Exploring The Academic Path: Background Strands of 2nd Year CCS Students in Senior High School

Members:

Maravilla, John Ace

Feliciano, Eric John

Buag, Neil

Manzanilla, Cedrick

Tena, Jhed

Larano, Zeon Mac

January 5, 2024

I. Introduction

The project aims to investigate and visualize the academic paths of 2nd year students specializing in College of Computer Studies Department about their chosen strand in their Senior High School. The primary goals include gaining insights into the diverse backgrounds of students within this academic strand, understanding their educational trajectories, and identifying any patterns or trends that may influence their academic journeys.

The visualized data focuses on academic backgrounds of second-year CCS students in senior high school, covering factors like sex, age, course, year, section, SHS strand, and relevant details. The dataset, potentially came from online surveys such as G-Forms, aims to provide a comprehensive overview of the students' academic profiles.

Statement of the Problem

The data analysis intended to answer the following questions:

1. What is the demographic profile of the respondents in terms of:
    1.1 Age
    1.2 Sex
    1.3 Course, Year & Section
    1.4 SHS Strand

2. What are the assessments of the students in terms of:

   2.1 Strand Alignment Satisfactory,

   2.2 Learning Adaptability

3. How does previous strand affect the chosen college course of a student?

II. Data

1. Rows and Columns

The dataset comprises 100 rows, and includes 12 columns capturing key details such as sex, age, course, year, section, SHS strand and relevant details for second-year CCS students in senior high school.

2. Data Types of Each Variable

- **Sex:** String
- **Age:** Integer
- **Course_Year_Section:** String
- **SHS_STRAND:** String
- **Strand_Alignment_Q1 to Q3:** Integer
- **Learning_Adaptability_Q1 to Q5:** Integer

3. Missing Values

Missing values in numeric columns were filled with their respective means. Additionally, the 'AGE' column was converted to numeric format, and rows with negative age values were filtered out for data cleanliness.

III. Visualization Techniques

Throughout the data analysis, we have used the following types of visualizations:

1. Pie Charts

We have used pie charts to visually represent respondent demographics, including Sex, Age, Course, Year & Section, and SHS Strand, effectively conveying proportional breakdowns. The pie charts offer a clear and intuitive showcase of category distributions, enhancing clarity in understanding the surveyed population's characteristics. Color distinguishes demographic categories, while size and shape convey proportional breakdowns for a clear visual representation of the demographic profile in the pie charts.

2. Bar Graphs

We have used bar charts to visually assess the alignment between students' learning adaptability and chosen SHS strands, aiding quick comprehension of trends. This visual tool facilitates informed decision-making in education by presenting a clear relationship between SHS strands and learning adaptability. The straightforward representation of mean values in the charts enhances data interpretability for educators and administrators. Color in the bar chart distinguishes between overall mean values and educational strands, aiding visual differentiation. Size and shape represent the magnitude and specific questions for each learning adaptability, facilitating easy comparison and providing a clear understanding of the dataset.

Additional Libraries and Packages Used in Visualization

1. **Matplotlib**

This library is used for creating static, animated, and interactive visualizations in Python. In our code, it is utilized for generating bar charts and pie charts to visualize the data.

IV. Implementation in Google Collab

1. **Data Loading and Cleaning**

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('background_strand_dataset.csv')
df.fillna(df.mean(numeric_only=True), inplace=True)
df['AGE'] = pd.to_numeric(df['AGE'], errors='coerce')
df = df[df['AGE'] >= 0]
df.head(101)
```

Using Pandas, we loaded the dataset ('background_strand_dataset.csv') into a DataFrame called df. We cleaned the data by filling missing values with means in numeric columns, converting the 'AGE' column to numeric format, and filtering out rows with negative 'AGE'. The resulting df.head() displays the cleaned and processed dataset.

2. Visualization Construction

2.1 Pie Chart 1: Distribution of Respondents According to their Sex

```
sex_counts = df['SEX'].value_counts()
plt.figure(figsize=(6, 6))
colors = ['blue', 'pink']

print(sex_counts)

plt.pie(sex_counts, labels=sex_counts.index, autopct='%1.1f%%', startangle=90, colors=colors)
plt.legend(sex_counts.index, loc="upper right", bbox_to_anchor=(1, 0, 0.5, 1))
plt.title('Pie Chart 1: Distribution of Respondents According to their Sex')
plt.show()
```

This code snippet utilizes Matplotlib to create a pie chart, analyzing the distribution of respondents' sex in the DataFrame df. By calculating counts of unique values in the 'SEX' column and generating a labeled pie chart, it offers a clear overview of the gender distribution among the dataset's respondents.

2.2 Pie Chart 2: Distribution of Respondents According to their Age

