LOAD DATA SET

```
import pandas as pd

data=pd.read_csv("/content/myexcel - myexcel.csv.csv")
data
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	06-Feb	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	06-Jun	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	06-May	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	06-May	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	06-Oct	231	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	06-Mar	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25	PG	24	06-Jan	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21	C	26	07-Mar	256	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24	C	26	7-0	231	Kansas	947276.0
457	Priyanka	Utah Jazz	34	C	25	07-Mar	231	Kansas	947276.0

458 rows × 9 columns

Next steps: Generate code with data | View recommended plots | New interactive sheet

```
import numpy as np
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 458 entries, 0 to 457
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      458 non-null    object
 1   Team      458 non-null    object
 2   Number    458 non-null    int64
 3   Position  458 non-null    object
 4   Age       458 non-null    int64
 5   Height    458 non-null    object
 6   Weight    458 non-null    int64
```

```
 7   College   374 non-null    object
 8   Salary    447 non-null    float64
dtypes: float64(1), int64(3), object(5)
memory usage: 32.3+ KB
```

```
#Replace height values

data['Height']=np.random.randint(150,181,size=len(data))
```

```
data
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	height
0	Avery Bradley	Boston Celtics	0	PG	25	171	180	Texas	7730337.0	168
1	Jae Crowder	Boston Celtics	99	SF	25	150	235	Marquette	6796117.0	158
2	John Holland	Boston Celtics	30	SG	27	178	205	Boston University	NaN	179
3	R.J. Hunter	Boston Celtics	28	SG	22	154	185	Georgia State	1148640.0	171
4	Jonas Jerebko	Boston Celtics	8	PF	29	158	231	NaN	5000000.0	169
...	...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	156	203	Butler	2433333.0	158
454	Raul Neto	Utah Jazz	25	PG	24	151	179	NaN	900000.0	158
455	Tibor Pleiss	Utah Jazz	21	C	26	177	256	NaN	2900000.0	154
456	Jeff Withey	Utah Jazz	24	C	26	162	231	Kansas	947276.0	177
457	Priyanka	Utah Jazz	34	C	25	169	231	Kansas	947276.0	153

458 rows × 10 columns

Next steps: Generate code with data | View recommended plots | New interactive sheet

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 458 entries, 0 to 457
Data columns (total 10 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      458 non-null    object
 1   Team      458 non-null    object
 2   Number    458 non-null    int64
 3   Position  458 non-null    object
 4   Age       458 non-null    int64
 5   Height    458 non-null    int64
 6   Weight    458 non-null    int64
```

```
 7   College   374 non-null    object
 8   Salary    447 non-null    float64
 9   height    458 non-null    int64
dtypes: float64(1), int64(5), object(4)
memory usage: 35.9+ KB
```

```
data.isnull().sum()
```

	0
Name	0
Team	0
Number	0
Position	0
Age	0
Height	0
Weight	0
College	84
Salary	11
height	0

```
data.count
```

```
pandas.core.frame.DataFrame.count
def count(axis: Axis=0, numeric_only: bool=False)
Parameters
----------
axis : {0 or 'index', 1 or 'columns'}, default 0
    If 0 or 'index' counts are generated for each column.
    If 1 or 'columns' counts are generated for each row.
numeric_only : bool, default False
```

```
data.describe()
```

	Number	Age	Height	Weight	Salary	height
count	458.000000	458.000000	458.000000	458.000000	4.470000e+02	458.000000
mean	17.713974	26.934498	165.085153	221.543668	4.833970e+06	164.860262
std	15.966837	4.400128	8.738566	26.343200	5.226620e+06	9.076752
min	0.000000	19.000000	150.000000	161.000000	3.088800e+04	150.000000
25%	5.000000	24.000000	158.000000	200.000000	1.025210e+06	157.000000
50%	13.000000	26.000000	165.000000	220.000000	2.836186e+06	165.000000
75%	25.000000	30.000000	172.000000	240.000000	6.500000e+06	173.000000
max	99.000000	40.000000	180.000000	307.000000	2.500000e+07	180.000000

