CLOUD COUNSELAGE
IT & MANAGEMENT CONSULTING SERVICES

Project Report
Data Analytics Assignment
Submitted to: Cloud Counselage IT & Management Consulting Services

Submitted by: Mayuri Khatpe

1. Introduction

The Data Analytics Assignment is focused on assessing skills in data cleaning, analysis, and visualization using Python and Power BI. This report highlights the execution and findings from various tasks, ranging from web scraping to building an insightful dashboard and addressing customer success queries.
This assignment also evaluates how effectively insights can be communicated through visualizations and storytelling to enable decision-making.

2. Project Objectives

1. Perform web scraping to collect relevant data (email IDs of colleges in the Mumbai region).
2. Clean the provided dataset to prepare it for analysis.
3. Create visualizations to uncover patterns and relationships.
4. Present insights and recommendations through data storytelling.
5. Answer specific questions related to customer success and career aspirations.

Tasks Overview

Task 1: Web Scraping

Objective:
Extract email IDs of colleges in the Mumbai region.

Process:

- Python libraries like requests, re, and PyPDF2 were used for fetching and parsing the PDF content.
- The script downloaded a public PDF file containing college details.
- Using regular expressions (re), email patterns were identified and validated within the extracted text.

Challenges:

1. Addressing encoding issues during text extraction from the PDF.
2. Refining the regex pattern to avoid false positives or invalid email matches.

Outcome:

- Successfully extracted a list of valid email IDs for colleges in the Mumbai region.
- The extracted data was saved or displayed for further reference.

Task 2: Data Cleaning

Objective: Clean and preprocess the dataset for analysis.

Steps Implemented in Excel:

1. **Removed Duplicate Entries:**
    - Duplicates were removed based on email IDs to avoid redundancy in the dataset.
    - Ensured only unique data entries were retained.
2. **Removed Null Values:**
    - Eliminated rows with missing or null data in critical columns, such as designation and email.
3. **Filtered Designation:**
    - Retained only rows where the designation was "Student" to focus on relevant data.

Steps Implemented in Python:

1. **Converted Family Income Ranges:**
    - Converted income ranges (e.g., "0-2 Lakh") into their mean values for quantitative analysis:
        - 0-2 Lakh -> 1 Lakh
        - 2-5 Lakh -> 3.5 Lakh
        - 5-7 Lakh -> 6 Lakh
        - 7 Lakh+ -> 7 Lakh
2. **Removed Extra Spaces and Special Characters:**
    - Cleaned text fields to ensure consistency and accuracy.
3. **Consolidated Responses:**
    - Combined responses from columns such as "How did you come to know about this event?" and "Specify in Others" into a single column labeled "Others" for better analysis.

Outcome:

- A cleaned dataset was prepared and The data was ready for advanced analysis and visualization.

