Alison Schnoes

Data Import Specialist Assessment

Keap

# Steps to accomplish the questions word file document:

I opened the email (Keap - Data Import Specialist Assessment) from Sue Lane and downloaded the zip file (Data\_Import\_Specialist\_Assessment) to my downloads.

Export zip file to Desktop’s Keap folder.

Open Test Requirement Word Document

Open Anaconda Navigator

Open Jupyter Notebook

Click on the Desktop and then the Keap folder

New -> Python 3 Notebook

# Question 1: Split the Name column in file 1 into First Name and Last Name

Loading Libraries

import pandas as pd

Load in the csv file 1 from the relative path in the Keap folder using pd.read\_csv

Show the first 5 columns in file 1 to ensure that it read properly using file1.head(5)

Then used the long way to code the split of the name column into first name and last name columns.

* New data frame with split value columns
  + file1new = file1["Name"].str.split(" ", n = 1, expand = True)
* Separate first name column from new data frame
  + file1["First Name"] = file1new[0]
* Separate last name column from new data frame
  + file1["Last Name"] = file1new[1]
* Dropping the old Name column
  + file1.drop(columns =["Name"], inplace = True)

Then show the first 5 columns to demonstrate the split of the first and last name in file 1.

# Question 2: Split the State column in file 1 into City and State

Then I googled on how to split with less code and found a string split online to use and adapt to the dataframe of file 1.

file1[['City', 'State']] = file1['State'].str.split(',', n=1, expand=True)

Cite: <https://www.geeksforgeeks.org/python-pandas-split-strings-into-two-list-columns-using-str-split/>

Surprise! It works! - as shown by showing the first 5 columns again from file 1

# Question 3: Set the Country column in file 1 as 'United States' for all rows

First, I had to find all of the value counts in the Country column using file1.value\_counts('Country'). I found out there were 28 counts of USA and the 13 counts of US.

Then I did a series loop function, to recode USA into United States, and US into United States with export at the end to value\_counts to confirm that they all were changed.

Also, verified again by showing the first 5 columns again from file 1

# Question 4: Set the Date Created column in file 2 to show as DD/MM/YYYY

Then I searched the web for the library to load (from datetime import datetime) and code that made sense.

Cite: <https://stackoverflow.com/questions/502726/converting-date-between-dd-mm-yyyy-and-yyyy-mm-dd>

Changed some words and dates around to adapt to the needs of this assessment. Using a series function and naming date\_formats, we were able to get this changed. It automatically listed data from the dataframe and it was confirmed by showing the first 5 rows of file 2

def conv\_dates\_series(df, col, old\_date\_format, new\_date\_format):

file2['Date Created'] = pd.to\_datetime(file2['Date Created'], format=old\_date\_format).dt.strftime(new\_date\_format)

return file2

test\_df = pd.DataFrame({"file2": ["01-01-0000", "01-01-9999"]})

old\_date\_format='%m/%d/%Y'

new\_date\_format='%d/%m/%Y'

conv\_dates\_series(test\_df, "file2", old\_date\_format, new\_date\_format)

# Question 6: Add 'Date Created' from file 2 and 'Address' column from file 3 into file 1 and make sure they are assigned to the correct record based on the Id

I had previously merged a file with the Id so merging file 2 into file 1 was easy only two columns in the dataframe.

merged\_data = file1.merge(file2,on=["Id"],)

merged\_data.head()

Then I did lots of reading and found numerous sources to test and try to figure out the syntax to add the Address column from file 3 into the previously merged dataframe with including the Id column to match. I found out that it was a left join using merging syntax.

file1 = pd.merge(merged\_data,file3[['Id','Address']],on='Id', how='left')

file1.head()

Download Jupyter Notebook file as Python and HTML files for viewing

Download Google Doc into Word File

Zip documents all together to submit