Business and Local Government
Data Research Centre

Project Title: Waste Strategy 2023
Organization: University of Essex

Data Science and AI Masters Project 2024

Summary Report

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

University of Essex

Author: Anurup Salokhe

Academic Supervisor: Prof Dr Ravi Shekhar

Date: 27-06-2024

Table of Contents

1 Background

The Essex Waste Partnership (EWP) consists of Essex County Council and the 12 district, city, and borough councils in Essex. The EWP is in the process of developing a new joint Waste Strategy for Essex. This strategy will provide a high-level, strategic framework for managing the waste and recycling generated by homes and businesses in the county over the next 30 years.

Essex County Council (ECC) facilitated a public consultation in autumn 2023 on behalf of the EWP. The purpose of the consultation was to gather feedback and views from residents, communities, businesses, and councils regarding the proposals outlined in the strategy.

The online survey was conducted by Essex County Council using Citizen Space. This platform allowed participants to express their agreement or disagreement with each section of the draft strategy. Additionally, respondents had the opportunity to provide feedback on each part of the strategy. Participants were given access to a range of documents online, along with the draft strategy, to help them provide well-informed responses to the consultation.

2 Research Objectives

- Developing certain topics through topic modelling using LDA (Latent Dirichlet Allocation).
- The main objective of this project is to generate Abstractive text summarizations of the comments given by the respondents regarding waste strategies.
- Describe the respondents' opinions on different strategies for improving the future environment.

3 Dataset

The initial step before cleaning the data is collecting it. For this report, we gathered the data from the online survey hosted by Essex County Council using Citizen Space, which allowed respondents to indicate whether they agreed or disagreed with each part of the draft strategy. Respondents were also provided with the opportunity to provide comments on each part of the draft strategy. The collected data typically includes raw text, which needs to be pre-processed.

3.1 Data Pre-processing:

- **Dropping NaN values:** If the missing values are substantial and could significantly affect the analysis, we can choose to drop the rows or columns containing these values.

- **Lower Casing:** Converting text to lowercase is a fundamental step in text preprocessing text analysis. It ensures uniformity and consistency in the text data by eliminating variations due to case differences.

- **Remove Special Characters:** Special characters in text data can introduce noise and affect the performance of analysis models. Removing these characters helps in cleaning the text data, making it more suitable for analysis.

- **Removing stopwords:** Stopwords are common words such as "is", "and", "the", etc., that often do not contribute significantly to the meaning of the text. Removing these words helps in focusing on the more meaningful words. Stopwords can be identified using predefined lists provided by natural language processing (NLP) libraries such as NLTK, SpaCy, or custom lists. Using the NLTK library, we can easily remove stopwords from our text data.

- **Lemmatization:** Lemmatization is the process of converting words to their base or dictionary form (lemma). Unlike stemming, which may simply cut off prefixes or suffixes, lemmatization considers the context and converts the word to its meaningful base form. After lemmatization, it is important to verify that the text data has been properly lemmatized and that words are in their base form.

- **Vectorization:** Vectorization is a crucial step in preprocessing text data. It transforms text into numerical representations that can be fed into machine learning models. By integrating vectorization using Gensim's corpora module into the preprocessing pipeline, we can ensure that the text data is ready for analysis and that the models can effectively learn from the data.

4 Methods/Methodology

4.1 Exploration by Numpy/Pandas

Numpy and Pandas are fundamental tools in the Python ecosystem for data analysis and manipulation:

1. **Numpy**: Primarily used for numerical computations on large, multi-dimensional arrays and matrices. It provides efficient functions for mathematical operations and linear algebra.
2. **Pandas**: Built on top of Numpy, Pandas is used for data manipulation and analysis. It offers data structures like DataFrame (for 2-dimensional data) and Series (for 1-dimensional data), which allow easy indexing, slicing, grouping, and aggregating data.

➢ **Versatility**: Both libraries are versatile and widely used in data science projects due to their rich functionalities and ease of use.
➢ **Data Handling**: Pandas excels in handling structured data, offering powerful tools for cleaning, transforming, and analyzing tabular data, which is often central to research projects.
➢ **Community Support**: Being open-source with a large community, Numpy and Pandas receive continuous updates, bug fixes, and support from developers worldwide.

4.2   Topic Modelling

Topic modeling is a statistical method that helps to reveal the hidden topics present in a set of documents. It is a type of unsupervised learning, meaning it doesn't need labeled data. Instead, it relies on the patterns of word co-occurrence within the documents to identify hidden topics.

Two popular techniques for topic modeling are Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA). Although both aim to reveal hidden semantic patterns in text data, their methods differ.

4.2.1 LDA (Latent Dirichlet Allocation)

LDA (Latent Dirichlet Allocation) is indeed a powerful tool used for topic modelling in Natural Language Processing (NLP). LDA is a Bayesian network, which means it is a generative statistical model that assumes documents are composed of words that help to identify the topics. As such, documents are associated with a set of topics by assigning each word in the document to different topics.

Building LDA Model for topic modelling:

▪ We added the preprocessed data, known as the corpus, to the LDA model. LDA generates probabilities for the words, using which the topics are formed, and eventually, the topics are classified into documents.

▪ Based on these word-topic assignments, LDA refines its understanding of the topics themselves, adjusting the word probabilities within each topic

- Specifying the optimal number of topics can be challenging. We have used number of topics to be 15 because too few might miss important themes, while too many could lead to overly granular or uninterpretable topics.

- Overall, LDA is a valuable tool for uncovering latent thematic structures within a corpus.

