Airbnb
A Tourist Guide, Boston MA
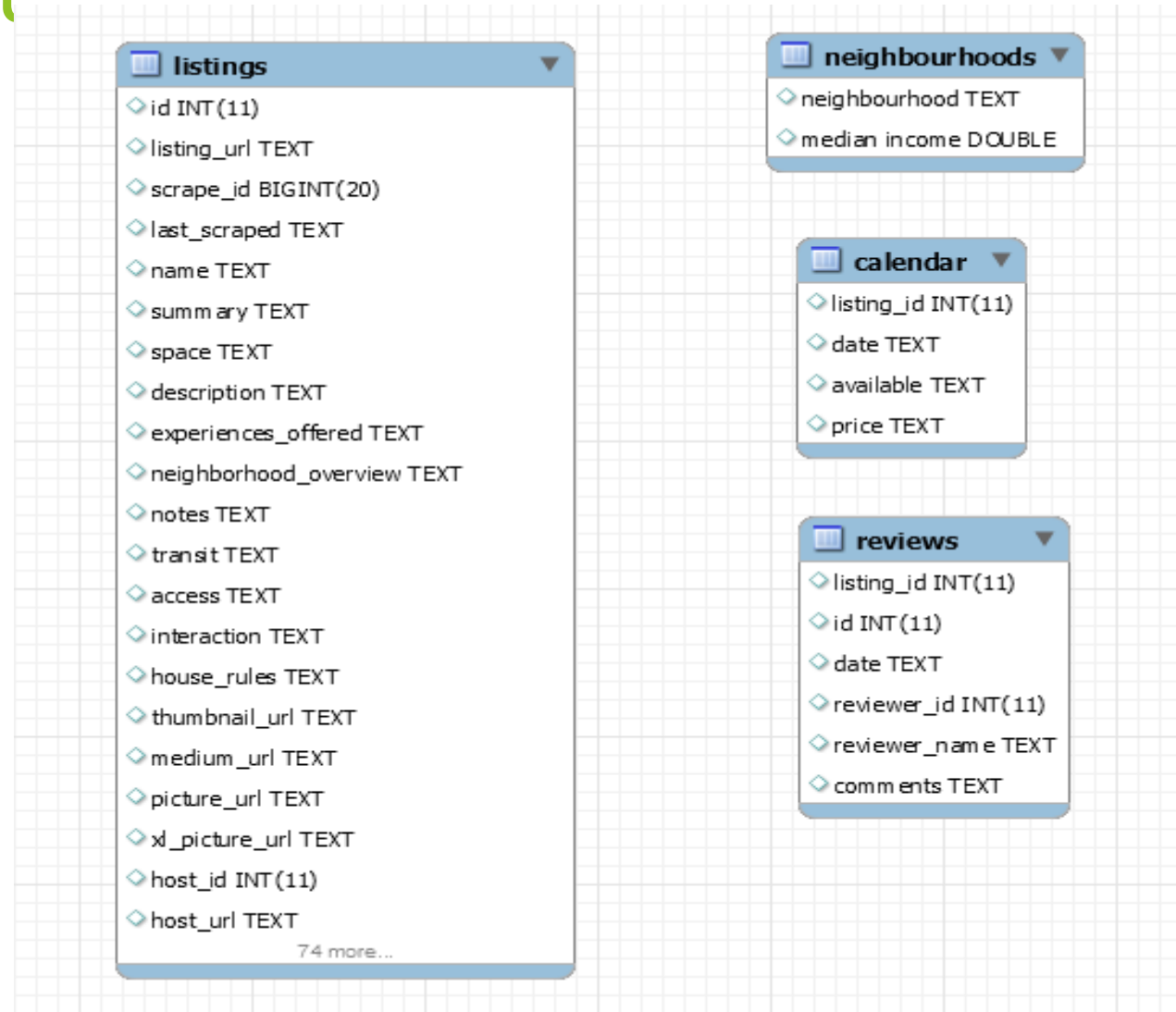
Exploring the neighbourhoods in and around Boston

Does the average pricing for each of the neighbourhood differ based on the demographics?
If, so which of them top the chart!!

Does the holiday or the festive season play a role in the pricing of the accommodations?
If yes, what is the maximum price increase one can expect throughout the year?

# What is the average property pricing for each of the neighbourhood?

▶ select distinct
neighbourhood,room_type,avg(price) as
avg_price,max(reviews_per_month) as reviews_per_month
from listings where
neighbourhood in (select distinct
neighbourhood from neighbourhoods)
group by neighbourhood;

neighbourhood	room_type	avg_price	reviews_per_month
Downtown	Entire home/apt	248.0969	10.56
Back Bay	Entire home/apt	229.7339	13.24
Charlestown	Private room	209.3109	11.64
Beacon Hill	Private room	201.9064	10.78
South Boston	Private room	195.7137	8.9
Fenway	Private room	190.3735	9.47
North End	Private room	183.9326	12.1
Jamaica Plain	Private room	140.2639	11
East Boston	Entire home/apt	112.28	15.1
West Roxbury	Private room	104.7818	9.59
Mission Hill	Private room	104.1039	10.66
Brighton	Private room	101.7265	11.81
Mattapan	Private room	100.697	6.1
Roslindale	Private room	97.3133	7.84
Allston	Private room	96.351	11.72
Dorchester	Private room	93.3571	10.68
Hyde Park	Private room	80.3684	10.29

Boston Neighbourhood has a wide ranges of accommodations!

A wide variety of accommodations are available!

