**Final Group Project**

1. **Group Name:** Data Pirates
2. **Project Title:** FIFA 22 Player Analysis
3. **Team Members:**

* Matthew Krivitskiy
* Jemil Patel
* Manita Batas
* Patricia Yrastorza

1. **Summary of group project, accomplishments and learning experience**

* We as a group worked on a project revolving around the player statistics from a popular video-game called FIFA 22. The topic itself made us very enthusiastic and eager to work. Throughout the semester, we learned Python from scratch and all the knowledge we acquired till the end helped us complete the project. We came up with some amazing questions to make our project more interesting and challenging. Finding a good dataset was a difficult task but we managed to find one that had a tremendous amount of information we could work on. Our group utilized a total of six tools that includes Pandas, Matplotlib, JSON processing, API, Web scraping, and Word cloud. After hours of trial and error, and holding numerous meetings each week we finally present our work in the form of a single menu-driven Python program.

1. **Summary of individual accomplishments and learning experience**

* Matthew: Our final group project to cap off CMSC 206 was a great way to practice all the different libraries and tools we learned in this course in Python. The project required our group to use at least six different tools, Pandas, JSON, Matplotlib, Wordcloud, API, and Web scraping. The tools utilized for my question were Pandas and Matplotlib.

My first question was to utilize the pandas function to answer the question, “Who are the highest overall players in FIFA (the actual federation, not the game) from the file “player\_22.csv”. I was able to achieve this using the pandas function I learned through the course.

For my second question I had to utilize matplotlib to showcase the top leagues based on the highest average overall players. Doing this was a simple task as it was also relevant to my first question. The part I struggled on was expanding the width of the name to fit the entire league names, however with some research I learned how to get this accomplished.

Finally my third and final question I once again utilized pandas and matplotlib to showcase the highest rated club team. I based this criteria on highest rated overall, another factor that should be considered is that these soccer leagues are a business and the highest rated teams should take account revenue, however this information was not in the file and is based solely on overall rating of players.

In conclusion this last and final project was a great example to show what the future will be like in completing Python projects with a team which is what we will be doing in the real world and this was a great first time experience to see the pros and cons of working with a team.

* Jemil: The group project has been an amazing learning experience for me. I learned a lot about Python and dealing with datasets and different tools. I utilized a total of four tools: Pandas, Matplotlib, JSON processing, and Web scraping. I used Pandas extensively throughout my questions because without it I could not make use of the dataset in an efficient manner.

For the first question, I played around with the dataset to display the top 10 most valuable players. I formatted the values so that they appear comma-separated which helps the user easily identify the amount. Next, I took all the data of those 10 players and exported it into a JSON file which can help transfer data to any other platform. I made sure to add indentation to make the file more human-readable.

I used Matplotlib in the second question to display a scatter plot between the BMI and overall score of the players. In order to do that, I first made a list of BMI of all the players using the weight and height provided in the dataset. Then, I inserted that list as a column to the dataset so that I could find the correlation value between player BMI and overall score.

The third question was very interesting for me, since I used web scraping to display the flag of the country with the maximum number of players. It took me a lot of trial and error, after which I was able to successfully scrape the flag of a country and display it. During the process I learned how to use the Inspect element feature of Chrome to get HTML tag names through which we can extract any particular thing. Using my common sense I noted that all flag photos had either a “Flag” or “flag” in their src which acted as the distinguishing factor for me from other images present on that page.

Overall I would say that the project was not easy as such since I was the group leader and managed all the meetings, documents, presentations, and code. However, I love coding and so when it comes down to it I can spend hours on it without getting bored. The project also helped me gain leadership and time-management skills.

* Manita: This group project has helped me to learn and use the different libraries and packages while working with datasets. I have used different tools for this project: pandas, matplotlib, JSON, APIs, and word cloud. I have used pandas methods and functions for all my questions to analyze the data and get the required output.

For my first question, I have used the seaborn barplot method to create a bar chart that displays the top 20 nations with the highest number of players. And, I have used wordcloud to generate the nationality names. For my second question, I have used corr() function to find the correlation between the player’s potential and the player’s age and drew a scatterplot using the scatter() function in the matplotlib library.

I ran into a number of errors while working on my third question, for which I wanted to display the different positions in FIFA and the total number of players. The total number of combinations of different players’ positions are 674. Since the total number of combinations was too high to display, I just displayed the different combinations of positions with the top 20 highest number of players. To create a bar chart for this, I used an API to create a URL object that accepts the JSON data via post code. I have entered the data manually to display the different positions with the total number of players. When a post successfully executes, it displays a success message and an url of the chart that was created.

I would like to thank Jemil for always taking the lead and for helpful suggestions regarding the project. I am also thankful to all the group members for their flexibility with the group meetings and discussions.

* Patricia: I learned how to read a csv file and explore DataFrames and Series in many ways using the pandas package so I can work on select columns. I was able to perform different statistical distributions like histograms and scatter plots to calculate correlations among columns. Matplotlib package helped visualize a result in the form of plotting and scaling with x and y axes. I also learned how to generate a word cloud from a column in the dataset.

1. **Next Steps:**

* If we had an additional four weeks, we would have made the project look more attractive and appealing. Our focus then would also be on formatting and displaying minute details.
* Matthew: I think my next step would be to use this information to see what data correlates the most with success. Possibly a program that shows what the key to success has been in previous years for winning teams.
* Jemil: I would have made something like a player card that includes a player’s photo along with some statistics below it. I would have used this idea while displaying the top 10 most valuable players. Other than that, I think I might not do anything differently.
* Manita: I would spend more time going through the dataset to analyze the data and come up with in-depth questions and answer them accordingly. I would also spend more time learning and making use of different tools and libraries to make the code concise and precise.
* Patricia: I would have incorporated a detailed process using numpy and seaborn to obtain the final results with bar graphs and more.