



Universidade Federal do Acre
Mestrado em Ciência da Computação

Entendendo o Comportamento de Pull Requests em Python: Um estudo de cenário de Tempo de Vida e Revisores

Apresentação: Álvaro Lopes Rios



Problema e Motivação

- Revisões de PR variam muito em tempo (lifetime) e de revisor.
- Estudos prévios (Daricélio) mostram padrões entre características do PR e esses desfechos
- Objetivo: replicar/Aplicar novas técnicas a abordagem em projetos Python e comparar com a referência



Escopo e Cenários

- Linguagem-alvo: Python
- Dois cenários de análise:
- Lifetime: tempo até a decisão do PR
- Reviewer (revisor): características do revisor
- Comparação: Bases Python × Referência (valores do estudo do Daricélio)

No.	Label	Count	Weight
1	dimagi/commcare-hq	6981	6981
2	ros/rosdistro	5739	5739
3	django/django	3730	3730
4	ipython/ipython	3231	3231
5	pydata/pandas	2232	2232



Pré-Processamento

Lifetime → very short, short, medium, lengthy (MergeManyValues)

Status → accepted / rejected (MergeManyValues)

Para Revisor foi dividido a base que estava unificada em projetos Python (RemoveWithValues)



weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm.

More

Capabilities

car False

classIndex -1

delta 0.05

doNotCheckCapabilities False

lowerBoundMinSupport 0.001

metricType Confidence

minMetric 0.001

numRules 1000000

outputItemSets False

removeAllMissingCols False

significanceLevel -1.0

treatZeroAsMissing False

upperBoundMinSupport 1.0

verbose False

Open... Save... OK Cancel

Remove columns with all missing va



LIFETIME



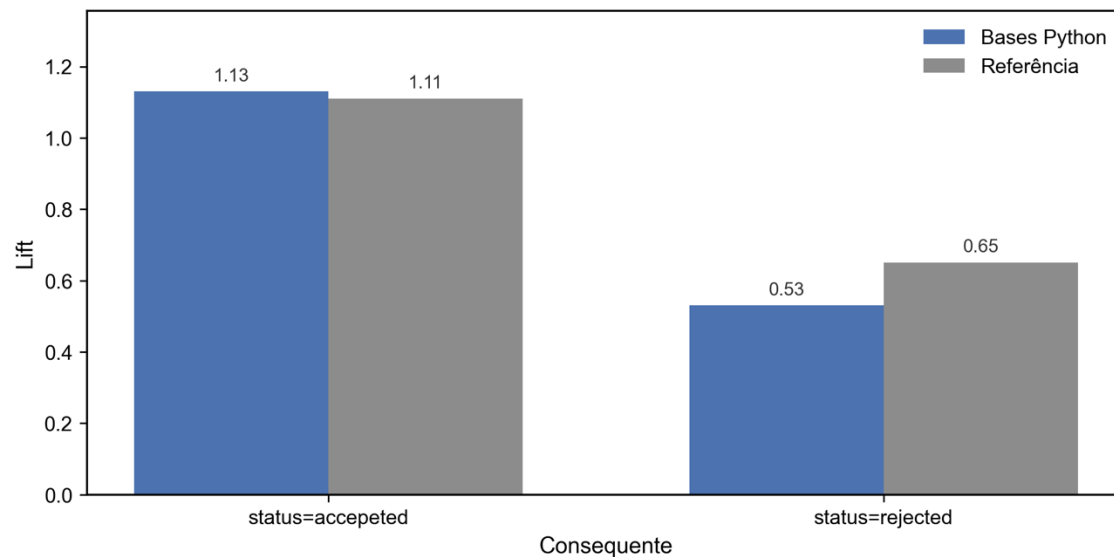
Antecedente	Consequente	Suporte	Confiança	Lift	Leverage	Conviction	Daricelio	diff	Mudou a dep	Check 1	Check 2
life_time=very short	status=accepted	0,6	0,89	1,13	0,07	1,9	1,11	-0,02	Não	Variação abaixo de 25%	Apenas Variação
life_time=short	status=accepted	0,1	0,73	0,92	-0,01	0,77	1,04	0,12	Sim	Variação abaixo de 25%	Apenas Variação
life_time=medium	status=accepted	0,04	0,66	0,83	-0,01	0,61	0,96	0,13	Não	Variação abaixo de 25%	Apenas Variação
life_time=lengthy	status=rejected	0,08	0,57	2,74	0,05	1,85	2,14	-0,6	Não	Variação entre 50% a 75%	Apenas Variação
life_time=lengthy	status=accepted	0,06	0,43	0,54	-0,05	0,36	0,66	0,12	Não	Variação abaixo de 25%	Apenas Variação
life_time=medium	status=rejected	0,02	0,34	1,64	0,01	1,2	1,13	-0,51	Não	Variação entre 50% a 75%	Apenas Variação
life_time=short	status=rejected	0,04	0,27	1,3	0,01	1,09	0,87	-0,43	Sim	Variação entre 25% a 50%	Mudança na Regra
life_time=very short	status=rejected	0,07	0,11	0,53	-0,07	0,89	0,65	0,12	Não	Variação abaixo de 25%	Apenas Variação
commitsPull_D=1 commit	life_time=very short	0,55	0,76	1,14	0,07	1,4	1,14	0	Não	Variação abaixo de 25%	Apenas Variação
commitsPull_D=some commits	life_time=very short	0,1	0,5	0,75	-0,03	0,67	0,75	0	Não	Variação abaixo de 25%	Apenas Variação
commitsPull_D=many commits	life_time=lengthy	0,03	0,39	2,84	0,02	1,4	1,92	-0,92	Não	Variação de 75% a 100%	Apenas Variação
commitsPull_D=many commits	life_time=very short	0,02	0,25	0,37	-0,03	0,44	0,51	0,14	Não	Variação abaixo de 25%	Apenas Variação
commitsPull_D=many commits	life_time=short	0,02	0,21	1,62	0,01	1,1	1,14	-0,48	Não	Variação entre 25% a 50%	Apenas Variação
commitsPull_D=some commits	life_time=short	0,04	0,21	1,6	0,02	1,1	1,24	-0,36	Não	Variação entre 25% a 50%	Apenas Variação
commitsPull_D=some commits	life_time=lengthy	0,04	0,18	1,36	0,01	1,06	1,28	-0,08	Não	Variação abaixo de 25%	Apenas Variação
commitsPull_D=many commits	life_time=medium	0,01	0,16	2,41	0,01	1,11	1,54	-0,87	Não	Variação de 75% a 100%	Apenas Variação
commitsPull_D=some commits	life_time=medium	0,02	0,1	1,59	0,01	1,04	1,29	-0,3	Não	Variação entre 25% a 50%	Apenas Variação
commitsPull_D=1 commit	life_time=short	0,07	0,1	0,76	-0,02	0,97	0,91	0,15	Não	Variação abaixo de 25%	Apenas Variação
commitsPull_D=1 commit	life_time=lengthy	0,07	0,09	0,69	-0,03	0,95	0,8	0,11	Não	Variação abaixo de 25%	Apenas Variação
commitsPull_D=1 commit	life_time=medium	0,03	0,04	0,68	-0,02	0,98	0,84	0,16	Não	Variação abaixo de 25%	Apenas Variação
changedFiles_D=1 file	life_time=very short	0,47	0,83	1,25	0,09	1,98	1,21	-0,04	Não	Variação abaixo de 25%	Apenas Variação
changedFiles_D=some files	life_time=very short	0,15	0,51	0,77	-0,04	0,68	0,87	0,1	Não	Variação abaixo de 25%	Apenas Variação
changedFiles_D=many files	life_time=lengthy	0,05	0,34	2,49	0,03	1,3	1,54	-0,95	Não	Variação de 75% a 100%	Apenas Variação
changedFiles_D=many files	life_time=very short	0,05	0,34	0,5	-0,05	0,5	0,68	0,18	Não	Variação abaixo de 25%	Apenas Variação
changedFiles_D=some files	life_time=lengthy	0,06	0,2	1,46	0,02	1,08	1,19	-0,27	Não	Variação entre 25% a 50%	Apenas Variação
changedFiles_D=many files	life_time=short	0,03	0,2	1,5	0,01	1,08	1,17	-0,33	Não	Variação entre 25% a 50%	Apenas Variação
changedFiles_D=some files	life_time=short	0,05	0,19	1,45	0,02	1,07	1,1	-0,35	Não	Variação entre 25% a 50%	Apenas Variação
changedFiles_D=many files	life_time=medium	0,02	0,13	2,01	0,01	1,07	1,36	-0,65	Não	Variação entre 50% a 75%	Apenas Variação



