

```
USE DATABASE FINAL_GROUP_PROJECT;

#DATA CLEANING - strip the $

UPDATE SEATTLE_AIRBNB
SET PRICE = SUBSTRING(PRICE,2,100);

UPDATE SEATTLE_AIRBNB
SET WEEKLY_PRICE = SUBSTRING(WEEKLY_PRICE,2,100);

UPDATE SEATTLE_AIRBNB
SET MONTHLY_PRICE = SUBSTRING(MONTHLY_PRICE,2,100);

UPDATE SEATTLE_AIRBNB
SET SECURITY_DEPOSIT = SUBSTRING(SECURITY_DEPOSIT,2,100);

UPDATE SEATTLE_AIRBNB
SET CLEANING_FEE = SUBSTRING(CLEANING_FEE,2,100);

UPDATE SEATTLE_AIRBNB
SET EXTRA_PEOPLE = SUBSTRING(EXTRA_PEOPLE,2,100);
```

```
# DATA CLEANING - Change datatype
```

```
ALTER TABLE SEATTLE_AIRBNB  
ADD COLUMN PRICE1 DECIMAL(10, 2);  
UPDATE SEATTLE_AIRBNB  
SET PRICE1 = CAST(REPLACE(PRICE,'','','') AS DECIMAL(10,2));  
ALTER TABLE SEATTLE_AIRBNB  
DROP COLUMN PRICE;  
ALTER TABLE SEATTLE_AIRBNB  
RENAME COLUMN PRICE1 TO PRICE;
```

```
ALTER TABLE SEATTLE_AIRBNB  
ADD COLUMN WEEKLY_PRICE1 DECIMAL(20, 2);  
UPDATE SEATTLE_AIRBNB  
SET WEEKLY_PRICE1 = CAST(REPLACE(WEEKLY_PRICE,'','','') AS DECIMAL(20,2));  
ALTER TABLE SEATTLE_AIRBNB  
DROP COLUMN WEEKLY_PRICE;  
ALTER TABLE SEATTLE_AIRBNB  
RENAME COLUMN WEEKLY_PRICE1 TO WEEKLY_PRICE;
```

```
ALTER TABLE SEATTLE_AIRBNB  
ADD COLUMN MONTHLY_PRICE1 DECIMAL(20, 2);  
UPDATE SEATTLE_AIRBNB  
SET MONTHLY_PRICE1 = CAST(REPLACE(MONTHLY_PRICE,'','','') AS DECIMAL(20,2));  
ALTER TABLE SEATTLE_AIRBNB  
DROP COLUMN MONTHLY_PRICE;  
ALTER TABLE SEATTLE_AIRBNB  
RENAME COLUMN MONTHLY_PRICE1 TO MONTHLY_PRICE;
```

```
ALTER TABLE SEATTLE_AIRBNB  
ADD COLUMN SECURITY_DEPOSIT1 DECIMAL(20, 2);  
UPDATE SEATTLE_AIRBNB  
SET SECURITY_DEPOSIT1 = CAST(REPLACE(SECURITY_DEPOSIT,'','','') AS DECIMAL(20,2));  
ALTER TABLE SEATTLE_AIRBNB  
DROP COLUMN SECURITY_DEPOSIT;  
ALTER TABLE SEATTLE_AIRBNB  
RENAME COLUMN SECURITY_DEPOSIT1 TO SECURITY_DEPOSIT;
```

```
ALTER TABLE SEATTLE_AIRBNB
ADD COLUMN CLEANING_FEE1 DECIMAL(20, 2);
UPDATE SEATTLE_AIRBNB
SET CLEANING_FEE1 = CAST(REPLACE(CLEANING_FEE,'','') AS
DECIMAL(20,2));
ALTER TABLE SEATTLE_AIRBNB
DROP COLUMN CLEANING_FEE;
ALTER TABLE SEATTLE_AIRBNB
RENAME COLUMN CLEANING_FEE1 TO CLEANING_FEE;

ALTER TABLE SEATTLE_AIRBNB
ADD COLUMN EXTRA_PEOPLE1 DECIMAL(20, 2);
UPDATE SEATTLE_AIRBNB
SET EXTRA_PEOPLE1 = CAST(REPLACE(EXTRA_PEOPLE,'','') AS
DECIMAL(20,2));
ALTER TABLE SEATTLE_AIRBNB
DROP COLUMN EXTRA_PEOPLE;
ALTER TABLE SEATTLE_AIRBNB
RENAME COLUMN EXTRA_PEOPLE1 TO EXTRA_PEOPLE;

ALTER TABLE SEATTLE_CALENDAR
ADD COLUMN PRICE1 DECIMAL(20, 2);
UPDATE SEATTLE_CALENDAR
SET PRICE = SUBSTRING(PRICE,2,100);
UPDATE SEATTLE_CALENDAR
SET PRICE1 = CAST(REPLACE(PRICE,'','') AS DECIMAL(20,2));
ALTER TABLE SEATTLE_CALENDAR
DROP COLUMN PRICE;
ALTER TABLE SEATTLE_CALENDAR
RENAME COLUMN PRICE1 TO PRICE;
```

