DATABASE 2
RDF Group Project

Team ASM

Alessandro Benetti Andrea Ceccato
Mohammad Sohail Sepide Bahrami

OUTLINE

Dataset

Graph

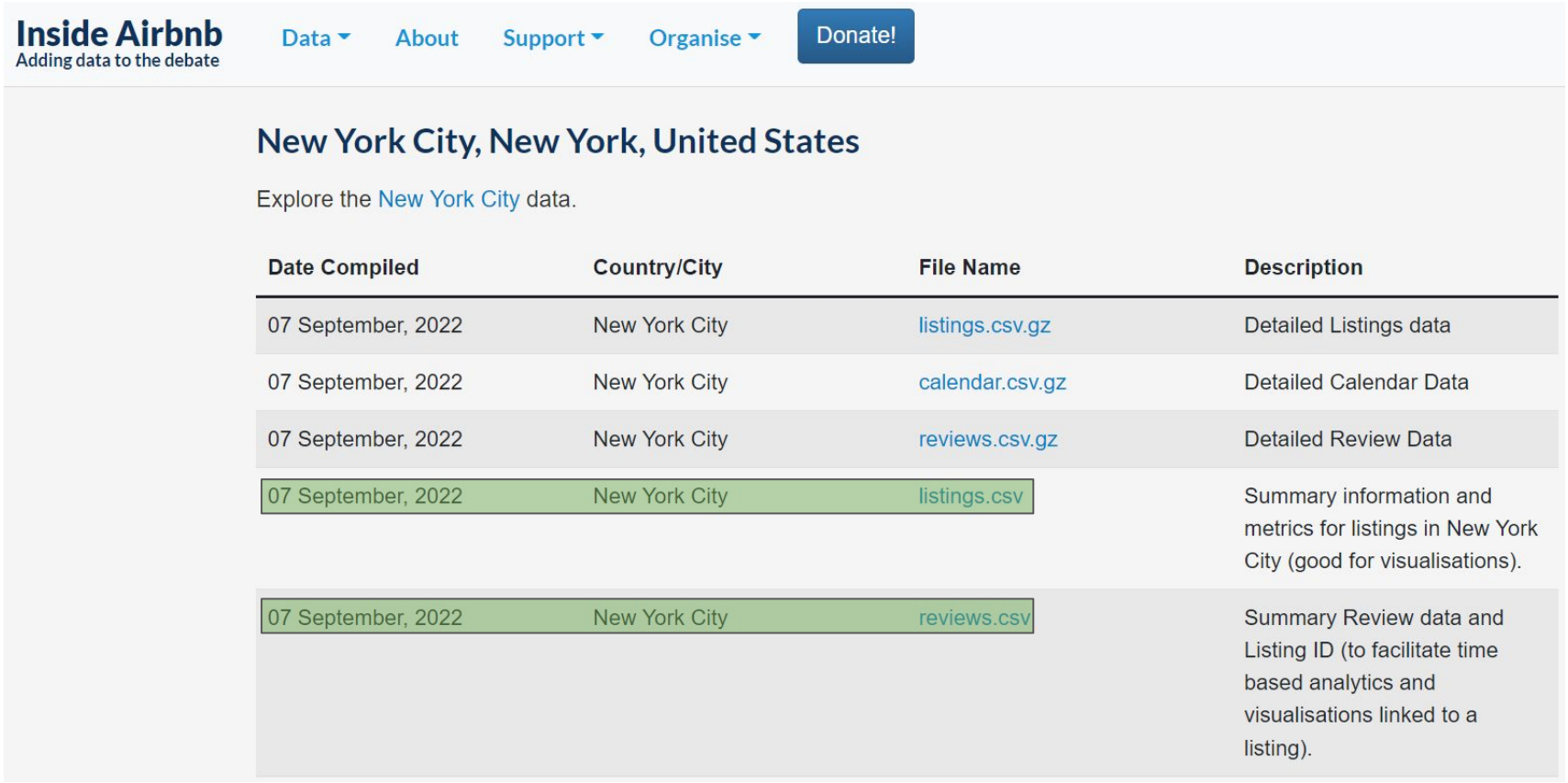
Ontology

Notebook

RDF Data

Queries

Dataset



Inside Airbnb
Adding data to the debate

Data ▾ About Support ▾ Organise ▾ Donate!

New York City, New York, United States

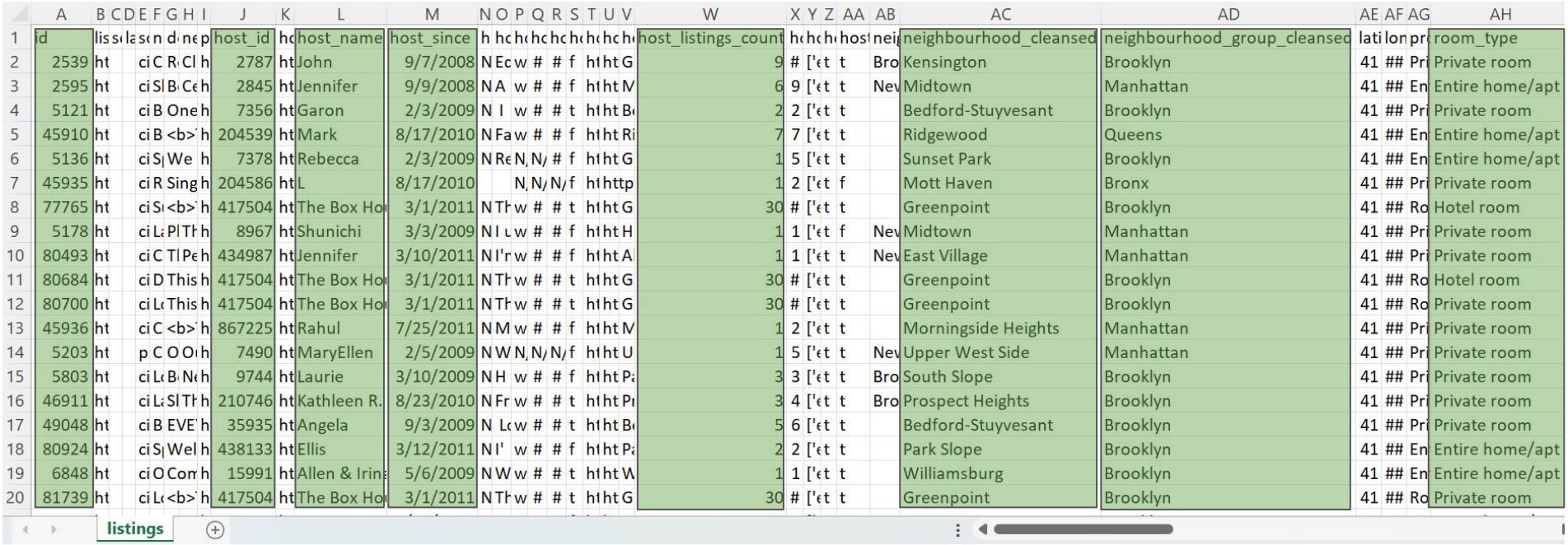
Explore the New York City data.

