

Отчет о проверке функциональности табличного редактора Р7-Офис в контексте книги "Data Smart"

Анализ возможностей для реализации методов очистки данных, классификации, кластеризации, регрессии, ансамблевых моделей, прогнозирования, оптимизации, обнаружения выбросов и перехода к языку R

Дата: 21 марта 2025 года

Автор: Станислав Leneve

Отчет о проверке функциональности в Р7-Офисе

Общие замечания

Р7-Офис предоставляет табличный редактор, схожий по базовым функциям с Microsoft Excel, однако он ограничен в поддержке современных инструментов, таких как динамические массивы (XLOOKUP, SEQUENCE, LET, LAMBDA), автоматизация (Power Query, VBA) и оптимизация (Solver). Тестирование проводилось на версии Р7-Офиса, доступной на 21 марта 2025 года. Отчет охватывает функциональность, необходимую для задач из глав 2–12 книги "Data Smart", включая очистку данных, наивный Байес, кластеризацию, сетевые графы, регрессию, ансамблевые модели, прогнозирование, оптимизационное моделирование, обнаружение выбросов, переход к языку R и практические рекомендации.

Проверенные разделы и результаты

1. Accessing Quick Descriptive Statistics (Доступ к быстрой описательной статистике)
 - Описание: Выделение диапазона для отображения суммы, среднего, количества в строке состояния.
 - Проверка в Р7-Офисе: Поддерживаются базовые статистики (сумма, среднее, количество), но настройка дополнительных (минимум, максимум) отсутствует.
 - Результат: Частично работает.
 - Ошибка: Ограниченный функционал строки состояния.
2. Excel Tables (Таблицы Excel)
 - Описание: Создание таблиц с фильтрацией, сортировкой и структурированными ссылками.
 - Проверка в Р7-Офисе: Таблицы создаются, фильтрация и сортировка работают, но структурированные ссылки (например, Table[Column]) не поддерживаются.
 - Результат: Частично работает.
 - Ошибка: "Неверная ссылка" при использовании структурированных ссылок.
3. Lookup Formulas (Формулы поиска)
 - VLOOKUP: Работает с диапазонами, но не с именами таблиц.
 - INDEX/MATCH: Работает с диапазонами, но не с таблицами.
 - XLOOKUP: Отсутствует.
 - Результат: Частично работает.
 - Ошибка: Отсутствие XLOOKUP; ограничения с таблицами.
4. PivotTables (Сводные таблицы)
 - Описание: Агрегирование данных (сумма, подсчет).
 - Проверка в Р7-Офисе: Полностью поддерживается.
 - Результат: Работает.
 - Ошибка: Отсутствует.
5. Using Array Formulas (Массивные формулы)
 - Описание: Использование SUMPRODUCT, TRANSPOSE.
 - Проверка в Р7-Офисе: Работает в старом стиле (Ctrl+Shift+Enter), динамические массивы не поддерживаются.
 - Результат: Частично работает.
 - Ошибка: "#VALUE!" при попытке использовать динамические массивы.

6. Solving Stuff with Solver (Решение задач с помощью Solver)

- Описание: Оптимизация параметров.
- Проверка в Р7-Офисе: Solver отсутствует.
- Результат: Не работает.
- Ошибка: "Надстройка Solver не найдена".

7. Power Query (Очистка данных, глава 2)

- Описание: Автоматизированная очистка данных (EmailContacts.xlsx).
- Проверка в Р7-Офисе: Power Query отсутствует, ручные операции возможны.
- Результат: Не работает как Power Query.
- Ошибка: Отсутствует аналог.

8. Naïve Bayes Classification (Наивный Байес, глава 3)

- Описание: Классификация твитов (Mandrill.xlsx).
- Проверка в Р7-Офисе: Требует Power Query и XLOOKUP, частично реализуемо вручную.
- Результат: Частично работает.
- Ошибка: Отсутствие Power Query и XLOOKUP.

9. Cluster Analysis with K-Means (К-средние, глава 4)

- Описание: Сегментация покупок вина (WineKMC.xlsx).
- Проверка в Р7-Офисе: Частично реализуемо, но без Solver оптимизация невозможна.
- Результат: Частично работает.
- Ошибка: Отсутствие Solver.

