Python Project

```
[ ]: #As a culminating project, working with a dataset of players, consisting of 458 rows and 9 columns.
     #The tasks include preprocessing the dataset, analyzing the data, and presenting the findings graphically.
```

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.
```

```
[7]: #Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[10]: pip install openpyxl
```

```
Requirement already satisfied: openpyxl in c:\users\hp\anaconda3\lib\site-packages (3.1.2)
Requirement already satisfied: et-xmlfile in c:\users\hp\anaconda3\lib\site-packages (from openpyxl) (1.1.0)
Note: you may need to restart the kernel to use updated packages.
```

```
[8]: #Load the dataset into pandas dataframe
df = pd.read_excel("C:\\Users\\hp\\Downloads\\myexcel.xlsx")
df
```

Home

Project Python

Project Python

localhost:8888/notebooks/Project%20Python.ipynb?

jupyter Project Python Last Checkpoint: 3 hours ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[8]:
```
#Load the dataset into pandas dataframe
df = pd.read_excel("C:\\Users\\hp\\Downloads\\myexcel.xlsx")
df
```

[8]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	2023-02-06 00:00:00	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	2023-06-06 00:00:00	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	2023-05-06 00:00:00	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	2023-05-06 00:00:00	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	2023-10-06 00:00:00	231	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	2023-03-06 00:00:00	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25	PG	24	2023-01-06 00:00:00	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21	C	26	2023-03-07 00:00:00	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	2023-03-07 00:00:00	231	Kansas	947276.0

458 rows × 9 columns

30°C
Mostly sunny

Search

ENG
IN

14:26
18-12-2024

```
#Correct height column by replacing it with random numbers between 150 and 180
df['Height'] = np.random.randint(150,181, size=len(df))
df
```

[48]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	175	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	167	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	178	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	166	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	174	231	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	179	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25	PG	24	172	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21	C	26	178	256	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24	C	26	177	231	Kansas	947276.0
457	Priyanka	Utah Jazz	34	C	25	177	231	Kansas	947276.0

458 rows × 9 columns

```
[50]: #Ensure data consistency and integrity by checking missing values ,duplicates and data types
print("Missing values:",df.isnull().sum())
print("Duplicates:",df.duplicated().sum())
print("Data types:" ,df.dtypes)
```

```
Missing values: Name         0
Team         0
Number       0
Position     0
Age          0
Height       0
Weight       0
College     84
Salary      11
dtype: int64
Duplicates: 0
Data types: Name         object
Team         object
Number        int64
Position     object
Age           int64
Height        int32
Weight        int64
College      object
Salary      float64
dtype: object
```

[56]:

```python
#Drop any rows with missing values
df.dropna(inplace=True)
df
```

[56]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	175	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	167	235	Marquette	6796117.0
3	R.J. Hunter	Boston Celtics	28	SG	22	166	185	Georgia State	1148640.0
6	Jordan Mickey	Boston Celtics	55	PF	21	179	235	LSU	1170960.0
7	Kelly Olynyk	Boston Celtics	41	C	25	168	238	Gonzaga	2165160.0
...	...	...	...	...	...	...	...	...	...
451	Chris Johnson	Utah Jazz	23	SF	26	163	206	Dayton	981348.0
452	Trey Lyles	Utah Jazz	41	PF	20	160	234	Kentucky	2239800.0
453	Shelvin Mack	Utah Jazz	8	PG	26	179	203	Butler	2433333.0
456	Jeff Withey	Utah Jazz	24	C	26	177	231	Kansas	947276.0
457	Priyanka	Utah Jazz	34	C	25	177	231	Kansas	947276.0

365 rows × 9 columns

ANALYSIS TASKS

Task-1 : Distribution of Players across each team

[92]:
```
#Determine the distribution of players across each team and calculate the percentage split relative to the total no:of Players
team_distribution = df['Team'].value_counts(normalize=True) * 100
team_distribution
```

[92]:
```
Team
New Orleans Pelicans      4.383562
Portland Trail Blazers    4.109589
Detroit Pistons           4.109589
Milwaukee Bucks           3.835616
Philadelphia 76ers        3.835616
Oklahoma City Thunder     3.835616
Los Angeles Clippers      3.835616
Washington Wizards        3.561644
Charlotte Hornets         3.561644
Phoenix Suns              3.561644
Sacramento Kings          3.561644
Memphis Grizzlies         3.561644
Brooklyn Nets             3.561644
Boston Celtics            3.287671
Dallas Mavericks          3.287671
Indiana Pacers            3.287671
Chicago Bulls             3.287671
Los Angeles Lakers        3.287671
```

Home
Project Python
Project Python

localhost:8888/notebooks/Project%20Python.ipynb?

jupyter Project Python Last Checkpoint: 3 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

