

JasonBot

(smart) Q&A Slack bot



Jason Salazer-Adams 🧐



bold *_italics_* ~strike~ ``code`` ````preformatted``` >quote

JasonBot

(smart) Q&A Slack bot




Jason Salazer-Adams 🧐



bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote



Chat bots...  |




bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

What's the weather like this weekend?



Are you on a boat? Because I was not able to find any results for that location.



Chat bots...  |



bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

What's the weather like this weekend?



Are you on a boat? Because I was not able to find any results for that location.

What's the weather like in Brooklyn this weekend?



The weather in Brooklyn, NY is 46°F and clear.



Chat bots...



bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

What's the weather like this weekend?



Are you on a boat? Because I was not able to find any results for that location.

What's the weather like in Brooklyn this weekend?



The weather in Brooklyn, NY is 46°F and clear.



Excusez-moi?

This weekend?

WEEKEND



Sorry, dozed off for a second. What were you saying?



+

Chat bots...  |

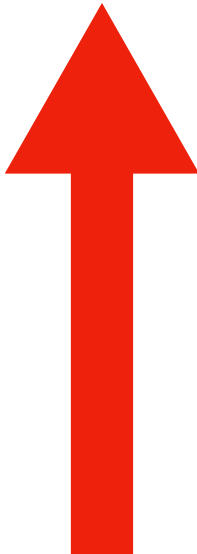
@ 

bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

What's the weather like this weekend?



Are you on a boat? Because I was not able to find any results for that location.



What's the weather like in Brooklyn this weekend?



The weather in Brooklyn, NY is 46°F and clear.



Excusez-moi?

This weekend?

WEEKEND



Sorry, dozed off for a second. What were you saying?





Tools and Workflow 



`*bold*` `_italics_` `~strike~` ``code`` ````preformatted```` `>quote`

Exploratory Data Analysis

- 400k pairs of questions
- Question pairs labeled as similar intent or not
- Actual Quora questions from a diverse set of topics

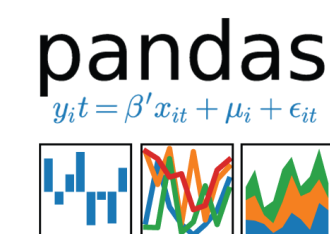
Feature Engineering

Modeling

Slack Bot Creation

kaggle

Quora



matplotlib
Version 3.0.2



Tools and Workflow 



bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

Exploratory Data Analysis

- Cleaned and lemmatized
- Calculated n-gram similarity between pairs

Feature Engineering

- Leveraged word embeddings
- Calculated distances between words for each question

Modeling

spaCy

NLTK 3.4



Slack Bot Creation

+

Similar Questions?? 🤔 💭 | @ 😊

`*bold*` `_italics_` `~strike~` ``code`` ```preformatted``` `>quote`

Base Question: How do I learn machine learning?

+

Similar Questions?? 🤔 💭 | @ 😊

`*bold*` `_italics_` `~strike~` ``code`` ````preformatted```` `>quote`

Base Question: How do I learn machine learning?

Compare Question: How do I become an astronaut?

