Développement d'un programme pour l'Extraction de Données depuis un Tableau à partir d'Images et de PDF

Soutenu par :

Benguerine Wassim
LAMHATTAT Redouane

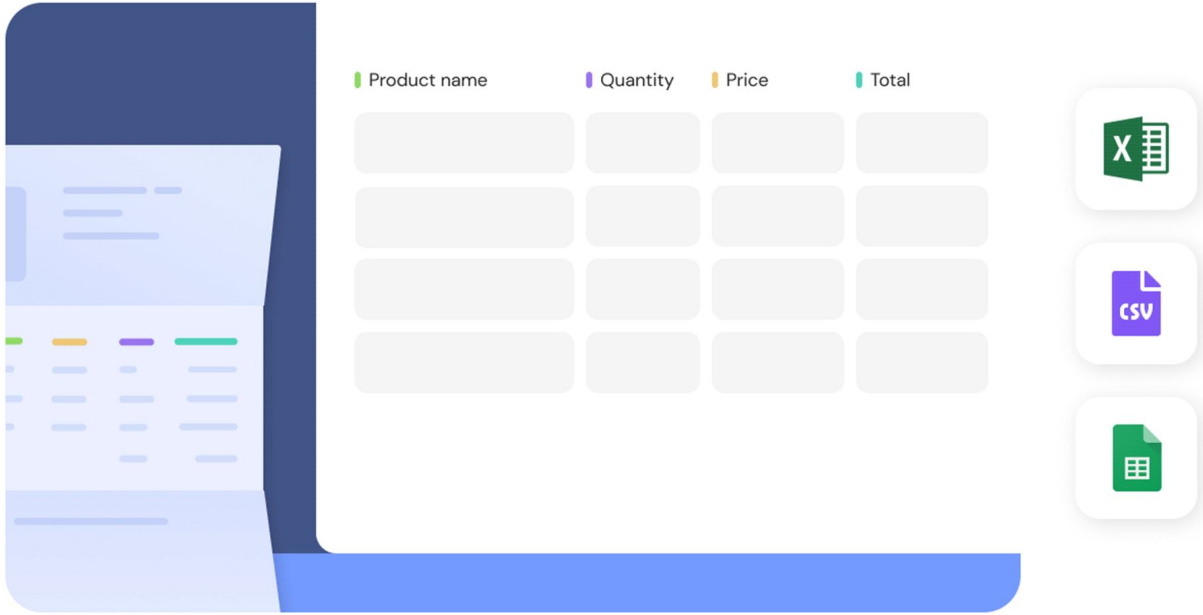
Sommaire

Qu'est-ce qu'une application d'extraction de données ?

Contexte général du Projet

- Présentation du projet
- Cadrage du projet

› Présentation du Projet



Objectifs du notre application

Réaliser une application qui permet de convertir un tableau présent dans une image en un format Excel ou CSV