```
Golden State Warriors     3.287671
Houston Rockets           3.013699
Cleveland Cavaliers       3.013699
San Antonio Spurs         3.013699
Atlanta Hawks             3.013699
New York Knicks           3.013699
Utah Jazz                 3.013699
Miami Heat                2.739726
Orlando Magic             2.739726
Toronto Raptors           2.739726
Denver Nuggets            2.465753
Minnesota Timberwolves    2.191781
Name: proportion, dtype: float64
```

Task-2: Segregation of players by position

•[26]:
```
#Segregate players based on their positions
Position_Segregation = df.groupby('Position') ['Name'].count()
Position_Segregation
```

[26]:
```
Position
C      79
PF    100
PG     92
SF     85
SG    102
Name: Name, dtype: int64
```

30°C Mostly sunny Search ENG IN 14:27 18-12-2024

Home
Project Python
Project Python

localhost:8888/notebooks/Project%20Python.ipynb?

jupyter Project Python Last Checkpoint: 3 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

Task-3: Predominant Age Group

•[176]:

```
#Identify the predominant age group  among players
Age_group = df.groupby('Age') ['Name'].count()
Age_group
 Predominant_age_group = Age_group.idxmax()
 Predominant_age_group
```

[176]:

```
Age
19     2
20    19
21    19
22    26
23    41
24    47
25    46
26    36
27    41
28    31
29    28
30    31
31    22
32    13
33    14
34    10
35     9
36    10
```

30°C Mostly sunny Search ENG IN 14:27 18-12-2024

```
40      3
Name: Name, dtype: int64
```

Task-4: Team and position with Highest Salary expenditure

[24]:

```
#Discover Which team and Position have the highest salary expenditure
Salary_expenditure = df.groupby(['Team','Position']) ['Salary'].sum()
Salary_expenditure
Highest_Salary_expenditure = Salary_expenditure.idxmax()
Highest_Salary_expenditure
```

[24]:

```
Team               Position
Atlanta Hawks      C           22756250.0
                   PF          23952268.0
                   PG           9763400.0
                   SF           6000000.0
                   SG          10431032.0
                                  ...
Washington Wizards C           24490429.0
                   PF          11300000.0
                   PG          18022415.0
                   SF          11158800.0
                   SG          11356992.0
Name: Salary, Length: 149, dtype: float64
```

Task-5: Correlation between Age and Salary

[86]:

```
#Investigate if there is any correlation between age and salary,and represent it visually
```

Task-5: Correlation between Age and Salary

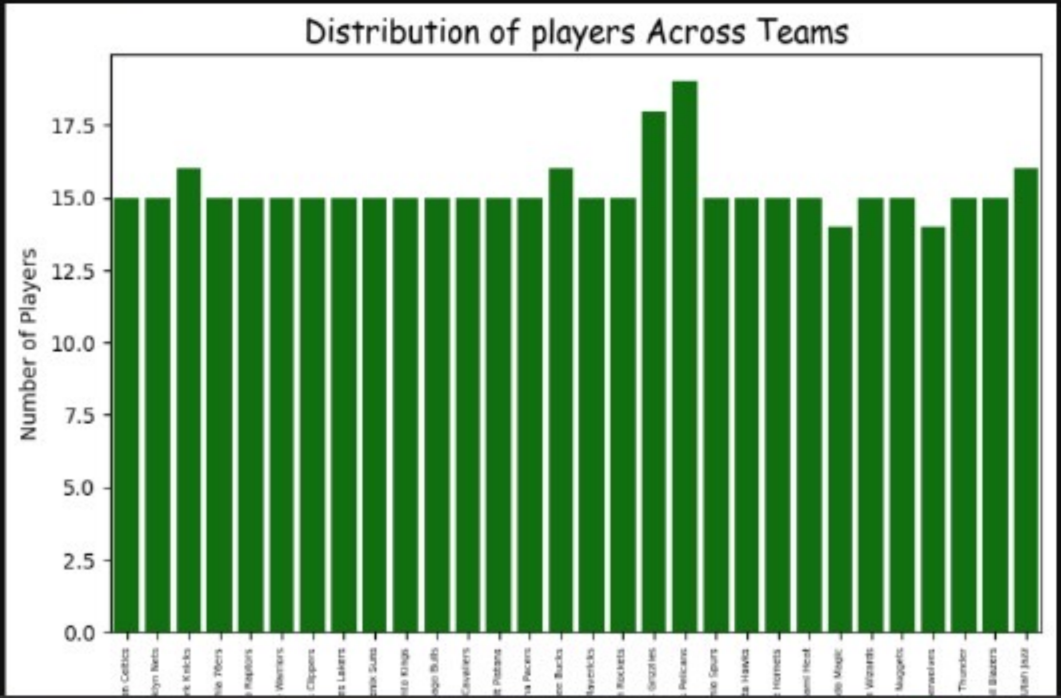
[86]:
```
#Investigate if there is any correlation between age and salary,and represent it visually
correlation=df['Age'].corr(df['Salary'])
correlation
plt.scatter(df['Age'],df['Salary'])
plt.xlabel('Age')
plt.ylabel('Salary')
plt.title('Correlation between Age and Salary' , fontdict={'fontname': 'Comic Sans MS' , 'fontsize': 20})
plt.show()
```

GRAPHICAL REPRESENTATION

[ ]:
```
#For each of five analysis tasks, Create appropriate visualizations to present the findings effectively
```

Distribution of players across each team

[274]:
```
plt.figure(figsize=(8,5))
sns.countplot(x='Team',data=df ,color='Green')
plt.title('Distribution of players Across Teams',fontdict={'fontname': 'Comic Sans MS' , 'fontsize': 15})
plt.xlabel('Team')
plt.ylabel('Number of Players')
plt.xticks(rotation=90, fontsize=5)
plt.show()
```



Home | Project Python | Project Python

localhost:8888/notebooks/Project%20Python.ipynb?

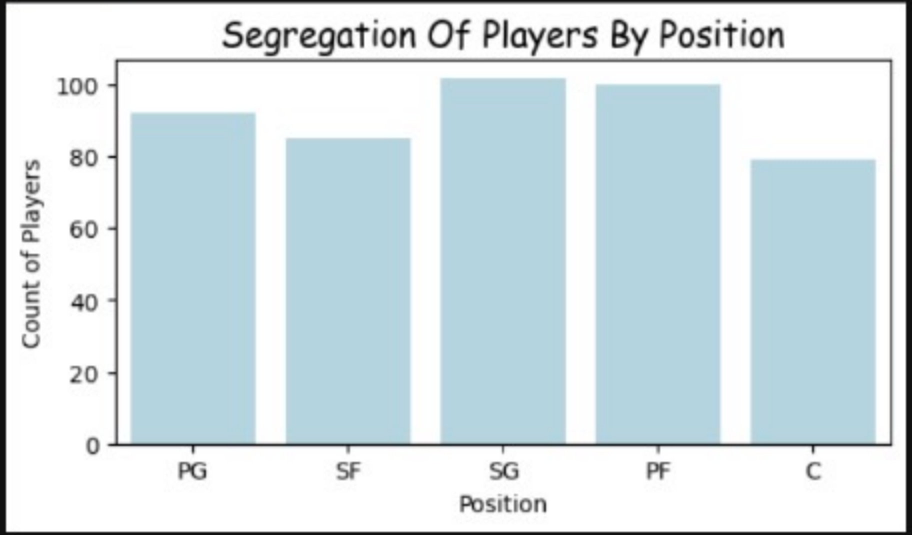
jupyter Project Python Last Checkpoint: 3 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

Segeration of Players by Position

[276]:
```
plt.figure(figsize=(6,3))
sns.countplot(x='Position' , data=df , color='lightblue')
plt.xlabel('Position')
plt.ylabel('Count of Players')
plt.title('Segregation Of Players By Position' , fontdict={'fontname':'Comic Sans MS','fontsize':15 })
plt.show()
```



30°C Mostly sunny Search ENG IN 14:30 18-12-2024

Home | Project Python | Project Python

