

STUDENT SURVEY ANALYSIS

SOI-2024-2410-0029

Date of Submission: 29 July 2024

Submitted By:

SOI-2024-0055

|  |  |
| --- | --- |
| 22023518 | Peh Yi Feng Javier |
| 22013343 | Tan Hong Bin |
| 22025338 | Liaw Hong Da |
| 22013361 | Daniel Sanjay Sadhnani |

**TABLE OF CONTENTS**

ACKNOWLEDGEMENTS..................................................................................................5

ABSTRACT..........................................................................................................................6

1 Introduction..................................................................................................................7

2 Project Specification and Plan.....................................................................................9

2.1 Project Overview...............................................................................................9

2.2 Project Outcome................................................................................................10

2.3 Project Plan........................................................................................................10

2.4 Tasks Allocation.................................................................................................13

3 Project Approach.........................................................................................................16

3.1 Methodology......................................................................................................16

3.2 Survey Questionnaire.........................................................................................16

4 Business Understanding...............................................................................................25

4.1 Business Questions............................................................................................25

5 Data Understanding....................................................................................................28

6 Data Preparation ........................................................................................................29

7 Modelling...................................................................................................................31

7.1 Data Visualisation...............................................................................................31

7.2 Clustering............................................................................................................34

7.3 Text Analysis.......................................................................................................37

7.4 Predictive Analysis.............................................................................................40

8 Conclusion and Recommendations..........................................................................43

9 User and Technical Documentations.........................................................................49

9.1 User Documentation/Guide/Manual.....................................................................49

10 Reflections....................................................................................................................52

10.1 Peh Yi Feng Javier............................................................................................52

10.2 Daniel Sanjay Sadhnani...................................................................................53

10.3 Tan Hong Bin Darius......................................................................................54

10.4 Liaw Hong Da..................................................................................................55

References.............................................................................................................................57

Appendices............................................................................................................................58

Project Poster.........................................................................................................................59

# ACKNOWLEDGEMENTS

Special Thanks to Our Supervisor Miss Nimmy Varghese for helping and guiding us to the right path for our project!

# ABSTRACT

Understanding student demographics, behaviours, and goals across educational institutions is the goal of the Students Survey Analysis Project. The research offers insights to improve educational programs and support services by examining data on academic performance, gaming habits, part-time work, co-curricular activities (CCAs), satisfaction levels, and career aspirations.

Firstly, a detailed survey was created and administered to 260 students, gathering comprehensive data on the specified areas. The data was then cleaned to guarantee correctness. To comprehend student profiles and their connections to gaming, part-time employment, CCA involvement, and satisfaction levels, data visualization approaches were used in the second phase. In the third stage, pupils were grouped according to shared characteristics using clustering techniques. Key insights were extracted from the analysis of replies concerning job alternatives and necessary abilities through the application of text analytics in the final phase.

The project has accomplished a detailed analysis of student profiles, identifying patterns that help educators design more effective programs and support services. Insights from data visualization and clustering analyses provide a deeper understanding of student behaviours and needs. Text analytics revealed key career aspirations and skill requirements, offering valuable information for career guidance.

The finished project consists of a thorough report, an intuitive dashboard, and multiple data and text analytics models. The report provides a more thorough and comprehensive insights, suggestions, and achievable actions. Given the circumstances, the project has met its goals and made a significant addition to our understanding of and efforts to enhance and improve student experiences and outcomes.

# Introduction

* The main problem aims to find out data on student behaviours, preferences, and future aspirations. This includes their academic performance, extracurricular activities, part-time work, and gaming habits. Understanding these factors can help tailor educational and career guidance more effectively to meet their needs and aspirations.
* By finding out factors that affect students GPA such as Gaming, Part-Time Work, CCA, School Satisfaction, the business can use this information to derive decisions that would potentially boost students GPA but addressing the factors in a positive way
* The major tasks and deliverables of the project are Survey Creation, Data Collection, Data Cleansing and Visualization, Data Analytics (Clustering), Text Analytics (Text analytics models, word clouds, and topics related to career options and skills.), Final Report
* The creation of the survey and data collecting, data cleansing and visualization, clustering analysis, and text analytics are the stages in which the project will be carried out. The outcomes will be combined into a report and shown through an intuitive dashboard, offering practical insights into the requirements and actions of the students.

# Project Specification and Plan

## Project Overview

* This project is such a crucial and personal project for us as this project helps to see what affects GPA based on a variety of factors such as Gaming, Part-Time Work, CCA and School Satisfaction. It is personal for us as being year 3 students, it can also help us in the real world and see how we can Increase GPA levels based on these factors. Which helps not only us but also the vast number of students seeking ways to do well in polytechnic
* Identify how Certain Factors such as Gaming, Part-Time Work, CCA and School Satisfaction Affect Students GPA In which we can use This Information To see how we can Improve Students GPA
* The project involves creating and conducting a survey to collect data from 150-200 participants on demographics, school performance, gaming habits, part-time work, CCA involvement, satisfaction levels, career aspirations, and required skills. The collected data will be cleansed via Knime and Python and visualized using Tableau or Power BI and Python, followed by clustering analysis using Knime, Orange, or RapidMiner. Text analytics will be performed on career and skills responses. Deliverables include a dashboard, data analytics and text analytics models, a user guide, and a final report with insights and recommendations.
* We think that the variables (gaming, part-time work, CCA, and school happiness) will affect GPA in a way that can be measured using text analytics and clustering, and that the models created will offer precise and useful information. Involving stakeholders is also essential; we anticipate that stakeholders, such as educators and students, will be open to discussing the results and suggestions, and that all parties who are meant to benefit from the final deliverables—the dashboard, models, user manual, and report—will be able to access and utilize them.

