McDonalds Review Sentiment (NLP)

Abstract:

Executive summary McDonald's is the **world's largest chain of fast-food restaurants**, serving around 68 million customers daily in 119 countries across more than 36,000 outlets. Founded in the United States in 1940, A McDonald's restaurant is operated by a franchisee, an affiliate, or the corporation itself. In this project we have done a sentiment analysis of negative McDonald's reviews. Contributors were given reviews culled from low-rated McDonald's from random metro areas and asked to classify why the locations received low reviews.

Design:

This project is one of the T5 Data Science Boot Camp requirements. Data provided by data. World. In this module we will be laying the foundation for our analysis by processing and exploring a large amount of data and preprocessing it by using text processing techniques and apply NLP technique on it. The dataset contains negative McDonald's reviews. *Get the data* here.

Understanding the dataset:

The dataset contains 1525 observations of 10 variables:

1. **unit_id**: id of record
2. **golden**: value *FALSE*
3. **unit_state**: value *finalized*
4. **trusted_judgments**: value *3*
5. **last_judgment_at**: time. Example *2/21/15 0:36*
6. **policies_violated**: the type of policies, violated. Example: *RudeService
7. **policies_violated.confidence**: the confidence of policies, violated. Example: *1.00.66670.6667*
8. **city**: City name
9. **policies_violated_gold**: value *NA*
10. **review**: review detail

Algorithms

- Data Collection
- Data Preprocessing (Very Important Step)
- Data Exploration and Visualization
- Model building (Of course the interesting part!!)
- Model Evaluation

Tools:

NLTK: Natural Language Toolkit, one of the leading tools for NLP, renders a whole set of programs and libraries to execute statistical and symbolic analysis in Python. This tool helps in separating a piece of text into smaller units (tokenization). Through this tool, you can recognize named entities and can tag some text. It is the leading tool of NLP and is easy to use.

SpaCy: This tool is a successor of NLTK. It comes with pre-trained statistical models and word vectors. It is a library created for use in Python and Cython. It supports tokenization for 49+ languages.it enables to break the text into semantic segments like articles, words, punctuation. It can be used for named entity recognition (NER) with pre-trained classes, recognizing dependencies in sentences. It provides the fastest and most accurate syntactic analysis than any NLP library.

Text Blob: This tool was designed based on NLTK. For the probationer, it is the best option to understand the complexities of NLP and designing prototypes for their projects. The tool enables sentiment analysis, tokenization, translation, phrase extraction, part-of-speech tagging, lemmatization, classification, spelling correction, etc.

GenSim: This service is designed for information extraction and natural language processing. It has many algorithms that can be deployed irrespective of the size of the collection of linguistic data. As it is dependent on NumPy and SciPy (Python packages for scientific computing), the user needs to install these two packages before installing Genism. The tool is extremely structured, and it has top-notch memory optimization and processing speed. It enables operating large text files even without loading the whole file in memory. Genism doesn't require costly annotations or hand tagging of documents because it uses unsupervised models.

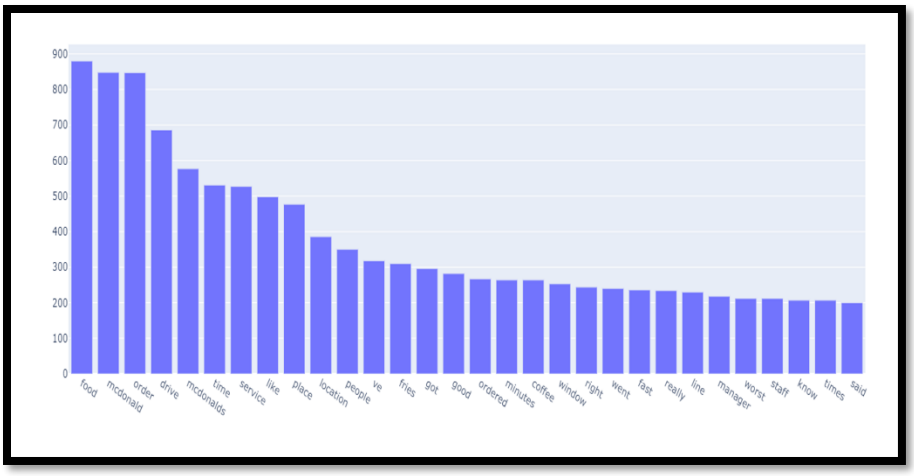
CoreNLP can be used to create linguistic annotations for text, such as:

- Token and sentence boundaries.
- Parts of speech.
- Named entities.
- Numeric and temporal values; dependency and constituency parser.
- Sentiment.
- Quotation attributions.
- Relations between words

Result:

After doing it text pre-processing we captured this World Cloud

- Here when we chose the top 30 unigrams in the MacDonalds review after removing stop words and lemmatization.



