

**Министерство науки и высшего образования Российской Федерации  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»  
(Университет ИТМО)**

Факультет      **Инфокоммуникационных технологий**

Образовательная программа **Интеллектуальные системы в гуманитарной сфере**

Направление подготовки **45.03.04 Интеллектуальные системы в гуманитарной сфере**

## **КУРСОВОЙ ПРОЕКТ**

Выбранная тема: «Разработка дашбордов для визуального анализа тенниса»

Обучающийся Шурубова Прасковья Михайловна, К3125

Работа выполнена с оценкой \_\_\_\_\_

Преподаватель (и):

\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
(подпись)

Дата 20.12.2022

Санкт-Петербург 2022

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
ОСНОВНАЯ ЧАСТЬ	4
1 Подготовка к созданию дашбордов	
1.1 Сбор данных	4
1.2 Разработка макетов дашбордов	5
2 Реализация дашбордов в Power BI	6 - 8
3 Создание дашбордов в Dash	9
4 Имитация теннисных матчей	10
5 Взаимодействие с командой и руководителем проекта	11
6 Результаты работы	12
ЗАКЛЮЧЕНИЕ	13
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	14
ПРИЛОЖЕНИЕ	15 – 18

## ВВЕДЕНИЕ

**Актуальность рассматриваемой темы.** Актуальность темы курсовой работы обусловлена повышенным интересом к теннису по всему миру, в том числе, в России. Визуализации смогут обеспечить всех причастных людей наглядной, полезной и понятной информацией. Тренеров – для выстраивания плана подготовки игрока с учетом его персональных и карьерных особенностей, особенностей будущего соперника по корту. Букмекеров – для формулировки заключений о форме игрока и шансах на победу. Болельщиков – для удовлетворения их интересов.

**Цель работы** – это реализованный набор интерактивных визуализаций для тенниса, позволяющих получать информацию о теннисистах, их результатах в карьере и персональных особенностях, а также реализация решения, позволяющего имитировать теннисный матч для формирования прогноза на игру.

### **Задачи:**

- Определиться с параметрами отбора информации про теннисистов,
- собрать и отсортировать данные с определёнными критериями,
- настроить API для ежедневного обновления данных,
- разработать макеты дашбордов,
- реализовать рабочие дашборды в DASH,
- реализовать рабочие дашборды в BI-системе,
- написать код, позволяющий проводить имитацию матча,
- протестировать готовые дашборды.

# ОСНОВНАЯ ЧАСТЬ

## 1 Подготовка к созданию дашбордов

### 1.1 Сбор данных

Перед непосредственным созданием базы данных требовалось определиться с параметрами игроков, по которым будет происходить парсинг и сортировка в реляционном формате с помощью Python (рисунок 1). Сбор информации шёл только по теннисистам-мужчинам. Были найдены ресурсы с подходящей статистикой по теннису и готовая БД, которую можно использовать в нашем проекте. Для прослеживания динамики теннисистов и для реализации взаимодействия таблиц между собой создавался программный интерфейс, благодаря которому при выборе одного параметра у игрока обновлялись и другие данные, связанные с ним. В итоге была реализована структура проекта, разбитого на три сервиса: БД, API для взаимодействия с БД и дашборды.

