|  |  |
| --- | --- |
|  | **CSCI/ISAT B104**  **Computer Programming Techniques, Practices, and Tools**  **Fall 2023**  **Project Contributions by Team Member** |

**Project Contributions by Team Member**

Document the contributions of each team member over the course of the project.

**Members and their Contact Information**

|  |  |  |
| --- | --- | --- |
| **Member** | **Email** | **Text** |
| Ian Bickford | [bickfori@email.uscb.edu](mailto:bickfori@email.uscb.edu) | (978) 895-9569 |
| Connor Floyd | [cwfloyd@email.uscb.edu](mailto:cwfloyd@email.uscb.edu) | (843) 816-3643 |

**Overall**

Relative Contribution of each member over the course of the entire project

|  |  |  |
| --- | --- | --- |
| **Member** | **Contribution** | **Total Hours** |
| Ian Bickford | 0 % | 0 |
| Connor Floyd | 0 % | 0 |

**YRBS Question Identification & Documentation**

|  |  |  |  |
| --- | --- | --- | --- |
| **Member** | **Contribution** | **Hours** | **Components** |
| Ian Bickford | 75 % | 3 | Identified multiple relationships, deciding on one relationship between marijuana and cocaine. Wrote step-by-step guide for query. Input pictures of all questions being explored. Thought of and wrote down hypothesizes to explore. |
| Connor Floyd | 25 % | 1 | Thought about different ideas for relationships between unrelated questions. |

**YRBS Data Winnowing & Retrieval**

|  |  |  |  |
| --- | --- | --- | --- |
| **Member** | **Contribution** | **Hours** | **Components** |
| Ian Bickford | 45 % | 2 | Focused on the step-by-step SQL query for exporting data from Microsoft Access to Microsoft Excel for easy access in python. |
| Connor Floyd | 55 % | 2 | Manipulated data and variables to create iterable columns out of the excel file, as well as a prototype for the heatmap made by seaborn. |

**Data Analysis (via Python)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Member** | **Contribution** | **Hours** | **Components** |
| Ian Bickford | 30 % | 3 | Analyzed data set and sought appropriate relationships within bar charts. |
| Connor Floyd | 70 % | 7 | Analyzed data set and sought appropriate relationships within a heat map and pie charts. |

**Graphics Creation (via Python)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Member** | **Contribution** | **Hours** | **Components** |
| Ian Bickford | 35 % | 6 | Created bar chart graphs, figured out how to manipulate text size and color in each of the graphs to be consistent. |
| Connor Floyd | 65 % | 16 | Created heat map and pie chart graphs, working on bar chart with multiple columns |

**Python Script Creation**

|  |  |  |  |
| --- | --- | --- | --- |
| **Member** | **Contribution** | **Hours** | **Components** |
| Ian Bickford | 75 % | 4 | Implemented input validation and logic. Created loop to continue asking user what graphs they want until they are done. Code explains why and what is wrong when there is an input error. Implemented a defined function for bar charts. Calls all charts within loop when appropriate. Commented throughout the python script to explain each portion of code. |
| Connor Floyd | 25 % | 1 | Implemented code to translate column headers into relevant information for use in our graphs. Implemented a defined function for pie charts and a heatmap. |

**Poster Development**

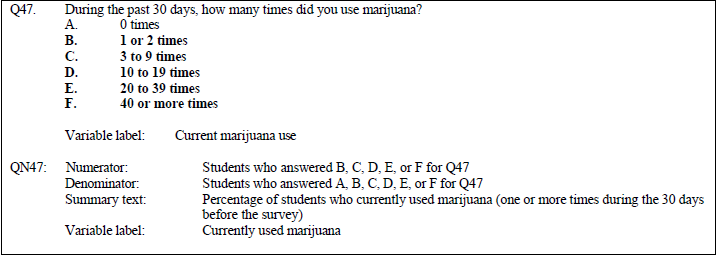
|  |  |  |  |
| --- | --- | --- | --- |
| **Member** | **Contribution** | **Hours** | **Components** |
| Ian Bickford | 0 % | 2 | Created prototype of poster through PowerPoint |
| Connor Floyd | 0 % | 0 |  |

