Выполнил: Федюкин Д.А.

Вариант: 10

Задание №1: Для набора данных проведите устранение пропусков для одного (произвольного) категориального признака с использованием метода заполнения наиболее распространенным значением.

Задание №2: Для набора данных проведите удаление повторяющихся признаков.

Задание №3: Для произвольной колонки данных построить гистограмму.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import itertools
import scipy.stats as ss
from sklearn.preprocessing import StandardScaler,  MinMaxScaler,
RobustScaler
from sklearn.impute import SimpleImputer
from sklearn.impute import MissingIndicator
from sklearn.impute import KNNImputer
```

```
df = pd.read_csv('restaurants.csv')
numerical = [var for var in df.columns if df[var].dtypes!='O']
categorical = [var for var in df.columns if df[var].dtypes=='O']
df.head()
```

```
                      place_id                                 name  \
0  ChIJx1i4PyCtqjARq5eQI4YeUFE           Jamal Store, Joykul Bazaar   
1  ChIJjyA9oZytqjAR6apb48G7hSY                 Salma Varaitis Store   
2  ChIJFYwq-zkLADoRf_tn0mu_rOQ                   হাজী বিরিয়ানি হাউজ   
3  ChIJPYyqnw8LADoRycl3-GrLje0         নিউ মুসলিম সুইটস এণ্ড বেকারি   
4  ChIJXU_rTB8LADoRYdOJ2LC_Vo4  মেসার্স সততা হোটেল এন্ড রেস্টুরেন্ট   

    latitude  longitude  rating  number_of_reviews  affluence  \
0  22.604275  90.094718     0.0                NaN        NaN   
1  22.619158  90.105594     5.0                1.0        NaN   
2  22.289046  89.958509     5.0                1.0        NaN   
3  22.288710  89.958482     5.0                4.0        NaN   
4  22.286784  89.958116     0.0                NaN        NaN   

                                              address  
0                  Unnamed Road, Kawkhali, Bangladesh  
1            Kawkhali bowlakanda, কাউখালি, Bangladesh  
2  Charkhali - Mathbaria – Patharghata Rd, Mathba...  
3                     সদর রোড, Mathbaria, Bangladesh  
4                     7XP5+P69, Mathbaria, Bangladesh  
```

Задание №1

```
df_new = df.copy()
df_new.affluence.isnull().sum()

10933

def impute_column(dataset, column, strategy_param):
    """
    Заполнение пропусков в одном признаке
    """
    temp_data = dataset[[column]].values
    imputer = SimpleImputer(strategy=strategy_param,
                            fill_value=None)
    all_data = imputer.fit_transform(temp_data)
    dataset[[column]] = all_data
    return dataset

df_new = impute_column(df_new, 'affluence', 'most_frequent')
df_new

                          place_id
name  \
0      ChIJx1i4PyCtqjARq5eQI4YeUFE           Jamal Store, Joykul
Bazaar
1      ChIJjyA9oZytqjAR6apb48G7hSY                 Salma Varaitis
Store
2      ChIJFYwq-zkLADoRf_tn0mu_rOQ                   হাজী বিরিয়ানি হাউজ
3      ChIJPYyqnw8LADoRycl3-GrLje0         নিউ মুসলিম সুইটস এণ্ড বেকারি
4      ChIJXU_rTB8LADoRYdOJ2LC_Vo4  মেসার্স সততা হোটেল এন্ড রেস্টুরেন্ট
...                            ...                                  ..
.
12698  ChIJJWgjO1vv-zkRZHwtUZQGu5I                          Matir
Manus
12699  ChIJiRR5BQbv-zkR5DYjeHQF2l8                      NR Home
Kitchen
12700  ChIJB-JYPK3v-zkRzte9zqVK9vY           Bindu Hotel And
Restaurant
12701  ChIJTVZsObPv-zkRc6h1tV7FzXI            Mostak Hotel &
Restaurant
12702  ChIJ4wTTKbjv-zkRuamuMmWSwJo      ডালাস হোটেল অ্যান্ড রেস্টুরেন্ট

        latitude  longitude  rating  number_of_reviews  affluence  \
0      22.604275  90.094718     0.0                NaN        2.0
1      22.619158  90.105594     5.0                1.0        2.0
2      22.289046  89.958509     5.0                1.0        2.0
3      22.288710  89.958482     5.0                4.0        2.0
4      22.286784  89.958116     0.0                NaN        2.0
...          ...        ...     ...                ...        ...
12698  24.374515  88.604166     0.0                NaN        2.0
12699  24.373602  88.600796     5.0                1.0        2.0
```

```
12700  24.374020  88.603169     3.8              689.0         1.0
12701  24.374332  88.608138     0.0                NaN         2.0
12702  24.374205  88.603853     0.0                NaN         2.0

