UNIVERSITY RESULT LEDGER SHEET ANALYSIS

Nageshkumar Mhetre, Sandeep Posham, Vitthal Narayankar, Sadguru Katakdhond

Computer Science & Engineering

N. B. Navale Sinhgad College Of Engineering, Solapur

Abstract

The traditionally universities declares the results in the form of PDF. When colleges get those PDF's, then the professors or the teachers has to take each record manually from the PDF for each student. Sometimes it takes 10-15 days to get the records from the PDF into an Excel. Generally there are thousands of students in a college and hundreds of students in a particular department. So it becomes difficult for the staff to analyze those data.

Introduction

The main purpose of this project is to provide an easy way that can help staff and students to get the data from PDF file which contains result of students. It becomes very difficult to scan the data and fetch the required data from that PDF because it contains a huge data that contains results of all the students who are studying in various colleges under the university.

So we have to do manual analysis on that PDF file which is not an easy task and it is very time consuming. That's why we are going to read the PDF file and extract required data from that file and convert it into excel file, so that the analysis becomes easier and faster than the previous.

In this project we're going to take input as a PDF file, process that file and extract all the required data like student's name, seat number, college name, department name, etc. Now this extracted data, we put that all data into an text file first and then convert that text

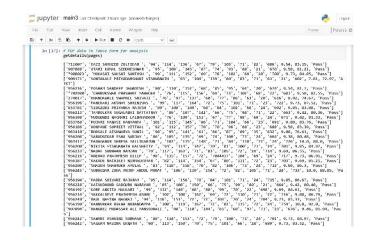
file into an Excel file. In Excel we're going show our output in the tabular form.

So that the user (staff/student) can easily work on that analyzed data in Excel file and they can compare the result of each student. So they'll know about the weakness and strength of students and they treat them accordingly.

Input PDF format:

Code	140	13410	Con	136.0	INC I	10	lbt 1365	POE C	rt Disco	1365	t Cot	150	1 63	1 200	150	al Res	Code	1 38	Man.	200	Man Dot	Man Dee	19331	PER DET	Sax	705 7531	Obt	Se I	42	269 12	10 20
Scott No.	(1243	4 1500	20170	32500	063568 853	C1 K 3	Laborer																								
Mi sue	66	OL/GI W	STORA 12	B 10.1	(a) (a) a 31	MORE.	-13																								
7004932	TE.	188		132		9 23		355		0		122					198662	Uses	Open I		244 144	100 (000)	155.3			120			9.00		0.05
2049112	TE	128		123						1-	1115	155					7051512	100	24		112 124	20 158				- 1		=	-		3 2
7044433	**			1		0 22		46	178	1	199	100					7044412	100	-		100				1175					50.001	
7044424	75					0 24				1							7786616	180					20			14					1 2
7054121	-	100		100			100	100		1		10	1991				7935519	1775	129 1	NO.	111 120	no ten	lan I			Dec 11			***		
7064411		Les	**	Des.	ine In		Les	100		i i war	178	160	1.60			1 8.2		1704	128 1	55	112:124	110 120	les I			lan I'	01	les I	***	441	RIX
774643-5					Ten In				1125		1104	Lo		00180-0	20.0	1.00	7064615	Late				10 123	135 1	40	175	Sec. 14					e 1 %
7044435					1					1							1 7064616	189.				10 124			125	1 12	4				9 1 5
7034416		1		1-						1				MITTO																	
H.K. N-a-	1	meal	Decks	. 24.	10	+2	Pr. 257.	60	10	ettes	6.71			Mink		070															
					abr 300+20												56.00 800						Coaler	**							
					AVI 362-25												50.00 EC														
		6216	DEMO	Prest	: MAT 2010											Credit:	\$1.00 BG														
ED IDEAS	No: 2										100000												GIRDA								
Count Tot	Mac C	17825	Scree	Donte	ed and Sev	colec					Per	ecute	pc: 8	. 51 -	180	A Marks	1 15		Itale	sec Se	oker Mill		GIADA								
Count Tot	Mac C	17825	Scree	Donte		colec	ó le Maria	es 253	Lec		lites 05	ecut o	pc: 8	. 51 -	180		1 15		Itale	sec Se			STREET.								
Great Tol.	Mar 9	11/825 corned	Screen 34.56	Ponts	od and Sov	enteci d dose	le Kum			741	95	ME/-21	P3. 4	6.91 · 73	180	A Marks	1 15		Itale	sec Se	oker Mill		GIADA:								
Great Tol.	No. 9	17825	14.16 14.16	22501	ed and Sev	enteur d over	tatera			. Cel	95	ME/-21	P3. 4	6.91 · 73	180	A Marks	1 15		Itale	sec Se	oker Mill		GT#1#								
Count Tol. Total Con- Deat For 216 Famou	11263 EA	10/825 ic gas same is	20170 20170	22501 8233	cd one Sec	entecr d mos or # 1 At 12	tateman MASSOCI	s Her	6924284		Lege Co	64 1 E	en a	901 - 713	180	A Marks	1 NA *14 ma		iBata Co	nec Se (Fo	skar Milli al Booles -	-									
CO ISSAT Score Tot Total Co- Deat Flor 200 Famous 7044417	Her P	17/825 10 808 10 808 14ZE III	20170 20170	22501 9233	od one Several Name of Several	enteur d over	Catenar STATES	n Hei	(125		167 167	ME/-21	en a	901 - 713	15	A Marks	1 84 1000 1 2004	prose	Itale	nec Se (Fo	skar Rull al Booles -	-	(44.)		1025	I 15				orai	
CO ISSAT Second Too Total Store Dead Flori 200 Warmer TO444417 TO444412	11263 EA	17/825 Grand 16 838 MADE 18 123	202.00 24.00 202.00 202.00 202.00 25	22901 9233	rel ese Secondarios de la consecución del consecución de la consec	entecr d (0.4) 6) 8 3 78 (2)	Catenar Catenar Castor	n Hei	(924284 (125		60 60 60 90	64) E	po. 4 serpce	-1 -	100	A Marks	1 NA 7000 1 7064411 1 7064412) C 200	Bata Go	inec Str	eke RJ) al deeles -	- I- 10 123	F-1		25	12	0:			1	9 1 0
CO ISSAN Grand Tot Total Store Dean Tot 230 Masser TO44411 TO44412 TO44412	Her C	17/825 Grand 16 808 MATE III 123 123	20270 20270 20270 20270	Bunda 1200 1200 122 122	colone Scotter Colone C	entecr d man or # 1 /# 123	tateman NASTOI	N Elet	1125 1125 1125		67 192 191	64 / E	bon 4 seroce	8.90 * .73	100	A Marks	1 NA Trans	prose	Bata	eec Se Fo	eker N.J.) al Novies -	10 123	[-]		125	12	0				1 2
Count Tot Total Total Deat Total 200 Minor TO44417 TO44412 Y004122 Y004122	Her C	17/825 16 808 16 808 123 123	20270 20270 20270 20270	22 bit 12	COUNCY TO	onteur d mos 61 E 2 78 123 9 22	Centeral Chateral Chateral	N Bet	(125 1100 1225 110		67 92 91 91 91	60 / E	20. 4 seroce	6.90 + -73 -1 - 50180.0	100	A Racks	1 NA 1 NA 1 NA 1 NA 1 NA 1 NA 1 NA 1 NA	91300 182 173	Bula	1000 Store	eker M.J.) al Mondes -	 10 10 10 1- 	[] []		125	- 12	0 0	0	Le.40	27,101	2 1 2
