**Data used**

* **Check-ins × business metadata**: After filtering to last 5 years, keeping only the businesses present in Daniel’s metadata and joining on business\_id, the working table (**joined**) has **2,100,624 rows** (each = one check-in) covering **39,409 unique businesses**.
* Daniel’s metadata lists **52,268** businesses, so **12,859** businesses have **0 check-ins** in this window.
* Time range of check-ins: **2018-01-01** to **2022-01-19** (UTC).

**What a “check-in” means**

* A **user-initiated action** in the Yelp app (“I’m here”)
* **Reviews/ratings are independent** of check-ins; a business can have stars/reviews but zero check-ins.

**Distribution of check-ins per business**

* **Unique businesses:** 39,409
* **Mean:** ~53.30 check-ins/business; **Median:** 17; **Max:** 9,284
* Percentiles (check-ins/business): **90%:** 129, **95%:** 211, **99%:** 528, **99.9%:** ~1,537
* Interpretation: **Heavy-tailed**—most businesses have modest counts; a small head is very active.

**Concentration (who dominates?)**

* **Top 1%** of businesses (~394 IDs) account for **18.16%** of all check-ins.
* **Top 5%** → **41.59%**; **Top 10%** → **56.97%**; **Top 20%** → **74.34%**; **Top 50%** → **94.11%**.
* Takeaway: A relatively small set of businesses contributes a **large majority** of check-ins.

A graph of a number of bars

AI-generated content may be incorrect.A graph of a bar graph

AI-generated content may be incorrect.

* The distribution is **highly skewed** (heavy-tailed).
* Most businesses are **low-engagement** (very few check-ins).

A screenshot of a graph

AI-generated content may be incorrect.

 A small fraction of **high-engagement businesses** dominate overall activity.

* This pattern is consistent with **Pareto’s 80/20 rule** (it’s even sharper: top 10% ≈ 57% of all check-ins).

A graph of a bar chart

AI-generated content may be incorrect.

Check-ins by year:

A graph of blue rectangular bars with numbers

AI-generated content may be incorrect.

 The distribution is **heavily front-loaded** (2018–2019 boom, followed by drop).

 More than **70% of all check-ins occurred before 2020**.

Check-ins by month:

A graph of a number of blue bars

AI-generated content may be incorrect.

* Stable, high check-in volumes in 2018–2019 (50–70k per month).
* **March–April 2020:** activity fell close to zero → aligned with global COVID lockdowns.
* 2021 showed partial recovery (~15–30k/month) but never returned to pre-COVID levels.

Check-ins by Day of Week:  
A graph of blue bars

AI-generated content may be incorrect.

* **Weekdays (Mon–Thu):** check-ins are fairly stable, each around **210k–240k**.
* **Friday:** check-ins jump to **~310k**.
* **Saturday:** the clear peak, with nearly **480k check-ins** — the busiest day of the week.
* **Sunday:** still very high at **~430k**, higher than any weekday.

**Check-ins by Hour of Day:**

A graph of a number of hours

AI-generated content may be incorrect.

 **Late-night spike (00:00–02:00):**

* Highest activity around midnight and early morning hours (200k+ check-ins).
* Suggests many check-ins happen **after evening social activities** (bars, clubs, late dining).

 **Morning lull (05:00–11:00):**

* Very few check-ins during early morning and late morning.
* Reflects downtime when most businesses are closed.

 **Afternoon build-up (12:00–16:00):**

* Gradual increase starts after lunch, peaking toward late afternoon.

 **Evening peak (17:00–23:00):**

* Strong second wave of check-ins during evening.
* Absolute peak around **23:00** — consistent with nightlife, late dining, entertainment.

A graph of a number of months

AI-generated content may be incorrect.