PHOTO DEDUPLICATOR

GROUP A

An SQL query goes into a bar, walks up to two tables and asks: "Can I join you?

Why don't programmers like nature?
It has too many bugs.

My programmer friend said I have a high IQ. He said it's 404.

Agenda

01. Project methodology

We will discuss the pros and cons of waterfall and agile methodologies and give our reasons for our chosen methodology.

03. Project Plan

We will show job roles and responsibilities for each job role throughout each of the development stages. We will also discuss specific features and business objectives.

02. Implementation team

We will discuss three job roles essential to the implementation team to cover all aspects of the delivery.

04. Prototype

We will demonstrate the prototype to illustrate its functionality and show you the code for it.

05. Testing

We will discuss our test table and what tests have passed and what needed a little more work.



METHODOLOGIES

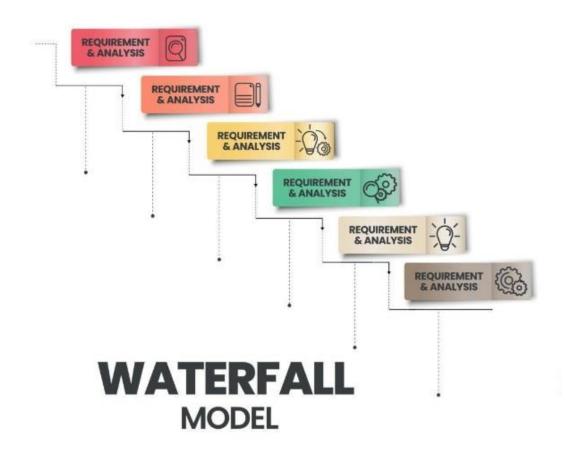
WATERFALL

PROS:

- Clear structure with defined phases which makes it easy to understand and manage each stage
- More accurate projection of time & cost
- Easier to manage and efficient as it has well-defined requirements that will not change.
- Each step is well documented ensuring that all team members are on the same page and requirements are clear.

CONS:

- Less adaptable and can be costly if changes required
- Testing late in the cycle could be costly if issues are found
- Feedback comes in late leading to significant rework if requirements were not clear
- Due to the sequential approach if any phase fails, it can jeopardize the project



Best suited for projects that have well defined requirements from start to finish. It is easy to manage however can lack flexibility. It gives predictability and structured documentation.

METHODOLOGIES

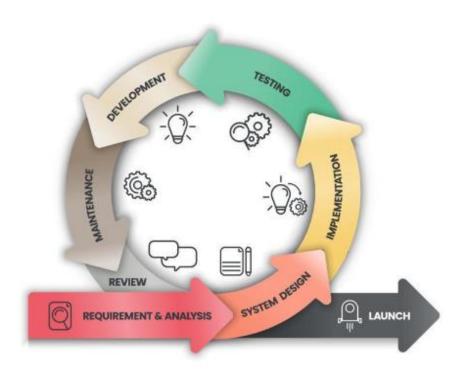
AGILE

PROS:

- Flexible and adaptable allowing for changes even late in development
- Regular feedback allows users to ensure product meets their needs
- Offers better communication among team allowing for better problem solving
- Early testing reduces risk of major issues down the line
- After every cycle, teams engage and reflect on ways to enhance their efficiency

CONS:

- Could lead to scope creep changes and additions could extend both timeline of project and costs
- Requires constant involvement from team members and customers which can be time- consuming
- Harder to predict timeline and costs to project that could lead to poor decision making



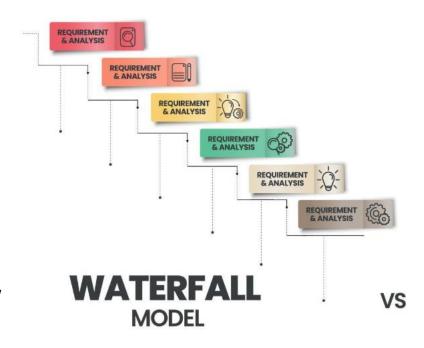
AGILE MODEL

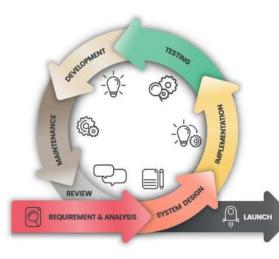
Great for projects that may need to evolve and feedback is required throughout the process. It is flexible and can be continuously improved with fast delivery. It can suffer from scope creep.

