classify

## Prompts

Part 1 - <https://chat.openai.com/c/e39d4cde-d4ef-434a-b32d-5b67ce52b72b>

Part 2 - <https://chat.openai.com/c/c6c3d27c-779a-45ba-9694-8e7b8605c057>

PArt 3 - <https://chat.openai.com/c/f6d2c233-6a5f-4783-a457-72803280cd40>

## Code

import os  
import pandas as pd  
import re  
import openpyxl  
import extract\_msg  
from extract\_msg.exceptions import InvalidFileFormatError  
  
*# Set the directory you want to search in*dir\_path = r'E:\onedrive\Documentos\Roberto'  
  
*# Load the existing Excel file as a DataFrame*excel\_path = r'E:\onedrive\Documentos\Roberto\projects\automation\email\_management\email\_management.xlsx'  
try:  
 df\_existing = pd.read\_excel(excel\_path)  
except FileNotFoundError:  
 df\_existing = pd.DataFrame(columns=["Email Subject", "Path", "Sender", "Recipients", "Archive", "Date Added"])  
  
*# Get the column widths from the existing Excel file*if os.path.exists(excel\_path):  
 wb\_existing = openpyxl.load\_workbook(excel\_path)  
 ws\_existing = wb\_existing.active  
 column\_widths = [ws\_existing.column\_dimensions[openpyxl.utils.get\_column\_letter(i+1)].width for i in range(len(df\_existing.columns))]  
  
*# Create a list to store the file names and paths*file\_list = []  
  
*# Initialize a counter for processed emails*processed\_emails = 0  
  
*# Iterate over all the files in the directory*for subdir, dirs, files in os.walk(dir\_path):  
 for file in files:  
  
 *# Check if the file has a .msg extension* if file.endswith(".msg"):  
  
 *# Get the full path to the file* file\_path = os.path.join(subdir, file)  
  
 *# Check if the file is already in the existing DataFrame* if not df\_existing[(df\_existing["Email Subject"] == file) & (df\_existing["Path"] == subdir)].empty:  
 continue  
  
 *# Remove leading number sequence and dash from file name* file\_name = re.sub(r"^\d+\s\*-\s\*", "", file)  
  
 *# Remove "re" or "rv" prefix and any leading whitespace from email subject* subject = re.sub(r"^\s\*(re|rv|fw[d]\*)\s\*:?\s\*", "", file\_name, flags=re.IGNORECASE)  
  
 *# Extract sender and recipient information from the .msg file, with error control* try:  
 with extract\_msg.Message(file\_path) as msg:  
 sender = msg.sender  
 recipients = msg.to  
 except InvalidFileFormatError:  
 print(f"InvalidFileFormatError: Skipping file {file\_path}")  
 continue  
  
 *# Add the modified subject, path, sender, recipients, and other information to the list* file\_list.append((subject.strip(), subdir, sender, recipients, None, pd.Timestamp.now()))  
  
 *# Increment the counter* processed\_emails += 1  
  
 *# Pause and ask for an "Enter" keypress after 1,000 emails processed* if processed\_emails % 1000 == 0:  
 input(f"Processed {processed\_emails} emails. Press 'Enter' to continue...")  
  
*# Create a DataFrame from the file list*df\_new = pd.DataFrame(file\_list, columns=["Email Subject", "Path", "Sender New", "Recipients New", "Archive New", "Date Added New"])  
  
*# Rename the "Archive", "Date Added", "Sender" and "Recipients" columns in the existing DataFrame to "Archive Existing", "Date Added Existing" "Sender Existing" and "Recipients Existing", respectively*df\_existing.rename(columns={"Archive": "Archive Existing", "Date Added": "Date Added Existing", "Sender": "Sender Existing", "Recipients": "Recipients Existing"}, inplace=True)  
  
*# Merge the existing and new DataFrames on the "Email Subject" and "Path" columns, keeping all rows*df\_all = pd.merge(df\_existing, df\_new, on=["Email Subject", "Path"], how="outer")  
  
*# Combine the "Sender Existing" and "Sender New" columns into a single "Sender" column*df\_all["Sender"] = df\_all["Sender Existing"].fillna(df\_all["Sender New"])  
df\_all.drop(["Sender Existing", "Sender New"], axis=1, inplace=True)  
  
*# Combine the "Recipients Existing" and "Recipients New" columns into a single "Recipients" column*df\_all["Recipients"] = df\_all["Recipients Existing"].fillna(df\_all["Recipients New"])  
df\_all.drop(["Recipients Existing", "Recipients New"], axis=1, inplace=True)  
  
