





Введение в анализ данных Лекция 3.3

Шевцов Василий Викторович, директор ДИТ РУДН, shevtsov_vv@rudn.university

Форматы. Условное форматирование. Работа с большими табличными массивами

Фильтрация данных. Срезы. Расширенный фильтр. Подведение промежуточных итогов. Консолидация данных





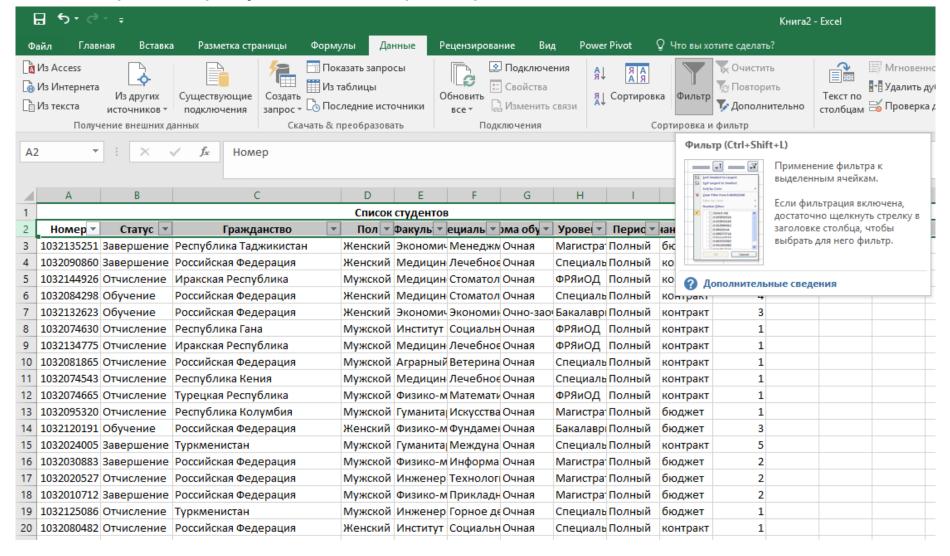
Фильтры



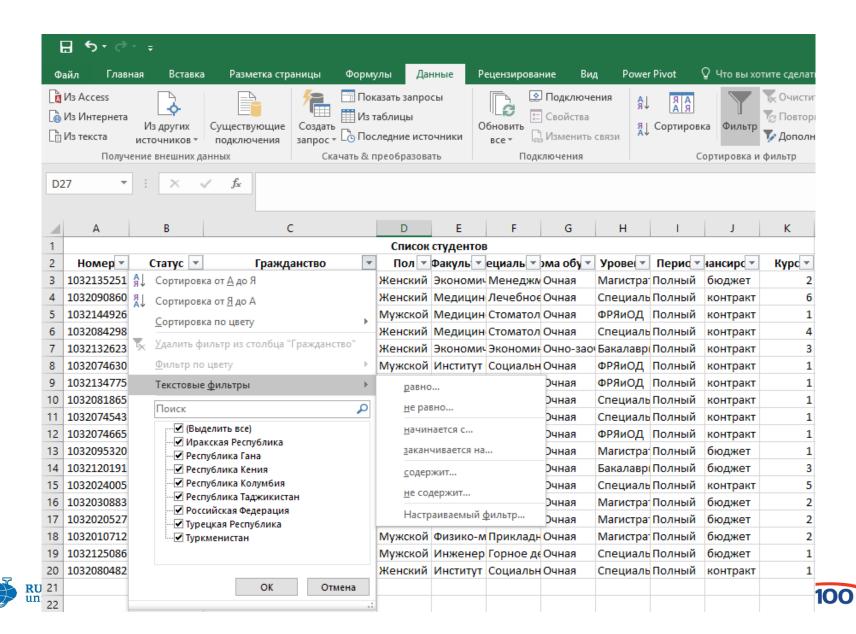


Фильтр

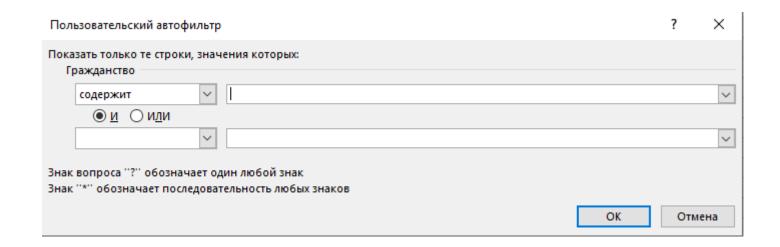
Выбрать строку заголовка фильтр, Ctrl+Shift+L



Текстовые фильтры



Пользовательский тектовый фильтр



Знак вопроса "?" означает один любой знак Знак звездочки "*" означает последовательность любых знаков

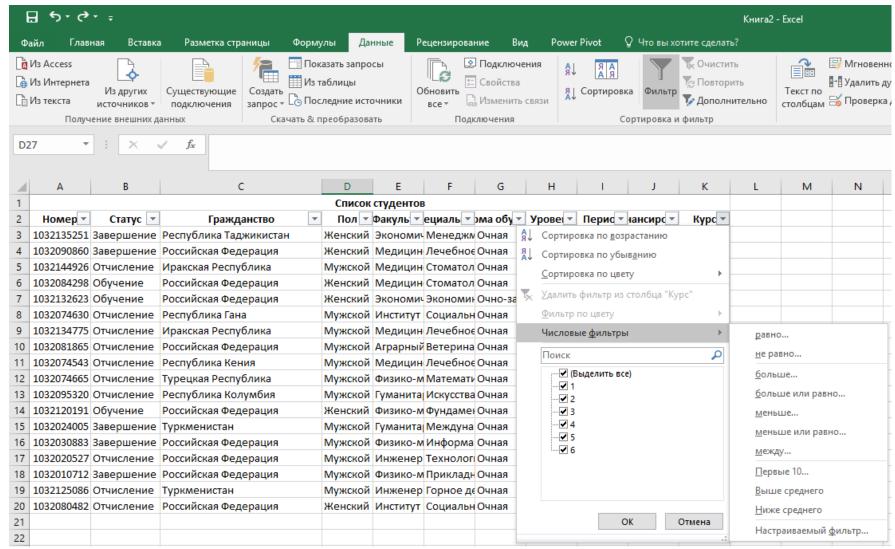


Подстановочные знаки применимы только к текстовым данным!





