```python
import pandas as pd
import numpy as np



df = pd.read_csv("C:\\Users\\GOKUL\\Downloads\\myexcel.csv.csv")

# Replace "Height" with random values between 150 and 180
np.random.seed(42)  # For reproducibility
df['Height'] = np.random.randint(150, 181, size=df.shape[0])

# Display the updated dataset
df.head()
```

```
          Name            Team  Number Position  Age  Height  Weight
\
0   Avery Bradley  Boston Celtics       0       PG   25     156     180

1     Jae Crowder  Boston Celtics      99       SF   25     169     235

2    John Holland  Boston Celtics      30       SG   27     178     205

3     R.J. Hunter  Boston Celtics      28       SG   22     164     185

4   Jonas Jerebko  Boston Celtics       8       PF   29     160     231


               College       Salary
0                 Texas   7730337.0
1             Marquette   6796117.0
2     Boston University          NaN
3         Georgia State   1148640.0
4                   NaN   5000000.0
```

```python
df.tail(5)
```

```
              Name        Team  Number Position  Age  Height  Weight
College  \
453   Shelvin Mack   Utah Jazz       8       PG   26     168     203
Butler
454      Raul Neto   Utah Jazz      25       PG   24     153     179
NaN
455    Tibor Pleiss  Utah Jazz      21        C   26     152     256
NaN
456     Jeff Withey  Utah Jazz      24        C   26     166     231
Kansas
457       Priyanka   Utah Jazz      34        C   25     166     231
Kansas


         Salary
453   2433333.0
```

```
454    900000.0
455   2900000.0
456    947276.0
457    947276.0
```

