CIS5100-93 Term Project

Authors: Adalberto Delgado, Tremon Everett, Angel Garcia, Nicholas Perez, Jessica Salazar, Jocelyn Salazar Andrade
Instructor: Jongwook Woo
Date: 04/03/2024

Lab Tutorial

{adelga126, teveret2, agarci772, nperez125, jherna364, jsalaz18, jwoo5}@calstatela.edu

Airbnb Data Analysis in Los Angeles County: Insights for Hosts using SAP Analytics Cloud

---

Objective

- In this hands-on lab, you will learn how to create:
  - Bar Chart
  - Pie Chart
  - Numeric Point Chart
  - Geo-Map
  - Scatterplot
  - Bubble Graph
  - Time Series with Linear Regression

Platform Spec

- SAP (SAC)

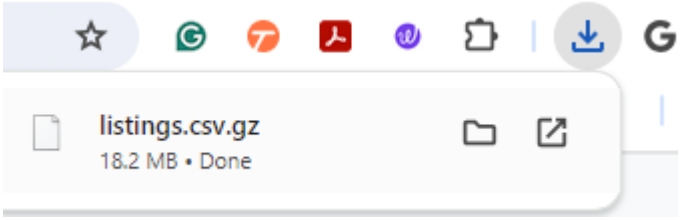
Data Required

- Inside Airbnb Listings.csv.gz

Step 1: Download the Dataset

1. Open a web browser and go to Inside Airbnb: http://insideairbnb.com/get-the-data/
2. Scroll to down until you find Los Angeles, California, United States
3. Click and download the data file listings.csv.gz

4. Your browser will now show the file being downloaded and say, 'Insecure download blocked'. Click 'Keep', and then 'Show in Folder'.

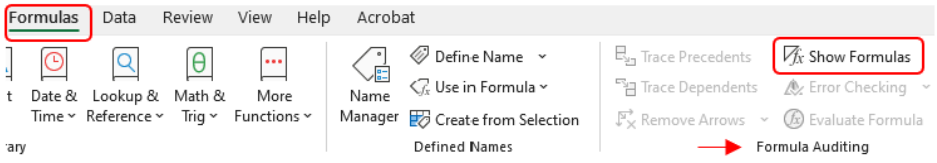


5. Once you select "Show in folder", you will see the file explorer to locate the file and the path. For example: C:\Users\hernanjp\Downloads

6. Extract the file to a folder of your choice and give it an appropriate name like Airbnb_Listings.

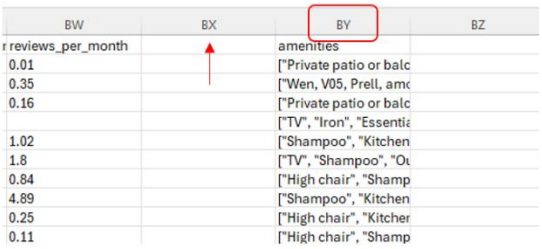
Step 2: Prepare Data in Excel

7. Open file using Excel. You will receive a pop-up message asking you to convert the data, click covert.
8. Go to the Formulas tab in the ribbon, and unselect Show Formulas in the Formula Auditing group as shown below:

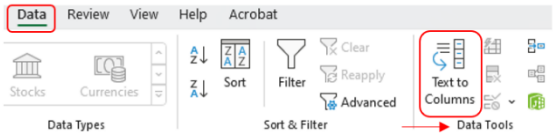


9. Perform the following tasks to create a count on the amenities:
    - Move the amenities column (column "AN") to the end of the dataset by cutting and pasting it into column "BY"

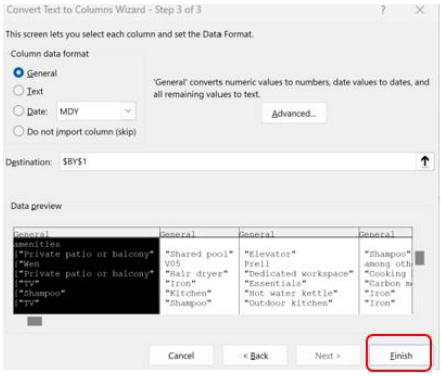
   (Note: This will leave an empty column between the last column of the dataset, and this newly pasted column)



a. While ensuring the entire "BY" column is highlighted, go to the Data tab in the ribbon, then select Text to Columns in the Data Tools group



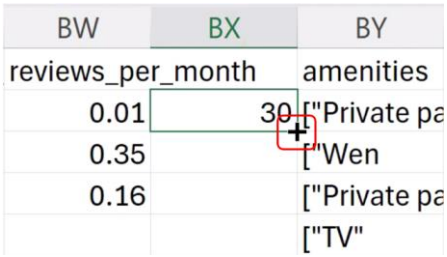
b. Select delimited and click next
c. Select comma as the delimiter and click next
d. Leave default selections and click finish



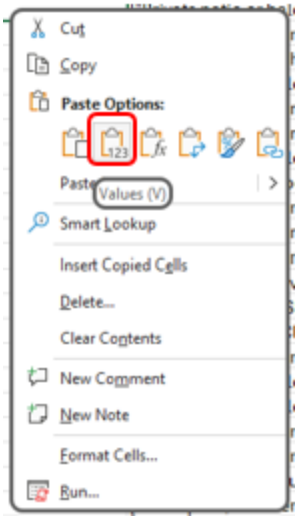
e. The result will be all amenities split into separate columns

10. Click into cell "BX2" to perform amenities count:

a. In cell "BX2" enter the following formula =SUMPRODUCT(COUNTIF(BY2:HD2,BY2:HD2))
b. Press Enter
c. Move cursor to the bottom right corner of cell "BX2" until fill handle cursor (+) appears and double click to autofill down on all rows



d. Highlight the entire "BX" column and select copy
e. Click on column "AN". Right click and select Paste Option: Values(V)

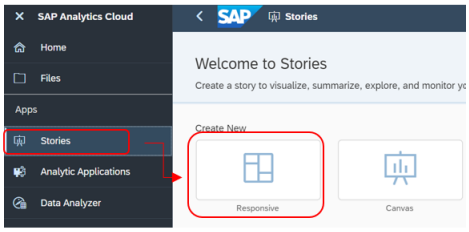


f. Title column “AN” as Amenities
g. Delete columns “BX” through “HD”

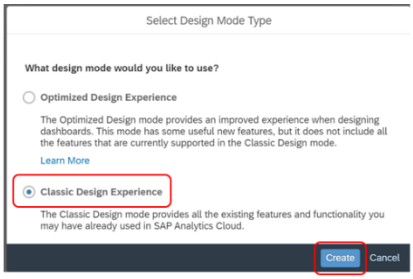
11. Delete host_about (column “O”)
12. Save file.
13. Now the data is ready to be uploaded to SAP.

Step 3: Upload Data to SAP and Clean

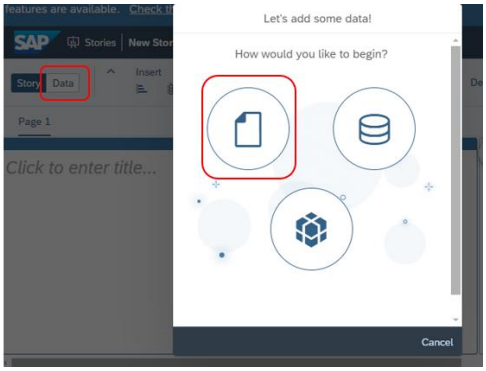
14. Go to SAP Analytics Cloud (SAC): https://higher-education.us10.sapanalytics.cloud
15. Select Stories from the menu on the left side of the screen.
16. Under Create New, select Responsive.



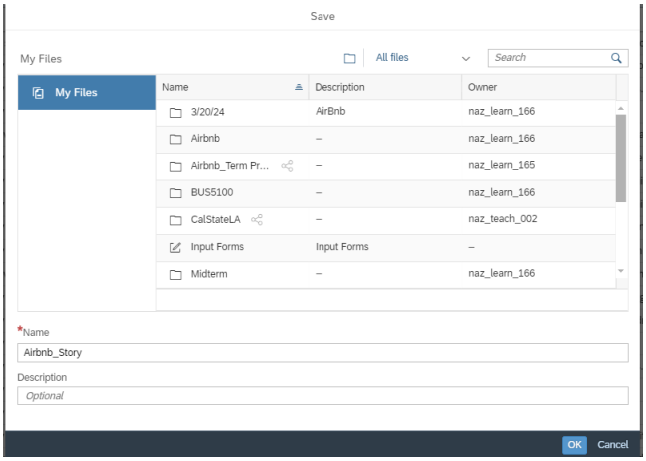
17. Pop-up window “Select Design Mode” will appear.
    a. Select Classic Design Experience
    b. Click Create



18. Click the Data tab at the upper left corner of the canvas. Then, choose Data uploaded from a file



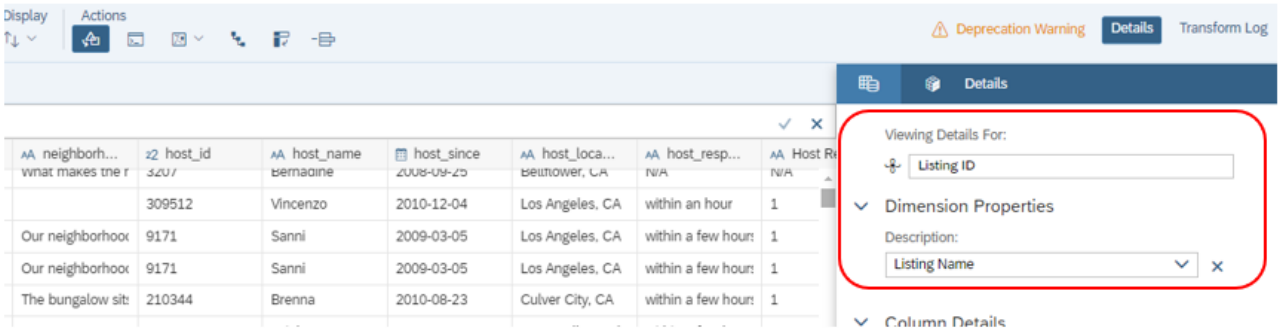
19. In the “Create Dataset from File” pop-up window, click Select Source File
20. Your File Explorer will appear. Navigate to your file location (from Step 6), select file, and click open
21. Ensure ‘Use first row as column headers’ is selected
22. Click Import, file will take about a minute or so to upload
23. Save Story in a location that makes sense to you and give it an appropriate name such as Airbnb_Story
24. Select OK