4.2.2 Pros and Cons of using the LDA method:

➢ LDA identifies topics and their distribution across documents but it cannot capture the latent semantic structure or reduce dimensionality.

➢ LDA is a generative model that explicitly models the document generation process. However, it is not able to transform the original matrix to capture latent relationships.

➢ LDA represents topics as probability distributions over words but cannot represent documents and terms in a lower-dimensional semantic space.

4.3 Text Summarization:

Summarization is a common task in Natural Language Processing (NLP). The vast amount of new content generated daily by billions of people and their smartphones has resulted in an overwhelming influx of data. With humans only being able to consume a limited amount of information, there is a need to filter out irrelevant information and extract what is truly important. Text summarization serves as a solution to this problem, helping to separate valuable insights from the noise and enabling meaningful actions to be taken.

4.3.1 Types of Summarizations:

1. Extractive:

Extractive summarization methods function in the way that the text is analyzed, sentences are ranked based on their understanding and relevance, and the most important sentences are presented.

This method does not create new words or phrases, it just takes the already existing words and phrases and presents only that. You can imagine this as taking a page of text and marking the most important sentences using a highlighter.

2. Abstractive:

Abstractive summarization involves attempting to understand the overall meaning of the text and conveying that meaning to the reader.

It creates new phrases, arranges them in a coherent manner, and includes the most important information from the original text. Therefore, abstractive summarization techniques are more intricate than extractive summarization techniques and are also more computationally demanding.

4.3.2 Developing Abstractive Text Summarization using Huggingface Transformers:

- We pre-processed the raw data again by performing all the mentioned techniques in Data Pre-processing, and finally, we tokenized the data in the form of sentences for text summarization.

- Using the topic modelling part, we tested all the sentences to determine the best score for a particular topic and assigned all the relevant sentences into a string for a particular topic.

- We have used a pre-trained summarization model into the pipeline from transformers library which allows you to create a pre-configured object to perform a specific NLP task. The pipeline function helps summarize a piece of text into a shorter version that captures the main points.

- By using the pipeline function, we summarized the sentences to obtain a short summary.

- To further summarize the short summary into one action point, we have used an API Call method. The open llama 7B model from the HuggingFace helps call the application without using API tokens.

- We have developed a prompt by passing the template to summarize the text to one action point and produced the result.

4.3.3 Benefits of using an API for abstractive text summarization:

- ➢ Many text summarization libraries like Huggingface are open source and do not require an API key.
- ➢ Many text summarization APIs offer access to pre-trained models that can be used without additional training.
- ➢ APIs can handle a large volume of requests, which is advantageous for performing text summarization on many documents.

5 Analysis and Findings

By using the Abstract text summarization and API call methods, we can find the sentences, which are the comments given by the respondents, generated into one action point. The results obtained are categorized into certain topics, which illustrate a particular strategy. The outcomes of the key findings are given below.

5.1 Summary of Key Findings

Response summary

A total of 4,545 responses to the consultation were received. This included 4,224 responses to the full survey and 321 responses to the Easy Read survey. 'Not Answered' was the most common response, accounting for 99.69% of participants. Additionally, notable responses were 'Paper copy' with 0.31%.

```
Column Question :  Is this consultation
being completed over the phone or
uploaded as a paper copy? - Phone/paper

Column Responses :  ['Not Answered' 'Not
Answered' 'Not Answered' ... 'Paper
copy' 'Paper copy' 'Paper copy']
```


Percentage of Unique Responses

Not Answered 99.69%   0.31% Paper copy
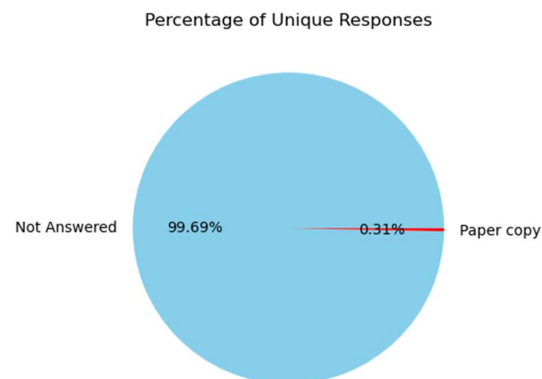
Vision

Respondents were asked the extent to which they agreed or disagreed with the vision statement: We aspire to be a zero-waste county. By working together we will reduce waste, protect the environment, and conserve resources.
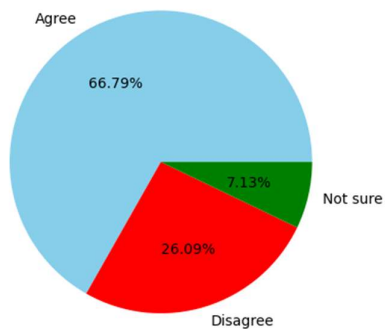
Figure 1 – Vision summary

```
Column Question :  To what extent do you agree or disagree with this vision
statement for the Waste Strategy for Essex? - vision
Column Responses :  ['Agree' 'Agree' 'Agree' ... 'Not sure' 'Agree' 'Agree']
```
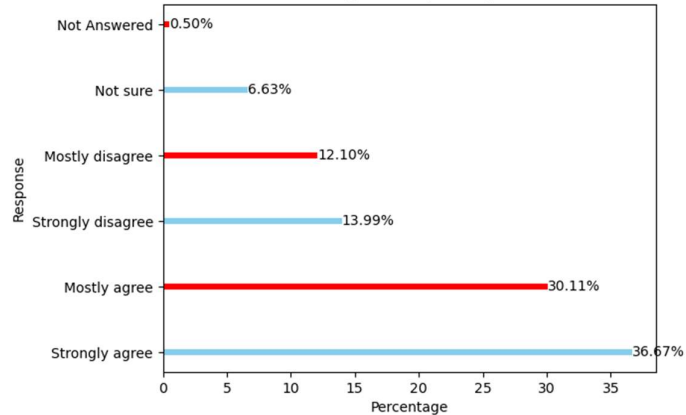
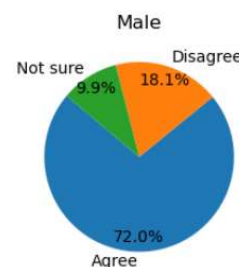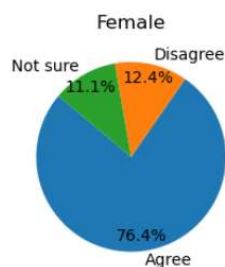Summary Overview:

