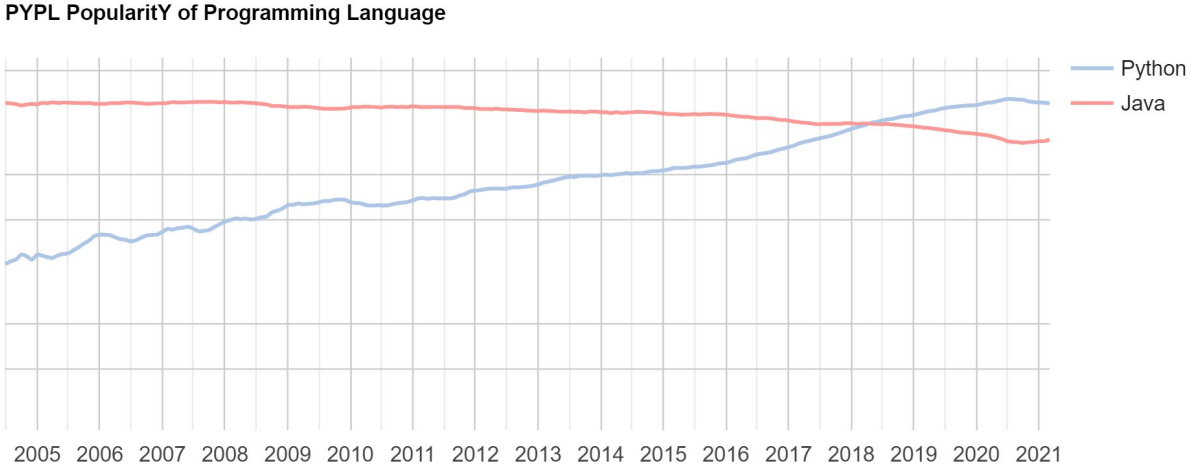
Evaluating Difficulties in Learning Python

Group 0:

By Mahmoud Maarouf, Vidya Kanekal, Hamed Mojtahed, Kevin Anderson, and Songlin Chen

Motivation



“Python grew the most in the last 5 years (-0.4%) and Java lost the most (13.7%)”

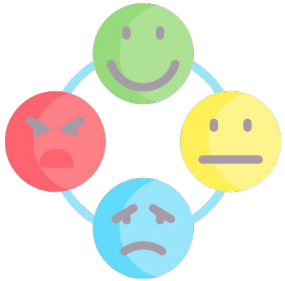
Worldwide, Feb 2021 compared to a year ago:

Rank	Change	Language	Share	Trend
1		Python	30.06 %	+0.3 %
2		Java	16.88 %	-1.7 %
3		JavaScript	8.43 %	+0.4 %
4		C#	6.69 %	-0.6 %
5	↑	C/C++	6.5 %	+0.5 %

Top 5 Languages by the PYPL Metric

Objective

Find gaps in knowledge

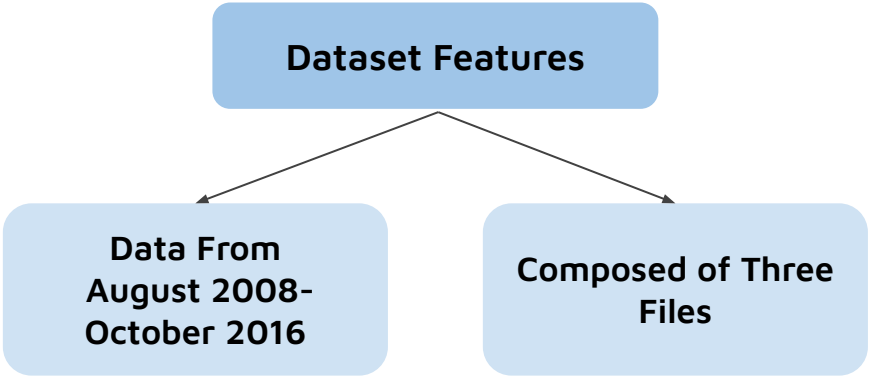


Explore sentiment of different topics

Look at trends over time

Dataset Overview

"Python Questions From Stack Overflow" Dataset



Questions.csv
Answers.csv
Tags.csv

Dataset Overview

Id	OwnerUserId	CreationDate	Score	Title	Body
469	147	2008-08-02T15:11:16Z	21	How can I find the full path to a font from its display name on a Mac?	I am using the Photoshop's javascript API to find the fonts in a given PSD. Given a font ...
502	147	2008-08-02T17:01:58Z	27	Get a preview JPEG of a PDF on Windows?	I have a cross-platform (Python) application which needs to generate a JPEG preview of the first ...
535	154	2008-08-02T18:43:54Z	40	Continuous Integration System for a Python Codebase	I'm starting work on a hobby project with a python codebase and would like to set up some form of...

Snapshot of Questions.csv file

Id	OwnerUserId	CreationDate	ParentId	Score	Body
497	50	2008-08-02T16:56:53Z	469	4	open up a terminal (Applications->Utilities->Terminal) and type this in: <pre><code>...
518	153	2008-08-02T17:42:28Z	469	2	I haven't been able to find anything that does this directly. I think you'll have to iterate thr...
536	161	2008-08-02T18:49:07Z	502	9	You can use ImageMagick's convert utility for this, see some examples in <a href="https://web.arc...