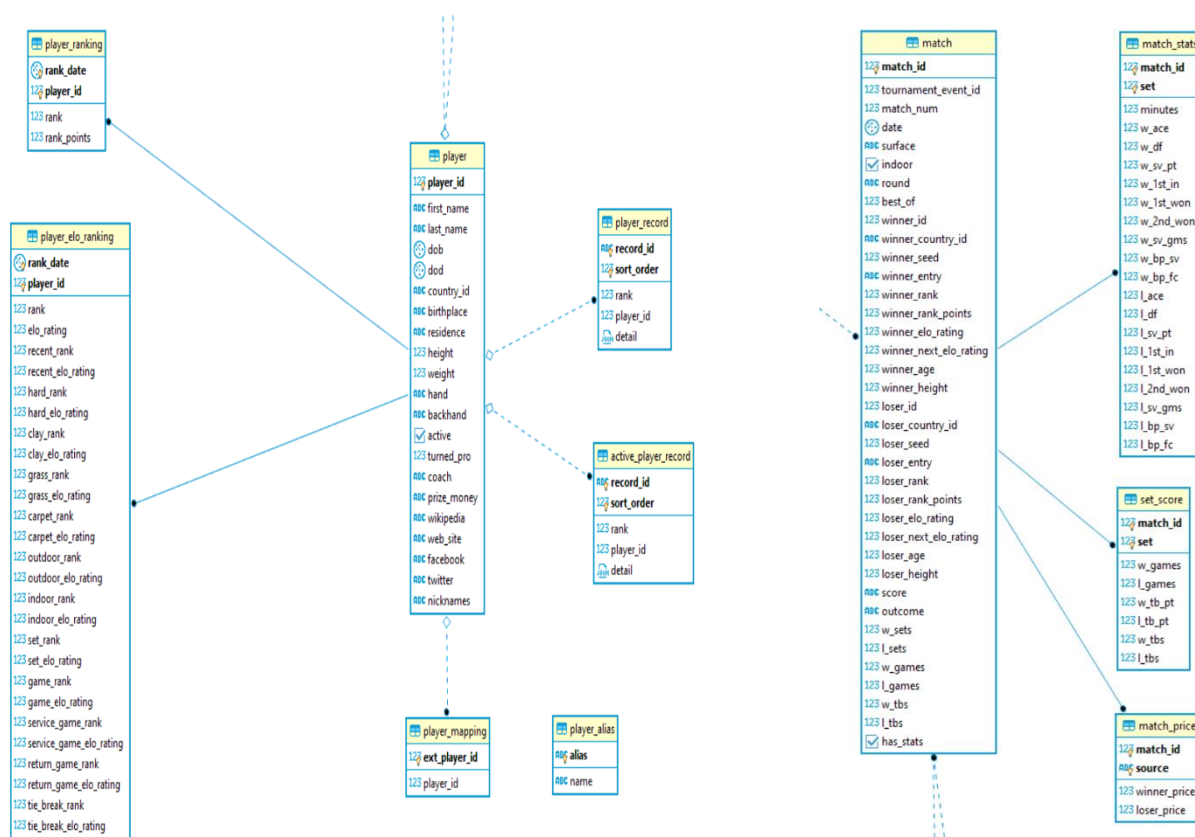


Рисунок 1 – Реализация базы данных в Python

## 1.2 Разработка макетов дашбордов

Для создания макетов в Dash и Microsoft Power BI необходимо было ознакомиться с функционалом каждой из систем. При разработке макетов в Dash ребята старались проанализировать данные о матчах каждого игрока, то есть визуализировать количество наград у игроков, распределение теннисистов по имеющимся титулам, количество побед по возрастам.

Так как я отвечала за часть реализации визуализаций в Power BI, то я проходила вводные курсы по работе в этом приложении, а также смотрела видео людей, которые уже давно работают на этой платформе. Это позволило мне перенять полезные советы по расположению графиков, определиться с шаблонами, подобрать цветовую гамму. Передо мной стояла задача отображения сводной аналитики, поэтому для своего макета я решила показать дату рождения, рост, вес, рабочую руку, ранг игрока, фотографию, страну, за которую он играет, а также суммарное количество баллов, которое принесли теннисисты своей стране (рисунок 2).