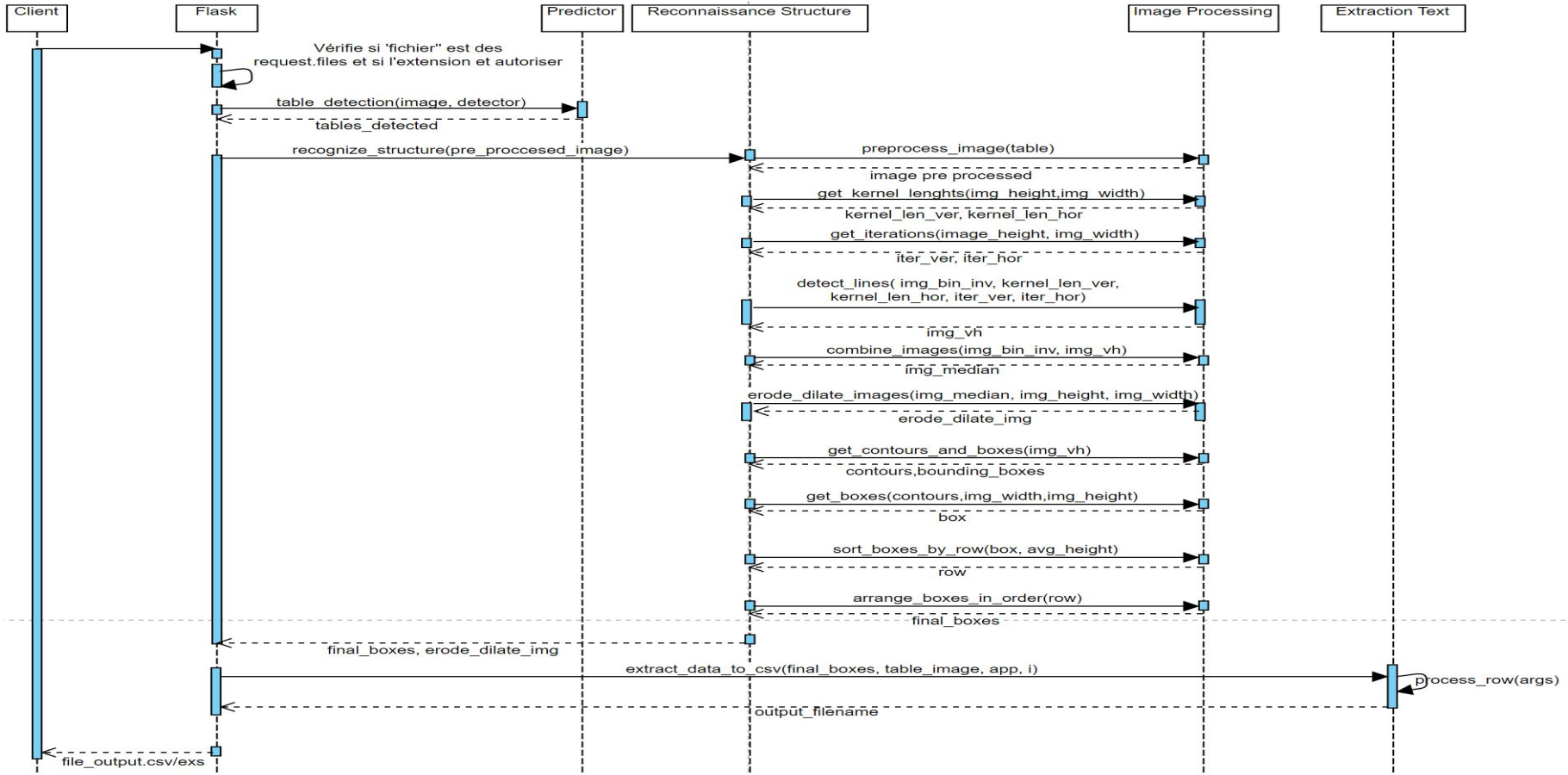
Analyse & Conception

- Étude fonctionnelle
- Étude technique

Analyse & Conception

› Étude fonctionnelle

Étude technique



- **Capture des besoins techniques**

python™

Detectron2

OpenCV

PaddleOCR

Flask

Algorithmes Utilisés

- Présentation des Algorithmes
- Fonctionnement des Algorithmes

Algorithmes Utilisés

› Présentation des algorithmes et leur fonctionnement

Algorithm: get_contours_and_boxes

```
Input: img_vh, a binary image
Output: Contours et boîtes de l'image
Procedure get_contours_and_boxes(img_vh)
    img_vh = add_border
    img_vh contours, _=findContours
    img_vh contours, bounding_boxes = sort_contourscontours
return contours, bounding_boxes
```

Algorithm: get_boxes procedure

```
Input: contours, image width (img_w), image height (img_h)
Output: Les boxes satisfaisant les conditions [éviter les boxes trop grandes]
Procedure get_boxes(contours, img_w, img_h)
    boxes = []
    for each contour c in contours do
        x, y, w, h = cv2.boundingRect(c)
        if w <0.9*img_w and h <0.9*img_h then
            boxes.append({'x': x, 'y': y, 'width': w, 'height': h})
        end
    end
return boxes
```

Algorithmes Utilisés

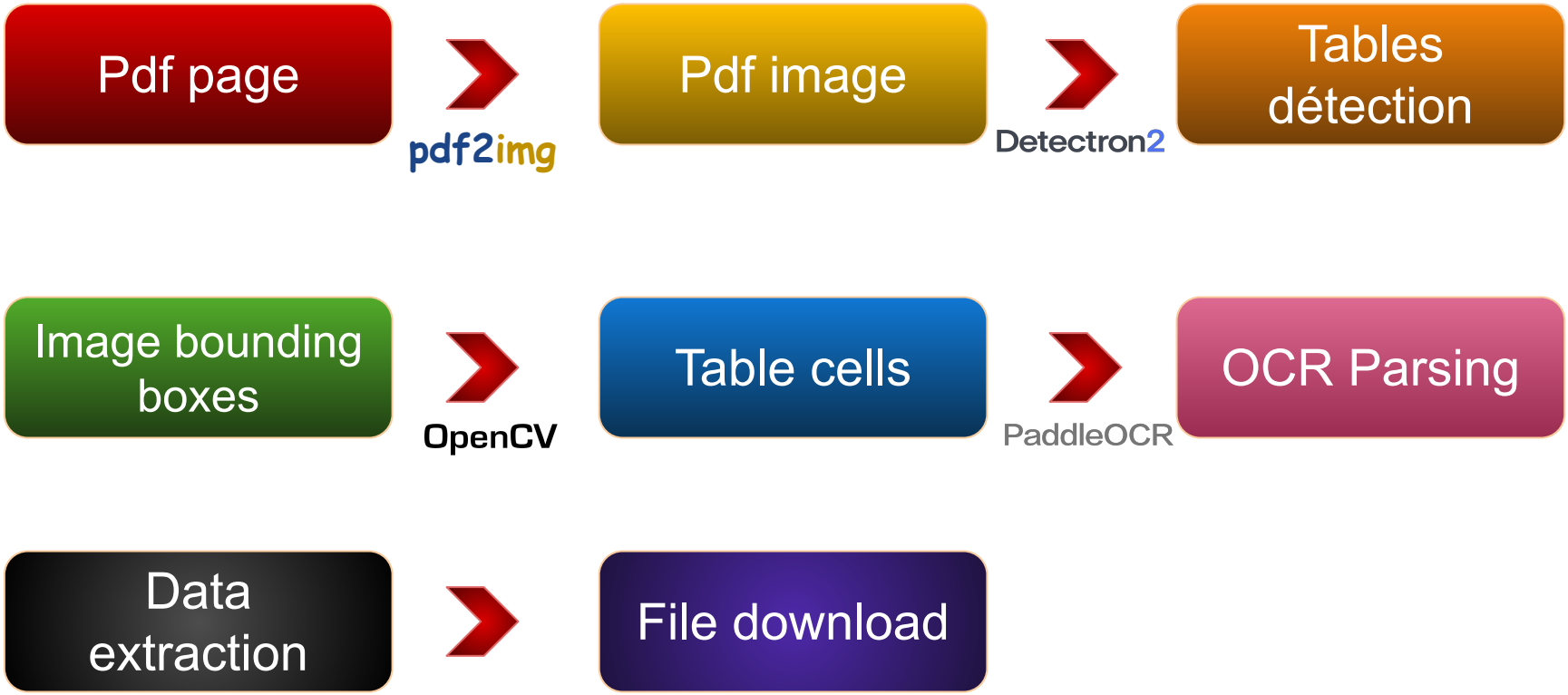
› Présentation des algorithmes et leur fonctionnement

```
Algorithm: arrange_boxes_in_order procedure
  Input: Row of boxes
  Output: Boxes organiser
  Procedure arrange_boxes_in_order(row)
    max_columns ← 0
    selected_index ← 0
    for i in range(0, length_of(row)) do
      if length_of(row[i]) >max_columns then
        max_columns ← length_of(row[i])
        selected_index ← i
      end
    end
    selected_row ← row[selected_index]
    centers ← empty_list
    for j in range(0, length_of(selected_row)) do
      center = int(selected_row[j]['x'] + selected_row[j]['width']/2)
      centers.append(center)
    end
    centers ← np.array(centers)
    centers.sort()
    organized_boxes ← empty_list
    for i in range(0, length_of(row)) do
      temp_list ← create_empty_list_with_length(max_columns)
      for j in range(0, length_of(row[i])) do
        distance ← calculate_distance_from_centers(centers, row[i][j])
        min_distance ← minimum_of(distance)
        index ← get_index_of(min_distance in distance)
        append row[i][j] to temp_list[index]
      end
      append temp_list to organized_boxes
    end
  return organized_boxes
```

```
Algorithm: extract_data_to_xls procedure
  Input: finalboxes, img, app, table_index
  Output: output_filename
  Procedure extract_data_to_xls(finalboxes, img, app, table_index)
    extracted_data ← empty_list
    With ProcessPoolExecutor() as executor do
    for row in finalboxes do
      futures ← [executor.submit(process_cells, cell, img) for cell in row if cell]
      row_data ← [f.result() for f in futures if f.result() is not None]
      append row_data to extracted_data
    end
    dataframe ← pd.DataFrame(extracted_data)
    styled_dataframe ← dataframe.style.set_properties(align="left")
    output_filename ← f"output_table_index.xlsx"
    output_filepath ← os.path.join(app.config['UPLOAD_FOLDER'], output_filename)
    styled_dataframe.to_excel(output_filepath)
  return output_filename
```