10. Cluster Analysis Part II: Network Graphs (Сетевые графы, глава 5)

- Описание: Обнаружение сообществ (WineNetwork.xlsx).
- Проверка в Р7-Офисе: Матрица смежности возможна, но без Solver и визуализации полная реализация невозможна.
- Результат: Частично работает.
- Ошибка: Отсутствие Solver и GiGraph.

11. Regression (Регрессия, глава 6)

- Описание: Линейная и логистическая регрессия (RetailMart.xlsx).
- Проверка в Р7-Офисе: Фиктивные переменные и базовые вычисления возможны, но без Solver, MMULT, MINVERSE и статистических функций полная реализация невозможна.
- Результат: Частично работает.
- Ошибка: Отсутствие Solver, MMULT, MINVERSE, DEVSQ, F.DIST, T.DIST, XLOOKUP, LET, SEQUENCE.

12. Ensemble Models: A Whole Lot of Bad Pizza (Ансамблевые модели, глава 7)

- Описание: Создание ансамблевых моделей (bagging и boosting) на основе данных RetailMart (Ensemble.xlsx).
 - Проверка в P7-Офисе:
 - Bagging: Ручная выборка через RAND() возможна, но без Power Query, CHOOSECOLS, CHOOSEROWS, SEQUENCE автоматизация невозможна. COUNTIFS, SUM, INDEX работают для базовых вычислений.
 - Boosting: Инициализация весов вручную, LN, EXP, SUM поддерживаются, но без VBA многоократное обучение невозможно.
 - Результат: Частично работает.
 - Ошибка: Отсутствие Power Query, CHOOSECOLS, CHOOSEROWS, SEQUENCE, VBA, XLOOKUP, LET.

13. Forecasting: Breathe Easy: You Can't Win (Прогнозирование, глава 8)

- Описание: Экспоненциальное сглаживание (SES, Holt's, Holt-Winters) для прогнозирования спроса (SwordForecasting.xlsx).
 - Проверка в P7-Офисе:
 - SES: IF, INDEX, SUM поддерживаются, но без Solver и IFS оптимизация alpha вручную.
 - Holt's: INDEX и SUM для уровня и тренда, но без Solver и LINEST инициализация и оптимизация ограничены.
 - Holt-Winters: Ручной расчет MA через AVERAGE, INDEX и MOD для сезонности, но без FILTER, LET, XLOOKUP сложен.
 - Autocorrelations: SUMPRODUCT работает, но без SEQUENCE лаги вручную.
 - Монте-Карло: RAND и NORM.INV поддерживаются, но без VBA и PERCENTILE ограничено.
 - Результат: Частично работает.
 - Ошибка: Отсутствие Solver, Power Query, VBA, IFS, LET, FILTER, SEQUENCE, XLOOKUP, LINEST, T.DIST.2T, PERCENTILE.

14. Optimization Modeling: Because That "Fresh-Squeezed" Orange Juice Ain't Gonna Blend Itself (Оптимизационное моделирование, глава 9)

- Описание: Оптимизация смеси апельсинового сока (OrangeJuiceBlending.xlsx) с использованием линейного программирования (Simplex LP), целочисленного программирования (Branch and Bound), нелинейной оптимизации (Evolutionary Solver), методов "Big M" и минимизации максимума (Minimax).

