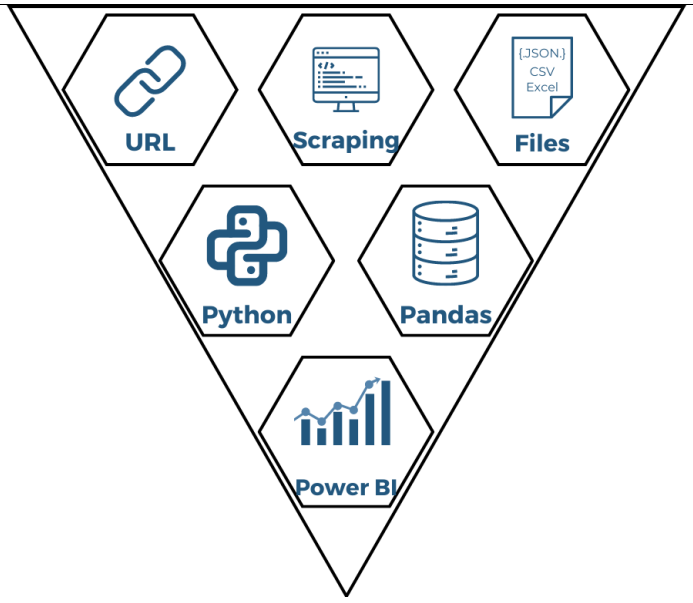
Post Graduate Diploma in Big Data Analytics (PG-DBDA)

March 2023 Batch

Group Number	Group - 5
Guide	Sunil Kumar
Group Members	1. Dhawal Dattani 2. Mohd Yusuf 3. Sudhanshu Prajapati
Title	Cricket World Cup – Team Selection
Objective	The problem entails selecting the optimal Cricket World Cup playing eleven using data from ESPNCricinfo. Employing Python & Bright Data for web scraping, Pandas for data cleaning, Power BI for visualization, and emphasizing National Cricket Development, the objective is to strategically compose the team, accounting for athlete performance and conducting comparative analysis for competitive advantage.
Abstract	This project delved into analyzing the ODI performances of players from the Indian Cricket Team, starting with data exploration of ESPNCricinfo dataset. Utilizing Python and Bright data, web scraping techniques were employed to collect extensive player performance data. Subsequent data cleaning processes were carried out using Google Colab and pandas, addressing missing values and standardizing formats. The project culminated in a comprehensive Power BI dashboard, showcasing insights via visualizations such as line charts, bar graphs, and scatter plots. This dashboard provides a valuable resource for informed decision-making, offering a deep understanding of player metrics and correlations, thus enabling data-driven strategies for enhancing the team's ODI performance

Project Architecture	
Scope of work	The Cricket World Cup team selection process is greatly aided by technology, including Python for web scraping, Pandas and MS Excel for data cleaning, and Power BI for visualization. Python efficiently gathers extensive player performance data from various sources, which is then meticulously cleaned and organized with Pandas and MS Excel for reliable insights. Power BI helps visualize data trends, allowing informed decisions on player selection, ensuring a strategically tailored squad for effective game plans and contingencies. These technological tools enhance the selection process, leading to a competitive team vying for the World Cup title.
Technologies used	**Web scraping –** Python, Bright Data **Data Cleaning–** Python Pandas, Google Colab & MS Excel **Dashboard/Visualization** – Power BI
Application	Sports Analytics Advancement: The methodology can be extended to analyse player performance in various sports, aiding team selection, talent identification, and strategic decision-making using data-driven insights. Player Development Strategies: Implement findings to design personalized training regimens, address weaknesses, and enhance athletes' skills, promoting data-backed development and maximizing potential. Tournament Performance Optimization: Apply the approach to analyse opponents' strengths and weaknesses, enabling tailored strategies for upcoming tournaments, leading to better team performance and success.

Project Timelines (Total:120 hours)	Group form: 14 August. Project title justification: 18 August. Abstract: 18 August. Discussed about database within team: 18 August. GitHub Repository: 18 August Distributed web scraping tasks: 19 August Started Web Scraping: 19 August Synopsis Review: 21 August Finalizing Web Scraped Data: 24 August Additional Data Scraping if required: 25 August Data Cleaning: 26 August Data Manipulation: 26 August Finalizing Power BI Data: 26 August Dashboard draft design: 29 August Dashboard creation: 29 August Final Dashboard: 01 September Report: 04 September Presentation: 05 September