

МИНОБРНАУКИ РОССИИ

Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Е. Е. КОТОВА Д. П. ПЛАХОТНИКОВ

**ВВЕДЕНИЕ В АНАЛИТИКУ
БОЛЬШИХ ДАННЫХ**

Учебно-методическое пособие

Санкт-Петербург
Издательство СПбГЭТУ «ЛЭТИ»
2020

УДК 004.67
ББК 32.972.13
В24

Котова Е. Е., Плахотников Д. П.

В24 Введение в аналитику больших данных: учеб.-метод. пособие. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2020. 48 с.

ISBN 000-0-0000-0000-0

Содержит методы базовой аналитики больших данных с помощью Microsoft Excel и Qlik Sense. Изучаются расширенные способы аналитики с помощью визуализации данных, построения листа прогноза, поиска оптимального решения с помощью Microsoft Excel. Предлагаются для изучения общие понятия и концепции, знакомство с аналитической платформой Qlik Sense, загрузкой данных, созданием приложений, способами визуализации данных и выполнения конкретных процедур.

Предназначено для подготовки студентов-бакалавров, обучающихся по направлениям подготовки: 09.03.02 – «Информационные системы и технологии», 27.03.04 – «Управление в технических системах (Компьютерные интеллектуальные технологии управления в технических системах)».

УДК 004.67
ББК 32.972.13

Рецензент к-т техн. наук А. А. Большаков (ООО «Газинформсервис»)

Утверждено
редакционно-издательским советом университета в качестве
учебно-методического пособия

ISBN 000-0-0000-0000-0

© СПбГЭТУ «ЛЭТИ», 2020

СОДЕРЖАНИЕ

| | |
|------------------------------|----|
| Лабораторная работа №1 | 4 |
| Лабораторная работа №2 | 13 |
| Лабораторная работа №3 | 21 |
| Лабораторная работа №4 | 29 |
| Лабораторная работа №5 | 33 |
| Лабораторная работа №6 | 41 |

Лабораторная работа №1

ВИЗУАЛИЗАЦИЯ ДАННЫХ С ГЕОГРАФИЧЕСКОЙ ПРИВЯЗКОЙ В EXCEL

Цель работы:

- ознакомиться со способами переноса информации в Microsoft Excel;
- изучить способы подготовки информации;
- изучить метод визуализации данных.

Минимальная необходимая версия – Microsoft Excel 2016 (либо установка блока инструментов, описанная в подразделе 1.1).

1.1. Включение блока инструментов для старых версий

Рассматриваемый пакет относится к категории надстроек. Соответственно, для включения пакета необходимо перейти в меню надстроек. Эта процедура выполняется следующим образом:

1. Необходимо открыть вкладку «Файл», расположенную в верхней части ленты интерфейса;
2. В левой стороне открывающегося меню найти раздел «Параметры» и кликнуть по нему;
3. В левой части окна открыть категорию надстроек (вторая снизу в списке), выбрать соответствующий пункт;
4. В раскрывающемся диалоговом меню найти пункт «Управление», кликнуть по нему левой кнопкой мыши;
5. Данное действие вызывает на экран диалоговое окно, в котором требуется выбрать раздел надстроек «Надстройки Excel», поменять его на обозначенное;
6. Нажать на экранную кнопку «Перейти» в разделе надстроек. В правой части выпадает список надстроек, которые устанавливает программа (рис. 1.1).

Далее необходимо активировать функции, предоставляемые надстройкой пакета:

1. В перечне надстроек, выпавшем после последовательного выполнения предыдущих операций, необходимо поставить знак «птички»

напротив раздела «Microsoft Power Map for Excel».

2. Выбрав активацию пакета, необходимо нажать клавишу «Ок», расположенную в верхней правой части диалогового окна (рис. 1.2).

После выполненных действий во вкладке «Вставка» должен появиться элемент «Карта» (рис. 1.3). Если данного элемента не появилось или произошла ошибка при установке компонента, можно воспользоваться онлайн надстройкой «Карты Bing» (рис. 1.4). В данном случае часть работы не может быть выполнена.

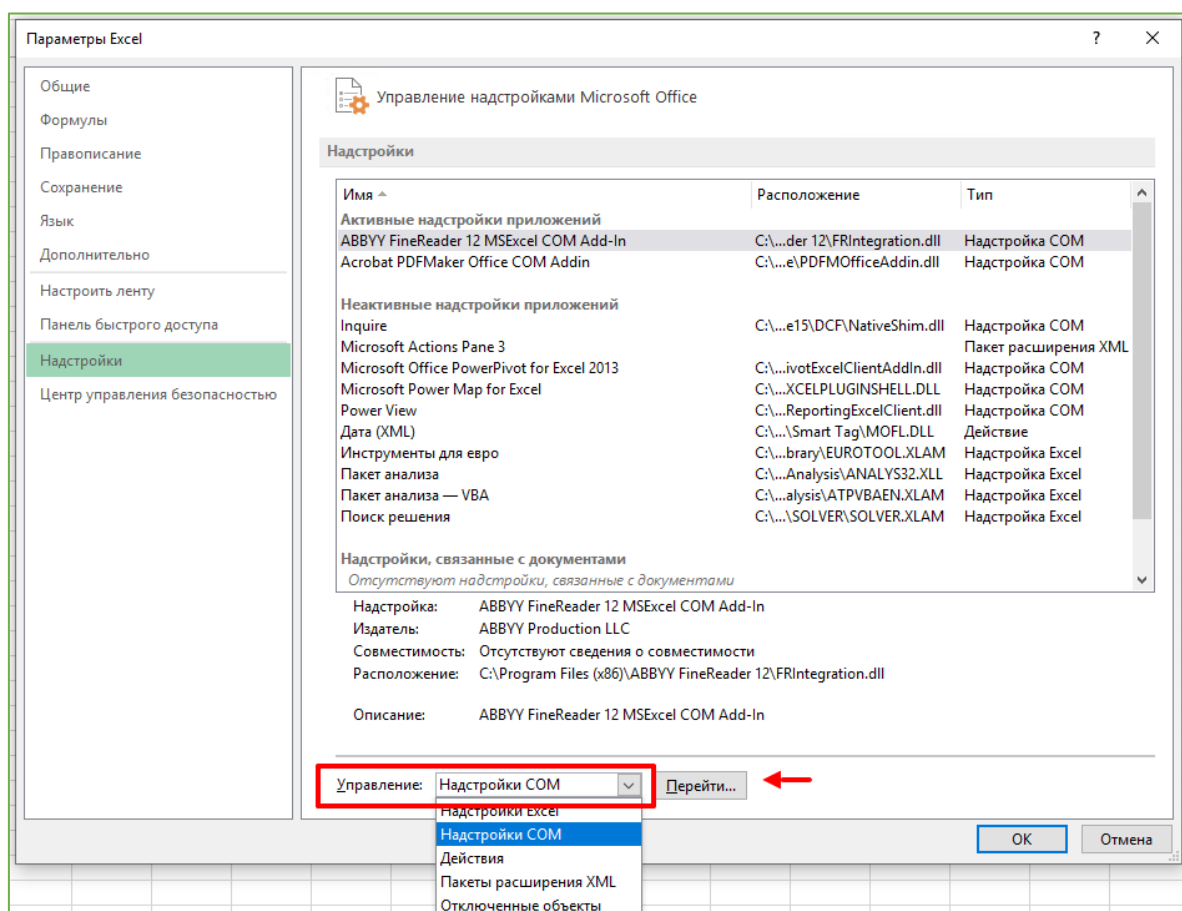


Рис. 1.1. Поиск пакета в надстройках Excel

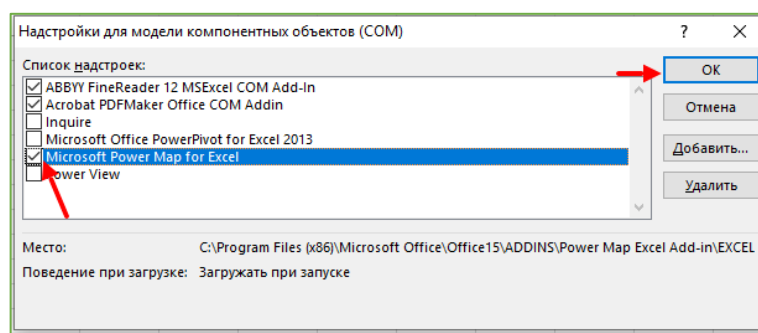


Рис. 1.2. Поиск пакета в надстройках Excel

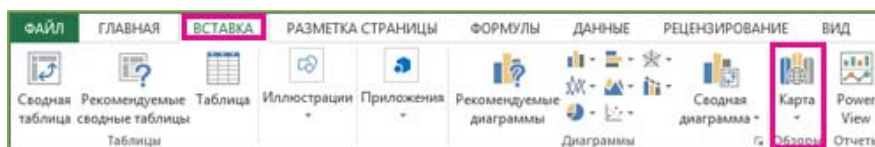


Рис. 1.3. Элемент «Карта»

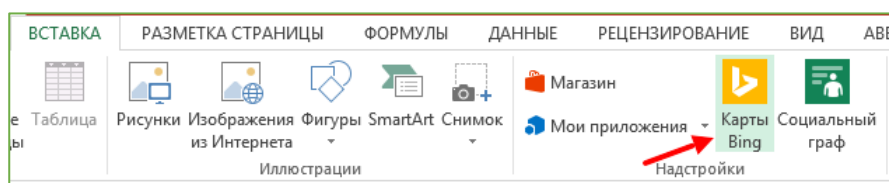


Рис. 1.4. Элемент «Карты Bing»

1.2. Визуализация данных

Визуализация данных обеспечивает пользователю наиболее эффективное представление результатов работы с данными. Визуализация данных широко применяется в научных и статистических исследованиях, таких, как прогнозирование, интеллектуальный анализ данных, бизнес-анализ, в различных новостных сводках и аналитических обзорах, а также для обучения и тестирования. Визуализация данных связана с визуализацией информации, инфографикой, визуализацией научных данных, разведочным анализом данных и статистической графикой.

1.3. Способ переноса информации в Microsoft Excel

В качестве исходных текстовых данных необходимо взять данные, включающие географические данные и числовую информацию. К примеру, это может быть таблица «Города России с населением свыше 100 тысяч жителей» из соответствующей статьи в Wikipedia (рис. 1.5)

(https://ru.wikipedia.org/wiki/Список_городов_России_с_населением_более_100_тысяч_жителей).

| Города России с населением свыше 100 тысяч жителей | | | | | | | | | | | | | | | | | | | | |
|---|--------------------------|-------------------------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| (численность населения по всероссийским переписям населения 1897—2002 годов по данным «Российского статистического ежегодника» — выпуск 2011 года ^[6] ; за 2010 год по окончательным данным Всероссийской переписи населения 2010 года, т.1 ^[6] ; за 2012 ^[7] , 2013 ^[8] , 2014 ^[9] , 2015 ^[10] , 2016 ^[11] , 2017 ^[12] , 2018 ^[13] , 2019 ^[14] , 2020 ^[15] годы — оценки Федеральной службы государственной статистики на 1 января) | | | | | | | | | | | | | | | | | | | | |
| № | Город | Население, тыс. человек | | | | | | | | | | | | | | | | | | |
| в 1897 ^[16] | | 1897 | 1926 | 1939 | 1959 | 1970 | 1979 | 1989 | 2002 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 1 | 2 Москва ^[17] | 1039 | 2080 | 4609 | 6133 | 7194 | 8057 | 8878 | 10126 | 11504 | 11541 | 11613 | 11980 | 12108 | 12198 | 12330 | 12381 | 12506 | 12615 | 12678 |
| 2 | 1 Санкт-Петербург | 1265 | 1737 | 3431 | 3390 | 4033 | 4569 | 4989 | 4661 | 4880 | 4899 | 4953 | 5028 | 5132 | 5192 | 5226 | 5282 | 5352 | 5384 | 5398 |
| 3 | Новосибирск | 8 | 120 | 404 | 885 | 1161 | 1309 | 1420 | 1426 | 1474 | 1475 | 1499 | 1524 | 1548 | 1567 | 1584 | 1603 | 1613 | 1618 | 1626 |
| 4 | 35 Екатеринбург | 43 | 140 | 423 | 779 | 1025 | 1210 | 1296 | 1294 | 1350 | 1353 | 1378 | 1396 | 1412 | 1428 | 1444 | 1456 | 1469 | 1483 | 1494 |
| 5 | 4 Казань | 130 | 179 | 406 | 667 | 869 | 989 | 1085 | 1105 | 1144 | 1145 | 1161 | 1176 | 1191 | 1206 | 1217 | 1232 | 1244 | 1252 | 1257 |
| 6 | 8 Нижний Новгород | 90 | 222 | 644 | 941 | 1170 | 1342 | 1400 | 1311 | 1251 | 1252 | 1255 | 1260 | 1264 | 1268 | 1267 | 1262 | 1259 | 1254 | 1252 |

Рис. 1.5. Таблица «Города России с населением более 100 тысяч жителей»

Далее, необходимо скопировать данную таблицу в Excel. Для этого необходимо выделить элементы (например, первые 100 элементов), нажать правой кнопкой мыши и выбрать пункт «Копировать» либо нажать комбинацию клавиш Ctrl+C (рис. 1.6).

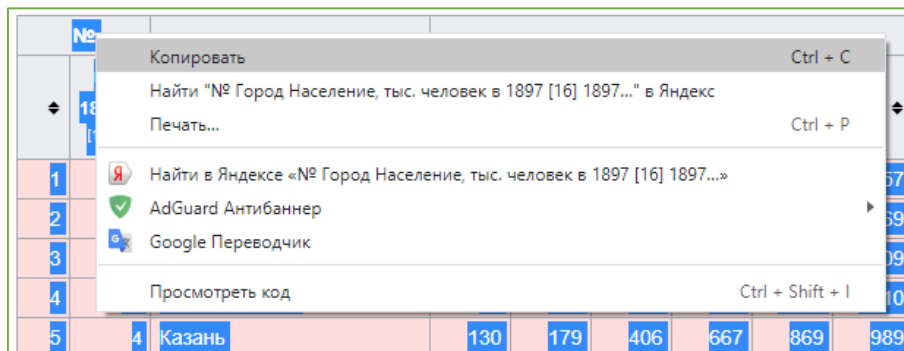


Рис. 1.6. Интерактивное меню копирования

После копирования в буфер обмена текстовой информации её необходимо перенести в Microsoft Excel. Для этого необходимо открыть Microsoft Excel, перейти в ячейку A1, нажать правой кнопкой мыши и выбрать параметр вставки «Использовать конечное форматирование» (рис. 1.7).

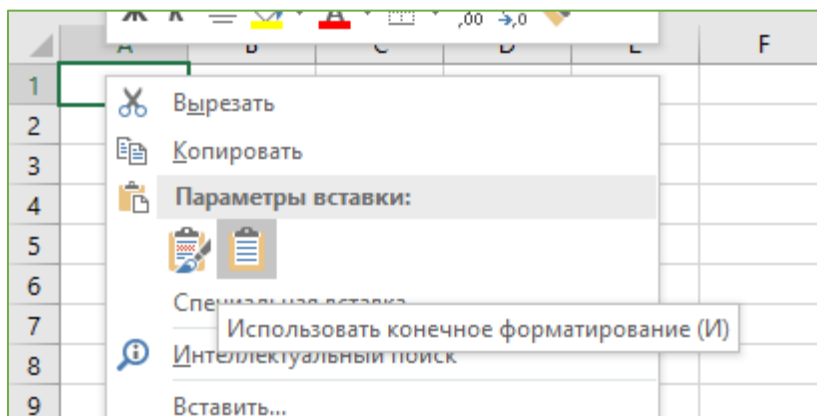


Рис. 1.7. Интерактивное меню вставки

Можно нажать комбинацию клавиш Ctrl+V и в параметрах вставки выбрать «Использовать конечное форматирование» (рис. 1.8).

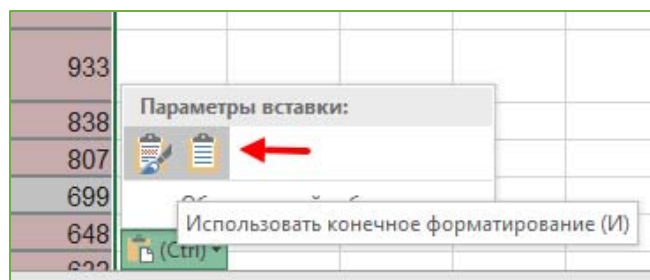


Рис. 1.8. Параметры вставки

1.4. Способ подготовки информации в Microsoft Excel

После переноса информации данные представлены в «сыром», неподготовленном виде. Имеются пропуски информации/данных: пустые строки и т. д. Необходимо их устранить. Для этого необходимо:

- 1) удалить все пустые строки (рис. 1.9). Для этого необходимо нажать правой кнопки мыши на необходимой строке и столбце и выбрать пункт «Удалить...» (рис. 1.10). Затем в меню удаления выбрать необходимый элемент (рис. 1.11).

| | | | | |
|----|-----|--------------|----|-----|
| 42 | 11 | Курск | 76 | 82 |
| 43 | 37 | Ставрополь | 42 | 59 |
| 44 | 20 | Севастополь | 54 | 75 |
| | | [Крым 1][19] | | |
| 45 | | Сочи | — | 13 |
| 46 | 177 | Улан-Удэ | 8 | 29 |
| 47 | 21 | Тверь | 54 | 108 |

Рис. 1.9. Пример пустой строки

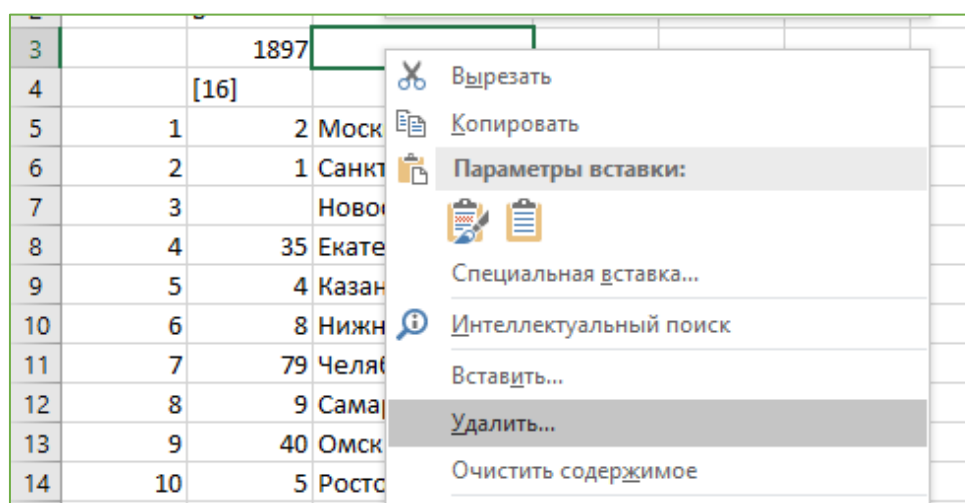


Рис. 1.10. Меню управления ячейкой

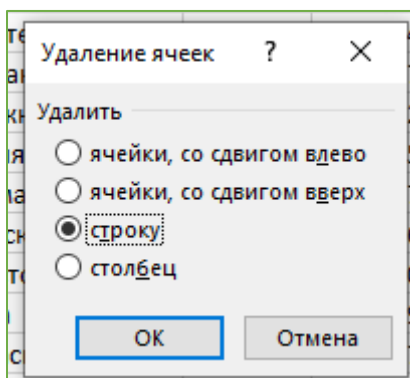


Рис. 1.11. Меню удаления ячеек

- 2) удалить сноски в названии городов (рис. 1.12).

