1. INTRODUCTION

Crop Production Analysis in India

Crop production analysis in India focuses on understanding the factors influencing agricultural productivity, such as soil quality, weather conditions, irrigation facilities, and government policies. Given that agriculture is a critical sector in India, contributing significantly to the economy and food security, data science plays a vital role in analyzing historical trends and predicting crop yields. By leveraging machine learning techniques like regression models and time series forecasting, along with tools like Python and Tableau, stakeholders can identify patterns, optimize resource allocation, and improve productivity. This analysis helps policymakers and farmers make data-driven decisions to enhance crop outputs and sustainability.

FIFA World Cup Analysis

FIFA World Cup analysis involves evaluating team and player performance using historical match data to uncover trends and predict future outcomes. By analyzing key metrics such as goals scored, possession percentages, player contributions, and match statistics, data science techniques like classification models and clustering algorithms provide insights into winning strategies and player performances. Tools like Python, along with libraries such as pandas and scikit-learn, are used to perform Exploratory Data Analysis (EDA) and build predictive models. This analysis helps teams, analysts, and fans better understand the dynamics of the game and anticipate match results.

IBM HR Analytics - Employee Attrition & Performance

Employee attrition and performance analysis focuses on identifying factors that lead to employee turnover and evaluating workforce productivity to enhance organizational performance. Using employee data such as demographics, job roles, salary, and satisfaction levels, data science models like Logistic Regression, Decision Trees, and clustering techniques predict attrition and highlight at-risk employees. These insights enable HR departments to address critical issues like job dissatisfaction, work-life balance, and career growth, helping reduce turnover rates and improve retention. Tools like Python, Power BI, and Tableau are widely used to analyze and visualize the findings for actionable strategies.

Heart Disease Diagnostic Analysis

Heart disease diagnostic analysis uses patient medical data to identify risk factors and predict the likelihood of heart-related conditions. By analyzing clinical indicators like blood pressure, cholesterol levels, age, and lifestyle habits, data science techniques such as Logistic Regression, Support Vector Machines (SVM), and Random Forest models help predict heart disease with high accuracy. This analysis assists healthcare professionals in diagnosing conditions early and developing preventive strategies. Leveraging datasets like the UCI Heart Disease Dataset, along with tools such as Python and visualization platforms like Power BI, this approach contributes significantly to improving patient care and reducing mortality rates.

1.1 About Industry

Unified Mentor, established in 2022, is a forward-thinking ed-tech company based in Gurgaon, India. It specializes in providing **online certification courses** and **live training sessions** in fields such as **data science**, **digital marketing**, and **web development**. Founded by Paras Grover and Sanket Patil, Unified Mentor aims to bridge the skill gap for professionals seeking career advancement in today's competitive market

Key Offerings:

1. **Certification Programs**:
    - Courses are tailored for beginners and professionals alike, focusing on hands-on learning.
    - The curriculum includes industry-relevant projects to ensure practical understanding.
2. **Internship Opportunities**:
    - Unified Mentor integrates internships into its programs, allowing students to apply theoretical knowledge in real-world settings.
3. **Live Training**:
    - Interactive sessions with industry experts provide real-time guidance and problem-solving support.

4. **Focus Areas**:
    - **Data Science**: Programs include machine learning, AI applications, and data analytics.
    - **Digital Marketing**: Training in SEO, social media management, and online branding strategies.
    - **Web Development**: Courses on modern frameworks, responsive design, and full-stack development

Mission and Vision:

Unified Mentor envisions creating a globally competitive workforce by making high-quality, accessible education available to all. Their emphasis on practical skills and mentorship ensures learners are well-equipped to meet industry demands.

Location and Accessibility:

The company's headquarters are located at SCO 17-18, Street 31C, Gurgaon, India. The programs are accessible globally, ensuring that learners from various regions can benefit from their services

2. PLAN OF INTERNSHIP PROGRAM

Week 1: Explored foundational Python concepts, including data types, variables, operators, lists, strings, and conditional statements, with practice and implementation.

Week 2: Focused on Python functions, loops, tuples, dictionary methods, and Numpy basics, including array creation, reshaping, indexing, slicing, and operations.

Week 3: Learned Pandas operations such as working with Series and DataFrames, concatenation, file handling, grouping, and creating date ranges.

Week 4: Delved into EDA by addressing missing and duplicate values, handling outliers using IQR, and creating visualizations with Seaborn and Matplotlib.

Week 5: Studied statistical measures of central tendency and dispersion, analyzing univariate, bivariate, and multivariate plots with relevant statistical insights.

Week 6: Applied learning to case studies on Wine Quality, Employee Dataset, and Telecom Dataset, integrating descriptive statistics and data visualization.

Week 7: Mastered Tableau by analyzing the Superstore dataset, learning charting techniques, formatting methods, and creating dashboards and stories.

Week 8: Advanced Tableau skills with additional charting techniques, filtering methods, calculated fields, and enhancing dashboards for practical business challenges.

Week 9: Conducted exploratory data analysis and visualizations on the Crop Production dataset, identifying trends and preparing predictive models.

Week 10: Finalized Crop Production dashboards, presented insights, and generated comprehensive reports with actionable solutions for stakeholders.

Week 11: Analyzed FIFA World Cup data through cleaning, EDA, and identifying patterns and trends using statistical and visualization techniques.

Week 12: Built Tableau dashboards for FIFA data, generated insights, created visual stories, and completed a detailed project report for presentation.

Week 13: Worked on the Heart Disease Diagnostic Analysis project, exploring clinical data, conducting EDA, and building predictive models to classify heart disease risks. Designed Tableau dashboards to visualize key insights, highlighting risk factors and their impacts.

Week 14: Focused on the IBM HR Analytics: Employee Attrition and Performance dataset, identifying trends in employee retention and exploring factors influencing attrition. Created interactive Tableau dashboards to present insights into employee performance and attrition trends.

Week 15: Finalized advanced analysis for Heart Disease and IBM HR data, refined Tableau dashboards, and presented key findings with actionable recommendations.

Week Range	Milestone	Key Focus
Week 1-4	Foundational Concepts	Python basics Data structures Numpy Pandas Introductory visualization
Week 5-7	Intermediate Analytics and Statistics	Statistical concepts Exploratory analysis Advanced visualization
Week 8	Tableau Mastery	Learning Tableau software and implementing case studies
Week 9-10	Crop Production Analysis	Dataset exploration Predictive modeling Dashboard creation
Week 11-12	FIFA World Cup Data Analysis	Visualization and reporting using Tableau and Python
Week 13–14	Heart Disease and IBM HR Analysis	Performed EDA, predictive modeling, and built Tableau dashboards for insights
Week 15	Advanced Dashboard Refinement	Enhanced dashboards, prepared reports, and presented key findings for both projects.

