In [1]:
```
import pandas as pd
import numpy as np
df = df = pd.read_csv('path', encoding = 'cp1251', sep = ';', decimal = ',')
```

In [2]:
```
# Первые строки
df.head()
```

Out[2]:

	new_id	Месяц	Трафик	Средний чек	Дата открытия, категориальный	Торговая площадь, категориальный	Нас
0	0	10	59662	823.060390	Средний по возрасту	Средний	К
1	0	5	56674	859.361975	Средний по возрасту	Средний	К
2	0	1	51488	763.937766	Средний по возрасту	Средний	К
3	0	6	56693	836.362309	Средний по возрасту	Средний	К
4	0	7	58128	845.257709	Средний по возрасту	Средний	К

In [3]:
```
df.shape
```

Out[3]:
```
(256723, 18)
```

In [4]:
```
df.dtypes
```

Out[4]:
```
new_id                                      int64
Месяц                                       int64
Трафик                                      int64
Средний чек                               float64
Дата открытия, категориальный              object
Торговая площадь, категориальный           object
Населенный пункт                           object
Регион                                     object
Численность населения                       int64
Количество домохозяйств                     int64
Трафик пеший, в час                       float64
Трафик авто, в час                        float64
Маркетплейсы, доставки, постаматы (100 м)   int64
Медицинские уч. и аптеки (300 м)            int64
Школы (300 м)                               int64
Остановки (300 м)                           int64
Продуктовые магазины (500 м)                int64
Пятерочки (500 м)                           int64
dtype: object
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 256723 entries, 0 to 256722
Data columns (total 18 columns):
 #   Column                                        Non-Null Count   Dtype  
---  ------                                        --------------   -----  
 0   new_id                                        256723 non-null  int64  
 1   Месяц                                         256723 non-null  int64  
 2   Трафик                                        256723 non-null  int64  
 3   Средний чек                                   256723 non-null  float64
 4   Дата открытия, категориальный                 256723 non-null  object 
 5   Торговая площадь, категориальный              256723 non-null  object 
 6   Населенный пункт                              256723 non-null  object 
 7   Регион                                        256723 non-null  object 
 8   Численность населения                         256723 non-null  int64  
 9   Количество домохозяйств                       256723 non-null  int64  
 10  Трафик пеший, в час                           256723 non-null  float64
 11  Трафик авто, в час                            256723 non-null  float64
 12  Маркетплейсы, доставки, постаматы (100 м)     256723 non-null  int64  
 13  Медицинские уч. и аптеки (300 м)              256723 non-null  int64  
 14  Школы (300 м)                                 256723 non-null  int64  
 15  Остановки (300 м)                             256723 non-null  int64  
 16  Продуктовые магазины (500 м)                  256723 non-null  int64  
 17  Пятерочки (500 м)                             256723 non-null  int64  
dtypes: float64(3), int64(11), object(4)
memory usage: 35.3+ MB
```

In [6]:
```
df.isna().sum()
```

Out[6]:
```
new_id                                        0
Месяц                                         0
Трафик                                        0
Средний чек                                   0
Дата открытия, категориальный                 0
Торговая площадь, категориальный              0
Населенный пункт                              0
Регион                                        0
Численность населения                         0
Количество домохозяйств                       0
Трафик пеший, в час                           0
Трафик авто, в час                            0
Маркетплейсы, доставки, постаматы (100 м)     0
Медицинские уч. и аптеки (300 м)              0
Школы (300 м)                                 0
Остановки (300 м)                             0
Продуктовые магазины (500 м)                  0
Пятерочки (500 м)                             0
dtype: int64
```

In [7]:
```
df['new_id'].nunique()
```

Out[7]:
```
21743
```

In [8]:
```
df.duplicated().sum()
```

Out[8]: np.int64(0)

In [9]:
```
df.duplicated(subset=['new_id', 'Месяц']).sum()
```

Out[9]: np.int64(12)

In [10]:
```
df[df.duplicated(subset=['new_id', 'Месяц'], keep = False)]
```

Out[10]:

	new_id	Месяц	Трафик	Средний чек	Дата открытия, категориальный	Торгова площадь категориальны
165905	13842	9	56253	813.800294	Новый	Средни
165906	13842	5	55744	774.269056	Новый	Средни
165907	13842	11	55034	887.869896	Новый	Средни
165908	13842	12	56601	983.275476	Новый	Средни
165909	13842	4	56410	765.066865	Новый	Средни
165910	13842	1	50437	768.432619	Новый	Средни
165911	13842	7	56573	799.565330	Новый	Средни
165912	13842	8	57029	805.408329	Новый	Средни
165913	13842	10	57715	827.479989	Новый	Средни
165914	13842	6	55787	761.978286	Новый	Средни
165915	13842	2	51525	779.700762	Новый	Средни
165916	13842	3	56140	805.221557	Новый	Средни
165917	13842	9	56109	806.125587	Новый	Средни
165918	13842	5	56025	782.441090	Новый	Средни
165919	13842	11	55719	890.582544	Новый	Средни
165920	13842	12	57147	972.845868	Новый	Средни
165921	13842	4	56097	768.535225	Новый	Средни
165922	13842	1	50699	744.972820	Новый	Средни
165923	13842	7	56671	794.015810	Новый	Средни

	new_id	Месяц	Трафик	Средний чек	Дата открытия, категориальный	Торгова площадь категориальны
165924	13842	8	57210	806.184034	Новый	Средни
165925	13842	10	56821	845.037807	Новый	Средни
165926	13842	6	55563	769.725726	Новый	Средни
165927	13842	2	52282	761.327144	Новый	Средни
165928	13842	3	56454	807.491308	Новый	Средни

In [11]:
```
df1 = df.groupby('new_id')['new_id'].value_counts().reset_index()
df1
```

Out[11]:

	new_id	count
0	0	12
1	1	12
2	2	12
3	3	12
4	4	12
...	...	...
21738	21738	5
21739	21739	5
21740	21740	6
21741	21741	6
21742	21742	5

21743 rows × 2 columns

In [12]:
```
df1[df1['count'] < 12]
```

Out[12]:

	new_id	count
727	727	10
914	914	8
915	915	8
973	973	11
976	976	3
...	...	...
21738	21738	5
21739	21739	5
21740	21740	6
21741	21741	6
21742	21742	5

1107 rows × 2 columns

In [13]:
```
df1[df1['count'] > 12]
```

Out[13]:

	new_id	count
13842	13842	24

In [14]:
```
df = df.drop_duplicates(subset=['new_id', 'Месяц'], keep=False)
```

In [15]:
```
df1 = df.groupby('new_id')['new_id'].value_counts().reset_index()
```

In [16]:
```
df1[df1['count'] > 12]
```

Out[16]:

	new_id	count

In [20]:
```
df2 = df[['new_id','Торговая площадь, категориальный','Населенный пункт','Реги
df2[(df2['Торговая площадь, категориальный'] > 1) | (df2['Населенный пункт'] >
```

Out[20]:

new_id	Торговая площадь, категориальный	Населенный пункт	Регион	Численность населения

In [21]:
```
df.to_csv('path', encoding = 'cp1251')
```