

DS_Project

2023-03-20

```
library(tidyverse)
library(modelr)
library(lubridate)
library(dplyr)

economy_data = read_csv("Flights Price Prediction Dataset\\economy.csv", show_col_types = FALSE)
business_data = read_csv("Flights Price Prediction Dataset\\business.csv", show_col_types = FALSE)
economy_data$time_taken = lubridate::hm(economy_data$time_taken)
business_data$time_taken = lubridate::hm(business_data$time_taken)
head(economy_data)

## # A tibble: 6 x 11
##   date      airline ch_code num_c~1 dep_t~2 from    time_ta~3 stop    arr_t~4 to
##   <chr>     <chr>   <chr>     <dbl> <time>   <chr> <Period> <chr> <time>   <chr>
## 1 11-02-2022 SpiceJ~ SG        8709 18:55  Delhi 2H 10M OS non-~ 21:05  Mumb~
## 2 11-02-2022 SpiceJ~ SG        8157 06:20  Delhi 2H 20M OS non-~ 08:40  Mumb~
## 3 11-02-2022 AirAsia I5       764  04:25  Delhi 2H 10M OS non-~ 06:35  Mumb~
## 4 11-02-2022 Vistara UK       995  10:20  Delhi 2H 15M OS non-~ 12:35  Mumb~
## 5 11-02-2022 Vistara UK       963  08:50  Delhi 2H 20M OS non-~ 11:10  Mumb~
## 6 11-02-2022 Vistara UK       945  11:40  Delhi 2H 20M OS non-~ 14:00  Mumb~
## # ... with 1 more variable: price <dbl>, and abbreviated variable names
## #   1: num_code, 2: dep_time, 3: time_taken, 4: arr_time

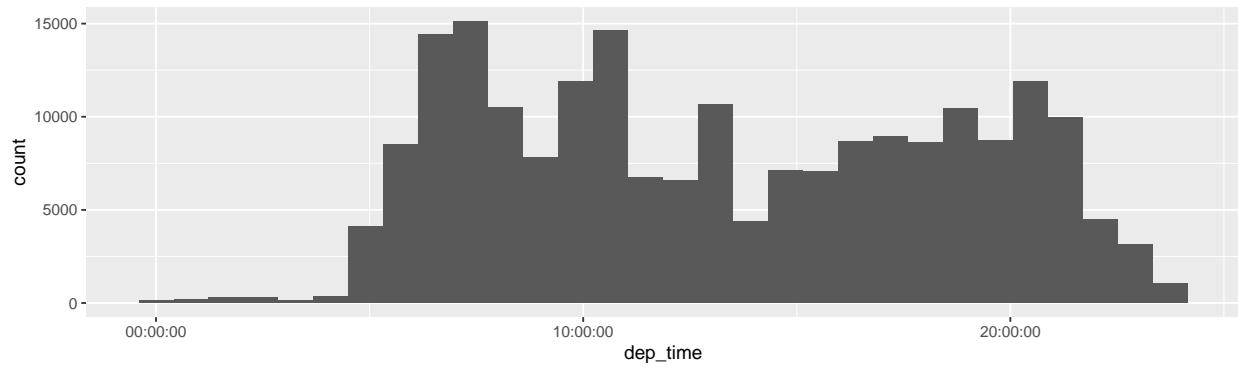
