

```
In [1]: import math
import pandas as pd
```

```
In [4]: def nCr(n,k):
        f=math.factorial
        return f(n)/(f(k)*f(n-k))
```

```
In [5]: nCr(5,2)
```

```
Out[5]: 10.0
```

▼ **Activity: Implement the birthday\_probability function**

```
In [6]: 1-((364/365)**nCr(10,2))
```

```
Out[6]: 0.11614023654879224
```

```
In [7]: def birthday_probability(number_of_people):
        return 1-((364/365)**nCr(number_of_people,2))
```

```
In [8]: birthday_probability(15)
```

```
Out[8]: 0.25028790861398265
```

▼ **NBA Birthday Paradox Analysis**

```
In [31]: df = pd.read_csv('nba_2017.csv', parse_dates=['Birth Date'])
```

```
In [32]: df.head()
```

```
Out[32]:
```

	Player	Pos	Age	Team	Birth Date
0	Alex Abrines	SG	23.0	Oklahoma City Thunder	1993-08-01
1	Quincy Acy	PF	26.0	Dallas Mavericks	1990-10-06
2	Quincy Acy	PF	26.0	Brooklyn Nets	1990-10-06
3	Steven Adams	C	23.0	Oklahoma City Thunder	1993-07-20
4	Arron Afflalo	SG	31.0	Sacramento Kings	1985-10-15

▼ **Activity: Create the Birth Date column**

```
In [33]: df['Birth Date'].dt.strftime("%Y-%m-%d").head()
```

```
Out[33]: 0    1993-08-01
         1    1990-10-06
         2    1990-10-06
         3    1993-07-20
         4    1985-10-15
         Name: Birth Date, dtype: object
```

```
In [34]: df["Birthday"] = df["Birth Date"].dt.strftime("%m-%d")
         df["Birthday"]
```

```
Out[34]: 0      08-01
         1      10-06
         2      10-06
         3      07-20
         4      10-15
         ...
        546     10-05
        547     01-17
        548     09-09
        549     02-18
        550     03-18
         Name: Birthday, Length: 551, dtype: object
```

```
In [35]: df.head()
```

```
Out[35]:
```

	Player	Pos	Age	Team	Birth Date	Birthday
0	Alex Abrines	SG	23.0	Oklahoma City Thunder	1993-08-01	08-01
1	Quincy Acy	PF	26.0	Dallas Mavericks	1990-10-06	10-06
2	Quincy Acy	PF	26.0	Brooklyn Nets	1990-10-06	10-06
3	Steven Adams	C	23.0	Oklahoma City Thunder	1993-07-20	07-20
4	Arron Affalo	SG	31.0	Sacramento Kings	1985-10-15	10-15

## ▼ Interlude: Combinatorics

For this project, you're free to use any technique that you prefer to answer how many players share a birthday for a given team. But, one recommendation would be to use combinatorics; specifically the *Combinations*, using the `itertools.combinations` function. Here's a quick example. Suppose we have these samples:

Name	Birthday
John	March 5th
Mary	Sept 20th
Rob	March 5th

Using combinations, we can take all the samples in pairs (  $r=2$  ) to compare them:

Person 1	Person 2
John	Mary
John	Rob
Mary	Rob

```
"""
"""
```

```
In [22]: from itertools import combinations
```

```
In [23]: names = ["John", "Mary", "Rob", "Susan", "Violet"]
birthdays = ["March 5th", "Sept 20th", "March 5th", "July 28th", "Sept 20th"]
```

```
In [24]: # Note: we need to wrap it in a list to force display
list(combinations(names, 2))
```

```
Out[24]: [('John', 'Mary'),
          ('John', 'Rob'),
          ('John', 'Susan'),
          ('John', 'Violet'),
          ('Mary', 'Rob'),
          ('Mary', 'Susan'),
          ('Mary', 'Violet'),
          ('Rob', 'Susan'),
          ('Rob', 'Violet'),
          ('Susan', 'Violet')]
```

```
In [25]: # Note: we need to wrap it in a list to force display
list(combinations(birthdays, 2))
```

```
Out[25]: [('March 5th', 'Sept 20th'),
          ('March 5th', 'March 5th'),
          ('March 5th', 'July 28th'),
          ('March 5th', 'Sept 20th'),
          ('Sept 20th', 'March 5th'),
          ('Sept 20th', 'July 28th'),
          ('Sept 20th', 'Sept 20th'),
          ('March 5th', 'July 28th'),
          ('March 5th', 'Sept 20th'),
          ('July 28th', 'Sept 20th')]
```