3. LEARNING EXPERIENCE

3.1 Knowledge Acquired

Analyzing Diverse Datasets : Acquired the ability to analyze large-scale datasets across multiple domains—agriculture, sports, healthcare, and HR—using Python and visualization tools such as Tableau and Power BI. This included interpreting data trends, predictive modeling, and deriving actionable insights to support domain-specific objectives.

• **Exploration of Historical Data :** Gained comprehensive insights by exploring datasets like FIFA World Cup statistics, clinical data for heart disease, and historical crop production trends. These analyses provided a deeper understanding of the evolution of sports tournaments, healthcare diagnostics, and agricultural patterns over time.

• **Understanding Key Influencing Factors :** Developed the skill to identify and evaluate critical factors influencing outcomes, such as crop yield determinants (climate, soil type), match outcomes (goal conversion rates, possession, player efficiency), and employee attrition (job satisfaction, performance ratings).

• **Predictive Modeling and Advanced Analytics :** Learned to implement machine learning models for predicting crop production and diagnosing heart disease risks. This involved handling data preprocessing, feature engineering, and creating accurate, scalable solutions for real-world challenges.

• **Data Visualization for Story telling :** Mastered the art of data storytelling by creating clear, engaging, and impactful visual narratives using Tableau dashboards. Visualizations were designed to convey complex data insights effectively to diverse audiences, such as policymakers, healthcare professionals, HR teams, and sports analysts.

• **Domain-Specific Insights**

1. **Sports Analytics:** Gained the ability to analyze match statistics, player performance, and team strategies using data-driven approaches to provide historical perspectives and identify winning patterns.
2. **Healthcare Analytics:** Acquired knowledge of identifying risk factors for heart disease and presenting actionable insights for diagnostics through predictive models and dashboards.
3. **Agriculture Analytics:** Learned to explore agricultural data, identify crop yield trends, and predict future production to assist in data-driven policymaking.
4. **Human Resource Analytics:** Gained insights into employee attrition and performance trends, empowering HR teams to implement retention strategies and improve productivity.

• **Integrated Approach to Problem-Solving :** Developed the ability to integrate tools and techniques across Python, Tableau, and machine learning to address domain-specific problems. This holistic approach to analytics ensured scalability, efficiency, and clarity in deriving and presenting solutions.

3.2 Relevant Description

- **Dataset Analysis**

  Worked extensively with comprehensive datasets, including CSV files which served as a foundation for analyzing tournament results and player performances. These datasets offered rich insights into match outcomes, goal-scoring patterns, and performance trends, making it possible to extract key takeaways related to team dynamics and success factors in the tournament.

  - **Data Cleaning and Preprocessing**

  Gained significant experience in data cleaning and preprocessing to ensure data accuracy and usability. This involved handling missing values, normalizing data formats, and transforming raw data into a structured format suitable for analysis. Through this process, I developed a keen understanding of the importance of data quality in obtaining meaningful insights and actionable conclusions.

  - **Trend Analysis**

  Conducted in-depth trend analysis on key metrics such as team performance over time, goals scored in different match conditions, and the impact of external factors like weather and match location on tournament outcomes. This analysis helped uncover hidden patterns that correlate with success in the World Cup, such as the role of team strength in specific environments or weather conditions.

  - **Interactive Dashboards**

  Developed visually intuitive and interactive dashboards using Tableau and Power BI, which summarized key insights such as player and team performance metrics, historical trends, and tournament predictions. These dashboards were designed to make complex data easily accessible and understandable, empowering stakeholders to explore trends and metrics with ease.

- **Power BI Dashboards:** Created tools for stakeholders to explore key metrics, including:
  - **Year-over-Year Trends:** Visualized player performance across different countries using line charts and area graphs, offering insights into changing trends over time.
  - **Top Goal Analysis:** Designed bar charts to highlight the most frequently scored goals, revealing regional and national patterns, and emphasizing trends in goal-scoring priorities during the tournament.
  - **Enhanced Data Exploration and Storytelling**

  The dashboards presented a seamless way for users to explore key metrics and trends. This not only made data-driven insights more accessible to a broad audience but also served as a tool for making informed predictions and decisions, ultimately enhancing the storytelling aspect of the data through compelling visual narratives.

3.3 Skills Learnt

Technical Skills:

Data Preprocessing: Mastered techniques for cleaning, structuring, and preparing data for analysis, ensuring high-quality inputs for models and visualizations.

Python Libraries: Gained proficiency in essential Python libraries like Pandas for data manipulation, NumPy for numerical computations, and Matplotlib/Seaborn for data visualization.

Visualization Tools: Learned to use Tableau and Power BI effectively for creating interactive and dynamic dashboards that communicate data insights in an engaging way.

Analytical Skills:

Key Performance Indicators: Developed the ability to identify and evaluate critical performance metrics that influence match outcomes and overall team success.

Trend Analysis: Acquired the expertise to analyze historical and real-time data for trends, helping to make data-driven predictions and insights.

- **Soft Skills:**

Storytelling with Data: Enhanced the ability to translate complex datasets into meaningful stories that resonate with diverse audiences, including technical teams, coaches, and fans.

Presentation Skills: Improved confidence and effectiveness in presenting findings clearly and concisely, whether through reports, dashboards, or oral presentations.

3.4 Tools Used

- **Programming Language:**

Python: Utilized Python extensively for data processing, analysis, and visualization. Its flexibility and wide range of libraries enabled efficient handling of large datasets, building custom workflows, and automating repetitive tasks to streamline the analysis process.

- **Visualization Tools:**

Tableau: Leveraged Tableau to create intuitive, user-friendly dashboards, providing deep insights into team and player performance, historical trends, and tournament highlights. Tableau's powerful visualization capabilities helped transform complex data into accessible and impactful visual narratives for diverse audiences. Key features utilized in Tableau included:

Geographic Visualizations: Integrated geographic data into maps to represent regional crop distributions, match outcomes, or player statistics, helping uncover regional trends and making data-driven decisions based on location.

Interactive Dashboards: Developed multi-dimensional views of data through interactive dashboards, allowing stakeholders to explore key metrics such as team performance, goals scored, and trends in real-time.

Calculated Fields and Dynamic Filters: Enhanced dashboards with calculated fields, dynamic filters, and tooltips, making it easy for users to engage with the data and extract relevant insights.

