Brief Raport, Mateusz Guściora, 02.04.2024

## Steps

1. Plan, design and Set hypothesis
2. Implement:
   1. Loading libraries and data files
   2. Quick Overview-general info
   3. Cleaning data and preliminary analysis
   4. Text preprocessing
   5. Sentiment analysis-adding new information to dataset
   6. Testing Hypothesis
   7. Data presentation and evaluation
3. Raport

## Data

Set used: “*job\_postings.csv*”

Attributes with possible interesting informations: *description(job), company\_description, views, applies, formatted\_work\_type, sponsored, skills\_desc, formatted\_experience\_level*

Analysis script: “*zadanko-linkledin-job-posting.ipynb*”

## Hypothesis

**Hypothesis 1** (H1): *Job postings with positive emotional tone in their descriptions are associated with a higher number of applicant views or/and applications, compared to those with neutral or negative tones.*

H0: *There is no association between the emotional tone of job postings (positive, neutral, negative) and the number of applicant views or applications.*

* if p\_value < 0.05 it supports H1
* if p\_value >=0.05 it supports H0 and therefore no significant findings
* Furthermore results can be either**: consistent findings** (if all methods show similar results) or **inconsistent findings** (if more than one of methods show different results)

## Questions and Answers during analysis

Q: Does sentiment have impact on number of views, applies?

A: Sentiment does not appear to significantly impact no. of views or applications

Q: Does methods for hypothesis testing are right for this data?

A: There is no normal distribution in data so ANOVA should not probably be used. The rest methods are appropriate but reveal no strong patterns in the data.

Q: How does the method of sentiment analysis (TextBlob vs. NLTK) impact evaluation?

A: Lead to different evaluation outcomes (sentiment scores) and different results in tests.

Q: Is there consistency in results? A: There is not. While most indicate no relation. Three results does.

Q: Can the applies-to-views ratio be a more insightful measure than raw applies or views counts?

A: It can, it shows relation between applies and views but not for this problem as there is not concluding pattern or association between sentiment and former.

Q: What additional factors, beyond sentiment, might influence the number of views and applications a job posting receives? (not tested))

## Evaluation of hypothesis and results

Analysis suggests a lack of consistent association between job posting sentiment and number of views or applications, with no strong evidence to support a significant impact of sentiment on applicant behavior, leaning towards the null hypothesis (H0). While there are instances where certain sentiment categories show potential significance, such as with the Chi-Square test for applications using TextBlob, these are not uniformly observed across all tests [*Table1*].

Obraz zawierający tekst, zrzut ekranu, menu, Czcionka

Opis wygenerowany automatycznie

Table1: P-value Results of testing hypothesis with evaluation (yes, no, dk)

The visualized data also does not indicate a distinct pattern of increase or decrease in average views or applications in relation to sentiment categories [*Figure1*].

Obraz zawierający tekst, zrzut ekranu, diagram, Czcionka

Opis wygenerowany automatycznie

Figure1: Distribution of avg. views, applies by sentiment category on two different methods

## Setup info

Setup: VC, jupyter, python, libraries (\*provided in the notebook), CPU

Methods for sentiment score calculation: textblop, nltk vader

Methods for hypothesis testing: ANOVA, Kruskal-Wallis , Chi-square-test

Run time of notebook: 818.93 seconds (13.60 minutes)

Sentiment method textblob: 167.05 seconds (2.70 minutes)

Sentiment method nltk: 640.49 seconds (10.60 minutes)