Санкт-Петербургский Государственный Университет

Калинина Екатерина Юрьевна

**АВТОМАТИЗАЦИЯ ПРОЦЕДУР ОБРАБОТКИ НЕСТРУКТУРИРОВАННЫХ МАССИВОВ ДАННЫХ В ЦЕЛЯХ МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЙ РЫНКА СУДОСТРОЕНИЯ**

Диссертация на соискание степени магистра

Направление: Бизнес-информатика

Магистерская программа: Информационная бизнес-аналитика

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись диссертанта

Научный руководитель :

Профессор кандидат технических наук,

Ботвин Геннадий Алексеевич \_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись научного руководителя

Санкт-Петербург

2017

Оглавление

1. [ВВЕДЕНИЕ 3](#_Toc482658833)
2. [АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ И ХАРАКТЕРИСТИКА ОБЪЕКТА ИССЛЕДОВАНИЯ 7](#_Toc482658834)
   1. [Общая информация об ИА ИНФОЛайн 7](#_Toc482658835)
   2. [Информация о готовом продукте ИА ИНФОЛайн 7](#_Toc482658836)
   3. [Характеристика отрасли 10](#_Toc482658837)
3. [ПОДГОТОВКА ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ 13](#_Toc482658838)
   1. [Консолидация данных 13](#_Toc482658839)
   2. [Очистка данных 15](#_Toc482658840)
4. [ОБОСНОВАНИЕ ВЫБОРА BI-СИСТЕМЫ 20](#_Toc482658841)
5. [РАЗРАБОТКА АЛГОРИТМА АВТОМАТИЗАЦИИ ПРОЦЕДУР ОБРАБОТКИ НЕСТРУКТУРИРОВАННЫХ МАССИВОВ ДАННЫХ 27](#_Toc482658842)
   1. [Определение бизнес-правил 27](#_Toc482658843)
   2. [Моделирование данных 28](#_Toc482658844)
   3. [Разработка приложения автоматизации процедур обработки неструктурированных массивов данных 29](#_Toc482658845)
      1. [Загрузочный скрипт 29](#_Toc482658846)
      2. [Пример решения сформулированных задач исследования в BI-системе «QlikView» 31](#_Toc482658847)
      3. [Краткое руководство для пользовательской работы с приложением в BI-системе 38](#_Toc482658848)
6. [ЗАКЛЮЧЕНИЕ 40](#_Toc482658849)
7. [СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 41](#_Toc482658850)
8. [Приложение 1. Код скрипта загрузки 43](#_Toc482658851)
9. [Приложение 2. Код макроса изменения цвета ярлыка с помощью VB макроса 44](#_Toc482658852)
10. [Приложение 3. Дашборды приложения 45](#_Toc482658853)
11. [Приложение 4. Структура анкеты 47](#_Toc482658854)

**ВВЕДЕНИЕ**

***Актуальность***

На современном этапе развития общества информационные технологии являются неотъемлемой частью в различных сферах человеческой деятельности: предпринимательская деятельность и государственное управление. С увеличением объемом данных, с которым общество сталкивается как в повседневной жизни, так и в профессиональной деятельности, очень важно правильно, грамотно и экономно обращаться с информацией и ее различными представлениями. Знание возможностей получения, анализа и представления данных, различной информации позволяет грамотно и экономно распорядиться имеющимися ресурсами, оптимизировать различные затраты и сэкономить самый ценный ресурс общества – время. Темпы развития современного мира находят свое отражение не только в различных сферах деятельности, применяемых технологиях и методах, но и предъявляют к ним определенные требования. Управление любой деятельностью, ее поддержка, сегодня неразрывно связаны с информационными системами, которые специалисты пытаются создавать с учетом возможных изменений в различных сферах.

В различных сферах, как бизнеса, так и государственного управления можно наблюдать различные подходы к автоматизации основных процессов деятельности. Часто при сложной структуре устройства предприятия отдельные подразделения решают свои задачи и цели различными путями, в рамках своих сил и возможностей. Зачастую возникают проблемы организации взаимосвязи между системами, обусловленные, как правило, разным временем внедрения систем, различными изначальными требованиями. В конечном итоге все это сказывается на уровне систематизации информации, прозрачности схемы работы организации, возможностях интеграции процедур обработки, хранения и представления данных и т.д.

Неструктурированные данные — это на сегодняшний день еще не до конца оцененный бизнес-актив, эффект от использования которого будет тем выше, чем совершеннее и «тоньше» будут соответствующие программные инструменты. По оценкам экспертов, около 80% внутрикорпоративного информационного контента имеют неструктурированный или частично структурированный вид — это файлы различных форматов (фото, аудио и видео, электронная почта), несущие в себе огромный потенциал для бизнес-анализа. Мало того, наиболее продвинутые компании начинают «сканировать» внешнее информационное поле в поисках дополнительных актуальных знаний о конъюнктуре рынка и поведении своего сегмента потребителей, надеясь найти новые решения для развития своего бизнеса. Однако основной массив данных, генерируемых существующими на планете организационными структурами (предприятиями, компаниями, семьями, группами по интересам и т. д.), остается вне фокуса внимания аналитиков — подобная информация не подпадает под критерии области анализа, отсутствует инструмент обработки, а поток данных избыточен.

Рынок продолжает расти вслед за ростом количества различных информационных систем, которые используют заказчики, и объема данных, обрабатываемых в этих системах. Постепенное накопление данных является драйвером спроса на BI-инструменты. И одним из важных драйверов, о котором рассказывают многие участники рынка, является накопление компаниями большого объема данных. В этом случае у бизнес-игроков появляется желание получить из них добавленную стоимость или дополнительную ценность.

В рамках обозначенной темы можно выделить деятельность ИА «ИНФОЛайн-Аналитика», которая состоит в подготовке ежегодных отраслевых обзоров по направлению строительство и индустриальные рынки, а также в области судостроения. Проблемой составления таких обзоров является разрозненность данных, которая существенно затрудняет процесс исследования. На решение этой проблемы уходят недели работы с данными, что очевидным образом свидетельствует об актуальности процесса автоматизации по их обработке. Данная проблема стала определяющим фактором при выборе темы исследовательской работы.

***Цель исследования***

Разработать алгоритм автоматизации обработки, хранения и предоставления изначально разнородных массивов данных рынка судостроительной отрасли в целях маркетинговых исследований с использованием таких программных продуктов, как MS Excel и QlikView.

***Задачи исследования***

1. Ознакомиться со спецификой новой предметной области на базе исследований, проведенных ИА «ИНФОЛайн-Аналитика», выделить ключевые информационные ресурсы, являющиеся основанием маркетинговых исследований по рынку судостроительной промышленности;
2. Разработать аналитический план работы с полученными данными, включающий в себя перечень возможных гипотез, строящихся на основании задействованных переменных, и их разрешения для маркетинговых исследований рынка судостроения;
3. Создать приложение в BI-системе QlikView, являющееся демонстрацией результатов маркетингового исследования.

В чем может заключаться интерес в данном исследовании непосредственно информационных агентств, таких как «ИНФОЛайн»:

* Прописываются «узкие» места работы с массивом данных;
* Автоматизируется процесс вычленения необходимой информации из текстовых полей данных;
* Визуализируется процесс выполнения комплексного анализа текущего состояния судостроительной отрасли в России.

***Объект исследования***

Информационно-аналитическое агентство ИНФОЛайн.

***Предмет исследования***

Процесс автоматизации процедур обработки данных в интересах информационно-аналитического агентства ИНФОЛайн.

***Практическая значимость результатов исследования***

Данное исследование отличается высокой практической значимостью. В ходе его создания был разработан алгоритм автоматизации обработки, хранения и предоставления неструктурированных массивов данных, позволяющий сделать процесс создания маркетинговых исследований ИА ИНФОЛайн наиболее результативным. Разработанный алгоритм рекомендуется внедрять в аналитический отдел.

***Новизна***

Научная новизна магистерского исследования состоит в разработке аналитического плана работы с неструктурированными данными на примере конкретных данных судостроительной отрасли в интересах информационно-аналитического агентства ИНФОЛайн.

# АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ И ХАРАКТЕРИСТИКА ОБЪЕКТА ИССЛЕДОВАНИЯ

## Общая информация об ИА ИНФОЛайн

Информационное агентство ИНФОЛайн было создано в 1999 году для оказания информационно-консалтинговых услуг коммерческим организациям. Осуществляет на постоянной основе информационную поддержку более 1000 компаний России и мира. Агентство ежедневно осуществляет мониторинг публикации в более 5000 СМИ и ежедневно ведет аналитическую работу 80 тематикам экономики РФ

ИНФОЛайн - информационно-аналитическое агентство, ориентированное на оказание услуг в сфере B2B, осуществляет оперативный сбор экономических событий, разработку и реализацию аналитических продуктов, организацию работы со средствами массовой информации, экспертную поддержку компаний, информационную и организационную поддержку форумов и конференций.

Начиная с 2010 года, агентство ИНФОЛайн по заказу клиентов и по собственной инициативе проводит кабинетные исследования рынков судостроения и судоремонта. При подготовке маркетингового исследования специалисты агентства используют уникальное информационное обеспечение и опираются на многолетний опыт работы с различными новостными потоками.

## Информация о готовом продукте ИА ИНФОЛайн

Одним из основных продуктов информационного агентства «ИНФОЛайн» является ежегодное исследование в сфере судостроения «Судостроительная промышленность России. Итоги 2015 г. Тенденции 2016 г. Прогноз до 2020 г.». Целью этого исследования является выполнение комплексного анализа текущего состояния судостроительной отрасли в России и разработка долгосрочных прогнозов направления развития отрасли и общего спроса на продукцию различных сегментов судостроения.

Также отдельный раздел посвящен рейтингам судостроительных предприятий и холдингов.

Судостроительный комплекс Российской Федерации исторически сложился в виде групп предприятий, тяготеющих к отдельным морским и внутренним водным бассейнам. По данным Минпромторга РФ судостроительные мощности расположены так, как показано на схеме 1.



**Северо-Запад**

предприятий 27%

кадры 58%

объем пр-ва 42%

**Центр**

предприятий 11%

кадры 9%

объем пр-ва 6%

**Юг (включая Крым)**

предприятий 23%

кадры 8%

объем пр-ва 11%

**Поволжье**

предприятий 18%

кадры 12%

объем пр-ва 32%

**Сибирь**

предприятий 5%

кадры 1%

объем пр-ва 1%

**Урал**

предприятий 5%

кадры 1%

объем пр-ва 1%

**Дальний Восток**

предприятий 12%

кадры 11%

объем пр-ва 7%

Схема 1. Схема распределения судостроительных мощностей на территории РФ[[1]](#footnote-1)

Целью составления рейтингов судостроительных предприятий России является изучение отраслевой динамики, определение приоритетных направлений развития судостроения, выявление факторов, определяющих направление работы основных российских верфей. Эти рейтинги получены как на основе данных отчетности предприятий, так и на основе базы данных «Заказы на строительство судов на верфях России», разработанной и поддерживаемой ИА ИНФОЛайн. Ранжирование выполнялось по финансовым показателям деятельности предприятий, по показателям совокупного тоннажа и совокупной стоимости сданных и строящихся судов.

Основная цель Аналитической базы «Заказы на строительство судов на верфях России: Итоги 2015-2016 годов»: мониторинг инвестиционных процессов и проектов строительства и реконструкции судов на российских верфях, включая гражданские и военные проекты, а также проекты по заказу зарубежных заказчиков.

При подготовке Исследования использовались следующие информационные источники:

База данных «Заказы на строительство судов на верфях России» - уникальный продукт, разработанный ИА ИНФОЛайн. Она содержат сведения о портфеле заказов около 60 основных судостроительных предприятий России, включая все крупнейшие отечественные верфи. По состоянию на 2015 год обновленная база содержала данные по более чем 2070 заказам, размещенным в отрасли начиная с 2000 года. База содержит записи о каждом строящемся судне с указанием предприятия-строителя, холдинга, к которому относится предприятие, заказчика строительства с указанием государственной принадлежности, типа заказа (военный/гражданский), типа судна, дедвейта, тоннажа, плановых сроков окончания строительства и текущего статуса заказа.

* экспертные опросы и анкетирование более 100 судостроительных и транспортных компаний;
* анализ новостной информации из подборки новостей, которые ИА ИНФОЛайн осуществляет в рамках услуги [Тематические новости: «Судостроительная промышленность РФ и стран ближнего зарубежья»](http://infoline.spb.ru/shop/tematicheskie-novosti/page.php?ID=22017&sphrase_id=54875);
* финансовая отчетность предприятий, пресс-релизы и презентации, материалы корпоративных интернет-сайтов;
* материалы СМИ: федеральной и региональной прессы, информагентств, электронных СМИ;
* материалы более 60 специализированных и отраслевых СМИ и сайтов, посвященных судостроительной и судоремонтной отрасли, военно-морскому флоту, а также водному транспорту;
* данные Федеральной службы государственной статистики;
* Материалы Министерства транспорта, Министерства промышленности и торговли, ФА по рыболовству;
* Нормативные акты Российской Федерации, Федеральные Целевые Программы Правительства РФ.

Как видно из информационных источников, данные достаточно разнородны, а также находятся в разных форматах. Например, экспертные опросы и анкетирование более 100 судостроительных и транспортных компаний представляют из себя набор текстовых файлов каждой компании за определенные год. Проблема свода их в один файл связана с различием переменных. Структура анкеты для судостроительных верфей представлена в приложении 4.

## Характеристика отрасли

В данном исследовании рассматривается *рынок* *судостроительной отрасли,*на котором действуют предприятия, осуществляющие строительство и крупные ремонты судов, а также организации, чья деятельность связана с поставками комплектующих и оборудования для судостроительной отрасли, проектные институты и конструкторские бюро.

Для магистерского исследования взят срез данных по корпорации АО «ОСК» (Объединенная судостроительная корпорация). Возникновение корпорации произошло путем объединения большинства крупнейших и конкурентоспособных судостроительных предприятий России в рамках единой госкорпорации АО «ОСК», которая контролирует по некоторым оценкам до 80% российского судостроительного рынка. По оценкам игроков рынка это было вынужденной мерой, так, например, АО «ОСК» способствовало сохранению работоспособности судостроительных предприятий, входивших в обанкротившуюся группу «Объединенная промышленная корпорация». Фактически АО «ОСК» выступила в качестве инструмента господдержки судостроительной отрасли, особенно по направлению военного судостроения.

