**SCHOOL OF COMPUTING (SOC)**

**IT8701 Introduction to Programming for Data Science**

**Practical 4 Submission Worksheet (Graded as part of CA3)**

|  |
| --- |
| **Instructoins:**   1. Submit this at Polymall “Lab Submissions->Practical 4 Submission” folder 2. Name your file “YourName-YourStudentID-YourLecturer.docx” |

|  |  |
| --- | --- |
| **Name** | Muhammad Iylia Bin Mohd Hutta |
| **Student ID** | P7474841 |
| **Your Class** | |  | | --- | | NSDDA1/CE/2220/4 | |

# Section 6 Tasks

## **Subset columns**

### Task 3

|  |
| --- |
| mydf = pd.read\_excel('/Users/iylia/Library/Mobile Documents/com~apple~CloudDocs/SP/Intro to Programming for DS/Lab 4/Lab4\_Dataset/singstats\_maritalstatus.xlsx', *index\_col*='Variables')  print('\*\*\* Data in 1980 column \*\*\*')  print(mydf['1980'])  print()  print(f"Number of rows more than 500k is {mydf['1980'][mydf['1980']>500000].count()}")  print(f"Number of rows less than 500k is {mydf['1980'][mydf['1980']<500000].count()}") |

|  |
| --- |
|  |

### Task 4

|  |
| --- |
| df\_2010\_and\_after = mydf.filter(*regex*='201.\*')  print(df\_2010\_and\_after) |

|  |
| --- |
| Graphical user interface, text, application  Description automatically generated |

## **Subset rows**

### Task 3

|  |
| --- |
| df\_rainfall = pd.read\_csv('/Users/iylia/Library/Mobile Documents/com~apple~CloudDocs/SP/Intro to Programming for DS/Lab 4/Lab4\_Dataset/rainfall-monthly-total.csv', *index\_col*='month')  df\_last12 = df\_rainfall[df\_rainfall['total\_rainfall']>300].sort\_values(*by*='total\_rainfall',*ascending*=True).tail(12)  df\_last12.plot.bar()  plt.show() |

|  |
| --- |
|  |

# Section 8 Tasks

## **Drop missing values with dropna**

### Task 1: Drop missing values with dropna

|  |
| --- |
| mydf = pd.read\_excel('/Users/iylia/Library/Mobile Documents/com~apple~CloudDocs/SP/Intro to Programming for DS/Lab 4/Lab4\_Dataset/singstats\_maritalstatus.xlsx', *index\_col*='Variables',*na\_values*='-')  print('\*\*\*\* First 10 rows of original dataset \*\*\*\*')  print(mydf.head(10))  print()  print('\*\*\*\* Remaining dataset after dropping columns with missing data \*\*\*\*')  print(mydf.dropna(*axis*=1).head(10)) |

|  |
| --- |
| Text  Description automatically generated  Graphical user interface, text  Description automatically generated  Graphical user interface, text, application, table, calendar  Description automatically generated  Graphical user interface, text, application  Description automatically generated |