



A PROJECT REPORT ON DATA ANALYTICS USING PYTHON

Submitted in project file the degree of MASTER OF COMPUTER APPLICATION (MCA)

Submitted by :- Neelanshi Singh

|  |  |  |  |
| --- | --- | --- | --- |
| **Created By:** | Neelanshi Singh | **Approved By:** | <Data analytics> |
| **Created On:** | 15/06/2025 | **Approved On:** |  |

**INDEX**

[**1** **PROJECT DETAILS** 2](#_Toc143445375)

[**2** **SUMMARY** 2](#_Toc143445376)

[**3** **INTRODUCTION** 2](#_Toc143445377)

[3.1 Background 2](#_Toc143445378)

[3.2 Stakeholders 2](#_Toc143445379)

[3.3 Objectives 2](#_Toc143445380)

[**4** **METHODOLOGY** 2](#_Toc143445381)

[4.1 Considerations & Assumption 3](#_Toc143445382)

[4.2 Approach 3](#_Toc143445383)

[4.3 Activities 3](#_Toc143445384)

[**5** **TARGETTED V/S ACHIEVED OUTPUT** 3](#_Toc143445385)

[**6** **CONCLUSION** 3](#_Toc143445386)

[**7** **APPENDICES** 4](#_Toc143445387)

[7.1 Appendix A – Title 4](#_Toc143445388)

**General Instructions for using the Live Project Report Template**

* This template and the subsequent document created using this template is a confidential document and is the intellectual property of Cloud Counselage Pvt. Ltd. Circulating it outside of the organisation without the consent of Cloud Counselage Pvt. Ltd. is the breach of company policies and will lead to legal actions
* This template is a guideline document to communicate the implementation of design ideas and the results of the work to the stakeholders.
* The **text between inequality (< >) is to be replaced** by relevant text
* Please **remove the yellow highlight on the Text** between the inequality (< >). This is done to help you notice the text to be changed/replaced
* The text in *italics* highlighted in grey is just for reference and should be removed after adding the relevant text

# **PROJECT DETAILS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Name** | Neelanshi Singh | | |
| **Project Sponsor** | Tushar Topale | | |
| **Project Manager** | Harshada Topale | | |
| **Start Date** | 15/06/25 | **Completion Date** | 21/06/25 |

# **SUMMARY**

The data analytics project aimed to conduct a comprehensive analysis of student interns to gain insights into the relationship between students' economic background, academic performance, competence, and expected salary. Leveraging the provided dataset containing attributes such as GPA, family income, Python programming experience, and leadership skills, we applied analytical techniques to extract meaningful insights. Our findings reveal key patterns in how economic background and academic performance influence expected salary, shedding light on factors affecting student success. Despite encountering potential limitations in data completeness, the project succeeded in identifying these relationships. Moving forward, recommendations include further research into additional influencing factors and enhanced data collection methods.

# **INTRODUCTION**

## Background

*Millions of students apply for internships/jobs every year, resumes play an important role in playing*

*the first impression. The recruiters spend a max of 2-3 minutes reviewing a resume after it landed in*

*their mailbox or Job board, ATS application. Surprising more than 70% of resumes get rejected in the*

*initial screening.*

This data analytics project is designed to tackle the critical gap in understanding how students' economic background, academic performance, competence, and expected salary interrelate, utilizing advanced data-driven techniques to reveal actionable insights. By examining a rich dataset that includes metrics like GPA, family income, Python programming experience, and leadership skills, this initiative aims to illuminate hidden patterns and correlations that can guide strategic decisions.

In an increasingly competitive environment, the power of data analytics to inform and innovate is essential. This project seeks to showcase its potential within the education and career development sector, positioning Cloud Counselage and CC-GAC Foundation as leaders in leveraging data to foster student success and career readiness.

## Stakeholders

Stakeholders play crucial roles in any data analytics project. Here's a breakdown of key stakeholders typically involved:

1. Business Leaders/Executives: Business leaders and executives are often the primary stakeholders in data analytics projects. They provide strategic direction, define project objectives, and make decisions based on the insights generated from the analysis.

2. Data Analysts/Data Scientists: Data analysts and data scientists are responsible for conducting the analysis, applying statistical techniques and machine learning algorithms to extract insights from data. They play a pivotal role in transforming raw data into actionable insights.

3.Data Engineers/Data Architects: Data engineers and data architects are involved in the design and implementation of data infrastructure, ensuring that the necessary data pipelines, databases, and data warehouses are in place to support the analytics process.

4. IT Department: The IT department is responsible for providing technical support, managing data security, and ensuring the smooth functioning of IT systems and infrastructure required for the data analytics project.

5. Business Analysts: Business analysts work closely with both business stakeholders and data analysts to understand business requirements, translate them into analytical solutions, and validate the relevance of insights generated from the analysis.

