Analyzing Employee Absenteeism & Identifying Employees at Risk
Project part 1&2

Mentor:Munna Pandey

Submitted:Deepak Venkat Sairam

Introduction:

Employee absenteeism poses significant challenges for organizations, affecting productivity and workforce management.
This project aims to analyze absenteeism data to identify trends, reasons for absences, and patterns among employees.
By leveraging data-driven insights, we seek to predict absence durations, identify employees at risk of high absenteeism, and group individuals with similar behaviors.
The ultimate goal is to optimize workforce strategies, improve efficiency, and create a supportive work environment.

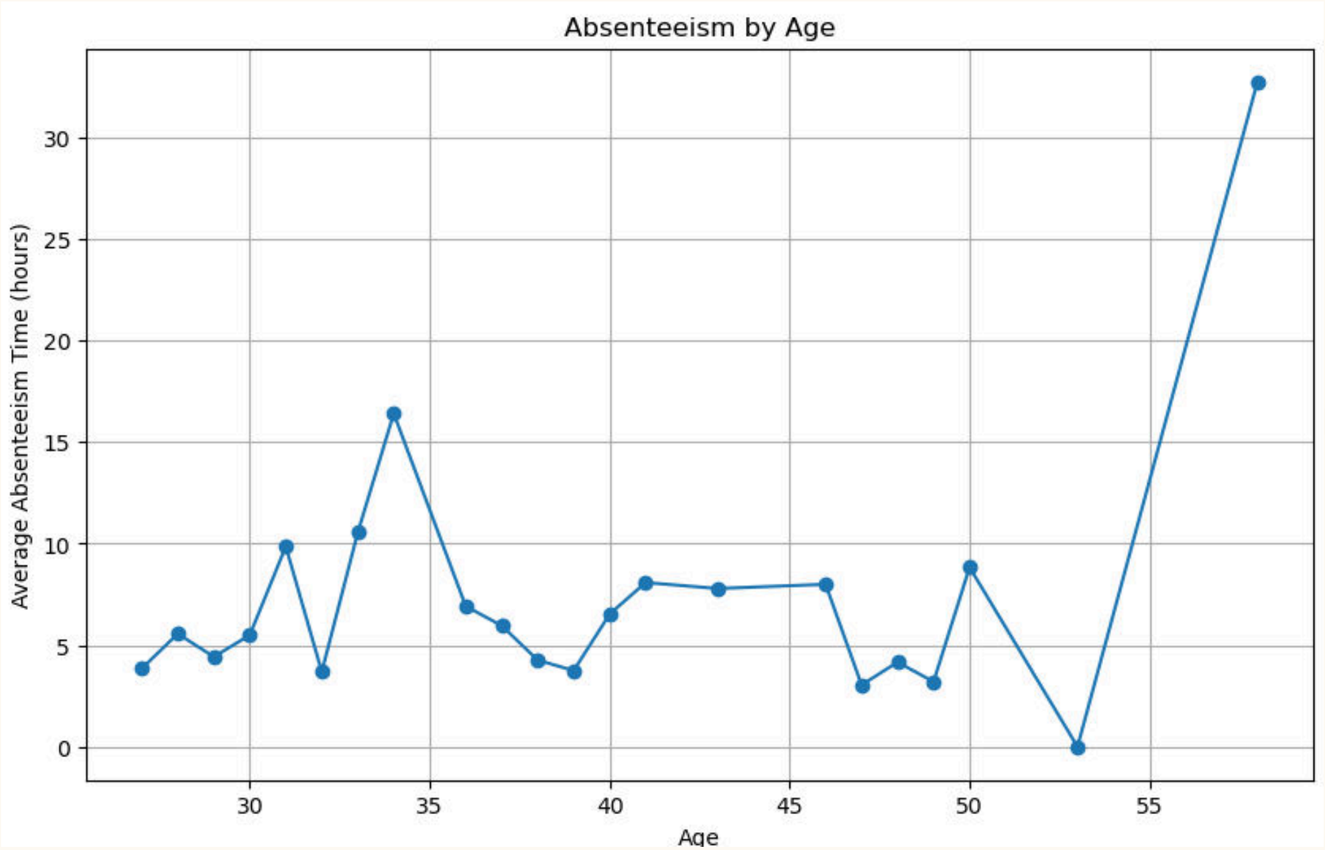
Approach:

- Data cleaning and preprocessing.
- Exploratory data analysis to uncover key absenteeism trends.
- Predictive modeling to foresee absence durations and high-risk employees.
- Clustering to group employees with similar absenteeism patterns.

Part1

Task2

How does absenteeism vary across different employee demographics?



Interpretation:

- **Younger Age Groups (30-35):** Absenteeism rates are relatively high in this range, with peaks around 32 and 34 years of age.
- **Mid-Age Groups (35-45):** Absenteeism generally decreases in this range, with a slight increase around 40 years of age.
- **Older Age Groups (45-55):** Absenteeism shows a more significant increase again, with a peak at 55 years of age.

