

Step 1 -- Covert the transcript samples from pdf to jpg format

Install two required modules: pdf2image and poppler

Install pdf2image

In [1]: `pip install pdf2image`

Requirement already satisfied: pdf2image in c:\users\edwar\anaconda3\lib\site-packages (1.15.1)
Requirement already satisfied: pillow in c:\users\edwar\anaconda3\lib\site-packages (from pdf2image) (8.0.1)
Note: you may need to restart the kernel to use updated packages.

Download poppler <https://anaconda.org/conda-forge/poppler/files>

Installing poppler

<https://stackoverflow.com/questions/53481088/poppler-in-path-for-pdf2image>

Convert two transcript samples from pdf format to image jpg format

Approach:

- Import the **pdf2image** module
- Store a PFD with **convert_from_path()**
- Save image with **save()**

```
In [2]: # import module
        from pdf2image import convert_from_path

        #poppler_path=r'C:\Program Files\poppler-0.68.0\bin'

        # Store Pdf with convert_from_path function
        file_path = 'sample_1.pdf'
        images = convert_from_path(file_path, 500 , poppler_path=r'C:\Program Files\poppler-0.6

        for i in range(len(images)):

        # Save pages as images in the pdf
            images[i].save('.\images\sample_'+ str(i+1) +'.jpg', 'JPEG')
```

```
In [3]: # Store Pdf with convert_from_path function
        file_path = 'sample_2.pdf'
        images = convert_from_path(file_path, 500 , poppler_path=r'C:\Program Files\poppler-0.6

        for i in range(len(images)):

        # Save pages as images in the pdf
            images[i].save('.\images\sample_'+ str(i+2) +'.jpg', 'JPEG')
```

Step 2. Using tesseraact, Open CV and Pytesseract to

convert image to text strings

Install tesseract & pytesseract

In [4]: `pip install tesseract`

Requirement already satisfied: tesseract in c:\users\edwar\anaconda3\lib\site-packages (0.1.3)

Note: you may need to restart the kernel to use updated packages.

In [5]: `pip install pytesseract`

Requirement already satisfied: pytesseract in c:\users\edwar\anaconda3\lib\site-packages (0.3.7)

Requirement already satisfied: Pillow in c:\users\edwar\anaconda3\lib\site-packages (from m pytesseract) (8.0.1)

Note: you may need to restart the kernel to use updated packages.

Transcript Sample #1

Course Flags (*:UC/CSU Approved, +:Weighted GPA) Column Labels (Mrk:Academic)

Remarks	I certify that the above information is correct.	Date this certificate issued
	Principal or Registrar	06/15/2015

```
Requirement already satisfied: opencv-python in c:\users\edwar\anaconda3\lib\site-packag
es (4.5.2.54)
Requirement already satisfied: numpy>=1.17.3 in c:\users\edwar\anaconda3\lib\site-packag
es (from opencv-python) (1.19.2)
Note: you may need to restart the kernel to use updated packages.
```

localhost:8888/nbconvert/html/OneDrive/Desktop/Project 2/Project-2-HIGH-SCHOOL-TRANSCRIPT-READER-main/Project2/Project 2 HIGH SCHO... 3/27

```

import pytesseract
from pytesseract import Output

# tell where the Tesseract engine is installed (Window).
pytesseract.pytesseract.tesseract_cmd = r"C:\Program Files\Tesseract-OCR\tesseract.exe"

image_path = './images/sample_1.jpg'

# Using openCV to read image file; can also use Pillow module, PIL.Image
img = cv2.imread(image_path)

text = pytesseract.image_to_string(img)
#print(text, file=open("output1.txt", "a"))
print(text)

```

| OPEL LWWA
 [_____] Bate of Birth _ Official Transcript

EEE EL XXX XXX)
 State Student 1p KOC OL ES POCONOS
 Place of Birth

Term Ending
 [ExitDate_] _ExitReason | Grad Dato Diploma Type

Issued To GPA Type GPA Crdt Atmpt
 Overall Weighted 190.00
 09/15/2016

1 of |

Total Credits Earned 190.00
 Nirk Credits |] C Mrk Credits || Graduation Requirements Short

Course IO Course ourse ID Course
 CCC CES Grd 09 HS Sem | 01/14 | eee Grd 11 HS Sem 1 01/1é | Comprehensive Grad Requireme
 nt
 412223 PE Course 1 A 5.00) *+737450 AP Biology (HP) : 5.00] English 10.00 40.00 30.00

