

ИСПОЛЬЗОВАНИЕ БИБЛИОТЕКИ DASH КАК BI-СИСТЕМЫ В УСЛОВИЯХ ИМПОРТОЗАМЕЩЕНИЯ

USING THE DASH LIBRARY AS A BI-SYSTEM IN CONDITIONS OF IMPORT SUBSTITUTION

**D. Trifanov
A. Prostakova
A. Popov**

Summary. In the article, the authors of the work examine the use of BI systems common in companies in the Russian Federation, as well as the Dash library of the Python programming language for their compliance with legislation, speed of data processing and ease of use. Special attention is paid to a review of the legislative reality in the field of information technology use at the moment; it is on the basis of this review that the Dash library is proposed as a means of developing interactive dashboards for a company due to its versatility and the absence of a licensed boxed tool, which reduces legislative risks to a minimum. The authors, using open data, by modeling and developing the same analytical dashboard using different tools, conduct a comparative analysis of popular BI systems of the Dash tool according to the following parameters: speed of data processing and building visualizations, complexity of building an analytical dashboard, the possibility of using third-party servers without specialized software, maximum volume of processed data. As a result of the analysis, conclusions are formed, the advantages and disadvantages of each tool are considered, and Dash is highlighted as a modern and technologically advanced way to create analytical reporting.

Keywords: BI system, Python, Dash, Python, Data Studio, Tableau, import substitution.

В настоящее время современное ведение бизнеса в эпоху цифровой трансформации претерпевает значительные изменения. Процесс развития научно-технического прогресса информационных технологий происходит с огромной скоростью. Вследствие чего увеличивается потребность бизнеса в различных сферах, в том числе в преобразовании и визуализации данных. Существует множество методов, с помощью которых можно удовлетворить эти потребности, при этом они имеют системный подход к решению проблем бизнеса. Наиболее популярной реализацией таких методов в современном мире являются системы аналитической отчетности и обработки данных BI-системы (Business Intelligence), которые представляет собой набор различных методов и технологий для сбора, анализа и обработки данных. Из-за ограниченности множества ресурсов

Трифанов Данила Сергеевич
ООО «ВК-ИТ»
trifanovrmk@gmail.com
Простакова Анастасия Александровна
ООО «ХКФ Банк»
prostakova.anastasia@yandex.ru
Попов Алексей Анатольевич
Кандидат технических наук, доцент, Российский
экономический университет им. Г.В. Плеханова
popov.aa@rea.ru

Аннотация. В статье авторами работы рассматривается использование распространенных в компаниях Российской Федерации BI-систем, а также библиотеки Dash языка программирования Python на предмет их соответствия законодательству, скорости обработки данных и удобства эксплуатации. Отдельное внимание уделяется обзору законодательной действительности в сфере использования информационных технологий на текущий момент, именно исходя из данного обзора предлагается библиотека Dash, как средство разработки интерактивных дашбордов для компании в виду её универсальности и отсутствия лицензированного коробочного инструмента, что сводит законодательные риски к минимуму. Авторами, с помощью использования открытых данных, путём моделирования и разработки одного и того же аналитического дашборда с помощью разных инструментов проводится сравнительный анализ популярных BI-систем инструмента Dash по следующим параметрам: скорость обработки данных и построения визуализаций, сложность построения аналитического дашборда, возможность использования стороннего сервера без специализированного ПО, максимальный объем обрабатываемых данных. В результате анализа формируются выводы, рассматриваются преимущества и недостатки каждого инструмента, выделяет Dash как современный и технологичный способ создания аналитической отчетности.

Ключевые слова: BI-система, Python, Dash, Python, Data Studio, Tableau, импортозамещение.

в рыночной экономике, в том числе в Российской Федерации, данные системы являются высоко эффективными.

