Associate data scientist technical exercise

Christopher Leonard

Outline

1) The data set

2) Exploratory data analysis

3) Future directions

The data set

- Inside Airbnb: activist project providing data and analysis about Airbnb

- Publicly available data compiled from airbnb.com, cleaned, and processed

- Airbnb listings in New York City (obtained 6 Aug 2019)

 – Host

  – Location

  – Date of last review

Data quality

- Pros: clean and consistent

  - Hosts have unique host_id

  - Neighbourhoods match with city definitions

- Cons: limited scope

  - No detailed rental history

How are hosts using Airbnb?

- Professionally or casually?

- How often is property available?

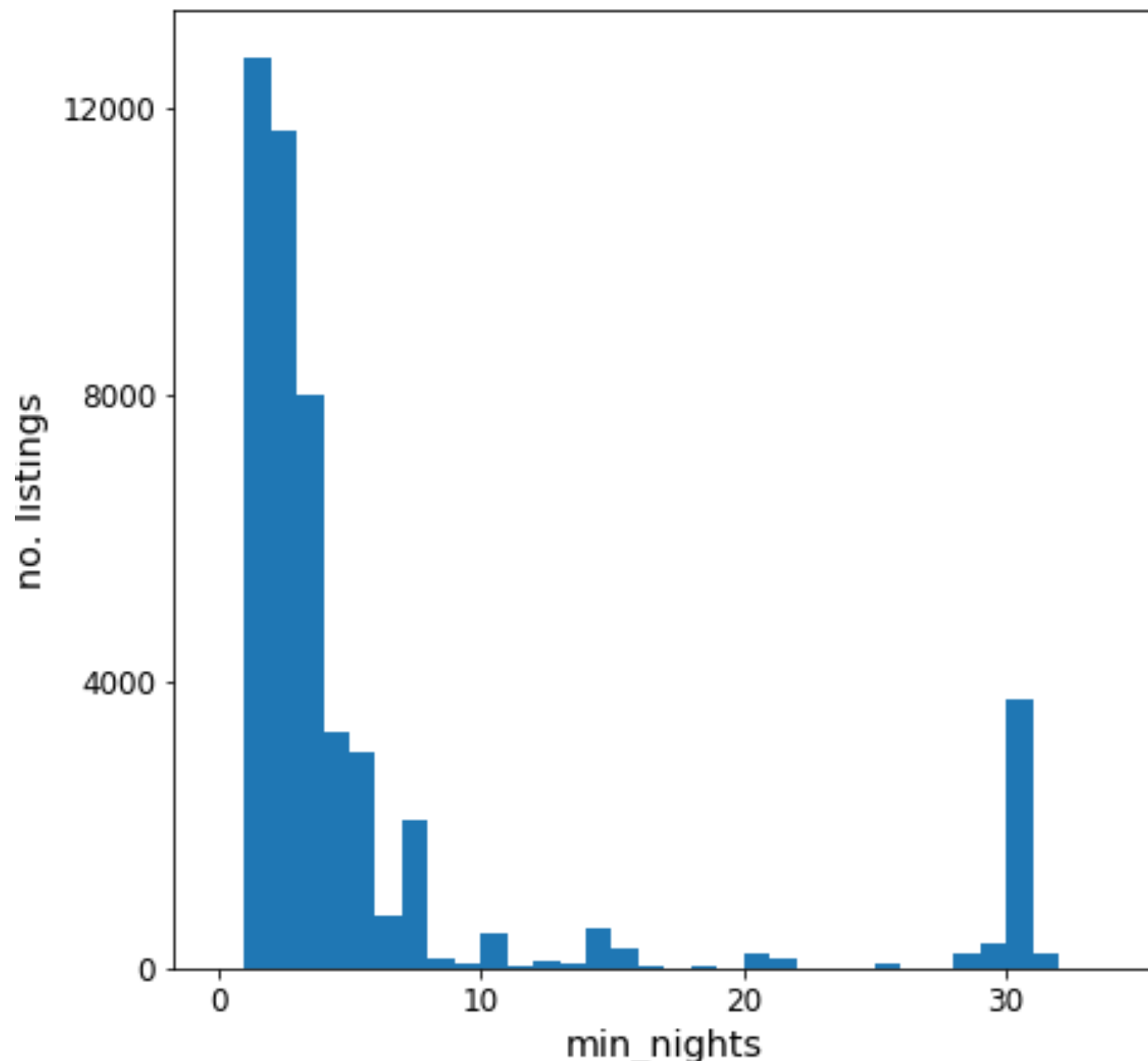- Offering long- or short-term rentals?