737500 Biology (P) B- 5.00 |+748380 AP Psychology (HP} 5.00] Global Studies 5.00 5.00
 729G21 Health Education B 5.00 | *748180 US Hist (P) 5.00 | World History 10.06 10.00
 *653635 Freshman English (P) C+ 5.00 | *653795 Junior English (P) 5.00] US History 10.00
 10.00
 *865100 Spanish | (P} B 5.00 | *655140 Spanish tll {P} 5,00] Civies 5.00 5.00 0.00
 *697110 Hon Geometry (P} Cc 5.00 | *+697230 AP Calculus AB {HP} 5.00 | Economics \$.00 5.
 00 0.00
 TERM: GPA 2.83 Credits 30.00 | TERM: GPA 2.83 Credits 30.00] Math 20.06 20.00
 CUMULATIVE; GPA 2.83 Credits 30.00 | CUMULATIVE: GPA 3.09 Credits 160.00 | Physical Scie
 nce 10.06 10.00
 Life Science 10.06 10.00
 LECCE CES Grd 09 HS Sem 2 0614 | FCCC Grd (1 HS Sen 2 06/16 | VPA/World Language 10.00 1
 0.00
 712223 PE Course 1 5.00 |*+737450 AP Biology (HP) 5.00 | Physical Education 10.00 20.00
 10.00
 *737500 Biology (P} 5.00] *+748380 AP Psychology (HP) 5.00] Health Education 5.00 5.00
 *748330 Global Studies {(P)} 5.00 | *748180 US Hist (P)} 5.00] Electives 10.00 89.00 79.
 00
 "653635 Freshman English (P) 5.00 | *653795 Junior English (PF) 0.00] Total Grad Credits
 40.00 230.06 190.00
 *865100 Spanish | (P} 5.00 | "655140 Spanish II] (P} 5.00) All Courses 190.06 190.00

*697110 Hon Geometry (P} 5.00 | *+697230 AP Calculus AB {HP} 5.00] Speech Prag Passed
 TERM: GPA 3.006 Credits 30.00 | TERM: GPA 3.60 Credits 25.06
 CUMULATIVE; GPA 2.92 Credits 60.00} CUMULATIVE: GPA 3.16 Credits 185.00
 EEC EEE EEE EEE LE
 Out of District/Transfer Credits Grd {0 SummerSe 08/14 | 73ers Grd (1 HS Summer 07/16 Ne
 rerexe)

*TRUC012 = Sum Alg2 (P) / Transfer Bf 10.00] "653795SS Junior English (P) B+ 5.00
 TERM: GPA 3.00 Credits 10.00 | TERM: GPA 3.00 Credits 5.00
 CUMULATIVE: GPA 2.93 Credits 70.00} CUMULATIVE: GPA 3.16 Credits 190.00

XCCECECCES Grd 10 HS Sem 1 OL/LS | Work In Progress

"737660 Chemistry (P) A. 500 | RR ISIererererer Grd (2
 748450 - Wrid Hist 2-3 (P) A- 5.00] 748220 Civics (P) 085/15 12/226

*653655 Sophomore English (P} A 5.00] €24250 Photography 1 (P} 0815/16 06/02/17
 *665120 Spanish II (P} A- 5.00 | 635210 Lit & Soc Just (P) 081516 06/02/17
 *+697370 AP Computer Sci (HP} Cc 5.00] 697145 AP Statistics (HP) 0815/16 o6/02/17
 *697120 Hon Pre-Calculus (HP} B- 5.00} 737700 Physics (P} 0813165 6/02/17
 TERM; GPA 3.67 Credits 30.00 | 748320 AP Human Geo (HP) 0815/16 a6/02/17
 CUMULATIVE; GPA 3.16 Credits 100.00 | 748260 Economics (P} 12/2316 06/02/17

K00 Grd (0 HS Sem 2 a6/15
 *737660 Chemistry (P} 5.00

*748460 Wrid Hist 2-3 (P} 5.00
 *853655 Sophomore English (P} 5.00
 *665120 Spanish || (P} 5.00
 *+897370 AP Computer Sci (HP} 5.00
 *697120 Hon Pre-Calculus (HP} C- 5.00
 TERM: GPA 3.17 Credits 30.00
 CUMULATIVE; GPA 3.15 Credits 130.00

Course Flags (*:UC/CSU Approved, +:Weighted GPA) Column Labels (Mrk: Academic)

I certify that the above information is correct. Date this certificate issued

06/15/2015

ipal ar Registrar

```
In [14]: d1 = pytesseract.image_to_data(img, output_type=Output.DICT)
d1.keys()
#print(text, file=open("output1.txt", "a"))
d1["text"]

while("" in d1["text"]) :
    d1["text"].remove("")

while(" " in d1["text"]) :
    d1["text"].remove(" ")

import json
jsonStr1 = json.dumps(d1['text'], indent=4)

print(jsonStr1)
```

```
[
  "|",
  "OPEL",
  "LWWA",
  "[_____|",
  "Bate",
  "of",
  "Birth",
  "-",
  "Official",
  "Transcript",
  "EEE",
  "EL",
  "XXX",
  "XXX)",
  "State",
  "Student",
  "1p",
  "KOC",
  "OL",
  "ES",
  "POCONOS",
  "Place",
  "of",
  "Birth",
  "Term",
  "Ending",
  "[ExitDate_|",
  "-ExitReason",
  "|",
  "Grad",
  "Dato",
  "Diploma",
  "Type",
  "Issued",
  "To",
  "GPA",
  "Type",
  "GPA",
  "Crdt",
  "Atmpt",
  "Overall",
  "Weighted",
  "190.00",
  "09/15/2016",
  "1",
  "of",
  "|",
  "Total",
  "Credits",
  "Earned",
  "190.00",
  "Nirk",
  "Credits",
  "|]",
  "C",
  "Mrk",
  "Credits",
  "||",
  "Graduation",
  "Requirements",
  "Short",
  "Course",
  "IO",
  "Course",

