KASDD Genap 2022/2023

Football Player Dataset

- Ainun Nur Rohmah - 2006597052
- Fitri 'Aliyah - 2006597115
- Hanif Ibrahim Syuaib - 2006482003
- Neiva Annur Syahistie - 2006531711

Kelompok Sebelah

Football Player

Business Understanding

Eksplorasi Data dilakukan untuk menjawab Pertanyaan berikut

1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?
2. Apakah pemain dengan umur lebih dari umur 30 masih bermain dengan kerja keras yang tinggi?
3. Apakah terdapat hubungan antara posisi pemain sepakbola dengan tingkat skill-nya?
4. Apakah negara sepakbola dengan starting eleven terbaik?
5. Apakah pemain dengan nilai rating yang diatas rata-rata memiliki nilai special yang tinggi?
6. Apa negara yang memiliki pemain dengan best overall rating tertinggi?

Football Player

Data Understanding

Dataset yang digunakan adalah Football Player. Dataset ini berisikan berbagai informasi seorang pesepak bola. Dataset ini mencakup pemain sepak bola, negara, klub, dan detail rating dari pemain tersebut.

Dataset ini memiliki 60 kolom dan 16710 baris. Setelah di transformasi, kami mengidentifikasi terdapat 45 kolom numerik dan 15 kolom kategorikal.

Data Numerik

```
Index(['Age', 'Overall', 'Potential', 'Best Overall Rating', 'Value', 'Wage',
       'Release Clause', 'Special', 'Height', 'Weight', 'Crossing',
       'Finishing', 'Heading Accuracy', 'Short Passing', 'Volleys',
       'Dribbling', 'Curve', 'FK Accuracy', 'Long Passing', 'Ball Control',
       'Acceleration', 'Sprint Speed', 'Agility', 'Reactions', 'Balance',
       'Shot Power', 'Jumping', 'Stamina', 'Strength', 'Long Shots',
       'Aggression', 'Interceptions', 'Positioning', 'Vision', 'Penalties',
       'Composure', 'Marking', 'Standing Tackle', 'Sliding Tackle',
       'GK Diving', 'GK Handling', 'GK Kicking', 'GK Positioning',
       'GK Reflexes', 'Defensive Awareness'],
      dtype='object')
```

Data Kategorikal

```
Index(['ID', 'Name', 'Nationality', 'Club', 'Joined', 'Contract Valid Until',
       'Position', 'Jersey Number', 'Preferred Foot',
       'International Reputation', 'Weak Foot', 'Skill Moves', 'Body Type',
       'Attacking Work Rate', 'Defensive Work Rate'],
      dtype='object')
```

Data Understanding

Penjelasan Data Numerik (1)

Terdapat total 45 atribut numerikal pada data.

- **Age:** Umur pemain sepak bola
- **Overall:** Rating pemain sepak bola
- **Potential:** Potensi rating pemain sepak bola
- **Best Overall Rating:** Rating tertinggi yang pernah didapatkan pemain sepak bola
- **Value:** Nilai harga pasar pemain sepak bola
- **Wage:** Gaji yang diperoleh pemain sepak bola
- **Release Clause:** Biaya transfer minimum pemain sepak bola
- **Special:** Nilai kespesialan pemain sepak bola
- **Height:** Tinggi badan pemain sepak bola
- **Weight:** Berat badan pemain sepak bola
- **Crossing:** Nilai kualitas pemain sepak bola dalam mengumpan silang
- **Finishing:** Nilai kualitas pemain sepak bola dalam mengeksekusi bola menjadi gol
- **Heading Accuracy:** Nilai kualitas akurasi pemain sepak bola dalam menyundul bola
- **Short Passing:** Nilai kualitas pemain sepak bola dalam mengumpan pendek
- **Volleys:** Nilai kualitas pemain sepak bola dalam tendangan voli

Data Understanding

Penjelasan Data Numerik (2)

- **Dribbling:** Nilai kualitas pemain sepak bola dalam menggiring bola
- **Curve:** Nilai kualitas pemain sepak bola dalam tendengan melengkung
- **FK Accuracy:** Nilai kualitas akurasi pemain sepak bola dalam tendangan bebas
- **Long Passing:** Nilai kualitas pemain sepak bola dalam mengumpan jauh
- **Ball Control:** Nilai kualitas pemain sepak bola dalam mengontrol bola
- **Acceleration:** Nilai kualitas pemain sepak bola dalam akselerasi
- **Sprint Speed:** Nilai kualitas pemain sepak bola dalam berlari
- **Agility:** Nilai kualitas pemain sepak bola dalam kelincahan
- **Reactions:** Nilai kualitas pemain sepak bola dalam bereaksi
- **Balance:** Nilai kualitas pemain sepak bola dalam keseimbangan
- **Shot Power:** Nilai kualitas pemain sepak bola dalam kekuatan tendangan
- **Jumping:** Nilai kualitas pemain sepak bola dalam melompat
- **Stamina:** Nilai kualitas pemain sepak bola dalam daya tahan
- **Strength:** Nilai kualitas pemain sepak bola dalam kekuatan
- **Long Shots:** Nilai kualitas pemain sepak bola dalam tendangan jarak jauh

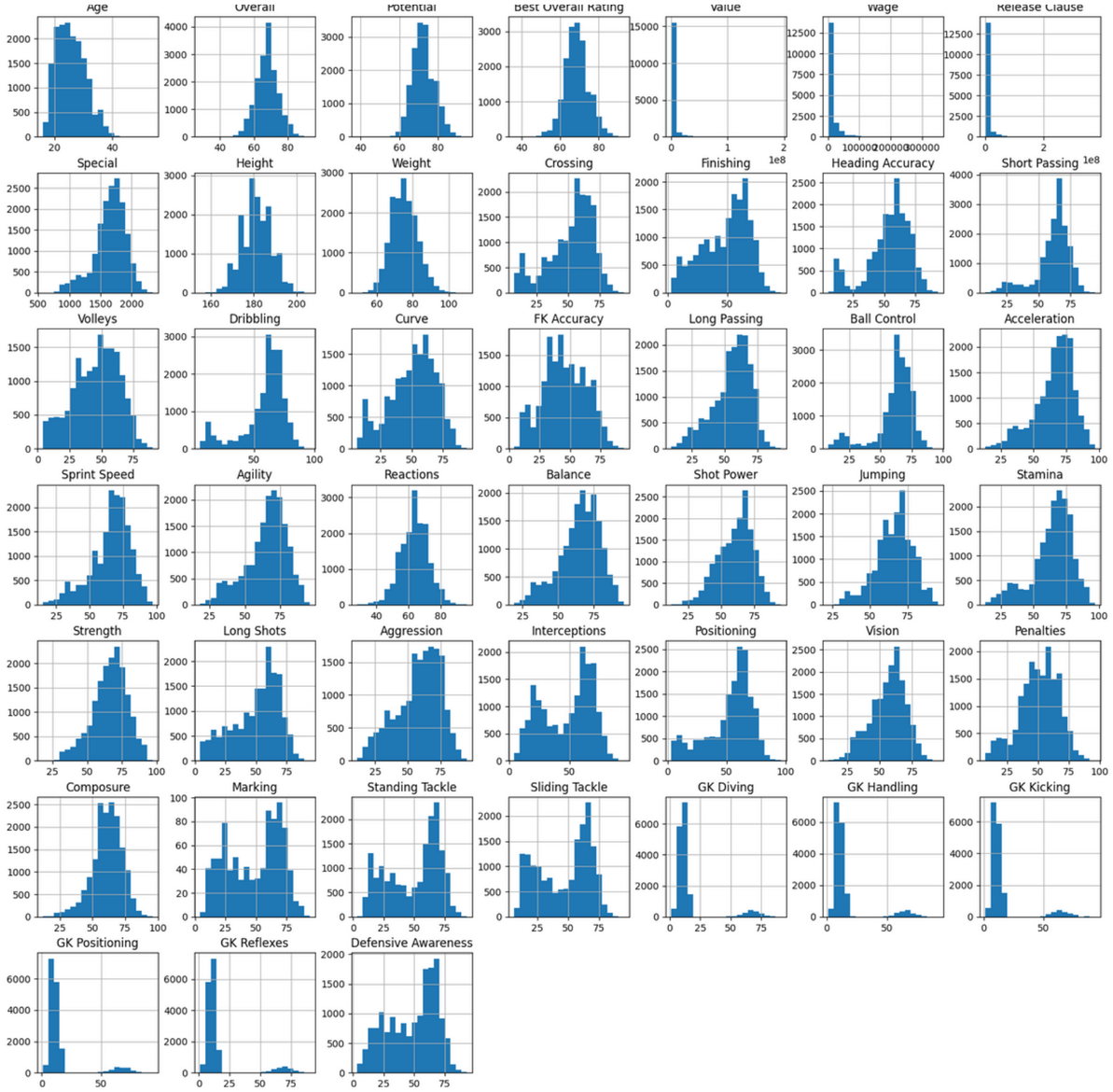
Data Understanding

Penjelasan Data Numerik (3)

- **Aggression:** Nilai kualitas pemain sepak bola dalam agresif dengan pemain lawan
- **Interceptions:** Nilai kualitas pemain sepak bola dalam mengintersep bola yang datang
- **Positioning:** Nilai kualitas pemain sepak bola dalam penempatan posisi
- **Vision:** Nilai kualitas pemain sepak bola dalam penglihatan
- **Penalties:** Nilai kualitas pemain sepak bola dalam tendangan pinalti
- **Composure:** Nilai kualitas pemain sepak bola dalam ketenangan
- **Marking:** Nilai kualitas pemain sepak bola dalam penjagaan terhadap pemain lawan
- **Standing Tackle:** Nilai kualitas pemain sepak bola dalam tackle berposisi berdiri
- **Sliding Tackle:** Nilai kualitas pemain sepak bola dalam tackle berposisi meluncur
- **GK Diving:** Nilai kualitas pemain sepak bola atau kiper dalam melompat untuk menipis bola
- **GK Handling:** Nilai kualitas pemain sepak bola atau kiper dalam menangkap bola
- **GK Kicking:** Nilai kualitas pemain sepak bola atau kiper dalam menendang bola
- **GK Positioning:** Nilai kualitas pemain sepak bola atau kiper dalam penempatan posisi
- **GK Reflexes:** Nilai kualitas pemain sepak bola atau kiper dalam reflek untuk menepis bola
- **Defensive Awareness:** Nilai kualitas kesadaran pemain sepak bola ketika bertahan

Data Understanding
Distribusi Data Numerik

Dapat dilihat pada visualisasi di sambing bahwa data cenderung skewed



Data Understanding

Penjelasan Data Kategorikal

Terdapat total 15 atribut kategorikal pada data.

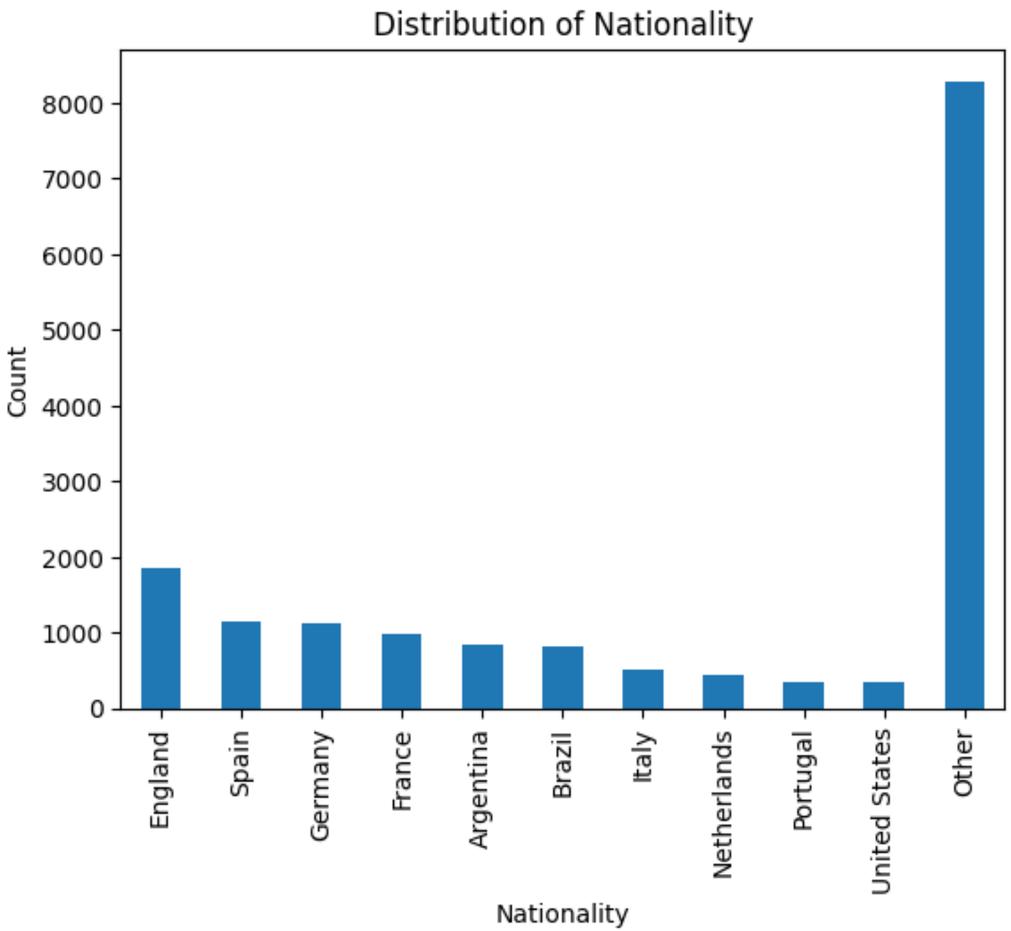
- **ID:** Unique identifier pemain sepak bola
- **Name:** Nama pemain sepak bola
- **Nationality:** Kewarganegaraan pemain sepak bola
- **Club:** Tim sepak bola yang dibela oleh pemain sepak bola
- **Joined:** Tanggal pemain sepak bola bergabung dengan tim sepak bola
- **Contract Valid Until:** Tanggal berakhirnya kontrak pemain sepak bola pada tim sepak bola
- **Position:** Posisi pemain sepak bola di lapangan
- **Jersey Number:** Nomor punggung yang digunakan pemain sepak bola
- **Preferred Foot:** Kaki yang paling sering digunakan oleh pemain sepak bola diantara kedua kaki
- **International Reputation:** Reputasi dari pemain sepak bola
- **Weak Foot:** Tingkat kemahiran pemain sepak bola pada menggunakan kaki yang jarang digunakan
- **Skill Moves:** Tingkat kemahiran pemain sepak bola dalam pergerakannya dengan bola
- **Body Type:** Tipe badan dari pemain sepak bola
- **Attacking Work Rate:** Tingkat kerja keras pemain sepak bola ketika menyerang
- **Defensive Work Rate:** Tingkat kerja keras pemain sepak bola ketika bertahan

Data Understanding

Distribusi Data Kategorikal

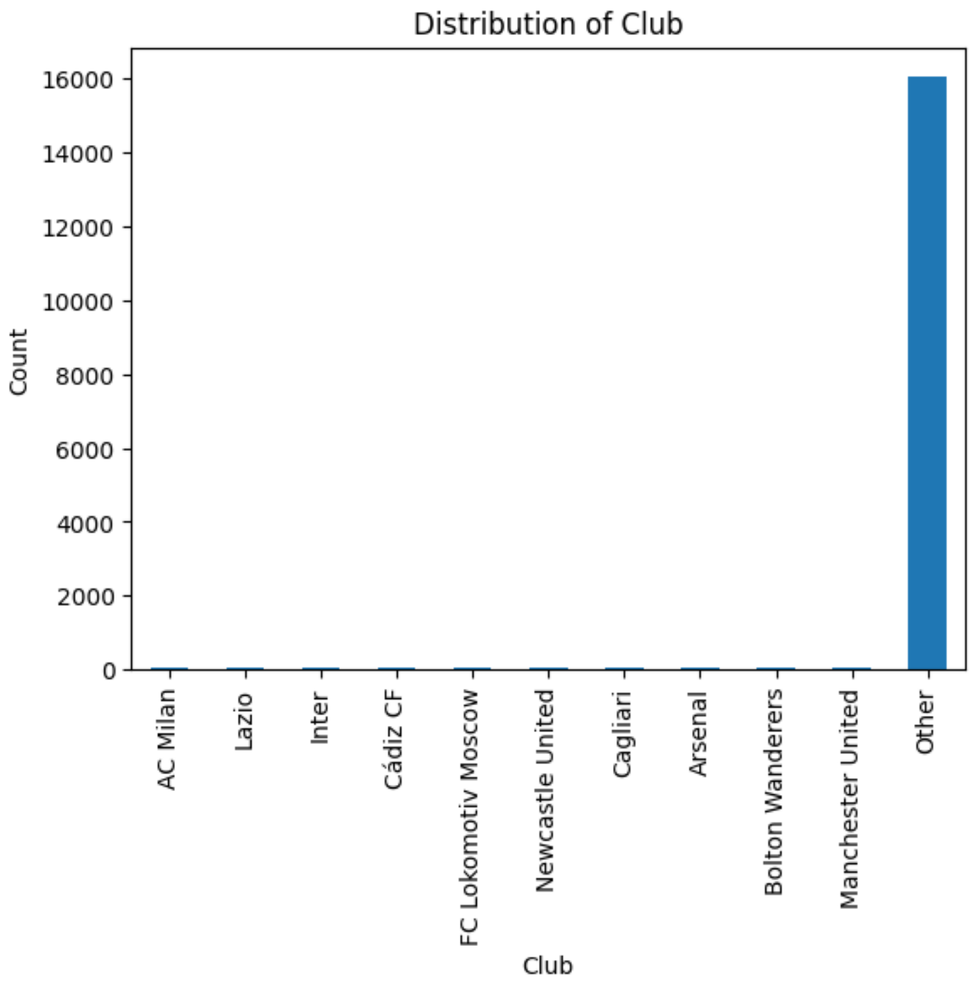
Selanjutnya akan ditampilkan bar chart untuk melihat distribusi masing-masing atribut kategorikal, kecuali 'ID', 'Name', 'Joined', dan 'Contract Valid Until'.

Nationality



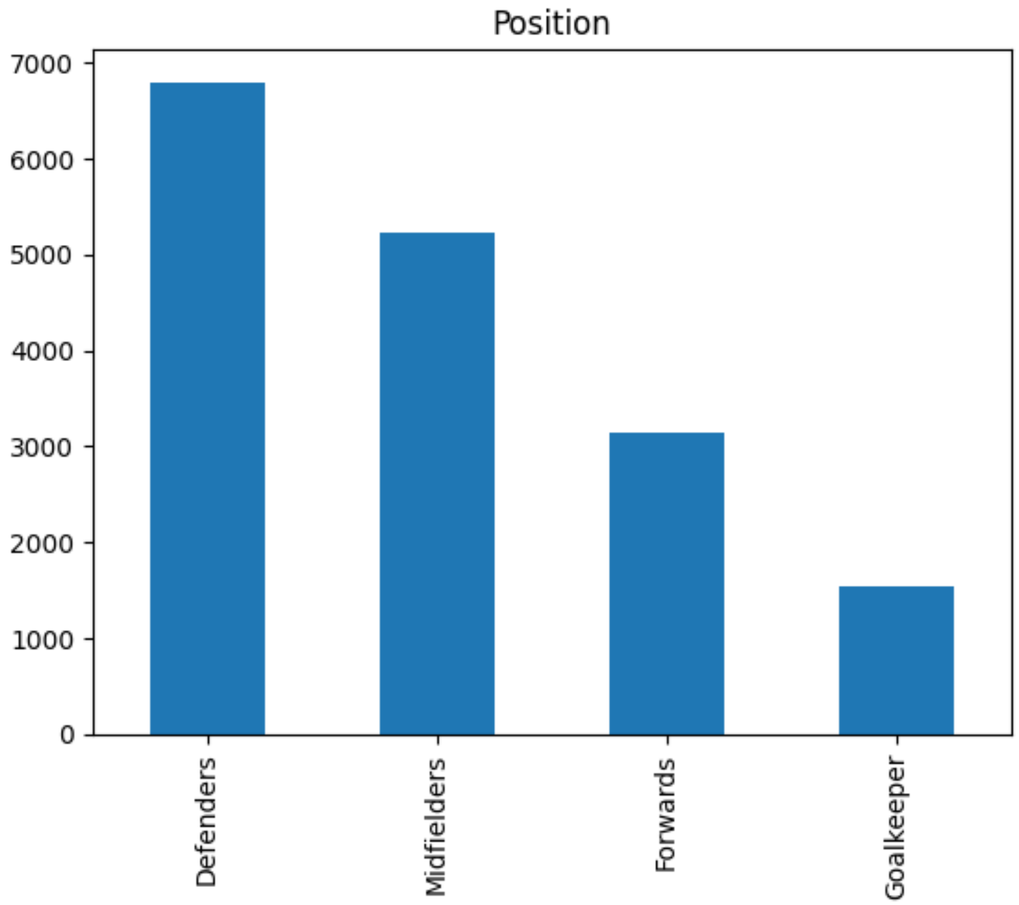
Unique Values: 168
Mode: England

Club



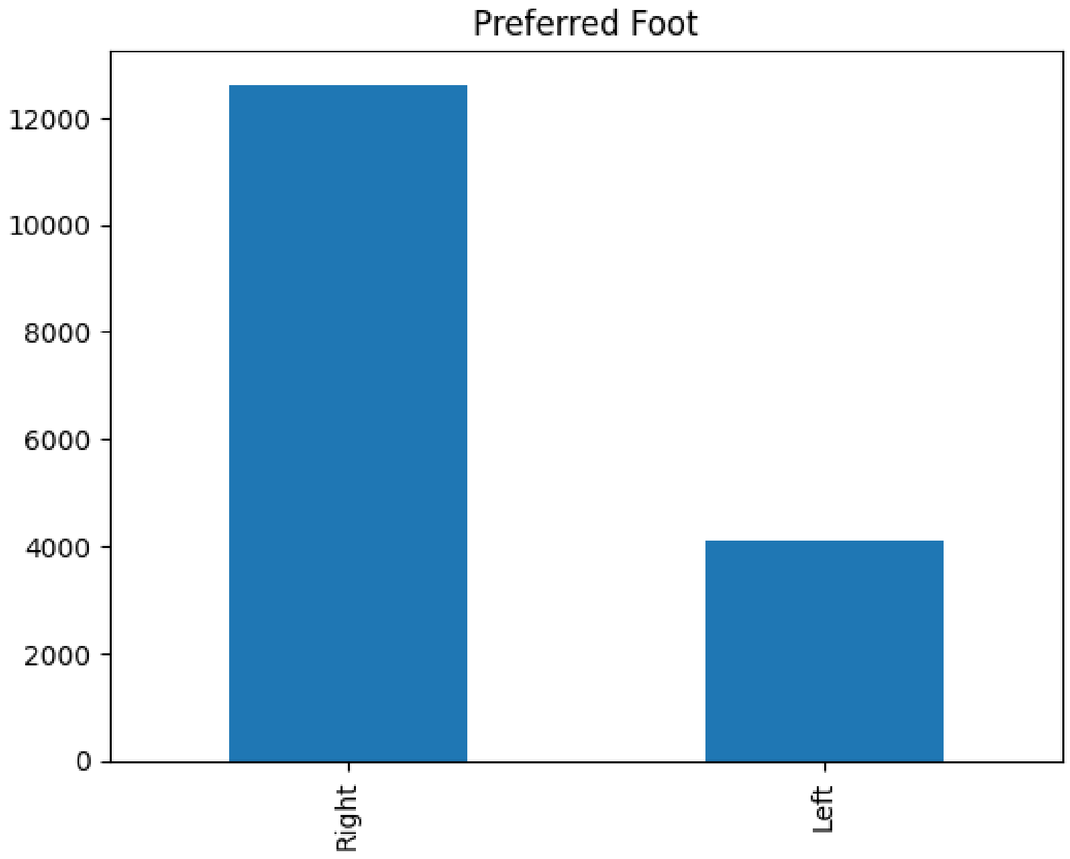
Unique Values: 869
Mode: AC Milan

Position



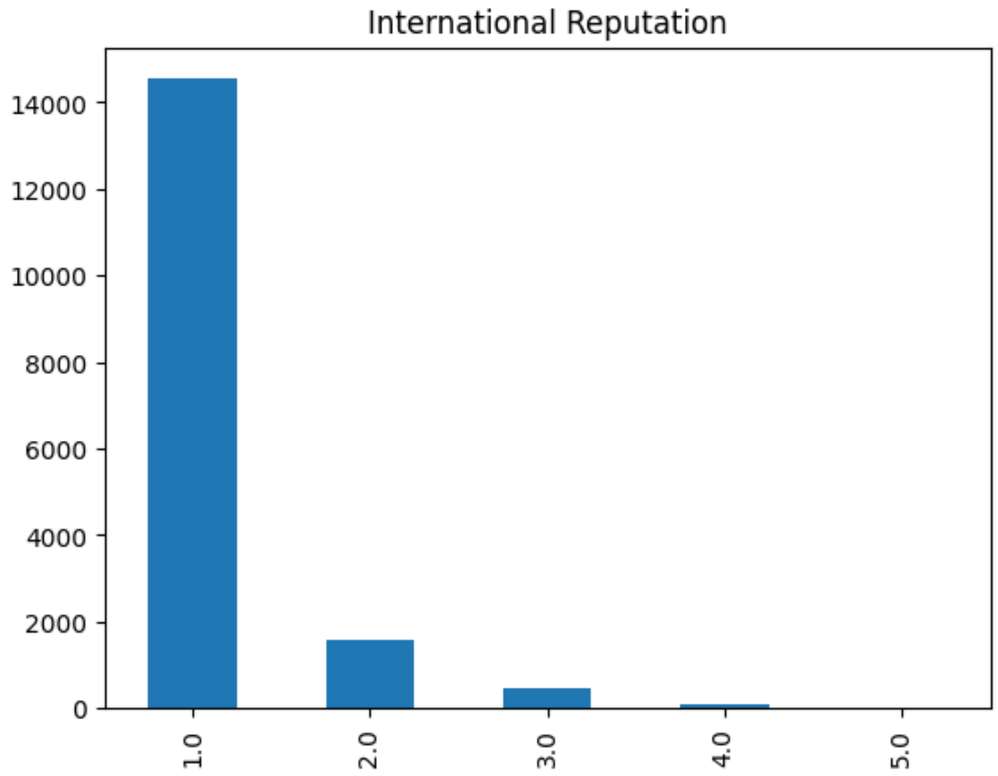
Unique Values: 4
Mode: Defenders

Preffered Foot



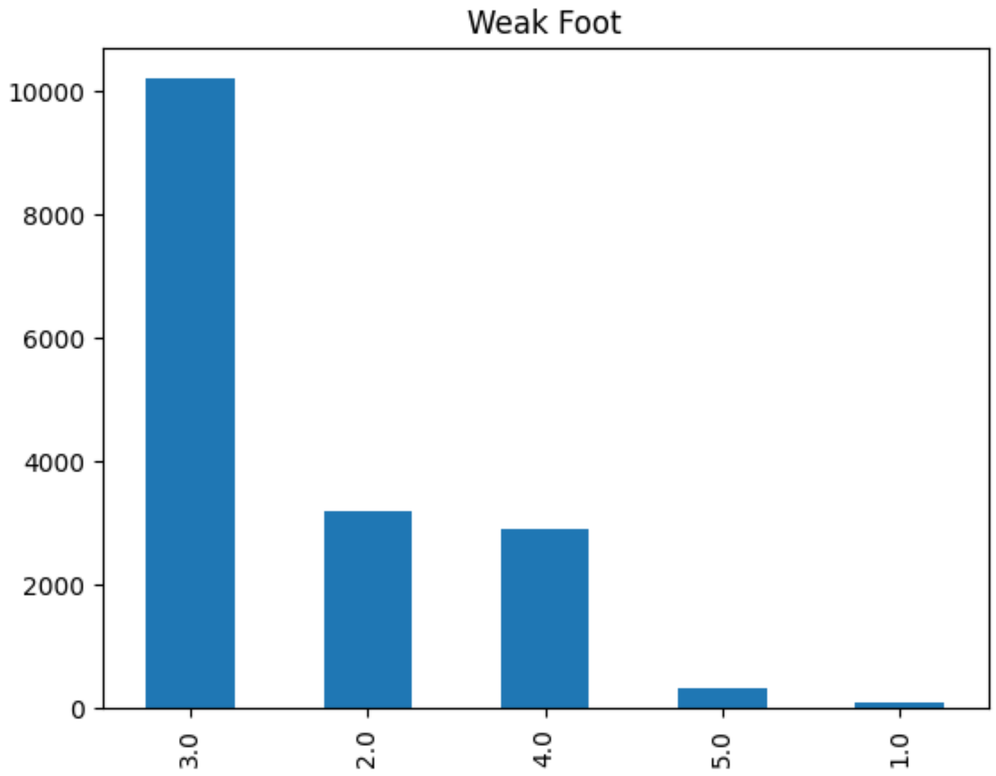
Unique Values: 2
Mode: Right

International Reputation



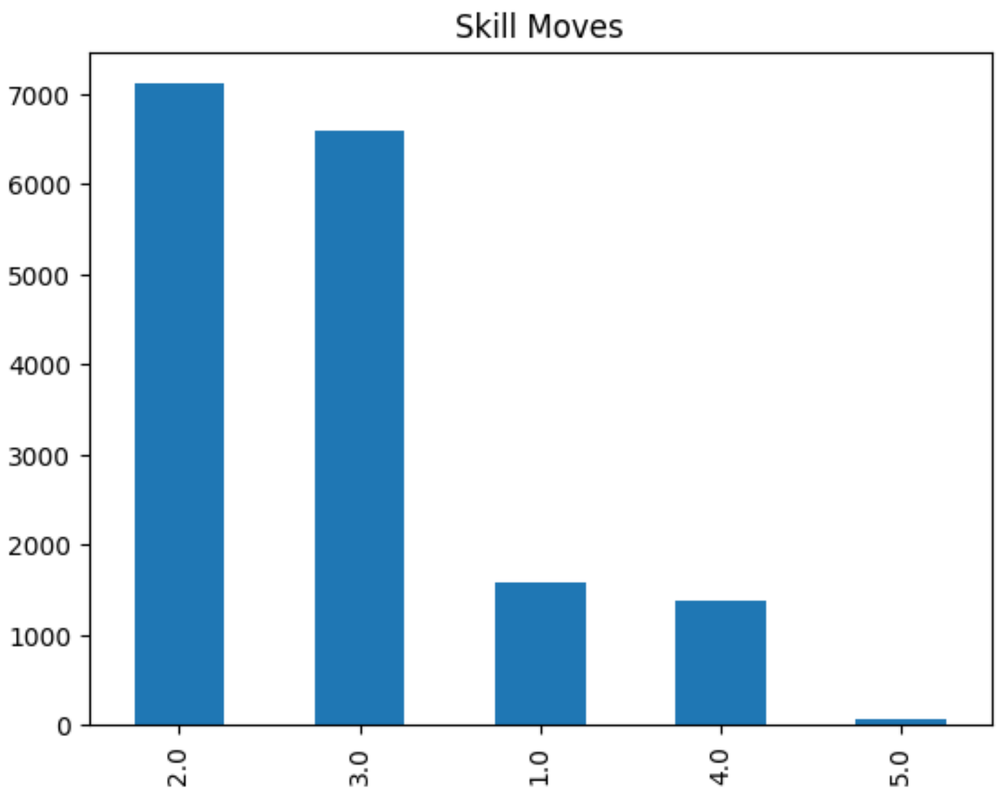
Unique Values: 5
Mode: 1.0

Weak Foot



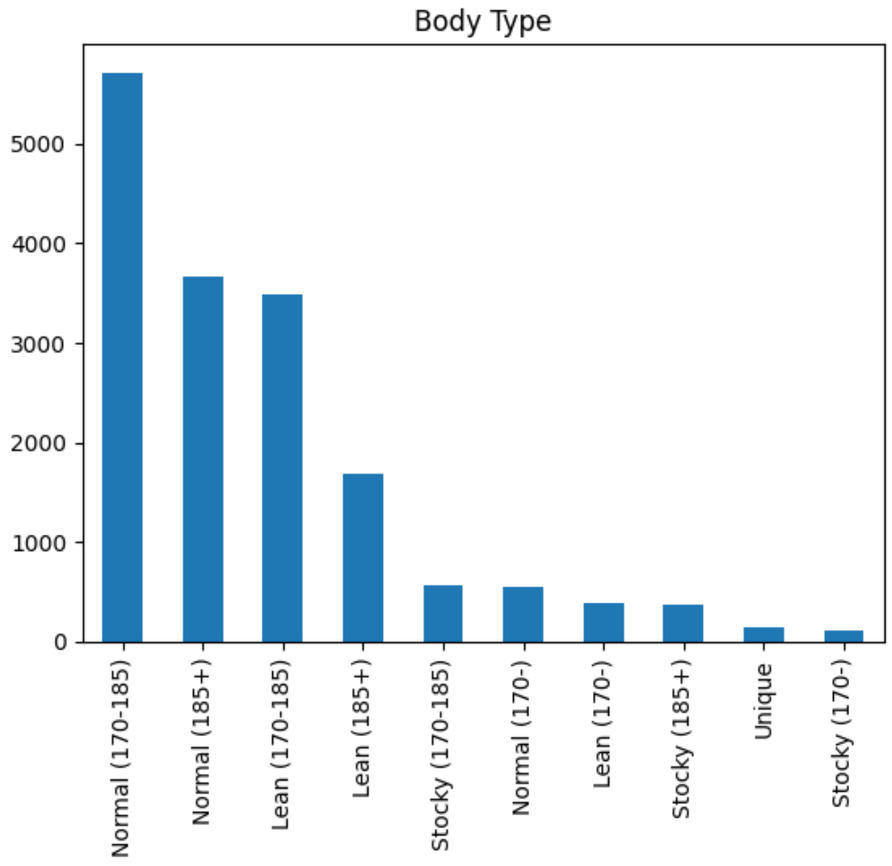
Unique Values: 5
Mode: 3.0

Skill Moves



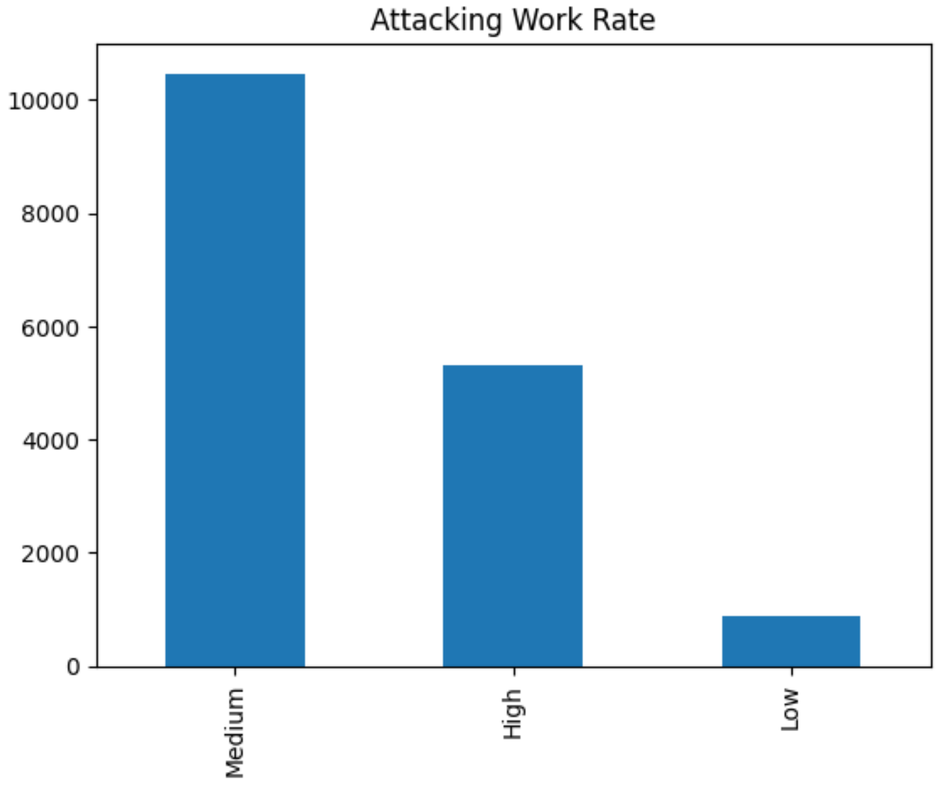
Unique Values: 5
Mode: 2.0

Body Type



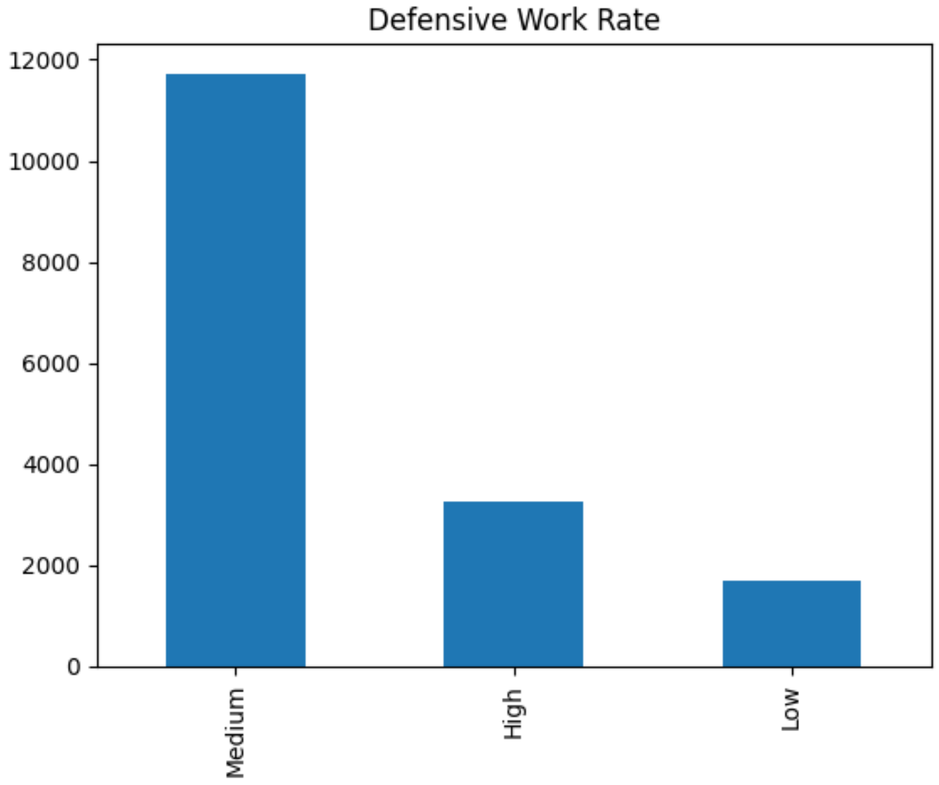
Unique Values: 10
Mode: Normal (170-185)