A wide variety of accommodations are available!!
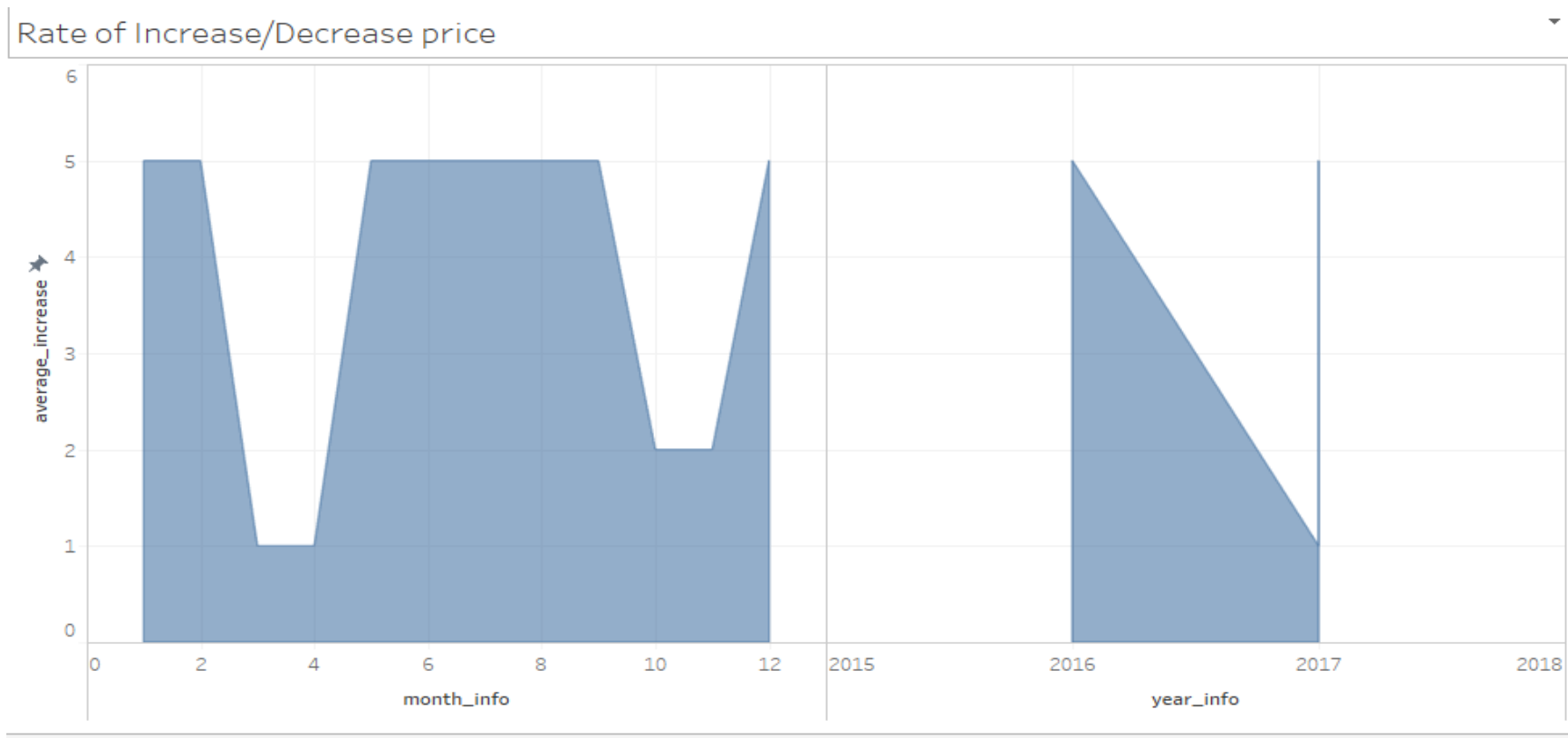
neighbourhood
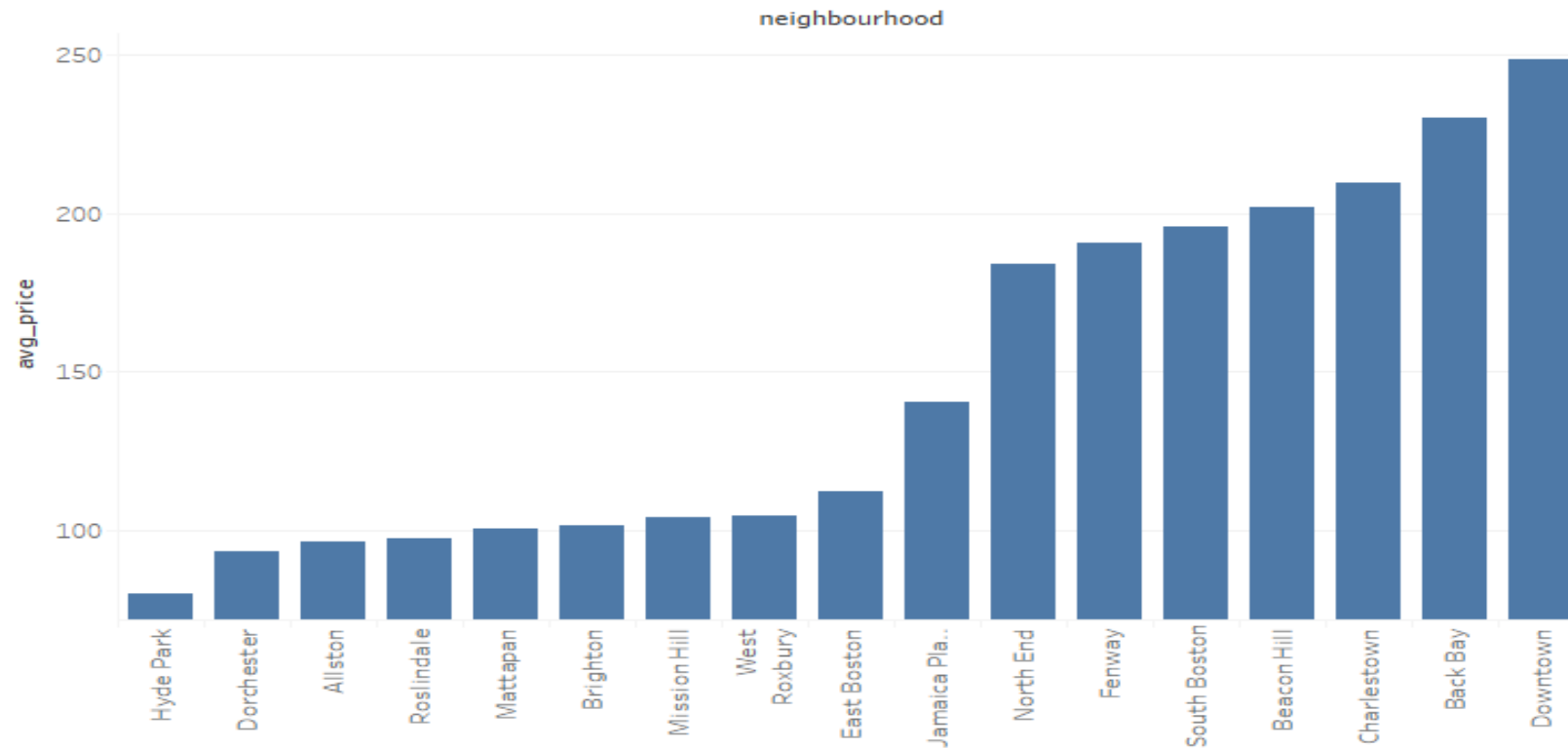
What is the rate of increase/decrease in rental price over a year?