```

"ourse",
"ID",
"Course",
"CCC",
"CES",
"Grd",
"09",
"HS",
"Sem",
"|",
"01/14",
"|",
"eee",
"Grd",
"11",
"HS",
"Sem",
"1",
"01/1\u00e9",
"|",
"Comprehensive",
"Grad",
"Requirement",
"412223",
"PE",
"Course",
"1",
"A",
"5.00)",
"*+737450",
"AP",
"Biology",
"(HP)",
":",
"5.00]",
"English",
"10.00",
"40.00",
"30.00",
" ",
"*737500",
"Biology",
"(P)",
"B-",
"5.00",
"*+748380",
"AP",
"Psychology",
"(HP}",
"5.00]",
"Global",
"Studies",
"5.00",
"5.00",
"729G21",
"Health",
"Education",
"B",
"5.00",
"|",
"*748180",
"US",
"Hist",
"(P)",
"5.00",

"|",
"World",
"History",
"10.06",
"10.00",
"*653635",
"Freshman",
"English",
"(P)",
"C+",
"5.00",
"|",
"*653795",
"Junior",
"English",
"(P)",
"5.00]",
"US",
"History",
"10.00",
"10.00",
"*865100",
"Spanish",
"|",
"(P}",
"B",
"5.00",
"|",
"*655140",
"Spanish",
"t11",
"{P}",
"5,00]",
"Civies",
"5.00",
"5.00",
"0.00",
"*697110",
"Hon",
"Geometry",
"(P}",
"Cc",
"5.00",
"|",
"*+697230",
"AP",
"Calculus",
"AB",
"{HP}",
"5.00",
"|",
"Economics",
"\u00a7.00",
"5.00",
"0.00",
"TERM:",
"GPA",
"2.83",
"Credits",
"30.00",
"|",
"TERM:",
"GPA",
"2.83",
"Credits",


```
"30.00]",  
"Math",  
"20.06",  
"20.00",  
"CUMULATIVE;",  
"GPA",  
"2.83",  
"Credits",  
"30.00",  
"|",  
"CUMULATIVE:",  
"GPA",  
"3.09",  
"Credits",  
"160.00",  
"|",  
"Physical",  
"Science",  
"10.06",  
"10.00",  
"Life",  
"Science",  
"10.06",  
"10.00",  
"LECCE",  
"CES",  
"Grd",  
"09",  
"HS",  
"Sem",  
"2",  
"0614",  
"|",  
"FCCC",  
"Grd",  
"(1",  
"HS",  
"Sen",  
"2",  
"06/16",  
"|",  
"VPA/World",  
"Language",  
"10.00",  
"10.00",  
"712223",  
"PE",  
"Course",  
"1",  
"5.00",  
"|*+737450",  
"AP",  
"Biology",  
"(HP)",  
"5.00",  
"|",  
"Physical",  
"Education",  
"10.00",  
"20.00",  
"10.00",  
"*737500",  
"Biology",  
"(P}",  
"5.00]",
```

"*+748380",
"AP",
"Psychology",
"(HP)",
"5.00]",
"Health",
"Education",
"5.00",
"5.00",
"*748330",
"Global",
"Studies",
"\u00a2(P))",
"5.00",
"|",
"*748180",
"US",
"Hist",
"(P))",
"5.00]",
"Electives",
"10.00",
"89.00",
"79.00",
"\653635",
"Freshman",
"English",
"(P)",
"5.00",
"|",
"*653795",
"Junior",
"English",
"(PF)",
"0.00]",
"Total",
"Grad",
"Credits",
"40.00",
"230.06",
"190.00",
"*865100",
"Spanish",
"|",
"(P)",
"5.00",
"|",
"\655140",
"Spanish",
"I1]",
"(P)",
"5.00)",
"All",
"Courses",
"190.06",
"190.00",
"*697110",
"Hon",
"Geometry",
"(P)",
"5.00",
"|",
"*+697230",
"AP",
"Calculus",