Percentage of Unique Responses

Percentage of Unique Responses

- Regarding the question, 'Agree' was the most common response, accounting for 66.79% of participants.
- Further, we observed these responses: 'Disagree' with 26.09%; 'Not sure' with 7.13%.

Demographic differences

Gender Demographic:

Female

Male

Summary by Gender:
- In the total of 2424, 'Female' group, 12.4% disagree, however 76.4% agreed and 11.1% are not sure.
- Out of total 1414, 'Male' category,72.0% of people in the Male category agree with the statement, few 18.1% disagree and rest were not sure .

Conclusion:
Through the gender response study, it is clear that females has almost similar but showed a slightly higher level of agreement compared to males, while males had a slightly higher level of disagreement, and Additionally, other groups had a moderate level of priorities and uncertainty compared to both females and males.
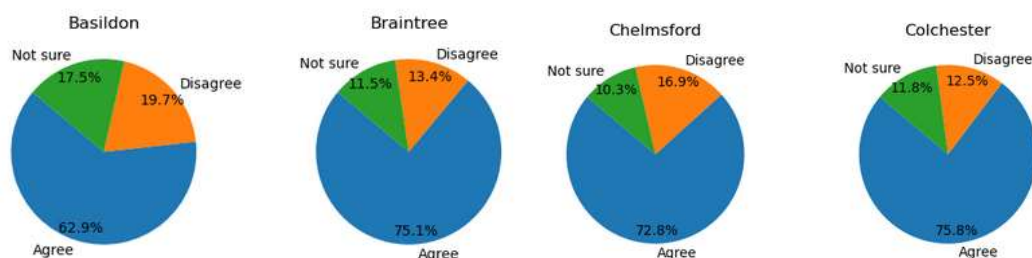
Age Demographic:

Percentage by Age Group

16-24
- Disagree 23.8%
- Not sure 0.0%
- Agree 76.2%

25-34
- Disagree 22.5%
- Not sure 6.7%
- Agree 70.8%

35-44
- Disagree 28.8%
- Not sure 6.7%
- Agree 64.5%

45-54
- Disagree 28.4%
- Not sure 6.0%
- Agree 65.6%

55-64
- Disagree 25.3%
- Not sure 5.9%
- Agree 68.8%

65+
- Disagree 22.8%
- Not sure 7.2%
- Agree 70.0%
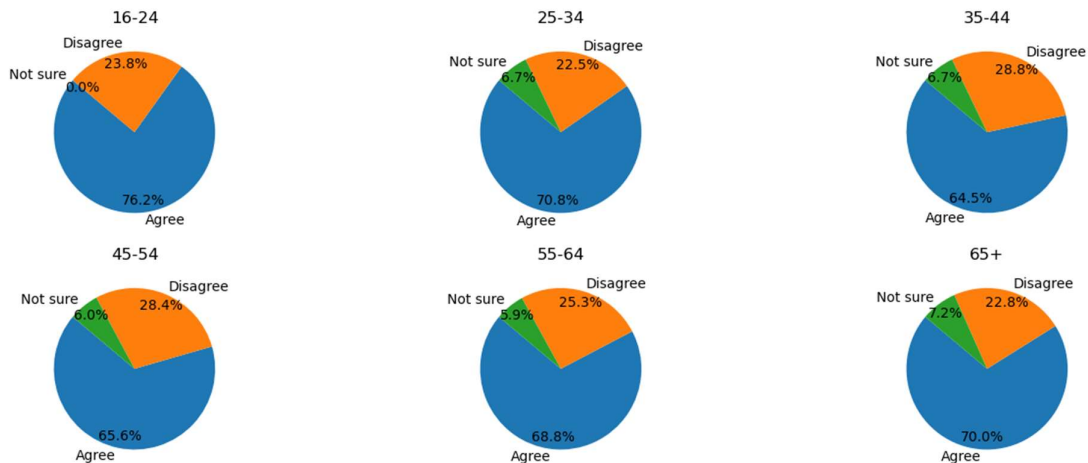
Summary by Age Group:
- In the 1054 people, '55-64' category, 73.4% of participants agree, 16.3% disagree and 10.2% are not sure.
- In the 1370 people, '65+' category, 76.6% of participants agree, 13.4% disagree and 10.1% are not sure.

Conclusion for Age Group:

The data show that candidates in 45-54 showed higher levels of agreement compared to 35-44 and Here 65+ showed higher levels of agreement compared to those aged 55-64, while while those aged 35-44 exhibited higher levels of disagreement and while those aged 55-64 exhibited higher levels of disagreement. Every age group exhibited moderate uncertainty levels.