- Проверка в P7-Офисе:
 - Linear Programming (Simplex LP): SUMPRODUCT для затрат и ограничения (например, =СУММПРОИЗВ(С6:С16,Л6:Л16)) поддерживаются, но без Solver оптимизация невозможна.
 - Integer Programming (Branch and Bound): Целочисленные ограничения вручную, без "bin" или "int" в Solver.
 - Nonlinear Optimization (Evolutionary Solver): IF для бонусов работает, но без Evolutionary Solver нелинейные задачи не решаемы.
 - Big M Constraint: Бинарные индикаторы и формулы (например, 672*C34 ≥ С6) задаются вручную, но без Solver не оптимизируются.
 - Minimax Formulation: MAX и ограничения (G27:G30 ≤ E2) вручную, без Solver не минимизируется.
 - Robust Optimization: NORM.INV для сценариев работает, но без SEQUENCE и Solver ограничения по сценариям не автоматизируются.
 - Результат: Частично работает.
 - Ошибка: Отсутствие Solver, SEQUENCE, LET, XLOOKUP, VBA; невозможность задать "bin" или "int" ограничения.
15. Outlier Detection: Just Because They're Odd Doesn't Mean They're Unimportant
(Обнаружение выбросов, глава 10)
- Описание: Обнаружение выбросов с использованием нормального распределения, Tukey's Fences, kNN-графов, k-distance, LOF (SupportCenter.xlsx).
 - Проверка в P7-Офисе:
 - NORM.DIST: Работает (=1-HOPM.РАСП(349,266,9,ИСТИНА)).
 - Tukey's Fences: QUARTILE поддерживается, PERCENTILE.INC отсутствует.
 - Standardizing Data: AVERAGE и STDEV.S работают, STANDARDIZE отсутствует.
 - kNN, k-Distance, LOF: SUMPRODUCT, RANK.EQ, COUNTIF поддерживаются, но без SEQUENCE и LET трудоемко.
 - Результат: Частично работает.
 - Ошибка: Отсутствие STANDARDIZE, LET, CHOOSECOLS, SEQUENCE, XLOOKUP, VBA, PERCENTILE.INC.
16. Moving on From Spreadsheets (Переход от таблиц, глава 11)
- Описание: Глава описывает переход от Excel к языку R для выполнения анализа данных, включая чтение данных (read.csv), очистку (tidyverse, dplyr), кластеризацию (skmeans), моделирование (glm, randomForest), прогнозирование (forecast), обнаружение выбросов (Rlof), визуализацию (ggplot). Используются пакеты tidyverse, dplyr, skmeans, randomForest, ROCR, forecast, Rlof.

- Проверка в P7-Офисе:
 - Чтение данных: Импорт CSV поддерживается вручную, но без read.csv требуется внешний инструмент (например, RStudio).
 - Очистка данных (dplyr): Операции select, replace, mutate вручную через фильтры и IF, но без %>% и автоматизации сложны.
 - Кластеризация (skmeans): SUMPRODUCT для косинусного расстояния и базовых вычислений, но без Solver и skmeans оптимизация невозможна.
 - Моделирование (glm, randomForest): Линейная регрессия через SLOPE и INTERCEPT возможна, но без Solver, MMULT, MINVERSE и randomForest полная реализация невозможна.
 - Прогнозирование (forecast): SES и Holt's вручную через IF и SUM, но без forecast и автоматизации Holt-Winters сложны.
 - Обнаружение выбросов (Rlof): QUARTILE и NORM.DIST для Tukey's Fences, но LOF требует Rlof и автоматизации (отсутствуют).
 - Визуализация (ggplot): Базовые графики (линии, столбцы) поддерживаются, но без ggplot сложные визуализации невозможны.
- Результат: Частично работает.
- Ошибка: Отсутствие R-интеграции, Power Query, Solver, VBA, MMULT, MINVERSE, SEQUENCE, LET, XLOOKUP, FILTER, LINEST, T.DIST.2T, PERCENTILE.INC, STANDARDIZE; невозможность использовать пакеты R (tidyverse, dplyr, skmeans, randomForest, ROCR, forecast, Rlof).

17. Conclusion (Заключение, глава 12)

- Описание: Глава обсуждает практические аспекты аналитики: понимание проблемы, коммуникация, избегание чрезмерной сложности, выбор инструментов, фокус на целях бизнеса. Не содержит конкретных вычислений, но подчеркивает переход от Excel к более мощным инструментам (например, R).
- Проверка в P7-Офисе:
 - Понимание проблемы: Ручная работа с данными (фильтры, сводные таблицы) возможна, но без Power Query и Solver ограничена.
 - Коммуникация: Базовые графики и таблицы для презентации поддерживаются.
 - Сложность и инструменты: Без R-интеграции и Solver сложные модели не реализуемы.
 - Цели бизнеса: Подсчет затрат и ограничений вручную через SUMPRODUCT и IF, но без оптимизации (Solver) и автоматизации (Power Query, VBA) эффективность низка.
- Результат: Частично работает.
- Ошибка: Отсутствие Power Query, Solver, VBA, R-интеграции, современных функций для автоматизации и оптимизации.

