



PROJECT REPORT

DATA ANALYTICS  
  
STUDENT INTERN COMPREHENSIVE ANALYSIS

|  |  |  |  |
| --- | --- | --- | --- |
| **Created By:** | Himaja G | **Approved By:** | Harshada Topale & Tushar Topale |
| **Created On:** | 28-06-2023 | **Approved On:** | 30-09-2023 |

Page left blank intentionally

**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** | **STUDENT INTERN COMPREHENSIVE ANALYSIS** | | |
| **Project Sponsor** | TUSHAR TOPALE | | |
| **Project Manager** | HARSHADA TOPALE | | |
| **Start Date** | 28-06-2023 | **Completion Date** | 30-09-2023 |

# **SUMMARY**

The project aimed to conduct an in-depth analysis of student interns to uncover insights regarding the relationship between academic performance, event participation, career aspirations, and factors influencing their success. A dataset containing attributes for each student served as the foundation for this analysis.

**Project Purpose:**

This project was initiated to address several key objectives:

1. *Gain Insights*: The primary goal was to gain a deeper understanding of the factors that contribute to the success of student interns within our organization.
2. *Improve Decision-Making*: By analyzing the data, we sought to make data-driven decisions regarding internship program enhancements, event planning, and career development initiatives.
3. *Enhance Student Experience*: We aimed to enhance the overall experience of student interns, providing them with opportunities for growth and development that align with their career aspirations.

**Long-Term Benefits:**

The comprehensive analysis of student interns has yielded several long-term benefits:

1. *Data-Driven Insights*: The project provided data-driven insights into the relationship between academic performance, event participation, and career aspirations, enabling us to tailor internship programs more effectively.
2. *Enhanced Program Planning:* By understanding the factors influencing success, we can better plan events and activities that align with the career goals and development needs of our interns.
3. *Improved Success Rates*: With a clearer understanding of success factors, we anticipate an increase in the overall success rates of our student interns, leading to a more satisfied and accomplished group of future professionals.
4. *Continuous Improvement*: The project has set the stage for ongoing data analysis and program improvements, ensuring that our organization remains at the forefront of student intern development.

# **INTRODUCTION**

## Background

Millions of students apply for internships/jobs every year, resumes play an important role in playing the first impression. The recruiters spend max of 2-3 minutes reviewing a resume landed in their mailbox or job board, ATS application. Surprising more than 70% of resumes get rejected in the initial screening.

In today's highly competitive job market, students and recent graduates often seek valuable internship experiences to kickstart their careers. Internships provide an opportunity for students to bridge the gap between academic learning and practical work experience. For organizations, interns bring fresh perspectives, enthusiasm, and potential future talent.

However, the success of student internships isn't solely dependent on the students' academic performance; it's a multifaceted equation. Interns' participation in events, networking, career aspirations, and various influencing factors also play a crucial role in determining their success and the overall value they bring to the organization.

Therefore, there is a growing need for data-driven insights to optimize the selection and management of student interns. This project aims to address this challenge by leveraging data analysis techniques to gain a deeper understanding of the factors that contribute to the success of student interns. By examining academic performance, event participation, career aspirations, and influential factors, the project seeks to provide organizations with actionable insights for improving their internship programs and enhancing the overall internship experience.

## Stakeholders

Some stakeholders related to the problem statement are:

1. Interns: The students participating in the internship program are direct stakeholders. They have a vested interest in the project's outcomes as it directly relates to their experiences and future career prospects.
2. Academic Institutions: The universities or colleges from which the interns originate are stakeholders since the project can impact their internship programs' effectiveness and reputation.
3. Employers/Organizations: Employers hosting the interns are stakeholders as they benefit from a better understanding of intern performance and factors affecting it. This knowledge can help them improve internship experiences.
4. Recruiters and HR Professionals: Personnel responsible for recruiting and managing interns have a stake in the project's findings, as it can inform their selection and mentoring processes.
5. Project Team: Members of the project team, including analysts, data scientists, and project managers, are stakeholders who are directly involved in the project's execution and outcome.
6. Management and Leadership: Executives and leaders within the organizations hosting interns have a stake in the project's results, as it can impact their decisions regarding internship programs.
7. Future Interns: Potential future interns could indirectly benefit from the project's insights as organizations may use these findings to enhance internship opportunities.
8. Industry and Professional Bodies: If applicable, industry associations and professional bodies related to the fields of study of the interns may have an interest in the project's outcomes.
9. External Partners: Any external partners or collaborators involved in the project are stakeholders, as they are invested in the project's success and outcomes.

## Objectives

The objective of the project is to conduct a comprehensive analysis of the student interns to gain insights about relationship between their academic performance, event participation, career aspiration and factors influencing their success by analyzing the dataset containing attributes for each student. Some of the preprocessing activities that are achieved are:

1. Data Collection: The project initiated by collecting a dataset containing attributes for each student intern. This dataset likely included academic performance data, event participation records, career aspirations, and various factors that could influence success.
2. Data Cleaning and Preparation: Before analysis could begin, data cleaning and preparation activities were conducted. This involved handling missing values, addressing outliers, and ensuring data quality and consistency.
3. Exploratory Data Analysis (EDA): EDA was performed to gain initial insights into the dataset. This included generating summary statistics, visualizing data distributions, and identifying potential correlations or patterns.
4. Data Visualization: Visualizations were created to represent the relationships between academic performance, event participation, career aspirations, and success factors. These visualizations aided in conveying insights effectively.

Also, eight questions are being answered from the *Basic questions* and eight questions are being answered from the *Moderate questions* given in the Problem statement, after a detailed analysis of the data.

# **METHODOLOGY**

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

## Considerations & Assumption

The considerations and assumptions should be followed throughout the project execution to manage risks, make informed decisions, and ensure the project's success while maintaining ethical and legal standards.

**Constraints:**

1. Data Availability: The availability and quality of the student intern dataset may be a constraint. Incomplete or inaccurate data can affect the accuracy and comprehensiveness of the analysis.
2. Privacy and Data Protection: Ensuring compliance with data privacy regulations and obtaining necessary permissions to analyze student data can be challenging.
3. Resource Constraints: Limited access to computing resources, software, or expertise in data analysis and machine learning can constrain the project's scope and capabilities.
4. Time Constraints: Meeting project deadlines, especially if there's a need for quick insights or to align with academic or internship program timelines, can be challenging.
5. Budget Constraints: Adequate funding may be required for data collection, analysis tools, and personnel, and budget limitations can affect project options.

**Challenges:**

1. Data Integration: If the student data comes from various sources, integrating and cleaning the data can be complex.
2. Ethical Considerations: Balancing the need for insights with ethical considerations related to student privacy and informed consent can be challenging.
3. Interdisciplinary Nature: The project may require expertise from various disciplines, including data science, education, and business, making collaboration and communication a challenge.
4. Complexity of Factors: Understanding the multitude of factors that influence student success is complex and may require advanced analytical techniques.

**Assumptions:**

1. Data Completeness: Assumption that the collected student data is sufficiently complete and accurate for analysis.
2. Relevance of Attributes: Assuming that the chosen attributes (academic performance, event participation, etc.) are relevant to student success.
3. Analysis Feasibility: Assuming that the selected analysis techniques and tools are feasible for the project's goals.

## Approach

A structured approach to solving a data analysis problem involves a systematic and organized methodology for extracting meaningful insights from data. Here is a step by step approach used in the project.

*Problem Definition and Understanding*: Started initially by clearly defining the problem or objective of the analysis. Understanding what insights or answers we seek to derive from the data. The objective is to gain insights into student intern behavior and success factors.

