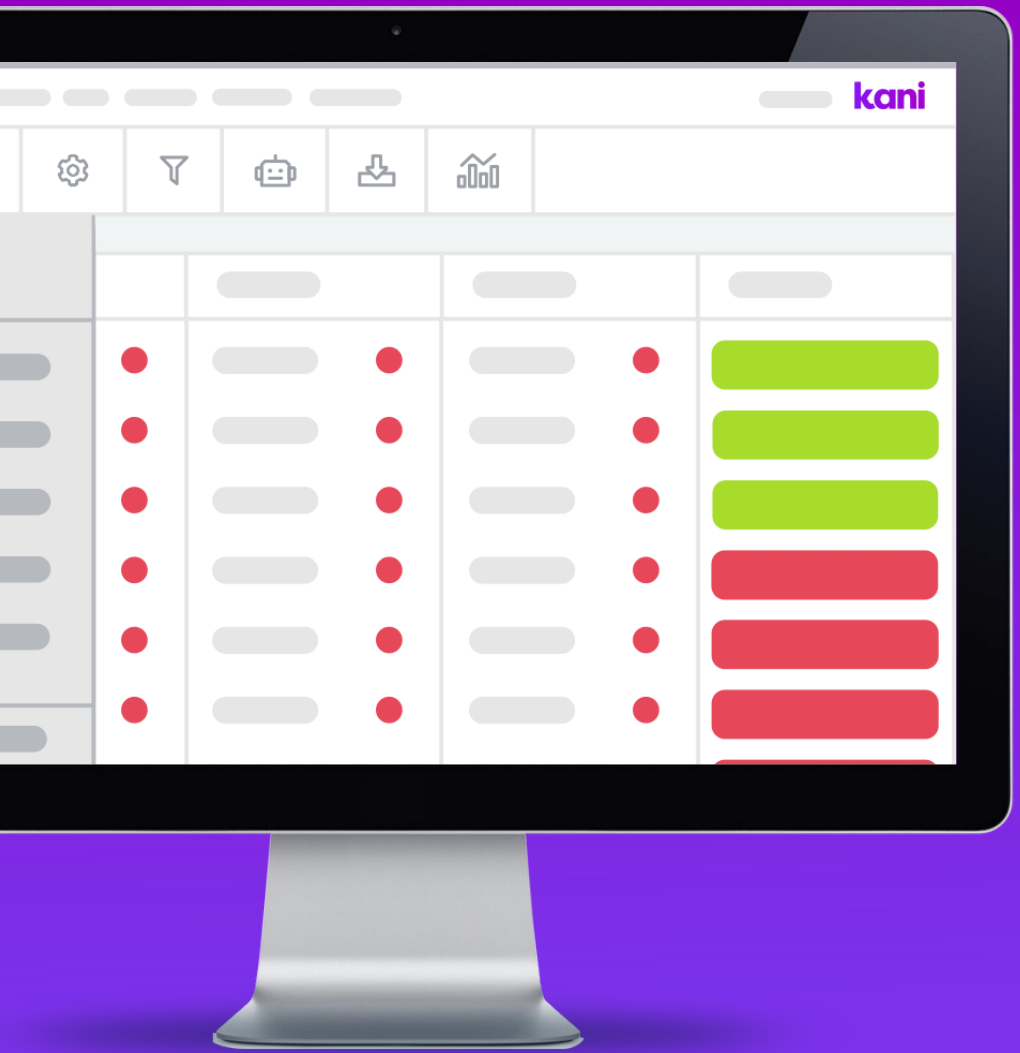
Data challenges in the payments industry

