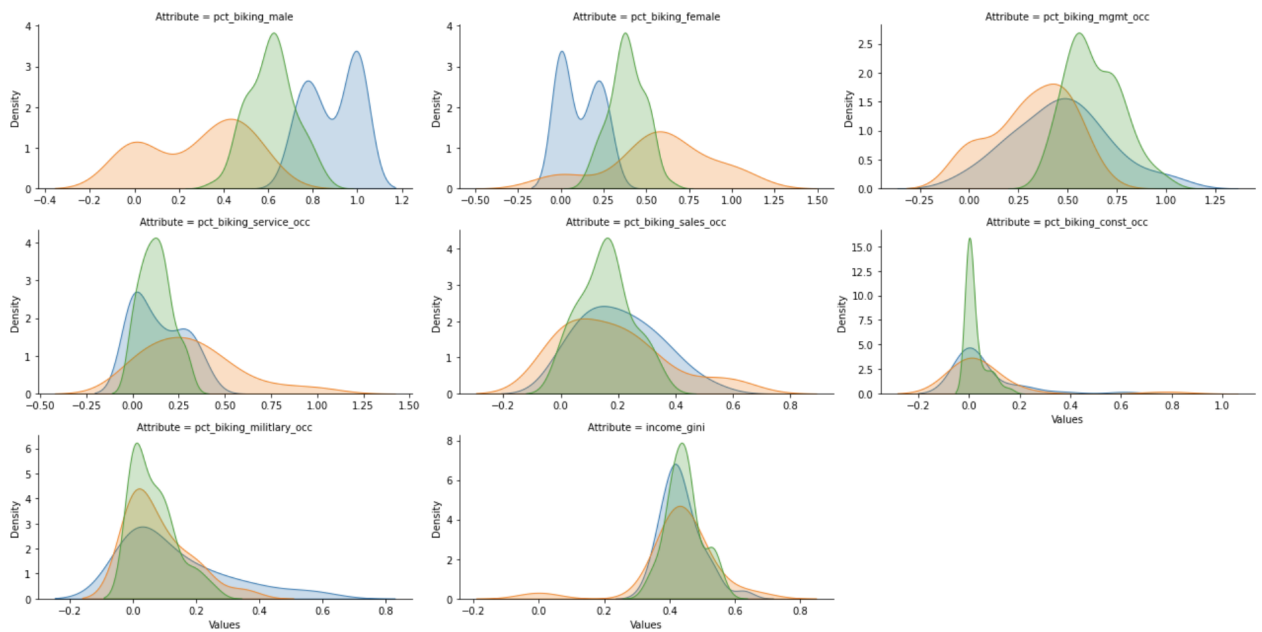
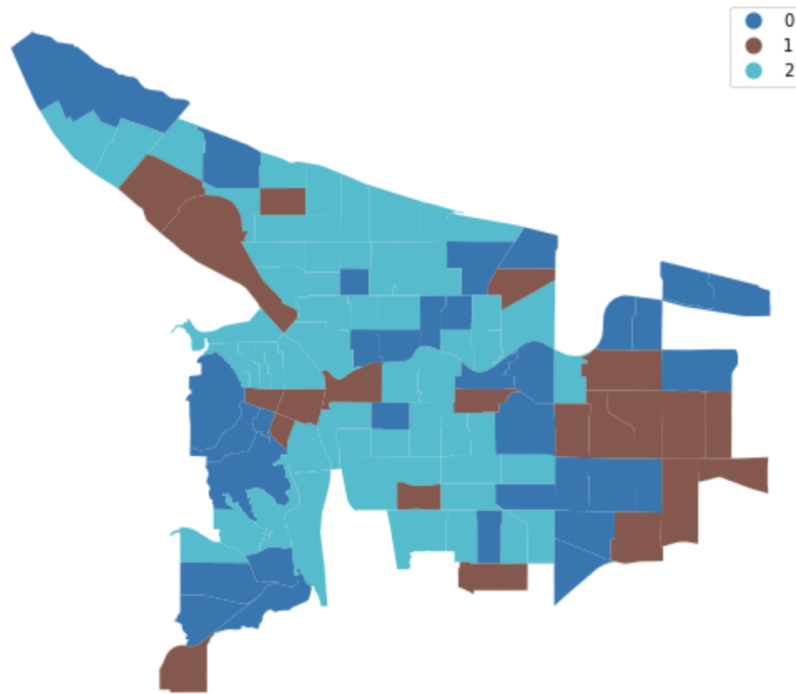