*Data Collection*: Gathering relevant data that will help address the problem. Ensuring the data is accurate, complete, and representative. We took the data on student attributes, event participation, academic performance, and career aspirations.

*Data Cleaning and Preprocessing*: Preparing the data for analysis by cleaning it. This includes handling missing values, outliers, and formatting issues. Data preprocessing ensures the data is in a suitable format for analysis.

*Exploratory Data Analysis (EDA)*: Conducting an exploratory analysis to understand the dataset's characteristics. This may involve summary statistics, data visualization, and identifying patterns or trends. EDA helps identify potential areas of interest or further analysis.

*Hypothesis Formulation*: Based on the analysis of the data and problem, formulating hypotheses or questions to investigate. Hypotheses guides the analysis and helps us to focus on specific aspects of the data.

*Data Analysis*: Performing statistical analyses, modeling, or other techniques to answer the given research questions or test hypotheses. This step often involves various data analysis tools and libraries.

*Visualization*: Creating visualizations to communicate findings effectively. Visualizations can include charts, graphs, and plots that provide a clear representation of insights.

*Interpretation*: Interpreting the results of our analysis in the context of the problem. What do the findings mean, and how do they relate to the project's objective?

## Activities

The activities done for the project are:

1. *Requirement Gathering*: The project's objectives and the questions to be answered were understood. The scope, data sources, and expected outcomes were clarified.
2. *Data Collection*: The necessary data was collected from various sources, ensuring alignment with the project's objectives.
3. *Data Cleaning and Preprocessing*: The data was cleaned by handling missing values, duplicates, and outliers. The data was formatted and structured for analysis.
4. *Exploratory Data Analysis (EDA)*: EDA was conducted to understand the data's characteristics. Summary statistics were generated, and data was visualized to identify patterns or trends.
5. *Hypothesis Formulation*: Hypotheses or research questions were formulated based on the initial understanding of the data.
6. *Data Analysis*: Statistical methods, modeling techniques, or algorithms were applied to analyze the data. Hypotheses were tested or research questions were answered.
7. *Visualization*: Data visualizations were created to present insights effectively.

Charts, graphs, and plots were used to communicate findings.

1. *Interpretation*: The results were interpreted in the context of the project's objectives and hypotheses.
2. *Conclusion and Recommendations*: Key findings were summarized, and recommendations or actionable insights were provided.
3. *Documentation*: The entire analysis process, including data sources, methods, and findings, was documented. Documentation was made clear and organized for reference.
4. *Presentation or Reporting*: A presentation or report was prepared to communicate the results to stakeholders or the intended audience. Visual aids and explanations were used to convey insights effectively.

# **TARGETTED V/S ACHIEVED OUTPUT**

In the project plan, the targeted outputs were defined to achieve specific goals and objectives. Here, I will outline the targeted outputs, what has been achieved, and the reasons for any deviations. It's important to note that deviations are common in projects and can be valuable for learning and improvement.

*Targeted Output 1***:** Clean and Preprocessed Data

*Achieved:* The data was successfully cleaned and preprocessed. Missing values, duplicates, and outliers were handled.

*Reason for Deviation*: No significant deviations occurred in this phase as it is a fundamental step in data analysis.

*Targeted Output 2*: Exploratory Data Analysis (EDA)

*Achieved*: EDA was conducted to understand data characteristics. Summary statistics and visualizations were created to identify patterns and trends.

*Reason for Deviation*: Some additional visualizations were created to gain deeper insights, which was not initially planned but added value to the analysis.

*Targeted Output 3*: Hypothesis Formulation

*Achieved*: Hypotheses and research questions were formulated based on the initial data exploration.

*Reason for Deviation*: No significant deviation occurred in this phase as it aligned with the project's objectives.

*Targeted Output 4*: Data Analysis

*Achieved*: Statistical methods and modeling techniques were applied to analyze the data. Hypotheses were tested, and research questions were answered.

