**Project Title: Digital Marketing Agency | Analysis of Opportunities to Invest in Spain – Bars & Restaurants**

**Team Members**

* Geraldine Carrillo Monroy

**Project Description**

In Project I, I analyzed BBVA’s databases, which included all BBVA card transactions and any other card transaction in BBVA POS terminals in “brick and mortar” stores in Madrid in October and November 2015.

PayStats offers anonymized and aggregated statistical data from millions of transactions performed with BBVA cards and any other cards in BBVA POS terminals, creating a virtual map comprised of consumers' habits, demographics and origins.

5 main dimensions to improve customer behaviour analysis: (i) Territorial unit, (ii) Time, (iii) Merchant Category, (iv) Gender and (v) Age.

**Main metrics**

Segmented by the main dimensions for analysis (territorial unit, time, merchant category, gender and age), the API offers different services to recover the following metrics:

* Purchase amount:
  + Maximum ticket
  + Minimum ticket
  + Average ticket
  + Standard deviation
  + Mean value
* Number of transactions
* Number of different cards (individuals)
* Peak and trough days and hours

**Data Protection**

Anonymization algorithm

All data is anonymous and compliant with Data Protection regulations (LOPD). To comply with the LOPD, any consultation of the API requires that the return value include a minimum of purchases, different cards or merchants POSs at the level/category consulted, otherwise the information will return information aggregated at a higher level. This way BBVA completely guarantees the anonymity of all transactions.

Preponderance algorithm

BBVA ensures that no data can be retrieved from an API query if the results could be attributable to a unique merchant.

**What kind of questions can PayStats answer?**

1. Is my business located on the best location?
2. What do my customers buy around my business?
3. Where do my clients come from?
4. How much do my clients spend on products in my sector? And in other sectors?
5. Is my advertising campaign optimized?

**Who can use PayStats?**

* Market analysis companies
* New businesses searching for the best location
* Franchises who want to optimize their network
* Companies looking to better understand their customers' consumption behavior
* Retail space management companies
* Consultants
* Business schools
* Research institutions
* ... Anybody!

**How do I consume PayStats?**

API call getting as a response to the REST call a JSON with the requested information.

API download requesting a dataset with all the information of a given area in a .csv file. Ideal for GIS and high-level analysis.

There are four main databases (monthly information – October 2015):

1. **Basic Stats**

This file provides statistics (average transaction amount, the number of transactions, number of merchants and the number of cards, etc.) for a particular area and business category.

* 1. Output layout:
     1. Zone: territorial unit id (in the territorial unit tile will be composed of two fields: latitude and longitude).
     2. Date: the date the data refers to (YYYY-MM format for month and YYYY-MM-DD for day).
     3. Card: Card type.
     4. Channel: Channel type.
     5. Category: commercial category label.
     6. Merchants: number of merchants with registered activity.
     7. Cards: number of different cards.
     8. Txs: number of physical purchase transactions.
     9. Avg. amount: average transaction amount.
     10. Max. amount: maximum transaction amount.
     11. Min. amount: minimum transaction amount.
     12. Std. amount: standard desviation transaction amount.
     13. Peak day: day of week with maximum number of transactions. Values: [sunday, monday, tuesday, wednesday, thursday, friday, saturday].
     14. Valley day: day of week with minimum number of transactions. Values: [sunday, monday, tuesday, wednesday, thursday, friday, saturday].
     15. Peak hour: hour of day with maximum number of transactions. Values: [00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23].
     16. Valley hour: hour of day with minimum number of transactions. Values: [00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23].
  2. Possible combinations:

SOURCE CHANNEL

* + 1. all pos / opos
    2. bbva pos / opos
    3. national bbva\_pos / bbva\_opos
    4. foreign bbva\_pos / bbva\_opos

Note: opos and bbva\_opos statistics are only available for the subregion territorial unit.

1. **Category Distribution**

This file provides category distributions statistics for a particular dimension. The file returns the average transaction amount and the number of transactions for each period and category or subcategory group as the output.

