Final_Project

Part I: Short and Simple

Question 1

We have started to see the impact of generative AI in our lives. Although it is not yet in our daily lives, it will be in the future and the discussions about it have started. The main question seems like whether or not AI will take our jobs. I do not think this is the case but there may be some jobs affected by the usage of AI negatively or positively. For some jobs, the time and effort to do a job by a single person will be decreased with these programs. This is efficient but may also result in lay offs in some departments. On the other hand, there may be new professions due to generative AI commonly used in our daily lives. Like the social media related professions that did not exist before the social media programs were used this much by everyone. I think there will be negative effects of generative AI, since these processes since we can identify some of the problems only when they occur. For example, right know some artists are not happy with the fact that using generative AI their art skills can be copied. What will be the legal consequences of this? Will the end result of the AI generated image using a specific artist's style belong to the artist or the person using the program? Will the artist be entitled to royalties if the product made using their art style used commercially? There are a lot of legal questions unanswered and I believe these questions will increase with generative AI's being more common. Overall, I think these programs will provide more efficient access to information for us and it looks like we can save a lot of time in learning, making projects or other areas. In this case, the communication skills will become more important. To be able to describe the end result we want clearly will be important in how effective these programs will be for us.

Question 2

For this project, the first step would be to clearly define the project. Questions such as data used in the project, definition of the variables used in the project, what will the filters be should be answered. If this is a project currently being made and we want to automate it, ideally the best situation would be using a program where we can upload the data and get the

end result in the same project. If possible the data upload part could be automated too such as making dashboards updated daily, weekly or monthly. Once every step of the project is clear we could make a single project that gives the end result with minimum effort. But here we should be careful that the data is correct. There may be problems in some variables and if these go unnoticed could result in a wrong end result. So we should be aware of what values our variables can take and we can add controls when there is an unexpected value in the data. Also there may be controls in the end result too. For example if we do not expect a negative result or a value higher than some upper bound in our report, we could add some warning for these cases after our project runs and gives the end result.

Question 3

This data gives the daily downloads of the packages tidyverse, shiny, rmarkdown and reticulate. Here, I would plot the graph of the downloads of each package for each day. With this graph we could see which package is downloaded more and also if there is a specific part of the month, for instance start or end of the month, that the number of downloads differs from the rest of the month. The plot is given below. Shiny looks like the most downloaded package but there is a significant decrease at the end of the month so to be sure we need to look at the download counts too.(Rmarkdown is the most downloaded package when we look at the counts.) Reticulate is the least downloaded package. And in the first half of the month number of downloads is higher. To see if that is a pattern we could look at the yearly data.

```r
#install.packages("cranlogs") # install package if not installed
library(cranlogs)
library(ggplot2)
data_cranlogs <- cranlogs::cran_downloads(
  packages=c("tidyverse","shiny","rmarkdown","reticulate"),
  from="2022-11-01",to="2022-11-30")
ggplot(data_cranlogs ,aes(x=date,y=count,color=package)) + geom_line()
```
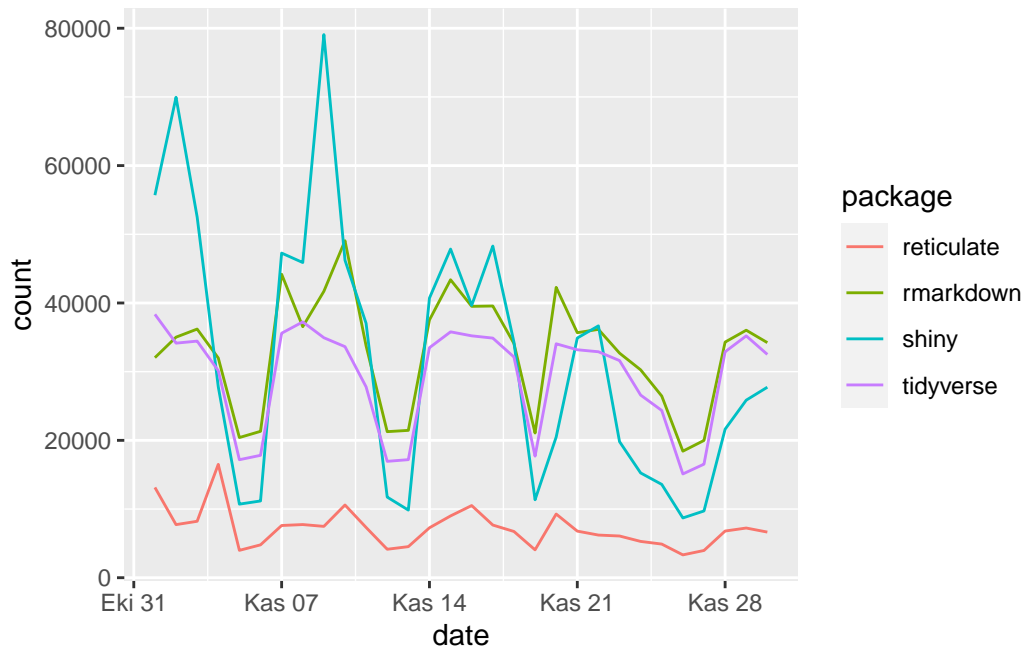
```
#number of downloads for each package
#data_cranlogs %>% group_by(package) %>% summarise(sum(count))
```

