Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 7, Issue 4, Pages 87-104, October-December 2023

Elevating Opportunities: Data Science Education as a Catalyst for Empowering Nigerian Youth and Tackling Unemployment

Ajao Isaac Oluwaseyi[1*] & Idoko Ifeoma Geraldine[2]

[1,2]Department of Statistics, The Federal Polytechnic, Ado-Ekiti, Nigeria.
Email: isaacoluwaseyiajao@gmail.com[*]

DOI: https://doi.org/10.46759/IIJSR.2023.7411

Article Received: 16 October 2023 *Article Accepted: 24 December 2023* *Article Published: 30 December 2023*

ABSTRACT

This article delves into the possibilities of leveraging data science education to empower Nigerian youths and combat unemployment challenges. It delves into the perceived advantages and obstacles associated with acquiring data science skills, while also exploring the impact of promoting skill acquisition and providing practical usage opportunities on youths' perspectives toward solving unemployment. The study further delves into the significance of background knowledge in statistics, mathematics, and computer science in the process of skill acquisition. Employing quantitative analysis and validity assessments, the article aspires to furnish valuable insights for policymakers and stakeholders aiming to formulate effective programs. These programs are intended to equip Nigerian youths with the necessary skills to thrive in the job market, thereby contributing to both personal success and overall economic growth.

Keywords: Data science education; Nigerian youths; Unemployment challenges; Skill acquisition; Quantitative analysis; Empowerment.

1. Introduction

The persistent challenge of youth unemployment remains a significant concern for developing economies worldwide, including Nigeria, which has a burgeoning young population. In the evolving digital age reshaping industries and labor markets, equipping the youth with relevant and in-demand skills has become crucial to address this pressing issue. Data science has emerged as a transformative domain, driving data-driven decision-making and offering immense potential for career opportunities. Against this backdrop, the present study aims to explore the promising role of data science education in empowering Nigerian youths and combating unemployment.

Literature on youth unemployment in Nigeria emphasizes the need for proactive interventions to bridge the skills gap and enhance employability prospects. Data science, with its interdisciplinary nature and emphasis on analytical and technical proficiencies, provides a promising avenue for equipping the youth with in-demand skills. Recent studies underscore the positive impact of data-driven technologies on various industries, highlighting the relevance of data science skills in the current job market.

A comprehensive literature review reveals that perceptions of acquiring data science skills play a pivotal role in shaping youths' interest and motivation to pursue this field. Studies in other contexts indicate that perceived benefits and career opportunities significantly influence individuals' decisions to engage in skill acquisition. Moreover, addressing barriers to skill acquisition is essential to ensure inclusivity and accessibility of data science education for all Nigerian youths.

Encouragement and support from stakeholders, including educational institutions, governmental bodies, and private sectors, are vital in nurturing a data science-focused ecosystem that fosters skill development among youths. Additionally, understanding the impact of a lack of background in relevant subjects such as statistics, mathematics, and computer science is crucial in devising targeted interventions that address knowledge gaps and facilitate a smooth transition into data science education.

The primary objective of this study is to investigate the relationship between data science education and its impact on combating youth unemployment in Nigeria. By delving into the perceptions of acquiring data science skills, the influence of encouragement and barriers, and the significance of background knowledge in statistics, mathematics, and computer science, this study seeks to offer evidence-based insights. Through rigorous analysis and validity assessments, the study aims to inform policymakers, educators, and stakeholders seeking effective strategies to promote data science education among Nigerian youths, fostering employability and contributing to the nation's socioeconomic development.

In light of these considerations, the study aims to shed light on essential aspects of data science skill acquisition that can effectively uplift the aspirations and employability prospects of Nigerian youths, forging a pathway towards a brighter and more inclusive future. Subsequent sections will present the research methodology, data analysis, and findings, providing a comprehensive understanding of the relationships between data science education and combating youth unemployment in Nigeria.

2. Literature Review

The authors Looman et al. (2023) in their paper discuss the importance of data science education in Nigeria and how it can empower Nigerian youths. They argue that data science education can help bridge the gap between the skills required by the job market and the skills possessed by Nigerian youths. They propose a curriculum that includes both theoretical and practical aspects of data science, such as data analysis, machine learning, and programming. The authors also suggest that data science education can help address social and economic challenges in Nigeria, such as unemployment and poverty. The paper highlights the need for collaboration between academia, industry, and government to provide Nigerian youths with access to data science education. The authors suggest that partnerships between universities and industry can help provide students with practical experience and job opportunities. They also propose that the government can support data science education by providing funding and creating policies that encourage the development of data science programs. Overall, the paper emphasizes the potential of data science education to empower Nigerian youths and contribute to the development of Nigeria as a whole. Investigation on the impact of education expansion on job opportunities for workers with and without the new education (Schultheiss, 2023). The authors exploit a quasi-random establishment of Universities of Applied Sciences (UASs) in Switzerland and analyze job advertisement data using machine learning methods. They find that the establishment of UASs led to an upskilling among workers who did not acquire the new educational degree, particularly in the STEM fields involving R&D. The improved job opportunities for vocationally trained workers translated into tangible improvements in labor market outcomes, and these workers experienced significant wage gains. Overall, the paper suggests that expanding tertiary education can have positive spillover effects on the job opportunities and wages of workers without the new educational degree.

This research paper by Smaldone et al. (2023) focuses on the employability criteria and prerequisites for data scientists in the digital labor market. The study aims to identify the skills, experience, and qualifications sought by employers for their data scientists by analyzing job advertisements published on US employment websites. The paper addresses three emerging implications concerning the employability of data scientists, including averting a skills gap in a rapidly evolving domain, the alignment between employers and academia, and the importance of employability for data scientists. The key contribution of this study is to provide a data-driven pathway for employability and avoiding skills gaps and mismatches in a profession that is pivotal in the Industry 4.0.

