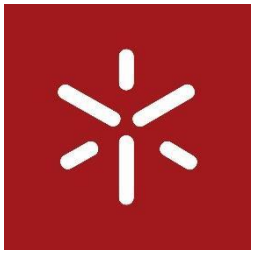


Universidade do Minho
Escola de Engenharia
Campus de Azurém

Engenharia e Gestão de Sistemas de Informação

(1º SEMESTRE / 3º ANO)

NBA

Relatório Final do Projeto Prático

Grupo 1 – Turno PL4

Engenharia de Dados para Suporte à Tomada de Decisão

Prof. João Carlos Gomes Rebelo

Ricardo Silva (a92948)

Duarte Brandão (a92938)

João Lemos (a92939)

Nuno Moreira (a92947)

Pedro Gonçalves (a92930)

Paulo Barros (a92929)

Tiago Lopes (a94123)

Índice:

Introdução:

No âmbito da UC de Engenharia de Dados para Suporte à Tomada de Decisão, o grupo decidiu, para o seu projeto prático, escolher como tema a “NBA” e irá usar este assunto e desporto como objeto de estudo para atingir os objetivos finais da UC e cumprir com os resultados de aprendizagem com sucesso. De seguida, consta-se o relatório final de Avaliação com a divisão em 3 subgrupos de trabalho.

1. SUBGRUPO 1

Use Case: A influência dos triplos na NBA

Ricardo Silva

A92948 - 3º ano LEGSI

a92948@alunos.uminho.pt

Paulo Barros

A92929 - 3º ano LEGSI

a92929@alunos.uminho.pt

1.1 Introdução

A partir do nosso Use Case (A influência dos triplos na NBA) extrapolámos 2 questões principais que usamos nas nossas 2 dashboards.

Vamos abordar principalmente a influência dos jogadores e dos triplos realizados por estes nas estatísticas da equipa e de seguida vamos abordar a influência dos triplos num panorama mais geral nas estatísticas e dados da equipa.

1.2 KPI's

Para este caso de uso o KPI será o seguinte:

- Número de Triplos

1.3 Analytical Questions

Qual a influência dos jogadores com mais triplos na classificação final da equipa na temporada?

Qual a influência da posição no número de triplos?

Qual a influência do número de triplos na classificação da equipa na respetiva conferência?

Qual a classificação final da equipa do jogador com mais triplos em dado ano?

Qual a influência dos triplos de dada equipa no número de títulos de campeão?

Qual a influência do número de triplos de um jogador no contributo do mesmo nas vitórias da equipa?

1.4 DataSets

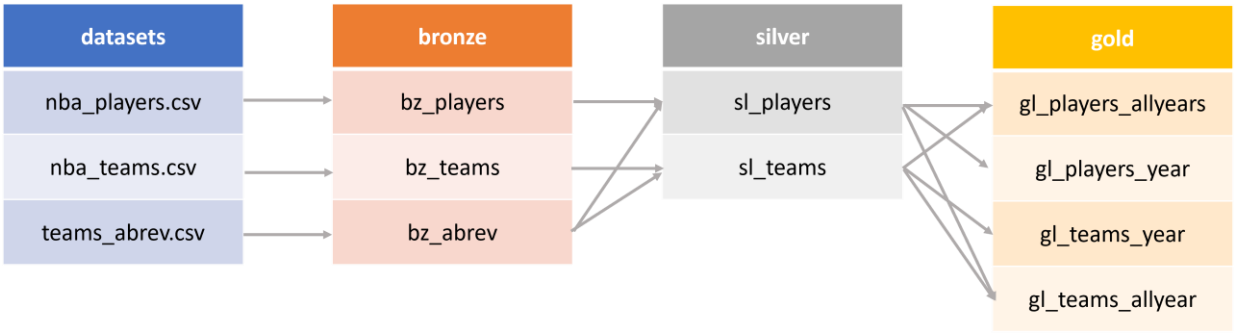
O grupo irá utilizar 2 datasets principais, sendo um deles referente àsstats de todos os jogadores em todas as épocas da NBA e o outro referente às classificações de todas as equipas em todas as épocas da história da NBA.

DataSet dos Jogadores: https://www.kaggle.com/drgilermo/nba-players-stats?select=Seasons_Stats.csv

DataSet das Equipas: https://www.kaggle.com/boonpalipatana/nba-season-records-from- every-year

Temos ainda um dataset auxiliar feito pelo grupo onde associamos o nome da equipa à sua abreviatura.

1.5 DataSet to Bronze to Silver to Gold



1.6 Data Quality

1.6.1 Análise ao DataSet dos Jogadores

A partir dos dados obtidos na análise podemos afirmar que o dataset é bastante sólido e que contêm a presença de poucos erros ou lacunas no mesmo. O grupo achou por bem referir alguns dos comentários a possíveis erros encontrados durante a análise, sendo estes os seguintes:

- Dos 70 valores distintos de abreviaturas do nome das equipas, 30 é o número de equipas atuais e os outros 40 valores correspondem a equipas extintas e a equipas que alteraram o nome uma ou mais vezes ao longo do tempo, pelo que os dados fazem sentido.
- Através de uma análise "Summary" a certos aspetos do dataset conseguimos confirmar que o período temporal corresponde ao que pretende ser analisado pelo grupo (1950-2017).
- Na coluna dos triplos marcados por cada jogador apareceu uma anomalia de 5764 linhas vazias, ou seja 23% dos dados. O grupo encara estes 23% de linhas sem valores como um sinal de validez e de confiança de dados do dataset pois o sistema de 3 pontos nem sempre existiu na NBA, sendo que estas linhas corresponderam a esse período temporal.

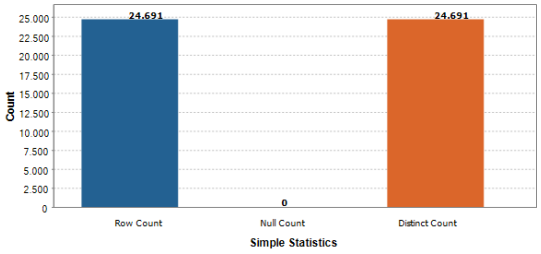
Quality Analysis:

- Row Count
- Null Count
- Distinct Count
- Minimum Range
- Maximum Range

Column: metadata.Column0

Simple Statistics

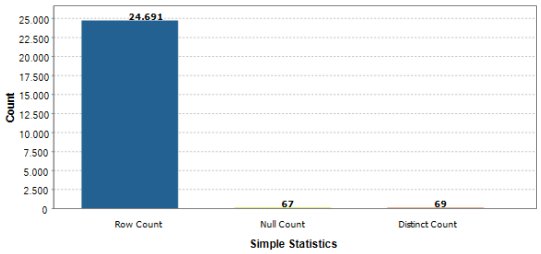
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	24691	100.00%



Column: metadata.Year

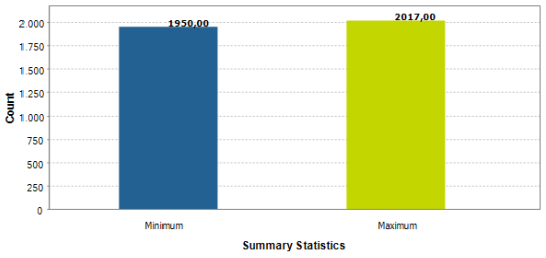
Simple Statistics

Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	69	0.28%



Summary Statistics

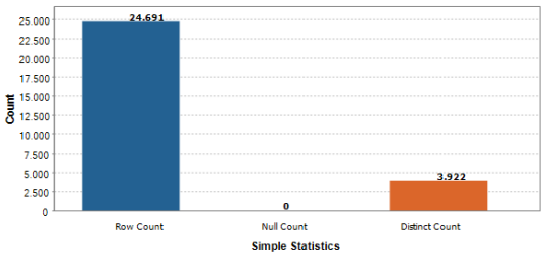
Label	Value
Range	67.0
Minimum	1950
Maximum	2017



Column: metadata.Player

Simple Statistics

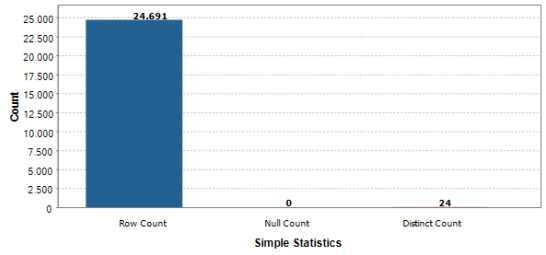
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	3922	15.88%



Column: metadata.Pos

Simple Statistics

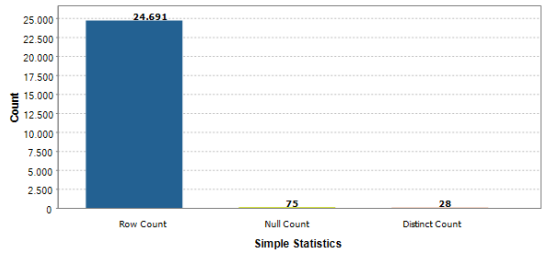
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	24	0.10%



Column: metadata.Age

Simple Statistics

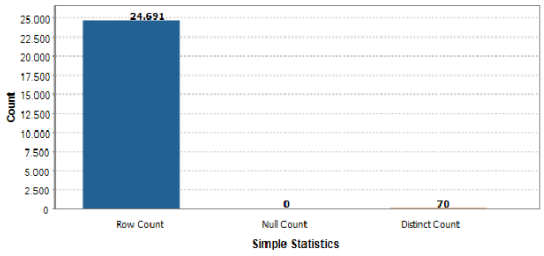
Label	Count	%
Row Count	24691	100.00%
Null Count	75	0.30%
Distinct Count	28	0.11%



Column: metadata.Tm

Simple Statistics

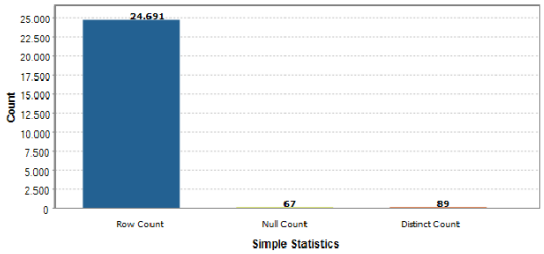
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	70	0.28%



Column: metadata.G

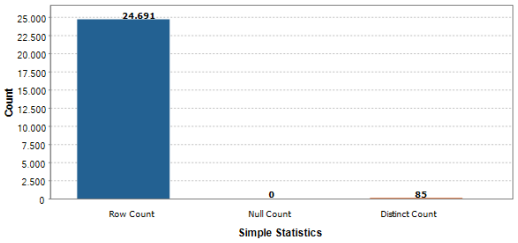
Simple Statistics

Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	89	0.36%



Simple Statistics

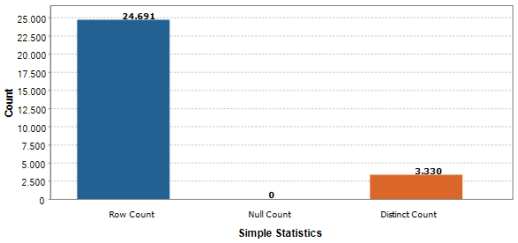
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	85	0.34%



Column: metadata.MP

Simple Statistics

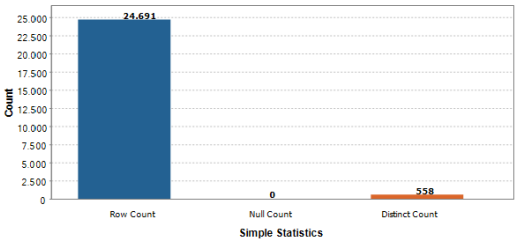
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	3330	13.49%



Column: metadata.PER

Simple Statistics

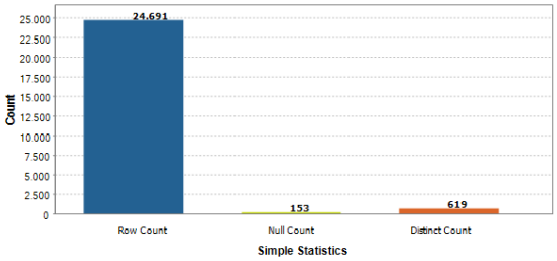
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	558	2.26%



Column: metadata.TS_

Simple Statistics

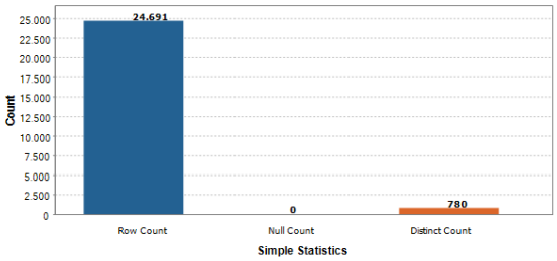
Label	Count	%
Row Count	24691	100.00%
Null Count	153	0.62%
Distinct Count	619	2.51%



Column: metadata._PAr

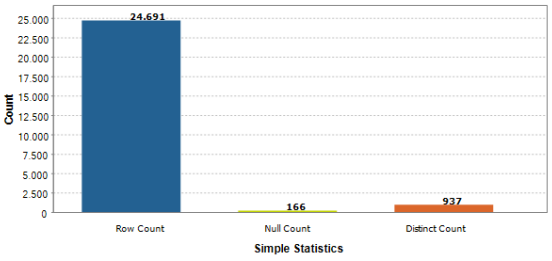
Simple Statistics

Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	780	3.16%



Simple Statistics

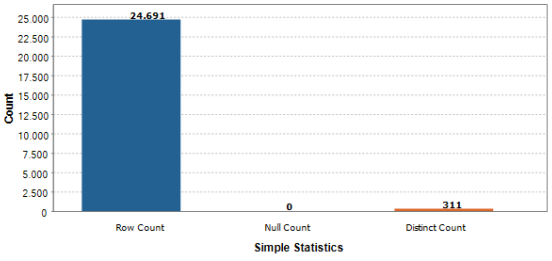
Label	Count	%
Row Count	24691	100.00%
Null Count	166	0.67%
Distinct Count	937	3.79%



Column: metadata.ORB_

Simple Statistics

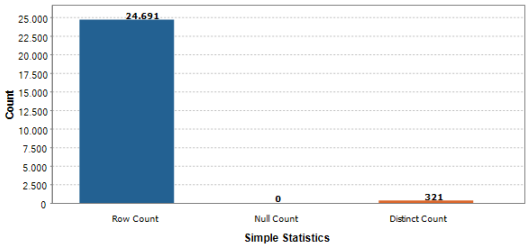
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	311	1.26%



Column: metadata.TRB_

Simple Statistics

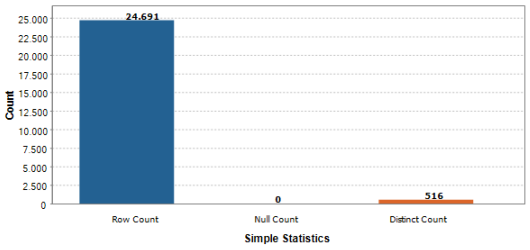
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	321	1.30%



Column: metadata.AST_

Simple Statistics

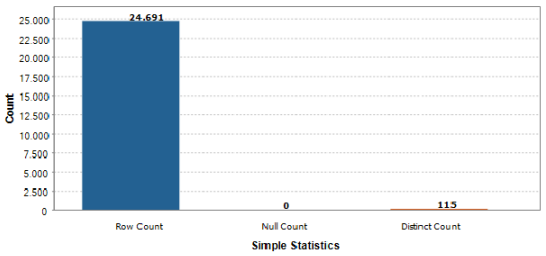
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	516	2.09%



Column: metadata.STL_

Simple Statistics

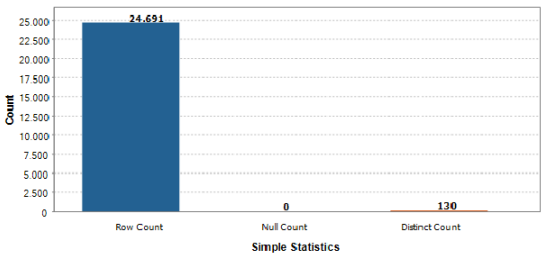
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	115	0.47%



Column: metadata.BLK_

Simple Statistics

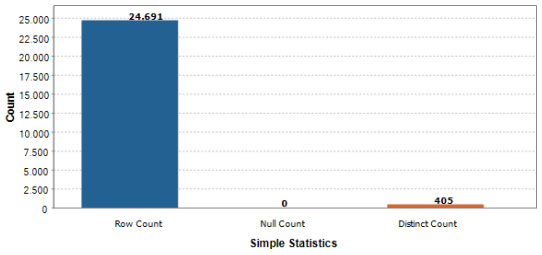
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	130	0.53%



Column: metadata.USG_

Simple Statistics

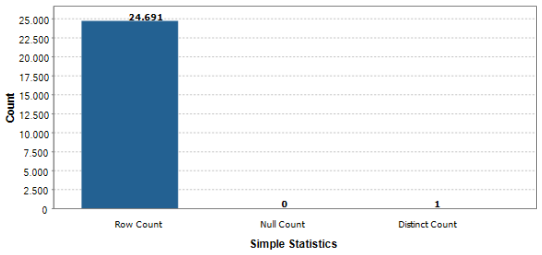
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	405	1.64%



Column: metadata.blanl

Simple Statistics

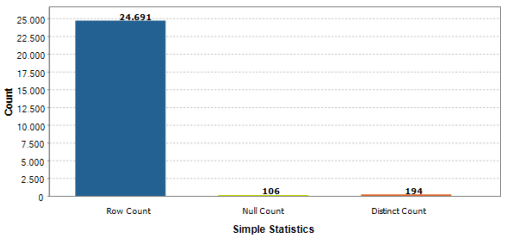
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	1	4.05E-3%



Column: metadata.OWS

Simple Statistics

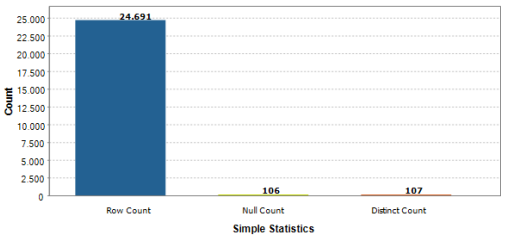
Label	Count	%
Row Count	24691	100.00%
Null Count	106	0.43%
Distinct Count	194	0.79%



Column: metadata.DWS

Simple Statistics

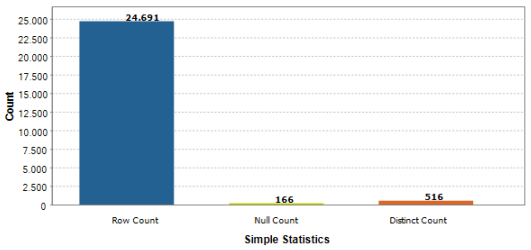
Label	Count	%
Row Count	24691	100.00%
Null Count	106	0.43%
Distinct Count	107	0.43%



Column: metadata.FG_

Simple Statistics

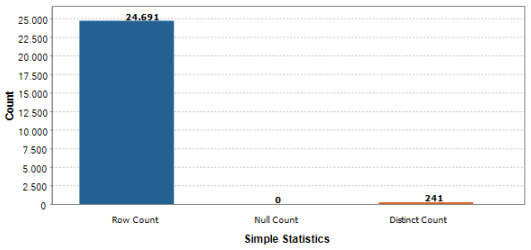
Label	Count	%
Row Count	24691	100.00%
Null Count	166	0.67%
Distinct Count	516	2.09%



Column: metadata._P

Simple Statistics

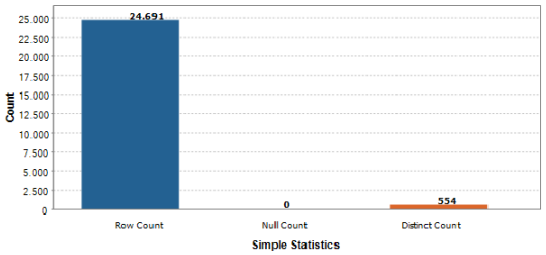
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	241	0.98%



Column: metadata._PA

Simple Statistics

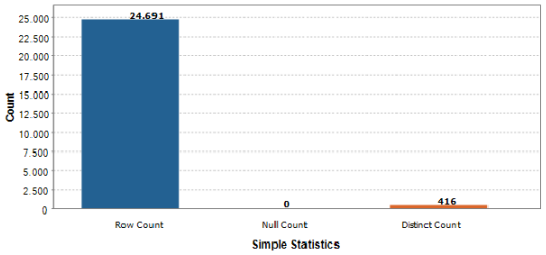
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	554	2.24%



Column: metadata.Column36

Simple Statistics

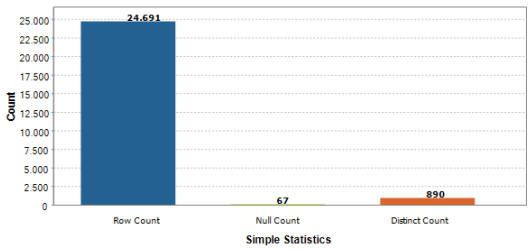
Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	416	1.68%



Column: metadata._P1

Simple Statistics

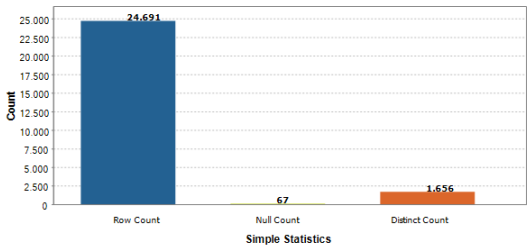
Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	890	3.60%



Column: metadata._PA1

Simple Statistics

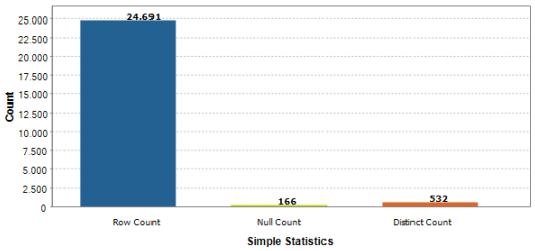
Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	1656	6.71%



Column: metadata.eFG_

Simple Statistics

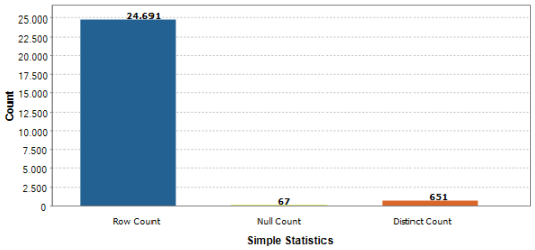
Label	Count	%
Row Count	24691	100.00%
Null Count	166	0.67%
Distinct Count	532	2.15%



Column: metadata.FT

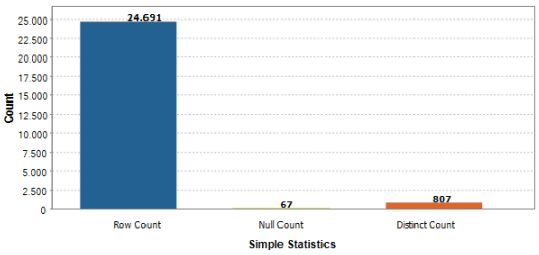
Simple Statistics

Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	651	2.64%



Simple Statistics

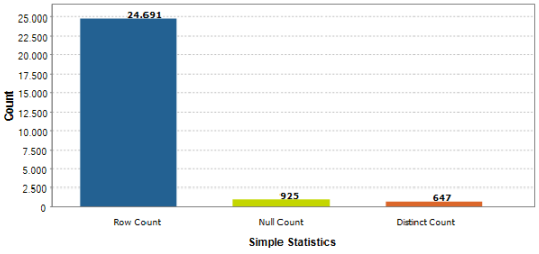
Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	807	3.27%



Column: metadata.FT_

Simple Statistics

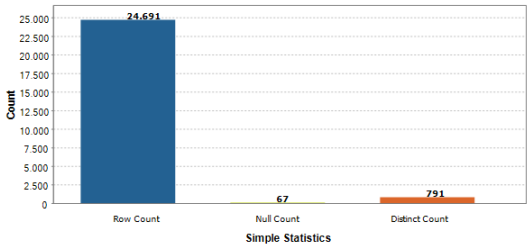
Label	Count	%
Row Count	24691	100.00%
Null Count	925	3.75%
Distinct Count	647	2.62%



Column: metadata.AST

Simple Statistics

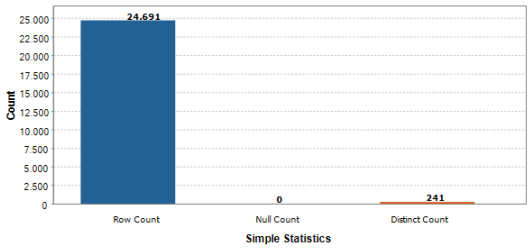
Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	791	3.20%



Column: metadata.STL

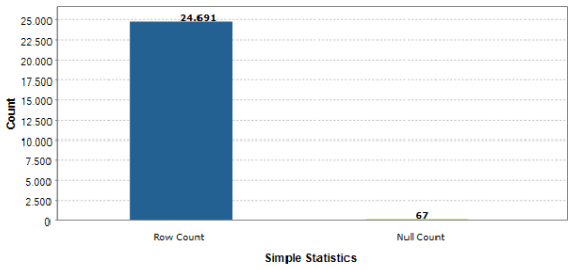
Simple Statistics

Label	Count	%
Row Count	24691	100.00%
Null Count	0	0.00%
Distinct Count	241	0.98%



Simple Statistics

Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%



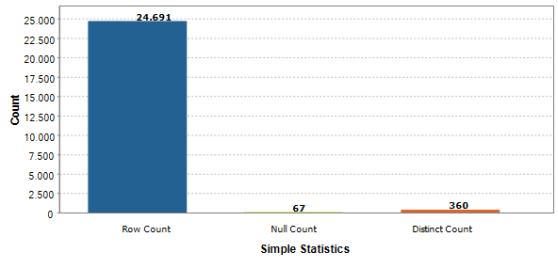
Summary Statistics

Label	Value
Mean	1992.5949886289798
Median	1996.0
Inter Quartile Range	26.0
Lower Quartile	1981.0
Upper Quartile	2007.0
Range	67.0
Minimum	1950
Maximum	2017

Column: metadata.PF

Simple Statistics

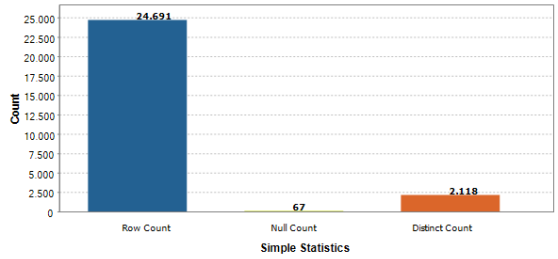
Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	360	1.46%



Column: metadata.PTS

Simple Statistics

Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	2118	8.58%



1.6.2 Análise ao DataSet das Equipas

A partir dos dados obtidos na análise podemos afirmar que o dataset é bastante sólido e que contêm a presença de poucos erros ou lacunas no mesmo. O grupo achou por bem referir alguns dos

comentários a possíveis erros encontrados durante a análise, sendo estes os seguintes:

- O número de épocas/anos desportivos presentes neste dataset contabiliza um total de 72 épocas. Destas retiramos 4 que correspondem a épocas não oficiais da NBA (do ano 46 até ano 50) e retiramos ainda a época de 2018 sendo que o outro dataset que temos só atinge até à época de 2017. Ficamos assim com um total de 67 épocas que é igual ao número de épocas presente no outro dataset.
- Em todas as linhas, a coluna correspondente à posição final da equipa está preenchida o que é essencial para o trabalho dado que a classificação final da equipa é um ponto crucial.
- O número de duplicados existentes em várias colunas são desprezáveis pois correspondem a estatísticas que várias equipas podem ter ao mesmo tempo, por exemplo o número de jogos, número de vitórias, etc.

