1. How was data collected?
2. Briefly, state relevant modules and toolkits (separate file for sparql queries)
3. Describe how the model was implemented
4. Summarize results in appropriate ways, including visualizations
5. Critically evaluate the models:
   1. Adequacy of the model (what are the reliant assumptions for the model, do they hold; are there factors which might invalidate conclusions)
   2. comparison with external data/published social/political science research on the subject
   3. What conclusions can be drawn from the model

A text on a white background

Description automatically generated

**Q1**. To what extend do Members of Parliament (MPs) tend to ask questions that directly reference their own constituency or a location in it? You should answer this question by identifying named entities that refer to places or identifiable geographical features (e.g. “Dartford Crossing”, “Reading Gaol”, etc.) in asked questions, and determining whether or not these are located in the MP’s constituency using data from Wikidata. You are free to choose any reasonable method in doing so – even if doing so will result in some false negatives – (e.g. relying on Wikidata property P131 “located in the administrative territorial entity”), however you should estimate how reliable you believe your chosen approach is. (N.B. it is not expected or required that your approach will result in 100% accuracy.)

* Identify named entities of regions within specific constituencies via wikidata. (cross reference with google maps?)
* Gather questions asked by MPs, categorised by constituency. (SPARQL)
* Add the question’s constituency reference (i.e. what constituency it is regarding) (COMBINED WIKIDATA AND SPARQL)
* Get data into QGIS, and visualise frequency of constituency specific questions in proportion to the total questions asked by the mp.

<https://medium.com/@langsamu/api-parliament-uk-7b87597019a4>

<https://github.com/ukparliament/Query/blob/master/Parliament.Data.Api.FixedQuery/Sparql/person_current_constituency.sparql>

<https://guides.statistics.gov.scot/article/22-querying-data-with-sparql>

<https://pythonprogramming.net/named-entity-recognition-stanford-ner-tagger/>

<https://stanfordnlp.github.io/stanza/tutorials.html>

<https://towardsdatascience.com/natural-language-processing-for-fuzzy-string-matching-with-python-6632b7824c49>

<https://www.investopedia.com/terms/c/central_limit_theorem.asp#:~:text=Key%20Takeaways-,The%20central%20limit%20theorem%20(CLT)%20states%20that%20the%20distribution%20of,for%20the%20CLT%20to%20hold>.

<https://medium.datadriveninvestor.com/an-attempt-to-extract-geo-location-from-text-c76cb6bd49d4>

see geodict and geotext, mordecai3

<https://www.parliament.uk/globalassets/documents/commons-information-office/p01.pdf>

<https://www.parliament.uk/globalassets/documents/commons-information-office/Brief-Guides/Parliamentary-Questions.pdf>

PREFIX : <https://id.parliament.uk/schema/>

PREFIX id: <https://id.parliament.uk/>

SELECT \*

WHERE {

?question :writtenQuestionIndexingAndSearchUin ?qnum ;

:questionText ?text ;

:questionAskedAt ?date .

?person :askingPersonHasQuestion ?question ;

:personGivenName ?name ;

:personFamilyName ?surname .

OPTIONAL {

?person :memberHasParliamentaryIncumbency ?seatIncumbency .

?seatIncumbency a :SeatIncumbency .

FILTER NOT EXISTS { ?seatIncumbency a :PastParliamentaryIncumbency . }

?seatIncumbency :seatIncumbencyHasHouseSeat ?seat .

?seat :houseSeatHasConstituencyGroup ?constituency .

?seat :houseSeatHasHouse ?house .

?house :houseName ?houseName .

?seatIncumbency :parliamentaryIncumbencyStartDate ?seatIncumbencyStartDate .

?constituency :constituencyGroupName ?constituencyName .

?constituency :constituencyGroupStartDate ?constituencyStartDate .

}

FILTER (?date >= "2023-01-01+00:00"^^xsd:dateTime && ?date < "2023-10-01+00:00"^^xsd:dateTime)

}

LIMIT 50LIMIT 10

**CONSTITUENCIES**

**WIKIDATA**

[tutorial](https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial)

[wikidata query](https://query.wikidata.org/#SELECT%20%3Fplace%20%3Fconstituency%20%3FplaceLabel%20%3FconstituencyLabel%0AWHERE%20%7B%0A%20%20%3Fplace%20wdt%3AP17%20wd%3AQ145%3B%0A%20%20%20%20%20%20%20%20%20rdfs%3Alabel%20%3FplaceLabel%20%3B%0A%20%20%20%20%20%20%20%20%20wdt%3AP131%20%3Fconstituency%20%3B%0A%20%20%20%20%20%20%20%20%20rdfs%3Alabel%20%3FconstituencyLabel.%0A%20%20FILTER%28LANG%28%3FplaceLabel%29%20%3D%20%22en%22%29.%0A%20%20FILTER%28LANG%28%3FconstituencyLabel%29%20%3D%20%22en%22%29.%0A%7D%0A%0A%20%20%20%20%20%20%20%20%20)

[Search](https://www.wikidata.org/w/index.php?search=a&title=Special:Search&profile=advanced&fulltext=1&ns0=1&ns120=1)

SELECT ?place ?constituency ?placeLabel ?constituencyLabel

WHERE {

?place wdt:P17 wd:Q145;

rdfs:label ?placeLabel ;

wdt:P131 ?constituency ;

rdfs:label ?constituencyLabel.

FILTER(LANG(?placeLabel) = "en").

FILTER(LANG(?constituencyLabel) = "en").

}

**Q2**. By applying LDA topic modeling and analyzing the results, what (if any) identifiable regional differences are there in the types of questions asked – e.g. do MPs representing, say, constituencies located in the North of England tend to ask more questions about certain topics than those in Southeast England? In answering this question, you should start by aggregating data into regions larger than an electoral district, such as those denoted by the property “region of England” https://www.wikidata.org/wiki/Q48091 (For simplicity, you may treat Scotland and Northern Ireland as two separate regions without further subdivisions, or alternatively use any reasonable administrative subdivisions for these regions as you see fit). Discuss the assumptions and limitations of your approach and analysis.

<https://aclanthology.org/E17-2069/>

<https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2>

<https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0>

<https://geoportal.statistics.gov.uk/>

<https://neptune.ai/blog/pyldavis-topic-modelling-exploration-tool-that-every-nlp-data-scientist-should-know>

<https://www.mygreatlearning.com/blog/understanding-latent-dirichlet-allocation/>

<https://medium.com/@lettier/how-does-lda-work-ill-explain-using-emoji-108abf40fa7d>

<https://towardsdatascience.com/6-tips-to-optimize-an-nlp-topic-model-for-interpretability-20742f3047e2>

<https://nicharuc.github.io/topic_modeling/>