Project Name: The productivity of male and female second year IT students that has their own devices.

Member(s):

1. Aguitez, Christian Joseph F.
2. Austria, Vic Maurice L.
3. Bernabe, Naomi Joy A.
4. Gonzales, Jerson Ray A.
5. Manahan, Mark Junelle S.
6. Rosales, Aicelle A.
7. Urriza, Rhenz Jhovanne C.

Date: January 2024

1. Introduction

This project aimed to gather insights about the relationship between internet connection, location, devices, quality of internet connection, and lifestyle factors and productivity among the respondents. Additionally, to provide an overview of the data for effective visualization, interpretation, and analysis,.

To achieve this, a survey was conducted among the 2nd year BSIT students to collect information about their internet connection, where they were located, the devices they used, the quality of their internet connection, and their lifestyle factors. The survey aimed to find patterns and relationships between these factors and the students' productivity.

The findings from the survey were then analyzed to understand which factors had an impact on the students' productivity. By examining the data, the project aimed to provide valuable insights into what influences productivity among BSIT 2nd-year students.

2. Data

- Details of the data set:
    - There are 170 rows and 11 columns
    - The data type of each variable is object
    - There are no missing values
    - Standardization is used in data transformation. This process involves scaling the values of numeric variables to a common scale. It ensures that variables with

different measurement units or scales are comparable and can be used effectively in various analytical techniques and models.

3. Visualization Technique(s)

The types of visualizations that we have used are bar charts and heat maps. We have used bar charts to visualize the distribution of responses for variables like gender, productivity, devices, location,connection, and lifestyle. Bar charts are suitable for displaying and comparing the counts or percentages of different categories within a single variable. They provide a clear view of the distribution and make it easy to compare values across variables.The x-axis represents the variables, and the y-axis represents the count or frequency. As for the colors, we have used different colors of the bar that distinguish different categories. In addition, we have also used heat maps to visualize the relationship between the variables as it is effective for showing the frequency observations across variables. The color intensity represents the count of observations in each combination. As for the additional libraries or packages used for visualization, we used matplotlib and seaborn libraries.

4. Implementation in Google Collab

- Provide snippets of the relevant Python code used to create the visualizations.

```
[ ] #Importing
    import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt
    import seaborn as sns
    plt.style.use('seaborn-whitegrid')

    #Dataset Load

    data = pd.read_csv('/content/drive/MyDrive/2ND YEAR/Colab Notebooks/DATASET SURVEY (Responses) - All Response - DATASET SURVEY (Responses) - All Response.csv')

    #Display
    data.head()
```

```
#bar graph
bars = pd.crosstab(index=data['Gender'], columns=data['1 to 5 Scale Productivity'])
bars.plot(kind="bar")
plt.xlabel('Productivity in quality of Internet Connection') #horizontal
plt.ylabel('Data') #vertical text title
plt.title('Productivity') #title of graph
plt.xticks(rotation= 0)
plt.tight_layout()
plt.show #show graph
```

- Explain the key steps involved in the code:
  - Data loading and cleaning

In data loading and cleaning we first start coding for importing necessary libraries and loading the dataset using pd.read_csv(). The data.head() statement is used to display the first few rows of the dataset.

- Visualization construction

  In our bar graphs, we have used pd.crosstab() function to create a contingency table between two variables and store the result in the bar variable.The bars.plot(kind="bar") statement is used to create a bar graph based on the contingency table.The plt.xlabel(), plt.ylabel(), and plt.title() functions are used to set the labels and title for the graph. The plt.xticks(rotation=0) statement is used to avoid rotating the x-axis labels. The plt.tight_layout() function adjusts the spacing between subplots. Finally, the plt.show() function is called to display the graph.

  Meanwhile, in our heatmaps we uses the sns.heatmap() function to create a heatmap. The pd.DataFrame() function is used to create a DataFrame that stores the counts/frequencies for each combination of categories. The DataFrame is created using a dictionary comprehension that iterates over the groups obtained by grouping the DataFrame _df_2, _df_14, _df_17, _df_20, and _df_32 by the 'Gender' variable. The resulting DataFrame is passed to the sns.heatmap() function along with the desired colormap (cmap parameter). The plt.xlabel() and plt.ylabel() functions set the labels for the x-axis and y-axis, respectively. The plt.subplots(figsize=(8, 8)) statement is used to set the size of the heatmap subplot.The plt.show() function is called to display the heatmap.