+

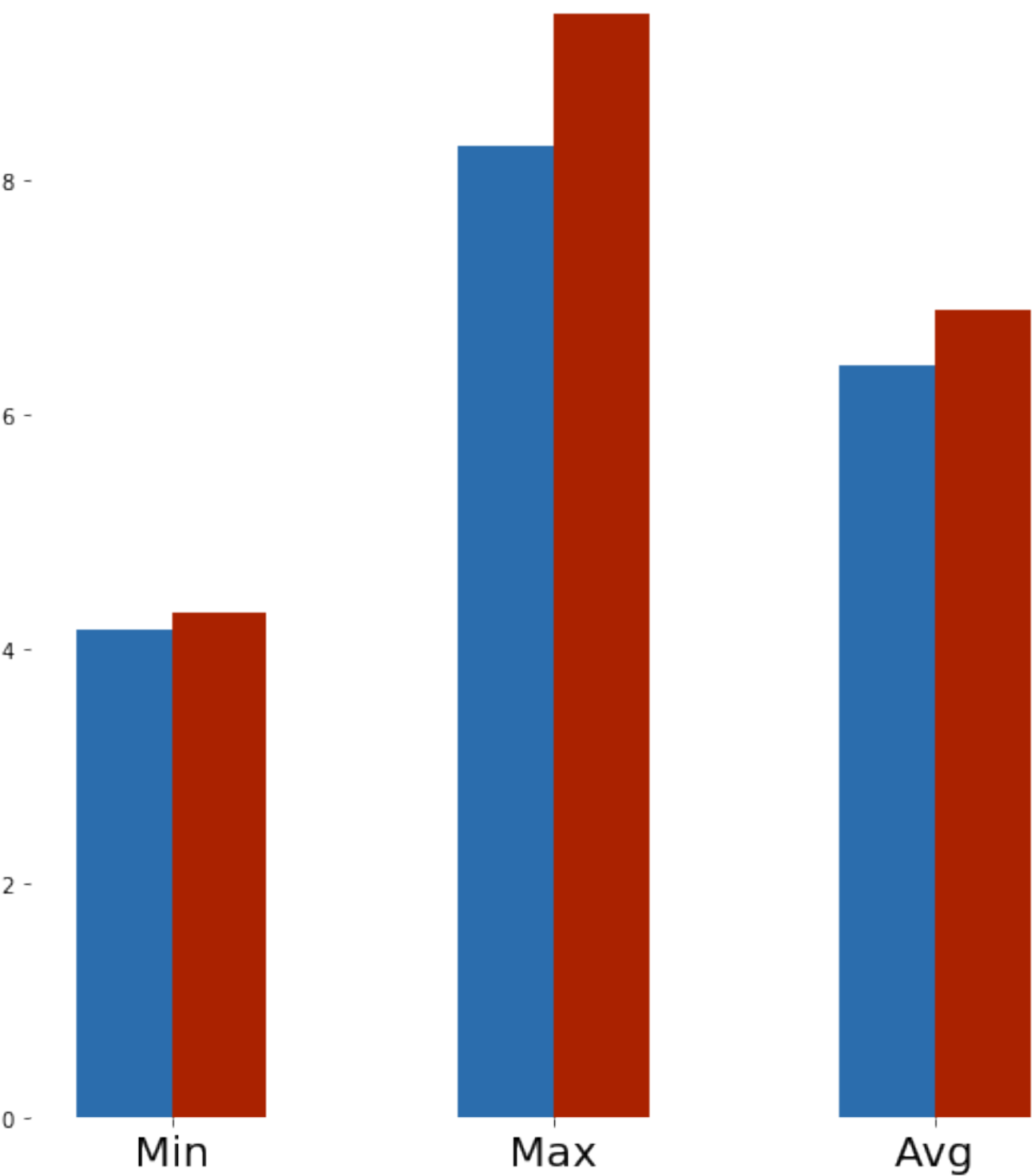
Similar Questions?? 🤔 ☁️ | @ 😊

`*bold*` `_italics_` `~strike~` ``code`` ````preformatted```` `>quote`

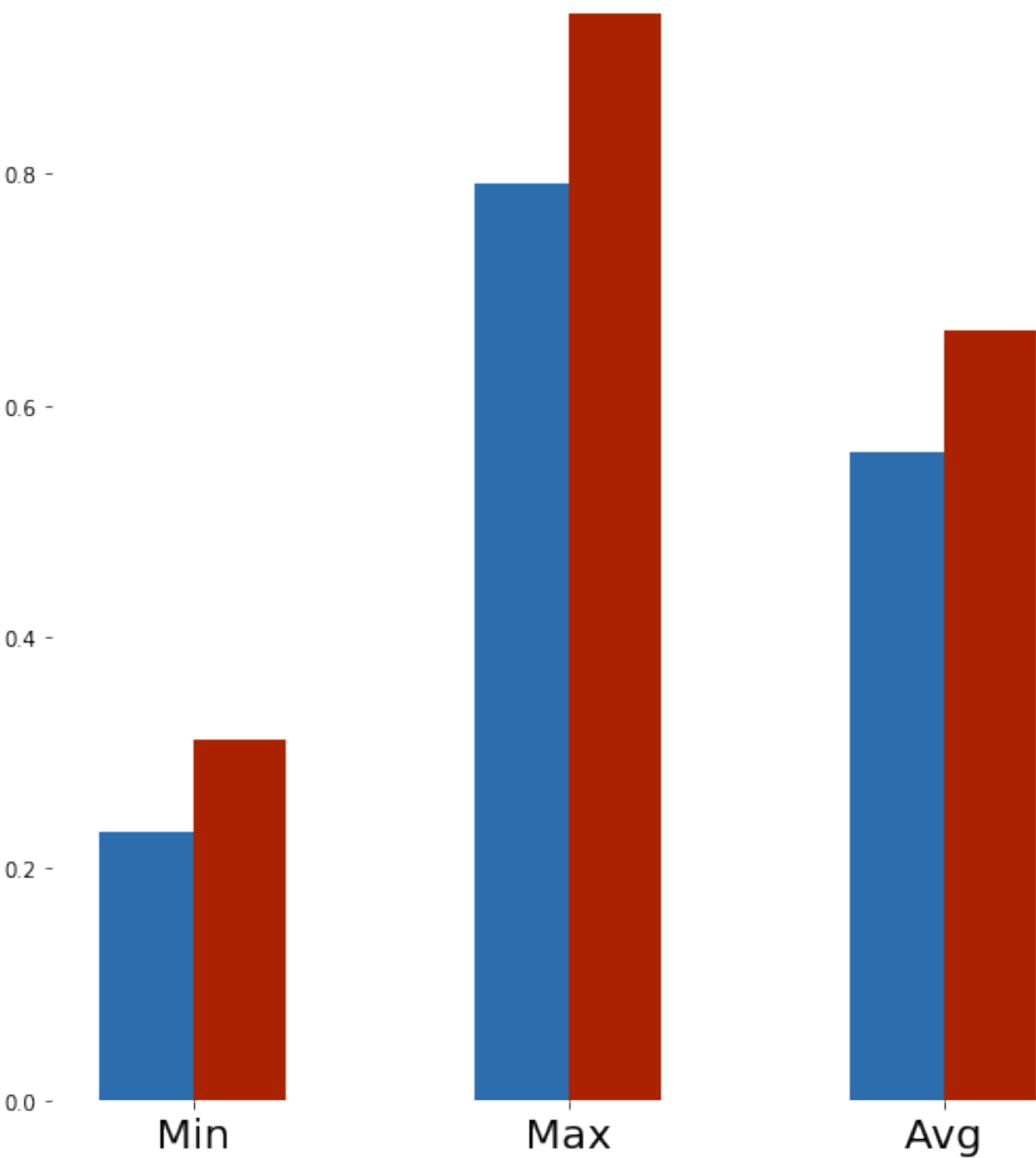
Base Question: How do I learn machine learning?

