

Lab 10:

1. Reading csv file

```
data = pd.read_csv('dataset.csv')
```

2. Printing rows and columns / shape of dataframe

```
print('We have {} rows.'.format(data.shape[0]))
print('We have {} columns'.format(data.shape[1]))
```

3. Checking null values

```
np.sum(pd.isnull(data))
```

4. Checking unique values of specific column

```
data['columnname'].unique()
```

5. Filling null values

```
num = data['columnname'].mode()[0]
data['education'] = data['columnname'].fillna(num)
```

6. Type conversion to int

```
data.columnname.astype(np.int64)
```

7. Dropping unnecessary columns

```
data.drop('id', axis=1, inplace=True)
```

8. Check datatypes of all columns

```
data.dtypes
```

9. Splitting into x and y

```
x = data.iloc[:, 0:-1]
x.shape
y = data.iloc[:, -1]
y.shape
```

10. Converting Object columns into Int columns

```
cat_columns = x.select_dtypes(['object']).columns
x[cat_columns] = x[cat_columns].apply(lambda x: pd.factorize(x)[0])
```

Lab 10 Task

1. (Continue from Lab 9):

- Data Pre-Processing
- Dealing with Null Values
- Changing Datatypes of Columns to be "Int"