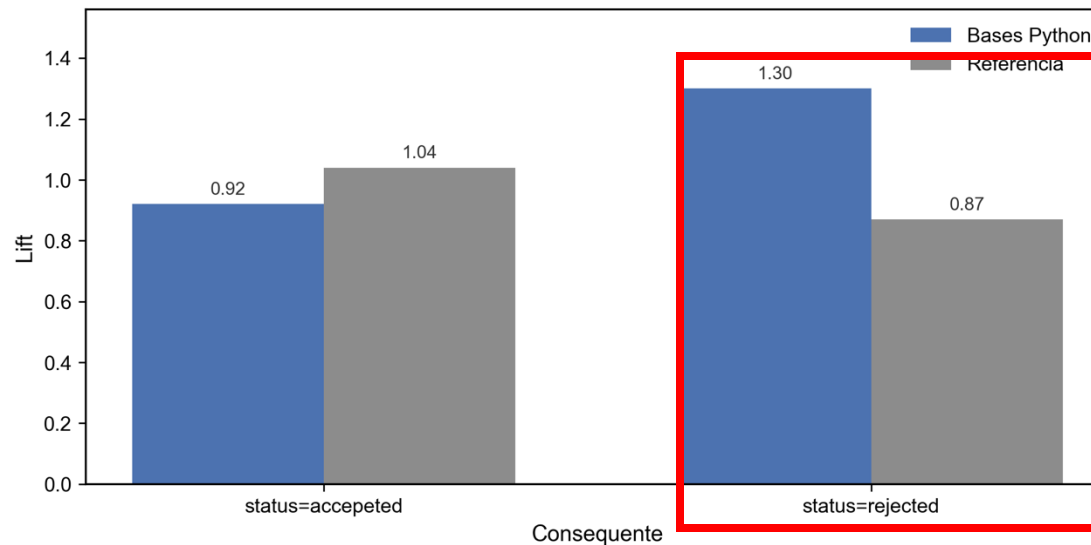
LIFE TIME -> STATUS

08

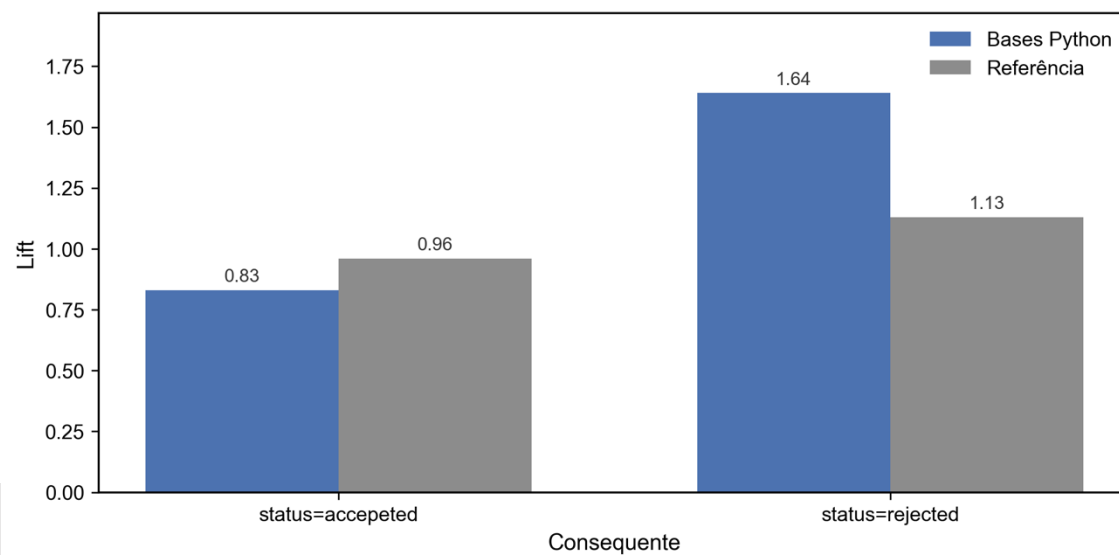
life time — very short



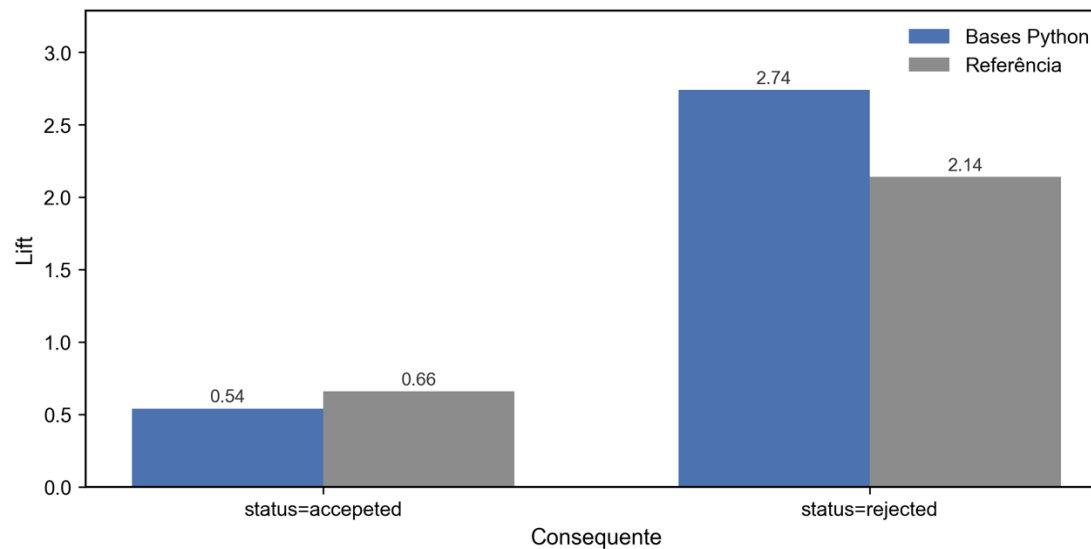
life time — short



life time — medium

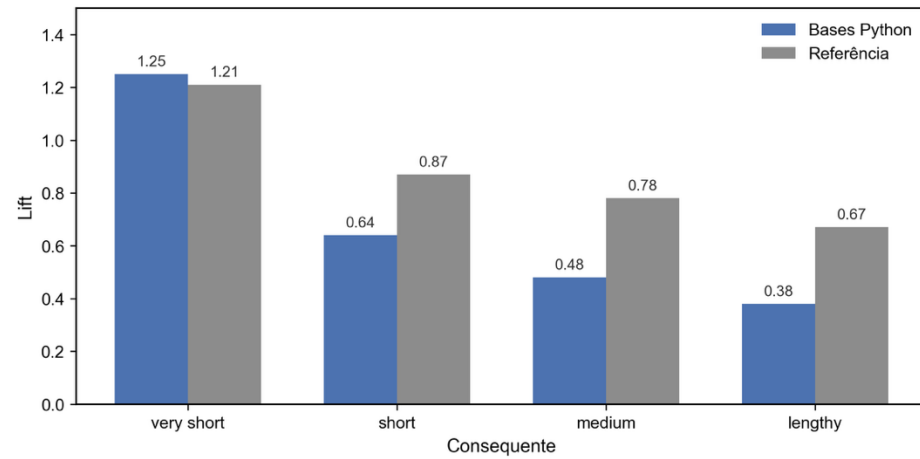


life time — lengthy

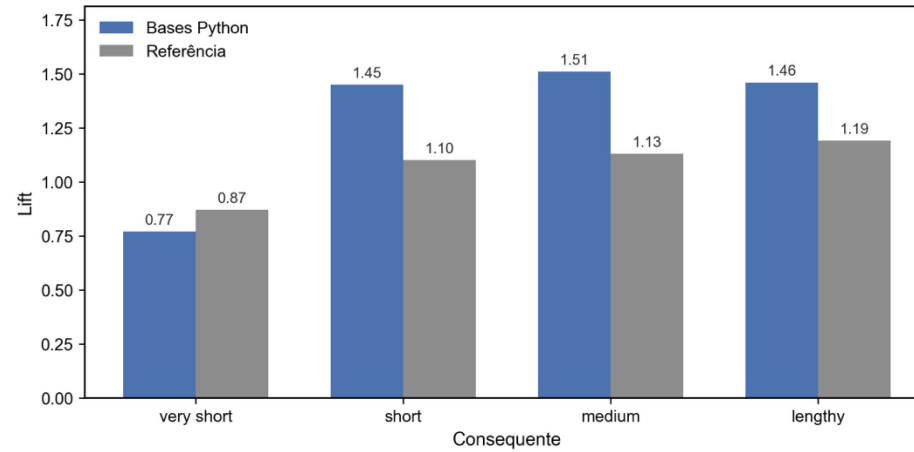




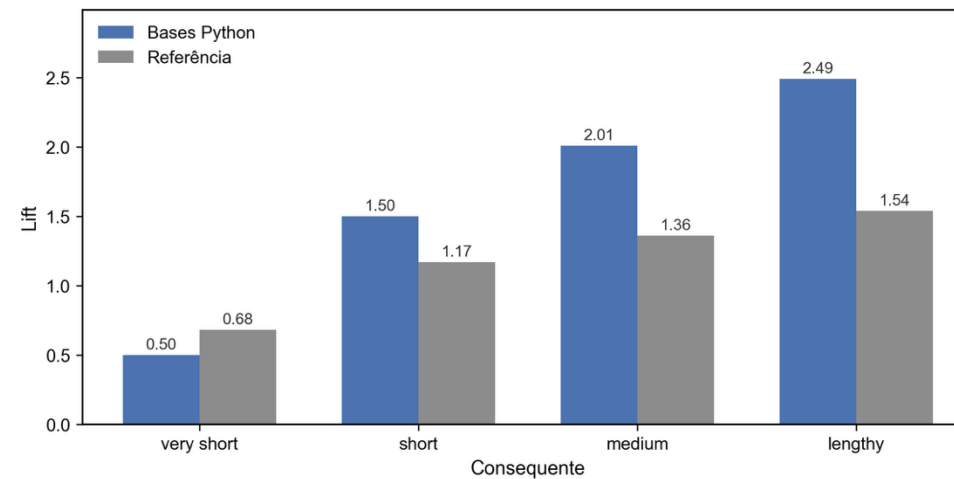
