

CSC343 Project Presentation

Airbnb Activities in Seattle in 2016

Jiaming Yang & Changyan Xu (Group_0096)



Domain

Domain

- travel home-stays, (regional) small business
- Dataset URL: <https://www.kaggle.com/airbnb/seattle>

The screenshot shows the dataset page for the Seattle Airbnb Open Data on Kaggle. The header features a dark background with the Seattle skyline and the text "Seattle Airbnb Open Data" and "A sneak peek into the Airbnb activity in Seattle, WA, USA". Below the header, there's an Airbnb logo and the text "updated 2 years ago (Version 2)". The main navigation bar includes "Data" (which is underlined), "Tasks (1)", "Notebooks (54)", "Discussion (2)", "Activity", and "Metadata". There are also "Download (19 MB)" and "New Notebook" buttons. Below the navigation, there are sections for "Usability 7.1", "License CC0: Public Domain", and "Tags travel, hotels and accommodations, united states". At the bottom, there's a "Data Explorer" section showing "calendar.csv (34.95 MB)" with file details like "85.94 MB", "Detail", "Compact", "Column", and "4 of 4 columns". Other files listed are "listings.csv" and "reviews.csv".

Results



Results

1. What is the peak tourist seasons for airbnbs in Seattle
and what is the average price for home-stays among off-peak seasons and peak tourist seasons?

SPRING BREAK

month	occupancy_rate	avgprice
1	0.567057878204972	124.7965304764153982
2	0.668534797612151	125.7856564248441121
3	0.71779495275572	130.7401723430447271
4	0.672790357925493	137.4980004704775347
5	0.680824241853013	142.6538667866475159
6	0.67845142439737	150.7168389319552110
7	0.626510992247697	155.4834608947477293
8	0.645679681425104	153.6413218013283702
9	0.673005978108677	145.8341956726246472
10	0.694296793051021	139.1497224532753909
11	0.71241018146389	137.3819919310722101
12	0.733508550685351	139.2375634680892088
(12 rows)		

month	occupancy_rate	avgprice
1	0.567057878204972	124.7965304764153982
2	0.668534797612151	125.7856564248441121
3	0.71779495275572	130.7401723430447271
4	0.672790357925493	137.4980004704775347
5	0.680824241853013	142.6538667866475159
6	0.67845142439737	150.7168389319552110
7	0.626510992247697	155.4834608947477293
8	0.645679681425104	153.6413218013283702
9	0.673005978108677	145.8341956726246472
10	0.694296793051021	139.1497224532753909
11	0.71241018146389	137.3819919310722101
12	0.733508550685351	139.2375634680892089
(12 rows)		

Note: occupancy rate is the rate of occupancy in a given month among all listings in Seattle.

CHRISTMAS!

Results

2. Who are the top-rating (i.e hosts who has ever got a highest rating) super-hosts for each neighbourhood in Seattle?

neighbourhood	host_id	host_name
Alki	38502645	Ron
Arbor Heights	5407579	Gina
Atlantic	13200167	Kumail
Atlantic	32807985	Ted
Ballard	53050	Bob
Ballard	7466082	Vera
Ballard	9330646	Carol
Belltown	4160438	Evan
Belltown	8534462	Daniela
Belltown	31668017	Andrew
Bitter Lake	19660299	Daniel And Teresa
Brighton	2536735	Akiko And Matt
Broadview	2907322	Archana
Broadway	4729705	Rob And Carol
Bryant	2409388	Lisa
Capitol Hill	31509	David
Capitol Hill	595610	Elena
Capitol Hill	658155	Joe
Capitol Hill	7673693	Shawn
Capitol Hill	9195336	Scooter
Capitol Hill	13309620	Craig

Result: Capitol Hill, Fremont, Leschi, Minor

— have >4 super-hosts

Note: "Superhost"s are the experienced and extraordinary hosts, where the official grants this title as the host satisfies specific criteria.

Results

3. How are the mainstream property types and non-mainstream property types associated with the listing's score?

property_type	avgscore	highest_score	lowest_score
Townhouse	96.625	100	80
Condominium	95.3188405797101	100	80
House	94.8335616438356	100	40
Apartment	93.9775910364146	100	20
(4 rows)			

main-stream property types

property_type	avgscore	highest_score	lowest_score
Yurt	100	100	100
Bungalow	97.6666666666667	100	94
Other	96.2631578947368	100	90
Cabin	96.2380952380952	100	84
Treehouse	96	100	92
Bed & Breakfast	95.7307692307692	100	85
Camper/RV	95.7	100	90
Boat	95.6	100	90
Loft	95.4545454545455	100	86
Tent	94.5	100	90
Dorm	90	93	87
Chalet	83.5	96	71
(13 rows)			

non-mainstream property types.

Note: "mainstream property type" is the property type that takes up at least 2% of all property types of the listings in Seattle.



