ProjTwo

January 8, 2026

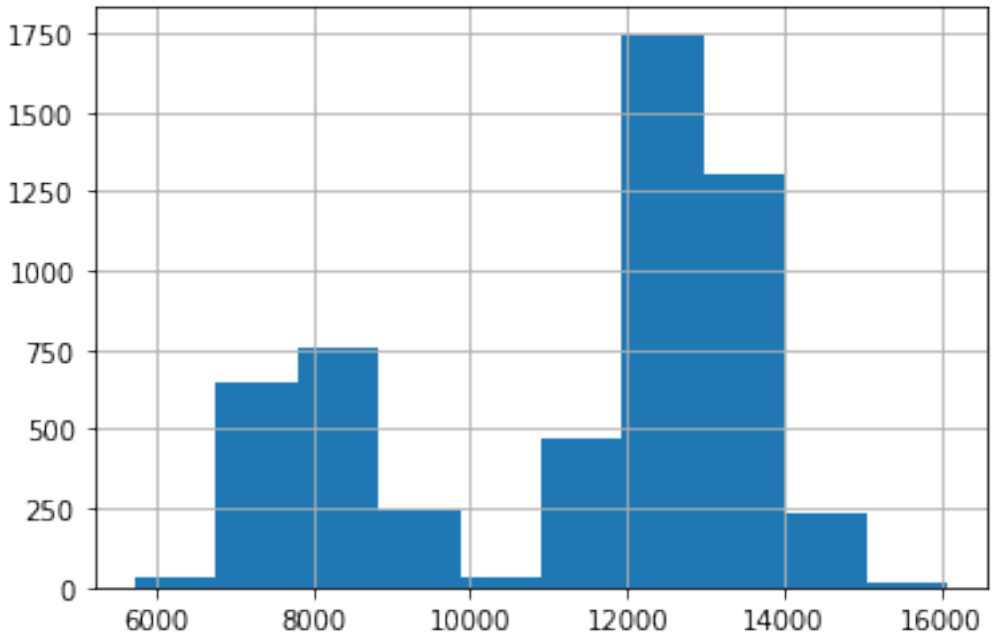
Analyzing Birth Trends in the U.S. (2000–2014): Patterns or Randomness?

Ashley Funes

Context: The dataset contains the number of births that occurred in the United States on each day from January 1st, 2000 through December 31st, 2014. I will investigate whether there are any trends in births or were born purely at random across the days of this 14 year period.

[4]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#df
stats_b = pd.read_csv("shared/data/US_births_2000-2014.csv", parse_dates␣
 ↪=['timestamp'])

stats_b["births"].hist()
```

[4]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7c3bd6887550>
```



The histogram's shape is bimodal. Although the histogram shape goes up, down, up, down, the left side is shorter than the right side.

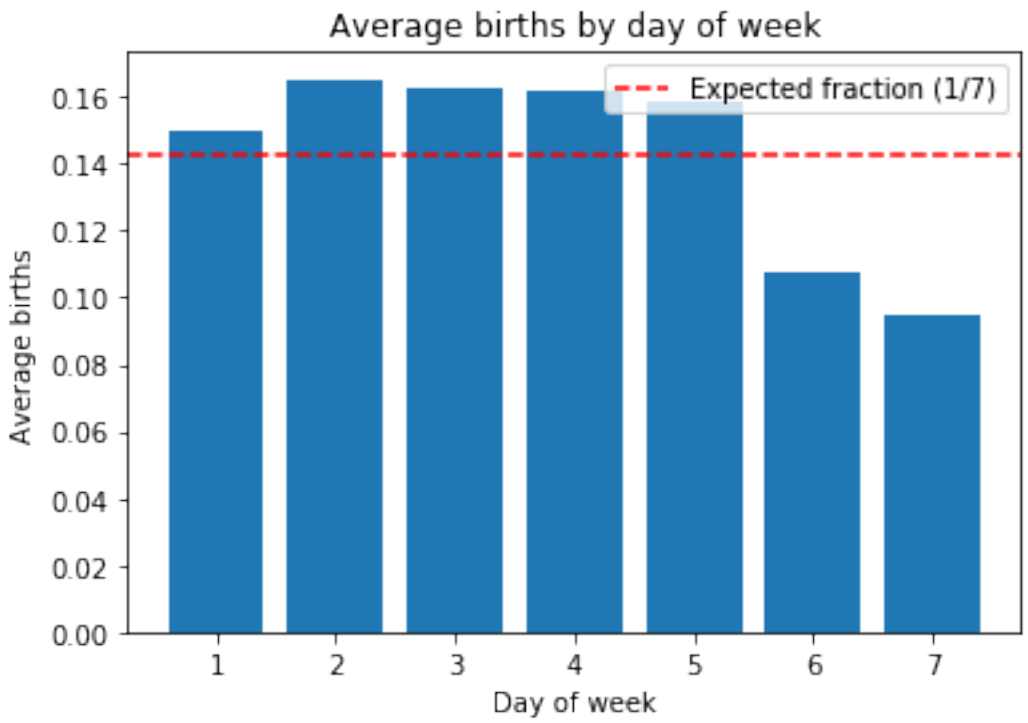
The proportion of all births on a given day of the week would be 1/7th because there are 7 days in a week. The proportion of births likely to occur in a given month would be 1/12th because there are 12 months in a year.

