DIRISA-Qualifiers-2024

Predicting Employment Trends Using Neural Networks

Project Overview

- The goal of this project is to predict employment trends in South Africa using various demographic, health, education, and economic data.
- We used data spanning from 2012 to 2022, focusing on thorough data cleaning, feature engineering, and visualization to extract meaningful insights before modeling.

Data Cleaning Process

The data cleaning process involved several steps:

Data Importation → Removing Missing Data → Data Standardization

Feature Engineering Pt.1

New features were engineered to enhance the dataset:

- Age and Age Groups: Unified under a single "Age" feature with categorized age groups.
- Gender: Standardized across all datasets.
- Disability Status: Merged disability indicators.
- Population Group: Combined "Race" and "Population" features.

Feature Engineering Pt.2

- Geographic Location: Merged data from GeoType*, Metro, Metro_code, and Prov.
- Chronic Health Conditions: Combined indicators for various chronic illnesses.
- Education Level: Categorized as "Primary," "Secondary," or "Tertiary."
- Employment Status: Created a unified employment feature from multiple indicators.

Data Overview (Post-Cleaning)

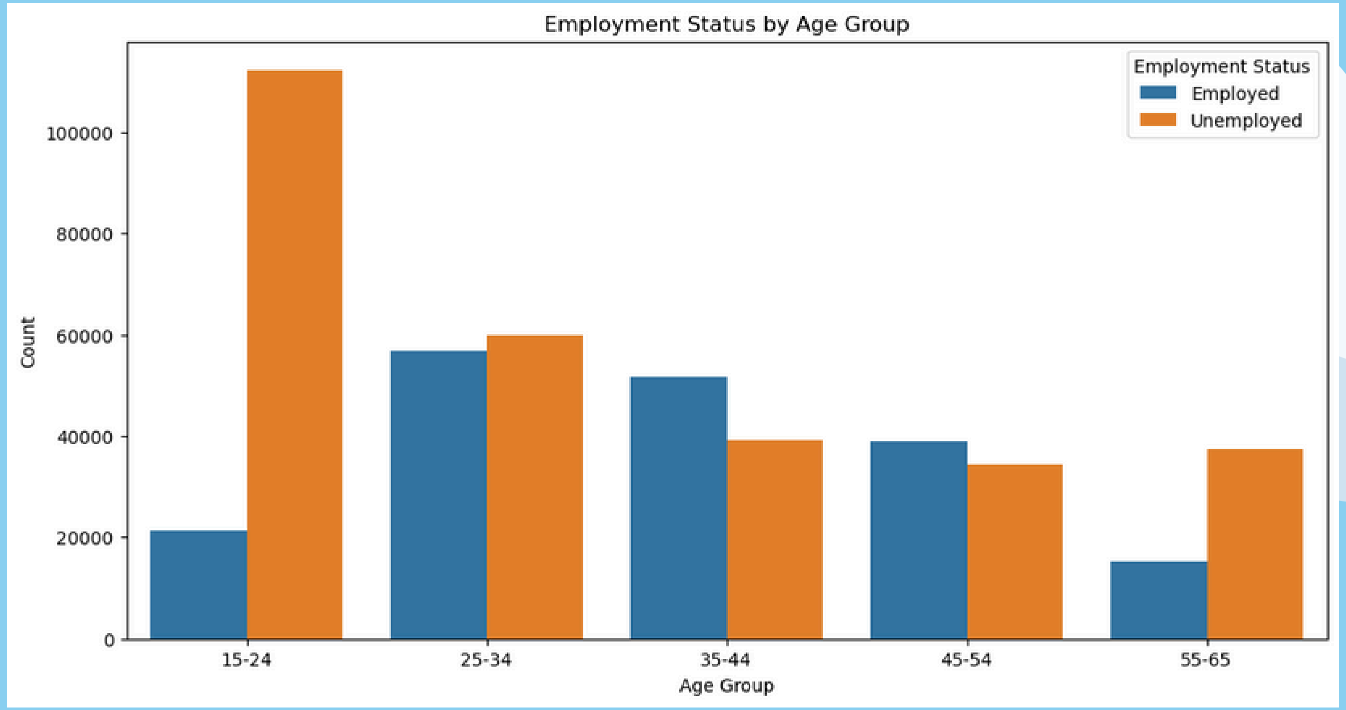
After cleaning, the dataset comprises 482,945 records with the following key features:

- Age: Focused on the working population (15-65 years).
- Gender: Male and Female categories.
- Race: Population groups.
- Province: Regional distribution.
- Chronic Illnesses and Disabilities: Health-related features.
- Education Level: Educational attainment.
- Employment Status: Consolidated employment indicators.