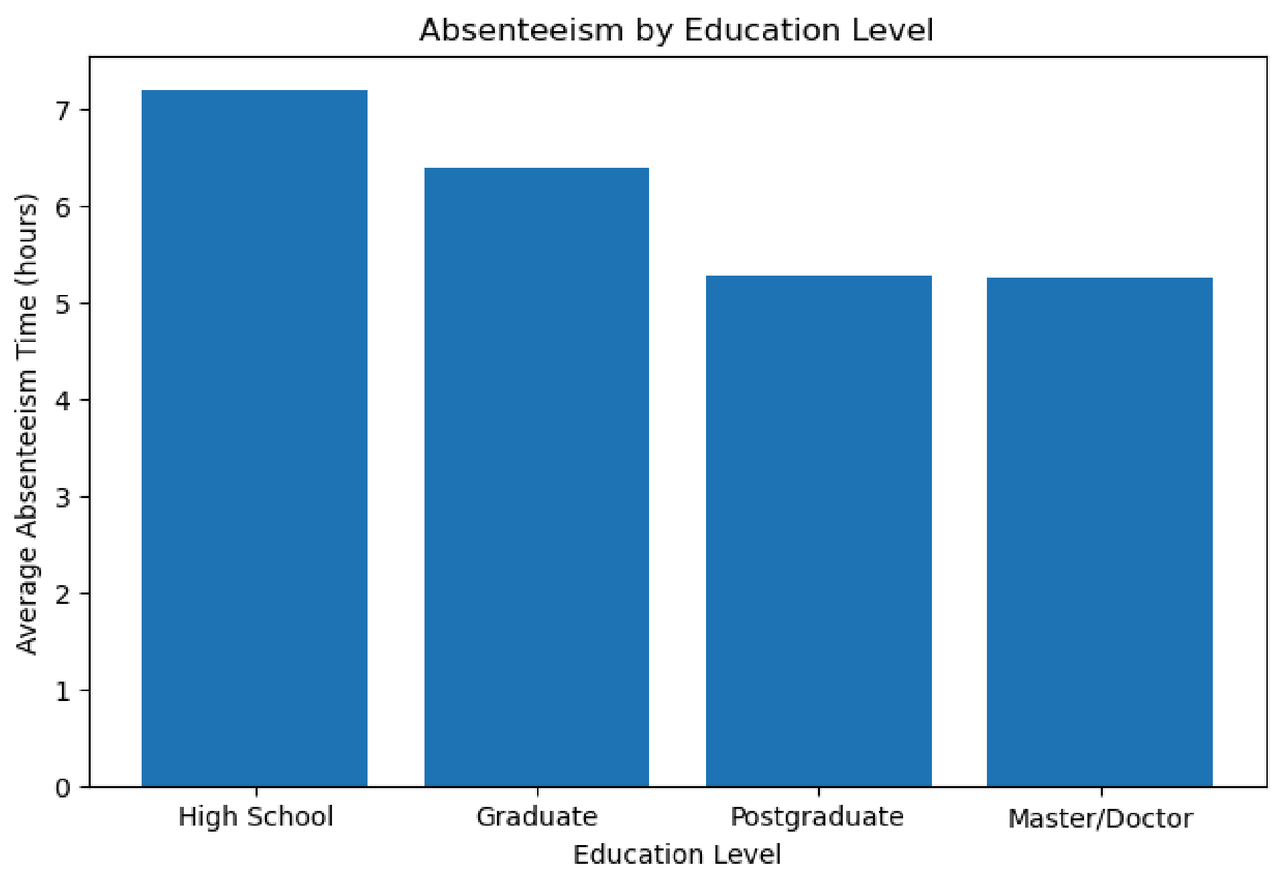
Possible Interpretations:

- **Younger Age Groups:** Higher absenteeism among younger workers could be due to factors like new job experiences, less established routines, or higher levels of social activities.
- **Mid-Age Groups:** The decrease in absenteeism in this range might be attributed to increased stability in career and family life.
- **Older Age Groups:** The rise in absenteeism among older workers could be linked to health issues, caregiving responsibilities, or approaching retirement.

Part1

Task2

How does absenteeism vary across different employee demographics?