"AB",
"{HP}",
"5.00]",
"Speech",
"Prag",
"Passed",
"TERM:",
"GPA",
"3.006",
"Credits",
"30.00",
"|",
"TERM:",
"GPA",
"3.60",
"Credits",
"25.06",
"CUMULATIVE;",
"GPA",
"2.92",
"Credits",
"60.00}",
"CUMULATIVE:",
"GPA",
"3.16",
"Credits",
"185.00",
"EEC",
"EEE",
"EEE",
"EEE",
"LE",
"Out",
"of",
"District/Transfer",
"Credits",
"Grd",
"{0",
"SummerSe",
"08/14",
"|",
"73ers",
"Grd",
"(1",
"HS",
"Summer",
"07/16",
"Nererexe)",
"*TRUC012",
"=",
"Sum",
"Alg2",
"(P)",
"/",
"Transfer",
"Bf",
"10.00]",
"\\"653795SS",
"Junior",
"English",
"(P)",
"B+",
"5.00",
"TERM:",
"GPA",

"3.00",
 "Credits",
 "10.00",
 "|",
 "TERM:",
 "GPA",
 "3.00",
 "Credits",
 "5.00",
 "CUMULATIVE:",
 "GPA",
 "2.93",
 "Credits",
 "70.00)",
 "CUMULATIVE:",
 "GPA",
 "3.16",
 "Credits",
 "190.00",
 "XCCECCECCES",
 "Grd",
 "10",
 "HS",
 "Sem",
 "1",
 "OL/LS",
 "|",
 "Work",
 "In",
 "Progress",
 "\u201c737660",
 "Chemistry",
 "(P)",
 "A.",
 "500",
 "|",
 "RR",
 "ISIererererer",
 "Grd",
 "(2",
 "748450",
 "\u2014",
 "Wrid",
 "Hist",
 "2-3",
 "(P)",
 "A-",
 "5.00]",
 "748220",
 "Civics",
 "(P)",
 "085/15",
 "12/226",
 "*653655",
 "Sophomore",
 "English",
 "(P}",
 "A",
 "5.00]",
 "\u2013",
 "Photography",
 "1",
 "(P}",
 "0815/16",
 "06/02/17",

"*665120",
"Spanish",
"I1",
"(P}",
"A-",
"5.00",
"|",
"635210",
"Lit",
"&",
"Soc",
"Just",
"(P)",
"081516",
"06/02/17",
"*+697370",
"AP",
"Computer",
"Sci",
"(HP}",
"Cc",
"5.00]",
"697145",
"AP",
"Statistics",
"(HP)",
"0815/16",
"06/02/17",
"*697120",
"Hon",
"Pre-Calculus",
"(HP}",
"B-",
"5.00}",
"737700",
"Physics",
"(P}",
"0813165",
"6/02/17",
"TERM;",
"GPA",
"3.67",
"Credits",
"30.00",
"|",
"748320",
"AP",
"Human",
"Geo",
"(HP)",
"0815/16",
"a6/02/17",
"CUMULATIVE;",
"GPA",
"3.16",
"Credits",
"100.00",
"|",
"748260",
"Economics",
"(P}",
"12/2316",
"06/02/17",
"K00",
"Grd",

"(0",
"HS",
"Sem",
"2",
"a6/15",
"*737660",
"Chemistry",
"(P}",
"5.00",
"*748460",
"Wrid",
"Hist",
"2-3",
"(P}",
"5.00",
"*853655",
"Sophomore",
"English",
"(P}",
"5.00",
"*665120",
"Spanish",
"||",
"(P}",
"5.00",
"*+897370",
"AP",
"Computer",
"Sci",
"(HP}",
"5.00",
"*697120",
"Hon",
"Pre-Calculus",
"(HP}",
"C-",
"5.00",
"TERM:",
"GPA",
"3.17",
"Credits",
"30.00",
"CUMULATIVE;",
"GPA",
"3.15",
"Credits",
"130.00",
"Course",
"Flags",
"(*:UC/CSU",
"Approved,",
"+:Weighted",
"GPA)",
"Column",
"Labels",
"(Mrk:",
"Academic)",
"I",
"certify",
"that",
"the",
"above",
"information",
"is",
"correct.",

```

    "Date",
    "this",
    "certificate",
    "issued",
    "06/15/2015",
    "ipal",
    "ar",
    "Registrar"
]

```

Output:

Student Name/Address/Phone Student IO BOK Official Transcript P00000QQQOM
 Lcenenenenimenl POCOOCO) POCO OOOO POOQOOCOOOO P000000000ON Counselor Term
 Ending] Class of Grade 07/28/2016] 2017 12 Exit Reason Grad Date Issued To Print Date GPA Type
 GPA Crat Atmpt Overall Weighted 3.16 190.00 09/15/2016 1 of 1 Total Credits Earned 190.00 Course
 10 Course fark Credits || Course (D_Course irk Credits | Graduation Requirements Short Req cmp]
 D00000000Y Grd 0) HSSem | 01/14 | eereren Grd) HSSem1 01716] Comprehensive Grad
 Requirement 712223 © PE Course 1 A 5.00 |+737450 AP Biology (HP) Cc 5.00] English 10.00 40.00
 30,00737500 Biology (P) B- 5.00 |+748380 AP Psychology (HP) 8. 5.00 Global Studies 5.00 \$,00
 728021 Health Education B 5.00] "748180 US Hist (P) A 5.00] World History 10.00 10.00 653635
 Freshman English (P) c+ 5.00]653795 Junior English (P) fom 5.00] US History 10.00 40,00 865100
 Spanish | (P) B 5.00] "685140 Spanish Ili (P} B 5.00] Civics 5.00 5.00 9,00 697110 Hon Geometry (P) c
 5.00 |+697230 AP Calculus AB (HP) D+ 5.00] Economics 5.00 5.00 0.00 TERM: GPA 2.83. Credits 30.00
 | TERM: GPA 2.83 Credits 30.00] Math 20.00 20.00 CUMULATIVE; GPA 2.83 Credits 30.00]
 CUMULATIVE: GPA 3.08 Credits 160.00 | Physical Science 10.00 10.00 Life Science 10.00 10.00
 D00000000F Grd 09 HSSen2 06/14 | TeTerere Grd (1 HSSem2 06/16] VPA/World Language 10.00
 10.00 712223 PE Course 1 A 5.00] +737450 AP Biology (HP) c 5.00] Physical Education 10.00 20.00
 10.00 "737500 _Biclogy (P) B+ 5.00 |+748380 AP Psychology (HP) B 5.00] Health Education 5.00 5,00
 "748330 Global Studies (P) B 5.00 | 748180 US Hist (P) A 5.00] Electives 10,00 80.00 70,00 "653635
 Freshman English (P) B+ 6.00 | "653795 dunior English (P) F 0.00} Total Grad Credits 40.00 230.00
 190.00 "865100 Spanish | (P) B+ 5.00 | "665140 Spanish fil (P) B 5.00] All Courses. 190.00 190.00
 "697110 Hon Geometry (P} C+ 5.00 |+697230 AP Calculus AB (HP) B 5.00] Speech Prof Passed TERM:
 GPA 3.00 Credits 30.00 | TERM: GPA 3.60 Credits 26.00 CUMULATIVE: GPA 2.92 Credits 60.00]
 CUMULATIVE: GPA 3.46 Credits 185.00 | RE ———| T Out af District/Transfer Credits Grd (0 Summer
 Se 08/14 | AS SSTeTeTS Grd (1 HS Summer 07/16] EE "TRUCO12 Sum Alg2 (P) / Transfer Bt 10.00 |
 "653795SS Junior English (P) B+ 5.00 TERM: GPA 3.00 Credits 10.00 | TERM: GPA 3.00 Credits 5.00
 CUMULATIVE: GPA 2.93 Credits 70.00) CUMULATIVE: GPA 3.16 Credits 190.00 D00000000F Grd 0
 HS Sen} O1/1S [Work In Progress Entry Exit °737660 Chemistry (P) A. 5.00 Grd (2 748450 Wrid Hist 2-
 3 (P) A 5.00] 748220 Civics (P) O8NSNS 12/2216 "653655 Sophomore English (P} A 5.00] 624250
 Photography 1 (P) 08/15/16 06/02/17 "665120 Spanish Il (P) Ae 5.00] 633210 Lit & Soc Just (P) Ossi
 06/02/17+697370 AP Computer Sci (HP) c 5.00] 697145 AP Statistics (HP) 08/15/16 06/02/17
 "697120 Hon Pre-Calculus (HP) B- 5.00] 737700 Physics (P) Osis 06/02/17 TERM: GPA 3.67 Credits
 30.00] 748320 AP Human Geo (HP) OB15/16 06/02/17 CUMULATIVE: GPA 315 Credits 100.00]
 748260 Economics (P) 12/2316 06/02/17 D00000000F Grd 10 HS Sen 2 06/15 "737660 Chemistry
 (P) B+ 5.00 748450 Wrid Hist 2-3 (P) 8 5.00 "853655 Sophomore English (P) B+ 5.00 665120 Spanish

II (P) A 5.00 +697370 AP Computer Sci (HP) B 5.00697120 Hon Pre-Calculus (HP) c- 5.00 GPA 3.17
Credits 30.00 GPA 3.15 Credits 130.00

