

Анализ качества вина с применением дашборда на основе открытых датасетов в Yandex DataLens

Пронин Артём Евгеньевич

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

В работе выполнен анализ качества вина на основе открытых датасетов с применением инструментов бизнес-аналитики. Результаты показали, что использование информационных технологий, в частности облачного сервиса Yandex DataLens, позволяет эффективно визуализировать и анализировать влияние физико-химических характеристик на итоговую оценку качества вина. Отмечается, что применение интерактивных дашбордов способствует повышению наглядности аналитических выводов, оптимизации процесса анализа данных и открывает новые возможности для принятия обоснованных решений в области оценки и контроля качества продукции.

Ключевые слова: бизнес-аналитика, Yandex DataLens, дашборд, визуализация данных, анализ данных, качество вина.

Analysis of wine quality using a dashboard based on open datasets in Yandex DataLens

Pronin Artem Evgenievich

Sholom-Aleichem Priamursky State University

Student

Abstract

The study analyzes wine quality based on open datasets using business intelligence tools. The results show that the use of information technologies, in particular the Yandex DataLens cloud service, enables effective visualization and analysis of the influence of physicochemical characteristics on the final wine quality score. It is noted that the use of interactive dashboards enhances the clarity of analytical conclusions, optimizes the data analysis process, and opens up new opportunities for making informed decisions in the field of product quality assessment and control.

Keywords: business intelligence, Yandex DataLens, dashboard, data visualization, data analysis, wine quality.

1 Введение

1.1 Актуальность

В условиях цифровизации и роста объёмов данных возрастает потребность в эффективных инструментах их анализа и визуализации. BI-системы позволяют оперативно обрабатывать большие массивы информации,

выявлять скрытые закономерности и поддерживать процесс принятия управленческих решений. Одним из востребованных инструментов визуальной аналитики является Yandex DataLens, обеспечивающий возможности построения интерактивных дашбордов и аналитических отчётов на основе различных источников данных. Анализ качества продукции, в том числе винодельческой, на основе открытых датасетов с применением BI-инструментов представляет практический интерес и позволяет наглядно оценивать влияние ключевых факторов на итоговые показатели качества.

1.2 Обзор исследований

В статье Д. Д. Акентьева представлен практический подход к проведению разведочного анализа данных в среде Yandex DataLens. Автор рассматривает процесс работы с датасетами, создание чартов и дашбордов, а также анализ изменений ключевых показателей компании на основе визуальной аналитики. [1].

В статье П. А. Полянской рассматриваются типы данных, используемые в BI-инструменте Yandex DataLens. Автор описывает особенности различных форматов данных и подчёркивает важность их корректной интерпретации для построения дашбордов, визуализации информации и принятия аналитических решений. [2].

В работе I. Sahipov, A. Zabashta и A. Filchenkov исследуется проблема симметрии матричного представления классификационных датасетов. Авторы предлагают методы стабилизации структуры данных, направленные на повышение эффективности генерации датасетов и выбора алгоритмов машинного обучения с применением генеративно-состязательных нейронных сетей. [3].

В статье Н. О. Плющ и Н. А. Федьковой рассматриваются основные возможности BI-сервиса Yandex DataLens, а также области его применения в бизнес-аналитике. Авторы анализируют преимущества и ограничения использования данного инструмента, уделяя внимание его роли в системах поддержки принятия решений и визуализации аналитических данных. [4].

В статье А. В. Бородина рассматривается применение BI-приложений для анализа рынка кино. Автор описывает использование сервиса Yandex DataLens для проверки выдвинутых гипотез на основе данных, полученных с платформы Kaggle, а также демонстрирует возможности визуализации и анализа информации с помощью дашбордов. [5].

1.3 Цель исследования

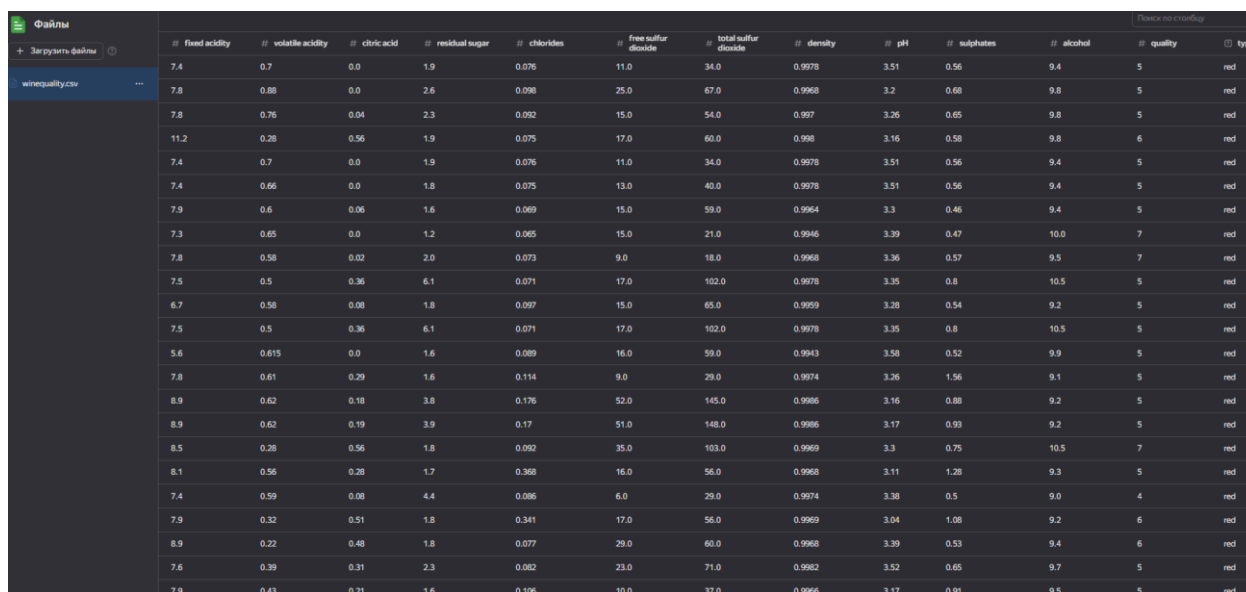
Целью исследования является разработка методики создания интерактивного дашборда для анализа качества красных и белых вин на основе датасетов Wine Quality с использованием возможностей сервиса Yandex DataLens.