In a study carried out by Donovan et al. (2023), their paper evaluates InPlace Learning, a blended learning approach to suicide prevention workforce education in youth services. The approach combines group video-guided workshops, follow-up Q&A via web-based 'office hours', and continual learning through job aids and brief refreshers. The study examines the feasibility of the approach, adoption and reach of InPlace Learning, as well as education outcomes for staff, including knowledge, self-efficacy, and transfer of learning. The findings indicate that InPlace Learning is a feasible and acceptable approach to suicide prevention workforce education in youth services with a positive impact on educational outcomes for staff.

The skill gap between the acquired university curriculum and the requirements of the job market Aljohani et al. (2023). While the paper does not explicitly mention youth skill acquisition, the findings and recommendations of the study can be relevant to young people who are preparing to enter the job market. The paper emphasizes the importance of aligning the curriculum with the skills required by the job market and identifying the gaps in the current system. This can help policymakers and educators to design programs that better equip young people with the skills they need to succeed in the job market.

In a research by Sánchez–Marco (2023) the focus is on the acquisition of nontechnical skills such as communication, leadership, teamwork, and situational awareness in emergency medical services and critical care units. The study evaluates the effectiveness of educational interventions on the acquisition of these skills using simulation interventions. The results showed that simulation interventions significantly improve the levels of knowledge, attitude, self-efficacy, and nontechnical skills performance of healthcare professionals.

Becker et al. (2023) shed light on the application of Partial Least Squares Structural Equation Modeling (PLS-SEM) in various disciplines, including hospitality management research, which has garnered attention from both methodological and applied researchers. As PLS-SEM is relatively new, researchers face uncertainties and challenges in its implementation. To address these issues, the authors conducted a text analysis of a renowned PLS-SEM discussion forum, focusing on the most significant questions and answers to offer valuable guidance for researchers struggling with its use in different contexts.

The paper by Kwong et al. (2015) discusses the use of SmartPLS software for Partial Least Squares Structural Equation Modeling (PLS-SEM) in marketing research. SmartPLS is user-friendly software with advanced reporting features that has gained popularity since its launch in 2005. The paper addresses the lack of instructional materials available for the software and provides guidance for beginners on how to use PLS-SEM in marketing research. The paper also provides recommendations for sample size determination, missing value replacement, and path modeling estimation.

The use of Smart PLS software for Partial Least Squares Structural Equation Modeling (PLS-SEM) in construction economics and management research was discussed by Yahaya (2019). The software is popular due to its user-friendly interface, advanced reporting features, and the ability to choose formative or reflective models. The paper provides a step-by-step method for measurement and structural model evaluation, as well as discussing model assessment and supplementary techniques for evaluating the robustness of outcomes. The aim of the paper is to help beginners in various fields of study to understand how PLS-SEM can be used in their research.

Law (2020) investigates the learning transfer of academic English among undergraduate students using partial least squares structural equation modeling (PLS-SEM). The study merged variables from various frameworks to create a network of latent variables and selected four main constructs for the structural model: content relevance, transfer outcome, understanding of learning, and transfer/applicability. The study measures the students' perceived degree of success in handling what is required in their CCCs. The paper also provides guidelines for evaluating PLS-SEM results.

The article by Deryabin & Glukhov (2022), examines the psychological and educational implications of data analysis practices for students in shaping their personal and professional futures and contributing to local communities. It focuses on a 10-day "Data Campus" Data Analysis Bootcamp, where students with limited mathematical training learn to conduct data research using the CRISP-DM cycle. The research, involving 600 students aged 14 to 18, employed an ascertaining experiment and demonstrated that students engaged emotionally and actively in projects related to regional and urban data. While students easily mastered technical programming skills, they faced challenges in developing productive solutions for research tasks due to a lack of experience in analyzing intersubject relationships and representing research objects in a multidimensional feature space. The article suggests that educational programs teaching data analysis methods in social sciences and humanities should provide students with methodological assistance, such as project mentoring, to help them conceptualize projects before working with data.

To implement educational programs teaching data analysis methods in social sciences and humanities, students should be provided with appropriate methodological assistance in the form of project mentoring and pre-project consultations to conceptualize the research project before working with data (Deryabin & Glukhov, 2022).

The analysis of the changes in MET due to the influence of the Fourth Industrial Revolution (4IR) shows that the demand for new competences such as data analysis and IoT has increased. The changes in youth's fields of interest and the increased popularity of informatics and data science are not as attractive in the context of MET (Mickienė & Valionienė, 2021).

Moon et al., 2023 work on the increasing prevalence of data and its influencing the way individuals engage with each other and the world. For young people growing up in a society driven by data, it is crucial that they acquire fundamental data literacy skills. An integral aspect of data literacy involves the capability to gather, analyze, visualize, and derive meaning from data. These activities are influenced and molded by the tools employed by young individuals in performing these data-related tasks. Recognizing the pivotal role tools play in facilitating and supporting youth in their interaction with and interpretation of data, comprehending the types of tools utilized and their application in educational settings is essential. This paper presents an analysis of the tools used for data collection and analysis in four widely adopted high school data science curricula. The analysis explores both the types of tools employed and the datasets they are utilized to examine. This research contributes to our comprehension of how young individuals are introduced to concepts and practices in the field of data science and the influence of tools on shaping these experiences.