availability_365: days listing is available over next year

- 36% of listings are currently unavailable

- 4% of listings available any day

- Uniform away from extremes
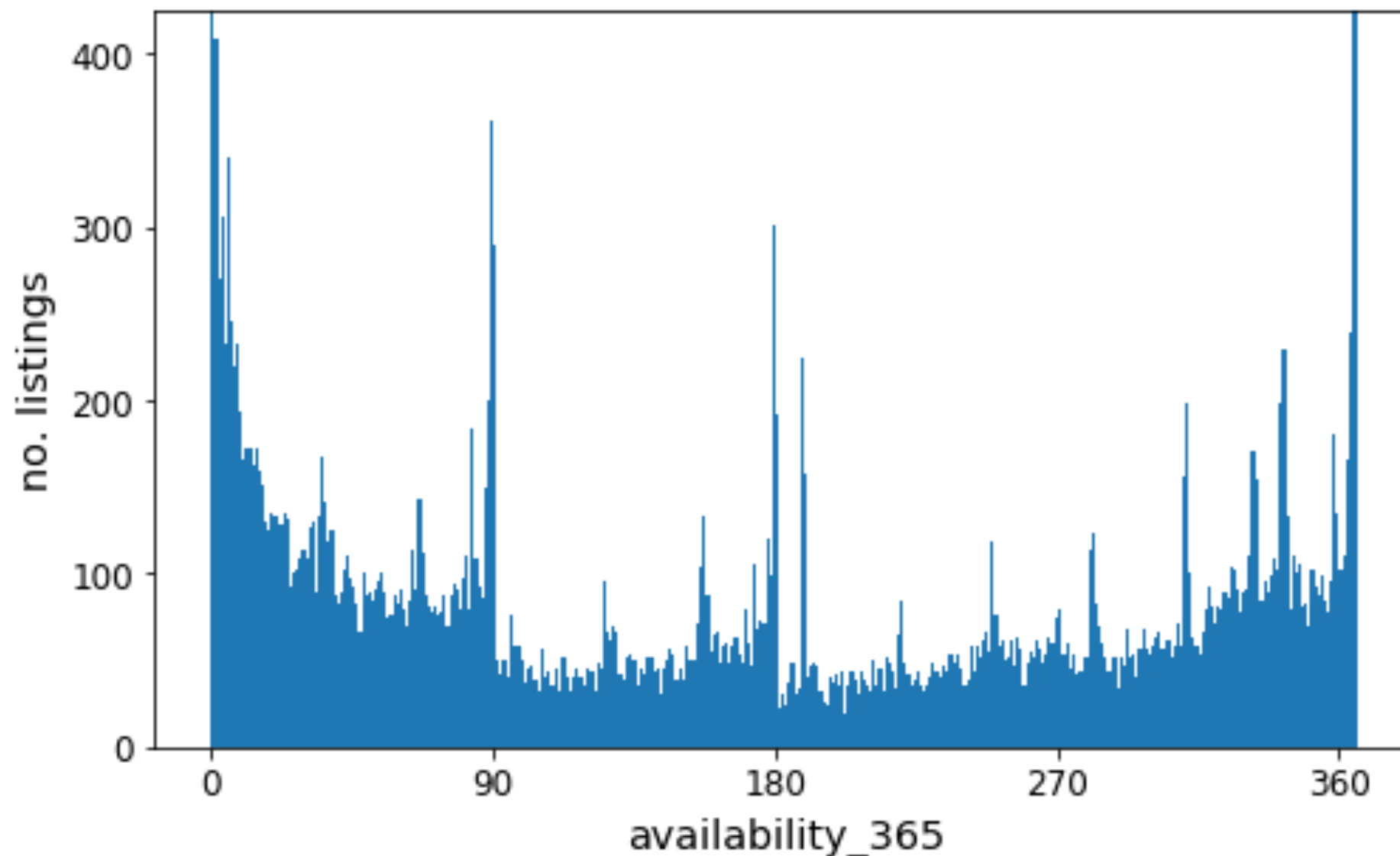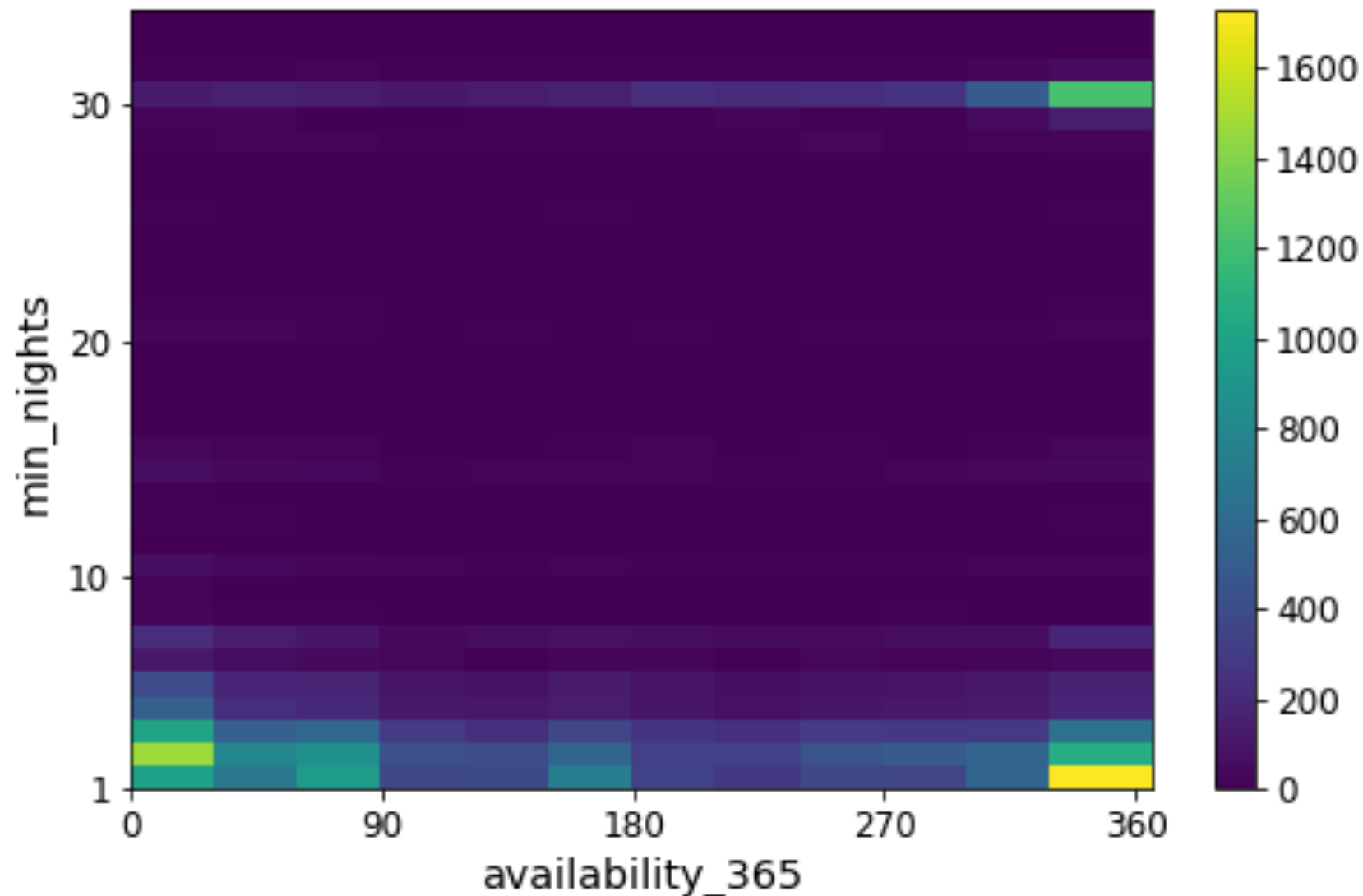