▶ select extract(year from cast(date as dateTime)) as year_info, extract(month from cast(date as dateTime)) as month_info,
    convert((min(price)-max(price)/min(price)*100) , decimal(7,2)) as average_increasefrom calendar
    where available = 't' and listing_id in (select id from listings group by neighbourhood) group by month_info order by month_info desc,average_increase;

year_info	month_info	average_increase
2016	12	5
2016	11	2
2016	10	2
2017	9	5
2017	8	5
2017	7	5
2017	6	5
2017	5	5
2017	4	1
2017	3	1
2017	2	5
2017	1	5

Rate of Increase/Decrease

Analyzing whether Hosts can influence a Neighbourhood's Bookings and Listing price.

▶ **Do Superhosts Dominate NonSuperhosts in bookings?**
▶ **Do hosts have influence on neighbourhood pricing?**
▶ **Does the number of Superhosts or Non Superhosts in a neighbourhood effect the Avg pricing.**
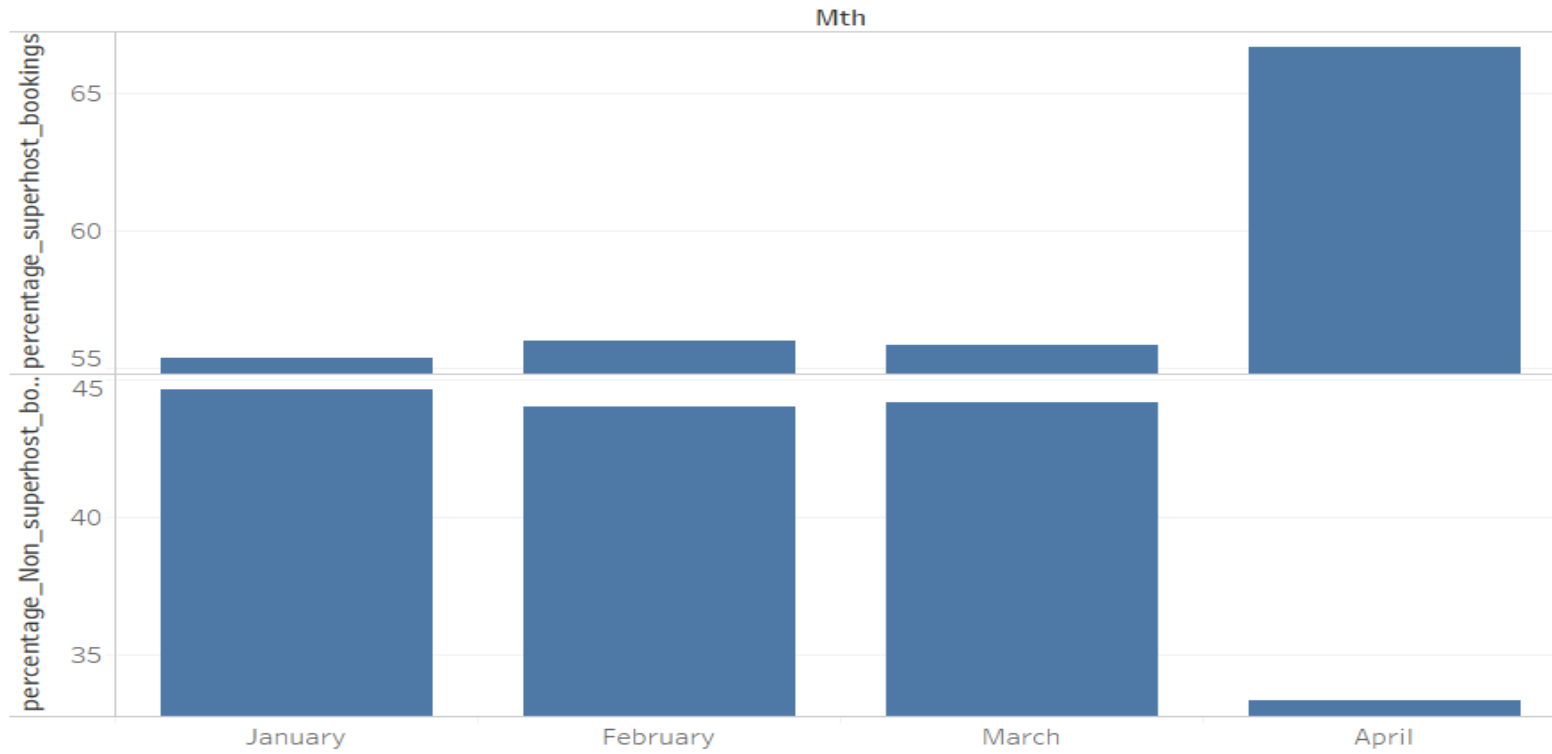
Percentage comparison of Bookings for Super Hosts and Non Super Hosts From Jan 17 to April 17

```
select t.Month,round((b*100)/(a+b),2) as Percentage_Non_Superhost_Bookings,
round((a*100)/(a+b),2) as Percentage_Superhost_Bookings from
(select sum(host_listings_count) a,monthname( c.date) as 'Month'
from listings l,calendar c where c.listing_id=l.id and c.date between '2017-01-
01' and '2017-04-01' and
host_is_superhost='t' and c.available='f' group by 2 order by 2 )t
inner join
 (select sum(host_listings_count) b,monthname( c.date) as 'Month'
from listings l,calendar c where c.listing_id=l.id and c.date between '2017-01-
01' and '2017-04-01' and
host_is_superhost='f' and c.available='f' group by 2 order by 2 )t1
on t.Month=t1.Month  ;
```

Result:

Month	Percentage_NonSuperhost_Bookings	Percentage_Superhost_Bookings
January	44.64	55.36
February	44	56
March	44.14	55.86
April	33.33	66.67

Percentage comparision of Bookings for Super Hosts and Non Super Hosts From Jan 17 to April 17

Findings:

- ❏ **Do Superhosts Dominate NonSuperhosts in bookings?**
- ❏ **Yes.Superhosts are more booked than Non Superhosts.**
- ❏ Percentage of listings that booked for SuperHost are greater than 55% of total bookings in all 4 months.
- ❏ Percentage of listings that booked for Non SuperHost are less than 45% of total bookings in all 4 months.

