Python for Urban Data Analysis: Text Processing and Sentiment Insights

Eric Delmelle[1,2,3]
delmelle@gmail.com

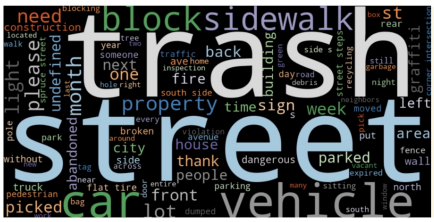
In this seminar, we will explore text processing and sentiment analysis using Python. Two datasets from Philadelphia (US) will be used: (a) 311 phone complaints from residents (typically about infrastructure issues) and (b) Google reviews for parks.

The class will cover generating wordclouds to visualize frequent complaints, basic and advanced sentiment analysis using models like DistilBERT and EMOBERT to assess emotional tones in the texts, and applying these techniques to understand public sentiment on urban issues.

There will be hands-on exercises, and you will gain practical skills in Natural Language Processing (NLP), discuss real-world applications, and engage in collaborative discussion as to how text analysis can inform public policy and community engagement.

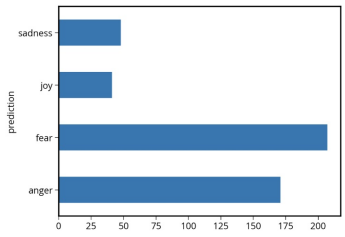
I will show basics of web data scraping (including APIs). If time allows, we will discuss how to incorporate Large Language Models (LLM) in your python environment.

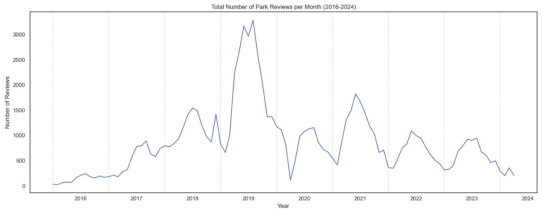
Requirement: some basic knowledge of python programming, and a google account (if you use Google Colab).




```
# The model
model = "bhadresh-savani/distilbert-base-uncased-emotion"

# Initialize our sentiment analyzer
emotion_classifier = pipeline(
    task="text-classification",  # The task we are doing
    model=model,  # The specific model name
    top_k=None,  # Predict all labels, not just top ones
    tokenizer=model,  # Tokenize inputs using model tokenizer
    truncation=True,  # Truncate text if we need to
)
```





[1] Vrije Universiteit Brussels
[2] University *of* North Carolina at Charlotte
[3] University *of* Pennsylvania