Compare Question: How do I become an astronaut?

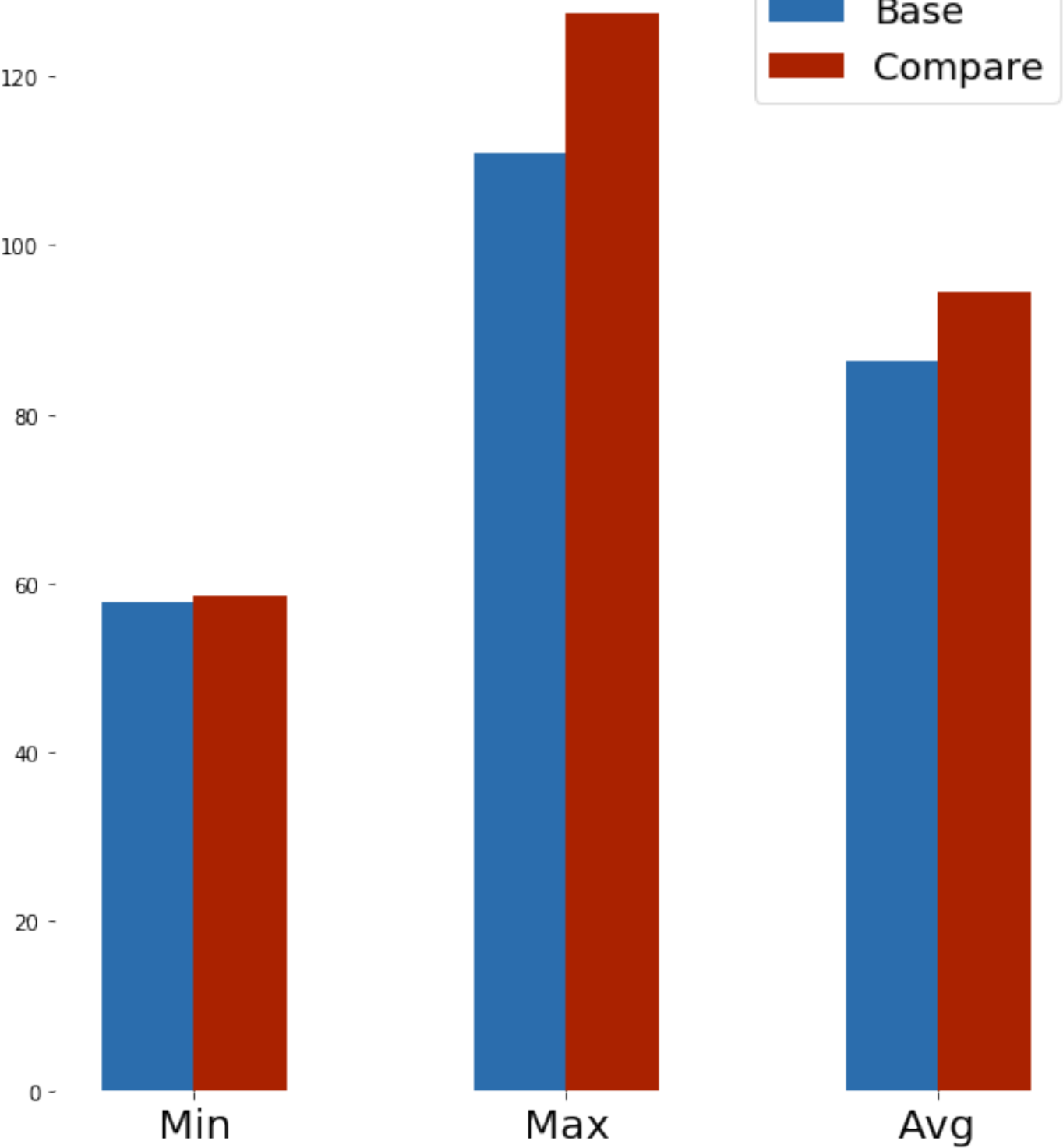
Euclidean



Cosine



Manhattan



+

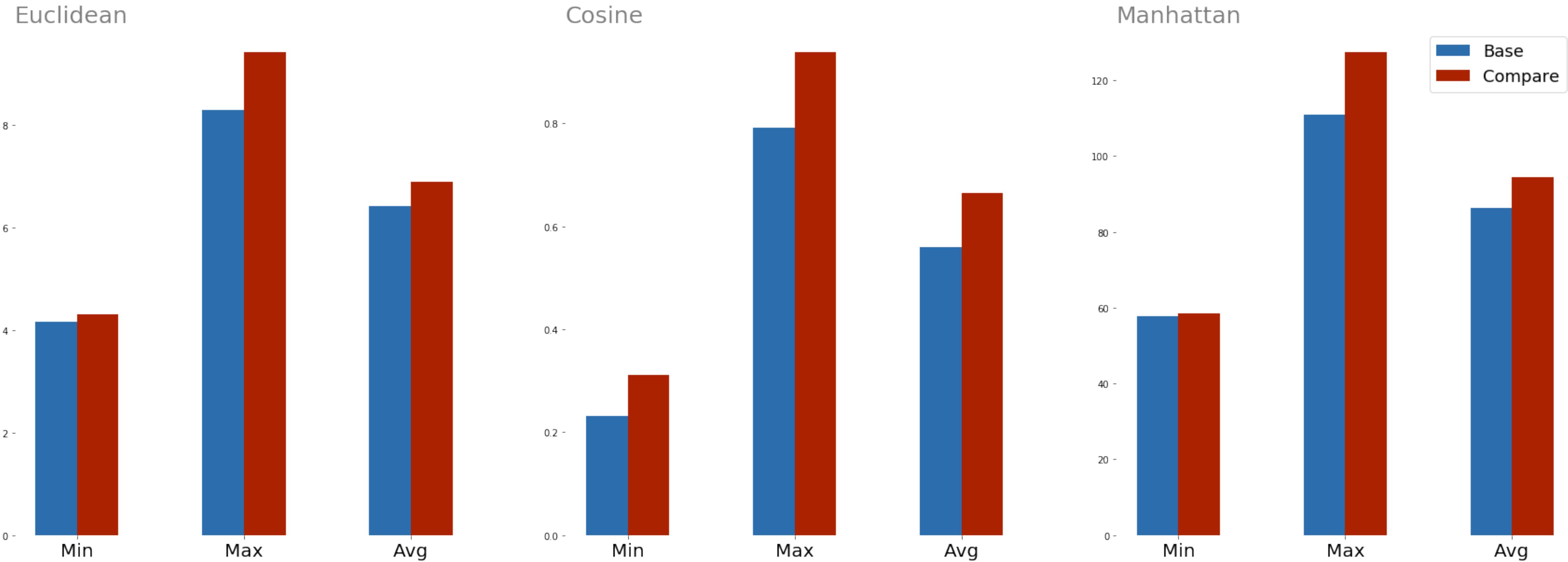
Similar Questions?? 🤔 ☁️ | @ 😊

`*bold*` `_italics_` `~strike~` ``code`` ````preformatted```` `>quote`

Base Question: How do I learn machine learning?

Compare Question: How do I become an astronaut?