Date Compiled	Country/City	File Name	Description
07 September, 2022	New York City	listings.csv.gz	Detailed Listings data
07 September, 2022	New York City	calendar.csv.gz	Detailed Calendar Data
07 September, 2022	New York City	reviews.csv.gz	Detailed Review Data
07 September, 2022	New York City	listings.csv	Summary information and metrics for listings in New York City (good for visualisations).
07 September, 2022	New York City	reviews.csv	Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing).

http://insideairbnb.com/get-the-data/

Dataset

listings.csv



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	id	lis	s	la	s	n	d	ne	p	host_id	h	host_name	host_since	h	h	h	h	h	h	h	h	h	host_listings_count	h	h	h	host	nei	neighbourhood_cleansed	neighbourhood_group_cleansed	lati	lon	pr	room_type
2	2539	ht			ci	C	R	Cl	h	2787	ht	John	9/7/2008	N	Ec	w	#	#	f	ht	ht	G	9	#	['et	t	t	Bro	Kensington	Brooklyn	41	##	Pri	Private room
3	2595	ht			ci	Sl	B	Ce	h	2845	ht	Jennifer	9/9/2008	N	A	w	#	#	f	ht	ht	M	6	9	['et	t	t	Nev	Midtown	Manhattan	41	##	En	Entire home/apt
4	5121	ht			ci	B	One	h		7356	ht	Garon	2/3/2009	N	I	w	#	#	t	ht	ht	B	2	2	['et	t	t		Bedford-Stuyvesant	Brooklyn	41	##	Pri	Private room
5	45910	ht			ci	B	<b>	h		204539	ht	Mark	8/17/2010	N	Fa	w	#	#	f	ht	ht	Ri	7	7	['et	t	t		Ridgewood	Queens	41	##	En	Entire home/apt
6	5136	ht			ci	S	We	h		7378	ht	Rebecca	2/3/2009	N	Re	N,	N/	#	f	ht	ht	G	1	5	['et	t	t		Sunset Park	Brooklyn	41	##	En	Entire home/apt
7	45935	ht			ci	R	Sing	h		204586	ht	L	8/17/2010			N,	N/	N/	f	ht	http		1	2	['et	t	f		Mott Haven	Bronx	41	##	Pri	Private room
8	77765	ht			ci	S	<b>	h		417504	ht	The Box Ho	3/1/2011	N	Th	w	#	#	t	ht	ht	G	30	#	['et	t	t		Greenpoint	Brooklyn	41	##	Ro	Hotel room
9	5178	ht			ci	L	Pl	Th	h	8967	ht	Shunichi	3/3/2009	N	I	w	#	#	f	ht	ht	H	1	1	['et	t	f	Nev	Midtown	Manhattan	41	##	Pri	Private room
10	80493	ht			ci	C	Tl	Pe	h	434987	ht	Jennifer	3/10/2011	N	I'r	w	#	#	f	ht	ht	A	1	1	['et	t	t	Nev	East Village	Manhattan	41	##	Pri	Private room
11	80684	ht			ci	D	This	h		417504	ht	The Box Ho	3/1/2011	N	Th	w	#	#	t	ht	ht	G	30	#	['et	t	t		Greenpoint	Brooklyn	41	##	Ro	Hotel room
12	80700	ht			ci	L	This	h		417504	ht	The Box Ho	3/1/2011	N	Th	w	#	#	t	ht	ht	G	30	#	['et	t	t		Greenpoint	Brooklyn	41	##	Ro	Private room
13	45936	ht			ci	C	<b>	h		867225	ht	Rahul	7/25/2011	N	M	w	#	#	f	ht	ht	M	1	2	['et	t	t		Morningside Heights	Manhattan	41	##	Pri	Private room
14	5203	ht			p	C	O	O	h	7490	ht	MaryEllen	2/5/2009	N	W	N,	N/	N/	f	ht	ht	U	1	5	['et	t	t	Nev	Upper West Side	Manhattan	41	##	Pri	Private room
15	5803	ht			ci	L	B	N	h	9744	ht	Laurie	3/10/2009	N	H	w	#	#	f	ht	ht	P	3	3	['et	t	t	Bro	South Slope	Brooklyn	41	##	Pri	Private room
16	46911	ht			ci	L	Sl	Th	h	210746	ht	Kathleen R.	8/23/2010	N	Fr	w	#	#	t	ht	ht	P	3	4	['et	t	t	Bro	Prospect Heights	Brooklyn	41	##	Pri	Private room
17	49048	ht			ci	B	EVE	h		35935	ht	Angela	9/3/2009	N	L	w	#	#	t	ht	ht	B	5	6	['et	t	t		Bedford-Stuyvesant	Brooklyn	41	##	Pri	Private room
18	80924	ht			ci	S	Wel	h		438133	ht	Ellis	3/12/2011	N	I'	w	#	#	f	ht	ht	P	2	2	['et	t	t		Park Slope	Brooklyn	41	##	En	Entire home/apt
19	6848	ht			ci	O	Com	h		15991	ht	Allen & Irin	5/6/2009	N	W	w	#	#	t	ht	ht	W	1	1	['et	t	t		Williamsburg	Brooklyn	41	##	En	Entire home/apt
20	81739	ht			ci	L	<b>	h		417504	ht	The Box Ho	3/1/2011	N	Th	w	#	#	t	ht	ht	G	30	#	['et	t	t		Greenpoint	Brooklyn	41	##	Ro	Private room