Births by Day of week

[10]:
```
day = stats_b.groupby("day_of_week")["births"].sum()

birth_totals = day.tolist()
print("birth_totals:", birth_totals)
print()

total = sum(birth_totals)
print("Total:", total)
print()

avg = day / total
print(avg)
print()

plt.bar(avg.index, avg.values)
plt.title("Average births by day of week")
plt.xlabel("Day of week")
plt.ylabel("Average births")
print(avg.index)
print()
expected_fraction = 1/7
plt.axhline(y=expected_fraction, color="red", linestyle='--', label ='Expected␣
 ↪fraction (1/7)')
plt.legend()
plt.show()
```

```
birth_totals: [9316001, 10274874, 10109130, 10045436, 9850199, 6704495, 5886889]

Total: 62187024

day_of_week
1    0.149806
2    0.165225
3    0.162560
4    0.161536
5    0.158396
6    0.107812
```

```
7    0.094664
Name: births, dtype: float64

Int64Index([1, 2, 3, 4, 5, 6, 7], dtype='int64', name='day_of_week')
```

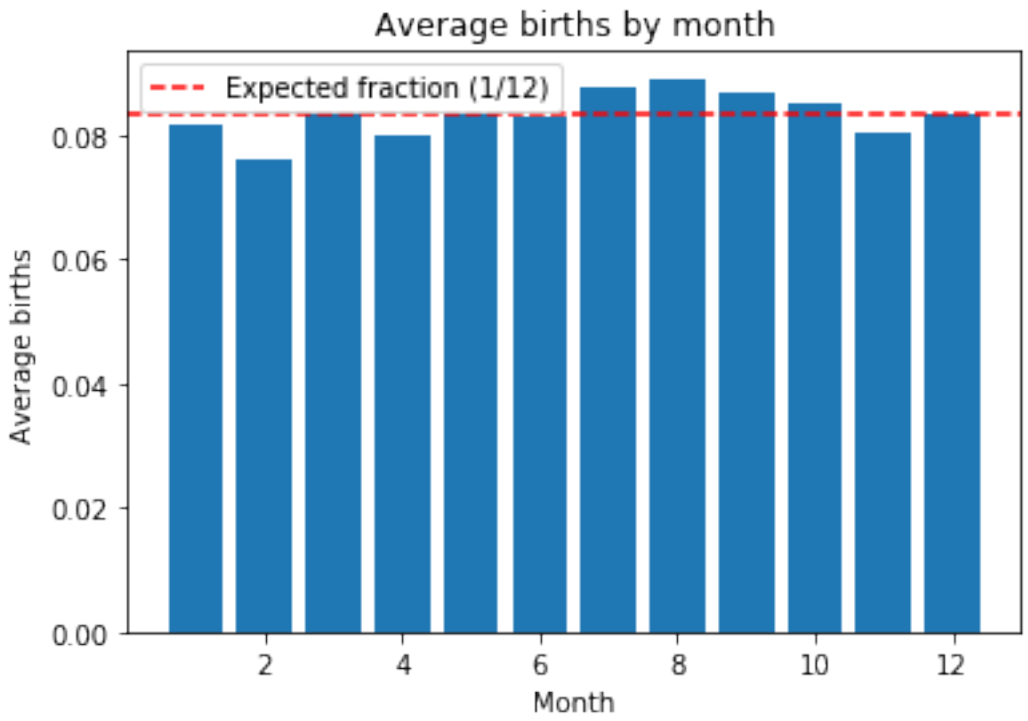


Births by Months

[9]:
```
month = stats_b.groupby("month")["births"].sum()

mbirth_totals = month.tolist()
print("month birth totals:", mbirth_totals)
print()

mtotal = sum(mbirth_totals)
print("Total:", mtotal)
print()

mavg = month / mtotal
print(mavg)

plt.bar(mavg.index, mavg.values)
plt.title("Average births by month")
plt.xlabel("Month")
```