```
age_counts = df['AGE'].value_counts()
age_counts = age_counts.sort_index()
colors = plt.cm.Paired(range(len(age_counts)))

plt.figure(figsize=(8, 8))
plt.pie(age_counts, labels=age_counts.index, autopct='%1.1f%%', startangle=90, colors=colors)
plt.legend(age_counts.index, loc="upper right", bbox_to_anchor=(1, 0, 0.5, 1))
plt.title('Pie Chart 2: Distribution of Respondents According to their Age')
plt.show()
```

Same as Pie Chart 1, this code generates a pie chart to illustrate the distribution of respondent ages in the DataFrame df. It calculates and displays the percentages for each age group, offering a visual representation of the diversity in ages within the surveyed population.

2.3 Pie Chart 3: Distribution of Respondents According to their Course, Year & Section

```
cys_counts = df['COURSE_YEAR_SECTION'].value_counts()
desired_order = ['BSIT 2A', 'BSIT 2B', 'BSIT 2C', 'BSIT 2D', 'BSIT 2E', 'BS COMSCI 2A', 'BS COMSCI 2B']

cys_counts = cys_counts.reindex(desired_order)
cys_counts = cys_counts.dropna()
colors = plt.cm.Paired(range(len(cys_counts)))

plt.figure(figsize=(8, 8))
plt.pie(cys_counts, labels=cys_counts.index, autopct='%1.1f%%', startangle=90, colors=colors)
plt.legend(cys_counts.index, loc="upper right", bbox_to_anchor=(1, 0, 0.5, 1))
plt.title('Pie Chart 3: Distribution of Respondents According to their Course, Year & Section')
plt.show()
```

This code snippet calculates counts of unique values in the 'COURSE_YEAR_SECTION' column of DataFrame df, reorganizes them, and generates a labeled pie chart with percentages and a title. The chart visually represents the distribution of respondents across various course, year, and section categories, offering insights into the composition of the surveyed population.

2.4 Pie Chart 4: Distribution of Respondent According to their SHS Strand

```
strand_counts = df['SHS_STRAND'].value_counts()
plt.figure(figsize=(8, 8))
colors = plt.cm.Paired(range(len(strand_counts)))

plt.pie(strand_counts, labels=strand_counts.index, autopct='%1.1f%%', startangle=90, colors=colors)
plt.legend(strand_counts.index, loc="upper right", bbox_to_anchor=(1, 0, 0.5, 1))
plt.title('Pie Chart 4: Distribution of Respondents According to their SHS Strand')
plt.show()
```

This code generates a pie chart using Matplotlib to visualize the distribution of respondents based on their Senior High School (SHS) strand. It calculates and displays the percentages for each SHS strand, offering a concise representation of the educational backgrounds within the surveyed population.

2.5 Bar Graph 1: Strand Alignment

```
overall_mean_values = df[['Strand_Alignment_Q1', 'Strand_Alignment_Q2', 'Strand_Alignment_Q3']].mean()
overall_weighted_mean = df[['Strand_Alignment_Q1', 'Strand_Alignment_Q2', 'Strand_Alignment_Q3']].mean().mean()
strands = ['ICT', 'STEM', 'HUMSS', 'GAS', 'TVL']
strand_mean_values = {}

print(f'\nWeighted Mean:')
print(overall_mean_values.round(2))
print(f'\nOverall Weighted Mean: ', overall_weighted_mean.round(2))
print()

bar_width = 0.1
bar_positions = np.arange(len(overall_mean_values))
plt.bar(bar_positions, overall_mean_values.values, width=bar_width, label='All Strands', alpha=0.7, color='black')

for i, strand in enumerate(strands):
    strand_df = df[df['SHS_STRAND'] == strand]
    strand_mean_values[strand] = strand_df[['Strand_Alignment_Q1', 'Strand_Alignment_Q2', 'Strand_Alignment_Q3']].mean()
    strand_overall_weighted_mean = strand_df[['Strand_Alignment_Q1', 'Strand_Alignment_Q2', 'Strand_Alignment_Q3']].mean().mean()
    plt.bar(bar_positions + (i + 1) * bar_width, strand_mean_values[strand].values, width=bar_width, label=strand, alpha=0.7)

    print(f'\n{strand} Weighted Mean:')
    print(strand_mean_values[strand].round(2))
    print(f'\n{strand} Overall Weighted Mean: ', strand_overall_weighted_mean.round(2))
    print()

print(f'\n')
combined_labels = list(overall_mean_values.index)
plt.xticks(bar_positions + (len(strands) / 2) * bar_width, combined_labels)
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.title('Bar Graph 1: Strand Alignment Satisfactory')
plt.xlabel('Questions')
plt.ylabel('Mean')
plt.show()
```

In this code, we calculate the overall mean values and weighted mean for strand alignment across three questions. We then iterate through specific Senior High School (SHS) strands, computing their mean values and overall weighted mean. The results are displayed in a bar graph, visually comparing the strand-wise alignment means and overall alignment means, providing insights into the alignment scores for different SHS strands.

2.6 Bar Graph 2: Learning Adaptability

