

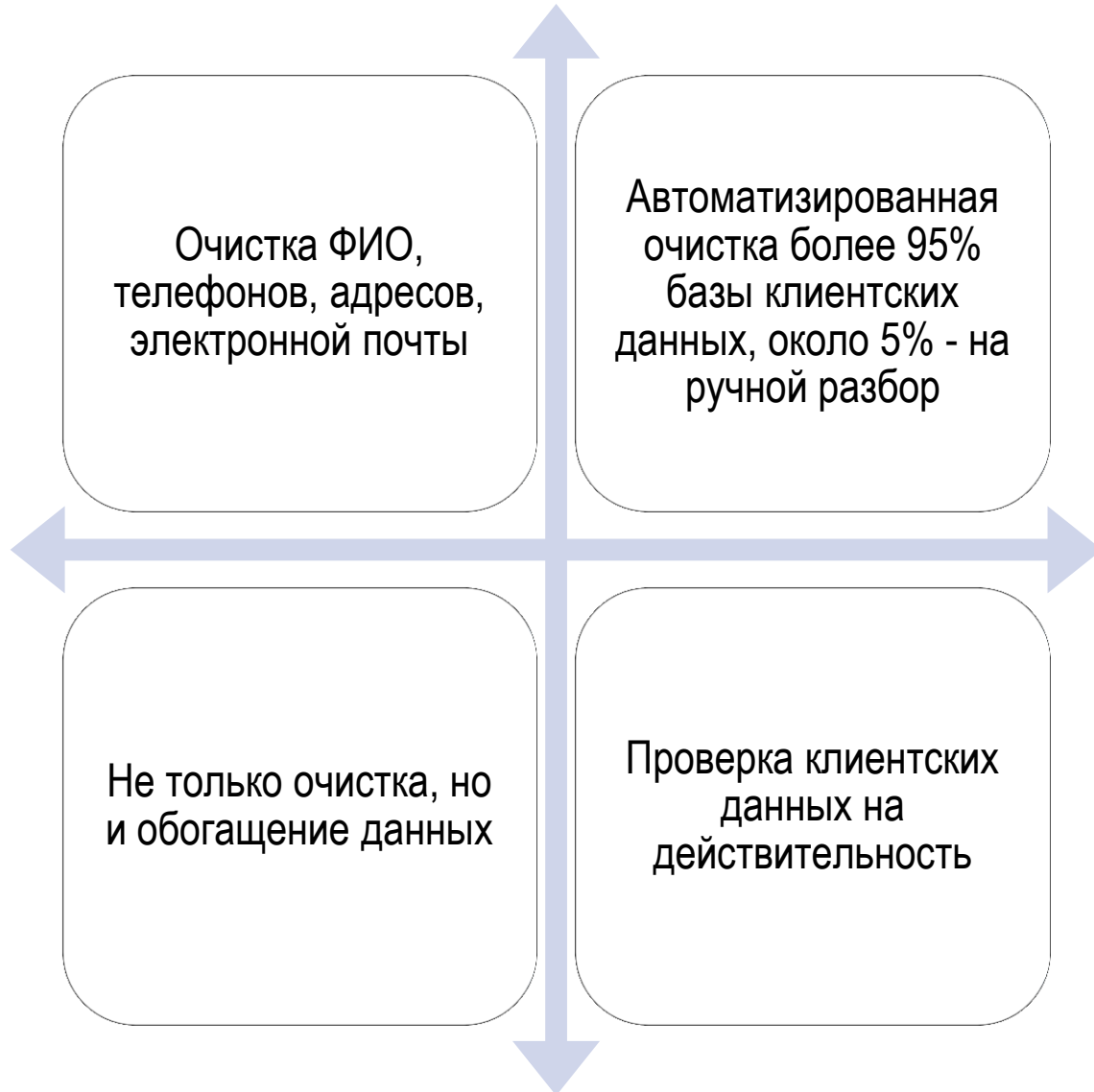
Чистая база

Решение для Вашего бизнеса

Москва - 2022



Концепция системы очистки клиентских данных



1

Профилирование – акт мониторинга и очистки данных - используется для проверки данных по стандартным статистическим мерам, выявления связей и проверки данных по сопоставимым описаниям

