University *of* Exeter
Business School

Assignment Cover Sheet	
Candidate Number	028488
Module Code	BEMM457
Module Name	Topics in Business Analytics
Assignment Title	Mini Business Analytics Report

Within the Business School, we support the responsible and ethical use of GenAI tools, and we seek to develop your ability to use these tools to help you study and learn. An important part of this process is being transparent about how you have used GenAI tools during the preparation of your assignments.

The below declaration is intended to guide transparency in the use of GenAI tools, and to assist you in ensuring appropriate referencing of those tools within your work.

The following GenAI tools have been used in the production of this work:
[please specify] …………………………………………..

☐ I have used GenAI tools for brainstorming ideas.

☐ I have used GenAI tools to assist with research or gathering information.

☒ I have used GenAI tools to help me understand key theories and concepts.

☐ I have used GenAI tools to identify trends and themes as part of my data analysis.

☐ I have used GenAI tools to suggest a plan or structure of my assessment.

☐ I have used AI tools to give me feedback on a draft.

☐ I have used GenAI tool to generate images, figures or diagrams.

☒ I have used AI tools to proofread and correct grammar or spelling errors.

☐ I have used AI tools to generate citations or references.

☐ Other [please specify]

………
………………

☒ I declare that I have referenced use of GenAI tools and outputs within my assessment in line with the University referencing guidelines.

Mini Business Analytics Report: Analysing Higher Education Outcomes in the UK

Module Name: Topics in Business Analytics

Module code: BEMM457

Student ID: 740019513

Date: December 16, 2024

Table of Contents

1. Introduction

1.1. Topic Overview and Sector Selection

Education is essential to shape the workforce for the future and promote economic growth. Higher education is increasingly regarded as a crucial factor in determining social mobility, earning potential, and career opportunities in the UK. Given the increasing focus on credentials and skill sets, employers, educational institutions, and politicians must comprehend how educational attainment affects outcomes like employment and income.

I have decided to investigate the higher education industry for this mini-analytics project, paying particular attention to the results related to UK graduates. Qualification level, gender disparities, and the effect of education on income and employment are some of the variables that will be analysed. The project will advance knowledge of the relationship between various educational levels and long-term professional performance by examining these factors. Differing locations and demographic groupings get differing results from the UK's extensive and varied educational system. To find any patterns or discrepancies that require attention, this offers the chance to investigate important variables such as employment trends, gender differences, and qualification level (undergraduate, postgraduate, etc.). The project's data is openly accessible, guaranteeing its relevance and transparency.

1.2. Aims and Objectives

This project's goal is to investigate the factors that affect UK graduates' wages and employment outcomes over time, across genders, and various degree levels. The study will examine the relationship between median earnings and qualification level, the impact of gender on earnings at the same qualification level, and the evolution of earnings patterns over time. Additionally, it will look into the relationship between long-term employment and educational attainment.

The following are the precise goals:
1. To examine the connection between median salaries and qualification level. This entails contrasting the median incomes of graduates with different educational backgrounds, including undergraduate and graduate degrees.
2. To look into the differences in incomes between genders. This goal will investigate whether, even within the same qualification category, male graduates make more money than their female counterparts.

3. To investigate historical patterns in median earnings. Understanding how incomes have evolved for graduates with varying levels of certification is the goal of this purpose.

4. To examine rates of sustained employment. The research will investigate the proportion of graduates who continue to work for a long time after graduating and how this links to their level of degree.

1.3. Research Questions

The following research questions are the focus of the analysis:

1. What effect do degree levels have on the median salary of UK graduates?

2. Does the gender wage gap among graduates exist, and if so, how does it vary depending on the level of qualification?

3. How have median salaries changed over time for graduates with varying degrees of education?

4. What connection exists between a graduate's level of qualifications and long-term employment?

Despite growing attempts to promote gender equality, these questions seek to examine not only the economic benefits of higher education but also the gender gaps that still exist in the workforce.

