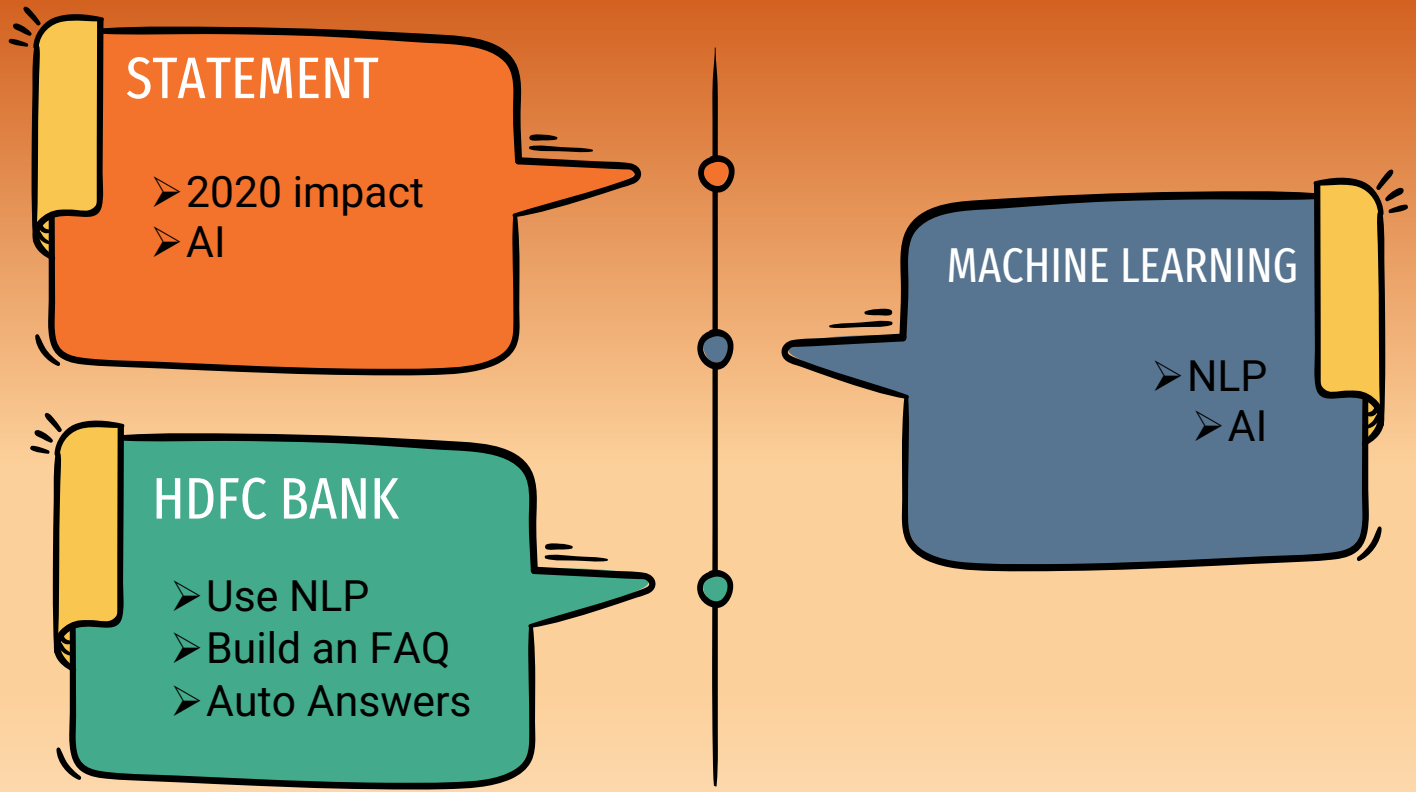
AUTO FAQ ANSWERING W/ NLP

BERNARD OPOKU
HDFC BANK

INTRODUCTION



DATA

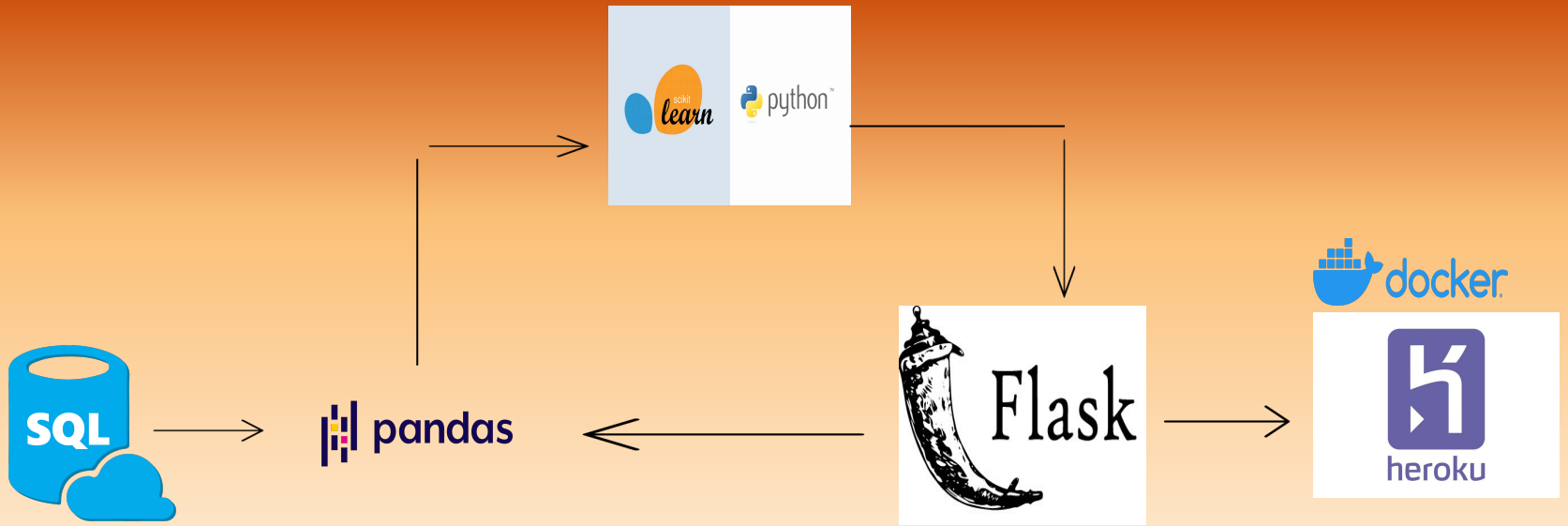
Kaggle.com

About 7000 observations

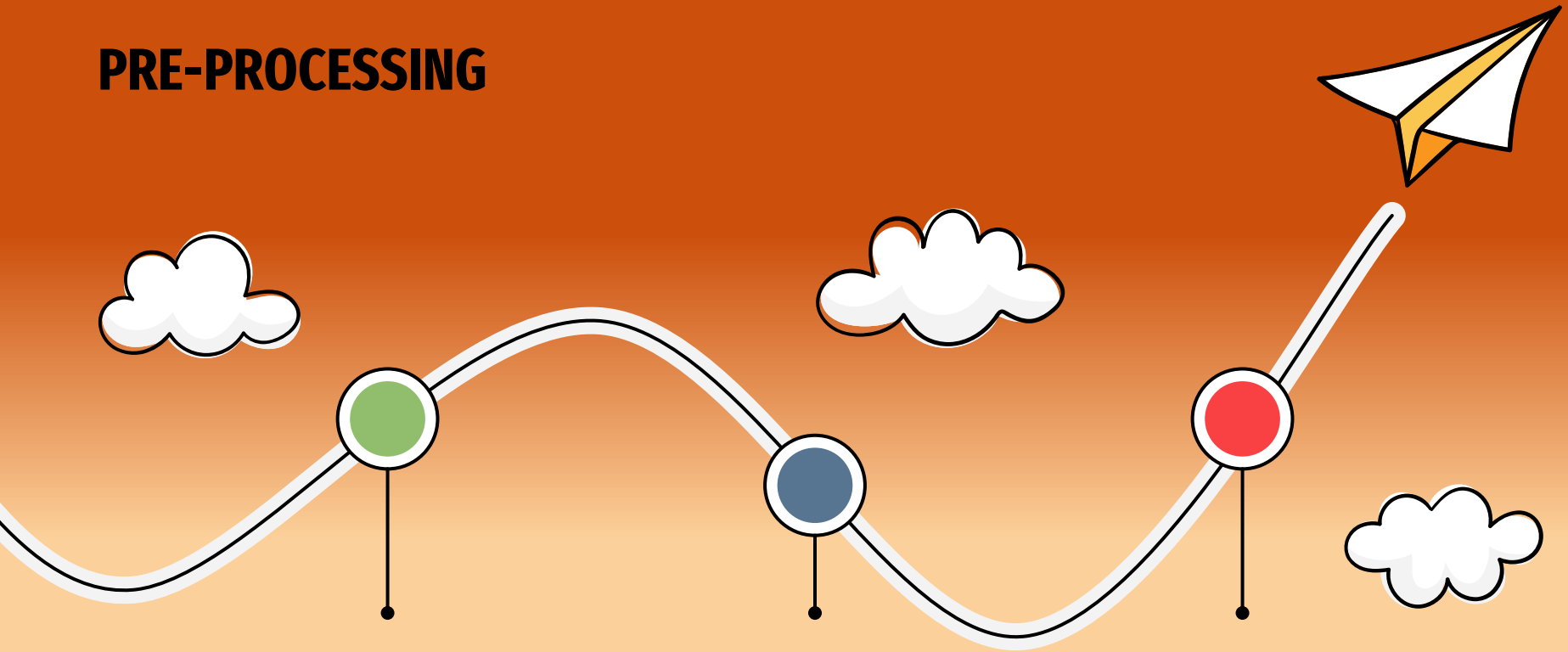
HDFC Bank

Customer query and answer pairs

END-END WORKFLOW



PRE-PROCESSING



Cleaning

- Replacing “/” with “or”
- Digits
- Extra Punctuations

Vectorization

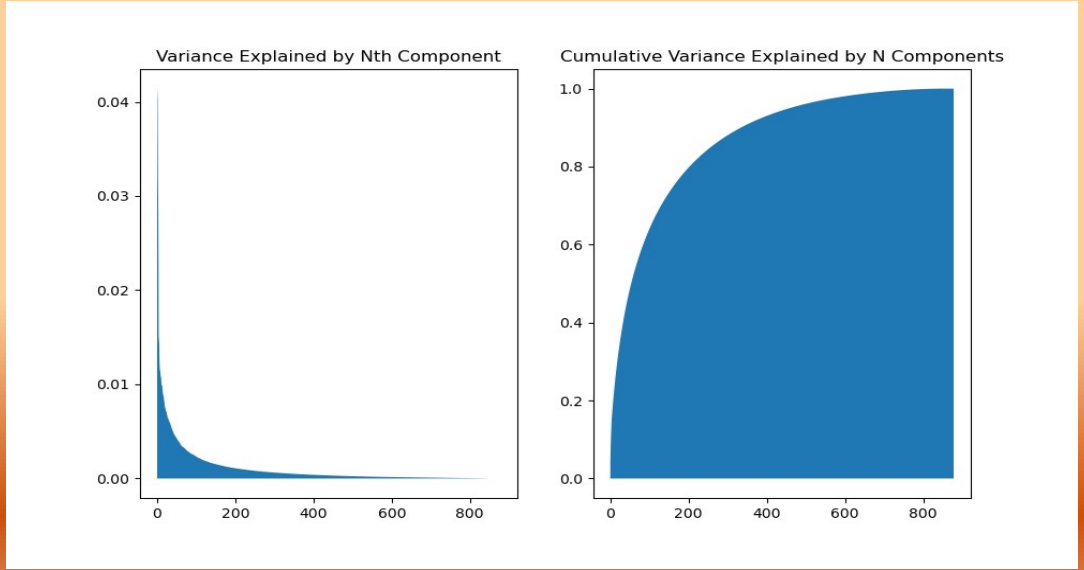
- TF-IDF
  - Penalize freq. words

Dimensionality reduction

