```
import os
import io
import pandas as pd
from google.cloud import vision
from pdf2image import convert from path
from PIL import Image
import PyPDF2
# Set up Google Cloud Vision client
os.environ["GOOGLE APPLICATION CREDENTIALS"] = "path/to/your/credentials.json"
client = vision.ImageAnnotatorClient()
def extract text from image(image path):
  with io.open(image path, 'rb') as image file:
    content = image file.read()
  image = vision.Image(content=content)
  response = client.text detection(image=image)
  texts = response.text annotations
  if texts:
    return texts[0].description
  return ""
def extract text from pdf(pdf path):
  pages = convert from path(pdf path, 300)
  text = ""
  for page number, page in enumerate(pages):
    temp image path = f"page {page number}.jpg"
    page.save(temp image path, 'JPEG')
    text += extract text from image(temp image path)
    os.remove(temp image path)
  return text
def process single document(file path):
  if file path.lower().endswith('.pdf'):
    return extract text from pdf(file path)
  elif file path.lower().endswith(('.jpg', '.jpeg', '.png')):
    return extract text from image(file path)
  return ""
def process multiple documents(directory path):
  data = []
  for file name in os.listdir(directory path):
    file path = os.path.join(directory path, file name)
    if file path.lower().endswith(('.pdf', '.jpg', '.jpeg', '.png')):
       text = process single document(file path)
       data.append({'file name': file name, 'extracted text': text})
  return pd.DataFrame(data)
# Example usage
if name == " main ":
  directory path = "path/to/your/documents"
  extracted data df = process multiple documents(directory path)
  extracted data df.to csv("extracted data.csv", index=False)
```