Final

Ugur Askar

2021-01-10

Part I: Short and Simple

1. Briefly describe the controversy around Timnut Gebru, her team's recent research and Google. What are the valid points of each side? What is your position?

Timnut Gebru told her colleagues her concerns about artificial intelligence bias between light skinned and darker skinned, male and female samples and told that she would consider resignation if their conditions are not met and last date for her employment after she returns a vacation. She was fired/her resignation got accepted before she came back from vacation.

In my opinion both of the parties have valid points in their thought process, Gebru was disappointed about AI was not recognizing dark skinned-females and wanted to fix this issue, however I think her approach about it was wrong because she demanded a resignation date and got frustrated after her e-mail considered as resignation letter and told public as she got fired.

I would say that Gebru's view about artificial intelligence is right because there is a discrimination in it however her approach to fix the issue was wrong.

2. How much of a decision be based on gut feeling and experience and how much of it should be based on data and forecasts? For example, suppose you are a seasoned portfolio manager investing in only the stock market. How would you build your portfolio?

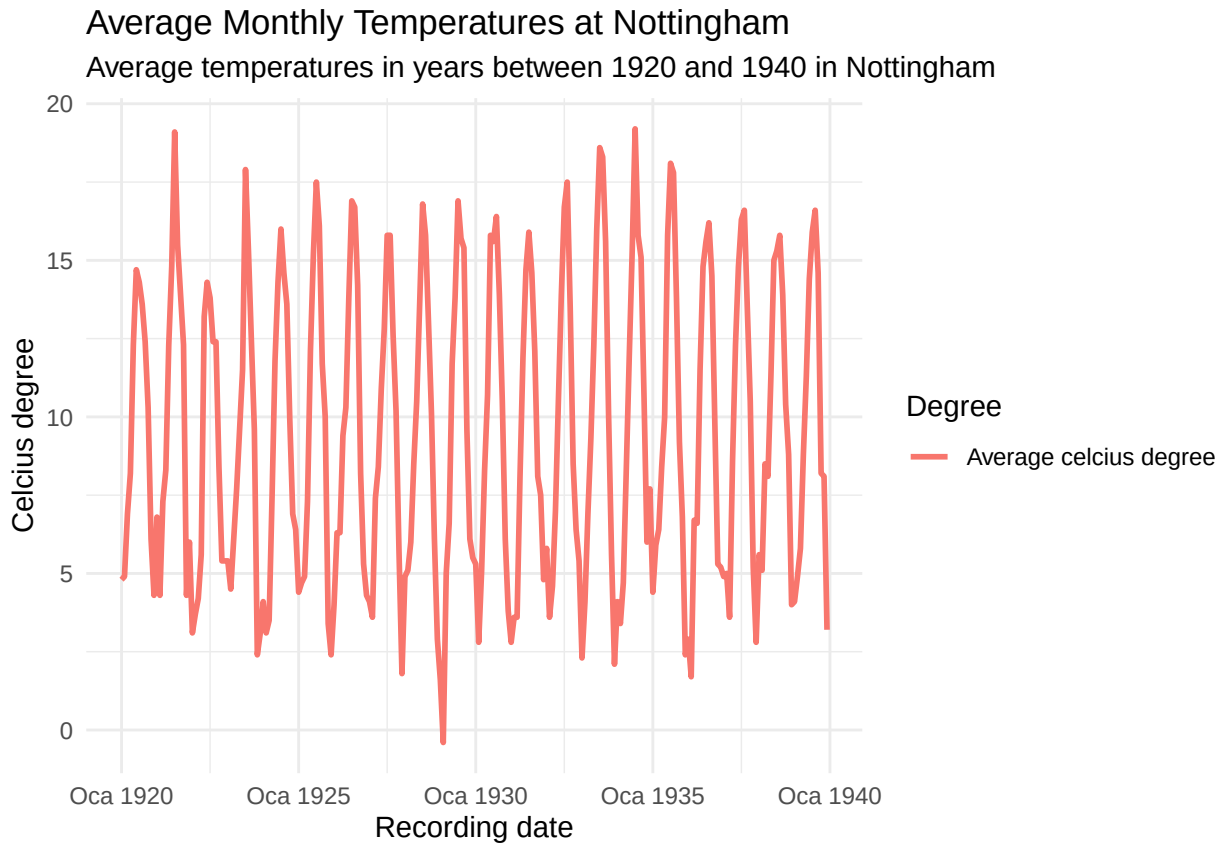
Decisions should be used with mix of both experience and data/forecasts thats one of the reasons why companies don't use **just** IT teams to make their decisions, when you combine both experience and data analysis your decisions become more effective. For example if you looked at data from 2019 and made your decisions with just only data avaible your guess for 2020 would be skewed because you wouldn't check corona cases for your financial decisions, however more experienced business units could predict that transportation companies would lose their revenues because of quarantine. For the example I would combine my experience with datasets to build my portfolio and use best of both worlds.

