



**Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)**

**Факультет «Информатика, искусственный и системы управления»  
Кафедра «Системы обработки информации и управления»**

Отчет по Лабораторной работе №2  
*«Обработка пропусков в данных,  
кодирование категориальных признаков,  
масштабирование данных.»*  
по дисциплине «Технология машинного обучения»

Выполнил:  
студент группы ИУ5-61Б  
И.А. Калинин

Проверил:  
Ю.Е. Гапанюк

2023 г.

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import math as mth
import matplotlib.patches as patches
from scipy import stats as st
plt.rcParams.update({'figure.max_open_warning': 0})
```

```
In [2]: df = pd.read_csv('Alzheimer Disease and Healthy Aging Data In US.csv')

C:\Users\redwh\AppData\Local\Temp\ipykernel_4056\209293884.py:1: DtypeWarning: Columns (13,14) have mixed types
. Specify dtype option on import or set low_memory=False.
df = pd.read_csv('Alzheimer Disease and Healthy Aging Data In US.csv')
```

```
In [3]: df.head()
```

Out[3]:

	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Unit	DataValueTypeID	...	Stratif
0	2020	2020	HI	Hawaii	BRFSS	Overall Health	Arthritis among older adults	Percentage of older adults ever told they have...	%	PRCTG	...	
1	2017	2017	ID	Idaho	BRFSS	Mental Health	Lifetime diagnosis of depression	Percentage of older adults with a lifetime dia...	%	PRCTG	...	
2	2017	2017	ID	Idaho	BRFSS	Overall Health	Arthritis among older adults	Percentage of older adults ever told they have...	%	PRCTG	...	
3	2018	2018	ID	Idaho	BRFSS	Overall Health	Physically unhealthy days (mean number of days)	Physically unhealthy days (mean number of days...	Number	MEAN	...	
4	2020	2020	IN	Indiana	BRFSS	Mental Health	Lifetime diagnosis of depression	Percentage of older adults with a lifetime dia...	%	PRCTG	...	

5 rows × 29 columns

```
In [4]: df['Stratification2'] = df['Stratification2']. fillna ('Male')
```

```
In [5]: df.head()
```

Out[5]:

	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Unit	DataValueTypeID	...	Stratif
0	2020	2020	HI	Hawaii	BRFSS	Overall Health	Arthritis among older adults	Percentage of older adults ever told they have...	%	PRCTG	...	
1	2017	2017	ID	Idaho	BRFSS	Mental Health	Lifetime diagnosis of depression	Percentage of older adults with a lifetime dia...	%	PRCTG	...	
2	2017	2017	ID	Idaho	BRFSS	Overall Health	Arthritis among older adults	Percentage of older adults ever told they have...	%	PRCTG	...	
3	2018	2018	ID	Idaho	BRFSS	Overall Health	Physically unhealthy days (mean number of days)	Physically unhealthy days (mean number of days...	Number	MEAN	...	
4	2020	2020	IN	Indiana	BRFSS	Mental Health	Lifetime diagnosis of depression	Percentage of older adults with a lifetime dia...	%	PRCTG	...	

5 rows × 29 columns

```
In [8]: df["StratificationCategoryID2"].value_counts()
```

Out[8]: RACE 134959  
GENDER 51834  
OVERALL 27669  
Name: StratificationCategoryID2, dtype: int64

```
In [10]: cleanup_nums = {"DataValueTypeID": {"PRCTG": 0, "MEAN": 1},  
                        "StratificationCategoryID2": {"RACE": 1, "GENDER": 2, "OVERALL": 3}}
```

```
In [11]: df = df.replace(cleanup_nums)
```

```
In [12]: pd.get_dummies(df, columns=["ClassID"]).head()
```

Out[12]:	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Unit	DataValueTypeID	...	Stratif
0	2020	2020	HI	Hawaii	BRFSS	Overall Health	Arthritis among older adults	Percentage of older adults ever told they have...	%	0	...	
1	2017	2017	ID	Idaho	BRFSS	Mental Health	Lifetime diagnosis of depression	Percentage of older adults with a lifetime dia...	%	0	...	
2	2017	2017	ID	Idaho	BRFSS	Overall Health	Arthritis among older adults	Percentage of older adults ever told they have...	%	0	...	
3	2018	2018	ID	Idaho	BRFSS	Overall Health	Physically unhealthy days (mean number of days)	Physically unhealthy days (mean number of days...	Number	1	...	
4	2020	2020	IN	Indiana	BRFSS	Mental Health	Lifetime diagnosis of depression	Percentage of older adults with a lifetime dia...	%	0	...	AGE_

5 rows × 35 columns

```
In [ ]:
```