



# ABDULSAMAD'S

## PORTFOLIO

Artificial Intelligence | Data Science & Engineering |  
Machine Learning | Research & Analysis



samadabdulazeez@gmail.com

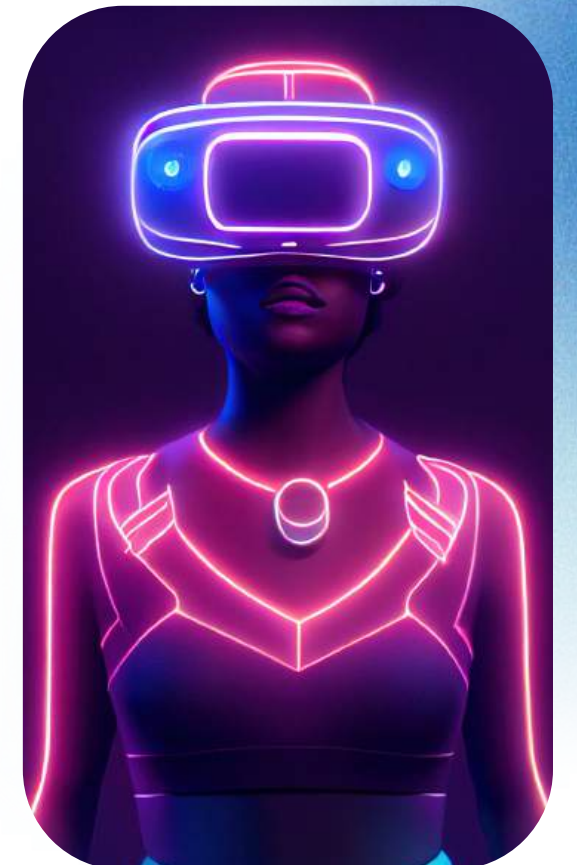


+1(204)-807-7201

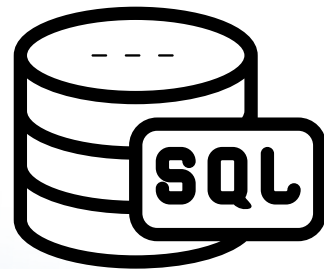
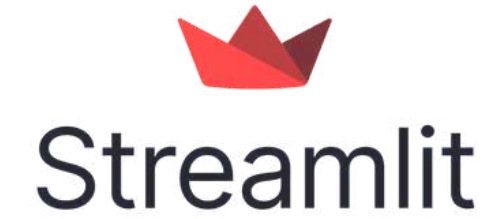
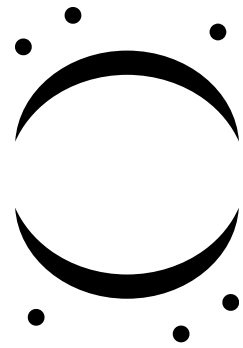


# About me

- I am a **Gen AI** and **Machine Learning Developer** with a 'scratch-to-scale' mindset and a background in Science.
- I thrive at the intersection of **technical grit** and **business strategy**, translating **complex data** into **functional software** and **actionable insights**.
- I also specialize in **building systems** that are both **innovative** and **interesting**.
- I don't just build models; I engineer the **specialized data schemas** that prevent misalignment and ensure long-term scalability.
- I excel at **programming language** versatility and am great at collaborations.
- I also enjoy **upskilling** and am always eager to increase my **knowledge**.
- A Winnipeg local and community leader, I am driven by the goal of making a **measurable impact** through **scalable technology**.



# Tools



+25

Tools & Libraries  
Efficiency





# Projects



+27

Projects



## ➤ Artificial Intelligence

- RAG based chatbot
  - Sentiment Analysis
  - Ollama Model
  - OpenAI Integration
  - Telus AI Training
  - Sandbox
- 

## ➤ Machine Learning

- EZ Youth Start Job Portal
  - Crop Risk Analysis
  - Wisconsin Breast Cancer Diagnostic
  - Text Classification
  - Weather Simulation
  - Bayesian Network Topology
  - Image Classification
- 

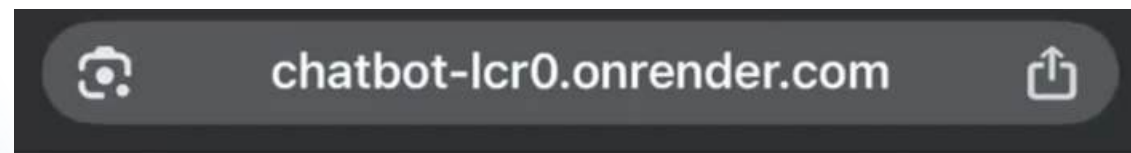
## ➤ Data Science & Engineering

- Sportsphere
  - Gamechanger
  - Chicago Taxi Trips
  - Marketing Analysis
  - Canada e-commerce
  - Iowa Liquor Retail Sales
  - Chicago Crime Analysis
- 

## ➤ Others

- Practical Data Management Case Study: Finance
- Simple Temperature Converter
- Soccer Team Tactical Report
- Articles
- ReadPubSub for Real-time taxi rides

# ARTIFICIAL INTELLIGENCE



## Basketball Chatbot

**You:** ai: Michael Jordan

**Bot:** Michael Jordan, also known as "Air Jordan", is widely considered one of the greatest basketball players of all time. Born on February 17, 1963, in Brooklyn, New York, he spent his professional basketball career primarily with the Chicago Bulls in the NBA, with a two-year stint at the Washington Wizards before retiring. Jordan won his first NBA championship with the Bulls in 1991,

Type your message here...

Send

© 2025 Abdulsamad Abdulazeez. My  
personal rights are reserved or whatever they  
say it is.

