



Airbnb Case Study

Oyo Data Science Academy

Agenda

- Problem Statement
- Data Pull
- Data Engineering
- Analysis and Discussion

Problem Statement

Using data on listings can we create a feature that suggests the property owner about the price to put, to have a better chance of renting out the property?

Problem Statement

Using data on listings can we create a feature that suggests the property owner about the price to put, to have a better chance of renting out the property?



Data

Data is about Airbnb Price listings, stored in the data warehouse, in the following tables

Data

Data is about Airbnb Price listings, stored in the data warehouse, in the following tables

Calendar

- listing_id
- date
- available
- price
- minimum_nights
- Maximum_nights

Data

Data is about Airbnb Price listings, stored in the data warehouse, in the following tables

Calendar

- listing_id
- date
- available
- price
- minimum_nights
- Maximum_nights

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Data

Data is about Airbnb Price listings, stored in the data warehouse, in the following tables

Calendar

- listing_id
- date
- available
- price
- minimum_nights
- Maximum_nights

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Hosts

- host_id
- host_name
- host_since
- host_location
- host_about

Data

Data is about Airbnb Price listings, stored in the data warehouse, in the following tables

Calendar

- listing_id
- date
- available
- price
- minimum_nights
- Maximum_nights

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Hosts

- host_id
- host_name
- host_since
- host_location
- host_about

Reviews

- listing_id
- review_id
- date
- reviewer_id
- reviewer_name
- comments

Data Pull

Calendar

- listing_id
- date
- available
- price
- minimum_nights
- Maximum_nights

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Hosts

- host_id
- host_name
- host_since
- host_location
- host_about

Reviews

- listing_id
- review_id
- date
- reviewer_id
- reviewer_name
- comments

What is going to be your dependent variable and what variables can you use to make predictions?

Data Pull

Calendar

- listing_id
- date
- available
- **price**
- minimum_nights
- Maximum_nights

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Hosts

- host_id
- host_name
- host_since
- host_location
- host_about

Reviews

- listing_id
- review_id
- date
- reviewer_id
- reviewer_name
- comments

What is going to be your dependent variable and what variables can you use to make predictions?

Data Pull

Calendar

- listing_id
- date
- available
- **price**
- minimum_nights
- Maximum_nights

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	6418099	2021-12-26	t	\$48.00	\$48.00	3	1125
1	11649861	2021-12-26	f	\$100.00	\$100.00	1	1125
2	11649861	2021-12-27	f	\$100.00	\$100.00	1	1125
3	11649861	2021-12-28	f	\$100.00	\$100.00	1	1125
4	11649861	2021-12-29	f	\$100.00	\$100.00	1	1125
5	11649861	2021-12-30	f	\$100.00	\$100.00	1	1125
6	11649861	2021-12-31	f	\$100.00	\$100.00	1	1125
7	11649861	2022-01-01	t	\$100.00	\$100.00	1	1125
8	11649861	2022-01-02	f	\$100.00	\$100.00	1	1125
9	11649861	2022-01-03	f	\$100.00	\$100.00	1	1125

What is going to be your dependent variable and what variables can you use to make predictions?

Data Pull

Calendar

- listing_id
- date
- available
- **price**
- minimum_nights
- Maximum_nights

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	6418099	2021-12-26	t	\$48.00	\$48.00	3	1125
1	11649861	2021-12-26	f	\$100.00	\$100.00	1	1125
2	11649861	2021-12-27	f	\$100.00	\$100.00	1	1125
3	11649861	2021-12-28	f	\$100.00	\$100.00	1	1125
4	11649861	2021-12-29	f	\$100.00	\$100.00	1	1125
5	11649861	2021-12-30	f	\$100.00	\$100.00	1	1125
6	11649861	2021-12-31	f	\$100.00	\$100.00	1	1125
7	11649861	2022-01-01	t	\$100.00	\$100.00	1	1125
8	11649861	2022-01-02	f	\$100.00	\$100.00	1	1125
9	11649861	2022-01-03	f	\$100.00	\$100.00	1	1125

What is going to be your dependent variable and what variables can you use to make predictions?