In youth-focused community and citizen settings, young individuals in both classrooms and community-based programs generate data that scientists, resource managers, and community members utilize. This interconnected data sets the context for learners' scientific activities within broader datasets, projects, and communities, impacting youth agency. To explore opportunities for active learning with data in these settings, we examine how youth engage with data across eight school and community projects and analyze their discussions about their data and contributions. Through 54 participant interviews, we found that youth perceived the data they produced as serving purposes such as broader scientific endeavors, their personal learning, and community initiatives. The significance of nested data uses became evident when youth interacted with end users, were exposed to the larger datasets they contributed to, or took actions tied to the data. However, not all youth recognized, believed in, or valued the idea that their data would be utilized by others. The framing of tasks, engagement with community users, data production protocols, and emphasis on youth-identified questions were identified as factors influencing youth perceptions and potentially shaping conditions for learner agency. These findings offer insights into the establishment of conditions that support expansive learning and agency in youth engagement with data (Harris et al., 2020).

Data science has garnered considerable interest by offering the potential to transform extensive data sets into valuable predictions and insights. This article explores the significance of data science for scientists, considering it from three viewpoints: statistical, computational, and human. While each perspective plays a crucial role in data science, the article contends that the true essence lies in effectively integrating all three components (Xu et al., 2021).

The literature review highlights the critical importance of skill acquisition and use of PLS-SEM in empowering Nigerian youths and combatting unemployment. By understanding youths' perceptions, addressing barriers, and providing stakeholder support, data science education holds immense potential to equip the youths with in-demand skills for the evolving job market. Additionally, addressing knowledge gaps and enhancing foundational knowledge in relevant subjects is essential for effective skill acquisition. The subsequent sections will present the research methodology, data analysis, and findings, aiming to contribute evidence-based insights for policymakers, educators, and stakeholders seeking to develop impactful strategies for youth empowerment and socioeconomic growth in Nigeria.

3. Research methodology

In Partial Least Squares Structural Equation Modeling (PLS-SEM), several equations are used to estimate model parameters and assess the relationships between latent variables and their indicators.

3.1. Measurement Model Equations

In the measurement (reflective or outer) model, the relationships between latent variables and their observed indicators are represented using the following equations:

For the k-th indicator of the i-th latent variable:

$$X_i = \lambda_i \xi_i + \varepsilon_i \quad (1)$$

Where:

X_i = Observed indicator for latent variable i

λ_i = Loading of the indicator on latent variable i

ξ_i = Latent variable i (construct)

ε_i = Measurement error of the indicator

3.2. Inner Model Equations

In the inner model, the relationships between latent variables are represented using path coefficients:

For the relationship between the i-th and j-th latent variables:

$$\xi_j = \beta_{ij} \xi_i + \zeta_{ij} \quad (2)$$

Where:

ξ_j = Dependent latent variable (j)

β_{ij} = Path coefficient representing the effect of latent variable i on latent variable j

ξ_i = Independent latent variable (i)

ζ_{ij} = Error term of the relationship between latent variables i and j

3.3. Total Effects Equations

Total effects represent the total impact of one latent variable on another, including both direct and indirect effects. They are calculated as the sum of direct and indirect effects:

Total effect of latent variable i on latent variable j

$$TE_{ij} = \beta_{ij} + \sum(\beta_{ik} \lambda_k \sum(\beta_{kl} \lambda_l \ldots \sum(\beta_{lm} \lambda_m))) \quad (3)$$

Where:

β_{ij} = Path coefficient representing the direct effect of latent variable i on latent variable j

λ_k = Loading of indicator k on latent variable k

Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 7, Issue 4, Pages 87-104, October-December 2023

β_{ik} = Path coefficient representing the effect of latent variable i on indicator k

β_{kl}, β_{lm}, etc. = Path coefficients representing the effects of latent variables on each other

λ_l, λ_m, etc. = Loadings of indicators on latent variables

These equations are fundamental in PLS-SEM for estimating model parameters, evaluating latent variable relationships, and assessing the overall model fit. They are used in conjunction with the algorithm of PLS-SEM to obtain model estimates and test hypotheses about the relationships between latent variables and their indicators.

4. Data Analysis

All analyses are carried out using Smart PLS-SEM 4 (Ringle et al., 2022) and R version 4.3.1 (R Core Team, 2023). The data used was collected through structured questionnaire among the Nigerian youths.

Table 1. Descriptive statistics of the latent variables

Constructs	Mean	Median	Std.dev	Kurtosis	Skewness	Cramér-von Mises test statistic	Cramér-von Mises p value
BAR	0.000	0.062	1.000	-0.126	0.186	0.682	0.000
BEN	0.000	-0.214	1.000	-0.293	-0.343	0.590	0.000
EMP	0.000	0.092	1.000	9.375	-2.403	1.006	0.000
ENC	0.000	0.008	1.000	-0.803	-0.361	0.650	0.000
KNOW	0.000	0.063	1.000	-0.571	-0.266	0.314	0.000
USE	0.000	0.044	1.000	1.969	-0.794	0.295	0.000

Where

EMP - Why acquiring skills in data science will solve unemployment among youths

USE- Extent of Usage of Python, R, Power BI, and Others in Acquisition of Data Science Skills

KNW - How Lack of Background in Statistics, Mathematics, and Computer Science Affects Acquisition of Data Science Skills

BAR - Barriers to Acquisition of Data Science

BEN - Perceived Benefits of Acquiring Data Science Skills

ENC- Ways to Encourage Acquisition of Data Science Skills

The table presents descriptive statistics of latent variables related to acquiring data science skills and its impact on unemployment among Nigerian youths. The constructs include Barriers to Acquisition of Data Science, Perceived Benefits of Acquiring Data Science Skills, Why acquiring data science skills will solve unemployment, Ways to Encourage Acquisition of Data Science Skills, Lack of Background in Relevant Subjects Affecting Skill

Acquisition, and Extent of Usage of Data Science Tools. All constructs have a mean close to zero, indicating neutral average responses. Significant non-normality is observed in all constructs, as indicated by low p-values in the Cramér-von Mises test. Further analyses, considering the non-normality, will be done to gain a deeper understanding of the relationships between these variables.

