Business Report for MALIK Financial Services (MFS)

Prepared by: Abdullah Akmal Business | Business Data Analyst

Table of Contents

1. Introduction

Purpose of the Report*:*

This report has been prepared to guide MALIK Financial Services through an important transformation by:

- **Understanding the Current Situation:** Exploring how our manual loan approval system impacts efficiency and service quality.
- **Proposing a Scalable Solution:** Introducing a partially automated process that leverages historical data insights and modern analytics.
- **Building Trust:** Explaining how our solution meets the needs of every stakeholder; from marketing and sales to internal audit and compliance.
- **Looking Ahead:** Offering recommendations for future developments, including a predictive model that further enhances decision-making.

2. Approach

2.1(a) Current Business Environment and Problem Statement:

In today's digital age, companies like MFS are experiencing a huge shift in how customers find and apply for loans. Our significant investment in digital marketing across platforms such as Google, Facebook, and LinkedIn have led to a more than 200% increase in loan applications. This digital success is exciting but also highlights a major challenge: our current manual review process is no longer efficient enough to cope with this growth.

2.1(b) Challenges We Face

The existing process requires applicants to fill out extensive forms covering personal and financial details. After submission, a team of dedicated loan officers reviews each application manually. This process presents several challenges:

- **Time-Consuming Workloads:** The sheer volume of applications leads to slower response times.
- **Resource Strain:** Our skilled team is stretched thin, limiting their ability to focus on high-value customer interactions.
- **Risk of Errors:** Manual reviews increase the potential for human error.

- **Operational and Control Risks:** Heavy reliance on manual processes makes it difficult to consistently meet regulatory standards.

2.1(c) The Impact on Our Business

Without change, continued inefficiencies may compromise customer satisfaction, increase the risk of errors, and hinder our ability to scale. A thoughtful, human-centered solution that automates repetitive tasks while supporting our team's strengths is urgently needed.

2.2 Project Requirements and Data Sources

2.2(a) What Our Users Need

After engaging with various departments, marketing, internal audit, compliance, sales, and customer service, we identified several key needs:

- **Streamlined Processes:** Automation should ease repetitive work and accelerate the loan approval process.
- **User-Friendly Tools:** The solution must be accessible to team members of all technical backgrounds.
- **Collaborative Approach:** The system should support the daily activities of various teams while aligning with our overall business strategy.
- **Regulatory Compliance:** Ensuring strict adherence to regulatory and ethical standards is non-negotiable.

2.2(b) Our Data Landscape

Our project is built on two critical data sources provided by our in-house database administrator:

- **MALIK_Loans_Database_Table.pdf:** Contains historical loan records with clear indicators (Yes/No) for loan approval outcomes.
- **MALIK Loans Data.xlsx:** An Excel file maintained by our Sales team, containing additional application details but subject to issues like duplication and missing values due to its shared storage.

Data Considerations

- **Historical Insights:** The PDF offers a reliable historical record essential for understanding trends and developing predictive models.

- **Data Quality Challenges:** The Excel file may require cleaning to address duplicates and gaps.
- **Integration Needs:** Seamlessly merging these two sources is crucial for comprehensive analysis and subsequent automation.

2.2(c) Tools and Technologies

Based on our discussions and the nature of our data, we have chosen the following tools:

- **Python:** Our language of choice for its versatility and strong data analytics ecosystem.
- **Libraries:**
    - **Pandas & NumPy:** For data manipulation and analysis.
    - **Pdfplumber or PDFMiner:** To extract data from the PDF.
    - **OpenPyXL or Pandas ExcelReader:** To work with Excel files.
    - **Matplotlib & Seaborn:** For visualizing trends and insights.
    - **Scikit-learn:** For laying the groundwork toward future predictive modeling.
- **Version Control (Git):** To manage code changes and facilitate team collaboration.
- **Centralized Data Storage:** We recommend migrating from a shared Excel file to a centralized database to reduce data duplication and improve integrity.

