|  |
| --- |
|  |
| Alumnos: |
| * Carlos Alcalá-Zamora Acero |
| * Dan Gonzalvo Guillén |
|  |
| Dirección: Julio Merelo Casado |
| **04/09/2017** |
|  |

# Análisis de hábitos de consumo en

# tarjetas de crédito usando Api BBVA.

Contenido

[Análisis de hábitos de consumo en tarjetas de crédito usando Api BBVA. 1](#_Toc492204385)

[Planteamiento del problema. 3](#_Toc492204386)

[Objetivo general y objetivos específicos, tecnológicos y de negocio. 3](#_Toc492204387)

[Aportaciones fundamentales esperadas. 3](#_Toc492204388)

[Breve estado del arte. 4](#_Toc492204389)

[Planteamiento de la propuesta. 4](#_Toc492204390)

[Metodología a seguir. 4](#_Toc492204391)

[Arquitectura inicial 8](#_Toc492204392)

[Desarrollo de la propuesta. 9](#_Toc492204393)

[Llamadas a la api PayStats 9](#_Toc492204394)

[Category Distribution 10](#_Toc492204395)

[Basics Stats 11](#_Toc492204396)

[OriginDistribution 11](#_Toc492204397)

[Age Distribution 12](#_Toc492204398)

[Gender Distribution 12](#_Toc492204399)

[Payment Distribution 13](#_Toc492204400)

[Destination Distribution 13](#_Toc492204401)

[Consumption Pattern 14](#_Toc492204402)

[Codificación Web 14](#_Toc492204403)

[Visualización 16](#_Toc492204404)

[Análisis Categorías 16](#_Toc492204405)

[Análisis Clientes 21](#_Toc492204406)

[Análisis Localización 28](#_Toc492204407)

[Implementación Web 30](#_Toc492204408)

[Control de versiones 31](#_Toc492204409)

[Arquitectura Online 32](#_Toc492204410)

[Conclusiones. 33](#_Toc492204411)

[Bibliografía. 34](#_Toc492204412)

# Planteamiento del problema.

BBVA PayStats recoge datos anónimos y estadísticos de millones de transacciones realizadas en las terminales POS de BBVA. Ofrece información de forma diaria desde 2014, disponible con datos geográficos o con código postal, podemos consultar datos básicos estadísticos de la factura de cada cliente, el número de transacciones o tarjetas usadas.

Con esta información podemos ser capaces de saber el tipo de consumidores que hay en una zona o sus patrones de consumo y en general estadísticas que puedan ayudar a la toma de decisiones para los negocios de la zona, como, por ejemplo, que tipos de productos y servicios se deben ofrecer o el horario que debería tener un comercio para optimizar sus ventas.

# Objetivo general y objetivos específicos, tecnológicos y de negocio.

Nuestro objetivo principal es la recopilación de la información para su tratamiento, obteniendo las métricas principales que nos permitan entender el comportamiento de los clientes. Una vez finalizado el proyecto podremos responder de forma clara y de un solo vistazo a cuestiones como:

* Tipos de compras realizadas.
* Origen de los clientes.
* Gasto de los clientes.
* Localización y tipo de gasto.

Para ello comenzaremos recopilando la información en formato JSON ofrecida por el api, al ser información ya recogida no necesitaremos un método de almacenamiento previo, por lo que enlazaremos el api con el tratamiento back en python. Recopilaremos toda la información y la filtraremos para generar conjuntos de datos que podamos representar y nos permitan contestar a preguntas como las anteriores. Una vez tratada la información ofreceremos una web sencilla en la que integraremos las gráficas y datos. Esta página web la publicaremos en una plataforma como Firebase para que sea accesible desde cualquier navegador web. De esta forma cualquier persona podrá visualizar las estadísticas de gasto de los usuarios, constituyendo una información muy valiosa para cualquier emprendedor que quiera abrir su negocio.

## Aportaciones fundamentales esperadas.

Nuestro objetivo es ofrecer una web clara y sencilla que nos muestre un conjunto de gráficas que interpreten los datos ofrecidos por el api. Esta web será accesible y nos mostrará los resultados para poder responder a las preguntas planteadas.

# Breve estado del arte.

Actualmente se disponen de aplicaciones privadas de pago que ofrecen una información similar para cada una de las entidades bancarias, pero de forma pública no se dispone de esta información. Gracias al api de BBVA podemos acceder a estos datos, y con este pequeño trabajo podremos ofrecer toda esta información a los nuevos emprendedores.

De esta manera el banco cumple con la directiva de servicios de pagos (PSD2) de la Unión Europea en las que las entidades financieras de la eurozona deben proporcionar acceso a la información de sus clientes a terceros cuando el cliente haya dado su consentimiento explícito. La fecha límite que ha dado la Unión Europea para llevar a cabo esta directiva es enero de 2018. Esta normativa pretende promover un mercado con mayor igualdad de oportunidades promoviendo la entrada de nuevos proveedores y servicios a la vez que protege de manera más fiable a los clientes disminuyendo el fraude.

# Planteamiento de la propuesta.

## Metodología a seguir.

Con una metodología Ágil, en un equipo de dos trabajadores y con la ayuda de un director de proyecto, hemos afrontado el reto. Para la gestión del proyecto hemos utilizado la herramienta Taiga de gestión de proyectos. En ella se ha creado el proyecto para la gestión del equipo y organización del desarrollo del proyecto. Taiga nos permite recoger todos los cambios realizados. Podemos visualizar todas las tareas directamente desde la interfaz de Taiga: <https://tree.taiga.io/project/danciff-apibbva/>

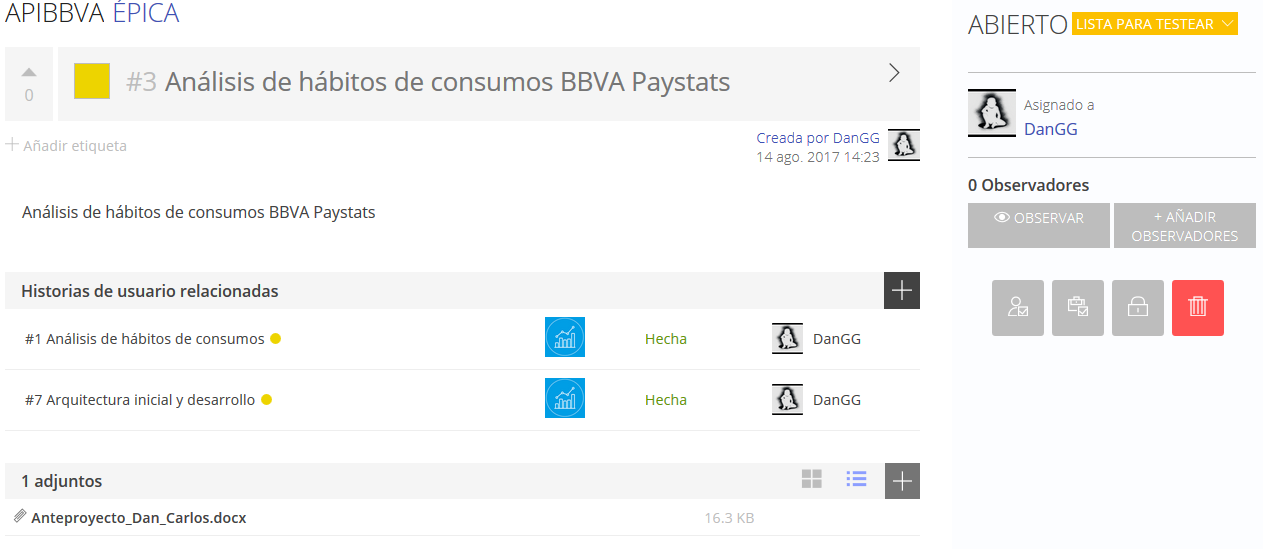


El proyecto se ha organizado en 4 tareas épicas, las cuales engloban todo el proyecto:



1. La primera de las tareas épicas es Análisis de hábitos de consumos BBVA Paystats, esta dispone de dos historias relacionadas:

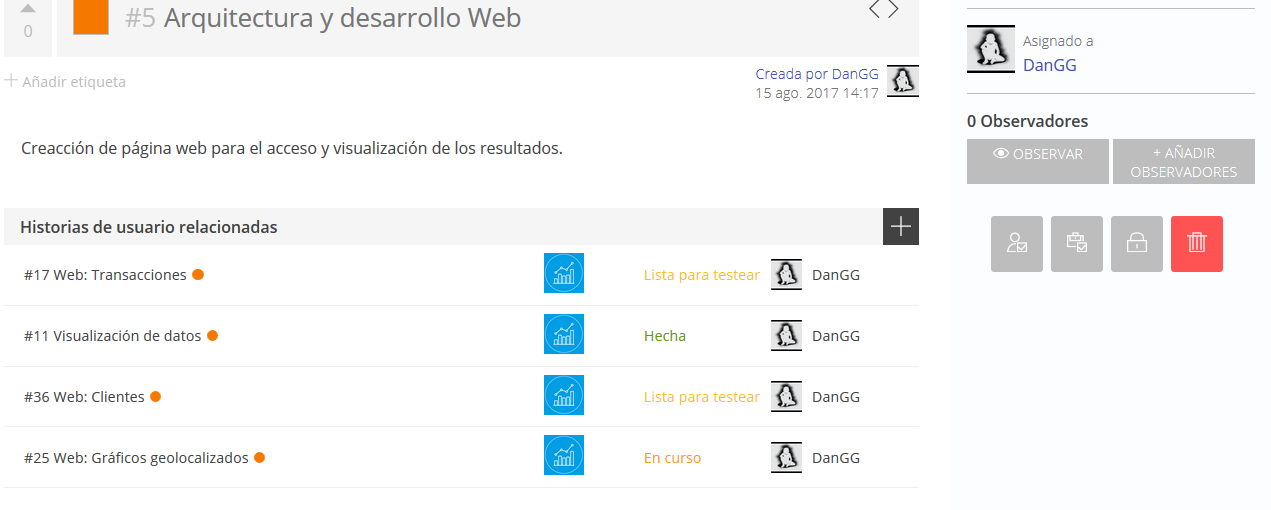
* Análisis de hábitos de consumos: Estudio de los datos ofrecidos y funcionamiento de Paystats.
* Arquitectura inicial y desarrollo: Diseño de la arquitectura inicial para cumplir los objetivos principales del proyecto.



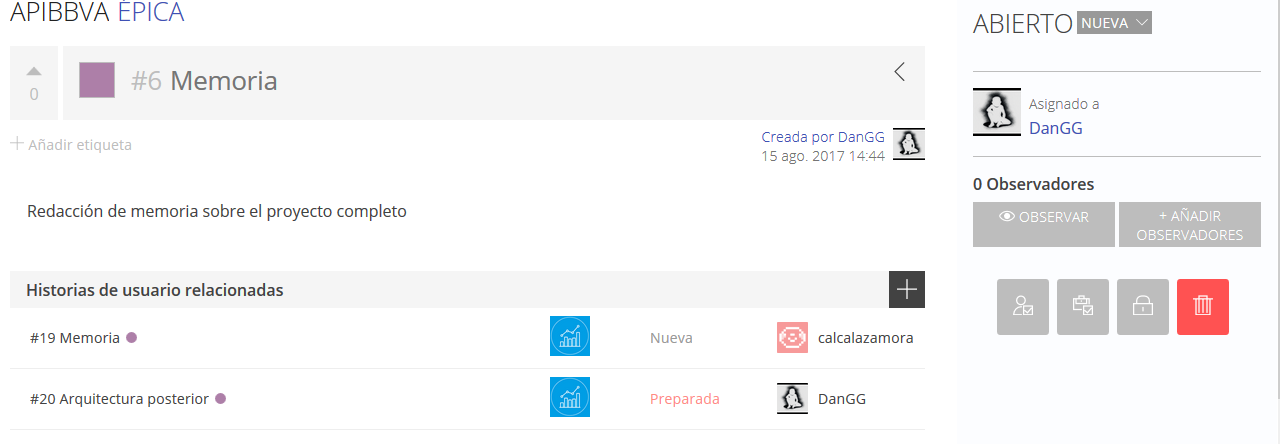
1. La segunda de las tareas se centra en el análisis y desarrollo de datos para la creación de gráficos, a su vez engloba un conjunto de tareas que referencian a los diferentes tipos de datos que podemos obtener:
   * Procesamiento de datos: Estudio general.
   * Localización: Relativo al estudio de los datos según el origen.
   * Categoría: Filtrado de datos de transacciones según el tipo.
   * Clientes: Relacionado con el análisis de las transacciones según el tipo de cliente.



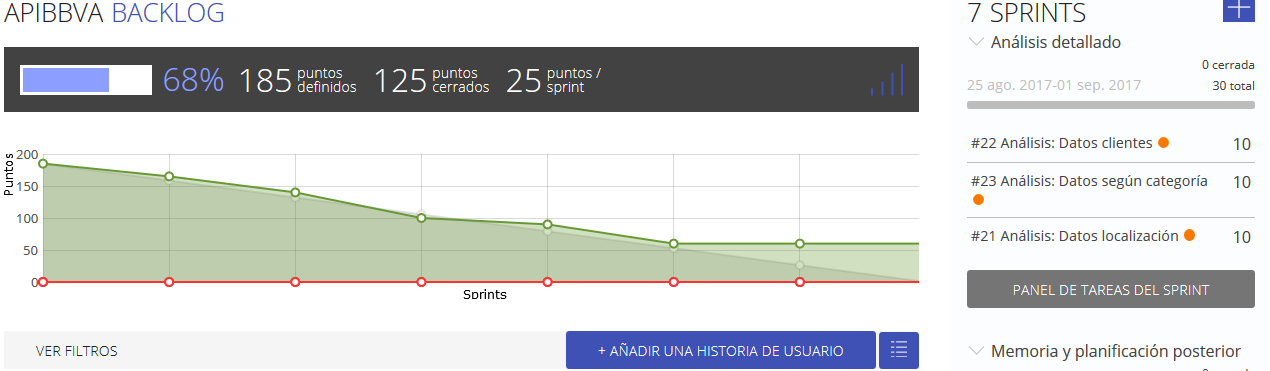
1. La tercera tarea está relacionada con la visualización de resultados, siguiendo el patrón de análisis, contiene 4 tareas:
   * Visualización: Corresponde a la estructura principal de la web.
   * Transacciones: Nos muestra el conjunto de gráficas relacionadas con los datos de transacciones analizados.
   * Clientes: Visualización de datos según el tipo de cliente.
   * Localización: Datos según el origen o destino de las transacciones.



