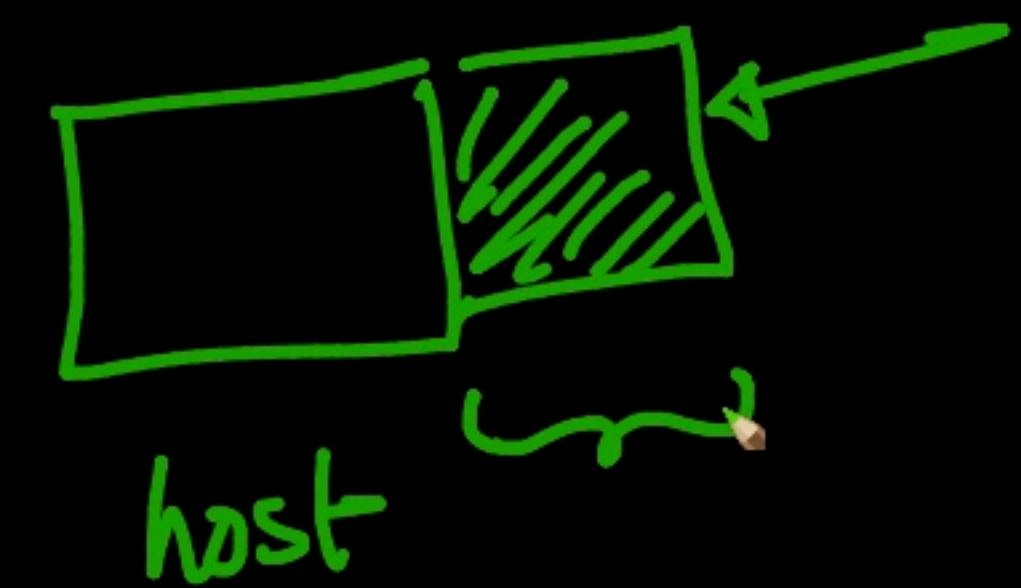


Data visualization → Tableau; Excel, Python
case study: Airbnb ← Simulated data

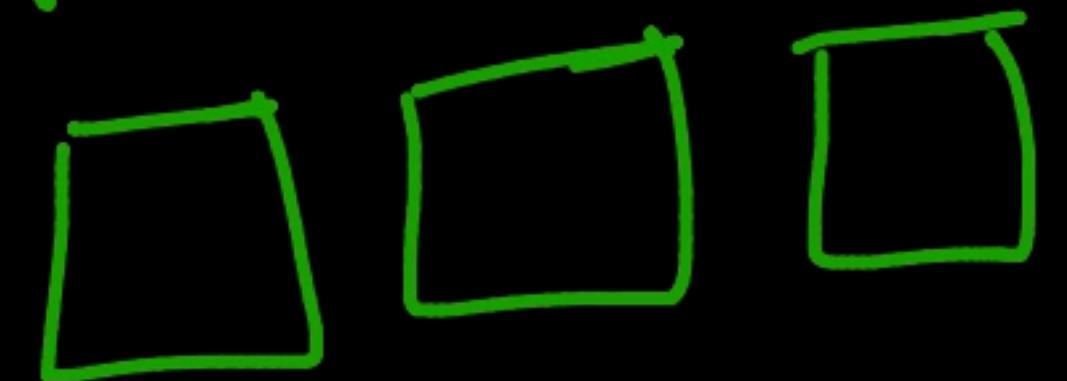
- class starts @ 9:03 PM -

- Flight overbooking → @ the end
- Sample given empirical distribution?
- user-segments; ToD/DOW segments

Airbnb



superhost:



3bedrooms

Images



improve your chances of getting a booking

6(a) min # images → ③

6(b) recommended range of # images

③

7-10

} to be posted by the host

Is one of the factors

reviews

price

location

* more images is tedious

docs.google.com/spreadsheets/d/1aBrE-8JaUsigK7eXVWpb-aGF1dVv918/edit#gid=1951333285

Dataset_1_Listings.csv.xlsx

Dataset_2_Open_Listings.csv

Dataset_3_Redundant_Listings

Dataset_1_Listings.csv XLSX

File Edit View Insert Format Data Tools Help

Share

150% 100% 123 Default

H7

	A	B	C	D	E	F	G	H	I	J
1	Listing_Id	Posting_Date	Posting_Time	Location	Images	Bookings	Host_Type			
2	4955430	13/05/2019	20:18:13	San Francisco		31	10	Regular		
3	3338094	04/01/2016	5:30:37	Amsterdam		6	48	Regular		
4	3132720	24/12/2018	12:26:15	Barcelona		25	10	Regular		
5	3196344	14/10/2014	15:09:27	Berlin		10	134	Superhost		
6	4588846	21/07/2016	9:26:48	Brussels		4	7	Regular		
7	4677052	14/05/2016	6:40:39	Kyoto		3	5	Regular		
8	3485608	04/04/2016	10:30:15	Madrid		7	44	Regular		
9	4980709	03/04/2017	9:44:58	Tokyo		12	52	Regular		
10	4761153	14/01/2018	15:44:54	Los Angeles		28	81	Superhost		
11	4991055	09/08/2017	10:26:32	Seattle		14	109	Regular		
12	5135019	06/01/2018	22:53:12	Mexico City		11	24	Regular		
13	5040928	16/01/2015	1:40:57	Milano		12	105	Regular		
14	3297315	05/01/2017	17:56:18	Paris		21	36	Regular		
15	5793190	20/12/2018	14:11:43	Osaka		2	3	Regular		
16	4435963	17/10/2015	1:14:07	Rio		23	234	Superhost		
17	3940988	28/07/2018	12:09:09	Seoul		10	51	Regular		
18	5433570	23/02/2019	9:22:07	Sao Paulo		10	16	Regular		
19	4920877	13/03/2015	2:23:16	Roma		22	71	Regular		
20	5008788	06/09/2017	19:14:52	Bengaluru		25	36	Regular		
21	5549595	20/06/2019	3:43:12	New Delhi		15	12	Regular		
22	4550407	28/02/2017	11:42:20	Chennai		15	32	Regular		
23	5228587	28/02/2017	19:40:17	Berlin		14	45	Superhost		

docs.google.com/spreadsheets/d/1aBrE-8JaUsig4k7eXVWpb-aGF1dVv918/edit#gid=1951333285

Dataset_1_Listings.csv XLSX

File Edit View Insert Format Data Tools Help

150% 10 Default 123

H7 fx

	A	B	C	D	E	F	G	H	I	J
1	Listing_Id	Posting_Date	Posting_Time	Location	Images	Bookings	Host_Type			
2	4955430	13/05/2019	20:18:13	San Francisco	31	10	Regular			
3	3338094	04/01/2016	5:30:37	Amsterdam	6	48	Regular			
4	3132720	24/12/2018	12:26:15	Barcelona	25	10	Regular			
5	3196344	14/10/2014	15:09:27	Berlin	10	134	Superhost			
6	4588846	21/07/2016	9:26:48	Brussels	4	7	Regular			
7	4677052	14/05/2016	6:40:39	Kyoto	3	5	Regular			
8	3485608	04/04/2016	10:30:15	Madrid	7	44	Regular			
9	4980709	03/04/2017	9:44:58	Tokyo	12	52	Regular			
10	4761153	14/01/2018	15:44:54	Los Angeles	28	81	Superhost			
11	4991055	09/08/2017	10:26:32	Seattle	14	109	Regular			
12	5135019	06/01/2018	22:53:12	Mexico City	11	24	Regular			
13	5040928	16/01/2015	1:40:57	Milano	12	105	Regular			
14	3297315	05/01/2017	17:56:18	Paris	21	36	Regular			
15	5793190	20/12/2018	14:11:43	Osaka	2	3	Regular			
16	4435963	17/10/2015	1:14:07	Rio	23	234	Superhost			
17	3940988	28/07/2018	12:09:09	Seoul	10	51	Regular			
18	5433570	23/02/2019	9:22:07	Sao Paulo	10	16	Regular			
19	4920877	13/03/2015	2:23:16	Roma	22	71	Regular			
20	5008788	06/09/2017	19:14:52	Bengaluru	25	36	Regular			
21	5549595	20/06/2019	3:43:12	New Delhi	15	12	Regular			
22	4550407	28/02/2017	11:42:20	Chennai	15	32	Regular			
23	5228587	28/02/2017	19:40:17	Berlin	14	45	Superhost			

docs.google.com/spreadsheets/d/1cilmQoWY2ri5d60d9OD3hBMbjcjUUMVK/edit#gid=1434565827

Dataset_2_Open_Listings.csv XLSX

File Edit View Insert Format Data Tools Help

F2 fx 38364

A B C D E F G H

Date Open_Listings_0_2 Open_Listings_3_5 Open_Listings_6_10 Open_Listings_11_15 Open_Listings_16

1 08/01/2018 51918 59583 9558 28830 38364

2 08/02/2018 50196 67897 10820 23024 31676

3 08/03/2018 56239 52148 24258 34328 7342

4 08/04/2018 57762 67057 15148 15531 23232

5 08/05/2018 59221 53113 18165 18387 8486

6 08/06/2018 51043 66962 18485 20555 27451

7 08/07/2018 51408 67209 9978 5250 14283

8 08/08/2018 56677 61590 13495 13531 38859

9 08/09/2018 57704 66742 25888 30466 35611

10 08/10/2018 52860 56609 9717 30089 13067

11 08/11/2018 50923 61888 10377 26318 26939

12 08/12/2018 51922 66660 23345 11538 22706

13 08/13/2018 58363 52087 33627 9313 31431

14 08/14/2018 58851 65549 34365 19516 27903

15 08/15/2018 50124 52314 22446 31637 35145

16 08/16/2018 58760 69791 25206 27806 39683

17 08/17/2018 54208 68304 25132 28452 9576

18 08/18/2018 50744 61919 16093 26516 36126

19 08/19/2018 56778 67252 8039 16672 25703

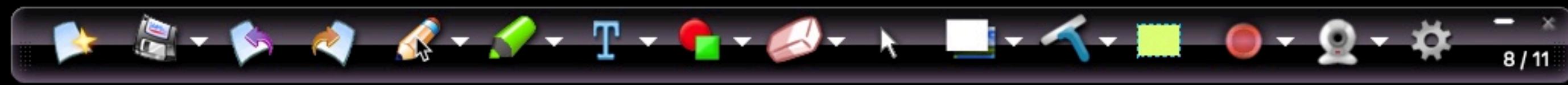
20 08/20/2018 58640 61927 8711 28163 14041

21 08/21/2018 55250 58742 30490 9119 25130

22 08/22/2018 53564 54316 15635 9601 36038

available but not booked
"open"
1 day

648af6c0-fee7-11e9-b53c-ebfc1c5



docs.google.com/spreadsheets/d/1OqO54nFQ9Afq4wU2UmaYtjCJmBfiW3AQ/edit#gid=2010248358

Dataset_3_Redundant_Listings.csv .XLSX

File Edit View Insert Format Data Tools Help

150% 10 Default 123

C2 fx 428368

	A	B	C	D	E	F	G	H	I	J
1	Property_Images	Total_Listings	Redundant_Listings							
2	0 to 2	563,246	428,368							
3	3 to 5	716,923	264,321							
4	6 to 10	1,341,164	98,874							
5	11 to 15	993,269	53,292							
6	15 to 20	873,269	61,382							
7	21 to 25	635,481	29,285							
8	26 to 30	498,456	42,473							
9	> 30	384,875	28,274							
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

671b8cb0-fee7-11e9-b53c-ebfc1c5 9/9

docs.google.com/spreadsheets/d/1OqO54nFQ9Afq4wU2UmaYtjCJmBfiW3AQ/edit#gid=2010248358