We can see how March 5th (John and Rob) are the same dates. Using Pandas:

```
In [26]: names_df = pd.DataFrame(combinations(names, 2), columns=["Person 1", "Person 2"],  
names_df
```

Out[26]:

	Person 1	Person 2
0	John	Mary
1	John	Rob
2	John	Susan
3	John	Violet
4	Mary	Rob
5	Mary	Susan
6	Mary	Violet
7	Rob	Susan
8	Rob	Violet
9	Susan	Violet

```
In [27]: birthdays_df = pd.DataFrame(combinations(birthdays, 2), columns=["Birthday 1",  
birthdays_df
```

Out[27]:

	Birthday 1	Birthday 2
0	March 5th	Sept 20th
1	March 5th	March 5th
2	March 5th	July 28th
3	March 5th	Sept 20th
4	Sept 20th	March 5th
5	Sept 20th	July 28th
6	Sept 20th	Sept 20th
7	March 5th	July 28th
8	March 5th	Sept 20th
9	July 28th	Sept 20th

Combining it:

```
In [28]: df = pd.concat([names_df, birthdays_df], axis=1)
```

In [29]: df

Out[29]:

	Person 1	Person 2	Birthday 1	Birthday 2
0	John	Mary	March 5th	Sept 20th
1	John	Rob	March 5th	March 5th
2	John	Susan	March 5th	July 28th
3	John	Violet	March 5th	Sept 20th
4	Mary	Rob	Sept 20th	March 5th
5	Mary	Susan	Sept 20th	July 28th
6	Mary	Violet	Sept 20th	Sept 20th
7	Rob	Susan	March 5th	July 28th
8	Rob	Violet	March 5th	Sept 20th
9	Susan	Violet	July 28th	Sept 20th

In [30]: df['Birthday 1'] == df['Birthday 2']

Out[30]:

0	False
1	True
2	False
3	False
4	False
5	False
6	True
7	False
8	False
9	False

dtype: bool

End of the interlude! Now, it's your turn to answer questions.

## ▼ Activities

- ▼ *How many pairs of players share a birthday for the Atlanta Hawks?*

```
In [36]: team_df=df.loc[df["Team"]=="Atlanta Hawks"]
team_df.head()
```

```
Out[36]:
```

	Player	Pos	Age	Team	Birth Date	Birthday
37	Kent Bazemore	SF	27.0	Atlanta Hawks	1989-07-01	07-01
42	DeAndre' Bembry	SF	22.0	Atlanta Hawks	1994-07-04	07-04
75	Jose Calderon	PG	35.0	Atlanta Hawks	1981-09-28	09-28
116	Malcolm Delaney	PG	27.0	Atlanta Hawks	1989-03-11	03-11
130	Mike Dunleavy	SF	36.0	Atlanta Hawks	1954-03-21	03-21

```
In [37]: names_df = pd.DataFrame(combinations(team_df["Player"], 2), columns=["Player 1", "Player 2"])
names_df
```

```
Out[37]:
```