3. If you had to plot a single graph using the nottem data (provided with base R) what would it be? Why? Make your argument, actually code the plot and provide the output. (You can find detailed info about the data set in its help file. Use ?nottem.)

In my opinion best plot for it would be a line graph with date series at x axis and degrees in y axis, of course first we need to convert it to celcius degrees because reports should be easily understandable by readers and we are not using fahrenheit degrees in general. Line graph will show time-wise changes more obvious than other graphs.

```
nottem_celcius = data.frame(year_month= melt(time(nottem)), celcius = melt(nottem))
colnames(nottem_celcius) = c('year_month','celcius')

nottem_celcius$celcius = round((nottem_celcius$celcius-32)*.5556,1)

ggplot(nottem_celcius, aes(x=as.yearmon(year_month), group=1)) +
  geom_line(aes(y = celcius, colour = "Average celcius degree") , size = 1) +
  theme_minimal() +
  labs(x = 'Recording date', y = 'Celcius degree', color = "Degree", title = "Average Monthly Temperatu
```

```
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
```



Part II: Extending Your Group Project

In our project we didn't made an analysis about funds total values, I am going to analyze top 5 most value gained funds.

To find top 5 most value gained funds I am going to find oldest funds in my dataset and compare it to recent ones by calculating percentage change then I am going to filter top 5 fund codes and filter my original dataset by these fund codes and visualize it in scatter plot.

```
df_monthly = readRDS("C:\\Users\\uguraskar\\Documents\\Github\\mef04g-champions\\mef04g-champions\\montl
newest_df = df_monthly %>%
  filter(date == ymd("2020-11-01")) %>%
  select(code, total_value)
```

```
## Adding missing grouping variables: `year`
```

