Skills Gap Analysis

# Skills Gap Analysis

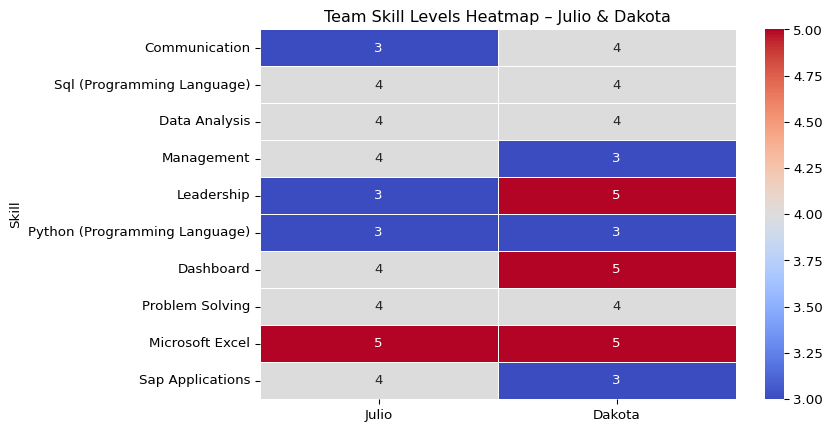
from pyspark.sql import SparkSession  
  
  
# Start a Spark session  
spark = SparkSession.builder.appName("JobPostingsAnalysis").getOrCreate()  
  
# Load the CSV file into a Spark DataFrame  
df = spark.read.option("header", "true").option("inferSchema", "true").option("multiLine","true").option("escape", "\"").csv("../data/lightcast\_job\_postings.csv")

WARNING: Using incubator modules: jdk.incubator.vector  
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties  
Setting default log level to "WARN".  
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).  
25/10/16 04:25:22 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
[Stage 1:> (0 + 1) / 1]

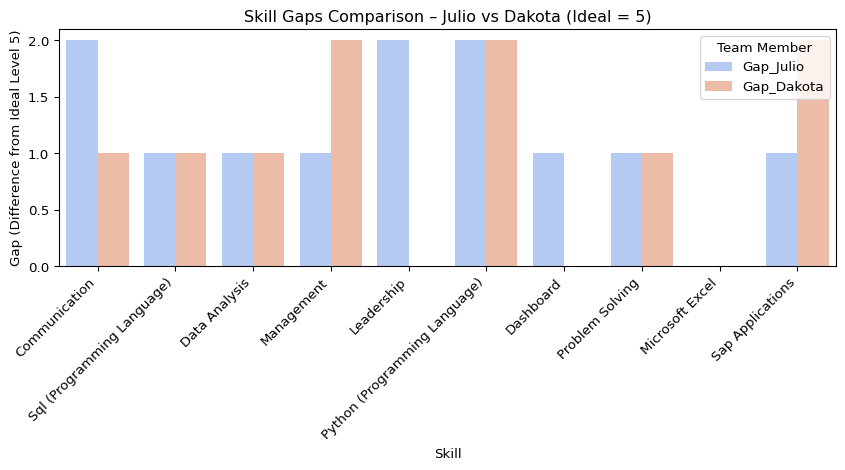
import pandas as pd  
  
columns = [  
 "COMPANY\_NAME", "COMPANY\_IS\_STAFFING", # Identification, company  
 "POSTED", "EXPIRED", "DURATION", "MODELED\_DURATION", # Dates, duration  
 "TITLE\_NAME", "EMPLOYMENT\_TYPE\_NAME", "IS\_INTERNSHIP", # Job title, contract type  
 "CITY\_NAME", "STATE\_NAME", "REMOTE\_TYPE\_NAME", # Geographic  
 "MIN\_YEARS\_EXPERIENCE", "MIN\_EDULEVELS\_NAME", "EDUCATION\_LEVELS\_NAME", # Education, experience  
 "SALARY", # Salary  
 "SKILLS\_NAME", "SPECIALIZED\_SKILLS\_NAME", "SOFTWARE\_SKILLS\_NAME", # Tech skills  
 "COMMON\_SKILLS\_NAME", # Common, soft skills  
 "CERTIFICATIONS\_NAME" # Certif  
]  
  
  
df\_columns = df.limit(4000).select([c for c in columns if c in df.columns])  
df\_columns\_pd=df\_columns.toPandas()  
  
skill\_cols = [  
"SKILLS\_NAME",  
"SPECIALIZED\_SKILLS\_NAME",  
"SOFTWARE\_SKILLS\_NAME",  
"COMMON\_SKILLS\_NAME",  
"CERTIFICATIONS\_NAME"  
]  
  
df\_columns\_pd["ALL\_SKILLS\_RAW"] = df\_columns\_pd[skill\_cols].fillna("").agg(" ".join, axis=1)  
df\_columns\_pd["ALL\_SKILLS\_RAW"] = (  
 df\_columns\_pd[skill\_cols]  
 .fillna("")  
 .agg(" ".join, axis=1)  
 .astype(str)  
 .str.replace(r"[{}\[\]'\"]", "", regex=True) # quita corchetes y comillas  
 .str.replace(r"\b[Nn]one\b|nan", "", regex=True) # quita None/nan  
 .str.replace(r"[;|/]", ",", regex=True) # normaliza separadores  
 .str.replace(r"\s\*,\s\*", ", ", regex=True) # limpia espacios entre comas  
 .str.replace(r"\s{2,}", " ", regex=True) # elimina espacios dobles  
 .str.strip() # quita espacios extra  
)  
print(df\_columns\_pd["ALL\_SKILLS\_RAW"].head(5))  
pd.set\_option('display.max\_colwidth', None); print(df\_columns\_pd["ALL\_SKILLS\_RAW"].head(5).to\_string(index=False))  
  