*# Combine the "Archive Existing" and "Archive New" columns into a single "Archive" column*df\_all["Archive"] = df\_all["Archive Existing"].fillna(df\_all["Archive New"])  
df\_all.drop(["Archive Existing", "Archive New"], axis=1, inplace=True)  
  
*# Combine the "Date Added Existing" and "Date Added New" columns into a single "Date Added" column*df\_all["Date Added"] = df\_all["Date Added Existing"].fillna(df\_all["Date Added New"])  
df\_all.drop(["Date Added Existing", "Date Added New"], axis=1, inplace=True)  
  
*# Remove duplicate email subjects and paths*df\_all.drop\_duplicates(subset=["Email Subject", "Path"], keep="first", inplace=True)  
  
*# Export the updated database to an Excel file*excel\_path = r'E:\onedrive\Documentos\Roberto\projects\automation\email\_management\email\_management\_updated.xlsx'  
df\_all.to\_excel(excel\_path, index=False)  
  
*# Apply the column widths to the final Excel file*if column\_widths:  
 wb\_final = openpyxl.load\_workbook(excel\_path)  
 ws\_final = wb\_final.active  
 for i, width in enumerate(column\_widths):  
 ws\_final.column\_dimensions[openpyxl.utils.get\_column\_letter(i+1)].width = width  
 wb\_final.save(excel\_path)

## Debug

"C:\Program Files\Python311\python.exe" E:\onedrive\Documentos\Roberto\projects\automation\email\_management.py

Processed 1000 emails. Press 'Enter' to continue...

Processed 2000 emails. Press 'Enter' to continue...

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\barcelona - mallorca 81\alquiler temporal\atbarcelona\clientes\2016-07\013 - Resumen alquiler ATBarcelona Resum lloguer ATBarcelona Rent summary ATBarcelona.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\barcelona - mallorca 81\alquiler temporal\fincas escarré\mariano benlliure 30\abogados\catala reinon\015 - documentación Roberto Ferraro - Catala-Reinon.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\barcelona - mallorca 81\alquiler temporal\fincas escarré\mariano benlliure 30\abogados\catala reinon\019 - maria benlliure 30 - informe pericial.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\barcelona - mallorca 81\alquiler temporal\fincas escarré\mariano benlliure 30\abogados\catala reinon\039 - RE juicio maria benlliure - fotos reparación.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\barcelona - mallorca 81\alquiler temporal\fincas escarré\mariano benlliure 30\abogados\catala reinon\selección documentos para juicio\040 - notas abogado.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\barcelona - mallorca 81\alquiler temporal\fincas escarré\mariano benlliure 30\abogados\catala reinon\selección documentos para juicio\anexos\001 - Fwd facturas.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\barcelona - mallorca 81\alquiler temporal\fincas escarré\mariano benlliure 30\abogados\catala reinon\selección documentos para juicio\anexos\019 - maria benlliure 30 - informe pericial.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\mariano benlliure 30\abogados\catala reinon\015 - documentación Roberto Ferraro - Catala-Reinon.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\mariano benlliure 30\abogados\catala reinon\019 - maria benlliure 30 - informe pericial.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\mariano benlliure 30\abogados\catala reinon\039 - RE juicio maria benlliure - fotos reparación.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\mariano benlliure 30\abogados\catala reinon\selección documentos para juicio\040 - notas abogado.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\mariano benlliure 30\abogados\catala reinon\selección documentos para juicio\anexos\001 - Fwd facturas.msg

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\house\mariano benlliure 30\abogados\catala reinon\selección documentos para juicio\anexos\019 - maria benlliure 30 - informe pericial.msg

Processed 3000 emails. Press 'Enter' to continue...

Processed 4000 emails. Press 'Enter' to continue...

Processed 5000 emails. Press 'Enter' to continue...

Processed 6000 emails. Press 'Enter' to continue...

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\investments\revisión\export\operaciones (no se utiliza)\20160924\001 - BOUGHT 73 EOA IBIS @ 6 797 (UXXX5785).msg

Processed 7000 emails. Press 'Enter' to continue...

InvalidFileFormatError: Skipping file E:\onedrive\Documentos\Roberto\areas\purchases\farmacia llambi\022 - Re minoxidil en crema.msg

Processed 8000 emails. Press 'Enter' to continue...

Processed 9000 emails. Press 'Enter' to continue...

Process finished with exit code 0

Archive

## Prompts

Part 4 - <https://chat.openai.com/c/44fb50fc-905f-4be2-bcae-a370fc5c6d75>

## Code

## debug