New Question: What steps should I follow to learn machine learning?



+

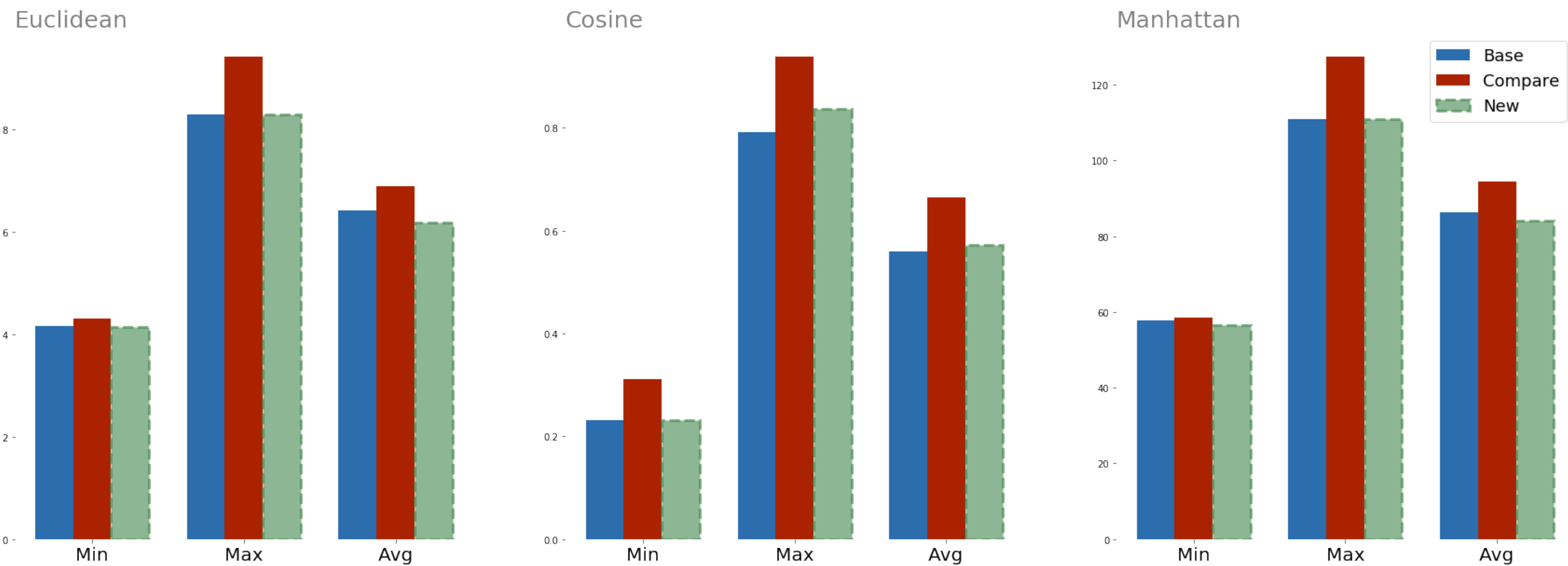
Similar Questions?? 🤔 💭 | @ 😊

`*bold*` `_italics_` `~strike~` ``code`` ````preformatted```` `>quote`

Base Question: How do I learn machine learning?

Compare Question: How do I become an astronaut?

New Question: What steps should I follow to learn machine learning?