                                                 address
0                      Unnamed Road, Kawkhali, Bangladesh
1                Kawkhali bowlakanda, কাউখালি, Bangladesh
2      Charkhali - Mathbaria – Patharghata Rd, Mathba...
3                          সদর রোড, Mathbaria, Bangladesh
4                         7XP5+P69, Mathbaria, Bangladesh
...                                                  ...
12698  BSCIC,Industrial Area,Sopura Rajshahi, রাজশাহী...
12699                       Ward-13, Rajshahi, Bangladesh
12700               Station Rd, Rajshahi 6000, Bangladesh
12701   Dhaka Bus Terminal, Seroil, Rajshahi, Bangladesh
12702              New Widened Road, Rajshahi, Bangladesh

[12703 rows x 8 columns]
```

```
df_new.affluence.isnull().sum()
```

```
0
```

Задание №2

```
# Поскольку повторяющихся колонок в датасете нет сгенерим повтор
df['rating_copy'] = df.rating.copy()
df
```

```
                          place_id
name  \
0      ChIJx1i4PyCtqjARq5eQI4YeUFE           Jamal Store, Joykul
Bazaar
1      ChIJjyA9oZytqjAR6apb48G7hSY                 Salma Varaitis
Store
2      ChIJFYwq-zkLADoRf_tn0mu_rOQ                   হাজী বিরিয়ানি হাউজ
3      ChIJPYyqnw8LADoRycl3-GrLje0         নিউ মুসলিম সুইটস এণ্ড বেকারি
4      ChIJXU_rTB8LADoRYdOJ2LC_Vo4  মেসার্স সততা হোটেল এন্ড রেস্টুরেন্ট
...                            ...                                  ..
.
12698  ChIJJWgjO1vv-zkRZHwtUZQGu5I                          Matir
Manus
12699  ChIJiRR5BQbv-zkR5DYjeHQF2l8                      NR Home
Kitchen
12700  ChIJB-JYPK3v-zkRzte9zqVK9vY           Bindu Hotel And
Restaurant
12701  ChIJTVZsObPv-zkRc6h1tV7FzXI            Mostak Hotel &
Restaurant
12702  ChIJ4wTTKbjv-zkRuamuMmWSwJo      ডালাস হোটেল অ্যান্ড রেস্টুরেন্ট
```

```
       latitude  longitude  rating  number_of_reviews  affluence  \
0      22.604275  90.094718     0.0                NaN        NaN
1      22.619158  90.105594     5.0                1.0        NaN
2      22.289046  89.958509     5.0                1.0        NaN
3      22.288710  89.958482     5.0                4.0        NaN
4      22.286784  89.958116     0.0                NaN        NaN
...          ...        ...     ...                ...        ...
12698  24.374515  88.604166     0.0                NaN        NaN
12699  24.373602  88.600796     5.0                1.0        NaN
12700  24.374020  88.603169     3.8              689.0        1.0
12701  24.374332  88.608138     0.0                NaN        NaN
12702  24.374205  88.603853     0.0                NaN        NaN

                                                 address  rating_copy

0                      Unnamed Road, Kawkhali, Bangladesh          0.0

1                Kawkhali bowlakanda, কাউখালি, Bangladesh          5.0

2      Charkhali - Mathbaria – Patharghata Rd, Mathba...          5.0

3                           সদর রোড, Mathbaria, Bangladesh          5.0

4                         7XP5+P69, Mathbaria, Bangladesh          0.0

...                                                  ...          ...

12698  BSCIC,Industrial Area,Sopura Rajshahi, রাজশাহী...          0.0
12699                       Ward-13, Rajshahi, Bangladesh          5.0

