**Summary**

**Data quality**

1. The Data was quite clean (without any missing values), only that it has a lot of categorical features which had to be handled before modeling.

2. According to the problem statement which intended to market top selling products, it was essential to have previous purchase history of the customers. Past behavior could have helped us in clustering the customers based on their liking or need which would have been more appropriate and accurate in this case.

**Solution**

1. In this problem, Geographical Clustering was chosen based on the available features. Features like country, street\_number, latitude and longitude have been used to categorize customers.

2. Key Points from Analysis:

* We see equal distribution of groups within title and department features.
* The dataset predominantly has Asian customers. China and Indonesia are the top Asian countries.
* Most of the customers are either Junior Executive or Analyst Programmer.
* In some of the companies, universities and professions we find either males or females but not both.
* Lifestyle, healthcare and visual industry are the most common skills among the customers.
* Continents were categorized and the corresponding latitude and longitude ranges were identified.

3. Conclusion:

**The business can focus on marketing its products based on geographic features. It can focus on the Asian market, particularly China and Indonesia, this would increase its target audience hence, increasing the chances of sales.**

**Next Steps**

* Try demographic segmentation using features like gender, skill, company, university.
* We can also use Silhouette score to evaluate the quality of clusters.
* Experiment with 2-3 other k values. Perhaps, trying with k=5 will yield better separable clusters.
* Check for outliers. Distance based algorithms are sensitive to outliers. The performance can be improved by handling the outliers if any