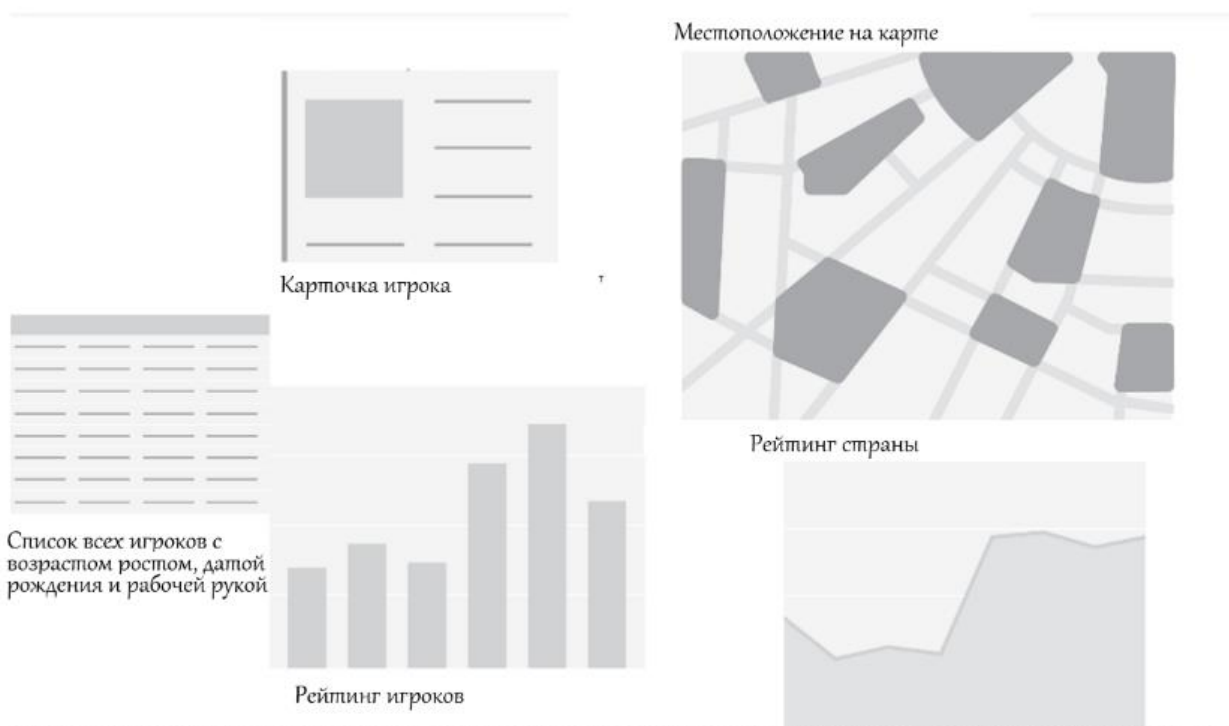


Рисунок 2 – Макет в Power BI

## 2 Реализация дашбордов в Power BI

Я была ответственна за реализацию дашбордов в BI-системе. При распределении обязанностей по выполнению курсового проекта я выбрала визуализации именно в этой системе, так как она решает ряд важнейших задач:

- Сбор, структурирование и хранение актуальных данных в одном хранилище,
- анализ большого объема данных для создания отчетов и бизнес-решений с учетом аналитики,
- моделирование всевозможных решений для бизнеса и удачного развития компании.
- создание стратегической и оперативной отчетности для оповещения руководства об отклонениях показателей эффективности работы.
- систематизация и сохранение полученных выводов для передачи в работу сотрудникам.

Конкретно передо мной стояли задачи:

- Ознакомиться с функционалом ведущих BI-систем,
- проанализировать BI-системы, доступные на российском рынке.
- опираясь на весовой коэффициент, подобрать наиболее подходящий продукт, в котором будет реализован набор общеприменимых инструментов,
- разобраться, как работать в BI-системе,
- определиться с вариантом развёртывания: локальная BI или облачная BI,
- объединить информацию в единую аналитику.

В ходе анализа различных BI-систем, которые доступны на российском рынке, а именно выявления коэффициентов по каждому из критериев, выбор

пал на Microsoft Power BI. Единство платформы и доступность мобильной версии стали решающими факторами. Ещё при создании макетов был проведён анализ различных систем, поэтому приступить к визуализации оказалось не так трудно, как импортировать данные из GitHub. При собирании образов из GitHub мои коллеги воспользовались методом контейнеризации, а именно Docker file. При копировании информации потребовалось обратиться к объектно-реляционной системе управления базой данных PostgreSQL. Это позволило определиться с форматом развёртывания – локально. Для связывания данных в Power BI было необходимо соединить каждую из таблиц в самом приложении через модель, создав связи между таблицами.

Так как информации было слишком много, более 14 тысяч человек, с коллегами из моей команды мы приняли решение взять мужчин-теннисистов с рангом не ниже 40, которые до сих пор принимают участие в турнирах. В итоге получилось 105 человек. Далее я начала собирать фотографии игроков, чтобы пользователь, который не может представить теннисиста только по его фамилии и имени, сразу имел представление о человеке. Снимки подбирались в горизонтальном формате, исходя из формата шаблона-карточка. В целом для реализации использовались такие макеты, как таблица для отображения имён игроков, их возраста, роста, веса и рабочей руки, диаграмма с областями с накоплениями для суммарных очков у каждой страны, гистограмма с группировкой для ранга интересующего нас теннисиста и его положения на общем фоне, карточка для фотографии и географическая карта для местоположения страны на карте. Цветовая гамма была подобрана для более комфортного восприятия пользователями: различные оттенки серого, синий, голубой и глубокий зелёный (рисунки 3,4).