Determine the distribution of employees across each team and calculate the percentage

```python
# Distribution of employees across each team
team_distribution = df['Team'].value_counts()
team_percentage = (team_distribution / len(df)) * 100

print(team_distribution)
print(team_percentage)
```

```
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
Los Angeles Clippers      15
Golden State Warriors     15
Toronto Raptors           15
Philadelphia 76ers        15
Dallas Mavericks          15
Orlando Magic             14
Minnesota Timberwolves    14
Name: count, dtype: int64
Team
```

```
New Orleans Pelicans      4.148472
Memphis Grizzlies         3.930131
Utah Jazz                 3.493450
New York Knicks           3.493450
Milwaukee Bucks           3.493450
Brooklyn Nets             3.275109
Portland Trail Blazers    3.275109
Oklahoma City Thunder     3.275109
Denver Nuggets            3.275109
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
```
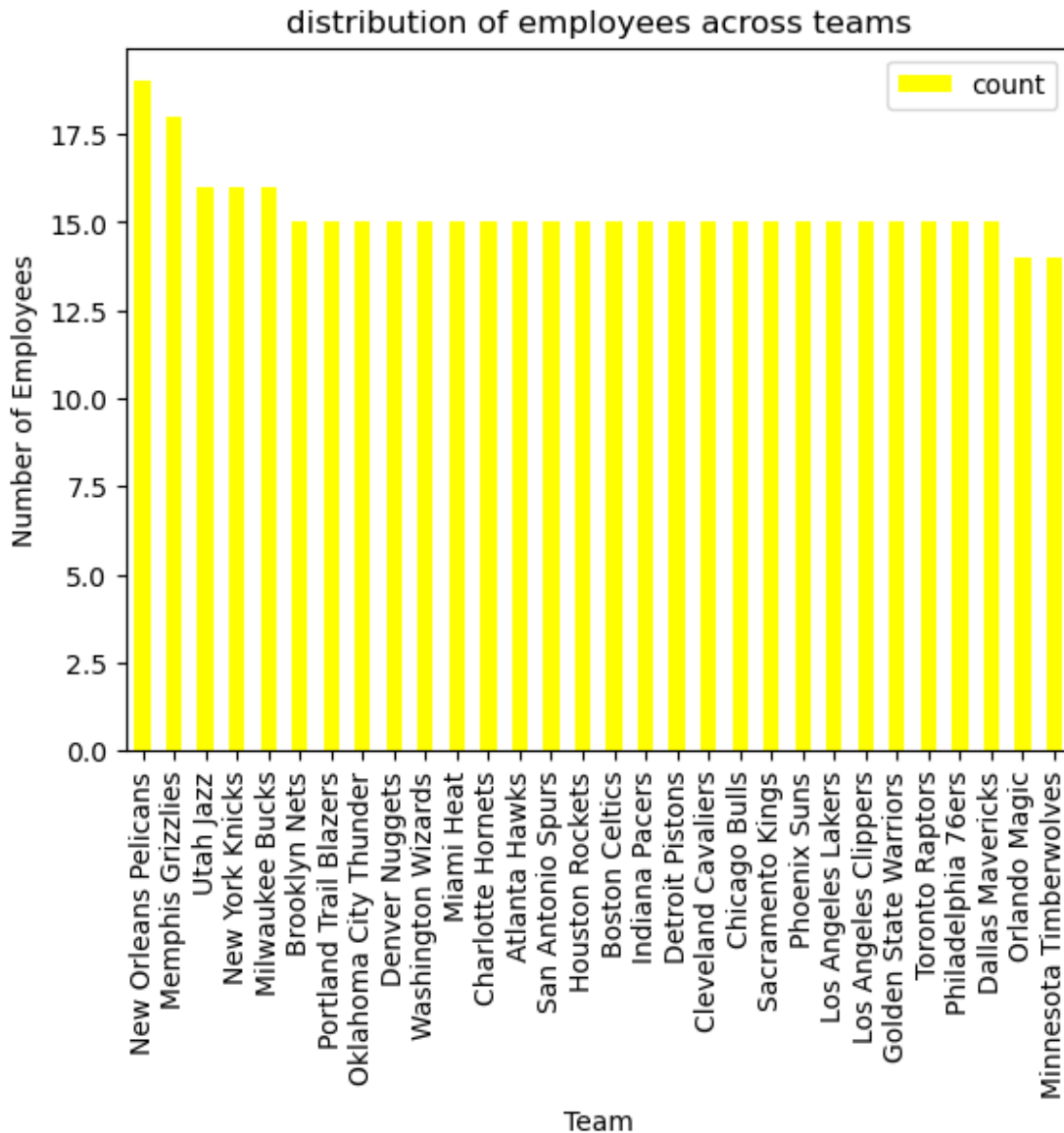
The above data showing which teams are the largest and smallest, indicating potential areas of focus or staffing need .

```python
import matplotlib.pyplot as plt


team_distribution.plot(kind='bar',title='distribution of employees across teams',color='yellow')
plt.xlabel('Team')
plt.ylabel('Number of Employees')
plt.legend()
plt.show()
```

distribution of employees across teams

Segregate employees based on their positions within the company.

```python
employees_distribution = df.groupby('Position')['Name'].apply(list)
for Position, Names in employees_distribution.items():
    print(f"Employees in {Position} Position:")
    for name in Names:
        print(name)
    print("\n")
```

```
Employees in C Position:
Kelly Olynyk
Jared Sullinger
```

Tyler Zeller
Brook Lopez
Henry Sims
Robin Lopez
Kevin Seraphin
Joel Embiid
Jahlil Okafor
Bismack Biyombo
Lucas Nogueira
Jonas Valanciunas
Andrew Bogut
Festus Ezeli
Marreese Speights
Cole Aldrich
DeAndre Jordan
Tarik Black
Roy Hibbert
Robert Sacre
Tyson Chandler
Alex Len
Alan Williams
Willie Cauley-Stein
DeMarcus Cousins
Kosta Koufos
Pau Gasol
Joakim Noah
Sasha Kaun
Timofey Mozgov
Tristan Thompson
Joel Anthony
Aron Baynes
Andre Drummond
Jordan Hill
Ian Mahinmi
Greg Monroe
Miles Plumlee
JaVale McGee
Salah Mejri
Zaza Pachulia
Dwight Howard
Josh Smith
Marc Gasol
Alexis Ajinca
Omer Asik
Kendrick Perkins
Matt Bonner
Boris Diaw
Tim Duncan
Boban Marjanovic

