|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Технологии и инструментарий анализа больших данных»

**Практическое занятие №2**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИНБО-07-21, Веригин Никита Алексеевич* | (подпись) | |
| Преподаватель | *Царёв Роман Юрьевич* | (подпись) | |
| Отчет представлен | «\_\_\_»\_\_\_\_\_\_\_\_2024 г. | |  | |

Москва 2024 г.

Введение

Цели работы:

1. Найти и выгрузить многомерные данные (с большим количеством признаков – столбцов) с использованием библиотеки pandas. В отчёте описать найденные данные.
2. Вывести информацию о данных при помощи методов .info(), .head(). Проверить данные на наличие пустых значений. В случае их наличия удалить данные строки или интерполировать пропущенные значения. При необходимости дополнительно предобработать данные для дальнейшей работы с ними.
3. Построить столбчатую диаграмму (.bar) с использованием модуля graph\_objs из библиотеки Plotly со следующими параметрами:
   1. По оси Х указать дату или название, по оси У указать количественный показатель.
   2. Сделать так, чтобы столбец принимал цвет в зависимости от значения показателя (marker=dict(color=признак, coloraxis="coloraxis")).
   3. Сделать так, чтобы границы каждого столбца были выделены чёрной линией с толщиной равной 2.
   4. Отобразить заголовок диаграммы, разместив его по центру сверху, с 20 размером текста.
   5. Добавить подписи для осей X и Y с размером текста, равным 16. Для оси абсцисс развернуть метки так, чтобы они читались под углом, равным 315. 3.6. Размер текста меток осей сделать равным 14.
   6. Расположить график во всю ширину рабочей области и присвоить высоту, равную 700 пикселей.
   7. Добавить сетку на график, сделать её цвет 'ivory' и толщину равную 2. (Можно сделать это при настройке осей с помощью gridwidth=2, gridcolor='ivory')
   8. Убрать лишние отступы по краям.
4. Построить круговую диаграмму (go.Pie), использовав данные и стиль оформления из предыдущего графика. Сделать так, чтобы границы каждой доли были выделены чёрной линией с толщиной, равной 2 и категории круговой диаграммы были читаемы (к примеру, объединить часть объектов)
5. Построить линейные графики, взять один из параметров и определить зависимость между другими несколькими (от 2 до 5) показателями с использованием библиотеки matplotlib. Сделать вывод.
   1. Сделать график с линиями и маркерами, цвет линии 'crimson', цвет точек 'white', цвет границ точек 'black', толщина границ точек равна 2.
   2. Добавить сетку на график, сделать её цвет 'mistyrose' и толщину равную 2. (Можно сделать это при настройке осей с помощью linewidth=2, color='mistyrose').
6. Выполнить визуализацию многомерных данных, используя t-SNE. Необходимо использовать набор данных MNIST или fashion MNIST (можно использовать и другие готовые наборы данных, где можно наблюдать разделение объектов по кластерам). Рассмотреть результаты визуализации для разных значений перплексии.
7. Выполнить визуализацию многомерных данных, используя UMAP с различными параметрами n\_neighbors и min\_dist. Рассчитать время работы алгоритма с помощью библиотеки time и сравнить его с временем работы t-SNE.
8. На основе проделанной работы составить отчёт с описанием и скриншотами полученных результатов, сделать выводы о выбранных данных на основе полученных графиков, сравнить библиотеки. Начиная с 6 пункта отчёт дополнительно должен содержать результаты визуализации для разных значений параметров и выводы.

ХОД РАБОТЫ

Код программ и результаты их выполнения представлены на Рисунках 1-15.

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

**Рисунок 1 *—* Задача 1**

Найдены данные по бейсбольным играм Главной лиги, собранными за 130 лет.