Sophie Harbisher

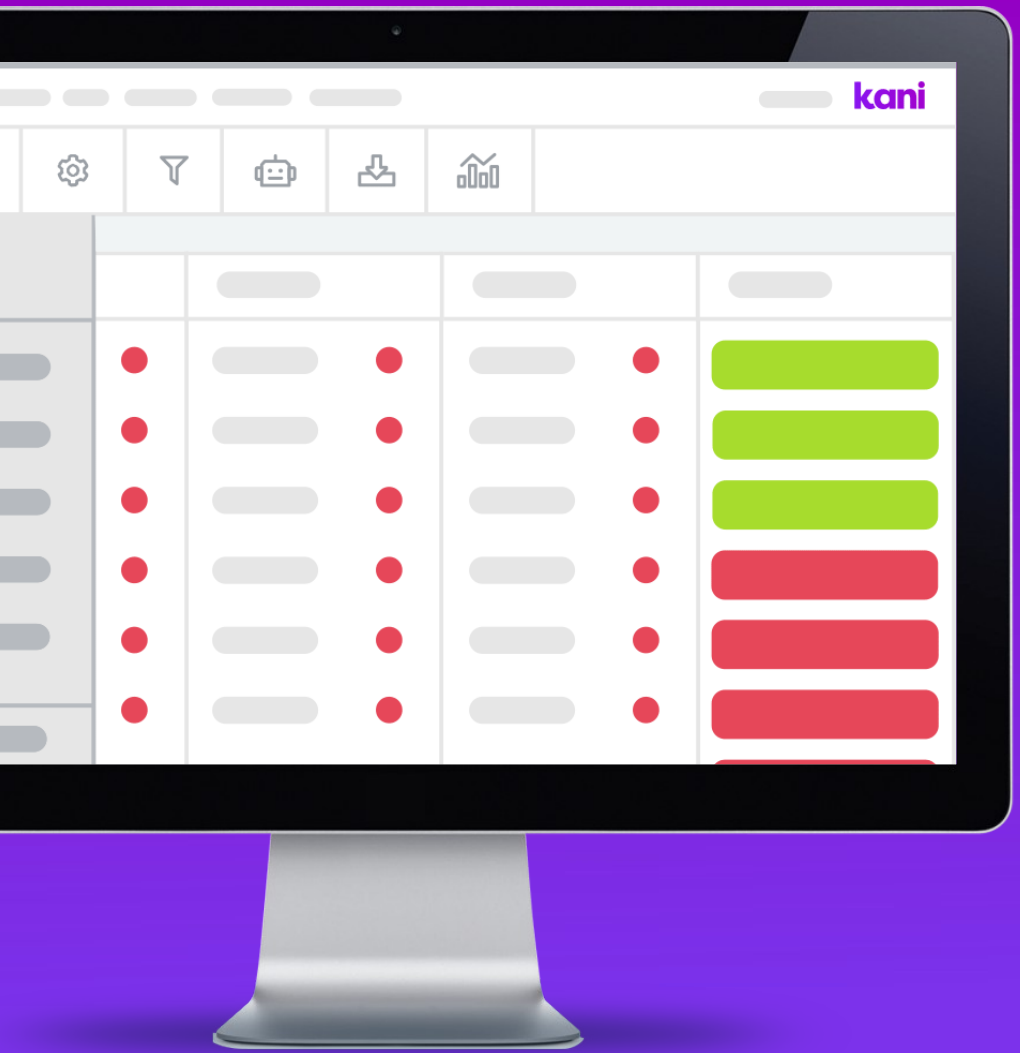
kani



Talk Outline

- Who are Kani and what do we do?
- Data challenges in the fintech industry
- Reconciling data

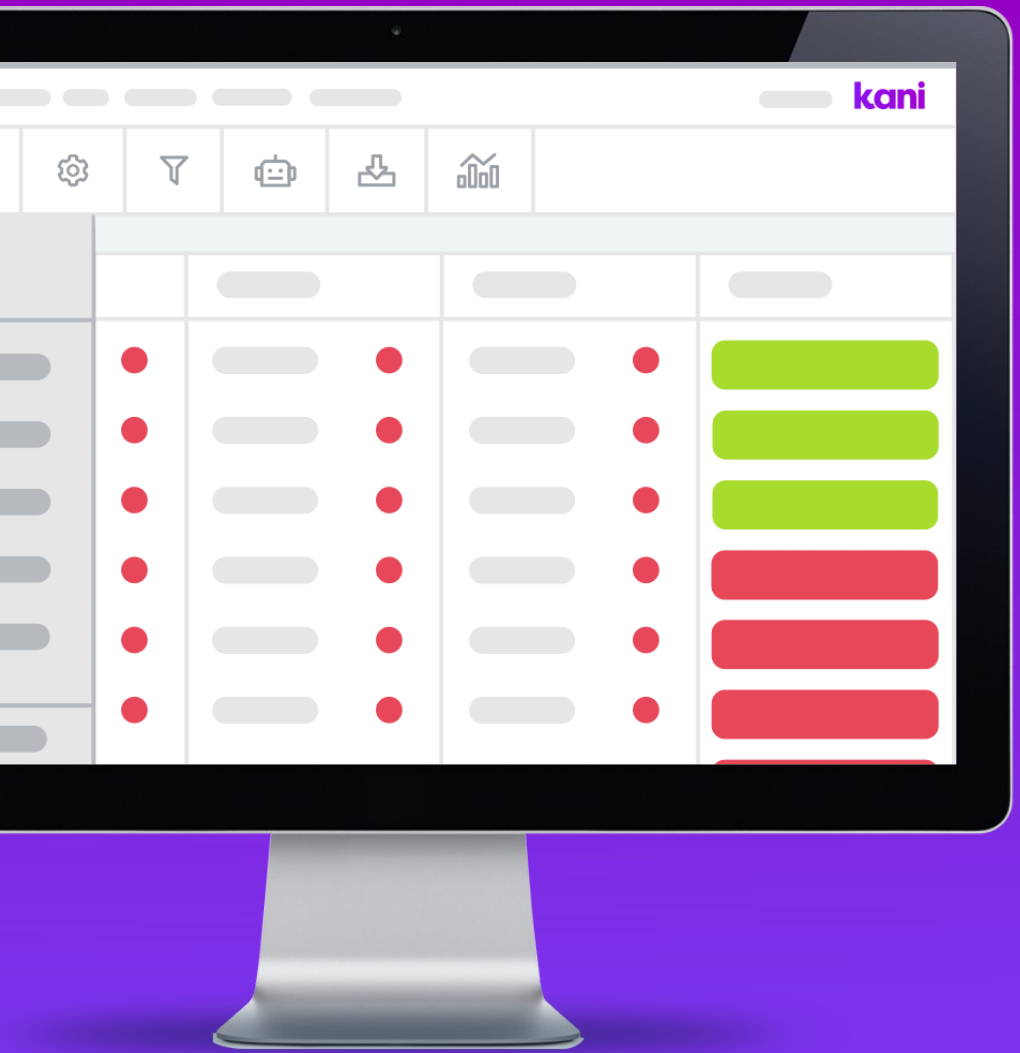
kani



A short guide to Kani

- Kani is a Fintech based in Newcastle
- Offer SasS to fintech companies
- The platform ingests data from multiple sources, handles reconciliations and provides BI reporting