Dataset_3_Redundant_Listings.csv .XLSX

File Edit View Insert Format Data Tools Help

Share +9

C2 | fx 428368

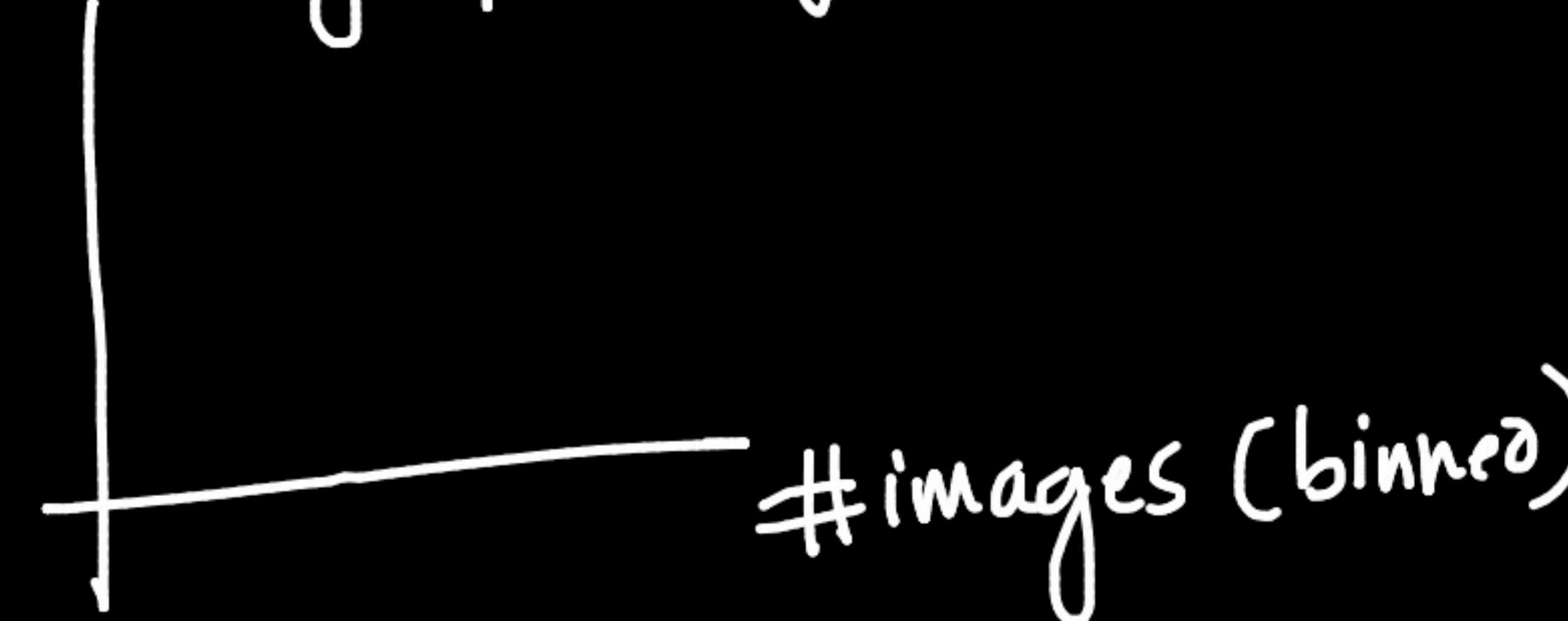
	A	B	C	D	E	F	G	H	I	J
1	Property_Images	Total_Listings	Redundant_Listings							
2	0 to 2	563,246	428,368							
3	3 to 5	716,923	264,321							
4	6 to 10	1,341,164	98,874							
5	11 to 15	993,269	53,292							
6	15 to 20	873,269	61,382							
7	21 to 25	635,481	29,285							
8	26 to 30	498,456	42,473							
9	> 30	384,875	28,274							
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

671b8cb0-fee7-11e9-b53c-ebfc1c5 10 / 10

Ideas:

percentage of bookings

①



②

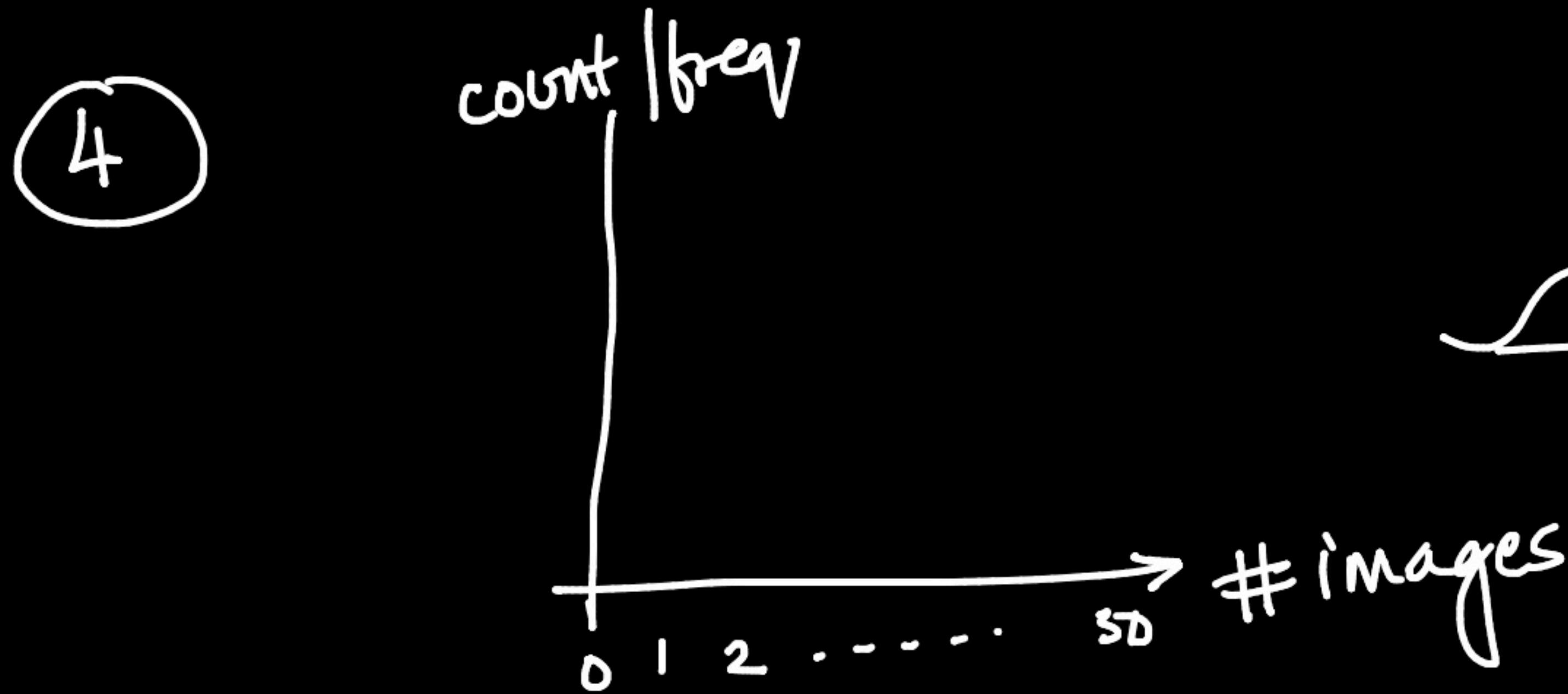
correlation b/w # images & booking count

- min # images
- recommend # images (range)

*

③

slice data by hostType



① min # images ✓
② recommended # images
(range)

⑤ ideally: other slices: location ; age of property

⑥

fog

① & ② X

#bookifs
==

⑦

#bookifs

slice: hostType

#images (bin)
==

docs.google.com/spreadsheets/d/1OqO54nFQ9Afq4wU2UmaYtjCJmBfiW3AQ/edit#gid=2010248358

Dataset_3_Redundant_Listings.csv .XLSX

B9 fx 384875

	A	B
1	Property_Images	Total_Listings
2	0 to 2	563,246
3	3 to 5	716,923
4	6 to 10	1,341,164
5	11 to 15	993,269
6	15 to 20	873,269
7	21 to 25	635,481
8	26 to 30	498,456
9	> 30	384,875
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		

Tableau Public - Airbnb_case_study_workbook_extract_new

Percentage of open lisitngs by number of images in last one year

Images Category	Percentage
Images 0-2	30.96%
Images 3-5	33.57%
Images 6-10	11.89%
Images 11-15	10.71%
Images >16	12.87%

Data Analytics

Listing_dataset
open listing data
Redundant listing data

Search

Tables

Date

Measure Names

Color
Size
Detail
Tooltip

Measure Values

Value

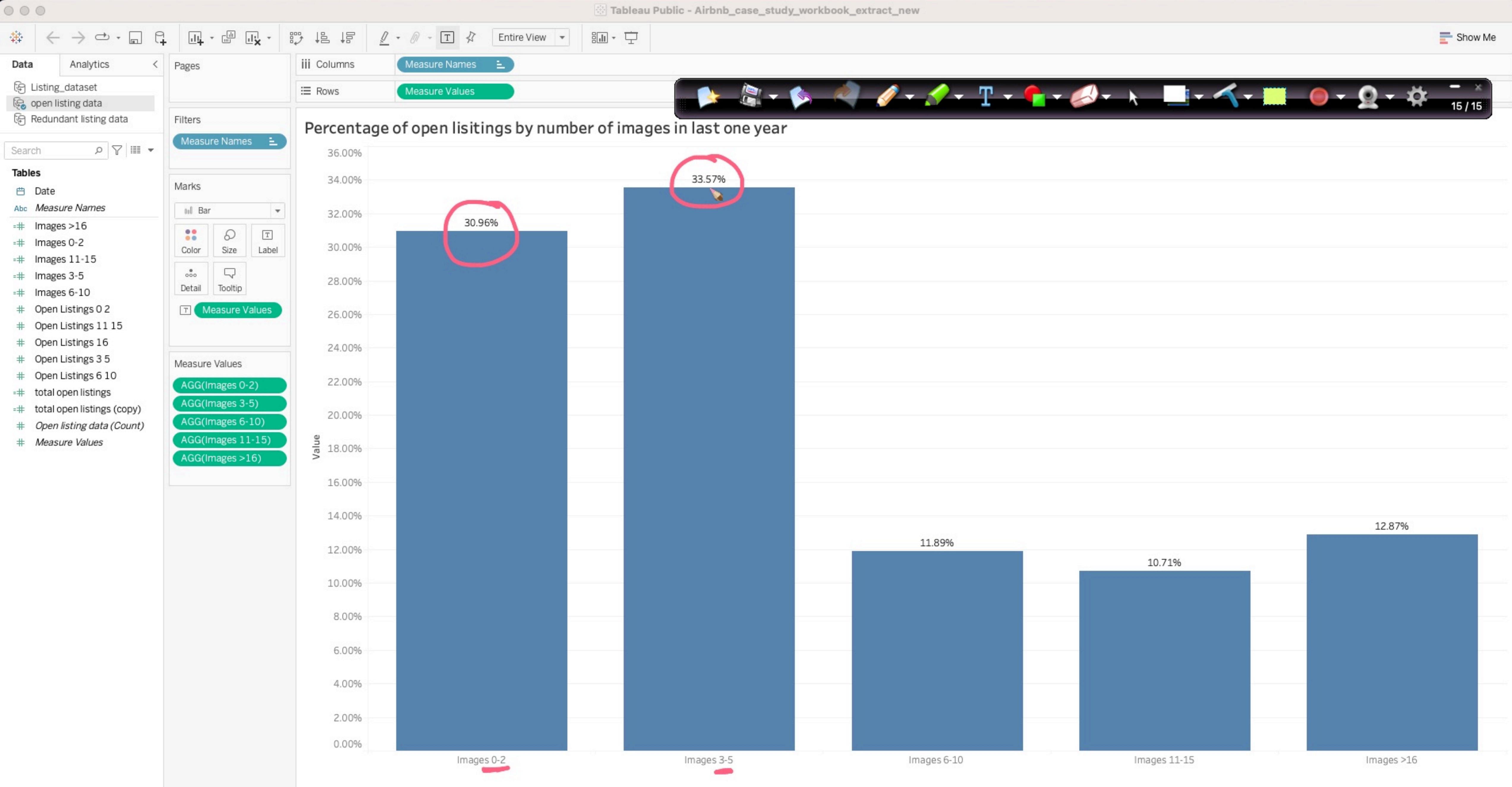
AGG(Images 0-2)
AGG(Images 3-5)
AGG(Images 6-10)
AGG(Images 11-15)
AGG(Images >16)