Attacking Work Rate



Unique Values: 3
Mode: Medium

Defensive Work Rate



Unique Values: 3
Mode: Medium

Data Understanding

Identifikasi Outlier pada Data Numerik (1)

Dengan menggunakan rumus IQR, RUB, dan LUB, didapatkan

34

dari 45 fitur numerik memiliki outliers

11

dari 45 fitur numerik tidak memiliki outliers

Data Understanding

Identifikasi Outlier pada Data Numerik (2)

AKAN TETAPI, berdasarkan hasil temuan mengenai setiap fitur tersebut melalui sumber-sumber yang terpercaya dan asumsi tim, data dianggap **NORMAL** sehingga **TIDAK PERLU DILAKUKAN HANDLE OUTLIER**

Selanjutnya, akan lebih dijelaskan mengapa kami menganggap data tersebut normal

Data Understanding

Identifikasi Outlier pada Data Numerik (3)

- **Value**: Dianggap normal karena walaupun range min dan max cukup jauh, pemain yang value-nya 0 adalah pemain yang tidak tergabung ke club manapun
- **Release Clause**: Dianggap normal karena biaya transfer minimum pemain bisa bervariasi
- **Wage**: Dianggap normal karena walaupun range min dan max cukup jauh, pemain yang wage-nya 0 adalah pemain yang tidak tergabung ke club manapun
- **Fitur terkait skill pemain**: Dianggap normal karena berdasarkan *fifaratings.com*, range skill pemain sepak bola adalah 0-100 dan semua nilai skill pemain pada dataset berada pada range tersebut
- **Weight**: Dianggap normal karena jarak min dan max masih berada di range berat badan normal seseorang
- **Height**: Dianggap normal karena jarak min dan max masih berada di tinggi badan normal seseorang
- **Age**: Dianggap mormal karena usia pemain pada dataset tidak ada yang terlalu muda atau terlalu tua (nilai min = 16, nilai max = 54)
- **Penalties**: Dianggap normal karena jumlah penalti yang diterima setiap pemain sepak bola bisa sangat bervariasi

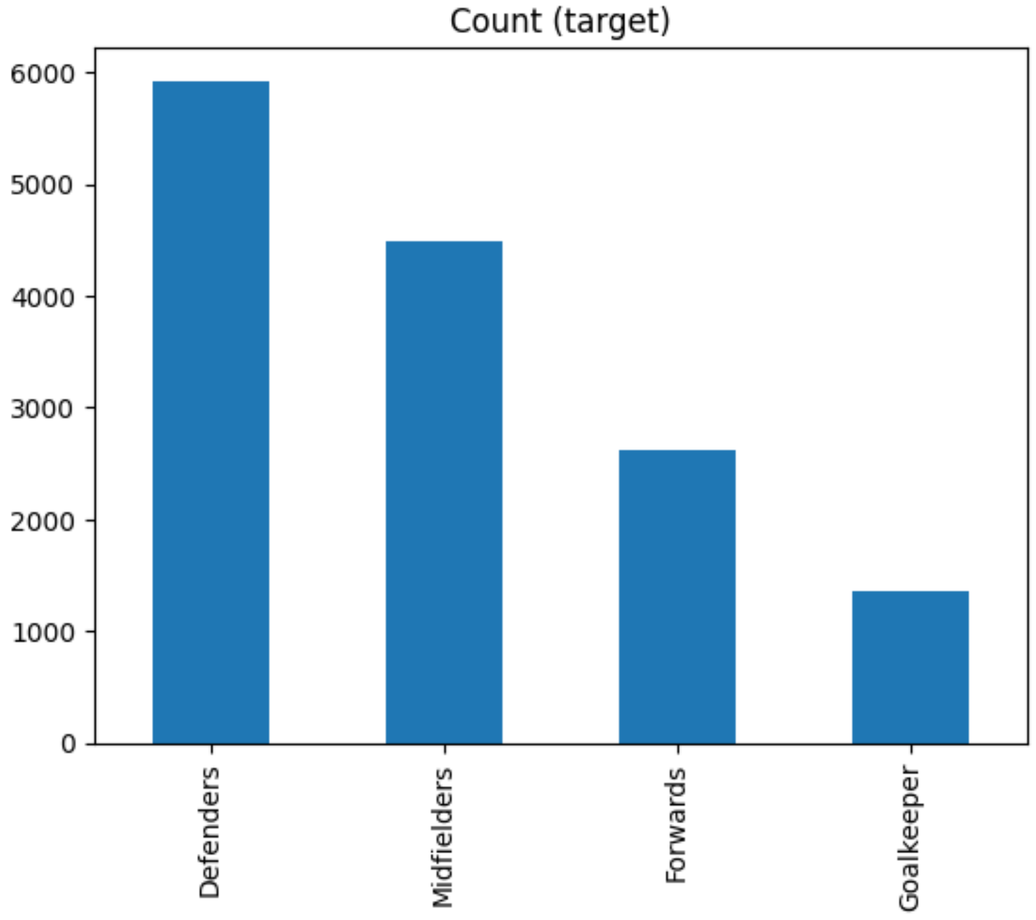
Data Understanding

Identifikasi Duplikasi Data

Dataset **TIDAK MEMILIKI DUPLIKASI** data, sehingga **TIDAK PERLU PENGHAPUSAN** baris yang duplikat

Data Understanding

Identifikasi Imbalanced Data



Berdasarkan visualisasi di samping, dapat diketahui bahwa dataset yang digunakan **SLIGHTLY IMBALANCED** sehingga **PERLU DILAKUKAN BALANCING** data dengan menggunakan teknik sampling.

Football Player

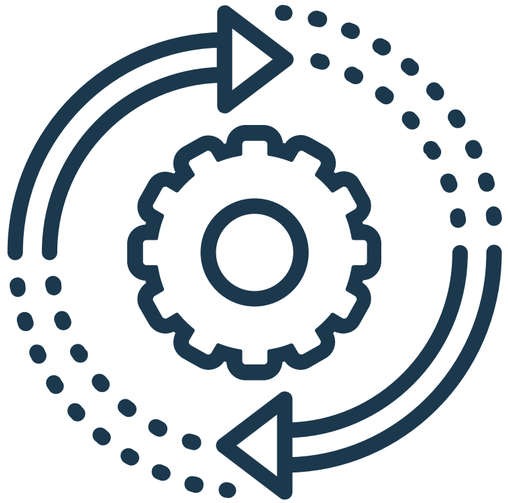
Data Preparation

Data Transformation

Handle Missing Value

Data Preparation

Data Transformation



Pada fitur 'Value', 'Wage', dan 'Release Clause'

- Tanda '€' dihilangkan
- Satuan 'K' diganti menjadi data * 1000
- Satuan 'M' diganti menjadi data * 1000000

Pada fitur 'Height'

- Tanda 'cm' dihilangkan

Pada fitur 'Weight'

- Tanda 'kg' dihilangkan

Tipe data kelima fitur di atas yang awalnya object diubah menjadi float

Melakukan Normalisasi pada data dengan menggunakan metode Min-Max karena outliers pada data cenderung sedikit

Data Preparation

Handle Missing Value

45	Positioning	float64	8	0.05	94	[87.0, 85.0, 91.0, 88.0, 81.0, 80.0, 83.0, 79....
46	Vision	float64	37	0.22	85	[90.0, 84.0, 94.0, 82.0, 85.0, 79.0, 86.0, 71....
47	Penalties	float64	0	0.00	89	[91.0, 60.0, 83.0, 76.0, 44.0, 71.0, 74.0, 79....
48	Composure	float64	251	1.50	83	[87.0, 82.0, 89.0, 85.0, 81.0, 86.0, 79.0, 90....
49	Marking	float64	15818	94.66	83	[nan, 81.0, 72.0, 63.0, 77.0, 75.0, 67.0, 86.0...
50	Standing Tackle	float64	0	0.00	87	[73.0, 85.0, 45.0, 65.0, 84.0, 79.0, 80.0, 87....
51	Sliding Tackle	float64	37	0.22	88	[65.0, 77.0, 38.0, 53.0, 82.0, 79.0, 75.0, 70....

Berdasarkan hasil pengecekan nilai null pada gambar diatas, kami memutuskan 2 hal, yaitu

1. Melakukan DROP pada fitur 'Marking' karena memiliki persentase null value sebesar 94,66%
2. Melakukan DROP pada seluruh baris fitur yang memiliki nilai null karena jumlah baris yang ada null hanya 2.312 baris atau 13,83% dari jumlah keseluruhan data

Data Visualization
Pertanyaan 1

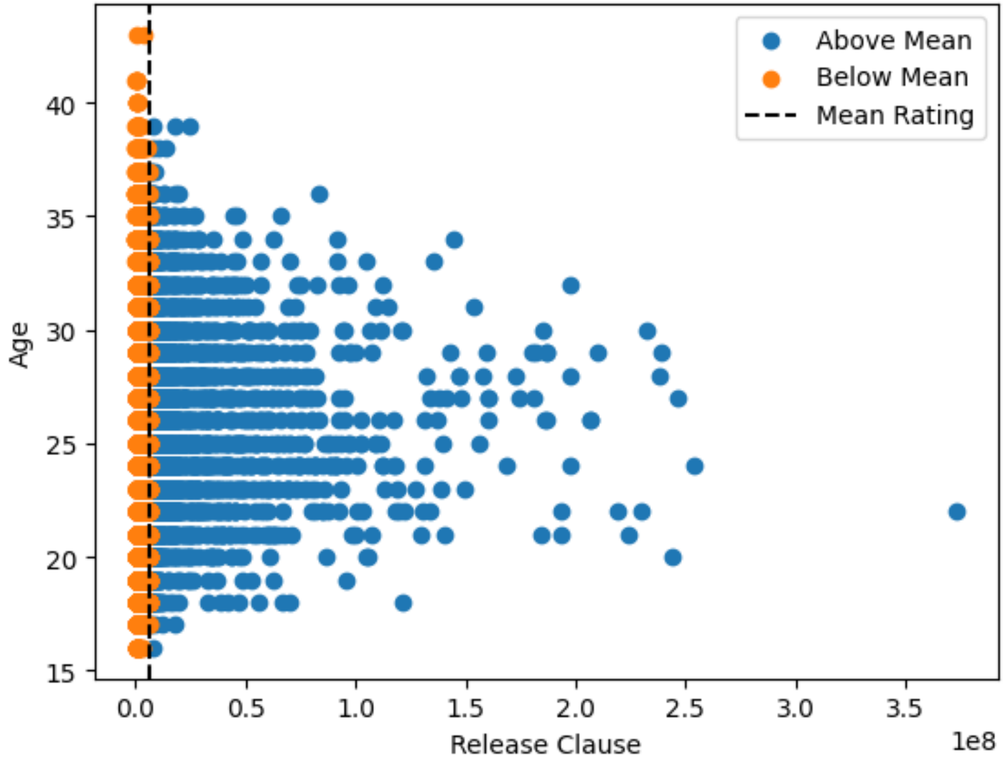
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

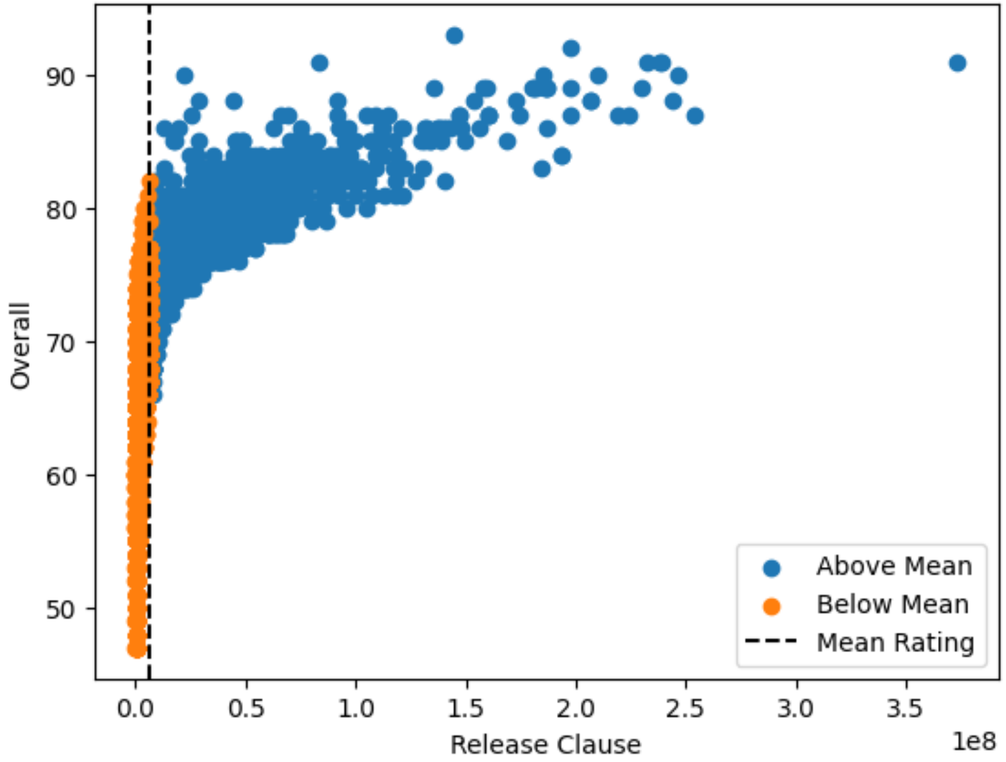
Untuk menjawab pertanyaan ini, berikut langkah yang kami lakukan:

1. Mencari hubungan untuk setiap atribut numerik dan kategorikal dengan atribut Release Clause
2. Untuk atribut numerik menggunakan scatter plot dengan membagi nilai biaya transfer minimum yang di atas dan di bawah rata-rata dan melihat hubungannya dengan atribut numeric
3. Untuk atribut categorical menggunakan pie chart. Pie chart dibuat dengan melihat hubungan biaya di atas rata-rata dengan atribut kategorikal.

Data Visualization
Pertanyaan 1

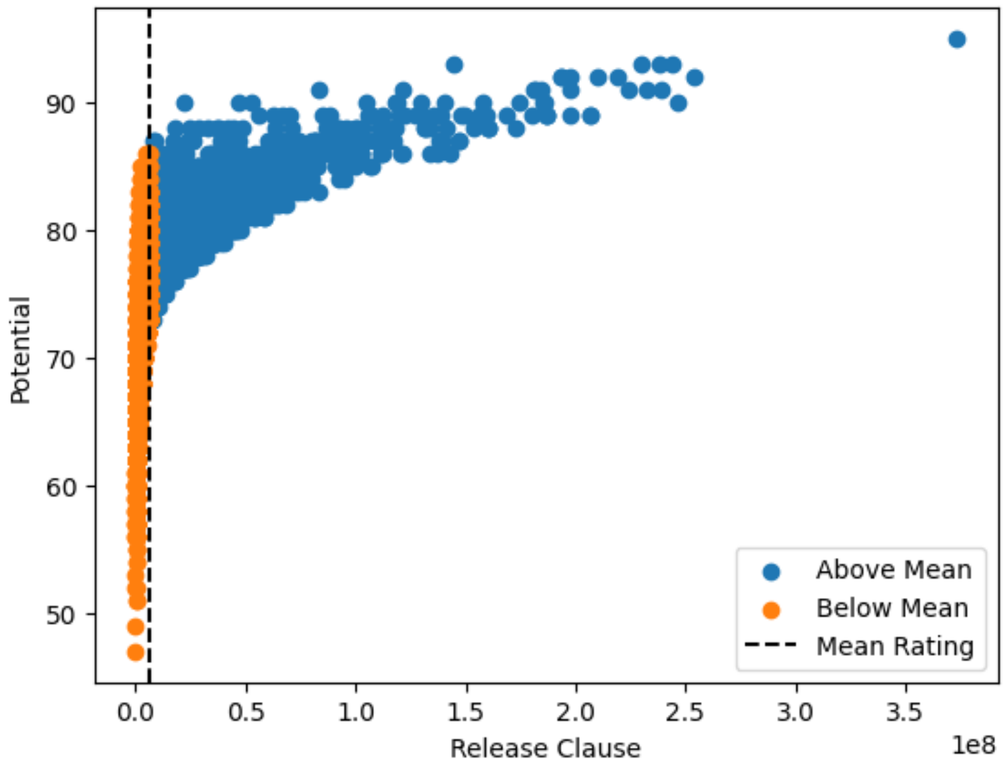
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

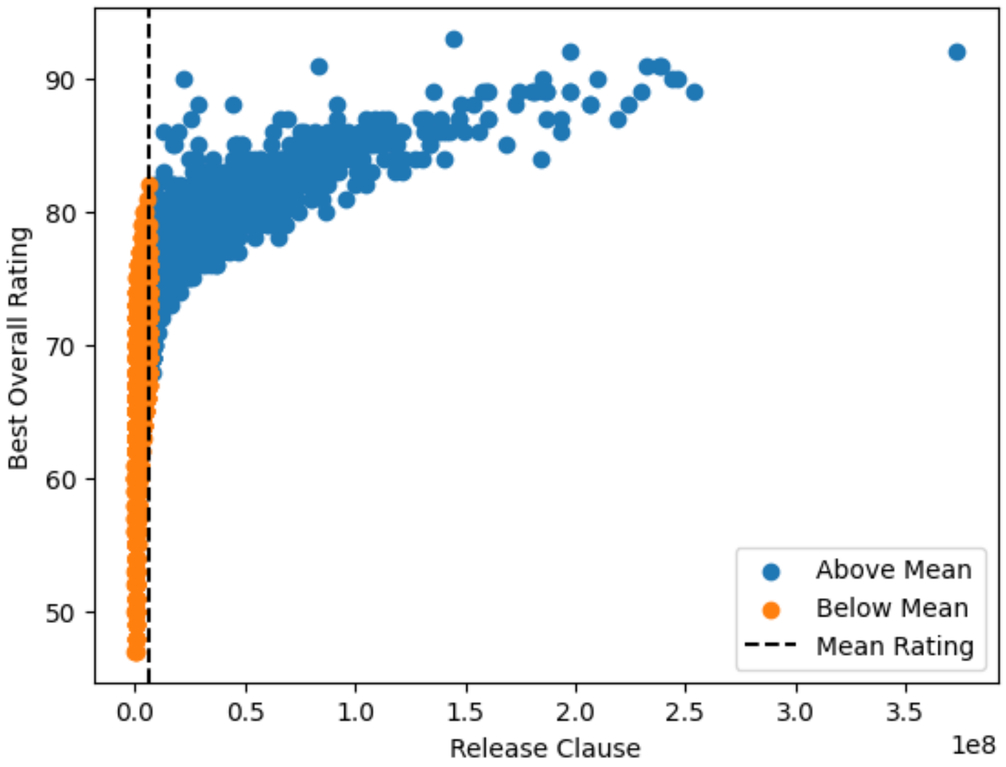




Data Visualization
Pertanyaan 1

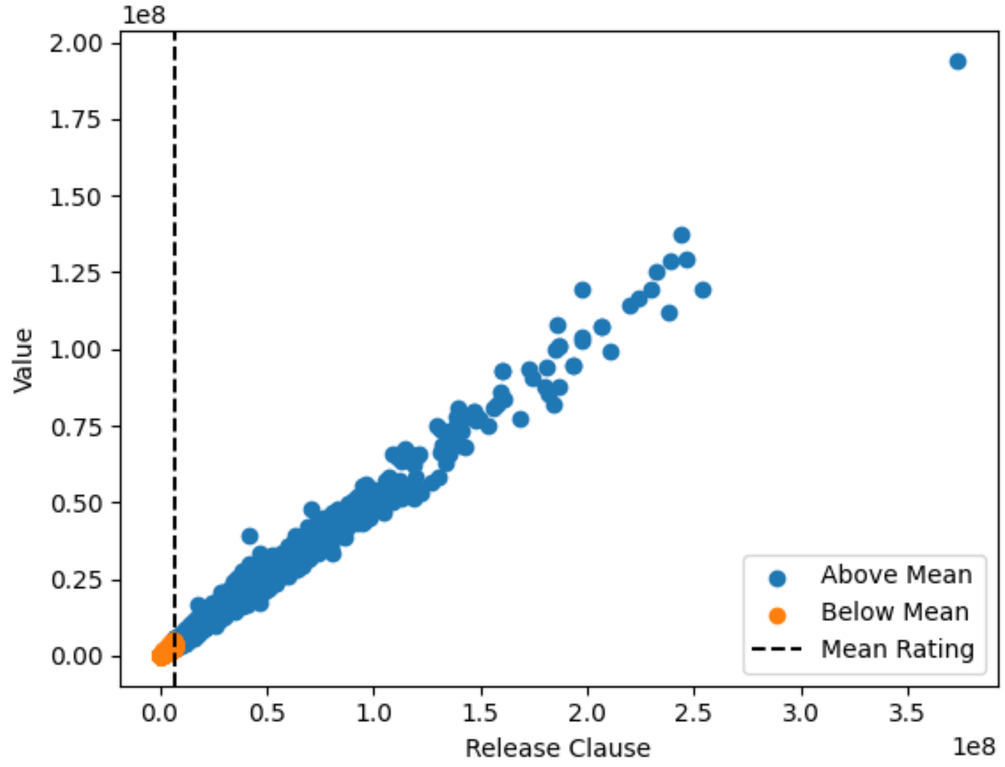
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

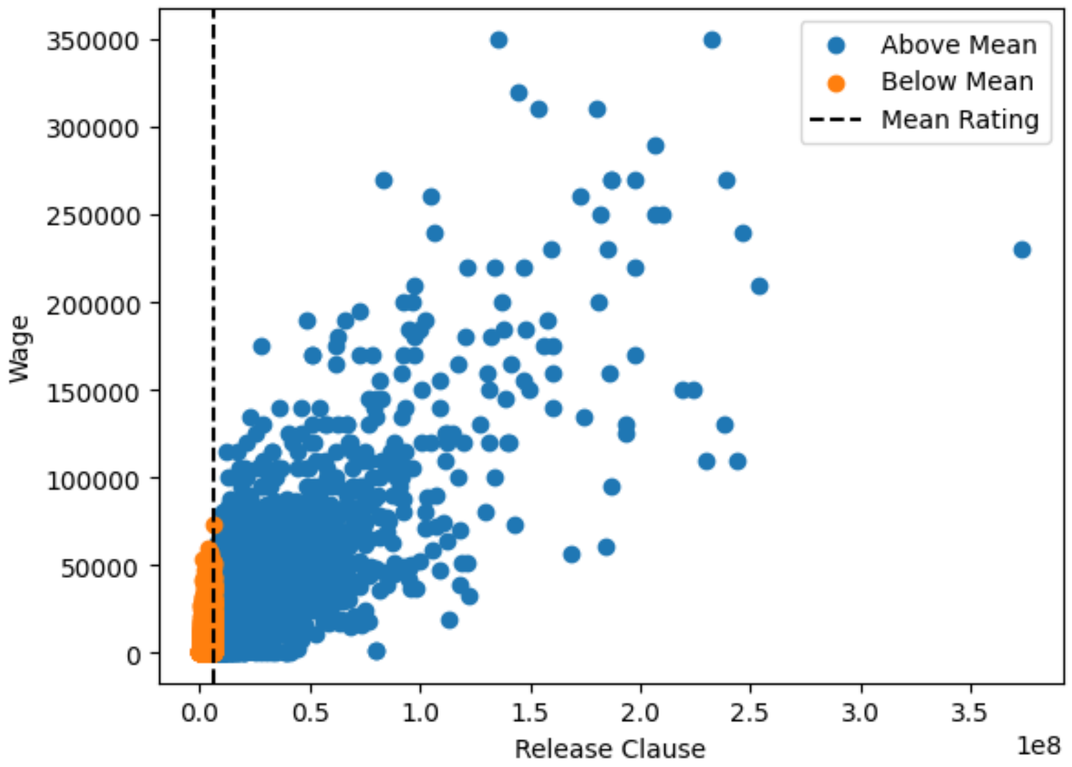




Data Visualization
Pertanyaan 1

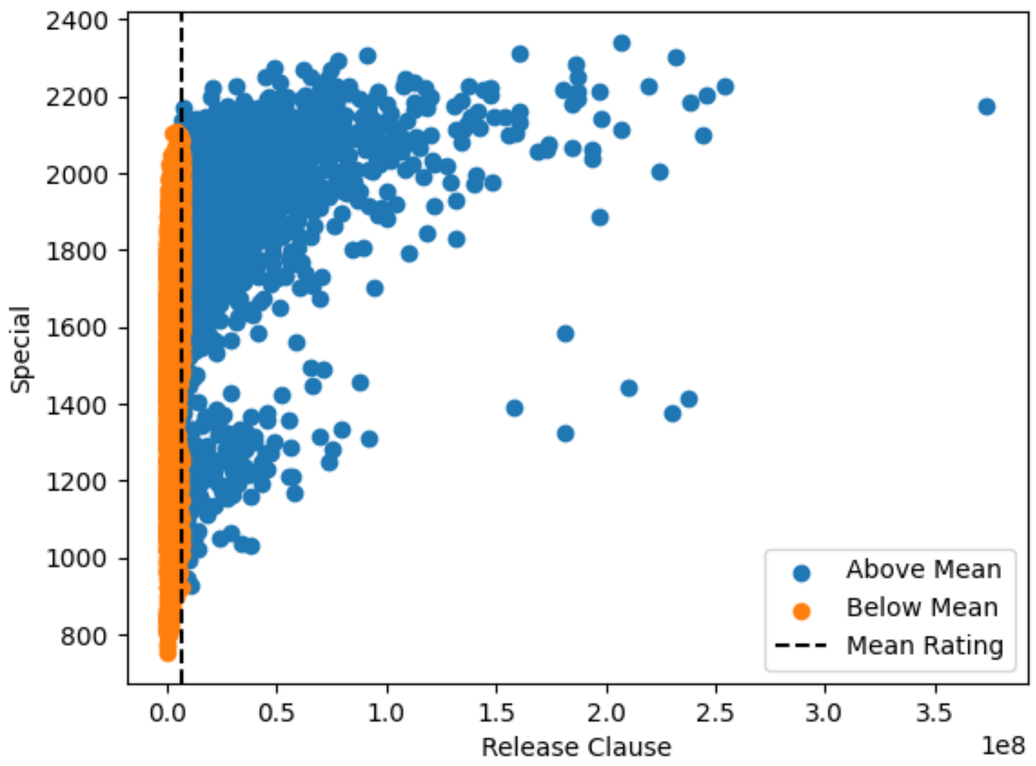
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

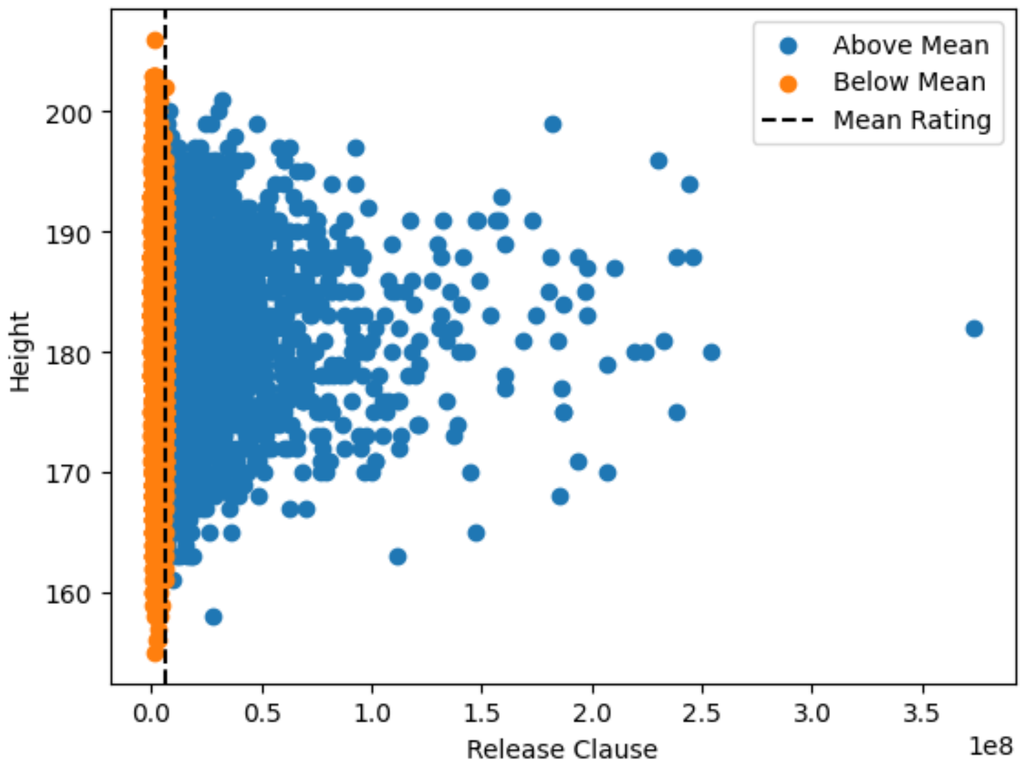




Data Visualization
Pertanyaan 1

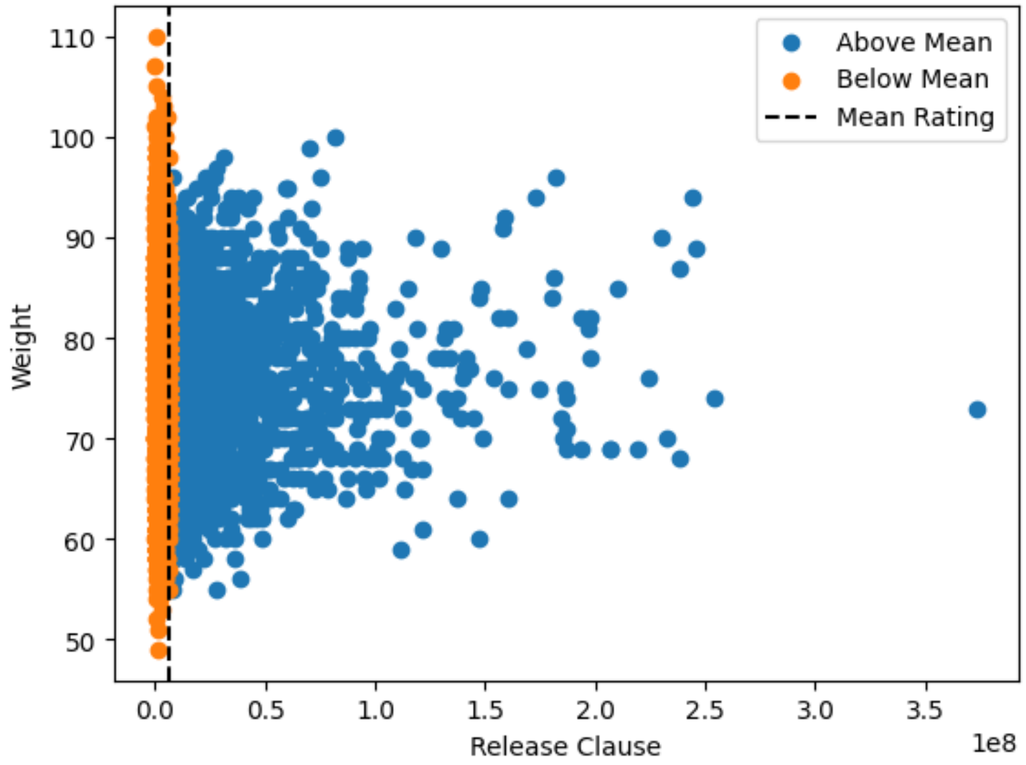
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

