EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS

**Conrad Lee, Derek Perkinson, Felix Krueger,
Kei Ichikawa, Martin Louzada, Vanessa Lopez Garcia**

INTRODUCTION

The analysis of tourists' credit card transactions can provide a deeper understanding on travel spending, allowing to identify trends and highlights in consumption patterns from visitors. This report focuses on the analysis of credit card transactions performed by tourists in the city of Madrid, Spain, between March 2nd and 3rd, 2012. The data available includes information about the amount, country of origin, spending category and timing of a transaction.

Within the 48 hours of data available, we found 110 different nationalities to be present, 10.793 transactions and €1.173.584 in purchases. As will be elaborated later in the report, the countries with the highest spending were the United States, Great Britain, China, France and Japan. In addition, most expenses during the 48 hours were concentrated among three main categories: "Fashion & Shoes", "Bars & Restaurants" and "Accommodation".

The exploratory analysis of the dataset revealed interesting patterns of consumer behavior of different nationalities, which present distinct characteristics in terms of spending category, timing of purchase and average ticket value, which will be further explained below.

Lastly, some potential data monetization strategies are discussed in the final part of the report, which focuses on how the discovered insights and patterns can be used as input for search engine keywords pricing strategies. The genesis for this project was to determine how Citibank can better monetize its current transactional data and present a set of recommendations at Citibank's annual convention. As a result of the data exploration, this report recommends a partnership and pilot program with Google using Citibank's transactional data.

KEY FINDINGS

- **Yankee Invasion:** US tourists are the ones spending more money in Madrid. In the 48 hours covered by the dataset, they left more than €180.000 at the Spanish capital. *¡Muchas gracias!*
- **Thailand, South Africa and Indonesia – highest average tickets:** visitors from those countries seem to be keen to making very expensive purchases. On average, each of their transactions is 4 times greater than the overall average.
- **Cover Page – Fashion in the spotlight with highest share of tourists' wallet: Tourists in** Madrid seem to love its fashion retail stores: in two days, they spent almost half a million euros in clothing and shoes.
- **Happy Hours – Bars & Restaurants cash in after sunset:** as the day is over in the Spanish capital, tourists make sure to enjoy its delicious food and *cañas:* 200 thousand euros were spent in two days, mostly after 6pm and well into the night.

SPENDING BY COUNTRY OF ORIGIN

The number one country in terms of transactions made in the given period is the United States, with 1.686 purchases, 15.6% of the total, followed by France (11.4%) and Great Britain (7.8%). In fact, out of the 110 different nationalities in the dataset, the top 25 account for over 83% of all transactions and 81% of all the money spent. Considering the

total money spent by tourists, the United States are again a clear winner, with almost 185 thousand Euros in transactions, or nearly 16% of the total. This is followed by Great Britain (7.6%), China (7.5%) and France (6.9%). Together, tourists from these four countries infused more than 440 thousand Euros to the Spanish capital's economy in just 48 hours. Furthermore, the removal from the analysis of some transactions which may be considered outliers does not alter such conclusions.

Interestingly, when we look at those transactions from an average ticket perspective, we see completely different countries topping the rankings. Considering only countries with at least 20 transactions made in the given period, we found out that tourists from Thailand spent on average €408 on each transaction. South African tourists, with an average of €395, Indians with around €358 and Malaysians with €332 per transaction, complete the top 5 in this category.

