

Gabriel de Person Pereira de Melo

Uma Análise Estatística do Cenário Competitivo Brasileiro de Valorant™

Belo Horizonte, Minas Gerais

2021

Gabriel de Person Pereira de Melo

Uma Análise Estatística do Cenário Competitivo Brasileiro de Valorant™

Projeto de pesquisa requisito para disciplina de Projeto Orientado a Computação do DCC/UFMG.

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Orientador: Pedro Olmo Stancioli Vaz De Melo

Belo Horizonte, Minas Gerais
2021

Resumo

O trabalho desenvolvido tem como objetivo aplicar métodos de ciências de dados para extração, processamento e análise de dados no âmbito de esportes eletrônicos, especificamente sobre dados de partidas competitivas de times brasileiros de ValorantTM. Na última década o mercado de *e-sports* tem crescido ano após ano, tanto em termos de receita e premiações de campeonatos como em adesão de público em arenas, plataformas de *streaming* e redes sociais. Grandes organizações tem adotado a cultura *data driven* e o uso de dados no direcionamento de decisões e estratégias para a evolução no cenário competitivo. A metodologia do trabalho consiste na extração e armazenamento local de dados web de partidas competitivas, pré-processamento dos conjuntos de dados e aplicação de técnicas de análises descritivas e testes estatísticos para avaliação das principais métricas de performance de times e jogadores. Os métodos de análise aplicados permitiram adquirir um melhor conhecimento sobre a natureza dos dados utilizados, quantificar relações entre atributos numéricos e categóricos e identificar diferenças significativas entre perfis de combate de jogadores. Esses resultados serão utilizados para o desenvolvimento de modelos de aprendizagem de máquina para predição de resultados de partidas na sequência do trabalho.

Palavras-chave: Ciência dos Dados. Esportes Eletrônicos. *Web Scraping*. Correlação de Dados. Testes Estatísticos. Processamento de Dados.

Sumário

1	INTRODUÇÃO	5
1.1	Objetivos Gerais	6
1.2	Objetivos Específicos	6
2	REFERENCIAL TEÓRICO	7
2.1	Conceitos de FPS Tático no Valorant™	7
2.2	<i>Web Scraping</i>	8
2.3	Pré-Processamento de Dados	8
2.4	Correlação de Dados	9
2.5	Testes de hipótese	9
3	METODOLOGIA	11
3.1	Recursos	11
3.2	Extração de Dados	11
3.3	Processamento de Dados	12
3.3.1	Importação e Agregação	12
3.3.2	Engenharia de Atributos	13
3.4	Análises Exploratórias e Estatísticas	14
3.4.1	Análise Exploratória	14
3.4.2	Correlações	14
3.4.3	Testes Estatísticos	15
4	RESULTADOS ESPERADOS	17
4.1	Base de Dados	17
4.2	Análises Descritivas e Exploratórias	18
4.2.1	Estatísticas Descritivas	18
4.2.2	Atributos categóricos	24
4.2.3	Correlações Atributos	25
4.2.4	Testes de Hipóteses	29
4.2.4.1	Testes de Normalidade - <i>Shapiro Wilk</i>	29
4.2.4.2	Testes de Independência	32
5	ETAPAS E CRONOGRAMA	36
	REFERÊNCIAS	37

APÊNDICE A – DESCRIÇÃO ATRIBUTOS BASES GERAIS . . .	39
--	-----------

1 Introdução

Na última década a indústria de Jogos Digitais e *E-Sports* tem crescido ano após ano tanto em termos de receita como de jogadores/espectadores, com mais organizações e grandes marcas investindo no mercado e pessoas consumindo mais produtos e assistindo mais campeonatos através de plataformas de *streaming*. Entre os anos de 2016 e 2017 essa receita chegou a crescer próximo de 41% e a indústria estabilizou-se bem no ano de 2020 em meio à pandemia do COVID-19, transitando as competições disputadas presencialmente em arenas para o meio digital porém limitando campeonatos internacionais às regiões (América do Norte, Europa, Ásia, etc...).



Figura 1 – Crescimento receita do mercado de *E-sports* ao longo dos anos

Aliado ao crescimento desta indústria tem-se a disseminação da cultura *Data Driven* em diversas outras indústrias no mundo todo, onde decisões de negócio são direcionadas e tomadas baseando-se em estudos e análise de dados e indicadores e não somente em suposições, inferências e experiências (DATA-DRIVEN,). No meio dos *E-Sports* a cultura *Data Driven* se estabeleceu através de organizações e times investindo na capacitação de jogadores e comissão técnica, principalmente através de técnicos e analistas que valorizam e utilizam dados em tomadas de decisão e estratégias (VALENS...,) e a adoção de plataformas como *shadow.gg* (*SHADOW...*,) e *mobalytics.gg* (*MOBALYTICS...*,) que utilizam dos dados gerados pelo próprios jogadores em partidas para gerar análises das principais métricas de performance auxiliando os jogadores no estudo de adversários, adoção das melhores estratégias e minimização de falhas (BIG-DATA...,).

Este trabalho se motiva nos métodos de análise de dados utilizados por essas plataformas e demais aplicações de ciência dos dados e na importância de sua utilização para a evolução do esporte eletrônico de alto nível. Os métodos de extração e análise de dados foram aplicados aos dados de partidas do jogo de FPS competitivo Valorant™ (VALORANT,

), título da Riot Games no mercado de jogos de FPS Tático lançado em Junho de 2020 com uma imensa adesão por fãs ao redor do mundo e com diversos campeonatos de alto nível sendo disputados desde seu lançamento com premiações cada vez maiores e tendo já distribuído aproximadamente 5 milhões de dólares em premiações ao redor do mundo (EARNINGS...) ao momento da produção deste texto.

O trabalho é direcionado ao uso de dados de partidas competitivas dos principais times brasileiros para o estudo e desenvolvimento de análises descritivas e estatísticas para identificação das principais métricas de performance de jogadores e times que determinam o sucessos destes.

1.1 Objetivos Gerais

O objetivo geral do trabalho consiste em utilizar métodos de *Web Scraping* junto à linguagem de programação Python na extração dados do *web-site* *www.vlr.gg* de partidas disputadas por times brasileiros, estruturar os dados extraídos em arquivos csv/excel, utilizá-los em processos e métodos de análise estatística e exploratória com a intenção de quantificar as principais métricas de performance de jogadores e times e desenvolver modelos de Aprendizado de Máquina de regressão e classificação para predição de resultados de partidas.

O trabalho será concluído no POC II com o desenvolvimento e avaliação de modelos de Aprendizagem de Máquina focados em regressão e classificação para a predição de resultados. O objetivo final do POC I é a consolidação do processo de extração de dados web para futuras atualizações e das análises e testes estatísticos das principais métricas de performance e diferenciação de perfis de combates entre jogadores de classes e personagens semelhantes. Nesta entrega serão abordados conceitos e métodos inerentes às etapas de extração, tratamento e análise dos dados.

1.2 Objetivos Específicos

Os objetivos específicos do trabalho em questão são:

- Estruturar conceitos de jogo para direcionamento das análises.
- Extração dos dados via *Web Scraping* e armazenamento local.
- Engenharia de atributos através de limpeza, criação e formatação destes.
- Uso de métodos de análises exploratórias e estatísticas para identificar as melhores métricas de performance.
- Desenvolver e avaliar modelos de Aprendizado de Máquina de regressão e classificação para predição de resultados de partidas.

2 Referencial Teórico

2.1 Conceitos de FPS Tático no Valorant™

Valorant™ é um jogo digital do gênero FPS (*First Person Shooter* - Tiro em Primeira Pessoa) onde os jogadores se enfrentam em um combate virtual simulado. O gênero é um dos mais populares do mundo desde a década de 1990 difundido em títulos como *Counter Strike* e *Quake* e popularizado ao longo dos anos com jogadores confrontando agentes inteligentes gerados pelo próprio jogo (modalidade *Single Player*) ou demais jogadores de outras partes do mundo em disputas *online* (modalidade *Multiplayer*).

A principal modalidade competitiva dentro de Valorant™ consiste em dois times de 5 jogadores definidos como Atacantes e Defensores disputando uma partida em um mapa definido pelo próprio jogo dentre um conjunto finito de mapas onde é declarado o time vencedor o primeiro a vencer 13 rodadas. O time atacante vence uma rodada ao eliminar todos os jogadores do time defensor ou ao ativar e garantir a detonação da *Spike* (semelhante à C4 do *Counter Strike*) em locais pré-definidos no mapa (*Spike Sites*) e o time defensor vence uma rodada ao eliminar todos os jogadores do time adversário, desativar a *Spike* caso essa tenha sido ativada ou evitar que o time atacante ative a *Spike*. Jogadores eliminados em uma rodada só renascem na rodada seguinte. Após 12 rodadas os times alternam entre atacantes e defensores, assim para vencer as 13 rodadas um time deve vencer pelo menos 1 rodada como atacante e defensor.

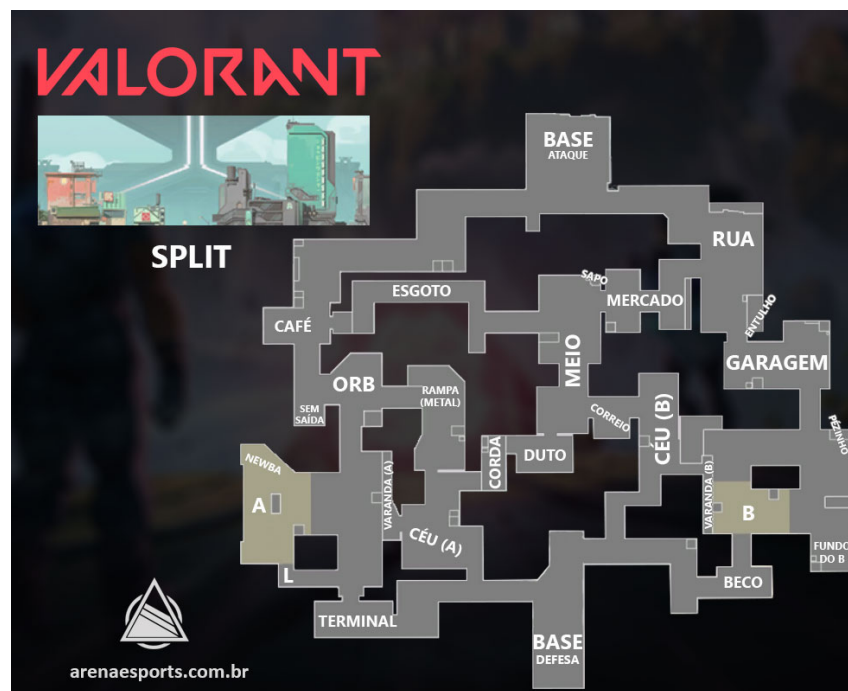


Figura 2 – Esboço do mapa SPLIT e suas regiões

Em Valorant™ cada jogador utiliza um personagem específico pertencente a uma das classes entre Duelista, Iniciador, Controlador e Sentinela e cada um desses possui um conjunto único de habilidades responsáveis porém não limitados a desempenhar funções em uma partida. Todos os personagens possuem os mesmos atributos básicos como velocidade, pontos de vida, altura, dano base, etc... prezando pela natureza tática e competitiva do jogo. A cada rodada jogadores recebem créditos baseado no resultado e no desempenho na rodada anterior (exceto nas primeiras rodadas de cada metade, onde todos começam com 800 créditos) e com esses créditos podem comprar armas de diferentes categorias, valores e habilidades únicas de seus personagens. As diferentes possibilidades de compras geram cenários de vantagens e desvantagens entre os times, direcionando as estratégias utilizadas para abater os adversários, dominar territórios no mapa e vencer a rodada.

2.2 Web Scraping

Web Scraping (MITCHELL, 2018) ou Raspagem de dados, em português, é definido como o método de extrair dados de páginas *web* sem necessariamente utilizar de uma API (*Application Programming Interface*) mas utilizando de *scripts* automatizados que acessam páginas públicas e recuperam dados contidos em *tags HTML* (*Hypertext Markup Language*). *Scrapers* ou raspadores possibilitam a extração e processamento de quantidades massivas de dados em tempo hábil principalmente para informações específicas e em domínios que não dispõe de uma API para esta finalidade.