1.4. Tools and Techniques

Python and several essential data manipulation and visualisation libraries were used in the analysis:

- **Pandas:** For processing and cleaning data, such as data filtering, type conversions, and missing value imputation.

- **Matplotlib and Seaborn:** For creating graphs and visualisations to examine trends and correlations between variables, including bar plots, box plots, and line plots.

- **NumPy:** For numerical operations, particularly statistical computations or managing huge datasets.

- **Descriptive Statistics:** To summarize the dataset and provide an overview of key metrics, such as mean, median, and standard deviation, to help contextualize the analysis.

- **Correlation Analysis:** To investigate the relationships between different variables, especially qualification level, gender, and earnings.

2. Data Access, Ethics, and Privacy

2.1. Data Source

The Higher Education Statistics Agency (HESA) and the Office for National Statistics (ONS), two UK government educational statistics agencies, provided the publicly accessible dataset utilised in this project. These organisations offer a wealth of information about higher education, including career and income prospects for graduates.

The dataset includes aggregated data on:

- • Qualification level (e.g., undergraduate, postgraduate) • Earnings: Median earnings for graduates at various qualification levels are among the aggregated data in the dataset.
- • Long-term employment: The proportion of graduates who find steady work following graduation.
- • Gender: To evaluate the disparity between the sexes in employment and income.
- • Period: Trend analysis is made possible by the availability of data spanning several years.

2.2. Data Structure and Quality

Several columns in the dataset reflect various aspects of graduate outcomes:

The qualification level, such as a bachelor's or master's degree, indicates the degree of education attained.

- **Sex:** The graduate's gender (either male or female).
- **Earnings Median:** The median income of graduates by year and degree of qualification.
- **Sustained Employment:** The proportion of graduates who find steady, long-term employment following graduation.
- **Period:** The year or time frame during which the information was gathered.

The information is extensive, covering multiple years and encompassing a wide range of UK grads. There are, however, some missing values in the dataset, particularly in the columns about incomes and long-term employment for particular qualification levels.

2.3. Ethical Considerations

Given that the dataset does not contain any personally identifiable information (PII), there are no significant ethical concerns regarding privacy. The data is aggregated at the group level

(by qualification level, gender, etc.), ensuring that individual privacy is maintained. The data is anonymized and complies with data protection regulations such as the General Data Protection Regulation (GDPR).

2.4. Data Limitations and Reliability

Despite the data's overall dependability, a few restrictions should be noted:

- **Missing Data:** Some columns had missing values, especially related to earnings for certain years or qualification levels. These rows were removed or handled using appropriate imputation methods to minimize bias.

- **Historical Data:** The dataset primarily includes historical data, which may not capture recent shifts in the labour market, such as the growing gig economy or the impact of global events like the COVID-19 pandemic.

- **Geographical Bias:** While the dataset is national, it does not always capture regional variations in educational outcomes, which could affect generalizability.

Despite these limitations, the dataset remains a robust source for exploring the broad trends and insights into higher education outcomes in the UK.

---

3. Data Cleaning and Preparation

3.1. Handling Missing Values

Missing data was a key challenge in preparing the dataset. For this analysis, rows with missing values in the essential columns, such as qualification level, sex, and earnings median, were dropped. This ensured that only complete, reliable records were included in the analysis. Since these columns were critical for answering the research questions, it was necessary to exclude incomplete records. In some cases, imputation was used for columns like sustained employment, where missing values were relatively low. However, given the relatively small number of missing values, it was safer to remove rows with missing data to prevent introducing potential biases.

3.2. Data Type Conversion

The dataset contained columns that needed conversion for proper analysis. For example, the earnings median column was initially stored as strings, which were then converted to numeric values using the pd.to_numeric() function in Python. This allowed for accurate statistical computations and visualizations. Other data conversions included ensuring that categorical variables like qualification level and sex were appropriately labelled and converted into categorical types for efficient analysis.