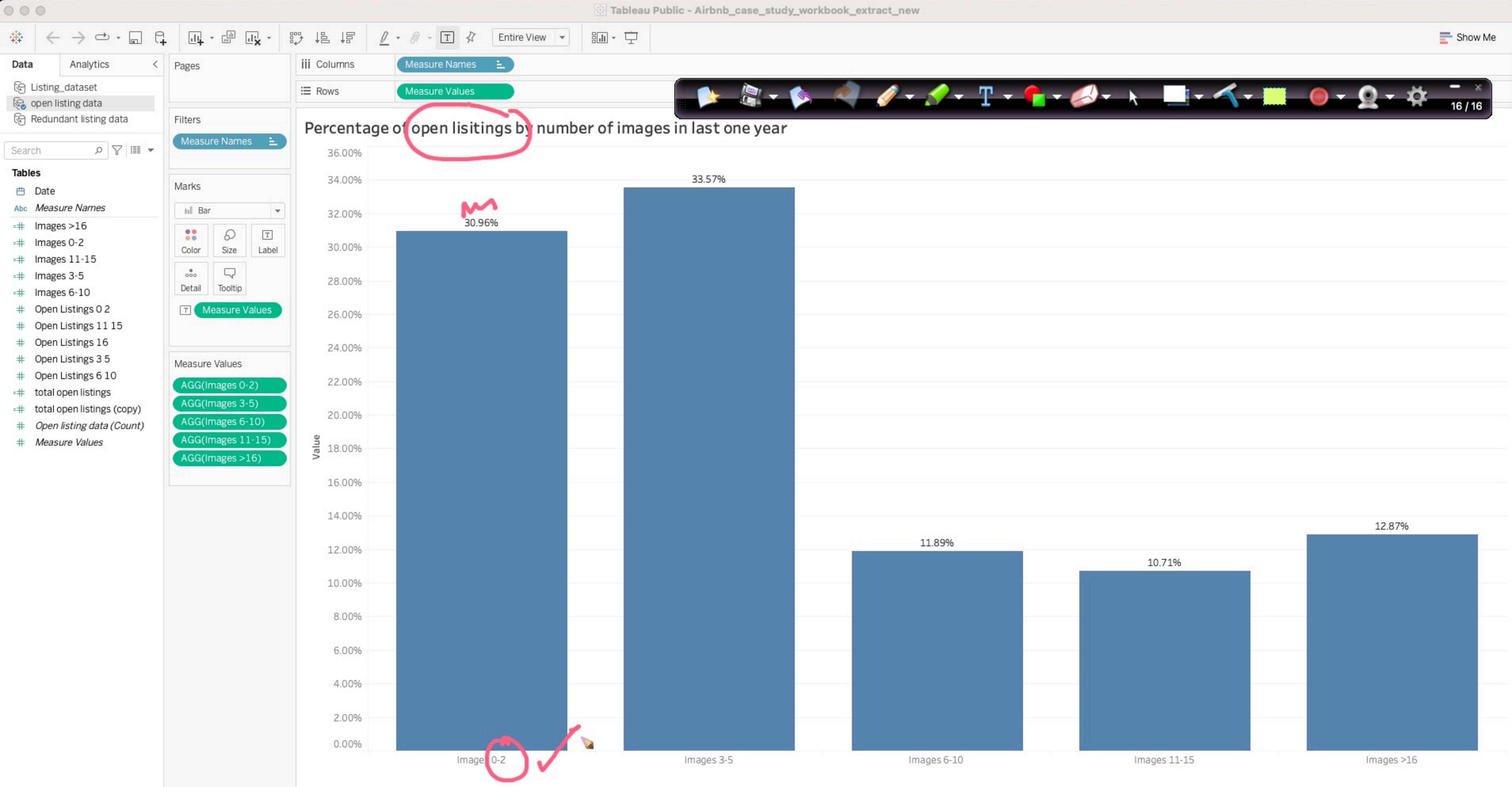
Data Source Image vs Open listing Total and redundant listing Optimal # images for regular hosr Optimal # of images for super h... Image bucket vs booking regular... Image bucket vs bookings for su... bookings vs image

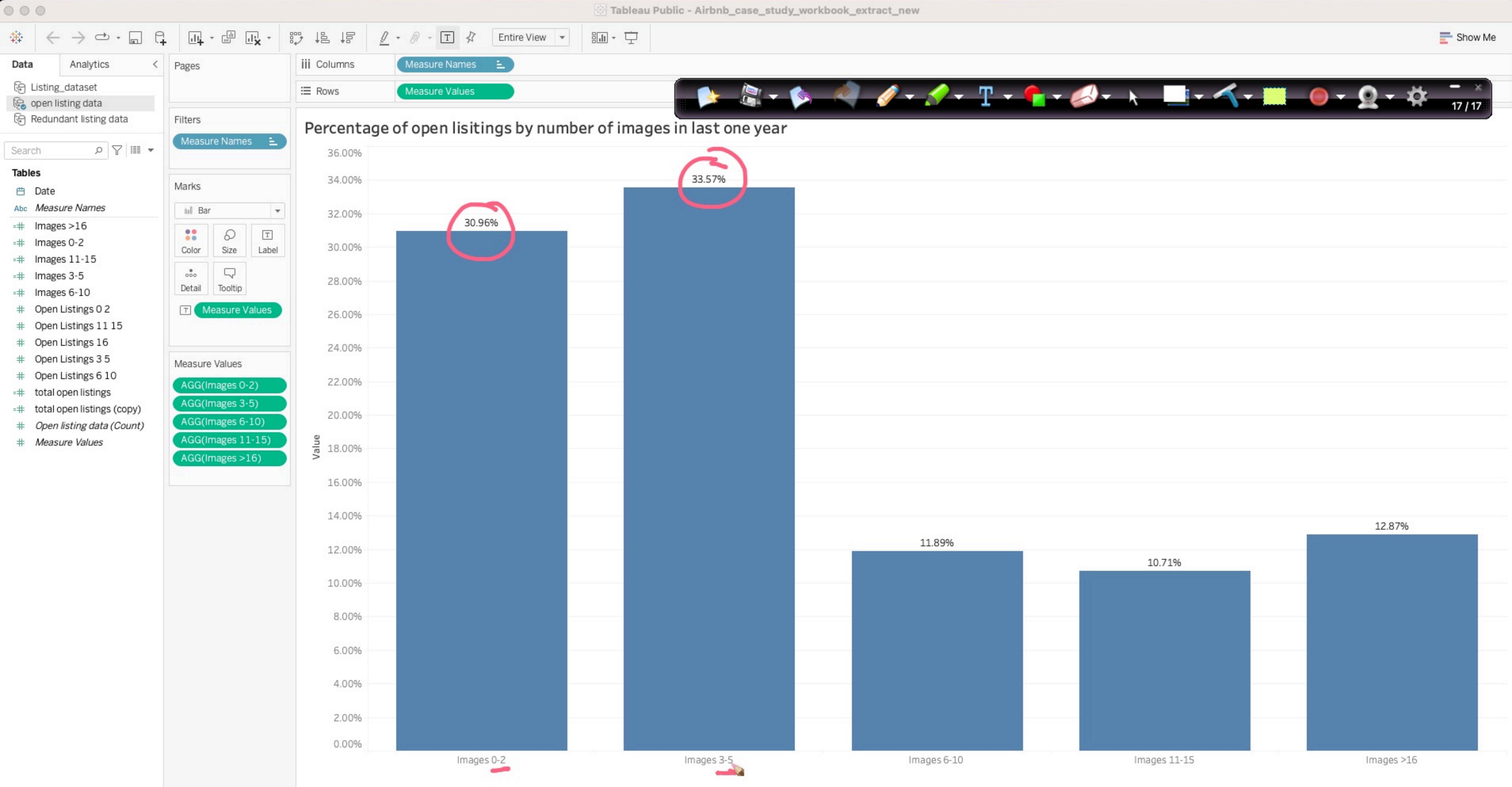
5 marks 1 row by 5 columns SUM of Measure Values: 100.00%