	Player 1	Player 2
0	Kent Bazemore	DeAndre' Bembry
1	Kent Bazemore	Jose Calderon
2	Kent Bazemore	Malcolm Delaney
3	Kent Bazemore	Mike Dunleavy
4	Kent Bazemore	Mike Dunleavy
...	...	...
226	Mike Scott	Edy Tavares
227	Mike Scott	Taurean Waller-Prince
228	Thabo Sefolosha	Edy Tavares
229	Thabo Sefolosha	Taurean Waller-Prince
230	Edy Tavares	Taurean Waller-Prince

231 rows × 2 columns

```
In [38]: birthdays_df = pd.DataFrame(combinations(team_df["Birthday"], 2), columns=["Bi
birthdays_df
```

```
Out[38]:
```

	Birthday 1	Birthday 2
0	07-01	07-04
1	07-01	09-28
2	07-01	03-11
3	07-01	03-21
4	07-01	09-15
...	...	...
226	07-16	03-22
227	07-16	03-22
228	05-02	03-22
229	05-02	03-22
230	03-22	03-22

231 rows × 2 columns

```
In [39]: check_df = pd.concat([names_df, birthdays_df], axis=1)
```

```
In [40]: check_df.head()
```

```
Out[40]:
```

	Player 1	Player 2	Birthday 1	Birthday 2
0	Kent Bazemore	DeAndre' Bembry	07-01	07-04
1	Kent Bazemore	Jose Calderon	07-01	09-28
2	Kent Bazemore	Malcolm Delaney	07-01	03-11
3	Kent Bazemore	Mike Dunleavy	07-01	03-21
4	Kent Bazemore	Mike Dunleavy	07-01	09-15

```
In [42]: (check_df['Birthday 1'] == check_df['Birthday 2']).sum()
```

```
Out[42]: 3
```

```
In [44]: check_df.loc[(check_df['Birthday 1'] == check_df['Birthday 2'])]
```

```
Out[44]:
```

	Player 1	Player 2	Birthday 1	Birthday 2
13	Kent Bazemore	Mike Muscala	07-01	07-01
106	Mike Dunleavy	Dennis Schroder	09-15	09-15
230	Edy Tavares	Taurean Waller-Prince	03-22	03-22

▼ **How many pairs of players share a birthday in the Cleveland Cavaliers?**

```
In [45]: team2_df=df.loc[df["Team"]=="Cleveland Cavaliers"]
team2_df.head()
```

```
Out[45]:
```

	Player	Pos	Age	Team	Birth Date	Birthday
11	Chris Andersen	C	38.0	Cleveland Cavaliers	1978-07-07	07-07
55	Andrew Bogut	C	32.0	Cleveland Cavaliers	1984-11-28	11-28
128	Mike Dunleavy	SF	36.0	Cleveland Cavaliers	1954-03-21	03-21
129	Mike Dunleavy	SF	36.0	Cleveland Cavaliers	1980-09-15	09-15
148	Kay Felder	PG	21.0	Cleveland Cavaliers	1995-03-29	03-29

```
In [46]: names2_df = pd.DataFrame(combinations(team2_df["Player"], 2), columns=["Player", "Player2"])
names2_df
```

```
Out[46]:
```

	Player 1	Player 2
0	Chris Andersen	Andrew Bogut
1	Chris Andersen	Mike Dunleavy
2	Chris Andersen	Mike Dunleavy
3	Chris Andersen	Kay Felder
4	Chris Andersen	Channing Frye
...	...	...
226	Edy Tavares	Deron Williams
227	Edy Tavares	Derrick Williams
228	Tristan Thompson	Deron Williams
229	Tristan Thompson	Derrick Williams
230	Deron Williams	Derrick Williams

231 rows × 2 columns



```
In [50]: birthdays2_df = pd.DataFrame(combinations(team2_df["Birthday"], 2), columns=["
birthdays2_df
```

```
Out[50]:
```

	Birthday 1	Birthday 2
0	07-07	11-28
1	07-07	03-21
2	07-07	09-15
3	07-07	03-29
4	07-07	05-17
...	...	...
226	03-22	06-26
227	03-22	05-25
228	03-13	06-26
229	03-13	05-25
230	06-26	05-25

