# Planning Application Analysis

## Application Types and Counts

Application Types and their Counts:

* Full planning permission: 605
* Householder planning permission: 490
* Change of use: 84
* Consent under Tree Preservation Orders: 74
* Consent to display an advertisement: 51
* Listed building consent (Alt/Ext): 43
* Pre Application (6 weeks): 34
* Lawful development: Proposed use: 30
* Lawful development: Existing use: 30
* Outline planning permission: Some matters reserved: 22
* Prior approval Part 1 Class A.1(ea): Larger home extension: 18
* Approval of details reserved by a condition: 17
* Non-Material Amendment: 15
* Prior notification: Demolition: 9
* Prior approval Part 3 Class MA: Commercial, business and service uses to dwellinghouses: 8
* Outline planning permission: All matters reserved: 8
* Pre Application (4 weeks): 8
* Approval of reserved matters: 8
* Removal/variation of conditions: 5
* Notification of proposed works to trees in a conservation area: 5
* Hybrid: 5
* Prior approval Part 18 Class A: Development under local or private Acts or Order: 4
* Prior approval Part 14 Class J: Installation or alteration etc of solar equipment on non-domestic premises: 3
* Prior notification: Development by telecoms operators: 3
* Prior approval Part 1 Class AA : Enlargement of a dwellinghouse by construction of additional storeys: 2
* Pre-Application enquiry: 1
* Prior approval Part 20 Class A: New dwellinghouses on detached block of flats: 1
* Planning Performance Agreement: 1
* Permission in Principle: 1

## Cluster Analysis of Invalid Reason Details

What is Cluster Analysis?

Cluster analysis is the process of grouping data points into clusters based on their similarities. The goal is to ensure:

- High intra-cluster similarity: Data points within the same cluster are closely related.

- Low inter-cluster similarity: Data points from different clusters are distinct.

Why is Cluster Analysis Useful Here?

In the context of invalid application reasons:

1. Simplifies Complexity: Clustering reduces the number of distinct reasons into manageable groups for analysis.

2. Identifies Patterns: Repeated issues such as missing documents or incorrect fees are highlighted.

3. Drives Actionable Insights: Groups can inform targeted actions, such as improving application guidelines.

Overview of the Script

This script is designed to analyse invalid planning application reasons using clustering techniques.

Key Steps:

1. Data Loading: Reads a CSV file containing invalid reasons and other details.

2. Theme-Based Clustering:

- Predefined themes (e.g., "Missing Documents") are matched using regular expressions.

- Reasons are assigned to the first matching theme, ensuring no overlap between clusters.

3. Refinement of Themes:

- The "Incorrect Fee" cluster is split into:

- Incorrect Fee - Underpayment: Includes reasons with "insufficient" or "further fee".

- Incorrect Fee - Other: All remaining reasons.

- Clusters are reordered to prioritise more actionable categories.

4. TF-IDF Vectorisation:

- Text data is transformed into numerical features using TF-IDF.

5. Dimensionality Reduction:

- Uses \*\*t-SNE\*\* to reduce high-dimensional data to two dimensions for visualisation.

6. KMeans Clustering:

- Groups reasons into 10 clusters based on their TF-IDF representation.

7. Outputs:

- Word Document: Summarises cluster counts and their descriptions.

- CSV File: Lists grouped reasons under their respective clusters.

- Scatter Plot: Visualises clusters using t-SNE.

Considerations:

1. Theme Accuracy: Ensure that regex patterns adequately capture the themes.

2. Dimensionality Reduction: t-SNE is sensitive to parameters (e.g., perplexity), which may influence the scatter plot.

3. Scalability: For larger datasets, clustering methods and parameters may need adjustments.

This script provides a streamlined approach to analyse and act upon recurring issues in invalid planning applications.

* Missing Reports (427 instances)
* Missing Forms (230 instances)
* Missing Plans (284 instances)
* Missing Details (23 instances)
* Missing Drawings (22 instances)
* Incorrect Fee - Underpayment (43 instances)
* Incorrect Fee - Other (62 instances)
* Validation Checklist (43 instances)
* Other (435 instances)