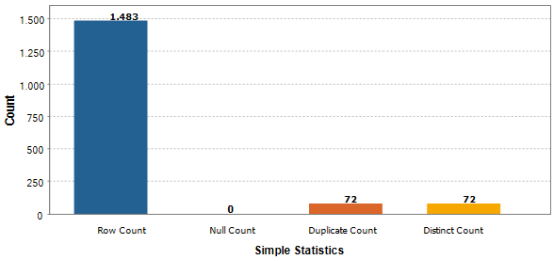
Quality Analysis:

- Row Count
- Null Count
- Distinct Count
- Duplicate Count
- Blank Count

Column: metadata._Season

Simple Statistics

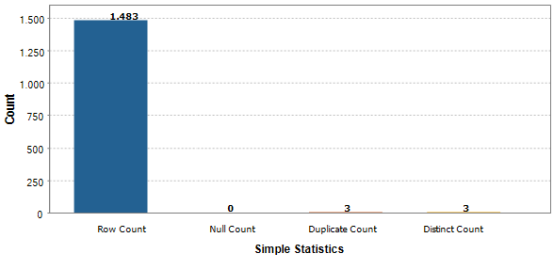
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	72	4.86%
Distinct Count	72	4.86%



Column: metadata.Lg

Simple Statistics

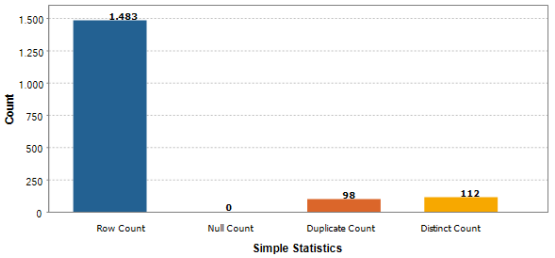
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	3	0.20%
Distinct Count	3	0.20%



Column: metadata.Team

Simple Statistics

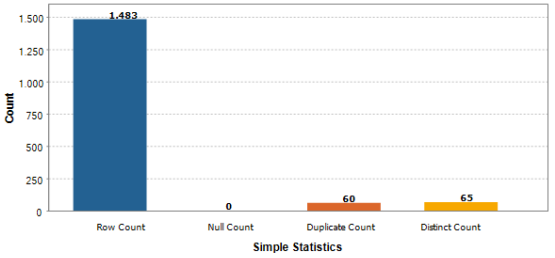
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	98	6.61%
Distinct Count	112	7.55%



Column: metadata.W

Simple Statistics

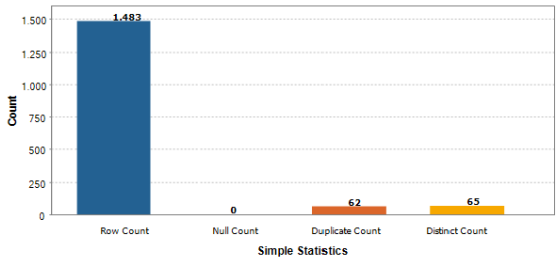
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	60	4.05%
Distinct Count	65	4.38%



Column: metadata.L

Simple Statistics

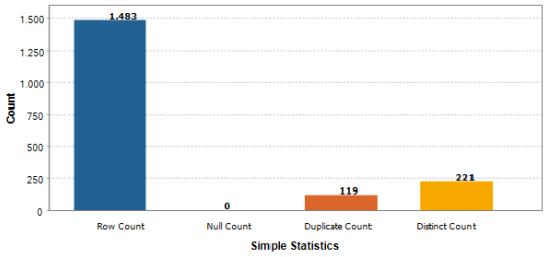
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	62	4.18%
Distinct Count	65	4.38%



Column: metadata.W_L_

Simple Statistics

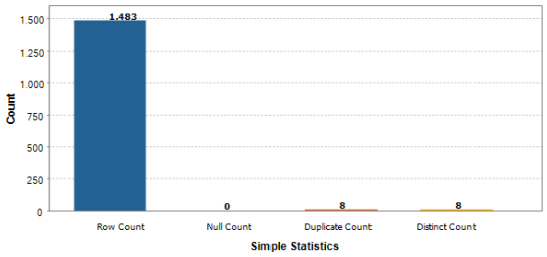
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	119	8.02%
Distinct Count	221	14.90%



Column: metadata.Finish

Simple Statistics

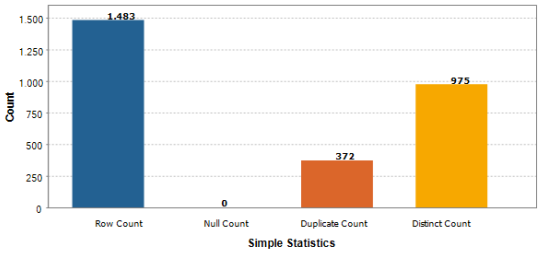
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	8	0.54%
Distinct Count	8	0.54%



Column: metadata.SRS

Simple Statistics

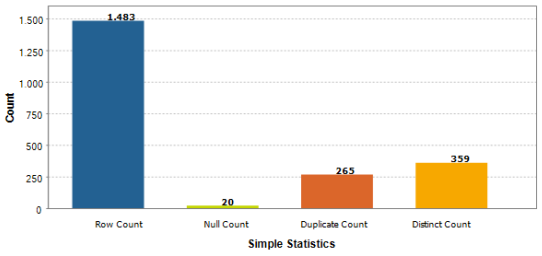
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	372	25.08%
Distinct Count	975	65.75%



Column: metadata.Pace

Simple Statistics

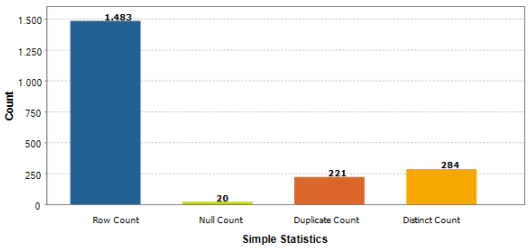
Label	Count	%
Row Count	1483	100.00%
Null Count	20	1.35%
Duplicate Count	265	17.87%
Distinct Count	359	24.21%



Column: metadata.ORtg

Simple Statistics

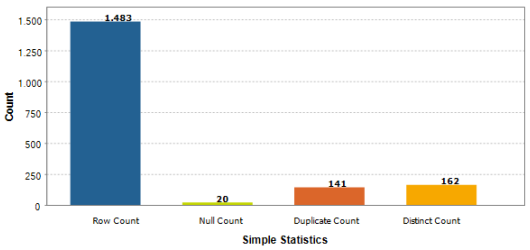
Label	Count	%
Row Count	1483	100.00%
Null Count	20	1.35%
Duplicate Count	221	14.90%
Distinct Count	284	19.15%



Column: metadata.Rel_ORtg

Simple Statistics

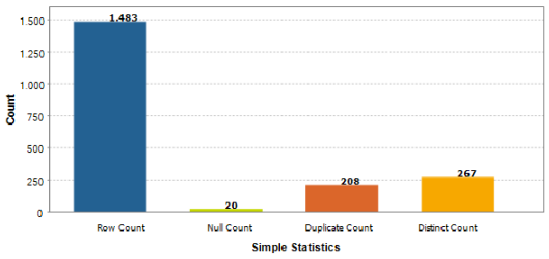
Label	Count	%
Row Count	1483	100.00%
Null Count	20	1.35%
Duplicate Count	141	9.51%
Distinct Count	162	10.92%



Column: metadata.DRtg

Simple Statistics

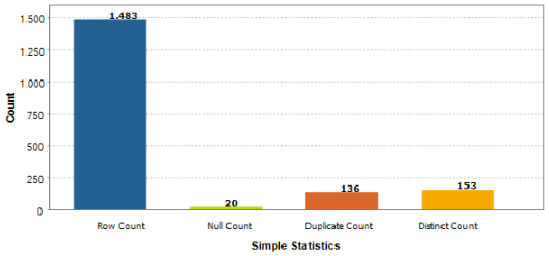
Label	Count	%
Row Count	1483	100.00%
Null Count	20	1.35%
Duplicate Count	208	14.03%
Distinct Count	267	18.00%



Column: metadata.Rel_DRtg

Simple Statistics

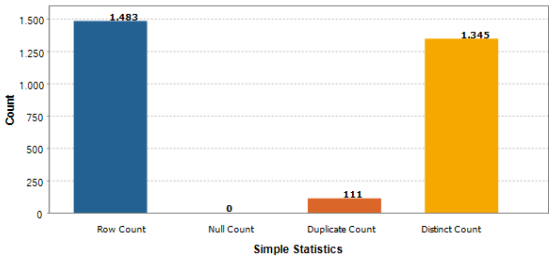
Label	Count	%
Row Count	1483	100.00%
Null Count	20	1.35%
Duplicate Count	136	9.17%
Distinct Count	153	10.32%



Column: metadata.Coaches

Simple Statistics

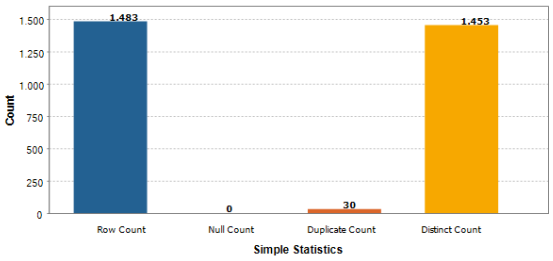
Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	111	7.48%
Distinct Count	1345	90.69%



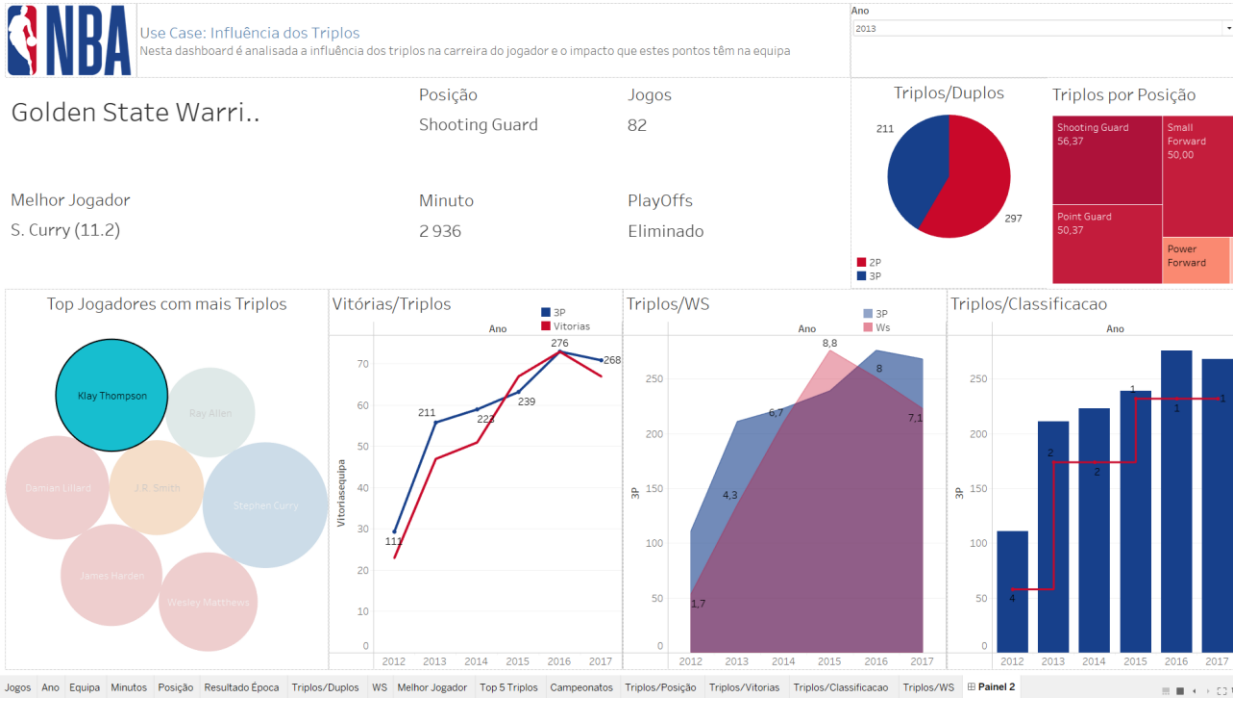
Column: metadata.Top_WS

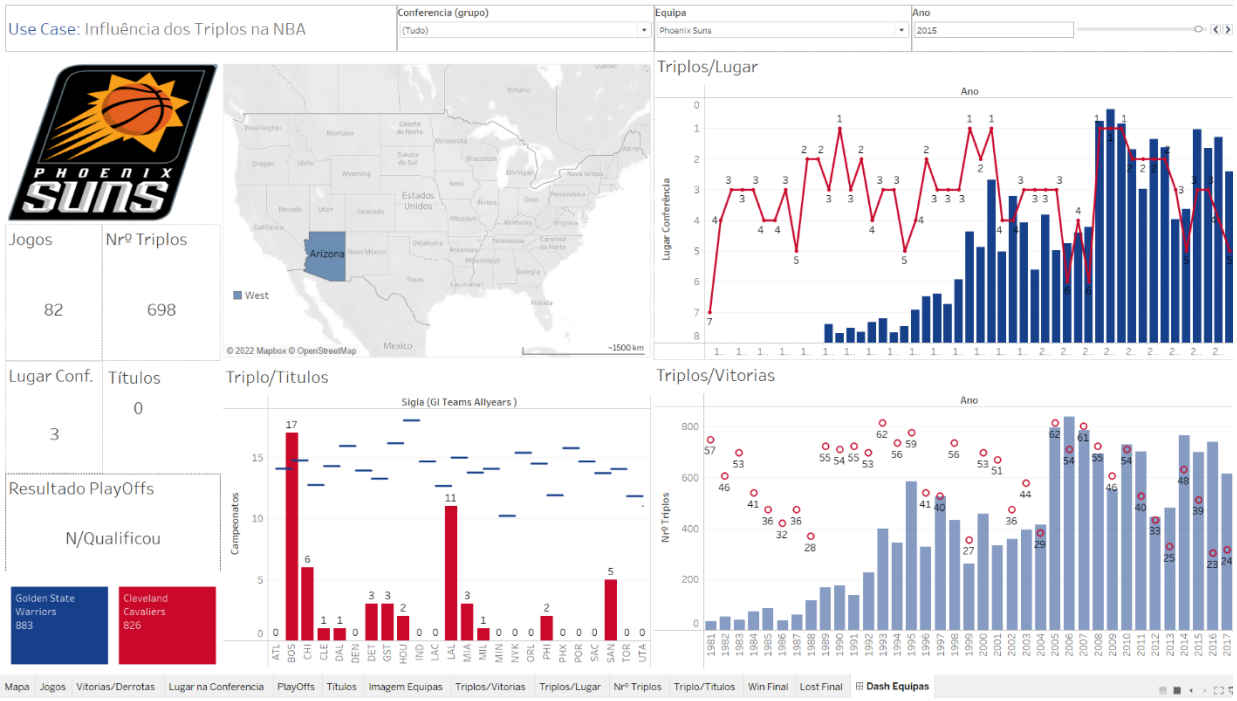
Simple Statistics

Label	Count	%
Row Count	1483	100.00%
Null Count	0	0.00%
Duplicate Count	30	2.02%
Distinct Count	1453	97.98%



1.7 Dashboard:





2. SUBGRUPO 2

Use Case: A influência da estatura física na NBA

Nuno Moreira

A92947 - 3º ano LEGSI

a92947@alunos.uminho.pt

João Lemos

A92939 - 3º ano LEGSI

a92939@alunos.uminho.pt

2.1 Introdução

A partir do nosso Use Case (A influência da estatura física na NBA) extrapolámos 3 questões principais que são abordados em dashboards diferentes.
Inicialmente pretendemos verificar se a altura é um dado influenciador nas estatísticas de cada jogador. Através desta conseguiremos verificar dados como as alterações de alturas médias ou a evolução de dados estatísticos ao logo das épocas.
Numa altura em que cada vez mais o desporto valoriza o IMC dos jogadores, pretendemos verificar se este tem uma influência determinante na velocidade, minutos e distância percorrida. Para isso utilizaremos a fórmula europeia para podermos comparar o IMC com parte do rendimento físico do jogador em campo.
Idade e lesões normalmente são dados fundamentais na continuação da carreira de um jogador, mas de que forma a vitalidade dos jogadores tem sido alterada e como isso tem sido decisivo na escolha da utilização ou não por parte do treinador.

2.2 KPI's

Para o desenvolvimento deste trabalho definimos como KPI indicadores relativos ao físico de cada jogador. Dessa forma selecionamos a Altura, Idade e IMC (altura, peso).

2.3 Analytical Questions

A influência da altura na eficácia dos Triplos ao longo dos anos;

A influência da altura no número de Assists, Field Goals e Free Throws;

A influência da altura nos Blocks;

Quais os jogadores de NBA mais altos;

Quais os jogadores de NBA mais baixos;

A influência da idade no número de jogos/titularidade;

A influência da idade nos minutos jogados;

A influência da idade no resultado dos jogos;

Quais os jogadores de NBA mais velhos;

Quais os estreantes na NBA mais velhos;

A influência do IMC na distância ofensiva e defensiva percorrida;

A relação entre o número de minutos e o IMC de cada jogador ao longo dos anos;

A influência do IMC na velocidade máxima e média;

Quais os jogadores de NBA com IMC mais elevado;

Quais os jogadores de NBA mais rápidos.

2.4 DataSets

Altura e Peso: https://www.kaggle.com/drgilermo/nba-players-stats?select=player_data.csv

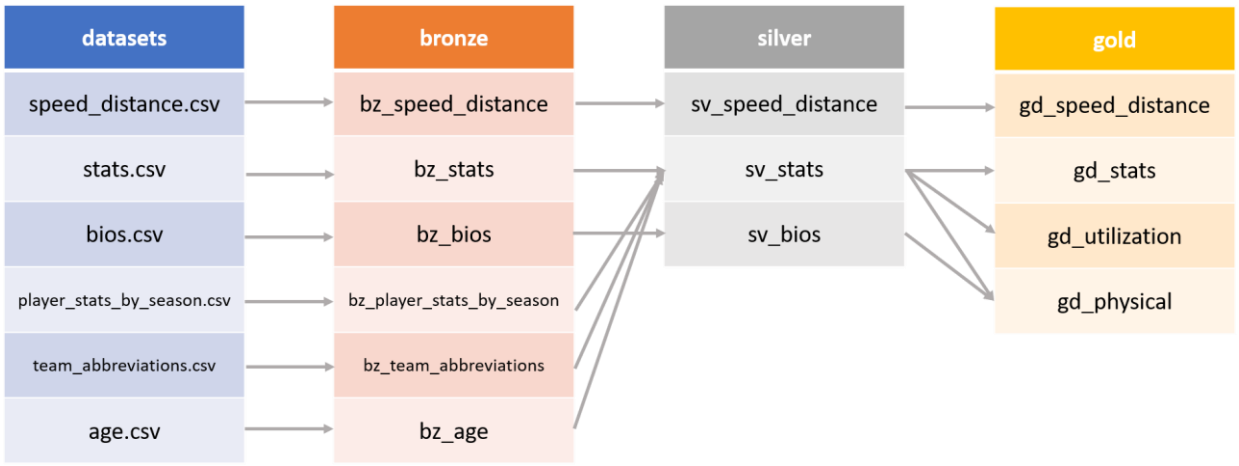
Dados estatísticos e Idade: https://www.kaggle.com/drgilermo/nba-players- stats?select=Seasons_Stats.csv

Velocidade e distância: https://www.nba.com/stats/players/speed-distance/ *

* Os datasets (speed_distance) são obtido através de um serviço que ajuda a fazer web scraping.

Para este Use Case o par pretendia utilizar apenas dados retirados da fonte oficial (stats.nba.com) através de web scraping. Devido a esta plataforma apenas facultar dados por época teríamos de utilizar um elevado número de datasets, já que como grupo estabelecemos um número elevado de anos igual para todos, de forma a promover coerência e um melhor trabalho de grupo.
Para isso tivemos de alteramos todos os datasets retirados de stats.nba.com para versões disponibilizadas no kaggle que agrupassem várias temporadas. No entanto, para usar a velocidade e distância como estatísticas tivemos de manter a ideia original, já que não encontramos nenhum dataset que garantisse esses dados relativos a várias épocas.

2.5 DataSet to Bronze to Silver to Gold



2.6 Data Quality

2.6.1 Análise ao DataSet dos Dados Estatísticos e Idade:

Este será o dataset utilizado para retirar todos os dados necessários para efetuar análises estatísticas e análises que incluam a idade como KPI. Qualquer erro detetado através desta fase será corrigido na transformação do processo ELT.

Quality Analysis:

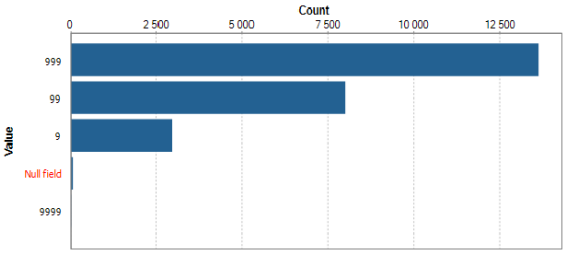
- Pattern Frequency
- Mean
- Median
- Range
- Minimun
- Maximum
- Value Frequency
- Null Count
- Distinct Count
- Unique Count
- Duplicate Count
- Row Count

Erros de Qualidade:

- Detetadas 67 linhas vazias (apenas com a coluna índex preenchidas). Desta forma, todas as colunas contam com um mínimo de 67 blank/null.
- Algumas colunas relativas a estatísticas de jogo têm linhas em branco, correspondentes aos anos em que essas estatísticas não eram registadas ou não existiam.
- Identificados alguns valores mínimos de 0 nas estatísticas de lançamentos (FTa, 3pa, etc) devido a existirem jogadores com um tempo de jogo reduzido que os impossibilitou de fazer qualquer lançamento ou ponto.
- Além das 67 linhas, existem alguns jogadores dos anos 50 sem dados relativos à idade. Este é um dado insignificante já que apenas usaremos dados a partir de meados do fim da época de 70.
- Games Started é um dado que apenas começou a ser registado em 1982, daí as mais de 6000 linhas em branco.

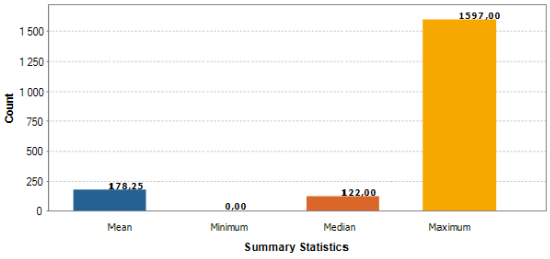
▾ **Pattern Frequency**

Value	Count	%
999	13640	55.24%
99	8003	32.41%
9	2965	12.01%
Null field	67	0.27%
9999	16	0.06%



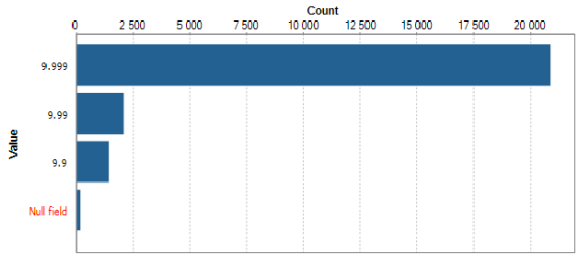
▾ **Summary Statistics**

Label	Value
Mean	178.25044671864848
Median	122.0
Range	1597.0
Minimum	0
Maximum	1597



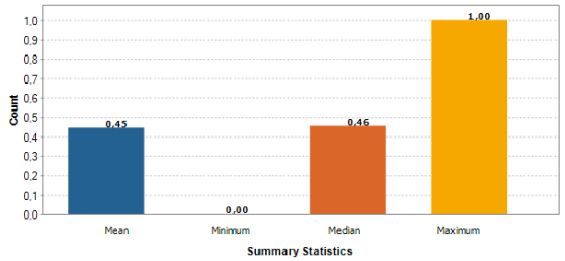
▾ **Pattern Frequency**

Value	Count	%
9.999	20919	84.72%
9.99	2129	8.62%
9.9	1448	5.86%
Null field	195	0.79%



▾ **Summary Statistics**

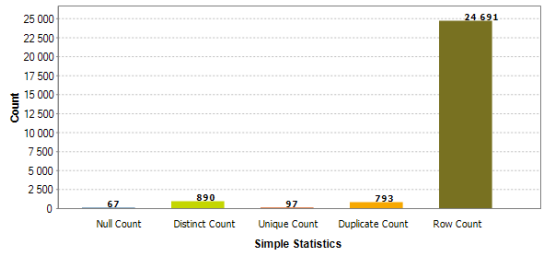
Label	Value
Mean	0.4453430764206401
Median	0.456
Range	1.0
Minimum	0.0
Maximum	1.0



▾ **Column: metadata_P1**

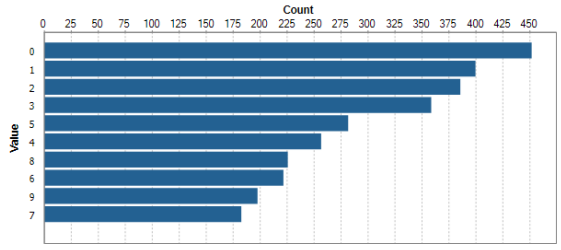
▾ **Simple Statistics**

Label	Count	%
Null Count	67	0.27%
Distinct Count	890	3.60%
Unique Count	97	0.39%
Duplicate Count	793	3.21%
Row Count	24691	100.00%



▾ **Value Frequency**

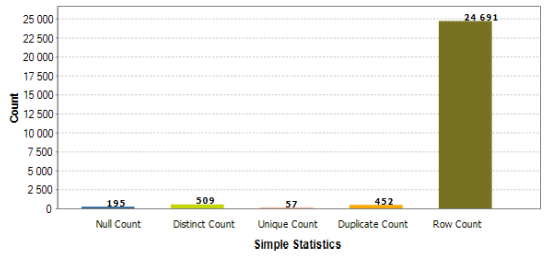
Value	Count	%
0	452	1.83%
1	400	1.62%
2	386	1.56%
3	359	1.45%
5	282	1.14%
4	257	1.04%
8	226	0.92%
6	222	0.90%
9	198	0.80%
7	183	0.74%



Column: metadata.Column39

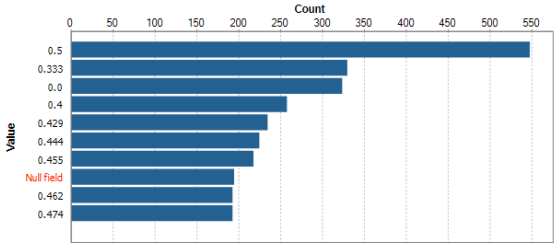
Simple Statistics

Label	Count	%
Null Count	195	0.79%
Distinct Count	509	2.06%
Unique Count	57	0.23%
Duplicate Count	452	1.83%
Row Count	24691	100.00%



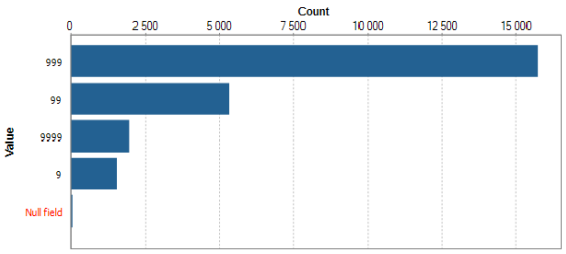
Value Frequency

Value	Count	%
0.5	548	2.22%
0.333	330	1.34%
0.0	324	1.31%
0.4	258	1.04%
0.429	235	0.95%
0.444	225	0.91%
0.455	218	0.88%
Null field	195	0.79%
0.462	193	0.78%
0.474	193	0.78%