Part II: Extending Group Project

For this part, I have extended the analysis for firm size comparison. When we look at the number of employees by year and firm size, we see that until 2017 the number of employees is increasing for all firm sizes except micro sized firms. After 2017 large, medium and small sized firms' employee count first decreases, then starts to increase again after 2020. Micro sized firms' number of employees continuously increases after 2017.
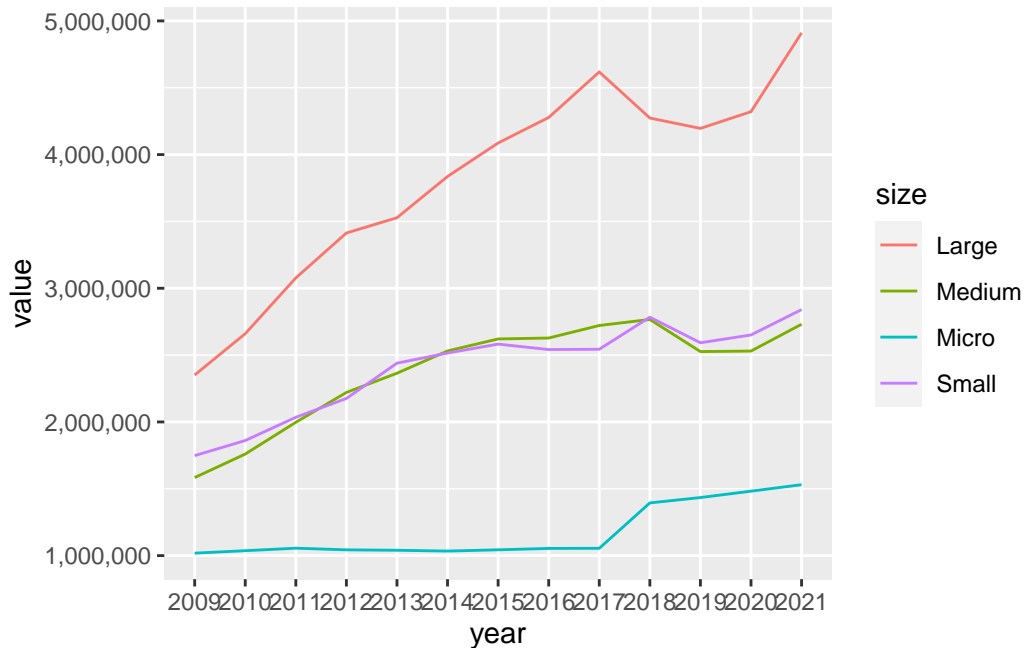
```r
library(scales)
library(tidyverse)

sector_data_org <- read_rds("Group_Project_Data/sector_information.rds")

sector_data <- sector_data_org %>%
  filter(number == 'Number_of_employees') %>%
    filter(size != 'Total') %>%
      group_by(size, year) %>%
```

3

```
        summarise(value=sum(value))


ggplot(sector_data ,aes(x=year,y=value,color=size, group=size)) +
    geom_line() + scale_y_continuous(labels = comma)
```


