Preprocessing:
Correct the data in the "height" column by replacing it with random numbers between 150 and 180. Ensure data consistency and integrity before proceeding with analysis.

Using this code, I randomize the height (between 150 & 180):

1. import pandas as pd
2. import numpy as np
3. import random
4. df = pd.read_csv(r"C:\Users\91807\Downloads\myexcel - myexcel.csv.csv")
5. for i in range(len(df)): df.loc[i, 'Height'] = random.randint(150, 180)

```
df.head(10)
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	169	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	178	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	169	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	174	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	163	231	NaN	5000000.0
5	Amir Johnson	Boston Celtics	90	PF	29	176	240	NaN	12000000.0
6	Jordan Mickey	Boston Celtics	55	PF	21	173	235	LSU	1170960.0
7	Kelly Olynyk	Boston Celtics	41	C	25	167	238	Gonzaga	2165160.0
8	Terry Rozier	Boston Celtics	12	PG	22	164	190	Louisville	1824360.0
9	Marcus Smart	Boston Celtics	36	PG	22	159	220	Oklahoma State	3431040.0

1. Determine the distribution of employees across each team and calculate the percentage split relative to the total number of employees.

   Ans:

```
df1 = df.Team.value_counts()
df1

total_employees = len(df)

team_percentage = (df1 / total_employees) * 100

team_stats = pd.DataFrame({
   'Number of Employees': df1,
   'Percentage of Total': team_percentage
})
team_stats
```

Team	Number of Employees	Percentage of Total
New Orleans Pelicans	19	4.148472
Memphis Grizzlies	18	3.930131
Utah Jazz	16	3.493450
New York Knicks	16	3.493450
Milwaukee Bucks	16	3.493450
Brooklyn Nets	15	3.275109
Portland Trail Blazers	15	3.275109
Oklahoma City Thunder	15	3.275109
Denver Nuggets	15	3.275109
Washington Wizards	15	3.275109
Miami Heat	15	3.275109
Charlotte Hornets	15	3.275109
Atlanta Hawks	15	3.275109
San Antonio Spurs	15	3.275109
Houston Rockets	15	3.275109
Boston Celtics	15	3.275109
Indiana Pacers	15	3.275109
Detroit Pistons	15	3.275109
Cleveland Cavaliers	15	3.275109
Chicago Bulls	15	3.275109
Sacramento Kings	15	3.275109
Phoenix Suns	15	3.275109
Los Angeles Lakers	15	3.275109
Los Angeles Clippers	15	3.275109
Golden State Warriors	15	3.275109
Toronto Raptors	15	3.275109
Philadelphia 76ers	15	3.275109
Dallas Mavericks	15	3.275109
Orlando Magic	14	3.056769
Minnesota Timberwolves	14	3.056769

2. Segregate employees based on their positions within the company.

Ans:

```
# Get the number of employees in each position
position_counts = df['Position'].value_counts()

# Get the total number of employees for percentage calculation
total_employees = len(df)

# Calculate the percentage of employees in each position
position_percentage = (position_counts / total_employees) * 100

# Combine the counts and percentages into a single DataFrame
position_stats = pd.DataFrame({
    'Number of Employees': position_counts,
    'Percentage of Total': position_percentage
```

})

position_stats

Position	Number of Employees	Percentage of Total
SG	102	22.270742
PF	100	21.834061
PG	92	20.087336
SF	85	18.558952
C	79	17.248908

3. Identify the predominant age group among employees.

Ans:

```
age = df.Age.value_counts()
age
```

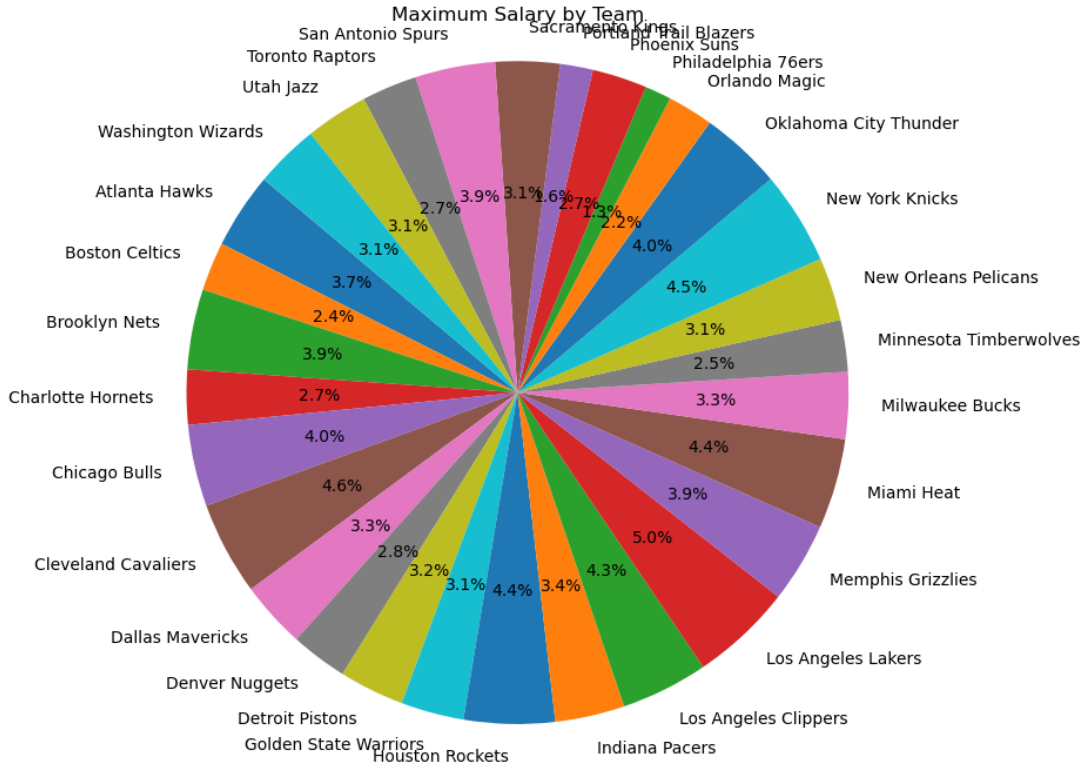
```
Age
24    47
25    46
27    41
23    41
26    36
28    31
30    31
29    28
22    26
31    22
20    19
21    19
33    14
32    13
34    10
36    10
35     9
37     4
38     4
40     3
39     2
19     2
Name: count, dtype: int64
```

4. Discover which team and position have the highest salary expenditure.

Ans:

```
max_salary = df.groupby('Team')['Salary'].max()
plt.figure(figsize=(8, 8))
plt.pie(max_salary, labels=max_salary.index, autopct='%1.1f%%', startangle=140)
```

```
plt.title('Maximum Salary by Team')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```



From the graph above we can see, Cleveland Cavaliers have the highest expenditure,

```
maxCleveland = g.get_group("Cleveland Cavaliers")
max_salary_position = maxCleveland[maxCleveland['Salary'] ==
maxCleveland['Salary'].max()]
max_salary_position
```

	Unnamed: 0	Name	Team	Number	Position	Age	Height	Weight	College	Salary
169	169	LeBron James	Cleveland Cavaliers	23	SF	31	156	250	NaN	22970500.0

From this we know LeBron James have the highest expenditure in Cleveland Cavaliers.