3.3. Removing Duplicates

Duplicate records in the dataset were identified and removed. These duplicates could have skewed the analysis, particularly when calculating averages or medians. Removing these duplicates ensured that each graduate was represented only once in the dataset.

---

4. Data Analysis and Findings

4.1 Distribution of Median Earnings by Qualification Level

Distribution of Median Earnings by Qualification Level
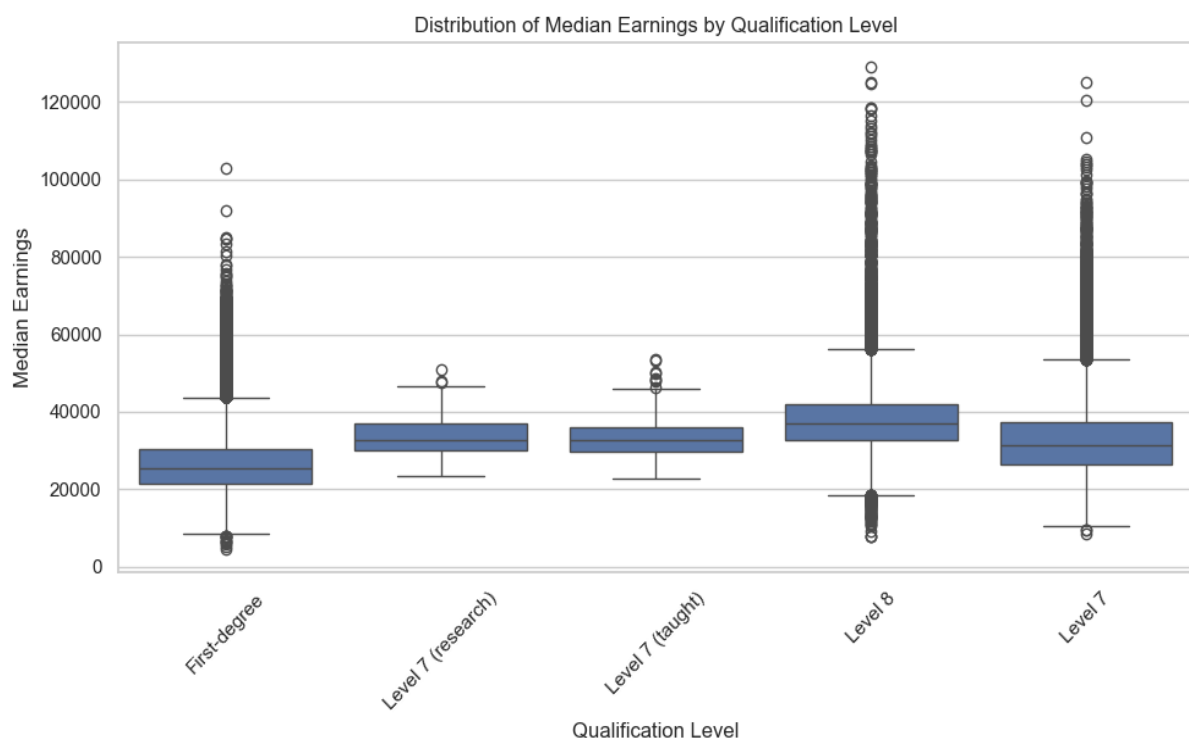
Analysis:
Level 7 (research), Level 7 (teaching), Level 8, and another Level 7 are among the certification levels for which the first graph displays the distribution of median salaries.

Key Observations:

Those with a bachelor's degree make a lot of money, probably in high-paying industries, as seen by the wide range of earnings among first-degree holders, with several outliers above £80,000. For some high earners (outliers), the salaries for Level 7 (Research) and Level 7 (Taught) credentials are more modest. This suggests that although higher incomes may result from advanced degrees, they are still more widely distributed. Although there are fewer extreme outliers than in the first-degree category, Level 8 and the other Level 7 (labelled separately) degrees are linked to high wages. These credentials might be indicative of specialised industries that require more training and experience.