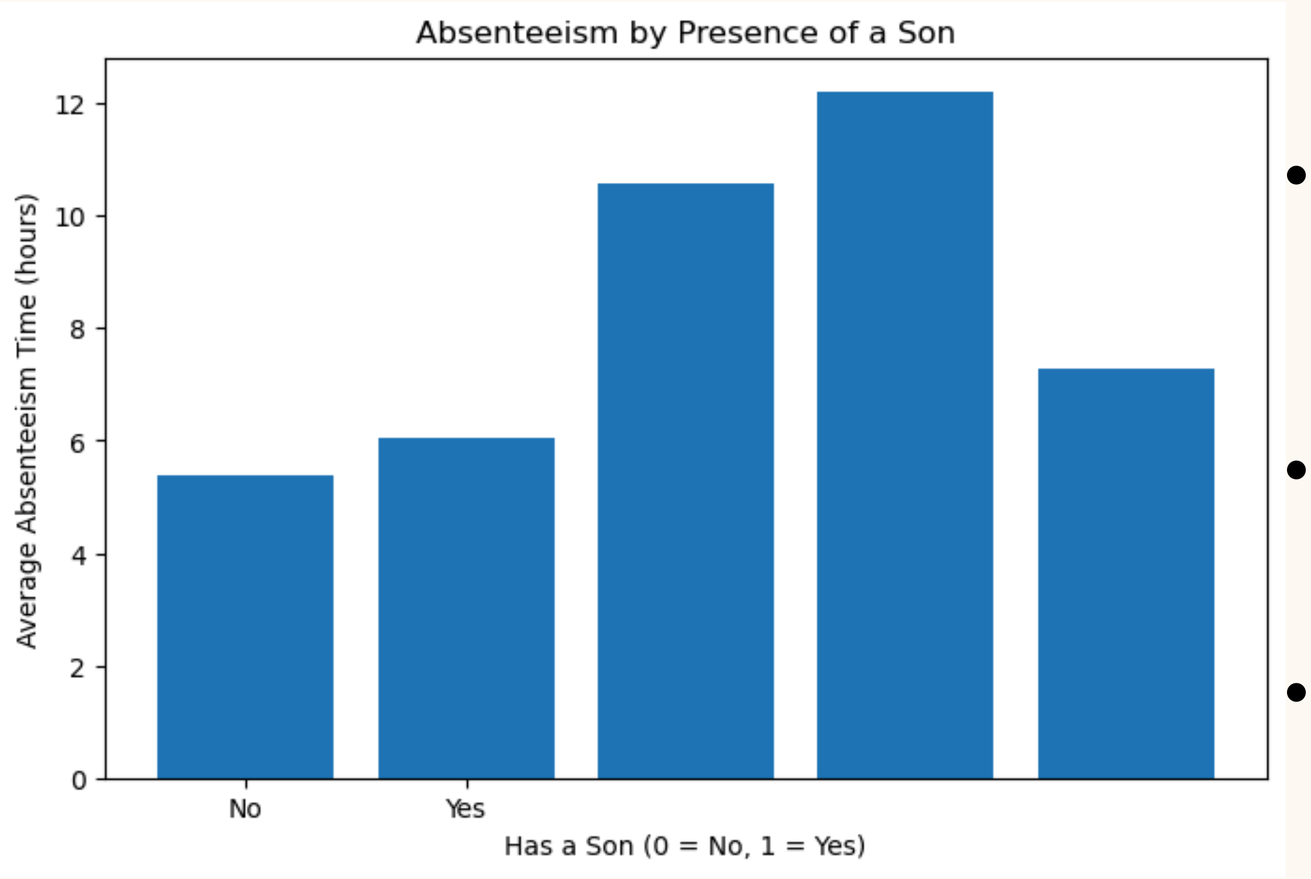
Interpretation:

- High School: This group has the highest average absenteeism time, around 7 hours.
- Graduate: The average absenteeism time decreases for graduates, around 6.5 hours.
- Postgraduate: A further decrease is observed for postgraduate education, with an average of 5.5 hours.
- Master/Doctor: The lowest average absenteeism time is seen for individuals with Master's or Doctorate degrees, around 5.5 hours.

Part1

Task2

How does absenteeism vary across different employee demographics?



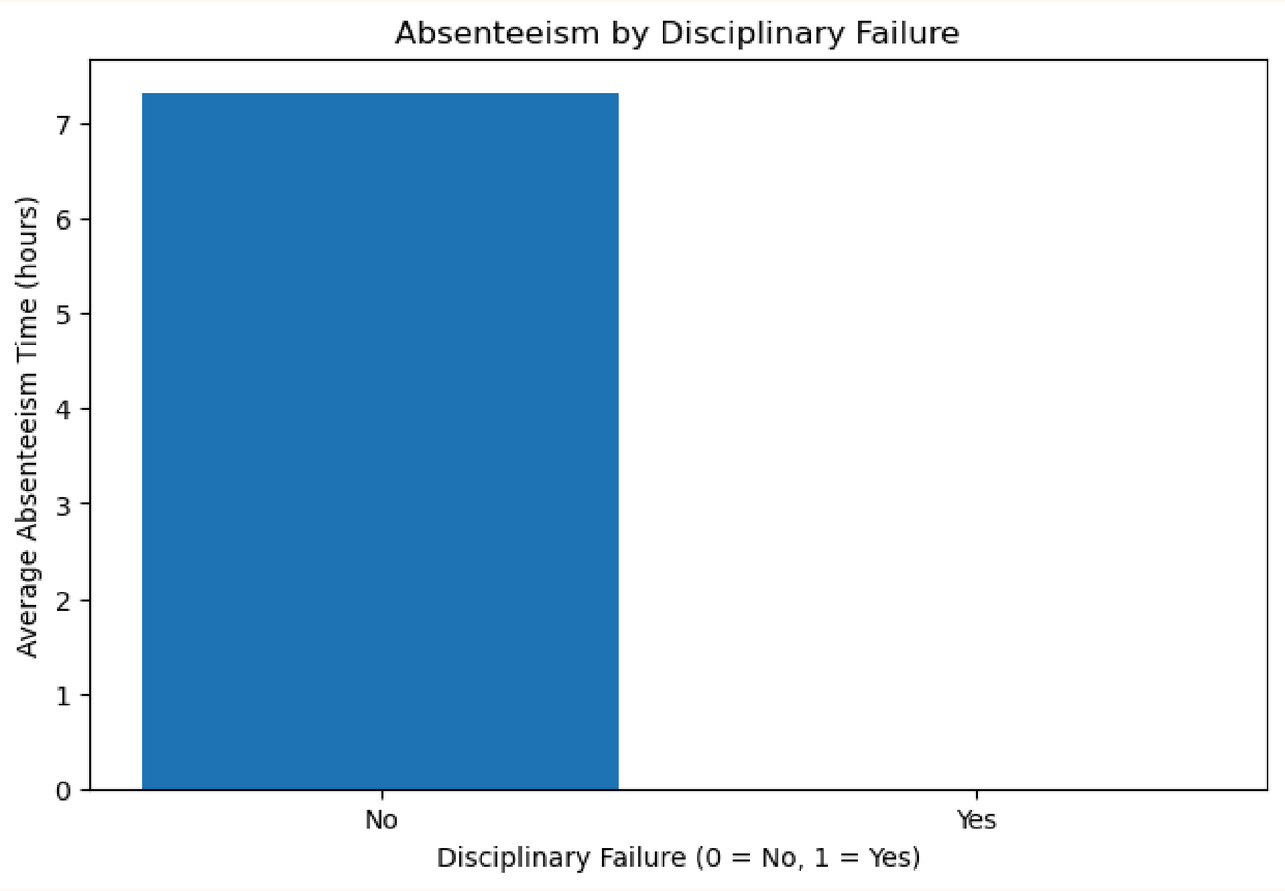
Interpretation:

- **Overall Trend:** Employees with sons generally have a higher average absenteeism time compared to those without sons.
- **Employees with Sons:** The average absenteeism time for employees with sons is approximately 12 hours.
- **Employees without Sons:** The average absenteeism time for employees without sons is approximately 5.5 hours.

Part1

Task2

How does absenteeism vary across different employee demographics?



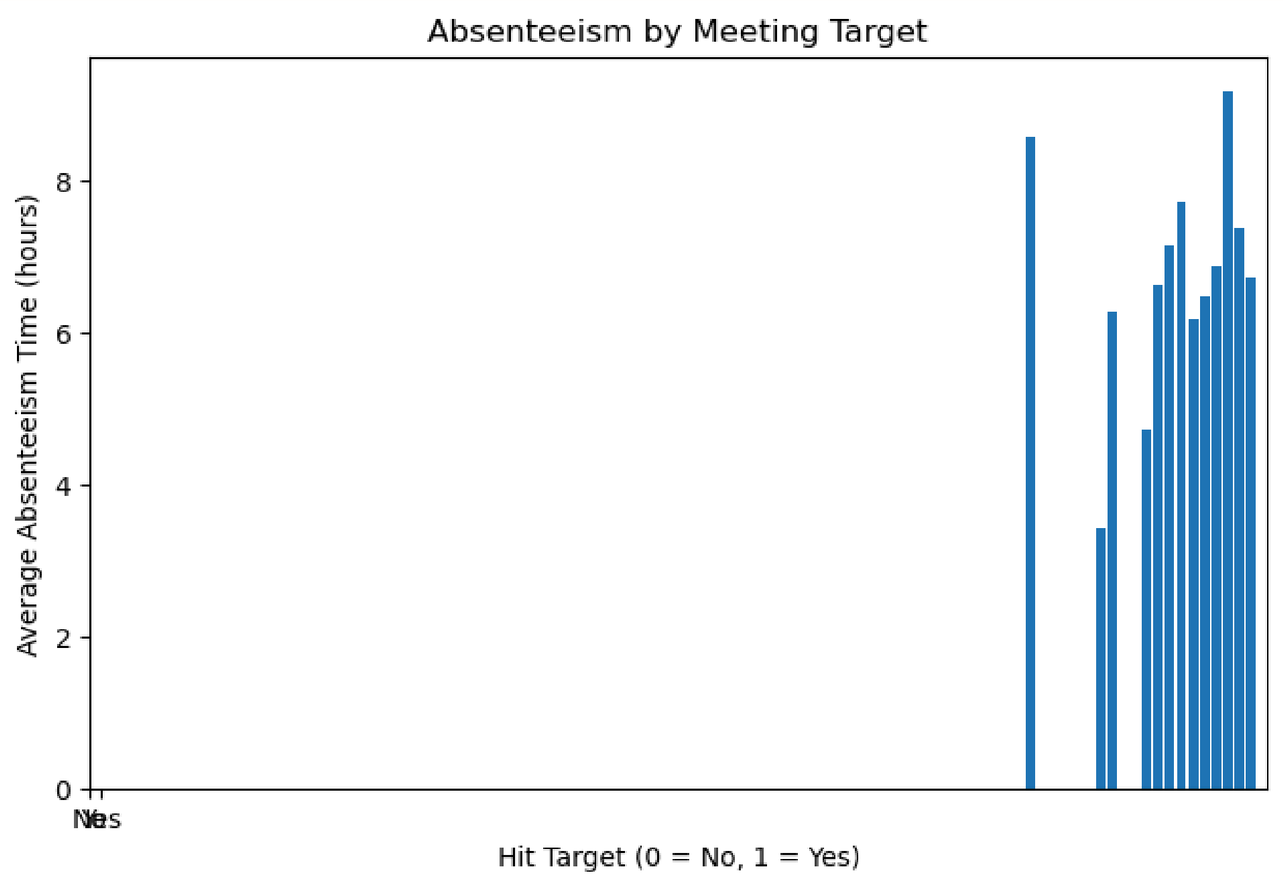
Interpretation:

- **Overall Trend:** Employees with a history of disciplinary failures have a significantly higher average absenteeism time compared to those without.
- **Employees with Disciplinary Failures:** The average absenteeism time for employees with disciplinary failures is approximately 7 hours.
- **Employees without Disciplinary Failures:** The average absenteeism time for employees without disciplinary failures is not explicitly shown in the chart.

Part1

Task2

How does absenteeism vary across different employee demographics?