| | |
|----|-----------------|
| 2 | Москва[17] |
| 1 | Санкт-Петербург |
| | Новосибирск |
| 35 | Екатеринбург |

Рис. 1.12. Пример сноски в названии города

- 3) привести заголовок к такому виду, чтобы он занимал только одну строку (рис. 1.13)

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|--------|-----------------|------|------|------|------|------|------|------|-------|-------|
| 1 | № | № 1897 | Город | 1897 | 1926 | 1939 | 1959 | 1970 | 1979 | 1989 | 2002 | 2010 |
| 2 | 1 | 2 | Москва | 1039 | 2080 | 4609 | 6133 | 7194 | 8057 | 8878 | 10126 | 11504 |
| 3 | 2 | 1 | Санкт-Петербург | 1265 | 1737 | 3431 | 3390 | 4033 | 4569 | 4989 | 4661 | 4880 |
| 4 | 3 | | Новосибирск | 8 | 120 | 404 | 885 | 1161 | 1309 | 1420 | 1426 | 1474 |

Рис. 1.13. Пример исправленного заголовка

- 4) исключить такие пропуски данных, как дефис (рис. 1.14). Для этого необходимо нажать сочетание клавиш Ctrl+N и заменить все значения «—» на ноль (рис. 1.15)

| | | | |
|---|-----|-----|-----|
| — | | 8 | 29 |
| | 5 | 9 | 31 |
| — | — | — | |
| | 115 | 155 | 285 |
| | 76 | 82 | 120 |
| | 42 | 59 | 85 |

Рис. 1.14. Пример нечисловых данных

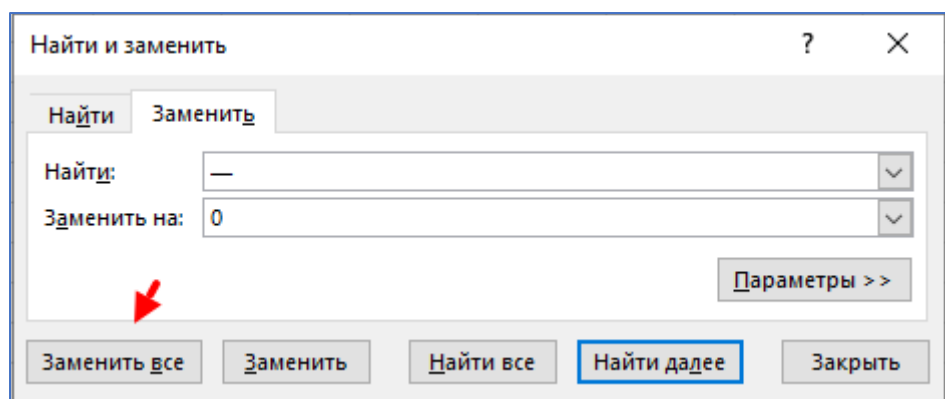


Рис. 1.15. Меню замены данных

После нажатия на кнопку должно появиться сообщение об успешной замене «Всё готово. Сделано замен: ...».

1.5. Построение 3D-карты

После подготовки данных можно построить 3D-карту. Для этого необходимо (рис. 1.16):

- 1) выделить необходимые для построения столбцы, содержащие географическую точку (город) и числовые данные (население в определённый год);
- 2) в верхней панели приложения выбрать пункт «Вставка»;
- 3) нажать на кнопку «3D-карта»;
- 4) из всплывающего меню выбрать пункт «Открыть 3D Maps».

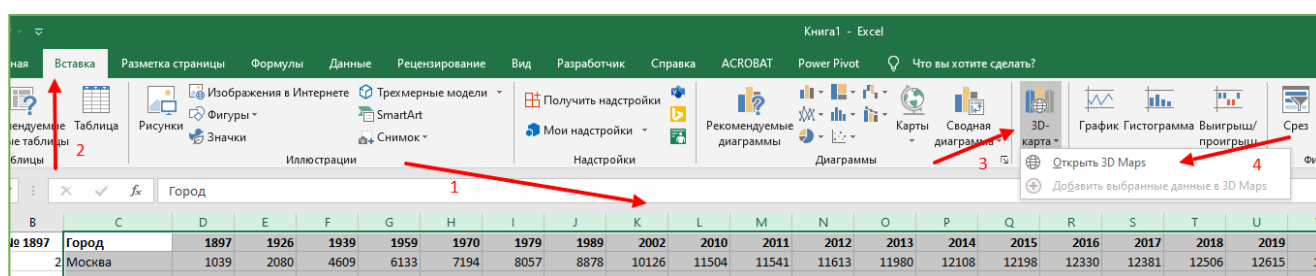


Рис. 1.16. Этапы создания 3D карт

После выполненных операций открывается компонент приложения 3D Maps. В нём можно создавать сцены, выбирать необходимые данные расположения, высоты, категории, настраивать тип графика и многое другое (рис. 1.17).

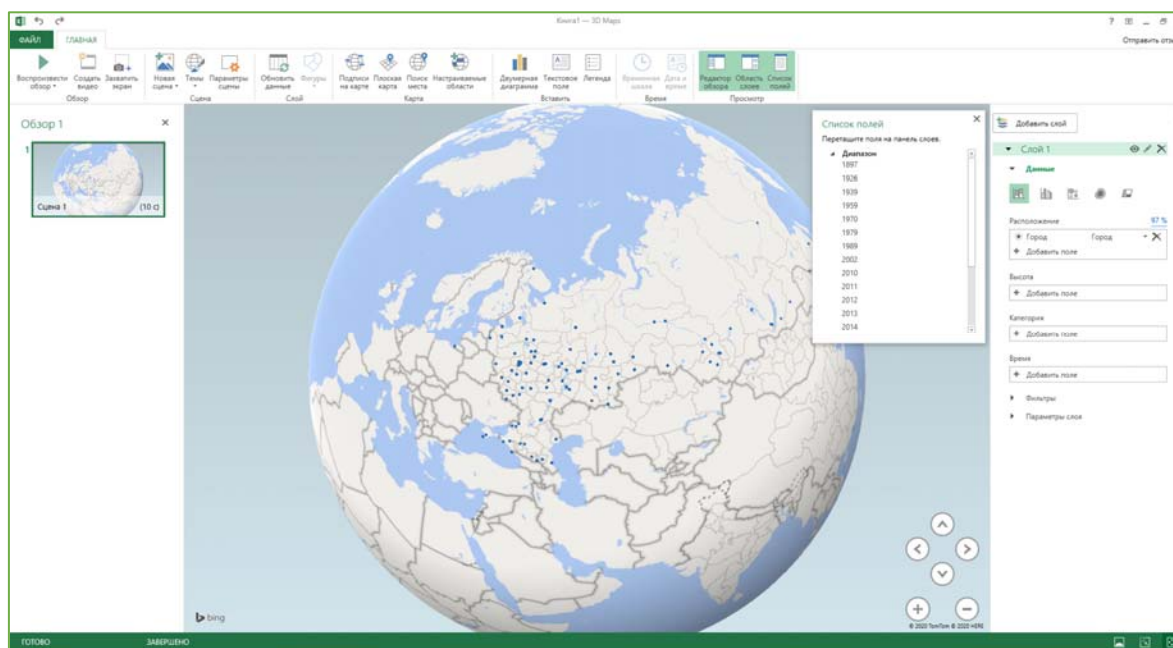


Рис. 1.17. Интерфейс модуля 3D Maps

Далее необходимо добавить данные для отображения на 3D карте. Для этого необходимо добавить соответствующие поля (рис. 1.18).

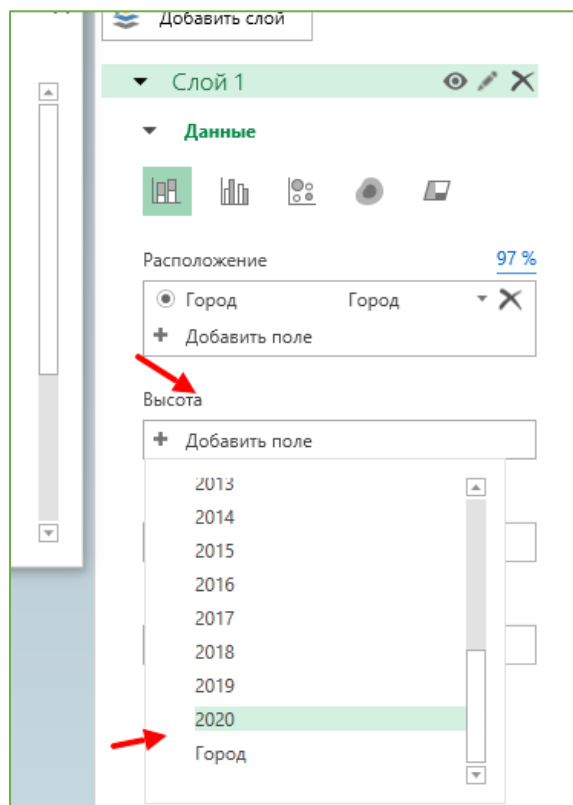


Рис. 1.18. Добавление полей на карту

После добавления полей их можно увидеть на полученной карте (рис. 1.19).

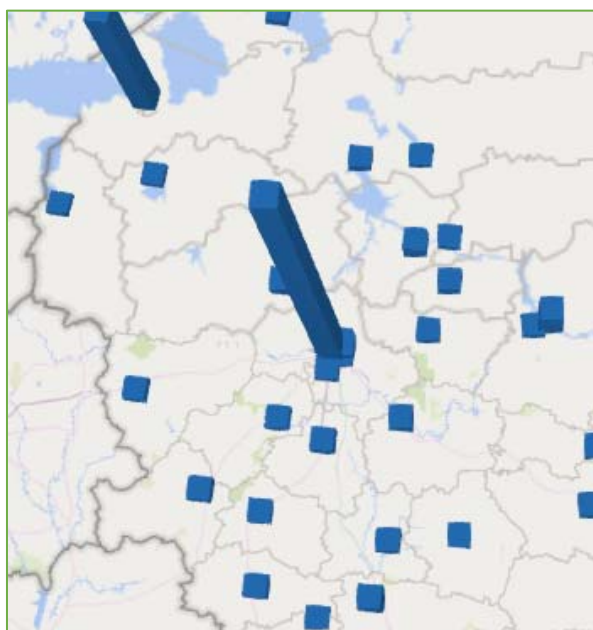


Рис. 1.19. Построенные поля

Если при добавлении данных недоступна агрегация по сумме, а доступны лишь «Количество непустых» или «Количество разных», значит имеются пробелы в данных и нужно их обнаружить и устранить (рис. 1.20).

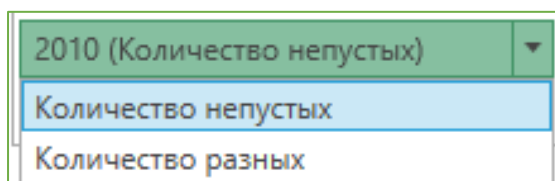


Рис. 1.20. Отсутствие нужных агрегаций

После добавления нужных слоёв данных (визуализации на карте), их можно настроить в параметрах слоя (рис. 1.21).

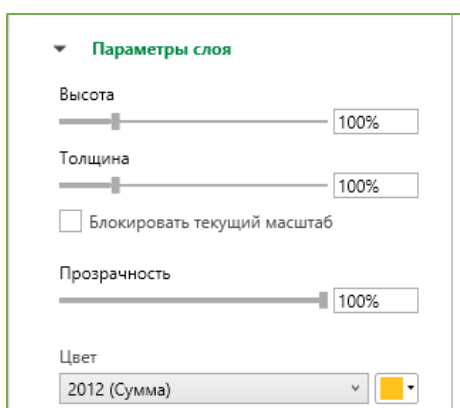


Рис. 1.21. Параметры слоя

В верхней области можно добавить дополнительные сцены, поменять темы и создать видео (рис. 1.22).

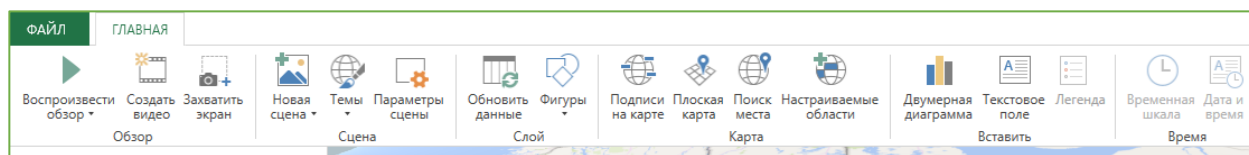


Рис. 1.22. Главная панель

1.6. Задание на лабораторную работу

Необходимо создать презентацию в виде видеоролика (с помощью кнопки «Создать видео» на главной панели), которая бы содержала несколько сцен, описывающая какие-либо числовые данные, привязанные к географическим объектам. Примерами данных могут служить бюджет регионов, население по странам мира, количество солнечных дней по городам и другое.

Лабораторная работа №2

ПОСТРОЕНИЕ ЛИСТА ПРОГНОЗА В EXCEL

Цель работы:

- ознакомиться со способами выгрузки информации из сети Интернет;
- изучить способы заполнения пропусков дат;
- изучить метод построение листа прогноза.

Минимальная необходимая версия – Microsoft Excel 2016.

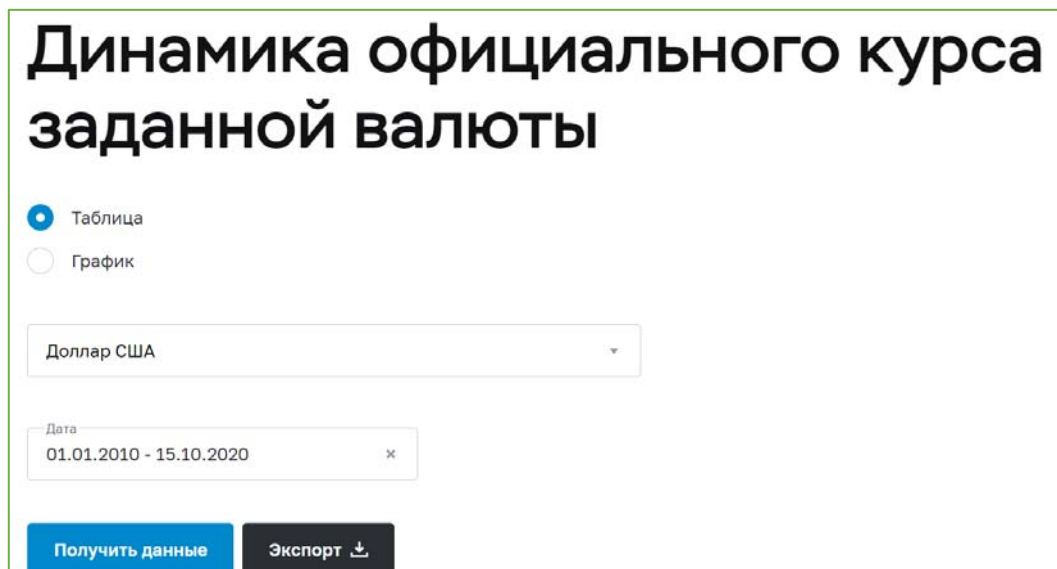
2.1. Лист прогноза

Инструмент прогнозирования, учитывающий сезонные закономерности в бизнес-процессах. Для расчёта прогноза необходимы данные за ранние периоды. Точность зависит от количества данных по периодам, при этом минимальное количество – 1 год. Необходимо использовать данные с одинаковыми интервалами времени между точками данных (например, месяц, неделя).

2.2. Получение данных для построения

Необходимо создать таблицу с данными за период и соответствующими ему показателями, например, от года.

Для примера перейдём на официальный сайт Центробанка в раздел "Динамика официального курса ЦБ" (https://cbr.ru/currency_base/dynamics/) (рис. 2.1).



Динамика официального курса заданной валюты

☒ Таблица
☐ График

Доллар США

Дата
01.01.2010 - 15.10.2020

Получить данные Экспорт

Рис. 2.1. Динамика официального курса заданной валюты

И скачать данные по курсу с начала с 2010 года. Для этого необходимо нажать кнопку «Получить данные», после «Экспорт» -> «Экспортировать в XLS» (рис. 2.2).

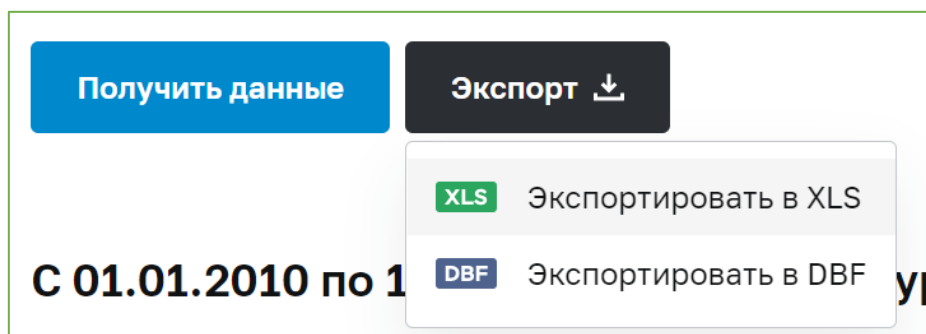


Рис. 2.2. Экспорт данных

Полученная таблица представлена на рисунке 2.3.

| | A | B | C | D | E |
|----|---------|------------|---------|------------|---|
| 1 | nominal | data | curs | cdx | |
| 2 | 1 | 01.01.2010 | 30.1851 | Доллар США | |
| 3 | 1 | 12.01.2010 | 29.4283 | Доллар США | |
| 4 | 1 | 13.01.2010 | 29.3774 | Доллар США | |
| 5 | 1 | 14.01.2010 | 29.6409 | Доллар США | |
| 6 | 1 | 15.01.2010 | 29.4299 | Доллар США | |
| 7 | 1 | 16.01.2010 | 29.5603 | Доллар США | |
| 8 | 1 | 19.01.2010 | 29.5963 | Доллар США | |
| 9 | 1 | 20.01.2010 | 29.5184 | Доллар США | |
| 10 | 1 | 21.01.2010 | 29.6941 | Доллар США | |
| 11 | 1 | 22.01.2010 | 29.7486 | Доллар США | |

Рис. 2.3. Полученная таблица

Далее необходимо разрешить редактирование данных (рис. 2.4).