listings

Dataset

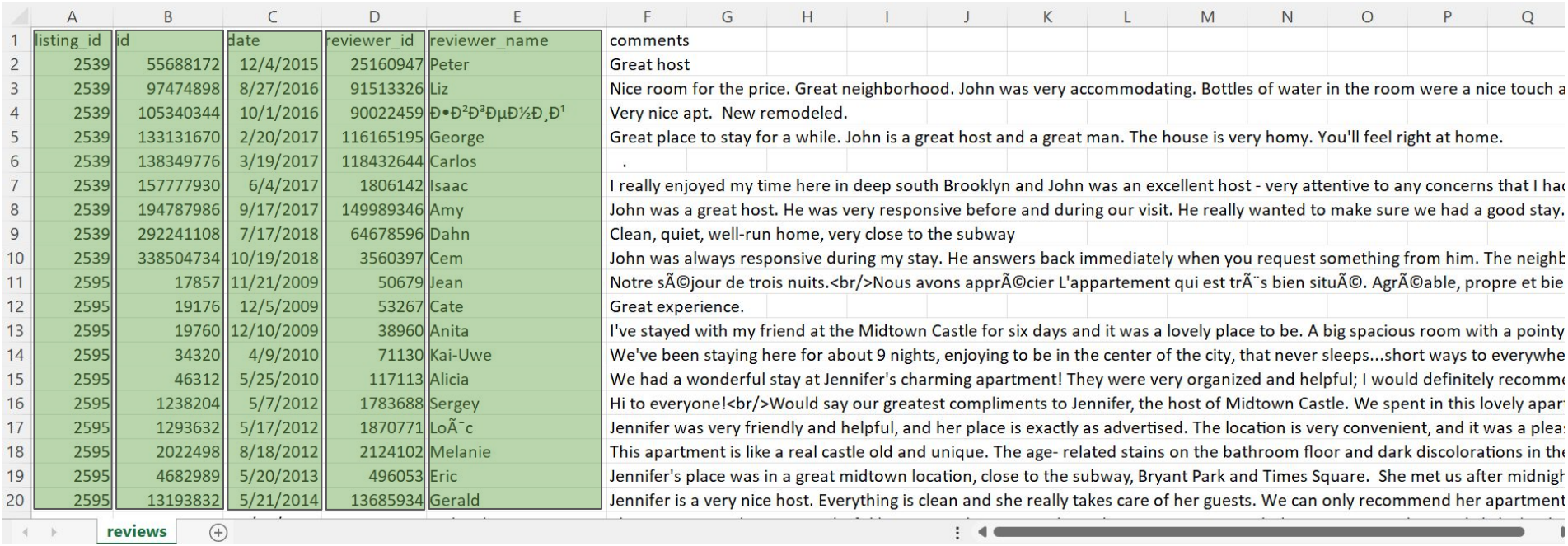
listings.csv

	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV	BW
1	price	minimum_nights	ma	mi	ma	mi	ma	mi	ma	cale	ha	ava	ava	ava	availability_365	cale	number_of_reviews	nur	nu	first	las	review_scores_rating	rev	rev	rev	rev	rev	rev	lice	ins	cal	cal	cal	cal	reviews_per_m
2	$299.00	30	##	30	30	##	##	30	##		t	21	51	81	356	##	9	0	0	##	#	4.89	5	5	5	5	5	5		f	9	1	6	2	0.11
3	$175.00	30	##	30	30	##	##	30	##		t	0	0	5	280	##	49	1	0	##	#	4.68	5	5	5	5	5	4		f	3	3	0	0	0.31
4	$60.00	30	##	30	30	##	##	30	##		t	5	30	60	335	##	50	0	0	##	#	4.52	4	4	5	5	4	5		f	2	0	2	0	0.31
5	$425.00	30	##	30	30	##	##	30	##		t	30	60	90	365	##	13	0	0	##	#	4.42	5	4	5	5	5	5		f	6	6	0	0	0.1
6	$275.00	21	##	21	21	##	##	21	##		t	0	0	0	179	##	3	1	1	##	#	5	5	5	5	5	5	5		f	1	1	0	0	0.03
7	$60.00	30	40	30	30	40	40	30	40		t	23	53	83	83	##	0	0	0											t	1	0	1	0	
8	$308.00	2	##	2	2	##	##	2	##		t	15	42	72	217	##	42	4	0	##	#	4.9	5	5	5	5	4	5		f	#	16	10	0	0.3
9	$68.00	2	14	2	2	14	14	2	14		t	2	4	7	160	##	550	56	2	##	#	4.23	4	4	5	4	5	4		f	1	0	1	0	3.39
10	$55.00	2	60	2	2	60	60	2	60		t	0	0	0	132	##	207	9	1	##	#	4.34	4	4	5	5	5	4		f	1	0	1	0	1.49
11	$529.00	2	##	2	2	##	##	2	##		t	25	55	85	100	##	10	2	0	##	#	5	5	5	5	5	5	5		f	#	16	10	0	0.08
12	$372.00	2	##	2	2	##	##	2	##		t	17	47	77	222	##	9	2	1	##	#	4.67	5	5	5	5	4	5		f	#	16	10	0	0.07
13	$75.00	31	##	31	31	##	##	31	##		t	0	0	0	219	##	135	4	0	##	#	4.65	5	4	5	5	5	5		f	1	0	1	0	0.95
14	$75.00	2	14	2	2	14	14	2	14		t	0	0	0	0	##	118	0	0	##	#	4.91	5	5	5	5	5	5		f	1	0	1	0	0.75
15	$118.00	4	21	4	4	21	21	4	21		t	0	0	24	299	##	216	22	1	##	#	4.7	5	5	5	5	5	5		f	3	1	2	0	1.33
16	$123.00	3	##	3	3	##	##	3	##		t	0	14	44	319	##	65	9	2	##	#	4.69	5	5	5	5	5	5		f	3	0	3	0	0.45
17	$90.00	30	##	30	30	##	##	30	##		t	0	0	0	249	##	23	2	0	##	#	4.36	4	4	4	4	4	4		f	5	0	5	0	0.17
18	$160.00	30	##	30	30	##	##	30	##		t	0	0	0	37	##	37	0	0	##	#	4.86	5	5	5	5	5	5		f	2	2	0	0	0.35
19	$84.00	30	##	30	30	##	##	30	##		t	5	8	12	274	##	185	4	0	##	#	4.56	5	5	5	5	5	5		t	1	1	0	0	1.14
20	$372.00	2	##	2	2	##	##	2	##		t	17	47	77	222	##	2	0	0	##	#	4	5	5	5	5	4	5		f	#	16	10	0	0.01