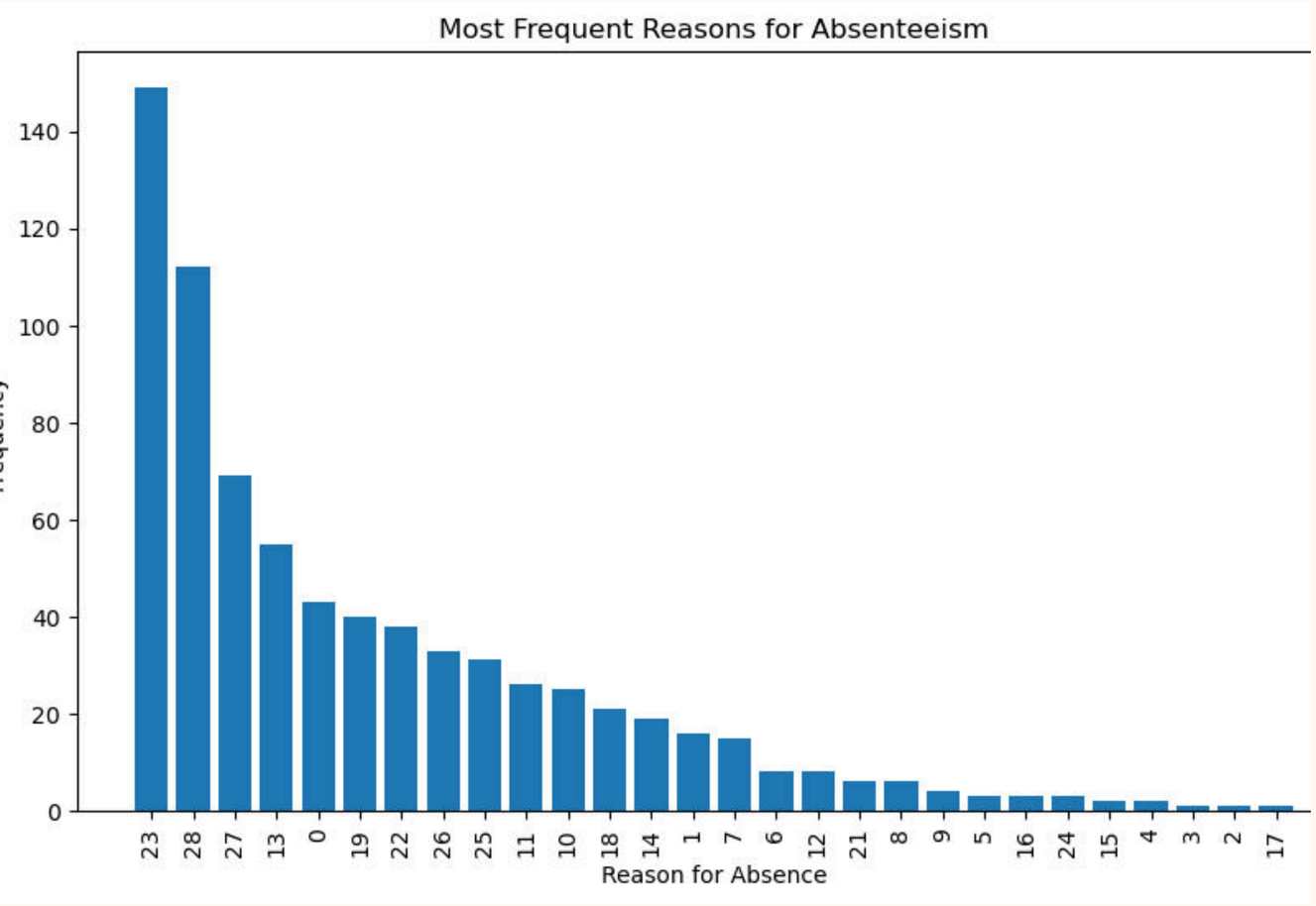
Interpretation:

- **Overall Trend:** Employees who met their targets generally have a lower average absenteeism time compared to those who did not.
- **Employees who Met Target:** The average absenteeism time for employees who met their targets is around 6-8 hours
- **Employees who Did Not Meet Target:** The average absenteeism time for employees who did not meet their targets is around 8-9 hours.

Part1

Task3

What are the most frequent reasons for employee absence? Are there any pattern in these reasons?