Al Horford
Tiago Splitter
Walter Tavares
Al Jefferson
Frank Kaminsky III
Cody Zeller
Hassan Whiteside
Dewayne Dedmon
Nikola Vucevic
Marcin Gortat
JJ Hickson
Nene Hilario
Nikola Jokic
Joffrey Lauvergne
Jusuf Nurkic
Gorgui Dieng
Nikola Pekovic
Karl-Anthony Towns
Steven Adams
Enes Kanter
Nazr Mohammed
Ed Davis
Chris Kaman
Mason Plumlee
Rudy Gobert
Tibor Pleiss
Jeff Withey
Priyanka

Employees in PF Position:
Jonas Jerebko
Amir Johnson
Jordan Mickey
Chris McCullough
Willie Reed
Thomas Robinson
Thaddeus Young
Lou Amundson
Kyle O'Quinn
Kristaps Porzingis
Derrick Williams
Elton Brand
Richaun Holmes
Carl Landry
Nerlens Noel
Christian Wood
James Johnson
Patrick Patterson

```
Luis Scola
Jason Thompson
Draymond Green
Anderson Varejao
Jeff Ayres
Blake Griffin
Luc Richard Mbah a Moute
Brandon Bass
Ryan Kelly
Larry Nance Jr.
Julius Randle
Jon Leuer
Mirza Teletovic
Duje Dukan
Eric Moreland
Cameron Bairstow
Cristiano Felicio
Taj Gibson
Nikola Mirotic
Bobby Portis
Channing Frye
Kevin Love
Marcus Morris
Anthony Tolliver
Lavoy Allen
Rakeem Christmas
Myles Turner
Shayne Whittington
John Henson
Johnny O'Bryant III
Jabari Parker
David Lee
Dirk Nowitzki
Dwight Powell
Charlie Villanueva
Clint Capela
Montrezl Harrell
Terrence Jones
Donatas Motiejunas
Chris Andersen
JaMychal Green
Jarell Martin
Zach Randolph
Alex Stepheson
Brandan Wright
Ryan Anderson
Dante Cunningham
Anthony Davis
LaMarcus Aldridge
```

David West
Kris Humphries
Paul Millsap
Mike Muscala
Mike Scott
Tyler Hansbrough
Spencer Hawes
Marvin Williams
Chris Bosh
Udonis Haslem
Josh McRoberts
Amar'e Stoudemire
Aaron Gordon
Ersan Ilyasova
Andrew Nicholson
Jason Smith
Drew Gooden
Markieff Morris
Darrell Arthur
Kenneth Faried
Nemanja Bjelica
Kevin Garnett
Adreian Payne
Greg Smith
Nick Collison
Serge Ibaka
Mitch McGary
Cliff Alexander
Meyers Leonard
Noah Vonleh
Trevor Booker
Derrick Favors
Trey Lyles

Employees in PG Position:
Avery Bradley
Terry Rozier
Marcus Smart
Isaiah Thomas
Jarrett Jack
Shane Larkin
Donald Sloan
Jose Calderon
Jerian Grant
Isaiah Canaan
Kendall Marshall
T.J. McConnell
Ish Smith

```
Cory Joseph
Kyle Lowry
Delon Wright
Stephen Curry
Shaun Livingston
Chris Paul
Pablo Prigioni
Austin Rivers
Jordan Clarkson
Marcelo Huertas
D'Angelo Russell
Eric Bledsoe
Brandon Knight
Phil Pressey
Ronnie Price
Darren Collison
Rajon Rondo
Aaron Brooks
Derrick Rose
Matthew Dellavedova
Kyrie Irving
Mo Williams
Steve Blake
Lorenzo Brown
Spencer Dinwiddie
Reggie Jackson
George Hill
Ty Lawson
Rodney Stuckey
Joe Young
Jerryd Bayless
Michael Carter-Williams
Tyler Ennis
Greivis Vasquez
J.J. Barea
Raymond Felton
Devin Harris
Deron Williams
Patrick Beverley
Andrew Goudelock
Mike Conley
Bryce Cotton
Jordan Farmar
Ray McCallum
Xavier Munford
Norris Cole
Toney Douglas
Tim Frazier
Jrue Holiday
```