## Functional Requirements

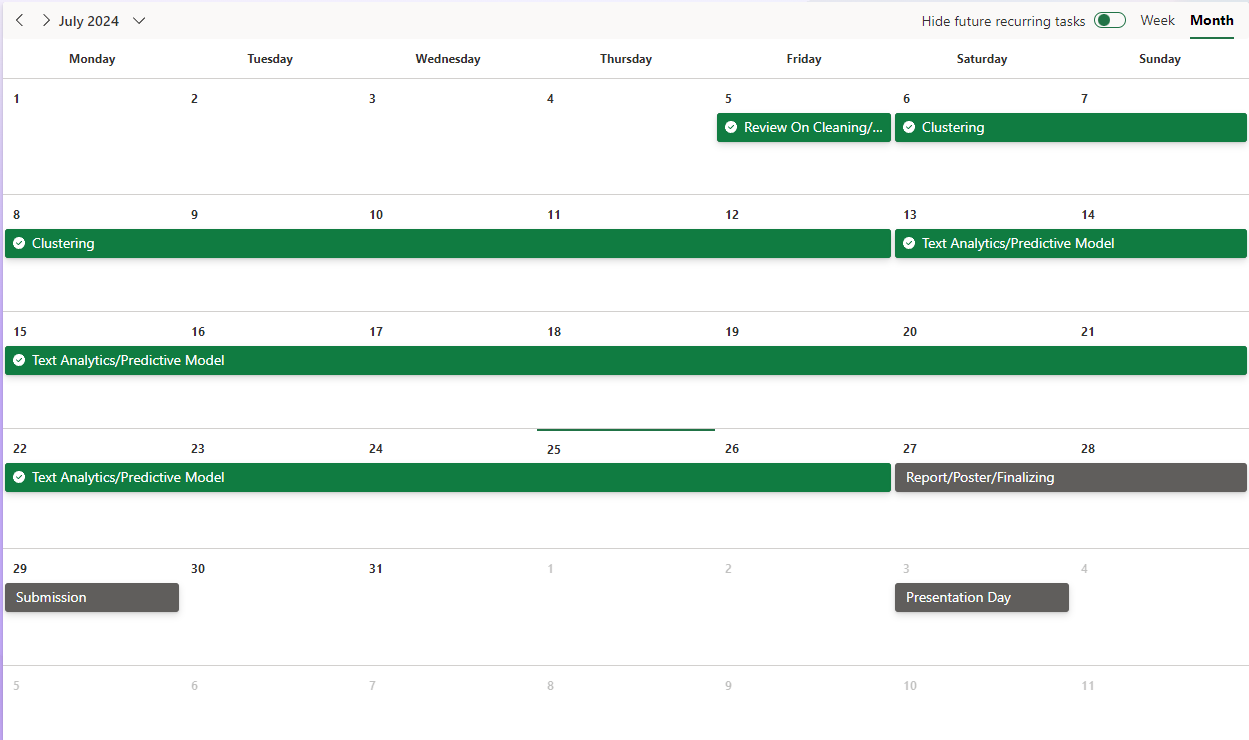
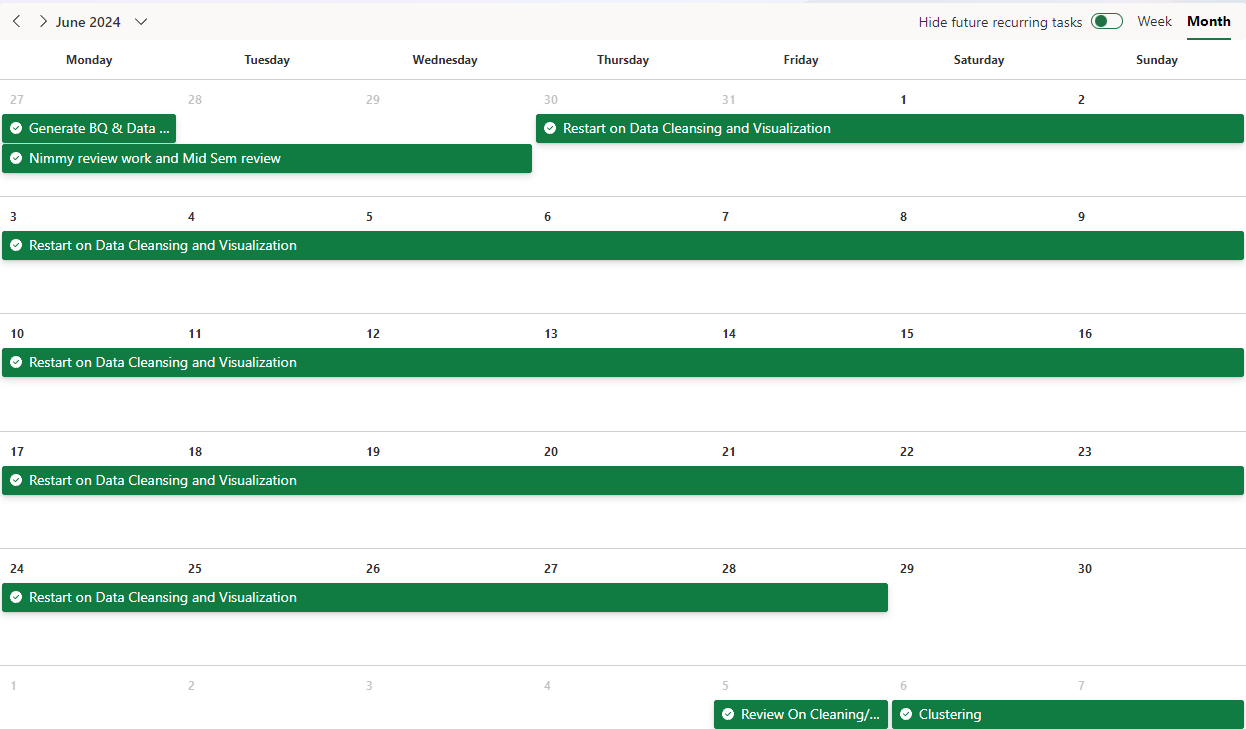
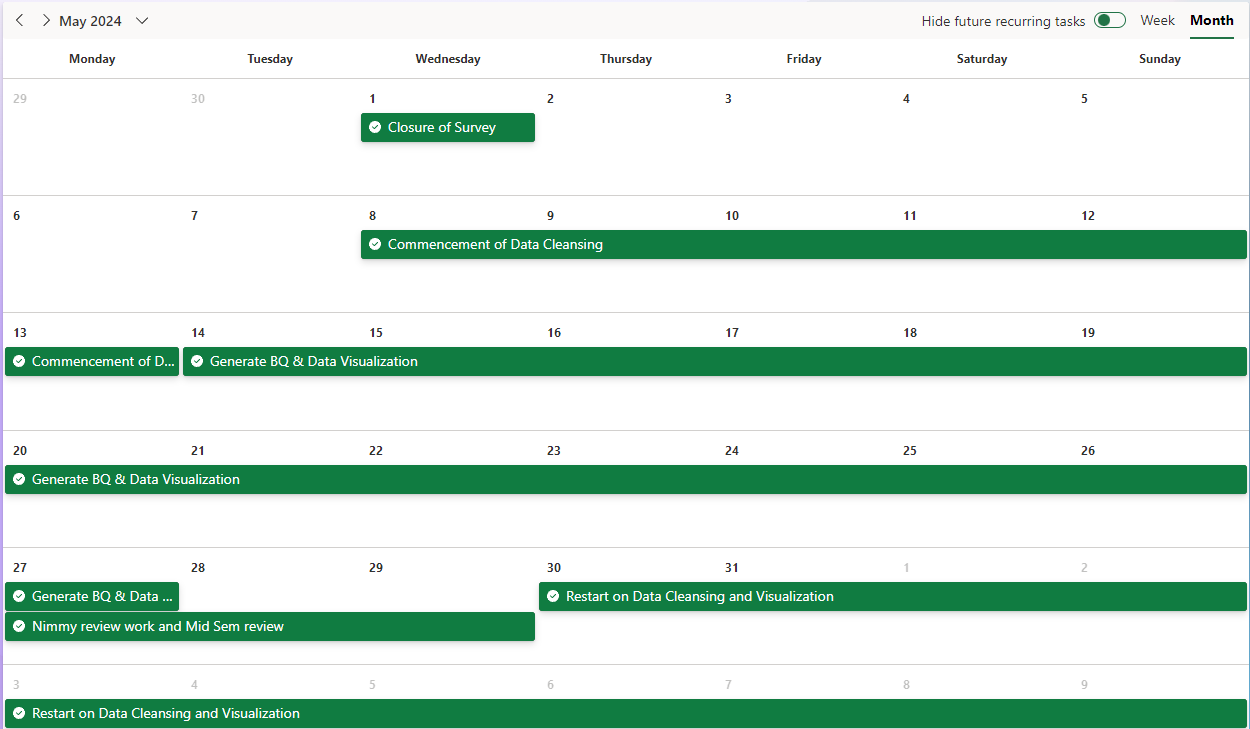
