

Tanya

Roll no – 101866008

4CS3

Input

```
In [4]: text
```

```
Out[4]: 'Success from two leading coronavirus vaccine programs likely means other frontrunners will also show strong protection against COVID-19, Bill Gates said Tuesday.\n\nThe fact that two coronavirus vaccines recently showed strong protection against COVID-19 bodes well for other leading programs led by AstraZeneca, Novavax, and Johnson & Johnson, Bill Gates said Tuesday.The billionaire Microsoft founder and philanthropist said it will be easier to boost manufacturing and distribute these other shots to the entire world, particularly developing nations.The vaccine space has seen a flurry of good news in recent days, marked by overwhelming success in late-stage trials by both Pfizer and Moderna. The studies showed both vaccines provided strong protection against the virus compared to a placebo. "With the very good news from Pfizer and Moderna, we think it's now likely that AstraZeneca, Novavax, and Johnson & Johnson will also likely show very strong efficacy," Gates told journalist Andrew Ross Sorkin. \n\nWhile Gates didn't delve into the scientific rationale behind that prediction, many scientists hold the same hope. All the leading vaccine candidates target the same part of the coronavirus in the spike protein. Early-stage clinical trials showed all the shots elicited varying levels of neutralizing antibodies, virus-fighting proteins that play a critical role in the body's immune response. But the only way to know if that translates to protection is by running late-stage trials that enroll tens of thousands of volunteers who receive either the experimental shot or placebo injections. The scientific success has turned the top challenges surrounding a COVID-19 vaccine to the manufacturing and distribution front. Gates noted that the world will be supply constrained for 2021, but these additional vaccines will prove valuable on that front.'
```

Output

## 2 Create the Frequency matrix of the words in each sentence.

```
In [6]: freq_matrix = _create_frequency_matrix(sentences)
print(freq_matrix)
```

```
{'Success from tw': {'success': 1, 'two': 1, 'lead': 1, 'coronaviru': 1, 'vaccin': 1, 'program': 1, 'like': 1, 'mean': 1, 'frontrunn': 1, 'also': 1, 'show': 1, 'strong': 1, 'protect': 1, 'covid-19': 1, 'bill': 1, 'gate': 1, 'said': 1, 'tuesday': 1, '': 1}, 'The fact that t': {'fact': 1, 'two': 1, 'coronaviru': 1, 'vaccin': 2, 'recent': 2, 'show': 1, 'strong': 1, 'protect': 1, 'covid-19': 1, 'bode': 1, 'well': 1, 'lead': 1, 'program': 1, 'led': 1, 'astrazeneca': 1, '': 5, 'novavax': 1, 'johnson': 2, '&': 1, 'bill': 1, 'gate': 1, 'said': 2, 'tuesday.th': 1, 'billionair': 1, 'microsoft': 1, 'founder': 1, 'philanthropist': 1, 'easier': 1, 'boost': 1, 'manufactur': 1, 'distribut': 1, 'shot': 1, 'entir': 1, 'world': 1, 'particularli': 1, 'develop': 1, 'nations.th': 1, 'space': 1, 'ha': 1, 'seen': 1, 'flurri': 1, 'good': 1, 'news': 1, 'day': 1, 'mark': 1, 'overwhelm': 1, 'success': 1, 'late-stag': 1, 'trial': 1, 'pfizer': 1, 'moderna': 1, '': 1}, 'The studies sho': {'studi': 1, 'show': 1, 'vaccin': 