DIRISA-Qualifiers-2024

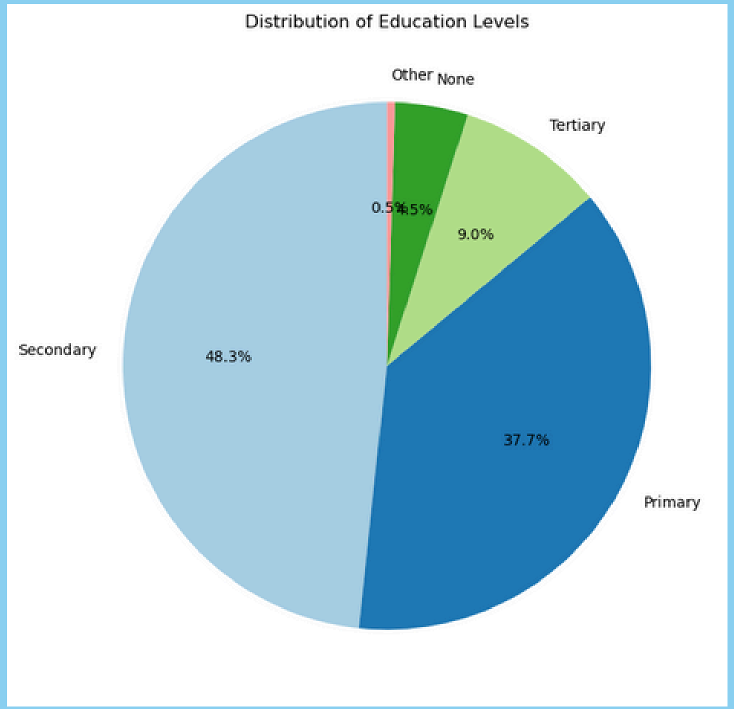
Data Visualizations

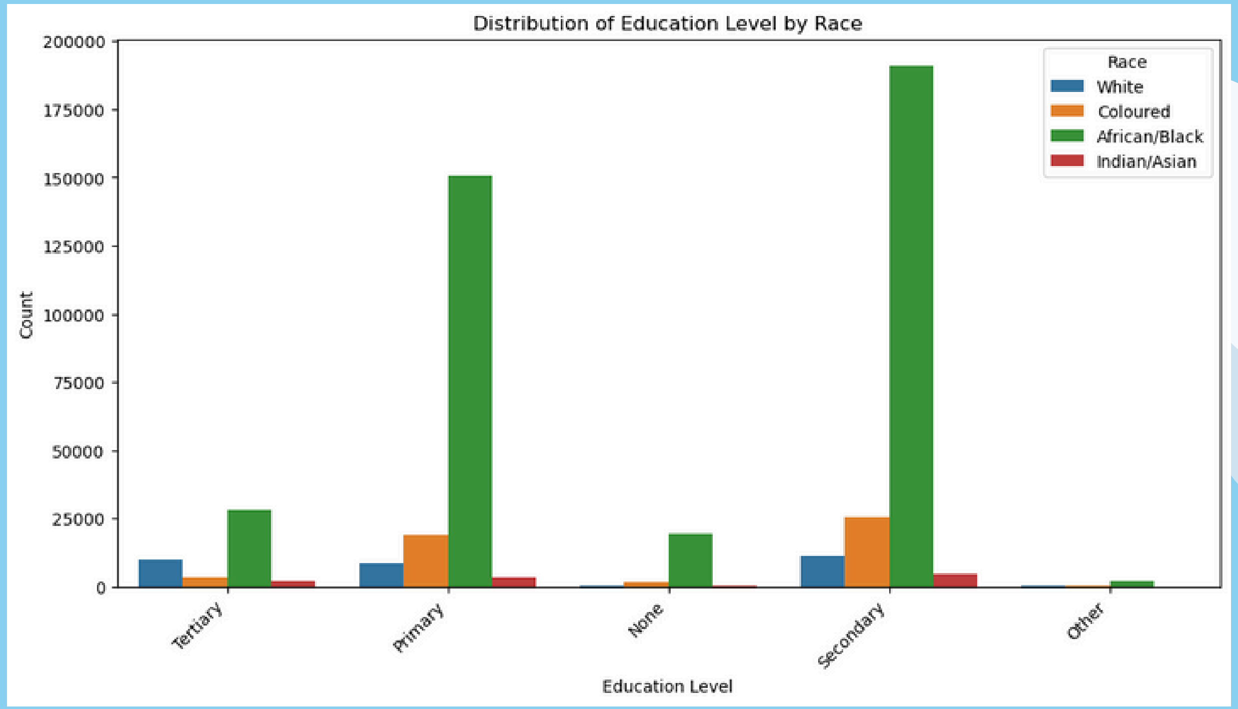
Distribution of Employment Status by Age



The plots above suggests that there are many young people who are unemployed. As much as unemployment is high in the age group 15-24, some people are still in school, we have to account for those people, only a few are employed or receive salary or wages which going to school. The violin plot shows that most people in their 30s are employed, which is reasonable as most adults have to provide for their families.

