

<epam>

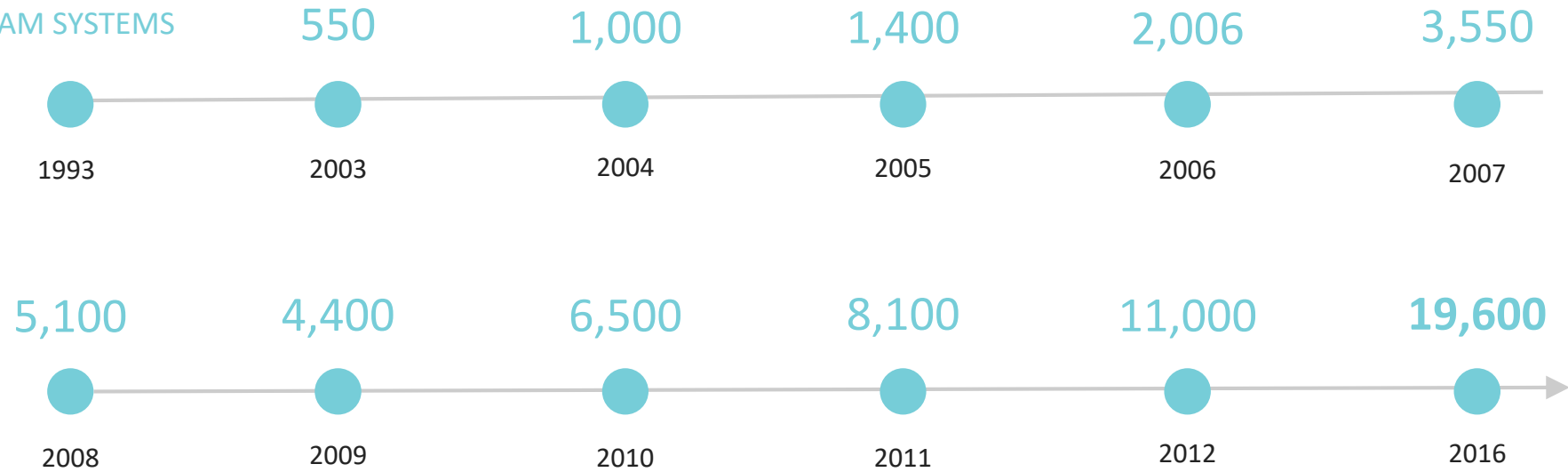
Data Science – из мифа в реальность рекрутёра или как чёрная кошка помогла нам найти уникальных кандидатов

