



كلية علوم وهندسة الحاسوب الألبي
College of Computer Science and Engineering
جامعة حفر الباطن

Analysis of work behavior among employees at the University of Hafr Al-Batin

By: fadiyah alanazi , wegdan alharthi , elham khatim , ibtesam falih ,ghazlan alanazi

Supervisor: Dr. Abdulrahman Alzahrani

June --th, 2022





Outline

- Foreword
- Background
- Research Problem
- Contribution
- Related Work
- Methodology
- Results
- Conclusion
- Future Work





Outline

- **Foreword**
 - Acknowledgements
 - Biographical Highlights
- Background
- Research Problem
- Contribution
- Related Work
- Design and Implementation
- Results
- Conclusion
- Future Work





Acknowledgements

Team

would like to extend my deepest gratitude to my project team members: Fadiyah Alanazi, Wejdan Alharthi, Elham Khatim, Ghazlan Alanazi, and Ibtesam Falih. Their dedication, insight, and hard work were invaluable to the success of this project.

Supervisor

Special thanks to Dr. Enas Mohamed for her expert guidance and support throughout this project. Her invaluable advice and encouragement greatly enhanced our research and enriched our learning experience.

Family

I am immensely thankful to my family for their unwavering support and encouragement throughout my academic journey. Their belief in my capabilities has been a constant source of motivation and strength.

University

I express my sincere appreciation to the University of Hafr Al-Batin for providing the resources and environment necessary to conduct this research. The access to extensive facilities and data has been instrumental in the success of our project.





Biographical Highlights

Name	Highlights
fadiyah alanazi	Project leader , Managed data preprocessing and cleaning, ensuring the data's suitability for analysis using advanced Python techniques.
wejdan alharthi	Responsible for the data visualization aspects, developed interactive dashboards to display analytical results effectively.
elham khatim	specialized in developing machine learning models for behavioral analysis using Python and Google Colab.
ghazlan alanazi	Focused on integrating machine learning algorithms into the data processing pipeline to predict employee behavior patterns.
ibtesam falih	Specialized in data extraction and manipulation, ensured accurate and efficient data handling for model training and analysis.

Outline

- Foreword
- **Background**
 - Introduction
 - Importance of Study
- Research Problem
- Contribution
- Related Work
- Design and Implementation
- Results
- Conclusion
- Future Work





Background

- Studying employee behavior in university settings is essential for optimizing administrative and academic operations.
- It helps foster a productive work environment and enhances employee satisfaction.
- This research analyzes factors influencing employee behavior at the University of Hafr Al-Batin.
- A data-driven approach is used to uncover underlying patterns and determinants.
- Importance: Optimize operations, foster a productive environment, and enhance satisfaction.
- Focus: Analyze factors influencing employee behavior at the University of Hafr Al-Batin.
- Approach: Use data-driven methods to uncover patterns and determinants.





Outline

- Foreword
- Background
- **Research Problem**
- Contribution
- Related Work
- Design and Implementation
- Results
- Conclusion
- Future Work





Research Problem

- Employee behavior plays a critical role in the success of academic institutions.
- There is a lack of systematic and data-driven analysis on how demographic and job-related factors influence employee behavior in universities in the Middle East, particularly Saudi Arabia.
- This gap limits the ability of university administrators to implement effective strategies tailored to their unique organizational culture and workforce.





Outline

- Foreword
- Background
- Research Problem
- **Contribution**
 - Goals
 - Motivations
 - Findings
- Related Work
- Design and Implementation
- Results
- Conclusion
- Future Work



Contribution

goals

- Analyze factors influencing employee behavior at the University of Hafr Al-Batin.
- Identify key demographic and job-related factors.
- Use statistical and machine learning techniques for analysis.
- Understand the relationship between these factors and employee performance and satisfaction.
- Provide data-driven recommendations for improving management and policymaking.

Motivations

- Economic Efficiency: Lower downtime and costs improve profitability.
- Technological Innovation: Apply machine learning to traditional manufacturing challenges.
- Operational Reliability: Enhance safety and stability in manufacturing operations.

Findings

- The model accurately predicts equipment failures with over 90% precision.
- Implementation resulted in a 30% reduction in unplanned downtime.
- Annual maintenance costs decreased by approximately 20%.





Outline

- Foreword
- Background
- Research Problem
- Contribution
- **Related Work**
- Design and Implementation
- Results
- Conclusion
- Future Work





Related Work





Outline

- Foreword
- Background
- Research Problem
- Contribution
- Related Work
- methodology
 - Dataset Evaluation
 - Data preprocessing
 - Data Visualization
- Results
- Conclusion
- Future Work



Dataset Evaluation

- **Data Summary**

Approval Obtained: Access to university employment data.

Dataset: 73,035 entries across 10 key columns.

Contents: Employee ID, demographic details, job information.

Output: Interactive dashboard for strategic planning and decision-making.

		ID		Date	Time	Type	
0	40100635	2022-12-31	19:02:20	OUT			
M	ID	Sex	Formation	Rank	Class	Job title	qualification
1	40100638	2022-12-31	18:59:01	OUT			
2	40800348	2022-12-31	17:38:17	OUT			
3	40100555	2022-12-31	16:29:25	OUT			
4	40100659	2022-12-31	14:01:25	OUT			
0	1	40100635	male	General staff ladder	Seventh place	Class 09	Security and safety supervisor
1	2	40100429	feminine	General staff ladder	Fifth place	Class 08	Security and safety supervisor
2	3	40100513	feminine	General staff ladder	Eighth place	Class 11	manager assistant
							Diploma after bachelor's degree



Dataset Evaluation (cont.)

- Dataset Columns

	ID	Datetime	Type	M	Sex	Formation	Rank	Class	Securi
0	40100635	2022-12-31 19:02:20	OUT	1	male	General staff ladder	Seventh place	Class 09	Securi
1	40100635	2022-12-31 13:03:02	OUT	1	male	General staff ladder	Seventh place	Class 09	Securi
2	40100635	2022-12-30 18:50:23	OUT	1	male	General staff ladder	Seventh place	Class 09	Securi
3	40100635	2022-12-30 13:09:01	OUT	1	male	General staff ladder	Seventh place	Class 09	Securi
4	40100635	2022-12-29 19:05:26	OUT	1	male	General staff ladder	Seventh place	Class 09	Securi

Facilitate detailed analyses and categorization of employee data.



