

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение высшего образования

**«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

ФГБОУ ВО «ИГУ»

**Институт математики и информационных технологий**

Кафедра алгебраических и информационных систем

**ОТЧЕТ ПО УЧЕБНОЙ ПРАКТИКЕ**

**Научно-исследовательская работа**

**(получение первичных навыков**

**научно-исследовательской работы)**

Руководитель:

Начальник отдела анализа данных и машинного обучения

ЭН+ ДИДЖИТАЛ ООО

Посохов А.Ю.

Студент:

Непокрытых Игорь Владимирович

гр. 02371-ДБ, 3 курс ФИ

Иркутск 2025

**ДНЕВНИК ПРОХОЖДЕНИЯ УЧЕБНОЙ ПРАКТИКИ**

|  |  |  |
| --- | --- | --- |
| Дата | Краткое содержание работы | Отметка о выполнении, |
| **26.05.25** | Получение первичного знакомства с рабочим отделом, местом работы и выполняемыми в нем задачами. | Выполнено |
| **27.05.25**  **по**  **28.05.25** | Получение задания, знакомство с программным обеспечением используемым в компании для выполнения задач аналитиков (Power BI). Знакомство с системой безопасности, получение доступа к необходимым ресурсам. | Выполнено |
| **29.05.25**  **по**  **30.05.25** | Разработка плана выполнения задания, уточнение деталей. Первое знакомство с разрабатываемой системой отслеживания данных | Выполнено |
| **02.06.25** | Изучение построения связей между таблицами данных.  Написание первых мер, исчислений и изучение работы встроенных функций даты программного обеспечения. Внесение изменений в исчисляемые поля для получения корректных результатов вычислений. Знакомство с языком запросов DAX | Выполнено |
| **03.06.25**  **по**  **04.06.25** | Выполнение первого пункта поставленного плана . Работа с датами, данными и связывание их для получения необходимых полей для выполнения следующих шагов. Также, начало ежедневных совещаний контроля выполнения поставленной задачи. | Выполнено |
| **05.06.25**  **по**  **06.06.25** | Связка всех графиков и полей с внесёнными изменениями  для корректного их отображения. Написание итогового отчета по практике. | Выполнено |
| **07.06.25** | Подготовка итогового отчёта по результатам практики. | Выполнено |

**Оглавление**

[1. Введение 4](#_Toc23423)

[2. Исследование предметной области 4](#_Toc28244)

[2.1 Введение в предметную область 4](#_Toc24515)

[2.2 Основные понятия и термины 5](#_Toc32542)

[2.3 Анализ текущего состояния предметной области 6](#_Toc12742)

[2.4 Тенденции и перспективы развития 7](#_Toc25744)

[2.5 Значение исследования предметной области для практики 7](#_Toc13602)

[3. Цели и Задачи 8](#_Toc27920)

[4. Выполненная работа 8](#_Toc23304)

[4. Анализ Результатов 13](#_Toc22475)

[5. Заключение 14](#_Toc13740)

[6. Список используемой Литературы 15](#_Toc26361)

[Приложение 1 17](#_Toc23543)

[Приложение 2 20](#_Toc25920)

[Приложение 3 21](#_Toc16244)

1. Введение

Data Science — это очень важное в настоящее время направление, которе необходимо во время быстрорастущей индустрии машинного обучения и искуственного интелекта. Data Scientist занимается извлечением полезной информации из больших массивов данных и созданием моделей, которые могут предсказывать события или оптимизировать процессы, сейчас практически все компании в целях оптимизации рабочего процесса прибегают к внедрению элементов машинного обучения, что является невозможной задачей без специалиста по данным. Главной задачей на практике являлось обучение новым навыкам и получение опыта работы с большим объёмом данных, а также изучение новых программных обеспечений в данной области.

1. Исследование предметной области

## 2.1 Введение в предметную область

Предметная область заключалась в разработке дашборда

Работа включает несколько ключевых этапов:

* Сбор данных. Это начальный этап, когда важно собрать данные из различных источников (базы данных, API, внешние сервисы)
* Очистка и обработка данных. Зачастую данные бывают грязными, неполными или неструктурированными поэтому одной из ключевых задач является их очищение и подготовка для анализа
* Анализ данных. На этом этапе специалист использует статистические методы, алгоритмы машинного обучения или визуализацию, чтобы изучить закономерности и тренды в данных
* Моделирование и прогнозирование. На основе анализа, строят модели, которые могут предсказать будущие события или закономерности, влияния
* Интерпретация и представление результатов. Специалисты должны уметь преобразовать результаты анализа в понятные бизнес-решения. Это включает в себя составление отчётов, построение графиков и презентаций для руководства
* Оптимизация процессов. На основе выведенных тенденций и закономерностей представляются способы решения или контроля различных проблем производства

