

```
import re
import pandas as pd

# Sample SSNs
ssns = ['123-45-6789', '987-65-4321', '456-78-9012', '333-22-4444', '111-22-3333',
        '999-88-7777', '777-66-5555', '555-44-3333', '222-33-4444', '888-99-1111']

# Creating a pandas DataFrame with the SSN column
df = pd.DataFrame({'SSN': ssns})
df
```



	SSN
0	123-45-6789
1	987-65-4321
2	456-78-9012
3	333-22-4444
4	111-22-3333
5	999-88-7777
6	777-66-5555
7	555-44-3333
8	222-33-4444
9	888-99-1111

0 123-45-6789

1 987-65-4321

2 456-78-9012

3 333-22-4444

4 111-22-3333

5 999-88-7777

6 777-66-5555

7 555-44-3333

8 222-33-4444

9 888-99-1111

Next steps: [Generate code with df](#) [View recommended plots](#)

```
#A SSN is comprised of 3 digits-2 digits-4 digits
pattern = r'\d{3}-\d{2}-\d{4}'
```

```
# Applying re.sub() to the 'SSN' column by using re.sub to replace part of the information
df['Hidden SSN'] = df['SSN'].apply(lambda x: re.sub(r'\d{3}-\d{2}', '***-**', x))
df
```



	SSN	Hidden SSN
0	123-45-6789	***-**-6789
1	987-65-4321	***-**-4321
2	456-78-9012	***-**-9012
3	333-22-4444	***-**-4444
4	111-22-3333	***-**-3333
5	999-88-7777	***-**-7777
6	777-66-5555	***-**-5555
7	555-44-3333	***-**-3333
8	222-33-4444	***-**-4444
9	888-99-1111	***-**-1111

0 123-45-6789

1 987-65-4321

2 456-78-9012

3 333-22-4444

4 111-22-3333

5 999-88-7777

6 777-66-5555

7 555-44-3333


8 222-33-4444

9 888-99-1111

Next steps: [Generate code with df](#) [View recommended plots](#)

✓ If it's a string rather than a column of values

```
# Extracting the 'SSN' column as text
ssn_text = '\n'.join(df['SSN'].tolist())
print(ssn_text)
print(type(ssn_text))
```



```
123-45-6789
987-65-4321
456-78-9012
```

```
333-22-4444
111-22-3333
999-88-7777
777-66-5555
555-44-3333
222-33-4444
888-99-1111
<class 'str'>
```

```
# Regular expression pattern to match the first 5 digits of SSN
pattern = r'\d{3}-\d{2}(?!-)'

# Replacing all characters except the last 4 digits with asterisks
masked_ssn_text = re.sub(pattern, "****", ssn_text)

print(masked_ssn_text)
```

```
→ ****-6789
****-4321
****-9012
****-4444
****-3333
****-7777
****-5555
****-3333
****-4444
****-1111
```

Note: The regular expression pattern `r'\d{3}-\d{2}(?!-)'` matches the first 5 digits of the SSN (xxx-xx) followed by a hyphen. The `(?!-)` is a positive lookahead assertion that ensures the hyphen is not included in the match.