Snapshot of Answers.csv file

Dataset Overview

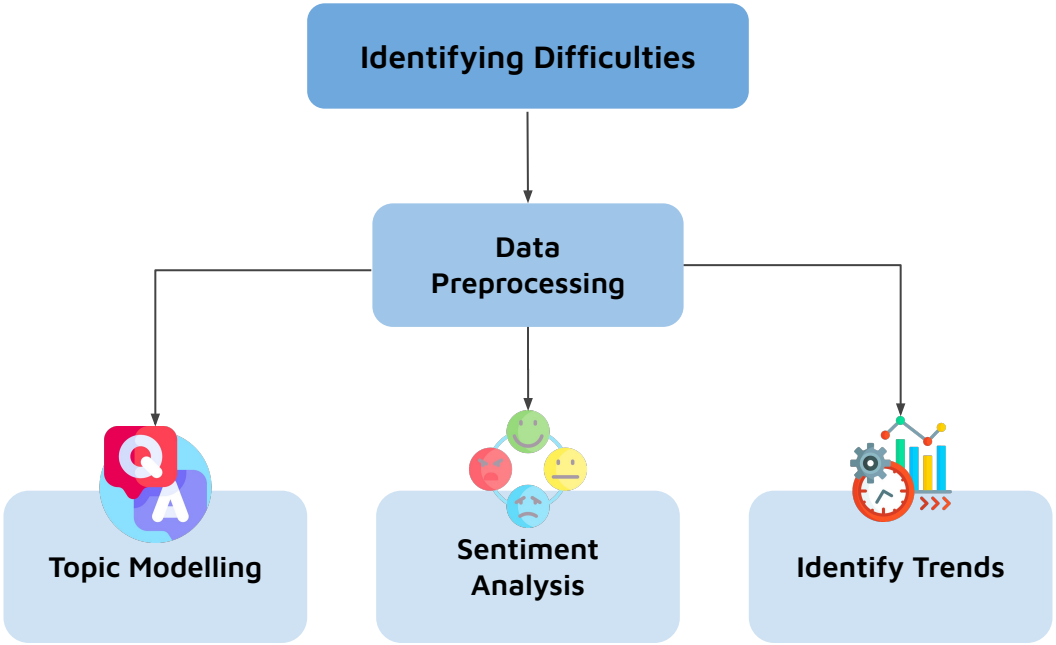
Id	OwnerUserId	CreationDate	Score	Title	Body
469	147	2008-08-02T15:11:16Z	21	How can I find the full path to a font from its display name on a Mac?	I am using the Photoshop's javascript API to find the fonts in a given PSD. Given a font ...
502	147	2008-08-02T17:01:58Z	27	Get a preview JPEG of a PDF on Windows?	I have a cross-platform (Python) application which needs to generate a JPEG preview of the first ...
535	154	2008-08-02T18:43:54Z	40	Continuous Integration System for a Python Codebase	I'm starting work on a hobby project with a python codebase and would like to set up some form of...

Snapshot of Questions.csv file

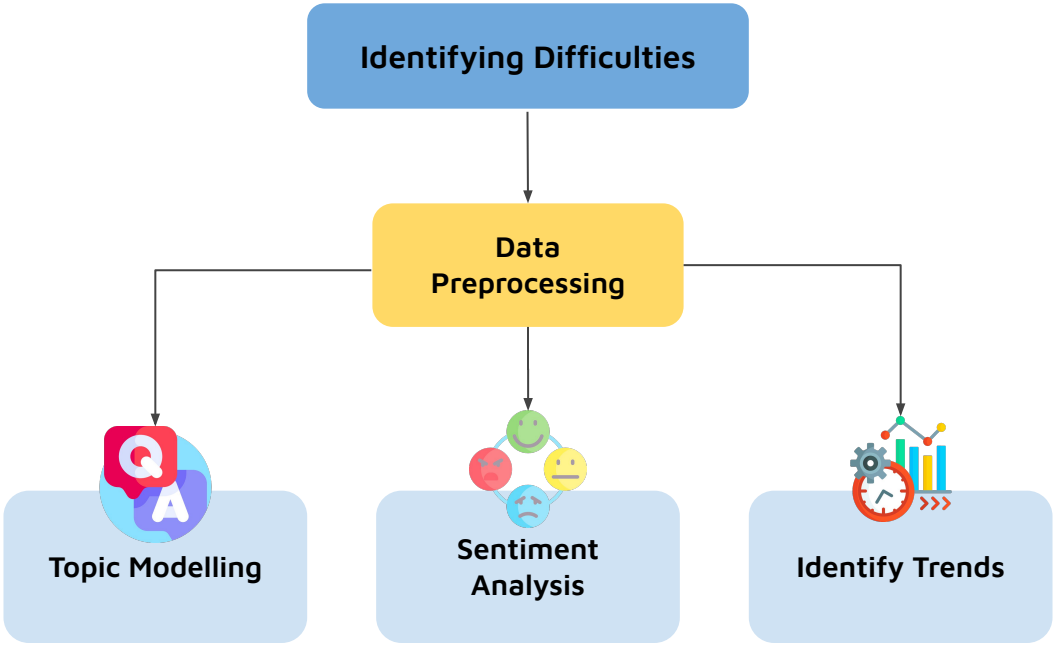
Id	Tag
469	python
469	osx
469	fonts
469	photoshop
502	python
502	windows
502	image
502	pdf
535	python
535	continuous-integration

Snapshot of Tags.csv file

Methodology



Methodology



Data Preprocessing

Start with HTML-formatted text

“ What are the difficulties in learning Python? ”

Data Preprocessing

Start with HTML-formatted text

“ What are the difficulties in learning Python? ”

Strip tags and punctuation

“what are the difficulties in learning python”

Data Preprocessing

Start with HTML-formatted text

“ What are the difficulties in learning Python? ”

Strip tags and punctuation

“what are the difficulties in learning python”

Removal of stop words

“difficulties learning python”

Data Preprocessing

Start with HTML-formatted text

“ What are the difficulties in learning Python? ”

Strip tags and punctuation

“what are the difficulties in learning python”

