



Основные форматы  
хранения и обмена  
табличной информацией

Данные.xlsx - Microsoft Excel

Разметка страницы    Формулы    Данные    Рецензирование    Вид

Фигуры    SmartArt    Снимок    Гистограмма    График    С областями    Круговая    Точечная    Линейчатая    Другие    Диаграммы    Sparkлайны    Срез    Гиперссылка    Фильтр    Ссылки

fx 27,1

В	С	Д	Е	Ф	
Регион	Территория, тыс. км <sup>2</sup>	Численность населения, тыс. чел.	Занятые в экономике, тыс. чел.	Денежные доходы (в месяц), руб.	Пот ие ме
Алгогородская область	27	1536	698	18800	
Аянская область	35	1264	561	15348	
Адимирская область	29	1432	704	14312	
Аронежская область	52	2332	1055	15871	
Вановская область	21	1054	491	13006	
Влужская область	30	1008	480	17557	
Встромская область	60	662	315	14823	
Врская область	30	1122	580	16387	
Впецкая область	24	1166	545	16811	
Всковская область	46	7199	2911	25605	
Воловская область	25	781	393	14824	
Взанская область	40	1149	502	14788	
Воленская область	50	981	496	15969	
Вмбовская область	35	1083	504	15151	
Верская область	84	1342	587	14943	
Вльская область	26	1545	771	16975	
Вославская область	36	1271	638	15509	
Москва	1	11613	6480	47319	
Вспублика Карелия	181	640	316	17543	
Вспублика Коми	417	890	461	23897	
Вхангельская область	590	1214	606	21455	
Влогодская область	145	1199	590	15638	
Влининградская область	15	947	472	16881	
Венинградская область	84	1734	748	15932	
Вурманская область	145	788	427	25304	
Вовгородская область	55	630	316	16981	
Всковская область	55	667	330	14185	
Санкт-Петербург	1	4953	2501	25995	

# Почему важно знать форматы табличных данных?

## Основа всех процессов

Табличные данные являются фундаментом для финансов, аналитики, отчетности и обмена информацией в любой сфере.

## Влияние на эффективность

Правильный выбор формата критически влияет на удобство работы, совместимость данных между системами и безопасность их хранения.

## Избегайте потери данных

Ошибки при выборе или использовании формата могут привести к необратимой потере данных или их искажению, что имеет серьезные последствия.

# CSV — простой и универсальный формат обмена

## Преимущества:

- Значения разделяются запятыми, что делает его легко читаемым и редактируемым вручную.
- Широко используется для передачи данных между различными приложениями и системами благодаря своей простоте.
- Поддерживается практически всеми программами для работы с данными.

## Ограничения:

- Отсутствует поддержка формул, форматирования, нескольких листов и сложных структур данных.
- Крайне важна правильная кодировка (например, UTF-8) и корректный выбор разделителей для русского языка, чтобы избежать проблем с отображением символов.





# XLS и XLSX — стандарты Microsoft Excel

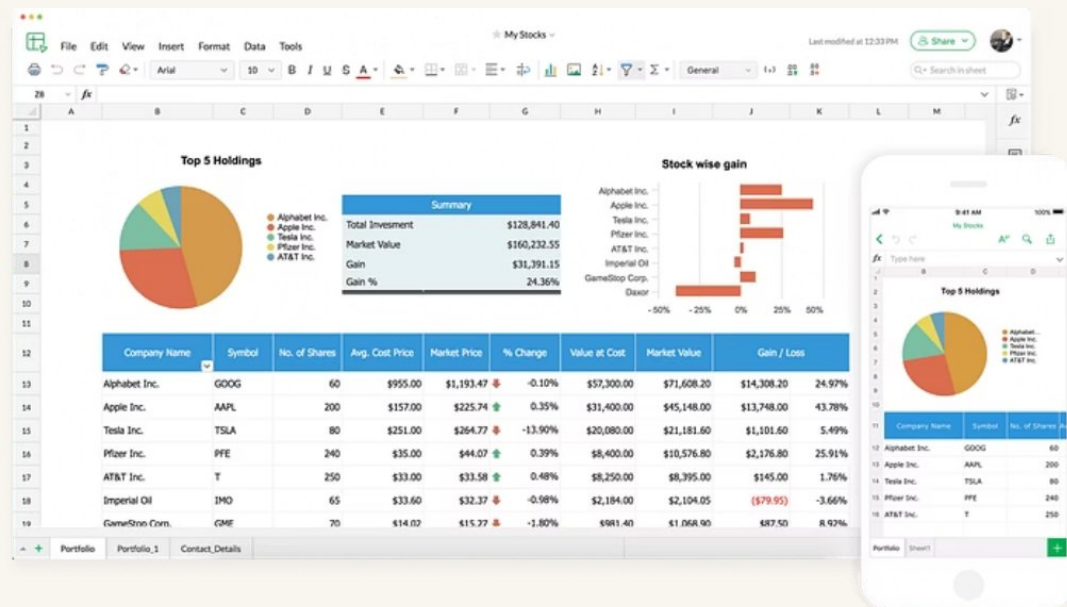
## XLS (До 2007)