6. Domain Experts/Subject Matter Experts (SMEs): Domain experts or subject matter experts possess in-depth knowledge of the industry or domain being analyzed. They provide valuable domain-specific insights, validate analysis results, and ensure that the findings are relevant and actionable.

7. Regulatory and Compliance Officers: In regulated industries, regulatory and compliance officers ensure that the data analytics project complies with relevant laws, regulations, and industry standards regarding data privacy, security, and ethical use of data.

8. Customers/Clients: In some cases, customers or clients may also be stakeholders in data analytics projects, especially if the analysis aims to improve customer experiences, personalize offerings, or address customer-related issues.

9. External Consultants/Vendors: External consultants or vendors may be engaged to provide specialized expertise, tools, or services to support the data analytics project, especially for complex analyses or advanced analytics capabilities.

## Objectives

*We aim to conduct a comprehensive analysis of our student interns to gain insights about relationship between their academic performance, event participation, career aspiration and factors influencing their success. We have collected a dataset containing various attributes for each student.*

The objective of a data analytics project typically revolves around leveraging data to achieve specific goals or address particular challenges within an organization or industry. Objectives can vary widely depending on the context and requirements of the project, but some common objectives include:

1. Optimizing Operations: Improve efficiency, productivity, and cost-effectiveness by identifying bottlenecks, streamlining processes, and optimizing resource allocation based on data-driven insights.

2. Enhancing Decision Making: Provide decision-makers with timely, accurate, and actionable insights to support strategic planning, risk management, and resource allocation, leading to better-informed decisions.

3. Understanding Customers: Gain deeper insights into customer behavior, preferences, and needs through data analysis, allowing for the development of targeted marketing strategies, personalized product recommendations, and improved customer experiences.

4. Predictive Analysis: Forecast future trends, outcomes, or events based on historical data and predictive modeling techniques, enabling proactive decision-making and risk mitigation.

5. Improving Product/Service Quality: Identify opportunities for product or service improvement, innovation, and differentiation by analyzing customer feedback, usage patterns, and performance metrics.

6. Detecting Fraud or Anomalies: Identify suspicious activities, fraudulent transactions, or unusual patterns within datasets, helping to mitigate risks, ensure compliance, and protect against financial losses.

7. Optimizing Marketing Campaigns: Measure the effectiveness of marketing campaigns, identify high-value customer segments, and allocate marketing resources more efficiently to maximize return on investment.

8. Forecasting Demand: Predict future demand for products or services based on historical sales data, market trends, and external factors, enabling better inventory management and production planning.

9. Risk Assessment and Management: Assess and mitigate various risks, such as credit risk, market risk, or operational risk, by analyzing relevant data and developing risk models to support decision-making processes.

10. Compliance and Regulatory Reporting: Ensure compliance with regulatory requirements and reporting standards by analyzing data and generating accurate, timely reports for regulatory bodies and stakeholders.

These objectives can be further refined and tailored to align with the specific needs, challenges, and strategic priorities of the organization or industry undergoing the data analytics project.

# **METHODOLOGY**

These conventions are all about the positions of line breaks, how many characters should go on a line, and everything in between.

In the methodology section of a data analytics project report, it's important to outline the considerations and assumptions made throughout the project. This helps provide transparency and context for the analysis. Here's a structured approach:

4. Methodology

4.1 Considerations & Assumptions

In conducting this data analytics project, several key considerations and assumptions were taken into account to ensure the validity and reliability of the analysis:

1. Data Quality: It was assumed that the data collected for analysis were of sufficient quality and accuracy. Efforts were made to address any data quality issues through data cleaning and preprocessing techniques.

2. Data Availability: The availability of relevant data sources was considered essential for conducting the analysis. Efforts were made to access and collect the necessary data from both internal and external sources.

3. Assumption of Statistical Significance: Statistical techniques were applied under the assumption that the observed patterns, trends, and correlations in the data were statistically significant and not purely due to chance.

4. Homogeneity of Data: It was assumed that the data collected were homogeneous and consistent across different sources, time periods, and geographical regions, unless otherwise specified.

5. Privacy and Ethics: Adherence to privacy regulations and ethical standards was assumed throughout the project. Any sensitive or personally identifiable information was handled with the utmost confidentiality and in compliance with relevant laws and regulations.

6. Representativeness of Sample: The representativeness of the sample data used for analysis was assumed to be indicative of the broader population or target audience, allowing for generalization of findings and insights.

7. Model Assumptions: Any statistical models or machine learning algorithms used for analysis were based on certain assumptions about the underlying data distribution, relationships, and dependencies. These assumptions were carefully considered and validated where possible.

8. External Factors: External factors such as market conditions, economic trends, and regulatory changes were assumed to have potential impacts on the analysis results. Efforts were made to account for these factors where feasible.

9. Expertise and Resources: It was assumed that the project team possessed the necessary expertise, skills, and resources to conduct the analysis effectively and interpret the findings accurately.