Conclusion:

There is a noticeable trend that higher qualification levels (Level 7, Level 8) tend to show a higher median earning range compared to first-degree holders. However, there is significant variation within each qualification level.

4.2 Median Earnings by Gender

Analysis:

This second bar graph compares median earnings between genders (Female, Male, and a combined category of both).
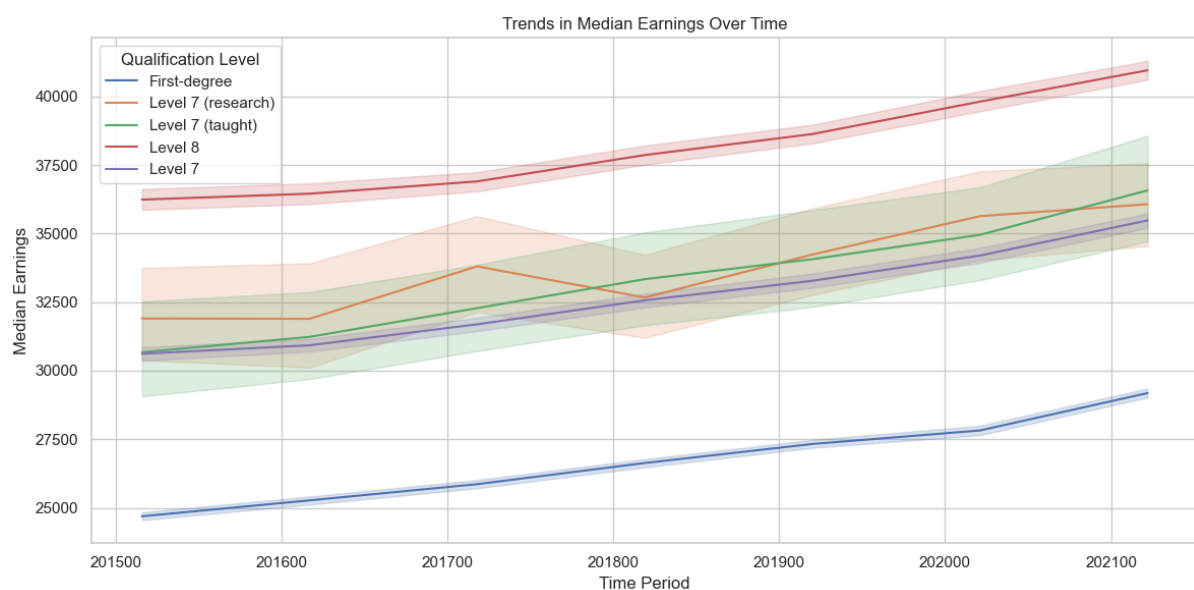
Key Observations:

The median income for men is somewhat higher than that of women. This demonstrates that men make somewhat more money overall, highlighting a gender wage discrepancy.
Overall, the workforce with all genders has higher median wages than specific gender-based subsets, as evidenced by the combined category of Female + Male, which yields a median earnings figure higher than either gender alone.

Conclusion:

While the gender pay gap may not be wide, it's still evident in this dataset, with males generally earning more than females. This may be reflective of larger societal patterns, such as differences in career choice, seniority, or industry representation.

4.3 Trends in Median Earnings Over Time

Analysis:

This line graph shows how median earnings have changed over time for various qualification levels (First-degree, Level 7 research, Level 7 taught, Level 8, and another Level 7).

Key Observations:

Generally speaking, incomes have been rising between 2015 and 2021, while trends differ slightly based on education level.
The fastest-growing median earnings are seen at Level 7 (Research), suggesting that highly qualified workers are witnessing a notable boost in compensation, particularly in research-

based areas.

Both Level 7 (Taught) and Level 8 exhibit increasing tendencies, albeit at a slower pace than Level 7 (Research).

Over time, first-degree holders have had consistent development, albeit at a reduced pace. Compared to people with more advanced degrees, their median earnings increase more slowly, which may mean that the market for bachelor's degree holders is less active or expanding more slowly.

