

PROBLEM STATEMENT

- TO BUILD A MODEL WHICH DEFINES THE PHYSICAL AND LOGICAL STRUCTURES
- THE MODEL IS USED TO DETECT TABLES AND TO ANALYZE AND DECOMPOSE

EXISTING SOLUTIONS

- CURRENT SOLUTIONS ARE BUILT AROUND PDFs
- THEY STRUGGLE WITH IMAGES AND ARE ALSO NOT EFFICIENT
UNDERSTANDING THE UNDERLYING TABLE STRUCTURE

CHALLENGE

BUILDING A MODEL WHICH DEALS WITH IMAGES AS EFFICIENTLY AS POSSIBLE
WITH PDFs AND TO UNDERSTAND AND EXTRACT WIDE VARIETY OF TABLES
TABLE STRUCTURES

COMPLEXITY

NO BOUNDARIES

Your organisation's country of establishment (indicate your country of residence if answering as an individual person):

	replies	%
AT - Austria	79	(4.6%)
BE - Belgium	142	(8.2%)
DE - Germany	258	(14.9%)
DK - Denmark	28	(1.6%)
EL - Greece	80	(4.6%)
ES - Spain	123	(7.1%)
FI - Finland	22	(1.3%)
FR - France	132	(7.6%)
IE - Ireland	33	(1.9%)
IT - Italy	147	(8.5%)
LU - Luxembourg	1	(0.1%)
NL - Netherlands	89	(5.1%)
PT - Portugal	81	(4.7%)
SV - Sweden	68	(3.9%)
UK - United Kingdom	157	(9.1%)
CY - Cyprus	12	(0.7%)
CZ - Czech Republic	10	(0.6%)

PARTIAL BOUNDARIES

Table A Annual growth rates of debt securities issued by euro area countries

(percentages; end of period)	2006	2007	2008
Total general government	2.5	2.8	8.1
Long-term	3.4	2.3	3.7
Fixed rate	3.4	2.0	3.5
Floating rate	3.4	5.4	5.1
Short-term	-8.8	9.6	61.9

Source: ECB.

Table B Structure of amounts outstanding of debt securities issued by euro area countries

(percentages of total debt securities issued by general government; end of period)	2006	2007	2008
Long-term	92.9	92.5	88.8
Fixed rate	83.9	83.2	79.7
Floating rate	8.0	8.2	8.1
Short-term	7.1	7.5	11.2
Total general government in EUR billions	4,710.7	4,841.8	5,266.2

Source: ECB.

COMPLEXITY

NESTED ROWS AND COLUMNS

Table 6
Correlations between the difficulties encounters and the perceived impacts on pupils, teachers and the school as a whole (Pearsons correlation coefficient*)

	Difficulties encountered with respect to		
	Lack of interest/ acceptance from colleagues	Lack of interest of pupils	Lack of interest of parents
Impacts on participating pupils			
Knowledge and awareness of different cultures	-0,1651	-0,2742	-0,1618
Foreign language competence	-0,0857	-0,1804	-0,1337
Social skills and abilities	-0,1237	-0,2328	-0,1473
Acquaintance of special knowledge	-0,1355	-0,2008	-0,1234
Self competence	-0,1291	-0,2466	-0,1636
Impacts on participating teachers			
Knowledge/appreciation of school system and education in the partner countries	-0,1505	-0,1636	-0,1349
Foreign language competence	-0,0545	-0,0997	-0,0519
Social skills and personal commitment	-0,2558	-0,2235	-0,1302
Professional knowledge and abilities	-0,2145	-0,2319	-0,1003
Impacts on the school as a whole			
European/International dimension of the school	-0,2438	-0,1945	-0,1030
School climate	-0,2976	-0,1810	-0,1012
Innovation in teaching and school management	-0,2586	-0,2557	-0,0928
Training of teachers	-0,1839	-0,1703	-0,0518
Involvement of external actors in the every day school-life	-0,2346	-0,2343	-0,2237
International mobility of pupils	**	**	-0,0583

* Significance p = 0,000

** No significant correlation

NOISY IMAGES

Payslip for the month of May 2017

Employee No.:	Location:
Name:	Department:
Date of Joining:	Designation:
Bank Name:	Total Days:
Bank Acc. No.	LOP:
PF No.:	LOP Reversal:
ESI No.:	Work Days:
PAN No.:	Effective Work Days:
UAN Number:	

