Lab Report – 1

CSE564: Visualization
Visual analysis of football players from FIFA 2022 database
Hritam Basak

Aim: The project aims to visualize a Data Set by generating bar-charts, histograms and scatter plots made using JavaScript & D3.js library. These plots help to visualize various attributes (both numerical and categorical) using bar-charts, histograms and scatterplots according to specifications in the lab-1 assignment of CSE-564 Visualization course at Stony Brook University.

Description of Dataset: The dataset has been scrapped from publicly available sofifa website, cleaned, and made public on Kaggle. The datasets provided include the players data for the Career Mode from FIFA 22. This data allows multiple comparisons for different players across 100+ attributes in the game.

The dataset contains information about 19,239 players and their 121 attributes. The dataset has been cleaned and formatted for the specific requirement of this project. Among these 26 attributes are selected for 550 players (randomly selected players). The attributes can be distributed in the following two categories:

1. Categorical Data:
    a. "league_level": Indicates what is the current level of player's club league, ranging from 1 to 5. A higher level indicates player quality is better.
    b. "nationality_rank": The international rank of player's country. Represents if the player is from a very reputed footballing country or not. Ranging from 1-10.
    c. "club_contract_valid_until": The year until which the player is retained in his current club. Ranging from 2022 – 2027.
    d. "work_rate": Defines if a player is hardworking or not. Ranges from 1-9, 1 indicating low work whereas 9 indicates very hardworking.

2. Numerical Data:
    a. "age": Age of player in Fifa 22 game.
    b. "overall": The overall rating of the player in Fifa 22. A good player has a high (>80 rating).
    c. "height_cm": The height of a player. It is positively correlated to player's heading skills.
    d. "weight_kg": The weight of player in kg.
    e. "pace": Indicates how fast a player can run. Interestingly some of the older players had very high pace values.
    f. "shooting": Shooting skills and accuracy of a player. Generally, a striker has high shooting skills, whereas defenders and goalkeepers do not have that.
    g. "passing": How good is the passing (both ground and lofted pass). Midfielders are known for their passing skills.
    h. "dribbling": How good the player can dribble. Midfielders and strikers are known as the best dribblers.
    i. "defending": Defending skill of a player. Needless to mention, defenders have very high value, as compared to strikers.
    j. "physic": Physical contact attribute of a player. Muscular players are having high score.
    k. "attacking_crossing": Crossing skill of a player. Generally, wingers have a high value.

l. "attacking_heading_accuracy": How accurate is a player's heading. Surprisingly defenders have somewhat high value in heading, as compared to the strikers, as heading is a part of defending as well.

m. "attacking_short_passing": Skill of short pass and move. Key attribute for hole-players.

n. "skill_dribbling": How likely is a player to dribble past a defender. Also relates to number of different dribbling tricks.

o. "skill_curve": A high value of this attribute indicates that a player can curl the ball well. Iconic players like David Beckham have very high value.

p. "power_shot_power": Attributes to the generated power while shooting of a player.

q. "power_jumping": How good can a player jump. Strikers and defenders have good jumping skills.

r. "mentality_aggression": This indicates how likely is a player to engage in attacking build-up.

s. "mentality_interceptions": This indicates how likely a player can intercept a ball. Key skill for defensive midfielders and defenders.

t. "mentality_vision": This is a very unique skill of some players, who can map the game really well, and has ability to create chances by providing through balls from interpreting the opponent's movements.

u. "mentality_penalties": How good penalty taker is the player.

v. "wage_eur": Wage of a player in euro in their current clubs. Generally international players have very high wages.

After cleaning the initial data using a python script (Preprocess_data.ipynb), the final data is saved in CSV format in the file final_data.csv.

Why did I think this dataset was interesting?: FIFA 22 player data contains a vast amount of information on the performance and attributes of individual players in the game. By analyzing this data, we can gain insights into player characteristics, strengths, weaknesses, and tendencies, which can be useful for a variety of projects, such as:

1. Building a winning team: By analyzing player data, we can identify players who have the skills and attributes needed to complement one another and build a strong team.

2. Developing player ratings: FIFA player ratings are determined by a complex algorithm that takes into account a wide range of factors, including player performance data. Analyzing this data can help us understand how player ratings are calculated and how they can be improved.

3. Predicting game outcomes: By analyzing player data and team statistics, we can make more accurate predictions about the outcome of FIFA matches.

4. Improving player performance: Coaches and players can use FIFA player data to identify areas of weakness and develop strategies for improving player performance.

Overall, FIFA 22 player data is a valuable resource for anyone interested in understanding the game of soccer and the factors that contribute to player and team success.

Project Learnings and beyond: Overall, this project that visualizes player attributes using JavaScript and D3.js is be a valuable learning experience that helped me develop skills in data visualization, web development, and data analysis. My key takeaways are:

1. Data visualization techniques: I could learn how to use different types of charts and graphs to represent data visually and convey insights effectively.

2. D3.js library: I could learn how to use the D3.js library, which is a popular JavaScript library for creating interactive data visualizations in the browser.

3. Player attributes in FIFA: I could learn about the different attributes that are used to describe players in FIFA, how they are measured, and how they affect player performance.

