D602 - Deployment
QBN1 Task 1: Business Case Analysis
Prepared by Shanay Murdock

A. Objectives

The primary objectives of implementing Machine Learning Operations (MLOps) at Kronkers are to consolidate and optimize the company's fragmented machine learning initiatives into a unified, scalable, and efficient system that directly supports business growth and competitive advantage.

Operational Consolidation and Efficiency: MLOps will unify the currently siloed machine learning efforts across Marketing and Procurement departments, eliminating redundant work and enabling shared resources. By standardizing workflows and automating repetitive tasks such as model training and deployment, Kronkers can significantly reduce the manual effort currently required for quarterly report generation and ad-hoc model execution.

Cross-Department Integration: Establish a unified platform that enables seamless collaboration between teams working on personalized product recommendations, demand forecasting, and customer segmentation. This integration will allow for the development of comprehensive business applications that leverage multiple models simultaneously, overcoming the current complexity barriers that have prevented such integration.

Scalable Model Deployment: Transform the current inconsistent deployment methods–ranging from individual laptop use to internal APIs–into a standardized, scalable deployment architecture that can accommodate future growth and new use cases as Kronkers expands into additional geographic regions.

Enhanced Business Intelligence: Enable more frequent and automated insights generation by moving away from quarterly manual processes to continuous model monitoring and prediction capabilities, allowing Kronkers to respond more rapidly to market changes and customer behavior patterns.

Intellectual Property Protection: Establish secure, controlled access to machine learning models and their outputs, protecting Kronkers' competitive advantages in personalized recommendations and demand forecasting from potential competitors while maintaining compliance with data governance requirements.

B. Constraints

Several significant constraints must be addressed when implementing MLOps at Kronkers:

Multi-Language Technology Complexity: The organization's use of multiple programming languages (Python, R, and Julia) across different projects creates integration challenges.

Current models developed in different languages cannot be easily combined into unified applications, and the MLOps solution must accommodate this heterogenous technology environment without forcing a complete rewrite of existing models.

Absence of Dedicated MLOps Expertise: Kronkers currently has no employees dedicated to machine learning model maintenance, deployment, or operations. While experienced programmers and data scientists exist, they lack specific MLOps knowledge and experience, creating a skills gap that must be addressed through training or hiring.

Senior Leadership Skepticism: Several senior leaders question the utility and value of MLOps investment, creating potential resistance to resource allocation and organizational change. This skepticism may limit budget approval, slow implementation timelines, and reduce employee adoption if leadership support appears uncertain.

Current Infrastructure Limitations: The existing deployment methods are highly inconsistent, ranging from manual laptop-based execution to basic internal APIs. The current OneDrive-based storage system lacks the structure, security, and scalability required for enterprise MLOps, necessitating significant infrastructure investment.

Manual Process Dependencies: Current models rely heavily on manual parameter updates and text file imports for changing business conditions. The quarterly execution cycle for most models reflects deep integration with manual business processes that will require restructuring.

Budget Constraints: While a budget exists for machine learning maintenance and quality control, the scope of MLOps implementation may exceed allocated resources, requiring careful prioritization and phased implementation to manage costs effectively.

Distributed Team Structure: With analysts spread across different departments (Marketing, Procurement), coordinating MLOps adoption and ensuring consistent practices across teams presents organizational and communication challenges.

C. Requirements

Functional Requirements

Multi-Language Model Integration: Develop a unified platform that can execute, manage, and integrate machine learning models written in Python, R, and Julia without requiring code rewrites.

Centralized Model Repository: Establish a comprehensive repository system to replace current OneDrive storage, capable of storing model artifacts, source code, datasets up to 500 MB, and maintaining complete version control with historical tracking of both model development and data versions used for training.

Automated Model Training and Retraining: Implement automated workflows for model training and retraining that can adapt to changing business conditions without manual

intervention. This includes scheduled retraining, performance-based triggers, and automated parameter optimization.

Standardized Deployment Framework: Create a unified deployment system that replaces the current mix of internal APIs and individual laptop-based execution with consistent, scalable deployment methods accessible to all authorized users across departments.

Cross-Departmental Access Control: Develop role-based access systems that allow Marketing, Procurement, and other departments to independently access, execute, and manage relevant models while maintaining appropriate security boundaries and audit trails.

Real-Time Model Monitoring: Implement comprehensive monitoring capabilities to track model performance, detect data drift, monitor data quality, and provide alerts when models require attention or retraining.

Automated Parameter Management: Replace manual parameter updates and text file imports with automated systems that can ingest changing business conditions and update model parameters accordingly.

Integrated Analytics Dashboard: Develop user-friendly interfaces that allow non-technical users to access model outputs, run predictions, and generate reports without requiring programming knowledge.

Data Pipeline Automation: Create automated data ingestion, processing, and validation pipelines that ensure consistent data quality and availability for all models.

CI/CD Pipeline Implementation: Establish continuous integration and deployment pipelines that enable seamless model updates, testing, and deployment without disrupting ongoing business operations.

Model Versioning and Rollback: Implement comprehensive versioning systems that allow for controlled model updates and the ability to rollback to previous versions if issues arise.

Cross-Model Integration Capabilities: Enable the integration of multiple models (recommendations, forecasting, segmentation) into unified business applications that can leverage insights from all models simultaneously.

Non-Functional Requirements

Security: Implement enterprise-grade security measures including data encryption, secure authentication, network security, and compliance with industry standards to protect intellectual property and customer data.

Scalability: Design the system to accommodate Kronker’s geographic expansion, supporting increased data volumes, additional users, and new model types without performance degradation.

Usability: Design intuitive interfaces that accommodate users with varying technical expertise, from data scientists to business analysts, with comprehensive documentation and training materials.

Reliability and Availability: Maintain high system uptime with appropriate backup systems to ensure business continuity, particularly for models supporting daily operations.

Performance: Ensure model execution times meet business requirements, with quarterly reports generated efficiently and real-time predictions available within acceptable response times.

Maintainability: Ensure the system can be easily maintained and updated by Kronkers' technical team, with clear documentation, modular architecture, and standardized procedures for system administration.

Interoperability: Maintain compatibility with existing IT infrastructure, business systems, and external data sources while supporting integration with feature technologies and platforms.

Cost-Effectiveness: Optimize resource utilization to provide maximum return on investment, including efficient use of computational resources and minimal ongoing operational costs.

Compliance and Governance: Ensure adherence to data governance policies, regulatory requirements, and internal audit standards while maintaining comprehensive logs and audit trails.

Disaster Recovery: Implement robust backup and recovery procedures to protect against data loss and ensure business continuity in case of system failures.

Training and Support: Provide comprehensive training programs and ongoing support to ensure successful adoption across all user groups and technical skill levels.

Extensibility: Design the architecture to accommodate future enhancements, new model types, additional programming languages, and evolving business requirements without requiring complete system redesign.

D. Resources

WGU Course Materials.