changedFiles D — 1 file



changedFiles D — some files

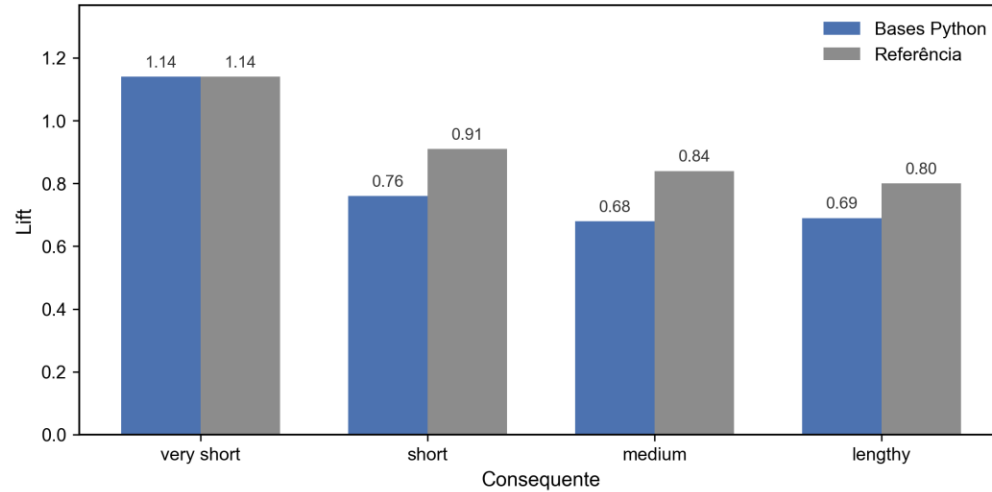


changedFiles D — many files

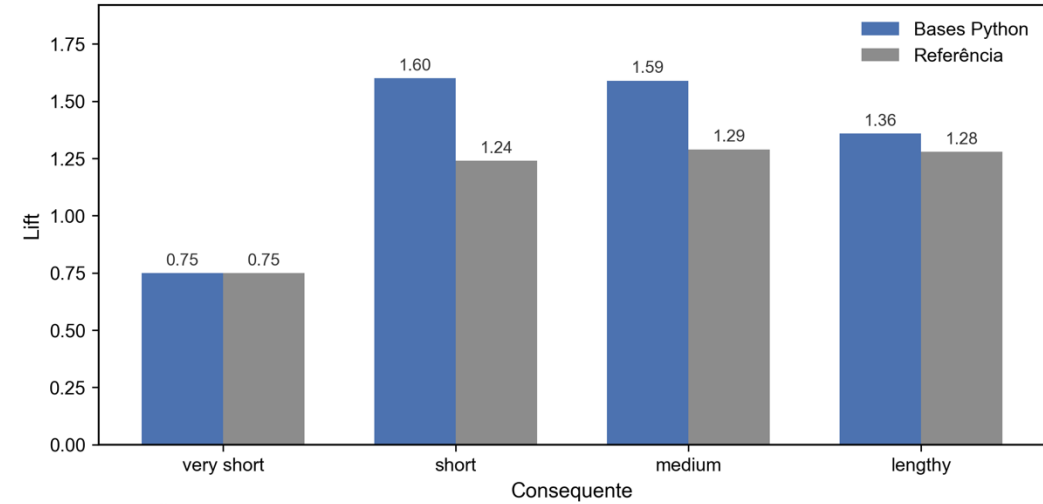




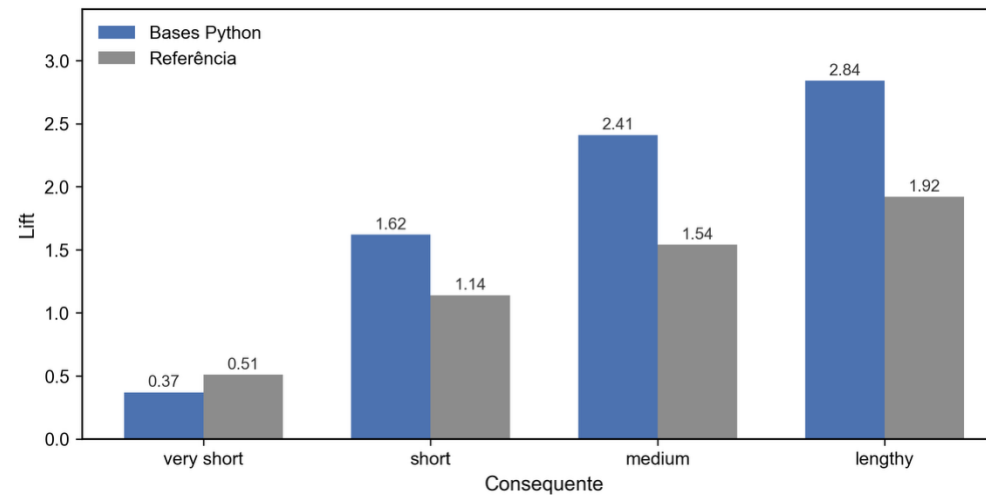
commitsPull D — 1 commit



commitsPull D — some commits



commitsPull D — many commits

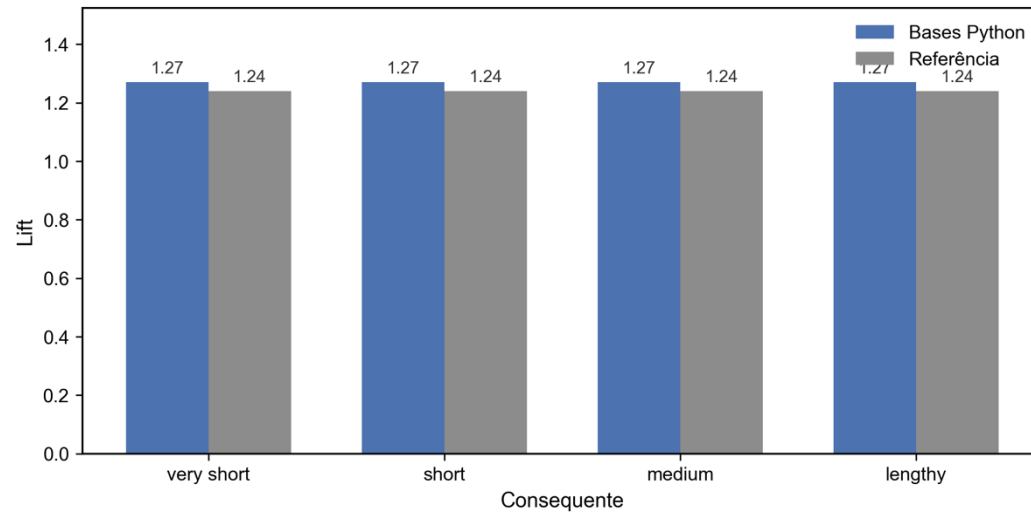




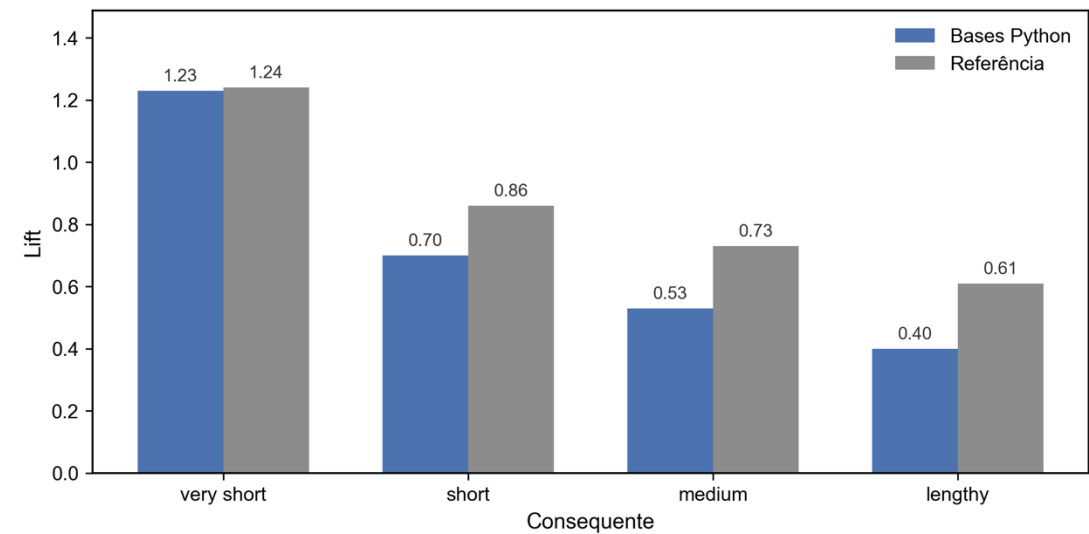
