**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Data Exploration and Preparation |
| **Assessment Title:** | CA2: |
| **Lecturer Name:** | Muhammad Iqbal |
| **Student Full Name:** | Muhammad Irfan, Nicholas Chibuike-Eruba, Deisane Pureza, Waseem Ullah |
| **Student Number:** | 2018426, 2021302, 2017328, 2019465 |
| **Assessment Due Date:** | 11/12/2021 |
| **Date of Submission:** | 11/12/2021 |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

**TABLE OF CONTENTS**

**ABSTRACT**

**INTRODUCTION DATA DESCRIPTION TASK DESCRIPTION**

**PART 1**

**DATA EXPLORATION AND PREPARATION**

**1**

*Discuss and describe the purpose and need for data exploration and mining within data analytics projects. Describe the possible consequences of relying on completely automatic data analysis tools rather than using manual statistical tools. Use illustrations to justify your answer.*

**2**

*Are outliers present in the above data set? Provide a strategy of removing outliers from the data set.*

**3** *Choose a relevant encoding technique to transform the categorical data to numeric format.*

**4**

*Formulate the questions for EDA (Exploratory Data Analysis) and perform EDA based on the data provided and address the issues to understand the impact of fire and the resulting burned area.*

**5**

*Explain the purpose of PCA (Principal Component Analysis) and how It is effective to reduce dimensionality of the dataset. Find two components of the above data set.*

**6**

*Analyse the benefits and drawbacks for feature selection and extraction techniques. Which features are important from the provided dataset? (Feature Selection Method)*

**7**

*Prepare the data for the machine learning model after encoding and feature selection, predict the burned area of forest fires using the machine learning model.*

**8**

*You are free to perform any process to enhance an understanding of data preparation and exploration of the Forest data set.*

**PART 2**

**DATA VISUALISATION AND COMMUNICATION**

*Analyse the data, select appropriate visualisations to summarise the data before cleaning it and perform  
the same visualisations after data cleaning. Summarise any differences that you discover. Briefly explain  
the advantages of preparing the data for further analysis in terms of graphic presentation.*

*● Use more than one interactive visualisation to highlight any impact that forest fire has on the  
environment. You are required to use additional data and merge it with the dataset provided to complete  
this task.*

*● Use statistical visualisations to describe and identify different aspects of the data (e.g., median, outliers,  
probabilistic histogram, etc...).*

*● Compare and contrast at least two different visualisation libraries that you used to complete the tasks of  
this assignment. Detail the differences and/or similarities between these libraries. Rationalise your  
design choices such as colour selection, font size, use of white spaces, etc...*

*● Pick a target audience of your choice, design your visualizations with this audience in mind.*

*● Include final visualisations relevant to the fire officers to illustrate your findings*

**PART 3**

**CONSTRAINT SATISFACTION PROBLEM:AI**

*1) In honour of the hard work that the fire officers are doing, the community has decided to build a  
new fire station for them. There are seven fire officers and four offices. Because there are more  
fire officers than offices, some fire officers have to be in the same offices as others. However, the  
fire officers are very picky about who they share with. The community administration is having  
trouble assigning fire officers to offices. They have asked you to plan where each fire officer goes.  
The fire officers are Phylis, Ann, Henry, Eva, Bill, Mark, and Bob.  
They have given you the plans of the fire station.  
Each numbered area is an office in the fire station. Multiple officers can go into the same office,  
and not all offices have to be filled.*

*Each fire person has restrictions about where they can be placed.  
1. Phylis and Eva dont get on, and do not want to be in the same office.  
2. Mark and Bob are best friends, and have to be in the same office.  
3. Henry listens to loud music. Only Eva will share his office.  
4. Eva doesn't talk to Mark, Bob, and Bill.  
5. Ann is always late but is a great fire officer. To hide that Ann is always late, Ann cannot be in  
either the same office or in an office adjacent to Phylis or Eva.  
6. Phylis annoys Bill, so Bill doesn't want to be in Phylis's office.  
7. Phylis is the fire chief, so she wants to be in office 1.  
Using any CSP (Constraint Satisfaction Problem) framework, discover if the above problem can be  
solved and if so detail who would be in each office.*

*2) Discuss in detail how using Constraint Satisfaction finds an answer or finds no solution to the  
problem in Tasks for Artificial Intelligence part 1. How does this differ from standard algorithmic  
solutions?*

*3) Can this problem be solved using any other algorithm we have studied in the module? Discuss your  
answer in detail including a proof of your hypothesis.*

*4) Support your arguments with citations/references in Harvard Style*