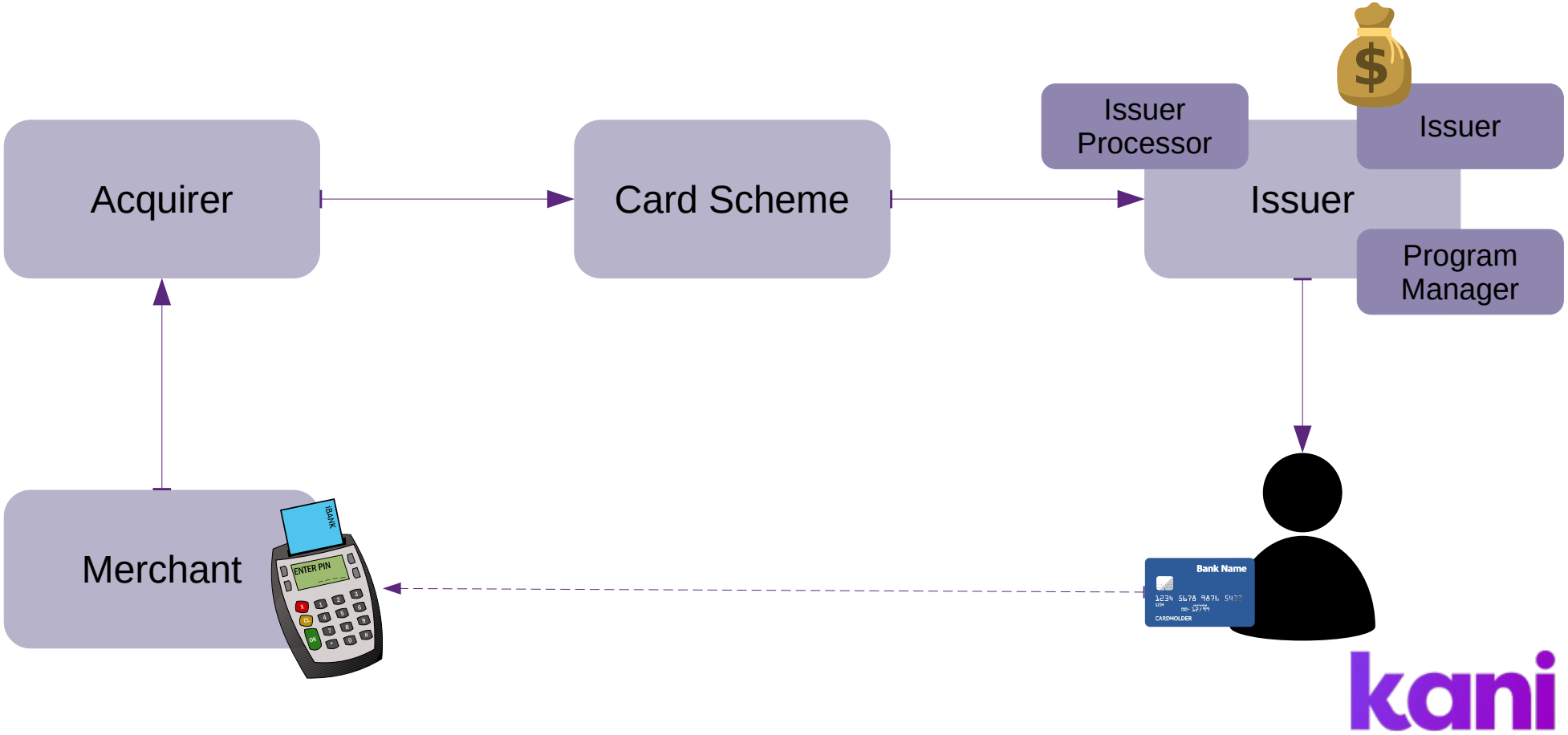
kani



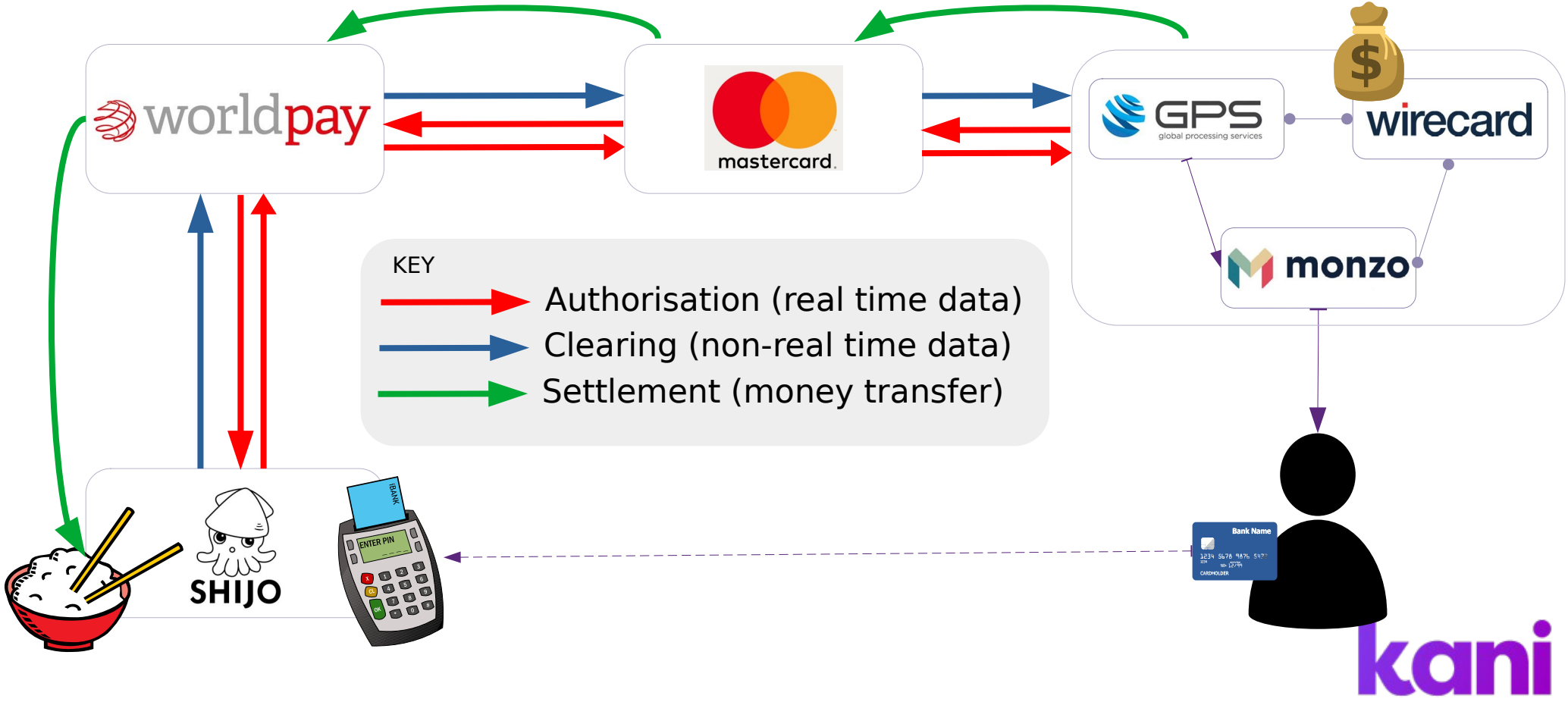
Payments Data

kani

Transaction Lifecycle



Transaction Lifecycle

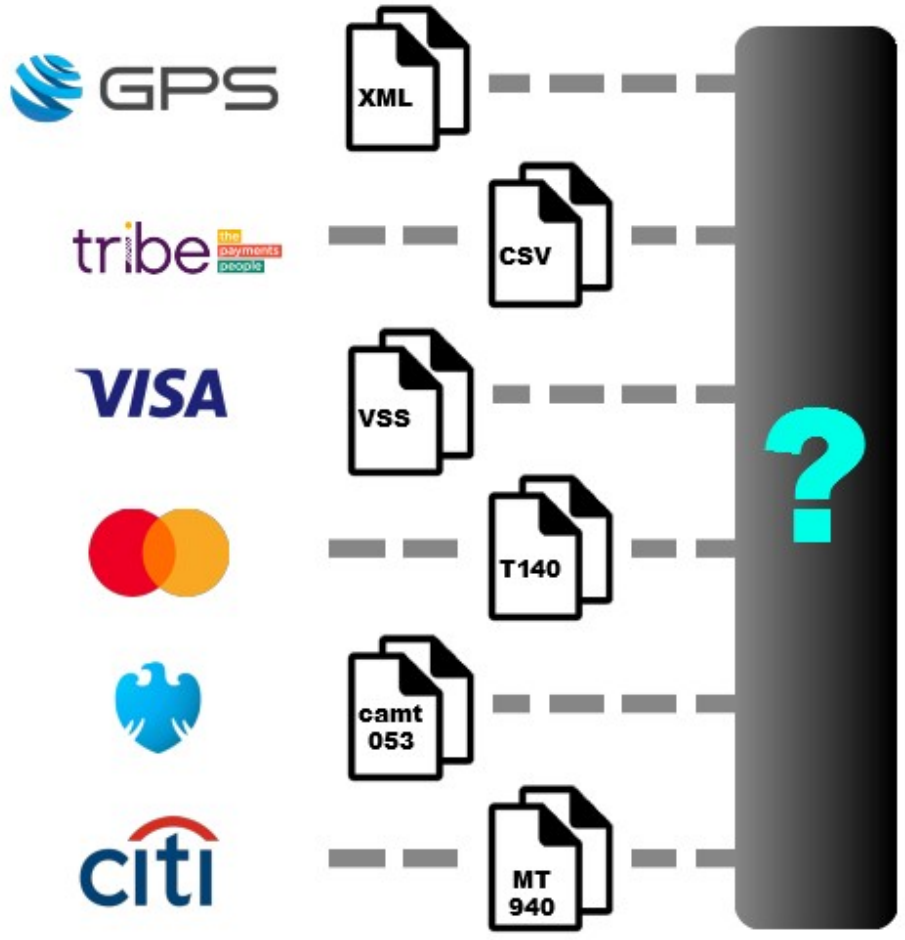


Transaction Lifecycle

- Messages are passed between acquirers, card schemes and issuers(/issuer processors)
- Settlements are aggregated (daily/over a clearing cycle) so that money transfers are made periodically rather than for every transaction
- Data passed around the transaction lifecycle isn’t necessarily consistent in format or interpretation
- Lack of understanding about various data elements

kani

Transaction Data



IT
build project needed; unfamiliar with what finance & compliance need to achieve

Finance
regular requests to IT due to lack of control over data, often fallback to MS Excel

Compliance
regular requests to IT due to lack of control over data, often fallback to MS Excel

Issues

- Finance companies need accurate and verifiable information
- Reporting and reconciliation is essential for payments companies
- This is difficult to achieve, time consuming and often inaccurate

kani

Record Matching

kani

Record Matching

- **If** a unique reference exists then reconciling across different data sources is easy
- However this is not always the case and we have to try to match based on similarity
- E.g. bank statement vs store receipt

kani

Record Matching

The record matching problem is often broken down into three main stages:

1. Similarity scores for individual record elements
2. Overall similarity score between records
3. Suggested matches

kani

Data Element Similarity

- There are 3 different data types we may want to assess similarity over:
  - Numeric quantities
  - Dates
  - Text
- The aim is to get a value which indicates which records are similar to each other and which are not

kani

Text scoring

- Classical comparisons between strings often belong to the family of *edit distances* (e.g. Levenshtein, Jaro-Winkler,...)
- E.g. the Levenshtein distance between **canny** and **kani** is 3

$$\text{canny} \xrightarrow{\text{del}} \text{cann} \xrightarrow{\text{sub}} \text{cani} \xrightarrow{\text{sub}} \text{kani}$$