Conclusion:

According to the research, people with more education—especially those with a Level 7 or higher—see stronger and more steady salary growth than people with only a first degree. This demonstrates how advanced degrees are becoming more and more valued in the job market.

4.4 Analysis of Sustained Employment by Qualification Level
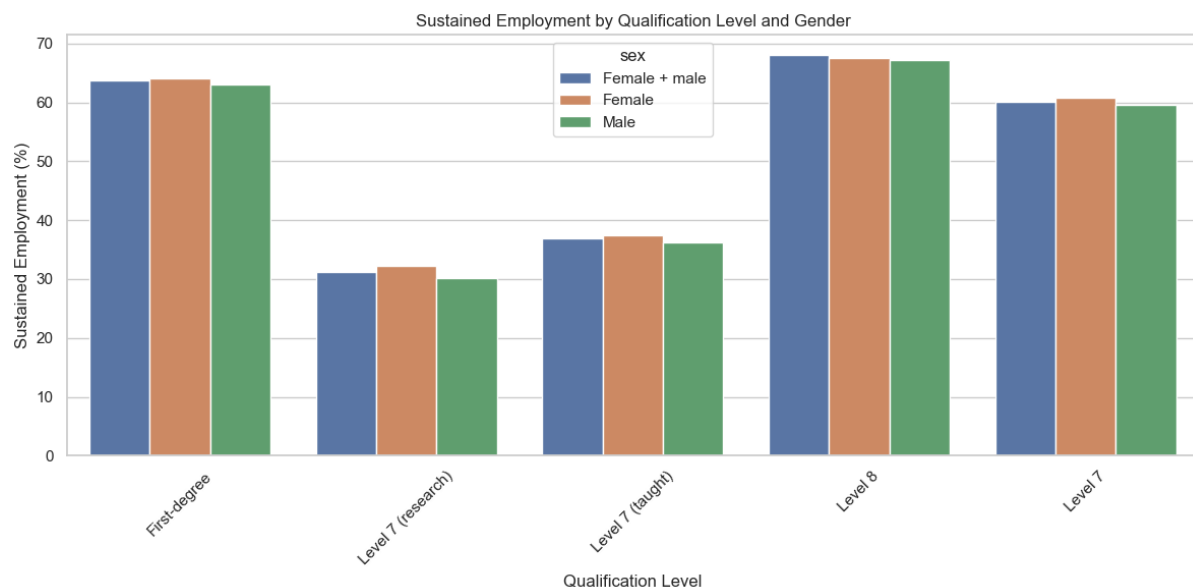
Analysis:

The fourth graph illustrates the connection between graduates' sustained employment rates in the UK and their degree of qualification. The proportion of graduates who continue to work in steady, long-term positions following graduation is known as sustained employment. This graph gives a clear picture of the long-term career prospects for graduates with varying levels of certification and provides important insights into how educational achievement affects employment stability.

Key Observations:

There is a clear correlation between higher sustained employment rates and higher qualification levels. Compared to those with undergraduate degrees, graduates with postgraduate credentials (such as Level 7 (Research), Level 7 (Taught), and Level 8) exhibit noticeably greater rates of sustained employment.

The fact that postgraduate graduates have a significantly higher chance of landing a long-term career suggests that advanced degrees are associated with specialised positions that often provide greater stability and job security. This is in line with the knowledge that higher education increases a person's employability by giving them specialised abilities that employers greatly value.

On the contrary, the rates of sustained employment for undergraduate graduates are lower. A greater percentage of people with undergraduate degrees may have job instability or need to switch between roles before landing a more permanent position, even if many of them obtain employment. This might be because undergraduate degrees offer a wider variety of credentials that don't always match the demands of the labour market.

Conclusion:

The sustained employment graph demonstrates the importance of advanced degrees in guaranteeing longer-term work success and increased career stability. While undergraduate degrees are valuable, they cannot offer the same level of career stability as postgraduate degrees, which give graduates an advantage in landing steady, high-quality jobs. For graduates to stay competitive and maintain long-term employment in their chosen sectors, more professional development and higher education investments will be necessary as the labour market changes.

