The Evolution of Data Analytics in Internal Audit

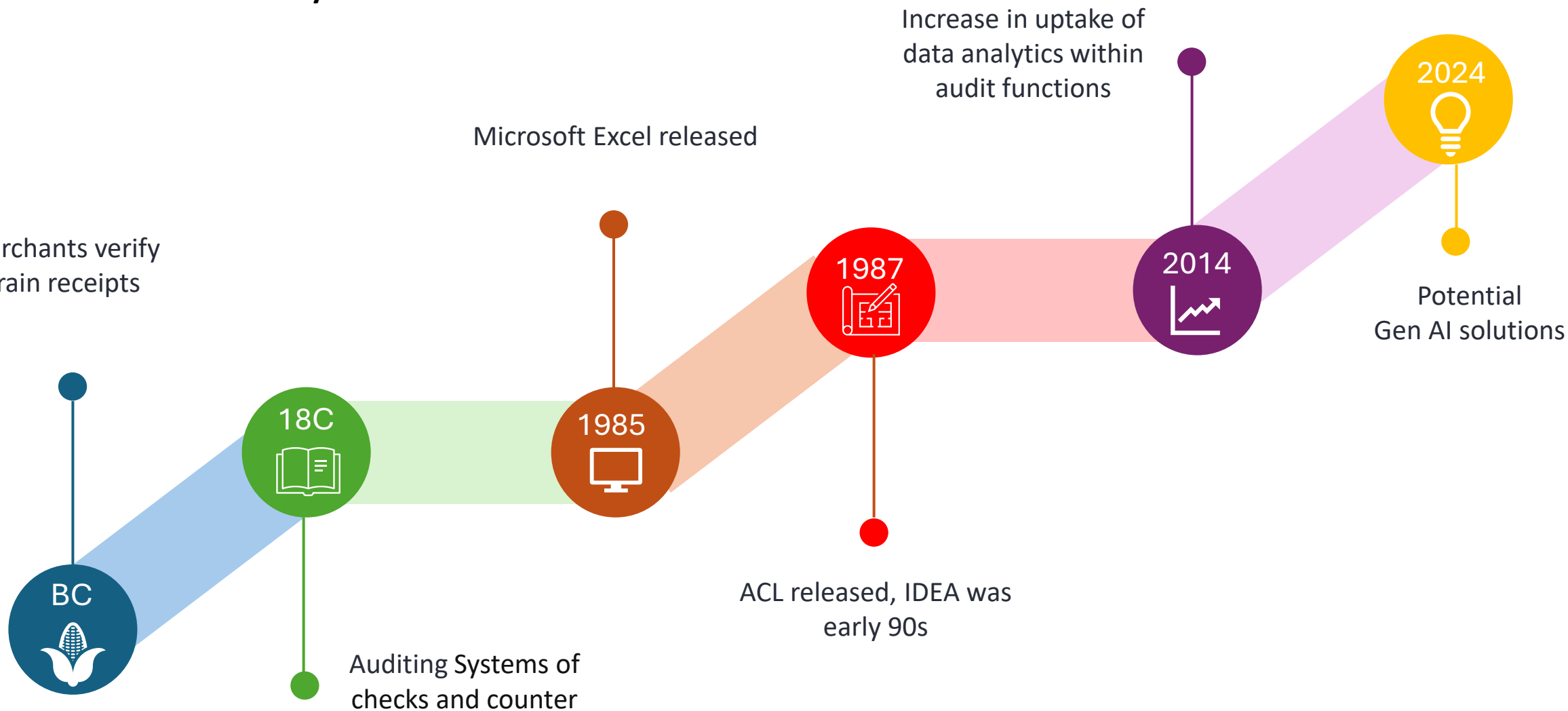
Zoe Adey

10 June 2025

Agenda

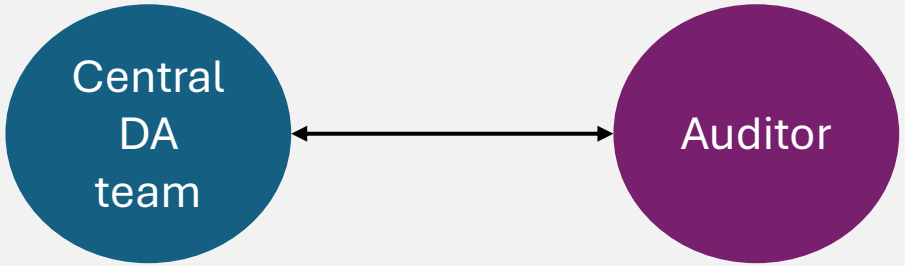
- Timeline of how DA has evolved in Internal Audit
- Types of Data Analytics
- Overview of Low Code Tools
- Examples of how to use low code tools for more advanced analytics:
  - K-means clustering in Knime
  - Sentiment analysis in Alteryx