Итог

- Базовые возможности: P7-Офис поддерживает VLOOKUP, INDEX/MATCH (с диапазонами), сводные таблицы, SUMPRODUCT (старый стиль), COUNTIFS, MIN, MATCH, SUM, LN, EXP, TRANSPOSE (ограниченно), RAND, NORM.INV, SLOPE, INTERCEPT, NORM.DIST, QUARTILE, STDEV.S, RANK.EQ, что достаточно для простых вычислений и частичной реализации задач из глав 2–12.
- Ограничения: Отсутствие Power Query, Solver, макросов VBA, R-интеграции, современных функций (XLOOKUP, CHOOSECOLS, CHOSEROWS, SEQUENCE, LET, LAMBDA, IFS, FILTER, LINEST, T.DIST.2T, PERCENTILE.INC, MMULT, MINVERSE, DEVSQ, F.DIST, STANDARDIZE) и динамических массивов делает невозможной полную автоматизацию, оптимизацию и переход к языку R без значительных ручных усилий или внешних инструментов.

В контексте глав:

- Глава 2 (Power Query): Не поддерживается.
- Глава 3 (Naïve Bayes): Частично возможно вручную.
- Глава 4 (K-Means): Частично, без оптимизации.
- Глава 5 (Network Graphs): Частично, без визуализации и оптимизации.
- Глава 6 (Regression): Частично, без Solver и статистических тестов.
- Глава 7 (Ensemble Models): Частично (один stump), без автоматизации.
- Глава 8 (Forecasting): Частично (SES и Holt's вручную), Holt-Winters, автокорреляции и Монте-Карло требуют усилий.
- Глава 9 (Optimization Modeling): Частично (линейные формулы), но без Solver задачи линейного, целочисленного и нелинейного программирования не решаемы полностью.
- Глава 10 (Outlier Detection): NORM.DIST и Tukey's Fences работают, kNN-графы, k-distance и LOF частично реализуемы, но трудоемки.
- Глава 11 (Moving on From Spreadsheets): Частично (базовые вычисления), но без R-интеграции и пакетов полная реализация невозможна.
- Глава 12 (Conclusion): Поддержка базовых функций для коммуникации, но без инструментов для сложных моделей и автоматизации.

Заключение

P7-Офис демонстрирует базовую функциональность для работы с таблицами, включая простые формулы (SUMPRODUCT, IF, VLOOKUP, INDEX/MATCH), сводные таблицы и базовые графики, что позволяет частично реализовать задачи из глав 2–12 книги "Data Smart". Однако отсутствие ключевых инструментов, таких как Power Query, Solver, VBA, R-интеграция и современные функции (XLOOKUP, SEQUENCE, LET, LAMBDA и др.), существенно ограничивает возможности автоматизации, оптимизации и перехода к более мощным аналитическим инструментам, таким как язык R, как описано в главе 11.

Для полного соответствия требованиям книги, включая очистку данных (глава 2), оптимизацию (глава 9), сложные модели (главы 6–8), обнаружение выбросов (глава 10) и переход к R (глава 11), P7-Офис требует интеграции с внешними инструментами (например, RStudio, Gephi) или значительных ручных усилий, что противоречит принципам эффективности, изложенным в главе 12. Таким образом, для задач аналитики данных, описанных в "Data Smart", рекомендуется использовать Microsoft Excel с Power Query, Solver и VBA или комбинировать его с R и Gephi, как предложено в книге.

Если P7-Офис планирует развиваться в направлении аналитики данных, необходимо внедрить:

1. Аналог Solver для линейной, целочисленной и нелинейной оптимизации.
2. Power Query для автоматизации очистки данных.
3. Поддержку VBA или альтернативного языка скриптов для автоматизации.
4. Интеграцию с R или Python для расширенного анализа.
5. Современные функции (XLOOKUP, SEQUENCE, LET, LAMBDA) и динамические массивы.

Без этих улучшений P7-Офис остается ограниченным инструментом для базовых задач и не может полностью заменить Excel или R в контексте "Data Smart".