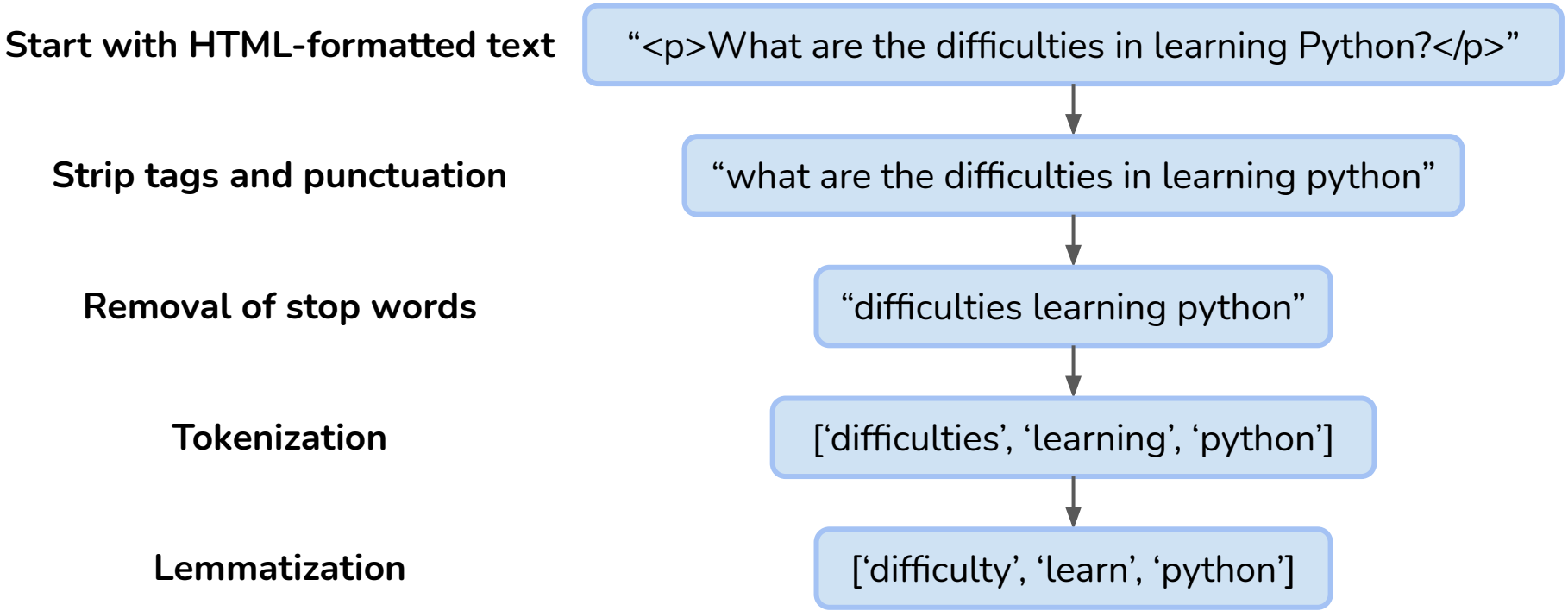
Removal of stop words

“difficulties learning python”

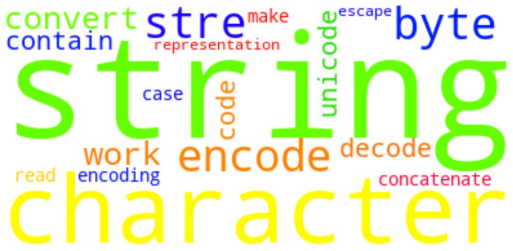
Tokenization

[‘difficulties’, ‘learning’, ‘python’]