O objetivo do uso de *Web Scraping* neste trabalho é identificar e extrair dados de partidas dos principais times e jogadores de Valorant™ do Brasil do *website* *www.vlr.gg* (VLR...), uma vez que o próprio jogo e o *website* não dispõem de APIs públicas para a incorporação destes dados.

2.3 Pré-Processamento de Dados

O Pré-processamento dos dados é realizado desde a etapa de extração dos dados até as análises exploratórias com o objetivo de modificar e organizar os dados extraídos obtendo melhores resultados nas análises descritivas e testes de hipóteses.

Algumas das principais operações realizadas durante esta etapa são a identificação das principais métricas de avaliação, tratamento/remoção de valores nulos, filtragem de dados com base em valores categóricos, agrupamento de valores numéricos contínuos por médias e somas. Esses processos são melhor apresentados na sessões 3.3 e 3.4

2.4 Correlação de Dados

Para a análise dos principais atributos de performance de natureza e distribuição contínua são utilizados medidas estatísticas de correlação, através das quais é possível identificar níveis de correlação entre dois ou mais atributos, ou seja, o quanto a variação um atributo influencia na variação dos demais.

Neste trabalho foram utilizados os métodos de correlação de *Pearson* e *Spearman* (TOWARDS...,) para identificação de correlações lineares e não-lineares respectivamente. Ambos os métodos possuem um conjunto de resultados de $[-1, 1]$, onde -1 representa correlação negativa absoluta entre os atributos (a variação positiva de um atributo implica na variação negativa do outro e vice-versa) e 1 representa correlação positiva absoluta (ambos os atributos variam na mesma taxa e na mesma direção, positiva ou negativa). O valor 0 (zero) representa ausência de correlação linear entre os atributos, podendo haver correlação não-linear. As formulas das correlações de *Pearson* e *Spearman* estão demonstrados logo abaixo:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Figura 3 – Fórmula método de correlação de *Pearson*

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Figura 4 – Fórmula método de correlação de *Spearman*

2.5 Testes de hipótese

A metodologia do teste de hipóteses tem por objetivo formular hipóteses sobre os parâmetros de um modelo de população, quando suspeita-se que haja alguma alteração em seu valor. Um teste sobre um parâmetro envolve uma hipótese nula (H_0) e a alternativa (H_1); uma estatística; um valor para o nível de significância e a região crítica. Se a estatística cair dentro da região crítica (RC), H_0 será rejeitada, em caso contrário, será mantida. A região crítica é determinada a partir do cálculo da probabilidade condicional de que ocorra o valor observado na estatística dado que H_0 é verdadeira. Se esta probabilidade

for menor que α (nível de significância do teste), então a hipótese nula é rejeitada. Caso contrário, a hipótese é mantida. (BRAGA,)

Neste trabalho são utilizados testes de hipóteses aliados ao processo de análise exploratória no estudo da natureza dos dados utilizados através de teste que validam distribuições Gaussianas e na escolha dos melhores atributos a serem utilizados em modelos de Aprendizagem de Máquina. Para os testes utilizados o valor de α foi definido em 0.05 (nível de significância 95%).

3 Metodologia

3.1 Recursos

Para o desenvolvimento do trabalho foi utilizado o sistema operacional Linux (Ubuntu 20.04) embutido no ambiente WSL (*Windows Subsystem for Linux*), módulo do sistema operacional Windows 10 que disponibiliza um ambiente Linux compatível com o sistema da Microsoft. Foi utilizado um computador pessoal com processador AMD Ryzen 5 5600X a 4.6 GHz, 16GB de RAM a 3200MHz, Placa de Vídeo Nvidia RTX 3060Ti, armazenamento SSD Nvme M.2 de 512GB.

Foram utilizadas as linguagens de programação **Bash** para a execução de *scripts* Python, linguagem Python para extração, limpeza e análise dos dados através de *scripts* e utilizando o ambiente de desenvolvimento *Jupyter Notebook*.

3.2 Extração de Dados

Os dados utilizados no desenvolvimento do trabalho foram extraídos do *website* www.vlr.gg (VLR...), considerado um dos maiores portais de divulgação de conteúdo e resultados do cenário competitivo de Valorant™ mundial, disponibilizando resultados de partidas, estatísticas de times e jogadores, ranqueamento de times por cada região de disputada, fóruns de discussão e notícias recentes sobre times e jogadores. Dados utilizados sujeitos à qualidade e disponibilidade destes no próprio *website*.


Os dados de cada partida selecionada foram extraídos utilizando-se as bibliotecas **Urllib** (URLLIB...), **Beautiful Soup** (BEAUTIFUL...), e **Pandas** (PANDAS...), onde cada página de interesse é acessada através de uma URL (*Uniform Resource Locator*) específica para cada série de partidas disputadas entre dois times e para cada partida são extraídos dados de interesse das abas *Overview*, *Performance* e *Economy*, previamente nomeadas pelos criadores do portal. A figura 5 exemplifica como os dados são dispostos na aba *Overview*.

As informações de desempenho dos jogadores para a série e cada mapa dispostas em formato de tabelas *html* são extraídas utilizando-se a função `read_html` da biblioteca **Pandas**, que identifica automaticamente os dados dessas tabelas e armazena em estruturas do tipo *Dataframes* no ambiente de execução. As demais informações dispostas na página, como rodadas vencidas por cada time, mapas disputados, versão do jogo e agentes utilizados são acessados e extraídos através de suas respectivas *tags html* e armazenados em listas e variáveis de ambiente.

Uma vez disponíveis no ambiente de desenvolvimento os *Dataframes* são pre-processados

através de tratamento de valores nulos, conversão de valores, criação de novos atributos e adição dos demais dados extraídos da página aos times correspondentes. Esse processo repete-se para cada URL lida de um arquivo texto de entrada do *script* Python e ao final da execução todos os *Dataframes* gerados são concatenados e salvos em um arquivo CSV ou Excel para cada aba da página.

Ao momento da produção deste texto e da entrega final do POC I a base de dados extraídos é composta por dados de 12 dos principais times do Brasil (Gamelanders, Furia, Black Dragons, Havan Liberty, Imperial, Ingaming, Pain Gaming, Rise Gaming, Sharks Esports, Slick, Team Vikings, Vorax) e seus adversários com partidas disputadas entre os meses de Setembro de 2020 e Março de 2021. Embora existam partidas oficiais datadas de desde Junho de 2020, mês do lançamento oficial de Valorant™, essas partidas não possuem informações suficientes disponíveis no *website*, portanto foram desconsideradas.



	ACS	K	D	A	+/-	ADR	HS%	FK	FD	+/-
mwzera GL	293	67	57	15	+10	183.2	19%	14	7	+7
Jonn GL	246	56	60	13	-4	146.6	24%	11	10	+1
Nyang GL	189	44	39	23	+5	113.2	19%	4	7	-3
JhoW GL	146	37	49	15	-12	91.6	16%	3	4	-1
fznnn GL	130	26	50	25	-24	110.9	17%	2	6	-4
saadhak VKS	253	59	43	35	+16	163.0	14%	6	5	+1
Sacy VKS	232	53	46	30	+7	140.9	34%	9	3	+6
gtznIN VKS	213	49	46	15	+3	126.2	13%	5	8	-3
frz VKS	210	50	44	9	+6	118.6	16%	10	6	+4
sutecas VKS	171	44	51	14	-7	119.9	32%	4	12	-8

Figura 5 – Aba *Overview* no website *www.vlr.gg*

3.3 Processamento de Dados

3.3.1 Importação e Agregação

Inicialmente os dados de cada time extraídos e salvos no disco rígido local são carregados em um ambiente *Jupyter Notebook* utilizando-se a função **pandas.read_csv**, assim para os arquivos para cada tipo de dado entre *Overview*, *Performance* e *Economy* estes são concatenados em *Dataframes* maiores contendo as informações de todos os times e jogadores.

O *Dataframe* **base_geral_jogadores** é resultante do agrupamento entre as bases gerais de *Overview* e *Performance* através da função **pandas.merge** (semelhante a um JOIN em SQL) utilizando-se os atributos comuns **match_id**, **Map** e **Player** para identificar ocorrências únicas de cada jogador em cada mapa para cada série. O *Dataframe* possui aproximadamente 3360 observações de 67 atributos de jogadores e de seus respectivos times, como ACS (*Average Combat Score*), Abates, Mortes, Assistências, Primeiros Abates (*First Kills*), rodadas vencidas e oponente. Este será utilizado nas análises de atributos e métricas de performance entre diferentes jogadores para diferentes personagens e diferentes classes.

O *Dataframe* **base_geral_times** é resultante do agrupamento entre a base de dados *Economy* de todos os times e a agregação de valores médios e somatórios de atributos de jogadores para cada time em cada mapa disputado de cada série. As agregações são realizadas através da função **pandas.group_by** utilizando-se os parâmetros *mean* (média) e *sum* (soma) para agregação de valores, totalizando 520 observações de 79 atributos. O agrupamento entre as bases é realizado através a função **pandas.merge** utilizando-se os atributos comuns **match_id**, **Map** e **Team** para identificar ocorrências únicas de cada time em cada mapa para cada série. Este será utilizado nas análises de principais métricas entre jogadores de um time que direcionam um time aos melhores resultados entre as partidas para diferentes mapas.

Descrição das colunas das bases gerais disponível nas tabelas 17 e 18 no Apêndice A.

3.3.2 Engenharia de Atributos

Aliado ao processo de agrupamento e agregação dos dados dos times e criação das bases maiores, novos atributos para estas bases são criados, estes normalizados/balanceados pela quantidade de rodadas disputados em cada partida para jogadores - abates/mortes/assistências por rodada, etc... - e times (porcentagem total de rodadas vencidas e por lado - atacantes ou defensores) com a intenção de equalizar estatísticas entre jogadores e times que tenham disputado um número maior ou menor de rodadas em cada mapa ou rodadas totais ao longo de todos os mapas.

Para os agentes utilizados por cada jogador são criadas colunas binárias, também conhecidas como *dummies* que representam se um agente foi escolhido ou não para cada jogador.

Para os principais times é realizada a padronização de seus nomes devido à frequentes mudanças de organizações e nomenclaturas no cenário (Ex.: O atual time da Pain Gaming era conhecido como No Org e No Org 2.0 antes de ser contratado pela organização).

3.4 Análises Exploratórias e Estatísticas

O processo de análises exploratórias, inferências estatísticas e teste de hipóteses foi conduzido com a intenção de prover melhores visualizações e conhecimento sobre os dados utilizados no desenvolvimento do projeto, através de correlações de atributos, contagem absoluta de ocorrências de valores, estatísticas descritivas - média, mediana, quartis, etc... - e distribuição dos dados.

O processo de análises é conduzido utilizando-se as bibliotecas **Pandas**, **Scipy** (SCIPY...), **Seaborn** (SEABORN...), as primeiras são utilizadas em manipulações, análises descritivas e testes estatísticos e a última na geração de gráficos para visualização dos dados em diferentes contextos, como gráficos de dispersão, densidade, contagem e distribuição dos dados.

3.4.1 Análise Exploratória

Na condução das análises exploratórias a biblioteca *Pandas* foi a principal ferramenta utilizada devido à diversidade de funções e métodos presentes na própria biblioteca para a condução dessas análises. Os métodos presentes na biblioteca variam desde amostras dos dados utilizados (**pandas.head**), descrição de métricas de distribuição dos dados - máximos, mínimos, medianas e quartis (**pandas.describe**) - e relatórios sobre tipos de valores e quantidades de valores não-nulos para cada atributo (**pandas.info**).

Esses métodos proveem um ponto de partida para as análises realizadas ao longo do projeto, sendo possível identificar valores e métricas importantes para cada atributo, assim como a possibilidade de replicar esses métodos para amostras filtradas da base de dados entre os possíveis valores de atributos categóricos desta (Mapas, Agentes, Times, Oponentes, Resultado).

As visualizações/gráficos que apoiam o processo de análise exploratória são conduzidos utilizando-se a biblioteca **Seaborn** com a qual foram geradas visualizações tanto para atributos únicos utilizando-se gráficos de distribuição/densidade (**seaborn.displot/seaborn.boxplot**) como para conjuntos de 2 ou mais atributos (**seaborn.scatterplot/seaborn.pairplot**). Estas visualizações são utilizadas de referência e apoio a testes de normalidade de distribuição dos atributos numéricos e contínuos, correlações entre atributos numéricos e testes de hipóteses de significância de independência entre amostras.