```
plt.ylabel("Average births")
print(avg.index)
print()

expected_fraction = 1/12
plt.axhline(y=expected_fraction, color="red", linestyle='--', label ='Expected␣
 ↪fraction (1/12)')
plt.legend()
plt.show()
```

```
month birth totals: [5072588, 4725693, 5172961, 4960750, 5195445, 5163360,
5450418, 5540170, 5399592, 5302865, 5008750, 5194432]

Total: 62187024

month
1     0.081570
2     0.075992
3     0.083184
4     0.079771
5     0.083545
6     0.083030
7     0.087646
8     0.089089
9     0.086828
10    0.085273
11    0.080543
12    0.083529
Name: births, dtype: float64
Int64Index([1, 2, 3, 4, 5, 6, 7], dtype='int64', name='day_of_week')
```

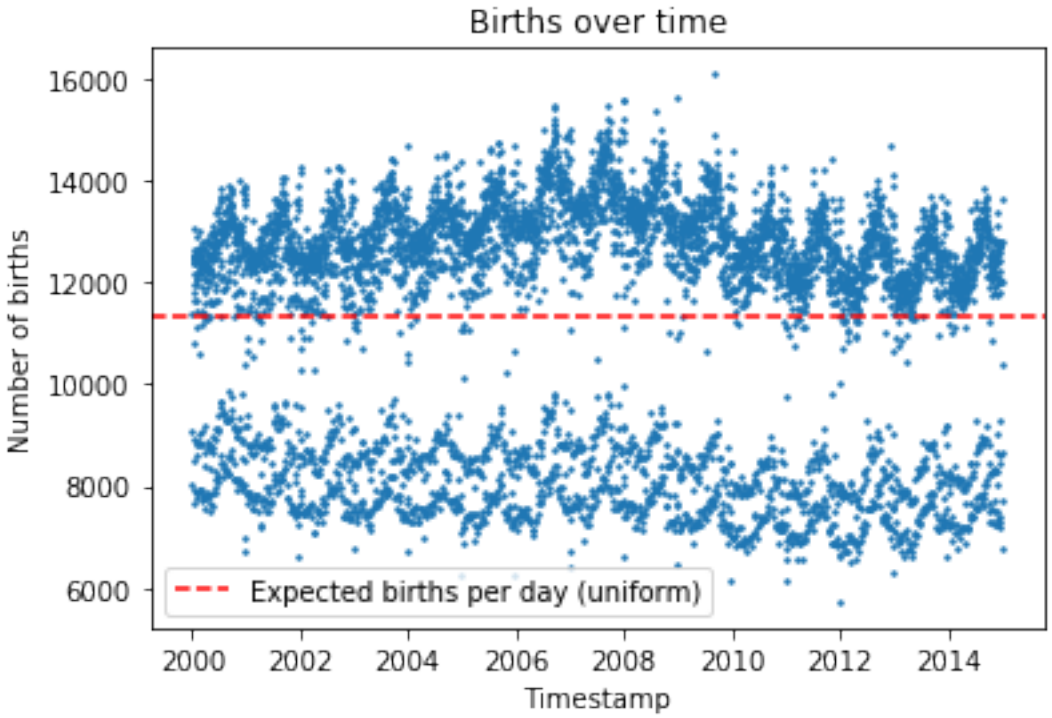


Births Over Time, Analysis

[121]:
```
y = stats_b["births"]
x = stats_b["timestamp"]


plt.scatter(x, y, s=2)
plt.title("Births over time")
plt.xlabel("Timestamp")
plt.ylabel("Number of births")

timestamp = stats_b.groupby("timestamp")["births"].sum()
print(timestamp)
print()

total = timestamp.sum()
print("Total:", total)

days_ = len(timestamp)
print("# of days:", days_)

e_births = total / days_
print("Expected births per day:", e_births)
```

```
plt.axhline(y=e_births, color="red", linestyle='--', label ='Expected births␣
 ↪per day (uniform)')
plt.legend()
plt.show()
```

