BitBucket Repository Analysis

Computer Science 594

John Duggan, Chunyan Tang, RJ Lusk

Problem Summary

Collect info on BitBucket repositories to:
- Discover opportunities for DeveloperParadise (DP) to grow
- Ensure that DP has a bright future

Objective

- **Target customers**: python developers
- **Identify** the most popular programming languages over time (2008-present).
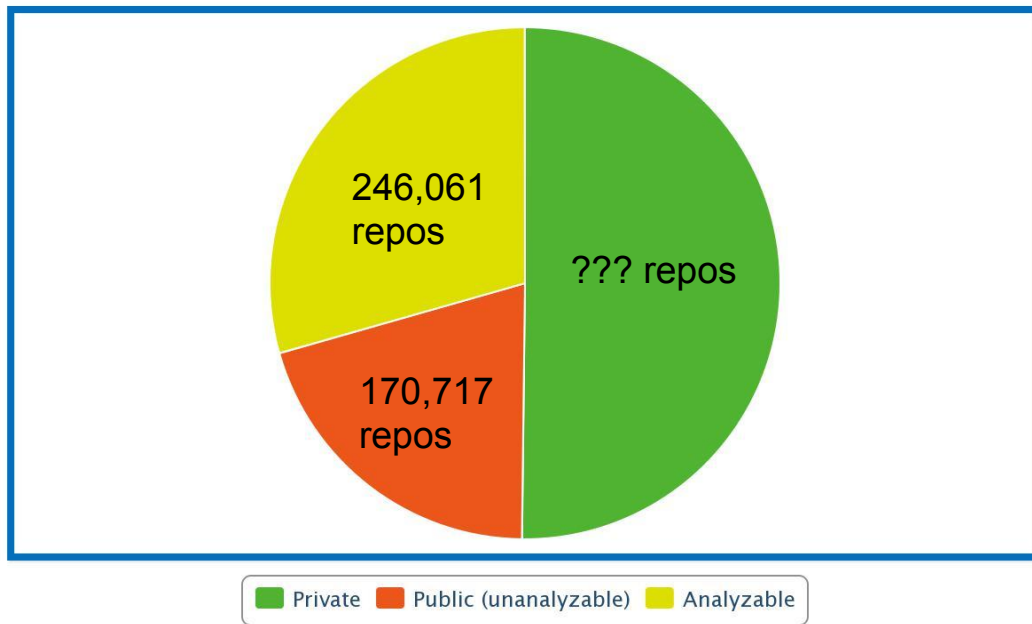- **Discover** relevant trends and patterns over time.

Methodology

- Collect all public repositories which have a language specified
- Segment the repos based upon year created
- Segment these further based upon programming language used