2 Материалы и методы

В качестве исходных данных использовались открытый датасет winequality.csv [6], содержащий информацию о физико-химических свойствах вина и экспертной оценке качества. Для анализа применялся облачный сервис Yandex DataLens, предоставляющий инструменты загрузки данных, создания вычисляемых полей, построения визуализаций и формирования интерактивных дашбордов.

3 Результаты и обсуждение

На первом этапе CSV-файлы с данными о красных и белых винах были загружены в Yandex DataLens через раздел «Подключения».



fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	type
7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	red
7.8	0.88	0.0	2.6	0.098	25.0	67.0	0.9968	3.2	0.68	9.8	5	red
7.8	0.76	0.04	2.3	0.082	15.0	54.0	0.997	3.26	0.65	9.8	5	red
11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.998	3.16	0.58	9.8	6	red
7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	red
7.4	0.66	0.0	1.8	0.075	13.0	40.0	0.9978	3.51	0.56	9.4	5	red
7.9	0.6	0.06	1.6	0.069	15.0	59.0	0.9964	3.3	0.46	9.4	5	red
7.3	0.65	0.0	1.2	0.065	15.0	21.0	0.9946	3.39	0.47	10.0	7	red
7.8	0.58	0.02	2.0	0.073	9.0	18.0	0.9968	3.36	0.57	9.5	7	red
7.5	0.5	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.8	10.5	5	red
6.7	0.58	0.08	1.8	0.087	15.0	65.0	0.9959	3.28	0.54	9.2	5	red
7.5	0.5	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.8	10.5	5	red
5.6	0.615	0.0	1.6	0.089	16.0	59.0	0.9943	3.58	0.52	9.9	5	red
7.8	0.61	0.29	1.6	0.114	9.0	29.0	0.9974	3.26	1.56	9.1	5	red
8.9	0.62	0.18	3.8	0.176	52.0	145.0	0.9986	3.16	0.88	9.2	5	red
8.9	0.62	0.19	3.9	0.17	51.0	148.0	0.9986	3.17	0.93	9.2	5	red
8.5	0.28	0.56	1.8	0.092	35.0	103.0	0.9969	3.3	0.75	10.5	7	red
8.1	0.56	0.28	1.7	0.368	16.0	56.0	0.9968	3.11	1.28	9.3	5	red
7.4	0.59	0.08	4.4	0.086	6.0	29.0	0.9974	3.38	0.5	9.0	4	red
7.9	0.32	0.51	1.8	0.341	17.0	56.0	0.9969	3.04	1.08	9.2	6	red
8.9	0.22	0.48	1.8	0.077	29.0	60.0	0.9968	3.39	0.53	9.4	6	red
7.6	0.39	0.31	2.3	0.082	23.0	71.0	0.9982	3.52	0.65	9.7	5	red
7.9	0.43	0.21	1.6	0.106	10.0	37.0	0.9966	3.17	0.91	9.5	5	red

Рис. 1 – Загрузка данных о винах

После загрузки DataLens автоматически определил типы полей (числовые и категориальные), такие как:

- фиксированная кислотность,
- летучая кислотность,
- содержание алкоголя,
- плотность,
- показатель качества (quality).

Создан рабочий воркбук. Для этого в левом навигационном меню сервиса выбран раздел «Воркбуки», после чего выполнена команда «Создать воркбук».