The Evolution of Data Analytics in Internal Audit



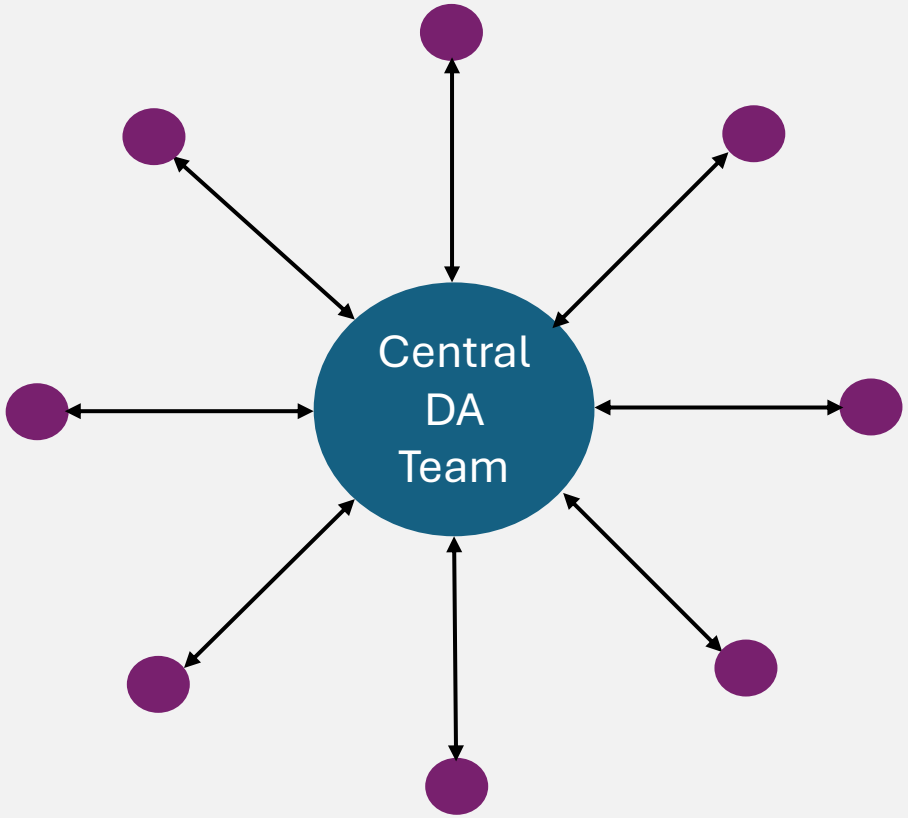
Team structure

Historic Structure



Central DA team perform majority of analytics and share results with auditors

Evolving Structure

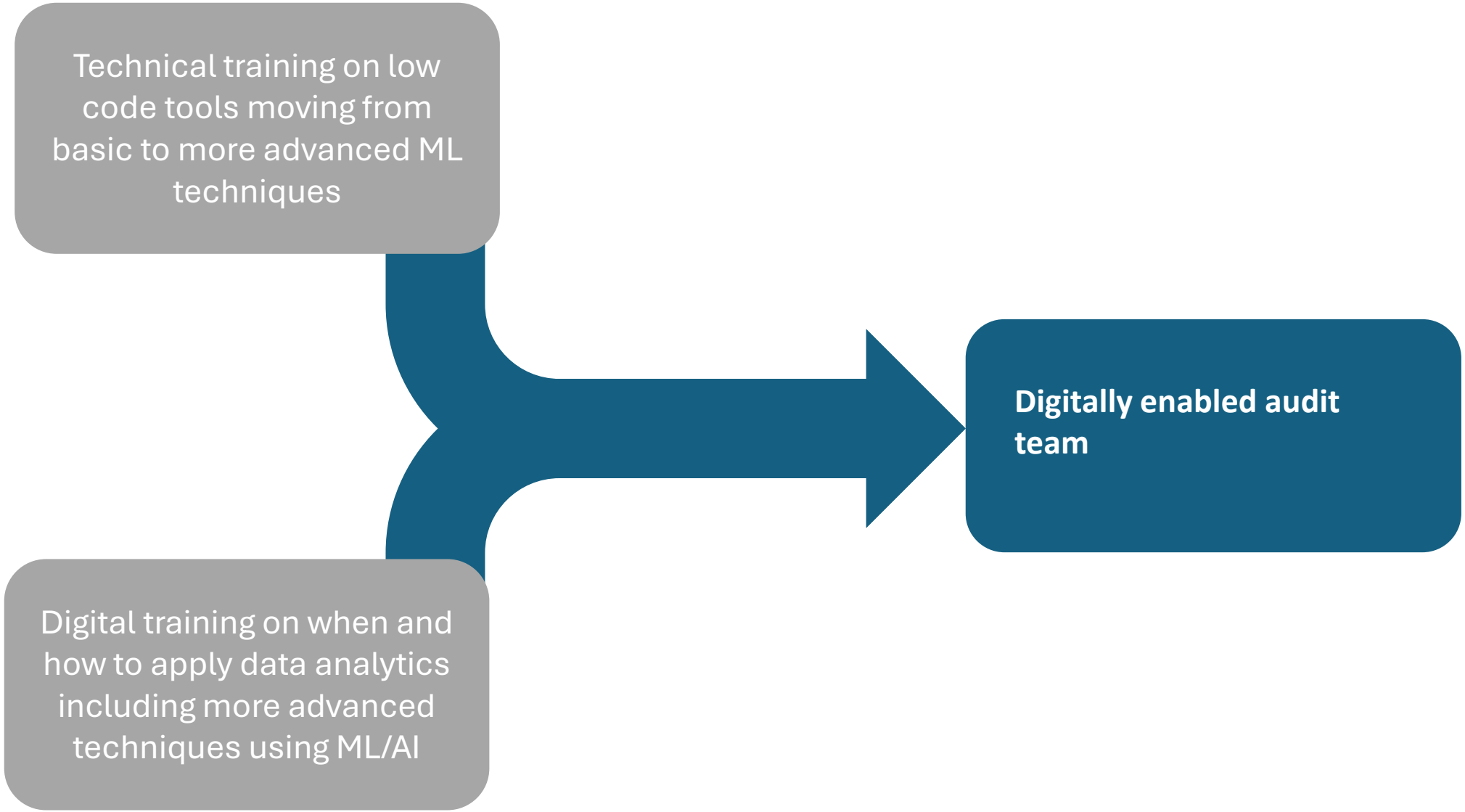


Auditors are data enabled, performing analytics with support on complex solutions from central DA team