Victoria Semenova
Talent Acquisition manager



EPAM

THE START OF
EPAM SYSTEMS



120

Fortune Global 2000 Customers

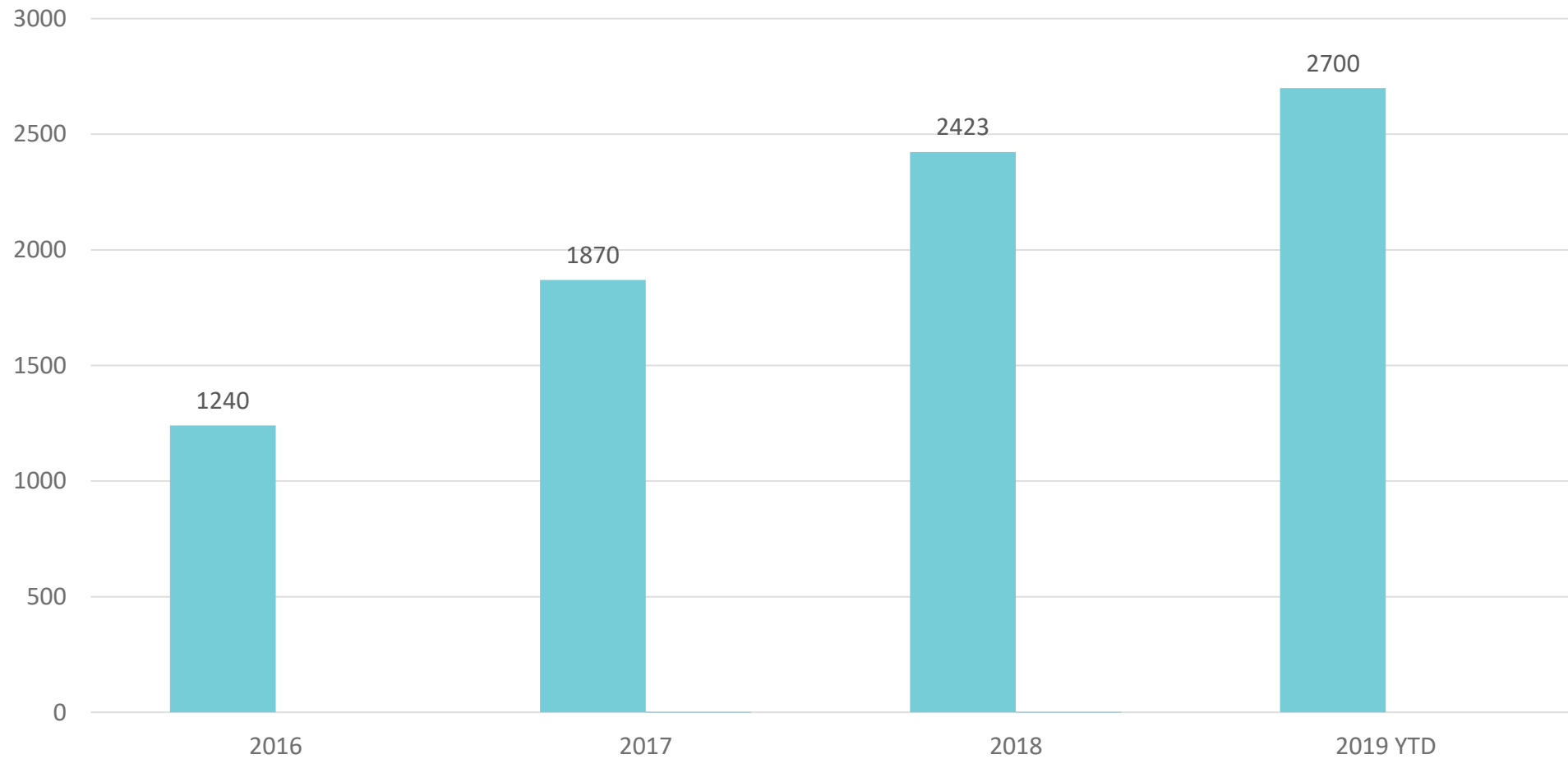
30,100+

Total EPAMers

25+

Contries

EPAM Saint-Petersburg



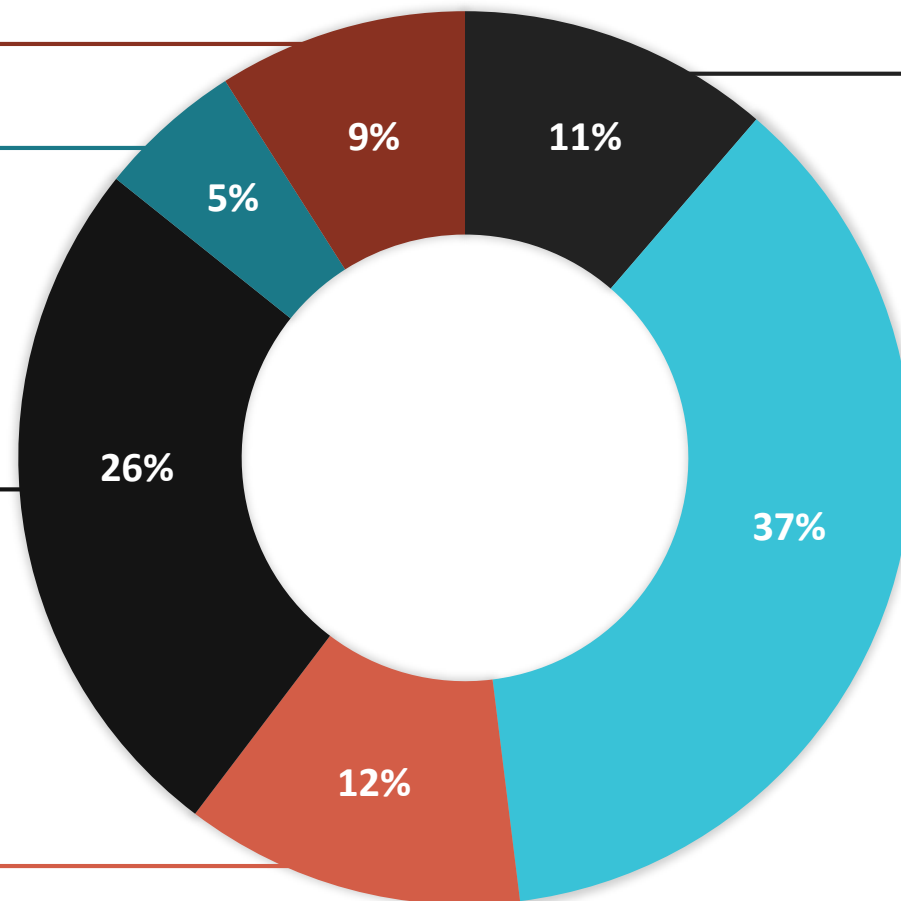
Other

External events

Internal Recruiting DB

Referral program

LI



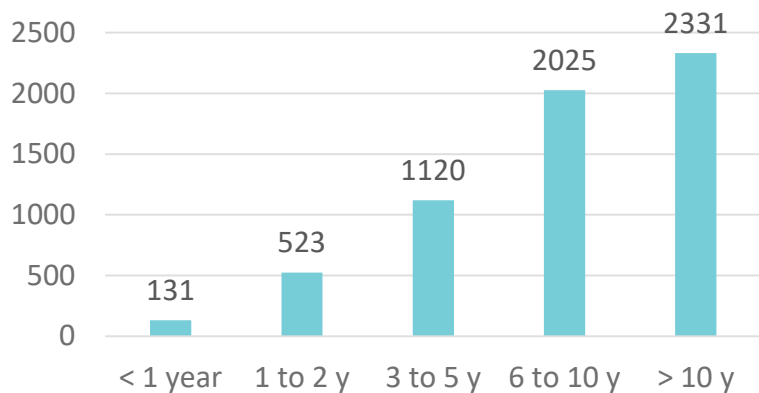
НН

Где мы находим наших кандидатов

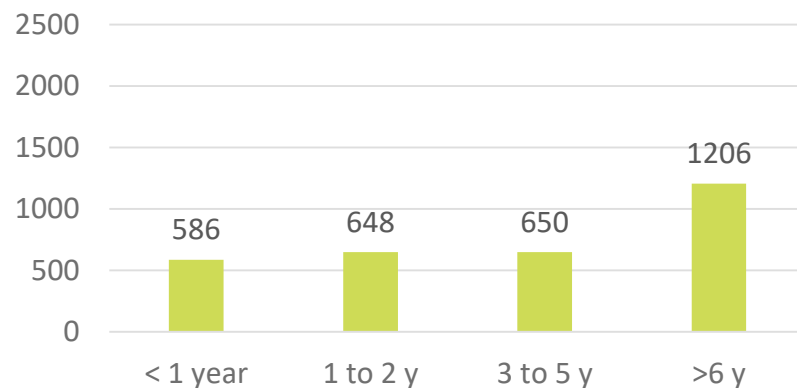
Среднее количество наймов 400-600 в год

Объём рынка в Санкт-Петербурге

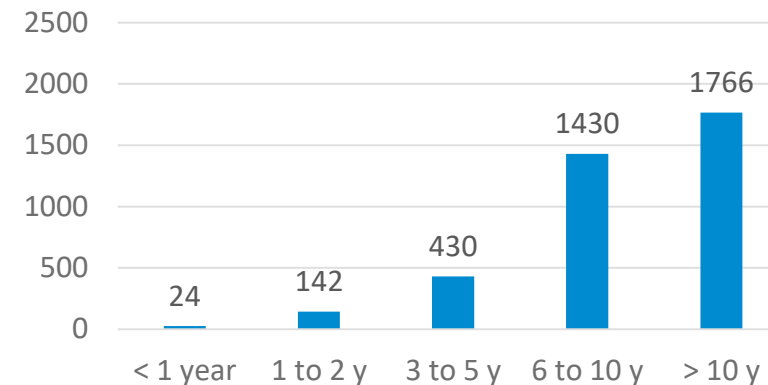
Java: 6130



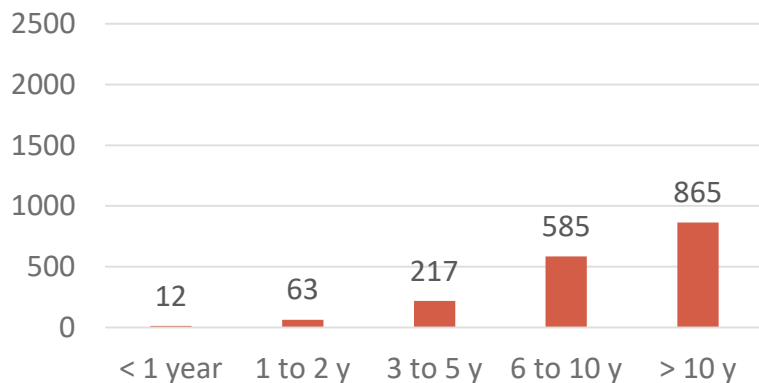
JavaScript: 3090



