In our Explanatory Data Analysis (EDA) of a dataset focused on orders placed on a specific day, we approached it as an unsupervised dataset, employing a K-means clustering approach. The selected clustering features included the time taken to ship orders and associated delays, leading to the identification of three clusters. The determination of the optimal K value was facilitated by examining an elbow curve.

As part of our analysis, our attention turned to Ports experiencing late shipping days, a critical factor in understanding customer satisfaction. Remarkably, Port 04 emerged as the sole port exhibiting delays. To understand the underlying reasons, we scrutinized carrier charges for Port 04, evaluating the performance of 'V444_0' and 'V444_1'. In comparison with Ports 05 and 09, we assessed the Minimum Average Cost, and Port 04 demonstrated the lowest minimum average cost, potentially contributing to its delayed shipping.

To validate this inference, we further investigated the rate per kilogram for each order. Consistently, Port 04 exhibited the lowest rate per kilogram compared to Ports 05 and 09. This observation strengthens our conclusion that the combination of lower overall cost and a reduced rate per kilogram for Port 04 may contribute to delayed shipments, making it a less reliable option for customers.

While our findings offer valuable insights into the factors influencing late deliveries, a more comprehensive analysis could be further enhanced with additional data, such as the type of transportation utilized. Our future plans include a more in-depth exploration of cluster properties for each identified cluster, aiming to refine our understanding and potentially uncover opportunities for improving transportation efficiency.