Laboratory work 4
Using Pandas for Data Analysis

 Goal: Learning main Pandas features for data analysis.

2. Tasks:

1. Download US Baby Names dataset from the site kaggle.com (https://www.kaggle.com/kaggle/us-baby-names?select=NationalNames.csv)

2. Output the first 8 rows of the dataset

Expected output:

Out[3]:

	Id	Name	Year	Gender	Count
0	1	Mary	1880	F	7065
1	2	Anna	1880	F	2604
2	3	Emma	1880	F	2003
3	4	Elizabeth	1880	F	1939
4	5	Minnie	1880	F	1746
5	6	Margaret	1880	F	1578
6	7	Ida	1880	F	1472
7	8	Alice	1880	F	1414

3. Output the last 8 rows of the dataset

Expected output:

Out[4]:

	Id	Name	Year	Gender	Count
1825425	1825426	Zo	2014	M	5
1825426	1825427	Zyeir	2014	M	5
1825427	1825428	Zyel	2014	M	5
1825428	1825429	Zykeem	2014	M	5
1825429	1825430	Zymeer	2014	M	5
1825430	1825431	Zymiere	2014	M	5
1825431	1825432	Zyran	2014	M	5
1825432	1825433	Zyrin	2014	M	5

4. Get the names of dataset columns

Expected output:

```
Out[4]: Index(['Id', 'Name', 'Year', 'Gender', 'Count'], dtype='object')
```

5. Get general information about data in the dataset

Expected output:

Out[5]:

	Id	Year	Count
count	1.825433e+06	1.825433e+06	1.825433e+06
mean	9.127170e+05	1.972620e+03	1.846879e+02
std	5.269573e+05	3.352891e+01	1.566711e+03
min	1.000000e+00	1.880000e+03	5.000000e+00
25%	4.563590e+05	1.949000e+03	7.000000e+00
50%	9.127170e+05	1.982000e+03	1.200000e+01
75%	1.369075e+06	2.001000e+03	3.200000e+01
max	1.825433e+06	2.014000e+03	9.968000e+04

6. Find the number of unique names in whole dataset

Expected output:

Out[33]:

93889

7. Calculate the number of unique female and male names in whole dataset

Out[37]:

	Name
Gender	
F	64911
M	39199

8. Find 5 the most popular male names in 2010

Expected output:

	Id	Name	Year	Gender	Count
1677392	1677393	Jacob	2010	M	22082
1677393	1677394	Ethan	2010	M	17985
1677394	1677395	Michael	2010	M	17308
1677395	1677396	Jayden	2010	M	17152
1677396	1677397	William	2010	M	17030

9. Find the most popular name based on the results of one year (the name for which Count is maximum)

Expected output:

```
The name is 'Linda' in 1947
```

10. Count the number of records with Count = minimum.

Expected output:

```
Out[10]: 254615
```

11. Count the number of unique names in each year

Expected output:

```
Out[26]:
```

	Name
Year	
1880	1889
1881	1830
1882	2012
1883	1962
1884	2158

12. Find the year with the most number of unique names.

Expected output:

```
Out[32]:
```

	Name
Year	
2008	32488

13. Find most popular name of the year with the most number of unique names (that is in 2008)

Expected output:

```
Out[24]:

'Jacob'
```

14. Find the year when the name "Jacob" was the most popular as a female name

Expected output:

	Id	Name	Year	Gender	Count
1455556	1455557	Jacob	2004	F	171

15. Find year, with the most number of gender neutral names (the same male and female names)

Expected output:

```
Out[19]:
```

	Gender_neutral_names
Year	
2008	2557

16. Find total births per year

Expected output of the first 5 rows:

Out[56]:

Year	Count
1880	201484
1881	192699
1882	221538
1883	216950
1884	243467

17. Find the year when the greatest number of children was born

Expected output:

Out[49]:

1957

18. Find the number of girls and boys that were born in each year

Expected output of the first 5 rows:

Out[50]:

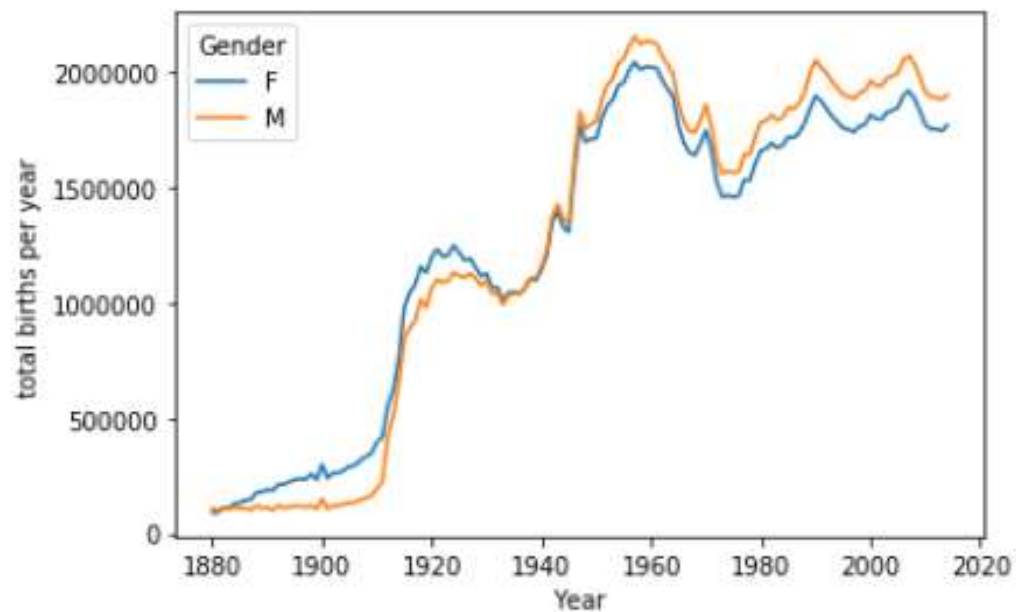| Gender Year	F	M
1880	90993	110491
1881	91954	100745
1882	107850	113688
1883	112321	104629
1884	129022	114445

19. Count the number of years when more girls were born than boys

Expected output:

Out[64]: 54

20. Draw the plot of total births per year of boys and girls

Expected output:

21. Count number of gender neutral names (same for girls and boys)

`Out[85]:` 10221

22. Count how much times boys were named as Barbara

`Out[99]:` 4139

23. Create 2 conditions for data analyzing of this dataset on your own and implement them.

3. The content of the report

1. Cover page of the report.
2. Topic and goal of the lab.
3. Progress of the work.
4. Link to the created Jupyter Notebook on GitHub, rendered by nbviewer.
5. Conclusions.