

Clean Tabular Data

AIM – To find out how Different open source OCR perform text detection.

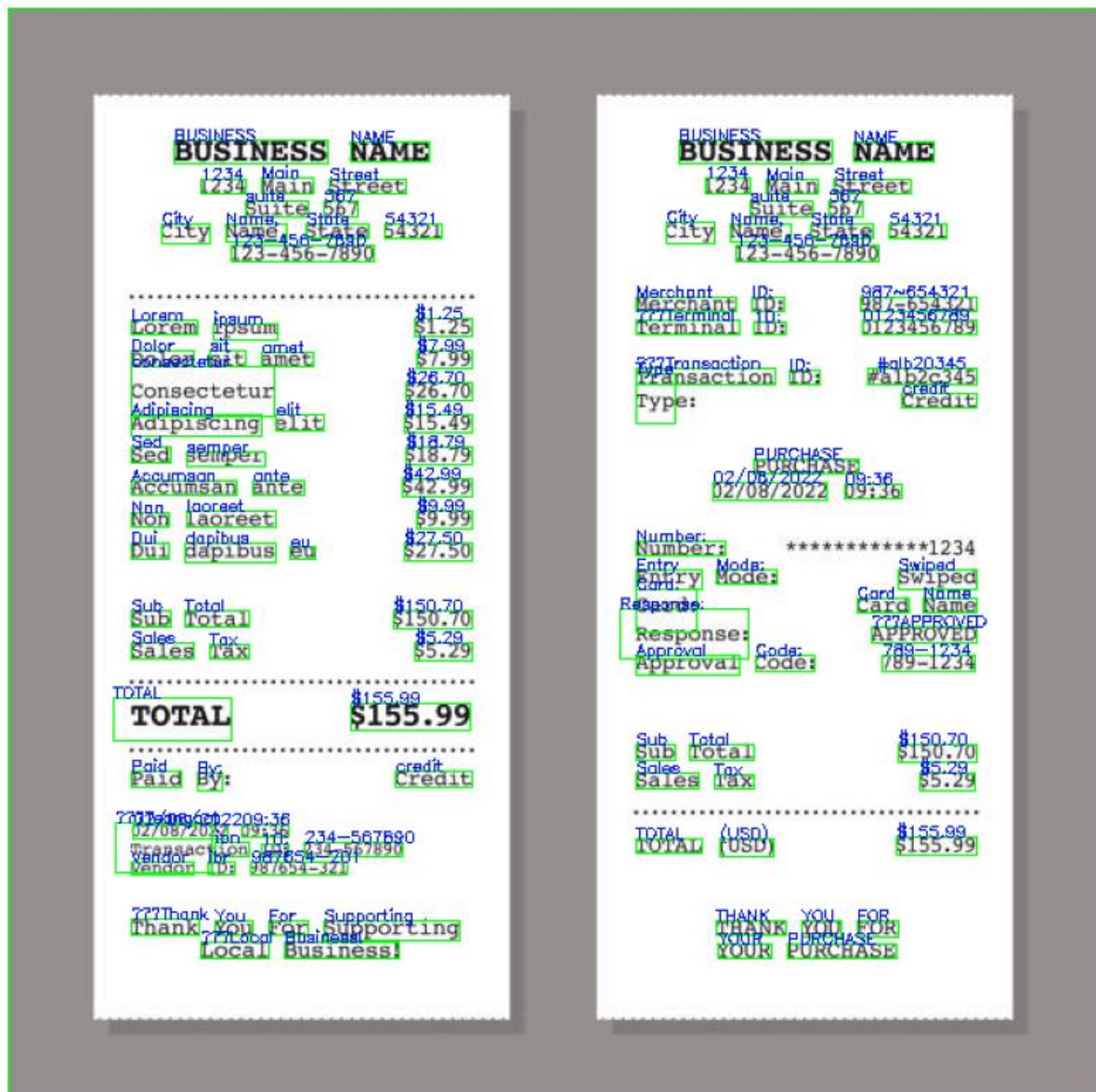
RAW IMAGE



TESEERACT OCR

Extraction –

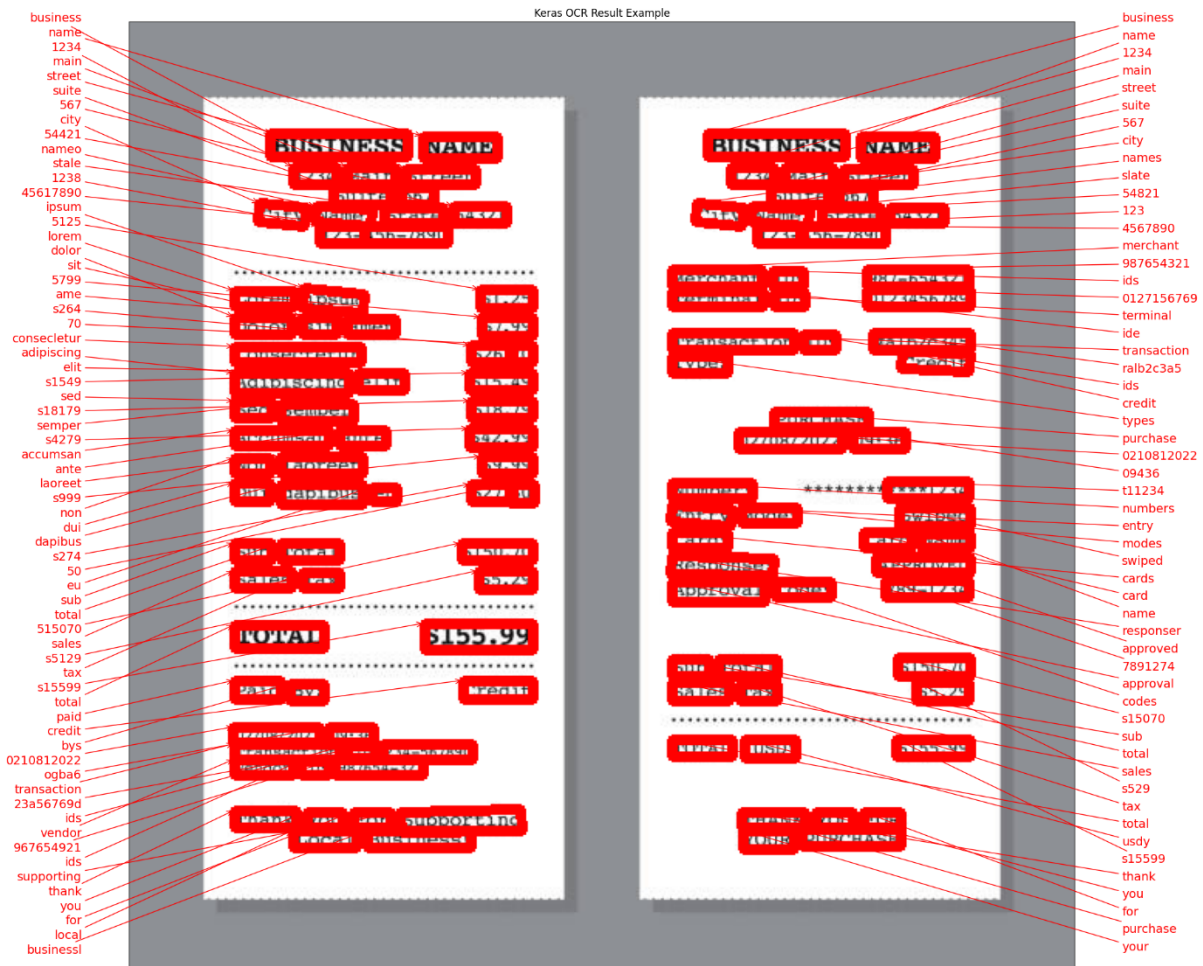
BUSINESS NAME 1234 Main Street suite 567 City Name, State 54321 123-456-7890 Lorem ipsum
\$1.25 Dolor sit amet \$7.99 consectetur \$26.70 Adipiscing elit \$15.49 Sed semper \$18.79 Accumsan
ante \$42.99 Non laoreet \$9.99 Dui dapibus eu \$27.50 Sub Total \$150.70 Sales Tax \$5.29 TOTAL
\$155.99 Paid By: credit 02/08/2022 09:36 'Peangact ion 10: 234-567090 Vendor 1Dr 987654-201
'Thank You For Supporting 'Local Business! BUSINESS NAME 1234 Main Street 'suite 567 City Name,
State 54321 123-456-7890 Merchant ID: 987~654321 'Terminal ID: 0123456789 'Transaction ID:
#alb2c345 Type! credit PURCHASE 02/08/2022 09:36 seeveeeeeei234 Swiped card Name
Response: APPROVED Approval Code: 789-1234 Sub Total \$150.70 Sales Tax \$5.29 TOTAL (USD)
\$155.99 THANK YOU FOR YOUR PURCHASE