2

Валидация - это процесс проверки данных различных типов по критериям корректности и полезности для конкретного применения

3

Обогащение – процесс позволяющий автоматически обогатить данные, если это возможно

4

Инциденты – создание инцидентов данных требующих ручного разбора

Перечень услуг включает в себя 3 блока

Очистка и стандартизация

- Создание полного адресного профиля клиента, проверка кода ФИАС на действительность
- Стандартизация и очистка ФИО, проверка на базе типовых и нетиповых имен
- Стандартизация и очистка даты и номеров паспортов, проверка на действительность
- Стандартизация и очистка адресов электронной почты, телефона, качественная проверка на существование, проверка по базе доменов
- Стандартизация и очистка ИНН, СНИЛС

Обогащение

- Построение витрины данных с учетом обогащения из внутренних (ЛКК, СУВК, СУЭД, 1С и пр.), внешних источников при наличии
- Дедупликация, обнаружение дублей и похожих записей, объединение их в группы

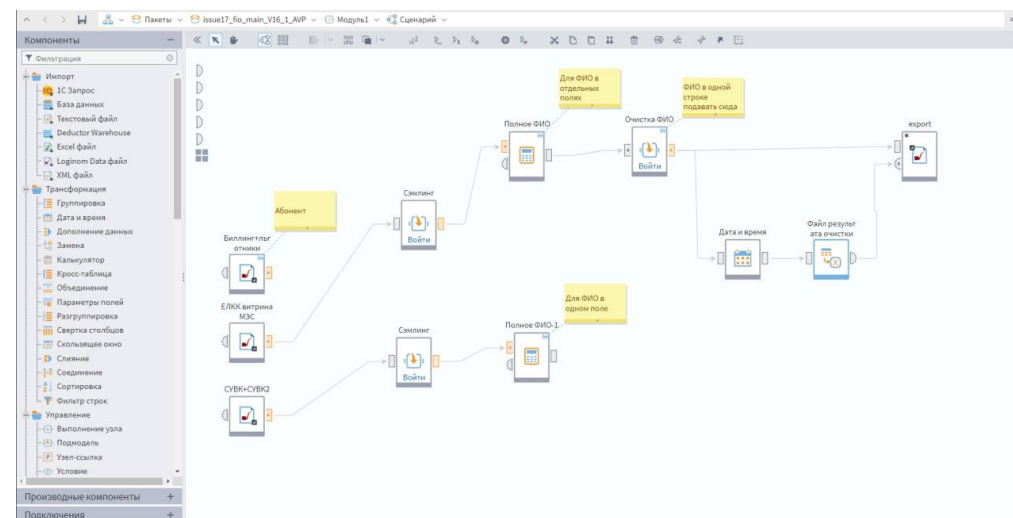
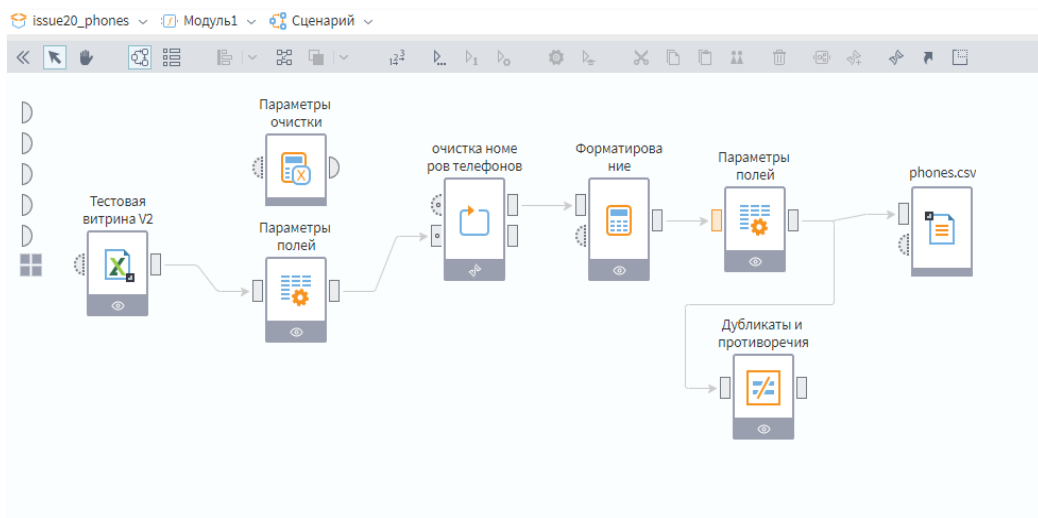
Загрузка