25. Delete the following columns
    - listing_url
    - scrape_id
    - last_scraped
    - source
    - description
    - neighbourhood_overview
    - picture_url
    - host_url
    - host_location
    - host_response_time
    - host_acceptance_rate
    - host_thumbnail_url

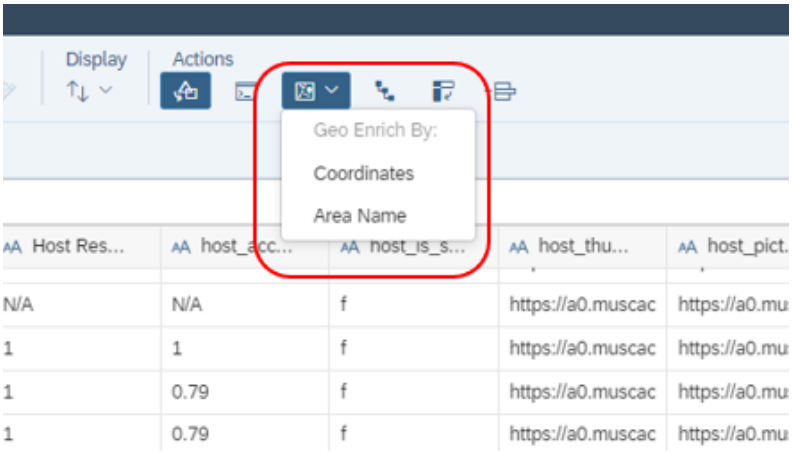
- host_picture_url
- host_neighbourhood
- host_verifications
- host_identity_verified
- neighbourhood
- neighnbourhood_group_cleansed
- Bathrooms
- bedrooms
- minimum_nights
- maximum_nights
- minimum_minimum_nights
- maximum_minimum_nights
- minimum_maximum_nights
- maximum_ maximum _nights
- minimum_nights_avg_ntm
- maximum_nights_avg_ntm
- calender_updated
- has_availabilty
- calender_last_scraped
- number_of_reviews_l10m
- number_of_reviews_l30d
- license
- calculated_host_listings_count
- calculated_host_listings_count_entire_homes
- calculated_host_listings_count_private_rooms
- calculated_host_listings_count_shared_rooms
- reviews per months

26. Rename the following columns:
    a. id à Listing ID
    b. name à Listing Name
    c. neighbourhood_cleansed à Neighbuorhood
    d. room_type à Room Type
    e. beds à Beds
    f. accommodates à Accommodates
    g. host_response_rate à Host Response Rate
    h. price à Price
    i. number_of_reviews à Number of Reviews
    j. host_has_profile_pic à Host Has Profile Pic
27. In the Details side panel, assign a Description to the following:
    a. Listing ID add Description: Listing Name

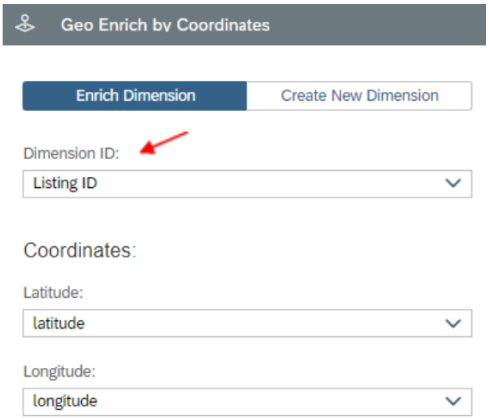


b. host_id add Description: host_name

28. Add a geo enrichment based on latitude and longitude by Listing ID
    a. In the Actions ribbon select Geo Enrichment > Coordinates



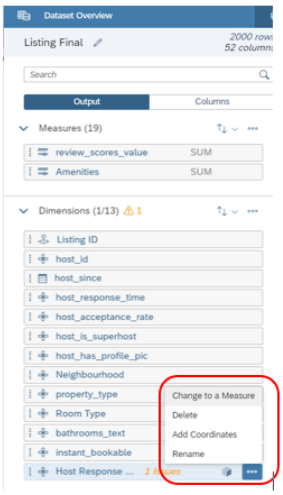
    b. Select the Enrich Dimension tab in the side panel
    c. Select Listing ID as the Dimension ID



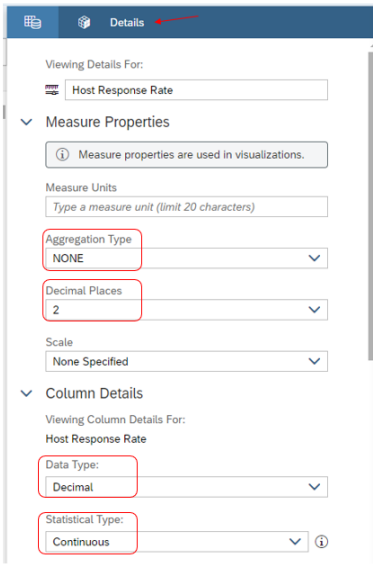
    d. Hit OK at the bottom right corner

29. In the Dataset Overview panel, perform the following actions for Host Response Rate:
    a. Select Host Response Rate (see screenshot below)
    b. Change from Dimension to Measure

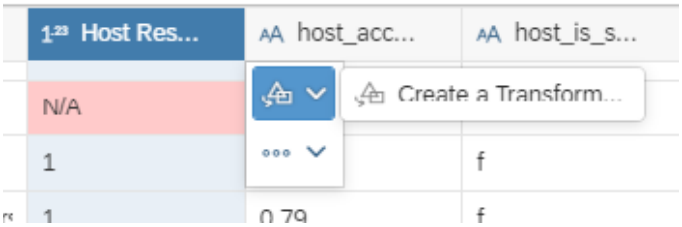
i. Click on (...) for Host Response Rate and a menu will appear. Select Change to a Measure.



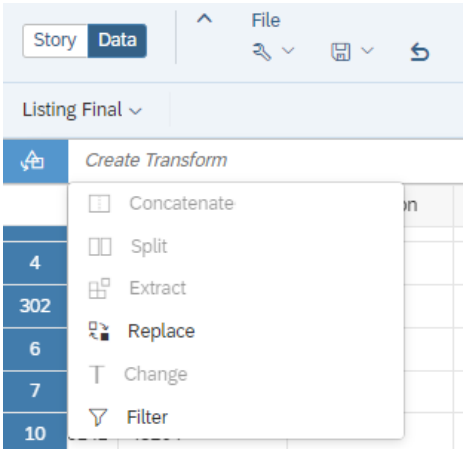
c. Expand the Details panel for Host Response Rate and match the following properties:
   i. Aggregation Type: None
   ii. Decimal Places: 2
   iii. Data Type: Decimal
   iv. Statistical Type: Continuous



d. The column is still indicating an issue, so perform the following steps to resolve:
   i. Select 'Create a Transform'

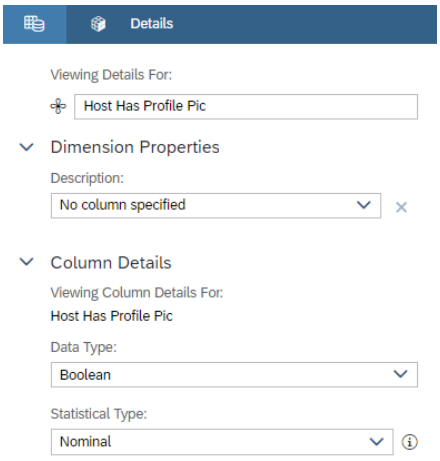


ii. Select ‘Replace’ from the pop-up

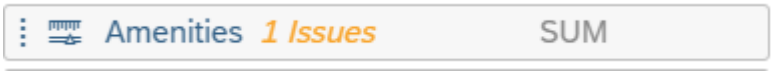


iii. Add ‘N/A’ as the first value and ‘0’ as the second value as follows. Hit the checkmark.