Task 3: Data Visualization

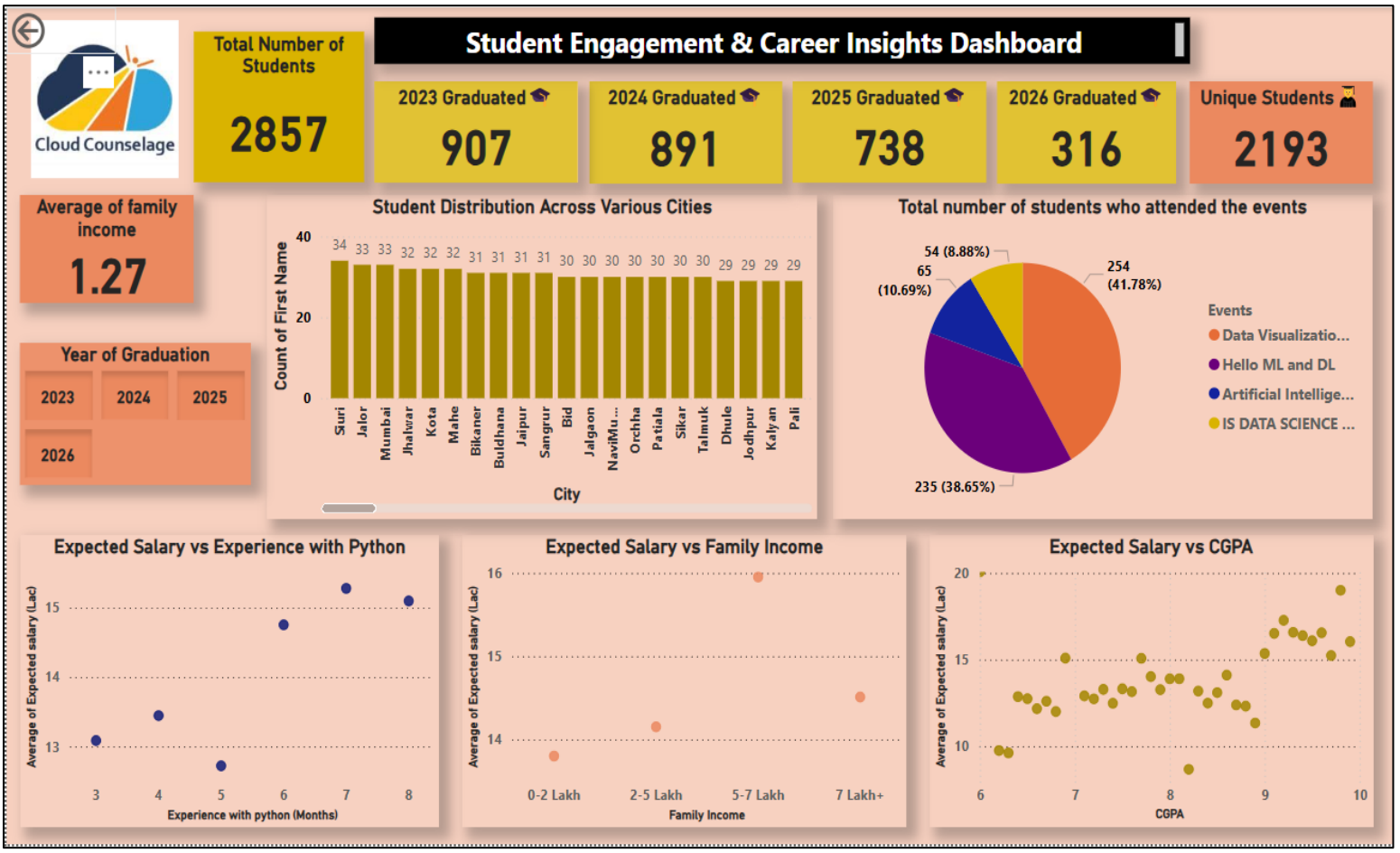
Objective: Create an interactive dashboard in Power BI to present insights.

Key Visualizations:

1. **Total Students and Unique Students:**
    - Total number of students: 2857
    - Unique students: 2193
2. **Student Distribution by City:**
    - Cities like Suri, Jalore, Mumbai, and Jhansi had participation counts ranging from 29 to 34 students each.
    - Bar chart representation for clarity.
3. **Expected Salary Analysis:**
    - Explored how expected salary correlates with:
        - **CGPA:** Higher CGPA generally aligns with higher salary expectations.
        - **Family Income:** Students from higher-income families tend to expect higher salaries.
        - **Python Experience:** A positive relationship exists between months of experience in Python and salary expectations.
4. **Graduation Year Analysis:**
    - Visualized the distribution of students graduating in 2023 (907), 2024 (891), 2025 (738), and 2026 (316).
5. **Event Participation:**
    - Data Visualization: 254 (41.78%)
    - ML and DL: 235 (38.65%)
    - AI: 65 (10.69%)
    - Data Science: 54 (8.88%)
6. **Average Family Income:**
    - Calculated as ₹1.27 Lakh overall, indicating the socio-economic status of students.

Outcome:

- An interactive Power BI dashboard was created, providing stakeholders with actionable insights through clear and dynamic visualizations.