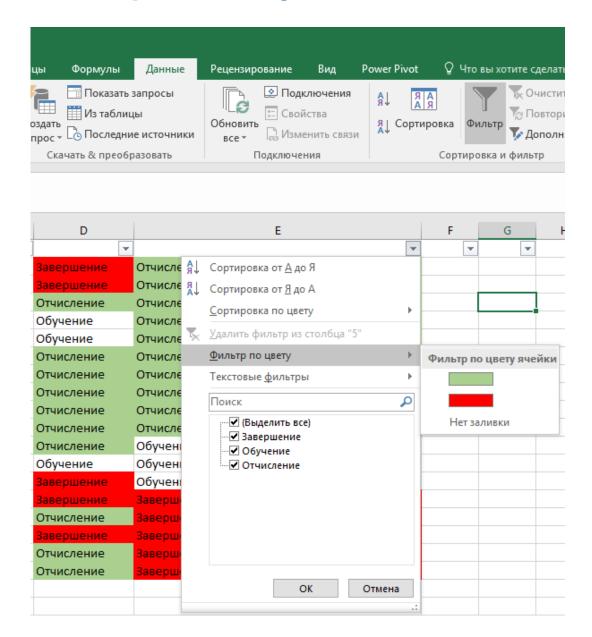
Числовой фильтр

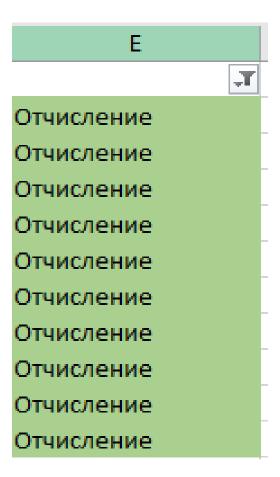






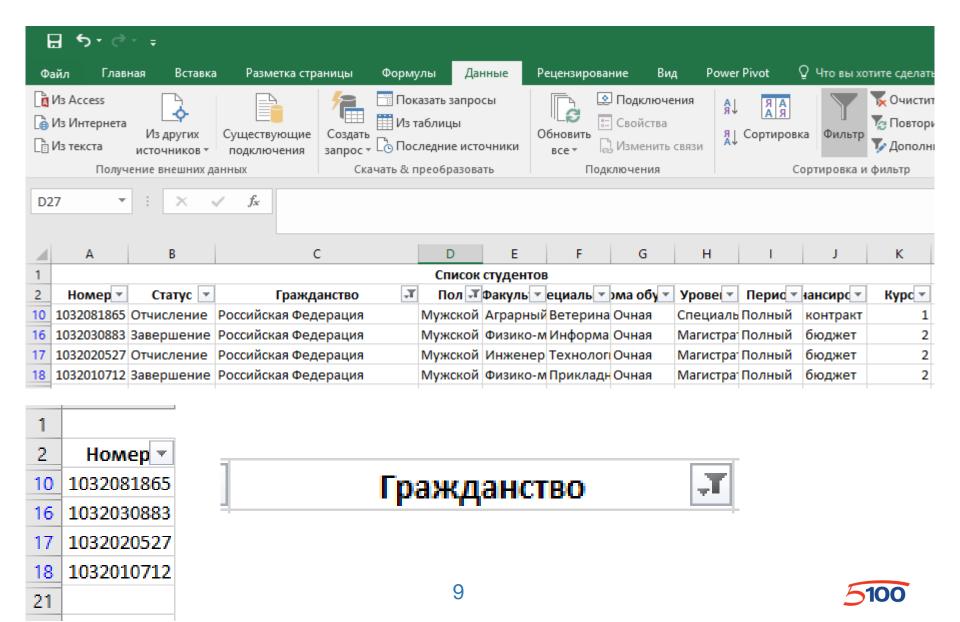
Фильтр по цвету ячейки



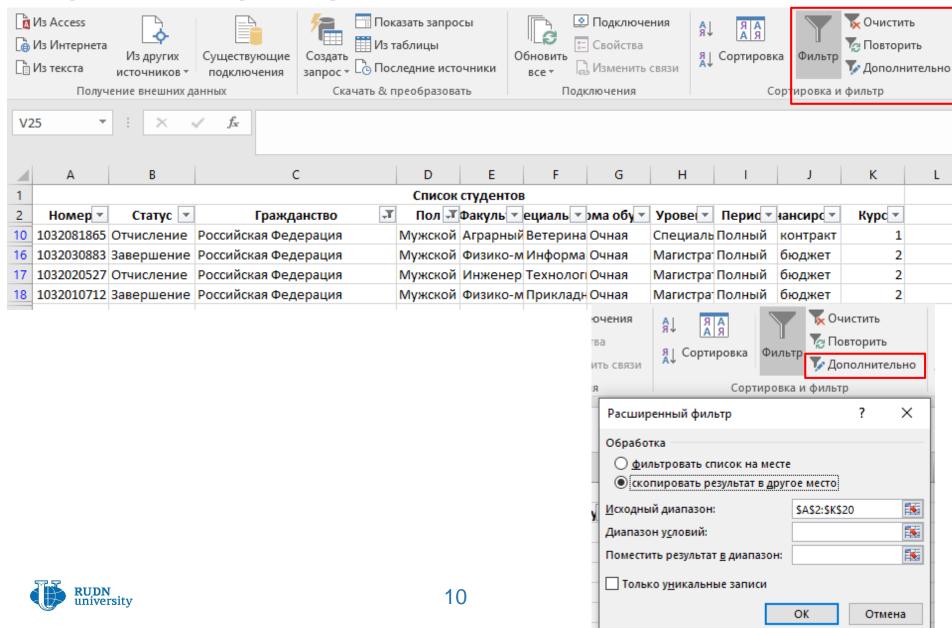




Применение фильтра



Управление фильтром

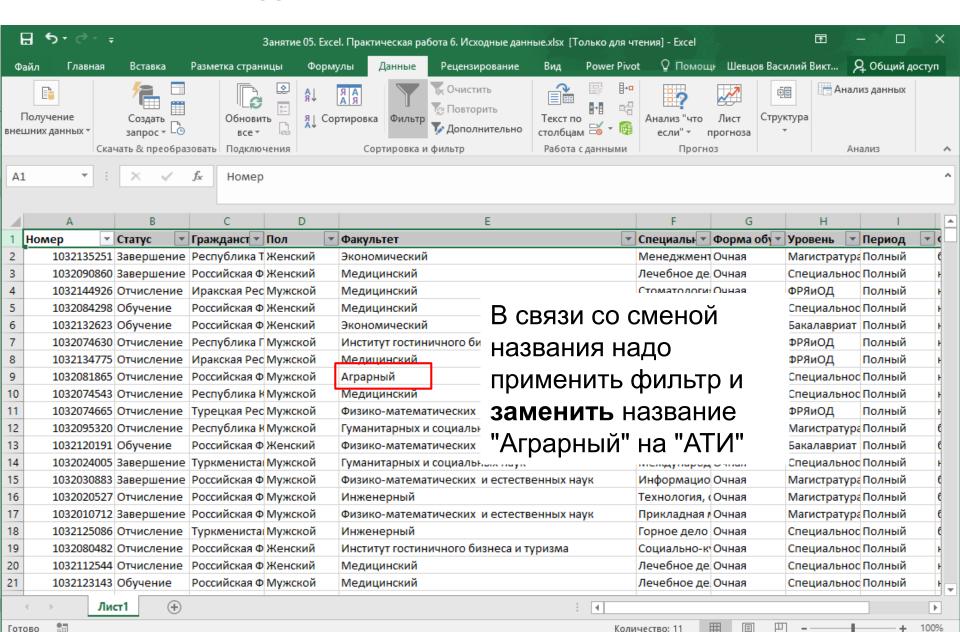


Особенности работы с фильтром

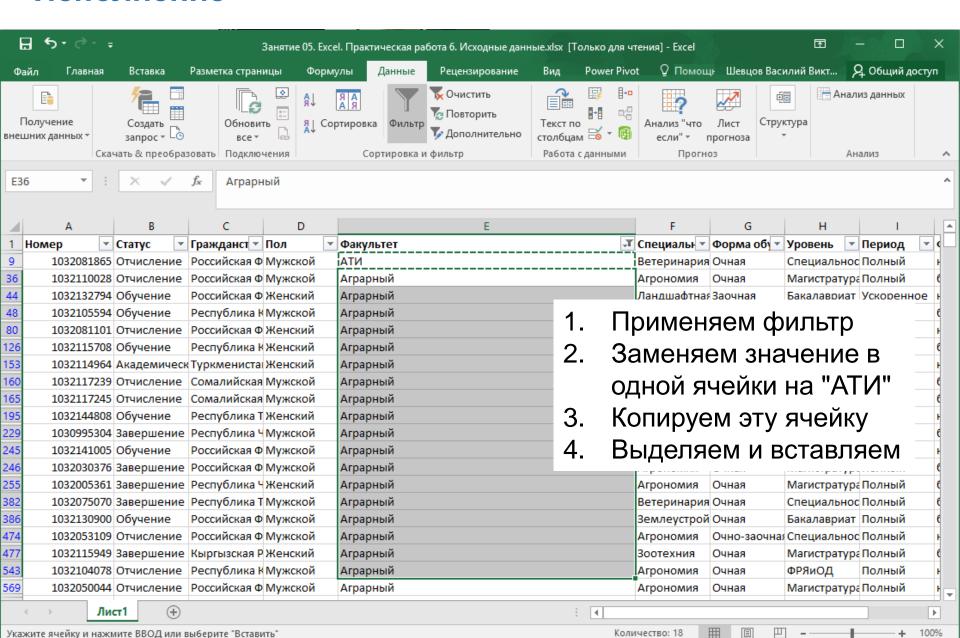




Постановка задачи

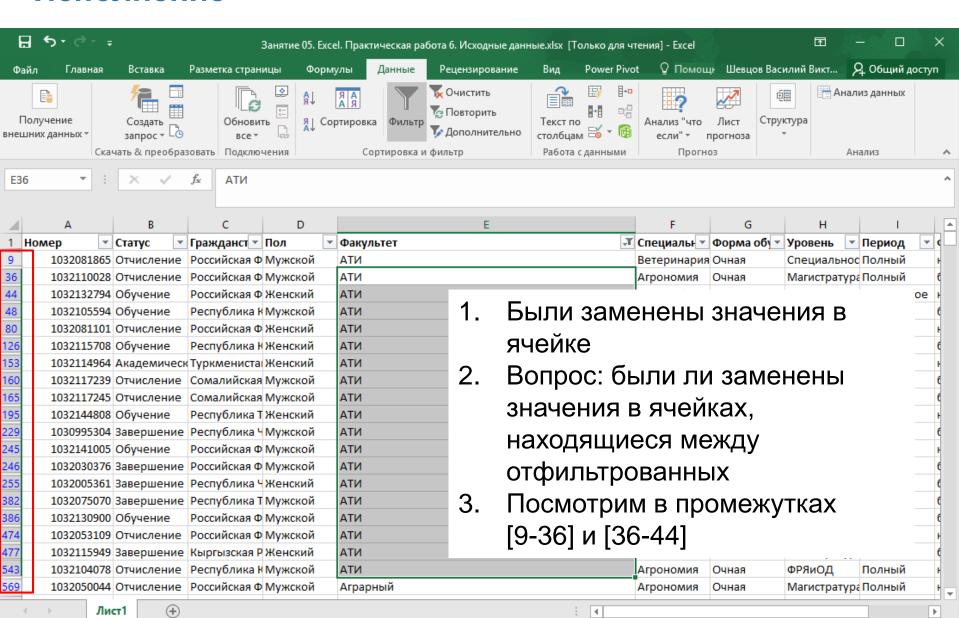


Исполнение



Исполнение

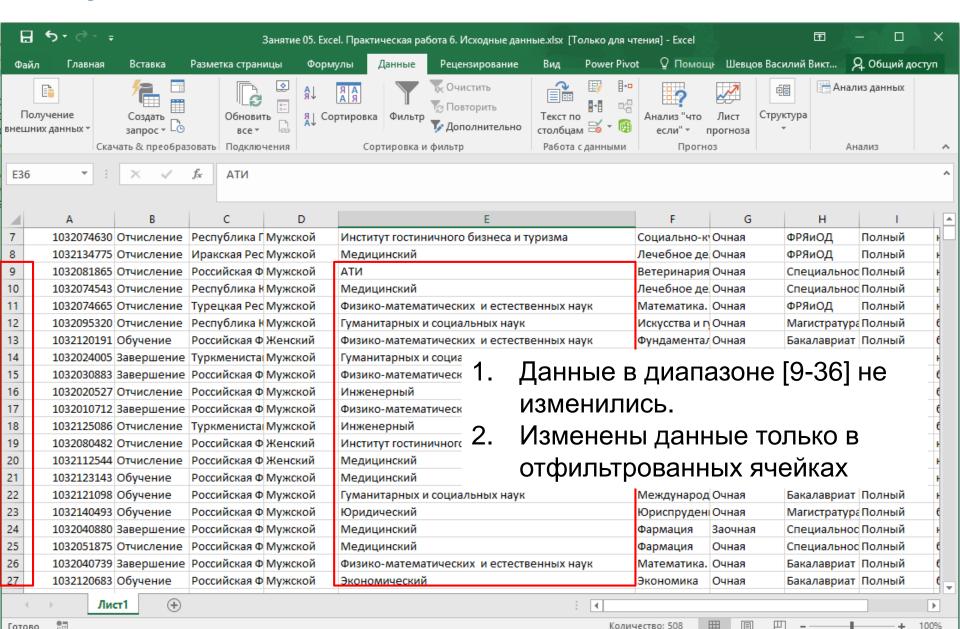
Найдено записей: 3749 из 69330



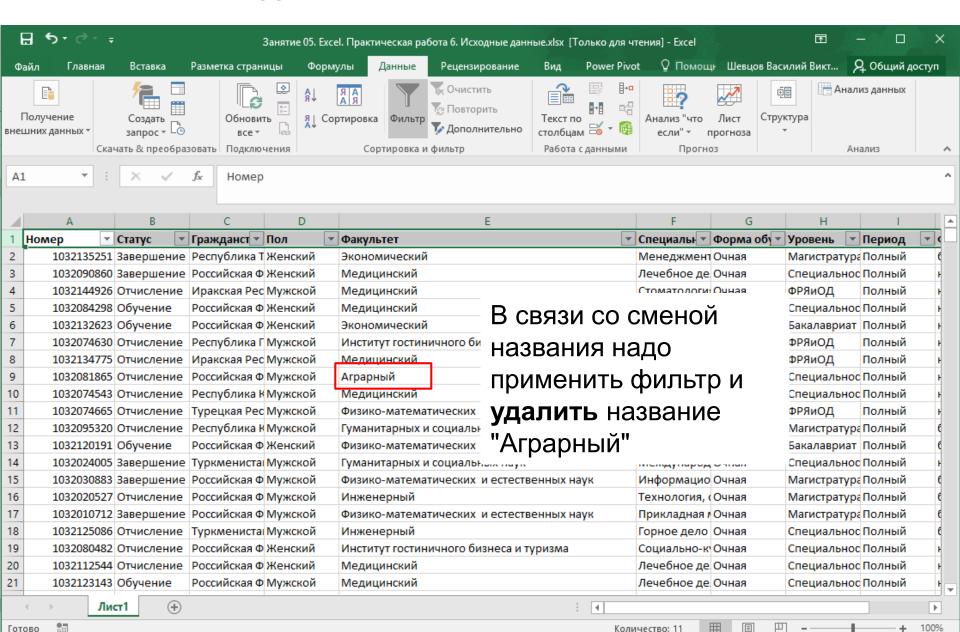
Количество: 18

Результаты

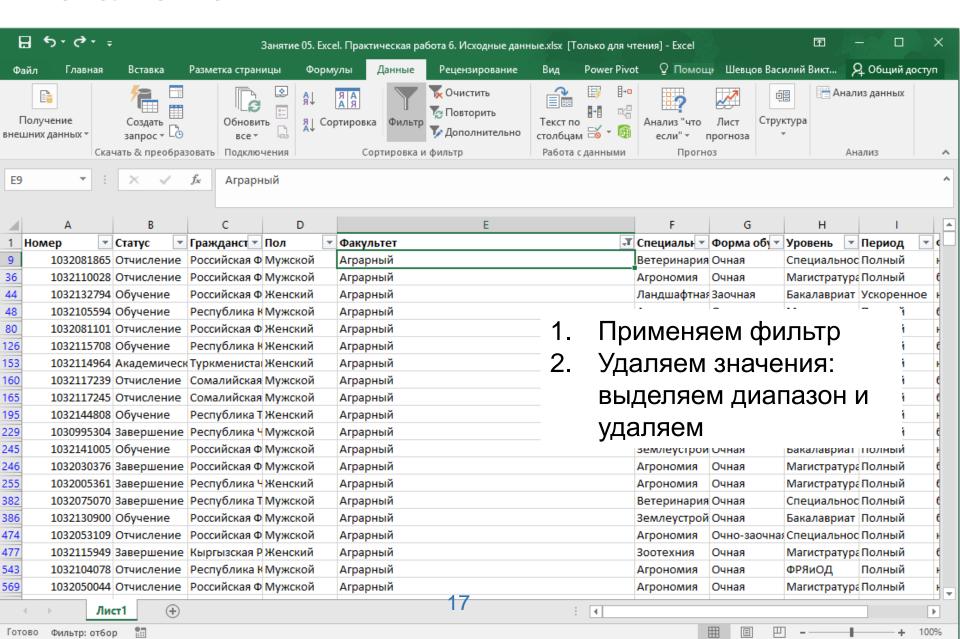
Готово



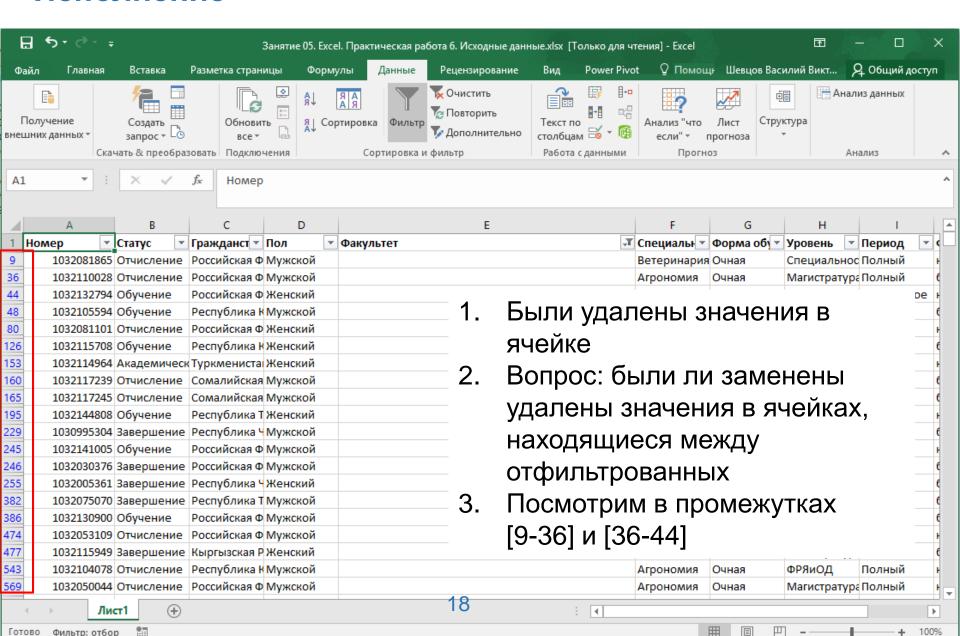
Постановка задачи



Исполнение

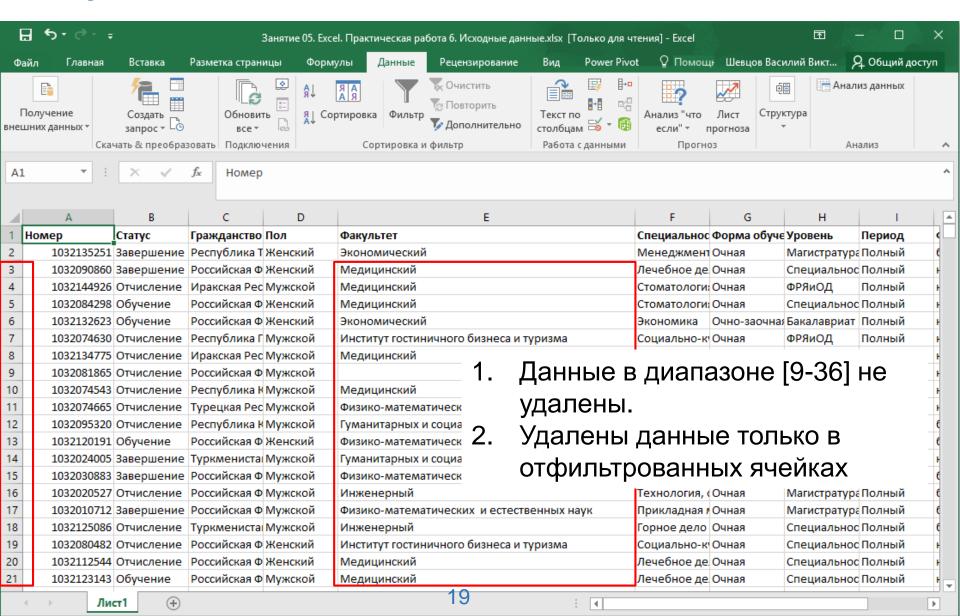


Исполнение



Результаты

Готово



100%

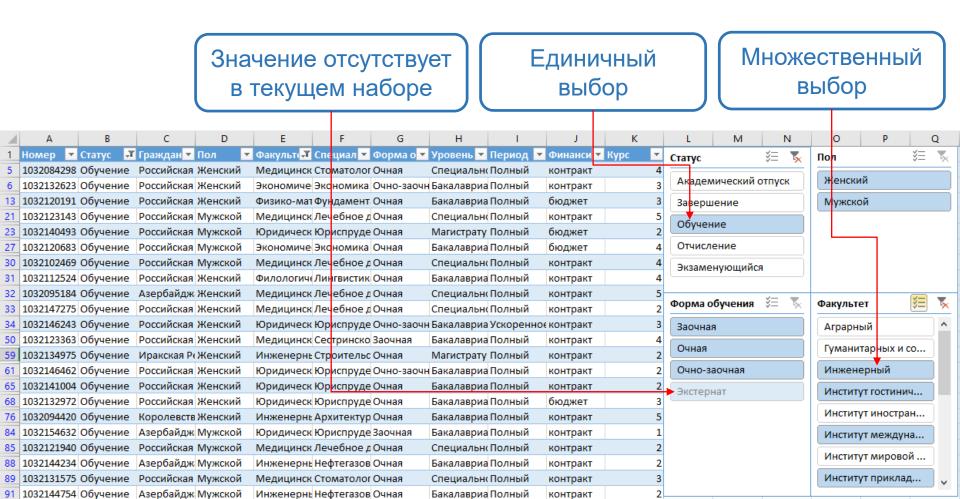
Срезы





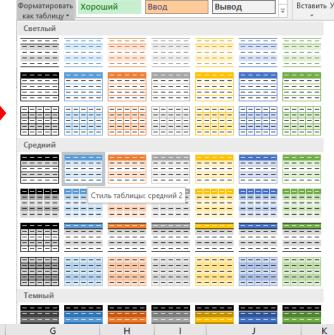
Срезы. Пример