```
overall_mean_values = df[['Learning_Adaptability_Q1', 'Learning_Adaptability_Q2', 'Learning_Adaptability_Q3', 'Learning_Adaptability_Q4', 'Learning_Adaptability_Q5']].mean()
overall_weighted_mean = df[['Learning_Adaptability_Q1', 'Learning_Adaptability_Q2', 'Learning_Adaptability_Q3', 'Learning_Adaptability_Q4', 'Learning_Adaptability_Q5']].mean().mean()
strands = ['ICT', 'STEM', 'HUMSS', 'GAS', 'TVL']
strand_mean_values = {}

print(f'\nWeighted Mean:')
print(overall_mean_values.round(2))
print(f'\nOverall Weighted Mean: ', overall_weighted_mean.round(2))
print()

bar_width = 0.1
bar_positions = np.arange(len(overall_mean_values))
plt.bar(bar_positions, overall_mean_values.values, width=bar_width, label='All Strands', alpha=0.7, color='black')

for i, strand in enumerate(strands):
    strand_df = df[df['SHS_STRAND'] == strand]
    strand_mean_values[strand] = strand_df[['Learning_Adaptability_Q1', 'Learning_Adaptability_Q2', 'Learning_Adaptability_Q3', 'Learning_Adaptability_Q4', 'Learning_Adaptability_Q5']].mean()
    strand_overall_weighted_mean = strand_df[['Learning_Adaptability_Q1', 'Learning_Adaptability_Q2', 'Learning_Adaptability_Q3', 'Learning_Adaptability_Q4', 'Learning_Adaptability_Q5']].mean()
    plt.bar(bar_positions + (i + 1) * bar_width, strand_mean_values[strand].values, width=bar_width, label=strand, alpha=0.7)

    print(f'\n{strand} Weighted Mean:')
    print(strand_mean_values[strand].round(2))
    print(f'\n{strand} Overall Weighted Mean: ', strand_overall_weighted_mean.round(2))
    print()