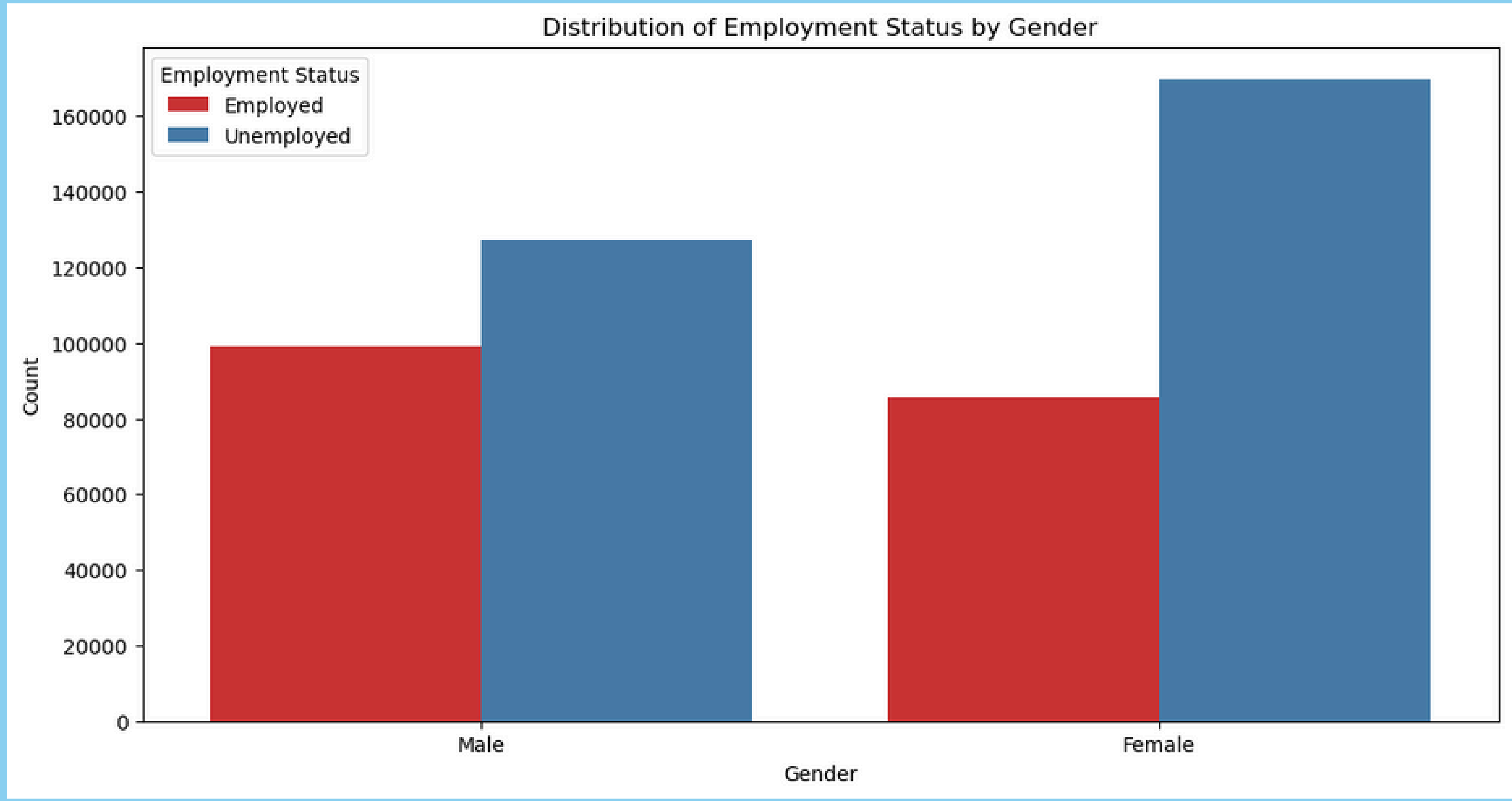
Distribution of Education Level by Race





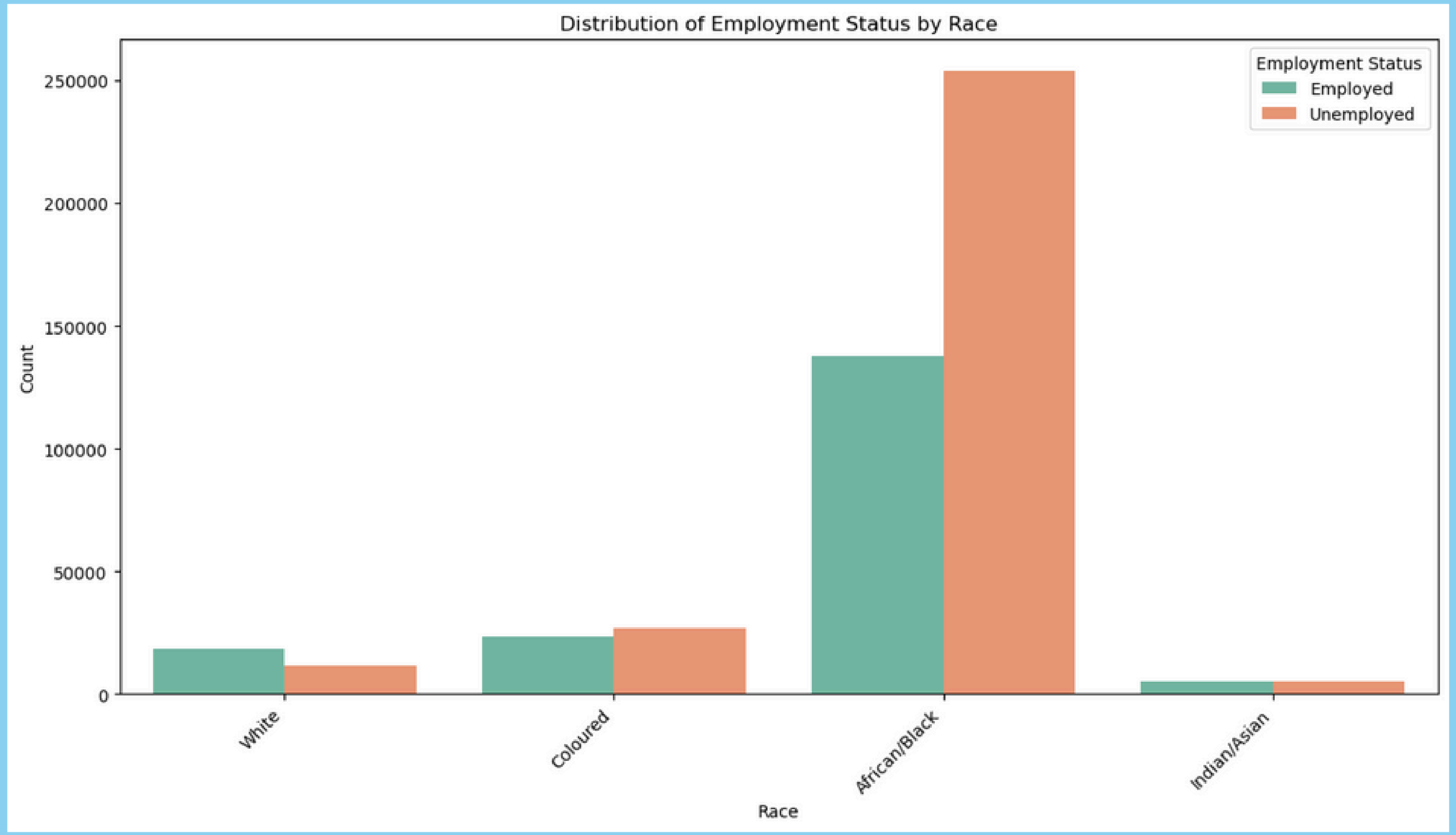
The bar graph above suggests that most African/black people only went tp school until Secondary. With the green bars very high, it shows that the population of African/Black is very high compared to other race. Also very less White and Indian/asian did not go to school.