## 2.2 Основные понятия и термины

В ходе выполнения практических задач были освоены следующие ключевые концепции и терминология:

1. Business Intelligence (BI)  
   Информационно-аналитическая система для сбора, обработки и визуализации данных с целью поддержки бизнес-решений. Включает инструменты для преобразования сырых данных в значимую информацию.
2. ETL (Extract, Transform, Load)  
   Процесс извлечения данных из источников, их трансформации (очистка, форматирование) и загрузки в целевую систему (на примере Power Query в Power BI).
3. Data Modeling  
   Построение структуры данных через создание таблиц и установление связей между ними.
4. DAX (Data Analysis Expressions)  
   Формульный язык для создания вычислений и мер в Power BI. Ключевые элементы:
   1. Меры (Measures) - динамические расчеты на основе контекста фильтра
   2. Вычисляемые столбцы (Calculated Columns) - статические вычисления на уровне строк
   3. Функции времени (Time Intelligence) - YTD, MTD, SAMEPERIODLASTYEAR
   4. Визуализация данных  
      Принципы представления информации через:
   5. Срезы (Slicers) - интерактивные фильтры отчетов
   6. Динамические диапазоны дат - пользовательские фильтры временных периодов
   7. Перекрестная фильтрация (Cross-filtering) - взаимодействие элементов визуализации
   8. Ключевые метрики  
      Основные показатели для анализа:
      1. KPI (Key Performance Indicator) - ключевые показатели эффективности
      2. YoY (Year-over-Year) - годовой рост показателей
      3. MoM (Month-over-Month) - месячная динамика
5. Системные термины
   1. Data Pipeline - процесс автоматизации потока данных
   2. Data Warehouse - централизованное хранилище данных
   3. Data Mart - тематическое подмножество хранилища
6. Процессные термины
   1. Спринт (Sprint) - период разработки (в контексте ежедневных совещаний)
   2. Требования (Requirements) - технические условия к решениям
   3. Документирование ошибок (Bug Tracking) - процесс регистрации неисправностей.

2.3 Анализ текущего состояния предметной области   
На данный момент в компании, программное обеспечение Microsoft Power bi начали использовать только в начале 2025 года и до него аналитики пользовались другим инструментом, а именно Apache Superset. Поэтому требуется время, ресурсы и люди для развития этого направления.

Если смотреть общую ситуацию в России, то на наших площадках заметно малое количество специалистов в работе с данным программным продуктом, так как найти информацию можно в основном только на англоязычных источниках, а все видеоуроки или инструкции либо же на иностранном языке, либо в недостаточном количестве и без глубокой реализации, что делало знакомство с программой весьма затруднительным и трудоёмким занятием.

В мире Power bi занимает 35% рынка Bi-иснтрументов, в последнее время растет спрос на мобильные дашборды, данные для которых можно постоянно обновлять и видеть самую последнюю статистику, в отличии от старых подходов один график - один промежуток. В связи с статистикой в мире явно видно, что именно этот продукт доминирует на рынке, и его малая распространённость в российском сегменте объясняется только тем, что в данный момент идёт сильный упор на отечественных производителей, в такой развивающейся области пытаются укрепиться такие гиганты российского рынка как яндекс со своим “Yandex DataLens”. Но пока на данный момент, программное обеспечение Microsoft остается самым производительным.

## 2.4 Тенденции и перспективы развития

Главная и большая задача состоит в будущем создавать автоматически генерируемые дашборды, так как Power bi позволяет прикреплять к модели данных свои модели искусственного интеллекта, что конечно ускорит работу с данными и уменьшит занятость аналитиков. Помимо глобальной задачи также стоит автоматизация контроля на дочерних предприятиях компании, в чем собственного и заключалась задача практики, для компании это постоянный мониторинг показателей и в следствии выявление тенденций и сокращения расходов топлива (как в моем примере). По прогнозам уже к 2026 году 80% всех мировых дашбордов будут генерироваться нейронными сетями.

## 2.5 Значение исследования предметной области для практики

Исследование предметной области (разработка BI-решений и аналитика данных) стало важным фундаментом для успешного выполнения практических задач. Его значение проявилось в следующих аспектах:

1. Осознание роли мер (Measures) в аналитике, осмысленное изучение DAX
2. Знание принципов data storytelling - создание интуитивных визуализаций вместо случайных графиков
3. Знание профессиональной терминологии аналитиков данных позволило участовать в собраниях аналитиков и понимать как вещи о которых они говрят, так и проблемы с которыми они сталкивались и какими способами их решали.
4. Анализ отраслевых проблем, таких как выбросы в данных позволил понять, какие ошибки есть в построении графиков при фильтрации.
5. Осознание логической последовательности построения дашбордов.
6. Правильно понимать, какие требования должны быть применены к моим фильтрам и как сделать их так, чтобы они не только не повлияли на исходные данные, но и правильно делали срезы для остальных визуализаций.
7. Цели и Задачи

