

DPSA_Data_Mining_Project

August 2, 2025

0.1 Introduction Summary

The page outlines a Python-based data mining project designed to extract job post information from PDF circulars published by the Department of Public Service and Administration (DPSA) in South Africa.

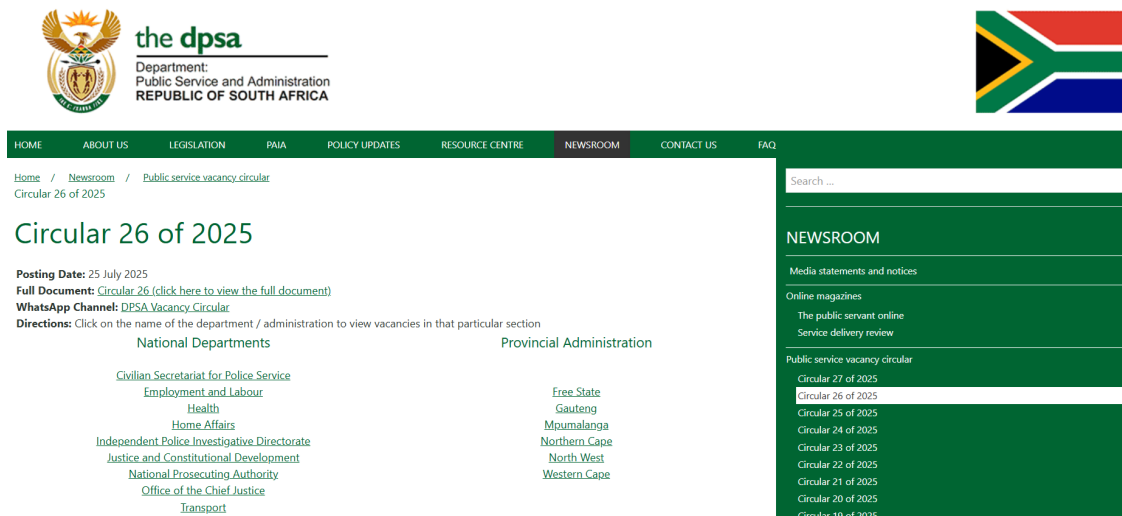
The project aims to automate the extraction of job vacancies from these circulars, which are typically published more than once a month. The extracted data includes details such as the post title, salary level, department, and closing date for applications.

0.2 Problem Statement

The **Department of Public Service and Administration (DPSA)** publishes most of its job vacancies on www.dpsa.gov.za in the form of a lengthy and unstructured PDF document on a monthly basis. This format is not user-friendly, especially considering the high youth unemployment rate in South Africa.


```
[4]: from PIL import Image
Image.open(r"C:\data_repo\DPSA Circular data mining\DPSA_Landing_Page.png")
```

[4]:



```
[5]: Image.open(r"C:\data_repo\DPSA Circular data mining\pdf_Vcancies.png")
```

[5]:

 <div>the dpsa Department of Public Service and Administration REPUBLIC OF SOUTH AFRICA</div>			
PUBLIC SERVICE VACANCY CIRCULAR			
PUBLICATION NO 26 OF 2025 DATE ISSUED 25 JULY 2025			
1. Introduction		POST 26/02	DIRECTOR: COMPLIANCE REF NO: CSP/05/2025
1.1 This Circular is, except during December, published on a weekly basis and contains the advertisements of vacant posts and jobs in Public Service departments.		SALARY	R1 266 714 per annum, (all-inclusive package)
1.2 Although the Circular is issued by the Department of Public Service and Administration, the Department is not responsible for the content of the advertisements. Enquiries about an advertisement must be addressed to the relevant advertising department.		CENTRE	Pretoria
2. Directions to candidates		REQUIREMENTS	Bachelor Degree in Public Administration/ Social Sciences or Development Studies (NQF 7) or relevant equivalent qualification. 5 years of experience at a middle/ senior managerial level. Understanding of monitoring and evaluation strategies, policies and procedures. Extensive knowledge and experience in monitoring and evaluation tools and systems. Sound knowledge of and interpretation of constitutional law. Knowledge of Civilian Secretariat for Police Act, SAPS Act, Public Service Act, Public Finance Management Act, SMS Handbook and community structures. Strategic capability and leadership, programme and project management, Change management, knowledge management and computer literacy. Problem solving and analysis. Verbal and written communication skills. Project management, presentation, planning and organising skills. A valid driver's license.
2.1 Applications on form Z83 with full particulars of the applicants' training, qualifications, competencies, knowledge and experience (on a separate sheet if necessary or a CV) must be forwarded to the department in which the vacancy/vacancies exist(s). NB. PROSPECTIVE APPLICANTS MUST PLEASE USE THE NEW Z83 WHICH IS EFFECTIVE AS AT 01 JANUARY 2021.		DUTIES	Manage the development of strategy tools and guidelines relevant to compliance monitoring. Manage the development and implementation of compliance monitoring plan. Review existing information needs for compliance monitoring. Ensure the development of information management system. Provide guidance on monitoring SAPS compliance with legislation, policing policies and Ministerial directives. Provide direction in the development of improvement plan to address non-compliance with legislation and policies. Manage the police compliance through oversight monitoring and evaluation (OME). Ensure development and management of strategies to monitor SAPS's compliance to legislation and policies. Provide guidance on the monitoring of implementation of policing legislation and policies to improve service delivery. Provide strategic guidance on compliance of the police in relation to policies and legislation relevant to GBV and other Ministers priorities. Conduct oversight on the police focusing on policing policies and legislation. Manage and enhance police compliance through oversight monitoring and evaluation. Manage the development of strategy tools and guidelines relevant to SAPS compliance monitoring. Manage consolidation of complex report on SAPS implementation and compliance with the Domestic Violence Act (DVA) to Parliament. Establish and manage Forums with SAPS to improve compliance with the DVA. Provide direction in the development of improvement plan to address non-compliance with the DVA. Manage and assess compliance level to policing policies by SAPS. Establish and manage relationship with relevant external stakeholders such as Civil Society, SAPS, Justice and other government departments. Establish and manage relationships with relevant internal stakeholders. Provide appropriate and timely reports with recommendations. Coordinate and manage the implementation of special projects. Management of resources (Financial and Physical).
2.2 Applicants must indicate the reference number of the vacancy in their applications.		ENQUIRIES APPLICATIONS	Me NM Sefili Tel No: (012) 4931 388 Can also be emailed to Recruitment42@csp.gov.za
2.3 Applicants requiring additional information regarding an advertised post must direct their enquiries to the department where the vacancy exists. The Department of Public Service and Administration must not be approached for such information.			OTHER POSTS
2.4 It must be ensured that applications reach the relevant advertising departments on or before the applicable closing dates.		POST 26/04	ASSISTANT DIRECTOR: EVALUATIONS REF: CSP/06/2025
3. Directions to departments			
3.1 The contents of this Circular must be brought to the attention of all employees.			
3.2 It must be ensured that employees declared in excess are informed of the advertised vacancies. Potential candidates from the excess group must be assisted in applying timely for vacancies and attending where applicable, interviews.			
3.3 Where vacancies have been identified to promote representativeness, the provisions of sections 15 (affirmative action measures) and 20 (employment equity plan) of the Employment Equity Act, 1998 should be applied. Advertisements for such vacancies should state that it is intended to promote representativeness through the filling of the vacancy and that the candidature of persons whose transfer/accommodation will promote			



When scrolling through the job listings, it would be more effective and accessible if the vacancies were presented in a structured table format, similar to other job portals. The current format — a long, text-heavy document — makes it difficult for users to quickly find relevant opportunities.