- Подготовка данных для загрузки в БД по согласованным шаблонам
- Контроль исполнения работ по корректировке и обогащению данными

Инфраструктура: Loginom, DaData, Yandex Cloud

#	ab Комментарий	ab Исходные данные Первый	ab Очищенный e-mail Перв...	ab id Перв...
1		glebovich84@mail.ru;krassilnikovfedor@gmail.com	glebovich84@mail.ru	1
2	Email успешно очищен. Домен найден в БД.	gromova.kseniya@gmail.com justelvi@mail.ru	gromova.kseniya@gmail.com	20
3	Email успешно очищен. Домен без точек найден в БД. Исправлено расположение точек в домене.	fa-gorobet@mail.ru	fa-gorobet@mail.ru	110
4	Email успешно очищен. Домен без точек найден в БД. Исправлено расположение точек в домене. Множественные точки заменены на одну.	tanya_landbest@mail.ru	tanya_landbest@mail.ru	108
5	Email успешно очищен. Домен без разделителей найден в БД. Исправлены разделители в домене.	taniana22@g-mail.com	taniana22@gmail.com	340
6	Email успешно очищен. Домен восстановлен с помощью нечеткого поиска.	dishy-ig2009@yaru	dishy-ig2009@ya.ru	146
7	Email успешно очищен. Ни один шаблон не подходит. Домен восстановлен с помощью нечеткого поиска.	isstan@mail.ru	isstan@mail.ru	116
8	Email очистить не удалось. Исправлено расположение точек в домене. Ни один шаблон не подходит.	@mail.ru	<null>	251
9	<u>Домен не найден ни в БД, ни во внешнем сервисе.</u>	minkina5@mail.ru	<null>	201
10	Email очистить не удалось. Выделенный email пустой. Ни один шаблон не подходит.	irihhka@mail.ru ls.a.s861986@mail.ru	<null>	12
11	Email очистить не удалось. Ни один шаблон не подходит.	@	<null>	253

Очищенный e-mail

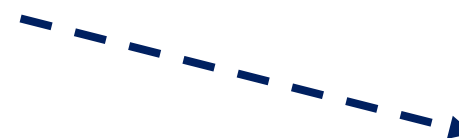
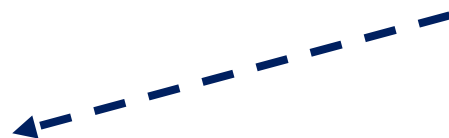


Сценарии очистки номера телефона (слева) и ФИО (справа)

Loginom – Self-service analytics



Loginom



1

IT-инфраструктура – возможность пользователя подключиться к разным данным (файлы (Excel, CSV, XML, Loginom Data File), базы данных (Oracle, MS SQL, PostgreSQL, ClickHouse, BigQuery, MySQL...) и т.д.

3

Анализ и создание отчетов – большой набор элементов визуализации, возможность автоматического создания показателей для визуализации, построения OLAP-кубов, возможность самостоятельной выгрузки отчета

