21162101012_CBA_Yash

Practical 3

Consider the given dataset of the employees of ABC organization. It has the details of the employees working for that organization. You need to find out the below mentioned information from the given dataset.

1. How many entries are there in the employee dataset?
2. How many departments are there in ABC organization?
3. Find out the maximum salary that is given in each department?
4. Find out the detail of the employee who has got the minimum salary in the entire organization?
5. Find out the total salary amount that is given in each department? (salary of employees working in the same department should be added and displayed)
6. Find out how many managers work in the organization?
7. Find out how many employees work in each department?
8. Find out what is the maximum salary that is given to an employee in this organization?
9. Find the details of all the employees whose Job_id = SA_MAN.
10. Find the average salary of each department?
11. Find the number of employees working under every manager in the organization.
12. Extract the name of the employee who has taken the maximum commission.
13. Extract designation-wise maximum salary for the employees.
14. Extract designation-wise total salary amount for the employees.

```
!pip install pandas
import pandas as pd
```

```
file_path = 'data_DMWp3.csv'
df = pd.read_csv(file_path)
print(df)
```

```
     EMPLOYEE_ID FIRST_NAME LAST_NAME     EMAIL  PHONE_NUMBER   HIRE_DATE  \
0            100     Steven      King     SKING  515.123.4567  1987-06-17
1            101      Neena   Kochhar  NKOCHHAR  515.123.4568  1987-06-18
2            102        Lex   De Haan   LDEHAAN  515.123.4569  1987-06-19
3            103  Alexander    Hunold   AHUNOLD  590.423.4567  1987-06-20
4            104      Bruce     Ernst    BERNST  590.423.4568  1987-06-21
..           ...        ...       ...       ...           ...         ...
102          202        Pat       Fay      PFAY  603.123.6666  1987-09-27
103          203      Susan    Mavris   SMAVRIS  515.123.7777  1987-09-28
104          204    Hermann      Baer     HBAER  515.123.8888  1987-09-29
105          205    Shelley   Higgins  SHIGGINS  515.123.8080  1987-09-30
106          206    William     Gietz    WGIETZ  515.123.8181  1987-10-01

          JOB_ID  SALARY  COMMISSION_PCT  MANAGER_ID  DEPARTMENT_ID
0       AD_PRES   24000             0.0           0             90
1         AD_VP   17000             0.0         100             90
2         AD_VP   17000             0.0         100             90
3       IT_PROG    9000             0.0         102             60
4       IT_PROG    6000             0.0         103             60
..          ...     ...             ...         ...            ...
102      MK_REP    6000             0.0         201             20
103      HR_REP    6500             0.0         101             40
104      PR_REP   10000             0.0         101             70
105      AC_MGR   12000             0.0         101            110
106  AC_ACCOUNT    8300             0.0         205            110

[107 rows x 11 columns]
```

```
# Question 1: Number of entries in the employee dataset
num_entries = df.shape[0]
print(f"\n1. Number of entries in the employee dataset: {num_entries}")
```

```
1. Number of entries in the employee dataset: 107
```

```
# Question 2: Number of departments in ABC organization
num_departments = df['DEPARTMENT_ID'].nunique()
print(f"\n2. Number of departments in ABC organization: {num_departments}")
```

```
2. Number of departments in ABC organization: 12
```

```
# Question 3: Maximum salary in each department
max_salary_per_department = df.groupby('DEPARTMENT_ID')['SALARY'].max()
print("\n3. Maximum salary in each department:")
print(max_salary_per_department)
```

```
3. Maximum salary in each department:
DEPARTMENT_ID
0       7000
10      4400
20     13000
30     11000
40      6500
50      8200
60      9000
70     10000
80     14000
90     24000
100    12000
110    12000
Name: SALARY, dtype: int64
```

```
# Question 4: Employee with the minimum salary in the entire organization
min_salary = df['SALARY'].min()
employee_with_min_salary = df[df['SALARY'] == min_salary]
print("\n4. Employee with the minimum salary in the entire organization:")
print(employee_with_min_salary)
```

```
4. Employee with the minimum salary in the entire organization:
    EMPLOYEE_ID FIRST_NAME LAST_NAME    EMAIL  PHONE_NUMBER   HIRE_DATE  \
32          132         TJ     Olson  TJOLSON  650.124.8234  1987-07-19

      JOB_ID  SALARY  COMMISSION_PCT  MANAGER_ID  DEPARTMENT_ID
32  ST_CLERK    2100             0.0         121             50
```

```
# Question 5: Total salary amount given in each department
total_salary_per_department = df.groupby('DEPARTMENT_ID')['SALARY'].sum()
print("\n5. Total salary amount given in each department:")
print(total_salary_per_department)
```

```
5. Total salary amount given in each department:
DEPARTMENT_ID
0        7000
10       4400
20      19000
30      24900
40       6500
50     156400
60      28800
70      10000
80     304500
90      58000
```