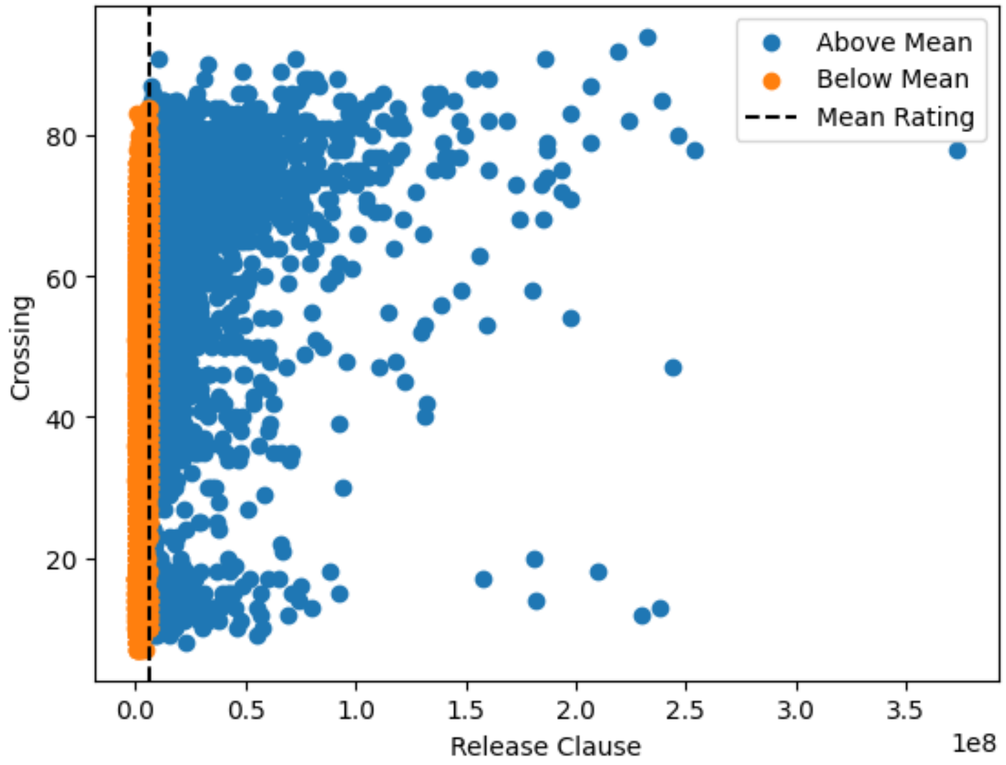




Data Visualization
Pertanyaan 1

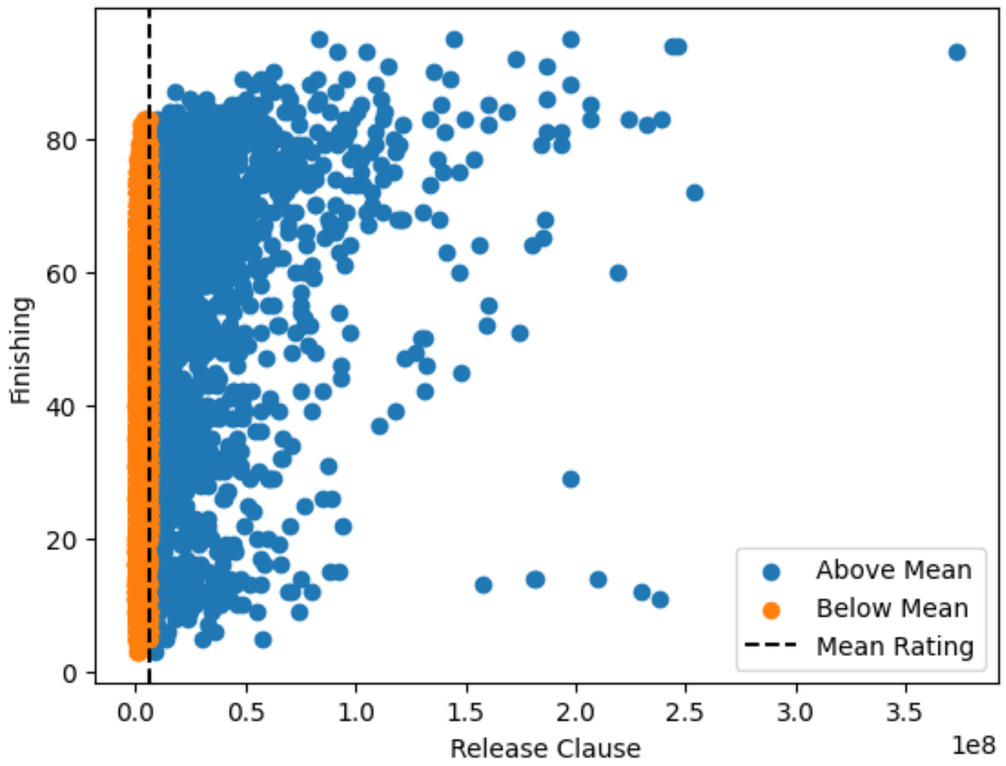
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

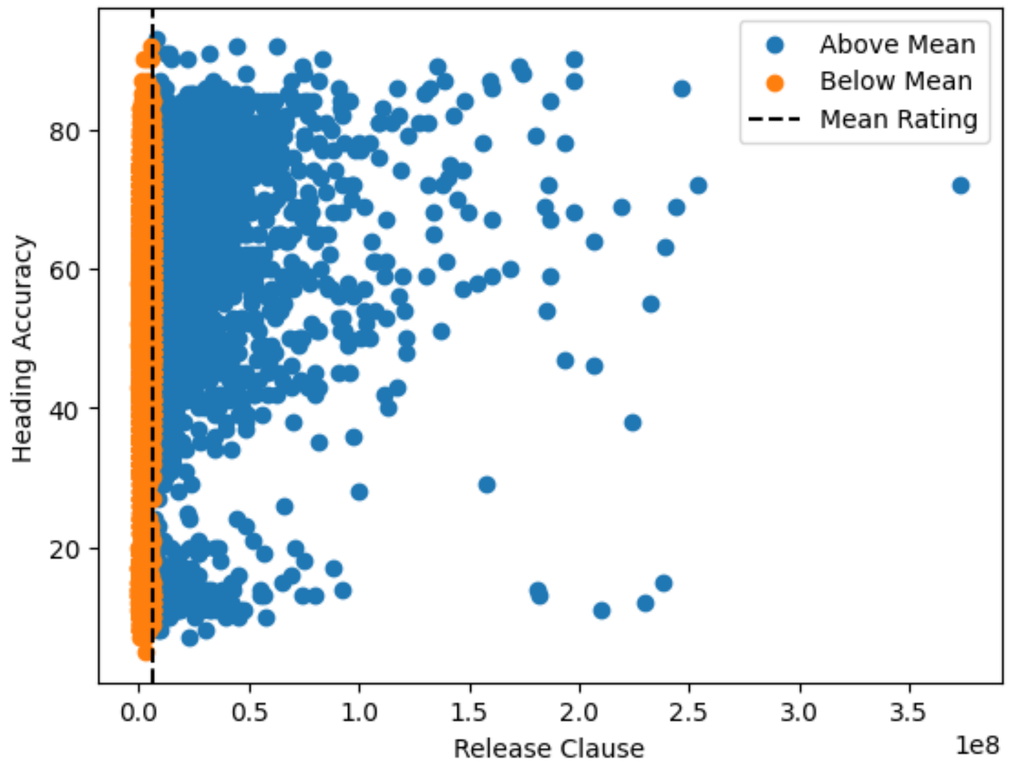




Data Visualization
Pertanyaan 1

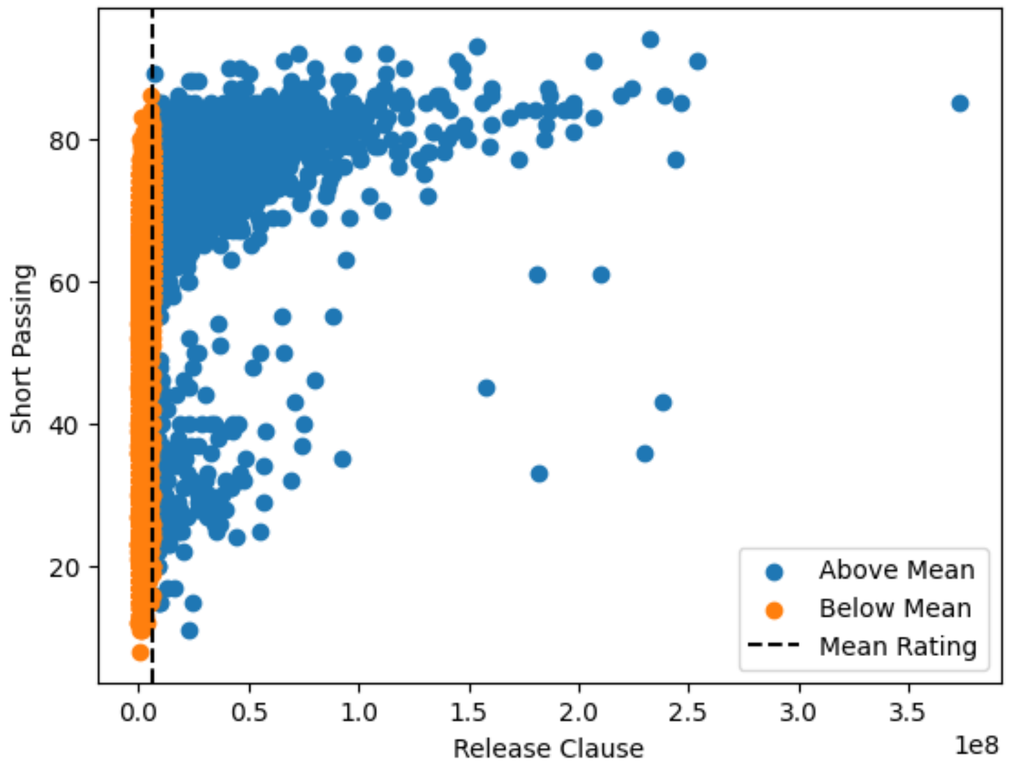
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

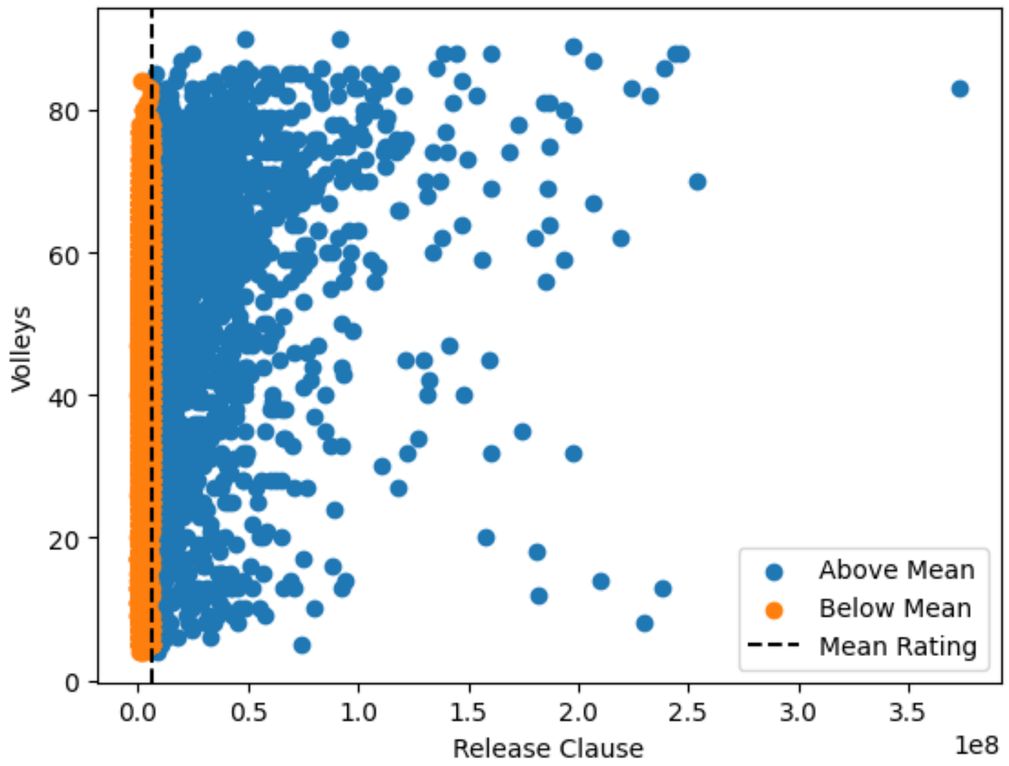




Data Visualization
Pertanyaan 1

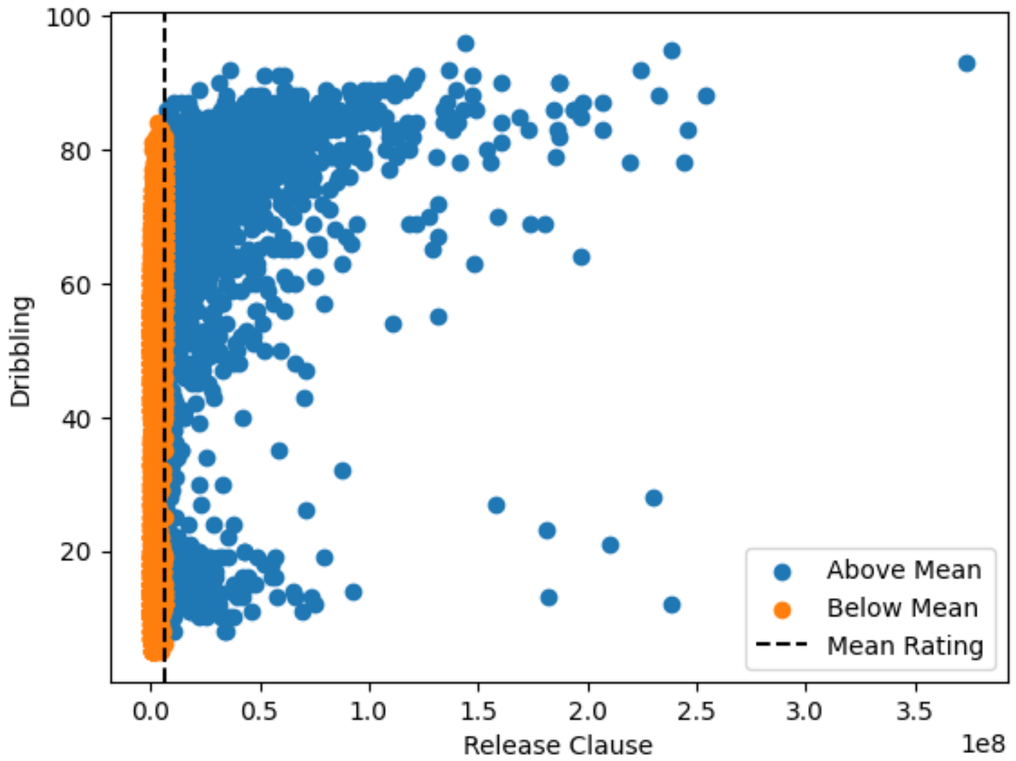
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

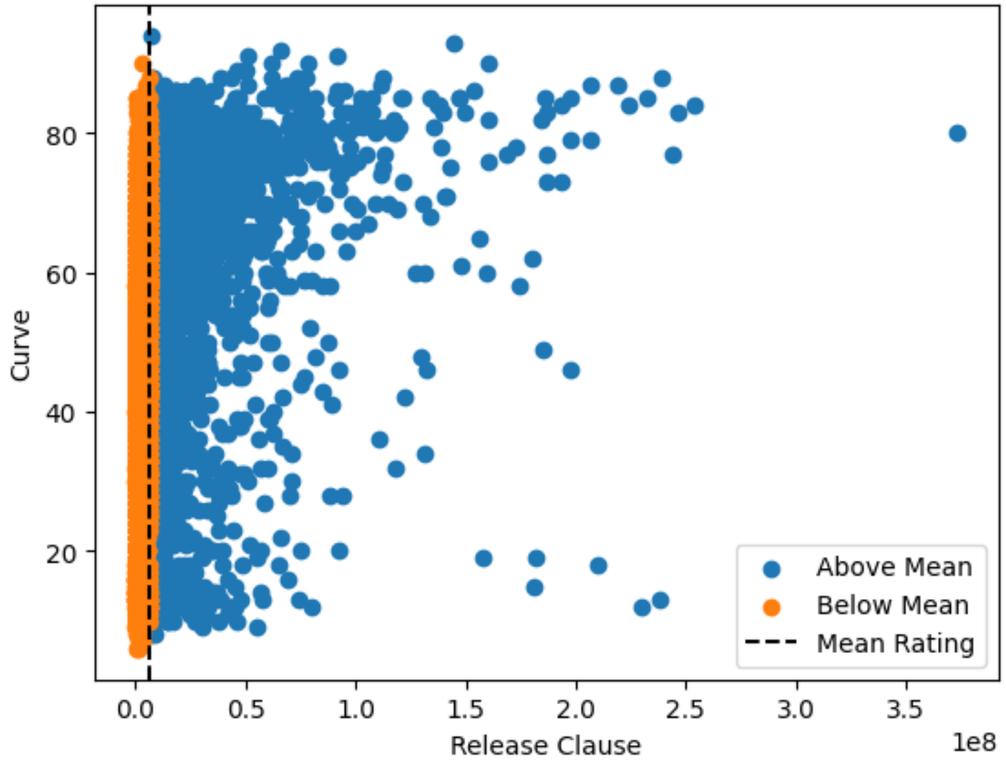




Data Visualization
Pertanyaan 1

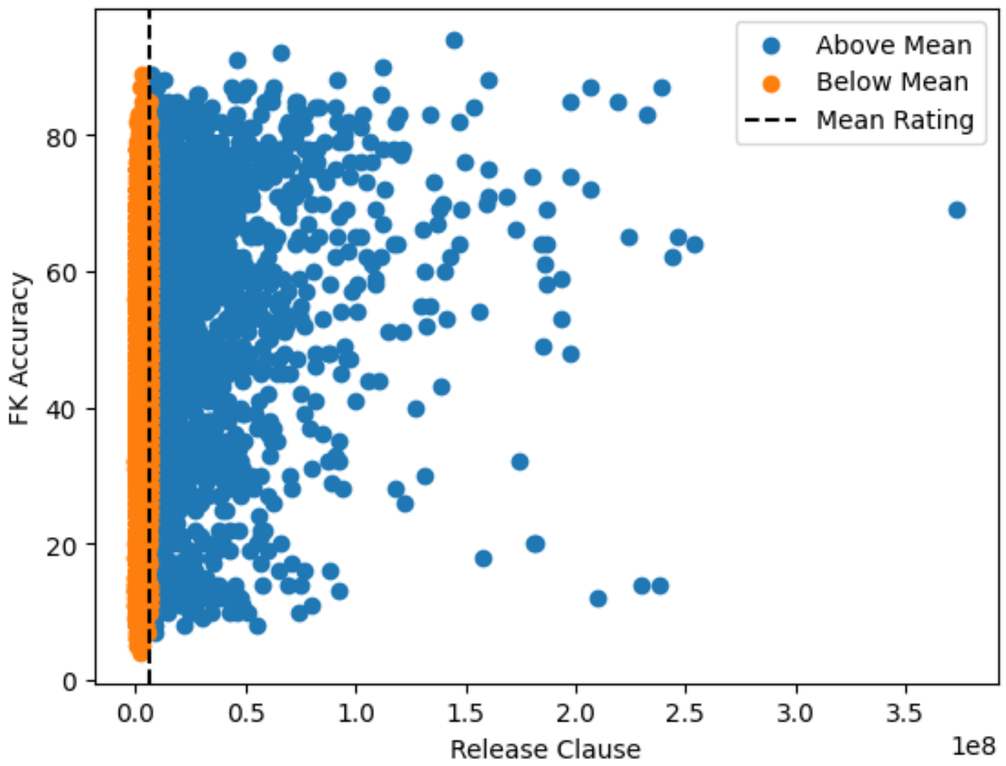
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

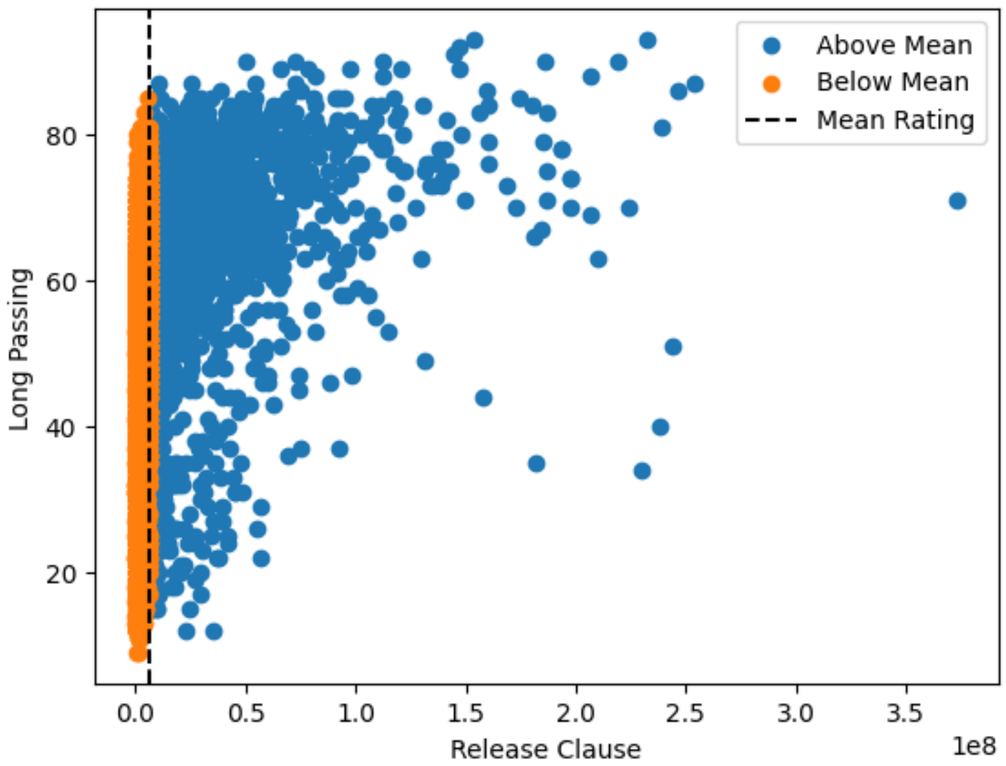




Data Visualization
Pertanyaan 1

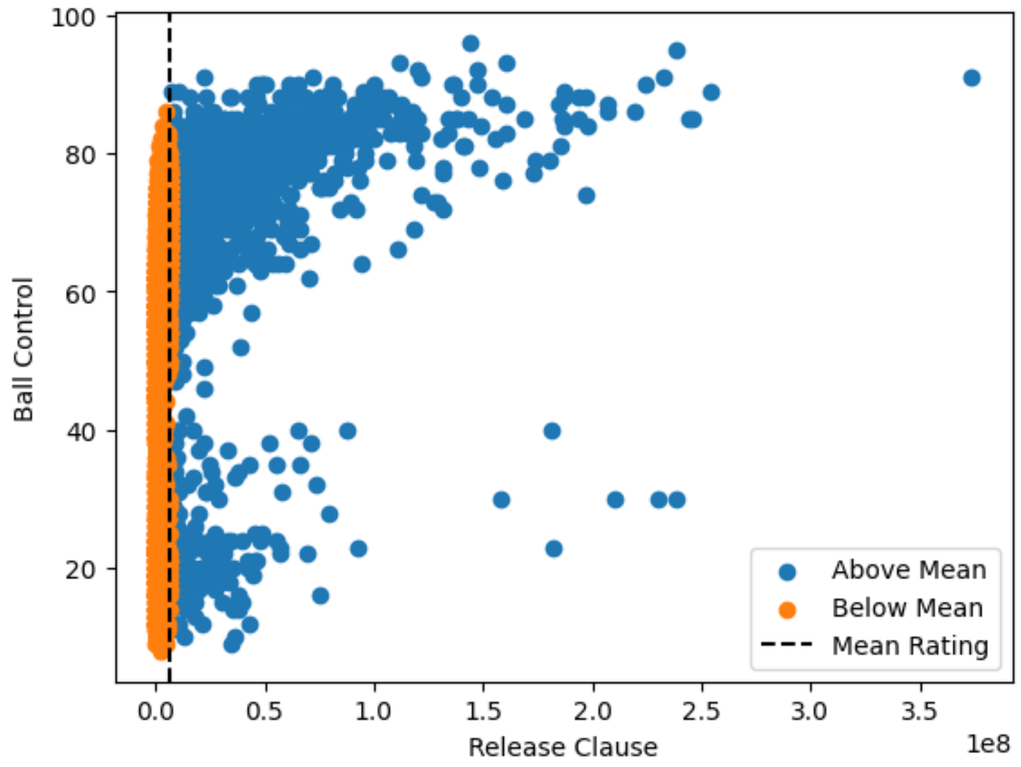
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

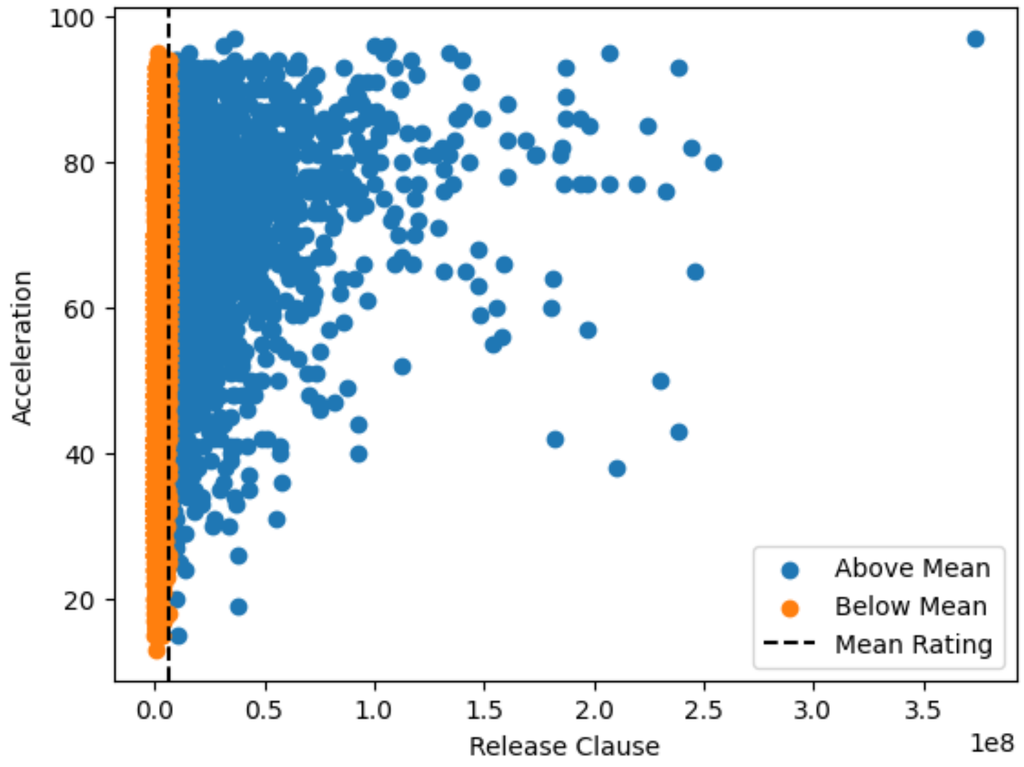




Data Visualization
Pertanyaan 1

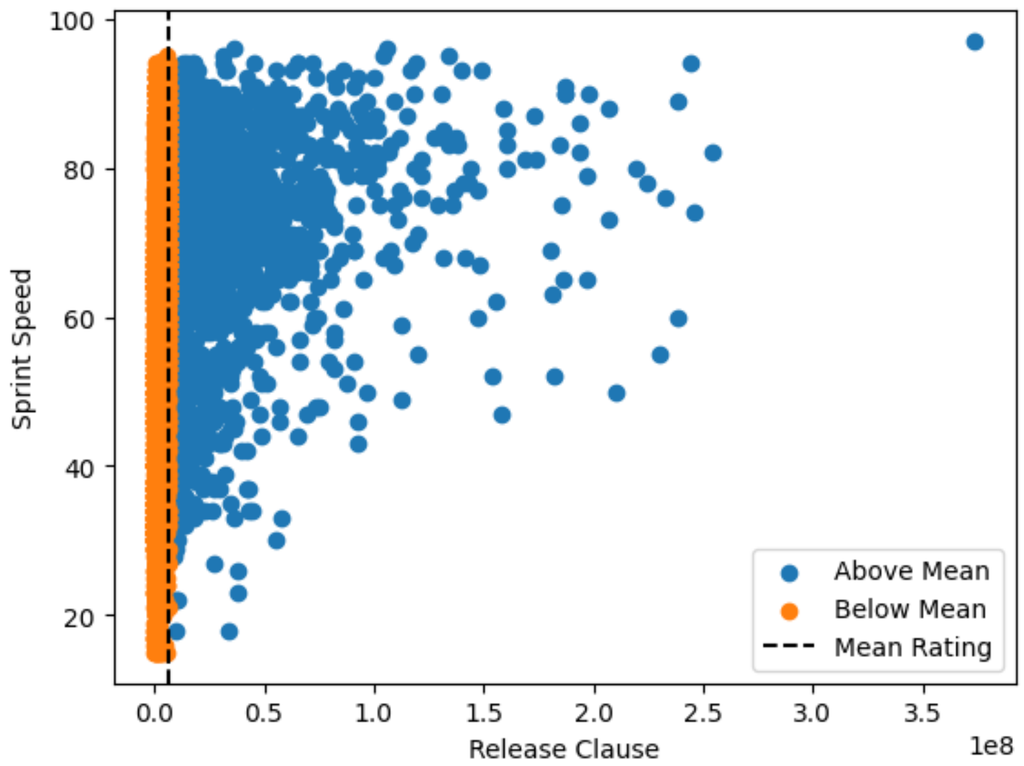
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

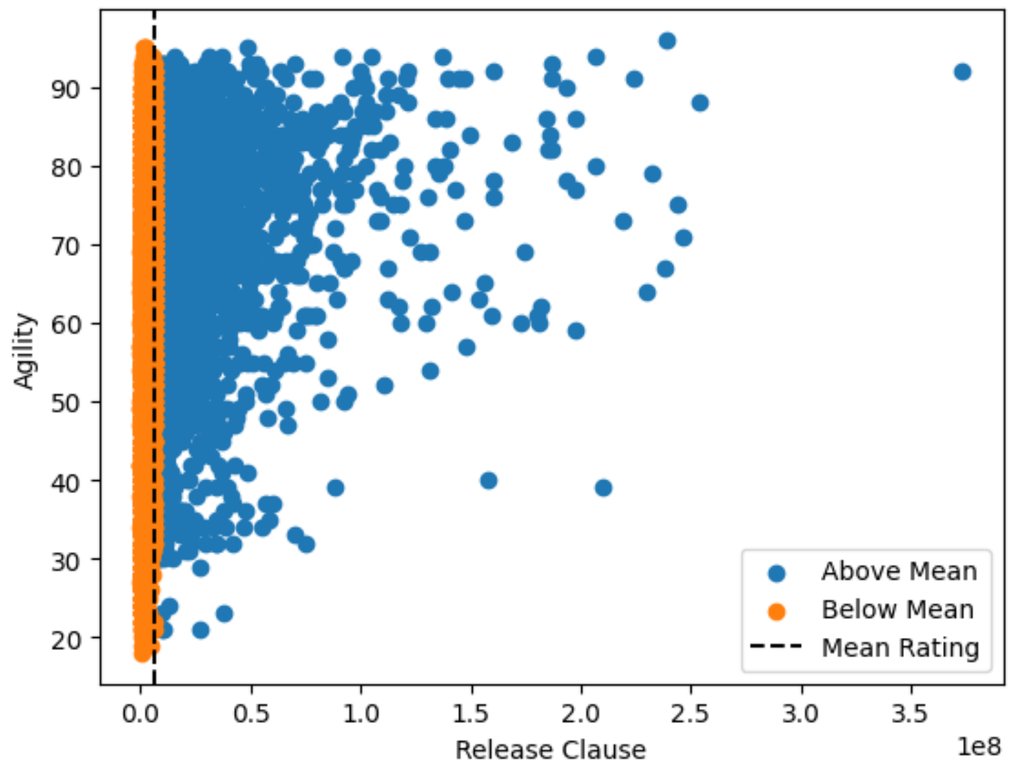




Data Visualization
Pertanyaan 1

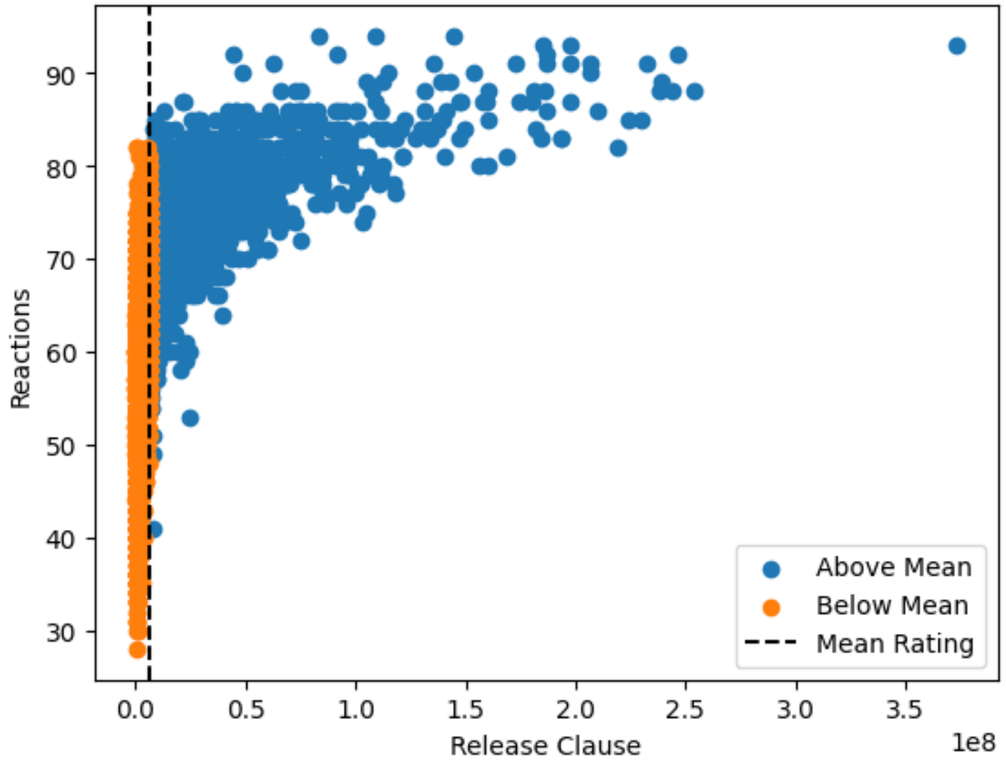
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