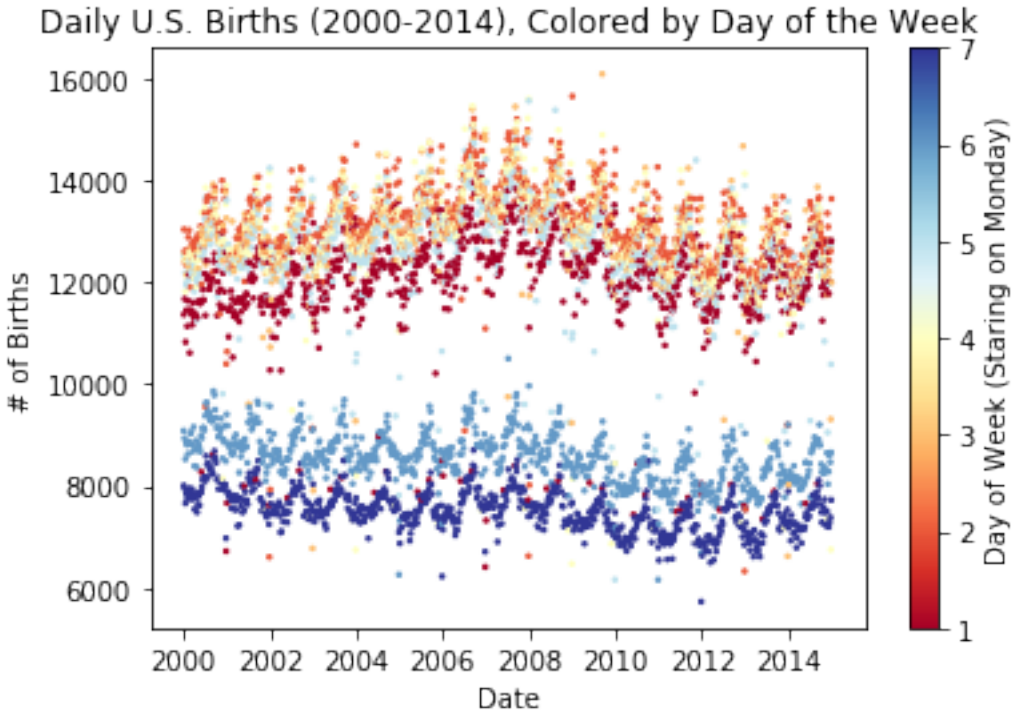
```
timestamp
2000-01-01     9083
2000-01-02     8006
2000-01-03    11363
2000-01-04    13032
2000-01-05    12558
              ...
2014-12-27     8656
2014-12-28     7724
2014-12-29    12811
2014-12-30    13634
2014-12-31    11990
Name: births, Length: 5479, dtype: int64

Total: 62187024
# of days: 5479
Expected births per day: 11350.068260631502
```



The group on top is the births occurring Monday through Friday; the bottom group is births

occurring Saturday to Sunday. The weekday group is more populated, meaning more births occur on weekdays. This is due to scheduling procedures like C-sections, which are preferably scheduled during a hospital's regular operating hours. While natural births populate all days, in conclusion, human intervention is a factor in the number of births on certain days of the week.

[122]:
```
plt.scatter(stats_b["timestamp"], stats_b["births"], c=stats_b["day_of_week"],␣
↪cmap="RdYlBu", s=2)

plt.colorbar(label="Day of Week (Staring on Monday)")
plt.title("Daily U.S. Births (2000-2014), Colored by Day of the Week")
plt.xlabel("Date")
plt.ylabel("# of Births")
plt.show()
```



It seems that the highest birth rate was between 2006-2007. This could be due to economic hardships from the 2008 recession that led to many families not having children.