- **We do three experiments of NLP Topic Modeling, Classification, Cluster:**
  - **Topic modeling: LSA, LDA, NMF**

✓ **The topic modeling for LSA is:**

```
Topic  0
order, mcdonald, food, drive, time, like, service, mcdonalds, place, people

Topic  1
mcdonald, like, review, mcwrap, fries, people, old, north, ξand, way

Topic  2
food, mcdonalds, place, fast, service, eat, like, good, people, want

Topic  3
food, order, fast, mcdonald, service, window, waiting, customers, time, longer

Topic  4
order, time, fries, manager, coffee, minutes, waiting, counter, ordered, chicken

Topic  5
fries, food, chicken, ordered, window, drive, asked, got, said, went
```

- ✓ **The topic modeling for NMF is:**

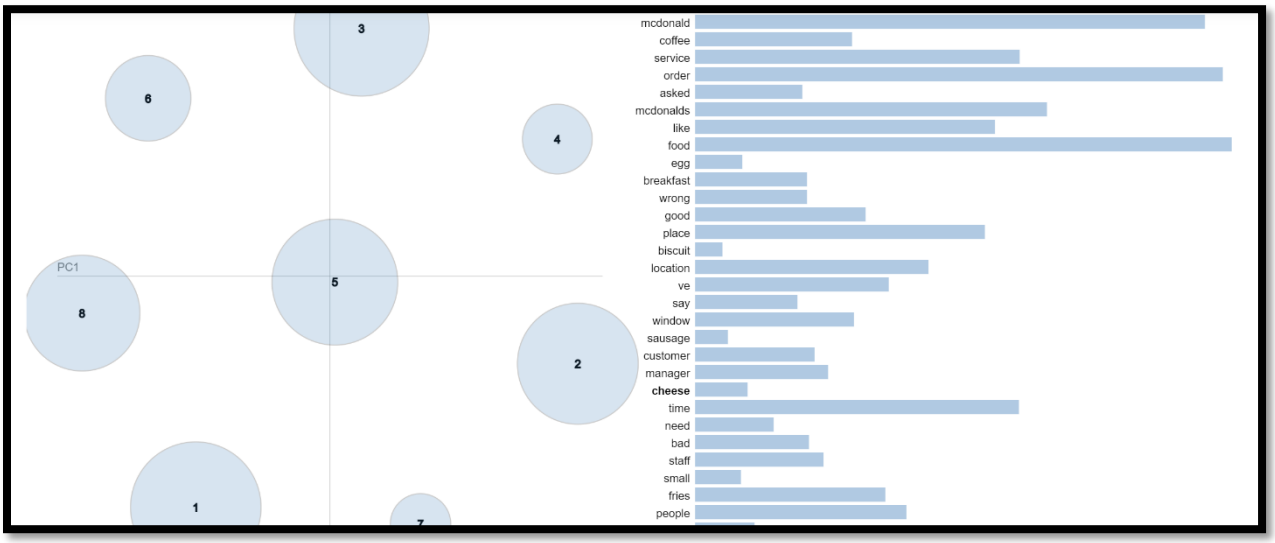
```
Topic  0
order, time, wrong, minutes, right, breakfast, wait, window, ordered, waiting

Topic  1
mcdonald, review, people, ve, know, way, line, ξthe, say, good

Topic  2
food, fast, time, waiting, eat, restaurant, customers, line, fresh, quick

Topic  3
mcdonalds, ve, breakfast, good, want, really, people, ξthe, employee, pretty

Topic  4
drive, window, car, way, inside, went, cars, location, sure, said

Topic  5
fries, chicken, ordered, got, said, ξthe, went, cold, asked, meal
```

- ✓ **The topic modeling for LDA is:**

```
Topic  0
coffee, mcdonald, place, like, food, good, ve, mcdonalds, drive, got

Topic  1
food, mcdonald, order, drive, like, fries, fast, time, mcdonalds, place

Topic  2
mcdonald, mcdonalds, food, like, time, location, place, good, fries, ve

Topic  3
service, order, drive, food, asked, manager, mcdonald, place, say, like

Topic  4
mcdonald, time, food, mcdonalds, order, breakfast, worst, orders, like, drive

Topic  5
order, drive, service, food, mcdonald, mcdonalds, time, window, place, minutes
```