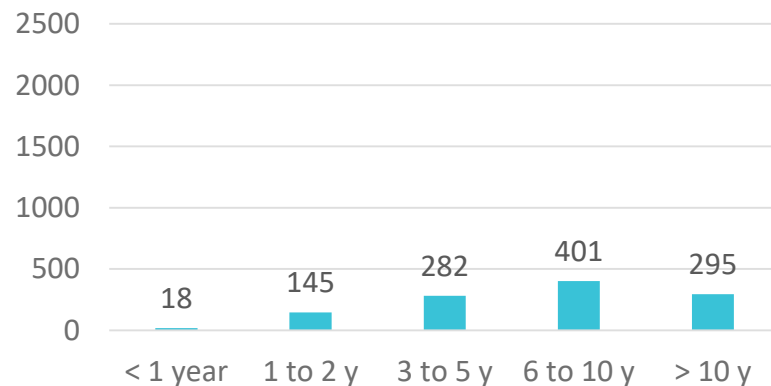
DevOps: 3790



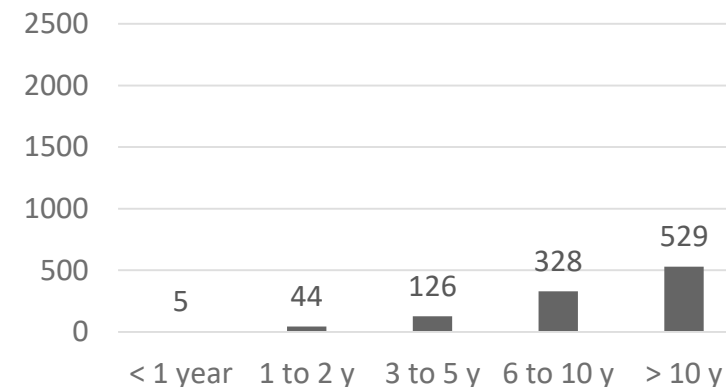
.NET: 1750



Python: 1140



Test Auto: 1030



Data Science Driven Sourcing – первые шаги



Data Science Driven Sourcing – первые шаги

REGION – NOVOSIBIRSK	NUMBER OF DONOR COMPANIES	NUMBER OF CANDIDATES
NOVOSIBIRSK	17	30
ROSTOV & TAGANROG	20	35
KAZAN	20	17
VORONEZH	8	28
CHELYABINSK	6	16
EKATERINBURG	7	11
OMSK	5	30

Total **167** profiles in LI

Data Science Driven Sourcing – первые шаги

Профессиональные
группы в VK

RELEVANCE
30%



Группы
компаний в VK

VK + MAGIC =
NEW CANDIDATES

Друзья Qas из LinkedIn

НАМ УДАЛОСЬ:

207

новых профилей
кандидатов в VK

VS

167

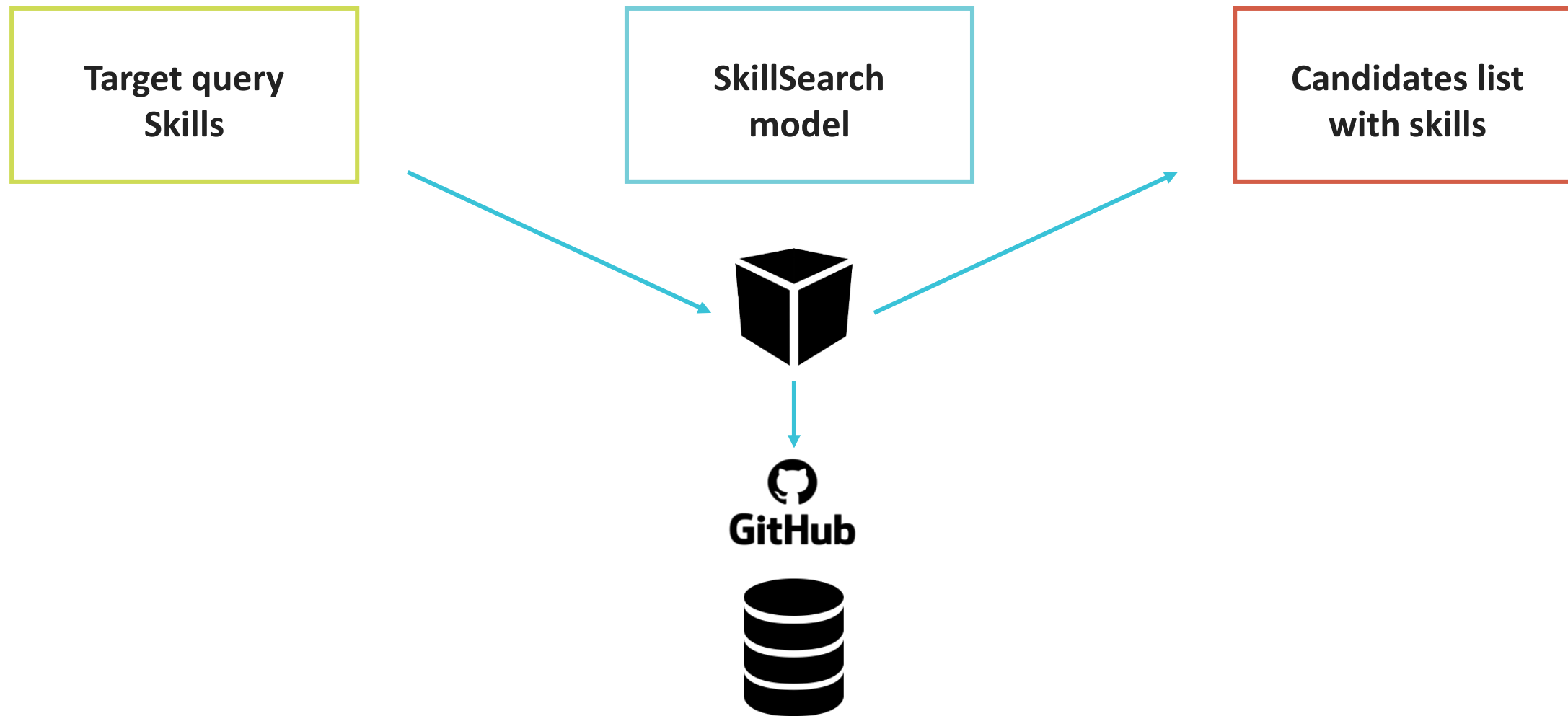
профилей в LI



Почему GitHub:

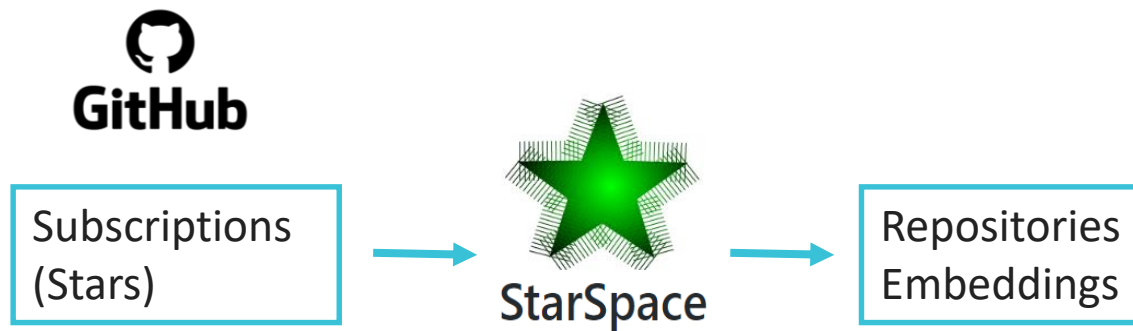
- Проф-ая соц. сеть для разработчиков
- Возможно сортировка по скилам
- Возможно видеть локацию, компанию
- Open source, готовые дампы

Data Science - из мифа в реальность рекрутера



Как работает алгоритм?

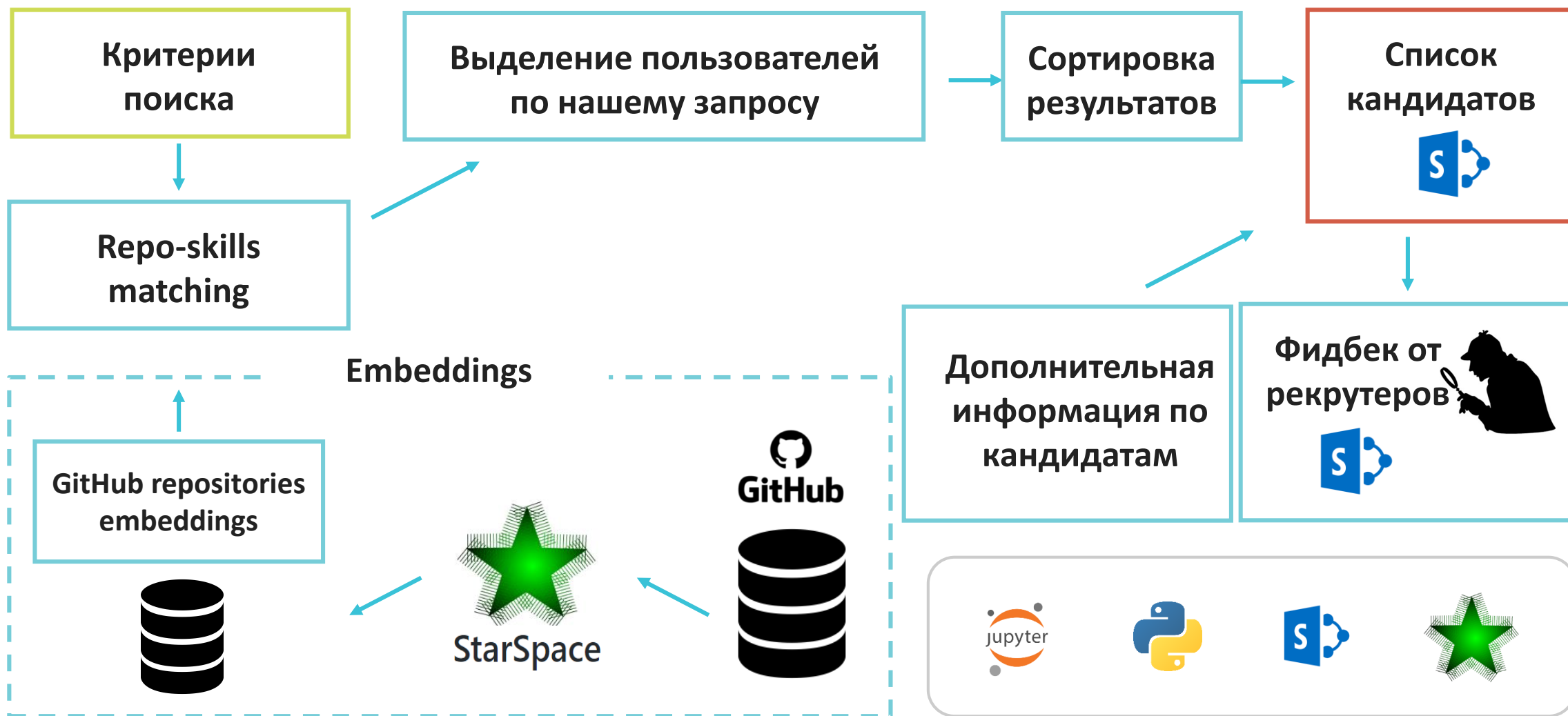
- Пространство «интересов» с координатами для каждого репозитория
- Далее помещаем юзеров и скилы в это же пространство