Key

Auditors

Training considerations

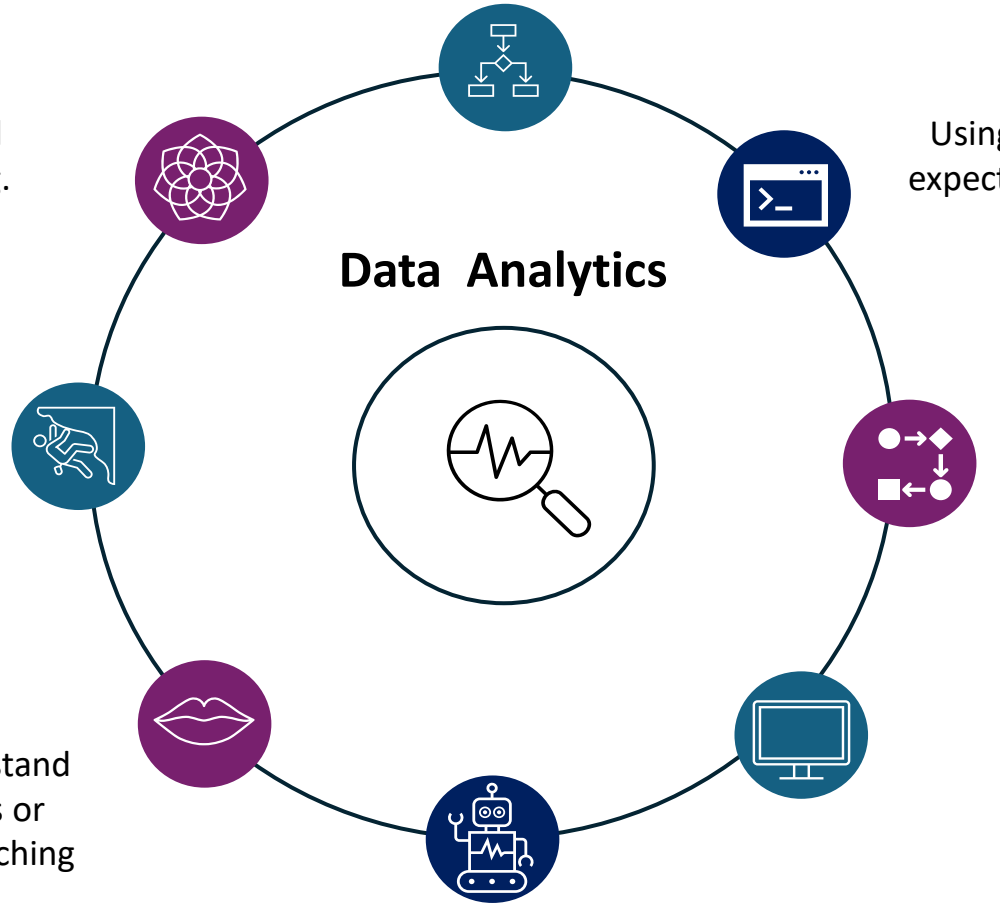


Types of Data Analytics?

Outcomes Testing
Using data to check if key processes are achieving the right outcome e.g. Identify any expenses exceeding approved thresholds

Pattern Mining
Using data to identify unusual patterns for further review e.g. network analysis

Controls Testing
Using data to check controls are operating as expected e.g. check system prevents users from changing critical details



Risk based sample testing
Using analytics to identify higher risk items for further review e.g. high value payments to specific countries

Process Mining
Using event data to identify missing controls or issues within a process e.g. customer onboarding journey

NLP
Leveraging NLP to understand text to identify patterns or exceptions e.g. fuzzy matching

ITGC testing
Using data to test standard IT controls e.g. backups, logging & monitoring, access etc

ML/AI
Using unsupervised e.g. k-means clustering to find outliers or supervised e.g. decision trees to identify potential fraudulent transactions

What are low code tools?

Low code tools are visual development platforms that enable users to build applications with minimal or no coding. They utilize graphical user interfaces (GUIs) and drag-and-drop features to automate parts of the development process, making it accessible to a wider range of users, including those with limited programming experience

Low code tools can be used for more advanced analytics including:

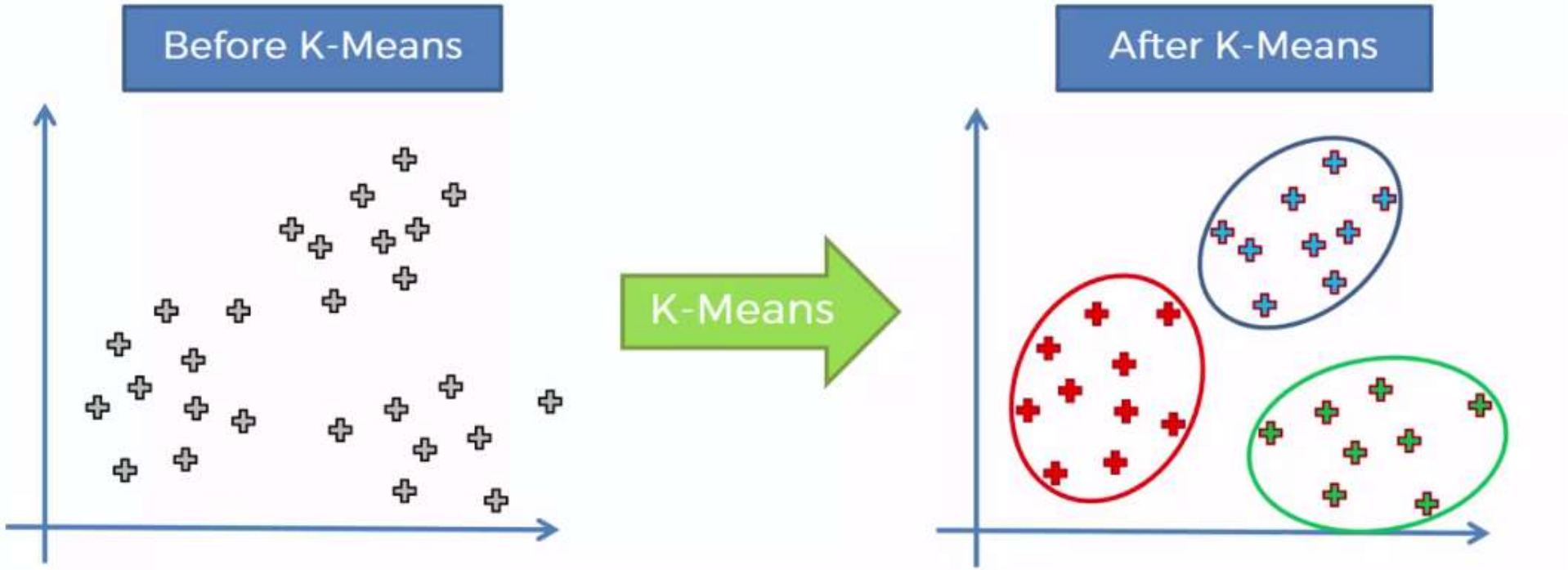
- Unsupervised learning e.g. K-means clustering
- Supervised learning e.g. Linear Regression. Decision trees etc
- Natural Language Processing (NLP) e.g. sentiment analysis

What is K-Means Clustering?

K-Means Clustering is an **unsupervised machine learning algorithm** used to group unlabelled data into clusters based on their similarities.

How K-Means Clustering Works

The algorithm works by partitioning a dataset into k clusters, where k is a predefined number of clusters



*** Image from medium.com**

Knime Example: Using K-Means clustering to look for anomalies in expense claims.

An auditor wants to review expenses for a hardware store and use k-means clustering to identify any potential outliers for further review a sample of the data is shown below.

Id	Travel	Accom	Meals	Entmt.	Days since submission
E1000	549.67	639.79	164.42	85.47	12
E1001	486.17	809.85	167.37	138.01	17
E1002	564.77	739.45	196.92	78.77	29
E1003	652.3	636.43	213.65	113.31	6
E1004	476.58	761.59	211.07	123.24	7
E1005	476.59	711.65	233.09	72.19	14
E1006	657.92	816.24	200.52	98.21	13
E1007	576.74	615.75	258.14	2.76	8
E1008	453.05	660.68	189.41	69.27	10
E1009	554.26	652.95	308.81	92.42	9