* 1. Output layout:
     1. **Zone:** territorial unit id (in the territorial unit tile will be composed of two fields: latitude and longitude).
     2. **Date:** the date the data refers to (YYYY-MM format for month and YYYY-MM-DD for day).
     3. **Card:** Card type.
     4. **Channel:** Channel type.
     5. **Merchants:** number of merchants with registered activity.
     6. **Cards:** number of different cards.
     7. **Txs:** number of physical purchase transactions.
     8. **Avg. amount:** average transaction amount.
     9. **Category level:** commercial category level type. Values: [categories, subcategories].
     10. **Category:** commercial category label.
     11. **Merchants by category:** number of merchants with registered activity of category or subcategory.
     12. **Cards by category:** number of different cards of category or subcategory.
     13. **Txs by category:** number of physical purchase transactions of category or subcategory.
     14. **Avg. amount by category:** average transaction amount of category or subcategory.
  2. Possible combinations:

SOURCE CHANNEL

* + 1. all pos
    2. bbva pos
    3. national bbva\_pos
    4. foreign bbva\_pos

1. **Consumption Pattern**

This file provides typical consumption patterns by the hour of the day and the day of the week for a particular area. For each one of the time units, consumption statistics by the hour of the day and the day of the week are provided in the response and it is only available for month grouping.

* 1. Output layout:
     1. **Zone:** territorial unit id (in the territorial unit tile will be composed of two fields: latitude and longitude).
     2. **Date:** the date the data refers to (YYYY-MM format for month and YYYY-MM-DD for day).
     3. **Card:** Card type.
     4. **Channel:** Channel type.
     5. **Category:** type of category or subcategory.
     6. **Merchants:** number of merchants with registered activity.
     7. **Cards:** number of different cards.
     8. **Txs:** number of physical purchase transactions.
     9. **Avg. amount:** average transaction amount.
     10. **Day:** day of the week. Values: [sunday, monday, tuesday, wednesday, thursday, friday, saturday].
     11. **Merchants by day:** number of merchants with registered activity of the day of the week.
     12. **Cards by day:** number of different cards of the day of the week.
     13. **Txs by day:** number of physical purchase transactions of the day of the week.
     14. **Avg. amount by day:** average amount of the day of the week.
     15. **Max. amount by day:** maximum transaction amount of the day of the week.
     16. **Min. amount by day:** minimum transaction amount of the day of the week.
     17. **Std. amount by day:** transaction standard deviation amount of the day of the week.
     18. **Hour:** hour of the day of the week. Values: [00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23].
     19. **Merchants by hour:** number of merchants with registered activity of the hour of the day of the week.
     20. **Cards by hour:** number of different cards of the hour of the day of the week.
     21. **Txs by hour**: number of physical purchase transactions of the hour of the day of the week.
     22. **Avg. amount by hour:** average amount of the hour of the day of the week.
     23. **Max. amount by hour:** maximum transaction amount of the hour of the day of the week.
     24. **Min. amount by hour:** minimum transaction amount of the hour of the day of the week.
     25. **Std. amount hour:** transaction standard deviation amount of the hour of the day of the week.
  2. Possible combinations:

SOURCE CHANNEL

* + - 1. all pos / opos
      2. bbva pos / opos
      3. national bbva\_pos / bbva\_opos
      4. foreign bbva\_pos / bbva\_opos

Note: opos and bbva\_opos statistics are only available for the subregion territorial unit.

1. **Origin Distribution by Categories**

This file provides origin distribution statistics by categories for a particular area. The file returns the average transaction amount, the number of transactions, number of merchants and the number of cards for each period and origin as the output.

* 1. Output layout:
     1. Zone: territorial unit id (in the territorial unit tile will be composed of two fields: latitude and longitude).
     2. Date: the date the data refers to (YYYY-MM format for month and YYYY-MM-DD for day).
     3. Card: Card type.
     4. Channel: Channel type.
     5. Merchants: number of merchants with registered activity.
     6. Cards: number of different cards.
     7. Txs: number of physical purchase transactions.
     8. Avg. amount: average transaction amount.
     9. Origin type: type of origin. Values: [subregions, localities, zipcodes, countries].
     10. Origin: ID of the origin.
     11. Merchants by origin: number of merchants with registered activity in this origin.
     12. Cards by origin: number of different cards in this origin.
     13. Txs by origin: number of physical purchase transactions in this origin.
     14. Avg. amount by origin: average amount in this origin.
     15. Category: id for the category.
     16. Merchants by category: number of merchants with registered activity of the category in this origin.
     17. Cards by category: number of different cards of the category in this origin.
     18. Txs by category: number of physical purchase transactions of the category in this origin.
     19. Avg. amount by category: average amount of the category in this origin.
  2. Possible combinations:

SOURCE CHANNEL ORIGIN\_TYPE

foreign bbva\_pos / bbva\_opos countries

bbva pos / opos localities

bbva pos / opos subregions

bbva pos / opos zipcodes

Note: opos and bbva\_opos statistics are only available for the subregion territorial unit.