KERAS OCR

Extraction –

business business name name 1234 main 1234 main street street suite 567 suite 567 city 54421 city
nameo stale names slate 54821 1238 45617890 123 4567890 merchant 987654321 ids ipsum 5125
0127156769 lorem terminal ide dolor sit 5799 ame transaction ralb2c3a5 ids s264 70 consecletur
credit types adipiscing elit s1549 sed s18179 semper purchase s4279 accumsan ante 0210812022
09436 laoreet s999 non t11234 dui numbers dapibus s274 50 eu entry modes swiped cards card
name sub total 515070 responser approved sales s5129 tax 7891274 approval codes s15599 total
s15070 sub total paid credit sales s529 bys tax 0210812022 ogba6 total usdy s15599 transaction
23a56769d ids vendor 967654921 ids thank supporting you for thank you for local businessl your
purchase



EASY OCR

Extraction-

BUSINESS NAME BUSINESS NAME 1234 Kain Street 1234 Kain Street Suite Suite 567 City Namc state 54321 City Namc state 54321 123-456-7890 123-456-7890 Merchant ID: 987-654321 Lorem ipsun 51.25 Terminal ID: 0123456789 Dolor sit amet 57.99 Transaction ID #alb2c345 Consectetur 526.70 Type: Credit Adipiscing elit 415.49 Sed semper 518.79 PURCHASE Accumsan ante 542.99 02/08/2022 09:36 Non laoreet \$9.99 Dui dapibus 527.50 Number: ***1*****+1234 Entry Mode: Swiped Card: Card Namc Sub Total 5150.70 Response: APPROVED Sales Tax 55.29 Approval Code: 789-1234 TOTAL \$155.99 Sub Total 5150.70 Paid By: Credit Sales Tax 55.29 02/08/2022 09:36 Transaction 234-567890 TOTAL (USD) \$155.99 vendor 987654-321 Thank You For Supporting THANK YOU FOR Local Business! YOUR PURCHASE

BUSINESS NAME 1234 Kain Street		BUSINESS NAME 1234 Kain Street	
BUSINESS NAME		BUSINESS NAME	
1234 Main Street		1234 Main Street	
Suite 567		Suite 567	
City Name, State 54321		City Name, State 54321	
123-456-7890		123-456-7890	
.....	Merchant ID: 987-654321	0123456789
.....	Terminal ID: 0123456789	0123456789
.....	Transaction ID #alb2c345	526.70
.....	Type: Credit	526.70
PURCHASE 02/08/2022 09:36		PURCHASE 02/08/2022 09:36	
Number: ***1*****+1234		Number: ***1*****+1234	
Entry Mode: Swiped		Entry Mode: Swiped	
Card: Card Name		Card: Card Name	
Response: APPROVED		Response: APPROVED	
Approval Code: 789-1234		Approval Code: 789-1234	
Sub Total Sales Tax 5150.70 55.29		Sub Total Sales Tax 5150.70 55.29	
Sub Total	\$150.70	Sub Total	\$150.70
Sales Tax	\$5.29	Sales Tax	\$5.29
TOTAL \$155.99		TOTAL (USD) \$155.99	
Paid By: Credit		Paid By: Credit	
02/08/2022 09:36 Transaction 234-567890		02/08/2022 09:36 Transaction 234-567890	
Vendor ID: 987654-321		Vendor ID: 987654-321	
Thank You For Supporting Local Business!		THANK YOU FOR YOUR PURCHASE	