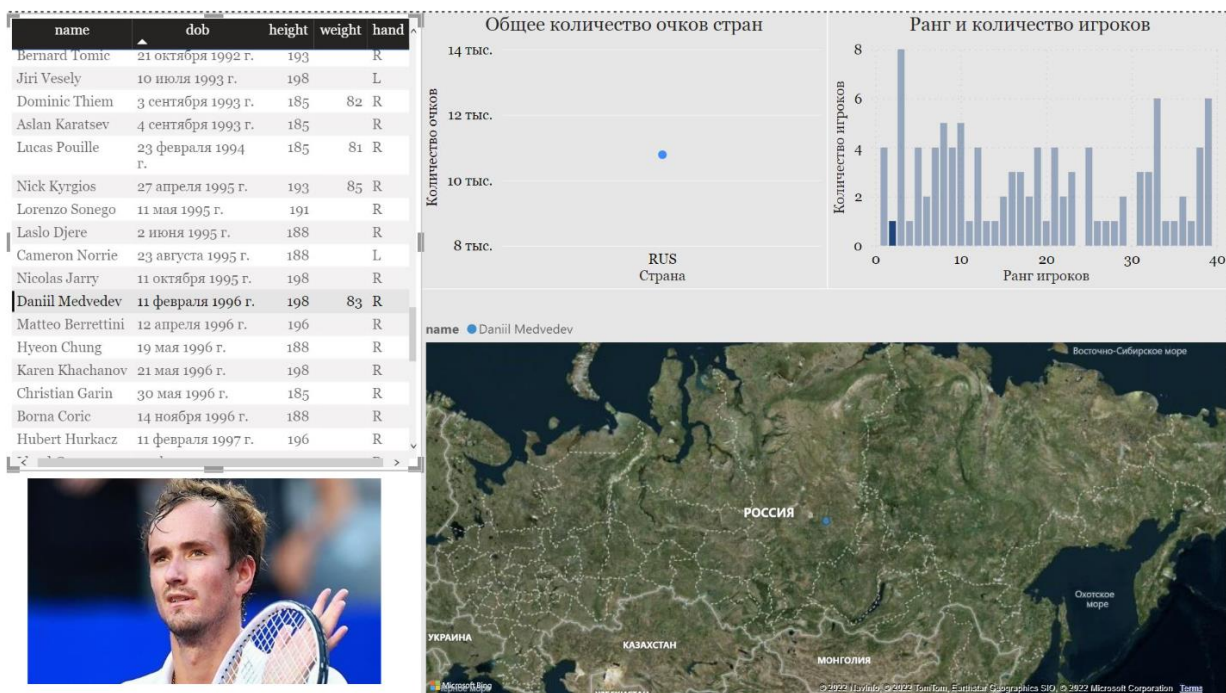


Рисунок 3 – Характеристика Даниила Медведева

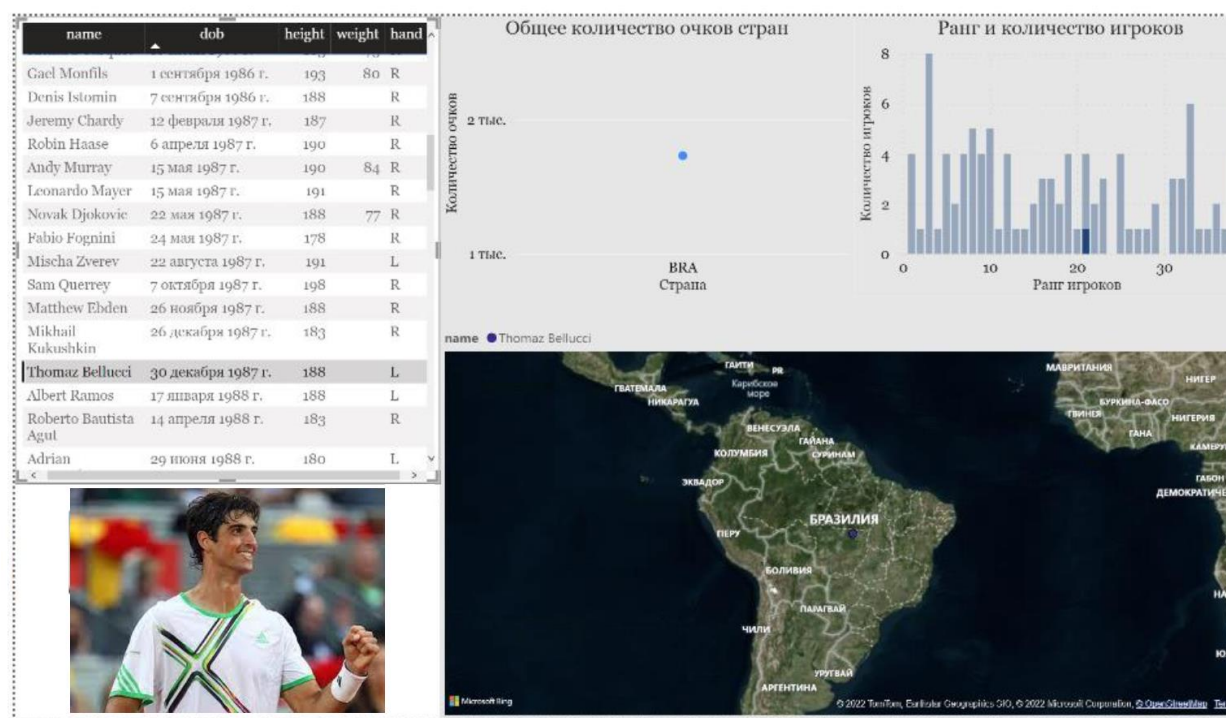


Рисунок 4 – Характеристика Томаса Беллуччи



### 3 Создание дашбордов в Dash

После сравнения функциональных возможностей библиотек Python для выбора оптимальной библиотеки визуального анализа тенниса Plotly, опираясь на созданные макеты, мои коллеги расположили визуализации на нескольких страницах, так как брали довольно-таки большой объём данных. Для отображения количество наград рассматривалось только 50 человек (рисунок 5), однако при визуализации титулов требовались данные практически обо всех теннисистах. Цветовая гамма подбиралась примерно такая же, как и для дашбордов в Power BI. При создании выпадающего списка с параметрами использовались практически все цвета из палитры, чтобы проще было найти страну или интересующего теннисиста.

