Assumptions / Methodology

* Suppliers in scope are in the file “supplier\_data.xlsx”.
* Any suppliers with latitude < 14 or > 15 are excluded.
* Similarly with longitude < 120.8 or > 121.2.
* Any suppliers with missing coordinates have those coordinates simulated by resampling from the existing supplier coordinates and adding Normal random noise with variance 4x10-5.
* Along with the supplier data we have the likelihood of a supplier delivering to non-shell sites (High Medium or Low), the number of times a supplier is expected to deliver to each site and the days of the week the supplier doesn’t deliver on.
* Where this information is missing we resample from the suppliers where it is present.
* If it is not present in any suppliers we do the following,
  + “the likelihood of a supplier delivering to non-shell sites” – sample one from ‘High’, ‘Medium’ or ‘Low’ with equal probability.
  + “the number of times a supplier is expected to deliver to each site” – sample from 1, 4, 8 or 25 times per month with equal probability.
  + “days of the week the supplier doesn’t deliver on” – default to delivering on all days.
* The internal sales data is then aggregated into deliveries with given frequency per month and random starting point ensuring the days not delivering on is respected.
* The dimension and weights of product categories are given in “product\_category\_dimensions.csv”, we assume all products in a category are of the same size and weight.
* Shell sales data used is a combination of internal records, “consolidated\_sales\_data\_v2.csv”, and those manually collected from sites, “additional\_sales.csv”.
* The Shell sites to be considered are in “Shell - Centro Asia SIte Name Mapping.csv”.
* The Centro Asia sales data used is in the sheet “centro\_sept.csv”.
* We assume any site which does not have “Shell“ in the name is not a Shell site.
* Using the Centro Asia data we pair up Shell and non Shell sites.
* We first calculate the sales value per category and then the distance between sites based on these sales.
* The probability of two sites being paired is proportional to 1 / distance.
* The non Shell sites considered are in the file “Suppliers Addresses\_CA Non-Shell Retailers.xlsx” and sheet “Centro Asia Non-Shell Retailers”.
* We remove sites based on coordinates in the same way as the suppliers above.
* We also simulate missing coordinates as above.
* We assume that non-Shell sites have the same order history as their paired Shell site.
* We then randomly remove orders for Non-Shell sites based on ““the likelihood of a supplier delivering to non-shell sites”, this is done randomly for each suppler / site pair separately.
* The probability that a non-Shell site does not order form a supplier is based on the values assigned,
  + “High” – 90% chance of ordering.
  + “Medium” – 50%.
  + “Low” – 10%.
* Tis probability is then multiplied by the proportion of Shell sites it delivers to, e.g. if a supplier has ‘High ‘ likelihood and delivers to 40% of Shell sites there is a 36% chance it will deliver to any non-Shell supplier.