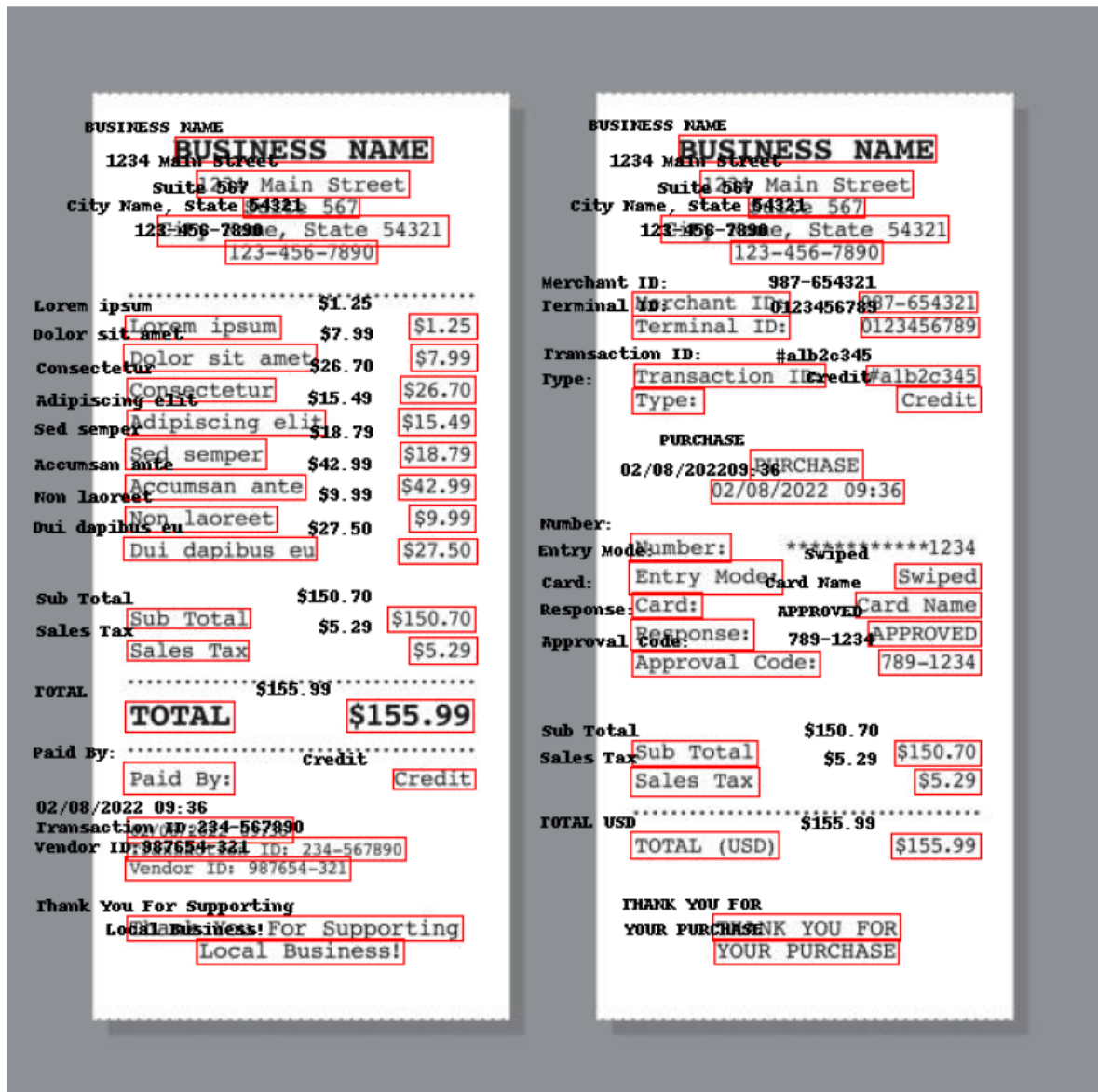
Paddle OCR

Extraction –

BUSINESS NAME BUSINESS NAME 1234 Main Street 1234 Main Street Suite 567 Suite 567 City Name, State 54321 City Name, State 54321 123-456-7890 123-456-7890 Merchant ID: 987-654321 Lorem ipsum \$1.25 Terminal ID: 0123456789 Dolor sit amet \$7.99 Consectetur \$26.70 Transaction ID: #alb2c345 Type: Credit Adipiscing elit \$15.49 Sed semper \$18.79 PURCHASE Accumsan ante \$42.99 02/08/202209:36 Non laoreet \$9.99 Dui dapibus eu \$27.50 Number: Entry Mode: Swiped Sub Total \$150.70 Card: Card Name Response: \$5.29 APPROVED Sales Tax Approval Code: 789-1234 TOTAL \$155.99 Sub Total \$150.70 Paid By: Credit Sales Tax \$5.29 02/08/2022 09:36 Transaction ID: 234-567890 TOTAL USD \$155.99 Vendor ID: 987654-321 Thank You For Supporting THANK YOU FOR Local Business! YOUR PURCHASE

