**Dummy Variables**

Instructions:

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name:**

**Batch Id: s**

**Topic: Preliminaries for Data Analysis**

**Problem Statement:**

Data is one of the most important assets. It is often common that data is stored in distinct systems with different formats and forms. Non-numeric form of data makes it tricky while developing mathematical equations for prediction models. We have the preprocessing techniques to make the data convert to numeric form. Explore the various techniques to have reliable uniform standard data, you can go through this link:

<https://360digitmg.com/mindmap-data-science>

1. Prepare the dataset by performing the preprocessing techniques, to have the all the features in numeric format.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | Animals | Gender | Homly | Types |
| 1 | Cat | Male | Yes | A |
| 2 | Dog | Male | Yes | B |
| 3 | Mouse | Male | Yes | C |
| 4 | Mouse | Male | Yes | C |
| 5 | Dog | Female | Yes | A |
| 6 | Cat | Female | Yes | B |
| 7 | Lion | Female | Yes | D |
| 8 | Goat | Female | Yes | E |
| 9 | Cat | Female | Yes | A |
| 10 | Dog | Male | Yes | B |

**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Type** | **Relevance** |
| Animals | Name of the animal | Quanlitative, Discrete, Nominal | Relevant |
| Gender | Gender of the animal | Quanlitative, Discrete, Nominal | Relevant |
| Homly | Shows its domestic or not | Quanlitative, Discrete, Nominal | Relevant |
| Types | Type of animal | Qualitative,  Discrete,ordinal | Relevant |

**BUSINESS PROBLEM**: Developing mathematical equations for prediction models.

**CONSTRAINT**: Non numeric form of data.

**DATA UNDERSTANDING:**

1. All the columns of the dataset have string type of data.
2. There are no missing values in the dataset.

**INSIGHTS FROM THE DATA:**

1. Dummy variables are created.
2. One hot encoding is applied which is converted to array then data frame.
3. The input columns and output column are segregated.
4. The non-numeric columns are converted to numeric using label encoder.
5. Then concatenated the input and output columns.

**Hints:**

For each assignment, the solution should be submitted in the below format

1. Work on each feature to create a data dictionary as displayed in the image displayed below:
2. Refer to animal\_category.csv data set.
3. Research and perform all possible steps for obtaining solution
4. All the codes (executable programs) should execute without errors
5. Code modularization should be followed
6. Each line of code should have comments explaining the logic and why you are using that function