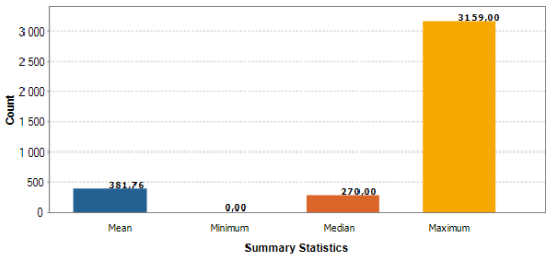
Pattern Frequency

Value	Count	%
999	15738	63.74%
99	5348	21.66%
9999	1980	8.02%
9	1558	6.31%
Null field	67	0.27%



Summary Statistics

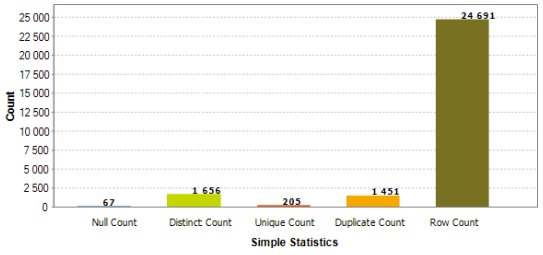
Label	Value
Mean	381.75678200129954
Median	270.0
Range	3159.0
Minimum	0
Maximum	3159



Column: metadata._PA1

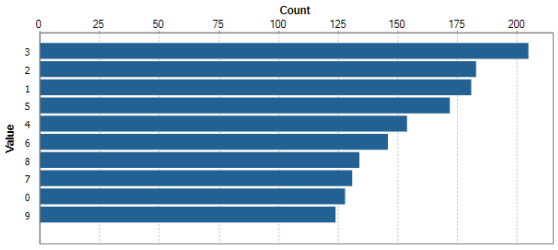
Simple Statistics

Label	Count	%
Null Count	67	0.27%
Distinct Count	1656	6.71%
Unique Count	205	0.83%
Duplicate Count	1451	5.88%
Row Count	24691	100.00%



Value Frequency

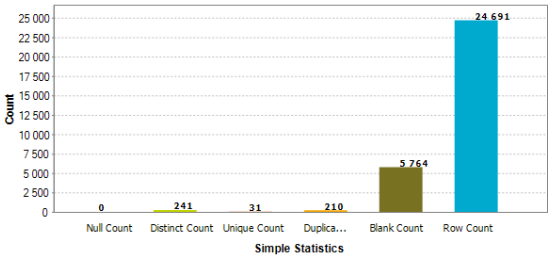
Value	Count	%
3	205	0.83%
2	183	0.74%
1	181	0.73%
5	172	0.70%
4	154	0.62%
6	146	0.59%
8	134	0.54%
7	131	0.53%
0	128	0.52%
9	124	0.50%



Column: metadata._P

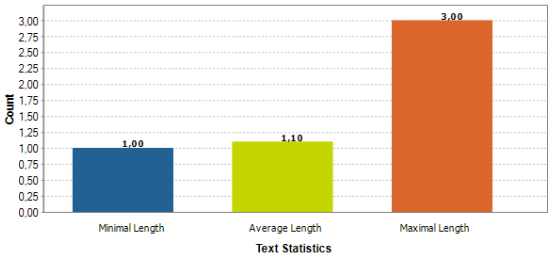
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	241	0.98%
Unique Count	31	0.13%
Duplicate Count	210	0.85%
Blank Count	5764	23.34%
Row Count	24691	100.00%



Text Statistics

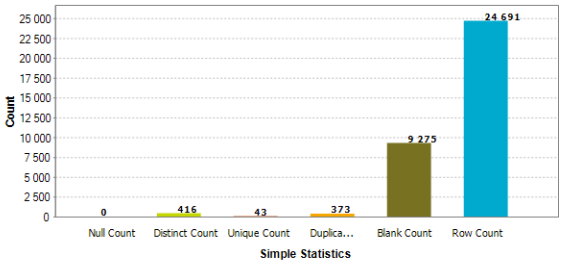
Label	Value
Minimal Length	1.00
Average Length	1.10
Maximal Length	3.00



Column: metadata.Column36

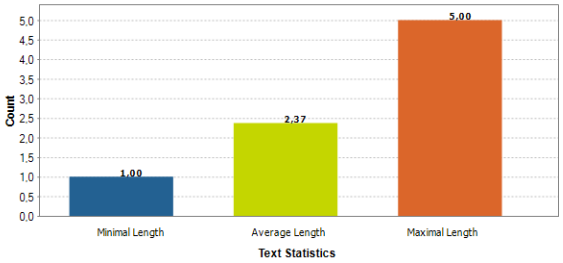
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	416	1.68%
Unique Count	43	0.17%
Duplicate Count	373	1.51%
Blank Count	9275	37.56%
Row Count	24691	100.00%



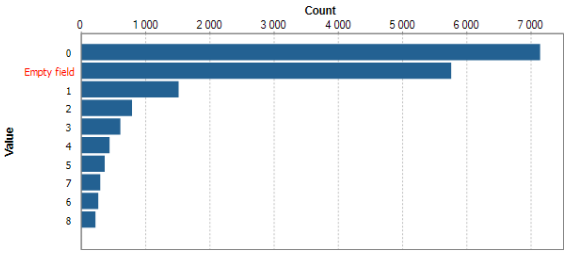
Text Statistics

Label	Value
Minimal Length	1.00
Average Length	2.37
Maximal Length	5.00



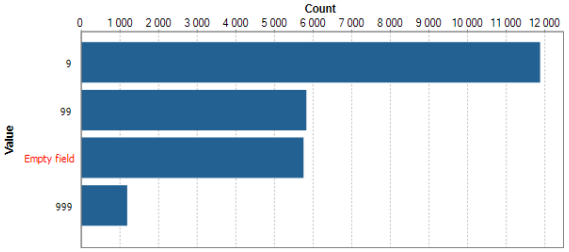
Value Frequency

Value	Count	%
0	7149	28.95%
Empty field	5764	23.34%
1	1521	6.16%
2	796	3.22%
3	614	2.49%
4	446	1.81%
5	371	1.50%
7	302	1.22%
6	271	1.10%
8	226	0.92%



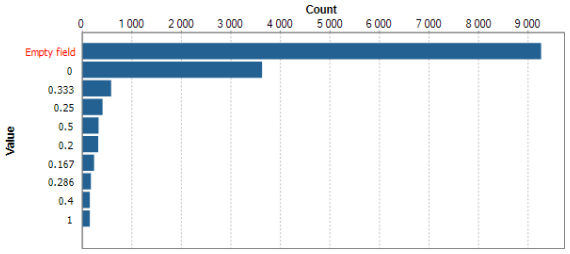
Pattern Frequency

Value	Count	%
9	11891	48.16%
99	5837	23.64%
Empty field	5764	23.34%
999	1199	4.86%



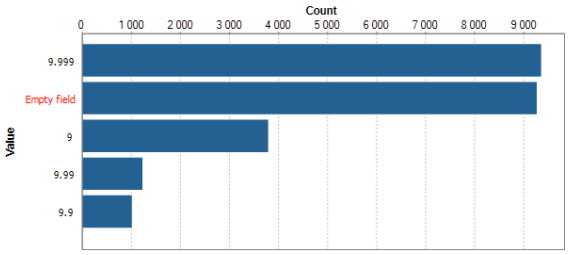
Value Frequency

Value	Count	%
Empty field	9275	37.56%
0	3638	14.73%
0.333	591	2.39%
0.25	417	1.69%
0.5	336	1.36%
0.2	330	1.34%
0.167	245	0.99%
0.286	181	0.73%
0.4	160	0.65%
1	160	0.65%



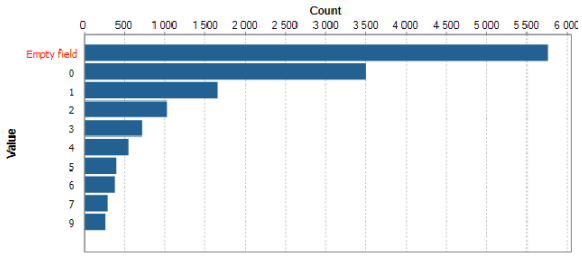
Pattern Frequency

Value	Count	%
9.999	9365	37.93%
Empty field	9275	37.56%
9	3798	15.38%
9.99	1234	5.00%
9.9	1019	4.13%



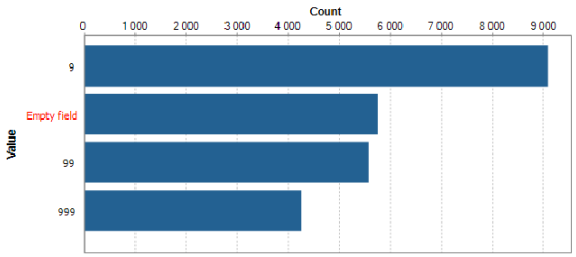
Value Frequency

Value	Count	%
Empty field	5764	23.34%
0	3511	14.22%
1	1664	6.74%
2	1036	4.20%
3	723	2.93%
4	556	2.25%
5	408	1.65%
6	389	1.58%
7	295	1.19%
9	265	1.07%



Pattern Frequency

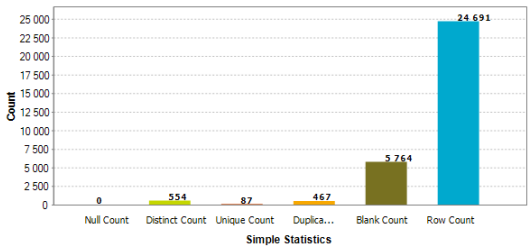
Value	Count	%
9	9093	36.83%
Empty field	5764	23.34%
99	5576	22.58%
999	4258	17.25%



Column: metadata_PA

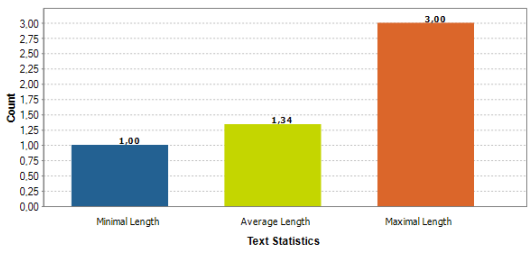
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	554	2.24%
Unique Count	87	0.35%
Duplicate Count	467	1.89%
Blank Count	5764	23.34%
Row Count	24691	100.00%



Text Statistics

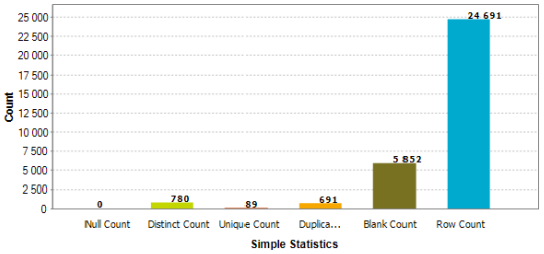
Label	Value
Minimal Length	1.00
Average Length	1.34
Maximal Length	3.00



Column: metadata._PAr

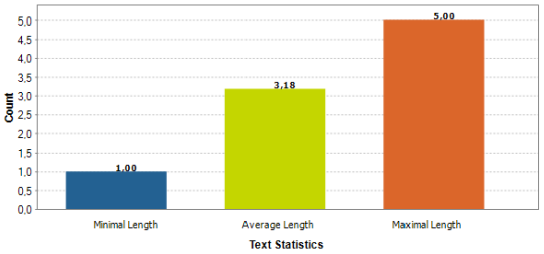
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	780	3.16%
Unique Count	89	0.36%
Duplicate Count	691	2.80%
Blank Count	5852	23.70%
Row Count	24691	100.00%



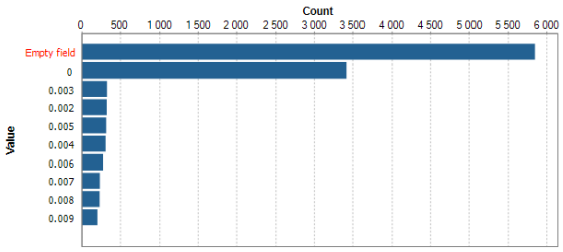
Text Statistics

Label	Value
Minimal Length	1.00
Average Length	3.18
Maximal Length	5.00



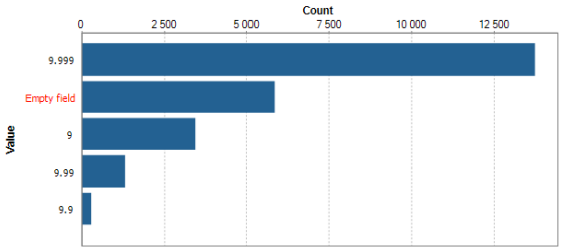
Value Frequency

Value	Count	%
Empty field	5852	23.70%
0	3423	13.86%
0.003	330	1.34%
0.002	327	1.32%
0.005	320	1.30%
0.004	313	1.27%
0.006	280	1.13%
0.007	237	0.96%
0.008	234	0.95%
0.009	206	0.83%



Pattern Frequency

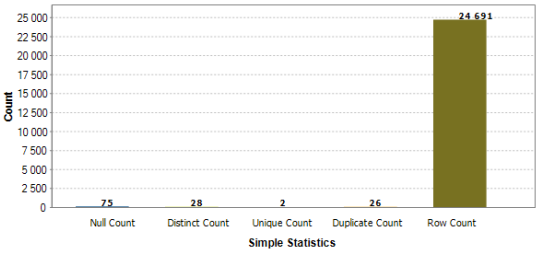
Value	Count	%
9.999	13760	55.73%
Empty field	5852	23.70%
9	3452	13.98%
9.99	1332	5.39%
9.9	295	1.19%



Column: metadata.Age

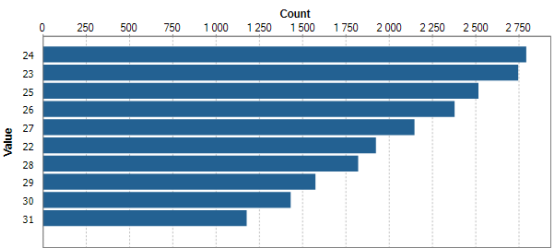
Simple Statistics

Label	Count	%
Null Count	75	0.30%
Distinct Count	28	0.11%
Unique Count	2	8.1E-3%
Duplicate Count	26	0.11%
Row Count	24691	100.00%



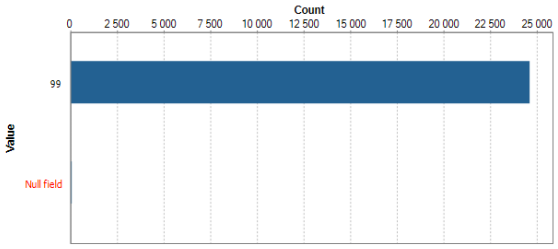
Value Frequency

Value	Count	%
24	2794	11.32%
23	2748	11.13%
25	2518	10.20%
26	2380	9.64%
27	2149	8.70%
22	1926	7.80%
28	1823	7.38%
29	1576	6.38%
30	1433	5.80%
31	1179	4.78%



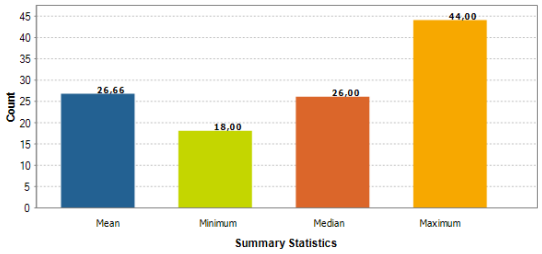
Pattern Frequency

Value	Count	%
99	24616	99.70%
Null field	75	0.30%



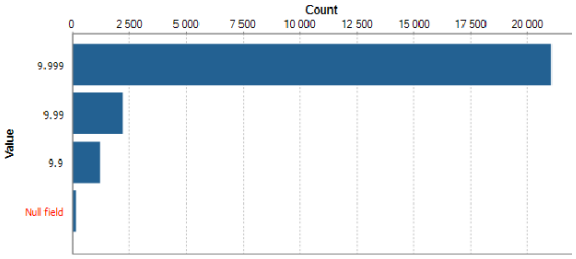
Summary Statistics

Label	Value
Mean	26.66440526486838
Median	26.0
Range	26.0
Minimum	18
Maximum	44



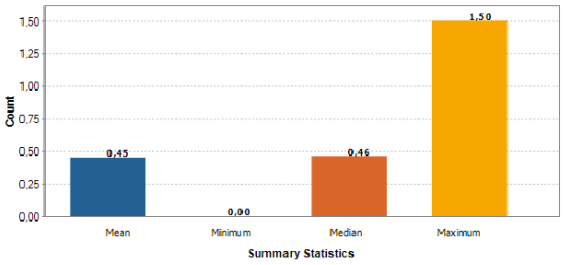
Pattern Frequency

Value	Count	%
9.999	21067	85.32%
9.99	2230	9.03%
9.9	1228	4.97%
Null field	166	0.67%



Summary Statistics

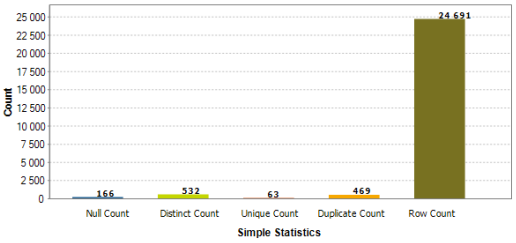
Label	Value
Mean	0.45065838939857294
Median	0.463
Range	1.5
Minimum	0.0
Maximum	1.5



Column: metadata.eFG_

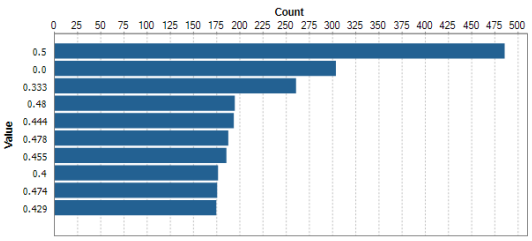
Simple Statistics

Label	Count	%
Null Count	166	0.67%
Distinct Count	532	2.15%
Unique Count	63	0.26%
Duplicate Count	469	1.90%
Row Count	24691	100.00%



Value Frequency

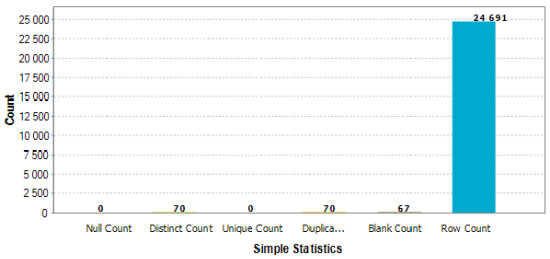
Value	Count	%
0.5	486	1.97%
0.0	304	1.23%
0.333	261	1.06%
0.48	195	0.79%
0.444	194	0.79%
0.478	188	0.76%
0.455	186	0.75%
0.4	177	0.72%
0.474	176	0.71%
0.429	175	0.71%



Column: metadata.Tm

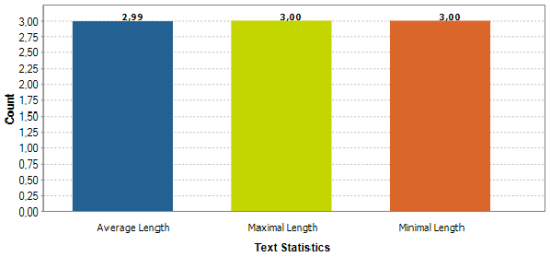
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	70	0.28%
Unique Count	0	0.00%
Duplicate Count	70	0.28%
Blank Count	67	0.27%
Row Count	24691	100.00%



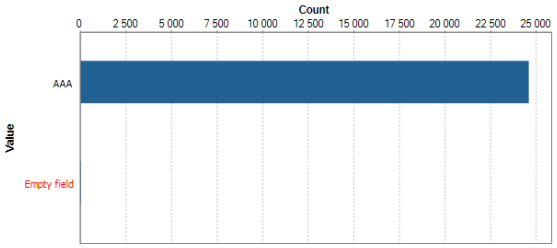
Text Statistics

Label	Value
Average Length	2.99
Maximal Length	3.00
Minimal Length	3.00



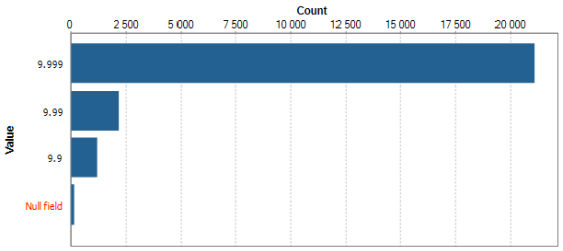
Pattern Frequency

Value	Count	%
AAA	24624	99.73%
Empty field	67	0.27%



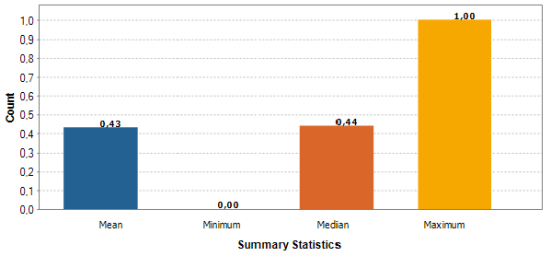
Pattern Frequency

Value	Count	%
9.999	21119	85.53%
9.99	2180	8.83%
9.9	1226	4.97%
Null field	166	0.67%



Summary Statistics

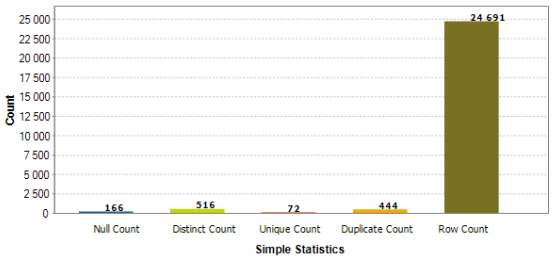
Label	Value
Mean	0.4308172884811417
Median	0.439
Range	1.0
Minimum	0.0
Maximum	1.0



Column: metadata.FG_

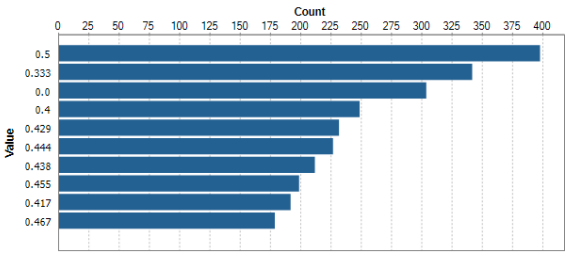
Simple Statistics

Label	Count	%
Null Count	166	0.67%
Distinct Count	516	2.09%
Unique Count	72	0.29%
Duplicate Count	444	1.80%
Row Count	24691	100.00%



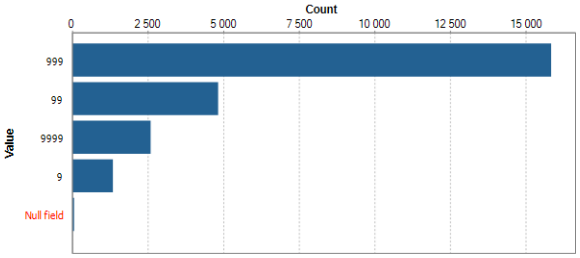
Value Frequency

Value	Count	%
0.5	398	1.61%
0.333	342	1.39%
0.0	304	1.23%
0.4	249	1.01%
0.429	232	0.94%
0.444	227	0.92%
0.438	212	0.86%
0.455	199	0.81%
0.417	192	0.78%
0.467	179	0.72%



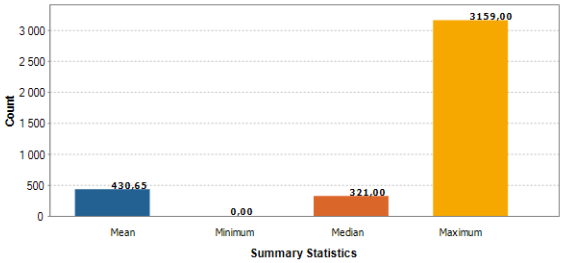
Pattern Frequency

Value	Count	%
999	15845	64.17%
99	4831	19.57%
9999	2599	10.53%
9	1349	5.46%
Null field	67	0.27%



Summary Statistics

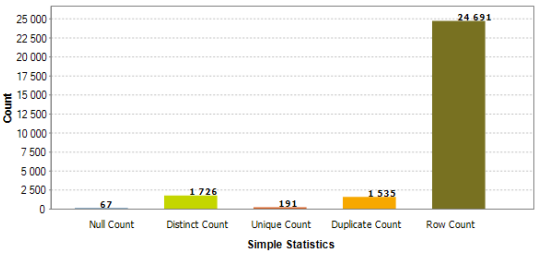
Label	Value
Mean	430.6457521117609
Median	321.0
Range	3159.0
Minimum	0
Maximum	3159



Column: metadata.FGA

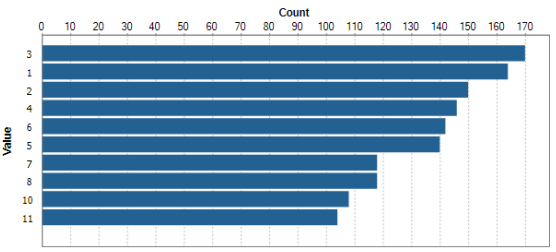
Simple Statistics

Label	Count	%
Null Count	67	0.27%
Distinct Count	1726	6.99%
Unique Count	191	0.77%
Duplicate Count	1535	6.22%
Row Count	24691	100.00%



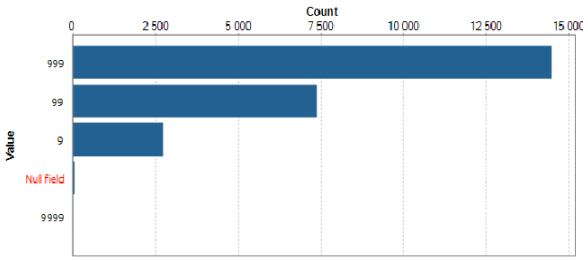
Value Frequency

Value	Count	%
3	170	0.69%
1	164	0.66%
2	150	0.61%
4	146	0.59%
6	142	0.58%
5	140	0.57%
7	118	0.48%
8	118	0.48%
10	108	0.44%
11	104	0.42%



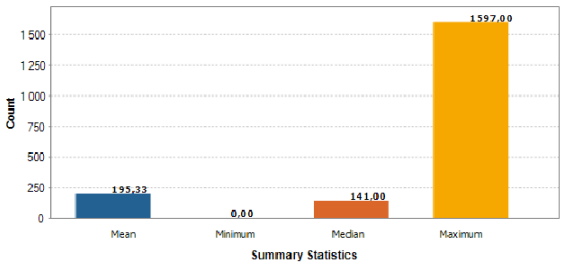
Pattern Frequency

Value	Count	%
999	14486	58.67%
99	7398	29.96%
9	2722	11.02%
Null field	67	0.27%
9999	18	0.07%



Summary Statistics

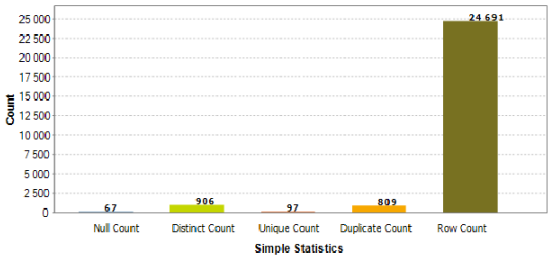
Label	Value
Mean	195.32582033788174
Median	141.0
Range	1597.0
Minimum	0
Maximum	1597



