RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

sales_analysis_Eigen.R | Data Acquisition.R | Data Acquisition_x.R | Data_Acquisation_Eigen.R | Data_wrangling_Eigen.R

Source on Save | Run | Source

```
assig_patassig_combined_tbl = patent_assignee_tbl %>%
  right_join(assignee_tbl, by = "assignee_id")



# Question 1 ----
assig_patassig_combined_tbl %>%
  group_by(organization) %>%
  summarise(
    count = n(),
    ) %>%
  ungroup() %>%
  arrange(desc(count))



# # A tibble: 464,605 x 2
# organization                                  count
# <chr>                                         <int>
#   1 International Business Machines Corporation 139092
# 2 NA                                          123331
# 3 Samsung Electronics Co., Ltd.                93562
# 4 Canon Kabushiki Kaisha                       75910
# 5 Sony Corporation                             54343
# 6 Kabushiki Kaisha Toshiba                     49443
# 7 General Electric Company                     47122
# 8 Hitachi, Ltd.                                45375
# 9 Intel Corporation                            42157
# 10 Fujitsu Limited                              37197
# # ... with 464,595 more rows



# Question 2 ----
setnames(patent_tbl, "id", "patent_id")
```

161:1 | Question 1 | R Script

Console | Terminal | Jobs

C:/Users/Stuwe/Desktop/GitHubClone/ws20-business-data-science-basics---lab-journal-JulianStu95/Business Data Science Basics/Exercise/DS_101/

```
>
> col_types <- list(
+     id = col_character(),
+     type = col_character(),
+     number = col_character(),
+     country = col_character(),
+     date = col_date("%Y-%m-%d"),
+     abstract = col_skip(),
+     title = col_skip(),
+     kind = col_character(),
+     num_claims = col_skip(),
+     filename = col_skip(),
+     withdrawn = col_skip()
+ )
>
> patent_tbl <- vroom(
+     file       = "patent.tsv",
+     delim      = "\t",
+     col_types  = col_types,
+     na         = c("", "NA", "NULL")
+ )
indexed 2.15GB in  1m, 73.65MB/s
```

Environment | History | Connections | Tutorial

Import Dataset | List

Global Environment

col_types	List of 11
con	Formal class SQLiteConnection
data_list	List of 5
html	List of 2
html_home	List of 2
html_rose	List of 2
html_ttx	List of 2
imdb_tbl	250 obs. of 6 variables
orderlines_tbl	15644 obs. of 7 variables
resp	List of 10
response	List of 10
rose_tbl	9 obs. of 2 variables
sales_by_state_tbl	12 obs. of 3 variables
sales_by_year_state_1_tbl	60 obs. of 4 variables
sp_500	505 obs. of 9 variables
test	32 obs. of 11 variables
tt_tbl	9 obs. of 2 variables
ttx_tbl	5 obs. of 2 variables
values	
alphavantage_api_url	"https://www.alphavantage.co/query"

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

C: > Users > Stuwe > Desktop > GitHubClone > ws20-business-data-science-basics---lab-journal-JulianStu95 > Business Data Science Basics > Exercise > DS_101

Name	Size	Modified
..		
.RData	3.3 KB	Nov 24, 2020, 2:00 AM
.Renviron	41 B	Nov 25, 2020, 3:36 PM
.Rhistory	12.5 KB	Nov 24, 2020, 12:03 PM
00_cheatsheet		
00_data		
00_scripts		
01_getting_started		
02_data_wrangling		
03_visualization		
04_modeling		
05_communicate		
assignee.tsv	62.7 MB	Aug 24, 2020, 9:46 PM
bike_data_cleaned_tbl.rds	1.9 KB	Nov 25, 2020, 10:14 PM
bike_data_colors_tbl.rds	2 KB	Nov 25, 2020, 10:14 PM
bike_data_tbl.rds	2.2 KB	Nov 25, 2020, 10:14 PM
Challenge_2.1.pdf	7.8 KB	Dec 24, 2020, 3:57 PM
Challenge_2.2.pdf	13.8 KB	Nov 25, 2020, 4:10 AM
DS_business_case.Rproj	218 B	Nov 24, 2020, 12:02 PM
inventor_gender.tsv	90.3 MB	Aug 24, 2020, 9:19 PM
patent.tsv	5.5 GB	Aug 26, 2020, 5:07 PM
patent_assignee.tsv	492.4 MB	Aug 26, 2020, 5:24 PM
Patents_DB_dictionary_bulk_downloads.xlsx	33.8 KB	Nov 27, 2020, 1:50 AM
Pfad Für Challenges.txt	1.1 KB	Dec 5, 2020, 3:56 PM
Revenue by year and main category.pdf	5.9 KB	Nov 25, 2020, 2:18 AM
Revenue by year.pdf	6.3 KB	Nov 25, 2020, 2:18 AM
Rplot.pdf	8.3 KB	Nov 25, 2020, 4:09 AM
test_df.csv	172.6 MB	Nov 27, 2020, 1:24 PM
uspc.tsv	963.4 MB	Nov 27, 2020, 1:55 AM
Worldmap_covid_Julian_Stuwe.pdf	444 KB	Dec 1, 2020, 1:31 PM

DE 16:51 24.12.2020