print(f'\n')
combined_labels = list(overall_mean_values.index)
plt.xticks(bar_positions + (len(strands) / 1.5) * bar_width, combined_labels, rotation=75, ha='center')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.title('Bar Graph 2: Learning Adaptability')
plt.xlabel('Questions')
plt.ylabel('Mean')
plt.tight_layout()
plt.show()
```

In this code, we personally analyze learning adaptability by calculating overall mean values and weighted mean across five questions. We iterate through specific Senior High School (SHS) strands, computing their mean values and overall weighted mean for learning adaptability. The results are presented in a labeled bar graph, facilitating a visual comparison of learning adaptability means for different SHS strands and overall scores.

Challenges

Throughout the development of our data analysis, we have faced the different challenges and overcome it in the following:

1. **Visualization of Data**

    1.1 Challenge

    Conveying complex information succinctly and selecting appropriate visualizations for diverse datasets.

    1.2 Solution

    We simplified visuals, focusing on key insights, and adapting visualization types to match the characteristics of the data for enhanced clarity.

2. **Interpretation of Patterns**

    2.1 Challenge

    Understanding complex data patterns.

    2.2 Solution

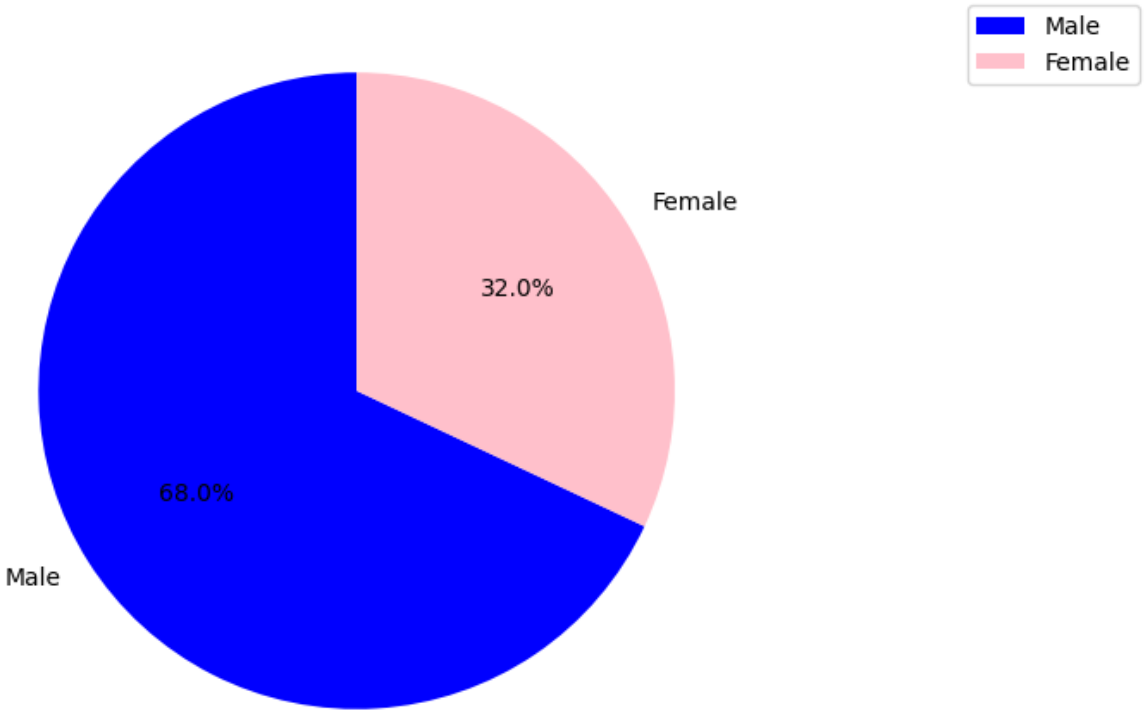
    We have used visualization techniques and collaborate with domain experts such as our seniors in school for deeper insights.

V. Results and Interpretation

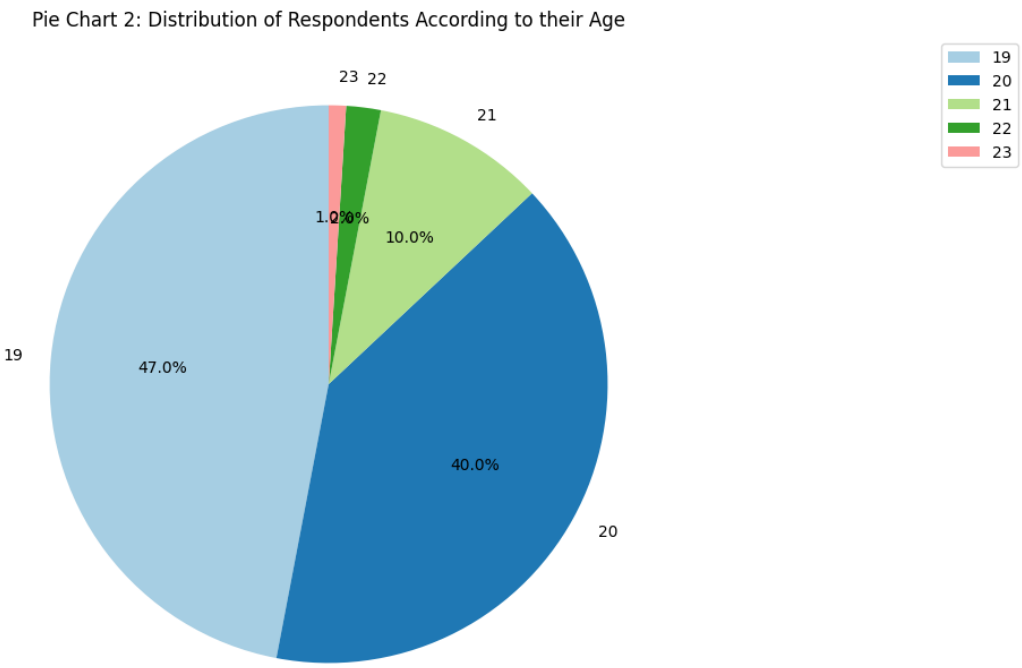
1. Profile of the Respondents

Pie Chart 1, Pie Chart 2, Pie Chart 3 and Pie Chart 4 present the demographic profile of the respondents of 2nd Year Students in CCS Department in Laguna State Polytechnic University – Sta Cruz Branch.

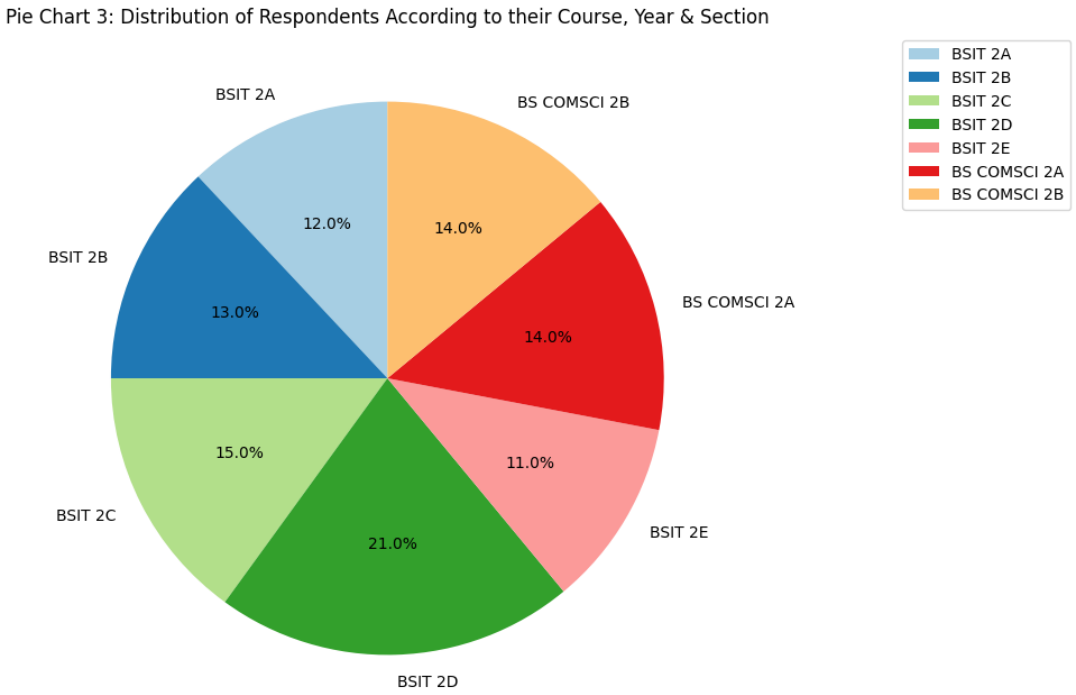
Pie Chart 1: Distribution of Respondents According to their Sex



According to data and the researchers gathered, it can be noted in Pie Chart 1, from 100 respondents 68% of them are male while 32% are female. This indicates that majority of respondents in 2nd year CCS students are male.

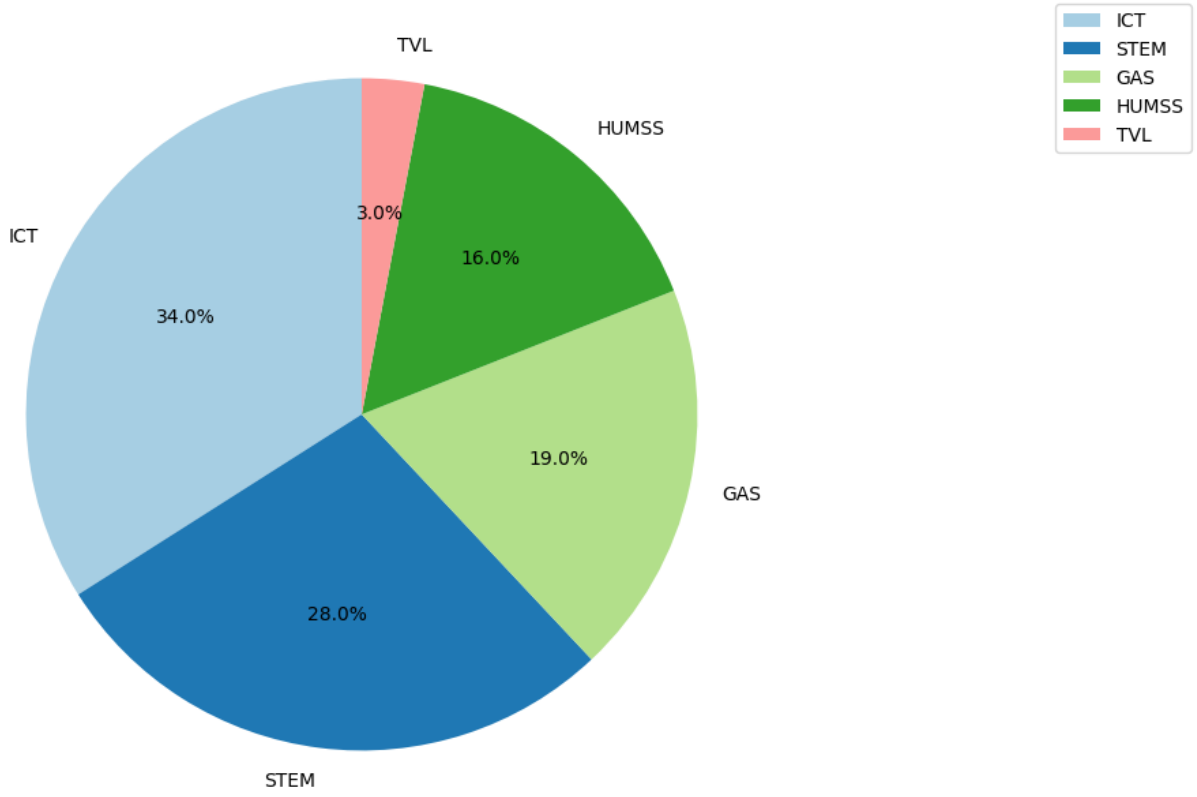


In Pie Chart 2, age of 19 years old have the highest with 47% while age of 23 have lowest with 1%. This indicates that majority of respondents in 2nd year CCS students are 19 years old.



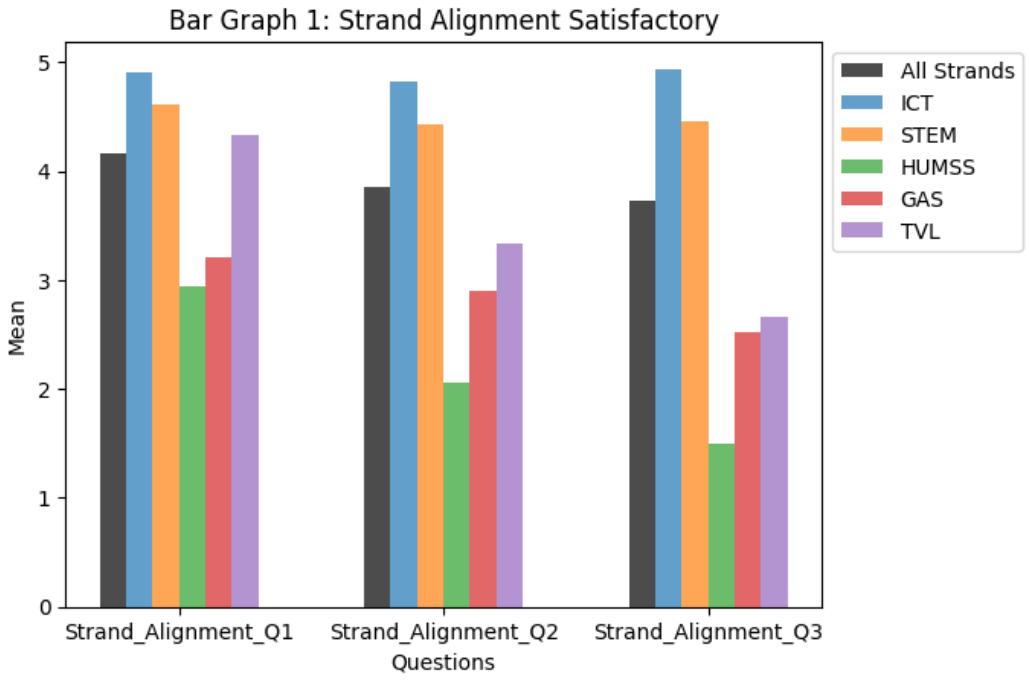