listings

Dataset

reviews.csv



	A	B	C	D	E	F
1	listing_id	id	date	reviewer_id	reviewer_name	comments
2	2539	55688172	12/4/2015	25160947	Peter	Great host
3	2539	97474898	8/27/2016	91513326	Liz	Nice room for the price. Great neighborhood. John was very accommodating. Bottles of water in the room were a nice touch a
4	2539	105340344	10/1/2016	90022459	Ð•Ð²Ð³ÐµÐ½Ð¸Ð¹	Very nice apt. New remodeled.
5	2539	133131670	2/20/2017	116165195	George	Great place to stay for a while. John is a great host and a great man. The house is very homy. You'll feel right at home.
6	2539	138349776	3/19/2017	118432644	Carlos	.
7	2539	157777930	6/4/2017	1806142	Isaac	I really enjoyed my time here in deep south Brooklyn and John was an excellent host - very attentive to any concerns that I hac
8	2539	194787986	9/17/2017	149989346	Amy	John was a great host. He was very responsive before and during our visit. He really wanted to make sure we had a good stay.
9	2539	292241108	7/17/2018	64678596	Dahn	Clean, quiet, well-run home, very close to the subway
10	2539	338504734	10/19/2018	3560397	Cem	John was always responsive during my stay. He answers back immediately when you request something from him. The neighb
11	2595	17857	11/21/2009	50679	Jean	Notre sÃ©jour de trois nuits. Nous avons apprÃ©cier L'appartement qui est trÃ¨s bien situÃ©. AgrÃ©able, propre et bie
12	2595	19176	12/5/2009	53267	Cate	Great experience.
13	2595	19760	12/10/2009	38960	Anita	I've stayed with my friend at the Midtown Castle for six days and it was a lovely place to be. A big spacious room with a pointy
14	2595	34320	4/9/2010	71130	Kai-Uwe	We've been staying here for about 9 nights, enjoying to be in the center of the city, that never sleeps...short ways to everywhe
15	2595	46312	5/25/2010	117113	Alicia	We had a wonderful stay at Jennifer's charming apartment! They were very organized and helpful; I would definitely recomme
16	2595	1238204	5/7/2012	1783688	Sergey	Hi to everyone! Would say our greatest compliments to Jennifer, the host of Midtown Castle. We spent in this lovely apar
17	2595	1293632	5/17/2012	1870771	LoÃ¯c	Jennifer was very friendly and helpful, and her place is exactly as advertised. The location is very convenient, and it was a pleas
18	2595	2022498	8/18/2012	2124102	Melanie	This apartment is like a real castle old and unique. The age- related stains on the bathroom floor and dark discolorations in the
19	2595	4682989	5/20/2013	496053	Eric	Jennifer's place was in a great midtown location, close to the subway, Bryant Park and Times Square. She met us after midnigh
20	2595	13193832	5/21/2014	13685934	Gerald	Jennifer is a very nice host. Everything is clean and she really takes care of her guests. We can only recommend her apartment

