lookalike

```
import pandas as pd
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.preprocessing import OneHotEncoder
import numpy as np
from datetime import datetime


customers_df = pd.read_csv(r"C:\Users\varun\Downloads\Customers.csv")
products_df = pd.read_csv(r"C:\Users\varun\Downloads\Products.csv")


customers_df['SignupDate'] = pd.to_datetime(customers_df['SignupDate'])

encoder = OneHotEncoder(sparse_output=False)
regions_encoded = encoder.fit_transform(customers_df[['Region']])

customers_df['Region_Encoded'] = list(regions_encoded)


customer_features = np.random.rand(len(customers_df), 10)


similarities = cosine_similarity(customer_features)

def get_top_lookalikes(similarities, customer_ids, top_n=3):
    lookalikes = {}

    customer_ids = customer_ids.astype(str)

    for customer_id in customer_ids:
        if isinstance(similarities, dict):
            similarity_scores = similarities[customer_id]
        elif isinstance(similarities, np.ndarray):
            customer_index = int(customer_id[1:])
            similarity_scores = similarities[customer_index]

        similar_indices = np.argsort(similarity_scores)[::-1][1:top_n+1]

        valid_indices = [i for i in similar_indices if i < len(customer_ids)]

        similar_customers = [(customer_ids.iloc[j], similarity_scores[j]) for
j in valid_indices]

        lookalikes[customer_id] = similar_customers

    return lookalikes
```

```
top_20_customers = customers_df['CustomerID'][:20]

lookalikes = get_top_lookalikes(similarities, top_20_customers)

lookalike_data = []
for cust_id, similar_customers in lookalikes.items():
    for similar_cust_id, score in similar_customers:
        lookalike_data.append([cust_id, similar_cust_id, score])

lookalike_df = pd.DataFrame(lookalike_data, columns=['CustomerID',
'LookalikeCustomerID', 'SimilarityScore'])
lookalike_df.to_csv('Lookalike.csv', index=False)

lookalike_df.head()
```

	CustomerID	LookalikeCustomerID	SimilarityScore
0	C0001	C0017	0.964993
1	C0002	C0007	0.902774
2	C0007	C0020	0.909963
3	C0016	C0018	0.970198
4	C0016	C0002	0.964993

```
import os

downloads_path = os.path.join(os.path.expanduser("~"), "Downloads",
"Lookalike.csv")

lookalike_df.to_csv(downloads_path, index=False)

print(f"Lookalike CSV file has been saved to {downloads_path}")
```

```
Lookalike CSV file has been saved to C:\Users\varun\Downloads\Lookalike.csv
```