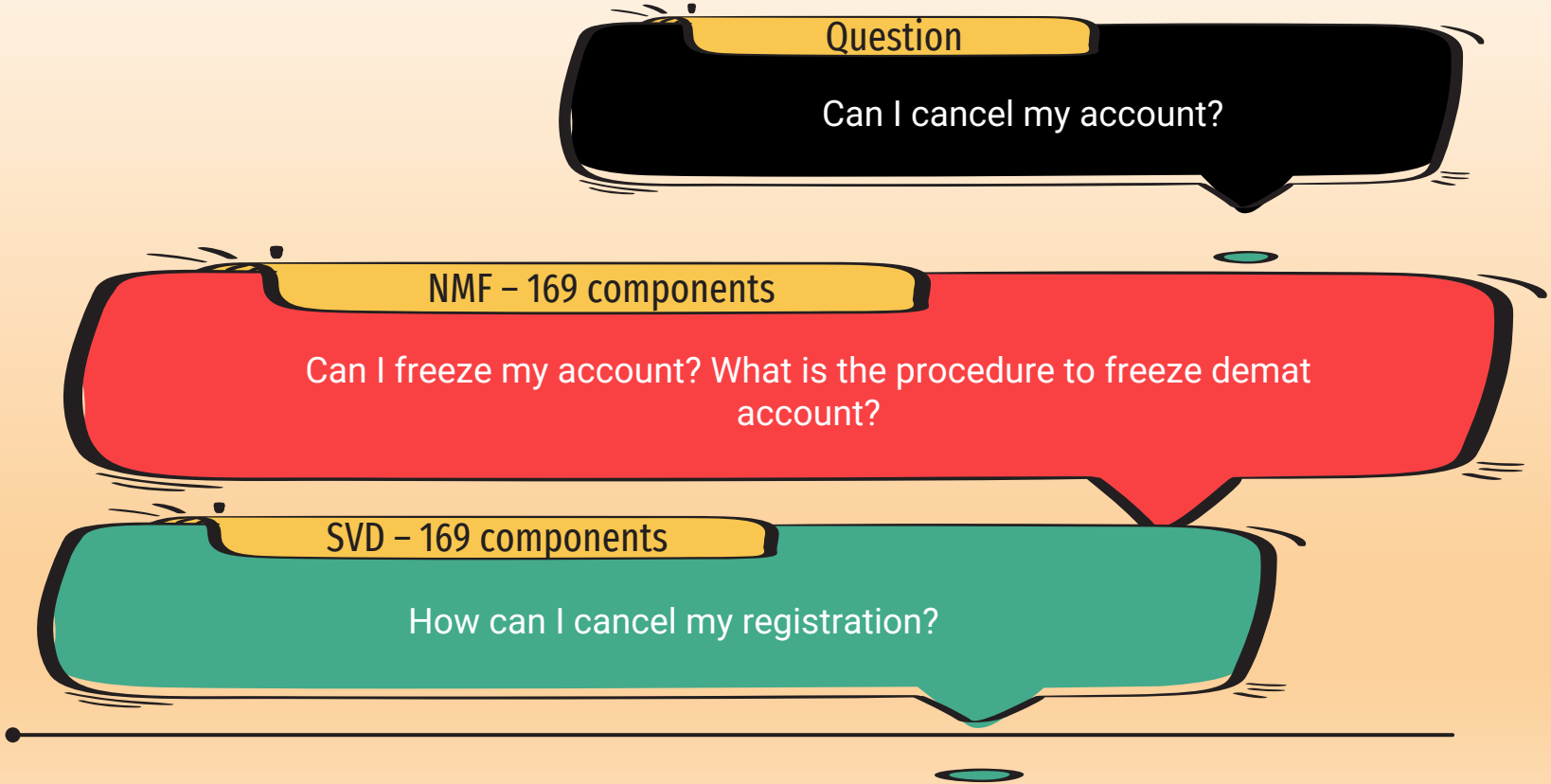
- OPTHT
- LSA (Truncated SVD)
- NMF

SVD and Optimal Threshold

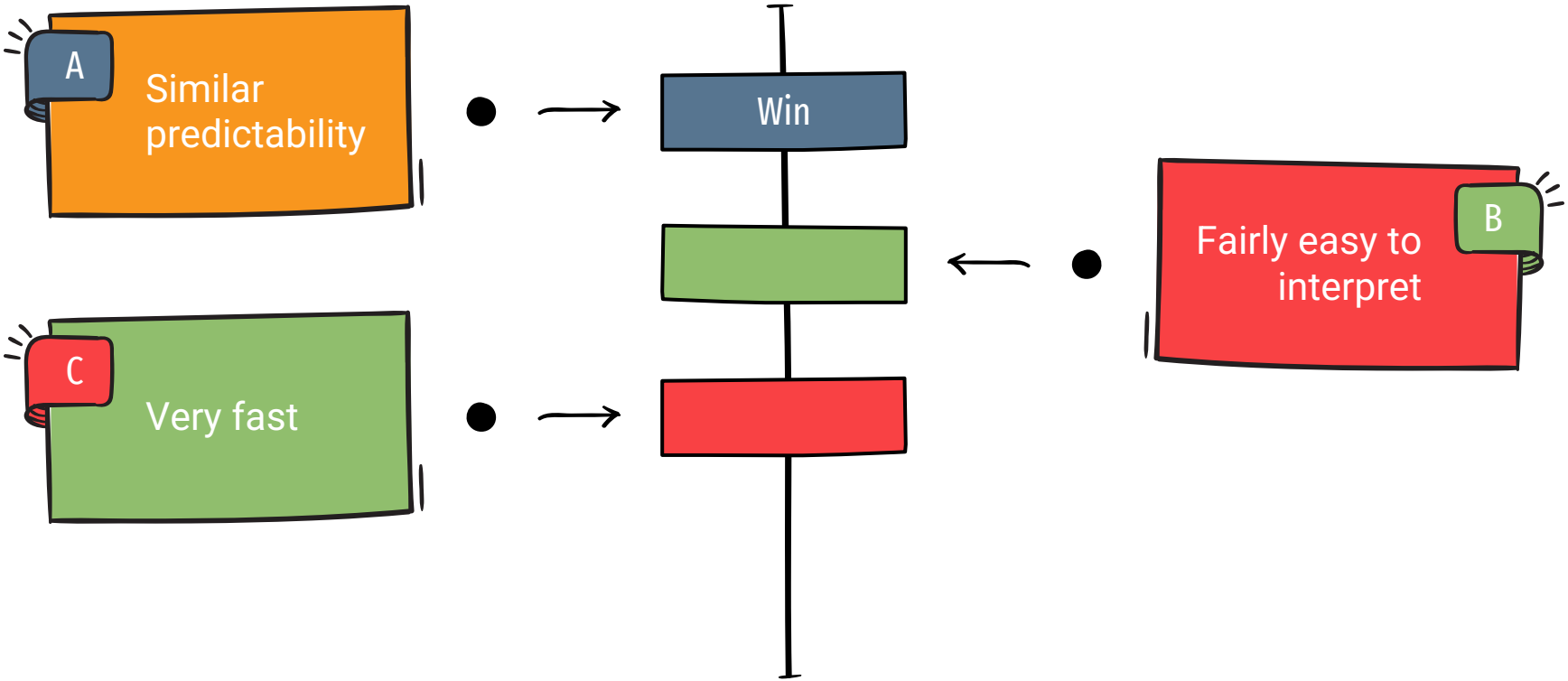
- OPTHT - IEEE (Gavin & Donoho)
  - 169 comps
- Explained Variance
  - Elbow method



RECOMMENDATION SYSTEM



LSA – Truncated svd



localhost:8890/lab/workspaces/auto-4/tree/NLP/nlp_proj/faq_modeling.ipynb

Apps | New Tab | internships | Physics II | The School of Pho | MIT ML | Spotify – Machine... | Tesla App | NNet From Scratc... | jobs | Other Bookmarks | Reading List