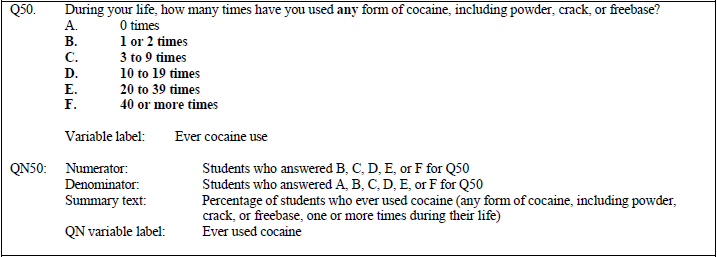
**Presentation Preparation & Practice**

|  |  |  |  |
| --- | --- | --- | --- |
| **Member** | **Contribution** | **Hours** | **Components** |
| Ian Bickford | 0 % | 0 |  |
| Connor Floyd | 0 % | 0 |  |

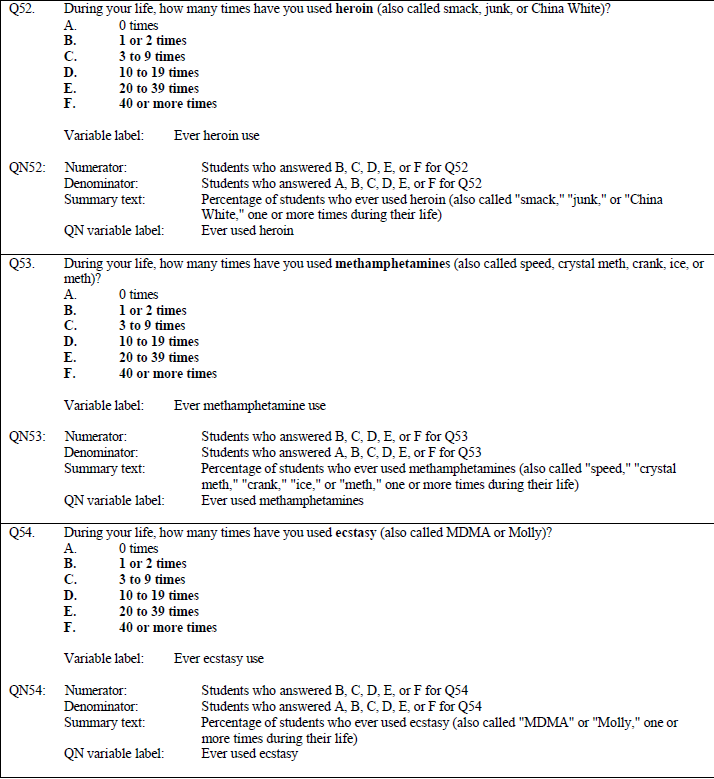
**YRBS Questions Explored**

Questions & responses explored over the course of the project.



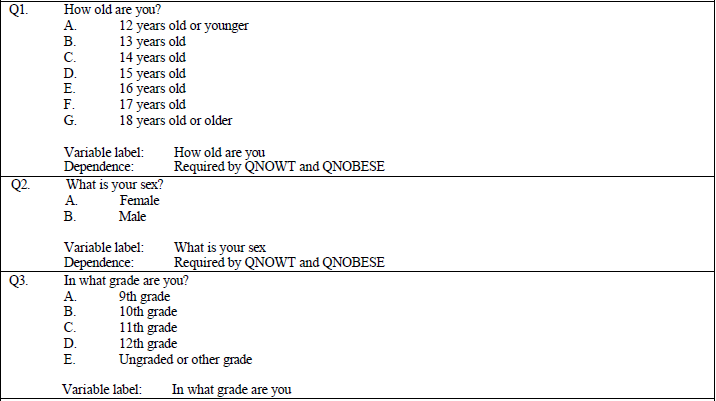


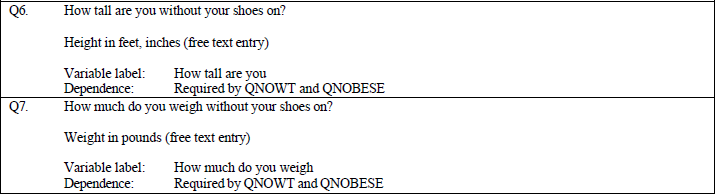
Continued On Next Page…



**YRBS Demographic Information Explored**

Demographic questions & responses explored over the course of the project.

**



**YRBS Data Winnowing & Retrieval**

How we winnowed & retrieved the data explored over the course of the project.

Write down a step-by-step guide explaining how you winnowed & retrieved the data from the YRBS site.