Обратим внимание на ситуацию непосредственно в РФ, где наиболее распространенными BI-системами являются MS Power BI, Data Studio, Tableau и Yandex Datalents, однако вследствие изменения законодательства могут возникнуть проблемы с их использованием [1]. В данной статье будет рассмотрена проблема соответствия различных BI-систем законодательству, скорости обработке данных и удобства взаимодействия, а также предложен вариант использования языка программирования Python и библиотеки Dash как самостоятельной BI-системы. Актуальность заключается в том, что в РФ для обработки данных можно использовать только

ко отечественное ПО, однако далеко не все BI-системы удовлетворяют данному критерию, поэтому необходимо искать пути решения, а непосредственно целью работы будет использование библиотеки Dash как решение исследуемой проблемы. Объектом исследования будут являться BI-системы, в том числе система на Python.

Для решения поставленной проблемы необходимо в рамках научного исследования применить такие методы, как

1. Юридический анализ, представляющий собой логический процесс разложения правовой нормы на отдельные элементы и признаки, а также формирование выводов соответствия правовому полю
2. Моделирование и разработка BI решений с помощью различного ПО для дальнейшего сравнительного анализа
3. Сравнительный анализ BI решений, необходимый для формирования выводов и заключения

Для начала стоит рассмотреть законодательную действительность Российской Федерации. В настоящий момент Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных» обязывает всех операторов сотовой связи обрабатывать персональные данные граждан РФ Исключительно при помощи баз данных, расположенных на территории России (ч. 5 ст. 18) [2]. Аналогичное правило предусмотрено и в Федеральном законе от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации» [3].

Хотелось бы отметить, что данный закон имеет исключения, а также разъяснение, согласно которому обрабатывать данные граждан России необходимо с использованием баз данных, не находящихся в России в ситуациях, если такая деятельность не подпадает под случаи, предусмотренные п.п. 2, 3, 4, 8 ч. 1 ст. 6 Закона N 152-ФЗ, и на территории России находятся используемые для обработки персональных данных базы данных, в которых содержится больший или равнозначный объем информации.

Такое разъяснение базируется на ч. 5 ст. 18 Закона N 152-ФЗ и общих положениях указанного закона, в которых не имеется разделения на «основную» базу с персональными данными и ее «копию» [2]. В обоих случаях речь идет о базах данных, с помощью которых обрабатываются персональные данные. Наряду с этим, закон не содержит указаний на полный запрет обработки данных граждан России с применением баз данных, не находящихся на территории России.

Далее рассмотрим некоторые системы на предмет соответствия ФЗ №152 следующих систем: MS Power BI, Data Studio и Tableau. В данный момент ни одна из этих

систем не удовлетворяет требованиям закона, так как сервера этих систем находятся в Европе и США. Именно поэтому использование напрямую облачных технологий становится невозможным, вследствие чего компании не могут в полной мере использовать все возможности рассмотренных BI-систем [4]. Это, безусловно усложняет развертывание систем в рамках компании и сокращает объем используемых бизнесом технологий. Более того, усложняет доступ компаний партнеров к дашбордам, например, в сфере аутсорсинга, ведь не каждая СОБ допустит стороннего клиента к использованию внутренних серверов компании. Следовательно, необходимо решение данной проблемы.

Возможным решением может являться разработка BI-решения с помощью инструмента, который позволяет самостоятельно разработать дашборд без коробочного решения [5]. Использование одного из высокоуровневых языков программирования и специализированной библиотеки может быть ключом к решению данной проблемы: для этого подходит язык программирования Python с библиотекой Dash, предназначенной для создания аналитических дашбордов [6]. Для оценки рациональности использования данного инструмента необходимо построить аналогичные аналитические дашборды на основании одинакового набора данных с помощью каждого из средств построения и провести сравнительный анализ по ключевым параметрам.

Изначально, необходимо определиться с общей структурой аналитического дашборда. Для построения визуализаций будет использовано 10 миллионов строк CSV файла открытого датасета об электронной коммерции. Датасет представляет собой набор данных со следующими атрибутами:

- Время целевого действия
- Тип события
- Категория товара
- Бренд
- Цена
- Пользовательский ID

На основании набора данных будут построены следующие визуализации:

