Understanding Happiness Using NLP

By:
Nadia & Manal

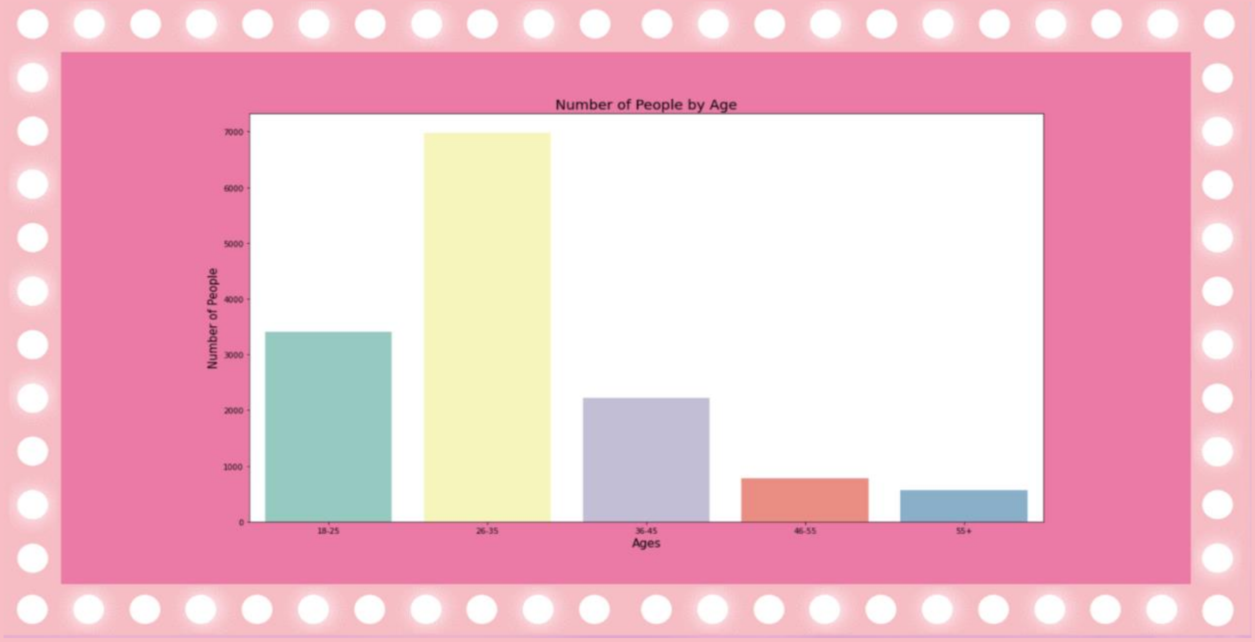
Abstract:

Being happy is one of the most fundamental requirements of a living being. Since the beginning of the human civilization, man has also been trying to develop new technologies, make new tools and improve his lifestyle for the sole purpose of attaining happiness. However, in its race of scientific endeavor and pursuit of money and luxuries, man is hardly aware of what exactly constitutes happiness.