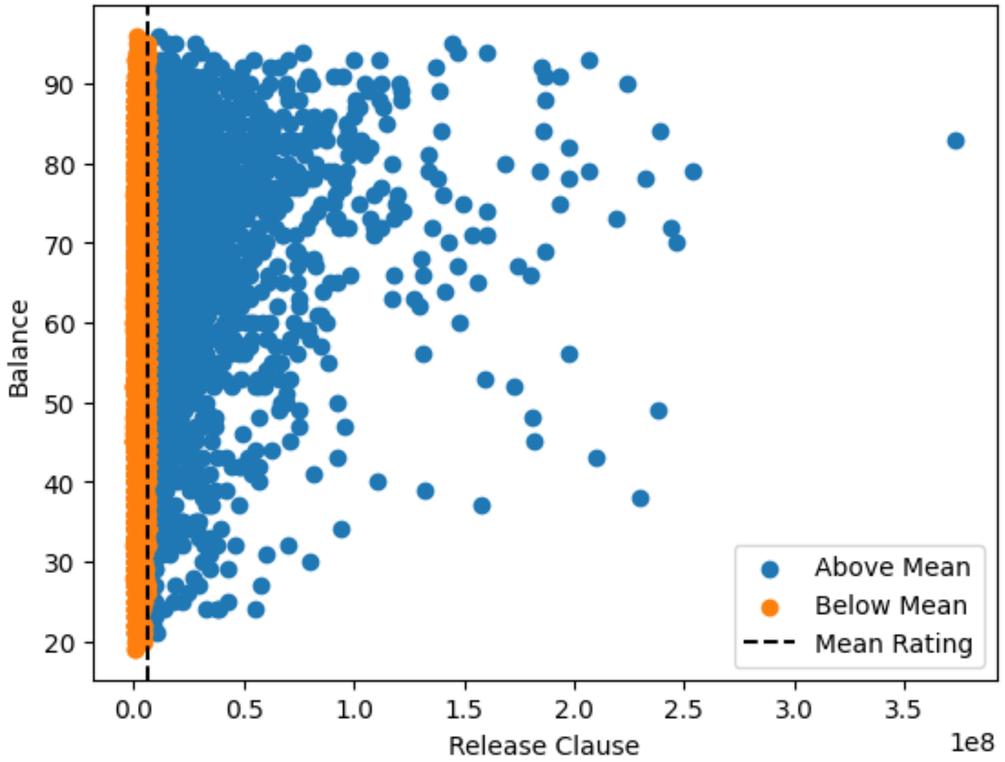




Data Visualization
Pertanyaan 1

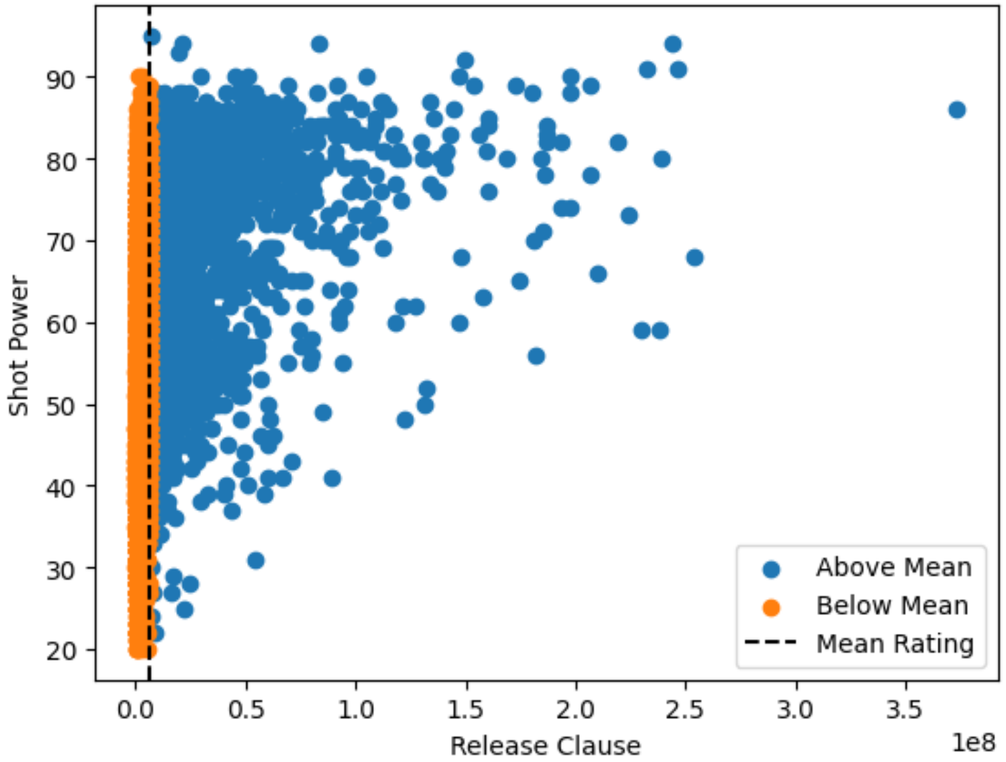
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

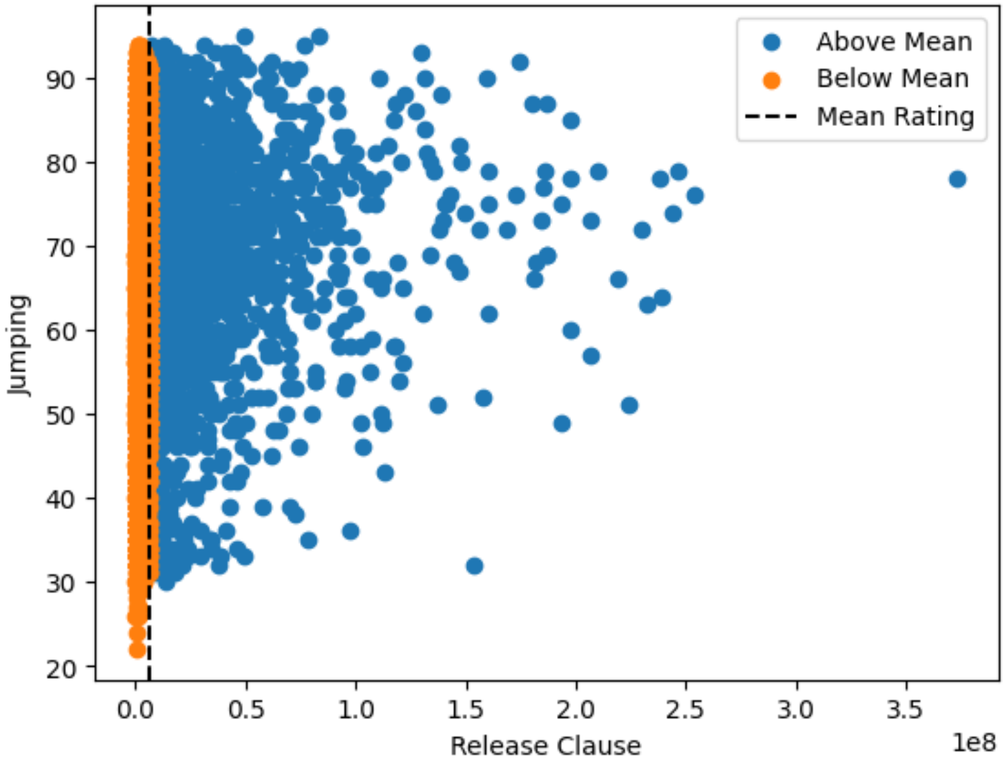




Data Visualization
Pertanyaan 1

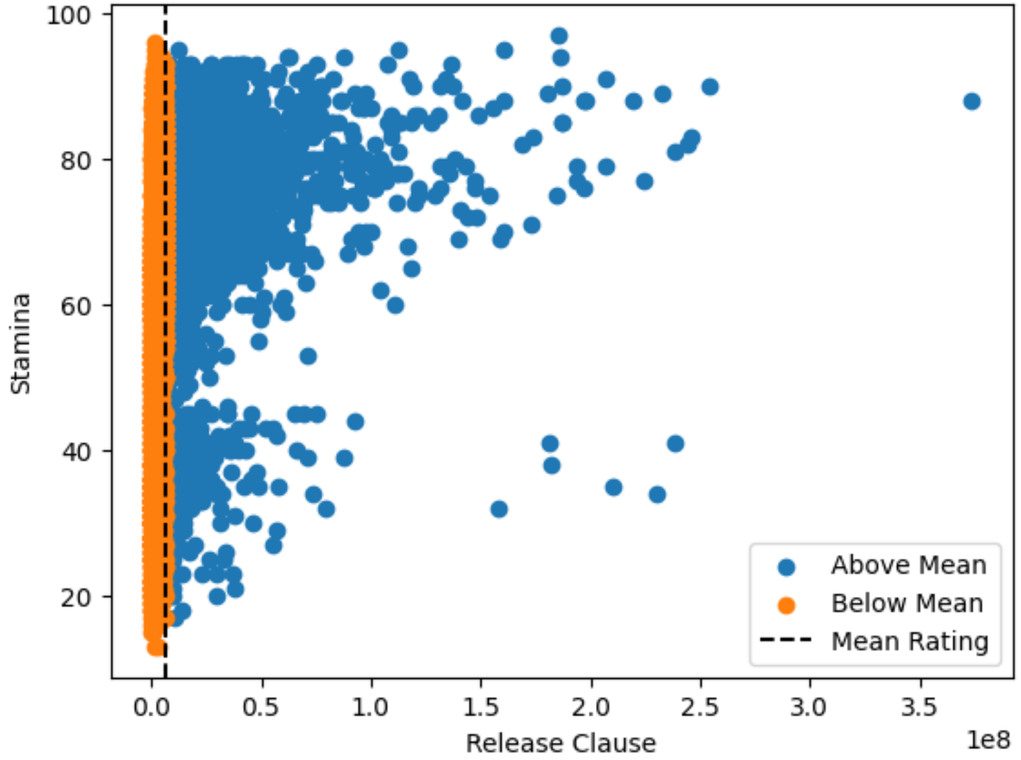
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

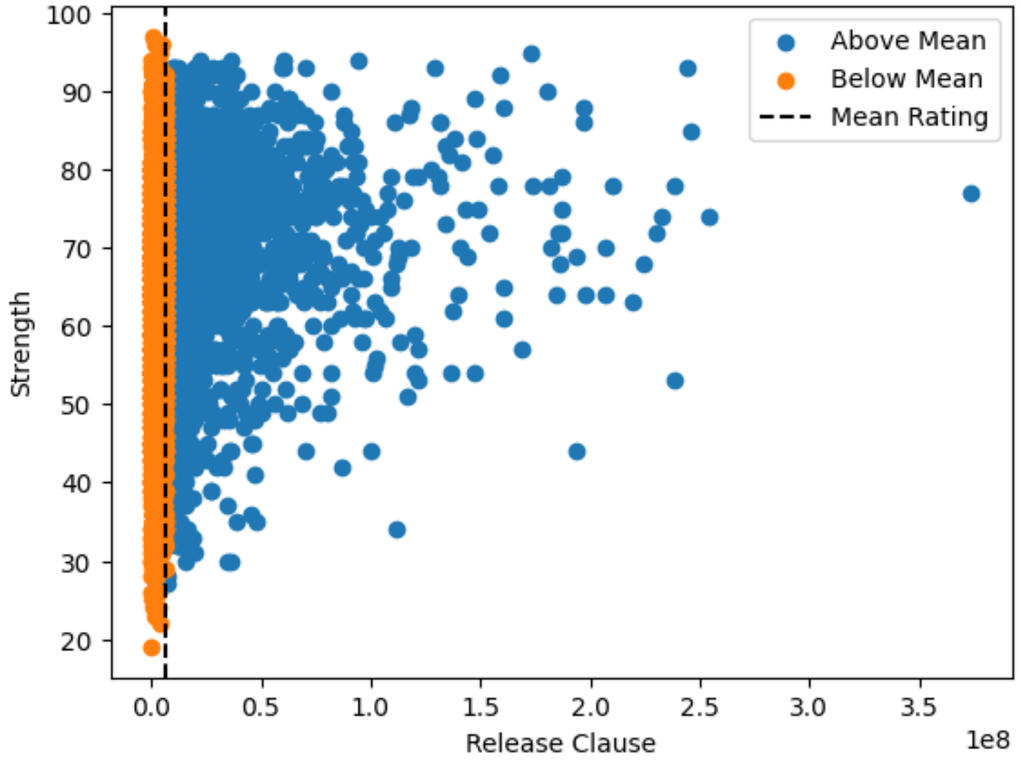




Data Visualization
Pertanyaan 1

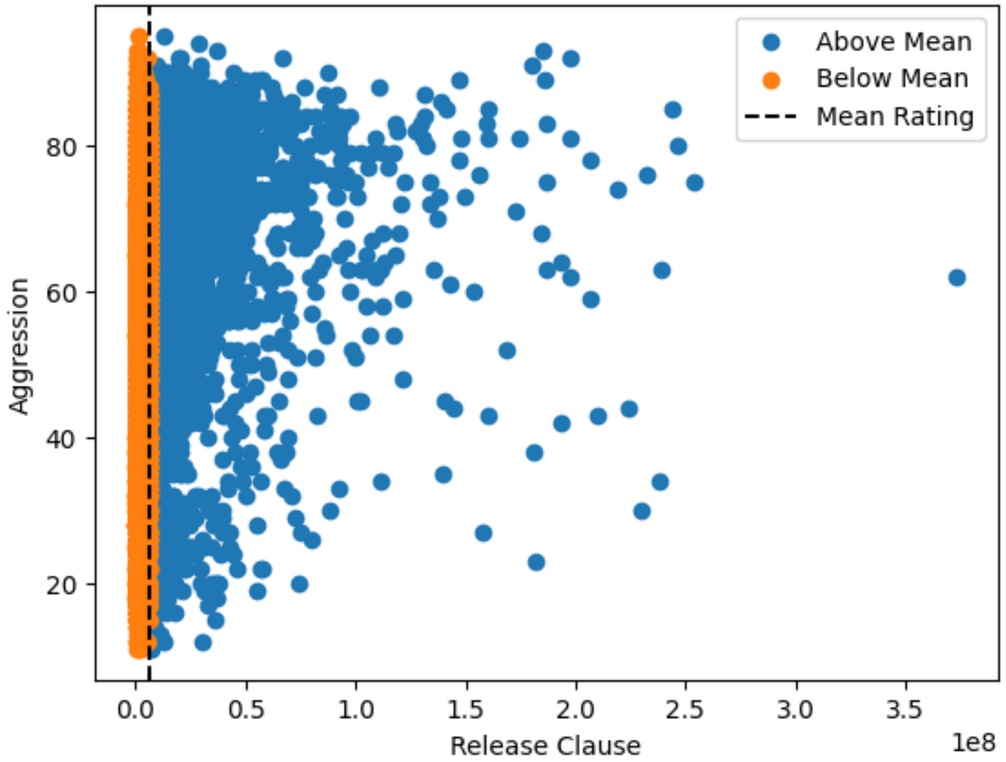
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

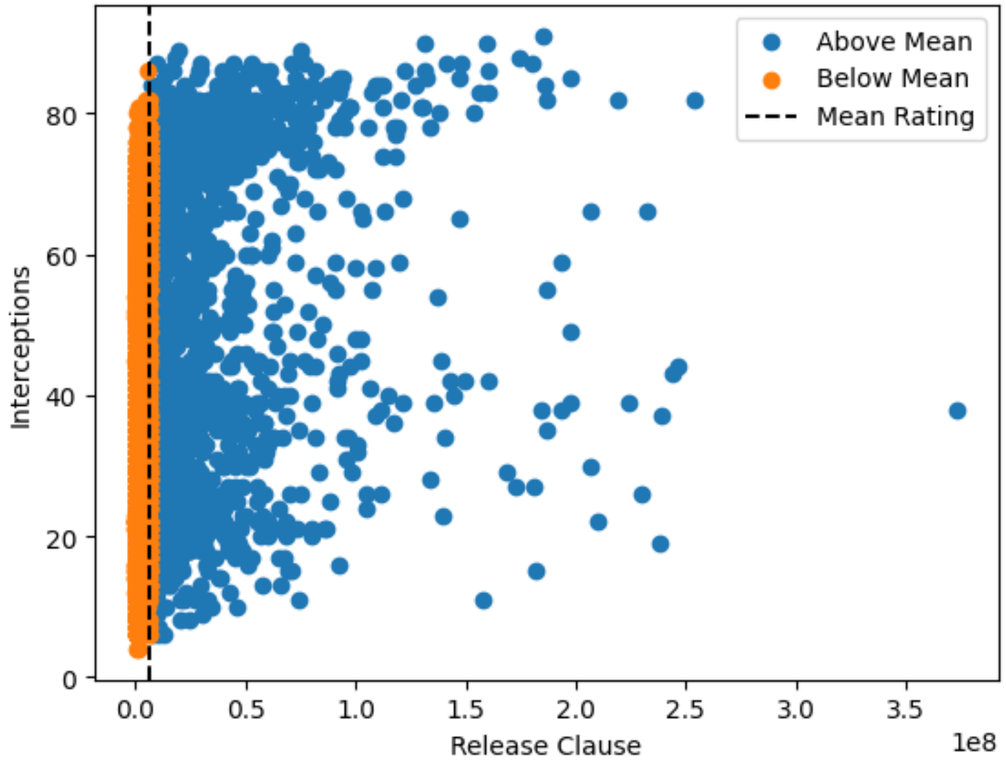




Data Visualization
Pertanyaan 1

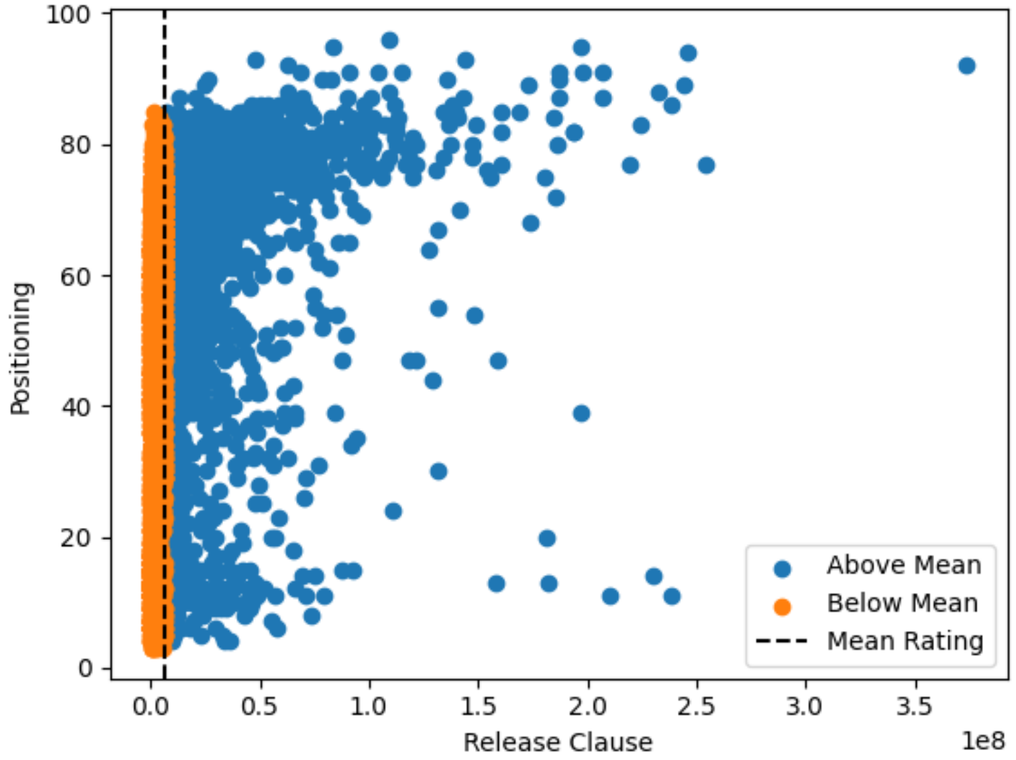
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

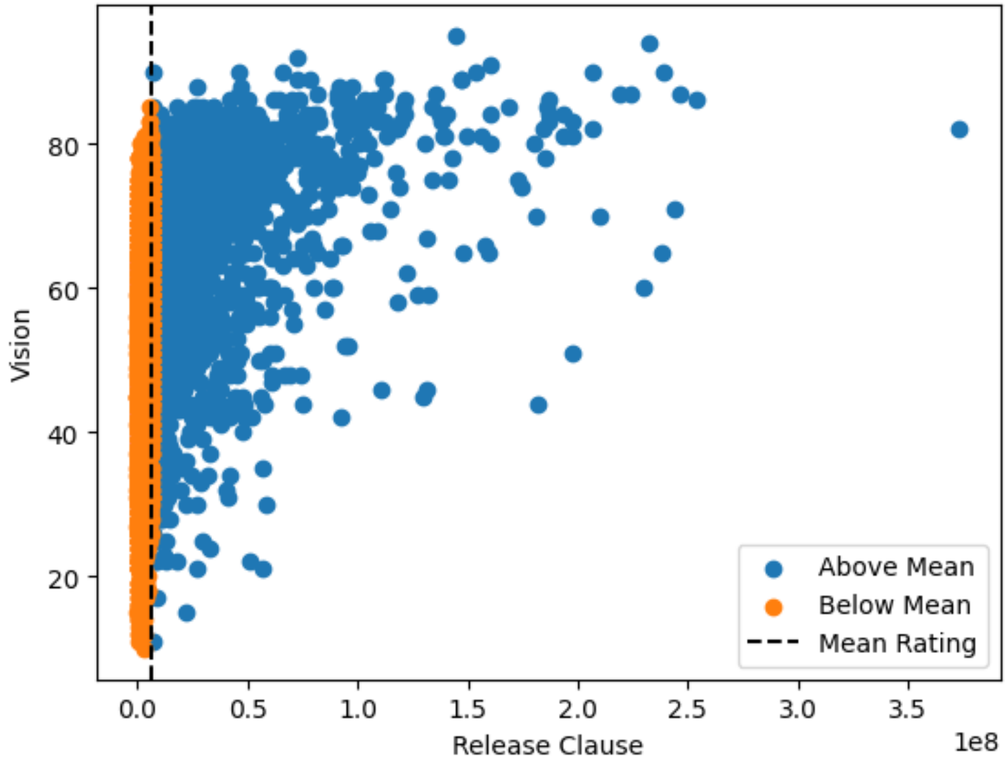




Data Visualization
Pertanyaan 1

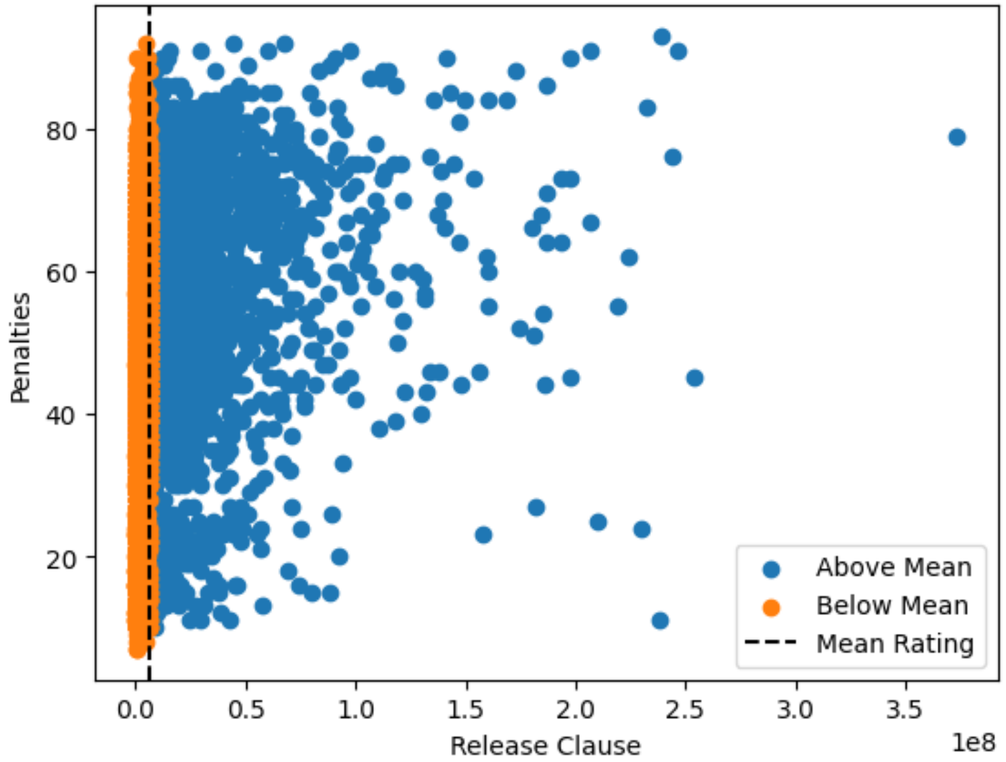
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

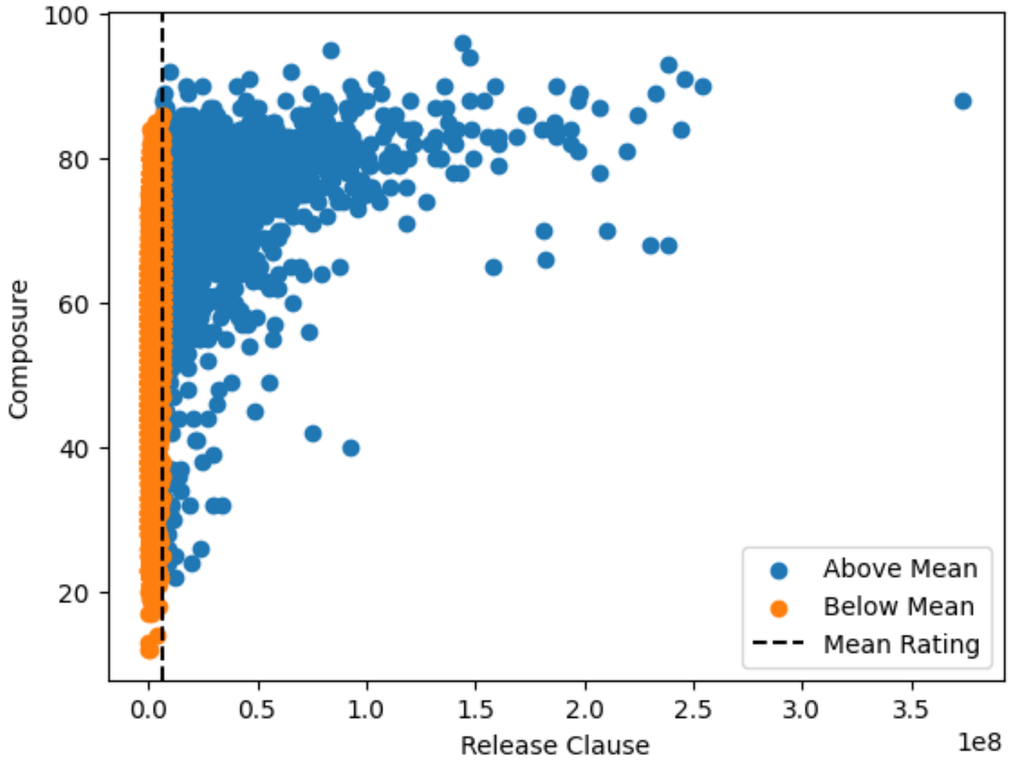




Data Visualization
Pertanyaan 1

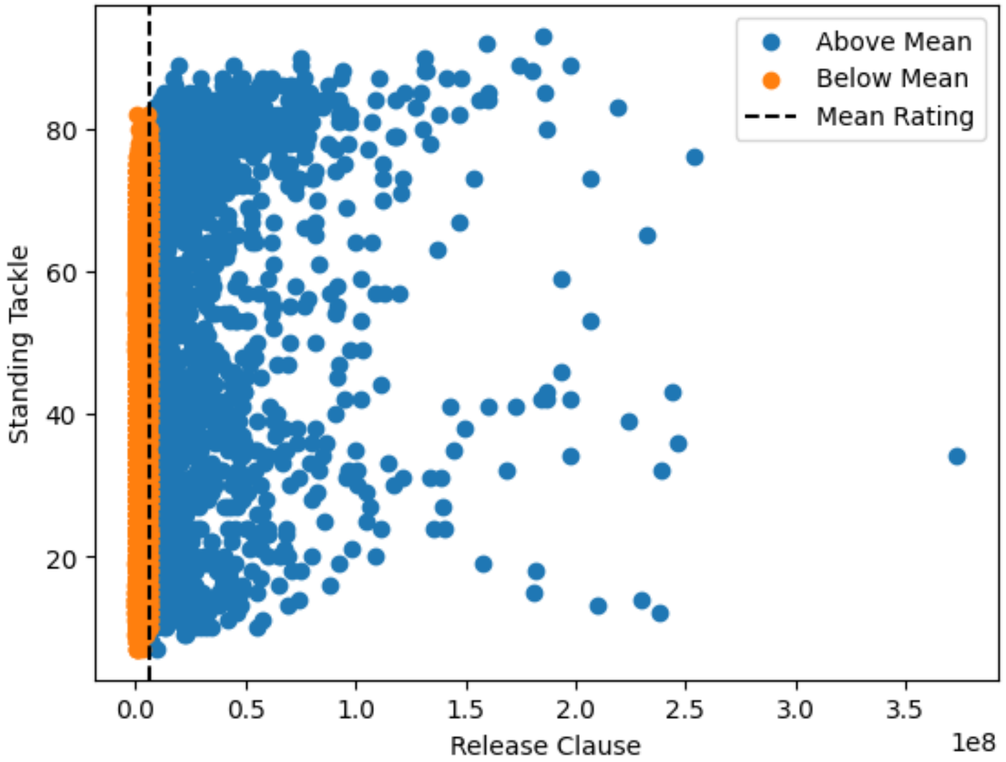
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

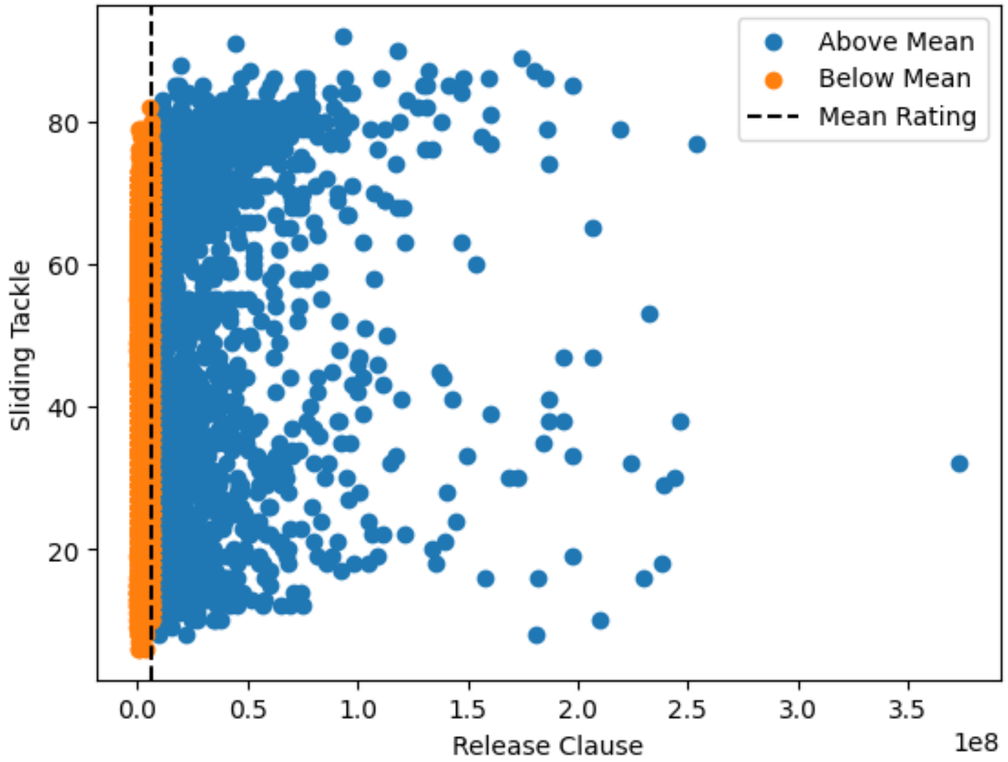




Data Visualization
Pertanyaan 1

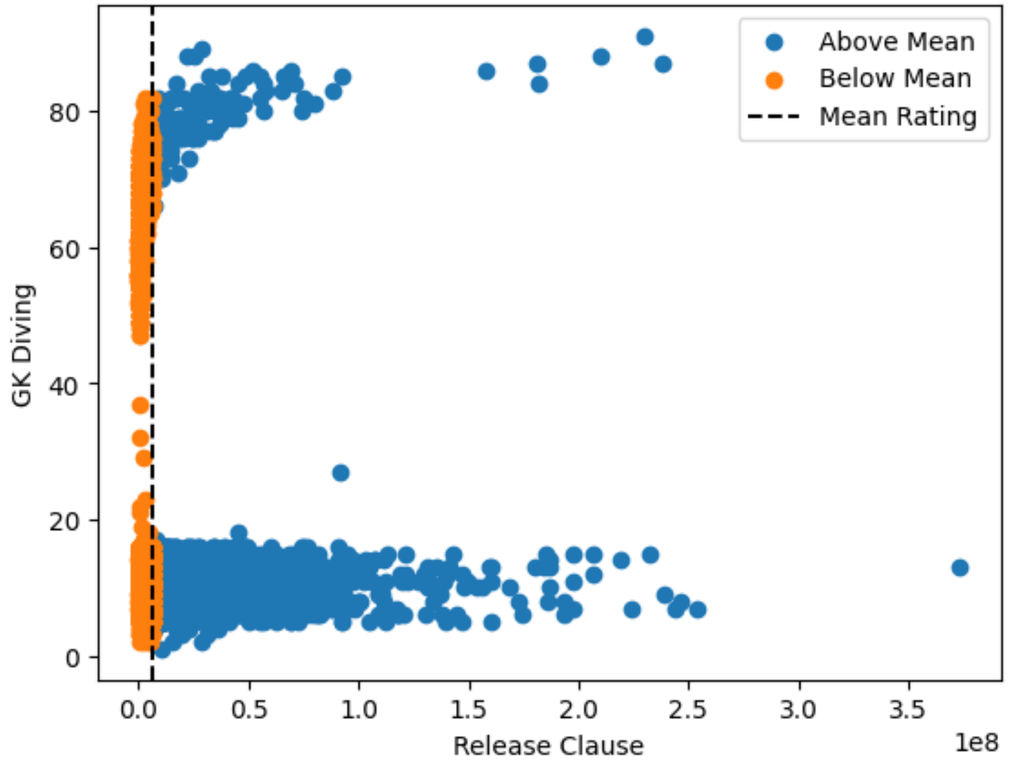
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