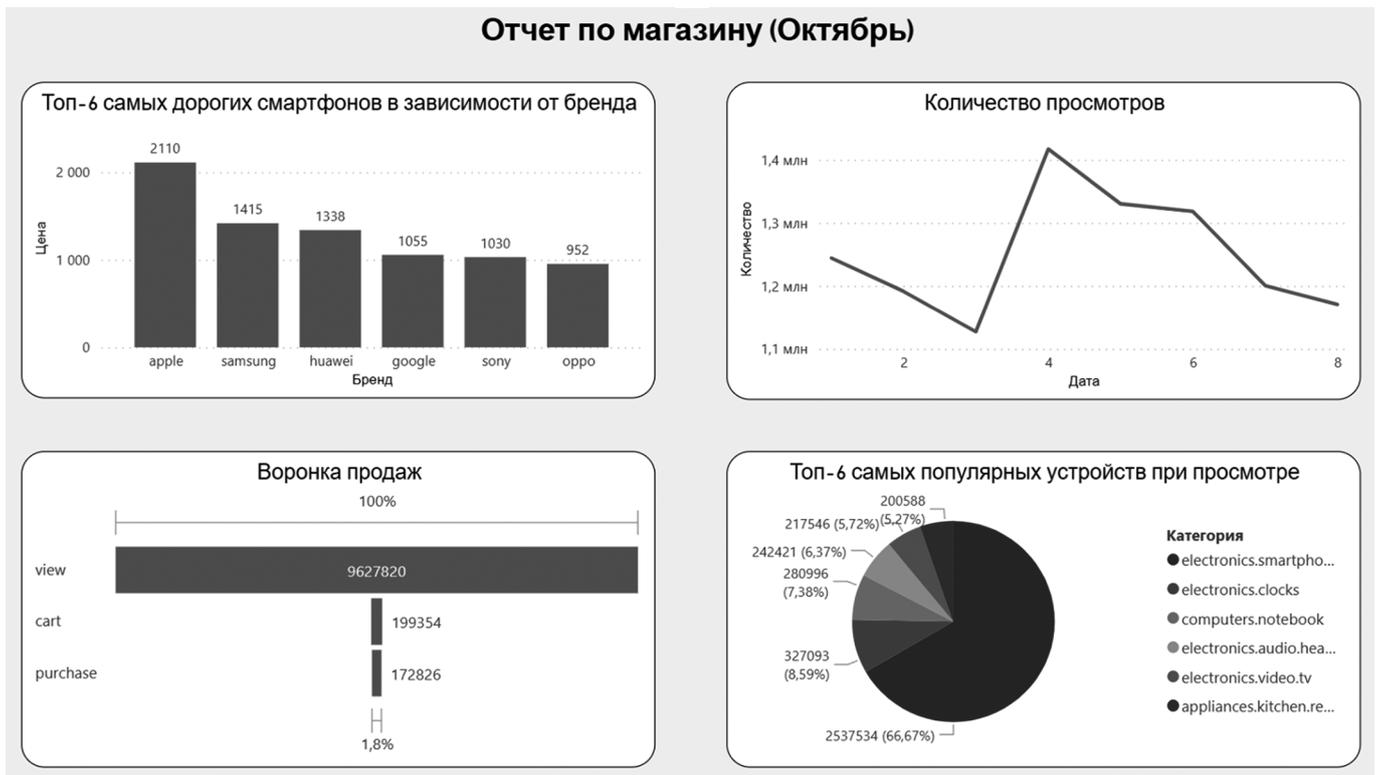
- гистограмма «Топ-6 самых дорогих смартфонов в зависимости от бренда»: считаем максимальную цену смартфона в каждом бренде, а также фильтруем первые 6;
- воронка «Воронка продаж»: строим воронку по типам событий и наблюдаем конверсию в покупку;
- график «Количество просмотров»: зависимость количества просмотров устройств от даты;
- круговая диаграмма «Топ-6 самых популярных устройств при просмотре».

Исходя из этого и будем строить аналитические дашборды с помощью различных средств визуализации.

При использовании построений визуализаций с помощью BI системы Tableau, которая подключается к источнику данных напрямую, время загрузки занимает значительное время и при размере данных больше 10Gb делает время построения визуализаций от 4 до 60 минут [7]. Более того, использование системы практически невозможно в виду сложившейся санкционной политики, в связи с этим, использование системы для подсчета скорости загрузки в современных реалиях не имеет практического значения.