1 Methodology & Results

This python script works with all pdfs published on [Department of public service](#) to convert it into a structured table format. The script uses the PyMuPDF library to read the PDF, extract text, and identify job postings based on specific patterns in the text. It then organizes this information into a pandas DataFrame for easy manipulation and analysis.

- **PDF Handling:** The script uses the PyMuPDF library (`fitz`) to open and read text from a DPSA PDF job circular stored on Google Drive.
- **Text Extraction & Splitting:** It processes the entire PDF, splits it by job post identifiers, and extracts structured information using regular expressions.
- **Captured Fields:**
 - Post title
 - Centre (location)
 - Salary
 - Requirements
 - Duties
 - Enquiries contact
 - Closing date (extracted from the full document or per post)
- **Output:** The collected data for each post is stored in a pandas DataFrame and can be saved or analyzed further.
- **Result:** The script reports a total of **313 job positions** extracted from the circular.

[6]: `pip install PyMuPDF`

Requirement already satisfied: PyMuPDF in c:\users\conrad
mogane\appdata\local\programs\python\python313\lib\site-packages (1.26.3)
Note: you may need to restart the kernel to use updated packages.

```
[7]: #Import necessary libraries
import os
import fitz # PyMuPDF
import re
import pandas as pd

# Define the function to extract job postings from DPSA PDF circulars
# This function reads a PDF file, extracts job postings, and returns them in a structured format

def extract_dpsa_jobs(pdf_path):
    """
    Extract job postings from DPSA PDF circulars.

    Args:
        pdf_path (str): Path to the PDF file

    Returns:
        pandas.DataFrame: DataFrame containing extracted job information
    """
    try:
        # Open the PDF
        doc = fitz.open(pdf_path)

        # Extract all text
        text = ""
        for page in doc:
            text += page.get_text()

        # Close the document
        doc.close()

        # Split text by each post
        posts = re.split(r"\nPOST\s+\d+/\d+\s*:\s*", text)[1:] # skip header

        data = []
        for post_text in posts:
            try:
                # Extract fields using regex patterns
                post_match = re.search(r"^(.*?)\n", post_text)
                centre_match = re.search(r"CENTRE\s*:\s*(.*)", post_text)
                salary_match = re.search(r"SALARY\s*:\s*(.*)", post_text)
```

```

        requirements_match = re.search(r"REQUIREMENTS\s*:\s*(.*?) (?:"
↪DUTIES\s*:)", post_text, re.DOTALL)
        duties_match = re.search(r"DUTIES\s*:\s*(.*?) (?:"
↪)", post_text, re.DOTALL)
        enquiries_match = re.search(r"ENQUIRIES\s*:\s*(.*)", post_text)
        closing_match = re.search(r"CLOSING DATE\s*:\s*(.*)", text)

        post = post_match.group(1).strip() if post_match else ""
        centre = centre_match.group(1).strip() if centre_match else ""
        salary = salary_match.group(1).strip() if salary_match else ""
        requirements = requirements_match.group(1).strip().
↪replace("\n", " ") if requirements_match else ""
        duties = duties_match.group(1).strip().replace("\n", " ") if
↪duties_match else ""
        enquiries = enquiries_match.group(1).strip() if enquiries_match
↪else ""
        closing_date = closing_match.group(1).strip() if closing_match
↪else ""

        data.append({
            "Post": post,
            "Centre": centre,
            "Salary": salary,
            "Requirements": requirements,
            "Duties": duties,
            "Enquiries": enquiries,
            "Closing Date": closing_date
        })
    except Exception as e:
        print(f"Error processing a post: {e}")
        continue

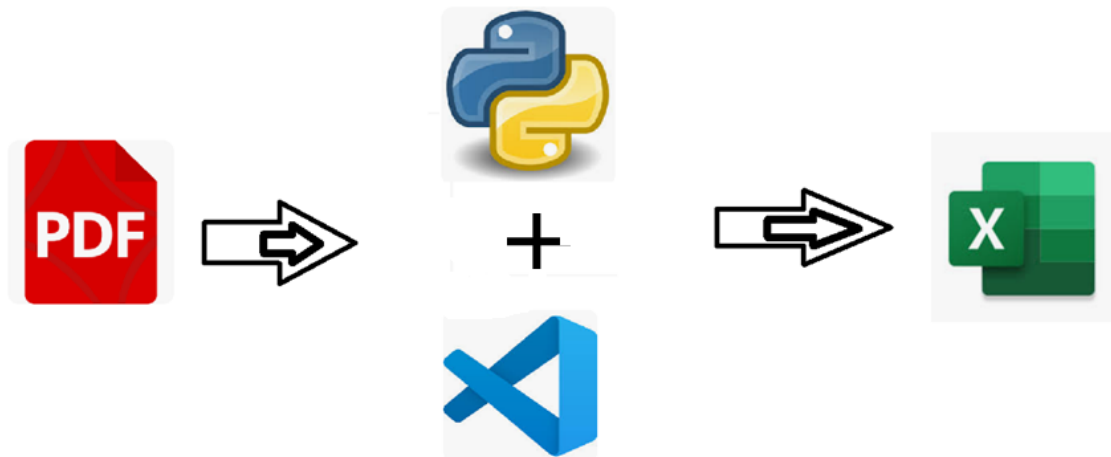
    return pd.DataFrame(data)

except fitz.FileNotFoundError:
    print(f"Error: PDF file not found at {pdf_path}")
    return pd.DataFrame()
except Exception as e:
    print(f"Error processing PDF: {e}")
    return pd.DataFrame()

# Display an image (if needed)
Image.open(r"C:\data_repo\DPSA Circular data mining\WORKFLOW.bmp")

