subramanian540_Project_Milestone4_Bkp

May 21, 2023

1 DSC 540 - Data Preparation

2 Week9 and Week 10

3 Project Milestone 4

4 Cleaning/Formatting API Source

5 Cleaning/Formatting API

Perform at least 5 data transformation and/or cleansing steps to your API data. The 5 data transformations that I will do are as follows:

Data transformation replaces the column headers in the "fraud_data" DataFrame with a new set of headers specified in the "new_headers" list. It then displays the first few rows of the DataFrame with the updated headers. This data transformation helps provide more descriptive and meaningful column names for the fraud detection data.

Data transformation converts them to float data type for 'distance_in_km' and 'distance_in_mile' columns with zero (0) . It rounds the values in the 'ip_latitude' and 'ip_longitude' columns to 6 decimal places. Lastly, it converts the values in the 'fraudlabspro_score' column to float data type. These transformations ensure that the data is in a more readable and appropriate format for further analysis or presentation

Data Transforamtion identifies and removes any duplicate rows in the "fraud_data" DataFrame. It creates a boolean mask of duplicated rows using the duplicated() function and then filters the DataFrame to keep only the rows that are not duplicates using the ~ operator. This data transformation ensures that the resulting DataFrame contains unique records, eliminating any duplicated entries.

Data Transformation modifies the DataFrame directly by removing any rows that have identical values in all columns, resulting in a DataFrame with unique records only.

Data transformation first drops rows from the "fraud_data" DataFrame where the columns 'ip_country', 'ip_region', and 'ip_city' have missing or empty values using the dropna() function with the subset parameter. This removes rows with missing values in these specific columns. Then, it further filters the DataFrame to drop rows where the values in these columns are '-'. The result is a DataFrame that excludes rows with missing or '-' values in the specified columns, improving the data quality and removing incomplete or irrelevant rows.

Data Transformation replaces missing values in the 'country_match' and 'ip_country' columns of the "fraud_data" DataFrame with the value 'Unknown'.This data transformation ensures that any missing values in the specified columns are replaced with the 'Unknown' value for better data completeness and consistency.

Data Transformation converts the values in the 'ip_country' column of the 'fraud_data' ensures consistent casing for the country names in the column. The resulting DataFrame will have the 'ip_country' values in uppercase format.

Data Transformation performs fuzzy matching to the 'ip_continent' column of the 'fraud_data' DataFrame. It replaces inconsistent values with their corresponding standardized values using the provided mapping dictionary. For example, 'North America' is replaced with 'NA', 'South America' with 'SA', and so on. The resulting DataFrame will have consistent and standardized values in the 'ip_continent' column.

Ethical Implications: Using the FraudLabs Pro API for order screening raises several ethical implications. One of the key concerns is privacy, as the dataset contains personal and sensitive information such as IP addresses, location details, email addresses, credit card information, and device identifiers. Safeguarding this data and ensuring compliance with privacy regulations are essential to protect individuals' privacy and prevent unauthorized access or data breaches. Transparency and informed consent are also important, as individuals should be aware of the purpose, collection, and usage of their data.Bias and discrimination are significant ethical considerations. It is crucial to ensure that the screening process and the algorithms used do not disproportionately impact certain individuals or groups based on factors like race, ethnicity, gender, or nationality. Compliance with legal and ethical standards is vital. Organizations using the dataset must adhere to data protection laws, privacy regulations, and consumer rights.Organizations should consider the ethical implications throughout their use of the dataset and take proactive measures to address privacy concerns, mitigate bias, ensure transparency, and uphold responsible data practices.

[3]:
```
## Import the required libraries

import pandas as pd
import json
import requests
import random
from concurrent import futures
```

[4]:
```
# Importing the FraudValidation class from the fraudlabspro.fraudvalidation␣
 ↪module
#API : https://www.fraudlabspro.com/developer/api/screen-order?ref=apilist.fun

from fraudlabspro.fraudvalidation import FraudValidation
```

The FraudValidation class allows you to access the fraud validation functionalities provided by the FraudLabs Pro API. By importing the FraudValidation class, you can access its methods and attributes to integrate the fraud validation capabilities into your code and leverage the features provided by the FraudLabs Pro API.

