Using daily Airbnb web scraped data to provide spatial and temporal understanding of short-term lets activity

UBDC Webinar 24th June 2021
Yang Wang and Mark Livingston
Outline of the Webinar

- Introduction to Airbnb
- Impacts of Airbnb
- Issues of measuring impact from Airbnb
- Increasing demand for fine-grained data and method
- Booking calendars and calendar activities
- Indicators including occupancy, revenue, reservations and cancellations
- Case study of cancellations during the pandemic in Edinburgh
Airbnb and sharing economy

- Emergence of the sharing economy
- Airbnb, originally a web site facilitated sharing your house, part of your house or a second home
- Allowed people to easily let space and properties
- Tapped into unfulfilled demand, especially in existing tourist cities
Airbnb sharing economy or not

- At the end of 2020 Airbnb valued at around $75 billion

- Airbnb no longer about the sharing economy

- In 2019 large majority of listings were for an entire property/apartment (75%) (Adamiak 2019)

- 59% of listings were managed by a host with multiple properties
Impacts of Airbnb (+ve & -ve)

- Increased supply of travel accommodation
  - Economic benefits of increased visitors to local economy
  - Greater choice for consumers
  - Reduced costs of accommodation

- Economic impacts on existing travel accommodation

- Impact on housing costs Rents/Sales

- Neighbourhood impacts of concentration of Airbnb lets
  - Pressure on resident population (tourist gentrification)
  - Change in neighbourhood services
Measuring the impact of Airbnb on neighbourhoods and housing

- To measure the impact of Airbnb we need data.
- Airbnb hold the only true and accurate record of activity (transaction data).

Others scrape the Airbnb web site for data:
- Inside Airbnb
- AirDNA

www.ubdc.ac.uk
Spatial analysis

• Concentration of Listings
  • Entire home rental
  • Active listings
  • New listings

• Comparing with e.g.
  • Housing stock
  • % of Rental/Residential Housing
  • Hotels, Services and other Amenities

Research using Reviews

- The number of reviews left by visitors as an estimate of Airbnb activity
- 50% -90% of visitors leave reviews (InsideAirbnb)
- Bias on the type of guest who leaves a review and those that don’t.

https://www.ubdc.ac.uk/news-media/2020/june/reviewing-airbnb/
Vulnerable Neighbourhoods

- Rent gap theory
- Limited by data quality affect estimation of Revenue
- Assumptions have to be made on
  - Occupancy,
  - Visits,
  - Length of visits
Demand for frequent and regular scraping

- More fine-grained Airbnb data will play an increasingly important role in capturing Airbnb market shifts,
  - indicating good timing for regulation,
  - and compare with LTR market indicators to better inform policy during post-epidemic recovery.

- An improved market balance can be achieved with improved open data evidence.
Data Analyses – Temporal Challenge

- Research often focus on **Spatial Analyses**
  - Spatial distribution of listings

- We contribute to address the **Temporal Challenge** by
  - clean, extract and encapsulate detailed changes on the listing availabilities
  - estimate the potential bookings/cancellations from the block/open of days in the listing calendars,
  - appraise the turnovers taking into account both nightly rental and additional fees.

https://www.talkwalker.com/blog/best-data-analytics-tools-on-the-market
Scraping our own data

- Scraping exercise
- Benefits:
  - Control the details, such as spatial and temporal intervals of scraping
  - Ensure data quality
  - Transparent data and method
- Challenges:
  - Not ‘preferred’ by Airbnb
  - Difficult to manage technically
    - Scales of data
    - API access limits
  - Cannot share data

3rd Party Data Education?

Airbnb Web Scraping Data Structure

- The original Airbnb web scrapes
  - API responses in JSON format
  - unstructured and
  - hard for effective and efficient information retrieval

- We extract/organize them into structuralized interconnected parts,
  - booking activities comprised of the daily calendar and booking policies,
  - customer reviews,
  - spatial location,
  - host status,
  - and structural and service amenities.
Booking Calendars

- 365-days availability planned in a booking calendar
- Hosts can take updates to them as frequently as required
  - InsideAirbnb provides these calendars monthly
  - AirDNA provides listing performances built from daily calendars
- The larger the gap in calendars, the less accurate we know about the bookings
- The finer scale of the calendars, the more complex to process and manage.

https://www.hostyapp.com/airbnb-multi-calendar/

www.ubdc.ac.uk
What does booking calendars tell us?