Column: metadata.FG

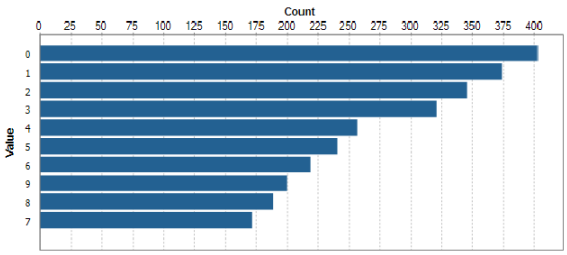
Simple Statistics

Label	Count	%
Null Count	67	0.27%
Distinct Count	906	3.67%
Unique Count	97	0.39%
Duplicate Count	809	3.28%
Row Count	24691	100.00%



Value Frequency

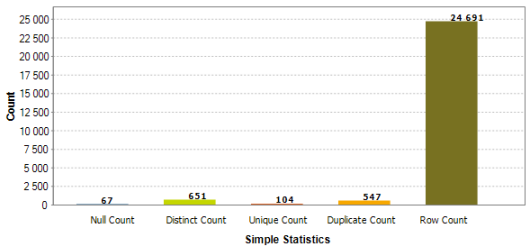
Value	Count	%
0	403	1.63%
1	374	1.51%
2	346	1.40%
3	321	1.30%
4	257	1.04%
5	241	0.98%
6	219	0.89%
9	200	0.81%
8	189	0.77%
7	172	0.70%



Column: metadata.FT

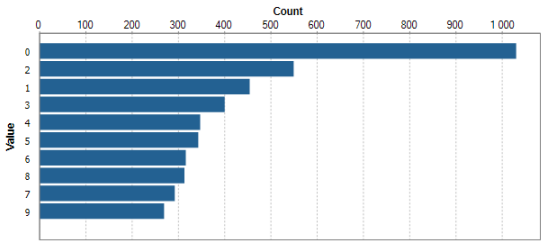
Simple Statistics

Label	Count	%
Null Count	67	0.27%
Distinct Count	651	2.64%
Unique Count	104	0.42%
Duplicate Count	547	2.22%
Row Count	24691	100.00%



Value Frequency

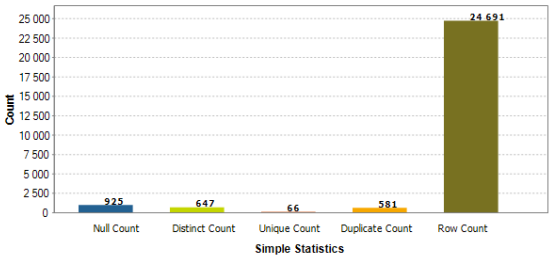
Value	Count	%
0	1031	4.18%
2	550	2.23%
1	455	1.84%
3	401	1.62%
4	348	1.41%
5	344	1.39%
6	317	1.28%
8	314	1.27%
7	293	1.19%
9	270	1.09%



Column: metadata.FT_

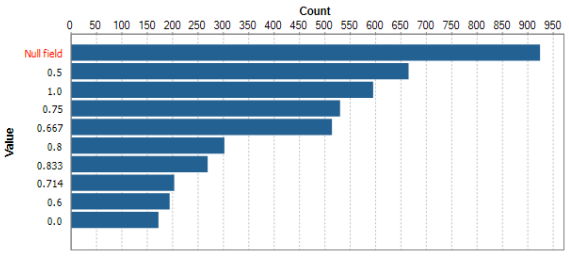
Simple Statistics

Label	Count	%
Null Count	925	3.75%
Distinct Count	647	2.62%
Unique Count	66	0.27%
Duplicate Count	581	2.35%
Row Count	24691	100.00%



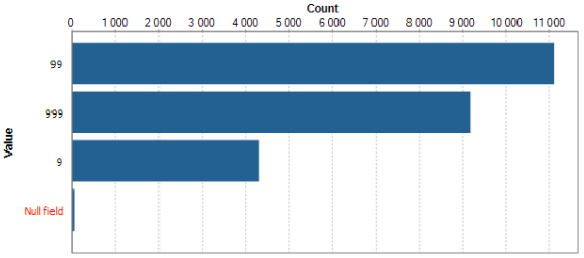
Value Frequency

Value	Count	%
Null field	925	3.75%
0.5	666	2.70%
1.0	596	2.41%
0.75	531	2.15%
0.667	515	2.09%
0.8	303	1.23%
0.833	270	1.09%
0.714	204	0.83%
0.6	195	0.79%
0.0	173	0.70%



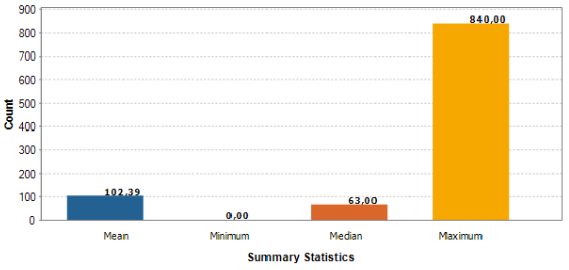
Pattern Frequency

Value	Count	%
99	11107	44.98%
999	9194	37.24%
9	4323	17.51%
Null field	67	0.27%



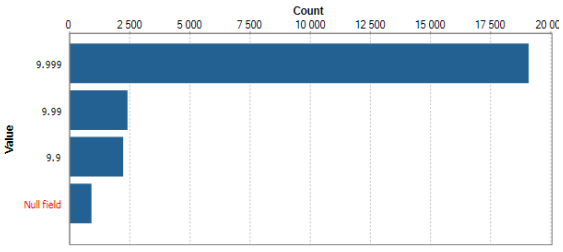
Summary Statistics

Label	Value
Mean	102.38933560753736
Median	63.0
Range	840.0
Minimum	0
Maximum	840



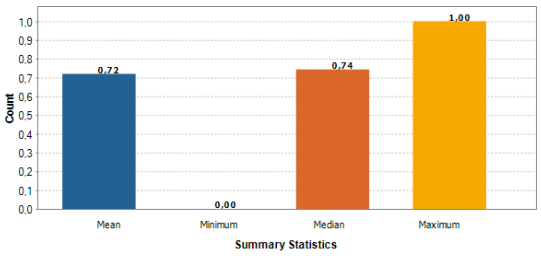
Pattern Frequency

Value	Count	%
9.999	19103	77.37%
9.99	2422	9.81%
9.9	2241	9.08%
Null field	925	3.75%



Summary Statistics

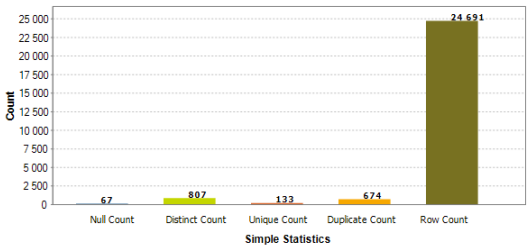
Label	Value
Mean	0.7192786333417487
Median	0.743
Range	1.0
Minimum	0.0
Maximum	1.0



Column: metadata.FTA

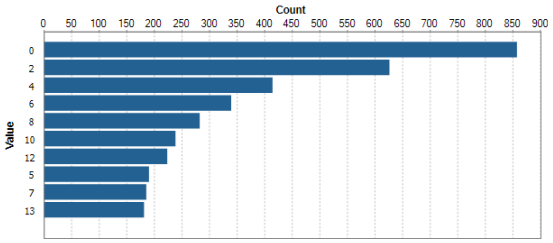
Simple Statistics

Label	Count	%
Null Count	67	0.27%
Distinct Count	807	3.27%
Unique Count	133	0.54%
Duplicate Count	674	2.73%
Row Count	24691	100.00%



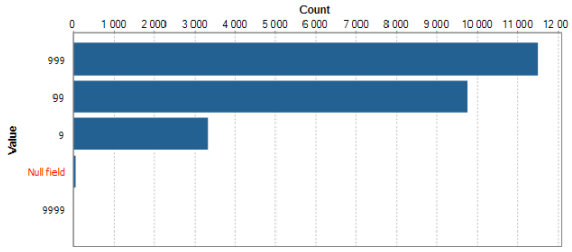
Value Frequency

Value	Count	%
0	858	3.47%
2	627	2.54%
4	415	1.68%
6	340	1.38%
8	283	1.15%
10	239	0.97%
12	224	0.91%
5	191	0.77%
7	186	0.75%
13	182	0.74%



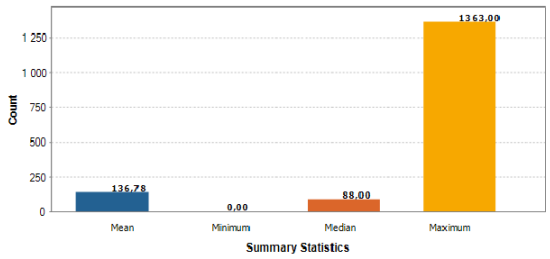
Pattern Frequency

Value	Count	%
999	11509	46.61%
99	9773	39.58%
9	3338	13.52%
Null field	67	0.27%
9999	4	0.02%



Summary Statistics

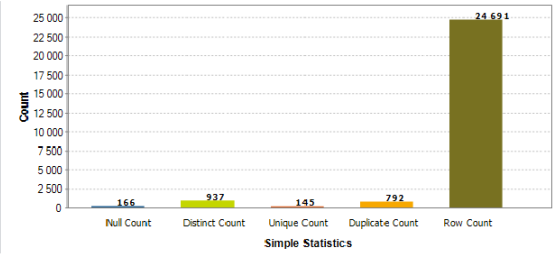
Label	Value
Mean	136.77521929824562
Median	88.0
Range	1363.0
Minimum	0
Maximum	1363



Column: metadata.FTr

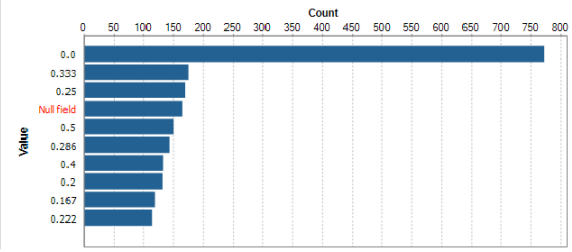
Simple Statistics

Label	Count	%
Null Count	166	0.67%
Distinct Count	937	3.79%
Unique Count	145	0.59%
Duplicate Count	792	3.21%
Row Count	24691	100.00%



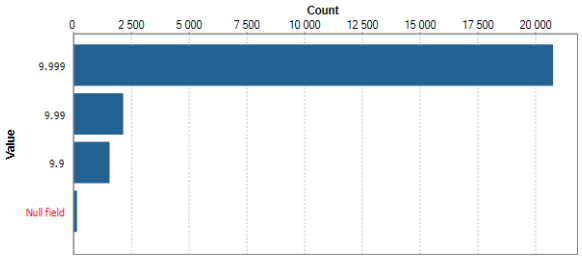
Value Frequency

Value	Count	%
0.0	772	3.13%
0.333	175	0.71%
0.25	171	0.69%
Null field	166	0.67%
0.5	151	0.61%
0.286	144	0.58%
0.4	133	0.54%
0.2	132	0.53%
0.167	119	0.48%
0.222	114	0.46%



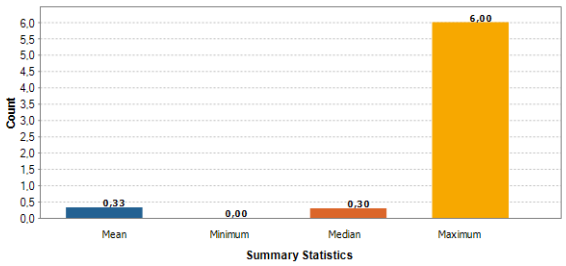
Pattern Frequency

Value	Count	%
9.999	20777	84.15%
9.99	2168	8.78%
9.9	1580	6.40%
Null field	166	0.67%



Summary Statistics

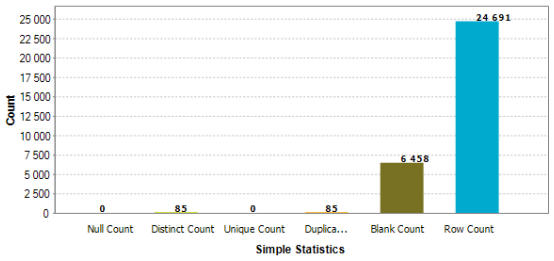
Label	Value
Mean	0.3254550458715596
Median	0.296
Range	6.0
Minimum	0.0
Maximum	6.0



Column: metadata.GS

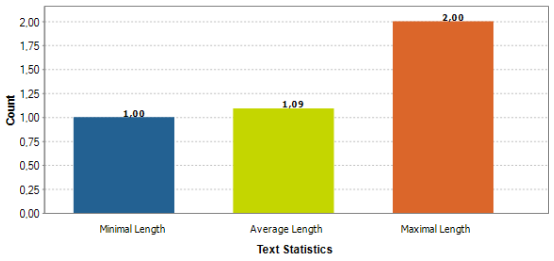
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	85	0.34%
Unique Count	0	0.00%
Duplicate Count	85	0.34%
Blank Count	6458	26.16%
Row Count	24691	100.00%



Text Statistics

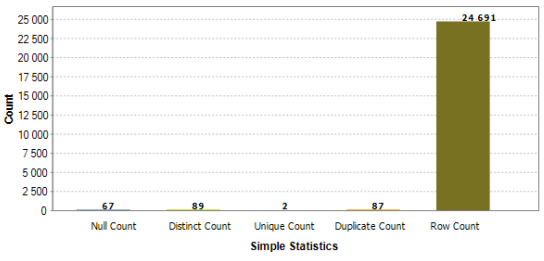
Label	Value
Minimal Length	1.00
Average Length	1.09
Maximal Length	2.00



Column: metadata.G

Simple Statistics

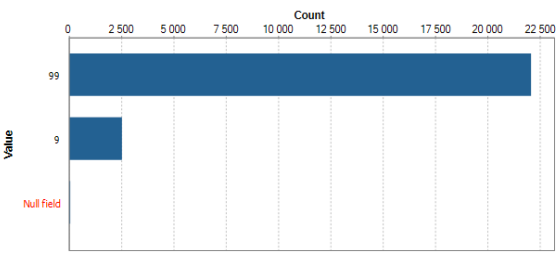
Label	Count	%
Null Count	67	0.27%
Distinct Count	89	0.36%
Unique Count	2	8.1E-3%
Duplicate Count	87	0.35%
Row Count	24691	100.00%



Value Frequency

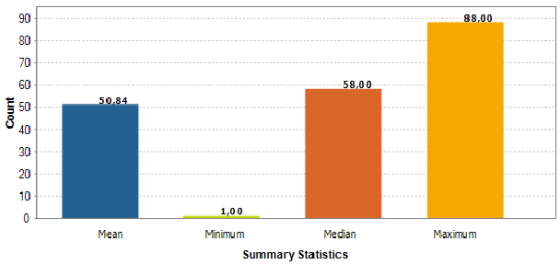
Pattern Frequency

Value	Count	%
99	22087	89.45%
9	2537	10.27%
Null field	67	0.27%



Summary Statistics

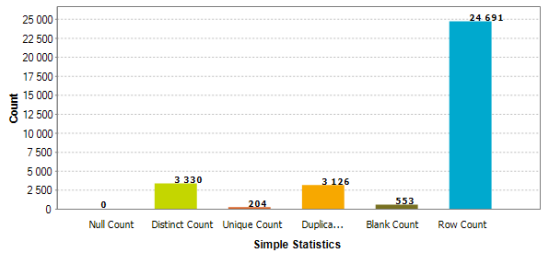
Label	Value
Mean	50.83711013645224
Median	58.0
Range	87.0
Minimum	1
Maximum	88



Column: metadata.MP

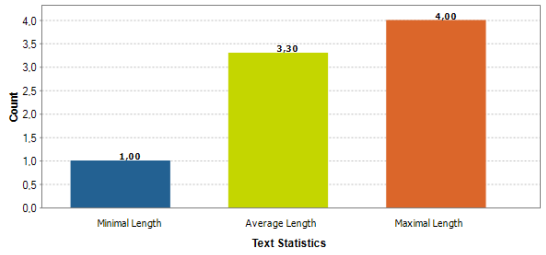
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	3330	13.49%
Unique Count	204	0.83%
Duplicate Count	3126	12.66%
Blank Count	553	2.24%
Row Count	24691	100.00%



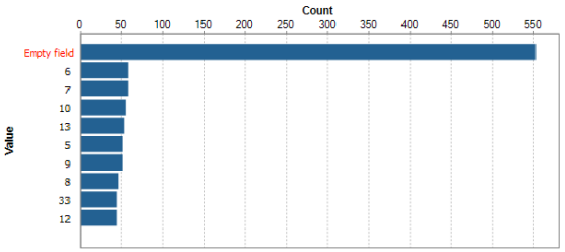
Text Statistics

Label	Value
Minimal Length	1.00
Average Length	3.30
Maximal Length	4.00



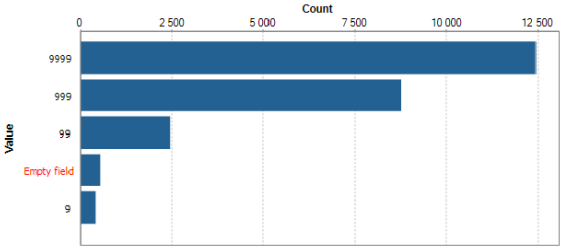
Value Frequency

Value	Count	%
Empty field	553	2.24%
6	59	0.24%
7	59	0.24%
10	56	0.23%
13	54	0.22%
5	52	0.21%
9	52	0.21%
8	47	0.19%
33	45	0.18%
12	45	0.18%



Pattern Frequency

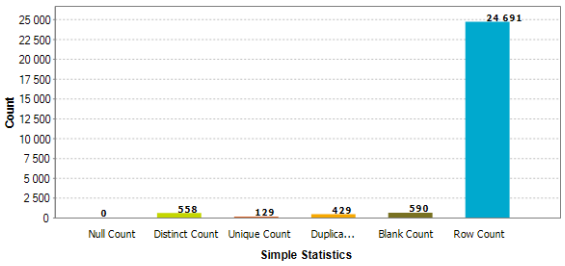
Value	Count	%
9999	12458	50.46%
999	8790	35.60%
99	2464	9.98%
Empty field	553	2.24%
9	426	1.73%



Column: metadata.PER

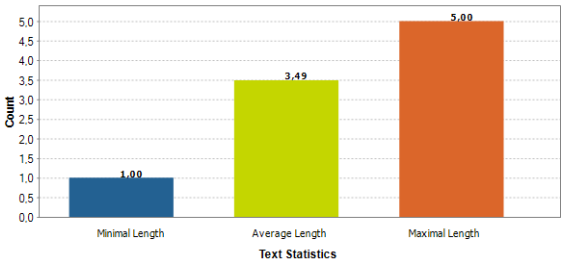
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	558	2.26%
Unique Count	129	0.52%
Duplicate Count	429	1.74%
Blank Count	590	2.39%
Row Count	24691	100.00%



Text Statistics

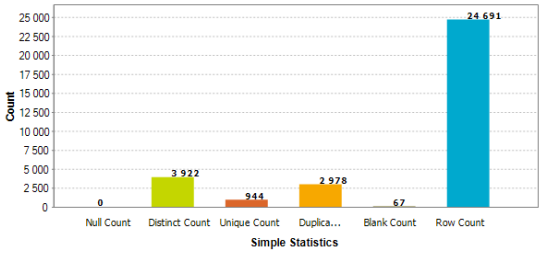
Label	Value
Minimal Length	1.00
Average Length	3.49
Maximal Length	5.00



Column: metadata.Player

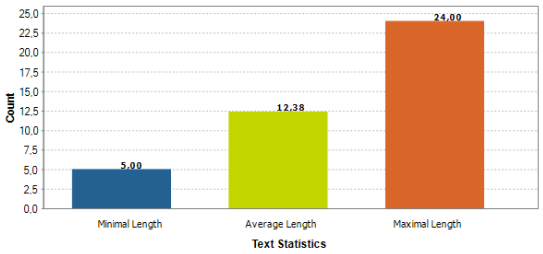
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	3922	15.88%
Unique Count	944	3.82%
Duplicate Count	2978	12.06%
Blank Count	67	0.27%
Row Count	24691	100.00%



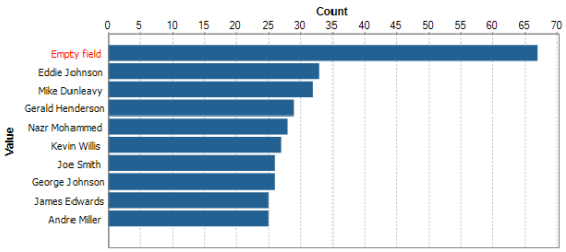
Text Statistics

Label	Value
Minimal Length	5.00
Average Length	12.38
Maximal Length	24.00



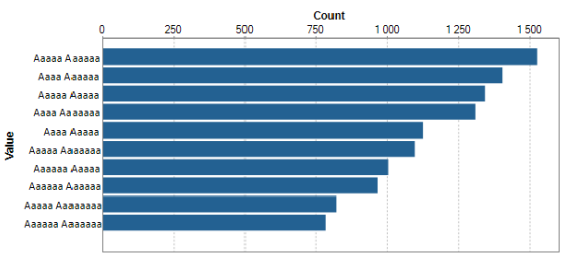
Value Frequency

Value	Count	%
Empty field	67	0.27%
Eddie Johnson	33	0.13%
Mike Dunleavy	32	0.13%
Gerald Henderson	29	0.12%
Nazr Mohammed	28	0.11%
Kevin Willis	27	0.11%
Joe Smith	26	0.11%
George Johnson	26	0.11%
James Edwards	25	0.10%
Andre Miller	25	0.10%



Pattern Frequency

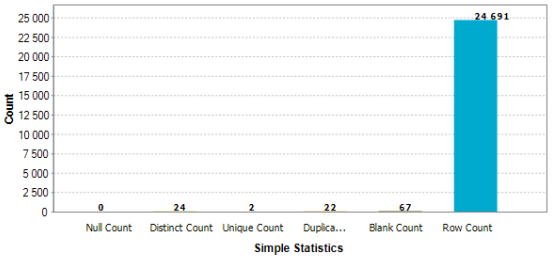
Value	Count	%
Aaaaa Aaaaaa	1527	6.18%
Aaaa Aaaaaa	1407	5.70%
Aaaaa Aaaaa	1345	5.45%
Aaaa Aaaaaaa	1311	5.31%
Aaaa Aaaaa	1125	4.56%
Aaaaa Aaaaaaa	1099	4.45%
Aaaaaa Aaaaa	1005	4.07%
Aaaaaa Aaaaaa	967	3.92%
Aaaaa Aaaaaaaa	821	3.33%
Aaaaaa Aaaaaaa	786	3.18%



Column: metadata.Pos

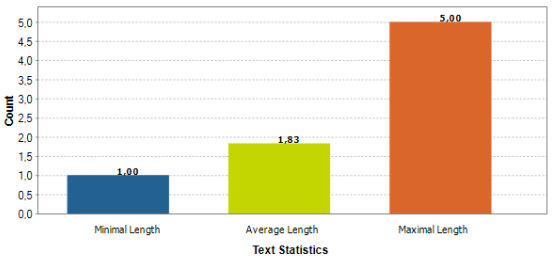
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	24	0.10%
Unique Count	2	8.1E-3%
Duplicate Count	22	0.09%
Blank Count	67	0.27%
Row Count	24691	100.00%



Text Statistics

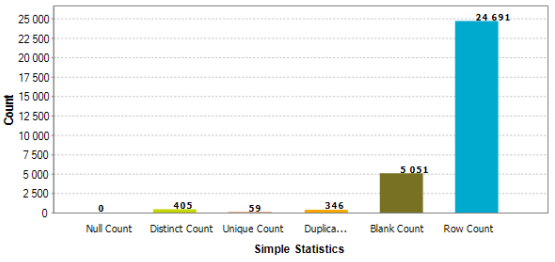
Label	Value
Minimal Length	1.00
Average Length	1.83
Maximal Length	5.00



Column: metadata.USG_

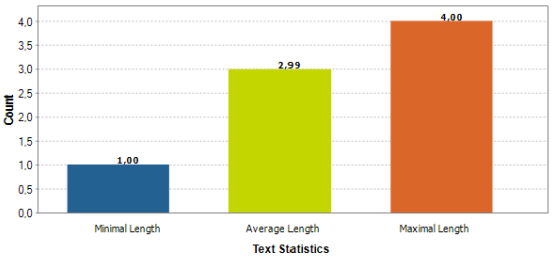
Simple Statistics

Label	Count	%
Null Count	0	0.00%
Distinct Count	405	1.64%
Unique Count	59	0.24%
Duplicate Count	346	1.40%
Blank Count	5051	20.46%
Row Count	24691	100.00%



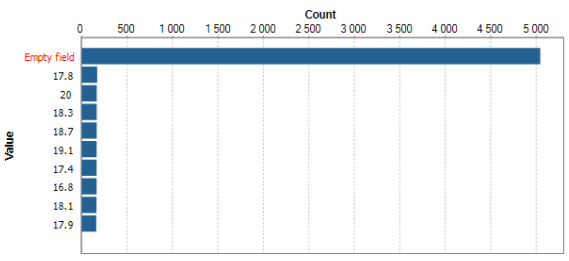
Text Statistics

Label	Value
Minimal Length	1.00
Average Length	2.99
Maximal Length	4.00



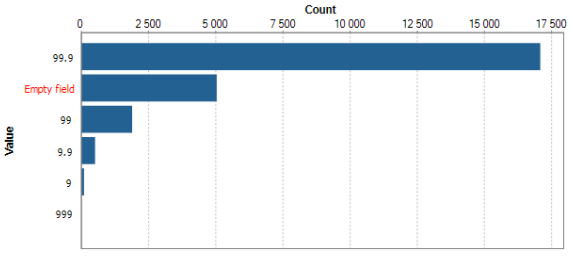
Value Frequency

Value	Count	%
Empty field	5051	20.46%
17.8	178	0.72%
20	173	0.70%
18.3	173	0.70%
18.7	173	0.70%
19.1	171	0.69%
17.4	171	0.69%
16.8	171	0.69%
18.1	170	0.69%
17.9	167	0.68%



Pattern Frequency

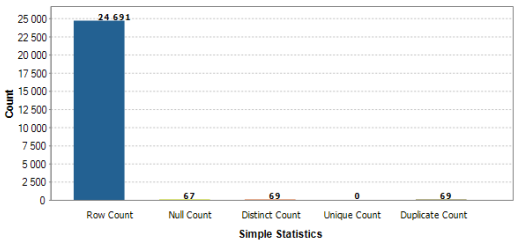
Value	Count	%
99.9	17098	69.25%
Empty field	5051	20.46%
99	1903	7.71%
9.9	525	2.13%
9	113	0.46%
999	1	4.05E-3%



Column: metadata.Year

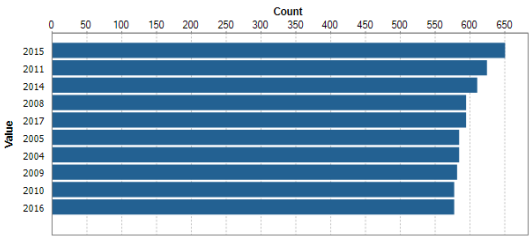
Simple Statistics

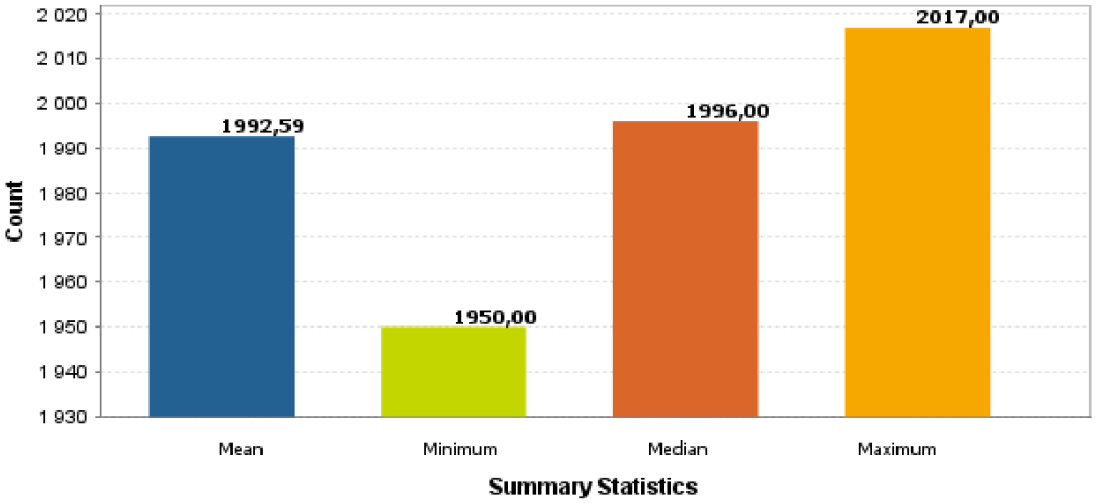
Label	Count	%
Row Count	24691	100.00%
Null Count	67	0.27%
Distinct Count	69	0.28%
Unique Count	0	0.00%
Duplicate Count	69	0.28%



Value Frequency

Value	Count	%
2015	651	2.64%
2011	625	2.53%
2014	611	2.47%
2008	595	2.41%
2017	595	2.41%
2005	585	2.37%
2004	585	2.37%
2009	582	2.36%
2010	578	2.34%
2016	578	2.34%





2.6.2 Análise ao DataSet da Velocidade e Distância

Este dataset foi retirado através de um serviço que ajuda a fazer web scraping, mas como referimos anteriormente apenas conseguimos obter estes dados época a época.

