



PROJECT REPORT

DATA ANALYTICS  
  
ABSENCE OF INSIGHTS FOR THE RELATIONSHIP BETWEEN STUDENT'S ECONOMIC BACKGROUND, ACADEMIC PERFORMANCE, COMPETENCE & EXPECTED SALARY.

|  |  |  |  |
| --- | --- | --- | --- |
| **Created By:** | Shivohum Pandita | **Approved By:** | <Data Analytics> |
| **Created On:** | 12-07-2025 | **Approved On:** |  |

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 3](#_Toc143445379)

[3.3 Objectives 3](#_Toc143445380)

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

[4.1 Considerations & Assumption 4](#_Toc143445382)

[4.2 Approach 5](#_Toc143445383)

[4.3 Activities 5](#_Toc143445384)

[**5** **TARGETTED V/S ACHIEVED OUTPUT** 6](#_Toc143445385)

[**6** **CONCLUSION** 7](#_Toc143445386)

[**7** **APPENDICES** 8](#_Toc143445387)

[7.1 Appendix A – Title 8](#_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** | ABSENCE OF INSIGHTS FOR THE RELATIONSHIP BETWEEN STUDENT'S ECONOMIC BACKGROUND, ACADEMIC PERFORMANCE, COMPETENCE & EXPECTED SALARY. | | |
| **Project Sponsor** | Tushar Topale | | |
| **Project Manager** | Harshada Topale | | |
| **Start Date** | 25-06-2025 | **Completion Date** | 13-07-2025 |

# **SUMMARY**

The goal of this data analytics project was to explore how students’ academic performance, skills, and background affect their career expectations—especially in terms of expected salary and success. I worked with student event registration data, cleaned it using Python to remove duplicates and fix inconsistent entries, and used Excel and Tableau to analyse and visualize the insights.

My findings reveal that students with higher CGPAs and more Python experience tend to expect higher salaries, shedding light on how technical proficiency and academic performance correlate with career optimism. Additionally, students from higher family income groups also reported higher expected salaries, suggesting socio-economic background plays a role in shaping career expectations.

One major challenge I faced was the poor quality of raw data—it took considerable time to clean and organize before analysis could begin. Still, the project achieved its goal of turning messy data into meaningful insights.

In the future, I recommend collecting cleaner, more structured data from the start and exploring more variables like domain interests or internship experience to deepen the analysis.

# **INTRODUCTION**

## Background

This data analytics project set out to understand how factors like academic performance, technical skills, and student background influence career-related expectations such as salary. By analysing student registration data from various events, the project aimed to uncover patterns and insights that could help guide students, educators, and organizations in making more informed decisions.

In today’s competitive job market, knowing what influences student success is more important than ever. This project shows how data analytics can help uncover meaningful trends in the education and career space—empowering institutions and stakeholders to support students more effectively and plan for future opportunities.

## Stakeholders

The key stakeholders involved in this project include:

* **Students:** As the primary data contributors, their academic records, experiences, and expectations form the core of the analysis. The insights generated can help them understand how their skills and background impact career outcomes.
* **Educational Institutions and Faculties:** Colleges and professors can use the findings to better guide students on academic and skill development paths aligned with industry trends.
* **Event Organizers (e.g., Cloud Counselage):** The data was collected through event registrations. Organizers are interested in understanding student demographics, participation patterns, and how to improve outreach and engagement strategies.
* **Career Counselors and Recruiters:** Insights about expected salary, CGPA, and experience can assist them in aligning hiring practices and advice with student readiness.
* **Data Analysts and Project Team:** Responsible for cleaning, analyzing, and presenting the data. Their goal is to ensure accuracy, reliability, and usability of the information.

## Objectives

The main objective of the project, as outlined in the Project Charter, was:

* **To analyse student data to understand how academic performance, technical skills, and personal background relate to career expectations and event participation.**

Though the Main Objective remains the same , there are some more objectives since Project Charter was approved, that is

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.  
  
Improving 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.

# **METHODOLOGY**

## Considerations & Assumption

While carrying out this data analytics project, I made the following considerations and assumptions to ensure the analysis remained relevant, clear, and accurate:

* **Data Quality**: I assumed that the raw data collected through event registrations was mostly accurate. However, I did notice issues like conflicting details (e.g., different graduation years for the same student), so I cleaned and standardized the data using Python (Pandas).
* **Data Cleaning Approach**: In cases where students had multiple entries with different values, I used majority voting (mode) for fields like graduation year and college name—except where a clear majority wasn’t possible. In such cases, I flagged them as conflicts and excluded them from sensitive analyses.
* **Scope of Analysis**: My analysis was based on student participation data, so the insights are limited to that context. I did not include other user roles (like professionals or faculty) when analysing trends related to students.
* **Event-Based Participation**: A student could participate in multiple events. For some analyses, I treated each event participation separately; for others, I consolidated based on unique students.
* **Trusted vs. Untrusted Data**: I performed integrity checks (e.g., variation in CGPA or multiple cities listed) to separate trusted and untrusted data. Wherever accuracy was critical (like calculating average CGPA), only trusted data was used.
* **No Advanced Modelling**: This project focused on descriptive analytics and visualization, not predictive models. I used statistical summaries, trends, and comparisons—not machine learning.
* **Time & Resource Limitations**: Due to time constraints, not all fields (like designation or course type) were deeply explored. I focused on the most meaningful and cleanable data.
* **Tool Capabilities**: I conducted all analysis using Excel, Python (for cleaning), and Tableau (for visualization). I assumed these tools were sufficient for delivering accurate insights.
* **Assumption of Honest Responses**: I assumed that participants filled in their details honestly—especially fields like expected salary, CGPA, and technical experience.
* **Generalizability**: The findings mostly apply to the specific dataset and may not reflect broader student populations beyond this event data.