- More recent developments in string comparisons leverage natural language processing

kani

Record Similarity & Matching

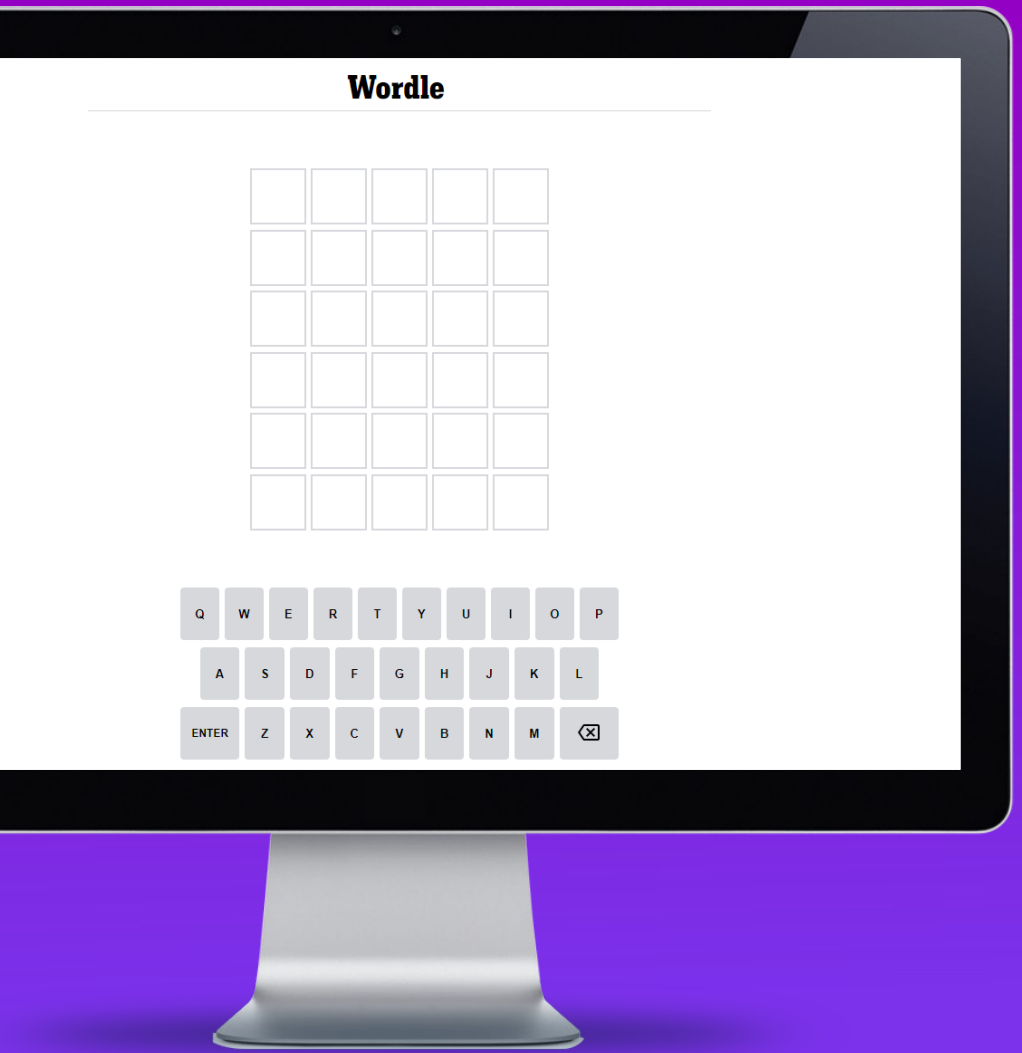
- Consideration of how to calculate an overall score given similarity scores of individual components
  - Logistic Regression
  - Naive Bayes Classification
  - Other…
- Suggesting which records are matches is difficult

kani

Record matching challenges

- Scalability & curse of dimensionality
- Ideally would want a probability of a match between records – lack of training data for supervised models, still issue of how to recommend matches
- Matching over different aggregations of numeric quantities
- Record matching/entity resolution is an active research area with some interesting approaches