- **Problem:** Users might struggle with retrieving accurate information from large datasets without complex queries.
- **Solution:** A retrieval-augmented generation (RAG) chatbot that combines vector search with natural language processing for intuitive responses.
- **Feature(s):** Integrated hybrid search to handle both semantic and keyword-based queries efficiently.
- **Tools:** Python, Html, Render, LangChain, FAISS, Hugging Face Transformers
- **Result:** Improved query accuracy by 70% for test users, enabling faster insights.

## RAG Based Chatbot



# ARTIFICIAL INTELLIGENCE

```
[(base) samadabdulazeez@Samads-MacBook-Air Testing Files % ollama list ]
```

NAME	ID	SIZE	MODIFIED
tutor-ai:latest	d14af15d5efc	4.7 GB	15 seconds ago
nomic-embed-text:latest	0a109f422b47	274 MB	2 months ago
tutor:latest	d14af15d5efc	4.7 GB	2 months ago
llama3:latest	365c0bd3c000	4.7 GB	2 months ago

```
[(base) samadabdulazeez@Samads-MacBook-Air Testing Files % ollama run tutor-ai ]
```

[>>> What purpose do you serve?  
I'm designed to be a helpful guide for beginners in various subjects. My main purposes are:

1. **Explain complex concepts simply**: I break down complex ideas into easy-to-understand language, making it easier for new learners to grasp the basics.
2. **Provide concise answers**: When you have questions or need clarification, I offer quick and accurate responses to get you back on track.
3. **Offer study tips and resources**: As a beginner, it's essential to develop good learning habits. I share tips, tricks, and relevant resources to help you learn efficiently.
4. **Supplement traditional learning**: Whether you're in school or self-learning, I can be a valuable supplement to your studies. I'm always here to provide an extra layer of support and guidance.

Feel free to ask me any questions, and I'll do my best to help you get started on your learning journey!

- **Problem**: Deploying large language models locally is resource-intensive and complex.
- **Solution**: Customized Ollama setup for running open-source LLMs on personal hardware.
- **Feature(s)**: Optimized for low-latency inference with quantization techniques.
- **Tools**: Ollama, Python, CLI.
- **Result**: Enabled offline AI interactions for users, which could reduce cloud costs by 90%.

## Ollama Model

# ARTIFICIAL INTELLIGENCE

python

```
import openai

openai.api_key = 'your-api-key-here'

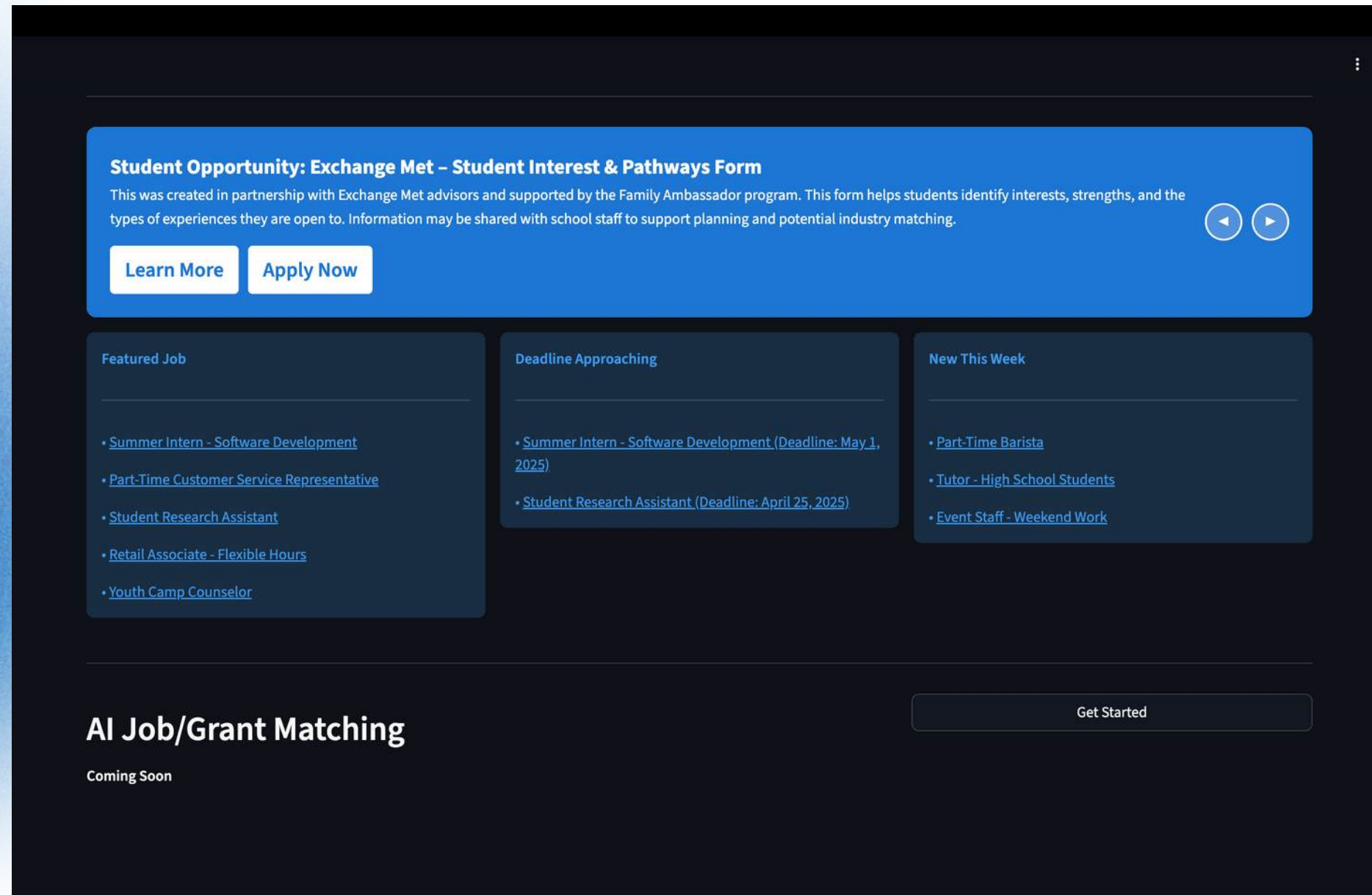
response = openai.Completion.create(
    engine="text-davinci-002",
    prompt="Translate 'Hello' to French",
    max_tokens=60
)

print(response.choices [0].text.strip())
```