В структуре производства российского судостроения преобладает строительство боевых кораблей и вспомогательного флота, как для ВМФ РФ, так и на экспорт. На рисунке 1 представлена структура судостроительной отрасли Российской Федерации, сложившейся к 2016 году. На схеме указаны типы предприятий и их численность по сведениям Минпромторг РФ.



Рисунок 1. Структура судостроительной отрасли РФ по состоянию на 2016 год

Основными терминами судостроительной отрасли, которые будут встречаться в работе:

*Дедвейт* – полная грузоподъемность судна, масса всех грузов, которые может принять судно. В дедвейт входит масса полезно груза (чистая грузоподъемность), судовых запасов (топливо, вода для парогенераторов, смазочное масло, пресная вода и провизия для экипажа, расходные материалы и пр.), экипажа и др. переменных в процессе эксплуатации судна грузов[[2]](#footnote-2);

*Тоннаж –* количественная характеристика судна соответствующая его весовому водоизмещению;

*IMO* – уникальный идентификационный номер, который присваивается морскому судну на основе Регистра судоходства Ллойда в соответствии с требованиями Международной морской организации (International Maritime Organization, IMO);

*PPP –* учет судна в Российском Речном Регистре[[3]](#footnote-3).

# ПОДГОТОВКА ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ

Кратко структуру алгоритма подготовки данных можно описать так:

1. Для каждого вида источников информации выделить приоритетные части, сузив область обработки данных;
2. Устранить ошибки в переменных, связанные с неправильным представлением в базе;
3. Проанализировать выбранные переменные на предмет информативности, если они содержат необходимую информацию не в явном виде, вычленить ее с последующим удалением ненужных значений;
4. Определить ключевые переменные;
5. Разработать аналитический план работы с данными, включающий описательные статистики и постановку важных как для клиента, так и для самих аналитиков ИНФОЛайн, гипотез;
6. Осуществить консолидацию всех данных в одном программном продукте с возможностью обновления данных из внешних источников;

## Консолидация данных

Ценность и достоверность информации, полученной в результате интеллектуального анализа бизнес-данных, зависит не только от эффективности используемых аналитических методов и алгоритмов, но и от того, насколько правильно подобраны и подготовлены исходные данные для анализа, поэтому прежде чем приступать к анализу данных, необходимо было выполнить ряд процедур, цель которых — доведение данных до приемлемого уровня качества и информативности.

*Консолидация* — комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразование в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему[[4]](#footnote-4).

Так как первоначально у нас были данные в разных источниках, а именно анкеты от множества судостроительных компаний, то необходимо было свести все анкеты в один файл для удобства работы с ними. Для этого был использован макрос для сбора всех листов из необходимых файлов в одну книгу Excel.

>Sub AddWorkbooks()

    Dim FilesToOpen

    Dim x As Integer

    Application.ScreenUpdating = False

    'вызываем диалог выбора файлов для импорта

    FilesToOpen = Application.GetOpenFilename \_

      (FileFilter:="All files (\*.\*), \*.\*", \_

      MultiSelect:=True, Title:="Files to Merge")

    If TypeName(FilesToOpen) = "Boolean" Then

        MsgBox "Не выбрано ни одного файла!"

        Exit Sub

    End If

    'проходим по всем выбранным файлам

    x = 1

    While x <= UBound(FilesToOpen)

        Set importWB = Workbooks.Open(Filename:=FilesToOpen(x))

        Sheets().Copy After:=ThisWorkbook.Sheets(ThisWorkbook.Sheets.Count)

        importWB.Close savechanges:=False

        x = x + 1

    Wend

    Application.ScreenUpdating = True

>End Sub

Следующим шагом необходимо было свести все анкеты в одну таблицу. Для этого была использована авторская надстройка для Excel – «Plex»[[5]](#footnote-5). Excel не имеет среди своих средств ничего подобного, а выполнение подобной задачи вручную, например, копированием через буфер, - весьма трудозатратно при большом количестве листов.

При помощи надстройки в книгу вставляется новый пустой лист, куда собираются данные со всех остальных листов книги. Сборка может производиться по нескольким критериям на выбор (рис. 2).

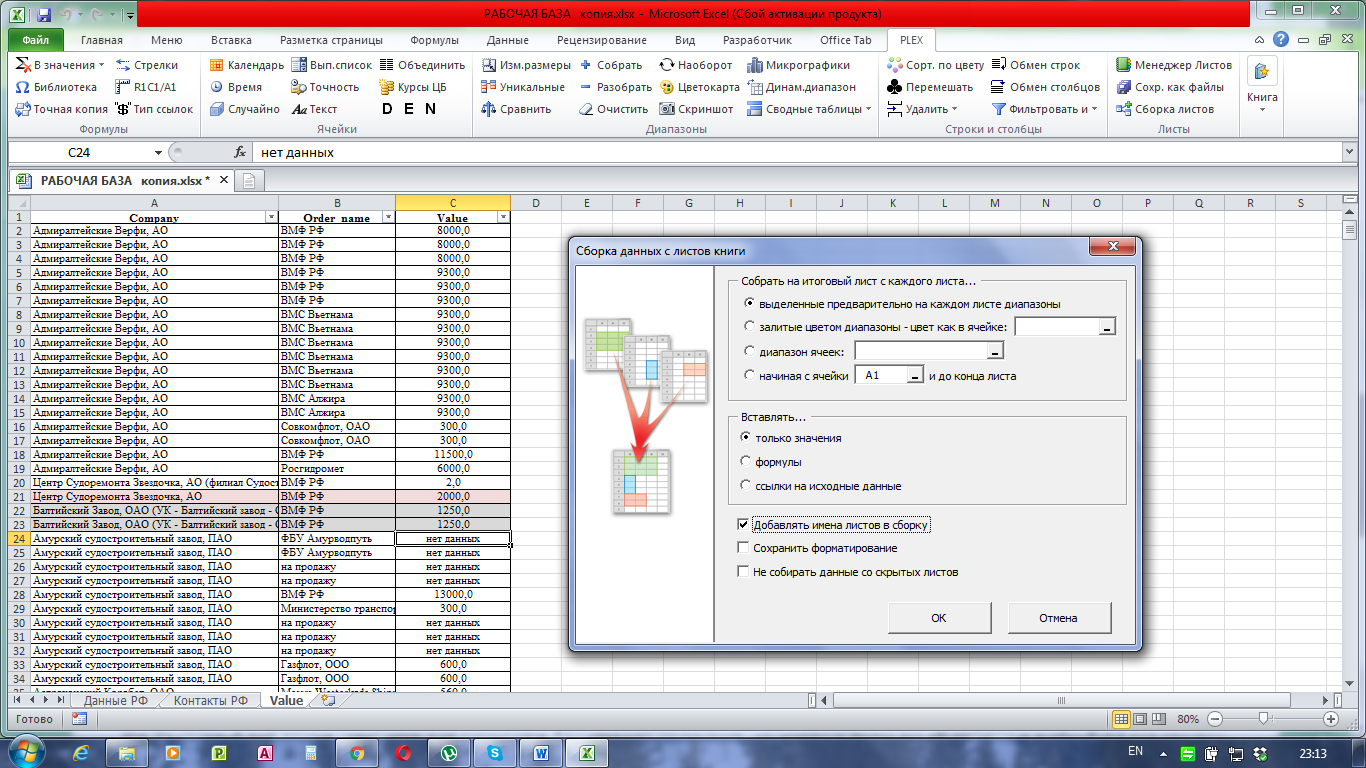


Рисунок 2. Сборка листов

## Очистка данных

На этапе подготовки данных также одним из важных пунктов была задача очистки данных.

*Очистка данных* — это процесс анализа качества данных в источнике данных с выполняемым вручную утверждением или отклонением рекомендаций, даваемых системой, и внесением изменений в данные[[6]](#footnote-6).

Извлечение данных из разных источников и перенос их в ИС с целью дальнейшей аналитической обработки связаны с рядом проблем, основными из которых являются нижеследующие:

* Исходные данные расположены в источниках самых разнообразных типов и форматов, в то время как для решения задач анализа данные должны быть преобразованы в единый универсальный формат, который поддерживается аналитическим приложением;
* Данные в источниках обычно излишне детализированы, тогда как для решения задач анализа в большинстве случаев требуются обобщенные данные;
* Исходные данные, как правило, содержат различные факторы, которые мешают их корректному анализу.

Чтобы подготовить данные к загрузке происходит сначала первичная очистка данных. Вторичная очистка в аналитической системе является пользовательской, она направлена на подготовку данных к решению конкретной аналитической задаче.

Для каждого структурного уровня данных — отдельной ячейки, записи, целой таблицы, отдельной БД или множества БД — характерны свои факторы, снижающие качество данных.

На уровне отдельной ячейки таблицы наиболее характерны следующие ошибки:

* Орфографические ошибки или опечатки могут привести к неправильному пониманию или искажению данных. Например, ошибка в названии судостроительной компании приведет к появлению двух разных фирм с одинаковыми реквизитами, что вызовет противоречия, ошибки в отчетах и т.д., а также приведет к неверным управленческим решениям;
* Пропуски данных — отсутствие данных в тех ячейках, где они должны быть. Пропуски могут быть вызваны ошибкой аналитика или отсутствием соответствующей информации;
* Фиктивные значения – Подобная ситуация возможна, если аналитик или заказчик, не располагая необходимой информацией, ввел в ячейку произвольную последовательность символов.

Формирование БД «Заказы на строительство судов на верфях России» производится несколькими сотрудниками, и, например, переменная адрес компании, несмотря на требование к заполнению, может представлять проблему визуализации отчетности, так как адрес фактического производства, юридический адрес и фактический адрес офиса могут и чаще всего отличаются. Также не исключены частые ошибки в заполнении данного поля. Это требует некоторых усилий для определения геолокации судостроительных компаний.

Для этого при помощи все той же надстройки «Plex» была произведена работа с содержимым ячеек (рис. 3).

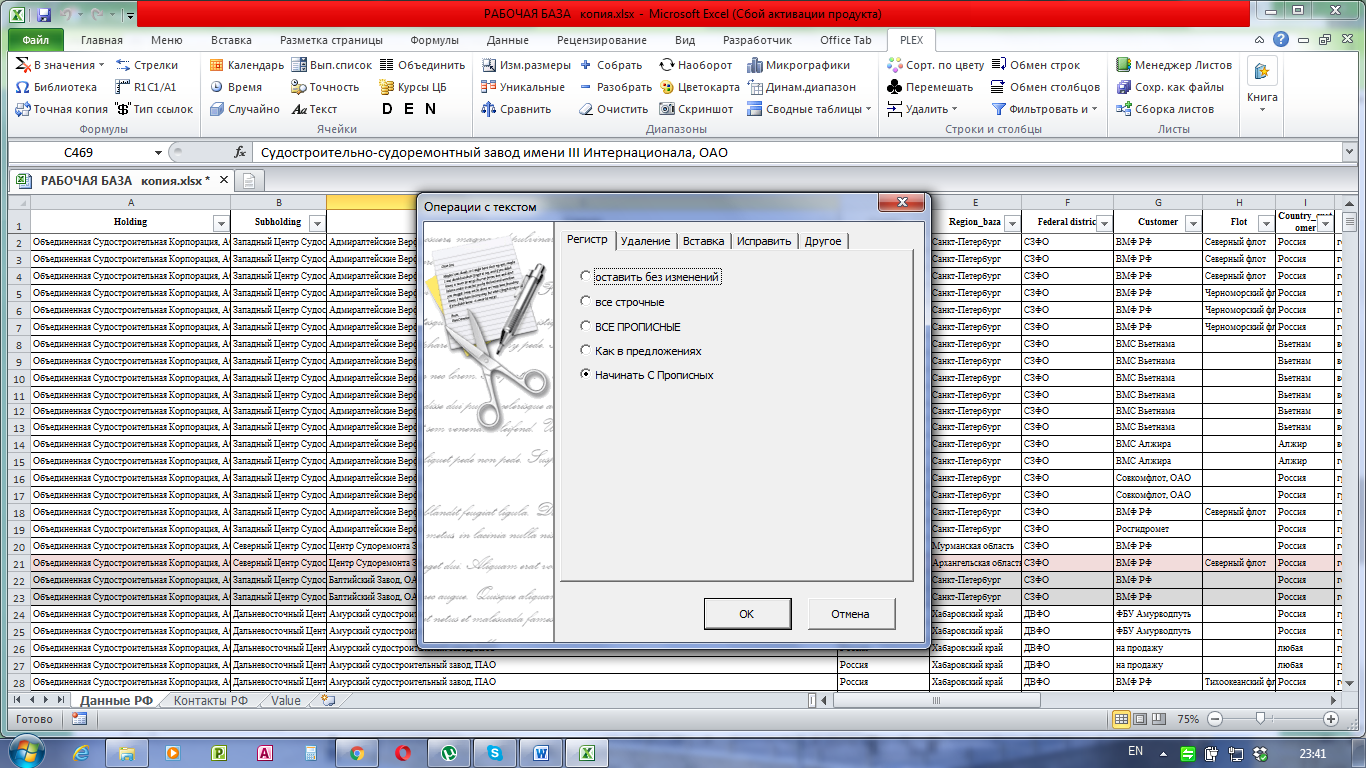


Рисунок 3. Операции с текстом в надстройке Plex

Вкладка *Регистр*позволила преобразовать регистр символов и привести его к нужному виду. Вкладка *Удаление*позволила быстро удалить лишние пробелы и непечатаемые символы. Вкладка *Исправить* позволила исправить числа из формата «текстовый» в формат «числовой».

После выполнения всех процедур над файлами мы получили готовый к загрузке рабочий файл.

Для наглядности смоделируем в виде бизнес-процесса. Воспользуемся для этого нотацией *BPMN*(англ. Business Process Model and Notation, нотация и модель бизнес-процессов) (рис. 4).

Нотация используется для описания процессов нижнего уровня. Диаграмма процесса в нотации BPMN представляет собой алгоритм выполнения процесса. На диаграмме могут быть определены события, исполнители, материальные и документальные потоки, сопровождающие выполнение процесса[[7]](#footnote-7).

|  |  |  |
| --- | --- | --- |
| **Младший аналитик** | | Сбор данных  Согласование данных  База данных «Заказы на строительство»  Экспертные опросы, Анкетирование  Материалы СМИ  Финансовая отчетность  **Х**  Подготовка данных  Определение ключевых переменных  Консолидация в QlikView  Разработка гипотез  Согласование  Сбор результатов в маркетинговое исследование  **Х**  Выставление на продажу  Согласовано |
| **Главный аналитик** | | Не согласовано |
| **Руководство** | | Согласовано  Не согласовано |
| **Менеджер** |  | |

Рисунок 4. Описание бизнес-процесса

# ОБОСНОВАНИЕ ВЫБОРА BI-СИСТЕМЫ

*Business intelligence* (сокращённо BI) — это методы и инструменты для перевода необработанной информации в осмысленную, удобную форму. Эти данные используются для бизнес-анализа. Технологии BI обрабатывают большие объёмы неструктурированных данных, чтобы найти стратегические возможности для бизнеса.

