Water Quality at Toronto Beaches*

An analysis of e-coli rates in toronto beaches 2007-2024

Jacob Gilbert

September 26, 2024

This study analyses E. coli contamination trends at two of Toronto's popular beaches, Sunnyside Beach and Marie Curtis Park East Beach, from 2007 to 2024. Using data sourced from Open Data Toronto, the analysis examines temporal patterns in contamination levels, assessing the frequency and severity of unsafe E. coli concentrations as defined by public health thresholds. The study reveals that while E. coli levels appear less erratic and generally lower in magnitude in recent years, the frequency of tests exceeding unsafe thresholds has not significantly diminished, indicating persistent contamination risks. These findings underscore the need for continued monitoring and targeted interventions to protect public health. This research aims to inform beach management practices and guide future efforts to improve water quality at Toronto's beaches, Aswell as provide insights for swimmers on the risks that they are exposed to.

Table of contents

*Code and data are available at: https://github.com/JfpGilbert0/toronto-beaches-water-quality.

Introduction

Water contamination is important in Toronto, especially in the summer months as locals flock to the nearby beaches as the weather turns from snow to sun. E-coli testing is common at most public swimming locations as high levels highlight risk to those embarking into the water. While E-coli is unlikely to cause illness on its own, If a water sample has high concentrations of E. coli, other more dangerous and infectious organisms may be present (Alliance, n.d.). Thus the levels at public beaches can affect public health, and as such action should be taken on the results. To either close public beaches that exhibit unsafe levels, or continuation of previous action that has led to lower levels.

Despite these public health measures, there is limited understanding of the specific patterns of E. coli contamination in Toronto's beaches. Most public advisories and beach management decisions rely on reactive measures rather than proactive analysis of historical trends. This paper aims to fill this gap by examining historical E. coli data from Open Data Toronto, focusing on two of the city's popular beaches: Sunnyside Beach and Marie Curtis Park East Beach. By analysing data over several years, this study seeks to identify temporal patterns in contamination, such as variations by day of the week and season, to help answer the critical question: When are beachgoers at the highest risk of swimming in contaminated waters?

Additionally, this analysis explores whether E. coli levels have shown improvement over time, informing on the effectiveness of current water management practices. Understanding the current and historic trends are paramount for city officials to understand if current water quality measures are sufficient. Or in the worse case, that interventions are needed in order to keep Toronto beaches safe for everyone.

This paper is structured as follows: First, we present the dataset, describing its sources, key variables, and the data cleaning process. Next, we analyse patterns in E. coli levels, examining how contamination varies across different times and locations. We then discuss the implications of these findings for public health and beach management, followed by a conclusion that summarises the key insights and recommendations for future action.

Referencing

In this paper, Python (Python Software Foundation 2024) and its packages are utilised as the primary programming language for the following analysis and displaying the data. Key packages used include pandas [McKinney, 2010], which facilitated efficient data cleaning, manipulation, and summarization of the dataset, NumPy (Harris et al. 2020) was employed for numerical operations, Matplotlib (Hunter 2007) and seaborn (Waskom 2021) were critical for creating informative visualisations.

Data

Data Source

Opendatatoronto is a publicly available collection of wide ranges of data. This resource was used to obtain the e-coli levels of the water at two of Toronto's many beaches. The data set included coli levels from the SunnySide and Marie Curtis park east beaches, from the 3rd of July 2007 to September 5th 2024. The large data range allows us to extrapolate trends well and further the recency of the data allows us to generate a good understanding of the present. We have a total of 3084 data points from the two beaches, 1370 and 1714 data points respectively. Each detailing the collectiondate, beachname, sitename, geometric location (given by longitude and latitude) and the e coli levels.

