EXECUTIVE SUMMARY

This project presents an end-to-end data analytics and machine learning solution developed on a banking customer dataset to analyse customer behaviour, assess financial risk, and support data-driven decision-making.

The project began with connecting a structured banking dataset stored in a relational database to a Jupyter Notebook environment using MySQL Workbench. This enabled efficient querying, exploration, and analysis of customer-level financial data. The dataset was then imported into Python for further processing and analysis.

An extensive **Exploratory Data Analysis (EDA)** phase was conducted to understand data distributions, detect anomalies, analyse customer demographics, and identify patterns across financial attributes such as account balances, credit usage, loans, and customer segmentation variables. Key insights derived from EDA were documented and preserved as a cleaned analytical dataset for downstream use.

To enhance stakeholder communication and enable interactive analysis, a **Power BI dashboard** was developed using the cleaned dataset. The dashboard provided visual insights into customer profiles, account balances, credit behaviour, and risk indicators, allowing non-technical users to explore trends and patterns through filters and drill-downs.

Following the analytical phase, the project transitioned into a **machine learning workflow** to build a predictive model for customer risk assessment. Feature engineering techniques were applied to create financially meaningful variables such as total account balance, credit utilization, and loan-to-income ratio. The target variable selected for modelling was **Risk Weighting**, representing customer-level financial risk categories.

Multiple machine learning models—including Logistic Regression, Random Forest, and Gradient Boosting—were trained and evaluated using appropriate preprocessing pipelines, class balancing strategies, and performance metrics such as accuracy, F1-score, and ROC-AUC. Among the tested models, Gradient Boosting demonstrated the strongest overall performance and robustness across risk classes.

The final model enables both **risk classification and probability-based predictions**, supporting practical use cases such as customer risk profiling, credit policy optimization, and targeted financial decision-making.

Overall, this project demonstrates a complete data lifecycle—from database connectivity and exploratory analysis to dashboarding and predictive modelling—showcasing the integration of data analytics, business intelligence, and machine learning in a real-world banking context.