localhost:8888/notebooks/Project%20Python.ipynb?

jupyter Project Python Last Checkpoint: 3 hours ago

File Edit View Run Kernel Settings Help

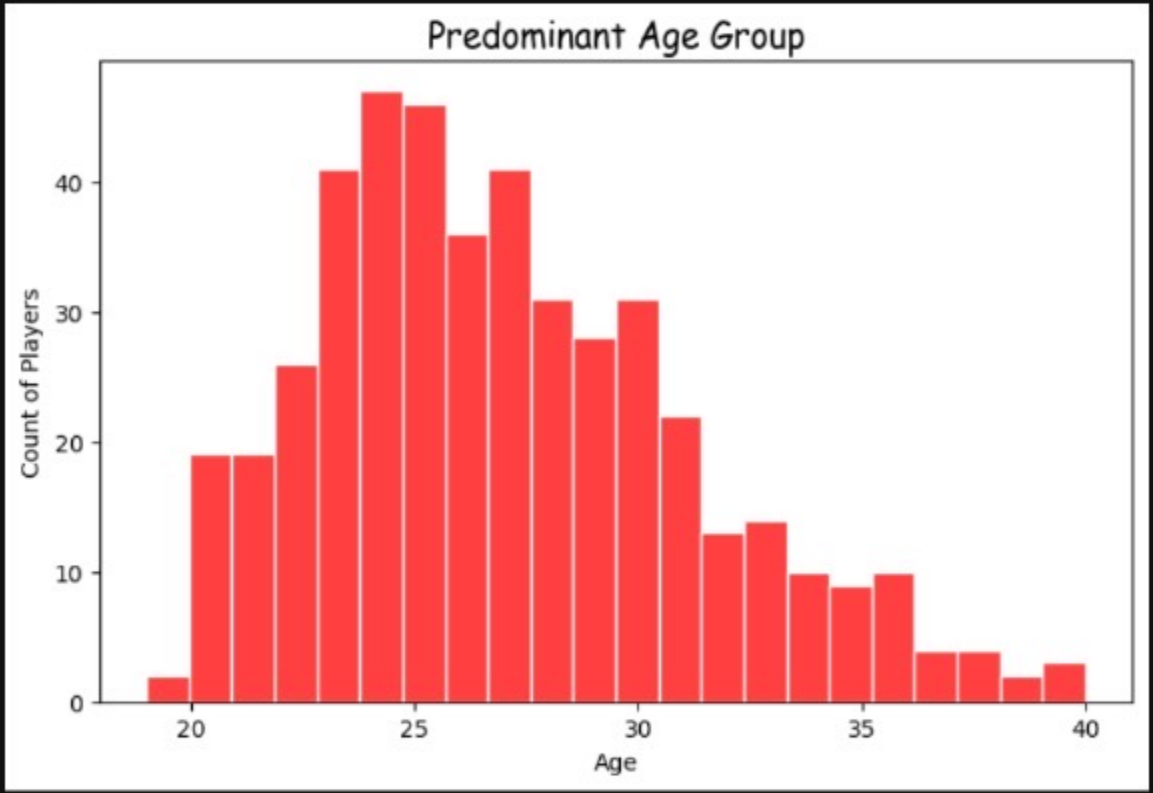
Trusted

Code

JupyterLab Python 3 (ipykernel)

Predominant Age Group

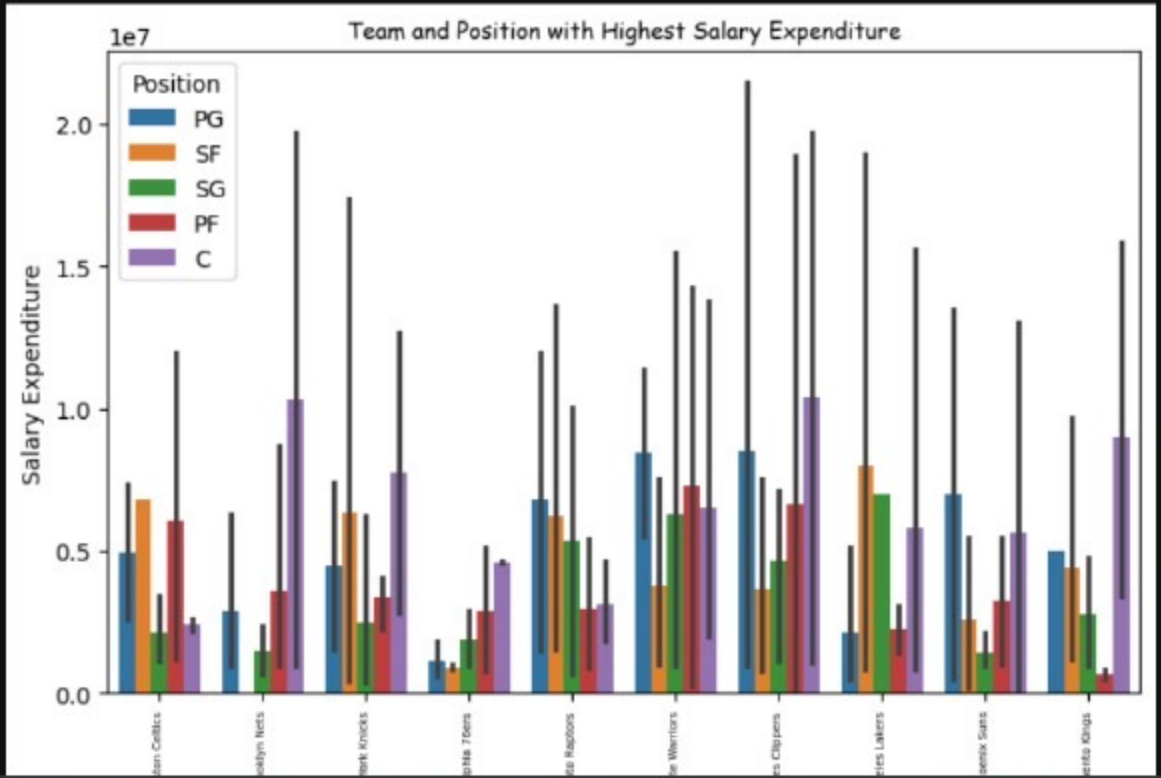
[278]:
```
plt.figure(figsize=(8,5))
sns.histplot(x='Age' , data=df , bins=22 , edgecolor='white',color='red')
plt.ylabel('Count of Players')
plt.title('Predominant Age Group' , fontdict={'fontname':'Comic Sans MS' ,'fontsize': 15})
plt.show()
```



30°C Mostly sunny Search ENG IN 14:30 18-12-2024

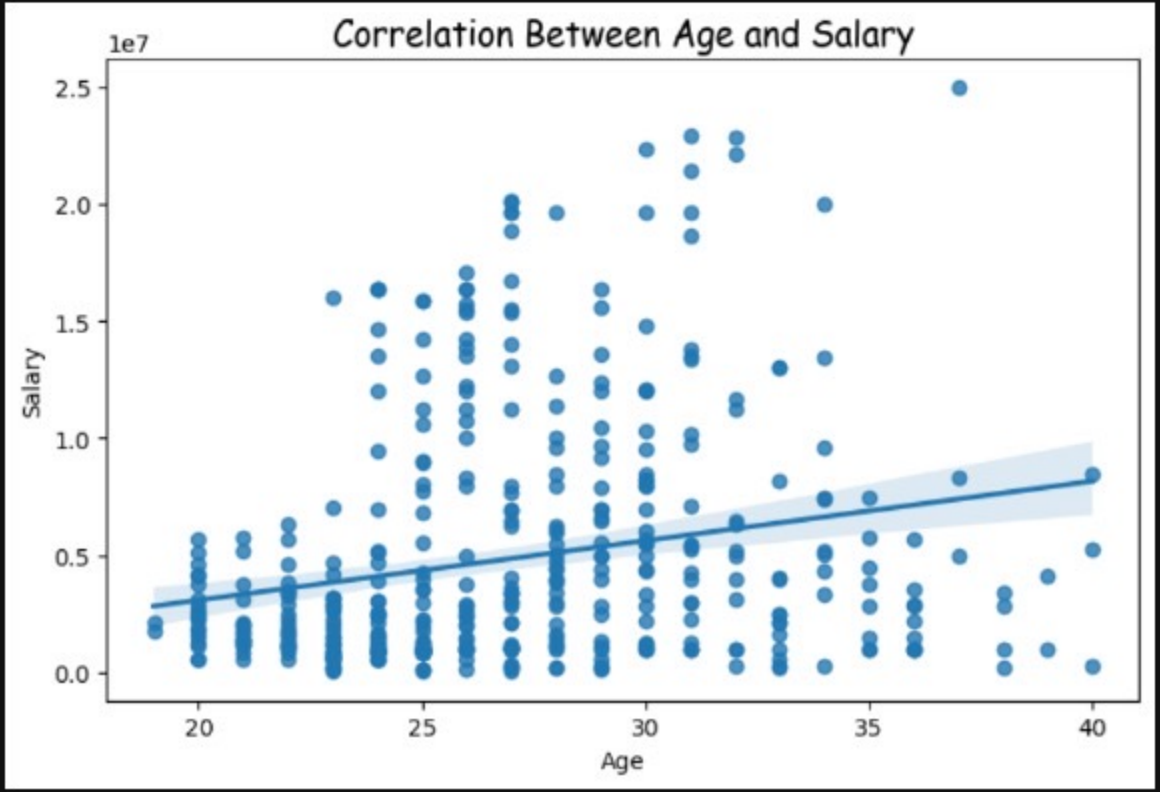
Team and Position with highest Salary Expenditure

[263]:
```python
plt.figure(figsize=(8,5))
sns.barplot(x='Team',y='Salary',hue='Position',data=df.head(150))
plt.xlabel('Team')
plt.ylabel('Salary Expenditure')
plt.xticks(rotation=90,fontsize=5)
plt.title('Team and Position with Highest Salary Expenditure',fontdict={'fontname':'Comic Sans MS','fontsize':10})
plt.show()
```



Correlation between Age and Salary

[268]:
```
plt.figure(figsize=(8,5))
sns.regplot(x='Age',y='Salary',data=df)
plt.title('Correlation Between Age and Salary',fontdict={'fontname':'Comic Sans MS', 'fontsize':15})
plt.show()
```



Home | Project Python | Project Python

localhost:8888/notebooks/Project%20Python.ipynb?

jupyter Project Python Last Checkpoint: 3 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

DATA STORY

[ ]:
```
#Provide Insights gained from the analysis ,Highlighting trends , Patterns and correlations within the dataset.
```