KNIME Analytics Platform

File Edit View Node Help

Zoe Adey

KNIME Explorer

My-KNIME-Hub (api.hub.knime.com)
EXAMPLES (knime@api.hub.knime.com)
LOCAL (Local Workspace)
binary
com.knime.features.enterprise.slave_root.cocoa.macosx.x86_64_4.16.1.v202301
com.knime.features.enterprise.slave_root.gtk.linux.x86_64_4.16.1.v20230131140
com.knime.features.enterprise.slave_root.win32.win32.x86_64_4.16.1.v20230131
org.eclipse.equinox.executable_root.cocoa.macosx.aarch64_3.8.1700.v2022050
org.eclipse.equinox.executable_root.cocoa.macosx.x86_64_3.8.1700.v20220509-
org.eclipse.equinox.executable_root.gtk.linux.x86_64_3.8.1700.v20220509-0833
org.eclipse.equinox.executable_root.win32.win32.x86_64_3.8.1700.v20220509-08
org.eclipse.rcp_root_4.24.0.v20220607-0700
org.knime.desktop.product.executable.cocoa.macosx.aarch64_4.7.1.v20230131
org.knime.desktop.product.executable.cocoa.macosx.x86_64_4.7.1.v202301311
org.knime.desktop.product.executable.gtk.linux.x86_64_4.7.1.v202301311353
org.knime.desktop.product.executable.win32.win32.x86_64_4.7.1.v20230131135
org.knime.features.product_root.cocoa.macosx.aarch64_4.7.1.v202301311311
org.knime.features.product_root.cocoa.macosx.x86_64_4.7.1.v202301311311

Workflow Coach

No node recommendations available.

Node Repository

IO
Manipulation
Views
Analytics
DB
Other Data Types
Structured Data
Scripting
Tools & Services
KNIME Labs
Workflow Control
Workflow Abstraction
Social Media
Reporting
Chemistry
Testing