(example gif: books on Wikipedia embedding)

<https://towardsdatascience.com/neural-network-embeddings-explained-4d028e6f0526>

Data Science - из мифа в реальность рекрутера



Результаты

■ - TA query skills relevance

■ - Skills

Excel Online | EPAM github_search_feedback > Python_Dev_long_exact | weighted_skills_plus_av... - Saved | Ilya

File Home Insert Data Review View Tell me what you want to do Open in Excel Share

P3 fx 0.966580065164662

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	python	flask	aiohttp	sanic	mesos	postgres	redis	rest	websocke	docker	swagger	sql	jenkins	django	total_w_relevance_by_skills	relevance_avg_weighted	total_score	user_id
1																		
2	0.10288	0.46965	0.47037	0.36449	0.03899	0.18383	0.2286	0.03456	0.46466	0.05237	0.11731	0	0.07256	0.40627	1	1	2	309538
3	0.08465	0.52752	0.44111	0.44082	-0.0105	0.18058	0.17802	0.01933	0.39693	0.09131	0.12496	0	0.00161	0.43687	0.987100125	0.966580065	1.95368019	137309
4	0.03763	0.48407	0.48279	0.34229	0.04728	0.09708	0.29093	0.08776	0.32495	0.08904	0.17869	0	0.02397	0.3562	0.960490302	0.929032349	1.889522651	6192
5	-0.015	0.35506	0.48718	0.50174	-0.0621	0.16276	0.24948	0.07901	0.46846	0.04139	0.07426	0	0.00352	0.33895	0.935374832	0.951758587	1.887133418	6898
6	0.02857	0.33	0.4022	0.40824	-0.0055	0.19025	0.15791	0.12547	0.37815	0.1368	0.1029	0	0.12809	0.41273	0.944220766	0.935131048	1.879351813	226786
7	0.00454	0.42294	0.35412	0.20387	0.01577	0.24271	0.23161	0.08143	0.32987	0.25884	0.11542	0	0.19942	0.43429	0.9609292	0.891897538	1.852826738	55026
8	-0.0224	0.42223	0.40265	0.28519	0.0498	0.31269	0.20894	0.06861	0.32911	0.15152	0.12776	0	0.01211	0.31518	0.921991738	0.919720385	1.841712122	68480
9	0.02856	0.40441	0.43255	0.34806	-0.0248	0.15988	0.11563	0.07701	0.42955	0.13331	0.1413	0	0.0747	0.30522	0.909632983	0.927535865	1.837168848	368372
10	0.11062	0.27673	0.52159	0.48066	-0.0343	0.10189	0.22553	0.03534	0.38667	0.05833	0.0356	0	0.0906	0.32153	0.911614119	0.925355692	1.836969811	828353
11	0.07265	0.37162	0.39474	0.3793	0.04271	0.1422	0.18862	0.09555	0.32873	0.11222	0.13014	0	0.0894	0.32468	0.916101155	0.909681138	1.825782293	36171
12	0.059	0.21158	0.38671	0.58798	0.07637	0.0922	0.1609	0.07932	0.36508	0.13127	0.06576	0	0.10024	0.28786	0.900884387	0.916523123	1.81740751	609598
13	0.07584	0.31625	0.4509	0.30096	0.00296	0.18076	0.09183	0.12232	0.4128	0.01242	0.10264	0	0.02681	0.42252	0.887811284	0.921999526	1.80981081	25278
14	0.07441	0.32919	0.38891	0.36024	-0.0822	0.16991	0.14436	0.09706	0.47691	-0.0225	0.10481	0	0.02307	0.45716	0.893571027	0.913494663	1.80706569	702603

Результаты

■ - TA query skills relevance

■ - Skills

Excel Online | EPAM github_search_feedback > Python_Dev_long_exact | weighted_skills_plus_av... - Saved | Ilya

File Home Insert Data Review View Tell me what you want to do Open in Excel Share