Earnings	Rs.	Deductions
Basic	14163	PF
HRA	7081	INCOME TAX
Special Allow	3982	PT
Shift Allowance	2500	
Incentive	14131	
Statutory Bonus	1400	
Total Earnings	43257	Total Deductions
Net Pay : Rs. 40280		

(Rupees Forty Thousand Two Hundred Eighty)

Handwritten signature

METHODOLOGY

PRE- PROCESSING

- SKEW-CORRECTION
- LINE EXTRACTION
- TEXT INVERSION

TABLE DETECTION

- BOUNDED FIGURES & TABLE DETECTION
- TEXT BLOCK FORMATION
- HORIZONTAL & VERTICAL ALIGNMENT VERIFICATION

TABLE EXTRACTION

- ROWS & COLUMNS IDENTIFICATION
- HEADERS' IDENTIFICATION
- MAPPING CELL'S ROW & COLUMN HEADER(S)

EXAMPLE

OUTPUT

INPUT

Contract number - RSM 01
Final Report

- o S12.5: Country Net Yearly Salary Average in EURO

Country	Net Yearly Salary average	Country	Net Yearly Salary average
Austria	21,522	Italy	21,522
Belgium	25,406	Latvia	9,150
Bulgaria	3,361	Lithuania	6,340
Croatia	12,477	Luxembourg	45,847
Cyprus	33,665	Malta	19,686
Czech Republic	11,514	Netherlands	36,791
Denmark	55,874	Norway	39,407
Estonia	7,865	Poland	8,059
Finland	26,288	Portugal	18,111
France	25,380	Romania	5,765
Germany	26,235	Slovakia	6,085
Greece	20,105	Slovenia	14,270
Hungary	9,423	Spain	22,500
Iceland	33,636	Sweden	28,075
Ireland	34,530	Switzerland	64,123
Israel	26,703	Turkey	14,565
		United Kingdom	37,001

Table 48 - Country Net Yearly Salary Average (2006, N=6,934, all currencies in EURO)

Note: A different analysis has been carried out for the net yearly salary costs obtained in the survey in order to detect unusual observations. The final sample had 7,018 correct replies, including 84 answers from Marie Curie fellowships. Marie Curie answers has not been considered for the calculation of the country net yearly salary average, as a result the sample had N=6,934.

[MAIN COLUMN HEADER/ROW HEADER/CONTENT]

	A	B	C
1	1		
2	Country Austria		Net Yearly S
3	Country Belgium		Net Yearly S
4	Country Bulgaria		Net Yearly S
5	Country Croatia		Net Yearly S
6	Country Cyprus		Net Yearly S
7	Country Czech Republic		Net Yearly S
8	Country Denmark		Net Yearly S
9	Country Estonia		Net Yearly S
10	Country Finland		Net Yearly S
11	Country France		Net Yearly S
12	Country Germany		Net Yearly S
13	Country Greece		Net Yearly S
14	Country Hungary		Net Yearly S
15	Country Iceland		Net Yearly S
16	Country Ireland		Net Yearly S
17	Country Israel		Net Yearly S
18	2		
19	Country Italy		Net Yearly S
20	Country Latvia		Net Yearly S
21	Country Lithuania		Net Yearly S
22	Country Luxembourg		Net Yearly S
23	Country Malta		Net Yearly S
24	Country Netherlands		Net Yearly S
25	Country Norway		Net Yearly S
26	Country Poland		Net Yearly S
27	Country Portugal		Net Yearly S
28	Country Romania		Net Yearly S
29	Country Slovakia		Net Yearly S
30	Country Slovenia		Net Yearly S
31	Country Spain		Net Yearly S
32	Country Sweden		Net Yearly S
33	Country Switzerland		Net Yearly S
34	Country Turkey		Net Yearly S
35	Country United Kingdom		Net Yearly S