Transcript Sample #2

District Name: High School Transcript

Student Number: Grade: 12

Generated Page 1 of 1

Student Information				Course				Credit Summary						
Student Number:		Grade:	12	Course	Mark	Weight	Credit	Curriculum Program: High School						
Birthdate:		Gender:	M	2016-2017 Grade 10 Term 1				High School	Attempted	Earned	Required	Remaining		
State ID:				7060 A Choir	A	5.0000	5	English	30.000	30.000	40.000	10.000		
Counselor:				3620 Chemistry H	A	5.0000	5	ELD	0.000	0.000	0.000	0.000		
Place of Birth:	United States			4440 Chinese 4 H	A	5.0000	5	Algebra	10.000	10.000	10.000	0.000		
Diploma Date:				4850 Law	B	5.0000	5	Geometry	20.000	20.000	10.000	0.000		
				2400 Math Analysis	B	5.0000	5	Life Science	20.000	20.000	10.000	0.000		
				1170 World Studies	B	5.0000	5	Physical Science	10.000	10.000	10.000	0.000		
				1770 World Studies SS	A	5.0000	5	World History	10.000	10.000	10.000	0.000		
				Credit: 35.000 U/W GPA: 3.571				US History	10.000	10.000	10.000	0.000		
GPA Summary				2016-2017 Grade 10 Term 2				Economics	0.000	0.000	5.000	5.000		
				7060 A Choir	A	5.0000	5	US Government	0.000	0.000	5.000	5.000		
Cumulative GPA (Unweighted)	3.400					3620 Chemistry H	B	5.0000	5	Physical Education	10.000	10.000	20.000	10.000
CA Cal Grant GPA	3.423					4440 Chinese 4 H	B	5.0000	5	World Language**	30.000	30.000	10.000	0.000
Unwghtd 10-12 A-G GPA	3.375					4850 Law	B	5.0000	5	Fine Arts**	20.000	20.000	10.000	0.000
Unwghtd 9-12 A-G GPA	3.324					2400 Math Analysis	B	5.0000	5	Applied Academics**	30.000	30.000	10.000	0.000
				1170 World Studies	B	5.0000	5	Electives	0.000	0.000	60.000	0.000		
				1770 World Studies SS	A	5.0000	5	Total	200.000	200.000	220.000	30.000		
				Credit: 35.000 U/W GPA: 3.286				Comments						
Enrollment Summary				2017-2018 Grade 11 Term 1				Unofficial transcript						
Start/End Date	Grade	School												
08/17/2015-06/02/2016	09		1130 Amer Lit/Writ B 5.0000 5											
08/15/2016-06/01/2017	10		3120 AP Biology A 5.0000 5											
08/14/2017-05/31/2018	11		4450 AP Chinese Lng/Cul A 5.0000 5											
08/20/2018-	12		1750 AP US History B 5.0000 5											
				2420 Pre-Calculus H B 5.0000 5										
				4870 Princ of Mktng A 5.0000 5										
				Credit: 30.000 U/W GPA: 3.500										
				2017-2018 Grade 11 Term 2										
				1130 Amer Lit/Writ B 5.0000 5										
				3120 AP Biology A 5.0000 5										
				4450 AP Chinese Lng/Cul B 5.0000 5										
				1750 AP US History B 5.0000 5										
				2420 Pre-Calculus H B 5.0000 5										
				4870 Princ of Mktng A 5.0000 5										
				Credit: 30.000 U/W GPA: 3.333										
				Work In Progress										
				2430 AP Calculus AB 5.000										
				1875 AP Macroeconomics 5.000										
				3750 AP Physics 1 5.000										
				4860 Money and Banking 5.000										
				1300 Myth/Folk/Writ 5.000										
				2740 PE Wt Training 5.000										


```

text = pytesseract.image_to_string(img)
#print(text, file=open("output2.txt", "a"))
print(text)

```

```

In [16]: d2 = pytesseract.image_to_data(img, output_type=Output.DICT)
d2.keys()
#print(text, file=open("output1.txt", "a"))
d2["text"]

while("" in d2["text"]) :
    d2["text"].remove("")

while(" " in d2["text"]) :
    d2["text"].remove(" ")

import json
jsonStr2 = json.dumps(d2['text'], indent=4)

print(jsonStr2)

```

```

[
  "|",
  "OPEL",
  "LWWA",
  "[_____|",
  "Bate",
  "of",
  "Birth",
  " ",
  "Official",
  "Transcript",
  "EEE",
  "EL",
  "XXX",
  "XXX)",
  "State",
  "Student",
  "1p",
  "KOC",
  "OL",
  "ES",
  "POCONOS",
  "Place",
  "of",
  "Birth",
  "Term",
  "Ending",
  "[ExitDate_|",
  "_ExitReason",
  "|",
  "Grad",
  "Dato",
  "Diploma",
  "Type",
  "Issued",
  "To",
  "GPA",
  "Type",
  "GPA",
  "Crdt",
  "Atmpt",
  "Overall",

```

"Weighted",
"190.00",
"09/15/2016",
"1",
"of",
"|",
"Total",
"Credits",
"Earned",
"190.00",
"Nirk",
"Credits",
"||",
"C",
"Mrk",
"Credits",
"||",
"Graduation",
"Requirements",
"Short",
"Course",
"IO",
"Course",
"ourse",
"ID",
"Course",
"CCC",
"CES",
"Grd",
"09",
"HS",
"Sem",
"|",
"01/14",
"|",
"eee",
"Grd",
"11",
"HS",
"Sem",
"1",
"01/1\u00e9",
"|",
"Comprehensive",
"Grad",
"Requirement",
"412223",
"PE",
"Course",
"1",
"A",
"5.00)",
"*+737450",
"AP",
"Biology",
"(HP)",
":",
"5.00]",
"English",
"10.00",
"40.00",
"30.00",
" ",
"*737500",
"Biology",