3. Recommendations

3.1(a) Step Towards Automation

Our plan focuses on reducing the manual workload by automating critical parts of the loan application process:

- **Extract and Integrate Data:** Automatically pull data from both the PDF and Excel sources, merging them into a single, reliable dataset.
- **Preliminary Screening:** Develop an algorithm to pre-screen applications using historical patterns and rule-based criteria.
- **Quality Checks:** Incorporate automated checks to flag duplicates and missing information, ensuring the data remains accurate.

- **User-Friendly Dashboards:** Create clear, accessible visual reports that enable stakeholders to quickly grasp key metrics and trends.

3.1(b) The Benefits of Our Approach

By implementing this solution, we will:

- **Boost Efficiency:** Automation will significantly reduce processing times, enabling our team to focus on higher-value customer interactions.
- **Improve Accuracy:** Systematic data cleaning will minimize errors.
- **Enhance Scalability:** The system is designed to grow with our business, handling increased volumes without additional strain.
- **Deepen Insights:** A unified, high-quality dataset will enable richer analytics and more informed decision-making.
- **Strengthen Control:** Automated processes reduce human error and improve compliance with regulatory standards.

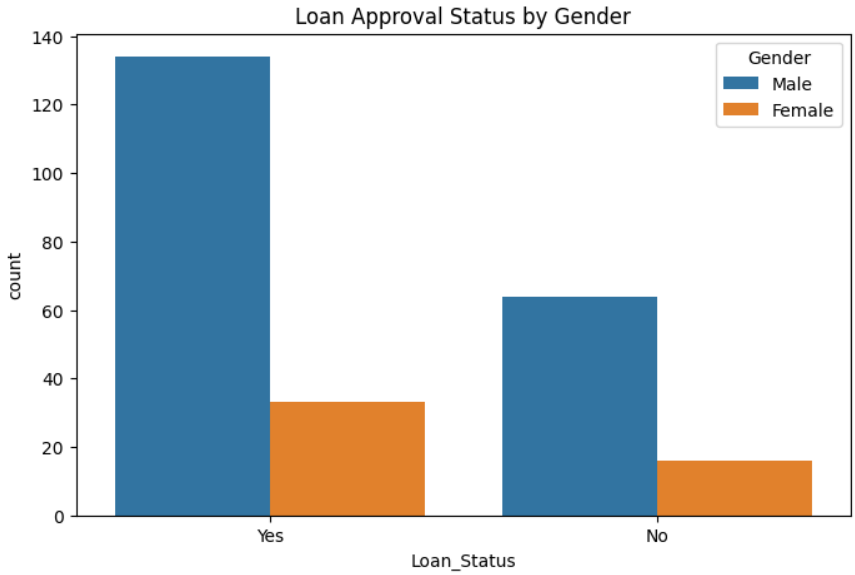
3.2(a) Exploratory Data Analysis (EDA)

A critical part of our project involved performing a thorough Exploratory Data Analysis (EDA) using the Task 2 Google collab notebook. This analysis helped us understand the underlying trends and patterns in our historical loan data, providing valuable insights that inform our automation strategy.