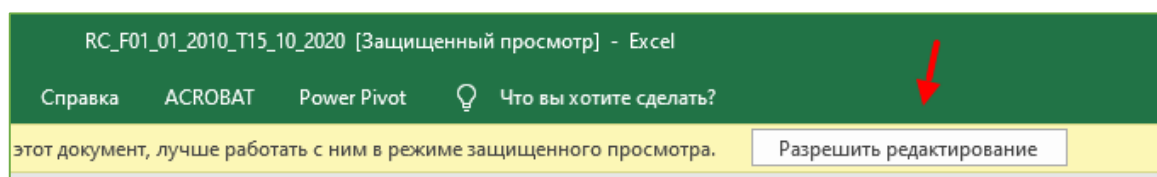


Рис. 2.4. Разрешение на редактирование данных

2.3. Преобразование данных для построения

Как можно увидеть, в данных имеются пропуски дат (праздники, выходные дни и т.д.) (рис. 2.5).

| | B | C | |
|------|------------|---------|------|
| data | | curs | cdx |
| 1 | 01.01.2010 | 30.1851 | Дол. |
| 1 | 12.01.2010 | 29.4283 | Дол. |
| 1 | 13.01.2010 | 29.3774 | Дол. |
| 1 | 14.01.2010 | 29.6409 | Дол. |
| 1 | 15.01.2010 | 29.4299 | Дол. |
| 1 | 16.01.2010 | 29.5603 | Дол. |
| 1 | 19.01.2010 | 29.5963 | Дол. |
| 1 | 20.01.2010 | 29.5184 | Дол. |

Рис. 2.5. Пропуски дат в данных

Чтобы их устранить необходимо создать новый лист в левой нижней части листа (рис. 2.6).

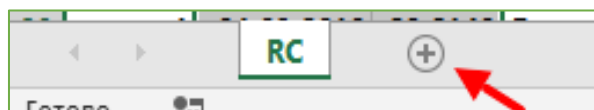


Рис. 2.6. Создание нового листа

В данном листе необходимо ввести первую дату диапазона данных (рис. 2.7).

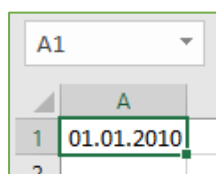


Рис. 2.7. Заполнение ячейки A1 нового листа

Далее необходимо заполнить даты с начала диапазона до конца. Для этого необходимо выбрать «Главная» -> «Заполнить» -> «Прогрессия» (рис. 2.8).

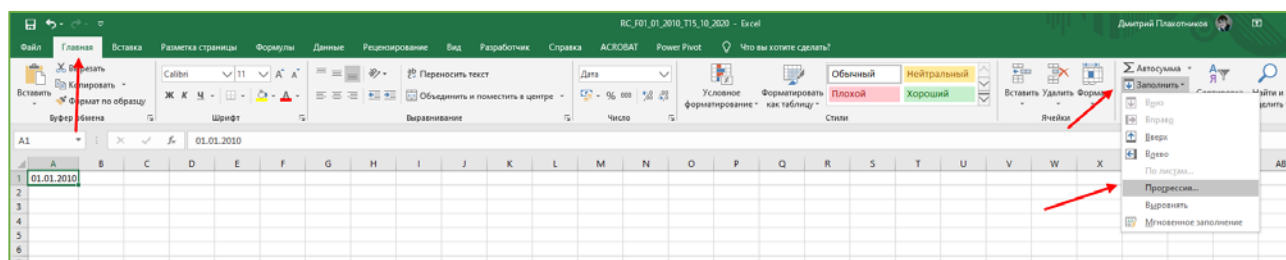


Рис. 2.8. Заполнение столбца датами

В появившемся окне необходимо выбрать расположение прогрессии «по столбцам», а также указать предельное значение в виде последней даты диапазона данных (рис. 2.9).

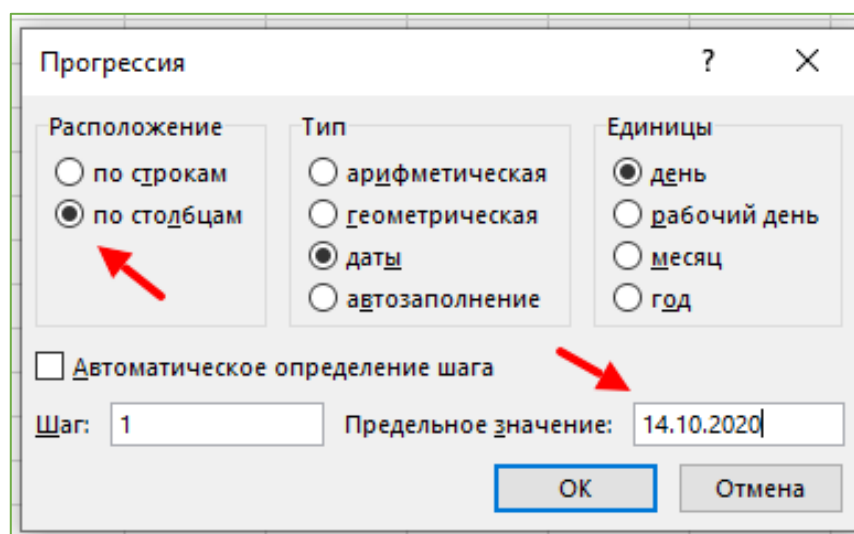


Рис. 2.9. Заполнение ячейки A1 нового листа

Далее необходимо перейти в крайнюю нижнюю ячейку A (для этого необходимо перейти в ячейку A1 нажать комбинацию Ctrl+стрелочка вниз). После этого перейти в соседнюю ячейку B (в данном случае это B3940) (рис. 2.10).

| | A | B |
|------|------------|---|
| 3936 | 10.10.2020 | |
| 3937 | 11.10.2020 | |
| 3938 | 12.10.2020 | |
| 3939 | 13.10.2020 | |
| 3940 | 14.10.2020 | |
| 3941 | | |

Рис. 2.10. Переход в крайнюю нижнюю ячейку

В выбранную ячейку необходимо вставить функцию (рис. 2.11).

| | A | B | C |
|------|------------|---|---|
| 3939 | 13.10.2020 | | |
| 3940 | 14.10.2020 | | |

Рис. 2.11. Вставка формулы

В появившемся окне необходимо выбрать функцию «ВПР» (рис. 2.12).

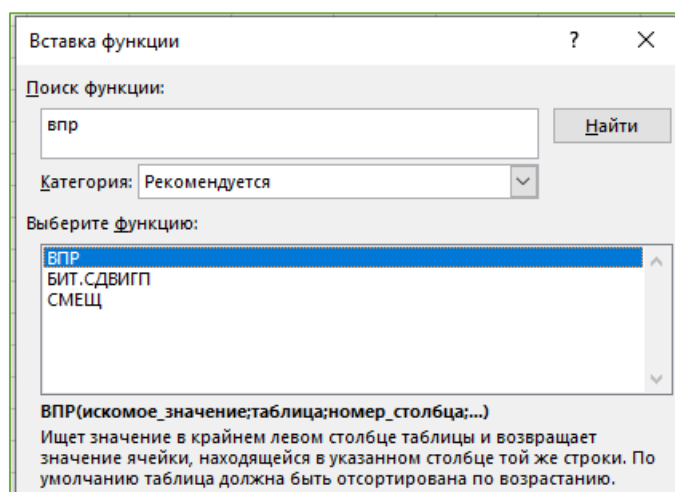


Рис. 2.12. Вставка функции

В аргументах функции необходимо написать значения (рис. 2.13), где:

A3940 – крайняя нижняя ячейка, **в зависимости от диапазона может принимать иное значение**;

RC – названия листа где хранятся необработанные данные;

\$B:\$C – содержащие дату и курс;

2 – номер столбца, из которого берутся значения (в данном случае, curs);

1 – использование интервального просмотра, то есть поиск значений, если нет точного совпадения, в соседних ячейках.

Получается формула вида «=ВПР(A3940;'RC'!\$B:\$C;2;1)».

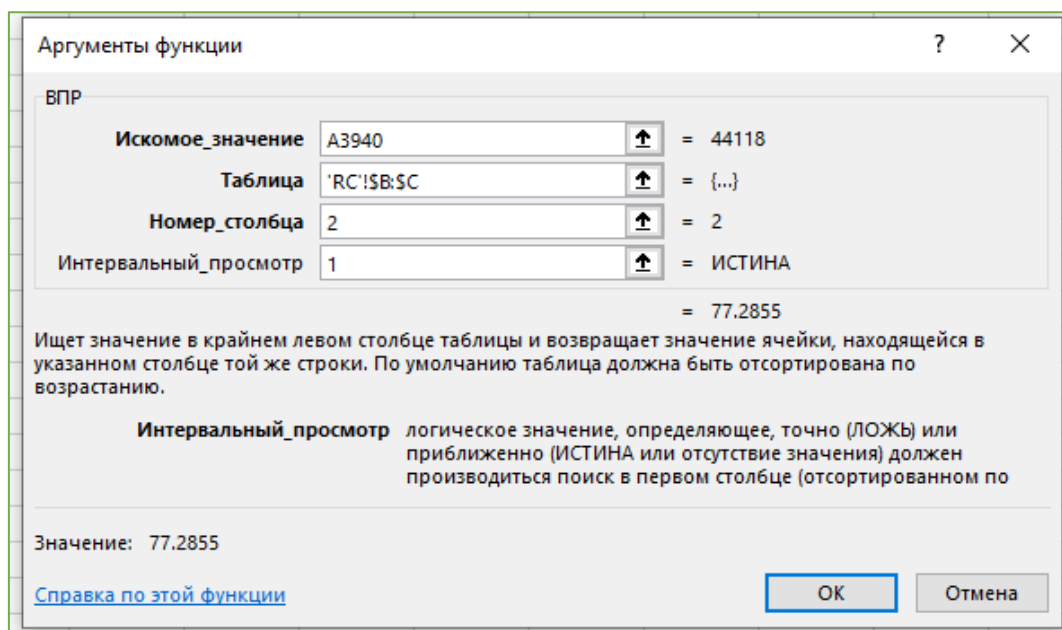


Рис. 2.13. Аргументы функции ВПР

Далее необходимо, находясь в данной ячейке, нажать комбинацию клавиш Ctrl+Shift+стрелочка вверх, чтобы выбрать весь диапазон для заполнения. Далее выбрать «Главная» -> «Заполнить» -> «Вверх» (рис. 2.14).

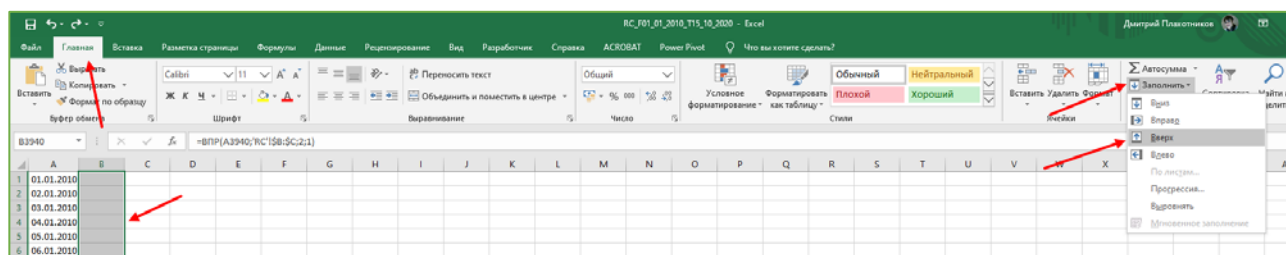


Рис. 2.14. Заполнение ячеек столбца В

Данные должны быть успешно заполнены (рис. 2.15).

| | | |
|----|------------|---------|
| 6 | 06.01.2010 | 30.1851 |
| 7 | 07.01.2010 | 30.1851 |
| 8 | 08.01.2010 | 30.1851 |
| 9 | 09.01.2010 | 30.1851 |
| 10 | 10.01.2010 | 30.1851 |
| 11 | 11.01.2010 | 30.1851 |
| 12 | 12.01.2010 | 29.4283 |
| 13 | 13.01.2010 | 29.3774 |
| 14 | 14.01.2010 | 29.6409 |
| 15 | 15.01.2010 | 29.4299 |
| 16 | 16.01.2010 | 29.5603 |
| 17 | 17.01.2010 | 29.5603 |
| 18 | 18.01.2010 | 29.5603 |
| 19 | 19.01.2010 | 29.5963 |
| 20 | 20.01.2010 | 29.5184 |

Рис. 2.15. Заполненные данные

2.4. Построение листа прогноза

Выделяем ячейки А и В новом листе проводя мышкой по верхней панели (рис. 2.16).

| | A | B |
|---|------------|---------|
| 1 | 01.01.2010 | 30.1851 |
| 2 | 02.01.2010 | 30.1851 |
| 3 | 03.01.2010 | 30.1851 |

Рис. 2.16. Выделение столбцов

На вкладке «Данные» в группе необходимо нажать кнопку «Лист прогноза» (рис. 2.17).

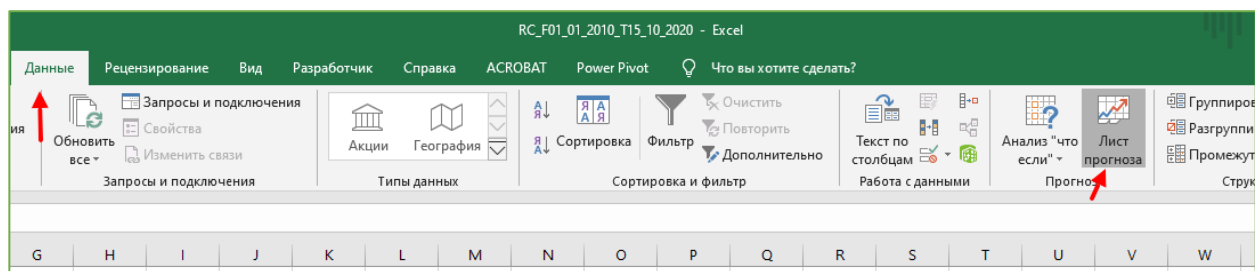


Рис. 2.17. Вызов листа прогноза

В окне «Создание листа прогноза» необходимо выбрать график или гистограмму для визуального представления прогноза и заполнить дату завершения прогноза. Так же при необходимости можно изменить дополнительные параметры (рис. 2.18)

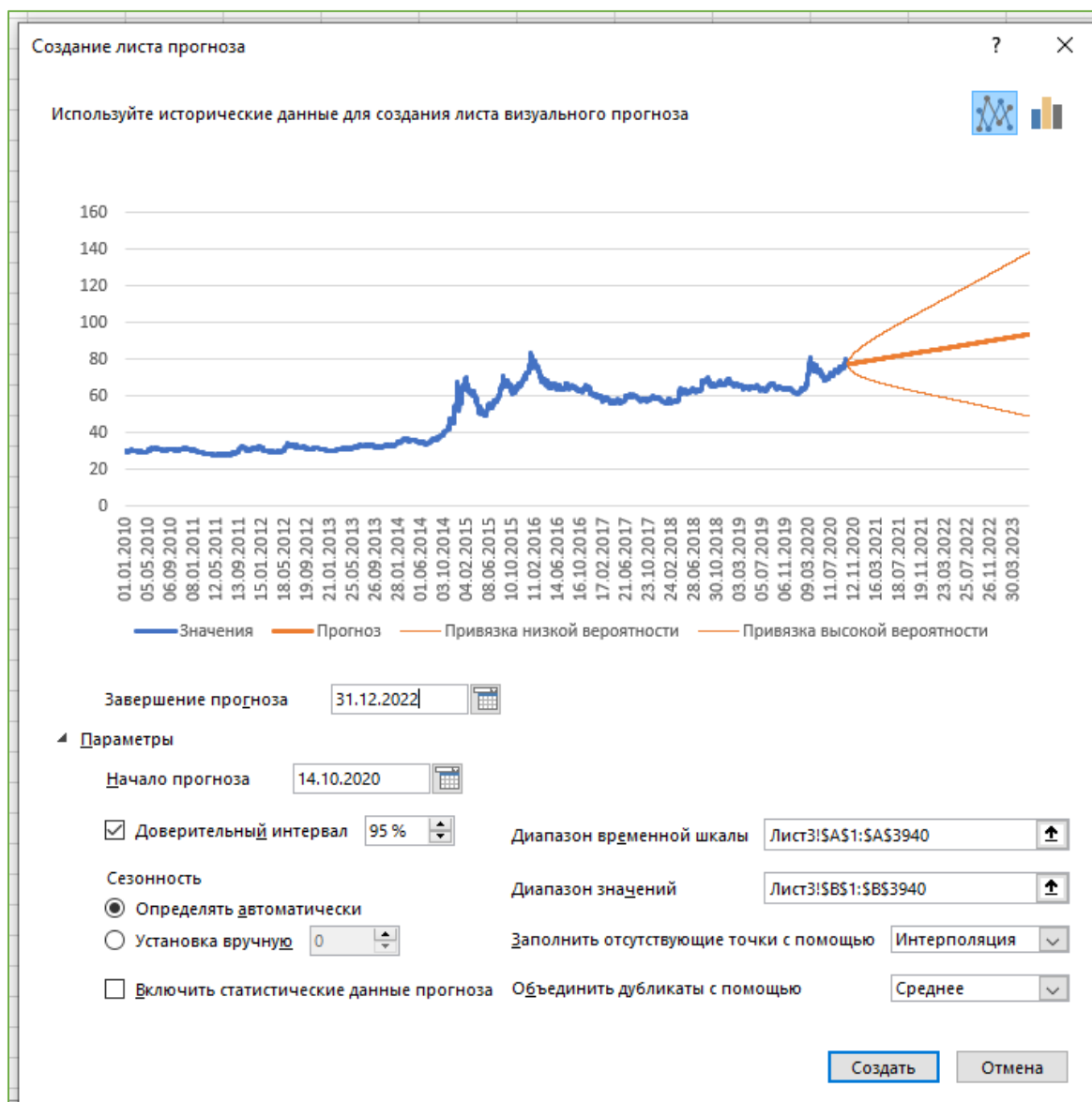


Рис. 2.18. Создание листа прогноза

После всех выполненных действий лист прогноза успешно построен – график (рис. 2.19) и прогнозная таблица (рис. 2.20).

