Nichol Ho

## I. Introduction

TechCorp faces the increasing competition for top-tier talent in the tech industry. They aim to refine their talent acquisition strategies by analyzing their historical hiring data. They want to uncover the key factors that contribute the successful hires, such as education background, technical skills, digital presence, and soft skills, as well as potential biases or areas for improvement in their hiring process.

*There are 3 main datasets provided from TechCorp.*

1. Candidate Interview and Hiring: provides candidate’s information such as age, gender, and communication skill level, cultural fit, preferred work environment.
2. Candidate Personal Data: provides candidate personal data such as role, years of experience, education level, graduate from elite college.
3. Candidate Technical Scores: provides candidate technical test score such as python, java, machine learning, etc and GitHub profile.

The goal of this project is to provide a detailed report highlighting the main factors decision in the previous hiring outcomes. A series of data visualization that encapsulate key insights, the results of hired candidates. Identify the potential areas of improvement or bias in the hiring process.

The report will focus on several characteristics such as education level, year of experience, technical skill score, interview score, communication skill, and digital profile. Data and exploratory analysis will be discussed in the next section. The conclusion and recommendations with key insights and potential areas to focus on will be provided at the end.

## II. Analysis

**all\_table (joined all 3 table Interview and Hiring, Personal Data, Technical Scores together by candidate.id) hired\_candidates is the table filter candidates are hired from all\_table.**

**2.1 Education level and Experience impacts**

*Do candidates graduated from elite college are more likely to be favored in hiring process or it depend on their experience?*

**Data**: use all\_table to summary to table 1 & visualize to figure 1.

**Method**: Summary data, group by graduation of elite college, education level and the metrics of average years of experience. Visualize the bar to show the hiring outcome based on education level.

**Analysis**:

Let candidates who do not graduate from elite college denote group 1. Let candidates graduate from elite college denote group 2.

In table 1. Those hired have the average years of experience of group 1: [5.4 – 6.16] years, and group 2: [6.12 – 7.28] years.

While group 2 those are not hired due to their below average of years’ experience. This show that graduate from elite is not the main decision factor in hiring process. We can see that the hiring outcome will depend on years of experience.

From figure 1 in the appendix shows the hiring outcome of candidates applied in the company with their education level. Due to a larger amount of applications, group 1 get more rejection compared to group 2 because they are not qualified (years of experience).

**Result**:

Regardless of whether they have an elite college degree or not, those who are hired have around [6-6.2] years of experience start from high school to bachelor level, [5.4 – 5.5] years if candidates have master’s or PhD.

**2.2 Technical skill score and Interview score**

*How much technical skill score and interview score candidates need in order to be hired?*

**Data**: use all\_table to summarize table 1 and visualize figure 2.

**Method**: Summary data, group by graduation of elite college, education level and the metrics of average technical score and interview score.Use geom\_violin to visualize the technical test score & interview score to the hiring outcome.

**Analysis**:

Let candidates who do not graduate from elite college denote group 1. Let candidates graduate from elite college denote group 2.

From table 1. The average of technical score in group 1: [8.94 – 9.12] point and average interview score [8.86 – 9.14] point. The average technical score in group 2 have a range from [7.8 – 9.00] point and average interview score from [7.72 – 8.46] point.

From figure 2. There’s a higher correlation between interview score and hiring.

**Technical score**: candidates with a wide range in technical test score [2.5 – 9.5] were hired. Indicating technical skill is important but may be less decisive.

**Interview score**: Candidates with interview score [7.5 – 10] are frequently hired, even if their technical score were below average.However, candidates with interview score below 7 required technical score of at least 6 to be considered for hiring.

Together, candidates must prepare well for their interview because it’s the main factors decision.

**Results**:

Candidates need to earn an interview score from 7 or above to highly security their offer. Candidates who can earn at that average or more for both technical & interview score will highly security their offer.

**2.3 Top 3 score of each average technical test based on specific role.**

*What are top 3 of average specific technical skill score needed for an applied role?*

**Data**: use hired\_candidates to calculate table avg\_score and visualize figure 3.

**Method**: Group by Position.Applied.For from hired\_candidates to do a pivot table including all of each technical skill and their score.From that table, calculate the average score group by position, visualize the table using geom\_point to show top 3 technical score of each role.

**Analysis**:

Top 3 average technical skill score of each role.

