МЕТОДИ ЗА АВТОМАТИЗАЦИЯ НА БИЗНЕС АНАЛИЗИ ЧРЕЗ ИНТЕГРИРАНЕ НА НЕЕДНОРОДНИ ДАННИ

Methods for Automating Business Analysis by Integrating Heterogeneous Data

Теодор Тодоров[1],
e-mail: ttodorov131@unwe.bg [1]

Абстракт
В съвременната бизнес среда няма компания, на която да не се налага да обработва и анализира данни. Независимо дали става дума за структура в частния или обществения сектор, с двама или десетки служители, обемът информация нараства, но с него расте и потенциалът бизнес единиците да извличат качествено знание. Най-разпространеното „решение“ за съхранение и обработка на първична информация дори и през 2025 остава Microsoft Excel. Наличието на напреднали инструменти с изкуствен интелект невинаги решава първичния проблем – как данните да бъдат интегрирани и представени, така че да могат да се използват занапред за задълбочен анализ и проследяване на тенденции. Настоящият доклад следва да предложи методология за автоматизация обработката на информация, която обединява процесите по събиране, уеднаквяване и трансформация на разнородни данни, извлечени от Excel източници. Ще бъдат предложени добри практики при първичното съхраняване на данни, ще бъдат разгледани съвременни low-code/no-code решения и ще се предложат идеи за приложение в малки и средни предприятия с ограничен ресурс.

Abstract
In the current business environment, there is no company that does not have to process and analyze data. Regardless of whether it is a structure in the private or public sector, with two or dozens of employees, the volume of information is growing, but with it the potential for business units to extract qualitative knowledge is also growing. The most common “solution” for storing and processing primary information even in 2025 remains Microsoft Excel. The availability of advanced artificial intelligence tools does not always solve the primary problem – how to integrate and present data so that it can be used in the future for in-depth analysis and trend tracking. This report should propose a methodology for automating information processing, which unites the processes of collecting, unifying and transforming heterogeneous data extracted from Excel sources. Good practices in primary data storage will be proposed, modern low-code/no-code solutions will be considered, and ideas for application in small and medium-sized enterprises with limited resources will be proposed.

Ключови думи: обработка на данни, изкуствен интелект, извличане на информация
JEL: C810

Увод

Годината е 1985 г. и компанията Microsoft пуска в продажба първата версия на системата за управление на електронни таблици MS Excel за компютри Apple Macintosh. Макар Lotus 123 – доминиращата платформа за изчисления по това време, да предлага възможности за изчисления, създаване на графики и макроси, Excel успява да се наложи с простотата на своето използване. На потребителите не се налага да имат сериозни познания по програмиране, за да въвеждат, пресмятат и визуализират информация. Excel надгражда идеята с по-интуитивен интерфейс и интеграция с графичен интерфейс – нещо ново за средата на 80-те. В този момент, едва ли някой предполага, че MS Excel ще се превърне в най-универсалния инструмент за управление на данни в света. През 1987 г. излиза версия и за новата MS Windows, а през 1993 г. Excel надминава по

[1] Докторант, катедра „Информационни технологии и комуникации“, факултет „Приложна информатика и статистика“, УНСС, гр. София, ORCID 0009-0002-5119-8873

пазарен дял Lotus 1-2-3, превръщайки се в синоним на електронна таблица (Walkenbach, 2013). С годините Excel се развива от инструмент за прости изчисления в универсална работна среда за данни. Финансисти, счетоводители, маркетолози и инженери започват да го използват не само за събиране и изваждане, но и за съхранение на големи обеми от информация, проследяване на запаси, проекти и бюджети. Така възниква парадокс - продукт, създаден като изчислителен инструмент, постепенно се превръща в *де факто* база от данни. Както посочва Реймънд Панко (2013), „повечето организации разчитат повече на електронни таблици, отколкото на официалните си бази данни“ – не защото това е най-доброто решение, а защото Excel е познат, достъпен и не изисква специална ИТ инфраструктура.