METHODOLOGIES

Our chosen methodology: AGILE

We have chosen the Agile Methodology for our Photo DeDuplicator project due to its flexibility and continuous feedback that can be obtained throughout all stages including the earlier testing stages. It has also allowed us to make any required changes throughout the development and will allow for any evolving features in the future.







IMPLEMENTATION TEAM

Lead Developer

- Responsible for coding, integration and testing.
- Ensures best coding practices and coding standards.
- Manages development team, assigns tasks and ensures project meets requirements

Product Owner

- Acts as voice of the customer
- Defines product requirements, and user stories.
- Meets with other team members to ensure quality and functionality meets customers expectations

Quality Assurance Engineer

- Responsible for quality and reliability
- Designs and runs test plans, test cases and test scripts.
- Ensures product meets the specified quality standards.

LEAD DEVELOPER

JOB DESCRIPTION

- Ensure adherence to best coding practices and standards
- Design and develop high quality code
- Collaborate with other team members/departments to define project requirements and deliver updates throughout the process
- Fix bugs and issues when and if required
- Create and update documentation to ensure it is up to date

- Proficiency in the relevant programming languages (e.g., HTML, Python, CSS)
- Have a strong understanding of software development methodologies (e.g., Agile, Waterfall, Scrum)
- Strong communication and interpersonal abilities
- Problem-solving and critical-thinking skills

PROJECT MANAGER

JOB DESCRIPTION

- Define the product vision and strategy and ensure it aligns with the business' goals and objectives
- Create and manage product requirements and user stories
- Gather requirements from stakeholders/customers and ensure all departments understand these requirements
- Take the lead over review meetings and sprint planning meetings
- Collect user feedback and use it to improve the product

- Have a strong understanding of software development methodologies and processes
- Be proficient in using product management systems (JIRA, Trello)
- Strong leadership and decision making skills
- Excellent communicator and analytical thinker
- Communicate with team members, ensuring sufficient progress is being made and making informed adjustments when necessary

QA ENGINEER

JOB DESCRIPTION

- Develop test plans and strategies and define the test objectives and criteria
- Write detailed test cases that cover all functional and non-functional requirements
- Execute manual tests to identify bugs and issues and then document and report the findings
- Conduct performance, load and stress tests and report findings
- Perform regression testing to ensure code changes do not create new defects
- Set up and maintain test environments

- Have a strong understanding of software development methodologies and processes
- Be proficient in using product management systems (JIRA, Trello)
- Detail-oriented with very strong analytical skills
- Excellent communication and collaboration abilities
- Problem- solving and critical- thinking skills

DEVELOPMENT STAGES

DEVELOPMENT STAGES								
ROLE	PLANNING	DESIGN AND DEVELOPMENT	TESTING AND IMPLEMENTATION	DEPLOYMENT	MONITORING AND UPDATE RELEASES			
PROJECT MANAGER	Provide detailed instructions for the planning, organizing, controlling, managing and reporting the project resources	Plan sprints, versions, and releases according to the Photo DeDuplicator project roadmap. Manage development iterations, organise the development events	Work with the development team, plan sprint increments	Coach the whole team, report on time/efforts and delivered features, remove obstructions for the team.	Ensure efficient communication between the Product Owner and the other members of the team. Tasks would end when project goals and objectives are successfully completed, product is delivered			
BUSINESS ANALYST	Analyse, communicate and document the requirements, business needs and goals for the Photo DeDuplicator with stakeholders, developers and other participants	Analyse and validate the detailed requirements of the project, creating and maintaining design artefacts (such as prototypes) and specifications. Oversee the changes and issued that arise during design.	Ensure that the final product meets the business requirements. Help the technical team to resolve requirement related roadblocks	Serve as the first point of contact for users when they face issues and provide feedback.	Coordinate the development team, organise future updates in the software or implements the required fixes.			
SOLUTIONS ARCHITECT	Try to understand constraints of the project. Considers technical and business factors which align with the project goal	Ensure all requirements are met during development and examine the limitations of the project to optimize alternatives, and reduce risks	Make crucial technical decisions which impact project's ultimate results (such as technology stack and integration with third-party systems)	Act as the link between the technical team and the project manager	Ensure continuous maintenance of the product developed so it works properly.			