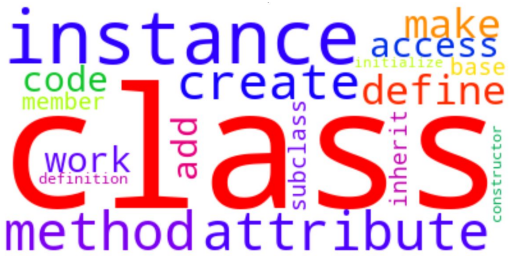
Data Preprocessing



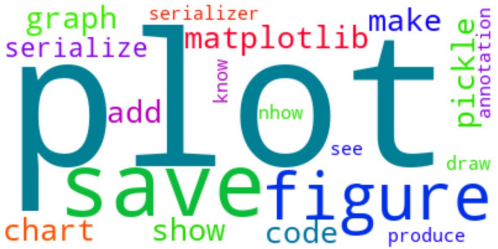
LDA Topic Modeling



Working with Strings

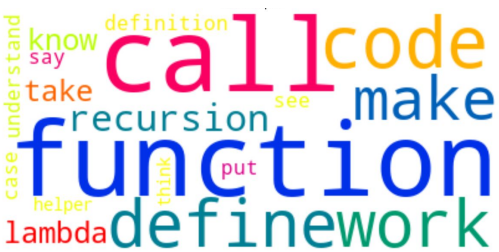


Object Oriented Programming



Data Visualization

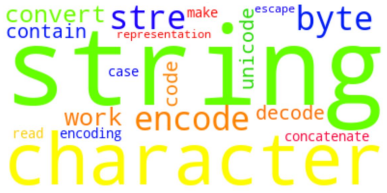
Loops and Iteration



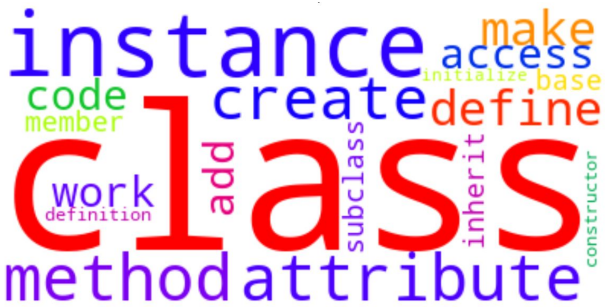
Functions

Database

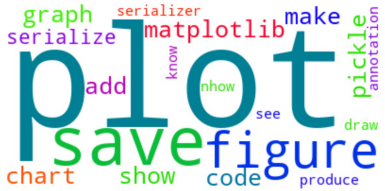
LDA Topic Modeling



Working with Strings

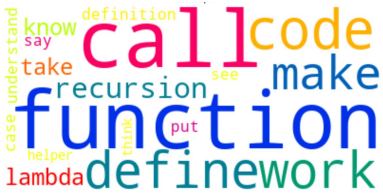


Object Oriented Programming



Data Visualization

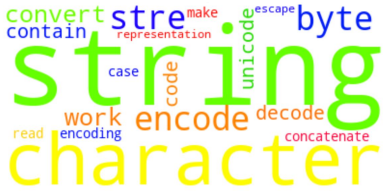
