Digital health

Report about the effect of smoking & drinking on diabetes and its INTERACTIVE WEBSITE

Use case:

I picked up this data set from Kaggle

The problem I found shows the relationship between both drinking alcohol and smoking to diabetes. This data highlights the relation between several categories within diabetes and smoking these include; (diabetic, pre diabetic), (regular smoker, and heavy smoker) respectively. The CDC shows that smoking is an important risk factor for developing diabetes. Whilst this data does not show categories within alcohol drinking it confirms the relationship between diabetes and drinking. It shows clearly the negative effect of alcohol drinking on diabetes.

As we know diabetic patients have secondary complication especially with drinking as both decrease immunity leaving the patient suspectable to infection. This dataset with its visualization will hopefully shine a light on these relationships to our stakeholders in public health and the WHO to conduct mass media intervention. Most importantly to encourage diabetic patients even those who are pre-diabetic to stop drinking alcohol and smoking.

The solution for this problem:

Sharing the visualization of this data to doctors, nonprofit organizations and the diabetes society to encourage people especially in the pre diabetic stage to stop smoking and drinking. Forming an online free registry for all current diabetic patients, this will help both healthcare professionals and patients with sharing information and easy access. This can also include people who are prediabetic. The main goal is to raise awareness, it can be by make an application for all information about the dangers, and risks of both smoking & drinking including means of cessation and available remedies like skin patches, gum and Champix tablets all of which proved to be effective in smoking cessation. It should include also access to helping communities such as alcohol addiction support groups. Pamphlets with available information and help call centers could be distributed along with medication and regular check ups at the health centers.

Also implement rules to ban smoking in public spaces and indoor where available which prevents second hand smoking. The ban for alcohol might include age

restrictions as well as space and time restrictions, for example people ages 21 and above can socially drink during evening only in dining places not cafes.

Making it harder to smoke by increasing tax on tobacco and alcohol beverages for drinking alcohol increase awareness in teenager and pre diabetic and diabetic.

FFPG	smoking	drinking
0.35926795540	0.244429843	0.37096286
-0.04928844897	0.261565582	0.15859890
0.24789299933	0.036824315	0.10563170
0.22121522241	0.139363947	0.18086596
0.16787388825	0.037214604	0.08491931
0.47393071783	0.278684848	0.34800780
0.10613063750	0.038657866	0.07508292
0.16593453650	-0.004733228	0.05588893
0.24448044530	0.262487122	0.29562698
0.24259652650	0.176180202	0.25400099
0.10075958310	-0.062541082	-0.01856681
0.11306505080	0.035583902	0.08386879
0.00024567150	-0.170457348	-0.13023054
1.00000000000	0.302942544	0.41014761
0.30294254360	1.000000000	0.76680424
0.41014760570	0.766804236	1.00000000
0.03660269190	-0.020773710	-0.04401282
0.58310693560	0.569857074	0.70658218

Implementation:

I started with getting the dataset from the internet and I imported the dataset in the R file program. I proceeded with the cleaning process and uploading the essential package (dplyr, shiny, vcd)

First, I did the matrix correlation to find which variable of the 18 has a relation with diabetes column, any variable with a 0.5 and more has a strong relation to diabetes.