[5]:
```python
# Configure your API key
api_key = 'W7HNWZ1PGIEX6DPG1LIP8ATLEU2DB3BL'

# Set up the fraud validation object
fraud_validation = FraudValidation(api_key)
```

[6]:
```python
# Set up an empty list to store the results
results_list = []
```

Generates random order details such as order ID, amount, and IP address. The process_order function takes an order ID, increments it, and generates random values for amount and IP address. It then calls the fraud validation method of the API (https://www.fraudlabspro.com/developer/api/screen-order?ref=apilist.fun) using the fraud_validation.validate function and stores the result in the result variable.

[21]:
```python
# Set up the initial order ID
order_id = 5000

# Define a function to process a single order
def process_order(order_id):
    # Increment the order ID
    order_id += 1

    # Set up the order details for each iteration
    order_id_str = str(order_id).zfill(5)  # Ensure 5-digit order ID

    # Generate random values for amount
    amount = round(random.uniform(0, 100), 2)  # Generate a random float␣
 ↪between 0 and 100

    # Generate random values for IP Address
    ip_address = f"{random.randint(0, 255)}.{random.randint(0, 255)}.{random.
 ↪randint(0, 255)}.{random.randint(0, 255)}"

    order_details = {
        'order_id': order_id_str,
        'ip': ip_address,
        'amount': amount,
    }

    # Call the validation method and parse the result as a dictionary
    result = json.loads(fraud_validation.validate(order_details))

    return result

# Define the number of orders to process
num_orders = 50000
```

```
# Define the number of threads to use
num_threads = 1000

# Perform fraud validation for orders in parallel
with futures.ThreadPoolExecutor(max_workers=num_threads) as executor:
    future_to_order = {executor.submit(process_order, order_id): order_id for␣
 ↪order_id in range(num_orders)}

    for future in futures.as_completed(future_to_order):
        order_id = future_to_order[future]
        result = future.result()
        results_list.append(result)

# Create a DataFrame from the results list
df = pd.DataFrame(results_list)

# Save the DataFrame to a CSV file
df.to_csv('fraud_results1.csv', index=False)

fraud_data = pd.read_csv('fraud_results1.csv', sep=",")
fraud_data
```