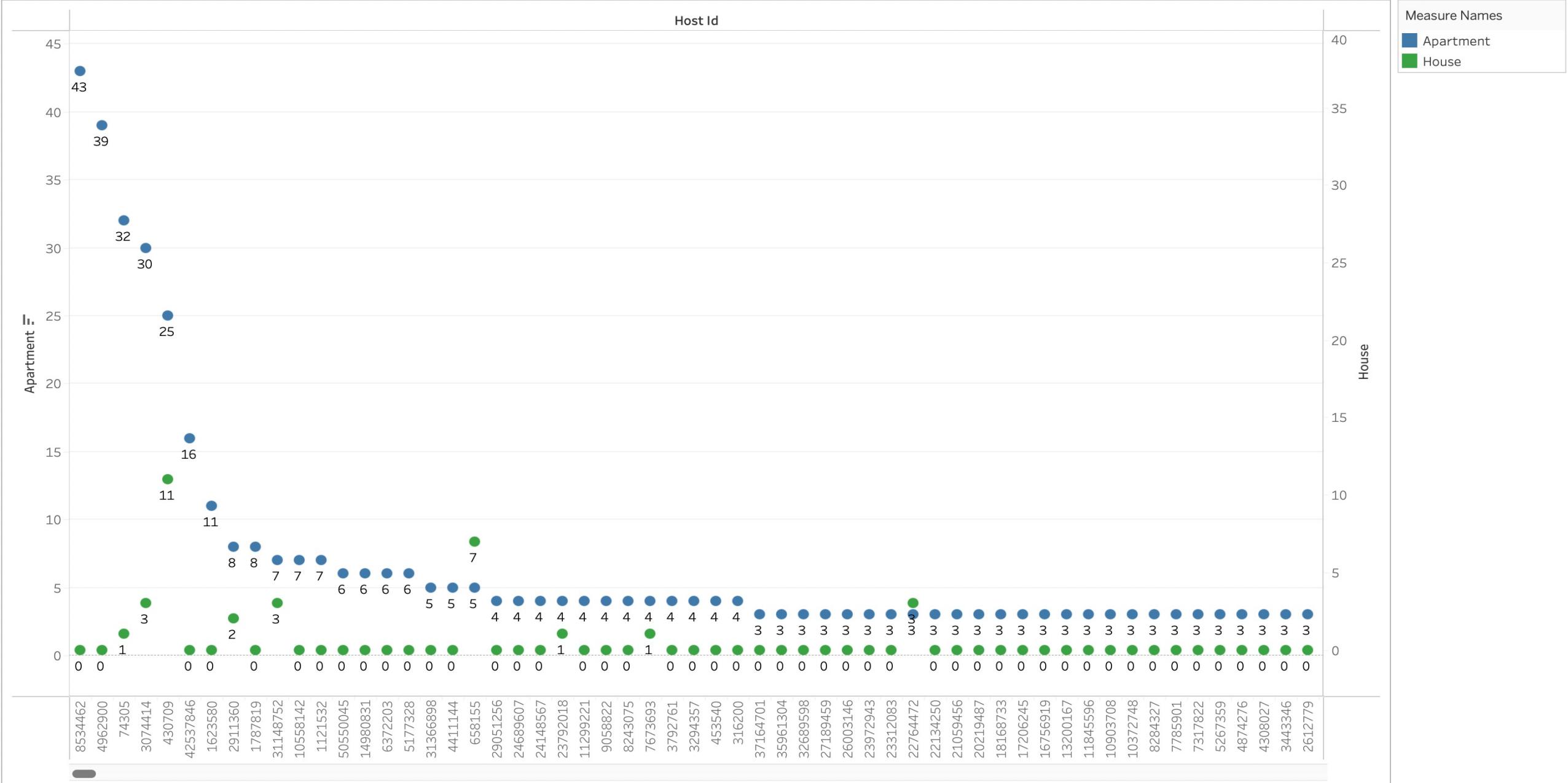
```
#DATA CLEANING - STRIP %
```

```
ALTER TABLE SEATTLE_AIRBNB
ADD COLUMN HOST_RESPONSE_RATE1 DECIMAL(20, 2);
UPDATE SEATTLE_AIRBNB
SET HOST_RESPONSE_RATE = NULL
WHERE HOST_RESPONSE_RATE = 'N/A';
UPDATE SEATTLE_AIRBNB
SET HOST_RESPONSE_RATE1 = CAST(REPLACE(HOST_RESPONSE_RATE,'%','') AS INT);
ALTER TABLE SEATTLE_AIRBNB
DROP COLUMN HOST_RESPONSE_RATE;
ALTER TABLE SEATTLE_AIRBNB
RENAME COLUMN HOST_RESPONSE_RATE1 TO HOST_RESPONSE_RATE;

ALTER TABLE SEATTLE_AIRBNB
ADD COLUMN HOST_ACCEPTANCE_RATE1 DECIMAL(20, 2);
UPDATE SEATTLE_AIRBNB
SET HOST_ACCEPTANCE_RATE = NULL
WHERE HOST_ACCEPTANCE_RATE = 'N/A';
UPDATE SEATTLE_AIRBNB
SET HOST_ACCEPTANCE_RATE1 = CAST(REPLACE(HOST_ACCEPTANCE_RATE,'%','') AS
INT);
ALTER TABLE SEATTLE_AIRBNB
DROP COLUMN HOST_ACCEPTANCE_RATE;
ALTER TABLE SEATTLE_AIRBNB
RENAME COLUMN HOST_ACCEPTANCE_RATE1 TO HOST_ACCEPTANCE_RATE;
```

#Rank each host based on the number of beds they have listed.

```
SELECT HOST_ID,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Apartment%' THEN 1 ELSE 0 END) AS APARTMENT,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%House%' THEN 1 ELSE 0 END) AS HOUSE,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Cabin%' THEN 1 ELSE 0 END) AS CABIN,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Condominium%' THEN 1 ELSE 0 END) AS CONDOMINIUM,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Camper%' THEN 1 ELSE 0 END) AS CAMPERRV,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Bungalow%' THEN 1 ELSE 0 END) AS Bungalow,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Townhouse%' THEN 1 ELSE 0 END) AS TOWNHOUSE,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Boat%' THEN 1 ELSE 0 END) AS BOAT,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Breakfast%' THEN 1 ELSE 0 END) AS BED_BREAKFAST,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Dorm%' THEN 1 ELSE 0 END) AS DORM,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Treehouse%' THEN 1 ELSE 0 END) AS TREEHOUSE,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Yurt%' THEN 1 ELSE 0 END) AS YURT,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Chalet%' THEN 1 ELSE 0 END) AS CHALET,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Tent%' THEN 1 ELSE 0 END) AS TENT,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Loft%' THEN 1 ELSE 0 END) AS LOFT,  
SUM(CASE WHEN PROPERTY_TYPE LIKE '%Other%' THEN 1 ELSE 0 END) AS OTHER  
FROM SEATTLE_AIRBNB  
GROUP BY HOST_ID;
```