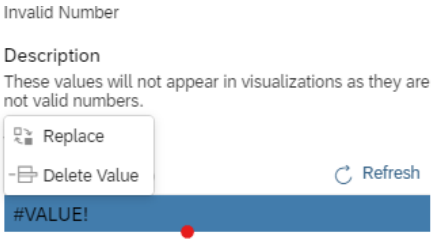
30. In the Dataset Overview panel, perform the following actions for Host Has Profile Pic:
    a. Select Host Has Profile Pic
    b. Expand the Details panel and match the following properties:
        i. Data Type: Boolean
        ii. Statistical Type: Nominal



31. Click Validate Full Dataset on bottom right corner of the Dataset Overview panel
    a. 2 issues should come up. 1 for Amenities and 1 for Host Has Profile Pic
32. In the Dataset Overview panel, click on '1 issue' next to Amenities



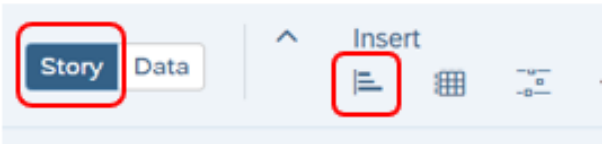
    a. The details panel will expand and the issue can be seen at the bottom
33. Select '#VALUE' and Delete Value



34. Repeat steps 18 and 19 for Host Has Profile Pic
35. Save. Do not close.

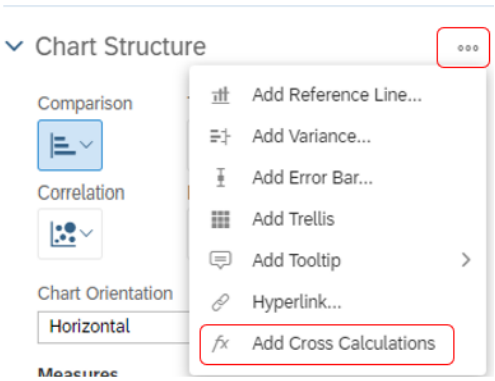
Step 4: Create a Bar Graph

36. Click on the Story tab, and then click on Insert > Chart.

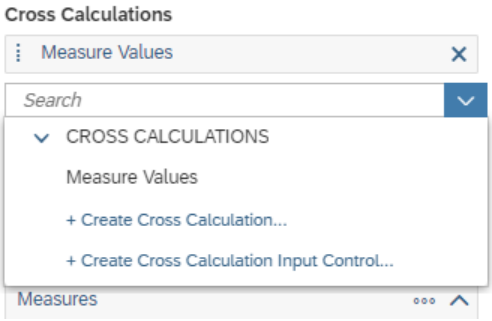


37. In the Builder panel, select Comparison > Bar/Column

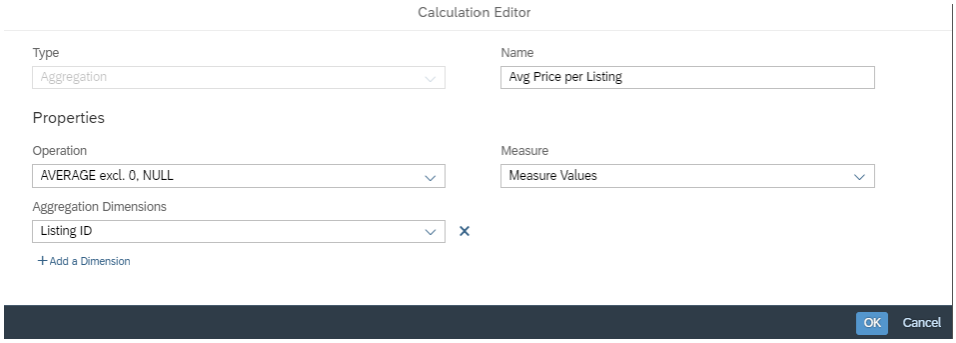
38. Chart Orientation should be Horizontal by default, if not change to Horizontal
39. Assign the following:
    a. Measures: Price
    b. Dimensions: Neighbourhood
40. Next to Chart Structure in the Builder panel, click (...) for 'Add/Remove Chart Component' and select 'Cross Calculations' from the menu that appears as shown in image below



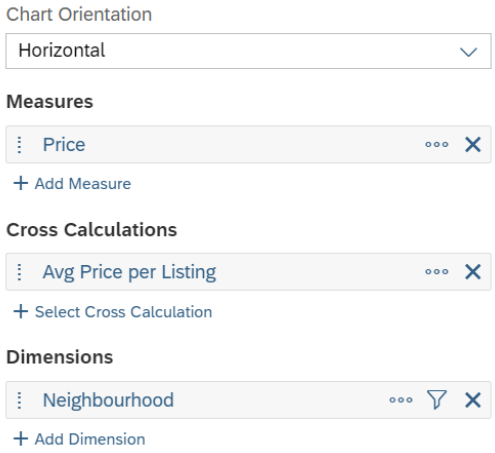
41. Under Cross Calculations click '+ Select Cross Calculation' > '+ Create Cross Calculation'



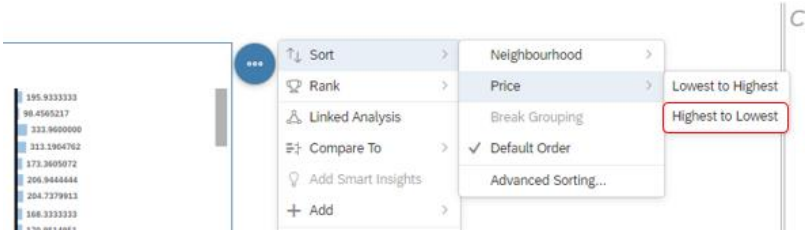
a. In the Calculation Editor pop-up, expand the dropdown menu for Type and select Aggregation
b. Enter the following values:
    iii. Name: Avg Price per Listing
    iv. Operation: AVERAGE exclude 0, NULL
    v. Aggregation Dimensions: Listing ID
c. Click OK



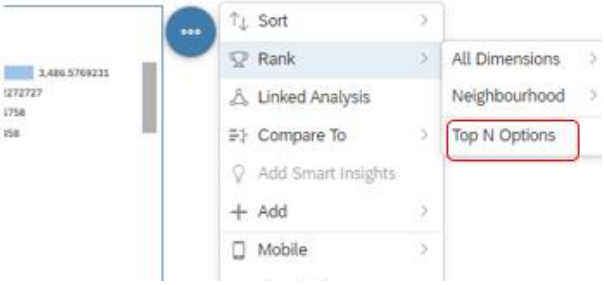
42. The Builder panel should look like the following image:



43. Sort the Price values
    a. Go to (...) More Actions on the chart > Sort > Price > Highest to Lowest

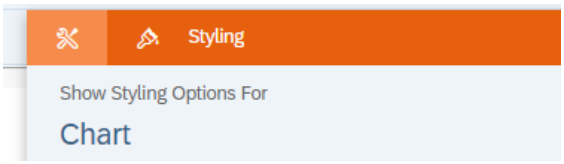


44. Assign a Rank
    a. Go to (...) More Actions on the chart > Rank > Top N Options

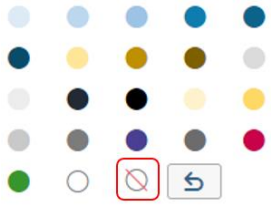


    b. Change Value to 10, select apply

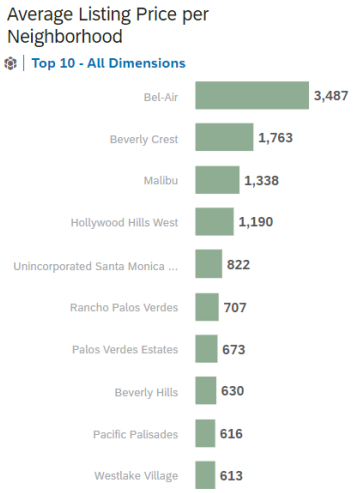
45. Apply styling
    a. Expand the Styling panel



    b. Scroll down to Fill Color and select Green
    c. Scroll down to Decimal and select 0 from the dropdown menu
    d. Scroll down to Axis > Axis Line Color. Select no color

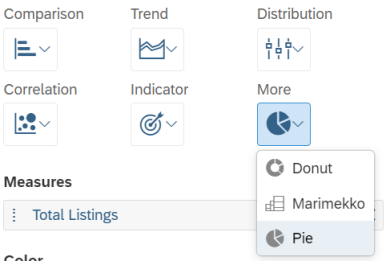


46. Rename bar chart
    a. Double-click on existing title name
    b. Rename to: Average Listing Price per Neighborhood
47. The bar chart should look as follows (you can enlarge the visualization by clicking and dragging the bottom right corner of the graph):



Step 5: Create a Pie Chart

48. Select lane 2, right-click and select Remove. We do not require separate lanes for this page.
49. Insert a new chart by choosing Insert > Chart
50. In the Builder panel, select More > Pie

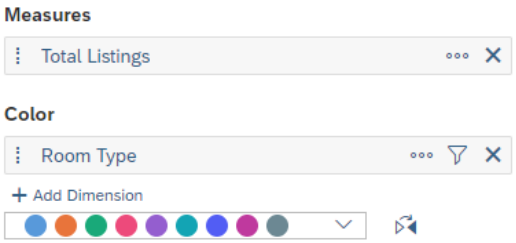


51. Under Measures select + Add Measure > + Create Cross Calculation
    a. In the Calculation Editor pop-up, expand the dropdown menu for Type and select Aggregation
    b. Enter the following values:
        vi. Name: Total Listings
        vii. Operation: Count Dimensions

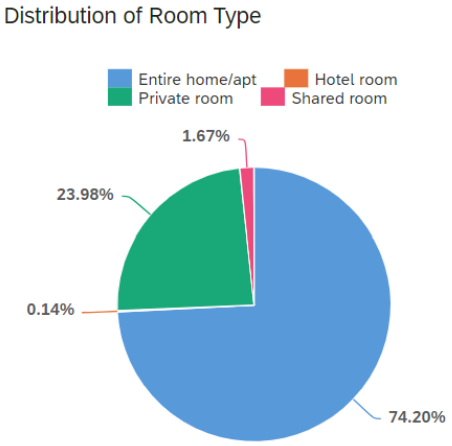
viii. Aggregation Dimensions: Listing ID

c. Click OK

52. Under Color, select Room Type as the dimension
53. Select a high contrast color palette such as the second one on the drop-down menu
54. Your Builder panel should look as follows:

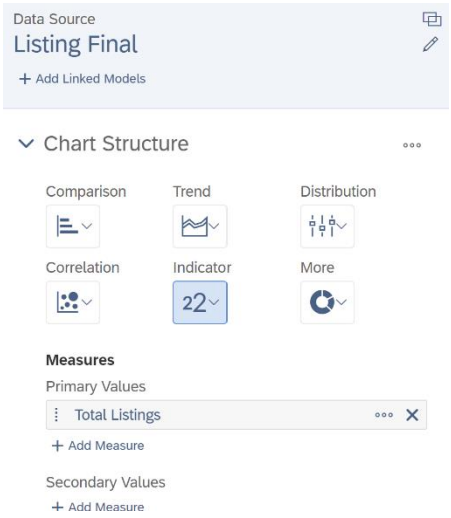


55. Rename pie chart: Distribution of Room Type
56. The following pie chart will appear (you may have to enlarge visualization to see the legend):