Distribution of Employment status against Gender



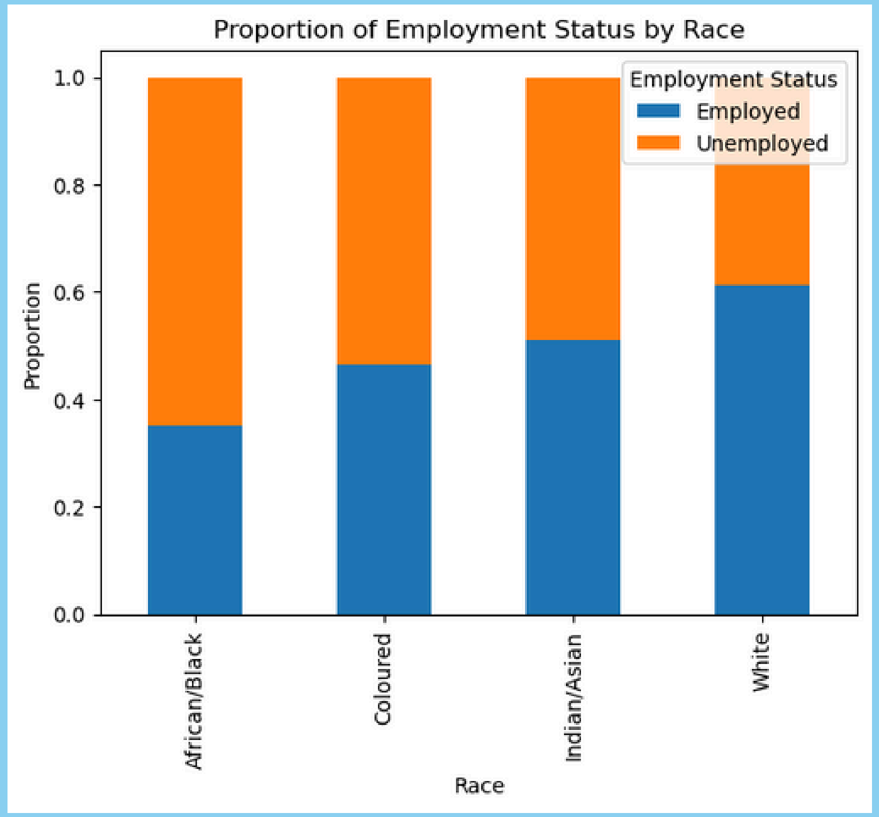
Most females are unemployed with number of employed females less that the number of employed females. Overall across both genders, number of unemployed individuals is more than number of employed individuals.