Comparison of total count of Super Hosts and Non Super Hosts and Avg price per night in the neighbourhoods of Boston

select t1.Neighbourhood,t.Total_SuperHosts,t1.Total_NON_SuperHosts
,round((t.a+t1.b)/2,2) as Avg_price_per_Night from
(select l.Neighbourhood ,count(host_is_superhost) as
Total_SuperHosts,round(sum(price)/count(host_is_superhost),2) as a from
listings l where host_is_superhost='t' and neighbourhood<>' ' group by
l.Neighbourhood)t,
(select l.Neighbourhood ,count(host_is_superhost) as
Total_NON_SuperHosts,round(sum(price)/count(host_is_superhost),2) as b
from listings l where host_is_superhost='f' and neighbourhood<>' ' group by
l.Neighbourhood)t1
where t.Neighbourhood=t1.Neighbourhood order by 4 ;

Results:

Neighbourhood	Total_Superhost	Total_NON_Superhosts	Avg_price_per_night
Hyde Park	2	23	82.87
Roslindale	19	31	94.62
Dorchester	26	167	98.74
West Roxbury	10	25	99.36
East Boston	6	111	103.22
Brookline	1	7	106.64
Mission Hill	4	99	114.46
Allston-Brighton	33	331	119.72
Jamaica Plain	76	237	134.96
Roxbury	12	102	167.4
Charlestown	14	65	200.28
Chinatown	2	76	203.08
Fenway/Kenmore	10	236	204.27
North End	17	106	206.22
West End	4	64	225.07
South End	47	243	228.4
South Boston	30	186	228.72
Beacon Hill	25	148	246.52
Leather District	3	5	253.64
Back Bay	21	267	259.9
Financial District	1	12	325.54
Theater District	2	31	346.8
Downtown Crossin	1	25	357.74

Highest:
★ Superhosts
★ NonSuperhosts
★ Avg Price

Comparision of Neighbourhoods by Avg Price per Night and total number of SuperHosts and Non SuperHosts in that neighbourhood.Sorted By Avg price to find costly neighbourhood.
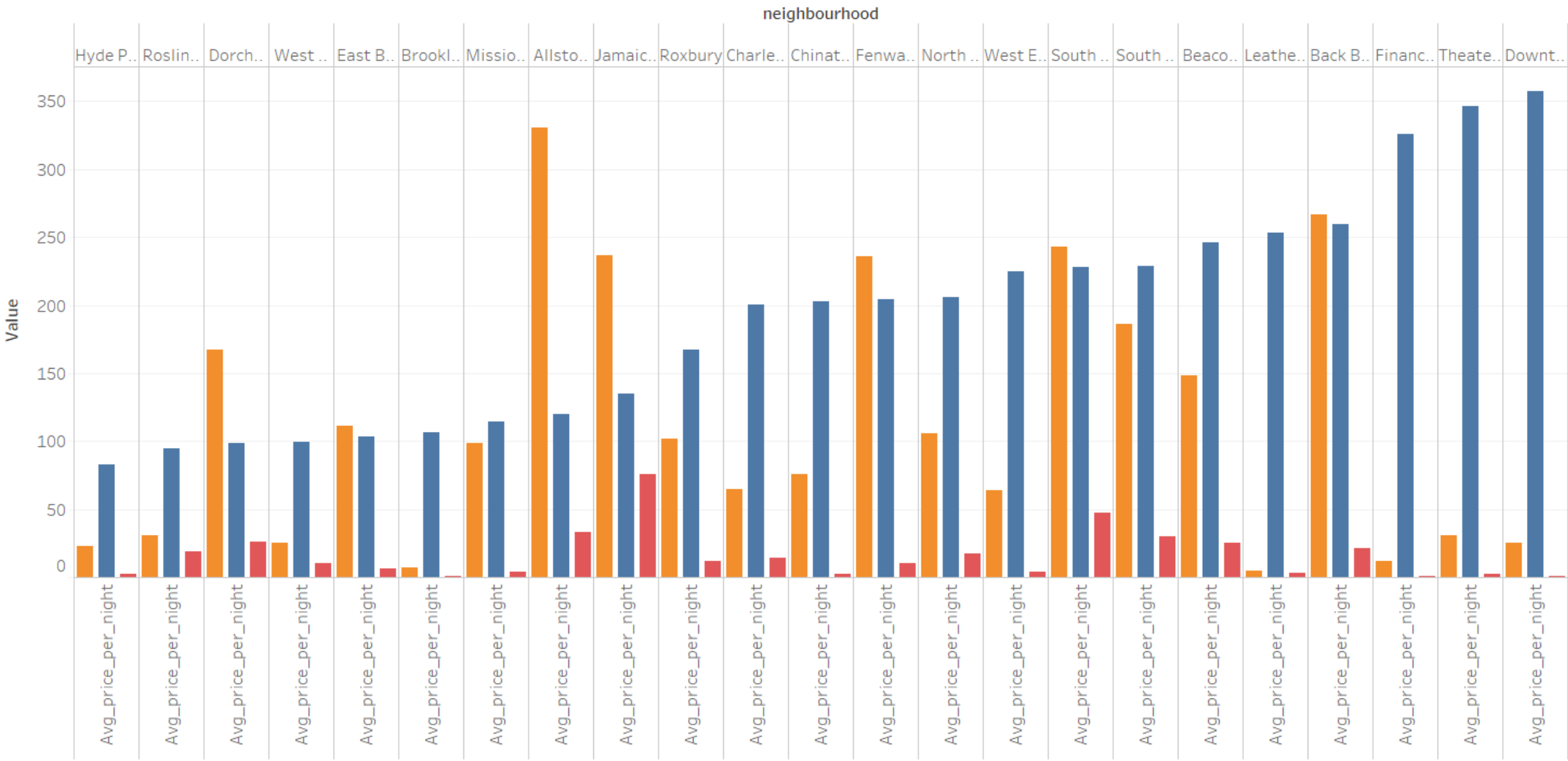
Findings:

In 23 neighbourhoods:

- ▶ Downtown Crossings has highest Avg price per night.
- ▶ Hyde Park had Least Avg Price.
- ▶ Jamaica Plain had highest number of SuperHosts.
- ▶ Allston Brighton had highest number of Non SuperHosts.
- ▶ **Do hosts have influence on neighbourhood pricing?**
- ▶ **No.**
- ▶ **Does the number of Superhosts or Non Superhosts in a neighbourhood effect the Avg pricing ?**
- ▶ **No.**
- ▶ Location of neighbourhood Can influence the pricing of listings.
- ▶ Neighbourhoods with high Avg price and low Avg price had almost same number of hosts around 26.So Hosts don't influence Neighbourhood pricing.

Find max and min price per night in each neighbourhood for Superhosts and Non Superhosts.

```
select t.Neighbourhood ,t.MaxPrice_SuperHost ,t.
MinPrice_SuperHost,t1.MaxPrice_Not_SuperHost
,t1.MinPrice_Not_SuperHost from
  (select Neighbourhood ,max(price) as MaxPrice_SuperHost
,min(price) as MinPrice_SuperHost  from listings where
host_is_superhost='t' and neighbourhood<>' ' group by
neighbourhood)t,
  (select Neighbourhood,max(price) as MaxPrice_Not_SuperHost
,min(price) as MinPrice_Not_SuperHost from listings where
host_is_superhost='f' and neighbourhood<>' ' group by
neighbourhood )t1
  where t.neighbourhood=t1.neighbourhood ;
```

Results:

Neighbourhood	MaxPrice_SuperHost	MinPrice_SuperHost	MaxPrice_Not_SuperHost	MinPrice_Not_SuperHost
Allston-Brighton	500	47	999	20
Back Bay	695	120	1250	40
Beacon Hill	849	100	769	75
Brookline	75	75	350	39
Charlestown	375	49	1000	39
Chinatown	179	159	399	93
Dorchester	325	40	395	29
Downtown Crossi	449	449	400	95
East Boston	120	50	359	30
Fenway/Kenmore	395	79	4000	30
Financial District	375	375	400	100
Hyde Park	85	55	269	31
Jamaica Plain	450	50	750	25
Leather District	400	159	390	139
Mission Hill	130	75	872	28
North End	345	69	559	45
Roslindale	229	40	285	45
Roxbury	625	78	1400	22
South Boston	500	95	3000	45
South End	1300	75	800	45
Theater District	600	350	500	80
West End	249	219	450	103
West Roxbury	119	50	349	45

Max and Min price per night in each neighbourhood for Superhosts and Non Superhosts.
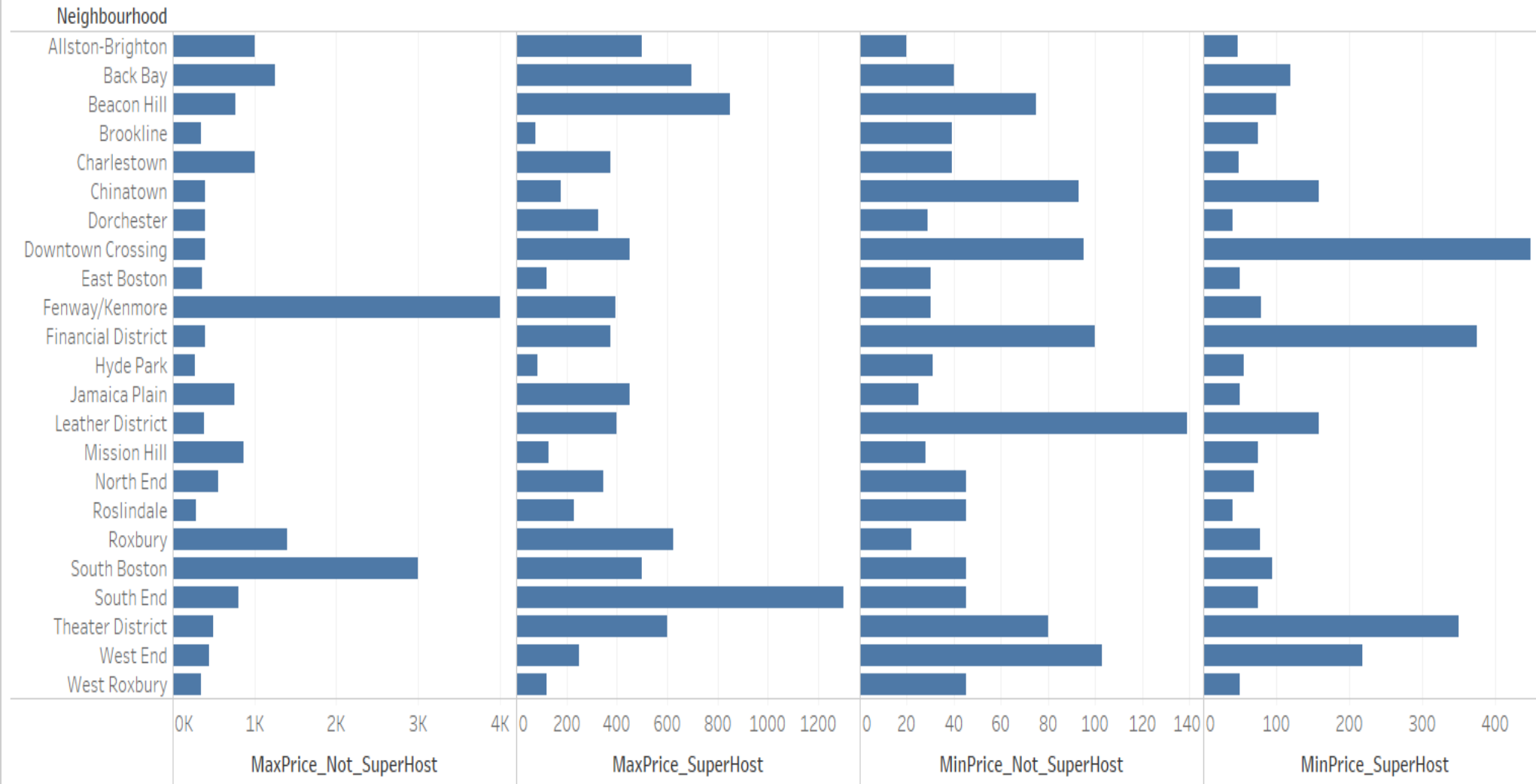
Findings:

Surprisingly, in all the neighbourhoods, maximum price of a superhost place is less than the maximum price of a Non Superhost place.
Whereas its the opposite with the minimum prices. The Minimum price of Non Superhosts are less than Superhosts.

Find top 10 places in Boston city that accommodates more than 3 guests with the host acceptance more than 50 with a trusted host.

SELECT host_acceptance_rate,
accommodates, host_id, host_name,city
FROM listings
where host_since>'4/15/2015' and city= 'boston' AND accommodates>3
AND host_acceptance_rate>50
in(select price from calendar c WHERE c.date>'2016/12/24')
order by host_acceptance_rate desc
limit 10

host_acceptance_...	accommodates	host_id	host_name	city
99	4	34,532,779	Christian	Boston
99	4	3,594,203	Shawn	Boston
99	6	3,594,203	Shawn	Boston
99	6	3,594,203	Shawn	Boston
99	4	3,594,203	Shawn	Boston
99	4	3,594,203	Shawn	Boston
99	6	43,959,997	Emma	Boston
99	5	36,981,635	Patt	Boston
99	6	43,959,997	Emma	Boston
99	4	43,959,997	Emma	Boston

Analysis of Guest Comments and Neighbourhoods

Purpose:
1. try to find out if there is a relationship between negative comments from guests and the median household income of the neighbourhood
   (assume listings in more affluent neighbourhoods would have less negative reviews)
1. try to find out if there a relationship between review patterns and the characteristics of the neighbourhood
   (assume there is a pattern of reviews left for certain neighbourhoods)

Negative Reviews v.s. Neighbourhood

```sql
select t3.neighbourhood, count(t2.comments) as number_of_bad_reviews, comment_count as number_of_tolReviews,
round((count(t2.comments)/comment_count)*100,2) as perc_of_bad_reviews, t3.median_household_income
from (select comments, listing_id from (select * from reviews
-- filter comments with bad words
where comments regexp 'dirty | disgusting | messy | dark |
rude | worse | disappointed | ghetto | trashyl skechy | Unfortunately |
 terrible | awful | honestly | beware') as t1
-- filter out mixed comments with good words
where comments
not regexp 'thank | wonderful | comfortable | pleasant |
amazing | satisfied | enjoyed | friendlyl welcoming | wow |
helpful | best | great | excellent | charmingl awesemel fantastic | sincerelyl pleased') as t2

-- joing listing talbe so that we can use neighbourhood as a foreign key to join next table
join listings l on l.id = t2.listing_id
-- create a table that made up of comement counts, listing id and neiggbourhood and median household income
 join (select count(comments) as comment_count, r.listing_id , n.neighbourhood, median_household_income
 from reviews r join listings l on r.listing_id=l.id
 join neighbourhoods n on n.neighbourhood=l.neighbourhood
 where l.neighbourhood is not null
 group by l.neighbourhood) t3 on l.neighbourhood= t3.neighbourhood
group by t3.neighbourhood
order by t3.median_household_income desc;
```

I'M SO ANGRY,
I CAN'T WAIT TO
WRITE A REVIEW!

Result:

neighbourhood	number_of_bad_reviews	number_of_tolReviews	perc_of_bad_reviews	median_household_income
Allston-Brighton	28	5156	0.54	52362
Back Bay	43	4333	0.99	103434
Beacon Hill	28	3367	0.83	103521
Cambridge	1	39	2.56	75909
Charlestown	7	1898	0.37	83926
Chinatown	9	651	1.38	46591
Dorchester	59	6339	0.93	42294
Downtown	1	91	1.1	60025
Downtown Crossing	5	954	0.52	141954
East Boston	66	4636	1.42	51832
Fenway/Kenmore	32	2776	1.15	60518
Financial District	1	56	1.79	125227
Hyde Park	1	267	0.37	53474
Jamaica Plain	38	8627	0.44	69341
Mattapan	1	148	0.68	48230
Mission Hill	27	1333	2.03	29559
North End	29	4026	0.72	116722
Roslindale	5	1304	0.38	81883
Roxbury	23	3117	0.74	34374
South Boston	25	4253	0.59	98301
South End	23	4736	0.49	88658
Theater District	18	605	2.98	56407
West End	2	268	0.75	72465
West Roxbury	1	672	0.15	101581