head(business_data)

## # A tibble: 6 x 11
##   date      airline ch_code num_c~1 dep_t~2 from    time_taken stop    arr_t~3 to
##   <chr>     <chr>   <chr>     <dbl> <time>   <chr> <Period> <chr> <time>   <chr>
## 1 11-02-20~ Air In~ AI        868  18:00  Delhi 2H 0M OS "non-~ 20:00  Mumb~
## 2 11-02-20~ Air In~ AI        624  19:00  Delhi 2H 15M OS "non-~ 21:15  Mumb~
## 3 11-02-20~ Air In~ AI        531  20:00  Delhi 24H 45M OS "1-s~ 20:45  Mumb~
## 4 11-02-20~ Air In~ AI        839  21:25  Delhi 26H 30M OS "1-s~ 23:55  Mumb~
## 5 11-02-20~ Air In~ AI        544  17:15  Delhi 6H 40M OS "1-s~ 23:55  Mumb~
## 6 11-02-20~ Vistara UK       985  19:50  Delhi 2H 10M OS "non-~ 22:00  Mumb~
## # ... with 1 more variable: price <dbl>, and abbreviated variable names
## #   1: num_code, 2: dep_time, 3: arr_time

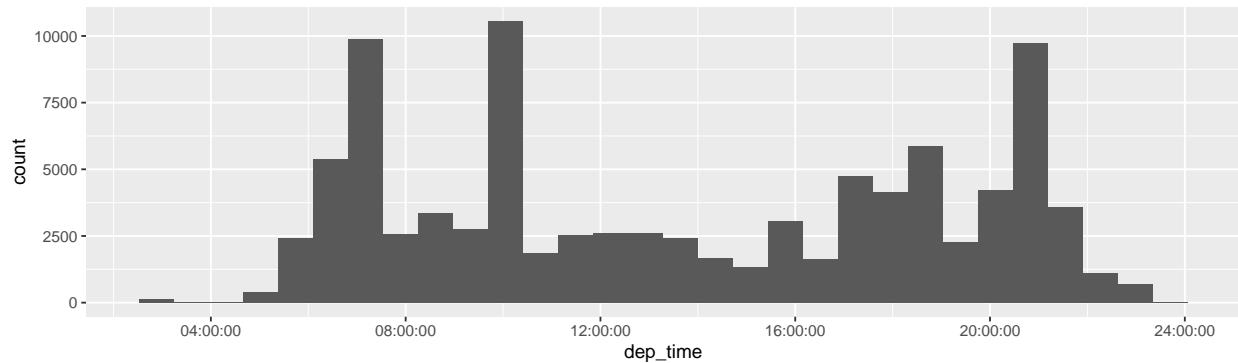
## distribution with respect to departure time

#distribution with respect to departure time

ggplot (data = economy_data) +
  geom_histogram(mapping = aes(x=dep_time))
```

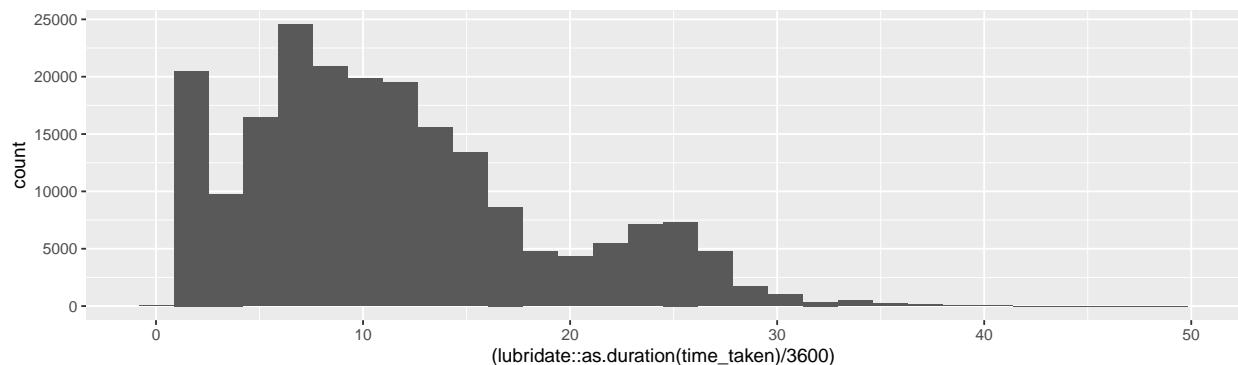


```
ggplot (data = business_data) +