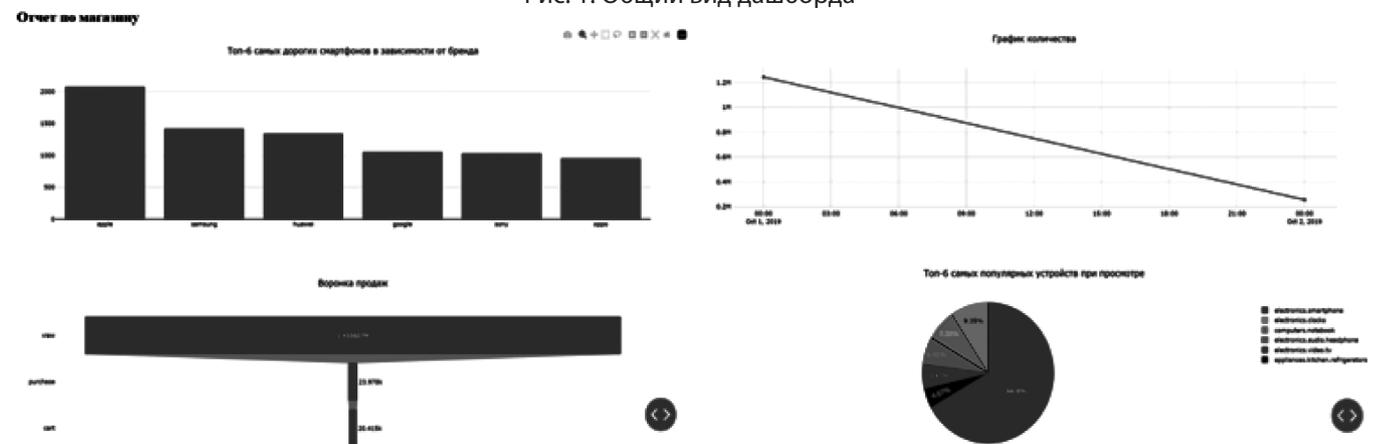
Для сравнения построим дашборды в MS Power BI, Data Studio и непосредственно с помощью Dash. Общая структура дашборда изображена на Рисунке 1.

Более подробно рассмотрим построение дашборда с помощью Python. Изначально необходимо импортировать необходимые для проведения разработки библиотеки. Далее, импортируем данные в переменную с помощью библиотеки Pandas. После этого на основании полученных данных обрабатываем их с помощью агрегации для построения визуализации.



Источник: Составлено авторами на основании [7], [8], [9]

Рис. 1. Общий вид дашборда



Источник: Составлено авторами на основании [10]

Рис. 2. Дашборд, построенный с помощью Python Dash

На основании полученных агрегаций сделаем хостинг аналитического дашборда на локальном компьютере, после чего увидим результат.

Исходя из проделанной работы, можно утверждать о том, что агрегированные данные с помощью различных средств визуализации одинаковы в результате, а также бизнес-смысле. На их основании можно проводить сравнительный анализ.

Сравнительный анализ необходимо построить на основании ключевых факторов, влияющих на скорость разработки BI решения, скорости его работы и использования в современных законодательных условиях (Табл. 1). Исходя из эмпирических наблюдений была составлена следующая сравнительная таблица, приведенная ниже.

Таблица 1.

Сравнительные характеристики BI-систем и Dash

	Скорость обработки данных и построения визуализаций	Сложность построения (max 3)	Возможности персонализации (max 3)	Возможность использования стороннего сервера без специализированного ПО	Максимальный объем данных
Power BI RS	4:23	1	2	–	1 ГБ (без Premium)
Datastudio	40:23	2	1	–	
Dash (Python)	1:01	3	3	+	Зависит от мощности компьютера

Источник: Составлено авторами на основании [7], [8], [9]

Исходя из таблицы можно сделать следующие выводы:
 — разработка дашборда с помощью языка программирования делает его более производительным, чем использование коробочного решения, которое само написано на одном из языков [11];

- при одинаковых вычислительных мощностях дашборд разработанный с помощью Dash показывает лучшие результаты, в том числе из-за внутренней агрегации данных [12];
- возможности персонализации дашборда на Dash являются самыми широкими за счет использования CSS, HTML и JS однако, в Power BI также используют популярные визуальные практики [13];
- для разработки BI-системы на основании Python требуются высококвалифицированные сотрудники, знающие инструменты обработки данных на Python, включая библиотеки Numpy и Pandas, а также имеющие навыки программирования JS, HTML.