[21]:

```
      is_country_match is_high_risk_country distance_in_km distance_in_mile  \
0                  NaN                    N              -                -
1                  NaN                    N              -                -
2                  NaN                    N              -                -
3                  NaN                    N              -                -
4                  NaN                    N              -                -
…                    …                    …              …                …
50967              NaN                    N              -                -
50968              NaN                    N              -                -
50969              NaN                    N              -                -
50970              NaN                    N              -                -
50971              NaN                    N              -                -

      ip_country   ip_continent                        ip_region  \
0             CA  North America                           Quebec
1             CH         Europe                           Zurich
2             BR  South America                   Santa Catarina
3             US  North America                       California
4             US  North America                          Florida
…            …              …                              …
50967         CL  South America  Region Metropolitana de Santiago
50968         US  North America                   North Carolina
50969         DE         Europe                Baden-Wurttemberg
50970         IT         Europe                            Lazio
```

```
50971         CL  South America  Region Metropolitana de Santiago

                 ip_city  ip_latitude  ip_longitude  … card_subtype  \
0               Montreal      45.5460      -73.6252  …          NaN
1                 Zurich      47.3668        8.5498  …          NaN
2      Balneario Camboriu     -26.9907      -48.6346  …          NaN
3              Cupertino      37.3166     -122.0465  …          NaN
4                  Miami      25.7743      -80.1936  …          NaN
…                    …          …             …   …          …
50967           Santiago     -33.4265      -70.5665  …          NaN
50968          Charlotte      35.2111      -80.8672  …          NaN
50969          Stuttgart      48.7825        9.1770  …          NaN
50970               Rome      41.8943       12.4843  …          NaN
50971           Santiago     -33.4265      -70.5665  …          NaN

       fraudlabspro_score fraudlabspro_distribution  fraudlabspro_status  \
0                     100                         0              APPROVE
1                     100                         0              APPROVE
2                     100                         0              APPROVE
3                     100                         0              APPROVE
4                     100                         0              APPROVE
…                     …                         …                 …
50967                 100                         0              APPROVE
50968                 100                         0              APPROVE
50969                 100                         0              APPROVE
50970                 100                         0              APPROVE
50971                 100                         0              APPROVE

       fraudlabspro_id  fraudlabspro_version fraudlabspro_error_code  \
0      20230521-SUAGLS                 1.5.1                     208
1      20230521-VZALEO                 1.5.1                     208
2      20230521-BSFT4Z                 1.5.1                     208
3      20230521-HYWZZZ                 1.5.1                     208
4      20230521-V8KLML                 1.5.1                     208
…                  …                   …                     …
50967  20230521-SYTEIJ                 1.5.1                     208
50968  20230521-YDS7ZB                 1.5.1                     208
50969  20230521-AWHT5R                 1.5.1                     208
50970  20230521-3JLK4N                 1.5.1                     208
50971  20230521-WGXLGN                 1.5.1                     208

         fraudlabspro_message fraudlabspro_credits  device_id
0      INVALID QUANTITY VALUE                  500        NaN
1      INVALID QUANTITY VALUE                  500        NaN
2      INVALID QUANTITY VALUE                  500        NaN
3      INVALID QUANTITY VALUE                  500        NaN
4      INVALID QUANTITY VALUE                  500        NaN
```

```
…                    …                             …         …
50967  INVALID QUANTITY VALUE                        500        NaN
50968  INVALID QUANTITY VALUE                        500        NaN
50969  INVALID QUANTITY VALUE                        500        NaN
50970  INVALID QUANTITY VALUE                        500        NaN
50971  INVALID QUANTITY VALUE                        500        NaN

[50972 rows x 56 columns]
```

[22]:

```python
#Data Transformatio : 1 Replace Headers

# Replace Headers
new_headers = ['country_match', 'high_risk_country', 'distance_in_km',␣
 ↪'distance_in_mile',
               'ip_country', 'ip_continent', 'ip_region', 'ip_city',␣
 ↪'ip_latitude', 'ip_longitude',
               'ip_timezone', 'ip_elevation', 'ip_domain', 'ip_mobile_mnc',␣
 ↪'ip_mobile_mcc',
               'ip_mobile_brand', 'ip_netspeed', 'ip_isp_name',␣
 ↪'ip_usage_type', 'free_email',
               'new_domain_name', 'domain_exists', 'proxy_ip_address',␣
 ↪'bin_found',
               'bin_country_match', 'bin_name_match', 'bin_phone_match',
               'bin_phone_country_match', 'bin_prepaid', 'address_ship_forward',
               'bill_ship_city_match', 'bill_ship_state_match',␣
 ↪'bill_ship_country_match',
               'is_bill_ship_postal_match', 'is_ship_address_blacklist',␣
 ↪'is_phone_blacklist',
               'ip_blacklist', 'email_blacklist', 'credit_card_blacklist',
               'device_blacklist', 'user_blacklist', 'high_risk_username',
               'export_controlled_country', 'malware_exploit', 'user_order_id',␣
 ↪'user_order_memo',
               'card_subtype', 'fraudlabspro_score',␣
 ↪'fraudlabspro_distribution', 'fraudlabspro_status',
               'fraudlabspro_id', 'fraudlabspro_version',␣
 ↪'fraudlabspro_error_code', 'fraudlabspro_message',
               'fraudlabspro_credits', 'device_id']

fraud_data.columns = new_headers
fraud_data.head()
```