Local differences

Geographical Demographic:

Basildon
- Not sure 17.5%
- Disagree 19.7%
- Agree 62.9%

Braintree
- Disagree 13.4%
- Not sure 11.5%
- Agree 75.1%

Chelmsford
- Disagree 16.9%
- Not sure 10.3%
- Agree 72.8%

Colchester
- Disagree 12.5%
- Not sure 11.8%
- Agree 75.8%

Summary by Essex Districts:
- out of 630 people in 'Basildon' district, the total of 19.7% of Basildon respondents disagree, 62.9% of respondents agree and unsure were 17.5%

- out of 691 people in 'Chelmsford' district, 72.8% of people in the Chelmsford district agree with the statement, few 16.9% disagree and rest were not sure .
- out of 578 people in 'Colchester' district, 75.8% of people in the Colchester district agree with the statement, few 12.5% disagree and rest were not sure .

The vision provides a clear basis for making future decisions regarding waste management.

- Establish firm deadlines for waste initiatives despite potential resistance and the challenges of changing resident behaviors.
- Essex Council should make recycling and waste disposal easier for residents to increase support for achieving zero waste.
- Develop alternatives to plastic for manufacturing to address the widespread presence of microplastics in the environment.
- I agree with the vision, but practicalities will likely require flexibility.

Targets

The draft strategy outlines targets that must be met to align with the national waste targets set by the government.

- EFDC deploys three teams to assess and address issues with fly-tipping and uncollected waste, including hazardous electronic items.
- Implement effective policies and reduce bureaucracy to prevent waste and illegal dumping, thereby safeguarding the planet by 2050.
- Councils should cease promoting recycling separation among residents, relying instead on specialized companies.
- Implement more convenient waste disposal options to reduce travel, time, and costs for residents.

Ambitions

Whilst the targets are the minimum that need to be achieved and are in line with the government's national targets, the draft strategy includes ambitions that aim to deliver greater change and impact more quickly.

- Support increasing the 10% waste reduction target or shortening the timeframe to achieve it by 2030.
- Advocate for tangible political action to achieve zero waste to landfill.
- Establish ambitious, long-term public goals with clear, measurable milestones.
- Review and adjust the timeframe and focus of the profit plan to ensure realistic goals and ethical practices.
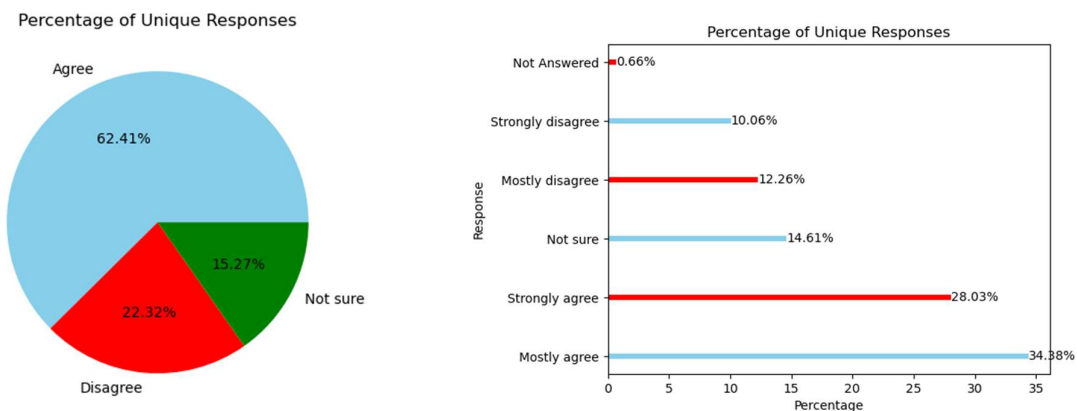
Circular Economy

Figure 2 – Move to a circular economy summary

The EWP believes that the most effective way to minimize the environmental impact of waste is by adopting a circular economy, conserving and utilizing finite resources efficiently.

```
Column Question :  To what extent do you agree or disagree with this priority?
- priority
Column Responses :  ['Agree' 'Agree' 'Agree' ... 'Not sure' 'Not sure'
'Agree']
```


'Agree' was the most common response, accounting for 62.41% of participants. Other significant responses include 'Disagree' with 22.32%; 'Not sure' with 15.27%.

- Develop circular economy practices that secure funding, gain resident support, and help small businesses cut costs to stay profitable.
- Increase transparency and local management in recycling to support the circular economy.
- ECC needs to improve the item restriction process at recycling centers as part of their waste management duties.
- Invest in technology to convert heat from Essex landfill fires into energy.

Recovery

Figure 3 – Recovery summary

```
Column Question :  To what extent do you agree or disagree that the EWP
should reduce the use of landfill? - landfill
Column Responses :  ['Agree' 'Agree' 'Agree' ... 'Not sure' 'Agree' 'Agree']
```