kani



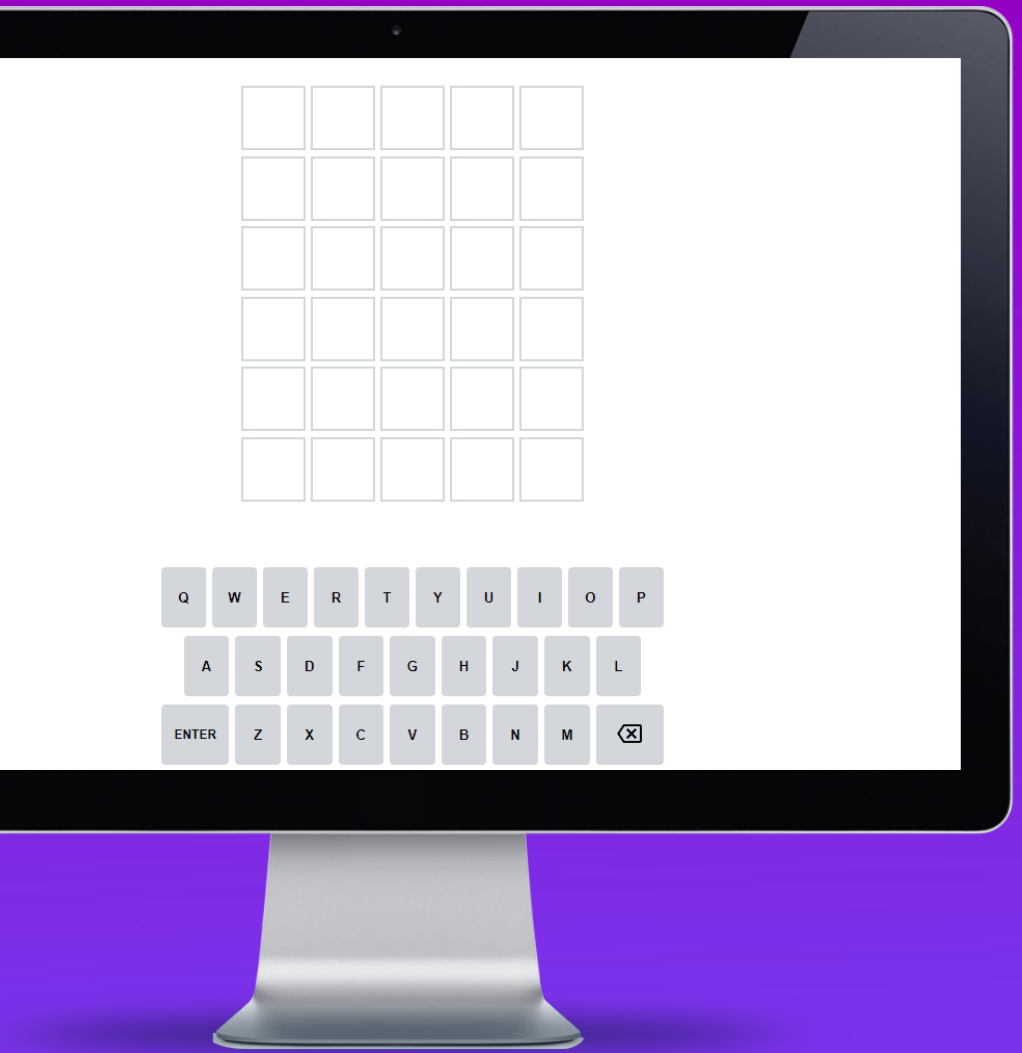
Application

kani

Wordle

- Aim is to guess what the five letter word is within 6 tries
- Every guess gives you some information
- Question: What is the best initial guess?

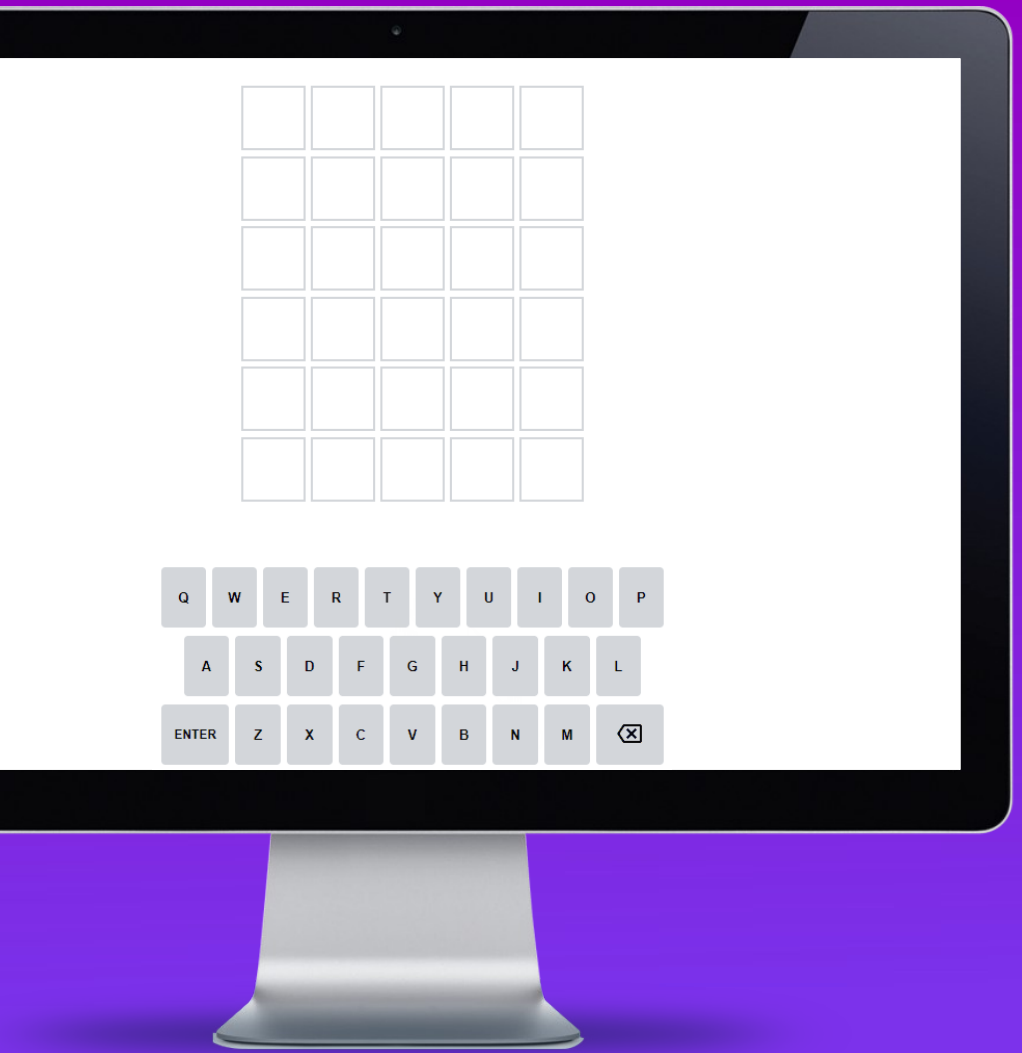
kani



Wordle

- Text similarity settings:
  - Correct letter, correct position gets similarity of 1
  - Correct letter, wrong position gets score of 0.5
  - Incorrect letter gets score of 0
- Aim to find the word that is on average closest to any other word