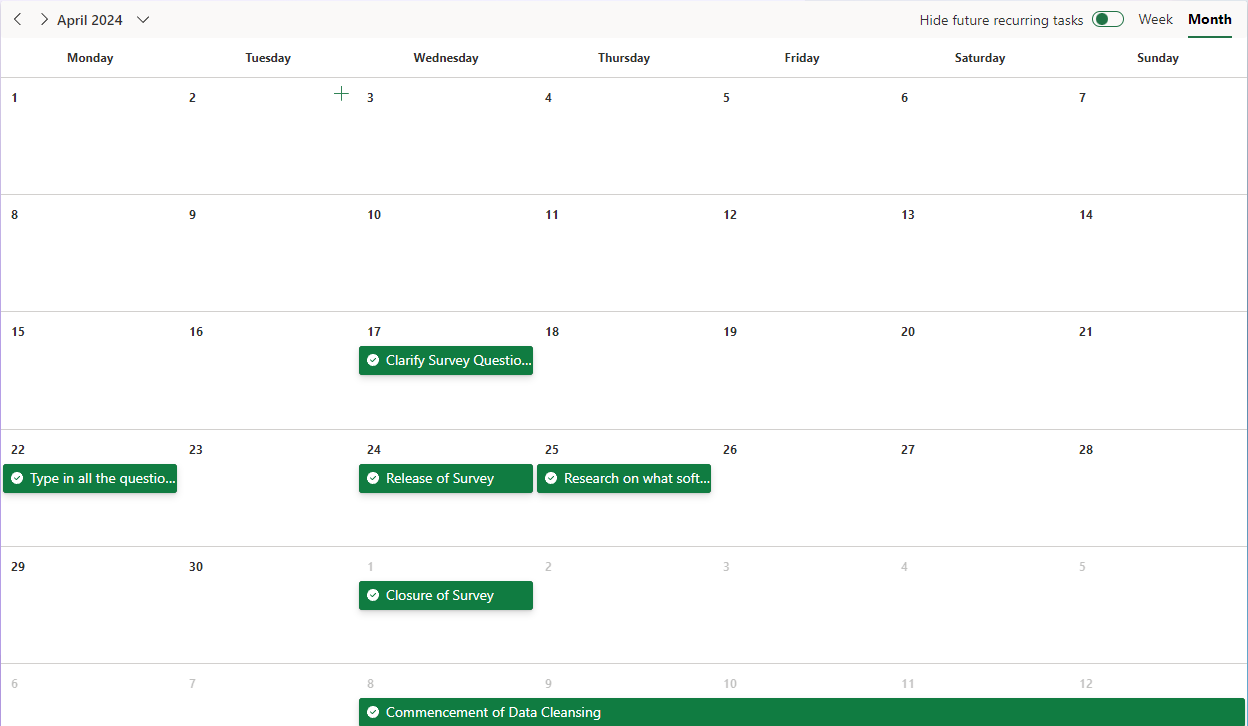
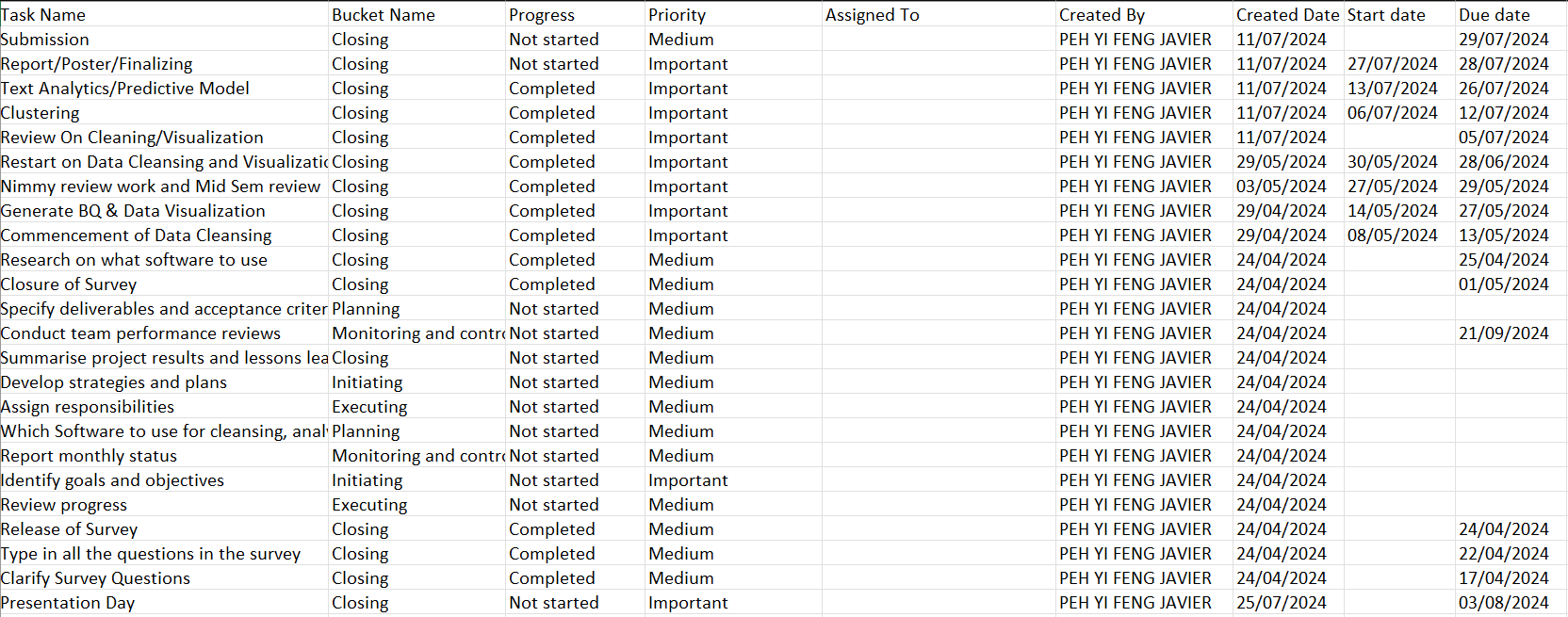
Major Functions of the system.

