




```
import pandas as pd
```

```
data=pd.read_csv("/content/CC.csv")
data.head()
```



	Sr No	Name	email	subject	salary	bonus	state	city	year	
0	1	John Doe	john.doe@enron.com	Meeting Reminder	50000	10000	California	Los Angeles	2001	
1	2	Jane Smith	jane.smith@enron.com	Re: Meeting Reminder	60000	15000	New York	New York	2001	
2	3	Mark Jones	mark.jones@enron.com	Project Update	50000	10000	Texas	Dallas	2001	
3	4	Paul Brown	paul.brown@enron.com	Re: Project Update	70000	20000	Florida	Miami	2001	
4	5	Sarah White	sarah.white@enron.com	New Proposal	60000	15000	Illinois	Chicago	2001	


Next steps:

[Generate code with data](#)

 [View recommended plots](#)


[New interactive sheet](#)

```
#1.Print columns names and their corresponding datatypes
print(data.dtypes)
```




```
Sr No      int64
Name      object
email     object
subject   object
salary    int64
bonus     int64
state     object
city      object
year      int64
dtype: object
```

```
#2.Find and print the number of rows and columns
print("Number of rows and columns are:",data.shape)
```



```
Number of rows and columns are: (10, 12)
```

```
#3. List all unique cities present in the dataset
print("All the unique cities present in dataset are:",data['city'].unique())
```



```
All the unique cities present in dataset are: ['Los Angeles' 'New York' 'Dallas' 'Miami' 'Chicago' 'Seattle' 'Las Vegas'
'Portland' 'Denver' 'Boston']
```

#4. Calculate the average salary of all the employees

```
a=(data['salary'].mean())
print("Average salary of all the employees:",a)
```

➦ Average salary of all the employees: 63000.0

#5. find minimum and maximum bonus

```
print("Maximun bonus:",data['bonus'].min())
print("Minimun bonus:",data['bonus'].max())
```

➦ Maximun bonus: 10000  
Minimun bonus: 25000

#6.List all employees with salary greater than \$60,000.

```
b =data[data['salary'] > 60000]
print("All the employees with salary greater than $60,000:",b)
```

➦ All the employees with salary greater than \$60,000:

Sr No	Name	email	subject	salary \
3	4 Paul Brown	<a href="mailto:paul.brown@enron.com">paul.brown@enron.com</a>	Re: Project Update	70000
5	6 Tom Green	<a href="mailto:tom.green@enron.com">tom.green@enron.com</a>	Re: New Proposal	80000
7	8 Chris Black	<a href="mailto:chris.black@enron.com">chris.black@enron.com</a>	Re: Follow Up	70000
9	10 Mike Brown	<a href="mailto:mike.brown@enron.com">mike.brown@enron.com</a>	Re: Final Review	80000

bonus	state	city	year	total_compensation \
3 20000	Florida	Miami	2001	90000
5 25000	Washington	Seattle	2001	105000
7 20000	Oregon	Portland	2001	90000
9 25000	Massachusetts	Boston	2001	105000

salary_percentage	domain
3 11.111111	enron.com
5 12.698413	enron.com
7 11.111111	enron.com
9 12.698413	enron.com

#7. Calculate the total salary paid to employees from New York.

```
total_salary_ny = data[data['state'] == 'New York']['salary'].sum()
print("Total salary paid to employees from New York:",total_salary_ny)
```

➦ Total salary paid to employees from New York: 60000

#8. Group by state and find the average salary.

```
avg_salary_by_state = data.groupby('state')['salary'].mean()
print("Average salary:",avg_salary_by_state)
```



```
Average salary: state
California      50000.0
Colorado        60000.0
Florida         70000.0
Illinois        60000.0
Massachusetts   80000.0
Nevada          50000.0
New York        60000.0
Oregon          70000.0
Texas           50000.0
Washington      80000.0
Name: salary, dtype: float64
```



```
#9. Identify the city with the highest average bonus.
avg_bonus_by_city = data.groupby('city')['bonus'].mean()
highest_bonus_city = avg_bonus_by_city.idxmax()
print("Highest average bonus:", highest_bonus_city)
```



```
Highest average bonus: Boston
```

```
#10. Add a column total_compensation = salary + bonus.
data['total_compensation'] = data['salary'] + data['bonus']
data.head()
```



	Sr No	Name	email	subject	salary	bonus	state	city	year	total_compensation	
0	1	John Doe	john.doe@enron.com	Meeting Reminder	50000	10000	California	Los Angeles	2001	60000	
1	2	Jane Smith	jane.smith@enron.com	Re: Meeting Reminder	60000	15000	New York	New York	2001	75000	
2	3	Mark Jones	mark.jones@enron.com	Project Update	50000	10000	Texas	Dallas	2001	60000	
3	4	Paul Brown	paul.brown@enron.com	Re: Project Update	70000	20000	Florida	Miami	2001	90000	
4	5	Sarah White	sarah.white@enron.com	New Proposal	60000	15000	Illinois	Chicago	2001	75000	