Percentage of Unique Responses
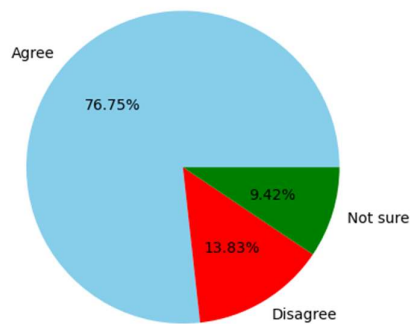
Percentage of Unique Responses
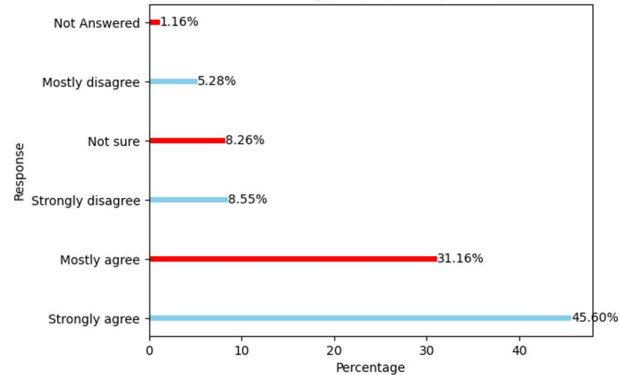
- 'Agree' was the most common response, accounting for 76.75% of participants. Other significant responses include 'Disagree' with 13.83%; 'Not sure' with 9.42%.
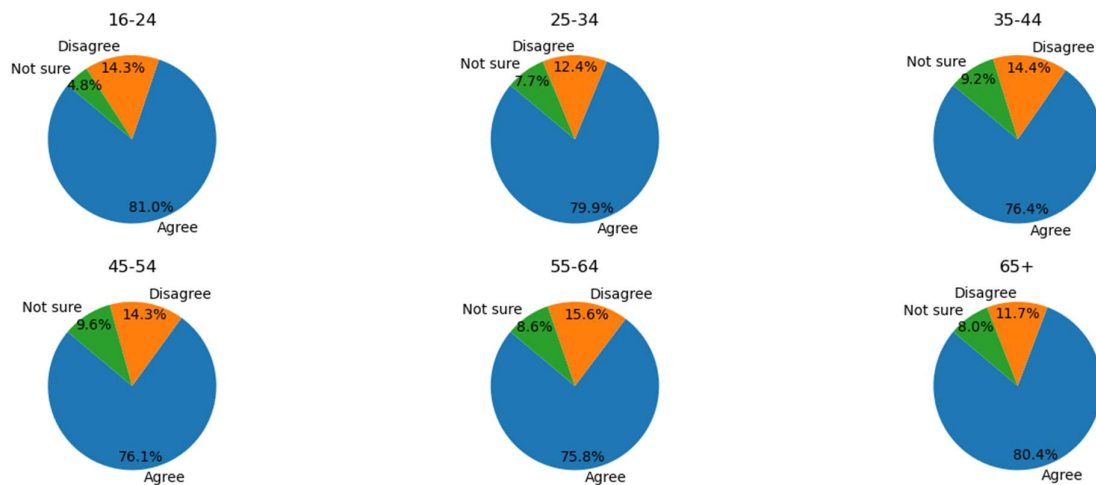
Percentage by Gender
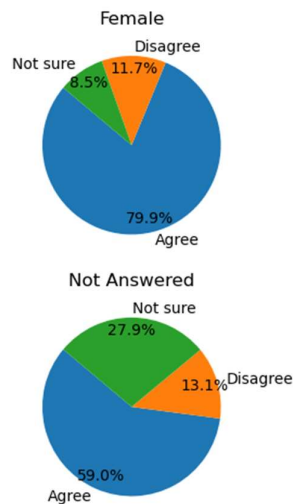
- Gender response data highlight females has almost similar but showed a slightly higher level of agreement compared to males, while males had a slightly higher level of disagreement, and additionally, other groups had a moderate level of priorities and uncertainty compared to both females and males.

Percentage by Age Group

- Our analysis of age-based responses indicates participants aged 35-44 showed higher levels of agreement with respect to 45-54 and Here 65+ showed higher levels of agreement compared to those aged 55-64, while those aged 35-44 exhibited higher levels of disagreement, and while those aged 55-64 exhibited higher levels of disagreement. Uncertainty remained moderate for all age groups.

The concept of recovering wasted resources should be a long-term goal that paves the way for cleaner production.

- Address illegal dumping by reallocating funds from garden waste disposal to glass recycling and revising tip appointment policies.
- Increase recycling and waste management efficiency in London to mitigate the environmental impact of high clothing consumption in the UK.
- Implement fines or incentives for businesses and residents to promote recycling and reuse.
- Implement recycling bins and improve public education on waste sorting and processing, similar to practices in London.

Collaboration

Figure 5 – Collaborate and innovate summary

To what extent do you agree or disagree that collaboration and innovation should be achieved through…

```
Column Question :  To what extent do you agree or disagree that Collaboration
and Innovation should be achieved through the following means: -
Collaboration - Work to reduce the carbon impact of waste operations by
```

increasing use of alternative fuels for our vehicles and equipment, and making waste transport routes as efficient as possible.
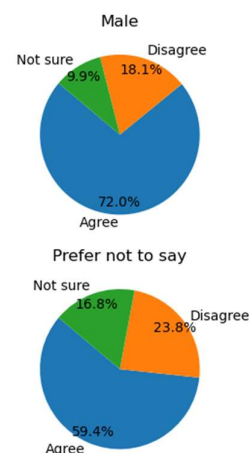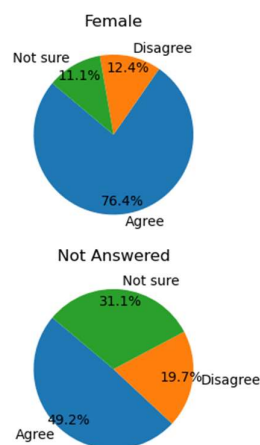
```
Column Responses : ['Agree' 'Agree' 'Agree' ... 'Not sure' 'Agree' 'Agree']
```