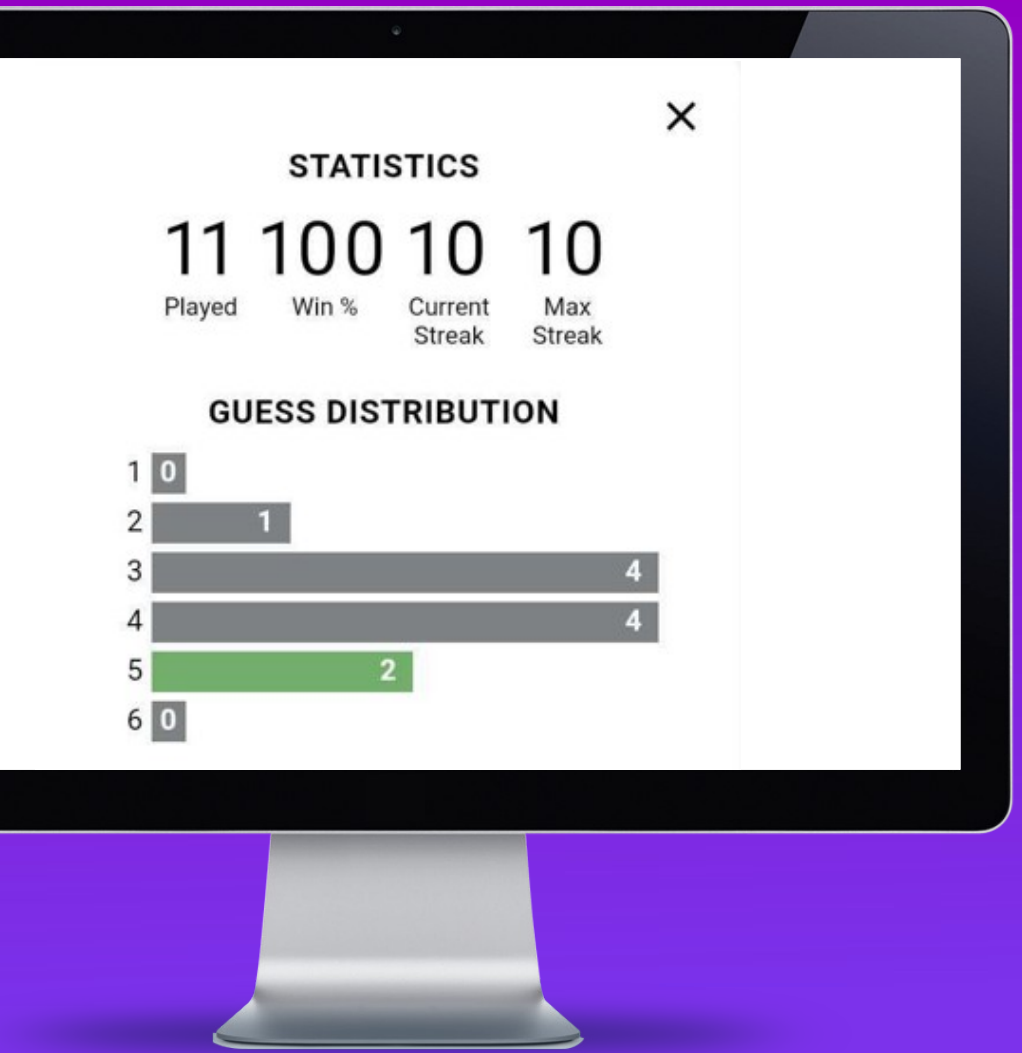
kani



Wordle

- Best initial guess is TARES
- Other good starting words:
  - tares 1.427
  - lares 1.421
  - rales 1.398
  - nares 1.397
  - cares 1.396