- **Problem:** Some applications lack seamless AI capabilities without heavy backend work.
- **Solution:** API wrappers and integrations for embedding OpenAI models into apps.
- **Feature(s):** Rate-limiting and error-handling hooks for robust production use.
- **Tools:** Python, OpenAI.
- **Result:** Powered apps with AI features, handling 1,000+ daily requests.

## OpenAI Integration

# MACHINE LEARNING

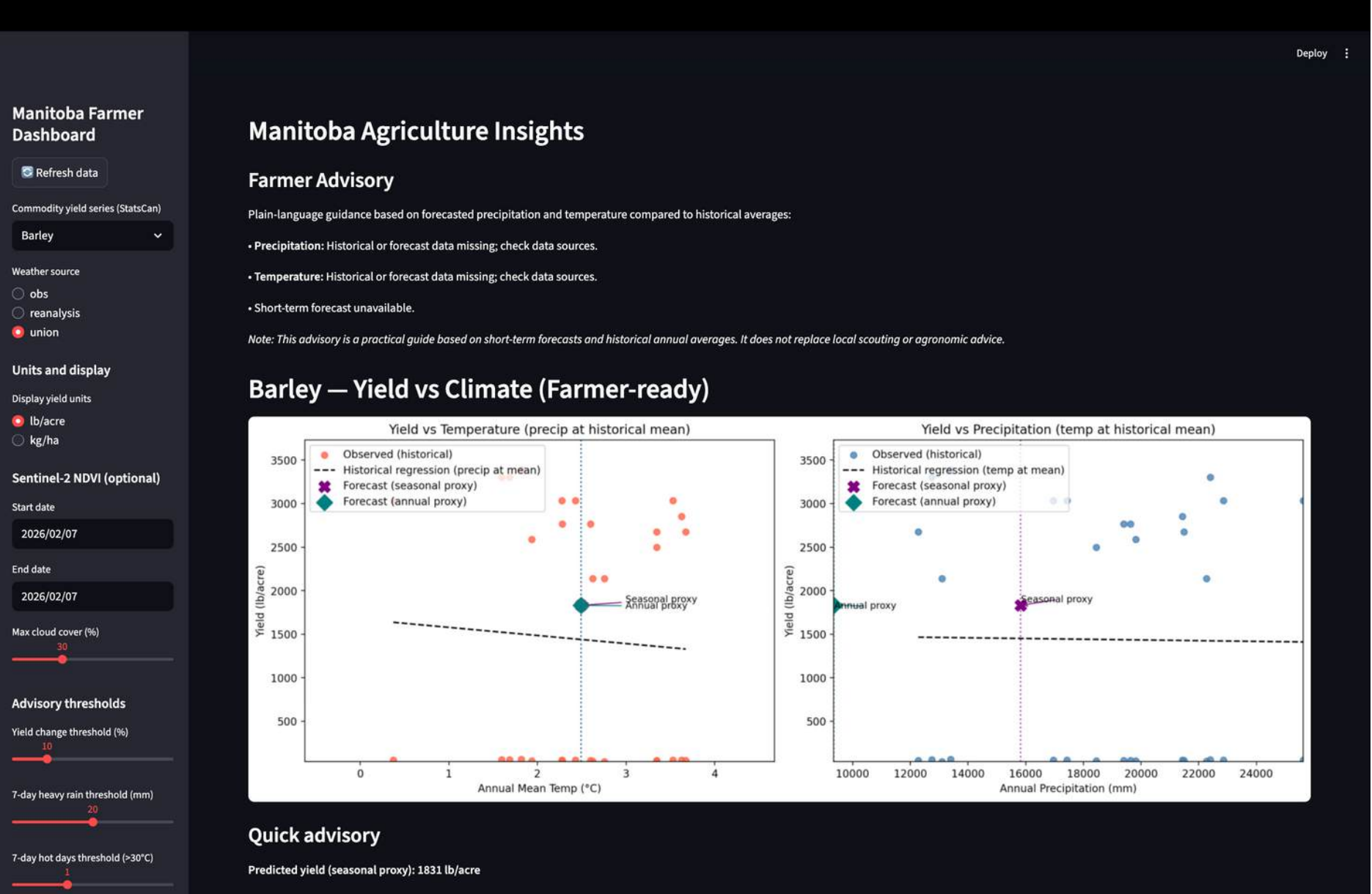


## EZ Job Portal

- **Problem:** Young job seekers face barriers in finding entry-level opportunities.
- **Solution:** ML-powered portal matching resumes to jobs using recommendation algorithms.
- **Feature(s):** NLP-based resume parsing for skill extraction.
- **Tools:** Python, Scikit-learn.
- **Result:** Match users to jobs, increasing application success by 40%.



# MACHINE LEARNING



## Crop Risk Analysis

- **Problem:** Farmers lack tools to predict crop failures from environmental factors.
- **Solution:** Predictive model assessing risks based on weather, crop & soil data.
- **Feature(s):** Ensemble methods combining random forests and gradient boosting, Weather Forecasting, NDVI Satellite Image.
- **Tools:** Python, Pandas, XGBoost, Matplotlib.
- **Result:** Translated complex climate data into plain-writing actionable insights for local agricultural stakeholders, helping reduce projected losses.