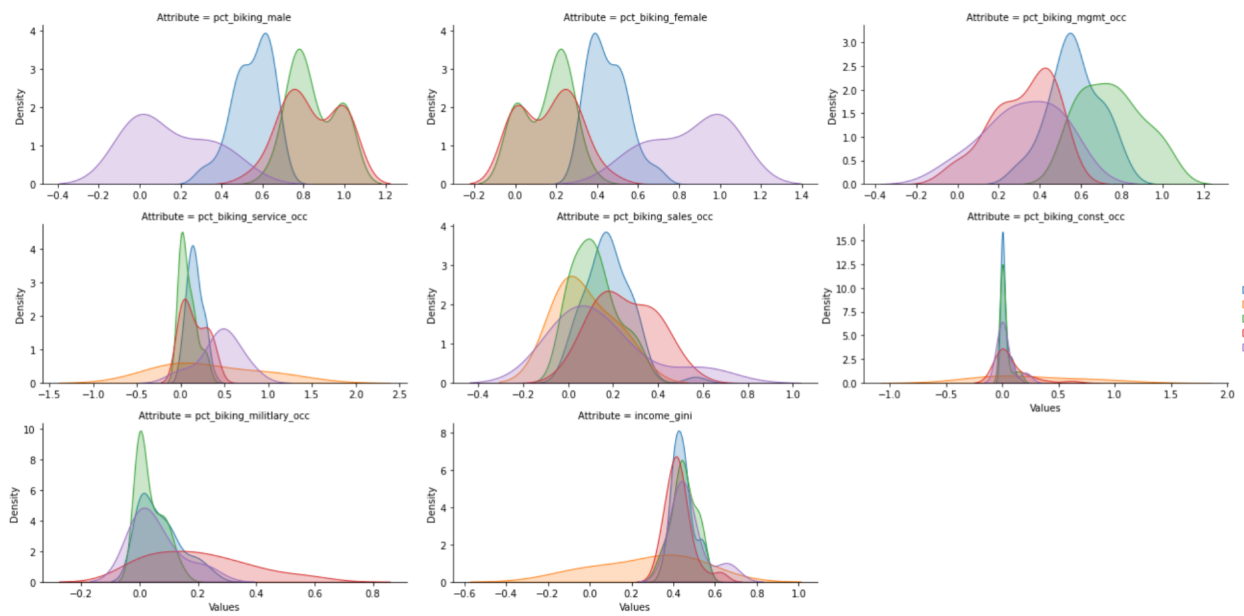
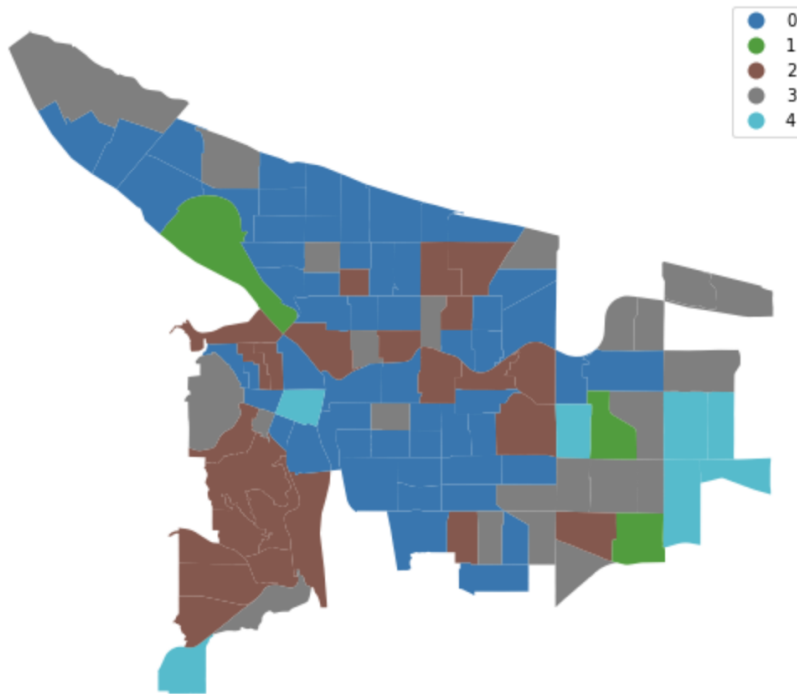
Q1.