TOTAL LINES -> LIFETIME

011

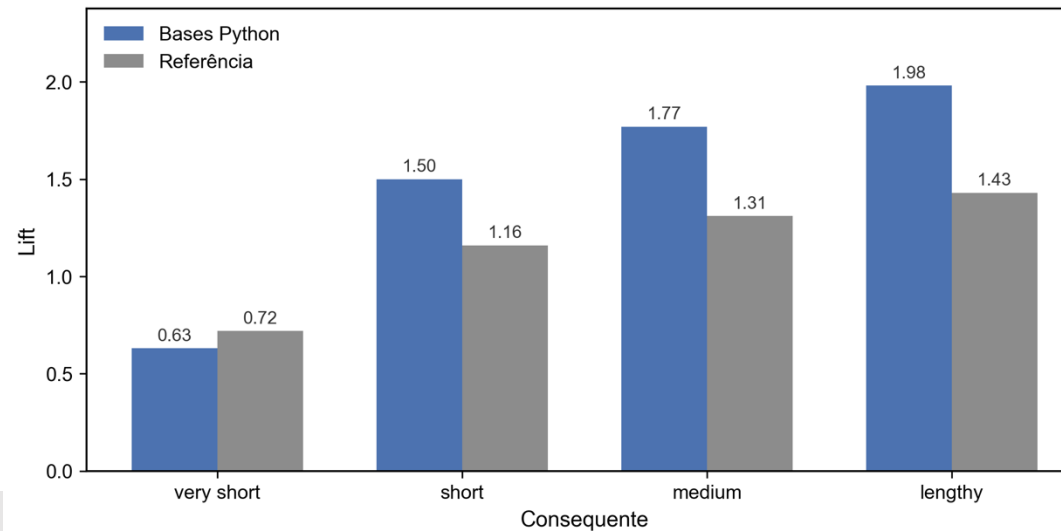
total lines D — 1 line



total lines D — some lines



total lines D — many lines

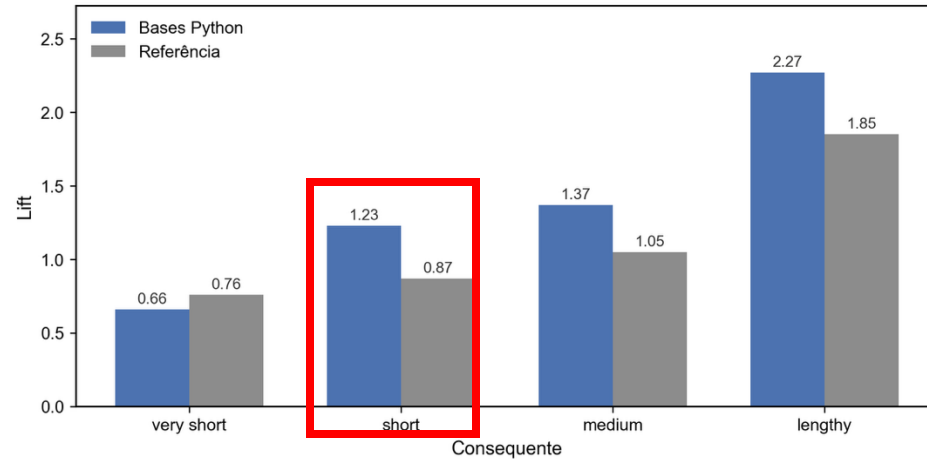




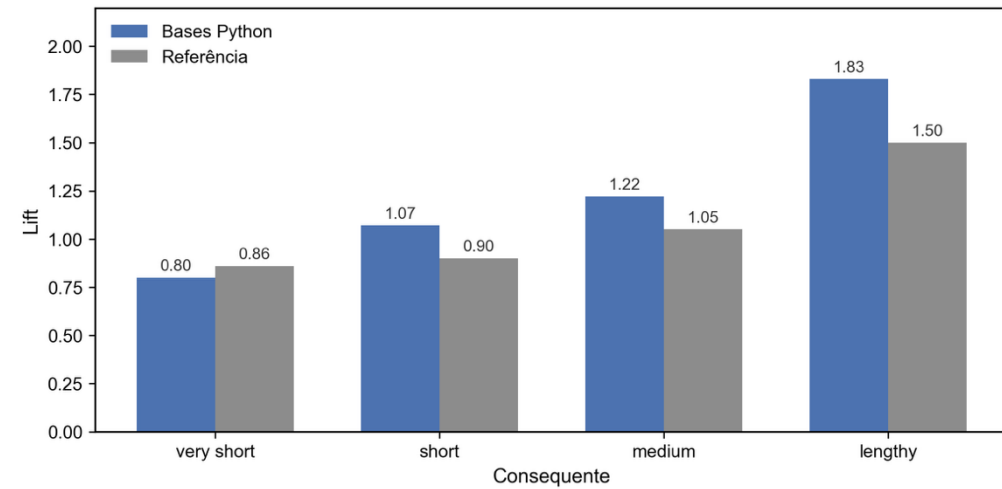
REQUESTER EXPERIENCE -> LIFETIME

012

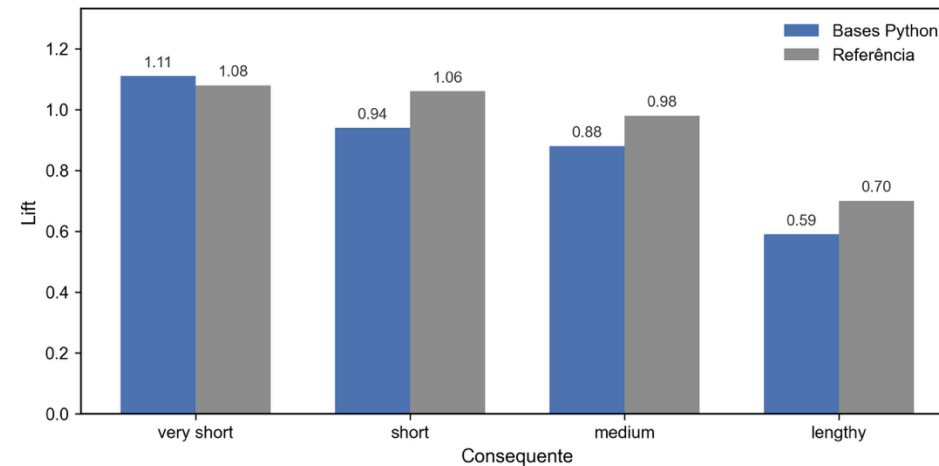
requester experience project — no contribution



