MID 1

17F 8333

11/16/2021

```
#
# for (i in 1:48) {
   pth=paste('C:/Users/HP/Downloads/E10/pdf (',').pdf',sep=as.character(i))
#
  print(pth)
#
  pdf_convert(
#
     pth,
#
     format = "png",
#
    pages = 2,
#
     filenames = NULL,
#
     dpi = 72,
#
     antialias = TRUE,
      opw = "",
upw = "",
#
#
#
      verbose = TRUE
#
#
# }
#for (i in 1:48) {
 #pth=paste('pdf (',')_2.png',sep=as.character(i))
raw_img=image_read('pdf (1)_2.png')
image_ggplot(raw_img)
```

COLK GAS RESULATION AUTOEP	Ψ		
E-11 Gasoline En-deput baie price Galer			
With afford from Depender \$1, 201	ų.		
		30/1200	
UMBOOKS ESD FALIG EST FALIGET FOR MEMORY		72.55	
February TeoDient pryprior automates by PSG		72(0)	
1005 part Filme diky (Indillary at the (1856 of Ti)		+5)	
DISC Supervisideges the Notice of principality of Asi	72.50	65.33	
-10 KD/P PTK: (-+ 0)		. 15.30	
Plan. Ex troust Sale Fride	M0645	E1066	
	Repail	Septi	Day
ME Connect Standay (16 12 Black Prints	37.55	72.61	53.9
S FEN		6.83	0.5
Submod (F+G)	12:342	704	73.0
Price Differential Carte			
Jubertal (8+1)	75,03	73,04	730
C DEC - Page	2.81	1.61	2.3
Epitus Sweets Ne	2.73	8.73	
Nember Dayserschol by Found 4 o 4	45.40	15.74	154
Subdecial () + H + L + M)	97.48	95,29	851
Salas Tax	16.55	1820	16.7
Hax 2-Dispot Sale Prior (H + O)	112,99	200,49	3114
Difference between MUSAS and RDS [Explained] 80 to Compare by Stage			
		2.19%	

```
raw_img %>%
   image_crop(geometry_area(350, 0,350, 10)) %>%
   ocr()

## [1] "UTORITY\nice Calculation\nr
01,2019\nRs/Liter\n7255\n72.00\n7.20\n7255 6530\n72.50\nMOGAS _E-
10Gasoline\nRetall Retail Direct\n755 72507250\n337054 _OS4\n7592 73:04
73.08\n7592 73:04 73.08\n281 2e1 281\n370 370\n15001574 19.48\n9743 95.29
95.29\n1656 1620-1620\n11399 11149 11149\nsy -\nDy Sage 2.19%\n"

raw_img %>%
   image_quantize(colorspace = "gray") %>%
   image_ggplot()
```

CHE K GAS REGULATION AUTORS	TY		
E-11 Gasotine to - deput hale price City With affort from Dependen #1, 20			
		36,52200	
PROGRESSED Weighted Amongstreet of surprise		72.55	
Entered Test entirely programmous market by PSC		72.00	
1005 and Filmed 4 to Catallary of the (10% of Ti)		*2)	
CO Separt Micros By Telland and Association (ADM)	72.50	62.33	
P-10 BROWNER (-+ D)		Y\$.50	
Nos. Ex thours Sale Erice	M0545	Etsentie.	
	Retail	Retall	Direct
162 Control Supply 716 h 18 Anna Price	37.45	72.61	2,22.5
FEN	2,97		6.5
Submed (7+G)	15:302	72,04	73.04
Price Differential Carte		1976	83.04
(ubestal (H+1)	75,93	73,04	73.0
CMI - Page	2.81	1.61	2.0
Epolas Sienticke	2.73	8.73	
Nemaleum begraphs all of by hourself to 4	45.03	15.74	15.4
Soldated (I + R = L + M)	97.43	95,29	25.1
String Tipe	10.55	1823	16.2
Max. In-Depot Sale Price (N + 0)	112,99	311,45	3014

```
fuzz_fun <- function(fuzz){
  raw_img %>%
    image_quantize(colorspace = "gray") %>%
    image_transparent(color = "white", fuzz=fuzz) %>%
    image_background("white") %>%
    image_crop(geometry_area(350, 0,350, 10))
}
fuzz_fun(20)
```