P3 fx 0.966580065164662

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	python	flask	aiohttp	sanic	mesos	postgres	redis	rest	websocke	docker	swagger	sql	jenkins	django	total_w_relevance_by_skills	relevance_avg_weighted	total_score	user_id
1															1	1	2	309538
3	0.08465	0.52752	0.44111	0.44082	-0.0105	0.18058	0.17802	0.01933	0.39693	0.09131	0.12496	0	0.00161	0.4368	0.987100125	0.966580065	1.95368019	137309
4	0.03763	0.48407	0.48279	0.34229	0.04728	0.09708	0.29093	0.08776	0.32495	0.08904	0.17869	0	0.02397	0.356	0.960490302	0.929032349	1.889522651	6192
5	-0.015	0.35506	0.48718	0.50174	-0.0621	0.16276	0.24948	0.07901	0.46846	0.04139	0.07426	0	0.00352	0.3389	0.935374832	0.951758587	1.887133418	6898
6	0.02857	0.33	0.4022	0.40824	-0.0055	0.19025	0.15791	0.12547	0.37815	0.1368	0.1029	0	0.12809	0.4127	0.944220766	0.935131048	1.879351813	226786
7	0.00454	0.42294	0.35412	0.20387	0.01577	0.24271	0.23161	0.08143	0.32987	0.25884	0.11542	0	0.19942	0.4342	0.9609292	0.891897538	1.852826738	55026
8	-0.0224	0.42223	0.40265	0.28519	0.0498	0.31269	0.20894	0.06861	0.32911	0.15152	0.12776	0	0.01211	0.3151	0.921991738	0.919720385	1.841712122	68480
9	0.02856	0.40441	0.43255	0.34806	-0.0248	0.15988	0.11563	0.07701	0.42955	0.13331	0.1413	0	0.0747	0.3052	0.909632983	0.927535865	1.837168848	368372
10	0.11062	0.27673	0.52159	0.48066	-0.0343	0.10189	0.22553	0.03534	0.38667	0.05833	0.0356	0	0.0906	0.3215	0.911614119	0.925355692	1.836969811	828353
11	0.07265	0.37162	0.39474	0.3793	0.04271	0.1422	0.18862	0.09555	0.32873	0.11222	0.13014	0	0.0894	0.3246	0.916101155	0.909681138	1.825782293	36171
12	0.059	0.21158	0.38671	0.58798	0.07637	0.0922	0.1609	0.07932	0.36508	0.13127	0.06576	0	0.10024	0.2878	0.900884387	0.916523123	1.81740751	609598
13	0.07584	0.31625	0.4509	0.30096	0.00296	0.18076	0.09183	0.12232	0.4128	0.01242	0.10264	0	0.02681	0.4225	0.887811284	0.921999526	1.80981081	25278
14	0.07441	0.32919	0.38891	0.36024	-0.0822	0.16991	0.14436	0.09706	0.47691	-0.0225	0.10481	0	0.02307	0.4571	0.893571027	0.913494663	1.80706569	702603

Результаты

■ - TA query skills relevance

■ - Skills

Excel Online | EPAM github_search_feedback > Python_Dev_long_exact | weighted_skills_plus_av... - Saved | Ilya

File Home Insert Data Review View Tell me what you want to do Open in Excel Share

P3 fx 0.966580065164662

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	python	flask	aiohttp	sanic	mesos	postgres	redis	rest	websocket	docker	swagger	sql	jenkins	django	total_w_relevance_by_skills	relevance_avg_weighted	total_score	user_id
1																		
2	0.10288	0.46965	0.47037	0.36449	0.03899	0.18383	0.2286	0.03456	0.46466	0.05237	0.11731	0	0.07256	0.40627	1	1	2	309538
3	0.08465	0.52752	0.44111	0.44082	-0.0105	0.18058	0.17802	0.01933	0.39693	0.09131	0.12496	0	0.00161	0.43687	0.987100125	0.966580065	1.95368019	137309
4	0.03763	0.48407	0.48279	0.34229	0.04728	0.09708	0.29093	0.08776	0.32495	0.08904	0.17869	0	0.02397	0.3562	0.960490302	0.929032349	1.889522651	6192
5	-0.015	0.35506	0.48718	0.50174	-0.0621	0.16276	0.24948	0.07901	0.46846	0.04139	0.07426	0	0.00352	0.33895	0.935374832	0.951758587	1.887133418	6898
6	0.02857	0.33	0.4022	0.40824	-0.0055	0.19025	0.15791	0.12547	0.37815	0.1368	0.1029	0	0.12809	0.41273	0.944220766	0.935131048	1.879351813	226786
7	0.00454	0.42294	0.35412	0.20387	0.01577	0.24271	0.23161	0.08143	0.32987	0.25884	0.11542	0	0.19942	0.43429	0.9609292	0.891897538	1.852826738	55026
8	-0.0224	0.42223	0.40265	0.28519	0.0498	0.31269	0.20894	0.06861	0.32911	0.15152	0.12776	0	0.01211	0.31518	0.921991738	0.919720385	1.841712122	68480
9	0.02856	0.40441	0.43255	0.34806	-0.0248	0.15988	0.11563	0.07701	0.42955	0.13331	0.1413	0	0.0747	0.30522	0.909632983	0.927535865	1.837168848	368372
10	0.11062	0.27673	0.52159	0.48066	-0.0343	0.10189	0.22553	0.03534	0.38667	0.05833	0.0356	0	0.0906	0.32153	0.911614119	0.925355692	1.836969811	828353
11	0.07265	0.37162	0.39474	0.3793	0.04271	0.1422	0.18862	0.09555	0.32873	0.11222	0.13014	0	0.0894	0.32468	0.916101155	0.909681138	1.825782293	36171
12	0.059	0.21158	0.38671	0.58798	0.07637	0.0922	0.1609	0.07932	0.36508	0.13127	0.06576	0	0.10024	0.28786	0.900884387	0.916523123	1.81740751	609598
13	0.07584	0.31625	0.4509	0.30096	0.00296	0.18076	0.09183	0.12232	0.4128	0.01242	0.10264	0	0.02681	0.42252	0.887811284	0.921999526	1.80981081	25278
14	0.07441	0.32919	0.38891	0.36024	-0.0822	0.16991	0.14436	0.09706	0.47691	-0.0225	0.10481	0	0.02307	0.45716	0.893571027	0.913494663	1.80706569	702603

Результаты

■ - TA query skills relevance

■ - Skills

Excel Online | EPAM github_search_feedback > Python_Dev_long_exact | weighted_skills_plus_av... - Saved | Ilya

File Home Insert Data Review View Tell me what you want to do Open in Excel Share