Срезы предоставляют кнопки, с помощью которых можно фильтровать данные таблиц или сводных таблиц. Помимо поддержки быстрой фильтрации срезы также указывают текущее состояние фильтрации, что позволяет легко понять данные в отфильтрованной сводной таблице.



Создание срезов. Форматировать как таблицу

	Α	В	С	D	E	F	G	Н	1	J	K
1	Номер	Статус	Гражданс	Пол	Факульте	Специаль	Форма об	Уровень	Период	Финансир	Курс
2	1032084298	Обучение	Российска	Женский	Медицин	Стоматол	Очная	Специаль	Полный	контракт	4
3	1032132623	Обучениє	Российска	Женский	Экономич	Экономин	Очно-заоч	Бакалаврі	Полный	контракт	3
4	1032120191	Обучениє	Российска	Женский	Физико-м	Фундамен	Очная	Бакалаврі	Полный	бюджет	3
5	1032123143	Обучениє	Российска	Мужской	Медицин	Лечебноє	Очная	Специаль	Полный	контракт	5
6	1032140493	Обучениє	Российска	Мужской	Юридиче	Юриспруд	Очная	Магистра	Полный	бюджет	2
7	1032120683	Обучениє	Российска	Мужской	Экономич	Экономин	Очная	Бакалаврі	Полный	бюджет	4
8	1032102469	Обучениє	Российска	Мужской	Медицин	Лечебноє	Очная	Специаль	Полный	контракт	4