Table 2. Path coefficients of the constructs

Constructs	Path coefficients
BAR -> USE	0.034
BEN -> USE	0.632
ENC -> USE	-0.081
KNOW -> USE	0.164
USE -> EMP	0.778
KNOW x ENC -> USE	0.288
KNOW x BAR -> USE	-0.236

Table 2 and Fig.1 present the path coefficients of the constructs in the study. The path coefficients represent the strength and direction of the relationships between the latent variables. Notably, the constructs "Perceived Benefits" (BEN) and "Extent of Usage of Data Science Tools" (USE) have a strong positive relationship with a path coefficient of 0.632. On the other hand, "Ways to Encourage Skill Acquisition" (ENC) and "Barriers to Acquisition" (BAR) show weaker associations with USE, with path coefficients of -0.081 and 0.034, respectively.

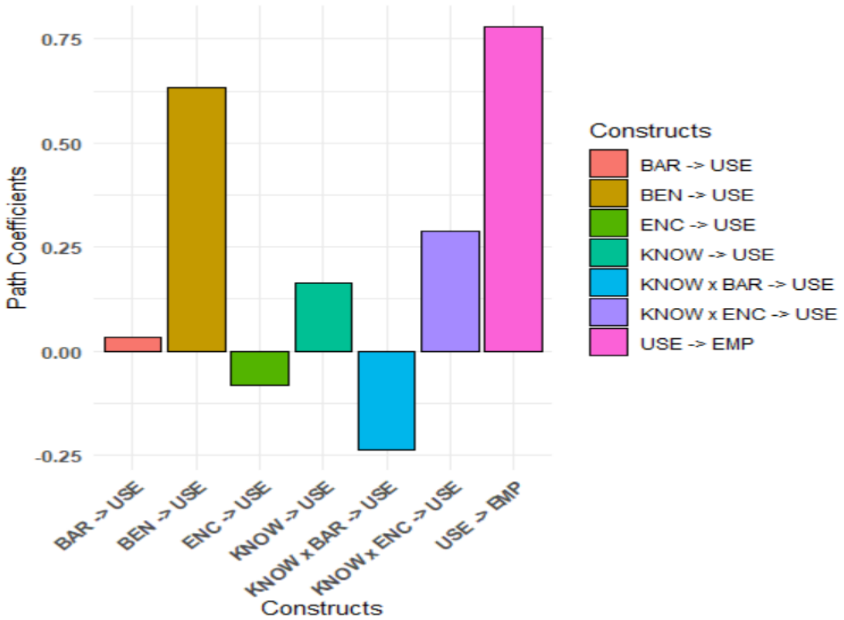

Figure 1. Bar chart showing the path coefficients

Additionally, there is a significant positive relationship between USE and the core construct "Why acquiring data science skills will solve unemployment" (EMP) with a path coefficient of 0.778. Moreover, the interaction between

Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 7, Issue 4, Pages 87-104, October-December 2023

"Lack of Background in Relevant Subjects Affecting Skill Acquisition" (KNOW) and "Ways to Encourage Skill Acquisition" (ENC) has a moderate positive effect on USE with a path coefficient of 0.288, while the interaction between KNOW and BAR has a moderate negative impact on USE with a path coefficient of -0.236. These path coefficients provide valuable insights into the complex relationships between the latent variables in the context of combating unemployment among Nigerian youths through data science skill acquisition.

Table 3. Total effects of the constructs

Constructs	Total effects
BAR -> EMP	0.026
BAR -> USE	0.034
BEN -> EMP	0.492
BEN -> USE	0.632
ENC -> EMP	-0.063
ENC -> USE	-0.081
KNOW -> EMP	0.128
KNOW -> USE	0.164
USE -> EMP	0.778
KNOW x ENC -> EMP	0.224
KNOW x ENC -> USE	0.288
KNOW x BAR -> EMP	-0.184
KNOW x BAR -> USE	-0.236

Table 3 displays the total effects of the constructs in the study. The total effects represent the overall influence of each latent variable on the dependent variable "Why acquiring data science skills will solve unemployment" (EMP) and the intermediate variable "Extent of Usage of Data Science Tools" (USE). Notably, "Perceived Benefits" (BEN) has a significant positive total effect on both EMP (0.492) and USE (0.632), indicating that perceiving benefits from acquiring data science skills positively impacts the perceived ability to combat unemployment. Conversely, "Ways to Encourage Skill Acquisition" (ENC) has a negative total effect on EMP (-0.063) and USE (-0.081), suggesting that the perceived encouragement to acquire data science skills might have a somewhat adverse effect on the overall perception of solving unemployment. "Lack of Background in Relevant Subjects Affecting Skill Acquisition" (KNOW) also has a positive total effect on EMP (0.128) and USE (0.164). The interaction between KNOW and ENC shows a moderate positive total effect on both EMP (0.224) and USE (0.288), while the interaction between KNOW and BAR has a moderate negative total effect on EMP (-0.184) and USE (-0.236). These total effects shed light on the complex relationships between the latent variables and their impact on combating unemployment among Nigerian youths through data science skill acquisition.