DEVELOPMENT STAGES

DEVELOPMENT STAGES								
ROLE	PLANNING	DESIGN AND DEVELOPMENT	TESTING AND IMPLEMENTATION	DEPLOYMENT	MONITORING AND UPDATE RELEASES			
LEAD DEVELOPER	Work closely with the Project Manager, Business Analyst, and other stakeholders to understand requirements and translate them into technical specifications	Provide technical direction and execution of the Photo DeDuplicator project. Oversee the entire development process, from design, coding, testing, and deployment	Manage the development team, assigns tasks, conduct code reviews, and ensure adherence to coding standards and best practices. Also guide the development team, create a collaborative environment, and promote continuous learning and improvement.	Oversee deployment process and ensure a smooth release	Stay up-to-date with the latest technologies, tools, and industry trends, and recommend appropriate updates and improvements for the project			
QA ENGINEER	Collaborate with the Business Analyst and stakeholders to understand the requirements	Review the design documents and provide feedback from a testing perspective. Participate in code reviews and provide feedback on code quality, testability, and maintainability	Develop and execute comprehensive test plans, test cases, and test scripts to validate the functionality, usability, and performance of the application. Identify and report defects, prioritize them based on severity and impact, and track them through resolution Develop and maintain test automation frameworks and scripts to streamline the testing process and improve test coverage.	Collaborate with cross- functional teams to ensure a smooth delivery and deployment process. Ensure compliance with industry standards, best practices, and quality assurance methodologies. Perform final testing and validation before release	Continuously monitor and test the application for any issues or defects. Update and maintain test cases and automation scripts for future releases.			

12

26 June, 2024

COMMITMENTS

FEATURES

- Application has cross-platform compatibility- can be used with various operating systems Windows and macOS
- The application supports a wide range of formats such as .png , .jpeg and .gif. It recognises all formats and is able to spot duplicates within a folder
- It has a friendly interface that is easy to understand no matter the skillset of the user.
- It will end up saving storage space as duplicate files can be deleted, freeing up storage space.

BUSINESS OBJECTIVES

- The project must be cost effective without sacrificing quality or functionality
- The application must comply with relevant regulations regarding safeguarding, user data and security
- There should be security measures within the application to protect the users sensitive data and the user's system
- The application should be able to handle big workloads without any effects on the performance

Future Features

Advanced Image Algorithms

Add techniques like Difference Hash (dHash) and Wavelet Hash (wHash) to improve the accuracy of finding duplicate images, especially for images that have been cropped or rotated.

Skills Required:

Knowledge of image processing libraries like OpenCV / Pillow; Algorithm development in Python.

Batch Processing

Introduce batch processing, this would allow users to select multiple folders or drives, and the program would scan all of them at the same time to find duplicate images.

Skills Required:

Knowledge of Python directories like : os.walk, os.listdir.

Data structures like lists or queues to store and manage the file paths.

UI design to integrate choice in menu.

Customizable Filters and Exclusions

This would allow users to define specific rules to include or exclude certain files from the deduplication process. For example user could exclude file types or files based on their size or date modified.

Skills Required:

Knowledge of Python.
Regular expressions for pattern
matching (e.g., re module in Python).
File system operations to check file size,
creation date, or modification date.
UI design to integrate these filters.

To implement these future features our Lead Developer will need to enhance their skills in Python programming, working with file systems, and creating user interfaces. Our QA Engineer's abilities to design test cases, perform different testing methods, and handle errors will be very important to make sure the Photo DeDuplicator application works reliably.

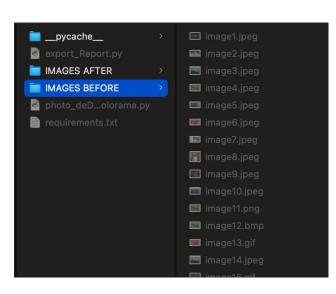
PROTOTYPE



This is the starting screen and main menu of the program. There are 5 options that the user can choose from that will all be explored in this section of our presentation.