* Survey: Should gather data on demographics, academic performance, lifestyle habits, satisfaction levels, and future career aspirations. (Google Forms)
* Cleansing: For the cleansing of data, Automated tools should be used for data cleansing to correct or remove corrupt, inaccurate, or extraneous data. The Data must be verified for accuracy and completeness before analysis. (Python and Excel)
* Visualization: Tools like PowerBI, Tableau and Python should be used for creating interactive dashboards. (Python)
* Clustering: Implement clustering algorithms to segment the student data. Clustering must identify meaningful patterns or groups within the data based on predefined criteria. (Python)
* Text Analytics: Text analytics must include keyword extraction, word cloud and clusters. Provide insights into common Career options and the required skills. Predictive Regression and Classification models to help in decision-making by categorizing data into predefined groups based on learned patterns (Python & Knime)

## Project Plan

Throughout the Student Survey Analysis project, our team employed a well-structured approach to ensure successful completion and timely delivery of all milestones. We initiated the project with a comprehensive kick-off meeting, where we outlined the objectives, assigned roles, and established clear communication protocols. Our project plan was meticulously documented, detailing the timeline, resources, and deliverables for each phase, from survey creation to text analytics. We used project management tools like Microsoft's 365 schedule to track progress, assign tasks, and adhere to deadlines effectively. Regular check-ins and milestone reviews were conducted to assess progress, address any challenges, and refine strategies, as necessary. All project documentation, including data, findings, and insights, was consistently updated, and shared on a common platform, ensuring seamless collaboration.

**



*In the Microsoft Project, you may include the following:*

* *List of tasks need to be assigned to a particular team member*
* *Ensure that there is adequate and fair involvement by each team member*
* *Project Plan must take into consideration the deliverables at specific milestone such as the 3 meetings*

## Task Allocation

|  |  |
| --- | --- |
| Task | Member Contribution (Based on Gantt Chart) |
| Survey Questionnaire |  |
| Data Cleansing |  |
| Business Questions & Data Visualisation |  |
| Clustering |  |
| Text/Predictive Model |  |

# Project Approach

## Methodology

* The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology is a structured approach for data mining projects. It consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.
* We choose CRISP-DM for our project because it provides a clear and organized framework, ensuring a thorough understanding of the business problem, systematic data handling, and robust model development, leading to reliable and actionable insights.

**3.2 Survey Questionnaire**

|  |  |
| --- | --- |
| **Team Member** | **Survey Questions** |
| Peh Yi Feng Javier | Section 1:   1. **What is your gender?**  * Male * Female  1. **How old are you this year?** (\*)  * 17 * 18 * 19 * 20 * 21 * 22 * Other…  1. **Which tertiary education institution are you in?** (\*)  * ITE * JC * University * Polytechnic  1. **Of what ethnicity are you?** (\*)  * Chinese * Malay * Indian * Eurasian * Other…  1. **Of what religion are you?** (\*)  * Islam * Buddhist * Taoist * Christian * Catholic * Free thinker * Hindu * Sikh * Other…  1. **Do your parents work?** (\*)  * Father Only * Mother Only * Both Work * Both don't work  1. **How many siblings do you have?** (\*)  * 0 * 1 * 2 * 3 * 4 * Other…  1. **What type of housing do you live in?** (\*)  * HDB * Condominium * Bungalow * Terrace * Studio apartment * Dorm * Other…   Section 2 (School Performance)   1. **What is your current cGPA? Please answer in 2 decimal points (e.g., 3.30)** (\*) 2. **What is your desired cGPA? Please answer in 2 decimal points (e.g., 3.30)** (\*) 3. **On average, how many times are you absent for class in a semester?** (\*)  * 0 * 1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 * 9 * 10 * Other… |
| Daniel Sanjay Sahdnani | Section 3 (Gaming):Do you play video games? (\*)  * + Yes   + No  Section 4 (Gaming):How many hours do you game per week? (\*)  * + 84   + 70 - 84   + 60 - 69   + 50 - 59   + 40 - 49   + 30 - 39   + 20 - 29   + 10 - 19   + < 9  What game genre do you mostly play? (\*)  * + Action (FPS, Shooter etc)   + Adventure   + Role-Playing   + Simulation   + Strategy   + Sports   + Battle Royale   + Survival   + Horror   + Rhythm   + MOBA   + Other…  Do you play games with friends? (\*)  * + Yes   + No  During what hours do you usually start playing video games? Please answer in the 24-hour time format (e.g., 0000, 1800, 2200 etc) (\*)Do your parents stop you from gaming? (\*)  * + Yes   + No  What is the main reason you started gaming? (\*)  * + Entertainment   + Escape from reality   + To make new friends   + A career option   + Influenced   + Other…   Section 5 (Part-Time Work):   1. **Are you working as a part-timer as of right now?** (\*)  * Yes * No   Section 6 (Part-Time Work):   1. **How many hours do you work a week?** (\*)  * 35 * 31 - 35 * 26 - 30 * 21 - 25 * 16 - 20 * 11 - 15 * 6 - 10 * < 5  1. **What job sector do you work in?** (\*)  * Retail * Food & Beverage * Healthcare * Education * Manufacturing * Warehouse * Entertainment * Wholesale & Trade * Other…  1. **Which days of the week do you work?** (\*)  * Monday * Tuesday * Wednesday * Thursday * Friday * Saturday * Sunday |
| Liaw Hong Da | Section 7 (CCA):   1. **Do you participate in any Co-Curricular Activities (CCAs) at your school?** (\*)  * Yes * No   Section 8 (CCA):   1. **How many Co-Curricular Activities (CCAs) are you currently participating in?** (\*)  * 1 * 2 * 3 * 4 * 4  1. **Please specify the name(s) of the CCA(s) you are involved in.** (\*) 2. **What role do you hold in your CCA(s)?** (\*)  * Member * Leader/President * Assistant Leader/Vice President * EXCO * Student Advisor * Other…  1. **How many hours per week do you dedicate to your CCA(s)?** (\*)   Section 9 (Satisfaction Level):   1. **On a scale of 1 to 10, how satisfied are you with your overall school experience? (1 being very unsatisfied, 10 being very satisfied)** (\*)  * 1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 * 9 * 10  1. **Which specific area of your school experience do you feel most satisfied with?** (\*)  * Results * Courses * Classmates * Lecturers * Social Life * Study Life * None * Other…  1. **Which specific area of your school experience are you not satisfied with?** (\*)  * Results * Courses * Classmates * Lecturers * Social Life * Study Life * None * Other…  1. **How do you rate the support provided by your school in achieving your personal and academic goals? (1 being very poor, 10 being excellent)** (\*)  * 1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 * 9 * 10  1. **Do you feel that your school fosters a sense of community and belonging among its students?** (\*)  * Yes * No |
| Tan Hong Bin | Section 10: Career & SkillsDo you have your future career planned already? (\*)  * + Planned   + Not yet   + Undecided  Section 11: Career & SkillsWrite down at least 3 career options after graduation (e.g., 1. teacher 2. chef 3. doctor) (\*)Section 12: Career & SkillsWrite down if there is any 3 career options after graduation that you would be interested to do (e.g., 1. teacher 2. chef 3. doctor) (\*)Section 13: Career & SkillsWith the career options that you have stated, briefly give the skills that are required for the options. (e.g., 1. Patience 2. Cooking 3. Precision) (\*) |

