# Analytics Startup Plan

**Synopsis: *This document provides a high-level walkthrough of the activities required to guide completion of the analysis.***

|  |  |
| --- | --- |
| **Project** | Revolutionizing Cash Flow Management with Predictive Analytics and Strategic Customer Segmentation |
| **Requestor** | *Centennial College* |
| **Date of Request** | *16th July 2024* |
| **Target Quarter for Delivery** | *2nd* |
| **Epic Link(s)** |  |
| **Business Impact** | *Implementing predictive analytics and strategic customer segmentation to enhance cash flow predictability will profoundly impact our business by significantly improving financial stability and operational efficiency. Accurately forecasting incoming payments will reduce uncertainty in cash flow management, allowing for better budgeting and financial planning. Tailoring collection strategies to individual customer segments based on their payment behaviors will lead to more effective debt recovery and optimized resource allocation. This approach not only minimizes the risk associated with high-risk customers but also streamlines collection efforts, resulting in reduced operational costs and improved overall profitability. By focusing on these data-driven strategies, the business can expect more reliable cash flow, enhanced decision-making, and stronger financial health.* |

## 1.0 Business Opportunity Brief

|  |  |
| --- | --- |
|  | Clearly articulated business statement of the Ask, opportunity, or problem you are trying to solve for. An important step is to understand the nature of the business, system or process and the desired problems to be addressed. This will be communicated back to All stakeholders for alignment.  To improve cash flow predictability and optimize collection strategies, we aim to predict customer payments based on current bill and opening balance, while also segmenting customers using factors like active month and payment history. This will help forecast incoming payments, tailor collection approaches, and focus resources on high-risk customer segments, enhancing operational efficiency. |

………

**The specific ask**

We will build a machine-learning model to predict customer payments and segment customers based on payment history and other relevant factors

## 1.1 Supporting Insights

|  |  |
| --- | --- |
|  | Define any supporting insights, trends and research findings. Where relevant, list key competitors in the market. What are their key messages, products & services? What is their share of the market, nationally and regionally?  ***Payment History Impact:*** *Customers with irregular payment histories are more likely to default on payments.*  ***Competitors:*** *Other utility companies leverage machine learning for similar predictive analytics to enhance collection efficiency and customer segmentation* |

## 1.2 Project Gains

|  |  |
| --- | --- |
|  | *Describe any revenue gains, quality improvements, cost and time savings (as applicable). What will you do differently and why would our customers care. What are the implications if we do nothing? This section is particularly key for prioritization against company goals and KPI’s.*  ***Revenue Gains:***   * *Improved cash flow management by predicting payments accurately.* * *Tailored collection strategies to reduce overdue accounts.*   ***Quality Improvements:***   * *Enhanced customer relationship management through personalized communication.* * *Improve customer experience through enhanced service delivery.*   ***Cost and Time Savings:***   * *Efficient allocation of resources by focusing on high-risk customers.*   ***Implications if Nothing is Done:***   * *Continued unpredictability in cash flows.* * *Inefficient use of collection resources.* * *Poor customer experience.* |

## *Note: Completion of the following sections is possible only after a careful assessment and triage of the Ask. This is required to determine scope, resource, time, priority, and data availability.*

## 2.0 Analytics Objective

|  |  |
| --- | --- |
|  | List the key questions, and assumptions and define the hypotheses. Often the deliverable may not just be an analysis output, but a recommended operating model or blueprint for a pilot, etc.  Note: Asking the right questions and truly understanding the problem will lead to the right data, right mathematics, and right techniques to be employed.  ***Key Questions:***   * *Which factors are the most predictive of (drives) customer payments?* * *How can we effectively segment customers based on payment behavior?* * *How can we ensure effective and efficient resource allocation?*   ***Hypotheses:***   * *Customers with a high number of active months and regular payment patterns are more likely to make payments on time.* * *Little or No opening balances and full current bill payments(current payment) indicate better payment behavior.* * *ROI is high when resources(huge investment) are allocated to areas with customers with good payment behavior.* |