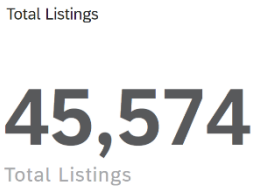


57. As an added element, we will build a count of the total listings. Insert a new chart by choosing Insert Chart at the top.
58. In the Builder panel, select Indicator > Numeric Point

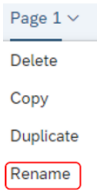
a. Primary Value Measure: Total Listings

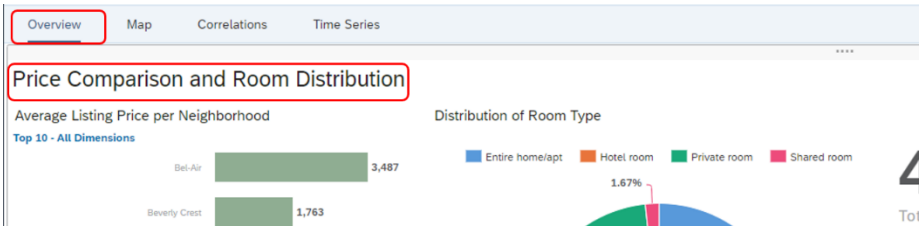


59. In the Styling panel, scroll down to Decimal Places and select “0”
60. The chart will look like the screenshot below:



61. Create a Linked Analysis
    a. Click on the bar chart
    b. Click on (...) More Actions > Linked Analysis
    c. In the Linked Analysis panel select ‘All Widgets on the Page’ and check ‘Filter on Data Point Selection’
    d. Click Apply
    e. Repeat steps a through d for the Pie Chart and Numeric Chart
62. Give this lane a title: Price Comparison and Room Distribution
63. Next to ‘Page 1’ select arrow down and click Rename
64. Rename page: Overview

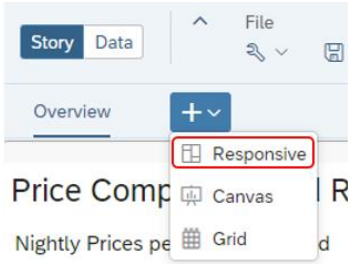




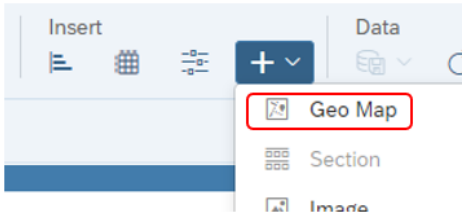
65. Save. Do not close SAC.

Step 6: Create a Geo Map

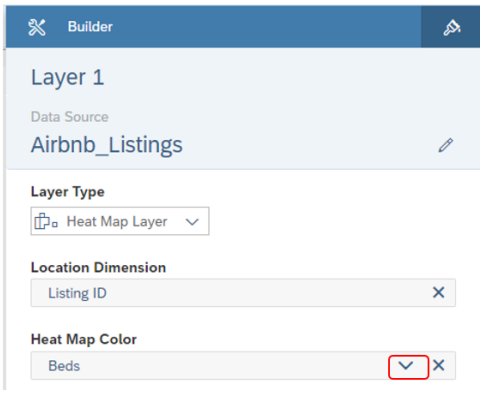
66. Hover over the area next to the “Overview” page tab and select + > Responsive, to add another page.



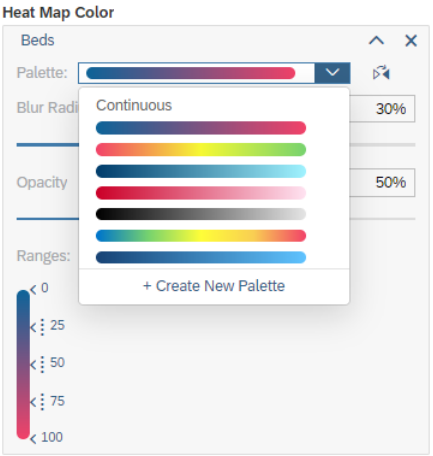
67. Delete one of the lanes by right-clicking and selecting Remove
68. Insert a Geo Map by choosing the Insert + > Geo Map



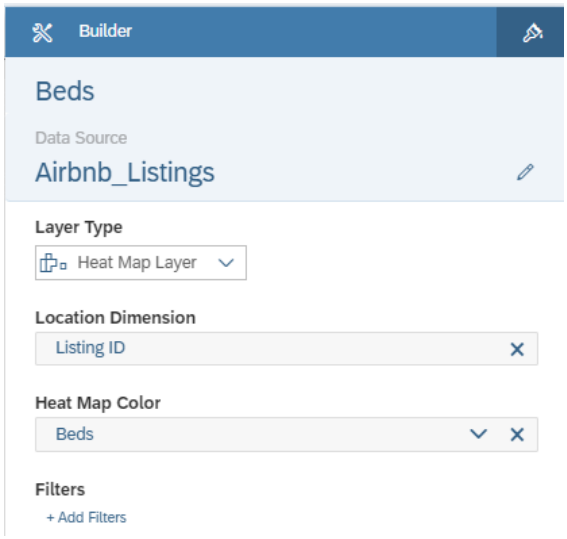
69. In the Builder Panel, create Heat Layer:
    a. Click + Add Layer
    b. Assign the following:
        i. Layer Type: Heat Map Layer
        ii. Location Dimension: Listing ID
        iii. Heat Map Color: Beds
    c. Expand the Heat Map Color section by clicking the downward arrow



    d. Apply the following:
        i. Change Color to Blue to Red
        ii. Blur Radius to 30%
        iii. Opacity to 50%
        iv. Range to 4

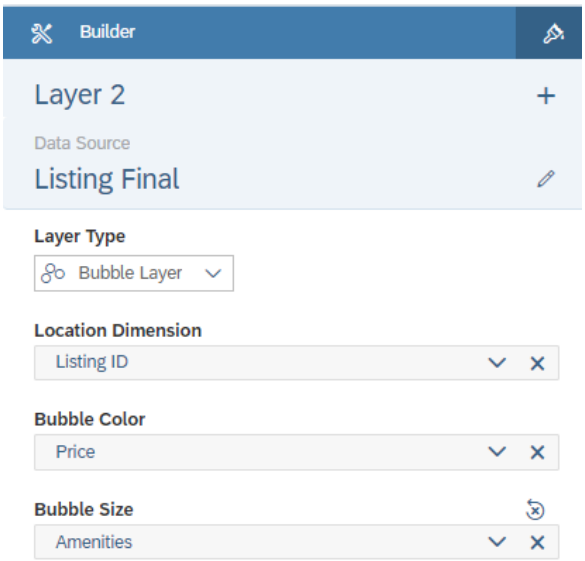


e. At the top of the Builder panel, rename from Layer 1 to: Beds

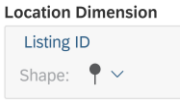


f. Select OK

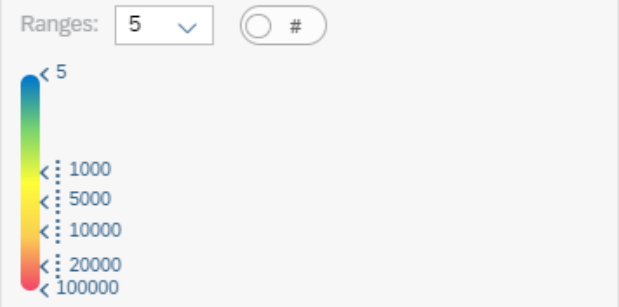
70. Under Content Layers, click + Add Layer
    a. Assign the following:
        i. Layer Type: Bubble Layer (default)
        ii. Location Dimension: Listing ID
        iii. Bubble Color: Price
        iv. Bubble Size: Amenities



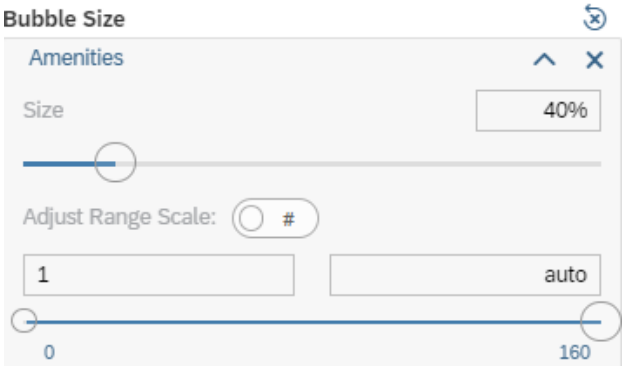
71. Expand the Location Dimension section by clicking the downward arrow
    a. Select the pin as the shape