Tools and Workflow 



bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

Exploratory Data Analysis

- Created modeling pipelines
- Trained classification models
- Cross-validated and hyper-tuned

Feature Engineering

Modeling



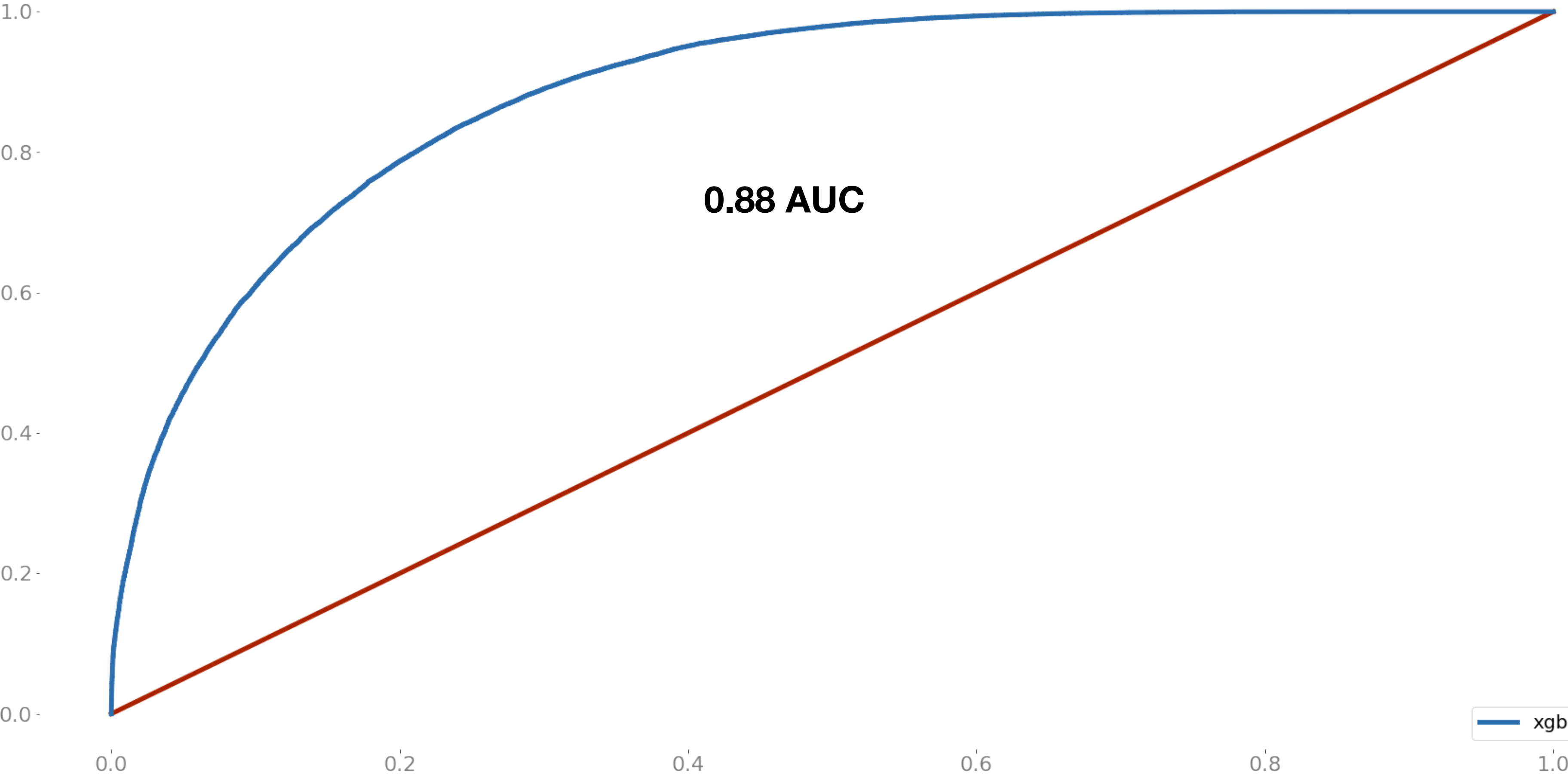
XGBoost



Slack Bot Creation

bold *italics* ~strike~ ``code`` ````preformatted```` >quote

True Positive Rate



False Positive Rate



Tools and Workflow 



bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

Exploratory Data Analysis

- Utilized Q&A data set from MS MARCO

- Developed Slack Bot

- Applied trained model to answer any question

Feature Engineering

Modeling



Slack Bot Creation

+

Demo! 🙌 |

@😊

bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

SEA FALL 2018

⌘1

METIS

⌘2

+

Metis Community

Jason Salazer-Adams

Jump to...

All Threads

Channels

admissions

ama

content-to-share

general

introductions

job-opportunities

python

q-and-a-global

random

Direct Messages

slackbot

Jason Salazer-Adams (...)

Apps

JasonBot

Jason Salazer-Adams (you)

☆ | ● active

i ⚙

Search

@ ☆ ⋮

Jason Salazer-Adams ●

This is your space. Draft messages, list your to-dos, or keep links and files handy. You can also talk to yourself here, but please bear in mind you'll have to supply both sides of the conversation.

+

Jot something down

@😊

+

Demo! 🙌 |

@😊

bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

SEA FALL 2018

⌘1

METIS

⌘2

+

Metis Community

Jason Salazer-Adams

Jump to...

All Threads

Channels

admissions

ama

content-to-share

general

introductions

job-opportunities

python

q-and-a-global

random

Direct Messages

slackbot

Jason Salazer-Adams (...)

Apps

JasonBot

Jason Salazer-Adams (you)

☆ | ● active

i ⚙

Search

@ ☆ ⋮

Jason Salazer-Adams ●

This is your space. Draft messages, list your to-dos, or keep links and files handy. You can also talk to yourself here, but please bear in mind you'll have to supply both sides of the conversation.