PRE-PROCESSING

SKEW-CORRECTION

LINE DETECTION

LINE REMOVAL

NOISE REDUCTION

• 512 • **Country Yearly Salary Average in U.S.D.**

Country	Yearly Salary Average	Country	Yearly Salary Average
Algeria	29,188	Denmark	39,847
Argentina	20,488	Egypt	9,425
Australia	32,474	France	35,267
Canada	32,474	Germany	49,765
China	17,814	Holland	39,847
Costa Rica	20,488	Italy	39,847
Czech Republic	30,584	Japan	54,171
Denmark	39,847	Peru	12,566
France	35,267	Poland	24,220
Germany	49,765	Spain	24,220
Greece	24,220	Sweden	49,765
Holland	39,847	Switzerland	54,171
India	20,488	Taiwan	49,765
Indonesia	20,488	Thailand	24,220
Italy	39,847	U.S.A.	54,171
Japan	54,171	U.S.S.R.	20,488
South Korea	39,847	U.S.S.R. (Russia)	20,488
Spain	24,220		
Sweden	49,765		
Switzerland	54,171		
Taiwan	49,765		
Thailand	24,220		
U.S.A.	54,171		
U.S.S.R.	20,488		
U.S.S.R. (Russia)	20,488		

Journal of Interpersonal Violence 26(10) 1997-2013
© The Author(s) 2011
Reprints and permissions: sagepub.com/journalsPermissions.nav
DOI: 10.1177/0886260511419111
jiv.sagepub.com

© 32.3. Country Net Yearly Salary Average in Euro

Country	Net Yearly Salary
Belgium	33,382
Bulgaria	25,492
Croatia	22,462
Czechia	17,477
Denmark	35,411
France	31,114
Germany	35,411
Greece	25,500
Italy	25,500
Netherlands	28,251
Norway	35,411
Poland	25,500
Portugal	25,500
Spain	25,500
Sweden	35,411
Switzerland	35,411
United Kingdom	25,500
United States	25,500


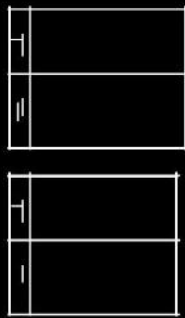
Created with  TabletoPy 1000000

Table 48 - Country Net Yearly Salary Average (2006, 44-934, all currencies in EUR)

[illegible]

Table 48 - Country Net yearly Salary Average (2006, March 9th, all currencies in EUR)



• 512 • **Country Yearly Salary Average in U.S.D.**

Country	Yearly Salary Average	Country	Yearly Salary Average
Albania	29,188	Dominican Republic	14,512
Algeria	20,000	Ecuador	15,425
Andorra	12,474	Egypt	49,267
Angola	12,474	El Salvador	15,766
Antigua	17,814	Equatorial Guinea	39,867
Armenia	35,864	Estonia	34,171
Australia	30,768	Ethiopia	12,506
Austria	31,535	Finland	34,171
Azerbaijan	19,423	France	24,203
Bahamas	19,423	Gabon	31,573
Bahrain	30,768	Ghana	14,512
Bangladesh	20,000	Greece	35,864
Barbados	29,188	Guatemala	14,512

Journal of Interpersonal Violence 26(10) 1997-2013
© The Author(s) 2011
Reprints and permissions: sagepub.com/journalsPermissions.nav
DOI: 10.1177/0886260511418811
jiv.sagepub.com

© 2013 Country Net Yearly Salary Average in Euro

Country	Net Yearly Salary
Algeria	33,482
Argentina	25,492
Australia	42,452
Austria	37,477
Canada	40,487
China	11,614
China-Taipei	25,471
Denmark	35,471
France	35,471
Germany	35,480
Ghana	22,581
Hong Kong	35,471
India	3,425
Indonesia	34,425
Ireland	34,425
Israel	25,303
Italy	34,425
Japan	34,425
Kenya	25,303
Malaysia	34,425
Malta	34,425
Mexico	34,425
Morocco	34,425
Netherlands	34,425
Norway	34,425
Peru	34,425
Portugal	34,425
Romania	34,425
Saudi Arabia	34,425
Spain	34,425
Sweden	34,425
Switzerland	34,425
Taiwan	34,425
Tanzania	34,425
Thailand	34,425
Turkey	34,425
Ukraine	34,425
United Kingdom	34,425
United States	34,425
Vietnam	34,425
Zimbabwe	34,425

Table 48 - Country Net Yearly Salary Average (2006, 44-934, all currencies in EUR)