Percentage of Unique Responses
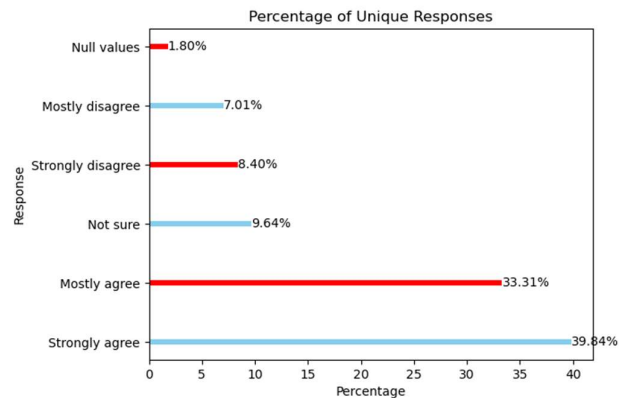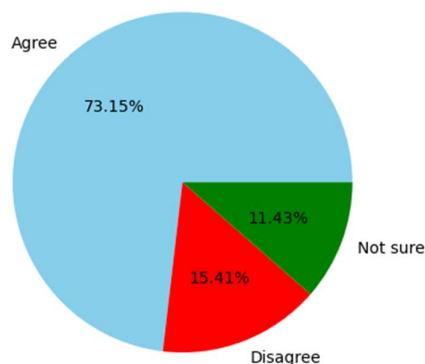
- The most frequent response was 'Agree', making up 73.15% of the total responses. Further, we observed these responses: 'Disagree' with 15.41%; 'Not sure' with 11.43%.
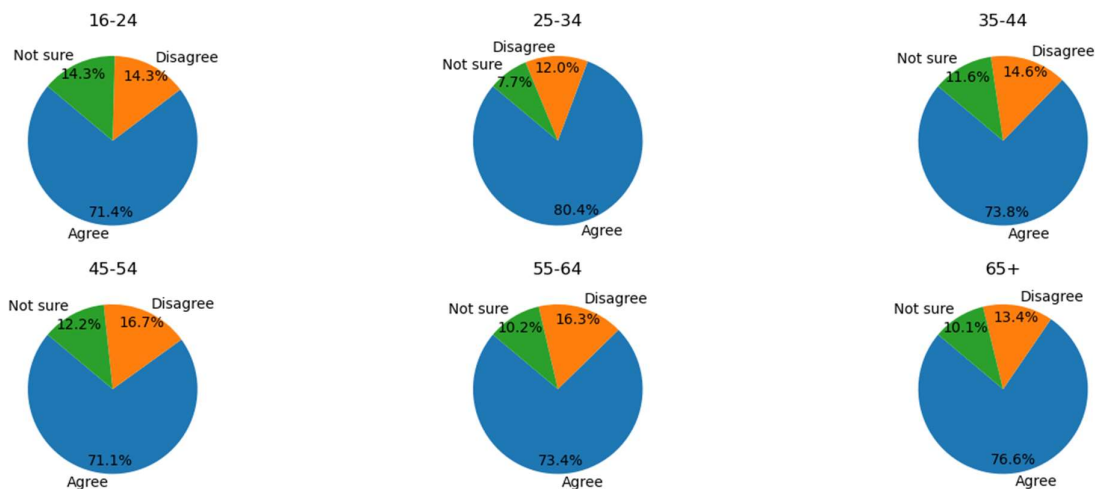

Percentage by Gender

- In the analysis of gender responses, females has almost similar but showed a slightly higher level of agreement compared to males, while males had a slightly higher level of disagreement. Additionally, other groups had a moderate level of priorities and uncertainty compared to both females and males.

- An analysis of the age groups highlights participants aged 35-44 showed higher levels of agreement with respect to 45-54 and Here 65+ showed higher levels of agreement compared to those aged 55-64, with while those aged 45-54 exhibited higher levels of disagreement and while those aged 55-64 exhibited higher levels of disagreement. Moderate uncertainty was present across all groups.

The EWP aims to innovate and collaborate with each other, governments, businesses, and institutions to establish a more sustainable waste management system.

- Expand separate waste collections per household.
- Optimize recycling infrastructure and transition waste vehicles to hydrogen fuel cell technology to reduce carbon emissions.
- Support volunteer-led reuse and recycling shops by encouraging retired individuals to contribute their time and efforts.
- Mandate glass recycling for all councils and encourage the use of hydrogen pellets for cars to address electricity shortages during cold weather.