```
data.duplicated().sum()
```

```
0
```

```
data.drop_duplicates()
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	height
0	Avery Bradley	Boston Celtics	0	PG	25	171	180	Texas	7730337.0	168
1	Jae Crowder	Boston Celtics	99	SF	25	150	235	Marquette	6796117.0	158
2	John Holland	Boston Celtics	30	SG	27	178	205	Boston University	NaN	179
3	R.J. Hunter	Boston Celtics	28	SG	22	154	185	Georgia State	1148640.0	171
4	Jonas Jerebko	Boston Celtics	8	PF	29	158	231	NaN	5000000.0	169
...	...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	156	203	Butler	2433333.0	158
454	Raul Neto	Utah Jazz	25	PG	24	151	179	NaN	900000.0	158
455	Tibor Pleiss	Utah Jazz	21	C	26	177	256	NaN	2900000.0	154
456	Jeff Withey	Utah Jazz	24	C	26	162	231	Kansas	947276.0	177
457	Priyanka	Utah Jazz	34	C	25	169	231	Kansas	947276.0	153

458 rows × 10 columns

```
data.index
```

```
RangeIndex(start=0, stop=458, step=1)
```

**TASK 1 **

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
```

```
data['Team'].value_counts()
```

Team	count
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

```
# %spliting w.r.t total employees

data['Team'].value_counts()/len(data)*100
```

Team	count
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

TASK 2

```
employees=data.groupby('Position')['Name'].apply(list)
for Position,Names in employees.items():
    print(f"employees in{Position}positions:")
    for name in Names:
        print("\n",name)
```

```
Andre Roberson

Dion Waiters

Pat Connaughton

Allen Crabbe

Gerald Henderson

C.J. McCollum

Luis Montero

Alec Burks

Rodney Hood
```

TASK 3

```
data['Age Group']=data['Age'].apply(lambda age:'20-29' if 20 <= age <30 else 0)
```

```
data
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	height	Age Group
0	Avery Bradley	Boston Celtics	0	PG	25	171	180	Texas	7730337.0	168	20-29
1	Jae Crowder	Boston Celtics	99	SF	25	150	235	Marquette	6796117.0	158	20-29
2	John Holland	Boston Celtics	30	SG	27	178	205	Boston University	NaN	179	20-29
3	R.J. Hunter	Boston Celtics	28	SG	22	154	185	Georgia State	1148640.0	171	20-29
4	Jonas Jerebko	Boston Celtics	8	PF	29	158	231	NaN	5000000.0	169	20-29
...	...	...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	156	203	Butler	2433333.0	158	20-29
454	Raul Neto	Utah Jazz	25	PG	24	151	179	NaN	900000.0	158	20-29
455	Tibor Pleiss	Utah Jazz	21	C	26	177	256	NaN	2900000.0	154	20-29
456	Jeff Withey	Utah Jazz	24	C	26	162	231	Kansas	947276.0	177	20-29
457	Priyanka	Utah Jazz	34	C	25	169	231	Kansas	947276.0	153	20-29

458 rows × 11 columns

Next steps: Generate code with `data` View recommended plots New interactive sheet

```
data['Age Group'].value_counts()
```

Age Group	count
20-29	334
0	124

TASK 4

```
spending_salary=data.groupby(['Team','Position'])['Salary'].sum()
spending_salary.idxmax()
```

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

TASK 5

```
correlation=data['Salary'].corr(data['Age'])
```

```
correlation
```