В процессе прохождения практики были поставлены следующие цели:

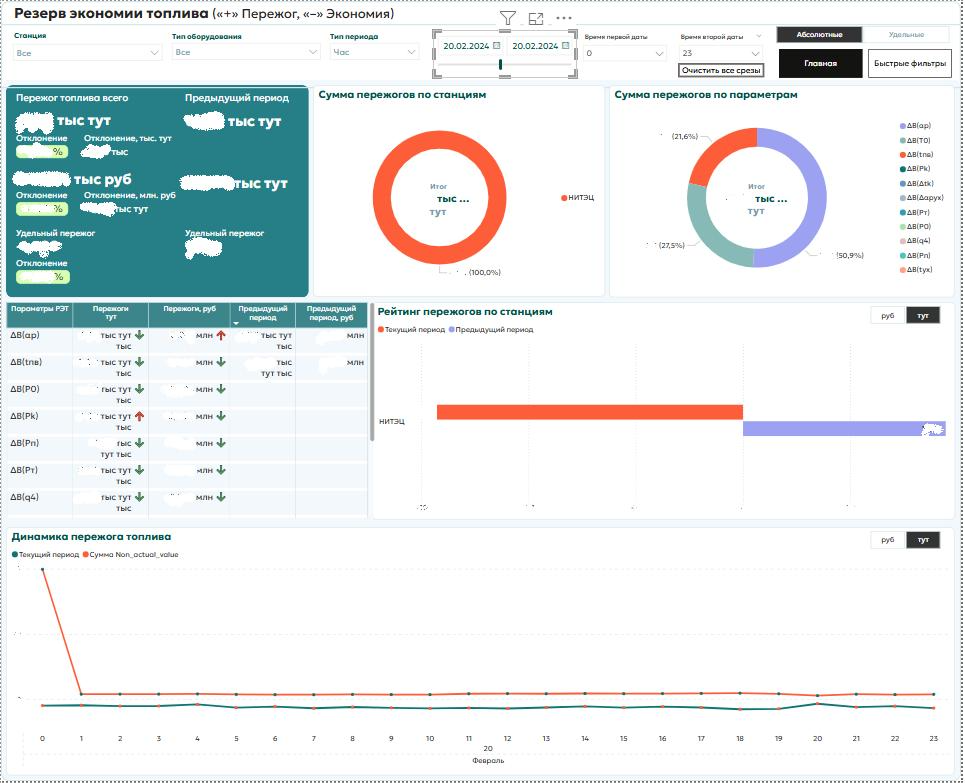
* Приобретение практического опыта. Освоение профессиональных навыков бизнес-аналитика в реальной среде компании
* Формирование компетенций в работе с Bi инструментами
* Решение настоящей бизнес проблемы

