Analysis of Gold Price Prediction Dataset

IMPORTING LIBRARIES

In [54]:
```
library(dplyr)
library(tidyr)
library(ggplot2)
```

IMPORTING DATA

In [55]:
```
data=read.csv("C:/Users/prathibha k s/Downloads/Gold Price.csv")
```

VIEWING THE FIRST FEW ROWS TO CHECK IF THE DATA IS LOADED CORRECTLY

In [56]:
```
head(data)
```

Date	Price	Open	High	Low	Volume	Chg.
2014-01-01	29542	29435	29598	29340	2930	0.25
2014-01-02	29975	29678	30050	29678	3140	1.47
2014-01-03	29727	30031	30125	29539	3050	-0.83
2014-01-04	29279	29279	29279	29279	0	-1.51
2014-01-06	29119	29300	29395	29051	24380	-0.55
2014-01-07	28959	29130	29195	28912	18710	-0.55

DATA DESCRIPTION

1. Date - Date in standard format
2. Price - It is close price which can be considered as final price
3. Open - Price at the time of market opening at that day
4. High - Highest price during whole day
5. Low - Lowest price during whole day
6. Volume - Traded Volume
7. Chg - % Change from previous price

MISSING VALUE ANALYSIS

In [98]:
```
which(is.na(data)) # no missing data
```

TO CHECK FOR SOME SUMMARY STATISTICS OF THE DATA

In [58]:
```
summary(data)
```

```
        Date          Price            Open            High      
 2014-01-01:   1   Min.   :24545   Min.   :24583   Min.   :24635  
 2014-01-02:   1   1st Qu.:28374   1st Qu.:28349   1st Qu.:28481  
 2014-01-03:   1   Median :30309   Median :30309   Median :30447  
 2014-01-04:   1   Mean   :33756   Mean   :33757   Mean   :33945  
 2014-01-06:   1   3rd Qu.:38181   3rd Qu.:38186   3rd Qu.:38273  
```

```
2014-01-07:   1   Max.   :56117   Max.   :56351   Max.   :56499
(Other)   :2020
     Low             Volume           Chg.
Min.   :24470   Min.   :     0   Min.   :-5.98000
1st Qu.:28170   1st Qu.:  6735   1st Qu.:-0.41000
Median :30117   Median : 11635   Median : 0.02000
Mean   :33562   Mean   : 13841   Mean   : 0.02773
3rd Qu.:37986   3rd Qu.: 18518   3rd Qu.: 0.46000
Max.   :55400   Max.   :106920   Max.   : 5.30000
```

DATA TYPES OF THE COLUMNS

In [59]:
```
str(data)
```

```
'data.frame':   2026 obs. of  7 variables:
 $ Date  : Factor w/ 2026 levels "2014-01-01","2014-01-02",..: 1 2 3 4 5 6 7 8 9 10
...
 $ Price : int  29542 29975 29727 29279 29119 28959 28934 28997 29169 29312 ...
 $ Open  : int  29435 29678 30031 29279 29300 29130 28916 28990 29030 29170 ...
 $ High  : int  29598 30050 30125 29279 29395 29195 29029 29053 29198 29330 ...
 $ Low   : int  29340 29678 29539 29279 29051 28912 28820 28865 28960 29133 ...
 $ Volume: int  2930 3140 3050 0 24380 18710 18140 15130 15810 13780 ...
 $ Chg.  : num  0.25 1.47 -0.83 -1.51 -0.55 -0.55 -0.09 0.22 0.59 0.49 ...
```