1, 'provid': 1, 'strong': 1, 'protect': 1, 'viru': 1, 'compar': 1, 'placebo': 1, '': 1}, '"With the very ': {'': 1, 'ver': 2, 'good': 1, 'news': 1, 'pfizer': 1, 'moderna': 1, '': 4, 'think': 1, "'s": 1, 'like': 2, 'astrazeneca': 1, 'novavax': 1, 'johnson': 2, '&': 1, 'also': 1, 'show': 1, 'strong': 1, 'efficaci': 1, "'": 1, 'gate': 1, 'told': 1, 'journalist': 1, 'andre': 1, 'ross': 1, 'sorkin': 1, '': 1}, 'While Gates did': {'gate': 1, 'n't': 1, 'delv': 1, 'scientific': 1, 'rational': 1, 'behind': 1, 'predict': 1, '': 1, 'mani': 1, 'scientist': 1, 'hold': 1, 'hope': 1, '': 1}, 'All the leading': {'lead': 1, 'vaccin': 1, 'candid': 1, 'target': 1, 'part': 1, 'coronaviru': 1, 'spike': 1, 'protein': 1, '': 1}, 'Early-stage cli': {'early-stage': 1, 'clinic': 1, 'trial': 1, 'show': 1, 'shot': 1, 'elicit': 1, 'vari': 1, 'level': 1, 'neutral': 1, 'antibodi': 1, '': 1}, 'virus-fight': 1, 'protein': 1, 'play': 1, 'critic': 1, 'role': 1, 'bodi': 1, "'s": 1, 'immun': 1, 'respons': 1, '': 1}, 'But the only wa': {'onli': 1, 'way': 1, 'know': 1, 'translat': 1, 'protect': 1, 'run': 1, 'late-stag': 1, 'trial': 1, 'enrol': 1, 'ten': 1, 'thousand': 1, 'volunt': 1, 'receiv': 1, 'either': 1, 'experiment': 1, 'shot': 1, 'placebo': 1, 'inject': 1, '': 1}, 'The scientific ': {'scientific': 1, 'success': 1, 'ha': 1, 'turn': 1, 'top': 1, 'challeng': 1, 'surround': 1, 'covid-19': 1, 'vaccin': 1, 'manufactur': 1, 'distribut': 1, 'front': 1, '': 1}, 'Gates noted tha': {'gate': 1, 'note': 1, 'world': 1, 'suppli': 1, 'constrain': 1, '2021': 1, '': 1, 'addit': 1, 'vaccin': 1, 'prove': 1, 'valuabl': 1, 'front': 1, '': 1}}
```

```
In [7]: tf_matrix = _create_tf_matrix(freq_matrix)
print(tf_matrix)
```

```
{'Success from tw': {'success': 0.05, 'two': 0.05, 'lead': 0.05, 'coronaviru': 0.05, 'vaccin': 0.05, 'program': 0.05, 'like': 0.05, 'mean': 0.05, 'frontrunn': 0.05, 'also': 0.05, 'show': 0.05, 'strong': 0.05, 'protect': 0.05, 'covid-19': 0.05, '': 0.05, 'bill': 0.05, 'gate': 0.05, 'said': 0.05, 'tuesday': 0.05, '': 0.05}, 'The fact that t': {'fact': 0.019230769230769232, 'two': 0.019230769230769232, 'coronaviru': 0.019230769230769232, 'vaccin': 0.038461538461538464, 'recent': 0.038461538461538464, 'show': 0.019230769230769232, 'strong': 0.019230769230769232, 'protect': 0.019230769230769232, 'covid-19': 0.019230769230769232, 'bode': 0.019230769230769232, 'well': 0.019230769230769232, 'lead': 0.019230769230769232, 'program': 0.019230769230769232, 'led': 0.019230769230769232, 'astrazeneca': 0.019230769230769232, '': 0.09615384615384616, 'novavax': 0.019230769230769232, 'johnson': 0.038461538461538464, '&': 0.019230769230769232, 'bill': 0.019230769230769232, 'gate': 0.019230769230769232, 'said': 0.038461538461538464, 'tuesday.th': 0.019230769230769232, 'billionair': 0.019230769230769232, 'microsoft': 0.019230769230769232, 'founder': 0.019230769230769232, 'philanthropist': 0.019230769230769232, 'easier': 0.019230769230769232, 'boost': 0.019230769230769232, 'manufactur': 0.019230769230769232, 'distribut': 0.019230769230769232, 'shot': 0.019230769230769232, 'entir': 0.019230769230769232, 