Table 4. Outer loadings of the indicators

Indicators	Outer loadings	Indicators	Outer loadings	Indicators	Outer loadings
BAR1 <- BAR	0.745	EMP1 <- EMP	0.900	KNW1 <- KNOW	0.602
BAR2 <- BAR	0.731	EMP2 <- EMP	0.809	KNW2 <- KNOW	0.500
BAR3 <- BAR	0.742	EMP3 <- EMP	0.859	KNW3 <- KNOW	0.587
BAR4 <- BAR	0.776	EMP4 <- EMP	0.883	KNW4 <- KNOW	0.626
BAR5 <- BAR	0.864	EMP5 <- EMP	0.847	KNW5 <- KNOW	0.782
BAR6 <- BAR	0.653	EMP6 <- EMP	0.844	KNW6 <- KNOW	0.538
BAR7 <- BAR	0.696	EMP7 <- EMP	0.872	KNW7 <- KNOW	0.669
BAR8 <- BAR	0.480	EMP8 <- EMP	0.784	KNW8 <- KNOW	0.513
BAR9 <- BAR	0.728	EMP9 <- EMP	0.728	KNW9 <- KNOW	0.013
BAR10 <- BAR	0.700	EMP10 <- EMP	0.848	KNW10 <- KNOW	0.603
BEN1 <- BEN	0.852	ENC1 <- ENC	0.844	USE1 <- USE	0.091
BEN2 <- BEN	0.829	ENC2 <- ENC	0.692	USE2 <- USE	0.294
BEN3 <- BEN	0.739	ENC3 <- ENC	0.765	USE3 <- USE	0.010
BEN4 <- BEN	0.710	ENC4 <- ENC	0.875	USE4 <- USE	0.252
BEN5 <- BEN	0.733	ENC5 <- ENC	0.843	USE5 <- USE	0.797
BEN6 <- BEN	0.818	ENC6 <- ENC	0.821	USE6 <- USE	0.687
BEN7 <- BEN	0.843	ENC7 <- ENC	0.809	USE7 <- USE	0.832
BEN8 <- BEN	0.837	ENC8 <- ENC	0.837	USE8 <- USE	0.878
BEN9 <- BEN	0.742	ENC9 <- ENC	0.731	USE9 <- USE	0.851
BEN10 <- BEN	0.869	ENC10 <- ENC	0.873	USE10 <- USE	0.851
				KNOW x ENC -> KNOWxENC	1.000
				KNOW x BAR -> KNOWxBAR	1.000

Table 4 and Fig. 2 display the outer loadings of the indicators for each latent variable in the study. Outer loadings represent the strength of the relationships between the latent variables and their corresponding observed indicators. For the latent variable "Barriers to Acquisition of Data Science" (BAR), all ten indicators (BAR1 to BAR10) have high outer loadings, ranging from 0.480 to 0.864, indicating that these indicators strongly relate to the BAR construct. Similarly, for "Perceived Benefits of Acquiring Data Science Skills" (BEN), the indicators BEN1 to

BEN10 show high outer loadings, ranging from 0.710 to 0.869, indicating a robust association with the BEN construct. The same pattern is observed for the "Ways to Encourage Acquisition of Data Science Skills" (ENC) latent variable, with outer loadings of ENC1 to ENC10 ranging from 0.692 to 0.873.

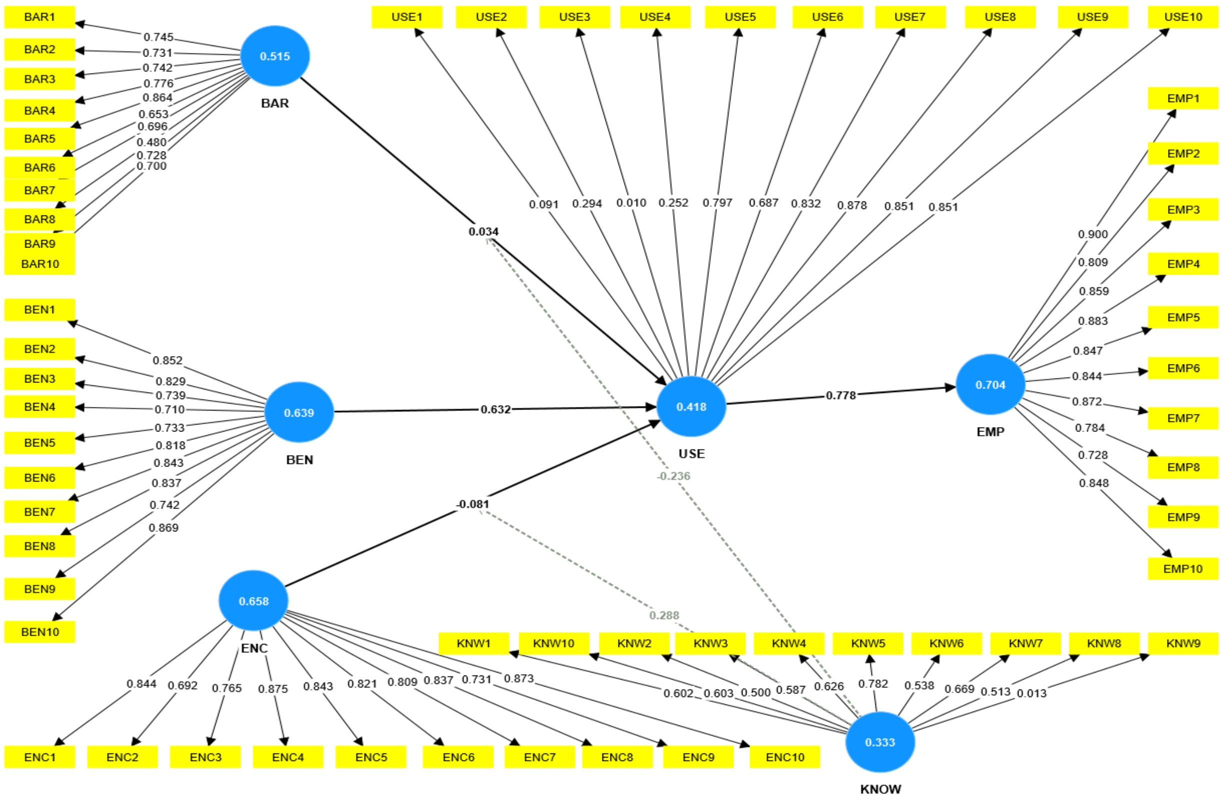

Figure 2. Graphical output showing the initial AVE of the constructs and outer loadings of the indicators