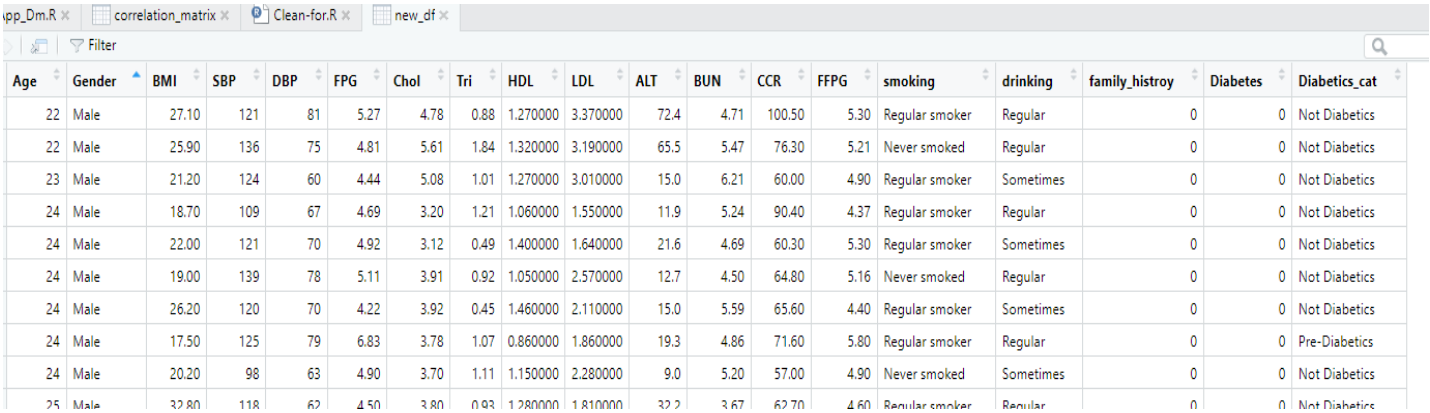
The following columns had a strong relation with diabetes; smoking, drinking, fasting plasma glucose (FPG) and final fasting plasma glucose (FFPG).

Then I changed some important columns using function (level); first I switched the gender from integer to character 1 & 2 to male & female respectively. Second, I changed the smoking variable from numeric 1,2,3 and 4 to never smoke, occasional smoker, regular smoker and heavy smoker. Finally, I changed the drinking variable from numeric to character 1,2,3 and 4 to never, some, regular and heavy.

I then created a new column named Diabetic categories, including non-diabetic, prediabetes and diabetic. This classification was based and mainly depends on

fasting plasma glucose now as the new-df data and it became 19 columns using the function MUTATE.

	Age	Gender	BMI	SBP	DBP	FPG	Chol	Tri	HDL	LDL	ALT	BUN	CCR	FFPG	smoking	drinking	family_histroy	Diabetes
1	26	1	20.10	119	81	5.80	4.36	0.86	0.90	2.43	12.0	5.40	63.8	5.40	3	3	0	0
2	40	1	17.70	97	54	4.60	3.70	1.02	1.50	2.04	9.2	3.70	70.3	4.10	1	1	0	0
3	40	2	19.70	85	53	5.30	5.87	1.29	1.75	3.37	10.1	4.10	61.1	4.85	3	3	0	0
4	43	1	23.10	111	71	4.50	4.05	0.74	1.27	2.60	36.5	4.38	73.4	5.30	2	3	0	0
5	36	1	26.50	130	82	5.54	6.69	3.49	0.91	3.64	69.3	3.86	67.5	5.53	3	3	0	0