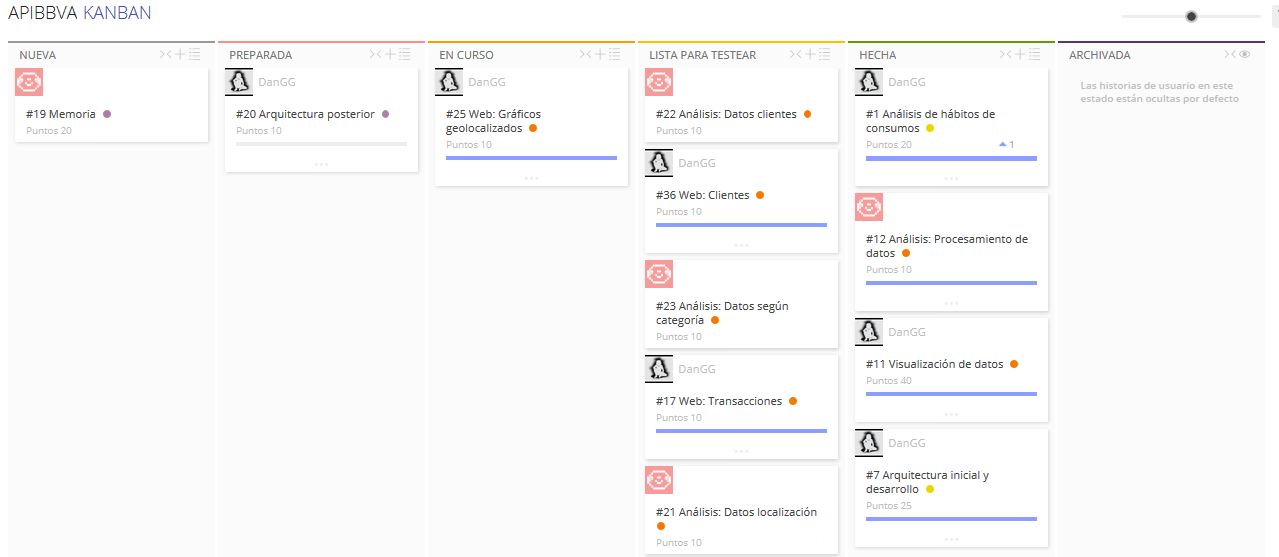
1. Por último tenemos la tarea de Memoria, que engloba la documentación del proyecto y su planificación futura:
   * Memoria: Generación de toda la documentación necesaria.
   * Arquitectura: Diseño del posible desarrollo para adaptar el proyecto para el funcionamiento con datos actualizados de forma continua.



El backlog del proyecto nos permite añadir y organizar las tareas, todas las tareas se han organizado en 7 Sprints, el inicio definitivo se ha realizado el 24 de Julio de 2017, y el proyecto finaliza el 4 de Septiembre de 2017.



Por último disponemos de una vista Kanban que nos permite de un solo vistazo comprobar el estado de las tareas y en qué fase de desarrollo se encuentran.



Todo lo anterior nos permite realizar una gestión clara y ordenada del proyecto, para conseguir el objetivo sin complicaciones.

## 

## Arquitectura inicial

Paystats nos ofrece toda la información por medio de consultas en formato json. Gracias a esta forma de recogida de datos, no necesitamos de un software específico como Kafka para recolectar los datos, ya que la aplicación nos ofrece los datos ya de forma ordenada. La información que nos ofrece es de un sandbox con todos los datos del año 2015, por lo que no necesitamos tratar información en tiempo real, y los datos no se van a actualizar una vez publicados.

Por todo lo anterior se va a realizar el procesamiento de toda la información y posteriormente una vez conseguidos los resultados se realizara la publicación en la web. Para ello se realizará una propuesta simple de arquitectura, primero se enlazará el entorno de trabajo en Python con la aplicación, en el que se realizará el procesamiento de los datos, y una vez generados los resultados se generarán archivos de datos, que se incrustarán en el desarrollo web para la generación de las diversas gráficas e informes que nos permitirán responder a las preguntas planteadas.

Jupyter

(Procesamiento de datos)

Paystats

(Origen de datos)

Web

(Visualizacion)

Ficheros de datos

Como podemos ver primero haremos uso de un entorno de desarrollo para trabajar los datos ofrecidos por paystats en python. Haciendo uso de un entorno de trabajo como Jupyter, se ha creado un proyecto de desarrollo en Python que permite realizar la conexión con la aplicación, desde el entorno de trabajo podemos conseguir todos los datos en formato json, se realiza el procesamiento y una vez obtenidos los datos necesarios para representar la información, estos son almacenados en archivos planos.

El segundo paso es la representación de la información procesada, para ello se ha desarrollado una web que permita visualizar de forma clara todos los datos. Primero se ha escogido una plantilla base con un estilo limpio y claro, a partir de ello se han realizado todas las modificaciones y se ha expandido la web para albergar todos los datos. Los datos se recogen de los ficheros de salida y se incrustan en la web, de forma que toda la ejecución se realiza desde el navegador. Para que esta web sea accesible se ha realizado el despliegue y hosting de la web gracias a Firebase.

# Desarrollo de la propuesta.

## Llamadas a la api PayStats

La api PayStats de BBVA nos proporciona una serie de llamadas desde las cuales poder acceder a los datos que nos ofrece.

Estas llamadas se dividen en dos, unas en las que la búsqueda de datos se hace en función de un código postal elegido y otras en las que la búsqueda de datos se lleva a cabo a través de coordenadas de latitud y longitud.

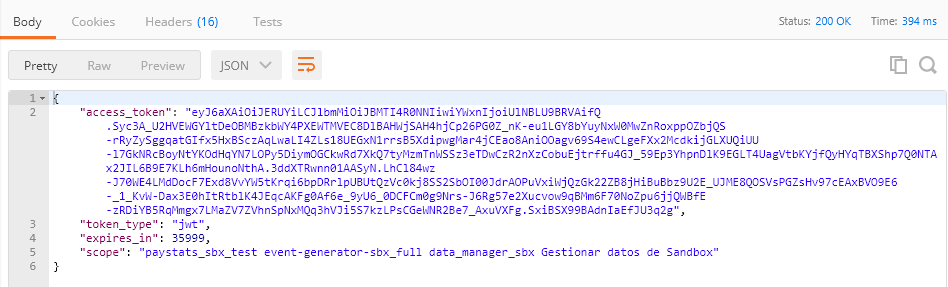
Para hacer la búsqueda por coordenadas la api ha dividido el mapa en secciones cuadradas de 500 metros cuadrados, al introducir una latitud y una longitud se seleccionarán datos de todo el cuadrado donde esté la coordenada introducida. Al hacer la búsqueda por código postal la api nos seleccionará los datos que se hayan generado dentro del código postal introducido.

Ambas formas cuentan con los mismos tipos de llamadas, sólo se diferencian en la forma de buscar los datos. Para el presente trabajo vamos a utilizar solamente el código postal para hacer las búsquedas de los datos que necesitaremos.

Las llamadas disponen de una serie de opciones que podemos usar al hacer la llamada para cambiar la forma en que se muestra la información o entre que fechas se buscarán los datos.

Según el tipo de llamada, esta contará con una serie de opciones. Las opciones que son obligatorias incluir en todas las llamadas son las fechas de inicio y final a partir de las cuales se buscarán los datos. En este trabajo usaremos como fecha de inicio enero del año 2015 y como fecha final diciembre del año 2015, ya que el modo sandbox, que es el que estamos utilizando, solo cuenta con datos del año 2015. Otra de las opciones más comunes es las categorías, incluyendo esta opción solo se mostrarían los datos de la categoría que hayamos indicado.

Para acceder a las llamadas se ha usado Python. Para hacer una llamada a la api de PayStats en el modo sandbox se necesita un “Access Token” el cual podemos conseguir creando una cuenta en “BBVA Api Market” y creando una nueva aplicación dentro de nuestra cuenta, estos datos lo introducimos en el programa “Postman” el cual nos proporciona una especie de contraseña que deberemos introducir en el código al hacer la llamada. Este “Access Token” solo tiene validez para un día y habrá que volverlo a obtener cada día que lo necesitemos.



La mayoría de las llamadas nos devolverán datos del número de tarjetas que se han utilizado, número de transacciones con tarjetas, número de establecimientos en los que haya habido al menos una transacción con tarjeta y la compra media de los clientes que usan tarjetas, todo esto dentro del código postal especificado. Estos son los datos más comunes, pero según el tipo de llamada de nos mostrarán unos datos u otros. El api de PayStats nos proporciona estos datos en formato Json.

A continuación, se van a explicar los distintos tipos de llamadas que dispone el api de PayStats.

### Category Distribution

Esta llamada nos proporciona los datos de cada categoría de compra media, número de transacciones, número de establecimientos y el número de tarjetas, todos estos agrupados por meses, del código postal introducido. Además, hemos calculado la venta total para ofrecer más información al usuario.



