Project_test

B203349

2022-12-11

Contents

```
library(sparklyr)
```

```
##
## Attaching package: 'sparklyr'

## The following object is masked from 'package:stats':
##
##     filter
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(cowplot)
library(knitr)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(tidyverse)
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v tibble  3.1.8     v purrr   0.3.4
## v tidyr   1.2.0     v stringr 1.4.0
```

```
## v readr     2.1.2      v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()         masks sparklyr::filter(), stats::filter()
## x kableExtra::group_rows() masks dplyr::group_rows()
## x purrr::invoke()         masks sparklyr::invoke()
## x dplyr::lag()            masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following object is masked from 'package:cowplot':
##
##     stamp
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
library(dbplot)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(broom)
library(formatR)
```

R Markdown

```
sc = spark_connect(master = 'local')
```

```
# diabetic_data =
# spark_read_csv(sc,
# '/Users/matt/Desktop/Dropbox/Home/College/Edinburgh
# - MSc Data Science/Big
# Data
# Analytics/diabetes_readmissions/RawData/diabetic_data.csv')
diabetic_data <- spark_read_csv(sc,
    "/home/jovyan/Matt/diabetes_readmissions/RawData/diabetic_data.csv")
```

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
# total number of patient encounters
total_number_of_patient_encounters <- pull(diabetic_data,
    patient_nbr) %>%
    length()
```

Table 1: Summary of patient encounters

	n
Total number of patient encounters:	101,766
Number of patient encounters that can be classified as 'repeat':	47,021
Number of patients with repeat encounters:	16,773

```r
# number of patients with repeat encounters
number_of_patient_with_repeat_encounters <- diabetic_data %>%
    group_by(patient_nbr) %>%
    filter(n() > 1) %>%
    tally() %>%
    sdf_nrow()

# number of patient encounters that can be
# classified as 'repeat'
number_of_patient_classed_as_repeat <- diabetic_data %>%
    group_by(patient_nbr) %>%
    filter(n() > 1) %>%
    sdf_nrow()

patient_encounters_table <- data.frame(n = c(total_number_of_patient_encounters,
    number_of_patient_classed_as_repeat, number_of_patient_with_repeat_encounters))

row.names(patient_encounters_table) <- c("Total number of patient encounters:",
    "Number of patient encounters that can be classified as 'repeat':",
    "Number of patients with repeat encounters:")

kable(patient_encounters_table, caption = "Summary of patient encounters",
    digits = 3, format.args = list(big.mark = ",",
        scientific = FALSE))
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.