1. **Origin Distribution by Ages and Genders**

This file provides origin distribution statistics by ages and genders for a particular area. The file returns the average transaction amount, the number of transactions, number of merchants and the number of cards for each period and origin as the output.

1. Output layout:
   1. Zone: territorial unit id (in the territorial unit tile will be composed of two fields: latitude and longitude).
   2. Date: the date the data refers to (YYYY-MM format for month and YYYY-MM-DD for day).
   3. Card: Card type.
   4. Channel: Channel type.
   5. Category (optional): commercial category label (when missing the row refers to all the existing categories)
   6. Merchants: number of merchants with registered activity.
   7. Cards: number of different cards.
   8. Txs: number of physical purchase transactions.
   9. Avg. amount: average transaction amount.
   10. Origin type: type of origin. Values: [subregions, localities, zipcodes].
   11. Origin: id for the origin type.
   12. Merchants by origin: number of merchants with registered activity in this origin.
   13. Cards by origin: number of different cards in this origin.
   14. Txs by origin: number of physical purchase transactions in this origin.
   15. Avg. amount by origin: average amount in this origin.
   16. Age: age range label by origin. Values: [0 (<=24), 1 (25-34), 2 (35-44), 3 (45-54), 4 (55-64), 5 (=65), U (Unknown)].
   17. Merchants by age: number of merchants with registered activity for age range.
   18. Cards by age: number of different cards for age range.
   19. Txs by age: number of physical purchase transactions for age range.
   20. Avg. amount by age: average amount for age range.
   21. Gender: gender label by age range and origin. Values: [F (Female), M (Male), E (Enterprise), U (Unknown)].
   22. Merchants by gender: number of merchants with registered activity for age and gender range.
   23. Cards by gender: number of different cards for age and gender range.
   24. Txs by gender: number of physical purchase transactions for age and gender range.
   25. Avg. amount by gender: average amount for age and gender range.
2. Possible combinations:

SOURCE CHANNEL ORIGIN\_TYPE

1. bbva pos / opos localities
2. bbva pos / opos subregions
3. bbva pos / opos zipcodes

Note: opos statistics are only available for the subregion territorial unit.

Destination Distribution

This file provides purchase destination distribution statistics for the clients who live in a particular area. The file returns the average transaction amount, the number of transactions, number of merchants and the number of cards for each period and destination as the output.

Not available for tiles.

Output layout:

Zone: territorial unit id (only subregions, localities and zipcodes).

Date: the date the data refers to (YYYY-MM format for month and YYYY-MM-DD for day).

Card: Card type.

Channel: Channel type.

Category (optional): commercial category label (when missing the row refers to all the existing categories)

Merchants: number of merchants with registered activity.

Cards: number of different cards.

Txs: number of physical purchase transactions.

Avg. amount: average transaction amount.

Destination type: type of purchase destination of customers living in the registration area. Values: [subregions, localities, zipcodes].

Destination: the ID of purchase destination.

Merchants by destination: number of different merchants by destination of the clients of the registry zone.

Cards by destination: number of different cards by destination of the clients of the registry zone.

Txs by destination: number of physical purchase transactions by destination of the clients of the registry zone.

Avg. amount by destination: average transaction amount by destination of the clients of the registry zone.

Possible combinations:

SOURCE CHANNEL DESTINATION\_TYPE

bbva pos localities

bbva pos subregions

bbva pos zipcodes

Territorial Units

Subregions

Specified by its two-digit identifier following the INE Subregion Codes

Localities

Specified by its five-digit identifier following the INE Locality Codes.

Zipcodes

The Spanish zipcode format is a 5 digit numeric field. Correos zip-codes finder.

List of relations between zipcodes, localities, subregions and regions (in CSV)

(Download).

Tiles

LONGITUDE and LATITUDE are the parameters that control the placement of the geographic tiles

LONGITUDE and LATITUDE are rounded to the third significant digit, but the third significant digit will be rounded to 0 or 5. For example: 40.336 to 40.335, 40.311 to 40.310, etc

To calculate statistics by tiles, Spain has been split into 500 x 500 meters quadrants. LONGITUDE and LATITUDE indicate the central point of each quadrant.

Note: This territorial unit is only available in the datasets grouped by month.

Tiles

List of relations between tiles, localities, subregions and regions (in CSV)

(Download).

Census Tracts

Census tracts are identified by a ten-digit code which is formed using the following INE identifiers:

Subregion Code (2 digits)

Locality Code (3 digits)

District Code (2 digits)

Census tract (3 digits)

Example: 28 (Subregion) + 123 (Locality) + 01 (District) + 024 (Census tract) = 2812301024

Census tract information can be found at the INE Census tract and Digital mapping info

Note: This territorial unit is only available in the datasets grouped by month.