10. Scope Limitations: The scope of the analysis was defined based on available resources, time constraints, and project objectives. Certain aspects of the data or variables of interest may have been excluded from the analysis due to these limitations.

By explicitly stating these considerations and assumptions, the methodology section provides a transparent framework for interpreting the analysis results and understanding the context in which the conclusions were drawn.

# **TARGETTED V/S ACHIEVED OUTPUT**

*5.1* ***Targeted Output***

*The targeted outputs of the data analytics project were defined to align with the objective of understanding the relationship between students' economic background, academic performance, competence, and expected salary. Key targeted outputs included:*

*-* ***Insightful Analysis****: Provide a detailed analysis of how economic background, GPA, and skills influence expected salary.*

*-* ***Actionable Recommendations****: Generate recommendations to guide career planning and internship strategies for students and stakeholders.*

*-* ***Improved Decision Making****: Enable data-driven decisions for Cloud Counselage and CC-GAC Foundation in supporting student success.*

*-* ***Enhanced Stakeholder Engagement****: Identify opportunities to optimize internship matching and policy development.*

*5.2 Achieved Output*

*The achieved outputs reflect the outcomes from analyzing the dataset with attributes like GPA, family income, Python experience, and leadership skills. Key achieved outputs include:*

*-* ***Insightful Analysis****: Uncovered patterns linking economic background and GPA to expected salary variations.*

*-* ***Actionable Recommendations****: Provided suggestions for tailoring internship opportunities based on academic and economic factors.*

*-* ***Improved Decision Making****: Delivered insights that supported strategic planning for stakeholder initiatives.*

*-* ***Enhanced Stakeholder Engagement****: Highlighted areas for improved internship matching, though full implementation awaits further data.*

*5.3* ***Discrepancies and Lessons Learned***

While the achieved outputs aligned closely with targets, some discrepancies were noted:

- **Scope Adjustments:** Analysis scope was refined due to limited data on certain attributes like leadership skills.

- **Unforeseen Challenges**: Data inconsistencies required additional cleaning, impacting initial timelines.

- **Continuous Improvement**: These lessons underscore the need for robust data validation and flexible planning in future projects.

# **CONCLUSION**

*This project will prove highly beneficial for stakeholders such as Cloud Counselage, CC-GAC Foundation, and educational institutions by providing data-driven insights that enhance student career planning, optimize internship matching, and inform policy decisions to support equitable opportunities. These insights can help recruiters refine hiring criteria, improve student employability, and bridge the gap between academic performance and expected salary, ultimately benefiting students and employers alike.*

*Future scope includes expanding the dataset to incorporate additional variables like industry trends, soft skills, and long-term career outcomes, enabling more comprehensive analyses. This could lead to the development of predictive models for student success, personalized career guidance tools, and broader research collaborations to address evolving educational and employment challenges.*

Acknowledgments:

- We extend our gratitude to all stakeholders, team members, and partners who contributed to the success of this data analytics project.

- Their dedication, expertise, and collaboration were instrumental in unlocking the full potential of data analytics and driving meaningful outcomes for [cloudcounselage].

# **APPENDICES**

## Appendix A – Title

In the appendices section of the data analysis project report, additional supplementary materials are included to provide further detail, context, or support for the analysis and findings presented in the main body of the report. The appendices may include:

1. Data Table: Detailed tables containing raw or processed data used in the analysis, including variables, observations, and any relevant metadata.

2. Code Snippets: Extracts of code or scripts used for data cleaning, preprocessing, analysis, or visualization, particularly if the analysis involved programming languages such as Python, R, or SQL.

3. Data Visualization: Additional charts, graphs, or visualizations that provide further insight into the data or support specific findings discussed in the main report.

4. Model Documentation: Detailed documentation of any statistical models, machine learning algorithms, or predictive models used in the analysis, including model specifications, assumptions, and performance metrics.

5. Survey Instruments: If survey data was collected as part of the analysis, the survey instruments, questionnaires, or survey response data may be included in the appendices.

6. References and Citations: A list of references, citations, or sources consulted during the project, including academic papers, textbooks, articles, and online resources.

7.Glossary of Terms: Definitions and explanations of key terms, acronyms, or technical terminology used throughout the report to assist readers in understanding the analysis.

8. Data Sources: Information on the sources of data used in the analysis, including data providers, data collection methods, and any relevant data agreements or permissions.

9. Ethical Considerations: Documentation of ethical considerations or approvals obtained for the collection, use, and analysis of data, particularly if the project involved sensitive or personally identifiable information.

10. Additional Analysis: Any supplementary analysis, sensitivity analyses, or alternative scenarios considered but not included in the main report due to space constraints.

Including appendices allows readers to delve deeper into the details of the analysis and provides transparency and accountability in the research process. Appendices should be clearly labeled and referenced in the main body of the report as needed.