"(P)",
"B-",
"5.00",
"|*+748380",
"AP",
"Psychology",
"(HP}",
"5.00]",
"Global",
"Studies",
"5.00",
"5.00",
"729G21",
"Health",
"Education",
"B",
"5.00",
"|",
"*748180",
"US",
"Hist",
"(P)",
"5.00",
"|",
"World",
"History",
"10.06",
"10.00",
"*653635",
"Freshman",
"English",
"(P)",
"C+",
"5.00",
"|",
"*653795",
"Junior",
"English",
"(P)",
"5.00]",
"US",
"History",
"10.00",
"10.00",
"*865100",
"Spanish",
"|",
"(P}",
"B",
"5.00",
"|",
"*655140",
"Spanish",
"t11",
"{P}",
"5,00]",
"Civies",
"5.00",
"5.00",
"0.00",
"*697110",
"Hon",
"Geometry",
"(P}",
"Cc",

"5.00",
"|",
"*+697230",
"AP",
"Calculus",
"AB",
"{HP}",
"5.00",
"|",
"Economics",
"\u00a7.00",
"5.00",
"0.00",
"TERM:",
"GPA",
"2.83",
"Credits",
"30.00",
"|",
"TERM:",
"GPA",
"2.83",
"Credits",
"30.00]",
"Math",
"20.06",
"20.00",
"CUMULATIVE;",
"GPA",
"2.83",
"Credits",
"30.00",
"|",
"CUMULATIVE:",
"GPA",
"3.09",
"Credits",
"160.00",
"|",
"Physical",
"Science",
"10.06",
"10.00",
"Life",
"Science",
"10.06",
"10.00",
"LECCE",
"CES",
"Grd",
"09",
"HS",
"Sem",
"2",
"0614",
"|",
"FCCC",
"Grd",
"(1",
"HS",
"Sen",
"2",
"06/16",
"|",
"VPA/World",

"Language",
"10.00",
"10.00",
"712223",
"PE",
"Course",
"1",
"5.00",
"|*+737450",
"AP",
"Biology",
"(HP)",
"5.00",
"|",
"Physical",
"Education",
"10.00",
"20.00",
"10.00",
"*737500",
"Biology",
"(P}",
"5.00]",
"*+748380",
"AP",
"Psychology",
"(HP)",
"5.00]",
"Health",
"Education",
"5.00",
"5.00",
"*748330",
"Global",
"Studies",
"\u00a2(P})",
"5.00",
"|",
"*748180",
"US",
"Hist",
"(P})",
"5.00]",
"Electives",
"10.00",
"89.00",
"79.00",
"\653635",
"Freshman",
"English",
"(P)",
"5.00",
"|",
"*653795",
"Junior",
"English",
"(PF)",
"0.00]",
"Total",
"Grad",
"Credits",
"40.00",
"230.06",
"190.00",
"*865100",

"Spanish",
"|",
"(P}",
"5.00",
"|",
"\\"655140",
"Spanish",
"I1]",
"(P}",
"5.00)",
"All",
"Courses",
"190.06",
"190.00",
"*697110",
"Hon",
"Geometry",
"(P}",
"5.00",
"|",
"*+697230",
"AP",
"Calculus",
"AB",
"{HP}",
"5.00]",
"Speech",
"Prag",
"Passed",
"TERM:",
"GPA",
"3.006",
"Credits",
"30.00",
"|",
"TERM:",
"GPA",
"3.60",
"Credits",
"25.06",
"CUMULATIVE;",
"GPA",
"2.92",
"Credits",
"60.00}",
"CUMULATIVE:",
"GPA",
"3.16",
"Credits",
"185.00",
"EEC",
"EEE",
"EEE",
"EEE",
"LE",
"Out",
"of",
"District/Transfer",
"Credits",
"Grd",
"{0",
"SummerSe",
"08/14",
"|",
"73ers",

"Grd",
"1",
"HS",
"Summer",
"07/16",
"Nererexe)",
"*TRUC012",
"=",
"Sum",
"Alg2",
"(P)",
"/",
"Transfer",
"Bf",
"10.00]",
"\653795SS",
"Junior",
"English",
"(P)",
"B+",
"5.00",
"TERM:",
"GPA",
"3.00",
"Credits",
"10.00",
"|",
"TERM:",
"GPA",
"3.00",
"Credits",
"5.00",
"CUMULATIVE:",
"GPA",
"2.93",
"Credits",
"70.00)",
"CUMULATIVE:",
"GPA",
"3.16",
"Credits",
"190.00",
"XCCECCECCES",
"Grd",
"10",
"HS",
"Sem",
"1",
"OL/LS",
"|",
"Work",
"In",
"Progress",
"\u201c737660",
"Chemistry",
"(P)",
"A.",
"500",
"|",
"RR",
"ISIererererer",
"Grd",
"2",
"748450",
"\u2014",