- Customization and styling

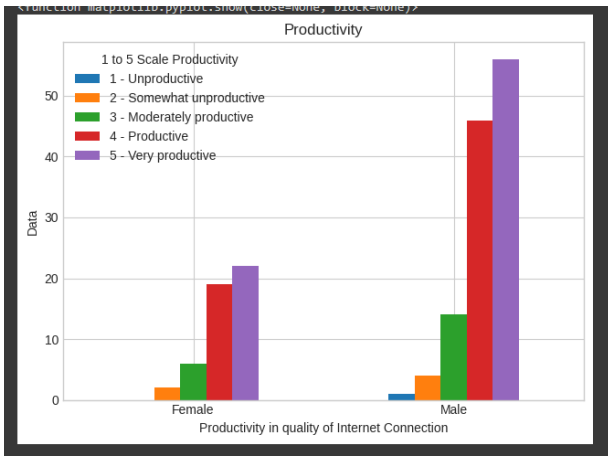
  As we customize our graphs we include various customization options such as setting the x-axis label, y-axis label, and title for each graph using plt.xlabel(), plt.ylabel(), and plt.title(). The plt.style.use('seaborn-whitegrid') statement sets the style of the plots to a seaborn whitegrid style. The plt.subplots(figsize=(8, 8)) statement is used to set the size of the subplots.

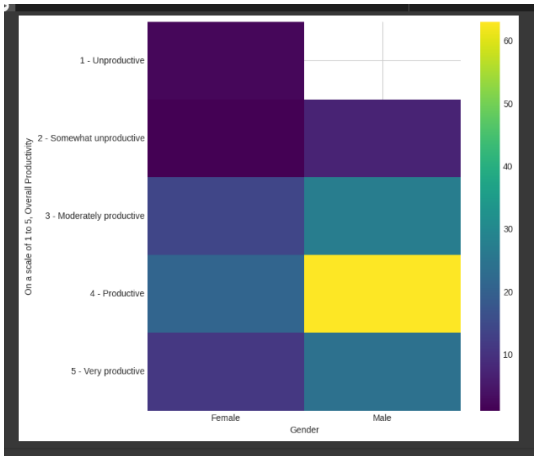
- Highlight any challenges faced and how they were overcome.

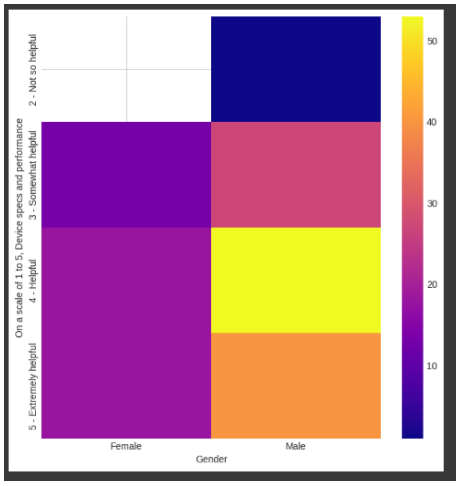
One of the challenges we have faced is the coding and execution on how the display the visualizations of the data we have gathered. Luckily, we have overcome this challenge as we have knowledge on it. In addition, we have browsed on how to do it.