	Metric	Value
0	Data points	3084
1	Date Range	2007-06-03 to 2024-09-05
2	Unique Beach Names	2
3	N0 of Unique Site Names	12
4	No of Unique Locations	12

Figure 0a Shows counts of unique variables and the date range of the data

	Statistic	Value
0	Count	3084.0
1	Mean	389.0
2	Standard Deviation	935.0
3	Min	10.0
4	25th Percentile (Q1)	10.0
5	Median (Q2)	10.0
6	75th Percentile (Q3)	10.0
7	Max	9000.0
8	Variance	873898.0

Figure 0b Shows summary statistics for e-coli levels over the whole dataset

Measurement

E-coli is measured in MPN/100ml, a measurement that reflects the e coli particles in every 100 ml of lake water. For our data presentation log values are used to give a better visualisation of the data. To obtain an accurate representation of the beach as a whole 12 different locations here used to gather data across the 2 beaches as to remove bias that one testing location may result in. The result is that we obtain an unbiased representation of e coli levels at each beach as a whole over more than a decade. Some missing data we encounter however is that although we have the collection date we do not observe the time that data was collected. Missing this data may lead to some bias as the time of day could impact the baseline levels of e-coli in the water. What this form of measurement does allow us to do is look at the differences in water quality at a date level, Weekly section shows there is a good distribution of data from each day of the week. There are some gaps in the data however, in the monthly data we only find results from the summer months. As during the summer the impact of water quality is at its highest this doesn't impact the usefulness of the data, however it does result in missing context of levels during the rest of the year.

Methods

Before we look at the daily differences in levels we will look at the data as a whole, observing the distribution of higher and lower levels of e coli in the water across time as well as by beach. The Ontario government sets their standard of 100 mpn/100ml as their water quality objectives (Ontario Ministry of the Environment, Conservation and Parks 1994). Thus we will use this as our benchmark for safe e-coli levels in the Beach water. Finally we will look at these distributions over the course of a week to answer the question of when water quality is at its best, and as a lake swimmer when is it safest for me to swim at these beaches. For average levels we will focus on data after 2017. This is because of the large difference in the variation of the data before and after this year. This could be due to a change in testing strategy or something else not represented in the data, making the actual levels reported on either side of this year not comparable. The years after 2017 were selected in order to compare more recent data to provide better context. The method used in this paper to look at if values are above and below a threshold allow us to compare and use the data as a whole.

Results

Time Series

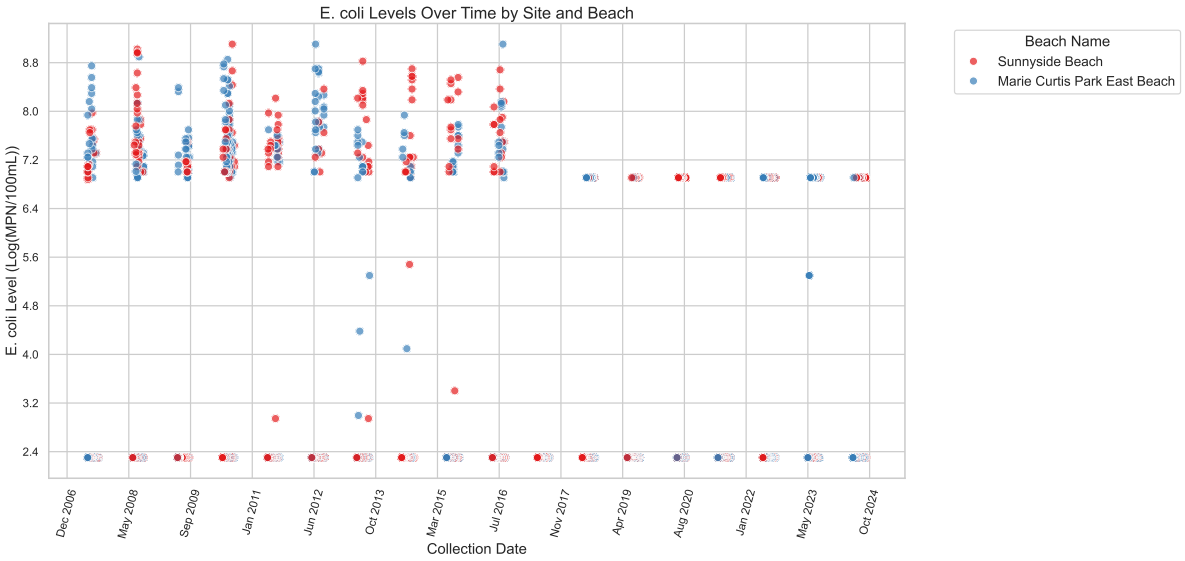

Figure 1

Figure 1 depicts E. coli contamination levels over time at Sunnyside Beach and Marie Curtis Park East Beach from 2008 to 2024. The scatter plot shows that E. coli levels were highly variable and frequently above the unsafe threshold in the early years, especially between 2008 and 2013. From 2017 onwards, E. coli levels appear less erratic, with fewer instances of extreme spikes