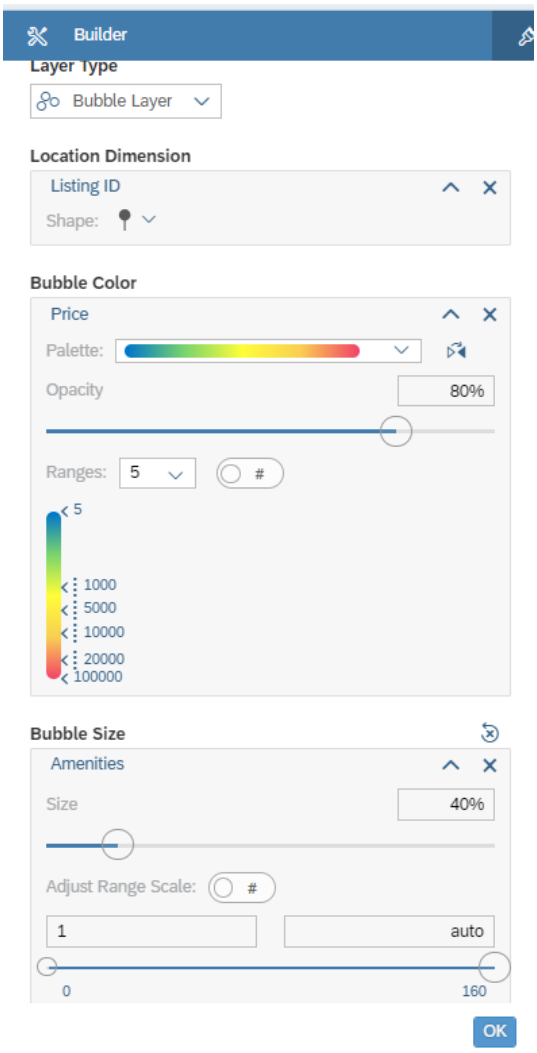
72. Expand the Bubble Color section by clicking the downward arrow
    a. Select the color Palette: Blue, Green, Yellow, Red
    b. Change Opacity to 80%
    c. Range to 5
    d. Toggle “Switch to use Actual Values”
    e. Enter range values and arrange on the scale
        i. Double click on the default value and enter values (5, 1000, 5000, 10000, 20000, 100000) as shown below
        ii. Click the ⋮ to adjust the values along the scale as shown below



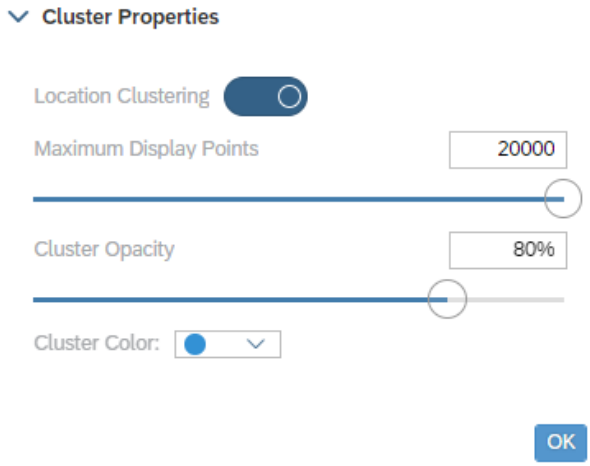
73. Expand the Amenities section by clicking the downward arrow
    a. Change Size to 40%
    b. In “Adjust Range Scale” toggle to “Switch to use Actual Values”
        i. Enter 1 to Auto



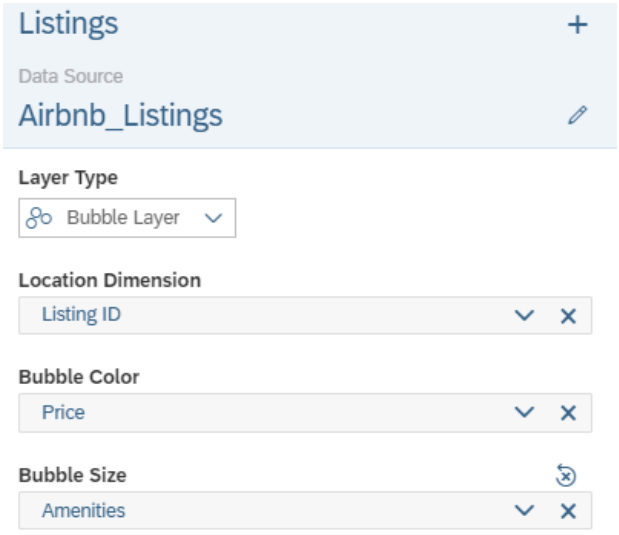
74. Your Builder panel should look like the following:



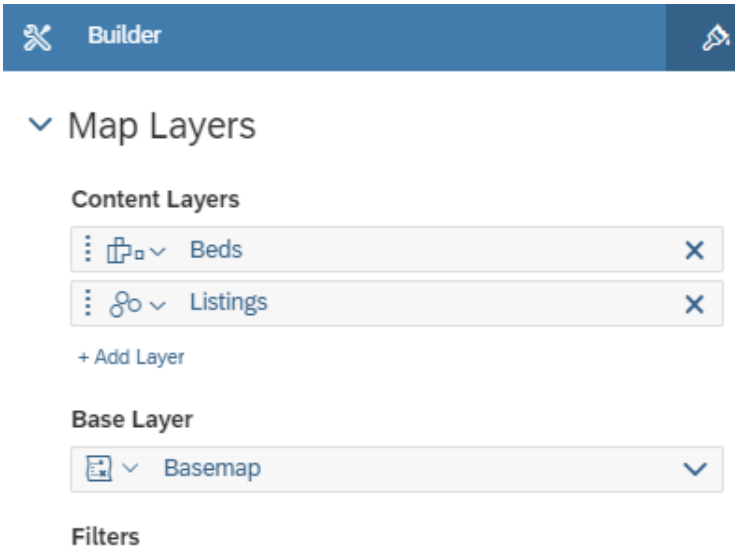
75. Scroll down to Cluster Properties
    a. Set Maximum Display points to 20,000
    b. Cluster Opacity to 80%



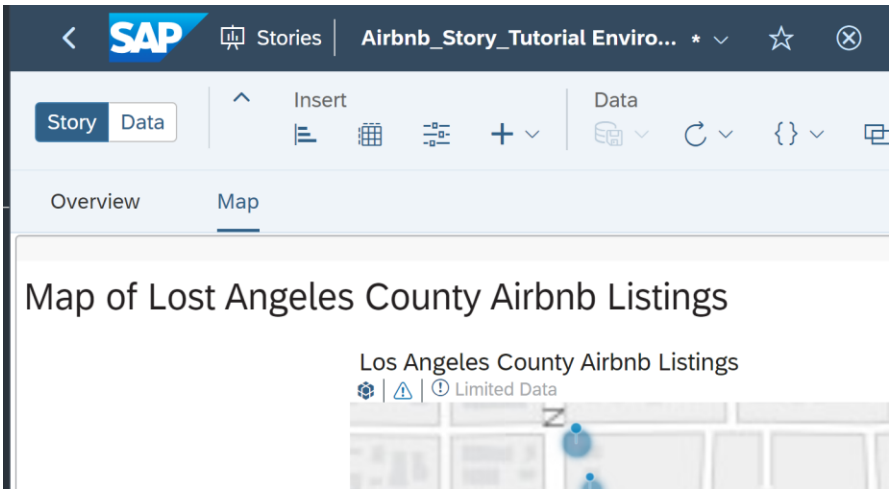
76. Rename Layer 2 to Listings



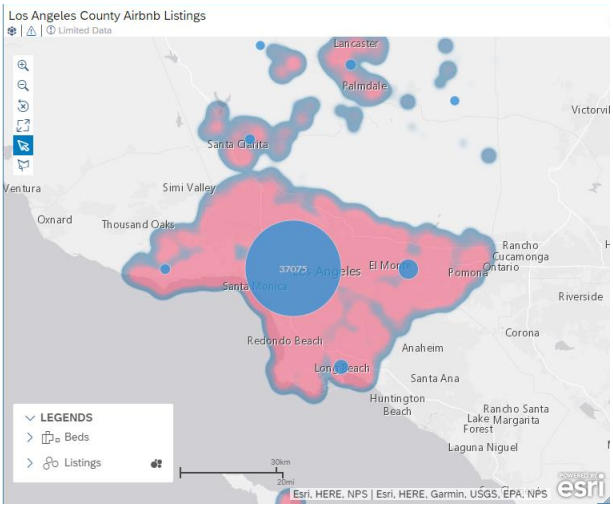
77. Select OK at the bottom of the Builder panel
78. Your Builder panel should look as follows



79. Click drop down arrow next to “Page 2” and rename tab to Map
80. Click the top of the panel on ‘Click to enter title...’ and title it “Map of Los Angeles County Airbnb Listings”



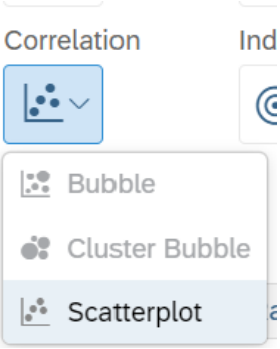
81. The Geo Map should look as follows:



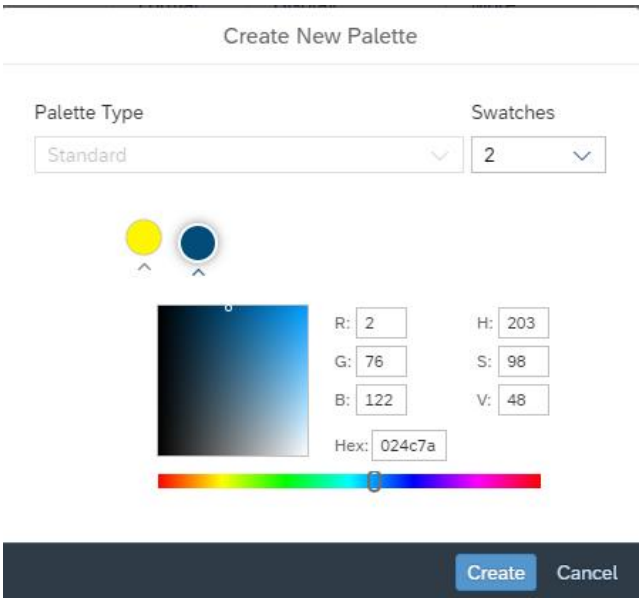
82. You can drag the bottom right corner to enlarge the visualization
83. Save. Do not close SAP (SAC).