Итак, на основании проведенного исследования, можно прийти к следующему выводу: действительно, в современной правовой реальности можно успешно разрабатывать дашборды с помощью Dash, при этом они смогут работать с большим объемом данных и иметь хорошие возможности для визуализации. Однако могут появиться сложности при создании, так как это требует высокой квалификации. При прочих возможностях тестирования различных BI решений с одного АРМ с присущими ему неизменными техническими характеристиками можно сделать вывод, что для обработки значительных и больших данных по объему без использования предварительной агрегации наиболее успешно использование не коробочных BI систем, а специализированных языков программирования в виду ограничений на объем принимаемых данных и отсутствия инструментов внутренней агрегации больших данных, как, например, Yandex Datalens, что усложняет первичный анализ данных. Однако, стоит отметить, что, например, использование специализированной библиотеки Dash языка программирования Python требует значительных человеческих ресурсов, как в квалификации разработчика BI решения, так и в его владения методиками внутренней агрегации данных и персонализации UI.

ЛИТЕРАТУРА

1. Указ Президента Российской Федерации от 30.03.2022 № 166 «О мерах по обеспечению технологической независимости и безопасности критической информационной инфраструктуры Российской Федерации».
2. Федеральный закон от 27.07.2006 N 149-ФЗ (ред. от 08.06.2020) «Об информации, информационных технологиях и о защите информации».
3. Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 30.12.2020) «О персональных данных» (с изм. и доп., вступ. в силу с 01.03.2021).
4. Даббас Э., Интерактивные дашборды и приложения с Plotly и Dash. Используем полноценный веб-фреймворк в Python на всю мощь — без JavaScript / перевод с английского А.Ю. Гинько. — Москва: ДМК Пресс, 2023. — 306 с.
5. Зыкова Г.В., Основы программирования на языке Python: учебно-методическое пособие / Г.В. Зыкова, А.С. Попов, Т.Н. Сапуглецева; научный редактор Г.В. Зыковой. — 2-е изд., стер. — Москва: ФЛИНТА, 2020. — 135 с.
6. Маккинни, У. Python и анализ данных / У. Маккинни; перевод с английского
7. Microsoft Power BI documentation [Электронный ресурс] — Режим доступа: <https://docs.microsoft.com/ru-ru/power-bi/> (Дата обращения 10.01.2024)
8. Tableau documentation [Электронный ресурс] — [Электронный ресурс] — Режим доступа: <https://docs.microsoft.com/ru-ru/power-bi/> (Дата обращения 10.01.2024)

9. Documentation of the Data Studio library [Электронный ресурс] — Режим доступа: <https://docs.microsoft.com/ru-ru/power-bi/> (Дата обращения 10.01.2024)
10. Dash Python Library Documentation [Электронный ресурс] — Режим доступа: <https://docs.microsoft.com/ru-ru/power-bi/> (Дата обращения 10.01.2024)
11. Мерчант, Б. Power BI: передовые методы оптимизации; перевод с английского А.Ю. Гинько. — Москва: ДМК Пресс, 2023. — 282 с.
12. Трифанов, Д.С. Архитектурное решение построения системы сквозной аналитики на основании open source ETL-систем / Д.С. Трифанов, А.А. Простакова // XXXV международные Плехановские чтения: Сборник статей дипломантов, Москва, 22–24 марта 2022 года. — Москва: Российский экономический университет имени Г. В. Плеханова, 2022. — С. 297–301. — EDN CNMPGY.
13. Черников Б.В., Попов А.А. Выбор информационной системы с учетом уровня готовности предприятия к информатизации // Информатизация и связь. 2016, №3, с. 152–159.

© Трифанов Данила Сергеевич (trifanovrmk@gmail.com); Простакова Анастасия Александровна (prostakova.anastasia@yandex.ru);
Попов Алексей Анатольевич (popov.aa@rea.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»