[22]:

```
   country_match high_risk_country distance_in_km distance_in_mile ip_country  \
0            NaN                 N              -                -         CA
1            NaN                 N              -                -         CH
2            NaN                 N              -                -         BR
3            NaN                 N              -                -         US
```

```
4             NaN                 N              -                 -         US

    ip_continent       ip_region             ip_city  ip_latitude  \
0  North America          Quebec            Montreal      45.5460
1         Europe          Zurich              Zurich      47.3668
2  South America  Santa Catarina  Balneario Camboriu     -26.9907
3  North America      California           Cupertino      37.3166
4  North America         Florida               Miami      25.7743

   ip_longitude  … card_subtype  fraudlabspro_score  \
0      -73.6252  …          NaN                 100
1        8.5498  …          NaN                 100
2      -48.6346  …          NaN                 100
3     -122.0465  …          NaN                 100
4      -80.1936  …          NaN                 100

  fraudlabspro_distribution  fraudlabspro_status  fraudlabspro_id  \
0                         0              APPROVE  20230521-SUAGLS
1                         0              APPROVE  20230521-VZALEO
2                         0              APPROVE  20230521-BSFT4Z
3                         0              APPROVE  20230521-HYWZZZ
4                         0              APPROVE  20230521-V8KLML

   fraudlabspro_version fraudlabspro_error_code    fraudlabspro_message  \
0                 1.5.1                     208  INVALID QUANTITY VALUE
1                 1.5.1                     208  INVALID QUANTITY VALUE
2                 1.5.1                     208  INVALID QUANTITY VALUE
3                 1.5.1                     208  INVALID QUANTITY VALUE
4                 1.5.1                     208  INVALID QUANTITY VALUE

  fraudlabspro_credits  device_id
0                  500        NaN
1                  500        NaN
2                  500        NaN
3                  500        NaN
4                  500        NaN

[5 rows x 56 columns]
```

[23]:

```
#Data Transformatio : 2 # Format data into a more readable format

fraud_data['distance_in_km'] = fraud_data['distance_in_km'].replace('-', 0).
 ↪astype(float)
fraud_data['distance_in_mile'] = fraud_data['distance_in_mile'].replace('-', 0).
 ↪astype(float)
fraud_data['ip_latitude'] = fraud_data['ip_latitude'].round(6)
fraud_data['ip_longitude'] = fraud_data['ip_longitude'].round(6)
```

```
fraud_data['fraudlabspro_score'] = fraud_data['fraudlabspro_score'].
 ↪astype(float)

fraud_data.head()
```

[23]:

```
  country_match high_risk_country  distance_in_km  distance_in_mile  \
0           NaN                 N             0.0               0.0
1           NaN                 N             0.0               0.0
2           NaN                 N             0.0               0.0
3           NaN                 N             0.0               0.0
4           NaN                 N             0.0               0.0

  ip_country   ip_continent       ip_region            ip_city  ip_latitude  \
0         CA  North America          Quebec           Montreal      45.5460
1         CH         Europe          Zurich             Zurich      47.3668
2         BR  South America  Santa Catarina  Balneario Camboriu    -26.9907
3         US  North America      California          Cupertino      37.3166
4         US  North America         Florida              Miami      25.7743

   ip_longitude  … card_subtype  fraudlabspro_score  \
0      -73.6252  …          NaN               100.0
1        8.5498  …          NaN               100.0
2      -48.6346  …          NaN               100.0
3     -122.0465  …          NaN               100.0
4      -80.1936  …          NaN               100.0

  fraudlabspro_distribution  fraudlabspro_status  fraudlabspro_id  \
0                         0              APPROVE  20230521-SUAGLS
1                         0              APPROVE  20230521-VZALEO
2                         0              APPROVE  20230521-BSFT4Z
3                         0              APPROVE  20230521-HYWZZZ
4                         0              APPROVE  20230521-V8KLML

   fraudlabspro_version fraudlabspro_error_code    fraudlabspro_message  \
0                 1.5.1                     208  INVALID QUANTITY VALUE
1                 1.5.1                     208  INVALID QUANTITY VALUE
2                 1.5.1                     208  INVALID QUANTITY VALUE
3                 1.5.1                     208  INVALID QUANTITY VALUE
4                 1.5.1                     208  INVALID QUANTITY VALUE

  fraudlabspro_credits  device_id
0                  500        NaN
1                  500        NaN
2                  500        NaN
3                  500        NaN
4                  500        NaN
```

```
[5 rows x 56 columns]
```