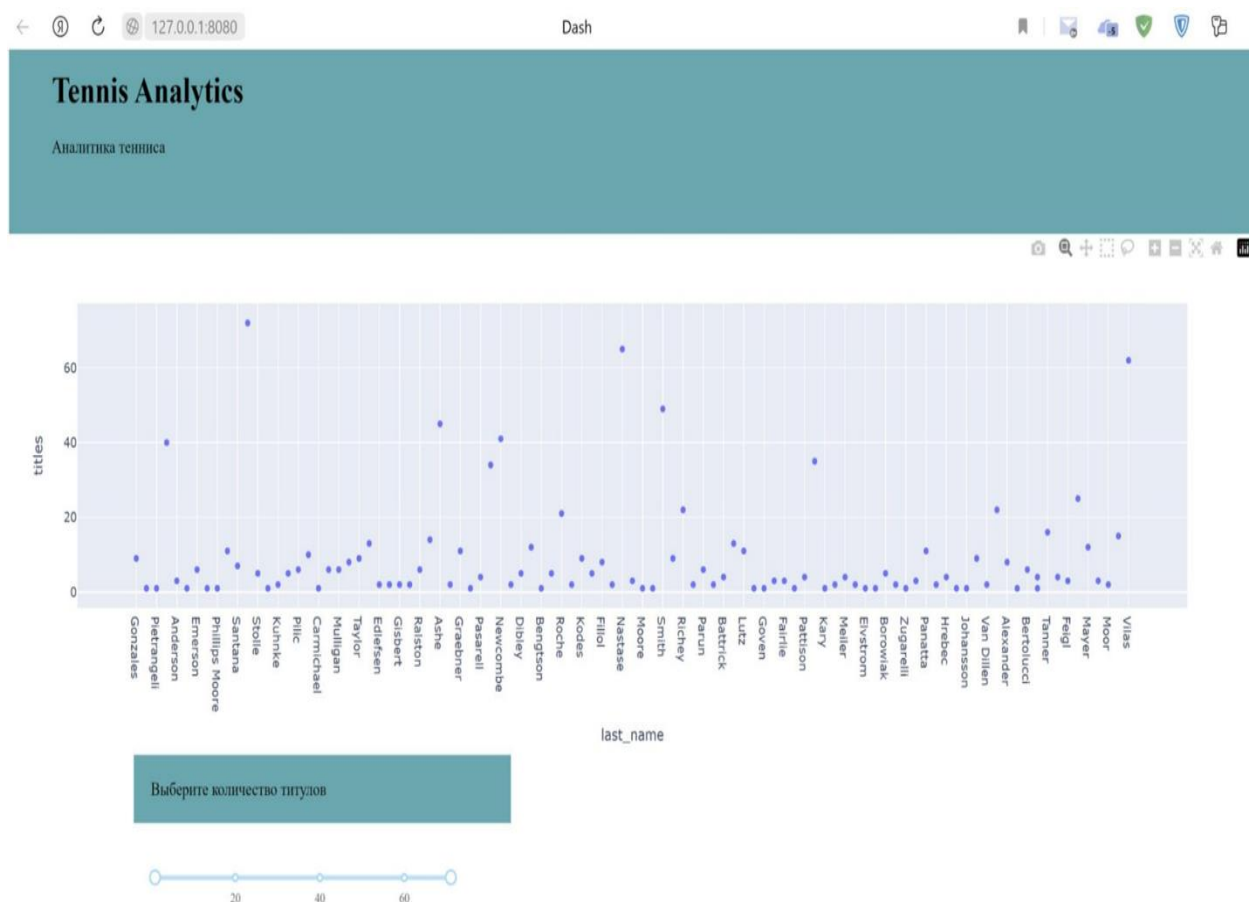


Рисунок 5 – Дашборд по количеству наград у 50 игроков

## 4 Имитация теннисных матчей

Существует много различных способов, применяемых для прогнозирования исхода игры. Перед моими коллегами стояла задача имитации теннисных матчей с помощью Python с целью прогнозирования того, как может складываться игра. На исход игры безусловно влияет множество факторов, но из-за ограниченности времени параметры предполагаемого матча были сужены до отслеживания таких показателей как вероятность того, что или 1, или 2 игрок выиграет очко на своей подаче. Код, который был написан для решения этой задачи, состоит из нескольких отдельных функций и главного цикла, который управляет этими функциями. Результат кода – отображение количества очков, набранных игроком 1 против игрока 2 (рисунок 6).

File - prediction

```
1 D:\dash\Scripts\python.exe D:\TennisDash\api\src\
  prediction.py
2 Player 1 0-0|[0-0]
3 Player 1 0-15|[0-0]
4 Player 1 15-15|[0-0]
5 Player 1 15-30|[0-0]
6 Player 1 30-30|[0-0]
7 Player 1 30-40|[0-0]
8 Player 1 0-0|[0-0]
9 game point
10 Player 1 A-D|[0-0]
11 Player 1 0-0|[0-0]
12 Player 1 0-A|[0-0]
13 Player 1 0-D|[0-0]
14 game point
15 Player 1 A-D|[0-0]
16 Player 1 0-D|[0-0]
17 game point
18 Player 1 A-D|[0-0]
19     Player 1: 8, Player 2: 6
20 Player 2 0-0|[0-1]
21 Player 2 0-15|[0-1]
22 Player 2 15-15|[0-1]
23 Player 2 15-30|[0-1]
24 Player 2 30-30|[0-1]
25 Player 2 30-40|[0-1]
26 Player 2 0-0|[0-1]
27 game point
28 Player 2 A-D|[0-1]
29     Player 2: 5, Player 1: 3
30 Player 1 0-0|[1-1]
31 Player 1 15-0|[1-1]
32 Player 1 30-0|[1-1]
33 Player 1 30-15|[1-1]
34 game point
35 Player 1 40-15|[1-1]
36 game point
37 Player 1 40-30|[1-1]
38     Player 1: 4, Player 2: 2
39 Player 2 0-0|[1-2]
40 Player 2 0-15|[1-2]
41 Player 2 15-15|[1-2]
42 Player 2 30-15|[1-2]
```