Por isso, optamos por não sobrecarregar este documento word com centenas de prints e colocamos apenas as representativas da última época. Verificamos todos os dados que necessitaremos e não identificamos qualquer erro.

Quality Analysis:

- Pattern Frequency
- Mean
- Median

- Range
- Minimun
- Maximum
- Value Frequency
- Null Count
- Distinct Count
- Unique Count
- Duplicate Count
- Row Count

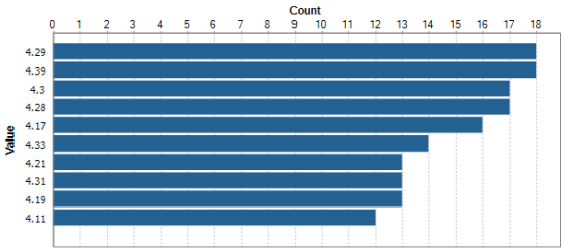
Erros de Qualidade:

- Nenhum erro detetado confirma a fiabilidade dos dados retirados do site oficial da NBA.

Column: metadata.AVG_SPEED

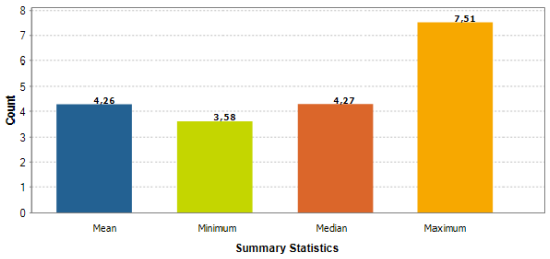
Value Frequency

Value	Count	%
4.29	18	3.33%
4.39	18	3.33%
4.3	17	3.15%
4.28	17	3.15%
4.17	16	2.96%
4.33	14	2.59%
4.21	13	2.41%
4.31	13	2.41%
4.19	13	2.41%
4.11	12	2.22%



Summary Statistics

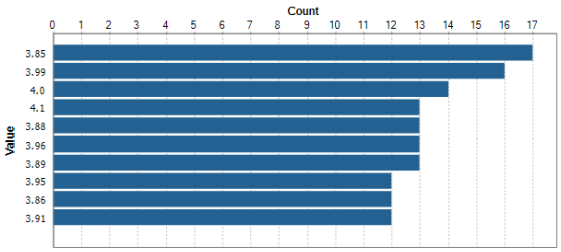
Label	Value
Mean	4.2561481481481485
Median	4.27
Range	3.9299999999999997
Minimum	3.58
Maximum	7.51



Column: metadata.AVG_SPEED_DEF

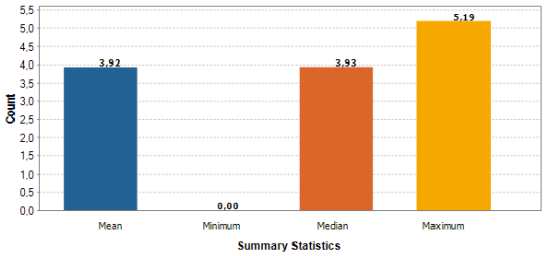
Value Frequency

Value	Count	%
3.85	17	3.15%
3.99	16	2.96%
4.0	14	2.59%
4.1	13	2.41%
3.88	13	2.41%
3.96	13	2.41%
3.89	13	2.41%
3.95	12	2.22%
3.86	12	2.22%
3.91	12	2.22%



Summary Statistics

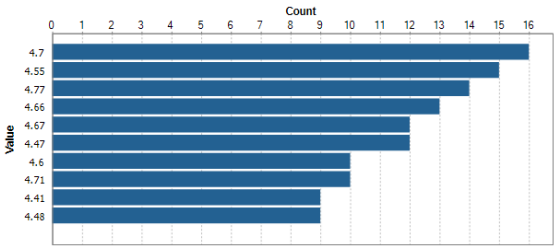
Label	Value
Mean	3.9225000000000003
Median	3.93
Range	5.19
Minimum	0.0
Maximum	5.19



Column: metadata.AVG_SPEED_OFF

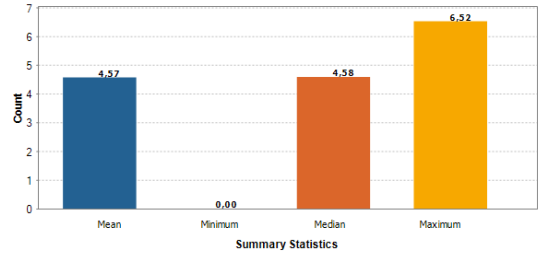
Value Frequency

Value	Count	%
4.7	16	2.96%
4.55	15	2.78%
4.77	14	2.59%
4.66	13	2.41%
4.67	12	2.22%
4.47	12	2.22%
4.6	10	1.85%
4.71	10	1.85%
4.41	9	1.67%
4.48	9	1.67%



Summary Statistics

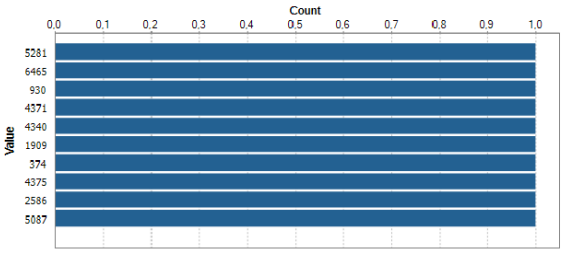
Label	Value
Mean	4.565018518518519
Median	4.58
Range	6.52
Minimum	0.0
Maximum	6.52



Column: metadata.DIST._FEET

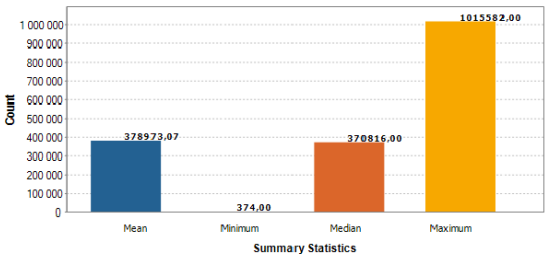
Value Frequency

Value	Count	%
5281	1	0.19%
6465	1	0.19%
930	1	0.19%
4371	1	0.19%
4340	1	0.19%
1909	1	0.19%
374	1	0.19%
4375	1	0.19%
2586	1	0.19%
5087	1	0.19%



Summary Statistics

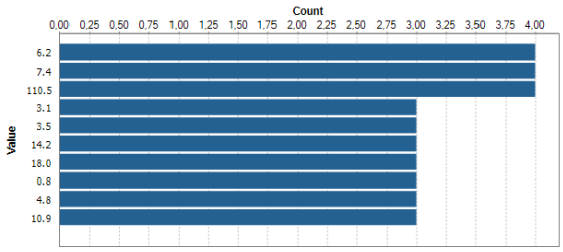
Label	Value
Mean	378973.07037037035
Median	370816.0
Range	1015208.0
Minimum	374
Maximum	1015582



Column: metadata.DIST._MILES

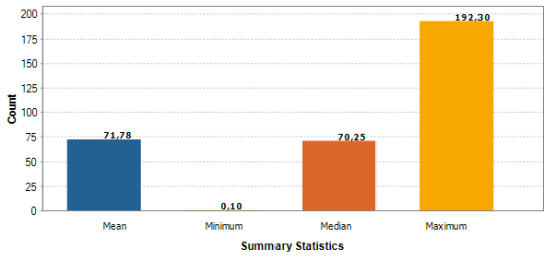
Value Frequency

Value	Count	%
6.2	4	0.74%
7.4	4	0.74%
110.5	4	0.74%
3.1	3	0.56%
3.5	3	0.56%
14.2	3	0.56%
18.0	3	0.56%
0.8	3	0.56%
4.8	3	0.56%
10.9	3	0.56%



Summary Statistics

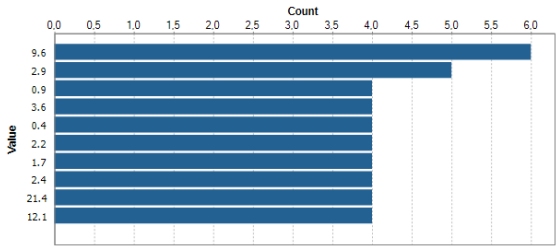
Label	Value
Mean	71.77611111111111
Median	70.25000076293945
Range	192.20000000000002
Minimum	0.1
Maximum	192.3



Column: metadata.DIST._MILES_DEF

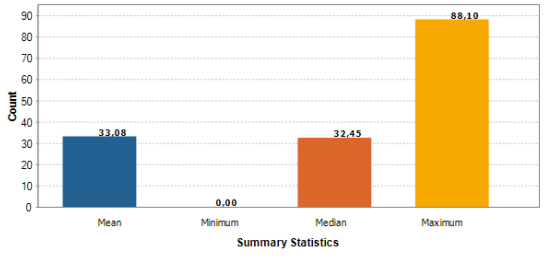
Value Frequency

Value	Count	%
9.6	6	1.11%
2.9	5	0.93%
0.9	4	0.74%
3.6	4	0.74%
0.4	4	0.74%
2.2	4	0.74%
1.7	4	0.74%
2.4	4	0.74%
21.4	4	0.74%
12.1	4	0.74%



Summary Statistics

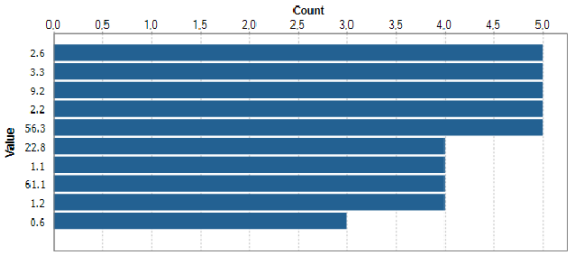
Label	Value
Mean	33.07833333333333
Median	32.45
Range	88.1
Minimum	0.0
Maximum	88.1



Column: metadata.DIST_MILES_OFF

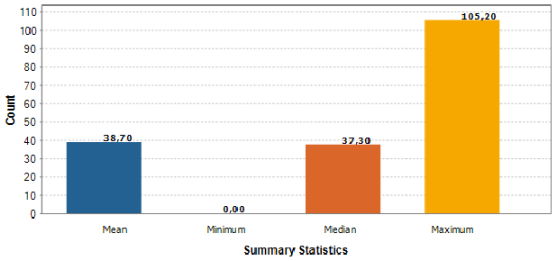
Value Frequency

Value	Count	%
2.6	5	0.93%
3.3	5	0.93%
9.2	5	0.93%
2.2	5	0.93%
56.3	5	0.93%
22.8	4	0.74%
1.1	4	0.74%
61.1	4	0.74%
1.2	4	0.74%
0.6	3	0.56%



Summary Statistics

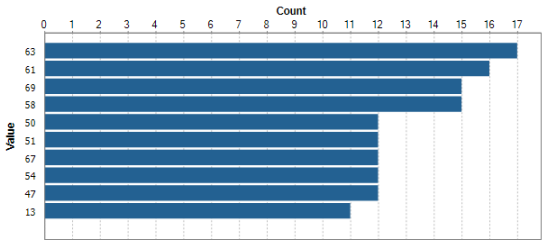
Label	Value
Mean	38.69592592592593
Median	37.3
Range	105.2
Minimum	0.0
Maximum	105.2



Column: metadata.GP

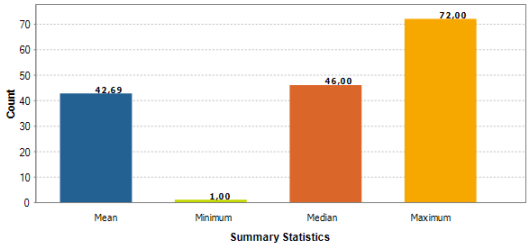
Value Frequency

Value	Count	%
63	17	3.15%
61	16	2.96%
69	15	2.78%
58	15	2.78%
50	12	2.22%
51	12	2.22%
67	12	2.22%
54	12	2.22%
47	12	2.22%
13	11	2.04%



Summary Statistics

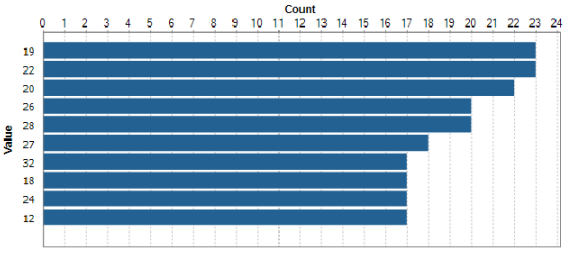
Label	Value
Mean	42.69259259259259
Median	46.0
Range	71.0
Minimum	1
Maximum	72



Column: metadata.I

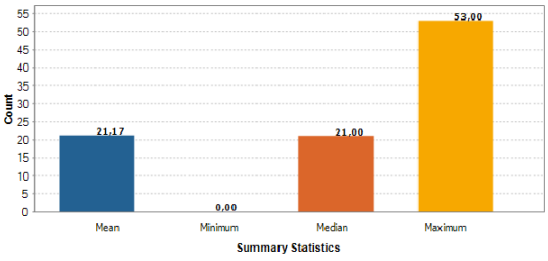
Value Frequency

Value	Count	%
19	23	4.26%
22	23	4.26%
20	22	4.07%
26	20	3.70%
28	20	3.70%
27	18	3.33%
32	17	3.15%
18	17	3.15%
24	17	3.15%
12	17	3.15%



Summary Statistics

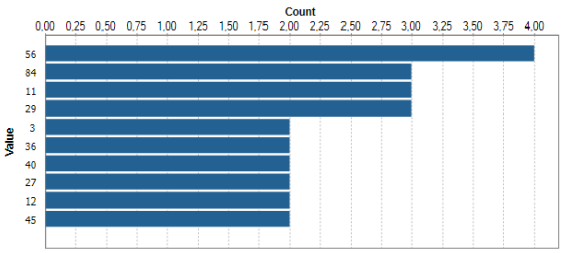
Label	Value
Mean	21.17037037037037
Median	21.0
Range	53.0
Minimum	0
Maximum	53



Column: metadata.MIN

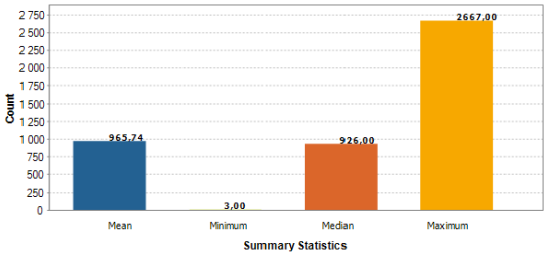
Value Frequency

Value	Count	%
56	4	0.74%
84	3	0.56%
11	3	0.56%
29	3	0.56%
3	2	0.37%
36	2	0.37%
40	2	0.37%
27	2	0.37%
12	2	0.37%
45	2	0.37%



Summary Statistics

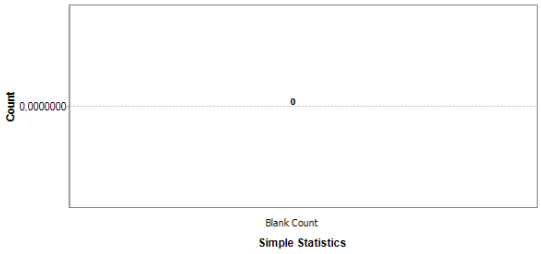
Label	Value
Mean	965.7407407407408
Median	926.0
Range	2664.0
Minimum	3
Maximum	2667



Column: metadata.Player

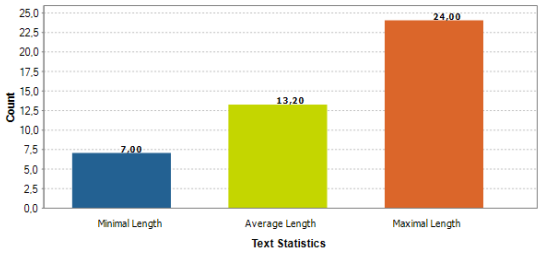
Simple Statistics

Label	Count	%
Blank Count	0	0.00%



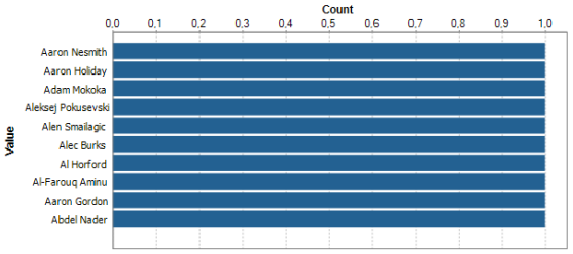
Text Statistics

Label	Value
Minimal Length	7.00
Average Length	13.20
Maximal Length	24.00



Value Frequency

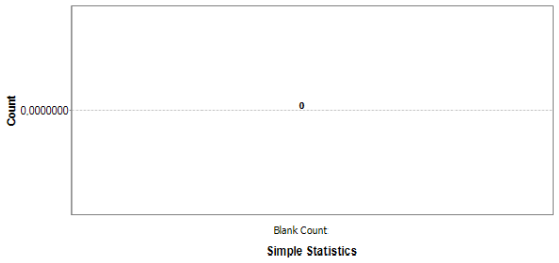
Value	Count	%
Aaron Nesmith	1	0.19%
Aaron Holiday	1	0.19%
Adam Mokoka	1	0.19%
Aleksej Pokusevski	1	0.19%
Alen Smailagic	1	0.19%
Alec Burks	1	0.19%
Al Horford	1	0.19%
Al-Farouq Aminu	1	0.19%
Aaron Gordon	1	0.19%
Abdel Nader	1	0.19%



Column: metadata.TEAM

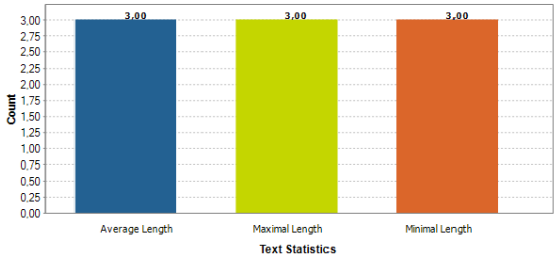
Simple Statistics

Label	Count	%
Blank Count	0	0.00%



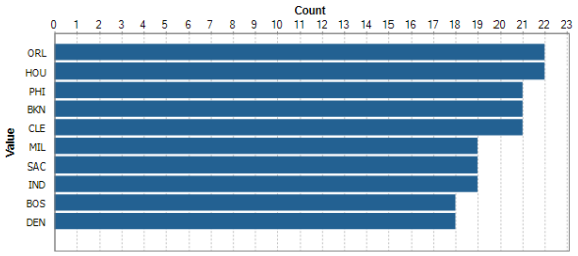
Text Statistics

Label	Value
Average Length	3.00
Maximal Length	3.00
Minimal Length	3.00



Value Frequency

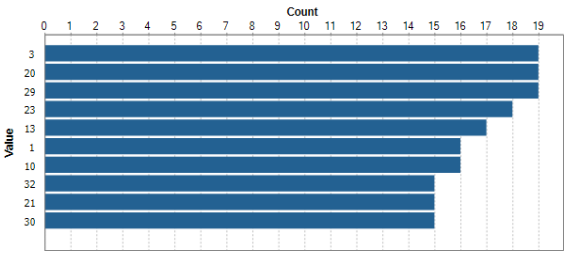
Value	Count	%
ORL	22	4.07%
HOU	22	4.07%
PHI	21	3.89%
BKN	21	3.89%
CLE	21	3.89%
MIL	19	3.52%
SAC	19	3.52%
IND	19	3.52%
BOS	18	3.33%
DEN	18	3.33%



Column: metadata.W

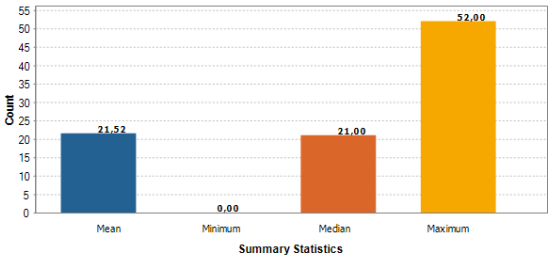
Value Frequency

Value	Count	%
3	19	3.52%
20	19	3.52%
29	19	3.52%
23	18	3.33%
13	17	3.15%
1	16	2.96%
10	16	2.96%
32	15	2.78%
21	15	2.78%
30	15	2.78%



Summary Statistics

Label	Value
Mean	21.522222222222222
Median	21.0
Range	52.0
Minimum	0
Maximum	52



2.6.3 Análise ao DataSet da Altura e Peso

Através deste dataset conseguimos garantir a altura e peso de todos os jogadores que precisamos para fazer as nossas análises.

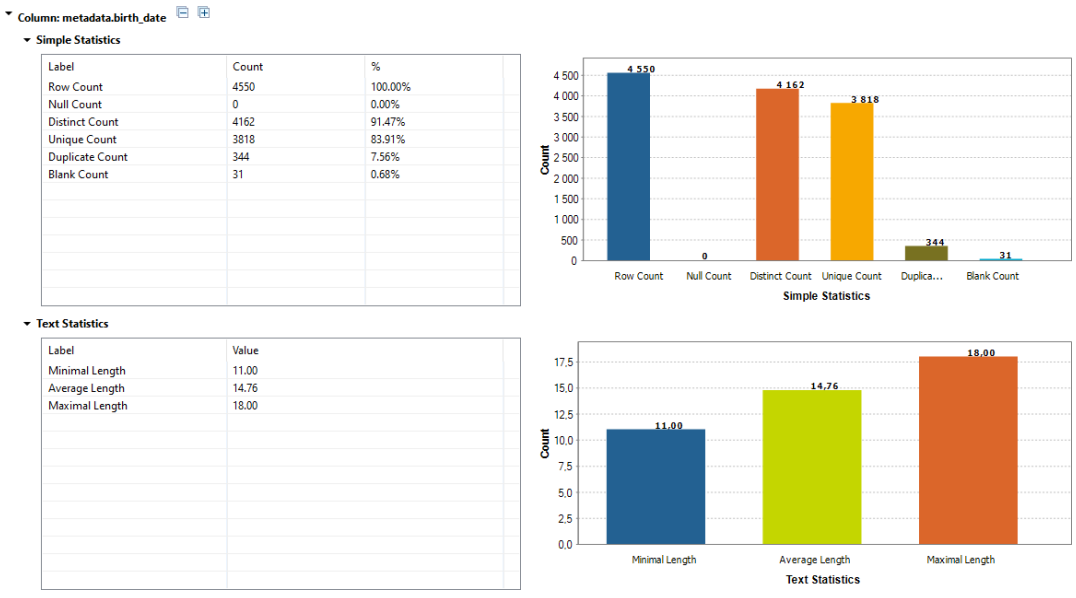
Qualquer erro detetado através desta fase será corrigido na transformação do processo ELT.

Quality Analysis:

- Pattern Frequency
- Mean
- Median
- Range
- Minimun
- Maximum
- Value Frequency
- Null Count
- Distinct Count
- Unique Count
- Duplicate Count
- Row Count

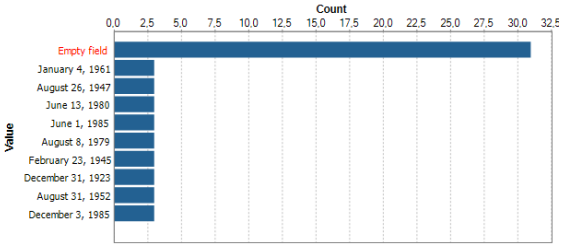
Erros de Qualidade:

- Existem alguns dados em Blank/Null mas como estes estão em número muito mínimo. Além do mais os únicos dados que nos interessam deste dataset são altura e peso e esses erros já estão identificados. Trata-se de um total de 6 jogadores com dados de peso ou altura em branco, no intervalo de 1950-1978.