Caption

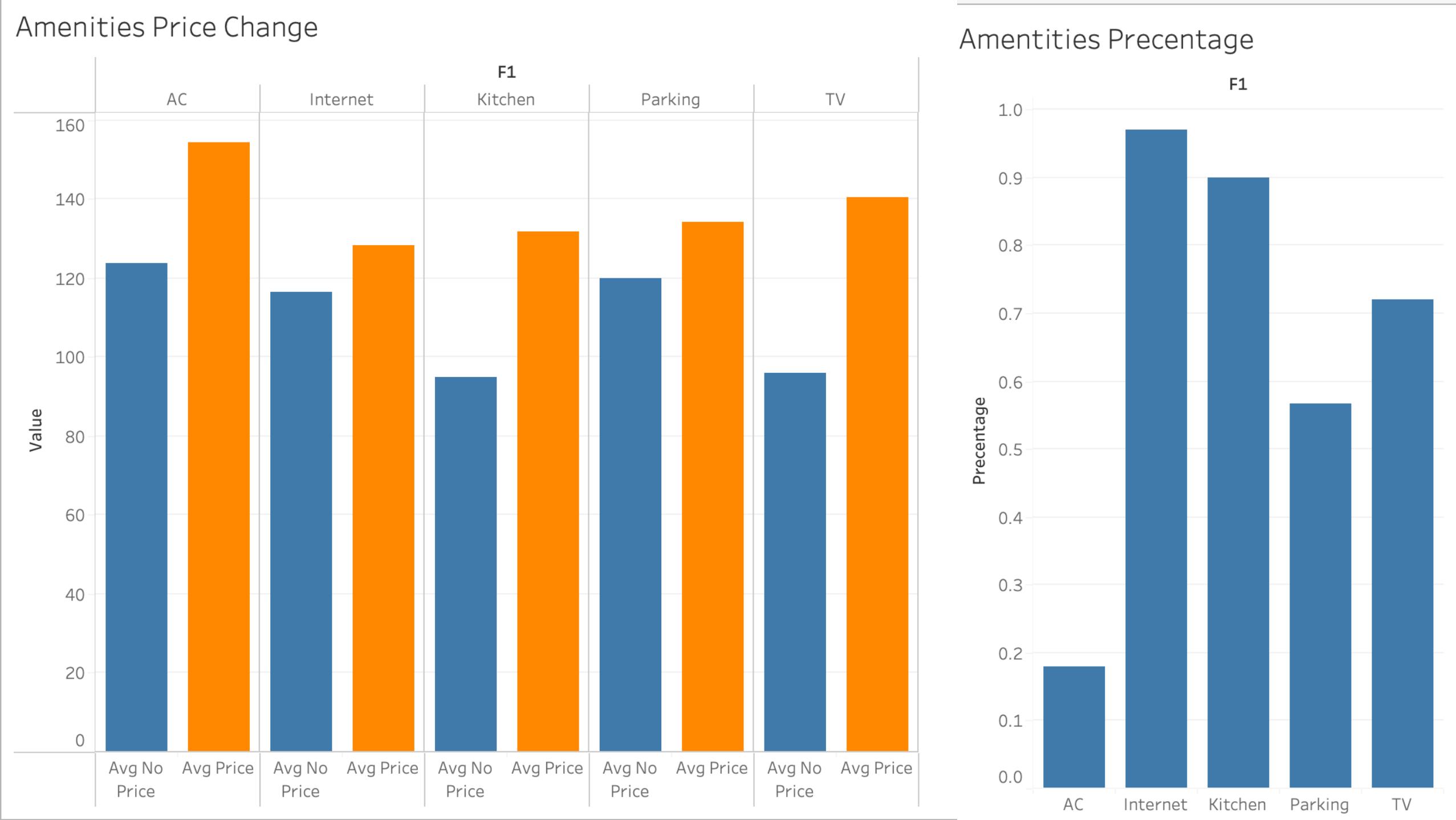
Apartment and House for each Host Id. Color shows details about Apartment and House. For pane Sum of Apartment: The marks are labeled by Apartment. For pane Sum of House: The marks are labeled by House.

Amenities analysis

```
SELECT NEIGHBOURHOOD,
AVG(CASE WHEN AMENITIES LIKE '%AIR CONDITION%' THEN PRICE ELSE NULL END) AS AVG_AC_PRICE,
AVG(CASE WHEN AMENITIES NOT LIKE '%AIR CONDITION%' THEN PRICE ELSE NULL END) AS
AVG_NOAC_PRICE,
AVG(CASE WHEN AMENITIES LIKE '%AIR CONDITION%' THEN REVIEW_SCORES_VALUE ELSE NULL END)
AS AVG_AC_REVIEW,
AVG(CASE WHEN AMENITIES NOT LIKE '%AIR CONDITION%' THEN REVIEW_SCORES_VALUE ELSE NULL
END) AS AVG_NOAC_REVIEW,
count(case when amenities like '%Air Condition%' then price else null end) / count(*) as ac_precentage
FROM SEATTLE_AIRBNB
GROUP BY NEIGHBOURHOOD;
```