In Pie Chart 3, we can see that most of respondents came from BSIT 2D with 21% while the lowest with 11 % came from BSIT 2E. This indicates that majority of respondents in 2nd year CCS students came from BSIT 2D

Pie Chart 4: Distribution of Respondents According to their SHS Strand



As we can see in Pie Chart 4, it implies that most of respondents have a strand of ICT Strand in their Senior High School with 34%. While, the lowest 3% are respondents from TVL Strand in their Senior High School. On the other hand, this shows that about 62% of respondents came from CCS Related Course which is ICT and STEM. Lastly, 38% came from Non-Related Course which is TVL, HUMSS and GAS.

2. Data Analysis



Legend:

4.21 - 5.00 - Strongly Agree

3.41 - 4.20 – Agree

2.61 - 3.40 – Neutral

1.81 - 2.60 – Disagree

1.00 - 1.80 - Strongly Disagree

Questions:

Strand_Alignment_Q1: I am satisfied with my choice of SHS strand in retrospect.

Strand_Alignment_Q2: My SHS strand influenced my college course selection.

Strand_Alignment_Q3: My SHS strand aligns well with my chosen college course.

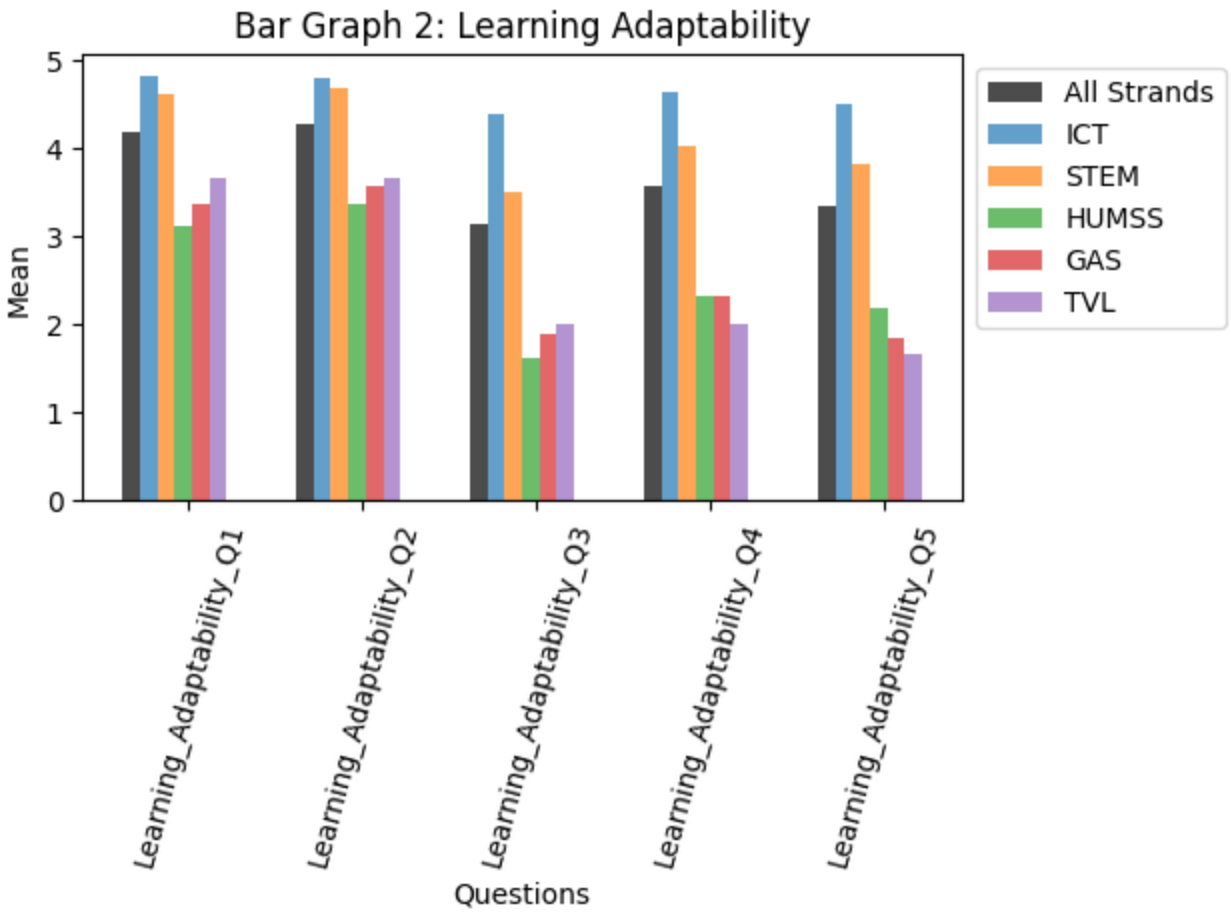
OVERALL WEIGHTED MEAN: 3.92

This Bar Graph 1 shows the Strand Alignment Satisfactory of 2nd year Students in CCS Department, the general weighted mean of 3.92 and interpreted as Agree. It indicates that students have a satisfactory level of strand alignment.

According to a survey that the researchers conducted, the indicators that have the highest attained mean rating, which is Strand_Alignment_Q1, “I am satisfied with my choice of SHS strand in retrospect” with the attained mean of 4.17 and interpreted as Agree. It means that students are satisfied of their chosen Strand in SHS. The lowest attained mean rating, which is Strand_Alignment_Q3 “My SHS strand aligns well with my chosen college course”, with the attained mean of 3.73 and interpreted as Agree. It means that students find their previous strands as aligned in their college courses.

In our further analyzation, it was found out that students who have taken ICT strand in their SHS have the highest attained mean rating with general weighted mean of 4.89 which is interpreted as Strongly Agree. It indicates that students who have taken ICT from their SHS are very satisfied with their chosen strand and is aligned with their courses.

On the other hand, students who have taken HUMSS strand in their SHS have the lowest attained mean rating with general weighted mean of 2.17 which is interpreted as Disagree. It indicates that students who have taken HUMSS from their SHS are unsatisfied with their chosen strand and is not aligned with their courses.