If the user inputs option '0', it will lead them to the page above which offers information on the Photo De-duplicator application's functionalities and also how to use it correctly. After viewing this, they can press 'Enter' to return to the main menu.



Option '1' in the main menu sets the folder destination that the program will run inside of. Once the user chooses this option, the user system's file manager will pop up to choose the folder that the user wants to select. In this case, it opens up the 'Finder' as this demonstration is being made on the macOS system.

After the folder is selected the destination of the folder will appear in the main menu above the options.

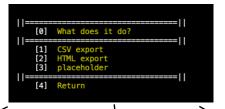
Current selected path is: /Users/elijahdeocampo/Deskto

image18.jpeg: 0f4f0702000407be image14.jpeg: ffffff7f0e000000 image21.bmp: 00082c067fffffff image19.jpeg: 8c3f37c1818dfbff image1.jpeg: 00007bff3f194848 image6.jpeg: 737e252727a3e3e1 image7.ipeg: ccc0c0c0f0fcfcfe image11.png: 78ffdfc2c0e1687c image10.jpeg: ffff21200a020000 image8.jpeg: 98b898883c3c3c3c image5.jpeg: f3a7c787170300ff image9.jpeg: 6044fcfc181c5e7a image2.jpeg: 0008280cfcffffff image20.jpeg: f1080038fe6e2603 image16.jpeg: 3e7f7f6170700000 image13.gif: f9f9d2e5e3e1e1c0 image17.jpeg: 0000b9fcf8fcec3e Duplicate Images: Would you like to delete duplicate images? YES or NO? YES Are you sure you would like to delete duplicate images? Whether or not they are recoverable will be based on your computer settings. YES or NO?

Option '2' in the main menu runs the De-duplicator code and created a list of all the images in the folder. It then creates a list of the duplicate images, regardless of their file types. It then asks the user if they want to delete any duplicate images. IF they choose 'YES', a warning is given to verify if the user wants the duplicates deleted.

[3] Clear history

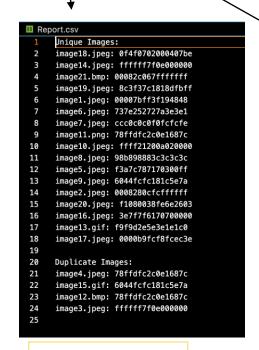
Option '3' will clear the history of the program. As specified in Option '0' (what does it do), this option should be inputted when the user wants to run the program on another file.



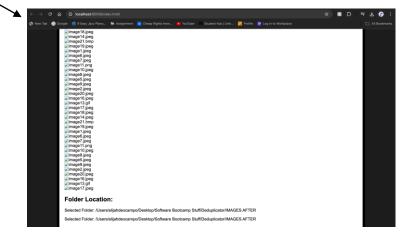
Option '4' is the where the export function is. When inputted, the user will be taken to another menu with different options. Option '0' will take the user to a section that explains what the different options do. This being a prototype, it currently has two working options which creates a report either in a CSV file within the program or on a HTML page. Option '4' will return to the main menu.

Press ENTER to move dialogue forward] In the future there may or may not be multiple export options. For now, there is only CSV A CSV file is a simplified version of the file format used by Excel, which stores Comma Seperate Values It can be used to contain simple strings of data, such as this. The program will create a new file in the same directory it is stored in. You can choose the files name, it is recommended that you make the names unique and meaningful. [Press ENTER to return to main menu]

Option '0' (Info)



Option '1' (CSV)