Loops and Iteration



Functions

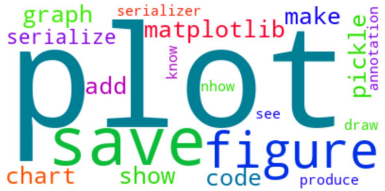
Database

LDA Topic Modeling



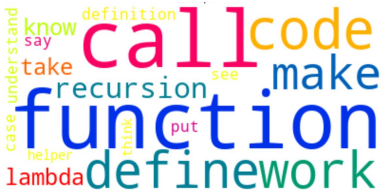
Working with Strings

Object Oriented Programming



Data Visualization

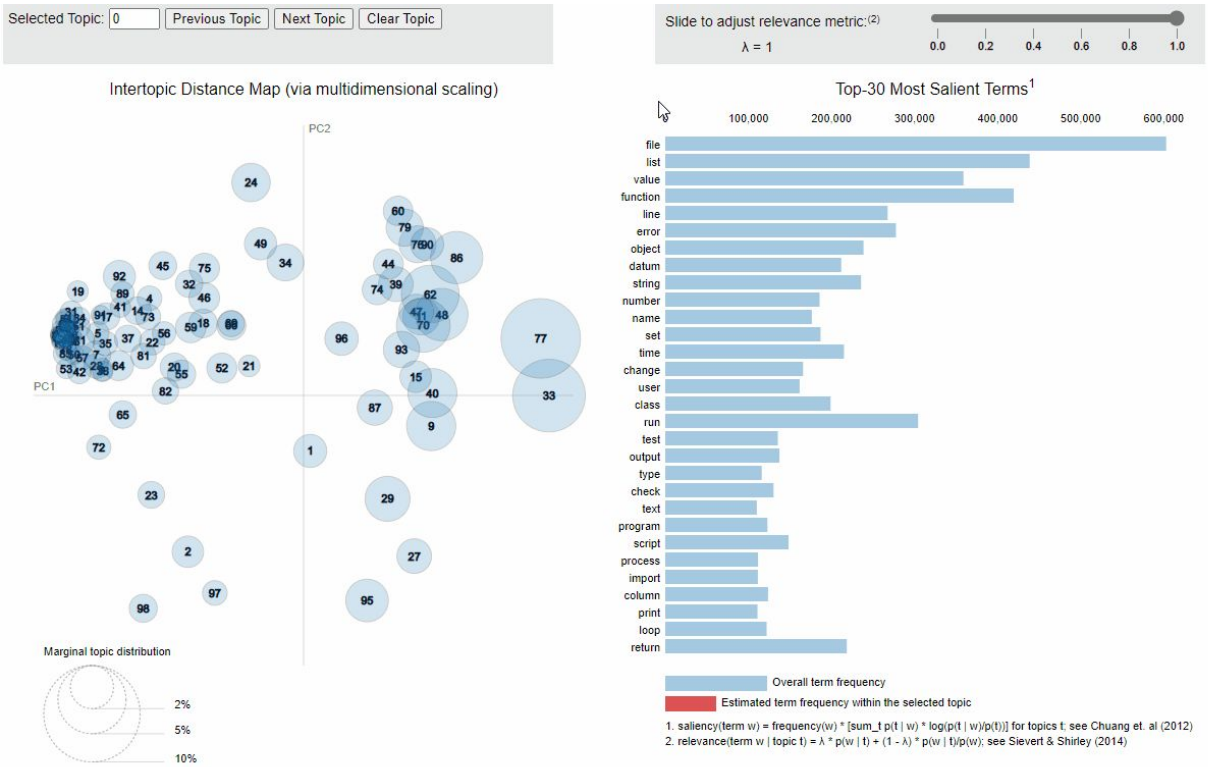
Loops and Iteration



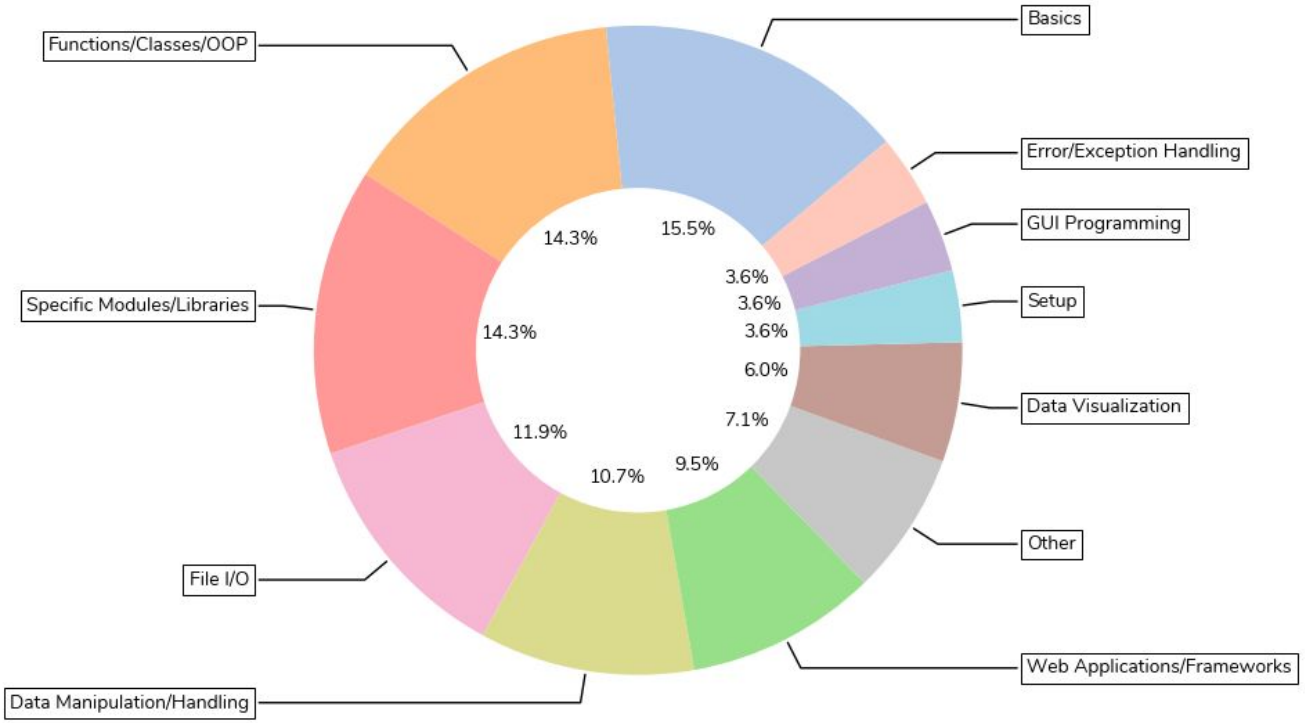
Functions

Database

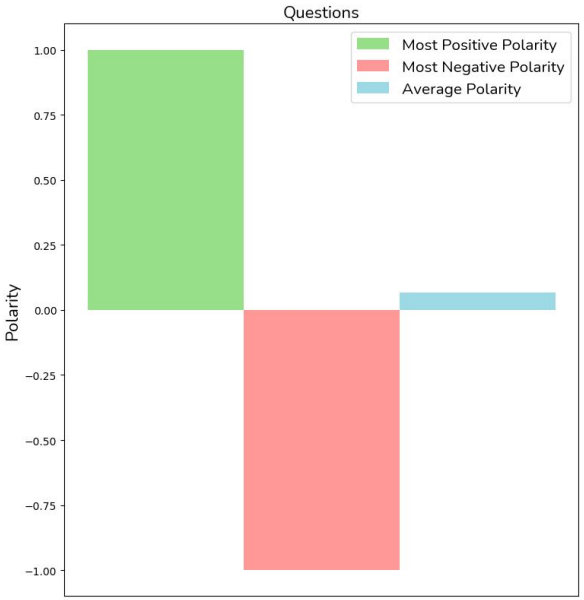
Distribution of Topics

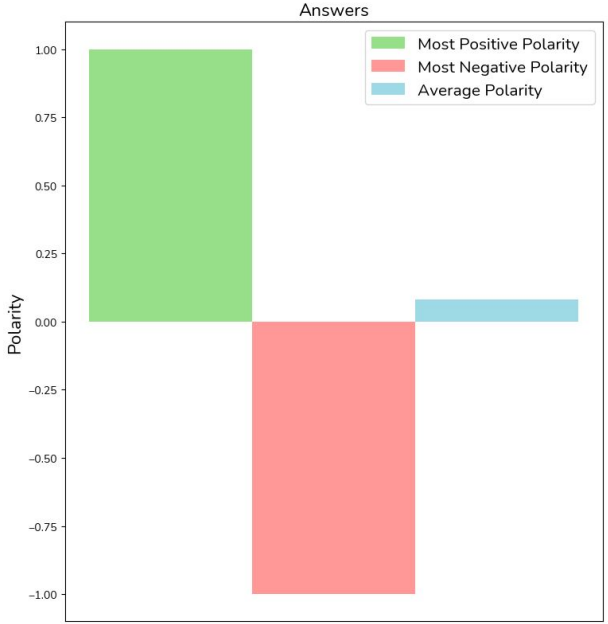


Frequency of Topics



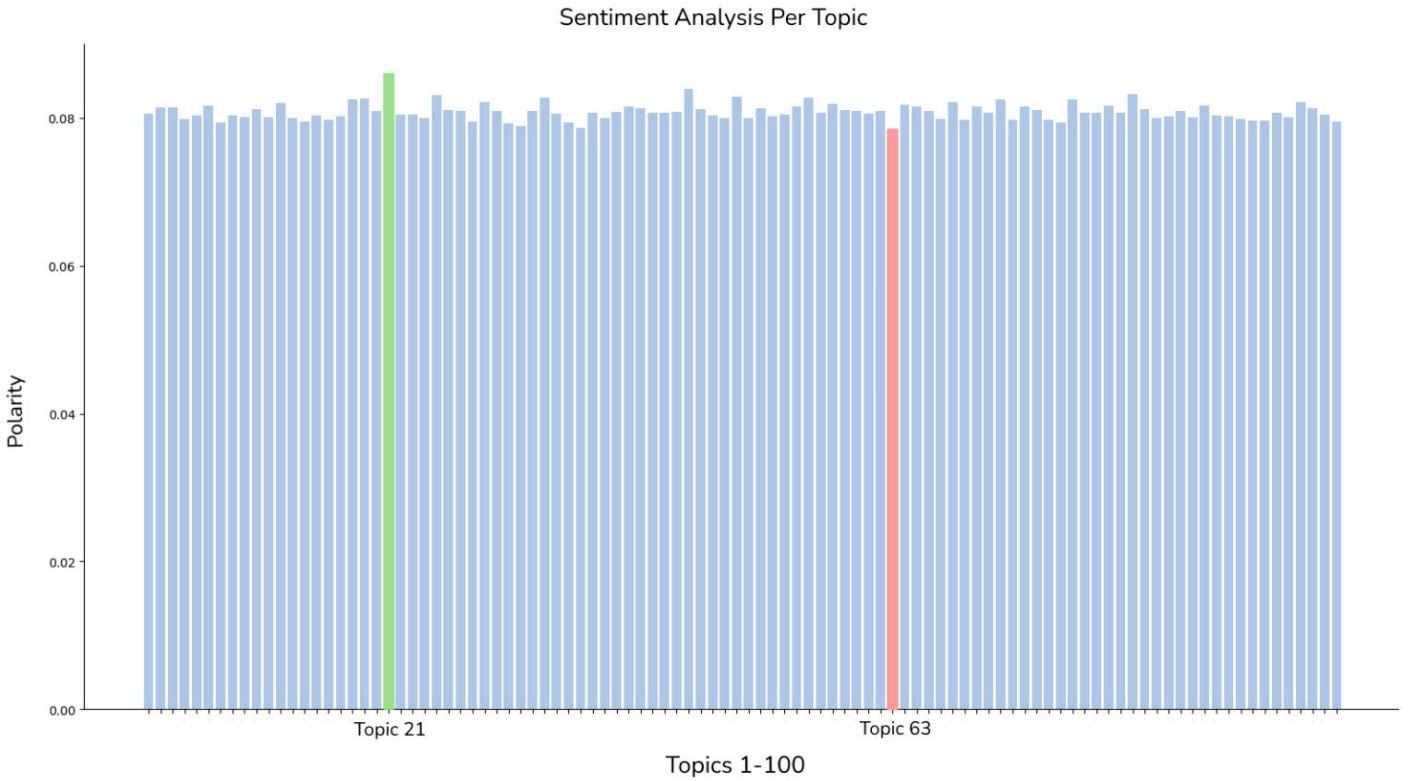
Sentiment Analysis on Questions and Answers