This indicates that, although not present in large numbers in Madrid, tourists from these countries represent an important source of tourism revenue, accounting for the most valuable transactions made by tourists in the city.
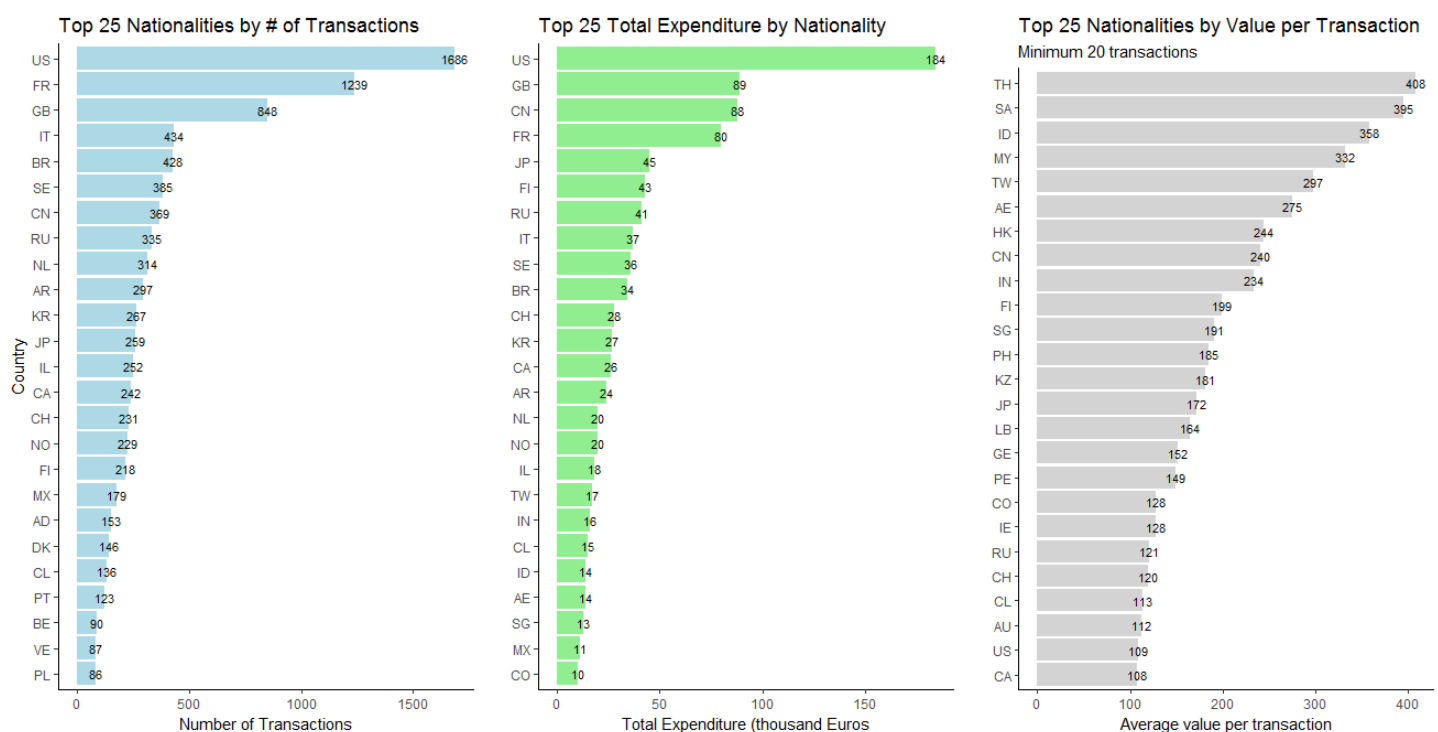

Figure 1 Top countries by number of transactions, total money spent and average ticket

SPENDING BY CATEGORY

Accounting for almost half a million Euros in just two days, "Fashion and Shoes" are the top category in terms of tourists spending in Madrid: its 5.372 transactions (49% of all transactions) represent 42% of the total credit card spending by tourists. The other two most important categories in terms of volume of transactions are "Bars &

Restaurants" as well as "Accommodation", which contribute, 24% and 10% of the number of transactions and 17% and 24% to the total spending respectively.

It is clear that "Fashion & Shoes" and "Bars & Restaurants" have a huge impact on tourism revenue in Madrid, largely because of the high number of purchases that occur in those categories, as they do not present significantly large value per transaction (€92 and €76). On the other hand, "Accommodation" presents an average ticket of €256, virtually tied for highest with "Culture & Leisure" and "Transportation".