3.4.2 Correlações

Em virtude da grande maioria de ocorrência de dados de natureza contínua na base de dados optou-se pelo uso de testes de correlação, mais especificamente *Pearson* e *Spearman*, para analisar e quantificar relações entre os diversos atributos. Essas correlações são

analisadas e calculadas através das próprias funções da biblioteca **Scipy** junto a gráficos de dispersão entre variáveis contínuas e de mapas de calor (**seaborn.heatmap**).

Entre os atributos que tangem a performance individual dos jogadores, foram utilizados os dois métodos de correlação e seus resultados foram posteriormente comparados. Jogadores geram impacto nas rodadas através de abates, assistências, dano infringido (através de armas ou habilidades), ativação e desativação da *Spike*, sucesso em *clutches* (situações onde o último jogador vence uma rodada contra 1 ou mais oponentes) dentre outras maneiras que não tangem somente ao combate. Esses valores são também utilizados pelo próprio Valorant™ no cálculo da principal métrica de performance, o ACS (*Average Combat Score* - Pontuação Média de Combate), porém vê-se a necessidade de quantificar melhor a relação destes atributos isolados com o ACS, tanto para jogadores como um todo, como para diferentes agentes e classes.

Tratando-se de performance dos times entre as partidas tem-se os atributos individuais de seus jogadores agregados em médias e somas para cada mapa disputado e atributos que tratam o time como uma unidade, como total e percentagem de rodadas vencidas por faixa de gasto econômico, total e percentagem de rodadas vencidas no mapa e por lado (atacante ou defensor). A principal métrica alvo utilizada para estas correlações é a percentagem de rodadas vencidas em um mapa para correlações lineares e o resultado em si da partida (Vitória ou Derrota) para correlações não-lineares.

3.4.3 Testes Estatísticos

Neste trabalho foram utilizados testes de hipóteses para validar/refutar inferências sobre a distribuição dos dados e independência entre amostras, mais especificamente se o conjunto de valores de um certo atributo assumem uma distribuição normal/Gaussiana e se as médias/medianas entre N grupos são iguais.

Para a validação de normalidade dos conjuntos de dados contínuos utiliza-se ferramentas de visualização de gráficos de densidade e distribuição (**seaborn.displot**) dos dados e o teste de *Shapiro Wilk*(SHAPIRO...,) da biblioteca **Scipy**, o qual testa a hipótese nula H_0 : Os dados estão distribuídos uniformemente em torno de um valor μ e com um desvio padrão σ^2 . (DISTRIBUICAO...,) O uso deste teste visa proporcionar um melhor entendimento dos dados utilizados neste trabalho, assim como um indicativo de qual o melhor teste de independência de amostras a ser utilizados.

Aliado à validação de normalidade de dados, testes de análise de variância entre amostras também são utilizados neste trabalho com a intenção de testar a hipótese nula H_0 : Não há uma diferença significativa entre as amostras dos grupos avaliados.

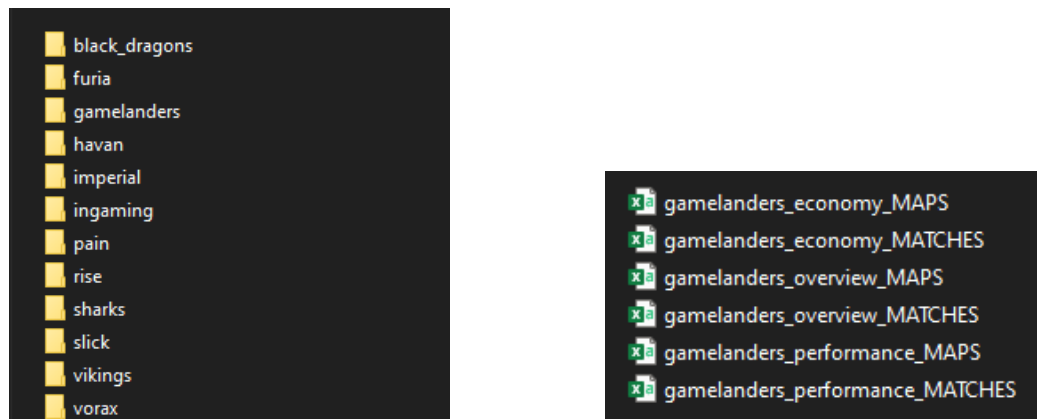
São utilizados os testes *ANOVA* (*Analysis of Variance*)(ANOVA...,) , *Kruskal Wallis* (KRUSKAL...,) e *chi² contingency*(CHI2...,) da biblioteca **Scipy**. *ANOVA* e *Kruskal Wallis* são utilizados para comparar distribuições contínuas de dados, porém o primeiro utiliza os valores médios das distribuições em amostras normalmente distribuídas para

validar ou rejeitar a hipótese nula, já o segundo compara as medianas das distribuições em amostras que não são normalmente distribuídas se tratando de um teste não-paramétrico. O teste *chi² contingency* é utilizado em amostras de natureza discreta e independentes através de comparação das frequências observadas.

4 Resultados Esperados

4.1 Base de Dados

Um dos principais objetivos no desenvolvimento deste trabalho é a estruturação de *scripts* Python para a extração via *Web Scraping* do website *www.vlr.gg* e de diretórios locais para armazenamento de dados dos principais times e jogadores de Valorant™ do brasil, uma vez que não há a possibilidade de se obter esses dados diretamente com a desenvolvedora Riot Games através de uma API pública.



Conjunto de diretórios de arquivos CSV de times Conjunto de arquivos do time Gamelanders

Figura 6 – Exemplo de gráficos de distribuição utilizados

Os dados utilizado neste trabalho tem sido atualizados ao longo de todo o processo de desenvolvimento do projeto, contendo partidas competitivas do cenário brasileiro de Valorant™ entre os meses de Setembro de 2020 e Março de 2021 dos principais times do país (Gamelanders, Furia, Black Dragons, Havan Liberty, Imperial, Ingaming, Pain Gaming, Rise Gaming, Sharks Esports, Slick, Team Vikings, Vorax).

Os dados extraídos e pré-processados ocupam um espaço de 1.15 MB (*Megabytes*) em armazenamento local. Estes serão atualizados ao longo do desenvolvimento do POC II devido ao surgimento e disputa de novos campeonatos pelos times selecionados.

4.2 Análises Descritivas e Exploratórias

Os métodos de análises exploratórias utilizados tem como objetivo estudar e compreender a natureza dos dados extraídos e utilizados neste trabalho. Nesta sessão são apresentados resultados do métodos de processamento e análise exposto nas sessões 3.3 e 3.4

4.2.1 Estatísticas Descritivas

A seguir são apresentados resumos sobre as principais bases de dados agregados utilizadas no trabalho - base geral de jogadores e base geral de times - como tipos e quantidades de valores não-nulos de cada atributo em cada base e as principais estatísticas descritivas sobre estas.

As tabelas 1 e 2 referenciam a base de dados de jogadores. Pela primeira tem-se 23 dos 67 atributos como numéricos de ponto flutuante, 22 numéricos inteiros (todos estes 55 atributos contínuos), 8 atributos categóricos e 14 discretos (binários), estes representando o uso ou não uso de certo agente pelo jogador. Pela segunda tem-se as estatísticas descritivas sobre os atributos numéricos da base como média, desvio padrão e intervalos de quantis, estes proporcionam uma primeira noção sobre a distribuição dos valores dos atributos. Todos os valores nulos foram previamente tratados.

#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype
0	Player	3366 non-null	object	34	2K	3366 non-null	int64
1	Agents	3366 non-null	object	35	3K	3366 non-null	int64
2	ACS	3366 non-null	int64	36	4K	3366 non-null	float64
3	K	3366 non-null	int64	37	5K	3366 non-null	float64
4	D	3366 non-null	int64	38	1v1	3366 non-null	float64
5	A	3366 non-null	int64	39	1v2	3366 non-null	float64
6	KD_DIFF	3366 non-null	int64	40	1v3	3366 non-null	float64
7	ADR	3366 non-null	float64	41	1v4	3366 non-null	float64
8	HS%	3366 non-null	int64	42	1v5	3366 non-null	float64
9	FK	3366 non-null	int64	43	ECON	3366 non-null	int64
10	FD	3366 non-null	int64	44	PL	3366 non-null	int64
11	FK_FD_DIFF	3366 non-null	int64	45	DE	3366 non-null	int64
12	match_id	3366 non-null	int64	46	Team_y	3366 non-null	object
13	Team_x	3366 non-null	object	47	Opp_Team_y	3366 non-null	object
14	Opp_Team_x	3366 non-null	object	48	Patch_y	3366 non-null	float64
15	rounds_won	3366 non-null	int64	49	total_mult_kill	3366 non-null	float64
16	rounds_lost	3366 non-null	int64	50	total_clutch	3366 non-null	float64
17	ct_rounds_won	3366 non-null	int64	51	MKPR	3366 non-null	float64
18	ct_rounds_lost	3366 non-null	int64	52	CPR	3366 non-null	float64
19	t_rounds_won	3366 non-null	int64	53	breach	3366 non-null	uint8
20	t_rounds_lost	3366 non-null	int64	54	brimstone	3366 non-null	uint8
21	Map	3366 non-null	object	55	cypher	3366 non-null	uint8
22	Patch_x	3366 non-null	float64	56	jett	3366 non-null	uint8
23	total_rounds	3366 non-null	int64	57	killjoy	3366 non-null	uint8
24	win_rate	3366 non-null	float64	58	omen	3366 non-null	uint8
25	KPR	3366 non-null	float64	59	phoenix	3366 non-null	uint8
26	DPR	3366 non-null	float64	60	raze	3366 non-null	uint8
27	APR	3366 non-null	float64	61	reyna	3366 non-null	uint8
28	FKPR	3366 non-null	float64	62	sage	3366 non-null	uint8
29	FKWR	3366 non-null	float64	63	skye	3366 non-null	uint8
30	FDPR	3366 non-null	float64	64	sova	3366 non-null	uint8
31	ct_wr	3366 non-null	float64	65	viper	3366 non-null	uint8
32	t_wr	3366 non-null	float64	66	yoru	3366 non-null	uint8
33	RESULT	3366 non-null	object				

Tabela 1 – Tipos dos atributos da base geral de jogadores

column	count	mean	std	min	25%	50%	75%	max
ACS	3366.00	200.63	63.83	12.00	156.00	196.00	239.00	640.00
K	3366.00	14.57	5.56	0.00	11.00	14.00	18.00	43.00
D	3366.00	14.59	3.90	0.00	13.00	15.00	17.00	32.00
A	3366.00	5.29	3.03	0.00	3.00	5.00	7.00	26.00
KD_DIFF	3366.00	-0.03	6.30	-16.00	-5.00	0.00	4.00	29.00
ADR	3366.00	127.02	39.53	12.60	99.22	125.20	151.10	332.50
HS%	3366.00	21.38	8.78	0.00	15.00	21.00	26.00	67.00
FK	3366.00	2.08	1.75	0.00	1.00	2.00	3.00	11.00
FD	3366.00	2.09	1.62	0.00	1.00	2.00	3.00	10.00
FK_FD_DIFF	3366.00	-0.00	2.17	-8.00	-1.00	0.00	1.00	10.00
rounds_won	3366.00	10.45	3.74	0.00	8.00	13.00	13.00	22.00