С разпространението на Microsoft Office през 90-те Excel става стандартен компонент на всеки офис компютър. В много компании той се превръща в първото и последното звено на анализа: данните се въвеждат ръчно, трансформират се с формули и макроси, а резултатите често се използват за управленски решения. Тази практика, макар и удобна, поставя сериозни ограничения: липса на контрол върху версиите, несъгласуваност на данните и трудности при обединяване на множество файлове. Kruck и Maher (2012) описват това като „илюзията на базата данни“ – Excel дава усещане за структура, но не осигурява нито устойчивост, нито връзки между данните, каквито предлагат релационните системи.
Днес, въпреки наличието на мощни BI и ERP решения, Excel остава повсеместен. Данните на Gartner (2024) показват, че над 80% от малките предприятия в световен мащаб продължават да използват електронни таблици като основен инструмент за съхранение и анализ на информация. Това е резултат от културна и икономическа реалност – Excel е познат, гъвкав и *„достатъчно добър“* за ежедневните нужди. Именно тук възниква и ключовият въпрос на настоящия доклад: как можем да запазим познатата среда, но да автоматизираме обработката, уеднаквяването и интегрирането на нейните нееднородни данни, така че тя да служи като основа за съвременен бизнес анализ.

Excel срещу базите данни

На теория, Excel и базите данни решават един и същ проблем — съхраняване и обработка на данни. На практика, обаче, те принадлежат към два напълно различни свята. Електронната таблица е създадена, за да бъде *гъвкав инструмент за изчисления*, докато базата от данни е *структурирана система за съхранение и взаимовръзки (Gartner)*. Разликата може да се назове и още по-просто - в Excel редовете „живеят“ самостоятелно, докато в една база от данни те имат взаимоотношения и зависимости. Excel не разполага с механизми за **референтна цялост** – няма първични и външни ключове, които да предотвратят дублиране или несъответствия. Една и съща стойност, например името на клиент или идентификатор на продукт, може да се появи по десет различни начина в различни файлове. В релационна база данни това би било невъзможно: всяка таблица е свързана чрез ключове, които гарантират логическа последователност и уникалност на данните (Inmon, 2005; Kimball & Ross, 2013). Това структурно различие обяснява защо толкова често бизнес анализът в Excel започва с часове или дни ръчно почистване – премахване на дубликати, корекция на грешни формати, уеднаквяване на имена и дати. Така възниква феноменът на т.нар. **нееднородни данни (heterogeneous data)** – множество файлове с различни колони, различни формати на датите, несъответстващи типове стойности и липсващи данни. Обикновено всеки файл отразява различен процес, екип или период, но без обща структура, която да ги обедини в единна картина. Според Power (2020), „в корпоративната практика нееднородните данни са по-скоро правило, отколкото изключение“, а най-често срещаните източници на такава нееднородност са именно електронни таблици и CSV файлове, създадени ръчно.

Най-простият пример за структурен конфликт е при форматирането на дати: в един файл датите са записани като „01/02/2025“, в друг като „1 февр. 2025“, а в трети – просто като пореден номер (числото, което Excel използва за броене на дни от 1900 година, например 43551). Подобни

разминавания водят до грешки при обединяване на файлове и – естествено – могат да доведат до погрешни изводи при анализ. В SQL база данни подобно разминаване е невъзможно – типът на данните се определя предварително (DATE, VARCHAR, INTEGER и др.), което предотвратява въвеждането на несъвместими стойности. Същият проблем възниква и при липсата на **метаданни** – информация за това какво означава всяка колона. В Excel често единствената „документация“ е заглавният ред, а той може да бъде променян, съкратен или преведен различно в различни файлове. Без стандарт за именуване и без контрол на версиите, обединяването на множество таблици се превръща в предизвикателство, което често надвишава времето за самия анализ. Както отбелязва Kruck и Maher (2012), електронната таблица „придава илюзия за структура“, но реално оставя потребителя да решава сам какво означава всяка колона и как трябва да се тълкуват данните. Всичко това води до необходимостта от **интеграционен слой** – процес или инструмент, който да обедини неоднородните източници и да наложи минимална форма на хармонизация. В миналото тази роля се е изпълнявала от релационните бази данни, но днес все по-често се използват гъвкави ETL инструменти и low-code решения, които позволяват да се създаде подобна структура *върху* Excel, без да се изисква преминаване към напълно нова платформа. Така Excel продължава да бъде неофициалната „входна точка“ към данните, но съвременните технологии дават възможност те да бъдат интегрирани, почистени и анализирани автоматично, без да се жертва достъпността на познатата среда.