Analyzable Repositories
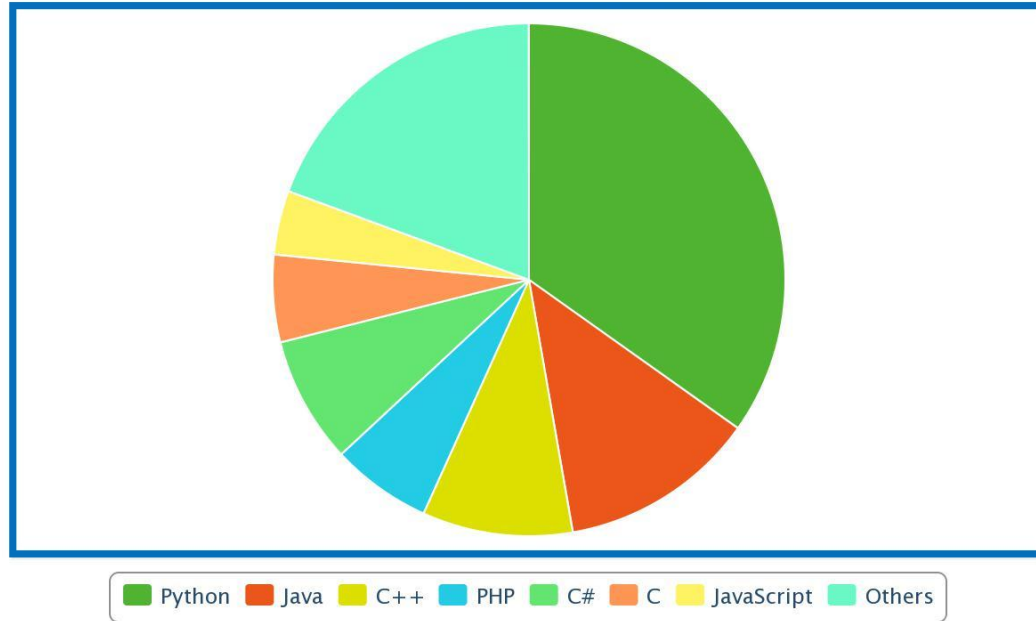

Breakdown of BB Repositories

246,061 repos

??? repos

170,717 repos

Private   Public (unanalyzable)   Analyzable

meta-chart.com

Example Distribution

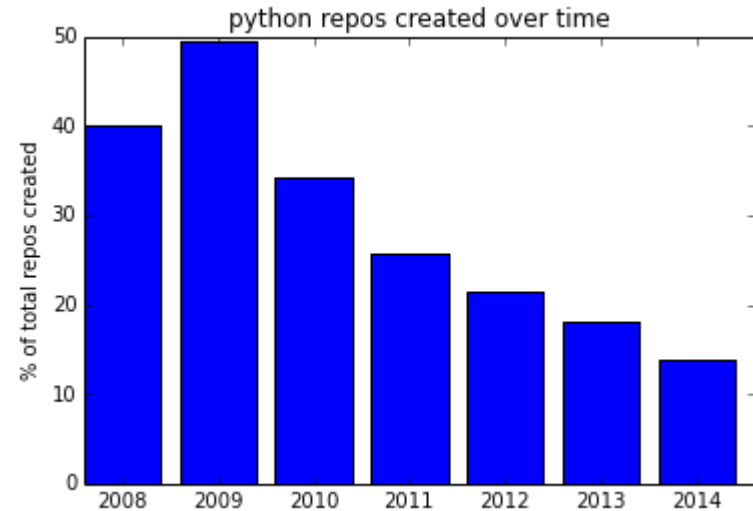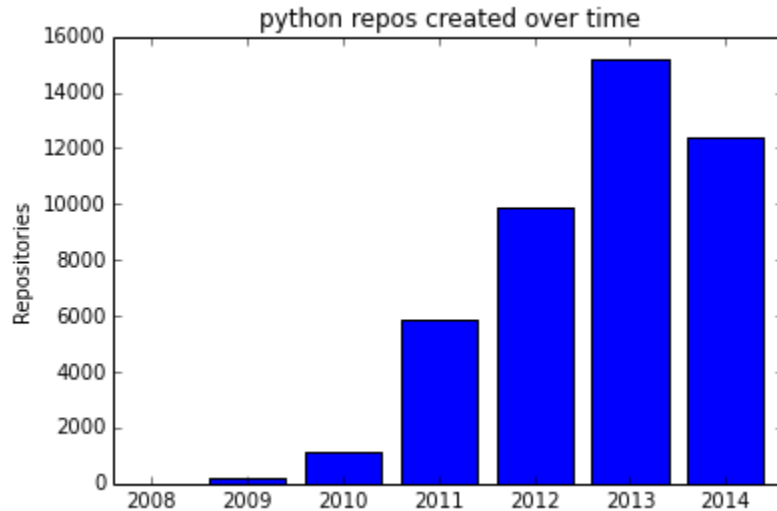Distribution of Repositories by Language (2010)