Figure 1 displays our data as a whole, we see very widely distributed values of what could be considered high levels of e-coli from 2007 to 2016. We see the highest density of data at 10 MLN/100ml but on the higher end levels range from 1000 to 9000. Closer to the present we do not see as much variation on the higher end. In fact post 2017 the data groups either at the low level of 10 or at around 1000 almost exclusively. The variations observed appear to not vary by beach in figure.

Figure 1 is a scatter plot of E. coli levels over time by site and beach, and displays the data as a whole. It highlights the trends in contamination levels at Sunnyside Beach and Marie Curtis Park East Beach from 2008 to 2024.the beach site itself does not seem to significantly impact the overall trend of E. coli levels over time, as both Sunnyside and Marie Curtis Park East show similar patterns of fluctuation. The magnitude of the fluctuations are very notable as prior to 2017 there are spikes in e-coli levels to many degrees, some extremely high. Whereas after 2017 we see levels almost exclusively sitting at 10 or around 1000 MNP/100ml. This

appears to indicate that although the risks are persistent over time they are lessened in recent years.

Figure 2 is a bar chart illustrating the percentage of tests above the unsafe threshold by year, providing an important contrast to the scatter plot. Although the scatter plot suggests a reduction in extreme E. coli levels in the later years, this bar graph reveals that the frequency of tests exceeding the 100 MPN/100mL threshold does not diminish in the same way. Years such as 2010, 2011, 2019, and 2021 see substantial spikes in the percentage of unsafe tests, reaching as high as over 50% in 2010. We do not see percentages moving consistently in one direction other than after 2017 where the rates appear to be climbing. 2017 is a strange anomaly here with no tests returning unsafe results. The combination of these two graphs paint a clear picture of water quality in Toronto over the years.

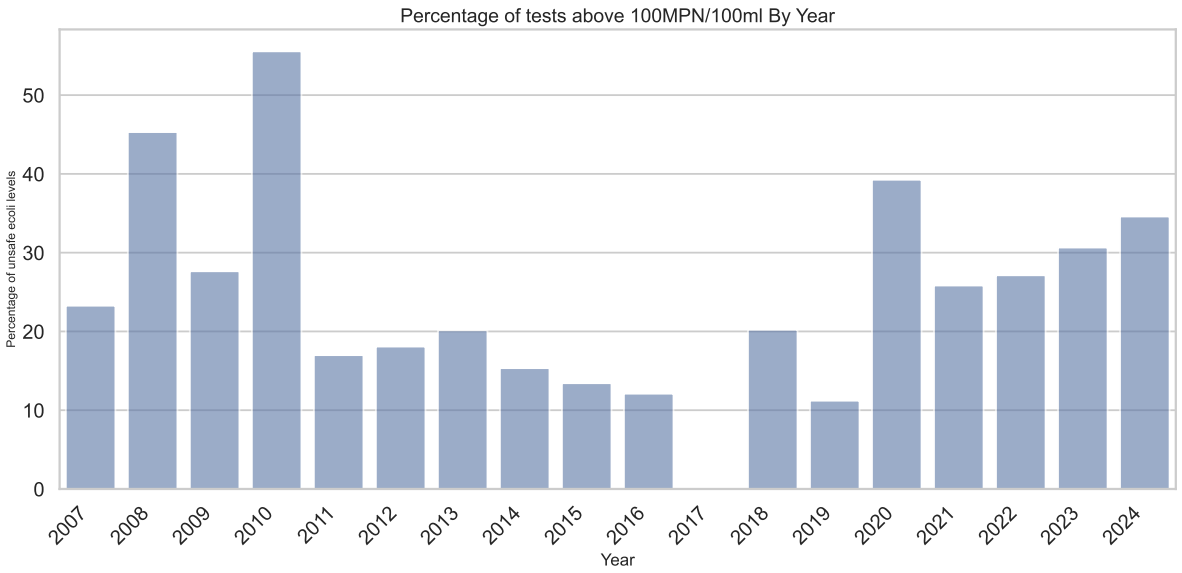

Figure 2

Figure 2: Bar graph reflecting the percentage of results each year where the water quality exceeds 100 MPN/100ml. There is no obvious pattern here although in later years we see a trend upwards and rates appear to trend down until 2017 where the rate is 0.