1) Go to https://www.cdc.gov/healthyyouth/data/yrbs/index.htm which is the YRBS website homepage.

2) From here you need to navigate through the left side drop down menu and select "Data and Documentation".

3) Scroll down to the most recent National YRBS Datasets and Documentation section which in our case was the 2021 dataset.

4) From there select the hyperlink labeled "Access" which will download a .zip file that contains the data in a Microsoft Access file.

5) In order to understand the data in the access file you must also download the Data User Guide that is directly below the access file hyperlink.

6) Once you have the access file and the user guide, place both these files into a folder that is named something you will remember.

7) Access the folder that you have placed both these files and right click the .zip file, use an unzipper such as 7zip and extract the files into your folder.

8) Once extracted open the unzipped file in Microsoft Access and select the XXHq table on the left-hand side.

9) Navigate to the "Create" tab on the top of Microsoft Access select "Query Design". Close the pop-up table that comes up.

10) From here it will open the "Design" tab for Query Tools. Navigate to the "View" icon at the top left and click the drop-down menu.

11) Next select "SQL View" and it will open a window to set the parameters of the query.

NOTE \*\*\* Depending on the questions that you are trying to compare in the data analysis, your code will be slightly different for the next step \*\*\*

12) For the purpose of our project, the code in order to query the specific questions we need for analysis and remove all null values is such:

SELECT q1, q2, q3, q6, q7, q47, q50, q52, q53, q54

FROM XXHq

WHERE q1 IS NOT NULL AND q2 IS NOT NULL AND q3 IS NOT NULL AND q6 IS NOT NULL AND q7 IS NOT NULL AND q47 IS NOT NULL AND q50 IS NOT NULL q52 IS NOT NULL AND q53 IS NOT NULL AND q54 IS NOT NULL

NOTE \*\*\* In the above code our specific questions that we are trying to analyze and compare is q47, q50, q52, q53, q54 replace these values in the first and third lines of the query

in order to utilize this query for questions you may want to compare.

13) Once the above code is implemented, navigate to the top left and select the "!" icon that is used to run the query.

14) Once done, your query will be made as "Query1", if you are trying to replicate our winnowing for this project ensure that there are 10118 rows of data.

15) Once you have your data you must export it and you have 2 options to do this. You may either export into a text file or an Excel file. We chose Excel as it was easier for us to work with.

16) In order to export this data, go to the top of Microsoft Access and select "External Data". From here you can select how you want to export it at the top of the page.

17) In our case we will select Excel. Change the file destination to the folder you made earlier and name the file something that makes sense.

18) Next check the box that says "Export Data with formatting and layout", then press ok at the bottom.

19) A pop up will occur asking if you'd like to save the export steps, just select close at the bottom of the pop-up.

20) Navigate to the saved folder for your project and ensure that the Excel file is present, and then open it.

21) Ensure that the number of rows is the same in your excel file and the Query in Microsoft Access. If this is the same you have successfully retrieved the applicable data.

**Research Question and/or Hypothesis**

The specific research questions we sought to answer.

What we are trying to investigate via our hypothesis found below is whether Marijuana IS a gateway drug statistically speaking. Our research does not indicate causation, but it may provide an avenue to investigate whether marijuana use is linked to the usage of other illegal drugs statistically speaking.

Hypothesis: When someone uses marijuana, their chances of using another illegal drug increases.

Hypothesis 2: When someone uses marijuana, their chances of using another illegal drug decreases.

Null Hypothesis: There is no relationship between the usage of marijuana and the usage of other illegal drugs.

**Bibliography**

Resources which we used over the course of the project.

<https://stackoverflow.com/questions/38438220/arrangement-of-pie-charts-using-matplotlib-subplot>

* Subplot formatting

<https://stackoverflow.com/questions/44076203/getting-percentages-in-legend-from-pie-matplotlib-pie-chart>

* Percentages and labels in legend instead of on the circle

<https://www.geeksforgeeks.org/change-font-size-in-matplotlib/>

* changing font size in the barb graphs

<https://datatofish.com/bar-chart-python-matplotlib/>

* making bar graphs

<https://stackoverflow.com/questions/14088687/how-to-change-plot-background-color>

* how to change background color of graphs

<https://python-graph-gallery.com/3-control-color-of-barplots/>

* changing color of bars in graph

<https://www.geeksforgeeks.org/change-font-size-in-matplotlib/>

* changing font size in graphs