© S&P 3. Country Risk Value Average in EuroQ

Country	1st Year Value	2nd Year Value
Belgium	22.98	23.97
France	22.98	23.97
Germany	22.98	23.97
Italy	22.98	23.97
Spain	22.98	23.97
United Kingdom	22.98	23.97
United States	22.98	23.97
Japan	22.98	23.97
Canada	22.98	23.97
Australia	22.98	23.97
Sweden	22.98	23.97
Norway	22.98	23.97
Denmark	22.98	23.97
Finland	22.98	23.97
Portugal	22.98	23.97
Greece	22.98	23.97
Switzerland	22.98	23.97
Austria	22.98	23.97
Netherlands	22.98	23.97
Luxembourg	22.98	23.97
Malta	22.98	23.97
Cyprus	22.98	23.97
Poland	22.98	23.97
Czech Republic	22.98	23.97
Slovakia	22.98	23.97
Slovenia	22.98	23.97
Hungary	22.98	23.97
Romania	22.98	23.97
Bulgaria	22.98	23.97
Greece	22.98	23.97
Spain	22.98	23.97
Portugal	22.98	23.97
Italy	22.98	23.97
France	22.98	23.97
Germany	22.98	23.97
United Kingdom	22.98	23.97
United States	22.98	23.97
Japan	22.98	23.97
Canada	22.98	23.97
Australia	22.98	23.97
Sweden	22.98	23.97
Norway	22.98	23.97
Denmark	22.98	23.97
Finland	22.98	23.97
Portugal	22.98	23.97
Greece	22.98	23.97
Switzerland	22.98	23.97
Austria	22.98	23.97
Netherlands	22.98	23.97
Luxembourg	22.98	23.97
Malta	22.98	23.97
Cyprus	22.98	23.97
Poland	22.98	23.97
Czech Republic	22.98	23.97
Slovakia	22.98	23.97
Slovenia	22.98	23.97
Hungary	22.98	23.97
Romania	22.98	23.97
Bulgaria	22.98	23.97
Greece	22.98	23.97
Spain	22.98	23.97
Portugal	22.98	23.97
Italy	22.98	23.97
France	22.98	23.97
Germany	22.98	23.97
United Kingdom	22.98	23.97
United States	22.98	23.97
Japan	22.98	23.97
Canada	22.98	23.97
Australia	22.98	23.97
Sweden	22.98	23.97
Norway	22.98	23.97
Denmark	22.98	23.97
Finland	22.98	23.97
Portugal	22.98	23.97
Greece	22.98	23.97
Switzerland	22.98	23.97
Austria	22.98	23.97
Netherlands	22.98	23.97
Luxembourg	22.98	23.97
Malta	22.98	23.97
Cyprus	22.98	23.97
Poland	22.98	23.97
Czech Republic	22.98	23.97
Slovakia	22.98	23.97
Slovenia	22.98	23.97
Hungary	22.98	23.97
Romania	22.98	23.97
Bulgaria	22.98	23.97
Greece	22.98	23.97
Spain	22.98	23.97
Portugal	22.98	23.97
Italy	22.98	23.97
France	22.98	23.97
Germany	22.98	23.97
United Kingdom	22.98	23.97
United States	22.98	23.97
Japan	22.98	23.97
Canada	22.98	23.97
Australia	22.98	23.97
Sweden	22.98	23.97
Norway	22.98	23.97
Denmark	22.98	23.97
Finland	22.98	23.97
Portugal	22.98	23.97
Greece	22.98	23.97
Switzerland	22.98	23.97
Austria	22.98	23.97
Netherlands	22.98	23.97
Luxembourg	22.98	23.97

Table 48 - Country Net yearly Salary Average (2006, March 9th, all currencies in EUR)

TABLE DETECTION

BOUNDED FIGURES &
TABLE DETECTION

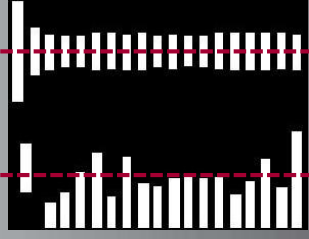
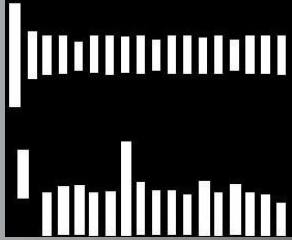


HORIZONTAL &
VERTICAL ALIGNMENT
VERIFICATION



Country	Net Yearly Salary average
Austria	31,552
Belgium	25,408
Bulgaria	3,361
Croatia	12,477
Cyprus	33,865
Czech Republic	11,514
Denmark	35,874
Estonia	7,685
Finland	26,551
France	30,030
Germany	28,351
Greece	9,423
Hungary	33,936
Iceland	34,930
Ireland	26,703
Israel	