We now look at the number of companies using the same structure we see that large and medium sized companies' count have not increased over the years, slight decrease could be seen after 2017. Small sized firms have experienced an increase until 2017, we see that it is decreasing after that point. Micro sized firms have the largest increase in all of them. This seem to correlate with the increase in the number of employees.

```
sector_data_2 <- sector_data_org %>%
    filter(number == 'Number_of_companies') %>%
    filter(size != 'Total') %>%
    group_by(size, year) %>%
    summarise(value=sum(value))


ggplot(sector_data_2 ,aes(x=year,y=value,color=size, group=size)) +
    geom_line() + scale_y_continuous(labels = comma)
```
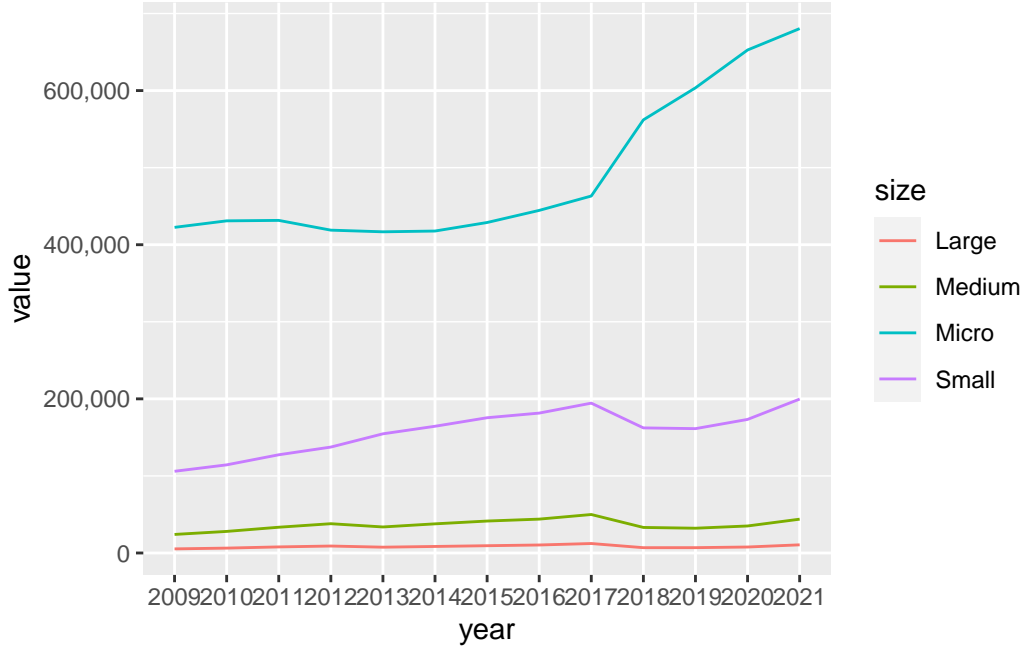
For a detailed analysis, we should look at the developments in the years prior to 2017. Combined with this knowledge, we could have clearer explanations for the reasons behind the sudden change in 2017. These reasons may have political or legal backgrounds such as changes in political systems or changes in legislation. Also, this analysis can be combined with employee wages over the years so that we could see how the growth of these firms have affected the employees in the sector.

Part III: Welcome to Real Life

Here I have used Landing and Takeoff numbers and total number of passengers in the "Domestic and International Transport and the Number of Arrivals-Departures on the Airports According to Classification of Statistical Region Units, Level 1, 2, 3" report data. By dividing total number passengers to the total number of landings and takeoffs we get a ratio we could use for comparing the airports' efficiency of passenger transportation . Therefore, in this analysis only the commercial flights are considered. Results show that Antalya, Muğla (Dalaman), Van (Ferit Melen), Ordu Giresun and Şırnak (Şerafettin Elçi) airports have the top 5 efficiency in terms of total landing & take-offs by passenger count. Some of these are touristic cities and have high numbers in population. If the nearing cities do not have an airport or do not have regular flights these airports could have been used by more than one city. The least 5 efficient airports are Çanakkale (Gökçeada), Balıkesir, Tokat, Uşak and Siirt. The reason for these

airports' efficiency numbers being low could be the low numbers in population, closeness to big cities, inaccurate planning of flight numbers.

```r
library(readxl)
library(tidyverse)

saveRDS(read_excel("Airport_data.xlsx"), "C:/Users/zyagm/R/mef06-gunerzy/Airport_data.rds"
airport_data_org <- readRDS("Airport_data.rds")

airport_data <- airport_data_org %>%
  mutate (commercial_landing_takeoff = Landings_Takeoffs_Total - Other)
airport_data <- airport_data %>%
  mutate(airport_effc =
           ifelse(Passengers_Total > 0 | commercial_landing_takeoff > 0, Passengers_Total/
#view(airport_data)
airport_data <- airport_data %>%
  group_by(IBBS_Detail) %>%
  filter(IBBS_Detail!="Türkiye") %>%
  mutate(count=n()) %>%
  summarise(airport_total_effc = sum(airport_effc/count))

airport_data_desc <- airport_data %>%  arrange(desc(airport_total_effc))

airport_data_desc<-airport_data_desc %>% slice(1:5)

airport_data_asc <- airport_data %>% arrange(airport_total_effc)

airport_data_asc<- airport_data_asc %>% slice(1:6)

airport_data_graph<-bind_rows(airport_data_desc,airport_data_asc)
#Since there are no commercial flights for Aydın, we consider the other airports.
airport_data_graph<-airport_data_graph%>%
  filter(airport_total_effc>0)

graph<-airport_data_graph%>%
  ggplot(aes(x=IBBS_Detail,y=airport_total_effc, fill=IBBS_Detail)) +
  geom_bar(stat='identity', position='dodge') +
  xlab("Airport Name") +
  ylab("Efficiency") +
  theme(axis.ticks.x = element_blank(), axis.text.x = element_blank())

graph
```