9	1032112524	Обучениє	Российска	Женский	Филологи	Лингвисть	Очная	Бакалаврі	Полный	контракт	4
10	1032095184	Обучение	Азербайд	Женский	Медицин	Лечебноє	Очная	Специаль	Полный	контракт	5
11	1032147275	Обучение	Российска	Женский	Медицин	Лечебноє	Очная	Специаль	Полный	контракт	2
12	1032146243	Обучение	Российска	Женский	Юридиче	Юриспруд	Очно-заоч	Бакалаврі	Ускоренн	контракт	3
13	1032123363	Обучение	Российска	Женский	Медицин	Сестринс	Заочная	Бакалаврі	Полный	контракт	4
14	1032134975	Обучение	Иракская	Женский	Инженер	Строител	Очная	Магистра	Полный	контракт	2
15	1032146462	Обучение	Российска	Женский	Юридиче	Юриспруд	Очно-заоч	Бакалаврі	Полный	контракт	2
16	1032141004	Обучение	Российска	Женский	Юридиче	Юриспруд	Очная	Бакалаврі	Полный	контракт	2
17	1032132972	-						Бакалаврі	Полный	бюджет	3
		_									



Нейтральный

Плохой

Обычный

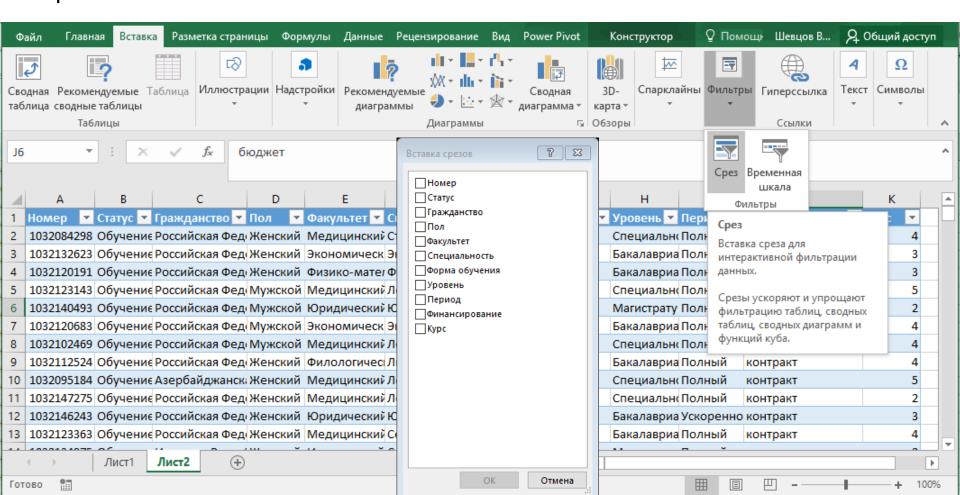
Для добавления среза необходимо диапазон с данными форматировать как таблицу