Those last two categories, however, had only 92 and 216 transactions each, less than 3% of all transactions, which indicates a potential opportunity to increase revenue by increasing sales and transactions by tourists in those areas. Another category with a potential to boost revenue is "Books and Music" with a relatively high ticket but very few transactions made.
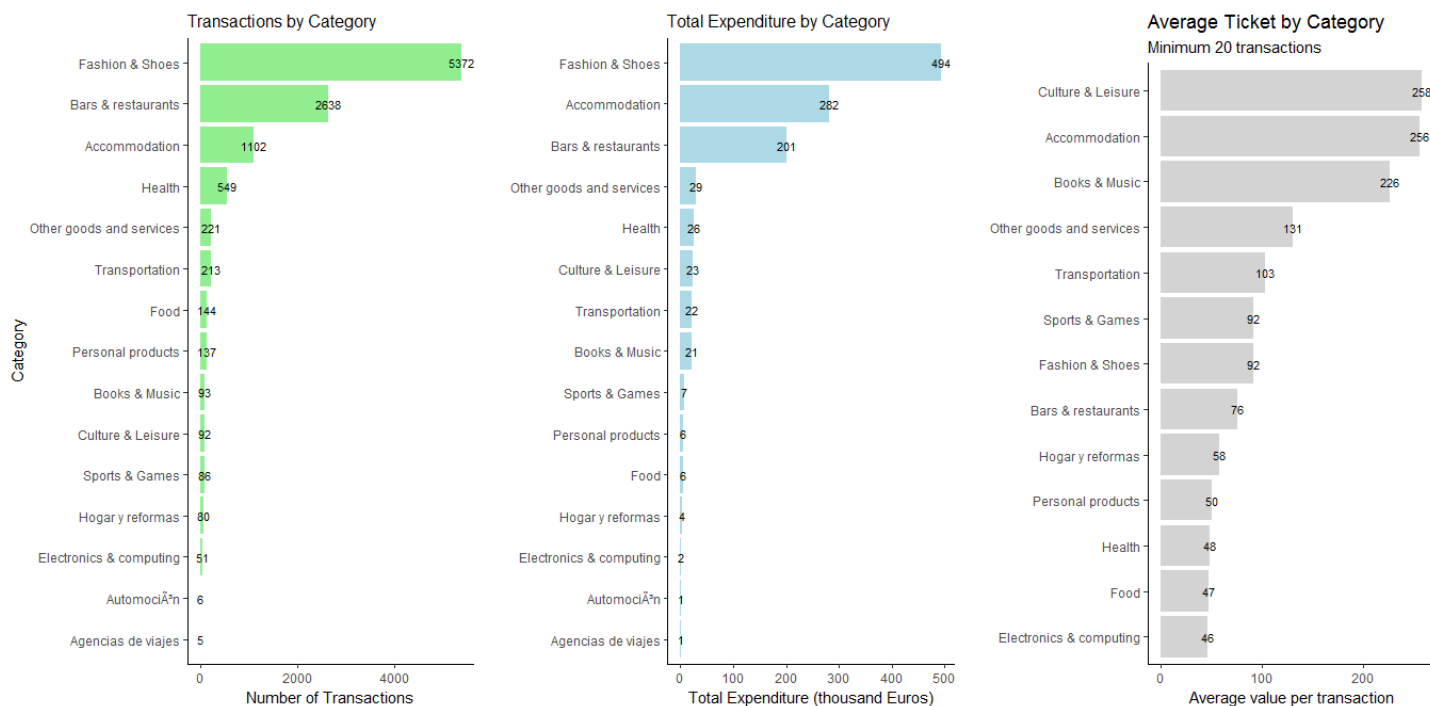

Figure 2 Categories by number of transactions, total money spent and average ticket

The analysis of spending by country within each category reveals interesting patterns. Let's take "Fashion and Shoes", "Bars and Restaurants" and "Accommodation" as our categories of interest, since it was previously stated that these are the main ones in terms of revenue and volume of transactions. In those categories, there are stark differences in average ticket prices among tourists from different nations, revealing some countries' tendencies to spend more for some items or to opt for cheaper options.