**Occupancy and Revenue**
- How many days a listing has been booked?
- How many days it has not been available?
- How many guests have been accommodated?
- How long did guests stay?

**Reservations and Cancellations**
- How many reservations have been made on a particular day?
- How many bookings have been cancelled?
- How many visits that potentially bring income but eventually failed to secure the deal?
Let’s see an example

– Bookings/Occupancy

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Calendar</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
<td>available</td>
</tr>
<tr>
<td>Activities</td>
<td>No bookings</td>
<td>Booked 11th-13th</td>
<td>Booked 15th-17th</td>
<td>Canceled 15th-17th</td>
<td>reach 11th visit</td>
<td>reach 12th visit</td>
<td>reach 13th visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visits</td>
<td>1 visit for 3 days</td>
<td>1 visit for 2 days</td>
<td>1 visit for 2 days</td>
<td>1 visit for 2 days</td>
<td>1 visit start</td>
<td>1 visit last</td>
<td>1 visit last</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Let’s see an example
– Reservation/Cancelation

<table>
<thead>
<tr>
<th>observed at</th>
<th>13-Mar-21</th>
<th>observed at</th>
<th>14-Mar-21</th>
<th>observed at</th>
<th>15-Mar-21</th>
<th>observed at</th>
<th>16-Mar-21</th>
<th>observed at</th>
<th>17-Mar-21</th>
<th>observed at</th>
<th>18-Mar-21</th>
<th>observed at</th>
<th>19-Mar-21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calendar</td>
<td>01-May-21 available</td>
<td>01-May-21 available</td>
<td>01-May-21 available</td>
<td>01-May-21 available</td>
<td>01-May-21 available</td>
<td>01-May-21 unavailable</td>
<td>01-May-21 unavailable</td>
<td>01-May-21 unavailable</td>
<td>01-May-21 unavailable</td>
<td>01-May-21 unavailable</td>
<td>01-May-21 unavailable</td>
<td>01-May-21 unavailable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>02-May-21 available</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td>02-May-21 unavailable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>03-May-21 available</td>
<td>03-May-21 available</td>
<td>03-May-21 available</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td>03-May-21 unavailable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>04-May-21 available</td>
<td>04-May-21 available</td>
<td>04-May-21 available</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td>04-May-21 unavailable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td>05-May-21 available</td>
<td></td>
</tr>
<tr>
<td></td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td>06-May-21 available</td>
<td></td>
</tr>
<tr>
<td></td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td>07-May-21 available</td>
<td></td>
</tr>
<tr>
<td></td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td>08-May-21 available</td>
<td></td>
</tr>
<tr>
<td>Activities</td>
<td>No bookings</td>
<td>Booked 2nd</td>
<td>Booked 7th-8th</td>
<td>Canceled 7th-8th</td>
<td>Booked 1st</td>
<td>Booked 3rd-4th</td>
<td>Booked 7th</td>
<td>No activities</td>
<td>Booked 5th</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Future Visits</td>
<td>No visits</td>
<td>1 visit for 1 day</td>
<td>1 visit for 2 days</td>
<td>canceled 1 visit</td>
<td>1 visit for 1 day</td>
<td>1 visit for 2 days</td>
<td>1 visit for 1 day</td>
<td>No new visits</td>
<td>1 visit for 1 day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Watch for Ambiguities

- ‘Unavailable’ can mean:
  - Booked – by guest
  - Blocked – by host, not available to book

- Airbnb does not distinguish them!
Calendar Activities - Occupancy

- **Occupancy:**
  - If the final status is ‘unavailable’
  - It is changed at least once from ‘available’ previously

- **An overestimation**
  - as the same calendar updates can suggest the host block the day for any bookings.

