EDA with Tableau

Fifa 18

Project Flow

1. Data Selection: The biggest challenge
   1.1. Contextual analysis of all 6 dataset options revealed significant weaknesses that preclude meaningful analysis.
   1.2. Subsequently, The Fifa 18 dataset was chosen as a mechanism for demonstrating data visualization methods from our coursework.

2. Cleaning and feature creation:
   2.1. Created meaningful categories for player positions: Player Type & Best Position
   2.2. Removed many superfluous attributes, such as player rating by different position.

3. Exploratory Data Analysis:
   3.1. Created many visualizations to demonstrate the relationship between player origin, rating, value, and position.

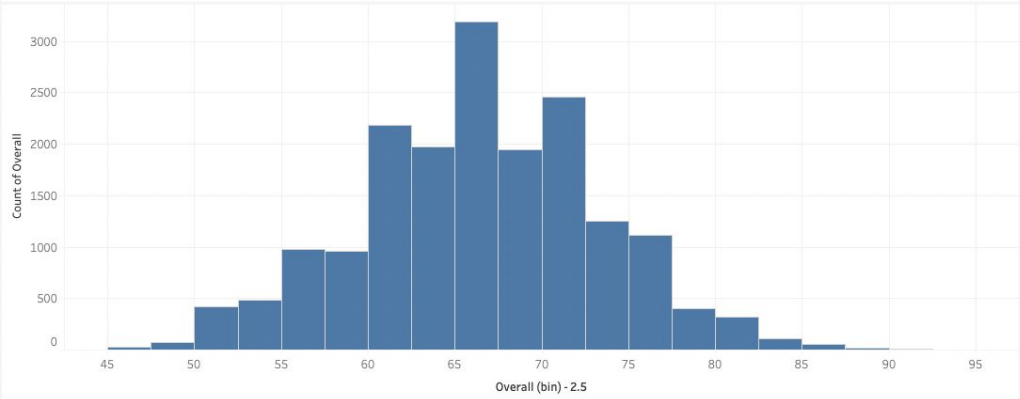
4. Main Question:
   4.1. Unable to support an answer to a significant question using the dataset.

Descriptive Visuals

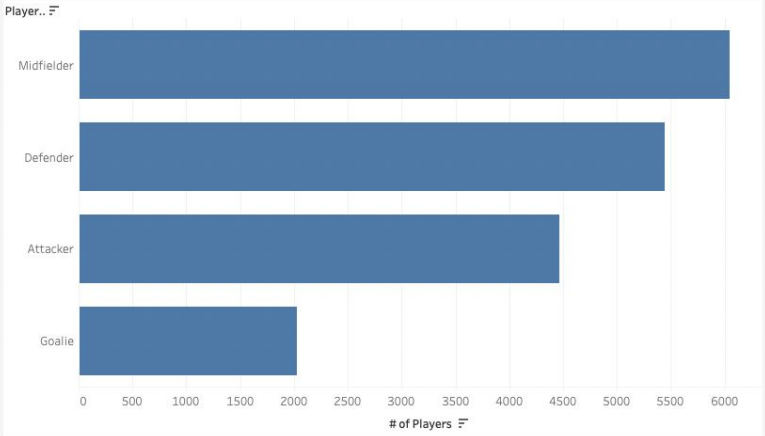
Players by Nationality

	Player Type			
	Attacker	Midfielder	Defender	Goalie
Median Overall	66	67	66	65
Median Agility	71	70	61	38
Median Short passing	62	67	60	26
Median Finishing	64	54	33	12
Median Interceptions	29	55	64	18
Median Ball control	66	66	58	20
Median Sprint speed	72	69	67	41
Median Vision	59	62	45	35
Median Strength	64	64	72	62