Potential Pitfalls

- Missing Data - do the missing repos have a similar PL distribution to those analyzed?
- Context - does BitBucket truly reflect the popularity of programming languages?
- Incorrect Data - is it likely for repos to have the wrong programming language specified?
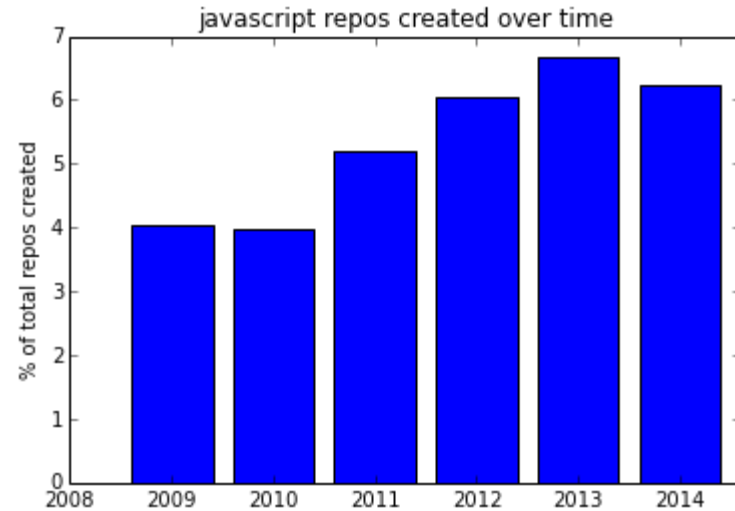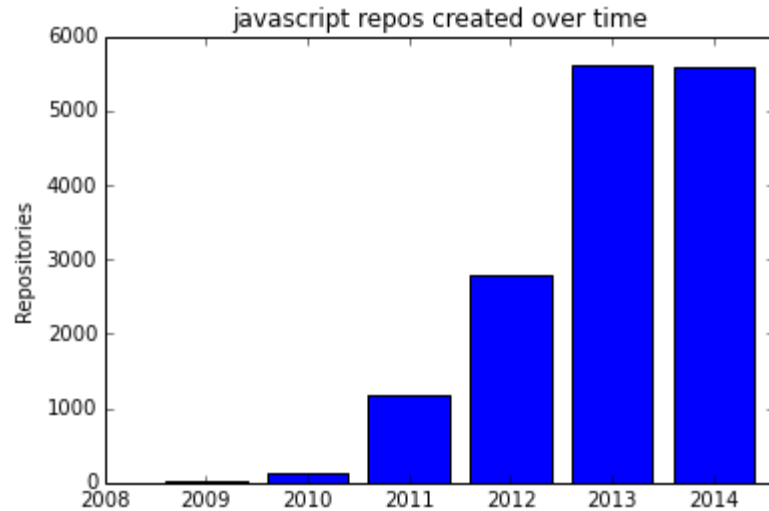
Results (Python)

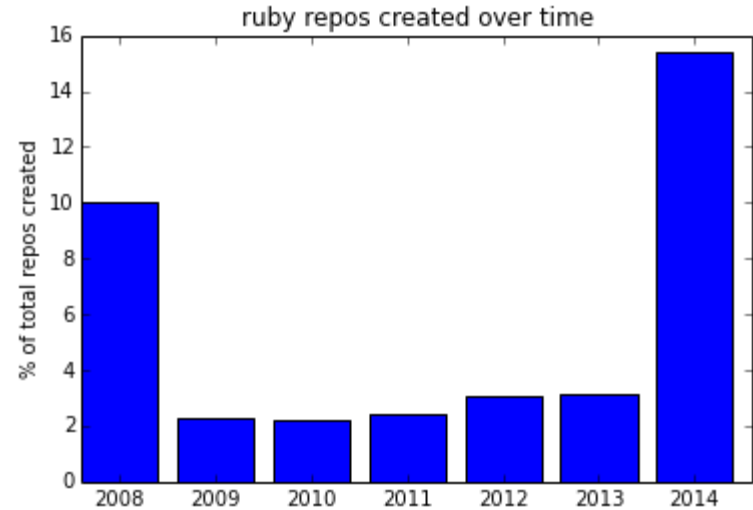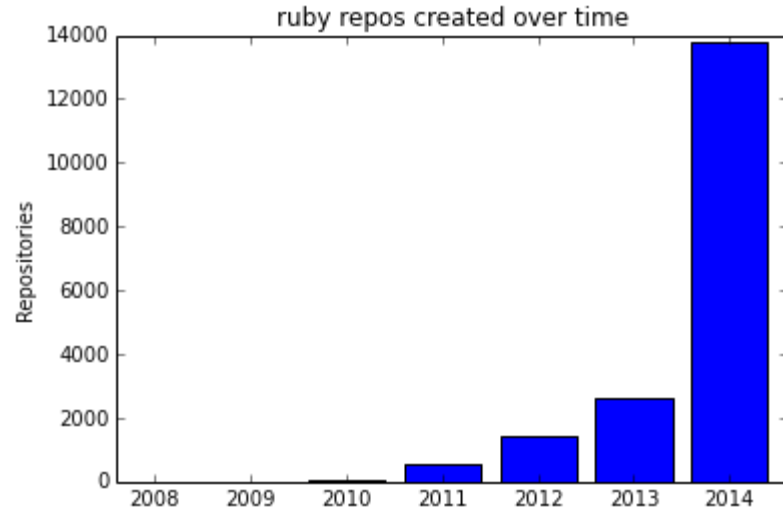
Projection
2015: 10.12%

Results (JavaScript)

Projection
2015: 6.26%

Results (Ruby)

Projection
2015: 20.52%

Conclusion

- DP's current customer base of python developers is still a strong one.

- DP should look into expanding its services to include JavaScript and Ruby developers.