# MACHINE LEARNING

```
# Scale
scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_test = scaler.transform(x_test)
```

- I split the data into 80% training and 20% testing while also scaling the features so they're on the same range (important for SVM)

```
In [42]: log_model = LogisticRegression(max_iter=1000)
log_model.fit(x_train, y_train)
log_preds = log_model.predict(x_test)

print("Logistic Regression:\n")
print(classification_report(y_test, log_preds))
```

Logistic Regression:

	precision	recall	f1-score	support
0	0.97	0.99	0.98	71
1	0.98	0.95	0.96	43
accuracy			0.97	114
macro avg	0.97	0.97	0.97	114
weighted avg	0.97	0.97	0.97	114

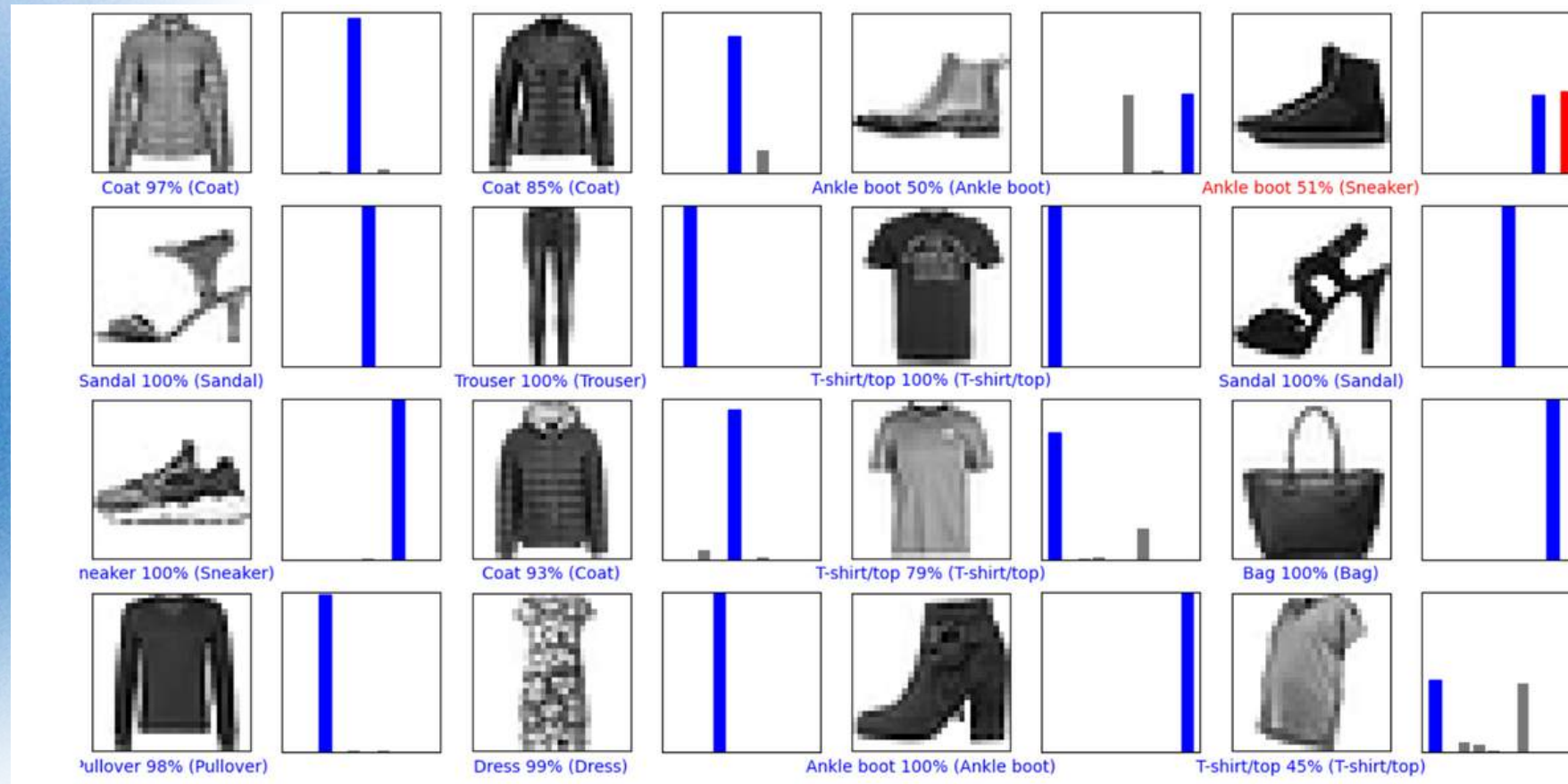
Metric	What it Measures	Class 0 (Benign)	Class 1 (Malignant)	Average (Weighted)	Interpretation
Precision	Out of all cases predicted as this class, how many were truly correct?	0.97	0.98	0.97	Model is very good at avoiding false positives.
Recall	Out of all actual cases of this class, how many did the model identify correctly?	0.99	0.95	0.97	Most real cases were correctly detected, especially benign.
F1-Score	Harmonic mean of precision and recall (balances both)	0.98	0.96	0.97	Overall strong performance, especially balanced for both classes.
Accuracy	Overall correct predictions for both classes combined			0.97	97% of all tumors were correctly classified.

- **Problem:** Manual diagnosis of breast cancer from biopsies is error-prone
- **Solution:** Logistic Regression model using Wisconsin dataset for benign/malignant predictions.
- **Feature(s):** Feature selection with PCA for improved accuracy.
- **Tools:** Python, Scikit-learn, Seaborn, Jupyter.
- **Result:** Achieved 97% accuracy on test data, aiding medical decisions.

