In [115…
```
import pandas as pd
```

In [116…
```
df = pd.read_csv('do_not_push_transactions.csv')
```

In [117…
```
df.drop('Description', axis=1, inplace=True)
```

In [118…
```
df['Original Description'] = df['Original Description'].str.lower()
df['Original Description'] = df['Original Description'].str.replace(' ', '')
df['Original Description'] = df['Original Description'].str.replace('-', '')
df['Original Description'] = df['Original Description'].str.replace('(', '')
```

In [119…
```
# remove transactions names and change them to Transaction 1, Transaction 2, etc.
for i in range(len(df)):
    df.loc[i, 'Original Description'] = 'Transaction ' + str(i+1)
```

This iteration area where the names of my accounts have been anonymized is redacted :)

```
df.reset_index(drop=True, inplace=True)

print(df['Account Name'].value_counts())
```

```
BCC1      67
BCC2      62
B         45
PP        33
C         29
```

```
CC2       14
CCC1      13
A1        11
W          8
A          7
CRYPTO     6
VEN        1
Name: Account Name, dtype: int64
```

In [120…

```
# reallocate categories from Mint.com https://mint.intuit.com/mint-categories/
Income =['Paycheck', 'Investment', 'Returned Purchase', 'Bonus', 'Interest Income', 'Reimb
Miscellaneous = ['Cash & ATM', 'Check', 'Miscellaneous']
Entertainment = ['Arts''Music','Movies & DVDs','Newspaper & Magazines', 'Entertainment']
Education= ['Tuition', 'Student Loan','Books & Supplies', 'Education']
Shopping = ['Clothing', 'Books', 'Electronics & Software', 'Hobbies', 'Sporting Goods', 'S
Personal_Care = ['Laundry', 'Hair', 'Spa & Massage', 'Personal Care']
Health_Fitness = ['Doctor', 'Dentist', 'Eye Care', 'Pharmacy', 'Health Insurance', 'Gym',
Kids= ['Activities', 'Allowance', 'Baby Supplies', 'Babysitter & Daycare', 'Child Support'
Food_Dining = ['Groceries', 'Coffee Shops', 'Fast Food', 'Restaurants', 'Alcohol', 'Food &
Gifts_Donations =['Gift', 'Charity', 'Donation']
Investments= ['Deposit', 'Withdrawal', 'Dividends & Cap Gains', 'Buy', 'Sell', 'Investment
Bills_Utilities = ['Television', 'Home Phone', 'Internet', 'Mobile Phone', 'Utilities', 'E
Auto_Transport= ['Auto & Transport', 'Gas & Fuel', 'Parking', 'Service & Auto Parts', 'Aut
Travel = ['Air Travel', 'Hotel', 'Rental Car & Taxi', 'Vacation','Ride Share', 'Travel','E
Fees_Charges = ['Bank Fees', 'Service Fee', 'Late Fee', 'Finance Charge', 'ATM Fee', 'Bank
Business_Services= ['Advertising', 'Office Supplies', 'Printing', 'Shipping', 'Legal', 'Bu
Taxes = ['Federal Tax', 'State Tax', 'Local Tax', 'Sales Tax', 'Property Tax', 'Taxes']
Transfer= ['Credit Card Payment']
```

In [121…

```
Main_Categories = [Income, Miscellaneous, Entertainment, Education, Shopping, Personal_Car
# iterate through the dataframe and assign the main categories, I unfortunately have to do
for i in range(len(df)):
    if df.loc[i, 'Category'] in Income:
        df.loc[i, 'Category'] = 'Income'
    if df.loc[i, 'Category'] in Miscellaneous:
        df.loc[i, 'Category'] = 'Miscellaneous'
    if df.loc[i, 'Category'] in Entertainment:
        df.loc[i, 'Category'] = 'Entertainment'
    if df.loc[i, 'Category'] in Education:
        df.loc[i, 'Category'] = 'Education'
    if df.loc[i, 'Category'] in Shopping:
        df.loc[i, 'Category'] = 'Shopping'
    if df.loc[i, 'Category'] in Personal_Care:
        df.loc[i, 'Category'] = 'Personal Care'
    if df.loc[i, 'Category'] in Health_Fitness:
        df.loc[i, 'Category'] = 'Health & Fitness'
    if df.loc[i, 'Category'] in Kids:
        df.loc[i, 'Category'] = 'Kids'
    if df.loc[i, 'Category'] in Food_Dining:
        df.loc[i, 'Category'] = 'Food & Dining'
    if df.loc[i, 'Category'] in Gifts_Donations:
        df.loc[i, 'Category'] = 'Gifts & Donations'
    if df.loc[i, 'Category'] in Investments:
        df.loc[i, 'Category'] = 'Investments'
    if df.loc[i, 'Category'] in Bills_Utilities:
        df.loc[i, 'Category'] = 'Bills & Utilities'
    if df.loc[i, 'Category'] in Auto_Transport:
        df.loc[i, 'Category'] = 'Auto & Transport'
    if df.loc[i, 'Category'] in Travel:
        df.loc[i, 'Category'] = 'Travel'
    if df.loc[i, 'Category'] in Fees_Charges:
        df.loc[i, 'Category'] = 'Fees & Charges'
    if df.loc[i, 'Category'] in Business_Services:
        df.loc[i, 'Category'] = 'Business Services'
```

```
    if df.loc[i, 'Category'] in Taxes:
        df.loc[i, 'Category'] = 'Taxes'
    if df.loc[i, 'Category'] in Transfer:
        df.loc[i, 'Category'] = 'Transfer'
```

In [122…

```
# drop labels and notes
df.drop(['Labels', 'Notes'], axis=1, inplace=True)
```

In [123…

```
# export csv and final cleaning
# drop transactions with income category
df = df[df['Category'] != 'Income']
df = df[df['Category'] != 'Transfer']
#replace ' & ' with '_'
df['Category'] = df['Category'].str.replace(' & ', '_')

# change column names to Date,Original Description,Amount,Transaction_Type,Category,Accou
df.columns = ['Date', 'Description', 'Amount', 'Transaction_Type', 'Category', 'Account_Na

df.to_csv('transactions_cleaned.csv', index=False)

df['Category'].value_counts()
```

Out[123…

```
Food_Dining          99
Shopping             45
Travel               12
Health_Fitness        6
Investments           6
Auto_Transport        3
Bills_Utilities       2
Education             2
Entertainment         1
Business Services     1
Name: Category, dtype: int64
```