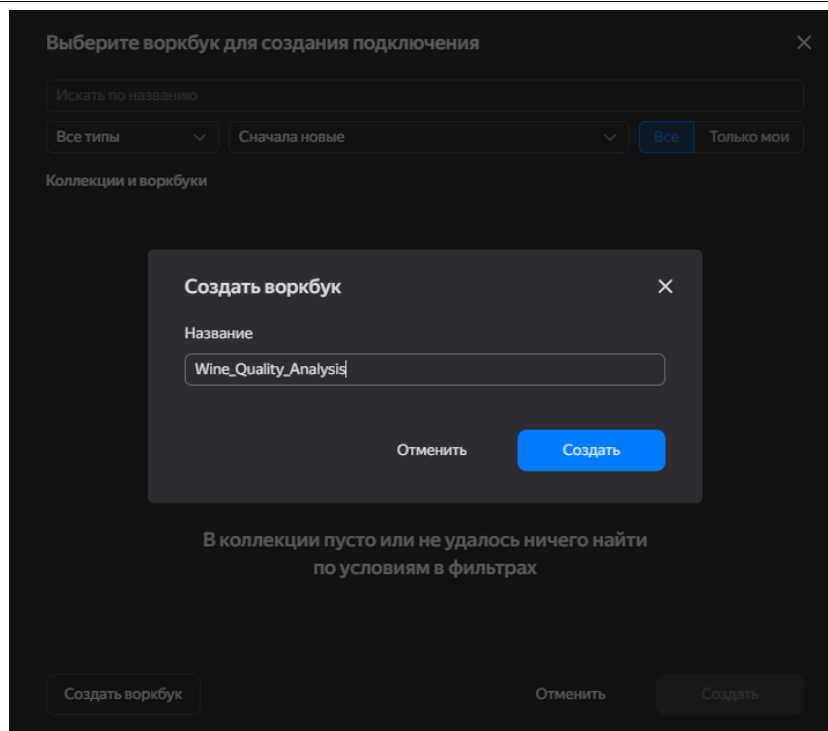


Рис. 2 – Создание воркбука

В качестве источника данных при создании датасета выбран ранее сформированный воркбук вин(рис.2), содержащий все загруженные файлы и аналитические объекты, относящиеся к исследованию. После выбора воркбука в структуру датасета добавлено ранее созданное подключение «вино»(рис.1), обеспечивающее доступ к исходным CSV-файлам с характеристиками вина. Добавление подключения позволило связать датасет с физическим источником данных, после чего система автоматически отобразила перечень доступных полей и их типы, что создало основу для дальнейшего создания вычисляемых показателей и построения аналитических визуализаций(рис.3).

#	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
1	7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
2	7.8	0.88	0.0	2.6	0.098	25.0	67.0	0.9968	3.2	0.68	9.8	5
3	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.997	3.26	0.65	9.8	5

Рис. 3 – Создание датасета

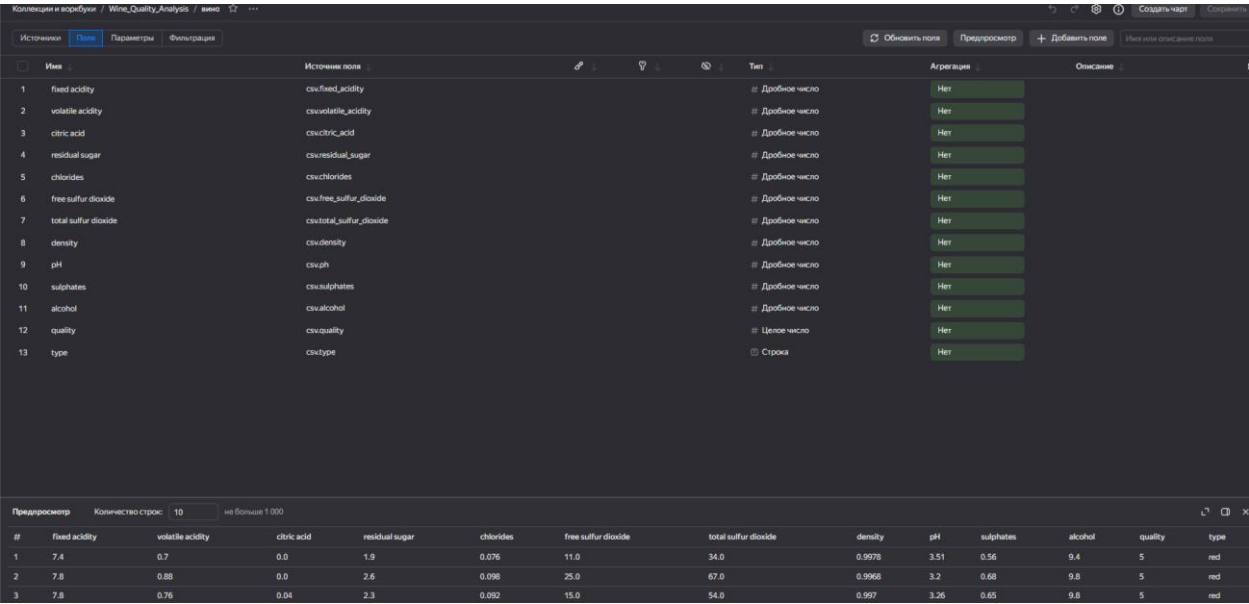


Рис. 4 – Созданный датасет

Для расширения аналитических возможностей в датасете созданы вычисляемые поля. Для этого в верхней панели датасета выбран пункт «Добавить поле»(рис.4), после чего в редакторе формул задавались названия полей и логика их расчёта на основе исходных показателей. Созданные вычисляемые поля использовались при построении аналитических визуализаций.

Поле «Тип вина» использовалось для разделения наблюдений на красное и белое вино(рис.5). Создание данного поля позволило проводить сравнительный анализ характеристик и качества вин в зависимости от их типа.