Power BI: Used Power BI to design interactive visualizations and reports, offering actionable insights into match statistics, player performance, and fan engagement. The tool's interactive features, such as slicers and filters, enabled stakeholders to explore data dynamically.

o **KPI Dashboards:** Designed Key Performance Indicator (KPI) dashboards to monitor progress against predefined goals. Metrics such as production efficiency, cost-effectiveness, and profitability were highlighted, enabling stakeholders to evaluate real-time performance.

o **Interactive Features:** Used slicers, filters, and dynamic visualizations to enable stakeholders to interact with the data. This allowed personalized exploration of the dataset, focusing on specific dimensions or attributes of interest and enhancing decision-making processes.

• **KPI Dashboards:** Designed Key Performance Indicator (KPI) dashboards to monitor progress against predefined goals. Metrics such as production efficiency, cost-effectiveness, and profitability were highlighted, enabling stakeholders to evaluate real-time performance.

o **Dashboard Design:** Created interactive dashboards that provided multi-dimensional views of the data. Features like calculated fields, dynamic filters, and tooltips enhanced the usability and interactivity, making it easier for users to extract insights.

• **Python Libraries:**

• **Pandas:** Employed Pandas for efficient data manipulation, including filtering, aggregating, and restructuring large datasets to make them suitable for analysis. This library was crucial for cleaning and preparing data for further analysis.

• **NumPy:** Applied NumPy for numerical computations, handling complex mathematical operations such as matrix calculations, statistical measures, and data transformations within datasets.

- **Matplotlib/Seaborn:** Used Matplotlib and Seaborn to create high-quality visualizations that represented trends, distributions, and relationships in the data, making it easier to interpret key insights.

- **Data Preprocessing:**

Handling Data Quality: Python played a crucial role in addressing data quality issues such as missing values, employing techniques like imputation or deletion to ensure clean datasets.

Data Scaling: Utilized scaling techniques like Min-Max Scaling and Standardization to ensure uniformity across numerical features, preparing the data for effective analysis.

Categorical Encoding: Employed encoding techniques such as one-hot encoding and label encoding to handle categorical variables efficiently, allowing them to be used in machine learning models and analytical processes.

Power BI:

- **Business-Oriented Reports**: Leveraged Power BI to develop detailed reports and dashboards, focusing on business-oriented metrics that emphasized actionable insights and decision-making support.
- **Interactive Features:** Used slicers, filters, and dynamic visualizations to enable stakeholders to interact with the data. This allowed personalized exploration of the dataset, focusing on specific dimensions or attributes of interest and enhancing decision-making processes.
- **Dashboard Design:** Created interactive dashboards that provided multi-dimensional views of the data. Features like calculated fields, dynamic filters, and tooltips enhanced the usability and interactivity, making it easier for users to extract insights.

- **Geographic Visualizations:**

Tableau Geographic Data: Tableau's ability to integrate geographic data was utilized to create intuitive maps showcasing regional crop distributions. This feature helped uncover regional trends and provided geographic insights for more informed decision-making in agricultural analytics.

- **Data Sources:**

Common CSV Files: The projects utilized various common types of datasets across different domains. These typically included:

Tournament Data: Historical data related to tournaments (e.g., FIFA World Cup), which includes match outcomes, team statistics, player performances, and more.

Match-Level Data: Detailed match-level data containing scores, player statistics, formations, and key events.

Healthcare Data: Clinical data, including patient characteristics, test results, and diagnostic information, used for predictive modeling in health-related projects.

HR Data: Employee datasets containing performance metrics, demographic information, and attrition-related factors used for analyzing employee engagement, retention, and performance.

Agricultural Data: Crop production data, including yield trends, regional agricultural patterns, and environmental factors influencing crop production.

4. OBJECTIVE OF INTERNSHIP ACTIVITY

4.1 Problem Identification and Solution

1. Heart Disease Diagnostic Analysis

Problem Identification: Heart disease risk factors are not always clearly identified, leading to delayed diagnosis and prevention.

Solution: Analyze clinical data to identify key factors influencing heart disease risk and create a predictive model. The solution will include a dashboard to visualize key insights and help healthcare professionals in early diagnosis and intervention.

2. Crop Production Prediction

Problem Identification: Accurate crop yield predictions are challenging due to fluctuating environmental factors, affecting food security and resource management.

Solution: Analyze historical crop production data to identify key factors impacting yields, and develop a predictive model to forecast future crop production for better planning and resource optimization.

3. FIFA World Cup Analysis

Problem Identification: Identifying key factors that influence World Cup victories remains complex, despite the availability of data.

Solution: Analyze historical World Cup data to uncover patterns in team and player performance, external factors like weather, and create interactive dashboards to visualize insights for better understanding of World Cup outcomes.

4. IBM HR Analytics: Employee Attrition & Performance

Problem Identification: High employee attrition and poor performance impact company productivity and morale.

Solution: Analyze HR data to identify factors contributing to attrition and low performance, and develop strategies to improve retention and employee satisfaction, supported by interactive dashboards for real-time insights.

4.2 Activity Carried Out and Relevant Details

Data Collection and Preprocessing:

- **Data Collection:** Gathered relevant datasets from reliable sources, ensuring the data is comprehensive and aligns with the project's objectives (e.g., medical, agricultural, sports, or HR data).
- **Data Cleaning:** Addressed missing values, outliers, and inconsistencies in the datasets using techniques like imputation, removal, or transformation to ensure data integrity.
- **Standardization:** Standardized data formats (e.g., team names, employee IDs, or crop types) to enable seamless analysis and consistency across datasets.
- **Feature Engineering:** Created and transformed new variables based on the dataset to better capture underlying trends or relationships (e.g., calculating goals per match, employee attrition rate, or crop yield trends).

Exploratory Data Analysis (EDA):

- **Descriptive Statistics:** Performed initial statistical analysis to understand distributions, correlations, and trends within the data.

- **Trend and Pattern Identification:** Identified significant trends and patterns by exploring different variables (e.g., heart disease rates by age/gender, crop production by region, performance metrics in football, or attrition rates in HR).
- **Data Visualization:** Utilized visualization libraries (such as matplotlib, seaborn) and tools like Tableau or Power BI to represent data visually. Created charts, graphs, and heatmaps to showcase patterns and relationships.

Key Metrics and Factor Identification:

- **Metrics Identification:** Identified key performance indicators (KPIs) or metrics that are critical to understanding the problem (e.g., goals per match, employee satisfaction score, or crop yield per hectare).
- **Factors Influencing Outcomes:** Analyzed the impact of various factors on outcomes, such as weather for crop production, player performance for FIFA, employee engagement for HR, or clinical factors for heart disease prediction.
- **Correlation Analysis:** Explored correlations between variables (e.g., heart disease risk factors, crop production metrics, or team performance stats) to uncover underlying relationships.