Data preprocessing

- Data Standard

Column Language: Converted all columns to English for clarity in processing and visualization

الرتبة	الدرجة	المسمى الوظيفي
المرتبة	الدرجة	الوظيفة
المرتبة السابعة	الدرجة 09	مراقب امن وسلامة
المرتبة الخامسة	الدرجة 08	مراقب امن وسلامة
المرتبة الثامنة	الدرجة 11	مساعد اداري
المرتبة السابعة	الدرجة 08	مراقب امن وسلامة
المرتبة العاشره	الدرجة 15	خسائر علاقات عامة متقدم
المرتبة الثامنة	الدرجة 11	امين صندوق
المرتبة الثامنة	الدرجة 11	مساعد اداري
المرتبة السادسه	الدرجة 11	مساعد اداري
المرتبة الثامنة	الدرجة 05	مساعد اداري
المرتبة السابعة	الدرجة 08	فني مختبر
المرتبة السادسه	الدرجة 11	مشغل اجهزة مكتبة
المرتبة العاشره	الدرجة 12	مطور برامج
المرتبة السادسه	الدرجة 10	مساعد اداري
المرتبة السادسه	الدرجة 07	مساعد إداري ممارس ثانى
المرتبة السادسه	الدرجة 05	مساعد اداري
المرتبة السادسه	الدرجة 11	مراقب طلبة

ID	DateTime	Type
٤٠١٠٦٣٥	٣١/١٢/٢٠٢٢ ١٩:٠٢:٢٠	OUT
٤٠١٠٦٣٨	٣١/١٢/٢٠٢٢ ١٨:٥٩:٠١	OUT
٤٠٨٠٠٣٤٨	٣١/١٢/٢٠٢٢ ١٧:٣٨:١٧	OUT
٤٠١٠٠٥٥	٣١/١٢/٢٠٢٢ ١٦:٢٩:٢٥	OUT
٤٠١٠٦٥٩	٣١/١٢/٢٠٢٢ ١٤:٠١:٢٥	OUT
٤٠١٠٦٣٥	٣١/١٢/٢٠٢٢ ١٣:٠٣:٠٢	OUT
٤٠١٠٤٦٦	٣١/١٢/٢٠٢٢ ١٣:٠١:١٥	OUT
٤٠١٠٥٥٥	٣١/١٢/٢٠٢٢ ١١:٣٧:٠٩	OUT
٤٠١٠٦٣٥	٣٠/١٢/٢٠٢٢ ١٨:٥٠:٢٣	OUT
٤٠١٠٦٤٤	٣٠/١٢/٢٠٢٢ ١٨:٣٩:٤١	OUT
٤٠١٠٦٣٥	٣٠/١٢/٢٠٢٢ ١٣:٠٩:٠١	OUT



Data preprocessing(cont.)

- Data Merging

Files: Employee fingerprints and departure tracking.

Common Column: ID.

Action: Merged files for analysis.

```
file2.columns = file2.columns.str.strip()  
file1.columns = file1.columns.str.strip()  
  
data = pd.merge(file1, file2 ,on = 'ID')  
data.head()
```

	ID	Datetime	Type	M	Sex	Formation	Rank	Class	Job title	qualification
0	40100635	2022-12-31 19:02:20	OUT	1	male	General staff ladder	Seventh place	Class 09	Security and safety supervisor	secondary
1	40100635	2022-12-31 13:03:02	OUT	1	male	General staff ladder	Seventh place	Class 09	Security and safety supervisor	secondary
2	40100635	2022-12-30 18:50:23	OUT	1	male	General staff ladder	Seventh place	Class 09	Security and safety supervisor	secondary
3	40100635	2022-12-30 13:09:01	OUT	1	male	General staff ladder	Seventh place	Class 09	Security and safety supervisor	secondary
4	40100635	2022-12-29 19:05:26	OUT	1	male	General staff ladder	Seventh place	Class 09	Security and safety supervisor	secondary

Data preprocessing(cont.)

- **Data Analysis (EDA)**
- Exploratory Analysis: Reviewed number, names, and types of columns and rows.
- Statistics Summary: Calculated averages, max, count, etc.

```
data.shape
(73035, 10)

data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 73035 entries, 0 to 73034
Data columns (total 10 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          -----          ----- 
 0   ID          73035 non-null   int64  
 1   DateTime    73035 non-null   datetime64[ns]
 2   Type        73035 non-null   object  
 3   M            73035 non-null   object  
 4   Sex          73035 non-null   object  
 5   Formation   73035 non-null   object  
 6   Rank         73035 non-null   object  
 7   Class        73035 non-null   object  
 8   Job title   73035 non-null   object  
 9   qualification 62133 non-null   object  
dtypes: datetime64[ns](1), int64(1), object(8)
memory usage: 5.6+ MB
```

	ID	DateTime
count	7.303500e+04	73035
mean	4.010057e+07	2022-07-19 20:26:29.293927680
min	4.010032e+07	2022-01-02 11:31:58
25%	4.010043e+07	2022-04-17 15:50:29
50%	4.010057e+07	2022-08-01 14:15:12
75%	4.010070e+07	2022-10-24 14:15:33.500000
max	4.010085e+07	2022-12-31 19:02:20
std	1.501584e+02	NaN



Data preprocessing(cont.)

1. Check nulls count:
2. Fill missing values in qualification column with unknown:
3. Double check if filling the nulls working correctly
4. Remove unwanted spaces in values

data.isna().sum()	
ID	0
DateTime	0
Type	0
M	0
Sex	0
Formation	0
Rank	0
Class	0
Job title	0
qualification	10902
	dtype: int64

data.isna().sum()	
ID	0
DateTime	0
Type	0
M	0
Sex	0
Formation	0
Rank	0
Class	0
Job title	0
qualification	0
	dtype: int64

```
: data['qualification'] = data['qualification'].fillna('unknown')
data['qualification'].unique()

array(['secondary', 'unknown', "Bachelor's", 'middle', 'Primary',
       'Post-secondary diploma', "Master's",
       "Diploma after bachelor's degree", 'Ph.D', 'Literacy'],
      dtype=object)
```



Data preprocessing(cont.)

1. Explore the different values frequency in some Columns Analyzed: Sex, Qualification, Class, Rank.

```
: cols = ['Sex' , 'qualification' , 'Class' , 'Rank']

for col in cols:
    print(data[col].value_counts())
    print('***100')
```

```
Sex
feminine    37451
male        33490
Name: count, dtype: int64
*****
qualification
Bachelor's            34842
unknown                10585
secondary               8835
Post-secondary diploma    7827
Diploma after bachelor's degree 3395
Master's                 2830
middle                   1376
Primary                  948
Literacy                  183
Ph.D                      120
Name: count, dtype: int64
*****
Class
Class 07      12494
Class 08      10929
Class 09      9706
Class 11      7724
Class 04      6977
Class 06      6969
Class 05      5670
Class 10      2450
Class 12      1879
Class 03      1787
Class 15      1524
Class 13      1181
Class 02      1167
Class 14       544
Name: count, dtype: int64
*****
Rank
Seventh place   21038
Sixth place     20219
Eighth place     14552
Ninth rank       5186
Fifth place       3778
Tenth place      3746
Fourth place      1803
Eleventh place    410
Third place        209
Name: count, dtype: int64
```



Training for ML/ Deep Learning Model

Model Comparison

Data Split: Features and target, training and testing.