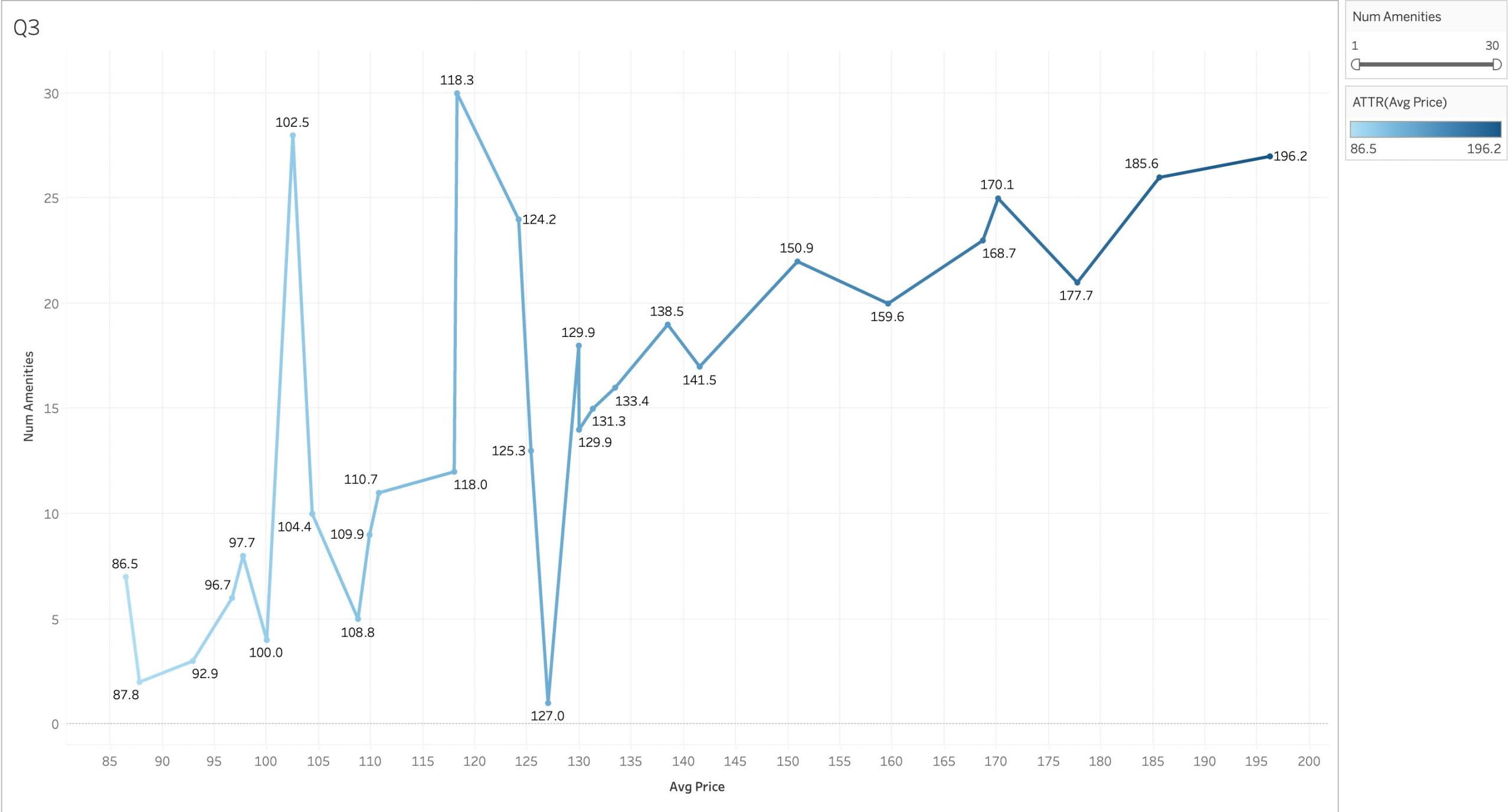
```
SELECT NEIGHBOURHOOD,
AVG(CASE WHEN AMENITIES LIKE '%PARKING%' THEN PRICE ELSE NULL END) AS AVG_PARK_PRICE,
AVG(CASE WHEN AMENITIES NOT LIKE '%PARKING%' THEN PRICE ELSE NULL END) AS
AVG_NOPARK_PRICE,
AVG(CASE WHEN AMENITIES LIKE '%PARKING%' THEN REVIEW_SCORES_VALUE ELSE NULL END) AS
AVG_PARK_REVIEW,
AVG(CASE WHEN AMENITIES NOT LIKE '%PARKING%' THEN REVIEW_SCORES_VALUE ELSE NULL END) AS
AVG_NOPARK_REVIEW,
count(case when amenities like '%Parking%' then price else null end) / count(*) as park_precentage
FROM SEATTLE_AIRBNB
GROUP BY NEIGHBOURHOOD;
```

```
SELECT NEIGHBOURHOOD,
ROUND(AVG(CASE WHEN AMENITIES LIKE '%TV%' THEN PRICE ELSE NULL END), 2) AS AVG_TV_PRICE,
ROUND(AVG(CASE WHEN AMENITIES NOT LIKE '%TV%' THEN PRICE ELSE NULL END), 2) AS
AVG_NOTV_PRICE,
ROUND((AVG_TV_PRICE - AVG_NOTV_PRICE) / AVG_TV_PRICE, 2) AS PRICE_DIFF,
ROUND(AVG(CASE WHEN AMENITIES LIKE '%TV%' THEN REVIEW_SCORES_VALUE ELSE NULL END), 2) AS
AVG_TV_REVIEW,
ROUND(AVG(CASE WHEN AMENITIES NOT LIKE '%TV%' THEN REVIEW_SCORES_VALUE ELSE NULL END),
2) AS AVG_NOTV_REVIEW,
ROUND(COUNT(CASE WHEN AMENITIES LIKE '%TV%' THEN PRICE ELSE NULL END) / COUNT(*), 2) AS
TV_PERCENTAGE
FROM SEATTLE_AIRBNB
GROUP BY NEIGHBOURHOOD;
```



#Average base on count of amenities

```
SELECT NUM_AMENITIES, AVG(PRICE) AS AVG_PRICE
FROM (SELECT PRICE,
LENGTH(AMENITIES) - LENGTH(REPLACE(AMENITIES,',',''))+1
AS NUM_AMENITIES
FROM SEATTLE_AIRBNB)
GROUP BY NUM_AMENITIES
ORDER BY AVG_PRICE DESC;
```