P3 fx 0.966580065164662

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	python	flask	aihttp	sanic	mesos	postgres	redis	rest	websocke	docker	swagger	sql	jenkins	django	total_w_relevance_by_skills	relevance_avg_weighted	total_score	user_id
1																		
2	0.10288	0.46965	0.47037	0.36449	0.03899	0.18383	0.2286	0.03456	0.46466	0.05237	0.11731	0	0.07256	0.40627	1	1	2	309538
3	0.08465	0.52752	0.44111	0.44082	-0.0105	0.18058	0.17802	0.01933	0.39693	0.09131	0.12496	0	0.00161	0.43687	0.987100125	0.966580065	1.95368019	137309
4	0.03763	0.48407	0.48279	0.34229	0.04728	0.09708	0.29093	0.08776	0.32495	0.08904	0.17869	0	0.02397	0.3562	0.960490302	0.929032349	1.889522651	6192
5	-0.015	0.35506	0.48718	0.50174	-0.0621	0.16276	0.24948	0.07901	0.46846	0.04139	0.07426	0	0.00352	0.33895	0.935374832	0.951758587	1.887133418	6898
6	0.02857	0.33	0.4022	0.40824	-0.0055	0.19025	0.15791	0.12547	0.37815	0.1368	0.1029	0	0.12809	0.41273	0.944220766	0.935131048	1.879351813	226786
7	0.00454	0.42294	0.35412	0.20387	0.01577	0.24271	0.23161	0.08143	0.32987	0.25884	0.11542	0	0.19942	0.43429	0.9609292	0.891897538	1.852826738	55026
8	-0.0224	0.42223	0.40265	0.28519	0.0498	0.31269	0.20894	0.06861	0.32911	0.15152	0.12776	0	0.01211	0.31518	0.921991738	0.919720385	1.841712122	68480
9	0.02856	0.40441	0.43255	0.34806	-0.0248	0.15988	0.11563	0.07701	0.42955	0.13331	0.1413	0	0.0747	0.30522	0.909632983	0.927535865	1.837168848	368372
10	0.11062	0.27673	0.52159	0.48066	-0.0343	0.10189	0.22553	0.03534	0.38667	0.05833	0.0356	0	0.0906	0.32153	0.911614119	0.925355692	1.836969811	828353
11	0.07265	0.37162	0.39474	0.3793	0.04271	0.1422	0.18862	0.09555	0.32873	0.11222	0.13014	0	0.0894	0.32468	0.916101155	0.909681138	1.825782293	36171
12	0.059	0.21158	0.38671	0.58798	0.07637	0.0922	0.1609	0.07932	0.36508	0.13127	0.06576	0	0.10024	0.28786	0.900884387	0.916523123	1.81740751	609598
13	0.07584	0.31625	0.4509	0.30096	0.00296	0.18076	0.09183	0.12232	0.4128	0.01242	0.10264	0	0.02681	0.42252	0.887811284	0.921999526	1.80981081	25278
14	0.07441	0.32919	0.38891	0.36024	-0.0822	0.16991	0.14436	0.09706	0.47691	-0.0225	0.10481	0	0.02307	0.45716	0.893571027	0.913494663	1.80706569	702603

Результаты

- Candidates info

- TA feedbacks

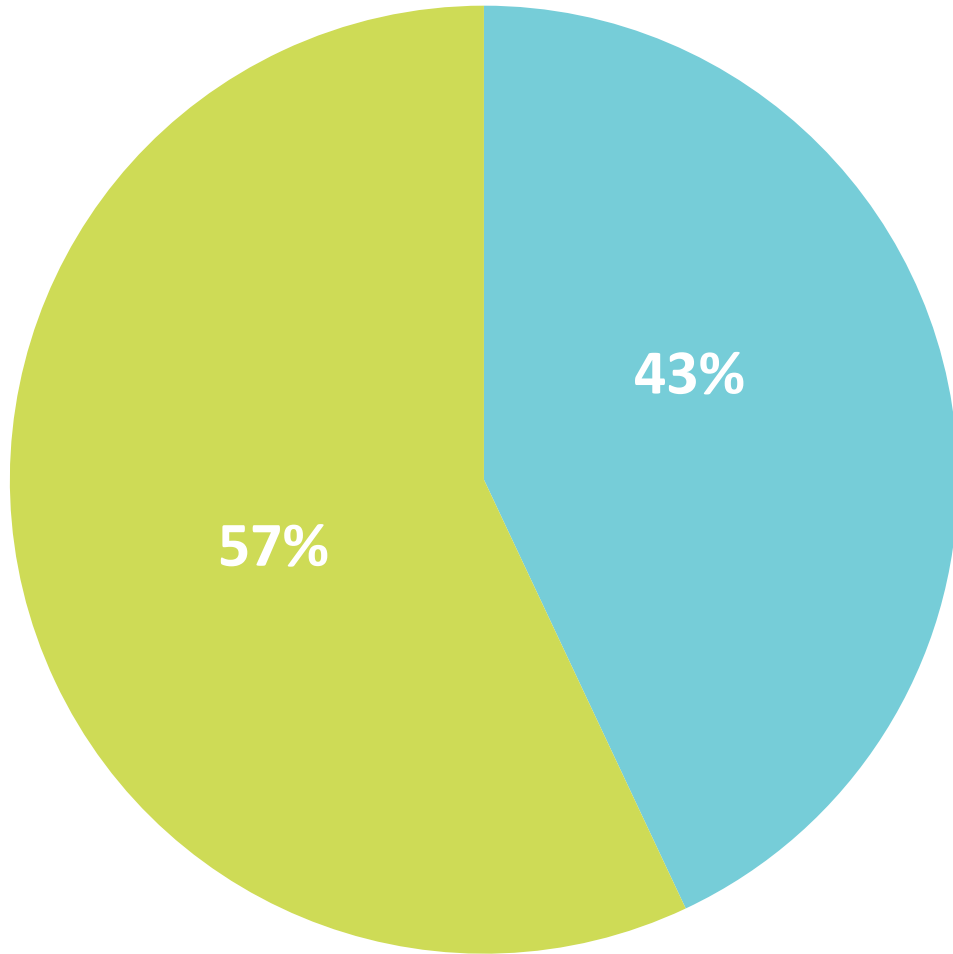
Excel Online | EPAM github_search_feedback > Python_Dev_long_exact | weighted_skills_plus_av... - Saved

File Home Insert Data Review View Tell me what you want to do Open in Excel It's just you here now Share