Термин впервые появился в 1958 году в статье исследователя из IBM Ханса Питера Луна. Он определил этот термин как: «Возможность понимания связей между представленными фактами»[[8]](#footnote-8)

BI в сегодняшнем понимании эволюционировал из систем для принятия решений, которые появились в начале 1960-х и разрабатывались в середине 1980-х.

В 1989 году Говард Дреснер (позже аналитик Gartner) определил Business Intelligence как общий термин, описывающий «концепции и методы для улучшения принятия бизнес-решений с использованием систем на основе бизнес-данных».

По оценкам TAdviser[[9]](#footnote-9), объем российского рынка BI-систем в 2015 году, включая услуги по внедрению и поддержке, увеличился на 5% и достиг 45,9 млрд. рублей.

Сегодня рынок информационных технологий предлагает большое количество BI-инструментов, которые способны удовлетворить практически любые требования и желания компании, решившей внедрить аналитические инструменты. Только Магический Квадрант Гартнера (Gartner Magic Quadrant for BI & Analytics Platforms), ежегодный рейтинг лидеров BI-отрасли, включает порядка 25 производителей.

В 2016 году Gartner сузил свое определение для этого квадрата, включив в него только самообслуживающиеся аналитические инструменты.

В феврале 2017 года Gartner публикует свой свежий рейтинг основных BI-систем. На рис. 1 сразу выделяются три лидирующих системы: Tableau, Microsoft и Qlik. Microsoft и его продукт Power BI занял лидирующую позицию по показателю «Полнота видения» и делит лидирующую позицию по показателю «Простота внедрения» с Tableau.



Рисунок 5. Магический квадрант рынка BI-систем и аналитических платформ

Для последующего выбора необходимой нам платформы стоит отметить сильные стороны Power BI:

* Стоимость;
* Простота использования и комплексный анализ;
* Полнота видения;
* Активное сообщество и поддержка.

Далее рассмотрим каждый из пунктов подробнее.

В 2015 году Microsoft существенно снизил стоимость Power BI (до 9,99$ в месяц за каждого пользователя), сделав его одним из самых дешевых решений на рынке. Низкая общая стоимость была признана вторым по важности критерием выбора клиента Microsoft Power BI. По словам Gartner, Microsoft выигрывает «снижением цен на рынке BI и аналитики». Он отмечает, что лицензия Power BI составляет одну треть стоимости постоянной лицензии для других инструментов и на 80% ниже других предложений облачных вычислений.

Под полнотой видения подразумеваются возможности по получению, преобразованию и представлению различных данных. Именно в этом направлении Микрософт занял первое место среди всех игроков рынка.

Также Microsoft имеет сильное сообщество партнеров, посредников и отдельных пользователей. Это сообщество расширяет продукт с готовыми приложениями, визуализациями и видео-уроками, расширяя контент, предоставляемый корпорацией Microsoft.

Но у данного продукта есть также и недостатки, рассмотрим основные из них:

* Одним из крупнейших недостатков пользователями отмечается то, что *продукт находится в стадии развития и доступен только в облачной инфраструктуре*. На данный момент у Microsoft отмечается слабая реализация функций и невозможность обрабатывать большие объёмы данных. Существенным пробелом в функционале является невозможность отображения данных в сводной таблице или подведение промежуточных итогов в рамках табличного отображения.
* Следующим недостатком отмечается *спектр применения*. Он отображает процент пользователей, использовавших продукт в различных стилях, включая просмотр отчетов, создание персонализированных информационных панелей, проведение простого AD-HOC анализа, выполнение сложных запросов, подготовку данных и использование моделей прогнозирования.
* Следующим недостатком пользователями отмечается *отсутствие единого стандарта:* Microsoft Power BI часто используется в комбинации с другими инструментами, однако подобная комбинированная стратегия усложняет использование продукта потребителями, желающими упростить работу с продуктом. В таком случае клиенты могут использовать Power BI в качестве недорого варианта для внедрения широко применяемых простых информационных панелей, совмещая приложение с решениями от других поставщиков, предоставляющих более широкий функционал.

Несмотря на то, что Microsoft Power BI стал лидером обзора Gartnera, Tableau и Qlik - или любой другой BI и аналитический инструмент очень перспективные и популярные системы. Gartner по-прежнему рекламирует Tableau и Qlik, как «золотой стандарт для интуитивного интерактивного исследования».

Одним из основных факторов, по которым Qlik в этом году немного опустился в своих позициях, стал переход рынка с унаследованного продукта QlikView на новую платформу QlikSense.

В настоящее время фокусировка на визуализации является одним из определяющих факторов сдвигов, которые произошли на рынках BI и аналитических инструментов. В то время как аналитики любят использовать свои инструменты визуализации, чтобы определять модели данных, Gartner видит «интеллектуальное обнаружение данных», основанное на алгоритмах машинного обучения, как тенденцию будущего.

В итоге, после рассмотрения трех лидеров обзора, изучения их достоинств и недостатков, в качестве среды разработки приложения была выбрана BI-платформа QlikView.

Платформа QlikView предназначена для бизнес-анализа, выполняемого конечными пользователями. QlikView является основным продуктом компании QlikTech, сокращённо Qlik, которая была основана в 1993 году в шведском городе Лунд. В 2012 году QlikTech, наряду с Apple и LinkedIn, вошла в первую тройку рейтинга 25 наиболее быстро растущих технологических компаний Америки по версии журнала Forbes.

Среди основных возможностей платформы QlikView можно выделить следующие:

* Консолидация в одном приложении данных из различных источников – распространённых СУБД, например, MS Access, MySQL, Oracle Database и других, файлов различных форматов, таких как MS Excel, XML-, HTML, а также из разнообразных информационных систем, таких как 1С или SAP, но при условии использования специальных коннекторов.
* Обеспечение совместной работы онлайн и в защищённой среде;
* Мощный аппарат визуализации с возможностью построения дашбордов.

Главной особенностью QlikView является то, что она реализует ассоциативную архитектуру с обработкой данных в оперативной памяти. Суть ассоциативного анализа состоит в том, что платформа объединяет таблицы из всех предоставленных информационных источников, ищет идентичные наименования полей и, сопоставляя их, устанавливает ассоциативные связи между сущностями. Таким образом, в оперативной памяти QlikView хранятся таблицы и определённые между ними ассоциативные связи.

Если сравнивать QlikView с его главным конкурентом Tableau, то основные различия представлены в таблице 1.

Таблица 1. Сравнительная характеристика QlikView и Tableau

| ***№ п/п*** | ***Характеристика*** | ***QlikView*** | ***Tableau*** |
| --- | --- | --- | --- |
| 1 | Онлайн источники данных | Да | Ограниченные возможности |
| 2 | Глобальный поиск | Да | Нет |
| 3 | Вычисления | Скрипты | IDE |
| 4 | Производительность | Миллиарды строк | Неограниченная |
| 5 | Масштабирование | Каждый объект настраивается отдельно | Все объекты масштабируются под размер экрана за счет помещения их в контейнер |
| 6 | ETL – обработчик | Есть, данные легко модифицировать в процессе загрузки | Нет, требуются уже обработанные данные |
| 7 | Принцип разработки | Выражения и скрипты | Через пользовательский интерфейс |
| 8 | Свободная разработка и размещение фильтров | Да | Ограниченные возможности |
| 9 | Модель доступа к данным | Скрипты | При помощи меню |
| 10 | Коммуникация между пользователями | Да | Ограниченные возможности |
| 11 | Кем создаются приложения | Разработчик, программист | Бизнес-пользователь |
| 12 | Дополнительные средства разработки | Визуализация, синтаксис выражений и т.п. | Ставка на визуализацию |
| 13 | Построение отчетов | Требуются знания SQL | Путем перемещения элементов (VizQL) |
| 14 | Встроенные карты | Нет, можно только использовать карту как картинку | Детализированный уровень встроенных карт (регионы, улицы) |

Исходя из результатов сравнения, можно сделать вывод о том, что, в то время как Tableau привлекает пользователей простотой загрузки данных и их визуализации, а также высокой скоростью смены многообразия разрезов для анализа, QlikView, предлагая расширенные возможности разработки, требует больше трудозатрат и определённого набора технических знаний от того, кто создаёт приложения, но при этом позволяет получать более детальные отчёты и выявлять более сложные зависимости между данными.

Также непосредственно в пользу данной платформы были предпочтения руководства, так не нужно будет производить дополнительные затраты на покупку аналитической платформы, потому что пакет Qlik уже куплен компанией.

***Лицензирование***

Бесплатная версия *QlikView Personal Edition* - является полноценным инсталлируемым приложением без ограничений по сроку использования, количеству таблиц, отчетов, объема данных и т.д., и включает средства ETL.

*Ограничение версии Personal Edition*:

Приложения (отчеты), созданные с помощью Personal Edition можно открывать только на том же рабочем месте, где они были созданы. Открывать приложения, сделанные другими пользователями QlikView нельзя. Все отчеты, созданные с помощью бесплатной версии Personal Edition, позже могут быть открыты с помощью любых полноценных версий QlikView, лицензия на которые приобретаются отдельно. Также невозможно осуществить полную загрузку из XML-файла.

Используя лицензии QlikView Server или QlikView Local Client (купленные лицензии) можно открыть любой документ, но пользователь лицензии для личного использования может открыть документ только с правильным ключом пользователя.

# РАЗРАБОТКА АЛГОРИТМА АВТОМАТИЗАЦИИ ПРОЦЕДУР ОБРАБОТКИ НЕСТРУКТУРИРОВАННЫХ МАССИВОВ ДАННЫХ

## Определение бизнес-правил

Система бизнес-правил понятие довольно широкое. В нашем случае это некий набор условий, которые управляют нашим процессом, для того, чтобы оно происходило так, как нужно для заказчика и разработчика. Важно также сформулировать условия, которые невозможны и нежелательны.

Определение необходимых требований к системе является одним из первостепенных этапов проектирования.

Основные бизнес-правила для работы с платформой:

* Все загруженные таблицы в приложении имеют конкретный путь, при перемещении файлов необходимо в скрипте переписать его;
* Для того чтобы понять интерпретацию данных и форматирование чисел в программе QlikView, необходимо знать, каким образом хранятся данные внутри программы;
* Адреса судостроительных компаний вносятся и указываются с указанием индекса и названия города;
* Стоимость контракта указывается в млн. рублей;
* Ячейка оставляется пустой при отсутствии данных, так как при попадании в систему она автоматически преобразуется в ноль;
* Чтобы выбрать несколько элементов в одном списке, следует удерживать клавишу *Ctrl* во время выбора дополнительных значений;
* Для того, чтобы на выбранный объект листа не действовали фильтры необходимо его *Открепить* в свойствах данного объекта;
* Информация исследования открыта для просмотра сотрудникам определенного подразделения аналитического агентства;
* Изменять данные и модифицировать приложение могут только старший аналитик и руководство.

Для анализа доступна информация, собранная за время функционирования крупнейшей судостроительной компании «Объединенная судостроительная корпорация» на рынке. Временной срез включает в себя контракты заключенные компанией с 2002 года.

## Моделирование данных

*Анализ входной информации*

Входная информация для разрабатываемого приложения поступает из трёх источников:

* База «Заказы на строительство судов на верфях России: Итоги 2015-2016 годов»;
* База стоимости контрактов судостроительных заказов;
* Контактный справочник по судостроительным компаниям.

База стоимостей контрактов судостроительных заказов и контактный справочник представляют собой электронные таблицы MS Excel. Ниже приведены их структуры (табл.2, 3).

Таблица 2. Структура таблицы «Стоимость\_контракта»

|  |  |
| --- | --- |
| **Атрибуты** | **Наименование полей таблицы** |
| Компания исполнитель | Company |
| Заказчик | Order\_name |
| Стоимость контракта | Value |

Таблица 3. Структура таблицы «Контакты\_компании»

| **Атрибуты** | **Наименование полей таблицы** |
| --- | --- |
| Наименование компании | Company |
| Тип компании | Type |
| Статус работы | Stat |
| Федеральный округ | FO |
| Регион | Region\_contacts |
| Адрес | Adress |
| Телефон | Telefon |
| Факс | Faks |
| E-mail адресс | @ |
| Web адрес страницы компании | Web |
| ФИО руководителя | Name\_director |
| Занимаемая должность | Position |

## Разработка приложения автоматизации процедур обработки неструктурированных массивов данных

Заметим, что данные представленные в данной работе, относятся к категории ретроспективной информации, так как База «Заказы на строительство судов на верфях России: Итоги 2015-2016 годов», как ясно из названия, содержит данные актуальные на начало 2016 года, к тому же срез данных осуществлен по одному холдингу судостроительной отрасли. Такой подход не снижает результатов исследований, так как при практическом применении приложения этот алгоритм можно переложить на релевантные и более емкие данные. В настоящее время в работе представлены результаты оценки эффективности и явное превосходство коммуникативной составляющей визуализации данных в выбранном программном продукте на основе имеющейся релевантной информации.

После организации, очистки и консолидации данных мы имеем четко очерченный пул информации, которую необходимо проверить на те или иные гипотезы, потенциально или реально важные для исследования компании.

### Загрузочный скрипт

Создание документа QlikView представляет собой извлечение данных из одного или нескольких источников, например из реляционной базы данных или из текстовых файлов, содержащих таблицы данных. Это извлечение выполняется с помощью созданного и выполненного скрипта, в котором указывается база данных, таблицы и поля, которые будут использоваться.

При формировании загрузочного скрипта во избежание возникновения циклических ссылок, нарушающих работу ассоциативной логики QlikView, изначально совпадающие наименования полей разных таблиц были изменены. Если приложением QlikView во время выполнения скрипта будет обнаружена циклическая ссылка, для таблиц будут использоваться слабые связи.

Полный текст скрипта загрузки вынесен в Приложение 1: Код скрипта загрузки.