## 4 Business Understanding

## 4.1 Business Questions

|  |  |
| --- | --- |
| **Team Member** | **Business Questions** |
| Daniel Sanjay Sadhnani | 1. How many hours per week do students with a GPA of 3.0 or higher game? 2. How many hours per week do students with a GPA of 3.0 or higher work? 3. How many CCA(s) do students with 3.0 or higher GPA take? 4. How do class absences impact students' academic perform? 5. How Does game timing affect students GPA? |
| Peh Yi Feng Javier | 1. Is there a correlation between non-gaming students and higher GPAs compared to their peers who engage in video gaming? 2. Which video game genres are most popular among male and female students respectively? 3. Does a higher GPA correlate with increased satisfaction with the school experience compared to students with lower GPAs? 4. What are the prevalent trends in gaming among students who play video games with friends versus those who play alone? 5. What are the primary reasons students begin playing video games? 6. What are the common areas of dissatisfaction among students regarding their school experience? 7. What percentage of students hold part-time jobs? 8. What is the distribution of CCA participation among students? |
| Liaw Hong Da | 1. What is the correlation between the number of CCAs participated in and the current GPA? 2. How does the time spent on CCAs (CCA Hours) affect the current GPA? 3. What is the distribution of CCA Roles among students? |
| Tan Hong Bin | 1. How does the Satisfaction Level of Students affect their average GPA? 2. How does the number of hours student work affect their average GPA? 3. Does participating in CCA affect the average GPA of students? 4. What effect does game timing have on the student's average GPA? |

# 5 Data Understanding

* The dataset comprises 260 records (rows) and 41 fields (columns), containing various attributes. These attributes include demographic information such as Gender, Age, Ethnicity, and Religion; academic performance indicators like Current GPA, Desired GPA, and Absences; and additional details about career interests and skills. Understanding these basic properties is crucial for guiding further analysis and ensuring all relevant data is considered.
* Digging deeper into the data involves querying, visualizing, and identifying relationships among the attributes. Initial descriptive statistics were generated to provide an overview of the dataset, highlighting key attributes like Gender, Age, Ethnicity, Current GPA, Desired GPA, and Absences. These statistics help to understand the distribution and central tendencies of the data. We use both tableau and Python to create charts for us to understand how the various attributes affect Current GPA.
* Significant missing values were found in multiple columns, and it will be necessary for us to clean those gaps. Additionally, there were multiple spelling errors and unusable values to be removed. We plan to achieve these through knime and python.

# 6 Data Preparation

|  |  |
| --- | --- |
| **Team Member** | **Data Cleansing/Preparation Action** |
| Daniel Sanjay Sadhnani | Knime:   * filtered out columns that are not in use (timestamp column) * Replaced? values with no * filtered out rows that had invalid data/responses * replaced certain data with another type of data (e.g. 1 half sibling 4 siblings to 5 siblings) * set string values in GPA column to double * set GPA to only have values between 0 to 4(filtered out the rest) * replaced string values in int/double columns with mean of data * replaced null values with something more appropriate (e.g. for game hours, if null replace with 0)   Python:   * filtered out columns that are not in use (timestamp column) * replaced? values with no * filtered out rows that had invalid data/responses * replaced certain data with another type of data (e.g. 1 half sibling 4 siblings to 5 siblings) * set string values in GPA column to double * set GPA to only have values between 0 to 4(filtered out the rest) * replaced string values in int/double columns with mean of data * replaced null values with something more appropriate (e.g. for game hours, if null replace with 0) |
| Peh Yi Feng Javier | * For Game genre, kept some main genres and classified minor ones or self-written ones as others * Cleansed Career Option, Interested Career and Required Skills removed foreign entries '$', '?' etc. * Split Career Option, Interested Career and Required Skills, into 3 different columns * Filled empty cells with 'Nan' |
| Liaw Hong Da | Knime:   * filtered out unusable data * filtered out useless data * reformat data that are need |
| Tan Hong Bin Darius | Knime:   * Reformatted data * Filtered unusable data |

# 7 Modelling

**7.1 Data Visualisation**

|  |  |  |
| --- | --- | --- |
| **Team Member** | **Business Question** | **Visualisation Charts** |
| Daniel Sanjay Sadhnani | 1. How many hours per week do students with a GPA of 3.0 or higher game? 2. How many hours per week do students with a GPA of 3.0 or higher work? 3. How many CCA(s) do students with 3.0 or higher GPA take? 4. How do class absences impact students' academic perform? 5. How Does game timing affect students GPA? | 1) 2) 3) 4) 5) |
| Peh Yi Feng Javier | 1. Comparison of the Average GPA of students that play video games to those who do not 2. Popularity of Game Genres Among Male and Female Students 3. Distribution of students that play games with friends and those who play alone 4. Average Current GPA by School Satisfaction Level 5. Distribution of the Main Reasons Students Started Gaming (Refined Categories 6. Distribution of students unsatisfied area in school 7. Distribution of Students with and without Part-Time Jobs 8. Distribution of the Number of CCA(s) Among Students | Inserting image... |
| Liaw Hong Da | 1. What is the correlation between the number of CCAs participated in and the current GPA? 2. How does the time spent on CCAs (CCA Hours) affect the current GPA? 3. What is the distribution of CCA Roles among students? | Inserting image... |
| Tan Hong Bin | 1. How does the Satisfaction Level of Students affect their average GPA? 2. How does the number of hours student work affect their average GPA? 3. Does participating in CCA affect the average GPA of students? 4. What effect does game timing have on the student's average GPA? | 1.  2.    3.  Inserting image...  4.  Inserting image... |

**7.2 Clustering**

|  |  |  |
| --- | --- | --- |
| **Team Member** | **Clustering Axis** | **Clustering Output** |
| Daniel Sanjay Sadhnani | 1. Agglomerative Clustering X-Axis – Current GPA   Y-Axis – Absences   1. Affinity Propagation Clustering   X-Axis – Current GPA  Y-Axis – School Satisfaction Level   1. BIRCH Clustering   X-Axis – Current GPA  Y-Axis – CCA Hours | 1)2)3)Inserting image... |
| Peh Yi Feng Javier | 1. K-means clustering   X-axis – Current GPA  Y-axis – Absences  2) K-means clustering  X-axis – Current GPA  Y-axis – Game Hours |  |
| Liaw Hong Da | 1. DBSCAN Clustering for Current GPA, Absences:   X-axis- Absences  Y-axis-Current GPA  2) DBSCAN Clustering for Desired GPA, Absences:  X-axis- Absences  Y-axis-Desired GPA  3) Mean Shift Clustering for Age and Current GPA:  X-axis- Age  Y-axis-Current GPA |  |
| Tan Hong Bin | 1. Agglomerative Cluster for Current GPA, Absences:   X-axis: Absences  Y-axis: Current GPA   1. Agglomerative Cluster for Desired GPA, Absences:   X-axis: Absences  Y-axis: Desired GPA | 1.    2. |

**7.3 Text Analysis/Text Clustering**

|  |  |
| --- | --- |
| Text Mining  (Javier) | **Word Cloud:**  Planned Career Option  Interested Career Option  Required Skills for the Jobs  **Frequency Chart:**  Frequency of mentioned planned career option    Frequency of mentioned Interested Career Option    Frequency of mentioned Required Skills for the Jobs |
| Text Clustering  (Hong Da) | **DBSCAN Clustering of Career**    DBSCAN Clustering of Skills |