The following image shows the different Census tracts which can be found in the Rivas Vaciamadrid locality:

census tract

Cards

Type of cards owner availabl:

all: all cards.

bbva: BBVA cards.

national: spanish cards.

foreign: non spanish cards.

Channels

Type of channel available.

pos: generic point of sale.

bbva\_pos: BBVA point of sale.

opos: online point of sale.

bbva\_opos: online BBVA point of sale.

**Categories**

Some of the provided statistics are separated by category. The following list contains all the categories that can be found across the downloads service:

* **es\_health:** Health
  + **es\_pharmacy:** Pharmacy
  + **es\_hospital:** Hospital and medical consultory
  + **es\_opticians:** Optician’s shops
* **es\_transportation:** Transport
  + **es\_parking:** Parkings
  + **es\_seatransport:** Sea transportation
  + **es\_carrental:** Car rental
  + **es\_urbantransport:** Urban transport: bus, metro and train
  + **es\_airline:** Airlines
  + **es\_boatrental:** Boat and aircraft renting
  + **es\_gas:** Gas Stations
  + **es\_bus:** Bus
  + **es\_toll:** Toll
  + **es\_train:** Trains
  + **es\_taxi:** Taxi
* **es\_tech:** Technology
  + **es\_photo:** Photography
  + **es\_musicalinstrument:** Musical instruments
  + **es\_techbig:** Computers and appliances: chains
  + **es\_techsme:** Computers and appliances: small shops
  + **es\_telephony:** Telephony: handset sales
* **es\_food:** Food
  + **es\_supermarket:** Supermarkets
  + **es\_foodsme:** Food: small shops
* **es\_home:** Home
  + **es\_floristsbig:** Gardening and Floristry: chains
  + **es\_homesme:** Furniture and decoration: small shops
  + **es\_diybig:** DIY: supermarkets
  + **es\_homebig:** Furniture and decoration: chains
  + **es\_floristssme:** Gardening and Floristry: small shops
  + **es\_diysme:** DIY: small shops
* **es\_sportsandtoys:** Sports and toys
  + **es\_sportandtoys:** Toys and sport articles
  + **es\_toys:** Toys: chains
  + **es\_sport:** Sportive activities
  + **es\_sportbig:** Sport: chains
* **es\_leisure:** Leisure
  + **es\_ticketsale:** Tickets
  + **es\_museum:** Museums and tourist attractions
  + **es\_leisuretime:** Shows
  + **es\_bet:** Bet
* **es\_contents:** Books and press
  + **es\_press:** Press
  + **es\_books:** Books
* **es\_fashion:** Fashion
  + **es\_jewelry:** Jewelry and watches
  + **es\_fashionsme:** Fashion: small shops
  + **es\_leather:** Leather goods stores
  + **es\_shoe:** Shoes
  + **es\_fashionbig:** Fashion: chains
* **es\_hotelservices:** Accomodation
  + **es\_hotel:** Accomodation
* **es\_travel:** Travel
  + **es\_travelagency:** Travel agency: local
  + **es\_travelweb:** Travel agency: web
* **es\_barsandrestaurants:** Bars and restaurants
  + **es\_fastfood:** Fast food restaurants
  + **es\_restaurant:** Restaurants
  + **es\_pub:** Pubs and night clubs
  + **es\_cafe:** Bars and cafes
* **es\_propertyservices:** Real state
  + **es\_realestate:** Real estate
* **es\_auto:** Auto
  + **es\_wash:** Carwash
  + **es\_car:** Dealers, garage and spare parts
  + **es\_cartest:** Vehicle roadworthiness
* **es\_bank:** Banking Services
  + **es\_atm:** ATM
  + **es\_branch:** Banking branches
* **es\_otherservices:** Other services
  + **es\_mail:** Packaging and storage
  + **es\_unknown:** Unknown
  + **es\_veterinarian:** Veterinary and pets
  + **es\_donation:** Donations
  + **es\_others:** Other
  + **es\_tobacconists:** Tobacconist
  + **es\_fplenish:** Phone plenish
  + **es\_insurance:** Insurance
  + **es\_video:** Video store and pay TV
  + **es\_tax:** Public tax
  + **es\_goods:** Bazaar
  + **es\_drycleaner:** Dry Cleaning & Laundry
  + **es\_phonebooth:** Cybercafe
  + **es\_education:** Teaching
  + **es\_funeral:** Funeral
  + **es\_dutyfree:** Duty free
* **es\_wellnessandbeauty:** Wellness and beauty
  + **es\_wellness:** Wellness and Beauty
  + **es\_beauty:** Hair and beauty
  + **es\_drugstorebig:** Cosmetics and perfumery: chains
  + **es\_drugstoresme:** Cosmetics and perfumery: small shops
* **es\_hyper:** Hypermarkets
  + **es\_mall:** Department stores
  + **es\_hypermarket:** Hypermarket