## 2.1 Other related questions and Assumptions:

|  |  |
| --- | --- |
|  | *List any assumptions that may affect the analysis*  ***Assumptions:***   * *Payment patterns and opening balances are reliable indicators of future payment behavior and resource allocation.* * *Data is accurately recorded and up to date.* |

## 2.2 Success measures/metrics

|  |  |
| --- | --- |
|  | *What does success look like? Define the key performance indicators (success definition/indicators, drivers and key metrics) against which the objectives will be analyzed. These should be drawn from the interlocking meeting with key stakeholders and will inform the approach and methodology for the analysis.*  ***Key Performance Indicators:***   * *Accuracy of the payment prediction model.* * *Reduction in overdue accounts.* * *Efficiency improvements in collection strategies.* * *ROI performance* * *CRM performance* * *Collection performance (Current payment/ Revenue billed)* |
|  |  |
|  |  |

## 2.3 Methodology and Approach

|  |  |
| --- | --- |
|  | *Now that you have a good understanding of the Ask and deliverable, detail the recommended approach/methodology.* |

**Type of Analysis**

* *Linear regression for current payment prediction*
* *Decision trees for understanding variable importance.*
* *Market Basket Analysis for Customer Segmentation*

**Methodology**

1. *Data Exploration: Understand the structure and quality of the data.*
2. *Feature Engineering: Create relevant features such as payment history patterns.*
3. *Model Building: Develop predictive models to forecast customer payments.*
4. *Customer Segmentation: Use clustering techniques to segment customers based on payment behavior.*

**Output**

* *Predictive model for customer payments.*
* *Customer segments with tailored collection strategies*

## 3.0 Population, Variable Selection, considerations

|  |  |
| --- | --- |
|  | Capture learning about the data available today location, structure, and reliability; this would include data in operational systems including dealer sourced, data warehouse and any CRM or email marketing systems available today. |

**Audience/population selection:** All active customers

**Observation window:** Data from January to July 2023

**Inclusions:** Account number, location, current opening balance, current closing balance, and other necessary information

**Exclusions:** Customer Name, address, staff number staff name, and other unnecessary features

**Data Available:** Customer information (e.g. account number, tariff), Payment history (e.g. current opening balance, current payment), and Activity metrics (e.g. active months, payment patterns)

**Data Sources:** Customer Information System (CIS)

**Audience Level:** All active customers within the network

**Variable Selection:** Features such as current opening balance, current payment, active months, and payment patterns will be used.

**Derived Variables:**

**Assumptions and data limitations:**

* Data is assumed to be complete and accurate.
* Missing values or inaccuracies in data recording might affect the model's performance

## 4.0 Dependencies and Risks

|  |  |
| --- | --- |
|  | Identification of key factors that may influence the outcome of the project and likelihood of it happening: |

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Likelihood (based on historical data)** | **Delay (based on historical data)** | **Impact** |
| *Data Quality Issues* | *Low* |  | *Low* |
| *Changes in customer behaviour.* | *Low* |  | *Low. 6 months of data should show good customer behaviour* |

## 5.0 Deliverable Timelines

|  |  |
| --- | --- |
|  | List key dates and timelines as a work-back schedule. Activate line items based on complexity and line-of-sight required. Will set the stakeholder expectations for the process. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item** | **Major Events / Milestones** | **Description** | **Scope** | **Days** | **Date** |
| 1. | Kick-off / Formal Request | *Data Begins* |  |  |  |
| 2. | EDA | *Assess data quality, perform in-depth analysis of statistical measures* |  | 7 | *July 8th 2024* |
| 3. | Modeling | Modeling |  | 14 | *July 15th 2024* |
| 4. | Governance | Documentation, Governance Plan, presentation peer review |  | 7 | *July 29th 2024* |
| 5. | Presentation | Documentation, Governance Plan, presentation peer review |  | *7* | *August 5th 2024* |
| 6. | Portfolio | Presentation, Portfolio |  | *7* | *August 12th 2024* |