### Basics Stats

Esta llamada nos proporciona datos de compra media, número de transacciones, número de establecimientos y el número de tarjetas de la fecha seleccionada agrupada por meses y también datos sobre la transacción máxima y mínima que se ha obtenido en un mes, la hora del día en que más y menos transacciones se llevan a cabo y el día de la semana en que más y menos transacciones se llevan a cabo del código postal indicado. Esta llamada incluye una opción para desglosar los datos según la categoría de negocio, la cual se ha utilizado para conseguir unos datos más útiles para el usuario final.



### OriginDistribution

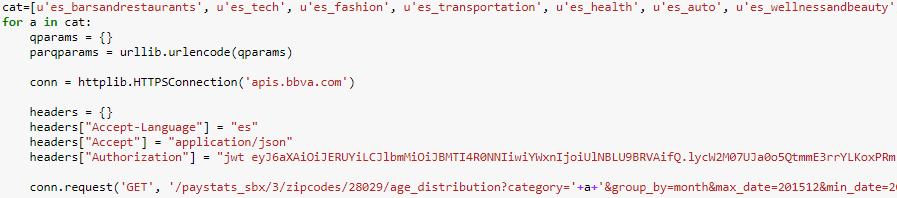
Esta llamada nos proporciona datos de compra media, número de transacciones, número de establecimientos y el número de tarjetas de la fecha seleccionada agrupada por meses del código postal indicado, pero, a diferencia de otras llamadas, no permite agrupar los datos por categorías. Esta llamada dispone de una opción para agrupar los datos en función al sexo, la edad o al sexo y la edad a la vez, por ejemplo, sería posible ver el número de transacciones que se han realizado en un código postal por mujeres de entre 35 y 44 años.



### Age Distribution

Esta llamada nos proporciona datos de compra media, número de transacciones, número de establecimientos y el número de tarjetas de la fecha seleccionada agrupada por meses del código postal indicado y nos agrupa estos datos por edad, además, se ha calculado la venta total para conseguir ampliar la información que nos proporciona. La edad se encuentra dividida en rangos de edades.

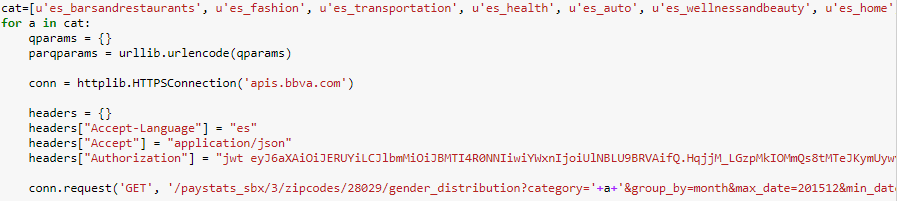
Esta llamada se diferencia de la llamada “Age.Gender Distribution” en que en esta sí que se puede agrupar por categoría de negocio para obtener datos más útiles para el usuario.



### Gender Distribution

Esta llamada se comporta de la misma manera que la de “Age Distribution” y proporciona los mismos tipos de datos, pero esta vez los agrupa por el sexo del cliente en vez de por la edad. Para sexo la aplicación nos proporciona 4 valores diferentes que son “F” para femenino, “M” para masculino, “U” para sexo desconocido (“Unknown” en inglés) y “E” si el cliente es una empresa.

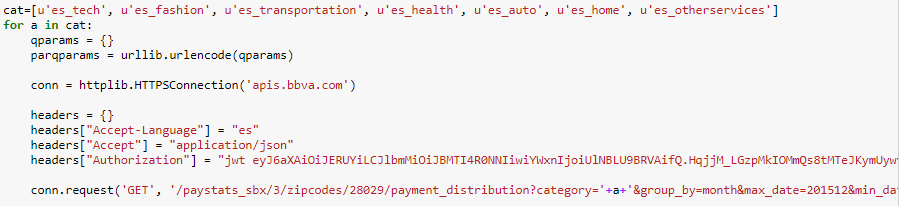
Al igual que en la llamada “Age Distribution” en esta también es posible agrupar por categoría de negocio para obtener unos datos más fáciles de interpretar.



### Payment Distribution

Esta llamada se comporta de la misma manera que la de “Age Distribution” y “Gender Distribution” y proporciona los mismos tipos de datos, pero esta vez los agrupa por el importe de la transacción. El importe de la transacción aparece dividido en una serie de rangos de cantidad, por ejemplo, transacciones de menos de 30 Euros o transacciones de entre 30 y 50 Euros.

Al igual que en la llamada “Age Distribution”y “Gender Distribution” en esta llamada también es posible agrupar por categoría de negocio, indicándolo en las opciones al hacer la llamada, para obtener unos datos más útiles y fáciles de interpretar.



### Destination Distribution

Esta llamada nos proporciona datos de compra media, número de transacciones, número de establecimientos y el número de tarjetas de la fecha seleccionada agrupada por meses del código postal indicado y los agrupa según el destino que tienen esas compras. Esta llamada tiene una opción que nos permite agrupar los datos según el código postal de destino y directamente con la localidad a la que va destinada la compra.

Esta llamada no dispone de opción para agrupar por categorías, pero tiene una opción para escoger que tipos de tarjetas se utilizan para obtener los datos, tarjetas de BBVA, todas las tarjetas o solo tarjetas nacionales. Para este trabajo hemos usado la opción por defecto que son las tarjetas de BBVA ya que el modo sandbox solo proporciona datos de este tipo de tarjetas.

****

### Consumption Pattern

Esta llamada proporciona patrones de consumo típicos por la hora del día y el día de la semana para un área en particular. Esta llamada nos muestra datos sobre compra media, número de transacciones, número de establecimientos, número de tarjetas, la transacción con importe máximo y mínimo y la desviación típica del importe de las transacciones, todos estos datos se encuentran agrupados por la hora y el día de la semana. Además, esta llamada dispone de una opción para indicar la categoría de negocio que queramos analizar y así expandir la información.

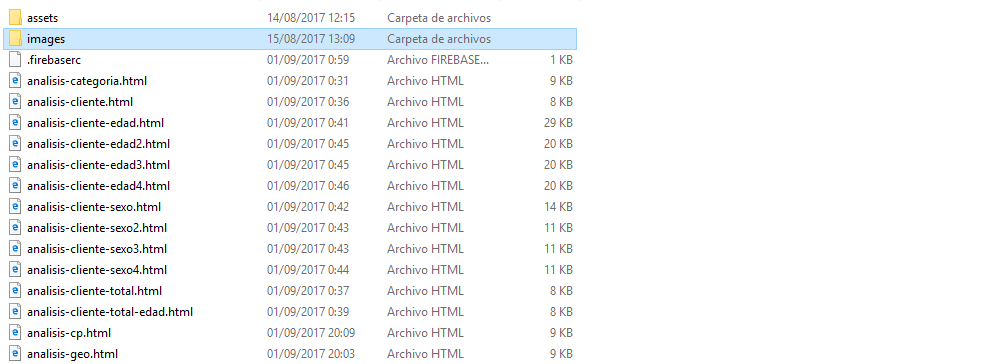


## Codificación Web

Para la visualización de los resultados obtenidos en el análisis de datos, se ha desarrollado una web basada en Html5. Para ello nos hemos basado en recursos abiertos que nos permiten obtener un resultado visual limpio, claro y efectivo para mostrar los resultados. Primero hemos partido de una plantilla web ofrecida por Html5up (<https://html5up.net/>), se ha escogido la plantilla Verti, que nos facilita todos los estilos principales de la web. Por otra parte se ha escogido Google Charts como librería para la generación de gráficos (<https://developers.google.com/chart/interactive/docs/quick_start>).

Con la ayuda de estos dos recursos se ha desarrollado la web disponible desde cualquier navegador: <https://appbbva-9b62b.firebaseapp.com/>

La web está estructurada en diferentes páginas, las cuales son referenciadas para mostrar todos los resultados obtenidos, además disponemos de los recursos para los estilos de la web, las imágenes utilizadas y archivos de configuración.