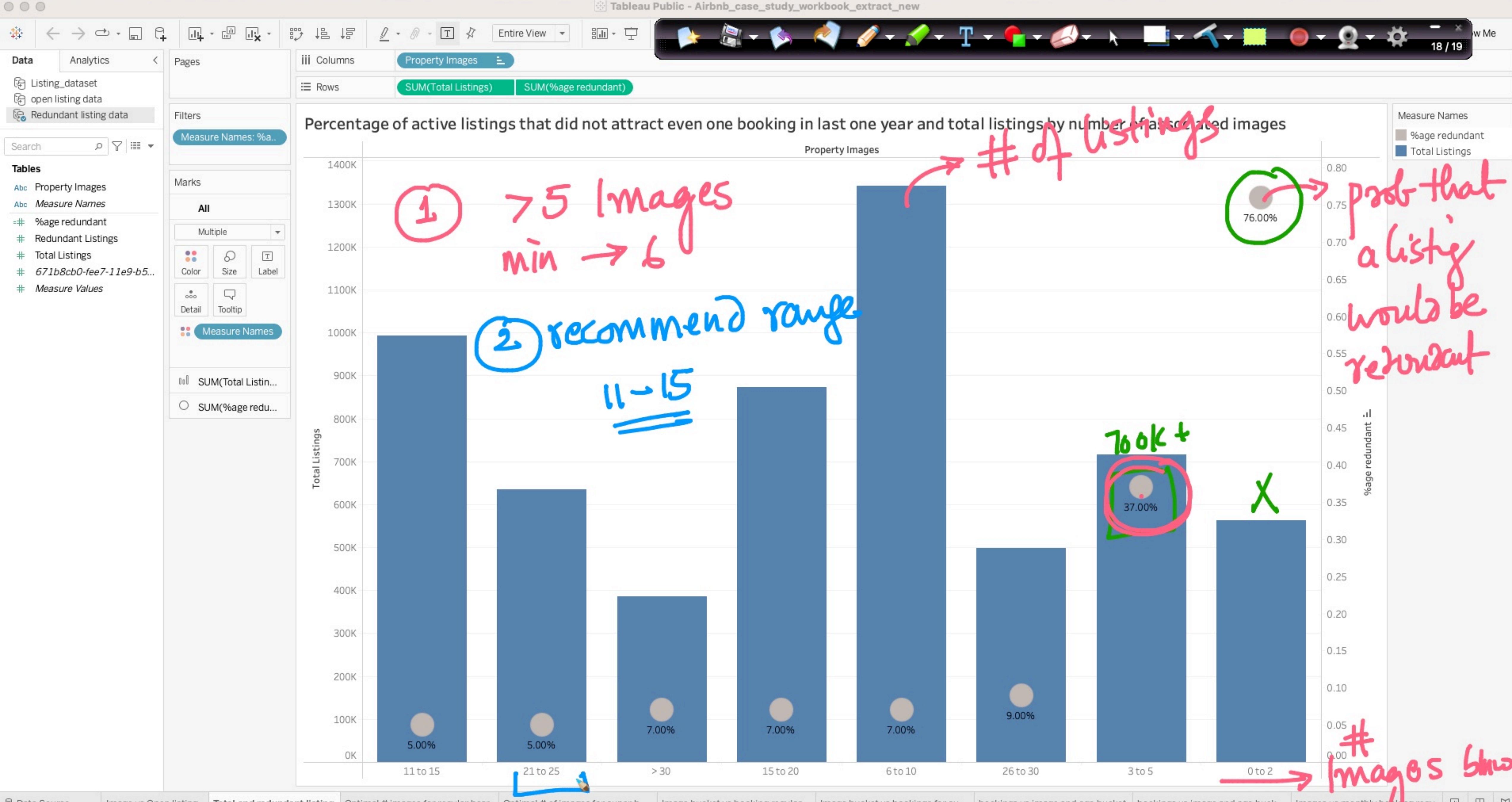
671b8cb0-fee7-11e9-b53c-ebfc1c5

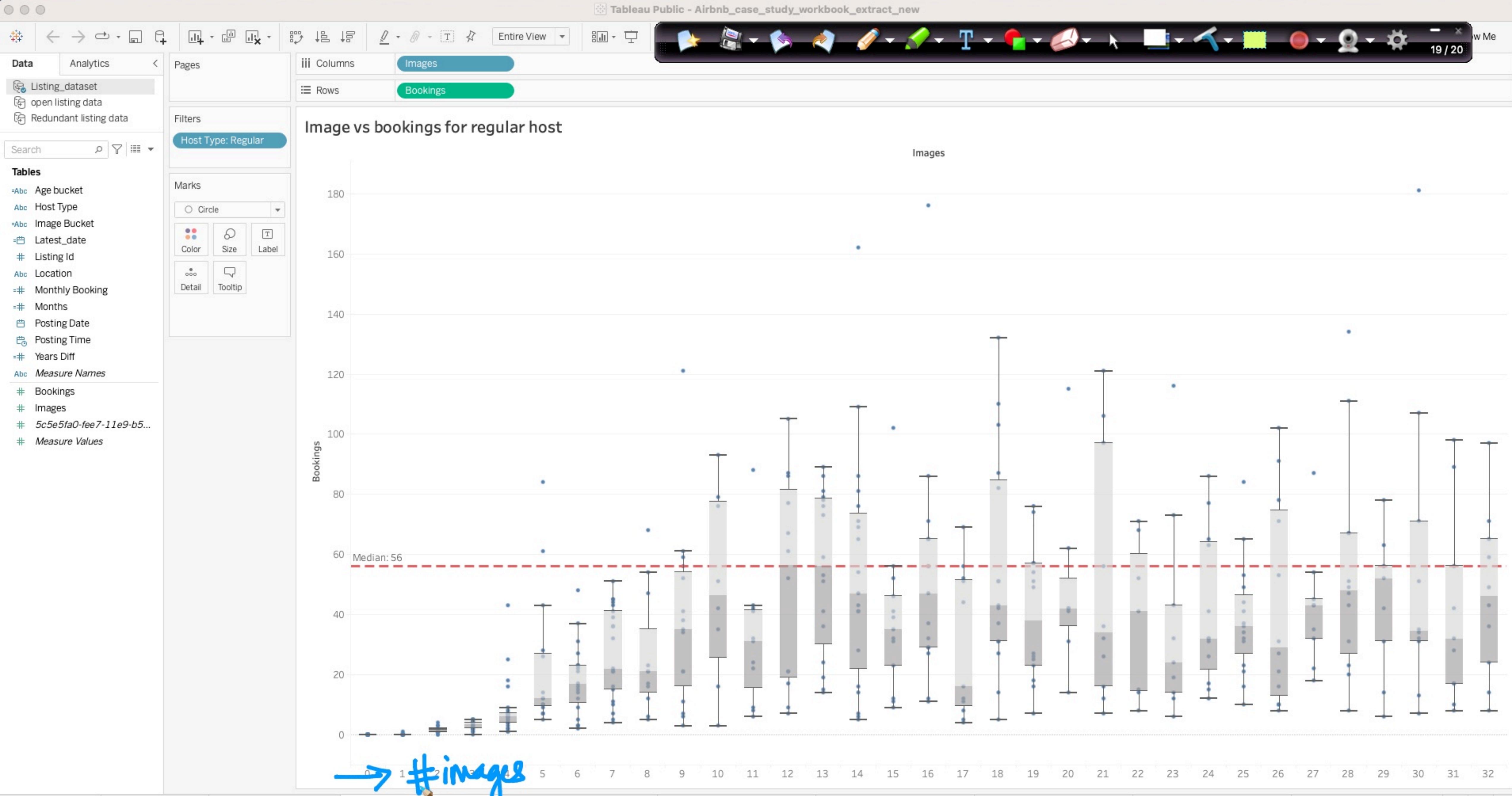
14 / 14

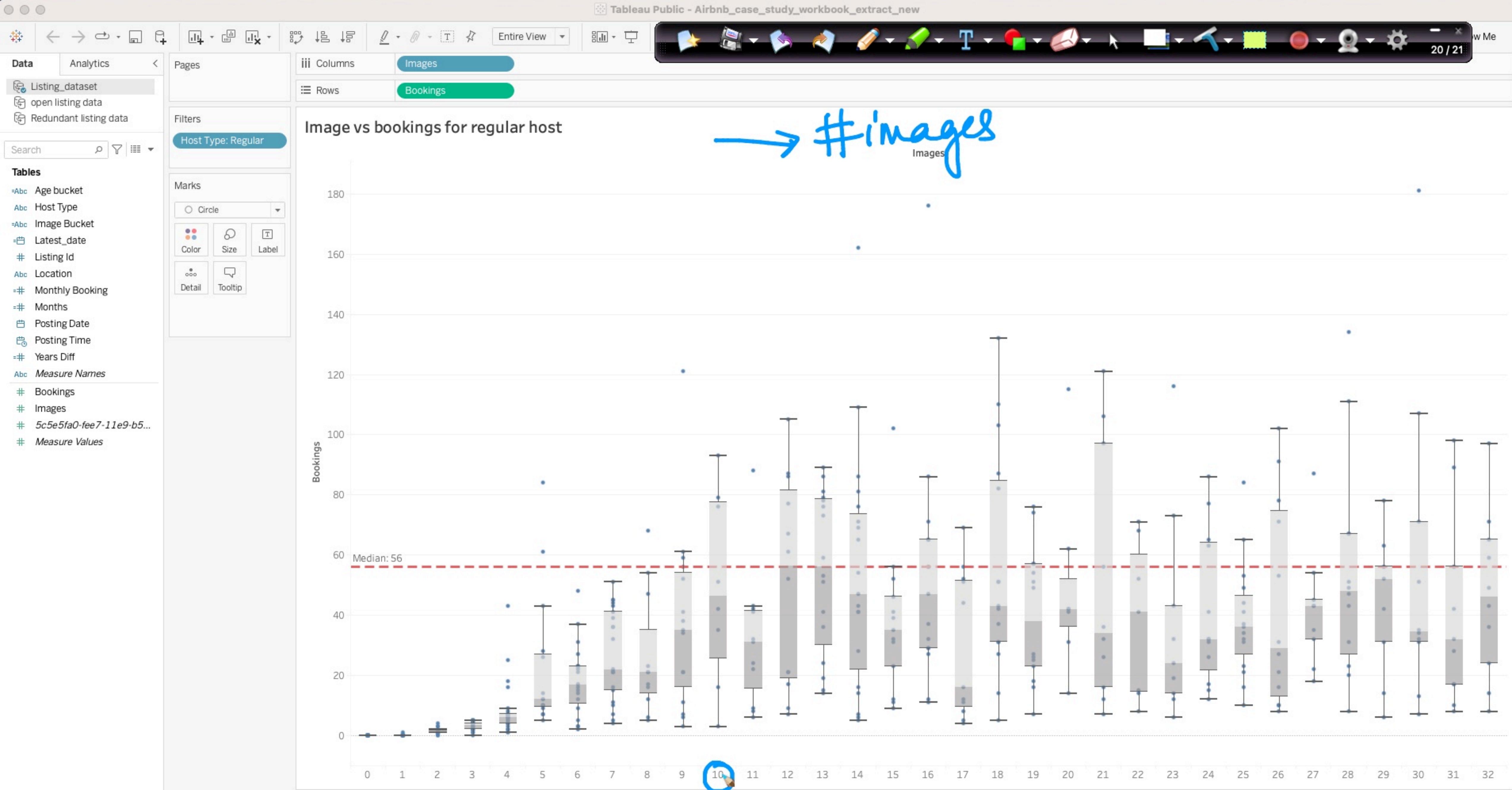




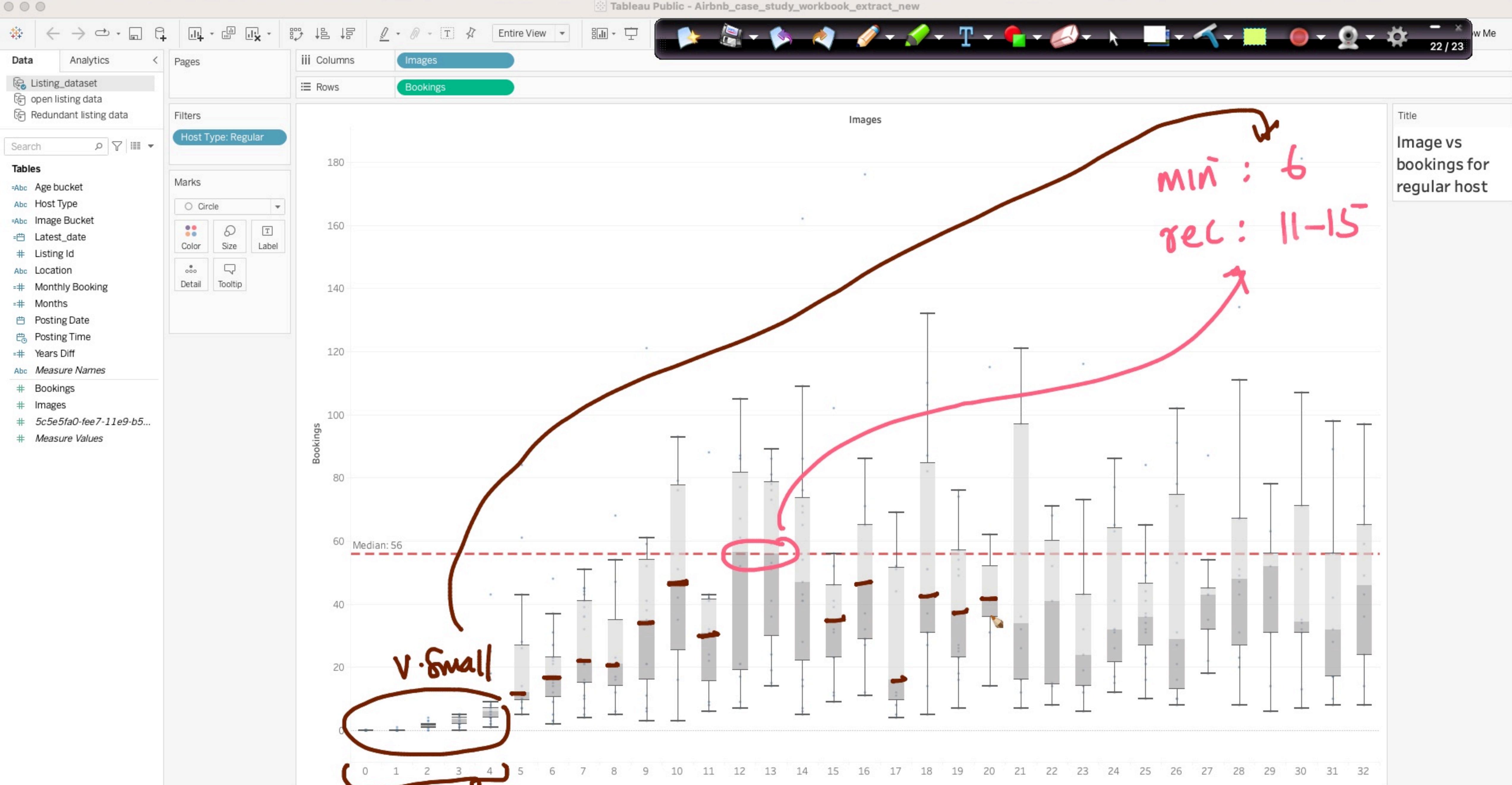




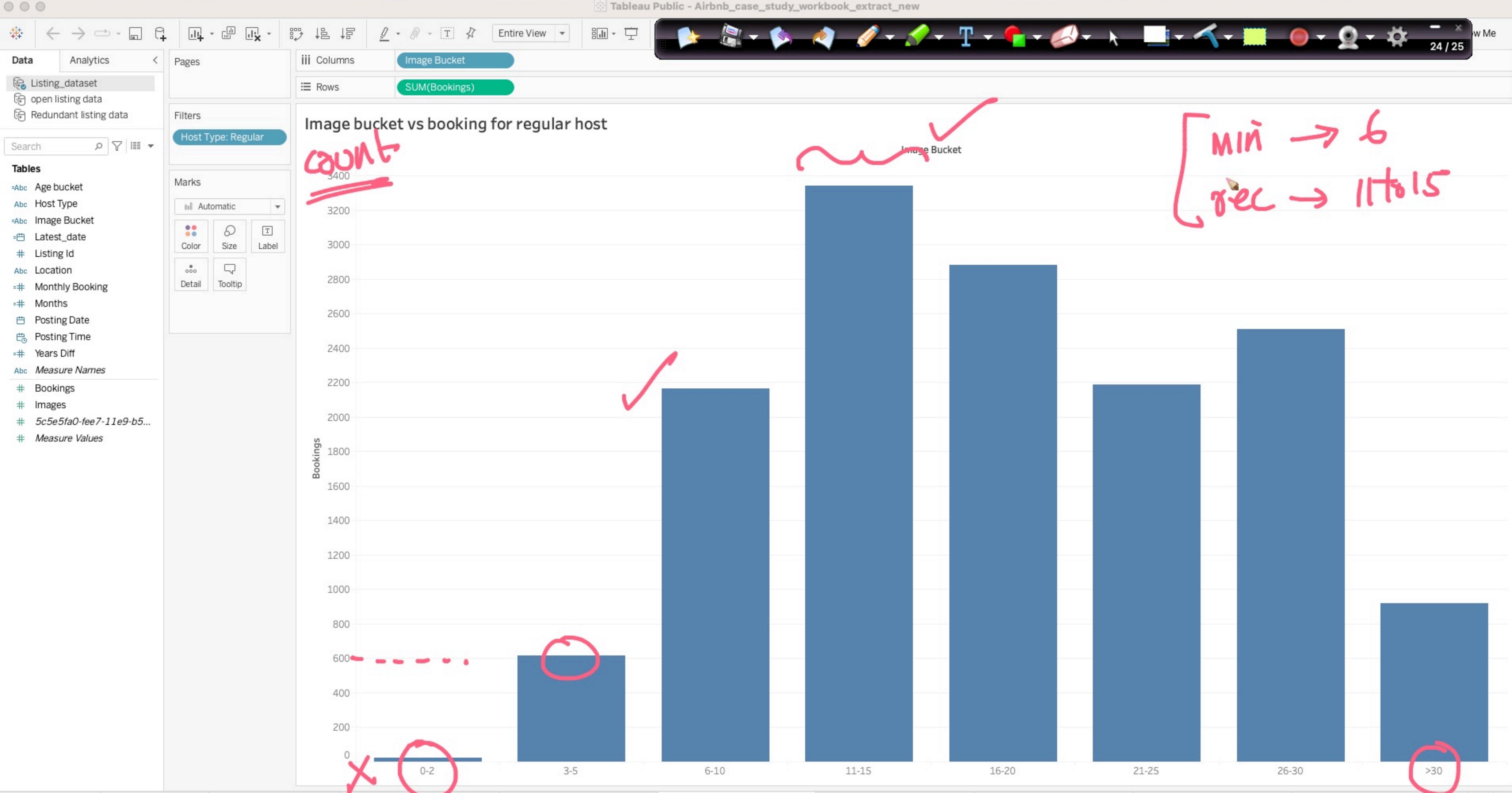












docs.google.com/spreadsheets/d/1aBrE-8JaUsig4k7eXVWpb-aGF1dVv918/edit#gid=1951333285

Dataset_1_Listings.csv XLSX

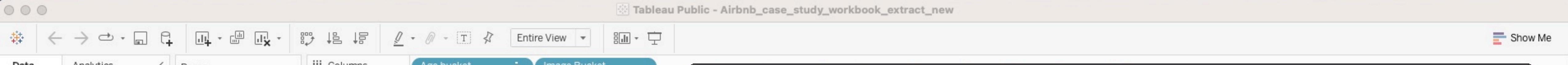
File Edit View Insert Format Data Tools Help

150% 10 Default... 123

B2 fx 13/05/2019

	A	B	C	D	E	F	G	H	I	J
1	Listing_Id	Posting_Date	Posting_Time	Location	Images	Bookings	Host_Type			
2	4955430	13/05/2019	20:18:13	San Francisco	31	10	Regular			
3	3338094	04/01/2016	5:30:37	Amsterdam	6	48	Regular			
4	3132720	24/12/2018	12:26:15	Barcelona	25	10	Regular			
6	4588846	21/07/2016	9:26:48	Brussels	4	7	Regular			
7	4677052	14/05/2016	6:40:39	Kyoto	3	5	Regular			
8	3485608	04/04/2016	10:30:15	Madrid	7	44	Regular			
9	4980709	03/04/2017	9:44:58	Tokyo	12	52	Regular			
11	4991055	09/08/2017	10:26:32	Seattle	14	109	Regular			
12	5135019	06/01/2018	22:53:12	Mexico City	11	24	Regular			
13	5040928	16/01/2015	1:40:57	Milano	12	105	Regular			
14	3297315	05/01/2017	17:56:18	Paris	21	36	Regular			
15	5793190	20/12/2018	14:11:43	Osaka	2	3	Regular			
17	3940988	28/07/2018	12:09:09	Seoul	10	51	Regular			
18	5433570	23/02/2019	9:22:07	Sao Paulo	10	16	Regular			
