Homework 1

Подтема

Автор

22 февраля 2022 г.

Содержание

```
library(tidyverse)
```

Домашняя работа 1

Для анализа был взят датасет с популярными для релокации городами. Предположим, что мы типичный IT-специалист, который подбирает место для будущей жизни и его интересуют только часть переменных из датасета

```
best_cities_for_a_workation <-
  read_csv("best cities for a workation.csv") %>% select(-"Ranking") %>% rename(
  remote_connection_speed = "Remote connection: Average WiFi speed (Mbps per second)",
  coffee_price = "Caffeine: Average price of buying a coffee",
  appartment_price = "Accommodation: Average price of 1 bedroom apartment per month",
  drinks_price = "After-work drinks: Average price for 2 beers in a bar",
  restaurant_price = "Food: Average cost of a meal at a local, mid-level restaurant",
  city = "City",
  country = "Country",
  coworking_space = "Co-working spaces: Number of co-working spaces",
  taxi_price = "Travel: Average price of taxi (per km)",
  sunshine_hours = "Climate: Average number of sunshine hours",
  tripadvisor_stats="Tourist attractions: Number of ‘Things to do’ on Tripadvisor",
  instagram_photos="Instagramability: Number of photos with #"
)
```

На основании имеющихся переменных подсчитаем сколько примерно можно потратить за вечер, проведённый в городе:

```
best_cities_for_a_workation <- best_cities_for_a_workation %>% mutate(
  average_evening_spends = taxi_price * 5 + drinks_price + restaurant_price
)
```

После обработки датасет выглядит так:

```
## # A tibble: 6 x 13
##   city      country  remote_connecti~ coworking_space coffee_price taxi_price
##   <chr>     <chr>               <dbl>           <dbl>        <dbl>      <dbl>
## 1 Bangkok   Thailand               28             117         1.56       0.82
## 2 New Delhi India                  12             165         1.42       0.19
## 3 Lisbon       Portugal            33           95       1.56      0.4
## 4 Barcelona    Spain               37          136       1.59      1.01
## 5 Buenos Aires Argenti~            17           67       1.22      0.47
## 6 Budapest     Hungary             37           40       1.2       0.72
## # ... with 7 more variables: drinks_price <dbl>, appartment_price <dbl>,
## #   restaurant_price <dbl>, sunshine_hours <dbl>, tripadvisor_stats <dbl>,
## #   instagram_photos <dbl>, average_evening_spends <dbl>
```

Сохраним полученный датасет в формат .rds

```
saveRDS(best_cities_for_a_workation, file="our_data.rds")
```