Case 1: $K = 3$

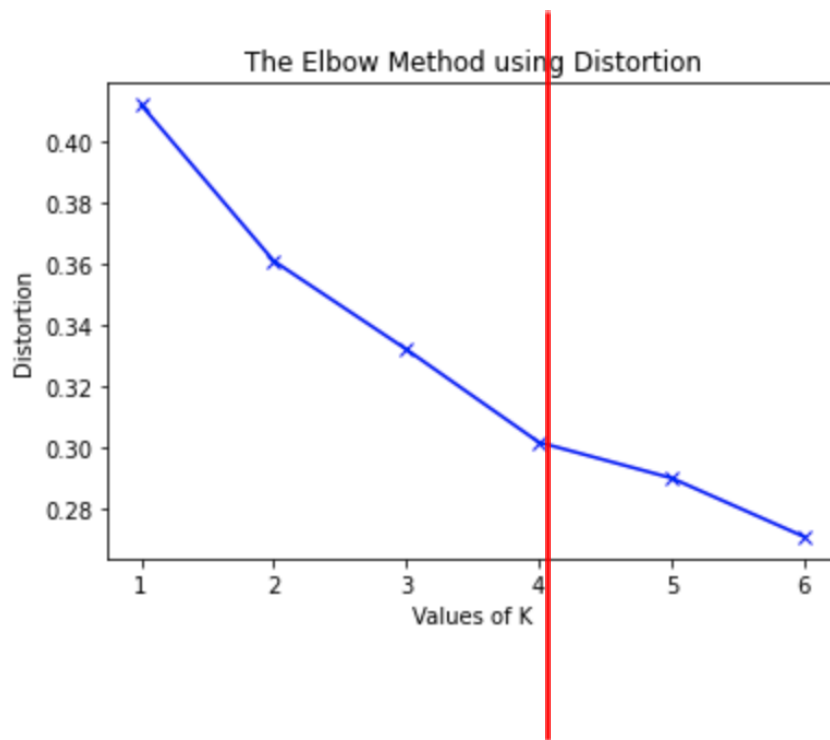


(case 2 continued on next page)