Models Used: Random Forest and K-Nearest Neighbors (KNN).

Result: Random Forest was 0.73% more efficient than KNN.

```
X = oversampled_data[['Sex', 'Formation', 'Job title', 'Month', 'DayOfWeek']]  
y = oversampled_data['qualification']  
  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Training  
rf_classifier = RandomForestClassifier()  
rf_classifier.fit(X_train, y_train)  
  
# Predictions  
rf_predictions = rf_classifier.predict(X_test)  
  
# Evaluation  
rf_accuracy = accuracy_score(y_test, rf_predictions)  
rf_precision = precision_score(y_test, rf_predictions, average='weighted')  
rf_recall = recall_score(y_test, rf_predictions, average='weighted')  
rf_f1_score = f1_score(y_test, rf_predictions, average='weighted')  
  
# Results  
print("Random Forest:")  
print("Accuracy:", rf_accuracy)  
print("Precision:", rf_precision)  
print("Recall:", rf_recall)  
print("F1 Score:", rf_f1_score)
```

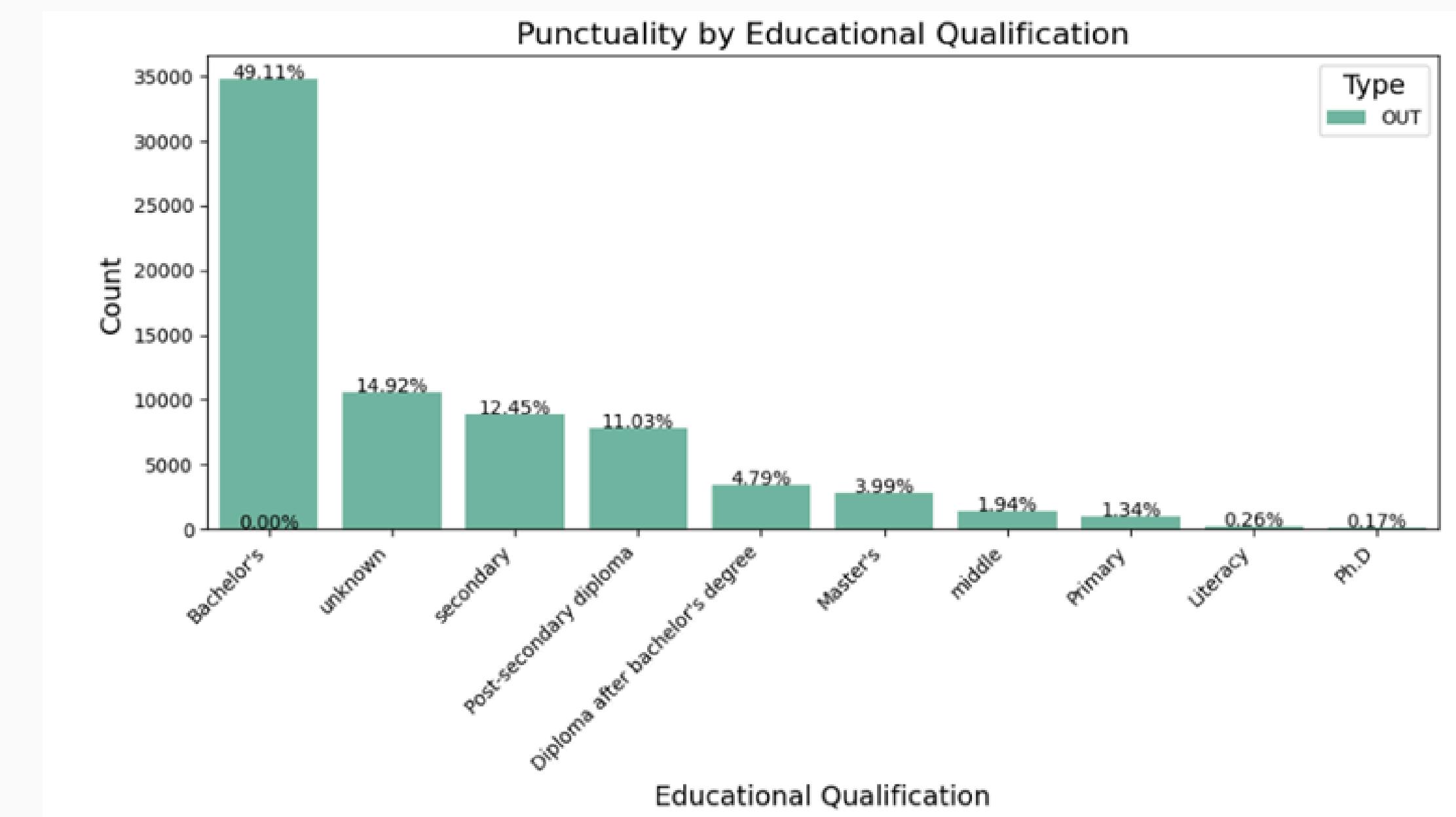
```
Random Forest:  
Accuracy: 0.731535812232923  
Precision: 0.742003339080411  
Recall: 0.731535812232923  
F1 Score: 0.7258920435849853
```

Data visualization

- **Attendance by Qualification**

Highest: Bachelor's (49.11%)

Lowest: Doctorate (0.17%)