  geom_histogram(mapping = aes(x=dep_time))
```

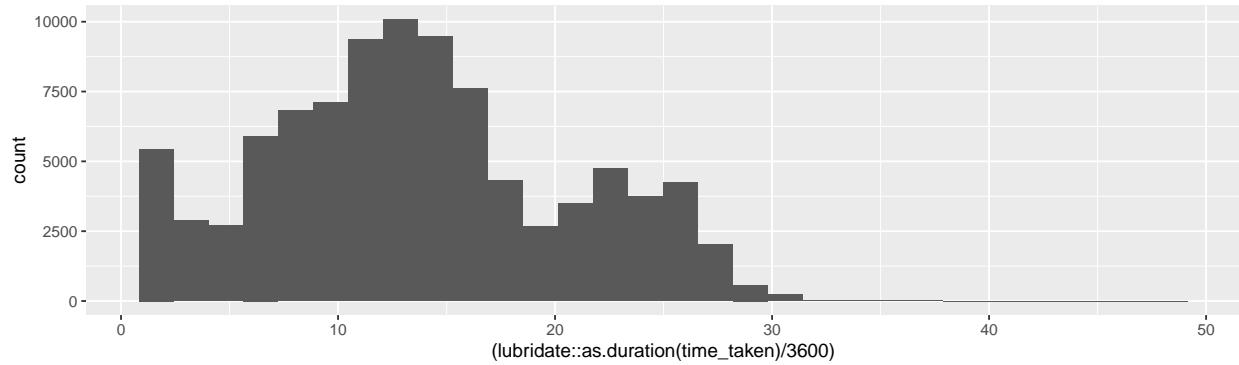


```
## distribution with respect to duration
```

```
ggplot (data = economy_data) +
  geom_histogram(mapping = aes(x=(lubridate::as.duration(time_taken)/3600)))
```



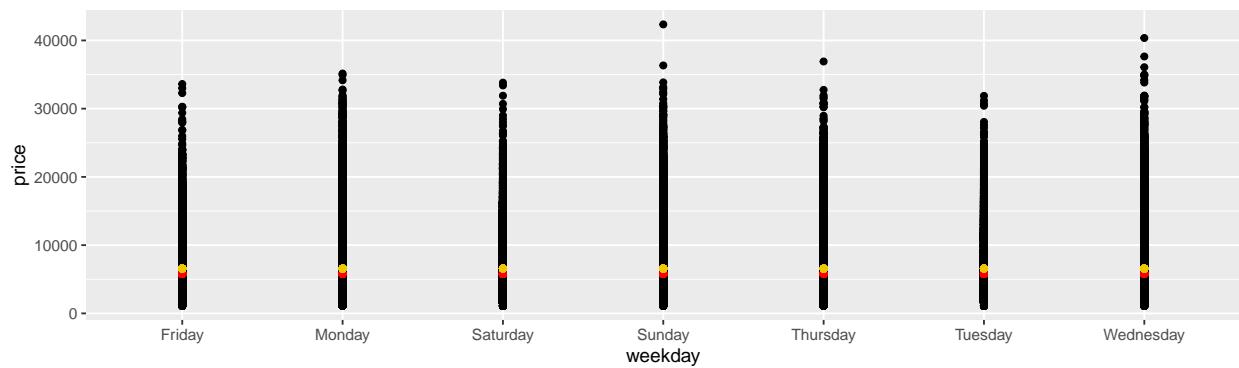
```
ggplot (data = business_data) +
  geom_histogram(mapping = aes(x=(lubridate::as.duration(time_taken)/3600)))
```



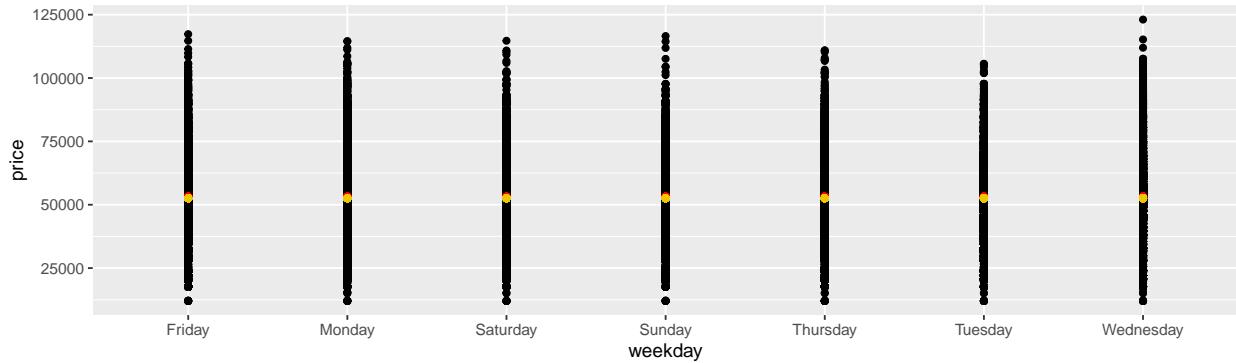
##Price distribution based on the day of the week

```
#Converting date to weekday
economy_data$date <- as.Date(economy_data$date)
economy_data$weekday <- strftime(economy_data$date, "%A")
business_data$date <- as.Date(business_data$date)
business_data$weekday <- strftime(business_data$date, "%A")

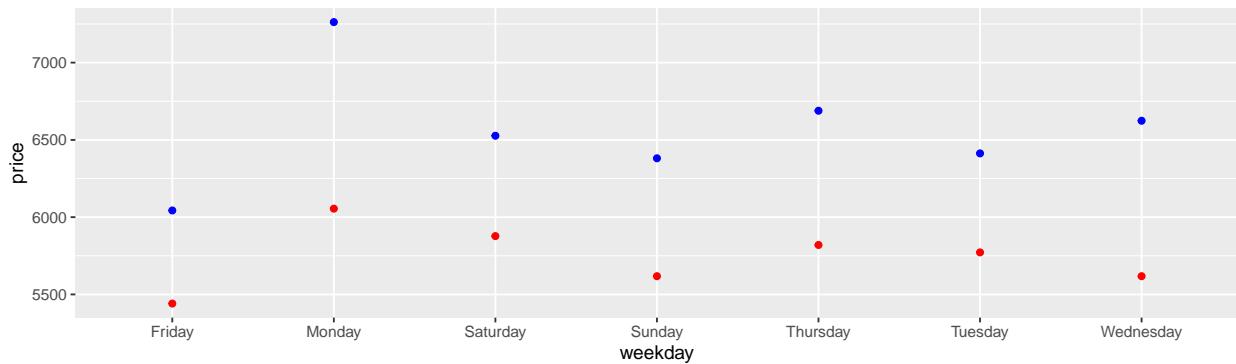
#Plotting price vs day of economy
weekday_graph_economy <- ggplot (data = economy_data) +
  geom_point(mapping = aes (x = weekday, y = price))+
  geom_point(mapping = aes (x = weekday, y = median(price)),color="red")+
  geom_point(mapping = aes (x = weekday, y = mean(price)),color="gold2")
(weekday_graph_economy)
```



```
#Plotting price vs day of business