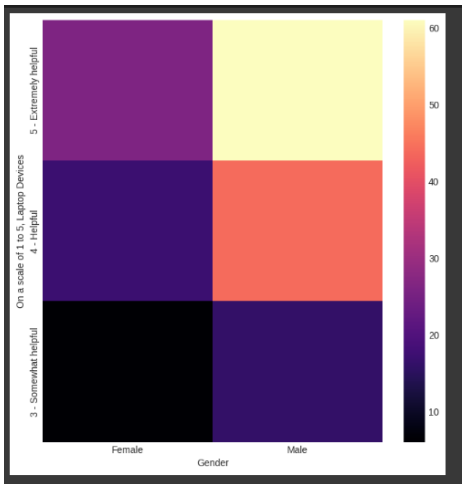
5. Results and Interpretation

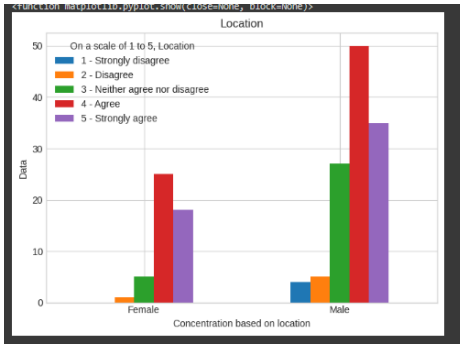
- Present the final visualizations with clear captions, labels, and relevant references.

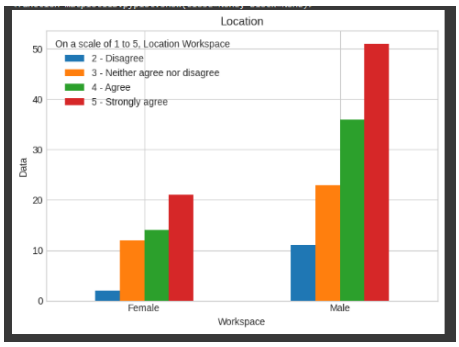

Productivity
1 to 5 Scale Productivity
1 - Unproductive
2 - Somewhat unproductive
3 - Moderately productive
4 - Productive
5 - Very productive
Data
0
10
20
30
40
50
Female
Male
Productivity in quality of Internet Connection

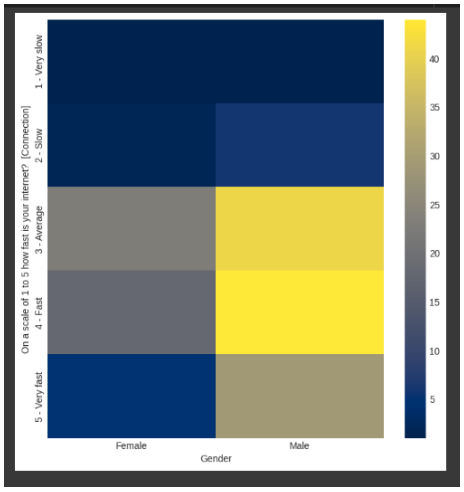

1 - Unproductive
2 - Somewhat unproductive
3 - Moderately productive
4 - Productive
5 - Very productive
On a scale of 1 to 5, Overall Productivity
Female
Male
Gender
10
20
30
40
50
60

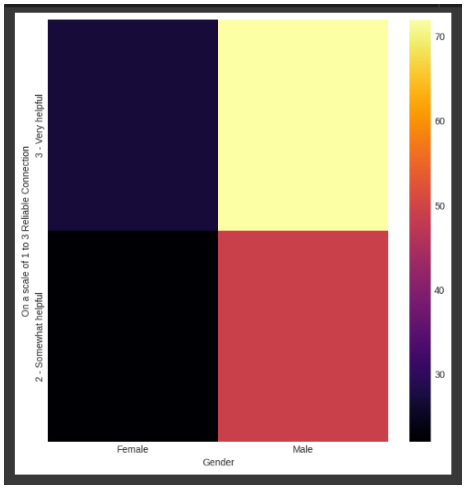

2 - Not so helpful
3 - Somewhat helpful
4 - Helpful
5 - Extremely helpful
On a scale of 1 to 5, Device specs and performance
Female
Male
Gender
10
20
30
40
50

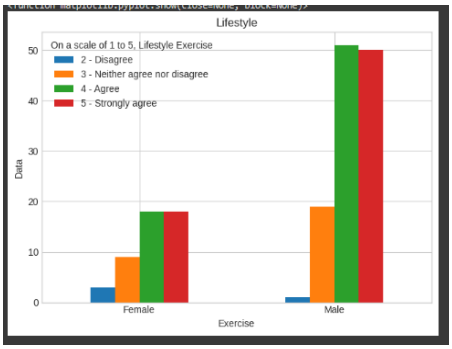

5 - Extremely helpful
4 - Helpful
3 - Somewhat helpful
On a scale of 1 to 5, Laptop Devices
Female
Male
Gender
10
20
30
40
50
60











- Describe what you see in the visualizations and what insights you can draw.