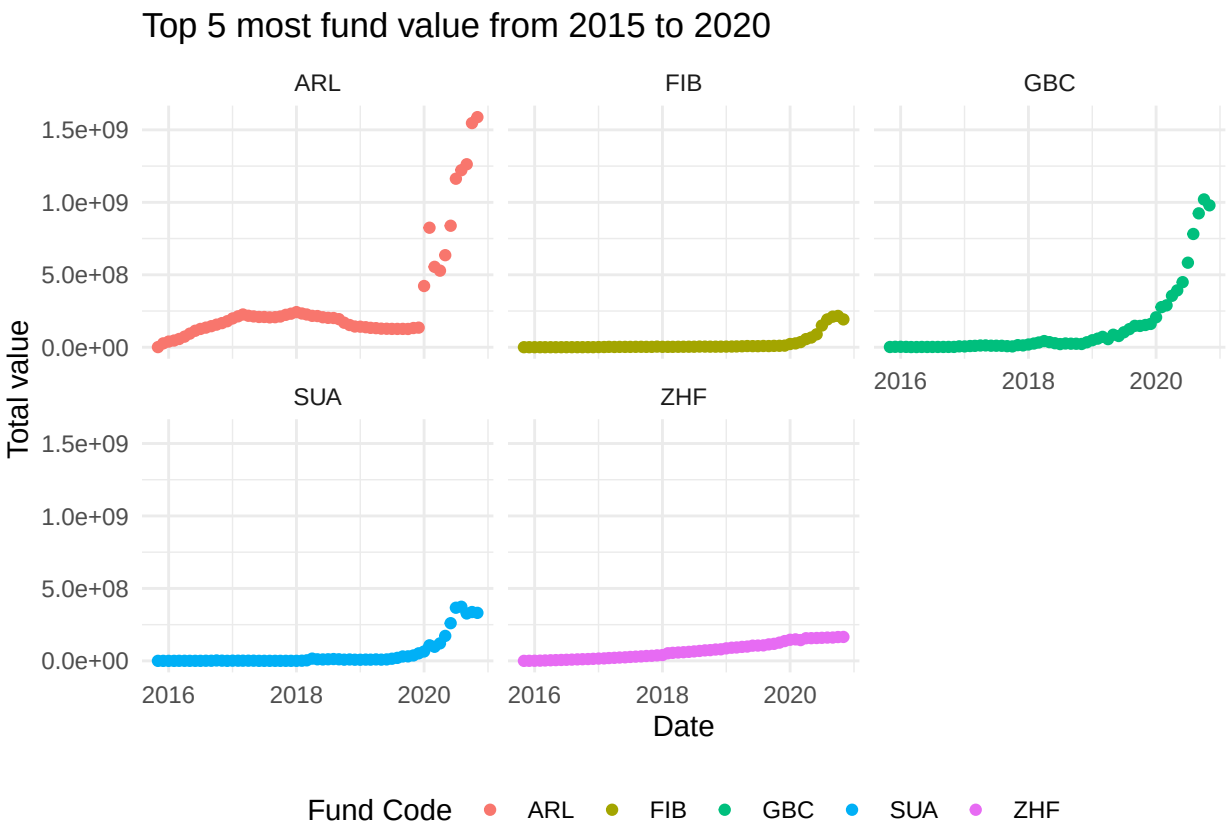
```
oldest_df = df_monthly %>%
  filter(date == ymd("2015-11-01")) %>%
  select(code, total_value)
```

```
## Adding missing grouping variables: `year`
```

```
price_change_df = inner_join(newest_df,oldest_df, by="code")

colnames(price_change_df) = c('new_year','code','new_total_value','old_year','old_total_value')

price_change_df = price_change_df %>%
  select(code, old_total_value, new_total_value) %>%
  mutate(value_change = round(((new_total_value-old_total_value)/old_total_value)*100)) %>%
  arrange(desc(value_change))

top_5_df = price_change_df[1:5,]

categories =
    top_5_df %>%
    distinct(code) %>%
    unlist(.,use.names = FALSE)

top_5_all_df = df_monthly %>%
  filter(code %in% categories) %>%
  select(code, date, total_value)
```

```
## Adding missing grouping variables: `year`
```

```
ggplot(top_5_all_df) +
  geom_point(aes(x=date, y=total_value, color=code)) +
  theme_minimal() +
  theme(legend.position="bottom") +
  facet_wrap(vars(code)) +
  labs(title = 'Top 5 most fund value from 2015 to 2020', x='Date', y='Total value', color='Fund Code')
```