Réalisation

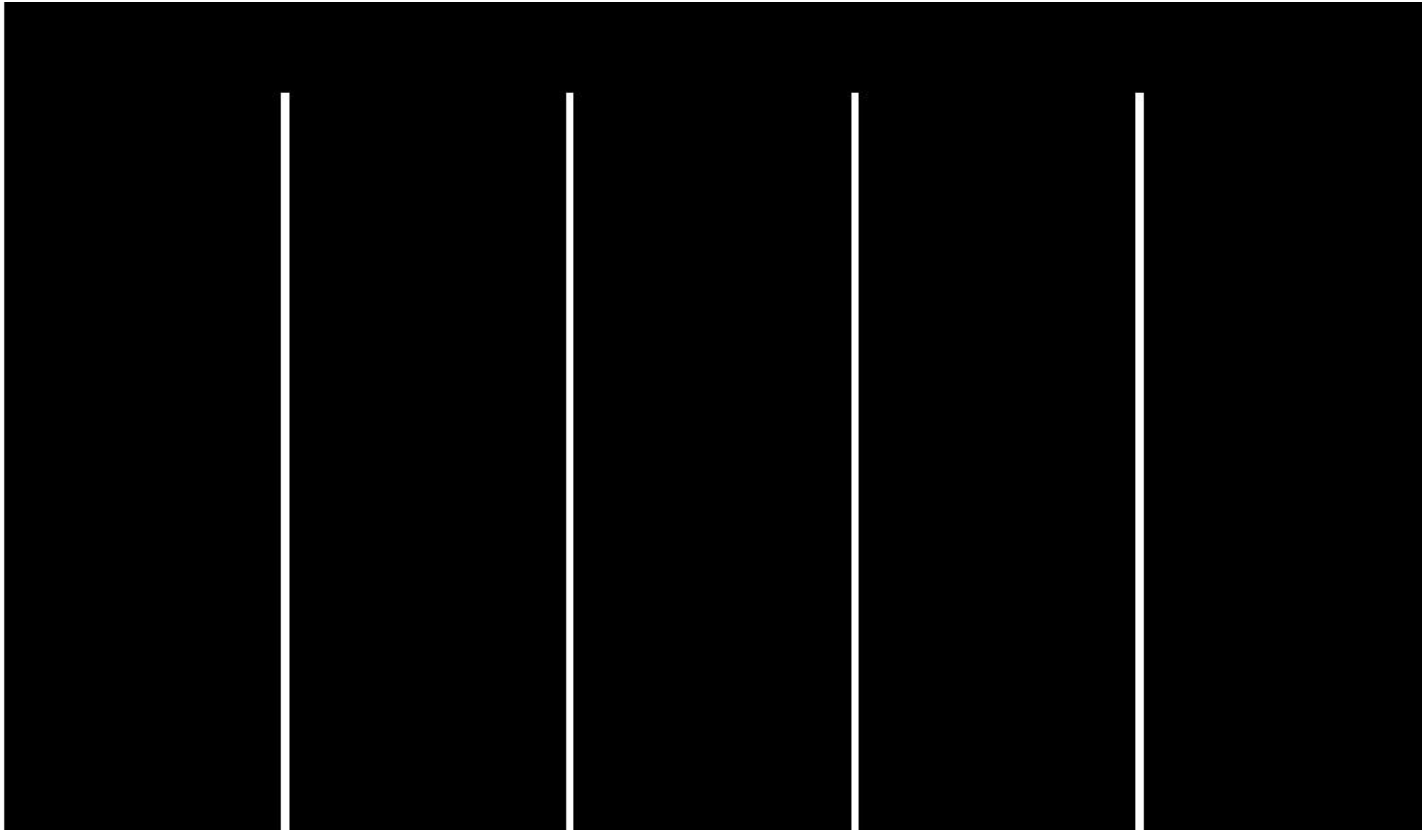
- Architecture utilisée
- Aperçu du projet
- Tests