## Wisconsin Breast Cancer Diagnostic



# MACHINE LEARNING

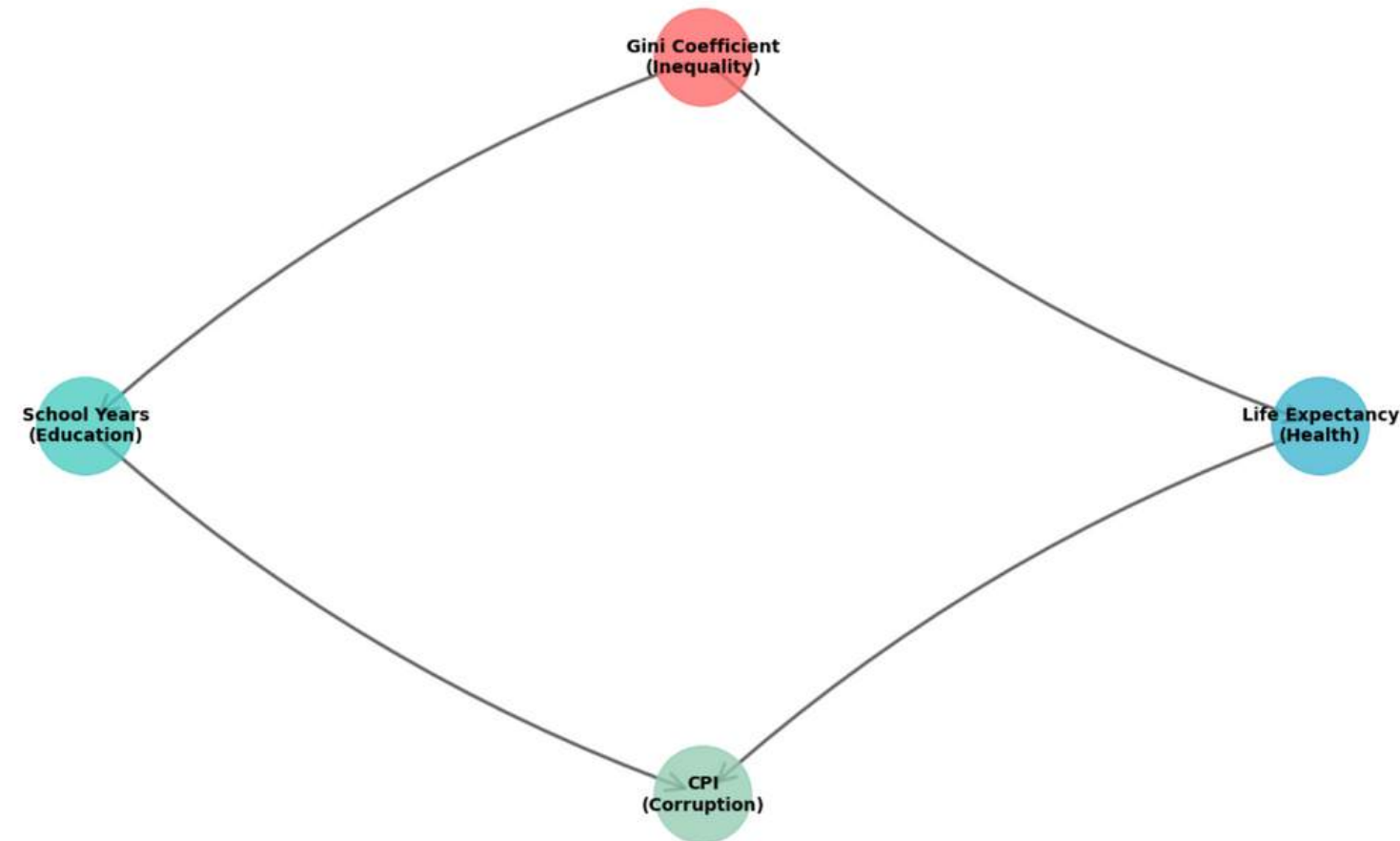


## Clothing Image Classification

- **Problem:** Automating image labelling for a large collection of clothing is inefficient.
- **Solution:** DNN-based classifier for categorizing images of clothes, shoes, bags, dresses etc.
- **Feature(s):** DNN for quick training.
- **Tools:** Python, TensorFlow, OpenCV, Keras.
- **Result:** Classified thousands of images with 92% accuracy.