Dashboard Development and Visualization:

- **Interactive Dashboards:** Developed interactive and user-friendly dashboards using Tableau or Power BI. These dashboards allowed users to explore data dynamically by providing drill-downs, filters, and sorting options.

- **Custom Visualization:** Incorporated various visualizations such as bar charts, line graphs, heatmaps, and pie charts to represent key metrics and insights effectively.
- **Real-Time Insights**: Enabled users to interact with the dashboards to focus on specific subsets of data (e.g., by year, region, team, or employee) for detailed exploration.

Integration and Testing:

- **Tool Integration:** Integrated data analysis with visualization tools (Tableau/Power BI) and predictive models into a cohesive workflow.
- **Testing:** Conducted rigorous testing to ensure the accuracy and reliability of the analysis, predictions, and dashboards. This included validation of the data models and the functionality of the visualizations.
- **User Feedback:** Gathered feedback from mentors, peers, and potential users to refine the tool, improve user experience, and ensure that the final output meets expectations.

Reporting and Findings:

- **Insights and Conclusions:** Provided actionable insights based on the analysis, highlighting critical findings, trends, and relationships (e.g., predicting heart disease risk, identifying performance gaps in HR, understanding factors leading to World Cup wins, or improving crop yield).
- **Final Report:** Compiled a final report with the methodology, data analysis, key findings, and recommendations for stakeholders. The report was supplemented by the developed dashboards to allow for further exploration of the data.

4.3 Outcome of Internship Activity

1. Placement:

- The internship enhanced my employability in analytics roles by providing hands-on experience in data analysis, visualization, and predictive modeling.
- The skills acquired during this project are directly applicable to roles in data science, business intelligence, and agricultural analytics.

2. Publication:

- The findings and methodologies developed during this internship have the potential to be submitted to reputed agricultural journals for publication.
- The insights generated from the tool can contribute to academic research and policy-making in the agricultural sector.

3. Upgradation:

- Acquired advanced skills in Tableau, including designing dynamic dashboards and implementing interactive visualizations.
- Gained expertise in predictive modeling techniques and machine learning frameworks.
- Improved proficiency in Python programming, focusing on libraries used .

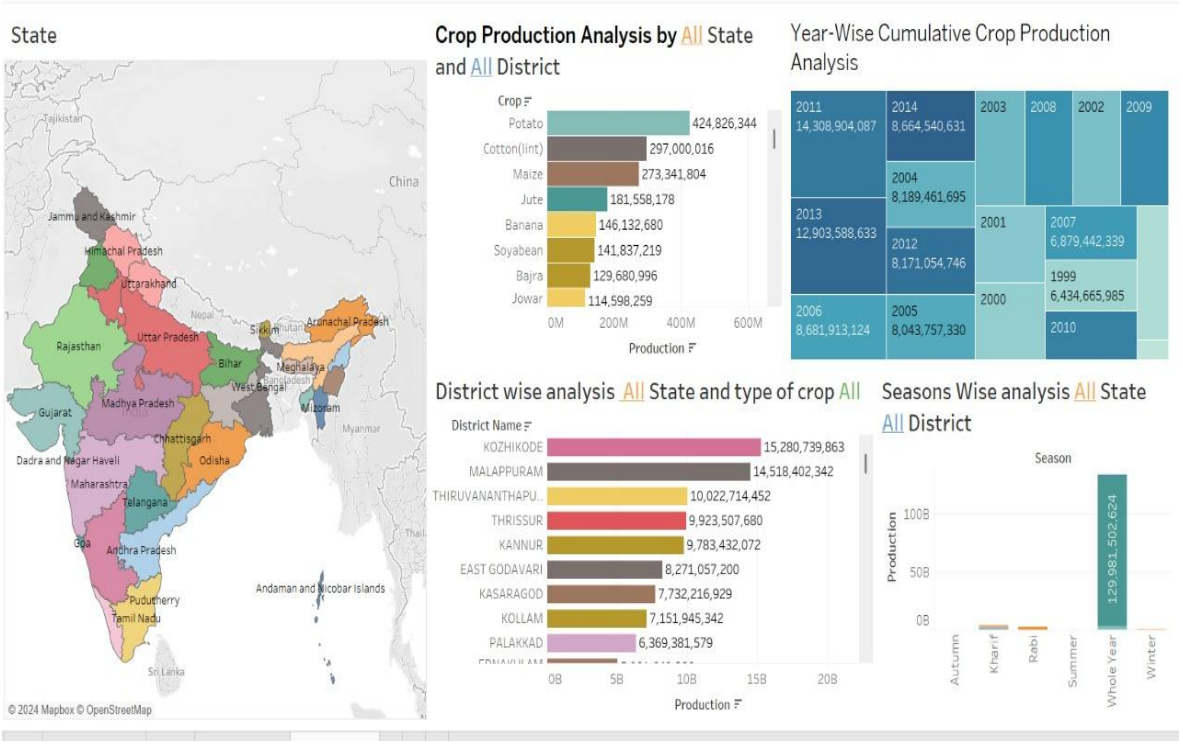
Participating in a data science internship offers transformative learning experiences by combining theoretical knowledge with real-world applications. Interns gain hands-on experience in handling datasets, building predictive models, and using tools such as Python, R, SQL, and machine learning frameworks like TensorFlow or scikit-learn. Through guided projects, interns learn to clean and preprocess data, explore patterns, and create actionable insights for businesses.