**7.4 Predictive Analysis**

|  |  |  |
| --- | --- | --- |
| **Team Member** | **Predictive Analysis Model** | **Model Accuracy** |
| Daniel Sanjay Sadhnani | Linear Regression Model (Knime) |  |
| Peh Yi Feng Javier | Linear Regression Model (Python) |  |
| Liaw Hong Da | Classification Logistic Model (Python) | **School Satisfaction Level:**    **Future Planned:** |
| Tan Hong Bin | Classification Logistic Model (Knime)  (Future Planned) |  |

# 8 Conclusion & Recommendations

**Conclusion:**

* **Recap:** The primary objective of the Student Survey Analysis project was to derive meaningful insights from a wide array of student data, encompassing demographics, school performance, gaming habits, part-time work, co-curricular activities (CCAs), satisfaction levels, career aspirations, and required skills. Utilizing the collected information, we performed extensive data cleansing, visualization, clustering, and text mining to deepen our understanding of student behaviours and needs. Additionally, we developed predictive models aimed at leveraging analytics to foster a more responsive, effective, and student-centred educational environment. This comprehensive approach not only highlighted key patterns and trends but also equipped us with actionable insights to enhance educational strategies and student engagement.
* **Summary of Key Findings:** Of the 260 responses gathered, 201 of the data were used in our final evaluation left as 59 of them were cleansed out due to the missing of significant values, data integrity, and duplicates. Removing these rows helps in cleaning the data, ensuring that the subsequent analyses are meaningful, and the findings are reliable. Of the 201 responses, our visualization provided information on students' profile in relation to gaming, part-time, CCA and satisfaction level.
* **Gaming:** we found out 83.6% students play video games and interestingly, we found out that the average GPA of students that play video games is higher compared to those that who do not play video games. For the most popular game genre, it was Action, tallying up to 57, 50 of which came from the males and 7 from females. The more popular game genre amongst females were classified under ‘others’ coming in at 14 responses.
* **Satisfaction:** The average GPA of each satisfaction level (1-10), people who has their School Satisfaction as a 10/10 has the highest average GPA (3.24) and for the lowest average GPA, it came from those which rated the satisfaction level 1/10(2.44). For the unsatisfied area students have about school life, most students are unsatisfied about their results, coming in at 37.3% and the lowest coming from classmates which they are unsatisfied about at 5.5%.
* **CCA:** 48.3% of students do not have CCA while the most common amount of CCA to have been one, at 31.8%. It is not common for students to have more than one CCA with 19.9% of students having two or more CCAs. 59.6% of students that have CCA(s) hold the role of member in their CCA while 23.15 are EXCOs, 10.6% being a leader/president and 6.7% are student advisors.
* **Part-time:** 69.2% of students do not have part-time jobs while 30.8% of students have part-time jobs. Students that work part-time, those that work 5 or lesser hours in a week has a higher average GPA to those that work more hours in a week.
* **Interpretation of Findings:**
* **Impact of Gaming on Academic Performance**: The higher average GPA among students who play video games challenges common perceptions regarding gaming and academics. This suggests that engagement in video games might be associated with skills that positively influence academic performance.
* **Role of Satisfaction in Academic Success**: The correlation between higher satisfaction levels and higher GPAs underscores the importance of a supportive and fulfilling educational environment. This highlights the potential benefits of initiatives aimed at improving student satisfaction.
* **Involvement in CCAs**: With a significant portion of students not participating in CCAs, there might be an untapped opportunity to enhance student engagement and skills development through these activities. The data suggests that a moderate level of involvement (one CCA) is the most common, potentially indicating a balance between the benefits of extracurricular engagement and the demands of academic responsibilities.
* **Part-Time Work and Academic Performance**: The finding that students working fewer hours per week have higher GPAs suggests that while part-time work provides valuable experience and financial benefits, excessive working hours might detract from academic success.
* **Impact of Career Planning on School Satisfaction**: Students who have planned careers report greater average satisfaction levels than those who have not made any preparations for their careers. This emphasizes how having the right career is to raising student satisfaction and improving their entire educational experience.
* **Correlation between GPA and School Satisfaction:** A positive connection can be shown in the chart examining the relationship between GPA and school satisfaction. In comparison to students with lower GPAs (<2.0), those with higher GPAs (3.5-3.9 and 4.0) report better levels of school satisfaction. This demonstrates how much a student's pleasure with their educational experience is influenced by their academic performance.
* **Effect of Part-Time Work on Academic Performance:** Students who work 16–20 hours per week have the highest average GPA, according to the bar chart comparing hours worked per week to current GPA. On the other hand, the average GPA of individuals who work more than 35 hours a week is the lowest. This implies that while working part-time can have advantages, working long hours can have a detrimental effect on one's scholastic achievement.

**Recommendations:**

* **Actionable Steps:**

Based on our findings, there is a positive correlation between gaming and academic performance so we can use it as a leverage. Another point, we found out is that we could enhance student satisfaction.in other words, we can improve student satisfaction to boost their academic performance. Moving on, if we want to increase student engagement and skills development, we could promote and facilitate more CCA participation. Lastly, we could ensure students part time work does not have any negative impact on their academic performance. With all the listed findings, we have the following suggested actionable steps.

**Gamification in Learning**

* + We could have some integration of gaming into learning. This way while students have learnt their materials from classes, they can apply it through playing games such as Kahoot.

**Monitor Gaming Habits**

* + Have a system that monitors and guide students in their gaming habits to ensure that they have a healthy balance with academic responsibilities.

**Conduct Regular Surveys**

* + Gather feedback on students' satisfaction with the school and identify the areas for improvement. This way when students are satisfied, they tend to do better in academic performance wise.

**Recognise and Reward**

* + Have a recognition system where it will celebrate all student achievement and contributions, this way it will foster a positive school environment.

**Flexible Scheduling**

* + Due to students having academic and personal commitments, flexible scheduling of CCA could help accommodate the students allowing them to have better time management.

**Worklife Balance Workshops**

* + Provide workshops on time management and WorkLife balance to help students manage their time and responsibilities effectively.

**Guidelines on Working Hours**

* + Create a guideline on the maximum recommended hours for students to work. This will give students a guide to not over work themselves and letting this WorkLife imbalance affect their studies.
* **Prioritization of Actions:**

|  |  |  |
| --- | --- | --- |
| Actions | Priority | Reason |
| Enhancing Students’ School Satisfaction Initiatives | High | It directly correlates with higher academic performance and overall, well being |
| Game Inclusive Educational Programmes | Medium | It capitalises on students’ interest to enhance the engagement and learning outcomes |
| Promoting and Facilitate CCA Participation | Medium | It enhances student engagement and skill development |
| Regulating part time work | Low | Although it is important to maintain academic performances, but the more important impact should be focused on students’ satisfaction and engagement initiatives |

* **Future Considerations:**

Studies on gaming and academic performance: We can assess the long-term effects of gaming on academic and other aspects of student life. This way we can find out how the integration of gaming in learning will improve students’ learning.