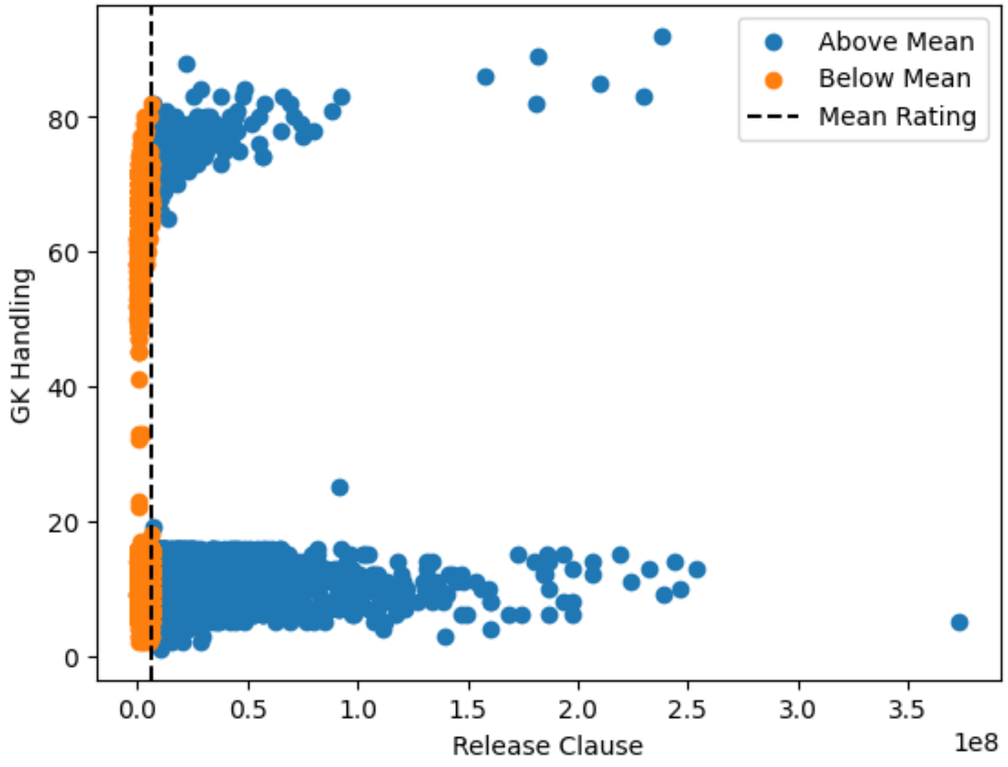




Data Visualization
Pertanyaan 1

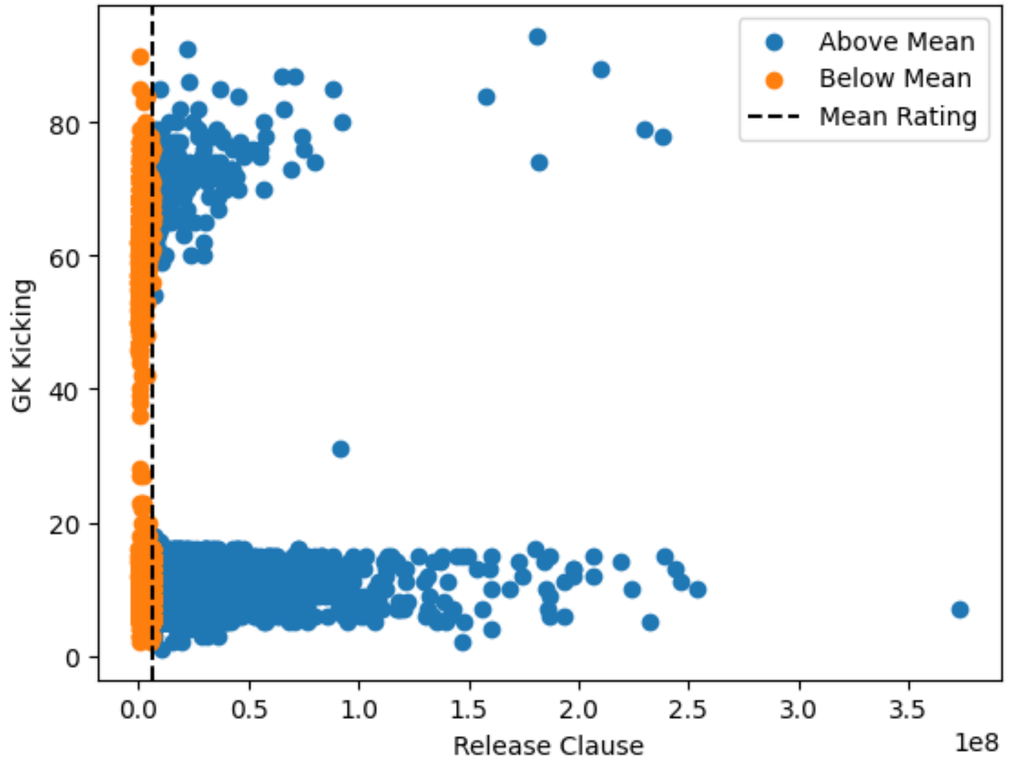
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

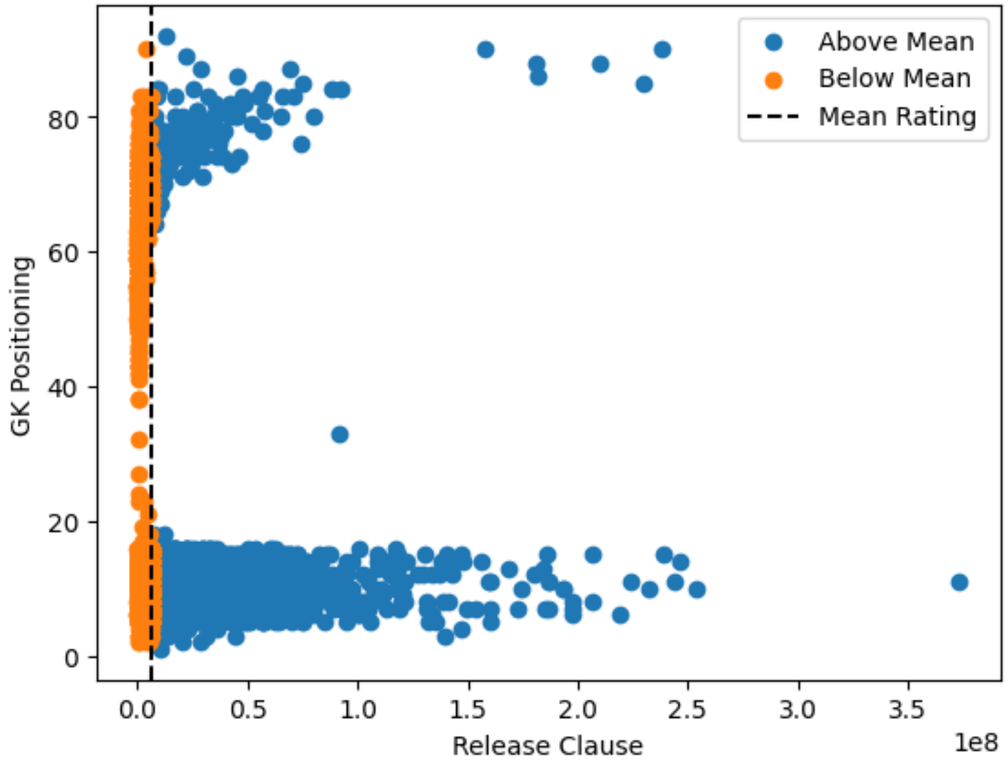




Data Visualization
Pertanyaan 1

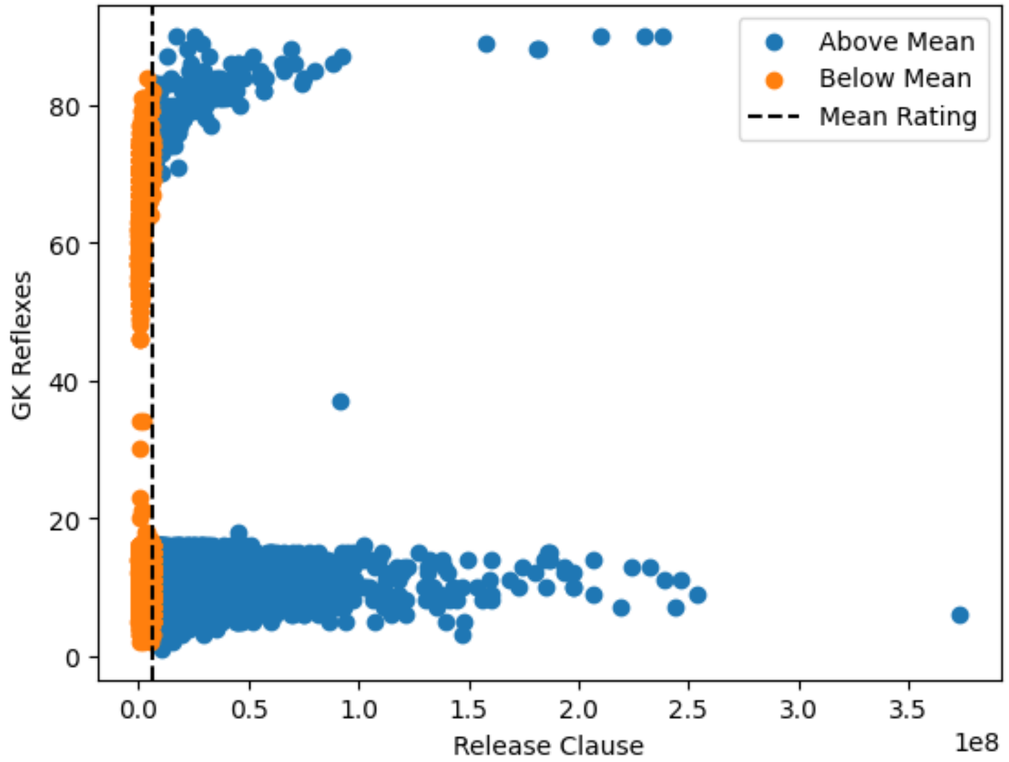
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

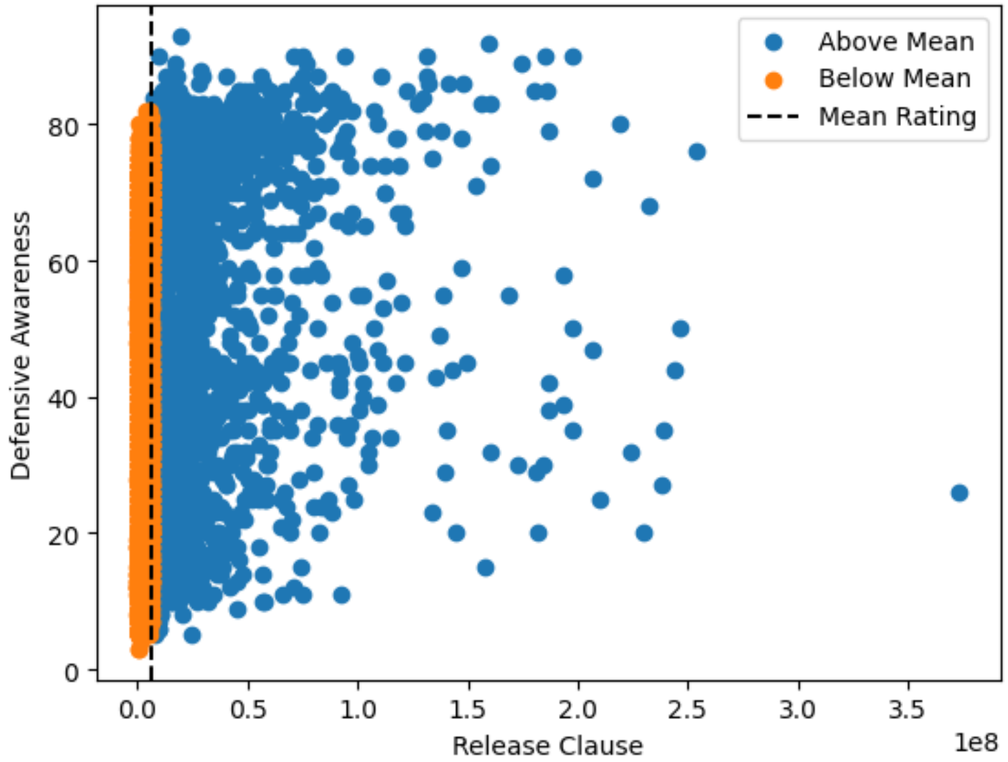




Data Visualization
Pertanyaan 1

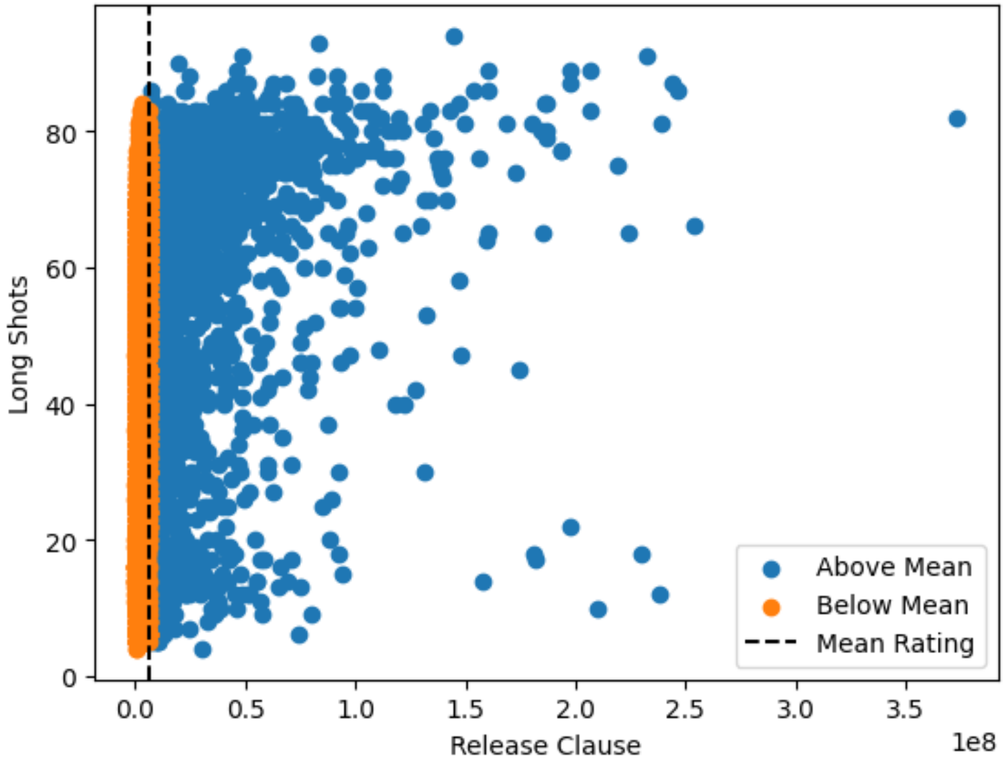
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?





Data Visualization
Pertanyaan 1

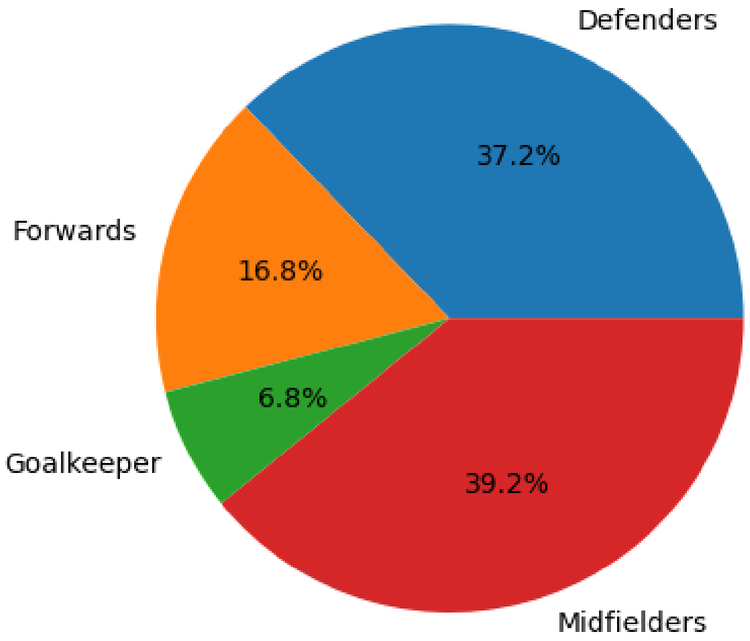
1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?



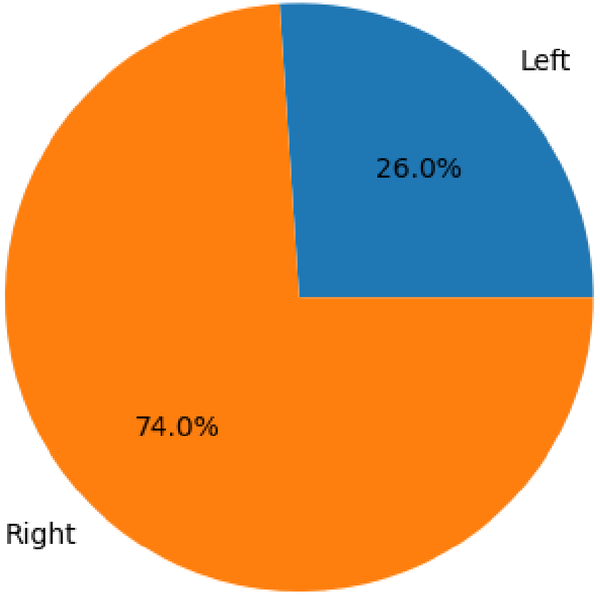
Data Visualization
Pertanyaan 1

1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

Position



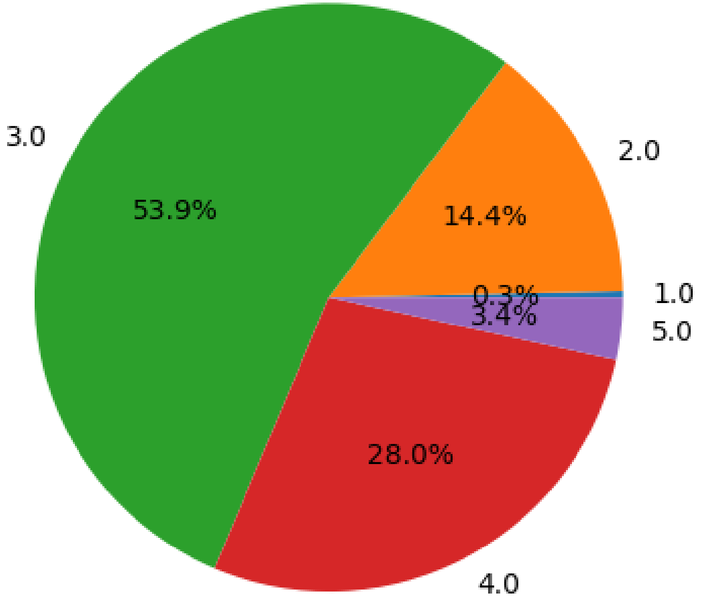
Preferred foot



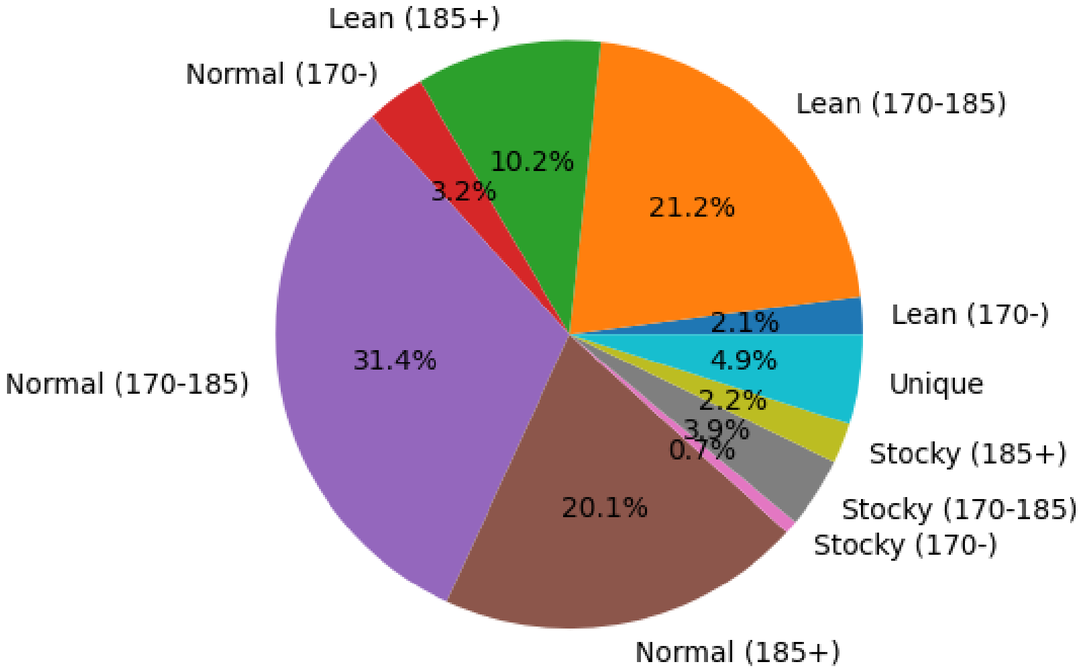
Data Visualization
Pertanyaan 1

1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

Weak foot



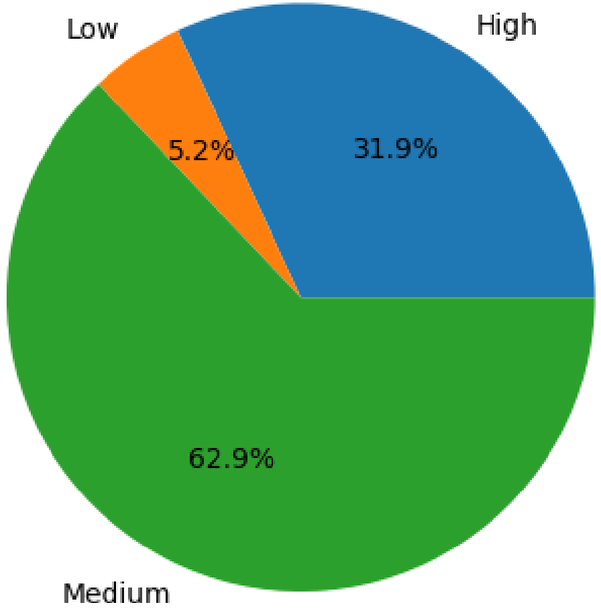
Body Type



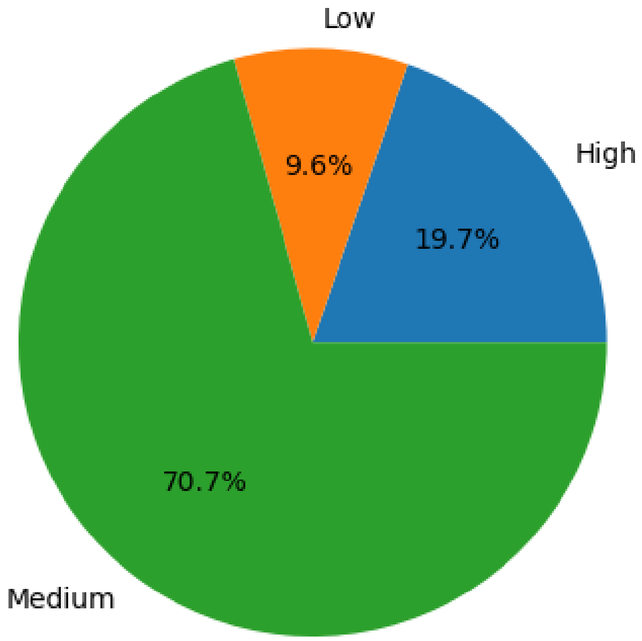
Data Visualization
Pertanyaan 1

1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

Attacking work rate

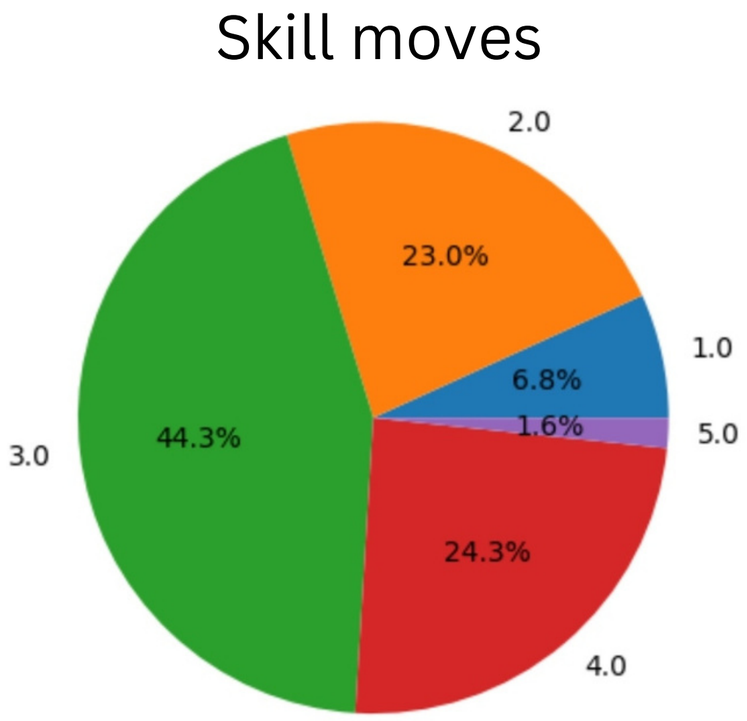


Defensive work rate



Data Visualization
Pertanyaan 1

1. Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?



Data Visualization
Pertanyaan 1

Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

Kesimpulannya adalah ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata antara lain

1. mempunyai nilai overall yang tinggi
2. mempunyai nilai potential yang tinggi
3. mempunyai nilai best overall rating yang tinggi.
4. mempunyai nilai value yang tinggi
5. mempunyai nilai wage yang tinggi

Data Visualization
Pertanyaan 1

Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

- mempunyai posisi midfielders
- mempunyai preffered foot right
- mempunyai weak foot 5.0
- mempunyai body type normal
- mempunyai Attacking Work Rate medium
- mempunyai Defensive Work Rate medium
- mempunyai skill moves 3.0
- mempunyai nilai reaction yang tinggi

Data Visualization
Pertanyaan 1

Apa ciri-ciri pemain sepakbola yang memiliki biaya transfer minimum yang diatas rata-rata?

- **Atribut yang tidak disebutkan tidak dapat dikatakan sebagai ciri-ciri tersebut karena tidak mempunyai nilai khas**

Data Visualization
Pertanyaan 2

Apakah pemain dengan umur lebih dari umur 30 masih bermain dengan kerja keras yang tinggi?

Berdasarkan deskripsi dataset yang disediakan di Scele, diketahui bahwa kerja keras pemain diwakilkan dengan **dua fitur,** yaitu **attacking work rate** (tingkat kerja keras seorang pemain dalam menyerang) dan **defensive work rate** (tingkat kerja keras seorang pemain dalam bertahan).

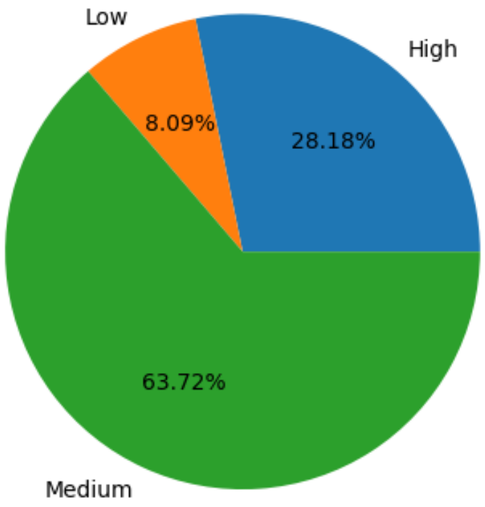
Untuk menjawab pertanyaan ini, berikut langkah yang kami lakukan:

1. Membuat dataframe baru dengan memfilter row dengan pemain berusia lebih dari 30 tahun
2. Membuat **pie chart** untuk menunjukkan distribusi tingkat kerja keras pemain yang berusia di atas 30 tahun

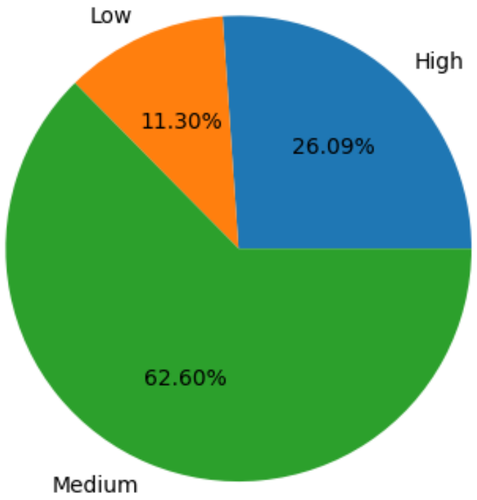
Data Visualization
Pertanyaan 2

Apakah pemain dengan umur lebih dari umur 30 masih bermain dengan kerja keras yang tinggi?

Persentase Tiap Attacking Work Rate pada Pemain Berusia Lebih dari 30 Tahun



Persentase Tiap Defensive Work Rate pada Pemain Berusia Lebih dari 30 Tahun



Data Visualization
Pertanyaan 2

Apakah pemain dengan umur lebih dari umur 30 masih bermain dengan kerja keras yang tinggi?

Berdasarkan **pie chart**, dapat dilihat bahwa rata-rata pemain berusia lebih dari 30 tahun bermain dengan kerja keras yang sedang (medium) dengan persentase 63.72% untuk Attacking Work Rate dan 62.60% untuk Defensive Work Rate.

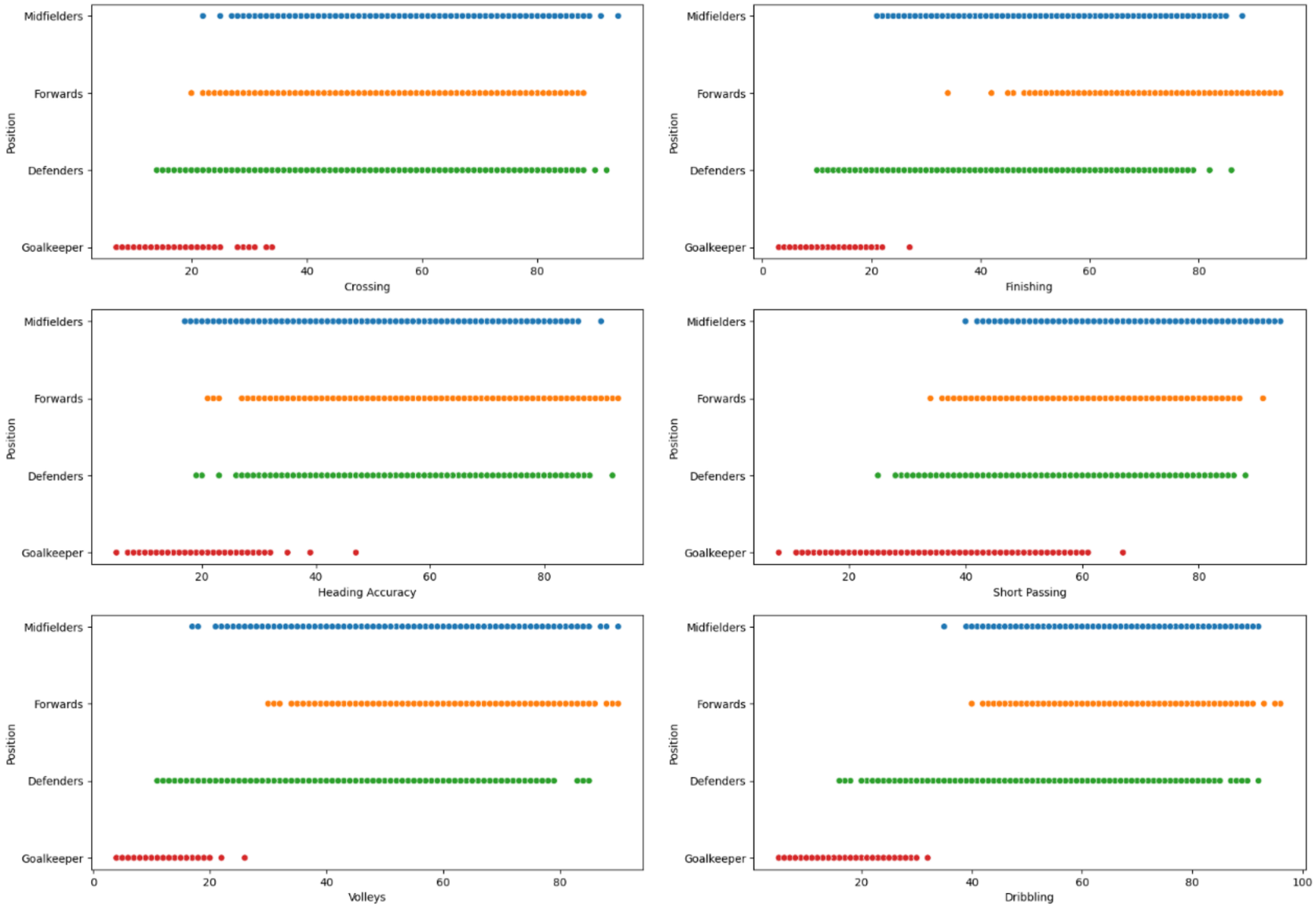
Data Visualization
Pertanyaan 3

Apakah terdapat hubungan antara posisi pemain sepakbola dengan tingkat skill-nya?