Additionally, for the "Extent of Usage of Data Science Tools" (USE) construct, the indicators USE1 to USE10 display moderate to high outer loadings, ranging from 0.010 to 0.878, suggesting a meaningful relationship with the USE variable. Lastly, the indicators representing the interaction terms, "KNOW x ENC -> KNOWxENC" and "KNOW x BAR -> KNOWxBAR," both have an outer loading of 1.000, indicating a perfect relationship with their corresponding latent variables. These outer loadings provide valuable insights into the connections between the latent constructs and their observed indicators, supporting the measurement validity of the study.

Table 5. Correlations among latent variables

	BAR	BEN	EMP	ENC	KNOW	USE	KNOW x ENC	KNOW x BAR
BAR	1.000	0.770	0.548	0.715	0.657	0.526	-0.215	-0.096
BEN	0.770	1.000	0.694	0.807	0.798	0.654	-0.342	-0.153
EMP	0.548	0.694	1.000	0.615	0.549	0.778	-0.398	-0.380
ENC	0.715	0.807	0.615	1.000	0.693	0.508	-0.349	-0.198

KNOW	0.657	0.798	0.549	0.693	1.000	0.573	-0.227	-0.049
USE	0.526	0.654	0.778	0.508	0.573	1.000	-0.149	-0.107
KNOW x ENC	-0.215	-0.342	-0.398	-0.349	-0.227	-0.149	1.000	0.846
KNOW x BAR	-0.096	-0.153	-0.380	-0.198	-0.049	-0.107	0.846	1.000

Table 5 shows the correlations among the latent variables in the study, indicating the strength and direction of their relationships: Barriers to Acquisition of Data Science (BAR), Perceived Benefits of Acquiring Data Science Skills (BEN), Why acquiring data science skills will solve unemployment (EMP), Ways to Encourage Acquisition of Data Science Skills (ENC), Lack of Background in Relevant Subjects Affecting Skill Acquisition (KNOW), Extent of Usage of Data Science Tools (USE), and the interaction terms between KNOW and ENC and between KNOW and BAR. Strong positive correlations are observed between BEN and ENC (0.807), EMP and USE (0.778), and the interaction term KNOW x ENC and KNOW x BAR (both 0.846). Additionally, moderate positive correlations are found between BAR and BEN (0.770), BEN and EMP (0.694), and ENC and KNOW (0.693). Negative correlations are observed between the interaction term KNOW x ENC and BAR (both -0.215), as well as between the interaction term KNOW x BAR and BEN (both -0.153). These correlations provide insights into the interrelationships between the latent variables, helping to understand their associations within the context of combating unemployment among Nigerian youths through data science skill acquisition.

Table 6. Construct reliability and validity

Constructs	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)
BAR	0.895	0.919	0.913	0.515
BEN	0.938	0.951	0.946	0.639
EMP	0.953	0.954	0.959	0.704
ENC	0.943	0.960	0.950	0.658
KNOW	0.771	0.840	0.816	0.333
USE	0.798	0.896	0.841	0.418

Table 6 presents the construct reliability and validity measures for the latent variables in the study: Barriers to Acquisition of Data Science (BAR), Perceived Benefits of Acquiring Data Science Skills (BEN), Why acquiring data science skills will solve unemployment (EMP), Ways to Encourage Acquisition of Data Science Skills (ENC), Lack of Background in Relevant Subjects Affecting Skill Acquisition (KNOW), and Extent of Usage of Data Science Tools (USE). The results indicate high construct reliability and validity for most constructs, as evidenced by the high values of Cronbach's alpha (ranging from 0.895 to 0.953) and composite reliability (rho_a and rho_c, both ranging from 0.919 to 0.960). Moreover, the Average Variance Extracted (AVE) values (ranging from 0.333

Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 7, Issue 4, Pages 87-104, October-December 2023

to 0.704) demonstrate that each construct accounts for a substantial amount of variance in its observed indicators, supporting the validity of the measurement model and indicating that the constructs are well-defined and distinct in the study. However, it is worth noting that the construct KNOW has relatively lower values of Cronbach's alpha (0.771) and composite reliability (rho_a and rho_c, both 0.840), indicating that the reliability of this construct might be somewhat weaker compared to the others. Nonetheless, the overall construct reliability and validity measures support the robustness and accuracy of the latent variables in the study.

Table 7. Discriminant validity - Heterotrait monotrait ratio (HTMT)

Constructs	BAR	BEN	EMP	ENC	KNOW	USE	KNOW x ENC	KNOW x BAR
BAR								
BEN	0.781							
EMP	0.567	0.690						
ENC	0.736	0.840	0.599					
KNOW	0.688	0.830	0.498	0.720				
USE	0.610	0.710	0.792	0.560	0.673			
KNOW x ENC	0.237	0.330	0.407	0.320	0.273	0.260		
KNOW x BAR	0.138	0.170	0.391	0.210	0.239	0.270	0.850	

Table 7 presents the Heterotrait Monotrait (HTMT) ratio values, which are used to assess the discriminant validity of the latent variables in the study: Barriers to Acquisition of Data Science (BAR), Perceived Benefits of Acquiring Data Science Skills (BEN), Why acquiring data science skills will solve unemployment (EMP), Ways to Encourage Acquisition of Data Science Skills (ENC), Lack of Background in Relevant Subjects Affecting Skill Acquisition (KNOW), Extent of Usage of Data Science Tools (USE), and the interaction terms between KNOW and ENC and between KNOW and BAR. The HTMT values are calculated for each pair of constructs, and they measure the extent to which the constructs share more variance with their own indicators (monotrait) than with the indicators of other constructs (heterotrait). For the most part, the HTMT values are below the threshold of 0.85, indicating good discriminant validity, except for the interaction term KNOW x BAR, which has a HTMT value of 0.850, suggesting a potential issue with discriminant validity between this interaction term and the BAR construct. Overall, the HTMT values support the discriminant validity of the majority of the latent variables in the study, but the potential lack of discriminant validity between the interaction term KNOW x BAR and the BAR construct should be further investigated and addressed.