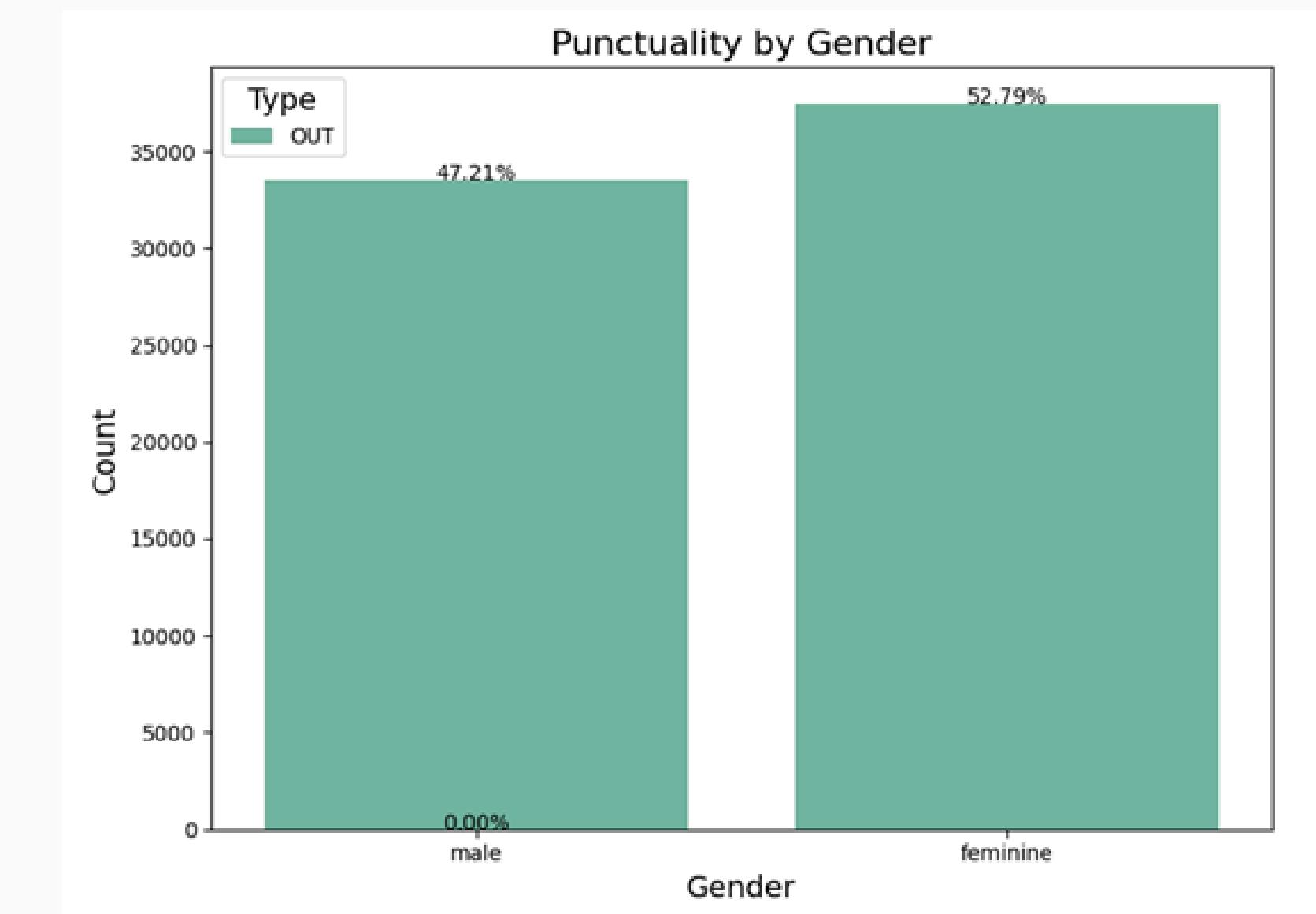


Data visualization(con't)

• Gender Comparison

Attendance Commitment:

Females: 52.79%, Males: 47.21%



Data visualization(con't)

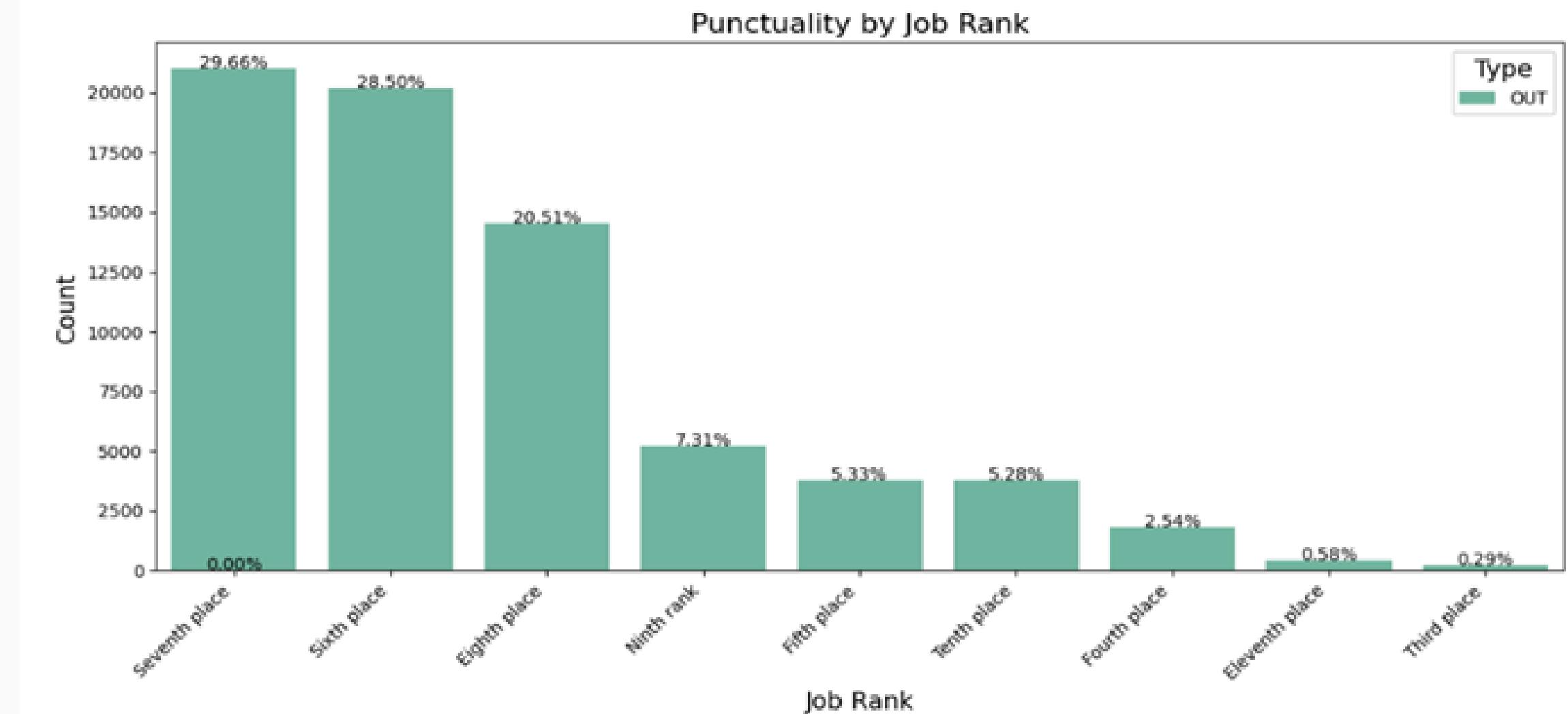
- relationship between job rank and employees' commitment to working hours