В результате исполнения загрузочного скрипта была сформирована ассоциативная модель сущность-связь (рис. 6).

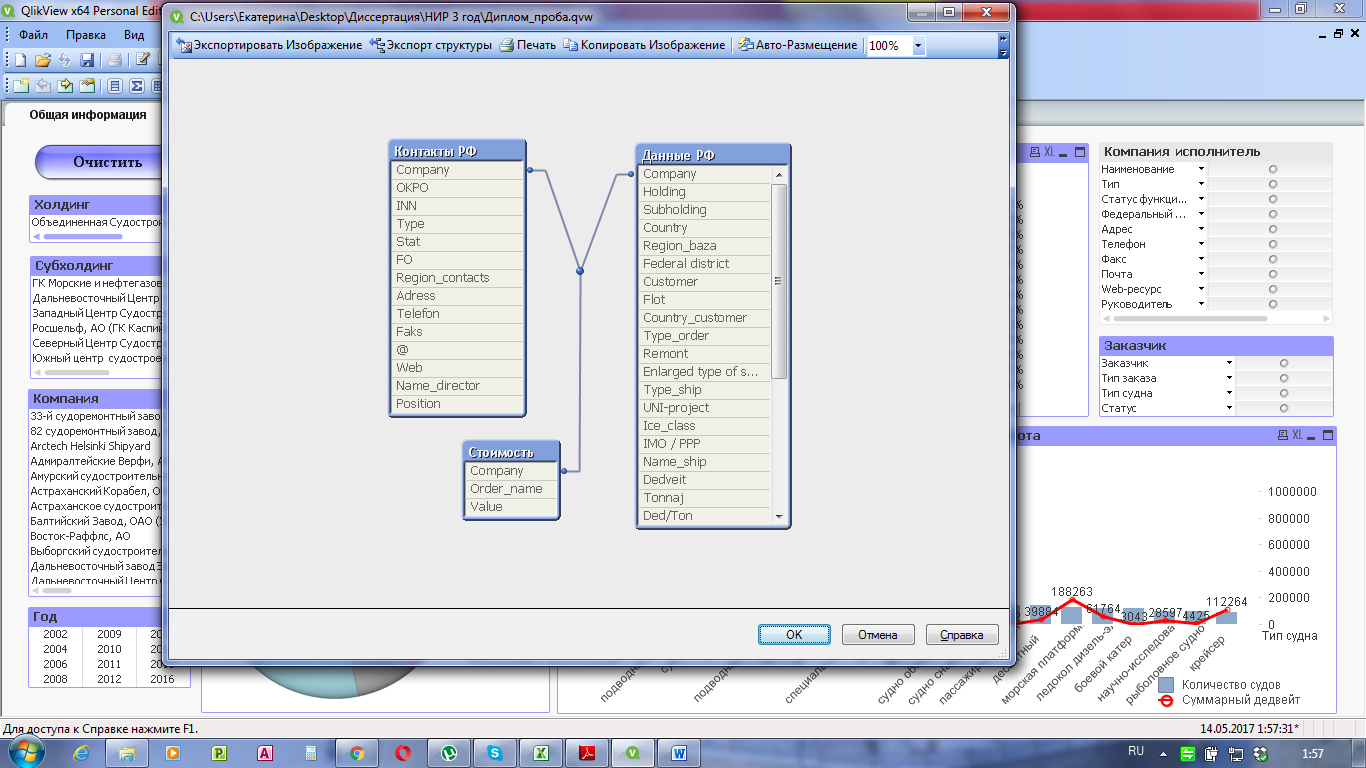


Рисунок 6. Ассоциативная модель данных

Так как платформа должна служить для увеличения скорости работы аналитика при написании кабинетных исследований, а также для минимизации вероятности допуска ошибок, то минимальное количество вопросов, которые, на мой взгляд, необходимо было вынести в рамках анализа следующие:

1. Определить структуру портфеля заказов, отображающую выборку по годам;
2. Количественный состав российского гражданского флота;
3. Определение самых частых и суммарно выгодных заказчиков;
4. Определить топ самых крупных заказчиков и исполнителей РФ на рынке судостроения;
5. Объём заключенных контрактов;
6. Динамика количества переданных заказчикам российскими верфями судов;
7. Посмотреть на типы судов и определить, какой тип более востребован для иностранных партнеров или наших производителей;
8. Определить суммарные затраты на строительство судов той или иной компанией;

### Пример решения сформулированных задач исследования в BI-системе «QlikView»

Демонстрация работы разработанного приложения целесообразно проводить в следующей последовательности.

1. **Определить структуру портфеля заказов, отображающую выборку по годам:**

На панели инструментов *Дизайн* выбираем пункт *Создать диаграмму.* Откроется первая страница мастера создания диаграммы — *Общие*. На этой вкладке можно выбрать необходимый тип диаграммы – *Круговая*. Далее на вкладке *Измерения* выбираем параметр *Enlarged type of ship* и перемещаем его в список отображаемых полей. Чтобы каждая часть круговой диаграммы соответствовала общему количеству выпущенных на воду типов укрупненных судов, выберем *Count*в раскрывающемся списке *Агрегирование*и *Enlarged type of ship* в списке *Поле*. Так как нам не столь важна количественная характеристика, отмечаем также пункт отображения значений в долях. Чтобы улучшить представление диаграммы, можно ограничить максимальное количество отображаемых столбцов, но в данном случае у нас стоит другая задача, поэтому ограничивать представление не будем.

Результат представлен на рис. 7.

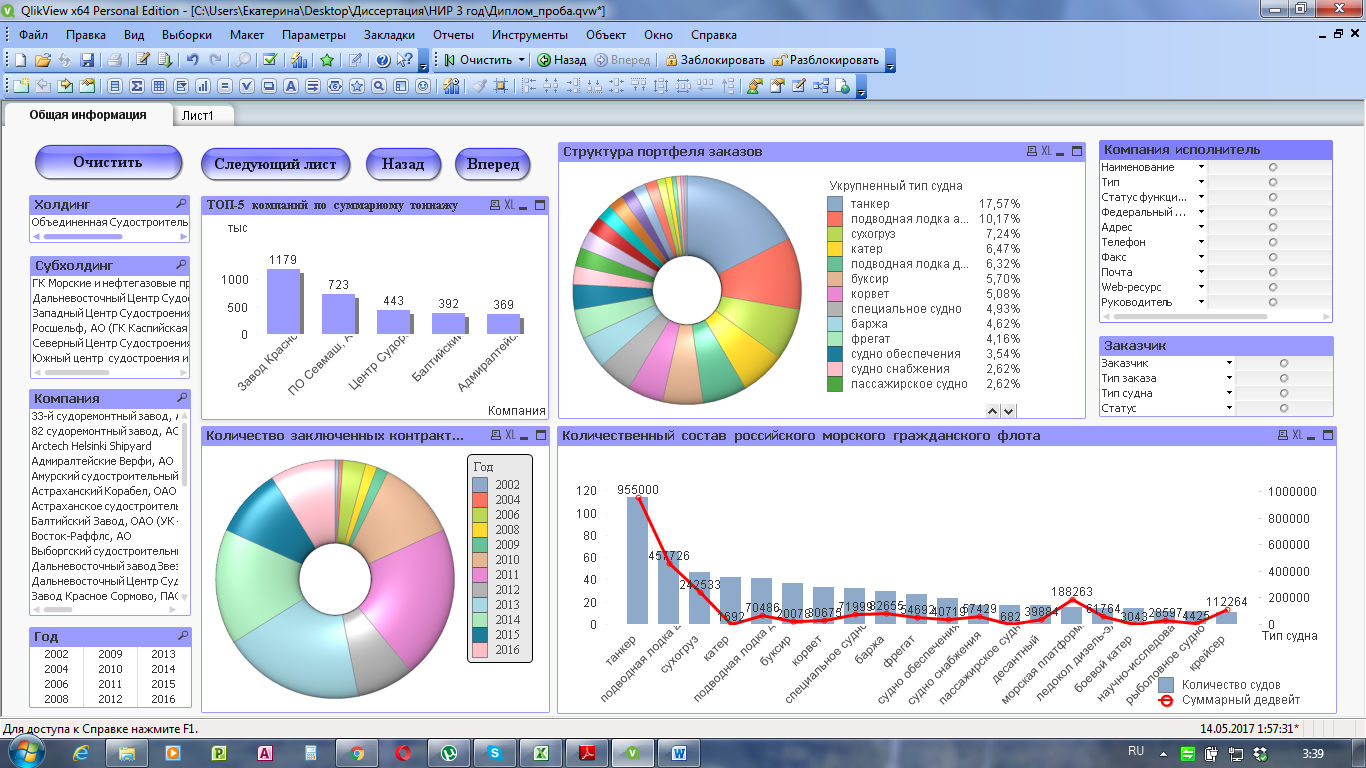


Рисунок 7. Диаграмма отображения структуры портфеля заказов

Данная структура отражает сырьевой характер российской экономики: в ней более 70 % заказов приходится на танкерный флот и только около 15% – на сухогрузные суда. В целом же, по состоянию на III квартал 2016 года, флот морских судов, контролируемых российскими предприятиями, насчитывает порядка 3 тысяч судов различного класса.

Так как ранее указывалось, что QlikView обладает ассоциативным поиском, то мы можем данную диаграмму фильтровать по годам, по компаниям исполнителям, по компаниям заказчикам, по типу заказа. При этом этому же фильтру будут подвержены и все остальные объекты листа, если только они не откреплены (рис. 8,9,10,11).

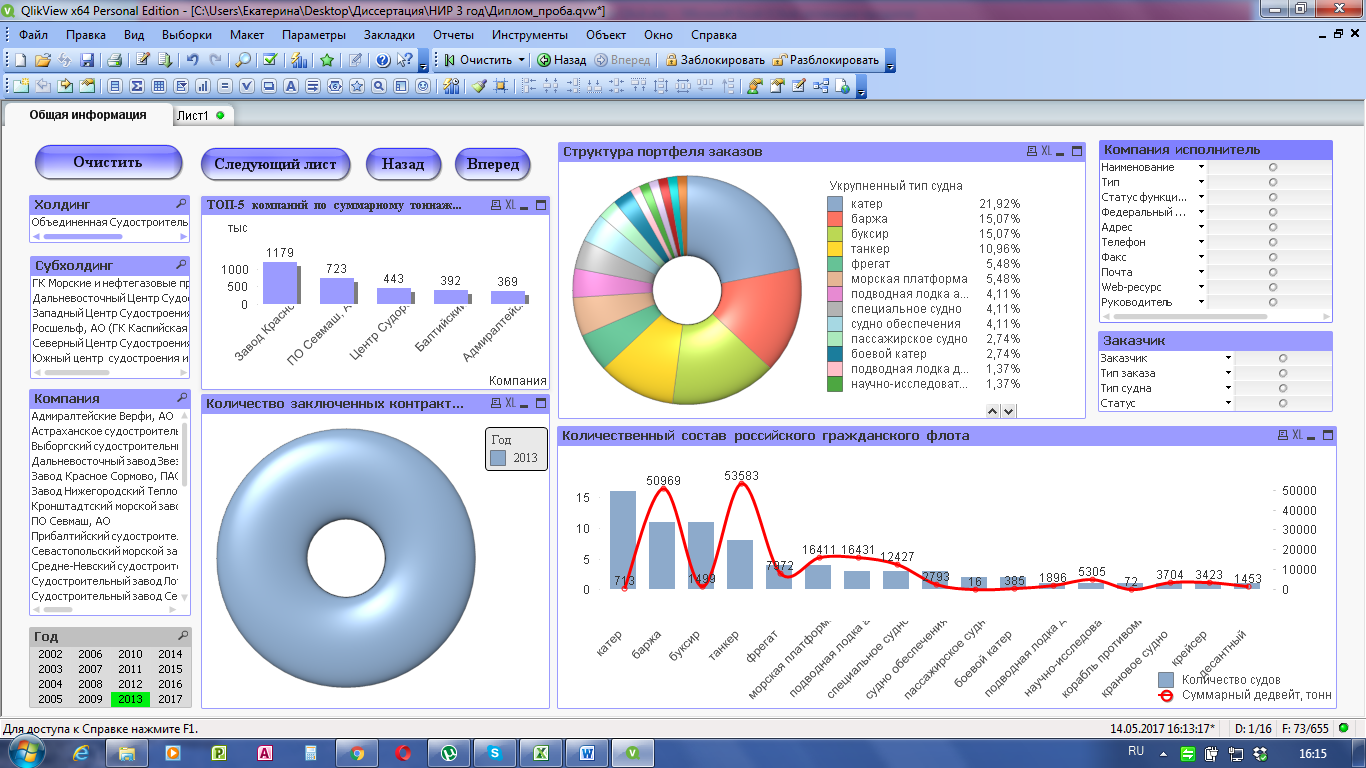
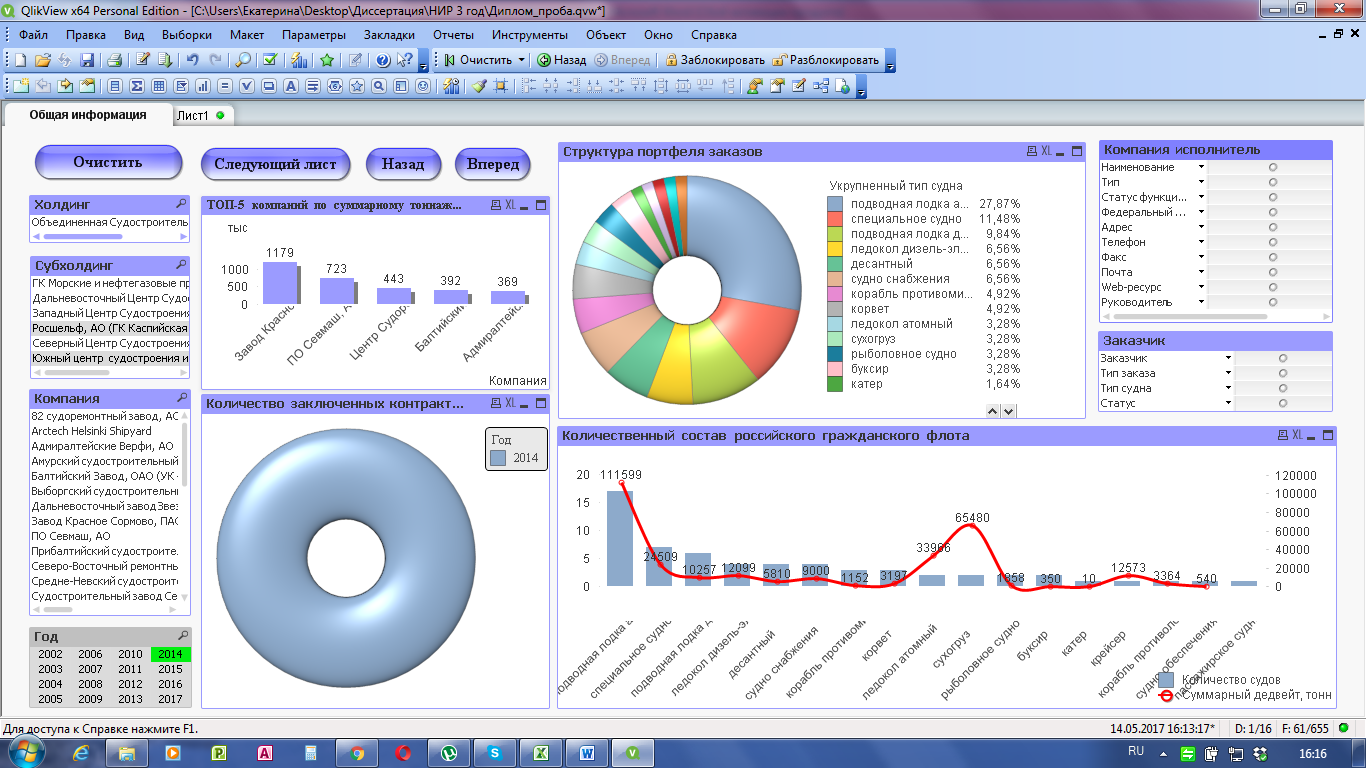
 

Рисунок 8. Структура портфеля заказов 2013 г. Рисунок 9. Структура портфеля заказов 2014 г.

