PROBLEM STATEMENTS ON GIVEN DATASET

1. **What is the highest salary in the dataset?**
2. **How many employees are from Pune?**
3. **How many employees are married?**
4. **What is the total salary expense?**
5. **How many employees have a post of Manager?**
6. **What is the lowest salary in the dataset?**
7. **How many employees are single and earn more than 90,000?**
8. **How many employees have a post of Sr. Manager and earn more than 100,000?**
9. **How many employees have a post of Supervisor and earn more than 80,000?**
10. **What is the average salary of divorced employees?**
11. **How many employees are there in each district?**
12. **How many employees are in each post category?**
13. **How many employees have a post of Manager and earn less than 95,000?**
14. **What is the average salary?**
15. **How many employees are there in each district?**
16. **What is the total salary for each post?**
17. **What is the average salary for each district?**
18. **What is the total salary for each status (single, married, divorced)?**
19. **How many employees have a post starting with "Manager"?**
20. **How many employees have a salary greater than 100,000?**

Reading the CSV File in python

```
[54] import pandas as pd
     data = pd.read_csv('data.csv')
```

What is the highest salary in the dataset?

```
[14] highest_salary = data['salary'].max()
     print("Highest Salary:", highest_salary)
```

```
Highest Salary: 150000
```

How many employees are from Pune?

```
[15] pune_employees = data[data['district'] == 'Pune'].shape[0]
     print("Number of employees from Pune:", pune_employees)
```

```
Number of employees from Pune: 6
```

How many employees are married?

```
[17] married_employees = data[data['status'] == 'married'].shape[0]
     print("Number of married employees:", married_employees)
```

```
Number of married employees: 3
```

What is the total salary expense?

```
[18] total_salary_expense = data['salary'].sum()
     print("Total salary expense:", total_salary_expense)
```

```
Total salary expense: 1052000
```

How many employees have a post of Manager?

```
[19] manager_count = data[data['post'] == 'Manager'].shape[0]
     print("Number of employees with the post of Manager:", manager_count)
```

```
Number of employees with the post of Manager: 4
```

What is the lowest salary in the dataset?

```
lowest_salary = data['salary'].min()
print("Lowest Salary:", lowest_salary)
```

```
Lowest Salary: 85000
```

How many employees are single and earn more than 90,000?

```
[21] single_high_earning_employees = data[(data['status'] == 'single') & (data['salary'] > 90000)].shape[0]
     print("Number of single employees earning more than 90,000:", single_high_earning_employees)
```

```
Number of single employees earning more than 90,000: 4
```

How many employees have a post of Sr. Manager and earn more than 100,000?

```
[23] sr_manager_high_earning_count = data[(data['post'] == 'Sr. Manager') & (data['salary'] > 100000)].shape[0]
     print("Number of employees with the post of Sr. Manager earning more than 100,000:", sr_manager_high_earning_count)
```

```
Number of employees with the post of Sr. Manager earning more than 100,000: 4
```

How many employees have a post of Supervisor and earn more than 80,000?

```
[25] supervisor_high_earning_count = data[(data['post'] == 'Supervisor') & (data['salary'] > 80000)].shape[0]
     print("Number of employees with the post of Supervisor earning more than 80,000:", supervisor_high_earning_count)
```

```
Number of employees with the post of Supervisor earning more than 80,000: 2
```

What is the average salary of divorced employees?

```
[26] divorced_average_salary = data[data['status'] == 'divorced']['salary'].mean()
     print("Average salary of divorced employees:", divorced_average_salary)
```

```
Average salary of divorced employees: 92500.0
```

How many employees are there in each district?

```
[27] district_counts = data['district'].value_counts()
     print("Employee count by district:\n", district_counts)
```

```
Employee count by district:
 Pune      6
Nashik    4
Name: district, dtype: int64
```

How many employees are in each post category?

```
[29] post_counts = data['post'].value_counts()
     print("Employee count by post:\n", post_counts)
```

```
Employee count by post:
 Manager        4
Sr. Manager    4
Supervisor     2
Name: post, dtype: int64
```

How many employees have a post of Manager and earn less than 95,000?

```
[31] manager_low_earning_count = data[(data['post'] == 'Manager') & (data['salary'] < 95000)].shape[0]
     print("Number of employees with the post of Manager earning less than 95,000:", manager_low_earning_count)
```

```
Number of employees with the post of Manager earning less than 95,000: 2
```

What is the average salary?

```
[44] average_salary = data['salary'].mean()
     print("Average Salary:", average_salary)
```

```
Average Salary: 105200.0
```

How many employees are there in each district?

```
[45] employee_count_by_district = data['district'].value_counts()
     print("Employee Count by District:\n", employee_count_by_district)
```

```
Employee Count by District:
 Pune      6
Nashik    4
Name: district, dtype: int64
```

What is the total salary for each post?

```
[46] total_salary_by_post = data.groupby('post')['salary'].sum()
     print("Total Salary by Post:\n", total_salary_by_post)
```

```
Total Salary by Post:
 post
Manager        381000
Sr. Manager    501000
Supervisor     170000
Name: salary, dtype: int64
```

What is the average salary for each district?

```
[47] average_salary_by_district = data.groupby('district')['salary'].mean()
     print("Average Salary by District:\n", average_salary_by_district)
```

```
Average Salary by District:
 district
Nashik     95500.000000
Pune      111666.666667
Name: salary, dtype: float64
```

What is the total salary for each status (single, married, divorced)?

```
total_salary_by_status = data.groupby('status')['salary'].sum()
print("Total Salary by Status:\n", total_salary_by_status)
```

```
Total Salary by Status:
 status
divorced    185000
married     341000
single      526000
Name: salary, dtype: int64
```

How many employees have a post starting with "Manager"?

```
[53] manager_count = data[data['post'].str.startswith('Manager')].shape[0]
     print("Number of Employees with Post starting with 'Manager'are", manager_count)
```

```
Number of Employees with Post starting with 'Manager'are 4
```

How many employees have a salary greater than 100,000?

```
[52] high_salary_employees = data[data['salary'] > 100000]
     high_salary_count = len(high_salary_employees)
     print("Number of Employees with Salary more than 100,000 are", high_salary_count)
```

```
Number of Employees with Salary more than 100,000 are 4
```

data.csv X

1 to 10 of 10 entries Filter

sr.no	name	district	post	salary	Gender	status
1	Sanvi	Pune	Manager	100000	female	single
2	Mrunmayee	Pune	Sr. Manager	150000	male	married
3	Jayesh	Nashik	Manager	90500	male	single
4	Gouri	Nashik	Sr. Manager	100500	female	married
5	Mahesh	Pune	Supervisor	85000	male	single
6	Pranav	Pune	Manager	100000	male	divorced
7	Saksham	Pune	Sr. Manager	150000	male	single
8	Raja	Nashik	Manager	90500	male	married
9	Sunil	Nashik	Sr. Manager	100500	male	single
10	Radha	Pune	Supervisor	85000	female	divorced

Show 10 per page