Evolution

Evolution

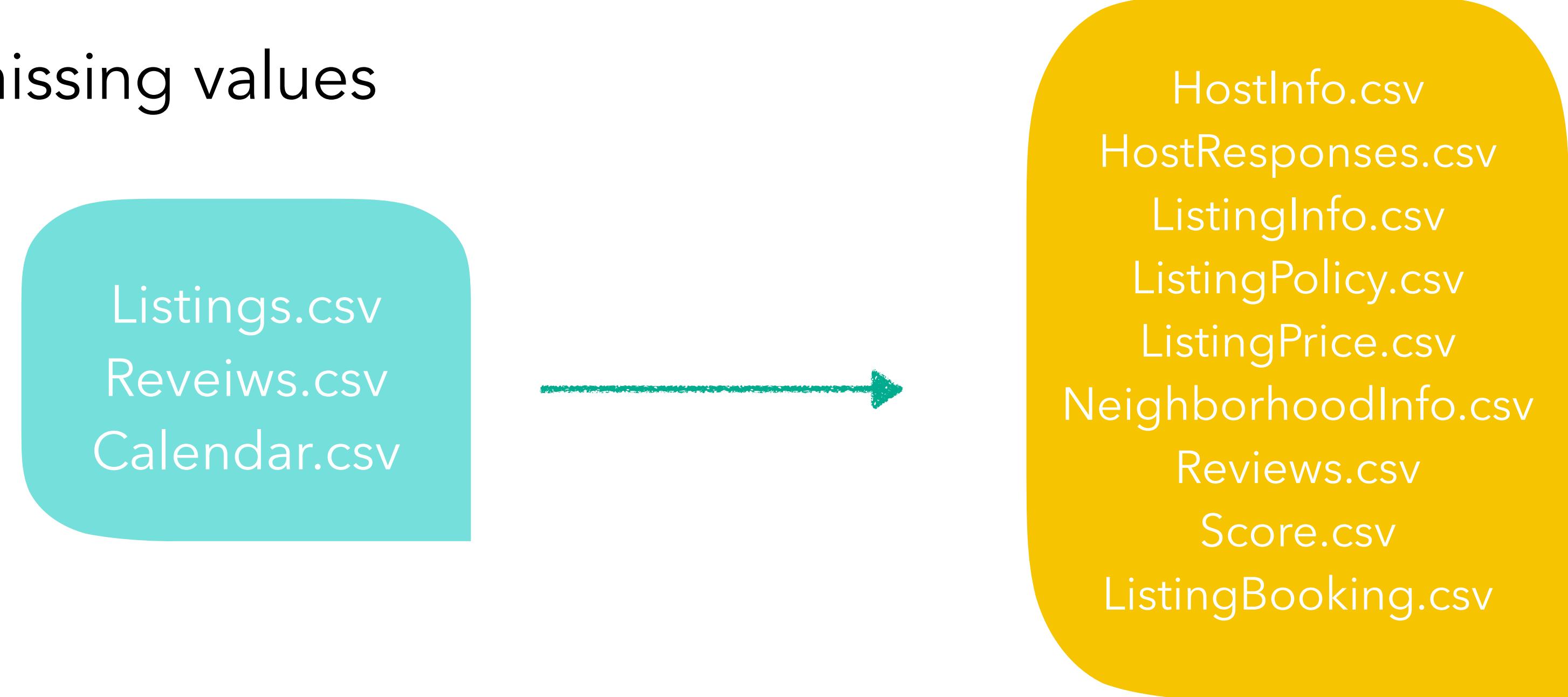
Changed the Schema (a tradeoff)

- table “Review”
- Which is better?
 - Approach 1:
 - Review(review_id, listing_id, reviewer_id, reviewer_name, date, comments)
 - Approach 2:
 - Review(review_id, listing_id, reviewer_id, date, comments)
 - ReviewerInfo(reviewer_id, reviewer_name)
- Final decision: NOT to break this Review table
 - extra table is too small and not quite useful
 - assume that a customer wouldn’t change their name (hold in most cases)
- If we have more information about the reviewers, then we should certainly have another table called ReviewerInfo.

```
--information about the review of the listing
Create table Review(
    --id of the review
    review_id Integer primary key,
    --id of the listing
    listing_id Integer not null,
    --id of the reviewer
    reviewer_id Integer not null,
    --name fo the reviewer
    reviewer_name varchar(1000) not null,
    --date of the review
    date Date,
    --content of the review
    comments varchar(100000),
    foreign key (listing_id) references ListingInfo);
```

Our design choices we made

- Splitting up tables
 - to avoid redundancy
 - to avoid missing values



- Remove useless attributes

Challenges and Lessons



Lessons

— lessons we learned about the database

- We learned how to figure things out by ourselves and make things work.
 - e.g. '\$' in price columns
 - Type cast: varchar::money::numeric

listing_id	price	weekly_price	monthly_price	security_deposit	cleaning_fee
241032	\$85.00				
953595	\$150.00	\$1,000.00	\$3,000.00	\$100.00	\$40.00
3308979	\$975.00			\$1,000.00	\$300.00
7421966	\$100.00	\$650.00	\$2,300.00		

Lessons

— lessons we learned about the database

- How to create a schema in sql using ddl.
- How to import a csv file into the database.
- How to clean the data using queries.

Thank you for listening!

Questions?