  
  
# skills text to list.  
df\_columns\_pd["ALL\_SKILLS\_LIST"] = df\_columns\_pd["ALL\_SKILLS\_RAW"].str.split(",")  
  
# list to row  
df\_skills = df\_columns\_pd.explode("ALL\_SKILLS\_LIST")  
  
# Clean up spaces and drop empty rows  
df\_skills = ( df\_skills.dropna(subset=["ALL\_SKILLS\_LIST"]).loc[df\_skills["ALL\_SKILLS\_LIST"].str.strip() != ""])  
df\_skills["ALL\_SKILLS\_LIST"] = ( df\_skills["ALL\_SKILLS\_LIST"] .str.strip() .str.title())  
  
# Count skills  
top\_skills = (df\_skills["ALL\_SKILLS\_LIST"].value\_counts().reset\_index().rename(columns={"index": "Skill", "ALL\_SKILLS\_LIST": "Frequency"}))  
  
# Show top 20  
print(top\_skills.head(20))  
  
# Frequency count  
# 0 Communication 3394  
# 1 Sql (Programming Language) 3134  
# 2 Data Analysis 2960  
# 3 Management 2116  
# 4 Leadership 2023  
# 5 Python (Programming Language) 1837  
# 6 Dashboard 1791  
# 7 Problem Solving 1788  
# 8 Microsoft Excel 1771  
# 9 Sap Applications 1658  
# 10 Operations 1550  
# 11 Project Management 1528  
# 12 Business Process 1484  
  
  
  
# 5 expert, 4 Advanced, 3 Intermediate, 2 Basic Knowledge, 1  
top10\_skills = [  
 "Communication",  
 "Sql (Programming Language)",  
 "Data Analysis",  
 "Management",  
 "Leadership",  
 "Python (Programming Language)",  
 "Dashboard",  
 "Problem Solving",  
 "Microsoft Excel",  
 "Sap Applications"  
]  
  
skills\_data = {  
 "Skill": top10\_skills,  
 "Julio": [3, 4, 4, 4, 3, 3, 4, 4, 5, 4],  
 "Dakota": [4, 4, 4, 3, 5, 3, 5, 4, 5, 3]  
}  
  
df\_team = pd.DataFrame(skills\_data)  
#df\_team  
  
  
import seaborn as sns  
import matplotlib.pyplot as plt  
  
plt.figure(figsize=(8,5))  
sns.heatmap(df\_team.set\_index("Skill"), annot=True, cmap="coolwarm", linewidths=0.5)  
plt.title("Team Skill Levels Heatmap – Julio & Dakota")  
plt.show()  
  
# SKILL GAP (Ideal vs our skills)  
df\_team["Gap\_Julio"] = 5 - df\_team["Julio"]  
df\_team["Gap\_Dakota"] = 5 - df\_team["Dakota"]  
  
# SHOT SKILLS  
print(df\_team[["Skill", "Julio", "Dakota", "Gap\_Julio", "Gap\_Dakota"]])  
  
  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
#Melt AND PLOT GAP  
df\_gaps\_melted = df\_team.melt(  
 id\_vars="Skill",  
 value\_vars=["Gap\_Julio", "Gap\_Dakota"],  
 var\_name="Member",  
 value\_name="Gap"  
)  
  
plt.figure(figsize=(9,5))  
sns.barplot(data=df\_gaps\_melted, x="Skill", y="Gap", hue="Member", palette="coolwarm")  
plt.title("Skill Gaps Comparison – Julio vs Dakota (Ideal = 5)")  
plt.xticks(rotation=45, ha="right")  
plt.ylabel("Gap (Difference from Ideal Level 5)")  
plt.xlabel("Skill")  
plt.legend(title="Team Member")  
plt.tight\_layout()  
plt.show()

[Stage 2:> (0 + 1) / 1]

