

```

1  import docx
2  import re
3  import pandas as pd
4  import shutil
5  import os
6  import tempfile
7
8  def extract_info_from_docx(docx_path):
9      # Create a temporary directory to store the copy of the docx file
10     with tempfile.TemporaryDirectory() as tmpdirname:
11         # Construct the path for the temporary copy
12         temp_docx_path = os.path.join(tmpdirname, os.path.basename(docx_path))
13
14         # Copy the original docx to the temporary directory
15         shutil.copy(docx_path, temp_docx_path)
16
17         # Open the Word document from the temporary copy
18         doc = docx.Document(temp_docx_path)
19         full_text = "\n".join([para.text for para in doc.paragraphs])
20
21         # Define regex patterns for the data to extract
22         name_pattern = re.search(r'(Full\s*Name|Name):?\s*(.*)', full_text, re.IGNORECASE)
23         id_pattern = re.search(r'(ID|Identification):?\s*(\d+)', full_text, re.IGNORECASE)
24         dob_pattern = re.search(r'(DOB|Date\s*of\s*Birth):?\s*([\d/]+)', full_text, re.IGNORECASE)
25         account_pattern = re.search(r'(Account\s*Number|Acc\s*No):?\s*(\d+)', full_text, re.IGNORECASE)
26         phone_pattern = re.search(
27             r'(Phone\s*Number|Contact):?\s*(\d{3})?[-.\s]?(\d{3})[-.\s]?(\d{4})', full_text, re.IGNORECASE)
28
29         # Extract information
30         full_name = name_pattern.group(2).strip() if name_pattern else ""
31         id_num = id_pattern.group(2).strip() if id_pattern else ""
32         dob = dob_pattern.group(2).strip() if dob_pattern else ""
33         account_num = account_pattern.group(2).strip() if account_pattern else ""
34         phone_num = phone_pattern.group(0).split()[-1] if phone_pattern else "" # To get the format "456-7890"
35
36         # Split the full name into first and last names
37         first_name, last_name = full_name.split()[:2] if len(full_name.split()) >= 2 else (full_name, "")
38
39         return first_name, last_name, id_num, dob, account_num, phone_num
40
41 def main():
42     text_file_path = r"C:\Users\Stephen\Documents\1. A Work\Test area\Filename holder.txt"
43
44     with open(text_file_path, 'r') as file:
45         docx_filename = file.read().strip()
46
47     docx_path = r"C:\Users\Stephen\OneDrive - Nomad\Apps\Caution_2024-09-08_.docx"
48     first_name, last_name, id_num, dob, account_num, phone_num = extract_info_from_docx(docx_path)
49
50     # Prepare data for Excel
51     data = {
52         "First Name": [first_name],
53         "Last Name": [last_name],
54         "ID 1": [id_num],
55         "DOB": [dob],
56         "AC No.": [account_num],
57         "Phone": [phone_num]

```

```
57     }
58
59     # Create a DataFrame
60     df = pd.DataFrame(data)
61
62     # Save the DataFrame to an Excel file
63     output_excel_path = r"C:\Users\Stephen\Documents\Extracted_Info.xlsx"
64     df.to_excel(output_excel_path, index=False)
65
66     print(f"Extracted information has been saved to {output_excel_path}")
67
68 if __name__ == '__main__':
69     main()
70
```