Distribution of Employment Status by Race



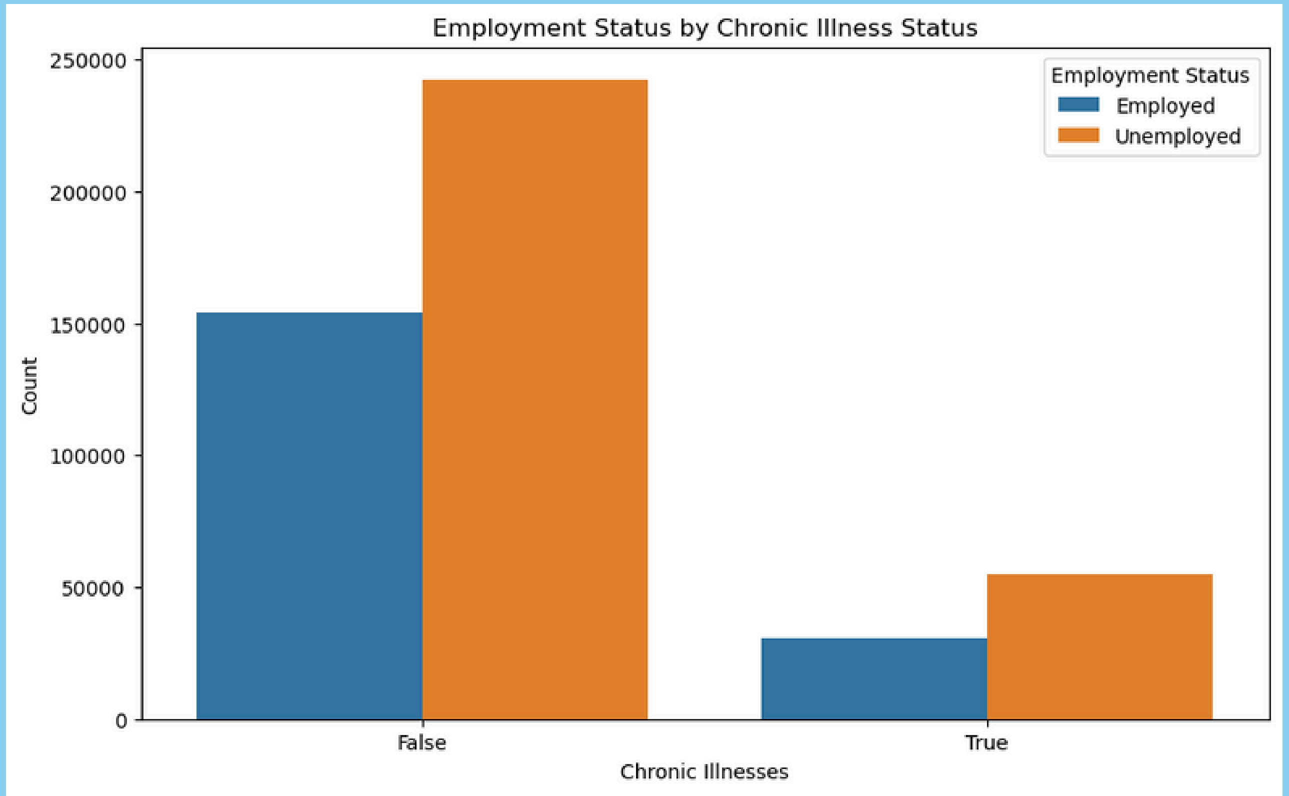
African/Black has the most unemployed and employed individuals.

Proportion of Employment Status within each race



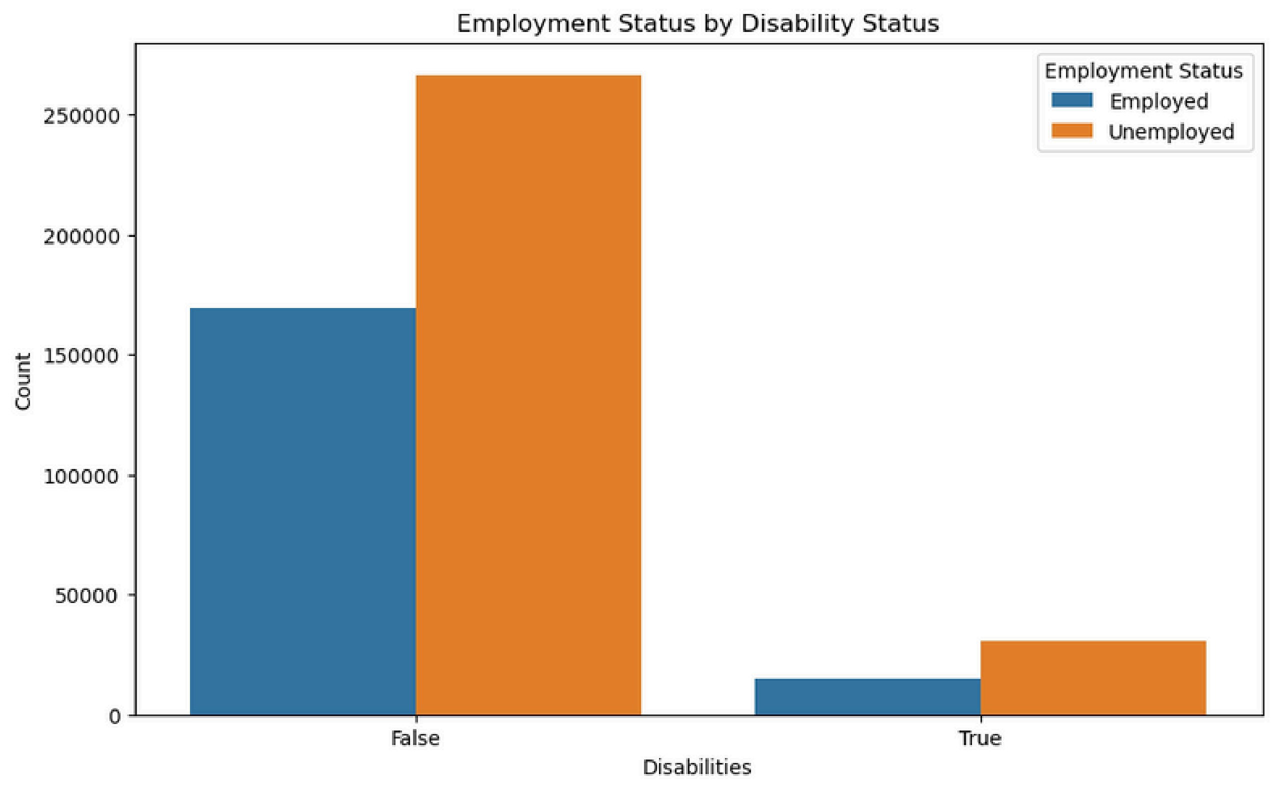
These two plots suggests that unemployment is high in both males and females, and across all races. The stacked bar plot shows that unemployment is high in all races, but higher in African/Black and Coloured.