Expanding the scope of satisfaction: We can identify and address factors that affects students’ satisfaction and performance. This way we could find out what students are not satisfied with and make amendments.

Impact of CCAs on skill development: We can conduct studies to compare the impact of the carious CCAs on students to understand how participation of the different CCAs influences skill development and academic success.

Effect of Part Time work: We can research on how part time work experiences contribute to skill development.

* **Implications:**
* Game: Educational strategies could potentially integrate game-based learning to capitalize on these engagement levels.
* Satisfaction: Enhance teacher-student relationships, offering more personalized learning experiences, and addressing specific areas of dissatisfaction like academic results.
* CCA: Strategies to encourage CCA participation could include making activities more accessible, relevant, and aligned with students' interests.
* Part-time: Excessive working hours might detract from academic success. This could inform policies on student employment, suggesting a need for guidelines or support systems to help students balance work and study effectively.

# 9 User & Technical Documentations

**9.1 User Documentation**

### 1. Introduction

**Purpose**: The purpose of this predictive model is to analyse and cluster career and skill data using the DBSCAN algorithm.

### 2. Data Description

**Data Source**: The data was collected from the last 9 columns of the survey. The survey included questions on career interest/options, and the skills required.

**Data Format**: The data was provided in an excel file and included text responses for career aspirations and skills.

**Key Attributes**:

* Career Data: Jobs that the respondent aspires to do.
* Skill Data: Descriptions of skills required.

**Data Preprocessing**: Text data was converted into numerical format using TF-IDF vectorization. Principal Component Analysis (PCA) was used to reduce dimensionality for visualization purposes.

### 3. Model Development

**Algorithm**: The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm was used for clustering.

**Training Process**: The DBSCAN algorithm was applied to the TF-IDF transformed data. Parameters such as eps (maximum distance between two samples for them to be considered as in the same neighbourhood) and min\_samples (minimum number of samples in a neighbourhood for a point to be considered a core point) were tuned for optimal clustering.

**Hyperparameters**:

* eps: 0.5
* min\_samples: 2

### 4. Usage Instructions

**Input Requirements**: Input data should be in a structured format, containing text descriptions of career aspirations and skills.

**Running the Model**:

* Preprocess the text data using TF-IDF vectorization.
* Apply the DBSCAN algorithm with appropriate parameters.
* Visualize the clusters using PCA for dimensionality reduction.

**Interpreting Output**: The model outputs clusters of careers and skills, which can be visualized in 2D plots. Each point represents a data sample, and colours indicate cluster membership.

### 5. Model Performance

**Evaluation Metrics**: The silhouette score was used to evaluate the clustering performance. This score measures how similar an object is to its own cluster compared to other clusters.

**Results:**

#### 1. DBSCAN Clustering of Career Data

* **Clusters Identification**: The chart shows several distinct clusters identified by different colours. Each cluster represents a group of similar career aspirations based on the textual descriptions provided by the respondents.
* **Cluster Cohesion**: The silhouette score for the career data clustering is 0.17, indicating moderate clustering quality. This suggests that while some groups are well-defined, others may overlap or contain noise.
* **Outliers**: There are points that appear isolated from the main clusters, which DBSCAN identifies as noise. These points represent career aspirations that do not fit well into the main clusters.
* **Principal Components**: The x-axis and y-axis represent the first and second principal components, respectively. These components are derived from PCA and help in visualizing the high-dimensional TF-IDF data in a two-dimensional space.

#### 2. DBSCAN Clustering of Skill Data

* **Clusters Identification**: The chart shows multiple distinct clusters, each indicated by different colours. Each cluster represents a group of similar skills based on the textual descriptions provided by the respondents.
* **Cluster Cohesion**: The silhouette score for the skill data clustering is 0.63, indicating good clustering quality. This suggests that the groups of skills are well-defined and more cohesive compared to the career data clusters.
* **Outliers**: Like the career data, there are isolated points identified as noise. These represent unique or less common skills that do not fit well into the main clusters.
* **Principal Components**: The x-axis and y-axis represent the first and second principal components, respectively. These components are derived from PCA and help in visualizing the high-dimensional TF-IDF data in a two-dimensional space.

# 10 Reflections

**10.1 Daniel Sanjay Sadhnani**

Looking back at the entire process of our Student Survey Analysis project, I can quite honestly say it was quite the emotional rollercoaster. In my case I joined this FYP team with short notice as i was originally supposed to go for Internship in sem 1 but after failing to seek a company, i opted to do my FYP in semester 1 instead. I knew 2 of my team members Hong Da and Darius after working with them in the year 2 semester 2 class but i had not met my team coordinator Javier yet.

Since the first meeting of our FYP all the way to the mid-sem break, I honestly was very proud of my team. We took each task seriously and wasted no time moving on to the next task with no forms of procrastination. We would ask questions from our supervisor if we had any doubts and none of us hesitated to expand our learning in the hopes of doing well for this FYP Project. But the one teammate I want to commend would be my Team Coordinator Javier. Not only would he follow up with our supervisor, but he would also often bring up any questions that we had regarding any tasks to the supervisor. He would also make sure we were following the deadlines strictly and for that i think he truly deserves merit.

After the Mid-Term Evaluation, this was where the roller coaster had hit the low. We thought we were making such good progress until we learned from our evaluators that in order to achieve a higher grade for FYP, we would need to include some form of programming in our project as the tools we were using for data cleansing & visualisation were little or no code softwares such as KNIME and Tableau. But did we let this set back stop us? Never.

During the mid-sem break, we decided to expand our learning by watching YouTube tutorials and websites on how we could do Data Cleansing and Data Visualisation Using python with Pandas. To our surprise, despite us thinking it was going to be difficult, after taking the time and effort to learn and understand, we realised it was not as difficult as we thought it would be. And we did not stop our learning there. We also applied coding with tasks like Clustering, Text Analysis/Clustering and Predictive Modelling which to our surprise worked like a charm.

In Conclusion, reflecting on our Student Survey Analysis project, it was an emotional rollercoaster. Joining the team on short notice and getting to know new members, including our standout coordinator Javier, was just the beginning. Initially proud of our efficient progress, we faced a setback when we needed to incorporate programming to achieve a higher grade. Undeterred, we embraced this challenge, quickly mastering Python and Pandas for data analysis. This new skill set not only improved our project but also proved more manageable than anticipated, leading to a successful outcome.

**10.2 Peh Yi Feng Javier**

Reflecting on our Student Survey Analysis project, I can confidently say it was a demanding yet rewarding journey that pushed us to our limits, both mentally and physically. The project was fraught with uncertainties, often leading us to seek clarifications from our supervisor regularly. Initially, the feedback we received during our mid-semester evaluation was quite challenging, as it was negative. This critique focused on our use of Knime and Excel for data cleansing, which led to a significant decision: restarting our project from scratch during our one-month holiday. This included reopening our survey to gather more data, adding another layer of complexity to our task.

In the second half of the semester, the demands of the project intensified. The second half was especially tough due to deadline coming closer and more stringent specifications. We found ourselves having to shift much of our workload to Python, a platform we were not familiar with. This required us to quickly learn new skills in data cleansing, visualization, and clustering skills we previously honed on different platforms like Tableau and Knime. Moreover, we had to delve into areas such as text mining and creating predictive models, which were new to us.