19	4920877	13/03/2015	2:23:16	Roma	22	71	Regular			
20	5008788	06/09/2017	19:14:52	Bengaluru	25	36	Regular			
21	5549595	20/06/2019	3:43:12	New Delhi	15	12	Regular			
22	4550407	28/02/2017	11:42:20	Chennai	15	32	Regular			
24	5645148	06/02/2016	3:04:32	Paris	3	2	Regular			
25	4757178	28/09/2016	16:23:09	San Jose	6	31	Regular			
26	4815422	13/06/2017	12:21:33	Amsterdam	16	65	Regular			
27	5112463	01/02/2017	17:55:46	Barcelona	17	44	Regular			

Avg 31st 2019
~~2018~~



Data Analytics <

Listing_dataset
open listing data
Redundant listing data

Search

Tables

- =Abc Age bucket
- =Abc Host Type
- =Abc Image Bucket
- Abc Latest_date
- # Listing Id
- Abc Location
- # Monthly Booking
- # Months
- Abc Posting Date
- Abc Posting Time
- # Years Diff
- Abc Measure Names
- # Bookings
- # Images
- # 5c5e5fa0-fee7-11e9-b5...
Measure Values





Tableau Public - Airbnb_case_study_workbook_extract_new

Entire View | Rows: SUM(Bookings) | Columns: Age bucket, Image Bucket

Pages: 28 / 29

Data Analytics < Listing_dataset open listing data Redundant listing data

Search

Tables

- =Abc Age bucket
- =Abc Host Type
- =Abc Image Bucket
- Abc Latest_date
- # Listing Id
- Abc Location
- # Monthly Booking
- # Months
- Posting Date
- Posting Time
- # Years Diff
- Abc Measure Names
- # Bookings
- # Images
- # 5c5e5fa0-fee7-11e9-b5...
- # Measure Values