'world': 0.019230769230769232, 'particularli': 0.019230769230769232, 'develop': 0.019230769230769232, 'nation.s.th': 0.019230769230769232, 'space': 0.019230769230769232, 'ha': 0.019230769230769232, 'seen': 0.019230769230769232, 'flurri': 0.019230769230769232, 'good': 0.019230769230769232, 'news': 0.019230769230769232, 'day': 0.019230769230769232, 'mark': 0.019230769230769232, 'overwhelm': 0.019230769230769232, 'success': 0.019230769230769232, 'late-stag': 0.019230769230769232, 'trial': 0.019230769230769232, 'pfizer': 0.019230769230769232, 'moderna': 0.019230769230769232, '': 0.019230769230769232}, 'The studies sho': {'studi': 0.1, 'show': 0.1, 'vaccin': 0.1, 'provid': 0.1, 'strong': 0.1, 'protect': 0.1, 'viru': 0.1, 'compar': 0.1, 'placebo': 0.1, '': 0.1}, '"With the very ': {'': 0.038461538461538464, 'veri': 0.07692307692307693, 'good': 0.038461538461538464, 'news': 0.038461538461538464, 'pfizer': 0.038461538461538464, 'moderna': 0.038461538461538464, '': 0.15384615384615385, 'think': 0.038461538461538464, 's': 0.038461538461538464, 'like': 0.07692307692307693, 'astrazeneca': 0.038461538461538464, 'novavax': 0.038461538461538464, 'johnson': 0.07692307692307693, '&': 0.038461538461538464, 'also': 0.038461538461538464, 'show': 0.038461538461538464, 'strong': 0.038461538461538464, 'efficaci': 0.038461538461538464, '': 0.038461538461538464, 'gate': 0.038461538461538464, 'told': 0.038461538461538464, 'journalist': 0.038461538461538464, 'andrew': 0.038461538461538464, 'ross': 0.038461538461538464, 'sorkin': 0.038461538461538464, '': 0.038461538461538464}, 'While Gates did': {'gate': 0.07692307692307693, 'n': 0.07692307692307693, 'delv': 0.07692307692307693, 'scientif': 0.07692307692307693, 'rational': 0.07692307692307693, 'behind': 0.07692307692307693, 'predict': 0.07692307692307693, '': 0.07692307692307693, 'mani': 0.07692307692307693, 'scientist': 0.07692307692307693, 'hold': 0.07692307692307693, 'hope': 0.07692307692307693, '': 0.07692307692307693}, 'All the leading': {'lead': 0.11111111111111111, 'vaccin': 0.11111111111111111, 'candid': 0.11111111111111111, 'target': 0.11111111111111111, 'part': 0.11111111111111111, 'coronaviru': 0.11111111111111111, 'spike': 0.11111111111111111, 'protein': 0.11111111111111111, '': 0.11111111111111111}, 'Early-stage cli': {'early-stag': 0.047619047619047616, 'clinic': 0.047619047619047616, 'trial': 0.047619047619047616, 'show': 0.047619047619047616, 'shot': 0.047619047619047616, 'elicit': 0.047619047619047616, 'vari': 0.047619047619047616, 'level': 0.047619047619047616, 'neutral': 0.047619047619047616, 'antibodi': 0.047619047619047616, '': 0.047619047619047616, 'virus-fight': 0.047619047619047616, 'protein': 0.047619047619047616, 'play': 0.047619047619047616, 'critic': 0.047619047619047616, 'role': 0.047619047619047616, 'bodi': 0.047619047619047616, 's': 0.047619047619047616, 'immun': 0.047619047619047616, 'respons': 0.047619047619047616, '': 0.047619047619047616}, 'But the only wa': {'onli': 0.05263157894736842, 'way': 0.05263157894736842, 'note': 0.05263157894736842, 'suppli': 0.05263157894736842, 'constrain': 0.05263157894736842, '2021': 0.05263157894736842, 'addit': 0.05263157894736842, 'prove': 0.05263157894736842, 'valuabl': 0.05263157894736842, 'front': 0.05263157894736842, '': 0.05263157894736842}