* 1: most positive, -1: most negative, 0: neutral

Sentiment Analysis on Topics

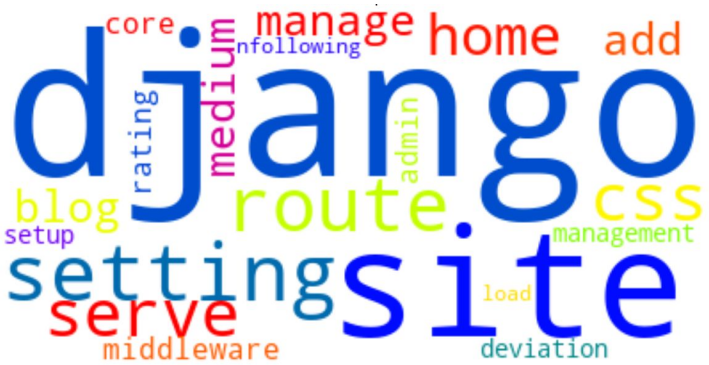


Sentiment Analysis on Topics

Most Positive: Topic 21

Programming Basics

Most Negative: Topic 63



Web Development

Identifying Trends

Specific Fields

Common Libraries

Additional Tags

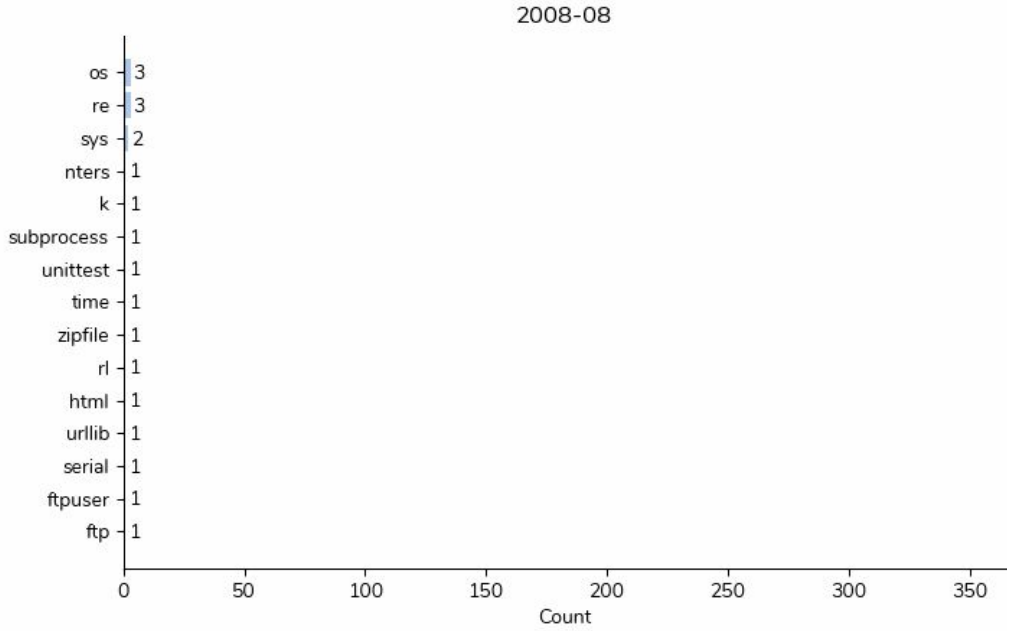
Popular Tools

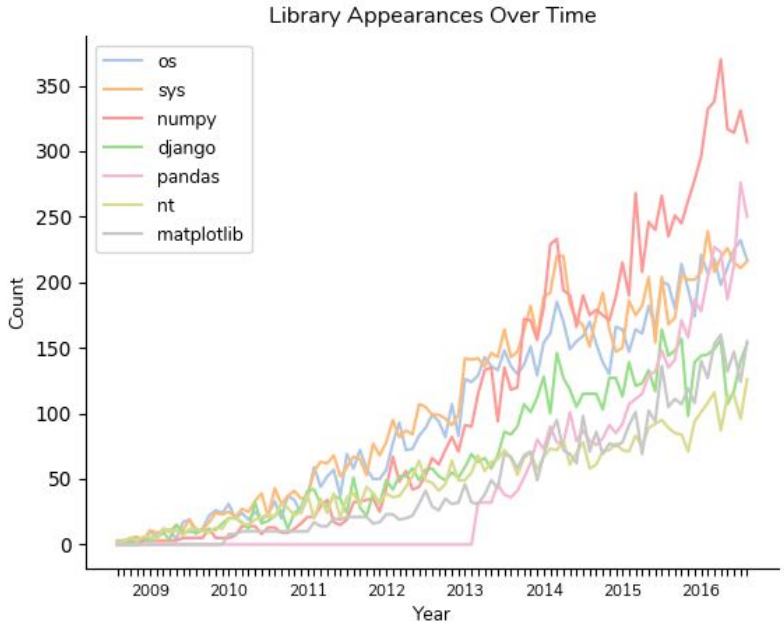
Operating Systems

IDE's

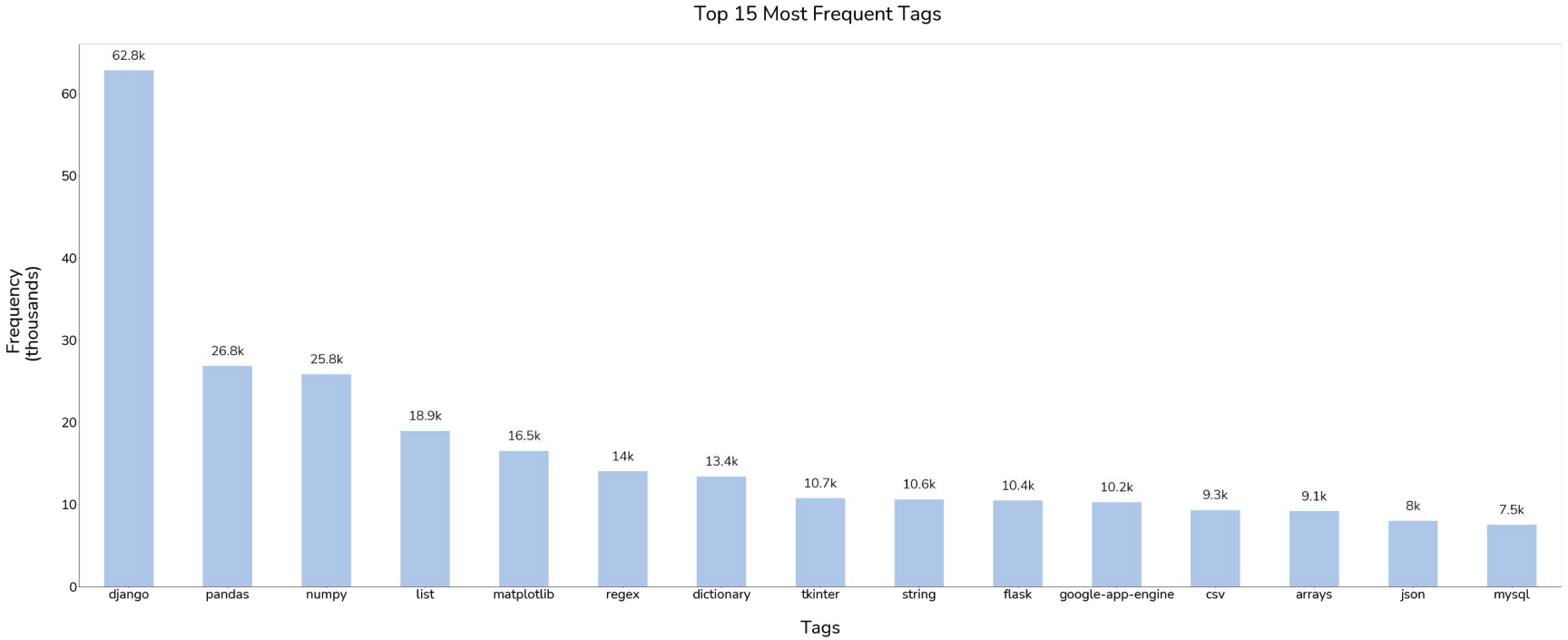
Package Managers

Most Frequent Libraries