*0: Kmeans Example

Console Error Log Progress

Workspace Log

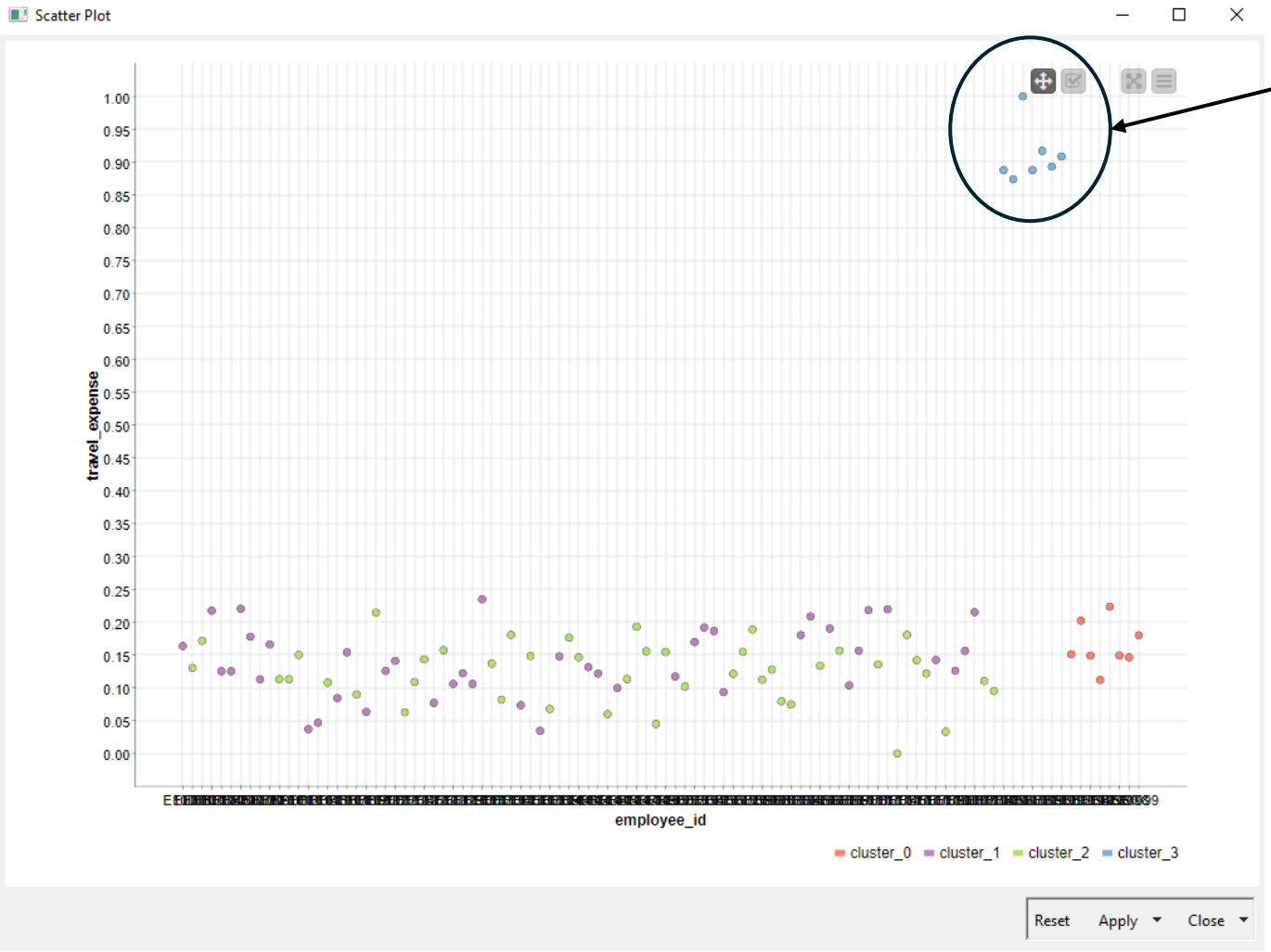
type filter text

Message	Plug-in	Date
Unable to load class 'org.knime.product.rcp.addons.ToolbarFixAddon' from bundle '1542'	org.eclipse.e4.ui.workbe...	13/05/2025, 16:48
Unhandled event loop exception	org.eclipse.ui	13/05/2025, 16:47
	org.eclipse.e4.ui.workbe...	13/05/2025, 16:47
Unhandled event loop exception	org.eclipse.ui	13/05/2025, 16:46

16:58
13/05/2025

Knime Example: Using clustering to look for anomalies in expense claims.

PROCESS



Cluster of high value expenses for different account types, K-means clustering is helping to identify similarities across multiple dimensions.

Selecting an example exception, we can see their expenses are much higher than average in all areas:

ID	Travel	Accom.	Meals	Entmt.
