Project Report

# GitHub URL

(insert URL here)

# Abstract

(Short overview of the entire project and features)

# Introduction

(Explain why you chose this project use case)

# Dataset

(Provide a description of your dataset and source. Also justify why you chose this source)

In advance of importing a dataset I developed a DataFrame in Python to practice the course learning as we proceed. The DataFrame showed the number of glasses of wine 4 individuals drank over a week. I call the DataFrame ‘Gwine’.

A screenshot of a computer

Description automatically generated

For my project the dataset I’m using for my project is a CSV file containing information about the times when individuals receive a personal assistant service. This is a sample file, not actual service delivery hours, as to work on a file with accurate information for my project could mean that individuals might be identified, which would be in breach of GDPR. I have modified a database from my work, I chose this database as learning how to manipulate and interrogate this database will allow me to do the same with live accurate information.

There are some abbreviations in the file headings:

* SU Ref = service user reference
* DON = disturbed overnights
* UON = undisturbed overnights

The dataset I’m using is made up of Integers, Floats and Strings.

A screenshot of a computer

Description automatically generated

# I

# Implementation Process

(Describe your entire process in detail)

I imported my CSV file from through a relative path in Python. File name ‘Project\_CSV.csv”. I uploaded it through Pandas, as ‘pahours’.

pahours = pd.read\_csv('Project\_CSV.csv')

I could have also used an:

* Absolute Pathway: C:\Users\michael.doyle\OneDrive - Irish Wheelchair Association\UCDPA
* Internet address in Github: [UCDPA/Project\_CSV.csv at master · Outlier22/UCDPA (github.com)](https://github.com/Outlier22/UCDPA/blob/master/Project_CSV.csv)

Exploring Functions in Pandas

I use the describe function to get an overview of the table I’m using. This showed the range of values in each column, it did not identify that there were any outliers in the table.

* pahours.describe

A screenshot of a computer

Description automatically generated

Using the functions in pandas I set SU Ref as the Index and ran a new datbase:

* UR=pahours.reset\_index('Su Ref')

SURef

A screenshot of a computer

Description automatically generated

To filter the dataset and to get a meaningful return I ran the original dataset over itself:

* Filtering within the file itself to show where ‘Hours per week’ that are less than 5 and ‘Totals Hrs’ are less than 10.

pahours[(pahours['Hours per week']<5) & (pahours['Total Hrs']<10)]

* In the screen shot I also sliced the dataset showing the first 5 rows with SU's and Total hours per week.

pahours[0:5][['Pub Hol Hrs','Sunday Hrs']]

A screenshot of a computer

Description automatically generated

This code - pahours[(pahours['Hours per week']<5) & (pahours['Sunday Hrs']==2) – ‘Hours per week’ less than 5 and ‘Sunday Hrs’ = to 2, but this work as ‘Sunday Hrs” are *Objects* and ‘Hours per week’ are *Floaters*.

Use the following quote to filter both these fields, renaming the database as ‘pahours\_filtted’:

* pahours\_filtered = pahours.loc[(pahours['Hours per week'] < 5.0) & (pahours['Sunday Hrs'] == '2')]

pahours\_filtered.head()

A screenshot of a computer

Description automatically generated

I carried out the cleaning DataFrame function of dropping duplicates the column ‘Hours per week’ using the following code, as I wanted to keep the duplicates I didn’t save the file as itself (writing over the original file) as I wanted to keep duplicates.

* pahours.drop\_duplicates(subset='Hours per week')

A screenshot of a computer

Description automatically generated

Carry out the function of checking for Nulls I ran the following code, which showed the summary of where there were blanks in my database.

* pahours.isna().sum()

A screenshot of a computer

Description automatically generated

To show the average hours per week, which is important information for planning resources etc, I ran the following code:

* ours\_per\_week=pahours.groupby('Hours per week').mean()

Hours\_per\_week.head()

A screenshot of a computer

Description automatically generated

Learning from the course:

# Results

(Include the charts and describe them)

The number of glasses of wine somebody drinks a day per week

# Insights

(Point out at least 5 insights in bullet points)

# References

(Include any references if required)