In the "Fashion and Shoes" sales, tourists from Thailand present the highest average spending per transaction at €475, or 516% of the overall average of the category. Also present in this category are visitors from India (€409) and Taiwan (€381). On a second tier, visitors from UAE, China, Singapore, Japan and Kazakhstan present average spending ranging from €270 to €178. These numbers suggest that targeting high value fashion items to those visitors is very likely to translate into sales with higher profit margins.

Regarding spending in Accommodation, tourists from Finland have by far the largest average ticket, €720, which is almost three times the average for the category. This points to the fact that Finish tourists are very likely to select luxury accommodations and therefore should be targeted with ads for that sector. Chinese visitors also present a high average Accommodation ticket and could also be subjects of interest to luxury hotels.

Finally, when it comes to "Bars and Restaurants", the average value spent by Chinese tourists is the largest one, with each transaction at around €179, almost 2.5 times the overall average ticket. Visitors from Japan, Italy, Poland, Russia and Switzerland belong to a second highest tier of spending, averaging around €120.

The next section analyzes how the consumption within these categories varies throughout the day, and assesses if different nationalities have different behaviors when it comes to the timing of their spending.
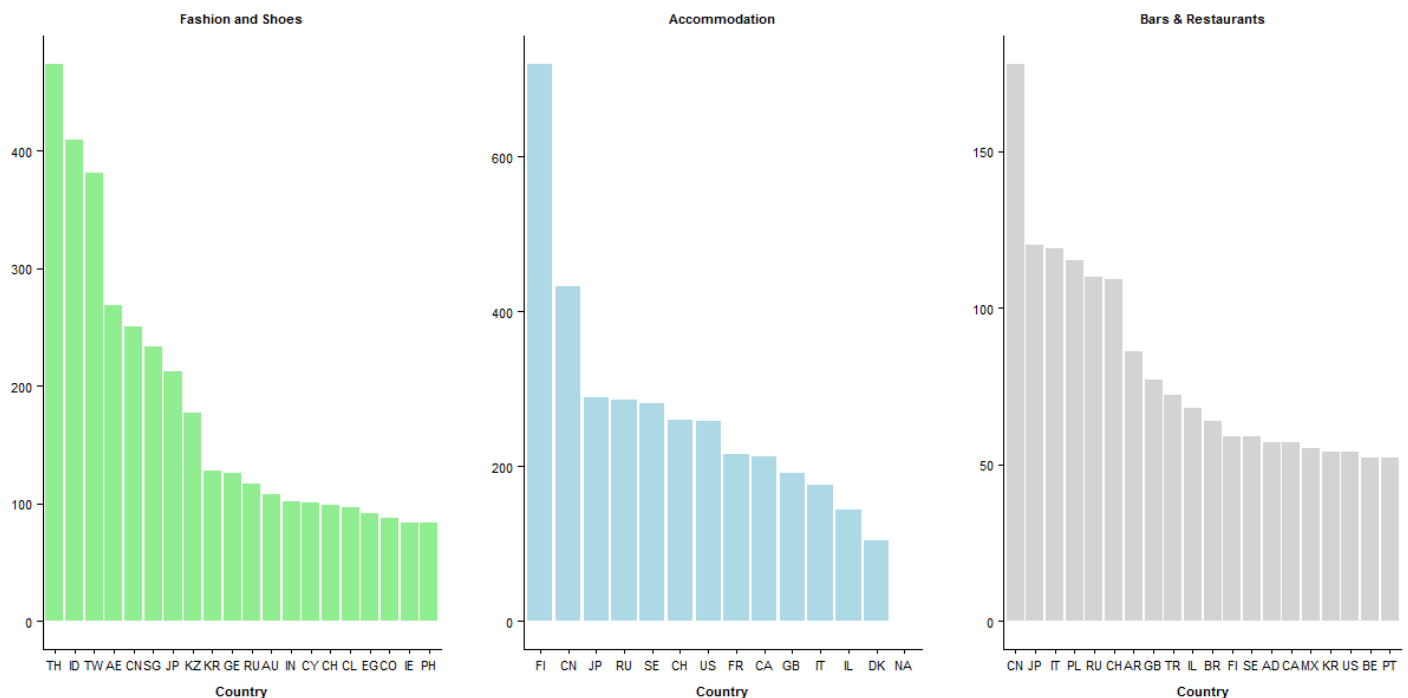