These considerations helped guide how I handled the data and ensured that the conclusions drawn were based on thoughtful assumptions and practical limitations.

## Approach

My approach began with understanding the dataset thoroughly examining the structure, key fields, and types of values present. I focused on identifying inconsistencies or conflicting information, especially in critical fields like graduation year, college name, and technical experience.

I cleaned the data using **Pandas** in Python, applying logic such as using the median for numeric fields (like CGPA) and mode for categorical fields (like college name). In cases where a clear consensus couldn't be determined (e.g., tied modes), I flagged those records with conflict indicators.

To maintain data reliability, I introduced an **“Is Trusted”** flag. If a student's CGPA varied by more than 1 point across records, or if they listed multiple cities, their data was marked as **untrusted** and excluded from sensitive analyses like average CGPA or city-based comparisons.

Once the data was cleaned and verified, I visualized and analysed it using tools like Excel and Tableau. This structured approach helped ensure insights were based on consistent and reliable data.

## Activities

To deliver the project effectively, I carried out the following activities:

* **Requirement Understanding**: I reviewed the problem statement and objectives to clarify what insights were expected from the analysis.
* **Data Understanding**: I explored the raw dataset to get familiar with its structure, types of fields, and potential issues.
* **Data Cleaning**: Using Python and Pandas, I removed duplicates, handled conflicting entries, resolved missing or inconsistent values, and flagged unreliable data based on defined rules.
* **Integrity Checks**: I implemented logic to identify records with unusually varying CGPA or multiple city entries and marked them as untrusted.
* **Data Transformation**: I applied aggregation techniques like median and mode to build a consistent profile for each student, especially where multiple event entries existed.
* **Exploratory Analysis**: I used Tableau and Excel to visualize the data, identify trends, and answer analytical questions based on academic performance, background, and career outcomes.
* **Insight Generation:** I interpreted the visualized data to draw conclusions about key metrics like graduation trends, CGPA distribution, event participation, and salary expectations.
* **Reporting:** I documented the findings, included conclusions for each question, and structured them into a final project file for submission.

# **TARGETTED V/S ACHIEVED OUTPUT**

**Targeted Output:**

The goal of the data analytics project was to understand how academic performance, skills, and background influence student outcomes—especially expected salary and participation in events. The targeted outputs included:

* **Insightful Analysis**: Identify meaningful trends in the dataset, such as CGPA distribution, graduation years, or student interest in different events.
* **Actionable Recommendations**: Generate insights that could help improve event planning, outreach, and targeting of students based on their profiles.
* **Improved Decision-Making**: Support data-driven decisions by presenting findings in a clear and interactive format using tools like Tableau.
* **Reliable & Clean Data**: Resolve conflicting information (e.g., inconsistent graduation years, CGPA, or college names) using appropriate techniques (median, mode, or conflict flags) and create a trustworthy dataset for analysis.

**Achieved Output:**

The project achieved its main goals:

* **Insightful Analysis:** The analysis uncovered that students graduating in 2023 were the most active, and those with higher CGPA and more Python experience expected better salaries. These trends were visualized through clean and interactive charts and dashboards.
* **Actionable Recommendations**: The data showed that channels like WhatsApp and Email were more effective for promoting events, and colleges like St. Xavier’s had the most engaged students. These findings helped recommend where to focus future outreach.
* **Improved Decision-Making:** Tableau dashboards made it easier to compare trends across colleges, cities, and events—helping stakeholders make smarter choices about planning and engagement.
* **Reliable & Clean Data**: Using Python (Pandas), the data was cleaned, and flags were added for entries with conflicting values. For students with major inconsistencies (e.g., in graduation year), integrity checks were done

# **CONCLUSION**

This data analytics project provides valuable insights into how students’ academic profiles, technical skills, and backgrounds influence their participation and expectations in career-oriented events. By cleaning and standardizing the dataset, identifying inconsistencies, and visualizing key trends using Tableau, I was able to create a reliable foundation for decision-making.

For stakeholders such as educational institutions, training providers, or event organizers, the insights can be used to:

* **Target the right student segments** based on CGPA, graduation year, or skill levels.
* **Refine promotional strategies** by identifying the most effective outreach channels.
* **Improve event planning** by understanding trends in student participation across different fields and cities.

**Future Scope**

* Adding more data sources (e.g., LinkedIn outcomes, placement data) could enhance predictions about career success.
* Implementing predictive models could help forecast expected salary or participation likelihood.
* More detailed skill-level tagging (beyond just Python experience) can provide deeper insights into employability.

# **APPENDICES**

## Appendix A – Title

In the appendices section of the data analysis project report, additional supporting materials are provided to enhance the understanding of the methodology and results presented. These materials serve to offer transparency and detail without disrupting the flow of the main report. The appendices may include:

* **Raw Data Samples**: Selected portions of the original dataset used in the project to illustrate formatting and structure.
* **Data Cleaning Scripts**: Python (Pandas) scripts used for cleaning, deduplication, and conflict resolution in the dataset.
* **Calculated Fields and Logic**: Definitions of calculated fields and derived metrics used in Tableau or Excel for analysis and visualization.
* **Visualizations**: Full-sized versions of key graphs, dashboards, and charts referenced throughout the report.
* **Conflict Resolution Examples**: Specific examples (e.g., for CGPA, Graduation Year) where data conflicts were identified and resolved using statistical techniques like mode or median.
* **Code Outputs**: Terminal or Google Colab outputs that validate the transformations or summarize data insights.
* **Additional Charts or Unused Visuals**: Visualizations generated during exploratory analysis that were not included in the main report but still hold informative value.

These appendices help ensure that the data analysis process is transparent, replicable, and understandable for stakeholders or future analysts referencing this project.