**FIFA 23 Data Analysis Tasks (Pandas & Python)**

**1. Introduction**

This document outlines a series of tasks for analyzing the FIFA 23 player dataset using **Python and Pandas**. The tasks start from basic data handling and gradually progress to advanced analysis, visualizations, and machine learning.

**2. Dataset**

* Dataset: FIFA 23 Player Data
* Download: FIFA 23 Dataset on Kaggle
* File format: CSV

**3. Tasks Overview**

**Level 1: Basic Data Handling (Easy)**

1. Load the dataset using Pandas (pd.read\_csv()).
2. Display the first five rows (df.head()).
3. Check for missing values (df.isnull().sum()).
4. Get basic statistics (df.describe()).
5. Find the number of unique clubs and nationalities (df['Club'].nunique(), df['Nationality'].nunique()).

**Level 2: Filtering & Aggregation (Medium)**

1. Find the **top 10 highest-rated players** based on the Overall rating.
2. List all players from a specific club (e.g., **Manchester United**).
3. Calculate the **average age of players** for each club (df.groupby('Club')['Age'].mean()).
4. Identify the **youngest & oldest player** in the dataset.
5. List the **top 5 most valuable players** based on Value.

**Level 3: Feature Engineering & Analysis (Advanced)**

1. Create a new column for **player experience**:

* Young (≤22 years)
* Prime (23-29 years)
* Veteran (≥30 years)

1. Find the **most common player position**.
2. Calculate the **average player rating per nationality** and list the **top 5 best nationalities**.
3. Identify clubs with the **highest average player rating for players under 23**.

**Level 4: Visualization & Trends (Expert)**

1. Plot a **bar chart** of the top 10 clubs with the highest average player rating using matplotlib and seaborn.
2. Create a **scatter plot** of Age vs. Overall rating.
3. Show a **distribution of player values** using a histogram.
4. Find trends in player attributes (e.g., how Pace and Dribbling correlate).

**Level 5: Machine Learning (Pro)**

1. Train a **Regression Model** to predict a player’s Value based on attributes (Overall, Age, Potential, etc.).
2. Train a **Classification Model** to predict if a player is a World-Class player (Overall > 85).

**Bonus Challenge: Power BI Integration**

1. Export the cleaned dataset to a CSV file (df.to\_csv('cleaned\_fifa23.csv', index=False)).
2. Import the CSV file into **Power BI** and create an **interactive dashboard** with insights on top clubs, player distributions, and values.

**4. Conclusion**

This structured approach ensures a progressive understanding of data analysis using **Pandas, Python, and Machine Learning**. As a next step, you can enhance visualizations, improve machine learning models, and integrate **Power BI** for better insights.