Monthly data

	month	n	average_ecoli	variance	unsafe_levels	percentage_unsafe
0	May	341	102.0	167326.0	21	6.16
1	June	1046	252.0	598139.0	146	13.96
2	July	910	548.0	1296202.0	277	30.44
3	August	681	549.0	1058926.0	224	32.89
4	September	106	271.0	230356.0	26	24.53

Figure 3: Monthly summary statistics indicate May exhibits the lowest average E. coli levels, while July and August have the highest, with over 30% of samples deemed unsafe.

Figure 3 looks at the monthly data at both beaches over the whole dataset, as mentioned in the data section only the summer months provide data due to the cold weather in Toronto limiting available data in the winter. Further we see a disproportionate number of data points in May and September, this could make the results here less reliable. However this data can still provide useful data. May exhibits the lowest average E. coli levels, at 101.76 MPN/100mL, which is just above the unsafe threshold. With only 6.16% of samples being unsafe. The low variance suggests that water quality is fairly consistent during this early summer month. August and July show extremely high averages, above 500 MPN/100ml, with over 30% of data collected here showing water quality above acceptable levels. Figure 4 shows that these trends are exhibited in both beaches to varying degrees post 2017. The high levels in the later 3 months are reflected at both beaches, though we do observe extremely high average levels at Sunnyside compared to Marie Curtis in this period. Contrarily at the beginning of the summer there is stellar water quality at sunnyside in May. Marie curtis on the other hand averages above threshold e-coli levels in all months. Overall, we see the average level as well as the likelihood of unsafe levels increase as the summer progresses and slightly taper off in september. Although the absence of a reasonable amount of data in September could contribute to the ending of this increasing trend.

figure 4 A bar graph showing the percent of results that are above the government target of 100 mpn/100ml grouped by month. The rates peak at both beaches in the middle months. Sunnysde experienced the highest rates of the two locations but also the lowest. Marie curtis had a bell-curve type of trend over the 4 summer months peaking in July.