```

## 4 Creating table for documents per words

```
In [8]: count_doc_per_words = _create_documents_per_words(freq_matrix)
print(count_doc_per_words)
```

```
{'success': 3, 'two': 2, 'lead': 3, 'coronaviru': 3, 'vaccin': 6, 'program': 2, 'like': 2, 'mean': 1, 'frontrunn': 1, 'also': 2, 'show': 5, 'strong': 4, 'protect': 4, 'covid-19': 3, '': 6, 'bill': 2, 'gate': 5, 'said': 2, 'tuesday': 1, '': 10, 'fact': 1, 'recent': 1, 'bode': 1, 'well': 1, 'led': 1, 'astrazeneca': 2, 'novavax': 2, 'johnson': 2, '&': 2, 'tuesday.th': 1, 'billionair': 1, 'microsoft': 1, 'founder': 1, 'philanthropist': 1, 'easier': 1, 'boost': 1, 'manufactur': 2, 'distribut': 2, 'shot': 3, 'entir': 1, 'world': 2, 'particularli': 1, 'develop': 1, 'nation.s.th': 1, 'space': 1, 'ha': 2, 'seen': 1, 'flurri': 1, 'good': 1, 'news': 2, 'day': 1, 'mark': 1, 'overwhelm': 1, 'late-stag': 2, 'trial': 3, 'pfizer': 2, 'moderna': 2, 'studi': 1, 'provid': 1, 'viru': 1, 'compar': 1, 'placebo': 2, '': 1, 'veri': 1, 'think': 1, 's': 2, 'efficaci': 1, '': 1, 'told': 1, 'journalist': 1, 'andrew': 1, 'ross': 1, 'sorkin': 1, 'n': 1, 'delv': 1, 'scientif': 2, 'rational': 1, 'behind': 1, 'predict': 1, 'mani': 1, 'scientist': 1, 'hold': 1, 'hope': 1, 'candid': 1, 'target': 1, 'part': 1, 'spike': 1, 'protein': 2, 'early-stag': 1, 'clinic': 1, 'elicit': 1, 'level': 1, 'neutral': 1, 'antibodi': 1, 'virus-fight': 1, 'play': 1, 'critic': 1, 'role': 1, 'bodi': 1, 'immun': 1, 'respons': 1, 'onli': 1, 'way': 1, 'know': 1, 'translat': 1, 'run': 1, 'enrol': 1, 'ten': 1, 'thousand': 1, 'volunt': 1, 'receiv': 1, 'either': 1, 'experiment': 1, 'inject': 1, 'turn': 1, 'top': 1, 'challeng': 1, 'surround': 1, 'front': 2, 'note': 1, 'suppli': 1, 'constrain': 1, '2021': 1, 'addit': 1, 'prove': 1, 'valuabl': 1}
```

## 5 Calculate IDF and generate a matrix

```
In [9]: idf_matrix = _create_idf_matrix(freq_matrix, count_doc_per_words, total_documents)
        print(idf_matrix)
```

```
{'Success from tw': {'success': 0.5228787452803376, 'two': 0.6989700043360189, 'lead': 0.5228787452803376, 'coronaviru': 0.5228787452803376, 'vaccin': 0.2218487496163564, 'program': 0.6989700043360189, 'like': 0.6989700043360189, 'mean': 1.0, 'frontrun n': 1.0, 'also': 0.6989700043360189, 'show': 0.3010299956639812, 'strong': 0.3979400086720376, 'protect': 0.3979400086720376, 'covid-19': 0.5228787452803376, '': 0.2218487496163564, 'bill': 0.6989700043360189, 'gate': 0.3010299956639812, 'said': 0.6989700043360189, 'tuesday': 1.0, '.': 0.0}, 'The fact that t': {'fact': 1.0, 'two': 0.6989700043360189, 'coronaviru': 0.5228787452803376, 'vaccin': 0.2218487496163564, 'recent': 1.0, 'show': 0.3010299956639812, 'strong': 0.3979400086720376, 'protect': 0.3979400086720376, 'covid-19': 0.5228787452803376, 'bode': 1.0, 'well': 1.0, 'lead': 0.5228787452803376, 'program': 0.6989700043360189, 'led': 1.0, 'astrazeneca': 0.6989700043360189, '': 0.2218487496163564, 'novavax': 0.6989700043360189, 