E1068	£1903.29	£2569.41	£716.99	£605.68
Average	£597	£816	£245	£136

What is Sentiment Analysis?

Sentiment analysis, also known as opinion mining, is a technique used to determine the sentiment behind a piece of text, whether it is positive, negative, or neutral. This process involves analysing large volumes of text data to extract meaningful insights that can guide business decisions. Sentiment analysis combines natural language processing (NLP) and machine learning (ML) to achieve this goal.

The cake was disgusting I hated every bite and had to send it back, never coming back very angry.

The cake was ok I might buy it again.

The cake was amazing it was full of chocolate; I enjoyed it so much that I had to take another slice.

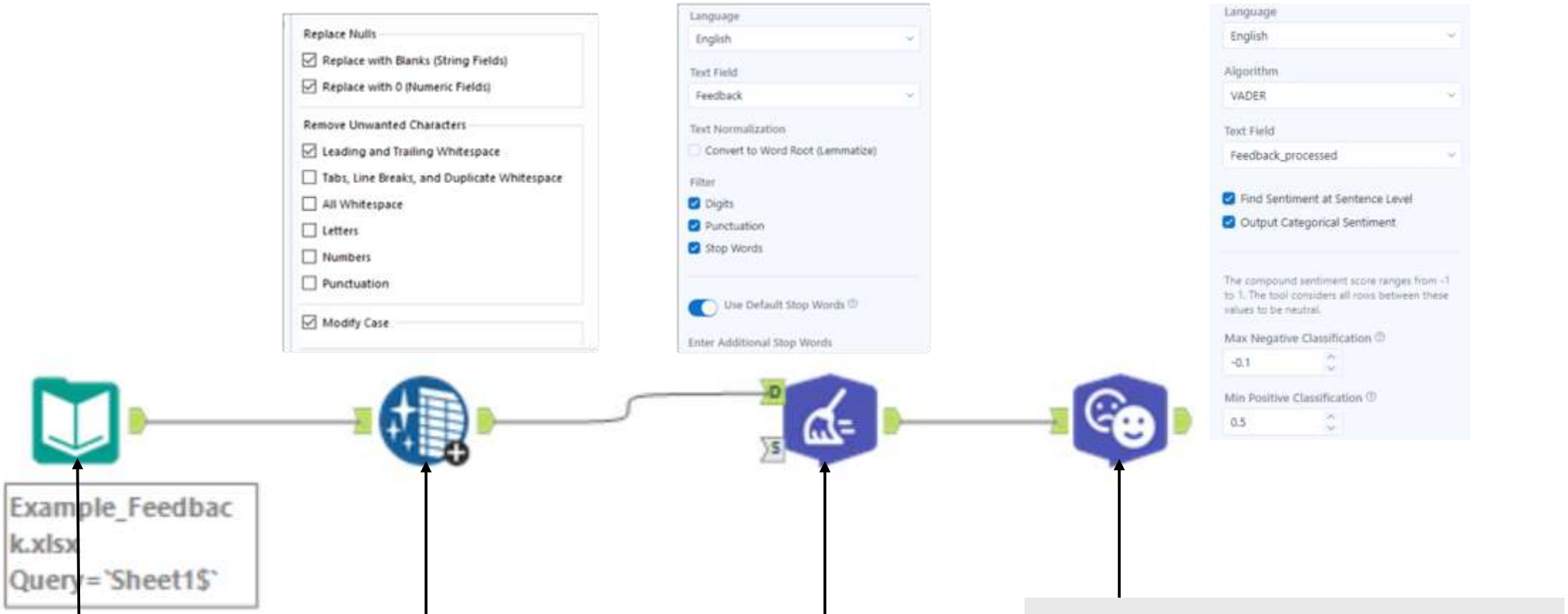
Alteryx Example: Using sentiment analysis to categorise complaints