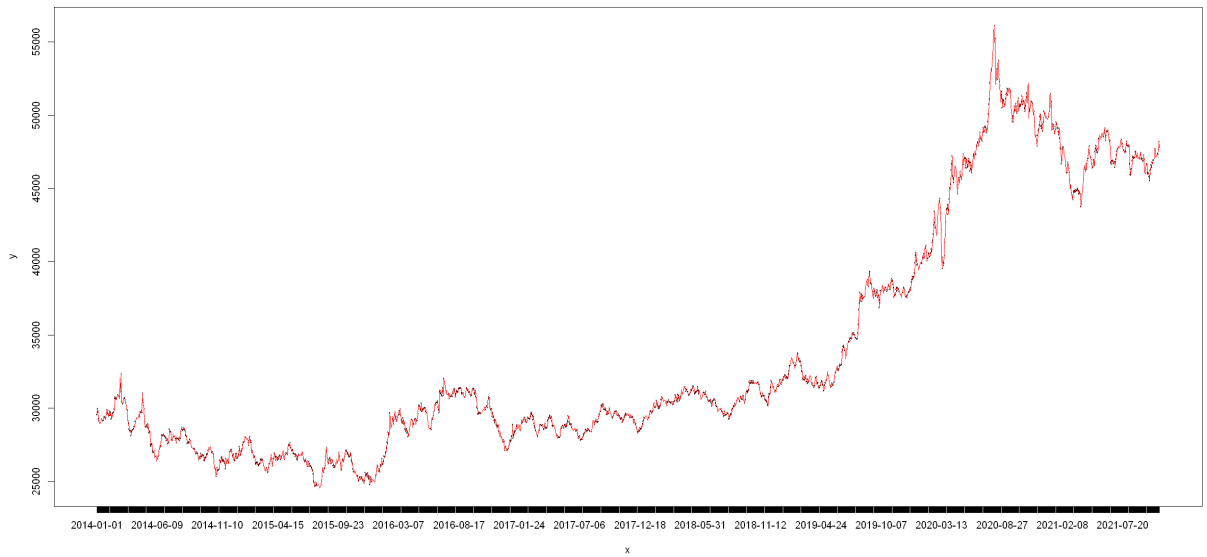
EDA

In [100…
```
options(repr.plot.width=20 ,repr.plot.height= 10)  # graph dimensions
```

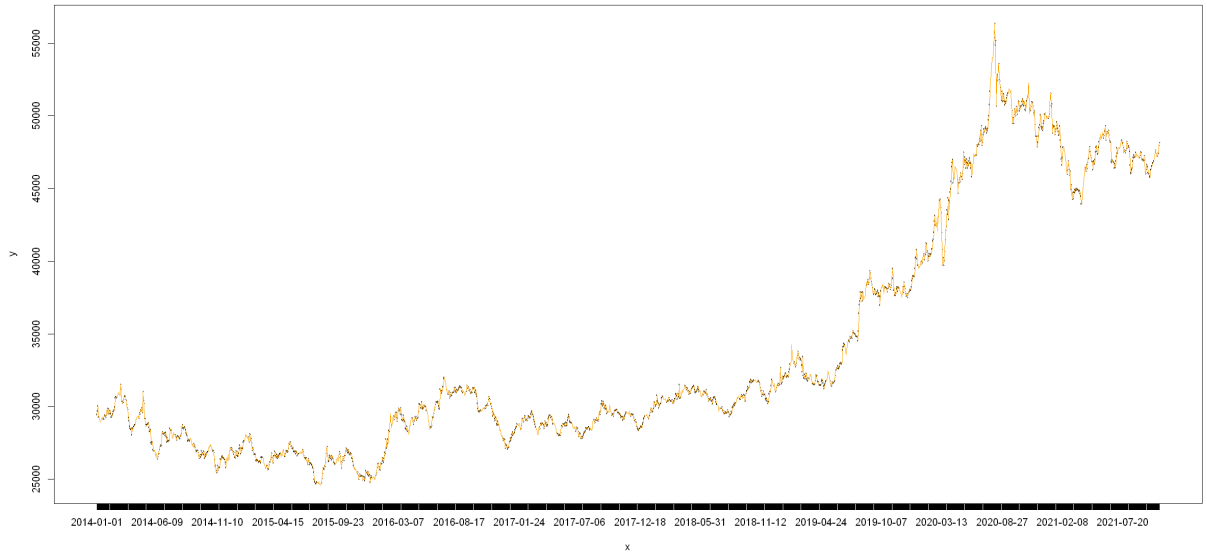
Line plots

- These plots help us analyse the trend of each variables over a time period from 2014 to 2021
- They help us conclude where there was a downfall in prices and where there was a rise