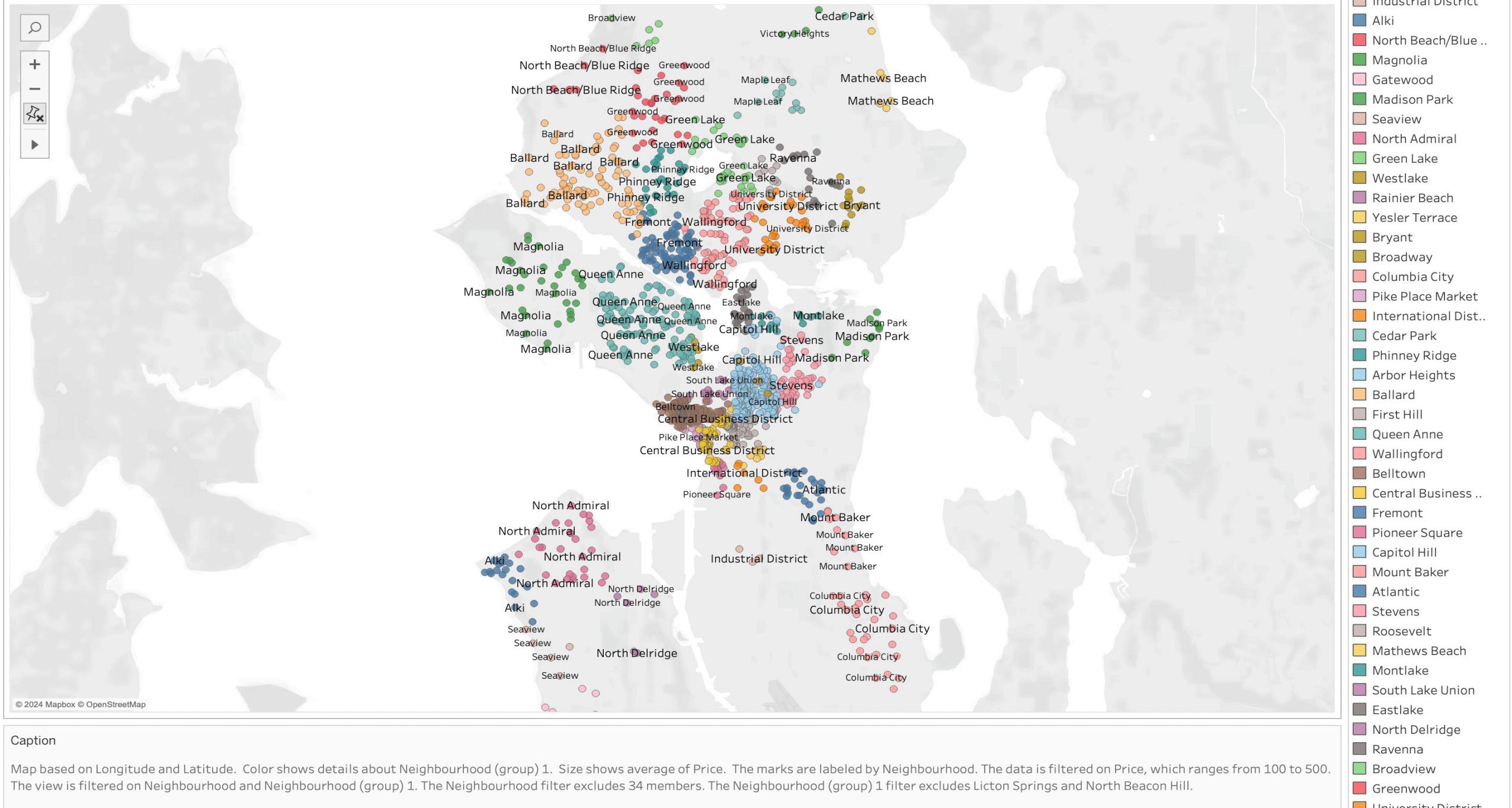


Caption

#Price by neighborhood

```
SELECT DISTINCT NEIGHBOURHOOD, AVG(LATITUDE), AVG(LONGITUDE),
ROUND(AVG(PRICE),2) AS AVG_PRICE,
ROUND(AVG(WEEKLY_PRICE),2) AS AVG_WEEKLY_PRICE,
ROUND(AVG(MONTHLY_PRICE),2) AS AVG_MONTHLY_PRICE
FROM SEATTLE_AIRBNB
GROUP BY NEIGHBOURHOOD
ORDER BY AVG_PRICE DESC;
```

Q4



#Identify listings with a high price relative to the average price in their neighborhood and property type

```
SELECT ID, NAME, PRICE, NEIGHBOURHOOD,  
CASE WHEN PRICE > 1.5* AVG(PRICE) OVER (PARTITION BY NEIGHBOURHOOD, PROPERTY_TYPE)  
THEN 'HIGH PRICE' ELSE 'NORMAL PRICE' END AS PRICE_CATEGORY  
FROM SEATTLE_AIRBNB;
```