Untuk menjawab pertanyaan tersebut, kami melakukan **visualisasi hubungan antara fitur 'Posisi' dengan setiap fitur yang kami asumsikan terkait dengan tingkat skill.**

Kami menggunakan 2 bentuk visualisasi, yaitu

1. **Scatter Plot** untuk melihat pola hubungan antara fitur kategorikal dan numerikal
2. **Bar Plot** untuk melihat pola hubungan antara fitur kategorikal dan kategorikal



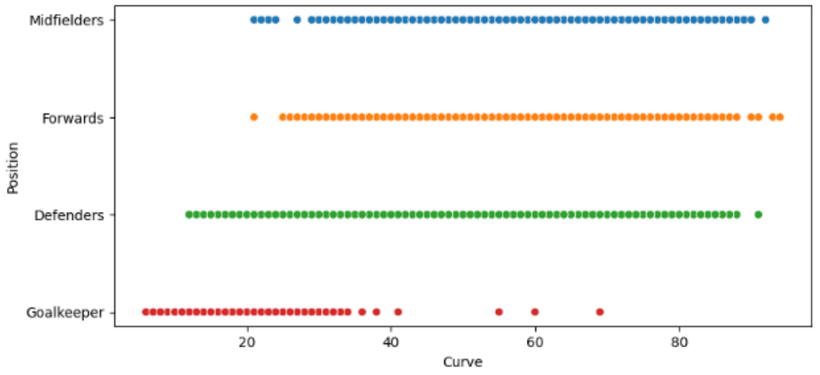
Dari 6 visualisasi di samping dapat diketahui bahwa **TERDAPAT HUBUNGAN** antara skill 'Crossing', 'Finishing', 'Short Passing', 'Volleys', dan 'Dribbling' dengan 'Position'

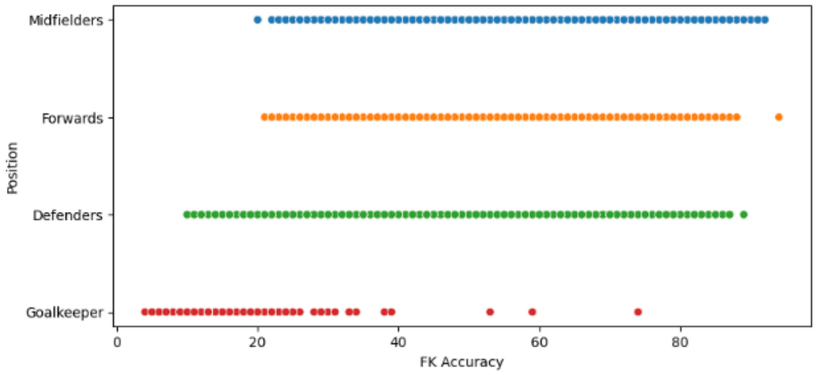
Seorang **Midlefielders** memiliki kemampuan Crossing dan Short Passing yang lebih baik daripada yang lain

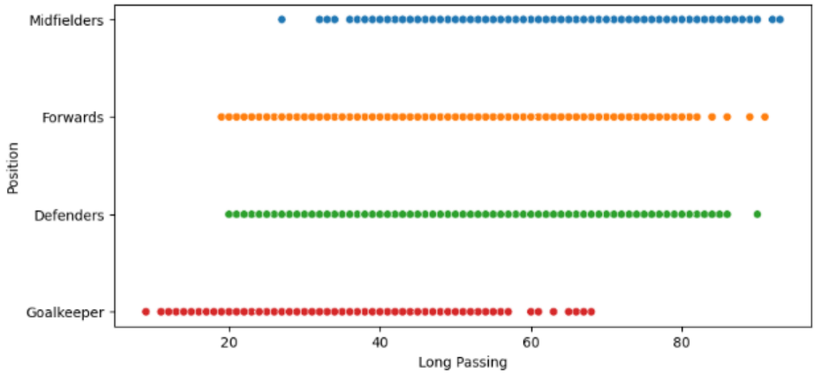
Seorang **Forwards** memiliki kemampuan Finishing dan Heading Accuracy yang lebih baik daripada yang lain

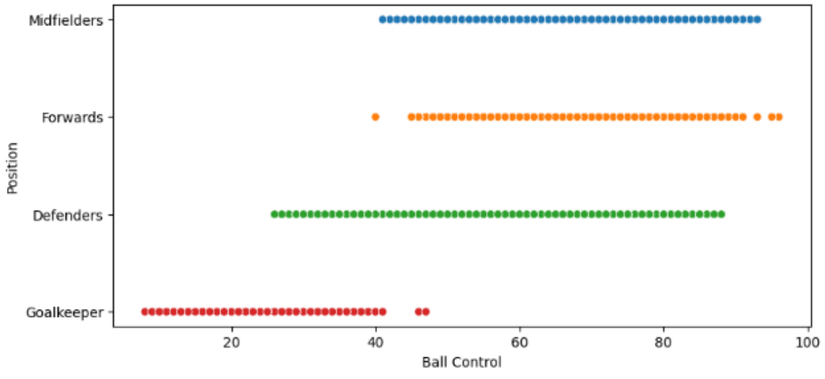
Seorang **Defenders** memiliki kemampuan Crossing, Heading Accuracy, Dribbling yang cukup baik tetapi lebih bervariasi

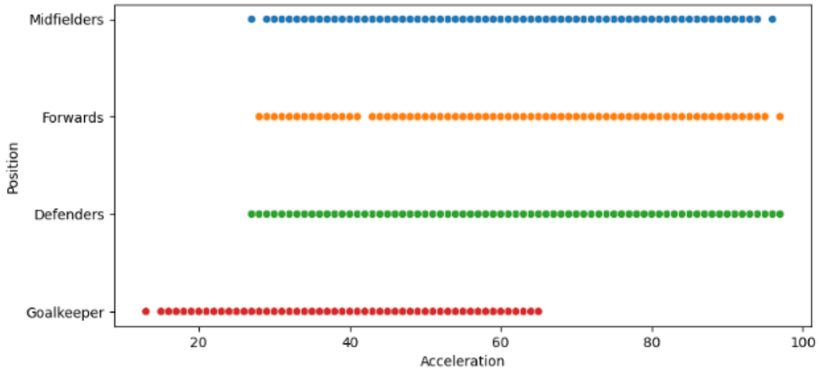
Seorang **Goalkeeper** memiliki kemampuan yang paling buruk dalam melakukan 6 skill yang disebutkan

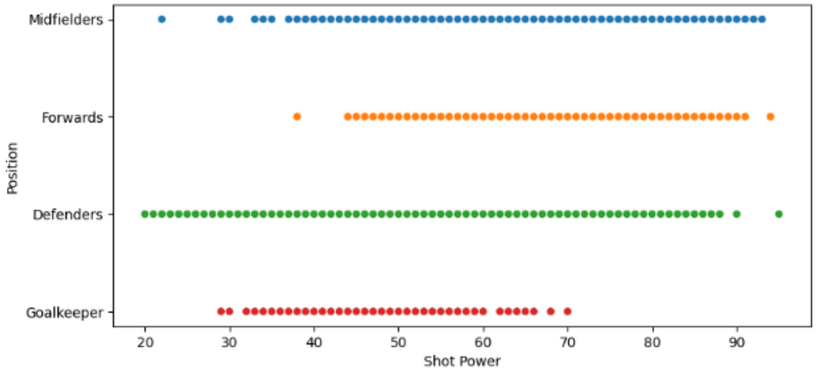












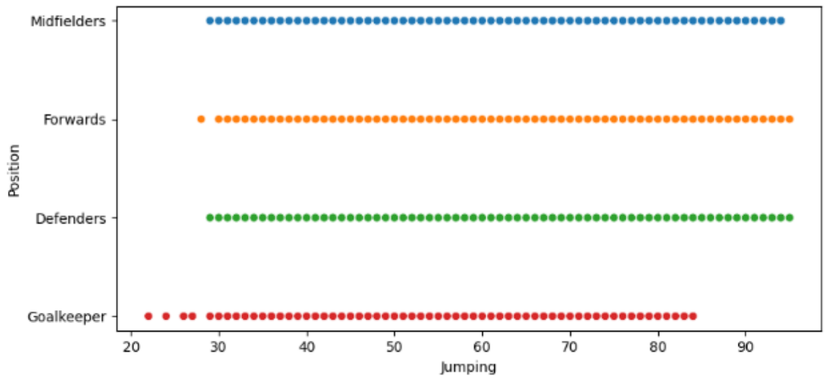
Dari 6 visualisasi di samping dapat diketahui bahwa **TERDAPAT HUBUNGAN** antara skill 'Curve', 'FK Accuracy', 'Long Passing', 'Ball Control', 'Acceleration', dan 'Shot Power' dengan 'Position'

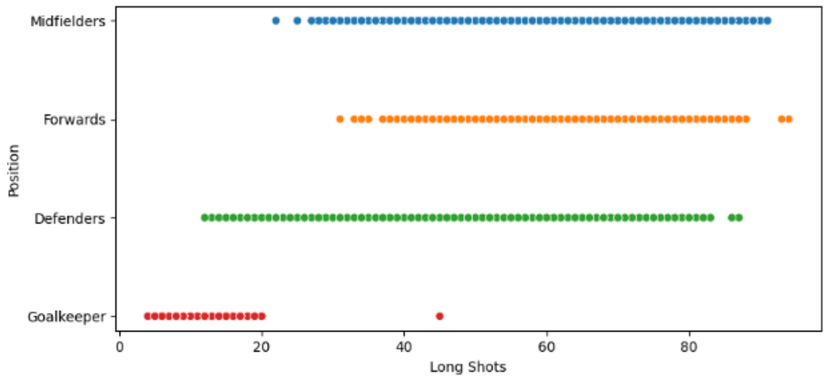
Seorang **Midlefielders** memiliki kemampuan Long Passing yang lebih baik daripada yang lain

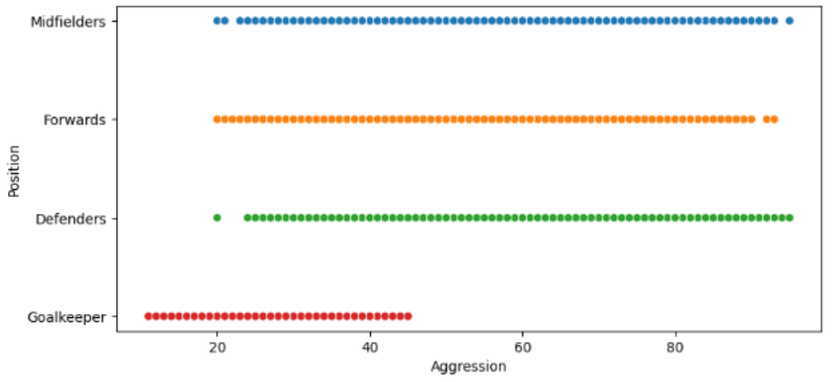
Seorang **Forwards** memiliki kemampuan Curve dan Ball Control yang lebih baik daripada yang lain

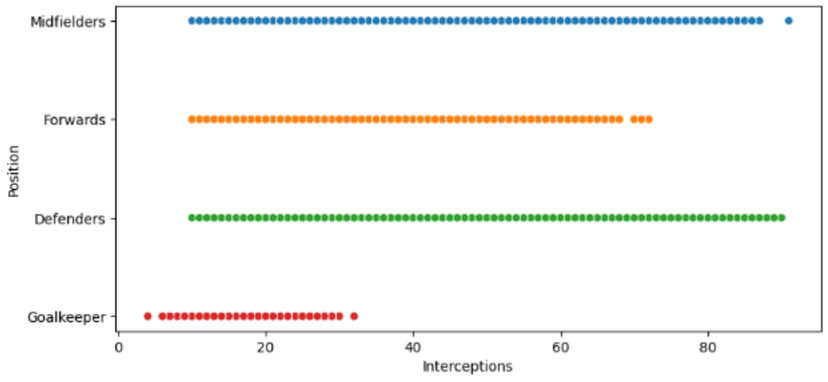
Seorang **Defenders** memiliki kemampuan Acceleration yang cukup lebih baik daripada yang lain

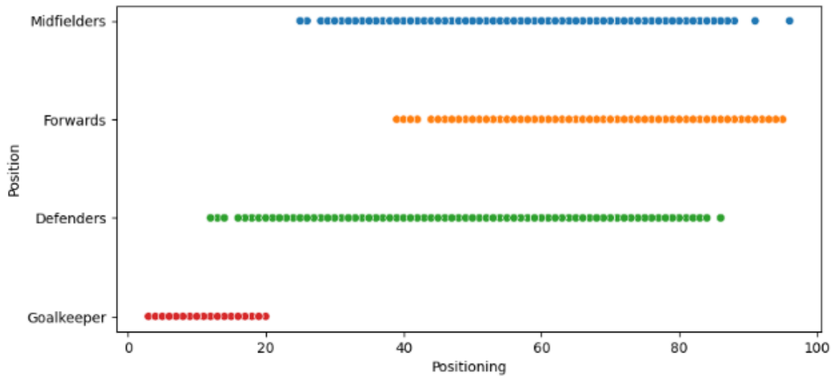
Seorang **Goalkeeper** memiliki kemampuan yang paling buruk dalam melakukan 6 skill yang disebutkan, kecuali Shot Power

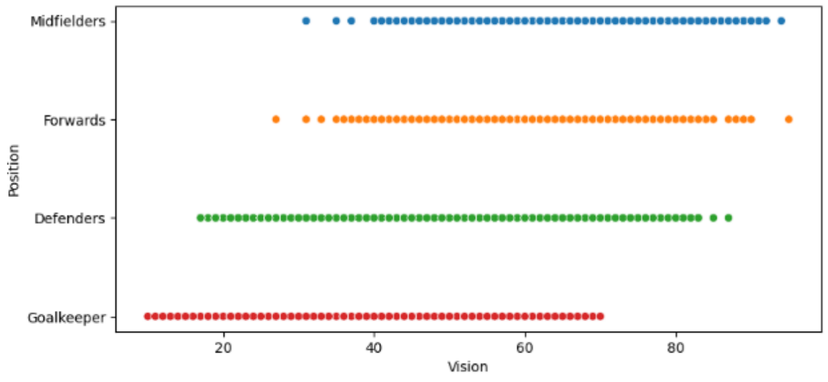












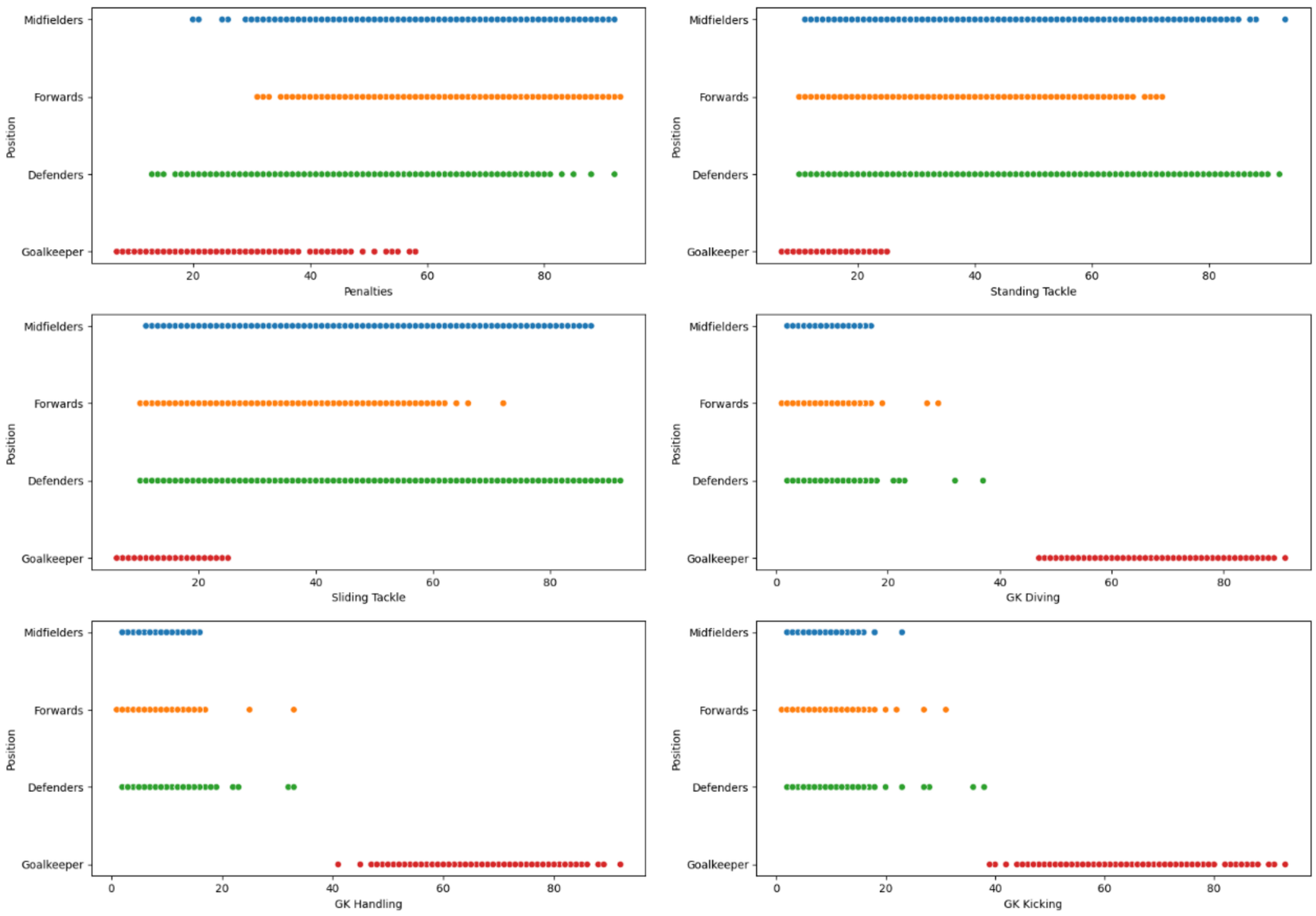
Dari 6 visualisasi di samping dapat diketahui bahwa **TERDAPAT HUBUNGAN** antara skill 'Jumping', 'Long Shots', 'Aggression', 'Interceptions', 'Positioning', dan 'Vision' dengan 'Position'

Seorang **Midlefielders** memiliki kemampuan Vision yang lebih baik daripada yang lain

Seorang **Forwards** memiliki kemampuan Long Shots dan Positioning yang lebih baik daripada yang lain

Seorang **Defenders** memiliki kemampuan Aggression yang lebih baik daripada yang lain

Seorang **Goalkeeper** memiliki kemampuan yang paling buruk dalam melakukan 6 skill yang disebutkan, tetapi cukup ahli dalam Vision dan Jumping



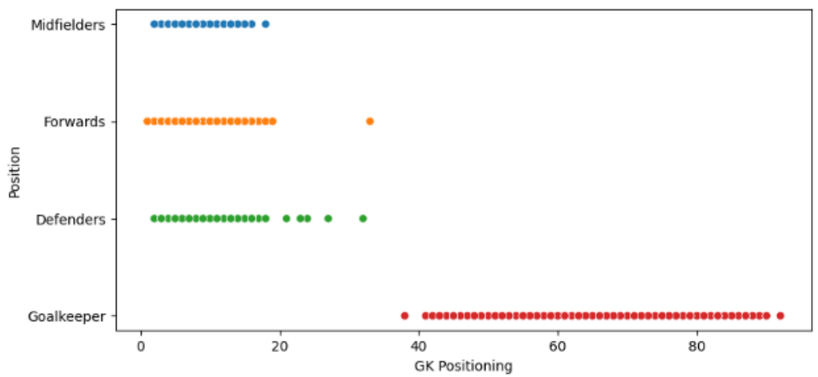
Dari 6 visualisasi di samping dapat diketahui bahwa **TERDAPAT HUBUNGAN** antara skill 'Penalties', 'Standing Tackle', 'Sliding Tackle, 'GK Driving', dan 'GK Handling' dengan 'Position'

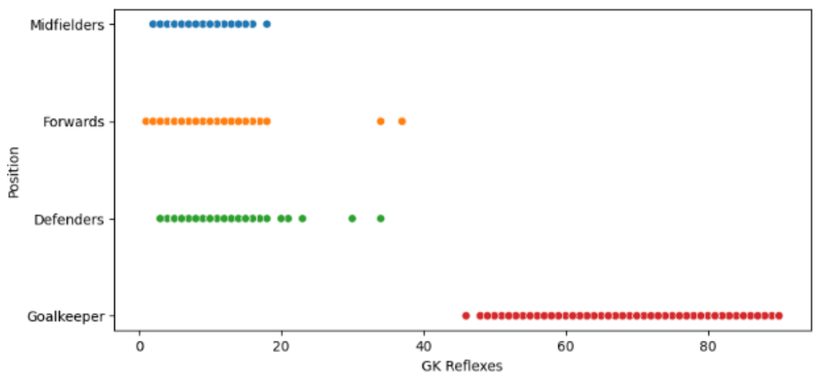
Seorang **Midlefielders** memiliki kemampuan Penalties kedua terbaik diantara seluruh posisi

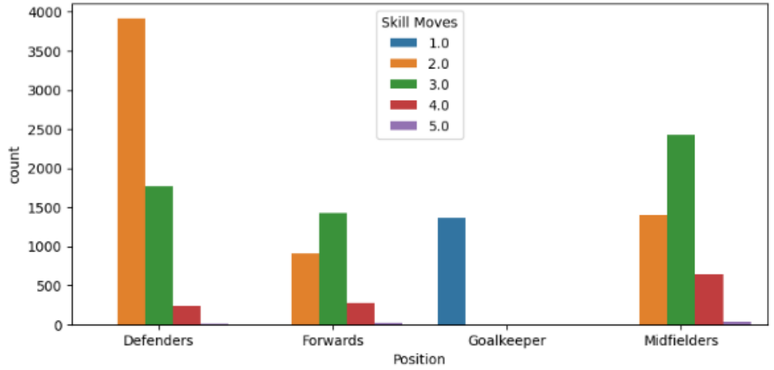
Seorang **Forwards** memiliki kemampuan Penalties lebih baik daripada yang lain

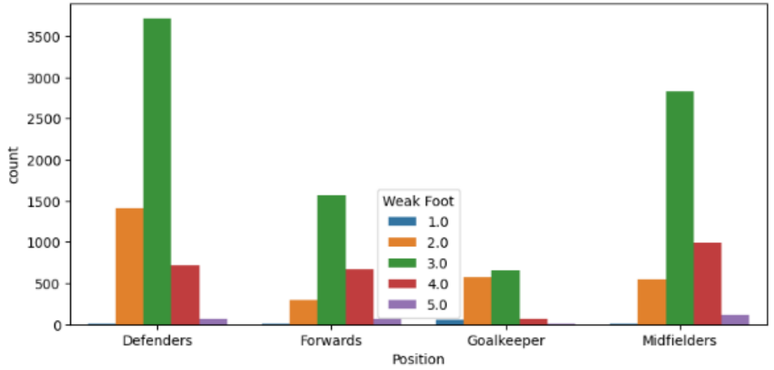
Seorang **Defenders** memiliki kemampuan Standing Tackle dan Sliding Tackle yang lebih baik daripada yang lain

Seorang **Goalkeeper** memiliki kemampuan yang paling buruk dalam melakukan Penalties, Standing Tackle, dan Sliding Tackel, tetapi kemampuan terbaik dalam melakukan GK Diving, GK Handling, dan GK Kicking









Dari 4 visualisasi di samping dapat diketahui bahwa **TERDAPAT HUBUNGAN** antara skill 'GK Positioning', 'GK Reflexes, 'Skill Moves, dan 'Weak Foot' dengan 'Position'

Seorang **Midlefielders** memiliki kemampuan Skill Moves dan Weak Foot yang lebih baik dibangingkan yang lain

Seorang **Forwards** dominan memilki kemampuan Skill Moves 2-3 dan Weak Foot 3-4

Seorang **Defenders** dominan memilki kemampuan Skill Moves 2 dan Weak Foot 3

Seorang **Goalkeeper** memiliki kemampuan yang paling buruk dalam Skill Moves dan Weak Foot, tetapi memiliki kemampuan terbaik dalam melakukan GK Positioning dan GK Reflex

Data Visualization
Pertanyaan 3

Apakah terdapat hubungan antara posisi pemain sepakbola dengan tingkat skill-nya?

Dapat disimpulkan bahwa **TERDAPAT HUBUNGAN** antara **POSISI** pemain dengan **TINGKAT SKILL**nya

- Seorang **Goalkeeper** memiliki kemampuan terbaik pada skill yang berkaitan dengan seorang kiper, yaitu GK Driving, GK Handling, GK Kicking, GK Positioning, dan GK Reflex.
- Seorang **Defenders** memiliki kemampuan terbaik pada skill yang berkaitan dengan pertahanan bola dari lawan, yaitu Aggression, Standing Tackle, dan Sliding Tackle
- Seorang **Forwards** memiliki kemampuan terbaik pada skill yang berkaitan dengan penyerangan bola yaitu Curve, Ball Control, Long Shots, Positioning, dan Penalties.
- Seorang **Midlefielders** memiliki kemampuan terbaik pada skill yang berkaitan dengan menjaga kuasa bola dari lawan dan menyerahkan bola ke Forwards, yaitu Long Passing, Crossing, Short Passing, dan Vision. Selain itu, memiliki kemahiran paling baik dalam menggunakan weak foot dan mengolah bola dengan pergerakan skillnya.

Data Visualization
Pertanyaan 4

Apakah negara sepakbola dengan starting eleven terbaik?

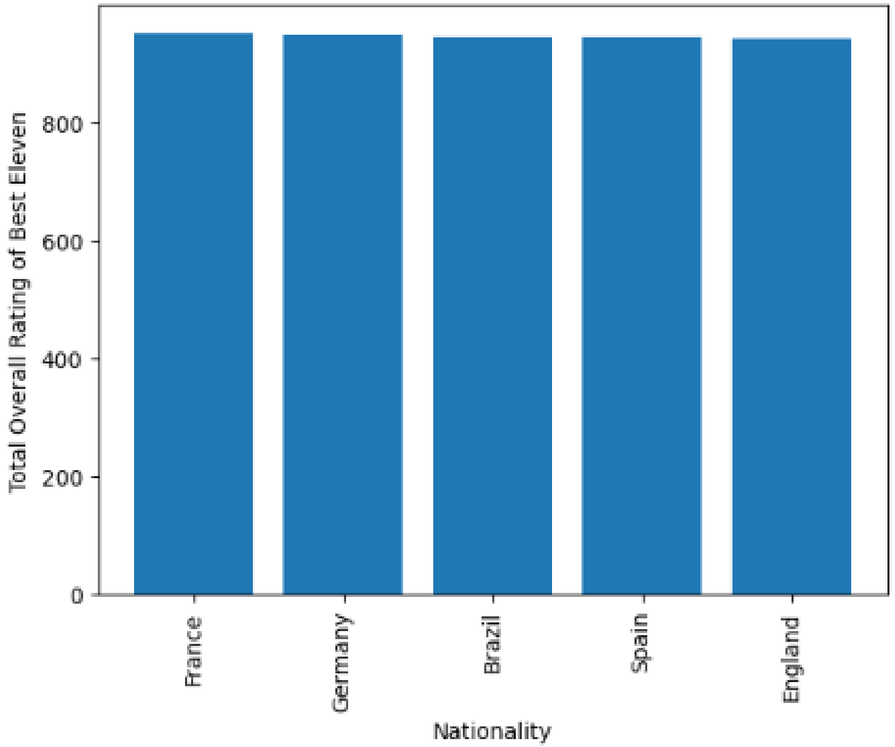
Untuk menjawab pertanyaan tersebut, kami melakukan visualisasi perbandingan 5 negara dengan total **Overall** rating terbaik, kemudian melakukan visualisasi perbandingan 5 negara dengan total **Value** tertinggi.

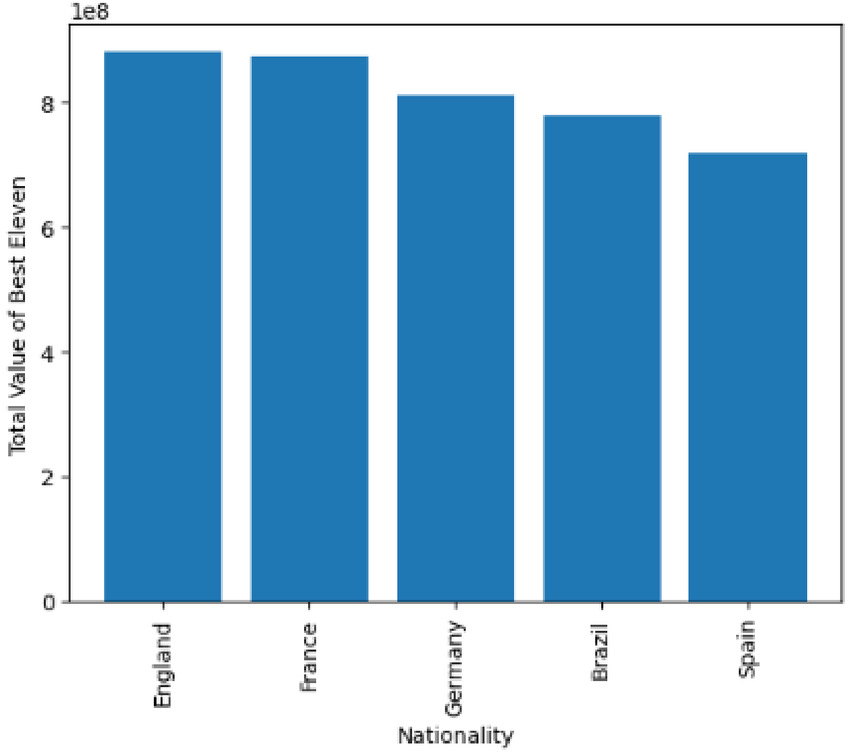
Kami menggunakan visualisasi, yaitu **Bar plot** untuk melihat perbandingannya.

Dalam pemilihan starting eleven, kami memastikan tepat 1 pemain sepak bola berposisi Goalkeeper.

Data Visualization
Pertanyaan 4

Apakah negara sepakbola dengan starting eleven terbaik?





Dapat dilihat pada visualisasi bagian kiri, **France memiliki Total Overall Rating paling terbaik**. Disusul oleh Germany, Brazil, Spain, England. Kemudian, dapat dilihat pada visualisasi bagian kanan, **England memiliki Total Value paling tertinggi**. **Disusul oleh France**, Germany, Brazil, Spain. Oleh karena itu, dapat disimpulkan bahwa negara sepakbola dengan starting eleven terbaik adalah **France**.

Data Visualization
Pertanyaan 5

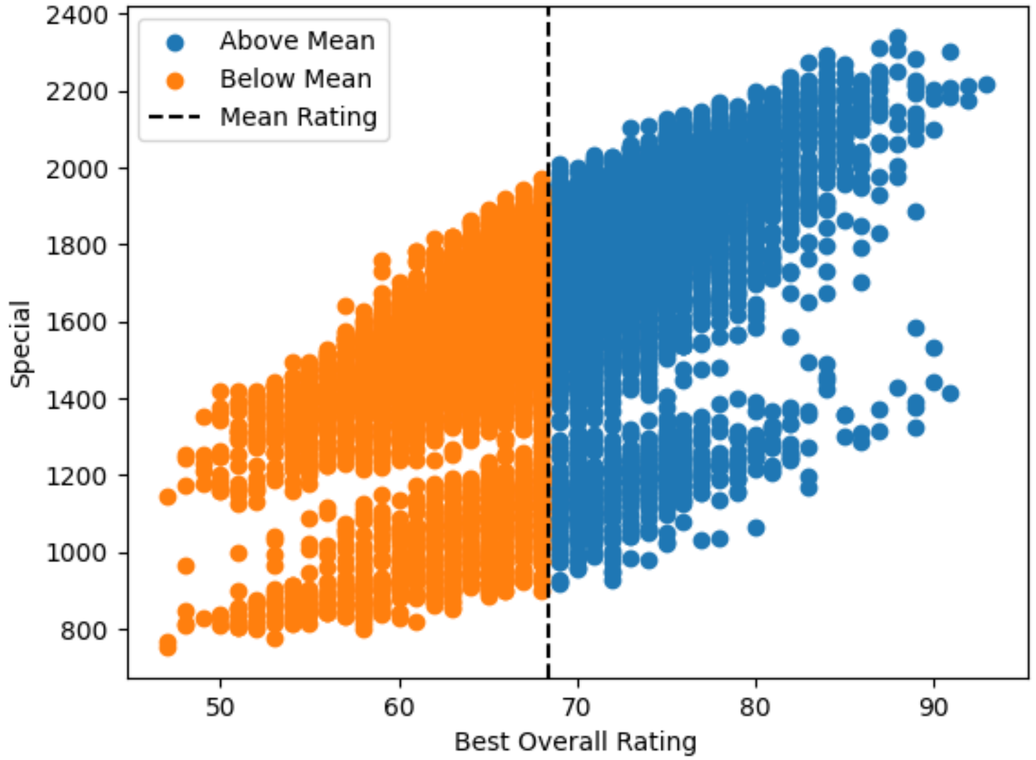
Apakah pemain dengan nilai rating yang diatas rata-rata memiliki nilai special yang tinggi?

Untuk menjawab pertanyaan tersebut, kami melakukan **visualisasi hubungan antara fitur 'Best Overall Rating' dengan fitur 'Special'**

Kami menggunakan visualisasi, yaitu **Scatter Plot** untuk melihat antara keduanya.

Data Visualization
Pertanyaan 5

Apakah pemain dengan nilai rating yang diatas rata-rata memiliki nilai special yang tinggi?



Dapat dlihat pada visualisasi disamping bahwa kedua variabel memiliki **korelasi yang cukup kuat dan positif**, dimana apabila nilai Best Overall Rating tinggi, nilai Special juga akan tinggi. Nilai Special yang memiliki rating di atas rata-rata cenderung lebih tinggi dibandingkat yang di bawa rata-rata. Oleh karena itu, dapat disimpulkan bahwa **pemain dengan nilai rating yang di atas rata rata cenderung memiliki nilai kespesialan yang tinggi**.

Data Visualization
Pertanyaan 6

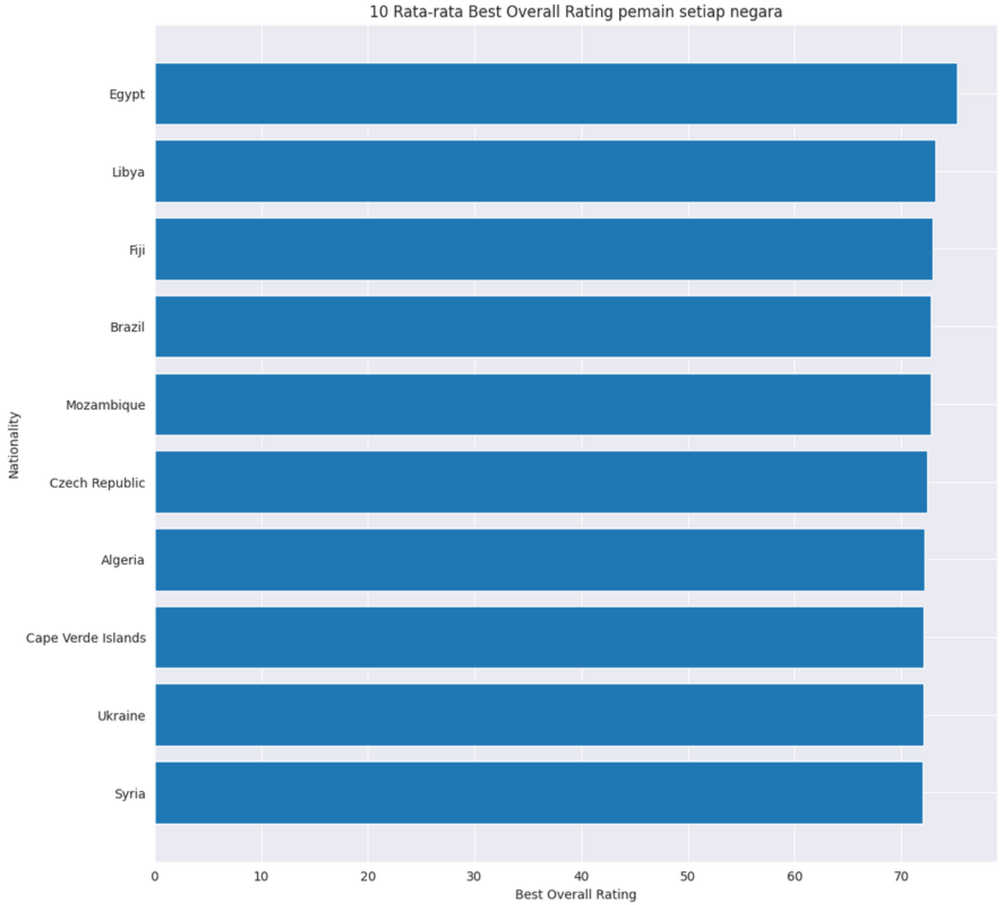
Apa negara yang memiliki pemain dengan best overall rating tertinggi?