En todas las páginas disponemos de un menú que nos permite navegar por las diferentes páginas de la web.



La web está distribuida para ver de forma clara los resultados, dispone de una estructura base de páginas, desde la página inicial, al resumen o la página de contacto. Todos los resultados se han englobado en la sección de análisis. Desde un menú desplegable tenemos acceso a cada una de las secciones, en cada una de ellas además de la descripción de los resultados y otros elementos web, se ha incluido un gráfico explicativo adecuado para la descripción de los datos.



En el cuerpo de la página hacemos referencia al gráfico generado, este puede ser un gráfico de área como el del ejemplo, cualquier otro tipo de gráfico, o incluso un tabla, el desarrollo será similar, por último añadimos los elementos necesarios para ver los resultados de forma clara. A continuación veremos de forma detallada cada uno de los gráficos y análisis generados.

## Visualización

A continuación, se ofrece una descripción de cada gráfica que aparece en la web y a partir de qué tipo de llamada se han obtenido los datos.

### Análisis Categorías

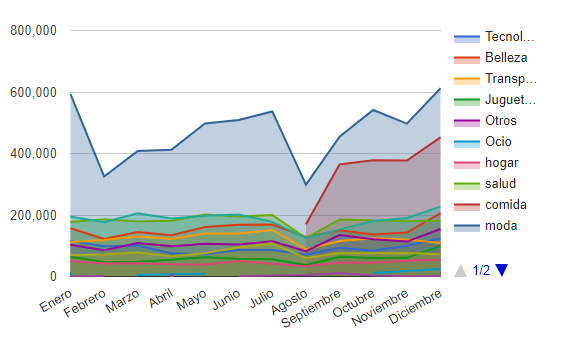
Este apartado incluye gráficas sobre el número de transacciones realizadas con tarjetas, el número de tarjetas que se han usado, el número de establecimientos en el que se hayan realizado al menos una transacción con tarjeta, la compra media y la venta total de las transacciones realizadas con tarjetas dentro del código postal que hemos analizado. En este apartado las gráficas están desglosadas por las categorías de negocio que el api PayStats nos ofrece.

Todas las gráficas que se muestran en esta sección se explican a continuación.

#### Ventas Categoría

Esta gráfica muestra las ventas por mes que tiene cada una de las categorías de negocio de las que dispone la api PayStats. La web nos ofrece una gráfica que muestra todas las categorías distinguidas por colores con el año del cual están disponibles los datos dividido por meses. Además de esto la web permite seleccionar sólo los meses que el usuario quiera analizar para una mejor visualización. Con estos datos el usuario puede ver cómo evolucionan las ventas a lo largo del tiempo para tomar decisiones en consecuencia.

Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Category Distribution”.

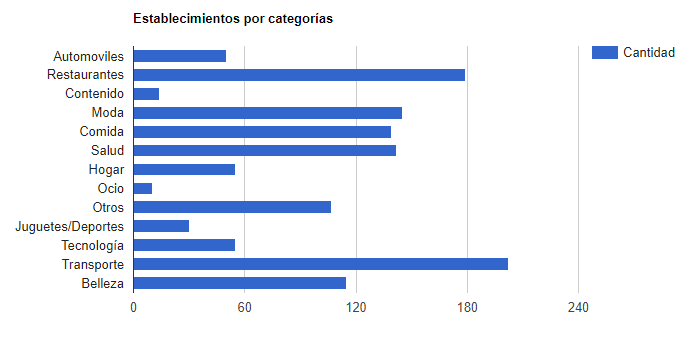


#### Número de Establecimientos por Categoría

Con esta gráfica el usuario puede ver el número de establecimientos que hay en cada categoría de negocio. Con esta información el usuario puede ver que categorías están más saturadas dentro del código postal que estamos analizando y en cual hay pocos establecimientos y, por lo tanto, podría encontrar una oportunidad de negocio gracias a estos datos. Esta información puede ser cruzada con la gráfica de ventas por categoría para conseguir más información al respecto.

Ya que en la aplicación los datos se obtienen por meses, para obtener el número de establecimientos se ha cogido el valor máximo del año de cada categoría ya que sólo se cuentan aquellos establecimientos en los que haya habido al menos una transacción. De esta manera se obtiene un número más aproximado al número real de establecimientos de una zona.

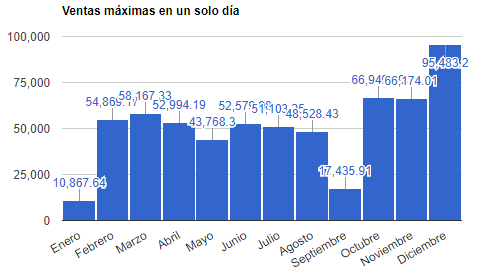
Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Category Distribution”.

****

#### Horario

Esta parte de la web muestra al usuario una gráfica con la venta máxima de cada mes que se ha conseguido en un solo día en la zona en la que se está haciendo el análisis, además también se incluye una tabla dónde se muestra por cada mes cual ha sido la hora y el día de la semana en la que ha habido mayor cantidad de ventas. Estos datos pueden ser útiles para saber cuál es el momento del día en que hace falta más personal que de costumbre.

Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Basic Stats” y “Consumption Pattern”.

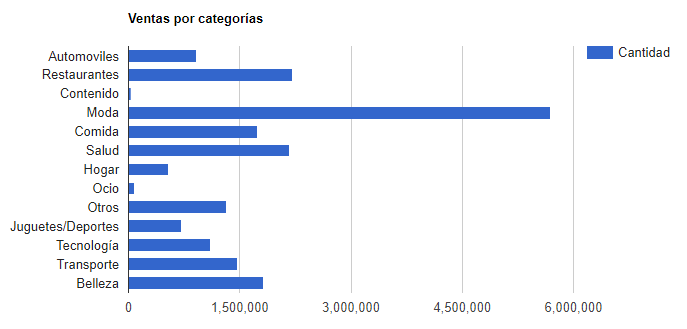




#### Venta Total

Esta parte de la web muestra un gráfico de barras con la venta total del último año para cada categoría de negocio. Esta información es útil para encontrar un lugar en el que abrir un nuevo negocio en un sitio que no esté ya muy saturado. Sería buena idea cruzar esta información con la del número de establecimientos por categoría para tener una idea más amplia de todo esto.

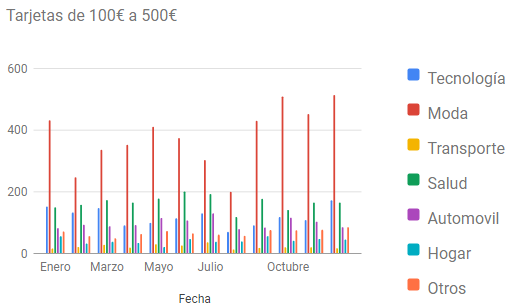
Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Category Distribution”.



#### Tarjetas por Cantidad

Estas gráficas muestran el número de tarjetas que han realizado al menos una transacción en cada una de las categorías durante cada mes. Hay una gráfica para cada rango del importe de la transacción. Por ejemplo, el usuario puede saber cuántas tarjetas distintas se han utilizado en abril, para realizar compras menores de 30 Euros en artículos de moda.

Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Payment Distribution”.

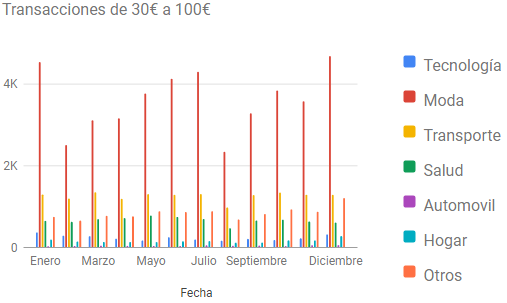


#### Transacciones por Cantidad

Estas gráficas muestran el número de transacciones que se han realizado en cada una de las categorías al mes dividida por rangos del importe de dicha transacción. Hay una gráfica para cada rango del importe de la transacción. Por ejemplo, el usuario puede saber cuántas transacciones se han realizado en diciembre, para realizar compras de entre 100 y 500 Euros en artículos de tecnología.