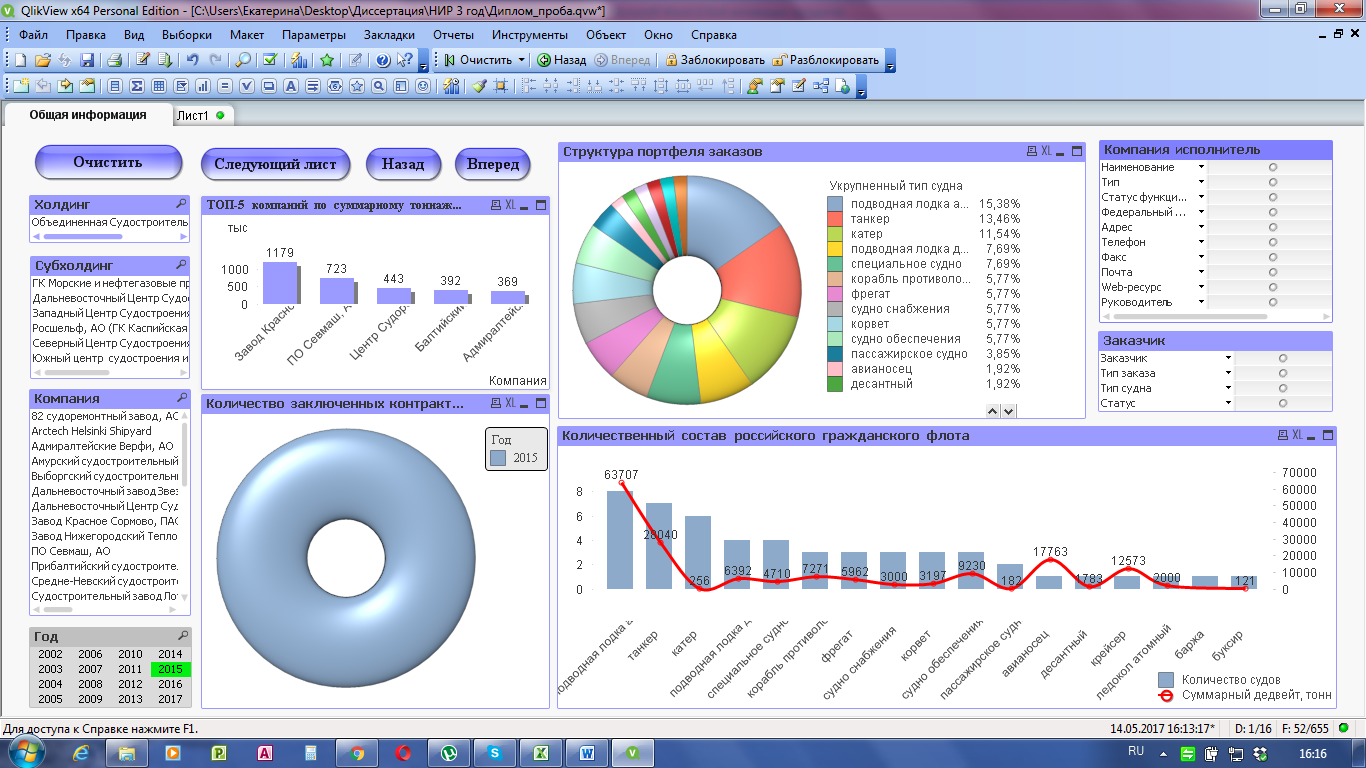
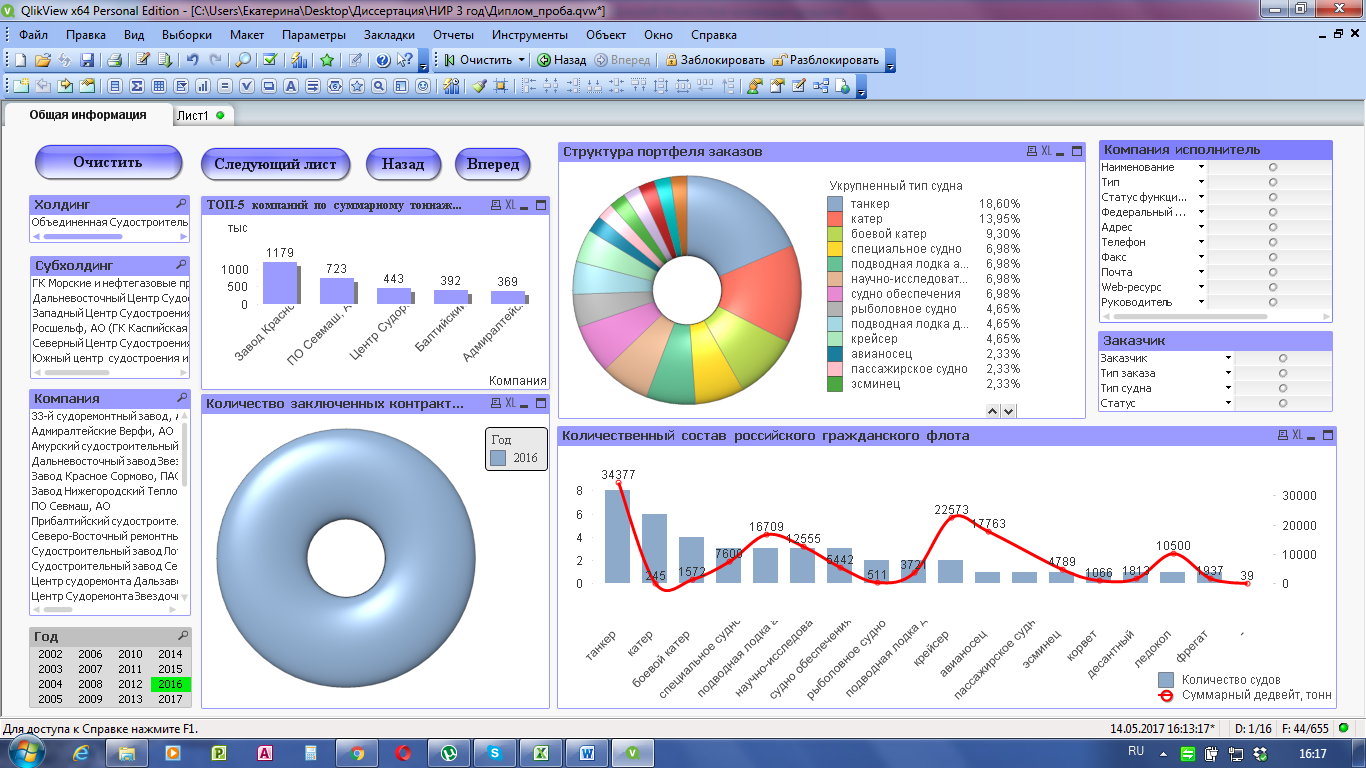
 

Рисунок 10. Структура портфеля заказов 2015 г. Рисунок 11. Структура портфеля заказов 2016 г.

1. **Количественный состав российского гражданского флота:**

Для отображения данного вопроса воспользуемся *Комбо* диаграммой. В комбо диаграмме можно сочетать элементы линейчатой диаграммы и линейного графика, например, одно выражение может быть представлено в виде столбцов, а другое — в виде линий, что и будет сделано в нашем случае.

Выполним следующие действия: на вкладке *Общие* выберем значок *Комбо диаграммы.* На вкладке *Выражения* наши поля *Count ([Enlarged type of ship])* и *Sum (Dedveit)* отобразятся списком. Для поля *Count ([Enlarged type of ship])* установим в опции *Параметры отображения* флажок *Столбец*. Во втором поле установим параметр *Символ* и *Линия* с отображением значений на точках графика. Данная диаграмма показывает соотношение количественного состава российского флота и его суммарного дедвейта относительно каждого из типов судов (рис.12).

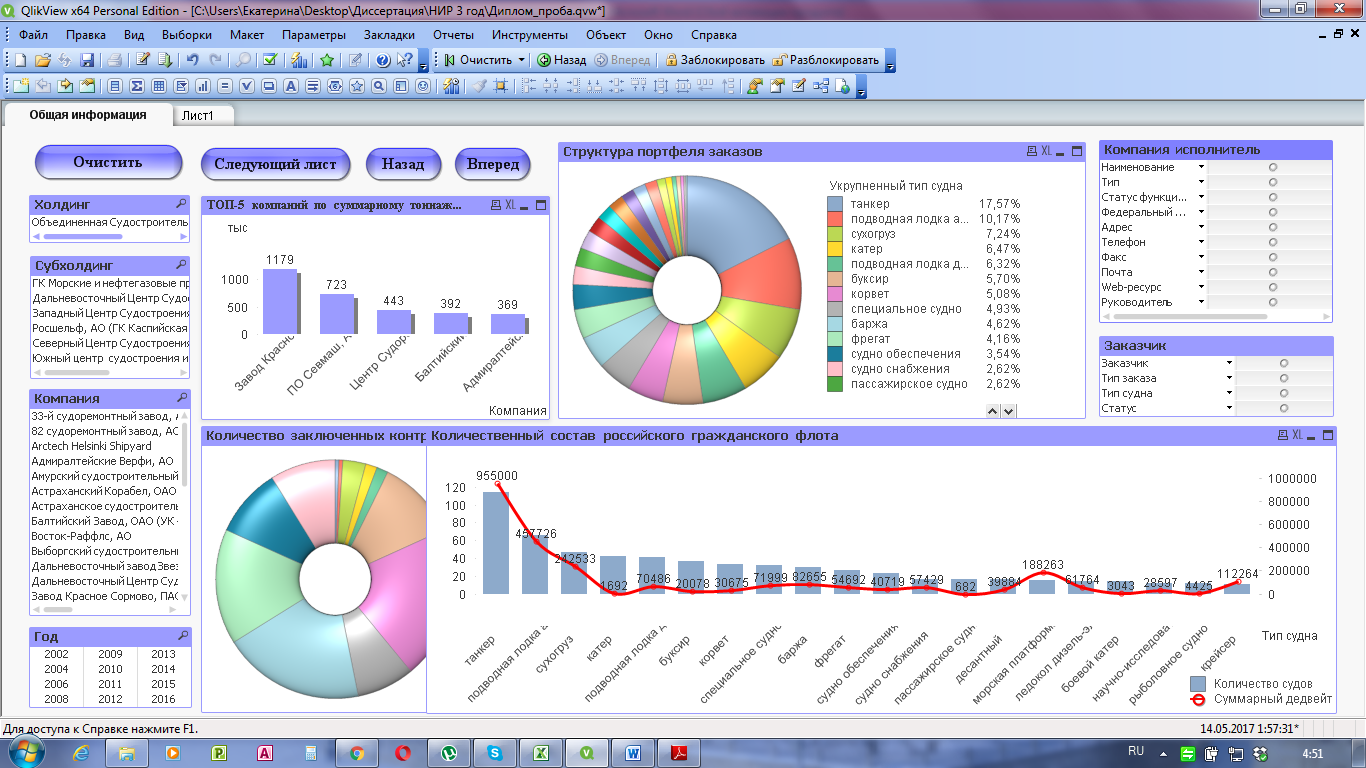


Рисунок 12. Количественный состав российского гражданского флота

Анализ данного вопроса интерес нам, в первую очередь, для понимания общей картины судостроительного рынка, для понимания ориентированности рынка и провозной способности морского транспортного флота, также для прогнозирования его пополнение новыми судами. Так, по информации ИА ИНФОЛайн, за период с 2010 до 2020 гг. в эксплуатацию должны быть введены транспортные суда общим дедвейтом более 734 тыс. тонн, а также должны быть построены и переданы заказчикам 56 единиц судов обеспечивающего флота.

1. **Определить топ самых крупных заказчиков и исполнителей РФ на рынке судостроения:**

Данная информация необходима для понимания расстановки сил в отрасли, ведущих компаний по судостроительству и ремонту, и суммарный тоннаж каждой из них (рис.13).

На вкладке *Измерения* из доступных полей выберем *Company* и добавим его в *Используемые измерения.* На вкладку *Выражения* добавляем вычисляемое поле *Sum (Tonnaj)*, которое вычисляет нам суммарный тоннаж спущенных на воду судов по каждой из компаний лидеров.