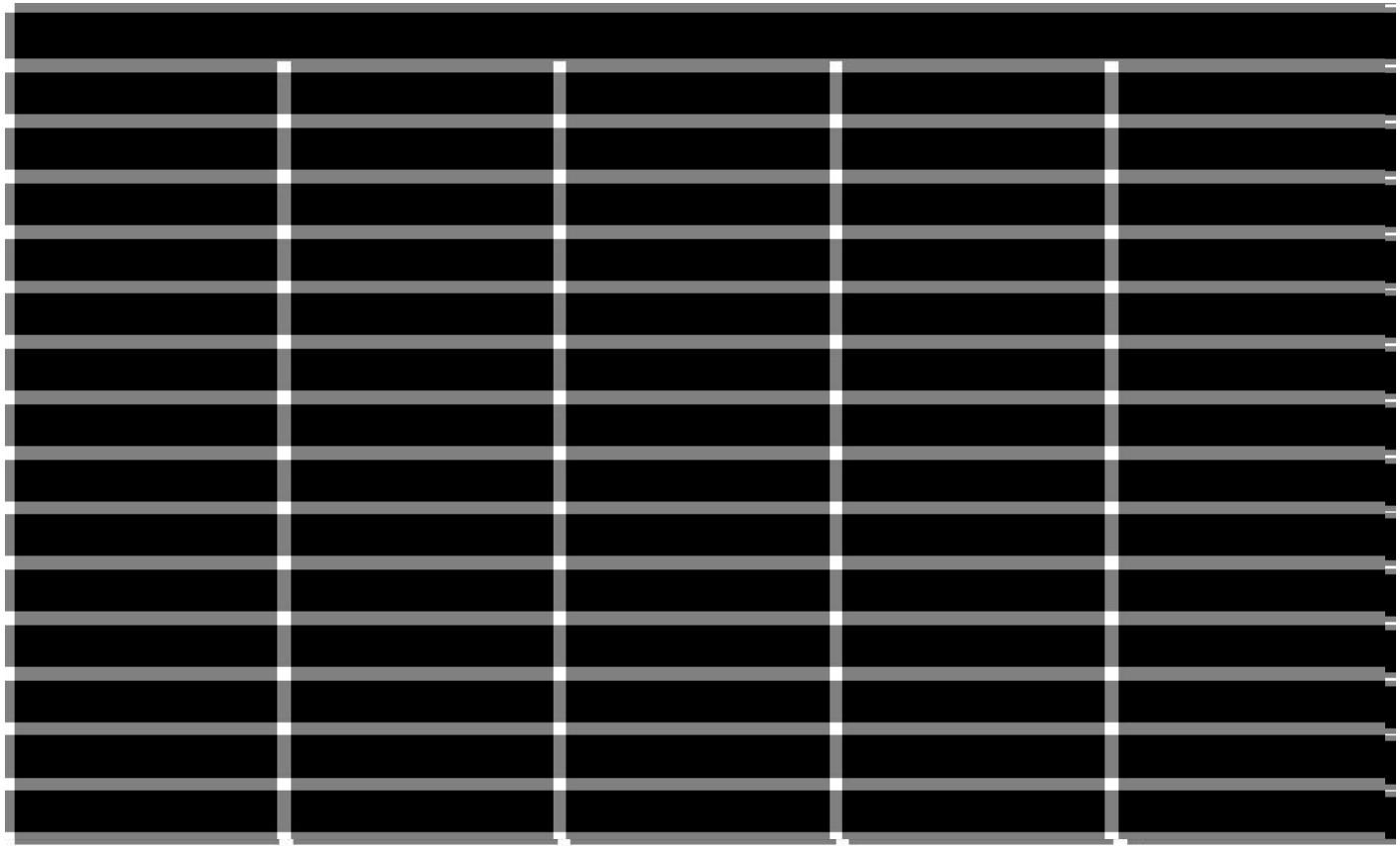

Pdf page
pdf2img
Pdf image
Detectron2
Tables détection
Image bounding boxes
OpenCV
Table cells
PaddleOCR
OCR Parsing
Data extraction
File download

Order Date	Order ID	Salesperson	Units	Order Amount
1/1/2015	10458	Bossk	13	$1,400.00
2/1/2015	10459	Dengar	14	$2,555.95
2/2/2015	10460	Dengar	7	$344.00
3/1/2015	10461	Bossk	8	$857.00
3/3/2015	10462	Greedo	9	$111.55
3/5/2015	10463	Fett	18	$823.00
4/1/2015	10464	Bossk	17	$24,455.50
4/28/2015	10465	Fett	13	$213.00
4/30/2015	10466	Fett	1	$10.00
5/1/2015	10467	Bossk	11	$789.70
5/2/2015	10468	Dengar	25	$21,286.60
5/3/2015	10469	Fett	8	$1,285.00
6/1/2015	10470	Bossk	11	$201.00
7/1/2015	10471	Bossk	13	$859.75

Order Date	Order ID	Salesperson	Units	Order Amount
1/1/2015	10458	Bossk	13	$1,400.00
2/1/2015	10459	Dengar	14	$2,555.95
2/2/2015	10460	Dengar	7	$344.00
3/1/2015	10461	Bossk	8	$857.00
3/3/2015	10462	Greedo	9	$111.55
3/5/2015	10463	Fett	18	$823.00
4/1/2015	10464	Bossk	17	$24,455.50
4/28/2015	10465	Fett	13	$213.00
4/30/2015	10466	Fett	1	$10.00
5/1/2015	10467	Bossk	11	$789.70
5/2/2015	10468	Dengar	25	$21,286.60
5/3/2015	10469	Fett	8	$1,285.00
6/1/2015	10470	Bossk	11	$201.00
7/1/2015	10471	Bossk	13	$859.75

[redacted]

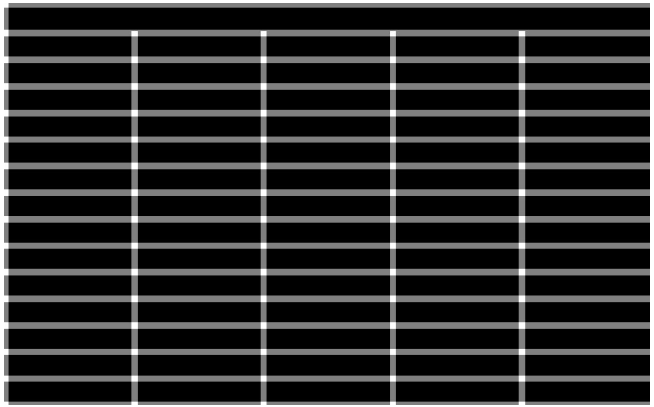




Order Date	Order ID	Salesperson	Units	Order Amount
1/1/2015	10458	Bossk	13	$1,400.00
2/1/2015	10459	Dengar	14	$2,555.95
2/2/2015	10460	Dengar	7	$344.00
3/1/2015	10461	Bossk	8	$857.00
3/3/2015	10462	Greedo	9	$111.55
3/5/2015	10463	Fett	18	$823.00
4/1/2015	10464	Bossk	17	$24,455.50
4/28/2015	10465	Fett	13	$213.00
4/30/2015	10466	Fett	1	$10.00
5/1/2015	10467	Bossk	11	$789.70
5/2/2015	10468	Dengar	25	$21,286.60
5/3/2015	10469	Fett	8	$1,285.00
6/1/2015	10470	Bossk	11	$201.00
7/1/2015	10471	Bossk	13	$859.75