min_nights: minimum nights required to book
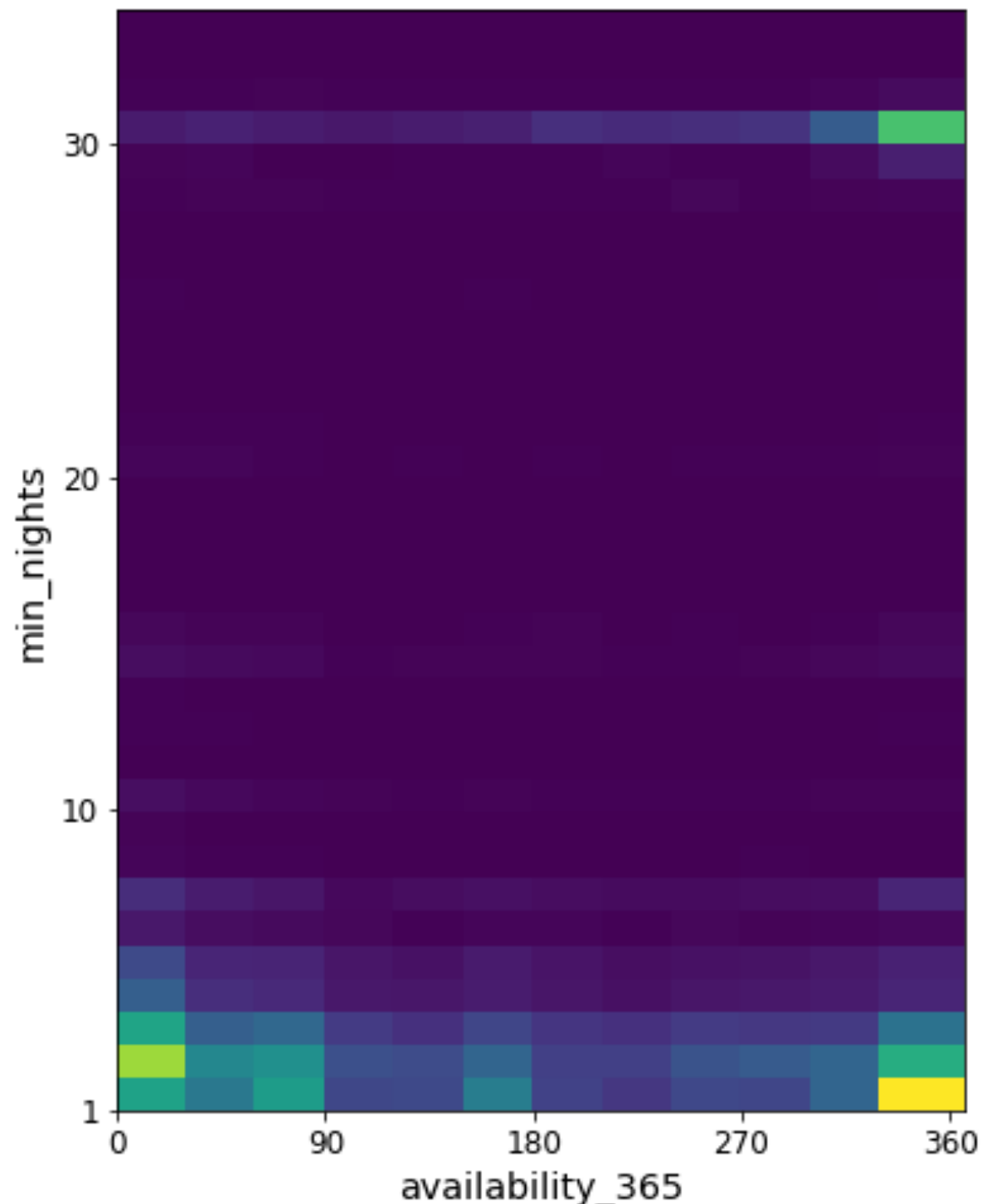
- 65% of listings allow short-term stays (< 4 nights)