pp_Dm.R | correlation_matrix | Clean-for.R | new_df

Filter

Age	Gender	BMI	SBP	DBP	FPG	Chol	Tri	HDL	LDL	ALT	BUN	CCR	FFPG	smoking	drinking	family_histroy	Diabetes	Diabetics_cat
22	Male	27.10	121	81	5.27	4.78	0.88	1.270000	3.370000	72.4	4.71	100.50	5.30	Regular smoker	Regular	0	0	Not Diabetics
22	Male	25.90	136	75	4.81	5.61	1.84	1.320000	3.190000	65.5	5.47	76.30	5.21	Never smoked	Regular	0	0	Not Diabetics
23	Male	21.20	124	60	4.44	5.08	1.01	1.270000	3.010000	15.0	6.21	60.00	4.90	Regular smoker	Sometimes	0	0	Not Diabetics
24	Male	18.70	109	67	4.69	3.20	1.21	1.060000	1.550000	11.9	5.24	90.40	4.37	Regular smoker	Regular	0	0	Not Diabetics
24	Male	22.00	121	70	4.92	3.12	0.49	1.400000	1.640000	21.6	4.69	60.30	5.30	Regular smoker	Sometimes	0	0	Not Diabetics
24	Male	19.00	139	78	5.11	3.91	0.92	1.050000	2.570000	12.7	4.50	64.80	5.16	Never smoked	Regular	0	0	Not Diabetics
24	Male	26.20	120	70	4.22	3.92	0.45	1.460000	2.110000	15.0	5.59	65.60	4.40	Regular smoker	Sometimes	0	0	Not Diabetics
24	Male	17.50	125	79	6.83	3.78	1.07	0.860000	1.860000	19.3	4.86	71.60	5.80	Regular smoker	Regular	0	0	Pre-Diabetics
24	Male	20.20	98	63	4.90	3.70	1.11	1.150000	2.280000	9.0	5.20	57.00	4.90	Never smoked	Sometimes	0	0	Not Diabetics
25	Male	32.80	118	62	4.50	3.80	0.93	1.280000	1.810000	32.2	3.67	62.70	4.60	Regular smoker	Regular	0	0	Not Diabetics

```
##############################  FIRST : IMPORTING DATA #################
# 1- Import Data:
data <- read.csv("diabetes.csv")
# 2- Select Important columns only
#########################################################################
# 3- Explore Data:
str(data)      # Structure of the data
summary(data)  # Summary statistics
head(data)     # Display the first few rows
## Correlation Matrix: ######################################################
correlation_matrix <- cor(data, use = "complete.obs")
############## Add New column: Diabetics Category ###########
library(dplyr)
new_df <-data %>%
  mutate(Diabetics_cat = case_when(FFPG<5.6 ~ "Not Diabetics",
                                   FFPG >= 5.6 & FFPG < 7 ~ "Pre-Diabetics",
                                   FFPG >= 7 ~ "Diabetic",
                                   TRUE ~ NA_character_))
## Change the type :
new_df$Diabetics_cat <- as.factor(new_df$Diabetics_cat)
new_df$smoking <- factor(new_df$smoking, levels =c(1,2,3,4.860753), labels = c("Never smoked", "Occasional smoker", "Regular smoker", "Heavy smoker"))
new_df$drinking <- factor(new_df$drinking, levels =c(1,2,3,4.860753), labels = c("Never", "Sometimes", "Regular", "Heavy"))
new_df$Gender <- factor(new_df$Gender, levels =c(1,2), labels = c("Male", "Female"))
```

Then I moved to the next step plotting this correlation in App Dm.R, loading shiny and vcd.

I used mosaic plot for exploring the association between two or more categorical variables, they display the relative frequencies of the categories as areas within rectangles. Then I found that there is a strong correlation between both diabetics and prediabetics who are heavy smokers as well as those who are heavy drinkers.

- Finally, I submit the App-Dm.R to shiny app account to visualize in webpage http://medawi.shinyapps.io/M-medawi

```
library(shiny)
library(vcd)

# Define UI for application
ui <- fluidPage(

  titlePanel("Effect of Smoking and Drinking on. Diabetics"),
  fluidRow(
    column(10, plotOutput("plot1")),
    column(10, plotOutput("plot2")),
    column(10, plotOutput("plot3"))
  )
)

# Define server logic
server <- function(input, output) {


  # Load the data from the R file
  source("Clean-for.R")
  # Sample data for demonstration

  # Mosaic plot 1:  Smoker vs. Diabetics
  output$plot1 <- renderPlot({
    mosaicplot(table(new_df$smoking, new_df$Diabetics_cat), main = "Smoker  vs.  Diabetics",color=c("blue", "green"))
  })

  # Mosaic plot 2: Drinking vs. Diabetics
  output$plot2 <- renderPlot({
    mosaicplot(table(new_df$drinking,new_df$Diabetics_cat), main = "Drinking  vs. Diabetics",color=c("blue", "green"))
  })

  # Mosaic plot 3: Gender vs. Diabetics
  output$plot3 <- renderPlot({
    mosaicplot(table(new_df$Gender,new_df$Diabetics_cat), main = "Gender  vs.  Diabetics ", color=c("blue", "green"))
  })
}

# Run the application
shinyApp(ui = ui, server = server)
```

Drinking vs. Diabetics