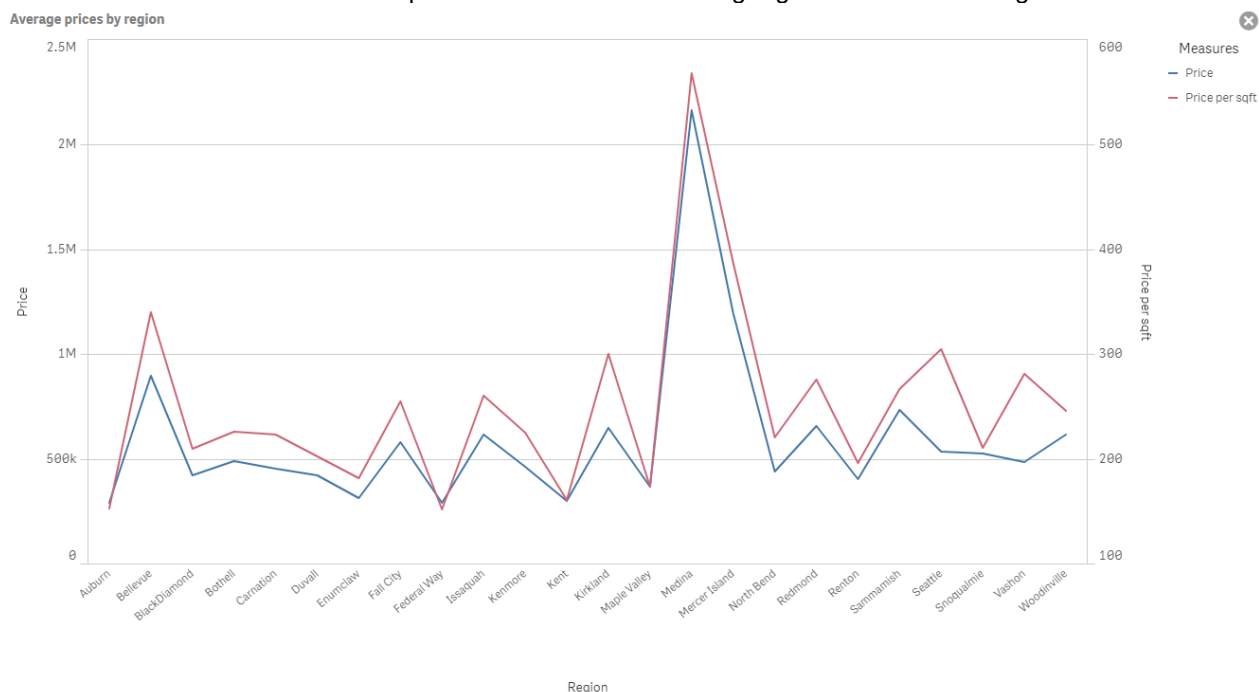


In order to perform reliable visualizations I performed Exploratory Data Analysis and cleansed the data before further analysis. As a result:

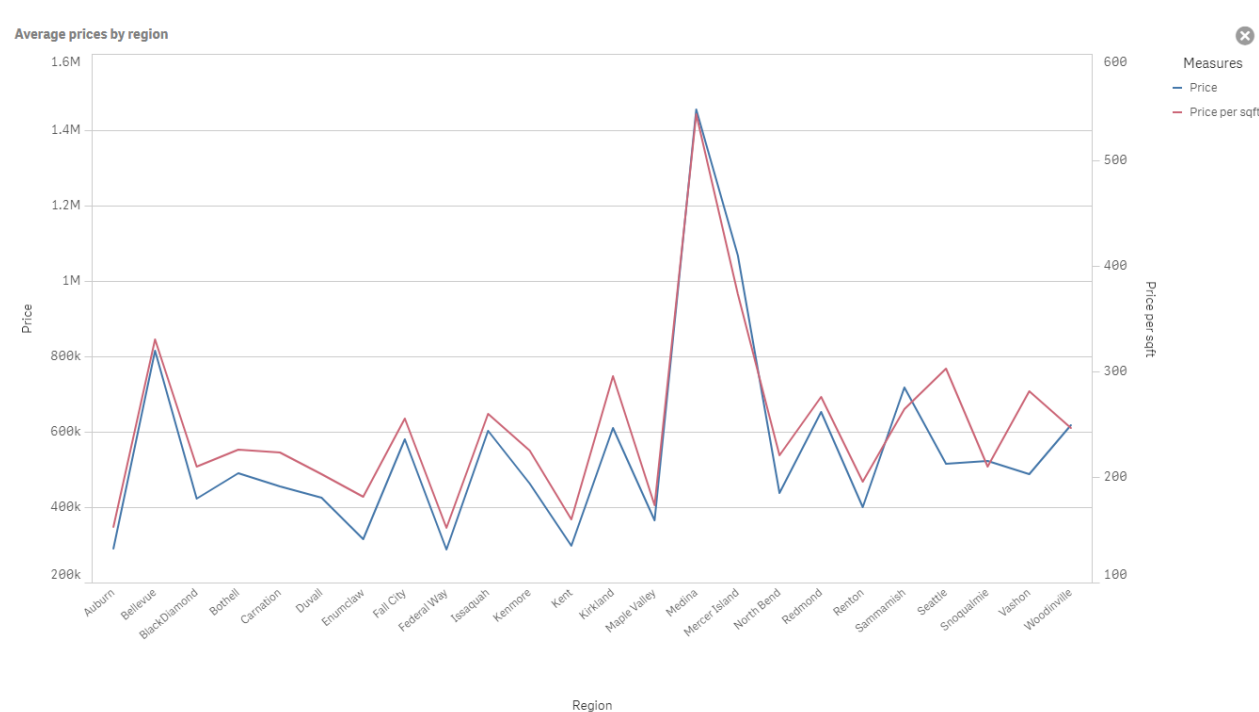
-I identified a record with 33 bedrooms. The price (640 k), living space (1620 sqft) and bathroom count (1.75) indicated that '33' is actually a typing error. I changed the number of bedrooms to 3.

-As the goal is to identify as much information as possible before moving to the region, I do not want to focus on outliers and very unusual cases. The bottom price is already limited by 75 k, which seems reasonable. However, upper prices can go up high without any limits. I decided to drop 1 % (216 rows) of the most expensive houses. I still get visualizations for 99 % of the data, which should correspond to the actual state of market.

Here's how the prices were distributed among regions before cleansing:



Here's the result after cleansing:



Dropping 1 % of the most expensive records has not affected the data in a way that can be easily noticed in these visualizations. Therefore, I decided to continue with 99 % of the records.

To create more insight I have also created a new field:
-price_per_sqft_living-Price per square foot of living space.

This was just a short note to introduce you to the small changes in the dataset. You are now all set to run the up. Have fun with the visualizations!