V4 {Python': 8, '\N': 4, 'Rust': 2, 'Go': 2}

	R	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	
	ore	user_id	company	city	languages({lang: r	top_popular_repos	count_stars	git_link	feedback	comments	status	GI	TI	MI
2	3095381	\N	\n	{'Python': 5, '\N'	1	https://github.com/divanovgh	4	нет информации						
3	1373096	\N	\n	{'Python': 12, '\N'	1	https://github.com/irvind	4	рассматривали						
4	61922	\N	novosibirsk	{'Python': 8, '\N'	22	https://github.com/unaxfromsibi	5	пригласила	reject by candidate (remote only)					
5	68982	Mail.Ru Gr	moscow	{'Python': 10, '\N'	3	https://github.com/driverx	4	пригласила						
6	2267868	\N	moscow	{'Python': 15, '\N'	18	https://github.com/renskiy	4	рассматривали						
7	550269	\N	\n	{'Python': 17, 'Jav	10	https://github.com/toxicwar	4	рассматриваем						
8	684806	\N	chelyabinsk	{'\N': 22, 'Pythor	35	https://github.com/clayman74	4	пригласила (нов)						
9	3683723	\N	nizhny nov	{'Python': 18, '\N'	2	https://github.com/notapresent	4	пригласила (нов.проф on hold)						
10	8283534	\N	\n	{'Python': 11, '\N'	7	https://github.com/pyvim	1	нет информации						
11	361715	\N	\n	{'Python': 5, 'Vim		https://github.com/oleynikandre	1	другой профиль						
12	6095983	\N	yekaterinb	{'\N': 9, 'Pythor	5	https://github.com/kurlov	4	пригласила						
13	252780	Uralairline	yekaterinb	{'Python': 20, '\N'	11	https://github.com/mnach	4	рассматривали						
14	7026035	@idaproje	moscow	{'\N': 20, 'Pythor	5	https://github.com/artinnok	4	другой профиль						

Результаты



5 technologies:

- Python developers
- Java
- JavaScript full-stack developers (Node.JS + React.JS)
- DevOps Engineers
- Data Science

~500 profiles

Результаты

13%

НОВЫХ КАНДИДАТОВ

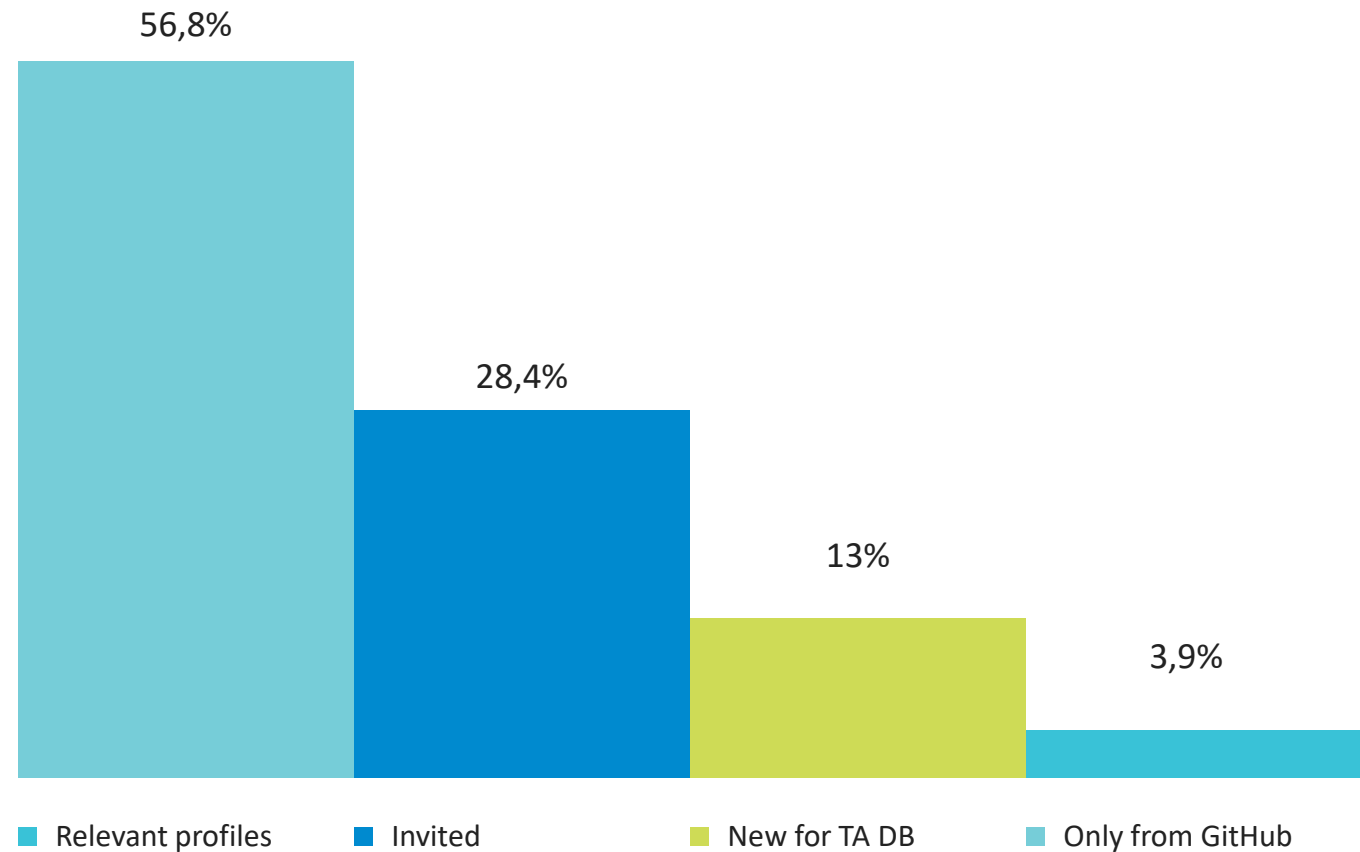
3.9%

смогли найти
только через GitHub

Профайлов в БД:

- 935к по миру
- 112к по России

Results in percentage



Что хорошего для TA team

5.5 min

для добавления нового
контакта через GH и e-mail

7.7 min

через LinkedIn

- Расширение базы кандидатов
- Использование для поиска редких кандидатов
- Экономия времени

Дальнейшие шаги

- **Создание приложения для рекрутеров**
Интеграция с Amazing Hiring plugin
- **Слияние с Stack Overflow data для увеличения точности**
Определение уровней кандидатов
- **Слияние с вн. БД кандидатов для экономии времени**
Распространение алгоритма на вн. БД кандидатов



THANK YOU