Figure 3 Average ticket by country on categories of interest

SPENDING BY DAY OF THE WEEK AND TIME OF THE DAY

Analyzing the evolution of tourist spending throughout the day shows a growing trend in the volume of transactions until it peaks between 5pm and 6pm, at which point they slowly start to decrease until there is a second peak around midnight. The same pattern occurs Thursday and Friday, although volume of transactions is larger on the later day.

It is clear that the first peak, at the end of afternoon, coincides with the peak of purchases in "Fashion and Shoes", however, this category presents a consistent number of transactions throughout all business hours.

From this time onwards, "Bars and Restaurant" start to concentrate tourism spending, with the period between 9pm and midnight being the most profitable for the sector.

On the other hand, Accommodation transactions peak earlier in the day, from 7am to 10am, which is likely the time when tourists check out and pay their hotel bills.
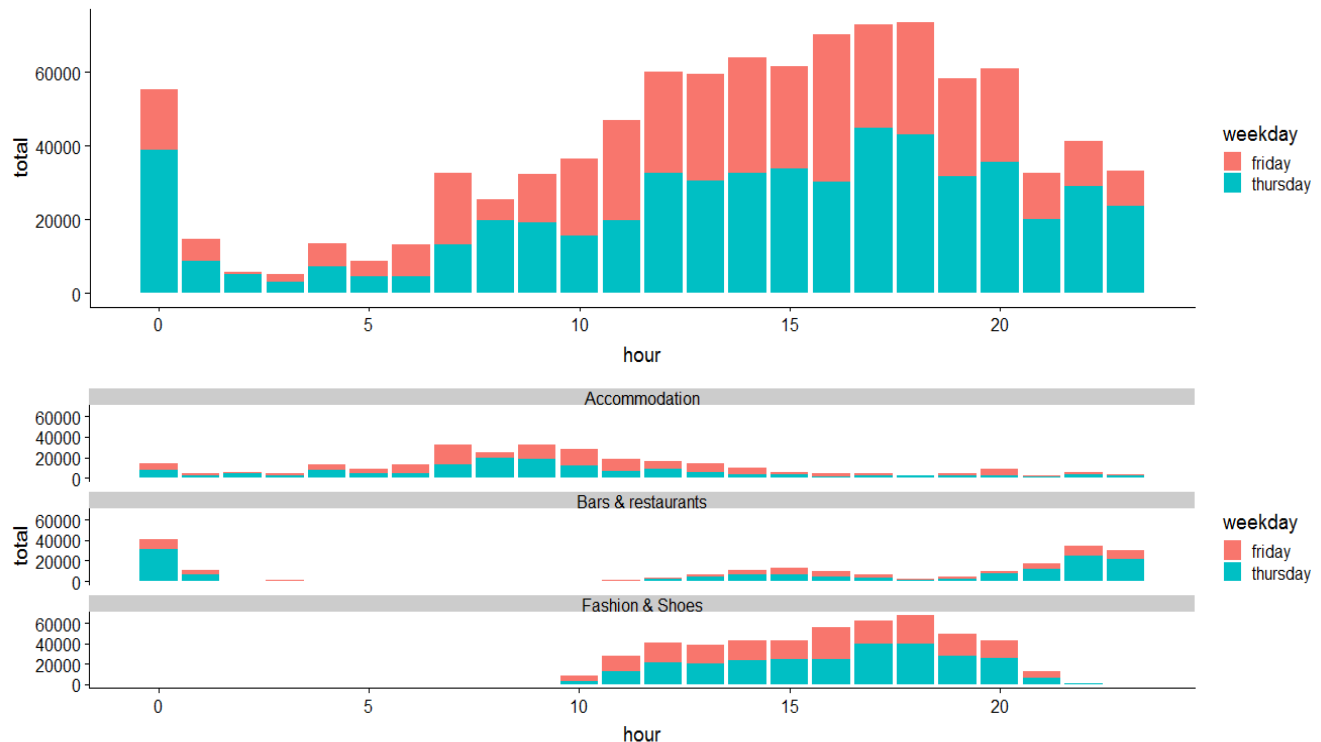
Figure 4 Tourist spending throughout the day and on main categories