```
100     51600
110     20300
Name: SALARY, dtype: int64
```

```
# Question 6: Number of unique managers in the organization
num_unique_managers = df[df['MANAGER_ID'] != 0]['MANAGER_ID'].nunique()
print(f"\n6. Number of unique managers in the organization: {num_unique_managers}")
```

```
6. Number of unique managers in the organization: 18
```

```
# Question 7: Number of employees in each department
employee_count_per_department = df['DEPARTMENT_ID'].value_counts()
print("\n7. Number of employees in each department:")
print(employee_count_per_department)
```

```
7. Number of employees in each department:
50     45
80     34
100     6
30      6
60      5
90      3
20      2
110     2
0       1
10      1
40      1
70      1
Name: DEPARTMENT_ID, dtype: int64
```

```
# Question 8: Maximum salary given to an employee in this organization
max_salary = df['SALARY'].max()
print(f"\n8. Maximum salary given to an employee in this organization: {max_salary}")
```

```
8. Maximum salary given to an employee in this organization: 24000
```

```
# Question 9: Details of employees with job ID 'SA_MAN'
sa_man_employees = df[df['JOB_ID'] == 'SA_MAN']
print("\n9. Details of employees with job ID 'SA_MAN':")
print(sa_man_employees)
```

```
9. Details of employees with job ID 'SA_MAN':
    EMPLOYEE_ID FIRST_NAME  LAST_NAME     EMAIL        PHONE_NUMBER  \
45          145       John    Russell   JRUSSEL  011.44.1344.429268
46          146      Karen   Partners  KPARTNER  011.44.1344.467268
47          147    Alberto  Errazuriz  AERRAZUR  011.44.1344.429278
48          148     Gerald  Cambrault  GCAMBRAU  011.44.1344.619268
49          149      Eleni    Zlotkey  EZLOTKEY  011.44.1344.429018

     HIRE_DATE  JOB_ID  SALARY  COMMISSION_PCT  MANAGER_ID  DEPARTMENT_ID
45  1987-08-01  SA_MAN   14000             0.4         100             80
46  1987-08-02  SA_MAN   13500             0.3         100             80
47  1987-08-03  SA_MAN   12000             0.3         100             80
48  1987-08-04  SA_MAN   11000             0.3         100             80
49  1987-08-05  SA_MAN   10500             0.2         100             80
```

```
# Question 10: Average salary of each department
average_salary_per_department = df.groupby('DEPARTMENT_ID')['SALARY'].mean()
print("\n10. Average salary of each department:")
print(average_salary_per_department)
```

```
10. Average salary of each department:
DEPARTMENT_ID
0       7000.000000
```

```
10      4400.000000
20      9500.000000
30      4150.000000
40      6500.000000
50      3475.555556
60      5760.000000
70     10000.000000
80      8955.882353
90     19333.333333
100     8600.000000
110    10150.000000
Name: SALARY, dtype: float64
```

```
# Question 11: Number of employees working under every manager
employees_per_manager = df.groupby('MANAGER_ID')['EMPLOYEE_ID'].nunique()
print("\n11. Number of employees working under every manager:")
print(employees_per_manager)
```

```
11. Number of employees working under every manager:
MANAGER_ID
0       1
100    14
101     5
102     1
103     4
108     5
114     5
120     8
121     8
122     8
123     8
124     8
145     6
146     6
147     6
148     6
149     6
201     1
205     1
Name: EMPLOYEE_ID, dtype: int64
```

```
# Question 12: Name of the employee with the maximum commission
max_commission = df['COMMISSION_PCT'].max()
employee_with_max_commission = df[df['COMMISSION_PCT'] == max_commission]
name_of_employee_with_max_commission = employee_with_max_commission[['FIRST_NAME', 'LAST_NAME']].values[0]
print("\n12. Name of the employee with the maximum commission:", ' '.join(name_of_employee_with_max_commissi
```

```
12. Name of the employee with the maximum commission: John Russell
```

```
# Question 13: Maximum salary for each designation
max_salary_by_designation = df.groupby('JOB_ID')['SALARY'].max()
print("\n13. Maximum salary for each designation:")
print(max_salary_by_designation)
```

```
13. Maximum salary for each designation:
JOB_ID
AC_ACCOUNT     8300
AC_MGR        12000
AD_ASST        4400
AD_PRES       24000
AD_VP         17000
FI_ACCOUNT     9000
FI_MGR        12000
HR_REP         6500
IT_PROG        9000
MK_MAN        13000
MK_REP         6000
PR_REP        10000
```

```
PU_CLERK       3100
PU_MAN        11000
SA_MAN        14000
SA_REP        11500
SH_CLERK       4200
ST_CLERK       3600
ST_MAN         8200
Name: SALARY, dtype: int64
```

```
# Question 14: Total salary amount for each designation
total_salary_by_designation = df.groupby('JOB_ID')['SALARY'].sum()
print("\n14. Total salary amount for each designation:")
print(total_salary_by_designation)
```

```
14. Total salary amount for each designation:
JOB_ID
AC_ACCOUNT      8300
AC_MGR         12000
AD_ASST         4400
AD_PRES        24000
AD_VP          34000
FI_ACCOUNT     39600
FI_MGR         12000
HR_REP          6500
IT_PROG        28800
MK_MAN         13000
MK_REP          6000
PR_REP         10000
PU_CLERK       13900
PU_MAN         11000
SA_MAN         61000
SA_REP        250500
SH_CLERK       64300
ST_CLERK       55700
ST_MAN         36400
Name: SALARY, dtype: int64
```