```

[7]:



```
[10]: # Example usage:
df = extract_dpsa_jobs(r"C:\data_repo\DPSA Circular data_
↳mining\Source_Data\PSV_CIRCULAR_25_OF_2025.pdf")
print(f"Total unique entries: {len(df)}")
```

Total unique entries: 154

```
[14]: df.to_excel(r"C:\data_repo\DPSA Circular data_
↳mining\Output\PSV_CIRCULAR_25_OF_2025.xlsx", index=False)
print("Data saved successfully!")
df # Display the first few rows of the DataFrame
```

Data saved successfully!

```
[14]:
```

	Post \
0	DEPUTY DIRECTOR: LABOUR CENTRE OPERATIONS REF NO:
1	ASSISTANT DIRECTOR: SUPPLY CHAIN MANAGEMENT RE...
2	ASSISTANT DIRECTOR: COMMUNICATIONS REF NO: HR4...
3	ASSISTANT DIRECTOR: LABOUR MARKET INFORMATION ...
4	ASSISTANT
..	...
149	STERILISATION PRODUCTION OPERATOR (5/8 POSTS)
150	SESSIONAL MEDICAL SPECIALIST GRADE 1 TO 3 (URO...
151	MEDICAL OFFICER GRADE 1 TO 3 (35 SESSIONS) NHI
152	MEDICAL OFFICER SESSIONAL (14 HOURS PER WEEK) ...
153	MEDICAL OFFICER: GRADE 1 TO 3 (35 SESSIONS - NHI)

	Centre \
0	Labour Centre: Knysna (Western Cape)
1	Provincial Office: Kimberley
2	Provincial Office: Western Cape

3 Provincial Office: Western Cape
 4 Provincial Office: Western Cape
 ..
 149 Caledon Hospital
 150 New Somerset Hospital
 151 Vredenburg Hospital
 152 Northern/Tygerberg Substructure
 153 Citrusdal Clinic, Cederberg Sub-district

Salary \

0 R1 059 105 per annum, (all inclusive)
 1 R468 459 per annum
 2 R468 459 per annum
 3 R468 459 per annum
 4 R468 459 per annum
 ..
 149 R102 300 per annum
 150 Grade 1: R612 per hour
 151 Grade 1: R482 per hour
 152 Grade 1: R482 per hour
 153 Grade 1: R482 per hour