[24]:
```python
#Data Transformatio : 3 # Find duplicates
duplicates = fraud_data.duplicated()
fraud_data = fraud_data[~duplicates]

fraud_data.head()

#Data Transformatio : 4 # Remove duplicate rows
fraud_data.drop_duplicates(inplace=True)
fraud_data.head()
```

[24]:
```
  country_match high_risk_country  distance_in_km  distance_in_mile  \
0           NaN                 N             0.0               0.0
1           NaN                 N             0.0               0.0
2           NaN                 N             0.0               0.0
3           NaN                 N             0.0               0.0
4           NaN                 N             0.0               0.0

  ip_country   ip_continent       ip_region            ip_city  ip_latitude  \
0         CA  North America          Quebec           Montreal      45.5460
1         CH         Europe          Zurich             Zurich      47.3668
2         BR  South America  Santa Catarina  Balneario Camboriu     -26.9907
3         US  North America      California          Cupertino      37.3166
4         US  North America         Florida              Miami      25.7743

   ip_longitude  … card_subtype  fraudlabspro_score  \
0      -73.6252  …          NaN               100.0
1        8.5498  …          NaN               100.0
2      -48.6346  …          NaN               100.0
3     -122.0465  …          NaN               100.0
4      -80.1936  …          NaN               100.0

  fraudlabspro_distribution  fraudlabspro_status  fraudlabspro_id  \
0                         0              APPROVE  20230521-SUAGLS
1                         0              APPROVE  20230521-VZALEO
2                         0              APPROVE  20230521-BSFT4Z
3                         0              APPROVE  20230521-HYWZZZ
4                         0              APPROVE  20230521-V8KLML

   fraudlabspro_version fraudlabspro_error_code    fraudlabspro_message  \
0                 1.5.1                     208  INVALID QUANTITY VALUE
1                 1.5.1                     208  INVALID QUANTITY VALUE
2                 1.5.1                     208  INVALID QUANTITY VALUE
3                 1.5.1                     208  INVALID QUANTITY VALUE
4                 1.5.1                     208  INVALID QUANTITY VALUE
```

```
  fraudlabspro_credits  device_id
0                   500        NaN
1                   500        NaN
2                   500        NaN
3                   500        NaN
4                   500        NaN

[5 rows x 56 columns]
```

[25]:
```python
#Data Transformatio : 5 # Replace missing values in specific columns

fraud_data['country_match'].fillna('Unknown', inplace=True)
fraud_data['ip_country'].fillna('Unknown', inplace=True)
fraud_data.head()
```

[25]:
```
  country_match high_risk_country  distance_in_km  distance_in_mile  \
0       Unknown                 N             0.0               0.0
1       Unknown                 N             0.0               0.0
2       Unknown                 N             0.0               0.0
3       Unknown                 N             0.0               0.0
4       Unknown                 N             0.0               0.0

  ip_country   ip_continent       ip_region            ip_city  ip_latitude  \
0         CA  North America          Quebec           Montreal      45.5460
1         CH         Europe          Zurich             Zurich      47.3668
2         BR  South America  Santa Catarina  Balneario Camboriu     -26.9907
3         US  North America      California          Cupertino      37.3166
4         US  North America         Florida              Miami      25.7743

   ip_longitude  … card_subtype  fraudlabspro_score  \
0      -73.6252  …          NaN               100.0
1        8.5498  …          NaN               100.0
2      -48.6346  …          NaN               100.0
3     -122.0465  …          NaN               100.0
4      -80.1936  …          NaN               100.0

  fraudlabspro_distribution  fraudlabspro_status  fraudlabspro_id  \
0                         0              APPROVE  20230521-SUAGLS
1                         0              APPROVE  20230521-VZALEO
2                         0              APPROVE  20230521-BSFT4Z
3                         0              APPROVE  20230521-HYWZZZ
4                         0              APPROVE  20230521-V8KLML

   fraudlabspro_version fraudlabspro_error_code    fraudlabspro_message  \
0                 1.5.1                     208  INVALID QUANTITY VALUE
1                 1.5.1                     208  INVALID QUANTITY VALUE
2                 1.5.1                     208  INVALID QUANTITY VALUE
```

```
3                 1.5.1                     208  INVALID QUANTITY VALUE
4                 1.5.1                     208  INVALID QUANTITY VALUE

  fraudlabspro_credits  device_id
0                  500        NaN
1                  500        NaN
2                  500        NaN
3                  500        NaN
4                  500        NaN

[5 rows x 56 columns]
```