Andre Miller
Patty Mills
Tony Parker
Dennis Schroder
Jeff Teague
Jorge Gutierrez
Jeremy Lin
Kemba Walker
Goran Dragic
Briante Weber
Brandon Jennings
Shabazz Napier
Elfrid Payton
C.J. Watson
Ramon Sessions
John Wall
D.J. Augustin
Emmanuel Mudiay
Jameer Nelson
Tyus Jones
Zach LaVine
Ricky Rubio
Cameron Payne
Russell Westbrook
Damian Lillard
Brian Roberts
Trey Burke
Dante Exum
Shelvin Mack
Raul Neto

Employees in SF Position:
Jae Crowder
Thanasis Antetokounmpo
Carmelo Anthony
Cleanthony Early
Lance Thomas
Robert Covington
Jerami Grant
Bruno Caboclo
DeMarre Carroll
Terrence Ross
Harrison Barnes
Andre Iguodala
Kevon Looney
James Michael McAdoo
Brandon Rush
Branden Dawson

```
Jeff Green
Wesley Johnson
Paul Pierce
Anthony Brown
Kobe Bryant
Metta World Peace
Nick Young
Chase Budinger
P.J. Tucker
T.J. Warren
Quincy Acy
Caron Butler
Omri Casspi
Rudy Gay
Doug McDermott
Tony Snell
LeBron James
Richard Jefferson
Reggie Bullock
Tobias Harris
Darrun Hilliard
Stanley Johnson
Paul George
Solomon Hill
C.J. Miles
Giannis Antetokounmpo
Damien Inglis
Steve Novak
Jeremy Evans
Chandler Parsons
Trevor Ariza
Michael Beasley
Sam Dekker
Matt Barnes
P.J. Hairston
Lance Stephenson
Luke Babbitt
James Ennis
Alonzo Gee
Quincy Pondexter
Kyle Anderson
Kawhi Leonard
Kent Bazemore
Thabo Sefolosha
Michael Kidd-Gilchrist
Luol Deng
Gerald Green
Joe Johnson
Justise Winslow
```

Dorell Wright
Devyn Marble
Jared Dudley
Kelly Oubre Jr.
Otto Porter Jr.
Marcus Thornton
Will Barton
Wilson Chandler
Danilo Gallinari
Shabazz Muhammad
Tayshaun Prince
Damjan Rudez
Kevin Durant
Josh Huestis
Kyle Singler
Al-Farouq Aminu
Maurice Harkless
Gordon Hayward
Joe Ingles
Chris Johnson

Employees in SG Position:
John Holland
R.J. Hunter
Evan Turner
James Young
Bojan Bogdanovic
Markel Brown
Wayne Ellington
Rondae Hollis-Jefferson
Sergey Karasev
Sean Kilpatrick
Arron Afflalo
Langston Galloway
Sasha Vujacic
Tony Wroten
Nik Stauskas
Hollis Thompson
DeMar DeRozan
Norman Powell
Leandro Barbosa
Ian Clark
Klay Thompson
Jamal Crawford
JJ Redick
C.J. Wilcox
Louis Williams
Devin Booker