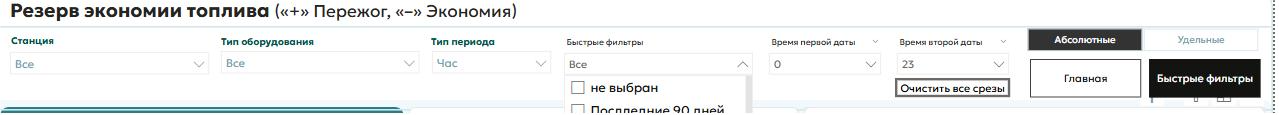
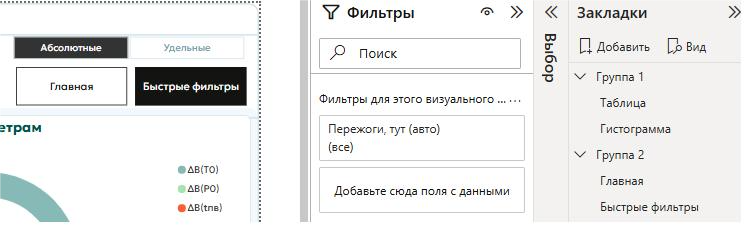
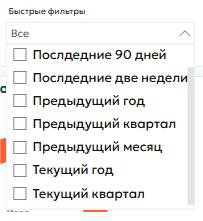
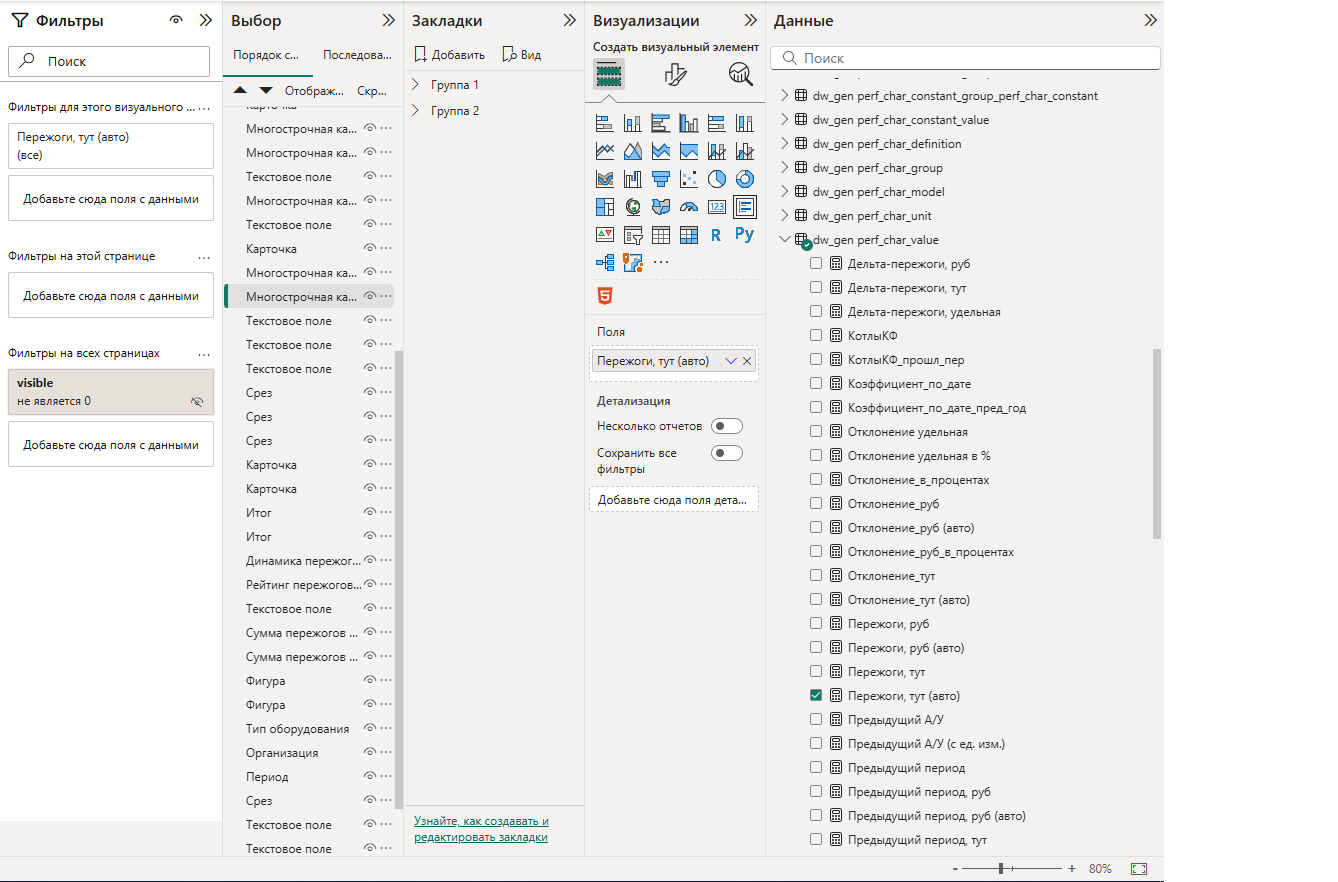