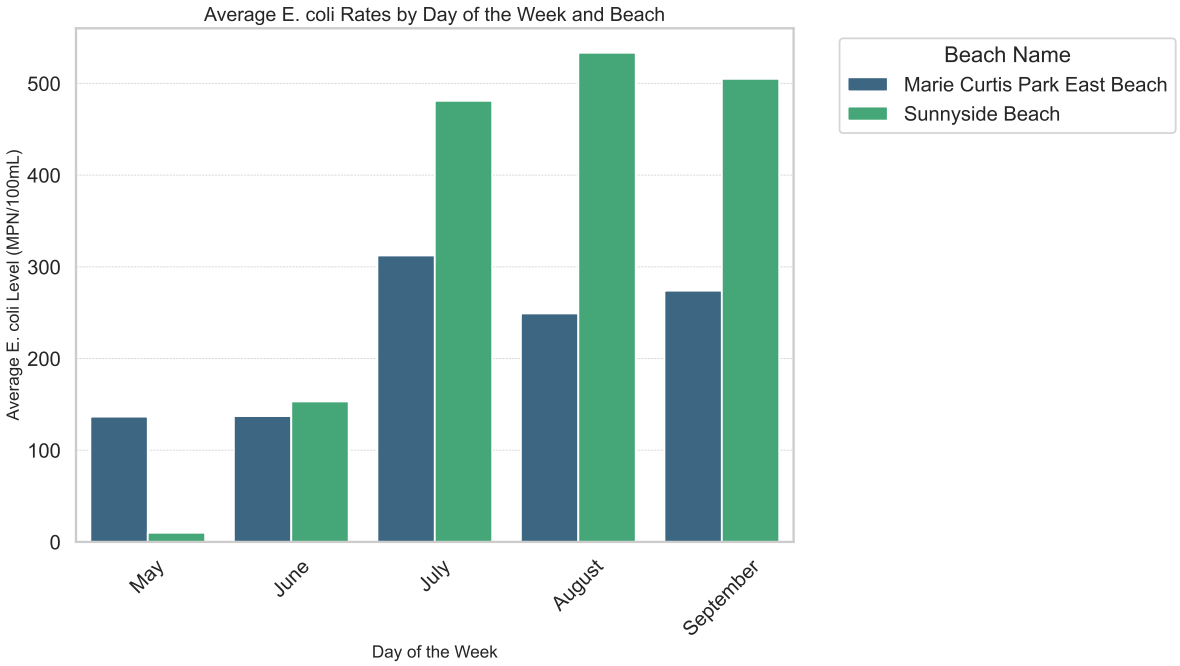

Figure 3

Weekly Data

	day	n	average_ecoli	variance	unsafe_levels	percentage_unsafe
0	Monday	456	426.0	793396.0	118	25.88
1	Tuesday	443	592.0	1457117.0	135	30.47
2	Wednesday	494	518.0	1463763.0	136	27.53
3	Thursday	488	358.0	644403.0	110	22.54
4	Friday	467	414.0	590344.0	138	29.55
5	Saturday	374	131.0	218206.0	31	8.29
6	Sunday	362	196.0	639871.0	26	7.18

Figure 5: Weekly summary statistics show the highest E. coli levels and unsafe probabilities on weekdays, particularly Tuesday and Wednesday, with much lower levels observed on weekends.

Figure 5 summarises E. coli data by day of the week, including the total number of samples (n), average E. coli levels (average_ecoli), variance, the number of samples with unsafe levels (unsafe_levels), and the percentage of samples deemed unsafe (percentage_unsafe). The weekdays, Monday through Friday, see the highest average amounts of e-coli, as well as the higher probability that the water is deemed unsafe. Tuesday and Wednesday see average levels

over 5x the 100 MPN/100ml unsafe , with relatively low variance highlighting these days as the worst for water quality over the course of a week. This point is further supported with both days having around 30% of their levels above the threshold. The weekend encounters the lowest levels by comparison, 134.91 MPN/100mL and 196.34 MPN/100mL on saturday and sunday respectively. with under 10% of tests having an unsafe result. This Is a positive sight as these days, it is assumed, encounter the most activity due to the 5 day work week. Figure 6 gives context to how the two beaches differ in their levels over the course of a week, in recent years after 2017. Marie Curtis see's consistently lower averages over the course of the weekdays, bu boh locations shar equally high levels of e-coli on friday. Sunnyside boasts levels over 300MPN/100ml during this period also. Highlighting it as having lower water quality consistently. Interestingly both locations share very low averages on both weekend days, especially when compared to the other days. Comparing this to Figure 5 highlights the improvement the water quality has seen in recent years as both averages now fall very near the observed minimum of 10MPN/100ml.