Based on the graph, in terms of productivity, male students are more productive than female students. As for productivity, researchers have found out that both male and female students only agree that their location affects their productivity. When it comes to having their personal workspace, they both strongly agree that having a workspace affects their productivity. Finally, the bar graph regarding lifestyle, womens have an equal result, agree and strongly agree. On the contrary, male students have an average of more than 50% who agree and exactly 50% who strongly agree that lifestyle affects their productivity.

- Discuss how the visualizations answer the initial project questions.

Well, these visualizations indicate that the dependent variables are related and really affect productivity of students. In addition it indicates the relationship of dependent and independent variables that includes the percentage of students who have that perception.

- Address any limitations or potential biases in the data or visualization.

This data is limited on validating whether the inputted data by our respondents is accurate on their true perspective.

6. Conclusion

- Summarize the key findings and takeaways from the data visualization.

1. Productivity and Internet Connection:
- The bar graph shows the distribution of productivity levels (on a scale of 1 to 5) based on the quality of the internet connection.

2. Overall Productivity and Gender:
- The heatmap shows the distribution of overall productivity ratings (on a scale of 1 to 5) based on gender.

3. Device Specs and Performance and Gender:
- The bar graph displays the distribution of ratings (on a scale of 1 to 5) for device specifications and performance based on gender.

4. Laptop Devices and Gender:
- The heatmap showcases the distribution of ratings (on a scale of 1 to 5) for laptop devices based on gender.

5. Concentration Based on Location and Gender:
- The heatmap illustrates the distribution of concentration levels (on a scale of 1 to 5) based on location and gender.

6. Workspace and Gender:
- The bar graph shows the distribution of workspace preferences based on gender.

7. Internet Connection Speed and Gender:
- The heatmap displays the distribution of internet connection speed ratings (on a scale of 1 to 5) based on gender.

8. Reliable Connection and Gender:
- The heatmap showcases the distribution of ratings (on a scale of 1 to 3) for reliable internet connections based on gender.

9.lifestyle Exercise and Gender:
- The bar graph illustrates the distribution of lifestyle Exercise(on a scale of 1 to 5) based on gender.

- Suggest potential future improvements or extensions to the project.

Data Cleaning and Preprocessing

- Before visualizing the data, it's important to clean and preprocess it. This includes handling missing values, removing outliers, and ensuring data consistency. Implementing data cleaning techniques will improve the accuracy and reliability of the visualizations.

Collaboration and Feedback

-Collaborate with users or stakeholders to gather feedback on the visualizations and understand their specific needs or areas of interest. This feedback can inform future improvements and ensure that the visualizations requirements of the target audience.

7. Appendix

- Include supporting materials such as:
    - Data dictionary

      The student ID's in our dataset serves as the identification of our respondents that he/she is really a student of LSPU- Sta. Cruz Campus. Then, gender is the second variable that serves as the gender identification of each student that gives us the understanding of the gender population of second year IT students that helps us identify the difference between male and female perception on productivity. Next, is the overall productivity that has to be rated on a scale of 1-5 that defines how productive they are. Then, device specs and performance using the same rating scale defines how these variables affect their productivity. Next, location with the same rating scale, will describe how their location is affecting their overall productivity. Next, workspace, with also the same rating scale will show how having their own workspace contributes to their productivity. Then, the internet connection variable that uses the same rating scale that defines if having a strong internet connection does affect their productivity. Lastly, their lifestyle, that is also in a rating scale of 1-5, shows that being involved in an exercise or physical activity does contribute to their productivity. With these variables, we can find an accurate results that can explain and describe those factors that affects a student's productivity.

Notes:

- Use screenshots or embed the visualizations directly into the document.
- Adjust the level of technical detail based on your audience.
- Proofread and ensure the documentation is clear, concise, and well-organized.