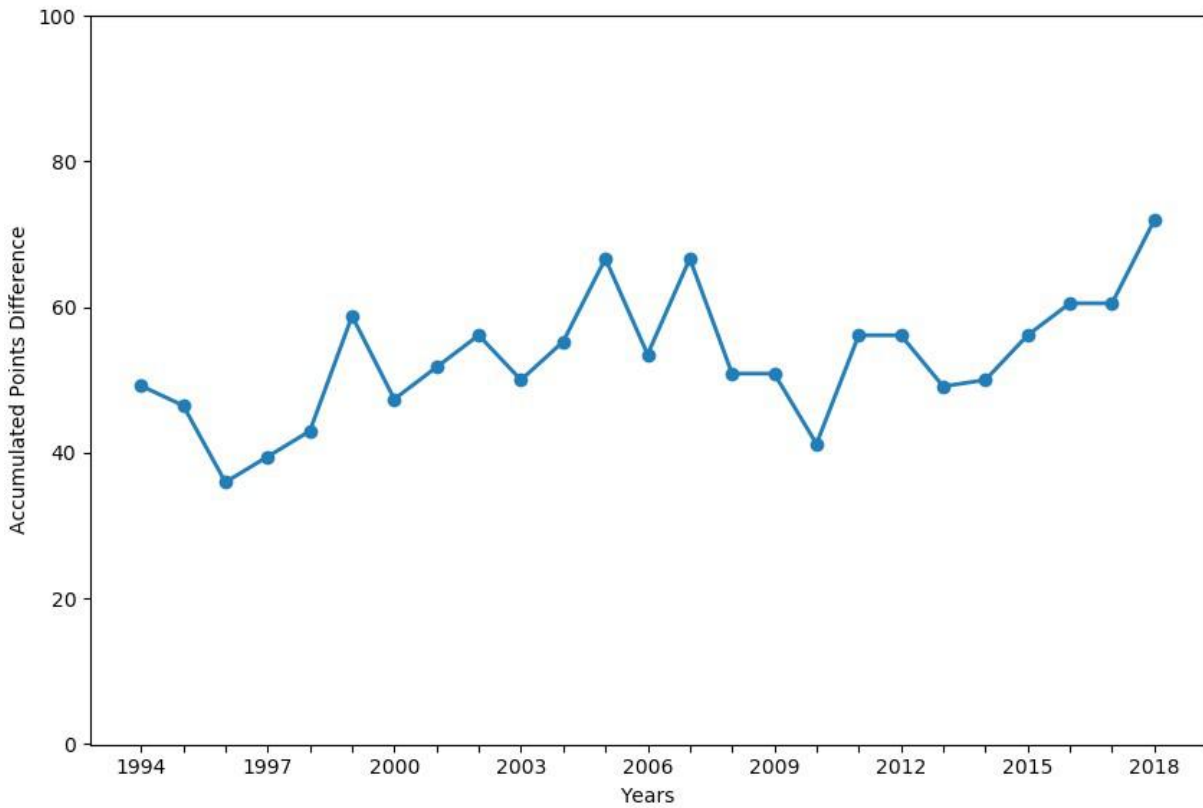
Evolução do valor da métrica C5 Index of Competitive Balance para o período de 1994 a 2018