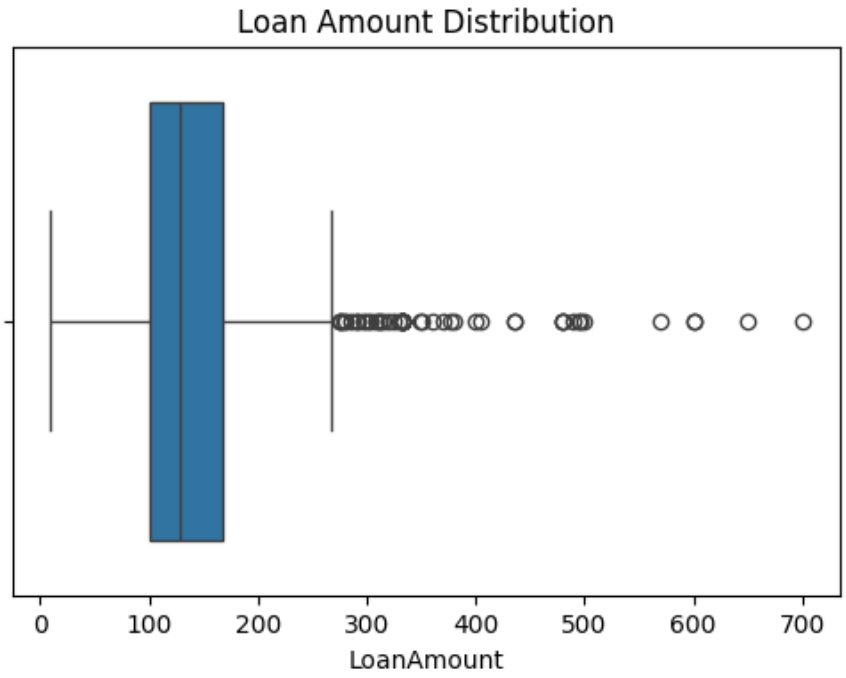
Using Python, our team analyzed the historical loan records with key objectives:

- **Data Quality Assessment:** Identified duplicates, missing values, and inconsistencies.
- **Descriptive Analysis:** Total Loaned Amount £ 95,557, Average Loan Amount is £148.15, Average Loan Term in Months is 335 and Maximum to Minimum Loan Ranger from £ 700 is £ 7.
- **Visualization:** Created charts and graphs to illustrate key trends, such as:

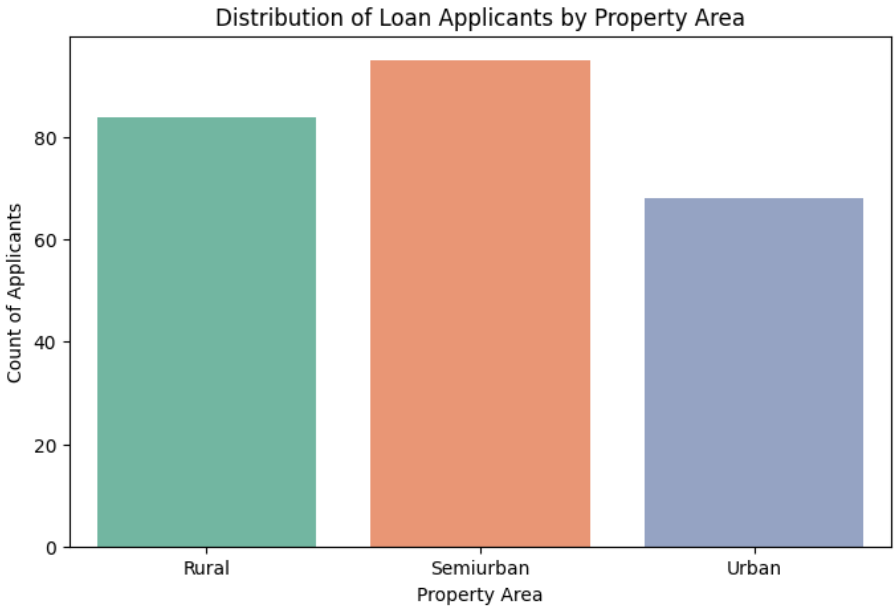
- **The distribution of loan approvals versus rejections.**



- **Loan Amount Distribution.**



- **Distribution of Loan Applicants by Property Area.**



3.2(b) Key Findings from the Analysis

The EDA yielded several actionable insights:

- **Approval Patterns:** Analysis shows that higher incomes, stronger educational backgrounds, and certain demographic factors are associated with higher loan approval rates.
- **Data Gaps:** The analysis confirmed that the Excel data contains duplicate entries, emphasizing the need for robust data cleaning.
- **Outliers in Financial Metrics:** The loan amount and income distribution have outliers which can be data errors or unique/high risk cases.
- **Future Predictive Modeling:** The insights gained from the EDA provide a strong foundation for developing a predictive model that could forecast loan outcomes with high accuracy.