rounds_lost	3366.00	10.46	3.73	0.00	8.00	13.00	13.00	22.00
ct_rounds_won	3366.00	5.26	2.74	0.00	3.00	5.00	7.00	12.00
ct_rounds_lost	3366.00	4.93	2.65	0.00	3.00	5.00	7.00	12.00
t_rounds_won	3366.00	4.92	2.65	0.00	3.00	5.00	7.00	12.00
t_rounds_lost	3366.00	5.27	2.74	0.00	3.00	5.00	7.00	12.00
total_rounds	3366.00	20.90	3.95	13.00	18.00	21.00	23.00	42.00
win_rate	3366.00	49.98	18.58	0.00	38.10	47.62	61.90	100.00
KPR	3366.00	0.70	0.24	0.00	0.53	0.68	0.83	2.46
DPR	3366.00	0.70	0.15	0.00	0.61	0.71	0.80	1.12
APR	3366.00	0.25	0.14	0.00	0.15	0.24	0.33	0.94
FKPR	3366.00	0.10	0.08	0.00	0.04	0.09	0.14	0.56
FKWR	3366.00	47.42	31.18	0.00	25.00	50.00	66.67	100.00
FDPR	3366.00	0.10	0.08	0.00	0.05	0.09	0.14	0.53
ct_wr	3366.00	51.11	24.38	0.00	33.33	50.00	66.67	100.00
t_wr	3366.00	48.84	24.40	0.00	33.33	50.00	66.67	100.00
2K	3366.00	2.65	1.67	0.00	1.00	2.00	4.00	9.00
3K	3366.00	0.81	0.96	0.00	0.00	1.00	1.00	6.00
4K	3366.00	0.14	0.39	0.00	0.00	0.00	0.00	3.00
5K	3366.00	0.02	0.14	0.00	0.00	0.00	0.00	1.00
1v1	3366.00	0.20	0.45	0.00	0.00	0.00	0.00	3.00
1v2	3366.00	0.10	0.31	0.00	0.00	0.00	0.00	3.00
1v3	3366.00	0.03	0.17	0.00	0.00	0.00	0.00	3.00
1v4	3366.00	0.00	0.07	0.00	0.00	0.00	0.00	1.00
1v5	3366.00	0.00	0.02	0.00	0.00	0.00	0.00	1.00
ECON	3366.00	52.37	19.00	5.00	40.00	50.00	62.00	225.00
PL	3366.00	1.38	1.79	0.00	0.00	1.00	2.00	13.00
DE	3366.00	0.43	0.68	0.00	0.00	0.00	1.00	5.00
total_mult_kill	3366.00	3.61	2.06	0.00	2.00	3.00	5.00	12.00
total_clutch	3366.00	0.33	0.57	0.00	0.00	0.00	1.00	4.00
MKPR	3366.00	0.17	0.10	0.00	0.10	0.17	0.24	0.77
CPR	3366.00	0.02	0.03	0.00	0.00	0.00	0.04	0.20
breach	3366.00	0.04	0.21	0.00	0.00	0.00	0.00	1.00
brimstone	3366.00	0.02	0.13	0.00	0.00	0.00	0.00	1.00
cypher	3366.00	0.15	0.36	0.00	0.00	0.00	0.00	1.00
jett	3366.00	0.14	0.35	0.00	0.00	0.00	0.00	1.00
killjoy	3366.00	0.05	0.21	0.00	0.00	0.00	0.00	1.00
omen	3366.00	0.18	0.39	0.00	0.00	0.00	0.00	1.00
phoenix	3366.00	0.03	0.18	0.00	0.00	0.00	0.00	1.00

raze	3366.00	0.14	0.34	0.00	0.00	0.00	0.00	1.00
reyna	3366.00	0.05	0.22	0.00	0.00	0.00	0.00	1.00
sage	3366.00	0.04	0.19	0.00	0.00	0.00	0.00	1.00
skye	3366.00	0.00	0.04	0.00	0.00	0.00	0.00	1.00
sova	3366.00	0.16	0.37	0.00	0.00	0.00	0.00	1.00
viper	3366.00	0.00	0.03	0.00	0.00	0.00	0.00	1.00
yoru	3366.00	0.00	0.02	0.00	0.00	0.00	0.00	1.00

Tabela 2 – Estatísticas descritivas da base geral de jogadores

As tabelas 3 e 4 referenciam a base de dados de times. Pela primeira tem-se 35 dos 79 atributos como numéricos de ponto flutuante, 24 numéricos inteiros (todos estes 55 atributos contínuos), 6 atributos categóricos e 14 discretos (binários), estes representando o uso ou não uso de certo agente pelo time. Pela segunda tem-se referencia estatísticas descritivas sobre os atributos numéricos da base como média, desvio padrão e intervalos de quantis, estes proporcionam uma primeira noção sobre a distribuição dos valores dos atributos. Todos os valores nulos foram previamente tratados.

Column	Non-Null Count	Dtype	Column	Non-Null Count	Dtype
Team_x	523 non-null	object	brimstone	523 non-null	uint8
match_id	523 non-null	int64	viper	523 non-null	uint8
Map	523 non-null	object	cypher	523 non-null	uint8
ACS	523 non-null	float64	sage	523 non-null	uint8
K	523 non-null	float64	killjoy	523 non-null	uint8
D	523 non-null	float64	breach	523 non-null	uint8
A	523 non-null	float64	sova	523 non-null	uint8
KD_DIFF	523 non-null	float64	skye	523 non-null	uint8
ADR	523 non-null	float64	Opp_Team_x	523 non-null	object
HS%	523 non-null	float64	rounds_won	523 non-null	int64
ECON	523 non-null	float64	rounds_lost	523 non-null	int64
FK	523 non-null	int64	ct_rounds_won	523 non-null	int64
FD	523 non-null	int64	ct_rounds_lost	523 non-null	int64
KPR	523 non-null	float64	t_rounds_won	523 non-null	int64
DPR	523 non-null	float64	t_rounds_lost	523 non-null	int64
APR	523 non-null	float64	Patch_x	523 non-null	float64
FKPR	523 non-null	float64	total_rounds	523 non-null	int64
FDPR	523 non-null	float64	win_rate	523 non-null	float64
2K	523 non-null	int64	ct_wr	523 non-null	float64
3K	523 non-null	int64	t_wr	523 non-null	float64
4K	523 non-null	float64	RESULT	523 non-null	object
5K	523 non-null	float64	Team	523 non-null	object
1v1	523 non-null	float64	Pistol_W	523 non-null	int64
1v2	523 non-null	float64	Pistol_P	523 non-null	int64
1v3	523 non-null	float64	Pistol_WR	523 non-null	float64
1v4	523 non-null	float64	Eco_0_5_P	523 non-null	int64
1v5	523 non-null	float64	Eco_0_5_W	523 non-null	int64
PL	523 non-null	int64	Eco_0_5_WR	523 non-null	float64
DE	523 non-null	int64	Semi_Eco_5_10_P	523 non-null	int64
total_mult_kill	523 non-null	float64	Semi_Eco_5_10_W	523 non-null	int64
MKPR	523 non-null	float64	Semi_Eco_5_10_WR	523 non-null	float64
total_clutch	523 non-null	float64	Semi_Buy_10_20_P	523 non-null	int64
CPR	523 non-null	float64	Semi_Buy_10_20_W	523 non-null	int64
jett	523 non-null	uint8	Semi_Buy_10_20_WR	523 non-null	float64
raze	523 non-null	uint8	Full_Buy_20_P	523 non-null	int64
phoenix	523 non-null	uint8	Full_Buy_20_W	523 non-null	int64
reyna	523 non-null	uint8	Full_Buy_20_WR	523 non-null	float64
yoru	523 non-null	uint8	Opp_team	523 non-null	object
omen	523 non-null	uint8	Num_maps	523 non-null	float64
Patch	523 non-null	float64			

Tabela 3 – Tipos dos atributos da base geral de times

column	count	mean	std	min	25%	50%	75%	max
ACS	523.00	204.12	27.81	47.00	187.60	206.60	223.00	268.60
K	523.00	14.89	3.35	3.80	13.20	15.20	16.80	29.20
D	523.00	14.33	3.64	3.60	12.20	14.80	16.40	29.40
A	523.00	5.45	1.69	1.20	4.20	5.40	6.60	13.20
KD_DIFF	523.00	0.57	4.02	-9.60	-2.60	1.00	3.80	9.60
ADR	523.00	128.90	17.27	22.28	119.80	130.00	139.69	179.66
HS%	523.00	21.51	3.93	12.60	18.60	21.40	23.80	35.00
ECON	523.00	53.41	10.41	29.20	47.00	51.80	58.10	113.80

column	count	mean	std	min	25%	50%	75%	max
FK	523.00	10.77	3.22	2.00	9.00	11.00	13.00	22.00
FD	523.00	10.17	3.45	1.00	8.00	10.00	12.00	22.00
KPR	523.00	3.57	0.60	1.36	3.16	3.60	3.98	5.00
DPR	523.00	3.42	0.63	1.36	3.04	3.41	3.80	5.93
APR	523.00	1.31	0.36	0.40	1.06	1.28	1.53	2.47
FKPR	523.00	0.52	0.14	0.10	0.43	0.52	0.61	0.94
FDPR	523.00	0.48	0.14	0.06	0.39	0.48	0.58	0.97
2K	523.00	13.58	4.46	1.00	10.00	14.00	17.00	28.00
3K	523.00	4.19	2.13	0.00	3.00	4.00	6.00	12.00
4K	523.00	0.76	0.85	0.00	0.00	1.00	1.00	4.00
5K	523.00	0.11	0.33	0.00	0.00	0.00	0.00	2.00
1v1	523.0	1.0	1.0	0.0	0.0	1.0	1.5	6.0
1v2	523.00	0.51	0.70	0.00	0.00	0.00	1.00	4.00
1v3	523.00	0.13	0.37	0.00	0.00	0.00	0.00	3.00
1v4	523.00	0.02	0.13	0.00	0.00	0.00	0.00	1.00
1v5	523.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PL	523.00	7.11	2.94	0.00	5.00	8.00	9.00	13.00
DE	523.00	2.20	1.58	0.00	1.00	2.00	3.00	8.00
total_mk	523.00	18.63	5.30	1.00	15.00	19.00	22.00	35.00
MKPR	523.00	0.89	0.23	0.07	0.73	0.90	1.05	1.58
total_clutch	523.00	1.66	1.26	0.00	1.00	2.00	2.00	8.00
CPR	523.00	0.08	0.06	0.00	0.04	0.07	0.12	0.40
jett	523.00	0.68	0.48	0.00	0.00	1.00	1.00	2.00
raze	523.00	0.71	0.45	0.00	0.00	1.00	1.00	1.00
phoenix	523.00	0.18	0.38	0.00	0.00	0.00	0.00	1.00
reyna	523.00	0.25	0.43	0.00	0.00	0.00	1.00	1.00
yoru	523.00	0.00	0.06	0.00	0.00	0.00	0.00	1.00
omen	523.00	0.92	0.27	0.00	1.00	1.00	1.00	1.00
brimstone	523.00	0.08	0.27	0.00	0.00	0.00	0.00	1.00
viper	523.00	0.01	0.09	0.00	0.00	0.00	0.00	1.00
cypher	523.00	0.78	0.41	0.00	1.00	1.00	1.00	1.00
sage	523.0	0.2	0.4	0.0	0.0	0.0	0.0	2.0
killjoy	523.00	0.20	0.41	0.00	0.00	0.00	0.00	2.00
breach	523.0	0.2	0.4	0.0	0.0	0.0	0.0	1.0
sova	523.00	0.79	0.41	0.00	1.00	1.00	1.00	1.00
skye	523.00	0.01	0.10	0.00	0.00	0.00	0.00	1.00
rounds_won	523.00	10.91	3.52	1.00	8.00	13.00	13.00	22.00
rounds_lost	523.00	10.04	3.86	0.00	7.00	11.00	13.00	22.00

column	count	mean	std	min	25%	50%	75%	max
ct_rounds_won	523.00	5.44	2.69	0.00	3.50	5.00	7.00	12.00
ct_rounds_lost	523.00	4.73	2.60	0.00	3.00	5.00	7.00	12.00
t_rounds_won	523.00	5.20	2.66	0.00	3.00	5.00	7.00	12.00
t_rounds_lost	523.00	5.05	2.71	0.00	3.00	5.00	7.00	12.00
total_rounds	523.00	20.95	3.97	13.00	18.00	21.00	23.00	42.00
win_rate	523.00	52.52	18.17	7.14	38.10	54.17	65.00	100.00
ct_wr	523.00	53.48	24.02	0.00	36.36	50.00	72.73	100.00
t_wr	523.00	51.52	24.21	0.00	33.33	50.00	66.67	100.00
Pistol_W	523.00	1.06	0.71	0.00	1.00	1.00	2.00	2.00
Pistol_P	523.0	2.0	0.0	2.0	2.0	2.0	2.0	2.0
Pistol_WR	523.00	52.87	35.44	0.00	50.00	50.00	100.00	100.00
Eco_P	523.00	3.02	1.07	2.00	2.00	3.00	3.00	15.00
Eco_W	523.00	1.16	0.82	0.00	1.00	1.00	2.00	9.00
Eco_WR	523.00	44.97	33.83	0.00	25.00	33.00	67.00	100.00
Semi_Eco_P	523.00	1.40	1.17	0.00	0.00	1.00	2.00	5.00
Semi_Eco_W	523.00	0.26	0.51	0.00	0.00	0.00	0.00	3.00
Semi_Eco_WR	523.00	13.91	29.27	0.00	0.00	0.00	0.00	100.00
Semi_Buy_P	523.00	4.14	2.02	0.00	3.00	4.00	5.00	11.00
Semi_Buy_W	523.00	2.28	1.40	0.00	1.00	2.00	3.00	8.00
Semi_Buy_WR	523.00	56.74	29.78	0.00	38.00	57.00	75.00	100.00
Full_Buy_P	523.00	12.41	4.05	1.00	10.00	12.00	14.50	35.00
Full_Buy_W	523.00	7.22	3.27	0.00	5.00	8.00	9.00	19.00
Full_Buy_WR	523.00	56.46	20.31	0.00	44.00	57.00	71.00	100.00

Tabela 4 – Estatísticas descritivas da base geral de jogadores

4.2.2 Atributos categóricos

Os gráficos 8 e 7 a seguir apontam as quantidades de escolhas de agentes e mapas entre as partidas respectivamente, indicando tendências de escolhas dos principais agentes e os mapas mais disputados. Vale ressaltar que o mapa Icebox foi adicionado ao jogo somente no mês de Outubro de 2020, assim como os agentes Skye e Yoru, adicionados ao jogo em Outubro de 2020 e Janeiro de 2021 respectivamente.