DEEP DIVE ON FASHION, BARS & RESTAURANTS AND ACCOMODATION SPENDING

Tourists from different countries have different consumer behavior not only in terms of what they buy and how much they are willing to spend, but also in regards of the timing of their purchases.

For instances, Fashion transactions occur earlier in the day for Emiratis, later for Taiwanese and Singaporeans, and even later for Indians. On the other hand, Chinese tourists shopped consistently throughout the day, and Japanese visitors concentrated their purchases on the first half of business hours.

Figure 5 Spending in fashion for 6 nationalities throughout the day

As for tourism transactions that occurred in Bars and Restaurants, it is possible to visualize that Britons and Russians are a constant presence in those places while Italians arrive later and stay longer, and Chinese and Japanese tourists prefer early, shorter night outs.
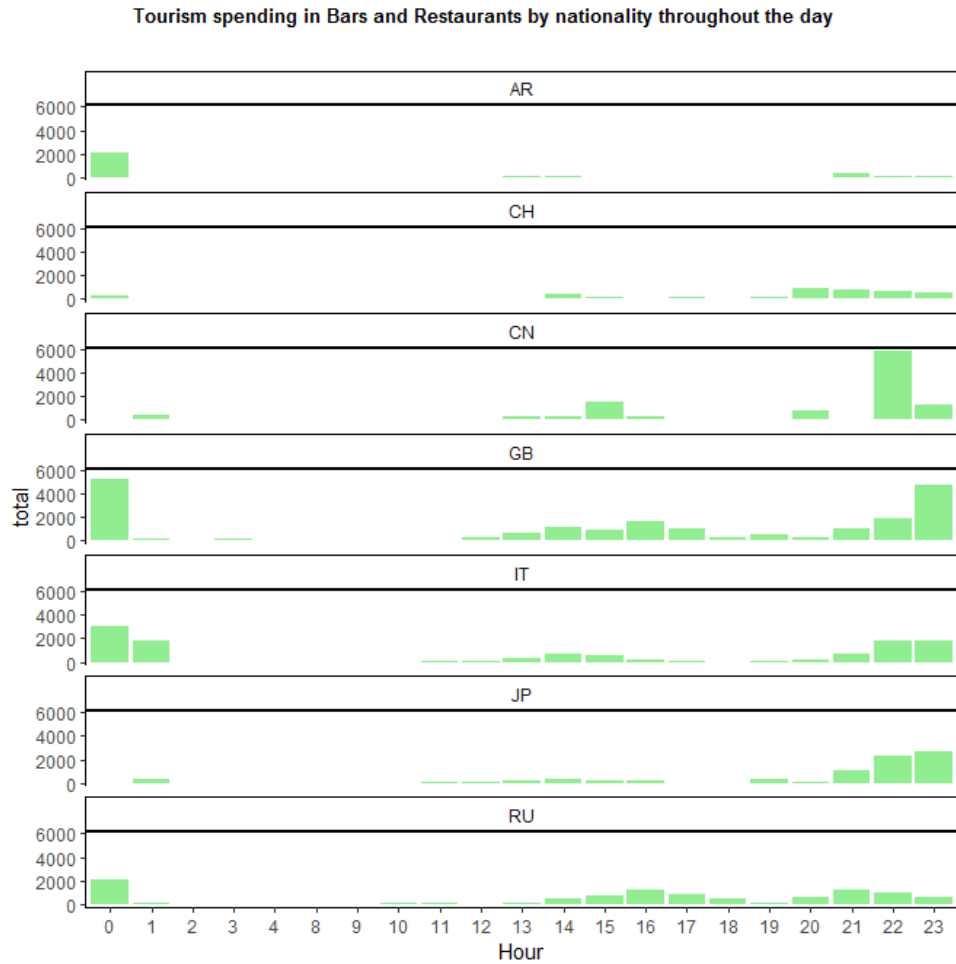
Figure 6 Spending in Bars & Restaurants for 5 nationalities throughout the day