Part III: Welcome to Real Life

a) Gather the data, join and bind them together and save it in an .RData file. You can make .RData file available online for everybody. Provide the data link in your analysis.

```
raw_summary_industry_df = read_xlsx(".\\data\\takipozet_sanayi.xlsx")
raw_estimation_industry_df = read_xlsx(".\\data\\takipTahmin_sanayi.xlsx")
raw_target_industry_df = read_xlsx(".\\data\\takiphedeffiyat_sanayi.xlsx")
raw_summary_insurance_df = read_xlsx(".\\data\\takipozet_sigorta.xlsx")
raw_estimation_insurance_df = read_xlsx(".\\data\\takipTahmin_sigorta.xlsx")
raw_target_insurance_df = read_xlsx(".\\data\\takiphedeffiyat_sigorta.xlsx")
raw_summary_bank_df = read_xlsx(".\\data\\takipozet_banka.xlsx")
raw_estimation_bank_df = read_xlsx(".\\data\\takipTahmin_banka.xlsx")
raw_target_bank_df = read_xlsx(".\\data\\takiphedeffiyat_banka.xlsx")
raw_summary_reit_df = read_xlsx(".\\data\\takipozet_gyo.xlsx")
raw_estimation_reit_df = read_xlsx(".\\data\\takipTahmin_gyo.xlsx")
raw_target_reit_df = read_xlsx(".\\data\\takiphedeffiyat_gyo.xlsx")
raw_summary_holding_df = read_xlsx(".\\data\\takipozet_holding.xlsx")
raw_estimation_holding_df = read_xlsx(".\\data\\takipTahmin_holding.xlsx")
raw_target_holding_df = read_xlsx(".\\data\\takiphedeffiyat_holding.xlsx")

clean_summary_industry_df = raw_summary_industry_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d%m%Y'), sector = 'Industry')
```

```
colnames(clean_summary_industry_df) = c('stock','recommendation','target_price_try','upside','record_da

clean_estimation_industry_df = raw_estimation_industry_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), EstimationUpdateDate = as.POSIXct(Esti
colnames(clean_estimation_industry_df) = c('stock','recommendation','target_price_try','profit_potentia

clean_target_industry_df = raw_target_industry_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'Industry')
colnames(clean_target_industry_df) = c('stock','recommendation','target_price_try','upside','record_dat

clean_summary_insurance_df = raw_summary_insurance_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'Insurance')
colnames(clean_summary_insurance_df) = c('stock','recommendation','target_price_try','upside','record_da

clean_estimation_insurance_df = raw_estimation_insurance_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), EstimationUpdateDate = as.POSIXct(Esti
colnames(clean_estimation_insurance_df) = c('stock','recommendation','target_price_try','profit_potentia

clean_target_insurance_df = raw_target_insurance_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'Insurance')
colnames(clean_target_insurance_df) = c('stock','recommendation','target_price_try','upside','record_da

clean_summary_bank_df = raw_summary_bank_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'Bank')
colnames(clean_summary_bank_df) = c('stock','recommendation','target_price_try','upside','record_date',

clean_estimation_bank_df = raw_estimation_bank_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), EstimationUpdateDate = as.POSIXct(Esti
colnames(clean_estimation_bank_df) = c('stock','recommendation','target_price_try','profit_potential','

clean_target_bank_df = raw_target_bank_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'Bank')
colnames(clean_target_bank_df) = c('stock','recommendation','target_price_try','upside','record_date','

clean_summary_reit_df = raw_summary_reit_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'REIT')
colnames(clean_summary_reit_df) = c('stock','recommendation','target_price_try','upside','record_date',

clean_estimation_reit_df = raw_estimation_reit_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), EstimationUpdateDate = as.POSIXct(Esti
colnames(clean_estimation_reit_df) = c('stock','recommendation','target_price_try','profit_potential','

clean_target_reit_df = raw_target_reit_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'REIT')
colnames(clean_target_reit_df) = c('stock','recommendation','target_price_try','upside','record_date','

clean_summary_holding_df = raw_summary_holding_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'Holding')
colnames(clean_summary_holding_df) = c('stock','recommendation','target_price_try','upside','record_dat

clean_estimation_holding_df = raw_estimation_holding_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), EstimationUpdateDate = as.POSIXct(Esti
colnames(clean_estimation_holding_df) = c('stock','recommendation','target_price_try','profit_potential
```