Despite these challenges, the dedication and teamwork of my groupmates were pivotal. Everyone was committed, regularly making time for meetings, and even traveling to school for discussions, despite living far from the school. The amount of commitment ensured that everyone was pulling their weight and contributing effectively, which made the project more manageable.

Looking back, this project not only improved and taught us more our technical skills but also strengthened our capabilities in teamwork and problem-solving. Overcoming these challenges together was incredibly fulfilling, deepening our understanding of data analytics' practical applications, and highlighting the value of resilience and collaboration in achieving complex objectives.

**10.3 Liaw Hong Da**

Looking back on my final year project, I'm happy to say that our group showed remarkable efficiency and collaboration. This is because everyone was dedicated to their roles, assignments were often finished on schedule. Our proactive approach to assigning tasks and our mutual support system made sure that we never missed a deadline, including the deadline for submitting deliverables. Our success was due to the team's cohesiveness and spirit of cooperation, and I am appreciative of the opportunity to have worked with such a committed bunch of people.

However, there were some obstacles along the way. We encountered an unforeseen difficulty during the mid semester evaluation when we were instructed by our evaluators to redo our work to incorporate Python coding for data cleansing and visualization, even though this was not originally specified in the project details. This was a huge obstacle for us because we had previously finished our assignments with the tools that was provided to us, KNIME and Tableau. As a result, we were forced to learn Python and redo our work during the school holidays, which at the time felt unfair and pointless. Still, we persisted, adjusting to the new specifications, and eventually getting past this obstacle.

Additionally, we also had trouble persuading people to fill out the survey that was required for our data analysis. We had a terrible time initially getting enough responses, which prevented us from beginning the critical steps of data cleansing and visualization. The timeliness and general success of our project were seriously jeopardized by the lack of data. Thankfully, our supervisor intervened and made it possible for the survey to be distributed to more people in RP (Republic Polytechnic). Our response rate increased dramatically because of this intervention, and we were able to gather enough information to continue with our research.

Furthermore, we were given more challenging tasks like clustering and creating predictive models. We were still learning how to apply these ideas in real-world situations even though we had only recently learnt about them. Nevertheless, this assignment proved to be an excellent chance for us to review the business analytics material from our C378 Module.

On the bright side, these challenges proved to be a worthwhile educational experience. I learned how to use Python to cleanse and visualise data, which will come in very handy for my next AI (Artificial Intelligence) and Machine Learning module (C338). My skill set expanded because of this, increasing my versatility in data analysis. I also learned about the Mean Shift clustering algorithm, a non-parametric technique that can effectively identify clusters in my data by iteratively shifting data points towards the mode of the distribution. This project also taught me resilience and the importance of adaptability in the face of obstacles.

Overall, this has been a fulfilling experience that has advanced my professional growth and improved my comprehension of data analysis. It even helped me secure internship opportunities, as interviewers were impressed by my project and the depth of my understanding. This reflection emphasizes the value of resilience and the unanticipated advantages of conquering obstacles.

**10.4 Tan Hong Bin Darius**

Looking back at the whole process of the Student Survey Analysis project, I am proud of whatever we could achieve up to this point. The group had all their dedicated job for cleansing, data visualization and clustering. All the jobs were finished on time and on schedule. I feel that with all the communication we had in the group made us not to miss any deadlines which made us on track.

However, I feel that there were some miscommunications and misunderstandings with the instructions stated in LEO and what the Mid Semester Evaluators commented. We were told that we had to incorporate Python coding for the data cleansing, visualizations though the instructed details did not state to include Python coding. Due to this change that we have to do, we found ourselves at a disadvantage as we were originally on schedule with all of our work. However, we still pushed through all our the hardships and at the end we got what we wanted to achieve.

Another point that may have slowed our progress was that the collection of surveys. At the beginning we thought maybe sending to our friends would help and get the responses quick but one thing that we overlooked was that most of them would not even give their response or even when given response its not the truthful ones. One thing that we can improve on should be the formatting of the question. But with all these problems we were still able to the quite a number of responses as we had help from our supervisor to get clearance to send to the whole school to get responses.

Overall, I feel that this experience to get a hands on Final Year Project is very effective to see if we learnt anything from our lessons. Reason being that partially of the Python coding that we had to incorporate, we were learning some from our Module C338 AI and Machine Learning. If we can complete this FYP with the incorporation of the Python codes it means that we have learnt and self taught ourselves with knowledge that exist but if we did not find out we would not have known it.

# References

* Cleansing
  + <https://www.w3schools.com/python/pandas/pandas_cleaning.asp>
  + <https://www.youtube.com/watch?v=bDhvCp3_lYw&pp=ygUYZGF0YSBjbGVhbnNpbmcgb24gcHl0aG9u>
* Visualization
  + <https://www.geeksforgeeks.org/data-visualization-with-python/>
  + <https://youtu.be/JpSMse3eVVg?si=wvLynhEOIdPX1Zd4>
* Clustering
  + <https://www.youtube.com/watch?v=iNlZ3IU5Ffw&t=725s&pp=ygUZZGF0YSBjbHVzdGVyaW5nIG9uIHB5dGhvbg%3D%3D>
  + <https://www.w3schools.com/python/python_ml_k-means.asp>
  + <https://www.geeksforgeeks.org/clustering-text-documents-using-k-means-in-scikit-learn/>
* Text Mining
  + <https://www.youtube.com/watch?v=TvfUjW0SEUI&pp=ygUSdGV4dCBtaW5pbmcgcHlodG9u>
  + <https://www.kdnuggets.com/2020/05/text-mining-python-steps-examples.html>
* Predictive Modelling
  + <https://365datascience.com/tutorials/python-tutorials/predictive-model-python/>
  + <https://www.youtube.com/watch?v=Cx8Xie5042M&pp=ygUecHJlZGljdGl2ZSBtb2RlbGxpbmcgb24gcHl0aG9u>
* CRISP-DM
  + <https://www.datascience-pm.com/crisp-dm-2/>

# Appendices

|  |  |
| --- | --- |
| Survey | Edit: <https://docs.google.com/forms/d/1u_KHrdl_WmLE0oxD1GzhncMwnBB98_Me8im5c-tP9Vs/edit>  Questionnaire: <https://forms.gle/8KBXRPk4uPc3uFJTA> |
| Original data | <https://docs.google.com/spreadsheets/d/17hpUl6pp1PJ_-1TrY-YAQPp3Rrz0ZJkgl_gje1jWNTg/edit?resourcekey=&gid=1483497035#gid=1483497035> |
| Finalized Data |  |
| Daniel Codes(Cleansing, visualization, clustering, predictive model) |  |
| Javier Codes  (Visualization, clustering, text analysis, predictive model) |  |
| Hong Da Codes  (Visualization, clustering, text clustering, predictive model) |  |
| Hong Bin Codes  (Visualization, clustering, predictive model) |  |
| Dashboard |  |
| Gantt Chart |  |
| Slides |  |

# Project Poster

**