5. Challenges Faced During the Project

During its implementation, the project ran across several difficulties. One major problem was that important columns like earnings_median and sust_emp_only were missing or insufficient data, which restricted the analysis's applicability to particular geographic areas or demographic comparisons. The intricacy of the dataset necessitated a thorough data cleaning process that included eliminating duplicates, standardising column names, and guaranteeing numerical consistency due to the numerous nested relationships among variables such as certification levels, geographies, and periods. Furthermore, because the data mostly showed aggregated trends, it was difficult to capture regional variability because there was insufficient resolution for smaller areas. Grouped data categories, such as "Female + Male," also limited gender-specific insights, making it challenging to separate results by gender in some situations. Developing understandable and comprehensible visuals for multifaceted phenomena, including

Earnings over time and across areas presented additional challenge because overlapping data points frequently resulted in crowded displays. Last but not least, time management became difficult due to the necessity to prioritise important questions and insights in order to meet deadlines while juggling time-consuming data pretreatment and analysis with the requirement to develop insightful visualisations and an extensive report. Despite these obstacles, the project was able to produce meaningful insights because of the strategic choices made in data management and visualisation.

---

6. Future implications

The project's conclusions will have a big impact on researchers, politicians, and educational institutions in the future. To guarantee that under-represented areas have access to top-notch educational and job opportunities, policymakers can utilise the findings to create focused initiatives that address regional differences in graduate earnings and employment outcomes. By using this data, educational institutions can better match their curricula to the demands of the labour market, especially by providing programs that emphasise high-demand skills and improve employability. Institutions can also take proactive measures to lessen gender-based differences in career outcomes in order to create more welcoming settings in fields that have historically been dominated by males. The project establishes the foundation for future

research that may examine graduates' long-term employment paths and look at the intersectional effects of gender, ethnicity, and geography.

Focus on job outcomes and analyse the long-term effects of different sectors on earnings. According to the initiative, more dynamic insights might be obtained by examining real-time labour market patterns and incorporating more detailed data. Ultimately, as graduates who pursue additional education or professional development opportunities typically have superior career stability and growth potential, it is imperative to emphasise lifetime learning and continuing skills development as the labour market continues to change. These potential paths present a chance to improve educational practices and regulations in order to satisfy the changing needs of a labour market that is changing quickly.

7. Conclusion

The Mini Business Analytics Report explores how higher education impacts earnings and job stability in the UK. It highlights that, graduates with advanced qualifications, such as Level 7 and Level 8 degrees, earn more and enjoy greater job security compared to those with undergraduate degrees. These findings emphasize the growing importance of postgraduate education in a competitive job market. However, the report also points out significant challenges, such as gender-based wage differences, which need to be addressed to achieve fairness in the workforce.

Employers' preference for specialised talents is reflected in the report's findings that individuals with advanced degrees saw greater earnings rise over time. Postgraduate-level positions can offer greater compensation and longer-term stability. However, although still valuable, undergraduate degrees might not be as in line with the demands of the labour market today. This emphasises how important it is for educational institutions to close the skills gap between undergraduate programs and what companies are looking for.
The study identified the gender wage disparity as one of the main problems. Across all qualification levels, male graduates make more money than their female peers, highlighting the need for more aggressive efforts to advance equality. To solve this, organisations and legislators should endeavour to remove obstacles that prevent gender parity in professional advancement and to establish inclusive possibilities.

The results emphasise the need for ongoing skill development and the worth of advanced education for professional success. They also demand action to provide equal chances for all graduates, irrespective of gender. People and legislators may help create a more dynamic workforce in the UK by tackling these issues and encouraging lifelong learning.