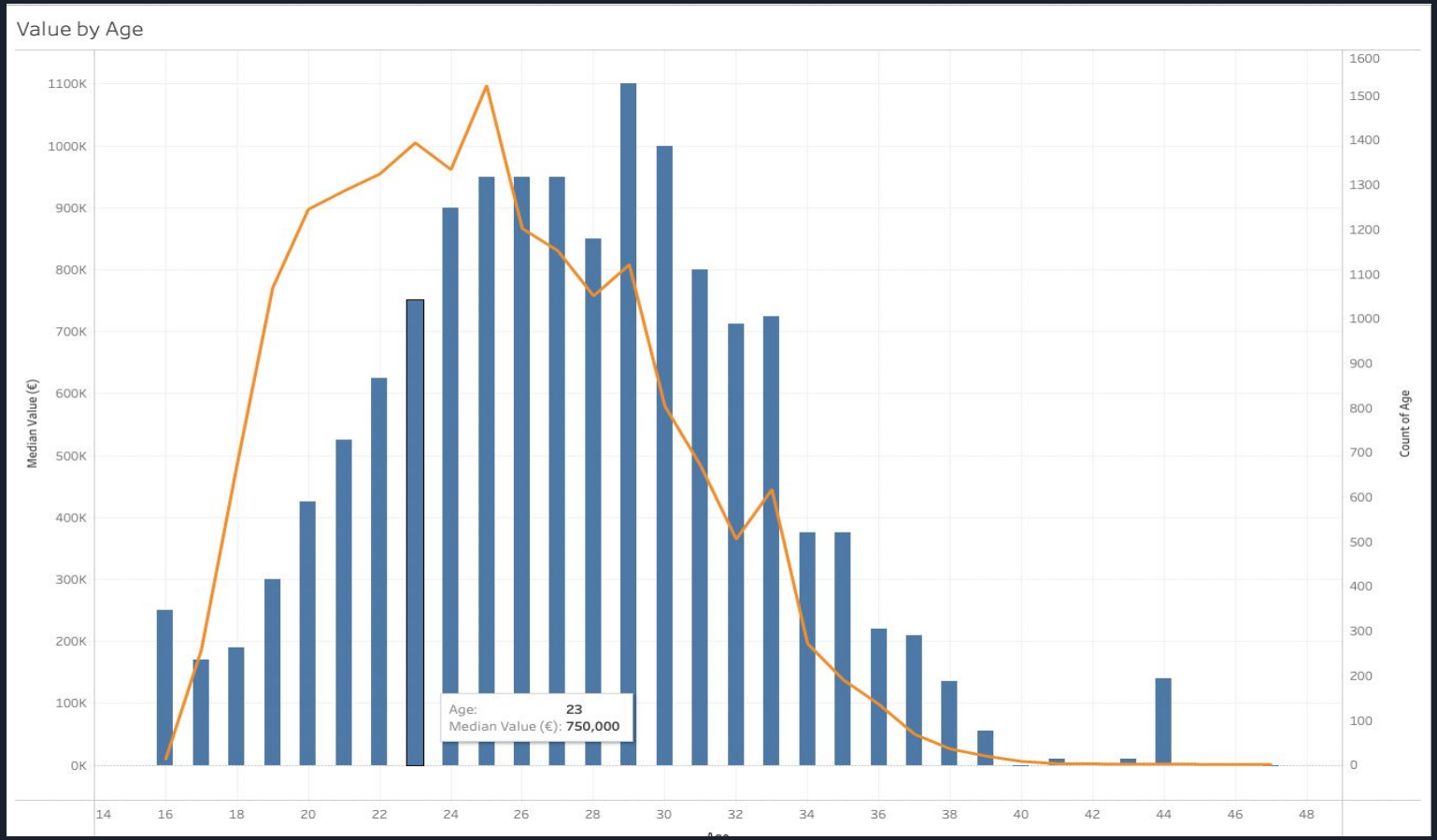
Ratings



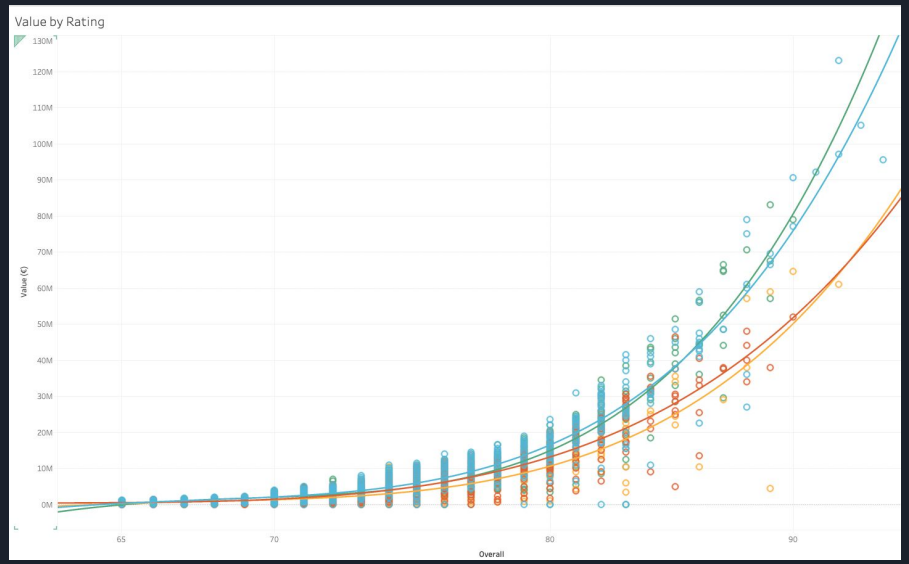
Positions

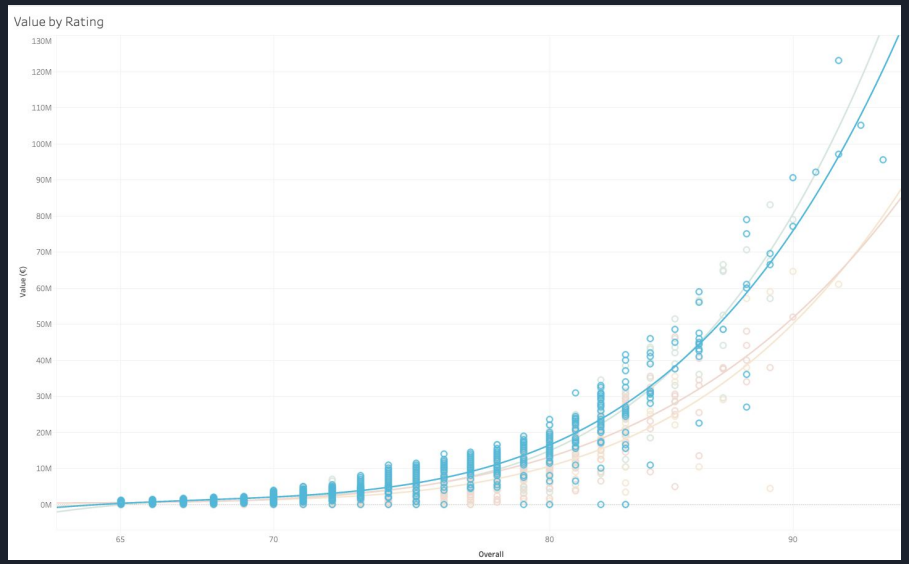


When to Invest

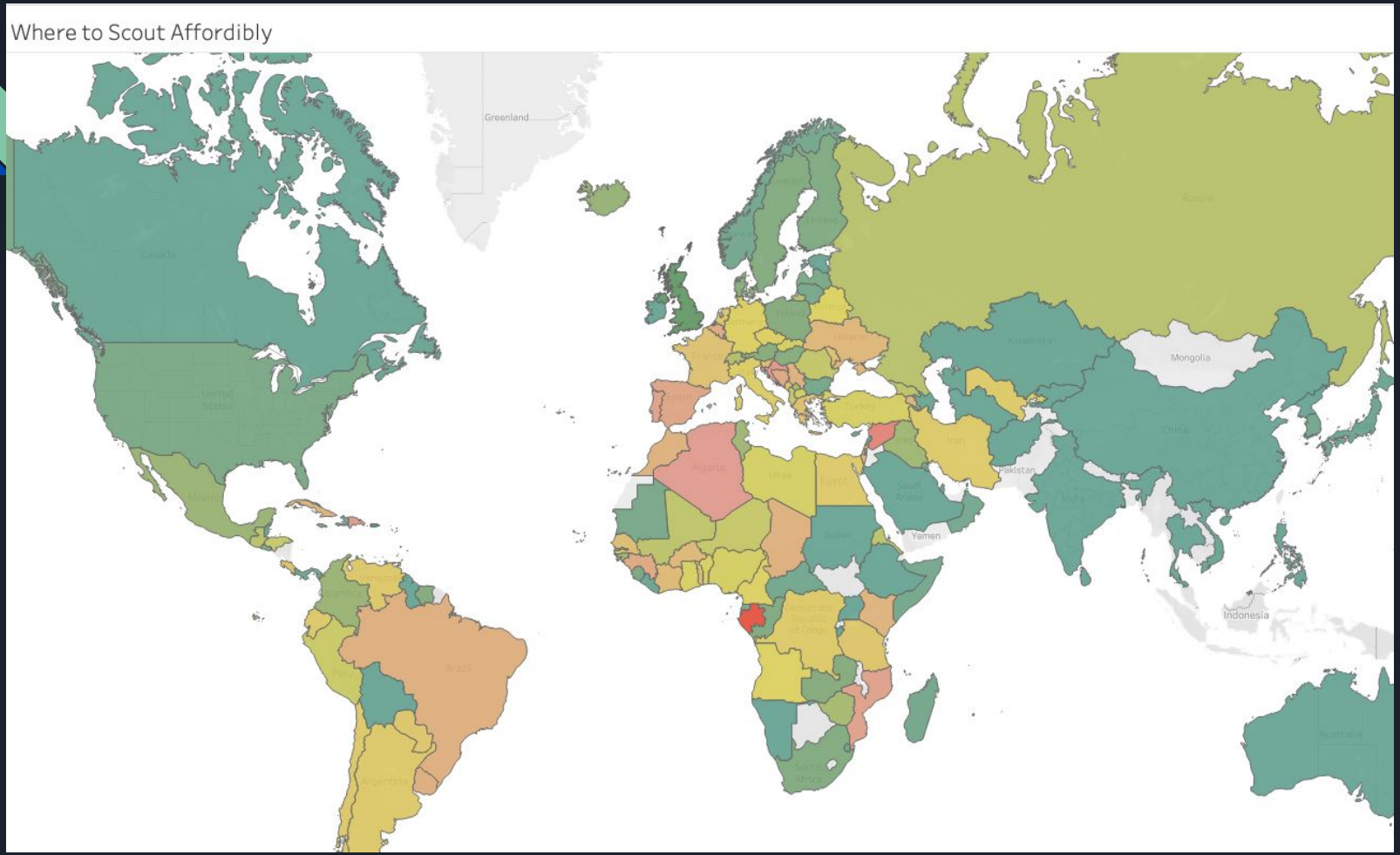


Value by Rating





Where to Scout Affordibly



Key Points

Player Scouting:

- Most affordable places to find talent are the UK, Canada, Scandinavia, and Asia.
- Best age to buy players is 17-19.
- A players highest value is ~25 years of age.
- Highest Value players are attackers & midfielders, with defenders & goalkeepers available for significantly less.
- Most players in the game comes from the UK, followed Eastern Europe, Brazil, and Argentina.

Challenges & Future Improvements

Challenges:

- Context:
  The dataset was curated and massaged for use in the video game. So trends regarding player value, growth, talent, and distribution were only meaningful within the context of the game.

- Modeling:
  The ability to visualize data quickly was useful but did not feel sufficient for meaningful analysis.

Future Improvements:

- Connectivity:
  Connecting to outside datasets, such as population distributions by country, video game sales, or player popularity ratings would provide additional meaningful layers for analysis.

- Question Selection:
  The context and limitations of the dataset felt insurmountable because I attempted to answer a meaningful question outside of the videogame. In actuality, I should have reduced my scope to instead answer how to best play the game.

- Data Curation:
  The majority of the players had ratings that were insignificant (<75 OVR), it may have been useful to focus on the highest rated players in the game since they will yield the most attention from video game players.