"Wrid",
"Hist",
"2-3",
"(P)",
"A-",
"5.00]",
"748220",
"Civics",
"(P)",
"085/15",
"12/226",
"*653655",
"Sophomore",
"English",
"(P}",
"A",
"5.00]",
"\u20ac24250",
"Photography",
"1",
"(P}",
"0815/16",
"06/02/17",
"*665120",
"Spanish",
"I1",
"(P}",
"A-",
"5.00",
"|",
"635210",
"Lit",
"&",
"Soc",
"Just",
"(P)",
"081516",
"06/02/17",
"*+697370",
"AP",
"Computer",
"Sci",
"(HP}",
"Cc",
"5.00]",
"697145",
"AP",
"Statistics",
"(HP)",
"0815/16",
"06/02/17",
"*697120",
"Hon",
"Pre-Calculus",
"(HP}",
"B-",
"5.00}",
"737700",
"Physics",
"(P}",
"0813165",
"6/02/17",
"TERM;",
"GPA",
"3.67",

"Credits",
"30.00",
"|",
"748320",
"AP",
"Human",
"Geo",
"(HP)",
"0815/16",
"a6/02/17",
"CUMULATIVE;",
"GPA",
"3.16",
"Credits",
"100.00",
"|",
"748260",
"Economics",
"(P)",
"12/2316",
"06/02/17",
"K00",
"Grd",
"()",
"HS",
"Sem",
"2",
"a6/15",
"*737660",
"Chemistry",
"(P)",
"5.00",
"*748460",
"Wrid",
"Hist",
"2-3",
"(P)",
"5.00",
"*853655",
"Sophomore",
"English",
"(P)",
"5.00",
"*665120",
"Spanish",
"||",
"(P)",
"5.00",
"*+897370",
"AP",
"Computer",
"Sci",
"(HP)",
"5.00",
"*697120",
"Hon",
"Pre-Calculus",
"(HP)",
"C-",
"5.00",
"TERM:",
"GPA",
"3.17",
"Credits",
"30.00",

```

"CUMULATIVE;",
"GPA",
"3.15",
"Credits",
"130.00",
"Course",
"Flags",
"(*:UC/CSU",
"Approved,",
"+:Weighted",
"GPA)",
"Column",
"Labels",
"(Mrk:",
"Academic)",
"I",
"certify",
"that",
"the",
"above",
"information",
"is",
"correct.",
>Date",
>this",
"certificate",
"issued",
"06/15/2015",
"ipal",
"ar",
"Registrar"
]

```

Output:

District Name: [eee eee eee ee eee) POQOOOKXXY) (aes High School Transcript Student Number:
gym Grade: 12 D000 GOO OOOO OOOOOOOOOOOOOOOO™DKDET Generated
EaEeerererererereres Page 1 of 1

Student Information XK KKKKK KKK KOKO Credit Summary Course Mark Weight Credit |Curriculum
Program: High School Student Number: (ae Grade: 12 Birthdate: POO) Gender: M 2016-2017 Grade
10 Term 1 High School Attempted Earned Required Remaining} State ID: POoooeg 7060 A Choir A
5.0000 5 [English 30.000 30,000 40.000 10.000 Counselor: POOOOO™XY 3620 Chemistry H A 5.0000
5 ELD 0.000 0.000 0.000 0.000 Place of Birth: United States 4440 Chinese 4H A 5.0000 5 —|Algebra
10.000 10.000 10.000 0.000 Diploma Date: 4850 Law B 5.0000 5 ~——|Geometry 20.000 20.000
10.000 0.000 2400 Math Analysis B 5.0000 5 —|Life Science 20.000 +~—«- 20.000 10.000 0.000 GPA
Summary 1170 World Studies B 5.0000 5 | Physical 1770 World Studies SS A 5.0000 5 —'|Science
10.000 10.000 10.000 0.000 Cumulative GPA (Unveighte) 3.400 Se NERD UStisery" 40.000 10000
10.000 0.000 2016-2017 Grade 10 Term 2 Economics 0.000 0.000 5.000 5.000 CA Cal Grant GPA
3.423 7060 A Choir A 5.0000 5 — Jus Government0.000 0.000 5.000 5.000 3620 Chemistry H B 5.0000
5 Physical Unwghtd 10-12 A-G GPA 3.375 4440 Chinese 4H B 5.0000 5 __—|Education 10.000 10.000
20.000 10.000 4850 Law B 5.0000 5 ~——|World 2400 Math Analysis B 5.0000 5—|Language **30.000**
+= 30.000 10.000 0.000 Unwalnid. 2:12. Ae1 GRA 8.824 1170 World Studies B 5.0000 5__s
Fine Arts 20.000 +=. 20.000 10.000 0.000 1770 World Studies SS A 5.0000 5 Applied Enrollment
Summary Credit: 35.000 U/W GPA: 3.286 [Academics** 30.000 30.000 10.000 0.000 Electives 0.000

Credit: 35.000 U/W GPA: 3.143