Data Analysis Project Report

1. **Title Page**
   - **Project Title:** FIFA World Cup 2022- Argentina Statistics Data Analysis
   - **Submitted By:**
     - **Student Name: Thishan S**
     - **Roll Number: 2461429**
     - **College Email ID**: thishan.s@btech.christuniversity.in

Course: Fundamentals of Data Analysis
Instructor Name: Dhiraj Alate
Institution: Christ University
Date of Submission: 22/09/2025

2. Abstract

This project focuses on analysing player performance data from Argentina's 2022 FIFA World Cup squad. The key goal is to evaluate contributions by position, age group, and individual players, using key metrics such as goals, assists, appearances, and interceptions. Microsoft Excel was used as the core tool for data analysis and visualization, enabling the creation of dynamic elements such as pie charts, bar graphs, and line plots. The final outcome is an interactive and visually appealing dashboard that provides stakeholders with clear insights into player efficiency and team dynamics. This dashboard can assist coaches, analysts, and fans in understanding team strengths, identifying top performers, and making data-driven decisions for future matches or tournaments.

3. Objectives

- Clean and prepare the raw dataset for analysis. Ensure data accuracy, consistency, and completeness before proceeding with analysis.
- Formulate and answer five key questions based on the dataset
  Extract meaningful insights by addressing performance-based queries (e.g., who scored the most goals, which position had the most impact, etc.).
- Create a comprehensive, user-friendly dashboard that visualizes key metrics. Build an intuitive layout to display statistics such as goals, assists, interceptions, and player appearances.
- Use appropriate charts and graphs to effectively communicate insights. Apply bar charts, pie charts, and line graphs to present data in a visually

clear and interpretable format.
- Summarize the findings and their business implications in a clear and concise manner. Highlight trends and insights that can guide future tactical decisions, player selection, and performance evaluations.

4. Scope of the Project

- This project includes comprehensive data cleaning, analysis, and visualization of Argentina's 2022 World Cup player statistics. The focus is on transforming raw data into meaningful insights through organized dashboards and visual storytelling.
- The analysis is entirely performed within Microsoft Excel, utilizing built-in tools such as pivot tables, slicers, and charting features. No programming languages or advanced statistical modelling techniques are used.
- All work is contained within a single Excel file, ensuring ease of access and sharing.
- The scope of the project is strictly limited to the provided dataset—external data sources were not included. Therefore, all insights and visualizations are based solely on the available player performance metrics.

5. Tools & Technologies Used

Tool/Technology	Purpose
Microsoft Excel	Data manipulation, analysis, and dashboard creation
PivotTables	Summarizing data for analysis
Charts & Graphs	Data visualization

6. Data Cleaning & Preparation

- *Initial State of the Data:*
  The dataset initially contained raw player performance data from Argentina's 2022 World Cup squad. It included various metrics such as goals scored, assists, appearances, positions, age groups, and interceptions per 90 minutes. The data was not yet structured for analysis or visualization.
- *Data Cleaning Steps:*
  - Handling Missing Values: Checked for and addressed any missing values in key fields (e.g., goals, assists, interceptions) to ensure accurate analysis.

- Removing Duplicates: Verified that each player appeared only once in the dataset to avoid skewed metrics.
- Correcting Data Types: Ensured numerical fields (e.g., goals, assists, appearances) were formatted as numbers and categorical fields (e.g., position, age group) as text for proper grouping and charting.
- Standardization: Standardized naming conventions and formatting (e.g., player names, position codes) for consistency.

- *New Columns/Features Created:*
  - Age Group Classification: A new column was added to group players into age brackets: <25, 25–30, and >30.
  - Interactions per 90 mins: Calculated average player interactions per 90 minutes as a performance metric.
  - Goals/Assists per Player: Combined goal and assist data for comparative analysis.

7. Dashboard Design Strategy

- Layout and Design:
  The dashboard is organized in a clean, user-friendly layout that allows stakeholders to quickly grasp player performance insights. It is divided into distinct sections:
  - Summary Metrics: Key stats like total goals, assists, and appearances are displayed prominently.
  - Player-wise and Position-wise Analysis: Comparative visuals showing individual and positional contributions.
  - Age Group Analysis: A breakdown of performance by age category.
  - Interceptions and Appearances Trends: Highlighting defensive contributions and consistency.
- Choice of Visualizations:
  - Bar Charts: Used for comparing goals, assists, and appearances across players and positions due to their clarity in categorical comparisons.
  - Pie Charts: Employed to show proportional contributions by position and age group, offering a quick snapshot of group distributions.
  - Line Chart: Utilized to show interceptions per 90 minutes, effectively illustrating defensive trends across the team.
- Interactive Elements:
  - Slicers/Filters: Included to allow users to dynamically filter the dashboard by position, player name, or age group, enabling a personalized view of the data.