In [73]:
```
plot(data$Date,data$Price)
lines(data$Price,col='red')
```

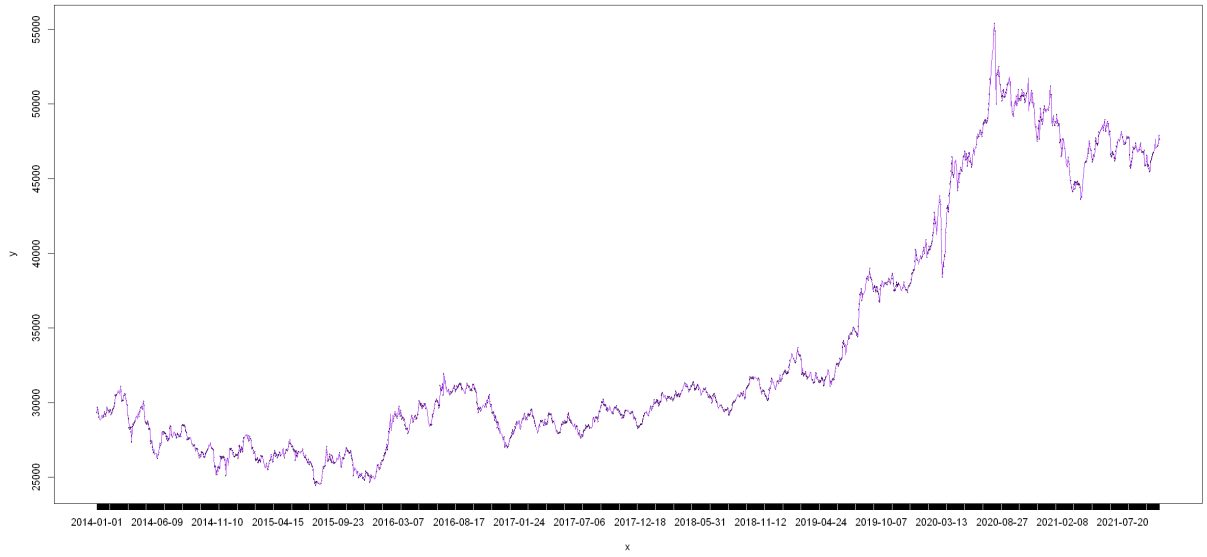


In [74]:
```
plot(data$Date,data$Open)
lines(data$Open,col='orange')
```



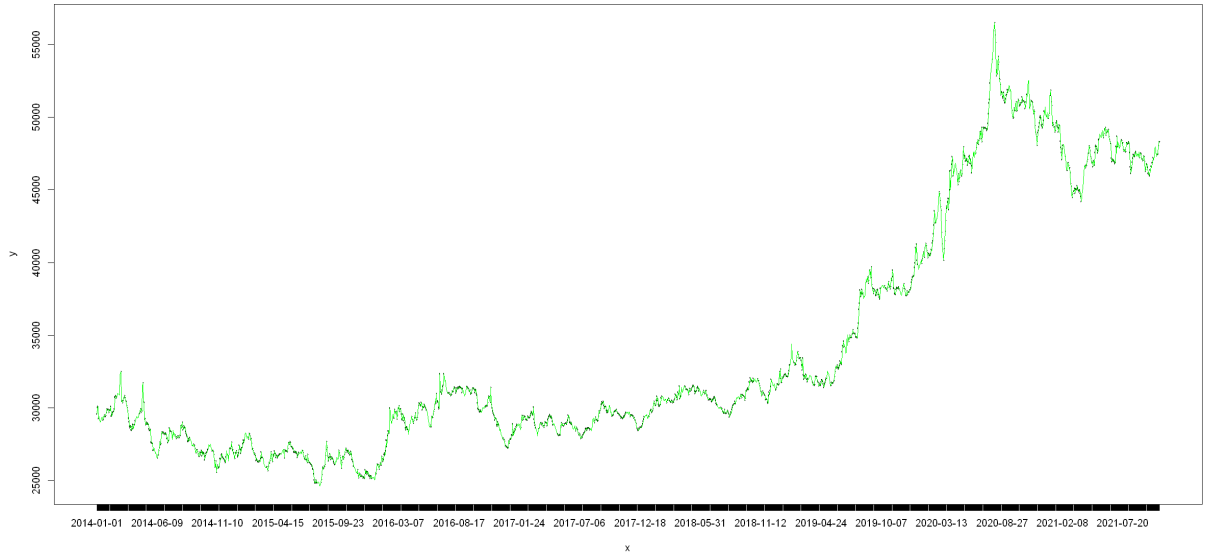
In [75]:

```
plot(data$Date,data$Low)
lines(data$Low,col='purple')
```