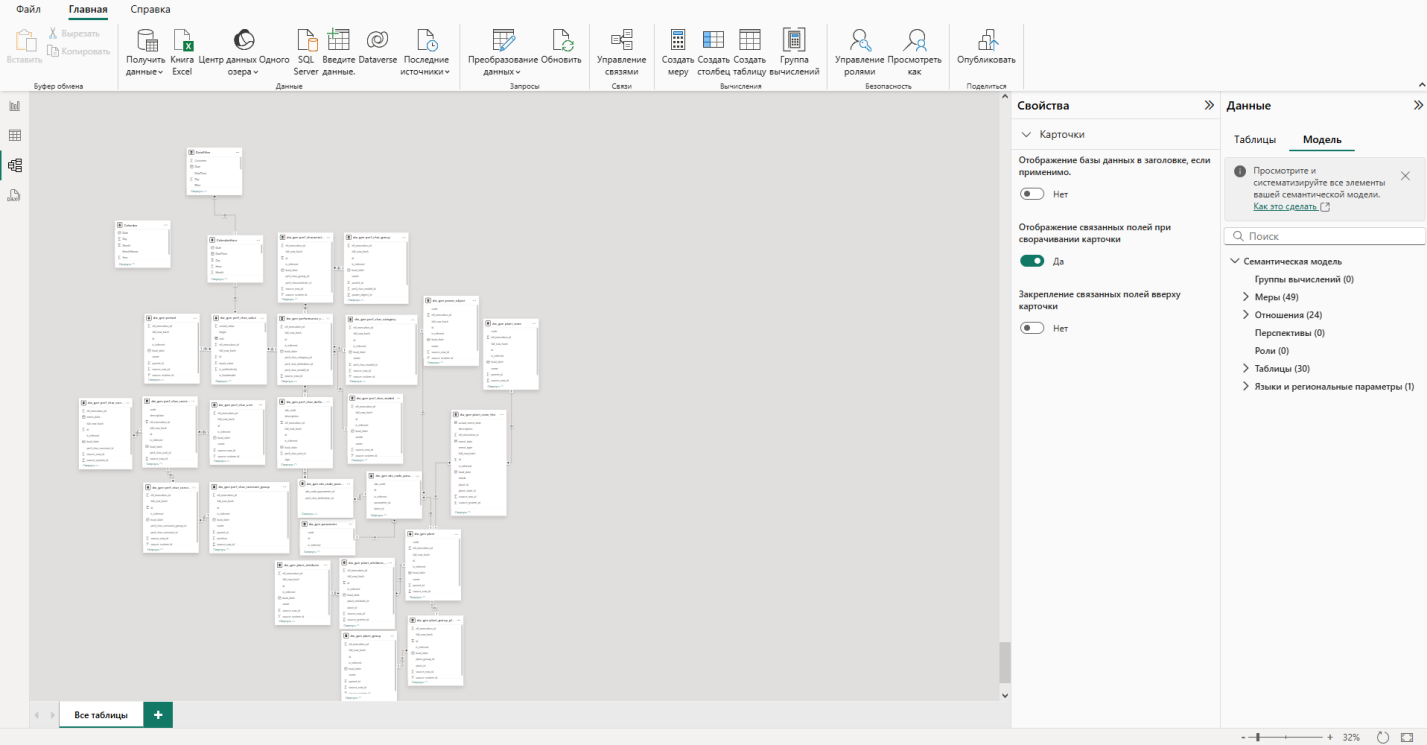
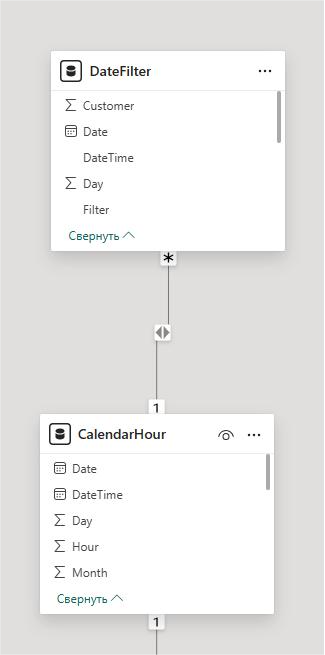
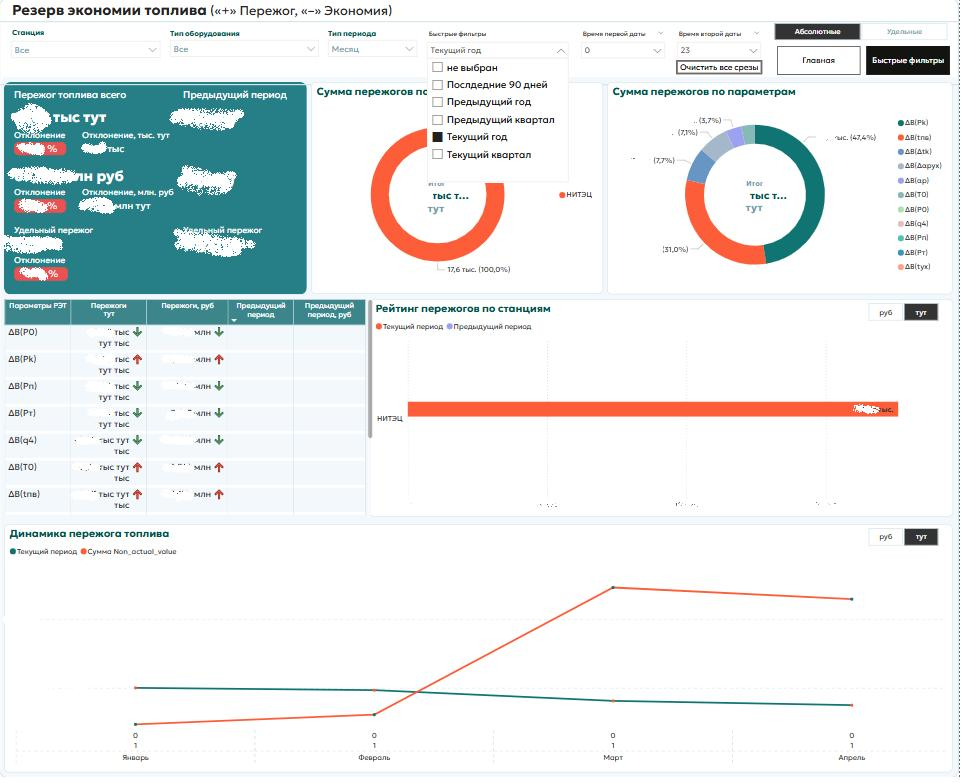
К задачам относились следующие пункты:

* Детализация отчета до почасового уровня
* Динамический пересчет показателей в выбранном интервале
* Создать набор предустановленных периодов
* Настройка корректного отображения всех визуализаций
* Оформление возвращаемых типов данных в том виде, в котором требует Power bi для их отображения.
* Синхронизация фильтров, чтобы календарный срез не конфликтовал с реализованными предопределёнными фильтрами и реализованным часовым селектором
* Настройка перекрестных фильтров между визуализациями
* Оптимизация производительности путём добавления расчётных таблиц и вычисления DAX запросов через подходящие функции
* Тестирование. А именно сравнение получившихся результатов с фактическими
* Выявление аномалий на часовом уровне (в некоторых моментах оборудование дает неверные результаты)
* И главная это изучение динамики пережога топлива оборудованием (пережог - излишний расход топлива при той же выработке электроэнергии)

1. Выполненная работа

Далее представлена главная итоговая страница построенного дашборда. На ней в целях безопасности данных спрятаны все отображаемые показатели :



* В первую очередь была переработана главная таблица дат модели, для работы с почасовыми данными. И добавлены два среза “Время первой даты” и ”Время второй даты” они отвечают за установление часа окончания временного интервала, первый как начальный час отчета, откуда мы смотрим статистику, второй как час дня до которого мы смотрим данные, также к примеру можно выбрать не только один день на рассмотрение, а также и промежуток с разницей в два дня к примеру 01.01.2024 2:00:00 до 03.01.2024 13:00:00, что естественно является необходимой возможностью в некоторых ситуациях.
* Была добавлена возможность переключения пользователя между двумя закладками “Главная страница” и ”быстрые фильтры”, на главной странице средством выбора интервала является календарь с выпадающими полями для удобного переключения пользователя на необходимый интервал, на странице быстрых фильтров есть возможность выбирать предопределённые периоды по типу : последние 90 дней, текущая неделя, текущий месяц, предыдущий месяц, текущий квартал, предыдущий квартал, последние две недели. И оба эти среза должны корректно взаимодействовать с фильтрацией по часам. Отсюда и переходим к следующему шагу разработки:  
    
  Закладки:  
    
    
  Интерфейс создания закладок:  
  
* Так выглядит выпадающий список предопределённых периодов  
    
  Сначала был создан его внешний вид, следующим шагом надо было написать DAX код для его работы
* Написание DAX запросов для таблицы быстрых фильров:  
  (Приложение 1), также была добавлена колонка customer, которая отвечает за отображение и скрытие некоторых элементов интерфейса
* Были добавлены некоторые меры в фактовую таблицу “dw\_gen perf\_char\_value”, для нового способа расчета показателей за новые интервалы с временем, это далеко не весь их список.
* Далее представлен вид связей всех необходимых таблиц и также новосозданных в процессе решения задачи:  
  Здесь мы можем увидеть количество созданных мер, таблиц и отношений
* Также здесь мы устанавливаем логику взаимодействия созданной нами талицы DateFilter для её взаимодействия со всеми визуальными элементами дашборда. Здесь они были связаны через поля DateFilter связью многие к одному, так как DateFilter содержит в себе повторяющиеся значения даты  
  
* Для корректной передачи данных в визуальные элементы была написана мера получения данных за предыдущий период (Предыдущий период использую все визуализации кроме круговых диаграмм)  
  (Приложение 3)
* Теперь проверяем работу созданных нами быстрых фильтров  
    
    
  на данном снимке экрана видно, что у нас правильно установился временной интервал для текущего 2025 года
* Также были созданы меры для формирования даты в таблице дат:  
  (Приложение 2)

4. Анализ Результатов

Анализ Результатов

В ходе прохождения практики были развиты и значительно улучшены следующие ключевые навыки и достигнуты важные результаты:

* Углубленное владение Microsoft Power BI:
* Освоены продвинутые методы моделирования данных
* Разработка специализированных DAX-мер для динамических расчетов показателей в почасовых интервалах
* Реализован механизм двухуровневой фильтрации (календарь + быстрые фильтры) с синхронизацией параметров
* Создан интуитивный интерфейс с переключением между режимами анализа ("Главная страница"/"Быстрые фильтры")
* Внедрена система предустановленных периодов,которая позволит сократить время анализа
* Обеспечена кросс-фильтрация визуализаций на единой панели
* Оптимизация данных и производительности:
* Реализована почасовая детализация данных через модификацию таблицы CalendarHour
* Созданы расчетные таблицы для ускорения обработки запросов
* Разработаны механизмы валидации данных для выявления аномалий оборудования
* Решение бизнес-задач:
* Внедрен инструмент мониторинга пережога топлива с почасовой детализацией
* Обеспечена возможность сравнения показателей с предыдущими периодами
* Реализация 100% поставленных задач в установленные сроки