Requirements \

0 Three (3) years National Diploma (NQF6)/ Under...
 1 Three (3) years' National Diploma (NQF 6)/ Und...
 2 Three (3) year National Diploma (NQF6)/ Underg...
 3 Three (3) year National Diploma (NQF6)/ Underg...
 4 Three (3) year qualifications National Diploma...
 ..
 149 Minimum educational qualification: General Edu...
 150 Minimum educational qualification: Appropriate...
 151 Minimum educational qualification: Appropriate...
 152 Minimum educational qualification: Appropriate...
 153 Minimum educational qualification: Appropriate...

Duties \

0 Manage the service delivery objectives as per ...
 1 Manage demand of goods and services in the Pro...
 2 Provide public relations and media liaison ser...
 3 Manage performance information of the province...
 4 Implement Fraud detections and Anti-Corruption...
 ..
 149 Collect and deliver soiled and clean linen and...
 150 Provision of clinical service delivery at spec...
 151 Clinical service provision. Forensic service p...
 152 Provide quality outpatient care to patients in...
 153 Clinical service provision (PHC). Surgical and...

	Enquiries \
0	Mr. Q Bowman Tel No: (021) 441 8120
1	Ms M Mogongoa Tel No: (053) 8381610
2	Mr. Q Bowman Tel No: (021) 441 8120
3	Mr. Q Bowman Tel No: (021) 441 8120
4	Mr. Q Bowman Tel No: (021) 441 8120
..	...
149	Mr S Roman Tel No: (028) 212 1071
150	Prof H Bougard Tel No: (021) 402-6444
151	Dr TV Zimri Tel No: (022) 709-7208
152	Mr. AE Patientia Tel No: (021) 815-8895
153	Dr C Adams Tel No: (027) 482-2166

	Closing Date
0	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
1	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
2	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
3	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
4	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
..	...
149	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
150	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
151	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
152	01 August 2025 at 16:00 (walk-in) and 00:00 (o...
153	01 August 2025 at 16:00 (walk-in) and 00:00 (o...

[154 rows x 7 columns]

2 Conclusion

This is an image from a page of the DPSA circular, showing how the job vacancies are presented in a structured format.

```
[15]: Image.open(r"C:\data_repo\DPSA Circular data mining\Vaccancy.png")
```

```
[15]:
```

