

# Deep and Statistical Learning WS 21-22

## Analysis of Berlins Airbnb Data using statistical and deep learning methods.

### Data

- Visual representation
- Get the data

Data last downloaded 10.11.21

### Possible analysis:

#### Fundamentals:

Taken from the about page of Insider Airbnb

By analyzing publicly available information about a city's Airbnb's listings, Inside Airbnb provides filters and key metrics so you can see how Airbnb is being used to compete with the residential housing market.

With Inside Airbnb, you can ask fundamental questions about Airbnb in any neighbourhood, or across the city as a whole. Questions such as:

- “How many listings are in my neighbourhood and where are they?”
- “How many houses and apartments are being rented out frequently to tourists and not to long-term residents?”
- “How much are hosts making from renting to tourists (compare that to long-term rentals)?”
- “Which hosts are running a business with multiple listings and where they?”

The tools are presented simply, and can also be used to answer more complicated questions, such as:

- “Show me all the highly available listings in Bedford-Stuyvesant in Brooklyn, New York City, which are for the ‘entire home or apartment’ that have a review in the last 6 months AND booked frequently AND where the host has other listings.”

These questions (and the answers) get to the core of the debate for many cities around the world, with Airbnb claiming that their hosts only occasionally rent the homes in which they live.

In addition, many city or state legislation or ordinances that address residential housing, short term or vacation rentals, and zoning usually make reference to allowed use, including:

- how many nights a dwelling is rented per year
- minimum nights stay
- whether the host is present
- how many rooms are being rented in a building
- the number of occupants allowed in a rental whether the listing is licensed

## Paper Inspiration

### Notizen für die QnA vom 15.11

- Ein wichtiger Teil ist Daten einlesen
- Minimum 3000 Beobachtungen verwenden
- Histogramme plotten und Korrelationen berechnen als Einstieg für Paper
- In der QnA wirkte es als wäre es gut wenn wir noch mehr Bilder aus dem Internet scrapen
- Bilder und comments analysieren wären gut
- Alle daten typen zusammen zu bringen ist die Stärke von neuronalen Netzen also sollten wir das machen
- Interpretierbarkeit der Ergebnisse soll möglich sein. Also keine reine Blackbox. Stichwort statistisches lernen
- Bis zur nächsten Session sollen wir mindestens die Daten runterladen und am besten die Bilder scrapen.
- Bilder sollen in reduzierter Pixelanzahl (64x64 oder so) runtergeladen werden
- Es gibt eine letzte Stadt mit der wir dann unsere Modelle gegeneinander antreten lassen
- Modelle sollen so geschrieben werden dass es sie auf die neue Stadt angepasst werden kann

### Anmeldedaten für DeepL Videos bei Manning (DeepLearning with R in motion):

- tobirinnert@gmail.com
- DeepLearningWS2021

### Dataset of labeled handwritten digits

- <http://yann.lecun.com/exdb/mnist/>
- <https://rpubs.com/Argaadya/nn-mnist> (Code)
- <https://www.kaggle.com/c/digit-recognizer/overview/description>
- <https://www.kaggle.com/russwill/build-your-own-neural-network-in-r>

### Important Dates

- QnA session: 13/12 17/01
- Group presentation 20/02 – 21/02, 09.00–12.00
- Term paper of 15 pages and Challenge-submittal:
- hand in by 06/03/2022, 23.59
- Revised term paper: hand in by 31/03/2022
- Publication of collective volume provisionally by end of 2022

Possible research questions: - Will a certain host list a new flat in the next weeks/month/years? - Which kind of pictures (given fixed categories for the pictures) should a host upload to be trust worthy / be able to demand a high price / increase the amount of requests