Step 7: Create a Scatterplot

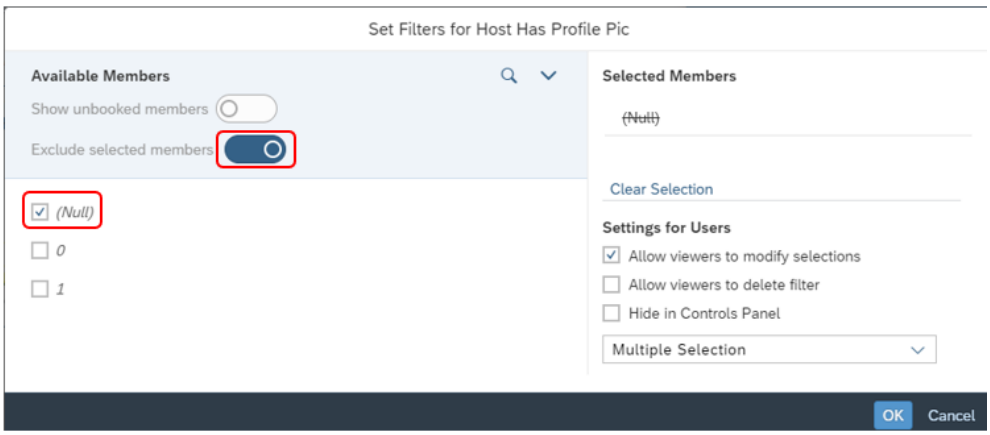
84. Add a Responsive page by hovering next to the Map page tab
    a. Click + > Responsive
    b. Rename new tab as Correlations
85. Select the first lane (lane on the left) and rename as Scatterplot
86. Insert a new chart by choosing Insert Chart at the top.
87. Inside the Builder panel click Correlation > Scatterplot



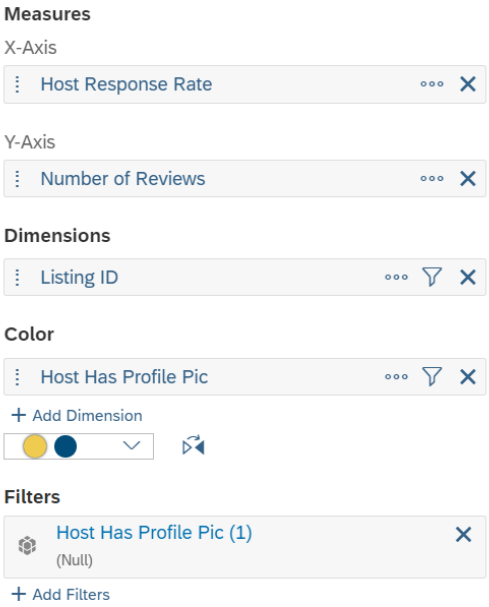
88. Assign the following:
    a. Measurements:
        i. X-Axis: Host Response Rate
        ii. Y-Axis: Number of Reviews
    b. Dimension: Listing ID
    c. Color > + Dimension: Host Has Profile Pic
        i. Under this section, expand the color palette drop down menu and select '+ Create New Palette'.
        ii. In the pop-up window, create a palette of 2 swatches with high contrasting colors. We chose a shade of yellow and blue.



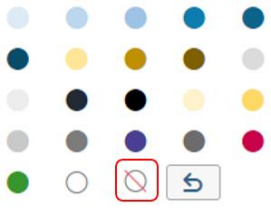
        iii. Click Create
89. In the Filter section of the builder panel, select + Add Filters > "Host Has Profile Pic (Member)"
90. Toggle Exclude Selected Member
91. Check: (null)
92. Click OK



93. Your builder should look as follows:



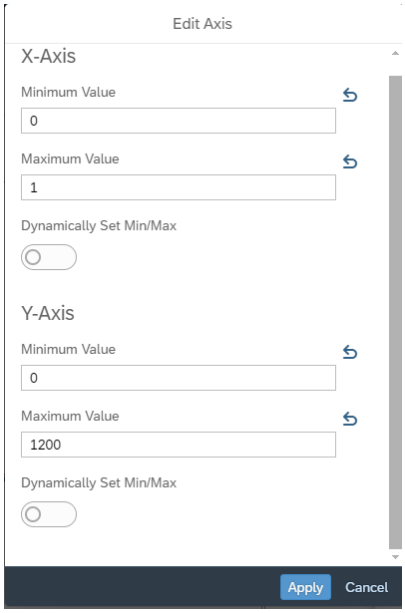
94. Expand the styling panel and apply the following changes:
    a. Data Marker Size: 80%
    b. Axis Line Color: No Color



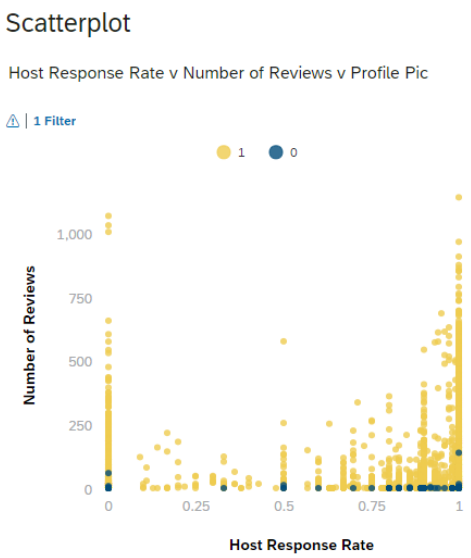
95. Edit X and Y Axis values
    a. Click on the graph and select (...) More Actions > Edit Axis
        i. In Edit Axis Window, enter the following values:
            - X-Axis:
                - Minimum Value: 0
                - Maximum Value: 1

- Y-Axis
  - Minimum Value: 0
  - Maximum Value: 1200

ii. Click Apply



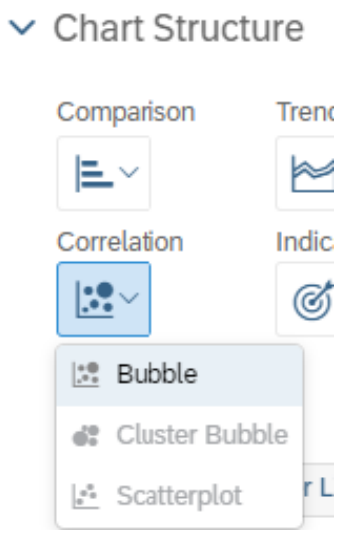
96. Rename Chart: Host Response Rate v Number of Reviews v Profile Pic
97. The final result should look like this:



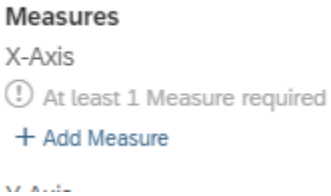
Step 8: Create a Bubble Graph

98. Select the second lane (lane to the right)
99. Insert a new chart by choosing Insert Chart at the top.

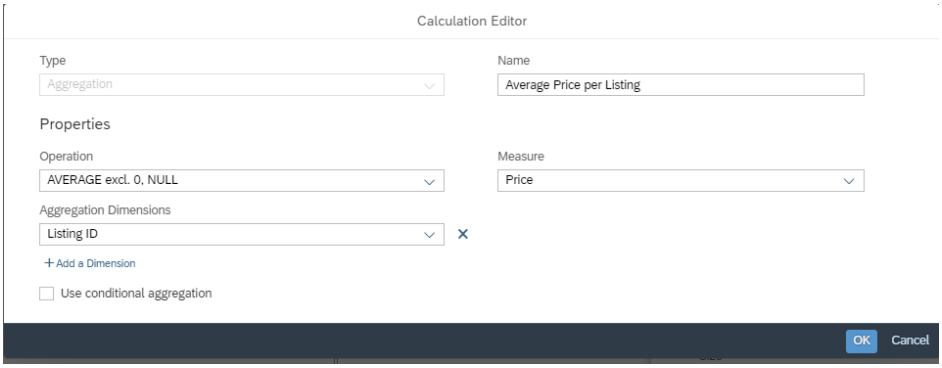
100. Inside the Builder panel select Correlation > Bubble



101. Click Add Measure under X-Axis

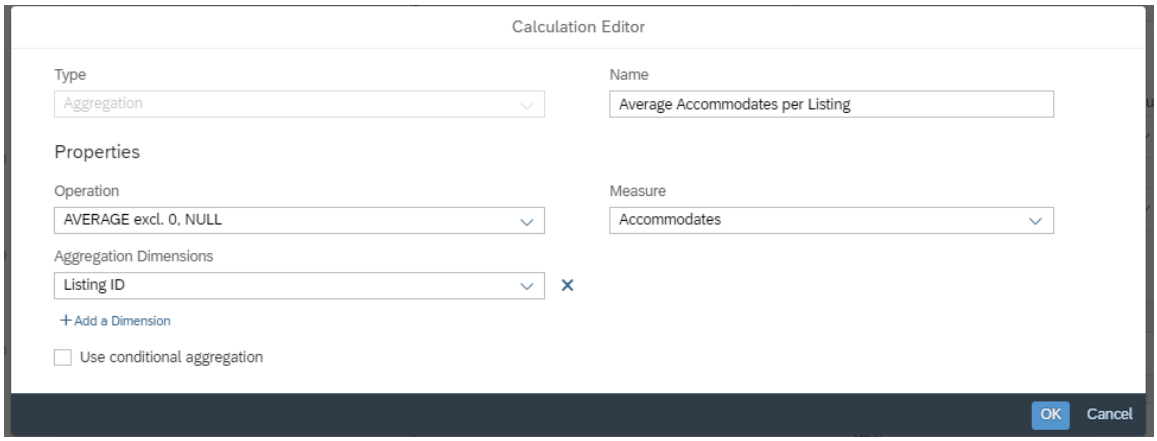


102. Select '+ Create Calculation…'
   a. Assign the following values:
      i. Type: Aggregation
      ii. Name: Average Price per Listing
      iii. Operation: AVERAGE exclude 0, NULL
      iv. Measure: Price
      v. Aggregation Dimensions: Listing ID
   b. Click OK

