Graduation Project Proposal Form

1. Project Information

• **Project Title:** Transforming E-Commerce Data into Actionable Insights: Optimizing Business Decisions and Customer Experience

• **Course/Track:** Microsoft Data Engineer

• **Team Members:**

1. Bassel Ashraf
2. Omar Khaled
3. Omar Elbanna
4. Mohamed Tarek
5. Ahmed Tarek

2. Project Overview

- **Objective:**
  To develop a comprehensive data engineering and machine learning solution that processes, analyzes, and visualizes e-commerce transaction data, providing valuable insights for business decision-making and enhancing customer experience.
- **Scope of Work:**
  -Ingest and process e-commerce transaction data from various sources
  -Design and implement a data warehouse using Azure SQL Database
  -Develop ETL pipelines using Python and Azure Data Factory
  -Create interactive dashboards and reports for data visualization
  -Implement data quality checks and monitoring systems
- **Expected Outcomes:**
  - A robust, scalable data pipeline for processing e-commerce data
  - A well-structured data warehouse for efficient querying and analysis
  -Machine learning models for predictive analytics (e.g., sales forecasting, customer segmentation)
  - Interactive dashboards showcasing key business metrics, trends and ML model insights
  - Automated reporting system for regular business insights
  - Improved data-driven decision-making capabilities for the e-commerce platform
  -Enhanced customer experience through personalized recommendations and targeted marketing

3. Problem Statement

The e-commerce industry generates vast amounts of transactional data, which often remains underutilized due to the lack of efficient data processing, analysis, and advanced predictive capabilities. This project aims to address the challenge of transforming raw e-commerce data into actionable insights using data engineering and machine learning techniques, enabling businesses to make informed decisions, optimize operations, and enhance customer experiences.

4. Proposed Solution

- **Technologies Used:**
  - Python: For data processing, analysis, and ETL operations
  - SQL: For data querying and management in the data warehouse

- Azure SQL Database: For storing structured data in a cloud-based data warehouse
- Azure Data Factory: For orchestrating and automating data workflows
- Azure Databricks: For big data processing and advanced analytics
- Power BI: For creating interactive dashboards and visualizations
-Machine Learning Libraries: scikit-learn, TensorFlow, or PyTorch for developing ML models

- **System Architecture:**

1. Data Ingestion Layer: Azure Data Factory for data extraction from various sources
2. Data Processing Layer: Azure Databricks for data transformation and cleaning
3. Data Storage Layer: Azure SQL Database for storing processed data
4. Analytics Layer: Python and SQL for data analysis and machine learning models
5. Visualization Layer: Power BI for creating interactive dashboards and reports

5. Resources Needed

- **Hardware/Software:**

- Azure subscription with access to Azure SQL Database, Azure Data Factory, and Azure Databricks
- Development environment with Python 3.x and necessary libraries (pandas, numpy, scikit-learn, etc.)
- SQL Server Management Studio or Azure Data Studio for database management
- Power BI Desktop for dashboard development

6. Approval

- **Instructor/Advisor: Moshira Ibrahim Ghaleb**
- **Signature:**

Moshira Ghaleb

5 sept 2024