<u>POST 25/03</u>	:	<u>ASSISTANT DIRECTOR: COMMUNICATIONS REF NO: HR4/4/10/832</u>
<u>SALARY</u>	:	R468 459 per annum
<u>CENTRE</u>	:	Provincial Office: Western Cape
<u>REQUIREMENTS</u>	:	Three (3) year National Diploma (NQF6)/ Undergraduate Bachelor Degree (NQF7) in Communications Science/ Marketing/ Public Relations/ Media Studies and Journalism. Two (2) years supervisory experience. Two (2) years functional experience in Media/ Public Relations/ Marketing/ Communication Services. Valid Driver's Licence. Knowledge: Departmental Policies and Procedures, Public Finance Management Act (PFMA), Basic knowledge of all legislations, Project Management, Batho Pele Principles. Skills: Planning and organizing, Interpersonal, Computer Literacy, Communication skills, Problem Solving, Listening and observation, Negotiation, Event Management.
<u>DUTIES</u>	:	Provide public relations and media liaison services at Provincial level including performing duties and responsibilities as spokesperson of DEL in the Province. Organise stakeholder briefings and exhibitions for the whole Province (Bi-weekly). Market the services of the DEL at Provincial level. Manage DEL internal communications such as notice boards, posting of information on intranet, updating staff on issues affecting the department in the media, production of internal newsletter, etc. Coordinate and facilitate all internal and external events in the province, such as Imbizo outreach, Programmes, outside broadcasts, national commemorative days, etc.
<u>ENQUIRIES</u>	:	Mr. Q Bowman Tel No: (021) 441 8120
<u>APPLICATIONS</u>	:	Jobs-WC3@LABOUR.gov.za Chief Director: Provincial Operations: PO Box 872, Cape Town, 8000 or hand deliver at: Department of Employment and Labour, No.9 Long Street, Cnr Riebeeck and Long Street, Cape Town.
<u>FOR ATTENTION</u>	:	Sub-directorate: Human Resources Management, Western Cape
<u>POST 25/04</u>	:	<u>ASSISTANT DIRECTOR: LABOUR MARKET INFORMATION STATISTICS REF NO: HR4/4/10/834</u>
<u>SALARY</u>	:	R468 459 per annum
<u>CENTRE</u>	:	Provincial Office: Western Cape
<u>REQUIREMENTS</u>	:	Three (3) year National Diploma (NQF6)/ Undergraduate Bachelor Degree (NQF7) in Social Science/Economics/Statistics. Two (2) years supervisory experience. Two (2) years functional experience in labour market data processing/statistics. Valid Driver's Licence. Knowledge: Public Service Act, Public Service Regulations and relevant prescripts, Departmental policies and procedures, Batho Pele Principles, Statistics and Data software packages, Functioning of DEL Labour Centres and Business Units, Labour market dynamics, Labour legislation. Skills: General management, Project management, Communication, Computer Literacy, Research, Interpersonal Relations, Leadership, Data analysis, Analytical, Innovative.
<u>DUTIES</u>	:	Manage performance information of the provinces in the Department of Employment and Labour, in order to ensure that the pre-determined plans and objectives of the organisation are provincially monitored and successfully achieved in line with the applicable government prescripts. Promote accountability and transparency by providing Government and the public with timely, accessible, accurate and quality assured provincial performance system; Monitor and evaluate implementation of Service Delivery Improvement Plans. Assist in consolidating and analysing trends recorded through quarterly performance reports against targets. Manage all the resources of the division.
<u>ENQUIRIES</u>	:	Mr. Q Bowman Tel No: (021) 441 8120
<u>APPLICATIONS</u>	:	Jobs-WC4@LABOUR.gov.za Chief Director: Provincial Operations: PO Box 872, Cape Town, 8000 or hand deliver at: Department of Employment and Labour, No.9 Long Street, Cnr Riebeeck and Long Street, Cape Town.

After running the script, the extracted job data is saved to an Excel file for further analysis or sharing. The structured format allows for easier access and understanding of available job opportunities, addressing the initial problem of PDF documents.

```
[16]: Image.open(r"C:\data_repo\DPSA Circular data mining\excel output.png")
```

```
[16]:
```