Cloud Counselage
Total Number of Students
2857
Student Engagement & Career Insights Dashboard
2023 Graduated
907
2024 Graduated
891
2025 Graduated
738
2026 Graduated
316
Unique Students
2193
Average of family income
1.27
Year of Graduation
2023
2024
2025
2026
Student Distribution Across Various Cities
Count of First Name
City
Suri Jalor Mumbai Jhalwar Kota Mahe Bikaner Buldhana Jaipur Sangrur Bid Jalgaon NaviMu... Orchha Patiala Sikar Talmuk Dhule Jodhpur Kalyan Pali
34 33 33 32 32 32 31 31 31 31 30 30 30 30 30 30 29 29 29 29
Total number of students who attended the events
54 (8.88%)
65 (10.69%)
254 (41.78%)
235 (38.65%)
Events
Data Visualizatio...
Hello ML and DL
Artificial Intellige...
IS DATA SCIENCE ...
Expected Salary vs Experience with Python
Average of Expected salary (Lac)
Experience with python (Months)
Expected Salary vs Family Income
0-2 Lakh
2-5 Lakh
5-7 Lakh
7 Lakh+
Family Income
Expected Salary vs CGPA
CGPA

Task 4: Data Storytelling

1. Average Family Income (Summary Card)

- **Story Element**: The income distribution highlights that most students come from financially constrained families, with an average family income of ₹1.27 Lakhs. This indicates the potential financial challenges faced by students that may affect their career aspirations and access to resources.
- **How the Dashboard Helps**: The summary card effectively draws attention to the socioeconomic background of students, underlining the need for targeted interventions such as financial support, scholarships, and counseling programs.

2. Graduation Years (Summary Card & Insights)

- **Story Element**: Engagement levels vary across graduation years, with the 2023 and 2024 cohorts showing higher participation. Students graduating in later years, like 2025 and 2026, display lower engagement levels, highlighting the need for early support to increase involvement and preparedness.
- **How the Dashboard Helps**: The graduation year summary enables stakeholders to identify patterns and tailor interventions to enhance engagement, particularly for younger cohorts. Programs can focus on skill-building and career readiness for less active groups.

3. Total Participations and Unique Students (Summary Card)

- **Story Element**: A total of 2,857 participations and 2,193 unique students reflect consistent interest among students, but also suggest opportunities to broaden the audience reach and deepen engagement.
- **How the Dashboard Helps**: Differentiating between total participations and unique students allows for a better understanding of attendance trends. Strategies can be designed to attract a wider and more diverse group of participants.

4. Student Distribution Across Cities (Bar Chart)

- **Story Element**: The city-wise distribution reveals active participation from cities like Suri, Jalor, Mumbai, and others, while some areas show comparatively lower involvement. This provides insight into regional disparities in engagement.
- **How the Dashboard Helps**: The bar chart clearly highlights geographic patterns, enabling stakeholders to focus outreach efforts on underrepresented cities by organizing localized events and collaborations with educational institutions.

5. Event Participation by Topic (Pie Chart)

- **Story Element**: Participation is balanced across topics like Artificial Intelligence, Machine Learning, and Data Science, reflecting students' broad interest in diverse technological domains.
- **How the Dashboard Helps**: The pie chart emphasizes the need to continue offering workshops across these topics and expand advanced training opportunities to sustain interest and develop deeper expertise.

6. Expected Salary vs. Python Experience (Scatter Plot)

- **Story Element**: A positive correlation between Python experience and expected salary shows that practical, hands-on experience plays a significant role in shaping career outcomes and salary expectations.
- **How the Dashboard Helps**: The scatter plot underscores the importance of skill development programs. Stakeholders can prioritize investments in real-world projects, internships, and Python training sessions.

7. Expected Salary vs. Family Income (Scatter Plot)

- **Story Element**: Students from lower-income families tend to set modest salary expectations, which could stem from limited access to career guidance or a lack of confidence.
- **How the Dashboard Helps**: This visualization highlights the need for mentorship and career counseling programs targeted at students from economically disadvantaged backgrounds. Such initiatives can help inspire higher aspirations and broaden career opportunities.

