# Rimilia - Customer Segmentation to Derive Customer Strategies

## Background:

Rimilia is a software company focused on providing full-stack AI for automated financial processes. These processes include cash payments/accounts receivables and providing insight on managing risk relating to their clientele's customers' financial behaviour. Thus, Rimilia needs to be able to strategize financial decisions based on an individual customer's payment behaviour.

This proposal highlights the viability, benefits, and methodology of exploring customer segmentation to derive the proposed customer strategies. Customer segmentation divides a customer base into groups of individuals who share common traits such as average time to pay, reliability/consistency of payment times, region, sector, etc. One may be able to predict a customer's future behaviour according to their segment(s), under the assumption their future behaviour will be similar to others in their segment.

Some potential strategies for customer segmentation include cluster analysis, attribute-based segmentation (i.e. firmographic), deep learning/neural networks, and principal component analysis.

## Problem:

The goal of this project is to identify what relationships exist, if any, between the information we have about a customer and how long they take to pay invoices. This benefits the company in that it allows them to sort customers into groups, letting them better predict customers' behaviour and deal with them in a more personalized manner. Customers whose characteristics correlate with paying sooner may be given preferential treatment. We are most interested in relations between a customer's behaviour and their region or business sector.

To accomplish this, we will be working with the company's new ML-based forecast engine. We aim to improve this engine's performance using upgraded statistical and ML methods. The objective will be to maximize the upgraded model's accuracy and flexibility, while also minimizing the time the model takes to run. This will be subject to the constraint that the upgraded model must be at least as accurate as the original.

## Data:

All the data the team will use will be provided by Rimilia. As Rimilia's head of data science is away from office, our ability to describe the datasets is significantly limited. However, he gave high-level information on what the data would consist of.

The primary raw data will be around customer payment and trade behaviours. The database can be divided into 6 main attributes. All of these can be analyzed on a daily, weekly, and monthly basis:

1. Customer Identifiers (anonymised).
2. Amount of outstanding debts per customer (split into invoices).
3. Raised and due dates of customer debts or invoices.
4. Payments against customer invoices or debts together with the dates.
5. Type of payments.
6. Any credit against customer accounts.

The head of data science told us that we could infer a lot of other information from the raw data. However, we won't know what this additional information is until we see the data itself.

**Methods:**

1. **Data cleaning, augmentation and normalization.** The anonymized data will need to be cleaned before we can do exploration & modeling on it.
2. **Exploratory data analysis.** This includes exploration of attributes and their relationships with methods that extend beyond basic statistics, as well as data visualization techniques.
3. **Design and implementation of data science models.** The primary technique will be clustering since we want to characterize customers into a behavior category. Other time sensitive ML models could also be employed since we can analyze the data periodically.
4. **Segmentation results.** The segments produced will indicate customer payment behavior, which could be useful for personalized customer understanding and treatment. For example, customers found to pay more reliably may be trusted with longer payment terms.

**Expectation:**

We expect our model to be able to identify different clusters of payment behaviour between customers, and to classify new customers into a cluster. We expect these clusters to give good information about payment characteristics and description about the customers in them.

If enough time/resources are available, we will attempt to investigate the impact of region and business on customers' payment behaviour. For this, we expect to recognize how well a customer's region and business sector correlate with their behaviour. This will be expressed as a table or chart showing typical behaviour per region and sector.

**Attribution table:**

| Team member | A. Kraft | A. Xie | D. Escoboza | A. Zhou | C. Kim |
|---|---|---|---|---|---|
| Contribution | Wrote problem statement and proof-read document | Wrote description of the data<br><br>Fixed some of the wording | Outlined expectations | Outlined the methods of the project | Explained the background |