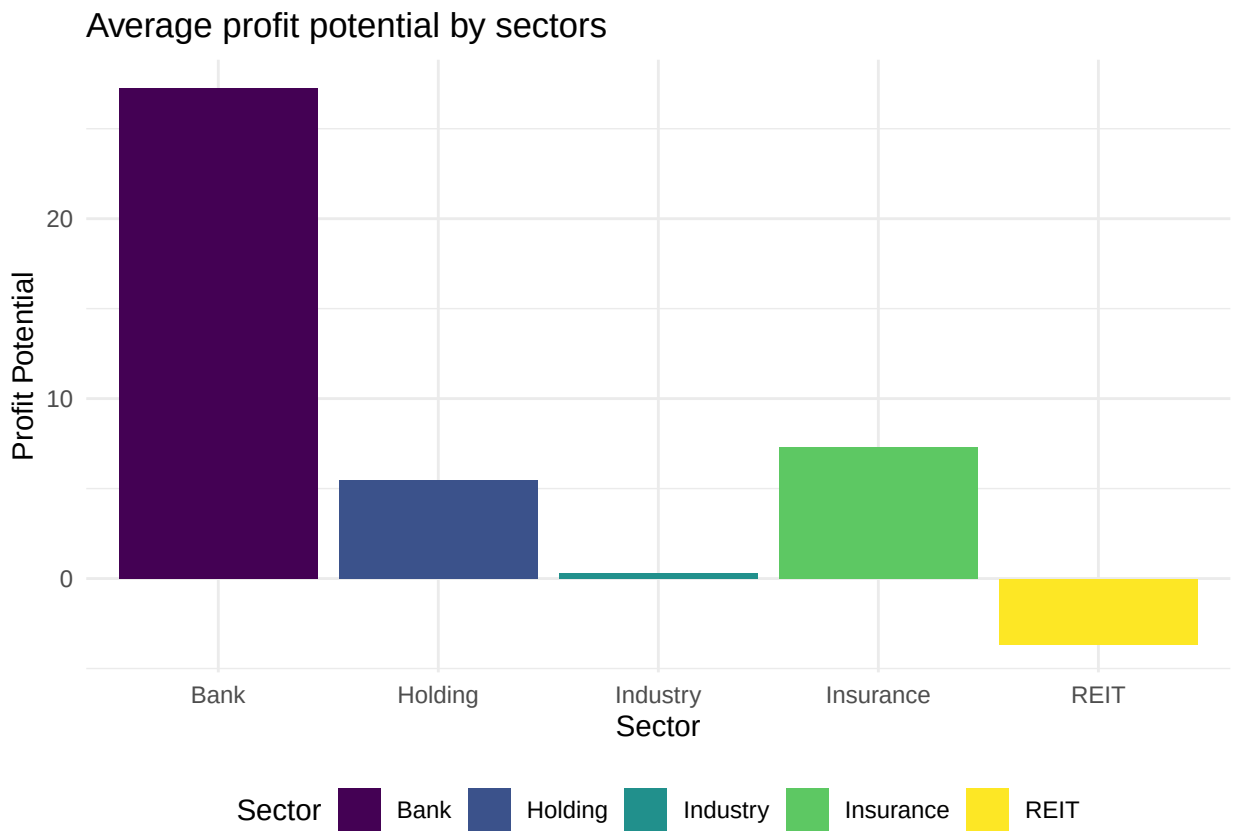
```
clean_target_holding_df = raw_target_holding_df %>%
  mutate("Rec. Date" = as.POSIXct(`Rec. Date`,format='%d.%m.%Y'), sector = 'Holding')
colnames(clean_target_holding_df) = c('stock','recommendation','target_price_try','upside','record_date

clean_summary_df = bind_rows(clean_summary_industry_df, clean_summary_insurance_df, clean_summary_bank_

clean_estimation_df = bind_rows(clean_estimation_industry_df, clean_estimation_insurance_df, clean_estim

clean_target_df = bind_rows(clean_target_industry_df, clean_target_insurance_df, clean_target_bank_df, 

isyatirim_df = inner_join(clean_summary_df,clean_estimation_df, by="stock") %>%
  inner_join(.,clean_target_df, by="stock")
colnames(isyatirim_df) = c('stock','recommendation','target_price_try','upside','record_date','close_try
isyatirim_df = isyatirim_df %>%
  select(stock, sector, record_date, recommendation, target_price_try, upside, close_try, market_cap_try

save(isyatirim_df, file = ".\\data\\isyatirim_df.RData")
```

You can view this dataset from this link : https://github.com/pjournal/mef04-uguraskar/blob/gh-pages/data/isyatirim_df.RData

b) Perform EDA on the data you collected based on the theme you decided on. Keep it short. One page is enough, two pages tops. Original and interesting work is important.