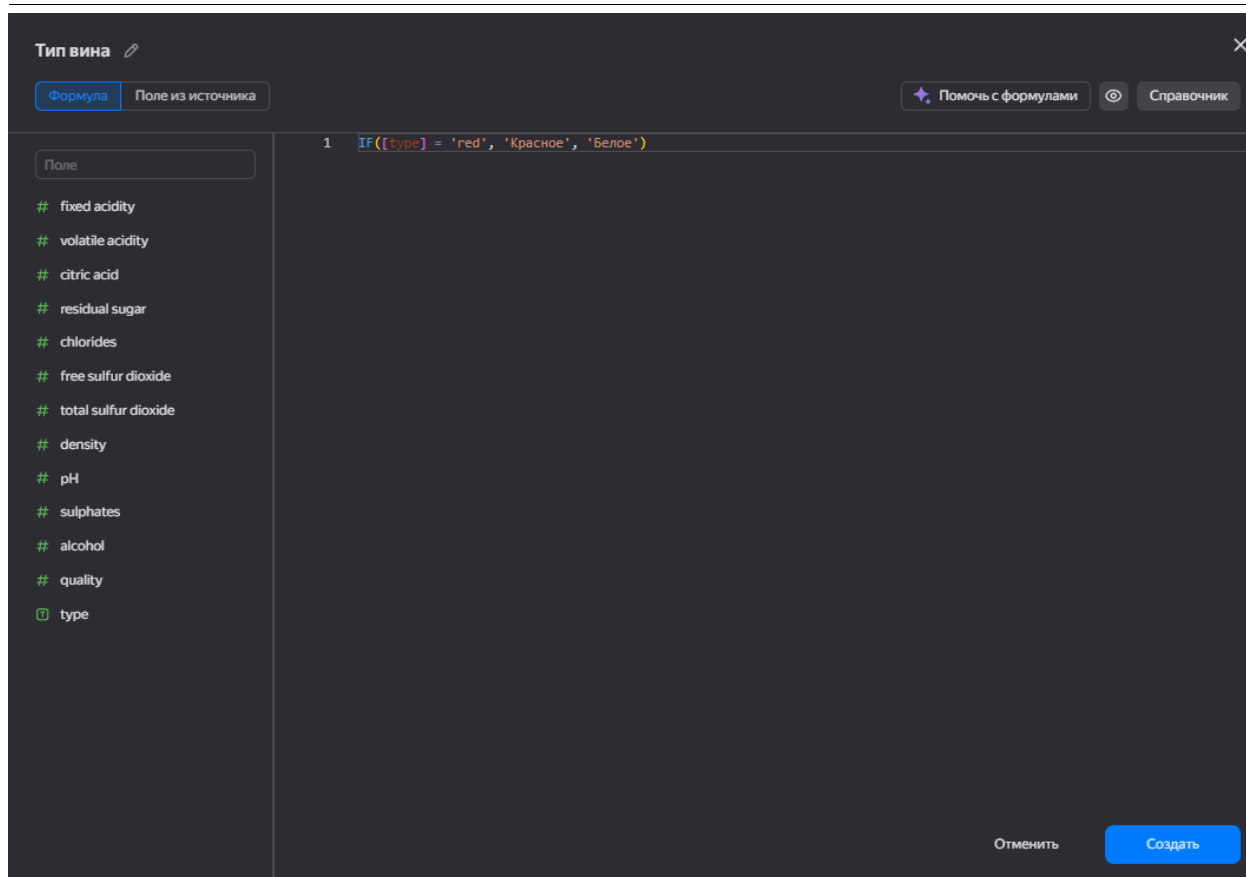


Рис. 5 – Формула для поля «Тип вина»

Для упрощения интерпретации показателя quality выполнена его категоризация(рис.6). Значения качества сгруппированы в три категории: низкое, среднее и высокое качество.

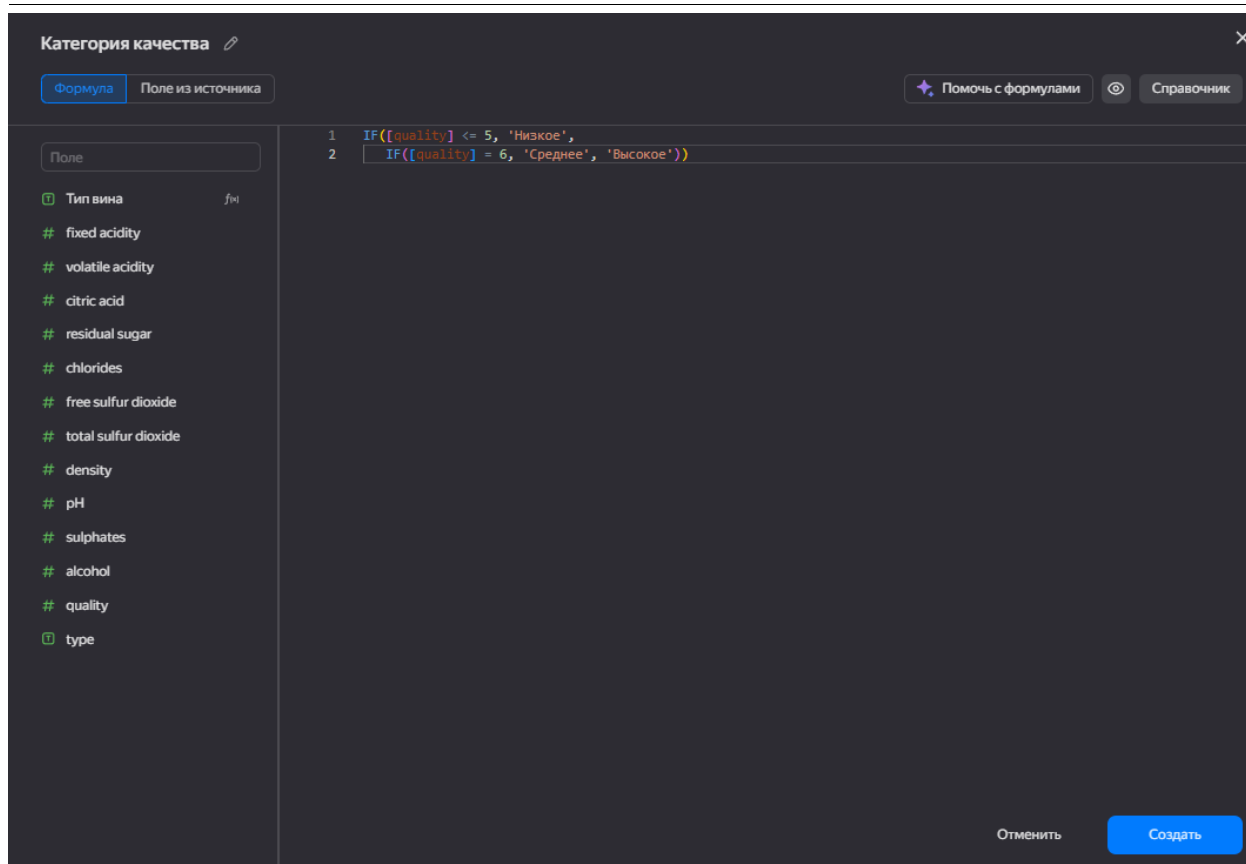


Рис. 6 – Формула для поля «Категория качества»