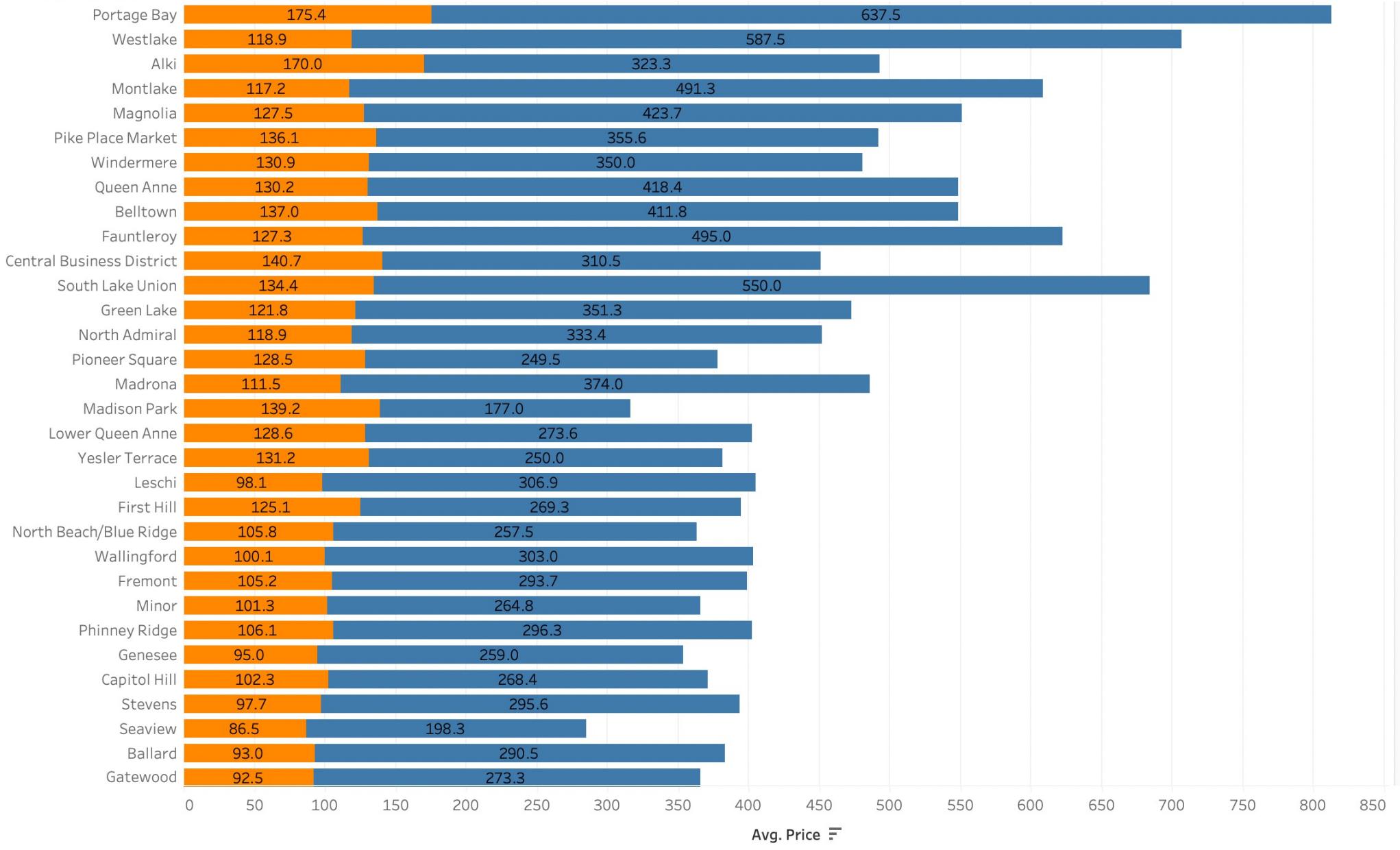
Q5

Neighbourhood

Price Category

HIGH PRICE

NORMAL PRICE



Caption

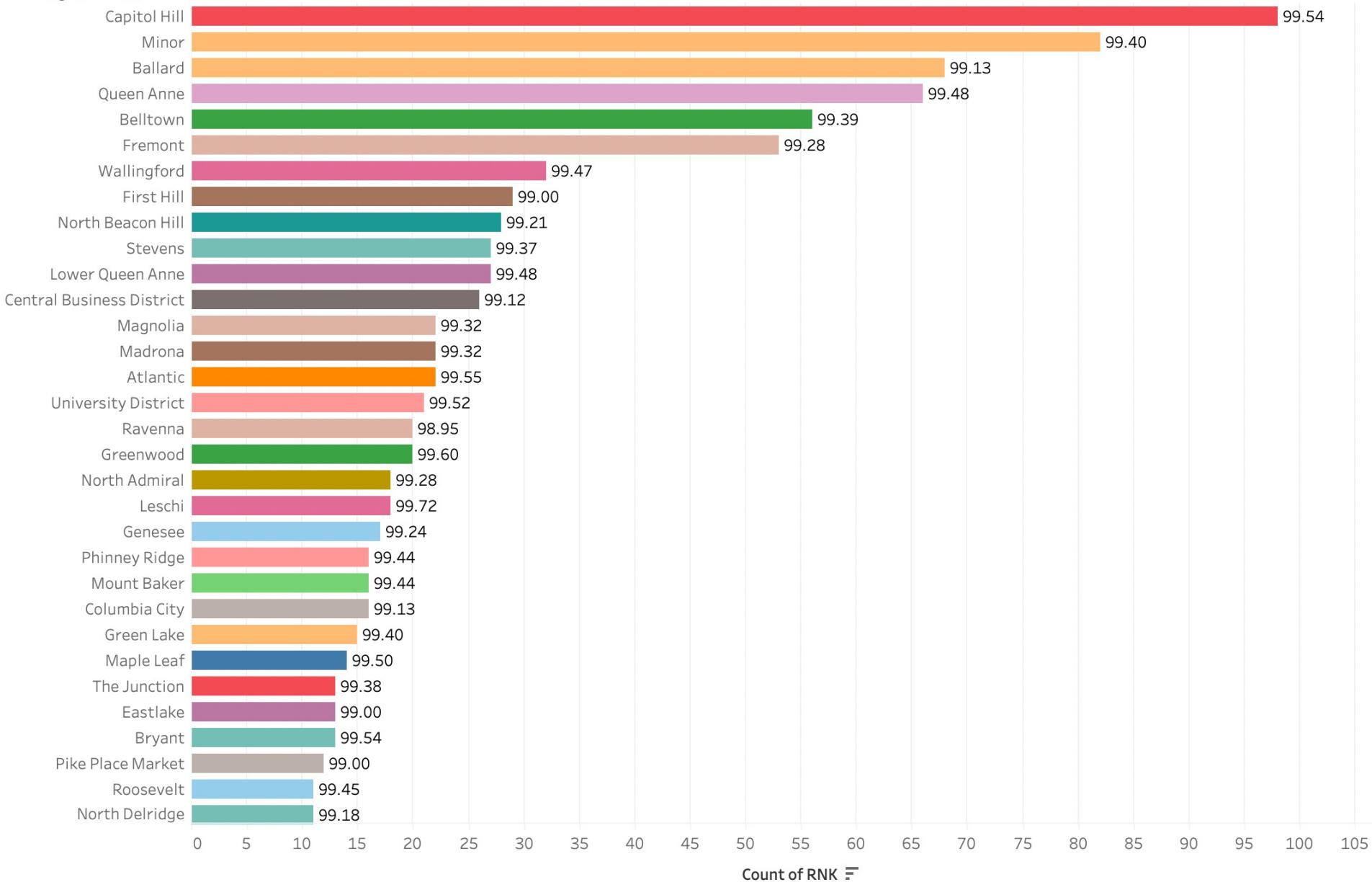
#Top 3 review score rating by neighborhood

```
WITH CTE AS(SELECT ID, NAME, NEIGHBOURHOOD,
REVIEW_SCORES_RATING,
DENSE_RANK() OVER(PARTITION BY NEIGHBOURHOOD ORDER BY
REVIEW_SCORES_RATING DESC) AS RNK
FROM SEATTLE_AIRBNB
WHERE REVIEW_SCORES_RATING IS NOT NULL)

SELECT ID, NAME, NEIGHBOURHOOD, REVIEW_SCORES_RATING, RNK
FROM CTE
WHERE RNK <=3
ORDER BY NEIGHBOURHOOD ASC;
```

Q6

Neighbourhood



AVG(Review Scores Rating)

85.67

99.72

Summary

Count:	79
AVG(Review Scores Rating)	
Average:	98.01
Minimum:	85.67
Maximum:	99.72
Median:	99.00
CNT(RNK)	
Average:	14.68
Minimum:	1
Maximum:	98
Median:	8.00

Neighbourhood

- Alki
- Arbor Heights
- Atlantic
- Ballard
- Belltown
- Bitter Lake
- Brighton
- Broadview
- Broadway
- Bryant
- Capitol Hill
- Cedar Park
- Central Business District
- Columbia City
- Crown Hill
- Dunlap
- Eastlake
- Fauntleroy
- First Hill
- Fremont
- Gatewood
- Genesee
- Georgetown
- Green Lake
- Greenwood

Caption

#Calculate the Median Price for Each Neighborhood and
Property Type Combination

```
SELECT NEIGHBOURHOOD, PROPERTY_TYPE,  
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY PRICE) AS  
MEDIAN_PRICE  
FROM SEATTLE_AIRBNB  
GROUP BY NEIGHBOURHOOD, PROPERTY_TYPE  
ORDER BY NEIGHBOURHOOD,MEDIAN_PRICE DESC;
```