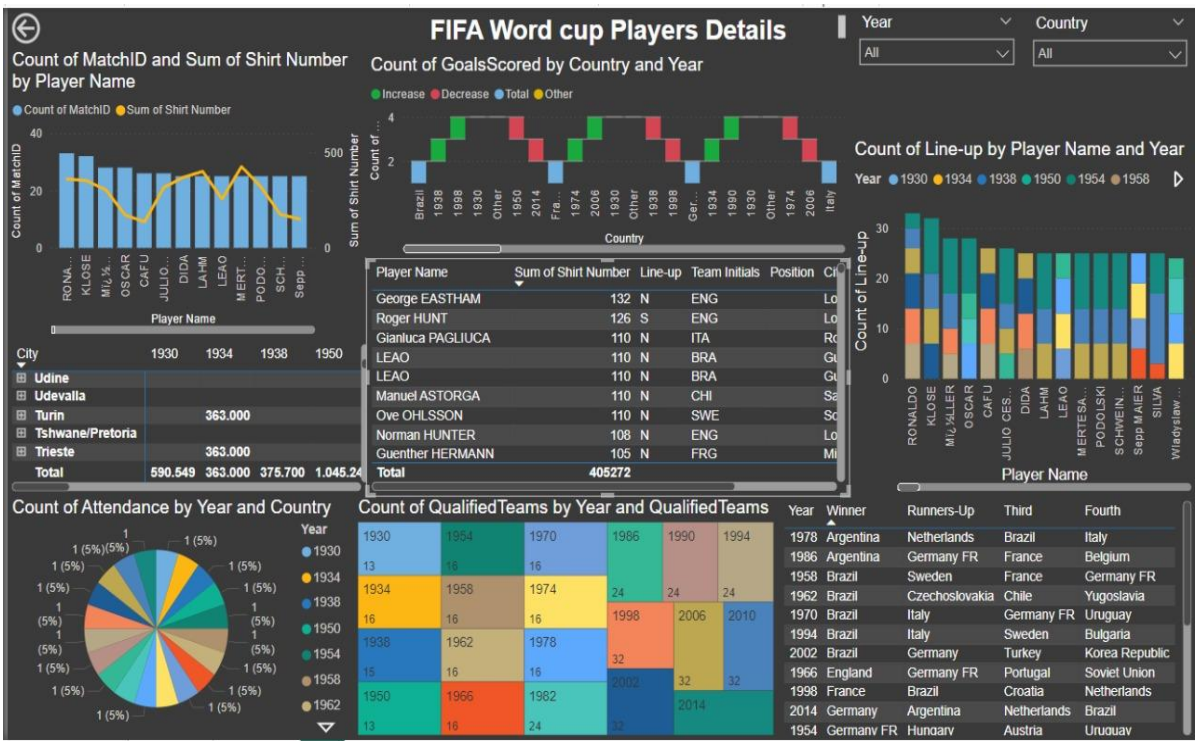
These activities enhance problem-solving skills, reinforce statistical and programming concepts, and improve collaboration by working on cross-functional teams. data science internship fosters innovation by encouraging interns to explore advanced techniques like deep learning, natural language processing (NLP), and big data analytics. Interns often work on diverse projects, such as building recommendation systems, predictive analytics for business optimization, or customer segmentation models.

The outcomes of such internships extend beyond technical growth. Interns develop a deeper understanding of the data science pipeline—from data acquisition to model deployment. Exposure to real-world challenges helps them build industry-relevant portfolios, boosting employability. Additionally, interns often sharpen their communication skills by presenting findings through data visualizations and reports, making them more effective at translating technical results into actionable business strategies. This holistic growth ensures readiness to tackle complex data problems in professional settings.