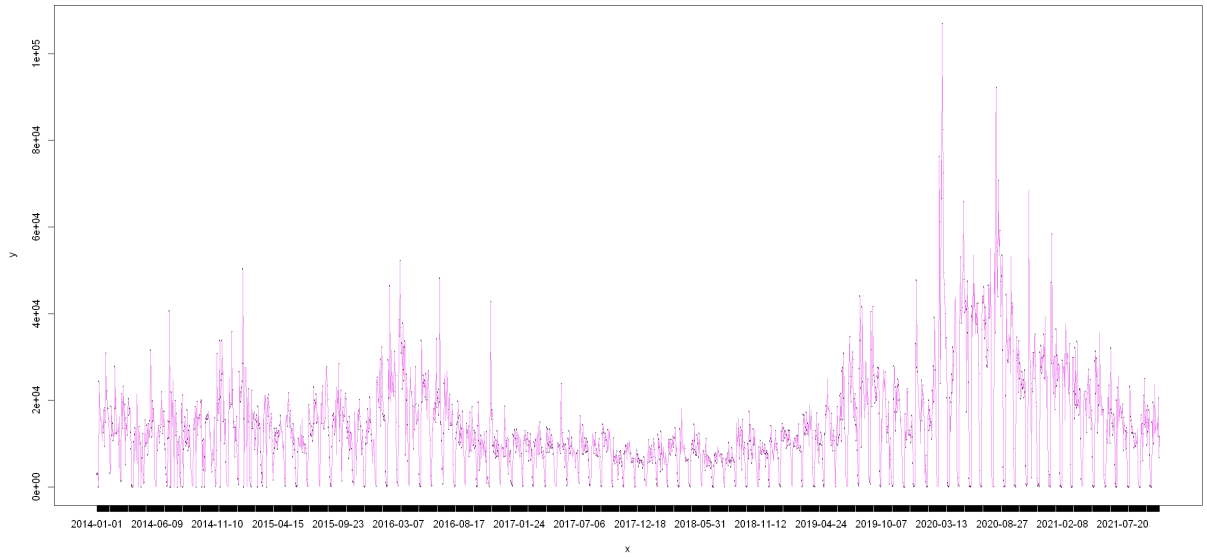
In [76]:

```
plot(data$Date,data$High)
lines(data$High,col='green')
```



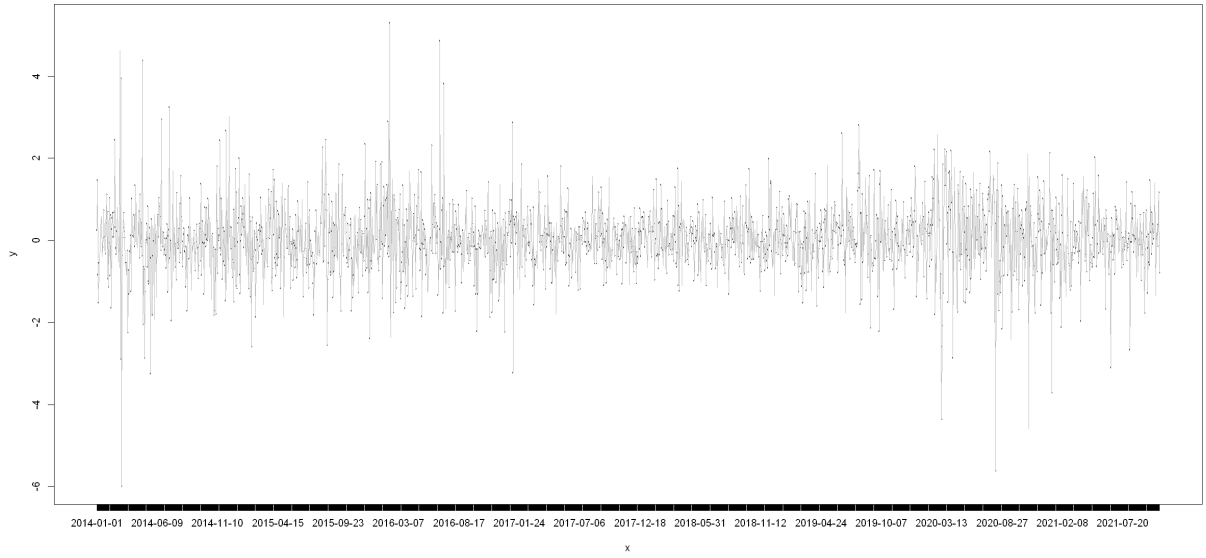
In [77]:

```
plot(data$Date,data$Volume)
lines(data$Volume,col='violet')
```