Legend:

4.21 - 5.00 - Strongly Agree

3.41 - 4.20 – Agree

2.61 - 3.40 – Neutral

1.81 - 2.60 – Disagree

1.00 - 1.80 - Strongly Disagree

Questions:

Learning_Adaptability_Q1: The SHS strand I selected significantly affects my ability to adapt to lessons in my current course.

Learning_Adaptability_Q2: My SHS strand contributed to my overall readiness in my course.

Learning_Adaptability_Q3: My SHS strand provided me with opportunities for hands-on experience or practical application of Computer/Technology concepts.

Learning_Adaptability_Q4: My SHS strand contribute to the development of my technological skills.

Learning_Adaptability_Q5: My SHS strand provided me a knowledge on basics of programming.

OVERALL WEIGHTED MEAN: 3.71

This Bar Graph 2 shows the Learning Adaptability of 2nd year Students in CCS Department, the general weighted mean of 3.71 and interpreted as Agree. It indicates that students have a satisfactory level of learning adaptability in their course based on their SHS strand.

According to a survey that the researchers conducted, the indicators that have the highest attained mean rating, which is Learning_Adaptability_Q2, “My SHS strand contributed to my overall readiness in my course”, with the attained mean of 4.27 and interpreted as Strongly Agree. It indicates that students’ previous strands are greatly contributed to their overall readiness in their course. The lowest attained mean rating, which is Learning_Adaptability_Q3 “My SHS strand provided me with opportunities for hands-on experience or practical application of Computer and Technology concepts”, with the attained mean of 3.15 and interpreted as Neutral. It indicates that students

have mixed opinions or are uncertain about the extent to which their SHS strand provided opportunities for hands-on experience or practical application of Computer and Technology concepts.

In our further analyzation, it was found out that students who have taken ICT strand in their SHS have the highest attained mean rating with general weighted mean of 4.63 which is interpreted as Strongly Agree. It indicates that students who have taken ICT from their SHS are easily adapts to the lessons and activities in their chosen course in college.