Esta apartado y el anterior ofrecen al usuario información muy importante sobre el tipo de compra según la categoría y la fecha.

Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Payment Distribution”.



### Análisis Clientes

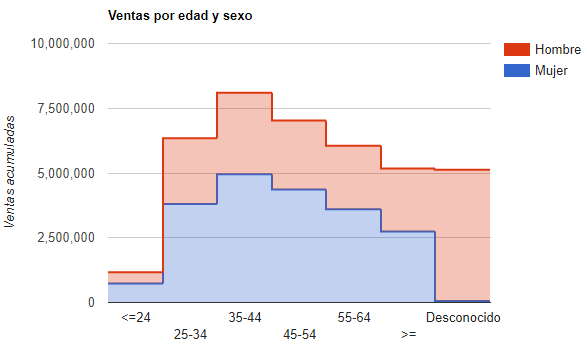
Este apartado incluye gráficas sobre el número de transacciones, el número de tarjetas que se han usado, el número de establecimientos en el que se hayan realizado al menos una transacción con tarjeta, y la compra media y la venta total de las transacciones realizadas con tarjetas dentro del código postal que estamos analizado. En este apartado las gráficas están desglosadas por rangos de edad y por el sexo de los clientes.

Todas estas gráficas se explican a continuación.

#### Ventas

Esta gráfica muestra la cantidad de ventas del año según la edad y el sexo del cliente. Para mostrar esta información se han realizado una serie de rangos para la edad y se mostrarán los datos de ambos sexos diferenciando cada uno por un color. Estos datos son muy útiles para que los empresarios puedan conocer que tipo de clientes son los que más compran en la zona y así poder hacer promociones u ofertas que beneficien al tipo de cliente que más compra o por el contrario para atraer a los tipos de clientes que menos compran y tratar de cambiar esta tendencia. También podría servir a la hora de abrir un nuevo negocio, ya que el empresario podría saber que tipo de clientes abundan más en la zona y ver si ese es el tipo de clientes que comprarían en su negocio o si debería irse a otra zona mejor.

Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Origin Distribution”.



#### Sexo

Con esta tabla el usuario puede ver los datos de número de transacciones, número de tarjetas, número de establecimientos, compra media y ventas del año separados según el sexo de los clientes. Estos datos, además de tener la misma utilidad que la gráfica anterior, permiten conocer cómo se comportan los clientes según su sexo, por ejemplo, se podría saber mirando el número de establecimientos que clientes suelen comprar en más tiendas diferentes y cuáles menos, o cuál es la compra media de cada uno de ellos.

Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Gender Distribution”.

La tabla dispone de cuatro valores para sexo diferentes:

* M: Masculino
* F: Femenino
* U: Desconocido
* E: Empresas

****

#### Edad

Con esta tabla el usuario puede ver los datos de número de transacciones, número de tarjetas, número de establecimientos, compra media y ventas del año separados según la edad de los clientes. La edad en esta tabla está dividida por intervalos de edad.

Estos datos, además de tener la misma utilidad que la gráfica de ventas por edad y sexo, permiten conocer cómo se comportan los clientes según su edad, por ejemplo, se podría saber mirando el número de establecimientos que clientes suelen comprar en más tiendas diferentes y cuáles menos o cuál es la compra media de cada uno de ellos.

Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Age Distribution”.



#### Ventas por Edad y por Sexo

En esta parte de la web se muestran la cantidad de ventas realizadas en la zona. A esta información se accede a través de dos apartados distintos en la web según como el usuario desee ver los datos, si los quiere desglosados según rangos de edad o, por el contrario, si los prefiere dividido por sexos. Para cada una de estas opciones habrá una gráfica para cada rango de edad, si se quiere desglosado por edad, o para cada sexo.

En este apartado los datos se muestran a través de una gráfica de columnas apiladas, en las que los colores representan a cada una de las categorías de negocio con una leyenda de las mismas en la parte superior de la gráfica. Además, si el usuario pasa el ratón por uno de esos trozos de la columna apilada se mostrará un pequeño cuadro en el que se indica el mes, la categoría y la cantidad de ventas, de este modo, es más fácil obtener estos datos.

Los datos utilizados para hacer esta gráfica se han obtenido de las llamadas “Age Distribution” y “Gender Distribution”.

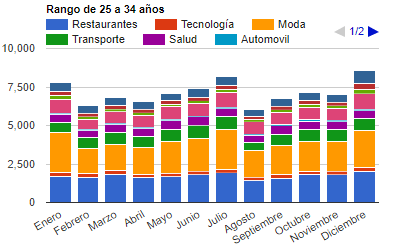
#### 

#### Tarjetas por Edad y por Sexo

En esta parte de la web se muestran la cantidad de tarjetas distintas que han sido usadas en la zona estudiada. A esta información se accede a través de dos apartados distintos en la web según como el usuario desee ver los datos, si los quiere desglosados según rangos de edad o, por el contrario, si los prefiere dividido por sexos. Para cada una de estas opciones habrá una gráfica para cada rango de edad, si se quiere desglosado por edad, o para cada sexo.

En este apartado los datos se muestran a través de una gráfica de columnas apiladas, en las que los colores representan a cada una de las categorías de negocio con una leyenda de las mismas en la parte superior de la gráfica. Además, si el usuario pasa el ratón por uno de esos trozos de la columna apilada se mostrará un pequeño cuadro en el que se indica el mes, la categoría y el número de tarjetas y, de este modo, sea más fácil obtener estos datos.

Los datos utilizados para hacer esta gráfica se han obtenido de las llamadas “Age Distribution” y “Gender Distribution”.

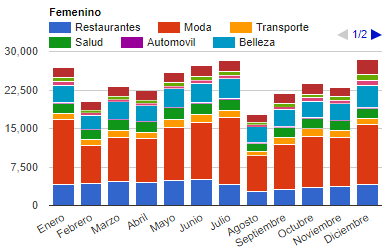


#### Transacciones por Edad y por Sexo

En esta parte de la web se muestran el número de transacciones a través de tarjetas que han sido realizadas en la zona estudiada. Como en el apartado anterior, el usuario puede elegir si quiere ver los datos desglosados según rangos de edad o si los prefiere dividido por sexos. Para cada una de estas opciones habrá una gráfica para cada rango de edad, si se quiere desglosado por edad, o para cada sexo.

Como en el apartado anterior, en esta parte de la web los datos se muestran a través de una gráfica de columnas apiladas, en las que los colores representan a cada una de las categorías de negocio. Además, si el usuario pasa el ratón por uno de esos trozos de la columna apilada se mostrará un pequeño cuadro en el que se indica el mes, la categoría y el número de transacciones.

Los datos utilizados para hacer esta gráfica se han obtenido de las llamadas “Age Distribution” y “Gender Distribution”.

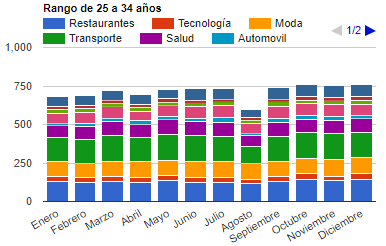
****

#### Establecimientos por Edad y por Sexo

En esta parte de la web se muestran el número de establecimientos que han realizado al menos una venta con tarjeta mostrando tanto el valor total como el de las distintas categorías de negocio en la zona estudiada. Como en los apartados anteriores, el usuario puede elegir si quiere ver los datos desglosados según rangos de edad o si los prefiere dividido por sexos. Para cada una de estas opciones habrá una gráfica para cada rango de edad, así como una para cada sexo.