requester experience project — some contributions

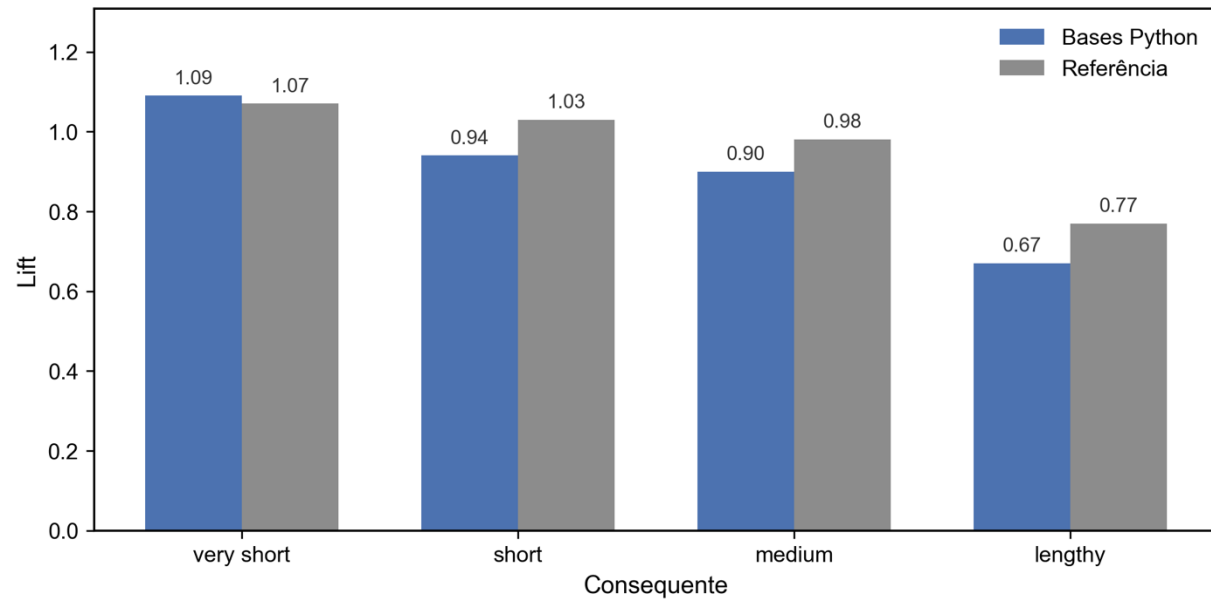


requester experience project — many contributions

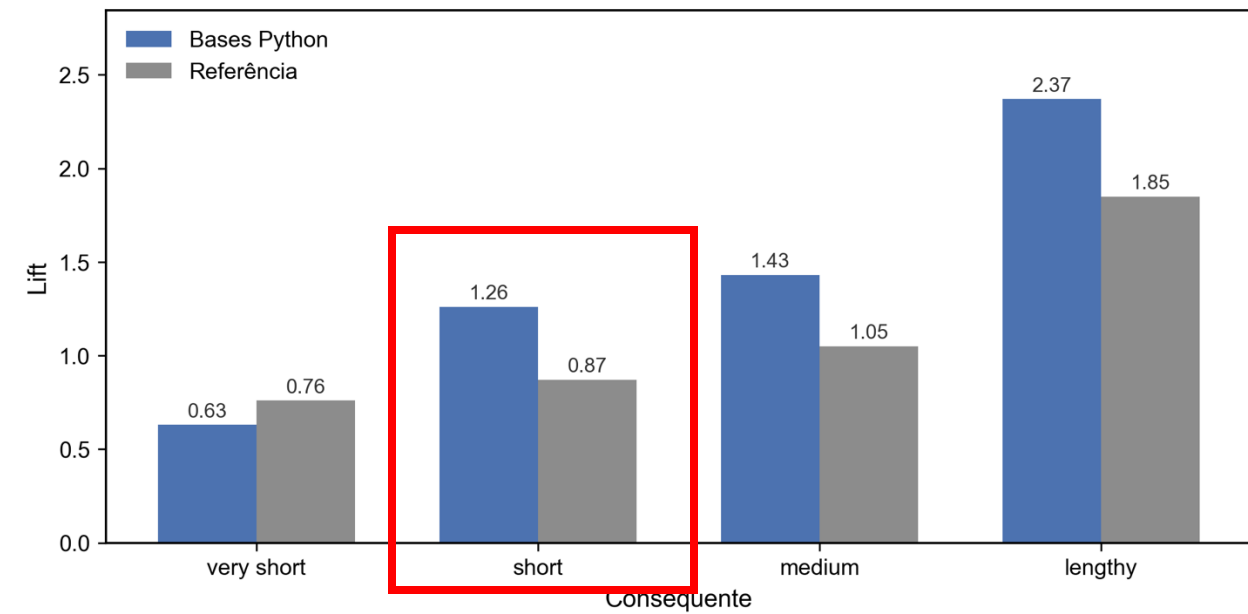




first Pull — False



first Pull — True

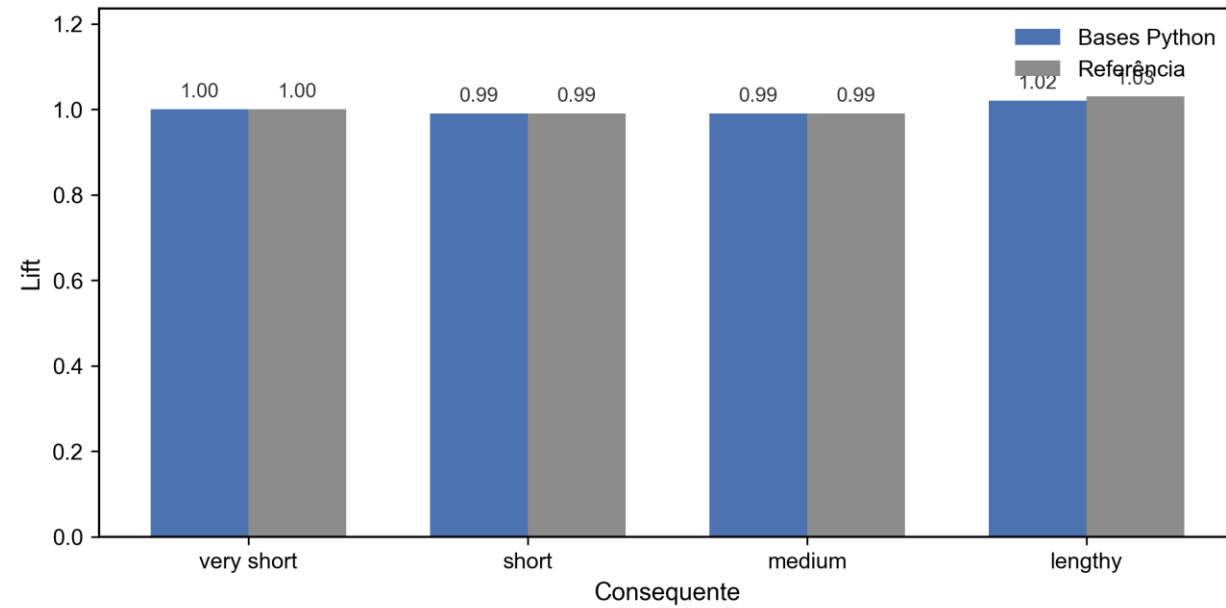




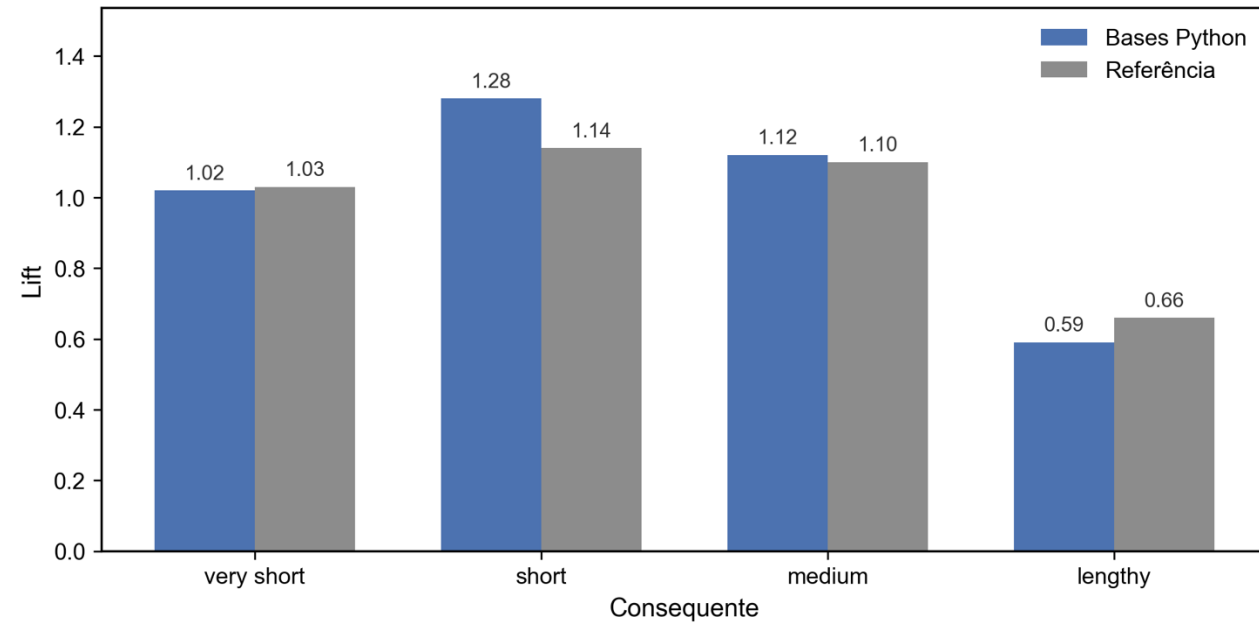
CORE TEAM FOLLOWS REQUESTER -> LIFETIME

014

coreTeamFollowsRequester — False

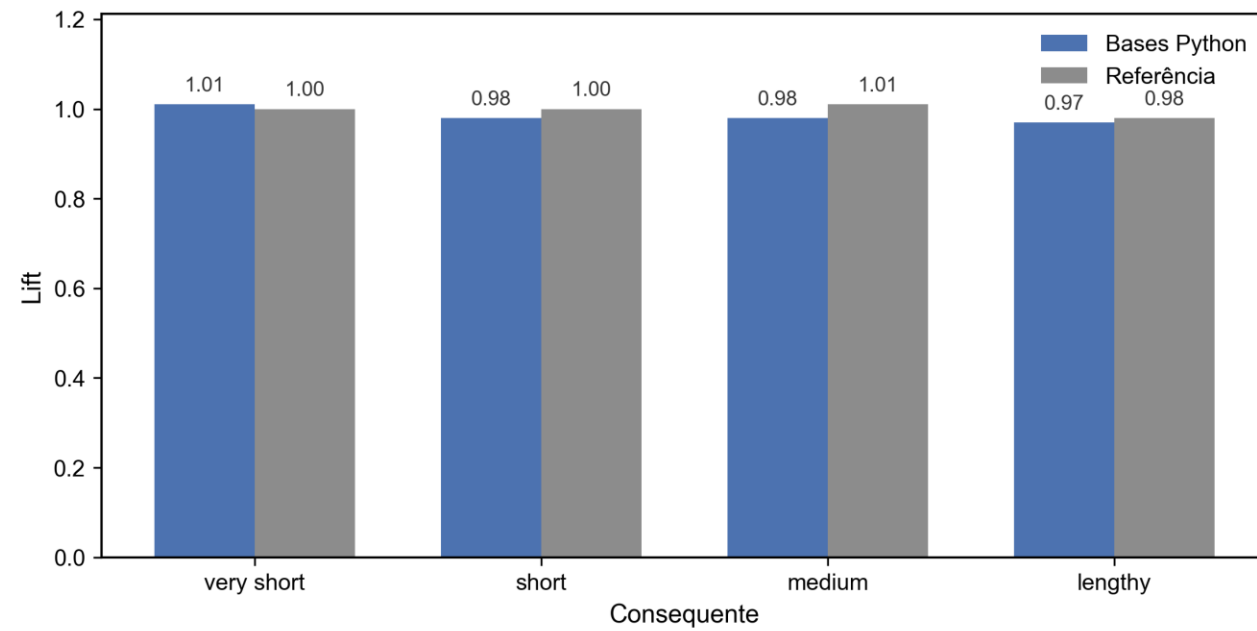


coreTeamFollowsRequester — True

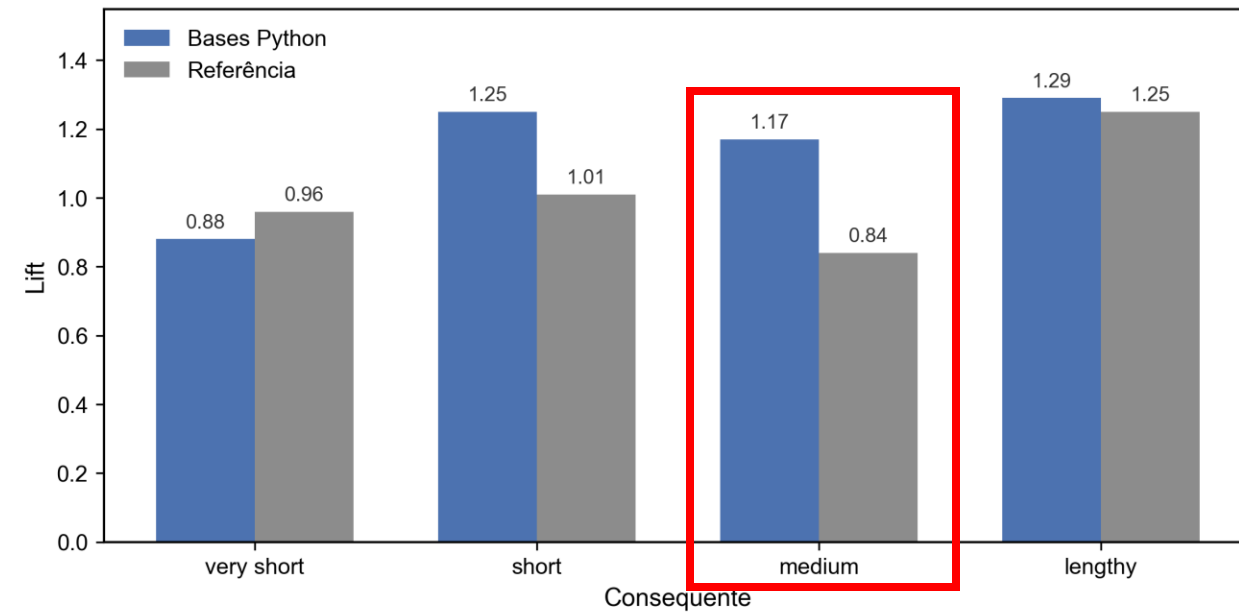




followers boolean — has followers

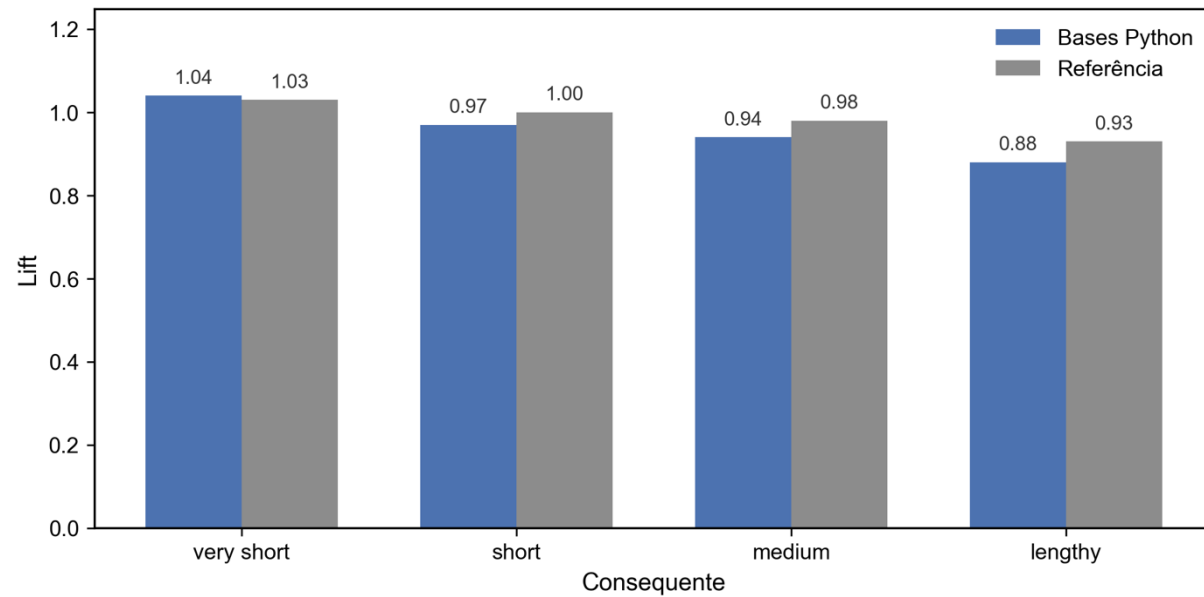


followers boolean — no followers

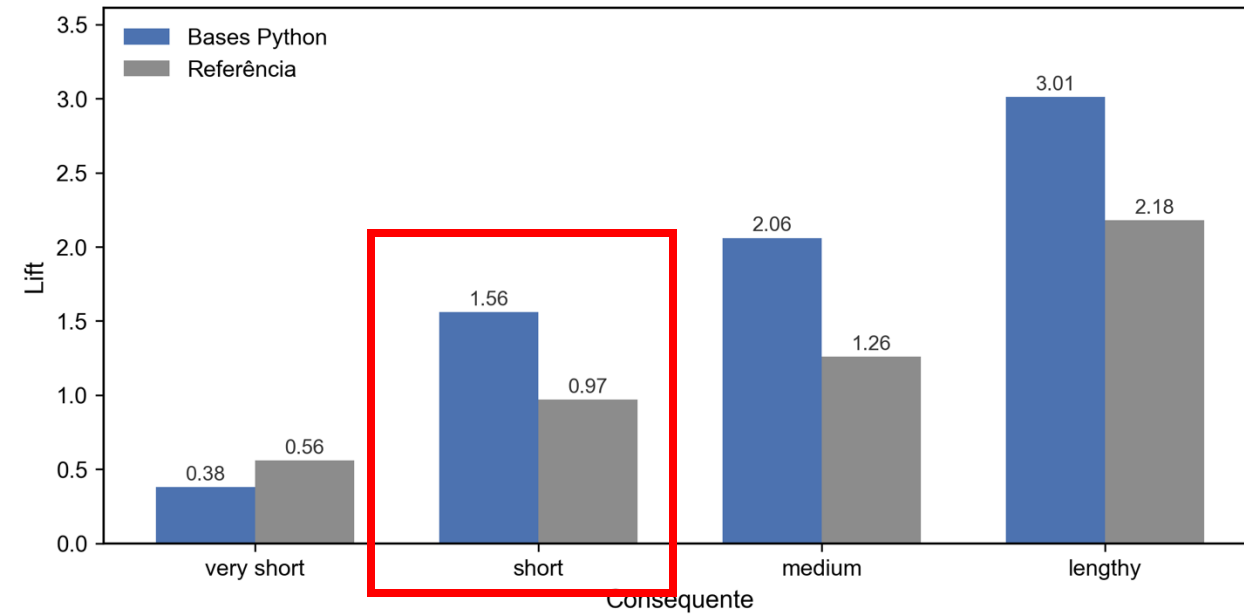




typeDeveloper — external



typeDeveloper — core





Variações	Quantidade de Regras
Variação abaixo de 25%	55
Variação entre 25% a 50%	19
Variação entre 50% a 75%	6
Variação de 75% a 100%	5
Total Geral	85

Variações	Quantidade de Regras
◦ Apenas Variação	80
Variação abaixo de 25%	55
Variação de 75% a 100%	5
Variação entre 25% a 50%	15
Variação entre 50% a 75%	5
◦ Mudança na Regra	5
Variação entre 25% a 50%	4
Variação entre 50% a 75%	1
Total Geral	85



REVIWER



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Commcare	changedFiles_D=1 file	closedBy=yedi	1,79%	3,72%	66,14%	1,38
Commcare	changedFiles_D=1 file	closedBy=dmyung	1,26%	2,62%	55,00%	1,14
Django	changedFiles_D=1 file	closedBy=ubernostrum	0,16%	0,39%	85,71%	2,08
Django	changedFiles_D=1 file	closedBy=kmtracey	0,13%	0,33%	71,43%	1,73
Ipython	changedFiles_D=1 file	closedBy=epatters	0,25%	0,51%	88,89%	1,84
Ipython	changedFiles_D=1 file	closedBy=bfroehle	1,30%	2,70%	66,67%	1,38