1. Заключение

Практика в компании "ЭН+ ДИДЖИТАЛ" стала важным этапом профессионального становления в области бизнес-аналитики. За время работы:

Достигнуты ключевые цели:

* Приобретен практический опыт разработки BI-решений в энергетической отрасли
* Освоен полный цикл создания аналитического дашборда: от проектирования до внедрения
* Решена актуальная бизнес-задача по мониторингу эффективности оборудования

Развиты профессиональные компетенции:

* Углубленное владение Power BI (моделирование данных, DAX, визуализация)
* Навыки проектирования пользовательских интерфейсов для аналитических систем
* Понимание специфики работы с временными рядами и производственными показателями

Получены значимые результаты:

* Внедренный инструмент анализа позволил выявить резервы для оптимизации расходов топлива

6. Список используемой Литературы

1. Документация Microsoft Power Bi  
   URL: <https://learn.microsoft.com/ru-ru/dax/addcolumns-function-dax> (Дата обращения: 07.06.2025)
2. Феррари А., Руссо М.  
   The Definitive Guide to DAX: Business intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel— 2nd ed. — Microsoft Press, 2023. — 768 с.  
    ISBN: 978-0137852686
3. Макдональд Б.  
   Power BI изнутри: Проектирование систем бизнес-аналитики— СПб.: Питер, 2024. — 512 с.I  
   SBN: 978-5-4461-2045-3

Приложение 1

DateFilter =

VAR \_today = TODAY()

RETURN

UNION(

ADDCOLUMNS(

FILTER(

CalendarHour,CalendarHour[Date] >= \_today - 7 && CalendarHour[Date] <= \_today

),

"Filter","текущая неделя",

"Customer", 1

),

ADDCOLUMNS(

FILTER(

CalendarHour,CalendarHour[Date] >= \_today - 14 && CalendarHour[Date] <= \_today

),

"Filter","Послдедние две недели",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour,CalendarHour[Date] >= \_today - 90 && CalendarHour[Date] <= \_today

),

"Filter","Послдедние 90 дней",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour,CalendarHour[Year] == YEAR(\_today) && CalendarHour[Date] >= DATE(YEAR(\_today),1,1) && CalendarHour[Date] <= \_today

),

"Filter","Текущий год",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour, CalendarHour[Date] >= DATE(YEAR(\_today)-1,1,1) && CalendarHour[Date] <= DATE(YEAR(\_today)-1,12,31)

),

"Filter","Предыдущий год",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour,CalendarHour[Date] >= ( DATE(YEAR(\_today),MONTH(\_today),1) ) && CalendarHour[Date] <= \_today

),

"Filter","Текущий месяц",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour,CalendarHour[Date] >= DATE(IF(MONTH(\_today)==1,YEAR(\_today)-1,YEAR(\_today)),MONTH(\_today)-1,1) && CalendarHour[Date] <= DATE(IF(MONTH(\_today)==1,YEAR(\_today)-1,YEAR(\_today)),MONTH(\_today)-1,DAY(EOMONTH(DATE(IF(MONTH(\_today)==1,YEAR(\_today)-1,YEAR(\_today)),MONTH(\_today)-1,1),0)))

),

"Filter","Предыдущий месяц",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour, CalendarHour[Date] >= IF(QUARTER(\_today)==1,DATE(YEAR(\_today)-1,10,1),IF(QUARTER(\_today)==2,DATE(YEAR(\_today),1,1),IF(QUARTER(\_today)==3,DATE(YEAR(\_today),4,1),DATE(YEAR(\_today),7,1)))) && CalendarHour[Date] <= IF(QUARTER(\_today)==1,DATE(YEAR(\_today)-1,12,31),IF(QUARTER(\_today)==2,DATE(YEAR(\_today),3,31),IF(QUARTER(\_today)==3,DATE(YEAR(\_today),6,31),DATE(YEAR(\_today),9,31))))

),

"Filter","Предыдущий квартал",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour, CalendarHour[Date] >= IF(QUARTER(\_today)==1,DATE(YEAR(\_today),1,1),IF(QUARTER(\_today)==2,DATE(YEAR(\_today),4,1),IF(QUARTER(\_toay)==3,DATE(YEAR(\_today),7,1),DATE(YEAR(\_today),10,1)))) && CalendarHour[Date] <= \_today