weekday_graph_business <- ggplot (data = business_data) +
  geom_point(mapping = aes (x = weekday, y = price))+
  geom_point(mapping = aes (x = weekday, y = median(price)),color="red")+
  geom_point(mapping = aes (x = weekday, y = mean(price)),color="gold2")
(weekday_graph_business)
```



```
#Grouping dataset by weekday
eco_grp_weekday = economy_data %>% group_by(weekday)  %>%
  summarise(mean_price = mean(price),
            median_price = median(price),
            .groups = 'drop')
bus_grp_weekday = business_data %>% group_by(weekday)  %>%
  summarise(mean_price = mean(price),
            median_price = median(price),
            .groups = 'drop')
#Mean and median price in economy by weekday
grp_economy <- ggplot (data = eco_grp_weekday) +
  geom_point(mapping = aes (x = weekday, y = mean_price),color="blue")+
  geom_point(mapping = aes (x = weekday, y = median_price),color="red")+
  ylab("price")
(grp_economy)
```



```
#Mean and median price in business by weekday
grp_business <- ggplot (data = bus_grp_weekday) +
  geom_point(mapping = aes (x = weekday, y = mean_price),color="blue")+
  geom_point(mapping = aes (x = weekday, y = median_price),color="red")+
  ylab("price")
(grp_business)
```