Page 1 of 10

File - prediction

```
43 game point
44 Player 2 40-15|[1-2]
45     Player 2: 4, Player 1: 1
46 Player 1 0-0|[2-2]
47 Player 1 15-0|[2-2]
48 Player 1 30-0|[2-2]
49 Player 1 30-15|[2-2]
50 Player 1 30-30|[2-2]
51 game point
52 Player 1 40-30|[2-2]
53     Player 1: 4, Player 2: 2
54 Player 2 0-0|[2-3]
55 Player 2 15-0|[2-3]
56 Player 2 30-0|[2-3]
57 Player 2 30-15|[2-3]
58 Player 2 30-30|[2-3]
59 game point
60 Player 2 40-30|[2-3]
61     Player 2: 4, Player 1: 2
62 Player 1 0-0|[3-3]
63 Player 1 15-0|[3-3]
64 Player 1 30-0|[3-3]
65 game point
66 Player 1 40-0|[3-3]
67 game point
68 Player 1 40-15|[3-3]
69     Player 1: 4, Player 2: 1
70 Player 2 0-0|[3-4]
71 Player 2 15-0|[3-4]
72 Player 2 15-15|[3-4]
73 Player 2 30-15|[3-4]
74 Player 2 30-30|[3-4]
75 game point
76 Player 2 40-30|[3-4]
77     Player 2: 4, Player 1: 2
78 Player 1 0-0|[4-4]
79 Player 1 0-15|[4-4]
80 Player 1 15-15|[4-4]
81 Player 1 15-30|[4-4]
82 Player 1 30-30|[4-4]
83 game point
84 Player 1 40-30|[4-4]
85     Player 1: 4, Player 2: 2
```

Page 2 of 10

Рисунок 6 – Результат работы программного кода

## **5 Взаимодействие с командой и руководителем проекта**

Я считаю, у нас получилась очень слаженная работа с моими коллегами. Изначально всё согласовалось и оговаривалось в общей группе. Мы совместно приняли решение, какие данные брать про теннисистов. В команде по Power BI я была вместе с Исмагиловой Мариной и Приваловым Кириллом. Ребята были на связи на протяжении всего курсового проекта. Несмотря на создание базы данных, Кирилл успевал помогать нашей команде и всегда был готов подсказать, как правильно расположить макеты. Шемчук Александр очень помог при импортировании данных в Microsoft Power BI из GitHub. Карина участвовала в нахождении фотографий, связывание данных и готовила вместе со мной выступление по курсовой. Румянцев Александр сопровождал нашу команду в течение всего проекта, помог определиться с BI-системой, давал полезные советы по оформлению промежуточных отчётов и по тексту выступления, заблаговременно напоминал нам о всех сроках сдачи.

## **6 Результаты работы**

В результате нашей работы были собраны данные про теннисистов, были созданы дашборды разных моделей в Power BI и Dash, в которых визуализированы сводная информация про игроков и подробные аналитические данные по турнирам. Также была написана программа на Python, которая способна предсказать исход игры, исходя из физических данных теннисистов. Кроме того, я научилась использовать метод контейнеризации, что очень востребовано при работе на разных операционных системах, и получила ценные навыки при работе с Power BI.

## **ЗАКЛЮЧЕНИЕ**

В заключении можно полностью утверждать, что цель курсового проекта была достигнута. Наша команда создала дашборды, которые показывают реальную динамику изменения про каждого из теннисистов и программа, которая прогнозирует исходы игр. Во время написания работы я очень хорошо ощутила свои сильные и слабые стороны, поняла, как важно уметь работать в команде, просить помощи, если никак не получается и, конечно получила массу новых знаний.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Изучаем Docker [Электронный ресурс] // URL:  
<https://habr.com/ru/company/ruvds/blog/438796/>
2. Как работать с Microsoft Power BI [Электронный ресурс] // URL:  
<https://netpeak.net/ru/blog/kak-rabotat-s-microsoft-power-bi-podrobnoe-rukovodstvo/>
3. Дашборды в Power BI [Электронный ресурс] // URL:  
<https://biprosto.ru/power-bi/dashboard.html>
4. Настройка и начало работы с PostgreSQL [Электронный ресурс] // URL: <https://learn.coderslang.com/ru/0119-setting-up-and-getting-started-with-postgresql/>

## **ПРИЛОЖЕНИЕ**

### **ТЕХНИЧЕСКОЕ ЗАДАНИЕ**

#### **1. Название проекта**

Разработка дашбордов для визуального анализа тенниса

#### **2. Цель (назначение)**

Реализовать набор интерактивных визуализаций (дашбордов) для тенниса, позволяющих получать информацию о теннисистах, их результатах в карьере и персональных особенностях. Также необходимо реализовать решение, позволяющее имитировать теннисный матч для формирования прогноза на игру.