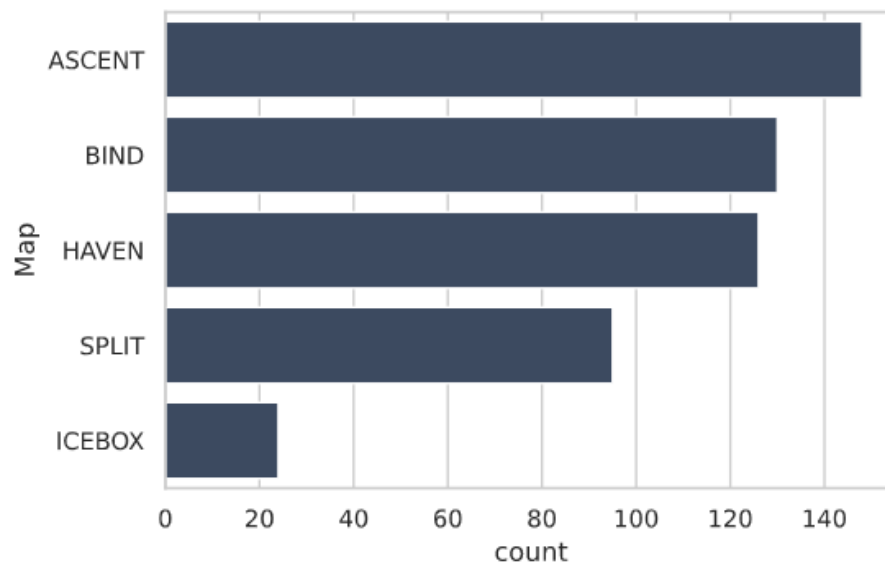


Figura 7 – Total de mapas jogados

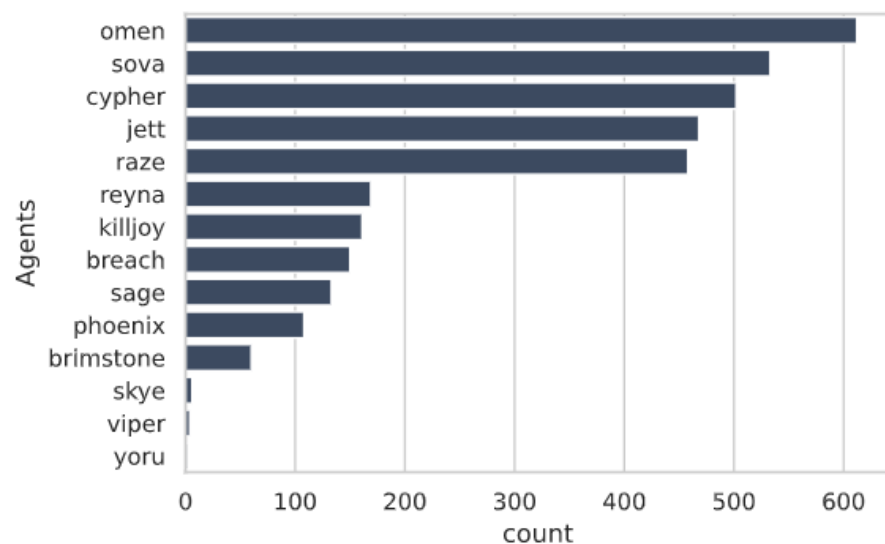


Figura 8 – Total de Agentes selecionados

4.2.3 Correlações Atributos

Os mapas de calor utilizam tanto do destaque visual como da visualização dos próprios resultados de correlação de *Pearson* e *Spearman* para indicar quais os atributos numéricos entre os escolhidos que melhor se relacionam.

A tabela 5 e figura 9 apresentam correlações lineares entre métricas de performance de jogadores e *win_rate* de seus times, sendo possível observar como as diversas métricas de performance dos jogadores impactam na porcentagem de rodadas vencidas em cada partida. A tabela 6 apresenta correlações não-lineares entre as mesmas métricas de performance de jogadores mas com o resultado final da partida - vitória ou derrota - apresentado em formato binário.

Através das tabelas apresentadas é possível observar melhor as correlações entre atributos relacionados ao próprio combate presente na dinâmica do jogo, onde times com mais abates, e principalmente aqueles que obtêm o primeiro abate em cada rodada - assegurando vantagem numérica - tendem a vencer mais rodadas e consequentemente mais partidas.

Atributos	ACS	ADR	KPR	FKPR	FDPR	MKPR	CPR	win_rate
ACS	1.00	0.90	0.92	0.67	-0.56	0.87	0.24	0.84
ADR	0.90	1.00	0.79	0.59	-0.47	0.75	0.20	0.71
KPR	0.92	0.79	1.00	0.67	-0.64	0.87	0.23	0.91
FKPR	0.67	0.59	0.67	1.00	-0.91	0.58	0.10	0.66
FDPR	-0.56	-0.47	-0.64	-0.91	1.00	-0.50	-0.09	-0.66
MKPR	0.87	0.75	0.87	0.58	-0.50	1.00	0.20	0.82
CPR	0.24	0.20	0.23	0.10	-0.09	0.20	1.00	0.26
win_rate	0.84	0.71	0.91	0.66	-0.66	0.82	0.26	1.00

Tabela 5 – Correlação Pearson Atributos Jogadores

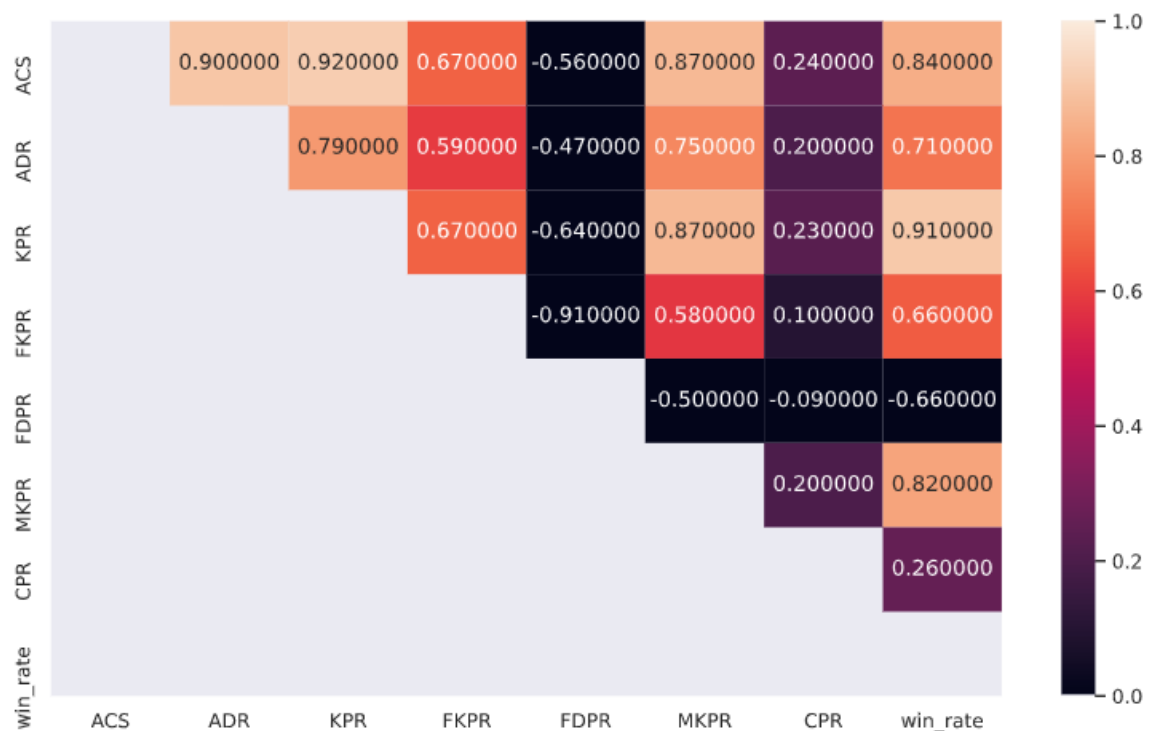


Figura 9 – Mapa de calor da correlação de Pearson entre atributos de Jogadores

Atributos	RESULT_BIN
ACS	0.77
ADR	0.65
KPR	0.78
FKPR	0.54
FDPR	-0.55
MKPR	0.71
CPR	0.18
RESULT_BIN	1.00

Tabela 6 – Correlação Spearman Atributos Jogadores

A tabela 7 e a figura 10 apresentam correlações lineares entre métricas de performance de times, principalmente percentagem de rodadas vencidas para cada lado no jogo (atacante ou defensor) e percentagem de rodadas vencidas para cada faixa de gasto de créditos do jogo com armas e habilidades.

Esses resultados permitem observar a importância de se vencer mais rodadas como atacante, uma vez que é dito que a maioria dos mapas e a própria dinâmica do jogo favorece o lado defensor, além de se vencer mais rodadas quando se possui mais créditos e consequentemente melhor poder de combate visto que rodadas onde os times dispõem de mais créditos são mais frequentes no decorrer da partida. Os gastos dos times se enquadram nas categorias *Eco* e *Semi-Eco* geralmente após perderem rodadas onde estavam completamente armados porém sem créditos reservas, necessitando de economizar esses créditos para melhores armas nas próximas rodadas.

A tabela 8 apresenta correlações não-lineares entre as mesmas métricas de performance de times mas com o resultado final da partida - vitória ou derrota - apresentado em formato binário.

