

```

1 packages <- c("plyr","dplyr","ggplot2")
2 lapply(packages, library, character.only = TRUE)
3 #load packages
4
5 df <- read.csv("D:/DATAMINING/ab_data.csv")
6 str(df)
7 #load dataframe and examine
8
9 colSums(is.na(df))
10 #check for NA
11
12 unique_id <- unique(df$user_id)
13 length(unique_id)
14 df <- df[!duplicated(df$user_id), ]
15 #check and remove duplicate user_id rows
16
17 summary(df$group)
18 summary(df$landing_page)
19 #explore columns
20
21 freqgrouplanding_page <- ddply(df,.(df$group, df$landing_page), nrow)
22 names(freqgrouplanding_page) <- c("group", "landing_page", "Freq")
23 freqgrouplanding_page
24 #check conditions
25
26 dfclean1 <- dplyr::filter(df, group == "control" & landing_page == "old_page")
27 dfclean2 <- dplyr::filter(df, group == "treatment" & landing_page == "new_page")
28 df <- rbind(dfclean1, dfclean2)
29 #clean dataframe, remove rows with control+new_page and treatment+old_page
30
31 dfgrouplanding_page <- ddply(df,.(df$group, df$landing_page), nrow)
32 names(dfgrouplanding_page) <- c("group", "landing_page", "Freq")
33 dfgrouplanding_page
34 #chek df yang dibersihkan
35
36 p <- ggplot(dfgrouplanding_page, aes(x = landing_page, y = Freq))+
37   geom_col(aes(fill = group), width = 0.7)
38 p
39 #plot group, landing_page, frequency
40
41 groupconvertfreq <- ddply(df,.(df$group, df$converted), nrow)
42 names(groupconvertfreq) <- c("group", "converted", "Freq")
43 groupconvertfreq
44
45 x <- matrix(c(126877, 127180, 17349, 17134), nrow=2)
46 chisq.test(x)
47 #Uji independensi chi-square. Karena nilai p kami lebih besar dari 0,05, kami gagal menolak
48 ##Itu menunjukkan tidak ada perbedaan konversi antara orang-orang yang melihat old_page dan
49
50
51
52 df$Date <- as.Date(df$timestamp)
53 df$Time <- format(as.POSIXct(df$timestamp),format="%H:%M:%S")
54 df$weekday <- weekdays(df$Date)
55 df$weekday <- ordered(df$weekday, levels=c("Monday",
56                                             "Tuesday",
57                                             "Wednesday",
58                                             "Thursday",
59                                             "Friday",
60                                             "Saturday",
61                                             "Sunday"))
62 barplot(table(df$weekday))

```