Booking vs image and age bucket for super host

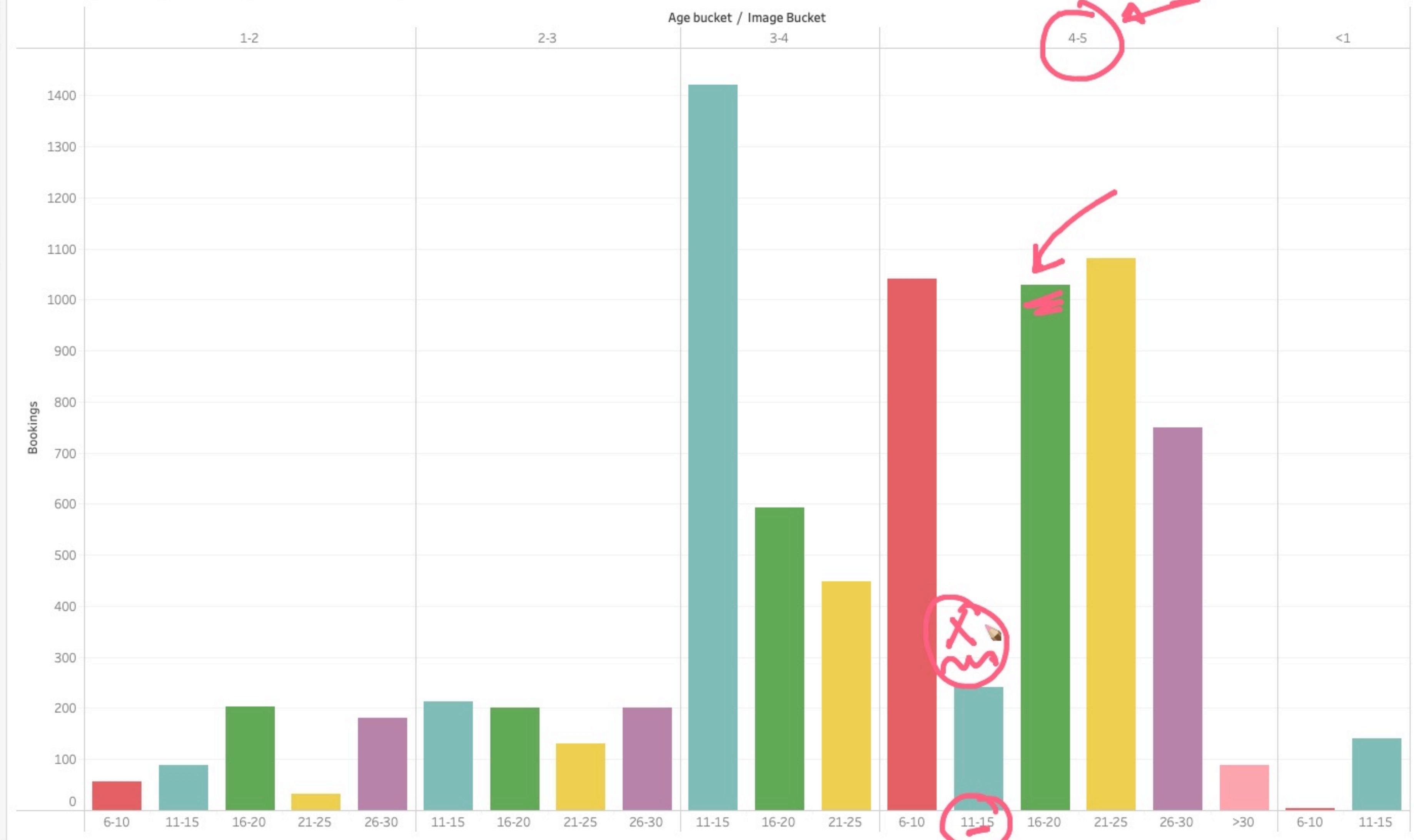


Image Bucket

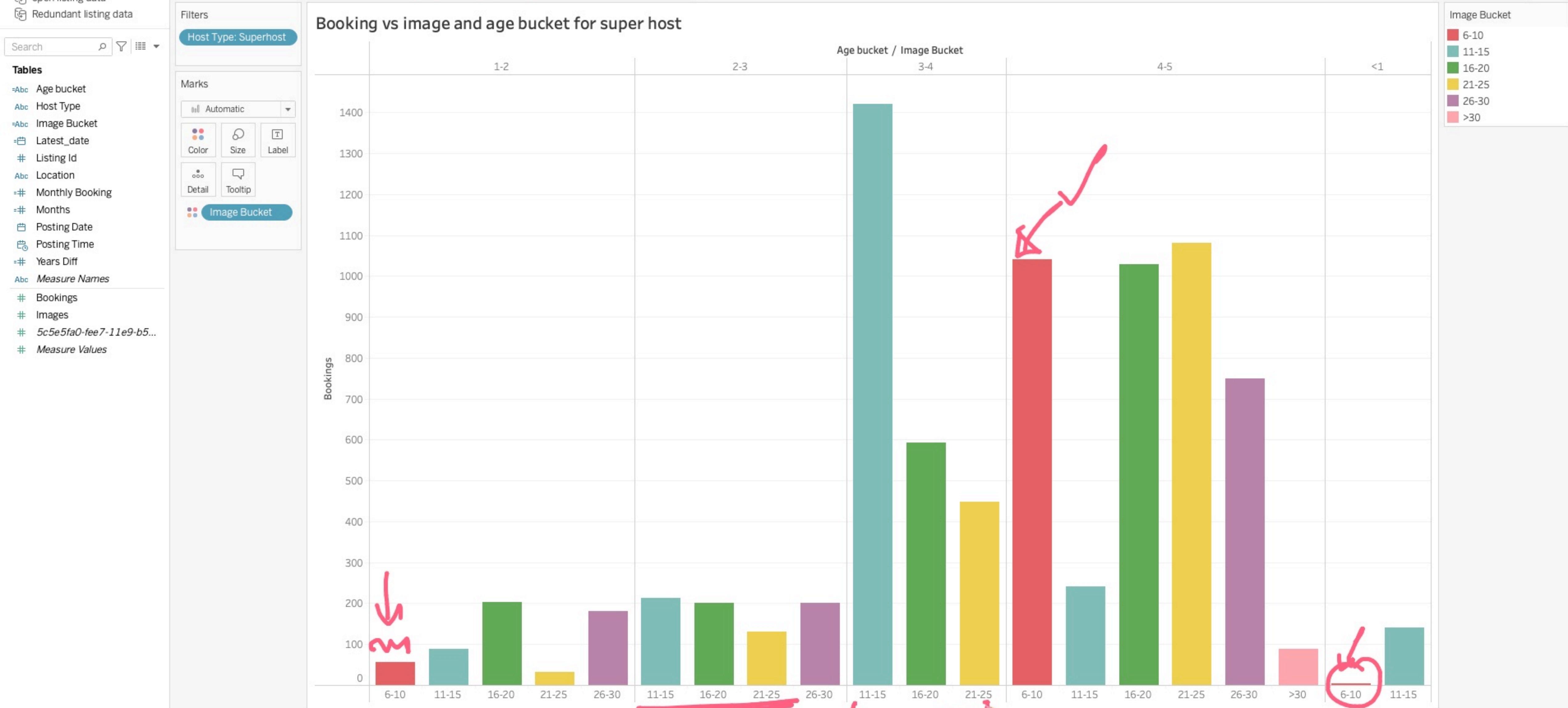
- 6-10
- 11-15
- 16-20
- 21-25
- 26-30
- >30

Tableau Public - Airbnb_case_study_workbook_extract_new

Entire View

Pages: Columns: Age bucket, Image Bucket
Rows: SUM(Bookings)

Show Me



docs.google.com/spreadsheets/d/1aBrE-8JaUsig4k7eXVWpb-aGF1dVv918/edit#gid=1951333285

Dataset_1_Listings.csv XLSX

File Edit View Insert Format Data Tools Help

150% 10 B I A F5 fx 134

	A	B	C	D	E	F	G	H	I	J
1	Listing_Id	Posting_Date	Posting_Time	Location	Images	Bookings	Host_Type			
2	4955430	13/05/2019		20:18:13	San Francisco	31	10	Regular		
3	3338094		04/01/2016		5:30:37	Amsterdam	6	48	Regular	
4	3132720	24/12/2018		12:26:15	Barcelona	25	10	Regular		
5	3196344	14/10/2014		15:09:27	Berlin	10	134	Superhost	59 Mmlhs	
6	4588846	21/07/2016		9:26:48	Brussels	4	7	Regular		
7	4677052	14/05/2016		6:40:39	Kyoto	3	5	Regular		
8	3485608		04/04/2016	10:30:15	Madrid	7	44	Regular		
9	4980709	03/04/2017		9:44:58	Tokyo	12	52	Regular		
10	4761153	14/01/2018		15:44:54	Los Angeles	28	81	Superhost		
11	4991055		09/08/2017	10:26:32	Seattle	14	109	Regular		
12	5135019	06/01/2018		22:53:12	Mexico City	11	24	Regular		
13	5040928	16/01/2015		1:40:57	Milano	12	105	Regular		
14	3297315		05/01/2017	17:56:18	Paris	21	36	Regular		
15	5793190	20/12/2018		14:11:43	Osaka	2	3	Regular		
16	4435963	17/10/2015		1:14:07	Rio	23	234	Superhost		
17	3940988	28/07/2018		12:09:09	Seoul	10	51	Regular		
18	5433570	23/02/2019		9:22:07	Sao Paulo	10	16	Regular		
19	4920877	13/03/2015		2:23:16	Roma	22	71	Regular		
20	5008788		06/09/2017	19:14:52	Bengaluru	25	36	Regular		
21	5549595	20/06/2019		3:43:12	New Delhi	15	12	Regular		
22	4550407	28/02/2017		11:42:20	Chennai	15	32	Regular		
23	5228587	28/02/2017		19:40:17	Berlin	14	45	Superhost		

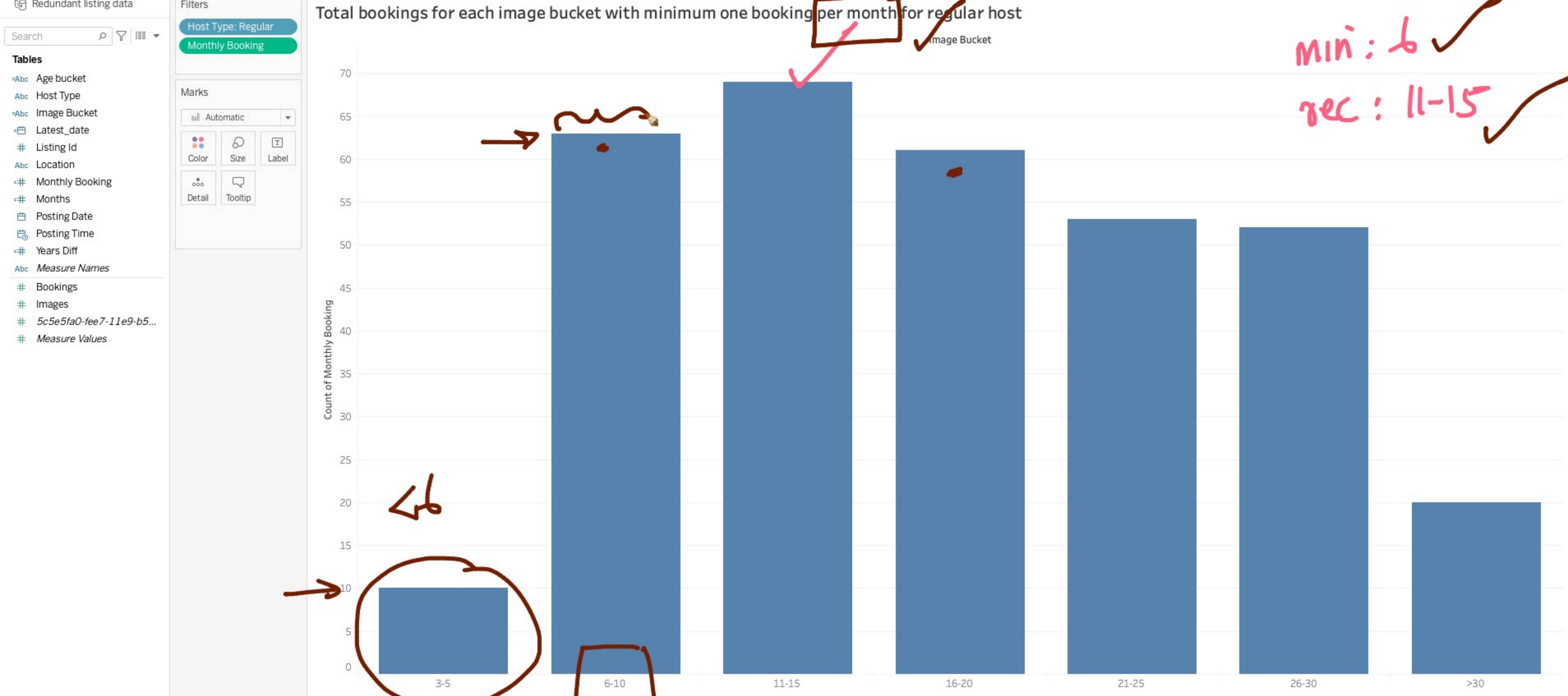
134
—
59 =

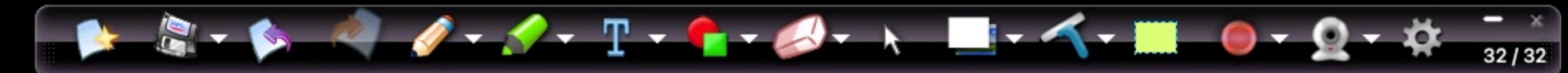
Tableau Public - Airbnb_case_study_workbook_extract_new

Entire View 31 / 31

Data Analytics Listing_dataset open listing data Redundant listing data

Pages Columns Image Bucket Rows CNT(Monthly Booking)