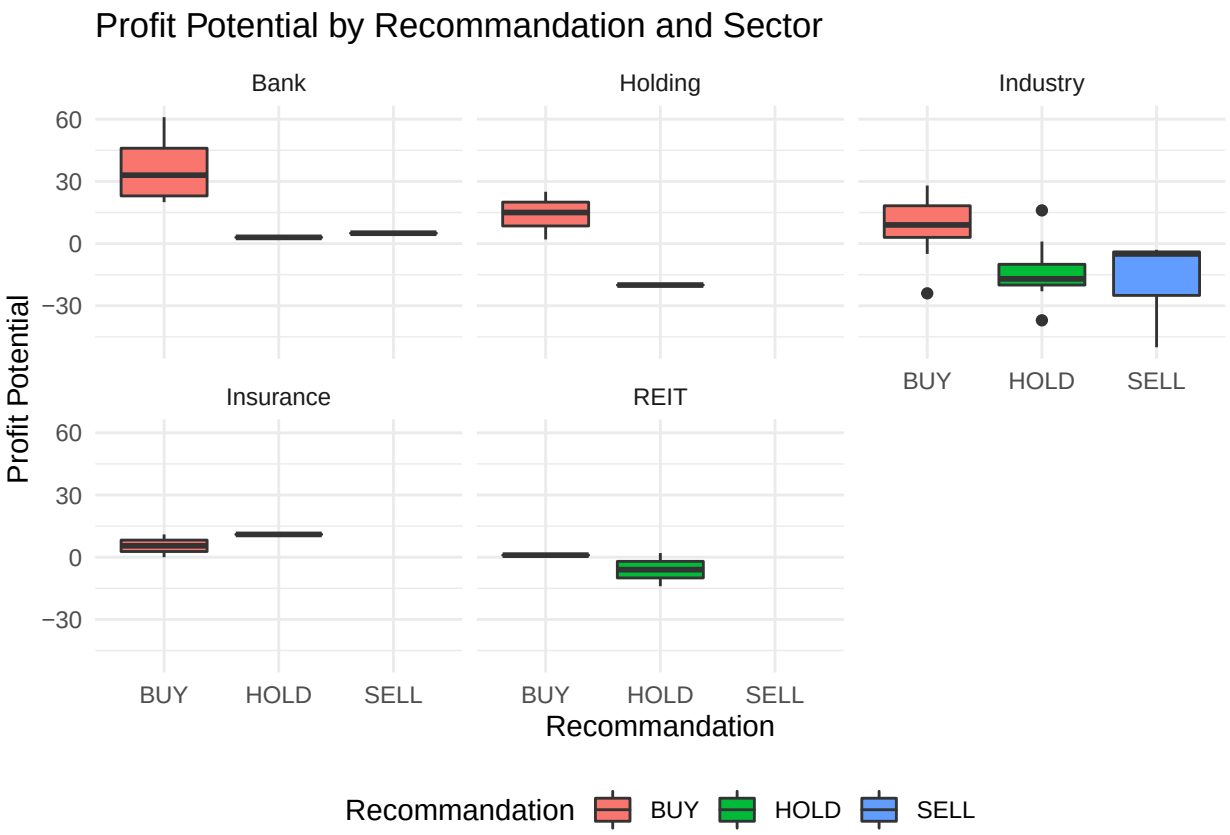
```
load(".\\data\\isyatirim_df.RData")

ggplot(isyatirim_df %>%
  group_by(sector) %>%
  summarise(average_profit_potential = mean(profit_potential, na.rm = TRUE))) +
 aes(x = sector, fill = sector, weight = average_profit_potential) +
 geom_bar() +
 scale_fill_viridis_d(option = "viridis") +
 labs(x = "Sector", y = "Profit Potential", title = "Average profit potential by sectors", fill = "Secto
 theme_minimal() +
 theme(legend.position = "bottom")
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



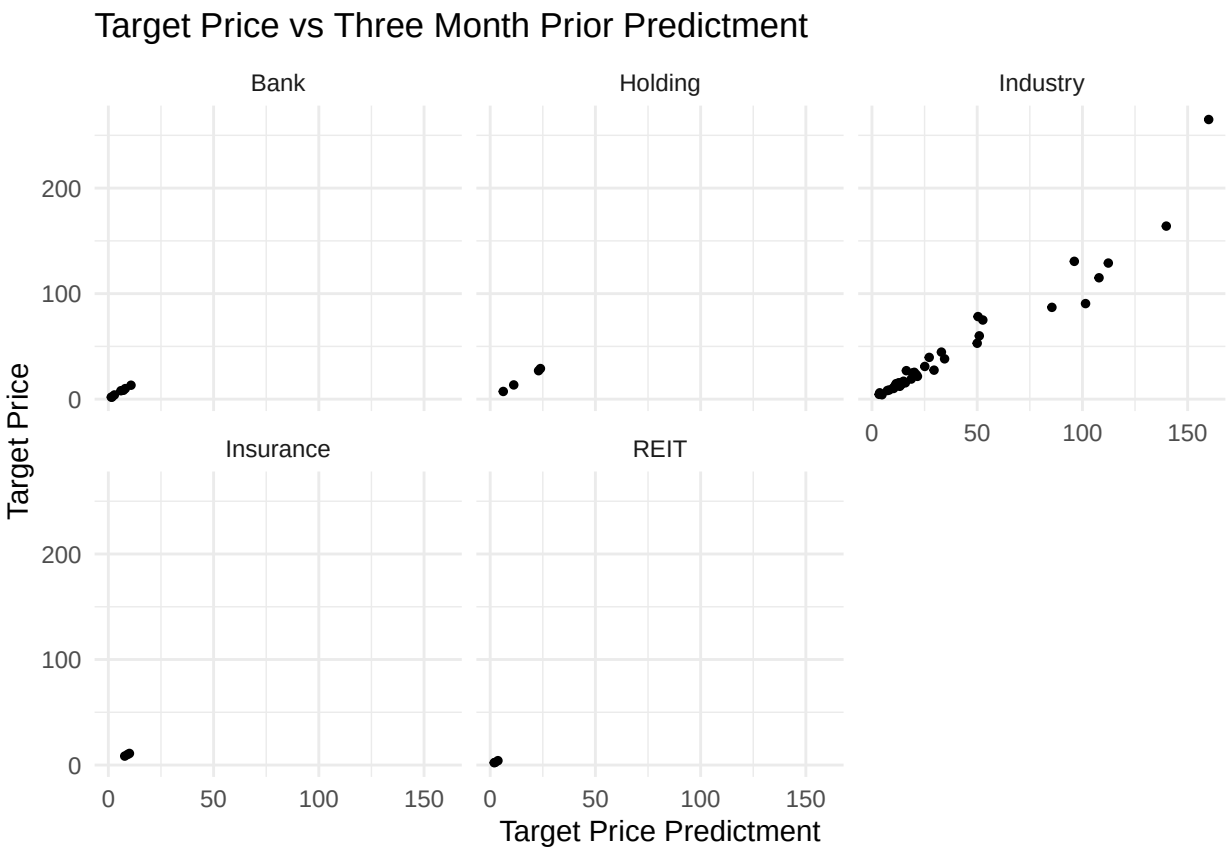
We can clearly see that currently investing in banking sector would yield more profit than others.

```
ggplot(isyatirim_df) +
 aes(x = recommendation, y = profit_potential, fill = recommendation) +
 geom_boxplot() +
 scale_fill_hue() +
 labs(x = "Recommandation", y = "Profit Potential", title = "Profit Potential by Recommandation and Sec
 theme_minimal() +
 theme(legend.position = "bottom") +
 facet_wrap(vars(sector))
```



We can see that generally you should actually sell at "hold" recommendations except for insurance sector. Also buying a stock in insurance sector doesn't look promising.

```
ggplot(isyatirim_df) +
 aes(x = three_month_target_price_try, y = target_price_try) +
 geom_point(size = 1L) +
 scale_color_hue() +
 labs(x = "Target Price Predictment", y = "Target Price", title = "Target Price vs Three Month Prior Pr
 theme_minimal() +
 facet_wrap(vars(sector))
```



We can see that only industry sector exceeded predictments made three months ago. Other sectors were way lower than expectations.