ASSISTANT DIRECTOR: COMMUNICATIONS REF NO: HR4/4/10/832						
A	B	C	D	E	F	G
Post	Centre	Salary	Requirements	Duties	Enquiries	Closing Date
ASSISTANT DIRECTOR: COMMUNICATIONS REF NO: HR4/4/10/832	Provincial Office: Western Cape	R468 459 per annum	Three (3) year National Diploma (NQF6)/ Undergraduate Bachelor Degree (NQF7) In Communications Science/ Marketing/ Public Relations/ Media Studies and Journalism. Two (2) years supervisory experience. Two (2) years functional experience in Media/ Public relations/ Marketing/ Communication Services. Valid Driver's Licence. Knowledge: Departmental Policies and Procedures, Public Finance Management Act (PFMA), Basic Knowledge of all legislations, Project Management, Batho Pele Principles. Skills: Planning and organizing, Interpersonal, Computer Literacy, Communication skills, Problem Solving, Listening and observation, Negotiation, Event Management.	Provide public relations and media liaison services at Provincial level including performing duties and responsibilities as spokesperson of DEL in the Province. Organise stakeholder briefings and exhibitions for the whole Province (Bi-weekly). Market the services of the DEL at Provincial level. Manage DEL internal communications such as notice boards, posting of information on Intranet, updating staff on issues affecting the department in the media, production of internal newsletter, etc. Coordinate and facilitate all internal and external events in the province, such as Imibizo outreach, Programmes, outside broadcasts, national commemorative days, etc.	Mr. Q. Bowman Tel No: (021) 441 8120	01 August 2025 at 16:00 (walk-in) and 00:00 (online)
ASSISTANT DIRECTOR: LABOUR MARKET INFORMATION STATISTICS	Provincial Office: Western Cape	R468 459 per annum	Three (3) year National Diploma (NQF6)/ Undergraduate Bachelor Degree (NQF7) In Social Science/Economics/Statistics. Two (2) years supervisory experience. Two (2) years functional experience in labour market data processing/statistics. Valid Driver's Licence. Knowledge: Public Service Act, Public Service Regulations and relevant prescripts, Departmental policies and procedures, Batho Pele Principles, Statistics and Data software packages, Functioning of DEL Labour Centres and Business Units, Labour market dynamics, Labour legislation. Skills: General management, Project management, Communication, Computer Literacy, Research, Interpersonal Relations, Leadership, Data analysis, Analytical, innovative.	Manage performance information of the provinces in the Department of Employment and Labour, in order to ensure that the pre-determined plans and objectives of the organisation are provincially monitored and successfully achieved in line with the applicable government prescripts. Promote accountability and transparency by providing Government and the public with timely, accessible, accurate and quality assured provincial performance system; Assist in consolidating and analysing trends recorded through quarterly performance reports against targets. Manage all the resources of the division.	Mr. Q. Bowman Tel No: (021) 441 8120	01 August 2025 at 16:00 (walk-in) and 00:00 (online)
ASSISTANT	Provincial Office: Western Cape	R468 459 per annum	Three (3) year qualifications National Diploma (NQF6)/ Undergraduate Bachelor Degree (NQF7) in Risk Management/ Internal Audit/ Accounting/ forensic investigations. Two (2) years supervisory experience. Two (2) years functional experience in anti-fraud/ corruption environment. Valid Driver's 7 Licence. Knowledge: Investigative Principles and Practices, Departmental and Fund's Policies and Procedures, All Labour legislations, Anti-Fraud and Corruption Policies, Legal environment: Court and Criminal Procedures, Fraud related administration and operations, Batho Pele Principles, Public Service Regulations, Public Financial Management Act (PFMA), Unemployment Insurance and Unemployment Insurance Contributions Act, Basic Conditions of Employment Act, Labour Relations Act, Public Service Act, CIDA. Skills: Planning and organizing, Time Management, Conflict Management, Analytical, Investigation, Communications (Verbal & Written), Computer Literacy, Presentation.	Implement Fraud detections and Anti-Corruption Strategies for Provincial Office. Conduct investigations on reported cases and Fraud and Corruption. Assist in recovery of all assets acquired fraudulently from the Provincial Office. Analyse systems capabilities to anti-fraud and corruption management programmes.	Mr. Q. Bowman Tel No: (021) 441 8120	01 August 2025 at 16:00 (walk-in) and 00:00 (online)
INSPECTOR: BCTA REF NO: HR4/4/10/832	Cape Town Labour		Three (3) year tertiary qualification National Diploma (NQF6)/ Undergraduate Bachelor Degree (NQF7) In Labour Relations/ BCOM Law/ L.B. Two (2) years functional experience in inspection and enforcement services. Valid driving license. Knowledge: Departmental	To plan and independently conduct substantive inspections with the aim of ensuring compliance with the Basic Conditions of Employment Act (BCEA). Sectoral Determination and the National Minimum Wage Act. Plan and execute investigations independently on suspected cases pertaining to contravention of	Mr. Q. Bowman Tel	01 August 2025 at

2.0.1 Reference: <https://github.com/ConradKatlegoMogane>