min_nights	% of listings
1	26%
2	24%
3	16%
4 - 6	15%
7 - 29	10%
30+	9%

min_nights against availability_365

- Long/short-term against how often it's available

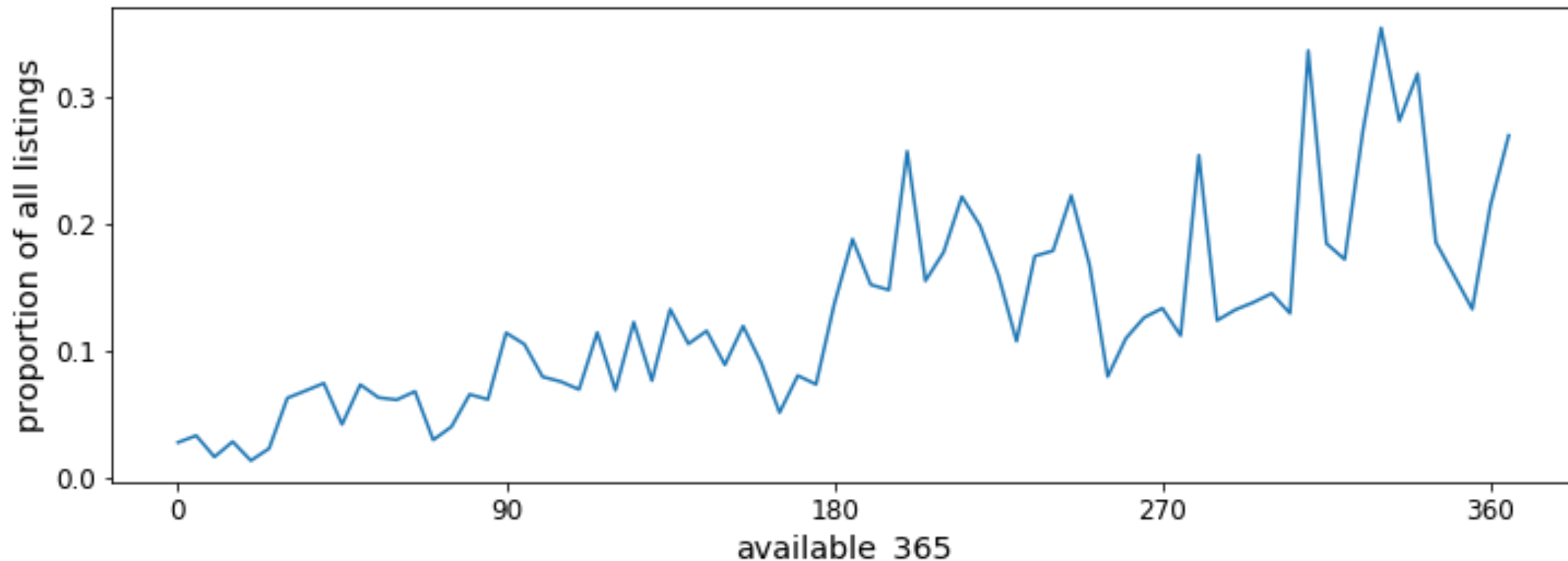- Colour intensity indicates number of listings