- Price?
- Avg Price?
- Avg Price Last 30 days?

Data Pull

Calendar

- listing_id
- date
- available
- **price**
- minimum_nights
- Maximum_nights

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	6418099	2021-12-26	t	\$48.00	\$48.00	3	1125
1	11649861	2021-12-26	f	\$100.00	\$100.00	1	1125
2	11649861	2021-12-27	f	\$100.00	\$100.00	1	1125
3	11649861	2021-12-28	f	\$100.00	\$100.00	1	1125
4	11649861	2021-12-29	f	\$100.00	\$100.00	1	1125
5	11649861	2021-12-30	f	\$100.00	\$100.00	1	1125
6	11649861	2021-12-31	f	\$100.00	\$100.00	1	1125
7	11649861	2022-01-01	t	\$100.00	\$100.00	1	1125
8	11649861	2022-01-02	f	\$100.00	\$100.00	1	1125
9	11649861	2022-01-03	f	\$100.00	\$100.00	1	1125

How will you extract 30 day average price for each listing?
Think over what the SQL query will look like.

Data Pull

Code Demo

Data Pull

Calendar

- listing_id
- date
- available
- **price**
- minimum_nights
- Maximum_nights

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	6418099	2021-12-26	t	\$48.00	\$48.00	3	1125
1	11649861	2021-12-26	f	\$100.00	\$100.00	1	1125
2	11649861	2021-12-27	f	\$100.00	\$100.00	1	1125
3	11649861	2021-12-28	f	\$100.00	\$100.00	1	1125
4	11649861	2021-12-29	f	\$100.00	\$100.00	1	1125
5	11649861	2021-12-30	f	\$100.00	\$100.00	1	1125
6	11649861	2021-12-31	f	\$100.00	\$100.00	1	1125
7	11649861	2022-01-01	t	\$100.00	\$100.00	1	1125
8	11649861	2022-01-02	f	\$100.00	\$100.00	1	1125
9	11649861	2022-01-03	f	\$100.00	\$100.00	1	1125

Once we are able to answer the question of how to work with the price variable, can you think of any features you can derive from this table for modelling purposes?

Data Pull

Calendar

- listing_id
- date
- available
- **price**
- minimum_nights
- Maximum_nights

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	6418099	2021-12-26	t	\$48.00	\$48.00	3	1125
1	11649861	2021-12-26	f	\$100.00	\$100.00	1	1125
2	11649861	2021-12-27	f	\$100.00	\$100.00	1	1125
3	11649861	2021-12-28	f	\$100.00	\$100.00	1	1125
4	11649861	2021-12-29	f	\$100.00	\$100.00	1	1125
5	11649861	2021-12-30	f	\$100.00	\$100.00	1	1125
6	11649861	2021-12-31	f	\$100.00	\$100.00	1	1125
7	11649861	2022-01-01	t	\$100.00	\$100.00	1	1125
8	11649861	2022-01-02	f	\$100.00	\$100.00	1	1125
9	11649861	2022-01-03	f	\$100.00	\$100.00	1	1125

Once we are able to answer the question of how to work with the price variable, can you think of any features you can derive from this table for modelling purposes?

Data Pull

Reviews

- listing_id
- review_id
- date
- reviewer_id
- reviewer_name
- comments

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	50904	31511792	2015-05-06	19482395	Jihae	Karin's "Aplace" is absolutely beautiful and c...
1	50904	470101024356869935	2021-10-10	333559	Emilie	Karin is a wonderful host, she was really help...
2	116134	972063	2012-03-05	928644	Aurélien	Amazing flat, really close from the MAS Museum,...
3	116134	1344011	2012-05-25	231288	Gail	This is a well equipped, very comfortable apar...
4	116134	7019975	2013-09-03	7984251	Marcel	This is a very nice appartement. We really lik...

