

Henry Dibie

Data scientist skilled in analysis, ML, data engineering, NLP, and workflow automation. Experienced in building AI tools, credit risk models, and scalable data pipelines using Python, R, SQL, and Prefect to drive business outcomes.

Stears, Data Research Analyst | (Automation & QA)

2024–2025

- Sourced, analyzed, and structured African market data using *Airtable* to drive insights for global investors.
 - Conducted Quality Assurance (QA) on priority company profile entries submitted by junior data research analysts before publication on the Stears website, ensuring accuracy and consistency.
 - Automated recurring market data extraction across 54 African countries using *Prefect* workflows, enhancing data reliability and reducing manual labor.
 - Built a generative AI pipeline for processing and summarizing sales call transcripts using *Whisper*, *OpenAI* embeddings, and *FAISS* vector similarity, improving sales insights automation.
 - Engineered a Python-based transaction monitor for African private equity deals, integrating *SQLAlchemy*, *webhooks*, and *Slack* bots for real-time alerts.
 - Developed a custom GPT-based private equity chatbot using *OpenAI*, *FAISS*, and prompt engineering, enabling natural language queries over investor reports and deal data.
 - Built NLP and document-processing pipelines that convert unstructured documents into validated, structured outputs, with emphasis on accuracy guarantees and automated QA.
 - Built and automated compliant data pipelines using *dbt*, *Python*, *PySpark*, and API integrations with version-controlled workflows, establishing a reliable data infrastructure key to scalable MLOps.
 - Led cross-functional collaboration across research, data & engineering teams & worked closely with external partners to scale & streamline the data collection process.
 - Created end-to-end NLP systems to extract macroeconomic indicators (e.g., inflation rates, CPI, FX rates) from government websites and convert them into structured datasets for investor use.
 - Designed and managed critical *Postgres* tables supporting PE transaction intelligence and internal analytics dashboards.
- **Result:** Accelerated research cycles by 40%, improved fund profile accuracy, and enabled AI-powered investor analysis.

Renmoney Microfinance Bank, Credit Risk Analyst (Data Science)

2022-2024

- Assessed individual and SME loan applications using quantitative and qualitative risk analysis.
- Designed and deployed credit risk models (*Logistic Regression*, *XGBoost*) to

[LinkedIn](#) | [GitHub](#) | [Medium](#)

Availability: Open to Remote, Hybrid or Global Roles

henrymorgan273@yahoo.com

SKILLS

Technical Skills

- **Analytics & Reporting Tools:** *Airtable*, *Excel*, *Power BI*, *Flourish*, *Metabase*, *Google Slides*, *Matplotlib*, *Seaborn*, *Plotly*, and *Dash*.
- **Programming & Scripting Languages:** *Python*, *R*, *SQL*, *DAX*, *HTML*, *CSS*, *JavaScript*, *Markdown*.
- **Frameworks & APIs:** *SQLAlchemy*, *PySpark*, *Pandas*, *Numpy*, *FastAPI*, *Streamlit*, *Flask*, *LangChain*, *Uvicorn*, *REST APIs*, *Azure Functions* (*Python*), *Azurite*, *OpenAPI/Swagger*.
- **ML & DL Libraries:** *Scikit-Learn*, *NLTK*, *TensorFlow*, *XGBoost*, *LightGBM*, *PyTorch*, *Keras*, *OpenCV*, *SpaCy*, *Falcon*, *Mistral*, & *Prompt Engineering*.
- **Statistical Tools:** *Pandas*, *NumPy*, *MATLAB*, *SNR calculation*, *Drift analysis*, *Linearity & Hysteresis modeling*.
- **Business Analytics & Productivity Tools:** *Microsoft Office*, *Google Analytics*, *Mixpanel*, & *Segment*.
- **Operating Systems:** *Linux*, *Unix*, *Windows*
- **Databases & Warehousing:** *PostgreSQL*, *MySQL*, *Microsoft SQL Server*, *Snowflake* & *BigQuery*.
- **Project Management & Collaboration:** *Git*, *GitLab*, *GitHub*, *Asana*, *G Suite*, *Slack*, *Microsoft Teams*, *ClickUp*, and *Discord*.
- **Cloud Computing:** *GCP*, *AWS*, *Azure* (*Functions*, *Storage*, *Azurite*) & *Prefect Cloud*.
- **Workflow Orchestration & CI/CD:** *dbt*, *Prefect*, *GitHub Actions*, *GitLab CI/CD*.
- **Web Scraping:** *BeautifulSoup*, *Selenium*, *Playwright*, *Scrapy*, *requests*, *APIs*
- **Risk Decisioning Software:** *Provenir*