reviews

OUTLINE

Dataset

Graph

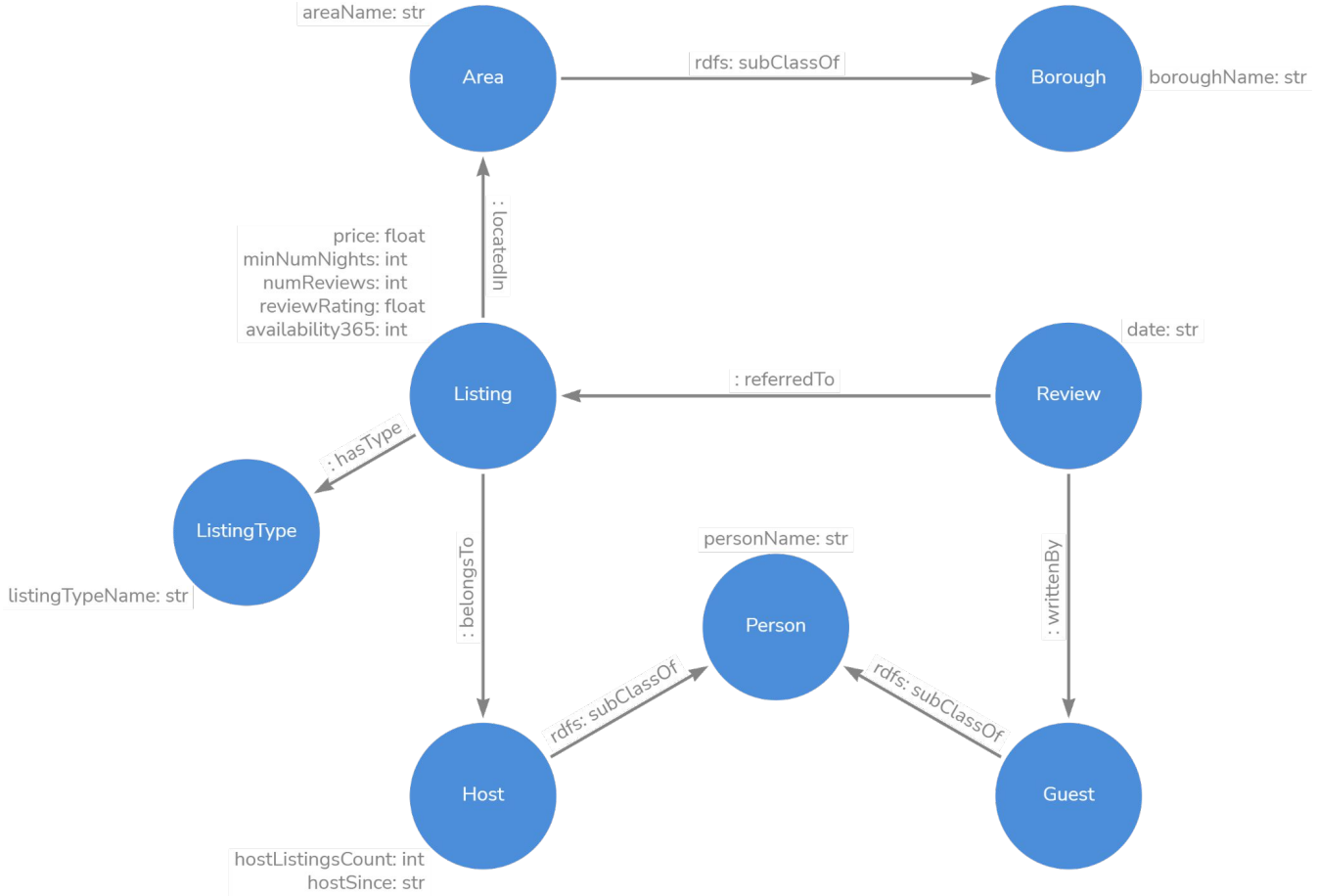
Ontology

Notebook

RDF Data

Queries

Graph



https://arrows.app/

OUTLINE

Dataset

Graph

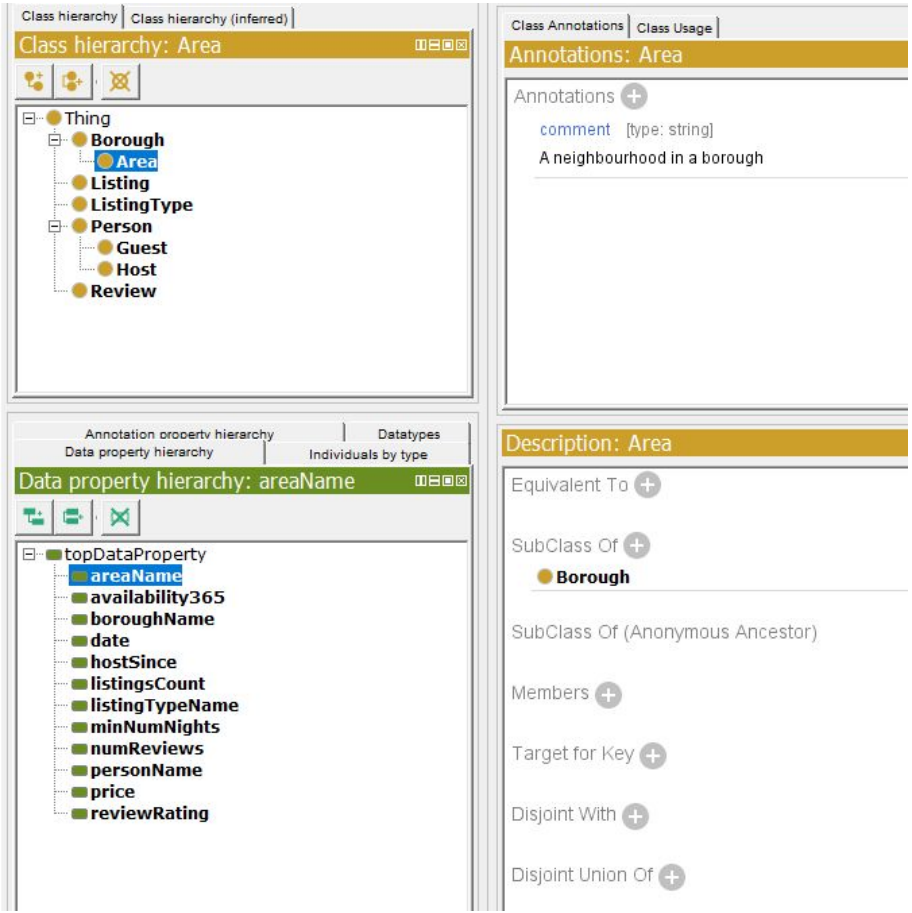
Ontology