Автоматизацията – от ръчна обработка към интелигентни потоци

В продължение на десетилетия работата с данни в предприятията следва една и съща логика — човекът е в центъра на процеса. Той копира, пренася, сравнява, филтрира и обединява файлове, за да изведе смисъл от информацията. Excel, уви, все още може да бъде основен инструмент на тази ръчна обработка, но именно тя днес е сериозно ограничение. Според проучване на Deloitte (2023), над 40% от времето на бизнес анализаторите в малките и средни предприятия се отделя за техническа подготовка на данни, а не за самия анализ. В тази среда автоматизацията не е просто удобство, а **методологична необходимост**. Тя означава прехвърляне на повторяемите и податливи на грешки стъпки – копиране, съвпадение на колони, проверка на формати – към система, която може да ги извършва сама, по предварително зададени правила. Най-простата форма на това е позната на всеки потребител на Excel: макросите. Те автоматизират рутинни операции чрез запис на последователни действия. Но макар и полезни, макросите остават „вътрешно решение“ – те работят само в рамките на един файл и не решават проблема с интегрирането на данни от множество източници.

С навлизането на **Power Query** (2010 г.) и по-късно на **Power Automate** и **Power BI**, Microsoft постепенно превърна Excel от изолирана електронна таблица в елемент от по-широка екосистема за обработка на данни. Power Query позволи потребителите да извличат, трансформират и зареждат информация (т.нар. ETL процес) от десетки източници — други Excel файлове, бази данни, уеб страници или API услуги. Тези инструменти, въпреки че все още изискват известна техническа настройка, вече въвеждат идеята за **„интелигентен поток на данни“ (data pipeline)** – процес, който може да се изпълнява автоматично и многократно при постъпване на нова информация. Следващата еволюционна стъпка е появата на **low-code и no-code платформи** – визуални среди, в които бизнес потребителите могат да изграждат автоматизирани потоци без програмиране. Примери за такива решения са **Make (Integromat)**, **Zapier**, **Airtable**, **KNIME** и **Talend Open Studio**. Чрез тях една компания може например:

- автоматично да извлича нови Excel файлове от споделена папка или имейл;
- да уеднаквява формати (дата, валута, мерни единици);
- да премахва дублирани записи;
- и накрая да записва резултата в централен отчетен файл или BI платформа.

Подобни сценарии обикновено вече не изискват програмиране, а могат да бъдат създадени от самия бизнес анализатор чрез „визуални блокове“. Gartner (2024) нарича това явление

„демократизация на интеграцията“ – процес, при който достъпът до автоматизацията напуска ИТ отдела и се премества към оперативните звена на организацията. По този начин малките и средните предприятия получават възможност да изграждат собствени потоци за обработка на данни без значителни инвестиции. Различните low-code инструменти прокламират на своите продуктови страници, че могат да съкращават времето за интеграция на данни с до 60% и да намалят риска от човешки грешки. Въпреки това, автоматизацията не е универсално решение на казуса. Нейната ефективност зависи от качеството на първичните данни и от яснотата на логиката, която системата трябва да следва. Ако във входните файлове липсват основни идентификатори, стойностите са непоследователни или форматите варират твърде драстично, нито един инструмент – колкото и интелигентен да е – не може напълно да елиминира нуждата от човешка намеса. Затова днешната цел не е пълна автоматизация, а **създаване на „интелигентна инфраструктура“**, която да комбинира повторяемостта на машината с експертизата на анализатора.

Така автоматизацията се превръща не просто в техническо решение, а в **нов тип организационно знание** – начин компаниите да мислят за данните си като за жив процес, който може да се поддържа, обновява и развива. Именно тази концепция стои в основата на следващата глава, в която ще бъде предложен методологичен модел за интеграция на нееднородни данни, адаптиран за нуждите на малките и средни предприятия.

