МИНОБРНАУКИ РФ

ФГБОУ Тверской государственный технический университет

Факультет информационных технологий

Кафедра «Программное обеспечение»

Лабораторная работа №1

Дисциплина: «Анализ больших данных»

Вариант №7

Работу выполнил: студент группы

ПИН. РИС.21.06

Койич Марко.

Тверь, 2025

3. Описание данных из базы данных Формулы 1

В данной базе данных о гонках Формулы 1 содержится несколько таблиц, каждая из которых содержит различные аспекты данных о гонках, гонщиках, командах и результатах. Ниже приведено описание основных таблиц и их полей, а также информация о признаках, их типах данных и шкалах.

1. Таблицы и их поля

1.1. Таблица Races

* raceId: уникальный идентификатор гонки (целое число).
* year: год гонки (целое число).
* round: номер гонки в сезоне (целое число).
* circuitId: идентификатор трассы (целое число).
* date: дата гонки (дата).
* name: название гонки (строка).

1.2. Таблица Drivers

* driverId: уникальный идентификатор гонщика (целое число).
* driverRef: ссылка на гонщика (строка).
* number: номер гонщика (целое число).
* code: код гонщика (строка).
* name: имя гонщика (строка).
* nationality: национальность гонщика (строка).

1.3. Таблица Constructors

* constructorId: уникальный идентификатор команды (целое число).
* name: название команды (строка).
* nationality: национальность команды (строка).

1.4. Таблица Results

* resultId: уникальный идентификатор результата (целое число).
* raceId: идентификатор гонки (целое число).
* driverId: идентификатор гонщика (целое число).
* constructorId: идентификатор команды (целое число).
* grid: стартовая позиция (целое число).
* position: позиция в гонке (целое число).
* points: полученные очки (число с плавающей точкой).

1.5. Таблица Circuits

* circuitId: уникальный идентификатор трассы (целое число).
* name: название трассы (строка).
* location: местоположение трассы (строка).
* country: страна (строка).

2. Признаки и их типы данных

2.1. Признаки

Признаки — это поля, которые могут быть использованы для анализа и построения моделей. В данной базе данных можно выделить следующие признаки:

* Таблица Races:
  + year (количественный)
  + round (количественный)
  + date (категориальный, временной)
  + name (категориальный)
* Таблица Drivers:
  + number (количественный)
  + name (категориальный)
  + nationality (категориальный)
* Таблица Constructors:
  + name (категориальный)
  + nationality (категориальный)
* Таблица Results:
  + grid (количественный)
  + position (количественный)
  + points (количественный)
* Таблица Circuits:
  + name (категориальный)
  + location (категориальный)
  + country (категориальный)

2.2. Типы данных и шкалы

* Типы данных:
  + Целые числа (int): raceId, year, round, circuitId, driverId, constructorId, grid, position, number.
  + Числа с плавающей точкой (float): points.
  + Строки (str): name, nationality, location, country, driverRef, code.
  + Дата (date): date.
* Шкала:
  + Количественные признаки (например, year, grid, position, points) имеют интервальную шкалу, так как они представляют собой числовые значения, которые можно сравнивать и выполнять арифметические операции.
  + Категориальные признаки (например, name, nationality, location, country) имеют номинальную шкалу, так как они представляют собой категории без естественного порядка.