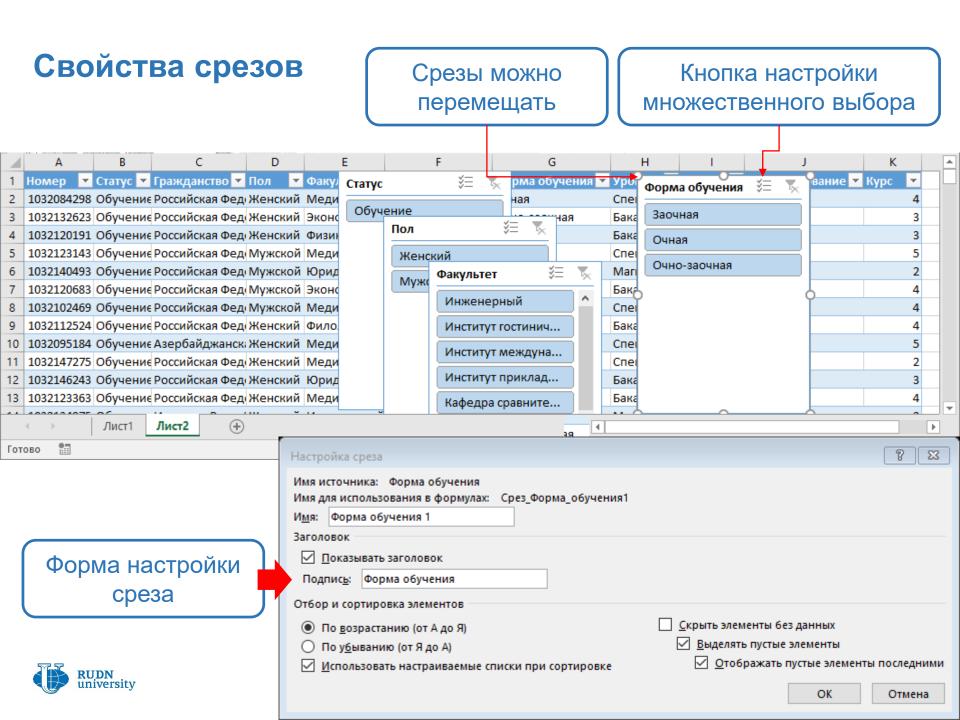


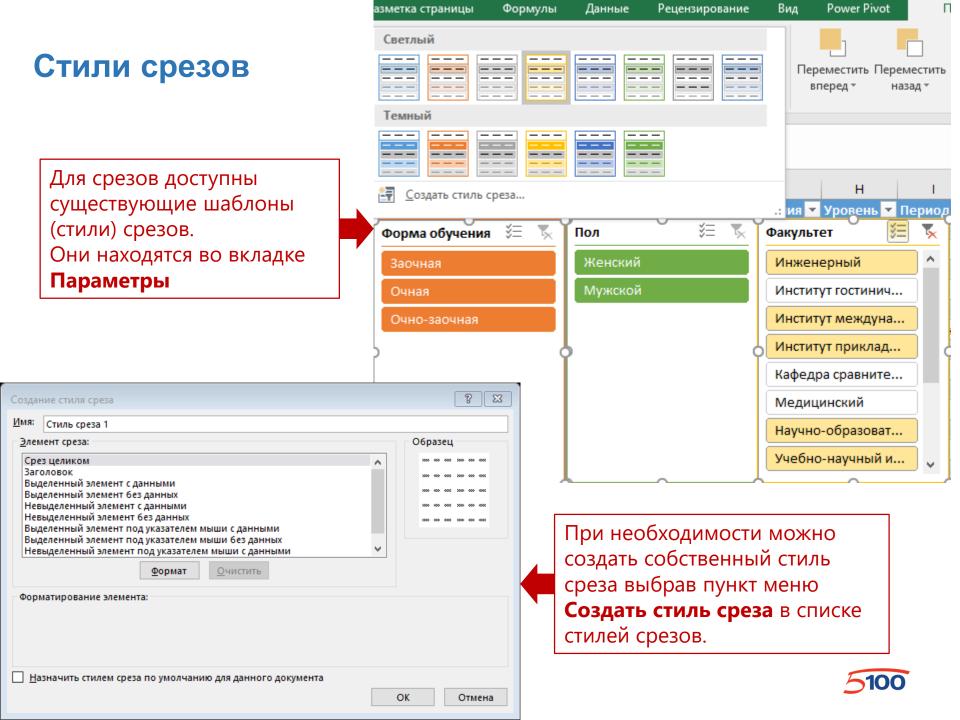
	Α	В	С	D	E	F	G	Н	1	J	K
1	Номер 💌	Статус 💌	Гражданство 💌	Пол 🔻	Факультет 💌	Специальность 💌	Форма обучения 💌	Уровень ▼	Период 🖸	Финансирование 💌	Курс 💌
2	1032084298	Обучениє	Российская Фед	Женский	Медицинский	Стоматология	Очная	Специальн	Полный	контракт	4
3	1032132623	Обучениє	Российская Фед	Женский	Экономическ	Экономика	Очно-заочная	Бакалавриа	Полный	контракт	3
4	1032120191	Обучениє	Российская Фед	Женский	Физико-мател	Фундаментальные	Очная	Бакалавриа	Полный	бюджет	3
5	1032123143	Обучениє	Российская Фед	Мужской	Медицинский	Лечебное дело	Очная	Специальн	Полный	контракт	5
	1032140493	Обучениє	Российская Фед	Мужской	Юридический	Юриспруденция	Очная	Магистрату	Полный	бюджет	2
	1032120683	Обучениє	Российская Фед	Мужской	Экономическ	Экономика	Очная	Бакалавриа	Полный	бюджет	4
8	1032102469	Обучениє	Российская Фед	Мужской	Медицинский	Лечебное дело	Очная	Специальн	Полный	контракт	4
9	1032112524	Обучениє	Российская Фед	Женский	Филологическ	Лингвистика	Очная	Бакалавриа	Полный	контракт	4
10	1032095184	Обучениє	Азербайджанск	Женский	Медицинский	Лечебное дело	Очная	Специальн	Полный	контракт	5
11	1032147275	Обучениє	Российская Фед	Женский	Медицинский	Лечебное дело	Очная	Специальн	Полный	контракт	2
12	1032146243	Обучениє	Российская Фед	Женский	Юридический	Юриспруденция	Очно-заочная	Бакалавриа	Ускоренн	о контракт	3
13	1032123363	Обучениє	Российская Фед	Женский	Медицинский	Сестринское дело	Заочная	Бакалавриа	Полный	контракт	4
14	1032134975	Обучениє	Иракская Респу	(Женский	Инженерный	Строительство	Очная	Магистрату	Полный	контракт	2
15	1032146462	Обучениє	Российская Фед	Женский	Юридический	Юриспруденция	Очно-заочная	Бакалавриа	Полный	контракт	2
16	1032141004	Обучениє	Российская Фед	Женский	Юридический	Юриспруденция	Очная	Бакалавриа	Полный	контракт	2
17	1032132972	Обучениє	Российская Фед	Женский	Юридический	Юриспруденция	Очная	Бакалавриа	Полный	бюджет	3
18	1032094420	Обучениє	Королевство Ма	Женский	Инженерный	Архитектура	Очная	Бакалавриа	Полный	контракт	5

Добавление срезов

Необходимо выделить любую ячейку или диапазон в форматированной таблице и вставить срез из ленточного меню. Возникает диалоговое окно для выбора столбцов для создания срезов.







Подведение промежуточных итогов.





Условия для создания промежуточных итогов

При анализе большого объема данных может потребоваться подведение итогов для части записей – так называемых промежуточных итогов.

Важное условие применения средств – значения организованы в виде списка или базы данных, одинаковые записи находятся в одной группе. При создании сводного отчета промежуточные итоги формируются автоматически.

Чтобы функция выдала правильный результат, проверьте диапазон на соответствие следующим условиям:

- Таблица оформлена в виде простого списка или базы данных.
- Первая строка названия столбцов.
- В столбцах содержатся однотипные значения.
- В таблице нет пустых строк или столбцов.