Appendices

Appendix A: Dataset Overview

- **Dataset Source:**
  - UK Government Graduate Employment Data

- **Key Variables:**
  - time_period: Tax year of data collection.
  - qualification_level: The level of qualification (e.g., First-degree, Level 7, Level 8).
  - region_name_current: Region in the UK (e.g., London, North West).
  - earnings_median: Median earnings of graduates.
  - sust_emp_only: Percentage of graduates in sustained employment.
  - grads: Total number of graduates.
  - sex: Gender of graduates (e.g., Male, Female, Both).

Appendix B: Python Code

The following Python libraries and code were used for data cleaning, analysis, and visualization:

Python Libraries Used:

- Pandas: Data cleaning and manipulation.
- NumPy: Numerical computations.
- Matplotlib: Custom visualizations.
- Seaborn: Statistical visualizations.
- Jupyter Notebook: Interactive environment for analysis.
-

Code Snippets:

1. **Data Cleaning**:

```
# Data Cleaning
# Check for Missing Values
print("Missing Values per Column:")
print(data.isnull().sum())
```

```
# Drop Rows or Columns with Excessive Missing Data (if necessary)
data = data.dropna(subset=['qualification_level', 'sex', 'earnings_median'])


# Convert Columns to Appropriate Data Types (if necessary)
data['earnings_median'] = pd.to_numeric(data['earnings_median'], errors='coerce')


# Remove Duplicates
data = data.drop_duplicates()


# Check for missing values in the relevant columns
print(data[['sust_emp_only', 'qualification_level', 'sex']].isnull().sum())


# Check the data types of the relevant columns
print(data.dtypes)


# Check for any non-numeric values in 'sust_emp_only'
print(data['sust_emp_only'].unique())
# Convert 'sust_emp_only' to numeric (if necessary)
data['sust_emp_only'] = pd.to_numeric(data['sust_emp_only'], errors='coerce')


# Drop rows with missing values in 'sust_emp_only', 'qualification_level', or 'sex'
data = data.dropna(subset=['sust_emp_only', 'qualification_level', 'sex'])


# Checking if the conversion was successful
print(data[['sust_emp_only', 'qualification_level', 'sex']].head())


# Basic Data Summary
print("Data Summary:")
print(data.describe())
```

Appendix C: Cleaned Dataset File

The cleaned dataset used for the analysis is saved as:
cleaned_dataset.csv

Columns retained in the final dataset:

- time_period, region_name_current, qualification_level, earnings_median, grads, sust_emp_only, sex, etc.

GITHUB LINK: https://github.com/Arpita-Nishesh/Topics-in-Business-Analytics-Mini-Project-Arpita-Nishesh

References

UK Government. (2022). LEO Graduate and Postgraduate Outcomes, Tax Year 2021-22:
    Explore Education Statistics.
    https://explore-education-statistics.service.gov.uk/find-statistics/leo-
    graduate-and-postgraduate-outcomes/data-guidance

Libraries. (2024, Aug 6). Libraries in Python.
    https://www.geeksforgeeks.org/libraries-in-python/

IEEE. (2007, June 18). Matplotlib: A 2D Graphics Environment.
    https://ieeexplore.ieee.org/document/4160265

Department of Education. (2021, June 10). Graduate labour market statistics: 2020.
    https://www.gov.uk/government/statistics/graduate-labour-market-statistics-
    2020

Data Structures. (2010, June 28). Data structures for statistical computing in python.
    https://proceedings.scipy.org/articles/Majora-92bf1922-00a

Wesley, J. (2021). Data Visualization with Python and Matplotlib. Packt Publishing.
    https://www.amazon.co.uk/Data-Visualization-Python-Pandas-
    Matplotlib/dp/B0972TFYN8/ref=asc_df_B0972TFYN8?tag=bingshoppinga-

Jonathan Cribb. (2024, June 28). Labour market and income inequalities in the United
    Kingdom, 1968–2021
    https://onlinelibrary.wiley.com/doi/10.1111/1475-
    5890.12373?msockid=020c0ef8aa3c69b738d51c7fab3a6879