Notebook

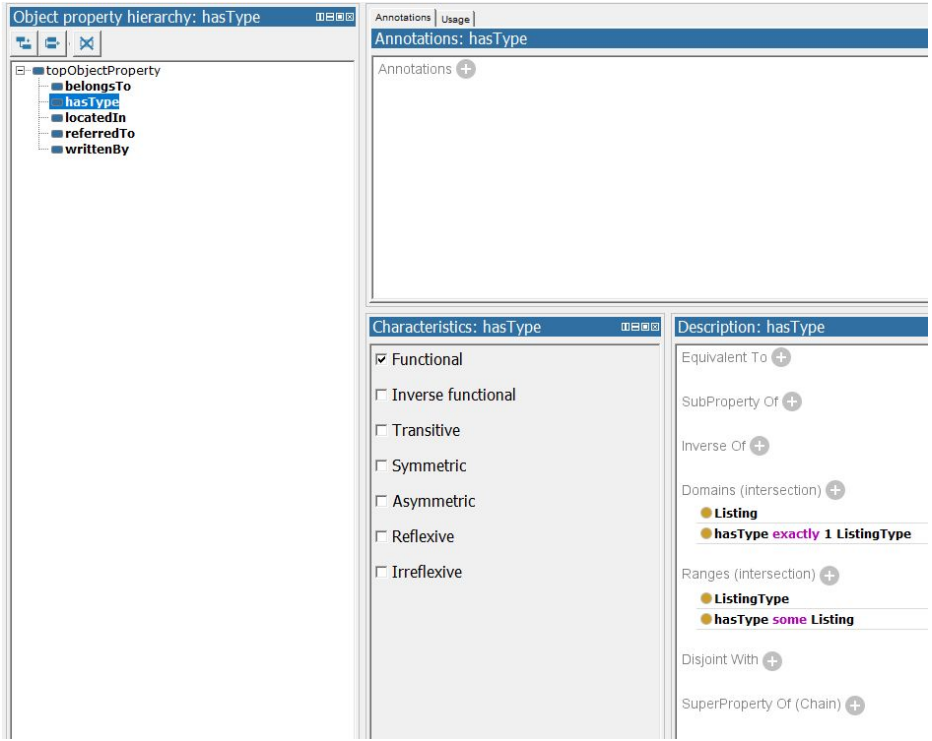
RDF Data

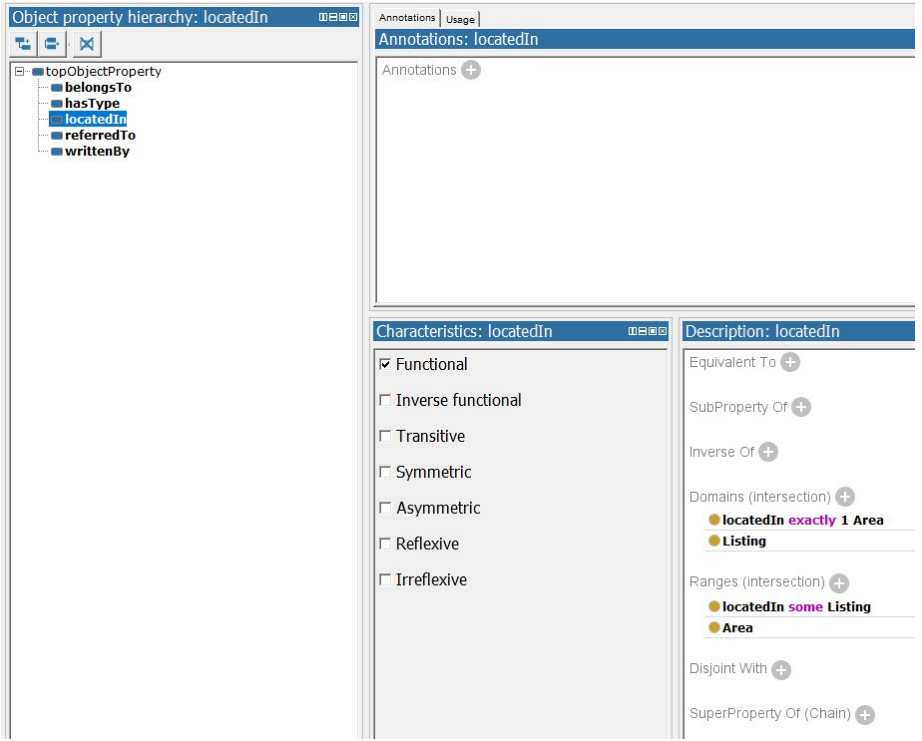
Queries

Ontology (Classes & Data Properties)

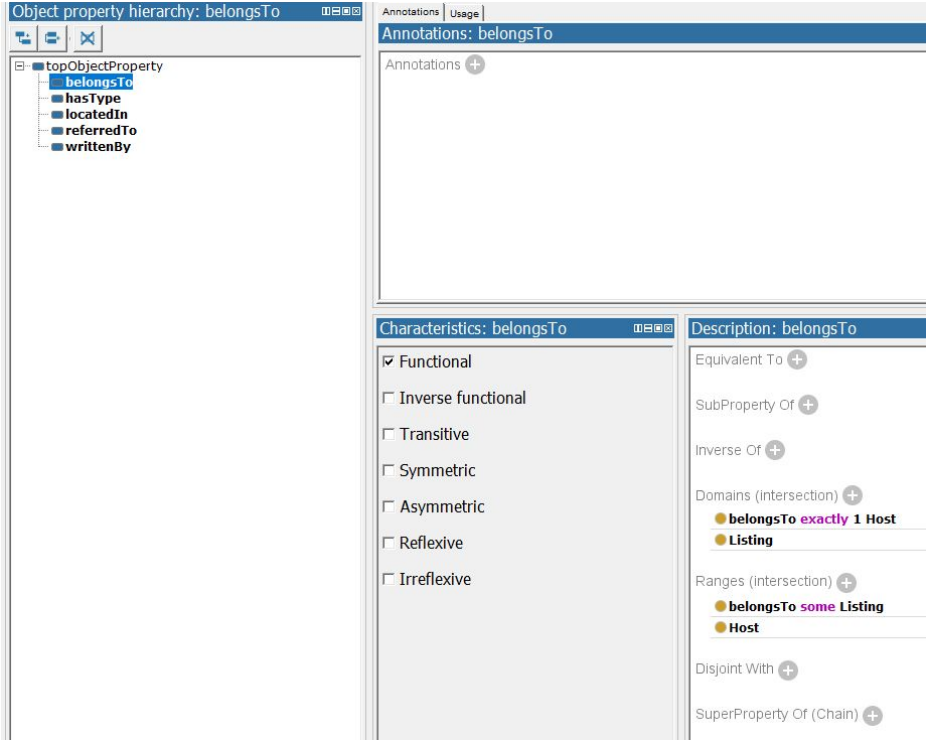


Ontology (Object Properties)