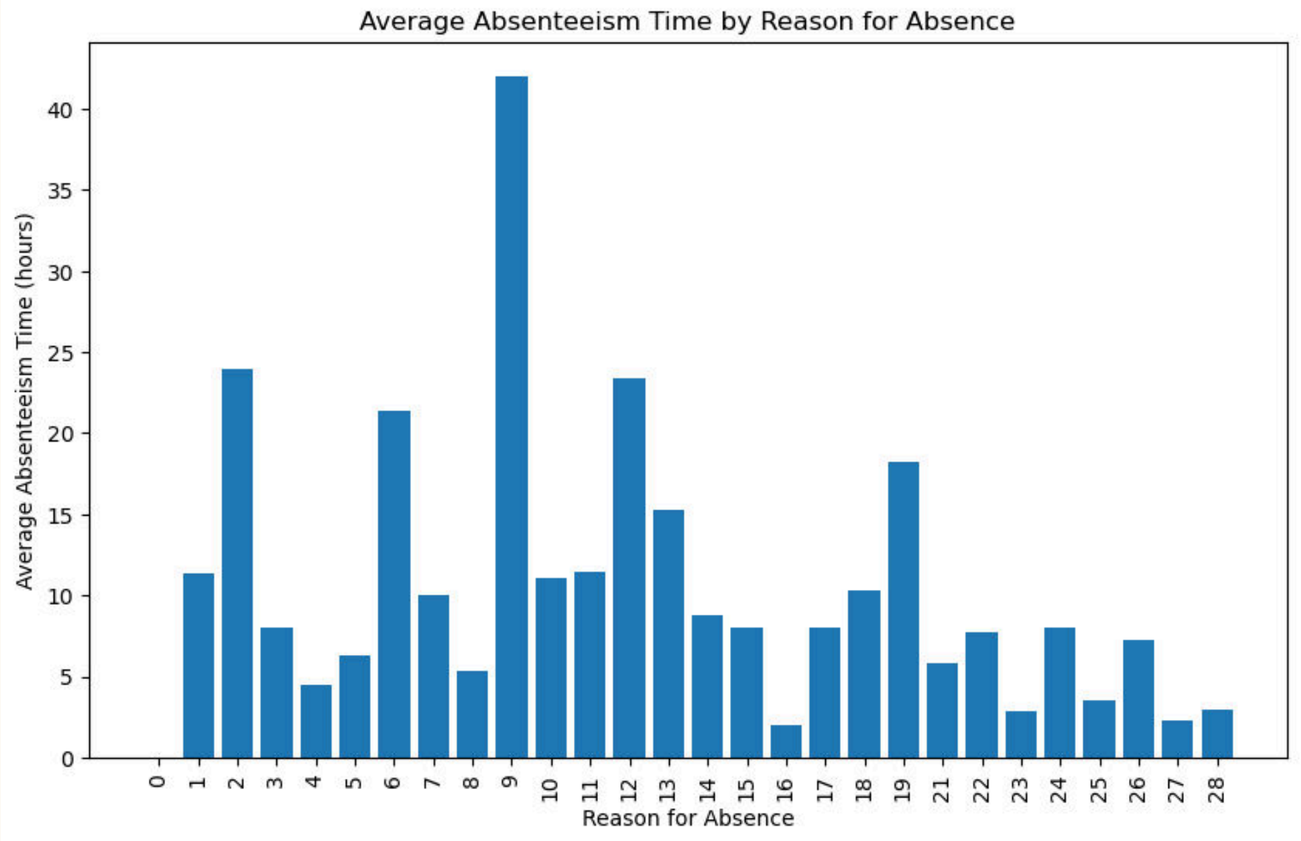
Interpretation:

- **Top Reason:** The reason labeled "23" is the most frequent cause of absenteeism, accounting for the highest number of occurrences.
- **Reason Distribution:** The frequencies of other reasons decline gradually, indicating a long tail of less frequent causes.
- **Data Limitations:** The specific reasons behind the numerical labels are not provided, making it difficult to draw precise conclusions about the nature of the absences.

Part1

Task3

What are the most frequent reasons for employee absence? Are there any pattern in these reasons?



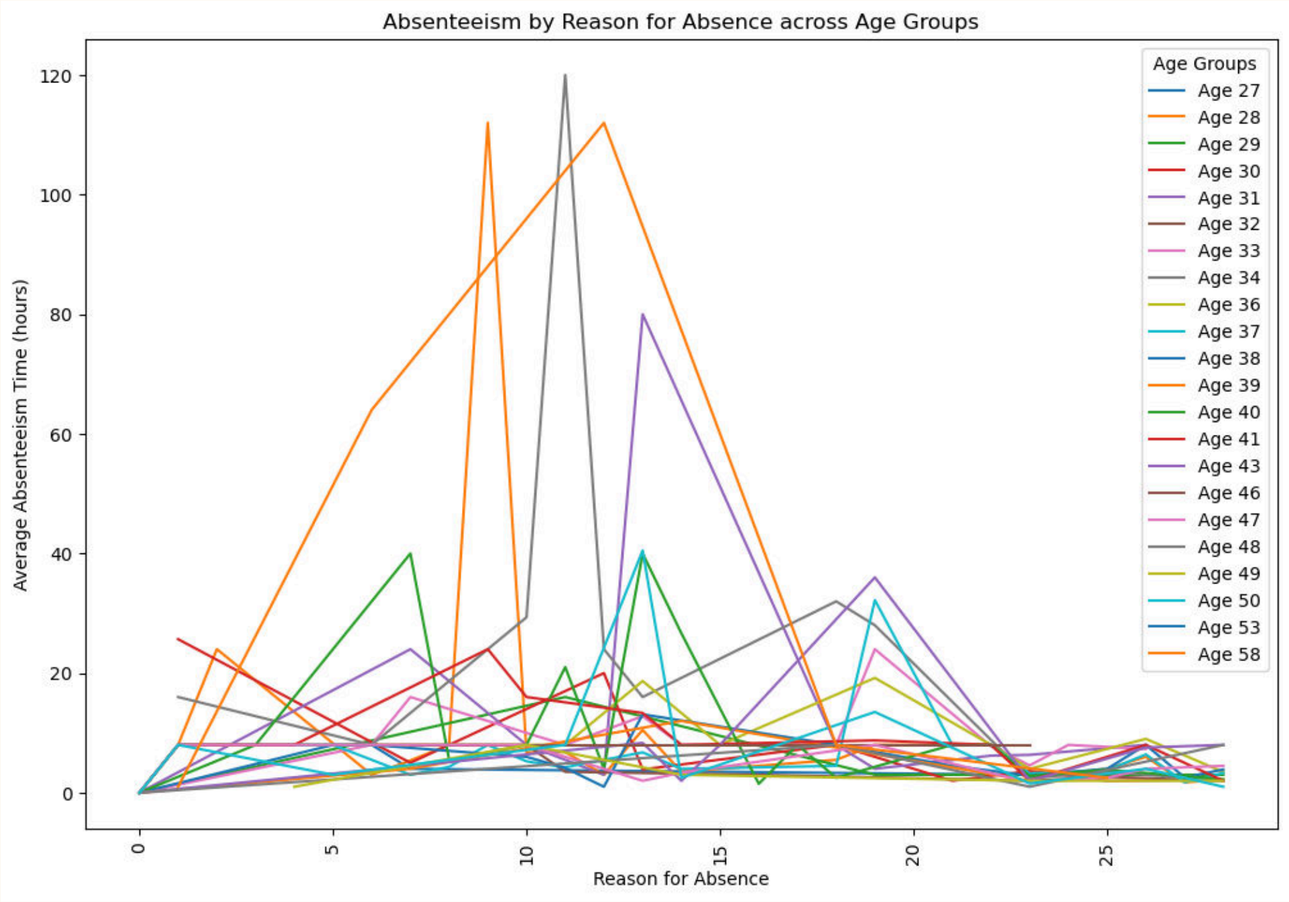
Interpretation:

- **Variability in Absence Duration:** The chart shows significant variation in the average absenteeism time across different reasons.
- **Top Reason:** The reason labeled "10" has the highest average absenteeism time, indicating that absences attributed to this reason tend to be longer.
- **Other Notable Reasons:** Reasons like "1", "2", "5", "12", and "19" also have relatively high average absenteeism times.

Part1

Task3

What are the most frequent reasons for employee absence? Are there any pattern in these reasons?



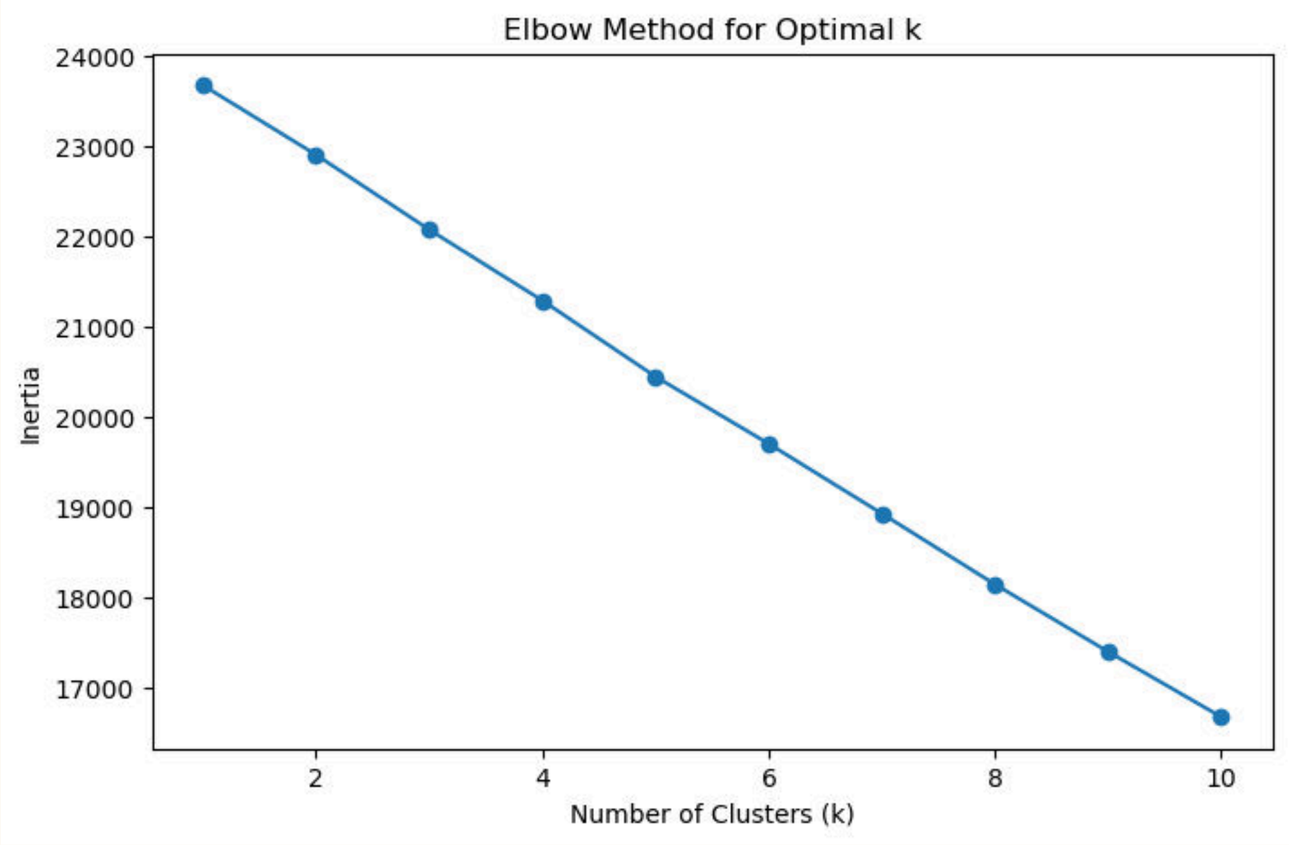
Interpretation:

- **Variability Across Age Groups:** The chart shows significant variation in the average absenteeism time for different reasons across different age groups.
- **No Clear Trend:** There doesn't seem to be a consistent pattern or trend in the absenteeism rates across age groups. Some age groups have higher absenteeism times for certain reasons, while others have lower rates.
- **Reason-Specific Variation**: The average absenteeism time for a particular reason can vary greatly across different age groups. For example, the reason labeled "25" shows a wide range of absenteeism times across different age groups.