Модел за автоматизация на данни

Съвременният бизнес анализ изисква не просто инструменти, а **методология** — последователен начин за организиране на данните така, че те да бъдат достъпни, съпоставими и полезни. За да се постигне това, в настоящия доклад се предлага опростен, но универсален модел за автоматизация на анализа на нееднородни данни, базиран на четири основни фази: **Extract (Извличане) -> Harmonize (хармонозиране, превръщане на данните в еднородни) -> Validate (валидиране на хармонизираните данни) -> Analyze (анализ; EHVA)**. Този процес на практика е базиран почти изцяло на класическата ETL парадигма, но е адаптиран за реалността на малките и средни предприятия, които работят с ограничен ресурс и разчитат предимно на Excel и сродни инструменти. EHVA е предназначен за интеграция и анализ на нееднородни бизнес данни в среда с ограничен технологичен ресурс. За разлика от традиционния ETL, EHVA не включва етапа на „зареждане“ на данни (Load), тъй като фокусът е върху уеднаквяването и проверката на данните, а не върху тяхното съхранение.

Извличане на данни

Първият етап от процеса включва системно събиране на всички файлове и таблици, които съдържат релевантна информация – обикновено Excel документи, CSV файлове или експорти от различни системи. Тук ключовият момент не е техническият инструмент, а **последователността на организацията**. Файловете трябва да бъдат събирани по ясен стандарт:

- Уеднаквени имена на колони (напр. „Дата“, „Приход“, „Разход“ вместо произволни наименования);
- Групиране според общ критерий – напр. период, обект или процес;
- Запазване на оригиналните файлове за проследимост.

Дори при ръчно събиране, този етап може да бъде частично автоматизиран чрез инструменти като **Power Automate**, **Zapier** или **Make**, които могат да извличат и подреждат файлове от имейл, облак или вътрешна система (Gartner, 2024).

Уеднаквяване на структурата

След извличането следва най-съществената стъпка – хармонизацията. Тук целта не е пълно уеднаквяване на всички данни, а постигане на **съвместимост**, достатъчна за анализ. Low-code решения като **Power Query**, **KNIME** или **Talend** позволяват на потребителите визуално да зададат как да се обединяват таблици с различни структури:

- Съпоставяне на колони (*schema matching*);
- Конвертиране на формати (дата, валута, мерни единици);
- Обединяване на таблици с различна дълбочина на детайла (напр. дневни и месечни отчети).

Този процес може да се „запише“ като шаблон и да се прилага автоматично при всеки нов набор от данни. Според Kim и Park (2022), подобна визуална настройка може да съкрати времето за интеграция на хетерогенни файлове с до 60% спрямо ръчната обработка.

Проверка и осигуряване на достоверност

След уеднаквяването идва моментът на валидиране — автоматична проверка за липсващи стойности, дублирани редове, несъвпадащи формати и логически грешки.

Дори без машинно обучение, low-code платформи могат да използват **вградени правила за валидация**:

- Проверка за празни клетки в ключови колони (ID, дати, стойности);
- Откриване на дублирани редове чрез сравняване на комбинации от полета;
- Известяване при отклонения от предварително дефинирани граници (например, необичайно висока стойност на разход).

Тези проверки могат да бъдат автоматично документирани, така че при следващ анализ системата да „знае“ какво се счита за допустимо.

Анализ и визуализация

След като данните са извлечени, уеднаквени и валидирани, те могат да бъдат визуализирани и анализирани в BI среда (напр. Power BI, Google Data Studio, Tableau). При всяка нова итерация на процеса — например добавяне на нов месец, нов файл или нова категория — системата може автоматично да опреснява анализите, без нужда от повторна ръчна обработка.

Така се постига **самообновяващ се аналитичен цикъл**, при който усилието е концентрирано в еднократната настройка, а не в постоянната поддръжка. Според Deloitte (2023), подобна архитектура може да намали оперативното време за изготвяне на периодични бизнес отчети с над 50%, като същевременно увеличава точността на данните.

EHVA моделът не изисква скъпа инфраструктура или специализирани екипи. Той може да бъде реализиран с достъпни и познати инструменти, а ключовото му предимство е **принципът на повторяемост** – веднъж настроен, процесът работи автономно и може да се прилага върху неограничен брой нови файлове. В контекста на малките и средни предприятия това означава, че усилието се измества от „правене на анализ“ към „изграждане на процес“, който сам подготвя данните за анализ.