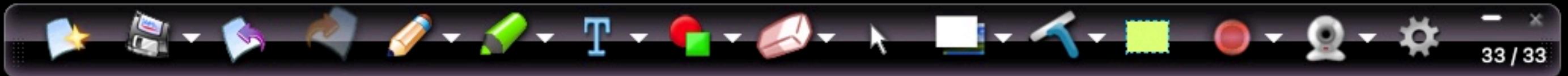


General

=

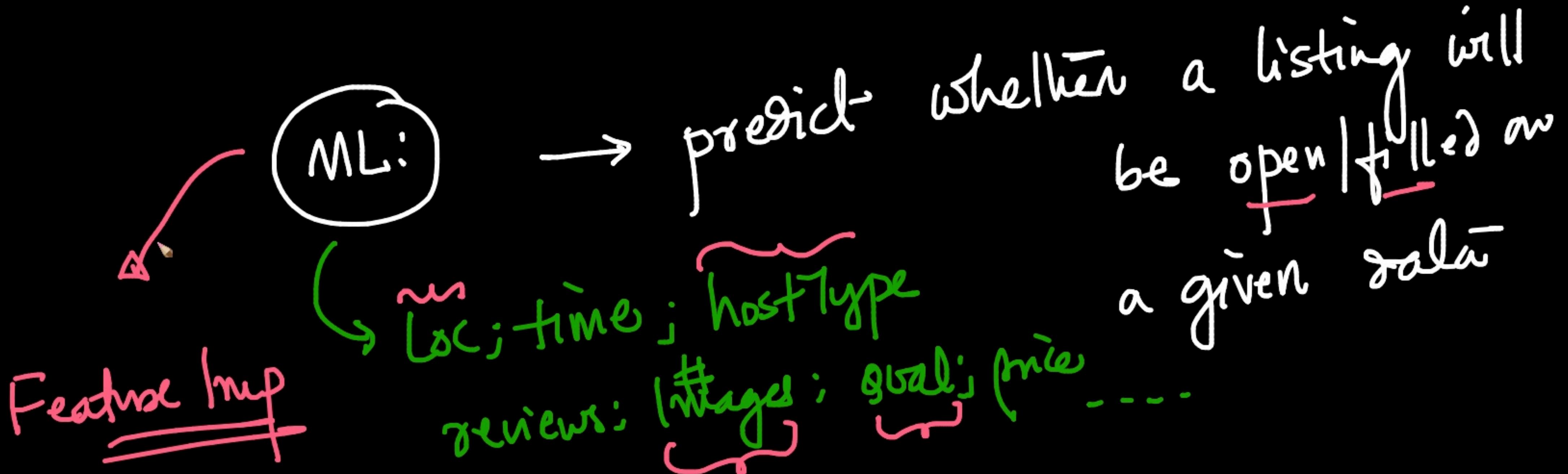
{ min → 6
rec.range → 11 to 15 } ✓

slice → age of posting
→ Superhost vs Regular



~~Key assumptions:~~

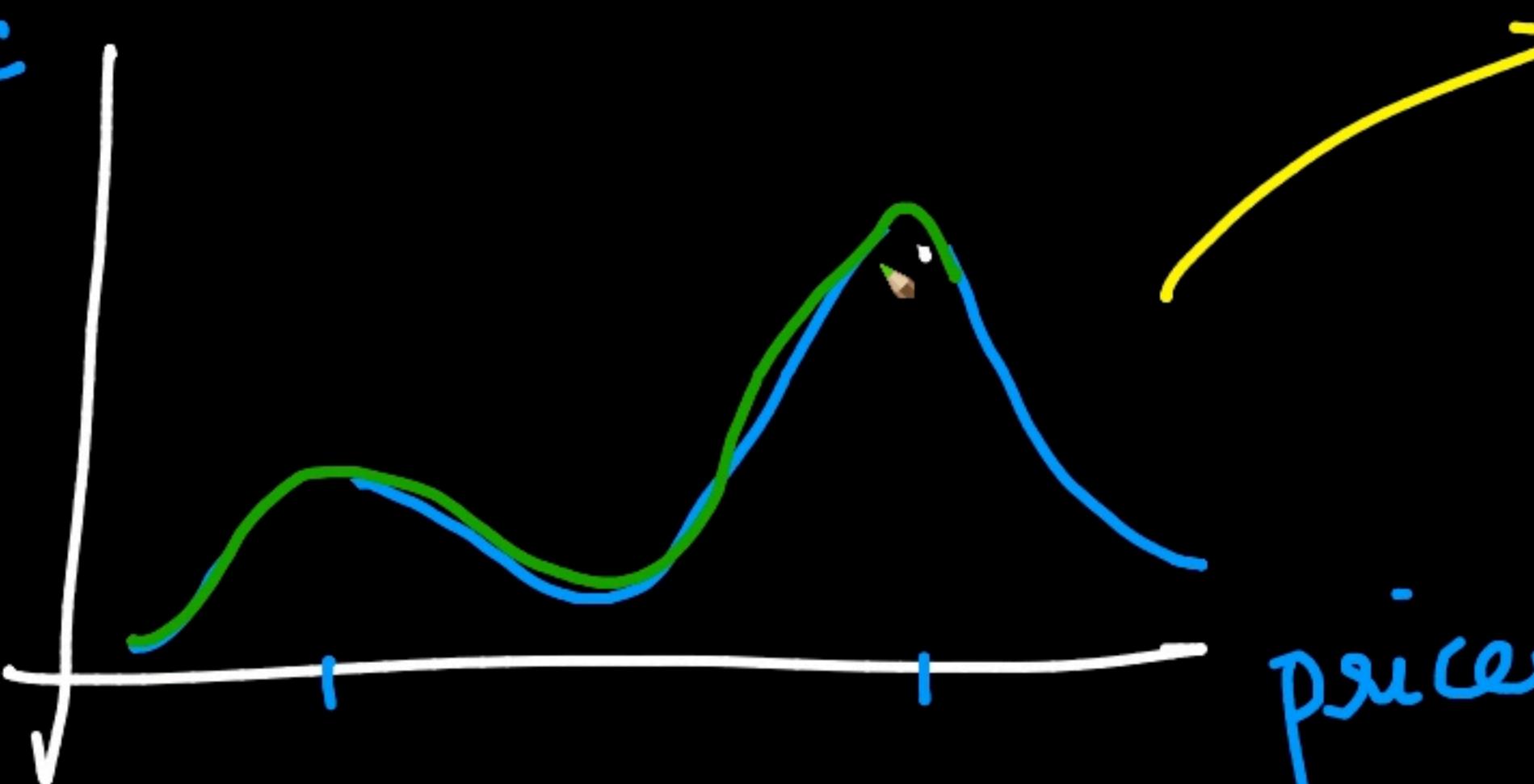
1. hosts & super host behave similarly
one of many factors
2. **Images:**