On the other hand, students who have taken GAS and TVL strand in their SHS have tied in the lowest attained mean rating with general weighted mean of 2.6 which is interpreted as Disagree. It indicates that students who have taken GAS and TVL from their SHS are hardly adapts to the lessons and activities in their chosen course.

3. Limitations and Potential Biases

Potential limitations in the data include response bias and reliance on self-reported perceptions, while the analysis primarily emphasizes quantitative measures, overlooking potential qualitative nuances and external factors influencing students' perspectives on strand alignment.

VI. Conclusion

Summary of Findings

The findings of the studies were summarized as by follows:

1. Demographic profile of the students.
   - 1.1 **Sex** – The percentage of respondents was 68% male and 32% female.
   - 1.2 **Age** – 19 years old was the largest respondents with 47%.
   - 1.3 **Course, Year & Section** – Majority of respondents came from BSIT 2D with 21%.
   - 1.4 **SHS Strand** – Most of respondents have a strand of ICT Strand in their Senior High School with 34%.
2. Assessments on Data Analysis
   - 2.1 In terms of Strand Alignment Satisfactory, the data present general weighted mean of 3.92 and interpreted as Agree.
   - 2.2 In terms of Learning Adaptability, the data present general weighted mean of 3.71 and interpreted as Agree.
3. Assessment on Data Analysis by SHS Strand
   - 3.1 In terms of Strand Alignment Satisfactory by SHS strand, the data present ICT as the highest general weighted mean of 4.89 and interpreted as Strongly Agree.
   - 3.2 In terms of Learning Adaptability by SHS strand, the data present ICT as the highest general weighted mean of 4.63 and interpreted as Strongly Agree.

Conclusion

Based from the result of the data analysis, our group concludes that:

1. The majority of students exhibit satisfaction with their chosen Senior High School strands, indicating a generally content and aligned experience. This positive trend highlights students' overall satisfaction with their strand choices, suggesting a positive connection between their Senior High School experiences and current academic pursuits.
2. As for learning adaptability of the students, on average, students perceive a satisfactory level of adaptability in their learning experiences, implying a positive reception of their Senior High School education in relation to their current academic courses.