FD 15mm: Total format Total for	Her C	17/825 10 808 10 808 123 123	102 YO SECRETARY	22 6 1 22 6 1 22 6 1 22 6 1 22 6 1 23 6 1 24 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	per see the see that the see th	enteur d mon 6: 8: 1 78: 123 0: 25 0: 25 0: 34	Cateral Cateral Casteral	8 He1	(125 (125 (120 (125 (125 (125	11111	167 191 111 124	64 / E	20. 4 sersos	6.90 *	100	A Racks (400)	7. NA 70000 7044411 7044412 7044412 7044414	0100 182 173	Bala 20 	eec Se	Char Milli	- 10 10 	[] 35		(25 (100 (110 (36	- 12	0 0 0 1			20,00	9 1 2 2
FD 15mm: Count Tol. Touch Com- Dems 70:1 200 Warner TO44411 TO44422 Y064922 Y064922 TO44416 TO44424	Her C	17/825 School School Sale is 123 123 124	Horea 14.16 19.16 10 10 10 10 10 10 10 10 10 10 10 10 10	22 6 1 22 6 1 22 6 1 22 6 1 22 6 1 22 6 1 23 6 1 24 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	of one too	contect d rose 6: 8: 1 78: 123 0: 22 0: 24	Catenar STATESTON	8 E91	1924284 1325 1200 1220 131 131 131	11111	167 191 1114 165 134 167	10 10 10 10 10 10 10 10 10 10 10 10 10 1	120.	00180 o	100	A Marks (************************************	1 7044411 1 7044412 1 7044412 1 7044412 1 7044413 1 7044414 1 7044414	(#20 #20 1203 #14 1203	18sts 2s	Sec. Sec. Sec. Sec. Sec. Sec. Sec. Sec.	 12 29 12 29 	- 10 15 - -	[] 35	 43	(25 (100 (110 (30 (100	- 1	0 0 0 0 0	0	10.40	x1.10	0.000
FD 15mm; Seast Tol Truel Stor- Deat Tol 210 Name: 7044417 7044412 704412 704412 704414 704414 704414	Her C	17/825 Grand In 808 Mare II 123 123	20170 20170	2000 1200 122 123 124 127	20 one 500	0 25 0 24	ie Koman Mastori I-	5 E91	1125 1100 1220 111 115 115 115	11111	00 167 191 110 161 161 161	10 10 10 10 10 10 10 10 10 10 10 10 10 1	20. 4 000000 120. 130. 130.	6.91 + 73.	0 0 0	8 Marks (400) 1 8,0 1 8,0 1 8,0 1 8,0 1 8,0 1 8,0	1 55 1 50 1 7044411 1 7044412 1 7044413 1 704441 1 704441 1 704441	(6130 193 193 194 193 193	Bata 	erc Se	 	- - 10 D - - - - - - 10 D	 35 		(25 (100 (17) (30 (100 (100	111	0 0 0 0 0 0		19.40	N.10	0011100
FD 15mm: Grant Tot Total Transit TO44411 TO44412 Y064122 Y064122 TO44413 TO44614 TO46617 TO46617 TO46617 TO46617	Her C	17/825 16 808 123 123 123	20170 20170	22 Miles	25 17 17 17 17 17 17 17 17 17 17 17 17 17	onteci d (from 6) E (78 123 0 25 0 25 0 34	Le Monte de la Contraction de	6 He i	1125 1100 120 120 131 131 115 110 110 110	1 111111	00 00 00 00 00 00 00 00 00 00 00 00 00	1 011000	120. 120. 120. 120. 120.	6.90 •		8 Karks (4 Karks (4 Karks (4 Karks (4 Karks (4 Karks (4 Karks (4 Karks) (4 Karks (4 Karks) (4 Karrs) (4 Ka	1 84 1 2044411 1 7044412 1 7044412 1 7044414 1 7044411 1 7044411 1 7044411	(A)	Bala 	erc Se		10 (13) 10 (13) 10 (13) 10 (14) 10 (15)	 		25 123 135 135 135 135 137	111111	0 6 0 1 2 0	0	10.50	81.18 1.18	O D D D D D D D D D D D D D D D D D D D
FD 15mm; Seast Tol Truel Stor- Deat Tol 210 Name: 7044417 7044412 704412 704412 704414 704414 704414	Her C	17/825 Grand In 808 Mare II 123 123	Server 19. le 19170 1917	2000 1200 122 123 124 127	25 17 25 17 25 17 17 17 17 17 17 17 17 17 17 17 17 17	0 25 0 24	Ctateman Cta	5 E91	1924284 1125 1200 1220 173 175 176 1200 1220 115	11111	160 191 110 61 24 61 61 111	B 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	20. 4 00000 120. 120. 120. 120. 120.	6.90 •		R Racks (1 K) 0 (2 K) 0 (2 K) 0 (3 K) 0 (4 K)	1 55 1 50 1 7044411 1 7044412 1 7044413 1 704441 1 704441 1 704441	(6130 193 193 194 193 193	Bata 	erc Se	 	- - 10 D - - - - - - 10 D	 35 		25 123 135 135 135 135 137	111	0 6 0 1 2 0		19.40	81.18 1.18	000000000000000000000000000000000000000
FD 15mme Total Pro- 1 Total Pro	No. 0	17820 Grand RC 800 PACE III	20270 20270 20270 20200 201	Daniel 122 122 122 122 122 122 122 122 122 12	25 17 25 17 18 18 18 18 18 18 18 18 18 18 18 18 18	colected from 61 8 12 22 22 22 22 22 22 23 24 24 24 24 24 24 24 24 24 24 24 24 24	Crateral	65 E	1924204 1125 1200 1225 171 175 1200 1225 175 125	111111111	60 160 191 110 161 171 171 171 171 173 27	B 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	20. 4 00000 120. 120. 120. 120. 120.	6.50 + .73 E D0180.0 D0120.0 D0120.0 D0120.0	100 mm m	K,	1 84 1 2044411 1 7044412 1 7044412 1 7044414 1 7044411 1 7044411 1 7044411	(A)	Bala 	erc Se		10 (13) 10 (13) 10 (13) 10 (14) 10 (15)	 		25 123 135 135 135 135 137	111111	0 6 0 1 2 0	0	10.50	81.18 1.18	9 1 2 2