[26]:

```
#Data Transformatio : 6 # Drop rows with missing or empty values in ip_country,␣
↪ip_region, and ip_city columns
fraud_data.dropna(subset=['ip_country', 'ip_region', 'ip_city'], inplace=True)
fraud_data.head()

#Data Transformatio : 6 # Drop rows with '-' values in ip_country, ip_region,␣
↪and ip_city columns
fraud_data = fraud_data[(fraud_data['ip_country'] != '-') &␣
↪(fraud_data['ip_region'] != '-') & (fraud_data['ip_city'] != '-')]
fraud_data.head()
```

[26]:

```
  country_match high_risk_country  distance_in_km  distance_in_mile  \
0       Unknown                 N             0.0               0.0
1       Unknown                 N             0.0               0.0
2       Unknown                 N             0.0               0.0
3       Unknown                 N             0.0               0.0
4       Unknown                 N             0.0               0.0

  ip_country   ip_continent       ip_region             ip_city  ip_latitude  \
0         CA  North America          Quebec            Montreal      45.5460
1         CH         Europe          Zurich              Zurich      47.3668
2         BR  South America  Santa Catarina  Balneario Camboriu     -26.9907
3         US  North America      California           Cupertino      37.3166
4         US  North America         Florida               Miami      25.7743

   ip_longitude  … card_subtype  fraudlabspro_score  \
0      -73.6252  …          NaN               100.0
1        8.5498  …          NaN               100.0
2      -48.6346  …          NaN               100.0
3     -122.0465  …          NaN               100.0
4      -80.1936  …          NaN               100.0

  fraudlabspro_distribution  fraudlabspro_status  fraudlabspro_id  \
0                         0              APPROVE  20230521-SUAGLS
1                         0              APPROVE  20230521-VZALEO
```

```
2                         0             APPROVE  20230521-BSFT4Z
3                         0             APPROVE  20230521-HYWZZZ
4                         0             APPROVE  20230521-V8KLML

   fraudlabspro_version fraudlabspro_error_code    fraudlabspro_message  \
0                 1.5.1                     208  INVALID QUANTITY VALUE
1                 1.5.1                     208  INVALID QUANTITY VALUE
2                 1.5.1                     208  INVALID QUANTITY VALUE
3                 1.5.1                     208  INVALID QUANTITY VALUE
4                 1.5.1                     208  INVALID QUANTITY VALUE

  fraudlabspro_credits  device_id
0                  500        NaN
1                  500        NaN
2                  500        NaN
3                  500        NaN
4                  500        NaN

[5 rows x 56 columns]
```

[27]:

```
#Data Transformatio : 7 # Fix casing or inconsistent values
fraud_data['ip_country'] = fraud_data['ip_country'].str.upper()
fraud_data.head()
```