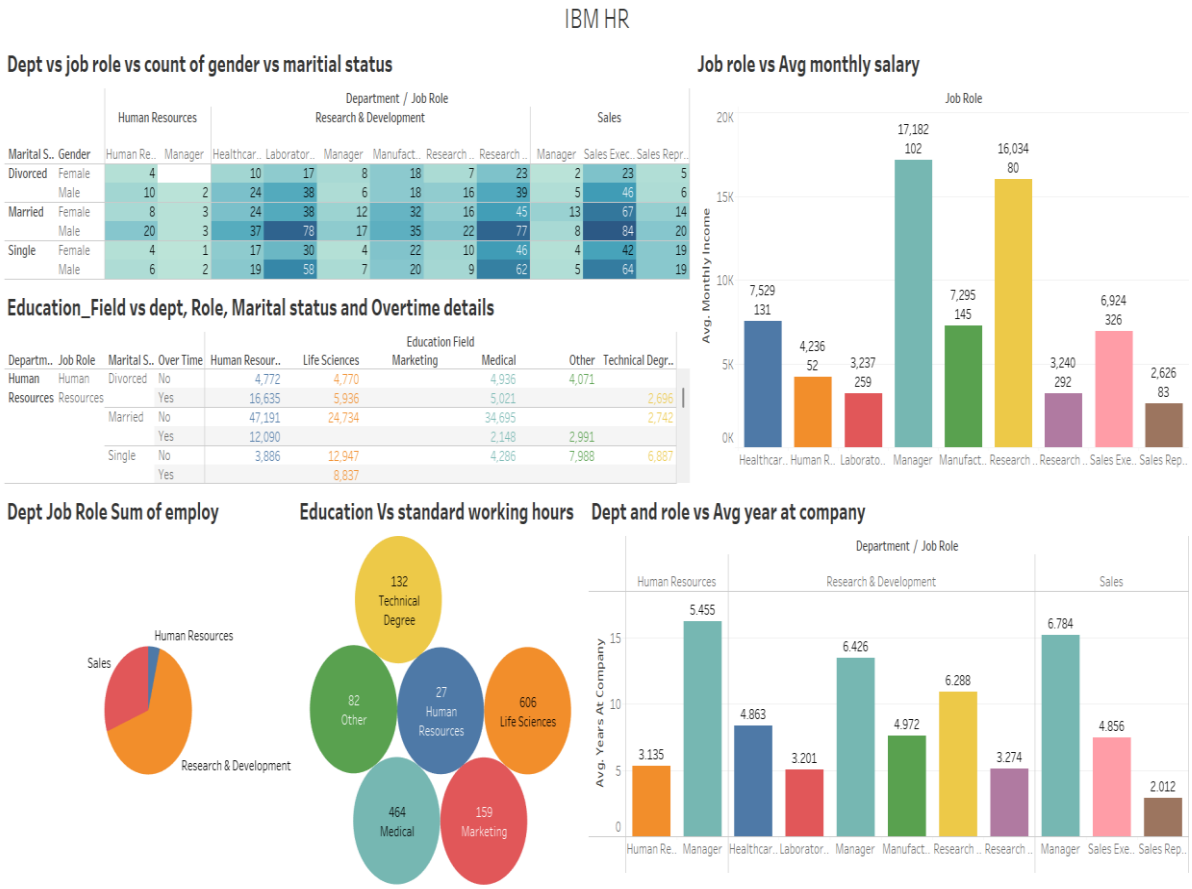
4.4 Relevant Screenshots

- **Dashboard Visualizations:**



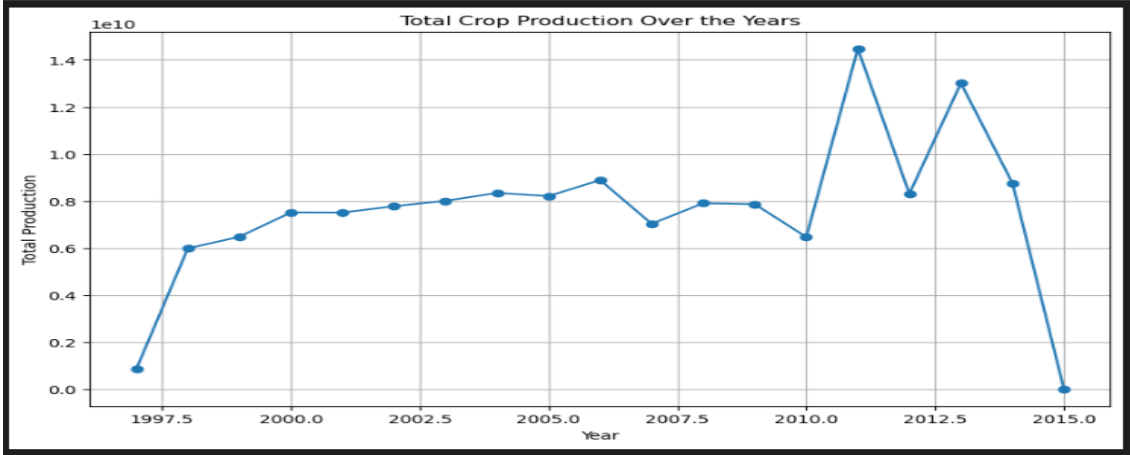



Heart Disease Diagnostic Analysis
Sex vs ST depression induced by exercise relative to rest
Sex
Female
Male
287.4
810.9
Age vs Exercise-induced angina
Exang
53.865
55.557
No
Yes
Serum cholesterol (mg/dl).
55.849
312
Female
53.815
713
Male
Age vs Sex
Sex
312
713
Female
Male
Chest pain type (categorical)
Disease
300
Male
No Disease
413
Male
Disease
226
Female
Age bin and Thalach(Maximum heart rate achieved)
Count of Age
Age (bin)
Resting blood pressure (mm Hg).
312
Female
41,715
713
Male
93,187
Exit full screen

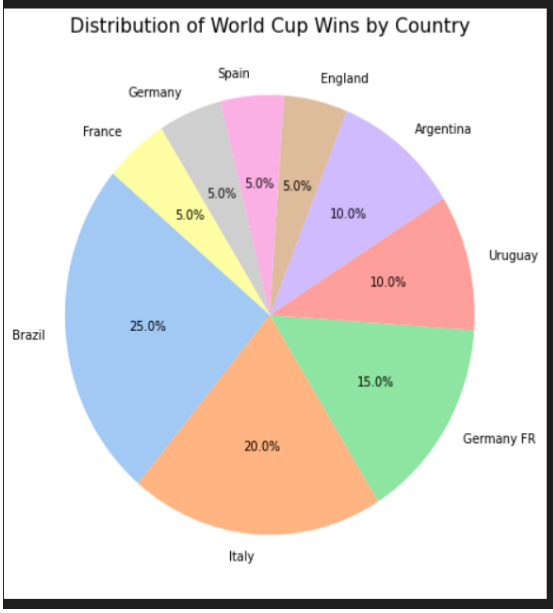

IBM HR
Dept vs job role vs count of gender vs maritial status
Department / Job Role
Human Resources
Research & Development
Sales
Marital S.. Gender Human Re.. Manager Healthcar.. Laborator.. Manager Manufact.. Research .. Research .. Manager Sales Exec.. Sales Repr..
Divorced Female 4 10 17 8 18 7 23 2 23 5
Male 10 2 24 38 6 18 16 39 5 46 6
Married Female 8 3 24 38 12 32 16 45 13 67 14
Male 20 3 37 78 17 35 22 77 8 84 20
Single Female 4 1 17 30 4 22 10 46 4 42 19
Male 6 2 19 58 7 20 9 62 5 64 19
Job role vs Avg monthly salary
Job Role
Avg. Monthly Income
7,529
131
4,236
52
3,237
259
17,182
102
7,295
145
16,034
80
3,240
292
6,924
326
2,626
83
Healthcar..
Human R..
Laborato..
Manager
Manufact..
Research ..
Research ..
Sales Exe..
Sales Rep..
Education_Field vs dept, Role, Marital status and Overtime details
Education Field
Departm.. Job Role Marital S.. Over Time Human Resour.. Life Sciences Marketing Medical Other Technical Degr..
Human Resources Human Resources Divorced No 4,772 4,770 4,936 4,071
Yes 16,635 5,936 5,021 2,696
Married No 47,191 24,734 34,695 2,742
Yes 12,090 2,148 2,991
Single No 3,886 12,947 4,286 7,988 6,887
Yes 8,837
Dept Job Role Sum of employ
Human Resources
Sales
Research & Development
Education Vs standard working hours
132
Technical
Degree
82
Other
27
Human
Resources
606
Life Sciences
464
Medical
159
Marketing
Dept and role vs Avg year at company
Department / Job Role
Human Resources
Research & Development
Sales
Avg. Years At Company
3.135
5.455
4.863
3.201
6.426
4.972
6.288
3.274
6.784
4.856
2.012
Human Re..
Manager
Healthcar..
Laborator..
Manager
Manufact..
Research ..
Research ..
Manager
Sales Exe..
Sales Rep..

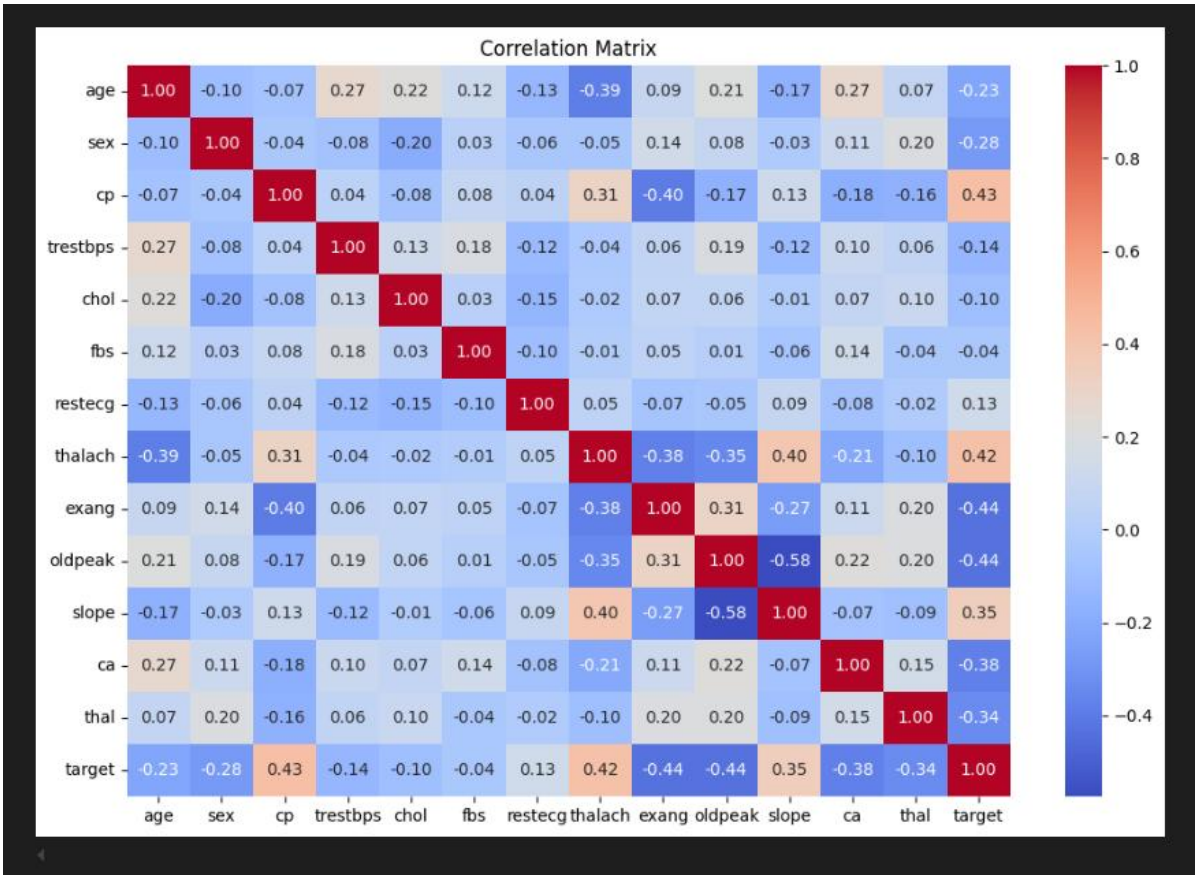
➢ Python Prediction Output:

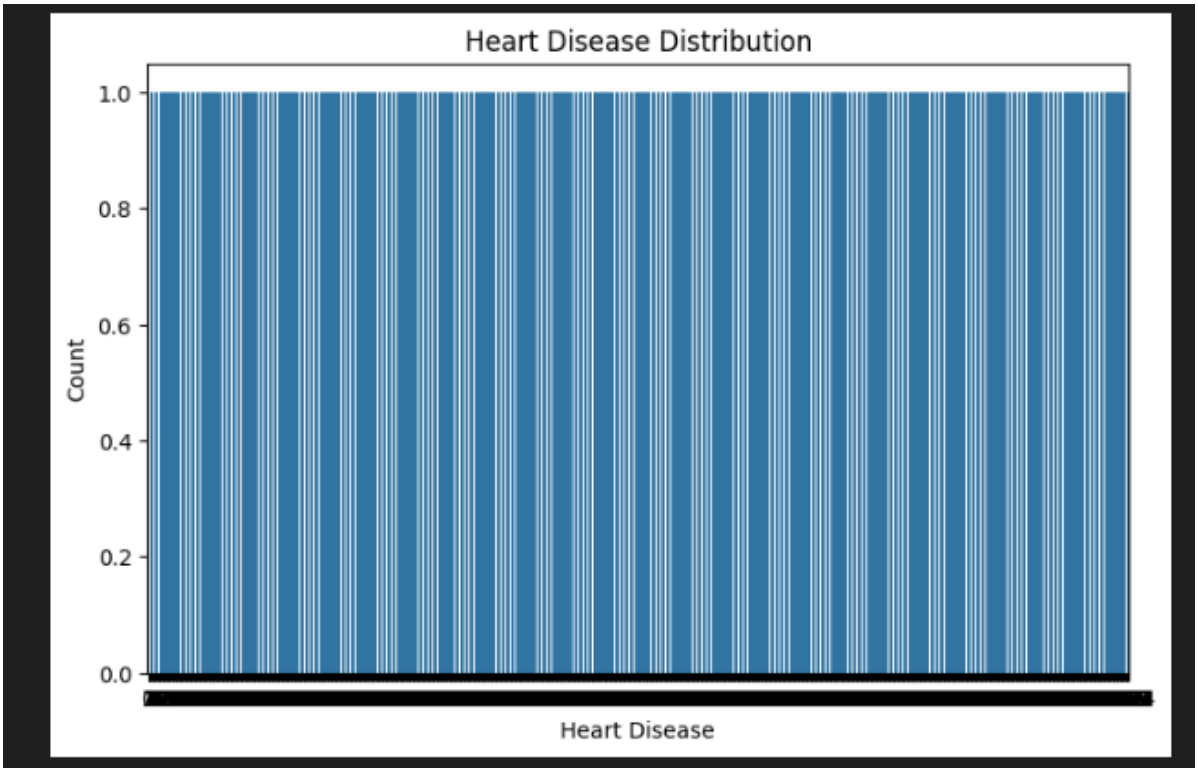
```
import matplotlib.pyplot as plt

# Group data by Crop_Year to see production trends over time
production_by_year = data.groupby('Crop_Year')['Production'].sum().reset_index()

# Plot
plt.figure(figsize=(10, 6))
plt.plot(production_by_year['Crop_Year'], production_by_year['Production'], marker='o')
plt.title('Total Crop Production Over the Years')
plt.xlabel('Year')
plt.ylabel('Total Production')
plt.grid()
plt.show()
```