Table 8. Discriminant validity - Fornell–Larcker criterion

Constructs	BAR	BEN	EMP	ENC	KNOW	USE
BAR	0.718					
BEN	0.770	0.799				
EMP	0.548	0.694	0.839			
ENC	0.715	0.807	0.615	0.810		
KNOW	0.657	0.798	0.549	0.690	0.577	
USE	0.526	0.654	0.778	0.51	0.573	0.646

Table 8 shows the Fornell–Larcker criterion results, which are used to assess the discriminant validity of the latent variables in the study: Barriers to Acquisition of Data Science (BAR), Perceived Benefits of Acquiring Data Science Skills (BEN), Why acquiring data science skills will solve unemployment (EMP), Ways to Encourage Acquisition of Data Science Skills (ENC), Lack of Background in Relevant Subjects Affecting Skill Acquisition (KNOW), and Extent of Usage of Data Science Tools (USE). The Fornell–Larcker criterion compares the square of the correlations between each construct and its indicators with the average variance extracted (AVE) for each construct. If the squared correlation is higher than the AVE, it indicates discriminant validity. Overall, the diagonal values (the squared correlations between the constructs and themselves) are higher than the corresponding AVE values for each construct, suggesting that the Fornell–Larcker criterion supports the discriminant validity of the latent variables in the study. The values on the diagonal (reflecting the constructs' correlations with themselves) should always be higher than the off-diagonal values (reflecting the correlations between constructs), indicating that the constructs have more shared variance with their own indicators than with other constructs' indicators, thus supporting their distinctiveness and discriminant validity.

Table 9. Collinearity statistics for inner model

Constructs	VIF
BAR -> USE	2.680
BEN -> USE	5.070
ENC -> USE	3.181
KNOW -> USE	2.890
USE -> EMP	1.000
KNOW x ENC -> USE	4.265
KNOW x BAR -> USE	3.858

Table 9 presents the collinearity statistics, specifically the Variance Inflation Factor (VIF), for the inner model of the study. The VIF measures the degree of multicollinearity between the independent variables (constructs) in the model. In general, a VIF value of 1 indicates no multicollinearity (perfectly uncorrelated variables), while higher values indicate increasing levels of multicollinearity. It is essential to address multicollinearity in regression models because high VIF values can lead to unstable coefficient estimates and decreased statistical power. In this table, the VIF values for the relationships between the latent variables and the dependent variable "Why acquiring data science skills will solve unemployment" (EMP) are relatively low, ranging from 1.000 to 5.070. VIF values below 10 are generally considered acceptable, suggesting that there is no severe multicollinearity issue among the predictors in the inner model. However, the VIF value of 5.070 for "Perceived Benefits of Acquiring Data Science Skills" (BEN) indicates some moderate multicollinearity with the "Extent of Usage of Data Science Tools" (USE).

5. Summary of the Study and findings

This study investigates the relationship between acquiring data science skills and combating unemployment among Nigerian youths. It focuses on six latent variables: Barriers to Acquisition of Data Science, Perceived Benefits of Acquiring Data Science Skills, Why acquiring data science skills will solve unemployment, Ways to Encourage Acquisition of Data Science Skills, Lack of Background in Relevant Subjects Affecting Skill Acquisition, and Extent of Usage of Data Science Tools. The study collected data and analyzed it using various statistical methods.

The latent variables "Perceived Benefits," "Ways to Encourage Acquisition," and "Extent of Usage of Data Science Tools" have strong positive relationships with "Why acquiring data science skills will solve unemployment," indicating that perceiving benefits, encouraging skill acquisition, and extensive usage of data science tools positively impact the perception of combating unemployment. "Lack of Background in Relevant Subjects Affecting Skill Acquisition" also positively influences "Why acquiring data science skills will solve unemployment," suggesting that addressing knowledge gaps is essential for solving unemployment. "Barriers to Acquisition of Data Science" and "Ways to Encourage Acquisition" have weak relationships with the extent of usage of data science tools, indicating that overcoming barriers and providing encouragement may not directly influence skill usage. The measurement model demonstrates good construct validity, as evidenced by high outer loadings, Cronbach's alpha, and composite reliability values. However, the construct "Lack of Background in Relevant Subjects Affecting Skill Acquisition" shows comparatively lower reliability, suggesting potential areas for improvement in its measurement.

The Fornell–Larcker criterion supports discriminant validity among most of the latent variables, indicating that they are distinct and well-defined constructs. However, the interaction term "KNOW x BAR" shows potential issues with discriminant validity and requires further investigation.

6. Conclusion and Recommendation

The findings of this study highlight the importance of acquiring data science skills in combating unemployment among Nigerian youths. Perceived benefits, encouragement, and extensive usage of data science tools positively influence the perception of addressing unemployment. Additionally, addressing knowledge gaps is crucial for successful skill acquisition. However, barriers to acquisition and some interactions between constructs require careful consideration. Overall, the study provides valuable insights for policymakers, educators, and stakeholders interested in promoting data science skill acquisition to combat youth unemployment in Nigeria. Further research should be conducted to address potential collinearity issues and investigate the causal relationships between these constructs more deeply.