0 Merchandising, Mathematics, Presentations, Pre...  
1 Procurement, Ficial Statements, Oracle Busines...  
2 Management, Exception Reporting, Report Writin...  
3 Exit Strategies, Reliability, User Story, Mana...  
4   
Name: ALL\_SKILLS\_RAW, dtype: object  
 Merchandising, Mathematics, Presentations, Predictive Modeling, Data Modeling, Advanced Analytics, Data Extraction, Statistical Analysis, Data Mining, Business Analysis, Fice, Algorithms, Statistics, SQL (Programming Language), Report Writing, Ad Hoc Reporting, Power BI, Relationship Building, Economics, Business Administration Merchandising, Predictive Modeling, Data Modeling, Advanced Analytics, Data Extraction, Statistical Analysis, Data Mining, Business Analysis, Fice, Algorithms, Statistics, SQL (Programming Language), Ad Hoc Reporting, Power BI, Economics SQL (Programming Language), Power BI Mathematics, Presentations, Report Writing, Relationship Building, Business Administration  
 Procurement, Ficial Statements, Oracle Business Intelligence (BI), OBIA, Oracle E-Business Suite, PL, SQL, Supply Chain, Business Intelligence, Oracle Fusion Middleware, Project Accounting Procurement, Ficial Statements, Oracle Business Intelligence (BI), OBIA, Oracle E-Business Suite, PL, SQL, Supply Chain, Business Intelligence, Oracle Fusion Middleware, Project Accounting Oracle Business Intelligence (BI), OBIA, Oracle E-Business Suite, PL, SQL, Oracle Fusion Middleware  
 Management, Exception Reporting, Report Writing, Security Clearance, Interpersonal Communications, Ability To Meet Deadlines, Presentations, Writing, Data Analysis, Organizational Skills, Negotiation, Data Integrity, Microsoft Office Exception Reporting, Data Analysis, Data Integrity Microsoft Office Management, Report Writing, Interpersonal Communications, Ability To Meet Deadlines, Presentations, Writing, Organizational Skills, Negotiation, Microsoft Office Security Clearance  
Exit Strategies, Reliability, User Story, Management, Strategic Planning, Hardware Configuration Management, On Prem, Agile Methodology, Solution Design, Advanced Analytics, Reengineering, Safety Assurance, Cross-Functional Collaboration, Requirements Elicitation, Business Analysis, Data Management, Data Architecture, Influencing Skills, Market Trend, Business Valuation, Creativity, Innovation, Goverce, Systems Development Life Cycle, Leadership, Test Planning, Multi-Tet Cloud Environments, Scrum (Software Development), Project Management, Operations, Data Migration, Regulatory Compliance, Product Roadmaps, SAS (Software), Troubleshooting (Problem Solving), Quality Assurance, Software As A Service (SaaS), Data Domain, Product Requirements, Data Goverce, Competitive Intelligence, Operations Architecture, Risk Appetite, Google Cloud Platform (GCP), User Feedback Exit Strategies, User Story, Hardware Configuration Management, On Prem, Agile Methodology, Solution Design, Advanced Analytics, Reengineering, Cross-Functional Collaboration, Requirements Elicitation, Business Analysis, Data Management, Data Architecture, Market Trend, Business Valuation, Systems Development Life Cycle, Test Planning, Multi-Tet Cloud Environments, Scrum (Software Development), Project Management, Data Migration, Regulatory Compliance, Product Roadmaps, SAS (Software), Software As A Service (SaaS), Data Domain, Product Requirements, Data Goverce, Competitive Intelligence, Operations Architecture, Risk Appetite, Google Cloud Platform (GCP), User Feedback SAS (Software), Google Cloud Platform (GCP) Reliability, Management, Strategic Planning, Safety Assurance, Influencing Skills, Creativity, Innovation, Goverce, Leadership, Operations, Troubleshooting (Problem Solving), Quality Assurance  
   
 Frequency count  
0 Communication 3394  
1 Sql (Programming Language) 3134  
2 Data Analysis 2960  
3 Management 2116  
4 Leadership 2023  
5 Python (Programming Language) 1837  
6 Dashboard 1791  
7 Problem Solving 1788  
8 Microsoft Excel 1771  
9 Sap Applications 1658  
10 Operations 1550  
11 Project Management 1528  
12 Business Process 1484  
13 Fice 1437  
14 Business Requirements 1415  
15 Planning 1211  
16 Presentations 1141  
17 Writing 1120  
18 Detail Oriented 1118  
19 Tableau (Business Intelligence Software) 1116



Skill Julio Dakota Gap\_Julio Gap\_Dakota  
0 Communication 3 4 2 1  
1 Sql (Programming Language) 4 4 1 1  
2 Data Analysis 4 4 1 1  
3 Management 4 3 1 2  
4 Leadership 3 5 2 0  
5 Python (Programming Language) 3 3 2 2  
6 Dashboard 4 5 1 0  
7 Problem Solving 4 4 1 1  
8 Microsoft Excel 5 5 0 0  
9 Sap Applications 4 3 1 2



# Key Takeaways

Julio: Focus on Communication, Leadership, Python, and Management to reduce gaps

Dakota: Focus on Python and Management. Leadership and Dashboard are strong.

## Recommended Actions

Technical Skills: SQL, Python, Dashboarding; Codecademy, DataCamp, Khan Academy, Coursera

SAP Applications: SAP Learning Hub, openSAP Courses

Soft Skills: Communication: LinkedIn Learning, Coursework Leadership and Management: Coursera Management courses

## Team Collaboartion to Bridge Skill Gaps

Cross Training among team members on strong skills, team projects to exercise the weaker skills, and track progress.