

Third-party Library Recommendations for Python Developers using Software Analysis Techniques

SMILESENG 2022

24/6/2022 – Córdoba

Pedro P. García Pozo – Aurora Ramírez – José Raúl Romero
Dept. Computer Science – University of Córdoba



UNIVERSIDAD
DE
CÓRDOBA

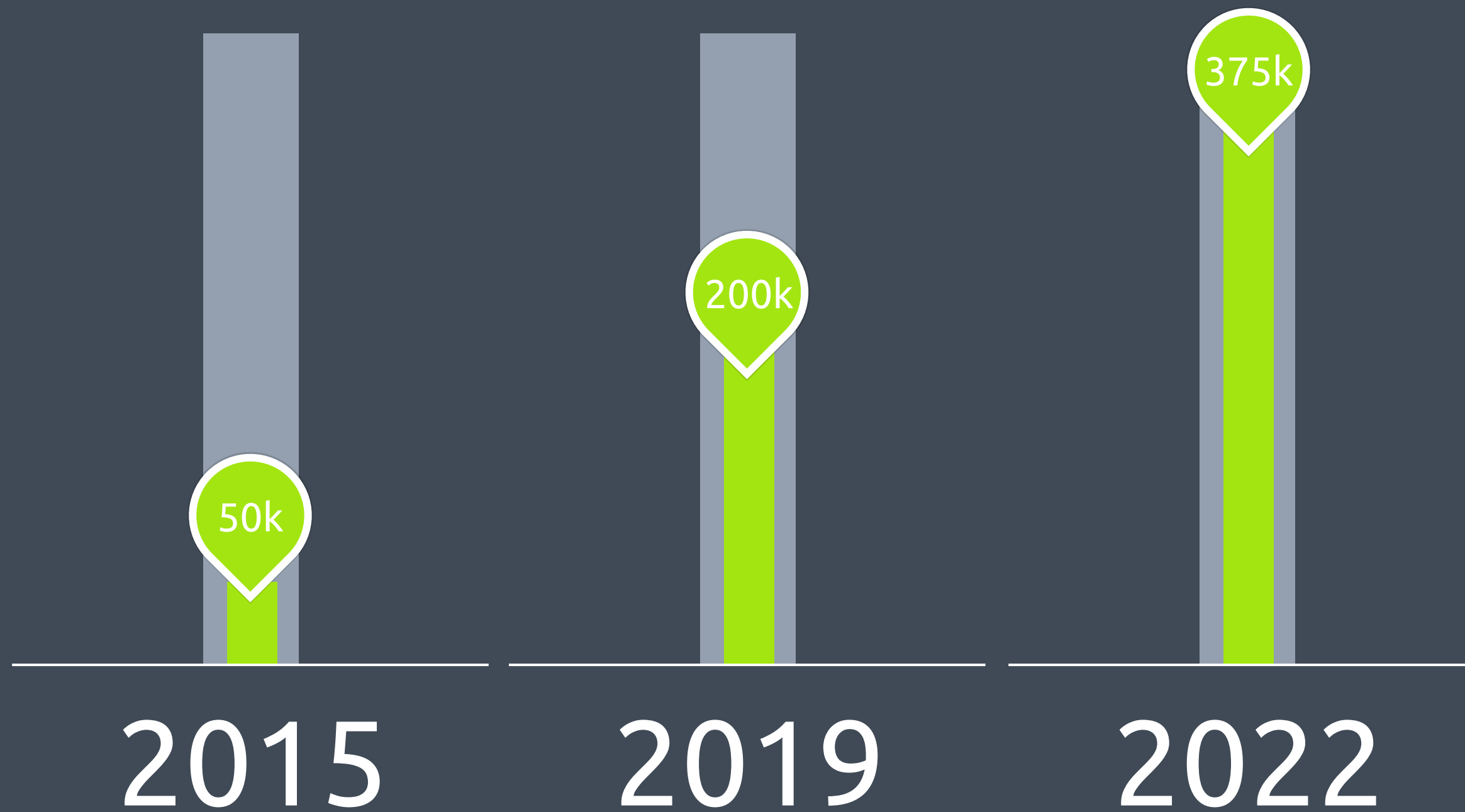
1

Introduction

Python™

Introduction

■ Number of Packages available in PyPi



Vast and Dynamic Ecosystem

Hard to select the most suitable library

- Development Support
- Dependencies
- Compatibility

Library Recommendation

Has already been widely studied in Java Systems

Our Vision



Our research vision is that current recommender systems do not exploit all the potential knowledge hidden in software repositories, and still require additional steps to support developers in the effective integration of the recommended libraries.