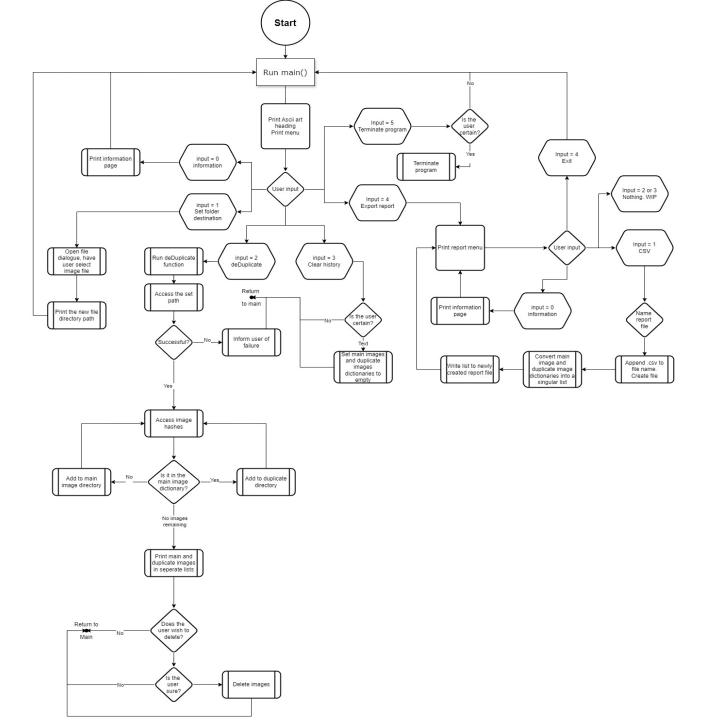
Option '2' (HTML)

26 June, 2024



Finally, option '5' will exit the program. Once inputted, the user will be asked if they are sure they would like to exit as if the report has not been exported, it will be deleted.

PYTHON CODE IMPLEMENTATION



```
from tkinter import filedialog
      from PIL import Image
3
      import imagehash
      import os
      from colorama import Fore, Back, Style
      from export_Report import exportCSV
      images = {}
      dupImages = \{\}
      path = "C:\\IMAGES" #location of the image folder on my computer, for testing purposes
10
11
      ## If you want a version without colorama, I can put a version of the code without it on my github ##
12
13 ∨ def reportOpt():
14
          os.system("cls")
15
          global images,dupImages
16
17
          print(f"||======||")
18
          print(f" [0] {Fore.YELLOW} What does it do? {Fore.RESET}")
          print(f"||======||")
19
20
          print(f"
                     [1] {Fore.YELLOW} CSV export{Fore.RESET}")
                     [2] {Fore.YELLOW} placeholder{Fore.RESET}")
21
          print(f"
22
          print(f"
                     [3] {Fore.YELLOW} placeholder{Fore.RESET}")
          print(f"||======||")
23
24
          print(f" [5] {Fore.YELLOW} Return\n{Fore.RESET}\n")
25
26
          choice = input(" ")
27
          print("\n\n")
28
29
          if choice == '0':
30
              os.system("cls")
31
              print("[Press ENTER to move dialogue forward]\n")
32
33
              print(Fore.YELLOW + "In the future there may or may not be multiple export options. For now, there is only CSV\n")
34
              input()
35
              print("A CSV file is a simplified version of the file format used by Excel, which stores Comma Seperate Values")
              print("It can be used to contain simple strings of data, such as this. The program will create a new file in\nthe same directory it is stored in.")
36
37
              print("You can choose the files name, it is recommended that you make the names unique and meaningful.")
38
              print(Fore.RESET) ### If other bluesky options get completed, then a description of them may be added here for the user's clarity. If you're unsure how
39
              print("[Press ENTER to return to main menu]")
40
              input()
```

```
elif choice == '1':
               exportCSV(images,dupImages)
42
43
              os.system("cls")
           elif choice == '5':
44
45
               return
46
           reportOpt()
47
48 ∨ def deDuplicate(path):
          os.system("cls")
49
50
           global images, dupImages
           for FILE in os.listdir(path):
51
52
               imgPath = os.path.join(path,FILE)
53
              HASH = str(imagehash.average_hash(Image.open(imgPath)))
54
              #print(FILE)
              #print(HASH)
55
56
              if HASH in images:
                  #print(f"Duplicate: {FILE}, {HASH}")
57
58
                  dupImages[FILE] = HASH
59
               else:
60
                  #print(f"Unique image: {FILE}, {HASH}")
                  images[HASH] = FILE
61
           print("Unique Images:")
62
63
           for key in images:
64
               print(f"{images[key]}: {key}")
           print("\n\nDuplicate Images:")
65
           for key in dupImages:
66
67
              print(f"{key}: {dupImages[key]}")
68
           print("Would you like to delete duplicate images?")
69
70
           CHOICE = input("YES or NO? ")
71
           if CHOICE.lower() == 'yes':
72
              print("\nAre you sure you would like to delete duplicate images?\nWhether or not they are recoverable will be based on your computer settings.")
73
              CHOICE = input("YES or NO? ")
              if CHOICE.lower() == 'yes':
74
75
                  for key in dupImages:
```