UTORIT	Y			
ice Calcu				
r 01, 201	.9			
	Rs/Liter			
		72.55		
		72.00		
		7.20		
	72.55	65.30		
		72.50		
	MOGAS	E-10 Gas	Gasoline	
	Retail	Retail	Direct	
	72.55	72.50	72.50	
	3.37	0.54	0.54	
	75.92	73.04	73.04	
	-			
	75.92	73.04	73.04	
	2.81	2.81	2.81	
	3.70	3.70		
	15.00	15.74	19.44	
	97.43	95.29	95.29	
	16.56	16.20	16.20	
	113.99	111.49	111.49	
	(2.50)	_		
by %age	(=.=0)	2.19%		

```
combo_fuzz <- c(
  fuzz_fun(100),
  fuzz_fun(100),
  fuzz_fun(100),
  fuzz_fun(100)
) %>%
  image_append(stack = TRUE)
image_ggplot(combo_fuzz)
```

```
no_grid <- raw_img %>%
  image_quantize(colorspace = "gray") %>%
  image_transparent(color = "white", fuzz=20) %>%
  image_background("white")

image_ggplot(no_grid)
```

OH K GAS RESPUATO	HY AUTORITY		
D-11: Gasotian Ex- depit 5 di	e price calculation		
With a Pour from Tears	when 11, 2019		
		hylados.	
 PROGASTSO FALLS INCOME A DESCRIPTION 		72.55	
Rije tiz nei Pe@extl organization.com by 190		72.00	
C Ith Shart Florand K. Dutallas in 19 (10% of 7)		10	
DISCORpart Mopra By Bellevier Indicas SEM at A (72.71	65.11	
E TO BERTHER U LUI		V\$230	
Number Project Sale Entire	100645	ELEGE	seller
Page 65 F-5 F-5 F-5 F-5 F-5 F-5 F-5 F-5 F-5 F-	Reput	Rept.	Direct
F. M.S. Comp. J. Science, 7 K. 15 Barra Price	32.55	71.64	13.70
G IFEN	2.07	0.5	6.54
H Submod (F+G)	15,000	(2.04)	73,04
Frice Differential Calife			
/ (ubestal (E+F)	75,53	73,04	730.9
R CV: - Page	S 81	1.6	2.31
L Borus Conniction	2.71	4,773	
Milliamilian Segment all by the less 1994	10.60	15.74	12.41
H Salanta (1+H - L - M)	9748	95.29	9524
O Sales Tive	16.55	1500	16.73
P. Max Endingson Sale Profess (M + O'	113,99	JL.,49	111.79
Difference between MUSAN and ROLE (Excluded)	(2.57)		
	ear by Stage	2,15%	

```
no_grid %>%
  image_negate() %>%
  image_ggplot()
```

```
no_grid %>%
  image_negate() %>% # negate
  image_morphology(method = "Thinning", kernel = "Rectangle:20x1") %>%
  image_negate() %>% # back to white
  image_ggplot()
```