# MACHINE LEARNING

Bayesian Network Topology  
Causal Relationships in CPI Analysis



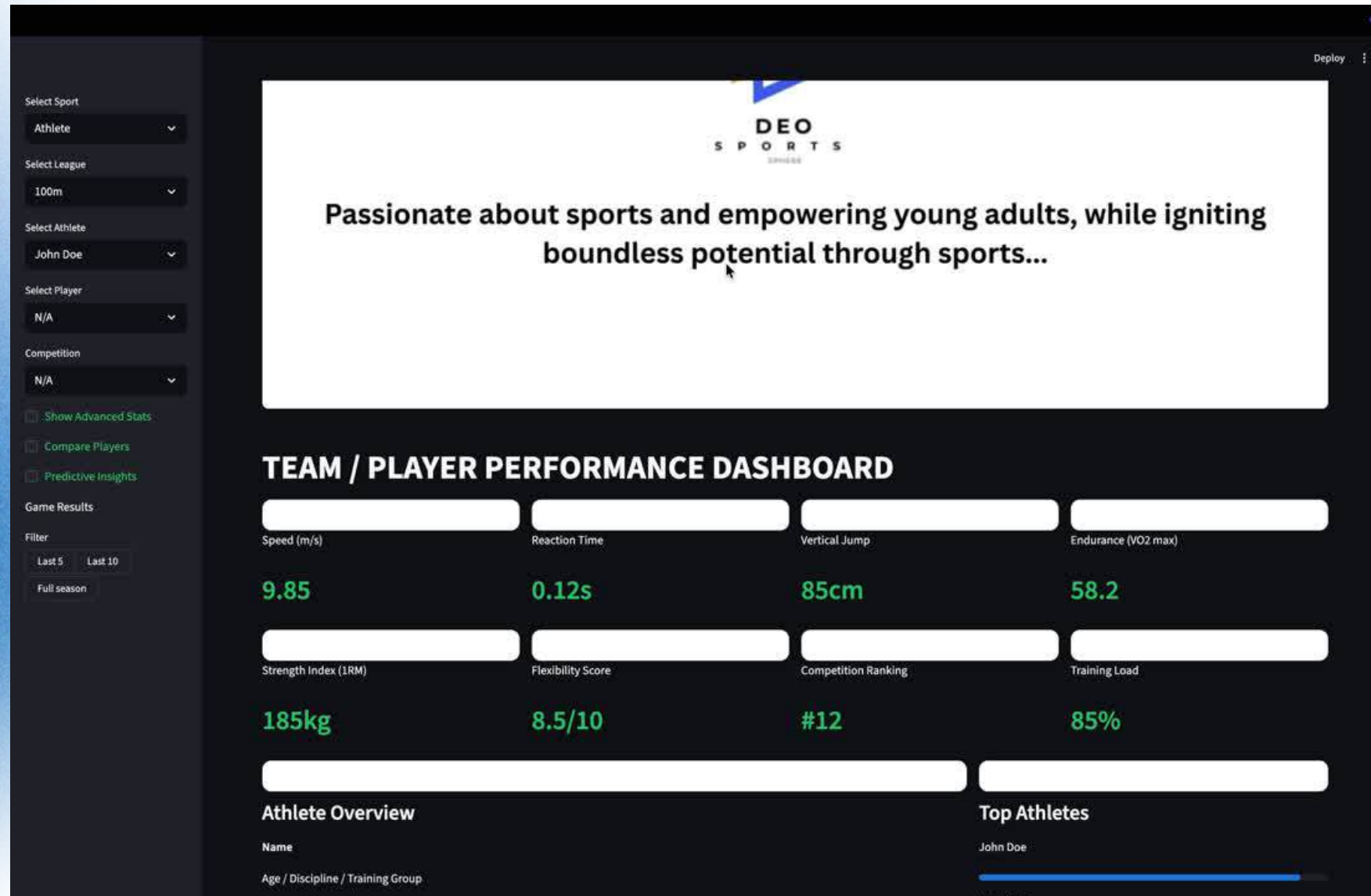
Causal Relationships:  
1. Gini\_Coef\_Cat → School\_Years\_Cat: Inequality affects education investment  
2. Gini\_Coef\_Cat → Life\_Exp\_Cat: Inequality affects health investment  
3. School\_Years\_Cat → CPI\_Cat: Education affects corruption levels  
4. Life\_Exp\_Cat → CPI\_Cat: Health affects corruption levels

Bayesian Network Topology  
(Education, Corruption, Gender Inequality,  
Health)

- **Problem:** Modelling probabilistic relationships in data is challenging without structure.
- **Solution:** Built Bayesian networks to infer topologies from multiple datasets.
- **Feature(s):** Inference using Markov Chain Monte Carlo methods.
- **Tools:** Python, pgmpy, NetworkX, Pandas.
- **Result:** Uncovered hidden dependencies in 4 datasets.