* Data scientist: Java [5.91], JavaScript [5.69], Python [5.65]
* Database Administrator: JavaScript [6.28], CSS [6.05], Java = Python [5.48]
* Software Engineer: Python [5.77], SQL [5.72], R [5.57]
* System Analyst: Python [6.35], JavaScript [6.08], CSS [5.94]
* Web Developer: R [6.31], HTML [6.04], CSS [5.81]

The analysis highlights that specific technical skills have a higher average score for different roles, which may indicate areas of expertise that are most valuable for each position.For instance, Data Scientists show strength in Java and JavaScript, while System Analysts excel in Python and JavaScript.

**Results**:

For future hiring, TechCorp must modify job descriptions to emphasize the top 3 or skills for each role, ensuring that candidates with the right expertise apply. TechCorp can focus on screening candidates who display strong performance in the top skills identified for each role. If a candidate is selected for a role where they lack one of the key technical skills, targeted training could bridge the gap.

**2.4 Soft skill and hiring outcome**

*How do communication skill level correlate and hiring success?*

**Data**: use all\_table to summary to table 4 & visualize to figure 4

**Method**: Use tabyl to see the summary of the soft skill range of candidates. Geom\_bar to visualize the proportion of candidates who are hired.

**Analysis**:

From table 2, there are around 670 – 730 candidates for each level of communication. The result of candidates who are hired at poor level of communication is 50, below average is 57, average is 35, good is 50 and excellent is 54.

Figure 2 can show more visualize the proportion is mostly the same each level. Candidates have lowest communication level are still being hire this show that the low level of communication is not affected to the hiring outcome.

**Result**:

Communication skill is not a factor that influences the hiring outcome.Some candidates have the lowest communication level but still get hired.

**2.5 Digital presence and hiring outcome**

*Do candidates with digital presence will get a higher hiring success?*

**Data**: use hired\_candidates to summary statistic table 5.

**Method**: summary of digital presence on hiring success, by counting LinkedIn Profile and GitHub Profile, hired.

**Analysis**:

From table 5, with candidates are hired with both accounts have 25, with LinkedIn profile only have 53 and have GitHub profile only have 72, with no profile have 96.The result show having a GitHub profile might be more beneficial for hiring outcome compared to LinkedIn alone or having both profiles. Company can check candidate project on GitHub, bring more information about that candidate’ skill.

**Result**:

Tech job is more flavor to candidates who have GitHub profile to check their projects. Having only LinkedIn project or both profiles appears to be less effective in this context.

**2.6 The impact of age and hiring outcome**

*Does candidate’ age and hiring outcome have a relationship?*

**Data**: use hired\_candidates to create age\_group table and from age\_group table visualize figure 6.

**Method**: group by age and count hired candidates from hired\_candidates table. Visualize bar to see if there’s a trend or relationship between age and hiring outcome.

**Analysis**:

From figure.6. The range of age candidates were hired from [22-32].The group age of 28 has the highest count of hired candidates (30) and there are 3 notable peaks of age [23 – 24], [27 – 29] and [31-32].The peak correlate with candidates who completed their education level: bachelor, master or PhD. Which aligns with TechCorp’s focus on youthful talent.

**Result**:

Indicates TechCorp’ strong attraction to younger, educated talent. The hiring trend suggest outreach to universities and higher education or campus recruitment events for maintain these age group.

## III.Conclusion/Discussion

**a. Revising Key Questions:**

* **Education Level vs. Experience**: Graduates from elite colleges show a slight edge in hiring, but years of experience are more critical.
* **Technical & Interview Scores**: Candidates need interview scores of 7 or above to secure offers. Technical scores are important but less decisive.
* **Top Technical Skills by Role**: Emphasizing key skills in job descriptions will help attract the right candidates.
* **Soft Skills & Hiring Success**: Communication skills show minimal impact on hiring outcomes; low-level candidates still get hired.
* **Digital Presence**: Having a GitHub profile is more beneficial for hiring success than just a LinkedIn profile or having both.
* **Age & Hiring Outcome**: TechCorp prefers younger, educated candidates, particularly those in the 22-32 age range.

**b. Additional Observations**:

TechCorp’s hiring strategies could benefit from focusing on experience and specific technical skills rather than solely educational pedigree. TechCorp should position itself as attractive by highlighting career growth in job posting to recent graduates, emphasizing growth opportunities.