Создание промежуточных итогов

4	А	В	С	D	Е	F	G	Н	1	J	
1,	Округ	Область	2005	2010	2011	2012	2013	2014	2015	2016	
2	Центральный федеральный округ	Белгородская область	1 512	1532	1536	1 541	1 544	1 548	1 550	1 553	
3	Центральный федеральный округ	Брянская область	1 327	1 275	1 264	1 254	1 242	1 233	1 226	1 221	
4	Центральный федеральный округ	Владимирская область	1 486	1 441	1 432	1 422	1 413	1 406	1 397	1 390	
5	Центральный федеральный округ	Воронежская область	2 361	2 335	2 332	2 330	2 329	2 331	2 333	2 335	
6	Центральный федеральный округ	Ивановская область	1 102	1 060	1 054	1 049	1 043	1 037	1 030	1 023	
7	Центральный федеральный округ	Калужская область	1 023	1 009	1 008	1 006	1 005	1 011	1 010	1 014	
8	Центральный федеральный округ	Костромская область	700	666	662	659	656	654	651	648	
9	Центральный федеральный округ	Курская область	1 178	1 126	1 122	1 119	1 119	1 117	1 120	1 123	
10	Центральный федеральный округ	Липецкая область	1 194	1 172	1 166	1 162	1160	1 158	1 156	1 156	
11	Центральный федеральный округ	Московская область	6 784	7 106	7 199	7 048	7 13 4	7 231	7 3 1 9	7 423	
12	Центральный федеральный округ	Орловская область	822	786	781	776	77 0	765	760	755	
13	Центральный федеральный округ	Рязанская область	1 189	1 152	1 148	1 144	1 141	Вып	○世間の	OTAD	o'
14	Центральный федеральный округ	Смоленская область	1 025	983	981	975	96 8	965	03 1117	953	//1
15	Центральный федеральный округ	Тамбовская область	1 139	1 090	1 082	1 076	1 069	CONT	เนทอ	DV2	диапазона
16	Центральный федеральный округ	Тверская область	1 415	1 350	1 342	1 334	1 325	င်ပျာ့၊	NPO	DIZ	дианазопа
17	Центральный федеральный округ	Тульская область	1 615	1 550	1 545	1532	1 522	T 1514	1506	1.499	O FIONDOFO
18	Центральный федеральный округ	Ярославская область	1 313	1 271	1 271	1 272		11023	наче	HINH	о первого
19	Центральный федеральный округ	г. Москва	10 924	11 541	11 613	11 980	12 10 8	12 197	£2 330	12 381	ELIOTABLII IO
20	Северо-Западный федеральный округ	Республика Карелия	676	643	640	637	63 4	CIOI	Юца	— ₆ 0 ₁	днотипные
21	Северо-Западный федеральный округ	Республика Коми	963	899	890	880	87 2	864	857	850	
22	Северо-Западный федеральный округ	Архангельская область	1 282	1 225	1 213	1 202	1 192	данн	ные	дол	ЖНЫ
23	Северо-Западный федеральный округ	Вологодская область	1 235	1 201	1 198	1 196	1 19 3	1 191	1 188	1 184	
24	Северо-Западный федеральный округ	Калининградская область	936	942	947	955	96 3	оказ	атьс	ся р	ядом.
25	Северо-Западный федеральный округ	Ленинградская область	1 685	1 719	1 734	1 751	176 4	1 776	1 779	1 792	. •
26	Северо-Западный федеральный округ	Мурманская область	839	794	788	780	771	766	762	757	
27	Северо-Западный федеральный округ	Новгородская область	666	633	630	626	62 3	619	616	613	
28	Северо-Западный федеральный округ	Псковская область	721	671	667	662	657	651	646	642	
29	Северо-Западный федеральный округ	г. Санкт-Петербург	4 713	4 899	4 953	5 028	5 132	5 192	5 226	5 282	
30	Южный федеральный округ	Республика Адыгея	441	440	443	445	446	449	451	454	
31	Южный федеральный округ	Республика Калмыкия	294	289	287	284	282	281	279	278	
32	Южный федеральный округ	Республика Крым						1 896	1 907	1 912	
33	Южный федеральный округ	Краснодарский край	5 127	5 230	5 284	5 330	5 404	5 454	5 514	5 571	
34	Южный федеральный округ	Астраханская область	1 003	1 010	1 015	1 014	1 017	1 021	1 019	1 019	
35	Южный федеральный округ	Волгоградская область	2 640	2 607	2 595	2 583	2 569	2 557	2 546	2 535	
36	Южный федеральный округ	Ростовская область	4 332	4 275	4 260	4 254	4 246	4 242	4 236	4 231	5100
37	Южный федеральный округ	г. Севастополь						399	416	429	

Создание промежуточных итогов

Выделить любую ячейку в диапазоне данных. Перейти во вкладку меню Данные Разгруппировать ▼ Промежуточный итог Округ Структура Промежуточный итог Сумма Быстрый подсчет промежуточных и общих итогов для строк взаимосвязанных данных и 2010 вставка полученных значений ✓ 2011 ✓ 2012 на лист. Поля, по которым будут рассчитываться промежуточные итоги

Настроить свойства Промежуточных **ИТОГОВ** \mathbb{Z} Промежуточные итоги <u>П</u>ри каждом изменении в: Операция: Добавить итоги по: Округ Область ✓ Заменить текущие итоги <u>Конец страницы между группами</u> ✓ Итоги под данными Убрать все OK Отмена

Категория данных, по которой приводятся промежуточные итоги

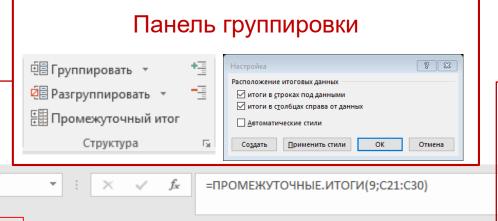
Групповая функция:

- сумма
- количество
- среднее
- минимум
- максимум
- произведение
- количество чисел
- смещенное отклонение
- несмещенное отклонение
- смещенная дисперсия
- несмещенная дисперсия





Промежуточные итоги. Результат



C31

Промежуточные итоги рассчитываются при помощи функции ПРОМЕЖУТОЧНЫЕ.ИТОГИ

	<u> </u>	_												
1	2 3	3	A	А	В	С		D	E	F	G	Н	1	J
			1	Округ Область				2010	2011	2012	2013	2014	2015	2016
Γ	+		20	Центральный федеральный округ Итог		38 16	9	38 445	38 538	38 679	38 820	38 951	39 104	39 209
	+		31	Северо-Западный федеральный округ	ьный округ Итог 13			13 626	13 660	13 717	13 801	13 844	13 854	13 899
	[·		32	Южный федеральный округ	Республика Адыгея	4/	41	440	443	445	446	449	451	454
			33	Южный федеральный округ	Республика Калмыкия	29	94	289	287	284	282	281	279	278
			34	Южный федеральный округ	Республика Крым							1 896	1 907	1 912
			35	Южный федеральный округ	Краснодарский край 5 :		27	5 230	5 284	5 330	5 404	5 454	5 514	5 571
			36	Южный федеральный округ	Астраханская область	100	03	1 010	1 015	1 014	1 017	1 021	1 019	1 019
			37	Южный федеральный округ	Волгоградская область	2 64	10	2 607	2 595	2 583	2 569	2 557	2 546	2 535
			38	Южный федеральный округ	Ростовская область	4 33	32	4 275	4 260	4 254	4 246	4 242	4 236	4 231
		ŀ	39	Южный федеральный округ	г. Севастополь							399	416	429
	_	1		Южный федеральный округ Итог		13 83	37	13 851	13 884	13 910	13 964	16 299	16 368	16 429
	[·	1	41	lри необходимости ст	РУКТУРУ стан	2 69	93	2 914	2 931	2 946	2 964	2 990	3 015	3 042
		1	42	Carana Varyascuuğ daranarılılığ ayrus	Воспублика Инпинати	41	17	415	430	442	453	464	473	481
			43	можно удалить, ост	АВИВ Кабардино-Балк арская	Pecr 86	56	860	859	859	859	861	862	865
			44	олько промежуточнь	I С ИТОГИ Р ССКАЯ	Респ 45	55	477	475	472	470	469	468	466
			45	Северо-Кавказский федеральный округ	Республика Сев ерная (Осети 70	07	712	709	706	704	706	704	703
		-	40	Северо-кавказский федеральный округ	чеченская Республика	1 15	52	1 275	1 302	1 325	1 346	1 370	1 394	1 415
			47	Северо-Кавказский федеральный округ	Ставропольский край	2 74	17	2 786	2 787	2 791	2 794	2 799	2 802	2 804
	_		48	Северо-Кавказский федеральный округ	Итог	9 03	37	9 439	9 493	9 541	9 590	9 659	9 718	9 776

Консолидация данных





Консолидация. Общие сведения

Консолидация — объединение, слияние двух или нескольких форм и/или данных.

Применяется, когда необходимо собрать и объединить данные из разных мест в одном. В качестве источников могут выступать листы книги или другие книги.

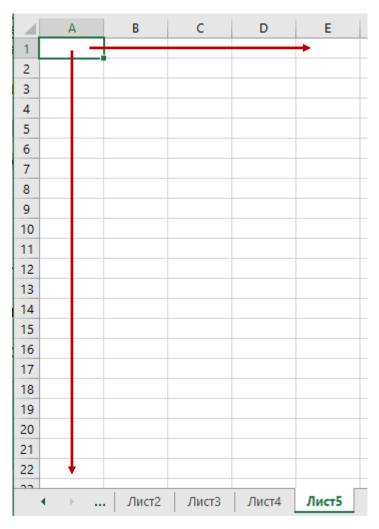
Требования к исходным данным:

- макеты всех таблиц были одинаковыми;
- названия столбцов идентичными (допускается перестановка колонок);

• нет пустых строк и столбцов.