Country	Net Yearly Salary average
Italy	21,821
Latvia	9,150
Lithuania	6,340
Luxembourg	45,847
Malta	19,688
Netherlands	36,791
Norway	39,407
Poland	8,059
Portugal	16,171
Romania	10,786
Slovakia	6,055
Slovenia	14,279
Spain	22,930
Sweden	28,075
Switzerland	64,123
Turkey	14,565
United Kingdom	37,001



Country	Net Yearly Salary average
Austria	31,552
Belgium	25,408
Bulgaria	3,361
Croatia	12,477
Cyprus	33,865
Czech Republic	11,514
Denmark	35,874
Estonia	7,685
Finland	26,551
France	30,030
Germany	28,351
Greece	9,423
Hungary	33,936
Iceland	34,930
Ireland	26,703
Israel	

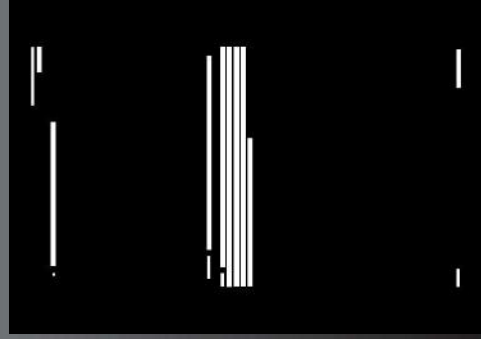
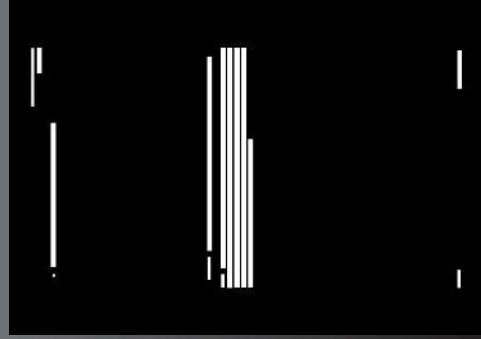


TABLE EXTRACTION

ROWS & COLUMNS
IDENTIFICATION



HEADER
IDENTIFICATION



MAPPING CELL'S
ROW & COLUMN
HEADER(S)



T

	A	1
1		Country Austria
2		Country Belgium
3		Country Bulgaria
4		Country Croatia
5		Country Cyprus
6		Country Czechia
7		Country Denmark
8		Country Estonia
9		Country Finland
10		Country France
11		Country Germany
12		Country Greece
13		Country Hungary
14		Country Iceland
15		Country Ireland
16		Country Israel
17		Country Italy
18		Country Latvia
19		Country Lithuania
20		Country Luxembourg
21		Country Malta
22		Country Netherlands
23		Country Norway
24		Country Poland
25		Country Portugal
26		Country Romania
27		Country Slovakia
28		Country Slovenia
29		Country Spain
30		Country Sweden
31		Country Switzerland
32		Country Turkey
33		Country United Kingdom
34		Country United States
35		Country United States
36		Country United States

CHALLENGES

NESTED HEADER MAPPING

Spain						
Scientific domain	Total Yearly Salary Costs of Researchers (data from the study)		Equivalent(s) profession(s) defined by ISCO classification	Category with available data in Eurostat	Total Yearly Salary Costs of similar professions from Eurostat	Male
	Male	Female				
Social and Human Sciences	27.301 €	16.806 €	2.4.4	2.4	46.657 €	46.657 €
Economics	42.978 €	33.669 €	2.4.1	2.4	42.138 €	42.138 €
Chemistry	39.091 €	20.296 €	2.1.1	2.1	37.111 €	37.111 €
Physics	27.197 €	16.067 €	2.1.1	2.1	42.138 €	42.138 €
Life Sciences	36.523 €	22.139 €	2.2.1	2.2	42.138 €	42.138 €
Mathematics	45.389 €	41.107 €	2.1.2	2.1	42.138 €	42.138 €
Information Sciences	26.059 €	25.893 €	2.1.3	2.1	42.138 €	42.138 €
Engineering Sciences	34.316 €	25.435 €	2.1.4	2.1	42.138 €	42.138 €
Environment and Geosciences	29.210 €	14.847 €	2.1.4	2.1	42.138 €	42.138 €