**c.Future Directions**:

* **Develop New Evaluation Metrics**: TechCorp should consider creating more nuanced evaluation metrics that assess practical skills through real-world simulations or projects in addition to traditional technical assessments.
* **Enhance Interview Processes**: Implementing structured interviews that focus on specific competencies can reduce bias and ensure that all candidates are evaluated consistently.

**d. Recommendations**:

* **Balanced Skill Evaluation**: TechCorp should place equal emphasis on technical and communication level in their hiring processes to ensure well-rounded hires.
* **Tailor Job Offerings**: create more opportunities for younger talent such as internship, partnership with university. Implement policies that promote work-life balance and professional development to retain employees.
* **Continuous Feedback Loop**: Establish a feedback loop that includes hiring managers and candidates to continuously refine the hiring process based on outcome and candidate experiences.
* **Encourage Digital Presence**: Use platforms like GitHub and LinkedIn as supplementary evaluation criteria, particularly for candidates with lower interview scores.

## IV.Appendix

## Code

# Load data  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(janitor)

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

# Load csv file  
interview\_hiring <- read.csv("/cloud/project/Datasets/CandidateInterviewandHiring.csv")  
  
personal\_data <- read.csv("/cloud/project/Datasets/CandidatePersonalData.csv")  
  
technical\_scores <- read.csv("/cloud/project/Datasets/CandidateTechicalScores.csv")  
  
# first table of join is from 2 table interview\_hiring and perosonal\_data, by candidate.id  
two\_table <- interview\_hiring %>%  
 left\_join(personal\_data, by = "Candidate.ID")   
  
# second table of join - all\_table is join the first join table background with the third table technical\_score, so we can have all three joined table by candidate.id  
  
  
all\_table <- two\_table %>%  
 left\_join(technical\_scores, by = "Candidate.ID")   
  
# Filter hired candidate from all\_table  
hired\_candidates <- all\_table %>%  
 filter(Hired == "Yes")

## Data summary & Visualize

# Table 1: Summary statistics showing the relationship between graduating from elite college, education level and hiring outcome. Summary table show the average & median years of experience, average technical and interview score of candidates.   
  
summarise\_background <- all\_table %>%  
 group\_by(Hired, Graduated.from.Elite.College,Education.Level) %>%  
 summarise(avg\_year = mean(Years.of.Experience),  
 median\_experience = median(Years.of.Experience),  
 count = n(),  
 avg\_technical\_score = mean(Technical.Test.Score),  
 avg\_interview\_score = mean(Interview.Score))

## `summarise()` has grouped output by 'Hired', 'Graduated.from.Elite.College'.  
## You can override using the `.groups` argument.

summarise\_background

## # A tibble: 16 × 8  
## # Groups: Hired, Graduated.from.Elite.College [4]  
## Hired Graduated.from.Elite…¹ Education.Level avg\_year median\_experience count  
## <chr> <chr> <chr> <dbl> <dbl> <int>  
## 1 No No Bachelor's 4.83 5 595  
## 2 No No High School 5.10 5 559  
## 3 No No Master's 4.81 5 595  
## 4 No No PhD 4.99 5 604  
## 5 No Yes Bachelor's 4.90 5 231  
## 6 No Yes High School 5.46 6 225  
## 7 No Yes Master's 5.21 5 223  
## 8 No Yes PhD 5.00 5 222  
## 9 Yes No Bachelor's 6 6 54  
## 10 Yes No High School 6.16 6 37  
## 11 Yes No Master's 5.41 5 41  
## 12 Yes No PhD 5.51 5 39  
## 13 Yes Yes Bachelor's 6.33 7 15  
## 14 Yes Yes High School 6.12 6.5 24  
## 15 Yes Yes Master's 6.86 7 22  
## 16 Yes Yes PhD 7.29 8.5 14  
## # ℹ abbreviated name: ¹​Graduated.from.Elite.College  
## # ℹ 2 more variables: avg\_technical\_score <dbl>, avg\_interview\_score <dbl>

# Figure 1. Hiring outcome by education level and graduated from elite college.Bars show different of hiring outcomes if candidate graduated from elite college.  
  
ggplot(all\_table, aes(x = Education.Level, fill = Hired)) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Hiring Outcomes by Education Level",  
 x = "Education Level", y = "Count of Candidates") +  
 facet\_wrap(~ Graduated.from.Elite.College) +theme(axis.text.x = element\_text(angle = 45, hjust = 1))

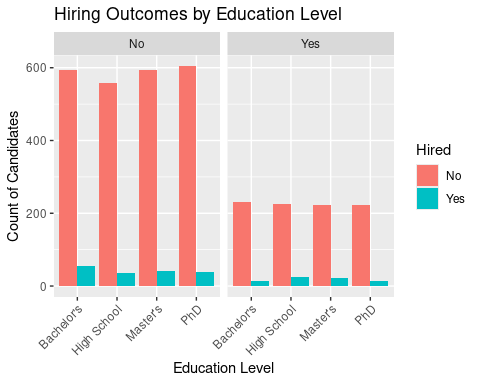


Figure 1: Education level vs hiring outcome.