Для анализа влияния содержания алкоголя на качество вина создан категориальный показатель уровня алкоголя(рис.7). Данное поле использовалось при построении визуализаций и интерактивных фильтров, позволяя выявлять зависимости между содержанием алкоголя и оценкой качества вина.

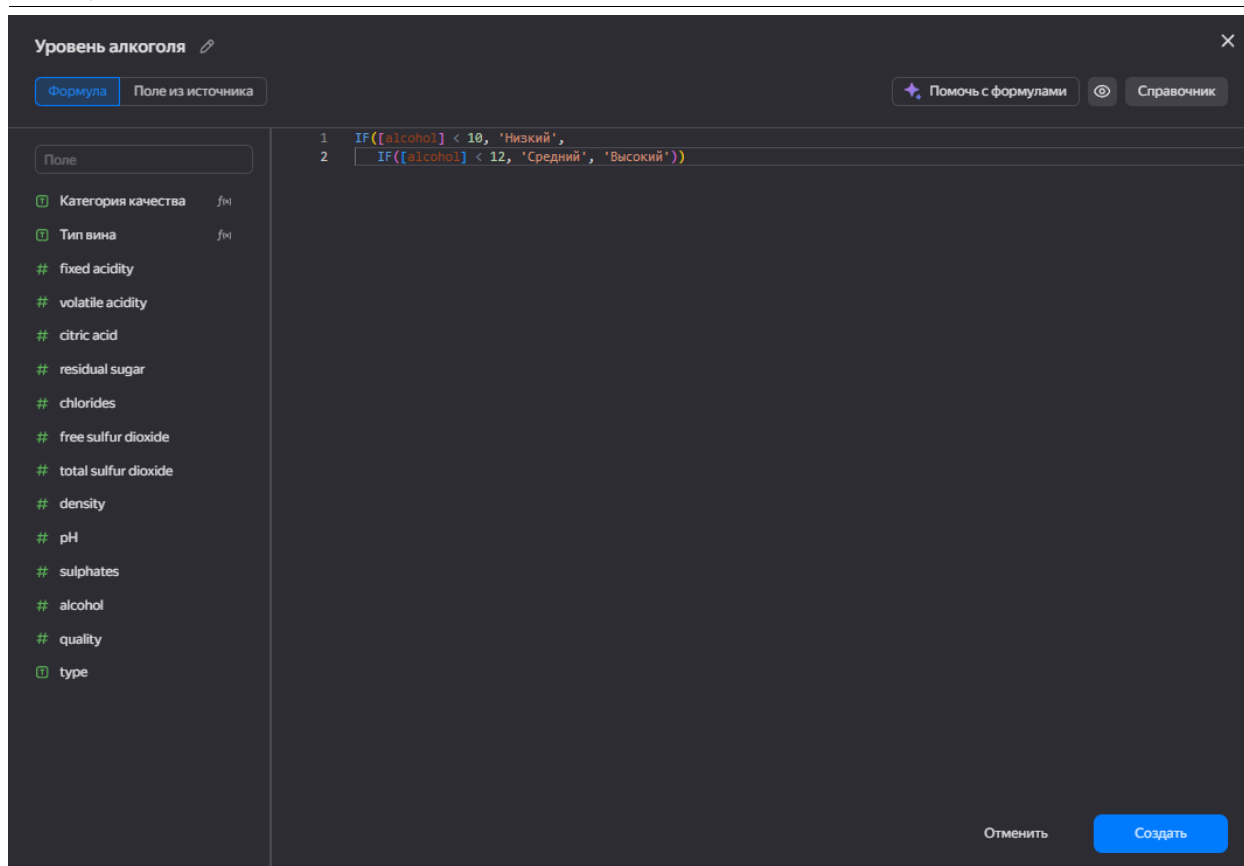


Рис. 7 – Формула для поля «Уровень алкоголя»

Для сравнительного анализа качества вин построена столбчатая диаграмма(рис.8). В качестве категориальной оси Х использовалось поле «Тип вина», а по оси Y отображалось среднее значение показателя качества. Построенная диаграмма позволила наглядно сравнить уровень качества красных и белых вин.