Теперь перейдём к разделению на группы, посмотрим на список стран, выберем из них несколько интересующих нас и выделим из датасета 5 стран с наибольшим количеством городов

```
top_five_countries <- best_cities_for_a_workation %>% group_by(country) %>% summarise(count=n()) %>% arrange
usa_cities <- best_cities_for_a_workation %>% filter(
  country =="United States"
) %>% select(-country)
print(usa_cities)
```

```
## # A tibble: 13 x 12
##    city     remote_connecti~ coworking_space coffee_price taxi_price drinks_price
##    <chr>               <dbl>           <dbl>        <dbl>      <dbl>        <dbl>
##  1 Los An~                58             105         3.39       1.21        10.1
##  2 Las Ve~                47              21         3.36       1.45         8.64
##  3 San Fr~                75              77         3.39       1.34        10.1
##  4 San Di~                74              53         3.06       1.34         8.6
##  5 Chicago                42             104         3.02       1.21         7.92
##  6 New Yo~                37             272         3.48       1.34        10.6
##  7 Houston                60              62         2.8        1.03         7.2
##  8 Miami                  40              59         3.25       1.16         8.64
##  9 Phoenix                44              35         3.32       1            7.18
## 10 New Or~                45              16         3.25       1.54         5.02
## 11 Washin~                68              59         3.3        1.87         8.6
## 12 Portla~                44              31         3.04       1.16         8.6
## 13 Boston                 33              41         3.2        1.34        10.1
## # ... with 6 more variables: appartment_price <dbl>, restaurant_price <dbl>,
## #   sunshine_hours <dbl>, tripadvisor_stats <dbl>, instagram_photos <dbl>,
## #   average_evening_spends <dbl>
```

```
germany_cities <- best_cities_for_a_workation %>% filter(
  country == "Germany"
) %>% select(-country)
print(germany_cities)
```

```
## # A tibble: 9 x 12
##   city      remote_connecti~ coworking_space coffee_price taxi_price drinks_price
##   <chr>                <dbl>           <dbl>        <dbl>      <dbl>        <dbl>
## 1 Berlin                  33             127         2.49       1.71         4.98
## 2 Hamburg                 41              65         2.46       1.63         7.26
## 3 Munich                  31              87         2.7        1.71         6.84
## 4 Cologne                 33              38         2.34       1.71         6.84
## 5 Dusseld~                25              44         2.59       1.88         6.84
## 6 Frankfu~                22              66         2.49       1.71         6.84
## 7 Stuttga~                33              25         2.51       1.45         6.82
## 8 Hannover                38               7         2.26       1.71         5.98
## 9 Dresden                 35               5         2.04       1.87         5.94
## # ... with 6 more variables: appartment_price <dbl>, restaurant_price <dbl>,
## #   sunshine_hours <dbl>, tripadvisor_stats <dbl>, instagram_photos <dbl>,
## #   average_evening_spends <dbl>
```

```
canada_cities <- best_cities_for_a_workation %>% filter(
  country == "Canada"
) %>% select(-country)
print(canada_cities)
```

```
## # A tibble: 6 x 12
##   city      remote_connecti~ coworking_space coffee_price taxi_price drinks_price
##   <chr>                <dbl>           <dbl>        <dbl>      <dbl>        <dbl>
## 1 Montreal                27              60         2.37       1.01         6.94
## 2 Toronto                 26             113         2.63       1.15         8.06
## 3 Vancouv~                40              43         2.6        1.08         8.06
## 4 Calgary                 24              34         2.44       1.16         8.1
## 5 Edmonton                30              10         2.69       1.04         6.94
## 6 Ottawa                  26              24         2.59       1.15         8.06
## # ... with 6 more variables: appartment_price <dbl>, restaurant_price <dbl>,
## #   sunshine_hours <dbl>, tripadvisor_stats <dbl>, instagram_photos <dbl>,
## #   average_evening_spends <dbl>
```

```
spain_cities <- best_cities_for_a_workation %>% filter(
  country == "Spain"
) %>% select(-country)
print(spain_cities)
```

```
## # A tibble: 6 x 12
##   city      remote_connecti~ coworking_space coffee_price taxi_price drinks_price
##   <chr>                <dbl>           <dbl>        <dbl>      <dbl>        <dbl>
## 1 Barcelo~                37             136         1.59       1.01         5.12
## 2 Madrid                  32             125         1.7        0.94        10.0
## 3 Valencia                30              39         1.51       0.85         5.1
## 4 Malaga                  26              17         1.27       0.73         4.26
## 5 Seville                 28               7         1.21       0.8          3.4
## 6 Palma d~                29              15         1.81       0.83         4.26
## # ... with 6 more variables: appartment_price <dbl>, restaurant_price <dbl>,
## #   sunshine_hours <dbl>, tripadvisor_stats <dbl>, instagram_photos <dbl>,
## #   average_evening_spends <dbl>
```

```
uk_cities <- best_cities_for_a_workation %>% filter(
  country == "United Kingdom"
) %>% select(-country)
print(uk_cities)
```

```
## # A tibble: 6 x 12
##   city      remote_connecti~ coworking_space coffee_price taxi_price drinks_price
##   <chr>                <dbl>           <dbl>        <dbl>      <dbl>        <dbl>
## 1 Liverpo~                26              17         2.66       0.93          3.5
## 2 London                  22             318         2.95       1.7          10
## 3 Manches~                33              38         2.81       1.22          8
## 4 Edinbur~                26              23         2.76       1.42          8.5
## 5 Glasgow                 26              16         2.77       1.06          7
## 6 Belfast                 26              13         2.74       1.07          9
## # ... with 6 more variables: appartment_price <dbl>, restaurant_price <dbl>,
## #   sunshine_hours <dbl>, tripadvisor_stats <dbl>, instagram_photos <dbl>,
## #   average_evening_spends <dbl>
```

Затем посчитаем основные описательные статистики для каждой из групп

```
usa_cities %>% select(average_evening_spends) %>% summarise_all(list(mean, median, sd, min, max)) %>% rename("M
```

```
## # A tibble: 1 x 5
##    Mean Median `Standard Deviation`   Min   Max
##   <dbl>  <dbl>                <dbl> <dbl> <dbl>
## 1  27.3   26.2                 3.01  23.0  31.7
```