'johnson': 0.6989700043360189, '&': 0.6989700043360189, 'bill': 0.6989700043360189, 'gate': 0.3010299956639812, 'said': 0.6989700043360189, 'tuesd ay.th': 1.0, 'billionair': 1.0, 'microsoft': 1.0, 'founder': 1.0, 'philanthropist': 1.0, 'easier': 1.0, 'boost': 1.0, 'manuf act ur': 0.6989700043360189, 'distribut': 0.6989700043360189, 'shot': 0.5228787452803376, 'entin': 1.0, 'world': 0.6989700043360189, 'particularli': 1.0, 'develop': 1.0, 'nations.th': 1.0, 'space': 1.0, 'ha': 0.6989700043360189, 'seen': 1.0, 'flurri': 1.0, 'good': 0.6989700043360189, 'news': 0.6989700043360189, 'day': 1.0, 'mark': 1.0, 'overwhelm': 1.0, 'success': 0.5228787452803376, 'late-stag': 0.6989700043360189, 'trial': 0.5228787452803376, 'pfizer': 0.6989700043360189, 'moderna': 0.6989700043360189, '': 0.0}, 'The studies sho': {'studi': 1.0, 'show': 0.3010299956639812, 'vaccin': 0.2218487496163564, 'provid': 1.0, 'strong': 0.3979400086720376, 'protect': 0.3979400086720376, 'viru': 1.0, 'compar': 1.0, 'placebo': 0.6989700043360189, '': 0.0}, 'With the very': {'': 1.0, 'veri': 1.0, 'good': 0.6989700043360189, 'news': 0.6989700043360189, 'pfizer': 0.6989700043360189, 'mod erna': 0.6989700043360189, '': 0.2218487496163564, 'think': 1.0, '": 0.6989700043360189, 'like': 0.6989700043360189, 'astraz eneca': 0.6989700043360189, 'novavax': 0.6989700043360189, 'johnson': 0.6989700043360189, '&': 0.6989700043360189, 'also': 0.6989700043360189, 'show': 0.3010299956639812, 'strong': 0.3979400086720376, 'efficaci': 1.0, '": 1.0, 'gate': 0.3010299956639812, 'told': 1.0, 'journalist': 1.0, 'andrew': 1.0, 'ross': 1.0, 'sorkin': 1.0, '.': 0.0}, 'While Gates did': {'gate': 0.3010299956639812, 'n't": 1.0, 'delv': 1.0, 'scientif': 0.6989700043360189, 'rational': 1.0, 'behind': 1.0, 'predict': 1.0, '': 0.2218487496163564, 'mani': 1.0, 'scientist': 1.0, 'hold': 1.0, 'hope': 1.0, '': 0.0}, 'All the leading': {'lead': 0.5228787452803376, 'vaccin': 0.2218487496163564, 'candid': 1.0, 'target': 1.0, 'part': 1.0, 'coronaviru': 0.5228787452803376, 'spike': 1.0, 'p rotein': 0.6989700043360189, '': 0.0}, 'Early-stage cli': {'early-stage': 1.0, 'clinic': 1.0, 'trial': 0.5228787452803376, 'sho w': 0.3010299956639812, 'shot': 0.5228787452803376, 'elicit': 1.0, 'vari': 1.0, 'level': 1.0, 'neutral': 1.0, 'antibodi': 1.0, '': 0.2218487496163564, 'virus-fight': 1.0, 'protein': 0.6989700043360189, 'play': 1.0, 'critic': 1.0, 'role': 1.0, 'bodi': 1.0, '": 0.6989700043360189, 'immun': 1.0, 'respons': 1.0, '': 0.0}, 'But the only wa': {'onli': 1.0, 'way': 1.0, 'know': 1.0, 'translat': 1.0, 'protect': 0.3979400086720376, 'run': 1.0, 'late-stag': 0.6989700043360189, 'trial': 0.5228787452803376, 'enro 1': 1.0, 'ten': 1.0, 'thousand': 1.0, 'volunt': 1.0, 'receiv': 1.0, 'either': 1.0, 'experiment': 1.0, 'shot': 0.5228787452803376, 'placebo': 0.6989700043360189, 'inject': 1.0, '': 0.0}, 'The scientific': {'scientif': 0.6989700043360189, 'success': 0.5228787452803376, 'ha': 0.6989700043360189, 'turn': 1.0, 'top': 1.0, 'challeng': 1.0, 'surround': 1.0, 'covid-19': 0.5228787452803376, 'vaccin': 0.2218487496163564, 'manufactur': 0.6989700043360189, 'distribut': 0.6989700043360189, 'front': 0.6989700043360189, '': 0.0}, 'Gates noted tha': {'gate': 0.3010299956639812, 'note': 1.0, 'world': 0.6989700043360189, 'suppli': 1.0, 'const rain': 1.0, '2021': 1.0, '': 0.2218487496163564, 'add': 1.0, 'vaccin': 0.2218487496163564, 'prove': 1.0, 'valuabl': 1.0, 'fr ont': 0.6989700043360189, '': 0.0}}