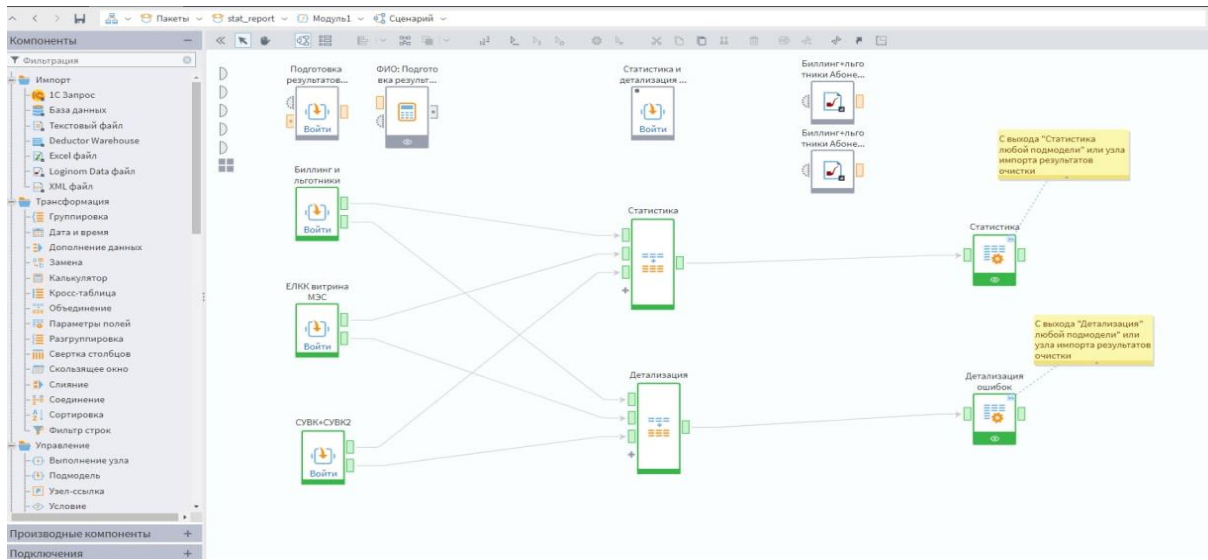
2

Анализ данных – предварительная обработка данных, более 60-и готовых компонентов обработки, как тривиальных, так и достаточно сложных. Например, основанных на алгоритмах машинного обучения

4

Управление данными – быстрый поиск, хранение, использование и публикацию объектов метаданных (показателей), возможность развертывания stand-alone (инсталляция на ПК), on-premise (установка на сервер), cloud (публичный облачный сервис)