This analysis is to examine trends, patterns, and correlations within the provided basketball players dataset. Key variables such as age, height, weight, salary, and player position are analyzed to identify insights into player demographics, positional trends, and salary distributions. By exploring relationships among these variables, the analysis sheds light on factors influencing player roles, physical attributes, and financial compensation in professional basketball.

TRENDS AND PATTERNS

Age Distribution:
  - The player's are between the ages of 19 to 40.The average age is approximately 27 years, indicating that most players are in their prime playing years. Younger players (aged 20-24) typically include many with less experience, while older players above 30 are often veterans with extensive careers
Salary Distribution:
  - There's a significant disparity in salaries among the players. 25% of players earn below $1 million, while the top earners exceed $20 million. indicating a potential correlation between skill level and compensation.
Position-related Trends:
  - Point Guards (PG) and Shooting Guards (SG) often have higher salaries compared to other positions like Small Forward (SF) and Power Forward (PF). This may reflect the current NBA trend where guards have significant influence over the game's pace and scoring ability.
College Influence:
  - A good number of players in the dataset are from top basketball programs like Duke, Kentucky which suggests a trend of NBA teams favoring players from these colleges, which may indicate strong scouting networks.
Team Composition:
  - Teams vary significantly in their overall salary structure; some, like the Cleveland Cavaliers and Golden State Warriors, have a high concentration of salary expenditure.
