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

▼ *1. What's the probability when `n = 10`?*

```
In [ ]:
```

▼ *2. What's the probability when `n` is 15?*

```
In [ ]:
```

▼ *3. Implement the `birthday_probability` function*

```
In [ ]:  def birthday_probability(number_of_people):
             pass
```

▼ NBA Birthday Paradox Analysis

```
In [17]:  df = pd.read_csv('nba_2017.csv', parse_dates=['Birth Date'])
```

```
In [18]:  df.head()
```

Out[18]:

	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

▼ *4. Create the `Birth Date` column*

In [26]:
```python
df['Birth Date'].dt.strftime("%m-%d")
```

Out[26]:
```
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: Birth Date, Length: 551, dtype: object
```

In [28]:
```python
df["Birthday"] = df['Birth Date'].dt.strftime("%m-%d")
df.head(15)
```

Out[28]:

	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 Afflalo	SG	31.0	Sacramento Kings	1985-10-15	10-15
5	Alexis Ajinca	C	28.0	New Orleans Pelicans	1988-05-06	05-06
6	Cole Aldrich	C	28.0	Minnesota Timberwolves	1988-10-31	10-31
7	LaMarcus Aldridge	PF	31.0	San Antonio Spurs	1985-07-19	07-19
8	Lavoy Allen	PF	27.0	Indiana Pacers	1989-02-04	02-04
9	Tony Allen	SG	35.0	Memphis Grizzlies	1982-01-11	01-11
10	Al-Farouq Aminu	SF	26.0	Portland Trail Blazers	1990-09-21	09-21
11	Chris Andersen	C	38.0	Cleveland Cavaliers	1978-07-07	07-07
12	Alan Anderson	SF	34.0	Los Angeles Clippers	1982-10-16	10-16
13	Justin Anderson	SF	23.0	Dallas Mavericks	1993-11-19	11-19
14	Justin Anderson	SF	23.0	Philadelphia 76ers	1993-11-19	11-19

▼ Interlude: Combinatorics

For this project, you're free to use any techinque 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

Name	Birthday
Mary	Sept 20th
Rob	March 5th

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

Person 1	Person 2
John	Mary
John	Rob
Mary	Rob

In [8]:
```python
from itertools import combinations
```

In [9]:
```python
names = ["John", "Mary", "Rob"]
birthdays = ["March 5th", "Sept 20th", "March 5th"]
```

In [10]:
```python
# Note: we need to wrap it in a list to force display
list(combinations(names, 2))
```

Out[10]: [('John', 'Mary'), ('John', 'Rob'), ('Mary', 'Rob')]

In [11]:
```python
# Note: we need to wrap it in a list to force display
list(combinations(birthdays, 2))
```

Out[11]: [('March 5th', 'Sept 20th'),
 ('March 5th', 'March 5th'),
 ('Sept 20th', 'March 5th')]

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

In [12]:
```python
names_df = pd.DataFrame(combinations(names, 2), columns=["Person 1", "Pe
names_df
```

Out[12]:

	Person 1	Person 2
0	John	Mary
1	John	Rob
2	Mary	Rob

In [13]:
```python
birthdays_df = pd.DataFrame(combinations(birthdays, 2), columns=["Birthd
birthdays_df
```

Out[13]:

	Birthday 1	Birthday 2
0	March 5th	Sept 20th
1	March 5th	March 5th
2	Sept 20th	March 5th

Combining it:

```
In [14]:  df_concat = pd.concat([names_df, birthdays_df], axis=1)
```

```
In [15]:  df_concat
```

Out[15]:

	Person 1	Person 2	Birthday 1	Birthday 2
0	John	Mary	March 5th	Sept 20th
1	John	Rob	March 5th	March 5th
2	Mary	Rob	Sept 20th	March 5th

```
In [16]:  df_concat['Birthday 1'] == df_concat['Birthday 2']
```

```
Out[16]:  0    False
          1     True
          2    False
          dtype: bool
```

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

▼   **Activities**

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

In [32]:
```python
filt = df['Team'] == 'Atlanta Hawks'
hawks_df = df[filt]
hawks_df
```

Out[32]:

	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
131	Mike Dunleavy	SF	36.0	Atlanta Hawks	1980-09-15	09-15
192	Tim Hardaway	SG	24.0	Atlanta Hawks	1966-09-01	09-01
193	Tim Hardaway	SG	24.0	Atlanta Hawks	1992-03-16	03-16
231	Dwight Howard	C	31.0	Atlanta Hawks	1985-12-08	12-08
234	Kris Humphries	PF	31.0	Atlanta Hawks	1985-02-06	02-06
241	Ersan Ilyasova	PF	29.0	Atlanta Hawks	1987-05-15	05-15
275	Ryan Kelly	PF	25.0	Atlanta Hawks	1991-04-09	04-09
279	Kyle Korver	SG	35.0	Atlanta Hawks	1981-03-17	03-17
344	Paul Millsap	PF	31.0	Atlanta Hawks	1985-02-10	02-10
358	Mike Muscala	C	25.0	Atlanta Hawks	1991-07-01	07-01
363	Gary Neal	SG	32.0	Atlanta Hawks	1984-10-03	10-03
393	Lamar Patterson	SG	25.0	Atlanta Hawks	1991-08-12	08-12
445	Dennis Schroder	PG	23.0	Atlanta Hawks	1993-09-15	09-15
447	Mike Scott	PF	28.0	Atlanta Hawks	1988-07-16	07-16
448	Thabo Sefolosha	SF	32.0	Atlanta Hawks	1984-05-02	05-02
472	Edy Tavares	C	24.0	Atlanta Hawks	1992-03-22	03-22
511	Taurean Waller-Prince	SF	22.0	Atlanta Hawks	1994-03-22	03-22

In [32]:
```python
filt = df['Team'] == 'Atlanta Hawks'
hawks_df = df[filt]
hawks_df
```

In [34]:
```python
names_df = pd.DataFrame(combinations(hawks_df['Player'], 2), columns=["P"
names_df
```

Out[34]:

	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 [37]:
```python
birthdays_df = pd.DataFrame(combinations(hawks_df['Birthday'], 2), colum
birthdays_df
```

Out[37]:

	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]:
```python
df_concat = pd.concat([names_df, birthdays_df], axis=1)
df_concat
```

Out[39]:

	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
...	...	...	...	...
226	Mike Scott	Edy Tavares	07-16	03-22
227	Mike Scott	Taurean Waller-Prince	07-16	03-22
228	Thabo Sefolosha	Edy Tavares	05-02	03-22
229	Thabo Sefolosha	Taurean Waller-Prince	05-02	03-22
230	Edy Tavares	Taurean Waller-Prince	03-22	03-22

231 rows × 4 columns

In [46]:
```python
(df_concat['Birthday 1'] == df_concat['Birthday 2']).sum()
```

Out[46]: 3

In [47]:
```python
filt = df_concat['Birthday 1'] == df_concat['Birthday 2']
df_concat.loc[filt]
```

Out[47]:

	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

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

In [ ]:

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

In [44]:
```python
filt = df['Team'] == 'Dallas Mavericks'
mavs_df = df[filt]
mavs_df
```

Out[44]:

	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
54	Andrew Bogut	C	32.0	Dallas Mavericks	1984-11-28	11-28
68	Nicolas Brussino	SF	23.0	Dallas Mavericks	1993-03-02	03-02
98	Quinn Cook	PG	23.0	Dallas Mavericks	1993-03-23	03-23
108	Seth Curry	PG	26.0	Dallas Mavericks	1990-08-23	08-23
152	Yogi Ferrell	PG	23.0	Dallas Mavericks	1993-05-09	05-09
153	Dorian Finney-Smith	PF	23.0	Dallas Mavericks	1993-05-04	05-04
170	Jonathan Gibson	PG	29.0	Dallas Mavericks	1987-11-08	11-08
191	A.J. Hammons	C	24.0	Dallas Mavericks	1992-08-27	08-27
198	Devin Harris	PG	33.0	Dallas Mavericks	1983-02-27	02-27
201	Manny Harris	SG	27.0	Dallas Mavericks	1989-09-21	09-21
247	Pierre Jackson	PG	25.0	Dallas Mavericks	1991-08-29	08-29
300	DeAndre Liggins	SG	28.0	Dallas Mavericks	1988-03-31	03-31
318	Wesley Matthews	SG	30.0	Dallas Mavericks	1986-10-14	10-14
337	Salah Mejri	C	30.0	Dallas Mavericks	1986-06-15	06-15
371	Nerlens Noel	C	22.0	Dallas Mavericks	1994-04-10	04-10
374	Dirk Nowitzki	PF	38.0	Dallas Mavericks	1978-06-19	06-19
412	Dwight Powell	C	25.0	Dallas Mavericks	1991-07-20	07-20
497	Jarrod Uthoff	PF	23.0	Dallas Mavericks	1993-05-19	05-19
524	Deron Williams	PG	32.0	Dallas Mavericks	1984-06-26	06-26

▼ The End!