Conclusion: scatter plot shows no correlation

Scatter Plot of Median Household Income of the Neighbourhood and Number of Bad Reviews
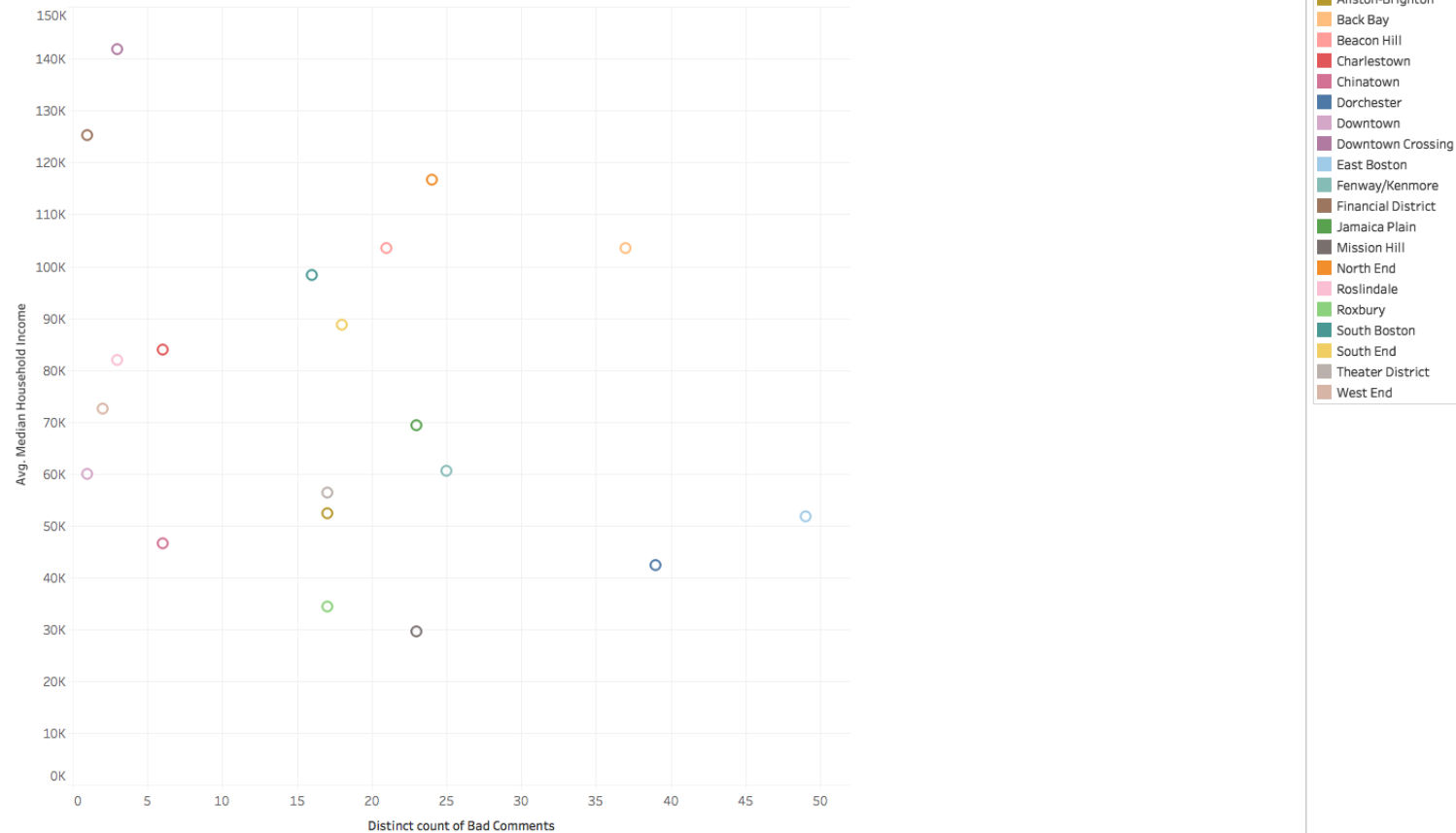
Review Pattern v.s. Neighbourhood

```sql
select comments, listing_id
from reviews r
join
(select id,city, state, n.neighbourhood, zipcode,
latitude, longitude, price from listings l
 join neighbourhoods n on n.neighbourhood =l.neighbourhood where l.neighbourhood<>'' and zipcode<>'') t
on t.id=listing_id
where r.comments
regexp 'superb | wonderful | comfortable | pleasant |
amazing | satisfied | enjoyed | wow | best | great | excellent | charming| aweseme| fantastic
| sincerely| pleasant';


select comments, listing_id
from reviews r
join
(select id,city, state, n.neighbourhood, zipcode,
latitude, longitude, price from listings l
 join neighbourhoods n on n.neighbourhood =l.neighbourhood where l.neighbourhood<>'' and zipcode<>'') t
on t.id=listing_id
where r.comments in
(select (select * from reviews
where comments regexp 'dirty | disgusting | messy | dark |