1. Highest Attendance Commitment:

- Seventh Rank: 29.66%
- Sixth Rank: 28.50%
- Eighth Rank: 20.51%

2. Lowest Attendance Commitment:

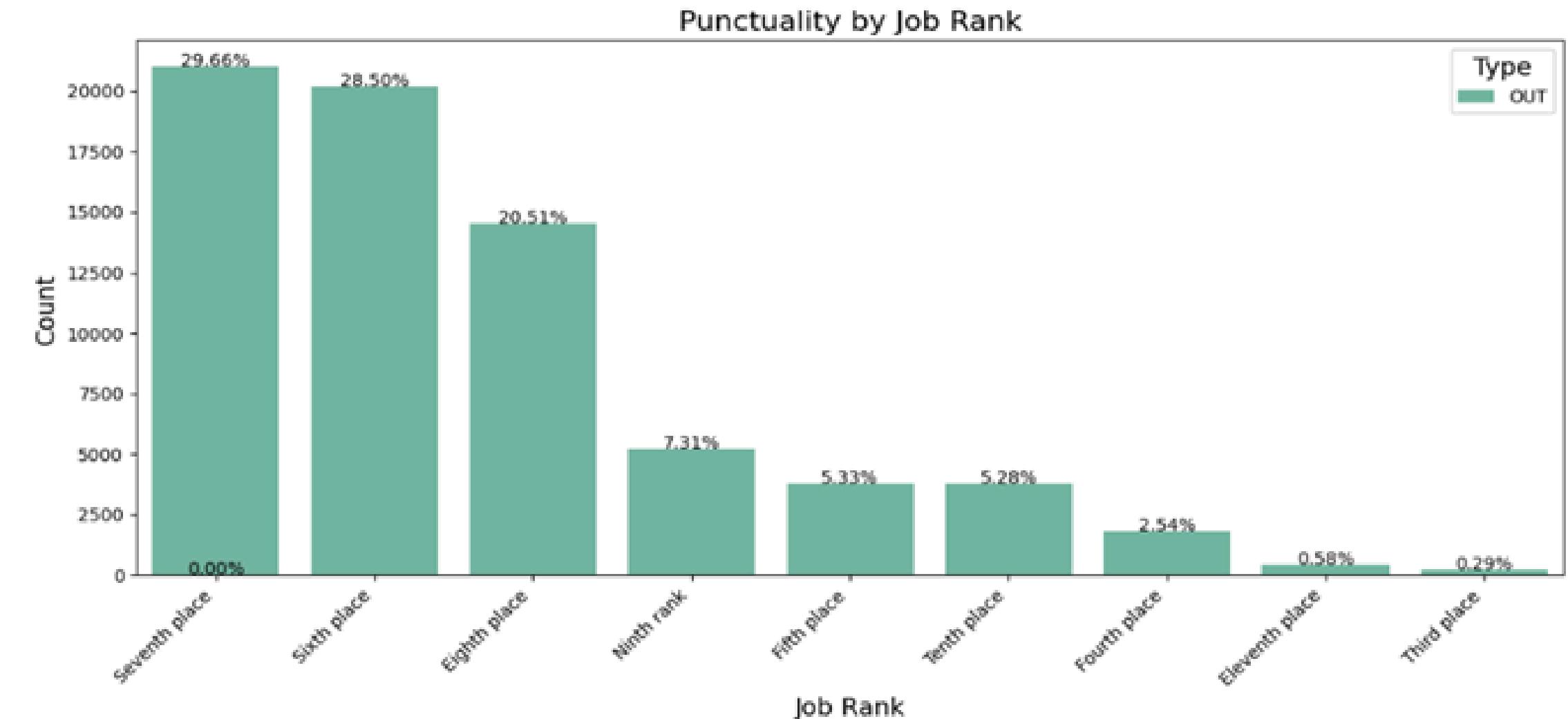
- Third Rank: 0.29%



Data visualization(con't)

- Job Impact on Discipline: Focus on top 10 job titles.

