Lab2.2

Alrrr1719

2023-05-22

Подключение библиотек

```
knitr::opts_chunk$set(echo = TRUE)
library(arrow)
```

```
## The tzdb package is not installed. Timezones will not be available to Arrow compute functions.
```

```
##
## Присоединяю пакет: 'arrow'
```

```
## Следующий объект скрыт от 'package:utils':
##
##     timestamp
```

```
library(dplyr)
```

```
##
## Присоединяю пакет: 'dplyr'
```

```
## Следующие объекты скрыты от 'package:stats':
##
##     filter, lag
```

```
## Следующие объекты скрыты от 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stringr)
library(lubridate)
```

```
##
## Присоединяю пакет: 'lubridate'
```

```
## Следующий объект скрыт от 'package:arrow':
##
##     duration
```

```
## Следующие объекты скрыты от 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
```

Загрузка данных

```
new_data <- arrow::read_csv_arrow("D:\\ProjectsRstudio\\2_sem_R\\LAB1\\gowiththeflow_20190826.csv",schema=schema
(timestamp=int64(),src=utf8(),dst=utf8(),port=int32(),bytes=int32()))
```

Решение задачи 2

Предварительная фильтрация

```
sel<-filter(new_data,    str_detect(src,"^((12|13|14)\\.)"),str_detect(dst,"^((12|13|14)\\.)",negate=TRUE))

sel$timestampp <- as.POSIXct(sel$timestamp/1000, origin = "1970-01-01")
sel$hour <- as.numeric(format(sel$timestampp, "%H"))
```

Кластерный анализ (фиксация 2-ух ключевых структур использования в рабочие и нерабочие часы

```
clust <- 2
kmeans_result <- kmeans(matrix(as.numeric(sel$hour), ncol = 1), centers = clust)
sel$cluster <- as.factor(kmeans_result$cluster)

centroid <- kmeans_result$centers
working_hours_cluster <- which.max(centroid)
working_hours_data <- sel %>%
  filter(cluster == working_hours_cluster)
start_time <- min(working_hours_data$hour)
end_time <- max(working_hours_data$hour)

filter(sel,(hour < start_time | hour > end_time)&str_detect(src,"13.37.84.125",negate=TRUE))%>%
select(src,bytes)%>%
group_by(src)%>%
summarise(bytes=sum(bytes))%>%
slice_max(bytes)%>%
select(src)
```

```
## # A tibble: 1 × 1
##   src
##   <chr>
## 1 13.48.72.30
```