- **Without Airbnb’s further clarifications on the meanings of ‘unavailable’, we cannot rule out these false positives.**
Calendar Activities - Cancellation

- Cancellation
  - If the final status is ‘available’
  - It has been marked ‘unavailable’ before
- Also an overestimation
  - The host makes the previously blocked date open for booking but failed to secure a visit at the end.
- This is also caused by the ambiguity of the ‘unavailable’ status.
Other Indicators

- Revenue
  - Occupancy X nightly rentals
- Visits
  - Days booked together
- Length of visits
  - Number of days a visit lasts
Let’s revisit before a case study

- **Occupancy**
  - Deriving these estimations of occupancy requires continuous monitoring of the listings
  - Beneficial to estimate
    - Revenue; Visits, length of visits
    - Customer preferences and market turnovers
  - Requires previous tracking records,

- **Market Activities**
  - Track the potential visits or agreed cancellations planned in the future.
  - Beneficial to anticipate
    - Market reactions to sudden interventions
    - Hosts’ behaviours in response to events

- **Watch out for the ambiguity!**

https://community.withairbnb.com/t5/Dicas-e-Tutoriais/Airbnb-Blog-Dicas-para-hospedagens-de-longa-dura%C3%A7%C3%A3o/td-p/369454
Let’s break here!

Any brief questions?
Edinburgh Market

- **Airbnb**
  - Accommodating around 140,000 visitors/year
  - Over 31% of the Airbnb daily £2M economic contributions to the Scottish economy

- **Distributions**
  - over 10,000 active Airbnb in Edinburgh
  - over 79% of listings (around 16% of the total dwellings in Edinburgh) are rented as entire homes

- **Regulations**
  - regulation consultation in 2019
    - later suggested a licensing regime
    - and prepared to give the power to local authorities to tackle the issue locally
  - U-turn happens in February 2021
    - local authorities demand extra support
    - COVID-19 Recovery
Edinburgh market during the Pandemic

Extenuating Circumstances Policy (ECPs)


Whether our data and method

- capture the fine details of the volume and scales of the cancelation triggered by each update of the ECPs?
- offer detailed and reliable data evidence to closely understand the local trends?

www.ubdc.ac.uk
Data

- Accumulated 10,489 listings since January 2020.
- 6564 entire home rentals
- Do not require definitions on ‘active listings’
- Focus on all the listings rented entirely on the market
ECPs during lockdown

- The key dates:
  - 16th March announced for coverage between 14th March and 14th April (Policy 1),
  - 30th March extended the coverage to 31th May (Policy 2),
  - 1th May 2020 further added reservations between 1th – 15th June (Policy 3),
  - then 1th June make it through 15th July (Policy 4),
  - and 15th June add days until 31st July (Policy 5).

Reinforced by blocking calendars

9th April - 15th July

UK lockdown

- 1st lockdown on 23rd March 2020
- Until 15th July, non-essential shops were allowed to open again
- By the end of July, no more death reported in Scotland.
This first wave of COVID-19 has passed.
Why this period is special?

- The market reactions to the ECPs fall into the top-right booking status change from unavailable to available.
- Locking the booking calendar makes the
  - Minimized the chances of switching blocked calendars to ‘available’ for new reservations
  - have a close-to-reality estimation of cancelations.
- Majority of other calendar availabilities either
  - (a) remains the same again as no new bookings are allowed, or
  - (b) altered to not open for bookings, shown in the other three types.
Retrieving Cancellations on ECPs
Tracking Daily Cancellations in Edinburgh - Volume
Tracking Daily Cancellations in Edinburgh – Turnover

![Graph showing policy coverage and cancelled turnover with values £0.8M and £0.6M]
Tracking Daily Cancellations in Edinburgh – Visits
Tracking Daily Cancellations in Edinburgh – Length of Cancelled Visits
Summary and Conclusion

In this Webinar we have tried to:
 Highlight some of the issues with increasing Airbnb growth

 Demonstrate the complexity of Airbnb data

 Show the value of more frequent (daily) scraping

 Demonstrate the development of new metrics

 Show how we can more accurately measure occupancy and activity with these metrics
References

- Analysis of Short-Term Lets Data for Edinburgh, 2017, Alasdair Rae, University of Sheffield.
- https://www.ubdc.ac.uk/news-media/2020/june/reviewing-airbnb/
Any Questions?