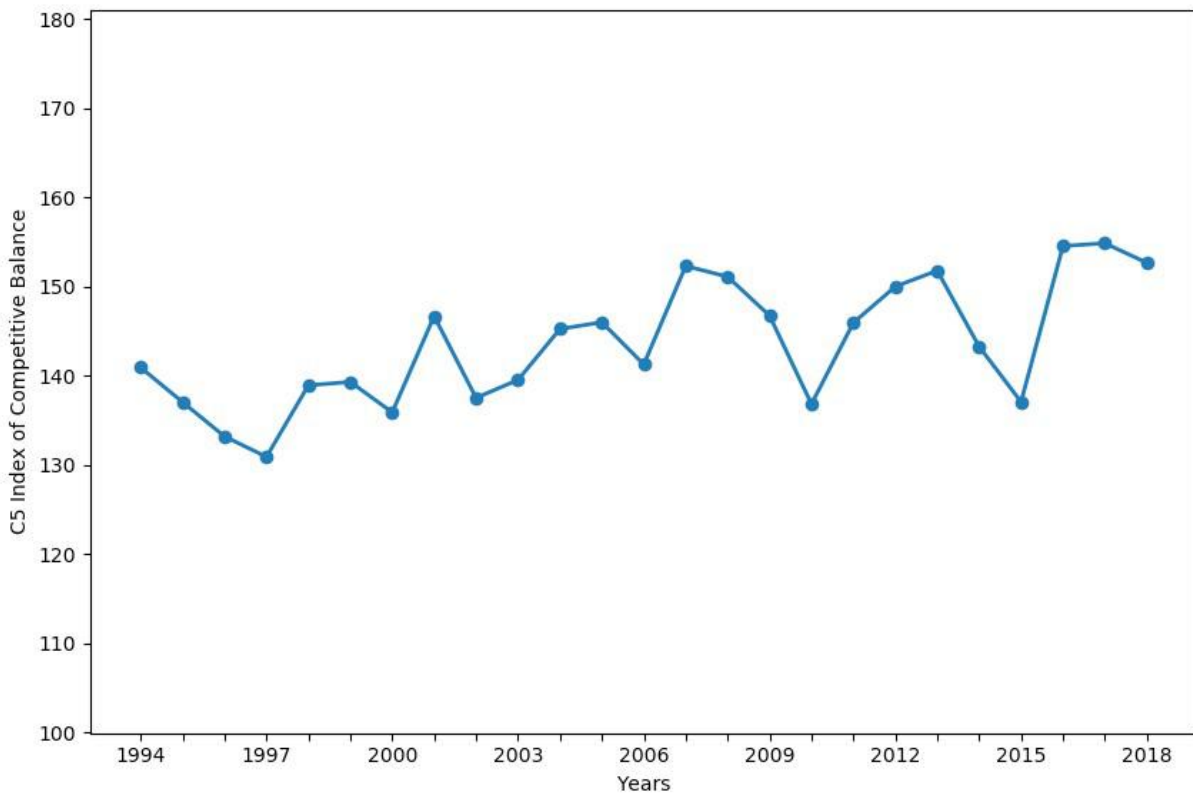
Evolução do valor da métrica Overall Aesthetic Metric no período de 1994 a 2018

As métricas mostram aproximadamente a mesma evolução para o campeonato português: o campeonato começa com uma avaliação boa, fica estável, tem um bom campeonato em 2004, mas o campeonato piora cada vez mais ao longo do tempo. Enquanto nos dois primeiros gráficos isso é mostrado de uma maneira mais drástica, o gráfico da métrica Overall Aesthetic Metric não mostra o mesmo declínio das outras métricas, mostrando alguns dos campeonatos com valores próximos do campeonato de 2004.