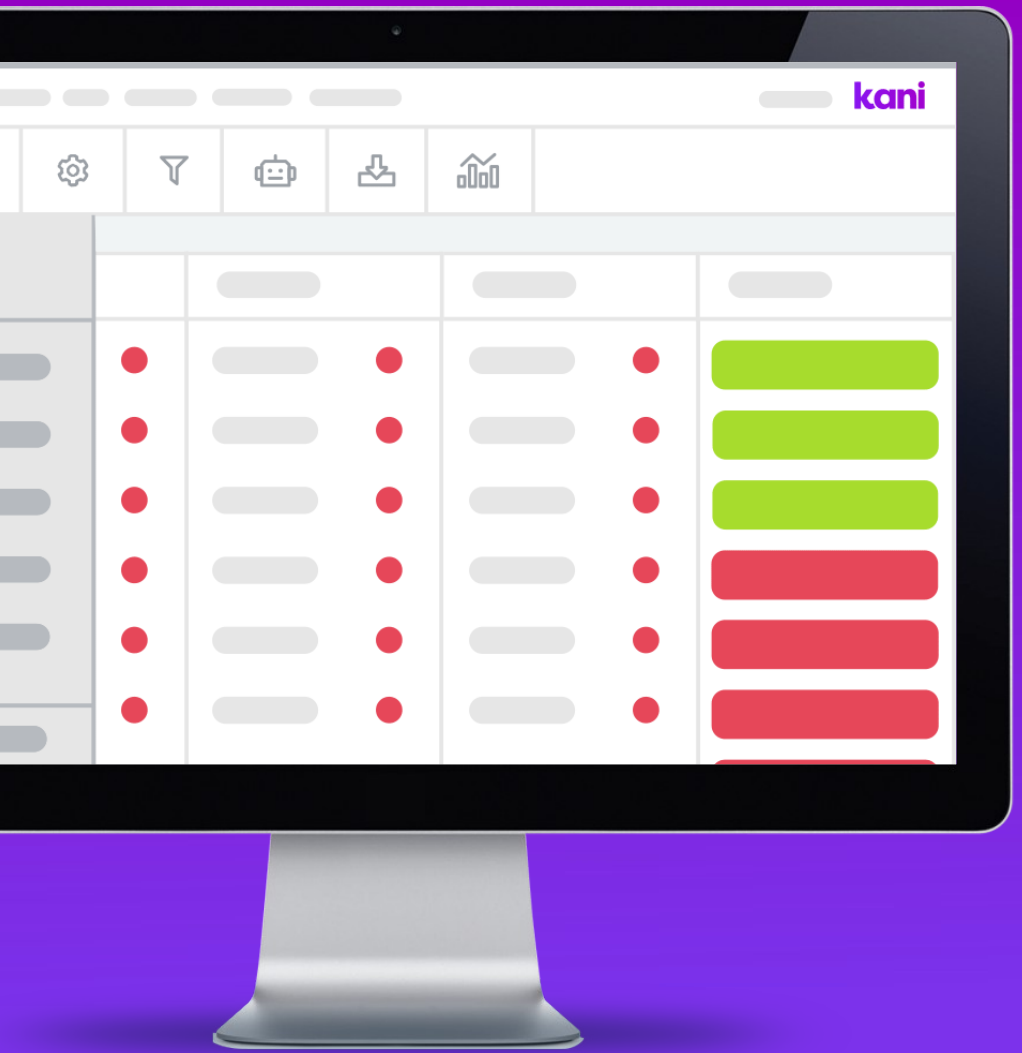
kani



Wordle

- Best initial guess is TARES
- Other good starting words:
  - tares 1.427
  - lares 1.421
  - rales 1.398
  - nares 1.397
  - cares 1.396

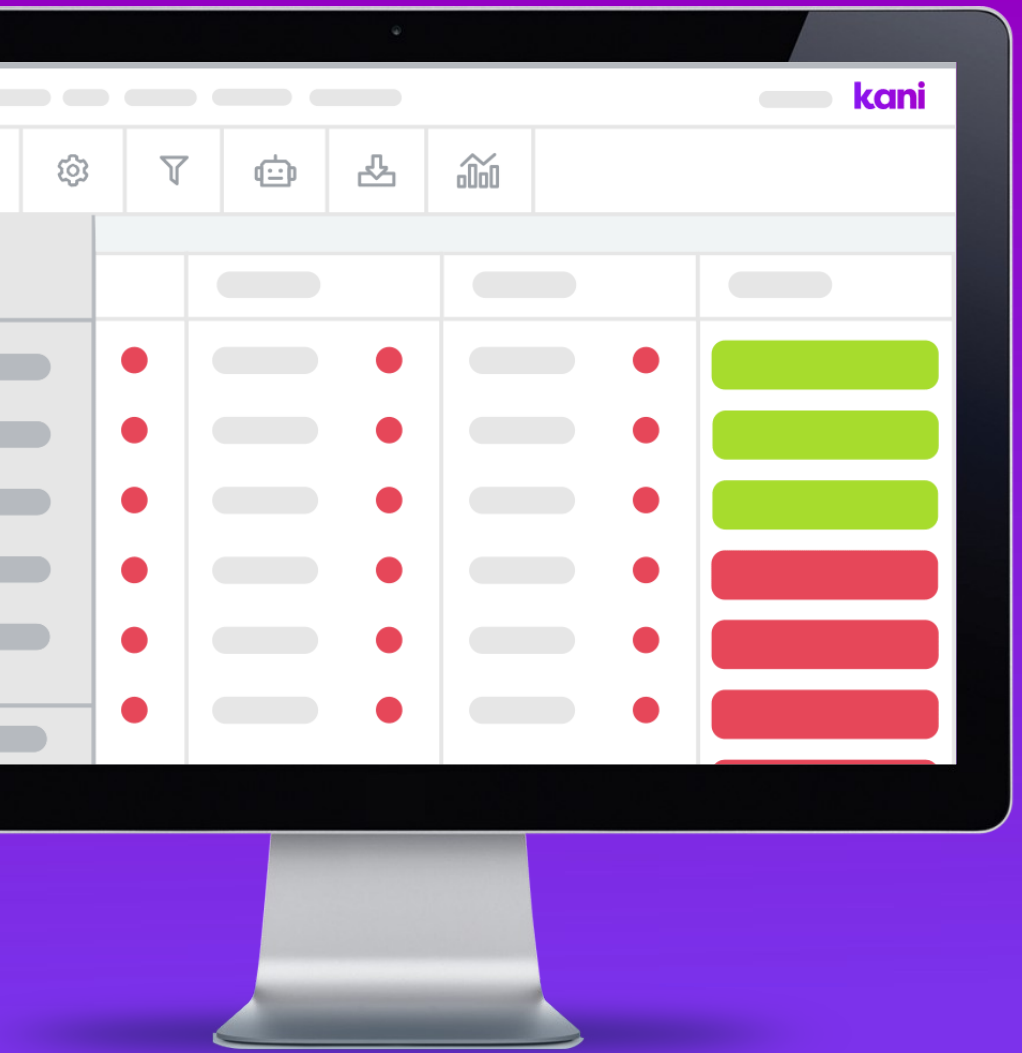
kani



Summary

- Understand the transaction lifecycle
- Reviewed the difficulties fintech companies face
- Brief overview of record matching and how we can attempt to match data
- How to make a good start to Wordle

kani



Thanks for listening

sophie@kanipayments.com

hello@kanipayments.com

kani