Grouped bar plot of Employment Status by Chronic Illness status



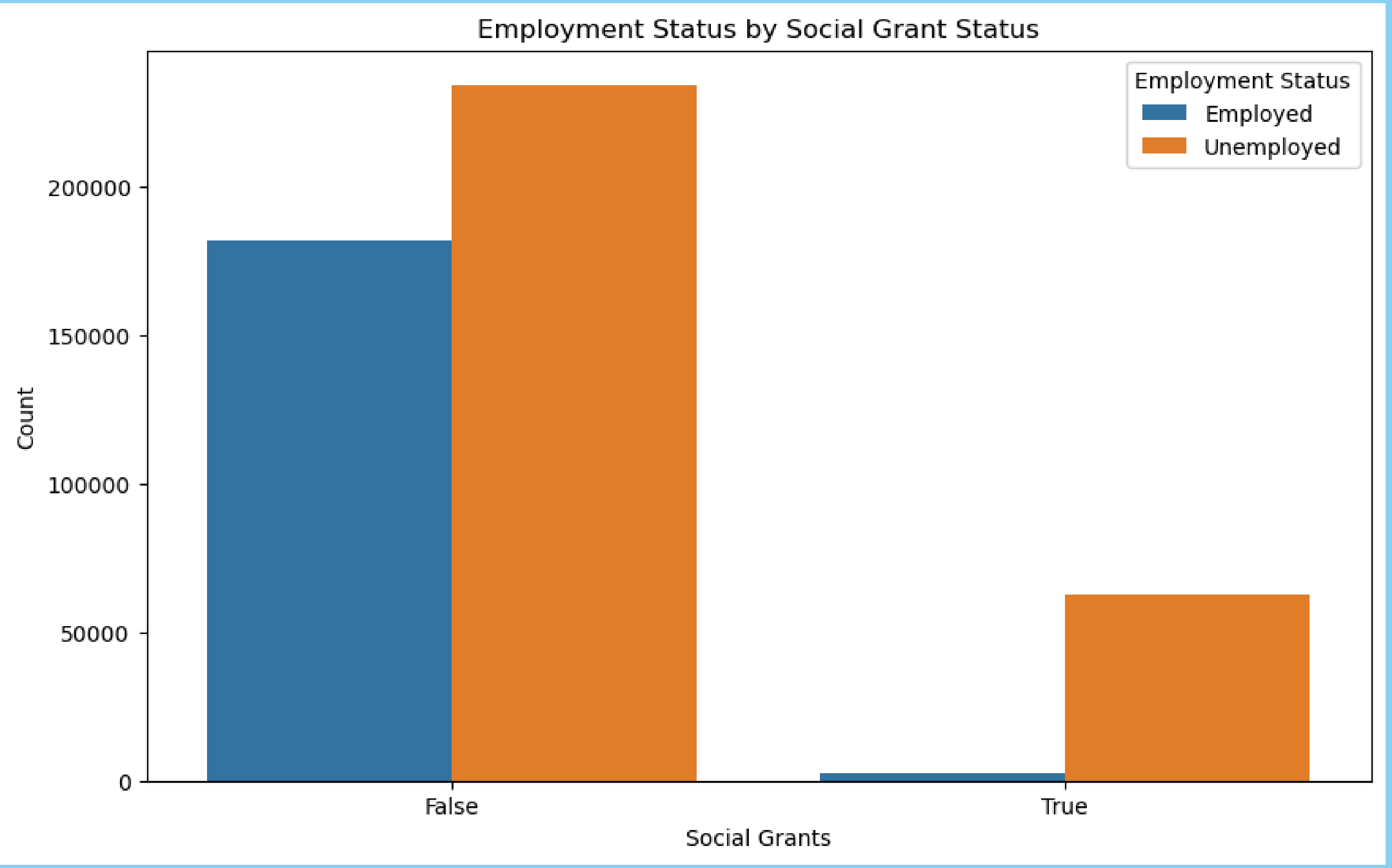
From the plot we notice that there are people with Chronic illnesses that are employed, the proportion is low compared to the individual who do not have any chronic illnesse and go to work.