2

Research Objectives

Research objectives



Analyze library ecosystem 01

- Analyze the use of Python libraries in repositories to extract hidden knowledge to make recommendations.



Enhance knowledge base 02

- Enhance the code knowledge base to be able to adapt the recommendations to the project context.



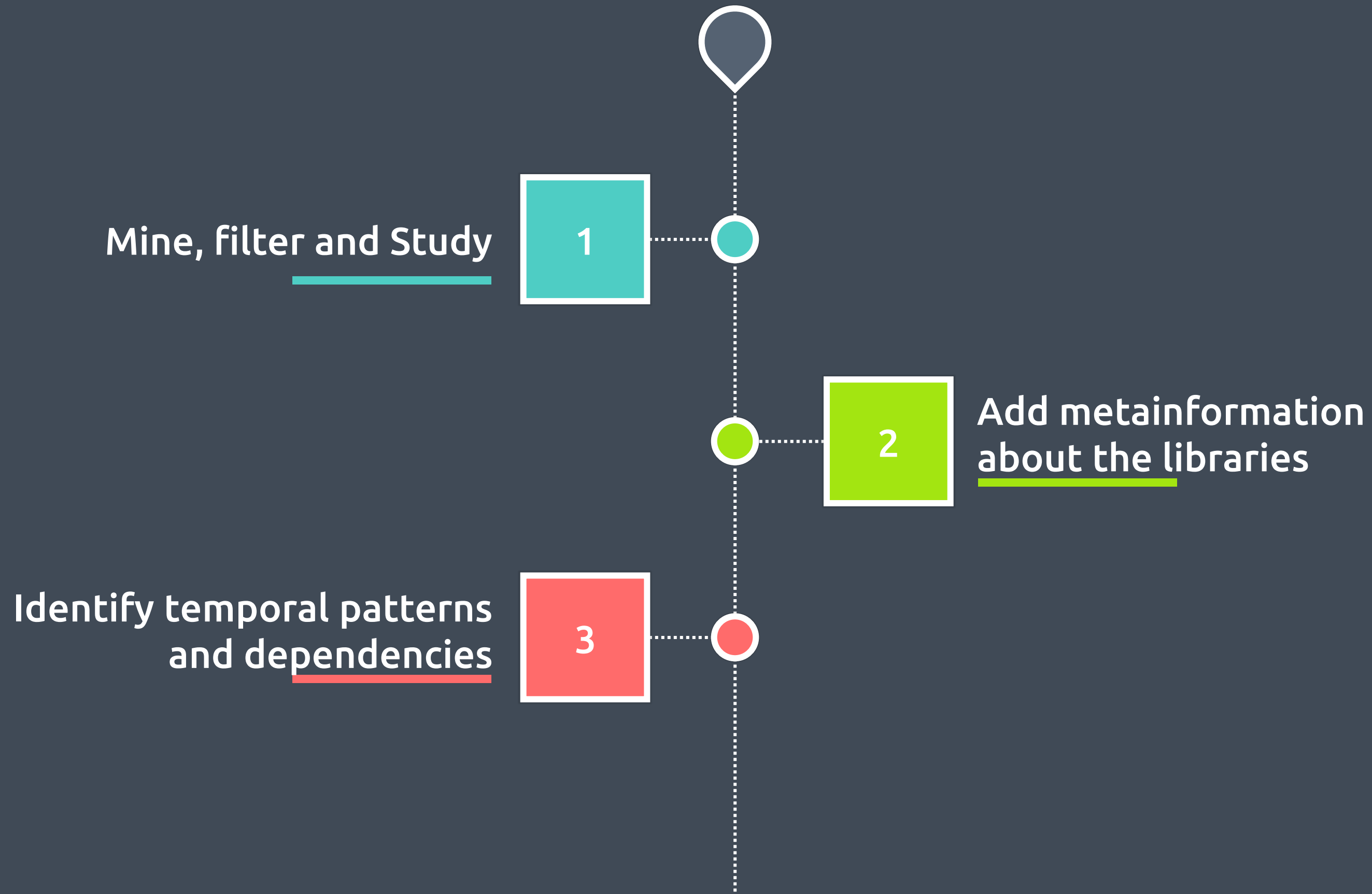
Assist in library integration 03

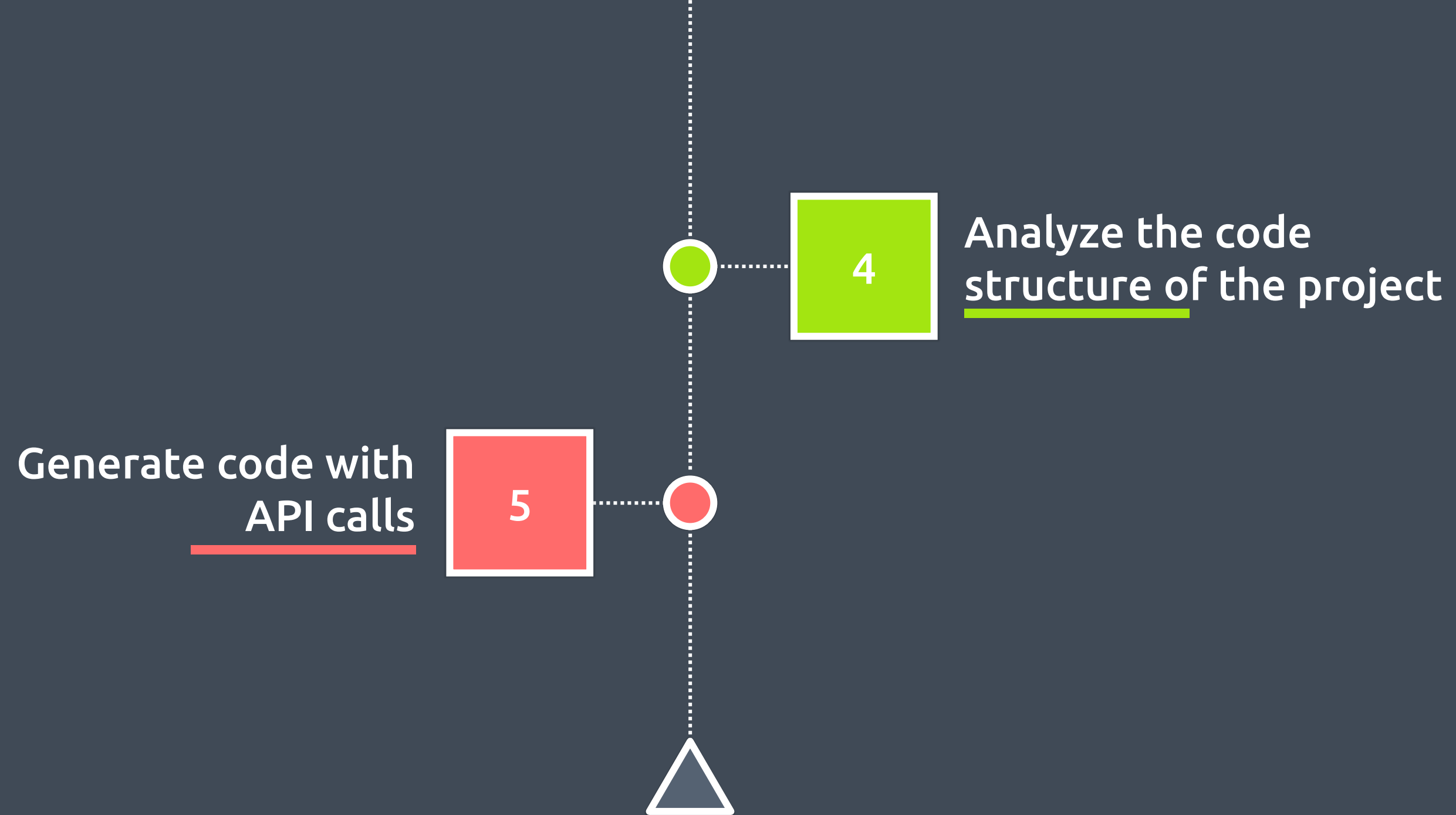
- Assist the developers regarding how their code should be combined or replaced by API calls to the library.