#### **3. Сроки выполнения**

Начало: 01 ноября 2022 г.

Конец: 20 декабря 2022 г.

#### **4. Исполнители проекта**

Исполнители проекта:

1. Румянцев Александр Андреевич (руководитель проекта);
2. Боровский Максим Олегович (К3126);
3. Исмагилова Карина Ильдусовна (К3126);
4. Кибенко Дарья Владимировна (К3126);
5. Лобова Полина Михайловна (К3126);
6. Макунина Арина Артуровна (К32421 - бывш. К3142);
7. Пожарицкий Павел Денисович (К3125);
8. Привалов Кирилл Алексеевич (К3124);

9. Шемчук Александр Евгеньевич (К3140);
10. Шурубова Прасковья Михайловна (К3125);
11. Чаккери Юрий (К3125).

## **5. Термины и сокращения**

Тѐннис или большóй теннис — вид спорта, в котором соперничают либо два игрока, либо две команды, состоящие из двух игроков;

Парсинг – процесс сбора большого массива информации с последующей обработкой и анализом;

Dash - веб-фреймворком для визуализации данных;

ВІ-система – это набор инструментов для сбора, анализа и наглядного представления данных;

Дашборд – это современный формат сбора и визуального представления массивов данных. Это аналитическая панель с понятным интерфейсом для интерактивного взаимодействия с огромным количеством постоянно изменяющихся показателей. Мощные средства аналитики обрабатывают данные, сравнивают цифры и выдают человеку перед монитором индивидуально настроенные визуализации.

## **6. Требования к проекту**

В Dash разрабатываются больше технические визуализации, направленные на анализ особенностей игры теннисиста, особенности его подачи, игры в обороне и атаке, зависимость его игры от типа корта.

В ВІ-системе разрабатываются дашборды, формирующие конечный взгляд, сводный отчет по спортсмену. Предназначается для удобной работы со статистической информацией.

Средства разработки:



- Парсер: Python / JavaScript;
- СУБД: MySQL, PostgreSQL, Microsoft SQL (на выбор участника);
  - Веб-фреймворк: Dash. Библиотеки – plotly, matplotlib, seaborn;
- BI-система: MS Power BI, QlikView, Tableau, Metabase (на выбор участника).

Требования к визуализациям:

- Дашборды не перегружены данными. Разбиение дашбордов по листам, в зависимости от тематики информации на них;
- Все показатели в динамике;
- Доступны сравнения;
- Данные одного типа – одним цветом.

Пользователи:

Тренерский штаб, теннисисты, букмекеры.

## 7. Содержание работы

№	Этапы плана	Сроки выполнения этапов	Ответственный за этап	Результат этапа
1	Разработка технического задания	14.11	Румянцев А.А.	Готовое ТЗ
2	Парсинг данных, подготовка БД	15.11 - 25.11	Шемчук А.Е.; Привалов К.А.; Лобова П.М.	Готовая БД с API для ежедневного обновления данных
3	Разработка макетов дашбордов	15.11 - 25.11	Чаккери Ю.; Пожарицкий П.Д.	Готовые макеты будущих дашбордов
4	Реализация дашбордов в Dash	25.11 - 07.12	Кибенко Д.В.; Боровский М.О.; Шемчук А.Е.	Рабочие дашборды в Dash, соответствующие описанным требованиям
5	Реализация дашборда в BI-системе	25.11 - 07.12	Исмагилова К.И.;	Рабочие дашборды в BI-системе, соответствующие

			Шурубова П.М.; Привалов К.А.;	описанным требованиям
6	Реализация имитации матча	25.11-07.12	Макунина А.А.	Готовый код, позволяющий проводить имитацию матча
7	Тестировани е, доработка	08.11 - 12.12	Все участники проекта	Протестированны е, готовые к использованию дашборды
8	Написание отчетов и презентации на защиту	13.12 - 18.12	Все участники проекта	Готовые презентации и отчеты, оформленные по ГОСТ
9	Защита проекта (сдача отчета и представлени е доклада с презентацией )	19.12	Все участники проекта	Защищенный на “Отлично” проект

## 8. Основные результаты работы и формы их представления

Готовые дашборды и готовый теннисный симулятор матча, при помощи которых возможно проводить анализ теннисистов, их персональных особенностей, изучать статистику, историю противостояний, прогнозировать исходы матчей.