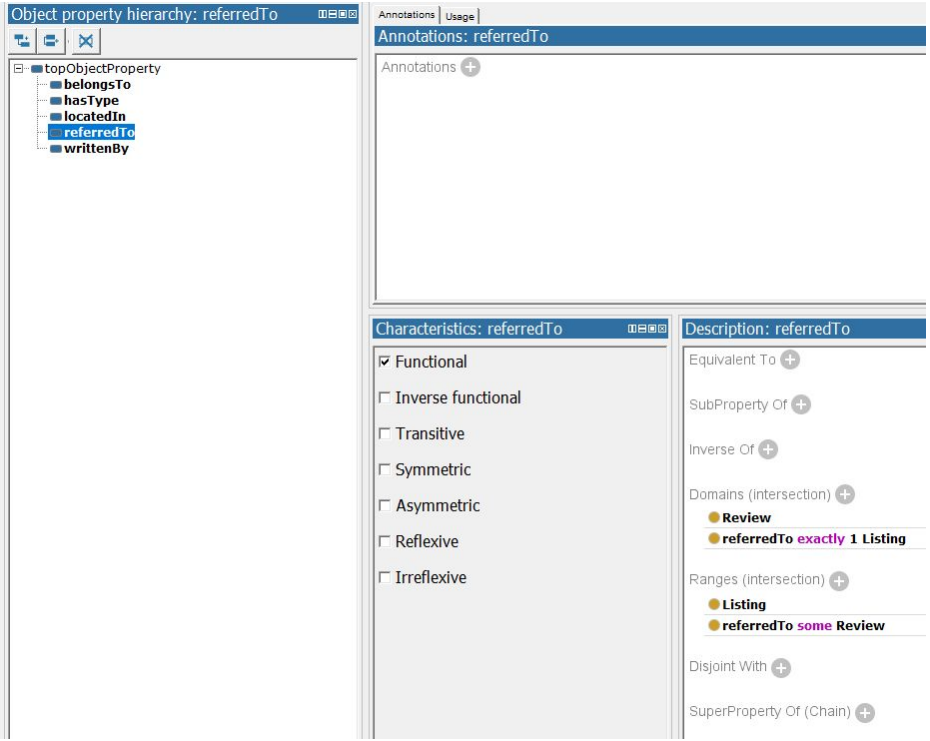


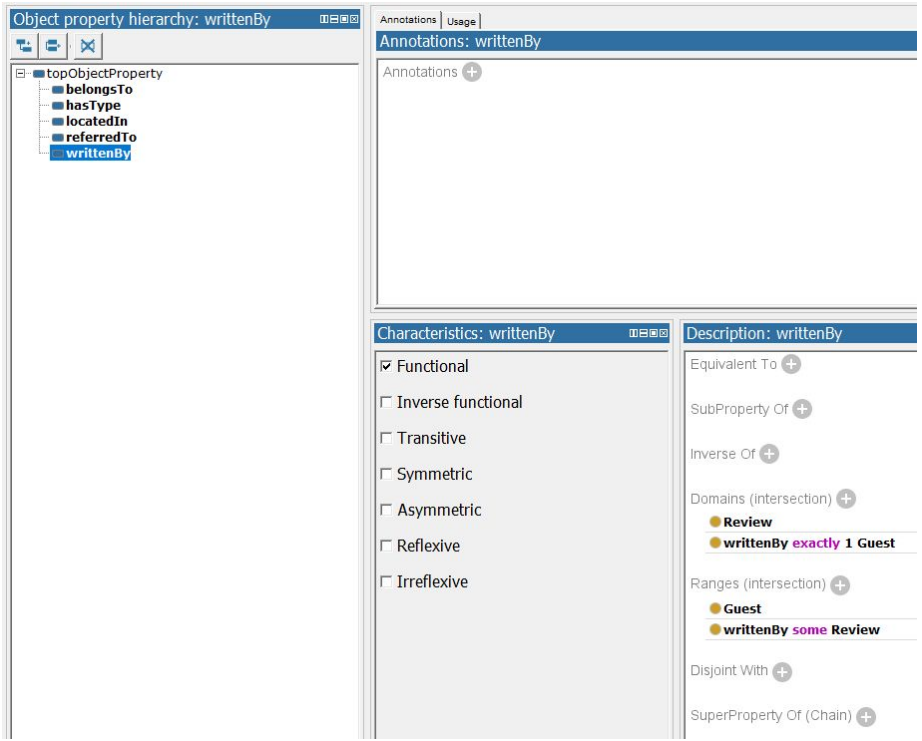


Ontology (Object Properties)



Ontology (Object Properties)





OUTLINE

Dataset

Graph

Ontology

Notebook

RDF Data

Queries

Notebook (Cleaning)

```
reviews = pd.read_csv(reviewsUrl, sep=',', index_col='id')
listings = pd.read_csv(listingsUrl, sep=',', index_col='id')

reviews.drop(['comments'], axis=1, inplace=True)
reviews.dropna(subset=['reviewer_name'], axis=0, inplace=True)
reviews.drop_duplicates(subset=['reviewer_id'], inplace=True)
reviews.drop(reviews[reviews.reviewer_name.str.contains(r'[^0-9a-zA-Z]')].index, inplace = True)

print("----------Reviews Done----------")

listings = listings[['host_id', 'host_name', 'host_since', 'host_listings_count',
                     'neighbourhood_cleansed', 'neighbourhood_group_cleansed', 'room_type',
                     'price', 'minimum_nights', 'availability_365', 'number_of_reviews', 'review_scores_rating']]

listings.dropna(subset=['host_name'], axis=0, inplace=True)
listings['review_scores_rating'].fillna(listings["review_scores_rating"].mean(), inplace=True)

print("----------Listings Done----------")
```

Notebook (Serialization)

```
%%time

g = Graph()
g.bind("xsd", XSD)
g.bind("ao", AO)

# iterate over the listings dataset
for index, row in listings.iterrows():
    # the node has the namespace + the listing_id as URI
    idU1 = "listing"+str(index)
    idU2 = "host"+str(row['host_id'])
    Listing = URIRef(AO[idU1])
    ListingType = URIRef(AO[clean_string(str(row['room_type']))])
    Area = URIRef(AO[clean_string(str(row['neighbourhood_cleansed']))])
    Host = URIRef(AO[idU2])

    # add triples
    g.add((Listing, RDF.type, AO.Listing))
    g.add((Listing, AO['price'], Literal(clean_number(row['price']), datatype=XSD.float)))
    g.add((Listing, AO['minNumNights'], Literal(row['minimum_nights'], datatype=XSD.int)))
    g.add((Listing, AO['numReviews'], Literal(row['number_of_reviews'], datatype=XSD.int)))
    g.add((Listing, AO['reviewRating'], Literal(row['review_scores_rating'], datatype=XSD.float)))
    g.add((Listing, AO['availability365'], Literal(row['availability_365'], datatype=XSD.int)))
    g.add((Listing, AO['locatedIn'], Area))
    g.add((Listing, AO['hasType'], ListingType))
    g.add((Listing, AO['belongsTo'], Host))
```