231 rows × 2 columns

```
In [51]: check2_df = pd.concat([names2_df, birthdays2_df], axis=1)
```

```
In [52]: check2_df.head()
```

```
Out[52]:
```

	Player 1	Player 2	Birthday 1	Birthday 2
0	Chris Andersen	Andrew Bogut	07-07	11-28
1	Chris Andersen	Mike Dunleavy	07-07	03-21
2	Chris Andersen	Mike Dunleavy	07-07	09-15
3	Chris Andersen	Kay Felder	07-07	03-29
4	Chris Andersen	Channing Frye	07-07	05-17

```
In [53]: (check2_df['Birthday 1'] == check2_df['Birthday 2']).sum()
```

```
Out[53]: 1
```

```
In [54]: check2_df.loc[(check2_df['Birthday 1'] == check2_df['Birthday 2'])]
```

```
Out[54]:
```

	Player 1	Player 2	Birthday 1	Birthday 2
219	Iman Shumpert	Deron Williams	06-26	06-26

▼ ***In the Dallas Mavericks, who shares a birthday with J.J. Barea?***

```
In [55]: team3_df=df.loc[df["Team"]=="Dallas Mavericks"]
team3_df.head()
```

```
Out[55]:
```

	Player	Pos	Age	Team	Birth Date	Birthday
1	Quincy Acy	PF	26.0	Dallas Mavericks	1990-10-06	10-06
13	Justin Anderson	SF	23.0	Dallas Mavericks	1993-11-19	11-19
28	J.J. Barea	PG	32.0	Dallas Mavericks	1984-06-26	06-26
29	Harrison Barnes	PF	24.0	Dallas Mavericks	1992-05-30	05-30
45	Ben Bentil	PF	21.0	Dallas Mavericks	1995-03-29	03-29

```
In [56]: names3_df = pd.DataFrame(combinations(team3_df["Player"], 2), columns=["Player", "Player2"])
names3_df
```

```
Out[56]:
```

	Player 1	Player 2
0	Quincy Acy	Justin Anderson
1	Quincy Acy	J.J. Barea
2	Quincy Acy	Harrison Barnes
3	Quincy Acy	Ben Bentil
4	Quincy Acy	Andrew Bogut
...	...	...
271	Dirk Nowitzki	Jarrod Uthoff
272	Dirk Nowitzki	Deron Williams
273	Dwight Powell	Jarrod Uthoff
274	Dwight Powell	Deron Williams
275	Jarrod Uthoff	Deron Williams

276 rows × 2 columns

```
In [57]: birthdays3_df = pd.DataFrame(combinations(team3_df["Birthday"], 2), columns=["birthdays3_df"])
```

```
Out[57]:
```

	Birthday 1	Birthday 2
0	10-06	11-19
1	10-06	06-26
2	10-06	05-30
3	10-06	03-29
4	10-06	11-28
...	...	...
271	06-19	05-19
272	06-19	06-26
273	07-20	05-19
274	07-20	06-26
275	05-19	06-26

276 rows × 2 columns

```
In [58]: check3_df = pd.concat([names3_df, birthdays3_df], axis=1)
```

```
In [59]: check3_df.head()
```

```
Out[59]:
```

	Player 1	Player 2	Birthday 1	Birthday 2
0	Quincy Acy	Justin Anderson	10-06	11-19
1	Quincy Acy	J.J. Barea	10-06	06-26
2	Quincy Acy	Harrison Barnes	10-06	05-30
3	Quincy Acy	Ben Bentil	10-06	03-29
4	Quincy Acy	Andrew Bogut	10-06	11-28

```
In [60]: (check3_df['Birthday 1'] == check3_df['Birthday 2']).sum()
```

```
Out[60]: 1
```

```
In [62]: check3_df.loc[(check3_df['Birthday 1'] == check3_df['Birthday 2'])]
```

```
Out[62]:
```