Pandas	changedFiles_D=1 file	closedBy=takluyver	0,09%	0,32%	100,00%	3,54
Pandas	changedFiles_D=1 file	closedBy=adamklein	0,40%	1,43%	69,23%	2,45
Rosdistro	changedFiles_D=1 file	closedBy=130s	0,12%	0,13%	100,00%	1,08
Rosdistro	changedFiles_D=1 file	closedBy=dirk-thomas	27,67%	29,88%	96,65%	1,04



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Django	changedFiles_D=some files	closedBy=jezdez	0,54%	1,44%	74,07%	1,98
Django	changedFiles_D=some files	closedBy=andrewgodwin	0,80%	2,15%	65,22%	1,75
Ipython	changedFiles_D=some files	closedBy=rgbkrk	0,28%	0,82%	39,13%	1,15
Ipython	changedFiles_D=some files	closedBy=ellisonbg	3,19%	9,34%	38,01%	1,11
Pandas	changedFiles_D=some files	closedBy=TomAugspurger	1,48%	2,88%	75,00%	1,46
Pandas	changedFiles_D=some files	closedBy=hayd	1,08%	2,09%	70,59%	1,37
Rosdistro	changedFiles_D=some files	closedBy=tfoote	2,13%	32,11%	10,02%	1,51
Rosdistro	changedFiles_D=some files	closedBy=isucan	0,14%	2,11%	8,89%	1,34



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Commcare	changedFiles_D=many files	closedBy=nickpell	1,27%	6,84%	30,38%	1,63
Commcare	changedFiles_D=many files	closedBy=snopoke	1,50%	8,07%	21,97%	1,18
Django	changedFiles_D=many files	closedBy=freakboy3742	0,19%	0,87%	33,33%	1,55
Django	changedFiles_D=many files	closedBy=loic	0,21%	1,00%	30,77%	1,43
Ipython	changedFiles_D=many files	closedBy=ellisonbg	2,23%	12,61%	26,57%	1,5
Ipython	changedFiles_D=many files	closedBy=fperez	3,25%	18,39%	25,93%	1,47
Pandas	changedFiles_D=many files	closedBy=jreback	14,02%	68,79%	24,43%	1,2
Pandas	changedFiles_D=many files	closedBy=wesm	4,30%	21,10%	20,08%	0,99