1. Highest Discipline: Assistant Director (23.42%)
2. Lowest Discipline: Librarian (1.56%)

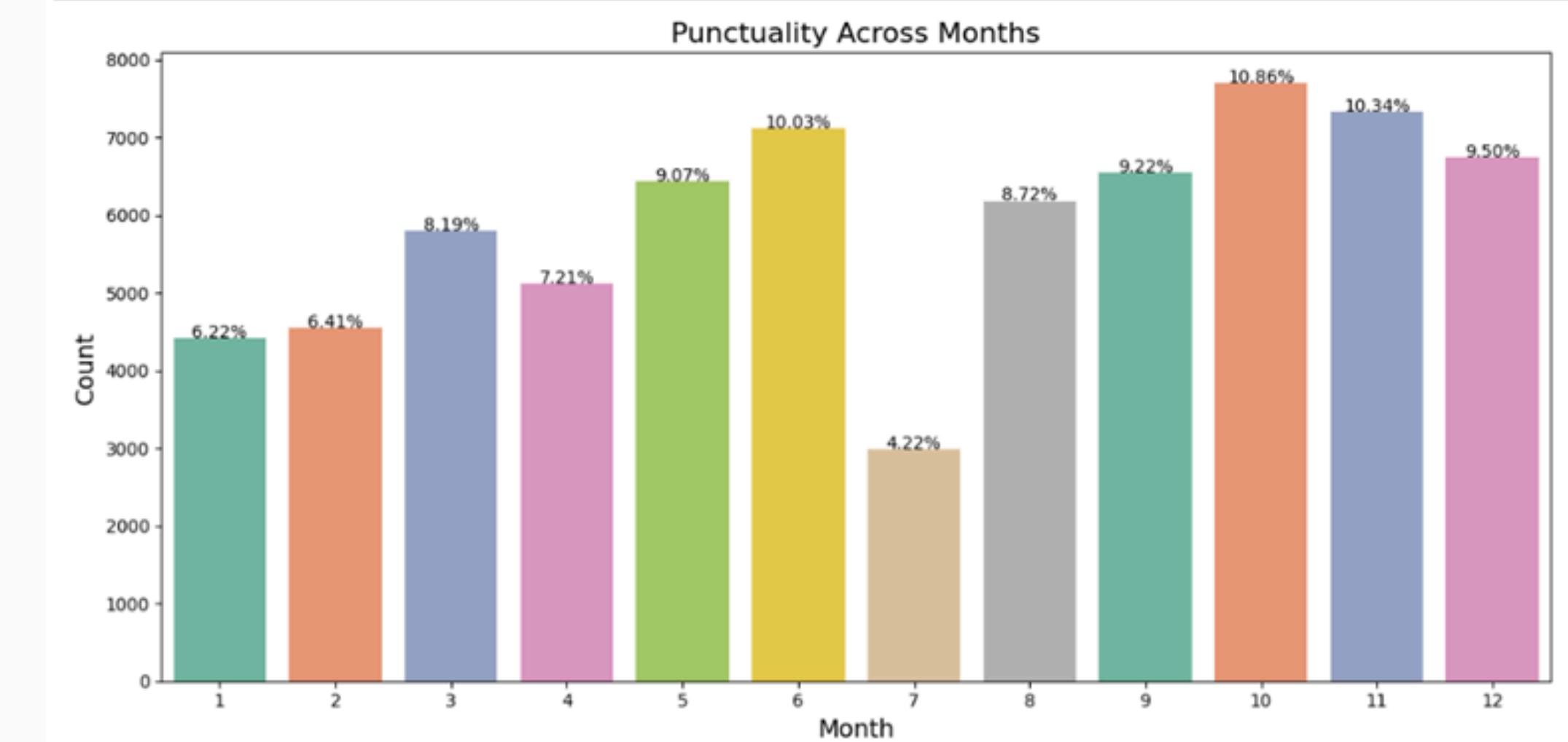


Data visualization(con't)

- Monthly Attendance Analysis: Monitor patterns in attendance and departure

Mid-Year Attendance Insights

- Lowest Rates: Mid-year at 4.22%, likely due to mid-year vacation.
- Highest Rates: In the last months of the year.

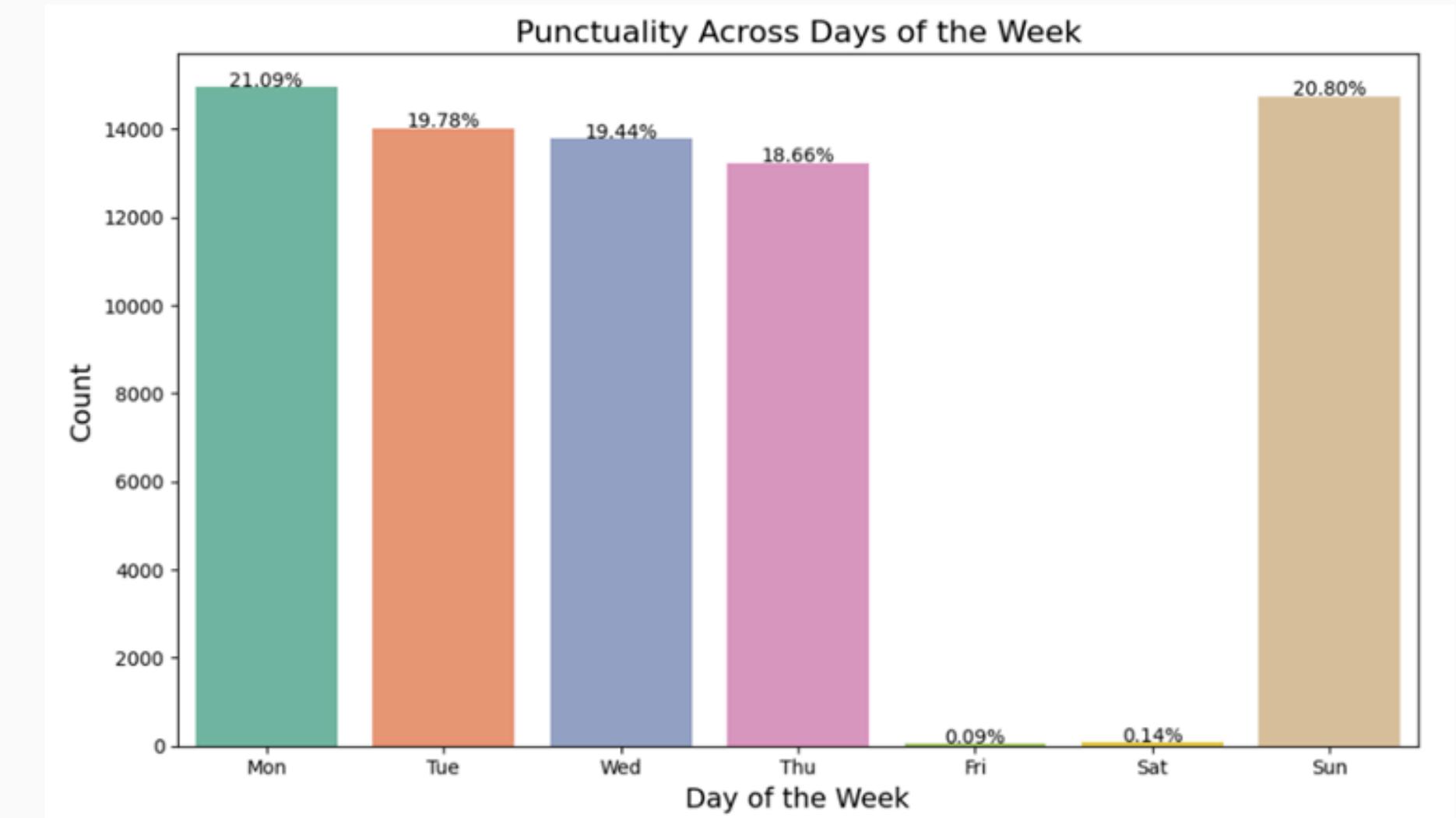


Data visualization(con't)