*Reason for Deviation*: Some additional analyses were performed to explore specific trends, which were not initially planned but contributed to a more comprehensive understanding of the data.

*Targeted Output 5*: Data Visualization

*Achieved*: Data visualizations were created to effectively communicate findings.

*Reason for Deviation*: The type and number of visualizations were adjusted during the process to better represent the insights, but this did not significantly deviate from the plan.

*Targeted Output 6*: Conclusion and Recommendations

*Achieved*: Key findings were summarized, and recommendations or actionable insights were provided.

*Reason for Deviation*: No significant deviation occurred in this phase as it aligned with the project's objectives.

*Targeted Output 7*: Documentation

*Achieved*: The entire analysis process, including data sources, methods, and findings, was documented.

*Reason for Deviation*: Additional documentation was added to clarify specific analysis steps and decisions.

*Targeted Output 8*: Presentation or Reporting

*Achieved*: A presentation or report was prepared to communicate the results to stakeholders or the intended audience.

*Reason for Deviation*: Some adjustments were made to the presentation format for better clarity and understanding.

# **CONCLUSION**

The comprehensive analysis of student interns and the insights gained from this project can be highly valuable to various stakeholders involved, including educational institutions, employers, students, and career counselors. Here's how it can prove useful to them:

1. Educational Institutions:

* Understanding Student Performance: Educational institutions can gain insights into the factors that influence student performance, such as GPA, field of study, and event participation.
* Curriculum Enhancement: Analysis of academic performance can help institutions identify areas where curriculum enhancements or additional support may be needed for students.
* Career Guidance: The project can assist institutions in providing more targeted career guidance to students based on their aspirations and academic backgrounds.

2. Employers:

* Recruitment Strategy: Employers can use the insights to tailor their recruitment strategies, focusing on institutions that produce high-performing and well-prepared interns.
* Skill Assessment: The analysis can help employers understand the skill sets and experiences of interns from different educational backgrounds.
* Salary Expectations: Insights into salary expectations can assist employers in setting competitive compensation packages.

3. Students:

* Career Planning: Students can benefit from understanding how their academic performance and extracurricular activities may impact their career aspirations.
* Event Participation: The analysis can highlight the value of participating in certain events or programs for career development.
* Salary Insights: Students can gain insights into salary expectations in their field of study and make informed career decisions.

4. Career Counselors:

* Guidance for Students: Career counselors can use the findings to provide more personalized guidance to students, helping them align their career goals with their academic performance and interests.
* Event Recommendations: Counselors can recommend specific events or activities to students based on their career aspirations.

*Future Scope:*

* Predictive Modeling: The project can be extended to develop predictive models that forecast the future career success of students based on their current academic performance and event participation.
* Feedback Mechanism: Implementing a feedback mechanism could provide ongoing insights to educational institutions and employers, allowing them to adapt their strategies based on real-time data.
* Incorporating External Factors: Future analyses could consider external factors such as economic conditions, industry trends, and market demand to provide a more holistic view of career prospects.
* Comparison Across Institutions: Benchmarking the performance and career outcomes of students across different institutions could offer valuable insights for improving educational programs.
* Longitudinal Analysis: Tracking the career progress of students over time could provide insights into the long-term impact of education and event participation.

# **APPENDICES**

## Appendix A – Title

**Appendix A: Components Table**

|  |  |  |
| --- | --- | --- |
| **Component Name** | **Description** | **Status** |
| Data Collection | Gathering student intern data | Completed |
| Data Cleaning | Preprocessing and handling missing data | Completed |
| Data Analysis | Exploratory data analysis | Completed |
| Visualizations | Creating charts and graphs | Completed |
| Statistical Testing | Hypothesis testing and correlation | Completed |
| Insights Generation | Deriving meaningful insights | Completed |
| Report Writing | Documenting findings and recommendations | Completed |
| Stakeholder Feedback | Gathering feedback from stakeholders | Pending |