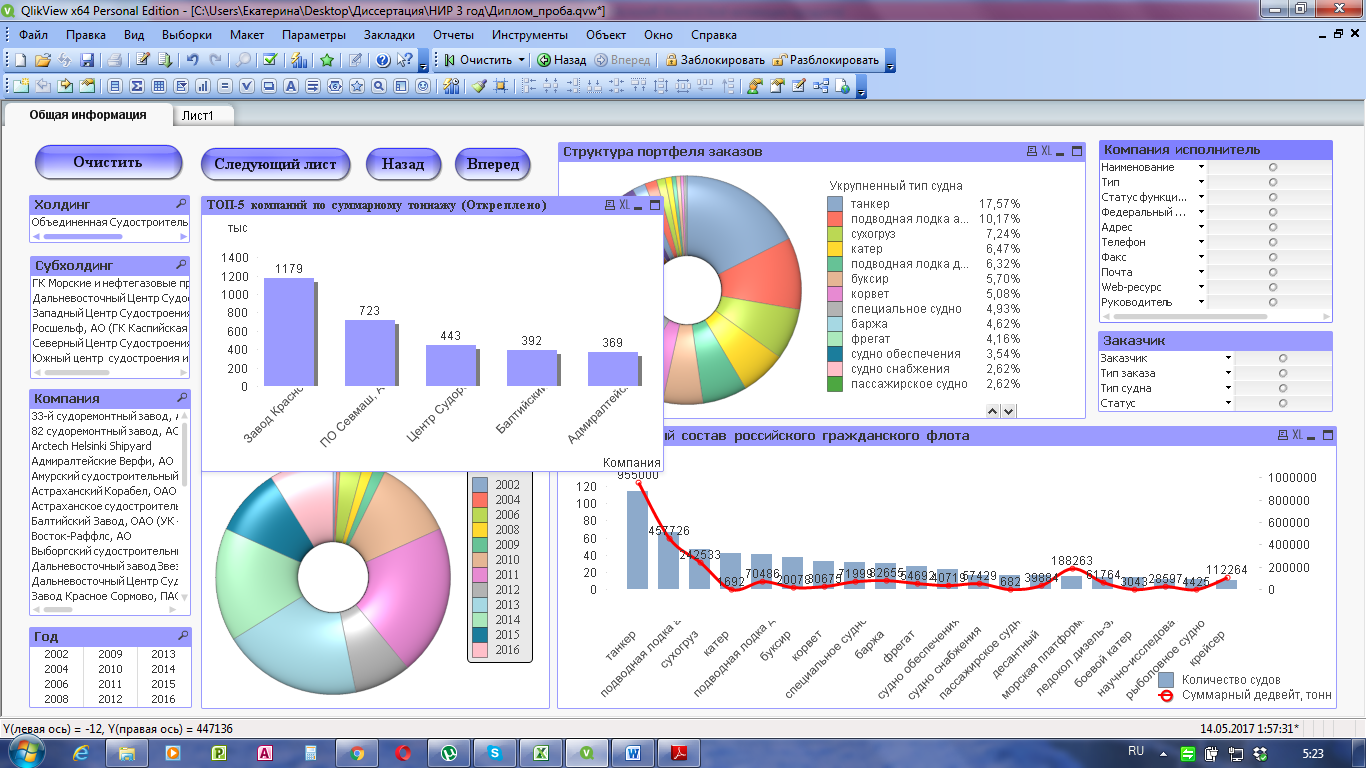


Рисунок 13. ТОП-5 компаний по суммарному тоннажу

1. **Динамика количества переданных заказчикам российскими верфями судов:**

Для построения такой диаграммы на вкладке *Измерения* выбираем поле *Type\_order* и добавляем его в используемые измерения. Во вкладке *Выражения* добавляем вычисляемое поле *Count (Year\_order)*, которое отображает нам общее количество сданных судов по годам. Сортируем *Y-значения* по убыванию и получаем диаграмму, представленную на рис.14.

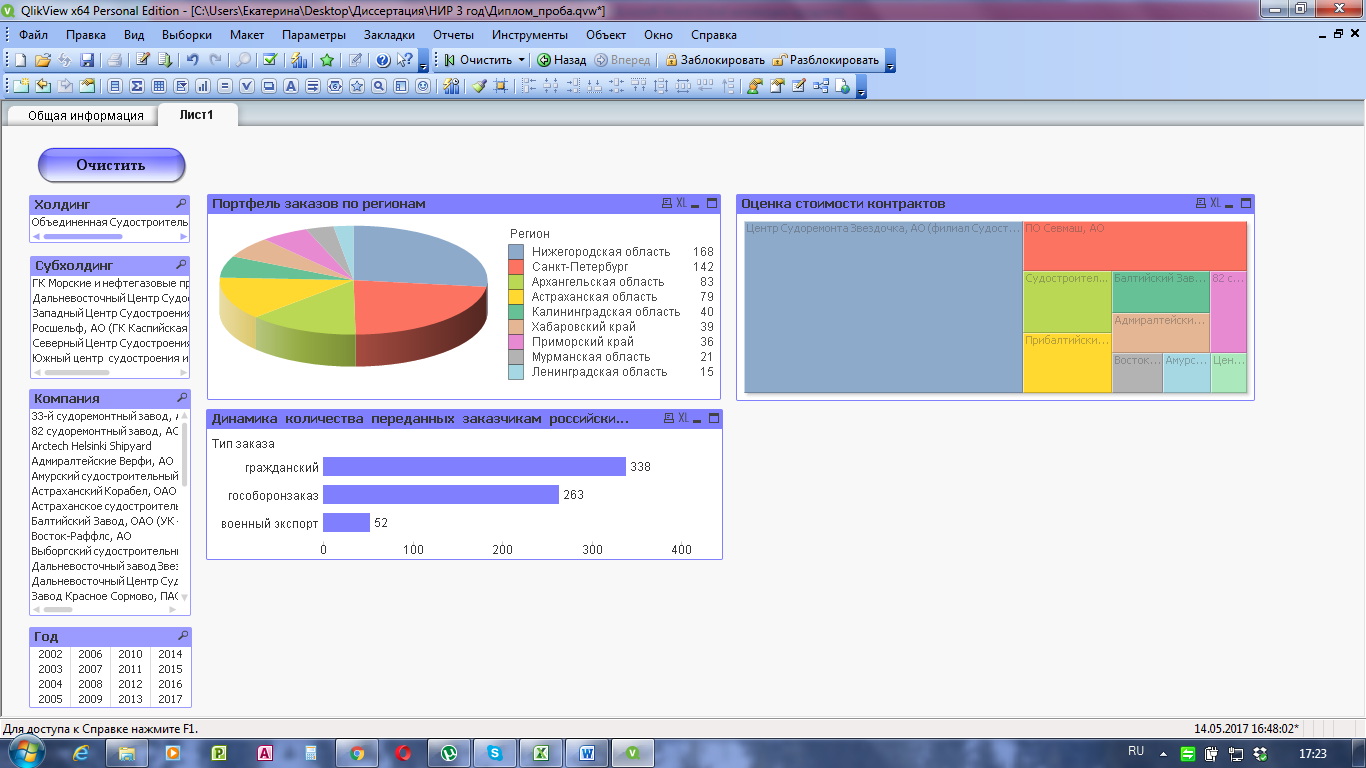


Рисунок 14. Динамика количества переданных заказчикам российскими верфями судов с 2002 года

В 2014 году впервые за последние шесть лет количество сданных военных кораблей оказалось больше, чем гражданских. Данное утверждение справедливо как для всей отрасли, так и для АО «ОСК», в срезе которой выполняется исследование. Военное кораблестроение в РФ уже с 2014 года опережает гражданское в стоимостном выражении за счет технологической сложности кораблей и масштабного гособоронзаказа, но гражданских заказов традиционно выполнялось больше, поскольку часто это серийные или простые в строительстве суда, которые не требуют долгих временных затрат.

По итогам работы корпорации «ОСК» 2014 года военные получили 44 крупных корабля и катеров, тогда как за этот же период судостроители сдали 17 гражданских судов. По итогам 2015 года военное кораблестроение также опережает гражданское. По данным специалистов ИА ИНФОЛайн, с учетом производственной программы крупнейших российских верфей в 2016-2017 гг. доля военного кораблестроения только продолжит расти. Отраслевые эксперты также подтверждают данный тезис, отмечая, что «перекос в сторону военного кораблестроения, очевидно, будет сохраняться в ближайшие годы не только на фоне большого гособоронзаказа, но и из-за макроэкономической ситуации в стране».

1. **Объём заключенных контрактов:**

Для расчета данного показателя во вкладке *Выражения* добавляем вычисляемое поле *Count (Year\_contract)*. Сортируем *Состояния* по убыванию и добавляем все необходимые настройки. Диаграмма представлена на рис. 15.

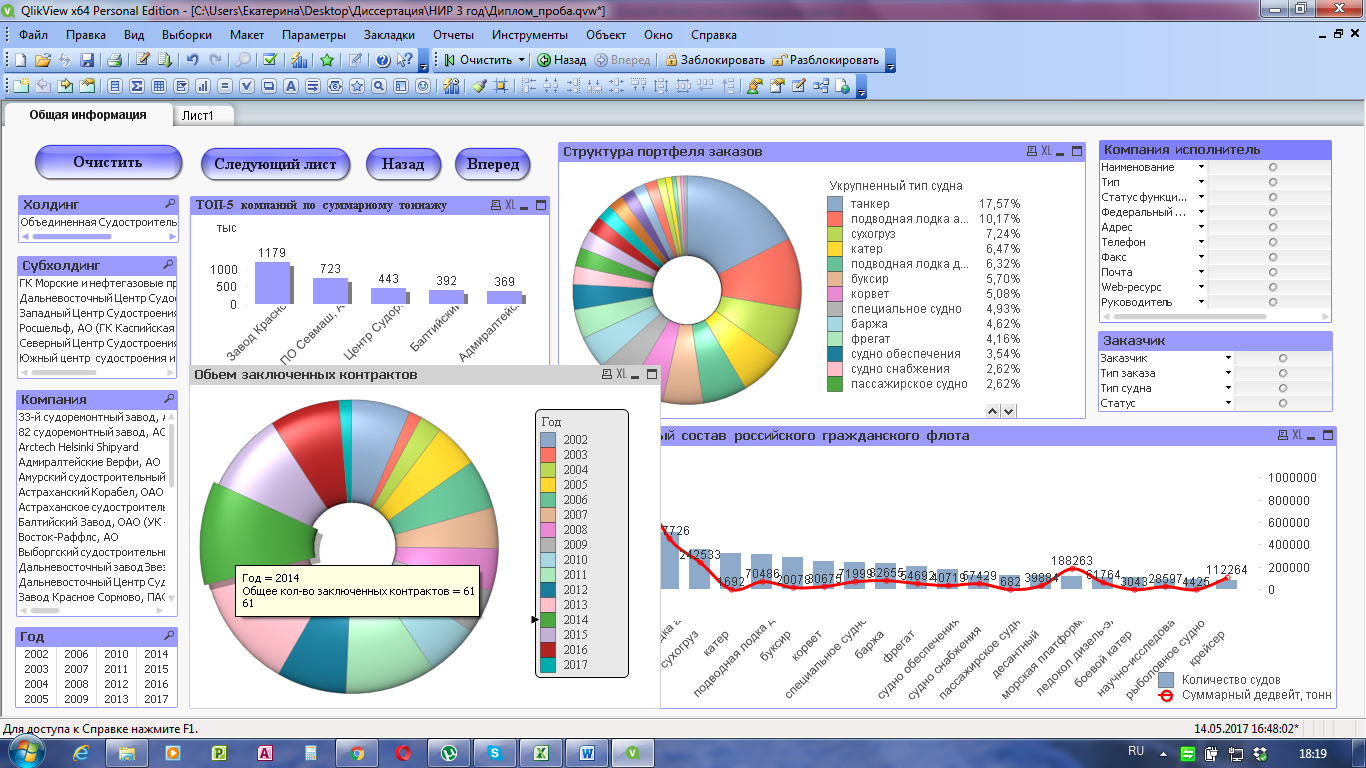


Рисунок 15. Объём заключенных контрактов

По данной диаграмме наглядно видны объёмы заключенных контрактов в каждый из определенных периодов. Также диаграмма фильтруется и по исполнителям данных контрактов, и по структуре портфеля заказов (укрупненным типам судов), так как QlikView поддерживает ассоциативный метод работы с данными. По диаграмме можно оценить, что объем заключенных контрактов на строительство крупных судов в 2015 году по сравнению с 2014 годом сократился в количественном выражении на 15%, что объясняется сложными рыночными условиями, так как инвесторы, по возможности, избегали строительства новых судов. Но также стоит отметить, что различные сегменты судостроительной отрасли демонстрировали разнонаправленную динамику.

1. **Определение самых частых и суммарно выгодных заказчиков:**

Для проведения анализа по этому вопросу было принято решение создать мультисписок. В его поля были добавлены следующие ячейки – *Customer, Type\_order, Enlarged type of ship, Status*.

Данный мультисписок при выборе компании заказчика фильтрует все объекты листа относительно нашей выбранной компании. Мы можем наглядно оценить структуру портфеля заказов в разрезе этой компании, произвести оценку стоимости каждого из контрактов и др. (рис.16).