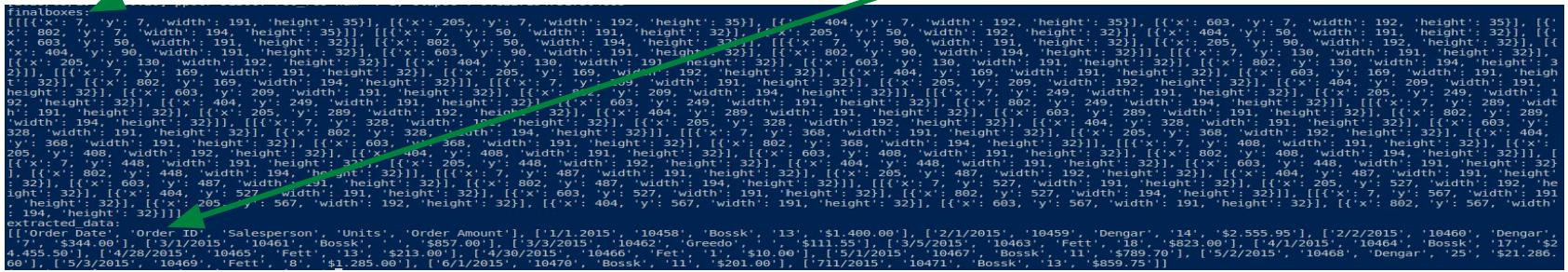
Réalisation

Order Date	Order ID	Salesperson	Units	Order Amount
1/1/2015	10458	Bossk	13	$1,400.00
2/1/2015	10459	Dengar	14	$2,555.95
2/2/2015	10460	Dengar	7	$344.00
3/1/2015	10461	Bossk	8	$857.00
3/3/2015	10462	Greedo	9	$111.55
3/5/2015	10463	Fett	18	$823.00
4/1/2015	10464	Bossk	17	$24,455.50
4/28/2015	10465	Fett	13	$213.00
4/30/2015	10466	Fett	1	$10.00
5/1/2015	10467	Bossk	11	$789.70
5/2/2015	10468	Dengar	25	$21,286.60
5/3/2015	10469	Fett	8	$1,285.00
6/1/2015	10470	Bossk	11	$201.00
7/1/2015	10471	Bossk	13	$859.75



Boxes detected

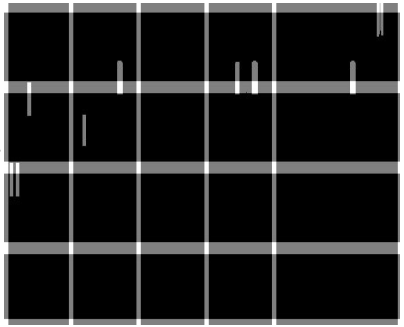
text extracted



```
finalboxes:
[[[{'x': 7, 'y': 7, 'width': 191, 'height': 35}], [{'x': 205, 'y': 7, 'width': 192, 'height': 35}], [{'x': 404, 'y': 7, 'width': 192, 'height': 35}], [{'x': 603, 'y': 7, 'width': 192, 'height': 35}], [{'
x': 802, 'y': 7, 'width': 194, 'height': 35}]], [[{'x': 7, 'y': 50, 'width': 191, 'height': 32}], [{'x': 205, 'y': 50, 'width': 192, 'height': 32}], [{'x': 404, 'y': 50, 'width': 191, 'height': 32}], [{'
x': 603, 'y': 50, 'width': 191, 'height': 32}], [{'x': 802, 'y': 50, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 90, 'width': 191, 'height': 32}], [{'x': 205, 'y': 90, 'width': 192, 'height': 32}], [{
'x': 404, 'y': 90, 'width': 191, 'height': 32}], [{'x': 603, 'y': 90, 'width': 191, 'height': 32}], [{'x': 802, 'y': 90, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 130, 'width': 191, 'height': 32}],
[{'x': 205, 'y': 130, 'width': 192, 'height': 32}], [{'x': 404, 'y': 130, 'width': 191, 'height': 32}], [{'x': 603, 'y': 130, 'width': 191, 'height': 32}], [{'x': 802, 'y': 130, 'width': 194, 'height': 3
2}]], [[{'x': 7, 'y': 169, 'width': 191, 'height': 32}], [{'x': 205, 'y': 169, 'width': 192, 'height': 32}], [{'x': 404, 'y': 169, 'width': 191, 'height': 32}], [{'x': 603, 'y': 169, 'width': 191, 'heigh
t': 32}], [{'x': 802, 'y': 169, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 209, 'width': 191, 'height': 32}], [{'x': 205, 'y': 209, 'width': 192, 'height': 32}], [{'x': 404, 'y': 209, 'width': 191, '
height': 32}], [{'x': 603, 'y': 209, 'width': 191, 'height': 32}], [{'x': 802, 'y': 209, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 249, 'width': 191, 'height': 32}], [{'x': 205, 'y': 249, 'width': 1
92, 'height': 32}], [{'x': 404, 'y': 249, 'width': 191, 'height': 32}], [{'x': 603, 'y': 249, 'width': 191, 'height': 32}], [{'x': 802, 'y': 249, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 289, 'widt
h': 191, 'height': 32}], [{'x': 205, 'y': 289, 'width': 192, 'height': 32}], [{'x': 404, 'y': 289, 'width': 191, 'height': 32}], [{'x': 603, 'y': 289, 'width': 191, 'height': 32}], [{'x': 802, 'y': 289,
'width': 194, 'height': 32}]], [[{'x': 7, 'y': 328, 'width': 191, 'height': 32}], [{'x': 205, 'y': 328, 'width': 192, 'height': 32}], [{'x': 404, 'y': 328, 'width': 191, 'height': 32}], [{'x': 603, 'y':
328, 'width': 191, 'height': 32}], [{'x': 802, 'y': 328, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 368, 'width': 191, 'height': 32}], [{'x': 205, 'y': 368, 'width': 192, 'height': 32}], [{'x': 404,
'y': 368, 'width': 191, 'height': 32}], [{'x': 603, 'y': 368, 'width': 191, 'height': 32}], [{'x': 802, 'y': 368, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 408, 'width': 191, 'height': 32}], [{'x':
205, 'y': 408, 'width': 192, 'height': 32}], [{'x': 404, 'y': 408, 'width': 191, 'height': 32}], [{'x': 603, 'y': 408, 'width': 191, 'height': 32}], [{'x': 802, 'y': 408, 'width': 194, 'height': 32}]], [
[{'x': 7, 'y': 448, 'width': 191, 'height': 32}], [{'x': 205, 'y': 448, 'width': 192, 'height': 32}], [{'x': 404, 'y': 448, 'width': 191, 'height': 32}], [{'x': 603, 'y': 448, 'width': 191, 'height': 32}
], [{'x': 802, 'y': 448, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 487, 'width': 191, 'height': 32}], [{'x': 205, 'y': 487, 'width': 192, 'height': 32}], [{'x': 404, 'y': 487, 'width': 191, 'height'
: 32}], [{'x': 603, 'y': 487, 'width': 191, 'height': 32}], [{'x': 802, 'y': 487, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 527, 'width': 191, 'height': 32}], [{'x': 205, 'y': 527, 'width': 192, 'he
ight': 32}], [{'x': 404, 'y': 527, 'width': 191, 'height': 32}], [{'x': 603, 'y': 527, 'width': 191, 'height': 32}], [{'x': 802, 'y': 527, 'width': 194, 'height': 32}]], [[{'x': 7, 'y': 567, 'width': 191
, 'height': 32}], [{'x': 205, 'y': 567, 'width': 192, 'height': 32}], [{'x': 404, 'y': 567, 'width': 191, 'height': 32}], [{'x': 603, 'y': 567, 'width': 191, 'height': 32}], [{'x': 802, 'y': 567, 'width'
: 194, 'height': 32}]]]
extracted_data:
[['Order Date', 'Order ID', 'Salesperson', 'Units', 'Order Amount'], ['1/1.2015', '10458', 'Bossk', '13', '$1.400.00'], ['2/1/2015', '10459', 'Dengar', '14', '$2.555.95'], ['2/2/2015', '10460', 'Dengar',
'7', '$344.00'], ['3/1/2015', '10461', 'Bossk', '8', '$857.00'], ['3/3/2015', '10462', 'Greedo', '9', '$111.55'], ['3/5/2015', '10463', 'Fett', '18', '$823.00'], ['4/1/2015', '10464', 'Bossk', '17', '$2
4.455.50'], ['4/28/2015', '10465', 'Fett', '13', '$213.00'], ['4/30/2015', '10466', 'Fett', '1', '$10.00'], ['5/1/2015', '10467', 'Bossk', '11', '$789.70'], ['5/2/2015', '10468', 'Dengar', '25', '$21.286.
60'], ['5/3/2015', '10469', 'Fett', '8', '$1.285.00'], ['6/1/2015', '10470', 'Bossk', '11', '$201.00'], ['711/2015', '10471', 'Bossk', '13', '$859.75']]
```

