Project Overview

In this project, we analyze employee data from ABC Company, which consists of 458 employees across various teams. The objective is to gain insights into the workforce, focusing on aspects such as team distribution, positions, salary expenditure, and correlations between age and salary. Through this analysis, we aim to present a comprehensive understanding of the company's workforce structure and identify key trends.

Objectives:

1. **Preprocessing**: Clean the data by correcting inconsistencies in the height column, ensuring data integrity before proceeding with deeper analysis.
2. **Team Distribution**: Determine how employees are distributed across teams and calculate the percentage split, highlighting the teams with the highest and lowest numbers.
3. **Position Segregation**: Classify employees based on their job positions and understand the composition of different roles within the company.
4. **Age Group Analysis**: Identify the predominant age group among the employees and observe age-related trends within the organization.
5. **Salary Expenditure**: Investigate which team and position incur the highest salary expenditures to understand budget allocation and resource distribution.
6. **Correlation Analysis**: Explore the relationship between employees' ages and their salaries to see if age is a determining factor in salary structure.

Methodology:

To accomplish the analysis, we utilized **Python** for data manipulation and visualization, leveraging libraries such as:

- **Pandas**: For data preprocessing, manipulation, and statistical analysis.
- **Matplotlib**: For creating various graphical representations such as bar charts, pie charts, and scatter plots.
- **Seaborn**: For more advanced statistical plots and visualizations.

Deliverables:

The project involves the following key steps:

1. **Data Preprocessing**: Fixing errors or inconsistencies in the dataset.
2. **Analysis Tasks**: Answering specific business questions through data exploration and calculation.
3. **Visualization**: Presenting findings with clear and meaningful charts.

4. **Insights and Recommendations**: Summarizing key trends and patterns observed, providing actionable insights for ABC Company.

1.

File Edit View Run Kernel Settings Help

Code

[35]:
```
# Count the number of employees in each team
team_distribution = df['Team'].value_counts()

# Calculate the percentage split
team_percentage = (team_distribution / len(df)) * 100

# Display the results
team_distribution, team_percentage
```

[35]:
```
(Team
 New Orleans Pelicans      19
 Memphis Grizzlies         18
 Utah Jazz                 16
 New York Knicks           16
 Milwaukee Bucks           16
 Brooklyn Nets             15
 Portland Trail Blazers    15
 Oklahoma City Thunder     15
 Denver Nuggets            15
 Washington Wizards        15
 Miami Heat                15
 Charlotte Hornets         15
 Atlanta Hawks             15
 San Antonio Spurs         15
 Houston Rockets           15
 Boston Celtics            15
 Indiana Pacers            15
 Detroit Pistons           15
 Cleveland Cavaliers       15
 Chicago Bulls             15
 Sacramento Kings          15
 Phoenix Suns              15
 Los Angeles Lakers        15
```

Markdown JupyterLab

[27]:
```
import numpy as np

# Generate random heights between 150 and 180
np.random.seed(0)  # For reproducibility
df['height'] = np.random.randint(150, 181, size=len(df))

# Verify the update
df[['height']].head()
```

[27]:

	height
0	162
1	165
2	171
3	150
4	153

2.Distribution of employees across each team:

[ ]:
```
team_distribution = df['team'].value_counts(normalize=True) * 100
```

[ ]:

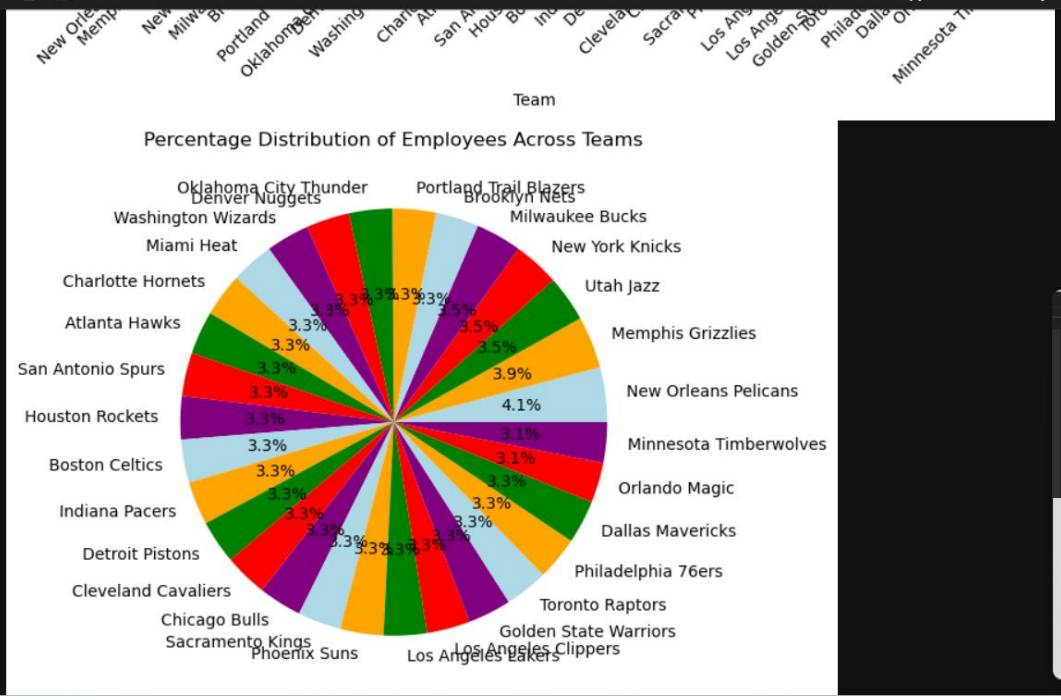
```
Washington Wizards      3.275109
Miami Heat              3.275109
Charlotte Hornets       3.275109
Atlanta Hawks           3.275109
San Antonio Spurs       3.275109
Houston Rockets         3.275109
Boston Celtics          3.275109
Indiana Pacers          3.275109
Detroit Pistons         3.275109
Cleveland Cavaliers     3.275109
Chicago Bulls           3.275109
Sacramento Kings        3.275109
Phoenix Suns            3.275109
Los Angeles Lakers      3.275109
Los Angeles Clippers    3.275109
Golden State Warriors   3.275109
Toronto Raptors         3.275109
Philadelphia 76ers      3.275109
Dallas Mavericks        3.275109
Orlando Magic           3.056769
Minnesota Timberwolves  3.056769
Name: count, dtype: float64)
```

[31]:
```
# Check the column names
df.columns
```

[31]:
```
Index(['Name', 'Team', 'Number', 'Position', 'Age', 'Height', 'Weight',
       'College', 'Salary', 'height'],
      dtype='object')
```