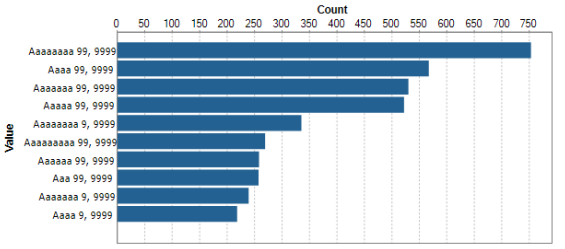
Value Frequency

Value	Count	%
Empty field	31	0.68%
January 4, 1961	3	0.07%
August 26, 1947	3	0.07%
June 13, 1980	3	0.07%
June 1, 1985	3	0.07%
August 8, 1979	3	0.07%
February 23, 1945	3	0.07%
December 31, 1923	3	0.07%
August 31, 1952	3	0.07%
December 3, 1985	3	0.07%



Pattern Frequency

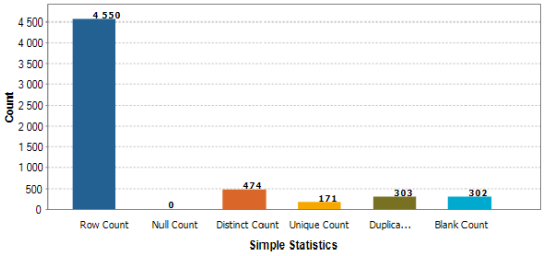
Value	Count	%
Aaaaaaaa 99, 9999	754	16.57%
Aaaa 99, 9999	568	12.48%
Aaaaaaa 99, 9999	531	11.67%
Aaaaa 99, 9999	523	11.49%
Aaaaaaaa 9, 9999	336	7.38%
Aaaaaaaaa 99, 9999	270	5.93%
Aaaaaa 99, 9999	259	5.69%
Aaa 99, 9999	258	5.67%
Aaaaaaa 9, 9999	240	5.27%
Aaaa 9, 9999	219	4.81%



Column: metadata.college

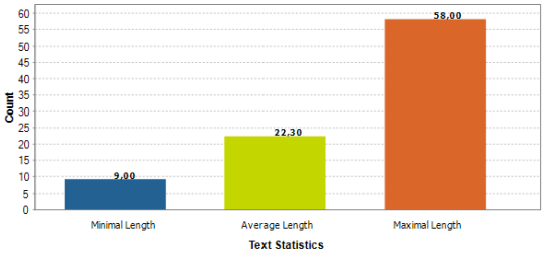
Simple Statistics

Label	Count	%
Row Count	4550	100.00%
Null Count	0	0.00%
Distinct Count	474	10.42%
Unique Count	171	3.76%
Duplicate Count	303	6.66%
Blank Count	302	6.64%



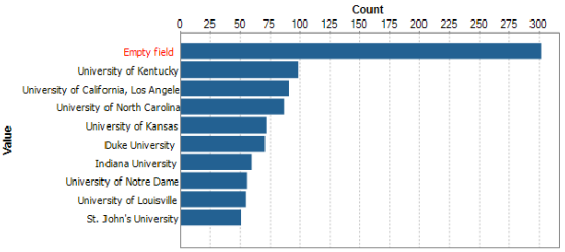
Text Statistics

Label	Value
Minimal Length	9.00
Average Length	22.30
Maximal Length	58.00



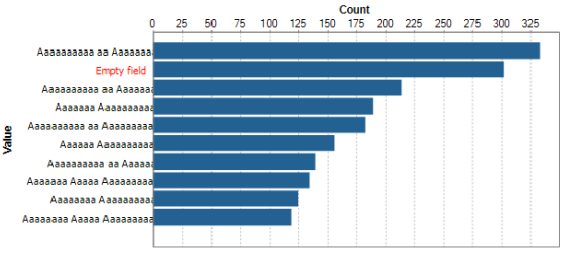
Value Frequency

Value	Count	%
Empty field	302	6.64%
University of Kentucky	99	2.18%
University of California, Los Angeles	91	2.00%
University of North Carolina	87	1.91%
University of Kansas	72	1.58%
Duke University	71	1.56%
Indiana University	60	1.32%
University of Notre Dame	56	1.23%
University of Louisville	55	1.21%
St. John's University	51	1.12%



Pattern Frequency

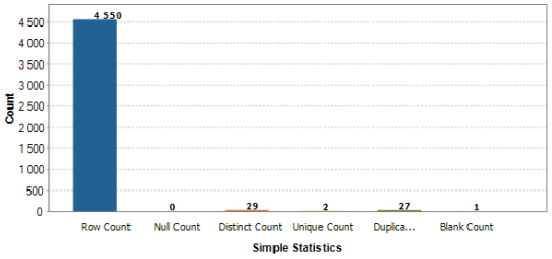
Value	Count	%
Aaaaaaaaa aa Aaaaaaa	333	7.32%
Empty field	302	6.64%
Aaaaaaaaaa aa Aaaaaaa	214	4.70%
Aaaaaaa Aaaaaaaaaa	189	4.15%
Aaaaaaaaaa aa Aaaaaaaaaa	183	4.02%
Aaaaaa Aaaaaaaaaa	156	3.43%
Aaaaaaaaaa aa Aaaaaa	140	3.08%
Aaaaaaa Aaaaa Aaaaaaaaaa	135	2.97%
Aaaaaaaa Aaaaaaaaaa	125	2.75%
Aaaaaaaa Aaaaa Aaaaaaaaaa	119	2.62%



Column: metadata.height

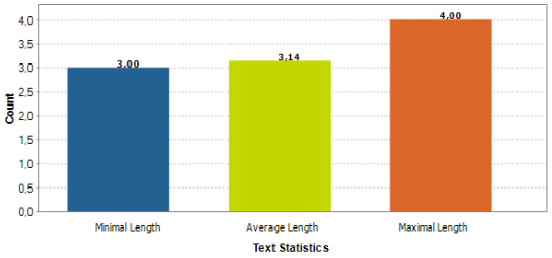
Simple Statistics

Label	Count	%
Row Count	4550	100.00%
Null Count	0	0.00%
Distinct Count	29	0.64%
Unique Count	2	0.04%
Duplicate Count	27	0.59%
Blank Count	1	0.02%



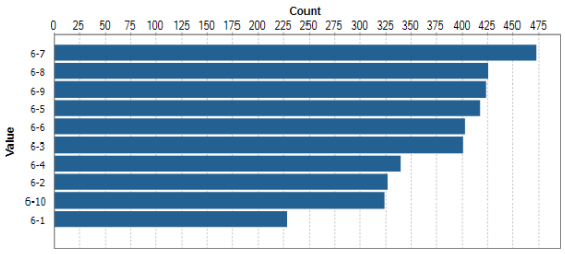
Text Statistics

Label	Value
Minimal Length	3.00
Average Length	3.14
Maximal Length	4.00



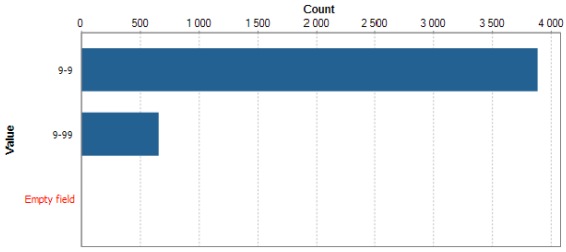
Value Frequency

Value	Count	%
6-7	473	10.40%
6-8	426	9.36%
6-9	424	9.32%
6-5	418	9.19%
6-6	403	8.86%
6-3	401	8.81%
6-4	340	7.47%
6-2	327	7.19%
6-10	324	7.12%
6-1	229	5.03%



Pattern Frequency

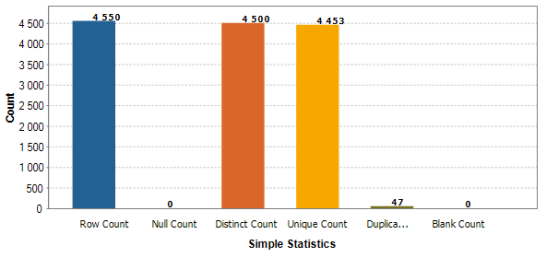
Value	Count	%
9-9	3887	85.43%
9-99	662	14.55%
Empty field	1	0.02%



Column: metadata.name

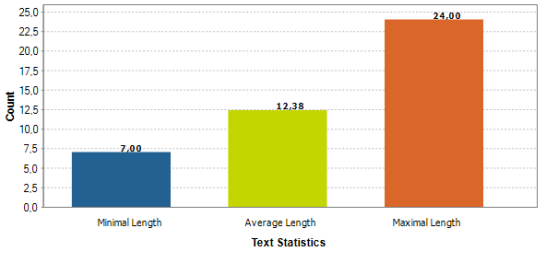
Simple Statistics

Label	Count	%
Row Count	4550	100.00%
Null Count	0	0.00%
Distinct Count	4500	98.90%
Unique Count	4453	97.87%
Duplicate Count	47	1.03%
Blank Count	0	0.00%



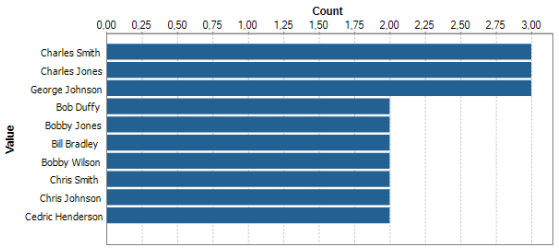
Text Statistics

Label	Value
Minimal Length	7.00
Average Length	12.38
Maximal Length	24.00



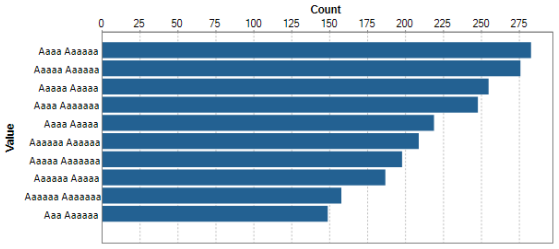
Value Frequency

Value	Count	%
Charles Smith	3	0.07%
Charles Jones	3	0.07%
George Johnson	3	0.07%
Bob Duffy	2	0.04%
Bobby Jones	2	0.04%
Bill Bradley	2	0.04%
Bobby Wilson	2	0.04%
Chris Smith	2	0.04%
Chris Johnson	2	0.04%
Cedric Henderson	2	0.04%



Pattern Frequency

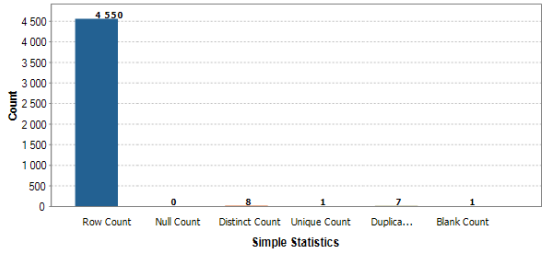
Value	Count	%
Aaaa Aaaaaa	283	6.22%
Aaaaa Aaaaaa	276	6.07%
Aaaaa Aaaaa	255	5.60%
Aaaa Aaaaaaa	248	5.45%
Aaaa Aaaaa	219	4.81%
Aaaaaa Aaaaaa	209	4.59%
Aaaaa Aaaaaaa	198	4.35%
Aaaaaa Aaaaa	187	4.11%
Aaaaaa Aaaaaaa	158	3.47%
Aaa Aaaaaa	149	3.27%



Column: metadata.position

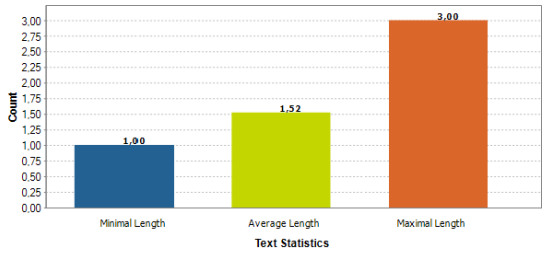
Simple Statistics

Label	Count	%
Row Count	4550	100.00%
Null Count	0	0.00%
Distinct Count	8	0.18%
Unique Count	1	0.02%
Duplicate Count	7	0.15%
Blank Count	1	0.02%



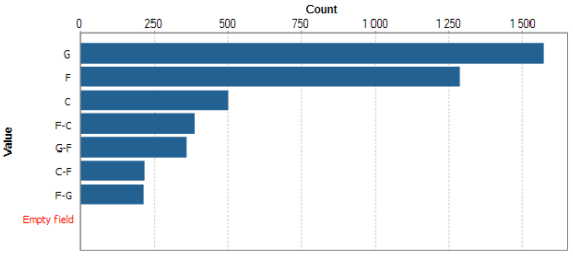
Text Statistics

Label	Value
Minimal Length	1.00
Average Length	1.52
Maximal Length	3.00



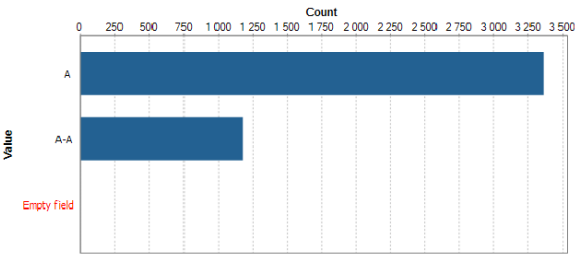
Value Frequency

Value	Count	%
G	1574	34.59%
F	1290	28.35%
C	502	11.03%
F-C	388	8.53%
G-F	360	7.91%
C-F	219	4.81%
F-G	216	4.75%
Empty field	1	0.02%



Pattern Frequency

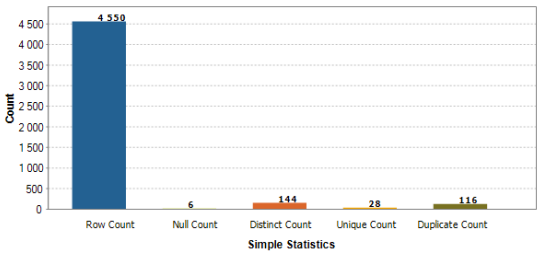
Value	Count	%
A	3366	73.98%
A-A	1183	26.00%
Empty field	1	0.02%



Column: metadata.weight

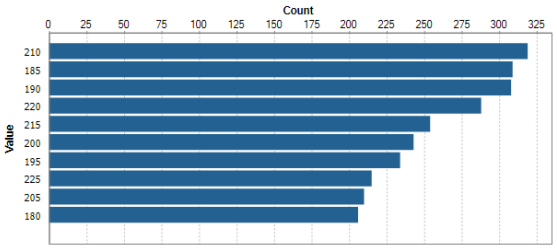
Simple Statistics

Label	Count	%
Row Count	4550	100.00%
Null Count	6	0.13%
Distinct Count	144	3.16%
Unique Count	28	0.62%
Duplicate Count	116	2.55%



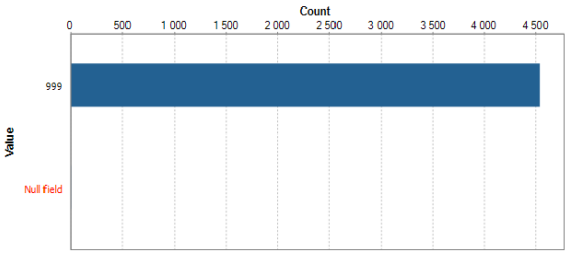
Value Frequency

Value	Count	%
210	319	7.01%
185	309	6.79%
190	308	6.77%
220	288	6.33%
215	254	5.58%
200	243	5.34%
195	234	5.14%
225	215	4.73%
205	210	4.62%
180	206	4.53%



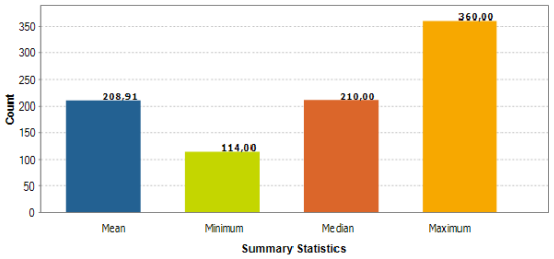
Pattern Frequency

Value	Count	%
999	4544	99.87%
Null field	6	0.13%



Summary Statistics

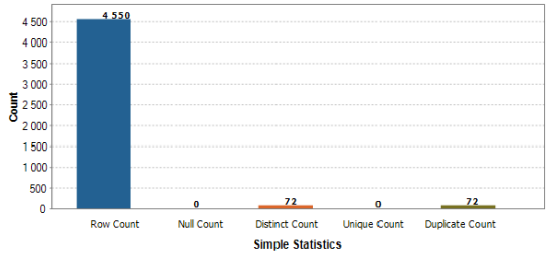
Label	Value
Mean	208.9080105633803
Median	210.0
Range	246.0
Minimum	114
Maximum	360



Column: metadata.year_end

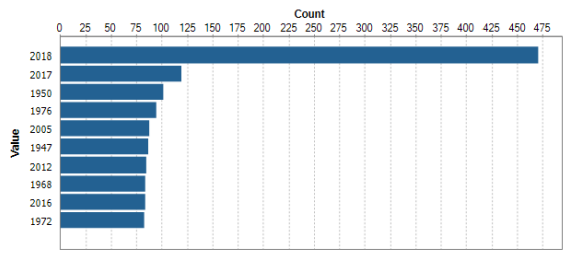
Simple Statistics

Label	Count	%
Row Count	4550	100.00%
Null Count	0	0.00%
Distinct Count	72	1.58%
Unique Count	0	0.00%
Duplicate Count	72	1.58%



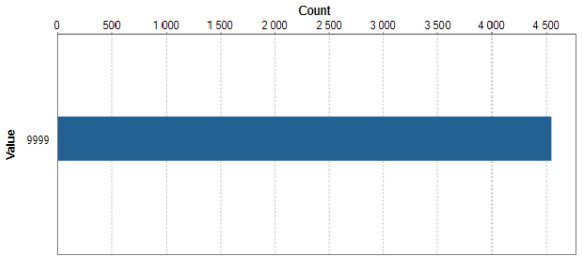
Value Frequency

Value	Count	%
2018	471	10.35%
2017	119	2.62%
1950	102	2.24%
1976	95	2.09%
2005	88	1.93%
1947	87	1.91%
2012	85	1.87%
1968	84	1.85%
2016	84	1.85%
1972	83	1.82%



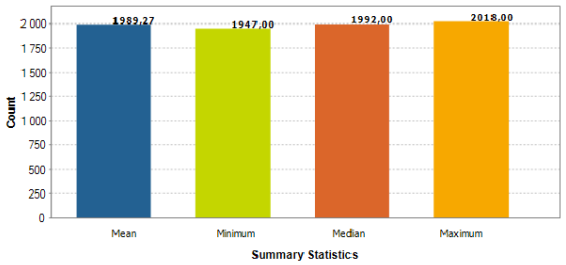
Pattern Frequency

Value	Count	%
9999	4550	100.00%



Summary Statistics

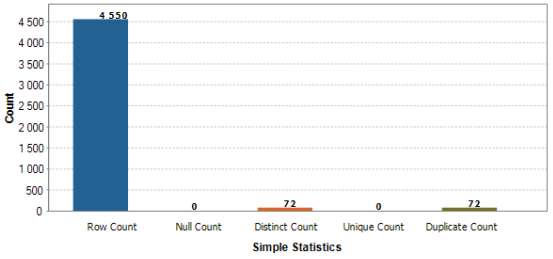
Label	Value
Mean	1989.2725274725274
Median	1992.0
Range	71.0
Minimum	1947
Maximum	2018



Column: metadata.year_start

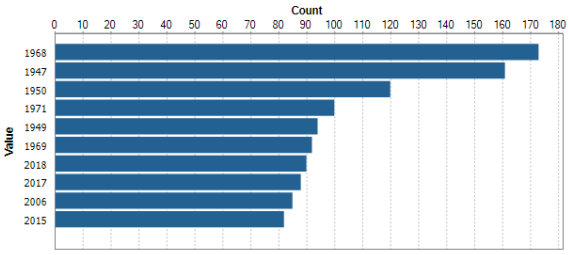
Simple Statistics

Label	Count	%
Row Count	4550	100.00%
Null Count	0	0.00%
Distinct Count	72	1.58%
Unique Count	0	0.00%
Duplicate Count	72	1.58%



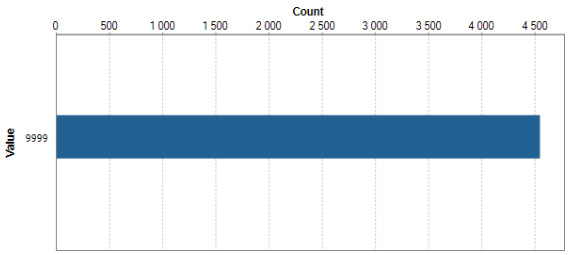
Value Frequency

Value	Count	%
1968	173	3.80%
1947	161	3.54%
1950	120	2.64%
1971	100	2.20%
1949	94	2.07%
1969	92	2.02%
2018	90	1.98%
2017	88	1.93%
2006	85	1.87%
2015	82	1.80%



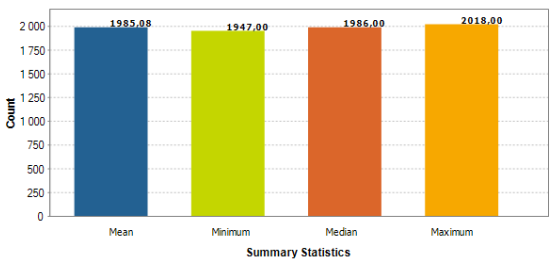
Pattern Frequency

Value	Count	%
9999	4550	100.00%

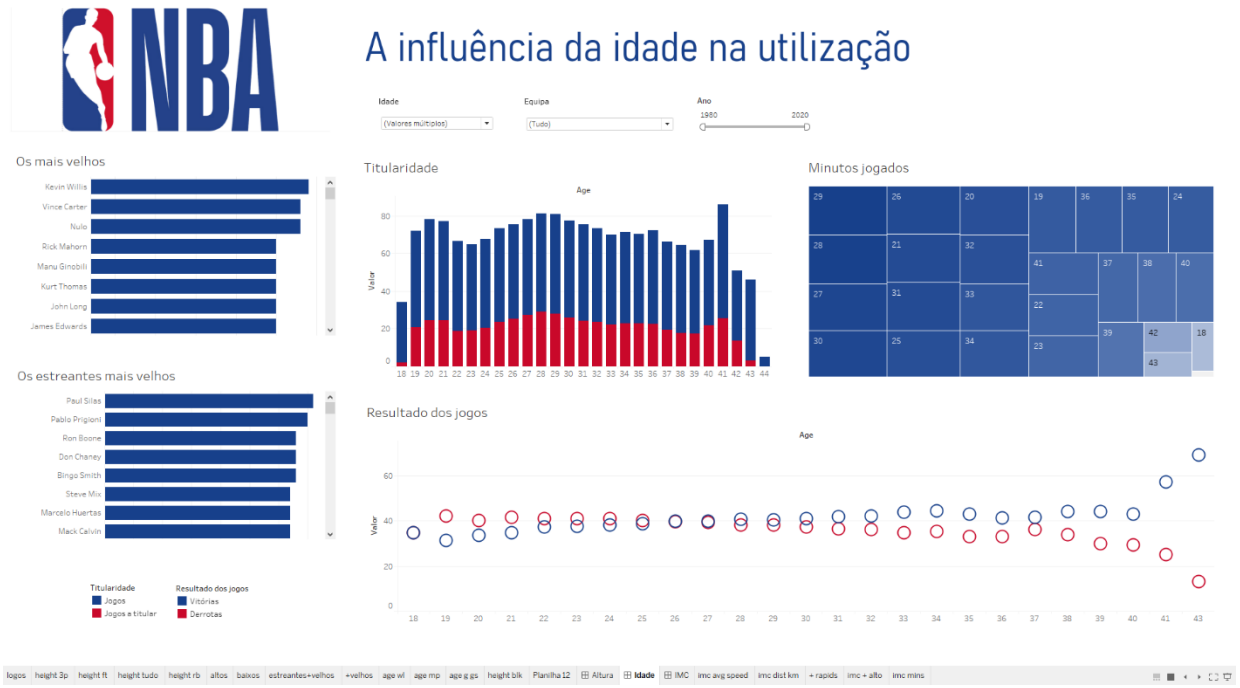


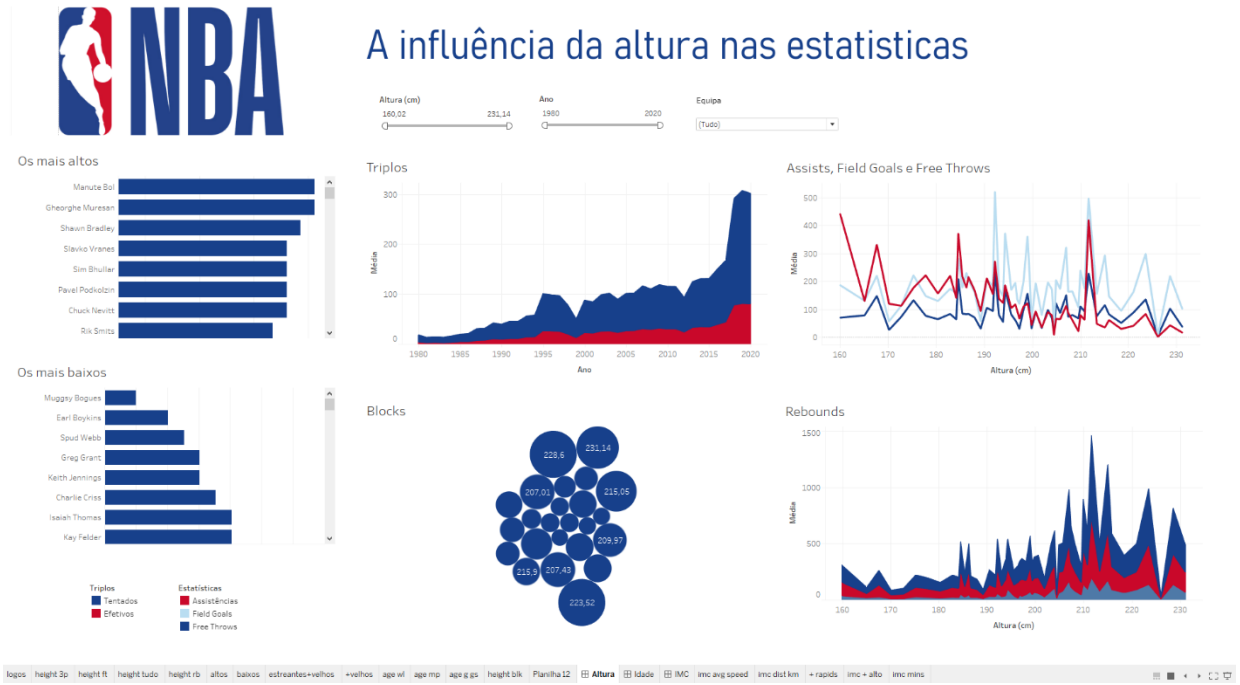
Summary Statistics

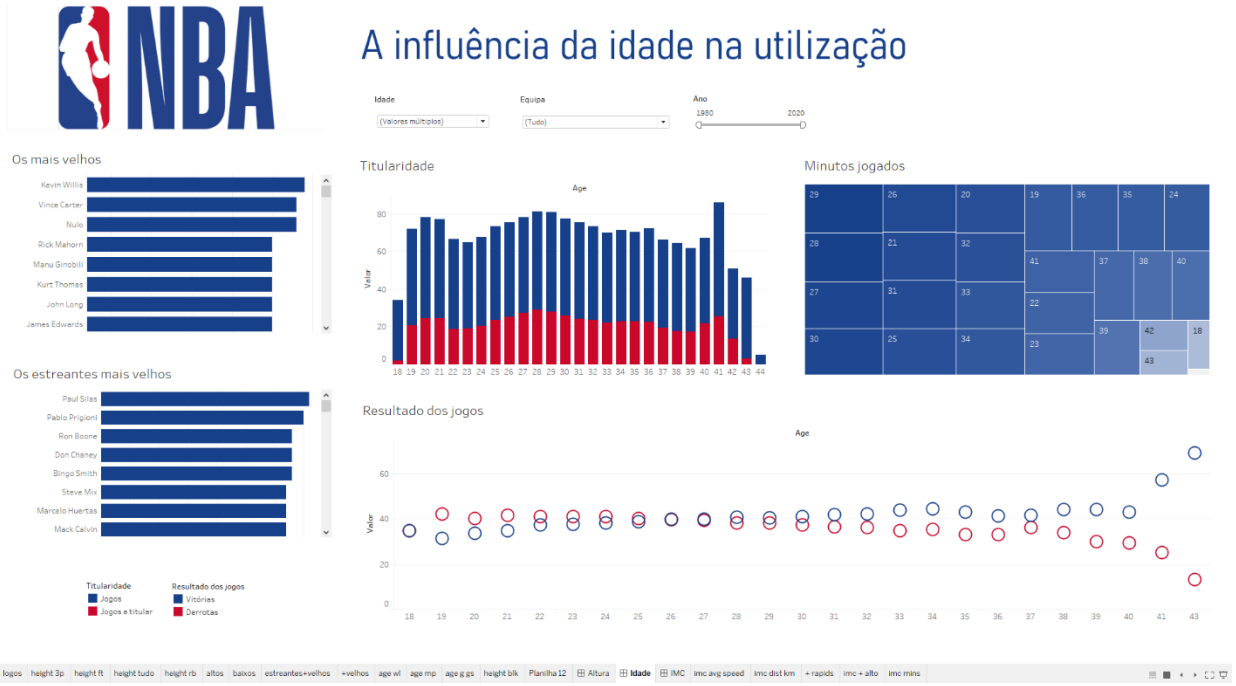
Label	Value
Mean	1985.0762637362636
Median	1986.0
Range	71.0
Minimum	1947
Maximum	2018



2.7 Dashboard:






A influência da idade na utilização
Idade
(Valores múltiplos)
Equipa
(Tudo)
Ano
1980
2020
Os mais velhos
Kevin Willis
Vince Carter
Nuls
Rick Mahorn
Manu Ginobili
Kurt Thomas
John Long
James Edwards
Titularidade
Age
Valor
Minutos jogados
Os estreantes mais velhos
Paul Silas
Pablo Prigioni
Ron Boone
Don Chaney
Bingo Smith
Steve Mix
Marcelo Huertas
Mack Calvin
Titularidade
Jogos
Jogos a titular
Resultado dos jogos
Vitórias
Derrotas
Resultado dos jogos
Age
Valor

3. SUBGRUPO 3

Use Case: Impacto do Draft na NBA

Duarte Brandão

A92938 - 3º ano LEGSI

a92938@alunos.uminho.pt

Pedro Gonçalves

A92930 - 3º ano LEGSI

a92930@alunos.uminho.pt

Tiago Lopes

A94123 - 3º ano LEGSI

a94123@alunos.uminho.pt

3.1 Introdução

Baseando o nosso trabalho no tema inicial (NBA) os elementos deste use case, dedicaram os seus esforços a analisar o draft deste desporto.