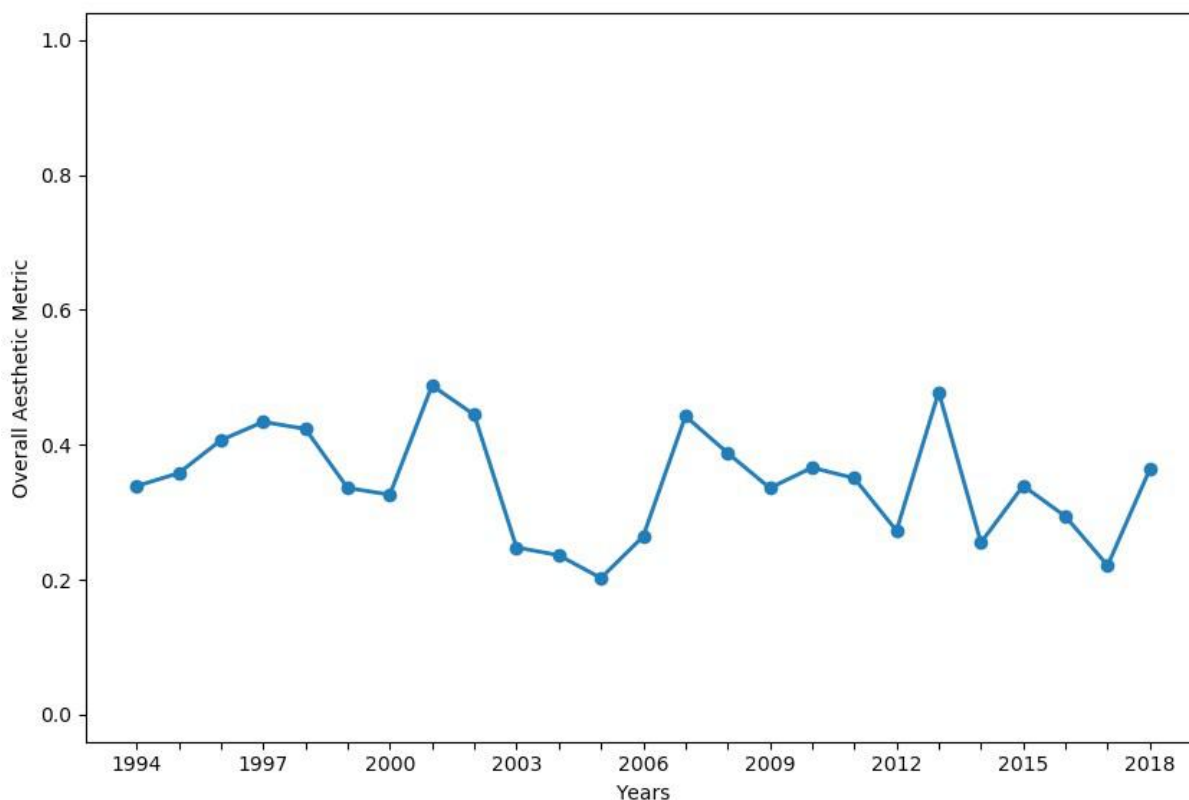
6.5 Campeonato inglês



Evolução do valor da métrica Accumulated Points Difference para o período de 1994 a 2018



Evolução do valor da métrica C5 Index of Competitive Balance para o período de 1994 a 2018



Evolução do valor da métrica Overall Aesthetic Metric para o período de 1994 a 2018

Nesse caso, as três métricas mostram evoluções diferentes para um mesmo campeonato de um país. A métrica APD mostra um campeonato que, de forma geral, só piora ao longo do tempo, enquanto a métrica C5ICB mostra um campeonato que piora ao longo do tempo, mas com alguns anos recuperando valores próximos de anos anteriores que apresentam valores muito bons. Já a métrica OAM mostra um campeonato com altos e baixos, apresentando épocas com valores mais baixos e épocas com valores mais altos, porém, o campeonato apresenta uma variação menor na avaliação do campeonato em relação a campeonatos de outros países, pois enquanto em outros campeonatos, em tempos ruins, os valores da métrica aproximavam-se de zero, o campeonato inglês mantém seus valores mínimos mais acima.

7 CONCLUSÃO

O presente projeto final se propôs a apresentar um programa que medisse o balanço competitivo de um campeonato. Foram escolhidas as ferramentas a serem utilizadas, como linguagem de programação e bibliotecas, as métricas que seriam utilizadas para avaliar o balanço competitivo de um campeonato.

A arquitetura necessária para o programa foi definida, com a definição de cada classe do programa, seus relacionamentos e funções, assim como foi definido um conjunto de dados para o teste do programa.

Um programa foi implementado, apresentado e utilizado para avaliar o balanço competitivo dos campeonatos.

Por fim, o que esse projeto final produziu foi um programa que pode ser utilizado para fazer a avaliação do balanço competitivo com a possibilidade da expansão do programa para abordar mais conjuntos de dados e mais métricas.

REFERÊNCIAS

- Brawl Stars - Android app on AppBrain.** Disponível em: <<https://www.appbrain.com/app/brawl-stars/com.supercell.brawlstars>>. Acesso em: 15 dez. 2021.
- BURAIMO, B.; SIMMONS, R. **Do Sports Fans Really Value Uncertainty of Outcome? Evidence from the English Premier League.** Disponível em: </paper/Do-Sports-Fans-Really-Value-Uncertainty-of-Outcome-Buraimo-Simmons/6703f7bdec2ec71479b09fc10ec97ae6ced22260>. Acesso em: 7 out. 2020.
- Call of Duty®: Mobile - Season 10: Shadows Return - Free Android app.** Disponível em: <<https://www.appbrain.com/app/call-of-duty%C2%AE-mobile-season-10-shadows-return/com.activision.callofduty.shooter>>. Acesso em: 15 dez. 2021.
- Cuju: conheça o jogo que originou o futebol.** Disponível em: <<https://www.ibrachina.com.br/cuju-conheca-o-jogo-que-originou-o-futebol/>>. Acesso em: 15 dez. 2021.
- DEJONGHE, T. **The evolution of Belgian football over the last decades.** UEFA/IASE Seminar: The economic impact of the UEFA Champions League on national football economies, Date: 2006/05/02-2006/05/02, Location: Nyon, Switzerland. **Anais...**2006.
- DRUMMOND, L.; JÚNIOR, A. A.; SHIKIDA, C. D. Campeonato Brasileiro de Futebol e balanço competitivo: uma análise do período 1971-2009. **Revista Brasileira de Futebol (The Brazilian Journal of Soccer Science)**, v. 3, n. 2, p. 73–87, 2013.
- FEDDERSEN, A. **Economic consequences of the UEFA Champions League for national championships: The case of Germany.** Hamburg: University of Hamburg, Chair for Economic Policy, 2006. Disponível em: <<http://hdl.handle.net/10419/49287>>.
- GASPARETTO, T.; BARAJAS, A. **ANÁLISIS DEL BALANCE COMPETITIVO POR LA DIFERENCIA DE PUNTOS ACUMULADA (DPA).** 6 fev. 2015.
- GERRARD, B. Still Up for Grabs? Maintaining the Sporting and Financial Viability of European Club Soccer in Fort, R. and Fizel, J. **International Sports Economics Comparisons**, 2004.
- JSON.** Disponível em: <<https://www.json.org/json-en.html>>. Acesso em: 15 dez. 2021.
- KAHANE, L. H. Comments on “Thinking about Competitive Balance”. **Journal of Sports Economics**, v. 4, n. 4, p. 288–291, 1 nov. 2003.
- KRINGSTAD, M.; GERRARD, B. The Concepts of Competitive Balance and Uncertainty of Outcome. 1 jun. 2004.
- LUDES | Laboratório de Ludologia, Engenharia e Simulação.** Disponível em: <<https://ludes.cos.ufrj.br/pt/ludes/>>. Acesso em: 15 dez. 2021.
- MANGELI, E. An Aesthetic Metric for Multiplayer Turn-based Games. p. 68, 2016.
- Matplotlib — Visualization with Python.** Disponível em: <<https://matplotlib.org/>>. Acesso em: 15 dez. 2021.
- MEEHAN, J. W.; NELSON, R. A.; RICHARDSON, T. V. Competitive Balance and Game Attendance in Major League Baseball. **Journal of Sports Economics**, v. 8, n. 6, p. 563–580, 1 dez. 2007.
- MICHIE, J.; OUGHTON, C. **Competitive balance in football: Trends and effects.** [s.l.] The sportsnexus London, 2004.
- More than half the world watched record-breaking 2018 World Cup.** Disponível em: <<https://www.fifa.com/tournaments/mens/worldcup/2018russia/media-releases/origin1904-p.cxm.fifa.commore-than->>