OFFICE AS RESULATIONS AUTO	RETY		
C-11: Gasotine Ex- depot safe price G	designed		
With a Poor from Dependen LL.	2019		
		B-78.4006	
PS0GAS750 Worg. 1.51 Co., reprint ht stephy		72.55	
lettered PoCoril pryprocumpanent by 190		72.00	
Strictor Floring C. Datallas, p. 16 (10% of 7)		1.70	
College to Modern to Andrew Advantage (Maria)	72.70	65.34	
L-10 82/9 VTC8 Q. + LIJ		V\$230	
Mass by Population Print	100645	ELIG	seller
	Retail	Retail.	Direct
9.5 Constituting, 78 College Price	32.55	77.64	13.7
FEN	2.37	0.5	0.5
(Subrical (F+G)	134,642	<2.04	73.0
Erke Dillmenti di Califi			
(ubodal (H+F)	75,53	75,04	7305
CM: - Mana	2.91	1.6.	2.3
Execute Committee in a	2.73	4,773	
Persieur ingransulial tylka av 104	10.60	15.74	14.4
Seasonal (II+M - L - M)	97.48	95.29	95.15
Sports	16.55	5520	16.7
Max 25-Bapor Sale Peler (M + 0"	113,99	JL.,49	
Difference between MUSAN and REG (Exchine)	(2.97)		
E 14/Java 4 by SA	44	2,15%	

```
no_grid %>%
  image_crop(geometry_area(350, 0,350, 10)) %>%
  image_ggplot()
```

```
no_grid_crop <- no_grid %>%
  image_crop(geometry_area(350, 0,350, 10))
no_grid_crop %>%
  image_ggplot()
```

```
### 100 FFF | 10
```

```
no_grid_crop %>%
  image_ocr()
## [1] "UTORITY\nice Calculation\nr01, 2019\nRs/Liter\n7255\n7200\n720\n7255
6530\n72.80\nMOGAS -_E-10Gasoline\nRetall Retail Direct\n7255 7250-7250\n337
0st OSH\n7592 73:08 73.08\n7592 78.04 73.08\n281 2e1 281\n370
370\n150015741948\n9743 (95.29 95.29\n1656 16.20 1620\n11399 11149
11149\nsy\nDy Mage 2.19%\n"
num_only <- tesseract::tesseract(</pre>
  options = list(tessedit_char_whitelist = c(".0123456789 "))
)
no_grid %>%
  image_quantize(colorspace = 'gray') %>%
  image threshold() %>%
  image_crop(geometry_area(350, 0,350, 10)) %>%
  ocr(engine = num_only)
## [1] "01 2019\n72.00\n758 65.30\n7250\n 10\n758 72807250\n337 5 088\n7592
7308 73.04\n7592 73.08 73.08\n281 21 2\n330 370\n00 15741944\n9743 98.29
9529\n1656 16.20 6.20\n11399 1114911149\n250\n2.199\n"
raw_text <- no_grid %>%
  image_quantize(colorspace = "gray") %>%
  image_transparent("white", fuzz = 22) %>%
  image_background("white") %>%
```

```
image threshold() %>%
  image_crop(geometry_area(300, 0,400, 10)) %>%
  ocr(engine = num_only)
raw_tibble <- raw_text %>%
  str split(pattern = "\n") %>%
  unlist() %>%
  tibble(data = .)
raw_tibble$data=gsub(')','',as.character(raw_tibble$data))
raw_tibble$data=gsub('-','',as.character(raw_tibble$data))
col1=raw tibble[1:7,]
raw_tibble=raw_tibble[-(1:9),]
raw_tibble=raw_tibble[-(11:13),]
col1=col1[-(1:2),]
col1=sub('.*? ','',col1$data)
col1=data.frame(col1)
col1$col1
## [1] "65.30"
                                                "75872807250" "5 088"
                     "7250"
                                   "10"
check=separate(raw_tibble,data ,into = c('retail1','retail2','direct'),sep =
'\\ ',extra = 'merge')
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 6 rows [2,
3, 6,
## 7, 8, 9].
col1=col1 %>% add_row(col1=check$retail2)
col1
##
             col1
## 1
            65.30
             7250
## 2
## 3
               10
## 4 75872807250
## 5
            5 088
## 6
              201
## 7
              370
         15781944
## 8
## 9
            98.29
            16.20
## 10
      1114911149
## 11
## 12
             <NA>
## 13
             <NA>
## 14
             <NA>
#}
```

R Markdown