**Sandbox**

Sandbox datasets will allow you to analyze the capabilities of our service. These datasets provide you with mock information generated from transactions that took place from October 1st, 2015 to November 30rd, 2015. The amount of information available depends on the territorial unit:

**Subregions:** Madrid Region

**Localities, Zipcodes and Tiles:** Madrid City

**Census tract:** Rivas Vaciamadrid (Madrid)

**Behaviour**

**Age Ranges**

Some statistics make use of age ranges to return the information. The following list shows the different age ranges that can be returned by the API:

* <=24
* 25-34
* 35-44
* 45-54
* 55-64
* >=65
* U: Unknown

**Gender Values**

Some statistics provide information grouped by gender. The following table shows the different genders considered by the API and the codes used to represent them:

Code Gender

F Female

M Male

E Enterprise

U Unknown

**Merchants Categories**

This service returns information about the business categories, with their literal description in Spanish or English.

**Data Filtering**

In order to prevent any entity identification and preserve anonymity, some rows of the datasets are filtered at different levels.

My purpose is the analysis of Spain’s patterns of consumption within the period of 2015 (specifically Madrid) in order to be able to provide recommendations to potential investors taking into account the main variables discussed earlier.

For the purposes of Project II, I’ll be focusing mainly on bars and restaurants categories, which are divided in four subcategories: (i) fast food restaurants, (ii) restaurants, (iii) pubs and night clubs, (iv) bars and cafes.

**Research questions to answer:**

1. **Database Consumption\_Pattern | En las y va la variable amount y en la x days of the week. La idea es que se vea el max amount y el min amount y con un punto el average amount. Acá se puede poner una caja que el punto mínimo y el máximo sea la cajita**
2. Database Consumption\_Pattern |Acá se van a tener dos gráficas, una de puros datos del lunes y otra de puros datos del viernes. En cada una se va a graficar en bolitas el monto de las transacciones. En las x van las categorías y subcategorías de restaurantes y en las y se gráfica transacciones
3. Database Consumption\_Pattern |en las x va la hora (sin separar por día de la semana) y en la y el número de transacciones y en bolas en monto del average de las transacciones
4. Origin\_distribution\_genders |en las x va gender y en la y el número de tarjetas eso se podría grafica de barra y poner en un punto el average amount y en punto también el número de transacciones indicado. Para esos dos puntos se podría utilizar una imagen para cada uno.  **La misma idea pero poner en la x la edad.**
5. Origin\_distribution\_categories |En las x poner lo de las categorías por restaurantes y en barras el numero de tarjetas (esta es la y) y el average amount y transacciones en punto marcado por cada x.
6. Origin\_distribution\_categories |En las y poner merchants by category y en la x poner categorías y subcategorias de restaurantes
7. Where do the customers, given each category for the local, come from?
8. Given each category, how does every Company performs?
9. Given each category, what is the percentage of foreign consumers?
10. What percentage of each population’s zone does each category cover?

**API sets to be used:**

* BBVA API\_Market
* Google Maps (location and recommendations)

**Rough breakdown of taks:**

1. Connect to the BBVA API and analyze the information given (cleaning it and analyze which columns are useful)
2. After analyzing BBVA’s information the datasets to use will be the following:
   1. Consumption Pattern
   2. Consumption Pattern:
3. After analyzing BBVA’s available information, take into account to use Google Maps API to be able to segment the information by location
4. Look if INE’s information available is useful for the analysis (i.e. average income by location)
5. Look up for the availability of Yelp’s information (e.g. available stars per category – this category is the one provided by BBVA, look if this applies)

**Link to datasets:**

<https://github.com/gcarrillomonroy/Opportunities-to-Invest-in-Spain-Bars-Restaurants>

**Link to the metadata:**

<https://www.bbvaapimarket.com/documentation/bbva/paystats-download>

<https://www.bbvaapimarket.com/documentation/bbva/paystats>

**A sketch of the final design:**

[TBD]

**Link to the primary GitHub repository I'll be housing my work in:**

<https://figshare.com/account/home#/projects/61610>