# Figure 2: Technical test score & Interview Score by hiring status Violin of technical test score vs Interview score. show the relationship of those two, and if it related to the hiring outcomes.  
   
ggplot(all\_table, aes(x = Technical.Test.Score, y = Interview.Score, fill = Hired)) +  
 geom\_violin() +  
 labs(title = "Violin of Technical Test Score Faceted by Interview Score",  
 x = "Technical Test Score", y = "Interview Score")

## Warning: `position\_dodge()` requires non-overlapping x intervals.

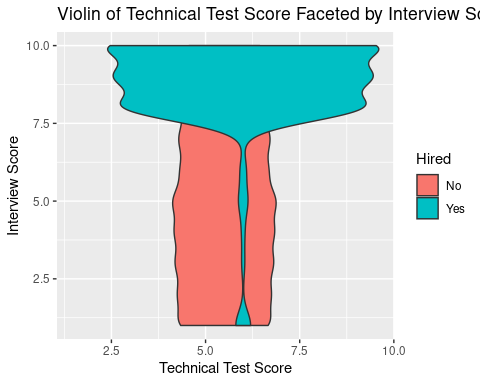


Figure 2: Technical test score, Interview score & Hiring outcome.

# Table all\_skill\_score have all technical test score, to see each specific technical score that a hired candidate and their roles.   
  
all\_skill\_score <- hired\_candidates %>%  
 group\_by(Position.Applied.For) %>%  
 pivot\_longer(cols = c(Python.Skill.Score,Java.Skill.Score, SQL.Skill.Score, R.Skill.Score, Machine.Learning.Skill.Score, HTML.Skill.Score, CSS.Skill.Score, JavaScript.Skill.Score),  
 names\_to = "Skill", values\_to = "Score") %>%  
 select(Candidate.ID, Position.Applied.For, Skill, Score)  
  
all\_skill\_score

## # A tibble: 1,968 × 4  
## # Groups: Position.Applied.For [5]  
## Candidate.ID Position.Applied.For Skill Score  
## <int> <chr> <chr> <int>  
## 1 35 System Analyst Python.Skill.Score 8  
## 2 35 System Analyst Java.Skill.Score 7  
## 3 35 System Analyst SQL.Skill.Score 10  
## 4 35 System Analyst R.Skill.Score 10  
## 5 35 System Analyst Machine.Learning.Skill.Score 4  
## 6 35 System Analyst HTML.Skill.Score 9  
## 7 35 System Analyst CSS.Skill.Score 2  
## 8 35 System Analyst JavaScript.Skill.Score 7  
## 9 79 System Analyst Python.Skill.Score 5  
## 10 79 System Analyst Java.Skill.Score 5  
## # ℹ 1,958 more rows

# Table 3: summary the average each skill of technical score group by position applied for.   
avg\_score <- all\_skill\_score %>%  
 group\_by(Position.Applied.For, Skill) %>%  
 summarise(avg\_score = mean(Score))

## `summarise()` has grouped output by 'Position.Applied.For'. You can override  
## using the `.groups` argument.

avg\_score

## # A tibble: 40 × 3  
## # Groups: Position.Applied.For [5]  
## Position.Applied.For Skill avg\_score  
## <chr> <chr> <dbl>  
## 1 Data Scientist CSS.Skill.Score 5.51  
## 2 Data Scientist HTML.Skill.Score 5.39  
## 3 Data Scientist Java.Skill.Score 5.92  
## 4 Data Scientist JavaScript.Skill.Score 5.69  
## 5 Data Scientist Machine.Learning.Skill.Score 5.29  
## 6 Data Scientist Python.Skill.Score 5.65  
## 7 Data Scientist R.Skill.Score 5.29  
## 8 Data Scientist SQL.Skill.Score 4.63  
## 9 Database Administrator CSS.Skill.Score 6.06  
## 10 Database Administrator HTML.Skill.Score 4.94  
## # ℹ 30 more rows