CHALLENGES

REPETITIVE COLUMN/ROW HEADERS

Country	Number of responses	% of total responses	Country	Number of responses	% of total responses
DE - Germany	258	14.9%	CH - Switzerland	25	1.4%
UK - United Kingdom	157	9.1%	FI - Finland	22	1.3%
IT - Italy	147	8.5%	LT - Lithuania	22	1.3%
BE - Belgium	142	8.2%	IL - Israel	17	1.0%
FR - France	132	7.6%	PL - Poland	16	0.9%
ES - Spain	123	7.1%	CY - Cyprus	12	0.7%
NL - Netherlands	89	5.2%	CZ - Czech Republic	10	0.6%
PT - Portugal	81	4.7%	EE - Estonia	10	0.6%
EL - Greece	80	4.6%	HU - Hungary	10	0.6%
AT - Austria	79	4.6%	BG - Bulgaria	8	0.5%
SV - Sweden	68	3.9%	SK - Slovak Republic	7	0.4%
TR - Turkey	44	2.5%	SL - Slovenia	6	0.3%
IE - Ireland	33	1.9%	MT - Malta	5	0.3%
NO - Norway	31	1.8%	LV - Latvia	4	0.2%
Other country	30	1.7%	LU - Luxembourg	1	0.1%
RO - Romania	29	1.7%	IS - Iceland	1	0.1%
DK - Denmark	28	1.6%	<i>TOTAL</i>	<i>1727</i>	<i>100%</i>

CHALLENGES

TABLE SPLITTING OVER MULTIPLE PAGES

Interactive Policy Making Online consultations		
Guidelines for future European Union policy to support research		
1777 responses received ¹ , 30/07/2004 - 15/10/2004		
Details		
I am answering as:	replies	%
individual person	540	(31.25)
governmental body	141	(8.23)
university/higher education	468	(26.43)
commercial organisation (including consultancy) more than 250 employees	115	(6.75)
commercial organisation (including consultancy) less than 250 employees	144	(8.13)
non-commercial organisation (e.g. trade association, trade union, employers' association, chamber of commerce, NGO)	113	(6.52)
other (please specify)	210	(12.63)
Your role in the organisation		
note - I am answering as an individual	replies	%
senior management	350	(20.25)
management	350	(20.25)
researcher	160	(10.43)
strategy/policy function	619	(35.83)
specialist expert	154	(9.83)
other (please specify)	178	(10.23)
210	12.13	
Your organisation's country of establishment (indicate your country of residence if answering as an individual person):		
AT - Austria	replies	%
BE - Belgium	79	(4.83)
DE - Germany	142	(8.23)
DK - Denmark	258	(14.93)
EL - Greece	28	(1.63)
ES - Spain	80	(4.83)
FI - Finland	123	(7.13)
FR - France	22	(1.33)
IE - Ireland	112	(7.43)
IT - Italy	33	(1.93)
LU - Luxembourg	149	(8.53)
NL - Netherlands	1	(0.13)
PT - Portugal	89	(5.13)
SE - Sweden	81	(4.73)
UK - United Kingdom	468	(26.43)
CY - Cyprus	152	(9.13)
CZ - Czech Republic	12	(0.73)
10	0.63	

¹ These are the overall figures based on all replies sent during the consultation. There are a number of countries included below. These are a group of countries from different countries that have not forwarded additional comments in favour of a specific policy. These countries are: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom, Cyprus, Czech Republic, and others. These countries have not specifically stated the reason for not submitting their comments.

Interactive Policy Making Online consultations		
Your organisation's geographical area of activities (indicate your area of activities if answering as an individual person):		
local	replies	%
regional	57	(3.23)
national	116	(6.73)
European	270	(15.43)
international	292	(16.93)
not applicable	153	(9.33)
40	2.33	
Your organisation's activity type (indicate your activity type if answering as an individual person):		
Higher Education	replies	%
Research	400	(23.13)
Industrial	818	(47.33)
Public Administration	199	(11.53)
Other (please specify)	30	(1.73)
730	42.33	

CHALLENGES

NOISY SCANNED DATA

Employee No		Bank Name	
Name		Bank Acc. No.	
Designation		IFSC Code	
Original DOJ		PF No.	
PAN		UAN	
Payable Days		Paid Days	