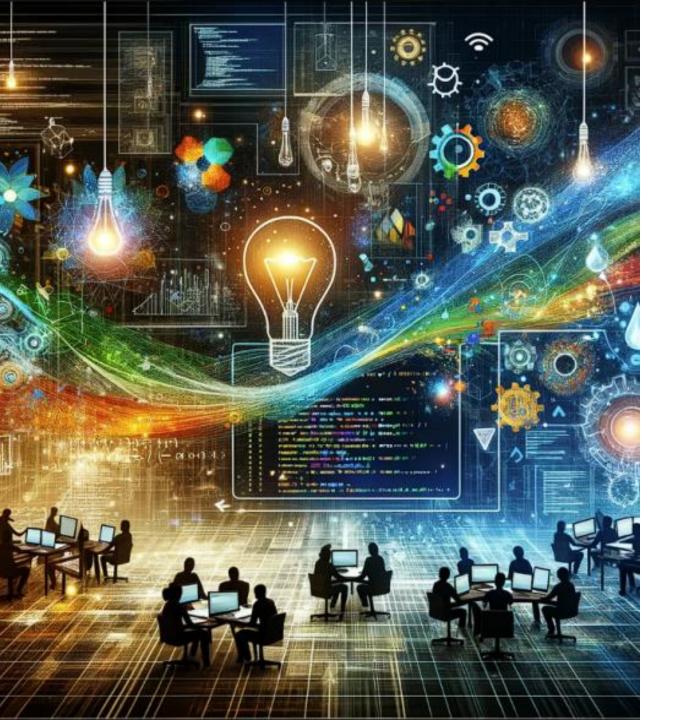
```
imgPath = os.path.join(path,key)
76
77
                print(f"Deleting: {imgPath}")
78
                try:
79
                   os.remove(imgPath) #this is the code to remove a file
80
                except:
                   print("Failed to remove one or more images.")
81
             print("Successfully deleted all duplicate images.")
82
83
84
85
86
87 ∨ def main():
88
        global path, images, dupImages
89
        os.system("cls")
90
        91
        print(rf"||
                   {Fore.RED}
                              _====_ /\ /\ //===\\ ====== //===\\ {Fore.RESET}
                                                                          11")
                                                                          11")
        print(rf"||
92
                   {Fore.RED}
                                                            {Fore.RESET}
                                                                          11")
        print(rf"||
93
                   {Fore.RED}
                                                            {Fore RESET}
                                                                          11")
94
        print(rf"||
                   {Fore RED}
                                         ш
                                                           (Fore RESET)
                             Ш
                                   || || \\===//
                                                                          11")
95
        print(rf"||
                   {Fore RED}
                                                print(rf"||------||")
96
        print(rf"||{Fore.BLUE} |-\ |-- |-\ | | |-\ | ---- /- /\ ---- /-\ |-\ {Fore.RESET} ||")
97
        print(rf"||{Fore.BLUE} | | |-- | | | | |-/ | | |
                                                  98
        99
        print(rf"||-----||","\n\n\n")
100
101
102
103
        print(f"||======||")
104
        print(f" {Fore.YELLOW}Current selected path is: {path} {Fore.RESET} ")
        print(f"||======||")
105
106
        print(f"
                [0] {Fore.YELLOW} What does it do? {Fore.RESET}")
        print(f"||=======||")
107
        print(f"
                [1] {Fore.YELLOW} Set folder destination{Fore.RESET}")
108
109
        print(f"
                [2] {Fore.YELLOW} Run Photo deDuplicator{Fore.RESET}")
110
        print(f"
                [3] {Fore.YELLOW} Clear history{Fore.RESET}")
```

```
111
            print(f"||======||")
112
           print(f"
                       [4] {Fore.YELLOW} Export report {Fore.RESET}(WIP)")
           print(f"||======||")
113
           print(f"
114
                       [5] {Fore.YELLOW} Exit deDuplicator\n{Fore.RESET}\n")
115
116
           ###after running program, print the report then have an option to export it as something. CSV, html, etc.
117
118
119
            choice = input(" ")
120
           print("\n\n")
121
122
123
           if choice == '0':
124
               os.system("cls")
125
               print("[Press ENTER to move dialogue forward]\n")
126
               input()
127
               print(Fore.YELLOW + "The Photo deDuplicator is built to find duplicate images in a specific folder. It has not been known to work on subfolders.\n")
128
               print("It does this by converting an image's data into a unique hash value, which is then compared by every other image stored.\n")
129
               input()
130
               print("If you wish to check multiple folders in one instance of the program, it is best to do them one at a time.\n")
131
               print("Between folders, you should clear the stored history. This is to protect against similar images stored in different folders,")
132
               print("which you may not wish to be deleted. If you want to make sure there are no duplicates at all, in every folder checked,")
133
               print("then you simply do not delete the history.")
134
               input()
135
               print("The program is currently able to export into a .csv file.\n\n" + Fore.RESET)
136
               print("[Press ENTER to return to main menu]")
137
               input()
138
139
            elif choice == '1':
140
               path = filedialog.askdirectory()
141
               print(path)
142
            elif choice == '2':
143
               deDuplicate(path) #runs the deDuplicate function, using the path as an input to carry information between functions
144
           elif choice == '3':
               print("Are you certain you wish to clear the system's history? Doing so will also clear the report, unless it's already been exported.\n\n")
145
               CHOICE = input("[yes/no]\n\n")
146
```

```
if CHOICE.lower() == 'yes':
147
                   images = {} #sets images and dupImages to empty dictionaries, effectively clearing them
148
149
                   dupImages = {}
            elif choice == '4':
150
151
                reportOpt()
            elif choice == '5':
152
153
               CHOICE = input("Are you sure? Leaving will delete report if not exported\n[yes/no]\n\n") ### make it check whether you've already exported the report
154
               if CHOICE.lower() == 'yes' or CHOICE == '':
                   os._exit(0)
155
156
            main() #reset once done
157
158
        main()
        #exportCSV(images,dupImages)
159
```