3 Hotspots - common situations

- Short-term, low availability

  - Traditional Airbnb

- Short-term, high availability

  - Traditional hotel

- Long-term, high availability

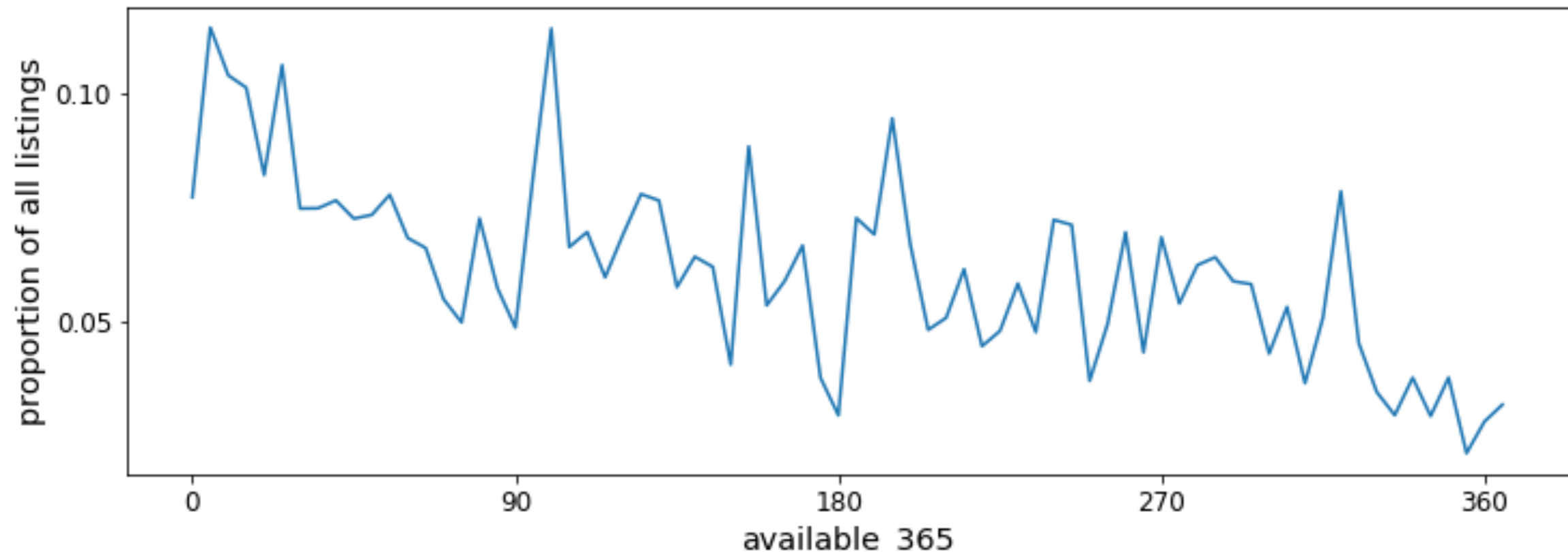 - Tenancy

Long-term listings (min_nights ≥ 30)

- Long-term listings are markedly different from others

- They increase as a proportion of all listings as available_365 increases

  - Unique to long-term listings

- Long-term rentals are treated differently under NY law

Further directions

- Examine short-term listings as a proportion of all listings (e.g. min_nights = 4)

- How does slope of line change with min_nights?

  – Perform linear regression analysis

  – Is dependence on min_nights statistically significant?

Questions?

Thank you for listening