+

Jot something down

@😊

+

Future Work 🧑‍🔧 |

@ 😊

bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote

- Explore other word embedding models
- Explore DL models
- Explore text generation for Slack bot

+

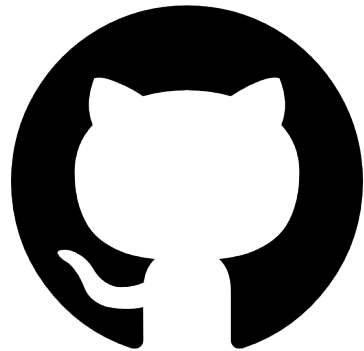
Thank You! 🧐 |

@😊

bold *_italics_* ~strike~ ``code`` ````preformatted```` >quote



jandlsalazeradams@gmail.com



[jason-sa](#)



[in/jason-salazer-adams](#)

Appendix

N-gram similarity

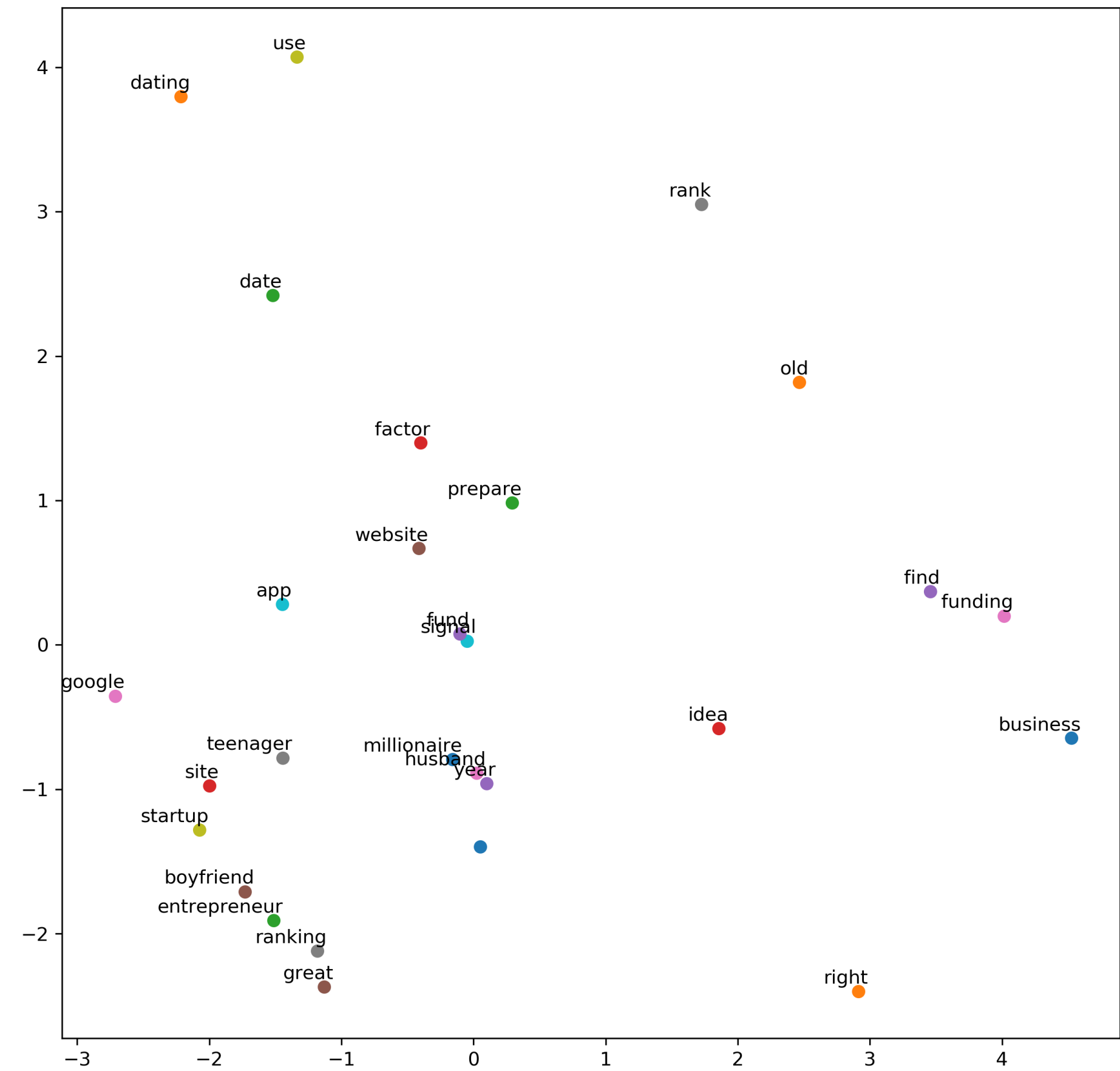
- Are the two questions using the same word or phrases?
- Sentence diagram showing n-gram similarity