TESTING

Test Case ID	Test Case Description	Test Data	Expected Result	Pass/Fail	Correction Plan
TC - 01	Test 'path' location	Add folder path	Display folder location and all items in location.	PASS	
TC - 02	Test De-Duplication functionality with images	Add folder path that holds unique and duplicated images	Display a list of all images and separate list of duplicates found.	PASS	
TC - 03	Test 'path' location if path is empty or does not exist	Add a non-existent path location	Display error message 'Path not found'	FAIL	Select folder location instead of typing it
TC - 04	Test functionality with different image formats	Add folder path that holds various image formats	Display images even if in different formats without errors.	PASS	
TC - 05	Test user interface to ensure instructions are clear	Try running on different screen sizes and resolutions	Ensure it is all positioned well, readable and easy to follow	PASS	
TC - 06	Test operating system cross-compatibility	Try running on Windows and MAC operating systems	Ensure runs well on both operating systems	PASS	
TC - 07	Test De-Duplication functionality with images that are similar but differ in size.	Add folder path that holds similar looking photos but different formats and sizes	Display a list of duplicates found.	FAIL	Implement in future updates as mentioned in "Future Features"



Thank you

Completing our final project wouldn't have been possible without the dedication, support, and hard work of this incredible group. Thank you all!

To our cohort - it has been an absolute pleasure working alongside each of you. You made these 13 weeks of intense effort truly enjoyable. Hope our paths cross again soon!

To John – our deepest gratitude goes to you! You've shared years of knowledge in such a short period and pushed us beyond what we thought we could achieve. We leave this bootcamp feeling confident and prepared for any challenge, thanks to your countless hours of guidance and support. You truly are an amazing tutor, and we are all very fortunate to have met you!