3. Analyzing Strand Alignment Satisfactory and Learning Adaptability by Senior High School strands, the data reveals exceptionally high satisfaction and alignment levels for students in the ICT strand. This emphasizes a positive connection between their chosen strand and current academic paths. In addition, students who pursued the ICT strand not only demonstrated exceptionally high satisfaction and alignment but also found the ICT background to be particularly beneficial in adapting to lessons, underscoring the practical relevance and positive influence of the chosen strand on their academic journey. This highlights the significant role of the ICT strand in fostering a seamless transition and enhanced adaptability for students pursuing courses in BS Information Technology and BS Computer Science.

Potential Future Improvements and Extensions

For future improvements our group suggest to incorporate quantitative interviews, implement a longitudinal study, explore additional demographics, establish a feedback mechanism, and align educational strategies with industry needs for a more comprehensive project. In addition, enhance data visualization techniques for accessibility and engagement, ensuring the continuous refinement of educational approaches based on both student expectations and real-world demands.

VII. Appendix

Data Dictionary

Data Set: background_strand_dataset.csv

Feature Name	Explanation	Data Type
SEX	sex	string
AGE	age	integer
COURSE_YEAR_SECTION	Course, year & section of student	string
SHS_STRAND	SHS strand of student	string
Strand_Alignment_Q1	I am satisfied with my choice of SHS strand in retrospect.	integer

Strand_Alignment_Q2	My SHS strand influenced my college course selection.	integer
Strand_Alignment_Q3	My SHS strand aligns well with my chosen college course.	integer
Learning_Adaptability_Q1	The SHS strand I selected significantly affects my ability to adapt to lessons in my current course.	integer
Learning_Adaptability_Q2	My SHS strand contributed to my overall readiness in my course.	integer
Learning_Adaptability_Q3	My SHS strand provided me with opportunities for hands-on experience or practical application of Computer/Technology concepts	integer
Learning_Adaptability_Q4	My SHS strand contribute to the development of my technological skills	integer
Learning_Adaptability_Q5	My SHS strand provided me a knowledge on basics of programming.	integer

Code for Data Cleaning and preprocessing

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('background_strand_dataset.csv')
df.fillna(df.mean(numeric_only=True), inplace=True)
df['AGE'] = pd.to_numeric(df['AGE'], errors='coerce')
df = df[df['AGE'] >= 0]
df.head(101)
```