Salary Distribution by Position:
Positions like Centers (C) and Shooting Guards (SG) have higher average salaries, likely due to their roles in scoring or defensive plays.

CORRELATION AND INSIGHTS

Experience vs. Salary:
  - There is a strong correlation between experience and salary, where, more experienced players generally command higher salaries.This is likely due to seniority or proven performance over time.

Height/Weight vs. Player Role:
  - The height and weight metrics indicate that taller, heavier players are concentrated in the Center and Power Forward positions, while guards fall into lower weight and height categories, which is reflective of the gameplay's pace and style.

30°C Mostly sunny Search ENG IN 14:30 18-12-2024

This may reflect the current NBA trend where guards have significant influence over the game's pace and scoring ability.
College Influence:
- A good number of players in the dataset are from top basketball programs like Duke, Kentucky which suggests a trend of NBA teams favoring players from these colleges, which may indicate strong scouting networks.

Team Composition:
- Teams vary significantly in their overall salary structure; some, like the Cleveland Cavaliers and Golden State Warriors, have a high concentration of salary expenditure.

Salary Distribution by Position:
Positions like Centers (C) and Shooting Guards (SG) have higher average salaries, likely due to their roles in scoring or defensive plays.

CORRELATION AND INSIGHTS

Experience vs. Salary:
- There is a strong correlation between experience and salary, where, more experienced players generally command higher salaries.This is likely due to seniority or proven performance over time.

Height/Weight vs. Player Role:
- The height and weight metrics indicate that taller, heavier players are concentrated in the Center and Power Forward positions, while guards fall into lower weight and height categories, which is reflective of the gameplay's pace and style.

Age vs. Salary:
- Older players tend to have slightly higher salaries, though the correlation is weak.Players aged 30 or older are likely in the higher salary bracket, reflecting experience

Potential and Production:
- Young players' salaries, particularly those below the age of 24, often reflect potential rather than on-court production. Players like Devin Booker and Karl-Anthony Towns are notable examples of young talent with increasing salaries as they demonstrate their value over time.

Trend of Versatility:
- The emergence of versatile players is notable, where players in the forward positions (SF/PF) are increasingly being able to handle multiple roles on the court. For instance, players like Draymond Green and Giannis Antetokounmpo serve multiple functions, thus impacting how teams shape their rosters.

These findings provide a clearer understanding of player characteristics and salary structures in professional basketball . The dataset reveals a wealth of insights into player demographics, salary trends, and positional importance within the NBA. Relationships between age, experience, height, weight, and salary are critical for roster construction and overall team performance, highlighting the importance of both immediate and long-term strategic thinking in professional basketball.