OUTLINE

Dataset

Graph

Ontology

Notebook

RDF Data

Queries

RDF Data

```
ao:listing10000070 a ao:Listing ;
    ao:availability365 "40"^^xsd:int ;
    ao:belongsTo ao:host51372003 ;
    ao:hasType ao:privateroom ;
    ao:locatedIn ao:flatbush ;
    ao:minNumNights "4"^^xsd:int ;
    ao:numReviews "15"^^xsd:int ;
    ao:price "109.0"^^xsd:float ;
    ao:reviewRating "4.6"^^xsd:float .
```

Listing

```
ao:allerton a ao:Area ;
    ao:areaName "Allerton"^^xsd:string ;
    rdfs:subClassOf ao:bronx .

ao:ardenheights a ao:Area ;
    ao:areaName "Arden Heights"^^xsd:string ;
    rdfs:subClassOf ao:statenisland .

ao:arrochar a ao:Area ;
    ao:areaName "Arrochar"^^xsd:string ;
    rdfs:subClassOf ao:statenisland .

ao:arverne a ao:Area ;
    ao:areaName "Arverne"^^xsd:string ;
    rdfs:subClassOf ao:queens .
```

Area

RDF Data

```
ao:entirehomeapt a ao:ListingType ;
    ao:listingTypeName "Entire home/apt"^^xsd:string .

ao:hotelroom a ao:ListingType ;
    ao:listingTypeName "Hotel room"^^xsd:string .

ao:privateroom a ao:ListingType ;
    ao:listingTypeName "Private room"^^xsd:string .

ao:sharedroom a ao:ListingType ;
    ao:listingTypeName "Shared room"^^xsd:string .
```

listingType

```
ao:bronx a ao:Borough ;
    ao:boroughName "Bronx"^^xsd:string .

ao:brooklyn a ao:Borough ;
    ao:boroughName "Brooklyn"^^xsd:string .

ao:manhattan a ao:Borough ;
    ao:boroughName "Manhattan"^^xsd:string .

ao:queens a ao:Borough ;
    ao:boroughName "Queens"^^xsd:string .

ao:statenisland a ao:Borough ;
    ao:boroughName "Staten Island"^^xsd:string .
```

Borough

OUTLINE

Dataset

Graph

Ontology

Notebook

RDF Data

Queries

Queries #1: Which listing type is the most repeated one in all boroughs?

```
select ?borough_name ?listing_type_name (COUNT (?Listing) as ?listing_count) where
{
    ?Listing ao:hasType ?listingType.
    ?listingType ao:listingTypeName ?listing_type_name.
    ?Listing ao:locatedIn ?Area.
    ?Area rdfs:subClassOf ?Borough.
    ?Borough ao:boroughName ?borough_name.
    {
        select ?Borough (MAX(?count) as ?max_count) where
        {
            select ?Borough ?listingType (COUNT (*) as ?count) where
            {
                ?Listing ao:hasType ?listingType.
                ?Listing ao:locatedIn ?Area.
                ?Area rdfs:subClassOf ?Borough.
            }
            GROUP BY ?Borough ?listingType
        }
        GROUP BY ?Borough
    }
}
GROUP BY ?borough_name ?listing_type_name ?max_count
HAVING (?listing_count = ?max_count)
```

Queries #1: Which listing type is the most repeated one in all boroughs?

	borough_name	listing_type_name	listing_count
1	"Bronx"	"Private room"	"793"^^xsd:integer
2	"Staten Island"	"Entire home/apt"	"273"^^xsd:integer
3	"Queens"	"Private room"	"3334"^^xsd:integer
4	"Brooklyn"	"Entire home/apt"	"8152"^^xsd:integer
5	"Manhattan"	"Entire home/apt"	"10837"^^xsd:integer

Queries #2: Do entire home/apt are less available than private room in a year?

```
ASK where
{
    {
        select ?listingType (SUM(?not_availability_per_listing) as ?sum_not_available_days_private) where
        {
            select ?Listing ?listingType ?not_availability_per_listing where
            {
                ?Listing ao:hasType ?listingType.
                ?listingType ao:listingTypeName "Private room"^^xsd:string.
                ?Listing ao:availability365 ?availability.
                BIND ("365"^^xsd:int - ?availability AS ?not_availability_per_listing)
            }
        }
        GROUP BY ?listingType
    }
    {
        select ?listingType (SUM(?not_availability_per_listing) as ?sum_not_available_days_entire) where
        {
            select ?Listing ?listingType ?not_availability_per_listing where
            {
                ?Listing ao:hasType ?listingType.
                ?listingType ao:listingTypeName "Entire home/apt"^^xsd:string.
                ?Listing ao:availability365 ?availability.
                BIND ("365"^^xsd:int - ?availability AS ?not_availability_per_listing)
            }
        }
        GROUP BY ?listingType
    }
    FILTER (?sum_not_available_days_entire > ?sum_not_available_days_private)
}
```

Queries #2: Do entire home/apt have more availability than private room in a year?

Query took 1.1s, moments ago.

NO

Queries #3: Find the total number of listings not located in Brooklyn

```
select (count(?Listing) as ?listing_count) where
{
    ?Listing rdf:type ao:Listing.
    MINUS
    {
      ?Listing ao:locatedIn ?Area.
      ?Area rdfs:subClassOf ?Borough.
      ?Borough ao:boroughName "Brooklyn".
    }
}
```

Queries #3: Find the total number of listings not located in Brooklyn

	listing_count
1	"24991"^^xsd:integer

Queries #4: Among listings of the host having max, which one got more reviews?

```
select ?Listing ?Host (count(?Review) as ?review_count) where
{
    ?Review ao:referredTo ?Listing.
    ?Listing ao:belongsTo ?Host.

    FILTER (?Host = ?inner_host) # host107434423
    {
        select ?inner_host (count(?Listing) as ?listing_count) where
        {
            ?Listing ao:belongsTo ?inner_host.
        }
        GROUP BY ?inner_host
        ORDER BY DESC(?listing_count)
        LIMIT 1
    }
}
GROUP BY ?Listing ?Host
ORDER BY DESC(?review_count)
LIMIT 1
```

Queries #4: Among listings of the host having max, which one got more reviews?

	Listing	Host	review_count
1	ao:listing46667739	ao:host107434423	"3"^^xsd:integer

Thank You
For Your Attention!