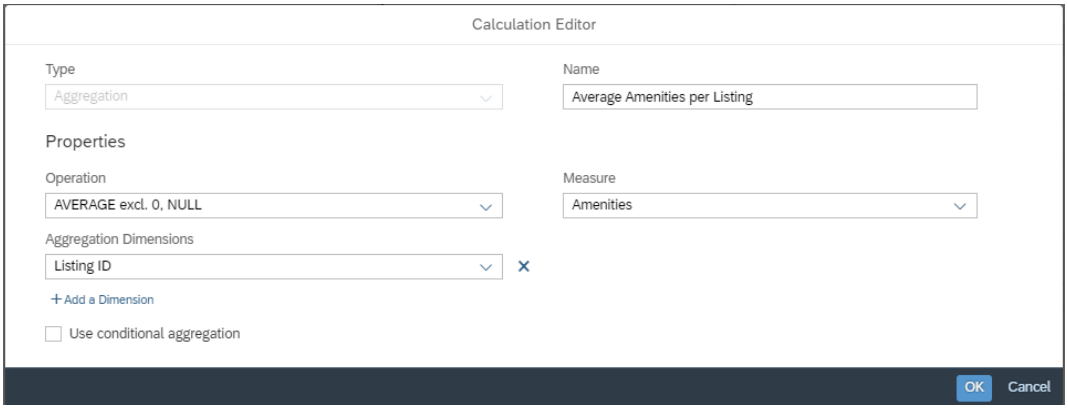


103. Click Add Measure under Y-Axis
104. Select '+ Create Calculation...'

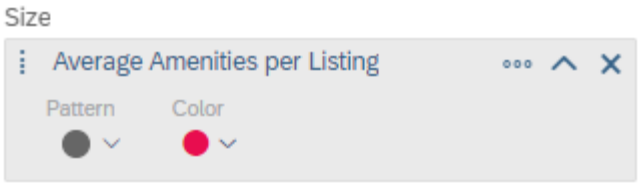
a. Assign the following values:
    i. Type Aggregation
    ii. Name: Average Accommodates per Listing
    iii. Operation: AVERAGE exclude 0, NULL
    iv. Measure: Accommodates
    v. Aggregation Dimensions: Listing ID
b. Click OK



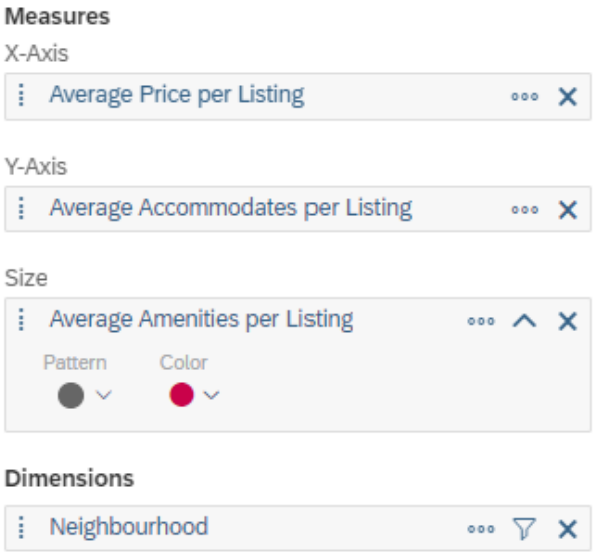
105. Click +Add Measure under Size
106. Select '+ Create Calculation...'
a. Assign the following values:
    i. Type Aggregation
    ii. Name: Average Amenities per Listing
    iii. Operation: AVERAGE exclude 0, NULL
    iv. Measure: Amenities
    v. Aggregation Dimensions: Listing ID
b. Click OK



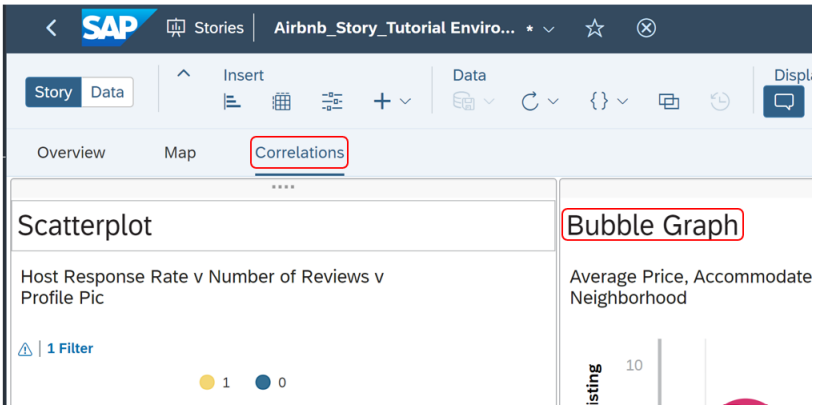
107. In the same section for Size, change color to Red



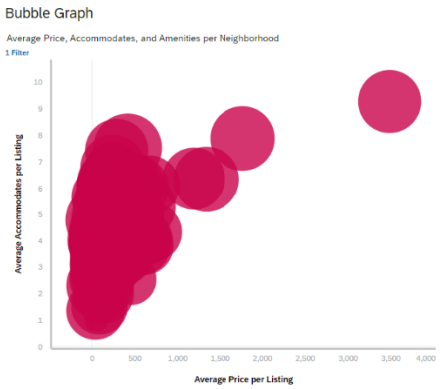
108. Scroll down to Dimensions and add Neighbourhood
109. Your builder should look as follows:



110. Click drop down arrow next to “Page 3” and rename tab to Correlations
111. Click the top of panel on ‘Click to enter title…’ and title it Bubble Graph



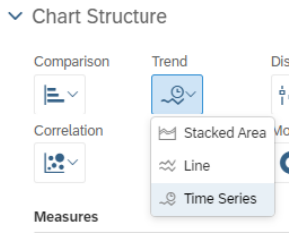
112. The bubble graph should look as follows:



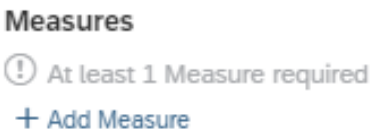
113. Save. Do not close SAP (SAC).

Step 9: Create a Time Series

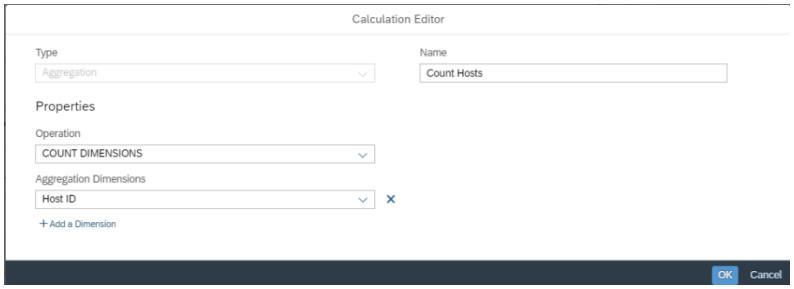
114. Add a Responsive page by hovering next to the Correlations page tab
    a. Click + > Responsive
    b. Rename new tab as Time Series
115. Delete one of the lanes by right-clicking and selecting Remove
116. Insert a new chart by choosing Insert Chart at the top.
117. Inside the Builder panel under Trend, select Time Series



118. Under Measures select +Add Measures



119. Select '+ Create Calculation...'
    a. Type Aggregation
    b. Name: Count Hosts
    c. Operation: Count Dimensions
    d. Aggregation Dimensions: Host ID
    e. Click OK



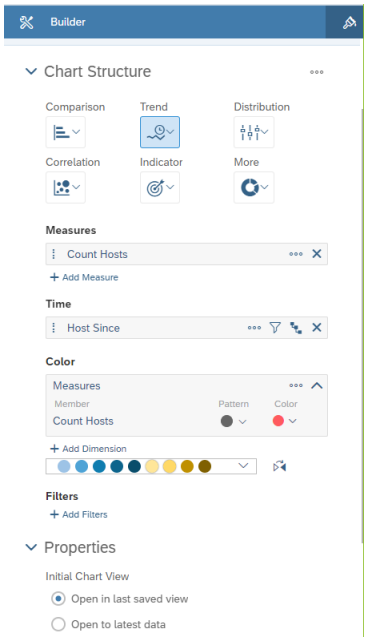
120. Under Time select Host Since

121. Under Color select Count Hosts

a. Select color: pink

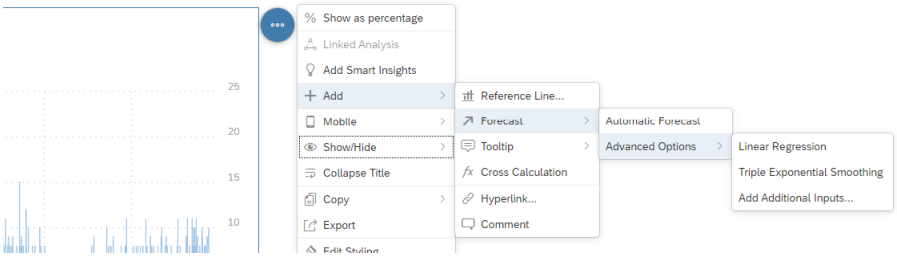
122. Under Properties, change Initial Chart View to Open in last saved view