Numa primeira fase, o draft consiste numa seleção anual de jogadores por parte de todas as equipas presentes na NBA. Atualmente este evento é composto por duas rondas onde são escolhidos 30 jogadores em cada etapa. Tanto a quantidade de rondas como o número de selecionados sofreram alterações ao longo dos anos. Inicialmente as equipas selecionavam todos os jogadores disponíveis, chegando mesmo as 21 rondas (ano de 1960 e 1968). Posteriormente foram diminuindo a quantidade de rondas tendo chegado a apenas duas mantendo-se atualmente. Esta decisão foi tomada para que os jogadores escolhidos tivessem mais oportunidades de jogo.

É importante ressalvar que neste use case houve uma abordagem diferente daquela existente nos outros dois. Ou seja, cada elemento do grupo se baseou em "zonas" e influencias diferentes do draft podendo desta forma tratar este use case de forma mais aprofundada e correta.

3.2 KPI's

Da mesma forma que existiu uma abordagem diferente ao longo do use case cada elemento também definiu KPI específicos. Assim os KPI utilizados foram a percentagem de minutos que os rookies jogaram nos seus primeiros anos, o ano do draft, picks no draft, nº títulos, nº runner ups, nº finais.

3.3 Analytical Questions

A melhor universidade do draft.

A média de idades do draft.

Os melhores jogadores do draft

Quantas foram as primeiras picks que jogaram e que não jogaram?

O melhor ano do draft?

Quais são as posições que após o draft têm melhor adaptabilidade?

Qual é a eficiência por pick?

As melhores picks têm mais liberdade para atirar?

Que equipas dão mais oportunidades aos rookies?

O número de picks nos drafts foi variando ou manteve se constante ao longo dos anos?

As primeiras 10 picks no draft são os jogadores que obtiveram mais prémios de All star durante a sua carreira?

A quantidade de jogadores que foram all stars é maior nas picks acima da 10ª posição?

A media dos minutos por jogo dos jogadores com a pick abaixo da 10ª posição é maior do que nos jogadores com a pick acima da 10ª posição?

A media dos pontos por jogo dos jogadores com a pick abaixo da 10ª posição é maior do que nos jogadores com a pick acima da 10ª posição?

As equipas que sairam beneficiadas no sorteio do draft são as que têm mais títulos?

As equipas que selecionaram mais picks que se tornariam all-stars são as que têm mais títulos?

3.4 DataSets

Os 7 datasets a utilizados foram os seguintes:

Todas as equipas vencedoras da National Basketball League: https://www.basketball-reference.com/playoffs/

Informações sobre todos os jogadores que jogaram, todos aqueles que foram escolhidos nas primeiras 14 picks e ainda o aniversario e data de nascimento de cada um deles

https://www.kaggle.com/skandasastry/nba-lottery-picks-from-1995-2020?select=birthdays-bbref-scraped.csv

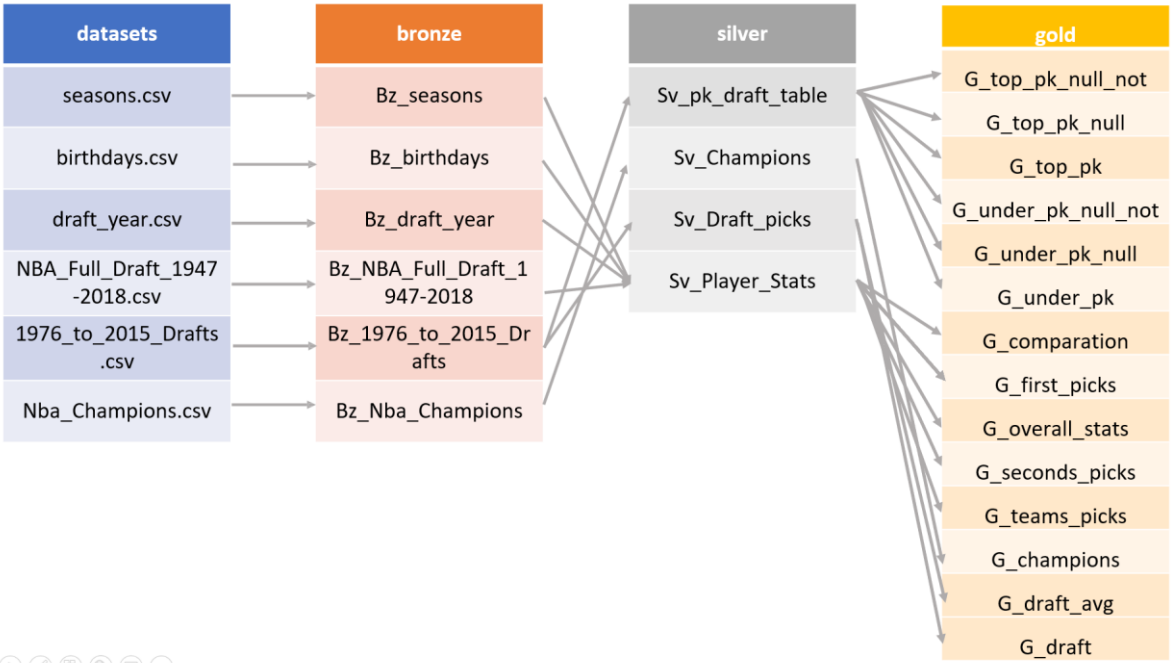
Os drafts:

https://www.kaggle.com/hrfang1995/nba-drafts-of-19472018

Estatísticas dos jogadores em todas as épocas da NBA:

https://datasetsearch.research.google.com/search?query=1976%20to%202015%20nba%20draft%20data&docid=L2cvMTFqbnlxcnJ2ag%3D%3D

3.5 DataSet to Bronze to Silver to Gold



3.6 Data Quality

3.6.1 Análise ao DataSet de Todas as equipas vencedoras da National Basketball League:

Este dataset será utilizado para retirar todos os dados necessários relativamente aos vencedores do NBA championship. Verificamos todos os dados que necessitaremos e não identificamos qualquer erro.

Quality Analysis:

- Row Count
- Null Count
- Distinct Count
- Unique Count
- Duplicate Count
- Blank Count

Erros de Qualidade:

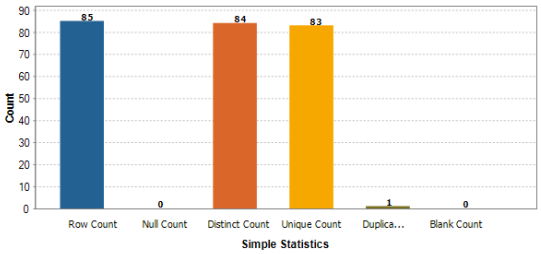
- Nenhum erro detetado confirma a fiabilidade dos dados

retirados do site BasketballReference. Sendo este um dos maiores sites relativos a estatísticas na NBA.

Column: metadata.Playoffs_Top_Performers2

Simple Statistics

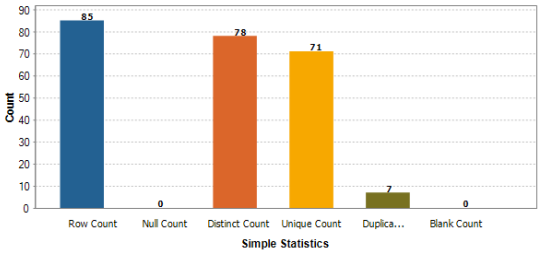
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	84	98.82%
Unique Count	83	97.65%
Duplicate Count	1	1.18%
Blank Count	0	0.00%



Column: metadata.Playoffs_Top_Performers3

Simple Statistics

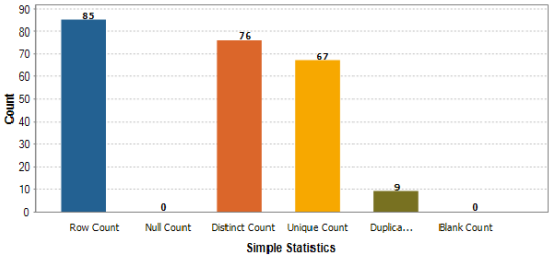
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	78	91.76%
Unique Count	71	83.53%
Duplicate Count	7	8.24%
Blank Count	0	0.00%



Column: metadata.Column0

Simple Statistics

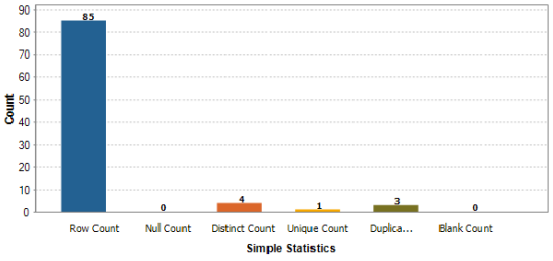
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	76	89.41%
Unique Count	67	78.82%
Duplicate Count	9	10.59%
Blank Count	0	0.00%



Column: metadata.Column1

Simple Statistics

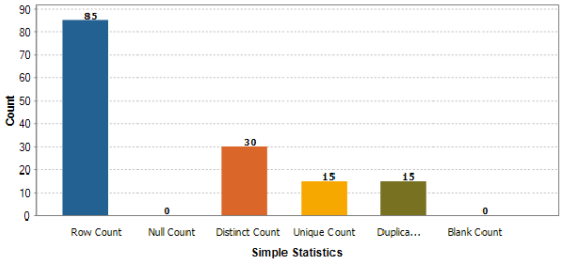
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	4	4.71%
Unique Count	1	1.18%
Duplicate Count	3	3.53%
Blank Count	0	0.00%



Column: metadata.Finals

Simple Statistics

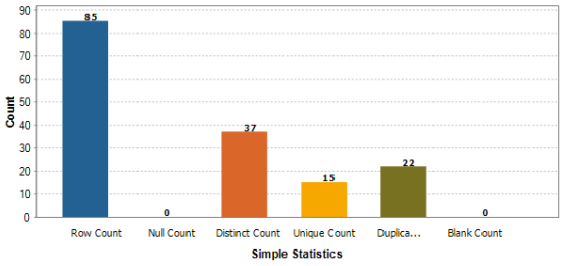
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	30	35.29%
Unique Count	15	17.65%
Duplicate Count	15	17.65%
Blank Count	0	0.00%



Column: metadata.Finals1

Simple Statistics

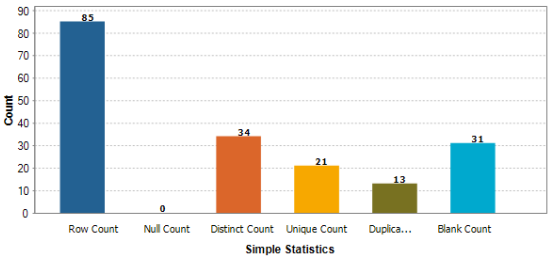
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	37	43.53%
Unique Count	15	17.65%
Duplicate Count	22	25.88%
Blank Count	0	0.00%



Column: metadata.Finals2

Simple Statistics

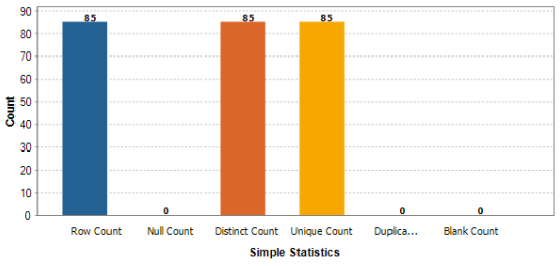
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	34	40.00%
Unique Count	21	24.71%
Duplicate Count	13	15.29%
Blank Count	31	36.47%



Column: metadata.Playoffs_Top_Performers

Simple Statistics

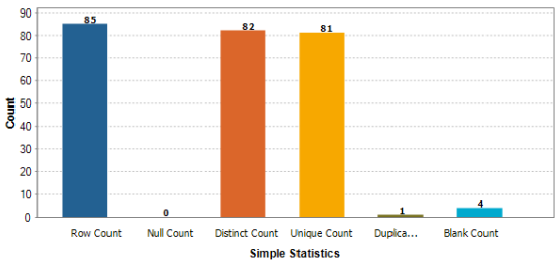
Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	85	100.00%
Unique Count	85	100.00%
Duplicate Count	0	0.00%
Blank Count	0	0.00%



Column: metadata.Playoffs_Top_Performers1

Simple Statistics

Label	Count	%
Row Count	85	100.00%
Null Count	0	0.00%
Distinct Count	82	96.47%
Unique Count	81	95.29%
Duplicate Count	1	1.18%
Blank Count	4	4.71%



3.6.2 Análise ao DataSet dos Drafts:

Este dataset será utilizado para retirar todos os dados necessários relativamente aos jogadores que foram escolhidos em todas as rondas e picks durante todos os anos.

Quality Analysis:

- Row Count
- Null Count
- Distinct Count
- Unique Count
- Duplicate Count
- Blank Count

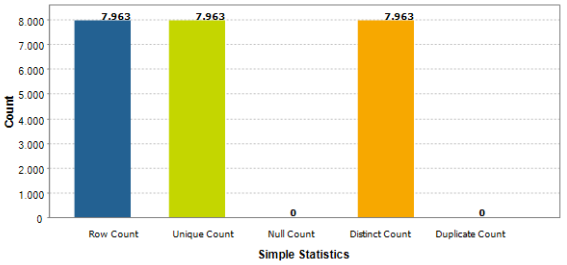
Erros de Qualidade:

- Nenhum erro detetado confirma a fiabilidade dos dados retirados do site BasketballReference. Sendo este um dos maiores sites relativos a estatísticas na NBA.

Column: metadata.Column0

Simple Statistics

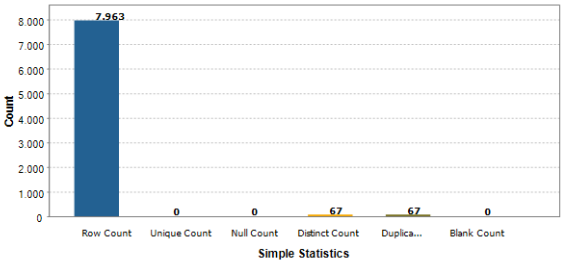
Label	Count	%
Row Count	7963	100.00%
Unique Count	7963	100.00%
Null Count	0	0.00%
Distinct Count	7963	100.00%
Duplicate Count	0	0.00%



Column: metadata.Team

Simple Statistics

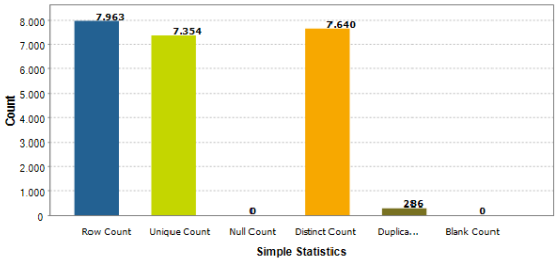
Label	Count	%
Row Count	7963	100.00%
Unique Count	0	0.00%
Null Count	0	0.00%
Distinct Count	67	0.84%
Duplicate Count	67	0.84%
Blank Count	0	0.00%



Column: metadata.Player

Simple Statistics

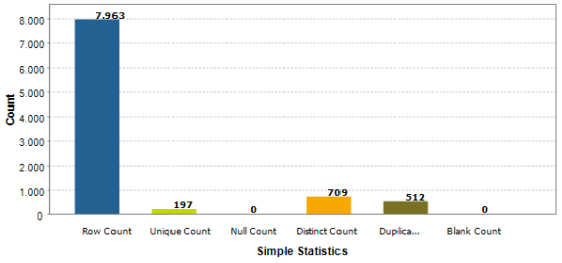
Label	Count	%
Row Count	7963	100.00%
Unique Count	7354	92.35%
Null Count	0	0.00%
Distinct Count	7640	95.94%
Duplicate Count	286	3.59%
Blank Count	0	0.00%



Column: metadata.College

Simple Statistics

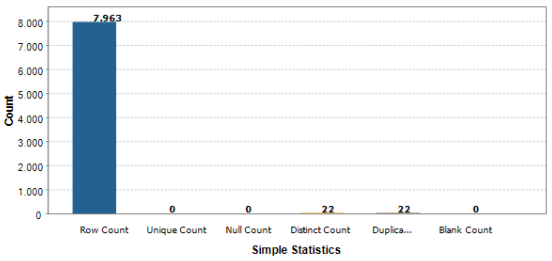
Label	Count	%
Row Count	7963	100.00%
Unique Count	197	2.47%
Null Count	0	0.00%
Distinct Count	709	8.90%
Duplicate Count	512	6.43%
Blank Count	0	0.00%



Column: metadata.Play_Yrs

Simple Statistics

Label	Count	%
Row Count	7963	100.00%
Unique Count	0	0.00%
Null Count	0	0.00%
Distinct Count	22	0.28%
Duplicate Count	22	0.28%
Blank Count	0	0.00%



3.6.3 Análise ao DataSet das Estatísticas dos jogadores em todas as épocas de NBA:

Este será o dataset mais usado, e todas as analitycal queries irão "beber" aqui. Desta forma, todo o ficheiro foi analisado utilizando o talend.

Quality Analysis:

- Row count;
- Null count;
- Distinct count;
- Unique count;
- Duplicate count;
- Blanke count;
- Value Frequency

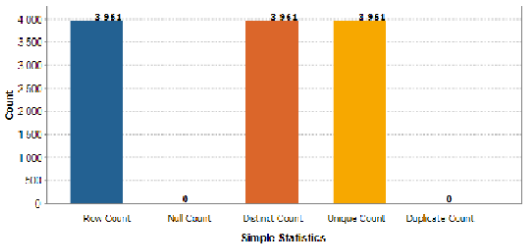
Erros de Qualidade/Observações:

- Numa primeira fase, obtivemos muitos valores “zero”, no entanto após alguma análise e pesquisa chegamos à conclusão de que havia sempre uma justificação para isso. Basicamente existem variadíssimos jogadores que foram escolhidos, no entanto não realizaram qualquer jogo ao longo da sua carreira. Partindo desse raciocínio e possível chegar à conclusão de que todas as colunas na tabela serão afetadas.
- Uma das garantias que tivemos, para que o dataset fosse legitimo, foi o facto de este não conter nenhum null ao longo da análise. Assim, e seguindo o raciocínio das aulas teóricas, parte-se do pressuposto que os dados são verídicos.
- De forma a realizar uma análise dos resultados ainda mais fidedigna analisou-se ainda com mais precisão as colunas que serão necessárias para a execução do trabalho utilizando o indicador Value Frequency.

▾ Column: metadata.Column0

▾ Simple Statistics

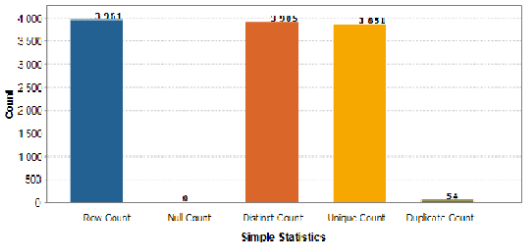
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	3961	100.00%
Unique Count	3961	100.00%
Duplicate Count	0	0.00%



▾ Column: metadata.Player

▾ Simple Statistics

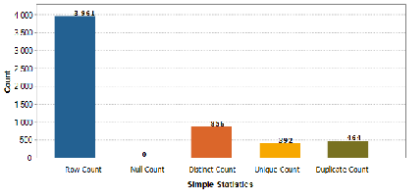
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	3905	98.59%
Unique Count	3851	97.22%
Duplicate Count	54	1.36%



▾ Column: metadata.Games

▾ Simple Statistics

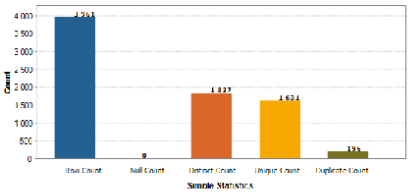
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	856	21.61%
Unique Count	392	9.90%
Duplicate Count	464	11.71%



▾ Column: metadata.Minutes_Played

▾ Simple Statistics

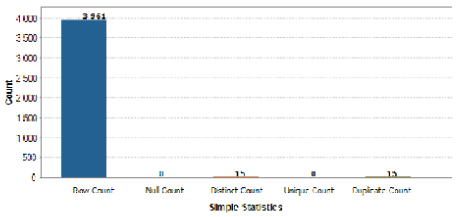
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	1827	46.12%
Unique Count	1631	41.18%
Duplicate Count	196	4.95%



▾ Column: metadata.All_NBA

▾ Simple Statistics

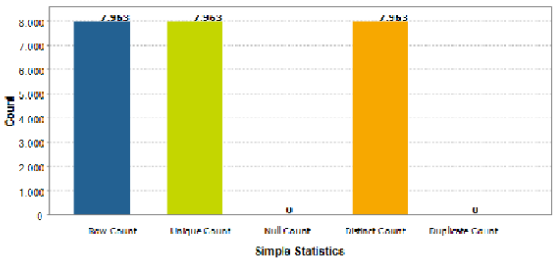
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	13	0.33%
Unique Count	0	0.00%
Duplicate Count	13	0.33%



▾ Column: metadata.Column0

▾ Simple Statistics

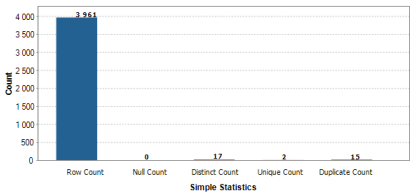
Label	Count	%
Row Count	7963	100.00%
Unique Count	7963	100.00%
Null Count	0	0.00%
Distinct Count	7963	100.00%
Duplicate Count	0	0.00%



Column: metadata.All_Star

Simple Statistics

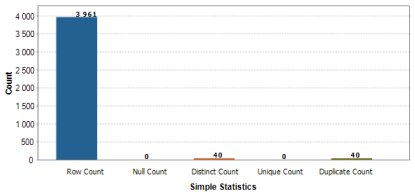
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	17	0.43%
Unique Count	2	0.05%
Duplicate Count	15	0.38%



Column: metadata.Draft_Yr

Simple Statistics

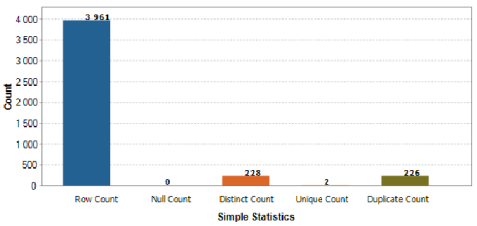
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	40	1.01%
Unique Count	0	0.00%
Duplicate Count	40	1.01%



Column: metadata.Pk

Simple Statistics

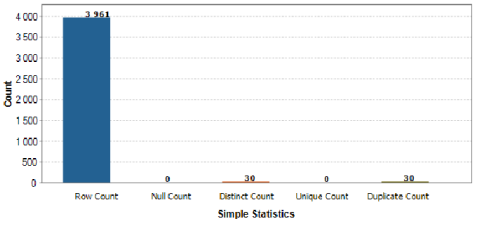
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	228	5.76%
Unique Count	2	0.05%
Duplicate Count	226	5.71%



Column: metadata.Team

Simple Statistics

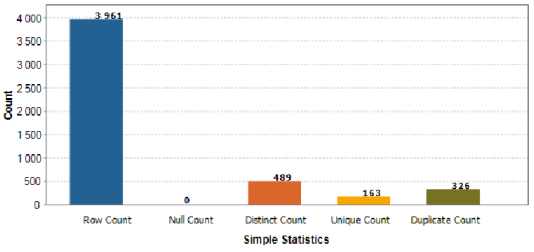
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	30	0.76%
Unique Count	0	0.00%
Duplicate Count	30	0.76%



Column: metadata.College

Simple Statistics

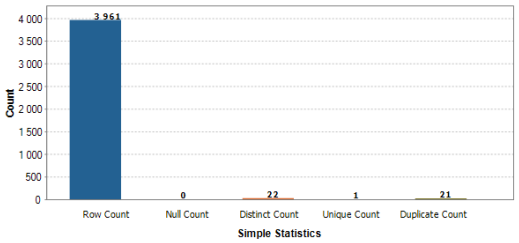
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	489	12.35%
Unique Count	163	4.12%
Duplicate Count	326	8.23%



Column: metadata.Yrs

Simple Statistics

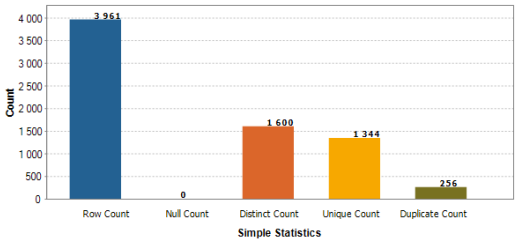
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	22	0.56%
Unique Count	1	0.03%
Duplicate Count	21	0.53%



Column: metadata.PTS

Simple Statistics

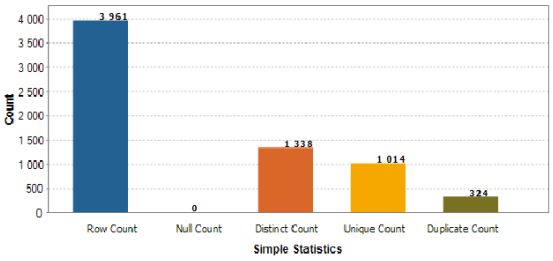
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	1600	40.39%
Unique Count	1344	33.93%
Duplicate Count	256	6.46%



Column: metadata.TRB

Simple Statistics

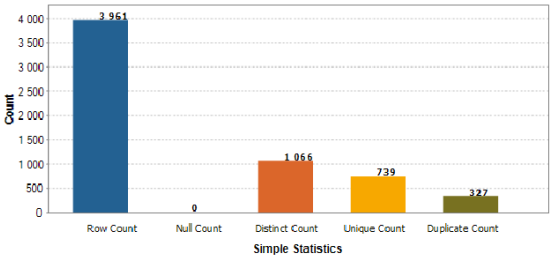
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	1338	33.78%
Unique Count	1014	25.60%
Duplicate Count	324	8.18%



Column: metadata.AST

Simple Statistics

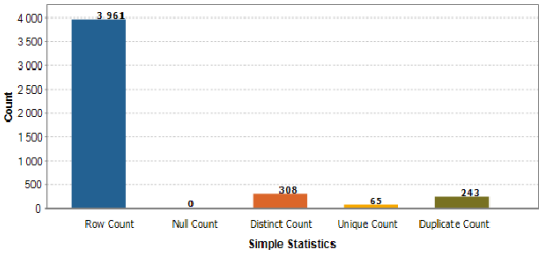
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	1066	26.91%
Unique Count	739	18.66%
Duplicate Count	327	8.26%