),

"Filter","Текущий квартал",

"Customer",1

),

ADDCOLUMNS(

FILTER(

CalendarHour, CalendarHour[Date]),

"Filter","не выбран",

"Customer", 0

)

)

Приложение 2

Текущая\_Дата = NOW()

EndDateTime = [MaxSelectedDate] + TIME([SelectedHourEnd],0,0)

IsInSelectedRange =

VAR CurrentDT = SELECTEDVALUE(CalendarHour[DateTime])

RETURN

IF(

CurrentDT >= [StartDateTime]

&&

CurrentDT <= [EndDateTime],

1,

0

)

MaxSelectedDate = CALCULATE(MAX(CalendarHour[Date]),ALLSELECTED(CalendarHour))

MinSelectedDate = CALCULATE(MIN(CalendarHour[Date]),ALLSELECTED(CalendarHour))

SelectedHourEnd = SELECTEDVALUE(RangeHourEnd[HoursEnd],0)

SelectedHourStart = SELECTEDVALUE(RangeHourStart[HourStart],0)

StartDateTime = [MinSelectedDate] + TIME([SelectedHourStart],0,0)

RangeHourEnd =

ADDCOLUMNS(

GENERATESERIES(0,23,1),

"HoursEnd",[Value]

)

RangeHourStart =

ADDCOLUMNS(

GENERATESERIES(0,23,1),

"HourStart",[Value]

)

Приложение 3

test =

VAR CurrentDates =

VALUES ( CalendarHour[Date] ) -- точки текущего контекста

VAR PrevYearDates =

SELECTCOLUMNS (

CurrentDates,

"Date", DATEADD (CalendarHour[Date], -1, YEAR )

)

/\* корректно берём границы из таблицы-переменной \*/

VAR MinDate = [StartDateTime] -- нижняя граница

VAR MaxDate = [EndDateTime] -- верхняя граница

VAR MinDateLy = DATE(YEAR(MinDate)-1,MONTH(MinDate),DAY(MinDate)) + TIME(HOUR(MinDate),MINUTE(MinDate),SECOND(MinDate))

VAR MaxDateLy = DATE(YEAR(MaxDate)-1,MONTH(MaxDate),DAY(MaxDate)) + TIME(HOUR(MaxDate),MINUTE(MaxDate),SECOND(MaxDate))

/\* диапазон прошлого года \*/// VAR per = DATESBETWEEN ( CalendarHour[Date], DATE(YEAR(MinDate) - 1,MONTH(MinDate),YEAR(MinDate)) , DATE(YEAR(MaxDate) - 1,MONTH(MaxDate),YEAR(MaxDate)) )

// VAR MinDateLy = MINX ( PrevYearDates, [Date] ) + TIME(HOUR(MinDate),MINUTE(MinDate),SECOND(MinDate)) -- нижняя граница

// VAR MaxDateLy = MAXX ( PrevYearDates, [Date] ) + TIME(HOUR(MaxDate),MINUTE(MaxDate),SECOND(MaxDate)) -- верхняя граница

RETURN

IF ( isblank ( MinDate) , BLANK(),

CALCULATE (

MIN( 'dw\_gen perf\_char\_value'[actual\_value] ),

//REMOVEFILTERS ( 'CalendarHour' ),

REMOVEFILTERS ( 'DateFilter' ),

FILTER(

ALL(CalendarHour[DateTime]), CalendarHour[DateTime] >= MinDateLy && CalendarHour[DateTime] <= MaxDateLy

),

/\* фильтр характеристик — как был \*/

TREATAS (

VALUES ( 'dw\_gen aks\_code\_param\_perf\_char\_def'[perf\_char\_definition\_id] ),

'dw\_gen performance\_characteristic'[perf\_char\_definition\_id]

)

)

)

DAX Создание столбца:

Non\_actual\_value =

VAR CurrentDate = 'dw\_gen perf\_char\_value'[begin]

VAR dat = DATE(YEAR('dw\_gen perf\_char\_value'[begin])-1,MONTH('dw\_gen perf\_char\_value'[begin]),DAY('dw\_gen perf\_char\_value'[begin])) + TIME(HOUR('dw\_gen perf\_char\_value'[begin]),MINUTE('dw\_gen perf\_char\_value'[begin]),SECOND('dw\_gen perf\_char\_value'[begin]))

VAR CurrentId = 'dw\_gen perf\_char\_value'[performance\_characteristic\_id]

RETURN

CALCULATE(

SUM('dw\_gen perf\_char\_value'[actual\_value]),

REMOVEFILTERS('dw\_gen perf\_char\_value'),

'dw\_gen perf\_char\_value'[begin] = dat,

'dw\_gen perf\_char\_value'[performance\_characteristic\_id] = CurrentId

)