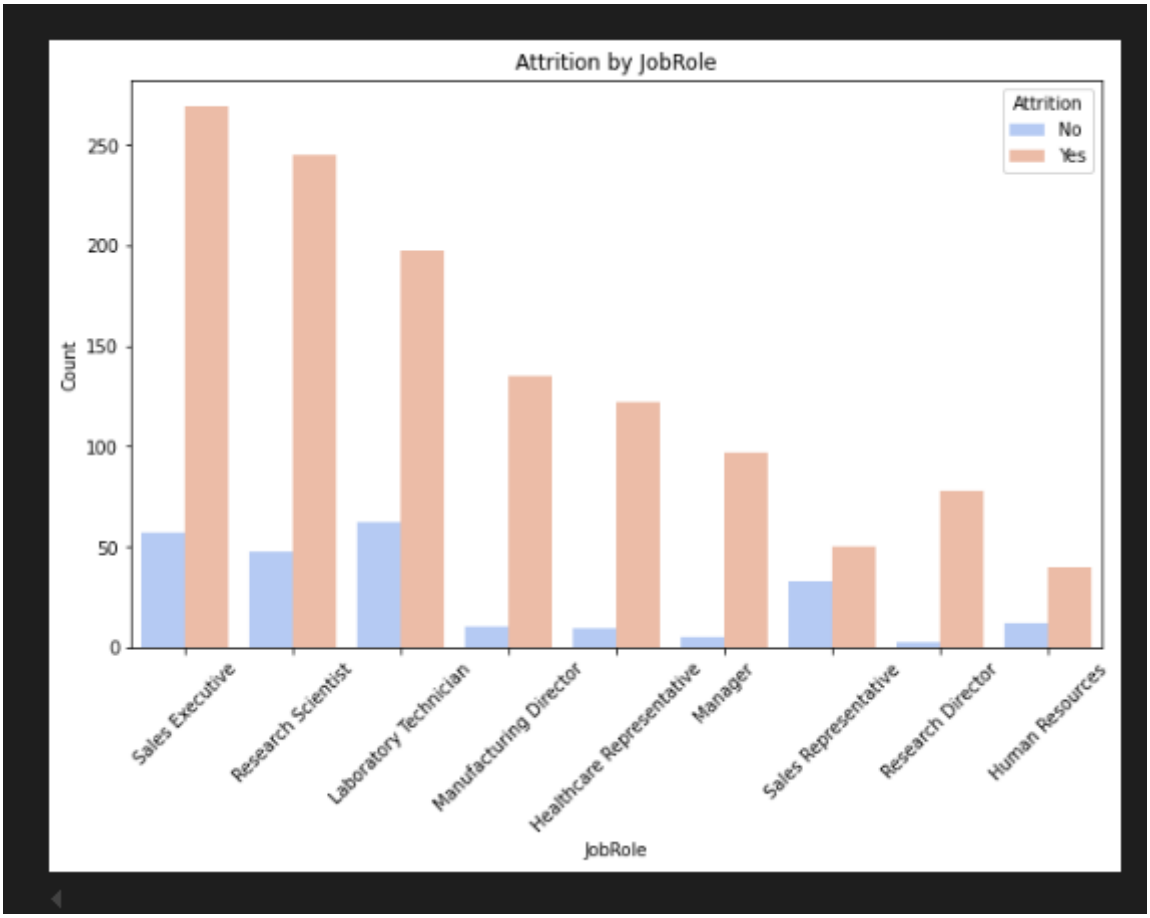



```
# 1. Pie Chart: Distribution of World Cup Wins by Country
plt.figure(figsize=(8, 8), facecolor='white')  # Set background color to white
cups_data['Winner'].value_counts().plot(kind='pie', autopct='%1.1f%%', startangle=140, colors=sns.color_palette('pastel'))
plt.title("Distribution of World Cup Wins by Country", fontsize=16)
plt.ylabel('')
plt.show()
```




Correlation Matrix
age 1.00 -0.10 -0.07 0.27 0.22 0.12 -0.13 -0.39 0.09 0.21 -0.17 0.27 0.07 -0.23
sex -0.10 1.00 -0.04 -0.08 -0.20 0.03 -0.06 -0.05 0.14 0.08 -0.03 0.11 0.20 -0.28
cp -0.07 -0.04 1.00 0.04 -0.08 0.08 0.04 0.31 -0.40 -0.17 0.13 -0.18 -0.16 0.43
trestbps 0.27 -0.08 0.04 1.00 0.13 0.18 -0.12 -0.04 0.06 0.19 -0.12 0.10 0.06 -0.14
chol 0.22 -0.20 -0.08 0.13 1.00 0.03 -0.15 -0.02 0.07 0.06 -0.01 0.07 0.10 -0.10
fbs 0.12 0.03 0.08 0.18 0.03 1.00 -0.10 -0.01 0.05 0.01 -0.06 0.14 -0.04 -0.04
restecg -0.13 -0.06 0.04 -0.12 -0.15 -0.10 1.00 0.05 -0.07 -0.05 0.09 -0.08 -0.02 0.13
thalach -0.39 -0.05 0.31 -0.04 -0.02 -0.01 0.05 1.00 -0.38 -0.35 0.40 -0.21 -0.10 0.42
exang 0.09 0.14 -0.40 0.06 0.07 0.05 -0.07 -0.38 1.00 0.31 -0.27 0.11 0.20 -0.44
oldpeak 0.21 0.08 -0.17 0.19 0.06 0.01 -0.05 -0.35 0.31 1.00 -0.58 0.22 0.20 -0.44
slope -0.17 -0.03 0.13 -0.12 -0.01 -0.06 0.09 0.40 -0.27 -0.58 1.00 -0.07 -0.09 0.35
ca 0.27 0.11 -0.18 0.10 0.07 0.14 -0.08 -0.21 0.11 0.22 -0.07 1.00 0.15 -0.38
thal 0.07 0.20 -0.16 0.06 0.10 -0.04 -0.02 -0.10 0.20 0.20 -0.09 0.15 1.00 -0.34
target -0.23 -0.28 0.43 -0.14 -0.10 -0.04 0.13 0.42 -0.44 -0.44 0.35 -0.38 -0.34 1.00
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Heart Disease Distribution
Count
1.0
0.8
0.6
0.4
0.2
0.0
Heart Disease


Attrition by JobRole
Attrition
No
Yes
Count
250
200
150
100
50
0
Sales Executive
Research Scientist
Laboratory Technician
Manufacturing Director
Healthcare Representative
Manager
Sales Representative
Research Director
Human Resources
JobRole


Total Goals Scored by Hosting Country
GoalsScored
250
200
150
100

CONCLUSION

During the internship, multiple projects were undertaken across diverse domains, showcasing the power of data analytics, visualization, and predictive modeling in addressing real-world challenges. Each project contributed to enhancing domain-specific insights and provided practical solutions for stakeholders. Below is a consolidated conclusion for all projects:

Crop Production Analysis

This project offered a deep understanding of agricultural trends using historical data. By developing predictive models and dashboards, key factors such as climate, soil type, and crop variety influencing crop yields were identified. These insights enable policymakers and stakeholders to implement data-driven strategies, ultimately boosting agricultural productivity and sustainability.

FIFA World Cup Data Analysis

The analysis of FIFA World Cup data successfully uncovered historical patterns, team and player performance metrics, and tournament trends. By transforming raw sports data into interactive Tableau dashboards, the project provided sports analysts and enthusiasts with an intuitive way to explore data. This demonstrated the utility of analytics in enhancing decision-making in the sports industry.

Heart Disease Diagnostic Analysis

This project highlighted the critical role of machine learning and visualization in healthcare. Clinical data was analyzed to build predictive models identifying key risk factors for heart disease. Tableau dashboards provided actionable insights to healthcare professionals, enabling better diagnostic decision-making. The project underscored the significance of integrating analytics into healthcare to improve patient outcomes.

IBM HR Analytics: Employee Attrition and Performance

The analysis of IBM HR data addressed workforce challenges by identifying factors influencing employee attrition and performance. Advanced dashboards provided HR teams with a comprehensive understanding of trends, enabling them to design effective strategies to improve employee retention and satisfaction. This project demonstrated the application of data-driven solutions in optimizing workforce management.

INTERNSHIP CERTIFICATE

GUIDE EVALUATION COMMENTS