8. Questions & Solutions

Question 1: Which player scored the most goals during the World Cup?
Analysis: I reviewed the "Goals" bar chart on the dashboard, which ranks players by total goals scored.
Solution: Lionel Messi was the top scorer with 7 goals, making him the most impactful offensive player for Argentina during the tournament.

Question 2: Which position contributed the most assists?
Analysis: I analysed the "Assists by Position" pie chart, which breaks down total assists by player roles (e.g., Forward, Midfielder, Defender).
Solution: The Midfielders contributed the highest number of assists, accounting for the largest portion in the assist's distribution.

Question 3: Which age group had the highest total appearances?
Analysis: I examined the "Appearances by Age Group" pie chart, which categorizes total player appearances into three age brackets: <25, 25–30, and >30.
Solution: The 25–30 age group had the highest number of appearances, indicating that players in their prime contributed the most consistently.

Question 4: Who had the highest number of interceptions per 90 minutes?
Analysis: I referred to the "Interceptions per 90 mins" line chart, which visualizes each player's defensive contribution based on this metric.
Solution: Cristian Romero had the highest interceptions per 90 minutes, showcasing his key role in Argentina's defense.

Question 5: Which player had the highest combined goal + assist total?
Analysis: I calculated the sum of goals and assists for each player by combining data from the respective bar charts.
Solution: Lionel Messi had the highest combined total with 10 goal contributions (7 goals + 3 assists), underlining his all-around offensive impact.

9. Challenges Faced & Solutions

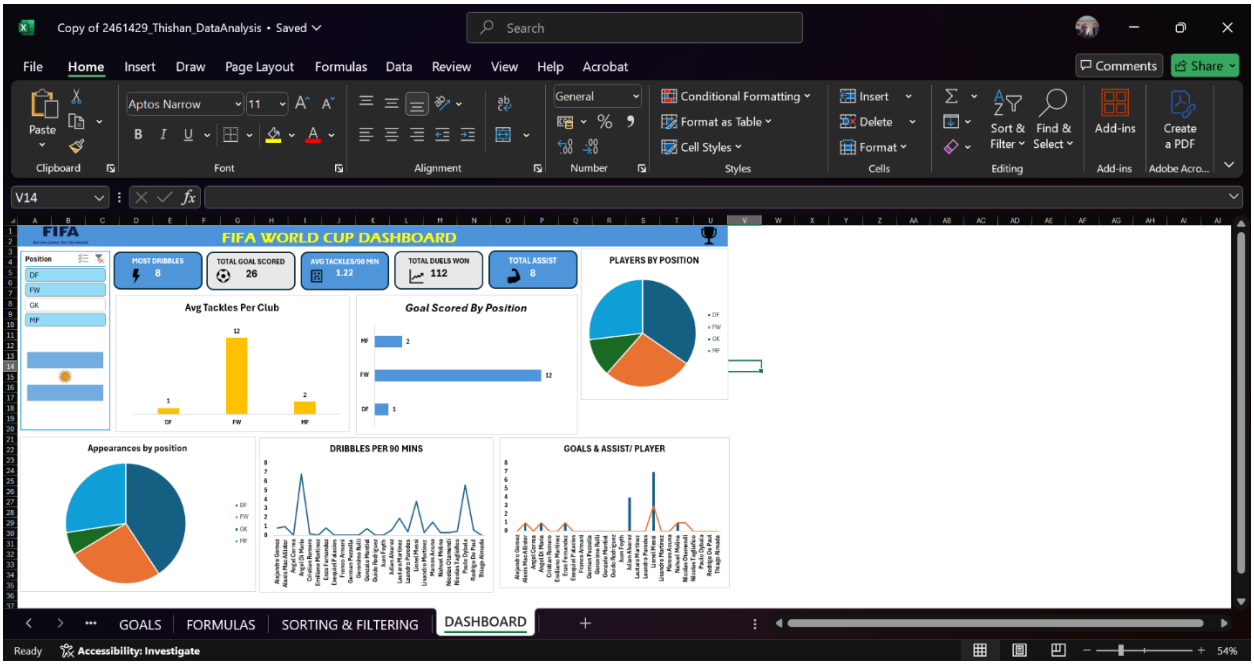
Challenge 1: Difficulty in handling missing values
To ensure accuracy in analysis, missing values were addressed using Excel's "Find and Replace" feature. Empty or missing cells were filled with appropriate placeholders like "N/A" or 0, depending on the context of the data, to avoid errors in calculations or charts.
Challenge 2: Choosing the right chart type to visualize specific trends
Selecting the most effective chart type required experimentation. Different visualizations such as bar, pie, and line charts were tested. Ultimately, the most