4. Patterns and trends in player data: By analyzing the visualizations created in the project, I could gain insights into patterns and trends in the player data, such as which attributes are most important for different positions or which attributes are most correlated with player performance.
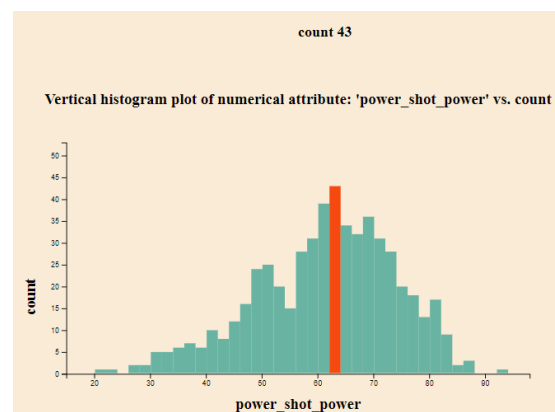
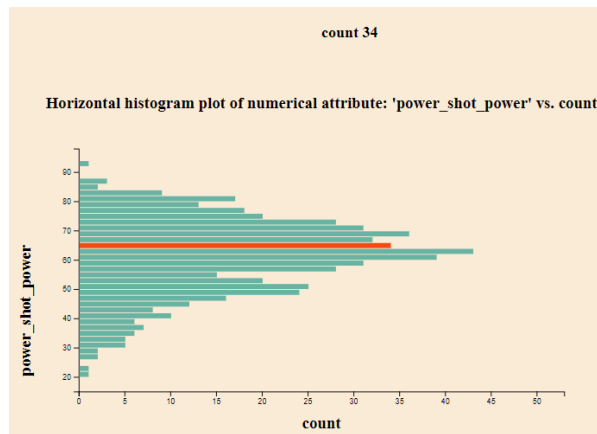A few snapshots of the webpage are attached as follows:
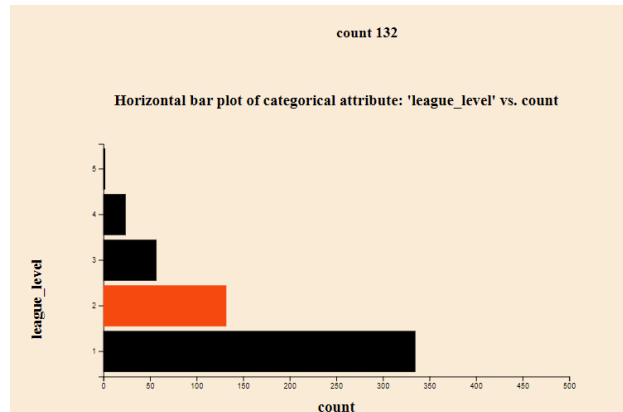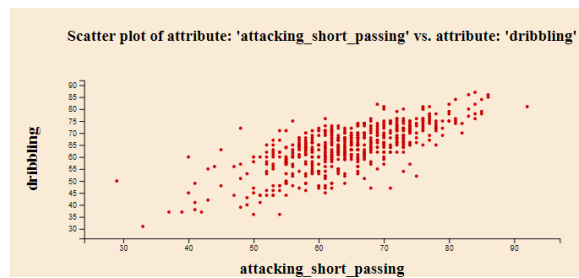

Categorical Data: Bar-chart in horizontal orientation
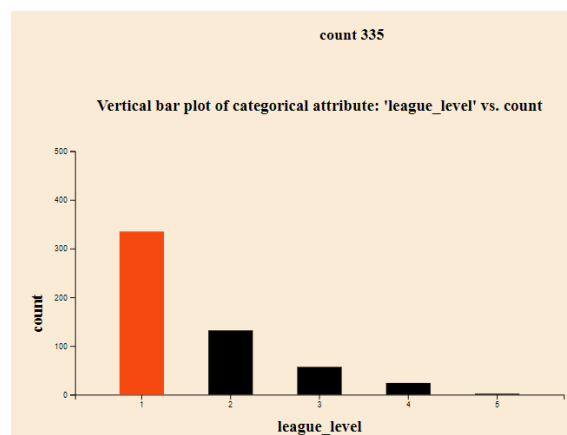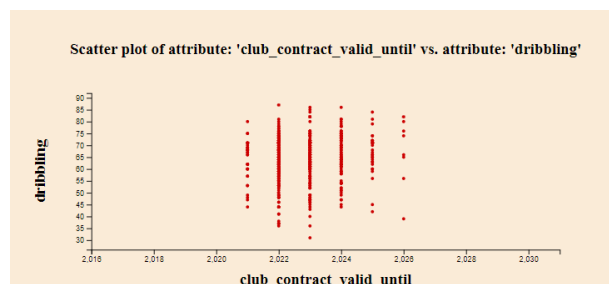

Categorical Data: Bar-chart in vertical orientation
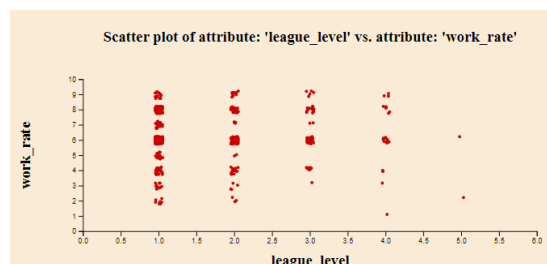

Numerical Data: Histogram in vertical orientation

Numerical Data: Histogram in horizontal orientation

Scatterplot of numerical vs numerical data

Scatterplot of numerical vs categorical data

Scatterplot of categorical vs categorical data