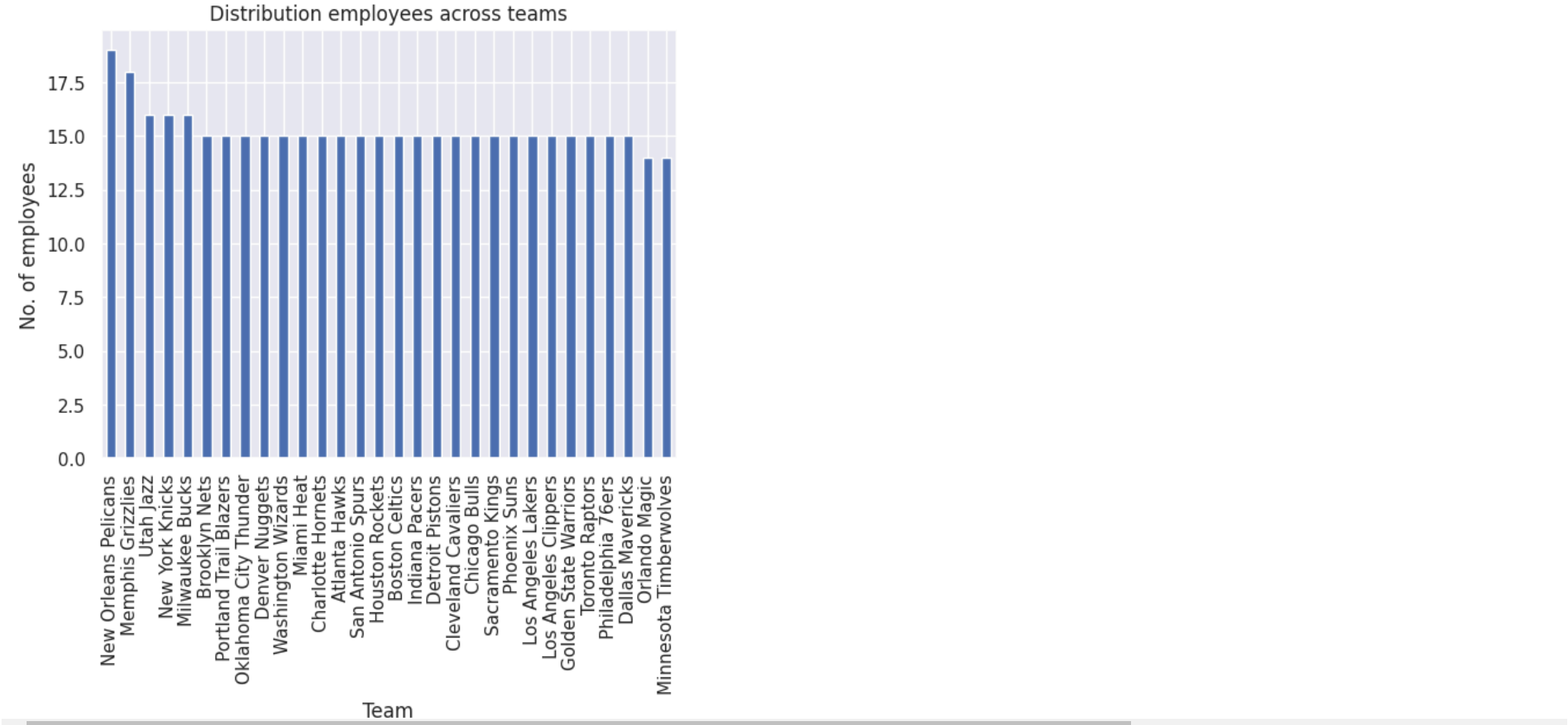
```
0.21400941226570974
```

```
sns.scatterplot(x="Age",y="Salary",data=data)
plt.ylabel("Salary")
plt.xlabel("Age")
plt.title("Correlation b/w salary and age")
plt.show()
```

GRAPHICAL REPRESENTATION

```
#TASK 1


data['Team'].value_counts().plot(kind="bar")
plt.title('Distribution employees across teams')
plt.xlabel("Team")
plt.ylabel("No. of employees")
plt.show()
```

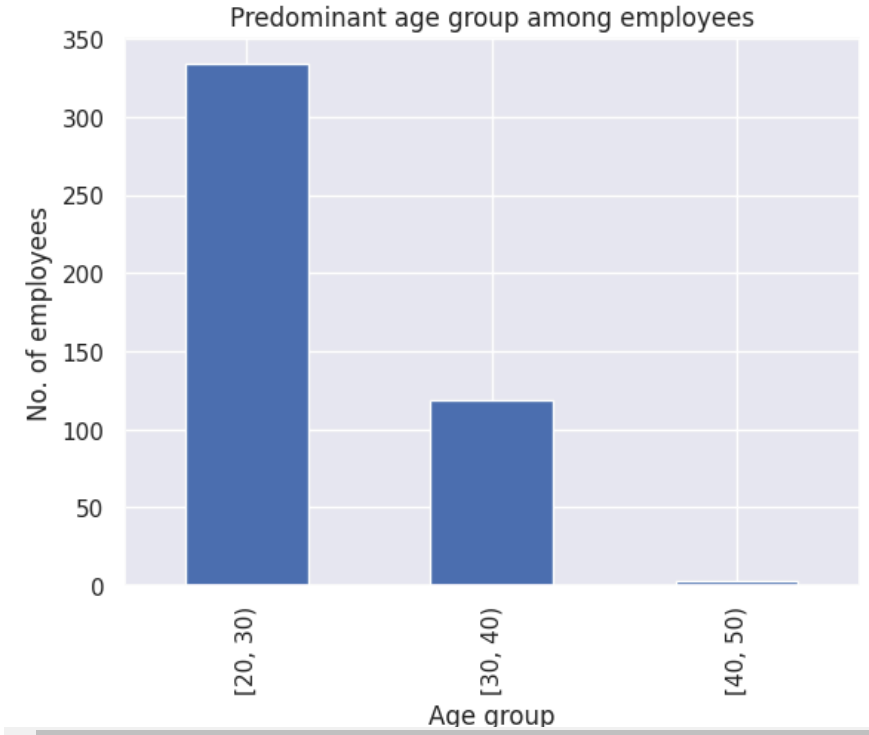



```
#TASK 2


position_dist=data['Position'].value_counts()
position_dist.plot(kind="bar")
plt.title('Employees segregation based on position')
plt.xlabel('Postion')
plt.ylabel('No. of employees')
plt.show()
```