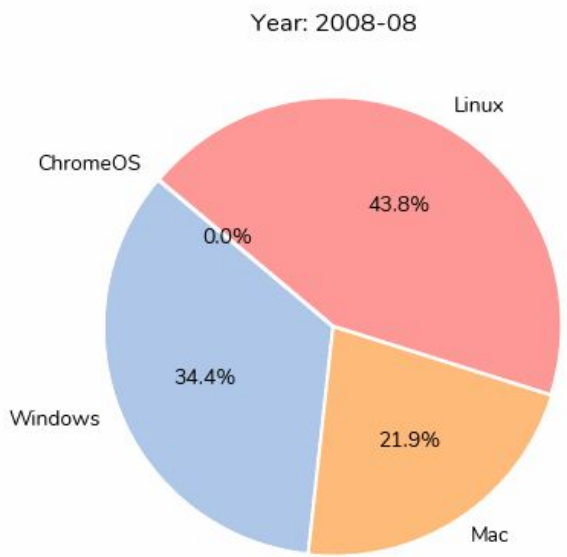


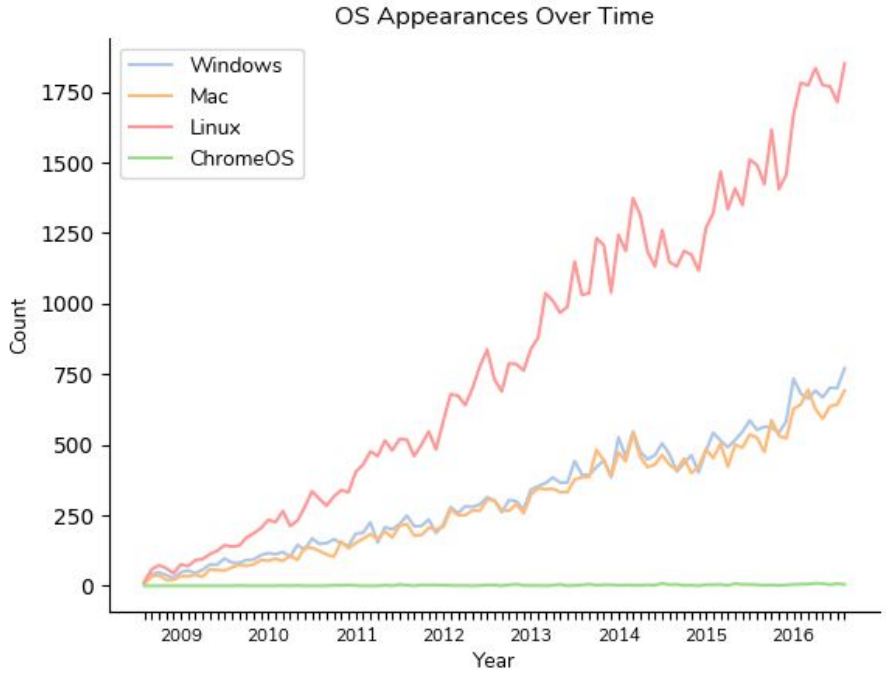


Most Frequent Tags

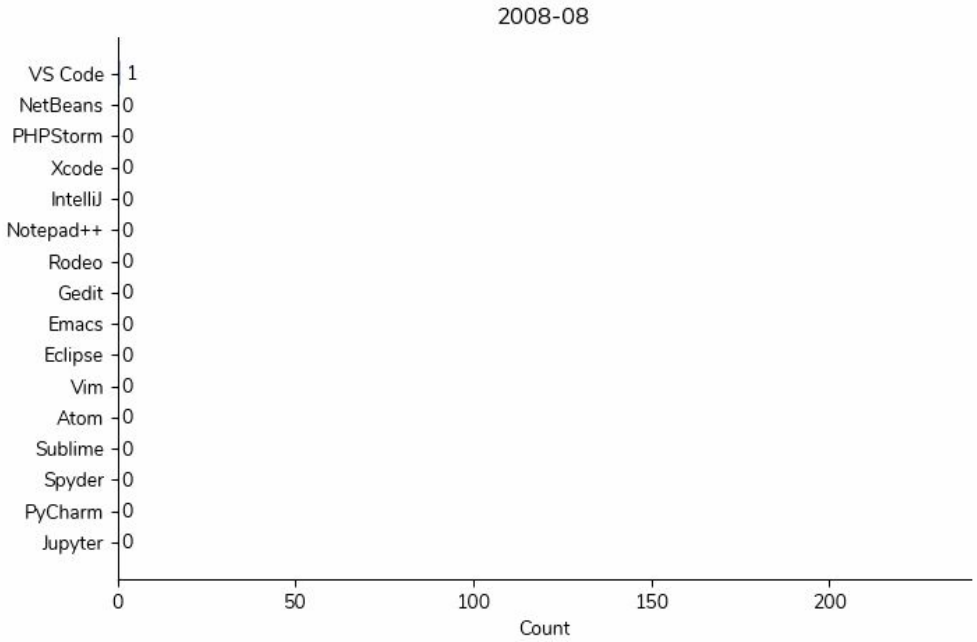


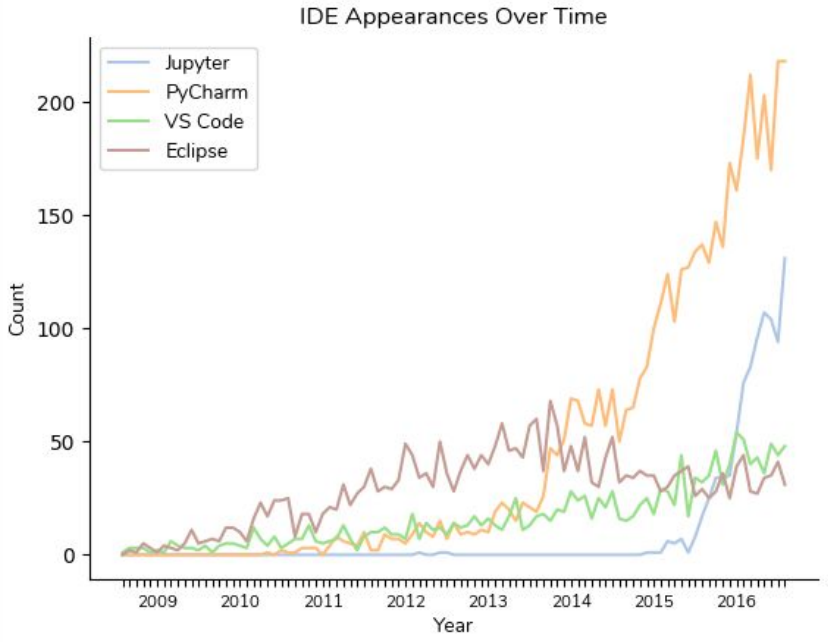
Most Frequent Operating System



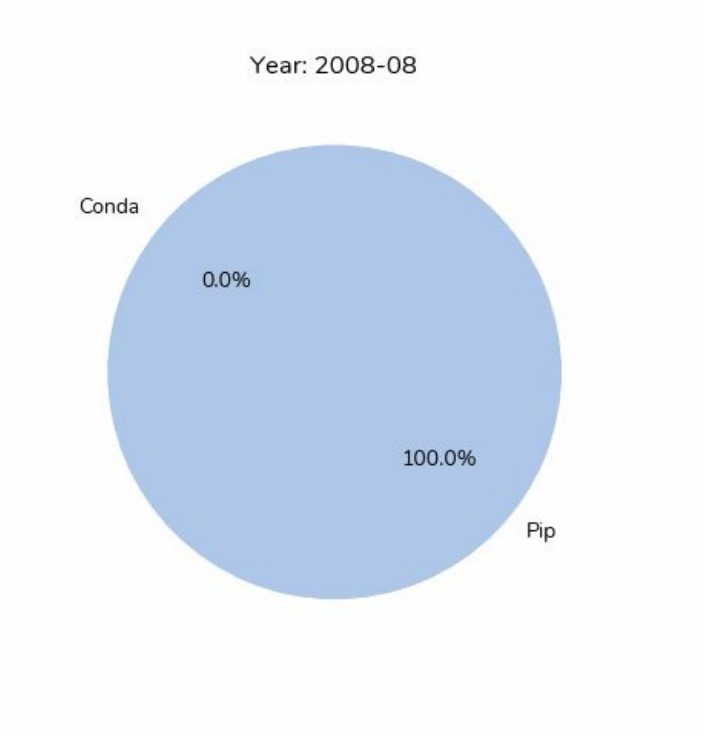


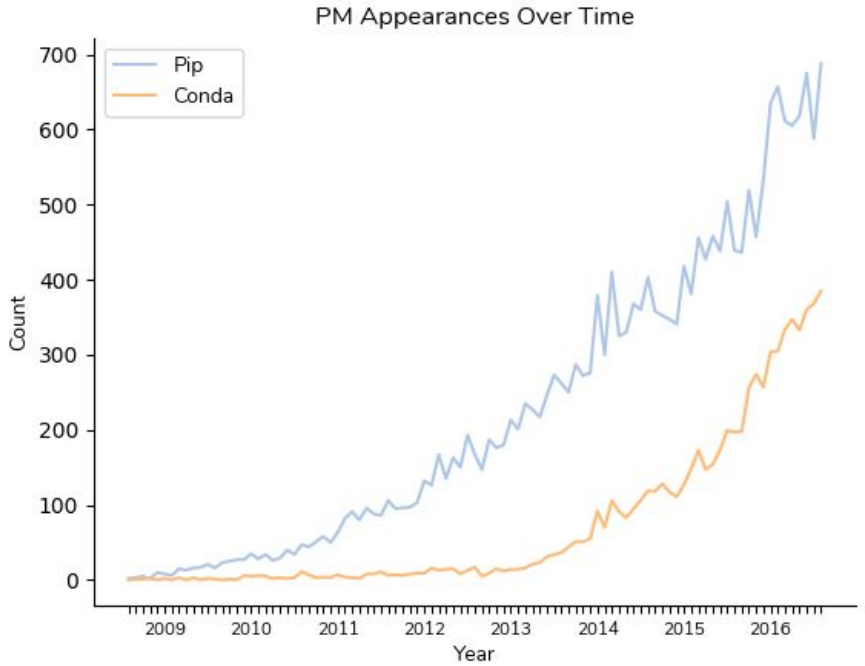
Most Frequent IDE





Package Managers - Pip vs Conda

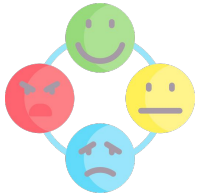




Conclusion

Most Common Topics

Basics
Functions/OOP
Specific Libraries



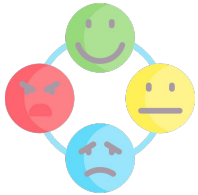
Sentiments

Trends over Time

Conclusion

Most Common Topics

Basics
Functions/OOP
Specific Libraries



Sentiments

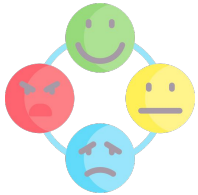
Most Negative - Web development
Most Positive - Basics

Trends over Time

Conclusion

Most Common Topics

Basics
Functions/OOP
Specific Libraries



Sentiments

Most Negative - Web development
Most Positive - Basics

Trends over Time

Top Fields- Numpy/Django
Top Tools - Linux/PyCharm/Pip

Next Steps

Updated Dataset

Guided LDA

Any Questions?