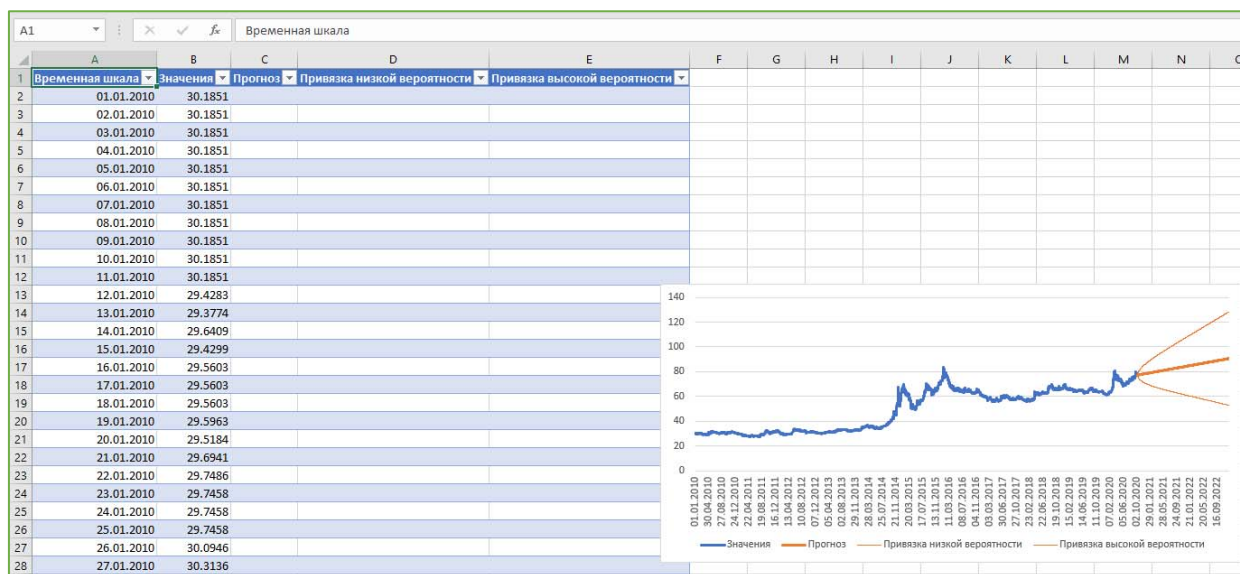


Рис. 2.19. Результаты построения

| Временная шкала | Значения | Прогноз | Привязка низкой вероятности | Привязка высокой вероятности |
|-----------------|----------|-----------|-----------------------------|------------------------------|
| 05.06.2022 | | 87.126466 | 57.43 | 116.83 |
| 06.06.2022 | | 87.142935 | 57.41 | 116.88 |
| 07.06.2022 | | 87.159405 | 57.38 | 116.93 |
| 08.06.2022 | | 87.175874 | 57.36 | 116.99 |
| 09.06.2022 | | 87.192344 | 57.34 | 117.04 |
| 10.06.2022 | | 87.208813 | 57.32 | 117.10 |
| 11.06.2022 | | 87.225283 | 57.30 | 117.15 |
| 12.06.2022 | | 87.241753 | 57.28 | 117.21 |
| 13.06.2022 | | 87.258222 | 57.26 | 117.26 |
| 14.06.2022 | | 87.274692 | 57.23 | 117.32 |
| 15.06.2022 | | 87.291161 | 57.21 | 117.37 |
| 16.06.2022 | | 87.307631 | 57.19 | 117.42 |
| 17.06.2022 | | 87.3241 | 57.17 | 117.48 |
| 18.06.2022 | | 87.34057 | 57.15 | 117.53 |
| 19.06.2022 | | 87.35704 | 57.13 | 117.59 |
| 20.06.2022 | | 87.373509 | 57.10 | 117.64 |
| 21.06.2022 | | 87.389979 | 57.08 | 117.70 |
| 22.06.2022 | | 87.406448 | 57.06 | 117.75 |
| 23.06.2022 | | 87.422918 | 57.04 | 117.81 |
| 24.06.2022 | | 87.439387 | 57.02 | 117.86 |
| 25.06.2022 | | 87.455857 | 57.00 | 117.91 |
| 26.06.2022 | | 87.472326 | 56.98 | 117.97 |
| 27.06.2022 | | 87.488796 | 56.95 | 118.02 |

Рис. 2.20. Созданная таблица данных прогноза

2.5. Выполнение практической работы

Необходимо повторить все действия, представленные в практической работе для курса валют в зависимости от варианта: 1 – евро, 2 – австралийский доллар, 3 – японская иена, 4 – норвежская крона, 5 – фунт стерлингов.

Лабораторная работа №3

ПОИСК НАИЛУЧШЕГО РЕШЕНИЯ С ПОМОЩЬЮ EXCEL

Цель работы:

- ознакомиться со способами включения дополнительных компонентов в Excel;
- изучить способы подготовки информации и использования формул;
- изучить метод поиска оптимального решения.

Минимальная необходимая версия – Microsoft Excel 2016 с включенным блоком инструментов «Поиск решений»

3.1. Включение блока инструментов для всех версий

Рассматриваемый пакет относится к категории надстроек. По умолчанию он отключён для всех версий. Соответственно, для включения пакета необходимо перейти в меню надстроек. Эта процедура выполняется следующим образом:

1. необходимо открыть вкладку «Файл», расположенную в верхней части ленты интерфейса;
2. с левой стороны открывающегося меню найти раздел «Параметры» и кликнуть по нему;
3. далее необходимо щелкнуть по подразделу «Надстройки». Здесь отображаются все надстройки программы, а внизу будет надпись «Управление». Справа от неё представлено выпадающее меню, в котором должны быть выбраны «Надстройки Excel», обычно уже установленные по умолчанию. Здесь необходимо нажать кнопку «Перейти» (рис. 3.1).
4. данное действие вызывает на экран диалоговое окно, в котором требуется выбрать раздел надстроек, если выставлено значение, отличное от «Поиск решения», поменять его на обозначенное и нажать «ОК» (рис. 3.2);
5. Требуемая функция появится на ленте в правой части вкладки «Данные» (рис. 3.3)

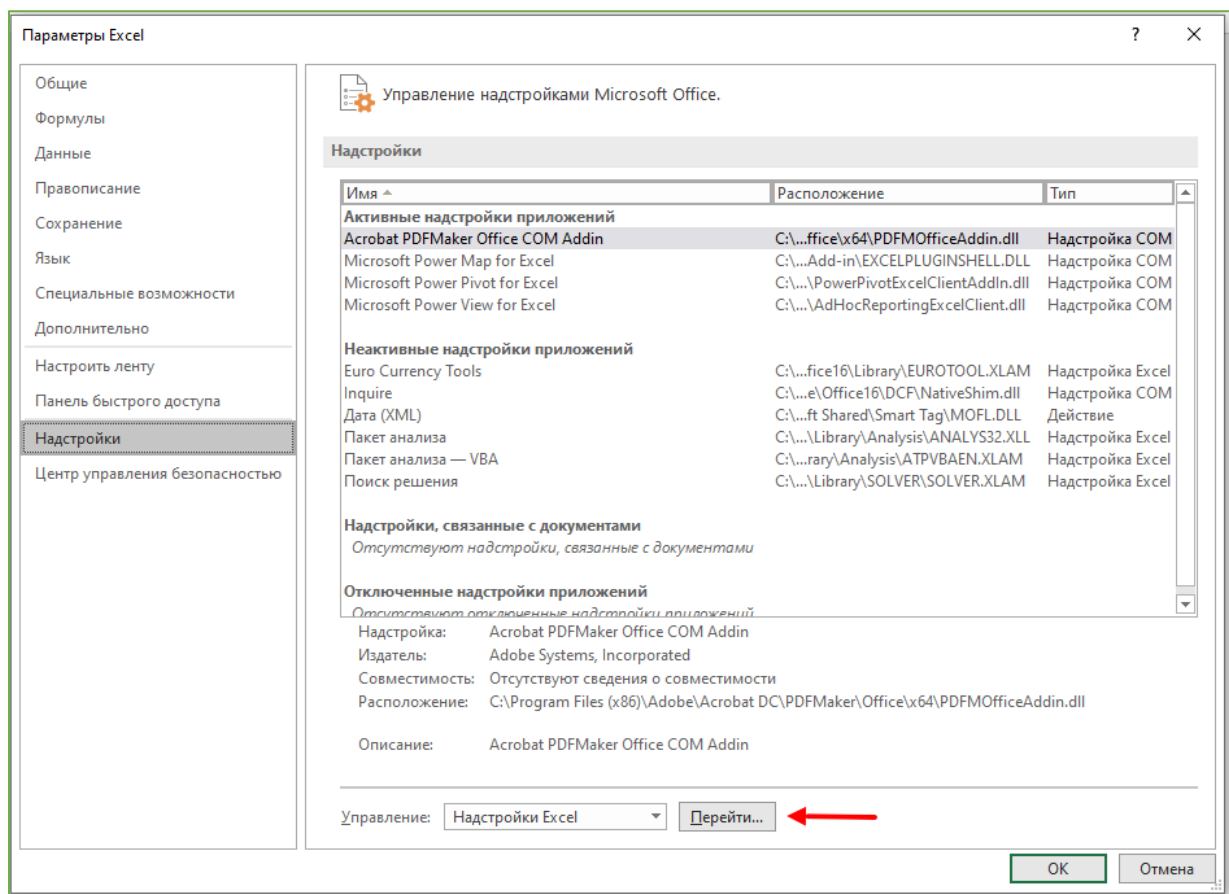


Рис. 3.1. Параметры Excel

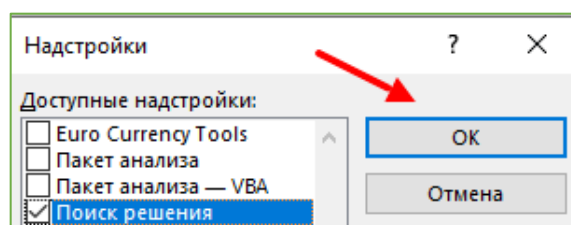


Рис. 3.2. Выбор надстройки

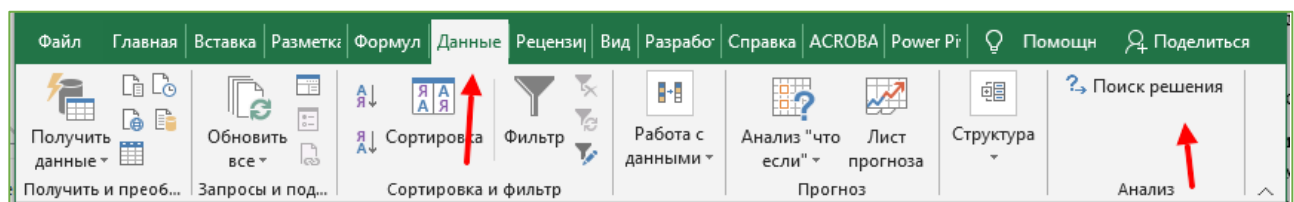


Рис. 3.3. Проверка функции

3.2. Подготовка исходных данных

В качестве примера решим следующую задачу – необходимо выполнить дополнительную выплату студентам в соответствии с их текущей стипендией. Фонд выплаты 5000 рублей. При заполнении ячеек «Итого» необходимо использовать формулу суммы (рис. 3.4).

| | | | | | |
|----|--------------|-----------------|--------------|---|-------------|
| B6 | | | | | |
| | A | B | C | D | E |
| 1 | Студент | Сумма стипендии | Доп. выплата | | Коэффициент |
| 2 | Иванов А.А. | 1400 | 0 | | |
| 3 | Сидоров И.И. | 2100 | 0 | | |
| 4 | Петров С.С. | 1800 | 0 | | |
| 5 | Попов Е.Е. | 3500 | 0 | | |
| 6 | Итого | 8800 | 0 | | |
| 7 | | | | | |

Рис. 3.4. Пример заполнения исходной таблицы

Целевая и искомая ячейка должны быть связаны друг с другом с помощью формулы. В данном случае, формула располагается в ячейке суммы доп. выплаты, и имеет следующий вид: « $=\$E\$2*B6$ », где $\$E\2 – абсолютный адрес искомой ячейки (коэффициент), а « $B6$ » – общая сумма стипендии, от которой производится расчет дополнительной выплаты студентам. Так же в ячейках доп. выплаты должна быть аналогичная формула (рис. 3.5).

| | | | | | |
|----|--------------|-----------------|--------------|---|-------------|
| B6 | | | | | |
| | A | B | C | D | E |
| 1 | Студент | Сумма стипендии | Доп. выплата | | Коэффициент |
| 2 | Иванов А.А. | 1400 | 0 | | |
| 3 | Сидоров И.И. | 2100 | 0 | | |
| 4 | Петров С.С. | 1800 | 0 | | |
| 5 | Попов Е.Е. | 3500 | 0 | | |
| 6 | Итого | 8800 | $=\$E\$2*B6$ | | |
| 7 | | | | | |
| 8 | | | | | |

Рис. 3.5. Формулы ячейки дополнительной выплаты

3.3. Запуск инструмента «Поиск решения»

После подготовки таблицы, находясь во вкладке «Данные», необходимо нажать на «Поиск решения», расположенным на ленте в блоке инструментов «Анализ». (рис. 3.6).

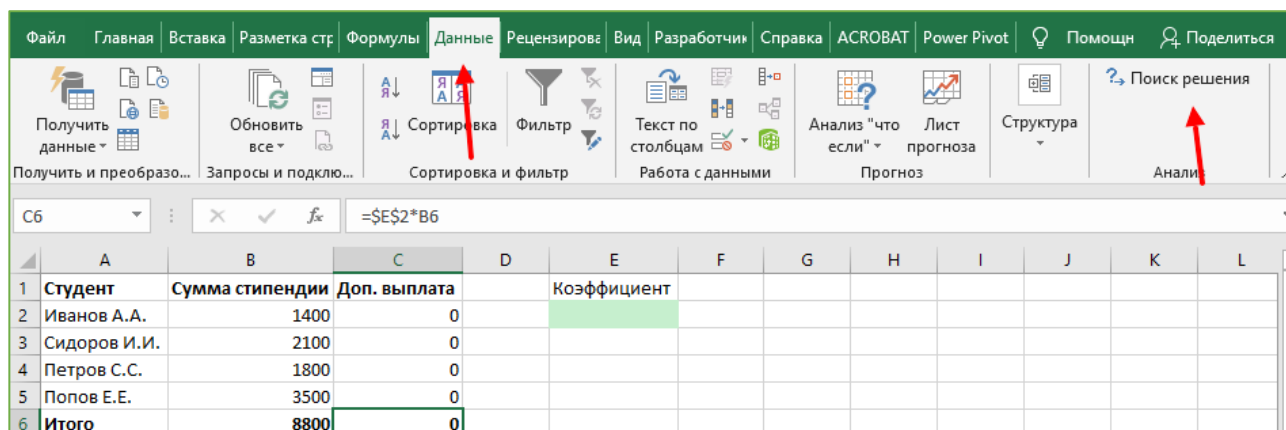


Рис. 3.6. Расположение «Поиск решения»

Открывается окно параметров, в котором нужно ввести необходимые данные. В поле «Оптимизировать целевую функцию» нужно ввести адрес целевой ячейки, где будет располагаться общая сумма дополнительной выплаты. Это можно сделать либо пропечатав координаты вручную, либо кликнув на кнопку, расположенную слева от поля ввода данных. Ниже расположено поле «Изменяя ячейки переменных», в котором необходимо задать адрес коэффициента, умножением на который стипендии будет рассчитана величина дополнительной выплаты. Адрес можно прописать аналогичным образом (рис. 3.7).

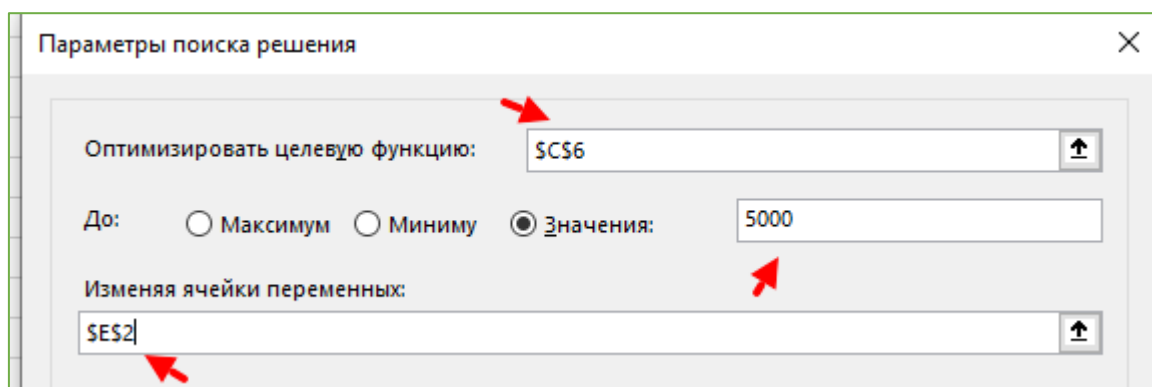


Рис. 3.7. Указание целевой и искомой ячейки

После этого, открывается окно добавления ограничения. В поле «Ссылка на ячейки» необходимо прописать адрес ячеек, относительно которых вводится ограничение. В данном случае, это искомая ячейка с коэффициентом. Далее необходимо задать нужный знак: «меньше или равно», «больше или равно», «равно», «целое число» и т.д. В данном случае, необходимо выбрать знак «больше или равно», чтобы сделать коэффициент положительным числом. Соответственно, в поле «Ограничение» необходимо указать число 0.

После этого необходимо нажать на кнопку «ОК», чтобы сохранить введенные ограничения (рис.3.8).

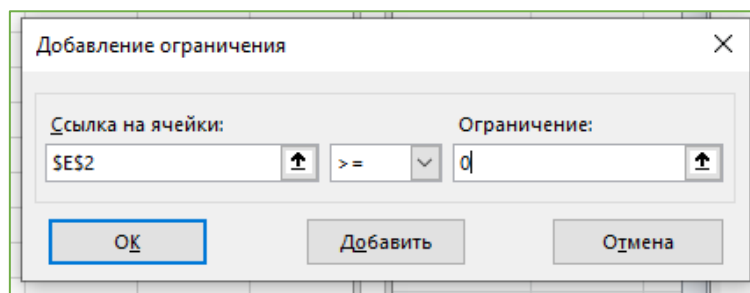


Рис. 3.8. Задание параметров поиска решения

После этого, ограничение появляется в соответствующем поле окна параметров поиска решения. Также, сделать переменные неотрицательными, можно установив галочку около соответствующего параметра чуть ниже (рис. 3.9). Необходимо, чтобы установленный параметр не противоречил тем правилам, которые указаны в ограничениях, во избежание конфликта параметров.