- Weekly Attendance Analysis: Monitor daily patterns

Weekly Attendance Insights

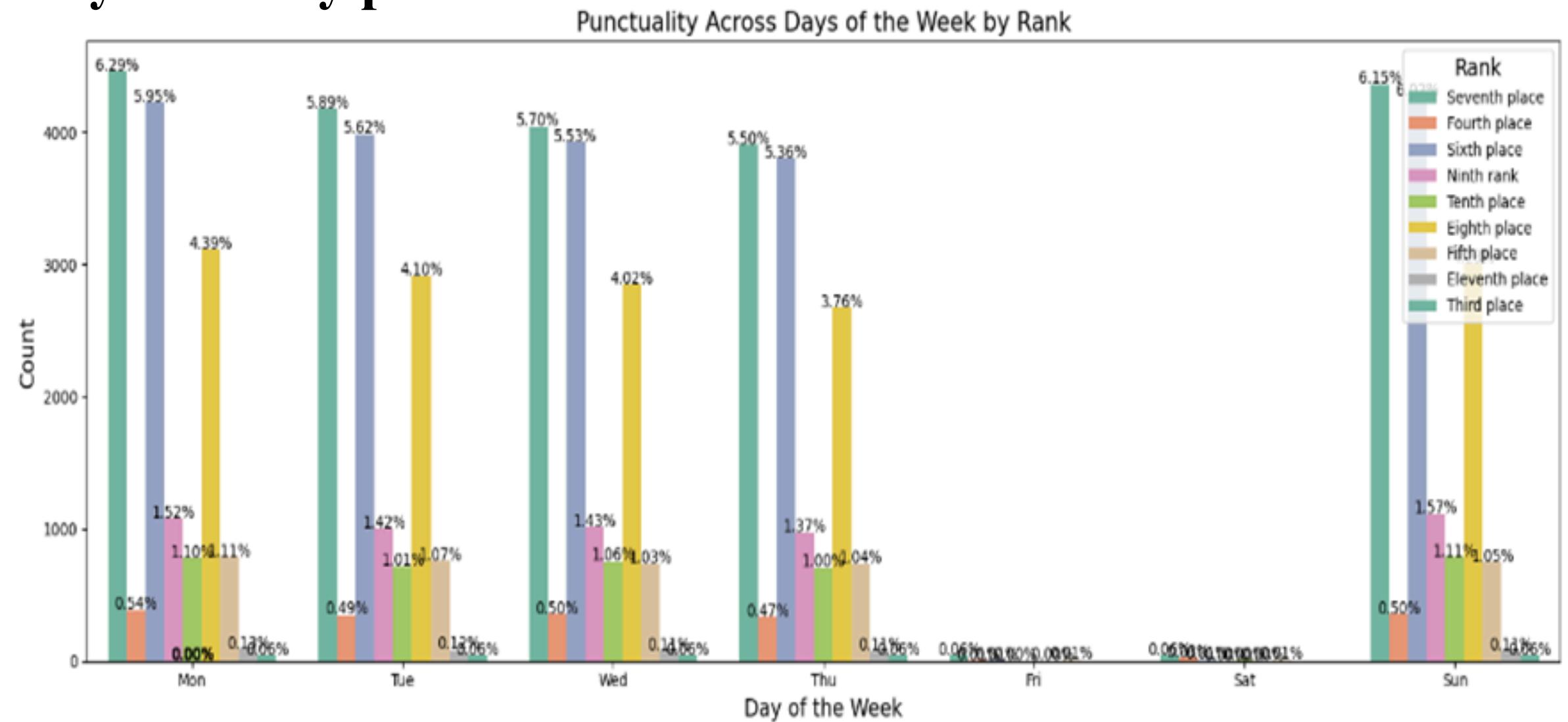
- Highest Attendance: Monday and Tuesday.
- Lowest Attendance: Friday and Saturday (weekend).



Data visualization(con't)

- Punctuality Rate by Job Rank: Analyze weekly patterns

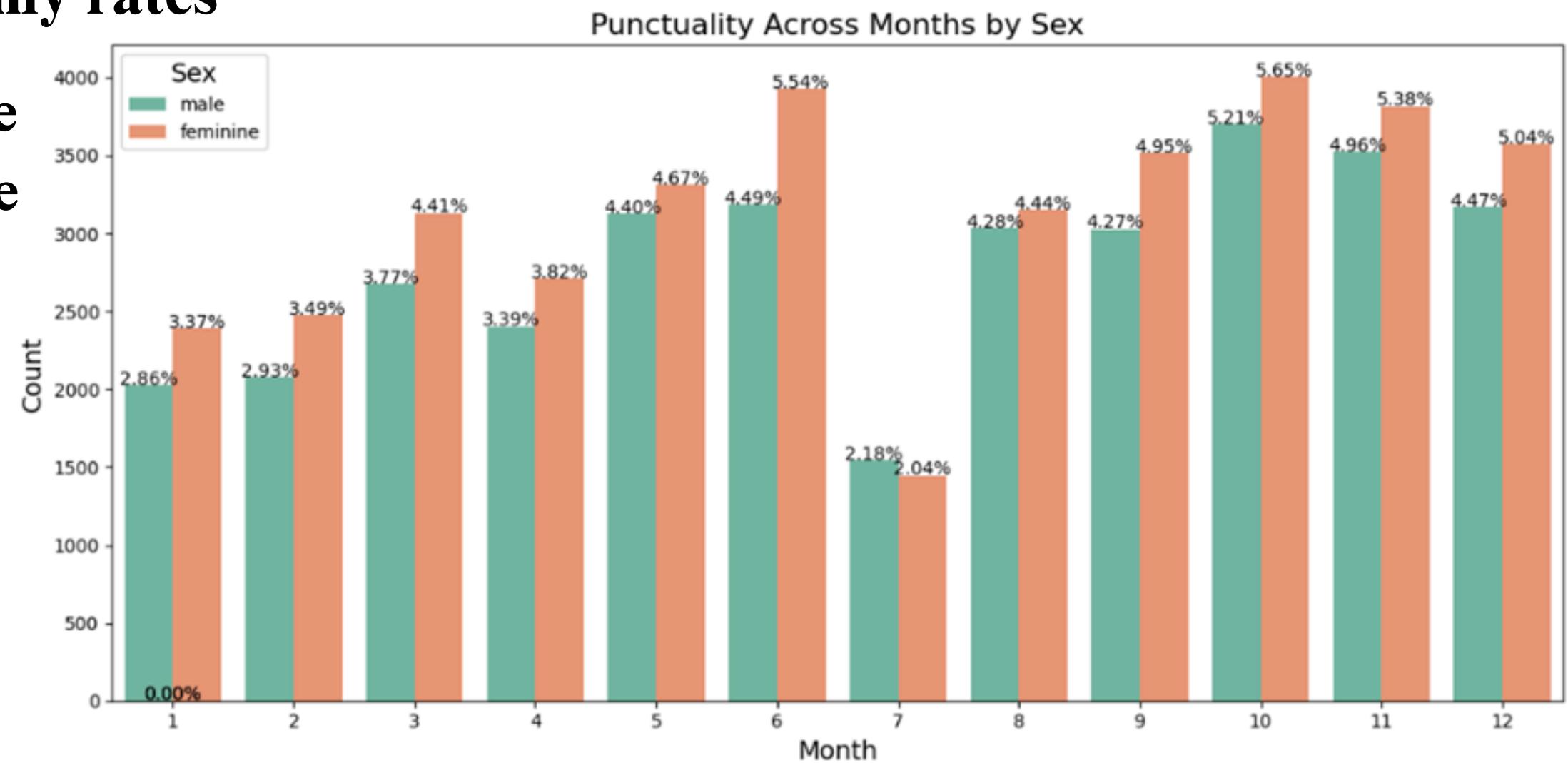
The seventh, sixth, and eighth ranks were the most committed throughout the week



Data visualization(con't)

- Punctuality by Gender: Analyze monthly rates

Females were the most punctual over the months compared to males who were the least