The auditor now wants to look to see what complaints have been received during the last month. The hardware store collects all feedback through one online form and the auditors wants to focus on the negative feedback. The below is a sample of the data they have received:

Feedback
The staff were incredibly helpful and friendly
Found everything I needed quickly – great store layout
The staff ignored me the whole time I was in the store
Poor product quality – the hammer broke after one use
Really appreciate the new weekend opening hours
Rude cashier and unhelpful service
My loyalty card which I have been using for the last year has now been discontinued and I've lost all my points you did not communicate this I will not be shopping here again
Store was dirty and disorganised
Prices are very high compared to other places
Staff were ok, prices average compared to others
THIS STORE IS TERRIBLE!

Alteryx Example: Using sentiment analysis to categorise complaints

PROCESS



Load data from Excel

Remove white space, blanks and change to upper case

Remove stop words e.g. an, a etc and punctuation/numbers

Using VADER (Valence Aware Dictionary and sEntiment Reasoner) this is a lexicon and rule-based sentiment analysis tool. Configure parameters e.g. raising Max Negative Classification makes the algorithm more sensitive to negative sentiment

Alteryx Example: Using sentiment analysis to categorise complaints

RESULTS

Feedback	Negative	Neutral	Positive	Category
The staff were incredibly helpful and friendly	0	0.233	0.767	positive
Found everything I needed quickly – great store layout	0	0.549	0.451	positive
The staff ignored me the whole time I was in the store	0.434	0.566	0	negative
Poor product quality – the hammer broke after one use	0.596	0.404	0	negative
Really appreciate the new weekend opening hours	0	0.597	0.403	neutral
Rude cashier and unhelpful service	0.5	0.5	0	negative
My loyalty card which I have been using for the last year has now been discontinued and I've lost all my points you did not communicate this I will not be shopping here again	0.195	0.508	0.297	neutral
Store was dirty and disorganised	0.592	0.408	0	negative
Prices are very high compared to other places	0	1	0	neutral
Staff were ok, prices average compared to others	0	0.645	0.355	neutral

Low Code Tools Overview

BENEFITS

01 Workflows are easy to reuse with updated data sets, making low code tools more efficient than Excel

02 Drag and drop functionality makes advanced analytics accessible to none specialists

03 Workflows can be shared easily enabling collaboration across teams

04 Human errors can be reduced as workflows enable reviews

CONSIDERATIONS

01 Workflows can become long and complex if the most efficient solution isnt planned

02 Training is required to ensure users do not become over reliant on the tool's automation and overlook data issues e.g. inconsistencies, null values

03 Training models and understanding the correct ML model to use can be complex and users may not fully understand what has been applied

04 With an increased use of Data Analytics overall there can be challenges obtaining required data.

Thank you