			-	. I							6
1	Α	В	С	A	В	С			А	В	С
1	Наименование	Количество	Выручка	1 Наименован	ие Количество	Выручка		1	Наименование	Количество	Выручка
2	Товар 1	15	8 445,00	2 Товар 1	15	8 445,00		2	Товар 16	95	829 825,00
3	Товар 2	38	309 206,00	3 Товар 2	38	309 206,00		3	Товар 17	20	151 460,00
4	Товар 3	133	118 104,00	4 Товар 3	133	118 104,00		4	Товар 18	87	235 248,00
5	Товар 4	193	1 183 669,00	5 Товар 4	193	1 183 669,00		5	Товар 19	148	662 300,00
6	Товар 5	197	154 448,00	6 Товар 5	197	154 448,00	-	6	Товар 20	74	136 752,00
	Товар б	149	829 036,00	7 Товар 6	149	829 036,00		7	Товар 21	159	1 052 898,00
	Товар 7	124	992 992,00	8 Товар 7	124	992 992,00		8	Товар 22	104	866 528,00
	Товар 8	186	1 750 818,00	9 Товар 8	186	1 750 818,00		9	Товар 23	60	149 460,00
	Товар 9	153	1 228 590,00	10 Товар 21	159	1 052 898,00		10	Товар 24	191	172 855,00
	Товар 10	11	53 394,00	 11 Товар 22	104	866 528,00		11	Товар 25	151	67 799,00
	Товар 11	157	212 892,00	12 Товар 23	60	149 460,00		12	Товар 26	48	74 592,00
	Товар 12	147	1 037 232,00	13 Товар 24	191	172 855,00		13	Товар 27	147	338 835,00
	Товар 13	168	1 355 088,00	14 Товар 25	151	67 799,00	- 1	14	Товар 28	44	330 484,00
	Товар 14	190	366 510,00	15 Товар 26	48	74 592,00	- 1	15	Товар 29	98	180 712,00
	Товар 15	109	475 894,00	16 Товар 27	147	338 835,00	- 1	16	Товар 30	90	224 010,00
17		103	473 834,00	 17 Topan 28	44	-	- 1	17	Topan 31	//3	404 802 DD
.,	1	іст1 Лист2	Лист3		Лист1 Лист2	Лист3			← → Л	ист2 Лист3	Лист4

Консолидация

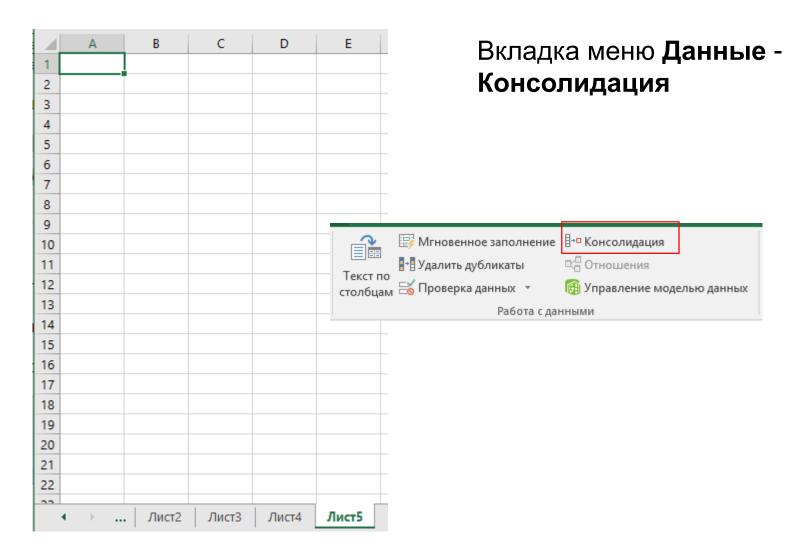


При консолидации данных происходит формирование нового диапазона данных, поэтому лучше проводить консолидацию на новом листе или новой книге.

Если консолидация производится на существующем листе, то надо убедиться в наличии свободного пространства для вставки данных.

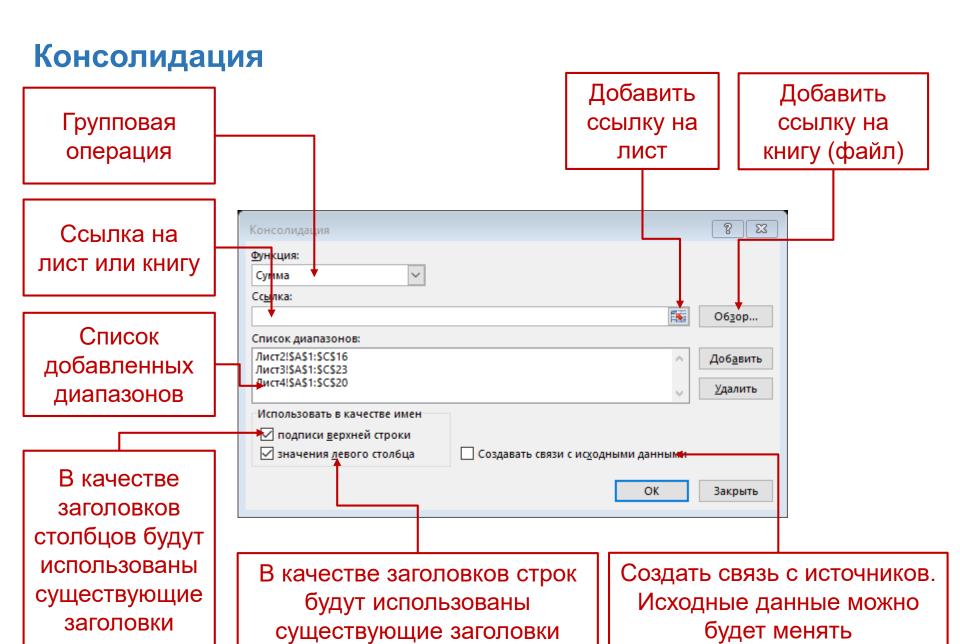


Консолидация











RUDN university Консолидация. Итог

В качестве заголовков строк выбраны существующие заголовки исходных диапазонов

4	Α	В	С		
1		Количество	Выручка		
2	Товар 1	30	16 890,00		
3	Товар 2	76	618 412,00		
4	Товар 3	266	236 208,00		
5	Товар 4	386	2 367 338,00		
6	Товар 5	394	308 896,00		
7	Товар 6	298	1 658 072,00		
8	Товар 7	248	1 985 984,00		
9	Товар 8	372	3 501 636,00		
10	Товар 9	153	1 228 590,00		
11	Товар 10	11	53 394,00		
12	Товар 11	157	212 892,00		
13	Товар 12	147	1 037 232,00		
14	Товар 13	168	1 355 088,00		
15	Товар 14	190	366 510,00		
16	Товар 15	109	475 894,00		
17	Товар 16	95	829 825,00		
18	Товар 17	20	151 460,00		
19	Товар 18	87	235 248,00		
20	Товар 19	148	662 300,00		
21	Товар 20	74	136 752,00		
22	Товар 21	318	2 105 796,00		
23	Товар 22	208	1 733 056,00		
24	Товар 23	120	298 920,00		
25	Товар 24	382	345 710,00		

В качестве заголовков категорий выбраны существующие заголовки исходных диапазонов

Групповая операция проводится среди всего набора уникальных значений строк

Произведена выбранная групповая операция (суммирование)





302

96

135 598,00

149 184,00

26 Товар 25

27 Товар 26

Ограничение

- Консолидация данных не работает с большим количеством строк.
- При превышении пограничного значения Консолидация данных не дает никакого результата.
- Ограничение находится в районе 70000 строк.





Регуля́рные выраже́ния (англ. regular expressions) — формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов (символов-джокеров, англ. wildcard characters). Для поиска используется строка-образец (англ. pattern, по-русски её часто называют «шаблоном», «маской»), состоящая из символов и метасимволов и задающая правило поиска. Для манипуляций с текстом дополнительно задаётся строка замены, которая также может содержать в себе специальные символы.

Регулярные выражения предоставляют мощный, гибкий и эффективный способ обработки текста. Комплексная нотация сопоставления шаблонов регулярных выражений позволяет быстро анализировать большие объемы текста в следующих целях:

- поиск определенных шаблонов символов;
- проверка текста на соответствие предопределенному шаблону (например, адресу электронной почты);
- извлечение, изменение, замена или удаление текстовых подстрок;
- добавление извлеченных строк в коллекцию для создания отчета.

Для многих приложений, которые работают со строками или анализируют большие блоки текста, регулярные выражения — незаменимый инструмент.

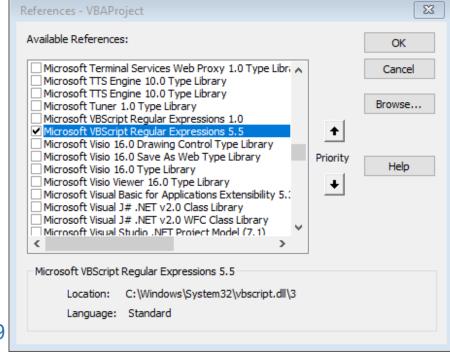




Microsoft Excel не имеет поддержки регулярных выражений как инструмента, доступного из меню, но есть реализация в VBA.