Réalisation

Size	Lethal to Operational Spacecraft	Number in Orbit	Trackable (i.e. can be cataloged)	Produces Lethal Fragments When Impacting An Operational Spacecraft
Small < 5mm	(Usually) Not	Millions	No	No
Medium 5mm – 10 cm	Usually	~ 500,000 in LEO	No	Maybe
Large > 10 cm	(Almost) Always	~ 21,000	Yes	Yes – 100s to 10,000s



Size	Lethato Operational Soacecraft	Number in Qrbit	Trackable (f.e. can be cataloged	Produces Lethal Fragm
Smal < 5mm	(Usually) Not	Miltions	No	No
Medim 5mm- 10 cm	Usually	1 500,000 in LEO	No	Maybe
Large > 10 cm	(Almost) Always	~ 21,000	Yes	Yes - 100s to 10,000s

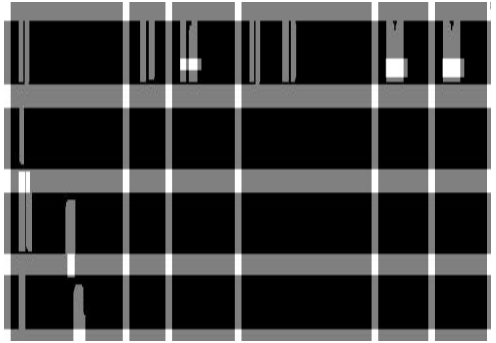
Boxes detected

text extracted

```
finalboxes:
[[[{'x': 10, 'y': 8, 'width': 140, 'height': 59}], [{'x': 158, 'y': 8, 'width': 146, 'height': 59}], [], [{'x': 312, 'y': 8, 'width': 147, 'height': 59}], [{'x': 467, 'y': 8, 'width': 146, 'he
ight': 59}], [{'x': 621, 'y': 8, 'width': 279, 'height': 59}]], [[{'x': 10, 'y': 75, 'width': 140, 'height': 59}], [{'x': 158, 'y': 75, 'width': 146, 'height': 59}, {'x': 180, 'y': 94, 'width'
: 8, 'height': 26}], [], [{'x': 312, 'y': 75, 'width': 147, 'height': 59}], [{'x': 467, 'y': 74, 'width': 146, 'height': 60}], [{'x': 621, 'y': 75, 'width': 279, 'height': 59}]], [[{'x': 10, '
y': 142, 'width': 140, 'height': 59}], [{'x': 158, 'y': 142, 'width': 146, 'height': 59}], [], [{'x': 312, 'y': 142, 'width': 147, 'height': 59}], [{'x': 467, 'y': 142, 'width': 146, 'height':
 59}], [{'x': 621, 'y': 142, 'width': 279, 'height': 59}]], [[{'x': 10, 'y': 209, 'width': 140, 'height': 56}], [{'x': 158, 'y': 209, 'width': 146, 'height': 56}], [], [{'x': 312, 'y': 209, 'w
idth': 147, 'height': 56}], [{'x': 467, 'y': 209, 'width': 146, 'height': 56}], [{'x': 621, 'y': 209, 'width': 279, 'height': 56}]]]
extracted_data:
[['Size', 'Lethato Operational Soacecraft', 'Number in Qrbit', 'Trackable (f.e. can be cataloged)', 'Produces Lethal Fragments When Impacting An Qperational Soacecraft'], ['Smal < 5mm', '(Usua
lly) Not', 'Miltions', 'No', 'No'], ['Medim 5mm- 10 cm', 'Usually', '1 500,000 in LEO', 'No', 'Maybe'], ['Large > 10 cm', '(Almost) Always', '~ 21,000', 'Yes', 'Yes - 100s to 10,000s']]
```