**Изображение выглядит как текст, снимок экрана, программное обеспечение

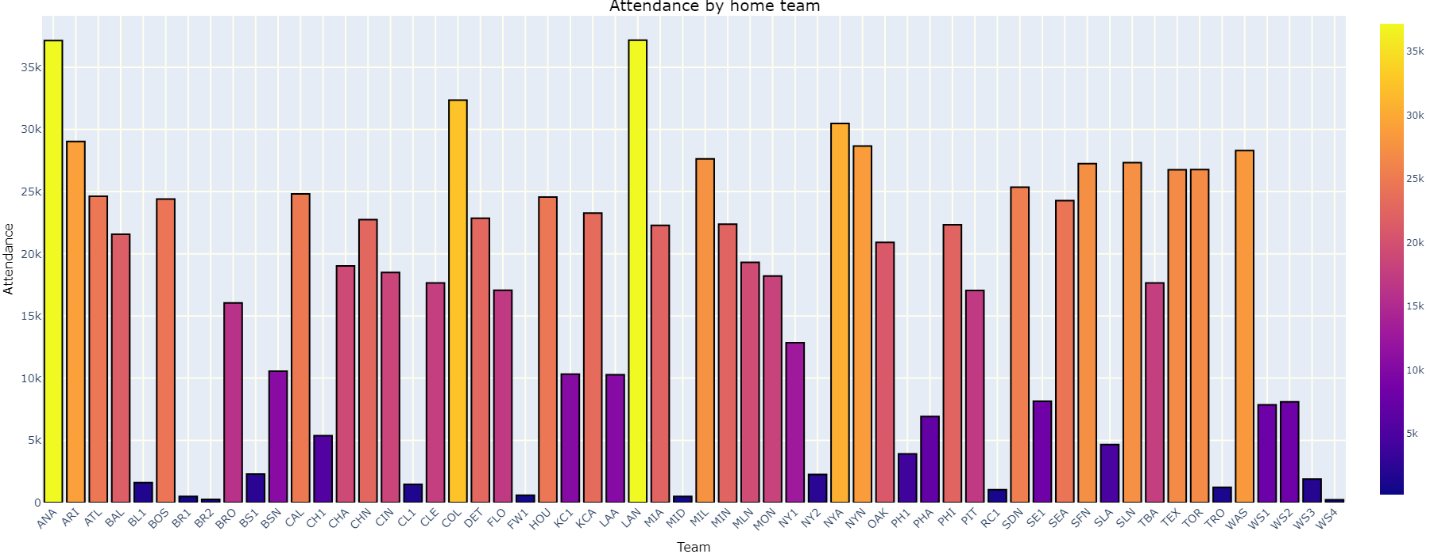
Автоматически созданное описание**

**Рисунок 2 *—* Задача 2**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 3 *—* Задача 3**

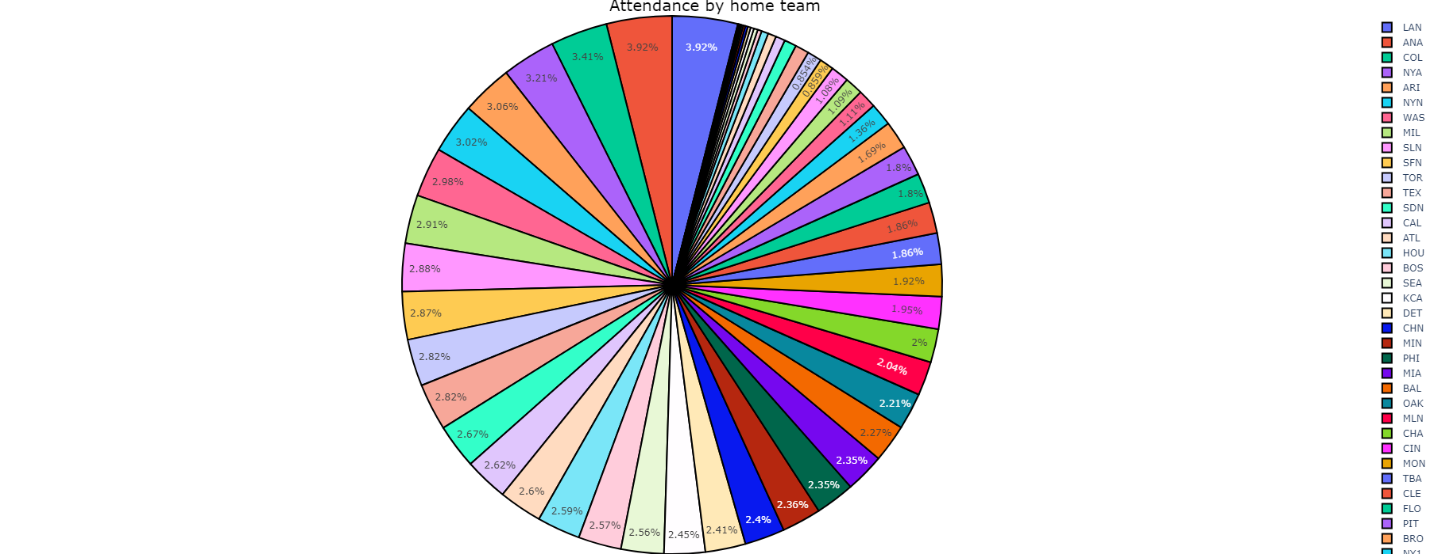
****

**Рисунок 4 *—* Столбчатая диаграмма**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 5 *—* Задача 4**

****

**Рисунок 6 *—* Круговая диаграмма**

**Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание**

**Рисунок 7 *—* Задача 5**

**Изображение выглядит как График, диаграмма, линия, текст

Автоматически созданное описание**

**Рисунок 8 *—* Линейные графики**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание**

**Рисунок 9 *—* Задача 6**

**Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание**

**Рисунок 10 *—* Задача 11**

**Изображение выглядит как текст, число, снимок экрана, Параллельный

Автоматически созданное описание**

**Рисунок 11 *—* Задача 12**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание**

**Рисунок 12 *—* Задача 7**

**Изображение выглядит как текст, снимок экрана, число, диаграмма

Автоматически созданное описание**

**Рисунок 13 *—* Задача 7**

**Изображение выглядит как текст, снимок экрана, число, График

Автоматически созданное описание**

**Рисунок 14 *—* Задача 7**

**Изображение выглядит как текст, снимок экрана, число, диаграмма

Автоматически созданное описание**

**Рисунок 15 — Задача 7**

ЗАКЛЮЧЕНИЕ

В ходе выполнения практической работы были изучены различные библиотеки визуализации данных (matplotlib, plotly, TSNE, UMAP) и особенности работы с ними в среде программирования Python.