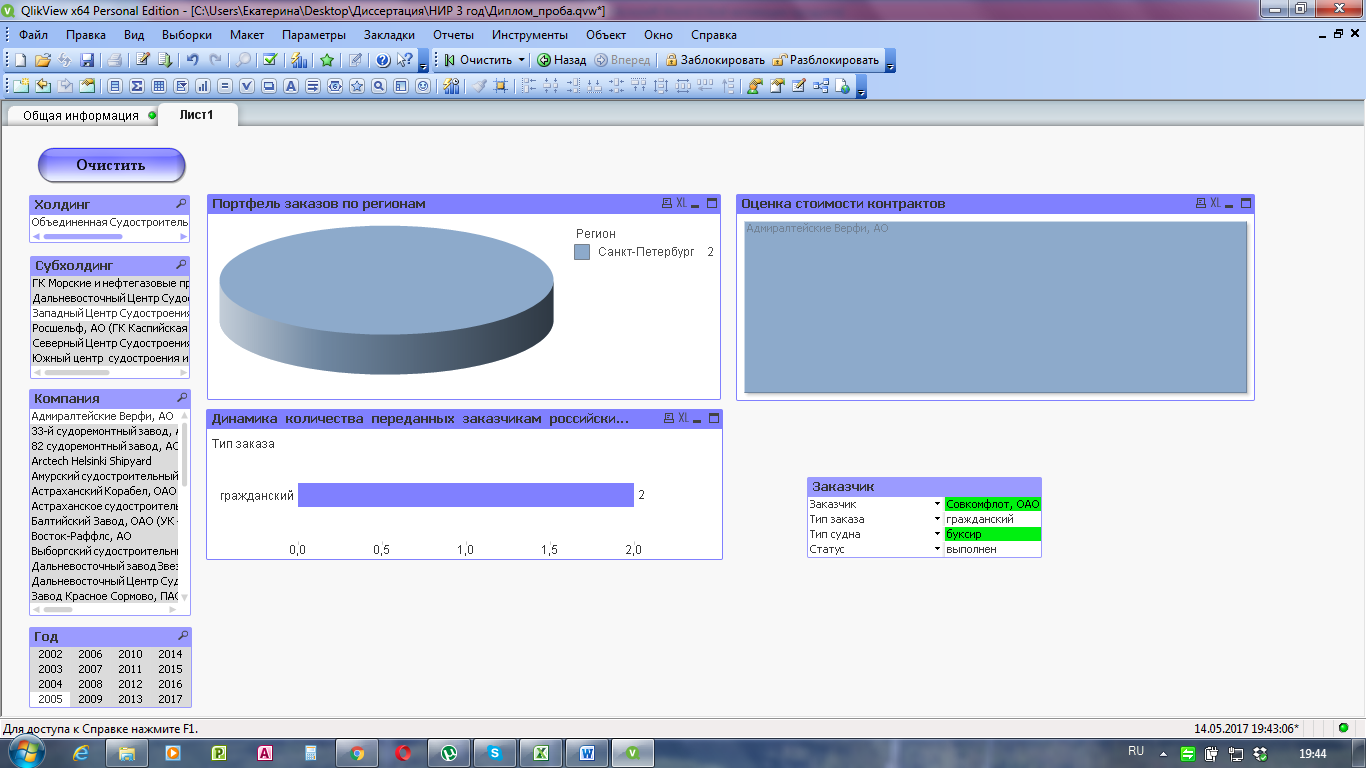


Рисунок 16. Объект листа - мультисписок «Заказчик»

1. **Оценка стоимости контрактов:**

Для визуализации этого вопроса использовалась *Блочная* диаграмма. Полем фильтрации выбираем *Company*, а на вкладке *Выражения* высчитываем общую сумму контрактов по каждой из компаний – *Sum (Value)*. При наведении курсора на каждый из блоков диаграммы, нам отображается вычисляемое значение (рис. 17).

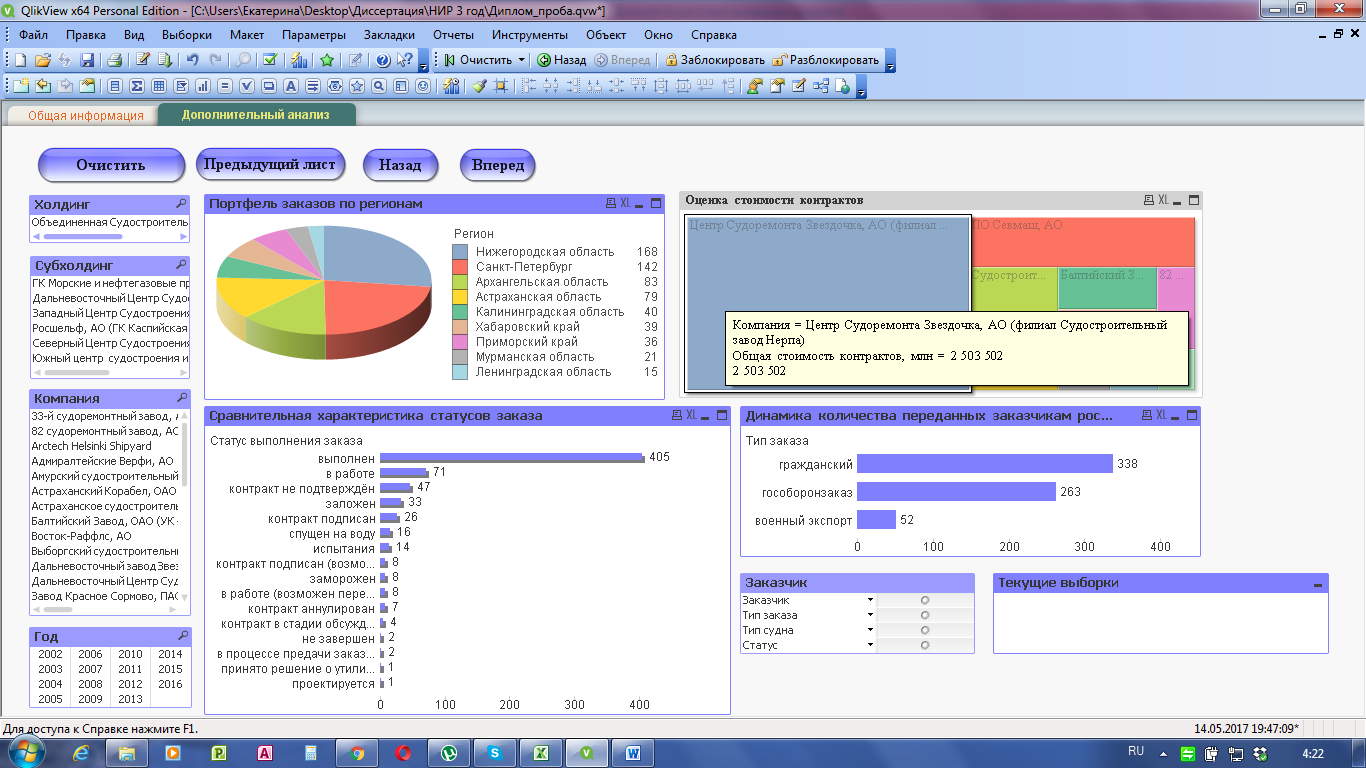


Рисунок 17. Оценка стоимости контрактов

1. **Сравнительная таблица статусов контрактов:**

Данная диаграмма наглядно показывает, какое количество контрактов выполнено той или иной компанией, какое количество контрактов находится еще в работе, а какое и вовсе аннулировано. Фильтруя списки по годам или по типам судов, мы мгновенно получаем необходимую для дальнейшего анализа информацию (рис. 18).

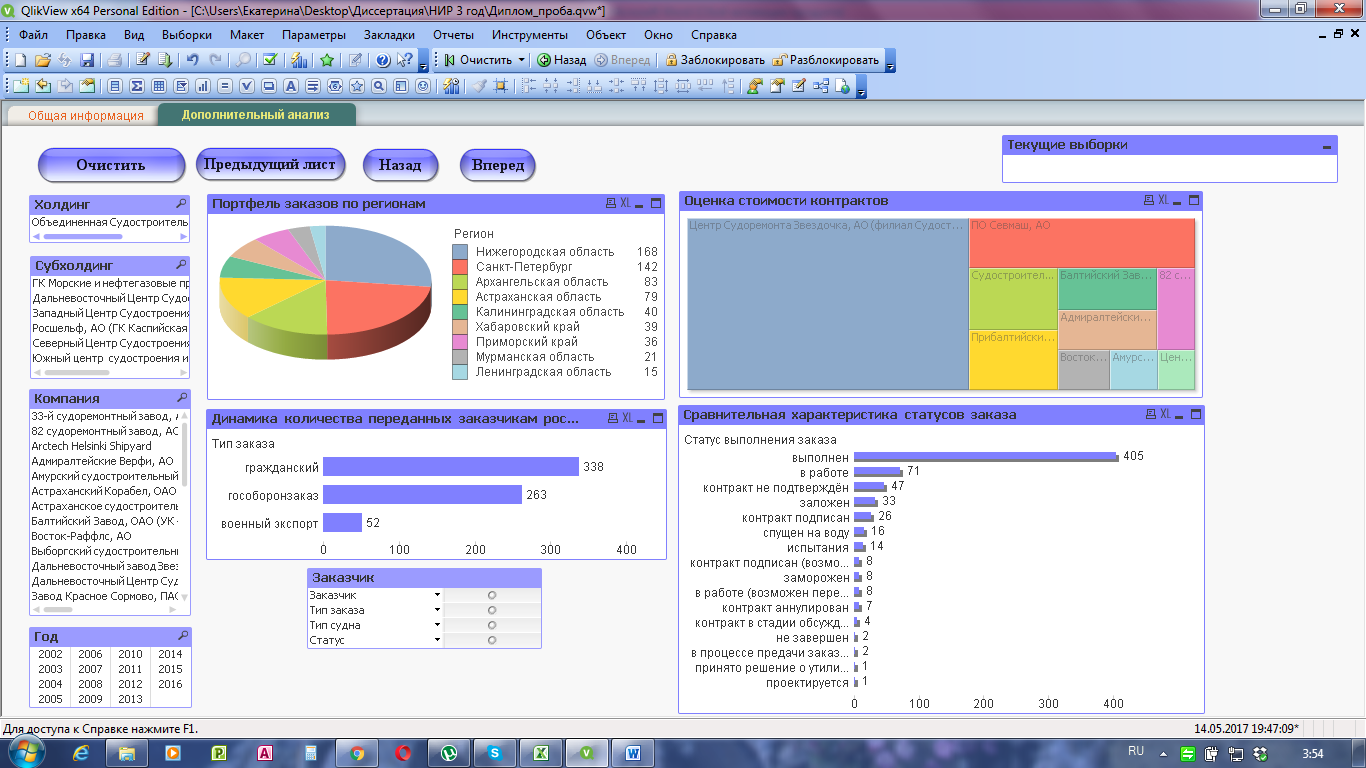


Рисунок 18. Количественная характеристика статусов выполнения заказа

### Краткое руководство для пользовательской работы с приложением в BI-системе

Все полученные в ходе работы обьекты можно объединить в группы или так называемые дашборды по тематике исследования тех или иных вопросов. Дашборд в переводе на русский означает панель индикаторов, шкал, приборов. Термин пришел из английского и ранее употреблялся в контексте автомобиля, сейчас же относится к различным механизмам и системам, включая компьютеры[[10]](#footnote-10). В рамках информационных систем дашборды понимаются как пользовательские интерфейсы взаимодействия между пользователем и системой и наоборот. Тщательно разработанный дашборд – мощный аналитический инструмент, который может быть одновременно красивым и удобным для пользователя.

Дашборды необходимы для лучшего визуального восприятия проблемы, так как можно объединить рабочие страницы, которые направлены на объект исследования с разными точками зрения.

Для улучшения навигации по приложению был прописан макрос для изменения цвета ярлыка активного листа. Он направлен на улучшение по дополнительной навигации и интерактивности при работе с приложением. Макрос был прописан в триггерах каждого листа в разделе *OnActivateSheet*. Полный текст кода для макроса прописан в Приложении 2.

Полные экраны дашбордов представлены в Приложении 3.

Данные дашборды носят больше ознакомительный характер, с целью показать возможность и простоту визуализации и вычленения информации из разнородного массива данных. Данный алгоритм применим к разным сферам и данным.

# ЗАКЛЮЧЕНИЕ

Таким образом, в данной работе была обоснована актуальность выпускной квалификационной работы, поставлены цель и задачи данного исследования, выделены объект и предмет исследования и даны их характеристики, дано описание понятия типового продукта и процесса его разработки, а также выделены основные сущности для дальнейшей автоматизации.

На основании данных ИА «ИНФОЛайн-Аналитика» сформирован алгоритм, включающий аналитический план работы с данными, что может позволить существенно снизить затраты на проведение маркетинговых исследований рынка судостроения.

Формирование подобного алгоритма - процесс трудоемкий, и в конечном счете, может не оправдать затрат на себя, так как существует вероятность того, что при масштабируемости или передаче на исполнение он окажется не пригодным по ряду причин, например, воспримется негативно будущими исполнителями или начальством. Тем не менее, стоит отметить, что проделанная работа может принести свои плоды, как минимум, по факту делегирования полномочий при исполнении, а также повышения эффективности работы.

Получившееся бизнес-приложение, как результат разработанного алгоритма, может служить начальным макетом для дальнейшего его расширения остальными массивами данных, для получения более широкой картины происходящего в отрасли. Также приложение поможет в выборе информации интересной для потенциальных покупателей продукта и информации, оставленной непосредственно для коммерческого продукта.

Результаты проделанной работы подтверждают успешность выполнения поставленных целей и задач исследования.

# Список используемых источников

**Монографии, учебники, учебные пособия**

1. Барсегян, А. А. «Анализ данных и процессов: учеб. пособие» /

. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009. — 512 с.;

1. В.Л. Аббакумов, Т.А. Лезина. Бизнес-анализ информации. Статистические методы. - М.: ЗАО «Издательство «Экономика», 2009. — 374; с. — (Учебники экономического факультета СПбГУ). ISВN 978-5-282-02918-5;
2. Паклин Н.Б. Орешков В.И. Бизнес-аналитика: от данных к знаниям: Учебное пособие. – СПб.: Питер, 2013. – 704с. ISBN 978-5-459-00717-6;
3. Пахомова Н.В., Мотовилов О.В, Маслова Е.В., Кузнецова А.С. - Методические указания по подготовке и защите магистерской диссертации по основным образовательным программам магистратуры по направлениям 38.04.01 «Экономика», 38.04.05 «Бизнес-информатика», 38.04.08 «Финансы и кредит» очной формы обучения. СПбГУ;
4. Яу Н. Искусство визуализации в бизнесе. Как представить сложную информацию простыми образами / Пер. с англ. Светланы Кировой. М., 2013;
5. QlikView Версия 11.2, SR5 для Microsoft Windows. Учебное пособие. QlikTech International AB, 2013.