Двоичный формат, поддерживающий формулы, графики, макросы и множество листов. Широко использовался для сложных отчетов и анализа данных.



## XLSX (Современный)

Основан на XML и является сжатым ZIP-архивом. Более устойчив к повреждениям, легче восстанавливается и поддерживается большинством современных приложений.





# Форматы для структурированных и связанных данных: JSON, XML, RDF

---

## JSON (JavaScript Object Notation)

Легковесный, удобочитаемый формат, идеальный для обмена данными между веб-серверами и клиентскими приложениями. Используется в API и конфигурационных файлах.

---

## XML (Extensible Markup Language)

Расширяемый язык разметки, который сохраняет как структуру данных, так и метаданные. Широко применяется в сложных интеграциях, документах и веб-сервисах.

---

## RDF (Resource Description Framework)

Формат для описания и объединения связанных данных из различных источников. Используется для построения "семантической паутины" и баз знаний.

# PDF и текстовые форматы для табличной информации



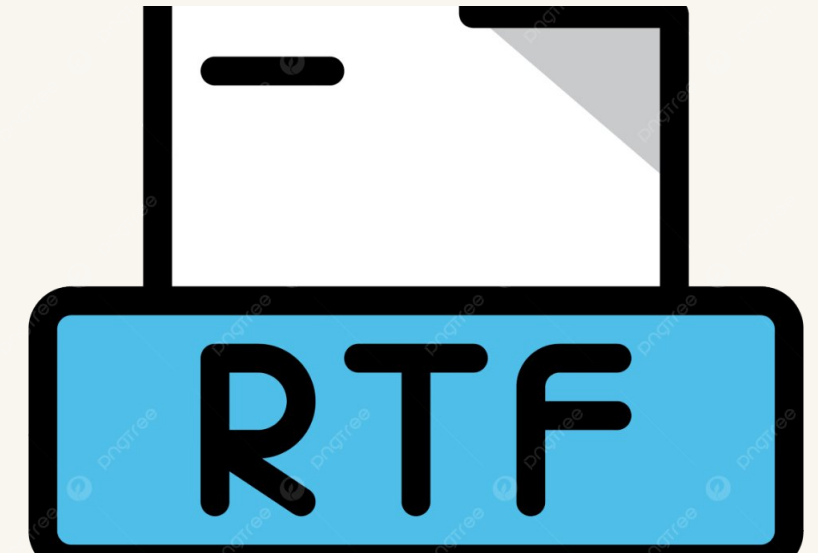
PDF (Portable Document Format)

Идеально подходит для презентаций, архивирования и распространения, поскольку точно сохраняет внешний вид таблиц. Однако он не предназначен для редактирования данных.



TXT (Plain Text)

Самый простой текстовый формат без какого-либо форматирования. Подходит для базового обмена данными, где не важна структура или внешний вид.



RTF (Rich Text Format)

Поддерживает базовое форматирование текста (жирный, курсив, шрифт), но не подходит для сложных табличных структур или формул.

# Современные форматы для больших данных и аналитики

## Parquet

Колоночный формат, оптимизированный для эффективного хранения и обработки больших объемов табличных данных в распределенных системах.



## Arrow

Формат для быстрой передачи табличных данных в памяти между различными аналитическими системами и языками программирования.

## Protobuf

Компактный, бинарный формат с поддержкой строго типизированных схем данных, используемый для высокопроизводительной сериализации и обмена в микросервисах.



# Заключение

1

## Избегайте ошибок

Знание особенностей каждого формата помогает предотвратить потерю данных и повысить общую продуктивность работы.

2

## Подбирайте под задачу

Всегда используйте формат, который наилучшим образом соответствует конкретной задаче и инструментам, которые вы используете.

3

## Контролируйте качество

Обращайте внимание на кодировку и структуру данных для обеспечения надежного и безопасного обмена информацией.

4

## Внедряйте инновации

Используйте современные форматы для создания масштабируемых аналитических решений и работы с большими данными.