Soft Skills

- Teamwork & Collaboration
- Problem-solving & Analytical Thinking
- High Emotional and Social intelligence

- evaluate borrower default probability & enhance loan approval precision by 15%.
- Conducted holistic appraisals of credit requests for retail and MSME borrowers, including analysis of bank statements, cash flow, and collateral documentation.
 - Led A/B testing and model explainability sessions with underwriters and product managers, improving stakeholder trust in automated decision-making.
 - Supported portfolio monitoring and performance reporting to guide dynamic credit decision rules.
 - Built segmentation frameworks for risk-based customer profiling and policy refinement, contributing to a 25% boost in model performance.
 - Contributed to Probability of Default (PD) model development and exposure analysis to strengthen portfolio risk assessment.
 - Created interactive Power BI dashboards that visualized model outcomes, reducing time-to-decision for business leaders by 20%.

Nupat Technologies, Data Scientist (Outsourced)

2020-2022

- Developed predictive models and anomaly detection systems to monitor internal metrics, resulting in marked improvement in overall data quality.
- Deployed anomaly detection and predictive models to flag high-risk customers and reduce default/claims exposure.
- Performed exploratory data analysis, feature engineering, and model performance monitoring using Python and SQL.
- Mentored junior data scientists in analysis & ML best practices, contributing to a collaborative, growth-driven team culture.
- Leveraged SQL and Python to build data pipelines for large transaction datasets.
- Worked with cross-functional teams to integrate statistical models into insurance risk and claims decision-making.

EDUCATION

Turing College, Vilnius, Lithuania

Professional Certification | Data Analytics | May 2024-Nov 2024

École Supérieure de Gestion et de Technologie

Bachelor of Science | Business Administration |

PROJECTS

1. [Optimizing Logistics: A Case Study on Clustering and MILP for Delivery Route Optimization](#)
2. [Case Study: Building a Robust Economic Policy Simulator with MLOps](#)
3. [Building Credit Risk Decision Engine with Rules + ML + SHAP](#)
4. [Predicting Customer Churn with Machine Learning and Power BI: An End-to-End Guide](#)
5. [How I Built an AI Assistant That Turns Business Questions into Instant Insights](#)
6. [Building an End-to-End AAPL Forecasting Pipeline with Deep Learning and Darts](#)

- Expert Attention to detail
- Effective Communication
- Critical Thinking and Creativity
- Time Management and Adaptability
- Leadership and ethical decision-making

AWARDS

I4G x DataCamp Scholarship—Ingressive for Good, 2023

Digital Explorers II Scholarship—Turing College, 2024

• *Recognized for high performance and initiative in accelerating business insight delivery.*

CERTIFICATIONS/LICENSES

- [Professional Diploma in Unit Economics Management | Udemy | 2024](#)
- [Google Sheets: The Complete Google Sheets Course | Udemy | 2024](#)
- [Professional Diploma in Agile & Project Management | Udemy | 2024](#)
- [ChatGPT for Data Analytics | Maven Analytics | 2023](#)
- [Natural Language Processing | DataCamp | 2023](#)
- [Data Visualization with Power BI Simplified | Udemy | 2024](#)
- [Complete Python and Machine Learning in Financial Analysis | Udemy | 2024](#)
- [Microsoft Excel with Java and Python & OpenPyXL | Udemy | 2024](#)
- [Keras Toolbox | DataCamp | 2023](#)
- [Advanced NLP with SpaCy | DataCamp | 2023](#)
- [Machine Learning Fundamentals with Python | DataCamp | 2023](#)
- [MLOps Deployment and Life Cycling | DataCamp | 2023](#)
- [Machine Learning with Tree-Based Models in Python | DataCamp | 2023](#)

7. [Building a High-Performance Realtime Chat Server with FastAPI](#)