Пример из практики

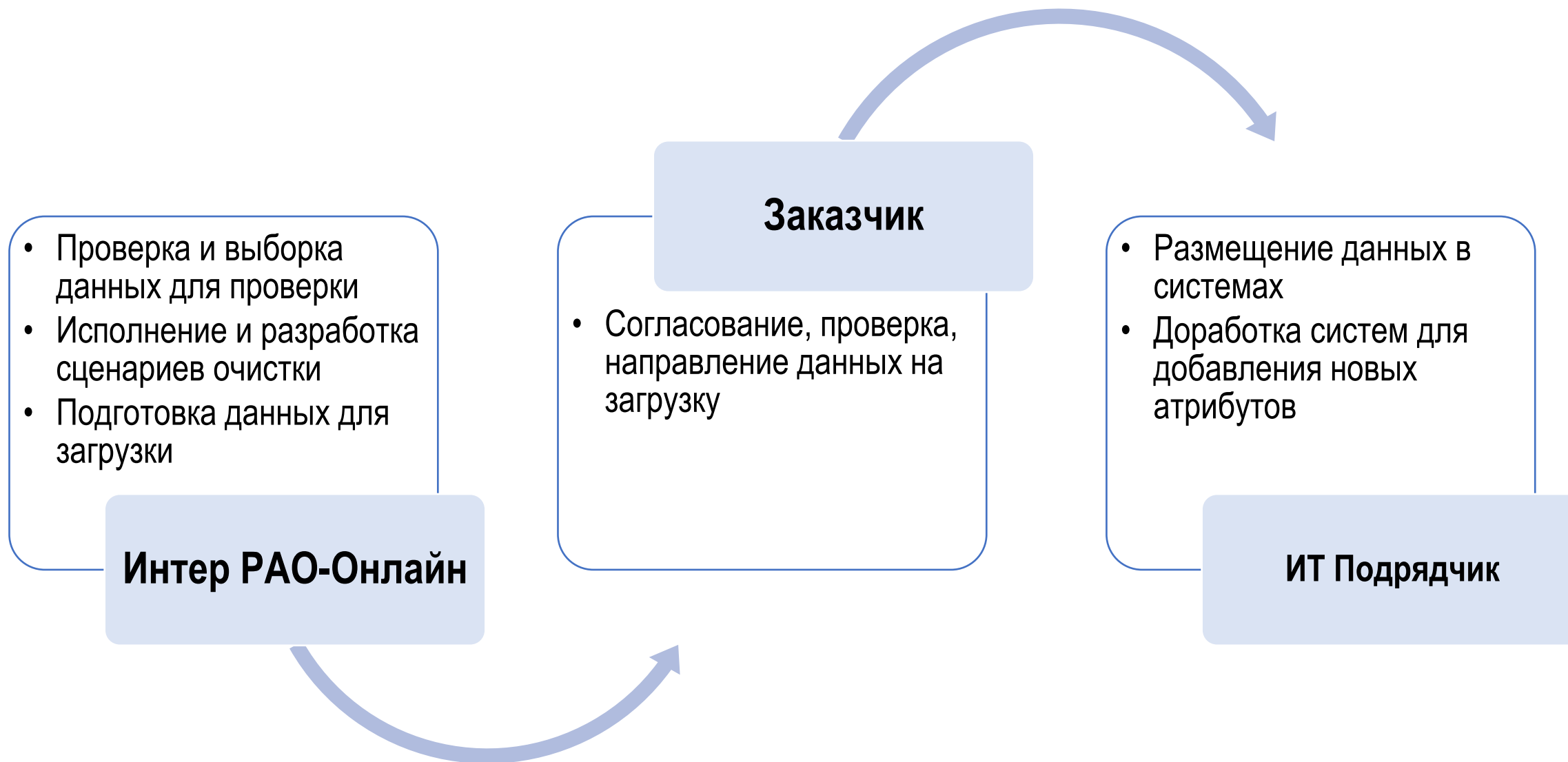


Выявлено и удалено 100%
фейковых номеров и Email.
Число новых ошибок упало на 30%

	> Email	> Дата выдач...	> Дата рожде...	> ИНН	> Серия и но...	> СНИЛС	> Телефон	> ФИО	Итого:
AO "Мосэнергобыв"									
> ВОСТОЧ...	1 075 088	8 648	460 574		41 091	3 863	3 842 604	840 750	6 272 618
> ЗАПАДН...	1 093 427	12 319	458 925	2	29 699	1 782	3 742 360	1 044 520	6 383 034
> ОТДЕЛ...	81 895	8 984	330 012		26 270	22	702 081	189 832	1 339 096
> ОТДЕЛ...	765 520	7 042	284 809	18	19 279	1 082	2 438 886	719 434	4 236 070
> СЕВЕРН...	932 163	38 094	405 085		38 667	2 111	3 541 010	1 077 179	6 034 309
> СЕВЕРО...	902 358	27 554	435 424		21 083	2 689	3 728 598	748 073	5 865 779
> СЕВЕРО...	677 339	21 610	362 670		50 460	2 040	2 452 609	591 284	4 158 012
> ЮГО-ВО...	830 390	23 112	382 360		43 326	2 671	3 464 339	850 011	5 596 209
> ЮГО-ЗА...	1 499 142	45 899	494 994		47 260	2 249	3 873 333	1 016 024	6 978 901
> ЮЖНОЕ ...	1 323 807	30 886	527 823		47 466	3 247	4 533 857	839 880	7 306 966
Итого:	9 181 129	224 148	4 142 676	20	364 601	21 756	32 319 677	7 916 987	54 170 994
AO "Мосэнергобыв"									
> ВОСТОЧ...	39 583	37 457	421 810		41 090	1	548 252	159 852	1 248 045
> ДМИТР...	23 950	65 666	337 570		77 083		619 326	128 191	1 251 786
> ЗАПАДН...	30 078	34 911	391 712	2	53 216		643 824	205 086	1 358 829
> КАШИР...	45 108	38 641	229 768		57 996		489 849	63 140	924 502
> КОЛОМ...	39 049	19 962	256 630	69	53 691		520 212	77 567	967 180
> МОЖАЙ...	35 582	30 886	213 336	1	40 844		399 188	72 040	791 877
> НОГИНС...	20 867	32 590	251 002		30 614		408 562	83 908	827 543
> ПОДОЛ...	39 515	78 598	443 244		56 401		647 720	150 995	1 416 473
> СЕВЕРН...	23 571	9 088	370 830		37 891	65	627 331	237 361	1 306 137
> СЕВЕРО...	13 856	18 550	335 204	2	13 260		465 245	215 385	1 061 502
> СЕВЕРО...	29 688	52 596	347 710		42 619		550 905	145 952	1 169 470
> ШАТУРС...	32 031	47 920	217 352	67	35 400		401 438	70 851	805 059
> ЮЖНОЕ ...	35 624	55 405	449 984	9	63 090	3	720 901	188 944	1 513 960
Итого:	408 502	522 270	4 266 152	150	603 195	69	7 042 753	1 799 272	14 642 363
Итого:	9 589 631	746 418	8 408 828	170	967 796	21 825	39 362 430	9 716 259	68 813 357