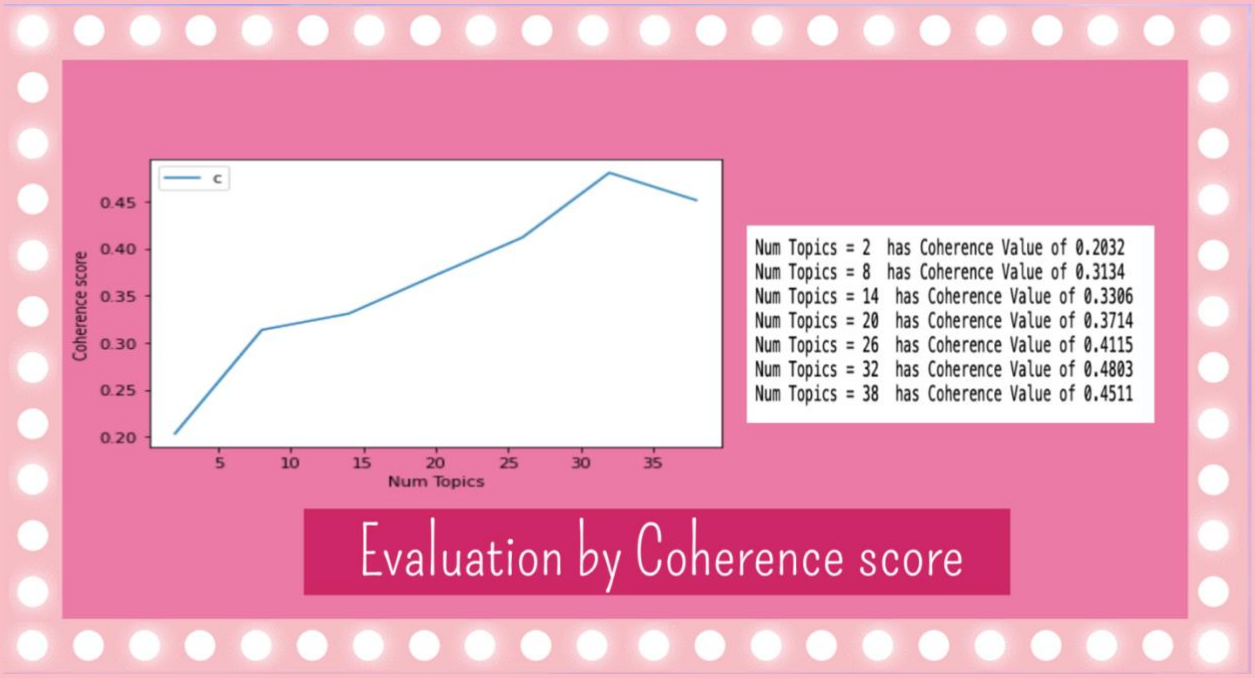
Project Goals: Topic modelling to understand what makes Amazon's workers happy for the previous 24 hours

Design:

After obtaining the data from Kaggle, we merged both files on common column and start applying EDA techniques to explore the dataset. One of the EDA's outcomes is the age group participated in:



After full understanding of the data, we start pre-processing journey to clean text sets. The journey included: cleaning texts from punctuations and numbers- Normalization/ regular expressions – Tokenization – Removing stop-words and finally Lemmatization. Then by building LDA model, we were able to cluster texts without label them. It helped in determining which word carries out more weight in each set of texts. After that, we visualized the topics using pyLDAvis to get more insights about our initialized topic numbers. Finally, we evaluated our decision regarding number of topics by using Coherence score which indicated that 32 topics is the optimal number of topics in our dataset.



```
Num Topics = 2  has Coherence Value of 0.2032
Num Topics = 8  has Coherence Value of 0.3134
Num Topics = 14  has Coherence Value of 0.3306
Num Topics = 20  has Coherence Value of 0.3714
Num Topics = 26  has Coherence Value of 0.4115
Num Topics = 32  has Coherence Value of 0.4803
Num Topics = 38  has Coherence Value of 0.4511
```

Evaluation by Coherence score

Data:

The data was obtained from: Kaggle
[https://www.kaggle.com/ritresearch/happydb];

“It is a dataset of more than 100,000 happy moments crowd-sourced via Amazon’s Mechanical Turk. Each worker was given the following task:
What made you happy today?
Reflect on the past 24 hours and recall three actual events that happened to you that made you happy. Write down your happy moment in a complete sentence. (Write three such moments.)
The second dataset contains demographic information of the respondents: age, gender, marriage status, etc. ‟

Algorithm:

EDA, NLP, Topic modelling and Coherence scores

Tools:

- Panda
- Numpy
- Matplotlib and Seaborn
- Gensim
- Spacy
- Nltk ' Natural Language Toolkit '
- pyLDAvis

Communication:

https://view.genial.ly/61939b92bdf28d0d7b8122fe/interactive-content-project4nlp