

## # Prediction of Moz customer's monthly payment

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
import statsmodels.api as sm
from sklearn.linear_model import Ridge
```

```
In [2]: # load data
df_train = pd.read_csv('s3://mozbi-sr-data-science/data/moz_customer_data_train.csv', sep='|')
df_test = pd.read_csv('s3://mozbi-sr-data-science/data/moz_customer_data_test.csv', sep='|')
```

## # 1 - EDA

Based on the training dataset of 48,000 customers, United States, Canada and United Kingdom consist of 78% of Moz's total customers from 8 countries across North America, Europe, and Asia. On average, the customers from these top 3 countries also pay most on monthly basis (Figure 1: Box plot of prior monthly paid by country) The customers from marketing industry are prominent with 37.7% followed by healthcare industry 24.9%, software industry 12.2% and the remaining allocated to unknown.

On average, customers spent about 33 minutes last month, ranging from 10 minute (25 percentile) to 38 minutes (75 percentile). Interestingly, one software company from United Kingdom spent 837 minutes last month, which is 25 times longer than the average. While the time customers spent on Moz app showed the outliers, the monthly payment is consistent in the distribution without outliers, ranging from \$64 (25 percentile), \$88 (median), and \$111 (75 percentile).