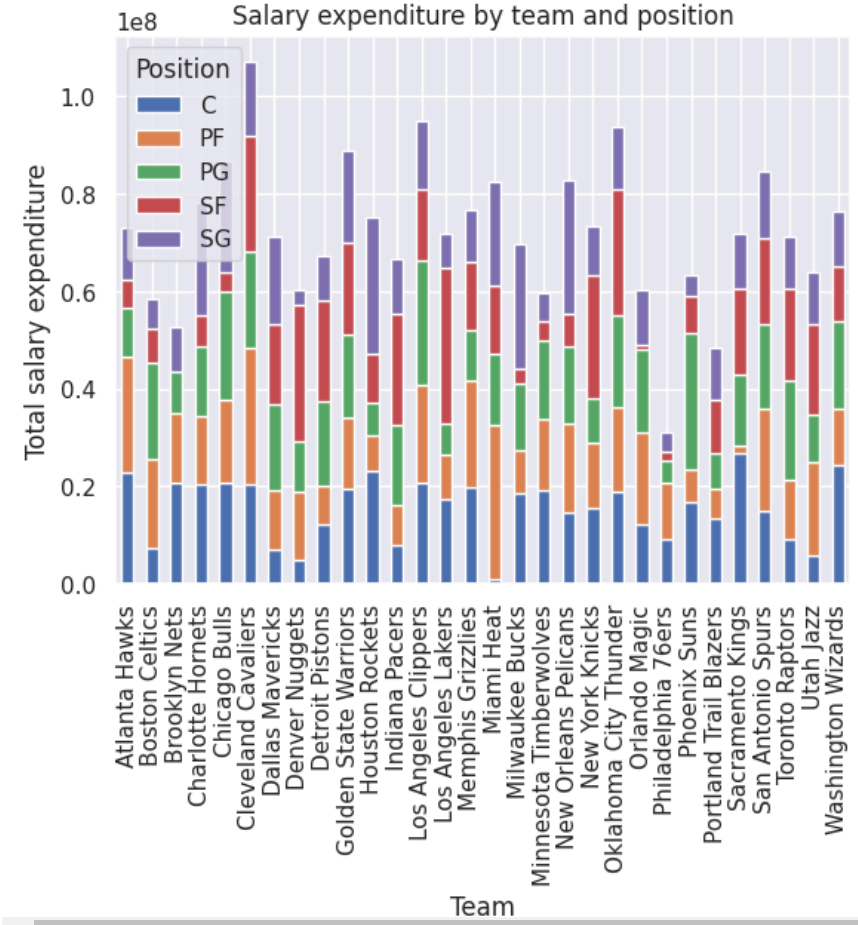

```
#TASK 3


age_group=pd.cut(data['Age'],bins=[20,30,40,50],right=False)
age_group_dist=age_group.value_counts()
age_group_dist.plot(kind="bar")
plt.title('Predominant age group among employees')
plt.xlabel('Age group')
plt.ylabel('No. of employees')
plt.show()
```




```
#TASK 4


spending_salary.unstack().plot(kind="bar",stacked=True)
plt.title("Salary expenditure by team and position")
plt.xlabel("Team")
plt.ylabel("Total salary expenditure")
plt.show()
```




```
#TASK 5


sns.regplot(x="Age",y="Salary",data=data)
plt.title("correlation b/w salary and age")
plt.xlabel("Age")
plt.ylabel("Salary")
plt.show()
```

Start coding or generate with AI.

Start coding or generate with AI.

Start coding or generate with AI.

Start coding or generate with AI.

Start coding or generate with AI.

Start coding or generate with AI.