```

## 6 Calculate TF-IDF and generate a matrix

```
In [10]: tf_idf_matrix = _create_tf_idf_matrix(tf_matrix, idf_matrix)
         print(tf_idf_matrix)
```

```
{'Success from tw': {'success': 0.026143937264016884, 'two': 0.034948500216800946, 'lead': 0.026143937264016884, 'coronaviru': 0.026143937264016884, 'vaccin': 0.01109243748081782, 'program': 0.034948500216800946, 'like': 0.034948500216800946, 'mean': 0.05, 'frontrunn': 0.05, 'also': 0.034948500216800946, 'show': 0.01505149978319906, 'strong': 0.01989700043360188, 'protect': 0.01989700043360188, 'covid-19': 0.026143937264016884, '': 0.01109243748081782, 'bill': 0.034948500216800946, 'gate': 0.01505149978319906, 'said': 0.034948500216800946, 'tuesday': 1.0, '.': 0.0}, 'The fact that t': {'fact': 0.019230769230769232, 'two': 0.013441730852615748, 'coronaviru': 0.010055360486160339, 'vaccin': 0.008532644216013708, 'recent': 0.038461538461538464, 'show': 0.005789038378153485, 'strong': 0.007652692474462262, 'protect': 0.007652692474462262, 'covid-19': 0.010055360486160339, 'bod e': 0.019230769230769232, 'well': 0.019230769230769232, 'lead': 0.010055360486160339, 'program': 0.013441730852615748, 'johnso n': 0.026883461705231496, '&': 0.013441730852615748, 'bill': 0.013441730852615748, 'gate': 0.005789038378153485, 'said': 0.026883461705231496, 'tuesday.th': 0.019230769230769232, 'billionair': 0.019230769230769232, 'microsoft': 0.019230769230769232, 'fou nder': 0.019230769230769232, 'philanthropist': 0.019230769230769232, 'easier': 0.019230769230769232, 'boost': 0.019230769230769232, 'manufactur': 0.013441730852615748, 'distribut': 0.013441730852615748, 'shot': 0.010055360486160339, 'entin': 0.019230769230769232, 'world': 0.013441730852615748, 'particularli': 0.019230769230769232, 'develop': 0.019230769230769232, 'nations.th': 0.019230769230769232, 'space': 0.019230769230769232, 'ha': 0.013441730852615748, 'seen': 0.019230769230769232, 'flurri': 0.019230769230769232, 'good': 0.013441730852615748, 'news': 0.013441730852615748, 'day': 0.019230769230769232, 'mark': 0.019230769230769232, 'overwhelm': 0.019230769230769232, 'success': 0.010055360486160339, 'late-stag': 0.013441730852615748, 'trial': 0.010055360486160339, 'pfizer': 0.013441730852615748, 'moderna': 0.013441730852615748, '': 0.0}, 'The studies sho': {'studi': 0.1, 's how': 0.03010299956639812, 'vaccin': 0.02218487496163564, 'provid': 0.1, 'strong': 0.03979400086720376, 'protect': 0.03979400086720376, 'viru': 0.1, 'compar': 0.1, 'placebo': 0.06989700043360189, '': 0.0}, 'With the very': {'': 0.0}, '0.038461538461538464, 'veri': 0.07692307692307693, 'good': 0.026883461705231496, 'news': 0.026883461705231496, 'pfizer': 0.026883461705231496, 'moder na': 0.026883461705231496, '': 0.03413057686405483, 