Atributos	ct_wr	t_wr	PistolWR	EcoWR	Semi_EcoWR	Semi_BuyWR	Full_BuyWR	win_rate
ct_wr	1.00	0.21	0.36	0.43	0.01	0.42	0.64	0.75
t_wr	0.21	1.00	0.37	0.37	0.04	0.53	0.62	0.76
PistolWR	0.36	0.37	1.00	0.92	0.03	0.24	0.15	0.45
EcoWR	0.43	0.37	0.92	1.00	0.03	0.27	0.17	0.49
Semi_EcoWR	0.01	0.04	0.03	0.03	1.00	-0.04	-0.04	0.04
Semi_BuyWR	0.42	0.53	0.24	0.27	-0.04	1.00	0.33	0.61
Full_BuyWR	0.64	0.62	0.15	0.17	-0.04	0.33	1.00	0.86
win_rate	0.75	0.76	0.45	0.49	0.04	0.61	0.86	1.00

Tabela 7 – Correlação Pearson Atributos Times

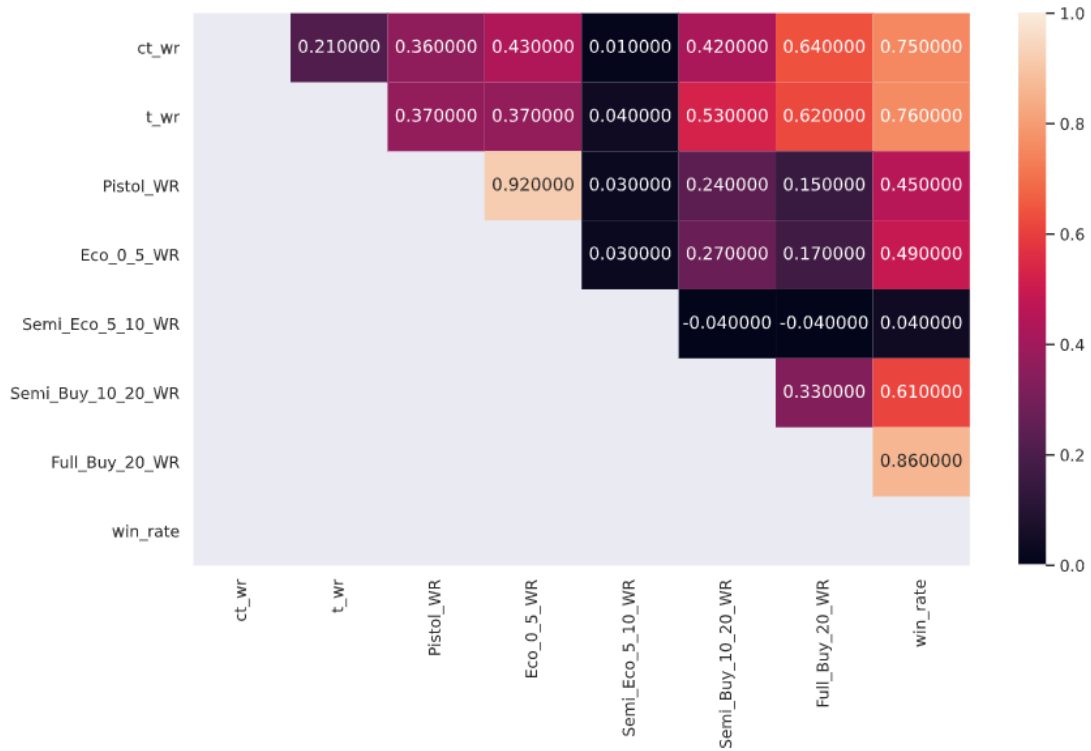


Figura 10 – Mapa de calor da correlação de Pearson entre atributos de Times

Atributos	RESULT_BIN
ct_wr	0.59
t_wr	0.63
PistolWR	0.30
EcoWR	0.33
Semi_EcoWR	0.01
Semi_BuyWR	0.50
Full_BuyWR	0.78
RESULT_BIN	1.00

Tabela 8 – Correlação Spearman Atributos Times

As tabelas 9 e 10 apresentam correlações de como o uso de certos agentes e a performance de classes de agentes impacta no resultado de partidas para cada mapa. Na primeira pode-se analisar como as escolhas de agentes (não há limitação na repetição de agentes de uma mesma classe) impacta na performance dos times para cada mapa, assim como tendências de escolhas de certos agentes para cada mapa, principalmente entre os times vencedores (correlações "mais positivas"). Valores nulos representam agentes nunca utilizados no respectivo mapa.

Na segunda tabela tem-se a relação entre o desempenho, em termos de ACS, entre as diferentes classes de agentes para cada mapa. Em um aspecto geral a performance de agentes duelistas gera mais impacto no desempenho do time, uma vez que esses são mais direcionados ao combate contra adversários enquanto demais classes desempenham

funções alternativas como ganho de informação e domínio de espaço nos mapas.

Jogadores podem gerar impacto em rodadas e jogos de diferentes maneiras em um FPS tático como Valorant™, neste trabalho o foco foi direcionado às métricas de combate, porém não limitam a performance e efetividade de um jogador.

Agente/Mapa	BIND	ASCENT	SPLIT	HAVEN	ICEBOX
jett	-0.140	-0.114	0.083	0.109	0.305
raze	0.132	0.024	0.060	-0.083	-0.020
phoenix	-0.150	-0.143	-0.295	-0.035	-0.069
reyna	0.294	0.091	-0.144	0.127	0.035
yoru	0.051	NaN	NaN	-0.004	NaN
omen	0.140	0.114	0.059	0.227	0.299
brimstone	-0.140	-0.161	-0.059	-0.227	0.023
viper	-0.106	NaN	-0.060	NaN	-0.526
cypher	0.231	0.129	0.148	0.071	0.086
sage	-0.197	0.054	-0.067	-0.040	0.102
killjoy	-0.238	-0.153	-0.044	-0.103	-0.410
breach	-0.029	0.021	0.105	-0.044	NaN
sova	0.100	0.066	0.028	-0.050	0.319
skye	NaN	NaN	0.073	-0.243	-0.335

Tabela 9 – Correlação de Pearson entre Agentes utilizados e rodadas vencidas em cada mapa.

classe/mapa	SPLIT	BIND	ASCENT	HAVEN	ICEBOX
duelist	0.571	0.484	0.44	0.482	0.421
controller	0.323	0.413	0.166	0.406	0.252
initiator	0.279	0.295	0.425	0.275	0.233
sentinel	0.349	0.379	0.289	0.344	0.235

Tabela 10 – Correlação Pearson entre ACS de cada classe com Win Rate

4.2.4 Testes de Hipóteses

As inferências sobre as distribuições são refutadas/validadas através dos testes estatísticos apoiados por gráficos de densidade na validação de normalidade nas distribuições e gráficos de dispersão em diferenças significativas entre amostras para melhores visualizações.

4.2.4.1 Testes de Normalidade - *Shapiro Wilk*

A tabela 11 e figura 11 apresentam os resultados da aplicação do teste estatístico de *Shapiro Wilk* para normalidade dos dados de atributos de jogadores e visualização de

distribuições (com valor médio na linha vertical vermelha) de alguns desses atributos respectivamente.

Através da aplicação do teste sobre os diversos atributos foi possível observar que este rejeitou a hipótese nula H_0 : Os dados da amostra assumem uma distribuição normal ou gaussiana, proporcionando um melhor direcionamento para a aplicação de testes não-paramétricos de diferenças entre amostras.

atributo	valor-teste	p-valor
ACS	0.984	0.000000
ADR	0.990	0.000000
HS%	0.969	0.000000
KPR	0.981	0.000000
DPR	0.989	0.000000
APR	0.966	0.000000
FKPR	0.914	0.000000
FDPR	0.935	0.000000
FKWR	0.934	0.000000
ECON	0.922	0.000000
MKPR	0.975	0.000000
CPR	0.621	0.000000
K	0.987	0.000000
D	0.983	0.000000
A	0.954	0.000000
KD_DIFF	0.989	0.000000
2K	0.944	0.000000
3K	0.782	0.000000
4K	0.407	0.000000
1v1	0.479	0.000000
1v2	0.340	0.000000
1v3	0.134	0.000000
1v4	0.038	0.000000
1v5	0.004	0.000000
total_mult_kill	0.962	0.000000
total_clutch	0.597	0.000000

Tabela 11 – Resultados Shapiro Wilk Atributos Jogadores

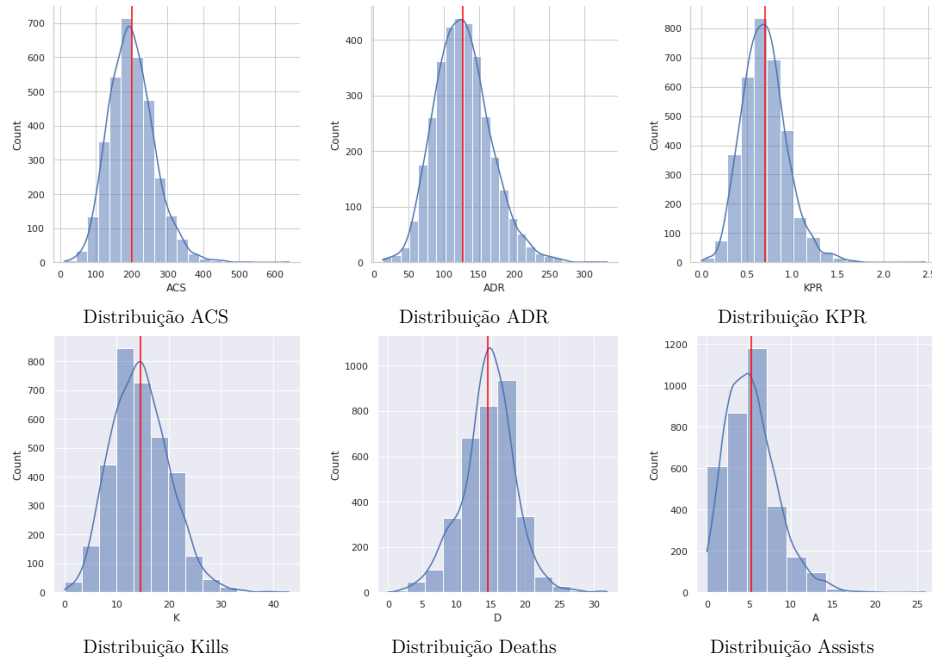


Figura 11 – Exemplo de gráficos de distribuição utilizados

Semelhante ao realizado sobre os dados de atributos de jogadores, a tabela 12 e figura 12 apresentam os resultados da aplicação do teste estatístico de *Shapiro Wilk* para normalidade dos dados de atributos de times e visualização de distribuições (com valor médio na linha vertical vermelha) de alguns desses atributos respectivamente.

Através da aplicação do teste sobre os diversos atributos foi possível observar que este rejeitou a hipótese nula H_0 : Os dados da amostra assumem uma distribuição normal ou gaussiana, proporcionando um melhor direcionamento para a aplicação de testes não-paramétricos de diferenças entre amostras. A não normalidade dos dados é melhor perceptível nos gráficos de densidade para os atributos dos times, com alguns valores únicos se destacando além dos demais.

atributo	valor-teste	p-valor
rounds_won	0.825	0.000000
rounds_lost	0.887	0.000000
ct_rounds_won	0.979	0.000001
ct_rounds_lost	0.975	0.000000
t_rounds_won	0.978	0.000000
t_rounds_lost	0.975	0.000000
total_rounds	0.946	0.000000
win_rate	0.992	0.004421
ct_wr	0.982	0.000005
t_wr	0.982	0.000004
Pistol_WR	0.807	0.000000
Eco_0_5_WR	0.862	0.000000

atributo	valor-teste	p-valor
Semi_Eco_5_10_WR	0.528	0.000000
Semi_Buy_10_20_WR	0.940	0.000000
Full_Buy_20_WR	0.989	0.000396

Tabela 12 – Resultados Shapiro Wilk Atributos Times

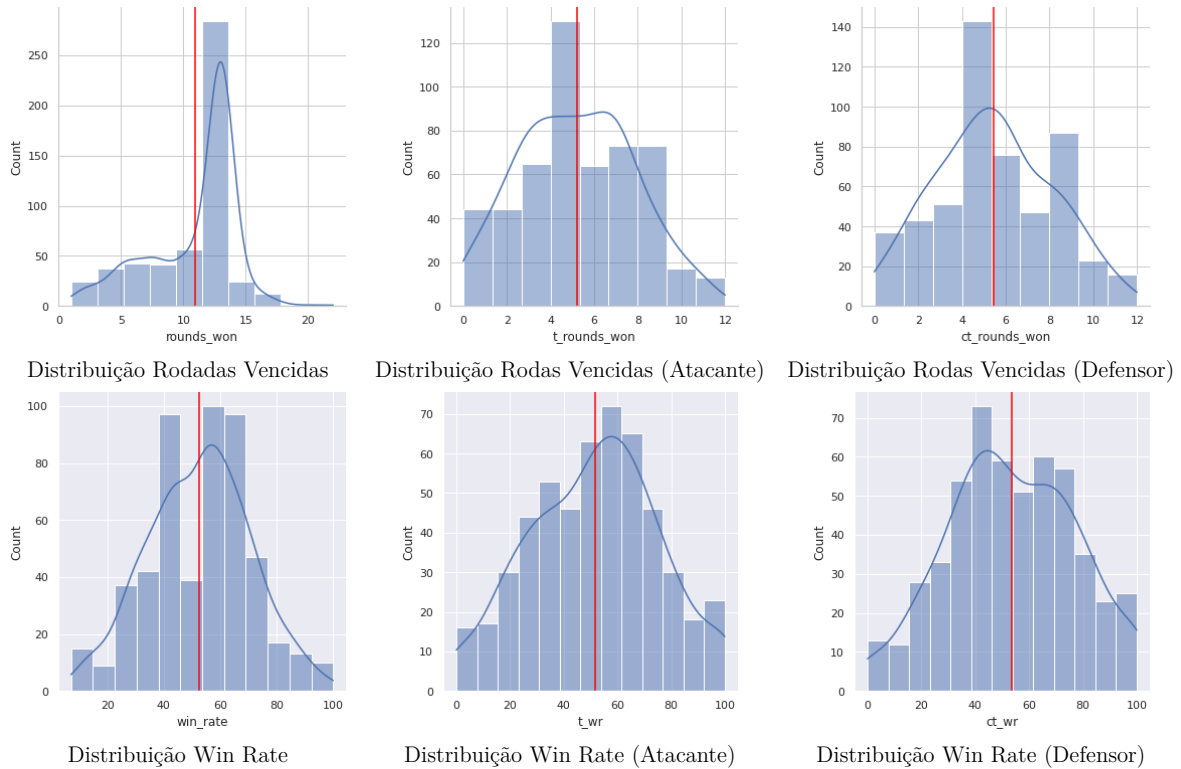


Figura 12 – Exemplo de gráficos de distribuição utilizados

4.2.4.2 Testes de Independência

A tabela 13 e as figuras 13 e 14 demonstram os resultados da utilização do teste estatístico *Kruskal Wallis* para o teste da hipótese nula H_0 : Não há uma diferença significativa entre os agentes de cada classe com relação à performances (ACS). Somente para a classe Controlador o teste não foi capaz de rejeitar H_0 , não sendo possível observar diferenças estatisticamente significativas entre as medianas das amostras. No caso das demais classes, a H_0 foi rejeitada pelo teste, assim tem-se a hipótese alternativa H_1 : Há pelo menos uma amostra que se difere das demais. A figura 13 exemplifica visualmente um caso de rejeição da hipótese nula de independência de performance entre agentes.