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ ... / nlp_proj / models /

Name	Last Modified
chatbot.mdl	18 hours ago
corpus.df	8 hours ago

faq_modeling.ipynb | fpl_modeling.ipynb | nlp_pipeline_pair_solu | Topic_Modeling_LSA_ | Untitled1.ipynb | auto_q_a.ipynb

Code | Python 3.9 (metis)

[ ]:
```
q = 'how can i cancel my account'
predict(svd169, question_topic, q)
```

[ ]:
```
q2 = 'what are the steps to cancel?'
predict(svd169, question_topic, q2)
```

[ ]:
```
q3 = 'how much loan can I qualify?'
predict(svd169, question_topic, q3)
```

$NMF - 169$

[ ]:
```
nmf169, question_topic_nmf169 = nmf_n(doc_term_mtx,169)
```

[ ]:
```
predict(nmf169, question_topic_nmf169, q)
```

I am pretty happy with both results and can be seen the prediction underfits and thinks almost evrything is similar to the question when when we go lower than 169 components and overfits when we go higher because we are adding more noise to the data. So this proves the Gavin and Donoho paper on $IEEE$

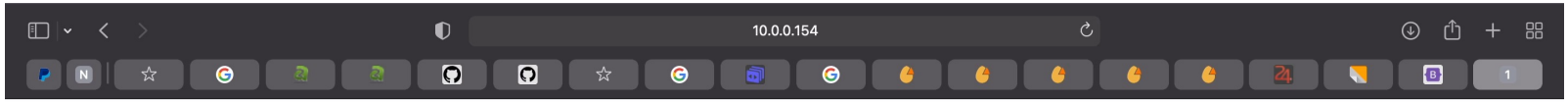
* Bonus

Standard scaling my data to see before dimensionatilty reduction

[ ]:
```
```

[ ]:
```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
scaled_doc_term_mtx_tfidf = sc.fit_transform(doc_term_mtx_tfidf.toarray())
```

Simple | 0 | 6 | Python 3.9 (metis) | Idle | Saving completed | Mode: Command | Ln 2, Col 36 | faq_modeling.ipynb



Welcome HDFC Bank Automatic FAQ

How may I assist you today?

Send

Future Work

- More Functionality
- More data
- Deep Learning/ Word2vec
- Better styling

Thank You

- Bernard