A scoring model to identify high-potential new markets for a FinTech company

Capstone Project Summary for CEU MS in Business Analytics

Sára Vargha

June 2022

Table of contents

 Project Description	1
 Project Summary 3.1 Data exploration and variable selection 3.2 Scoring model 	1
3.1 Data exploration and variable selection3.2 Scoring model	2
3.2 Scoring model	2
	2
3.3 Country ranking and selection	2
4. Summary	3
4.1 Further research and limitations	3
4.2 Conclusion	3

1. Client introduction

The client is a Hungarian startup helping independent investors and traders to find an online broker fitting to their needs. They compare online brokers - be it a local one or a large international platform - across more than 300 dimensions. The company was founded in 2016 and successfully built up an operation of 35 people without external financing. However, to further scale the company, they are planning to expand the business outside their primary markets (e.g., the United Kingdom and the United States). The company generates its revenues in a commission fee model, meaning that they get a certain % either after the user opens an account at an online brokerage or after they click through from the client's website to the online broker's website.

2. Project Description

The aim of the project was to identify new target markets, e.g., countries, where the client could gain significant additional revenue by starting a paid online marketing campaign. Although the company already had detailed data on country-level user behavior and traffic from Google Analytics, they lacked information on macroeconomic indicators (such as the mean and median wealth per adult, GINI-index and GDP of countries, ratio of English-speaking population, Digital Adoption Index, etc.) that might influence the business potential of a new

market. The scope of the project was to formulate business recommendations on countries to target and further research.

3. Project Summary

To identify high potential new markets, a scoring model was built to rate the potential of countries. To achieve this, business specific data from Google Analytics and the company's reports was integrated with various macroeconomic data mainly extracted from the Credit Suisse Global Wealth Databook 2021 and the World Bank.

3.1 Data exploration and variable selection

Building on the resulting analytical layers, a thorough explanatory data analysis was conducted to get deeper insights into the data and to discover components that are correlated with the main target variables, the company's revenue, and end-to-end conversion rate (revenue generated per session).

In case of macro variables, EDA was complemented by a set of simple regressions and several multiple regression models to understand the pattern of association between the target variables and the different independent macro variable combinations. In case of business-related variables, the selection process was based on the findings of the EDA as well as the client's inputs on the most important business metrics.

3.2 Scoring model

Then to scoring model was built up from two components: a macro and a business component. The macro score was built upon the variables that were found significant in the simple and explanatory regression models. These variables included the ratio of English-speaking population, the total wealth, World Bank's Digital Adoption Index and income group classification of counties as well as the ratio of financial wealth in the countries. While these variables were proven to be positively associated with the company's revenue and the number of sessions spent on their website, they had no significant effect on revenue per session.

The business score was created from the revenue per session, the ratio of desktop and returning users, the number of sessions spent on the best converting sites of the client's website and the number of responses to the client's revenue driver customer survey. In total, the maximum points to achieved were 5-5 points in both score categories. The final score was calculated as the sum of the macro score and the business score multiplied by two, to give a bigger weight to business aspects in line with the project scope. The maximum final score was thus 15 points.

3.3 Country ranking and selection

Finally, countries were rated with the scoring model and ranked accordingly in each score categories. While in terms of macro score highly developed and wealthy countries ranked the highest (such as the Scandinavians, Switzerland, Canada or Israel), Central and Eastern European countries (such as Slovenia, Bulgaria, Hungary, Turkey, Greece) performed the best according to the business scores. The countries that scored among the top ten in either of the

two subscores as well as in the final score were selected as "focus countries". A detailed data visualization dashboard and a brief country profile for the selected countries were provided to the client to serve as basis for further growth strategy building and research.

4. Summary

4.1 Further research and limitations

Further research would be needed to uncover which variables explain the variance in the endto-end conversion rates. Nevertheless, this research identified a range of macro and Google Analytics variables that are positively associated with both revenue and sessions. Building upon the findings we can conclude the variance in the number of sessions explains most of the variance in revenue, thus the company should be able to increase revenues in selected countries by boosting traffic with paid marketing campaigns. Furthermore, to get a better understanding on how big the revenue potential of selected countries would be, various revenue prediction techniques could be applied to quantify growth potential.

Scoring model could be enhanced by including (1) a wider time range and (2) more variables with higher variance in the analysis (e.g., using deciles instead of quantiles to avoid equal scores). Depending on the business incentives, experimenting with more complex scoring formulas and the application of different weights might lead to better results.

Finally, the study had several limitations meanly related to the availability and completeness of data, as many data points were available only for a limited time period or not at all for less-developed countries where overall data quality was very low. Other data points, such as the ratio of English-speaking population in countries were based on estimations which decreases reliability. As a result of countries with poor data quality, lot of observations with many missing values had to be dropped that significantly decreased the sample size. Therefore, the external validity of results needs to be handled with caution.

4.2 Conclusion

To conclude, this analysis discovered a set of macroeconomic factors that have statistically significant positive effect on both the number of sessions spent on the website and the generated revenue. However, this study failed to identify variables that might explain the revenue generated by session (also known as the end-to-end conversion rate). Nevertheless, the results might serve as good basis for further research on end-to-end conversion rate.

Furthermore, based on the combination of the outcomes of data exploration and simple regression models and the client's business inputs, a scoring model was set up to rank potential target countries based on macroeconomic and business metrics. The model in its initial form provided a list of lucrative target markets that might serve as the basis of the client's growth strategy and a tool that can be further improved depending on the business objectives of the company.