3.3 Implementation Approach

3.3(a) Following a Proven Framework: CRISP-DM

Our development and implementation follow the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, which is both methodical and accessible:

1. **Business Understanding:**
   We began by mapping out how our manual process affects operations and set clear goals for automation.
2. **Data Understanding:**
   A detailed review of both the PDF and Excel files allowed us to identify data strengths and weaknesses.
3. **Data Preparation:**
   This phase involved extracting, cleaning, and merging the data. The work done during the EDA played a crucial role here by informing us of our strategies for handling duplicates and missing values.
4. **Modeling:**
   Initially, we applied rule-based methods for pre-classification. This foundation will eventually support more sophisticated predictive models.
5. **Evaluation:**
   Testing our automated process against historical data ensured accuracy. Feedback from our loans team and stakeholders is central to refining the system.
6. **Deployment and Maintenance:**
   We will implement the solution alongside our existing process, with continuous updates supported by version control (Git) and clear documentation.

3.3(b) Step-by-Step Data Preparation and Code Insights

Key steps include:

- **Data Extraction:**
  Using libraries such as Pdfplumber and Pandas, we automated data extraction from both the PDF and Excel files.
- **Data Cleaning:**
  The process involved deduplication, imputation for missing values, and normalization of data fields.

- **Data Integration:**
  We merged cleaned datasets based on common identifiers, creating a unified dataset for analysis and automation.
- **Preliminary Screening:**
  A rule-based algorithm was developed to classify applications based on historical patterns. This step was validated using the insights from our EDA.
- **Visualization and Reporting:**
  The dashboards and charts generated during the EDA have been integrated into our reporting framework, ensuring stakeholders can easily monitor key metrics.

3.3(c) Testing and Maintenance

- **Testing:**
  Validation against historical data and Excel-based verification were key to ensuring our solution meets operational needs. User Acceptance Testing (UAT) is planned to gather feedback and refine the system.
- **Maintenance:**
  Ongoing support will be provided through a robust version control system (Git), comprehensive documentation, and modular code design to ease future updates.

3.4 Challenges and Mitigation Strategies

3.4(a) Tackling Operational and Control Risks

We recognize that change comes with challenges:

- **Manual Bottlenecks:**
  The current manual review process is a known issue. Our automation will reduce these bottlenecks by cutting processing times and minimizing errors.
- **Data Quality Concerns:**
  The Excel file's issues with duplicates values are addressed through our detailed data cleaning routines.
- **Resource Strain:**
  By automating repetitive tasks, our skilled team can focus on higher-value, customer-centric activities.

- **Regulatory and Ethical Standards:**
  Our solution includes built-in safeguards to ensure data is handled securely and ethically, meeting all regulatory requirements.

3.4(b) The Value of Reusable Code

Reusable, modular code is an investment in our future:

- **Long-Term Efficiency:**
  Clear, well-documented code makes future updates and enhancements much easier.
- **Enhanced Collaboration:**
  Version control via Git fosters collaboration and ensures that knowledge is shared across the team.
- **Cost-Effective Maintenance:**
  A reusable codebase reduces long-term development costs and minimizes disruption during updates.

3.5 Recommendations for Future Work

3.5(a) Moving Toward Predictive Modeling

Beyond partial automation, our historical data and EDA insights provide an excellent basis for developing a predictive model:

- **Developing a Predictive Model:**
  Utilizing machine learning techniques (e.g., logistic regression, decision trees, ensemble methods) with libraries such as Scikit-learn or TensorFlow.
- **Evaluation Metrics:**
  Future models will be evaluated using accuracy, precision, recall, and F1-score to ensure robust performance.
- **Data Quality Enhancements:**
  Transitioning from shared Excel files to a centralized database will further improve data reliability.
- **Compliance and Ethics:**
  Future predictive efforts will continue to prioritize data security and adherence to regulatory standards.