123. Your builder should look as follows:

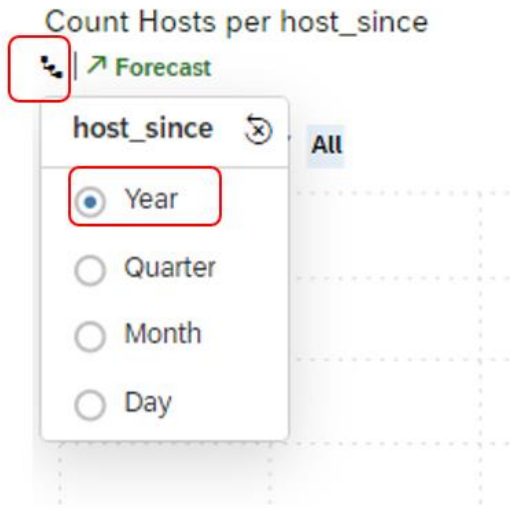


124. Add a Forecast to the chart

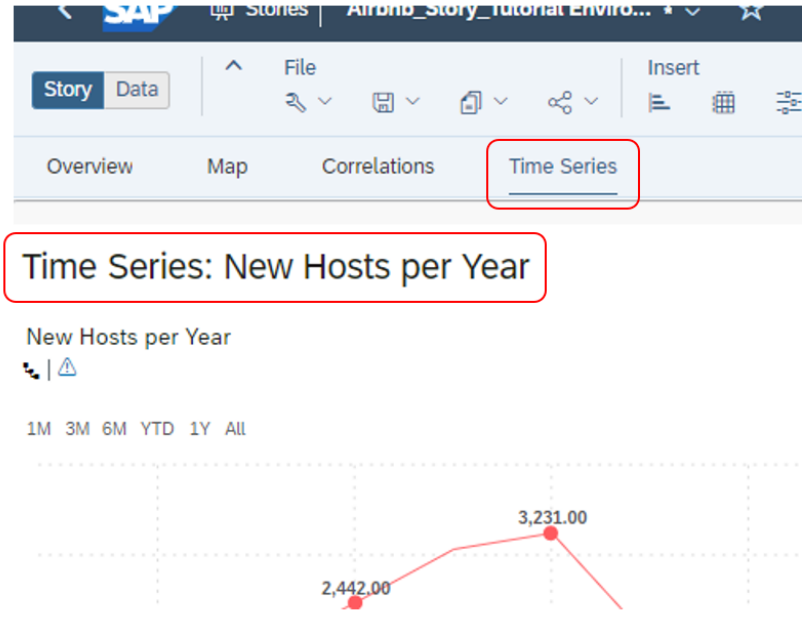
a. Click on (…) More Actions > + Add > Forecast > Advanced Options > Linear Regression



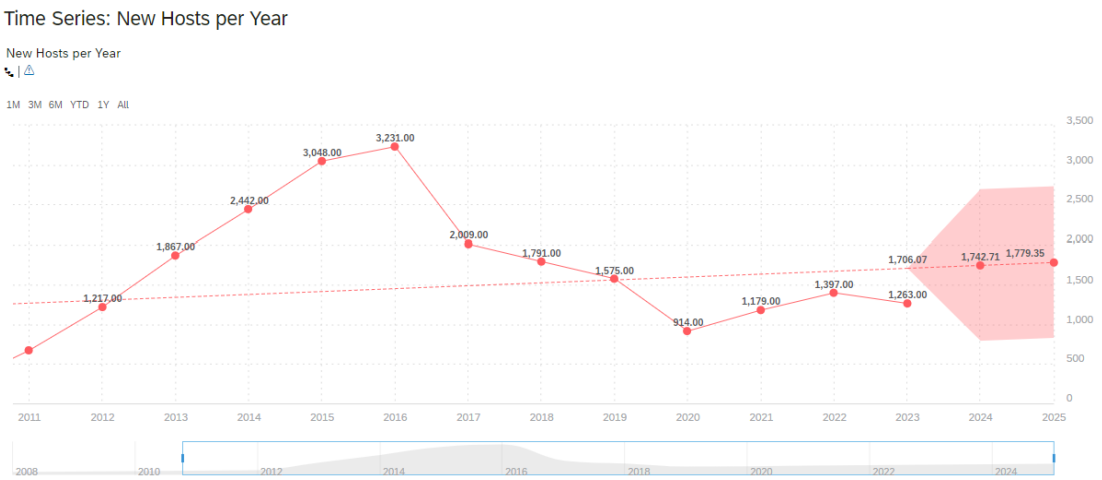
125. In the chart, click set drills and select Year



126. Enlarge visual by dragging the bottom right corner of chart to see full values
127. Click drop down arrow next to “Page 4” and rename tab to Time Series
128. Click the top of the panel on ‘Click to enter title...’ and title it “Time Series: New Hosts per Year”



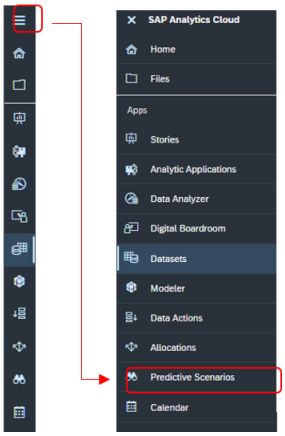
129. The following time series will appear:



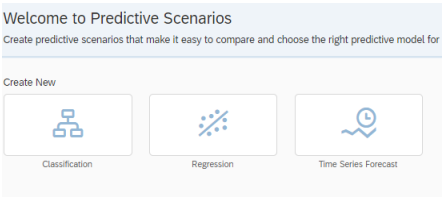
130. Save. Do not close SAP (SAC).

Step 10: Create a Regression Analysis

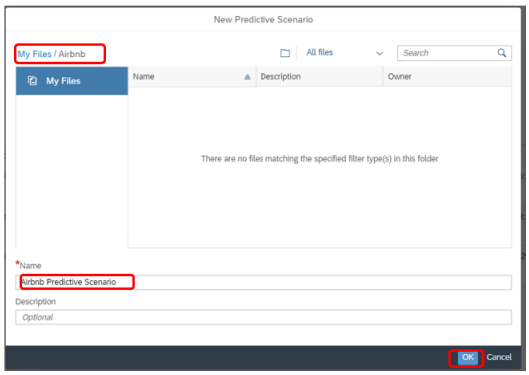
131. Go to the menu, and select Predictive Scenarios



132. Select Regression

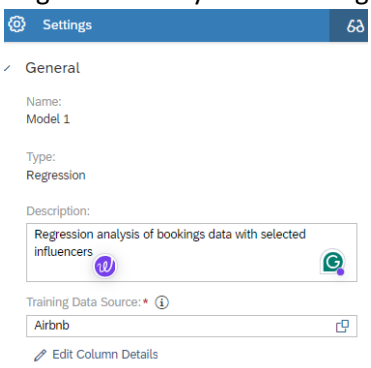


133. Save your Regression Model in your Airbnb folder as Airbnb Predictive Scenario, and select OK. That is My Files > Airbnb

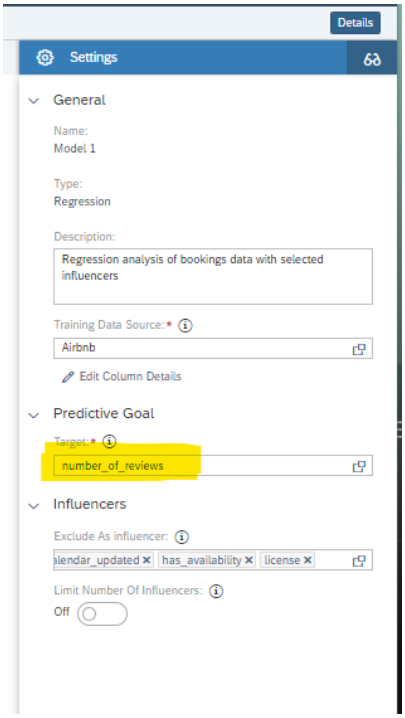


134. Select the dataset “Listings” from your Airbnb folder

135. Fill in the description as “Regression analysis of bookings data with selected influencers”

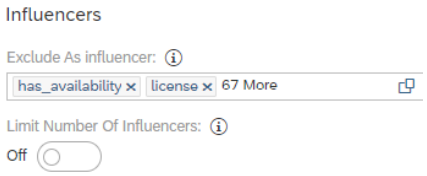


136. Select number_of _reviews as the predictive goal target

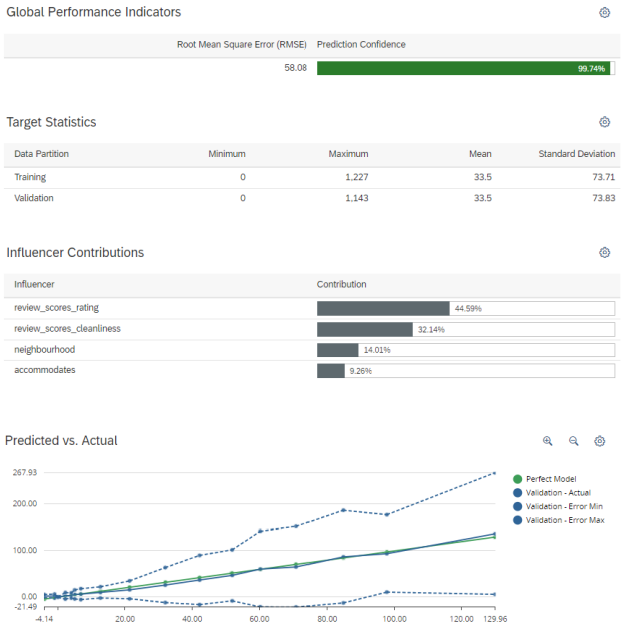


137. Exclude all influencers except for the following:
    a. Review_scores_rating
    b. Review_scores_cleaniness
    c. Neighbourhood

d. Accommodates
e. Instant_bookable



138. Select OK
139. Select Train
140. The following regression analysis will appear:



This is the end of the lab

References

1. URL of Data Source: https://insideairbnb.com/get-the-data/
2. URL of your Dataset: http://data.insideairbnb.com/united-states/ca/los-angeles/2023-12-03/visualisations/listings.csv
3. URL of Github: https://github.com/salazarjp/airbnbgroup4
4. URL of SAP Analytics Cloud Story and Model Location: https://higher-education.us10.sapanalytics.cloud/sap/fpa/ui/tenants/81639#view_id=contentLib;resourceId=B9988A08D89F7BA8E5A912D5AB36BAA3;resourceType=FOLDER