#Calculate the Median Price for each neighborhoods with the highest number of reviews

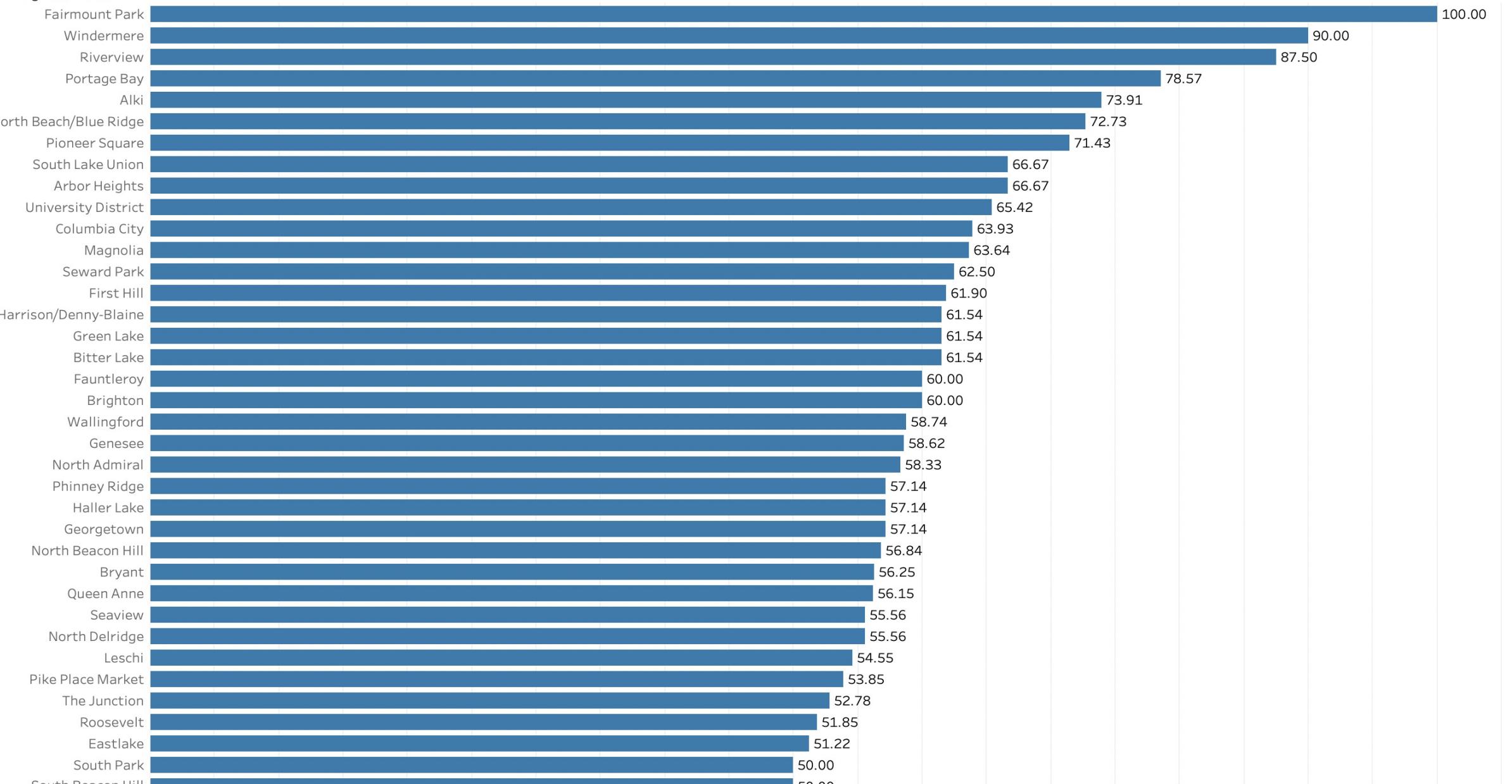
```
WITH CTE AS(SELECT NEIGHBOURHOOD  
FROM SEATTLE_AIRBNB  
WHERE NUMBER_OF_REVIEWS > (SELECT  
AVG(NUMBER_OF_REVIEWS) FROM SEATTLE_AIRBNB)  
GROUP BY NEIGHBOURHOOD  
ORDER BY COUNT(*) DESC  
LIMIT 3)
```

```
SELECT ID, NAME, NEIGHBOURHOOD, PRICE,  
ROUND(PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY PRICE)  
OVER (PARTITION BY NEIGHBOURHOOD),2) AS MEDIAN_PRICE  
FROM SEATTLE_AIRBNB  
WHERE NEIGHBOURHOOD IN (SELECT NEIGHBOURHOOD FROM CTE);
```

#Security Deposit Percentage by Neighborhood

```
SELECT NEIGHBOURHOOD,  
       (COUNT(CASE WHEN SECURITY_DEPOSIT IS NOT NULL THEN 1 END)  
        / COUNT(*)) *100 AS SECURITY_DEPOSIT_PERCENTAGE  
    FROM SEATTLE_AIRBNB  
   GROUP BY NEIGHBOURHOOD  
ORDER BY SECURITY_DEPOSIT_PERCENTAGE DESC;
```