File Edit View Run Kernel Settings Help

Code JupyterLab Python [con

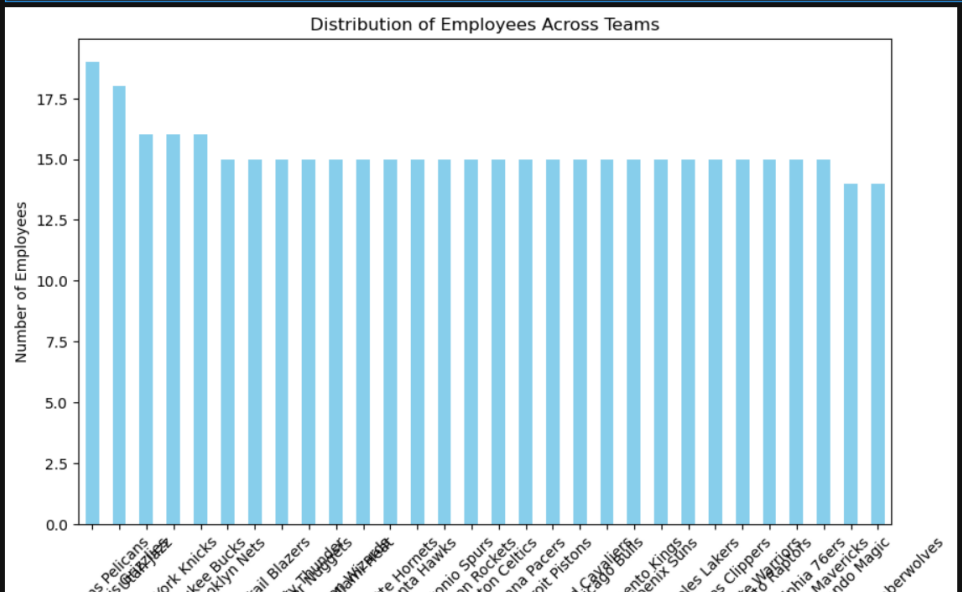


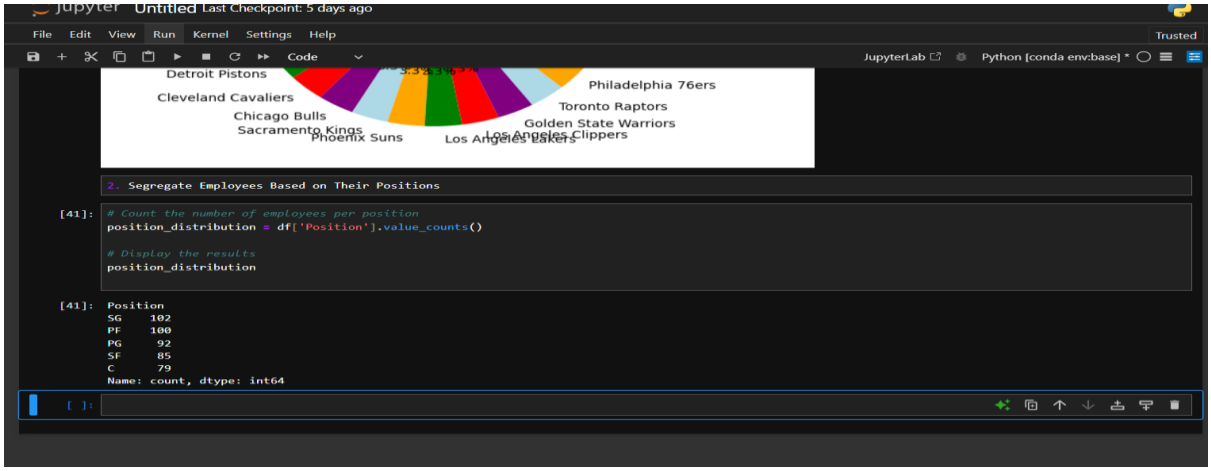
```
plt.figure(figsize=(10,6))
position_distribution.plot(kind='bar', color='lightgreen')
plt.title('Segregation of Employees by Position')
plt.xlabel('Position')
plt.ylabel('Number of Employees')
plt.xticks(rotation=45)
plt.show()
```

jupyter Untitled Last Checkpoint: 5 days ago

File Edit View Run Kernel Settings Help

Code JupyterLab Python [conda env:base] *





2. Segregate Employees Based on Their Positions

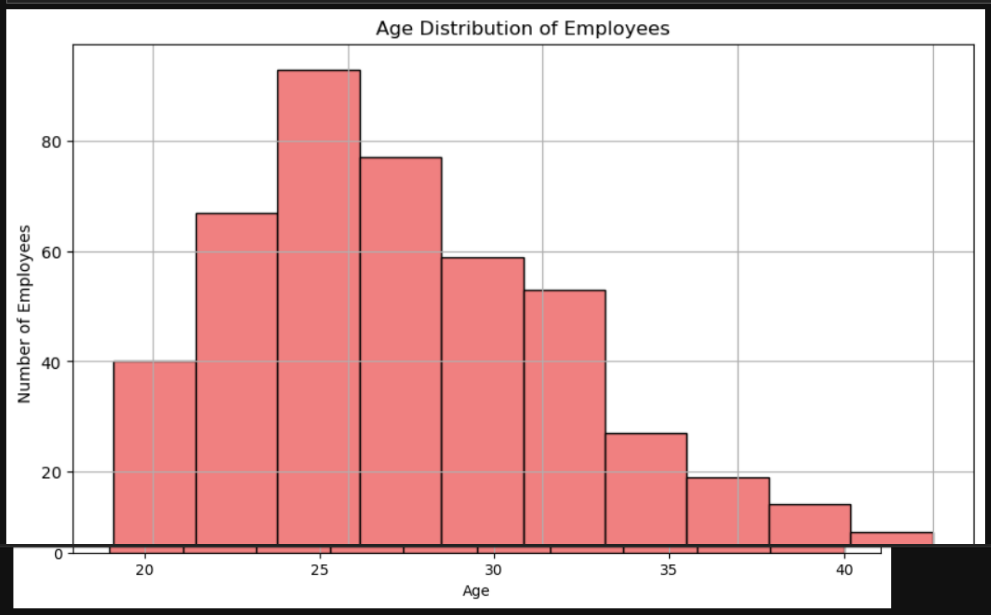
[41]:
```
# Count the number of employees per position
position_distribution = df['Position'].value_counts()

# Display the results
position_distribution
```