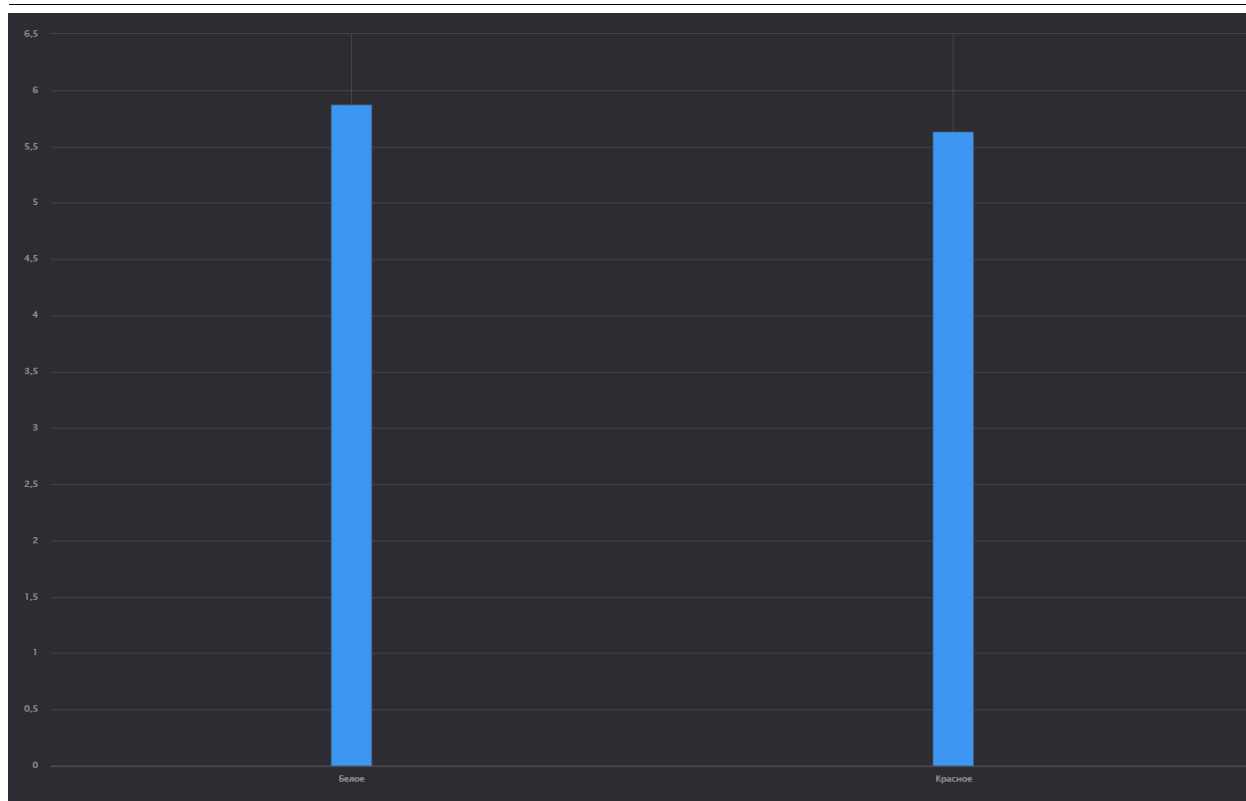


Рис. 8 – Среднее качество по типам вина

В результате построения столбчатой диаграммы установлено, что среднее значение оценки качества белых вин несколько выше по сравнению с красными (рис.8). Полученная визуализация наглядно демонстрирует различия в уровне качества в зависимости от типа вина.

Для анализа влияния содержания алкоголя на оценку качества вина построена линейная диаграмма (рис.9). По оси X отображалось значение показателя содержания алкоголя, по оси Y — среднее значение оценки качества вина.