Research, Planning and Execution

The EWP proposes to comprehensively review this strategy every five years to ensure alignment with any changes in national policy and legislation, trends in waste generation, and the development of new approaches and technologies.
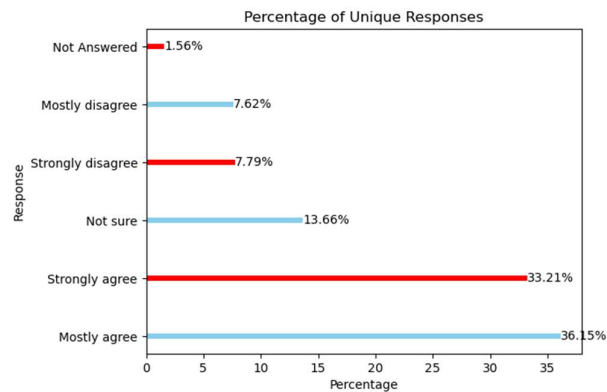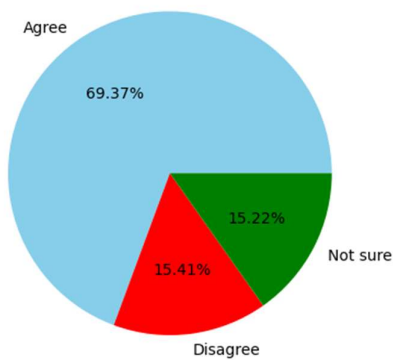
Figure 7 – Research, planning and performance monitoring summary

```
Column Question :  To what extent do you agree or disagree with this approach
to research, planning and performance monitoring?   - agree
Column Responses :  ['Agree' 'Agree' 'Agree' ... 'Not sure' 'Agree' 'Agree']
```
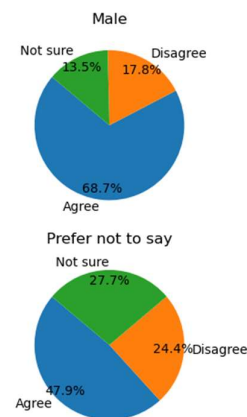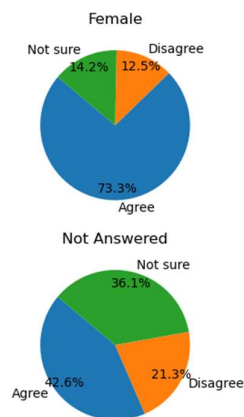
Percentage of Unique Responses

- Regarding the question, 'Agree' was the most common response, accounting for 69.37% of participants. Additionally, notable responses were 'Disagree' with 15.41%; 'Not sure' with 15.22%.
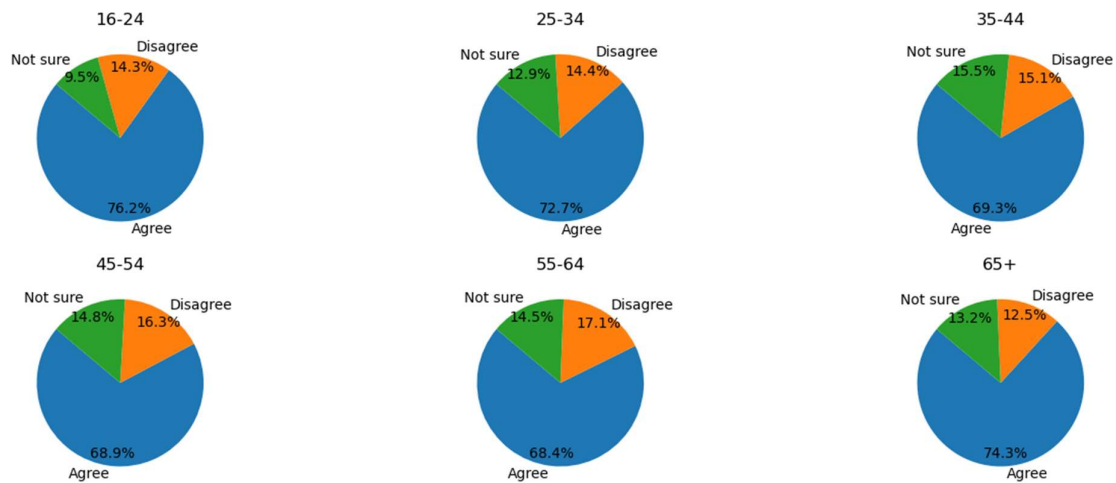
Percentage by Gender

- Through the gender response study, it is clear that females has almost similar but showed a higher level of agreement compared to males, while males had a slightly higher level of disagreement, and Additionally, other groups had a moderate level of priorities and uncertainty compared to both females and males.

Percentage by Age Group

- Analyzing the responses by age groups, we observed participants aged 35-44 showed higher levels of agreement with respect to 45-54 and Here 65+ showed higher levels of agreement compared to those aged 55-64, and also while those aged 45-54 exhibited higher levels of disagreement and while those aged 55-64 exhibited higher levels of disagreement. Each age group showed moderate levels of uncertainty.
- Implement continuous, active review processes to provide timely feedback and ensure progress.
- Improve waste disposal efficiency by switching to quarterly reporting for better accountability.
- Focus council spending on establishing local recycling centres.
- Prioritize frequent reviews for meaningful progress over target achievement.

Educate and Engage

The EWP proposes to Engage with residents to gather feedback and provide information and initiatives that promote changes in attitudes and behaviors to minimize waste and increase recycling.
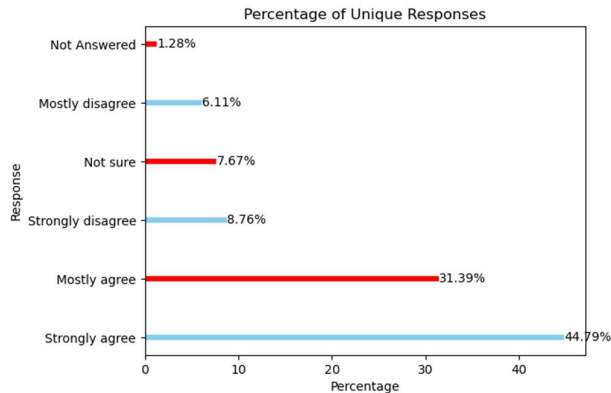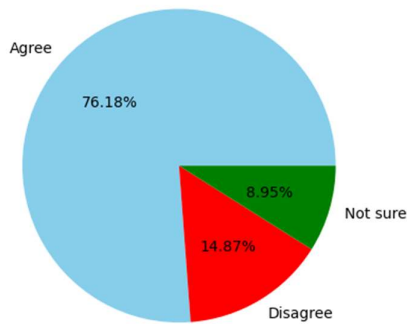
Figure 6 – Educate and engage summary

```
Column Question :  To what extent do you agree or disagree with this priority?

Column Responses :  ['Agree' 'Agree' 'Agree' ... 'Not sure' 'Agree' 'Agree']
```
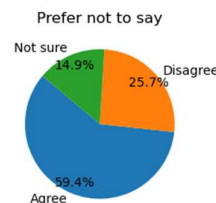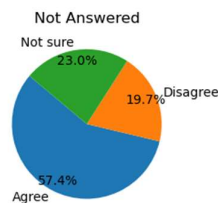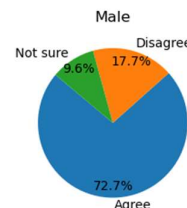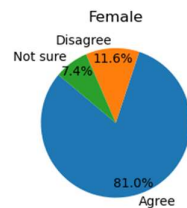
Percentage of Unique Responses