Grouped bar plot of Employment Status by Disability status



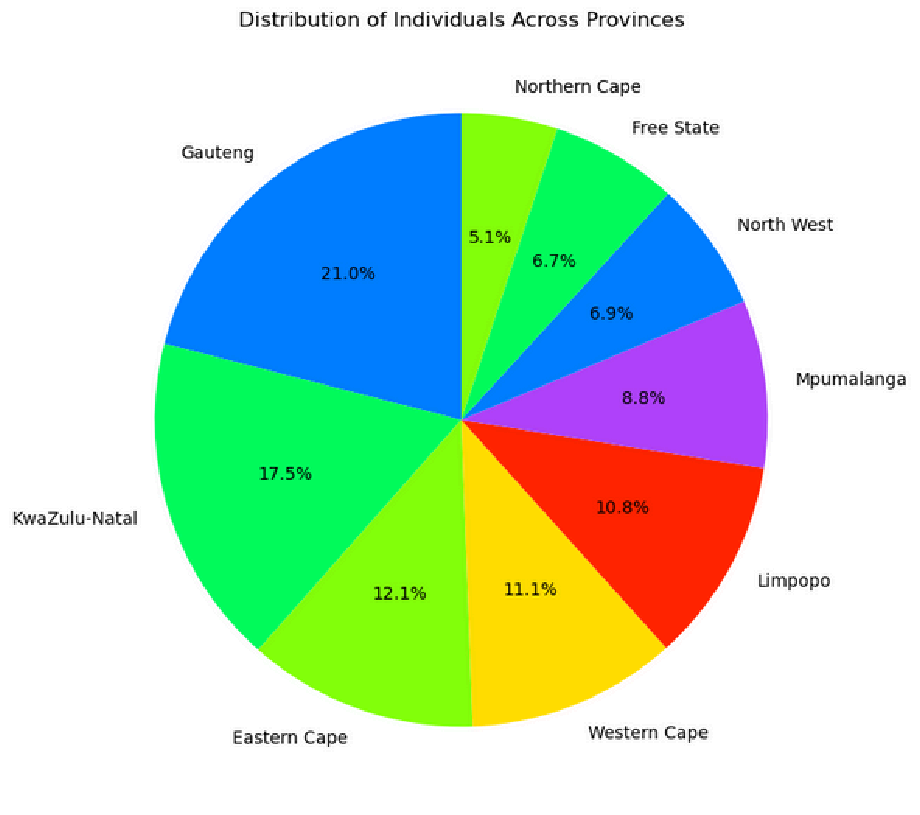
The plots above suggests that unemployment is high in both people with chronic illness and disabilities.

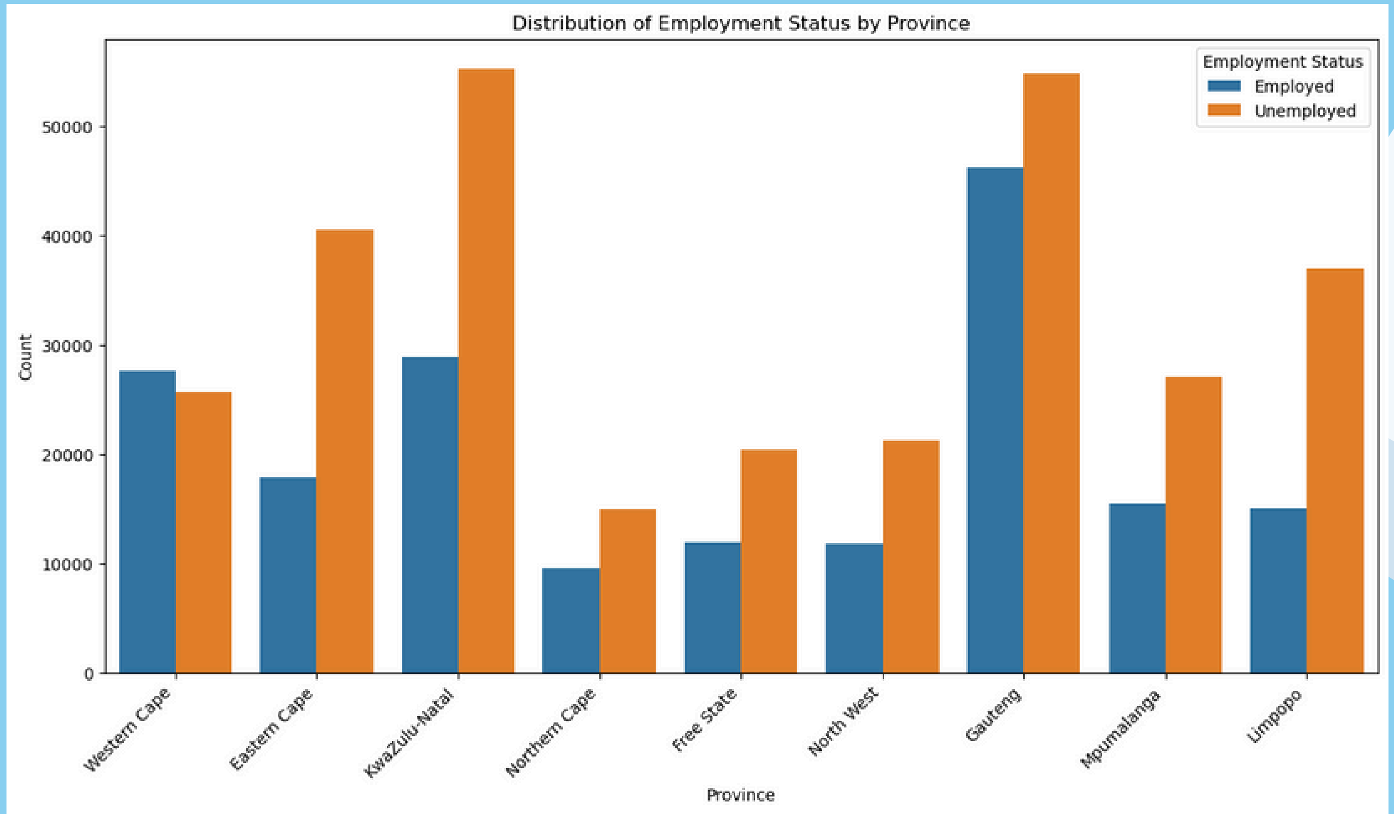
Impact of Social Grants on Employment



The plot above suggests that most people who receive social grants are unemployed.