Rosdistro	changedFiles_D=many files	closedBy=tfoote	0,38%	48,89%	1,81%	2,3
Rosdistro	changedFiles_D=many files	closedBy=wjwwood	0,14%	17,78%	0,69%	0,87



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Commcare	commitsPull_D=1 commit	closedBy=yedi	2,11%	3,40%	77,78%	1,26
Commcare	commitsPull_D=1 commit	closedBy=dmyung	1,63%	2,64%	71,25%	1,15
Django	commitsPull_D=1 commit	closedBy=erikr	0,24%	0,32%	100,00%	1,31
Django	commitsPull_D=1 commit	closedBy=evildmp	0,19%	0,25%	100,00%	1,31
lpython	commitsPull_D=1 commit	closedBy=epatters	0,25%	0,46%	88,89%	1,66
lpython	commitsPull_D=1 commit	closedBy=jdmarch	0,15%	0,29%	71,43%	1,33
Pandas	commitsPull_D=1 commit	closedBy=takluyver	0,09%	0,12%	100,00%	1,39
Pandas	commitsPull_D=1 commit	closedBy=jorisvandenbossche	5,56%	7,71%	91,18%	1,27
Rosdistro	commitsPull_D=1 commit	closedBy=130s	0,12%	0,13%	100,00%	1,1
Rosdistro	commitsPull_D=1 commit	closedBy=dirk-thomas	27,18%	29,85%	94,95%	1,04



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Django	commitsPull_D=some commits	closedBy=bmispelon	0,70%	4,04%	23,85%	1,38
Django	commitsPull_D=some commits	closedBy=freakboy3742	0,13%	0,78%	23,81%	1,38
Ipython	commitsPull_D=some commits	closedBy=rgbkrk	0,34%	1,09%	47,83%	1,53
Ipython	commitsPull_D=some commits	closedBy=fperez	4,49%	14,40%	35,80%	1,15
Pandas	commitsPull_D=some commits	closedBy=hayd	0,54%	2,55%	35,29%	1,67
Pandas	commitsPull_D=some commits	closedBy=wesm	7,17%	33,97%	33,47%	1,59
Rosdistro	commitsPull_D=some commits	closedBy=kwc	0,14%	1,72%	34,78%	4,28
Rosdistro	commitsPull_D=some commits	closedBy=isucan	0,21%	2,58%	13,33%	1,64