Заключение

В динамичната бизнес среда през 2025 година, данните са не само в стратегически ресурс, но в ежедневен предизвикателен обект за управление. За малките и средни предприятия това предизвикателство често се изразява в липса на хармонизирани източници, ограничен човешки и технически капацитет и зависимост от инструменти като Excel, които, макар и гъвкави, не са проектирани за обработка на големи и неоднородни масиви от информация. Настоящият доклад показва, че автоматизацията на бизнес анализа невинаги изисква непременно сложна инфраструктура или високоспециализиран персонал – тя може да бъде постигната чрез комбинация от съществуващи low-code решения и ясна методология. Предложеният модел по своята същност не е нещо ново и нечувано, но представлява адаптирана рамка, ориентирана към реалните условия на българските малки и средни фирми, които боравят с разнообразни източници и периодично натрупващи се данни. Тази методология извежда фокуса от традиционната трансформация и съхранение (ETL) към процес на **хармонизиране и повторяем анализ**, който може да се реализира в достъпна среда. По този начин EHVA не просто описва технически етапи, а предлага нов организационен подход към обработката на данни – при който усилието е концентрирано в еднократната настройка, а не в рутинното ръчно изпълнение. Внедряването на подобен модел носи няколко съществени предимства: съкращава времето за подготовка на информация, намалява риска от човешки грешки и осигурява по-висока повторяемост на резултатите. Най-съществено обаче е, че то позволява на предприятията да **превърнат данните си в инструмент за придобиване на знание**, а не просто в отчетен ресурс. Съчетаването на машинното обучение, low-code технологиите и ясно дефинирана логика на обработка създава възможност за устойчива трансформация – такава, която не замества човека, а го освобождава да мисли стратегически.

Автоматизацията на бизнес анализите чрез интегриране на неоднородни данни не е само технологична тенденция, а форма на бизнес интелигентност. В свят, в който скоростта на промяната надвишава способността ни да реагираме, именно структурираното, хармонизирано знание ще бъде най-ценният капитал на малкия и средния бизнес.

Източници

1. Berthold, M.R., Cebron, N., Dill, F., et al. (2022). *KNIME: the Konstanz information miner,* KNIME Press.
2. Deloitte (2023). *The Rise of Low-Code/No-Code Development: Democratizing Application Development.* Deloitte Insights.
3. Gartner (2024). *Hype Cycle for Data Integration Tools.* Gartner Inc. Available at: https://www.gartner.com/en/articles/data-integration (Accessed: October 2025).
4. Gartner (2024). *Market Guide for Data Quality Tools.* Gartner Inc. Available at: https://www.gartner.com/en/data-analytics/topics/data-quality (Accessed: October 2025).
5. Gartner (2024). *Enterprise Low-Code Application Platforms.* Gartner Peer Insights. Available at: https://www.gartner.com/reviews/market/enterprise-low-code-application-platform (Accessed: October 2025).
6. Inmon, W.H. (2005). *Building the Data Warehouse* (4th ed.). Wiley.
7. Elshan, E., Dickhaut, E. and Ebel, P. (2023). *An Investigation of Why Low Code Platforms Provide Answers and New Challenges*
8. Kimball, R. and Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley.
9. Panko, R.R. and Ordway, N. (2005). *Sarbanes–Oxley: What About All the Spreadsheets? Proceedings of the European Spreadsheet Risks Interest Group (EuSpRIG 2005)*

10. Microsoft (2023). *Power Query and Power Automate Documentation.* Microsoft Learn. Available at: https://learn.microsoft.com/power-query (Accessed: October 2025).
11. Panko, R.R. (2008). *Spreadsheet Errors: What We Know. What We Think We Can Do.*
12. Power, D.J. (2015). *Data Science Supporting Decision Making.* Available at: https://www.tandfonline.com/doi/full/10.1080/12460125.2016.1171610 (Accessed: October 2025)
13. Walkenbach, J. (2013). *Excel Bible* (10th Anniversary ed.). Wiley.