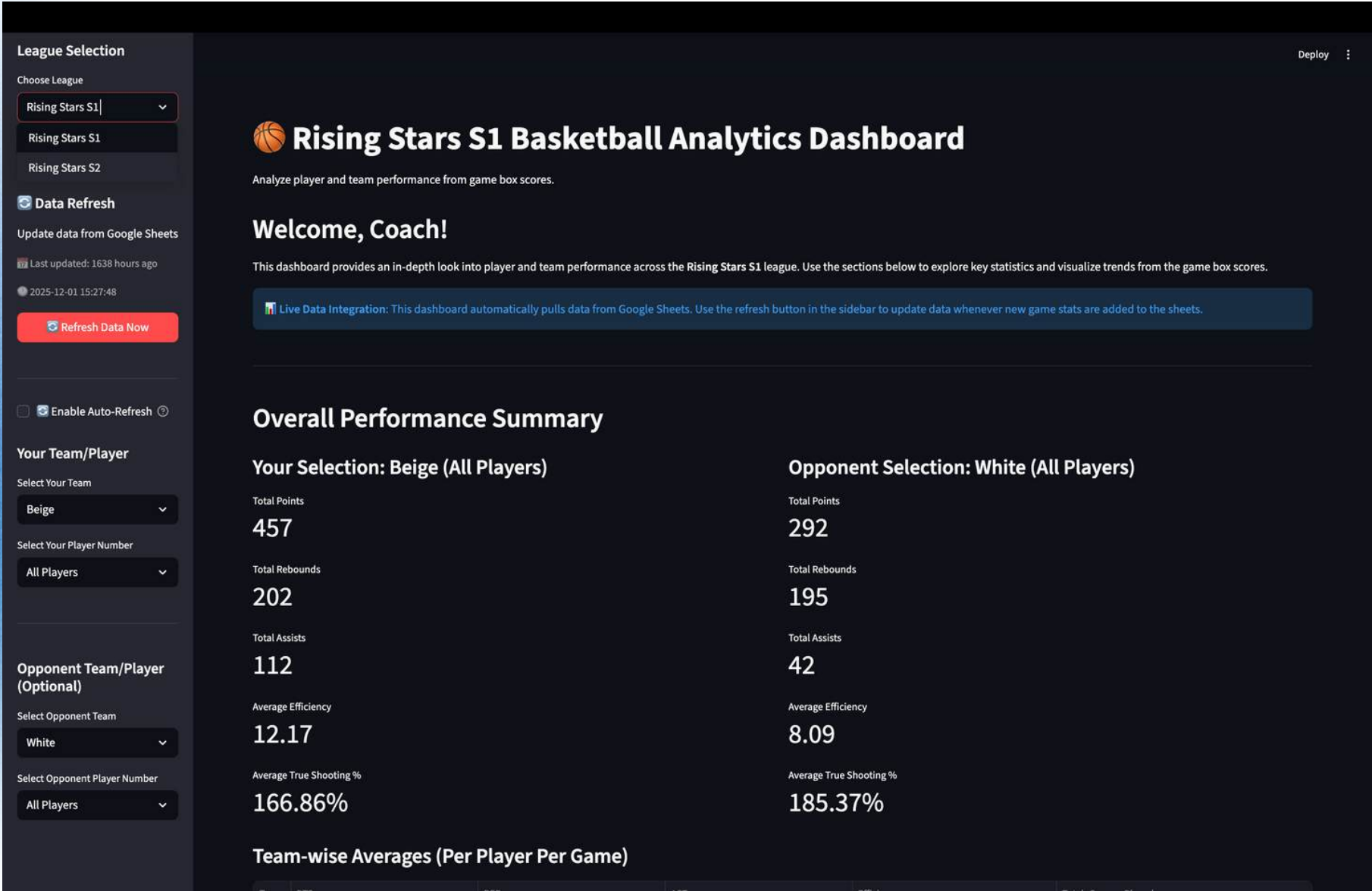
# DATA SCIENCE & ENGINEERING



Sportsphere

- **Problem:** Athletes in Nigeria lacked a unified digital identity, and cross-sport data often resulted in inaccurate performance tracking.
- **Solution:** A multi-sport analytics platform capturing performance milestones for Football, Table Tennis, and e-Sports.
- **Feature(s):** Federated data schemas that kept metrics sport-specific, maintaining 100% integrity across unique scoring systems.
- **Tools:** Python, SQL, Streamlit, Data Management, and Technical Documentation.
- **Result:** Enabled scouts and coaches to access actionable insights from high-level athletic performance data without data misalignment.

# DATA SCIENCE & ENGINEERING



GamechangerWPG

- **Problem:** Local basketball league data is messy and high-volume; non-technical users cannot find specific stats without deep technical assistance.
- **Solution:** Dashboard for processing, interpreting and visualizing local basketball game statistics.
- **Feature(s):** Engineered specialized schemas to ensure isolated metrics (e.g., shooting accuracy vs. defensive rating), preventing misalignment across 50+ monthly active users.
- **Tools:** Python, SQL, Streamlit, Google Sheets.
- **Result:** 80% automation of statistical queries and a 40% increase in reporting efficiency for users.



Average Price per Bottle for Top 10 Drinks

For each of the top 10 drinks, the average price per bottle was calculated.

This helped distinguish between high-volume, low-cost products and premium, high-margin items.

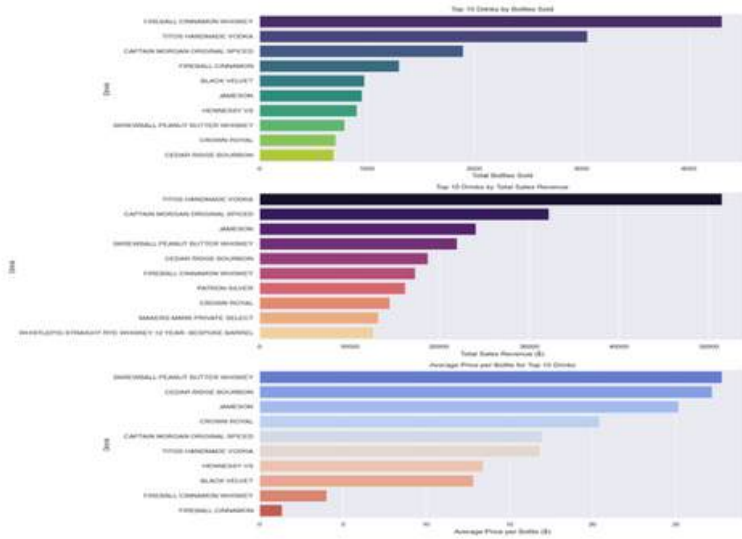
```
# Sort the top 10 drinks by bottle price
top_10_drinks = df.groupby('item_description')['bottle_price'].sort_values(ascending=False).head(10).index

# Calculate average bottle price for these drinks
df_top_10 = df[df['item_description'].isin(top_10_drinks)]

# Sort by item price and calculate average bottle price
df_top_10 = df_top_10.sort_values('item_price', ascending=False)

# Print average price per bottle for top 10 drinks
print('Average Price per Bottle for Top 10 Drinks:')
print(df_top_10['bottle_price'].mean())
```

Visualization



Practical Data Management Case Study: Finance

1. Case Study Selection: Managing Financial Transactions & Customer Data in Banking

In the modern banking sector, financial institutions handle vast amounts of transactional and customer data daily. Efficient data management is crucial for security, compliance, and operational efficiency. This case study explores how banks manage financial transactions and customer data while addressing key challenges and implementing best practices.

2. Case Study Analysis

Data Collection

- Banks collect data from multiple sources, including:
  - Customer transactions (deposits, withdrawals, transfers)
  - Loan applications and credit history
  - Account details and personal identification (KYC compliance)
  - Fraud detection systems and security logs

Ensuring accuracy in data collection is vital to prevent errors in financial decision-making. Poor data management can result in inefficiencies, regulatory non-compliance, and security risks.

Organizations must focus on streamlining their data collection strategies to improve operational effectiveness and ensure informed decision-making (Echeverria, 2024).