'think': 0.038461538461538464, '": 0.026883461705231496, 'like': 0.05376692341046299, 'astrazeneca': 0.026883461705231496, 'novavax': 0.026883461705231496, 'johnson': 0.05376692341046299, '&': 0.026883461705231496, 'also': 0.026883461705231496, 'show': 0.01157807675630697, 'strong': 0.015305384948924524, 'efficaci': 0.038461538461538464, '": 0.038461538461538464, 'gate': 0.01157807675630697, 'told': 0.038461538461538464, 'journalist': 0.038461538461538464, 'andrew': 0.038461538461538464, 'ross': 0.038461538461538464, 'sorkin': 0.038461538461538464, '': 0.0}, 'While Gates did': {'gate': 0.02315615351261394, 'n't": 0.07692307692307693, 'delv': 0.07692307692307693, 'scientif': 0.05376692341046299, 'rational': 0.07692307692307693, 'behind': 0.07692307692307693, 'predict': 0.07692307692307693, '': 0.017065288432027415, 'man i': 0.07692307692307693, 'scientist': 0.07692307692307693, 'hold': 0.07692307692307693, 'hope': 0.07692307692307693, '': 0.0}, 'All the leading': {'lead': 0.05809763836448195, 'vaccin': 0.024649861068484043, 'candid': 0.11111111111111111, 'target': 0.11111111111111111, 'part': 0.11111111111111111, 'coronaviru': 0.05809763836448195, 'spike': 0.11111111111111111, 'protein': 0.0776633381511322, '': 0.0}, 'Early-stage cli': {'early-stage': 0.047619047619047616, 'clinic': 0.047619047619047616, 'trial': 0.024898987870492266, 'show': 0.047619047619047616, 'shot': 0.024898987870492266, 'elicit': 0.047619047619047616, 'vari': 0.047619047619047616, 'level': 0.047619047619047616, 'neutral': 0.047619047619047616, 'antibodi': 0.047619047619047616, '': 0.010564226172207447, 'virus-fight': 0.047619047619047616, 'protein': 0.0332842859207628, 'play': 0.047619047619047616, 'critic': 0.047619047619047616}}
```

## 7 Important Algorithm: score the sentences

```
In [11]: sentence_scores = _score_sentences(tf_idf_matrix)
print(sentence_scores)

{'Success from tw': 0.027817431287605544, 'The fact that t': 0.015668854959339104, 'The studies sho': 0.06017728766960432, '"Wi
th the very ': 0.032505924637327904, 'While Gates did': 0.06048431212790743, 'All the leading': 0.07366143511744506, 'Early-sta
ge cli': 0.03847296200569854, 'But the only wa': 0.04388265237646744, 'The scientific ': 0.05184885362045636, 'Gates noted th
a': 0.05409862428147178}
```

## 8 Find the threshold

```
In [12]: threshold = _find_average_score(sentence_scores)
print(threshold)

0.04586183380833235
```

## 9 Generate the summary

```
In [13]: summary = _generate_summary(sentences, sentence_scores, 1 * threshold)
print(summary)

The studies showed both vaccines provided strong protection against the virus compared to a placebo. While Gates didn't delve
into the scientific rationale behind that prediction, many scientists hold the same hope. All the leading vaccine candidates ta
rget the same part of the coronavirus in the spike protein. The scientific success has turned the top challenges surrounding a
COVID-19 vaccine to the manufacturing and distribution front. Gates noted that the world will be supply constrained for 2021, b
ut these additional vaccines will prove valuable on that front.
```