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Django	commitsPull_D=some commits	closedBy=bmispelon	0,70%	4,04%	23,85%	1,38
Django	commitsPull_D=some commits	closedBy=freakboy3742	0,13%	0,78%	23,81%	1,38
Ipython	commitsPull_D=some commits	closedBy=rgbkrk	0,34%	1,09%	47,83%	1,53
Ipython	commitsPull_D=some commits	closedBy=fperez	4,49%	14,40%	35,80%	1,15
Pandas	commitsPull_D=some commits	closedBy=hayd	0,54%	2,55%	35,29%	1,67
Pandas	commitsPull_D=some commits	closedBy=wesm	7,17%	33,97%	33,47%	1,59
Rosdistro	commitsPull_D=some commits	closedBy=kwc	0,14%	1,72%	34,78%	4,28
Rosdistro	commitsPull_D=some commits	closedBy=isucan	0,21%	2,58%	13,33%	1,64



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Django	total_lines_D=1 line	closedBy=adrianholovaty	0,11%	5,48%	10,00%	5,11
Django	total_lines_D=1 line	closedBy=charettes	0,16%	8,22%	6,67%	3,41
Ipython	total_lines_D=1 line	closedBy=Carreau	0,46%	18,29%	3,03%	1,19
Ipython	total_lines_D=1 line	closedBy=minrk	0,90%	35,37%	2,91%	1,14
Pandas	total_lines_D=1 line	closedBy=cpccloud	0,09%	7,14%	5,41%	4,31
Pandas	total_lines_D=1 line	closedBy=y-p	0,18%	14,29%	3,77%	3,01
Rosdistro	total_lines_D=1 line	closedBy=wjwwood	0,58%	30,28%	2,83%	1,49
Rosdistro	total_lines_D=1 line	closedBy=tfoote	0,58%	30,28%	2,71%	1,43



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Django	total_lines_D=some lines	closedBy=erikr	0,19%	0,40%	77,78%	1,66
Django	total_lines_D=some lines	closedBy=kmtracey	0,13%	0,29%	71,43%	1,52
Ipython	total_lines_D=some lines	closedBy=epatters	0,22%	0,44%	77,78%	1,56
Ipython	total_lines_D=some lines	closedBy=jdmarch	0,15%	0,31%	71,43%	1,43
Pandas	total_lines_D=some lines	closedBy=takluyver	0,09%	0,25%	100,00%	2,8
Pandas	total_lines_D=some lines	closedBy=changhiskhan	0,72%	2,01%	69,57%	1,95
Rosdistro	total_lines_D=some lines	closedBy=130s	0,12%	0,13%	100,00%	1,07
Rosdistro	total_lines_D=some lines	closedBy=dirk-thomas	27,27%	29,25%	95,25%	1,02



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Django	total_lines_D=many lines	closedBy=jphalip	0,51%	1,00%	82,61%	1,61
Django	total_lines_D=many lines	closedBy=jezdez	0,54%	1,05%	74,07%	1,45
Ipython	total_lines_D=many lines	closedBy=ellisonbg	4,98%	10,45%	59,41%	1,25
Ipython	total_lines_D=many lines	closedBy=fperez	7,43%	15,58%	59,26%	1,24
Pandas	total_lines_D=many lines	closedBy=TomAugspurger	1,43%	2,27%	72,73%	1,15
Pandas	total_lines_D=many lines	closedBy=jtratner	1,16%	1,85%	72,22%	1,14
Rosdistro	total_lines_D=many lines	closedBy=kwc	0,10%	2,14%	26,09%	5,35
Rosdistro	total_lines_D=many lines	closedBy=tfoote	1,46%	30,00%	6,90%	1,41



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Commcare	typeDeveloper=core	closedBy=czue	0,13%	45,00%	0,36%	1,25
Django	typeDeveloper=core	closedBy=MarkusH	0,48%	2,35%	38,30%	1,86
Django	typeDeveloper=core	closedBy=ramiro	0,24%	1,17%	30,00%	1,46
Ipython	typeDeveloper=core	closedBy=rgbkrk	0,09%	1,70%	13,04%	2,39
Ipython	typeDeveloper=core	closedBy=bfroehle	0,15%	2,84%	7,94%	1,46
Pandas	typeDeveloper=core	closedBy=shoyer	0,31%	2,97%	17,95%	1,7
Pandas	typeDeveloper=core	closedBy=jorisvandenbossche	0,90%	8,47%	14,71%	1,39
Rosdistro	typeDeveloper=core	closedBy=tfoote	0,24%	29,17%	1,15%	1,37
Rosdistro	typeDeveloper=core	closedBy=wjwwood	0,16%	18,75%	0,77%	0,92



Base	Antecedente	Consequente	Suporte	Confianca x -> y	Confianca y -> x	Lift
Commcare	typeDeveloper=external	closedBy=czue	35,80%	35,90%	99,64%	1
Commcare	typeDeveloper=external	closedBy=dannyroberts	14,37%	14,41%	99,70%	1
Django	typeDeveloper=external	closedBy=jacobian	0,24%	0,30%	100,00%	1,26
Django	typeDeveloper=external	closedBy=kmtracey	0,19%	0,24%	100,00%	1,26
lpython	typeDeveloper=external	closedBy=epatters	0,28%	0,29%	100,00%	1,06
lpython	typeDeveloper=external	closedBy=jdmarch	0,22%	0,23%	100,00%	1,06
Pandas	typeDeveloper=external	closedBy=takluyver	0,09%	0,10%	100,00%	1,12
Pandas	typeDeveloper=external	closedBy=TomAugspurger	1,88%	2,10%	95,45%	1,07
Rosdistro	typeDeveloper=external	closedBy=130s	0,12%	0,12%	100,00%	1,01
Rosdistro	typeDeveloper=external	closedBy=dirk-thomas	28,45%	28,69%	99,39%	1



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Commcare	coreTeamFollowsRequester=true	closedBy=kennknowles	0,23%	2,44%	41,03%	4,37
Commcare	coreTeamFollowsRequester=true	closedBy=czue	8,74%	93,13%	24,32%	2,59
Django	coreTeamFollowsRequester=true	closedBy=jezdez	0,16%	46,15%	22,22%	63,76
Ipython	coreTeamFollowsRequester=true	closedBy=ivanov	2,07%	32,52%	82,72%	12,97
Ipython	coreTeamFollowsRequester=true	closedBy=rgbkrk	0,50%	7,77%	69,57%	10,91
Pandas	coreTeamFollowsRequester=true	closedBy=wesm	6,27%	93,96%	29,29%	4,39
Pandas	coreTeamFollowsRequester=true	closedBy=cpcloud	0,18%	2,68%	10,81%	1,62
Rosdistro	coreTeamFollowsRequester=true	closedBy=130s	0,12%	12,73%	100,00%	104,35
Rosdistro	coreTeamFollowsRequester=true	closedBy=wjwwood	0,77%	80,00%	3,77%	3,93



Base	Antecedente	Consecuente	Suporte	Confianza x -> y	Confianza y -> x	Lift
Commcare	coreTeamFollowsRequester=false	closedBy=dannyroberts	14,41%	15,90%	100,00%	1,1
Commcare	coreTeamFollowsRequester=false	closedBy=snopoke	6,85%	7,56%	100,00%	1,1
Django	coreTeamFollowsRequester=false	closedBy=timgraham	62,90%	63,12%	100,00%	1
Django	coreTeamFollowsRequester=false	closedBy=aaugustin	5,23%	5,25%	100,00%	1
Ipython	coreTeamFollowsRequester=false	closedBy=takluyver	23,46%	25,06%	99,87%	1,07
Ipython	coreTeamFollowsRequester=false	closedBy=fperez	12,53%	13,39%	100,00%	1,07
Pandas	coreTeamFollowsRequester=false	closedBy=jreback	57,39%	61,50%	100,00%	1,07
Pandas	coreTeamFollowsRequester=false	closedBy=y-p	4,75%	5,09%	100,00%	1,07
Rosdistro	coreTeamFollowsRequester=false	closedBy=dirk-thomas	28,63%	28,91%	100,00%	1,01
Rosdistro	coreTeamFollowsRequester=false	closedBy=vrabaud	27,55%	27,81%	100,00%	1,01



Conclusão

Para Lifetime, no geral, ela segue o mesmo comportamento da Tese de Soares, porém em 5 das 85 regras analisadas tiveram mudança de tendencia, logo se expandir a pesquisa para mais atributos, novos comportamentos diferentes de Soares podem ser encontrados.

Com relação ao revisor, é possível analisar que o comportamento em bases de Python reforça o que Soares em sua tese investigou, que em certos domínios, mesmo sendo códigos open-source onde todos deveriam atuar de forma igualitária, existem sim comportamentos em que determinados revisores “Dominam” aquele ambiente, fazendo com que na grande maioria dos casos ele seja o revisor.



Universidade Federal do Acre
Mestrado em Ciência da Computação

Entendendo o Comportamento de Pull Requests em Python: Um estudo de cenário de Tempo de Vida e Revisores

Apresentação: Álvaro Lopes Rios