BUSINESS NAME	
1234 Main Street	
Suite 567	
City Name	State 54321
City	123-456-7890
123-456-7890	
.....	\$1.25
.....	\$1.25
.....	\$7.99
.....	\$7.99
.....	\$26.70
.....	\$26.70
.....	\$15.49
.....	\$15.49
.....	\$18.79
.....	\$18.79
.....	\$42.99
.....	\$42.99
.....	\$9.99
.....	\$9.99
.....	\$27.50
.....	\$27.50
Sub Total	\$150.70
Sub Total	\$150.70
Sales Tax	\$5.29
Sales Tax	\$5.29
.....	\$155.99
TOTAL	\$155.99
TOTAL	\$155.99
Paid By:	Credit
Paid By:	Credit
02/08/2022 09:36	
Transaction ID: 234-567890	
Vendor ID: 987654-321	
Vendor ID: 987654-321	
Thank You For Supporting	
Thank You For Supporting	
Local Business!	
Local Business!	

BUSINESS NAME	
1234 Main Street	
Suite 567	
City Name	State 54321
City	123-456-7890
123-456-7890	
Merchant ID:	987-654321
Merchant ID:	987-654321
Terminal ID:	0123456789
Terminal ID:	0123456789
Transaction ID:	#alb2c345
Transaction ID:	#alb2c345
Type:	Credit
Type:	Credit
PURCHASE	
PURCHASE	
02/08/2022 09:36	
02/08/2022 09:36	
Number:	*****1234
Number:	*****1234
Entry Mode:	Swiped
Entry Mode:	Swiped
Card:	Card Name
Card:	Card Name
Response:	APPROVED
Response:	APPROVED
Approval Code:	789-1234
Approval Code:	789-1234
Sub Total	\$150.70
Sub Total	\$150.70
Sales Tax	\$5.29
Sales Tax	\$5.29
Sales Tax	\$5.29
TOTAL USD	\$155.99
TOTAL USD	\$155.99
TOTAL (USD)	\$155.99
TOTAL (USD)	\$155.99
THANK YOU FOR	
THANK YOU FOR	
YOUR PURCHASE	
YOUR PURCHASE	



DONUT OCR

Extraction –

```
<s_cord-v2><s_menu><s_nm> BUSINESS NAME</s_nm><s_price> NAME</s_price><sep/><s_nm> 1234 Main  
Street</s_nm><s_price> 1234 Main Street</s_nm><s_price> 1234</s_price><sep/><s_nm> Suite 567</s_nm><s_price>  
Suite 567</s_price><sep/><s_nm> City Name, State 54321</s_nm><s_price> 123-456-7890</s_price><sep/><s_nm> 123-  
454</s_nm><s_price> 123-21</s_price></s_menu><s_sub_total><s_subtotal_price> 987-  
654321</s_subtotal_price><s_discount_price> 0123456789</s_discount_price><s_tax_price> $7.99</s_tax_price><s_etc>  
#alb2345<sep/><s_nm> Adipiscing elit $15.49</s_nm><s_price> Credit</s_price><sep/><s_nm> sed  
semper</s_nm><s_price> $18.79</s_price><sep/><s_nm> Accumsan ante $42.99</s_nm><s_price>  
02/08/2022</s_price><sep/><s_nm> Non laoreet</s_nm><s_price> 09:36</s_price><sep/><s_nm>  
$9.99</s_nm><s_price> Dui dapibus eu</s_nm><s_price> $27.50</s_price></s_menu><s_sub_total><s_subtotal_price>
```

Number:</s_nm><s_price> Swiped</s_nm><s_price> Swided</s_price></s_menu><s_sub_total><s_subtotal_price>
\$150.70</s_subtotal_price><s_discount_price> 789-1234</s_discount_price><s_service_price>
\$5.29</s_service_price><s_tax_price> \$150.70</s_tax_price><s_etc>
85.29</s_etc></s_sub_total><s_total><s_total_price> \$155.99</s_total_price><s_total_etc> THANK YOU
FOR</s_total_etc><s_changeprice> Local Business!</s_nm><s_price> YOUR PURCHASE</s_changeprice></s_total></s>

After Cleaning

BUSINESS NAME NAME 1234 Main Street 1234 Main Street 1234 Suite 567 Suite 567 City
Name, State 54321 123-456-7890 123-454 123-21 987-654321 0123456789 \$7.99 #alb2345
Adipiscing elit \$15.49 Credit sed semper \$18.79 Accumsan ante \$42.99 02/08/2022 Non
laoreet 09:36 \$9.99 Dui dapibus eu \$27.50 Number: Swiped Swided \$150.70 789-1234 \$5.29
\$150.70 85.29 \$155.99 THANK YOU FOR Local Business! YOUR PURCHASE