[27]:

```
  country_match high_risk_country  distance_in_km  distance_in_mile  \
0       Unknown                 N             0.0               0.0
1       Unknown                 N             0.0               0.0
2       Unknown                 N             0.0               0.0
3       Unknown                 N             0.0               0.0
4       Unknown                 N             0.0               0.0

  ip_country   ip_continent       ip_region             ip_city  ip_latitude  \
0         CA  North America          Quebec            Montreal      45.5460
1         CH         Europe          Zurich              Zurich      47.3668
2         BR  South America  Santa Catarina  Balneario Camboriu     -26.9907
3         US  North America      California           Cupertino      37.3166
4         US  North America         Florida               Miami      25.7743

   ip_longitude  … card_subtype  fraudlabspro_score  \
0      -73.6252  …          NaN               100.0
1        8.5498  …          NaN               100.0
2      -48.6346  …          NaN               100.0
3     -122.0465  …          NaN               100.0
4      -80.1936  …          NaN               100.0

  fraudlabspro_distribution  fraudlabspro_status  fraudlabspro_id  \
0                         0             APPROVE  20230521-SUAGLS
```

```
1                      0             APPROVE  20230521-VZALEO
2                      0             APPROVE  20230521-BSFT4Z
3                      0             APPROVE  20230521-HYWZZZ
4                      0             APPROVE  20230521-V8KLML

   fraudlabspro_version fraudlabspro_error_code    fraudlabspro_message  \
0                 1.5.1                     208  INVALID QUANTITY VALUE
1                 1.5.1                     208  INVALID QUANTITY VALUE
2                 1.5.1                     208  INVALID QUANTITY VALUE
3                 1.5.1                     208  INVALID QUANTITY VALUE
4                 1.5.1                     208  INVALID QUANTITY VALUE

  fraudlabspro_credits  device_id
0                  500        NaN
1                  500        NaN
2                  500        NaN
3                  500        NaN
4                  500        NaN

[5 rows x 56 columns]
```

[28]:

```python
#Data Transformatio : 8 # Conduct Fuzzy Matching (Example: fixing inconsistent␣
 ↪values in 'ip_continent')
fuzzy_mapping = {'North America': 'NA', 'South America': 'SA', 'Europe': 'EU',␣
 ↪'Asia': 'AS', 'Africa': 'AF', 'Oceania': 'OC'}
fraud_data['ip_continent'] = fraud_data['ip_continent'].replace(fuzzy_mapping)
fraud_data.head()
```

[28]:

```
  country_match high_risk_country  distance_in_km  distance_in_mile  \
0       Unknown                 N             0.0               0.0
1       Unknown                 N             0.0               0.0
2       Unknown                 N             0.0               0.0
3       Unknown                 N             0.0               0.0
4       Unknown                 N             0.0               0.0

  ip_country ip_continent       ip_region            ip_city  ip_latitude  \
0         CA           NA          Quebec           Montreal      45.5460
1         CH           EU          Zurich             Zurich      47.3668
2         BR           SA  Santa Catarina  Balneario Camboriu    -26.9907
3         US           NA      California          Cupertino      37.3166
4         US           NA         Florida              Miami      25.7743

   ip_longitude  … card_subtype  fraudlabspro_score  \
0      -73.6252  …          NaN               100.0
1        8.5498  …          NaN               100.0
2      -48.6346  …          NaN               100.0
3     -122.0465  …          NaN               100.0
```

```
4      -80.1936  …           NaN               100.0

  fraudlabspro_distribution  fraudlabspro_status  fraudlabspro_id  \
0                         0              APPROVE  20230521-SUAGLS
1                         0              APPROVE  20230521-VZALEO
2                         0              APPROVE  20230521-BSFT4Z
3                         0              APPROVE  20230521-HYWZZZ
4                         0              APPROVE  20230521-V8KLML

   fraudlabspro_version fraudlabspro_error_code    fraudlabspro_message  \
0                 1.5.1                     208  INVALID QUANTITY VALUE
1                 1.5.1                     208  INVALID QUANTITY VALUE
2                 1.5.1                     208  INVALID QUANTITY VALUE
3                 1.5.1                     208  INVALID QUANTITY VALUE
4                 1.5.1                     208  INVALID QUANTITY VALUE

  fraudlabspro_credits  device_id
0                  500        NaN
1                  500        NaN
2                  500        NaN
3                  500        NaN
4                  500        NaN

[5 rows x 56 columns]
```