Column: metadata.FG_Percentage

Simple Statistics

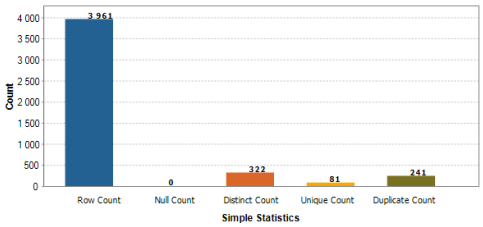
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	308	7.78%
Unique Count	65	1.64%
Duplicate Count	243	6.13%



Column: metadata.TP_Percentage

Simple Statistics

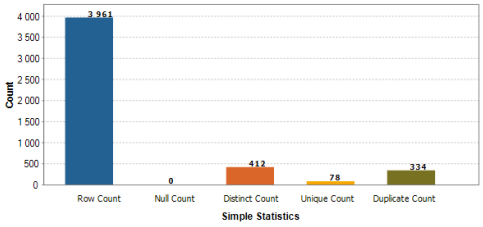
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	322	8.13%
Unique Count	81	2.04%
Duplicate Count	241	6.08%



Column: metadata.FT_Percentage

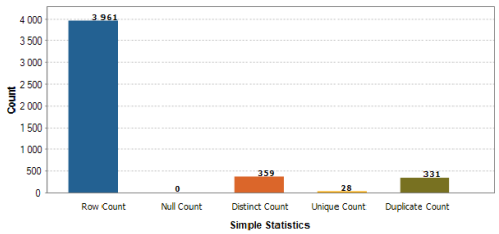
Simple Statistics

Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	412	10.40%
Unique Count	78	1.97%
Duplicate Count	334	8.43%



Simple Statistics

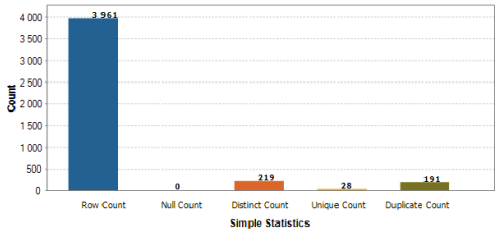
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	359	9.06%
Unique Count	28	0.71%
Duplicate Count	331	8.36%



Column: metadata.Points_per_Game

Simple Statistics

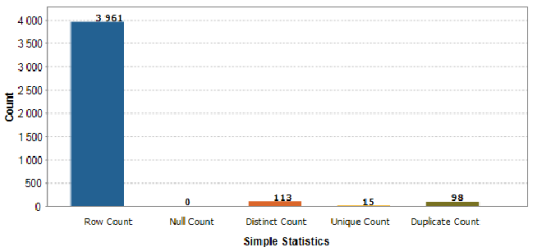
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	219	5.53%
Unique Count	28	0.71%
Duplicate Count	191	4.82%



Column: metadata.TRB_per_game

Simple Statistics

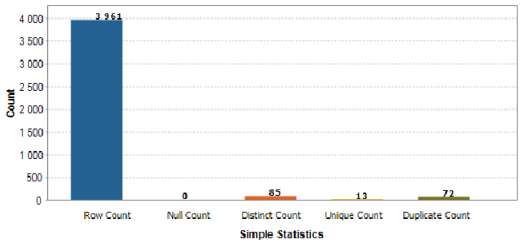
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	113	2.85%
Unique Count	15	0.38%
Duplicate Count	98	2.47%



Column: metadata.Assits_per_Game

Simple Statistics

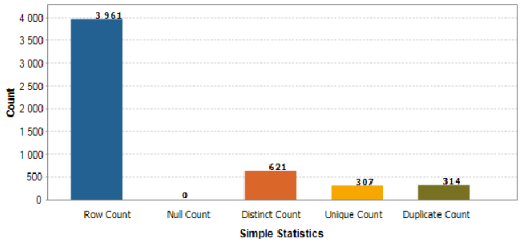
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	85	2.15%
Unique Count	13	0.33%
Duplicate Count	72	1.82%



Column: metadata.Win_Share

Simple Statistics

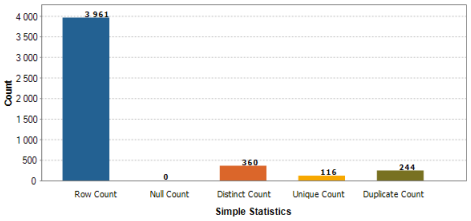
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	621	15.68%
Unique Count	307	7.75%
Duplicate Count	314	7.93%



Column: metadata.WS_per_game

Simple Statistics

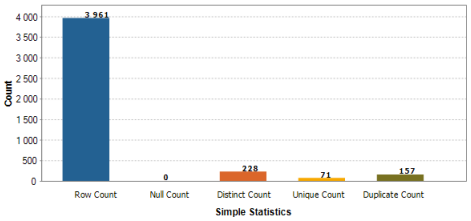
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	360	9.09%
Unique Count	116	2.93%
Duplicate Count	244	6.16%



Column: metadata.BPM

Simple Statistics

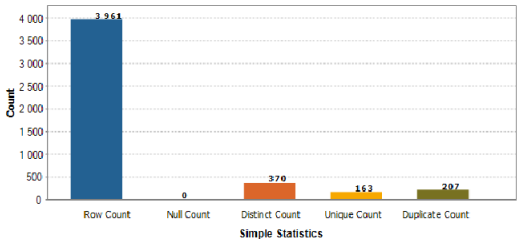
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	228	5.76%
Unique Count	71	1.79%
Duplicate Count	157	3.96%



Column: metadata.VORP

Simple Statistics

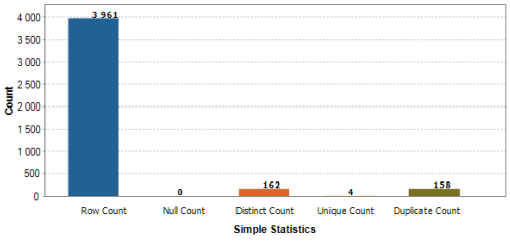
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	370	9.34%
Unique Count	163	4.12%
Duplicate Count	207	5.23%



Column: metadata.Executive

Simple Statistics

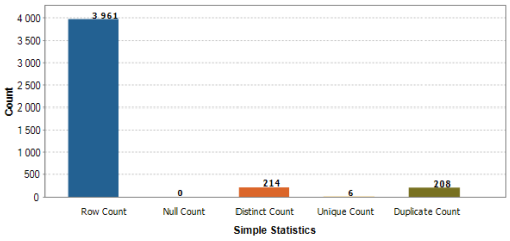
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	162	4.09%
Unique Count	4	0.10%
Duplicate Count	158	3.99%



Column: metadata.Tenure

Simple Statistics

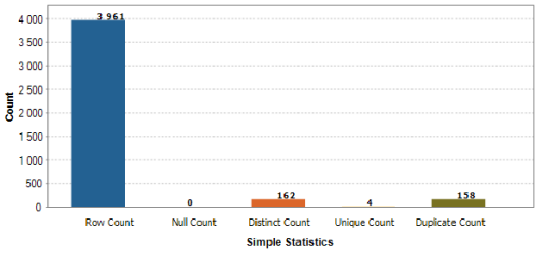
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	214	5.40%
Unique Count	6	0.15%
Duplicate Count	208	5.25%



Column: metadata.Exec_ID

Simple Statistics

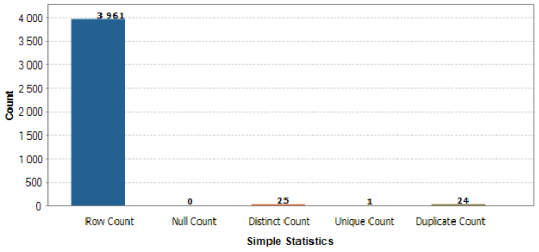
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	162	4.09%
Unique Count	4	0.10%
Duplicate Count	158	3.99%



Column: metadata.Exec_draft_exp

Simple Statistics

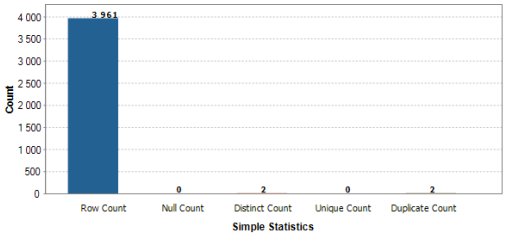
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	25	0.63%
Unique Count	1	0.03%
Duplicate Count	24	0.61%



Column: metadata.attend_college

Simple Statistics

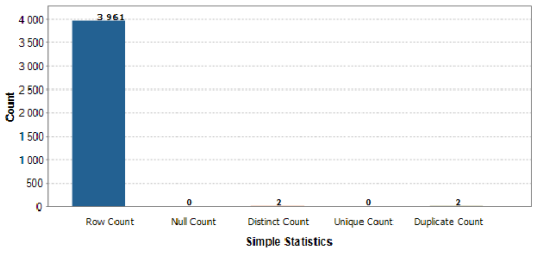
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	2	0.05%
Unique Count	0	0.00%
Duplicate Count	2	0.05%



Column: metadata.first_year

Simple Statistics

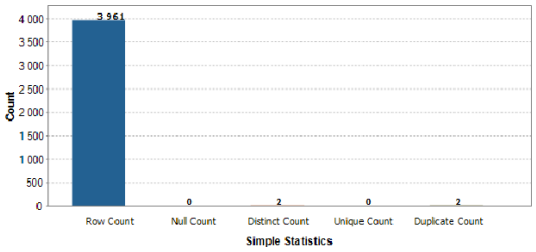
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	2	0.05%
Unique Count	0	0.00%
Duplicate Count	2	0.05%



Column: metadata.second_year

Simple Statistics

Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	2	0.05%
Unique Count	0	0.00%
Duplicate Count	2	0.05%

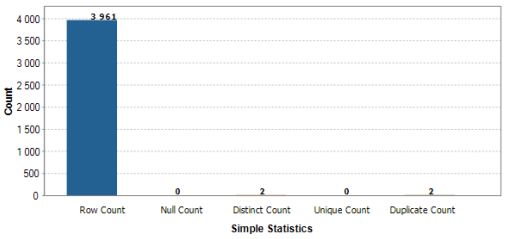


Column: metadata.third_year

Column: metadata.third_year

Simple Statistics

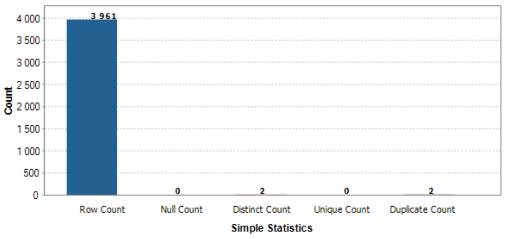
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	2	0.05%
Unique Count	0	0.00%
Duplicate Count	2	0.05%



Column: metadata.fourth_year

Simple Statistics

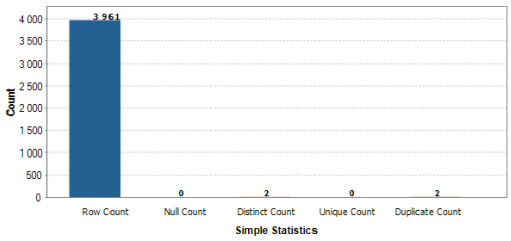
Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	2	0.05%
Unique Count	0	0.00%
Duplicate Count	2	0.05%



Column: metadata.fifth_year

Simple Statistics

Label	Count	%
Row Count	3961	100.00%
Null Count	0	0.00%
Distinct Count	2	0.05%
Unique Count	0	0.00%
Duplicate Count	2	0.05%

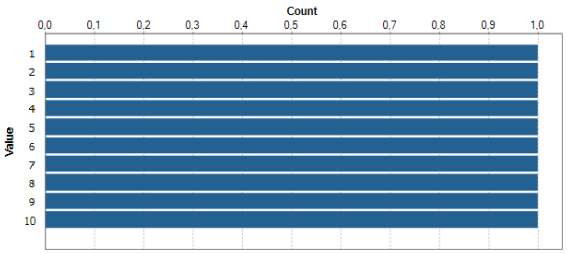


Analysis Results

Column: metadata.Column0

Value Frequency

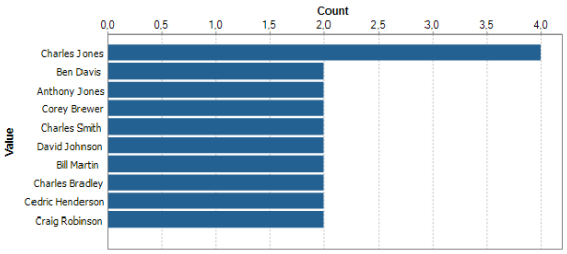
Value	Count	%
1	1	0.03%
2	1	0.03%
3	1	0.03%
4	1	0.03%
5	1	0.03%
6	1	0.03%
7	1	0.03%
8	1	0.03%
9	1	0.03%
10	1	0.03%



Column: metadata.Player

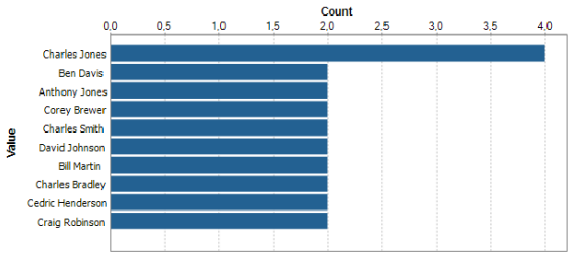
Value Frequency

Value	Count	%
Charles Jones	4	0.10%
Ben Davis	2	0.05%
Anthony Jones	2	0.05%
Corey Brewer	2	0.05%
Charles Smith	2	0.05%
David Johnson	2	0.05%
Bill Martin	2	0.05%
Charles Bradley	2	0.05%
Cedric Henderson	2	0.05%
Craig Robinson	2	0.05%



Value Frequency

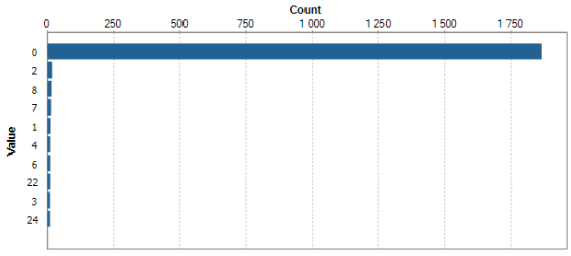
Value	Count	%
Charles Jones	4	0.10%
Ben Davis	2	0.05%
Anthony Jones	2	0.05%
Corey Brewer	2	0.05%
Charles Smith	2	0.05%
David Johnson	2	0.05%
Bill Martin	2	0.05%
Charles Bradley	2	0.05%
Cedric Henderson	2	0.05%
Craig Robinson	2	0.05%



Column: metadata.Games

Value Frequency

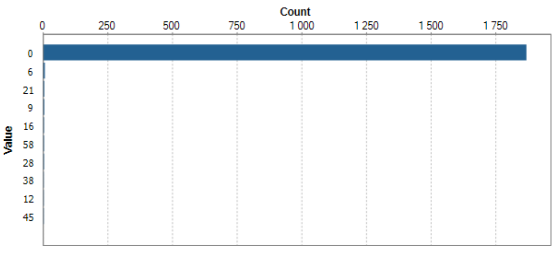
Value	Count	%
0	1869	47.19%
2	20	0.50%
8	18	0.45%
7	16	0.40%
1	14	0.35%
4	13	0.33%
6	13	0.33%
22	13	0.33%
3	12	0.30%
24	12	0.30%



Column: metadata.Minutes_Played

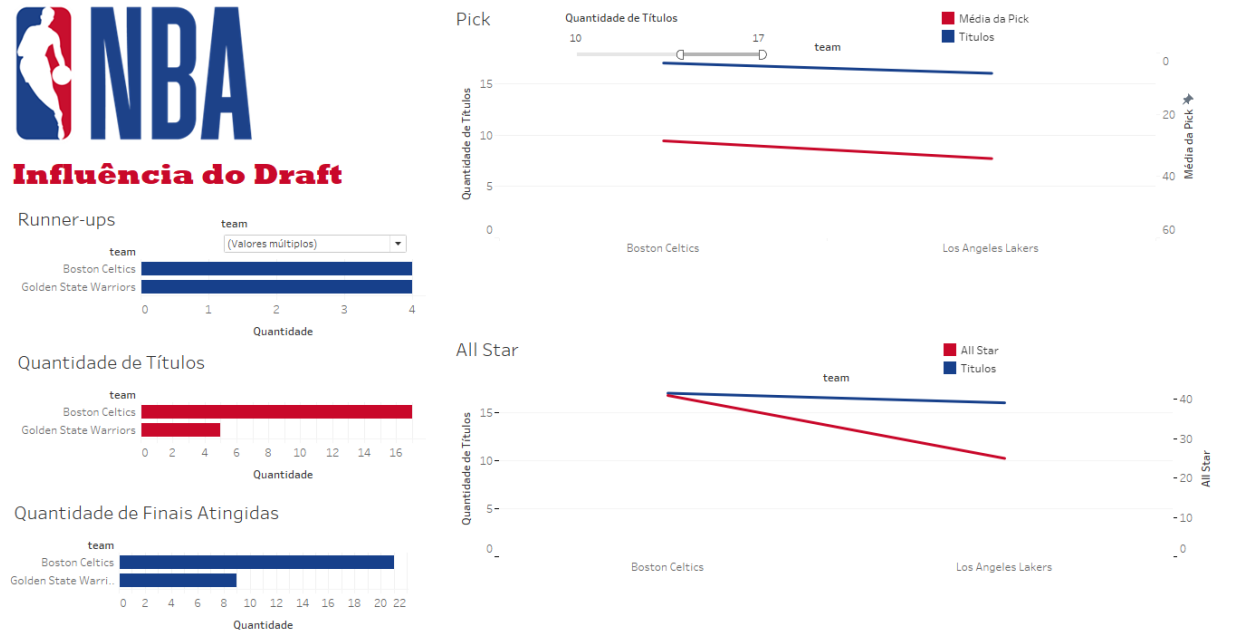
Value Frequency

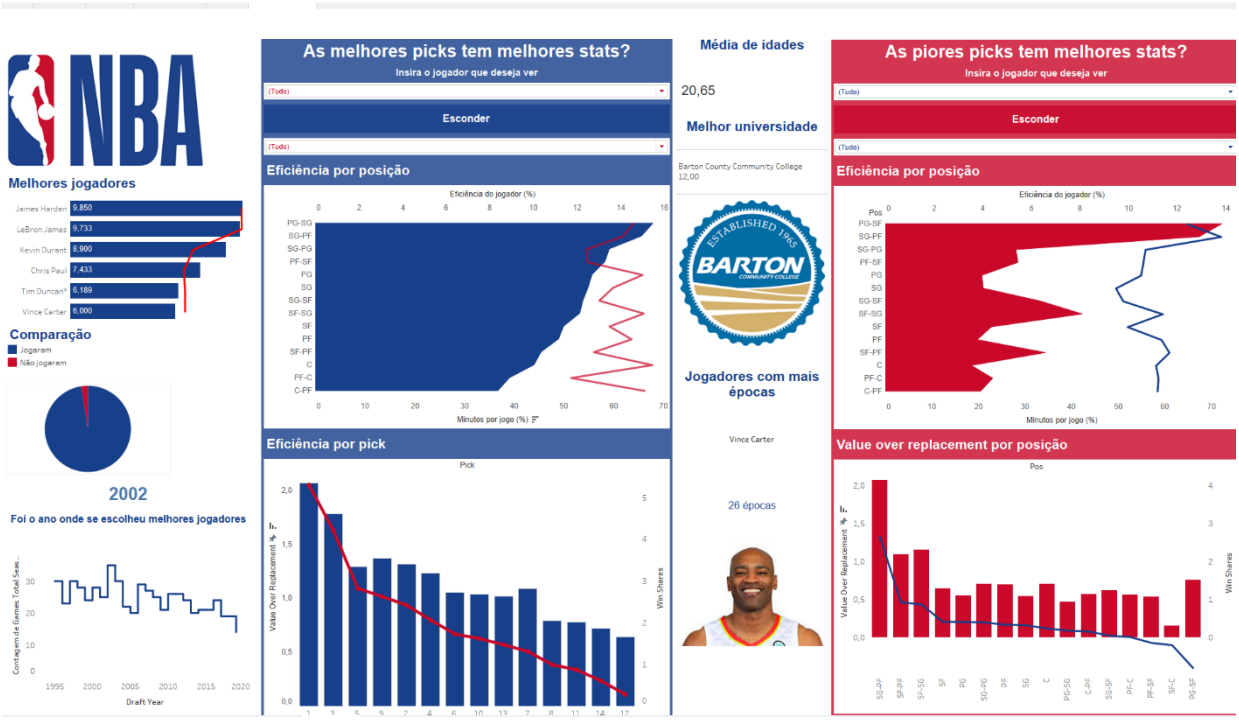
Value	Count	%
0	1870	47.21%
6	8	0.20%
21	6	0.15%
9	6	0.15%
16	5	0.13%
58	5	0.13%
28	5	0.13%
38	4	0.10%
12	4	0.10%
45	4	0.10%

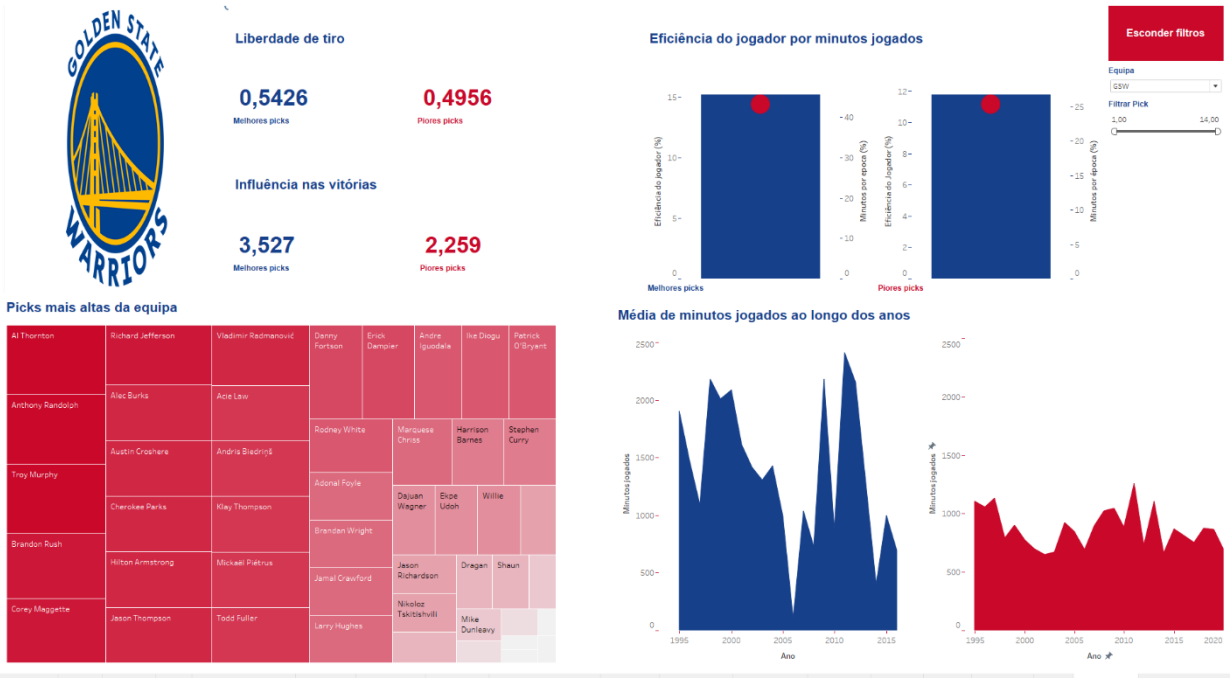


Go to page 1/1

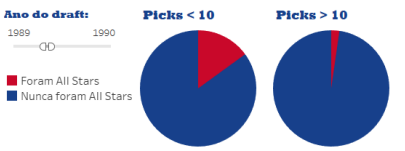
3.7 Dashboard:

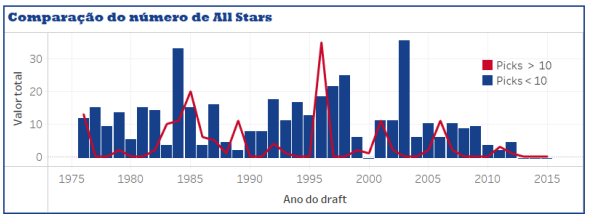


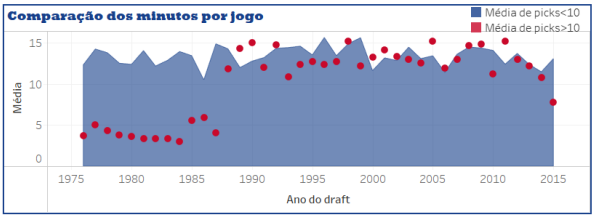


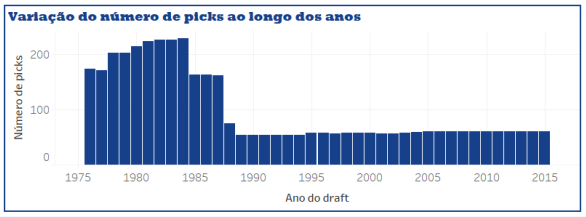

Liberdade de tiro
0,5426
Melhores picks
0,4956
Piores picks
Influência nas vitórias
3,527
Melhores picks
2,259
Piores picks
Eficiência do jogador por minutos jogados
Esconder filtros
Equipa
GSW
Filtrar Pick
1,00
14,00
Melhores picks
Piores picks
Picks mais altas da equipa
Al Thornton
Anthony Randolph
Troy Murphy
Brandon Rush
Corey Maggette
Richard Jefferson
Alec Burks
Austin Croshere
Cherokee Parks
Hilton Armstrong
Jason Thompson
Vladimir Radmanović
Acie Law
Andris Biedriņš
Klay Thompson
Mickaël Piétrus
Todd Fuller
Danny Fortson
Erick Dampier
Andre Iguodala
Ike Diogu
Patrick O'Bryant
Rodney White
Marquese Chriss
Harrison Barnes
Stephen Curry
Adonal Foyle
Dajuan Wagner
Ekpe Udoh
Willie
Brandan Wright
Jason Richardson
Dragan
Shaun
Jamal Crawford
Nikoloz Tskitishvili
Mike Dunleavy
Larry Hughes
Média de minutos jogados ao longo dos anos
Minutos jogados
Ano

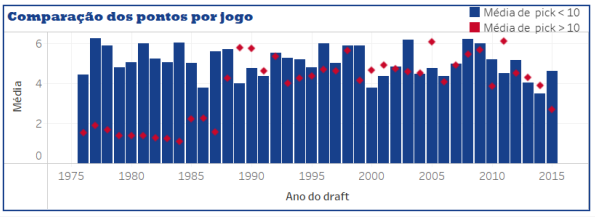

NBA
Serão as primeiras 10 picks no draft os melhores jogadores?


Ano do draft:
1989
1990
Foram All Stars
Nunca foram All Stars
Picks < 10
Picks > 10


Comparação do número de All Stars
Valor total
Picks > 10
Picks < 10
Ano do draft


Comparação dos minutos por jogo
Média de picks<10
Média de picks>10
Média
Ano do draft


Variação do número de picks ao longo dos anos
Número de picks
Ano do draft


Comparação dos pontos por jogo
Média de pick < 10
Média de pick > 10
Média
Ano do draft

Distribuição de Notas:

Após discussão entre todos os elementos do grupo, achamos que a distribuição correta e justa das notas a cada elemento deve ser a seguinte:

(a92948) Ricardo Silva : **N**

(a92938) Duarte Brandão : **N**

(a92939) João Lemos : **N**

(a92947) Nuno Moreira : **N**

(a92930) Pedro Gonçalves : **N**

(a92929) Paulo Barros : **N**

(a94123) Tiago Lopes : **N**