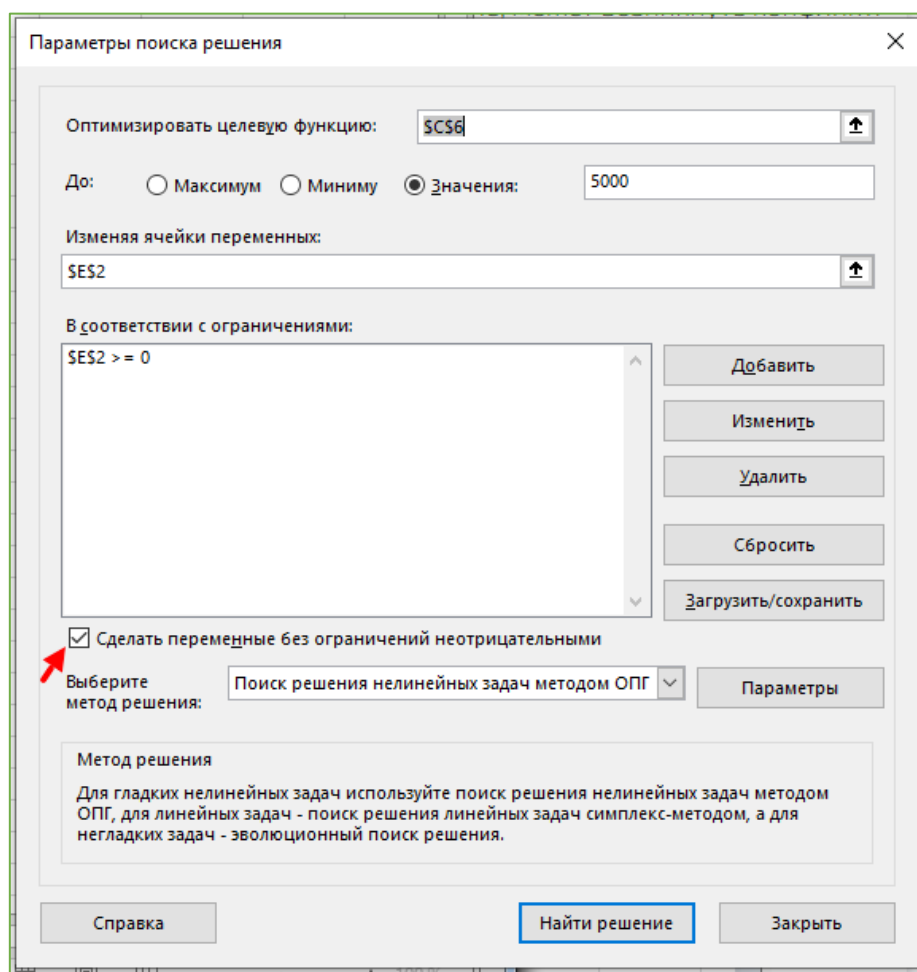


Рис. 3.9. Задание параметров поиска решения

Дополнительные настройки можно задать, кликнув «Параметры». Здесь можно установить точность ограничения и пределы решения. Когда нужные данные введены, необходимо нажать «ОК» (рис. 3.10).

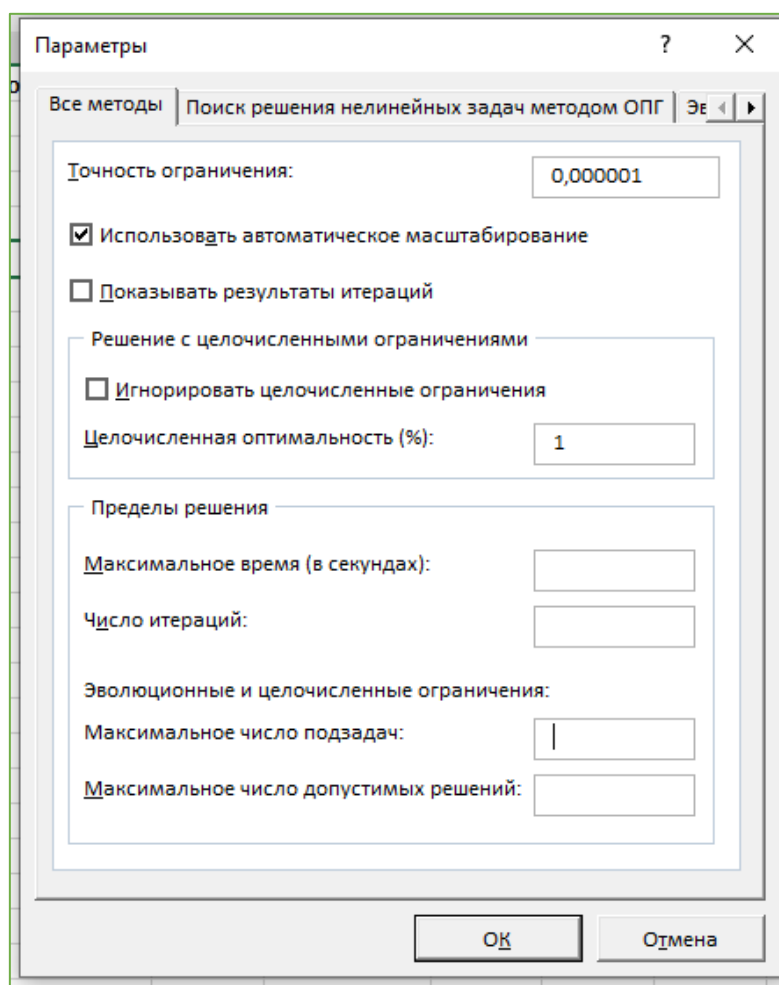


Рис. 3.10. Дополнительные параметры

После настройки всех параметров необходимо нажать кнопку «Найти решение» в правом нижнем углу окна (рис. 3.11).

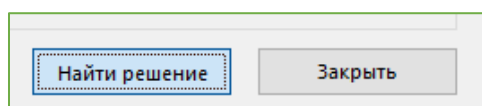


Рис. 3.11. Кнопка «Найти решение»

Далее, Excel выполнит расчеты. После выполнения расчётов открывается окно, в котором можно сохранить найденное решение либо восстановить исходные значения, переставив переключатель в соответствующую позицию. Так же можно вернуться в диалоговое окно параметров, поставив соответствующую галочку. После того, как выставлены галочки и переключатели, необходимо нажать на кнопку «ОК» (рис. 3.12).

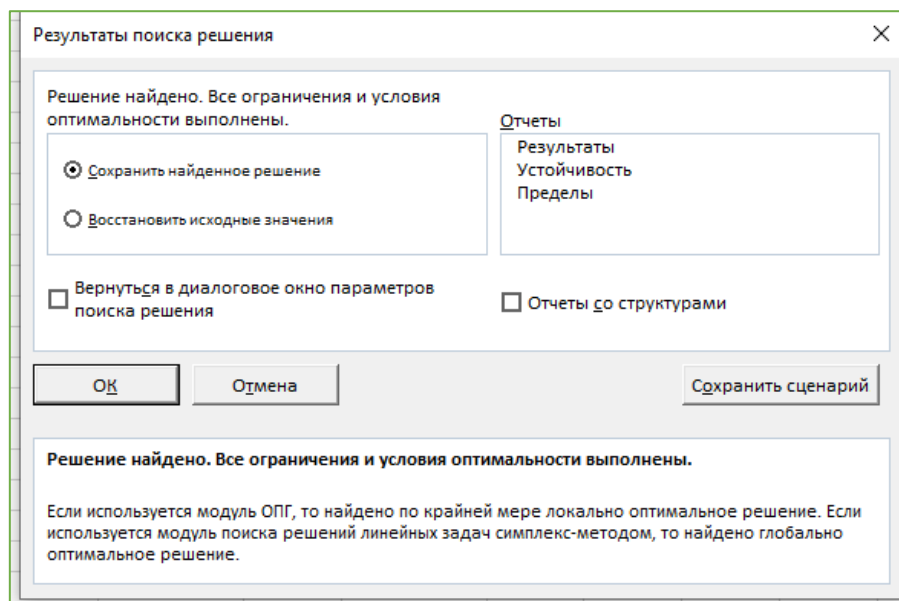


Рис. 3.12. Окно «Результаты поиска решения»

Далее, необходимо проверить наличие формулы в ячейках дополнительной выплаты. Формула должна располагаться в ячейке каждой суммы доп. выплаты, и иметь следующий вид: « $=\$E\$2*B2$ », где $\$E\2 – абсолютный адрес искомой ячейки (коэффициент), а « $B2$ » – общая сумма стипендии, от которой производится расчет дополнительной выплаты студентам. После написания данной формулы в ячейку C2 её необходимо «растянуть» её на весь ряд, потянув за нижний угол (рис. 3.13).

| | A | B | C | D | E | F |
|---|--------------|-----------------|--------------|---|-------------|---|
| 1 | Студент | Сумма стипендии | Доп. выплата | | Коэффициент | |
| 2 | Иванов А.А. | 1400 | 795.4545455 | | 0.568181818 | |
| 3 | Сидоров И.И. | 2100 | | | | |
| 4 | Петров С.С. | 1800 | | | | |
| 5 | Попов Е.Е. | 3500 | | | | |
| 6 | Итого | 8800 | 5000 | | | |

Рис. 3.13. Заполнение формулы

3.4. Проверка и корректировка результатов решения

Если результаты поиска решений не являются удовлетворительными, или при их подсчете программа выдаёт ошибку, то, в таком случае, необходимо вернуться в, описанное выше, в диалоговое окно параметров. Повторно проверить введенные данные и в случае, если ошибки нет, то перейти к параметру «Выберите метод решения» и попробовать решить поставленную задачу, выбрав другой метод (рис. 3.14).

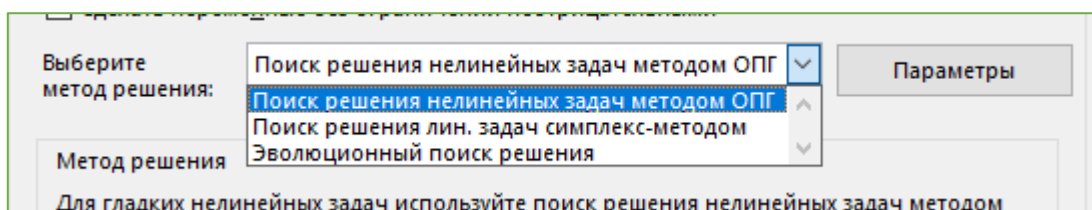


Рис. 3.14. Методы решения задачи

После успешного поиска результатов необходимо задать формат чисел. Для этого выделяем необходимый диапазон и выбираем формат «Денежный» и количество десятичных знаков 2 (рис. 3.15).

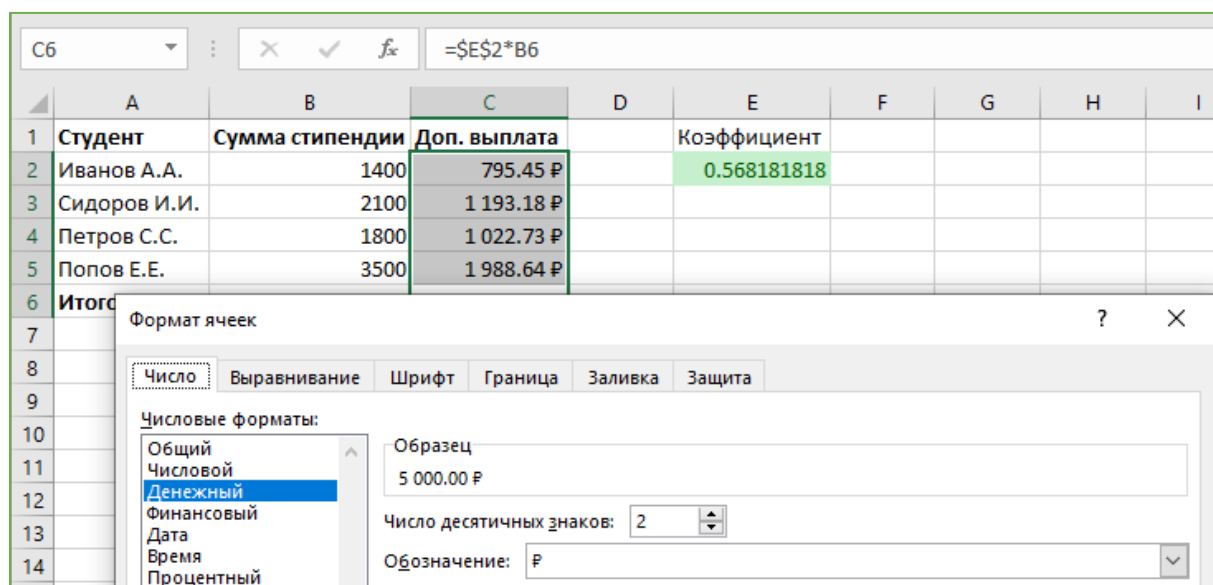


Рис. 3.15. Формат ячеек

3.5. Задание на лабораторную работу

Выполнить задачу поиска оптимального решения для придуманной задачи – это может быть распределение ресурсов, денежных средств, часов или что-либо другое. Обязательно должны быть минимум 3 ограничения и минимум 5 значений для расчёта.

Лабораторная работа №4

ЗНАКОМСТВО С АНАЛИТИЧЕСКОЙ ПЛАТФОРМОЙ QLIK SENSE

Цель работы:

- ознакомиться с особенностями платформы Qlik Sense;
- получить доступ к онлайн версии платформы Qlik Sense;
- изучить возможности платформы.

4.1. Платформа Qlik Sense

Qlik Sense – это представитель BI-платформ для самостоятельного проведения бизнес-анализа, предназначенная для всех корпоративных бизнес-пользователей. С помощью данного продукта можно анализировать данные и использовать полученные результаты для поддержки решений.

4.2. Получение доступа к облачной версии

Для получения доступа к бесплатному 30 дневному триальному доступу необходимо перейти по ссылке <https://www.qlik.com/ru-ru/try-or-buy/join-qlik-sense-cloud> и нажать на кнопку «Попробуй бесплатно» (рис. 4.2).

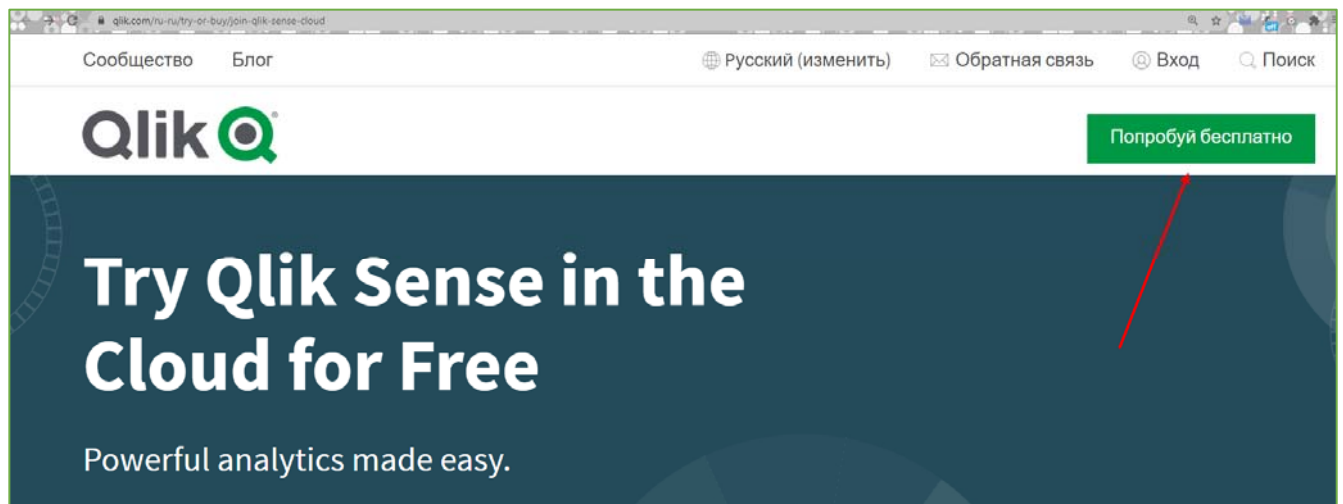


Рис. 4.2. Получение доступа к Qlik Sense в облаке

После нажатия на кнопку происходит перенаправление на форму регистрации в которой необходимо заполнить все поля и нажать «Register Now» (рис. 4.3)

Get Started for Free

Already have a Qlik account? [Login](#)

First name

Last name

Business email

Company

Job title

Telephone

Country

State

☐ Don't miss out! – your registration includes updates on free help, business trends, and updates from Qlik. You may opt-out of receiving further communications at any time. For further information, please see our [Privacy Policy](#).

Register Now

Рис. 4.3. Регистрационная форма

После нажатия кнопки, если всё было заполнено правильно, должно появиться сообщение об успешной регистрации (рис. 4.4)

Thank You for your interest in Qlik Sense!

Please check your e-mail on how to access your trial.

Рис. 4.4. Сообщение об успешной регистрации

Далее необходимо перейти на почту, открыть письмо «Qlik Sense On Cloud Services Setup» и в нём нажать кнопку «Let's get started» (рис. 4.5).