- **We take the world cloud of best model LDA:**

✓ **We Visualize the top words per topic:**



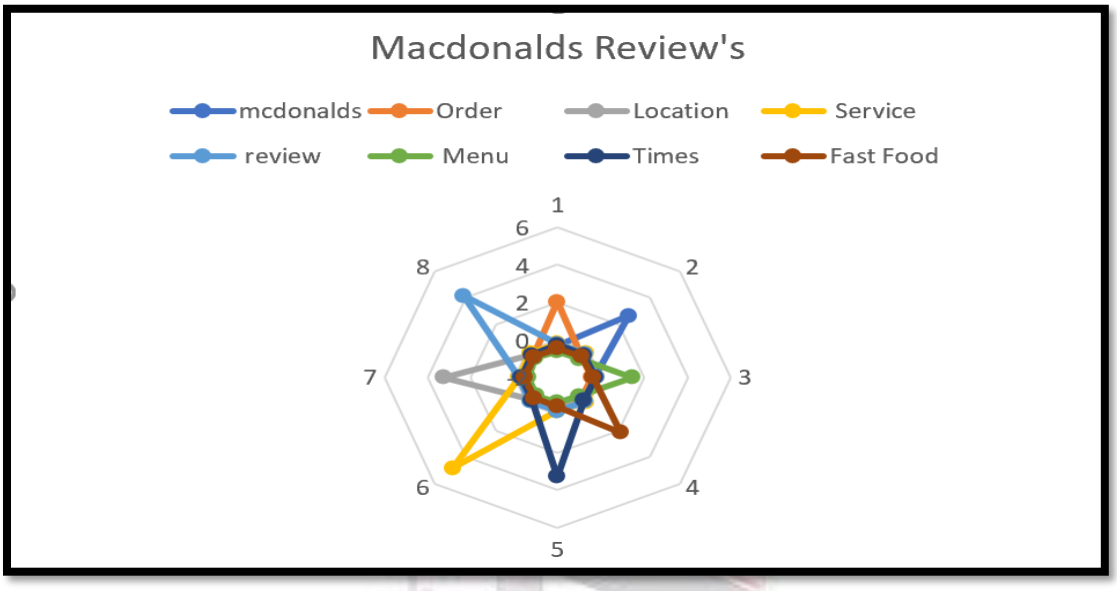
✓ **Show the top words for each topic in a dataframe:**

	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10	Word 11	Word 12	Word 13	Word 14
Topic 0	coffee	mcdonald	place	like	food	good	ve	mcdonalds	drive	got	time	staff	location	service	inside
Topic 1	food	mcdonald	order	drive	like	fries	fast	time	mcdonalds	place	location	people	ordered	got	went
Topic 2	mcdonald	mcdonalds	food	like	time	location	place	good	fries	ve	service	drive	people	think	make
Topic 3	service	order	drive	food	asked	manager	mcdonald	place	say	like	wrong	customer	cheese	location	way
Topic 4	mcdonald	time	food	mcdonalds	order	breakfast	worst	orders	like	drive	sauce	location	coffee	people	work
Topic 5	order	drive	service	food	mcdonald	mcdonalds	time	window	place	minutes	customer	line	like	ve	wait
Topic 6	mcdonald	food	like	service	time	people	breakfast	biscuit	place	egg	drive	ξthe	ordered	fries	way
Topic 7	order	food	mcdonalds	drive	location	time	people	place	service	right	mcdonald	like	really	times	ve

- ✓ **CorEx**

```
0: ordered,order,said,window,bag,time,asked,got,told,know
1: recent,dont,old,write,children,chaotic,numbers,staying,friends,building
2: pretty,mcdonald,review,walked,visit,mcd,case,ξthis,week,thing
3: exit,star,cash,turn,customers,trying,ξand,lanes,ξit,ξi
4: line,manager,place,store,later,new,work,think,guy,fact
5: egg,sausage,cheese,ξthe,biscuit,wrong,mcmuffin,bacon,black,hash
```

- ✓ **Cluster:**



Communication: The slides will be provided here, feel free to any pull requests besides details are provided at the readme of the project.