En ISaat Grant Tol Tourist From Dear Hai 200 Manual 7044411 7044412 704412 704413 704413 704413 704413 704413 704413 D.E. Doar	No. 0	17/820 0 200 0 200 0 200 123 123 123 123 123 124 125 125 125 125 125 125 125 125 125 125	Service 18. In 18. In 1	Posts 12 12 12 12 12 12 12 1	20 one 50	enteca d re-se 61 # 123 0 27 0 25 0 34	Contention of the Contention o	65 CC	1725 1200 1200 1215 171 175 120 120 125		160 180 191 110 101 101 101 101 101 101 101 10	64 E	120. 120. 120. 120. 120. 120. 120. 120.	6.50 5 -73 E -1 -1 -1 -1	100 E E E E E E E E E E E E E E E E E E	K, E E, C	7 854 7000 1 7044412 1 7044412 1 7044413 1 7044411 1 7044411 1 7044411 1 7044411 1 7044411	PER STREET	18sts Co	G G G		- 10 133 10 123 10 145 16 123	 30 		(28 (100 (100 (100 (100 (100 (100 (100 (10	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 1 0 2 2		19.40	0.00	ON STREET
ED (Seat Dol No. 2 To 1 To 2 To 2 To 2 To 2 To 2 To 2 To	No. 0	17825 School School Sales III 123 123 123 123 123 123 123 123 124 125 125 125 125 125 125 125 125 125 125	Screen 19. To a control of the contr	22 bit 122 iii ii	25 17 17 17 17 17 17 17 17 17 17 17 17 17	enteur d (00 en 6) E (00 78 (00 0 00 0 00 0 00 0 00 0 00 0 00 0 00	Contention of the Contention o	65 CC	1725 1200 1200 1215 171 175 120 120 125		160 180 191 110 101 101 101 101 101 101 101 10	64 E	120. 120. 120. 120. 120. 120. 120. 120.	6.50 5 73 6 6 6 7 7 7 7 8 7 8 7 8 9 9 9 9 9 9 9 9 9 9 9	EC	K,	1 85 1 2044411 1 7044412 1 7044412 1 704413 1 704413 1 704441 1 704441 1 704441 1 704441 1 704441 1 704441	STORE	18sts	GE CE		- 19 13 10 15 10 16 10 16 10 16	 30 20	 +3 18	(28 (100 (100 (100 (100 (100 (100 (100 (10	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 1 0 2 2		19.40	0.00	ON STREET
ED (Seat Dol No. 2) Common Tool	Her C	17925 00 808 80 808 123 123 123 123 123 123 123 123 123 123	Server 19.19 and	No. 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	25 17 25 17	enteur d (00 en 6) E (00 7) 27 0 25 0 34 0 34	Contention of the Contention o	65 CC	1725 1200 1200 1215 171 175 120 120 125		160 180 191 110 101 101 101 101 101 101 101 10	64 E	120. 120. 120. 120. 120. 120. 120. 120.	E 50 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	E0	K,	1 884 1 7044411 1 7044412 1 7044412 1 704413 1 704413 1 704413 1 704413 1 704413 1 704413 1 704413 1 704413 1 704413	92 (75.	Bata Co	G G G		- 1 10 123 - 1 - 1 10 123 10 125 10 128	 30 20 	4.3	(28 (100 (100 (100 (100 (100 (100 (100 (10	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 1 0 2 2		19.40	0.00	ON STREET
Court Tolly Company Co	No. C	17925 000000 100000000000000000000000000000	Service Servic	22 Miles	00 one 500 one	entecid (100 a) (27 a)	Contention of the Contention o	65 CC	1725 1200 1200 1215 171 175 120 120 125		160 180 191 110 101 101 101 101 101 101 101 10	64 E	120. 120. 120. 120. 120. 120. 120. 120.	6.50 5 73 6 6 6 7 7 7 7 7 8 7 8 10 10 10 10 10 10 10 10 10 10 10 10 10	E0	A Racks (18,0) (7 NA 44411 7044412 7044412 7044412 7044413 7044414 7044411 7044411 7044411 7044414 7044416	(192 193 193 193 193 193 193 193 193 193 193	18 sta 0	Sec. Sec. Sec. Sec. Sec. Sec. Sec. Sec.	Char R.I.I. al Sharks - 12 120 12 1	- 10 13 - 10 12 10 15 10 15	 30 20 		(28 (100 (100 (100 (100 (100 (100 (100 (10	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 1 0 2 2		19.40	0.00	ON STREET
Court Tolly Company Co	No. C	17925 000000 100000000000000000000000000000	Service Servic	22 Miles	25 17 25 17	entecid (100 a) (20 a)	Contention of the Contention o	65 CC	1725 1200 1200 1215 171 175 120 120 125		160 160 151 110 161 171 171 171 171 171 171 171	64 E	120. 120. 120. 120. 120. 120. 120. 120.	6.50 5 73 6 6 6 7 7 7 7 7 8 7 8 10 10 10 10 10 10 10 10 10 10 10 10 10	E0	A Racks (18,0) (7 NA 44411 7044412 7044412 7044412 7044413 7044414 7044411 7044411 7044411 7044414 7044416	(192 193 193 193 193 193 193 193 193 193 193	18 sta 0	100 Sec. 300		- 10 13 - 10 12 10 15 10 15	 30 20 		(28 (100 (100 (100 (100 (100 (100 (100 (10	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0 0 0 0 0 0 1 0 2 2		19.40	0.00	9 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1