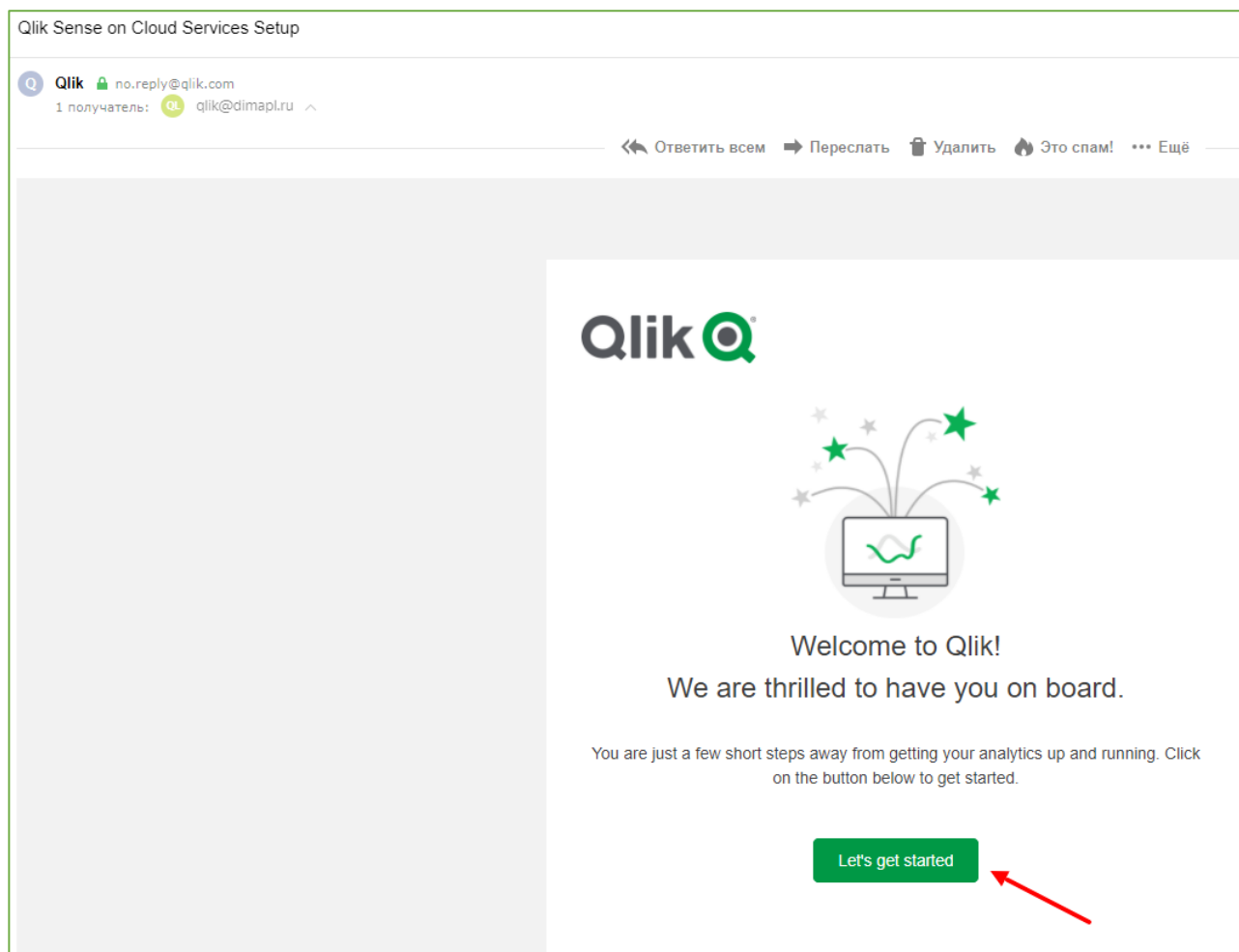


Рис. 4.5. Письмо «Qlik Sense On Cloud Services Setup»

После перехода по ссылке откроется форма настройки учётной записи. В ней необходимо создать пароль, согласиться с правилами и нажать кнопку «Далее» (рис. 4.6).

The image shows a web form titled "Настройка учетной записи" (Account Setup). It has a dark blue sidebar with a rocket and charts. The main area has a subtitle "Установите пароль и примите условия и положения." (Set password and accept terms and conditions). There are two input fields: "Адрес электронной почты" (Email address) with "qlik@dimapl.ru" and "Создать пароль" (Create password) with masked text. Below is a checked checkbox "Я прочитал(а) и принимаю Условия и положения." (I have read and accept the Terms and Conditions). A red arrow points to the checkbox. At the bottom right is a green button "Далее" (Next), with a red arrow pointing to it.

Рис. 4.6. Настройка учётной записи

После этого необходимо выбрать регион (желательно оставить выбранный автоматически) и нажать кнопку «Начало работы» (рис. 4.7).



Рис. 4.7. Выбор региона сервера

Доступ к облачной версии Qlik Sense получен (рис. 4.8).

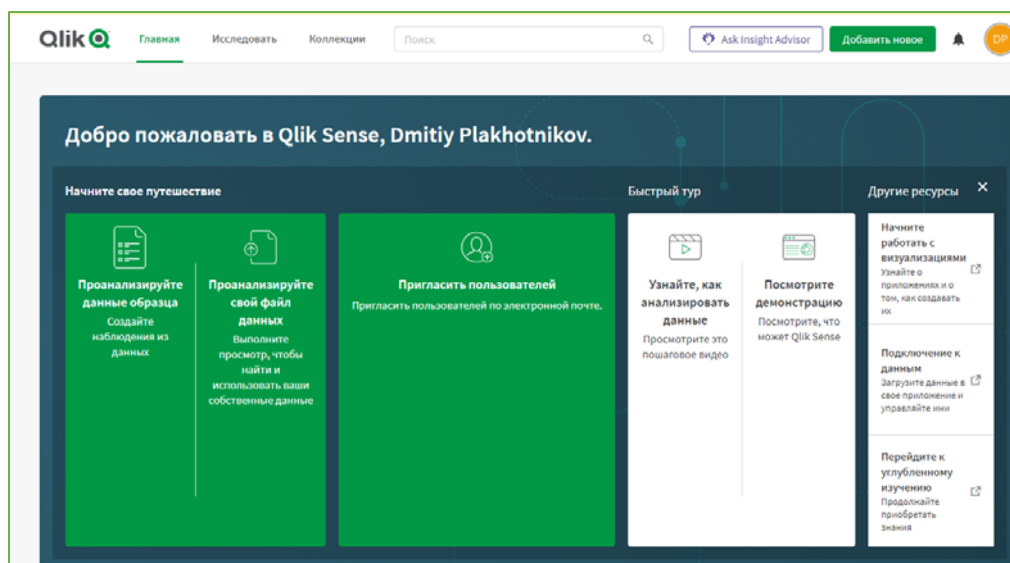


Рис. 4.8. Qlik Sense Cloud

4.3. Задание на лабораторную работу

Необходимо получить доступ к платформе, и ознакомиться с ней, с помощью ссылок на главной странице в подпунктах «Быстрый тур» (рис. 4.8).

Так же учебное пособие по Qlik Sense на русском языке доступно по ссылке: https://help.qlik.com/ru-RU/cloud-services/Subsystems/Hub/Content/Sense_Hub/Tutorials/introduction-beginning-with-basics.htm

Лабораторная работа №5

СОЗДАНИЕ ПРИЛОЖЕНИЙ И ЗАГРУЗКА ДАННЫХ В АНАЛИТИЧЕСКОЙ ПЛАТФОРМЕ QLIK SENSE

Цель работы:

- научиться создавать приложения в аналитической платформе;
- ознакомиться со способами загрузки данных в Qlik Sense;
- изучить способы изменения полей в скрипте загрузке данных.

5.1. Создание аналитического приложения

Для начала, необходимо перейти в платформу по ссылке <https://www.qlikcloud.com/> и авторизоваться с помощью аккаунта, зарегистрированного в прошлой лабораторной работе (рис 5.1).

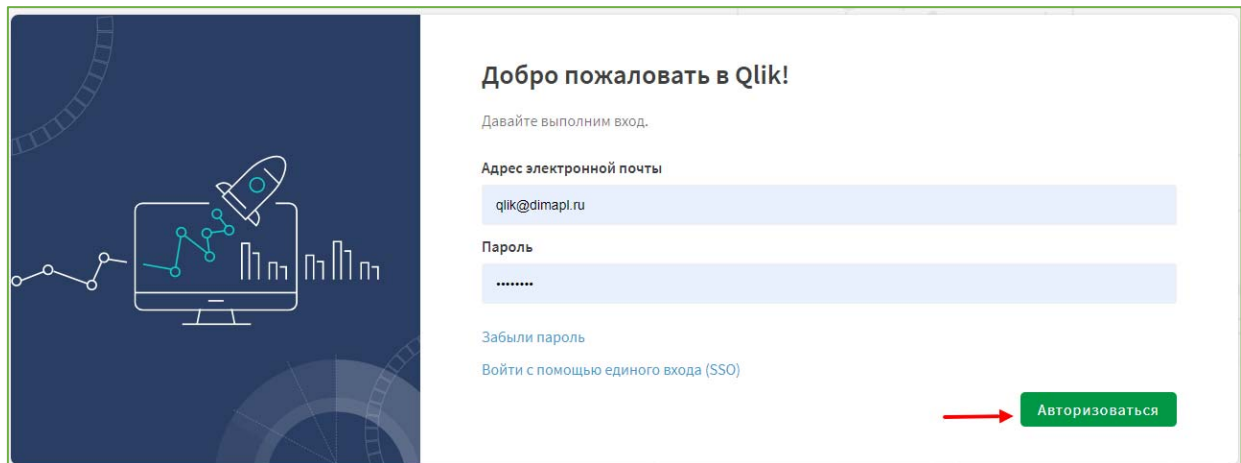


Рис. 5.1. Окно авторизации

Далее в правом верхнем углу необходимо нажать кнопку «Добавить новое» (рис. 5.2).

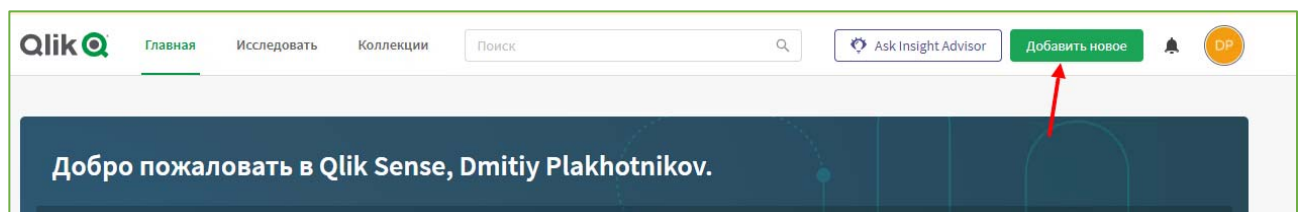


Рис. 5.2. Кнопка добавления аналитического приложения

И в появившемся окне необходимо выбрать «Новое аналитическое приложение» (рис. 5.3).

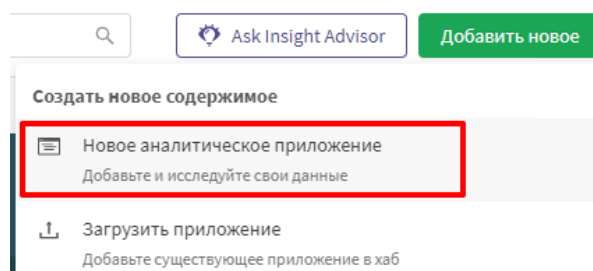


Рис. 5.3. Список добавления приложения

Далее необходимо заполнить форму создания нового приложения и нажать кнопку «Создать» (рис. 5.4).

Рис. 5.4. Форма создания приложения

После создания приложения открывается пространство приложений. В нём необходимо нажать на созданное приложение (рис. 5.5).

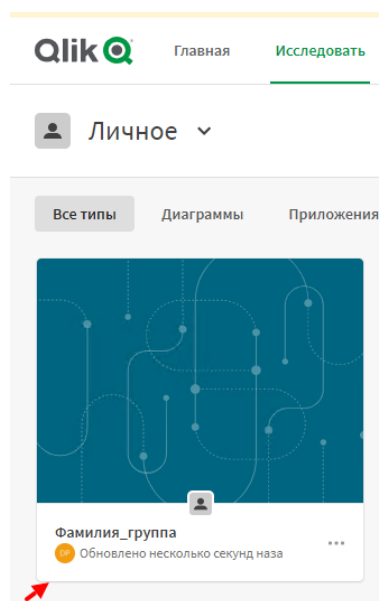


Рис. 5.5. Созданное приложение

5.2. Загрузка данных в аналитическое приложение

В приложение необходимо добавить данные. Для этого требуется нажать кнопку «Добавьте данные из файлов и других источников» (рис. 5.6).

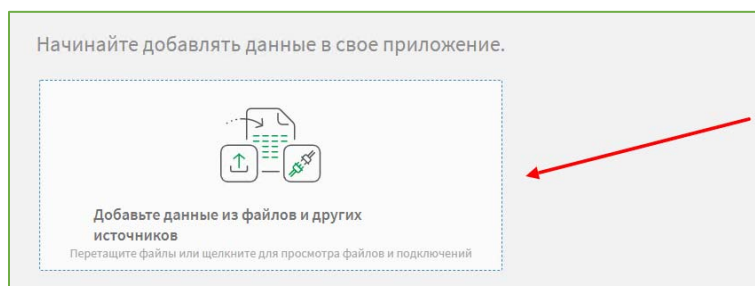


Рис. 5.6. Приложение без данных

Открывается окно «Добавить данные в приложение» (рис. 5.7).

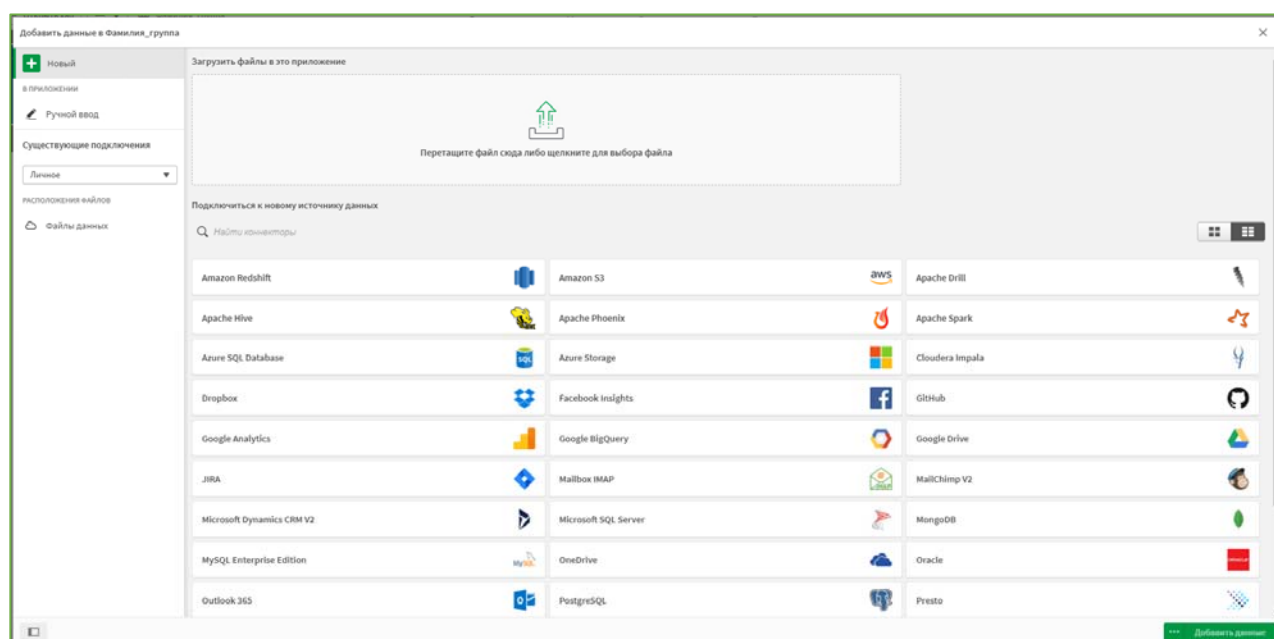


Рис. 5.7. Окно добавления данных

В качестве исходных данных необходимо скачать набор данных по ссылке <https://gist.github.com/jwalsh/ce1dc0436aba5b7a5c9666f47fa5a380>. Для этого необходимо нажать «Download ZIP» (рис. 5.8).

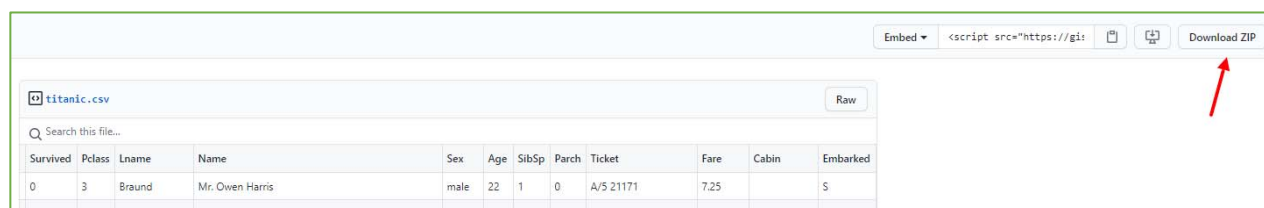


Рис. 5.8. Загрузка набора данных

После загрузки файла необходимо вернуться к окну «Добавить данные в приложение» и нажать на кнопку выбора файла (рис. 5.9).

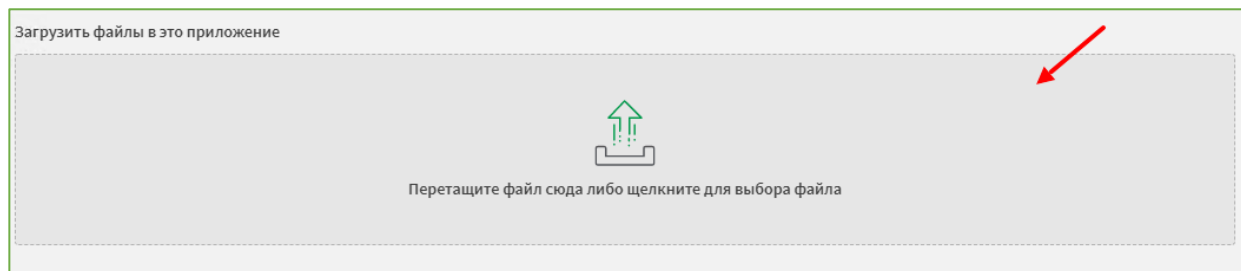


Рис. 5.9. Загрузка файла в приложение

После загрузки файла аналитическая платформа сама изучает файл и определяет формат файла, имена полей (в данном файле – встроенные), разделитель (в данном файле – запятая), и т.д. После анализа файла платформой необходимо нажать кнопку «Добавить данные» в нижнем правом углу (рис. 5.10).