Next steps:

Generate code with data

 View recommended plots

New interactive sheet

```
#11. Find the top 3 employees with highest total compensation.
top_3 = data.sort_values(by='total_compensation', ascending=False).head(3)
print("Top three employees with highest total compensation:", top_3)
```



		Top three employees with highest total compensation:	Sr No	Name	email	subject	salary	bonus	\
9	10	Mike Brown	<a href="mailto:mike.brown@enron.com">mike.brown@enron.com</a>	Re: Final Review	80000	25000			
5	6	Tom Green	<a href="mailto:tom.green@enron.com">tom.green@enron.com</a>	Re: New Proposal	80000	25000			

```
7      8 Chris Black chris.black@enron.com Re: Follow Up 70000 20000
```

```
      state      city year total_compensation salary_percentage \
9 Massachusetts Boston 2001          105000          12.698413
5 Washington Seattle 2001          105000          12.698413
7 Oregon Portland 2001          90000          11.111111
```

```
      domain
9 enron.com
5 enron.com
7 enron.com
```

```
#12 Find employees with salary > $70,000 or bonus > $20,000.
```

```
filtered_employees = data[(data['salary'] > 70000) | (data['bonus'] > 20000)]
print(filtered_employees)
```

```
⇒ Sr No      Name      email      subject salary bonus \
5      6 Tom Green tom.green@enron.com Re: New Proposal 80000 25000
9     10 Mike Brown mike.brown@enron.com Re: Final Review 80000 25000
```

```
      state      city year total_compensation
5 Washington Seattle 2001          105000
9 Massachusetts Boston 2001          105000
```

```
#13. Calculate salary as a percentage of the total salary.
```

```
total_salary = data['salary'].sum()
data['salary_percentage'] = (data['salary'] / total_salary) * 100
print(data[['Name', 'salary', 'salary_percentage']])
```

```
⇒      Name salary salary_percentage
0 John Doe 50000 7.936508
1 Jane Smith 60000 9.523810
2 Mark Jones 50000 7.936508
3 Paul Brown 70000 11.111111
4 Sarah White 60000 9.523810
5 Tom Green 80000 12.698413
6 Emily Davis 50000 7.936508
7 Chris Black 70000 11.111111
8 Anna White 60000 9.523810
9 Mike Brown 80000 12.698413
```

```
#14. Find how many emails were sent each year.
```

```
emails_per_year = data.groupby('year').size()
print("Emails were sent each year:", emails_per_year)
```

```
➦ Emails were sent each year: year
2001    10
dtype: int64
```

```
#15.Which employee has the lowest salary?
lowest_salary = data[data['salary'] == data['salary'].min()]
print(lowest_salary[['Name', 'salary']])
```

```
➦
```

	Name	salary
0	John Doe	50000
2	Mark Jones	50000
6	Emily Davis	50000

```
#16.How many employees work in each city?
city_employee_counts = data['city'].value_counts()
print("Number of employees work in each city:",city_employee_counts)
```

```
➦ Number of employees work in each city: city
Los Angeles    1
New York       1
Dallas         1
Miami          1
Chicago        1
Seattle        1
Las Vegas      1
Portland       1
Denver         1
Boston         1
Name: count, dtype: int64
```

```
#17.Which employees are from California?
california_employees = data[data['state'] == 'California']
print(california_employees[['Name', 'city']])
```

```
➦
```

	Name	city
0	John Doe	Los Angeles

```
#18.Find the median bonus paid in each city.
median_bonus_by_city = data.groupby('city')['bonus'].median()
print("Median bonus paid in each city:",median_bonus_by_city)
```

```
➦ Median bonus paid in each city: city
Boston    25000.0
```

```
Chicago      15000.0
Dallas       10000.0
Denver       15000.0
Las Vegas    10000.0
Los Angeles  10000.0
Miami        20000.0
New York     15000.0
Portland     20000.0
Seattle      25000.0
Name: bonus, dtype: float64
```

```
#19.Which two cities have the highest total employee compensation?
data['total_compensation'] = data['salary'] + data['bonus']
city_total_comp = data.groupby('city')['total_compensation'].sum()
top_2_cities = city_total_comp.sort_values(ascending=False).head(2)
print("Two cities have the highest total employee compesation:",top_2_cities)
```

```
➦ Two cities have the highest total employee compesation: city
Boston      105000
Seattle     105000
Name: total_compensation, dtype: int64
```

```
#20.Find the most common email domain.
data['domain'] = data['email'].str.split('@').str[1]
most_common_domain = data['domain'].value_counts().idxmax()
print("Most common domain is:",most_common_domain)
```

```
➦ Most common domain is: enron.com
```