3

Overview of the
approach

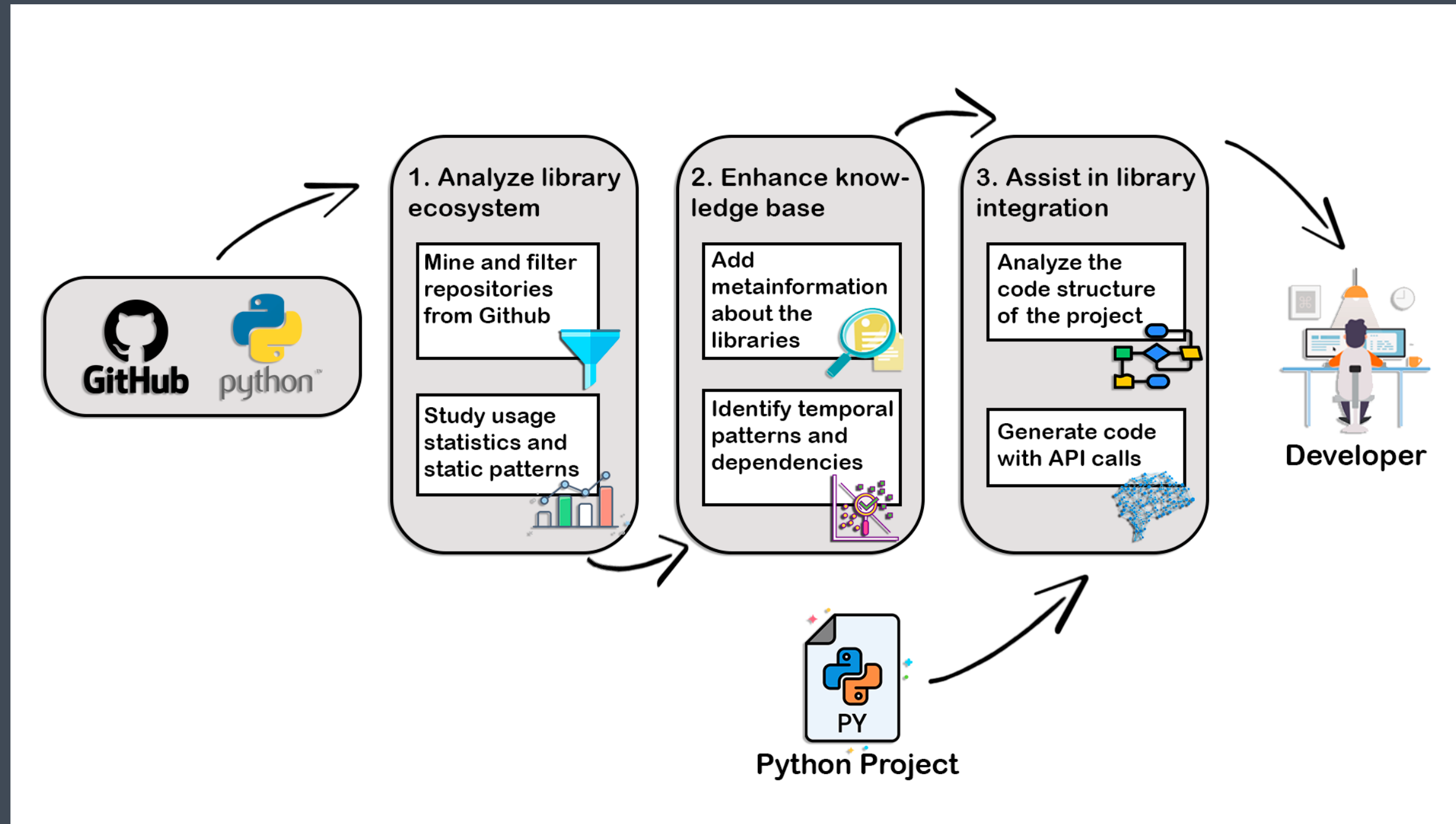
Overview of the approach





Effective recommendation

A picture is worth a thousand words





4

First Results

First results



+2k
Libraries

Only appear 1 time

15.5
Libraries

Of average and standard deviation (24.6) clearly indicate that the usage of libraries across repositories presents a skewed distribution

10
Libraries

Of them (requests, numpy, sphinx, sphinx, six, pyyaml, scipy, python-dateutil, jinja2, matplotlib and pytz) are included in 100 repositories or more

232
Libraries

Is the maximum number of different libraries in one project

32
Libraries

of the 3,300 libraries (0.97%) appear in 50 or more repositories

5

Conclusion and
future work



Conclusion and future work

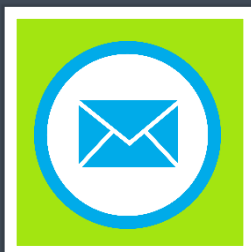
- Useful knowledge about the current adoption of Python libraries in Github repositories
We must enlarge the dataset and apply other unsupervised techniques
- IDE plug-in/extension ?





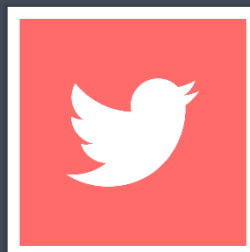
Pedro Pablo García Pozo

IT Student (Computer Science + Software Engineering)
Dept. Computer Science – University of Córdoba



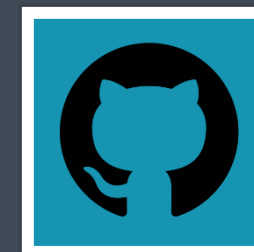
Email

i82gapop@uco.es



Twitter

twitter.com/half_twilight



Github

<https://github.com/i82gapop>

Thank you!



Any questions?