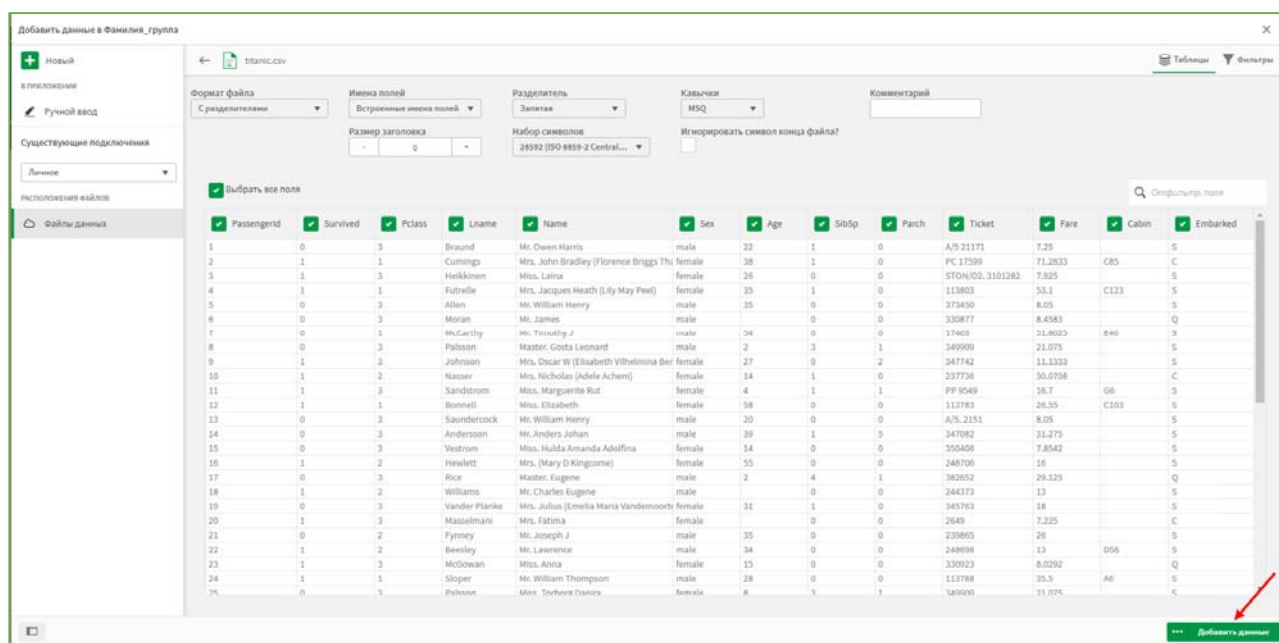


Рис. 5.10. Анализ добавленного файла

После успешного добавления должно появиться окно, сообщающее о том, что данные успешно загружены, которое можно закрыть, нажав соответствующую кнопку в правой нижней окне.

После этого раскрыть с помощью стрелочки вниз раздел «Подготовить» и выбрать «Редактор загрузки данных» (рис. 5.12).

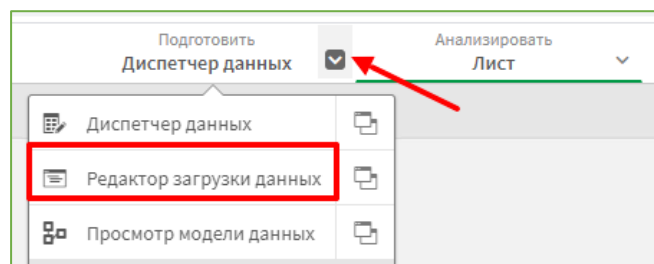


Рис. 5.12. Переход в редактор загрузки данных

В нём необходимо перейти в автоматически созданный раздел и нажать кнопку разблокировать (рис. 5.13).

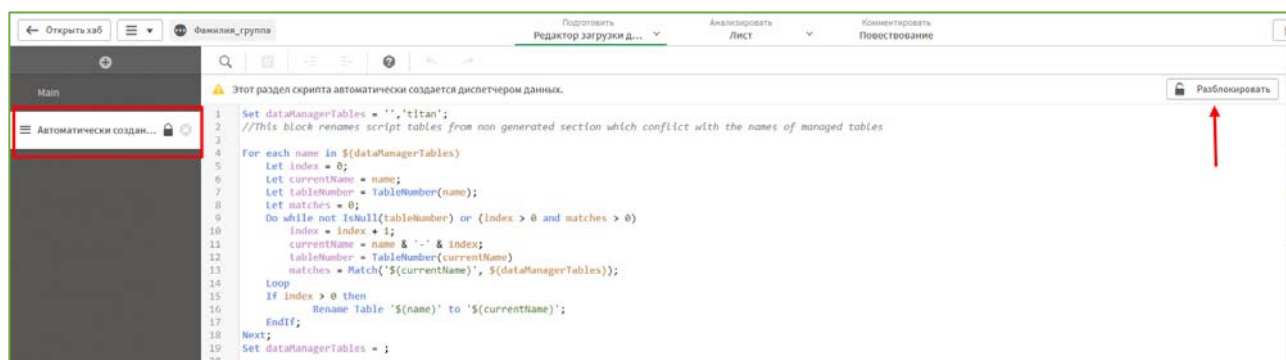


Рис. 5.13. Разблокировка скрипта загрузки данных

Далее необходимо подтвердить разблокировку скрипта (рис. 5.14).

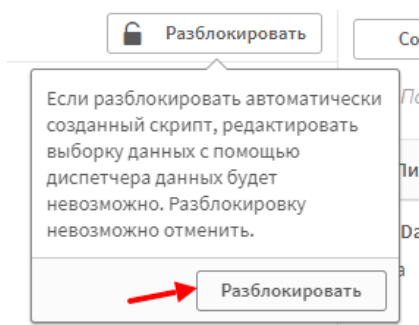


Рис. 5.14. Подтверждение разблокировки скрипта загрузки данных

Теперь можно создавать собственные поля с помощью специальных функций.

Наиболее часто используемой является функция условия – if.

Функция if имеет три параметра, condition, then и else, все из которых являются выражениями. Два остальных, then и else, могут относиться к любому типу. Синтаксис функции – *if(condition, then, else)*. Описание аргументов представлено в таблице 5.1.

Таблица 5.1 – Аргументы функции if

| Аргумент | Описание |
|-----------|---|
| condition | Выражение, которое интерпретируется логическим образом. |
| then | Выражение, которое может быть любого типа. Если элемент condition равен True, функция if возвращает значение выражения then. |
| else | Выражение, которое может быть любого типа. Если элемент condition равен False, функция if возвращает значение выражения else. |

Внутри выражения могут быть использованы логические операторы. Все логические операторы интерпретируют операнды в соответствии с определенной логикой и выдают результат True (-1) или False (0).

Список операторов:

- Not - логическое отрицание. Один из нескольких унарных операторов. Операция возвращает логическое отрицание операнда.
- And - Логическое И. Операция применяет логическое И к операндам.
- Or - Логическое ИЛИ. Операция возвращает логическое ИЛИ операндов.
- Xor - Логическое исключающее ИЛИ. Операция возвращает результат операции логического исключающее ИЛИ операндов.

Пример использования условной функции и логических операторов представлен на рисунке 5.15. В первой созданной строке говорится, что если возраст (*Age*) пассажира больше или равен (\geq) 18 лет, то он *Взрослый*, иначе он *Ребёнок* и это поле называется (*as*) *Возрастная_группа*.

Во второй строчке уже используются логические операторы. В ней описано условие, что если возраст (*Age*) пассажира меньше 3 лет либо пассажир 1 класса (*Pclass=1*) и возраст (*Age*) меньше 12, то поле *Бесплатная_няня* будет содержать *Да*, иначе *Нет*.

```
[Embed_kcs],
if(Age>=18,'Взрослый','Ребёнок') as Возрастная_группа,
if(Age<=3 or Pclass=1 and Age<12,'Да','Нет') as Бесплатная_няня
FROM [lib://DataFiles/titanic.csv]
(txt, codepage is 28592, embedded labels, delimiter is ',', msq);
```

Рис. 5.15. Пример использования функции и операторов

Так же стоит обратить внимание на синтаксис. После каждого поля должна ставиться запятая, кроме последнего.

После создания поля необходимо заново загрузить данные нажав на кнопку «Загрузить данные» в правом верхнем углу (рис. 5.16).

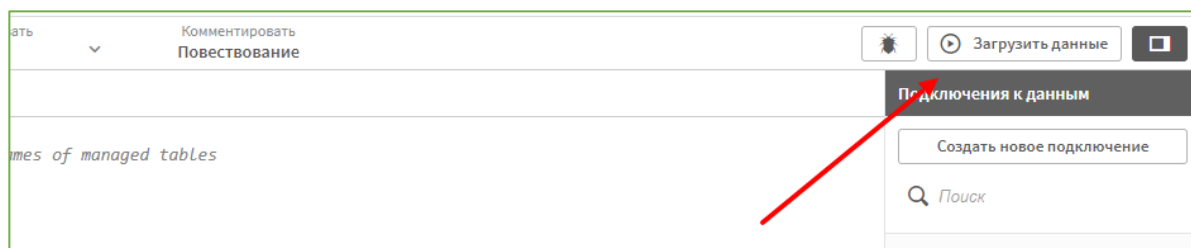


Рис. 5.16. Расположение кнопки «Загрузить данные»

Через несколько секунд должно появиться сообщение об успешной загрузке данных (рис. 5.17). Если появилось сообщение о ошибке загрузки значит была допущена ошибка в синтаксисе. После появления сообщения необходимо нажать кнопку «Закрыть» в правом нижнем углу окна.

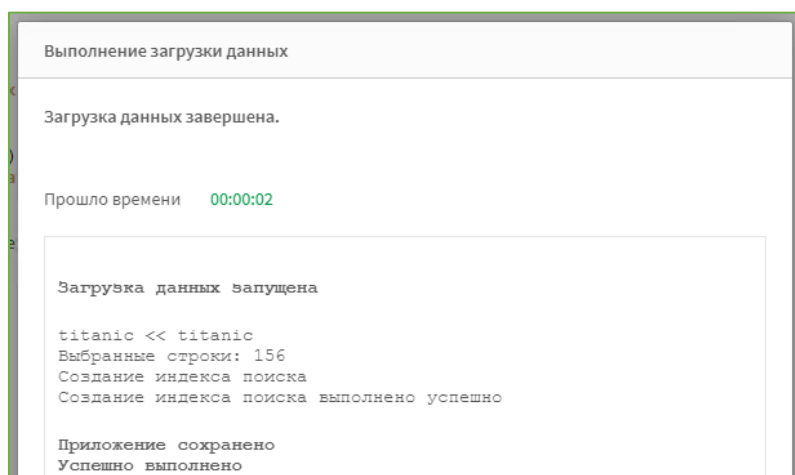


Рис. 5.17. Успешная загрузка данных

Далее необходимо раскрыть подменю «Подготовить» с помощью стрелки и открыть «Просмотр модели данных» (рис. 5.18).

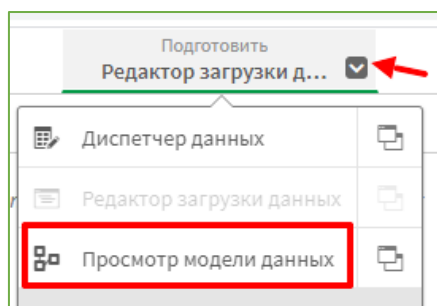


Рис. 5.18. Переход к инструменту «Просмотр модели данных»

Далее необходимо выбрать таблицу «titanic» щёлкнув по ней левой кнопкой мыши, раскрыть предварительный просмотр так же щёлкнув левой кнопкой мыши и можно будет увидеть новые созданные поля (рис. 5.19).

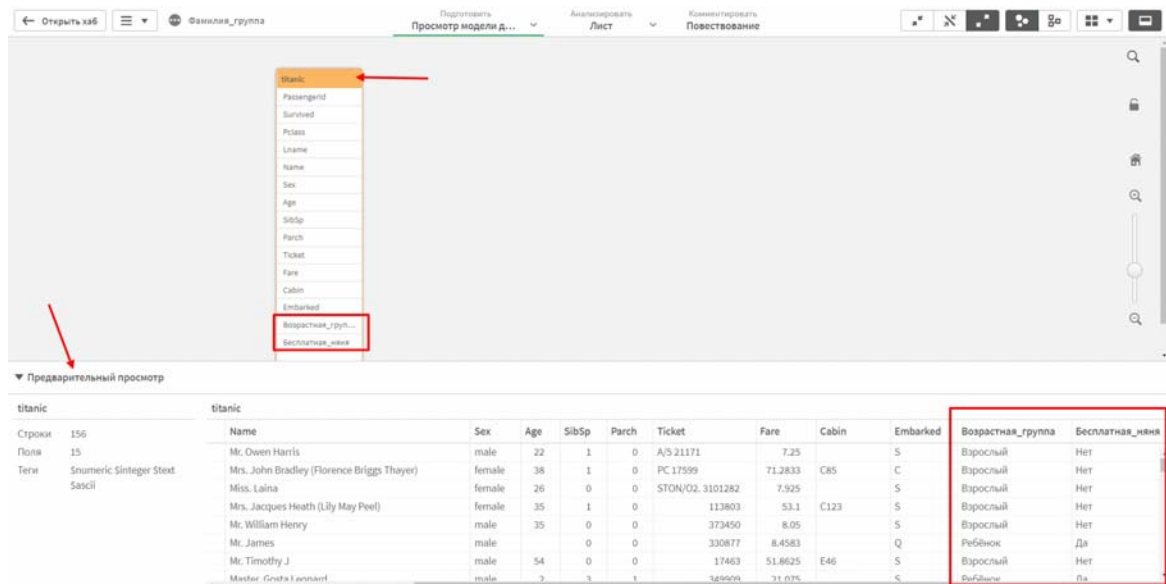


Рис. 5.19. Просмотр модели данных

5.3. Задание на лабораторную работу

Необходимо придумать свой вариант поля в скрипте загрузки данных на основе загруженных данных. В данном поле должен быть задействован минимум один логический оператор. Данные после создания поля должны быть успешно загружены.

Лабораторная работа №6

ПОСТРОЕНИЕ ВИЗУАЛИЗАЦИЙ В АНАЛИТИЧЕСКОЙ ПЛАТФОРМЕ QLIK SENSE

Цель работы:

- ознакомиться с визуальными элементами Qlik Sense;
- научиться строить визуализации;
- изучить язык анализа множеств.

6.1. Переход к редактированию визуальных элементов

Для начала, необходимо перейти в платформу по ссылке <https://www.qlikcloud.com/> и авторизоваться с помощью аккаунта, зарегистрированного в четвёртой лабораторной работе (рис 6.1).

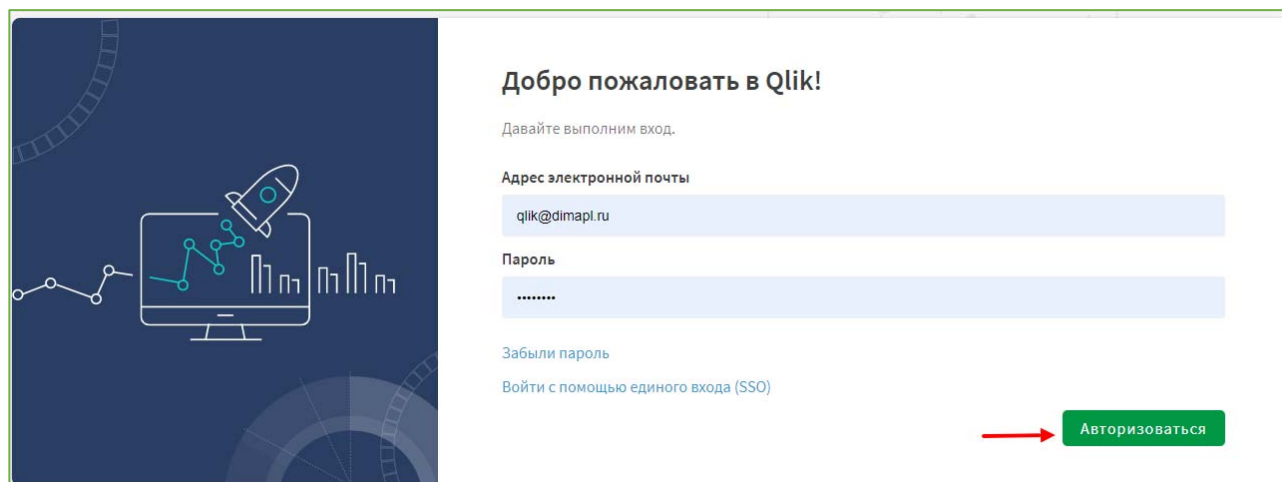


Рис. 6.1. Окно авторизации

Далее необходимо перейти в пункт «Исследовать» на главной странице (рис. 6.2).

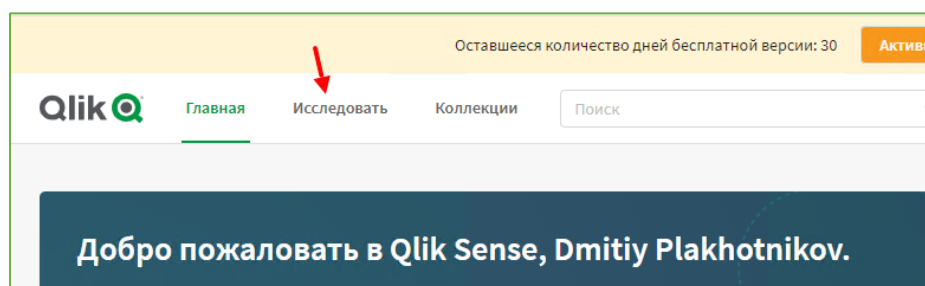


Рис. 6.2. Пункты главной страницы

В разделе приложений необходимо открыть созданное в прошлой лабораторной работе приложение (рис. 6.3).

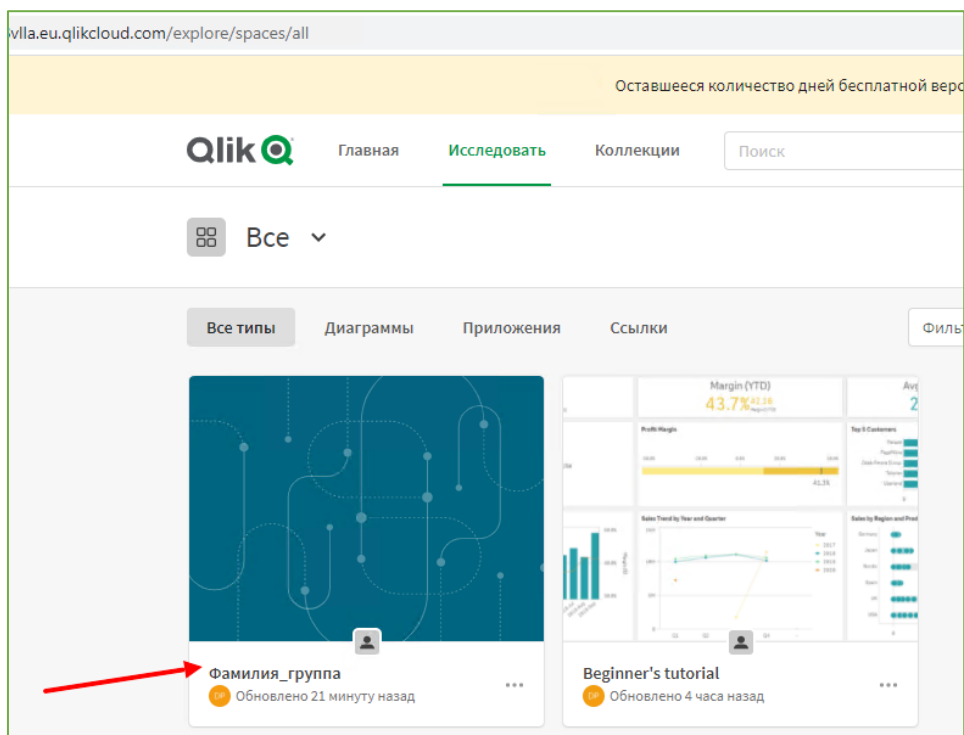


Рис. 6.3. Аналитические приложения

Далее необходимо перейти в пункт «Лист» раздела «Анализировать» (рис. 6.4).