[29]:

```
#Data Transformatio : 9 # Reset the index after dropping rows

fraud_data.reset_index(drop=True, inplace=True)
fraud_data.rename(columns={'index': 'Index_ID'}, inplace=True)
fraud_data
```

[29]:

```
      country_match high_risk_country  distance_in_km  distance_in_mile  \
0           Unknown                 N             0.0               0.0
1           Unknown                 N             0.0               0.0
2           Unknown                 N             0.0               0.0
3           Unknown                 N             0.0               0.0
4           Unknown                 N             0.0               0.0
…               …                 …              …                …
43823       Unknown                 N             0.0               0.0
43824       Unknown                 N             0.0               0.0
43825       Unknown                 N             0.0               0.0
43826       Unknown                 N             0.0               0.0
43827       Unknown                 N             0.0               0.0

      ip_country ip_continent                        ip_region  \
0             CA           NA                           Quebec
1             CH           EU                           Zurich
```

```
2             BR           SA                    Santa Catarina
3             US           NA                        California
4             US           NA                           Florida
…          …          …                                 …
43823         CL           SA  Region Metropolitana de Santiago
43824         US           NA                    North Carolina
43825         DE           EU                 Baden-Wurttemberg
43826         IT           EU                             Lazio
43827         CL           SA  Region Metropolitana de Santiago

                   ip_city  ip_latitude  ip_longitude  … card_subtype  \
0                 Montreal      45.5460      -73.6252  …          NaN
1                   Zurich      47.3668        8.5498  …          NaN
2       Balneario Camboriu     -26.9907      -48.6346  …          NaN
3                Cupertino      37.3166     -122.0465  …          NaN
4                    Miami      25.7743      -80.1936  …          NaN
…                      …            …             …  …          …
43823             Santiago     -33.4265      -70.5665  …          NaN
43824            Charlotte      35.2111      -80.8672  …          NaN
43825            Stuttgart      48.7825        9.1770  …          NaN
43826                 Rome      41.8943       12.4843  …          NaN
43827             Santiago     -33.4265      -70.5665  …          NaN

       fraudlabspro_score fraudlabspro_distribution  fraudlabspro_status  \
0                   100.0                         0              APPROVE
1                   100.0                         0              APPROVE
2                   100.0                         0              APPROVE
3                   100.0                         0              APPROVE
4                   100.0                         0              APPROVE
…                     …                       …                  …
43823               100.0                         0              APPROVE
43824               100.0                         0              APPROVE
43825               100.0                         0              APPROVE
43826               100.0                         0              APPROVE
43827               100.0                         0              APPROVE

       fraudlabspro_id  fraudlabspro_version fraudlabspro_error_code  \
0      20230521-SUAGLS                 1.5.1                     208
1      20230521-VZALEO                 1.5.1                     208
2      20230521-BSFT4Z                 1.5.1                     208
3      20230521-HYWZZZ                 1.5.1                     208
4      20230521-V8KLML                 1.5.1                     208
…                  …                   …                      …
43823  20230521-SYTEIJ                 1.5.1                     208
43824  20230521-YDS7ZB                 1.5.1                     208
43825  20230521-AWHT5R                 1.5.1                     208
43826  20230521-3JLK4N                 1.5.1                     208
```

```
43827  20230521-WGXLGN                 1.5.1                      208

         fraudlabspro_message fraudlabspro_credits  device_id
0      INVALID QUANTITY VALUE                  500        NaN
1      INVALID QUANTITY VALUE                  500        NaN
2      INVALID QUANTITY VALUE                  500        NaN
3      INVALID QUANTITY VALUE                  500        NaN
4      INVALID QUANTITY VALUE                  500        NaN
…                         …                  …        …
43823  INVALID QUANTITY VALUE                  500        NaN
43824  INVALID QUANTITY VALUE                  500        NaN
43825  INVALID QUANTITY VALUE                  500        NaN
43826  INVALID QUANTITY VALUE                  500        NaN
43827  INVALID QUANTITY VALUE                  500        NaN

[43828 rows x 56 columns]
```