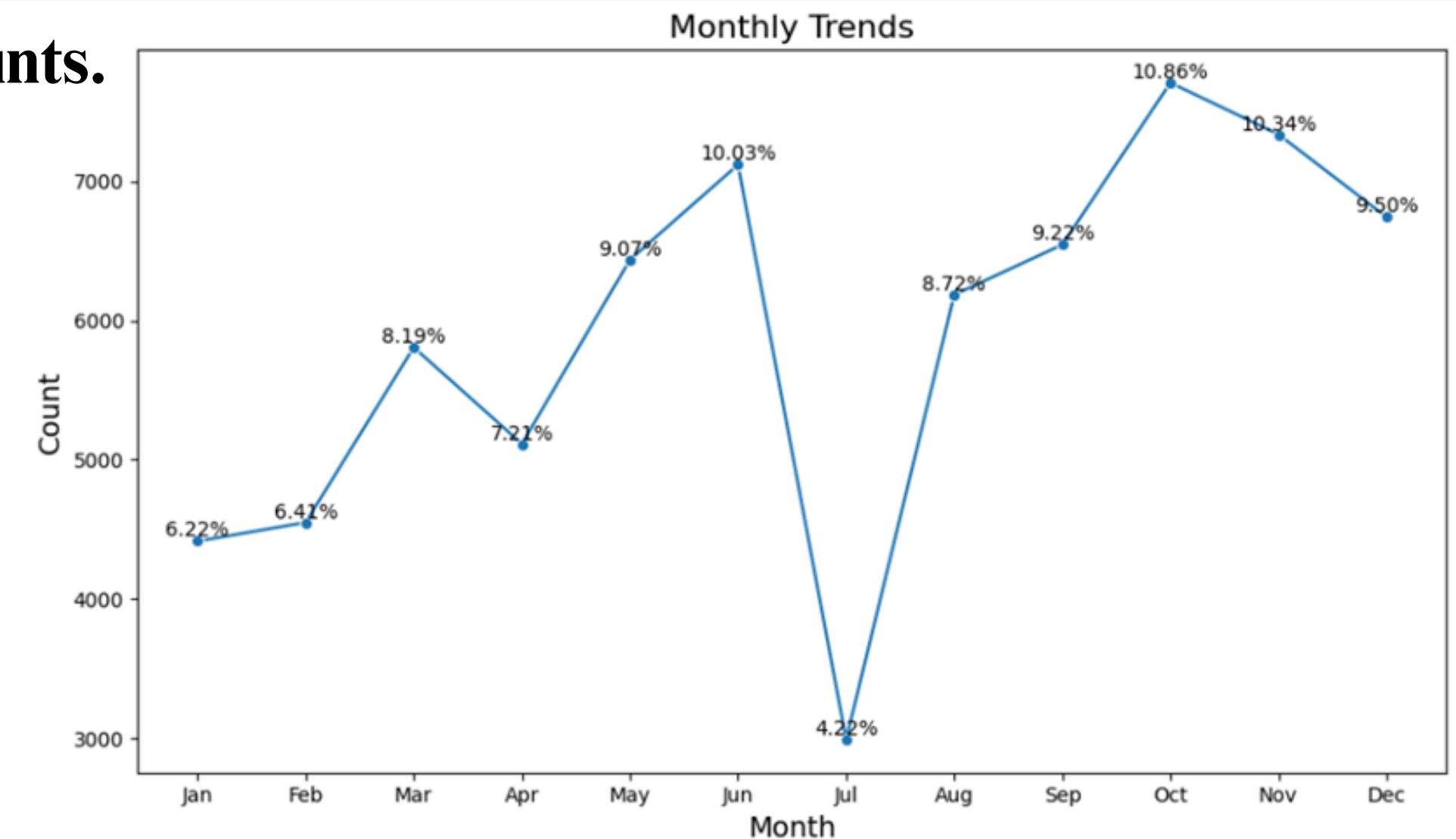


Data visualization(con't)

- Monthly Entry Trends: Analyze entry counts.

Yearly Attendance Insights

- Lowest Rate: July (4.22%)
- Highest Rate: October (10.86%)





Outline

- Foreword
- Background
- Research Problem
- Contribution
- Related Work
- Design and Implementation
- **Results**
- Conclusion
- Future Work





Results

- Model Performance: Random Forest accuracy of 70.5%.
- Findings: Significant differences in exit times based on job rank and gender, influenced by workplace culture and policies.





Outline

- Foreword
- Background
- Research Problem
- Contribution
- Related Work
- Design and Implementation
- Results
- Conclusion
- Future Work





Conclusion

- **Insights Gained:** Data analysis and the Random Forest model provided valuable insights into employee behaviors and patterns.
- **Potential:** Advanced analytics can enhance understanding of workplace factors.
- **Recommendations:** Emphasize investing in data technologies and developing supportive policies for job performance and satisfaction.
- **Future Work:** Focus on improving models, exploring new methods, and enhancing technological infrastructure.
- **Significance:** This project is a foundation for comprehensive analytical initiatives, highlighting data's value in strategy formulation for sustainable growth and success.





Outline

- Foreword
- Background
- Research Problem
- Contribution
- Related Work
- Design and Implementation
- Results
- Conclusion
- **Future Work**





Future Work

- Model Improvement: Add variables and explore advanced algorithms.
- Gender Analysis: Study and address gender disparities.
- Longitudinal Studies: Analyze job ranks and qualifications over time.
- Policy Assessment: Evaluate and improve workplace policies.
- Tech Integration: Use AI and big data for better analysis.





References

- Bass, B. M., & Avolio, B. J. (1994). Improving organizational effectiveness through transformational leadership. Thousand Oaks, CA: Sage.
- García-Morales, V. J., Jiménez-Barrionuevo, M. M., & Gutiérrez-Gutiérrez, L. (2012). Transformational leadership influence on organizational performance through organizational learning and innovation. *Journal of Business Research*, 65(7), 1040-1050.
- Herzberg, F., Mausner, B., & Snyderman, B. B. (1959). *The motivation to work* (2nd ed.). New York, NY: John Wiley & Sons.
- Johnson, S., & Smith, P. (2013). Gender differences in workplace behavior and conflict resolution. *Journal of Business Studies Quarterly*, 5(1), 12-22.
- Kotter, J. P., & Heskett, J. L. (1992). *Corporate culture and performance*. New York: Free Press.
- Leith wood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, 38(2), 112-129.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370-396.
- Nguyen, T., Taylor, J., & Bradley, S. (2016). Job autonomy and job satisfaction: new evidence. *Education Economics*, 24(3), 282-307.
- Smith, M., & Robertson, J. (2015). Demographic influences on employee satisfaction in higher education. *Teaching in Higher Education*, 20(5), 513-527.
- Turner, B. A., & Müller, R. (2005). The influence of project managers on



Questions