rude | worse | disappointed | ghetto | trashy| skechy | Unfortunately | honestly | beware') t2
where comments
not regexp 'thank | wonderful | comfortable | pleasant |
amazing | satisfied | enjoyed | friendly| welcoming | wow |
helpful | best | great | excellent | charming| aweseme| fantastic | sincerely| pleased');
```
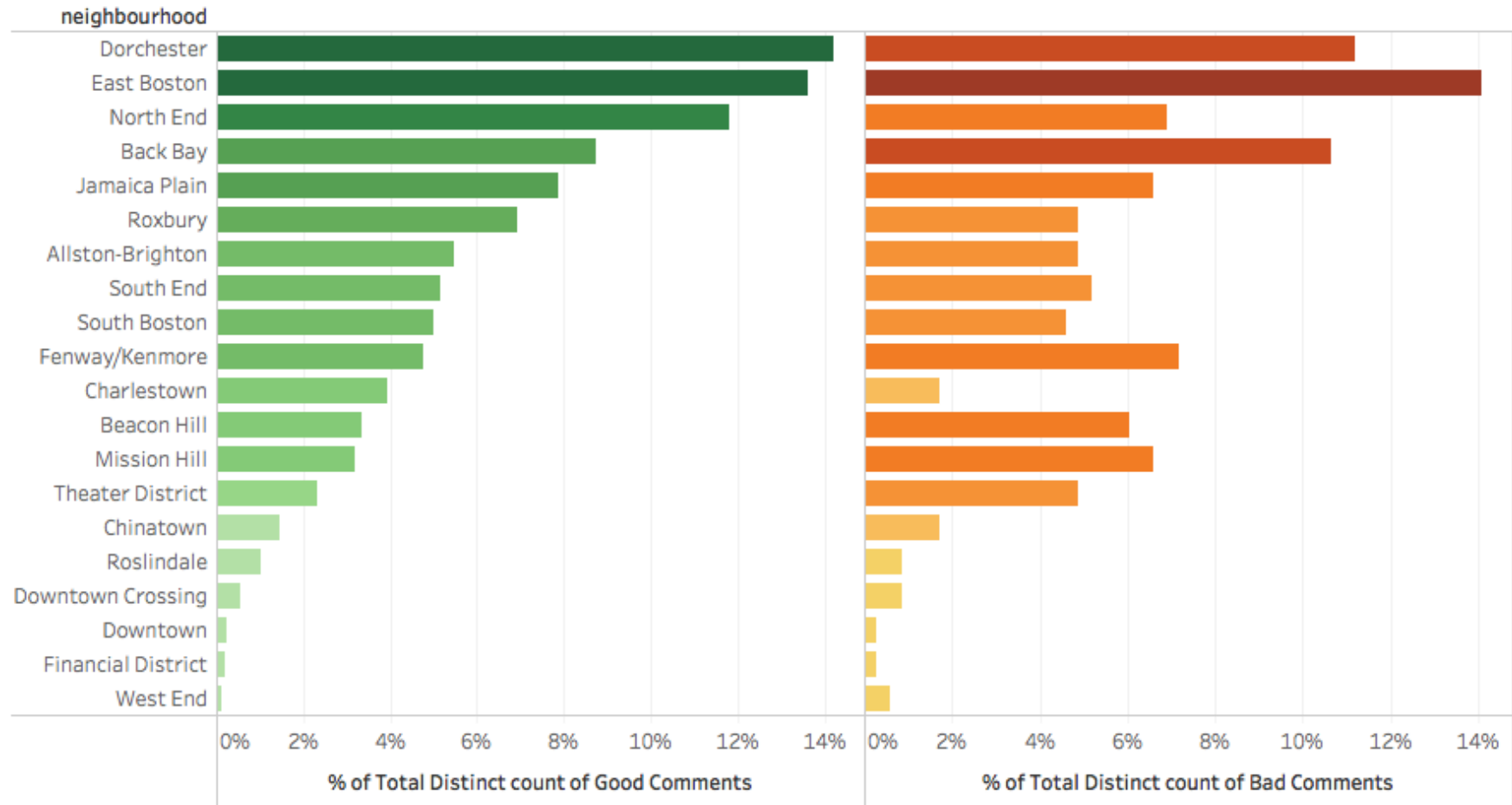
Result: Maps
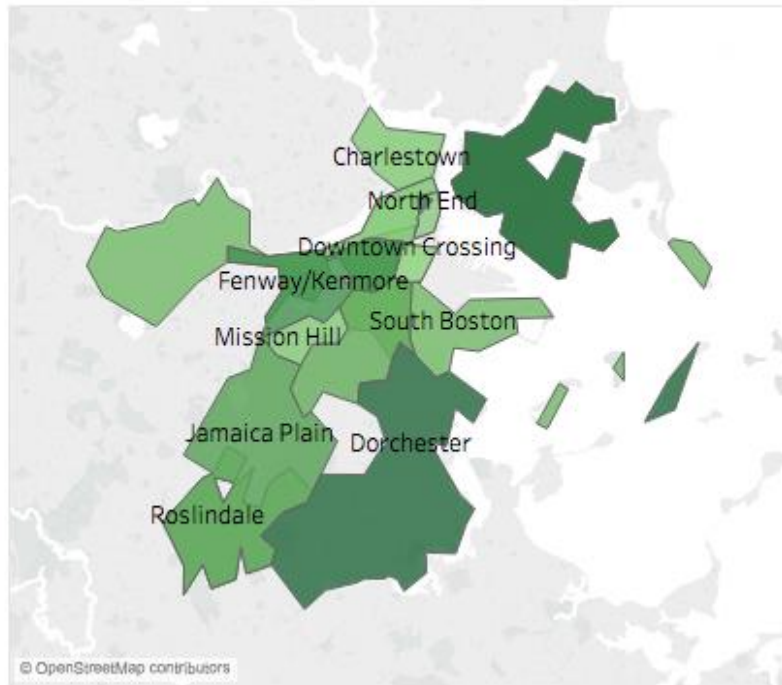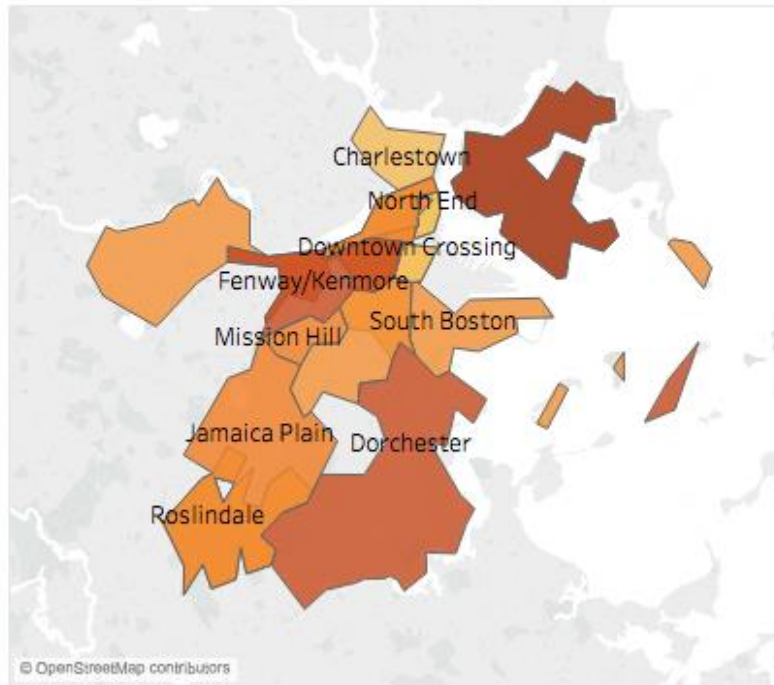
Result: Horizontal Bars

Comparison bwt Good & Bad Reviews by Neighbourhood

Summary:

1. There isn't any correlation between the median household income of a neighbourhood and the number of bad reviews from guests who stayed in that neighbourhood
2. the number of bad reviews seem correlated to the number of good reviews: the more good reviews the more bad review
3. guests tend to review more negatively on listings in financial district such as Back Bay area and East Boston where the Logan airport located
4. guests tend to review more positively on historic area of Boston such as Dorchester, North End, and Charlestown