8. Expected Salary vs. CGPA (Scatter Plot)

- **Story Element**: A positive correlation between CGPA and expected salary emphasizes the role of academic performance in shaping career ambitions.
- **How the Dashboard Helps**: The scatter plot calls attention to the potential benefits of academic support programs to help students with lower CGPAs achieve better career outcomes and raise their expectations.

9. General Observations and Trends

- **Balanced Interest in Technology**: Equal participation in topics like AI, ML, and Data Science reflects students' broad curiosity, providing opportunities to introduce more specialized and advanced sessions.
- **Key Correlations**: The dashboard highlights the interplay between factors like Python experience, family income, and CGPA in shaping students' career aspirations. These insights form a foundation for targeted interventions.

Actionable Recommendations

The dashboard simplifies complex data into intuitive visuals, enabling actionable insights:

- **City-level Targeting**: Use the bar chart insights to focus outreach campaigns in underrepresented regions.
- **Career Readiness**: Leverage summary cards and scatter plots to prioritize support for students from lower-income families or with lower GPAs.
- **Skill Development Focus**: Scatter plots linking Python expertise with salary expectations justify offering advanced workshops and real-world training opportunities.

Conclusion

The dashboard effectively combines data visualization and key metrics to narrate a compelling story about student engagement, aspirations, and potential interventions. It empowers stakeholders to make data-driven decisions, maximize impact, foster inclusivity, and better prepare students for successful careers.

Task 5: - Data Analysis for Customer Success

Customer success is about helping customers achieve their goals with a product or service. By supporting them and ensuring they get value, businesses build stronger relationships, leading to happier customers. It's important because satisfied customers are more likely to stay and recommend the business.
• Are you interested in pursuing a career in customer success? Why or why not?

Ans: Yes, I would like to pursue a career in customer success as it aligns directly with my passion for problem-solving, communication, and value creation. I find this exciting and fulfilling because customer success is all about helping customers maximize the value they get from a product or service, which is exciting and fulfilling to build a relationship with.

I think a customer journey is the most important aspect of the success story of any business, and I am enthused about being part of the journey. By utilizing data, feedback, and proactive support, I can ensure customers reach their objectives, which leads to stronger relationships and greater retention. Additionally, working in customer success allows one to be continually learning and growing since it is constantly interacting with the diverse needs, challenges, and goals of the customers, thus making every day unique.

It is a key intersection of both technical and interpersonal elements with strong communication skills and a data-driven mindset that allows me to see customer success as a perfect blend. I am eager to help businesses enhance their relationships with customers while contributing to the overall success and satisfaction of clients

Outputs and Deliverables

1. **Cleaned Dataset (Excel):**
    - Includes sheets detailing steps like removing duplicates, handling null values, and filtering student designations.
2. **Python Code:**
    - The script (data_analytics_assignment.py) automates tasks like web scraping, data cleaning, and column consolidation.
3. **Power BI Dashboard:**
    - An interactive dashboard visualizing:
        - Student distribution across cities.
        - Expected salary trends with CGPA, income, and experience.
        - Graduation timelines and event participation.
4. **Web Scraping Results:**
    - A comprehensive list of email IDs for colleges in the Mumbai region.

Conclusion

The project successfully demonstrated skills in data collection, cleaning, analysis, visualization, and storytelling. The interactive dashboard provides actionable insights into student demographics, skill correlations, and event participation.

Recommendations

1. **Support Low-Income Families**:
    - Provide scholarships or financial aid programs to alleviate financial barriers for students.
2. **Upskilling Opportunities**:
    - Encourage technical training in Python and data analysis tools to enhance career readiness.
3. **Targeted Outreach**:
    - Focus efforts on underrepresented cities to increase participation and engagement.
4. **Mentorship and Guidance**:
    - Implement mentorship programs to help students from lower-income families and those with lower GPAs aim higher in their career aspirations.
5. **Specialized Workshops**:
    - Offer advanced workshops in AI, ML, and Data Science to cater to students' interests and sustain their enthusiasm for technological domains.