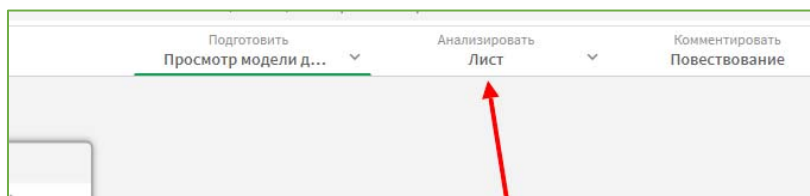


Рис. 6.4. Пункты меню аналитического приложения

После открытия листа для изменения визуализаций необходимо нажать «Изменить лист» (рис. 6.5).

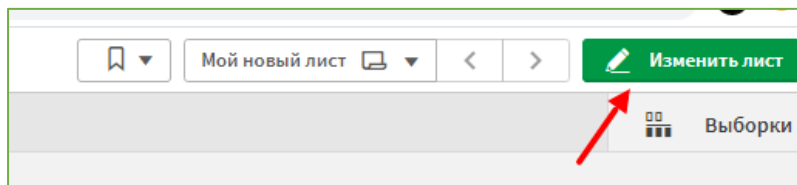


Рис. 6.5. Кнопка изменения листа

6.2. Добавление визуализаций на лист аналитического приложения

После нажатия кнопки появляется список объектов. В них необходимо выбрать «Диаграммы», далее нажать на «Таблицы» и, зажав левую кнопку мыши, переместить на лист, расположенный справа (рис. 6.6).

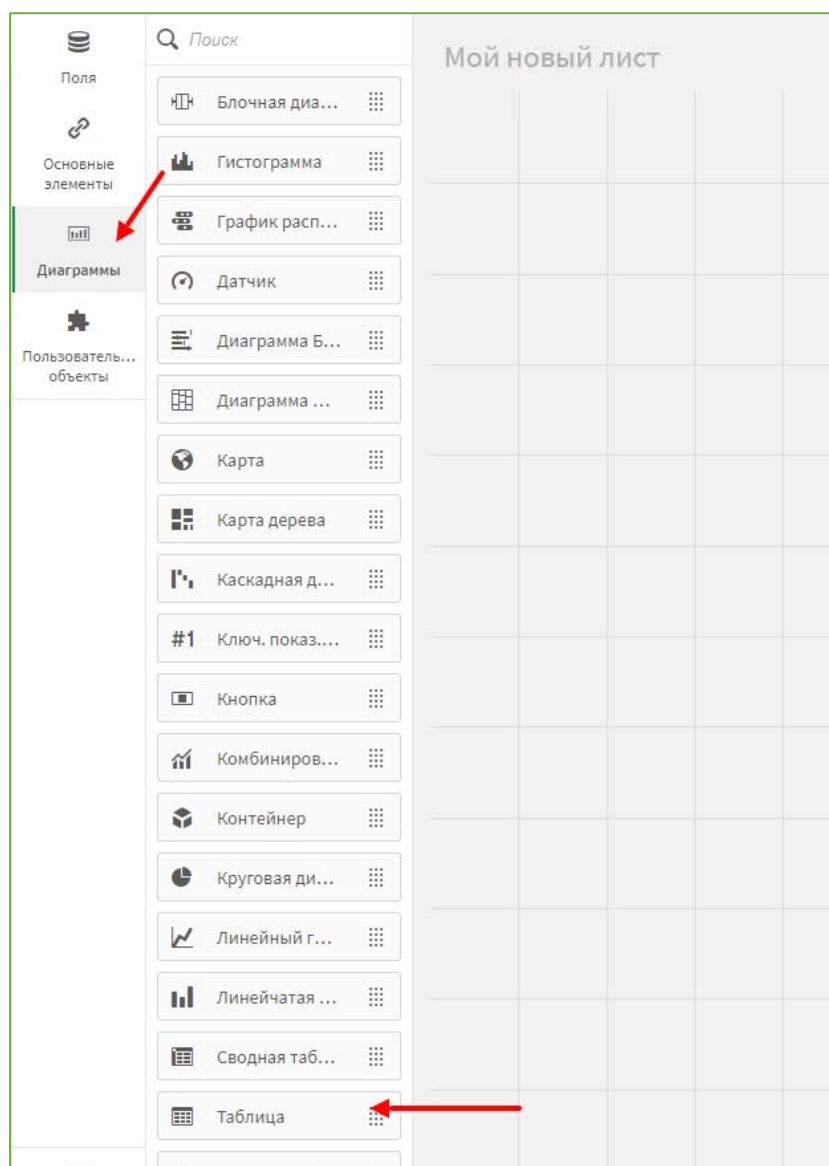


Рис. 6.6. Выбор необходимой визуализации

После добавления визуализации на ней имеются две кнопки «Добавить измерения» и «Добавить меры».

Меры – это значения данных, которые вы хотите отобразить на визуализации. Меры создаются на основе функций агрегации, например, Sum (сумма) или Avg (среднее), то есть все меры агрегируются из исходного источника по правилам подсчета их суммы или среднего (или минимума, максимума, количества).

Измерения – это категории, по которым вы хотите сгруппировать ваши данные; уровень, на котором вы хотите показать меры.

Для обычной таблицы необходимо добавить измерения. Для начала, можно добавить PassengerId (рис. 6.7).

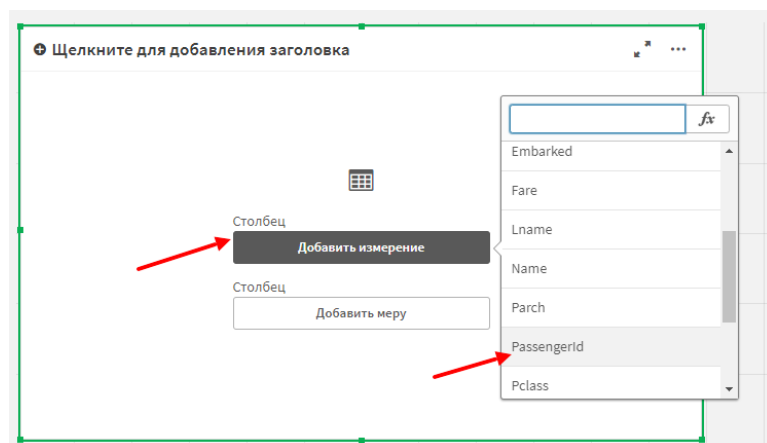


Рис.6.7. Добавление измерений в таблицу

В таблице отобразились идентификаторы пассажиров (рис. 6.8).

| PassengerId |
|-------------|
| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| 6 |
| 7 |
| 8 |
| 9 |

Рис. 6.8. Идентификаторы пассажиров

Далее необходимо добавить основные столбцы в таблицу. Для этого необходимо справа в подразделе «Данные» выбрать «Добавить столбец» -> «Измерение» (рис. 6.9). И поочерёдно повторить операцию для основных столбцов.

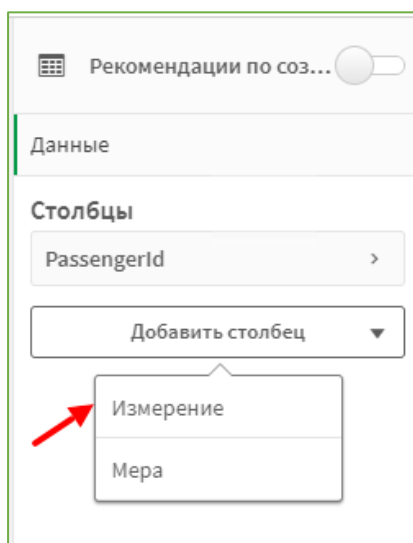


Рис. 6.9. Добавление измерения

После добавления измерений получаем таблицу пассажиров (рис. 6.10).

| PassengerId | Name | Sex | Pclass | Age |
|-------------|--|--------|--------|-----|
| 1 | Mr. Owen Harris | male | 3 | 22 |
| 2 | Mrs. John Bradley (Florence Briggs Thayer) | female | 1 | 38 |
| 3 | Miss. Laina | female | 3 | 26 |
| 4 | Mrs. Jacques Heath (Lily May Peel) | female | 1 | 35 |
| 5 | Mr. William Henry | male | 3 | 35 |
| 6 | Mr. James | male | 3 | |
| 7 | Mr. Timothy J | male | 1 | 54 |
| 8 | Master. Gosta Leonard | male | 3 | 2 |
| 9 | Mrs. Oscar W (Elisabeth Vilhelmina Berg) | female | 3 | 27 |
| 10 | Mrs. Nicholas (Adele Achem) | female | 2 | 14 |
| 11 | Miss. Marguerite Rut | female | 3 | 4 |
| 12 | Miss. Elizabeth | female | 1 | 58 |
| 13 | Mr. William Henry | male | 3 | 20 |
| 14 | Mr. Anders Johan | male | 3 | 39 |
| 15 | Miss. Hulda Amanda Adolfina | female | 3 | 14 |
| 16 | Mrs. (Mary D Kingcome) | female | 2 | 55 |
| 17 | Master. Eugene | male | 3 | 2 |
| 18 | Mr. Charles Eugene | male | 2 | |
| 19 | Mrs. Julius (Emelia Maria Vandemoortele) | female | 3 | 31 |
| 20 | Mrs. Fatima | female | 3 | |

Рис. 6.10. Таблица «Пассажиры»

Теперь можно добавить КРІ (ключевой показатель эффективности). Для этого в меню объектов нужно выбрать «Диаграммы» и левой кнопкой мыши перенести «Ключ. показ. эфф.» на лист (рис. 6.11).

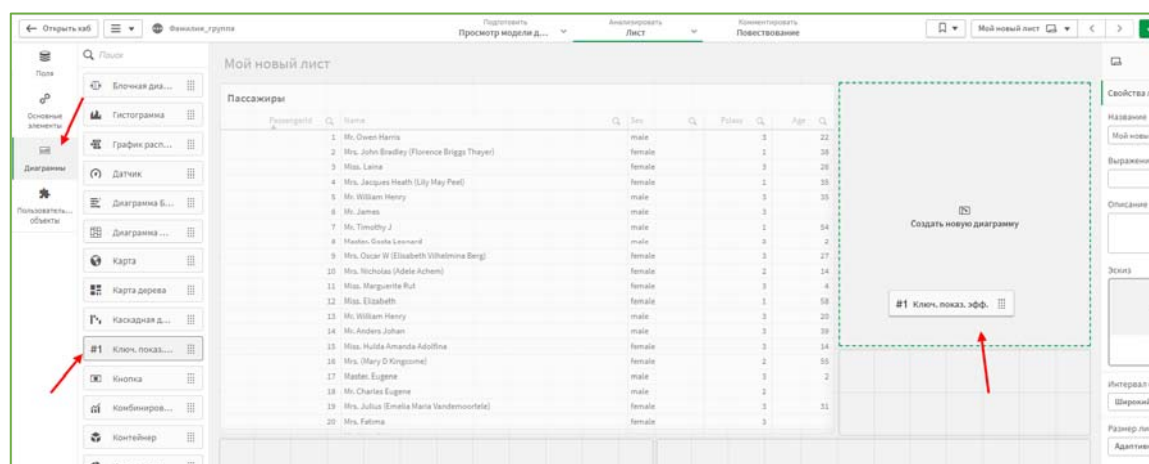


Рис. 5.11. Добавление КРІ на лист

После добавление на КРІ необходимо нажать единственную кнопку «Добавить меру» (рис. 6.12).

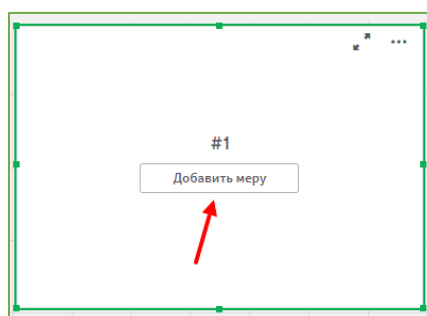


Рис. 6.12. Кнопка «Добавить меру» на КРІ

К примеру, можно выбрать меры по возрасту (*Age*) (рис. 6.13).

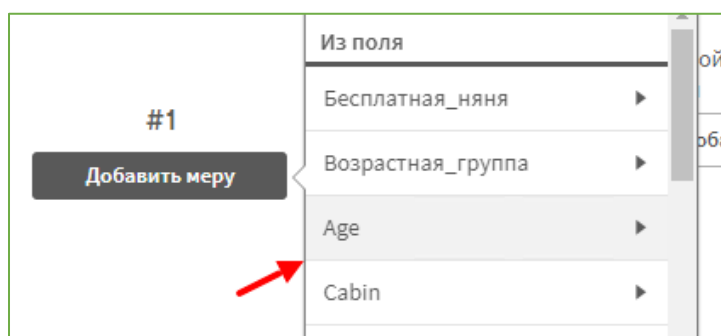


Рис. 6.13. Выбор меры для KPI

И выбрать среднее (*Avg*) от возраста (*Age*) (рис. 6.14).

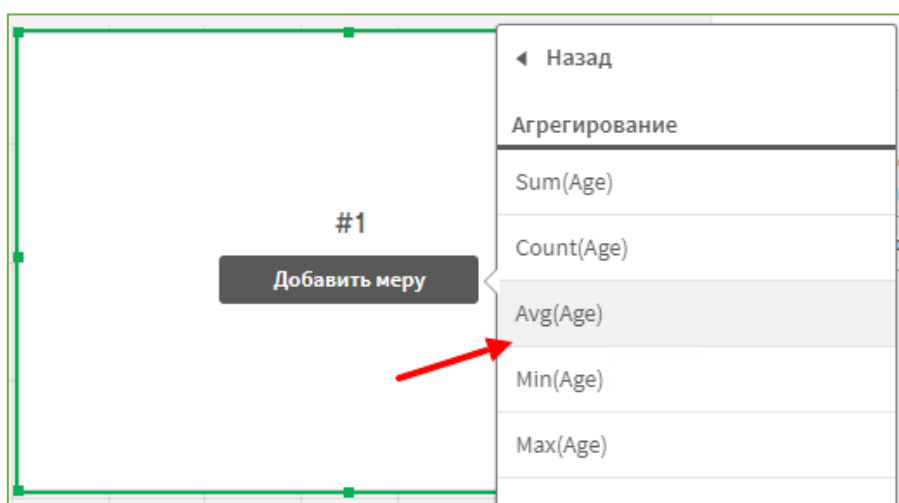


Рис. 6.14. Выбор агрегирования для возраста

Получается KPI отображающий средний возраст всех пассажиров. Для того чтобы поменять формулу необходимо щёлкнуть на KPI и справа в пункте «Данные» открыть соответствующую меру и нажать на знак «fx» (рис. 6.15).

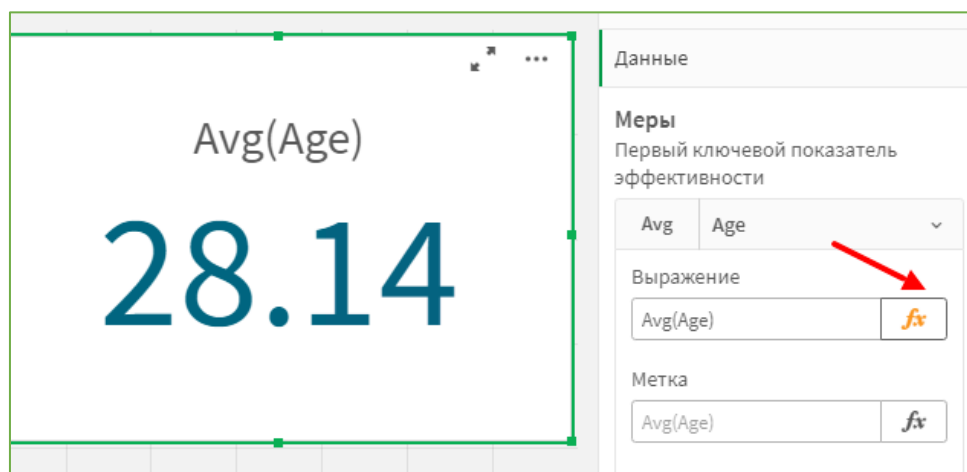


Рис. 6.15. Открытие меню редактирования меры

Для анализа ограниченного набора данных используется специальный язык анализа множеств (Set Analysis). Множества описываются при помощи фигурных скобок {}. Множество может быть изменено дополнительной или переобозначенной выборкой. Подобное изменение может быть записано в выражении множества. Модификатор состоит из одного или нескольких имен полей, за каждым из которых следует выборка, которая должна быть составлена на основе поля и заключена в «<» и «>». Имена и значения полей могут цитироваться как обычно.

Пример такой формулы представлен на рисунке 6.16.

A screenshot of a formula bar in a software application. The formula entered is `Avg({$<[Pclass]={ '1' }, [Sex]={ 'male' }>}Age)`. The text is color-coded: `Avg` is blue, the opening curly brace is blue, `{` is blue, `$` is blue, `<` is blue, `[Pclass]` is blue, `=` is blue, `{` is blue, `'1'` is green, `}` is blue, `,` is blue, `[Sex]` is blue, `=` is blue, `{` is blue, `"male"` is green, `}` is blue, `>` is blue, `}` is blue, and `Age` is blue. The closing parenthesis is blue.

Рис. 6.16. Использование анализа множеств

\$ – говорит о том, что используется текущая выборка,

Pclass={'1'} – показатель считается только для пассажиров 1 класса,

[Sex]="male" – и только для мужчин

Таким образом, с помощью данной формулы считается средний возраст мужчин в первом классе обслуживания (рис. 6.17).

A screenshot of a software interface showing the result of a calculation. At the top, the formula `Avg({$<[Pclass]='1',[Sex]='...}` is partially visible. Below it, the result `39.32` is displayed in a large, bold, blue font.

Рис. 6.17. KPI среднего возраста мужчины в 1 классе обслуживания

6.3. Задание на лабораторную работу

Необходимо придумать свой вариант ключевого показателя эффективности на основе загруженных данных. В мере должен быть обязательно задействован анализ множеств содержащий, минимум один аргумент. Дополнительно можно добавить другие визуализации.

Котова Елена Евгеньевна,
Плахотников Дмитрий Петрович

Введение в аналитику больших данных

Учебно-методическое пособие

Редактор Г. Г. Петров

Подписано в печать 30.10.20. Формат 60×84 1/16.

Бумага офсетная. Печать цифровая. Печ. л. 3,0.

Гарнитура «Times New Roman». Тираж 72 экз. Заказ 000.

Издательство СПбГЭТУ «ЛЭТИ»
197376, С.-Петербург, ул. Проф. Попова, 5