# Figure 3. Top 3 of the average of each skill based on the role.   
  
ggplot(avg\_score, aes(x = Position.Applied.For, y = avg\_score, color = Skill)) +  
 geom\_point(size = 3, position = position\_dodge(width = 0.5)) +  
 labs(  
 title = "Average Technical Skill Scores by Position Applied For",  
 x = "Position Applied For",  
 y = "Average Skill Score",  
 color = "Technical Skill"  
 ) +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

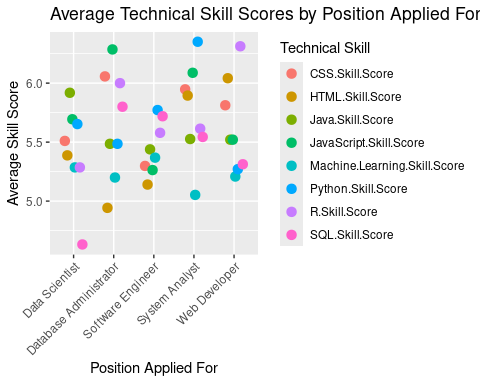


Figure 3: Top skills based on role.

# Table 4: Summary statistics of communication skills and hiring outcomes.Summary of communication skill levels of candidates who are hired.   
  
tabyl(all\_table, Hired,  
 Communication.Skill.Level)

## Hired Average Below Average Excellent Good Poor  
## No 662 621 659 685 627  
## Yes 35 57 54 50 50

# Figure 4: communication skills and hiring outcomes. (Bars show the amount of candidates who are hired or not based on their communication level)  
  
ggplot(all\_table, aes(y = Communication.Skill.Level, fill = Hired)) +  
geom\_bar(position = "fill") +  
ylab("Communication Level") +  
xlab("Proportion of Candidates") +  
labs(title = "Communication level of Candidates") +  
theme(axis.text.x = element\_text(angle = 0, hjust = 1))

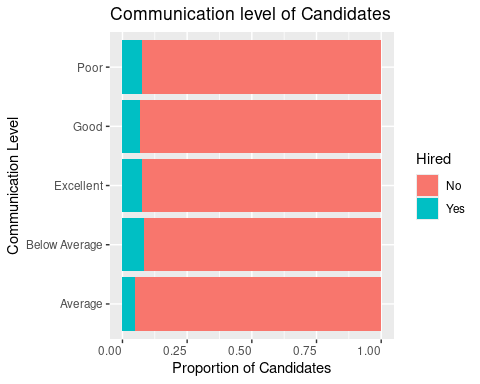


Figure 4 :Communication skill level & Hiring outcome.

# Table 5: Summary statistics of Candidates use digital presence profile. Summary of digital presence, show all candidates if they are hired, if they have linked profile or github profile.   
  
digital\_presence\_summary <- hired\_candidates %>%  
 count(LinkedIn.Profile, GitHub.Profile, Hired) %>%  
 arrange(desc(n))  
  
digital\_presence\_summary

## LinkedIn.Profile GitHub.Profile Hired n  
## 1 No No Yes 96  
## 2 No Yes Yes 72  
## 3 Yes No Yes 53  
## 4 Yes Yes Yes 25

#Age group table, group by age and count the hired to visualize figure.6.  
  
age\_group <- hired\_candidates %>%  
 group\_by(Age) %>%  
 summarise(count = n())

# Figure 6. Visualize bar of age by hiring outcome.   
  
ggplot(age\_group, aes(x = Age, y = count)) +  
 geom\_bar(stat = "identity", fill = "lightblue")+ # Stacked bar chart  
 labs(title = "Hiring Outcomes by Age",  
 x = "Age",  
 y = "Hired Candidates")

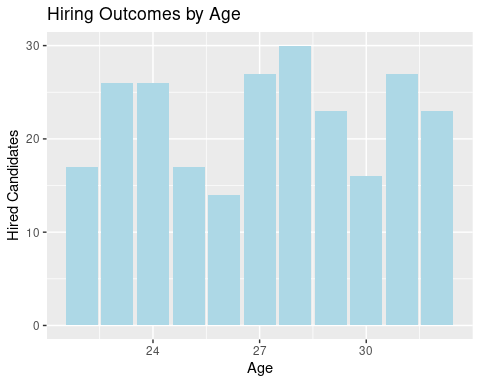


Figure 6. Age and Hiring outcome.