**Статьи в журналах и других периодических изданиях**

1. Иванов П.Д., Вампиловв В.Ж. Технологии Big Data и их применение на современном промышленном предприятии. Инженерный журнал: наука и инновации, 2014, вып. 8. URL: http://engjournal.ru/catalog/it/asu/1228.html
2. Официальный сайт международной компании «Прогноз», статья: «Российский рынок BI в 2016 году: чего ожидать и что делать» – URL: http://www.prognoz.ru/blog/market/russian-bi-market-2016/;
3. Big Data: What It Is and Why You Should Care. White Paper. – IDC, 2011;

**Нормативно-правовые документы**

1. ГОСТ 2.105–95. Единая система конструкторской документации. Общие требования к текстовым документам [Электронный ресурс] // Федер. агентство по техн. регулированию и метрологии – URL: http://protect.gost.ru/v.aspx?control=8&baseC=-1&page=0&month=-%201&year=-1&search=&RegNum=1&DocOnPageCount=15&id=126445;

**Интернет – ресурсы**

1. Блог об использовании Microsoft Power BI для бизнес-аналитики и визуализации данных – URL: http://powerbirussia.ru;
2. Документация Business Studio – URL: http://www.businessstudio.ru/wiki/;
3. Интернет ресурс «Открытые системы» – URL: http://www.osp.ru/os/2013/06/13036849/;
4. «ИНФОЛайн» — информационно-аналитическое агентство. [Электронный ресурс]. – URL: http://ИНФОЛайн.spb.ru/;
5. Николай Павлов, Planetaexcel, 2006-2017. – URL: http://www.planetaexcel.ru/;
6. Отраслевой портал морского и речного судостроения РФ – URL: http://sudostroenie.info/;
7. Официальный сайт консалтинговой компании Gartner. – 2016. – URL: https://www.gartner.com/doc/reprints?id=1-2XXKCD7&ct=160204&st=sb;
8. Официальный сайт компании BI Consult – URL: http://biconsult.ru/;
9. Официальный сайт Минпромторга России – URL: http://minpromtorg.gov.ru/;
10. Официальный сайт разработчиков Microsoft – URL: https://msdn.microsoft.com/ru-ru;
11. Портал BI-платформы QlikView – URL: http://www.qlik.com/ru-ru;
12. Российский Речной Регистр – URL: http://www.rivreg.ru/activities/;
13. Судостроительный портал «Корабел.ру» – URL: https://www.korabel.ru/;

# Приложение 1. Код скрипта загрузки

**LOAD** OKPO,   
 INN,   
 Company,   
 Type,   
 Stat,   
 FO,   
 Region\_contacts,   
 Adress,   
 Telefon,   
 Faks,   
 @,   
 Web,   
 Name\_director,   
 Position  
FROM  
[C:\Users\Екатерина\Desktop\Диссертация\Данные\Контакты\_Компании.xlsx]  
(ooxml, embedded labels, table is [Контакты РФ]);  
  
**LOAD** Holding,   
 Subholding,   
 Company,   
 Country,   
 Region\_baza,   
 [Federal district],   
 Customer,   
 Flot,   
 Country\_customer,   
 Type\_order,   
 Remont,   
 [Enlarged type of ship],   
 Type\_ship,   
 [UNI-project],   
 Ice\_class,   
 [IMO / РРР],   
 Name\_ship,   
 Dedveit,   
 Tonnaj,   
 [Ded/Ton],   
 Year\_order,   
 Quarter\_order,   
 Status,   
 Year\_contract,   
 Quarter\_contract,   
 Year\_start,   
 Quarter\_start,   
 Year\_water,   
 Quarter\_water,   
 Datef\_update,   
 Period,   
 Update,   
 [Calculation principle]  
FROM  
[C:\Users\Екатерина\Desktop\Диссертация\Данные\РАБОЧАЯ БАЗА Исправл для Qlick.xlsx]  
(ooxml, embedded labels, table is [Данные РФ]);

Стоимость:   
**LOAD** Company,   
 Order\_name,   
 Value  
FROM  
[C:\Users\Екатерина\Desktop\Диссертация\Данные\Стоимость\_контракта.xlsx]  
(ooxml, embedded labels, table is Value);

# Приложение 2. Код макроса изменения цвета ярлыка с помощью VB макроса

1. sub ChangeColorActive
3. 'Обходим все листы приложения и задаем им "неактивный" цвет
4. for i=0 to ActiveDocument.NoOfSheets-1
6. set vSheet = ActiveDocument.GetSheet(i)
8. set vSheetProp = vSheet.GetProperties
10. vSheetProp.TabAttr.BgColor.PrimaryCol.Col = RGB(241,238,229)
12. vSheetProp.TabAttr.FgColor.PrimaryCol.Col = RGB(232,98,32)
14. vSheetProp.TabAttr.Mode = 2
16. vSheet.SetProperties vSheetProp
18. next
20. 'Определяем активный лист и задаем ему "активный" цвет
21. set vSheet = ActiveDocument.ActiveSheet
23. set vSheetProp = vSheet.GetProperties
25. vSheetProp.TabAttr.BgColor.PrimaryCol.Col = RGB(67,117,115)
27. vSheetProp.TabAttr.FgColor.PrimaryCol.Col = RGB(250,245,120)
29. vSheetProp.TabAttr.Mode = 2
31. vSheet.SetProperties vSheetProp
33. End sub

# Приложение 3. Дашборды приложения

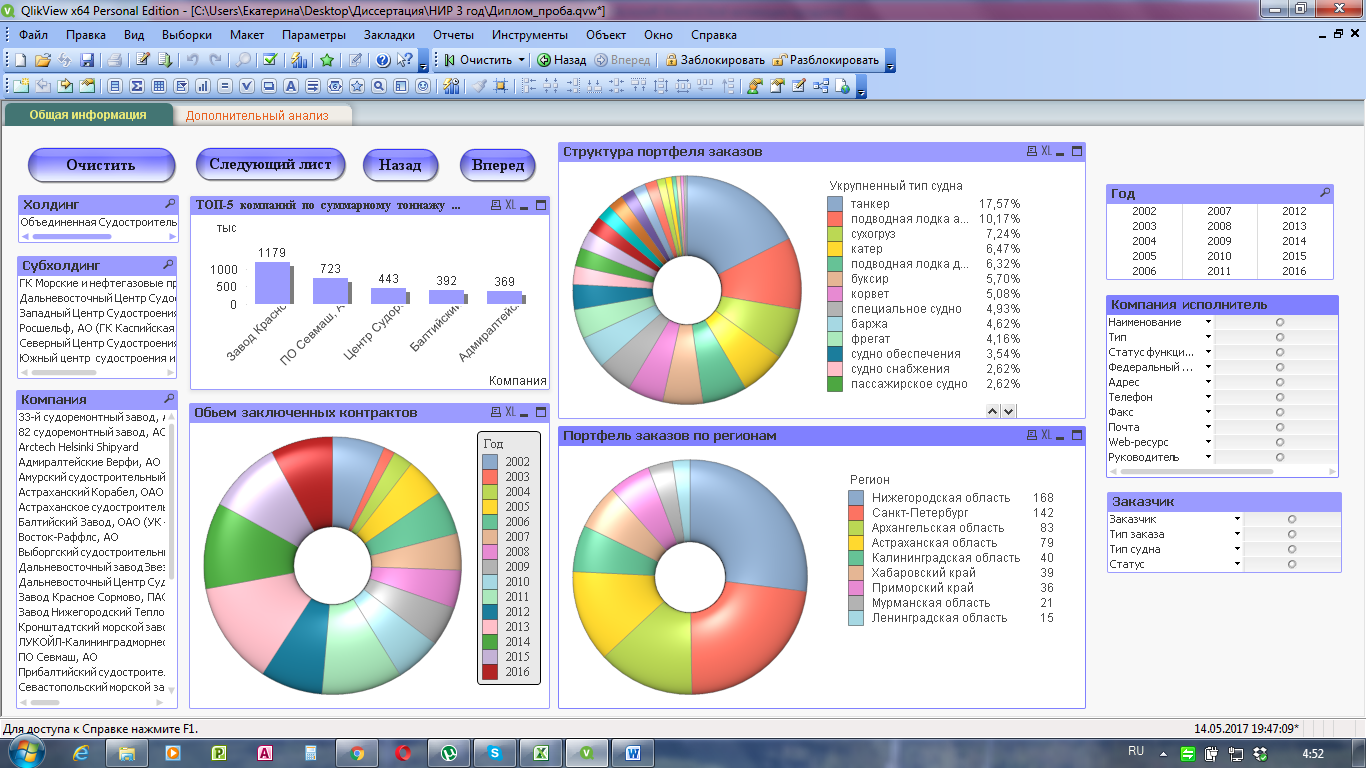


Рисунок 19. Лист «Общая информация»

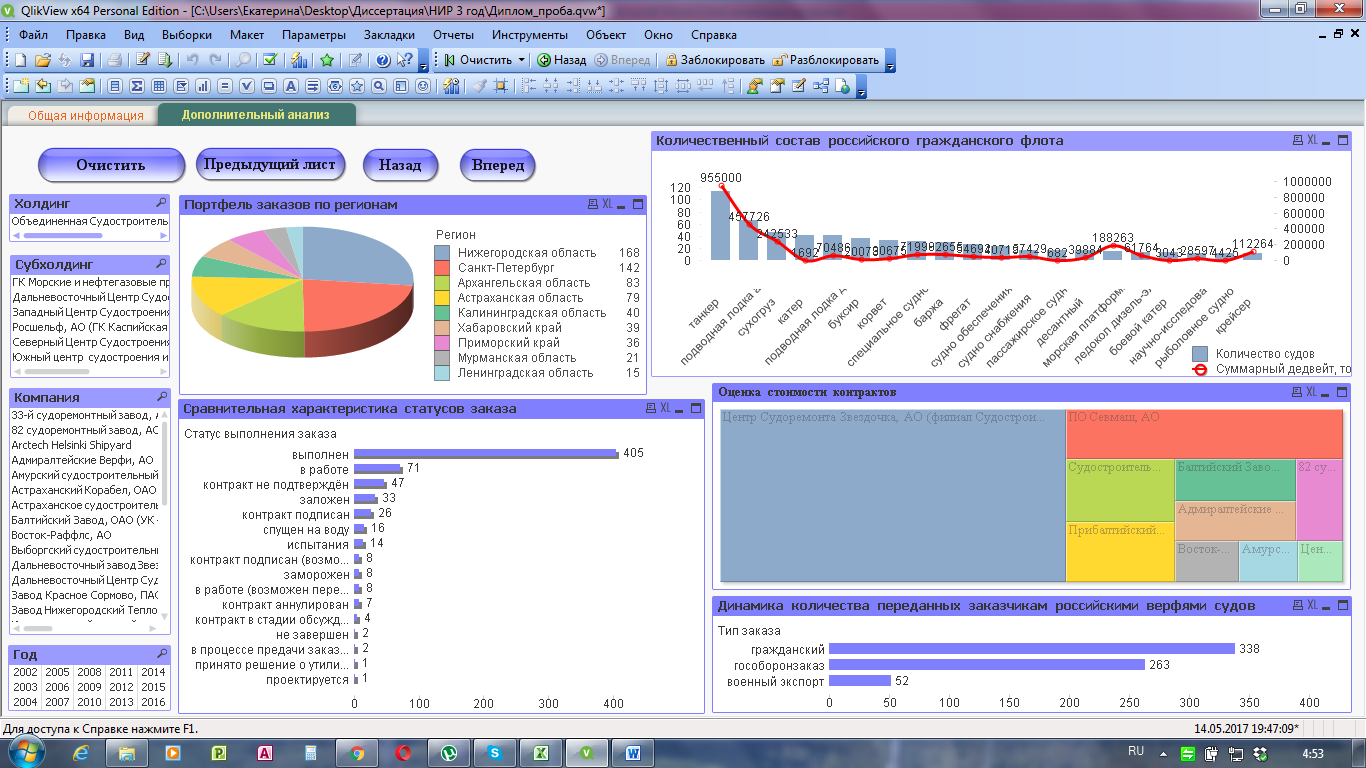


Рисунок 20. Лист «Дополнительный анализ»

# Приложение 4. Структура анкеты

**Анкета для судостроительных верфей «INFOLine Shipbuilding Russia»**

Название компании:

Ф.И.О. и должность руководителя:

Телефон:

Факс:

E-mail:

Выполненные за 2012-2015 гг. заказы ПОДТВЕРЖДАЮ: **да**  🞏 **нет**  🞏

| **Заказчик** | **Тип судна  и номер проекта** | **Название, строительный  номер** | **Дедвейт в море, тонн** | **Тоннаж, тонн** | **Год**  **сдачи** | **Квартал сдачи** |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

*Если выполнено больше, чем строк в таблице, просим выслать отдельным файлом*

Портфель заказов на март 2015 года ПОДТВЕРЖДАЮ

| **Заказчик** | **Тип судна  и номер проекта** | **Название, строительный  номер** | **Дедвейт в море, тонн** | **Тоннаж, тонн** | **Текущий статус заказа\*** | **Плановый срок сдачи** |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

*\* контракт в стадии обсуждения, контракт подписан, заложен, в работе, спущен на воду, испытания, сдан заказчику, заморожен*

*Если заказов больше, чем строк в таблице, просим выслать отдельным файлом*

Информация о лице, заполнившем анкету

Ф.И.О.: Должность:

Телефон / факс: Электронная почта:

**Мы готовы ответить на любые Ваши вопросы по телефонам: (812) 322-68-48, доб. 310 или 206,   
или по электронной почте** [**shipbuilding.infoline@yandex.ru**](mailto:shipbuilding.infoline@yandex.ru?subject=Анкета%20для%20участия%20в%20рейтинге%20INFOLine%20Shipbuilding%20Russia%20TOP)

1. Доля предприятий, расположенных в каждом из округов, рассчитана в количественном выражении на основании данных Аналитической базы "Заказы на строительство судов на верфях РФ: Итоги 2015-2016 годов". Доля объема производства рассчитана на основании данных Аналитической базы "Заказы на строительство судов на верфях РФ: Итоги 2015-2016 годов" о тоннаже сданных в России судов в период за 2010-1 пол. 2016 гг. Данные о количестве кадров приведены по данным СМИ. [↑](#footnote-ref-1)
2. Отраслевой портал «Корабел.ру». 2002-2017. [↑](#footnote-ref-2)
3. Федеральное автономное учреждение Российский Речной Регистр. [↑](#footnote-ref-3)
4. Интернет-проект «Корпоративный менеджмент», 1998–2017. [↑](#footnote-ref-4)
5. Николай Павлов, Planetaexcel, 2006-2017 (http://www.planetaexcel.ru/). [↑](#footnote-ref-5)
6. Официальный сайт разработчиков Microsoft; [↑](#footnote-ref-6)
7. Документация [Business Studio](http://www.businessstudio.ru/wiki/docs/current/doku.php/start) - http://www.businessstudio.ru/wiki/ [↑](#footnote-ref-7)
8. *H. P. Luhn.* A Business Intelligence System. *IBM Journal* (Октябрь 1958) [↑](#footnote-ref-8)
9. TAdviser – это крупнейшая в России база знаний о технологиях, ИТ-проектах и профессионалах отрасли. http://www.tadviser.ru/index.php/Статья:Business\_Intelligence\_(рынок\_России) [↑](#footnote-ref-9)
10. The Web's Largest Resource for Definitions & Translations. Электронный источник: <http://www.definitions.net/definition/Dashboard> [↑](#footnote-ref-10)