Archie Goodwin
John Jenkins
James Anderson
Marco Belinelli
Seth Curry
Ben McLemore
Jimmy Butler
Mike Dunleavy
Justin Holiday
E'Twaun Moore
Dahntay Jones
James Jones
Jordan McRae
Iman Shumpert
J.R. Smith
Kentavious Caldwell-Pope
Jodie Meeks
Monta Ellis
Glenn Robinson III
Jared Cunningham
O.J. Mayo
Khris Middleton
Rashad Vaughn
Justin Anderson
Wesley Matthews
Corey Brewer
James Harden
K.J. McDaniels
Jason Terry
Jordan Adams
Tony Allen
Vince Carter
Bryce Dejean-Jones
Tyreke Evans
Eric Gordon
Jordan Hamilton
Orlando Johnson
Manu Ginobili
Danny Green
Kevin Martin
Jonathon Simmons
Tim Hardaway Jr.
Kirk Hinrich
Kyle Korver
Lamar Patterson
Nicolas Batum
Troy Daniels
Aaron Harrison
Jeremy Lamb

```
Courtney Lee
Tyler Johnson
Josh Richardson
Dwyane Wade
Evan Fournier
Mario Hezonja
Victor Oladipo
Alan Anderson
Bradley Beal
Jarell Eddie
Garrett Temple
Gary Harris
Mike Miller
JaKarr Sampson
Axel Toupane
Andrew Wiggins
Randy Foye
Anthony Morrow
Andre Roberson
Dion Waiters
Pat Connaughton
Allen Crabbe
Gerald Henderson
C.J. McCollum
Luis Montero
Alec Burks
Rodney Hood
```

```python
import pandas as pd
import matplotlib.pyplot as plt

employees_distribution = df.groupby('Position')['Name'].apply(list)

plt.figure(figsize=(5, 2))
plt.bar(employees_distribution.index,
employees_distribution.apply(len), color='black', width=0.5)
plt.xlabel('Position')
plt.ylabel('Number of Employees')
plt.title('Employee Distribution by Position')
plt.xticks(rotation=45)
plt.show()
```

Employee Distribution by Position
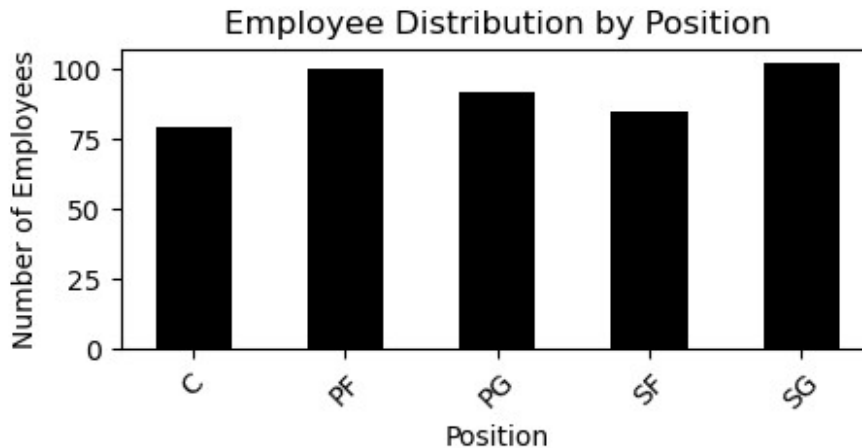
Here employees are segregate by the position of C,PF,PG,SF,SG

Identify the predominant age group among employees.

```
df['Age group']=df['Age'].apply(lambda age:'20-25' if 20<=age<=25 else
('26-30' if 26<= age<=30 else ('31-35' if 31<= age<=35 else'36 and
above')))
df
```

