

Практическое задание 4

alrrr17
2022-12-17

Основы обработки данных с помощью R

Цель работы

1. Закрепить практические навыки использования языка программирования R для обработки данных
- 2.Закрепить знания основных функций обработки данных экосистемы tidyverse языка R
- 3.Развить прктические навыки использования функций обработки данных пакета dplyr - функции select(), filter(), mutate(), arrange(), group_by()

Задание

Проанализировать встроенные в пакет nycflights13 наборы данных с помощью языка R и ответить на вопросы:

Задание 1

1. Сколько встроенных в пакет nycflights13 датафреймов?

```
library(nycflights13)
library(tidyverse)

#nycflights13::airlines
#nycflights13::airports
#nycflights13::flights
#nycflights13::planes
#nycflights13::weather
5

## [1] 5
```

Задание 2

2. Сколько строк в каждом датафрейме?

```
airlines <- nycflights13::airlines
airports <- nycflights13::airports
flights <- nycflights13::flights
planes <- nycflights13::planes
weather <- nycflights13::weather
nrow(airlines)

## [1] 16

nrow(airports)

## [1] 1458

nrow(flights)

## [1] 336776

nrow(planes)

## [1] 3322

nrow(weather)

## [1] 26115
```

Задание 3

3. Сколько столбцов в каждом датафрейме?

```
length(airlines)

## [1] 2

length(airports)

## [1] 8

length(flights)

## [1] 19

length(planes)

## [1] 9

length(weather)

## [1] 15
```

Задание 4

4. Как посмотреть примерный вид датафрейма?

```
glimpse(airlines)

## Rows: 16
## Columns: 2
## $ carrier <chr> "9E", "AA", "AS", "B6", "DL", "EV", "F9", "FL", "HA", "MQ", "O...
## $ name <chr> "Endeavor Air Inc.", "American Airlines Inc.", "Alaska Airline...
```

Задание 5

5. Сколько компаний-перевозчиков (carrier) учитывают эти наборы данных (представлено в наборах данных)?

```
airlines %>% nrow()

## [1] 16
```

Задание 6

6. Сколько рейсов принял аэропорт John F Kennedy Intl в мае?

```
faa <- airports %>% filter(name == "John F Kennedy Intl") %>% select(faa) %>% paste(sep='')
flights %>% filter(month == 5, dest == faa) %>% nrow()

## [1] 0
```

Задание 7

7. Какой самый северный аэропорт?

```
airports %>% filter(lat == max(lat)) %>% select(name)

## # A tibble: 1 × 1
## name
## <chr>
## 1 Dillant Hopkins Airport
```

Задание 8

8. Какой аэропорт самый высокогорный (находится выше всех над уровнем моря)?

```
airports %>% filter(alt == max(alt)) %>% select(name)

## # A tibble: 1 × 1
## name
## <chr>
## 1 Telluride
```

Задание 9

9. Какие бортовые номера у самых старых самолетов?

```
planes %>% filter(year == min(year,na.rm = TRUE)) %>% select (tailnum)

## # A tibble: 1 × 1
## tailnum
## <chr>
## 1 N381AA
```

Задание 10

10. Какая средняя температура воздуха была в сентябре в аэропорту John F Kennedy Intl (в градусах Цельсия).

```
weather %>% filter(origin == "JFK",month == 9) %>% summarise(avg_temp = mean(5/9*(temp - 32), na.rm=TRUE))

## # A tibble: 1 × 1
## avg_temp
## <dbl>
## 1 19.4
```

Задание 11

11. Самолеты какой авиакомпании совершили больше всего вылетов в июне?

```
carr <- flights %>% filter(month == 6) %>%
  group_by(carrier) %>%
  summarise(n_flights=n()) %>%
  arrange(desc(n_flights)) %>%
  head(1) %>%
  select(carrier) %>% paste(sep='')
airlines %>% filter(carrier == carr)

## # A tibble: 1 × 2
## carrier name
## <chr> <chr>
## 1 UA United Air Lines Inc.
```

Задание 12

12. Самолеты какой авиакомпании задерживались чаще других в 2013 году?

```
carr <- flights %>% filter(dep_delay != 0 | arr_delay != 0) %>%
  group_by(carrier) %>%
  summarise(n_flights=n()) %>%
  arrange(desc(n_flights)) %>%
  head(1) %>%
  select(carrier) %>% paste(sep='')
airlines %>% filter(carrier == carr)

## # A tibble: 1 × 2
## carrier name
## <chr> <chr>
## 1 UA United Air Lines Inc.
```