DATA MONETIZATION RECOMMENDATIONS

The structure of the partnership between Citibank and Google will have Citibank selling real-time transactional data to Google allowing it to charge a premium on keywords for products typically associated with impulse purchases. The size of Citibank's credit card customer base and the fact that the majority of Citibank's credit card customers use their card as the primary card for purchases will allow Google to augment a Citibank credit card user's current customer's search history (indicating potential purchase intent) with Citibank transactional data (indicating the actual purchase). The resulting combination will allow Google to subsequently target ads more effectively. Google could also augment both types of data with other data that it possesses (location-based information for example) to further enhance the value associated with the keyword advertisements.

The examples used while pitching to Google included targeting according to nationality. One example pitched was that consumers from Thailand, South Africa and Indonesia all purchased products at a higher price point compared

to tourists from other countries. The Citibank research focused primarily on Fashion & Shoes, Bar and Restaurants, and Accommodations to determine broad trends. We focused on these categories due to their importance to the Madrid economy and the relatively short time frame from intent to actual purchase. For each of these categories, tourists from each of these countries spent significantly more on average compared with other tourists and would allow justifying a significant increase in the keyword pricing.

Another example would be related to Americans and travel. Americans disproportionately often traveled first class (for this exercise we estimated first class based on the travel price, but we would actually know if the card holder traveled first class based upon the ticket code sent by the airline). Americans also disproportionately often booked travel early in the morning. Both pieces of information would be good to know when appropriately pricing keyword searches. A similar approach could be taken in relation to Finish travelers when deciding on accommodation, giving that they are more prone to select exclusive hotels.

One additional point to illustrate the potential use of the given data to guide keywords prices is the different behavior when it comes to going to bars and restaurants. For example, British and Russian tourists seem to decide on which places to go earlier in the day, so prices could surge for these nationalities around the time they search for a recommendation. On the other hand, Japanese and Chinese customers chose to go out later so prices would also vary accordingly.

A final illustration of the pitch that we (Citibank) gave to Google was based on Fashion & Shoes. Despite other cities being recognized as shopping destinations for fashion (Milan, Paris, and London), Madrid has a disproportionate percentage of purchases that are connected to fashion. The importance for Google is that both, the amount per purchase and the total number of purchases, are high compared to other categories. This helps us (Citibank) by suggesting that tourist from all nations disproportionately buy Fashion compared to other categories, thus allowing Google to target tourists indiscriminately (thus resulting in a higher overall revenue gain). This is an illustration of one of the biggest takeaways from the given data – the ability to target specific demographic groups by category for certain information while other categories allow Google to indiscriminately price keywords at the highest possible price based upon the insights provided by the data.

COUNTRY CODES

AD - Andorra, AE - United Arab Emirates, AU - Australia, BE - Belgium, BR - Brazil, CA - Canada, CH - Switzerland, CL - Chile, CN - China, CO - Colombia, CY - Cyprus, DK - Denmark, EG - Egypt, FI - Finland, FR - France, GB - Great Britain, GE - Georgia, HK - Hong Kong, ID - Indonesia, IE - Ireland, IL - Israel, IN - India, IT - Italy, JP - Japan, KR - South Korea, KZ - Kazakhstan, LB - Lebanon, MX - Mexico, MY - Malaysia, NL - Netherlands, NO - Norway, PE - Peru, PH - Philippines, PL - Poland, PT - Portugal, RU - Russia, SA - South Africa, SE - Sweden, SG - Singapore, TH - Thailand, TW - Taiwan, US - United States, VE - Venezuela