In [78]:

```
plot(data$Date,data$Chg)
lines(data$Chg,col='gray')
```



Data Transformation

Coverting the Date column from chr datatype to Date format

In [109…

```
data$Date = as.Date(data$Date,"%Y-%m-%d")
data1= mutate(data, Year = format(Date,"%Y")) # sorting by year
data2= mutate(data, Month = format(Date,"%m")) # sorting by month
```

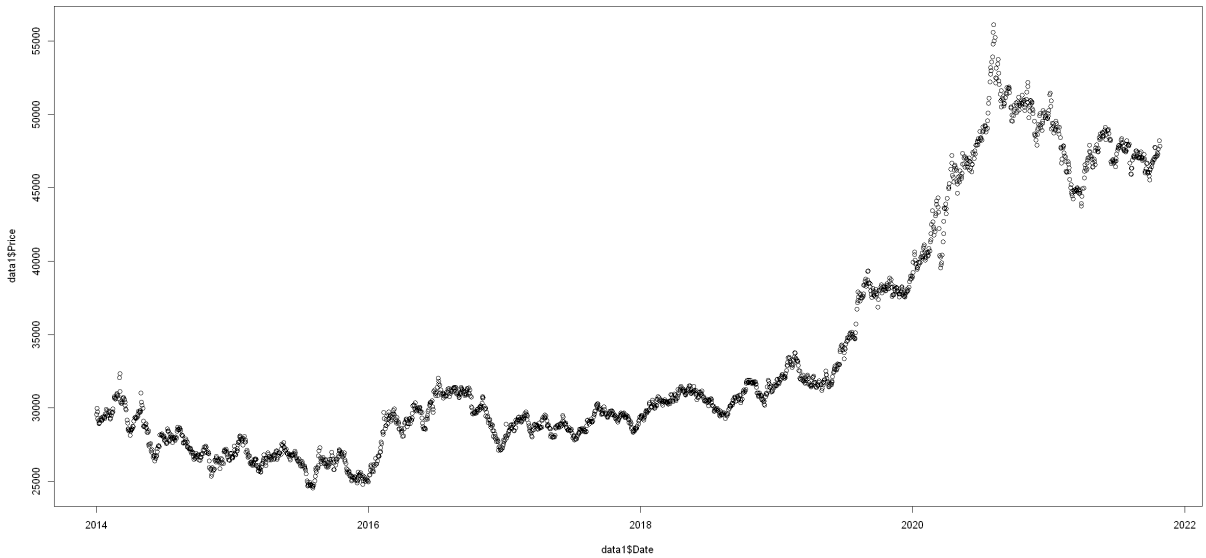
The transformed data can be used for further analysis

In [111…

```
##Sample
```

In [96]:

```
plot(data1$Date,data1$Price)
lines(data1$Price,col='red')
```



To check for when there was the highest % change in the Gold price value

In [86]:

```
max(data$Chg)
```

5.3

In [ ]:
```
data[which(data1$Chg == 5.3),] ## 2016 marks highest % change in gold price value
```

To check when there was maximum volume traded

In [88]:
```
max(data$Volume)
```

106920

In [107…
```
data[which(data1$Volume == 106920),] # maximum volume trading happened during march
```

	Date	Price	Open	High	Low	Volume	Chg.
1612	2020-03-16	39548	40802	41291	38419	106920	-2.07

ANALYSING THE TARGET COLUMN

In [92]:
```
data[which(data1$Price == min(data$Price)),] # we can see the gold price was minimum
```

	Date	Price	Open	High	Low	Volume	Chg.
426	2015-08-05	24545	24614	24635	24500	100	-0.38

In [93]:
```
data[which(data1$Price == max(data$Price)),] # the gold price was maximum during aug
```

	Date	Price	Open	High	Low	Volume	Chg.
1711	2020-08-06	56117	55405	56199	55400	45660	0.94

In [ ]:
```

```