Откройте редактор Visual Basic с вкладки Разработчик (Developer) или сочетанием клавиш Alt+F11. Затем вставьте новый модуль через меню Insert - Module и скопируйте туда текст вот такой макрофункции:

```
Public Function RegExpExtract(Text As String, Pattern As String, Optional Item As Integer = 1) As String
    On Error GoTo ErrHandl
   Set regex = CreateObject("VBScript.RegExp")
   regex.Pattern = Pattern
   regex.Global = True
   If regex.Test(Text) Then
       Set matches = regex.Execute(Text)
        RegExpExtract = matches.Item(Item - 1)
        Exit Function
    End If
ErrHandl:
    RegExpExtract = CVErr(xlErrValue)
End Function
```





```
Доступна функция регулярных выражений =RegExpExtract( Txt ; Pattern ; Item ) где
```

- Тхt ячейка с текстом, который мы проверяем и из которого хотим извлечь нужную нам подстроку
- Pattern маска (шаблон) для поиска подстроки
- Item порядковый номер подстроки, которую надо извлечь, если их несколько (если не указан, то выводится первое вхождение)





Паттерн	Описание
	Самое простое - это точка. Она обозначает любой символ в шаблоне на указанной позиции.
\s	Любой символ, выглядящий как пробел (пробел, табуляция или перенос строки).
\S	Анти-вариант предыдущего шаблона, т.е. любой НЕпробельный символ.
\d	Любая цифра
\D	Анти-вариант предыдущего, т.е. любая НЕ цифра
\w	Любой символ латиницы (А-Z), цифра или знак подчеркивания
\W	Анти-вариант предыдущего, т.е. не латиница, не цифра и не подчеркивание.
[символы]	В квадратных скобках можно указать один или несколько символов, разрешенных на указанной позиции в тексте. Например ст[уо]л будет соответствовать любому из слов: стол или стул. Также можно не перечислять символы, а задать их диапазоном через дефис, т.е. вместо [ABDCDEF] написать [A-F]. или вместо [4567] ввести [4-7]. Например, для обозначения всех символов кириллицы можно использовать шаблон [а-яА-ЯёЁ].
[^символы]	Если после открывающей квадратной скобки добавить символ "крышки" ^, то набор приобретет обратный смысл - на указанной позиции в тексте будут разрешены все символы, кроме перечисленных. Так, шаблон [^ЖМ]уть найдет Путь или Суть или Забудь, но не Жуть или Муть, например.
1	Логический оператор ИЛИ (OR) для проверки по любому из указанных критериев. Например (счет счёт invoice) будет искать в тексте любое из указанных слов. Обычно набор вариантов заключается в скобки.
٨	Начало строки
\$	Конец строки
\b	Край слова

Если мы ищем определенное количество символов, например, шестизначный почтовый индекс или все трехбуквенные коды товаров, то на помощь нам приходят квантификаторы или кванторы - специальные выражения, задающие количество искомых знаков. Квантификаторы применяются к тому символу, что стоит перед ним:

Квантор	Описание
?	Ноль или одно вхождение. Например .? будет означать один любой символ или его отсутствие.
+	Одно или более вхождений. Например \d+ означает любое количество цифр (т.е. любое число от 0 до бесконечности).
*	Ноль или более вхождений, т.е. любое количество. Так \s* означает любое количество пробелов или их отсутствие.
{число} или {число1,число2}	Если нужно задать строго определенное количество вхождений, то оно задается в фигурных скобках. Например \d{6} означает строго шесть цифр, а шаблон \s{2,5} - от двух до пяти пробелов





Извлечение числа из текста

Нужно извлечь из буквенно-цифровой каши первое число, например мощность источников бесперебойного питания из прайс-листа:

=RegExpExtract(A2;\$B\$1)

4	A	В
1	НФА	\d+
2	Источник бесперебойного питания Smart- UPS 700NET сч. SD-042070	700
3	Источник бесперебойного питания APC 800 VA сч.1012	800
4	Устройство бесперебойного питания UPS Back-UPS CS 500 сч.7Q0015	500
5	Источник бесперебойного питания АРС	#3HAЧ!
6	Источник бесперебойного питания APC 800 VA сч.1012	800
7	Источник бесперебойного питания SUA1500I	1500
8	Источник бесперебойного питания АРС	#3HAЧ!
9	Источник бесперебойного питания АРС	#3HAЧ!
10	Источник бесперебойного питания UPS 3000VA Smart On-Line APC сч.191	3000
11	Источник бесперебойного электропитания Back-UPS CS 500VA 230V	500
12	Источник бесперебойного питания Back-UPS CS 500VA 530V сч.61730714/1	500
13	Источник бесперебойного питания UPS 3000VA Smart On-Line APC сч.238	3000
14	Источник бесперебойного питания Back-UPS CS 500VA 530V сч.61730714/1	500





Почтовый индекс

Поиск ровно шести цифр подряд.

Используется спецсимвол \d для цифры и квантор {6} для количества знаков. Однако, возможна ситуация, когда левее индекса в строке стоит еще один большой набор цифр подряд (номер телефона, ИНН, банковский счет и т.д.) Тогда регулярное выражение выдернет из нее первых 6 цифр, т.е. сработает некорректно:

1	A	В
16		\d{6}
17	РОССИЯ, 119071, Москва г, Ленинский пр-кт, дом № 20	119071
18	117198, Москва г, Миклухо-Маклая ул, дом № 17, корпус 7	117198
19	РОССИЯ, 125475, Москва г, Зеленоградская ул, дом № 27А	125475
20	тел.89031234567890 РОССИЯ, 127562, Москва г, Каргопольская ул, дом № 6	890312
21	РОССИЯ, 127322, Москва г, Милашенкова ул, дом № 18, квартира 195	127322

Необходимо добавить в регулярное выражение по краям модификатор \b означающий конец слова. Это даст понять Excel, что нужный фрагмент (индекс) должен быть отдельным словом, а не частью другого фрагмента (номера телефона):

4	A	В
23		\b\d{6}\b
24	РОССИЯ, 119071, Москва г, Ленинский пр-кт, дом № 20	119071
25	117198, Москва г, Миклухо-Маклая ул, дом № 17, корпус 7	117198
26	РОССИЯ, 125475, Москва г, Зеленоградская ул, дом № 27А	125475
27	тел.89031234567890 РОССИЯ, 127562, Москва г, Каргопольская ул, дом № 6	127562
28	РОССИЯ, 127322, Москва г, Милашенкова ул, дом № 18, квартира 195	127322
	RUDN	



Телефон

Проблема с нахождением телефонного номера среди текста состоит в том, что существует очень много вариантов записи номеров - с дефисами и без, через пробелы, с кодом региона в скобках или без и т.д. Поэтому проще сначала вычистить из исходного текста все эти символы с помощью нескольких вложенных друг в друга функций ПОДСТАВИТЬ (SUBSTITUTE), чтобы он склеился в единое целое, а потом регулярным выражением \d{11} вытаскивать 11 цифр подряд:

	A	В	С
30			\d{11}
31	Иванов, +7(495) 955-08-78, доб. 3 878	Иванов,+74959550878,доб.3 878	74959550878
32	Петров, 8(926) 1880303	Петров,89261880303	89261880303
33	Сидоров А.А. +7(495) 433-80-29, доб. 1 164	СидоровА.А.+74954338029,доб.1 164	74954338029
34	Захаров Алексей +7(495) 433-06-00, доб.	ЗахаровАлексей+74954330600,доб.	74954330600

Функция склеивания:

=ПОДСТАВИТЬ(ПОДСТАВИТЬ(ПОДСТАВИТЬ(ПОДСТАВИТЬ(А31;" ";"");"(";"");");"-";"")



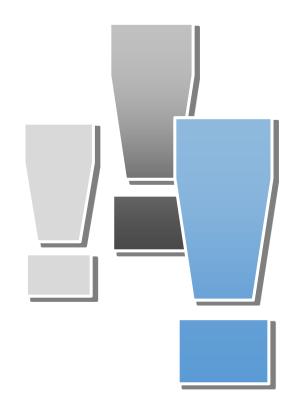


инн	\d{12} \d{10}
Артикулы товаров, SAP-коды, SKU	[A-Z]{3}-\d{3}
Денежные суммы, которые указываются через дефис	\d+-\d{2}
Автомобильные номера	[ABEЛMHOPCTУX]\d{3} [ABEЛMHOPCTУX]{2}\d{2,3}
	([0-1]\d 2[0-3]):[0-5]\d
Время	[0-5]\d - любое число в интервале 00-59 [0-1]\d - любое число в интервале 00-19 2[0-3] - любое число в интервале 20-23
Город из адреса	Γ*?





Спасибо за внимание!



Шевцов Василий Викторович

shevtsov_vv@rudn.university +7(903)144-53-57