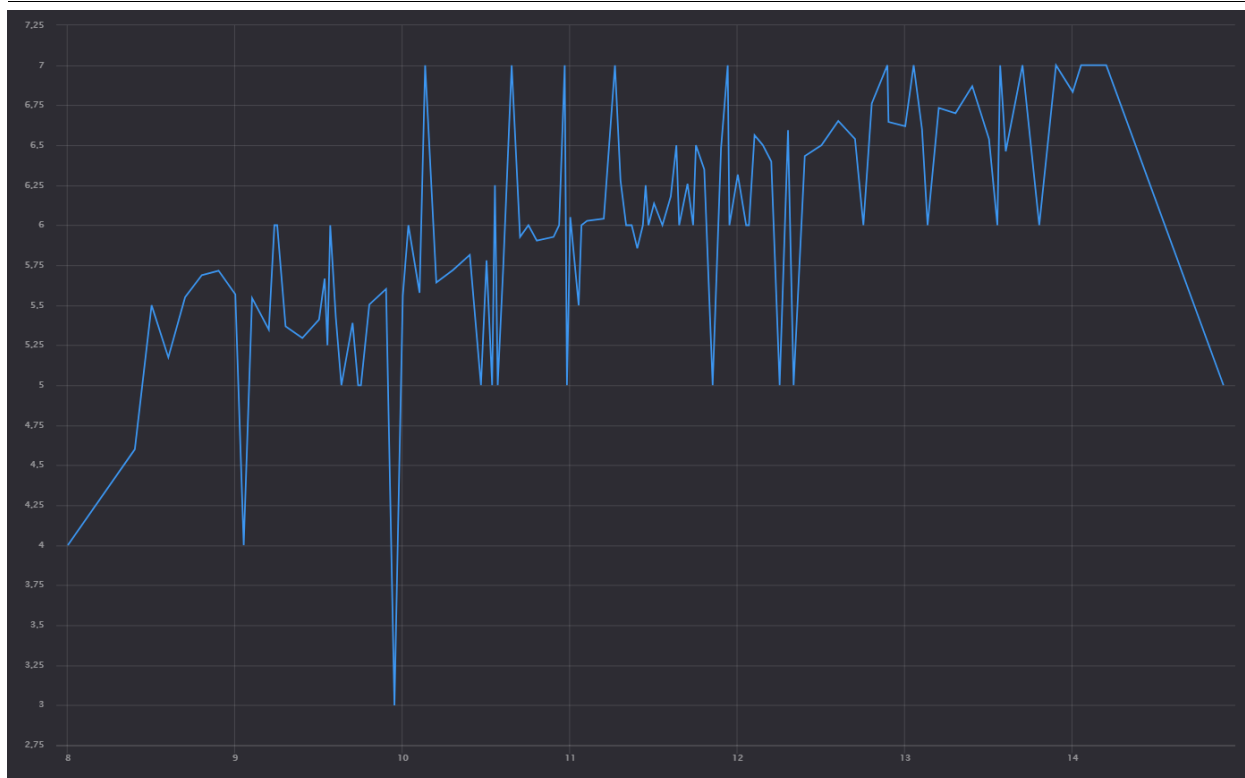


Рис. 9 – Зависимость качества вина от содержания алкоголя

Построенная линейная диаграмма (рис.9) демонстрирует общую положительную зависимость между содержанием алкоголя и средней оценкой качества вина. С увеличением доли алкоголя наблюдается тенденция роста среднего значения показателя качества, несмотря на наличие локальных колебаний.

Для детального анализа характеристик вина создана таблица(рис.10), включающая основные физико-химические показатели, такие как кислотность, содержание сахара, плотность, уровень алкоголя, а также экспертную оценку качества. Таблица поддерживает сортировку по выбранным показателям, что позволяет оперативно выявлять образцы с наибольшими и наименьшими значениями качества и анализировать их характеристики.

Тип вина	alcohol	fixed acidity	citric acid	sulphates	pH	volatile acidity	residual sugar	density	quality
Белое	8,00	4,20	0,23	0,44	3,42	0,21	5,10	1,00	3
Белое	8,00	4,50	0,21	0,42	3,34	0,19	0,95	0,99	5
Белое	8,40	6,20	0,23	0,48	3,16	0,31	3,30	0,99	5
Белое	8,40	6,40	0,12	0,46	3,06	0,37	5,90	1,00	4
Белое	8,50	6,10	0,35	0,50	3,37	0,43	9,10	1,00	5
Белое	8,50	6,30	0,30	0,40	3,47	0,20	5,90	1,00	6
Белое	8,50	6,40	0,48	0,61	2,98	0,18	18,00	1,00	6
Белое	8,50	6,40	0,49	0,48	3,14	0,37	13,30	1,00	6
Белое	8,50	6,50	0,16	0,39	3,19	0,36	1,30	0,99	5
Белое	8,50	6,50	0,48	0,61	2,98	0,18	18,00	1,00	6
Белое	8,50	7,60	0,49	0,44	3,01	0,28	20,15	1,00	5
Белое	8,50	7,80	0,46	0,76	3,24	0,18	12,60	1,00	8
Белое	8,50	9,10	0,38	0,38	3,23	0,59	1,60	1,00	3
Белое	8,60	6,00	0,29	0,45	3,39	0,21	13,10	1,00	5
Белое	8,60	6,10	0,31	0,48	3,16	0,34	12,00	1,00	5
Белое	8,60	6,30	0,30	0,43	3,05	0,28	3,10	0,99	5
Белое	8,60	6,50	0,19	0,35	3,25	0,27	4,20	1,00	4
Белое	8,60	6,50	0,33	0,49	3,25	0,14	7,60	1,00	5
Белое	8,60	6,80	0,20	0,45	3,13	0,44	16,00	1,00	5
Белое	8,60	6,80	0,24	0,46	3,03	0,22	4,90	1,00	6
Белое	8,60	6,80	0,28	0,38	3,03	0,27	13,30	1,00	6
Белое	8,60	6,90	0,49	0,47	2,98	0,27	23,50	1,00	5
Белое	8,60	7,20	0,62	0,49	3,08	0,40	10,80	1,00	4
Белое	8,60	7,30	0,39	0,36	3,21	0,24	17,95	1,00	5
Белое	8,60	7,30	0,54	0,49	3,06	0,36	13,30	1,00	4
Белое	8,60	7,30	0,69	0,45	3,06	0,31	10,20	1,00	5
Белое	8,60	7,70	0,58	0,49	3,06	0,34	11,10	1,00	5

Рис. 10 – Детальная таблица характеристик вина

Для итогового аналитического представления в левом навигационном меню сервиса выбран пункт «Дашборды → Создать дашборд». В качестве рабочей области выбран соответствующий воркбук, после чего создан новый дашборд(рис.11). На дашборд последовательно добавлены все ранее созданные чарты, включая индикаторы, столбчатые и линейные диаграммы, а также детальную таблицу.