Como en los apartados anteriores, en esta parte de la web los datos se muestran a través de una gráfica de columnas apiladas, en las que los colores representan a cada una de las categorías de negocio. Además, si el usuario pasa el ratón por uno de esos trozos de la columna apilada se mostrará un pequeño cuadro en el que se indica el mes, la categoría y el número de establecimientos que hayan realizado al menos una venta a través de una tarjeta.

Los datos utilizados para hacer esta gráfica se han obtenido de las llamadas “Age Distribution” y “Gender Distribution”.

****

### Análisis Localización

Este apartado incluye una gráfica que muestra el destino de las ventas realizadas con tarjetas dentro del código postal que estamos analizando, mostrando el porcentaje del total de las ventas que tienen como destino otros códigos postales. Además, también se incluye un mapa con datos de venta total y compra media de las transacciones realizadas con tarjetas dentro de un conjunto de ciudades.

La gráfica y el mapa que se incluyen en esta sección están explicados con mayor detalle a continuación.

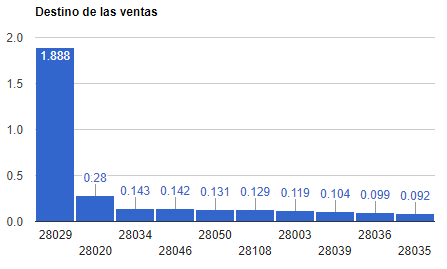
#### Destino por Código Postal

Con esta gráfica el usuario puede ver el destino que tienen las ventas que se realizan dentro de un código postal. En esta gráfica aparecen en el eje x los 10 códigos postales en los que mayor cantidad de ventas han sido destinadas con origen en el código postal elegido por el usuario, estos datos están representados en porcentaje sobre el total de las ventas.

A parte de estos datos,también se muestran a la derecha de la tabla el porcentaje de ventas de esa zona cuyo destino es un código postal dentro de la ciudad de Madrid, la Comunidad de Madrid o cuyo destino sea fuera de la Comunidad de Madrid.

Esta información podría ser útil para saber el código postal de los clientes más habituales, lo que podría ofrecer una oportunidad de negocio. Si, por ejemplo, hay un porcentaje alto de clientes que provienen de otro código postal, el usuario podría hacer un estudio para saber si sería beneficioso abrir un nuevo establecimiento en esa zona.

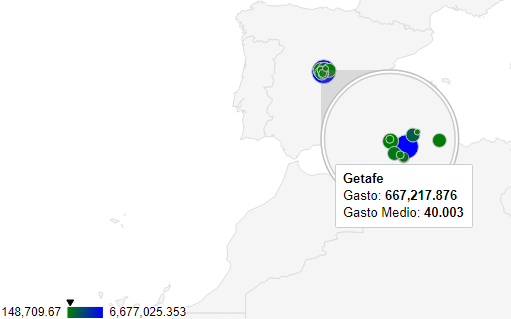
Los datos utilizados para hacer esta gráfica se han obtenido de la llamada “Destination Distribution”.



#### Gasto por región

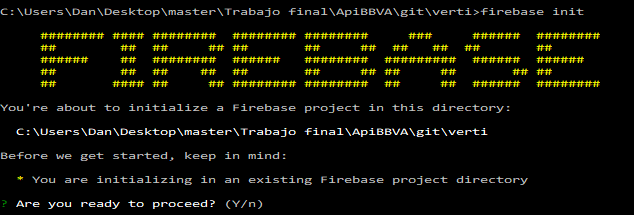
En este apartado se muestra un mapa en el que el usuario puede pasar el ratón por encima de una de las ciudades para mostrar un cuadro en el que aparecen la venta total y la compra media que se han realizado con tarjetas en esa ciudad.

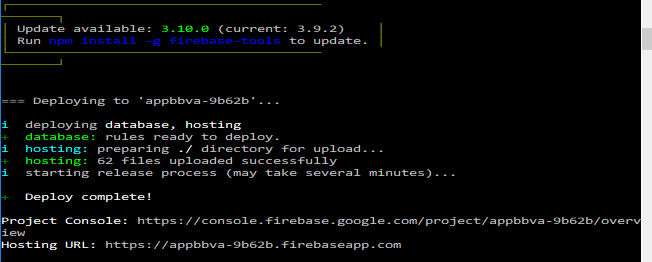
En este mapa aparecen las ciudades disponibles marcadas con un punto de color cuyo color va desde el verde al azul según el dato de la cantidad de ventas de cada ciudad.



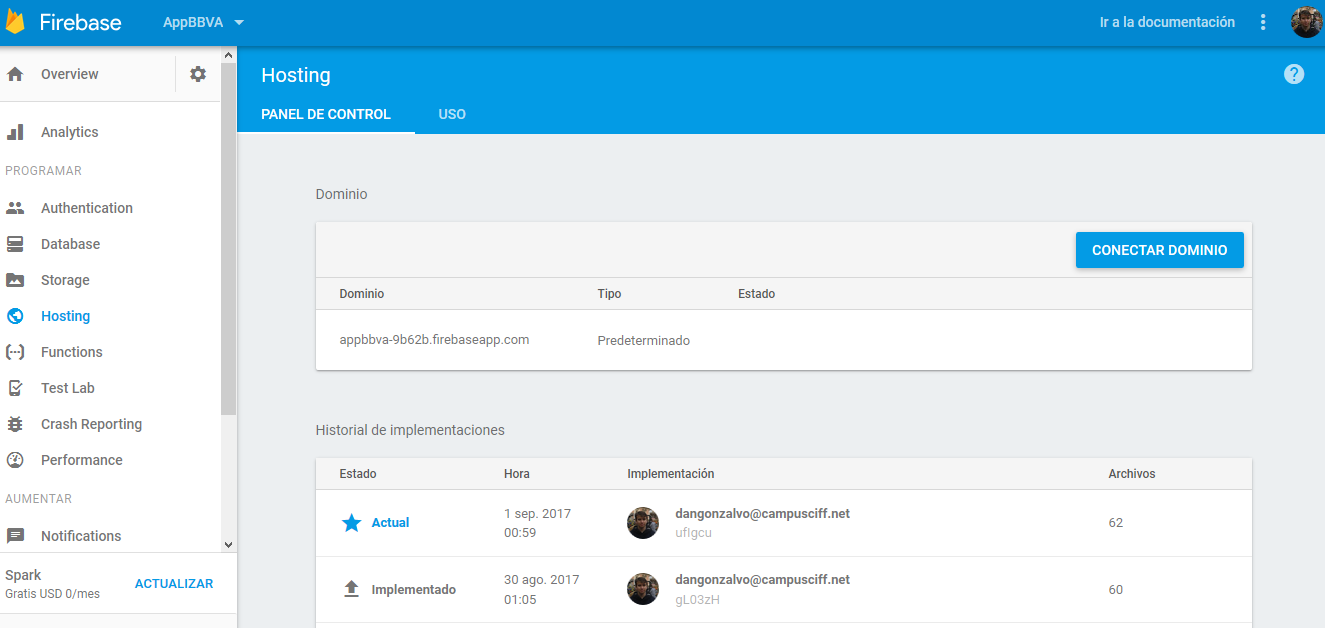
## Implementación Web

Todo el proyecto se ha desplegado gracias a Firebase. Una vez creado el proyecto podemos almacenar y hospedar la web en el entorno de Firebase.