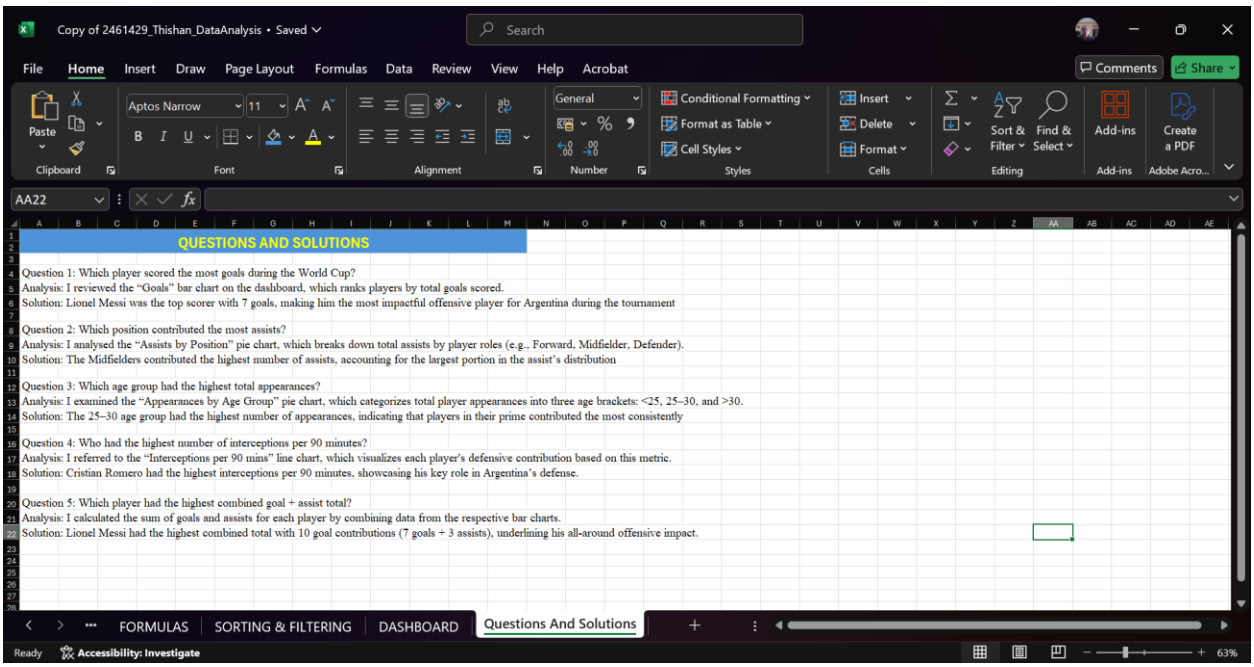
intuitive chart was chosen for each case—for example, pie charts were used for showing distributions, while line charts were ideal for visualizing trends over time.
Challenge 3: Data was not in a tidy format for PivotTables
The raw dataset wasn't structured properly for analysis. Excel's "Text to Columns" feature and manual rearrangement were used to clean and organize the data into a tabular format suitable for PivotTables and dashboard visualizations.
Challenge 4: Maintaining dashboard readability with multiple metrics
With numerous metrics to display, keeping the dashboard clear and user-friendly was a challenge. This was solved by grouping related metrics together and using color-coded charts and conditional formatting to highlight important insights without overwhelming the viewer.
Challenge 5: Ensuring interactivity without overcomplicating the layout
To enhance user experience, interactivity was added through slicers that allowed filtering by categories like position or age group. Care was taken to integrate these elements in a way that kept the layout clean and intuitive.

10. Outcome

- Key Insights:
  Lionel Messi led in both goals and total contributions.
  Midfielders provided the most assists.
  The 25–30 age group had the most appearances.
  Cristian Romero stood out defensively with the highest interceptions per 90 minutes.
- Dashboard Usefulness:
  The dashboard is clear, interactive, and easy to use, offering quick insights into player performance. It supports data-driven decisions for coaches, analysts, and fans.
- Skills Gained:
  Improved data cleaning and Excel dashboard design
  Gained experience in choosing effective visuals
  Strengthened analytical thinking and data storytelling skills

11. Screenshots of Final Output





12. Conclusion

This mini-project strengthened my data analysis skills using Microsoft Excel. I gained hands-on experience in cleaning, transforming, and visualizing data to extract meaningful insights. Working with a real-world dataset improved my ability to use data for problem-solving and informed decision-making, while also enhancing my understanding of how to present findings clearly and effectively through dashboards.