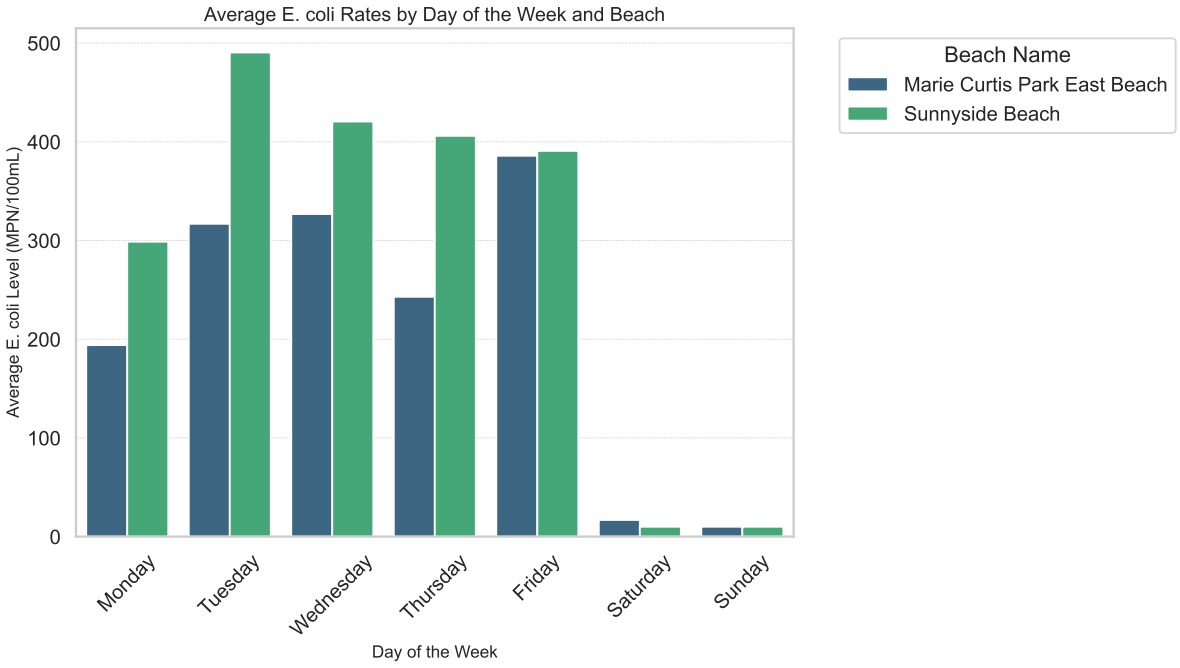

Figure 4

Figure 6: E. coli levels by day of the week from data after 2017, shows significant differences between the two beaches, with Sunnyside exhibiting extraordinarily high levels on Monday and Tuesday compared to Marie Curtis. but both see very low averages on the weekend.

Conclusion

This analysis of E. coli contamination levels at Sunnyside Beach and Marie Curtis Park East Beach reveals critical insights into the patterns and trends of water quality in Toronto's public swimming areas. While the data shows a reduction in the variability of E. coli levels in recent years (assuming this change in variability is not due to some bias), the frequency of unsafe readings remains concerning, particularly during peak summer months. The persistent risk highlighted by figure 2 indicates that ongoing contamination remains a public health challenge, despite efforts to improve water quality.Average tests that are above the target level is still over 30% in recent years. Policy makers should improve their efforts to reduce the number of days that water quality is unacceptable over the summer months. The latter months are showing the greatest risk as shown in the monthly results section, thus efforts should be made to try and carry the lower levels shown earlier throughout the summer. From the perspective of swimmers the results displayed here give insights on how to lower the risks of visiting the beaches when water quality is lower. Firstly overall we see that in recent years the extremes are much lower which is good news for lake swimmers in toronto. We also see that choosing

to swim in the earlier months of the summer appears to bring with them better overall water quality than later months, May is especially low risk. Similarly the weekends bear a far lower risk to swimmers, which is fortunate as it is likely that these are the beaches busier times. The weekly data does have warning signs also, avoiding swimming during the week appears to be the consensus, especially at Sunnyside. The weekends, which are assumed to be busier for Toronto's beaches actually exhibit low risk, especially over recent years, this is a very positive sign for public health. If this is a result of policy then the data shows that it has been very successful in raising water quality on those days. Overall the results here are fairly concerning when it comes to the percentage of results well above and acceptable threshold. There are silver linings however, and the insights gathered here can aid swimmers in choosing when and where they can minimise health risks from low water quality. The positive data points in the results show where efforts to improve water quality have been successful and where these efforts should be duplicated.

Alliance, Clean Lakes. n.d. "E. Coli and Beach Pathogens." https://www.cleanlakesalliance.org/e-coli/.

Harris, Charles R., K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, et al. 2020. "Array Programming with NumPy." *Nature* 585: 357–62. https://doi.org/10.1038/s41586-020-2649-2.

Hunter, J. D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. https://doi.org/10.1109/MCSE.2007.55.

Ontario Ministry of the Environment, Conservation and Parks. 1994. *Water Management: Policies, Guidelines, Provincial Water Quality Objectives*. Queens Printer for Ontario. https://www.ontario.ca/page/water-management-policies-guidelines-provincial-water-quality-objectives.

Python Software Foundation. 2024. *Python*. https://www.python.org/.

Waskom, Michael. 2021. *Seaborn: Statistical Data Visualization*. https://seaborn.pydata.org/.