Untuk menjawab pertanyaan tersebut, kami melakukan **visualisasi** perbandingan 10 negara dengan mean 'Best Overall Rating' terbaik

Kami menggunakan visualisasi, yaitu **Horizontal Bar Plot** untuk melihat perbandingannya.

Data Visualization
Pertanyaan 6

Apa negara yang memiliki pemain dengan best overall rating tertinggi?



Dapat dlihat pada visualisasi disamping bahwa **Egypt** memiliki rata-rata Best Overall Rating terbaik. Disusul oleh Libya, Fiji, Brazil, Mozambique, dan seterusnya

KASDD Genap 2022/2023

KLASIFIKASI

Football Player

Modeling

Berikut adalah langkah-langkah yang kami lakukan untuk pemodelan klasifikasi

1. Scaling data menggunakan MinMaxScaler
2. Meng-encode kategorikal data menjadi numerik
3. Membagi variabel dependen dan independen
4. Split data test dan train
5. Mencari model terbaik
6. Features Reduction menggunakan k-best
7. Mencari best parameter
8. Mencari sampling terbaik

Football Player

Mencari Model Terbaik

	Classifier	Accuracy
0	XGB	0.889390
1	SVC	0.884702
2	GBM	0.882271
3	LR	0.872981
4	LiSVC	0.872027
5	RF	0.872026
6	SGD	0.861782
7	LDA	0.857441
8	QDA	0.836951
9	KNN	0.803353
10	DT	0.803265
11	GNB	0.753950
12	ET	0.744227

Dapat dlihat pada visualisasi di samping bahwa kami telah melakukan modeling dengan berbagai macam model, kemudian kami urutkan dari yang memiliki akurasi tertinggi. Penggunaan model XGB menghasilkan akurasi tertinggi (senilai 0.889) dibandingkan lainnya seperti SVC, GBM, LR, dan seterusnya.

Oleh karena itu, kami akan menggunakan model XGB.

Football Player

Melakukan Features Reduction

	Classifier	Acc normal	Acc k-best
0	XGB	0.88939	0.893992

Dapat dlihat pada visualisasi di atas bahwa kami telah melakukan features reduction dengan k-best. Penggunaan model XGB dengan features reduction menggunkaan k-best memiliki akurasi tertinggi dibandingkan dengan jika tidak dilakukan feature reduction.

Oleh karena itu, kami akan menggunakan features yang sudah direduksi dengan menggunakan k-best. Kami menggunakan 53 features sebagai X_train dan X_test.

Football Player

Best Parameter

```python
from sklearn.model_selection import RandomizedSearchCV, GridSearchCV
from sklearn.model_selection import StratifiedKFold

param_grid = {
    'max_depth': range(3,10,2),
    'min_child_weight': range(1,6,2),
    'gamma': [0.1, 0.2, 0.3, 0.4],
    'subsample': [0.5, 0.6, 0.7, 0.8, 0.9],
    'colsample_bytree': [0.5, 0.6, 0.7, 0.8, 0.9],
    'learning_rate': [0.01, 0.05, 0.1, 0.2],
    'n_estimators': [50, 100, 200, 300]
}

skf = StratifiedKFold(n_splits=5)
xgb = XGBClassifier()
xgb_random_search = RandomizedSearchCV(estimator = xgb, param_distributions = param_grid, n_iter = 50, cv = skf, verbose=2, random_state=42, n_jobs = -1)
xgb_random_search.fit(X_train_selected, y_train)
```

```
Fitting 5 folds for each of 50 candidates, totalling 250 fits
```

```
▸ RandomizedSearchCV
▸ estimator: XGBClassifier
    ▸ XGBClassifier
```

Football Player

Sampling

	sampler	accuracy
0	BorderlineSMOTE	0.896528
1	TomekLinks	0.896181
2	RandomOverSampler	0.895486
3	SVMSMOTE	0.893403
4	SMOTE	0.892708
5	RandomUnderSampler	0.864931
6	NearMiss2	0.767014
7	NearMiss1	0.744792
8	NearMiss3	0.640625

Dapat dlihat pada visualisasi di samping bahwa kami telah melakukan sampling dengan berbagai macam teknik, kemudian kami urutkan dari yang memiliki akurasi tertinggi. Penggunaan teknik BorderlineSMOTE, TomekLinks, dan Random OverSampler menghasilkan akurasi tertinggi dibandingkan lainnya.

Oleh karena itu, kami akan menggunakan ketiga teknik tersebut untuk dicari stratified cross validationnya.

Football Player

Hasil Evaluasi

Didapatkan hasil akurasi tertinggi menggunakan sampler RandomOverSampler dengan nilai sbagai berikut.

Hasil akurasi dengan Stratified Cross Validation:

0,95

Hasil akurasi dengan Memprediksi X_test:

0,89

```
Hasil Evaluasi berdasarkan classification report

              precision    recall  f1-score   support

           0       0.91      0.89      0.90      1171
           1       0.87      0.89      0.88       501
           2       1.00      1.00      1.00       258
           3       0.86      0.87      0.86       950

    accuracy                           0.89      2880
   macro avg       0.91      0.91      0.91      2880
weighted avg       0.90      0.89      0.89      2880
```

Confusion Matrix

prediction	0	1	2	3
actual				
0	1043	28	0	100
1	15	446	0	40
2	0	0	258	0
3	82	38	0	830

```
Butuh informasi lebih lengkap? silakan simak di bawah ini :
Accuracy Average: 0.8947916666666667
F1 Macro Average: 0.9119439241842916
F1 Micro Average: 0.8947916666666667
Precision Macro Average: 0.9104190334486345
Precision Micro Average: 0.8947916666666667
Recall Macro Average: 0.9136488719715498
Recall Micro Average: 0.8947916666666667
```

```
[ ] skf = StratifiedKFold(n_splits=5)
    cv_scores = cross_val_score(best_model, X_train_sample, y_train_sample, cv=skf, scoring='accuracy')
    mean_acc = cv_scores.mean()
    print(mean_acc)

    0.9508112361693349
```

Football Player

Kesimpulan

Model yang telah dibuat untuk mengklasifikasikan posisi dari seorang pemain sepak bola meraih nilai akurasi rata-rata sekitar

95%

pada subset training data yang telah di cross validasi menggunakan metode StratifiedKFold.

Dapat disimpulkan bahwa akurasi dari model sudah tinggi untuk memprediksi klasifikasi dari posisi pemain sepak bola.

Football Player

Saran untuk Meningkatkan Performa

- Melakukan features engineering dengan membuat variabel baru yang menggabungkan beberapa features yang sejenis
- Melakukan stratified cross validation dalam menjadi Select K-Best
- Mencoba metode features reduction lain
- Mencoba metode encode lain
- Mencoba metode scaling lain

KASDD Genap 2022/2023

REGRESI

Football Player

Modeling

Berikut adalah langkah-langkah yang kami lakukan untuk pemodelan regresi

1. Scaling data menggunakan MinMaxScaler
2. Meng-encode kategorikal data menjadi numerik
3. Membagi variabel dependen dan independen
4. Split data test dan train
5. Mencoba melakukan features reduction
6. Mencari Best Parameter
7. Melakukan data sampling untuk mengubah imbalanced data menjadi balanced

Football Player

Mencari Model Terbaik

	Model	MSE	R2
0	RF	1.015503e+08	0.807551
1	LR	1.411366e+08	0.732530
2	L	1.411785e+08	0.732451
3	R	1.415297e+08	0.731785
4	TS	1.441896e+08	0.726744
5	HR	1.560167e+08	0.704331
6	DT	2.013140e+08	0.618487
7	SGD	3.671728e+08	0.304166
8	KNN	4.012461e+08	0.239593
9	EN	4.381977e+08	0.169566
10	RANSAC	5.017465e+08	0.049134

Dapat dlihat pada visualisasi di samping bahwa kami telah melakukan modeling dengan berbagai macam model, kemudian kami urutkan dari yang memiliki akurasi tertinggi. Penggunaan model RF menghasilkan nilai r2 tertinggi (senilai 0.807) dibandingkan model lainnya seperti LR, L, R, TS, dan seterusnya.

Oleh karena itu, kami akan menggunakan model RF.

Football Player

Melakukan Features Reduction

	Classifier	R2 normal	R2 k-best
0	RF	0.782793	0.644355

Dapat dlihat pada visualisasi di atas bahwa kami membandingkan nilai r2 pada data normal (tidak dilakukan features reduction) dengan data yang dikurangi fiturnya menggunakan k-best.

Penggunaan model RF tanpa features reduction menghasilkan nilai r2 yang lebih tinggi (senilai 0.78) dibandingkan pada model dengan feature reduction menggunakan k-best.

Oleh karena itu, kami memutuskan untuk tidak melakukan features reduction.

Football Player

Best Parameter

```
param_distribution = {
    'n_estimators': [100, 200, 500],
    'max_features': ['auto', 'sqrt'],
    'max_depth': [None, 5, 10],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}

rf = RandomForestRegressor()
skf = StratifiedKFold(n_splits=5)
random_search = RandomizedSearchCV(rf, param_distribution, cv=skf)
random_search.fit(X_train, y_train)
```

Football Player

Hasil Evaluasi

Setelah melakukan sampling, didapatkan hasil r2 terbaik menggunakan Random Undersampling, dengan nilai sebesar:

0,809

	sampler	r2
0	RandomUnderSampler	0.809080
1	RandomOverSampler	0.808291

Football Player

Kesimpulan

Nilai r2 hasil random undersampling sebesar 0,809 pada model yang telah dibuat untuk memprediksi wage dari seorang pemain sepak bola menunjukkan bahwa sekitar

80.9%

variansi dari wage pemain sepak bola dapat dijelaskan oleh fitur-fitur yang digunakan dalam model.

Dengan nilai r2 sebesar 0.809, dapat disimpulkan bahwa model mampu memberikan prediksi yang cukup baik terkait wage pemain sepak bola berdasarkan fitur-fitur yang ada.

Football Player

Saran untuk Meningkatkan Performa

- Melakukan features engineering dengan membuat variabel baru yang menggabungkan beberapa features yang sejenis
- Melakukan stratified cross validation dalam mencari Select K-Best
- Mencoba metode features reduction lain
- Mencoba metode encode lain
- Mencoba metode scaling lain

KASDD Genap 2022/2023

CLUSTERING

Football Player

Clustering

Mengklasterisasi pemain sepak bola dan melakukan analisis jenis-jenis seorang pemain sepak bola yang terdapat dalam suatu cluster

Tahapan:

1. Memilih fitur yang relevan
2. Melakukan visualisasi Elbow Method
3. Melakukan visualisasi Silhouette Score
4. Melakukan analisis jenis pemain dalam suatu cluster

Football Player

Clustering

Berikut adalah fitur-fitur yang dipilih

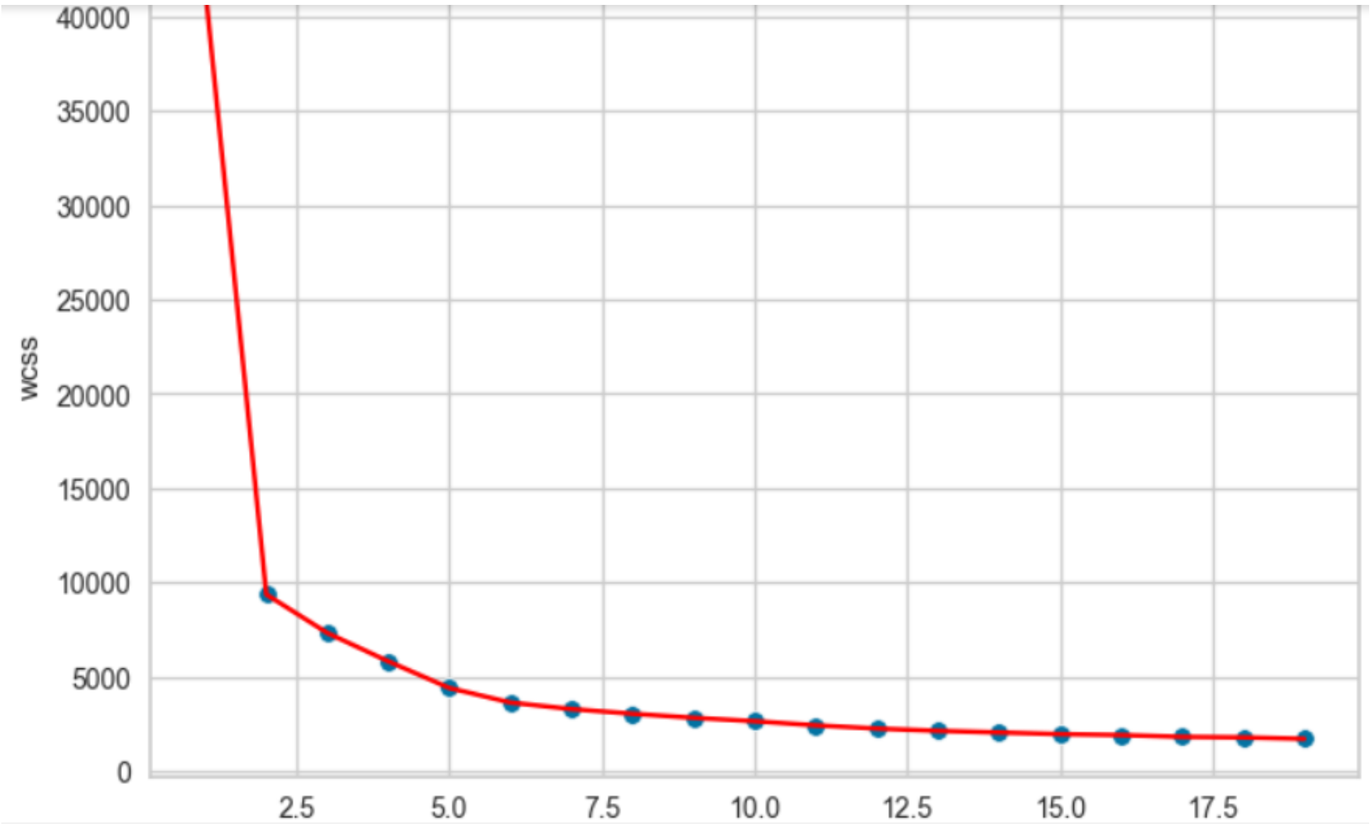
- Position
- GK Diving
- Aggression
- Curve
- Long Passing
- Finishing
- Short Passing
- Dribbling
- Sprint Speed
- Shot Power
- Stamina
- Interceptions
- GK Handling

Football Player

Clustering

Visualisasi cluster menggunakan Elbow Method menemukan range nilai k 2-6.

Kemudian, untuk memastikan memilih jumlah cluster yang optimal maka akan mencoba menggunakan silhouette method.



Football Player

Clustering

Berikut ini clustering yang terbuat dari hasil analisis silhoutte coefficient

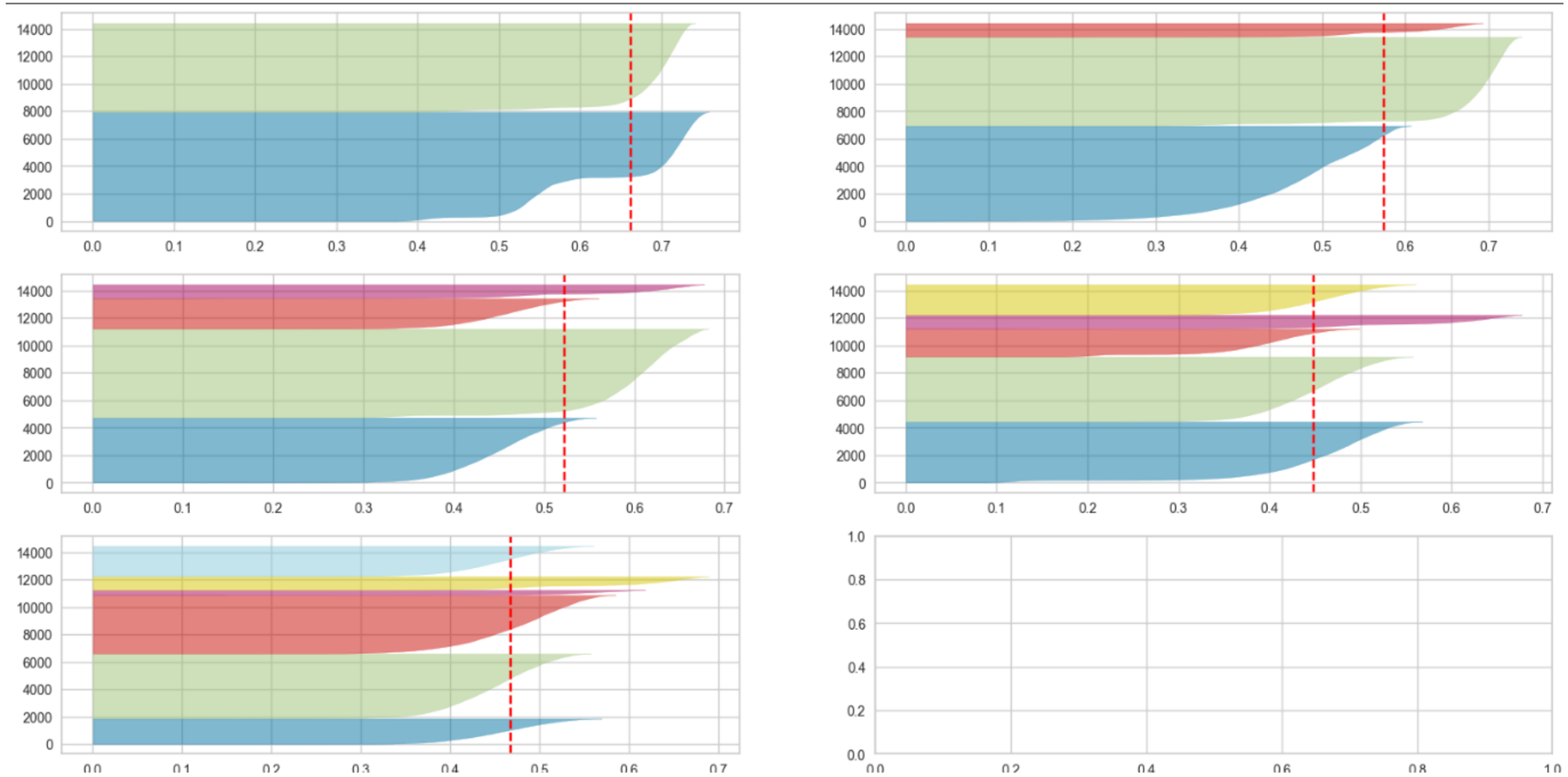
```
Komparasi silhouette coefficient untuk k berbeda:
==================================================
c:\Users\FITRI\AppData\Local\Programs\Python\Python310\lib\site-packages\sk
  warnings.warn(
Untuk k = 2, rata-rata silhouette_coefficient adalah: 0.6622052763131298
c:\Users\FITRI\AppData\Local\Programs\Python\Python310\lib\site-packages\sk
  warnings.warn(
Untuk k = 3, rata-rata silhouette_coefficient adalah: 0.5748590384194737
c:\Users\FITRI\AppData\Local\Programs\Python\Python310\lib\site-packages\sk
  warnings.warn(
Untuk k = 4, rata-rata silhouette_coefficient adalah: 0.5230429496603839
c:\Users\FITRI\AppData\Local\Programs\Python\Python310\lib\site-packages\sk
  warnings.warn(
Untuk k = 5, rata-rata silhouette_coefficient adalah: 0.4482360855812633
c:\Users\FITRI\AppData\Local\Programs\Python\Python310\lib\site-packages\sk
  warnings.warn(
Untuk k = 6, rata-rata silhouette_coefficient adalah: 0.46783853363932476
```

Hasil clustering menunjukkan paling optimal ketika k=2.

Hasil yang diperoleh masih kurang maksimal bisa karena beberapa hal seperti kurang tepatnya pemilihan fitur atau kurang banyak dalam pemilihan fitur sehingga informasi yang didapat masih kurang.

Football Player
Clustering

Visualisasi kluster dengan k = 2-6

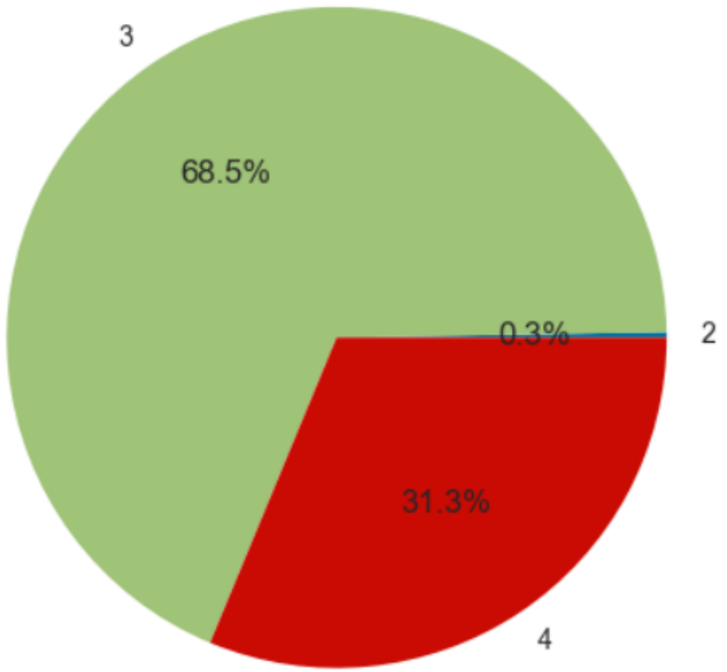


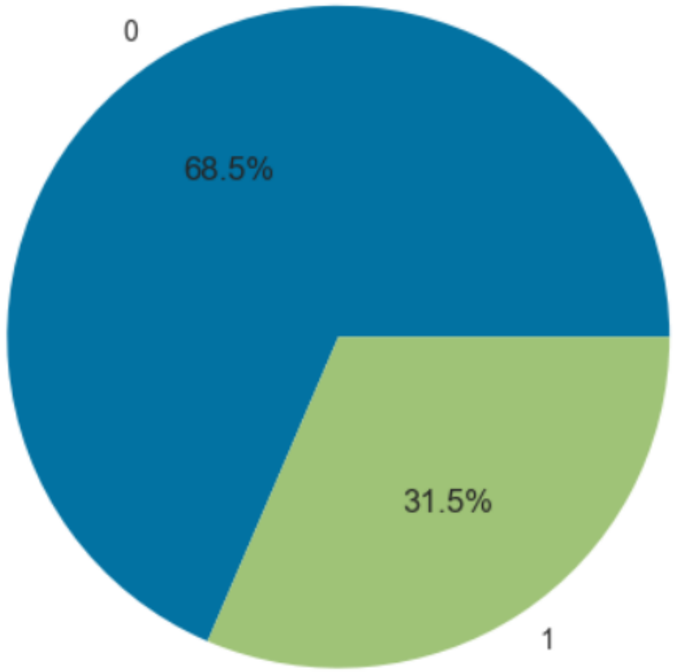
Football Player

Clustering

INTERPRETASI CLUSTERING :

Position

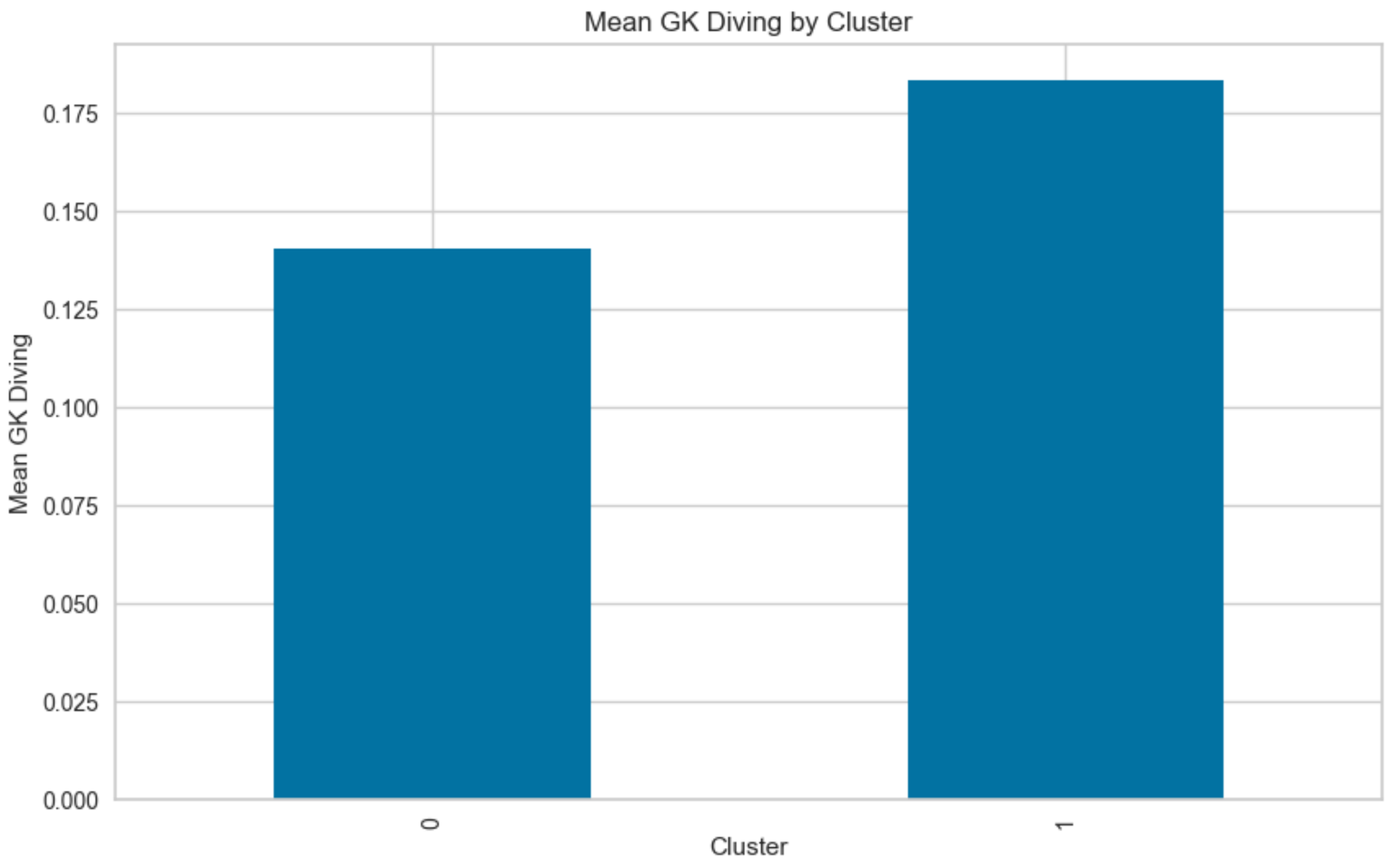




- **Cluster 0**
- **Cluster 1**

Football Player
Clustering

INTERPRETASI CLUSTERING :
GK Diving

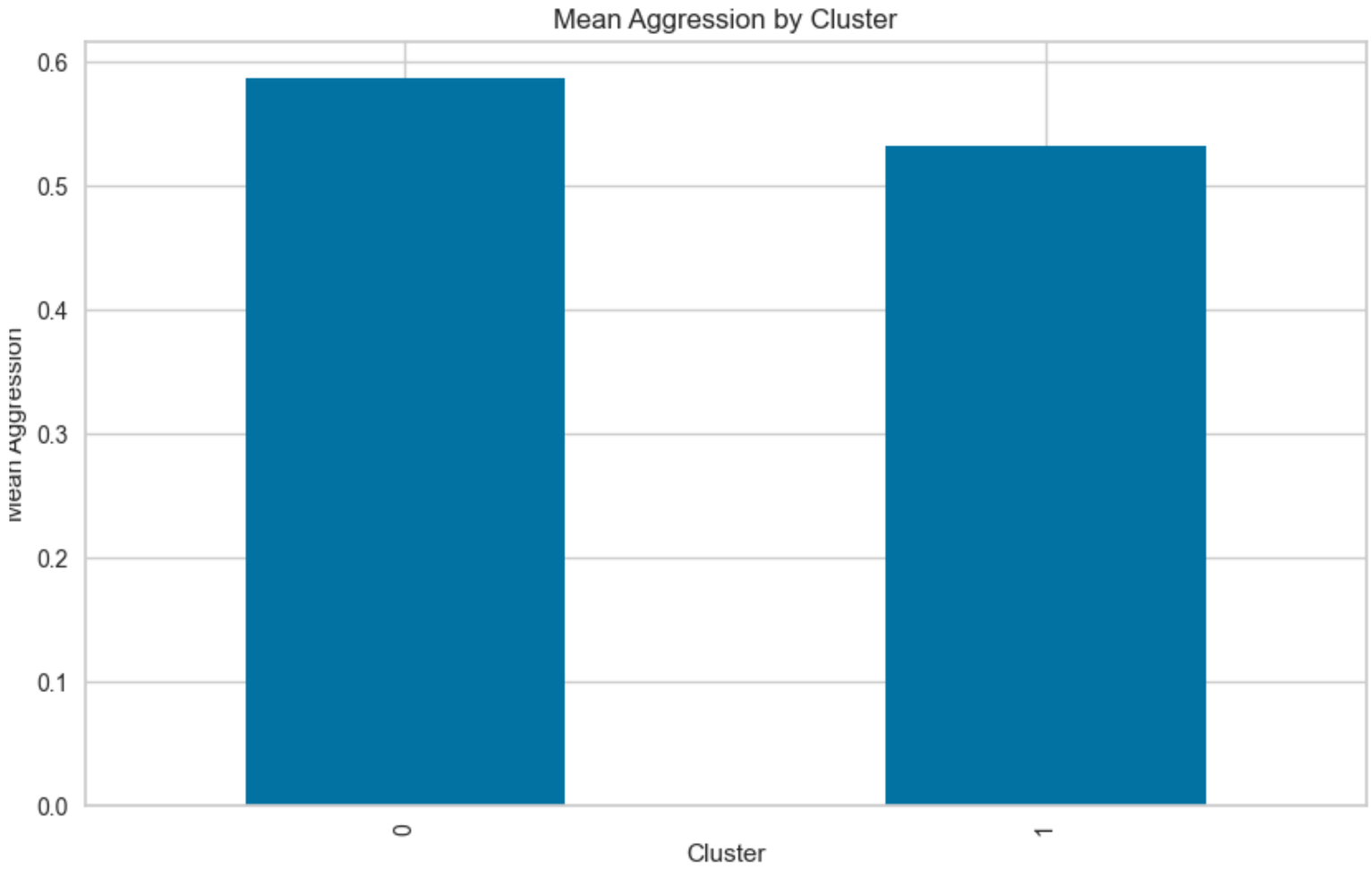


Football Player

Clustering

INTERPRETASI CLUSTERING :

Aggression

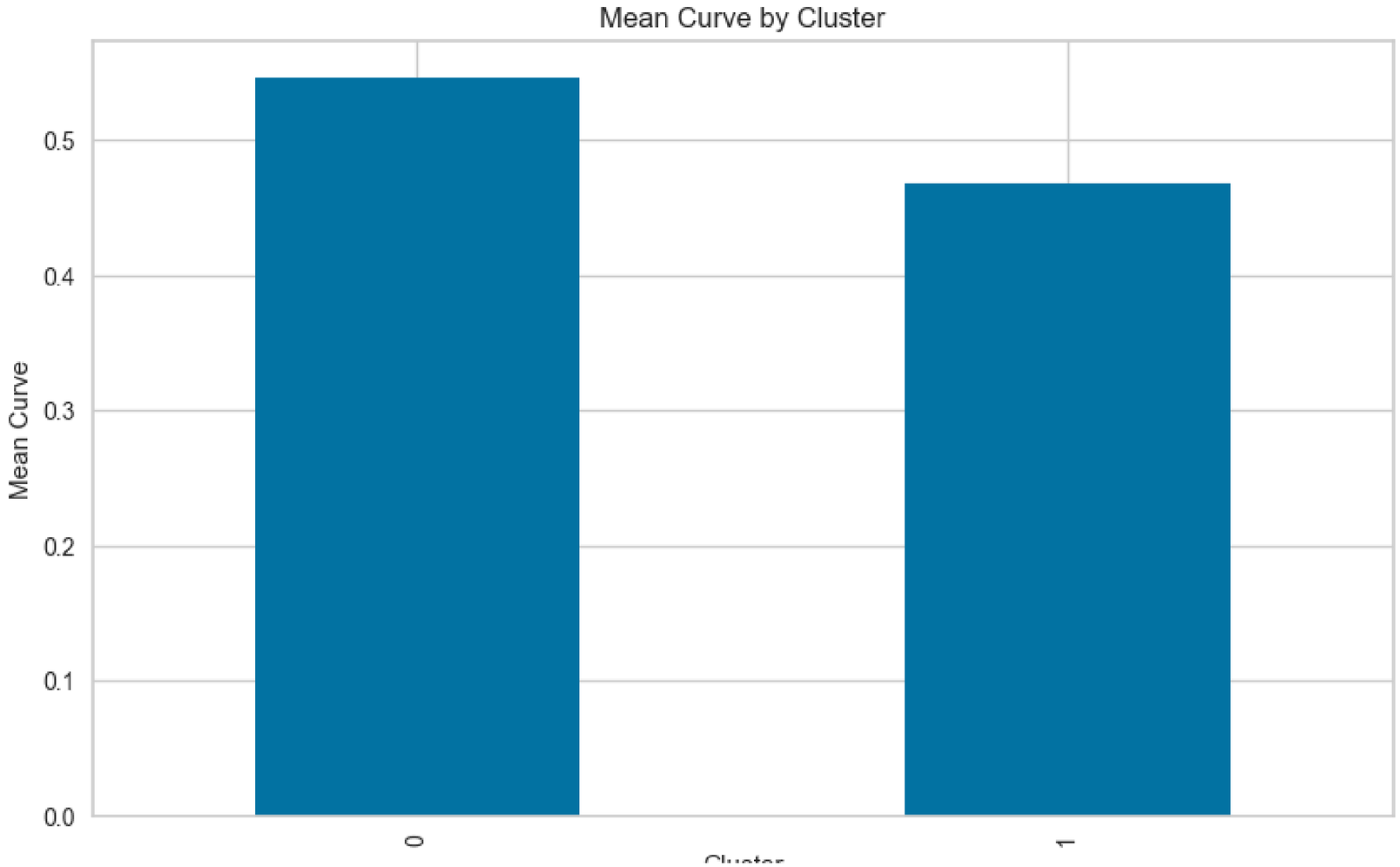


Football Player

Clustering

INTERPRETASI CLUSTERING :