	Player 1	Player 2	Birthday 1	Birthday 2
65	J.J. Barea	Deron Williams	06-26	06-26

```
In [68]: df["Team"].value_counts()
```

```
Out[68]: Team
New Orleans Pelicans    27
Dallas Mavericks        24
Cleveland Cavaliers    22
Philadelphia 76ers      22
Atlanta Hawks          22
Brooklyn Nets          21
Milwaukee Bucks        20
Oklahoma City Thunder  19
Denver Nuggets         19
Charlotte Hornets      19
Los Angeles Lakers     19
Sacramento Kings       19
Orlando Magic          19
Phoenix Suns           18
Washington Wizards     18
Houston Rockets        18
Chicago Bulls          18
Golden State Warriors  17
Toronto Raptors        17
Memphis Grizzlies      17
Indiana Pacers         17
San Antonio Spurs      17
Minnesota Timberwolves 16
New York Knicks        16
Miami Heat             15
Los Angeles Clippers   15
Portland Trail Blazers  15
Detroit Pistons        15
Utah Jazz              15
Boston Celtics         15
Name: count, dtype: int64
```

```
In [69]: team3_df=df.loc[df["Team"]=="New Orleans Pelicans"]
team3_df.head()
```

```
Out[69]:
```

	Player	Pos	Age	Team	Birth Date	Birthday
5	Alexis Ajinca	C	28.0	New Orleans Pelicans	1988-05-06	05-06
22	Omer Asik	C	30.0	New Orleans Pelicans	1986-07-04	07-04
64	Anthony Brown	SF	24.0	New Orleans Pelicans	1992-10-10	10-10
83	Omri Casspi	SF	28.0	New Orleans Pelicans	1988-06-22	06-22
99	Quinn Cook	PG	23.0	New Orleans Pelicans	1993-03-23	03-23

```
In [70]: names3_df = pd.DataFrame(combinations(team3_df["Player"], 2), columns=["Player", "Player"])
names3_df
```

Out[70]:

	Player 1	Player 2
0	Alexis Ajinca	Omer Asik
1	Alexis Ajinca	Anthony Brown
2	Alexis Ajinca	Omri Casspi
3	Alexis Ajinca	Quinn Cook
4	Alexis Ajinca	DeMarcus Cousins
...	...	...
346	Hollis Thompson	Reggie Williams
347	Hollis Thompson	Reggie Williams
348	Axel Toupane	Reggie Williams
349	Axel Toupane	Reggie Williams
350	Reggie Williams	Reggie Williams

351 rows × 2 columns

```
In [71]: birthdays3_df = pd.DataFrame(combinations(team3_df["Birthday"], 2), columns=["Birthday", "Birthday"])
birthdays3_df
```

Out[71]:

	Birthday 1	Birthday 2
0	05-06	07-04
1	05-06	10-10
2	05-06	06-22
3	05-06	03-23
4	05-06	08-13
...	...	...
346	04-03	03-05
347	04-03	09-14
348	07-23	03-05
349	07-23	09-14
350	03-05	09-14

351 rows × 2 columns

```
In [72]: check3_df = pd.concat([names3_df, birthdays3_df], axis=1)
```

```
In [73]: check3_df.head()
```

```
Out[73]:
```

	Player 1	Player 2	Birthday 1	Birthday 2
0	Alexis Ajinca	Omer Asik	05-06	07-04
1	Alexis Ajinca	Anthony Brown	05-06	10-10
2	Alexis Ajinca	Omri Casspi	05-06	06-22
3	Alexis Ajinca	Quinn Cook	05-06	03-23
4	Alexis Ajinca	DeMarcus Cousins	05-06	08-13

```
In [74]: (check3_df['Birthday 1'] == check3_df['Birthday 2']).sum()
```

```
Out[74]: 0
```