12700                 Station Rd, Rajshahi 6000, Bangladesh          3.8

12701   Dhaka Bus Terminal, Seroil, Rajshahi, Bangladesh          0.0

12702                New Widened Road, Rajshahi, Bangladesh          0.0


[12703 rows x 9 columns]
```

```
# Удаляем повтор
df.drop(['rating_copy'], inplace=True, axis=1)
df
```

```
                          place_id
name  \
0      ChIJx1i4PyCtqjARq5eQI4YeUFE           Jamal Store, Joykul
Bazaar
1      ChIJjyA9oZytqjAR6apb48G7hSY                  Salma Varaitis
```

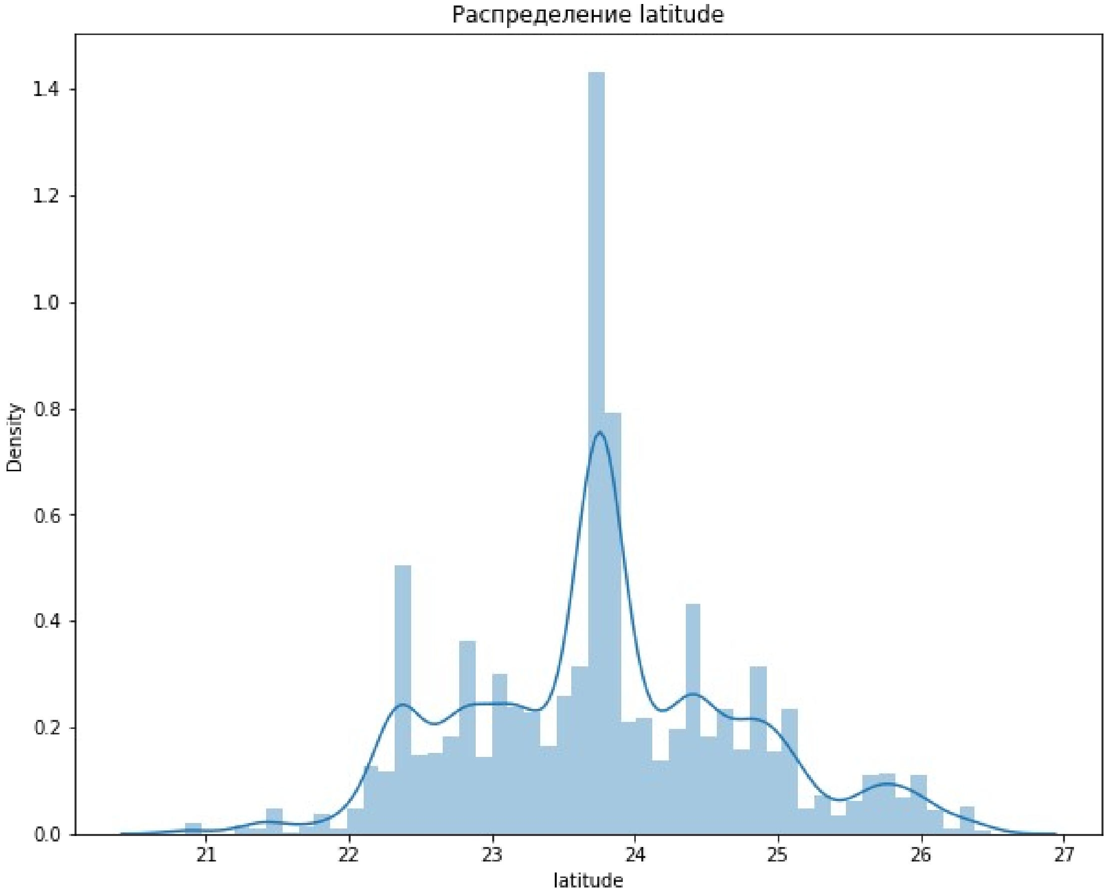
```
Store
2      ChIJFYwq-zkLADoRf_tn0mu_rOQ                   হাজী বিরিয়ানি হাউজ
3      ChIJPYyqnw8LADoRycl3-GrLje0         নিউ মুসলিম সুইটস এণ্ড বেকারি
4      ChIJXU_rTB8LADoRYdOJ2LC_Vo4  মেসার্স সততা হোটেল এন্ড রেস্টুরেন্ট
...                            ...                                  ..
.
12698  ChIJJWgjO1vv-zkRZHwtUZQGu5I                          Matir
Manus
12699  ChIJiRR5BQbv-zkR5DYjeHQF2l8                      NR Home
Kitchen
12700  ChIJB-JYPK3v-zkRzte9zqVK9vY           Bindu Hotel And
Restaurant
12701  ChIJTVZsObPv-zkRc6h1tV7FzXI            Mostak Hotel &
Restaurant
12702  ChIJ4wTTKbjv-zkRuamuMmWSwJo      ডালাস হোটেল অ্যান্ড রেস্টুরেন্ট

        latitude  longitude  rating  number_of_reviews  affluence  \
0      22.604275  90.094718     0.0                NaN        NaN
1      22.619158  90.105594     5.0                1.0        NaN
2      22.289046  89.958509     5.0                1.0        NaN
3      22.288710  89.958482     5.0                4.0        NaN
4      22.286784  89.958116     0.0                NaN        NaN
...          ...        ...     ...                ...        ...
12698  24.374515  88.604166     0.0                NaN        NaN
12699  24.373602  88.600796     5.0                1.0        NaN
12700  24.374020  88.603169     3.8              689.0        1.0
12701  24.374332  88.608138     0.0                NaN        NaN
12702  24.374205  88.603853     0.0                NaN        NaN

                                                 address
0                     Unnamed Road, Kawkhali, Bangladesh
1               Kawkhali bowlakanda, কাউখালি, Bangladesh
2      Charkhali - Mathbaria – Patharghata Rd, Mathba...
3                          সদর রোড, Mathbaria, Bangladesh
4                        7XP5+P69, Mathbaria, Bangladesh
...                                                  ...
12698  BSCIC,Industrial Area,Sopura Rajshahi, রাজশাহী...
12699                      Ward-13, Rajshahi, Bangladesh
12700              Station Rd, Rajshahi 6000, Bangladesh
12701   Dhaka Bus Terminal, Seroil, Rajshahi, Bangladesh
12702             New Widened Road, Rajshahi, Bangladesh

[12703 rows x 8 columns]
```

Задание №3

```
f, ax = plt.subplots(figsize=(10,8))
x = df['latitude']
ax = sns.distplot(x)
```

```
ax.set_title("Распределение latitude")
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```


Распределение latitude
Density
latitude