Curve

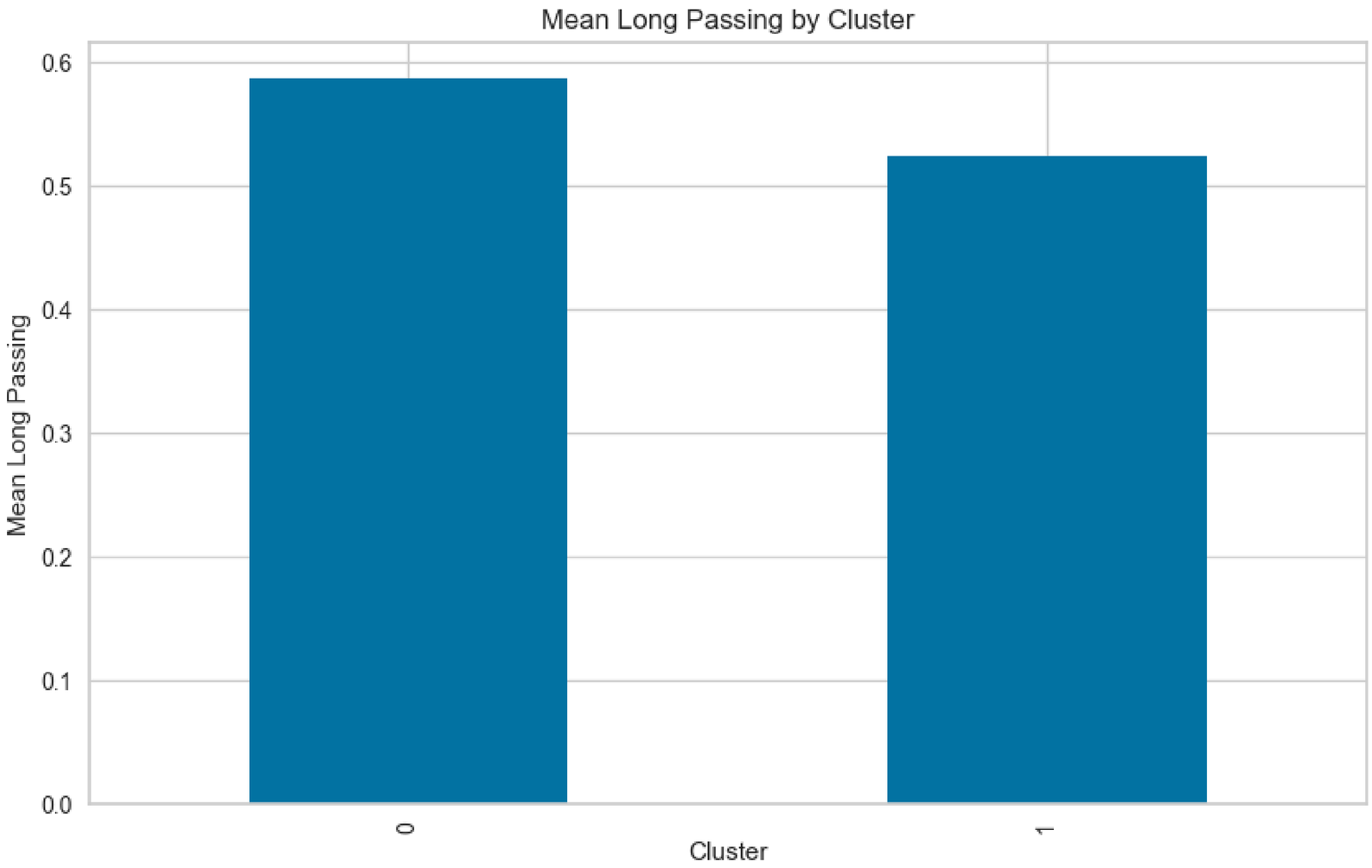


Football Player

Clustering

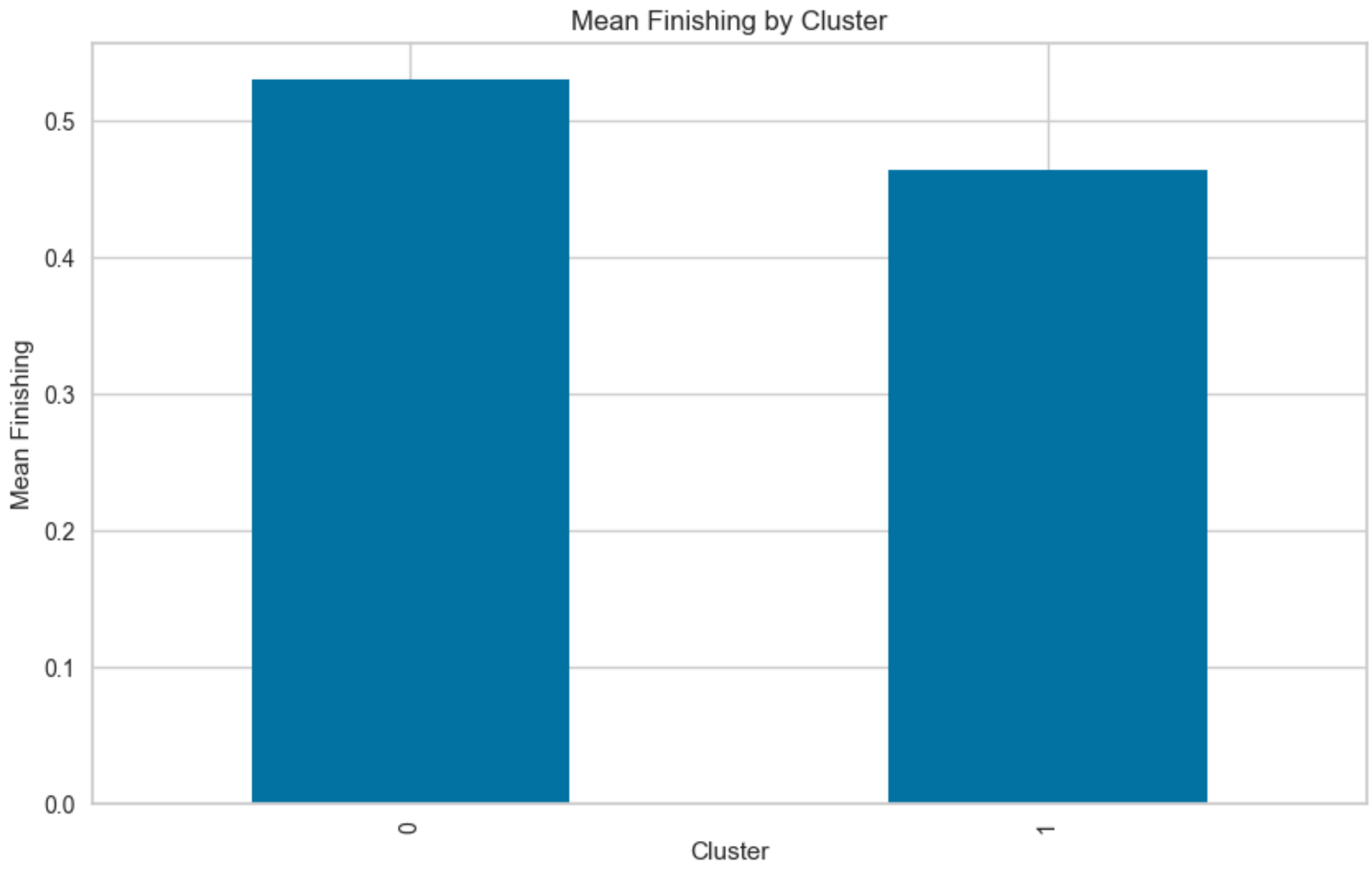
INTERPRETASI CLUSTERING :

Long Passing



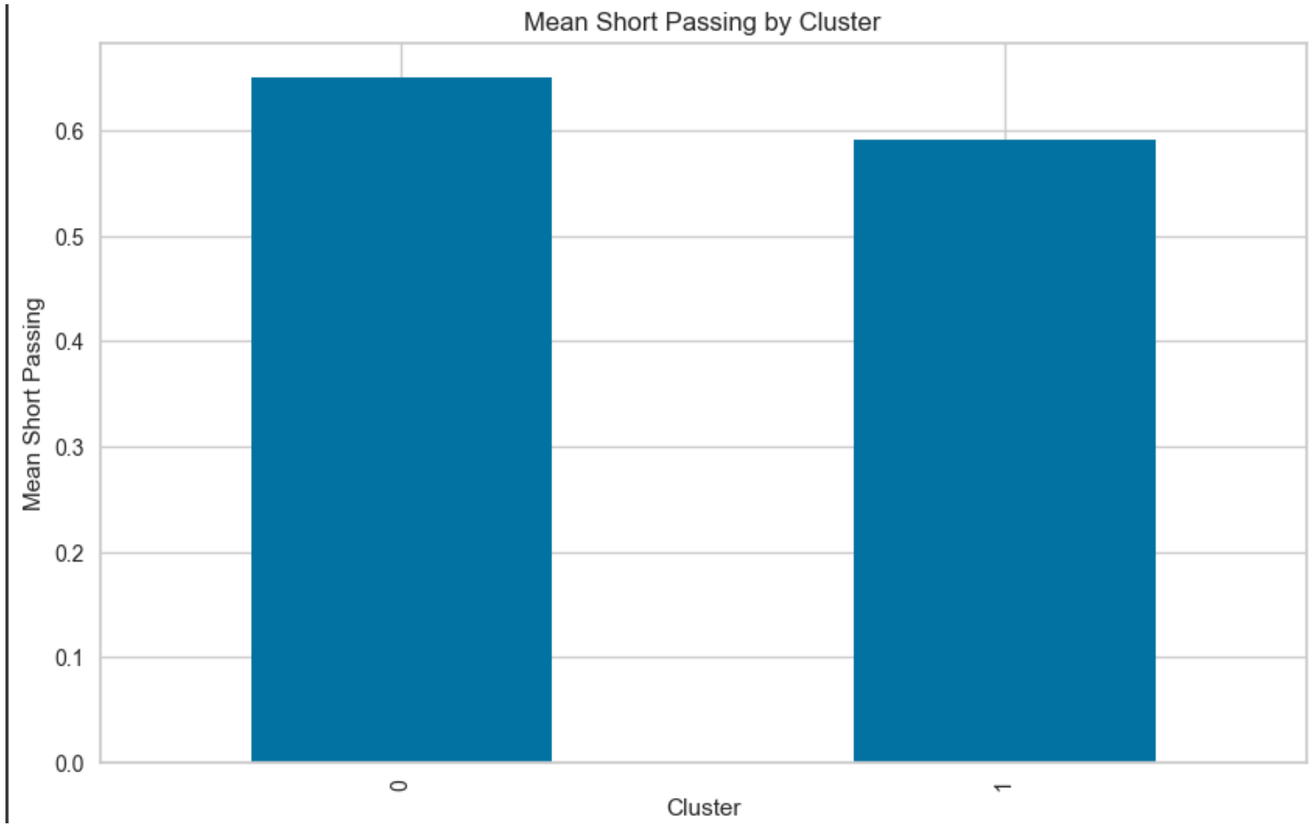
Football Player
Clustering

INTERPRETASI CLUSTERING :
Finishing



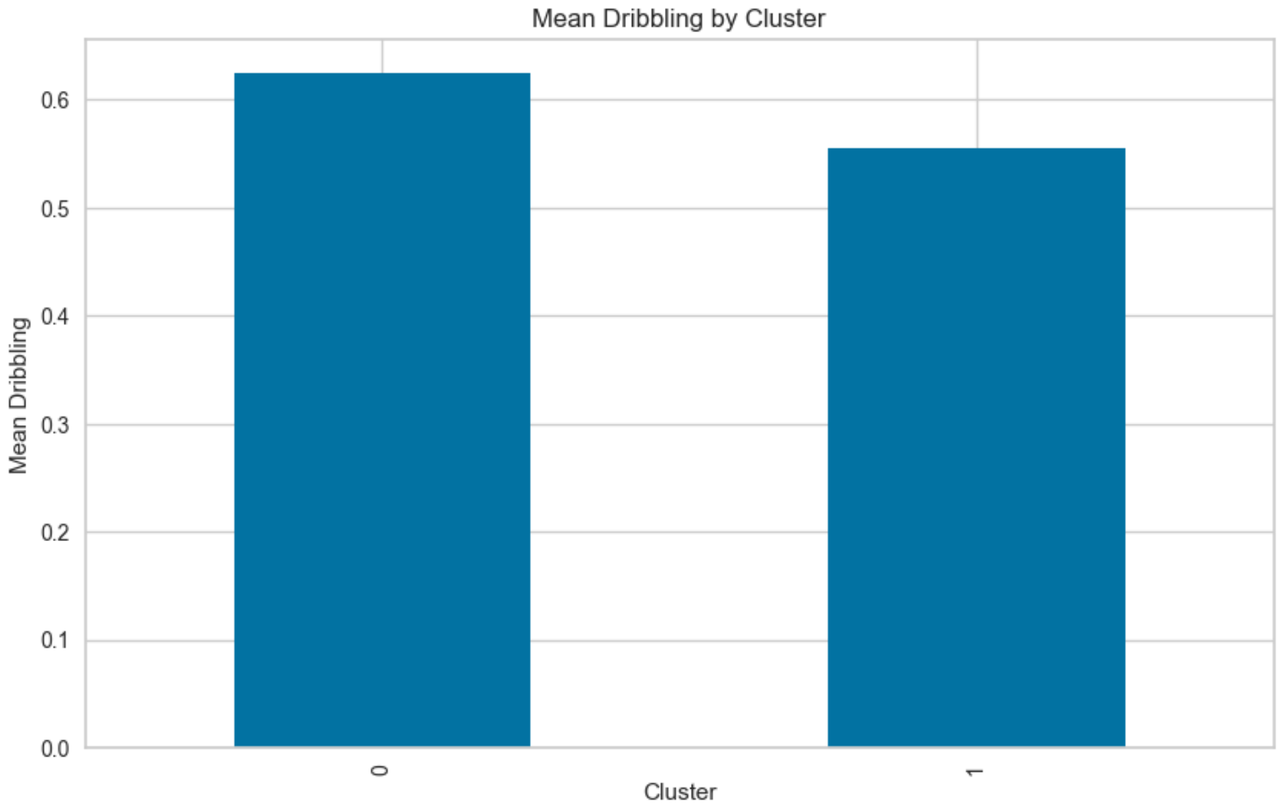
Football Player
Clustering

INTERPRETASI CLUSTERING :
Short Passing



Football Player
Clustering

INTERPRETASI CLUSTERING :
Dribbling

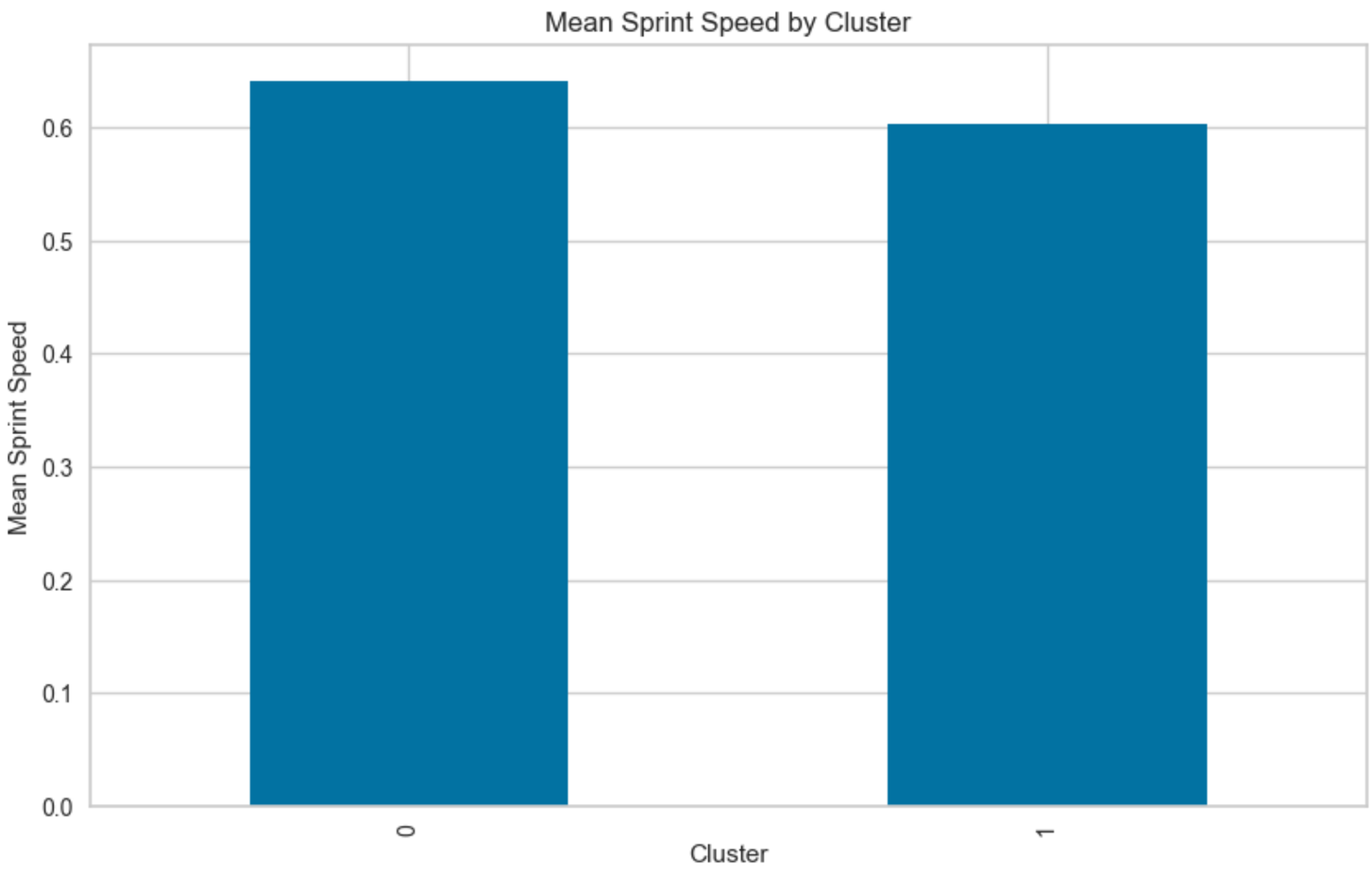


Football Player

Clustering

INTERPRETASI CLUSTERING :

Sprint Speed

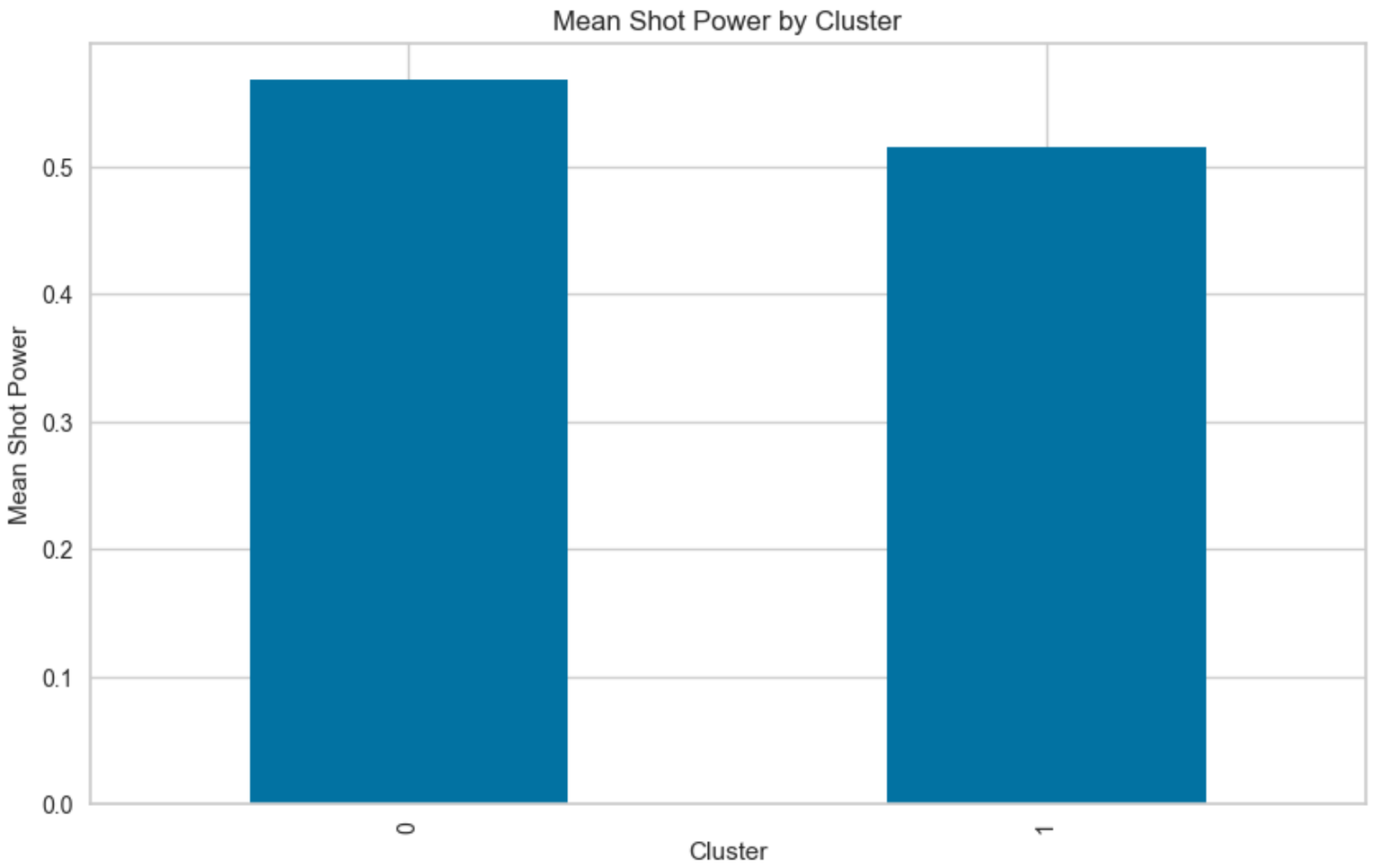


Football Player

Clustering

INTERPRETASI CLUSTERING :

Shot Power

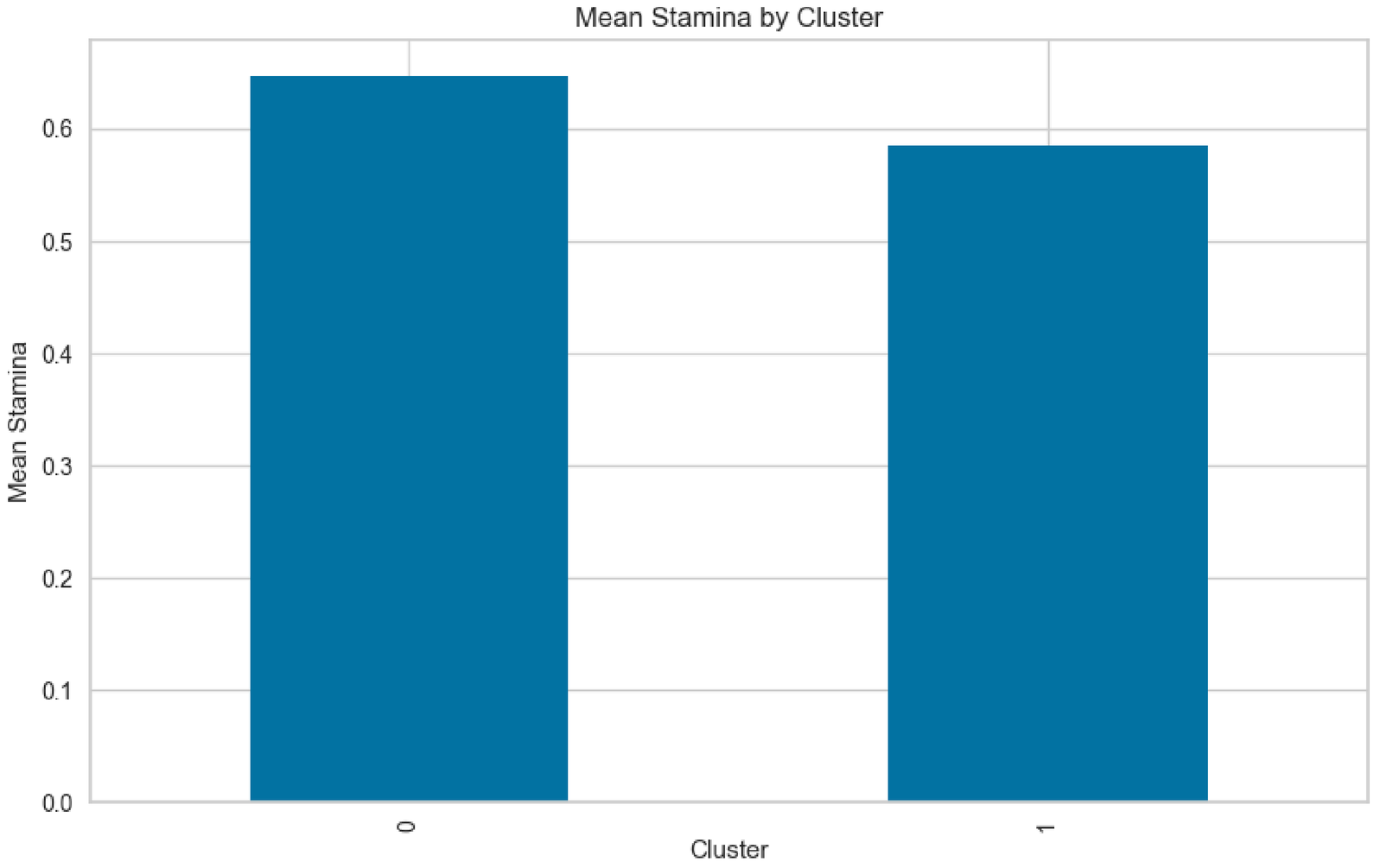


Football Player

Clustering

INTERPRETASI CLUSTERING :

Stamina

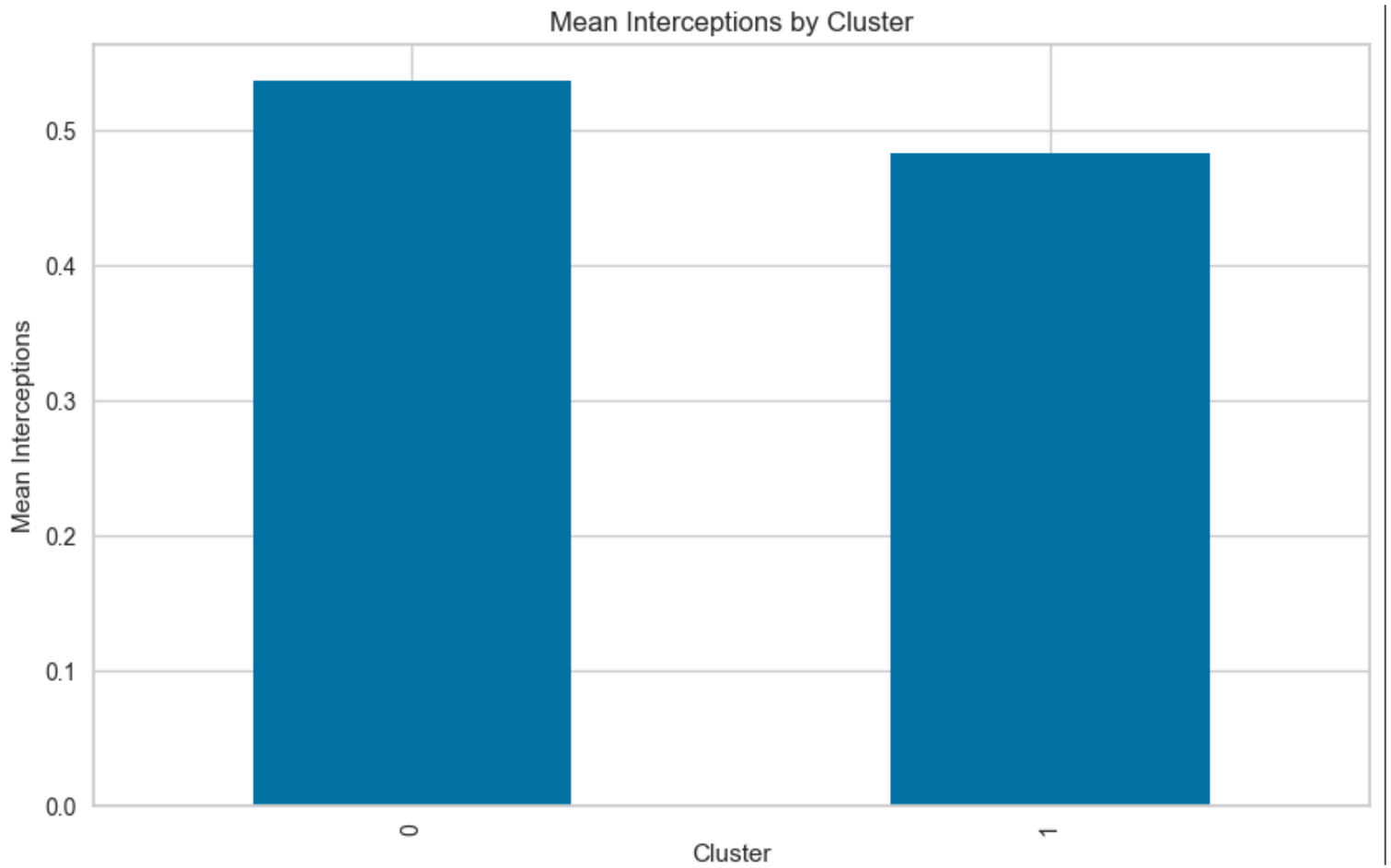


Football Player

Clustering

INTERPRETASI CLUSTERING :

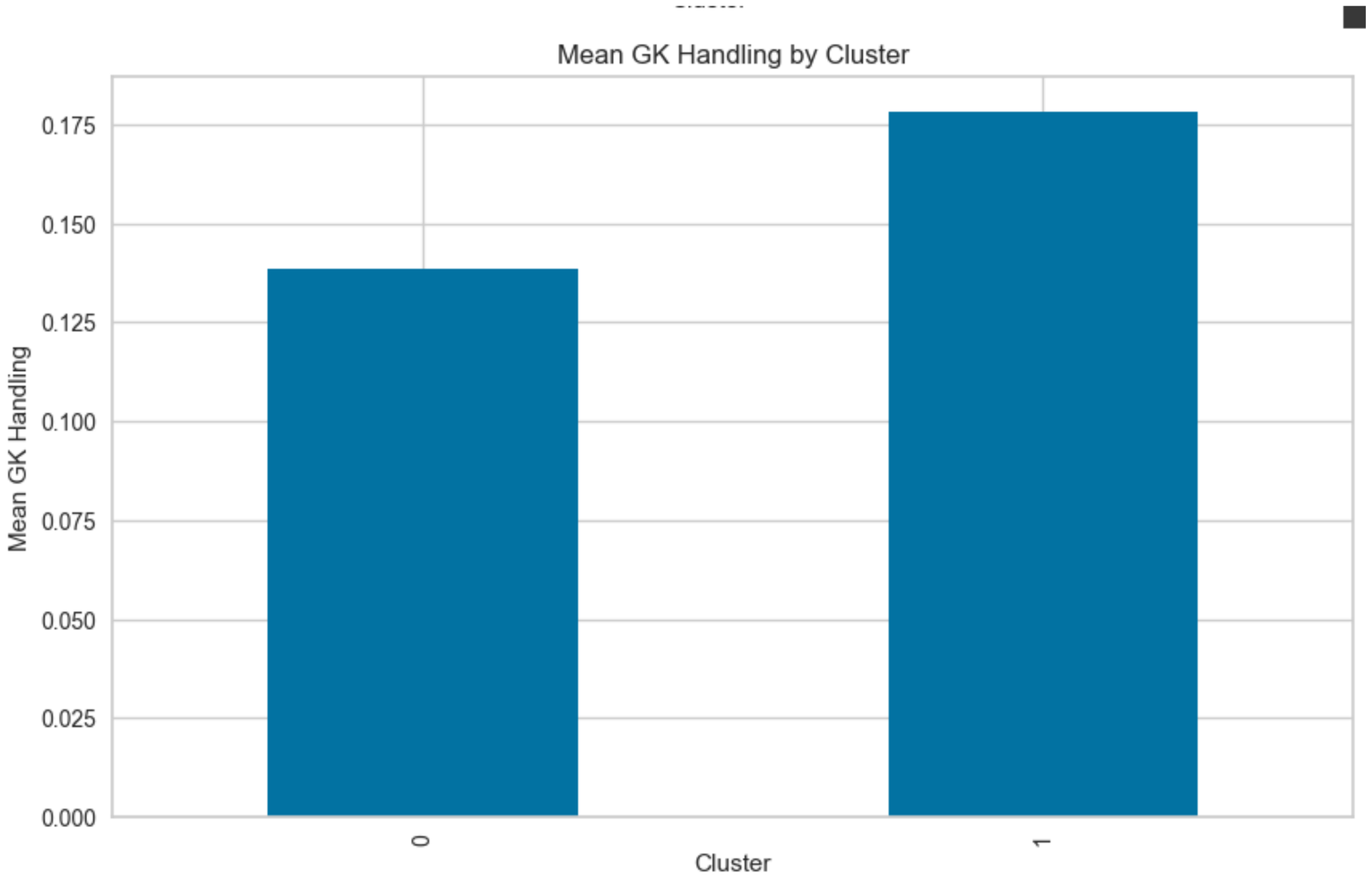
Interceptions



Football Player
Clustering

INTERPRETASI CLUSTERING :

GK Handling



Football Player

Kesimpulan

Berdasarkan hasil clustering diperoleh k=2 yang paling optimal. Hal ini diambil dari average silhouette_coefficient nya adalah **0,66 dan** elbow method di titik **k=2.**

Football Player

Saran untuk Meningkatkan Performa

- Melakukan features engineering dengan membuat variabel baru yang menggabungkan beberapa features yang sejenis
- Mencoba memilih fitur lain yang lebih tepat untuk dapat menggambarkan cluster
- Mencoba metode features reduction lain
- Mencoba metode encode lain
- Mencoba metode scaling lain

KASDD Genap 2022/2023

Thank You

For Your Attention & Cooperation

Kelompok Sebelah