half-the-world-watched-record-breaking-2018-world-cup>. Acesso em: 15 dez. 2021.

MÜLLER, P. (ED.). Introduction. In: **Modular Specification and Verification of Object-Oriented Programs**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2002. p. 1–37.

NAGHSHBANDI, S. et al. The comparison of competitive balance in Football Premier Leagues of England, Germany, Spain, France, Italy and Iran: A case study from 2009-2010 Season. **Journal of Human Sport and Exercise**, v. 6, n. 4, p. 673–681, 2011.

NBA vs WNBA: Revenue, Salaries, Attendance, Ratings. Disponível em: <<https://www.wsn.com/nba/nba-vs-wnba/>>. Acesso em: 15 dez. 2021.

PEI, A. **This esports giant draws in more viewers than the Super Bowl, and it's expected to get even bigger**. Disponível em: <<https://www.cnn.com/2019/04/14/league-of-legends-gets-more-viewers-than-super-bowl-whats-coming-next.html>>. Acesso em: 15 dez. 2021.

Pesquisa. Disponível em: <<https://ludes.cos.ufrj.br/pt/pesquisa/>>. Acesso em: 15 dez. 2021.

QUIRK, J.; FORT, R. D. **Pay Dirt: The Business of Professional Team Sports**. [s.l.] Princeton University Press, 1997.

RASCHER, D.; SOLMES, J. Do Fans Want Close Contests? A Test of the Uncertainty of Outcome Hypothesis in the National Basketball Association. **Kinesiology (Formerly Exercise and Sport Science)**, 1 ago. 2007.

SANDY, R.; SLOANE, P. J.; ROSENTRUB, M. S. The economics of sport: an international perspective. **The economics of sport: an international perspective.**, 2004.

SILVA, C. D. DA et al. COMPETITIVE BALANCE IN FOOTBALL: A COMPARATIVE STUDY BETWEEN BRAZIL AND THE MAIN EUROPEAN LEAGUES (2003-2016). **Journal of Physical Education**, v. 29, 2018.

TEAM, T. WXPYTHON. **Overview of wxPython**. Disponível em: <<https://wxpython.org/pages/overview/index.html>>. Acesso em: 15 dez. 2021.

WEGNER, P. Object-oriented programming (OOP). In: **Encyclopedia of Computer Science**. GBR: John Wiley and Sons Ltd., 2003. p. 1279–1284.

Why is Python a dynamic language and also a strongly typed language - Python Wiki. Disponível em: <<https://wiki.python.org/moin/Why%20is%20Python%20a%20dynamic%20language%20and%20also%20a%20strongly%20typed%20language>>. Acesso em: 15 dez. 2021.

ZIMBALIST, A. Competitive Balance in Sports Leagues: An Introduction. **Journal of Sports Economics**, v. 3, p. 111–121, 1 jan. 2003.