Com isso há indicativos de que os agentes em algumas classes se diferem quanto ao ACS e que a escolha de determinados agentes pode influenciar no desempenho do jogador e consequentemente no resultado da partida.

Classe	Valor-Teste	p-valor
Duelista	45.773	0.000000
Controlador	3.003	0.222822
Iniciador	32.329	0.000000
Sentinela	6.190	0.045279

Tabela 13 – Teste Kruskal Wallis de ACS entre Agentes de uma mesma classe

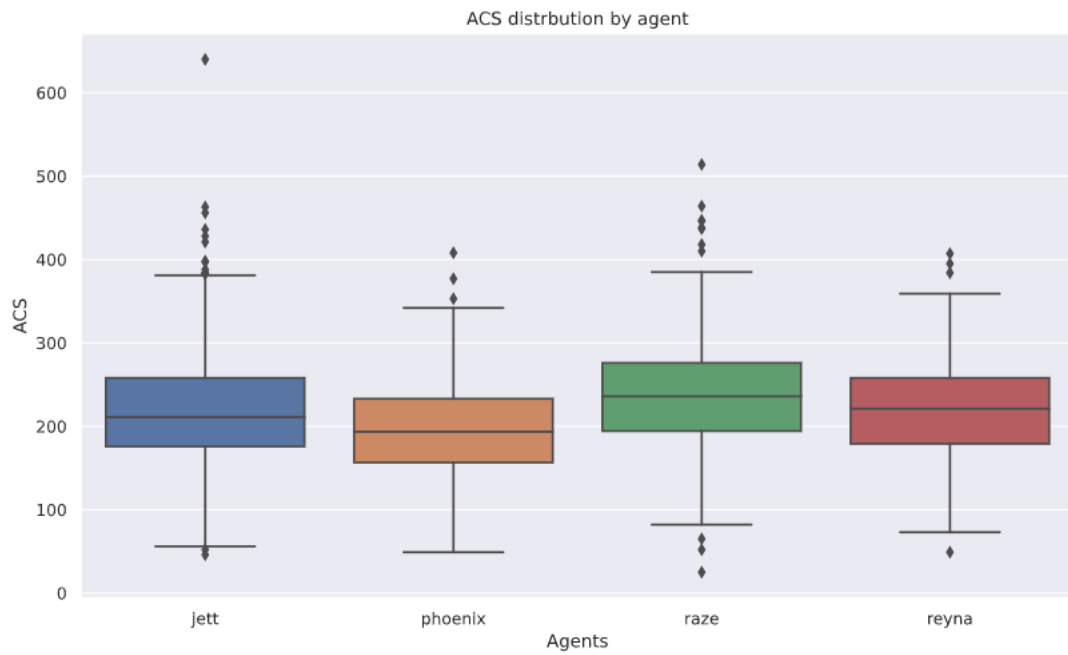


Figura 13 – Gráficos de Dispersão de ACS entre Agentes da classe Duelista

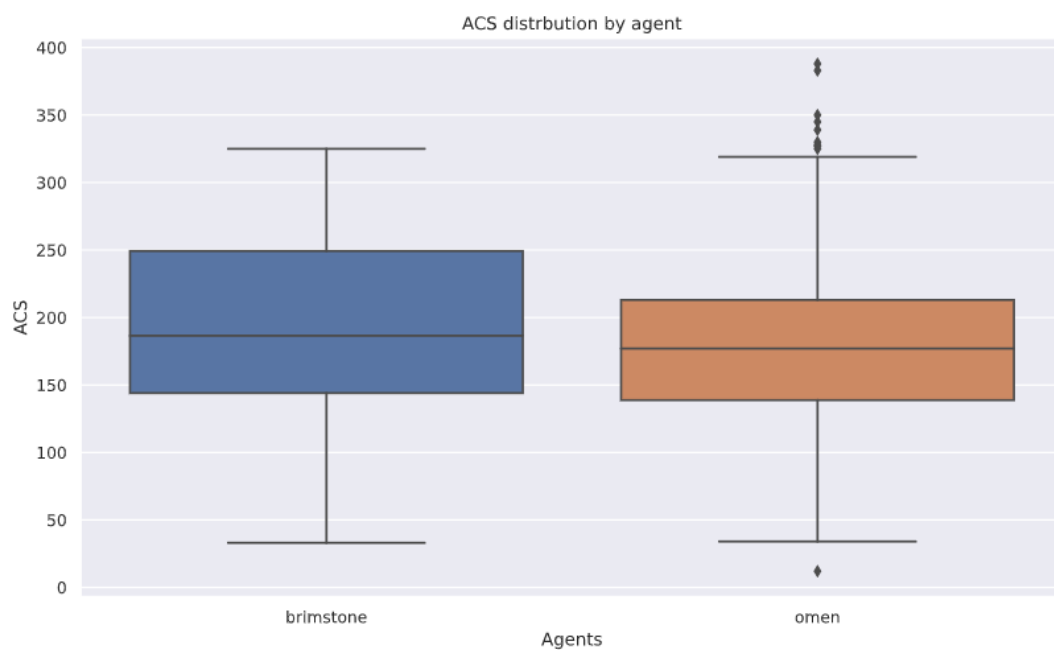


Figura 14 – Gráficos de Dispersão de ACS entre Agentes da classe Controlador

O teste de *Kruskal Wallis* também foi utilizado para avaliar as performances de cada agente para cada mapa do jogo - Ascent, Haven, Split, Bind e Icebox. A tabela 14 contém os resultados do teste para a hipótese nula H_0 : Não há uma diferença significativa entre as performances de cada agente entre os mapas. Foram considerados somente os agentes utilizados em pelo menos 4 mapas no jogo e dentre estes o teste foi capaz de rejeitar H_0 para os agentes Raze e Sage, surgindo assim a hipótese alternativa H_1 : Há pelo menos uma amostra que se difere das demais.

As figuras 15 e 16 ilustram melhor as distribuições de ACS para cada mapa dos agentes Raze e Sage, apoiado os resultados do teste.

Agente	Valor-Teste	p-valor
jett	4.239	0.3746617084
raze	10.766	0.0293210967
omen	1.442	0.8367846875
cypher	5.420	0.2468837484
breach	2.382	0.4970309239
killjoy	7.290	0.1213103428
sova	1.812	0.7702580997
reyna	5.347	0.2535491344
sage	16.799	0.0021146159
phoenix	5.681	0.2242344062
brimstone	5.251	0.2624910558

Tabela 14 – Teste Kruskall Wallis – ACS Agentes entre Mapas

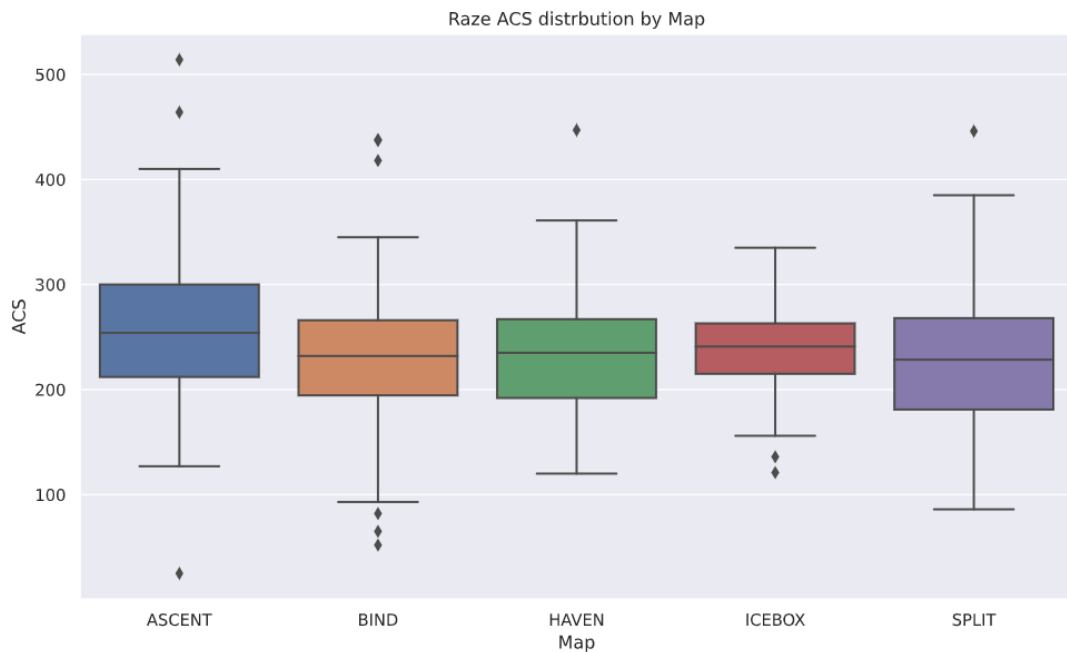


Figura 15 – Gráficos de Dispersão de ACS entre Mapas do agente Raze

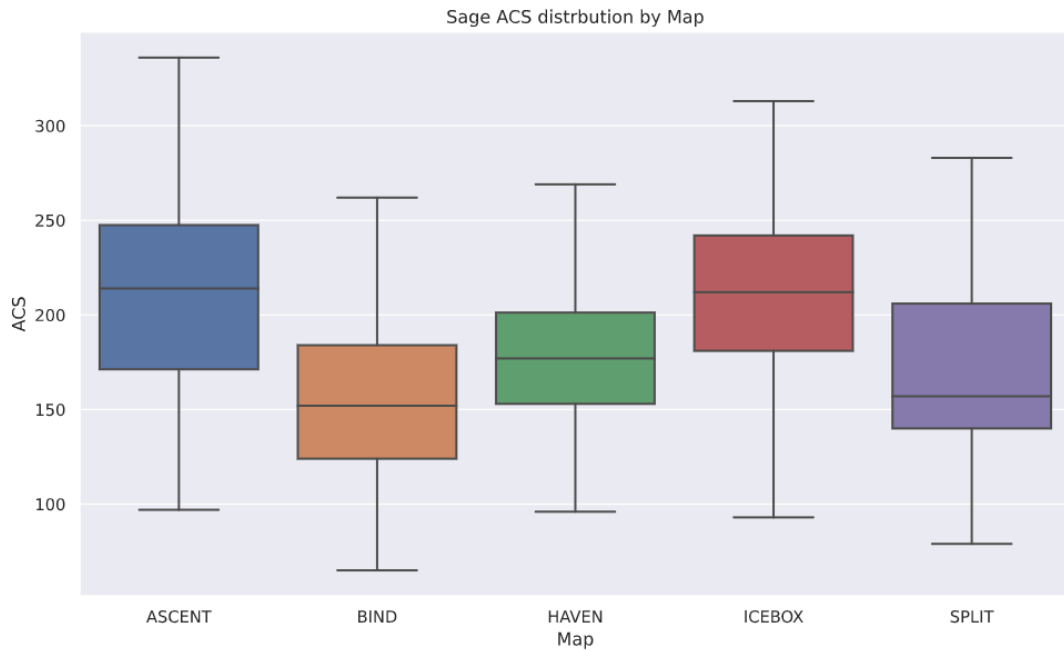


Figura 16 – Gráficos de Dispersão de ACS entre Mapas do agente Sage

Na validação de independência de dados de natureza discreta foi utilizado o teste $\chi^2_{contingency}$. A tabela 15 contém os resultados do teste sobre a hipótese nula H_0 : A escolha de um agente não influencia no resultado da partida. O teste compara as frequências de ocorrências de vitórias e derrotas para cada agente e determina se há uma diferença estatisticamente significativa entre as frequências.

