SMA EXPERIMENT NO - 3

Yash Ashok Shirsath BE AI & DS - 40 / B2

```
!pip install praw
import praw
import csv

client_id = "DkVBFHRvhbUbRpS0H0JYVQ"
client_secret = "POky-u3gAkHw_lOSTgp80S87xx6K7g"
user_agent = "u/Beginning-Shape-8543"

reddit = praw.Reddit(
    client_id=client_id,
    client_secret=client_secret,
    user_agent=user_agent
)

def extract_comments_to_csv(post_url, csv_filename="reddit_comments.csv"):
    """
    Extract comments from a Reddit post and save them to a CSV file.

    Args:
        post_url (str): URL of the Reddit post.
        csv_filename (str): Name of the CSV file to save.
    """
    try:
        submission = reddit.submission(url=post_url)
        submission.comments.replace_more(limit=None)  # Load all comments

        with open(csv_filename, 'w', newline='', encoding='utf-8') as csvfile:
            writer = csv.writer(csvfile)

            writer.writerow(["Title", "Post Author", "Comment Author", "Comment", "Score"])

            for comment in submission.comments.list():
                post_author = submission.author.name if submission.author else "Deleted/Unavailable"
                comment_author = comment.author.name if comment.author else "Deleted/Unavailable"
                writer.writerow([submission.title, post_author, comment_author, comment.body, comment.score])

        print(f"Comments saved to {csv_filename}")

    except Exception as e:
        print(f"An error occurred: {e}")

post_url = "https://www.reddit.com/r/thugeshh/comments/1i5fp6b/dont_u_think_its_too_much_of_a_concern_now/?utm_source=share&utm_medium=we
extract_comments_to_csv(post_url)

# Or specify a different filename:
# extract_comments_to_csv(post_url, "my_comments.csv")
```

```python
import pandas as pd
import re
from IPython.display import display
```

```python
df = pd.read_csv('reddit_comments.csv')
df.to_csv('raw_reddit_comments.csv', index=False)
```

```python
print(df.head)
```

```
<bound method NDFrame.head of                                                       Title        Post Author  \
0    Don't u think it's too much of a concern now ???  Deleted/Unavailable
1    Don't u think it's too much of a concern now ???  Deleted/Unavailable
2    Don't u think it's too much of a concern now ???  Deleted/Unavailable
3    Don't u think it's too much of a concern now ???  Deleted/Unavailable
4    Don't u think it's too much of a concern now ???  Deleted/Unavailable
..                                                ...                  ...
638  Don't u think it's too much of a concern now ???  Deleted/Unavailable
639  Don't u think it's too much of a concern now ???  Deleted/Unavailable
640  Don't u think it's too much of a concern now ???  Deleted/Unavailable
641  Don't u think it's too much of a concern now ???  Deleted/Unavailable
642  Don't u think it's too much of a concern now ???  Deleted/Unavailable

          Comment Author                                            Comment  \
0        OkEstablishment1881  post this is benguluru sub and see how they at...
1        ProfessionalWear2575          kannada mein resignation letter de do
2               RelevantRick  kanadigas are the stupidest being in india tbh...
3        Deleted/Unavailable  lol some people from bengaluru sub will come d...
4        AlarmingPsychology52  \n\nyes really too much to share the hatred wi...
..                       ...                                                ...
638             TacoSlayer66  say what you want but clearly this hate is com...
639             TacoSlayer66  okay and what else have you achieved in analys...
640               ArunMKumar  na i never beleived it matters to you\nthat is...
641                BlurrFrost        your comprehension skills need practice lol
642             TacoSlayer66  didnt realise this sub was handing out english...

     Score
0      325
1       26
2       65
3       20
4       40
..     ...
638      1
639      1
640      1
641      1
642      1

[616 rows x 5 columns]>
```

```python
df['Comment Author'].fillna("Unknown", inplace=True)
df['Comment'].fillna("No Comment", inplace=True)

def clean_text(text):
    if isinstance(text, str): # Check if it's a string
        text = re.sub(r'http\S+', '', text)
        text = re.sub(r'[^a-zA-Z0-9\s]', '', text)
```

```
            text = text.lower()  # Convert to lowercase
            return text
        else:
            return "" # or handle non-string values differently

df['Comment'] = df['Comment'].apply(clean_text)

df.drop_duplicates(subset=['Comment'], inplace=True)

df_filtered = df[df['Score'] > 5]  # Adjust filter as needed

df_filtered.to_csv('cleaned_reddit_comments.csv', index=False
```

a cleaning and storage complete!
ython-input-12-bef0b97ed9f2>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assign
 behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting val

 example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].m

f['Comment Author'].fillna("Unknown", inplace=True)
ython-input-12-bef0b97ed9f2>:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assign
 behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting val

 example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].m

f['Comment'].fillna("No Comment", inplace=True)

◄ | | ►

```
try:
    df_raw = pd.read_csv('raw_reddit_comments.csv')
except FileNotFoundError:
    print("Raw data file not found.")
    exit()

try:
    df_cleaned = pd.read_csv('cleaned_reddit_comments.csv')
except FileNotFoundError:
    print("Cleaned data file not found.")
    exit()

columns_to_display = ['Comment Author', 'Comment']

print("Raw Data:")
display(df_raw[columns_to_display].head(10))

print("\nCleaned Data:")
display(df_cleaned[columns_to_display].head(10))
```

Raw Data:
None   Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

Cleaned Data:

	Comment Author	Comment
0	OkEstablishment1881	post this is benguluru sub and see how they at...
1	ProfessionalWear2575	kannada mein resignation letter de do
2	RelevantRick	kanadigas are the stupidest being in india tbh...
3	Deleted/Unavailable	lol some people from bengaluru sub will come d...
4	AlarmingPsychology52	\n\nyes really too much to share the hatred wi...
5	YipeeKaiYayMoF	this is kannada imposition why cant we speak i...
6	Deleted/Unavailable	they will ban him
7	pikaboii	post in bangalore sub bengaludu sub is cringe
8	Powerful-Captain-362	i hope ai kills it
9	SolRon25	but no one will speak that talk in kannada in ...

No charts were generated by quickchart
Error: Runtime no longer has a reference to this dataframe, please re-run this cell and try