Problem Statement:-

You are provided with a file "main.csv" in the input folder. Following operation needs to be provided on the file:-

1. Create an output file "filteredCountry.csv" inside a folder named "output". This file should contain only those records where country contains the word USA.

Code:

filteredCountry.csv

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plot


data = pd.read_csv(“filteredCountry.csv”)

data

df = pd.DataFrame(data)

df.describe()


import pandas as pd

data = pd.read_csv(“filteredCountry.csv”)

df1 = data.groupby([‘SKU’, ‘Description’, ‘Year’)] [ ‘capacity’]. sum()

df2 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [ ‘URL’].sum()

df3 = data.groupby([‘SKU’, ‘Description’, ‘Year)]’, [‘ price’].sum()

df4 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [‘Seller Information’].sum()

df5 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [‘Offerr Information’].sum()

df6 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [‘Country’].sum()
```

```
print (df1)

print (df2)

print (df3)

print (df4)

print (df5)

print (df6)


import pandas as pd

data = pd.read_csv("filteredCountry.csv")

df3 = data.groupby(['SKU', 'Description', 'Year)]', [' price'].sum()

df4 = data.groupby(['SKU', 'Description', 'Year')], ['Seller Information'].sum()

print(df3)

print(df4)


import pandas as pd

data = pd.read_csv("filteredCountry.csv")

df5 = data.groupby(['SKU', 'Description', 'Year')], ['Offerr Information'].sum()

df6 = data.groupby(['SKU', 'Description', 'Year')], ['Country'].sum()

print(df5)

print(df6)


import math

def status2(x):

    if x>10000 :

return ("filteredCountry")

data['filteredCountry_status'] = data['status'].apply(status2)

data
```

dataframe=pd.Dataframe(data)

array = dataframe.values

array

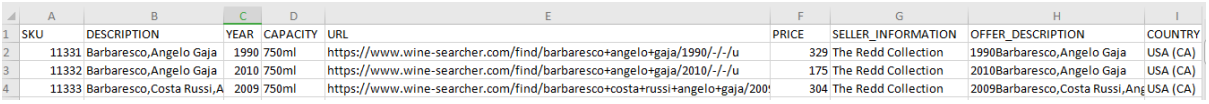
x = array[:,9:3]

print(x)

y = array[“,3]

print(y)

output:



	A	B	C	D	E	F	G	H	I
1	SKU	DESCRIPTION	YEAR	CAPACITY	URL	PRICE	SELLER_INFORMATION	OFFER_DESCRIPTION	COUNTRY
2	11331	Barbaresco,Angelo Gaja	1990	750ml	https://www.wine-searcher.com/find/barbaresco+angelo+gaja/1990/-/-/u	329	The Redd Collection	1990Barbaresco,Angelo Gaja	USA (CA)
3	11332	Barbaresco,Angelo Gaja	2010	750ml	https://www.wine-searcher.com/find/barbaresco+angelo+gaja/2010/-/-/u	175	The Redd Collection	2010Barbaresco,Angelo Gaja	USA (CA)
4	11333	Barbaresco,Costa Russi,A	2009	750ml	https://www.wine-searcher.com/find/barbaresco+costa+russi+angelo+gaja/200	304	The Redd Collection	2009Barbaresco,Costa Russi,An	USA (CA)

2. Now consider "filteredCountry.csv" as the input file. For each group of "SKU" find 2 minimum prices and store this result in "lowestPrice.csv" inside a folder named "output"

Code:

Input file---“filteredCountry.csv”

Create a new file—“lowestPrice.csv”

import pandas as pd

data = pd.read_csv(“filteredCountry.csv”)

df1 = data.groupby([‘SKU’, ‘Description’, ‘Year’)] [‘capacity’]. sum()

df2 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [‘URL’].sum()

df3 = data.groupby([‘SKU’, ‘Description’, ‘Year)]’, [‘ price’].sum()

df4 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [‘Seller Information’].sum()

df5 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [‘Offerr Information’].sum()

df6 = data.groupby([‘SKU’, ‘Description’, ‘Year’)], [‘Country’].sum()

print (df1)

print (df2)

print (df3)

print (df4)

print (df5)

print (df6)

dataframe=pd.Dataframe(data)

array = dataframe.values

array

minimum prices = array[:,9:3]

print(minimum prices)

output:

“lowest Prices.csv”

	A	B	C
1	SKU	FIRST_MINIMUM_PRICE	SECOND_MINIMUM_PRICE
2	11333	304	359.99
3	11334	329	339
4	11335	304	315