flight overbooking

①

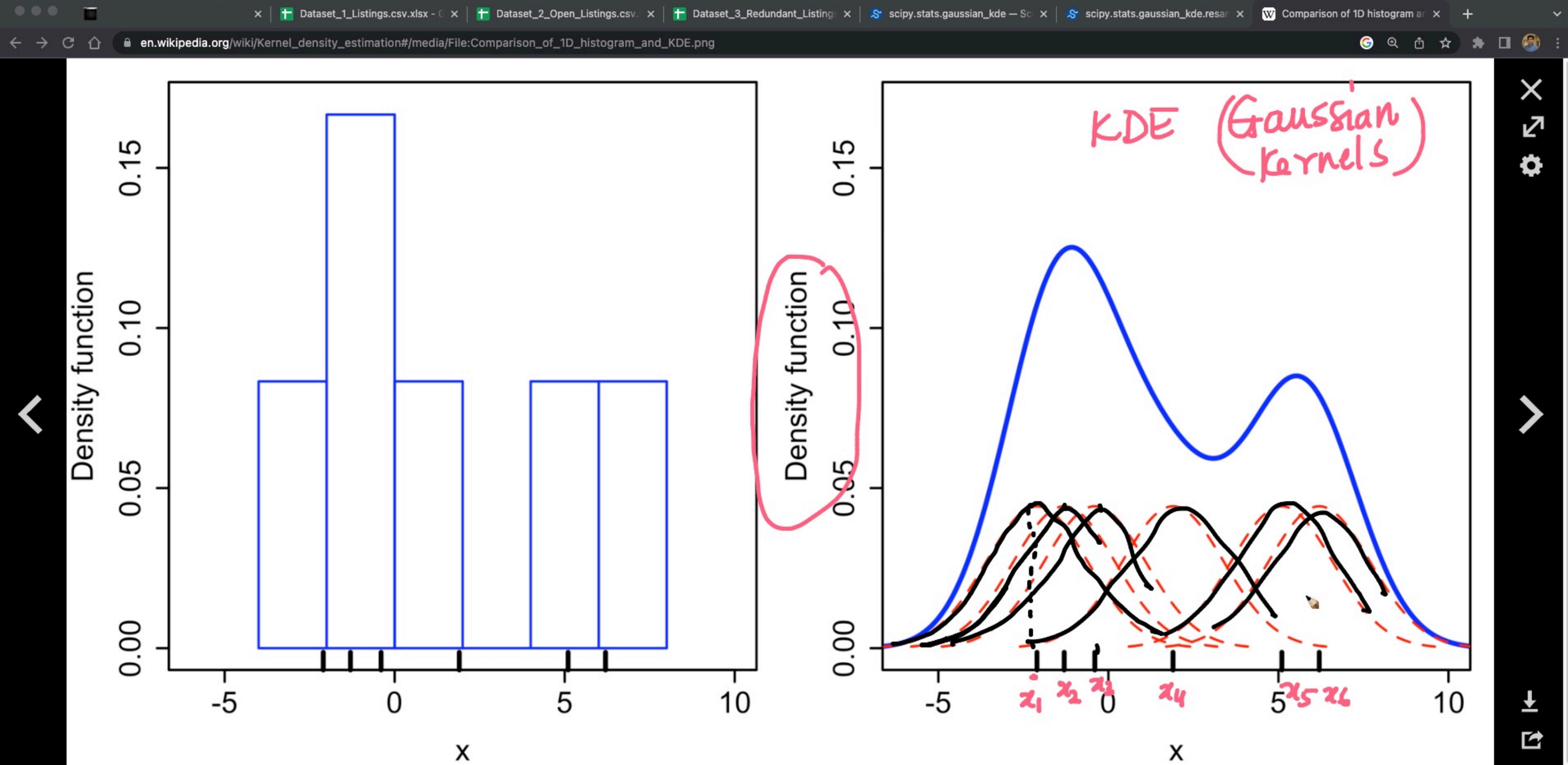
PDF



empirical disb

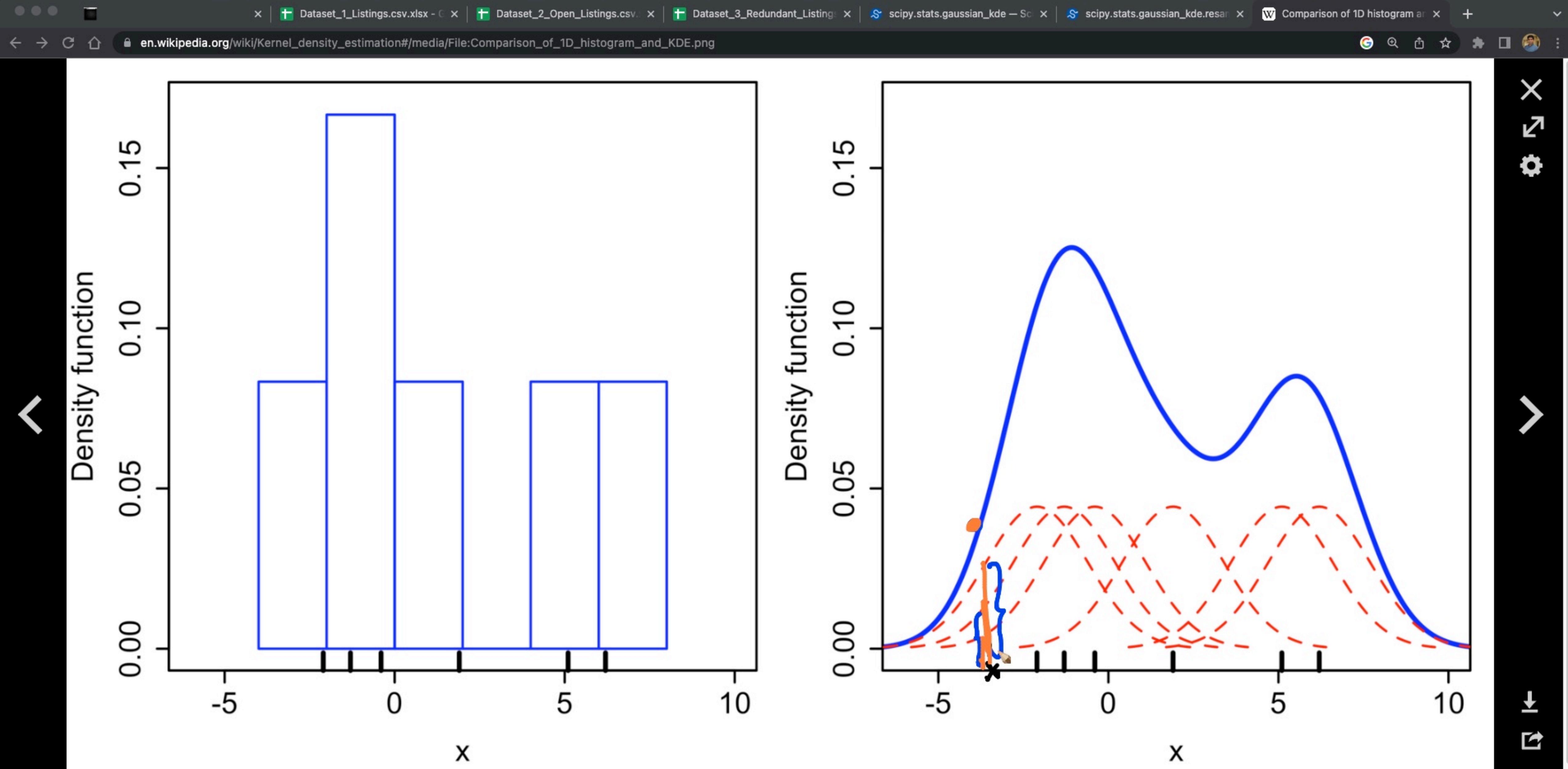
sample
1000 values





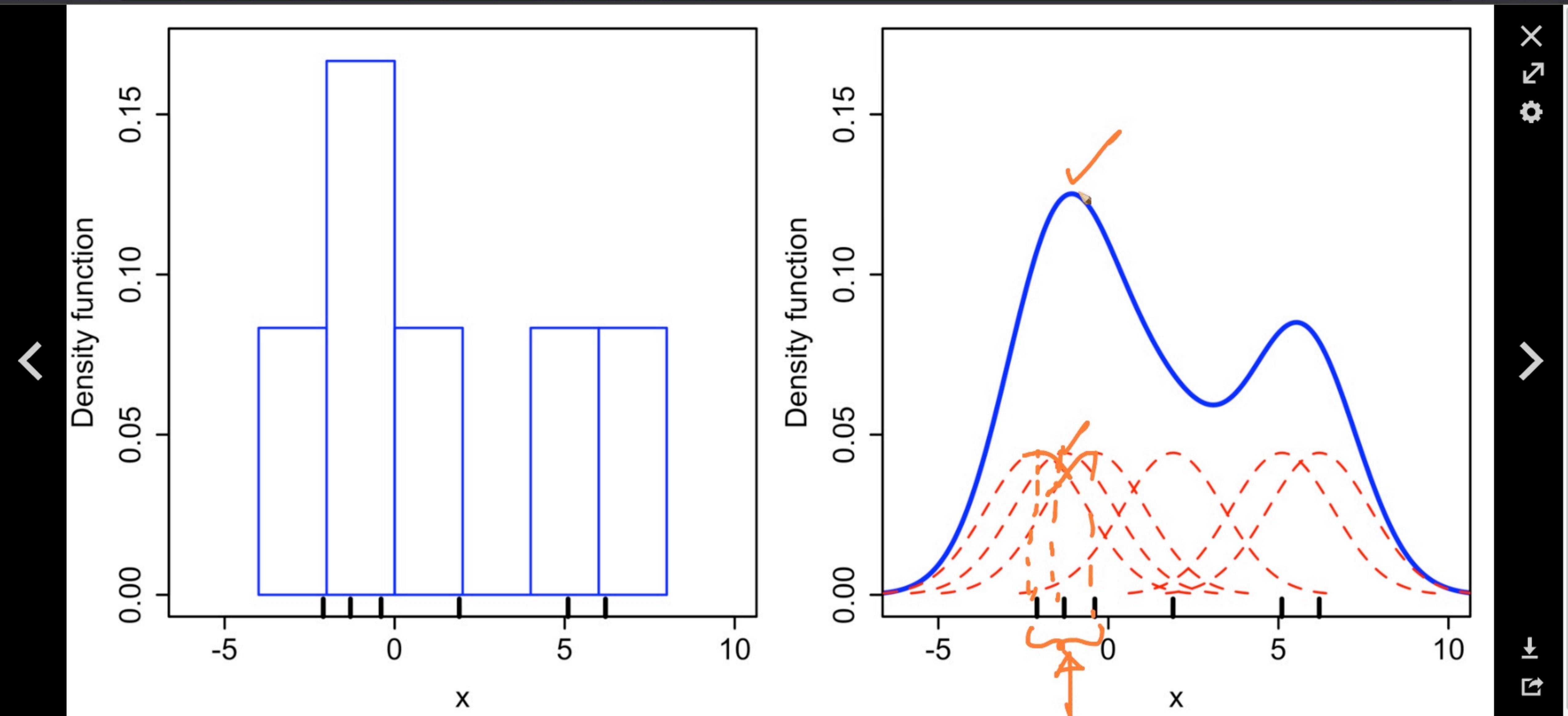
Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The six individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot.

More details



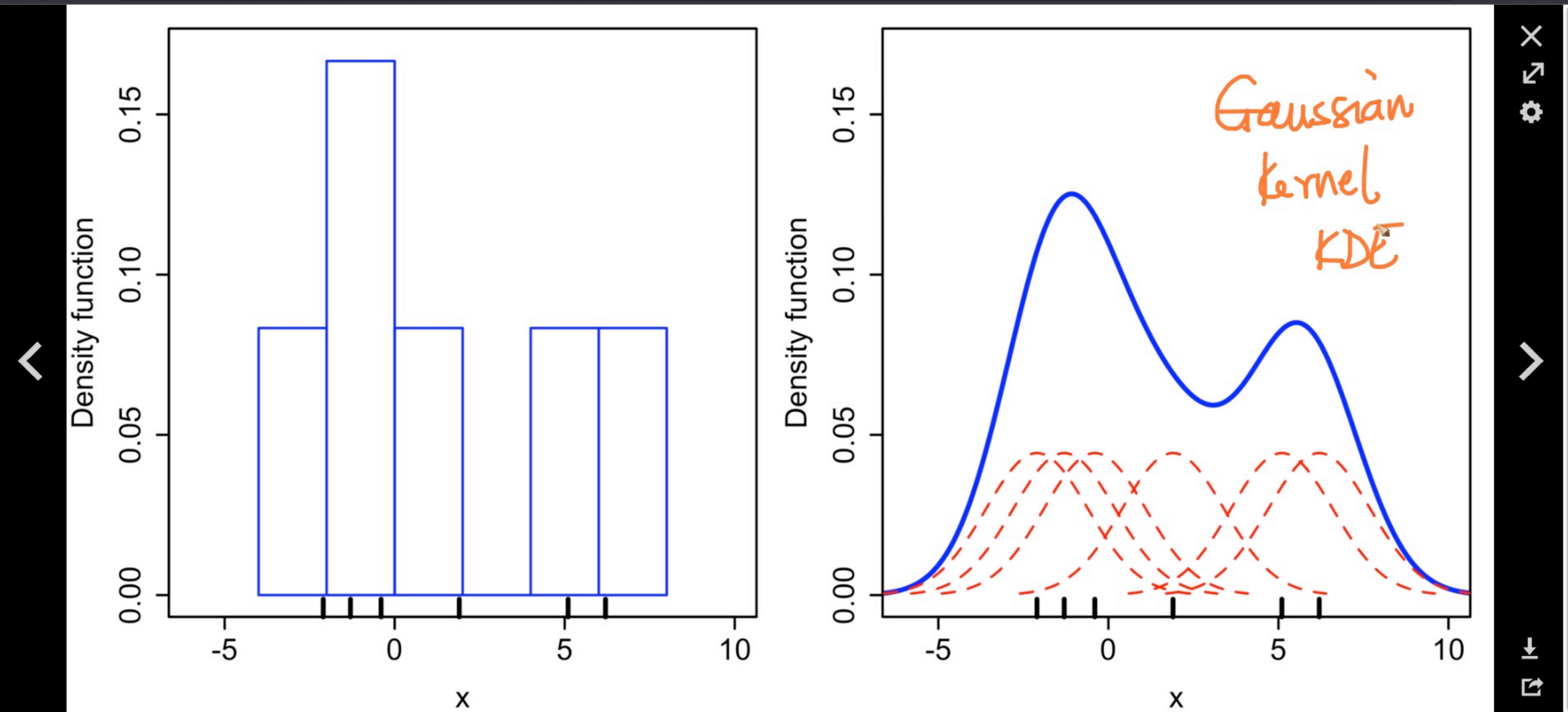
Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The six individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot.

[More details](#)



Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The six individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot.

More details



Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The six individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot.

More details

User Seg in flight booking

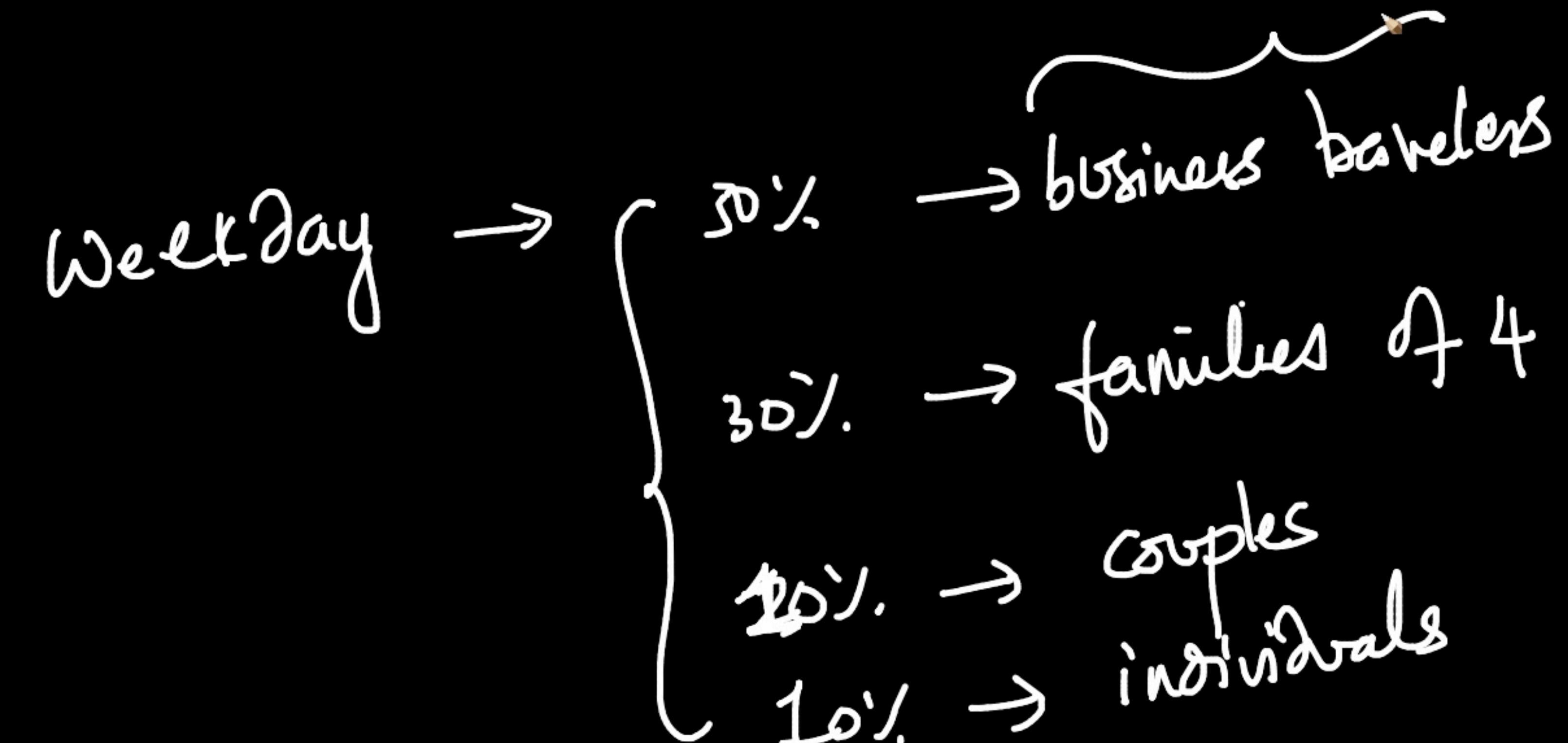


Tableau Public - Airbnb_case_study_workbook_extract_new

Entire View | Show Me

Pages: **Image Bucket**

Rows: **CNT(Monthly Booking)**

Filters: **Host Type: Superhost**, **Monthly Booking**

Tables:

- Abc Age bucket
- Abc Host Type
- =Abc Image Bucket
- Abc Latest_date
- # Listing Id
- Abc Location
- # Monthly Booking
- # Months
- Posting Date
- Posting Time
- # Years Diff
- Abc Measure Names
- # Bookings
- # Images
- # 5c5e5fa0-fee7-11e9-b5...
- # Measure Values

Count of Monthly Booking

Image Bucket

6-10, 11-15, 16-20, 21-25, 26-30, >30

Monthly bookings for each image bucket with minimum one booking per month for Super host

Image Bucket	Count of Monthly Booking
6-10	7
11-15	11
16-20	14
21-25	14
26-30	7
>30	1

6 marks 1 row by 6 columns SUM of CNT(Monthly Booking): 54

41 / 41