	Name	Team	Number	Position	Age	Height	Weight
0	Avery Bradley	Boston Celtics	0	PG	25	156	180
1	Jae Crowder	Boston Celtics	99	SF	25	169	235
2	John Holland	Boston Celtics	30	SG	27	178	205
3	R.J. Hunter	Boston Celtics	28	SG	22	164	185
4	Jonas Jerebko	Boston Celtics	8	PF	29	160	231
..	...	...	...	...	...	...	..
453	Shelvin Mack	Utah Jazz	8	PG	26	168	203
454	Raul Neto	Utah Jazz	25	PG	24	153	179
455	Tibor Pleiss	Utah Jazz	21	C	26	152	256
456	Jeff Withey	Utah Jazz	24	C	26	166	231
457	Priyanka	Utah Jazz	34	C	25	166	231

```
            College      Salary Age group
```

```
0              Texas   7730337.0      20-25
1          Marquette   6796117.0      20-25
2    Boston University        NaN     26-30
3       Georgia State   1148640.0     20-25
4              NaN     5000000.0      26-30
..              ...         ...        ...
453           Butler   2433333.0      26-30
454           NaN      900000.0       20-25
455           NaN     2900000.0       26-30
456           Kansas    947276.0      26-30
457           Kansas    947276.0      20-25
```

[458 rows x 10 columns]

```python
df['Age group'].value_counts()
```

```
Age group
20-25          198
26-30          167
31-35           68
36 and above    25
Name: count, dtype: int64
```

```python
import pandas as pd
import matplotlib.pyplot as plt

age_groups = pd.Series({
    '20-25': 198,
    '26-30': 167,
    '31-35': 68,
    '36 and above': 25
})

# Create a bar chart
plt.figure(figsize=(5, 2))
plt.bar(age_groups.index, age_groups.values)
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Distribution of Employees by Age Group')
plt.show()
```

Distribution of Employees by Age Group
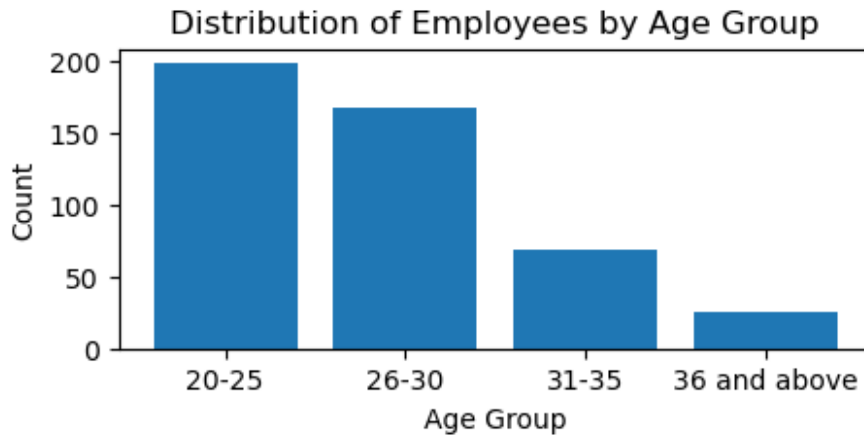

```
20-25 age groups are predominant as per the data.
```

Discover which team and position have the highest salary expenditure

```python
team_salary=df.groupby(['Team','Position'])['Salary'].sum()
team_salary.idxmax()
```

```
('Los Angeles Lakers', 'SF')
```

```python
import pandas as pd
import matplotlib.pyplot as plt

positions = ['C', 'PF', 'PG', 'SF', 'SG']
salaries = [10000000, 20000000, 21000000, 25000000, 15000000]  # SF
has the highest salary

# Create a DataFrame
df = pd.DataFrame({'Position': positions, 'Salary': salaries})

df.plot(x='Position', y='Salary', kind='bar', title='Total Salary
Expenditure by Position', color='purple')
plt.xlabel('Position')
plt.legend()
plt.ylabel('Total Salary')
plt.show()
```

Total Salary Expenditure by Position

```
Highest salary expenditure team is Los Angel Lakers    in the position
of SF.
```

Investigate if there's any correlation between age and salary, and represent it visually

```python
import matplotlib.pyplot as plt
import seaborn as sns

# Scatter plot for Age vs Salary
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Age', y='Salary', data=df)
plt.title('Correlation Between Age and Salary')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```

Correlation Between Age and Salary

Illustrate the relationship between age and salary, the age group 25-35 shows higher salary trend.