Case 2: K = 5

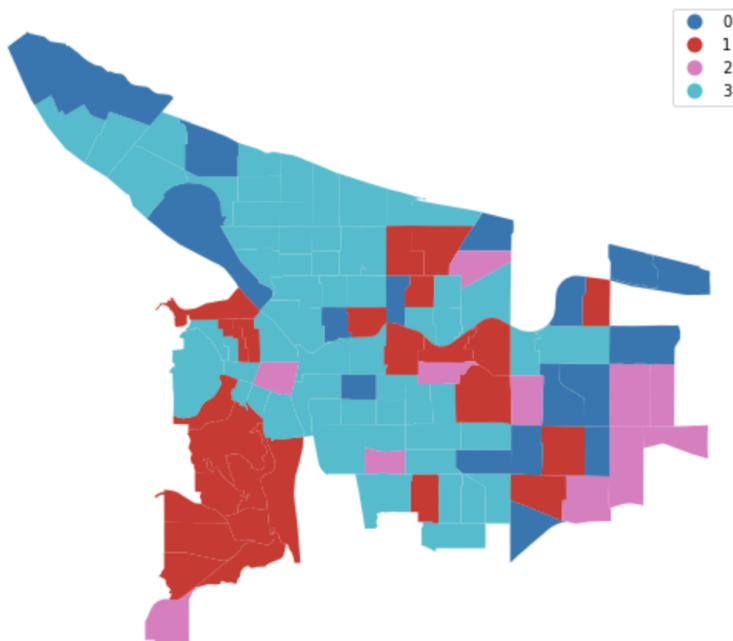


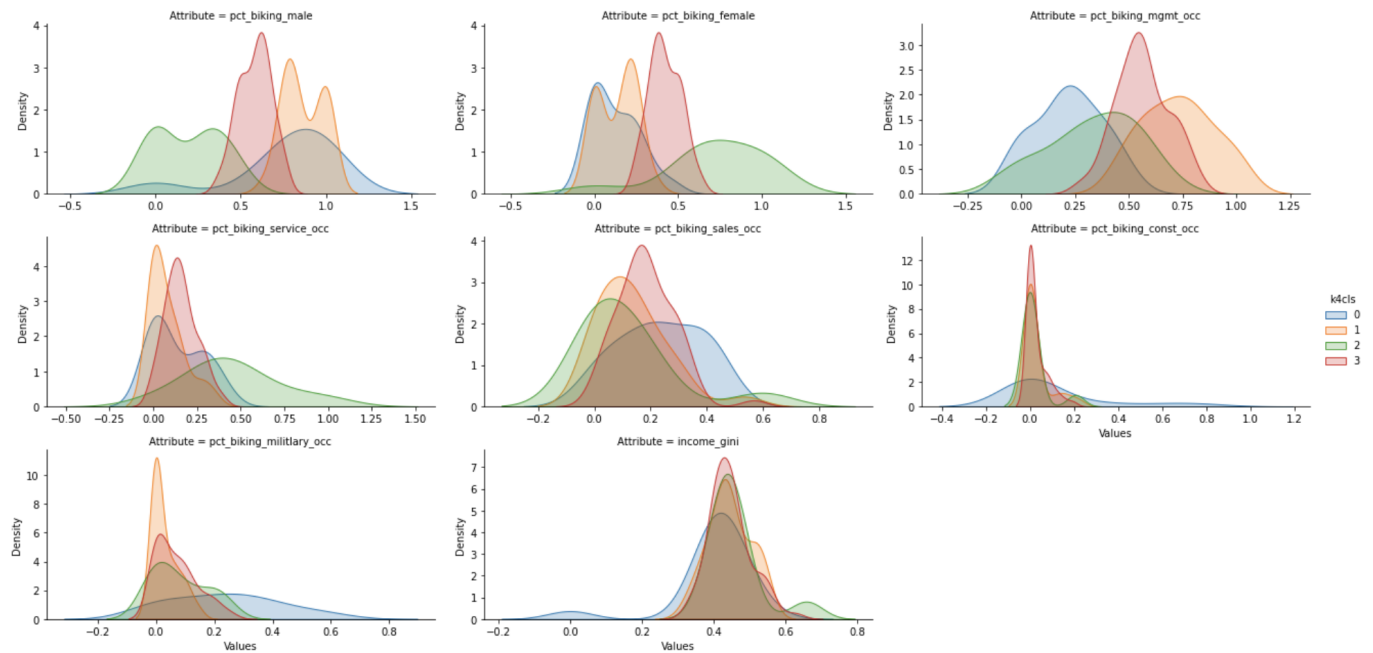
Q2. Plotting different clustering values of k , and by placing a vertical line through $k = 4$, I have decided to choose $k = 4$ as the number of clusters to perform my analysis on in question 3.



Q3.

I'm going to start by plotting a colored map for $k = 4$, and then I'm going to plot the distributions for the clustered variables. With this data analyzed, I will be able report the characteristics of this clustering configuration.





Unfortunately, with the downside of analyzing undisruptive, unchallenging data, the real differences in the 4 biking clusters come from their composition of men and their composition of educated driven occupations which also correspond geographically to areas that are not economically disadvantaged. These attributes define clusters 0 and 1 and correspond to wealthier, advantaged people who have the luxury and desire to commute on bicycle. Contrasted with cluster 3, this group, by analyzing its density charts, is economically disadvantaged but still occupies a large population of bicycle commuters. This suggests that there also exists a group, still existing in education-based occupation, that does not make a lot of money but still also chooses to biking as their mode of commute. The reason other types of occupations had little influence on the clusters is that geographically it is unlikely (compared to business/education occupations) that people live close enough to their jobs to be able to bicycle to them safely, even in a very bike friendly city like Portland, Oregon.