- How can you use this table?
- What aggregations can you do?

Data Pull

Reviews

- listing_id
- review_id
- date
- reviewer_id
- reviewer_name
- comments

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	50904	31511792	2015-05-06	19482395	Jihae	Karin's "Aplace" is absolutely beautiful and c...
1	50904	470101024356869935	2021-10-10	333559	Emilie	Karin is a wonderful host, she was really help...
2	116134	972063	2012-03-05	928644	Aurélien	Amazing flat, really close from the MAS Museum,...
3	116134	1344011	2012-05-25	231288	Gail	This is a well equipped, very comfortable apar...
4	116134	7019975	2013-09-03	7984251	Marcel	This is a very nice appartement. We really lik...

- How can you use this table?
- What aggregations can you do?

Data Pull

Reviews

- listing_id
- review_id
- date
- reviewer_id
- reviewer_name
- comments

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	50904	31511792	2015-05-06	19482395	Jihae	Karin's "Aplace" is absolutely beautiful and c...
1	50904	470101024356869935	2021-10-10	333559	Emilie	Karin is a wonderful host, she was really help...
2	116134	972063	2012-03-05	928644	Aurélien	Amazing flat, really close from the MAS Museum,...
3	116134	1344011	2012-05-25	231288	Gail	This is a well equipped, very comfortable apar...
4	116134	7019975	2013-09-03	7984251	Marcel	This is a very nice appartement. We really lik...

- How can you use this table?
- What aggregations can you do?

- Total Num reviews till date
- Reviews in last 30/40/50 days

Data Pull

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Hosts

- host_id
- host_name
- host_since
- host_location
- host_about

- How can we use these two tables?

Data Pull

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Hosts

- host_id
- host_name
- host_since
- host_location
- host_about

- How can we use these two tables?
- Create feature on number of listings per host etc

Data Pull

Calendar

- listing_id
- date
- available
- price
- minimum_nights
- Maximum_nights

Listings

- listing_id
- listing_url
- name
- description
- lat
- long
- property_type
- room_type
- accommodates
- bathrooms
- bedrooms
- beds
- amenities
- host_id

Hosts

- host_id
- host_name
- host_since
- host_location
- host_about

Reviews

- listing_id
- review_id
- date
- reviewer_id
- reviewer_name
- comments

listing_id	price_30	#_reviews	#_reviews_30	availability_30	availability_60	long	lat	#_amenities
...

Data Engineering

listing_id	price_30	#_reviews	#_reviews_30	availability_30	availability_60	long	lat	...	#_amenities
...

Data Engineering

listing_id	price_30	#_reviews	#_reviews_30	availability_30	availability_60	long	lat	...	#_amenities
...

neighbourhood_group	neighbourhood
	Amandus - Atheneum
	Borgerhout Extra Muros
	Borgerhout Intra Muros Noord
	Borgerhout Intra Muros Zuid
	Brederode
	Centraal Station
	Dam
	Deurne Dorp - Gallifort
	Deurne Noord
	Deurne Oost
	Deurne Vlieghaven
	Deurne Zuid Oost
	Deurne Zuid West

Data Engineering

listing_id	price_30	#_reviews	#_reviews_30	availability_30	availability_60	long	lat	...	#_amenities
...

neighbourhood_group	neighbourhood
	Amandus - Atheneum
	Borgerhout Extra Muros
	Borgerhout Intra Muros Noord
	Borgerhout Intra Muros Zuid
	Brederode
	Centraal Station
	Dam
	Deurne Dorp - Gallifort
	Deurne Noord
	Deurne Oost
	Deurne Vlieghaven
	Deurne Zuid Oost
	Deurne Zuid West

- What is the city center?
- How far is each listing from city center?

Analysis and Discussion

listing_id	price_30	#_reviews	#_reviews_30	availability_30	availability_60	long	lat	...	#_amenities
...

- What kind of model will you make?
- How will you communicate results to stakeholders?
- How will you monitor and re-train