



**Московский государственный технический университет
им. Н.Э. Баумана
(МГТУ им. Н.Э. Баумана)
Радиотехнический факультет (РТ)**

Отчёт по рубежному контролю №1

По дисциплине

«Технологии машинного обучения»

Проверил:

Преподаватель кафедры ИУ-5

Гапанюк Ю.Е.

Подпись: _____

«__» _____ 2020 г.

Выполнил:

студент группы РТ5-61Б

Ануров Н.С.

Подпись: _____

«__» _____ 2020 г.

Москва, 2020

Ануров Никита РК-1 Вариант №3

Задача №1.

Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

Для студентов группы РТ5-61Б - для пары произвольных колонок данных построить график "Jointplot".

```
In [14]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [15]: dt=pd.read_csv('toy_dataset.csv')
```

```
In [16]: dt.shape
```

```
Out[16]: (150000, 6)
```

```
In [8]: dt.head()
```

```
Out[8]:
```

	Number	City	Gender	Age	Income	Illness
0	1	Dallas	Male	41	40367.0	No
1	2	Dallas	Male	54	45084.0	No
2	3	Dallas	Male	42	52483.0	No
3	4	Dallas	Male	40	40941.0	No
4	5	Dallas	Male	46	50289.0	No

Проверим датасет на наличие пропусков

```
In [9]: dt.isnull().sum()
```

```
Out[9]: Number      0
City      0
Gender     0
Age        0
Income     0
Illness    0
dtype: int64
```

```
In [7]: dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150000 entries, 0 to 149999
Data columns (total 6 columns):
Number      150000 non-null int64
City        150000 non-null object
Gender      150000 non-null object
Age         150000 non-null int64
Income      150000 non-null float64
Illness     150000 non-null object
dtypes: float64(1), int64(2), object(3)
memory usage: 6.9+ MB
```

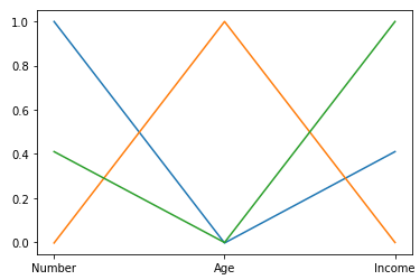
Построим корреляционную матрицу

```
In [17]: df=dt.corr()
df
```

```
Out[17]:
```

	Number	Age	Income
Number	1.000000	-0.003448	0.410460
Age	-0.003448	1.000000	-0.001318
Income	0.410460	-0.001318	1.000000

```
In [18]: plt.plot(df);
```



Построим график Jointplot

```
In [13]: g = sns.jointplot(x="Age", y="Income", data=dt)
```