Réalisation

Statistic	N	Mean	St. Deviation	Min	Max
Complaints	30	64.63	12.73	40	85
Rating	30	65.84	13.35	37	90
Privileges	30	47.95	11.72	30	83



A	B	C	D	E	F	G
	0	1	2	3	4	5
0	tatistic	N	ean	t. Deviation	Min	M ax
1	Gomplaints	30	64.63	12.73	40	85
2	Rating	30	65.84	13.35	37	90
3	rivileges	30	47.95	11.72	30	83

Boxes detected

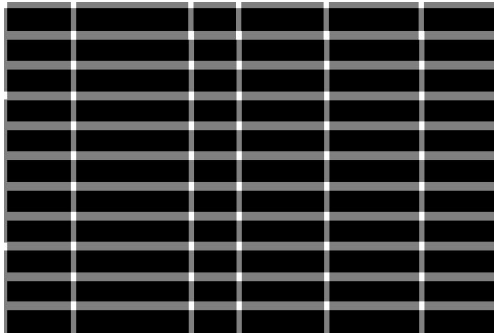
text extracted

```
finalboxes:
[[[{'x': 10, 'y': 8, 'width': 23, 'height': 29}], [{'x': 37, 'y': 8, 'width': 143, 'height': 29}], [{'x': 188, 'y': 8, 'width': 56, 'height': 29}], [{'x': 273, 'y': 8, 'width': 10, 'height
8}, {'x': 252, 'y': 8, 'width': 96, 'height': 29}], [], [{'x': 356, 'y': 8, 'width': 23, 'height': 29}], [{'x': 383, 'y': 8, 'width': 170, 'height': 29}], [{'x': 583, 'y': 8, 'width': 8, '
ht': 6}, {'x': 561, 'y': 8, 'width': 77, 'height': 29}], [], [{'x': 668, 'y': 8, 'width': 8, 'height': 6}, {'x': 646, 'y': 8, 'width': 80, 'height': 29}], []], [[{'x': 10, 'y': 45, 'width'
0, 'height': 28}], [], [{'x': 188, 'y': 45, 'width': 56, 'height': 28}], [{'x': 252, 'y': 45, 'width': 96, 'height': 28}], [], [{'x': 356, 'y': 45, 'width': 197, 'height': 28}], [], [{'x':
, 'y': 45, 'width': 77, 'height': 28}], [], [{'x': 646, 'y': 45, 'width': 80, 'height': 28}], []], [[{'x': 10, 'y': 81, 'width': 170, 'height': 28}], [], [{'x': 188, 'y': 81, 'width': 56,
ght': 28}], [{'x': 252, 'y': 81, 'width': 96, 'height': 28}], [], [{'x': 356, 'y': 81, 'width': 197, 'height': 28}], [], [{'x': 561, 'y': 81, 'width': 77, 'height': 28}], [], [{'x': 646, '
81, 'width': 80, 'height': 28}], []], [[{'x': 10, 'y': 116, 'width': 14, 'height': 25}], [{'x': 32, 'y': 116, 'width': 148, 'height': 25}], [{'x': 188, 'y': 116, 'width': 56, 'height': 25}
{'x': 252, 'y': 116, 'width': 96, 'height': 25}], [], [{'x': 356, 'y': 116, 'width': 197, 'height': 25}], [], [{'x': 561, 'y': 116, 'width': 77, 'height': 25}], [], [{'x': 646, 'y': 116, '
h': 80, 'height': 25}], []]]
extracted data:
[['tatistic', 'N', 'ean', 't. Deviation', 'Min', 'M ax'], ['Gomplaints', '30', '64.63', '12.73', '40', '85'], ['Rating', '30', '65.84', '13.35', '37', '90'], ['rivileges', '30', '47.95', '
2', '30', '83']]
```

Réalisation

Architecture utilisé | Aperçu du projet | **Tests**

PPGENDER	PPEDUC	Count	Mean.FWBScore	Median.FWBScore	SD.FWBScore
Male	less than high school	169	47.02367	48	13.32647
Male	high school degree/GED	703	53.85064	53	14.63737
Male	some college/associate	1025	54.65366	55	13.61152
Male	bachelor's degree	790	59.32405	59	12.93089
Male	graduate/professional degree	665	62.36842	61	13.11986
Female	less than high school	260	48.79615	49	12.68633
Female	high school degree/GED	919	53.96409	54	14.61242
Female	some college/associate	908	53.88326	53	13.97973
Female	bachelor's degree	522	58.37931	59	12.97368
Female	graduate/professional degree	433	61.05543	61	12.93061



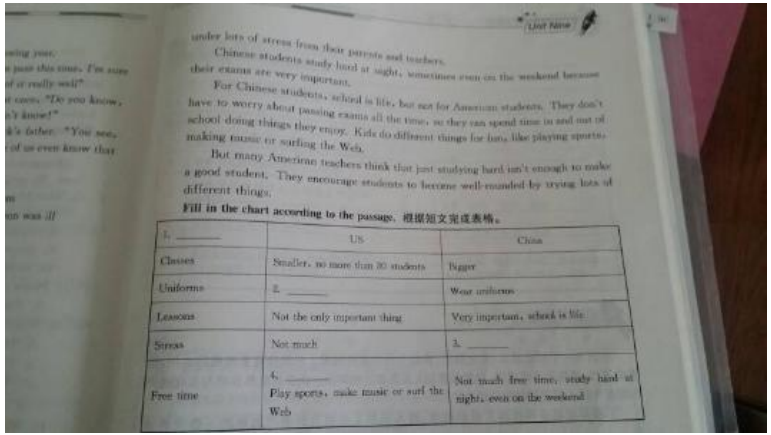
A	B	C	D	E	F	G
	0	1	2	3	4	5
0	PPGENDER	PPEDUC	Count	Mean.FWBScore	Median.FWBScore	SD.FwBScore
1	Male	less than high school	169	47.02367	48	13.32647
2	Male	high school degree/GED	703	53.85064	53	14.63737
3	Male	some e collegefassociate	1025	54.65366	55	13.61152
4	Male	bachelor's degree	790	59.32405	59	12.93089
5	Male	graduate/professional dep	665	62.36842	61	13.11986
6	Female	less than high school	260	48.79615	49	12.68633
7	Female	high school degree/GED	919	53.96409	54	14.61242
8	Female	some e college/associate	908	53.88326	53	13.97973
9	Female	bachelor's degree	522	58.37931	59	12.97368
10	Female	graduate/professional dep	433	61.05543	61	12.93061