Pelo teste tem-se que a escolha do agente Killjoy possui maior influência no resultado final da partida, onde somente neste caso o teste rejeitou a hipótese nula.

Agente	Valor-Teste	p-valor
jett	0.8189	0.36547
raze	0.6322	0.42654
phoenix	0.221093	0.63820
reyna	2.08664	0.14859
yoru	0.496738	0.48093
omen	0.033422	0.85494
brimstone	0.142027	0.70627
cypher	2.68971	0.10099
sage	0.74613	0.38770
killjoy	4.31288	0.03782
breach	0.3162	0.57384
sova	0.054622	0.81520
skye	1.49198	0.22190

Tabela 15 – Teste χ^2 de vitórias entre agentes utilizados.

5 Etapas e Cronograma

Semana	Tarefa
06/12	Estudo da estruturação da página web
13/12	Desenvolvimento do processo de extração
20/12	Extração e Análise Primária dos Dados
03/01	Preparação e Entrega do Pitch parcial
10/01	Estudo de materiais relacionados e de apoio
31/01	Aprimoramento do processamento de dados
21/02	Desenvolvimento das análises
07/03	Desenvolvimento das análises
14/03	Preparação Pitch e Texto final
21/03	Preparação Pitch e Texto final

Referências

- ANOVA Scipy. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html#scipy.stats.f_oneway>. Accessed: 2021-03-20.
- BEAUTIFUL Soup. <<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>>. Accessed: 2021-03-20.
- BIG-DATA at E-Sports. Acessado em 19/03/2021. Disponível em: <<https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/big-data-powered-esports.html>>.
- BRAGA, L. **Compreendendo Probabilidade e Estatística**. E-PAPERS. ISBN 9788576502821. Disponível em: <<https://books.google.com.br/books?id=jtUOfHuhme8C>>.
- CHI2 Contingency. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html>. Accessed: 2021-03-20.
- DATA-DRIVEN. Acessado em 19/03/2021. Disponível em: <<https://neilpatel.com/br/blog/data-driven/>>.
- DISTRIBUICAO Gaussian. <<http://www.portalaction.com.br/probabilidades/62-distribuicao-normal>>. Accessed: 2021-03-20.
- EARNINGS Valorant. Acessado em 19/03/2021. Disponível em: <<https://liquipedia.net/valorant/Portal:Statistics>>.
- KRUSKAL Wallis Scipy. <<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html#scipy.stats.kruskal>>. Accessed: 2021-03-20.
- MITCHELL, R. **Web Scraping with Python: Collecting More Data from the Modern Web**. O'Reilly Media, 2018. ISBN 9781491985526. Disponível em: <<https://books.google.com.br/books?id=TYtSDwAAQBAJ>>.
- MOBALYTICS GG. Acessado em 19/03/2021. Disponível em: <<https://mobalytics.gg/>>.
- PANDAS Documentation. <<https://pandas.pydata.org/pandas-docs/stable/index.html>>. Accessed: 2021-03-20.
- SCIPY Documentation. <<https://docs.scipy.org/doc/scipy/reference/index.html>>. Accessed: 2021-03-20.
- SEABORN Documentation. <<https://seaborn.pydata.org/index.html>>. Accessed: 2021-03-20.
- SHADOW GG. Acessado em 19/03/2021. Disponível em: <<https://shadow.gg>>.
- SHAPIRO Wilk Scipy. <<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html>>. Accessed: 2021-03-20.

TOWARDS Data Science - Spearman x Pearson. <<https://towardsdatascience.com/clearly-explained-pearson-v-s-spearman-correlation-coefficient-ada2f473b8>>. Accessed: 2021-03-20.

URLLIB Package. <<https://docs.python.org/3/howto/urllib2.html>>. Accessed: 2021-03-20.

VALENS hired by Evil Geniuses. Acessado em 19/03/2021. Disponível em: <<https://www.hltv.org/news/28332/valens-joins-evil-geniuses-as-head-of-data-science>>.

VALORANT. Acessado em 19/03/2021. Disponível em: <<https://playvalorant.com/pt-br/>>.

VLR.GG. Acessado em 19/03/2021. Disponível em: <<https://www.vlr.gg/>>.

APÊNDICE A – Descrição Atributos Bases Gerais

Tabela de descrição das colunas/atributos da base geral de dados de jogadores

Coluna	Descrição Coluna
Player	Nome do jogador
Agents	Agente utilizado pelo jogador
ACS	Average Combat Score - Pontuação Média de Combate
K	Kills - Abates do jogador na partida
D	Deaths - Mortes do jogador na partida
A	Assists - Assitências do jogador na partida
KD_DIFF	Diferença entre Abates e Mortes do jogador
ADR	Dano médio por rodada infringido pelo jogador
HS%	Porcentagem de Headshots (tiros na cabeça)
FK	First Kill - Total de primeiros abates pelo jogador
FD	First Deaths - Total de primeiras mortes pelo jogador
FK_FD_DIFF	Diferença entre primeiros abates e primeiras mortes
match_id	ID única da série de partidas
Team_x	Nome do time
Opp_Team_x	Nome do time oponente
rounds_won	Total de rodadas vencidas
rounds_lost	Total de rodadas perdidas
ct_rounds_won	Total de rodadas vencidas como defensor
ct_rounds_lost	Total de rodadas perdidas como defensor
t_rounds_won	Total de rodadas vencidas como atacante
t_rounds_lost	Total de rodadas perdidas como atacante
Map	Nome do mapa
Patch	Versão atualização do jogo
total_rounds	Total de rodadas disputadas
win_rate	Porcentagem de rodadas vencidas
KPR	Kills per round - Abates por rodada
DPR	Deaths per round - Mortes por rodada
APR	Assists per round - Assitências por rodada
FKPR	First Kills per round - Primeiros abates por rodada
FKWR	First Kills Win Rate - Porcentagem de vitórias de primeiros abates
FDPR	First Deaths per round - Primeiras mortes por rodada

Coluna	Descrição Coluna
ct_wr	Percentagem de rodadas vencidas do lado defensor
t_wr	Percentagem de rodadas vencidas do lado atacante
RESULT	Resultado da partida - Vitória ou Derrota
2K	Total de rodadas com abates duplos
3K	Total de rodadas com abates triplos
4K	Total de rodadas com abates quadruplos
5K	Total de rodadas com abates quintuplos
1v1	Total de duelos 1v1 ganhos
1v2	Total de duelos 1v2 ganhos
1v3	Total de duelos 1v3 ganhos
1v4	Total de duelos 1v4 ganhos
1v5	Total de duelos 1v5 ganhos
ECON	Pontuação economica - Pontuação por créditos gastos
PL	Total de Spikes ativadas
DE	Total de Spikes desativadas
total_mult_kill	Total de abates multiplos na partida
total_clutch	Total de duelos 1vX ganhos na partida
MKPR	Abates multiplos por rodada
CPR	Duelos 1vX ganhos por rodada
breach	Escolha do agente Breach - binário
brimstone	Escolha do agente Brimmstone - binário
cypher	Escolha do agente Cypher - binário
jett	Escolha do agente Jett - binário
killjoy	Escolha do agente Killjoy - binário
omen	Escolha do agente Omen - binário
phoenix	Escolha do agente Phoenix - binário
raze	Escolha do agente Raze - binário
reyna	Escolha do agente Reyna - binário
sage	Escolha do agente Sage - binário
skye	Escolha do agente Skye - binário
sova	Escolha do agente Sova - binário
viper	Escolha do agente Vipe - binário
yoru	Escolha do agente Yoru - binário
RESULT_BIN	Identificador binário de vitória

Tabela 17 – Descrição colunas Base Geral Jogadores

Tabela de descrição das colunas/atributos da base geral de dados de times. Maioria dos resultados são agregações de valores médios e somas da base geral de jogadores.

Coluna	Descrição da coluna
Team	Nome do time
match_id	ID único da serie de partidas
Map	Nome do mapa
ACS	ACS médio dos jogadores/time
K	K médio dos jogadores/time
D	D médio dos jogadores/time
A	A médio dos jogadores/time
KD_DIFF	KD_DIFF médio dos jogadores /time
ADR	ADR dos jogadores/time
HS%	HS% médio dos jogadores/time
ECON	ECON médio dos jogadores/time
FK	FK total dos jogadores/time
FD	FD total dos jogadores/time
KPR	KPR total dos jogadores/time
DPR	DPR total dos jogadores/time
APR	APR total dos jogadores/time
FKPR	FKPR total dos jogadores/time
FDPR	FDPR total dos jogadores/time
2K	2K total dos jogadores/time
3K	3K total dos jogadores/time
4K	4K total dos jogadores/time
5K	5K total dos jogadores/time
1v1	1v1 total dos jogadores/time
1v2	1v2 total dos jogadores/time
1v3	1v3 total dos jogadores/time
1v4	1v4 total dos jogadores/time
1v5	1v5 total dos jogadores/time
PL	PL total dos jogadores/time
DE	DE total dos jogadores/time
total_mult_kill	total_mult_kill total dos jogadores/time
MKPR	MKPR total dos jogadores/times
total_clutch	total_clutch dos jogadores/times
CPR	CPR total dos jogadores/time
breach	Escolha do agente Breach - binário
brimstone	Escolha do agente Brimmstone - binário
cypher	Escolha do agente Cypher - binário

Coluna	Descrição da coluna
jett	Escolha do agente Jett - binário
killjoy	Escolha do agente Killjoy - binário
omen	Escolha do agente Omen - binário
phoenix	Escolha do agente Phoenix - binário
raze	Escolha do agente Raze - binário
reyna	Escolha do agente Reyna - binário
sage	Escolha do agente Sage - binário
skye	Escolha do agente Skye - binário
sova	Escolha do agente Sova - binário
viper	Escolha do agente Vipe - binário
yoru	Escolha do agente Yoru - binário
rounds_won	Total de rodadas vencidas
rounds_lost	Total de rodadas perdidas
ct_rounds_won	Total de rodadas vencidas como defensor
ct_rounds_lost	Total de rodadas perdidas como defensor
t_rounds_won	Total de rodadas vencidas como atacante
t_rounds_lost	Total de rodadas perdidas como atacante
total_rounds	Total de rodadas disputadas
win_rate	Porcentagem de rodadas vencidas
ct_wr	Porcentagem de rodadas vencidas do lado defensor
t_wr	Porcentagem de rodadas vencidas do lado atacante
RESULT	Resultado da partida - Vitória ou Derrota
Pistol_W	Rodadas Pistol Vencidas
Pistol_P	Rodadas Pistol jogadas
Pistol_WR	Porcentagem de vitórias de rodadas pistol
Eco_0_5_P	Rodadas Eco jogadas
Eco_0_5_W	Rodadas Eco Vencidas
Eco_0_5_WR	Porcentagem de vitórias Rodadas Eco
Semi_Eco_5_10_P	Rodadas Semi-Eco jogadas
Semi_Eco_5_10_W	Rodadas Semi-Eco Vencidas
Semi_Eco_5_10_WR	Porcentagem de vitórias rodadas Semi-Eco
Semi_Buy_10_20_P	Rodadas Semi-Buy jogadas
Semi_Buy_10_20_W	Rodadas Semi-Buy Vencidas
Semi_Buy_10_20_WR	Porcentagem de vitórias rodadas Semi-Buy
Full_Buy_20_P	Rodadas Full-Buy jogadas
Full_Buy_20_W	Rodadas Full-Buy Vencidas
Full_Buy_20_WR	Porcentagem de vitórias rodadas Full-Buy
Opp_team	Nome do time oponente

Coluna	Descrição da coluna
Patch	Versão atualização do jogo
RESULT_BIN	Identificador binário de vitória

Tabela 18 – Descrição colunas Base Geral Jogadores