Distance Distribution

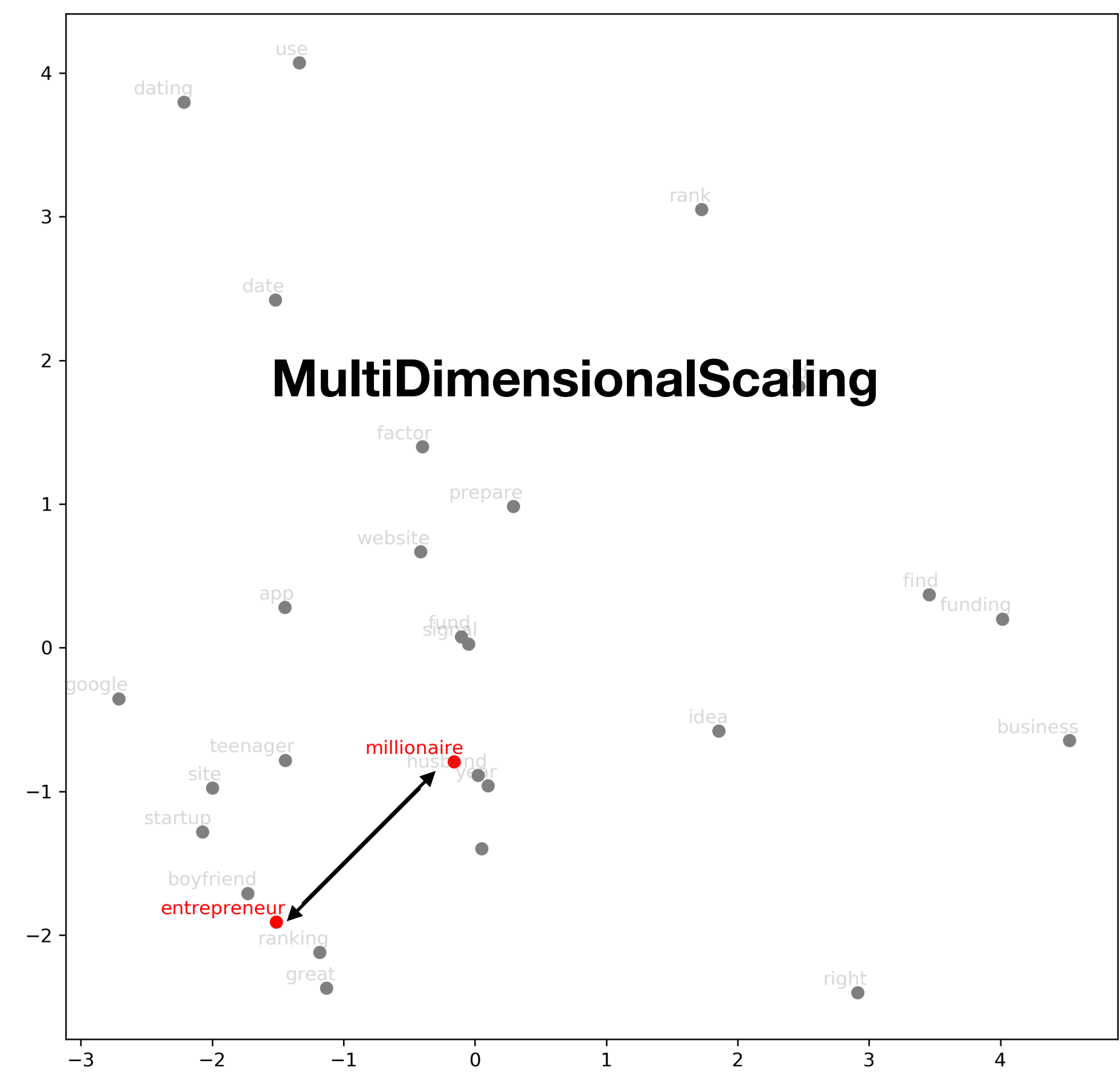
- Project 300d vectors into 2d space
- Highlight the similar words and how they are close
- Highlight the distance distribution in each question is similar