Percentage of Unique Responses

- Regarding the question, 'Agree' was the most common response, accounting for 76.18% of participants. Further, we observed these responses: 'Disagree' with 14.87%; 'Not sure' with 8.95%.
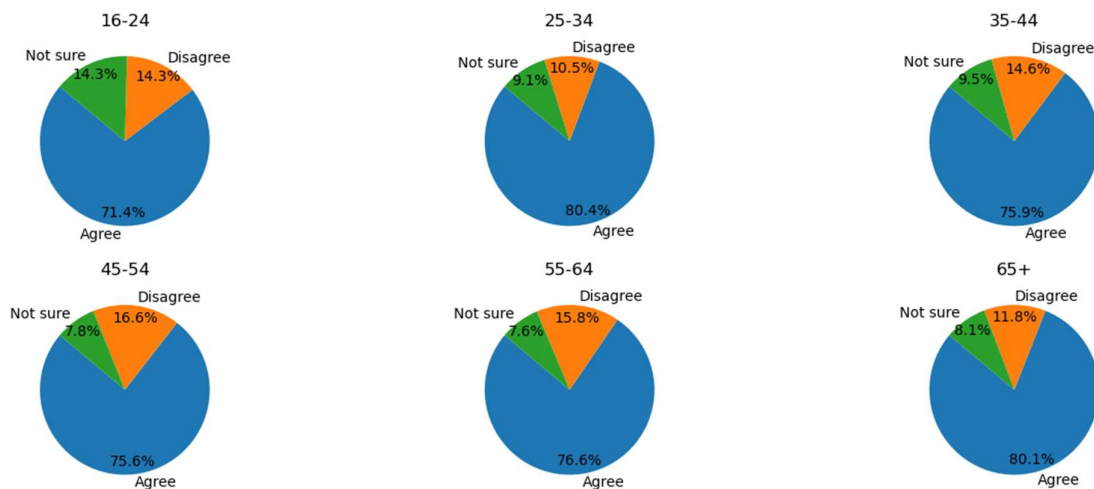
Percentage by Gender

- Examining the responses, we see that females has almost similar but showed a slightly higher level of agreement compared to males, while males had a slightly higher level of disagreement, with Additionally, other groups had a moderate level of priorities and uncertainty compared to both females and males.

Percentage by Age Group

- Based on the age group data, participants aged 35-44 showed higher levels of agreement with respect to 45-54, Here 65+ showed higher levels of agreement compared to those aged 55-64, while those aged 45-54 exhibited higher levels of disagreement, and while those aged 55-64 exhibited higher levels of disagreement. Uncertainty was moderate across the age spectrum.
- Promote proper plastic disposal to prevent littering and contamination.
- Engage with residents to gather information, while implementing fines and incentives for compliance.
- Politicians should prioritize education by supporting local focus groups and community initiatives.
- Educate disabled people on recycling and enforce compliance.

6 Conclusions and Next Steps

6.1 Conclusion Summary

Some of the comments were proposed to advocate for ambitious waste reduction policies with specific, measurable goals to achieve zero waste to landfill by 2030 and improve waste management in Essex by enhancing recycling transparency, refining item restrictions at recycling centers, and investing in technology to convert landfill fire heat into energy.

There is a preference for achieving ambitions and targets sooner. Some believe elements of the strategy, especially zero waste, are unachievable or too ambitious, leading to disagreement, though this is a minority view.

Many respondents suggested that Waste collection and recycling services must be convenient and user-friendly to meet targets and goals. Some people find it difficult

to understand certain concepts in the recovery priority, especially for the food waste treatment.

Education and support for residents with their waste and recycling is considered important. This should also involve engaging with residents and listening to their feedback.

Respondents remarked to expand and optimize waste management and recycling systems, incorporating volunteer support and hydrogen technology to reduce carbon emissions and promote proper plastic disposal through community engagement, education, and enforcement.

6.2   Future steps to be taken

Implementing regular reviews and quarterly reporting to enhance waste disposal efficiency, establish local recycling facilities, and ensure timely feedback and progress are necessary.

Businesses need to focus on making additional efforts, especially when it comes to reducing packaging. Services need to be convenient and easy to use, with frequent collections, no charge for garden waste, increased recycling materials, and convenient access to recycling centres. We need more items that are easy or inexpensive to repair. It is important to change people's mindsets and address the throwaway culture.

Business and Local Government Data Research Centre
Parkside 2C
Knowledge Gateway
University of Essex
Wivenhoe Park
Colchester CO4 3SQ

T: 01206 873859
E: BLGDataResearch@essex.ac.uk
W: www.BLGdataresearch.org
@BLGDataResearch