Based on the findings of the study, a recommendation would be to develop and implement targeted programs and initiatives that focus on promoting data science skill acquisition among Nigerian youths. These programs should emphasize the perceived benefits of acquiring these skills, provide encouragement and support to overcome barriers, and offer hands-on opportunities to use data science tools effectively. Additionally, it is essential to address the knowledge gaps by providing relevant educational resources and training in statistics, mathematics, and computer science. By doing so, policymakers and stakeholders can enhance the employability of Nigerian youths and empower them with valuable skills that align with the demands of the modern job market, ultimately contributing to the reduction of unemployment rates in the country.

Declarations

Source of Funding

This study has not received any funds from any organization.

Conflict of Interest

The authors declare that they have no conflict of interest.

Consent for Publication

The authors declare that they consented to the publication of this study.

Authors' Contribution

Both the authors took part in literature review, analysis, and manuscript writing equally.

Acknowledgement

The authors appreciate the Tertiary Education Trust Fund's (TET Fund) Institutional Based Research (IBR) Intervention line for the sponsorship of this research and the Committee of Research (Centre for Research and Innovative Development, CRID) of the Federal Polytechnic, Ado-Ekiti for the approval.

References

Aljohani, N.R., Aslam, A., Khadidos, A.O., & Hassan, S.U. (2022). Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature. Journal of Innovation and Knowledge, 7(3). doi: https://doi.org/10.1016/j.jik.2022.100190.

Becker, J.M., Cheah, J.H., Gholamzade, R., Ringle, C.M., & Sarstedt, M. (2023). PLS-SEM's most wanted guidance. International Journal of Contemporary Hospitality Management, 35(1): 321–346. doi: https://doi.org/10.1108/IJCHM-04-2022-0474.

Deryabin, A.A., & Glukhov, P.P. (2022). Regional and Urban Data Science Projects for Citizen and Youth Engagement. doi: https://doi.org/10.15405/epes.22043.4.

Donovan, S., Maggiulli, L., Aiello, J., Centeno, P., John, S., & Pisani, A. (2023). Evaluation of sustainable, blended learning workforce education for suicide prevention in youth services. Children and Youth Services Review, 148. doi: https://doi.org/10.1016/j.childyouth.2023.106852.

Harris, E.M., Dixon, C.G.H., Bird, E.B., & Ballard, H.L. (2020). For Science and Self: Youth Interactions with Data in Community and Citizen Science. Journal of the Learning Sciences, 29(2). doi: https://doi.org/10.1080/10508406.2019.1693379.

Kwong, K., & Wong, K. (2015). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. http://www.researchgate.net/publication/268449353.

Law, L. (2020). Applying partial least squares structural equation modeling (PLS-SEM) in an investigation of undergraduate students' learning transfer of academic English Creativity: A Corpus Linguistic Systemic Functional Multimodal Discourse Analysis Approach View project Corpus of Hotel Press Release (CorHPR) (2014–Present). doi: https://doi.org/10.13140/RG.2.2.15180.64641/1.

Looman, W.S., Eull, D., Geiger-Simpson, E., Lee, J., Nersesian, P. v., Bell, A.N., & Miller, A.L. (2023). Assessment of Youth Coping and Appraisals Through a Timelining Activity: Validation and Utility of a Tool for Pediatric Research and Practice. Journal of Pediatric Health Care. doi: https://doi.org/10.1016/j.pedhc.2023.02.007.

Mickienė, R., & Valionienė, E. (2021). Modelling the effectiveness index of digital marketing strategy oriented to increase the popularity of maritime education. Trans Nav., 15(3). doi: https://doi.org/10.12716/1001.15.03.08.

Moon, P.F., Israel-Fishelson, R., Tabak, R., & Weintrop, D. (2023). The Tools Being Used to Introduce Youth to Data Science. Proceedings of IDC 2023–22nd Annual ACM Interaction Design and Children Conference: Rediscovering Childhood. doi: https://doi.org/10.1145/3585088.3589363.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ringle, C.M., Wende, S., & Becker, J. (2022). SmartPLS 4. Oststeinbek: SmartPLS.

Sánchez-Marco, M., Escribano, S., Rubio-Aparicio, M., Juliá-Sanchis, R., & Cabañero-Martínez, M.J. (2023). Effectiveness of nontechnical skills educational interventions in the context of emergencies: A systematic review and meta-analysis. In Australian Critical Care, Elsevier Ireland Ltd., doi: https://doi.org/10.1016/j.aucc.2023.01.007.

Schultheiss, T., Pfister, C., Gnehm, A.S., & Backes-Gellner, U. (2023). Education expansion and high-skill job opportunities for workers: Does a rising tide lift all boats? Labour Economics, 82. doi: https://doi.org/10.1016/j.labeco.2023.102354.

Smaldone, F., Ippolito, A., Lagger, J., & Pellicano, M. (2022). Employability skills: Profiling data scientists in the digital labour market. European Management Journal, 40(5): 671–684. doi: https://doi.org/10.1016/j.emj.2022.05.005.

Yahaya, M.L. (2019). Partial Least Square Structural Equation Modeling (PLS-SEM): *A Note for Beginners* Transaction cost analysis of the Nigerian construction Industry in the context of Public Procurement Act (2007). http://www.casirmediapublishing.com.

Xu, Z., Tang, N., Xu, C., & Cheng, X. (2021). Data science: connotation, methods, technologies, and development. Data Science and Management, 1(1). doi: https://doi.org/10.1016/j.dsm.2021.02.002.