Part2

Task2

Use K-Means clustering to segment customers into groups based on features like demographics, financial attributes, and engagement factors.



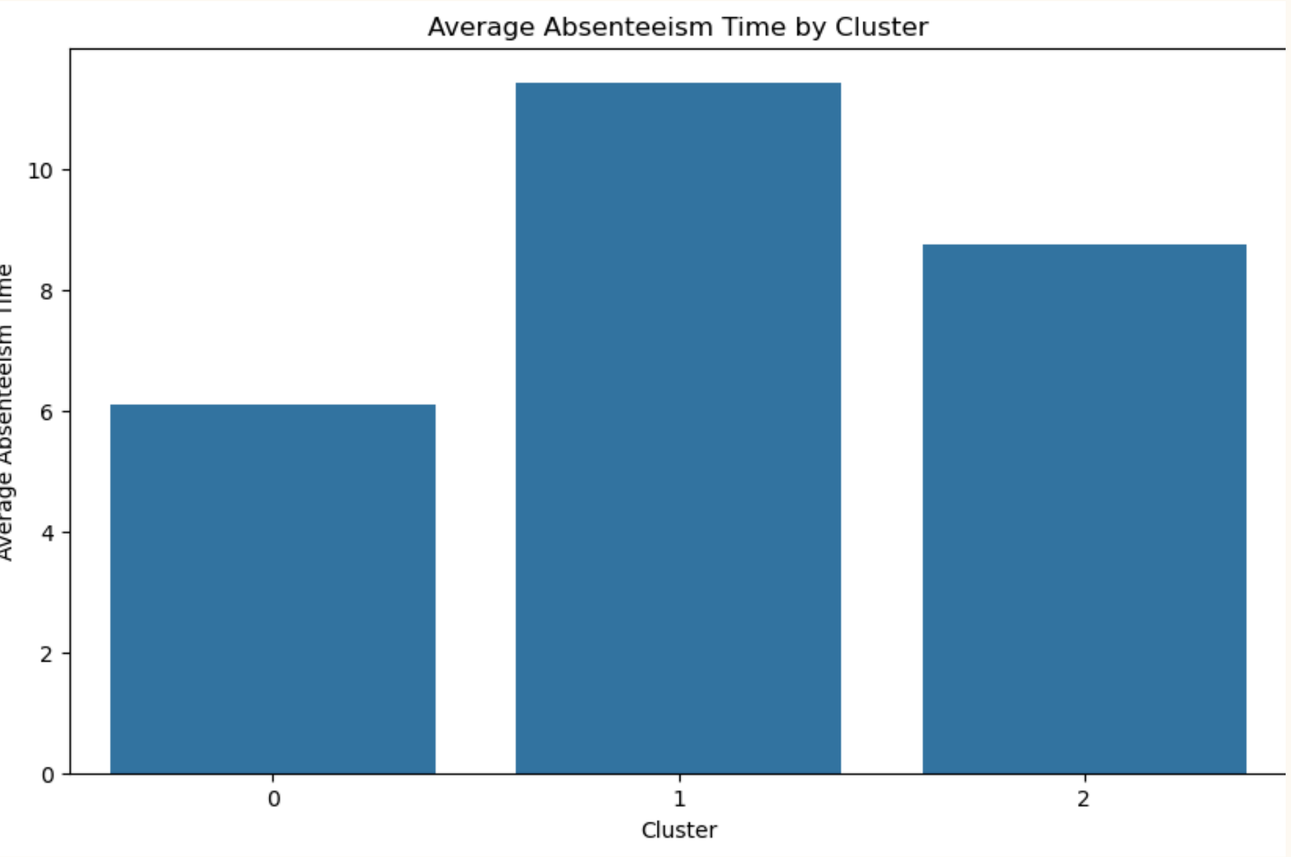
Interpretation:

- **Inertia:** The y-axis represents inertia, which is a measure of within-cluster variation. Lower inertia indicates that the data points are more tightly clustered.
- **Number of Clusters (k):** The x-axis shows the number of clusters (k) being considered.

Part2

Task2

Use K-Means clustering to segment customers into groups based on features like demographics, financial attributes, and engagement factors.



Interpretation:

- **Cluster Variation:** The chart shows variation in the average absenteeism time across the three clusters.
- **Cluster 1:** Cluster 1 has the highest average absenteeism time, indicating that employees in this cluster tend to have longer absences.
- **Cluster 0 and 2:** Clusters 0 and 2 have lower average absenteeism times, suggesting that employees in these clusters tend to have shorter or less frequent absences.

Conclusion:

Overall Conclusion

Absenteeism is a complex issue influenced by various factors, including individual characteristics, work environment, and organizational policies. The provided charts offer valuable insights into the patterns and trends of absenteeism within a specific organization.

Key Findings:

- **Individual Factors:** Age, gender, and parental status can impact absenteeism rates.
- **Work Environment:** Job satisfaction, workload, and work-life balance can influence employee attendance.
- **Organizational Factors:** Company policies regarding leave, flexible work arrangements, and employee support programs can affect absenteeism.
- **Clustering:** Clustering analysis can help identify distinct groups of employees with similar absenteeism patterns, enabling targeted interventions.

THANK YOU