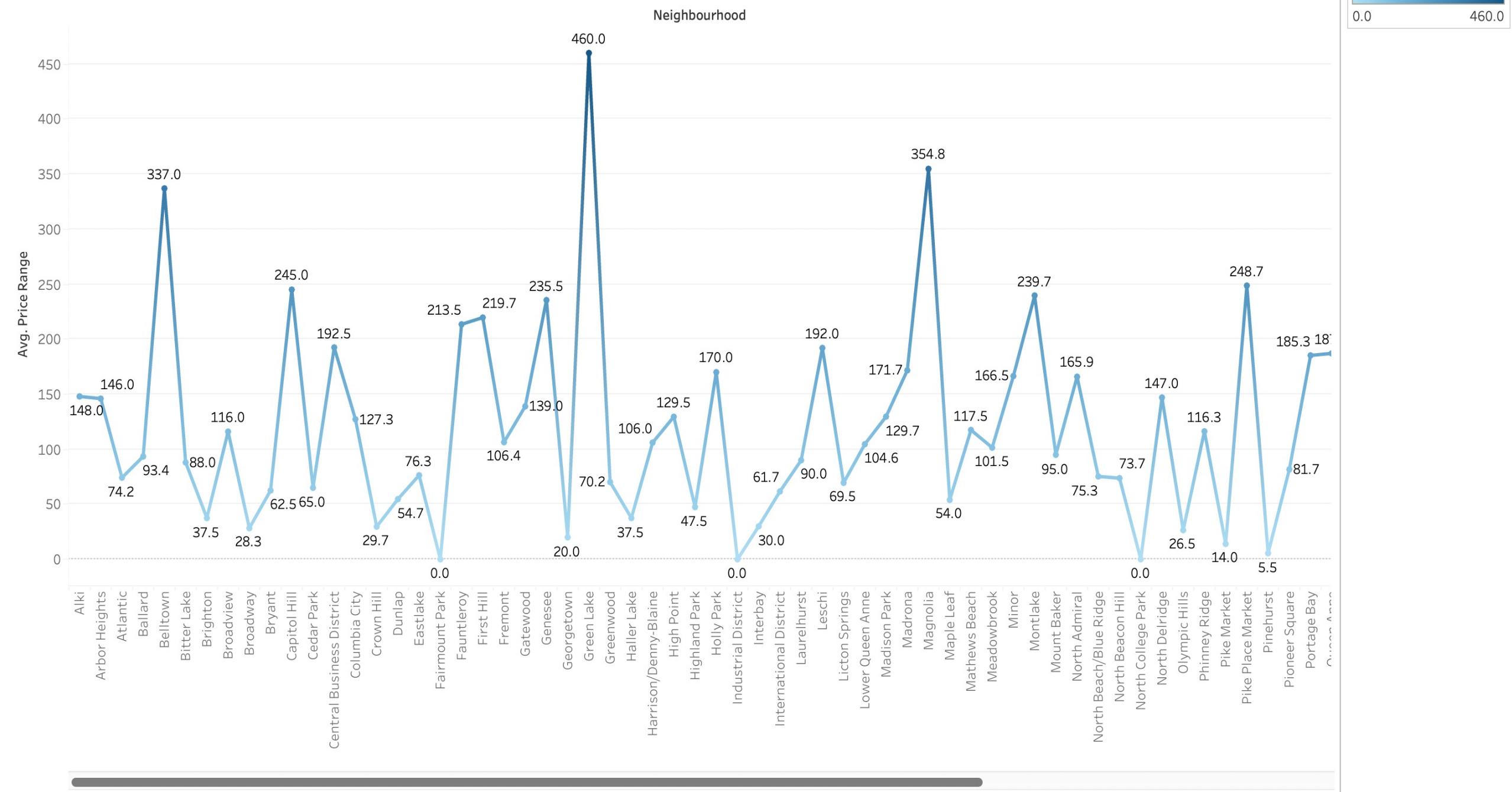
Neighbourhood



#Identify Neighborhoods with the Highest Price Fluctuations
for each property type.

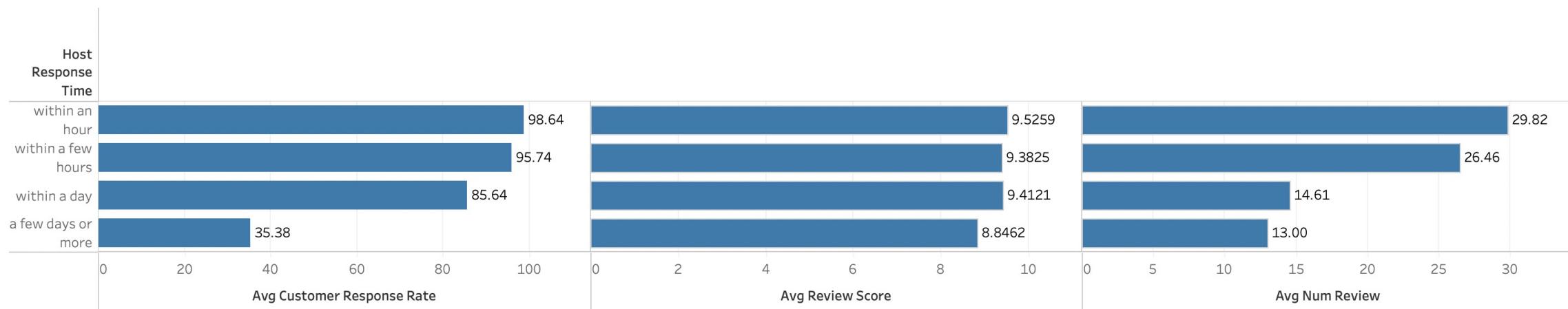
```
SELECT NEIGHBOURHOOD, PROPERTY_TYPE, PRICE_RANGE
FROM (SELECT NEIGHBOURHOOD, PROPERTY_TYPE,
MAX(PRICE) - MIN(PRICE) AS PRICE_RANGE
FROM SEATTLE_AIRBNB
GROUP BY NEIGHBOURHOOD, PROPERTY_TYPE) AS SUB
WHERE PRICE_RANGE IS NOT NULL
ORDER BY NEIGHBOURHOOD, PRICE_RANGE DESC;
```

Q10



```
#Correlation between response speed to customer  
booking count  
SELECT HOST_RESPONSE_TIME,  
AVG(HOST_RESPONSE_RATE) AS AVG_RESPONSE_RATE,  
AVG(HOST_ACCEPTANCE_RATE) AS AVG_ACCEPT_RATE,  
AVG(NUMBER_OF_REVIEWS) AS AVG_NUM_REVIEW,  
AVG(REVIEW_SCORES_VALUE) AS AVG REVIEW SCORE,  
COUNT(*)  
FROM SEATTLE_AIRBNB  
WHERE HOST_RESPONSE_RATE IS NOT NULL AND  
HOST_ACCEPTANCE_RATE IS NOT NULL  
GROUP BY HOST_RESPONSE_TIME  
ORDER BY HOST_RESPONSE_TIME;
```

Q11



#Relationship between Neighborhood, Room type, average price, and number of Airbnb's.

```
SELECT NEIGHBOURHOOD, ROOM_TYPE, AVG(a.PRICE), COUNT(*)AS  
NUMBER_OF_AIRBNB  
FROM SEATTLE_AIRBNB A  
JOIN SEATTLE_CALENDAR C ON A.ID = C.LISTING_ID  
GROUP BY NEIGHBOURHOOD, ROOM_TYPE  
ORDER BY NEIGHBOURHOOD;
```

Q12