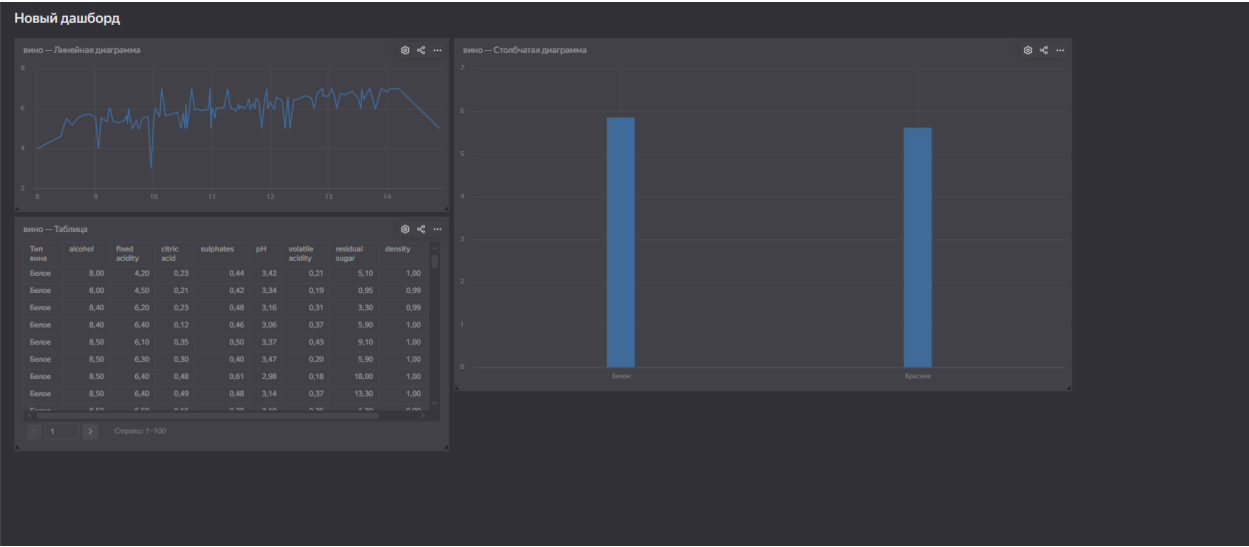


Рис. 11 – Компоновка элементов дашборда

На дашборде настроены интерактивные селекторы(рис.12), обеспечивающие возможность динамического изменения представления данных. В качестве полей фильтрации использовались тип вина, категория качества и уровень алкоголя. При изменении значений селекторов автоматически обновлялись все визуальные элементы дашборда, что позволило оперативно анализировать данные в различных разрезах и выявлять зависимости между характеристиками вина и его качеством.

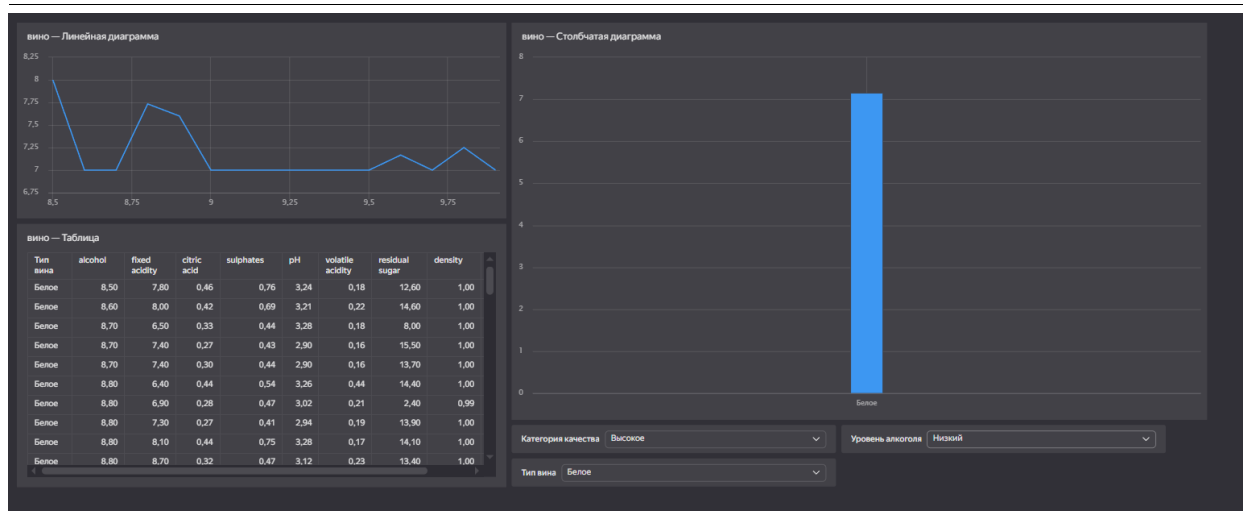


Рис. 12 – Добавление фильтров на дашборд

4 Выводы

В результате работы разработана универсальная методика построения интерактивного аналитического дашборда для анализа качества вина в среде Yandex DataLens. Использование открытых датасетов и low-code инструментов позволило наглядно продемонстрировать влияние физико-химических характеристик на оценку качества вина. Предложенный подход может быть адаптирован для анализа данных в других предметных областях и использован в учебных целях при изучении визуализации и аналитики данных.

Библиографический список

1. Акентьев Д. Д. Разведочный анализ в DataLens // Постулат. 2024. № 6(104). С. 69. URL: <https://elibrary.ru/item.asp?id=80665923>
2. Полянская П. А. Типы данных в BI-инструменте Yandex DataLens // Студенческий. 2024. № 42-2(296). С. 39–40. URL: <https://elibrary.ru/item.asp?id=78127334>
3. Sahipov I., Zabashta A., Filchenkov A. Stabilization of dataset matrix form for classification dataset generation and algorithm selection // Lecture Notes in Computer Science. 2020. Vol. 12490. P. 66–75. URL: <https://elibrary.ru/item.asp?id=45099682>
4. Бородин А. В. Использование BI-приложения для анализа рынка кино // Вестник науки. 2024. Т. 4. № 6(75). С. 1080–1087 URL: <https://elibrary.ru/item.asp?id=67889103>
5. Плющ Н. О., Федькова Н. А. Проблемы и перспективы Yandex DataLens на рынке бизнес-аналитики // Вестник образовательного консорциума «Среднерусский университет». Информационные технологии. 2022. № 2(20). С. 22–26. URL: <https://elibrary.ru/item.asp?id=50036404>
6. Wine Quality Dataset URL: <https://archive.ics.uci.edu/dataset/186/wine+quality>