		Телефон			ФИО			Итого:				
		Незаполне...	Исправл...	Количеств...	Сошибка...	Незаполне...	Исправл...	Количеств...	Сошибка...	Незаполне...	Исправл...	Количеств...
AO "Мосэнергобыв"												
> ОТДЕЛ...		40,44	57,10	2 096 403	13,83	13,33	2,78	70 591	20,39	51,50	16,47	2 495 753
> СЕВЕРН...		40,30	57,47	3 105 068	14,77	12,89	6,23	395 996	20,91	51,27	17,68	3 985 578
> СЕВЕРО...		39,16	59,14	3 323 242	10,99	12,21	3,30	153 462	19,91	51,48	17,10	3 977 755
> СЕВЕРО...		36,93	60,77	2 144 301	10,90	13,77	3,78	144 702	20,04	51,57	17,15	2 700 315
> ЮГО-ВО...		41,42	55,90	2 999 814	12,37	12,82	3,92	206 916	19,86	52,55	16,66	3 654 452
> ЮГО-ЗА...		38,31	59,39	3 370 969	12,88	13,10	4,53	265 740	21,30	50,20	17,66	4 227 861
> ЮЖНОЕ ...		37,95	60,33	4 000 550	10,06	12,18	3,25	165 880	20,16	50,60	17,28	4 779 182
Итого:		39,31	58,48	28 386 149	12,03	13,36	3,94	1 891 592	20,17	51,61	17,03	35 024 977
AO "Мосэнергобыв"												
> ВОСТОЧ...		37,10	61,88	540 776	10,56	18,15	4,51	47 016	21,34	53,23	19,35	1 047 761
> ДМИТР...		34,48	64,70	613 351	8,81	21,33	3,60	42 450	19,93	54,09	18,04	1 059 958
> ЗАПАДН...		34,57	64,95	639 246	10,06	21,28	3,66	47 423	19,37	54,81	17,29	1 114 189
> КАШИР...		31,31	67,21	483 271	6,89	21,36	3,42	26 244	19,66	54,39	17,81	778 547
> КОЛОМ...		30,53	67,43	512 998	7,36	21,42	3,44	27 516	19,33	52,22	17,34	818 150
> МОЖАЙ...		31,45	67,47	394 594	7,79	21,45	3,44	24 199	19,81	52,65	17,84	663 785
> НОГИНС...		33,89	64,69	403 814	8,30	20,34	3,53	30 952	20,14	54,30	18,32	719 147
> ПОДОЛ...		34,20	64,75	640 257	8,81	20,33	4,30	68 326	20,85	51,46	19,01	1 231 603
> СЕВЕРН...		34,00	65,03	621 475	10,22	20,19	3,48	42 672	19,16	54,81	17,05	1 044 741
> СЕВЕРО...		37,97	61,73	462 484	13,57	20,80	2,72	27 280	19,09	58,58	16,55	844 167
> СЕВЕРО...		34,46	64,99	545 800	9,50	21,21	4,25	62 095	19,97	52,53	18,19	1 009 321
> ШАТУРС...		34,65	63,34	395 465	9,67	17,98	4,55	28 740	23,33	50,89	21,28	690 302
> ЮЖНОЕ ...		33,88	65,33	715 288	9,69	21,32	4,40	85 506	19,62	54,83	17,74	1 307 345
Итого:		33,78	65,10	6 968 819	9,15	20,65	3,78	560 419	20,03	53,76	18,06	12 329 016
Итого:		35,68	62,83	35 354 968	10,02	18,44	3,83	2 452 011	20,07	53,16	17,77	47 353 993
Итого:		35,68	62,83	35 354 968	10,02	18,44	3,83	2 452 011	20,07	53,16	17,77	47 353 993

Бизнес-процесс актуализации данных



Спасибо за внимание

Директор по данным
Анисимов Владимир
+7 916 040 87 92