text extracted

Boxes detected

```
idth': 133, 'height': 26}]], [[{'x': 6, 'y': 283, 'width': 122, 'height': 27}], [{'x': 136, 'y': 283, 'width': 214, 'height': 27}], [{'x': 358, 'y': 283, 'width': 82, 'height': 27}], [{'x': 448, 'y':
, 'width': 157, 'height': 27}], [{'x': 613, 'y': 283, 'width': 171, 'height': 27}], [{'x': 792, 'y': 283, 'width': 133, 'height': 27}]], [[{'x': 6, 'y': 318, 'width': 122, 'height': 25}], [{'x': 136,
: 318, 'width': 214, 'height': 25}], [{'x': 358, 'y': 318, 'width': 82, 'height': 25}], [{'x': 448, 'y': 318, 'width': 157, 'height': 25}], [{'x': 613, 'y': 318, 'width': 171, 'height': 25}], [{'x': 7
 'y': 318, 'width': 133, 'height': 25}]], [[{'x': 6, 'y': 351, 'width': 122, 'height': 28}], [{'x': 136, 'y': 351, 'width': 214, 'height': 28}], [{'x': 358, 'y': 351, 'width': 82, 'height': 28}], [{'x
448, 'y': 351, 'width': 157, 'height': 28}], [{'x': 613, 'y': 351, 'width': 171, 'height': 28}], [{'x': 792, 'y': 351, 'width': 133, 'height': 28}]]]
extracted_data:
[['PPGENDER', 'PPEDUC', 'Count', 'Mean.FWBScore', 'Median.FWBScore', 'SD.FwBScore'], ['Male', 'less than high school', '169', '47.02367', '48', '13.32647'], ['Male', 'high school degree/GED', '703', '
85064', '53', '14.63737'], ['Male', 'some e collegefassociate', '1025', '54.65366', '55', '13.61152'], ['Male', "bachelor's degree", '790', '59.32405', '59', '12.93089'], ['Male', 'graduate/profession
degree', '665', '62.36842', '61', '13.11986'], ['Female', 'less than high school', '260', '48.79615', '49', '12.68633'], ['Female', 'high school degree/GED', '919', '53.96409', '54', '14.61242'], ['Fe
e', 'some e college/associate', '908', '53.88326', '53', '13.97973'], ['Female', "bachelor's degree", '522', '58.37931', '59', '12.97368'], ['Female', 'graduate/professional degree', '433', '61.05543'
61', '12.93061']]
```

Realisation



under lots of stress from their parents and teachers.

Chinese students study hard at night, sometimes even on the weekend because their exams are very important.

For Chinese students, school is life, but not for American students. They don't have to worry about passing exams all the time, so they can spend time in and out of school doing things they enjoy. Kids do different things for fun, like playing sports, making music or surfing the Web.

But many American teachers think that just studying hard isn't enough to make a good student. They encourage students to become well-rounded by trying lots of different things.

Fill in the chart according to the passage. 根据短文完成表格。

1. _____	US	China
Classes	Smaller, no more than 30 students	Bigger
Uniforms	2. _____	Wear uniforms
Lessons	Not the only important thing	Very important, school is life
Stress	Not much	3. _____
Free time	4. _____ Play sports, make music or surf the Web	Not much free time, study hard at night, even on the weekend

Boxes detected

No text extracted

```
finalboxes:
[[[{'x': 0, 'y': 0, 'width': 10, 'height': 97}], [{'x': 17, 'y': 0, 'width': 70, 'height': 15}, {'x': 15, 'y': 22, 'width': 71, 'height': 12}], [], [{'x': 94, 'y': 5, 'width': 114, 'height': 14}, {'x': 9
4, 'y': 23, 'width': 115, 'height': 14}], [], [{'x': 216, 'y': 8, 'width': 118, 'height': 11}, {'x': 217, 'y': 24, 'width': 118, 'height': 13}], [], [{'x': 341, 'y': 0, 'width': 9, 'height': 52}]], [[],
[{'x': 14, 'y': 41, 'width': 72, 'height': 12}], [], [{'x': 93, 'y': 41, 'width': 40, 'height': 13}], [], [{'x': 218, 'y': 42, 'width': 120, 'height': 12}], [], []], [[], [{'x': 12, 'y': 60, 'width': 73,
 'height': 13}], [], [{'x': 92, 'y': 60, 'width': 119, 'height': 13}], [], [{'x': 218, 'y': 59, 'width': 123, 'height': 13}], [], []], [[], [{'x': 11, 'y': 80, 'width': 74, 'height': 13}], [], [{'x': 92,
 'y': 80, 'width': 120, 'height': 12}], [], [{'x': 219, 'y': 78, 'width': 124, 'height': 13}], [], []], [[{'x': 6, 'y': 100, 'width': 78, 'height': 45}], [], [], [{'x': 90, 'y': 99, 'width': 125, 'height
': 46}], [], [{'x': 220, 'y': 96, 'width': 130, 'height': 42}], [], []]]
extracted_data:
[[' ', ' ', ' ', ' ', ' '], [' ', ' ', ' '], [' ', ' ', ' '], [' ', ' ', ' '], [' ', ' ', '2']]
```

Conclusion & Perspective

- Conclusion
- Perspective

- Ce que nous avons appris de ce projet

Pour la version 2.0 :

- L'étude des tableaux plus complexes
- Réduire le temps de traitement

Merci pour votre attention