Regional Disparities in Employment





The plots above suggests that unemployment is high in all provinces, but higher in Eastern Cape, KwaZulu-Natal, and Limpopo. Gauteng with the highest number of employed individuals.

DIRISA-Qualifiers-2024

Model Training

Data Pre-processing

We used the following methods to prepare our data to be in a suitable format for training and test

Shuffling and Splitting → Encoding Categorical Variables

Neural Network Architecture

Model Architecture:

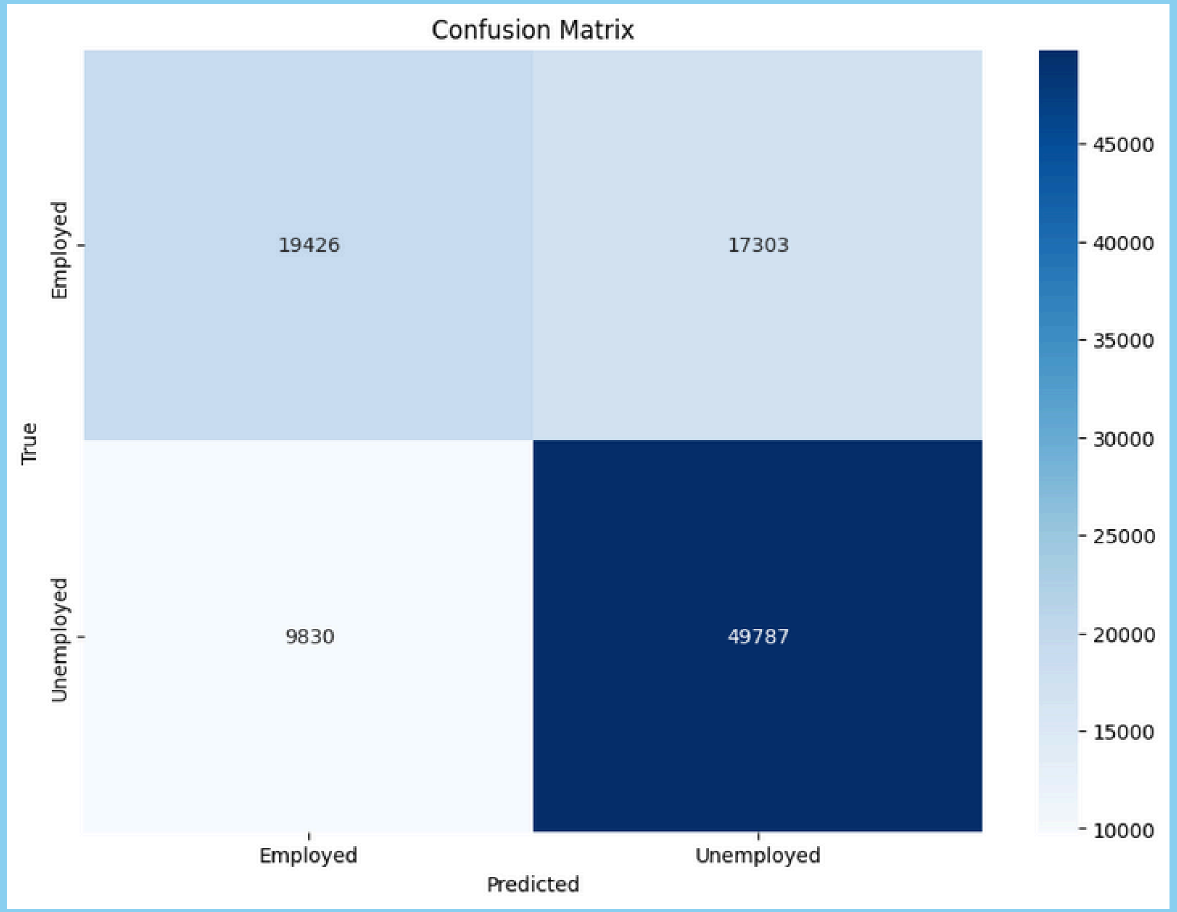
- Built a Sequential model with 4 layers:
- Input Layer with 128 neurons, activation: ReLU
- Hidden Layers with 64 and 32 neurons, activation: ReLU
- Output Layer with 2 neurons (for binary classification), activation: Softmax

Model Compilation:

- Compiled using Adam optimizer and Categorical Crossentropy loss function.
- Displayed the model summary showing a total of 13,602 trainable parameters.

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

Model Evaluation



Evaluated the model using the test set, achieving a test accuracy of 71.84%

DIRISA-Qualifiers-2024

Summary

Model Optimization: Changes and Rationales

In the process of refining our neural network model to predict employment status, several changes were implemented between the initial version (v1) and the improved version (v2). These adjustments aimed to enhance model accuracy and generalization. Here are the detailed changes and their rationales:

1. **Increased Model Complexity**
2. **Incorporation of Dropout Layers**
3. **Introduction of Early Stopping**
4. **Adjustment of Training Epochs**
5. **Changes in Hyperparameters**

These enhancements are designed to provide a robust framework for the neural network, aiming to improve its ability to generalize well to new, unseen data while mitigating the risk of overfitting.

DIRISA-Qualifiers-2024

Thats all from us!