3.5(b) Enhancing Data Quality and Collection

- **Centralized Data Repository:**
  Moving to a centralized system reduces duplication and ensures data integrity.
- **Automated Data Validation:**
  Real-time checks can prevent errors at data entry, ensuring consistent quality.
- **Regular Audits:**
  Scheduled data audits will help maintain high standards over time.

3.5(c) Investing in Our People

- **Ongoing Training:**
  Regular training sessions will keep everyone up to date with the latest tools and methodologies.
- **Change Management:**
  A strong change management strategy ensures a smooth transition and continuous improvement.
- **Enhanced Communication:**
  Open feedback channels will support collaboration and ongoing refinement of the system.

4. Conclusion

In summary, this report outlines a balanced, compassionate strategy to transform our loan application process through automation. By integrating insights from a thorough EDA documented in our Task 2 Google Collab Notebook with a scalable automation plan, we can significantly reduce processing times, improve accuracy, and empower our team to focus on building strong customer relationships.

Our proposed solution is not just about technology, it's about creating an environment where every team member feels supported, every customer is valued, and our business can continue to grow sustainably. The benefits of streamlined processes, deeper data insights, and a future-ready platform are clear. Without these changes, operational bottlenecks and compliance risks will only grow.

By embracing this plan, MALIK Financial Services will not only meet today's challenges but also be well-prepared for tomorrow's opportunities.

5. Programming Solution

5.1 System Design and Data Preparation

The technical solution is developed using Python, which offers a versatile ecosystem for data analysis and automation:

- **Data Extraction:**
  We use Pdfplumber2 or PDFMiner to extract data from the PDF and Pandas (with OpenPyXL) to read the Excel file.
- **Data Cleaning:**
  The process includes deduplication, imputation for missing values, and standardization of categorical fields (e.g., gender, marital status).
- **Data Integration:**
  Cleaned datasets are merged based on common identifiers, forming a unified dataset for both preliminary screening and future predictive modeling.

5.2 Pseudocode

Below is a simplified pseudocode outline of our approach:

```
BEGIN
  // Data Extraction
  LOAD pdf_data FROM 'MALIK_Loans_Database_Table.pdf' USING PyPDF2
  LOAD excel_data FROM 'MALIK Loans Data.xlsx' USING Pandas ExcelReader
  // Data Cleaning
  FOR each record IN pdf_data DO
    IF record IS duplicate OR has missing values THEN
      HANDLE record (impute or flag)
    ENDIF
  ENDFOR
  FOR each record IN excel_data DO
    IF record IS duplicate OR has missing values THEN
      HANDLE record
```

```
    ENDIF
  ENDFOR
 // Data Integration
  MERGE pdf_data AND excel_data INTO unified_dataset ON common_identifier
  // Preliminary Screening
  FOR each application IN unified_dataset DO
    APPLY screening_rules BASED ON historical patterns
    CLASSIFY application AS 'approved' OR 'rejected'
  ENDFOR
  // Visualization and Reporting
  GENERATE charts and graphs FROM unified_dataset
  // Output Processed Data
  EXPORT unified_dataset TO centralized database
END
```

Here is Google Collab Notebook link for task 2 EDA.

6. References

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., & Wirth, R. (2000) *CRISP-DM 1.0: Step-by-step data mining guide.* SPSS Inc.
- Davenport, T. H. & Harris, J. G. (2007) *Competing on Analytics: The New Science of Winning*. Harvard Business School Press.
- Provost, F. & Fawcett, T. (2013) *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Python Software Foundation (2024) *Python Documentation*. Available at: https://docs.python.org/3/.
- McKinsey & Company (2018) *The analytics advantage: How data science is transforming business.* Available at: https://www.mckinsey.com