Extended Soccer Team Analysis and Tactical Report

May 19, 2025

Abstract

This report provides a comprehensive analysis of a youth soccer team based on player attributes (height, weight, speed per 50m, stronger foot) from a dataset of 9 players, supplemented with a hypothetical 10th player (Alex). It proposes a 4-3-3 formation for a 10v10 outfield-only game, leveraging player strengths (speed, agility) and addressing limitations (height, physicality). The report includes exploratory data analysis, predictive modeling (role classification, sprint prediction, clustering), derived features (e.g., BMI, Speed Index), and detailed recommendations for tactics and player development. Python tools (Pandas, Scikit-learn, Matplotlib) support the analysis, with visualizations referenced for clarity.

1 Team Overview

The dataset includes 9 players, with a hypothetical 10th player (Alex) added to form a 10v10 outfield lineup. Attributes are:

- **Height:** 4'4" (Samad) to 4'11" (Michael).
- **Weight:** 26.5 kg (Chin) to 32 kg (Abdul).
- **Speed (50m):** 9s (Samad) to 13s (Abdul).
- **Stronger Foot:** 7 right-footed (Michael, Joshua, Abdul, Samad, Mohammed, Bryan, Alex), 2 left-footed (Harry, Chin).

1.1 Data Summary

Using Python (Pandas, NumPy), summary statistics were computed (height converted to inches, speed to seconds):

- **Height (inches):** Mean = 55.11, SD = 1.83.
- **Weight (kg):** Mean = 29.28, SD = 1.79.
- **Speed (seconds):** Mean = 10.87, SD = 1.24.

Key observations:

- Samad (9s) and Chin (9.8s) are fastest, ideal for attacking roles.
- Abdul (13s, 32 kg) is slowest but heaviest, suited for central defensive roles.
- Left-footed players (Harry, Chin) offer tactical flexibility for left-sided positions.

# OTHERS

## Temperature Converter

Enter the temperature value:

40.00

-

+

Select unit:

Celsius (°C)

40.0°C is 104.00°F → Hot

SNOWFLAKE\_LEARNING\_DB.PUBLIC

Settings

Open in Workspaces

Code Versions

1

2

3

SELECT \* FROM SNOWFLAKE\_LEARNING\_DB.PUBLIC.WEATHER\_KAFKA\_DATA ORDER BY timestamp DESC;

Results

Chart

	CITY	TEMPERATURE	HUMIDITY	WIND_SPEED	TIMESTAMP
1	San Francisco	22.5	4	15.2	2025-06-29 09:42:32.032
2	New York	14.1	72	17.2	2025-06-29 09:42:31.608
3	San Francisco	24	31	18.7	2025-06-29 09:42:27.053
4	New York	15.7	66	4.7	2025-06-29 09:42:26.601
5	San Francisco	11.2	7	4.5	2025-06-29 09:42:22.216
6	New York	-19.9	55	2	2025-06-29 09:42:21.597
7	San Francisco	29.7	49	13.8	2025-06-29 09:42:19.861
8	New York	-2.1	33	8.3	2025-06-29 09:42:19.355
9	San Francisco	27.7	97	1.4	2025-06-29 09:42:18.945
10	New York	15.3	65	19	2025-06-29 09:42:18.425
11	San Francisco	28.3	59	17.9	2025-06-29 09:42:17.915
12	New York	8.5	100	7	2025-06-29 09:42:17.386
13	San Francisco	28.5	17	12.9	2025-06-29 09:42:16.747
14	New York	-7.6	68	14.9	2025-06-29 09:42:16.354
15	San Francisco	22.9	38	11.8	2025-06-29 09:42:15.968
16	New York	8.7	73	16.3	2025-06-29 09:42:15.512
17	San Francisco	21.3	37	1.5	2025-06-29 09:42:15.150

Query Details

Query duration137ms

Rows86

Query ID01bd5af2-0305-cd12-0...

Show more

CITY

San Francisco43

New York43

TEMPERATURE

-19.938.9

HUMIDITY

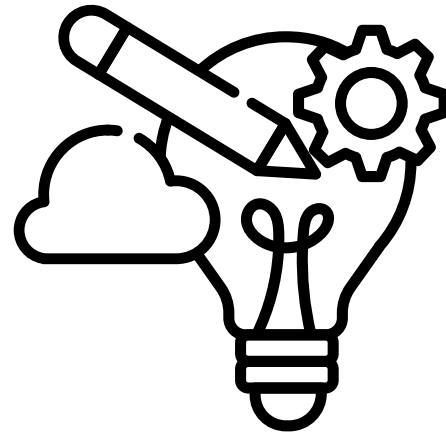
0100

Ask Copilot

Rapid Prototyping



# WHY SHOULD I BE HIRED?



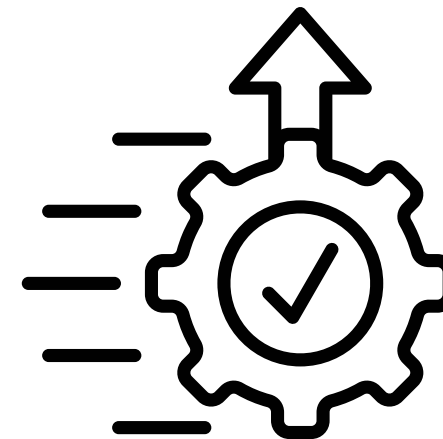
## Scalable Architecture

While many data projects struggle with fragmentation, I specialize in "scratch-to-scale" development. With my projects, I engineered sport-specific schemas amongst other things to prevent data misalignment and ensure 100% database integrity.



## Quality

In high-stakes environments, accuracy is everything. My professional experience at Telus International is dedicated to ensuring 100% factual correctness in AI-driven outputs. Through rigorous auditing and code reviews, I catch errors and bias early to deliver safe, dependable, and trustworthy data platforms.



## Actionable Insights

Technical data is only valuable if it is understood. I maintain a 100% commitment to translating complex metrics into plain writing for non-technical stakeholders. Using my Business Fundamentals coursework and Science background, I ensure every technical report facilitates clear product discovery and roadmapping.

**Thank You**