Distance Distribution - Parallel coordinates

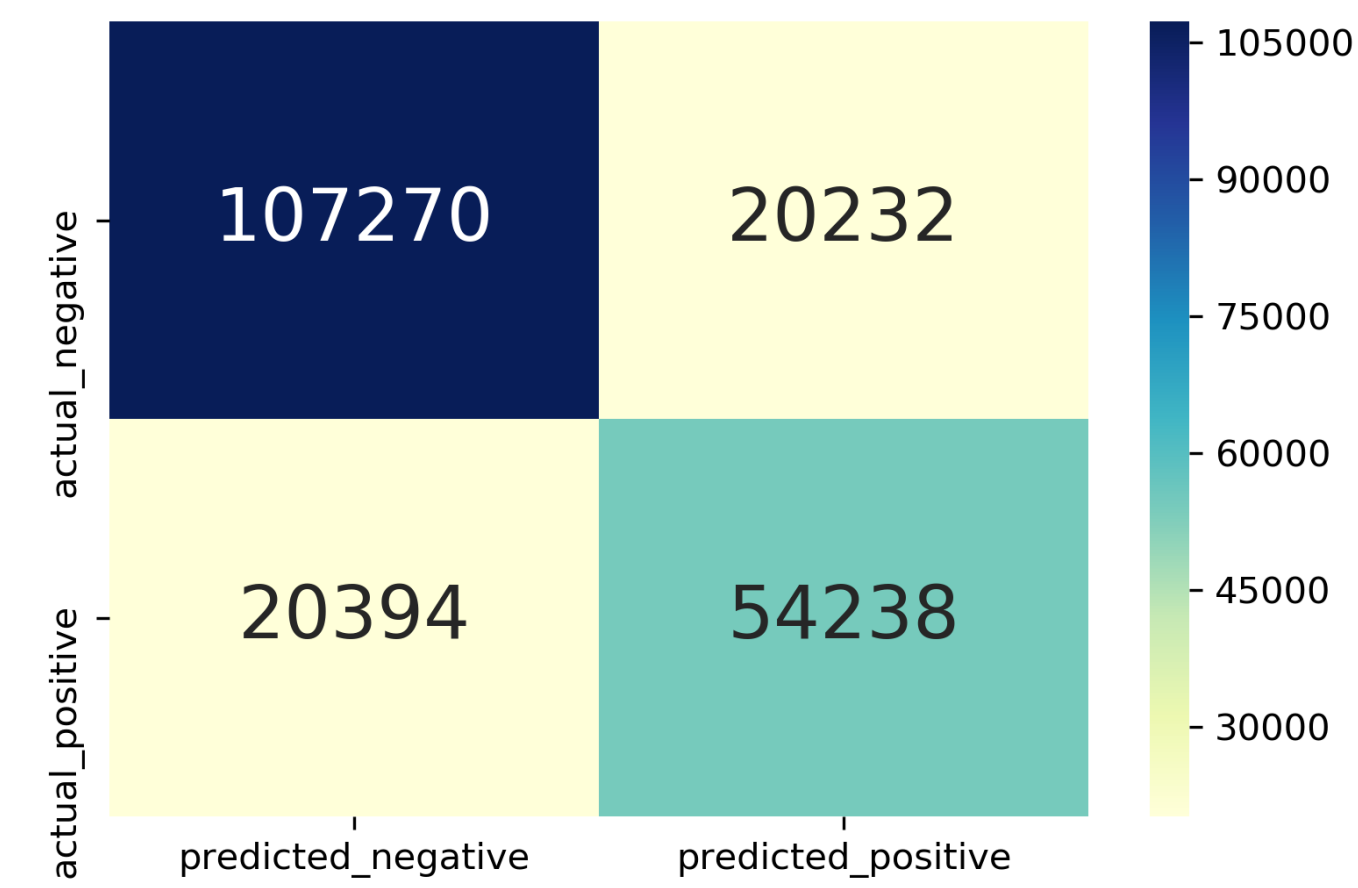
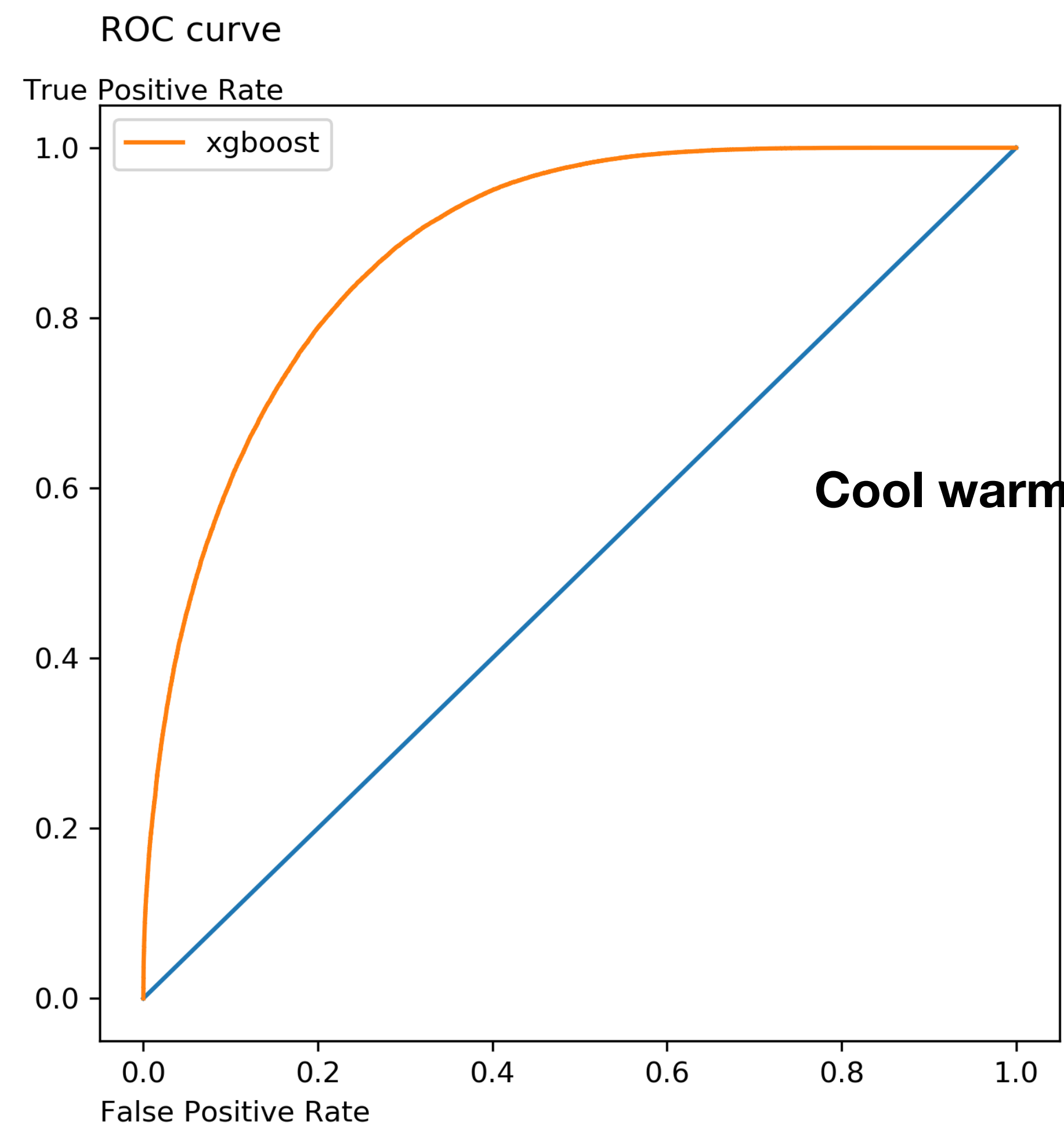
AdjustText



Distance Distribution



Results



Best model balances
sensitivity and specificity