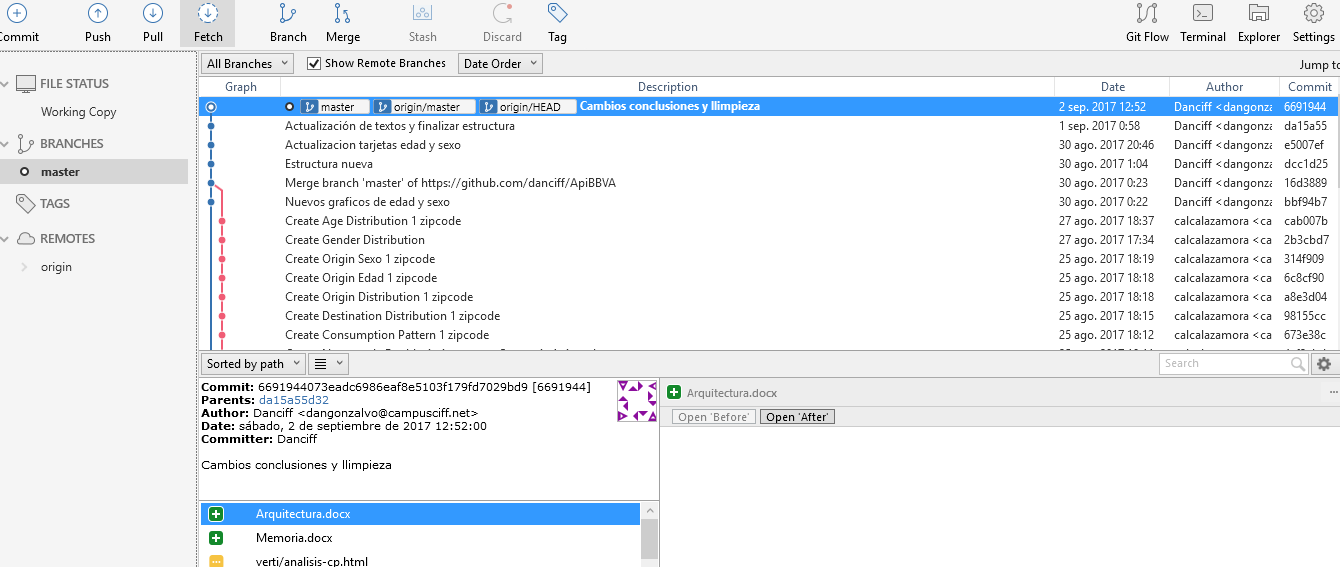


Nos permite desplegar nuestra web para la visualización de los resultados, y que esta sea accesible por cualquier usuario desde su navegador.

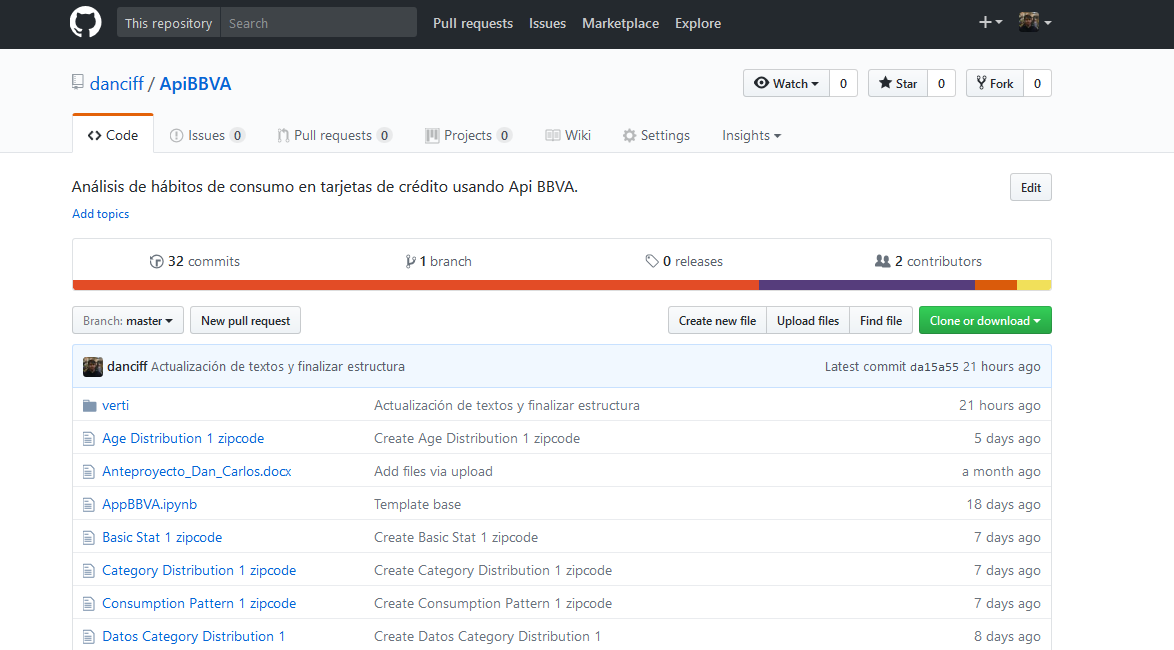


## Control de versiones

Todo el proyecto está disponible desde Github, todos los cambios y control de versiones se ha realizado durante el desarrollo del proyecto.



Se ha utilizado Git como repositorio de toda la codificación desarrollada y de la documentación necesaria: <https://github.com/danciff/ApiBBVA>



## Arquitectura Online

Paystats también nos ofrece la opción de tener acceso a todos los datos posteriores al año 2015 hasta la actualidad. Si se realiza un acuerdo comercial con BBVA podemos tener acceso a estos nuevos datos, y la información se actualizaría de forma semanal. En este caso habría que tener en cuenta esta actualización de los datos por lo que habría que implementar una arquitectura más ambiciosa.

Aunque no se trata de una actualización en tiempo real, sí que habría que hacer una actualización de forma periódica por lo que sería adecuado usar una arquitectura de tipo Kappa. Como en el desarrollo realizado la información la ofrece directamente Paystats por lo que no sería necesario un recolector de información, pero el procesamiento sí que sería necesario realizarlo de forma automática con cada actualización de los datos, y también tendríamos que conectar de forma directa la web con los resultados obtenidos.

Cassandra

(Almacenamiento de datos)

Spark

(Procesamiento de datos)

Paystats

(Origen de datos)

Web

(Visualizacion)

Consultas

En esta arquitectura haríamos uso directamente de la tecnología Spark, como sistema de procesamiento, gracias a PySpark podemos directamente migrar el desarrollo en Python realizado, lo que nos permitirá procesar los datos de Paystats, los resultados en lugar de ser ofrecidos en ficheros, tendrán que ser almacenados en una base de datos como Cassandra. Para recoger los datos podemos hacer uso de servicios Rest, que puedan ser leídos desde la web. Igualmente habría que modificar la web para recepcionar los datos desde servicios rest. Además habría que realizar el hosting no solo de la web sino del procesamiento y almacenamiento.

# Conclusiones.

Disponemos de **330000** clientes, de más de **1815000** transacciones, las cuales alcanzan los **20000000**€. Esto nos permite conocer el tipo de clientes de una zona e incluso sus patrones de comportamiento.

Por ejemplo, con las ventas se podría averiguar que sexo es el que más compra en una zona para hacer promociones u ofertas dedicadas sólo a ese sexo, o al contrario si lo que quiere el empresario es atraer a ese grupo al que todavía no logra vender lo suficiente e intentar cambiar esta tendencia, o con el número de establecimientos se podría saber qué tipo de cliente compra todo en unas pocas tiendas de confianza o grandes superficies, y por lo tanto tendrá un número bajo de establecimientos, o si por el contrario le gusta ir mirando por distintas tiendas para ver si hay algo que le guste, este último caso tendrá un número elevado de establecimientos y los clientes que estén en este grupo son más susceptibles a promociones o anuncios. En resumen, estas gráficas son muy útiles para ayudar a obtener el público objetivo para llevar a cabo una acción de marketing.

# Bibliografía.

* <https://www.bbvaapimarket.com/products/paystats>
* <https://bbvaopen4u.com/es/>
* <https://www.bbva.com/>
* <https://taiga.io/>
* <https://github.com/>
* <https://developers.google.com/chart/>
* <https://html5up.net/>
* <https://www.getpostman.com/>
* <http://jupyter.org/>