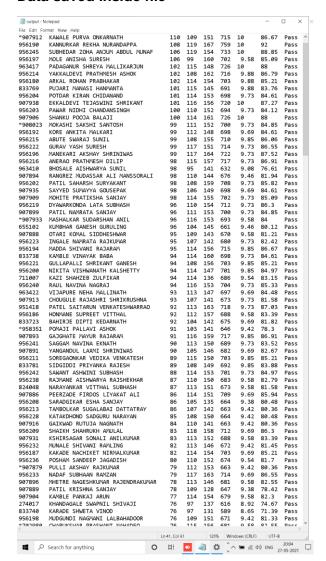
Extracted data from PDF file:



Literature Survey

Final Output:

Data saved inside file



PDF is a versatile document format, but it can be difficult to read and edit it. To easily extract a table or edit text in a spreadsheet format, you can to convert PDF to editable Excel spreadsheets. There few applications or online PDF to Excel convertors, but these applications convert the whole PDF in Excel. Many of the times PDF has some irrelevant data that also get converted to Excel.

To perform analysis the data must be in appropriate format with required relevant data.

An Excel/CSV file store the data in tabular form that helps analyst to analyze data without any difficulty on other hand these things cannot be performed on PDF. Traditionally data analysis performed on data in Excel/CSV, but there is not a reliable way available to get PDF data into Excel.

Implementation

University declares the result which is in the format of ledger sheet. If staff wants to analyze the result then it'll be difficult because the result is in the pdf format.

So in our project we are making the process of analysis easy.

But we need data in tubular form like Excel/CSV to do analysis, so we're converting that pdf file into Excel/CSV file so the user can analyze result of each student.

We're doing this project using python language and using ide as anaconda jupyter. Python has libraries that helps to convert data from one file format to another. We're currently using PyPDF2 and Numpy libraries for our project.

PyPDF2 is used for reading the data from pdf file and Numpy is used for working with arrays and lists that we are using for analysis.

At the beginning we take input file which is in the pdf format and we read all the data included in that file using PyPDF2 library.

Then we extract the only data which we required and store it in another file, after that we convert the data in the file in tables and there we do our final step as analysis.

In that we analyze student data such as marks gained by student in each Subject, Students Percentage/SGPA, Overall Marks, etc.

CONCLUSION:

After successful completion of this project we reached to the following conclusion.

- While working on this project, we worked with some unstructured data & tried to format the result in a structured tabular manner.
- ✓ We worked with many python libraries like PyPDF2, Numpy, tkinter, pandas, & tabula to achieve our goal and to solve the one of the problem in our institute and simplified the task of user for analysis purpose.

References

- https://pypi.org/
- https://www.anaconda.com

Future Scope

Now our project supports only the specific format of input PDF file.

In future many applications like this can be build which are robust enough to support multiple input formats.

This project is a standalone application till now but we can make it a web application by adding some extra features, So that any of the user can access this product remotely on browser/web.

Guidance:

Prof. H. T. Gurme (CSE)