[41]:
```
Position
SG    102
PF    100
PG     92
SF     85
C      79
Name: count, dtype: int64
```

[ ]:

```
plt.figure(figsize=(10,6))
df['Age'].hist(bins=10, color='lightcoral', edgecolor='black')
plt.title('Age Distribution of Employees')
plt.xlabel('Age')
plt.ylabel('Number of Employees')
plt.show()
```



4. Discover Which Team and Position Have the Highest Salary Expenditure

[52]:
```
# Group by team and sum the salary for each team
team_salary_expenditure = df.groupby('Team')['Salary'].sum()

# Find the team with the highest salary expenditure
highest_salary_team = team_salary_expenditure.idxmax()
highest_salary_team_value = team_salary_expenditure.max()

# Group by position and sum the salary for each position
position_salary_expenditure = df.groupby('Position')['Salary'].sum()

# Find the position with the highest salary expenditure
highest_salary_position = position_salary_expenditure.idxmax()
highest_salary_position_value = position_salary_expenditure.max()

highest_salary_team, highest_salary_team_value, highest_salary_position, highest_salary_position_value
```

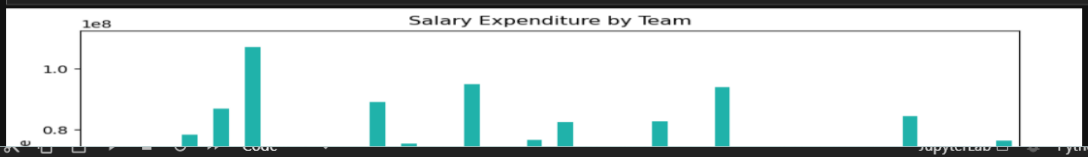
[52]:
```
('Cleveland Cavaliers', 106988689.0, 'C', 466377332.0)
```

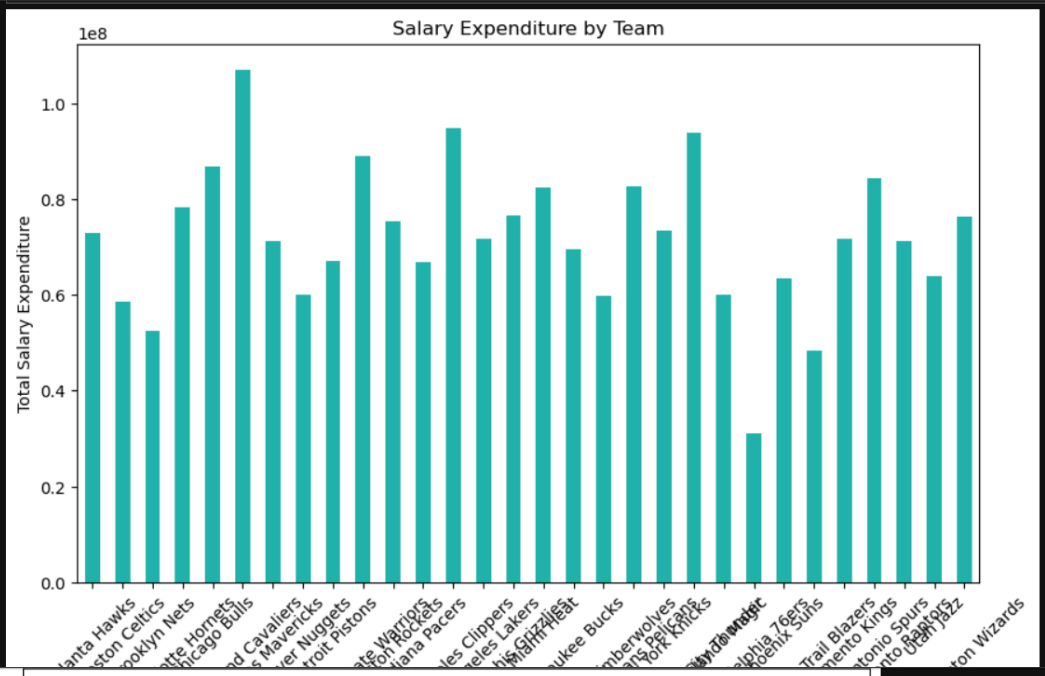
[ ]:

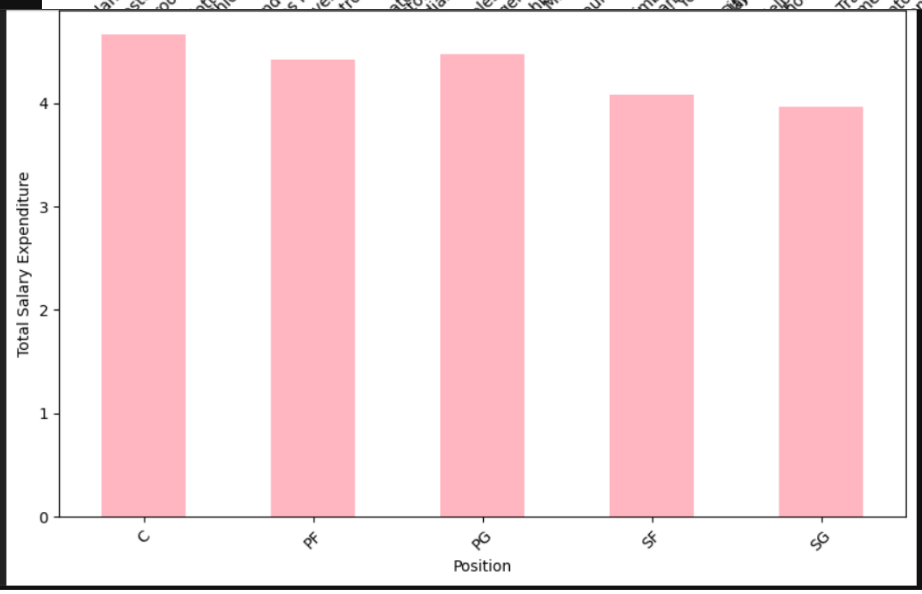
[55]:

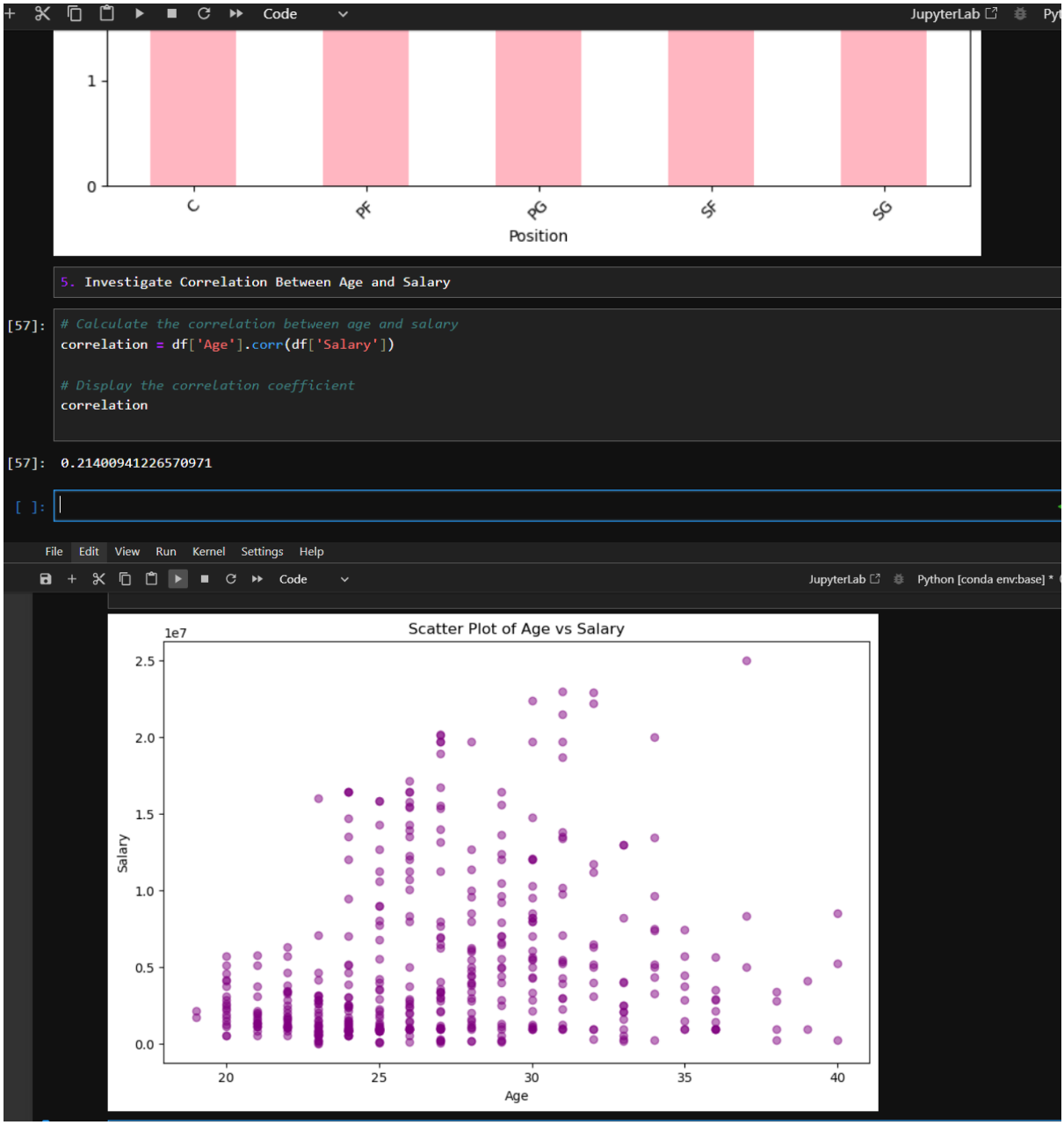
```
# Plot the salary expenditure by team
plt.figure(figsize=(10,6))
team_salary_expenditure.plot(kind='bar', color='lightseagreen')
plt.title('Salary Expenditure by Team')
plt.xlabel('Team')
plt.ylabel('Total Salary Expenditure')
plt.xticks(rotation=45)
plt.show()

# Plot the salary expenditure by position
plt.figure(figsize=(10,6))
position_salary_expenditure.plot(kind='bar', color='lightpink')
plt.title('Salary Expenditure by Position')
plt.xlabel('Position')
plt.ylabel('Total Salary Expenditure')
plt.xticks(rotation=45)
plt.show()
```








Code
JupyterLab
1
0
C
PF
PG
SF
SG
Position
5. Investigate Correlation Between Age and Salary
[57]:
# Calculate the correlation between age and salary
correlation = df['Age'].corr(df['Salary'])
# Display the correlation coefficient
correlation
[57]: 0.21400941226570971
[ ]:
File
Edit
View
Run
Kernel
Settings
Help
Code
JupyterLab
Python [conda env:base]
Scatter Plot of Age vs Salary
1e7
2.5
2.0
1.5
1.0
0.5
0.0
Salary
20
25
30
35
40
Age